Ya. B. Zeldovich and I. M. Yaglom

Higher Math

for Beginners

Higher Math

for Beginners

Я. Б. Зельдович, И. М. Яглом

ВЫСШАЯ МАТЕМАТИКА ДЛЯ НАЧИНАЮЩИХ ФИЗИКОВ И ТЕХНИКОВ

Издательство «Наука» Москва

Ya. B. Zeldovich and I. M. Yaglom

Higher Math

for Beginners
(Mostly Physicists and Engineers)



Mir Publishers Moscow

Translated from the Russian
by Eugene Yankovsky

First published 1987
Revised from the 1982 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

© Издательство «Наука», 1982
© English translation, Mir Publishers, 1987

Foreword

This book is the joint attempt of a physicist and a mathematician to write an entirely new type of book for future scientists and engineers. This calls for a few words about its purpose and scientific ideology.

Many physicists, as is known, are displeased with the presentation of the basics of mathematical analysis that was introduced by mathematicians in the first half of the 20th century. The reasons are understandable. The development of mathematical research connected with the elaboration of the logical foundations of our science left its imprint on the style of presentation even of the very first chapters of textbooks for beginners. Exact definitions of such concepts as a real number, limit, and continuity were the result of a prolonged and very nontrivial logical analysis of theories that were already created and were intuitively clear (on the scientific level of rigor). These definitions, which are not at all simple for the beginner, came to be used in the wrong context. Textbooks presented them before any explanation was given of the theory and its applications, thereby complicating an understanding of things that were intuitively clear.

It must be borne in mind that theoretical physicists make extensive use in their research of all, or nearly all, the methods of classical mathematics. They often have occasion to employ the most modern ideas that have sprung from the depths of abstract mathematics or even to devise their own new mathematical methods. But the ultimate result of the investigations of a physicist must always be effective. It must be expressed as a number or formula relating to the quantities observed. The physicist feels it is unimportant to justify and establish the logical structure of the foundations: he has faith in mathematics and does not doubt that its methods contain no internal contradictions. Theoretical physicists are not satisfied with the existing textbooks and popular scientific literature on mathematics written by mathematicians themselves (especially by those who are far removed from physics) they feel it necessary to make known to the beginner their own concepts of how a scientist should use the mathematical apparatus and in what simple way it is possible to grasp, "to a first approximation," the methods he will have need of. It seems to me that their tremendous mathematical experience gives them this right.

Of course, such a view of mathematics is somewhat one-sided, but surely there is no harm in that. Those who go on deeper into mathematics will be able better to grasp the meaning of the foundations of mathematics since they

will already have an intuitive understanding of the essence of the matter and will have developed the techniques of problem solving. As I see it, many courses in mathematics are spoiled by the desire to start the teaching of a new theory by confronting the student with the foundations and the formal language. As a mathematician who has worked with physicists for many years, I take the liberty to express the opinion that it would be advisable not only for future physicists and engineers but also for future professional mathematicians (and their teachers) to learn to understand better the style of mathematical thinking of the natural scientist and to grasp the problems that confront him.

This text is the result of a great effort to improve the book *Higher Mathematics for Beginners and Its Application to Physics* by the well-known Soviet physicist Academician Ya. B. Zeldovich. The purpose of this book is to enable the future physicist (chemist, engineer, etc.) to use higher mathematics in his or her work by mastering its methods without going into a full logical substantiation of them, allowing the student to view mathematics as a section of natural sciences and to solve as many concrete problems as possible.

The previous book had a number of technical drawbacks. Much has been done in this book to eliminate the shortcomings and improve the presentation. Considerable new material has been added, and the text has been carefully edited. All this work was done jointly by the two authors: the physicist Academician Ya. B. Zeldovich and the mathematician I. M. Yaglom, Doctor of Physico-Mathematical Sciences.

It seems to me that the book achieves its objective. I would confidently recommend it to the thinking young reader who wishes to master the methods of higher mathematics informally and to apply them to problems in physics and technology.

Academician S. P. Novikov

Note to the Reader

There are probably not many people who have never heard of Pushkin or Leo Tolstoy. But very many grown-up people, one-half or even nine-tenths, would not be able to explain what a *derivative* or an *integral* is. Yet without these concepts it is impossible to describe or investigate the variables or functions that characterize the dependence of quantities on one another. The laws of nature are formulated in the language of higher mathematics, the language of derivatives and integrals.

It's our view, though we are not impartial, that higher mathematics is as beautiful as the poetry of Pushkin and as profound as the prose of Dostoevsky or Tolstoy. But is higher mathematics seen by the general public as a cultural asset? Has it really entered the public consciousness as such? Now that modern science has greatly extended the boundaries of the visible world by discovering the secrets of the structure of the atom and the many mysteries of the starry sky, the concepts of derivative and integral have become necessary elements of general culture. And in everyday life, too, it is quite useful to understand the rate of change of a quantity (the derivative) and the overall effect of the action of a factor (the integral). This expands one's outlook and can be used in a great variety of situations.

The traditional methods of teaching higher mathematics have complicated matters. It would seem simple to understand "Crime and Punishment" or "War and Peace" and to grasp the sphericity of the earth or the atomic structure of matter, yet difficult to master differentiation or integration. Students of the first half of the 20th century were terrified by the theory of limits and the language of infinitesimals, which is so unlike ordinary arithmetic and school algebra. When higher mathematics was introduced in the traditional way in the first year of study at institutions of higher learning, it proved to be the main subject that caused students to drop out. The only way to change the attitude of the student toward this subject is to change the way it is taught. Higher mathematics must be transformed from a dry and difficult subject into a set of clear and natural concepts that open the way to the study of physics, chemistry, and engineering disciplines.

The first problem confronting the authors of this book was to give the student an understandable introduction to higher mathematics unencumbered by an unwieldy apparatus or logical subtleties. We believe that anyone acquainted with the basics of arithmetic and who had above-average marks in algebra in secondary school will find it easy (and, we hope, interesting) to read this

book. We seek for a friendly student who has no doubts but believes and wants to take up this book with a view to learning something new, undeterred by the fact that at times the authors offer simple examples in place of "very sientific" theory.

The book contains many concrete examples (perhaps even too many for some). There are calculations and problems connected with natural phenomena. We hope that these examples will make it easier for the student in his studies. If we had expected the reader to be given to arguing, we would have written it in a different way. Perhaps for such a reader the standard presentation with all the usual reservations would have been more suitable, but we have no wish to lose a hundred possible friends just because of a single one who likes to argue.

This book is intended for beginners, that is, for high-school students in the upper grades, students of trade schools and vocational schools, and students in the first years of college. We also have in mind anyone who by himself wishes to become better acquainted with higher mathematics, say, people who finished school some time ago.

The place of mathematics in life and in science is determined by the fact that it permits one to translate an everyday intuitive approach to reality based on purely qualitative and hence approximate descriptions into the language of exact definitions and formulas from which quantitative conclusions can be drawn. So true it is that the scientific level of any discipline is determined by the extent to which it uses mathematics. But the real language of science is not at all elementary algebra or geometry. Higher mathematics plays a far greater role. It studies variables and processes for which algebra or geometry do not suffice. It also investigates other more complicated divisions of mathematics which are only touched on in this book. It is no accident that the development of differential and integral calculus by Newton was intimately bound up with his development of the foundations of theoretical (we can also say mathematical) physics. He saw the concepts of higher mathematics as a (literal!) translation of the basic concepts of mechanics into mathematical language.[1]

The first chapters of this book give an idea both of the essence of higher mathematics and of its possible applications. So we can imagine a reader who wishes to restrict himself to a selective study of Chapters 1 to 7. But it was not for him that we wrote this book. To grasp the basics of higher mathematics it is necessary to have a good understanding of how this apparatus is used. Knowledge not used is not real knowledge. It is grasped with difficulty and easily forgotten. This book is intended for people especially interested in physics and engineering. For this reason the applications are largely taken from these areas. It would be possible today to compile a textbook of higher mathematics for a biologist (or future biologist), geologist, or economist, but that would be an entirely different book and other people must write it.

The second part of this book deals with a number of topics from physics and engineering. In these areas the power of higher mathematics shows itself to the full. A great many highly important physical phenomena can be described with a degree of completeness that is unattainable without the use of a derivative or an integral. By way of illustration we can mention the theory of *radioactivity* (Chapter 8). Another example is the phenomenon of *resonance*, which in this book appears in two different forms, as one of the effects appear-

[1] It is also no accident that the other inventor of higher mathematics, Leibniz, worked out evolutionary conceptions that were closely linked to his mathematical ideas and were considerably ahead of their time. These conceptions touched on biology, geology, and physics; for instance, Leibniz conceived of the concept of energy, a quantity that transforms from one form to another but never vanishes in physical phenomena.

ing in the theory of mechanical oscillations (Chapter 10), and as one of the properties of electric circuits (Chapter 13). Although some of the sections of physics are elaborated in Part 2 in great detail and others are only outlined or even omitted altogether, the amount of information about physics presented here is so great that as a whole the book can be viewed not only as a text in mathematics but also as an important addition to the ordinary physics textbook.

Chapters 8 to 13 (Part 2 of the book) are devoted to physical problems and are almost independent of each other. They can be read in any order. The only exception is Chapter 10, which is devoted to oscillations. To understand this chapter it is necessary to read Sections 9.1 to 9.6 of the chapter devoted to mechanics. Although Part 1 of the book, called "Elements of Higher Mathematics," contains many examples from physics and mechanics, various sections of Part 2, called "Higher Math Applied to Problems in Physics and Engineering," will be of interest to *all* categories of readers.

The large size of this book would seem to contradict the word "beginners" in the title. But don't be afraid of the number of pages. We are well aware of all the shortcomings (and also of the merits) of the detailed and frankly rather wordy system of presentation that we have chosen. The book contains a great many examples and calculations. Some problems are elaborated several times and from different angles. Much space is devoted to the history of the theories presented. All the information contained in this book could, of course, be compressed and presented on far fewer pages. But then it would require a greater effort to read the book, and anything omitted or incorrectly understood would hamper further progress (or even make it impossible to proceed). Of course, there are situations in life when the best source of information is a brief telegram, but there is also charm in the art of unhurried presentation with many seemingly unnecessary details, which was so popular in the 18th and 19th centuries and which has almost been forgotten. We think the telegraph style is not at all suitable for the beginner. The extra information, which in some cases can be ignored, is not burdensome, whereas insufficient information or its extreme concentration in a book for the inexperienced reader is quite impermissible.

It is also necessary to bear in mind that this book contains considerable material for persons interested in an in-depth study of the basics of higher mathematics. During the first reading much of this material can be ignored. Say, all the small print, including the story about the origin of higher mathematics, and the starred sections. The same goes for the entire third part of the book ("Some Additional Topics") and the Conclusion ("What Next?"). And, finally, we have already emphasized the possibility of a selective acquaintance with the second (physics) part of the book (Chapters 8 to 13).

The large first chapter ("Functions and Graphs") mainly contains material familiar from school. A quick run-through should suffice for most readers. Chapters 2 and 3 present the foundation ideas of higher mathematics (differential and integral calculus), while Chapters 4 and 5 elaborate the basic techniques of higher mathematics sufficient for elementary applications. Chapter 6 is devoted to certain more advanced areas of mathematical analysis (series and differential equations), and Chapter 7 is concerned with purely mathematical applications of the concepts of a derivative and an integral (physical applications are considered in Part 2 of the book). Elements of the history of higher mathematics are found in the concluding sections of Chapters 3 and 6. These are separated from the main text by three asterisks. And finally the Conclusion ("What Next?"), which discusses some of the newer divisions of mathematical science now widely used in applications.

The third part of the book (notably the Conclusion) is of a slightly different character than the first and second. This last part is devoted to two trends in higher mathematics that considerably expand (and generalize) the classical differential and integral calculus of Newton and Leibniz. The first is *complex numbers* and *functions of a complex variable*, which have long played a major role in physics and engineering. It is indeed remarkable that many facts connected with ordinary real functions of real arguments become clearer when we move into the complex domain. The reader will find some examples of this in Chapters 14, 15, and 17. The second trend involves *generalized functions* (or distributions) $f(x)$, and first of all the Dirac delta function, which is always equal to zero except when x equals zero, in which case it is equal to infinity (!). Such functions also play a big role in physics and engineering. In fact, their origin is much more closely connected with physics than with mathematics. We consider it most desirable that future physicists and engineers become acquainted with these remarkable functions as early as possible.

Chapter 14 to 17, which are devoted to functions of a complex variable and to generalized functions, are intended for the especially interested reader. They are written in more of an outline form than the rest of the book (although they can be understood by a sufficiently persevering beginner). They can be read in greater or lesser detail, as one wishes.

The Conclusion ("What Next?") is an even more general survey. We do not expect the reader to grasp the material fully. It is only intended to arouse the reader's curiosity and encourage further work. It would be naive to think that the topics embraced by this book can complete the mathematical education of a physicist or engineer. In the Conclusion we touch on certain other trends in science that found wide application in physics and technology only in the second half of this century.[2] Although the Conclusion and, in fact, the entire third part of the book can easily be omitted (it is not needed for an understanding of the main part of the book), we believe that most readers will wish, at least cursorily, to familiarize themselves with these, since they reveal prospects for the future and explain certain ideas that are important to and characteristic of modern science.

In short, this book can be used in a variety of ways. The reader interested largely in mathematics will naturally devote most of his attention to Part 1. He will undoubtedly find it useful, however, to familiarize himself with various sections of Part 2, say, Chapter 8 or the first portions of Chapter 9 and some of Chapter 10, or with Chapter 13. Anyone who wishes to deepen his knowledge in mathematics is advised to turn to the concluding sections of Chapter 10 and/or the third part of the book. The reader interested in the functions of a complex variable and the delta functions should not ignore Chapter 17, which is devoted to the applications of the corresponding theories. Additional information about mathematics can be derived from the Conclusion, but for an in-depth grasp of the topics tuched on in the Conclusion it is necessary to turn to other literature. Finally, a reader interested in physics more than in mathematics need only run through the first part of the book, ignoring the text in small print and omitting most of the starred sections (but not Section 6.7, which shows how differential equations can be used to analyze phenomena in the natural sciences). Such a reader need not dwell long on Chapter 7, which is purely mathematical, but he should give careful study to Chapter 6 because its applied significance is great.

---

[2] Here we draw particular attention to the theory of groups, which mathematicians developed in the 19th century and which at that time was considered far removed from any physics. At present the physics of elementary particles owes its achievements largely to the theory of groups.

The reader interested in physics will probably make a careful study of Part 2 (and perhaps also Part 3, which is likewise connected with profound physical theories). He can study Chapters 8 to 13 in any order, though in some respects it is best not to allow for too big a break in time between Chapter 10 and Chapter 13.

The list of literature given at the end of the book includes certain other publications that outline the basics of higher mathematics, also textbooks that cover more material, and books that deal with topics that go far beyond the material presented here. As a natural continuation of this book we recommend [15].[3] On the other hand, the books [19], [21] are far removed from the main content of this book: they are devoted to branches in mathematical science that relate to trends mentioned in the Conclusion.

In the literature on physics we especially recommend an excellent three-volume course [34] (there are also problems and exercises relating to the whole course). In books [22-30, 37] higher mathematics is not used.

In Appendices 1 to 4 at the end of the book the reader will find short tables of derivatives, integrals, and numerical series (which are discussed in Part 1) and also extracts from numerical tables, which, in a sense, round out the book. They will save the reader the trouble of turning to other textbooks when doing the exercises that complete each section of the book. The physics examples are based on the now accepted international system of units (SI). The reader insufficiently acquainted with this system will find Appendix 5 helpful. Finally, the concluding Appendix 6 presents the Greek alphabet, which is widely used in this book.

Like any book on mathematics, this one presupposes an active reader (one with pencil and paper) willing to work the many exercises that are given here. Better still if you have a pocket calculator or a programmable calculator. Then you will be able to verify and substantiate the general theorems with calculations and make freer use of the method of "arithmetical experimentation," that is, approximate computations that reveal the content of the concepts introduced and the meaning of the formulas. The many specific examples will help you to evaluate the exactitude of the approximate formulas scattered throughout the book, and you will better appreciate the wisdom of our forerunners who created methods and concepts without modern computer techniques. At the same time you will feel the new power which we children of the 20th century derive from computers.

Ya. B. Zeldovich
I. M. Yaglom

[3] Students of technical colleges who have studied our book in the first year are advised, in the second year, to turn to [15].

Teaching Notes

This book is intended not only for students but also for teachers who, while familiar with the subject matter, might be interested in the methods that we recommend for teaching higher mathematics to future physicists and technicians. Our textbook is not designed for any particular category of students: it can be used in teaching mathematics or physics in high school, technical school, vocational school, or college. High-school teachers will find it helpful in refreshing their ideas, obtained at college, about the relationships and interconnections between mathematics and physics. They can use it in elective classes or in physics and mathematics study circles; some parts of it can be conveyed to students directly in the classroom. The college or university instructor can use the material in practical classes in the calculus course (higher mathematics) or in the general physics course. Our book will also be useful to instructors during work on exercises in differential and integral calculus with first-year students of technical colleges or in the physics department, even if a different trend predominates in the lecture course in higher mathematics. Work on topics in Part 2 of the book (possibly Part 3 as well) will help students to get a balanced idea of mathematics and facilitate assimilation of certain parts of physics and technical subjects (the theory of oscillations, for instance). Most important of all, of course, is the fact that this book will help them not only to grasp the substance of differential and integral calculus but also to realize how and why calculus arose and why this key subject of mathematics is needed.

The Selected Readings at the back of the book are intended for both students and teachers. Some of the books listed [4-6, 12-14] demonstrate possible ways to achieve an integrated approach to the teaching of higher mathematics and its applications.

This book consists of three parts, the first of which (Chapters 1 to 7) chiefly reminds readers of facts learned at school. Here we set forth the ideas and methods of higher mathematics. Part 2 (Chapters 8 to 13) deals with applications of higher mathematics in physics and engineering; it is a fundamental part of the book, unlike Part 3 (Chapters 14 to 17), which is supplementary and can, in principle, be skipped. The inseparable connection between the first part of the book, devoted to mathematics, and the second part, dealing with physics, is underlined by the fact that, for instance, the section on Fourier series is included in Part 2 (in the chapter on oscillations), while the process of water flow is studied in Chapter 6 de-

voted to series and differential equations.

In writing this text we did not specifically take into account the possible use by present-day students of personal computers. Naturally, these would further simplify the teacher's explanations. For one thing, they make "numerical experiments," such as the one in Section 4.7, much easier and faster (and also, therefore, more natural). They enable students to employ numerical (approximate) differentiation and integration on a much broader scale. Computers should, naturally, also be taken into account when teaching students the applications of mathematics.

The present volume is linked up with a book by one of the authors (Zeldovich). It came out in English in 1973 (*Higher Mathematics for Beginners and Its Application to Physics*, Mir Publishers, Moscow).[1] One of the reasons for writting it was that in those years differential and integral calculus was not even mentioned in the Soviet Union's high schools: the analytical section of the school mathematics course was limited to the fundamentals of algebra and of trigonometry, then still a separate subject. As for the college course, this taught the fundamentals of calculus according to a system far removed from the needs of physicists and technicians. We remind our reader that the main college textbook in higher mathematics in the Soviet Union throughout the 1930s was V. Grenville's well-known book, which had been translated into Russian and then later revised by Academician N. Luzin. Revised editions of this simple book went through a large number of printings and were highly popular among engineering students. However, the harsh criticism leveled at the textbook by "pure" mathematicians, who accused it of being slipshod, led to its complete replacement by textbooks of an altogether different trend. In the 1950s it became traditional to open the calculus course with the theory of limits based on a refined delta-epsilon technique using profound concepts whose value was at first only stated but not explained. All this created a situation in which the concepts of the derivative and integral, intuitively fairly clear and simple, seemed, to the uninitiated, to be something very profound and aroused a kind of mystical awe (in the case of first-year students there was also the real fear of not passing the exam). Indeed, many mathematicians, it is to be regretted, promoted that feeling by stressing the logical subtlety of the concept of limits and the need to use as a basis one of the variations of the theory of real numbers (though it is common knowledge that substantiations of this kind, of necessity fairly complicated, appeared only in the second half of the 19th century, that is, two centuries after the birth of higher mathematics). The students' logical criteria were developed on the basis of theorems that seemed clear to them without any proof. For instance, the fact that a variable can have only one limit[2] was traditionally established without any regard for the intuitive idea of a limit, which clearly tells us that if a variable moves back and forth between two "limits", it does not tend to either of them.

The situation today is indisputably better. A course in the rudiments of differential and integral calculus is obligatory in high schools in the Soviet Union. And among the new college and university textbooks there are some based on an approach similar to ours. For example, the author of the textbook [9], which has gone through a number of editions, does not attempt to prove a single theorem on limits, preferring, instead, to replace strict logic

[1] The first Russian edition came out in 1960, while the fifth and last, from which the English translation was done, came out in 1970.

[2] Compare this with a theorem from a present-day American school textbook on geometry: E.E. Moise and F.L. Downes, Jr., *Geometry*, 2nd ed., Addison-Wesley, Menlo Park, Cal., 1971: p. 51, Theorem 2-3: "every segment has exactly one midpoint." This is obviously intended only for future mathematicians.

with a frank appeal to common sense. And even the textbooks which traditionally open with an exposition of the theory of limits are read in an entirely different way by students who learned the fundamentals of the derivative and the integral while still in high school.

Nevertheless, we feel it necessary to detail the motivation behind our method, especially since this method undoubtedly does not meet with general approval. Although elements of higher mathematics have won a firm place in high-school mathematics, the question of how these elements should be presented is far from settled. An "engineering" approach or a "natural science" approach is quite possible here, but some take a fundamentally different view. Nor can it be said that the principles underlying the mathematics course in technical and vocational schools are clear and definite, though we believe that a more rigorous treatment than the one applied in this book is hardly needed for these schools.

Clearly, before starting to plan a system of teaching mathematics we must give thought to the reasons why we are teaching it. The answer to "how" can only be arrived at after we answer the question "why." There can be the following reasons for studying mathematics:

(1) to pass an exam;

(2) to develop one's capacity for abstract and logical thinking;

(3) to learn to use computers;

(4) to apply it in studying engineering, physics, biology, the microcosm of atoms and elementary particles, or the macrocosm of stars and galaxies.

These aims differ greatly, and they necessitate completely different instructional methods and systems. Indisputably practical though it is, the first of the above-listed aims can of course be dismissed at once. It is clear that the examiner must meet the needs of the students rather than the other way around (although, unfortunately, this is not always the case). The second aim cannot and should not be wholly ignored. Since it is connected with the training of future mathematicians, the growing importance of mathematics compels us to pay due attention to this point. But it would be absurd to maintain that teaching shaped by aim (2) should be greatly expanded; after all, mathematicians will never comprise a significant part of the population. From this point of view the trend in schools in many countries to teach the new math, understood as the fundamentals of symbolic logic and of the general theory of sets and the doctrine of abstractly introduced functions and so-called binary relations, strikes us as harmful. Nor can aim (3) be the main guideline in planning large-scale mathematics instruction; the training of future programmers and other computer experts is important, yet for the most part it is only a special case.

Our book is meant for future engineers and natural scientists who want to know mathematics for aim (4). In our opinion, this undoubtedly large category of students should, as quickly as possible, be taught the information they need, without cluttering it up with superfluous logical subtleties or striving for maximum scope or a completely rigorous approach. For them, after all, mathematics will always be a tool, a language with which to describe phenomena, and not something in which they are interested for itself. Moreover, we feel that, as was the case in the 19th century and will probably be the case in the 21st century too, the main stress should be put on the elements of mathematical analysis, differential and integral calculus, and the simplest differential equations.

The approach to the main concepts of analysis should proceed from the idea that Nature is arranged fairly simply, that "God may be subtle, but He is not malicious," as Einstein put it. Hypertrophy of the theorems of existence can only discourage students who are endowed with a healthy feeling for physics. It is clear that a moving particle has a velocity, and that a curved line

has a tangent; to demonstrate the existence of velocity is just as unnecessary as it is to demonstrate that a flat figure has an area. Just take some paper or cardboard, cut out a figure and weight it, divide its mass by the mass per square meter of the material, and you have the area. In other words, in the initial stage of the study of analysis the existence theorems of the derivative and integral are not needed (in fact, they are harmful): both of these concepts have a natural physical meaning, and hence they exist.[3] It would be absurd to start learning grammar by studying the exceptions instead of the general rules, or to start a language course by studying the grammar instead of first building up a vocabulary. Yet that is the same, unfortunately, as the fairly widespread system of learning the "mathematical language" in which the concept (and properties) of the limit precedes any substantive application of this concept.

A fundamental law of biology is that "ontogeny recapitulates philogeny," that is, the development of an individual repeats, in one degree or another, the evolution of the entire group. For instance, at a certain period of its growth the human embryo has gill slits, because life originated in the sea and the ancestors of mammals, including man, were fish. We feel that teachers, too, should remember this law of biology. The future physicist or engineer should first study mathematics in the form in which it was created by the great scientists of the 17th and 18th centuries instead of receiving it immediately in the form that had been worked out by the second half of the 20th century after a long period of evolution.

This, it goes without saying, should not be understood as a call to reproduce exactly, in present-day textbooks, the fairly archaic reasoning of Newton, with its long-forgotten terms and symbols, or of Leibniz, with his characteristic "metaphysics of infinitesimals." However, the ideas of a course for beginners could, it seems to us, be based on familiar classical graphic ideas instead of on the modern theories of real numbers or on strict definitions developed over the centuries. Of course, there comes a time when the student has to be told about possible complications, but not about exotic "continuous functions that do not have any derivatives anywhere,"[4] but simply about the possibilities of a gap or a break in the curve. However, we believe such warnings should be given not before but *after* the student has intuitively assimilated the basic ideas, just as it would be absurd to *begin* the study of a new mechanism by learning the instructions about a possible breakdown. Furthermore, we are inclined to believe that a far more instructive (and informative) approach than the traditional nihilistic is the modern "constructive" approach to peculiarities of the kind under consideration, which declares that every continuous function (even if it's not smooth) has a derivative which, however, may prove to be a delta function instead of an ordinary function; we devote much attention to this approach in Chapter 16 of the book. We also take advantage of every opportunity to stress such important general ideas as *linearity* and the *principle of superposition*, which are sometimes impermissibly left in the shade in books

[3] When students of the famous William Thomson, Lord Kelvin (1824-1907), tried to determine the derivative by the Cauchy procedure, he grew irritated. "Oh, forget Totgentner", he said. "The derivative is just the velocity." (Totgentner was a professor of "pure" mathematics who in Kelvin's time taught a course in calculus at Cambridge.)

[4] Here we are inclined to concur with the French mathematician Charles Hermite, one of the most prominent analysts of the second half of the 19th century, who said, in a letter to his friend T.J. Stieltjes, the Dutch analyst, that "I turn away with fright and horror from this lamentable evil of functions which do not have derivatives." A book by the French mathematician B. Mandelbrot (*Fractals*, W.H. Freeman, San Francisco, 1977), which takes a diametrically opposite stand, is highly interesting but clearly intended for far more sophisticated readers than those for whom our elementary textbook is designed.

intended for physicists and engineers or students training for these fields.

Mathematics is developing by leaps and bounds, and so is the body of knowledge needed by those who use it. Physicists and engineers have to be taught more today than ever before. Some of the new scientific disciplines are named in the Conclusion entitled "What Next?". Computational mathematics, based on the enormous potentials of the electronic computer, is hardly touched on in this book, but it too is acquiring tremendous importance. We think it quite possible to acquaint beginners more quickly with the initial sections of higher mathematics so that they can apply their knowledge immediately and have more time to learn new substantive mathematical theories.

The pedagogical approach taken by the authors of many familiar calculus textbooks is reminiscent of the debates conducted by medieval Scholastics. Such authors regard the student as an experienced opponent eager to find weak points in the views of the teacher and they believe that the work of the teacher is to refute all possible objections. We, on the contrary, regard the student as a friend and ally who is ready to believe the teacher or the textbook and whose primary concern is to be able as quickly as possible to employ, in the study of nature and technology, the new methods learned. It is chiefly through an analysis of many examples and applications that the student comes to understand the substance of the new tools and gains the intuition that forewarns when the tools refuse to function. The strictly logical approach bypasses the question of the origin, significance, and benefit of the concepts and theorems under study. This book focuses mainly on "rough" mathematical ideas and their connections with natural phenomena.

We find support for our positions in the writings of many prominent scientists. The famous Russian naval architect Aleksei Krylov (1863-1945), Member of the Academy of Sciences, outstanding scientist, technician and engineer whose work and achievements covered an exceptionally broad range, author of the first Russian book on numerical methods in mathematics, translator of Newton's *Mathematical Principles of Natural Philosophy*, and an authority on celestial mechanics, wrote the following: "That on which Newton based the whole of the modern theory of the Universe and his incontrovertible proofs of the structure of the system of the world cannot be regarded as insufficiently rigorous for a 16-year-old high-school student." Krylov pointed out that beginners often accept the refined rigor of mathematical proofs as "a triumph of science over common sense." In his well-known *Autobiographical Notes* Albert Einstein, one of the greatest physicists of all times, wrote:

> At the age of 12-16 I familiarized myself with the elements of mathematics together with the principles of differential and integral calculus. In doing so I had the good fortune of hitting upon books which were not too particular in their logical rigor, but which made up for this by permitting the main thoughts to stand out clearly and synoptically. This occupation was, on the whole, truly fascinating; climaxes were reached whose impression could easily compete with that of elementary geometry—the basic idea of analytical geometry, the infinite series, the concepts of the differential and the integral....

A similar view was held by the distinguished Soviet theoretical physicist Lev Landau (1908-1968), Lenin Prize and Nobel Prize winner, Member of the USSR Academy of Sciences, a man who combined outstanding intuition in physics with a high level of mathematical technique. When asked for his opinion about a mathematics syllabus for a Moscow physics institute, he wrote:

> Unfortunately, your syllabus suffers from the same shortcomings that are common to the mathematics syllabuses that turn half of the study of mathe-

matics by physicists into a tiring waste of time. Important as mathematics is to physicists, what they need, as we know, is a calculating analytical mathematics; however, for reasons I cannot understand, mathematicians force logical exercises on us as an obligatory part of the course.... I think it is high time to teach physicists what they themselves think they need instead of trying to save their souls against their own wishes. I have no desire to discuss the idea worthy of medieval scholasticism that people learn to think logically by studying things they do not need. I most emphatically maintain that all theorems of existence, highly rigorous proofs, etc., should be completely expelled from the mathematics studied by physicists.

These differences of opinion between mathematicians and physicists led to a paradoxical situation in which three (!) Nobel Prize winners in physics and chemistry (Walther Nernst, Svante Arrhenius, and Hendrik Antoon Lorentz) found it necessary to write elementary textbooks in higher mathematics "for natural scientists and physicians," as the title of the book by Arrhenius says, or "for natural scientists, with special attention paid to the needs of chemists," as Nernst pointed out to his readers. The best known of these books is the *Course in Differential and Integral Calculus, With Applications in Natural Science* by H. A. Lorentz, which went through several editions between 1901 and 1926 and was still highly popular in the 1930s as a textbook for future engineers and also for self-instruction. The ideas in the present book are fairly similar to those set forth in all the above-named books.

It is worth noting that many "pure" mathematicians who do not specialize in abstract algebra or mathematical logic and have a taste for applications think along the same lines as the physicists mentioned in the above paragraph. For instance, the famous German (and later American) mathematician Richard Courant (1888-1972), founder of New York University's Courant Institute of Mathematical Sciences, wrote in 1964 that mathematicians had for a very long time accepted Euclidean geometry as a model of a strictly logical approach, of strict logical deduction. But here is what he says further (*Scientific American*, September 1964, p. 43):[5]

> But emphasis on this [axiomatic, logical] aspect of mathematics is totally misleading if it suggests that construction, imaginative induction and combination and the elusive mental process called intuition play a secondary role in productive mathematical activity or genuine understanding. In mathematical education, it is true, the deductive method starting from seemingly dogmatic axioms provides a shortcut for covering a large territory. But the constructive Socratic method that proceeds from the particular to the general and eschews dogmatic compulsion leads the way more surely to independent productive thinking.

All the above provides a sufficiently full explanation of the basic guidelines followed in our textbook. Its principles and underlying propositions coincide with those typical of the textbook *Higher Mathematics for Beginners and Its Application to Physics*; however, the details of the exposition differ substantially from the old one, even in the sections that were part of the above-named book. There are many new topics here; this refers both to a number of particular points (the second differential of a function of one variable or the total differential of a function of two variables; the idea of inertial reference frames) and to entire large areas of science (Fourier series, the theory of functions of a complex variable, or the design of a laser).

When using this textbook, the teacher can freely draw on material from the physics sections to illustrate mathematical concepts and theorems or, conversely, can include separate mathematics topics in physics lessons. For example, in the text we illustrate the concept of the derivative by the connection between the path traversed and

[5] Another excellent book, G. Polya's *Mathematics and Plausible Reasoning*, 2 vols, (Princeton Univ. Press, Princeton, N.Y., 1954), follows the same approach.

the speed, or between the quantity of heat needed to heat a body and the body's heat capacity; but nothing prevents the teacher from referring here to, say, the connection between the electromotive force and the current intensity for a coil (or for a capacitor, where the relationship is the reverse; cf. Section 13.1). It goes without saying that the number of such examples can be increased indefinitely. We believe that integrated teaching of calculus and its applications is very important for future technicians and physicists because a mathematical technique acquires deep meaning only in its application. The physics sections of the book are doubly beneficial: they not only give the student important information in physics but shed additional light on the mathematical constructions in the initial chapters and help the student to gain a new understanding of how natural and necessary these constructions are. During instruction, part of this material can be omitted of course; however, it seems to us that some use of physical considerations in the exercises in mathematics (and partly in the lectures as well) is very useful, and that a well-thought-out coordination of the elementary courses in mathematics and physics is absolutely essential.

In conclusion we would like to call the attention of the teacher (and student) to textbook [11], which is close to our ideas in spirit but is more mathematical in content. The modern "computerization era" makes it necessary to mention another excellent (but more complicated) book [20], which is aimed at acquainting the reader with the applications of computers in mathematics and mathematical physics.

Ya. B. Zeldovich
I. M. Yaglom

Contents

Part 1. Elements of Higher Mathematics

CHAPTER 1. FUNCTIONS AND GRAPHS 23

1.1 The Functional Relationship 23
1.2 Coordinates. Distances and Angles Expressed in Terms of Coordinates 26
1.3 Graphical Representation of Functions. The Equation of the Straight Line 30
1.4 Inverse Proportionality and the Hyperbola. The Parabola 34
1.5 Higher-Order Parabolas and Hyperbolas. The Semicubical Parabola 42
1.6 The Inverse of a Function. Graphs of Inverse Functions 47
1.7 Transforming Graphs of Functions 50
1.8 Parametric Representation of a Curve 59
1.9* Some Additional Topics from Analytic Geometry 61

CHAPTER 2. WHAT IS A DERIVATIVE? 67

2.1 Motion, Distance, and Velocity 67
2.2* Specific Heat Capacity of an Object. Thermal Expansion 70
2.3 The Derivative of a Function. Simple Examples of Calculating Derivatives 72
2.4 Properties of Derivatives. Approximating the Values of a Function by Means of a Derivative 75
2.5 A Tangent to a Curve 80
2.6 Increase and Decrease of Functions. Maxima and Minima 85
2.7 The Second Derivative of a Function. Convexity and Concavity of a Curve. Points of Inflection 89

CHAPTER 3. WHAT IS AN INTEGRAL? 92

3.1 Determining Distance from the Rate of Motion. The Area Bounded by a Curve 92
3.2 The Definite Integral 96
3.3 The Relationship Between the Integral and the Derivative 101
3.4 The Indefinite Integral 104
3.5 Properties of Integrals 109
3.6 Examples and Applications 112

CHAPTER 4. CALCULATION OF DERIVATIVES 127

4.1 The Differential 127
4.2 Derivatives of a Sum and of a Product of Functions 132
4.3 The Composite Function. The Derivative of the Fraction of Two Functions 134
4.4 The Inverse Function. Parametric Representation of a Function 135
4.5 The Power Function 138
4.6 Derivatives of Algebraic Functions 139
4.7 The Exponential Function 140
4.8 The Number e 142
4.9 Logarithms 145

4.10 Trigonometric Functions 148
4.11 Inverse Trigonometric Functions 151
4.12 Differentiating Functions Dependent on a Parameter and Functions of Several Variables. Partial Derivatives 153
4.13 The Derivative of an Implicit Function 158

CHAPTER 5. INTEGRATION TECHNIQUES 162

5.1 Statement of the Problem 162
5.2 Elementary Integrals 162
5.3 General Properties of Integrals 164
5.4 Integration by Parts 166
5.5 Change of Variables in Integration 168
5.6 Change of Variable in a Definite Integral 170
5.7 Integrating Functions Dependent on a Parameter 173

CHAPTER 6. SERIES. SIMPLE DIFFERENTIAL EQUATIONS 177

6.1 A Series Representation of a Function 177
6.2 Computing the Values of Functions by Means of Series 184
6.3 Cases Where Series Expansions Cannot be Applied. The Geometric Progression 187
6.4 The Binomial Theorem for Integral and Fractional Exponents 192
6.5 The Order of Increase and Decrease of Functions. L'Hospital's Rule 194
6.6 First-Order Differential Equations. The Case of Variables Separable 198
6.7* The Differential Equation for Water Flow from a Vessel 202

CHAPTER 7. INVESTIGATION OF FUNCTIONS. SOME PROBLEMS FROM GEOMETRY 214

7.1 Smooth Maxima and Minima 214
7.2 Other Types of Maxima and Minima. Salient Points and Discontinuities. The Left and Right Derivatives of a Function 221
7.3 Investigating Maxima and Minima of Functions Dependent on a Parameter 227
7.4* Convex Functions and Algebraic Inequalities 234
7.5 Computing Areas 241
7.6* Estimating Sums and Products 245
7.7* More on Natural Logarithms 250
7.8 Average, or Mean, Values 253
7.9 Arc Length 260
7.10 Curvature and the Osculating Circle 265
7.11 Solid Geometry Applications of Integral Calculus 271
7.12 Curve Sketching 275

Part 2. Higher Math Applied to Problems of Physics and Engineering

CHAPTER 8. RADIOACTIVE DECAY AND NUCLEAR FISSION 281

8.1 The Basic Characteristics of Radioactive Decay 281
8.2 Measuring the Mean Lifetime of Radioactive Atoms 283
8.3 Series Disintegration (Radioactive Family) 289
8.4 Investigating the Solution for a Radioactive Family (Series) 291
8.5 The Chain Reaction in the Fission of Uranium 294
8.6 Multiplication of Neutrons in a Large System 295
8.7 Escape of Neutrons 297
8.8 Critical Mass 298
8.9 Subcritical and Supercritical Mass for a Constant Source of Neutrons 299
8.10 The Critical Mass 301

CHAPTER 9. MECHANICS 303

9.1 Force, Work, and Power 303
9.2 Energy 308
9.3 Equilibrium and Stability 312
9.4 Newton's Second Law 316

9.5 Impulse 317
9.6 Kinetic Energy 320
9.7 Inertial and Noninertial Reference Frames 321
9.8* The Galilean Transformations. Energy in a Moving Reference Frame 324
9.9* The Path of a Projectile. The Safety Parabola 327
9.10 The Motion of a Body in Outer Space 331
9.11 Jet Propulsion and Tsiolkovsky's Formula 334
9.12 The Mass, Center of Gravity, and Moment of Inertia of a Rod 337
9.13*Centers of Gravity of a String and of a Plate 342
9.14 The Motion of a Body in a Medium that Resists this Motion with a Force Dependent Solely on the Velocity 346
9.15*The Motion of a Body in a Fluid 350

CHAPTER 10. OSCILLATIONS 354

10.1 Motion Under the Action of an Elastic Force 354
10.2 The Case of a Force Proportional to Deviation. Harmonic Oscillations 357
10.3 Pendulums 360
10.4 Oscillation Energy. Damped Oscillations 363
10.5 Forced Oscillations and Resonance 366
10.6 On Exact and Approximate Solutions of Physical Problems 368
10.7 Combining Oscillations. Beats 372
10.8 The Vibrations of a String 375
10.9 Harmonic Analysis. Fourier Series 380

CHAPTER 11. THE THERMAL MOTION OF MOLECULES. THE DISTRIBUTION OF AIR DENSITY IN THE ATMOSPHERE 387

11.1 The Condition for Equilibrium in the Atmosphere 387
11.2 The Relationship Between Density and Pressure 388
11.3 Density Distribution 389
11.4 The Molecular Kinetic Theory of Density Distribution 391
11.5 The Brownian Movement and Kinetic-Energy Distribution of Molecules 393
11.6 Rates of Chemical Reactions 395
11.7 Evaporation. The Emission Current of a Cathode 396

CHAPTER 12. ABSORPTION AND EMISSION OF LIGHT. LASERS 399

12.1 Absorption of Light: Statement of the Problem and a Rough Estimate 399
12.2 The Absorption Equation and Its Solution 400
12.3 The Relationship Between Exact and Approximate Absorption Calculations 400
12.4 The Effective Cross Section 402
12.5 Attenuation of a Charged-Particle Flux of Alpha and Beta Rays 403
12.6* Absorption and Emission of Light by a Hot Gas 405
12.7* Radiation in Thermodynamic Equilibrium 407
12.8* Emission Probability and the Conditions for Thermodynamic Equilibrium 410
12.9* Lasers 414

CHAPTER 13. ELECTRIC CIRCUITS AND OSCILLATORY PHENOMENA IN THEM 418

13.1 Basic Concepts and Units of Measurement 418
13.2 Discharge of a Capacitor Through a Resistor 423
13.3 Oscillations in a Capacitance Circuit with Spark Gap 425
13.4 The Energy of a Capacitor 428
13.5 Inductance Circuit 431
13.6 Breaking an Inductance Circuit 433
13.7 The Energy of Inductance 435
13.8 The Oscillatory Circuit 439
13.9 Damped Oscillations 441
13.10*The Case of a Large Resistance 443
13.11 Alternating Current 444
13.12 Average Quantities. Power and Phase Shift 447
13.13 An Alternating-Current Oscillatory Circuit. Series Resonance 449
13.14 Inductance and Capacitance in Parallel. Parallel Resonance 451

13.15 General Properties of Resonance in a Linear System 452
13.16*Displacement Current and the Electromagnetic Theory of Light 453
13.17*Nonlinear Resistance and the Tunnel Diode 454

Part 3. Some Additional Topics

CHAPTER 14. COMPLEX NUMBERS 458

14.1 Basic Properties of Complex Numbers 458
14.2 Raising a Number to an Imaginary Power and the Number *e* 463
14.3 Trigonometric Functions and the Logarithm 466
14.4* Trigonometric Functions of a Purely Imaginary Independent Variable. Hyperbolic Functions 469

CHAPTER 15. FUNCTIONS THE PHYSICIST NEEDS 475

15.1 Analytic Functions of a Real Variable 475
15.2 The Derivative of a Function of a Complex Variable 478

CHAPTER 16. DIRAC'S REMARKABLE DELTA FUNCTION 486

16.1 Various Ways of Defining a Function 486
16.2 Dirac and His Function 487
16.3 Discontinuous Functions and Their Derivatives 489
16.4 Representing the Delta Function by Formulas 493

CHAPTER 17. APPLYING FUNCTIONS OF A COMPLEX VARIABLE AND THE DELTA FUNCTION 496

17.1 Complex Numbers and Mechanical Oscillations 496
17.2 Integrals in the Complex Plane 500
17.3 Analytic Functions of a Complex Variable and Liquid Flow 507
17.4 Application of the Delta Function 512

Conclusion. What Next? 516
Selected Readings 530
Appendices 532
1. Derivatives 532
2. Integrals of Some Functions 532
3. Series Expansions 534
4. Numerical Tables 534
5. The International System, or SI 536
6. Greek Alphabet 536
Hints, Answers, and Solutions 537
Name Index 555
Subject Index 557

Part 1

Elements of Higher Mathematics

Chapter 1 **Functions and Graphs**

This chapter is of an introductory nature. It contains material that the majority of the readers already know and is devoted to functional relationships between quantities. In it the reader will find some specific functional relationships that we will often use in our narrative. With other important functional relationships the reader will get acquainted in the other chapters.

1.1 The Functional Relationship

Nature and technology abound in ***functional relationships***, that is, relationships between various quantities. It is natural then that mathematics pays so much attention to such relationships. A functional relationship between one quantity (y) and another (x) signifies that to every value of x there corresponds a definite value of y. The quantity x in this case is called the ***independent variable***, and y is the ***function*** of this variable. We also sometimes say that x is the *argument* of the function y.

Here are a few examples taken from geometry and physics.

(1) The area S of a square is a function of the length a of the square's side: $S = a^2$.

(2) The volume V of a sphere is a function of the sphere's radius R: $V = (4/3)\,\pi R^3$.

(3) The volume V of a cone with a given altitude h is a function of the radius r of the cone's base:

$$V = \frac{1}{3}\pi r^2 h. \tag{1.1.1}$$

On the other hand if we assume that r is fixed, then formula (1.1.1) expresses the volume V of the cone as the function of its altitude h.

(4) The distance z traversed by a freely falling body depends on the time t that elapses from the beginning of fall. This relationship is expressed by the formula

$$z = \frac{gt^2}{2}, \tag{1.1.2}$$

where $g \simeq 9.8$ m/s^2 is the acceleration of gravity.

(5) The current i depends on the resistance R of a conductor: for a given potential difference u we have

$$i = \frac{u}{R}. \tag{1.1.3}$$

This list could be extended without end.

In mathematics, functional relationships are ordinarily defined by formulas, for example,

$$\begin{gathered} y = 2x + 3, \quad y = x^2 + 5, \\ y = 3x^3 - x^2 - x, \\ y = \frac{x-1}{x+1}, \quad y = \sqrt{3x+7}. \end{gathered} \tag{1.1.4}$$

(Here y is everywhere a function of argument x). A formula enables us to compute the values of the function for each given value of the independent variable. Sometimes this method requires using tables or other technical devices. For instance, in the last example in (1.1.4) we must extract the square root of a number to find the value of the function. This can be done via tables of square roots or, even simpler, with a pocket calculator.

In physics and technology, functional relationships are usually recorded by measuring devices; say, the pointer of the speedometer in your car shows, together with the hand of your watch, the functional relationship of the car's speed and time. But in this case, too, the relationship can often be represented by a simple formula, which yields sufficiently accurate results. For instance, if the potential difference in an electric circuit remains constant, the experimentally observed (via the scale of the ammeter) relationship between the electric current i and the resistance R of the circuit is described by formula (1.1.3) fairly well. If we have a formula that relates the values of quantities that appear in physics and technology, we say that we have established a *law* for the phenomenon in question. In relation to formula (1.1.3) the law is ***Ohm's law.***

It is characteristic that in most cases in nature and technology the quantity of interest (the function) depends on *several* other quantities. In Example (5) the current depends on two quantities: the potential difference u and the resistance R of the conductor. The cone's volume in Example (3) depends on altitude h and base radius r. Assuming all quantities except one to be given and constant, we study the dependence of the function on a single variable. In this book we will confine ourselves mainly to functions of *one* variable. For example, taking a given storage battery with a definite potential difference u, we will vary the resistance R of the conductor and measure the current i. In this experiment the current depends only on the resistance, the quantity u in (1.1.3) being regarded as a constant coefficient.

Sometimes the formula gives only values of y for x varying within certain limits. Say, the formula $y = \sqrt{x}$ enables finding the values of y only for nonnegative values of x, that is, for $x \geqslant 0$. If $y = \log_2(x - 2)$, then y exists only for $x - 2 > 0$, that is, for $x > 2$. In the formula $y = (x - 1)/(x + 1)$ in (1.1.4) we must assume that $x \neq -1$, while in the last formula in (1.1.4) we must assume that $3x + 7 \geqslant 0$, that is, $x \geqslant -7/3 \simeq -2.33$.

In formulas that appear in physical, engineering, geometric, and other problems, the restrictions on the possible values of the independent variable sometimes follow from the very meaning of the problem considered. In the examples at the beginning of this section, say, when calculating the area of a square and the volume of a cone, we assumed that the side of the square and the base radius of the cone and its altitude are positive quantities. Often the formulas "tell" us about the possible restrictions on the values of the independent variable. For instance, the formula for the cone's volume yields a negative value if we take h negative, which result is of course unnatural; in many formulas of the theory of relativity, the speed v of a moving body appears in expressions of the type $\sqrt{c^2 - v^2}$, where c the speed of light, from which it follows that $v < c$; and so on.

In the first three formulas of (1.1.4) and, obviously, in many other formulas, no restrictions are imposed on the independent variable: x can be large or small, positive or negative. The same is true of many physical quantities; say, electric current can conveniently be assumed positive when the electrons move in a specified direction and negative when they move in the opposite direction.

Knowing the formula that states the dependence of y on x, we can easily construct a table of values of y for several arbitrarily chosen values of x.

By way of an illustration, we will set up a table for the function $y = 3x^3 - x^2 - x$. The upper row contains the values of x that we choose, the lower row, under each value of x, contains the appropriate value of y:

x	−3	−2	−1	0	1	2	3
$y=3x^3-x^2-x$	−87	−26	−3	0	1	18	69

Using this formula, we can make a more detailed table specifying, say, the values $x = 0, 0.1, 0.2, \ldots$. Thus a formula is stronger, so to say, than any table. The formula not only contains the information necessary to compile the given table but also enables one to find the values of the function for values of the independent variable not contained in the table. On the other hand, a table is convenient in that it immediately gives the value of y for any given value of x, provided that the needed x is given in the table. A table is also more pictorial than a complicated formula, by using which it is often difficult to estimate the values that the function assumes. However, a simple formula, with certain experience on the part of the researcher, provides a faster means for determining the behavior of the function than the inexpressive sequence of numbers in a table.

In natural sciences and technology it often happens that the theory of the phenomenon of interest to us is lacking and the physicist (or chemist, biologist, engineer) is only able to supply experimentally obtained facts—the dependence of the quantity of interest upon the quantity that was given in the experiment. This is what happens, say, in studies of the relationship between the resistance of a conductor and the temperature of the conductor. Here the functional relationship can only be given in the form of a table containing the experimental data.

Experiments show that for a given conductor (of a given material, a given cross section, and a given length) the electrical resistance depends on the temperature of the conductor. For each value of the temperature T, the conductor has a definite resistance R, so that we can speak of a relationship in which resistance R is a function of temperature T. Carrying out experiments, we can find the values of R for various T and thus find the dependence (relationship or function) R versus T, or $R = R(T)$. Here, the results of experiments are given in the table below (values of R for distinct values of T):

T (degrees Celsius)	0	25	50	75	100
R (ohms)	112.0	118.4	124.6	130.3	135.2

If we are interested in the values of R for other temperatures which do not appear in the table, additional measurements are required because there is no exact formula defining the dependence of R on T. Practically speaking, however, we can always offer an approximate formula that is in good agreement with the experimental data at temperatures at which the measurements were made. Let us take, for example, the formula

$$R = 112.0 + 0.272T - 0.0004T^2 \tag{1.1.5}$$

and compile the following table using this formula:

T (degrees Celsius)	0	25	50	75	100
R (ohms)	112.0	118.55	124.6	130.15	135.2

We see that the formula yields values of R that are very close to the experimental values for those temperatures at which the measurements were made, and so one is justified in assuming that at intermediate temperatures (say, at $T = 10$, 80, or 90 °C) the formula will likewise give a correct description of the functional depen-

dence of R on T. Mathematicians say in this case that the established dependence of R on T does not yield large errors in ***interpolation***, which is a process in which we go from the known values of R to new, intermediate, values (the Latin word *interpolare* means to refurbish, alter). The dependence of resistance R on temperature T found via formula (1.1.5) is called an ***empirical dependence***, or an ***empirical formula.*** (The adjective "empirical," from the Greek word *empeiria* meaning experience, means relying on experience or observation, based on observation or experience, or capable of being verified or disproved by observation or experiment.) However, an empirical formula, of course, requires verification, since the error introduced by using it may be considerable (clearly, the reliability of an empirical formula is the higher the thicker the grid formed by those values of the variable from which the empirical formula was constructed). An empirical formula becomes quite unreliable when we use it outside the range of the independent variable for which it was constructed (such continuation of the range of the empirical formula is known as ***extrapolation***, from the Latin *extra* meaning outside); the errors in this case may be very large. For instance, we cannot apply formula (1.1.5) outside the range of the investigated interval, say, at $T = -200$ °C or at $T = 500$ °C, since there are no grounds to expect that for *all* T $R(T)$ will be expressed by a *quadratic* trinomial in T, or formula (1.1.5).

1.2 Coordinates. Distances and Angles Expressed in Terms of Coordinates

Rectangular Cartesian ***coordinates*** are used as a pictorial way of representing functional relationships by means of drawings (graphs). Draw two perpendicular straight lines in a plane. Call the horizontal line the x axis (also known as the ***axis of abscissas***), the vertical line the y axis (also known as the

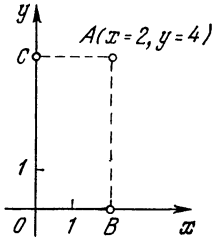


Figure 1.2.1

axis of ordinates), and the point of intersection of the two lines (point O in Figure 1.2.1) the ***origin of coordinates***, or simply the ***origin***. It is customary to picture the plane with the x and y axes not flat on a table but vertically in front of the reader, just as if it were drawn on the wall opposite you. The arrow of the x axis is from left to right and the arrow of the y axis is pointed upward. These arrows show that, say, the segment OB of the x axis is taken to be positive if B lies to the right of point O (as in Figure 1.2.1) and negative if B lies to the left of point O; similarly, the segment OC of the y axis is taken positive if C lies above O (as in Figure 1.2.1) and negative if C lies below O.

A definite pair of values of x and y, say, $x = 2$ and $y = 4$, is represented in the coordinate plane by a single point. To construct this point we plot on the axes of abscissas and ordinates the segments $OB = 2$ and $OC = 4$ (Figure 1.2.1), with both segments being positive because they are plotted from left to right and upward, respectively, and the x and y are positive. If we erect perpendiculars at points B and C on the x and y axes, the point at which they intersect is the sought point $A(x = 2,\ y = 4)$ or, briefly, $A(2,4)$. Conversely, to find the coordinates of an arbitrary point A it is sufficient to drop perpendiculars AB and AC on the coordinate axes and "read" the values of x and y corresponding to points B and C (these values are equal to the lengths of segments OB and OC taken with the appropriate signs). The coordinate axes divide the plane of the draw-

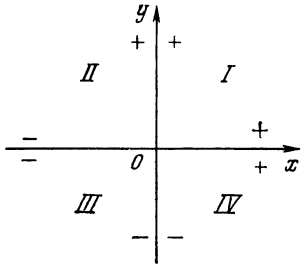


Figure 1.2.2

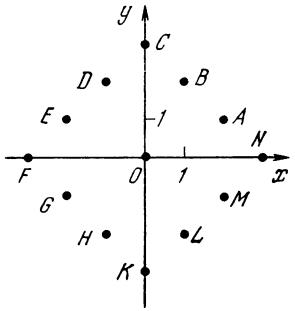


Figure 1.2.4

ing into four parts or quadrants, I, II, III, and IV, with the signs in each quadrant being as follows: (+, +), or positive abscissas and positive ordinates, in quadrant I, and (—, +), (—, —), and (+, —) in quadrants II, III, and IV, respectively (Figure 1.2.2). The axis of abscissas, Ox, corresponds to the value $y = 0$, while the axis of ordinates, Oy, corresponds to the value $x = 0$. The origin O has coordinates (0, 0). Figure 1.2.3 shows a few examples of points for which the corresponding values of the coordinates x and y are given in parentheses.

An important piece of practical advice: get into the habit of evaluating approximately the values of the coordinates of points in the plane and of finding the points from the known values of the coordinates. For practical work it is often convenient to use squared paper (it usually has a grid of millimeter squares) to locate points and plot curves. For instance, we advise the reader to try and determine the coordinates of points A to N in Figure 1.2.4 without resorting to additional lines and designations.

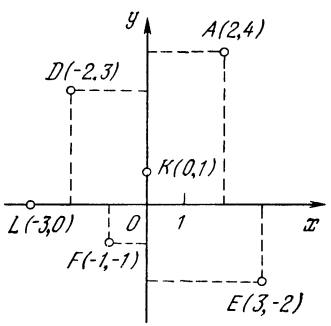


Figure 1.2.3

Thus, we can conclude that fixing two numbers, values of x and y, determines the position of a point in the plane. Therefore, all geometric quantities referring to this point can also be expressed in terms of the coordinates of the point.

Let us find, for instance, the ***distance*** r from the origin to the point A with coordinates x and y, that is, the length r of the line segment OA joining the origin O and the point A (Figure 1.2.5), and the ***angle*** α between the straight line OA and the axis of abscissas. From point A we drop the perpendiculars AB and AC on the axes of coordinates. The length of OB is equal to $|x|$, and that of segment AB is equal to the length of OC, or $|y|$. From the right triangle OAB, by the Pythagorean theorem, we have

$$(OA)^2 = r^2 = (OB)^2 + (AB)^2 = x^2 + y^2,$$

or

$$r = \sqrt{x^2 + y^2}. \qquad (1.2.1)$$

By the definition of the tangent,

$$\tan\alpha = \frac{AB}{OB} = \frac{y}{x}. \qquad (1.2.2)$$

For example suppose that $x = 2$ and $y = 3$ (see Figure 1.2.5). Then

$$r = \sqrt{13} \simeq 3.6, \quad \alpha = \arctan\frac{3}{2} \simeq 56°.$$

Note that the angle α is reckoned from the *positive* direction of the x axis *coun-*

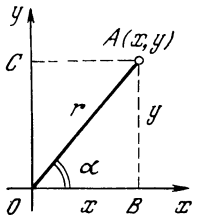


Figure 1.2.5

terclockwise. For this reason, if, say, $x = -2$ and $y = 2$ (point A in Figure 1.2.6), the angle α is obtuse: than $\alpha = 2/-2 = -1$ and $\alpha = 135°$. When a point lies below the x axis, the angle α proves to be greater than 180°. In Figure 1.2.6 we have two instances of this case: point B (2, —2) for which $\alpha = 315°$ and point C (—3, —3) for which $\alpha = 225°$. Since such large angles often prove to be inconvenient, it is customary to reckon them from the same positive direction of the x axis but *clockwise*, taking α negative. For point B we then have $\alpha = -45°$, and for point C we have $\alpha = -135°$. In other words, we can assume that for any point the angle lies either within the range from 0 to 360° (without the value of 360°, since it is more convenient to take $\alpha = 0°$ instead of $\alpha = 360°$) or within the range from —180° to 180°, with the values $\alpha = -180°$ and $\alpha = 180°$ belonging to points on the negative "half" of the x axis (here both values of α have an equal status).

In this respect formula (1.2.2) is incomplete since it does not enable us to determine whether a point lies in the I or III quadrant (or in the II or IV quadrant)—in these two quadrants the tangent has the same sign. To determine α completely, we must take into account the signs of x and y, which make it possible to determine the quadrant in which the particular point lies, or to use more complete formulas than (1.2.2), which are also more complicated:

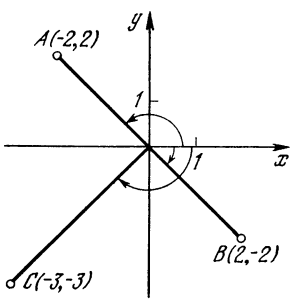


Figure 1.2.6

$$\cos\alpha = \frac{x}{r}, \quad \sin\alpha = \frac{y}{r}, \quad r = \sqrt{x^2 + y^2}. \tag{1.2.2a}$$

It is easy to solve the inverse problem: suppose we are given a point A at a given distance r from the origin O, with the line segment OA forming an angle α with the x axis (the positive direction of the x axis is assumed as usual). It is required to find the coordinates of point A. Looking at Figure 1.2.5, we see that

$$x = r\cos\alpha, \quad y = r\sin\alpha \tag{1.2.3}$$

These formulas are valid, without exception, for arbitrary positive and negative angles α and yield the proper signs of x and y in any quadrant.

Clearly, the position of a point A in a plane can be fixed by specifying the Cartesian coordinates x and y; however, instead of these two numbers we can use the distance r and the angle α. These two numbers, r and α, are known as ***polar coordinates*** of point A, point O is called the ***pole*** of polar coordinates in the plane, and the ray Ox is called the ***polar axis*** (the Oy axis does not take part in the definition of polar coordinates). Thus, formulas (1.2.2a) and (1.2.3) define the transition from rectangular Cartesian coordinates to polar coordinates and vice versa.

Let us now examine problems involving two points, A_1 and A_2. We denote the coordinates of the first point by x_1 and y_1, and the coordinates of the second point by x_2 and y_2 (Figure 1.2.7). We wish to find the distance r_{12} between these points and the angle α_{12} between

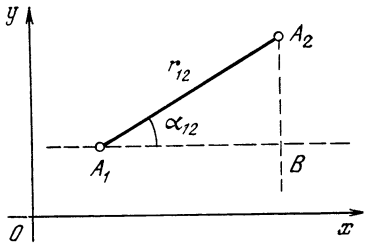


Figure 1.2.7

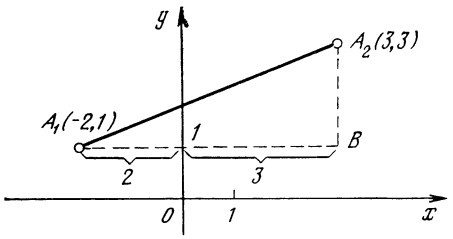


Figure 1.2.8

line segment A_1A_2 and the x axis.[1.1]

It is convenient to draw through A_1 a straight line parallel to the x axis and through A_2 a line parallel to the y axis (in Figure 1.2.7 they are shown as dashed lines and their point of intersection is B). In the triangle A_1A_2B the line segment A_1B is equal to $|x_2 - x_1|$ and the segment A_2B is equal to $|y_2 - y_1|$. The construction of triangle A_1A_2B is similar to the construction of triangle OAB in Figure 1.2.5.

By the Pythagorean theorem,

$$r_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \qquad (1.2.4)$$

The angle α_{12} is found from the condition

$$\tan \alpha_{12} = \frac{y_2 - y_1}{x_2 - x_1}, \qquad (1.2.5)$$

or (cf. (1.2.2a))

$$\cos \alpha_{12} = \frac{x_2 - x_1}{r_{12}} = \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}},$$

$$\sin \alpha_{12} = \frac{y_2 - y_1}{r_{12}} = \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}. \qquad (1.2.5a)$$

[1.1] The subscripts on the letters are known as *indices* (the Latin word *index* means to indicate). They are read: "x sub one" for x_1, "A sub two" for A_2. The same letter with different indices is used in place of a variety of letters to emphasize that we are dealing with similar (yet different) quantities. For instance, x_1 and x_2 are quantities on the x axis (both are abscissas), but they refer to distinct points. Quantities denoted by different letters but the same index refer to one and the same point: A_1 denotes a certain point, x_1 denotes the abscissa of that point, and y_1 denotes the ordinate of that same point. We sometimes use double-index notation: r_{12}, which is read "r sub one two" and not "r sub twelve", is the distance between the first point, A_1, and the second, A_2.

The reader should assure himself that the formulas (1.2.4), (1.2.5), and (1.2.5a) hold true for arbitrary signs of all four quantities x_1, y_1, x_2, y_2 and for any relationships between the coordinates: $x_1 > x_2$ or $x_1 < x_2$, $y_1 > y_2$ or $y_1 < y_2$. For example, in Figure 1.2.8 we have the case $x_1 < 0$ and $x_2 > 0$, with $A_1 = A_1(-2, 1)$ and $A_2 = A_2(3, 3)$. The length of segment A_1B is equal to the sum of the absolute values $|x_1| = 2$ and $|x_2| = 3$, which is strictly in accord with the general formula $A_1B = x_2 - x_1 = 3 - (-2) = 5$. Consequently, the expressions for r_{12} and than α_{12} are also correct.

Exercises

1.2.1. Plot the points (1, 1), (−1, 1), (−1, −1), and (1, −1).

1.2.2. Plot the points (1, 5), (5, 1), (−1, 5), (−5, 1), (−5, −1), (1, −5), and (5, −1).

1.2.3. Plot the points (0, 4), (0, −4), (4, 0), and (−4, 0).

1.2.4. Find the distance from the origin and the angle α for the points (1, 1), (2, −2), (−3, −3), and (−4, 4).

1.2.5. Find the distance between the following pairs of points: $A_1(1, 1)$ and $A_2(1, -1)$, $A_1(1, 1)$ and $A_2(-1, -1)$, $A_1(2, 4)$ and $A_2(4, 2)$, and $A_1(-2, -4)$ and $A_2(-4, -2)$.

1.2.6. Write out the coordinates of the vertices of a square with side a if the diagonals of the square coincide with the x and y axes.

1.2.7. Write out the coordinates of the vertices of a regular hexagon with side a if one of the diagonals coincides with the x axis and the center lies at the origin.

1.2.8. (a) Write out the coordinates of the vertices of an equilateral triangle with side a, with the base on the x axis, and with the vertex of the subtended angle on the y axis; (b) the same if the base lies on the x axis and the vertex of one of the angles lies at the origin.

1.2.9. Given a point A_1 with coordinates x_1 and y_1. Write out the coordinates of point A_2 symmetric to A_1 about the x axis; the same for A_3 symmetric to A_1 about the y axis; the same for A_4 symmetric to A_1 with respect to the origin.

1.3 Graphical Representation of Functions. The Equation of the Straight Line

In Section 1.2 it was shown that each pair of values x, y is associated with a definite point in the plane. If it is given that y is a definite function of x, then this means that to every value of x there corresponds a definite value of y. Therefore, if we are given a range of values of x, we can find the various corresponding values of y, and these pairs of values, (x, y), will yield many points in the plane. If we increase the number of distinct values of x by taking them closer and closer together, then finally the points will merge into a solid curve. This curve is called the ***graph*** of the function.

Actually, only a few points suffice to plot a graph, the intermediate points and the whole graph (curve) of the function being obtained by joining the points with a smooth curve. However, in order to avoid crude errors, we must have a general picture of the form of curves representing various functions. We begin with a few of the more typical and important functions.

We consider the ***linear relationship***, or ***linear function***,

$$y = kx + b. \qquad (1.3.1)$$

Suppose, say, $y = 2x + 1$. We construct a few points for which x and y are given in the table below:

x	0	1	2	3
y	1	3	5	7

Now plot these points on the graph in Figure 1.3.1. It is immediately seen that these points lie on one straight line. In this case, we draw the straight line (whence the term "linear relationship" or "linear functions") and obtain

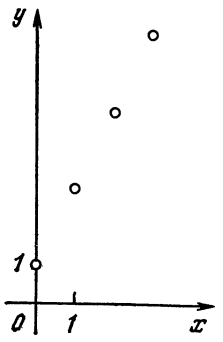


Figure 1.3.1

the entire graph of the function: for any x, the corresponding point (x, y) lies on the straight line that connects any two points of the graph.

How do we prove that, for any function of the form $y = kx + b$ (for arbitrary k and b), all points of the graph lie on a single straight line? In other words, how can we determine, without construction, merely by computing from the values of coordinates, that three points, A_1, A_2 and A_3, lie on a single line? It is clear that if the angle α_{12} between the segment A_1A_2 and the x axis is equal to the angle α_{13} between the segment A_1A_3 and the x axis, this means that the line segments A_1A_2 and A_1A_3 belong to one straight line, and if this is not so, then the segments belong to different straight lines. In Figure 1.3.2 we have the case where α_{13} is greater than α_{12} and point A_3 lies above the extension of A_1A_2, the same figure shows that if α_{13} were equal to α_{12}, then A_3 would be

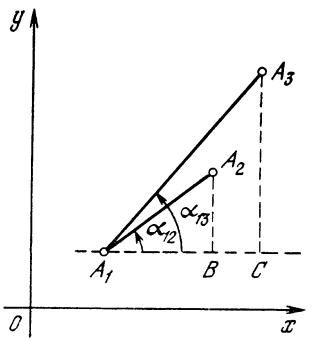


Figure 1.3.2

on the straight line that is the extension of A_1A_2.

From the expression of the tangent of angle α_{12} (1.2.5), it follows that at $\alpha_{12} = \alpha_{13}$ we have the following relationship between the coordinates of points A_1, A_2, and A_3:

$$\frac{y_2-y_1}{x_2-x_1} = \frac{y_3-y_1}{x_3-x_1}. \qquad (1.3.2)$$

Without using trigonometry, we can say that condition (1.3.2) is a condition of similarity of two right triangles A_1A_2B and A_1A_3C (see Figure 1.3.2). The similarity of the triangles indicates that the angles at the vertex A_1 are equal.

The relation (1.3.2) is also applicable in the case where point A_1 lies between points A_2 and A_3 (Figure 1.3.3); if the three points lie on one line, then from the similarity of the triangles A_1A_2B and A_1A_3C follows the proportion (1.3.2). In the example given in Figure 1.3.3, $x_3 - x_1 < 0$ and $y_3 - y_1 < 0$, but their ratio is positive and equal to the ratio of two positive quantities, $x_2 - x_1$ and $y_2 - y_1$.

Now let us verify that condition (1.3.2) remains valid for any triple of points of the linear function (1.3.1). Consider two points, $A(x_1, y_1)$ and $B(x_2, y_2)$, whose coordinates obey Eq. (1.3.1). In this case $y_1 = kx_1 + b$ and $y_2 = kx_2 + b$, with the result that $y_2 - y_1 = kx_2 + b - (kx_1 + b) = k(x_2 - x_1)$, whence

$$\frac{y_2-y_1}{x_2-x_1} = k.$$

The ratio proves to be independent of x_1 and x_2. Hence, for any other pair of points of the graph, in particular for the points $A(x_1, y_1)$ and $C(x_3, y_3)$, we also get

$$\frac{y_3-y_1}{x_3-x_1} = k = \frac{y_2-y_1}{x_2-x_1}.$$

This means that for any three points of the graph, $A(x_1, y_1)$, $B(x_2, y_2)$, and $C(x_3, y_3)$, the relation (1.3.2) is valid, which means that any three points of the graph lie on a single straight line and, hence, *all* points of the graph of the function $y = kx + b$ belong to one straight line. Thus, the graph of the function $y = kx + b$ is a straight line, which we will often call, for the sake of brevity, "the straight line $y = kx + b$" (we will also call it "the straight line (1.3.1)").

The equation $y = kx + b$ is called the ***equation of the straight line.*** The coefficient k determines the angle between the straight line and the x axis. Substituting $x = 0$ into the equation yields $y = b$, which means that one of the points of the straight line is the point $(0, b)$. This point lies on the y axis at a distance b above the origin if b is positive or at a distance b below the origin if b is negative. Thus, b is the ordinate of the point of intersection of the straight line and the y axis (sometimes b is called the initial ordinate of the straight line), and $|b|$ is the length of the line segment cut off by the straight line on the axis of ordinates (in Figure 1.3.1, $b = 1$), which is called the y intercept.

To construct a straight line corresponding to a given equation one need not compute the coordinates of a large number of points and plot them on the graph: it is clear that the construction of two points fully determines the straight line passing through these two points. For instance, we can always take two points, $y = b$ for $x = 0$ and $y = b + k$ for $x = 1$, and draw the line. For the second point we could also take the point of intersection of the straight line and the x axis, that is, the point with $x = x_0$ and $y = 0$. From the condition $y = kx_0 + b = 0$ we find

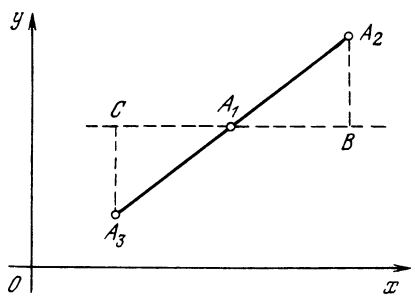


Figure 1.3.3

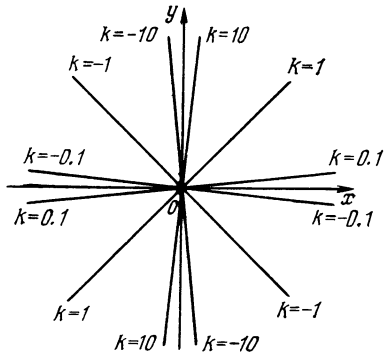


Figure 1.3.4

that $x_0 = -b/k$, which is known as the x intercept.

It is useful to do some drilling in the construction of graphs so as to be able to glance at an equation and picture roughly the variation and the position of the curve in question.

This is easy to do when we have a linear function whose graph is a straight line. The line depends only on two quantities, k and b, of the equation. Thus, not so many variants have to be examined: k can be positive or negative, k can be large or small in absolute value (greater than 1 or less than 1), and b can be positive or negative or even zero. Let us see how to carry out such an investigation.

We start with the case $b = 0$, or the equation $y = kx$. The straight line here will clearly pass through the origin, that is, through the point with $x = 0$ and $y = 0$. Figure 1.3.4 depicts several straight lines with different k's, whose values stand at the endpoints of each straight line. Check the correctness of each line and you will feel sure of the following general conclusions:

(1) If k is positive, the straight line lies in the first and third quadrants, while if k is negative, the straight line lies in the second and fourth quadrants.

(2) By the foregoing, if $k = 1$, the line lies in the first and third quadrants. The part of the straight line in the first quadrant forms an angle $\alpha = 45°$ with the x axis, which means that it bisects the angle between the x and y axes. The "angle with the x axis" here stands for the angle with the positive direction of the x axis (the one with the arrowhead). An extension of the straight line lying in the third quadrant forms an angle $\alpha = -135°$ with the x axis. The entire straight line bisects the angle between the x and y axes in the first and third quadrants.

(3) For $k = -1$, the part of the straight line lying in the second quadrant forms an angle $\alpha = 135°$ with the x axis, while the extension of the line in the fourth quadrant forms an angle $\alpha = -45°$. The entire straight line bisects the angle between the x and y axes in the second and fourth quadrants.

(4) If $|k| < 1$, the straight line is sloping, that is, is closer to the x axis than to the y axis, and the smaller the $|k|$, the closer the line is to the x axis. If $|k| > 1$, the straight line is steep, it is closer to the y axis than to the x axis, and the greater the $|k|$, the closer the line is to the y axis.

The quantity k is called the ***slope*** of the line; it is fixed by the value of angle α between the straight line and the x axis.

Now that this is clear, let us investigate the general case of a straight line with b different from zero. Suppose we have two straight lines: $y = kx$ (say, with $k = 0.5$; see Figure 1.3.5), that is, $b = 0$ in Eq. (1.3.1), and the straight line with the same slope k but with $b \neq 0$ (say, $y = 0.5x + 2$ in Figure 1.3.5). For the sake of convenience we introduce the

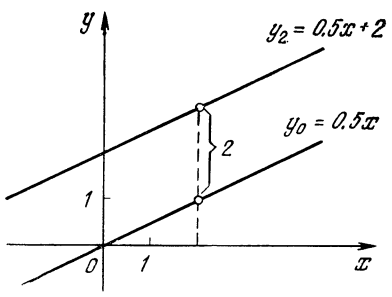


Figure 1.3.5

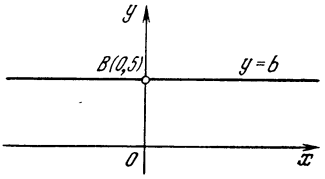


Figure 1.3.6

notation $y_0 = 0.5x$ and $y_2 = 0.5x + 2$. [1.2] For each given x, the quantity y_2 is two units greater than y_0. To summarize, then, the points of line y_2 are obtained from the points of line y_0 with the same x by an elevation of two units. The straight line y_2 is therefore parallel to y_0 and lies two units above it. Quite obviously, this rule holds true for any b (if b is negative, the line lies *below* the origin and below the corresponding straight line $y = kx$).

Now that we see how straight lines with equations $y = kx$ are located for distinct k's, we can readily picture the general positions of straight lines $y = kx + b$ with arbitrary k and b. Exercises that will help you to drill this material are given at the end of this section. In the particular case of $k = 0$ the equation is $y = b$ (it is assumed that $y = b$ for *all* values of x), which is associated with a horizontal straight line with a slant (slope) of zero (Figure 1.3.6).

We can imagine a man walking from left to right in the direction of increasing values of x. If $k > 0$, then he walks uphill (a positive slope), while if $k < 0$, the man walks downhill (a negative slope); if $k = 0$ (zero slope), the man walks along a horizontal path.

The slope k indicates the ratio of the variation of the function to the respective variation of the independent variable. Indeed, for any two points on the straight line, $A(x_1, y_1)$ and $B(x_2, y_2)$, we have

$$\frac{y(x_2)-y(x_1)}{x_2-x_1} = \frac{kx_2+b-(kx_1+b)}{x_2-x_1} = k$$

We have already calculated this ratio when we proved that a linear function on a graph is depicted by a straight line. In the general case of an arbitrary function $y(x)$ the similar quantity, $[y(x_2) - y(x_1)]/(x_2 - x_1)$ is the tangent of the angle between the segment connecting points $(x_1, y(x_1))$ and $(x_2, y(x_2))$ and the x axis (Figure 1.3.7*a*).

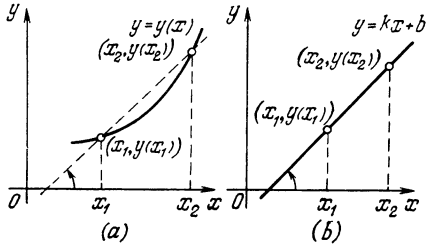


Figure 1.3.7

A linear function is distinguished by the fact that this ratio is the same for any two points: it depends neither on x_2 nor on x_1 (Figure 1.3.7*b*). For this reason, all points of a linear function belong to a single straight line.

Note, in addition, that a straight line parallel to the y axis (such a straight line does not represent any function) is written as $x = a$ (it is assumed that $x = a$ with y arbitrary; see Figure 1.3.8, where a is negative). It is natural to assume that the k of such a straight line is *infinite* ($k = \infty$), since in this case

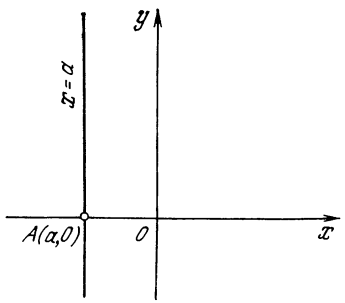


Figure 1.3.8

[1.2] The subscripts here are used somewhat differently from our earlier practice: y_0 refers to the entire line and not to the ordinate of a point—it is the ordinate of an arbitrary point on a line with given k and $b = 0$. In the same manner, y_2 is the ordinate of an arbitrary point on a line with given k and $b = 2$. In other words, neither y_0 nor y_2 are numbers; they are functions: $y_0 = y_0(x)$ and $y_2 = y_2(x)$.

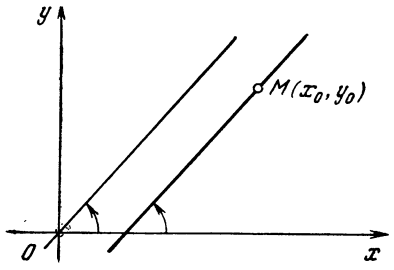


Figure 1.3.9

the ratio $k = (y_2 - y_1)/(x_2 - x_1)$, where points (x_1, y_1) and (x_2, y_2) belong to the straight line, has the form $c/0$ (the straight line $x = a$ is one with an infinitely great slope).

Now we can easily find the equation of a straight line *passing through a given point* $M(x_0, y_0)$ *and having a given slope* k. Such an equation will have the form of (1.3.1), with k known, but b has yet to be found, and the values x_0 and y_0 must satisfy this equation (Figure 1.3.9). This suggests the form of the sought equation:

$$y - y_0 = k(x - x_0), \qquad (1.3.3)$$

or

$$y = kx + (y_0 - kx_0). \qquad (1.3.3a)$$

Indeed, the slope of the straight line (1.3.3a) or (1.3.3) is k; on the other hand, if we substitute $x = x_0$ and $y = y_0$ into both sides of Eq. (1.3.3), we arrive at an identity, $0 = 0$.

Exercises

1.3.1. Determine whether the point triples lie on a straight line: $A_1(0, 0)$, $A_2(2, 3)$, $A_3(4, 6)$; $A_1(0, 0)$, $A_2(2, 3)$, $A_3(-2, -3)$; and $A_1(2, -3)$, $A_2(4, -6)$, $A_3(-2, 3)$.

1.3.2. Construct the straight lines $y = 3x$, $y = 3x + 2$, $y = 3x - 1$, $y = 2 - x$, $y = 2 - 0.5x$, and $y = x - 3$.

1.3.3. Find the equation of (a) a straight line that passes through point A (1, —1) and has a slope equal to —1, and (b) a straight line that passes through point B (2, 3) and has a slope equal to 2.

1.3.4. Prove that the equation of each straight line that has a nonzero slope (this slope must not be infinitely large), that is, a straight line not parallel to either axis, can be written in the form $x/a + y/b = 1$ (the intercept form). What geometrical meaning have the quantities a and b in this equation?

1.4 Inverse Proportionality and the Hyperbola. The Parabola

Let us recall the idea of ***direct proportionality*** of two quantities. The formula

$$y = kx \qquad (1.4.1)$$

(geometrically, as we already know, this formula represents a straight line) means that y is proportional to x: an increase in x by a certain factor leads to an increase in y by the same factor. In other words, for any two points $A(x_1, y_1)$ and $B(x_2, y_2)$ whose coordinates satisfy (1.4.1) we always have $y_2/y_1 = x_2/x_1$ (Figure 1.4.1*a*). It is also commonly said that formula (1.4.1) expresses the fact of *direct proportionality* between y and x (with the ***proportionality factor***, or coefficient, k). The more general equation

$$y = kx + b \qquad (1.4.1a)$$

characterizes the proportionality (or direct proportionality) between the *increment* $x_2 - x_1$ of the independent variable and the corresponding *increment* $y_2 - y_1$ of the function (dependent variable): for any two increments $x_2 - x_1$ and $x_4 - x_3$ of the independent variable of the function (1.4.1a), the corresponding increments of the function (dependent variable) will be proportional to the increments of the independent variable (Figure 1.4.1*b*):

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{y_4 - y_3}{x_4 - x_3} = k.$$

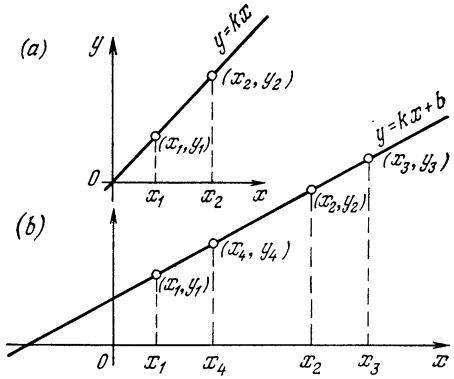


Figure 1.4.1

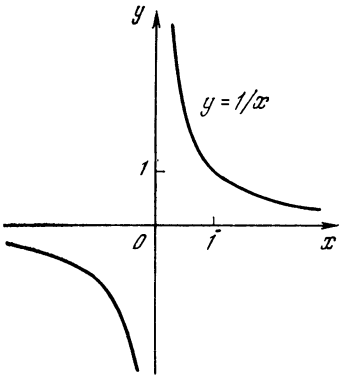


Figure 1.4.2

Another often encountered relationship between y and x is

$$y = \frac{k}{x}. \qquad (1.4.2)$$

Such a dependence is known as ***inverse proportionality*** (with coefficient k) between y and x: if points $A(x_1, y_1)$ and $B(x_2, y_2)$ satisfy Eq. (1.4.2), then x_2 is *greater* than x_1 by a factor by which y_2 is *less* than y_1, or $x_2/x_1 = y_1/y_2$ (since $y_2 \div y_1 = k/x_2 \div k/x_1 = x_1 \div x_2$).

Note that direct and inverse proportionalities are *reciprocal*: if y is directly (inversely) proportional to x, then x is directly (inversely) proportional to y, and vice versa. However, while the proportionality factor in direct proportionality (the factor that connects x with y) is the *inverse* of the proportionality factor in the (direct) proportionality relation that connects y with x (if $y = kx$, then $x = (1/k)\, y$), in the inverse proportionality between y and x the coefficients of (inverse) proportionality connecting y with x and x with y are the *same* (if $y = k/x$, then $x = k/y$).

The curve corresponding to Eq. (1.4.2) is known as the ***hyperbola***.[1.3] Figure 1.4.2 depicts a "unit" hyperbola

$$y = \frac{1}{x}, \qquad (1.4.2a)$$

and below we give the values of y for some values of x for this hyperbola:

x	-1	-0.1	-0.01	-0.001	0.001	0.01	0.1	1
y	-1	-10	-100	-1000	1000	100	10	1

The hyperbola has the peculiarity that for small positive x the value of y is very large, while for small negative x the value of y is also negative and very large in absolute value.

This property of curve (1.4.2a) (or the general curve (1.4.2)) is often expressed by stating that at $x = 0$ we have $y = \pm\infty$, where the plus or minus sign is chosen depending on the side from which we approach $x = 0$. The exact meaning of "$y = \pm\infty$ at $x = 0$" is that for a sufficiently small x the value of y become *arbitrarily large* (in absolute value), with it being positive or negative depending on the sign of x (the symbol ∞ is read "infinity" and is of course not a number). The smaller the value of x we choose (in absolute value, of course), the higher (or the lower in the case of $x < 0$) the corresponding point lies on the hyperbola: both branches of hyperbola (1.4.2a) "go off to infinity" (in the positive or negative direction) along the y axis (upward or downward) and approach the y axis without limit (but never intersect it), since the abscissa can be arbitrarily small. Similarly, as x increases without limit (in absolute value), that is, when x "tends to $+\infty$" or "tends to $-\infty$" (we remind the reader once more that the symbol ∞ does not correspond to a number), the value of $y = 1/x$ becomes arbitrarily small (in absolute value): the hyperbola "goes off to infinity along the x axis", approaching this axis without limit but never intersecting it (the two branches approach the x axis from different directions). This property of the hyperbola is expressed briefly by saying that the straight lines $x = 0$ (the y axis)

[1.3] We are speaking here only of *equilateral hyperbolas*, since by "hyperbola" mathematicians mean a somewhat more general curve than the one specified by Eq. (1.4.1). However, nowhere in our narrative do we use such general hyperbolas, and so the adjective "equilateral" will always be dropped.

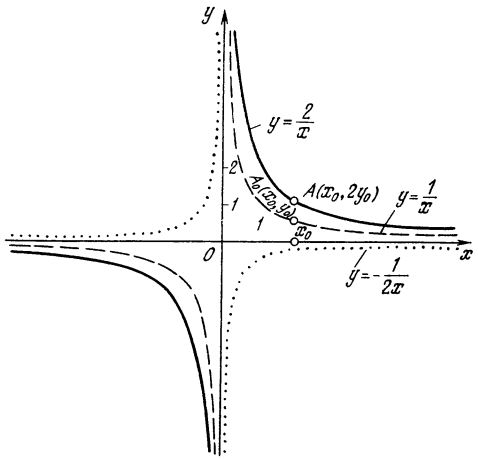


Figure 1.4.3

and $y = 0$ (the x axis) are the ***asymptotes*** to the hyperbola (from the Greek word *asymptotos* meaning "not meeting").

As seen from Figure 1.4.2, the hyperbola (1.4.2a) consists of two parts, or branches, corresponding to $x > 0$ and $x < 0$; these branches are separated, that is, they do not intersect.

The arbitrary curve (1.4.2) can be obtained from the curve (1.4.2a) by a simple transformation. Suppose that $A_0\,(x_0,\ y_0)$ is a point of the curve (1.4.2a), that is, $y_0 = 1/x_0$ (see Figure 1.4.3, in which the curve (1.4.2a) is shown by a dashed curve). In this case the same value $x = x_0$ has corresponding to it on curve (1.4.2), where we assume that k is positive, a point A with coordinates x_0 and $y = k/x_0 = k(1/x_0)$, that is, $A(x = x_0,\ y = ky_0)$. This point was obtained from point A_0 by a "stretching" transformation along the y axis (or to put it differently, by a "stretching" transformation away from the x axis) with a transformation coefficient equal to k, that is, a k-fold increase in all vertical dimensions—it is farther away from the x axis than point A_0 by a factor of k.

Here we assume that $k > 1$; when k is positive but less than unity, the curve (1.4.2) is obtained from the curve (1.4.2a) by a "shrinking" transformation along the y axis (or by a "shrinking" transformation toward the x axis), since each point of the curve (1.4.2) is closer to the x axis than the respective point of the curve (1.4.2a) by a factor of k (in this case the ratio $y \div y_0 = AP \div A_0P = k$ is less than unity).[1.4]

Figure 1.4.3 shows the curve $y = 1/x$ (the dashed curve), the curve $y = 2/x$ (the solid curve), and the curve corresponding to the function $y = (-1/2)/x$ which corresponds to a negative value of k equal to $-1/2$ (the dotted curve); this last function is negative for positive x and positive for negative x.

The direct and inverse proportionalities are often encountered in physical laws. For instance, *Ohm's law* (1.1.3) states that current i in a conductor changes in direct proportion to the potential difference u and in inverse proportion to the resistance R of the conductor. For a given R, the current i is directly proportional to u (with a proportionality coefficient equal to $1/R$); on the other hand, for a given u the current i is inversely proportional to R, or $i = k/R$, where potential difference u plays the role of k. A similar simple relationship $s = vt$ exists between the distance s traveled in uniform motion with a velocity v in the course of a time interval t. It shows that $t = s/v$, or that the time of traversal is directly proportional to distance s and inversely proportional to velocity v. Finally, in Boyle's law, pressure is in inverse proportion to volume of gas. Examples abound. The formulas $s = vt$ and $t = s/v$ show that, in uniform motion, the distance s traveled is proportional to the time t of traversal with the coefficient of proportionality equal to v, while t is proportional to s with coefficient $1/v$, that is, if, say, $v = 20$ m/s, then the above-noted coefficients of proportionality are 20 m/s and 0.05 s/m, respectively. The situation with Boyle's law $pV =$ constant

[1.4] On the connection between curves (1.4.2) and (1.4.2a) see Exercise 1.4.3.

is different. Here, if the temperature is 0 °C and the volume V and pressure p are measured in m^3 and Pa (pascal; see Appendix 5: 1 Pa = 1 N/m^2 = 1.450377 $\times 10^{-4}$ lb (wt)/in. ^{2}), it can be proved (see Section 11.2) that pV = 0.8737 N·m, and we see that p is inversely proportional to V, and V is inversely proportional to p, and in both cases the coefficients of proportionality are equal to 0.8737 N·m.

These examples proved a good illustration of the difference between the coefficients in direct proportionality $y = kx$ and inverse proportionality $y = k_1/x$. If x is measured in units of e_1 and y in units of e_2, then k has the dimensions of e_2/e_1, whereby there is no way in which k can serve as the coefficient k' in $x = k'y$, since the latter coefficient must be measured in units of e_1/e_2. The situation is opposite with k_1, which has the dimensions of e_1e_2. This implies that k_1 may (and actually does) coincide with the coefficient k_1' in $x = k_1'/y$.

Let us now turn to the *quadratic* function $y = ax^2 + bx + c$. We start with the simplest case:

$$y = ax^2, \qquad (1.4.3)$$

where coefficient a may be arbitrary. For the sake of simplicity we take $a = 1$, that is, we consider the curve[1.5]

$$y = x^2. \qquad (1.4.3a)$$

What general properties does this function have?

1. It is always true that $y > 0$, both for $x > 0$ and for $x < 0$. Only for $x = 0$ do we have $y = 0$. This means that the entire curve lies *above* the x axis and touches the x axis only at the origin.

2. y has a *minimum* (smallest value) at $x = 0$. The minimum is equal to 0. On the graph, the minimum is the lowest point of the curve.

3. Associated with two values of x identical in absolute value but with opposite signs are values of y identical

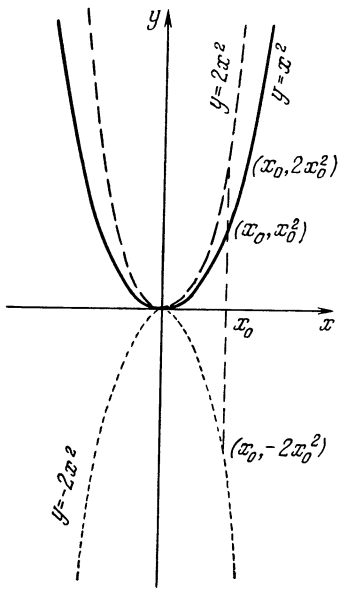


Figure 1.4.4

both in sign and absolute value. This means that the curve is *symmetric about the y axis.*

The curve (1.4.3a) is shown in Figure 1.4.4 by a solid line. It is called a ***parabola***, and this term will be used for all curves (1.4.3) with *arbitrary a* (and for curves of a broader class, as we will see below). For every positive value of a, the curve (1.4.3) possesses the same properties 1 to 3 as the curve (1.4.3a). Indeed, the transition from Eq. (1.4.3a) to Eq. (1.4.3) in all respects is similar to the transition from the "unit" hyperbola (1.4.2a) to the "general" hyperbola (1.4.2) (with a positive coefficient k): curve (1.4.3) is obtained from curve (1.4.3a) by stretching all the dimensions along the y axis a-fold (see Figure 1.4.4, where we have depicted parabolas $y = x^2$ and $y = 2x^2$).

What will happen if $a < 0$? Consider an example with $a = -2$, that is, the curve $y = -2x^2$. The dotted line (the one below the x axis) in Figure 1.4.4 depicts this curve. The properties of this curve are:

[1.5] In a certain sense this case can be thought of as general (see Section 1.8).

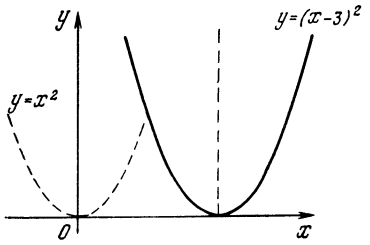


Figure 1.4.5

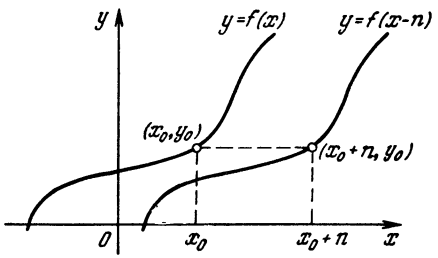


Figure 1.4.6

1. $y < 0$ for arbitrary $x \neq 0$. The whole curve lies *below* the x axis and touches the x axis at the origin.

2. The function has a *maximum* value at $x = 0$. The maximum is equal to 0. On the graph the maximum is the uppermost point of the curve.

3. The curve is symmetric about the y axis, just as in the case of a positive.

Now let us consider a more general equation

$$y = a\,(x - n)^2. \qquad (1.4.4)$$

We take $a = 1$ and $n = 3$. The respective curve is depicted in Figure 1.4.5. This is the same parabola as $y = x^2$ but it has been displaced rightward three units along the x axis.

This simple fact is not usually realized as readily as it should be. If a function $y = f(x)$ is given and we compare it with another function $y = f(x - n)$, the graph of the second function is shifted *rightward* from that of the first by n units (assuming n is positive). It is assumed here that in both cases f is one and the same function. In our example, the symbol f denotes a squaring of the independent variable, that is, the quantity inside the parentheses:

$f(x) = x^2$, $f(-x) = (-x)^2 = x^2$,

$f(x - 2) = (x - 2)^2$,

$f(x - n) = (x - n)^2$,

$f(x^2) = (x^2)^2 = x^4$, and so on.

But why is the graph shifted to the right? We will go into this in more detail. Suppose the graph of the function $y = f(x)$ has some kind of characteristic point $x = x_0$ (a kind of notch, so to say). For example, at this point the function may have, say, a salient point or a maximum or merely assume some definite value y_0 (Figure 1.4.6). Then that same value y_0 or the same salient point or the maximum will appear on the graph of the new function $y_1 = f(x - n)$ when the independent variable in the function f is equal to the old value x_0, that is, at $x - n = x_0$. This means that now the coordinates of the notch are $x = x_0 + n$, $y_0 = f(x_0)$. It is clear then that any notch, as it were, moves together with the whole graph to the right: point (x_0, y_0) on the first graph has corresponding to it on the second graph the point $(x_0 + n, y_0)$ (see Figure 1.4.6 and the solid and dashed parabolas in Figure 1.4.5, where for the notch we can take the point $x_0 = 0$, $y_0 = f(x_0) = 0^2 = 0$).

All this is very simple and elementary, but it is extremely important and the student should not merely learn it but fully comprehend the meaning of it. The first urge of most students is to say that when we replace $y = x^2$ with $y = (x - 3)^2$ the curve is displaced to the left because we subtract 3 from the value of x. It is well worth your time to make a detailed analysis of the examples, which demonstrate this common error.

Now we can state the general rules:

1. The curve $y = a\,(x - n)^2$ has the vertical line $x = n$ for its axis of symmetry.

2. This curve, for $a > 0$, lies above the x axis and has a minimum $y = 0$ at $x = n$, while for $a < 0$ it lies below

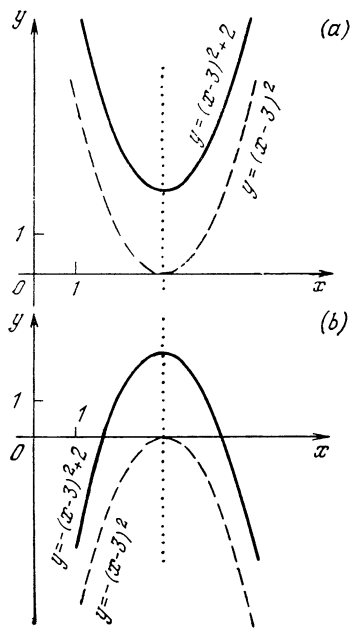


Figure 1.4.7

the x axis and has a maximum $y = 0$ at $x = n$.

Finally, there is yet another modification of this equation that does not alter the shape of the curve. Let us consider the function

$$y = a\,(x - n)^2 + m. \qquad (1.4\ 5)$$

This curve (also a parabola) clearly differs from the preceding one (without m) solely in the vertical displacement by the quantity m. The position of the axis of symmetry of the curve remains unchanged; for $a > 0$ the function has a minimum at $x = n$ and the value of the function at the minimum is equal to $y = m$ (the minimum, together with the whole curve, was shifted by the amount m), while for $a < 0$ the point $(x = n,\ y = m)$ is the point of maximum of the curve.

Two examples will suffice (Figure 1.4.7): $y = (x - 3)^2 + 2$ and $y = -(x - 3)^2 + 2$. The axes of symmetry in both parabolas are depicted in Figure 1.4.7 by dotted straight lines: the minimum point in Figure 1.4.7*a* and the maximum point in Figure 1.4.7*b* lie at the intersection of the respective curve and its axis of symmetry.

Removing the brackets in the expression $y = a\,(x - n)^2 + m$ yields

$$y = ax^2 - 2anx + an^2 + m. \qquad (1.4.5a)$$

On the right-hand side of (1.4.5a) we have a polynomial of degree two, which in its most general form has the notation

$$y = ax^2 + bx + c. \qquad (1.4.6)$$

This formula can be transformed as follows:

$$\begin{aligned} y &= a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) \\ &= a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) + a\left(\frac{c}{a} - \frac{b^2}{4a^2}\right) \\ &= a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right). \end{aligned}$$

Hence,

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right),$$

which implies that curve (1.4.6) is also a parabola with an axis of symmetry $x = -b/2a$ and a minimum point or maximum point $(-b/2a,\ c - b^2/4a)$.

Using the graph of a parabola, we can investigate the solution of a *quadratic equation* and the various cases that arise in this connection. We can approach the solution of the quadratic equation

$$ax^2 + bx + c = 0$$

this way: consider the curve

$$y = ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right)$$

and find the points of intersection of this curve with the x axis. At these points we have $y = 0$, and so the values of x corresponding to the points of intersection are the roots of the quadratic equation.

But we know that the curve $y = ax^2 + bx + c$ is a parabola. We also know that this parabola has an axis

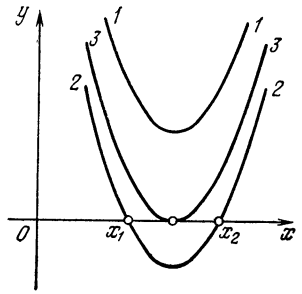


Figure 1.4.8

of symmetry, the vertical line $x = -b/2a$, and that for $a > 0$ the parabola has a minimum point on the axis of symmetry and the altitude (ordinate) of this minimum is $y = c - b^2/4a$ (if we glance at the right-hand side of the last formula, we see that it has the customary form $a(x - n)^2 + m$). For $a > 0$, the limbs of the parabola point upward.

It is clear that if the minimum lies above the x axis, the parabola does not intersect the x axis at any point (Figure 1.4.8, curve *1*). This means that when $a > 0$ and $c - b^2/4a > 0$ simultaneously, the quadratic equation has no roots.[1.6] But if the minimum lies below the x axis and the limbs of the parabola point upward, the parabola will definitely intersect the x axis at two points; these points will be symmetric with respect to the straight line $x = n = -b/2a$, the axis of symmetry of the parabola (curve *2* in Figure 1.4.8). This means that when $a > 0$ and $c - b^2/4a < 0$ simultaneously, the equation has two roots x_1 and x_2 as shown in Figure 1.4.8.

Finally, there may be an intermediate case where the parabola touches the x axis (curve *3* in Figure 1.4.8). This case occurs when $c - b^2/4a = 0$. If we gradually move curve *2* upward, it will finally coincide with curve *3*, the two roots x_1 and x_2 will come closer to each other and, ultimately, at the instant of tangency, will merge. That is why, when $c - b^2/4a = 0$, we speak not of one root but of *two equal* (coincident) roots of the equation.

The case $a < 0$ is considered in a similar manner. The respective curve has a maximum and the limbs point down. The reader is advised to draw the curves and to verify that (a) when $a < 0$ and $c - b^2/4a < 0$ simultaneously, there are no real roots, (b) when $a < 0$ and $c - b^2/4a > 0$, there are two real roots, and (c) when $a < 0$ and $c - b^2/4a = 0$, there are two equal roots (tangency).

The ordinary formula for the roots of a quadratic equation is

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The equation has two real roots when we are able to take a square root of $b^2 - 4ac$, that is, when $b^2 - 4ac > 0$. If we write this as

$$b^2 - 4ac = -4a\left(c - \frac{b^2}{4a}\right) > 0,$$

we see that this condition is satisfied when

$$(1)\ a > 0,\ c - \frac{b^2}{4a} < 0, \text{ and } (2)\ a < 0,\ c - \frac{b^2}{4a} > 0.$$

These are the two cases of the existence of two roots, which were obtained earlier from a consideration of the curves corresponding to the function $y = ax^2 + bx + c$.

We can approach the question of solving quadratic equations from a somewhat different angle. Let us divide all terms in the equation $ax^2 + bx + c = 0$ (where, of course, $a \neq 0$) by a. The result is

$$x^2 + px + q = 0, \text{ or } x^2 = -px - q, \quad (1.4.7)$$

where $p = b/a$ and $q = c/a$. Let us plot in the xy-plane two functions:

$$y_1 = x^2 \quad (1.4.7a)$$

(this is the "unit" parabola (1.4.3a)) and

$$y_2 = -px - q \quad (1.4.7b)$$

(a straight line). Clearly, for the values of x that satisfy Eq. (1.4.7) we have $y_1 = y_2$, that is, these values of x correspond to the

[1.6] Here by "root" we mean a *real* root of the equation; for the time being, until Chapters 14 and 15, we will simply ignore *complex-valued* roots.

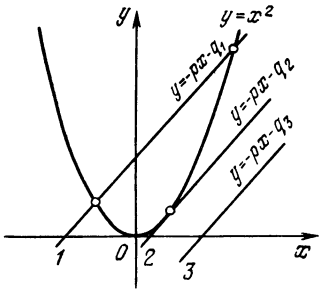


Figure 1.4.9

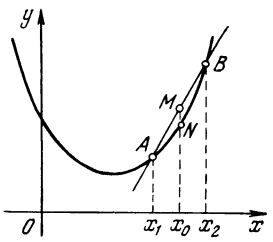


Figure 1.4.10

points of intersection of parabola (1.4.7a) with straight line (1.4.7b). But parabola (1.4.7a) can be plotted with great accuracy (using, say, graph paper), so that to solve any quadratic equation it suffices to draw the straight line (1.4.7b) (say, finding any two points of the straight line and using a ruler) on the same graph paper. The roots can then be "read off" the paper. The straight line (1.4.7b) may intersect the parabola (1.4.7a) at two points (as line *1* in Figure 1.4.9 does) or touch it at a single point (as line *2* in Figure 1.4.9; two coincident points of intersection with the parabola) or even have no points in common with parabola (1.4.7a) (as line *3* in Figure 1.4.9). All this corresponds to cases where Eq. (1.4.7) has two distinct roots or two coincident roots (i.e. one root) or no roots at all.

The aforesaid also enables us to conclude the following. We know that the equation $ax^2 + bx + c = 0$ has a single root if and only if $c - b^2/4a = 0$, whereby the equation $x^2 + px + q = 0$ (the case with $a = 1$) has a unique root if and only if

$$q - \frac{p^2}{4} = 0. \qquad (1.4.8)$$

Thus, (1.4.8) is the condition for tangency of the straight line (1.4.7b) and parabola (1.4.7a); in other words, the straight line $y = kx + b$ touches parabola $y = x^2$ if and only if [1,7]

$$k^2 + 4b = 0. \qquad (1.4.8a)$$

In Chapter 2 we derive the same condition (1.4.8a) for tangency of a straight line and a parabola using another method.

Observe, finally, that, depending on the sign of the coefficient a of x^2 in the equation of the parabola (1.4.6), the curve is convex down ($a > 0$) or convex up ($a < 0$). This property does not depend on the values and signs of b and c in the equation of the parabola (1.4.6).

An exact definition of convexity is this: take two points $A(x_1, y_1)$ and $B(x_2, y_2)$ on a curve and draw a line through them (Figure 1.4.10). If the portion of the curve between the two points lies below the straight line (below the chord AB of the curve), we say that the curve is ***convex down***, while if the portion of the curve between the two points lies above the straight line (above chord AB), we say that the curve is ***convex up***.

The convexity of a parabola is readily seen in a drawing, but we can also define it algebraically. Take arbitrary values x_1 and x_2 of the abscissa x. They are associated with points on the parabola, $A(x_1, y_1)$ and $B(x_2, y_2)$, where $y_1 = ax_1^2 + bx_1 + c$ and $y_2 = ax_2^2 + bx_2 + c$.

We wish to find the coordinates of point M lying at the midpoint of the line segment AB (Figure 1.4.10). It may be demonstrated geometrically that if $AM = MB$, the coordinates of point $M(x_0, y_0)$ are arithmetic means of the coordinates of A and B:

$$x_0 = \frac{x_1 + x_2}{2} \quad \text{and} \quad y_0 = \frac{y_1 + y_2}{2}.$$

(This follows from the fact that in Figure 1.4.10 the length of the mean-line Mx_0 of the trapezoid ABx_2x_1 is equal to one-half of the sum of the lengths of the bases Ax_1 and Bx_2 of the trapezoid.) Now let us find the coordinates of the point $N(X, Y)$ lying on the parabola for the same value

[1,7] It is clear that if in (1.4.8) we substitute k for $-p$ and b for $-q$, we arrive at (1.4.8a).

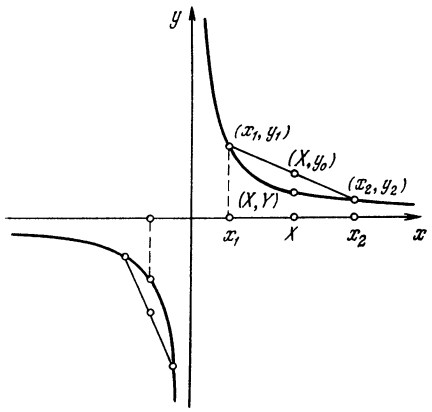


Figure 1.4.11

$x_0 = (x_1 + x_2)/2$ of the abscissa. Here

$$Y = aX^2 + bX + c$$
$$= a\left(\frac{x_1+x_2}{2}\right)^2 + b\frac{x_1+x_2}{2} + c.$$

The reader can assure himself that

$$Y - y_0 = a\left(\frac{x_1+x_2}{2}\right)^2 - \left(a\frac{x_1^2}{2} + a\frac{x_2^2}{2}\right)$$
$$= -a\left(\frac{x_1-x_2}{2}\right)^2$$

(since the terms involving b and c cancel out). But the quantity $\left(\frac{x_1 - x_2}{2}\right)^2$ is positive for arbitrary x_1 and x_2. Consequently, for $a > 0$ we have $Y < y_0$, or the point on the parabola lies below the corresponding point on the straight line (i.e. having the same abscissa X), that is, the parabola is convex down. On the other hand, for $a < 0$ we have $y_0 < Y$, and the parabola is convex up.

The hyperbola $y = k/x$ (where we assume that k is positive) consists of two branches. The reader can clearly see that the branch of the hyperbola corresponding to positive x's is convex down, while the second branch of the hyperbola is convex up (see Figure 1.4.11 and Exercise 1.4.4).

Exercises

1.4.1. Plot the following curves: $y = 3/x$, $y = -0.5/x$, and $y = 1/x + 3$.

1.4.2. Plot the following curves: $y = x^2 - 2x + 2$ and $y = 2x^2 + 4x$.

1.4.3. Show that the hyperbola $y = k/x$, where k is positive, can be obtained from the hyperbola $y = 1/x$ by (a) a stretching transformation from the origin O (a homothetic transformation) with a coefficient $\sqrt{k}$ (for $k < 1$ it is more appropriate to speak of a shrinking transformation), which means that if M' and M are points belonging to the curves $y = k/x$ and $y = 1/x$ and lying on the straight line that passes through the origin O, then $OM' \div OM = \sqrt{k}$; and (b) a stretching transformation along the x axis (or away from the y axis) with a coefficient k (here, too, for $k < 1$ it is more appropriate to speak of a shrinking transformation).

1.4.4. Prove algebraically that for $k > 0$ the hyperbola $y = k/x$ is convex down when $x > 0$ and convex up when $x < 0$.

1.5 Higher-Order Parabolas and Hyperbolas. The Semicubical Parabola

The curve given by the graph of the function

$$y = ax^n \tag{1.5.1}$$

(where n is a positive integer) is often called a ***parabola of order n***. For instance, Figure 1.5.1 depicts a *third-order parabola* (or a cubical parabola)

$$y = x^3 \tag{1.5.2}$$

and a *fourth-order parabola*

$$y = x^4. \tag{1.5.3}$$

We see that a fourth-order parabola (1.5.3) resembles an ordinary (second-order) parabola: it has a symmetry axis (the y axis), at point (0, 0) it has a minimum (the only one), and at the origin O it touches the x axis that does not intersect it (just as the parabola $y = x^2$).

The cubical parabola $y = x^3$ possesses quite different properties. It has neither maxima nor minima: an increase in x *always* brings about an increase in y, that is, as a point moves from left to right on the curve, it always rises. It is said that the function (1.5.2) *increases* for all values of x (what we

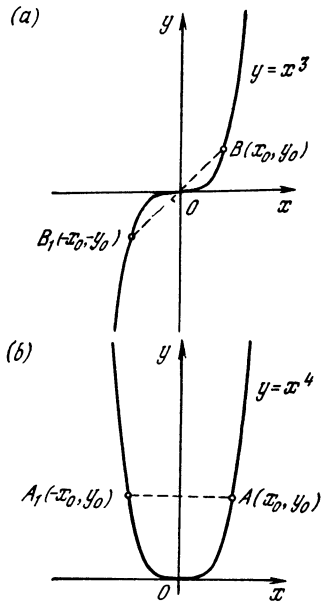


Figure 1.5.1

mean by this is that the function increases with the independent variable and that this behavior does not change). Curve (1.5.2) has no axis of symmetry: the values of y in this case are negative for negative x and positive for positive x. Instead it has a *center of symmetry*, the origin O (0, 0). Indeed, if we take the curve (1.5.3), we can see that to every two values of the independent variable that differ in sign but not in absolute value, x_0 and $-x_0$, there corresponds only *one* value $y_0 = x_0^4$ of the function (or two coincident values), that is, to each point $A(x_0, y_0)$ on curve (1.5.3) there corresponds a point $A_1(-x_0, y_0)$ *symmetric about the y axis* (Figure 1.5.1*b*). Now if we take the function (1.5.2), we see that to every two values of the independent variable that differ in sign but not in absolute value, x_0 and $-x_0$, there correspond two values of the function that *differ in sign* but not in absolute value, namely, $y_0 = x_0^3$ and $-y_0 = -x_0^3$, so that to each point $B(x_0, y_0)$ on the parabola (1.5.2) there corresponds a point $B_1(-x_0, -y_0)$ *symmetric* with respect to the origin O, which evenly divides the line segment BB_1 into two parts (Figure 1.5.1*a*).

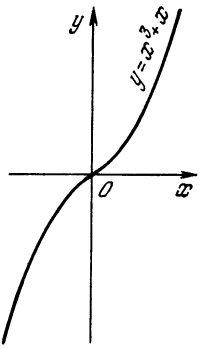


Figure 1.5.2

Figure 1.5.2 depicts a curve defined by the formula

$$y = x^3 + x. \tag{1.5.4}$$

This curve also has the distinguishing feature that on any portion of it an increase in x brings about an increase in y and the curve constantly rises from left to right, just as the function $y = x^3$ does. The curve (1.5.4) has neither maxima nor minima. Quite obviously, such a curve cuts the axis of abscissas only once, at point O.

Figure 1.5.3 shows a curve constructed on the basis of the formula

$$y = x^3 - x. \tag{1.5.5}$$

As is evident from the graph, this curve has two portions where y increases with x: for negative $x < -0.58$ and for positive $x > 0.58$. Between them, on the interval $-0.58 < x < 0.58$, the function is decreasing: y decreases as x grows.

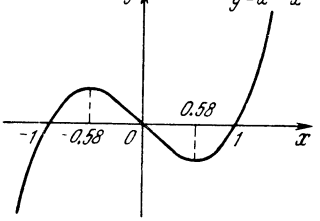


Figure 1.5.3

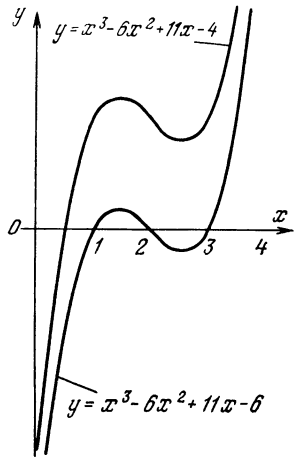


Figure 1.5.4

The function (1.5.5) has a *maximum* at ($x \simeq -0.58$, $y \simeq 0.38$).[1.8] In this context, the word "maximum" does not mean that in the given case $y \simeq 0.38$ is the greatest possible value of y given by the expression (1.5.5), since it is clear that for large positive values of x the quantity y will assume *arbitrarily large* values. So what is so conspicuous about the maximum point ($x \simeq -0.58$, $y \simeq 0.38$) of the function (1.5.5)?

As the graph of this function shows, at this point y is greater than at *adjacent* points. The point of maximum separates the portion of the curve where the function is growing (to the left of the maximum) from the portion where the function is decreasing (to the right of the maximum). This is what is called the ***local*** (or ***relative***) maximum: the value of y at this point is greater than the values of y at other points, but only for values of x that are not too far from x_{max} (in our case $x_{max} \simeq -0.58$). Similarly, at point ($x \simeq +0.58$, $y \simeq -0.38$) the function has a local (relative) minimum.

In Figure 1.5.4 we have two more examples of curves describing polynomials of order three. The cubic equation that we get by equating the respective polynomial to zero has one (real) solution (or root), $x \simeq 0.48$, in the case of the upper curve, and three roots, $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$, in the case of the lower curve. It is easy to see that a cubic equation always has *at least one real root*: to be sure of this, the reader is advised to examine the behavior of the curve $y = ax^3 + bx^2 + cx + d$ for very large (in absolute value) positive and negative values of x.

The graph of the cubical parabola (1.5.2) can be used to solve (approximately) an arbitrary equation of degree three. Let us write the general equation of degree three (the cubic equation) as

$$x^3 + 3ax^2 + bx + c = 0 \qquad (1.5.6)$$

(we will find it convenient to denote the coefficient of x^2 by $3a$ rather than by a). We transform Eq. (1.5.6) thus:

$$(x^3 + 3ax^2 + 3a^2x + a^3) + (b - 3a^2)\,x + (c - a^3) = 0,$$

or

$$(x + a)^3 - (3a^2 - b)\,(x + a) - [(a^3 - c) - a\,(3a^2 - b)] = 0,$$

or, finally,

$$X^3 - KX - B = 0, \qquad (1.5.6a)$$

where $X = x + a$, $K = 3a^2 - b$, and $B = ab - c - 2a^3$.

Clearly, solving Eq. (1.5.6) is the same as solving Eq. (1.5.6a), which can also be rewritten as

$$X^3 = KX + B. \qquad (1.5.7)$$

Hence, if we can sketch the cubical parabola $y = X^3$ with great accuracy (say, using graph paper), we need only construct on the same graph the straight line $y_1 = KX + B$ and find the points (or point) of intersection of the straight line with the cubical parabola—the result is the solution, or the approximate values of the roots of Eq. (1.5.6a) (and hence of Eq. (1.5.6)).

For example, in the case of the equation $x^3 - 6x^2 + 11x - 4 = 0$ (see Figure 1.5.4) we have $a = -2$, $b = 11$, $c = -4$, whence $K = 3a^2 - b = 3 \times 4 - 11 = 1$ and $B = ab - c - 2a^3 = -2 \times 11 + 4 - 2 \times (-8) = -2$. Thus, to find the solutions to our equation we need only find the points of intersection of the parabola $y = X^3$ and the straight line $y_1 = X - 2$. Figure 1.5.5, yields the approximate solution: $X \simeq -1.5$, whence $x = X - a \simeq -1.5 + 2 \simeq 0.5$.

[1.8] In Section 2.6 we will learn how to find the maxima and minima of a function. For example, in the case of formula (1.5.5) we will find that $x_{max} = -1/\sqrt{3} \simeq 0.57735$, $y_{max} = 2/(3\sqrt{3}) \simeq 0.3849$.

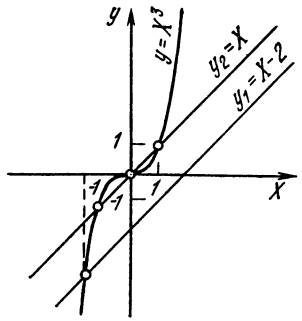


Figure 1.5.5

Similarly, in the case of the equation $x^3 - 6x^2 + 11x - 6 = 0$ (see Figure 1.5.4) we have $a = -2$, $b = 11$, $c = -6$, $K = 3a^2 - b = 1$, and $B = ab - c - 2a^2 = 0$. Here, obviously, the points of intersection of the cubical parabola $y = X^3$ with the straight line $y_2 = X$ correspond to values of X that are -1, 0, and $+1$, which implies that the roots $x = X - a = X + 2$ of the initial equation are equal to 1, 2, and 3 (see Figure 1.5.5).

The curve that represents the function

$$y = \frac{k}{x^n} \left(= kx^{-n}\right), \tag{1.5.8}$$

where n is a positive integer, is sometimes called a *hyperbola of order* $n + 1$ or a *hyperbola of degree* n. For example, Figure 1.5.6 depicts a *hyperbola of degree two*

$$y = \frac{1}{x^2} \tag{1.5.9}$$

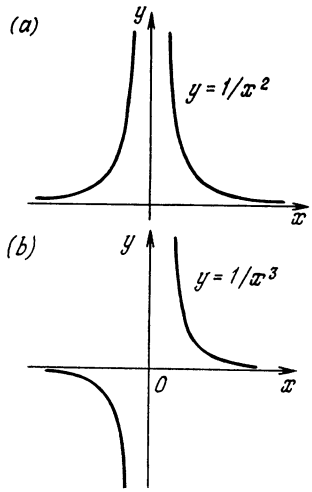


Figure 1.5.6

and a *hyperbola of degree three*

$$y = \frac{1}{x^3}. \tag{1.5.10}$$

As one can easily understand, curve (1.5.10) resembles an ordinary hyperbola (1.4.2a): both consist of two branches (symmetric about the origin O), and at $x = 0$ the value of y is $+\infty$ or $-\infty$ (compare the results of Section 1.4 concerning the hyperbola (1.4.2a) with this result). On the other hand, the hyperbola (1.5.9) differs from an ordinary hyperbola, since here y is positive for x both positive and negative. This curve also consists of two branches, which are symmetric about the straight line $x = 0$ (the y axis) rather than about O. For the function $y = 1/x^2$ we can symbolically write $y = +\infty$ at $x = 0$ (this means that for small values of x the quantity y may become as large as desired).

The graph of the function

$$y = \pm x^{3/2} = \pm \sqrt{x^3}, \tag{1.5.11}$$

or, more precisely, the curve with the equation

$$y^2 = x^3 \tag{1.5.11a}$$

is known as a ***semicubical parabola*** (Figure 1.5.7). Since no values of y correspond to negative values of x (for negative x's the right-hand side of (1.5.11a) is negative, which is impos-

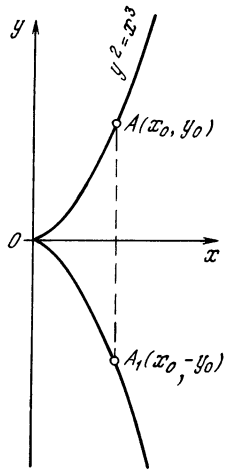


Figure 1.5.7

sible since y^2 is never negative), the entire curve lies in the right half-plane. Since to each value of x there correspond two values of y that differ in sign but not in absolute value, precisely, $+\sqrt{x^3}$ and $-\sqrt{x^3}$, the curve is *symmetric* with respect to the x axis: to each point on the curve, $A(x_0, y_0)$, there corresponds a point symmetric about the axis of abscissas, $A_1(x_0, -y_0)$; below, in Section 2.5, we prove with sufficient rigor that the semicubical parabola at the origin O *touches* the x axis.

These examples illustrate the behavior of all *power functions*, that is, functions of the type

$$y = x^n, \tag{1.5.12}$$

where the exponent n may be positive or nonpositive, greater or smaller than unity in absolute value, an integral or a fractional number.

If $n > 1$, the curves representing (1.5.12) behave like the quadratic parabola (1.4.3a) (we will consider only nonnegative values of x since the very definition of the quantity x^n with x negative and n a noninteger presents certain difficulties; for instance, $x^{1/2} = \sqrt{x}$ does not exist for x negative). All curves of this type touch the x axis at the origin O and pass through point Q (1, 1), and the higher the value of n, the closer the corresponding curve (1.5.12) is to the x axis in the vicinity of the origin and the steeper is the curve in the vicinity of point Q. In the interval $0 < n < 1$, on the other hand, curves (1.5.12) touch the y axis, and the smaller the value of n, the closer they are to the y axis; these curves also pass through point Q (1, 1), and the smaller the value of n, the sharper is their turn away from the origin to this point (see Figure 1.5.8a, which shows the different curves, which correspond to different positive values of n in (1.5.12), and the straight line $y = x$, which corresponds to $n = 1$ and separates the curves (1.5.12) with $n < 1$ from those with $n > 1$).

The curves (1.5.12) with n negative behave quite differently. These curves, specified by the equation

$$y = x^{-m} = \frac{1}{x^m}, \quad m > 0, \tag{1.5.12a}$$

also pass through point Q (1, 1), but, in contrast to the case where n is positive, they do not enter the square $CAQB$ with the diagonal OQ and lie completely outside it. The x and y axes are the asymptotes to these curves, and

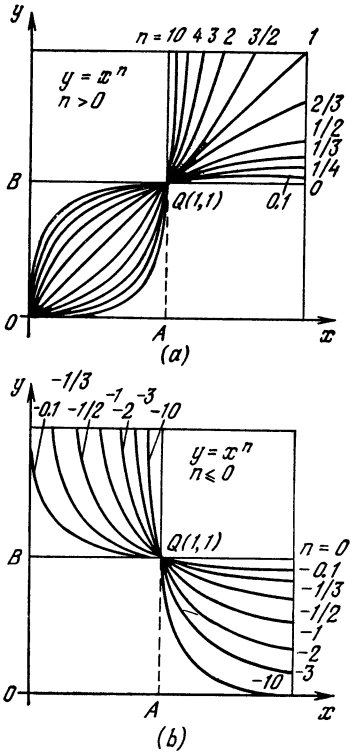


Figure 1.5.8

the greater the absolute value of n, that is, the greater the positive number $m = -n$, the faster these curves tend to merge with the x axis (a complete merge never happens, however) and the slower they approach the y axis (see Figure 1.5.8*b*, where we have depicted curves (1.5.12) with the following values of n: $-1/10$, $-1/3$, $-1/2$, -1, -2, -3, -10, that is, $m = 1/10$, $1/3$, $1/2$, 1, 2, 3, 10 in Eq. (1.5.12a) and the "limit line" $y = x^0 = 1$ corresponding to $n = 0$).

Note also that through each point $M(x, y)$ of the first quadrant (i.e. points with x and y positive) for which $x \neq 1$ there passes *only one* curve (1.5.12); if $x - 1$ and $y - 1$ have the same sign, that is, if $x > 1$ and $y > 1$ or $x < 1$ and $y < 1$, the value of n for this curve is positive, while if $x - 1$ and $y - 1$ have different signs, that is, if one of the numbers x, y is greater than unity and the other is smaller than unity, then $n < 0$ for this curve. As for the points $N(1, y)$ with $y > 0$, *not a single* curve of (1.5.12) passes through such a point N that differs from Q, while all curves specified by (1.5.12) pass through point Q.

As for the behavior of the curves specified by (1.5.12) at great absolute values of n, see Exercise 1.5.3. (Mathematicians love to speak

of asymptotic behavior, or simply asymptotics, of the functions (1.5.12) when $n \to \infty$ or $n \to -\infty$.)[9.1]

Exercises

1.5.1. Construct the curves given by the equations (a) $y = x^4 + 1$, (b) $y = -x^4 + 1$, and (c) $y = x^4 + x^2$.

1.5.2. Construct the curves given by the equations (a) $y = -x^{-3} + 1$ and (b) $y = -x^{-2} + 4$.

1.5.3. What shape have the curves (a) $y = x^{2n}$, (b) $y = x^{2n+1}$, (c) $y = x^{-2n}$, and (d) $y = (x^{2n+1})^{-1}$ for very large values of the positive integer n?

1.6 The Inverse of a Function. Graphs of Inverse Functions

By fixing a quantity y as a function of another quantity x we mean that to each x there is assigned a definite value of y. But this dependence can be inverted, namely, we can fix y and then find x. For example, the law of motion of a train traveling from one station to another can be fixed by specifying the position z of the train at every moment t, where z can be, say, the distance traveled from the initial station. This fact, of course, can be stated mathematically as $z = f(t)$, and the engine-driver uses the function in his timetable. But a passanger is more interested in the "inverse" timetable, the dependence of the time t_1 at which the train will be at a specified station determined by the coordinate z_1. The dependence $t = g(z)$ is called the ***inverse*** of $z = f(t)$ or, in other words, g is a ***function*** that ***is inverse to*** f. Note that, of course, f is the inverse of g.

Here are some examples, where the left column lists the function $y = y(x)$ and the right column the function $x = x(y)$ (in the last instance the independent argument is denoted by y and the function by x contrary to tradition):

$$\begin{aligned} &y = x + a, && x = y - a; \\ &y = 3x + 2, && x = \frac{1}{3}y - \frac{2}{3}; \\ &y = 1 - x, && x = 1 - y; \\ &y = x^2, && x = \pm\sqrt{y}; \\ &y = x^3 + 1, && x = \sqrt[3]{y - 1}. \end{aligned} \qquad (1.6.1)$$

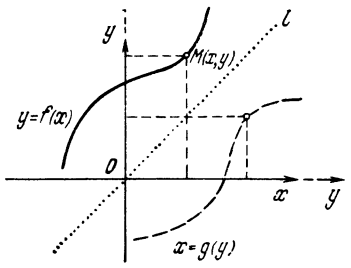


Figure 1.6.1

It is easy to grasp how the graphs of the functions f and g are connected. Suppose that we have the graph of the function $y = f(x)$ (Figure 1.6.1). For this graph to represent the function $x = g(t)$, we must view it at a different "angle", precisely, the y axis must be considered the axis of abscissas (the independent variable) and the x axis the axis of ordinates. If we wish the axis of the independent variable to remain horizontal and the axis representing the values of the function vertical, we must rotate the graph together with the coordinate axes through 180° about bisector of the first and third quadrant angles (the straight line l). After such a rotation the y axis becomes horizontal and the x axis vertical. The result of such a rotation is depicted in Figure 1.6.1 by a dashed curve, and the continuations of the x and y axes are also depicted by dashed segments to stress the fact that the axes have changed places.

In Figure 1.6.2 the two parts of Figure 1.6.1 (the solid curve and the dashed) have been separated; in Figure 1.6.2*b* the abscissas are denoted by x and the ordinates by y (as is custo-

[9.1] The word "asymptotics" is of the same origin as "asymptote" since the statement that the function $y = 1/x$ has two asymptotes $y = 0$ and $x = 0$ characterizes the behavior of this function for small and large absolute values of x.

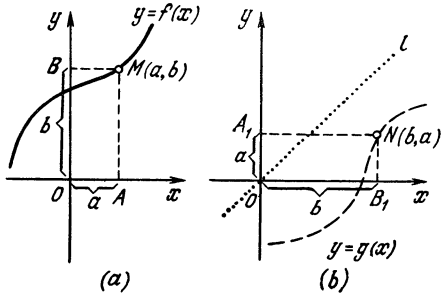


Figure 1.6.2

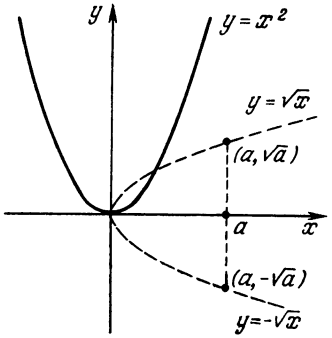


Figure 1.6.4

mary). We see that while in Figure 1.6.2*a* the value $a = OA$ of the independent variable corresponds to $b = OB$ of the function (where $b = f(a)$), in Figure 1.6.2*b*, depicting the graph of g, we have $a = g(b)$. A rotation through 180° about the bisector l maps the segments $OA = a$ and $OB = b$ of Figure 1.6.2*a* into the segments $OA_1 = a$ and $OB_1 = b$ of Figure 1.6.2*b*. Thus *the graphs of two functions each of which is the inverse of the other are symmetric about the bisector l of the first and third quadrant angles.* For example, Figure 1.6.3 depicts graphs for a pair of such functions: $y = 3x + 2$ and $y = x/3 - 2/3$. In particular, if the graph of a function is symmetric about the bisector l the function coincides with its inverse. Such are, say, the functions $y = 1 - x$ and $y = 1/x$ (see Figure 1.4.2). Indeed, $y = 1 - x$ implies that $x = 1 - y$, while $y = 1/x$ implies that $x = 1/y$.

Let us now turn to Figure 1.6.4, which depicts the graphs of the functions $y = x^2$ (the solid curve) and $y = \sqrt{x}$ (the dashed curve). The reader can see that the function $y = \pm\sqrt{x}$, which is the inverse of the function $y = x^2$, is two-valued: to each positive value of the independent variable x there correspond two values of the function, $y = \sqrt{x}$ and $y = -\sqrt{x}$. On the other hand, there are no values of y corresponding to negative values of x, since the graph of the function $y = \pm\sqrt{x}$ lies entirely in the right half-plane corresponding to x positive. Therefore, the statement that to each value of x there corresponds a value of the inverse function y will be incorrect in the given case, since there are values of x ($x < 0$) for which no values of y exist, while for other values of x ($x > 0$) we have even an excess of values of y (two values of y for each value of $x > 0$).

The situation with the curve of a function $f(x)$ depicted in Figure 1.6.5 is even more complicated (the dashed curve corresponds to the graph of the inverse function g). Here to a value of, say, $x = a$ there correspond *four* val-

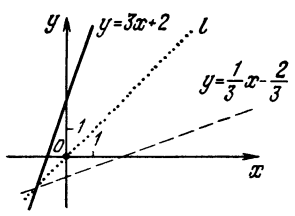


Figure 1.6.3

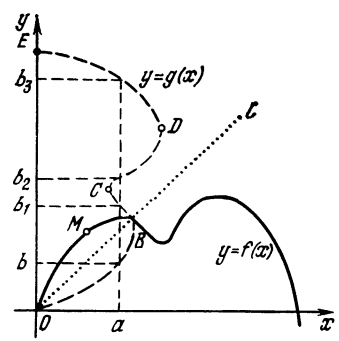


Figure 1.6.5

ues of $g(x)$, namely, b, b_1, b_2, and b_3. This, however, is not surprising—there is no rule by which a function that is the inverse of a given function $f(x)$ must be single-valued. If we take the example of the train discussed above, the points (stations) through which the train passes (the position of each point is defined by the distance z from the initial point O) are determined uniquely by the value of t (time), and it is this dependence $z = z(t)$ that is given in the timetable for the engine-driver. But if a given suburbian train does several runs a day, then it passes a certain point on its trip back and forth several times in the course of the day, which implies that the inverse dependence of time on distance, $t = t(z)$, is *multiple-valued*—to one value of z (the distance of a point from the initial point) there correspond several moments in time when the train is exactly at a point distant z from O.

So where does the difference lie between a function like $y = 3x + 2$ (see Figure 1.6.3; here the inverse function is single-valued) and a function like $y = x^2$ (see Figure 1.6.4; the inverse function is two-valued) and, all the more, a function like the one depicted in Figure 1.6.5 (the inverse function is multiple-valued)? The answer is obvious. The inversion, so to say, of a function, that is, finding the inverse of a given function, consists in reconstructing the values of abscissas from the respective values of the ordinates, say, reconstructing the time t from the distance z traveled by a train. We are lucky if to each value of the function y there corresponds exactly *one* value of the independent variable x, that is, if the inverse function is single-valued, but such good "behavior" rarely happens. However, an inverse function is always single-valued if the initial (single-valued) function $y = f(x)$ is *monotonic*, that is, it either always increases or always decreases. Here, as the independent variable x changes, the function runs through new values of y and to each such value of y there corresponds only one value of x. For example, there is no difficulty in determining the inverse of $y = 3x + 2$, since this (linear) function increases monotonically. The same is true for the dependence of the distance traveled by a train on the time, $z = f(t)$, if the train moves without stops and in one direction. But if the direction in which the function changes is reversed (say, the train starts moving in the opposite direction), for example, the function first decreases, as $y = x^2$, and then, after passing through the minimum point, begins to increase, each value of the function is passed a second time, and therefore we cannot guarantee that it is a certain value of x that corresponds to a given value of y, since there can be several values of x corresponding to a single value of y (two values in the case of the function $y = x^2$). Finally, if we take the function $y = a$ (see Figure 1.3.6), there is no way in which we can define the inverse, since y does not depend on x and, hence, x cannot be reconstructed from y; in other words, the function $y = a$ has no inverse.

In the case of other functions the situation is not as bad as one would think. If we wish to construct a single-valued function that is the inverse of, say, $y = x^2$, we need only confine ourselves to one of the monotonic parts of this function, say consider only positive values of the independent variable x (in Figure 1.6.4 this corresponds to the part of the curve $y = x^2$ lying in the first quadrant). This monotonic section of the function has a single-valued inverse (the graph of the single-valued inverse function $y = \sqrt{x}$ corresponding to this section is depicted by the dashed curve lying in the first quadrant of Figure 1.6.4; it is this function that is known as the *arithmetic value* of the square root of x and is denoted by $\sqrt{x}$).[1.10] Similarly, if only

[1.10] One must bear in mind, however, that the quantity $+\sqrt{x}$, where by $\sqrt{x}$ we mean the arithmetic value of the root is not

a point M (x_0, y_0) on the graph is not a maximum or a minimum of a function $y = f(x)$, like point M in Figure 1.6.5, then in the neighbourhood of point M a monotonic interval can be specified.[1.11] To this interval there corresponds a single-valued branch of the inverse function g (see the parts of the solid curve and dashed curve from point O to point B in Figure 1.6.5). In general, a multiple-valued function, like the function $y = g(x)$ depicted in Figure 1.6.5 by a dashed curve, usually splits into separate single-valued branches (in our case these are the branches depicted in Figure 1.6.5 by arcs OB, BC, CD, and DE). The difficulty here may arise only when we wish to choose the "principal" branch, but from the viewpoint of mathematics this question is irrelevant.

Exercises

1.6.1. What are the inverses of (a) $y = 2x + 4$, (b) $y = x^2 - 2$, (c) $y = x^3 + 3x^2 + 3x$, and (d) $y = x^4 + 2x^2$?

1.6.2. Employ Figures 1.5.2, 1.5.3, 1.5.6*a*, and 1.5.6*b* to construct the graphs of the functions that are the inverse of (a) $y = x^3 + x$, (b) $y = x^3 - x$, (c) $y = x^{-2} = 1/x^2$, and (d) $y = x^{-3} = 1/x^3$. Separate, where possible, the intervals on which the functions are monotonic, which permits arriving at single-valued inverse functions.

1.6.3. Find the functions that are the inverses of (a) an arbitrary linear function $y = ax + b$, (b) an arbitrary quadratic function $y = ax^2 + bx + c$, and (c) the function $y = (ax + b)/(cx + d)$.

1.6.4. Suppose that $f(x)$ and $g(x)$ are two functions each of which is the inverse of the other, with f defined on the interval $a \leqslant x \leqslant b$ and g on the interval $\alpha \leqslant x \leqslant \beta$ where $f(a) = \alpha$ and $f(b) = \beta$ and the function $y = f(x)$ increases monotonically on the interval $a \leqslant x \leqslant b$. Prove that (a) $f(g(x)) = x$, $g(f(x)) = x$, and (b) the functions $F(x) = f(f(x))$ and $G(x) = g(g(x))$ constitute another pair of such functions. What are the domains of these two functions?

the complete inverse of $y = x^2$, only the *two-valued* function $y = \pm\sqrt{x}$ is the exact inverse of $y = x^2$.

[1.11] In this connection one speaks of the function $y = f(x)$ as being *locally monotonic* at point M, that is, locally monotonic in a certain neighbourhood of point M.

1.7 Transforming Graphs of Functions

Above we repeatedly encountered the problem of transforming graphs, that is, the transformation of one graph into another graph that is similar to the initial graph. We will now return to this problem having in mind a more systematic approach.

The simplest case is the translation of a curve. Suppose that we have the graph of a function $y = f(x)$ and we would like to know the shape of the curve obtained through shifting this graph b units upward. The answer is almost obvious: it is clear that if we consider two curves, $y = f(x)$ and $y = f(x) + b$, then to each point $A(x_0, y_0)$ of the first graph there corresponds a point A_1 (x_0, y_1) of the second graph, which lies b units above point A (see Figure 1.7.1, with b positive). Putting it differently, we can say that if we have two functions, $y = f(x)$ and $y - b = f(x)$, then the graph of the second is b units higher than the graph of the first function, that is, replacing y with $y - b$ in the equation of a curve is equivalent to shifting the curve b units upward.

It could seem that there is no sense in formulating the same statement in two different ways that differ very little from each other, namely, lifting a curve means performing the following transformation of the equation of the curve: go over from $y = f(x)$ to $y = f(x) + b$ or from $y = f(x)$ to $y - b = f(x)$. However, the second formulation, which involves replacing y with $y - b$ is more convenient than the first when the curve is specified implicitly, that is, by an equation of

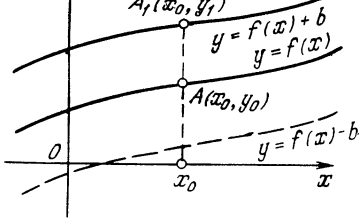


Figure 1.7.1

the type $F(x, y) = 0$, which is not resolved for y. For instance, the equation of a circle S of radius 1 with center at the origin is conveniently written as

$$x^2 + y^2 = 1,$$

or

$$F(x, y) = x^2 + y^2 - 1 = 0 \qquad (1.7.1)$$

(compare with formula (1.2.1) for the distance OA, with $A = A(x, y)$). Replacing y with $y - b$, we get

$$x^2 + (y - b)^2 - 1 = 0,$$

which is the equation of a unit circle centered at the point $Q_1(0, b)$, an equation obtained from S by shifting S upward by b units (Figure 1.7.2).

Similarly (as was discussed in great detail in Section 1.4), replacing in the equation $F(x, y) = 0$ the independent variable x with $x - a$ is equivalent to shifting the curve a units to the right. For instance, the equation $(x - a)^2 + y^2 - 1 = 0$ defines a circle of radius 1 centered at point $Q_2(a, 0)$, that is, an equation obtained from S by shifting S rightward a units (see Figure 1.7.2).

It is clear that the quantities a and b may be negative; for instance, transforming the equation $y = f(x)$ into the equation $y = f(x) - b$ (where, as usual, b is positive) or replacing y in the

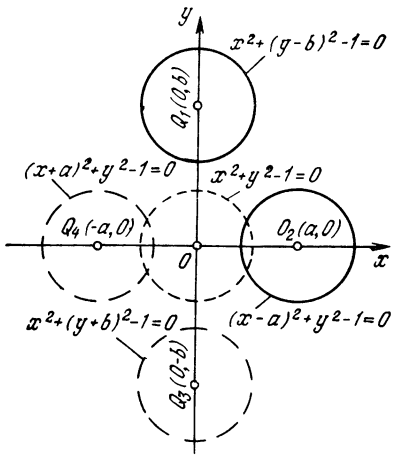


Figure 1.7.2

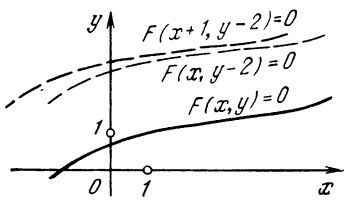


Figure 1.7.3

equation of the curve with $y + b$ means that we shift the curve *downward* b units (the dashed curve in Figure 1.7.1). Similarly, the curve specified by the equation $F(x + a, y) = 0$ is obtained from the curve $F(x, y) = 0$ by shifting the latter a units to the *left*. For instance, the centers of the circles $x^2 + (y + b)^2 - 1 = 0$ and $(x + a)^2 + y^2 - 1 = 0$ are the points $Q_3(0, -b)$ and $Q_4(-a, 0)$ (see Figure 1.7.2). These translations may all be combined; for instance, the curve $F(x + 1, y - 2) = 0$ is obtained from the curve $F(x, y) = 0$ by shifting the latter one unit to the left and two units upward (Figure 1.7.3).

Example. Let us consider the *linear-fractional* function

$$y = \frac{ax + b}{cx + d}, \quad c \neq 0. \qquad (1.7.2)$$

Many experimentally established relationships obey, either exactly or approximately, this function, whereby it is important to know how to simplify the function and construct its graph.

It is clear that if the numerator and denominator of the fraction on the right-hand side of (1.7.2) are such that $a \div c = b \div d$, or $ad = bc$, the function specified by (1.7.2) is a constant, or $y = k$, where $k = a/c = b/d$; whereby the only interesting case is where $ad \neq bc$. We state that in this case the graph of the function (1.7.2) is a *hyperbola* and, hence, the function (1.7.2) represents inverse proportionality of two quantities.

We leave it to the reader to examine the general case (see Exercise 1.7.6) and consider the particular case

$$y = \frac{19 - 6x}{2x - 5}. \qquad (1.7.2a)$$

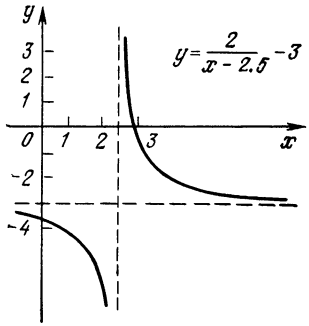


Figure 1.7.4

In the right-hand side we isolate the integral part:

$$y = \frac{(15-6x)+4}{2x-5} = \frac{15-6x}{2x-5} + \frac{4}{2x-5}$$

$$= -3 + \frac{4}{2x-5}.$$

The last equation can be rewritten thus:

$$y+3 = \frac{2}{x-5/2}, \qquad (1.7.2b)$$

from which we see that the curve (1.7.2a), or (1.7.2b), is the graph of inverse proportionality (with coefficient 2) between $y + 2$ and $x - 5/2$ (Figure 1.7.4). This completes the solution of the problem.

Now let us see how the equation of a curve must be changed so that all vertical dimensions (along the y axis) are increased c-fold.[1.12] Obviously, in place of the equation $y = f(x)$ we must take a new equation $y = cf(x)$ (we will write $y_0 = f(x)$ and $y_1 = cf(x)$, since these are two different curves). Then for the same x the quantity y_1 will be c times greater than before, that is to say, c times y_0, and the curve will be stretched in the vertical direction c-fold.

As an example, recall the equations of straight lines passing through the origin. The equation of the bisector of the first and third quadrant angles is $y_0 = x$. The equation $y_1 = 10x$ corresponds to a straight line that is more steeply slanted: for a given x the ordinate is 10 times greater (see Figure 1.3.4).

The law by which $y_0 = f(x)$ is transformed into $y_1 = cf(x)$ may also be described thus: in the equation of the curve $y_0 = f(x)$ replace y_0 by y_1/c, that is, write $y_1/c = f(x)$. Then the dependence of y_1 on x (the new y) is characterized by the fact that the curve $y_1(x)$ is elongated c-fold vertically as compared to the curve $y_0(x)$ (the old y).

Again the need for two formulations whose equivalence is quite obvious (indeed, it is clear that the equations $y = cf(x)$ and $y/c = f(x)$ are the same equations) follows from the fact that it is more convenient to replace y with y/c if we are forced to "stretch" c-fold away from the x axis a curve whose equation $F(x, y) = 0$ is not resolved for y. Just substitute y/c for y in the new equation, and we have the sought result: $F(x, y/c) = 0$.

Let us again turn to the unit circle S whose equation is $x^2 + y^2 - 1 = 0$, that is, a unit circle centered at the origin. How would you write the equation of the curve obtained by stretching the circle along the vertical axis 3-fold (Figure 1.7.5; the new curve is denoted by $y_1(x)$)? By the rule that we have just stated, in the equation of the cir-

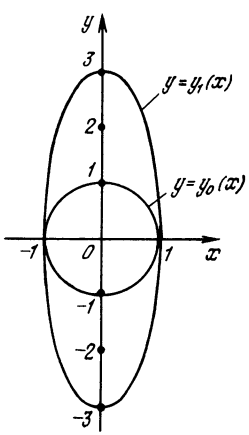


Figure 1.7.5

[1.12] For the sake of simplicity, we from now on assume that c is greater than unity; the case with c less than unity (and even negative) we discuss below.

cle we must replace y with $y/3$. This yields

$$x^2 + \left(\frac{y_1}{3}\right)^2 - 1 = 0 \qquad (1.7.3)$$

(here we write y_1 instead of y to distinguish between curve (1.7.3) and circle (1.7.1)).

The curve obtained as a result of stretching the unit circle away from the horizontal (or vertical) diameter (see Exercise 1.7.5) is an example of an ***ellipse***.[1.13] We have therefore transformed circle (1.7.1) (in Figure 1.7.5 this cicrle is denoted by $y_0(x)$) into an ellipse, (1.7.3).

In the given case Eqs. (1.7.1) and (1.7.3) can easily be solved: $y_0 = \sqrt{1 - x^2}$ and $y_1 = 3\sqrt{1 - x^2}$, which clearly shows that $y_1 = 3y_0$ for equal x. But the rule, or statement, by which replacing y with y/c leads to a c-fold stretching of the curve along the vertical axis holds true also for curves defined by a complicated equation $F(x, y) = 0$, that is, an equation that cannot be resolved algebraically for y, say

$$x + y \log y = 0. \qquad (1.7.4).$$

The curve that results from stretching curve (1.7.4) 3-fold along the vertical axis is described by the following equation

$$x + \frac{1}{3} y \log\left(\frac{y}{3}\right) = 0. \qquad (1.7.4a)$$

The statement concerning the replacement of y with y/c is readily extended to the x-coordinate. When we replace x_0 with x/c in the equation of the curve, $F(x_0, y) = 0$, that is, when we go over to the equation $F(x_1/c, y) = 0$, the initial curve stretches along the x axis c-fold, which is to say, for equal y the value of x_1 is c times the value of x_0.

We begin with examples instead of a proof: $y = x_0$ and $y = x_1/10 = 0.1x_1$

[1.13] The stretching coefficient, or ratio, may be equal to 1 (of course, such an identity transformation does not change the shape of a curve); accordingly with this a circle is usually considered as a particular case of an ellipse.

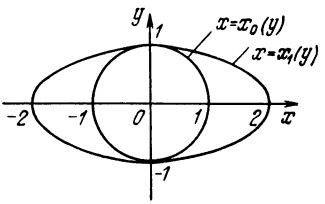


Figure 1.7.6

(see Figure 1.3.4). The first straight line slants at an angle of 45° to the x axis, the second straight line is less steep.

Another illustration: $x_0^2 + y^2 - 1 = 0$ and $(x_1/2)^2 + y^2 - 1 = 0$. The first equation corresponds to a circle S of radius 1 centered at the origin, while the second to the curve obtained from that circle via 2-fold stretching along the x axis (Figure 1.7.6). It is clear that the new curve is an *ellipse*. The proof of this is almost obvious. If we solve the equations of the first and second curves,

$$y = f(x_0) \text{ and } y = f(x_1/c), \qquad (1.7.5)$$

for x, we obtain

$$x_0 = \varphi(y) \quad \text{and} \quad x_1/c = \varphi(y),$$

$$\text{i.e.} \quad x_1 = c\varphi(y) = cx_0, \qquad (1.7.5a)$$

where φ is the inverse of function f (see Section 1.6). This corresponds to the initial statement that substitution of x/c for x stretches the curve along the x axis c-fold. The important thing here is that f is one and the same function in formulas involving x_0 and x_1, (1.7.5). Therefore φ is also the same in the formulas involving x_0 and x_1, (1.7.5a).

Here is another example:

$$y = 10^{x_0} \quad \text{and} \quad y = 10^{x_1/2}. \qquad (1.7.6)$$

The inverse of a power function is a logarithmic function, whereby

$$x_0 = \log y \quad \text{and} \quad x_1/2 = \log y,$$

$$\text{i.e.} \quad x_1 = 2 \log y. \qquad (1.7.6a)$$

Thus, the graph of function $y = 10^{x/2}$ is obtained from the graph of function

$y = 10^x$ by a 2-fold stretching of the latter along the x axis.

What do we do if in the equation $y = f(x)$ the quantity kx is substituted for x? To take advantage of the above-stated rule, let us recall that multiplication by k is the same as division by $1/k$, since $kx = x/(1/k)$. Hence, $1/k$ plays the role of c in the earlier formulas, and if, say $k = 1/2$, then $c = 1/k = 2$. This means that the substitution of $0.5x_1$ for x_0 is the same as replacing x_0 with $x_1/2$ and leads to a stretching of the curve along the x axis by a factor of 2. If $k = 3$, then $c = 1/k = 1/3$, and the replacement of x with $3x$ is the substitution of $x/(1/3)$ for x.

What do all these substitutions mean geometrically? In the case where $c > 1$ the result can be stated thus: when, in the equation of the curve, y is replaced with y/c, the curve is stretched vertically by a factor of c, while in the substitution of x/c for x the curve is stretched horizontally by a factor of c.

When $0 < c < 1$, nothing really changes in our reasoning, with the exception that the word "stretching" cannot be used, since stretching 2-fold means increasing the dimensions by a factor of 2, while stretching "1/3-fold" means multiplying the dimensions by 1/3, which means not multiplying but dividing by 3, or "shrinking" the curve (in a certain direction) 3-fold. Hence, when we substitute y/c for y, the vertical dimensions change by a factor of c, which results in the curve "shrinking" in the vertical direction; for instance, at $c = 0.5$, a transition from curve $y = f(x)$ to curve $y = cf(x) = 0.5f(x)$ means an actual reduction in height by one half. The same goes for the substitution $x \to x/c$; for $0 < c < 1$ this substitution amounts to a "shrinking" of the curve along the x axis.

Now let us establish the meaning of the substitutions $y \to y/c$ and $x \to x/c$ when c is *negative*. Such a replacement can be conducted in two stages. We introduce $c = -b = (-1)\,b$, where b is positive, and carry out subsequently

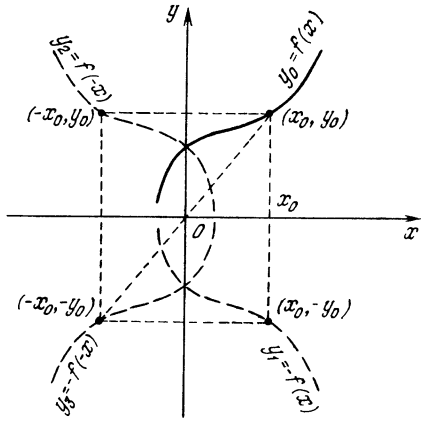


Figure 1.7.7

two substitutions:

$$y_0 \to \frac{y_1}{b} \to \frac{y_2}{(-1)\,b}$$

$$\left(= -\frac{y_2}{b}, \text{ so that } y_2 = -y_1\right).$$

The first operation, substitution of y_1/b for y_0, where $b > 0$, has already been discussed—it leads to a change in the vertical dimensions by a factor of b. What remains to be clarified is the meaning of the change in sign of y, that is, the meaning of the substitution of $-y$ for y. Obviously, the points of the curve $y_0 = f(x)$ and $y_1 = -f(x)$ corresponding to each other lie symmetrically about the x axis (Figure 1.7.7). Therefore, the curve $y_1 = -f(x)$ is obtained from the curve $y_0 = f(x)$ as the mirror reflection of the latter with respect to the x axis, and the same can be said of the curves $F(x, y_0) = 0$ and $F(x, -y_1) = 0$, where the second equation is obtained from the first by reversing the sign of y.

Reasoning in a similar manner, we can say that since points (x, y) and $(-x, y)$ are symmetric with respect to the y axis, the replacement of x with $-x$ in the equation of a curve (the transition from curve $F(x, y_0) = 0$ to curve $F(-x, y_2) = 0$) replaces the initial curve with a curve symmetric to the initial one about the y axis. Finally, the transition from equation

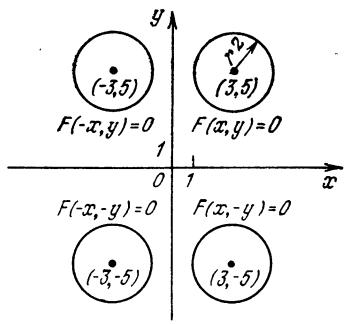


Figure 1.7.8

$F(x, y_0) = 0$ to equation $F(-x, -y_3) = 0$ reduces to successive alternations of the sign of y (symmetry with respect to the x axis) and of the sign of x (symmetry with respect to the y axis). But two successive reflections, from the x axis and the y axis, are equivalent to a reflection with respect to the origin $O(0, 0)$ (see Figure 1.7.7). Thus, simultaneous substitution of $-x$ for x and of $-y$ for y in the equation of a curve is equivalent to a symmetry transformation with respect to O.

Notwithstanding the simplicity of the above reasoning, beginners (for which this book is intended) often make mistakes concerning the various transformations we have just discussed. To clarify matters we give an example:

$$F(x, y) = (x - 3)^2 + (y - 5)^2 - 4 = 0.$$

This is the equation of a circle of radius 2 centered at a point with coordinates $x = 3$ and $y = 5$. Figure 1.7.8 shows the initial curve and also the following curves:

$$F(x, -y) = (x - 3)^2 + (-y - 5)^2 - 4 = 0$$

$$F(-x, y) = (-x - 3)^2 + (y - 5)^2 - 4 = 0,$$

$$F(-x, -y) = (-x - 3)^2 + (-y - 5)^2 - 4 = 0$$

The letter F in all four cases denotes one and the same function. See what happens to the curve (circle) under the substitution of $-x$ for x or the substitution of $-y$ for y or under the simultaneous substitutions of $-x$ for x and $-y$ for y. A firm grasp of these rules will make it possible, after you have built a curve $y = f(x)$ or $F(x, y) = 0$, to picture the graphs of all functions of the type

$$\frac{y-b}{d} = f\left(\frac{x-a}{c}\right) \quad \text{or} \quad F\left(\frac{x-a}{c}, \frac{y-b}{d}\right) = 0$$

with arbitrary values of the constants a, b, c, and d.

Here are two more simple (but important) examples. In Figure 1.7.9 we have two curves:

$$y = \sin x \quad \text{and} \quad y = \sin 3x, \qquad (1.7.7)$$

where, of course, x is measured in radians. The second curve has been compressed horizontally by a factor of three.

The relation $y = \sin x$ is a *periodic* function: at $x = 2\pi \simeq 6.3$ radians (which corresponds to an angle of 360°) the sine has the same value as for $x = 0$. (Adding 2π to any angle leaves the value of the sine unchanged.) Thus, the graph of the first function, $y = \sin x$, transforms into itself under a translation along the x axis (in either direction) by 2π, which simply means that $\sin x$ is a periodic function with a period of 2π. The function $y = \sin 3x$ is also periodic, but the period here is less by a factor of 3. If x varies by $2\pi/3 \simeq 2.1$ rad, $3x$ (the angle whose sine is laid off on the axis of ordinates) varies by 2π, and $\sin 3x$ returns to the same value:

$$\sin 3x = \sin 3\left(x + \frac{2\pi}{3}\right).$$

Use this example to think over the general assertion that the substitution of kx for x in a periodic function re-

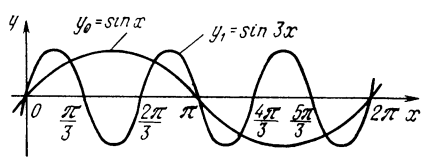


Figure 1.7.9

sults in the decrease in the *period* by a factor of k and an increase in the *frequency* by the same factor (the frequency is the number of periods per unit length).

The second example deals with logarithmic (and exponential) functions. Take two curves:

$$y_1 = \log_a x \quad \text{and} \quad y_2 = \log_b x, \tag{1.7.8}$$

or, which is the same,

$$x = a^{y_1} \quad \text{and} \quad x = b^{y_2}, \tag{1.7.8a}$$

where it is assumed that both a and b are greater than unity. It can be established, by a method similar to that employed in the analysis of (1.7.6) and (1.7.6a) (see Section 1.4.9) that the second curve in (1.7.8) is obtained from the first by stretching the latter along the y axis by a factor $c = \log_b a = (\log_a b)^{-1}$, where c is the *modulus* of base a with respect to base b. Similarly, any two exponential functions, $y = a^x$ and $y = b^x$, are related in the same manner: the second is obtained from the first by stretching the first curve along the x axis by a factor $c = \log_b a$.

Two simple but highly important concepts are related to what we have just said. A curve given by the equation $F(x, y) = 0$ that retains its form under the substitution of $-x$ for x (say, the parabola $y = ax^2$) is symmetric about the y axis. A curve whose equation retains its form under the substitution of $-x$ for x and $-y$ for y (say, the hyperbola $y = 1/x$ or the cubical parabola $y = x^3$) is symmetric with respect to the origin. Finally, a curve whose equation $F(x, y) = 0$ maps into itself under the substitution of $-y$ for y (say, the semicubical parabola $y^2 = x^3$; see Figure 1.5.7) is symmetric with respect to the x axis. All these statements follow directly from the geometrical meaning of the respective substitutions.

A function $y = f(x)$ whose graph is symmetric with respect to the y axis (Figure 1.7.10*a*) is said to be *even*, while if the graph of a function is symmetric with respect to origin O (Figure 1.7.10*b*), the function is said to be *odd*. These names originate in the fact that parabolas of even order, $y = ax^{2m}$, are even functions, while parabolas of odd order, $y = ax^{2m+1}$, are odd functions (cf. Section 1.5).

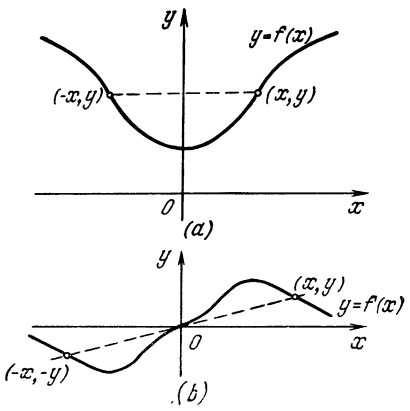


Figure 1.7.10

Finally, we note another important fact. It is necessary to understand (this is often a neglected aspect) that in applied problems the procedure in which all ordinates y or abscissas x are changed in a given ratio (and nothing more) is related to a change in the units of measurement of y or x, so that the graphs obtained as a result of such a change of scale represent one and the same dependence of y on x to within, so to say, a change in scale (or the system of units).[1.14]

Suppose that on the axis of abscissas, Ox, we plot time and on the axis of ordinates, Oy, we plot distance (Figure 1.7.11); the function $y = f(x)$ (or, as we will often write, $z = f(t)$) fixes the law of motion. But since y and x are measured in different units, no comparison of the two quantities is possible (indeed, there is no way that we can compare 1 cm with 1 s or say which of the two quantities is great-

[1.14] However, if x and y are measured in the *same* units, the substitution of cy for y has a physical meaning, since the functions $y = f(x)$ and $y = cf(x)$ are different.

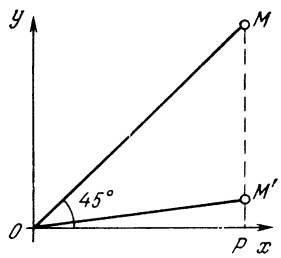


Figure 1.7.11

er). The statement, say, that for a given point M (x, y) the segment OM lies at an angle of 45° to the axis of abscissas (see Figure 1.7.11) is meaningless, that is, if we change the scale of the y axis (go over from centimeters to meters), point M is replaced with a new point M' corresponding to the same moment in time, but now $\angle POM'$ is not 45°, that is $\angle POM' \neq 45°$, while in all other respects M' is equivalent to point M. Thus, in physical problems the ellipses in Figures 1.7.5 and 1.7.6 do not differ from the circles depicted in the same figures, and if we are interested in a particular curve (ellipse), then it is wise to select units of measurement such that the ellipse is simply a circle.

Similarly, any parabola $y = ax^2 + bx + c$ can be reduced, as we have seen earlier, to a parabola $y_1 = ax_1^2$ by a translation, or shift (i.e. by a transition from coordinates x and y to a set of new coordinates, $x_1 = x + p$ and $y_1 = y + q$, where p and q must be chosen appropriately). But physically such a translation means nothing else but a change in position of the origin from which the physical quantities x and y are reckoned, say the initial moment $t = 0$ and the initial position $z = 0$, and so does not influence in any way the physical process. After this we can transform $y_1 = ax_1^2$ into the "unit" parabola $Y = \pm X^2$ (or even to $Y = X^2$ if we select the positive direction along the y axis[1.15]) by choosing a new system of units along the y_1 axis (changing the scale on this axis); it is in this sense that, as mentioned in footnote 1.5, the parabola $y = x^2$ can be considered as the general case of a parabola.

In the same way, any linear-fractional function $y = (kx + l)/(mx + n)$ can be reduced, by changing the position of the origin on the x and y axes and the scale on the y axis, to the "unit" hyperbola" $Y = 1/X$ (see Exercise 1.7.6). Similarly, any cubic function (polynomial) $y = ax^3 + bx^2 + cx + d$ can be reduced to one of the three functions: $y = x^3 + x$, $y = x^3 - x$, or $y = x^3$ (no further reduction is possible). The list of examples can be continued.

Let us discuss a more complicated example involving cubic functions (polynomials):

$$y = ax^3 + bx^2 + cx + d$$
$$= a\,(x^3 + px^2 + qx + r), \qquad (1.7.9)$$

where $p = b/a$, $q = c/a$, and $r = d/a$ (of course, $a \neq 0$). In Section 1.5 we saw that the substitution $X = x + p/3$ transforms (1.7.9) into

$$y = a\,(X^3 + PX + Q), \qquad (1.7.9a)$$
$$\text{or } Y = aX^3 + BX = a\,(X^3 + kX),$$

where $Y = y - aQ$, $B = aP$, and $k = B/a\ (= P)$, with $P = q - p^2/3$ and $Q = (2/27)\,p^3 + r - (1/3)\,pq$ (see formulas (1.5.6) and (1.5.6a)). At $k = 0$ (i.e. if $B = 0$) the function (1.7.9a) has the form $Y = aX^3$, or

$$y_1 = x_1^3, \qquad (1.7.10a)$$

where $x_1 = X$ (the X-coordinate does not change) and $y_1 = Y/a$. If $k > 0$ (i.e. B and a are of the same sign), we put $x_1 = X/\sqrt{k}$, i.e. $X = \sqrt{k}x_1$. Then (1.7.9a) becomes

$$Y = a\,(k^{3/2}x_1^3 + k^{3/2}x_1) = ak^{3/2}\,(x_1^3 + x_1),$$

that is, we have the law

$$y_1 = x_1^3 + x_1, \qquad (1.7.10b)$$

where $y_1 = Y/ak\sqrt{k}$. Finally, if $k < 0$ (a and B are of opposite sign), a substitution that is similar to the one used above, $x_1 = X/\sqrt{|k|}$, that is, $X = \sqrt{|k|}\,x_1$, transforms (1.7.9a) into

$$Y = a\,(|k^{3/2}|x_1^3 + k|k|^{1/2}x_1),$$

[1.15] While in the case of time, t, we have a natural direction of growth, which corresponds to the simple fact that the future and the past are not equivalent, for distance, z, this is not the case, and we are free to choose any direction for the increase in z.

that is,

$$y_1 = x_1^3 - x_1, \tag{1.7.10c}$$

where $y_1 = -Y/ak\sqrt{|k|}$. Thus, for a physicist any scientific function of the type (1.7.9) that links two quantities of different dimensions, y and x, is equivalent to one of the three functions (1.7.10a)-(1.7.10c) (see Figures 1.5.1*a*, 1.5.2, and 1.5.3 and Exercises 1.7.8 and 1.7.9).

Let us use the following example to illustrate the aforesaid. It is well known that the time dependence of the altitude h of a stone thrown upward from a tower of height w with an initial velocity v is given by the formula

$$h = w + vt - \frac{g}{2}t^2, \tag{1.7.11}$$

where $g \simeq 9.8$ m/s^2 is the acceleration of gravity. If, however, for the zero level we take not the ground level but the greatest altitude reached by the stone, h_0, and for time zero we take the moment t_0 when this altitude is reached, then in the new coordinates, $h_1 = h - h_0$ and $t_1 = t - t_0$, Eq. (1.7.11) describing the motion of the stone simplifies considerably:

$$h_1 = -\frac{gt_1^2}{2}. \tag{1.7.11a}$$

But the coefficient $-g/2$ in Eq. (1.7.11a) has nothing to do with the physics of the process and is related solely to the choice of the system of units. Equation (1.7.11a) can be simplified still further by dropping the minus sign on the right-hand side; for this we need only agree that h is measured from point $h = h_0$ not upward but downward. Next, we can go over to a "natural" system of units in which the acceleration of gravity is equal to 2 (time can be measured in seconds, as usual, while distance can be measured in units of $g/2$, that is, approximately, 4.9 m). Then, in the new coordinates, $H = -(2/g)(h - h_0)$, $T = t_1 = t - t_0$, and the law (1.7.11) or (1.7.11a) assumes the "canonical" form $H = T^2$.

In a similar way we can always transform logarithmic and exponential functions often encountered in scientific laws into such a form that the bases acquire a convenient form, since the transition from one base to another means actually changing only the units of measurement of the quantities (cf. the above material on logarithmic and exponential curves). Therefore, each time we encounter the logarithmic function $y = \log_e x$, we can freely assume the base to be equal to 10 or to $e \simeq 2.272$ (see Section 1.4.9) or even to two;[1.16] when exponential functions are considered, $y = d^x$, the base d is usually assumed equal to e.

Exercises

1.7.1. Construct the curves $\frac{x^2}{4} + \frac{y^2}{9} - 1 = 0$, $\frac{(x+3)^2}{4} + \frac{(y-5)^2}{9} - 1 = 0$, and $\frac{(x-3)^2}{9} + \frac{(y+5)^2}{4} - 1 = 0$ (it is natural here to employ the fact that $x^2 + y^2 - 1 = 0$ is the equation of a unit circle centered at the origin).

1.7.2. Construct the curve $y = \sin x$ by taking, for example, eight values of x in the interval from $-\pi$ to $+\pi$ by 0.25π each. Employing this graph, sketch the graphs of the functions (a) $y = 2 \sin x$, (b) $y = \sin 0.5x$, (c) $y = 3 \sin 3x$, (d) $y = \cos x$, (e) $y = \cos x + \sin x$ $(= \sqrt{2} \sin (x + \pi/4))$, (f) $y = \cos^2 x$ $(= 1/2 + (1/2) \cos 2x)$, (g) $y = \sin^2 x$ $(= 1/2 - (1/2) \cos 2x)$. [*Hint.* All these curves may be obtained via translation, stretching, and shrinking the sinusoid $y = \sin x$; to solve Exercises 1.7.2d, f, g, employ the identity $\cos x = \sin (x + \pi/2)$.]

1.7.3. Plot the following curves: (a) $y = \pm\sqrt{x^2 - 1}$, (b) $y = \pm\sqrt{x^2 + 1}$, (c) $y = 2 \pm \sqrt{(x-1)^2 - 1}$, and (d) $4y^2 + 4y - x^2 = 0$. [*Hint.* Construct curve (a) using twenty values of x in the interval from -5 to $+5$ with each value differing from the adjacent one by 0.5 (and each being a multiple of 0.5). Transform the equations of the curves (a) and (b) to $y^2 - x^2 + 1 = 0$ and $x^2 - y^2 + 1 = 0$, respectively, and note that (b) is obtained from

[1.16] For instance, in the ***theory of information***, where the amount of information is expressed by logarithmic functions of the various variables, all logarithms are assumed *binary*, which is of no importance, of course, and is connected solely with the system of units employed (with the fact that information is measured in *bits*). The interested reader can refer to the book of A.M. Yaglom and I.M. Yaglom, *Probability and Information*, Reidel, Dordrecht, 1983.

(a) by interchanging x and y. To construct curve (d), transform the equation of this curve into $4(y + 1/2)^2 - x^2 - 1 = 0$ and then, by translating and shrinking curve (b), you will arrive at the result.]

1.7.4. Write the equations of the curves obtained from the curve $y = x^3 - x$ (see Figure 1.5.3) by (a) shrinking along the axis Oy with a ratio 1/3 (i.e. diminishing all vertical dimensions 3-fold), (b) substituting $x/2$ for x, and (c) substituting $-y/2$ for y. Sketch the three curves.

1.7.5. Prove that stretching (or shrinking) the unit circle $x^2 + y^2 - 1 = 0$ to any straight line with any ratio results in an ellipse. [*Hint.* This problem may be formulated as follows: stretching (or shrinking) from a circle $(x - a)^2 + (y - b)^2 - r^2 = 0$ away (or toward) the x axis, that is, substituting ky for y, transforms the circle into an ellipse.]

1.7.6. Prove that the graph of an arbitrary linear-fractional function $y = (ax + b)/(cx + d)$ with $c \neq 0$ and $\Delta = ad - bc \neq 0$ may be obtained from the graph of inverse proportionality $y = k/x$, where $k = -\Delta/c^2$, by shifting the latter to the left by d/c units (or by $|d/c|$ units to the right if $d/c < 0$) and upward by a/c units.

1.7.7. Prove that every function may be represented by the sum of an odd and even function. [*Hint.* Employ the identity

$$f(x) = \frac{1}{2}(f(x) + f(-x)) + \frac{1}{2}(f(x) - f(-x)).]$$

1.7.8. Prove that no two curves in (1.7.10a)-(1.7.10b) can be reduced to each other by applying substitutions of the form $x' = ax + p$ and $y' = by + q$ (where, of course, $a \neq 0$ and $b \neq 0$), which are equivalent to changing the position of the origin for x and y (the translation of the origin) and the units of measurement for x and y.

1.7.9. Find the rule that enables us to establish from the coefficients a, b, c and d whether (1.7.9) can be reduced to (1.7.10a) or (1.7.10b) or (1.7.10c).

1.8 Parametric Representation of a Curve

Let us now study the motion of a material point, $M = M(x, y)$. Each of the coordinates, x and y, changes in the course of time t, and the motion of point M is specified by two functions $x = x(t)$ and $y = y(t)$, say,

$$x = \cos t, \quad y = \sin t. \tag{1.8.1}$$

These relations can be depicted graphically as two curves by plotting t

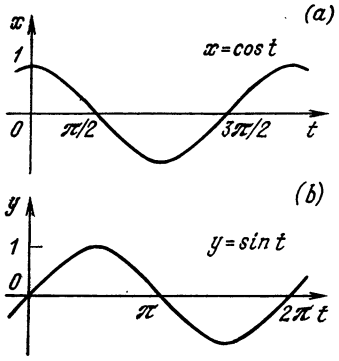


Figure 1.8.1

on the axis of abscissas and x on the axis of ordinates in one drawing, and t on the axis of abscissas and y on the axis of ordinates in the other (Figure 1.8.1).

Let us investigate the *trajectory* of point M. To each value of t there correspond a value of $x(t)$ and a value of $y(t)$. What curve will point $M = M(x(t), y(t)) = M(t)$ describe as t varies? To answer this question, we can eliminate t from the two equations $x = x(t)$ and $y = y(t)$. This yields an expression that will involve only y and x, that is, an equation of the type $y = y(x)$ or $F(x, y) = 0$. Then we construct the curve in the usual fashion by specifying various x's and finding the corresponding y's. For instance, in our example,

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1,$$

that is,

$$y = \pm\sqrt{1 - x^2}, \text{ or } x^2 + y^2 - 1 = 0,$$

so that the curve (1.8.1) is a *circle* of unit radius in the xy-plane (Figure 1.8.2*a*).

However, it often happens that even comparatively simple expressions for $x(t)$ and $y(t)$ lead to such complicated expressions when trying to eliminate t that it makes no sense to tackle them. For instance, if

$$x = x(t) = a_1t^4 + b_1t^3 + c_1t^2 + d_1t + e_1, \; y = y(t) = a_2t^4 + b_2t^3 + c_2t^2 + d_2t + e_2,$$

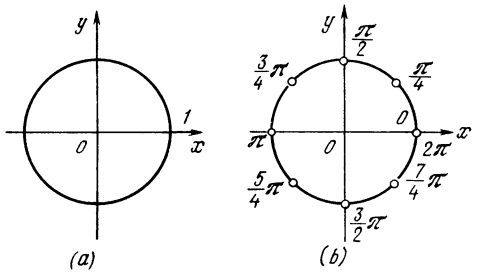


Figure 1.8.2

then to eliminate t we would have to solve a quadratic equation, and this lead to expressions that in view of their unwieldiness can simply not be written. Yet it is possible to construct the curve in the xy-plane without eliminating t: it suffices to specify various values of t and find x and y for each of them. To illustrate this point, we construct the following table of values of x (t) and y (t) corresponding to the simple example (1.8.1):

t	0	$\pi/4$	$\pi/2$	$3\pi/4$	π
$x=\cos t$	1	0.7	0	−0.7	−1
$y=\sin t$	0	0.7	1	0.7	0

t	$5\pi/4$	$3\pi/2$	$7\pi/4$	2π
$x=\cos t$	−0,7	0	0.7	1
$y=\sin t$	−0.7	−1	−0.7	0

It is clearly not necessary to take t greater than 2π because the values of x and y repeat. Using this table, we can plot the points of the curve. In so doing we employ only the values of x and y. Those values of t for which the x's and y's have been computed are no longer needed for plotting the points. "The Moor has done his duty; the Moor can go", as the German proverb has it.

This method of representing a curve or, what is the same thing, of specifying a functional relationship $y = y(x)$, by two functions, x (t) and y (t), is called ***parametric representation***, and t is called a ***parameter***.[1.17] In physical problems, a parameter often has a definite physical meaning. For instance, in the example with which we started this section, t may be the time variable, and in this case both x (t) and y (t) are functions of time. Of course, one can be interested only in the shape of the trajectory described by point $M(x, y)$ but it is also interesting to know the velocity with which the point moves at a certain moment of time and the position of the point at various moments in time. To find these two quantities, we must retain the values of the parameter and for each point $M = M(t)$ write the number t (it goes without saying that only a finite number of points can be plotted on a drawing, e.g. see Figure 1.8.2*b*). In this way, a "parametrized" curve yields more information than the trajectory of point M alone does (compare Figures 1.8.2*a* and 1.8.2*b*).

Let us take the following curves:

(1) $x = \cos t$, $y = \sin t$;

(2) $x = \cos t$, $y = -\sin t$;

(3) $x = \cos 3t$, $y = \sin 3t$;

(4) $x = \sin 3t$, $y = \cos 3t$.

In each case, if we eliminate t, we arrive at the equation of a unit circle in the xy-plane, or $x^2 + y^2 - 1 = 0$. So in what respect are all these four cases different?

In the first three cases, at $t = 0$ the point lies on the axis of abscissas, while in the fourth case it lies on the axis of ordinates. In cases (1) and (3) the point (one can imagine a heavy material point) is moving along a circle counterclockwise, while in cases (2) and (4) it is moving clockwise. In absolute value, the angular velocity of rotation (or the rate of rotation) or

[1.17] The word "parameter" was invented by mathematicians of Ancient Greece for just this purpose and has no exact translation. *Webster's Third New International Dictionary* gives the following definition: "An arbitrary constant characterizing by each of its values some member of a system (as of expressions, curves, surfaces, functions)."

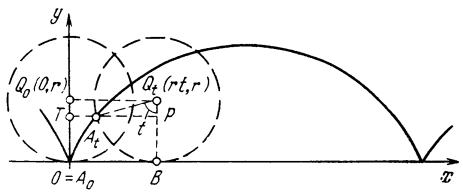


Figure 1.8.3

even the linear speed $|v|$ (since the radius of the circle is equal to unity) is 1 rad/s (i.e. $|v| = 1$ m/s if the radius of the circle is 1 m) in cases(1) and (2), and 3 rad/s (or $|v| = 3$ m/s) in cases (3) and (4).

Here is another example. Let us build the curve that a point on the rim of a bicycle wheel describes when the cyclist is moving along a straight line, the point A on the circumference of a circle of radius r as the circle rolls along a straight line. The curve is called a ***cycloid***.[1.18]

Let us take a horizontal straight line along which our "wheel" is moving as the x axis, and let us assume that at the beginning of the motion, $t = 0$, the center of the wheel, $Q = Q_0$, lies on the y axis at the point $(0, r)$. We will also assume that the wheel rolls with a constant velocity. For instance, let us assume that in unit time the wheel rotates through a unit angle. Then, in the course of time t, the wheel will travel a distance rt in the positive direction of the x axis (note that the wheel does not slip in the process of motion, so that, rotating through an angle t, it travels a distance equal to the length of the arc BA_t in Figure 1.8.3, which amounts to rt). The center (or axis) of the wheel Q_0, will occupy the position $Q_t(rt, r)$, and the point A_0 will occupy A_t. (We wish to find the coordinates of point A_t.

From the right triangle Q_tA_tP (see Figure 1.8.3), in which $\angle A_tQ_tP = t$ and $Q_tA_t = r$, we get $A_tP = r \sin t$ and $Q_tP = r \cos t$. Since $BO = PT = rt$ and $Q_tB = Q_0O = r$, we get

$$x = TA_t = TP - A_tP = rt - r \sin t,$$
$$y = OT = BP = Q_tB - Q_tP = r - r \cos t.$$

The final result is

$$x = r(t - \sin t),\ y = r(1 - \cos t) \qquad (1.8.2)$$

(the solid curve in Figure 1.8.3 depicts the cycloid).

Exercises

1.8.1. Construct the curves given by the following equations: (a) $x = \cos t$, $y = \sin 2t$ and (b) $x = \cos t$, $y = \sin 3t$. [*Hint.* Since $\sin 3t$ varies rapidly, you must take close-lying values of t, say, $t = 0, 0.1, 0.2, \ldots$, or, if you have no trigonometric tables with a column for radians, $t = 0, 5°, 10°, 15°, \ldots$]

1.8.2. Construct the curve $x = \cos(5t + 1)$, $y = \sin(5t + 1)$.

1.8.3. Construct the curve $x = \cos t$, $y = \cos(t + \pi/4)$.

1.8.4. Construct the curve $x = \cos t$, $y = \cos t$.

1.8.5. Express the parametric dependence (1.8.2) explicitly, in the form of a function $x = \varphi(y)$.

1.8.6. A circle s of radius r is rolling without slipping along a circle S of radius R. Write the parametric equations of the curve described by a point A on the moving circle if (a) s lies outside S and (b) s lies inside S or (for $R < r$) S lies inside s. The curve (a) is called an ***epicycloid***, and the curve (b) a ***hypocycloid***.]

Consider the following particular cases: in case (a) $r = R$, $r = R/2$, and $r = 2R$; in case (b) $r = R/2$, $r = R/3$, and $r = R/4$. (At $r = R$ the epicycloid is called a ***cardiod*** (from the Greek kardia meaning "heart"), at $r = R/3$ and $r = R/4$ the hypocycloids are called the ***Steiner***[1,19] ***curve*** and the ***astroid*** (from the Latin word astrum meaning "star"). What is the appearance of the hypocycloid at $r = R/2$?)

1.9* Some Additional Topics from Analytic Geometry

Above (in Section 1.2) we considered a number of geometrical problems solved by the coordinate method, or by methods of *analytic geometry*.

Suppose that $A_1(x_1, y_1)$, $A_2(x_2, y_2)$, and $A_3(x_3, y_3)$ are three points on the coordinate plane (Figure 1.9.1).

[1.18] From the Greek *kykloeides* meaning circular.

[1.19] ***Jacob Steiner*** (1796-1863), an outstanding German geometer.

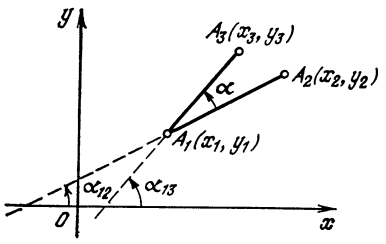


Figure 1.9.1

How does one find the angle $\angle A_2A_1A_3 = \alpha$? In Section 1.2 we found that

$$\tan \alpha_{12} = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{and} \quad \tan \alpha_{13} = \frac{y_3 - y_1}{x_3 - x_1},$$

where α_{12} and α_{13} are the angles formed by the x axis and the segments A_1A_2 and A_1A_3 (see formula (1.2.5)). But Figure 1.9.1 clearly shows that

$$\alpha = \angle A_2A_1A_3 = \alpha_{13} - \alpha_{12}.$$

Since, according to a well-known formula from trigonometry,

$$\tan(\beta - \gamma) = \frac{\tan\beta - \tan\gamma}{1 + \tan\beta \tan\gamma},$$

we have

$$\tan\alpha = \tan(\alpha_{13} - \alpha_{12}) = \frac{\tan\alpha_{13} - \tan\alpha_{12}}{1 + \tan\alpha_{13}\tan\alpha_{12}} = \frac{\dfrac{y_3 - y_1}{x_3 - x_1} - \dfrac{y_2 - y_1}{x_2 - x_1}}{1 + \dfrac{y_3 - y_1}{x_3 - x_1}\dfrac{y_2 - y_1}{x_2 - x_1}},$$

or, finally,

$$\tan\alpha = \frac{(y_3 - y_1)(x_2 - x_1) - (y_2 - y_1)(x_3 - x_1)}{(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1)}. \tag{1.9.1}$$

Thus, knowing the coordinates of points A_1, A_2, and A_3, we can find the angle $\alpha = \angle A_2A_1A_3$.

Combining (1.9.1) with the well-known formulas $\sin\alpha = [\tan^2\alpha/(1 + \tan^2\alpha)]^{-1/2}$ and $\cos\alpha = (1 + \tan^2\alpha)^{-1/2}$, we arrive at the following formulas:

$$\sin\alpha = \frac{(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} \tag{1.9.2a}$$

$$\cos\alpha = \frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} \tag{1.9.2b}$$

The comparatively complicated formula (1.9.1) (and also the formulas (1.9.2a) and (1.9.2b)) are rarely employed. All these formulas hold true for any values (arbitrary both in size and sign) of the coordinates of the points considered and the differences of these coordinates provided that we agree to reckon α from A_1A_2 to A_1A_3 and that α can be both positive and negative.

Since $\tan\alpha$ is meaningless when $\alpha = \pi/2$ (the tangent of α tends to infinity as $\alpha \to \pi/2$ and is said to become infinite, the denominator in the expression for $\tan\alpha$ must be zero at $\alpha = \pi/2$), formula (1.9.1) yields the following condition necessary and sufficient for the straight lines A_1A_2 and A_1A_3 to be *perpendicular to each other*:

$$(x_3 - x_1)(x_2 - x_1) + (y_3 - y_1)(y_2 - y_1) = 0. \tag{1.9.3}$$

The following conclusions can be made on the basis of (1.9.1), (1.9.2a), and (1.9.2b). Note that the expressions in the denominators in the right members of (1.9.2a) and (1.9.2b) are the distances A_1A_2 and A_1A_3 (cf. (1.2.4)):

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = A_1A_2,$$

$$\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} = A_1A_3.$$

From the school course of geometry the reader should know that

$$S_{\Delta A_1A_2A_3} = \frac{1}{2} A_1A_2 \times A_1A_3 \times |\sin\alpha|,$$

where $S_{\Delta A_1A_2A_3}$ is the area of triangle $A_1A_2A_3$. This yields the following formula for the area:

$$S_{\Delta A_1A_2A_3} = \frac{1}{2}|(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)|. \tag{1.9.4}$$

On the other hand, $S_{\Delta A_1A_2A_3} = (1/2)A_1A_2 \times h$, where h is the distance

[1.20] In formulas (1.9.4) and (1.9.5) you can drop the vertical bars and assume that $S_{\Delta A_1A_2A_3}$ and h_{A_3, A_1A_2} possess a sense of direction, so to say, that is, their sign determines on which side of A_1A_2 point A_3 lies. We will not dwell on this any longer.

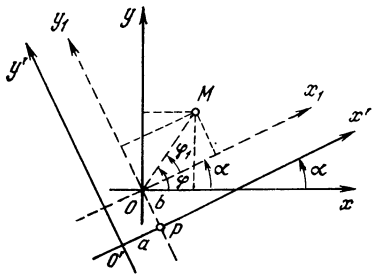


Figure 1.9.2

from point A_3 to the straight line A_1A_2. In view of (1.9.4), we then have the following formula for the distance $h = h_{A_3, A_1A_2}$ from point A_3 to straight line A_1A_2:

$$h_{A_3, A_1A_2} = \frac{|(x_2-x_1)(y_3-y_1)-(x_3-x_1)(y_2-y_1)|}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}. \quad (1.9.5)$$

We note also that the s stem of coordinates in a plane for geometrical problems is rather arbitrary: to specify a system of coordinates we must fix the point O (the origin) and the two axes Ox and Oy (perpendicular to each other), which, however, in all other respect can be chosen at our will.

For example, suppose that xOy and $x'O'y'$ are two different coordinate systems (Figure 1.9.2). To establish the relationship between the "old" coordinates (x, y) and the "new" coordinates (x', y') of *one and the same point* M, we introduce an "intermediate" coordinate system x_1Oy_1 whose origin coincides with the origin of the old system while the directions of the x_1 and y_1 axes coincide with the directions of the x' and y' axes, respectively. To relate the coordinates (x, y) and (x_1, y_1) of a point, it is convenient to consider two systems of polar coordinates with the same pole O and the polar axes Ox and Ox_1, respectively (see Figure 1.9.2). If $\angle xOx_1 = \alpha$, then point $M(r, \varphi)$ (i.e. point M with the polar coordinates in the first system of units being r and φ) has coordinates r_1 and φ_1 in the second system, where, obviously, $r_1 = r$ and $\varphi_1 = \varphi - \alpha$. Therefore (see Eqs. (1.2.3)), $x = r\cos\varphi$, $y = r\sin\varphi$ and $x_1 = r_1\cos\varphi_1 = r\cos(\varphi - \alpha)$, $y_1 = r\sin(\varphi - \alpha)$, that is,

$$\begin{aligned} x_1 &= r\cos(\varphi-\alpha) = r(\cos\varphi\cos\alpha \\ &\quad + \sin\varphi\sin\alpha) \\ &= r\cos\varphi\cos\alpha + r\sin\varphi\sin\alpha \\ &= x\cos\alpha + y\sin\alpha, \\ y_1 &= r\sin(\varphi-\alpha) = r\sin\varphi\cos\alpha \\ &\quad - \cos\varphi\sin\alpha) \\ &= -r\cos\varphi\sin\alpha + r\sin\varphi\cos\alpha \\ &= -x\sin\alpha + y\cos\alpha. \end{aligned}$$

On the other hand, from the same Figure 1.9.2 it follows that

$$x' = x_1 + a, \quad y' = y_1 + b,$$

where $a = O'P$ and $b = PO$ are the coordinates of the old origin O in the new system of coordinates $x'O'y'$.

Thus, the final result is[1.21]

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha + a, \\ y' &= -x\sin\alpha + y\cos\alpha + b. \end{aligned} \quad (1.9.6)$$

These formulas are fundamental to geometry. The method of cooridnates makes it possible to characterize every point in a plane by a pair of numbers, x and y, the coordinates of the point; if we have two points, then two pairs of coordinates are needed; and a set of points can be replaced with a set of pairs of numbers. For instance, a curve in a plane has corresponding to it a *one-parameter* set of pairs of numbers, $x(t)$ and $y(t)$, that depend on parameter t (cf. Section 1.8). The transition

[1.21] As is customary in analytic geometry we restrict our discussion only to *right-handed* coordinate systems, that is, systems in which a rotation of the positive semiaxis Ox through an angle of 90° that maps the semiaxis into the positive semiaxis Oy must be performed in the *positive* direction, or counterclockwise. The reader can easily show that if xOy is a right-handed coordinate system and $x'O'y'$ a left-handed one, then

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha + a, \\ y' &= x\sin\alpha - y\cos\alpha + b. \end{aligned}$$

(What is the meaning here of angle α and line segments a and b?)

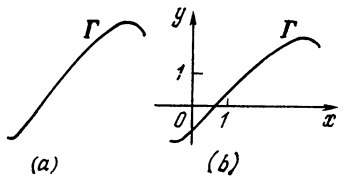


Figure 1.9.3

from one system of coordinates to another in no way affects the geometric properties of curves, and for this reason in geometry the only relationships between coordinates of points that have any meaning are those that retain their form under any transformation of the (1.9.6) type, that is, under a transfer from coordinate system xOy to any other coordinate system $x'O'y'$. And these are the relations that were studied above.

The aforesaid can be worded in a different manner. The equation $y = f(x)$ or $F(x, y)$ of a certain curve Γ depends, obviously, not only on the shape of the curve but also on the choice of the coordinate system; in other words, the equation describes not the curve alone (Figure 1.9.3*a*) but a much more complicated object: the curve Γ and the coordinate axes (the axis of abscissas and the axis of ordinates; and not only the axes but also the unit segments on the axes that fix the units of measurement for x and y; see Figure 1.9.3*b*). But of geometric meaning (and we are interested here in geometry and not in physics[1.22]) are only those expressions connected with the equation of a curve that retain their form under a substitution of variables via formulas (1.9.6). For instance, in the equation of a circle of radius r centered at point $Q(a, b)$,

$$(x - a)^2 + (y - b)^2 = r^2, \text{ or}$$

$$(x - a)^2 + (y - b)^2 - r^2 = 0, \quad (1.9.7)$$

the quantities a and b characterize (only) the position of the circle in the coordinate plane, with the number r related to the "geometry" of the curve (with the size of the circle; see Exercise 1.9.6).

To demonstrate the aforesaid, let us turn to formula (1.2.4) for the distance r_{12} between two points. If $A_1(x_1, y_1)$ and $A_2(x_2, y_2)$ are two points, then, in view of (1.2.4), the square of the distance between them, r_{12}^2, is equal to $(x_2 - x_1)^2 + (y_2 - y_1)^2$. Let us introduce new coordinates, x' and y'. Then, according to (1.9.6),

$$x_1' = x_1 \cos\alpha + y_1 \sin\alpha + a,$$

$$y_1' = -x_1 \sin\alpha + y_1 \cos\alpha + b,$$

and, similarly, we can also find the new coordinates x_2' and y_2' of point A_2 in terms of the old coordinates x_2 and y_2. We then have

$$x_2' - x_1' = (x_2 - x_1)\cos\alpha + (y_2 - y_1)\sin\alpha,$$

$$y_2' - y_1' = -(x_2 - x_1)\sin\alpha + (y_2 - y_1)\cos\alpha$$

and, respectively,

$$\begin{aligned}(x_2' - x_1')^2 + (y_2' - y_1')^2 &= [(x_2 - x_1)\cos\alpha + (y_2 - y_1)\sin\alpha]^2 \\ &+ [-(x_2 - x_1)\sin\alpha + (y_2 - y_1)\cos\alpha]^2 \\ &= (x_2 - x_1)^2(\cos^2\alpha + \sin^2\alpha) \\ &+ (y_2 - y_1)^2(\sin^2\alpha + \cos^2\alpha) \\ &= (x_2 - x_1)^2 + (y_2 - y_1)^2,\end{aligned}$$

which is proof of the independence of r_{12} of the choice of the system of coordinates.

Clearly, two straight lines $y = k_1x + b_1$ and $y = k_2x + b_2$ (Figure

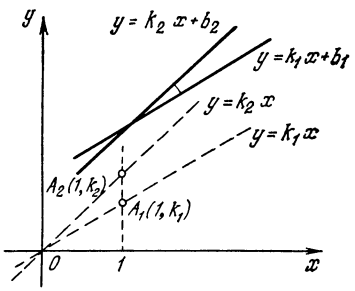


Figure 1.9.4

[1.22] In physics the situation is somewhat different, in this connection see Section 9.8.

1.9.4) are parallel if and only if their slopes are the same, that is, if $k_1 = k_2$. Let us now consider two straight lines $y = k_1x$ and $y = k_2x$, which pass through the origin O and are parallel to our straight lines (which are not parallel to each other, in general; see the dashed lines in Figure 1.9.4). Since there are two points $A_1\,(1, k_1)$ and $A_2\,(1, k_2)$ that belong to the new straight lines (i.e. the ones that pass through O), the angle α, equal to the angle $\angle A_1OA_2$ between our straight lines (it is unimportant whether these are the new lines or the old lines), is determined via formula (1.9.1):

$$\tan\alpha = \frac{(k_2-0)\,(1-0)-(k_1-0)\,(1-0)}{(1-0)\,(1-0)+(k_2-0)\,(k_1-0)}$$

(since here the points $O\,(0, 0)$, $A_1\,(1, k_1)$, and $A_2\,(1, k_2)$ play the role of points $A_1\,(x_1, y_1)$, $A_2\,(x_2, y_2)$, and $A_3\,(x_3, y_3)$ in formula (1.9.1)). Thus, the final result is

$$\tan\alpha = \frac{k_2-k_1}{1+k_2k_1}. \qquad (1.9.8)$$

In particular, two straight lines $y = k_1x + b_1$ and $y = k_2x + b_2$ are perpendicular to each other if and only if

$$k_1k_2 + 1 = 0, \text{ or } k_1k_2 = -1. \qquad (1.9.9)$$

(Why?)

Now, suppose that $y = kx + b$ and $y = kx + b_1$ are two parallel straight lines l and l_1 whose slopes are the same, k, and which intersect the y axis at points $B\,(0, b)$ and $B_1\,(0, b_1)$ (Figure 1.9.5). In this case, obviously, $BB_1 = |b_1 - b|$, on the other hand, the perpendicular B_1P dropped from point B_1 onto l forms an angle α with the y axis that is equal to the angle between l (or l_1) and the x axis (see Figure 1.9.5,

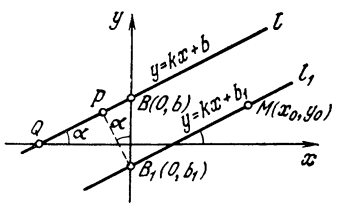


Figure 1.9.5

where $\angle BB_1P$ and $\angle xQl$ are angles with mutually perpendicular sides). But $\tan\angle xQl = \tan\alpha = k$, by the very definition of the slope of a straight line, which means that $\tan\angle BB_1P = k$ and, hence,

$$\cos\angle BB_1P = \cos\alpha = \frac{1}{\sqrt{1+\tan^2\alpha}} = \frac{1}{\sqrt{1+k^2}}.$$

From triangle BB_1P (see Figure 1.9.5) we obtain

$$B_1P = BB_1\cos\alpha = \frac{|b_1-b|}{\sqrt{1+k^2}}.$$

We have found that the distance $d = B_1P$ between parallel lines l and l_1 is given by the formula

$$d = \frac{|b_1-b|}{\sqrt{1+k^2}}. \qquad (1.9.10)$$

If l is a straight line given by the equation $y = kx + b$ and $M\,(x_0, y_0)$ is an arbitrary point (see Figure 1.9.5), then another straight line l_1 passing through point M is given by the equation $y - y_0 = k\,(x - x_0)$, or $y = kx + b_1$, where $b_1 = y_0 - kx_0$ (see formula (1.3.3)). Hence, in view of (1.9.10) the distance h from point M to straight line l is given by the formula

$$h = \frac{|y_0-kx_0-b|}{\sqrt{1+k^2}} \qquad (1.9.11)$$

(note that in the numerator on the right-hand side of (1.9.11) we have the absolute value of the result of substituting the coordinates of point M into the left-hand side of the equation of the straight line, $y - kx - b = 0$).

Exercises

1.9.1. Given a point M and a straight line l: (a) $M = M\,(1, -1)$, $y = x + 1$; (b) $M = M\,(0, -1)$, $y = 4$; (c) $M = M\,(-2, 0)$, $y = -x - 2$; and (d) $M = M\,(0, 0)$ (the origin), $y = (3/4)\,x - 1/4$. Draw a line p that passes through M and is perpendicular to l.

1.9.2. Given three points A_1, A_2, A_3: $A_1\,(0, 0)$, $A_2\,(4, 3)$, $A_3\,(-6, 8)$; $A_1\,(4, -4)$, $A_2\,(4, 0)$, $A_3\,(-1, 1)$; and $A_1\,(1, 2)$, $A_2\,(2, 1)$, $A_3\,(-1, 1)$. (a) Find $\angle A_1A_2A_3$,

(b) calculate $S_{\Delta A_1A_2A_3}$, and (c) find h_{A_1,A_2A_3}.

1.9.3. Given two parallel straight lines l and l_1: (a) $y = x + 2$, $y = x - 2$; and (b) $3x - 4y - 1 = 0$, $-3x + 4y = 0$. Find the distance d between the lines.

1.9.4. Given the point M and the straight line l of Exercise 1.9.1. Find the distance h between M and l.

1.9.5. Suppose that $A_1 = A_1\ (x_1,\ y_1)$, $A_2 = A_2\ (x_2,\ y_2)$, $A_3 = A_3\ (x_3,\ y_3)$, and $A_4 = A_4\ (x_4,\ y_4)$ are four points in the xy-plane. Prove that $A_1A_2 \perp A_3A_4$ if and only if $(x_2 - x_1)(x_4 - x_3) + (y_2 - y_1)(y_4 - y_3) = 0$.

1.9.6. Establish how the equation (1.9.7) of a circle transforms under a transformation to new coordinates x' and y' via (1.9.6). Verify that the quantities a and b in (1.9.7) change (i.e. that the new equation will be of the same form but with different a and b) while r remains the same. [*Hint.* To solve this problem it proves convenient to "reverse" formulas (1.9.6), that is, to express the old coordinates x and y in terms of the new coordinates x' and y'.]

1.9.7. Prove that under a transition (1.9.6) to another system of coordinates the following conditions and formulas retain their form: (a) the condition $k_1 = k_2$ that two straight lines are parallel, (b) the condition (1.9.9) that two straight lines are perpendicular to each other, (c) the formula (1.9.8) for the angle between two straight lines, (d) the formula (1.9.10) for the distance between two parallel straight lines, (e) the formula (1.9.11) for the distance from a point to a straight line, (f) the formula (1.9.4) for the area of a triangle, (g) the formula (1.9.1) for an angle in a triangle, (h) the formula (1.9.5) for the distance from a point to a straight line, and (i) the conditions for two straight lines being perpendicular to each other (see Exercise 1.9.5) [*Hint.* See the hint to Exercise 1.9.6.]

Chapter 2 What is a Derivative?

2.1 Motion, Distance, and Velocity

Let us examine the translational motion of an object, or body, along a straight line. Denote the distance of some point M of the body to a specific point O on the line (the coordinate of point M) by z. We will consider the distance in one direction to be positive and in the opposite direction negative. For example, suppose that the straight line along which our body is moving is vertical. Points above O will correspond to positive values of z, and those below O to negative values of z. It is often convenient to assume that the body is small, so that we can simply speak of a material point or particle, M, and of the distance of this point from a definite point O on the straight line, or the origin of our coordinate system.

Problems on the motion of bodies with constant velocities v lead to simple arithmetic and algebraic computations based on the fact that distance is velocity (speed) multiplied by time, that is, the elementary formula $s = vt$, where s is distance and t is time. However, in nature as a rule we deal with motions whose velocities *vary* with time. Studies of such motions lead to two important physical concepts, *distance* and *velocity*, as functions of time, and already at this stage there appear the two basic concepts of higher mathematics, the concept of the ***derivative*** and the concept of the ***integral***. Starting with this chapter, we consider these concepts in great detail.

The process of motion of a particle M consists in the fact that the z-coordinate of this point changes with time. The motion of the body (or point M) is determined by the dependence of z on t, that is, is characterized by the function $z = z(t)$. Knowing this function, we can find the position of the body at any moment in time. The function $z(t)$ may be represented graphically by laying off time on the axis of abscissas (the t axis) and the distance from point M to point O on the axis of ordinates (the z axis).

In *uniform* motion, with constant velocity v, the distance s covered by the body in time t is directly proportional to t, with the coefficient of proportionality being v (here $s = vt$). Denote by z_0 the coordinate of the body at time $t = 0$. The distance s covered in time t will then be equal to the difference $z(t) - z_0$. Thus.

$$z(t) = z_0 + vt. \tag{2.1.1}$$

Hence, in *uniform* motion, the dependence of the coordinate z on time t is given by a *linear* function. The graph of the function $z = z(t)$ in this case is a straight line in the tz-plane (Figure 2.1.1).

In the case of nonuniform motion, the function $z = z(t)$ is expressed by more involved formulas and the coorresponding graph is some kind of curve. Let us analyze the following basic problem, *given a function* $z = z(t)$, or the dependence of the coordinate of the body on the time variable, *find the rate of motion* (or velocity) v *of the body*. In the general case of nonuniform motion, the velocity does not remain constant, it changes in the course of time. This means that the velocity v is also a function of the time variable t, or $v = v(t)$, and our problem consists in expressing this function in terms of the known function $z = z(t)$.

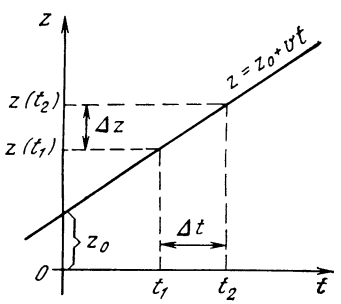


Figure 2.1.1

Everything is simple in the particular case of *uniform* motion. The velocity is defined as the distance covered in unit time. Let us find, for instance, the distance traversed in one second from time t_1 seconds to time $t_1 + 1$ seconds. This distance (equal numerically to the velocity) is equal to the difference between the coordinates $z\,(t_1 + 1)$ and $z\,(t_1)$:

$$z\,(t_1 + 1) - z\,(t_1) = [z_0 + v\,(t_1 + 1)] - [z_0 + vt_1] = v.$$

Instead of this we can take an arbitrary instant of time between t_1 and t_2 and divide the distance traveled $z_2 - z_1$ by the magnitude of the interval $t_2 - t_1$:

$$\frac{z_2 - z_1}{t_2 - t_1} = \frac{(z_0 + vt_2) - (z_0 + vt_1)}{t_2 - t_1} = v \quad (2.1.2)$$

It is precisely because the velocity is constant that we are able to take *any* interval $t_2 - t_1$ to compute the velocity, and the answer will be independent of both the time t_1 and the magnitude of the time interval, $| t_2 - t_1 |$. The situation is different in the case of a *variable* rate of motion.

Before going over to the more general case, it will be convenient to change the notation. We will write $t_1 = t$ and $t_2 = t + \Delta t$, so that the difference $t_2 - t_1$ (the time interval) is denoted by Δt (see Figure 2.1.1).[2.1] Similarly, we write Δz to denote the difference $z\,(t_2) - z\,(t_1) = z\,(t + \Delta t) - z\,(t) = \Delta z$.

In this notation, formula (2.1.2) can be rewritten thus:

$$\frac{\Delta z}{\Delta t} = v. \quad (2.1.2a)$$

[2.1] Note that Δ is not a factor but a symbol, and Δt is not the product of Δ by t, just as $\sin\alpha$ is not the product of sin by α, and so Δ cannot be canceled from the numerator and denominator on the right-hand sides of (2.1.2a) and (2.1.3), in the same way as we cannot cancel sin from the numerator and denominator in the fraction $\sin\alpha/\sin\beta$. Δ is the capital Greek letter delta; Δt is read "delta t" and Δz is read "delta z"; these are also spoken of as the increment, or change, in time and the increment, or change, in path (or distance).

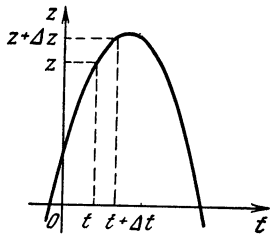


Figure 2.1.2

In the general case of *nonuniform* motion, the right-hand side of (2.1.2a) yields the ***average velocity*** v_{av} in the interval Δt between t and $t + \Delta t$ (the size of the interval is therefore Δt) is

$$v_{av} = \frac{\Delta z}{\Delta t}. \quad (2.1.3)$$

We speak here of the average velocity because the velocity itself can change over the interval Δt.

Let us consider an example in which $z\,(t)$ is given by a *quadratic* function instead of the linear function shown in Figure 2.1.1:

$$z\,(t) = z_0 + bt + ct^2. \quad (2.1.4)$$

Figure 2.1.2 illustrates a possible graph corresponding to a function of the form (2.1.4). Let us compute the average velocity v_{av} over the interval Δt using the formula (2.1.3). In the case at hand we have

$$z\,(t) = z_0 + bt + ct^2,$$
$$z\,(t + \Delta t) = z_0 + b\,(t + \Delta t) + c\,(t + \Delta t)^2 = z_0 + bt + b\Delta t + ct^2 + 2ct\Delta t + c\,(\Delta t)^2$$

and, hence,

$$\Delta z = z\,(t + \Delta t) - z\,(t) = b\Delta t + 2ct\Delta t + c\,(\Delta t)^2.$$

From this we get

$$v_{av} = \frac{\Delta z}{\Delta t} = b + 2ct + c\Delta t, \quad (2.1.5)$$

Compare the results of (2.1.2a) and (2.1.5) for the average velocity when the motion obeys the law (2.1.1) and the law (2.1.4). The second example differs from the first in that here the

average velocity depends both on the time t and on the time interval Δt. How can we find the *instantaneous* velocity at time t which depends solely on this moment in time?

The velocity changes gradually, and so the smaller the time interval over which the distance traveled is measured, the smaller the change in velocity and, hence the closer will the average velocity be to the instantaneous value. Formula (2.1.5) for v_{av} contains two terms that do not depend on the size of Δt and one term that is proportional to Δt. For very small Δt the product $c\Delta t$ will also be small, whence the last term on the right-hand side of (2.1.5) can be ignored and the formula for v_{av} will transfer into a formula for the instantaneous velocity:

$$v_{inst} = b + 2ct. \qquad (2.1.6)$$

For instance, suppose $z_0 = 1$, $b = 2$, and $c = -1$ in (2.1.4). We wish to find the velocity at $t = 0$. Equation (2.1.5) then yields

$$v_{av} = 2 + 2 \times (-1) \times 0 + (-1) \times \Delta t = 2 - \Delta t.$$

We can construct the following table:

Δt (from 0 to Δt s)	1	0.5	0.1	0.05	0.01
v_{av}	1	1.5	1.9	1.95	1.99

Thus, the smaller the size of Δt, the closer v_{av} is to 2, and it is natural to assume that $v_{inst} = 2$.

The attentive reader will have most likely recognized the expressions (2.1.4) and (2.1.6) from the school course of physics to be the formulas for uniformly accelerated motion:

$$z(t) = z_0 + v_0 t + \frac{at^2}{2}, \quad v(t) = v_0 + at. \qquad (2.1.7)$$

All that is needed is to substitute for b in (2.1.4) the *initial velocity* v_0 (i.e. the velocity at time $t = 0$) and for c in (2.1.4) the "half-acceleration" $a/2$ (since a is the *acceleration* of the body).

We have computed the instantaneous velocity at time t on the basis of the average velocity over the interval from t to $t + \Delta t$. Now let us try to compute it by choosing the interval in a somewhat different way. We find the average velocity in the interval from $t_1 = t - (3/4)\,\Delta t$ to $t_2 = t + (1/4)\,\Delta t$. As before, the duration of the interval is Δt. From formula (2.1.4) we get

$$z(t_1) = z_0 + b\left(t - \frac{3}{4}\Delta t\right) + c\left(t - \frac{3}{4}\Delta t\right)^2,$$

$$z(t_2) = z_0 + b\left(t + \frac{1}{4}\Delta t\right) + c\left(t + \frac{1}{4}\Delta t\right)^2$$

and

$$z(t_2) - z(t_1) = b\Delta t + 2ct\Delta t - \frac{1}{2}c(\Delta t)^2.$$

Whence it follows that

$$v_{av} = \frac{z(t_2) - z(t_1)}{\Delta t} = b + 2ct - \frac{1}{2}c\Delta t. \qquad (2.1.8)$$

Comparing (2.1.5) and (2.1.8), we see that the average velocities over the interval from t to $t + \Delta t$ and over the interval from $t - (3/4)\,\Delta t$ to $t + (1/4)\,\Delta t$ differ by $c\,\Delta t\,[1 - (-1/2)] = (3/2)\,c\,\Delta t$. But if we want to find the *instantaneous* velocity, we have to take a very small time interval Δt. Then the difference between the two expressions for the average velocity will vanish and we again obtain for the instantaneous velocity $v_{inst} = b + 2ct$. However, the value of the average velocity over the interval from $t - (3/4)\,\Delta t$ to $t + (1/4)\,\Delta t$ given by formula (2.1.8) is *closer* to the instantaneous velocity (2.1.6) than the value of the average velocity over the interval from t to $t + \Delta t$ given by formula (2.1.5). This implies that (with greater precision) the choice of the interval from t to $t + \Delta t$ is less convenient: here v_{av} is a worse approximation for v_{inst}. It is even better to select the interval in such a way that the value of t we are interested in lies at the center of the interval; in this case, for the quadratic dependence (2.1.4) between z and t, the value of v_{av} will coincide with v_{inst} (see Exercise 2.1.2). We will dwell on this fact again in Section 6.1.

We have considered the concept of instantaneous velocity for two specific cases: uniform and uniformly accelerated motion. In Section 2.3 we give a more exact definition of instantaneous velocity for an arbitrary law of motion.

Exercises

2.1.1. The following dependence of the distance (z) traveled by a material particle M along a straight line on time (t) was observed: (a) $z = t^3 + t$, and (b) $z = (1 + t^2)^{-1}$. What is the average velocity v_{av} of each motion over the time interval from t to Δt? What is the instantaneous velocity v_{inst} at time t?

2.1.2. For the quadratic dependence of z on t given by formula (2.1.4), find v_{av} using the formula

$$v_{av} = \frac{z\left(t+\frac{1}{2}\Delta t\right)-z\left(t-\frac{1}{2}\Delta t\right)}{\Delta t}$$

and compare the result with the value of v_{inst} at time t.

2.2* Specific Heat Capacity of an Object. Thermal Expansion

Note that a procedure similar to the one discussed in the previous section is often encountered in many physical problems. We will demonstrate this using two simple examples and the knowledge that a school physics course supplies.

The reader will recall that the ***specific heat capacity*** of a substance is the quantity of heat (in joules) required to change the temperature of 1 kg of the substance (water, iron, gold, etc.) by 1 °C. But for different initial temperatures the quantity of heat required to change the temperature of 1 kg of substance by 1 °C is different, so that the specific heat capacity is a function of temperature T, or $c = c(T)$. For instance, to heat 1 kg of iron taken at a temperature of 0 °C up to 1 °C we must supply 440.857 J of heat, while to heat the same amount of iron taken at 50 °C up to 51 °C we must supply 470.583 J. Then how does one define the heat capacity of a body corresponding to a fixed temperature T?

The content of Section 2.1 suggests the following method for defining the quantity we are interested in, $c = c(T)$ (the specific heat capacity at temperature T). Here the quantity of heat Q (in joules) that must be delivered to the specified amount (1 kg) of the substance in order to heat the substance at the initial temperature (it is unimportant which initial temperature specifically) up to temperature T is the analog of the distance z of Section 2.1, and it is clear that this quantity depends on T, or $Q = Q(T)$. Heating the 1 kg of the substance from temperature T_1 to temperature T_2 requires $Q(T_1, T_2) = Q(T_2) - Q(T_1)$ joules of heat, while heating the same amount of substance from T to $(T + \Delta T)$ °C requires $Q(T + \Delta T) - Q(T) = \Delta Q$ joules of heat. Therefore, the *average* specific heat capacity c_{av} over the temperature interval from T to $T + \Delta T$ is naturally defined as

$$\frac{Q(T+\Delta T)-Q(T)}{\Delta T} = \frac{\Delta Q}{\Delta T}, \qquad (2.2.1)$$

and the *instantaneous* specific heat capacity c_{inst} (the word "instantaneous" refers in this case not to a definite moment in time but to a definite temperature T) is defined as a value of the average specific heat capacity c_{av} over a very small temperature interval ΔT, and the narrower the interval the closer c_{av} will be to c_{inst}. Note that in the majority of cases the value of 1 °C for ΔT is sufficiently small to yield an exact value of $c = c(T)$. Here the expression "sufficiently small" means that the value of c obtained in this manner will for all practical purposes coincide with the value at which we arrive by selecting a smaller (and even considerably smaller) interval ΔT of temperature variation.

Example 1. The quantity of heat $Q = Q(T)$ required to heat 1 kg of iron from 0 °C to T °C is given by the following empirical formula (which quite sufficiently represents the process we are interested in, at least for $T < 200$ °C):

$$Q(T) = 440.857T + 0.29725T^2. \qquad (2.2.2)$$

In full accordance with the material of Section 2.1 this yields (note that

the function (2.2.2) is quadratic, just as (2.1.4) is)

$$c_{av} = \frac{\Delta Q}{\Delta T} = \frac{440.857\,(T+\Delta T)+0.29725\,(T+\Delta T)^2}{\Delta T} - \frac{440.857T+0.29725T^2}{\Delta T} = 440.857 + 0.5945T + 0.29725\Delta T \qquad (2.2.3)$$

and, hence

$$c_{inst} = 440.857 + 0.5945T. \qquad (2.2.4)$$

We see that $c\,(0) = 449.857$ J/kg × °C, as we noted earlier, while, say, $c\,(100) = 440.857 + 0.5945 \times 100 = 500.307$ J/kg·°C.

A similar situation arises in all problems where the quantity we are interested in is the rate of variation of another quantity. For instance, the coefficient k of (thermal) ***linear expansion*** is usually defined as the number that shows by what amount a thin rod made of the substance of interest (the length of the rod is assumed equal to 1 cm) changes its length when its temperature is increased by 1 °C. Here too the number k is not a constant: it depends on the initial temperature T of the rod, whereby, strictly speaking, our definition, which involves heating the rod from T to $(T + 1)$ °C, is not quite correct, since in the process of heating the quantity (or function) $k = k\,(T)$ varies.

Then how are we to determine k_{inst}, or the quantity k corresponding to a *definite* temperature T? Let us take a rod 1 cm long at 0 °C and gradually heat it. The length of the rod, L, will change in the process and, hence, is a function of the rod's temperature T, or $L = L\,(T)$. If we heat the rod from temperature T to temperature $T + \Delta T$, where the temperature increment ΔT is assumed small, the length of the rod will increase by

$$\Delta L = L\,(T + \Delta T) - L\,(T),$$

so that the following average elongation of the rod corresponds to an increase in temperature by 1 °C:

$$\frac{L\,(T+\Delta T)-L\,(T)}{\Delta T} = \frac{\Delta L}{\Delta T}.$$

However, this elongation refers not to a 1-cm rod but to a rod of length L, so that per unit length of the rod we have (on the average)

$$k_{av} = \frac{1}{L}\,\frac{\Delta L}{\Delta T}. \qquad (2.2.5)$$

The value of k_{av} given by formula (2.2.5) is the approximate value[2.2] of the (instantaneous) value of the coefficient of linear expansion $k = k\,(T)$ of the substance we are interested in, and the smaller the temperature interval ΔT we select, the closer k_{av} will be to k_{inst}.

Example 2. For platinum, within a broad temperature range, the function $L = L\,(T)$ we are discussing here obeys fairly well the following empirical formula:

$$L\,(T) = 1 + 8.806 \times 10^{-6}T + 1.95 \times 10^{-9}T^2 \qquad (2.2.6)$$

(note that the dependence of L on T is again quadratic). We wish to express the coefficient of linear expansion $k = k\,(T)$ of platinum as a function of T and find the values $k\,(0)$ and $k\,(1000)$.

Answer. Just as we did above (check all the calculations yourself), we find that

$$k\,(T) = \frac{1}{L}\,\frac{\Delta L}{\Delta T} \simeq \frac{8.806\times10^{-6}+3.90\times10^{-9}T}{1+8.806\times10^{-6}T+1.95\times10^{-9}T^2}, \qquad (2.2.7)$$

from which it follows that $k\,(0) = 8.806 \times 10^{-6}$ (1/°C) and $k\,(1000) = 12.57 \times 10^{-6}$ (1/°C).

Exercises

2.2.1. It is known that the amount of heat $Q\,(T)$ (joules) required to heat a 1-kg diamond (!) from 0 to T °C can be expressed fairly

[2.2] The approximate nature of formula (2.2.5) manifests itself in the fact that we relate the elongation ΔL (more precisely the ***specific elongation*** $\Delta L/\Delta T$) to the length $L = L\,(T)$ of the rod at the initial temperature, although in reality the rod's length changes continuously in the entire heating process. (Why should we divide $\Delta L/\Delta T$ by $L\,(T)$ and not by $L\,(T + \Delta T)$?) This inaccuracy in formula (2.2.5) for k is unimportant. (Why?)

well by the following empirical formula (which is valid in the 0 to 800 °C temperature range):

$$Q(T) = 0.3965T + 2.081 \times 10^{-3}T^2 - 5.024 \times 10^{-7}T^3.$$

Find the average specific heat capacity c_{av} of diamond over the interval from T to $(T + \Delta T)$ °C and the "instantaneous" specific heat capacity $c = c(T)$ as a function of the temperature T. What are the values of c (0), c (100), and c (500) for diamond?

2.2.2. The quantity of heat $Q(T)$ required for heating 1 kg of water from 0 to T °C is fairly well represented by the following empirical formula:

$$Q(T) = 4186.68T + 8373.36 \times 10^{-5}T^2 + 1256 \times 10^{-6}T^3.$$

What is the specific heat capacity $c(T)$ of water? Find the values of c (10) and c (90).

2.2.3. Prove that (for every substance and at any temperature T) the ***coefficient of volume expansion*** (also known as the ***bulk expansion coefficient***), that is, the "instantaneous rate" of increase in volume of substance resulting from an increase in temperature per unit volume, is three times the coefficient of linear expansion.

2.3 The Derivative of a Function. Simple Examples of Calculating Derivatives

In Section 2.1 we considered the problem of instantaneous velocity and examined ratios of the form $[z(t_2) - z(t_1)]/(t_2 - t_1)$, where the values t_1 and t_2 must be considered very close-lying. The expression "close-lying" is not exact, that is, mathematically it is not rigorous. The exact formulation is this. It is necessary to find the *limit* to which the ratio

$$\frac{z(t_2) - z(t_1)}{t_2 - t_1} \qquad (2.3.1)$$

tends as t_2 tends to t_1. Using the designations Δt and Δz, we can rewrite this ratio as

$$v_{av} = \frac{\Delta z}{\Delta t}, \qquad (2.3.2)$$

and the condition $t_2 \to t_1$ now takes the form $\Delta t \to 0$.

In (2.3.2), the quantities Δt and Δz are related: any time interval $\Delta t = t_2 - t_1$ can be selected, but after the denominator Δt has been selected, it is assumed that Δz (the numerator) is not just any distance interval but precisely that distance that corresponds to the time interval Δt. This was obvious in formula (2.3.1) from the way the ratio was written; the numerator is the difference $z(t_2) - z(t_1)$ of the values of the function $z = z(t)$ at t_2 and at t_1; the right-hand side of (2.3.2) is simply an abbreviation of the ratio (2.3.1).

Thus, the quantity that interests us, the instantaneous velocity $v_{inst} = v(t_1)$ at time t_1, is the limit of the ratio $\Delta z/\Delta t$ as Δt tends to zero (where $\Delta t = t_2 - t_1$). This statement can be written as

$$v(t_1) = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t}.$$

Here $v(t_1)$ is precisely the instantaneous velocity and lim stands for "limit". The particular kind of limit we have in mind is indicated underneath lim— when Δt approaches zero, and the arrow stands for "approaches". The quantity to the right of lim is the one whose limit is being sought. The notation is read as follows: "the limit of Δz over Δt with Δt approaching zero," where the word "over" replaces the phrase "divided by the corresponding value of."

What meaning do we attribute to the terms "limit" and "approaching the limit"? The calculations carried out in Section 2.1 served as an illustration of these notions. We saw that for small intervals Δt the value of v_{av} in Example 2 in Section 2.1 differed from the value of v_{inst} by a quantity proportional to Δt. For small Δt this quantity is small, too: the smaller the Δt, the smaller the quantity, and for infinitely small values of Δt (i.e. so small that they can be assumed negligible against the background of b and $2ct$ in (2.1.5)) the quantity also becomes infinitely small. It is precisely for this reason that we can neglect the term with Δt in the expression for v_{av} when Δt is small.

Thus, the ratio

$$\frac{\Delta z}{\Delta t} = \frac{z(t_2) - z(t_1)}{t_2 - t_1} \qquad (2.3.3)$$

or, as we predominantly wrote in Section 2.1 by replacing t_1 with t and t_2 with $t + \Delta t$,

$$\frac{\Delta z}{\Delta t} = \frac{z(t+\Delta t) - z(t)}{\Delta t}, \qquad (2.3.3a)$$

tends to a definite limit when $\Delta t = t_2 - t_1$ tends to zero.[2.3] The corresponding limit is the instantaneous velocity v, which is also a function of t:

$$\lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = v(t). \qquad (2.3.4)$$

Why is that, when computing the velocity from the given formula $z(t)$, we have to carry out so many calculations and find Δz for distinct Δt and only then find the limit $\lim_{\Delta z \to 0} \frac{\Delta z}{\Delta t}$? Couldn't we simply take the value $\Delta t = 0$ from the very start? No, because in this case we would simply get $\Delta z = 0$, since $\Delta z = z(t + \Delta t) - z(t)$, and if $\Delta t = 0$, then $\Delta z = 0$. By this thoughtless mode of operations we would get $\Delta z/\Delta t = 0/0$, which means we would get nothing definite.

When computing velocity, the whole idea is to take *small* Δt and *small* Δz *that correspond to* Δt. In this way, we obtain a very definite ratio $\Delta z/\Delta t$ each time. When Δt is reduced, (tends to zero), then Δz diminishes in approximate proportion to Δt, and so the ratio remains approximately constant, that is, the ratio $\Delta z/\Delta t$ approaches a definite limit when Δt tends to zero.[2.4] The magnitude of this limit is the instantaneous velocity $v(t)$ for the case where t is time and z is distance.

The limit of the ratio of the increment of the function to the increment of the independent variable as the increment of the independent variable tends to zero is of prime importance in higher mathematics and its numerous applications. We have already seen, for example, that such an important concept as the (instantaneous) velocity of motion is found with the aid of the limit of such a ratio. (In Section 2.2 we reduced to a similar procedure the problems of calculating the specific heat capacity and the coefficient of linear expansion; think over these examples.) This is why the limit of this ratio has a special name: the ***derivative of the function*** or, simply, the ***derivative***. This name is due to the fact that if z is a function of t, or $z = z(t)$, then the limit (2.3.4) depends both on the type of function $z = z(t)$ and on the value of t at which this limit is calculated; in other words, this limit is also a function of t, a new function *derived* from the old function $z(t)$. It is natural then to speak of this function as the derivative, where the word *derivative* stresses the dependence of this new function on the *basic* function $z = z(t)$.

We have special notations for the derivative. One notation (differential notation) is

$$\frac{dz}{dt}\left(= \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t}\right).$$

Here the quantity dz/dt (it is read "$d\,z\,d\,t$") is not a fraction but is simply an abbreviated way of writing the limit on the right. The quantity dz/dt is written in the form of a fraction to remind us that it is obtained from the fraction $\Delta z/\Delta t$ by a passage to the limit.[2.5]

A different notation for derivatives is the prime notation, $v = z'(t)$, or, for example, for the function $y = y(x)$,

$$y' = y'(x) = \frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

[2.3] We could have also started from the initial formula (2.3.3) for the ratio and assumed that t_2 and t_1 tend to a definite value of t, so that $\Delta t = t_2 - t_1$ tends to zero (in particular, see the text in small type at the end of Section 6.1).

[2.4] In some "unnatural" laws of motion $z = z(t)$ this limit may not exist; the respective restrictions will be considered later. We advise the reader not to think, at least for the time being, about such exceptions.

[2.5] Below (in Section 4.1) we will see that the expression dz/dt (or dy/dx) can also be interpreted as a fraction; for the time being, however, the reader must interpret this notation as simply a symbol.

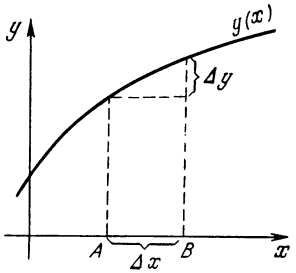


Figure 2.3.1

Occasionally, in place of the function symbol, one gives the expression of the function: if $z = at^2 + b$, then instead of dz/dt we can write directly $d(at^2 + b)/dt$ or $(at^2 + b)'$.

Thus (and this is extremely important), *the derivative of a function is defined as the limit of the ratio of the increment of the function to the increment of the independent variable as the latter tends to zero*:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}. \qquad (2.3.5)$$

The instantaneous velocity of motion of a body is equal to the derivative of the body's coordinate with respect to time. By analogy with this, when x is not time and y is not distance, we nevertheless speak of the derivative dy/dx as the rate of change (or variation) of the function y under the variation of the indepedent variable x. For instance, using the notations of Figure 2.3.1, we can say that the ratio $\Delta y/\Delta x$ is the "average rate" of increase of the function y on the interval AB within which the independent variable x varies, while the limit of this ratio, with point B approaching point A, expresses the rate of growth of y at point A on the x axis.

Let us now find algebraically the derivative of the function

$$z = t^2 \qquad (2.3.6)$$

(we have performed this operation in Section 2.1). Form the ratio

$$\frac{\Delta z}{\Delta t} = \frac{(t+\Delta t)^2 - t^2}{\Delta t}.$$

Next we remove the brackets in the numerator,

$$\Delta z = (t + \Delta t)^2 - t^2 = t^2 + 2t\Delta t + (\Delta t)^2 - t^2 = 2t\Delta t + (\Delta t)^2,$$

and get

$$\frac{\Delta z}{\Delta t} = \frac{2t\Delta t + (\Delta t)^2}{\Delta t} = 2t + \Delta t. \qquad (2.3.7)$$

Since the first term on the right-hand side of (2.3.7) is independent of Δt, where only one term is left that does not decrease as we send Δt to zero, and we can write

$$\frac{dz}{dt} = \frac{d(t^2)}{dt} = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \lim_{\Delta t \to 0} (2t + \Delta t) = 2t. \qquad (2.3.6a)$$

Let us consider another example:

$$z = t^3. \qquad (2.3.8)$$

Here

$$\Delta z = (t+\Delta t)^3 - t^3 = t^3 + 3t^2\Delta t + 3t(\Delta t)^2 + (\Delta t)^3 - t^3 = 3t^2\Delta t + 3t(\Delta t)^2 + (\Delta t)^3,$$

$$\frac{\Delta z}{\Delta t} = 3t^2 + 3t\Delta t + (\Delta t)^2,$$

and

$$\frac{dz}{dt} = \frac{d(t^3)}{dt} = \lim_{\Delta t \to 0} [3t^2 + 3t\Delta t + (\Delta t)^2] = 3t^2. \qquad (2.3.8a)$$

The limit was readily found in these examples because Δt canceled out when we computed the ratio $\Delta z/\Delta t$. Let us consider a more complicated example:

$$z = \frac{1}{t}. \qquad (2.3.9)$$

In this case we have

$$\frac{\Delta z}{\Delta t} = \frac{\frac{1}{t+\Delta t} - \frac{1}{t}}{\Delta t}.$$

Can we disregard the quantity Δt in the first fraction, in the expression $1/(t + \Delta t)$, when we pass to the limit? No, we cannot, because we have not yet canceled out the quantity Δt in the denominator. By substituting $1/t$

for $1/(t + \Delta t)$ when Δt is small, we introduce a small error (if Δt is small) in one of the summands of the numerator of the fraction $\Delta z/\Delta t$. But in this fraction both numerator and denominator are small if Δt is small. For this reason we cannot allow for a small error in the numerator (it transforms the numerator into zero).

Here is the proper way to do this:

$$\Delta z = \frac{1}{t+\Delta t} - \frac{1}{t} = \frac{t-(t+\Delta t)}{t(t+\Delta t)} = -\frac{\Delta t}{t(t+\Delta t)}, \quad \frac{\Delta z}{\Delta t} = -\frac{1}{t(t+\Delta t)}.$$

Now we can find the limit (the derivative) by dropping Δt in the denominator:

$$\frac{dz}{dt} = \frac{d(1/t)}{dt} = \lim_{\Delta t \to 0}\left[-\frac{1}{t(t+\Delta t)}\right] = -\frac{1}{t^2}. \tag{2.3.9a}$$

These examples illustrate a very important property, the fundamental property of limits (this property is commonly taken as the *definition* of a limit). As Δt is made smaller and smaller, the difference between the value of the ratio $\Delta z/\Delta t$ and the limit of this ratio (this limit is the derivative), $\lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \frac{dz}{dt}$, may be made as small as we please, which is to say, less than any given number. What is essential here is that the difference then remains smaller than the given number in the further process of variation of the ratio $\Delta z/\Delta t$.

An example will serve to illustrate this point. For $z = 1/t$ we have

$$\frac{dz}{dt} = -\frac{1}{t^2}, \quad \frac{\Delta z}{\Delta t} = -\frac{1}{t(t+\Delta t)}.$$

Let us take, say, $t = 2$. Then $dz/dt = -0.25$. Can we choose Δt such that $\Delta z/\Delta t$ differs from the limit, or -0.25, by less than 0.0025? What we mean is that Δt will have to be chosen such that

$$\frac{\Delta z}{\Delta t}\left(= -\frac{1}{2(2+\Delta t)} = -\frac{1}{4+2\Delta t}\right)$$

lies between $-0.25 + 0.0025 = -0.2475$ and $-0.25 - 0.0025 = -0.2525$.

But this, as can easily be seen, is sure to be the case if Δt in absolute value is made smaller than 0.02 (and as Δt is made smaller than 0.02 we will never encounter a value of the difference greater than 0.0025).

The same goes for other functions as well: the approach to a limit as $\Delta t \to 0$ signifies the opportunity of choosing Δt such that any degree of closeness to the limit is attainable (this limit is then never lost).

Finding the derivative in the special case of $z = t$ is particularly simple: quite obviously, $\Delta z = \Delta t$ and $\Delta z/\Delta t = 1$, that is, the ratio $\Delta z/\Delta t$ is equal to unity for arbitrary (large or small) Δt and hence in the limit as well. Thus,

$$\text{if } z = t, \quad \text{then} \quad \frac{dz}{dt} = \frac{dt}{dt} = 1. \tag{2.3.10}$$

Finally, if we consider a constant, $z = c$, it can also be regarded as a special case of a function—the graph of this function is a straight line parallel to the x axis (see Figure 1.3.6). In this case clearly $\Delta z = 0$ for all Δt, whereby the following rule holds true:

$$\text{if } z = c, \quad \text{then} \quad \frac{dz}{dt} = \frac{dc}{dt} = 0. \tag{2.3.11}$$

2.4 Properties of Derivatives. Approximating the Values of a Function by Means of a Derivative

Here are some general properties of derivatives.

If a function is multiplied by a constant factor, then the derivative is multiplied by the same factor. For instance,

$$\text{if } z = 3t^2, \quad \text{then} \quad \frac{dz}{dt} = \frac{d(3t^2)}{dt} = 3\frac{d(t^2)}{dt} = 3 \times 2t = 6t.$$

In general,

$$\text{if } z(t) = ay(t), \text{ then } \frac{dz}{dt} = a\frac{dy}{dt}. \tag{2.4.1}$$

It is also obvious that *the derivative of the sum of two functions is equal to*

the sum of the derivatives of the two functions:

if $z(t) = x(t) + y(t)$,

then $\frac{dz}{dt} = \frac{dx}{dt} + \frac{dy}{dt}$. (2.4.2)

The last rule can easily be generalized so as to include the sum of three, four, and, in general, of any number of functions.

Using the above two rules, we find that *the derivative of a sum of several functions taken with constant* (but, generally speaking, different) *coefficients is equal to the sum of the derivatives of these functions with the same coefficients*:

if $z(t) = ax(t) + by(t) + cu(t)$,

then $\frac{dz}{dt} = a\frac{dx}{dt} + b\frac{dy}{dt} + c\frac{du}{dt}$. (2.4.3)

Each of these rules is readily proved starting directly from the definition of the derivative: these rules hold true for the increment $\Delta z = z(t + \Delta t) - z(t)$ of the function $z = z(t)$ (for any Δt), whereby they are valid for the ratio $\Delta z/\Delta t$ and for the limit of the ratio, or the derivative dz/dt. For instance, if $z(t) = ay(t)$, with a constant, then

$$\Delta z = z(t + \Delta t) - z(t) = ay(t + \Delta t) - ay(t) = a[y(t + \Delta t) - y(t)] = a\Delta y$$

and, hence,

$$\frac{\Delta z}{\Delta t} = a\frac{\Delta y}{\Delta t} \text{ for all } \Delta t, \text{ i.e. } \frac{dz}{dt} = a\frac{dy}{dt}.$$

If $z(t) = x(t) + y(t)$, then

$$\Delta z = z(t + \Delta t) - z(t)$$
$$= [x(t + \Delta t) + y(t + \Delta t)] - [x(t) + y(t)]$$
$$= [x(t + \Delta t) - x(t)]$$
$$+ [y(t + \Delta t) - y(t)] = \Delta x + \Delta y$$

and

$$\frac{\Delta z}{\Delta t} = \frac{\Delta x + \Delta y}{\Delta t} = \frac{\Delta x}{\Delta t} + \frac{\Delta y}{\Delta t},$$

whence $\frac{dz}{dt} = \frac{dx}{dt} + \frac{dy}{dt}$, and so on.

It is now easy to find the derivative of a *polynomial*. We already know that

$$\frac{dc}{dt} = 0, \quad \frac{dt}{dt} = 1, \quad \frac{d(t^2)}{dt} = 2t, \quad \frac{d(t^3)}{dt} = 3t^2. \tag{2.4.4}$$

This yields

$$\frac{d(a + bt + ct^2 + et^3)}{dt} = \frac{da}{dt} + b\frac{dt}{dt} + c\frac{dt^2}{dt} + e\frac{dt^3}{dt} = b + 2ct + 3et^2. \tag{2.4.5}$$

For instance, in Section 2.1 we considered the function $z(t) = z_0 + bt + ct^2$ (see formula (2.1.4)). In view of (2.3.5), where we must put $a = z_0$ and $e = 0$, we have

$$v(t) = \frac{dz}{dt} = b + 2ct,$$

which is the result we arrived at in Section 2.1 without turning to the general properties of derivatives.

The technique for finding derivatives (also called the *differentiation of functions*) is given in detail at the beginning of Chapter 4.

Running ahead a bit, we may point out that finding derivatives of functions given by formulas is a relatively simple job, much easier, say, than the solution of algebraic equations. The formulas for the derivatives of functions are often even simpler (and never more complicated) than the formulas defining the functions. For instance, if the function is a polynomial, its derivative is also a polynomial, and the new polynomial is simpler than the initial polynomial in the sense that its degree is lower (see the above example of a third-degree polynomial, whose derivative proved to be a quadratic function of the independent variable; a similar situation arises for polynomials of *other* degrees). If the function is an algebraic fraction, then the derivative is also a fraction. If the function contains roots of fractional powers, then the derivative

also contains them. The derivatives of trigonometric functions are also trigonometric functions, and in some cases (the logarithmic function or inverse trigonometric functions, for instance) the derivative proves to be a simpler function (an algebraic fraction for the logarithm and the arctangent).

Finding derivatives does not require any kind of special ingenuity or imagination. The problem is always solved in a neat fashion through the use of the simple rules given above. As we have already said, other rules and examples of application will be given in Chapter 4.

So far, all the functions we have considered are defined by formulas. One should not think, however, that this is absolutely necessary for the existence of a derivative. For example, we can regard the dependence of distance covered upon time as having been found from experiments, in the form of very extensive tables. It is clearly possible, using these tables, to compute the instantaneous velocity (i.e. the derivative) by forming the ratios $\frac{\Delta z}{\Delta t} = \frac{z(t_2) - z(t_1)}{t_2 - t_1}$ for various Δt and observing how the ratios behave when we send Δt to zero. Of course, here we cannot speak of the *limit* of the ratio $\Delta z/\Delta t$ with Δt approaching zero, since for a function specified by a table there is no way in which we can make Δt as small as desired (Δt cannot be made smaller than the step interval of the table), whereby we cannot find the exact value of the derivative. In the majority of cases, however, tables yield a sufficiently good estimate for the derivative dz/dt, which exists for all "good," or "well-behaved" (or "smooth"), functions, regardless of the origin of the functions and the way in which the functions are specified. We note also that the "arithmetic calculation of the derivative," or finding the values of the ratio $\Delta z/\Delta t$ (or, for the function $y = f(x)$, the ratio $\Delta y/\Delta x$) for a number of decreasing values of Δt (or Δx) and observing the behavior of the values of the ratio, is simplified considerably if one employs a pocket calculator.

The derivative dz/dt is defined as the limit of the ratio of the increments $\Delta z/\Delta t$ as $\Delta t \to 0$. When Δt is not equal to zero, the ratio of the increments $\Delta z/\Delta t$ is not equal in general to the derivative dz/dt, but this ratio is *approximately* equal to dz/dt and the approximation is the better the smaller Δt is. Therefore, we can write the *approximate relationship*

$$\frac{\Delta z}{\Delta t} \simeq \frac{dz}{dt} = z'(t),$$

$$\Delta z \simeq \frac{dz}{dt}\Delta t = z'(t)\,\Delta t. \qquad (2.4.6)$$

From this we can find the approximate value of the function $z(t + \Delta t)$:

$$z(t+\Delta t) = z(t) + \Delta z \simeq z(t) + \frac{dz}{dt}\Delta t$$
$$= z(t) + z'(t)\,\Delta t. \qquad (2.4.7)$$

Note that in (2.4.7) the first equality sign is exact in accord with the definition of Δz and the second one denotes approximate equality.

Let us now return to the designations $t_2 = t + \Delta t$ and $t_1 = t$ used earlier. We can then rewrite Eq. (2.4.7) as

$$z(t_2) \simeq z(t_1) + z'(t_1)(t_2 - t_1). \qquad (2.4.7a)$$

Thus, given a small difference $t_2 - t_1$, that is, t_2 is close to t_1, the function $z(t_2)$ can be expressed by an approximate formula involving the value of the function $z(t)$ and its derivative $z'(t)$ at $t = t_1$. Note that this formula is *linear* in t_2 (to the first power).

Formula (2.4.7) (we will return to this formula in Section 4.1) is extremely important, whereby we discuss it in detail. Let us take an example. Suppose $v = x^3$, and for the sake of clarity we assume that we are speaking of a cube of volume v and edge x. The derivative dv/dx, as we know, is equal to $3x^2$, with the result that (2.4.7) is

$$v(x + \Delta x) = (x + \Delta x)^3 \simeq x^3 + 3x^2\Delta x,$$

that is,

$$\Delta v = v(x + \Delta x) - v(x) \simeq 3x^2 \Delta x, \tag{2.4.8}$$

while the exact expression for $v(x + \Delta x)$ is, in view of the well-known formula,

$$v(x + \Delta x) = (x + \Delta x)^3$$
$$= x^3 + 3x^2 \Delta x + 3x(\Delta x)^2 + (\Delta x)^3,$$

whence

$$\Delta v = 3x^2 \Delta x + 3x(\Delta x)^2 + (\Delta x)^3. \tag{2.4.8a}$$

Suppose that the edge is 1 m long. Then the cube's volume is, obviously, 1 m^3. But what can we say about the volume if it is found that in measuring the edge length we introduced an error, that the edge is somewhat longer than 1 m, say, by 1 or 5 mm or even by 1 or 2 cm?

If our error amounts to 1 mm, then actually $x = 1.001$ m, and the initial value of the volume $v = 1$ must be increased by Δv given by formula (2.4.8a). Let us estimate the contribution of each term on the right-hand side of (2.4.8a):

$$3x^2 \Delta x = 3 \times 1^2 \times 0.001 = 0.003,$$
$$3x(\Delta x)^3 = 3 \times 1 \times (0.001)^2 =$$
$$= 0.000003,$$
$$(\Delta x)^3 = (0.001)^3 = 0.000000001;$$

here

$$v(1.001) = 1.003003001.$$

$$\Delta v = 0.003003001. \tag{2.4.9}$$

But is there any sense in writing out all the digits in v? The third term on the right-hand side of (2.4.8a) is smaller than the first term by a factor of 3 000 000 (three million!), whereby retaining the third term in this case cannot be justified. The second term is also smaller than the first term, by a factor of 1000, so that the precision it provides for the final result is illusory, since if the error in measuring the cube's edge is not precisely 1 mm but 1.1 mm (which, of course, is quite a realistic assumption), the first term on the right-hand side of (2.4.8a) will be equal to 0.0033 instead of the initial value of 0.003, and all digits in (2.4.9) will become unreliable. So is there any reason to perform unnecessary work and clutter up the calculations with unnecessary figures?

A similar picture arises at other values of Δx (values that can still be considered small). For instance, at $\Delta x = 0.005$ m (= 1/2 cm) we have

$$3x^2 \Delta x = 3 \times 1^2 \times 0.005 = 0.015,$$
$$3x(\Delta x)^2 = 3 \times 1 \times (0.005)^2$$
$$= 0.000075,$$
$$(\Delta x)^3 = (0.005)^3 = 0.000000125.$$

Thus, the final result is

$$v(1.005) = (1.005)^3 = 1.015075125,$$

where, of course, we can also confine ourselves (and even more, this is necessary, as a rule) to the approximation $v \simeq 1.015$.

The results of numerical calculations associated with this example are listed in detail in the table below, from which the reader can see that even at $\Delta x = 0.02$ m = 2 cm, or with low requirements concerning the accuracy of the result and at $\Delta x = 0.05$ m = 5 cm, replacing $(x + \Delta x)^3$ with $x^3 + 3x^2 \Delta x$ is quite admissible:

$x = 1 + \Delta x$	$1 + 0 = 1$	1.001	1.005	1.01
$v = (1 + \Delta x)^3$	1	1.003003	1.015075	1.0303
$1 + 3\Delta x$	1	1.003	1.015	1.03

$x = 1 + \Delta x$	1.02	1.05	1.1	1.5	2
$v = (1 + \Delta x)^3$	1.061?	1.1576	1.3310	3.375	8
$1 + 3\Delta x$	1.06	1.15	1.30	2.50	4

Another example is $y = \sqrt{x}$ (say, y is the side of a square plate with area or mass[2.6] x). We find the values

[2.6] The mass m of a homogeneous square plate whose side is y is equal to ρy^2, where ρ is a constant (the plate's density); the function $m = \rho y^2$ differs from $x = y^2$ by an unimportant factor ρ (in this connection see Section 1.7).

of the function y for x close to 4. In this case $y(4) = \sqrt{4} = 2$. The derivative $y'(x) = 1/2\sqrt{x}$ (see Exercise 2.4.2), whereby $y'(4) = 1/2\sqrt{4} = 1/4$, and the approximate formula (2.4.7) is of the form

$$y(x) = \sqrt{4+\Delta x} \simeq 2 + 0.25\Delta x.$$

We again compare the approximate and exact values of $y(x)$:

$x = 4+\Delta x$	4	4.1	4.5	5
$y = \sqrt{4+\Delta x}$	2	2.02485	2.1213	2.24
$2+0.25\Delta x$	2	2.025	2.125	2.25

$x = 4+\Delta x$	6	7	8	9
$y = \sqrt{4+\Delta x}$	2.45	2.65	2.83	3
$2+0.25\Delta x$	2.50	2.75	3.0	3.25

Let us now return to our basic example concerning distance, time, and velocity, that is, we assume that t is the time, $z(t)$ is the distance covered in time t, $z'(t) = dz/dt$ is the instantaneous velocity, and Δz is the increment in distance, that is, the distance covered during the small time interval Δt. The formula

$$\Delta z \simeq z'(t)\,\Delta t \qquad (2.4.10)$$

then signifies that the distance covered is equal to the product of the instantaneous velocity and the time interval. But the instantaneous velocity itself varies with time. Therefore, the approximation expression (2.4.10) is true only when the instantaneous velocity does not perceptibly change during the time Δt. Hence, the faster $z'(t)$ varies, the smaller Δt must be taken in (2.4.10); and conversely, the slower $z'(t)$ varies, the larger Δt may be taken. That is to say, the magnitude of the increment Δt for which formula (2.4.10) still yields a small error depends on the rate of change of the derivative over the interval Δt. Of course, the same holds for an arbitrary function $y = f(x)$ of an arbitrary independent variable x, where the derivative dy/dx may still be considered as the rate of variation of function y at a given value of independent variable x (see Figure 2.3.1 and the text referring to this figure).

The cases we have examined (where we denote by t the independent variable and by z the value of the function and assume that t is the time and z is the distance covered) confirm this conclusion. In the first example, $z = t^3$, when t varies from 1 to 2 (at $\Delta t = 1$), the derivative $z'(t) = 3t^2$ varies from 3 to 12 (which is to say, by a factor of 4). In the second example, $z = \sqrt{t}$, when t varies from 4 to 9 ($\Delta t = 5$), the derivative $z'(t) = 1/2\sqrt{t}$ varies from 0.25 to 0.167 (or roughly by 30%). Therefore, in the latter instance formula (2.4.7) or (2.4.10) yields a good result for larger values of Δt; for example, at $\Delta t = 4$, which constitutes a 100% increase against the initial value of t, the error introduced by this formula constitutes only 6% of the true value of the function, while at $\Delta t = 5$, which constitutes a 125% increase against the initial value of t, the error constitutes $1/12 \simeq 8\%$ of the true value. On the other hand, if we turn to the function $z = t^3$, we see that (2.4.7) becomes invalid already at $\Delta t = 0.5$, when the error constitutes 25% of the true result (at $\Delta t = 1$ the error constitutes a 100% of the true result). Of course, what we have said here is valid for both positive and negative values of Δt (see Exercise 2.4.5).

A detailed discussion of the range of application of formula (2.4.7), of the question of estimating the error introduced by this formula, and of the various refinements is given in Chapter 6.

Anticipating the first sections in Chapter 6, we note that for small differences $t_2 - t_1$ the function $z(t_2)$ can be represented as

$$z(t_2) = z(t_1) + a(t_2 - t_1) + b(t_2 - t_1)^2 + c(t_2 - t_1)^3 + \ldots, \qquad (2.4.11)$$

and this relationship can be considered exact if we assume that the sum on the right-hand side consists of an *infinite* number of summands (the exact meaning of this assertion will be discussed in Chapter 6). Since we assume that $t_2 - t_1$ is small, each term in (2.4.11) is smaller than the preceding one, since it contains a higher power of the small difference $t_2 - t_1$. The first two terms of the general formula (2.4.11) coincide with the right-hand side of the approximate formula (2.4.7a), since $a = z'(t_1) = dz(t_1)/dt$. The difference between the approximate formula and the exact one lies in the fact that the latter contains a term proportional to $(t_2 - t_1)^2$ and terms with higher powers of this small quantity.

The common terminology here is as follows: $t_2 - t_1$ and $a(t_2 - t_1)$ are said to be (provided that $t_2 - t_1 = \Delta t$ is small) quantities of the *first order of smallness* (here and in what follows we assume that t_2 tends to t_1 and Δt tends to zero); the quantities $(t_2 - t_1)^2$ and $b(t_2 - t_1)^2$ are said to be quantities of the *second order of smallness*; $(t_2 - t_1)^3$ and $c(t_2 - t_1)^3$ are of the *third order of smallness*; etc. With this in mind, we can say that (2.4.7a) is exact to within first-order terms, while formula (2.4.7a) is exact to within second-order terms (the word "smallness" is usually dropped).

If we employ (2.4.7a) to derive an approximate expression for the derivative, we will be forced to divide both sides of (2.4.7a) by $t_2 - t_1$, which reduces the order of smallness by one. The equality

$$z'(t_1) \simeq \frac{z(t_2) - z(t_1)}{t_2 - t_1}$$

is approximate, and the error introduced by it is not of the second but of the first order of smallness, that is, the error is proportional to $t_2 - t_1$; nevertheless, as $t_2 \to t_1$, that is, as $\Delta t = t_2 - t_1 \to 0$, the error tends to zero. This fact follows from the definition of the derivative given above.

Exercises

2.4.1. Find the derivatives of the following functions: (a) $y = x^4$, (b) $y = 4x^3 - 3x^2 + 2x - 1$, (c) $y = (2x + 1)^2$, (d) $y = 1/x^2$, (e) $y = a(x + 1/x)$, and (f) $y = ax^2 + b/x^2$.

2.4.2. Prove that the derivative of $y = \sqrt{x}$ is equal to $1/2\sqrt{x}$. [*Hint.* Multiply the numerator and denominator of $(\sqrt{x + \Delta x} - \sqrt{x})/\Delta x$ by the sum $\sqrt{x + \Delta x} + \sqrt{x}$.]

2.4.3. Prove that the derivative of $y = \sqrt{x^3}$ is equal to $(3/2)\sqrt{x}$. [*Hint.* Use the method employed in Exercise 2.4.2.]

2.4.4. Find $(1.2)^2$, $(1.1)^2$, $(1.05)^2$, and $(1.01)^2$ using formula (2.4.7). Compare the results with the exact values.

2.4.5. Using the expression for the derivative of the function $z(t) = 2 + 20t - 5t^2$, find $z(1.1)$, $z(1.05)$, and $z(0.98)$. Compare the results with the exact values. [*Hint.* In the last case take $t = 1$ and $\Delta t = -0.02$.]

2.5 A Tangent to a Curve

Using the concept of a derivative, we can solve an important geometric problem: to find the *tangent line* to a curve given by the equation $y = f(x)$. The coordinates of the point A of tangency are given: $x = x_0$ and $y = y_0 = f(x_0)$.

From the viewpoint of analytic (coordinate) geometry introduced by Descartes (see Chapter 1), to find the tangent line means to find the equation of the line. It is clear that the tangent line is one of the straight lines passing through the point of tangency. But the equation of any straight line passing through a given point $A(x_0, y_0)$ can be written as $y - y_0 = k(x - x_0)$ (see formula (1.3.3)). In order to find the equation of the tangent line, it remains to determine the quantity k, the slope of the tangent line (its "steepness"). To do this, we first find the slope of the line passing through two given points A and B of the curve at hand (Figure 2.5.1). We call this line a ***secant line.*** When these two points of the curve approach each other, the line approaches the tangent line.

In Figure 2.5.1 we see two secant lines through points A and B and through points A and C, with C lying closer to A than B. The closer point B

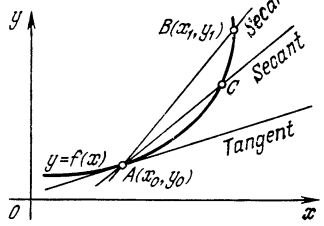


Figure 2.5.1

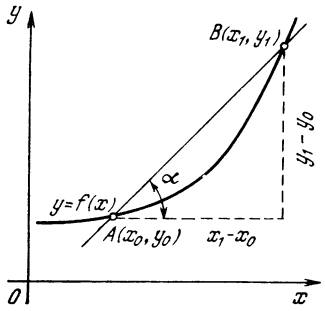


Figure 2.5.2

or C is to A, the closer the secant line AB or AC is to the tangent line. Therefore, the slope of the tangent line is equal to the *limit* approached by the slope of the secant line as the distance between the two points of intersection of the secant and the curve tends to zero (or the points of intersection approach each other). This fact can be taken as the *definition* of the tangent line; of course, it would be more precise to say that point B tends to point A or that the two points B and C tend to A and the secant line BC tends to the tangent line at the chosen point A.

The slope k_s of a secant line can easily be expressed in terms of the coordinates of the intersection points. For one of the points of intersection of the secant and the curve we take the point $A(x_0, y_0)$ at which we desire to draw a tangent line to the curve; we denote by x_1 and y_1 the coordinates of the second point of intersection, B. Since both points lie on the curve whose equation is $y = f(x)$ ("belong to this curve" is usual phrase, it follows that $y_0 = f(x_0)$ and $y_1 = f(x_1)$. As can be seen from Figure 2.5.2 the slope of the secant line is

$$k_s = \tan\alpha = \frac{y_1 - y_0}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

(see Section 1.3).

In order to obtain the slope of the tangent line at the point (x_0, y_0), we must take point B closer and closer to A, which means that x_1 must approach x_0. Consequently, the slope k of the tangent line is equal to the limit of k_s as x_1 tends to x_0:

$$k = \lim_{x_1 \to x_0} k_s = \lim_{x_1 \to x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

We denote the difference $x_1 - x_0$ by Δx; then $x_1 = x_0 + \Delta x$ and $\Delta f = f(x_1) - f(x_0) = f(x_0 + \Delta x) - f(x_0)$. In this notation, the slopes k_s and k of the secant and tangent lines, respectively, are given by the formulas

$$k_s = \frac{\Delta f}{\Delta x}, \quad k = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}.$$

Thus, *the slope of the tangent line is equal to the derivative of the function* $f(x)$ *at point* x_0:

$$k = \frac{df}{dx} = f'(x_0). \qquad (2.5.1)$$

We know that the derivative $f'(x)$ of a function $f(x)$ is itself a function of x. Since we sought the slope of the tangent line at a fixed point $A(x_0, y_0)$, we assumed, in computing the limit of $\Delta f/\Delta x$, that $x = x_0$ is fixed. That is why in the final formula we have $f'(x_0)$, which is the value of the derivative $f'(x)$ at $x = x_0$. On the other hand, the function $f'(x)$, obviously, expresses the slope of a tangent line at a variable point (x, y) or $(x, f(x))$ of this curve.

Let us consider the *example* of a *parabola* $y = x^2$, that is, $f(x) = x^2$. We set up the equation of the tangent line at the point $x_0 = 2$, $y_0 = f(x_0) = 4$.

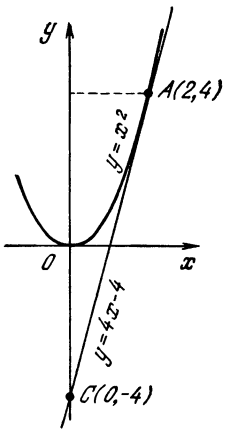


Figure 2.5.3

We know the derivative of x^2:

$$f'(x) = \frac{df}{dx} = \frac{dx^2}{dx} = 2x$$

(cf. (2.3.6a)). Consequently, at the point of interest the slope of the tangent line is

$$k = f'(x_0) = 2x_0 = 4,$$

while the equation of the tangent line is (Figure 2.5.3)

$$y - y_0 = k\,(x - x_0), \quad \text{i.e.}$$

$$y - 4 = 4\,(x - 2),$$

or

$$y = 4x - 4.$$

In general, the slope of the tangent line to the parabola $y = x^2$ at point (x_0, x_0^2) is equal to $2x_0$ (since here $dy/dx = 2x$); hence, the equation of the tangent line is

$$y - x_0^2 = 2x_0\,(x - x_0), \quad \text{or} \quad y = 2x_0 x - x_0^2. \tag{2.5.2}$$

We have thus arrived at the result established in Section 1.4 by a different method, namely, that the straight line $y = kx + b$ touches the parabola $y = x^2$ if and only if the coefficients k and b in the equation of the straight line can be represented as $2x_0$ and $-x_0^2$, which depend on the parameter x_0; in other words, the condition for a straight line $y = kx + b$ touching a parabola $y = x^2$ can be expressed as $b = -(k/2)^2$ (cf. (1.4.8a)).

Here is another example. Take the *semicubical parabola* $y = \sqrt{x^3}$ (see Section 1.5). Here $y' = (3/2)\sqrt{x}$ (see Exercise 2.4.3), whence $k = f'(x_0) = (3/2)\sqrt{x_0}$. In particular, at the origin the slope of the tangent line is $k = f'(0) = (3/2)\sqrt{0} = 0$; whence, at point 0 the semicubical parabola touches the x axis, as depicted in Figure 1.5.7.

Without the aid of derivatives, it is rather difficult to draw a tangent line to a curve given by the equation $y = f(x)$: you have to compute a large number of points of the curve, then, using a French curve, draw the curve through these points and, by eye, apply a ruler to the curve at the given point and pay special attention to see that you do not intersect the curve near the point of tangency. Using derivatives, we find the equation of the tangent line, then from this equation we find two points lying on the straight line given by this equation, and then we draw the straight line (tangent line) with a ruler (through the two points). For one of the two points it is natural to take the point of tangency itself, $A\,(x_0, y_0)$. The second point C may be taken on the straight line at a good distance from A (the farther the better); we can then more accurately determine the slope and the position of the tangent as the straight line passing through the two points A and C.

For example, above we found the equation of a straight line tangent to the parabola $y = x^2$ at the point $x_0 = 2$, $y_0 = 4$. It has the form $y = 4x - 4$. Let us find the coordinates of two points on this line: at $x = 2$ we find that $y = 4$. This is the point of tangency $A\,(2, 4)$; the coordinates need not have been computed since the tangent must pass through it. For the second point, C, we choose the point of intersection of the tangent line and the y axis. Putting $x = 0$, we find $y = -4$, so that $C = C\,(0, -4)$ (see Figure 2.5.3).

Note the curious fact that *for* $x = 0$, $y = -y_0$ *the point* C *of intersection of the tangent line with the* y *axis lies below the* x *axis just as much as the point of tangency itself lies above the* x *axis.* This is not accidental. The rule holds true for all tangents to quadratic parabolas with an equation $y = ax^2$ (with $a > 0$). Indeed, if the tangent is drawn to the point $A(x_0,\ y_0 = ax_0^2)$, its equation is

$$y - y_0 = 2ax_0\,(x - x_0) \tag{2.5.2a}$$

(cf. (2.5.2)), and for $x = 0$ we get $y - y_0 = -2ax_0^2$, or $y = y_0 - 2ax_0^2 = y_0 - 2y_0 = -y_0$. Thus, the tangent passes through the points $A\,(x_0,\ y_0 = ax_0^2)$ and $C\,(0,\ y = -y_0 = -ax_0^2)$.

When plotting a curve by points it is hard to construct the curve if there are few points. Using derivatives, you can draw the tangents to the curve at these points beforehand, and then the curve itself can be drawn with greater ease and accuracy.

Pictorially, it is clear that the tangent is horizontal at the points of *maximum* and *minimum*; in future we will return to this aspect many times. The equation of a horizontal straight line is y = constant, and the slope of the horizontal straight line is $k = 0$.

Hence, we arrive at the following theorem: *the derivative of a function* $y = f(x)$ *(the graph of which is a curve) is zero at the points of minimum and maximum of the curve.*

Using this theorem, one can find the x-coordinates of the points of maximum and minimum of the curve. The respective y-coordinates can then be easily found by substituting the established values of x into the equation of the curve. It is also obvious that knowing the coordinates of the points of maximum and minimum we can draw the curve itself more accurately.

It is a useful exercise to draw a curve $y(x)$ freehand and then, at least approximately but rapidly, draw the curve $y'(x)$, noting the sign of $y'(x)$ and the points where $y'(x)$ vanishes. This is illustrated in Figure 2.5.4*a* (the graph of $y(x)$) and Figure 2.5.4*b* (the graph of the derivative $y'(x)$).

For the derivative $y'(x)$, the points at which $y(x)$ vanishes are of no interest. If the curve $y(x)$ is moved upward or downward (in Figure 2.5.4*a* the curve has been moved upward and the result is depicted by the dashed curve) by an arbitrary segment b, the curve $y'(x)$ does not change in any way because in translation in the vertical direction all the slopes remain the same (to be precise, we must speak of the slopes of the tangent lines to this curve at all points: compare the upper and lower curves in Figure 2.5.4*a* and, in particular, the tangents to these curves at points A and B). This result is in accord with the property of derivatives: the graphs of the functions $y = f(x)$ and $y_1 = f(x) + b$ (the graphs of these functions are obtained through translation upward or downward) have equal derivatives at corresponding points.

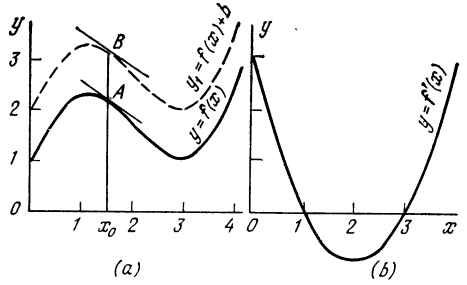


Figure 2.5.4

Another mathematical game is this: draw freehand the graph of the derivative and then give a rough construction of the graph of the function. Here you have to specify in arbitrary fashion one point $(x_0, y(x_0))$ on the curve and then draw the graph up or down (depending on the sign of the derivative).

In conclusion we must dwell on one important point closely related, as we will see below, with the question discussed concerning the connection between the slope of a tangent line to a curve and the derivative (above we ignored this point on purpose). Note that in physical problems, the quantities x and $f(x)$ are usually dimensional (for example, x or t is time and $y = f(x)$ or $z = f(t)$ is distance). But this means that dy/dx is dimensional, too. Clearly, when z is measured in centimeters and t in seconds, $dz/dt = v(t)$ is the velocity, with dimensions cm/s. If, say, y is expressed in kilograms and x in months (say, $y = f(x)$ represents the weight of a baby, or its mass, as a function of time), the derivative dy/dx (= lim $\Delta y/\Delta x$) expresses the rate of weight gain by the baby and has the dimensions kg/month. If y is the current in a conductor and x is the conductor's resistance, then the derivative $y' = dy/dx$ is measured in A/Ω. If x is the edge of a cube and y its volume, then dy/dx has the dimensions of

$cm^3/cm = cm^2$. Examples of this type abound. In general, if the independent variable x is measured in units of e_1 and the function $y = f(x)$ in units of e_2, then the derivative dy/dx has the dimensions of e_2/e_1.

But, as we know, the trigonometric function $\tan\alpha$ is dimensionless (being equal to the ratio of the lengths of two line segments). Therefore we cannot simply write $dy/dx = \tan\alpha$ since the left and right members have different dimensions. Only in the rare instances when the two quantities, x and y, have the same dimensions (say, when we are studying the speed of a yacht as a function of the speed of wind) or both are dimensionless (say, the function $y = \sin x$, where x is measured in radians) does the relationship $dy/dx = \tan\alpha$ have a meaning.

How can we get round this difficulty? Note that above we assumed the scales along the x and y axes to be the same, that is, that the unit of measurement for x and the unit of measurement for y are depicted by the same segments; this assumption seemed so natural that we did'nt even spell it out. But in the case where x and y have different dimensions (and this is the general case) the above assumption is not only unnatural but really has no meaning, since there is no way in which we can assume that 1 s = 1 cm. Therefore, in reality the scales along the two axes are *different*, and because of this we cannot write $dy/dx = \tan\alpha$.

Suppose that y is the distance traveled and x is the time. We construct the graph of the position of a body depending on the time, $y = y(x)$. On the axis of ordinates we lay off y using the scale: 1 meter of distance equals 1 cm in the drawing. On the axis of abscissas we lay off time using the scale: 1 second of time equals 1 cm in the drawing. Then the velocity v expressed in meters per second and equal to the derivative dy/dx will indeed be equal to $\tan\alpha$, the tangent of the angle formed by the tangent line and the x axis. But if we choose a different scale for x say, 1 s = l cm in the drawing, we get

$$\tan\alpha = \frac{dy}{l\,dx} = \frac{1}{l}\frac{dy}{dx}.$$

If $l = 5$, we have $\tan\alpha = (1/5)\,dy/dx = (1/5)\,v$.

In the general case, if one x unit in the drawing is laid off to a scale of l_1 cm and one y unit is laid off to a scale of l_2 cm, then

$$\tan\alpha = \frac{l_2}{l_1}\frac{dy}{dx}. \qquad (2.5.3)$$

The scale factors l_1 and l_2 in this formula improve the situation—they make the formula proper from the standpoint of dimensions. Thus, in the example with a baby's weight, l_1 has the dimensions of cm/month (1 cm in the drawing per month of age) and l_2 has the dimensions of cm/kg (1 cm on the graph per kilogram of weight), so that $(l_2/l_1)\,dy/dx$ is dimensionless: in the formula all the dimensions cancel out and the formula becomes meaningful (and correct).

All this should be borne in mind when comparing the derivative and the slope of a curve representing the basic function (i.e. the slope of the tangent line to the graph at the point of interest).

Exercises

2.5.1. Construct the graph of the function $y = x^2 + 1$ within the range from $x = -1.n$ to $x = 2.5$ and draw the tangent lines at the points $x = -1$, $x = 0$, $x = 1$, and $x = 2$.

2.5.2. Construct the graph of the function $y = x^3 - 3x^2$, $-1 < x < 3.5$; draw the tangent lines at $x = -1$, 0, 3. Find the points with horizontal tangent lines.

2.5.3. On the curve $y = x^3 - x + 1$ find points having horizontal tangents. Construct the curve for $-2 < x < 2$. [*Hint.* In Exercises 2.5.1 to 2.5.3 it is advisable to use graph paper and a large scale.]

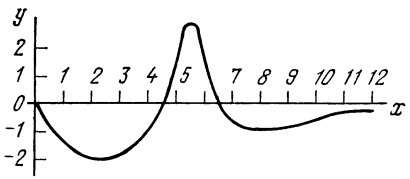


Figure 2.5.5

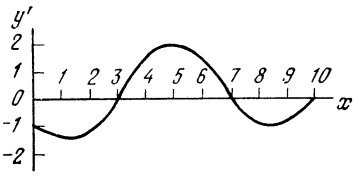


Figure 2.5.6

2.5.4. Construct (freehand) the curve $y'(x)$ for the function $y(x)$ depicted in Figure 2.5.5.

2.5.5. The graph of a derivative $y'(x)$ is depicted in Figure 2.5.6. Construct (freehand) the graph of the function $y(x)$ passing through the point (5, 0). At what angle will $y(x)$ intersect the axis of ordinates? At what angle will $y(x)$ intersect the axis of abscissas at point $x = 5$? [*Hint.* In Exercises 2.5.4 and 2.5.5, first copy Figures 2.5.5 and 2.5.6 on a fresh sheet of paper and then construct the respective graphs; it is advisable to construct these graphs on the same sheet of paper with the initial graphs, say, strictly below the respective graph.]

2.5.6. Set up the equations of the tangent lines to the cubical parabola $y = x^3$ at the points with (a) $x = 0.5$ and (b) $x = 1$. Find the points of intersection of the tangent lines with the x and y axes.

2.5.7. Find the general rule that enables determining the points of intersection with the x and y axes of the straight lines tangent to the curves (a) $y = ax^2$ and (b) $y = bx^3$ at point $(x_0, y(x_0))$.

2.6 Increase and Decrease of Functions. Maxima and Minima

Observing a curve that depicts the behaviour of a function $y = f(x)$ (say, the dependence of temperature $T(t)$ on time t), one can easily see the points at which the function *grows*, or *increases* (say, point A in Figure 2.6.1), the points at which the function *decreases* (say, point B), the points of *maximum*, where growth of the function is replaced with its decrease (point C), and points of *minimum*, where decrease of the function is replaced with its growth (point D). But how must we define such concepts with sufficient mathematical rigor and how to determine, without looking at the graph, the behavior of a function at, say, point $x = x_0$ (drawing a graph always involves errors and therefore constitutes a nonreliable method)?

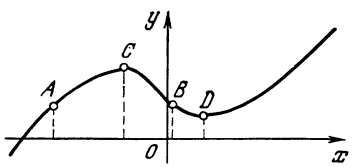


Figure 2.6.1

Without a knowledge of derivatives, we have to seek the answer to these questions numerically, that is, we must take the temperature T at a given time t and then take it at some following time t_1 and see if it has increased or decreased. This is clearly not a reliable approach: even if $T(t_1)$ is greater than $T(t)$, it still might be that at time t the temperature fell, then soon afterward (after t but prior to t_1) it reached a minimum, and only then began to grow and by t_1 had risen above $T(t)$.

Using derivatives we get an exact solution: we have to find the derivative dy/dx. If $dy/dx = y'(x)$ is positive for a given x, then $y(x)$ is an increasing function, that is, if x increases by a small quantity Δx, the value of y increases by a small amount $\Delta y \simeq y'(x)\,\Delta x$ (as was clarified earlier, the smaller the value of Δx the more exact the equation). We consider $\Delta x > 0$ (say, time x increases). If $y'(x) > 0$, $\Delta x > 0$, then, of course, $\Delta y > 0$, that is, the value of y increases with x (say, the temperature increases in time). The numerical value of the (positive) derivative shows *how fast* y (the temperature) rises: in the example with the temperature we find that if $T'(t) = 10$, then in the vicinity of t the temperature increases 10 times faster than time (for example, by 10 °C each second), while if $T'(t) = 0.1$, then the temperature increases 10 times slower than time (however, compare this with what was said in Section 2.5 about the dimensions of the derivative and its connection with the choice of units for the independent variable and the function). But if $y'(x) < 0$, $\Delta x > 0$, then $\Delta y < 0$; for instance, if $T'(t) < 0$, the temperature $T(t +$

$+ \Delta t$) at $t + \Delta t$ will be *lower* than T (t) at the given moment in time. Thus, a *positive* derivative indicates that the ***function is increasing*** and a *negative* derivative that the ***function is decreasing***.

The expressions "increasing function" and "decreasing function" are applied to any functions y (x) and not only to those that depend on time (functions of time); here the independent variable may be an arbitrary quantity (with or without dimensions). An *increasing* (*decreasing*) function is one in which y increases as the independent variable x increases (decreases). The derivative dy/dx is what indicates the ***rate of growth***, that is, the ratio of the variation in y to the corresponding (small) variation in x. A negative rate of growth means a decrease in y as x increases, and if $dy/dx < 0$, then $| dy/dx | = -dy/dx$ is the rate at which the function decreases.

The expression "the quantity y has a large negative derivative with respect to x" means that y decreases rapidly as x increases, while a positive dy/dx means that y increases with x. In this way the derivative of y points to the *tendencies* in the variation in y and enables us to know what to expect from further variations in the independent variable. This constitutes the main reason for studying the derivative of a function.

Physicists and mathematicians, especially those in the making (who have just learned what a derivative is), frequently put it to use in everyday life like this: "the derivative of my mood with respect to time is positive" in place of "my mood is definitely improving."

Solve this joke problem: what sign does the derivative of my mood have with respect to the distance from the dentist's chair? My mood deteriorates, "decreases," becomes "negative" as the distance decreases; hence, the derivative is positive.

The serious editor may complain about abuse of the English language, but actually this free-style use of mathematical concepts is good practice for further serious applications of mathematics.

There are functions that have the same sign of the derivative for any values of the variable: such is the property of the linear function $y = kx + b$, since here the derivative $dy/dx = k$ is a constant. Later on we will see that in the case of the exponential function $y = a^x$ the derivative has a constant sign (although it is not constant in magnitude) for arbitrary x. However, a derivative need not have a constant sign; the sign of the derivative of a given function may be different for different values of the independent variable.

Let us imagine a function y (x) whose derivative $y'(x)$ is positive for $x < x_0$ and negative for $x > x_0$; in short, $y'(x) > 0$, $x < x_0$, and $y'(x) < 0$, $x > x_0$.

What can we say about such a function? We begin with $x < x_0$. As x increases to x_0, the function y (x) will increase; however, as x continues to increase the derivative becomes negative and y falls. The conclusion is that for $x = x_0$ the function y (x) has a *maximum*.

Consider the contrary case: $y'(x) < 0$ for $x < x_0$ and $y'(x) > 0$ for $x > x_0$. Reasoning as before, we conclude that in this case y (x) has a *minimum* at $x = x_0$.

If a function y (x) is defined by a formula associated with a smooth curve, so that $y'(x)$ also varies smoothly with x, then the different signs of $y'(x)$ for $x < x_0$ and $x > x_0$ in both cases signify that the derivative *vanishes* at $x = x_0$:

$$y'(x_0) = \frac{dy(x_0)}{dx} = 0. \qquad (2.6.1)$$

Thus, as already noted in Section 2.5, *by equating the derivative to zero we can find those values of the independent variable for which the function has a maximum or a minimum.*[2.7] Now we can re-

[2.7] More precisely, the values at which the function *may* have (but also may not have) a maximum or a minimum. For example, the

fine the general theorem discussed in Section 2.5: at the points where a function attains *maxima* the derivative changes its sign from *plus to minus*, while at the points where a function attains *minima* the derivative changes its sign from *minus to plus*.

Here are some examples. First we turn to the function $y = 3x^3 - x^2 - x$ discussed in Section 1.1 (see the table for this function in Section 1.1). Judging by this table, one might think that the function increases for all values of x since every increase in x by unity caused an increase in y.

Let us calculate the derivative, however:

$$y'(x) = 9x^2 - 2x - 1.$$

Taking $x = 0$, we find that $y'(0) = -1 < 0$, which means that when $x = 0$, the function is *decreasing*. This refutes the supposition (obtained from a glance at the table) that the function is an everywhere increasing function.

We equate $y'(x)$ to zero. Solving the equation $y'(x) = 9x^2 - 2x - 1 = 0$ yields two roots: $x_1 \simeq -0.24$ and $x_2 \simeq 0.46$.

Now we form a detailed table including the maximum and minimum points just found:

x	-3	-2	-1	-0.3	-0.24	-0.18
y	-87	-26	-3	0.129	0.140	0.131

x	0	0.40	0.46	0.52	1	2
y	0	-0.372	-0.381	-0.370	1	18

We see that, true enough, on the portion from $x \simeq -0.24$ to $x = 0.46$ the function y falls from $+0.14$ to -0.38. A comparison of the value $y(-0.24)$ with the adjacent values $y(-0.30)$ and $y(-0.18)$ confirms the fact that when $x \simeq -0.24$, y reaches a (local) maximum, the adjacent values of y being smaller. The graph of the function $y = 3x^3 - x^2 - x$ is given in Figure 2.6.2.

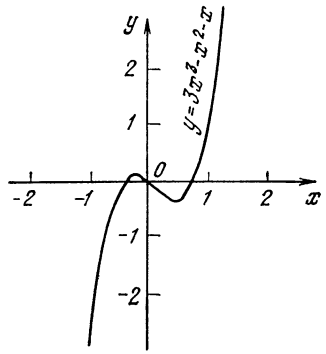


Figure 2.6.2

This example shows once more that the word "maximum" should not be understood as meaning the largest of all possible values of y. Indeed, at the maximum point $y(-0.24) \simeq 0.14$, while $y = 1$ for $x = 1$, $y = +18$ for $x = 2$, $y = 269$ for $x = 10$, and so on, y rapidly increasing without bound as x increases without bound. In what way does the maximum point $x_{max} \simeq -0.24$, $y_{max} \simeq 0.14$ that we found differ from all other points of the curve?

The difference is that for *close-lying* values of x, both larger than x_{max}, and less than x_{max} the value of y is less than $y_{max}(=y(x_{max}))$. This peculiarity of x_{max} is clearly seen in the table (e.g. compare $y(-0.30)$, $y(-0.24)$, and $y(-0.18)$). The same arguments can be applied to the minimum $x_{min} \simeq 0.46$, $y_{min} \simeq -0.381$: for large (in absolute value) negative x's the value of y decreases without bound and becomes less than y_{min} (and, in general, y can be made less than any negative number), but x_{min} differs in that the value y_{min} $(= (y(x_{min}))$ is less than the values of y for x close to x_{min}. The condition of a vanishing derivative enables finding just such (*local*) maxima and minima.

For the second example we take the function $y = x^3 - x$ discussed in Sec-

function $y = x^3$ has a derivative, $y' = 3x^2$, which vanishes at $x = 0$; however, at this point the function (see its graph in Figure 1.5.1*a*) has neither a maximum nor a minimum. (Other exceptions to our rule that refer to curves that are not smooth are discussed in Section 7.2.)

tion 1.5. The derivative of this function is $y' = 3x^2 - 1$, whereby $y'(x) = 0$ at $x = \pm 1/\sqrt{3}$. Let us see how the sign of the function $y'(x) = 3x^2 - 1$ varies in the vicinity of points $x = 1/\sqrt{3} \simeq 0.577$ and $x = -1/\sqrt{3} \simeq -0.577$. We will readily find that at $x = -1/\sqrt{3}$ the function $y(x)$ has a (local) maximum and at $x = 1/\sqrt{3}$ the function $y(x)$ has a (local) minimum (see Figure 1.5.3).

Now we can give a complete explanation why the graphs of the functions $y = x^3 + x$ and $y = x^3 - x$ (Figures 1.5.2 and 1.5.3) differ. Let us consider the general equation

$$y = x^3 + cx, \qquad (2.6.2)$$

where c is an arbitrary number. In this case $y' = 3x^2 + c$. It is clear that for $c > 0$ the derivative y' is positive everywhere, which means that y increases for all values of x, that is, there are neither maxima nor minima. On the other hand, for $c < 0$ the derivative y' vanishes at $x = \pm\sqrt{-c/3} = \pm x_0$, and in the interval from $x = -\infty$ to $x = -x_0$ the derivative y' is positive (the function is increasing), for $-x_0 < x < x_0$ the derivative is negative (the function is decreasing), and for $x_0 < x < \infty$ the derivative is again positive (the function is again increasing); here the symbols $-\infty$ and ∞ denote, as usual, very large negative and positive numbers. Thus, at $x = -x_0$ the function has a maximum and at $x = x_0$ it has a minimum. If we make c smaller and smaller (in absolute value), the maximum and minimum approach each other: at $c = 0$ the points $\pm x_0 = \pm\sqrt{-c/3}$ merge into a single "critical" point $x = 0$, the point where the derivative of y vanishes. In this case (for the curve $y = x^3$) the critical point is neither a maximum nor a minimum, since if we consider "neighboring" curves corresponding to negative values of c that are small in absolute value, we might decide that our point is simultaneously a maximum and a minimum,

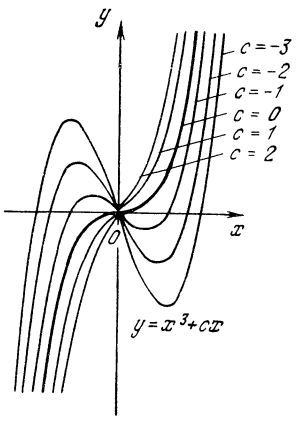


Figure 2.6.3

which is clearly impossible, and it is clearly unjustified to expect that the point is a maximum (and not a minimum) or a minimum (and not a maximum). All these peculiarities of the curves specified by Eq. (2.6.2) are shown in Figure 2.6.3. We see that there are two distinct cases: $c < 0$ and $c > 0$; the "intermediate" case corresponds to the "limit" curve $y = x^3$, for which $c = 0$.

Let us return to Exercise 2.2.1 devoted to the specific heat capacity of diamond. There we saw that there exists an empirical relationship between the temperature T and the quantity of heat $Q = Q(T)$ (in joules) required to heat 1 kg of diamond from 0 to T °C:

$$Q(T) = 0.3965T + 2.081 \times 10^{-3}T^2 - 5.024 \times 10^{-7}T^3,$$

which implies that the specific heat capacity of diamond, $c = c(T)$ (in J/kg·°C), is expressed by the formula

$$c(T) = 0.3965 + 4.162 \times 10^{-3}T - 15.072 \times 10^{-7}T^2$$

(see the solution to Exercise 2.2.1). To analyze the behavior of the specific heat capacity of diamond in the temperature interval in which the empirical formula is valid, we need the expression for the derivative of c:

$$c'(T) = 4.162 \times 10^{-3} - 30.144 \times 10^{-7}T.$$

It is clear that $c'(T)$ is positive for $T < (4.162/30.144) \times 10^4 = T_0 \simeq 1380$ °C, is zero at $T = T_0$, and is negative for $T > T_0$.

But, as noted earlier, the above empirical formula is valid only up to 800 °C. All the above formulas imply that $c'(T)$ vanishes at

$T \simeq 1380$ °C, that $c\,(T)$ vanishes at $T \simeq 2850$ °C, and that the quantity of heat $Q\,(T)$ vanishes at $T \simeq 4320$ °C. Of course, all these results are meaningless and prove that empirical formulas cannot be employed outside their range of application.

Determining maxima and minima arithmetically (by computing and comparing the values of the function for different values of the independent variable) is many times more arduous and less exact. However, if you employ a pocket calculator, then an "arithmetic experiment" enables finding a maximum or a minimum by meaningfully comparing the values of the function at different points (a calculator must be used, of course, in a proper manner), and our problem is considerably simplified. But even in this case the use of derivatives makes the calculations more transparent and simple. Higher mathematics is not only a remarkable achievement of the mind. Practical computational problems are resolved much more easily by the methods of higher mathematics.

Exercises

2.6.1. Find the values of x for which the following functions have a maximum or a minimum. In each case determine whether the minimum or maximum is involved. For functions involving constants give the answer for various values of these constants (in particular, with different signs): (a) $y = ax^2$, (b) $y = x + 1/x$, (c) $y = ax + b/x$, (d) $y = x^3 - 3x + 100$, (e) $y = x^3 + px^2 + qx + r$, (f) $y = x^4 + ax^2 + b$, and (g) $y = ax^2 + b/x^2$.

2.6.2. The temperature dependence of the volume of 1 g (or cm^3) of water is given by the following empirical formula: $v\,(t) = 1 + 8.38 \times 10^{-6}\,(t - 4)^2$. At what temperature will the volume be minimal?

2.6.3. Solve Exercise 2.6.2 with the following refinements (suggested by various scientists) of the above empirical formula: (a) $v\,(t) = 1 - 61.045 \times 10^{-6}t + 77.183 \times 10^{-7}t^2 - 37.34 \times 10^{-9}t^3$, and (b) $v\,(t) = 1 - 57.577 \times 10^{-6}t + 75.601 \times 10^{-7}t^2 - 35.07 \times \times 10^{-9}t^3$.

2.7 The Second Derivative of a Function. Convexity and Concavity of a Curve. Points of Inflection

Let us once more study the behavior of a function near its "critical" points (i.e. points at which it might have a maximum or a minimum) $x = x_0$ of the curve $y = y\,(x)$ representing the function, or points at which

$$y'(x_0) = 0. \tag{2.7.1}$$

How to distinguish a maximum from a minimum if condition (2.7.1) is satisfied? We know that condition (2.7.1) holds true both for (local) maxima and (local) minima, the difference being in the sign of $y'(x)$ for $x < x_0$ and for $x > x_0$.

But how is it possible to determine the sign of $y'(x)$ for x close to x_0 without computing y' directly for other values of x? Let us start with the case of a *maximum* of the function $y\,(x)$, when $y'(x) > 0$ for $x < x_0$ and $y'(x) < 0$ for $x > x_0$. We see that here the derivative $y'(x)$ is itself a *decreasing* function: as x increases, the derivative, which was at first positive (for $x < x_0$), vanishes (when $x = x_0$) and, continuing to fall, becomes negative when $x > x_0$. But we already know how to distinguish a decreasing function from an increasing function: its derivative is *negative*. Hence, at a value $x = x_0$ at which y has a *maximum* the derivative $y'\,(x_0)$ vanishes and the derivative of the derivative is *negative*. This quantity, the derivative of a derivative, which by the ordinary rules can be written as a "double-decker" fraction,

$$\frac{dy'}{dx} = \frac{d\left(\frac{dy}{dx}\right)}{dx},$$

is called the *second derivative*. It has the notation $y''(x)$ or d^2y/dx^2 and is read "d two y over d x squared."

It is clear that if $z = z\,(t)$ is a function that specifies the dependence of distance on time (see Section 2.1), then $z'(t) = v$ is the velocity of the motion and $z''(t) = dv/dt$ is the rate of change of velocity, or the *acceleration* (see also Chapter 9). If y has the dimensions of distance (cm) and x is time (s), d^2y/dx^2 has the dimensions of acceleration (cm/s^2); if y is measured in units

of e_1 and x in units of e_2, the derivative d^2y/dx^2 has the dimensions of e_1/e_2^2, as indicated by the notation $d^2y/(dx)^2$. It is advisable to bear in mind the dimensions of the second derivative when dealing with physical problems that involve $y''(x)$.

It is the meaning of the second derivative, acceleration, in our distance-time problem that makes the idea of a second derivative so important to physics, since by Newton's second law the acceleration is the basic characteristic of motion. In Chapter 9 we discuss all these matters in greater detail.

Let us revert to the examples given above. If $y(x) = 3x^3 - x^2 - x$, then $y'(x) = 9x^2 - 2x - 1$ and, hence,

$$y''(x) = (y'(x))' = 18x - 2.$$

At $x \simeq -0.24$ we have $y' = 0$ and $y'' \simeq -6.3 < 0$, and, indeed, point $x \simeq -0.24$, $y \simeq 0.14$ is a *maximum* of $y(x)$. At $x \simeq 0.46$ we have $y' = 0$ and $y'' \simeq 6.3 > 0$, that is, point $x \simeq 0.46$, $y \simeq -0.38$ is a *minimum* of the function.

Similarly, if $y = x^3 - x$, then $y' = 3x^2 - 1$ and $y'' = 6x$. Therefore, at $x = -1/\sqrt{3} \simeq -0.577$ the second derivative y'' is negative, and at $x = +1/\sqrt{3}$ it is positive: at the first point the function has a *maximum* and at the second a *minimum*. (This result could also be predicted geometrically, since at $x = -1$ and at $x = 0$ the function $y = x^3 - x = -(x - x^3)$ vanishes and at the intermediate points it is positive, so that at $x \simeq -0.577$ there can be only a *maximum* and not a minimum.) In exactly the same way, in the example with the specific heat capacity of diamond (see Section 2.6) the second derivative $c''(T) \simeq -3.0144 \times 10^{-6}$ is always negative, which would mean that the curve representing $c(T)$ has a maximum if our formulas could be applied at T_0 where $c'(T_0) = 0$ (actually the specific heat capacity of diamond, just as that of other solids, increases with T and tend to a constant value at high temperatures).

To summarize, then, if at a certain value of x the function $y(x)$ is such that

$$y'(x) = 0, \quad y''(x) < 0, \tag{2.7.2a}$$

then at this point the function has a *maximum*, and if at a certain value of x the function $y(x)$ is such that

$$y'(x) = 0, \quad y''(x) > 0, \tag{2.7.2b}$$

then at this point the function has a *minimum*. (The reader is advised to return to the exercises accompanying Section 2.6 and apply the second-derivative criterion, which enables distinguishing between a maximum and a minimum.)

We see that the sign of the derivative of a function $y(x)$ has a simple geometrical meaning: the fact that the derivative is *positive* means that the function is *increasing*, that is, its graph is directed upward if we move along the x axis from left to right, while if $y'(x) < 0$, the function is *decreasing*, that is, its graph is directed downward. The sign of the second derivative $y''(x)$ also has a simple geometrical meaning. If the second derivative of a function is positive, or $y''(x) > 0$, this means that the first derivative $y'(x)$ is increasing or, in other words, the angle α (where $\tan \alpha = k = y'(x)$) formed by the tangent line to the curve and the x axis increases. But in this case, as we move from left to right along the x axis, the tangent line rotates counterclockwise (point A in Figure 2.7.1). And this, obviously, means that our curve is *convex downward* (cf. Section 1.4); sometimes it is said that the curve is *concave upward*.

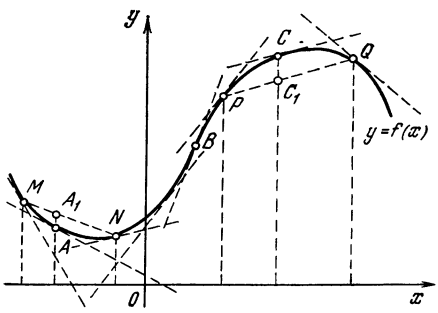


Figure 2.7.1

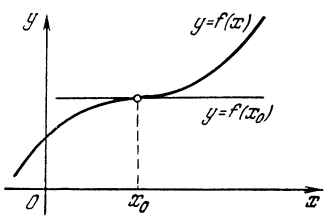


Figure 2.7.2

Similarly, the fact that $y''(x) < 0$ means that the first derivative $y'(x)$ of the function $y = f(x)$ decreases, and so does the angle α. Thus, at the points where $y''(x) < 0$, the tangent line rotates clockwise as we move from left to right along the x axis, that is, the curve is *covvex upward*, or *concave downward* (point C in Figure 2.7.1). The points at which $y''(x) = 0$ can be generally characterized as the points where the second derivative changes its sign, that is, points at which concavity is replaced with convexity or the other way round.[2.8] At such points the tangent to the graph of the function goes from one side of the curve to the other, that is, the tangent *intersects* the curve. Points at which the tangent intersects the curve (point B in Figure 2.7.1) are known as ***points of inflection.***[2.9]

For instance, when the graph of a function $y = f(x)$ is intersected by a horizontal tangent at point $x = x_0$ (Figure 2.7.2), this point is neither a maximum nor a minimum of the function, while the derivative at the point is zero (the tangent line is horizontal). Thus, we see that the cases where condition (2.7.1) does not work, that is, the point where the derivative vanishes is neither a maximum nor a minimum, are due to the fact that in addition to (2.7.1) the following condition holds true at this point:

$$y''(x_0) = 0, \tag{2.7.3}$$

which characterizes a point of inflection. (On the other hand, compare what we have just said with the footnote 2.8, where the function $y = x^4$ satisfies both conditions, (2.7.1) and (2.7.3), at point $x = 0$ and nevertheless has a maximum at this point.)

In Section 7.4 we will once more encounter the concept of convexity of a function.

Exercises

2.7.1. Find the second derivatives of the following functions: $y = x^2$, $y = x^3$, $y = x^4$, and $y = ax^2 + bx + c$.

2.7.2. Find the acceleration of a point moving along a straight line according to the following law: $z = at^2 + bt + c$. What can be said about the acceleration?

2.7.3. Specify the ranges of convexity and concavity for the following functions: (a) $y = x^3$, (b) $y = x^3 + px^2 + qx + c$, and (c) $y = x + 1/x$.

[2.8] The rare instances of the type of the origin for the curve $y = x^4$ (see Figure 1.5.1*b*; here $y' = 4x^3$, $y'' = 4 \times 3x^2 = 12x^2$, so that $y'' = 0$ at $x = 0$, while the curve does not change the direction of convexity at this point) refer exclusively to the points at which the first, second, and third derivatives vanish simultaneously (or the second derivative has a salient point or a discontinuity). (The third derivative is the derivative of the second derivative and is written as y''' or d^3y/dx^3 and has the dimensions of e_1/e_2^3, where e_1 and e_2 are the units of measurement of y and x. Below we will encounter third and higher-order derivatives.)

[2.9] Note that at a point of inflection the tangent intersects the curve smoothly, touching it without a break, contrary to the case where the y axis intersects the parabola $y = x^2$.

Chapter 3 What is an Integral?

3.1 Determining Distance from the Rate of Motion. The Area Bounded by a Curve

The problem of determining the instantaneous rate of motion, or velocity, $v(t)$ from a given dependence of the position of a body upon the time, $z = z(t)$, led us to the concept of the derivative:

$$v(t) = \frac{dz}{dt}.$$

The inverse problem consists in determining the position of a body, $z = z(t)$, that is, the distance covered by a body in a given interval of time t, when we know the instantaneous velocity as a function of the time, $v = v(t)$. This problem brings us to the second most important concept of higher mathematics, that of the ***integral***.

Let us agree on some convenient notation. We consider the distance traveled during time from t_1 to t_2. So as to avoid subscripts, let us denote the beginning of the time interval by a and the end of the interval by b, so that $t_1 = a$ and $t_2 = b$. We denote the distance covered from time a to time b by $z(a, b)$. Remember that when the two quantities, a and b, stand under the function symbol z in parentheses, then $z(a, b)$ is the distance covered during the interval of time from $t = a$ to $t = b$, whereas $z(t)$ with the single quantity t in parentheses is the position (coordinate) of the body at a specified time t. These quantities are related in a simple manner:

$$z(a, b) = z(b) - z(a), \text{ or}$$
$$z(b) = z(a) + z(a, b) \qquad (3.1.1)$$

(here we assume that $b > a$). We see that the distance covered during time from a to b is equal to the difference between the coordinate at the end of the time interval under consideration, $z(b)$, and that at the beginning of the interval, $z(a)$.[3.1]

Now let us compute $z(a, b)$. In the most elementary case, when the velocity is constant, or

$$v(t) = \text{constant} = v_0, \qquad (3.1.2)$$

the distance covered will obviously be equal simply to the product of the time of motion into the velocity:

$$z(a, b) = (b - a)\, v_0. \qquad (3.1.3)$$

Taking advantage of the graph of velocity versus time, we find that a constant velocity is associated with a horizontal straight line (Figure 3.1.1). The distance covered is clearly equal to the hatched area because the area of a rectangle is equal to the product of the base $(b - a)$ by the altitude (v_0). What do we do in the general case where the instantaneous velocity is not a constant quantity?

Let us make a detailed study of a numerical example. Suppose the velocity is given by the formula[3.2] $v = t^2$. We seek the distance covered during the time interval from $t = a = 1$ to $t = b = 2$.

We partition the entire interval from a to b into ten subintervals and set up the table of velocity:

t	1.0	1.1	1.2	1.3	1.4	1.5
v	1.0	1.21	1.44	1.69	1.96	2.25

t	1.6	1.7	1.8	1.9	2.0
v	2.56	2.89	3.24	3.61	4.0

Why is it difficult to compute the distance at a velocity $v(t)$ given by a

[3.1] For the sake of simplicity we consider only motion in one direction in such pictorial considerations, although possibly, the velocity varies in time.

[3.2] If the velocity v is expressed in units of cm/s and time t in units of s, then, in order to maintain the requirements of dimensionality, we write $v = kt^2$, where the factor k has the dimensions of cm/s^3. Here we assume that $k = 1$ cm/s^3.

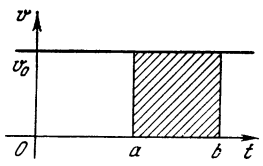


Figure 3.1.1

formula? Clearly because the velocity is variable (for a constant velocity, the answer is trivial). In the case at hand, the velocity changes four times over the time interval from $t = 1$ to $t = 2$. However, after this interval is partitioned into ten parts (subintervals), the velocity varies less over each subinterval of duration 0.1 second (only 10 to 20 percent). Therefore, in the subintervals the velocity can roughly be taken to be constant, and we can compute the distance covered during such a subinterval of time as the product of that subinterval by the velocity. In what follows, we denote the subintervals into which the interval from $t = a$ to $t = b$ is divided by Δt (the time increment); for each such subinterval Δt we will assume the distance to be simply proportional to Δt, with the velocity assumed constant over Δt.

To compute the distance covered during each subinterval Δt, equal to 0.1 s, we utilize the *initial* velocity in the given subinterval: 1 cm/s in Δt from 1 to 1.1 s, 1.21 cm/s in Δt from 1.1 to 1.2 s, and so on; finally, 3.61 cm/s in the last subinterval Δt extending from 1.9 to 2.0 s. The total distance covered during the time interval from $t = 1$ to $t = 2$ is then computed to be

$$z\,(1, 2) \simeq 0.1 + 0.121 + 0.144 + \ldots + 0.361 = 2.185 \text{ cm}. \qquad (3.1.4)$$

It is quite clear that we have *reduced* the actual distance covered because the velocity here increases with time and so the velocity at the beginning of each subinterval is less than the average velocity. Each of the ten terms into which the entire distance was partitioned is slightly less than the actual value, and so the result has a deficit, too.

Let us now compute the distance somewhat differently, namely, in each subinterval Δt we will take the value of velocity at the *end* of the subinterval. For the first subinterval Δt from 1 to 1.1 this velocity is equal to 1.21 cm/s, for the last one from 1.9 to 2.0 s it is 4 cm/s. We then get the following estimate for the entire distance traveled:

$$z\,(1, 2) \simeq 0.121 + 0.144 + \ldots + 0.400 = 2.485 \text{ cm}. \qquad (3.1.5)$$

This calculation clearly yields the distance $z\,(1, 2)$ with an *excess*.

Hence, the true value lies between 2.185 and 2.485 cm. The difference between these numbers amounts to about 15%. Rounding off the boundary values for z yields $2.18 < z\,(1,2) < 2.49$.

These computations can be illustrated by means of a graph. We construct the graph (Figure 3.1.2) with time laid off on the axis of abscissas and velocity on the axis of ordinates. In the figure we divide the time interval into five intervals instead of ten (as we did in the table), which makes each part (step) stand out more clearly. Each term in the first sum (3.1.4) is the area of a narrow rectangle with the corresponding subinterval Δt as the base and the velocity at the beginning of the subinterval as the altitude. Thus, the sum is the area under the polygonal (step-like) line (the hatched area in Figure 3.1.2*a*).

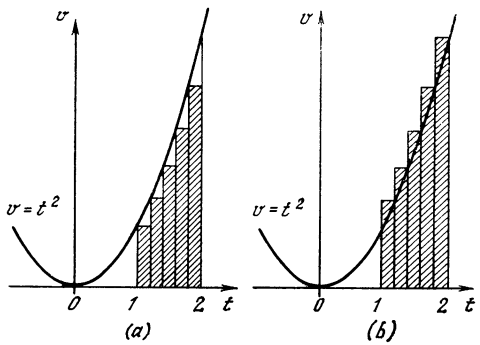


Figure 3.1.2

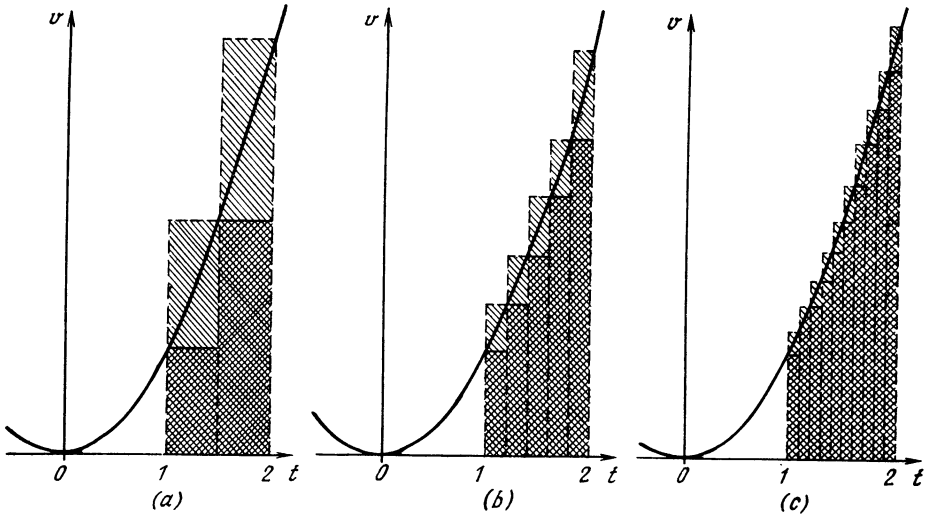


Figure 3.1.3

The second sum (3.1.5), in which the velocity in each subinterval is taken at the end of the subinterval, corresponds to the hatched area in Figure 3.1.2*b*.

How can we make a more accurate computation of the distance covered during a given time (from $t = a = 1$ s to $t = b = 2$ s in our example)? Clearly, the distance between the lower and upper estimates, that is, the difference between 2.18 and 2.49 in our case, depends on the variation of velocity within the limits of each subinterval Δt. For this difference to become smaller, we must simply see to it that the velocity varies less rapidly within each subinterval; but since the velocity is a quantity over which we have no power, we have to partition the time interval into smaller subintervals. This approach suits us perfectly since it brings us closer to our goal: if all the subintervals are made smaller (while the total number of the subintervals increases proportionately), the two estimates for the distance $z\ (a, b)$ obtained through the method discussed above move closer to each other.

For instance, if we split up the 1-to-2-s interval into 20 subintervals (instead of 10) of 0.05 s duration each, using the method discussed above and the *initial* velocities in each Δt, we compute the distance to be

$$z\ (1, 2) \simeq 0.05 + 0.05 \times 1.1025 + \ldots + 0.05 \times 3.8025 = 2.25875. \tag{3.1.4a}$$

Using the *terminal* velocities in each subinterval, we get the distance

$$z\ (1,2) \simeq 0.05 \times 1.1025 + 0.05 \times 1.21 + \ldots + 0.05 \times 4 = 2.40875. \tag{3.1.5a}$$

The difference between 2.25875 and 2.40875 now amounts to about 7%. The range within which $z\ (1, 2)$ lies has narrowed down. Rounding off these figures, we get $2.26 < z\ (1, 2) < 2.41$.

As we reduce the subintervals Δt, the result approaches the *true* value of the distance covered. It will be computed later on and proves to be equal to $z\ (1, 2) = 2\frac{1}{3} \simeq 2.333$. As we reduce Δt, the difference between the initial and terminal velocities in each subinterval Δt decreases, and so also does the relative error in each summand. It is because we are dealing here with the *relative* error we can be sure that although the number of terms (that is, the num-

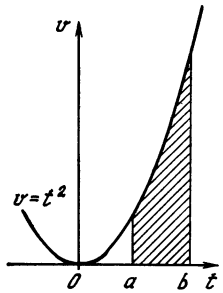


Figure 3.1.4

ber of errors) increases, nevertheless the precision with which we estimate the entire sum of the distances, that is, the precision with which we estimate z (1, 2) is sure to increase.

To illustrate the above reasoning, in Figure 3.1.3 we depict the constructions connected with the estimate of z (1, 2) from the given velocity $v = t^2$ at $\Delta t = 0.5$, 0.2, and 0.1 s. Geometrically, it is obvious that as we increase the number of subintervals Δt and reduce the duration of each one, the dimensions of each step in Figure 3.1.3 become smaller, whereby the step-line comes closer and closer to the curve v versus t.

We thus conclude that the distance covered during time from $t = a$ to $t = b$, given an arbitrary dependence of the instantaneous velocity on time, $v = v(t)$, is equal to the *area* bounded by the curve $v = v(t)$, the vertical lines $t = a$ and $t = b$, and the t axis (Figure 3.1.4).

This conclusion yields a method for practical computation of the distance: we can construct the graph on graph paper and determine the hatched area either by counting the squares, or, for instance, by cutting out the area of paper, weighing the sheet, and comparing its weight with the weight of a rectangular or square piece of the same paper of known area.[3.3]

[3.3] A pocket calculator, however, makes the method of constructing sums of the type (3.1.4) and (3.1.5) (or (3.1.4a) and (3.1.5a)) much easier to carry out than in the case of primitive weighing of paper or cardboard figures.

Such methods are convenient and quite justified when the velocity is not known exactly, that is, is given in the form of a table or graph obtained empirically (in an experiment) rather than by a formula. We will not dwell on these approximate methods and will attempt to express the distance by a *formula* when the velocity is given by a *formula*. Note also that in some cases when, say, the velocity is expressed by the formula $v(t) = (1 - t)^{-1}$, with t varying from 0 to 1 s, and increases without limit as $t \to 1$ s, the concept of distance becomes meaningless (the distance tends to infinity). For the time being we will not consider such cases (but see Section 5.2).

We can also make more precise the numerical method used above to determine the distance from the velocity. We will assume that in each subinterval Δt the velocity is constant and equal to the ***arithmetic mean*** (half-sum) of the initial and terminal velocities in the given subinterval. In this approach, with a partition into ten subintervals, the velocity in the first subinterval from 1 to 1.1 s is taken equal to $(1 + 1.21)/2 = 1.105$ cm/s (see the table at the beginning of this section) and the distance covered during this subinterval of time is 0.1105 cm, the distance covered during the second subinterval is $0.1\ (1.21 + 1.44)/2 = 0.1325$ cm, and so on. Adding all the distances obtained in this manner, we get the distance covered during the time interval from $a = 1$ s to $b = 2$ s: z (1, 2) = 0.1105 + 0.1325 + ... = 2.335 cm. If we divide the interval into 20 subintervals, we get (using the same half-sum of velocities) z (1, 2) = 2.33375 cm. These values are much closer to the true value 2.33333 cm than those computed on the basis of the initial and terminal values of velocity for the same number of subintervals: for ten subintervals, the error is equal to 0.07% instead of 15% in the earlier method, and for 20 subintervals the error is only 0.02% instead of 7%.

The new method for finding the distance can also be displayed vividly on a graph. The product of the half-sum of the velocities at the beginning and end of an interval by the magnitude of the interval of time is equal to the trapezoid $ABCD$ (Figure 3.1.5). With bases AB and DC and altitude AD, the area is

$$\frac{AB+DC}{2}\,AD=\frac{v(t_1)+v(t_2)}{2}\,(t_2-t_1).$$

For this reason, determining the distance on the basis of the half-sum of the velocities is known as the ***trapezoid method.*** For the shape of the curve shown in Figure 3.1.5, the area of the trapezoid is somewhat greater than the area bounded by the straight lines BA, AD, DC and the portion BC of the curve $v = v(t)$. The difference between the area of the trapezoid and the area bounded by the arc of the curve is equal to the area of the crescent-like

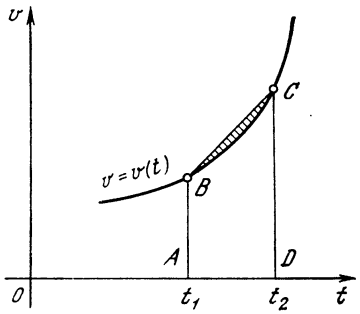


Figure 3.1.5

figure formed by the chord BC and the portion BC of the curve shown hatched in Figure 3.1.5). The sum of all such areas yields the error, that is, the difference between the true value of the distance and the value computed by the trapezoid method. A comparison with Figures 3.1.2 and 3.1.3 shows vividly that the error in the trapezoid method should be less than that when formulas (3.1.4) and (3.1.5) are employed (the step method or, as mathematicians call it, the ***rectangular method.***

Of course, when one compares the distance with the area under the graph representing the v versus t dependence, it is important to take into account the scale used (compare to what was said in this connection in Section 1.7). Suppose that 1 cm along the axis of abscissas on the graph corresponds to a time interval of T seconds and 1 cm on the axis of ordinates corresponds to a velocity of V cm/s. Then, if the motion is at a constant velocity v_0 during a time from $t = a$ to $t = b$, the distance covered is equal to $v_0 (b - a)$, and the area of the (hatched) rectangle on the graph (Figure 3.1.1) is equal to $S = (v_0/V)\ ((b - a)/T)$ cm^2.

Thus, here we have $z(a, b) = SVT$.

This relationship between the distance covered and the area on the graph of velocity bounded by the curve $v(t)$, the axis of abscissas, and two vertical lines is preserved in the case of a *variable* velocity v, that is, in the case of an arbitrary function $v = v(t)$.

3.2 The Definite Integral

In the preceding section, two problems—finding the distance traversed by a body and the equivalent problem of finding the area under a curve—led to a consideration of sums of a special type with a large number of small terms (summands). A rigorous approach to such problems leads to the concept of the integral.

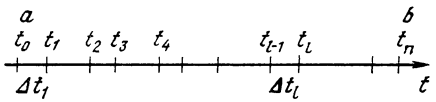


Figure 3.2.1

The distance $z = z(a, b)$ found from a given velocity $v = v(t)$ is called the ***definite integral*** of the function $v(t)$ (velocity) with respect to the variable t (time) taken from a to b (the same quantity is sometimes called the ***integral of function*** $v(t)$).

We now give a mathematical definition of the integral that corresponds to the ideas illustrated by the numerical example of Section 3.1. The definition will remain valid when we consider physical or mathematical quantities of a nature different from velocity and distance.

Suppose that we have a function $v = v(t)$. To find its integral from a to b we partition the interval into a large number of subintervals, n. We denote the values of the independent variable t at the endpoints of the subintervals by $t_0, t_1, t_2, \ldots, t_{n-1}$, where, obviously, $t_0 = a$ and $t_n = b$ (Figure 3.2.1). The lengths Δt of the small subintervals of time are equal to the difference between adjacent values of t.[3.4] Thus, for an arbitrary l (with $l = 1, 2, \ldots, n$) $\Delta t_l = t_l - t_{l-1}$.

The subscripts on the quantities t and Δt are merely number labels, or indices, as they are sometimes called (see footnote 1.1).

[3.4] If the interval from a to b is specially partitioned into n equal parts, then each subinterval Δt equals $(b-a)/n$. It is not obligatory, however, that the subintervals be equal; the only thing we require is that *each* subinterval be small. The reader will see the truth of this if he or she thinks through the distance-velocity example of Section 3.1 see also Exercise 3.2.3.

The approximate value of the integral $z(a, b)$ is given by the formula

$$z(a, b) \simeq \sum_{l=1}^{l=n} v(t_{l-1})\,\Delta t_l. \qquad (3.2.1)$$

$\sum_{l=1}^{l=n}$ signifies that the expression standing to the right of the summation symbol[3.5] be taken for all values of l from 1 to n and then all these expressions are to be added together. For example, if $n = 10$, then $\sum_{l=1}^{l=10} v(t_{l-1})\,\Delta t_l = v(t_0)\,\Delta t_1 + v(t_1)\,\Delta t_2 + \ldots + v(t_9)\Delta t_{10}$. In the example of Section 3.1 (see the table in that section), $t_0 = 1$, $t_1 = 1.1$, $t_2 = 1.2$, . . ., and $z(1, 2) \simeq \sum_{l=1}^{l=10} t_{l-1}^2 \Delta t_l = 2.185$.

In the approximate expression (3.2.1), the value of the function $v(t)$ in each subinterval was taken at the *beginning* of the subinterval, at the point t_{l-1}. A different approximate expression is obtained if we take the value of the function at the *endpoint* of each subinterval:

$$z(a, b) \simeq \sum_{l=1}^{l=n} v(t_l)\,\Delta t_l. \qquad (3.2.2)$$

In the example of Section 3.1, this sum for $n = 10$ was equal to 2.485.

The ***definite integral*** of a function $v(t)$ taken from a to b is the *limit* approached by the sums (3.2.1) and (3.2.2) as all subintervals Δt_l tend to zero. The integral is written

$$z(a, b) = \int_a^b v(t)\,dt \qquad (3.2.3)$$

(read: $z(a, b)$ equals the definite integral of $v(t)$ from a to b, dt). The integral sign $\int$ is merely an elongated S (the first letter of the word *summa*, which is the Latin for sum).

Unlike Δt, the symbol dt signifies that in order to obtain the exact value of the integral it is necessary to pass to the limit as all subintervals Δt tend to zero (just as the derivative dz/dt is obtained from the ratio $\Delta z/\Delta t$ if we send Δz and Δt to zero and pass to the limit). The formulas (3.2.1) and (3.2.2), in which the Δt are small but finite, only yield *approximate* values of the integral, just as the ratio $\Delta z/\Delta t$ only yields an approximate value for dz/dt for finite Δz and Δt.

When the subintervals Δt become smaller and smaller, it is immaterial whether we take the value of the function $v(t)$ at the beginning, at the end, or in the middle of the subintervals, which is to say that it is immaterial whether we proceed from (3.2.1) or from (3.2.2) or from some other expression for the "integral sum" similar to the one in the right-hand sides of (3.2.1) and (3.2.2) (say, we could have added the terms $v(t_m)\,\Delta t$, where t_m is the middle of the corresponding subinterval Δt). For this reason formula (3.2.3) simply has $v(t)$, which is a value of the function in the subinterval Δt without any indication of the value of t being taken inside (or at the beginning or end) of the subinterval.

Another way in which the integral (3.2.3) differs from the sums (3.2.1) and (3.2.2), which yield approximate values of the integral, lies in the fact that as the Δt becomes smaller and smaller and the number of subintervals increases, we no longer label them. For this reason, we indicate on the integral only the limits of variation of t (range of t) from a to b. The quantity a is placed at the bottom of the integral sign and is termed the ***lower limit*** of integration, while b is placed at the top and is called the ***upper limit*** of integration.[3.6]

[3.5] The symbol $\sum$ is the capital Greek letter sigma. It corresponds to S in the Latin alphabet, the first letter of the term sum.

[3.6] In this section the word "limit" is used in two meanings. First, the integral is the limit of a sum in the same sense that the derivative is the limit of a ratio. Here, limit corresponds to the sign lim. Second, we speak of the limits of variation of t from a to b, the ***limits of integration*** a and b. The meaning

The range of t from a to b is called the ***domain of integration.*** The function $v(t)$ in the expression of the integral is called the ***integrand***, t being the ***variable of integration.***

Thus, the integral is defined as the limit approached by the sum of products of the values of the function multiplied by the difference of the values of the independent variable when all differences of the independent variable tend to zero:[3.7]

$$\lim_{\Delta t_l \to 0} \sum_{l=1}^{l=n} v(t_l)\,\Delta t_l$$

$$= \lim_{\Delta t_l \to 0} \sum_{l=1}^{l=n} v(t_{l-1})\,\Delta t_l = \int_a^b v(t)\,dt. \quad (3.2.4)$$

Although the first and second sums in (3.2.4) are different for a finite number l of small subintervals Δt_l, their limits coincide when all subintervals Δt decrease without limit (tend to zero). These limits yield the value of the integral.

As Δt tends to zero, each separate summand tends to zero, but on the other hand the number of terms in the sum increases and approaches infinity. The sum itself tends to a very definite limit, which is the solution of the problem and is termed the *integral.* If the function is the instantaneous velocity, this limit, the integral of the function, is equal to the distance covered. If the integrand represents the ordinates of the points of the graph $y = f(x)$, the integral $\int_a^b f(x)\,dx$ is equal to the area bounded by our graph, the x axis, and the vertical lines $x = a$ and $x = b$.

Naturally, not just any sum of a large number n of small terms tends to a definite limit as $n \to \infty$. Say, the sum of n terms each of which is $1/\sqrt{n}$ is equal to $\sqrt{n}$ and tends to infinity as $n \to \infty$, which means that this sum has no finite value. We shall try to explain why in our case the limit must exist.

Let us partition the interval from a to b into n equal subintervals, the length of each being $(b - a)/n$. If for the sake of simplicity we take the velocity v as being constant, we get a sum of n terms each being equal to $v\Delta t = v(b - a)/n$. The total sum (that is, the distance covered) is equal to $nv\Delta t = nv(b - a)/n = v(b - a)$, which means that it is independent of n. What is important here is that each separate term diminishes in exactly the same proportion (in proportion to $1/n$) as does the number of terms n increase. It is also clear that for the case of a variable velocity and the partition of the interval from a to b into n equal small segments the same is true, that is, the decrease in the length of each subinterval is inversely proportional to the growth in the number of such subintervals. For instance, when each small subinterval Δt is divided into two halves, $\Delta_1 t$ and $\Delta_2 t$, the terms in the integral sum corresponding to these halves prove to be approximately half the initial, "large", summand, but the total number of terms doubles, so that the order of magnitude of the entire sum does not change.[3.8] The reader that is not convinced by all this reasoning should do the exercises at the end of this section.

This explanation is useful if we are dealing with the mathematical definition of the integral as the limit of a certain sum. In physical problems, how-

here is different. The attentive reader will readily see which of the two meanings is used in any given case.

[3.7] To be more precise, we should have written

$$\lim_{\Delta t_1,\ \Delta t_2,\ \ldots,\ \Delta t_n \to 0} \sum_{l=1}^{l=n} v(t_l)\,\Delta t_l \quad \text{(and}$$

similarly for the second sum), but such notation is too unwieldy.

[3.8] Note that in the example with n summands each of which is equal to $1/\sqrt{n}$, a 2-fold increase in the number of terms is accompanied by a decrease in the magnitude of each term only by a factor of $\sqrt{2}$. In the example of the sum of terms each of which is $1/n^2$, a 2-fold increase in the number of terms leads to a 4-fold decrease in the magnitude of each term, so that the new sum is only half the old one.

ever, the existence of the integral, that is, the limit of the "integral sums" considered, is as a rule quite obvious. For instance, there is no doubt that a body traveling with a finite (constant or variable) velocity will, over a finite time interval, traverse a certain (finite) path, that is, it will travel a certain distance. In what follows we will explain how by employing the concept of the integral one can calculate the area of curvilinear figures (this question was discussed in brief above). Here too there can be no doubt that the problem has a solution, that is, a figure must have an area, and so the appropriate integral exists.

Since the variable of integration can assume the values a and b, it is clear that the limits of integration have dimensions and their dimensions are those of the variable of integration (in the distance-velocity example, the limits of integration have the dimensions of *time*). (As for the dimensions of the integral, see the end of this section.)

Clearly, the value of a definite integral depends on the values of the function under the integral sign solely *within* the domain of integration. Values of the function outside the domain of integration have no effect on the magnitude of the integral and are of no interest to us. For instance, this distance covered, $z\ (a, b)$, depends of course only on the value of the velocity $v = v\ (t)$ inside the domain of integration and does not depend on the velocity before $t = a$ or after $t = b$.

It was pointed out in Section 3.1 that the distance can be determined by computing the area on the graph in which the velocity is a function of time. The problem of finding the *area* S bounded above by a curve with a specified equation $y = y\ (x)$, below by the axis of abscissas (the x axis), and on the sides by the vertical lines $x = a$ and $x = b$ (Figure 3.2.2), that is, of calculating the area of the *curvilinear trapezoid* $ABCD$, whose bases are the parallel line segments AC and BD of the straight lines $x = a$ and $x = b$ and

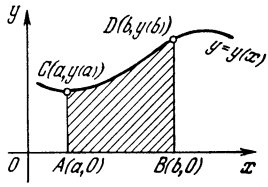


Figure 3.2.2

whose lateral sides are the line segment AB of the axis of abscissas and the portion CD of the curve $y = y(x)$ reduces to computing the integral

$$S = \int_a^b y\ (x)\ dx \qquad (3.2.5)$$

To explain this, recall Figure 3.1.2. Imagine that the values of an (arbitrary) independent variable x are laid off on the axis of abscissas, the values of a function $y\ (x)$ on the axis of ordinates, and both x and y are assumed to be the distances of a variable point $M(x, y)$ from the axes of coordinates (see Section 1.2) and in no way are related to the physical concepts of time and distance. The sum of the areas of the hatched rectangles in Figure 3.1.2*a* is equal to $\sum_{l=1}^{l=n} y\ (x_{l-1})\ \Delta x_l$; the same sum in Figure 3.1.2*b* is equal to $\sum_{l=1}^{l=n} y\ (x_l)\ \Delta x_l$. In the limit, as $\Delta x_l \to 0$, these sums are, by definition, equal to the integral, and the sum of the areas of the rectangles tends to the area bounded by the curve $y\ (x)$, since the smaller the subintervals Δx_l the closer to the curve is the polygonal (step-like) line bounding the rectangles.

In conclusion we note that the definite integral depends on the integrand and the limits of integration, but it is independent of the designation of the variable of integration. Judge for yourself. Suppose that we have the integrand in the form $v\ (t) = 3t^2 + 5$. Substituting x for t, we get $v\ (x) = 3x^2 + 5$. When we compute the integral, it is immaterial how the variable

of integration is designated, the only important thing being over what range it varies and what are the values of the function. For this reason,

$$z(a, b) = \int_a^b v(t)\,dt = \int_a^b v(x)\,dx$$
$$= \int_a^b v(u)\,du = \int_a^b v(\lambda)\,d\lambda,$$

or even

$$z(a, b) = \int_a^b v(o)\,do$$

(the last form is never used, of course). Any letter will do for the variable of integration—the result will be the same.

A variable which (like the variable of integration) does not appear in the final result is called a ***dummy variable.*** The variable of integration under the integral sign can be changed to any letter without disrupting the validity of the formulas. An ordinary (not dummy) variable can be replaced by a different letter only in all parts of a formula; for instance, in the formula $(x + 1)^2 = x^2 + 2x + 1$ one cannot write $(x + 1)^2 = t^2 + 2t + 1$, but in integrals we can write $z(a, b) = \int_a^b v(r)\,dt$ or $z(a, b) = \int_a^b v(x)\,dx$.

Finally, let us turn to the question of *dimensions.* By definition,

$$\int_a^b y(x)\,dx = \lim_{\Delta x_l \to 0} \sum_{l=1}^{l=n} y(x_l)\,\Delta x_l. \quad (3.2.6)$$

If x is measured in units of e_1 and y in units of e_2, then each term in the sum on the right-hand side of (3.2.6) has the dimensions of e_1e_2 (the factor Δx_l has the dimensions of e_1, and $y(x_l)$ has the dimensions of e_2), whereby the limit of the sum (the integral) has the same dimensions. For instance, if the velocity $v(t)$ is measured in units of cm/s and t in units of s, then the distance $z(a, b) = \int_a^b v(t)\,dt$ is measured in units of (cm/s)·s = cm. If the abscissa x and the ordinate y are measured in centimeters, the area $S = \int_a^b y(x)\,dx$ is measured in square centimeters (cm·cm = cm^2). A transfer to new units of measurement, say, to e_1' in x and to e_2' in y, results in the multiplication of integral (3.2.5) by a factor of k_1k_2, where $e_1 = k_1e_1'$ and $e_2 = k_2e_2'$. The integral increases in the process by a factor of k_1k_2, but it expresses the same quantity as before (only in new units, $e_1'e_2'$). For intance, in going over from centimeters o millimeters, we increase the numerical value of the area $S = \int_a^b y(x)\,dx$ by a factor of $10 \times 10 = 100$, while in going over from cm/s to m/min we are forced to multiply the value of the distance (which before was measured in centimeters) $z(a, b) = \int_a^b v(t)\,dt$ by a factor of $(0.01 \div 1/60) \times 1/60 = 0.01$ (the new value, of course, will be expressed in meters).

Exercises

3.2.1. Consider the case $v = kt + s$ (uniformly accelerated motion). Find the distance covered during the time interval from a to b by dividing this interval into m equal subintervals; take advantage of the fact that the terms of the sum form an arithmetic progression. Find the limit of the sum as $m \to \infty$. Compare the resulting expression with the area, equal to the distance covered, of a trapezoid in the vt-plane.

3.2.2. Consider the case $v = t^2$ and for this case find the distance covered in the time interval from $t = 1$ to $t = 2$, in other words, find the integral $\int_1^2 t^2\,dt$. To do this, partition the interval from 1 to 2 into m equal parts and compute the sum $\sum t_{l-1}^2 \frac{1}{m}$ or the sum

$\sum t_i^2 \frac{1}{m}$. Compare these two sums. [*Hint.* To calculate these sums, employ the formula $1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$.]

3.2.3. Evaluate the integral $\int_1^2 x^3\,dx$. To this end divide the interval from 1 to 2 into m small subintervals Δx whose lengths form a *geometric progression*, and pass to the limit as $m \to \infty$.

3.3 The Relationship Between the Integral and the Derivative

In the preceding sections we considered, separately, the concepts of the derivative and the integral. In the present section we examine (using the distance-velocity example) the relationship between these two concepts. This step will unite the calculus of derivatives and the calculus of integrals into a single science that abounds with applications, the ***differential and integral calculus***, or, as it is sometimes called, ***mathematical analysis.***

We assume as given and known the instantaneous velocity as a function of time $v = v(t)$. We regard as constant the time $t = a$ at the beginning of the motion (a train leaves a certain station at a known time $t = a$), and consider the distance covered in the time interval from a to b as a *function* of the terminal instant b. To emphasize this, we will denote the terminal (undefined) moment not by b but say, by the letter u, since it is customary (but not mandatory) to denote variable quantities by the last letters of the English alphabet. The result is

$$z(a, u) = \int_a^u v(t)\,dt. \tag{3.3.1}$$

We know that

$$z(a, u) = z(u) - z(a) \tag{3.3.2}$$

(in the example with the train it is natural to assume that $z(a)$ is zero). Let us take the derivative of the left and right members, regarding u as the (independent) variable and a as a constant, with the result that $-z(a)$ on the right-hand side is also a constant and $d(-z(a))/du = 0$. This yields $dz(a, u)/du = dz(u)/du$. But we know that the derivative of the coordinate of a body with respect to time is nothing but the instantaneous velocity of that body, so that $dz(u)/du = v(u)$ is the velocity at time u and, hence

$$\frac{dz(a, u)}{du} = v(u). \tag{3.3.3}$$

Substituting here the expression (3.3.1) for distance $z(a, u)$ in the form of an integral, we get

$$\frac{d}{du}\left(\int_a^u v(t)\,dt\right) = v(u). \tag{3.3.4}$$

This equation is the most important general property of the definite integral. In the given form, this equation is a general mathematical theorem. Its validity is independent of whether $v(t)$ is velocity (and the integral is distance) or $v(t)$ is some other quite different quantity. For any function, say $y(x)$, we have

$$\frac{d}{db}\left(\int_a^b y(x)\,dx\right) = y(b), \tag{3.3.4a}$$

where we have replaced the upper (variable) limit of integration with b.

The theorem is stated thus: *the derivative of a definite integral with respect to the upper limit is equal to the value of the integrand at the upper limit.*

Because the theorem is so important (it is called the ***Newton-Leibniz theorem***[3.9]), we give a different derivation of it based on a consideration of area. We will compute the derivative by first principles, that is, as the limit of the ratio of the increment of the

[3.9] In some old textbooks on differential and integral calculus this theorem is called the ***fundamental theorem of higher mathematics.*** This name sounds rather high-flown, but it is quite justified.

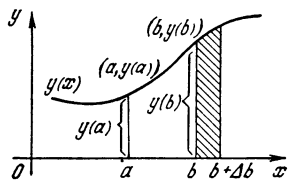


Figure 3.3.1

function to the increment of the independent variable. We consider

$$I(a, b) = \int_a^b y(x)\,dx.$$

This integral is the area bounded from above by the curve $y(x)$, from below by the x axis, on the left by the vertical line $x = a$, and on the right by the vertical line $x = b$ (see Figure 3.2.2).

How do we find the increment ΔI of the integral I caused by a small increment Δb of the upper limit b? By the definition of an increment, $\Delta I = I(a, b + \Delta b) - I(a, b)$. The area equal to the integral $I(a, b + \Delta b)$ differs from the area equal to the integral $I(a, b)$ in that the right vertical is displaced rightward by a small Δb (compare Figure 3.3.1 with Figure 3.2.2). Consequently, the increment ΔI is the difference between two areas: that with the base from a to $b + \Delta b$ and that with the base from a to b. ΔI is clearly the area of the thin strip that is hatched in Figure 3.3.1. The base of this strip on the x axis is a line segment of length Δb.

The desired derivative is equal to the limit

$$\frac{dI(a, b)}{db} = \lim_{\Delta b \to 0} \frac{\Delta I}{\Delta b}.$$

Quite obviously, as Δb tends to zero, the area of the strip approaches $y(b)\,\Delta b$, and the ratio $\Delta I/\Delta b$ approaches the quantity $y(b)$. We have thus once again given a pictorial proof of the Newton-Leibniz theorem:

$$\frac{d}{db}\left(\int_a^b y(x)\,dx\right) = y(b).$$

The definite integral of a known function $y(x)$ or $v(t)$ is a function of the limits of integration a and b, that is, a function of two variables, a and b. Formula (3.3.4) or (3.3.4a) yields the value of the derivative of this function at the upper limit of integration, the variable b. Below, in Section 3.5, we will find the value of its derivative at the lower limit of integration, a.[3.10]

The definition of an integral as the limit of a sum, which was given in the preceding section, explains the role the integral concept plays in the solution of physical or geometric problems: in computing the distance traveled when a body is moving with a (variable) velocity that is known, in determining the area bounded by a curvilinear trapezoid, and in other problems that lead to the construction of "integral sums" (below we will encounter many problems of this type). But this definition does not yield a convenient general method for computing an integral or finding the value of the integral in the form of a formula, as a function of the limits of integration.[3.11]

A method for finding such a formula follows from the Newton-Leibniz theorem proved above, a theorem that concerns the derivative of an integral. Here, besides the property of the derivative of an integral, we make use of yet a second property of the definite integral: *the definite integral is equal to zero when the upper and lower limits of integration coincide*:

$$z(a, a) = \int_a^a v(t)\,dt = 0. \qquad (3.3.5)$$

[3.10] Here, of course, we are speaking of the *partial* derivatives of the function $I(a, b) = \int_a^b y(x)\,dx$ of *two* variables, a and b (see Section 4.13).

[3.11] Only in rare cases and with great difficulty is one able to sum an arbitrary number of small summands present in the definition of the integral (e.g. see Exercises 3.2.2 and 3.2.3).

This property is obvious because the distance is equal to zero if the time in transit is zero.

The formula which yields the value of the integral as a function of the limits of integration will be derived in this fashion in Section 3.4.

Let us now study the question of the integral of a derivative. In the integral

$$\int_a^b v(t)\,dt \tag{3.3.6}$$

the integrand may be an arbitrary function, and of course there is no universal method that enables finding the integral for an arbitrary function $v(t)$. But suppose we are lucky and have found a function F such that the integrand $v(t)$ is the derivative of this function F:

$$\frac{dF(t)}{dt} = v(t). \tag{3.3.7}$$

In such a favorable case everything turns out wonderful: here we can easily find the exact value of the integral. To write it out, we recall the approximate expression for the increment of function (cf. (2.4.6)):

$$\Delta F \simeq F'(t)\,\Delta t = v(t)\,\Delta t,$$

or in the more common notation, which refers to the integral $\int_a^b f(x)\,dx$, with

$$f(x) = dF(x)/dx,$$

$$\Delta F \simeq F'(x)\,\Delta x = f(x)\,\Delta x. \tag{3.3.8}$$

The quantity in the right member of the equation is precisely one of the summands whose sum is equal to the integral. And so, by putting, say, $\Delta x_l = x_l - x_{l-1}$ and, hence, $\Delta F = F(x_l) - F(x_{l-1})$, we can write, approximately,

$$\Delta F = F(x_l) - F(x_{l-1}) \simeq f(x_l)\,(x_l - x_{l-1}) = f(x_l)\,\Delta x_l. \tag{3.3.9}$$

As already mentioned, (3.3.9) is approximate and its accuracy increases as the difference between x_l and x_{l-1},

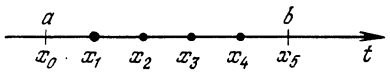


Figure 3.3.2

that is, the increment Δx_l becomes smaller. But as the difference $x_l - x_{l-1} = \Delta x_l$ decreases, that is, as x_l approaches x_{l-1}, the difference between $f(x_l)$ and $f(x_{l-1})$ also diminishes. For this reason, we have just as much right (the degree of accuracy is the same) to put $f(x_{l-1})$ in the right-hand side of (3.3.9) as we do $f(x_l)$, as was done above, where we selected the latter case.

Let us now write formulas like (3.3.9) for *all* subintervals into which the domain of integration, that is, the interval from a to b, is partitioned. Suppose that the interval is partitioned into five subintervals (Figure 3.3.2) so that $x_0 = a$ and $x_5 = b$. Then

$$\begin{aligned}
&F(x_1) - F(x_0) \simeq f(x_1)\,(x_1 - x_0),\\
&F(x_2) - F(x_1) \simeq f(x_2)\,(x_2 - x_1),\\
&F(x_3) - F(x_2) \simeq f(x_3)\,(x_3 - x_2),\\
&F(x_4) - F(x_3) \simeq f(x_4)\,(x_4 - x_3),\\
&F(x_5) - F(x_4) \simeq f(x_5)\,(x_5 - x_4).
\end{aligned} \tag{3.3.10}$$

Now add them together. In the left members, all values of the function F for intermediate values of x cancel out leaving only

$$F(x_5) - F(x_0) = F(b) - F(a).$$

On the right-hand side we have precisely those sums with the aid of which we approximately expressed the integral in Section 3.1 (expressing the distance $z(a, b)$ for a given velocity $v(t)$). Thus,

$$F(b) - F(a) \simeq \sum_{l=1}^{l=5} f(x_l)\,(x_l - x_{l-1})$$

$$= \sum f(x_l)\,\Delta x_l \simeq \int_a^b f(x)\,dx,$$

where $f(x) = dF/dx$.

The smaller each increment Δx, or the difference $x_l - x_{l-1}$, the more exact the expression (3.3.9) for the increment ΔF. But as all the differences $\Delta x_l = x_l - x_{l-1}$ decrease (and the number of subintervals Δx_l grows without limit), the sums tend to the integral. Therefore, the equation

$$F(b) - F(a) = \int_a^b f(x)\,dx \quad \text{for } f(x) = \frac{dF}{dx} \tag{3.3.11}$$

is now exact.

The last statement seems to be quite convincing, but it can be justified by even a more exact computation. The thing is that the error introduced by a formula similar to (3.3.8) or (3.3.9) is, generally speaking, *quadratic*: it is approximately proportional to the second power of the length Δx of the interval, that is, has the form $c\,(\Delta x)^2$, where the quantity $c = c(x)$ depends on x but weakly and with a high degree of accuracy can be assumed to be a constant (all this will be explained in Chapter 6). Thus, for $\Delta x = (b-a)/n$, each expression in (3.3.10) (the number of such expressions must be taken to be n instead of five) introduces an error proportional to $(\Delta x)^2 = (1/n^2)(b-a)^2$. The total error obtained through summation of all such approximate equations will be proportional to $n[(1/n^2)(b-a)^2] = (1/n)(b-a)^2$, from which it follows that it tends to zero as $n \to \infty$.

Formula (3.3.11) establishes a relationship between the integral and the derivative. From this formula it follows that if we are able to find a function F whose derivative is equal to the integrand f, the problem of evaluating the integral is solved, for all that remains is to substitute a and b for the independent variable of this function and find the difference $F(b) - F(a)$.

Since formula (3.3.11) is so important, in the sections that follow we will give a different derivation of it on the basis of a more detailed consideration of the properties of the integral and the function F.

3.4 The Indefinite Integral

In the preceding sections we introduced the concept of a definite integral as the limit of a sum of a large number of small terms. In Section 3.3 we elucidated the principal property of the definite integral: the derivative of a definite integral with respect to the upper limit is equal to the integrand, that is,

$$\text{if } z(a, b) = \int_a^b v(t)\,dt, \text{ then } \frac{dz(a, b)}{db} = v(b). \tag{3.4.1}$$

We now wish to take advantage of this property to compute a definite integral.

We seek a function of b whose derivative is the known function $v(b)$. We denote this function by $F(b)$. Then, by definition,

$$\frac{dF(b)}{db} = v(b). \tag{3.4.2}$$

This equation does not fully define the function $F(b)$. We know that the addition of any constant to a function does not change the derivative of that function. Hence, if $F(b)$ satisfies (3.4.2), so will the function $G(b) = F(b) + C$ for any constant C.

The function $F(b)$ which satisfies Eq. (3.4.2) is called the *indefinite integral* of the function $v(b)$. Similarly, if

$$\frac{dF(x)}{dx} = f(x), \tag{3.4.2a}$$

it is said that $F(x)$ is the ***indefinite integral*** of the function $f(x)$. This term reflects two properties of the function $F(x)$. First, the function F has the same derivative with respect to b as the (definite) integral $z(a, b)$, or $\int_a^b f(x)\,dx$ (cf. (3.4.1), (3.4.2), and (3.4.2a)); for this reason F is also called an integral. On the other hand, condition (3.4.2) or (3.4.2a) does not define this function completely, since we can always add a constant quantity C

to it, whence the modifying adjective "indefinite."

Any solution to Eq. (3.4.2) or (3.4.2a) can differ from a solution $F(b)$ (or $F(x)$) solely by a *constant*. Indeed, if there is another solution to (3.4.2), $G(b)$, then for their difference $F - G$ we have

$$\frac{d}{db}[F(b) - G(b)] = v(b) - v(b) = 0.$$

But only the derivative of a constant is equal to zero for arbitrary values of the argument, that is, only a body at rest has a zero velocity all the time.

According to (3.4.1), the definite integral $z(a, b)$ is also one of the solution to (3.4.2). Hence, $z(a, b)$ can likewise be represented as

$$z(a, b) = F(b) + C, \tag{3.4.3}$$

where $F(b)$ is a solution to (3.4.2) and C is a constant, and it only remains to determine the constant. To do this, we take advantage of the second property of the definite integral: it is equal to zero when the upper limit coincides with the lower limit,

$$z(a, a) = 0 \tag{3.4.4}$$

(see (3.3.5)). Substituting a for b in (3.4.3) and using (3.4.4), we get $0 = F(a) + C$, or $C = -F(a)$. From this we finally have

$$z(a, b) = F(b) - F(a), \tag{3.4.5}$$

or

$$\int_a^b f(x)\,dx = F(b) - F(a), \tag{3.4.5a}$$

where $F'(x) = f(x)$.

It is to be noted that the "indeterminacy" of the function $F(b)$ does not in any way hamper computation, with its aid, of a definite integral from formula (3.4.5) or (3.4.5a). Indeed, in place of $F(b)$ take some other solution to Eq. (3.4.2), say $G(b)$, which differs from $F(b)$ by a constant: $G(b) = F(b) + C$. We will evaluate the definite integral using (3.4.5), taking G instead of F:

$$\begin{aligned} z(a, b) &= G(b) - G(a) \\ &= [F(b) + C] - [F(a) + C] \\ &= F(b) - F(a). \end{aligned}$$

The result is the same as above.

In the distance-time problem it is convenient to denote the indefinite integral by the same letter z as we used for the definite integral. For a given integrand $v(t)$, the definite integral depends on the upper and lower limits of integration, which is to say, it is a function of two variables: $z = z(a, b)$. The indefinite integral is a function of one variable, which we denote by t. Thus, the indefinite integral $z(t)$ is a function that satisfies the equation

$$z'(t) = \frac{dz(t)}{dt} = v(t). \tag{3.4.6}$$

Using this function, we can find the definite integral $z(a, b)$ of the function $v(t)$ from the formula

$$z(a, b) = \int_a^b v(t)\,dt = z(b) - z(a). \tag{3.4.7}$$

The following compact notation is used to state the difference between the values of one and the same function for two different values of the variable:

$$z(t)\Big|_a^b = z(b) - z(a). \tag{3.4.8}$$

Here, on the left is the function of a dummy variable t, which is followed by a vertical line at the top of which is the value of the variable at which we desire to take the function with a plus sign, and below, the value for which we want to take the function with a minus sign.

Substituting, under the integral sign in (3.4.7), $v(t)$ expressed in terms of $z(t)$ in accord with (3.4.6) and putting expression (3.4.8) into the right-hand side of (3.4.7) we get the identity

$$\int_a^b z'(t)\,dt = z(t)\Big|_a^b, \tag{3.4.9}$$

or, in other notation,

$$\int_a^b F'(x)\,dx = F(x)\Big|_a^b. \qquad (3.4.9a)$$

It will be seen that the positions of a and b on the left and on the right are the same, which is a good mnemonic device for memorizing (3.4.9) or (3.4.9a).

We have studied in sufficient detail the concept of the indefinite integral. It is now time to examine some examples. Let us consider the problem of the distance covered during the time interval from a to b at a velocity equal to $v(t) = t^2$. The sought distance is equal to the definite integral

$$z(a, b) = \int_a^b t^2\,dt.$$

In this problem, the indefinite integral $z(t)$ is obtained by solving the equation

$$\frac{dz(t)}{dt} = v(t) = t^2. \qquad (3.4.10)$$

But we know that $d(t^3)/dt = 3t^2$, hence $d(t^3/3)/dt = (1/3)(3t^2) = t^2$. Consequently, the equation is satisfied by the function

$$z(t) = \frac{t^3}{3} + C, \qquad (3.4.11)$$

where C is an arbitrary constant that can be taken to be zero, since we know that this does not affect in any way the difference of two values of function z for two values of the independent variable.

Substituting (3.4.10) and (3.4.11) into (3.4.9) yields

$$\int_a^b t^2\,dt = \frac{t^3}{3}\Big|_a^b = \frac{b^3}{3} - \frac{a^3}{3}.$$

The special case of $a = 1$ and $b = 2$ yields

$$\int_1^2 t^2\,dt = \frac{8}{3} - \frac{1}{3} = \frac{7}{3} \simeq 2.333.$$

Thus, using the indefinite integral, we obtain in a few lines the *exact* result which we laboriously approached in Section 3.1 by numerical computations.

The definite integral is the limit of a sum of the form

$$v(t_0)(t_1 - t_0) + v(t_1)(t_2 - t_1) + \ldots$$

as each term tends to zero and the number of terms increases correspondingly. In an approximation of this sum, one has to partition the domain of integration into a number of subintervals, find the approximate value of the distance $v\Delta t$ in each subinterval, and then add these values. To obtain good accuracy requires numerous arithmetic operations. But if we know the indefinite integral $z(t)$, that is, if we know the function whose derivative is equal to the integrand $v(t)$, then any definite integral $\int_a^b v(t)\,dt$ is obtained immediately from formula (3.4.9). The possibility of finding functions with a given derivative (indefinite integrals) has "suddenly" put at our disposal a powerful method for computing sums (definite integrals).

The indefinite integral is sometimes called a ***primitive function*** (***antiderivative***). This term is understood as the inverse of a derivative: we are speaking of a function whose derivative is known. The term is used in textbooks where the problem of finding a function from the known derivative of the function is solved before definite integrals are considered, that is, where integral calculus is started from the concept of an indefinite integral rather than that of a definite integral. We do not use this approach here.

An indefinite integral can always be expressed in terms of a definite integral:

$$z(t) = C + \int_{a_0}^t v(x)\,dx. \qquad (3.4.12)$$

Applying the rule concerning the derivative of a definite integral with

respect to the upper limit, it is easy to verify that $z(t)$ given by Eq. (3.4.12) satisfies (3.4.6) for arbitrary constants C and a_0.

In all problems, the answer always involves the difference $z(b) - z(a)$, and this is independent of C and a_0. Therefore (3.4.12) can be written more compactly:

$$z(t) = \int^{t} v(x)\,dx.$$

This is frequently abridged still further to

$$z(t) = \int v(t)\,dt. \qquad (3.4.13)$$

This notation is widely used and we will employ it, but one should bear in mind that, properly speaking, it is not correct.[3.12] It may be compared with those grammatically loose expressions that occur in everyday speech and are clear to all (except children and pedants), for example, expressions like "to eat another plate" or "I'm ready to throw the whole thing over". Notations and expressions that are not altogether precise but are sufficiently clear are frequently used in mathematics, they do not usually create any difficulties for anyone.

The notation (3.4.13) violates the rule by which the dummy variable of integration does not appear in the result. Therefore, when using the abridged notation (3.4.13) one should always bear in mind that this is a conventional contraction of the exact expression (3.4.12), in which $C = z(a_0)$.

The familiar formulas for derivatives yield a table of indefinite integrals:

$$\int dt = t, \quad \int t\,dt = \frac{t^2}{2}, \quad \int t^2\,dt = \frac{t^3}{3},$$
$$\int \frac{dt}{t^2} = -\frac{1}{t}.$$

In addition, the results of Exercises 2.4.2 and 2.4.3 enable us to write

$$\int \frac{dt}{\sqrt{t}} = 2\sqrt{t}. \quad \int \sqrt{t}\,dt = \frac{2}{3}\sqrt{t^3}.$$

To verify any of these formulas, it suffices to find the derivative of the right-hand side. If in doing so we get the function under the integral sign, the formula is correct.

Methods for finding indefinite integrals of a variety of functions are considered in detail in Chapter 5. Thanks to the relationship which exists between an integral and a derivative, we are able to find the integrals of a large number of functions.

The problem of integration is technically a much more complicated job than the problem of finding derivatives. The complexity is due, for one thing, to the fact that in the integration of rational (i.e. not containing radicals) algebraic expressions we get answers involving logarithms and inverse trigonometric functions. In the integration of algebraic expressions with radicals the result is sometimes expressible only with the aid of new, nonelementary, functions that cannot be expressed in terms of a finite number of operations involving algebraic, power, and trigonometric functions (see Section 5.5). However, the difficulties of expressing integrals by formulas should not eclipse the fundamental simplicity and clarity of the integral concept. If it is impossible (or difficult) to evaluate an integral by formula (3.4.5a), it is always possible to evaluate an integral by means of cumbersome, yet fundamentally very simple, computations.

In the last few years, with the advent of programmable pocket calculators, it has become very easy to calculate the sum of 10, 20, or even 50 terms—this requires 10 to 30 minutes depending on the complexity of the integrand. In many cases, therefore, a direct calculation of an integral is preferable over trying to obtain complex formulas. In the diary of the famous physi-

[3.12] Note that in the left-hand side of (3.4.13) we have a *function* $z(t)$, while in the right-hand side we have an indefinite integral, that is, a *family of functions* $F(t) + C$ differing from one another by a constant (since in $F(t) + C$ the constant C is arbitrary).

cist Enrico Fermi there is a calculation of a rather simple definite integral. Notwithstanding the fact that the corresponding indefinite integral could be expressed in terms of the inverse sine and the natural logarithm, Fermi preferred to find the integral numerically, not employing formula (3.4.5a).

Many books are devoted to fast and exact methods of numerical integration. The reader wishing to familiarize himself with these methods can turn to the book [15].

Exercises

3.4.1. Evaluate the following integrals (a) $\int_0^1 t^2 dt$, (b) $\int_1^{1.1} t^2\,dt$, (c) $\int_1^2 dt/t^2$, and (d) $\int_1^3 dt/\sqrt{t}$.

3.4.2. (a) Using an integral, find the area of a right triangle having a base b and an altitude h. [*Hint.* Put the origin of coordinates at the vertex of an acute angle of the triangle and the right angle on the x axis (Figure 3.4.1*a*). Find the equation of the hypotenuse in this system of coordinates and find the area as an integral.]

(b) Find the area of the same triangle by placing the right angle at the origin and an acute angle of the triangle at the point $(b, 0)$ (Figure 3.4.1*b*). [*Hint.* In integrating, make use of the obvious property of the integral of a sum of two terms, $\int (f+g)\,dx = \int f\,dx + \int g\,dx$.]

Remark. Do not be indignant that a lot of effort is put into finding the familiar answer $S_\Delta = (1/2)\,bh$ because the method of integration will be used later on in cases where elementary methods of finding areas do not suffice.

3.4.3. (a) Find the area under the parabola $y = ax^2$ (Figure 3.4.2*a*) bounded by the vertical line $x = x_0$ and the axis of abscissas. Express the area in terms of the coordinates x_0 and y_0 of the end of the arc OA of the parabola.

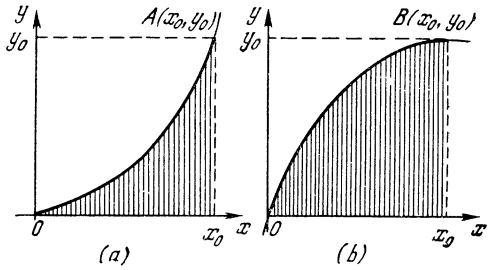


Figure 3.4.2

(b) The same for a parabola passing through the origin and having a horizontal tangent at point B (x_0, y_0) (Figure 3.4.2*b*). [*Hint.* The answer may be obtained at once by using the result of Exercise 3.4.3 (a). Still and all, take the hard way and do all the operations in their requisite order, without resorting to a clever trick. Seek the equation of the parabola in the form $y = kx^2 + px + q$, where k, p, and q are found from the condition of passage through the points (x_0, y_0) and the origin (0, 0) and from the condition that the tangent at the point (x_0, y_0) is horizontal. Express the area in terms of x_0 and y_0. If you are not able to use the result of Exercise 3.4.3 (a), then first try the above-described procedure and the result will suggest the relationship between Exercises 3.4.3 (a) and 3.4.3 (b).

3.4.4. Write the expression for the area of a semicircle of radius r (Figure 3.4.3) in the form of a definite integral. [*Hint.* Use the equation of a circle, $x^2 + y^2 = r^2$.]

3.4.5. Evaluate the integral $\int_0^1 dx/(1+x^2)$ using the trapezoid formula (see Section 3.1), putting first $n = 5$ and then $n = 10$. Carry out the computations to the fourth decimal place.

Remark. The exact value of the integral is $\pi/4 \simeq 0.785\,398$ (see Exercise 5.2.4). An approximate evaluation of the integral enables us to obtain an approximate value for the number π.

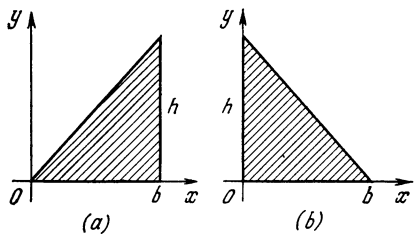


Figure 3.4.1

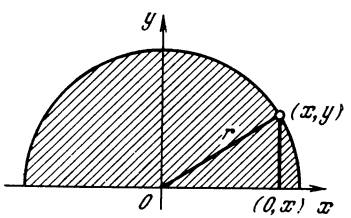


Figure 3.4.3

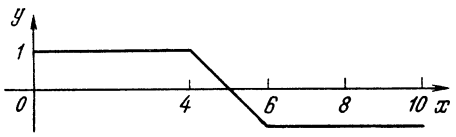


Figure 3.4.4

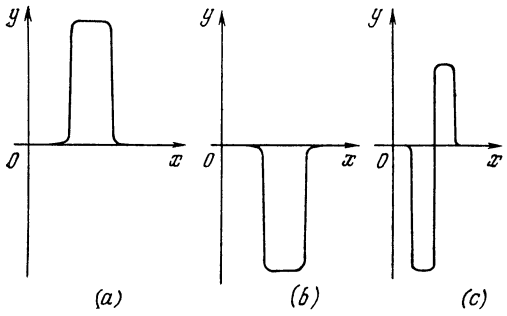


Figure 3.4.5

3.4.6. Construct the graph of the function $F(x) = \int_a^x y(x)\,dx$. The function $y(x)$ is given graphically (Figure 3.4.4). For values of a take $a = 0$, $a = 4$, and $a = 8$.

3.4.7. Construct the graph of the function $F(x) = \int_0^x y(x)\,dx$, where the (possible) graphs of the function $y(x)$ are depicted in Figure 3.4.5.

3.4.8. Construct the curves $F(x) = \int_0^x \varphi(x)\,dx$, where the functions $\varphi(x)$ are given by the curves which appear in Figures H.6 and H.7 at the end of this book. Compare $F(x)$ with the curves given in Figures 2.5.5 and 2.5.6.

3.5 Properties of Integrals

Above we considered the simplest case of a definite integral having a positive integrand and with the upper limit of integration greater than the lower limit:

$$z(a, b) = \int_a^b v(t)\,dt, \quad v(t) > 0 \text{ and } b > a.$$

The integral here is clearly positive, since it is equal to the limit of a sum of positive terms. The integral has the simple physical meaning of a *distance* covered (if $v = v(t)$ is the velocity) or an *area* (if $v = v(t)$ is the equation of a curve). But what is the sign of the integral of a *negative* function, that is, in the case of $v(t) < 0$?

For the time being, we leave the condition $b > a$. In the expression for the sum (which in the case of passage to the limit becomes an integral), the factor Δt in each term is positive, the factor $v(t)$ is negative, which means that each summand is negative, the sum is negative, hence the integral is negative, too. To summarize, if $v(t)$ is negative and $a < t < b$ (so that $b > a$), then $\int_a^b v(t)\,dt < 0$. In the case of motion, the meaning of the answer is simple: a negative $v(t)$ signifies that the motion occurs in a direction opposite to the positive direction, or the direction of increasing z coordinate. The distance traversed in the negative direction will always be regarded as negative. In this case, z decreases, $z(b) < z(a)$. The general formula

$$z(b) = z(a) + z(a, b) = z(a) + \int_a^b v\,dt,$$

or

$$\int_a^b v(t)\,dt = z(b) - z(a),$$

remains valid, only in the second form both sides are negative.

In the case of a velocity that *changes sign*, it may happen that, for one thing, $\int_a^b v(t)\,dt = 0$, although $b > a$ and $b \neq a$. This occurs if part of the time between a and b the body moves in one direction, and during another time interval it moves in the opposite direction, with the result that by time b it returns to the position it occupied at time a.

Let us examine the problem of the area under a curve. For $b > a$ and

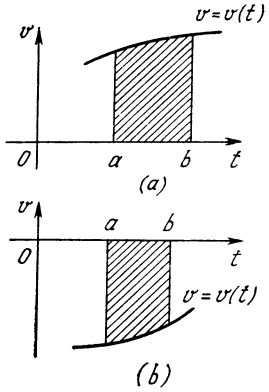


Figure 3.5.1

$v(t) > 0$, the integral is equal to the area bounded by the curve $v = v(t)$, the t axis, and the vertical lines $t = a$ and $t = b$ (Figure 3.5.1*a*). For $b > a$ and $v(t) < 0$, we know that $\int_a^b v\,dt < 0$. In this case the curve lies below the axis of abscissas (Figure 3.5.1*b*). Thus, in order to preserve the law by which the area is equal to an integral, it is necessary to regard the area as *negative* if the curve $v = v(t)$ lies *below* the axis of abscissas.

If we take a function that changes sign, say, $v(t) = \sin t$ (Figure 3.5.2), then, as can easily be seen, the area under such a curve over an interval equal to the period of the function, that is, from $t = 0$ to $t = 2\pi$, will be equal (by our definition) to zero. This means that the area under the first half-wave, which we assume to be positive, and the negative area under the second half-wave cancel each other exactly.[3.13]

The definite integral also generalizes to the case when the upper limit is

[3.13] If our problem is to find how much paint is required to paint over the hatched portions in Figure 3.5.2, such a definition of area is no good. In this case we have to partition the entire integral into parts over which v does not change sign (in our case there will be two parts, from 0 to π and from π to 2π), then compute the integral over each portion separately, and finally add the absolute values of the integrals of the separate parts.

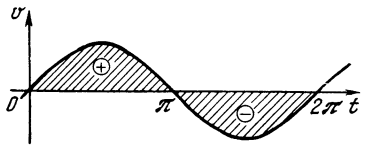


Figure 3.5.2

less than the lower. In this case we will no longer speak of the distance, time, and velocity (cf. Section 3.1) and will regard the definition of the integral as a sum (see Section 3.2). Referring to Figure 3.5.3, we again partition the interval between a and b into n parts by the intermediate values $t_1, t_2, \ldots$ $\ldots, t_{n-1}$ and convince ourselves that all the Δt are now negative. It is now easy to see that

$$\int_a^b v(t)\,dt = -\int_b^a v(t)\,dt, \tag{3.5.1}$$

since in any partition of the interval from a to b the corresponding sums will differ as to signs of the subintervals Δt in all terms.

An essential property of the integral consists in the fact that the domain of integration may be divided into parts: the distance covered in the time interval between a (beginning) and b (end) may be represented as the sum of the distances traversed between time a to c (an intermediate point in time) and between c and b (Figure 3.5.4*a*):

$$\int_a^b v(t)\,dt = \int_a^c v(t)\,dt + \int_c^b v(t)\,dt. \tag{3.5.2}$$

With the aid of (3.5.1) we can extend formula (3.5.2) to the case where c lies *outside* the interval from a to b instead of inside the interval. Suppose

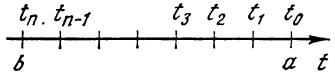


Figure 3.5.3

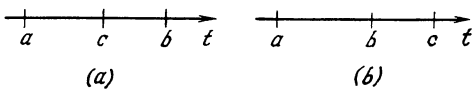


Figure 3.5.4

that $c > b > a$ (Figure 3.5.4*b*). Then, obviously,

$$\int_a^c v\,(t)\,dt = \int_a^b v\,(t)\,dt + \int_b^c v\,(t)\,dt. \tag{3.5.3}$$

Transpose the last term to the left and take advantage of (3.5.1). The result is

$$\int_a^c v\,dt - \int_b^c v\,dt = \int_a^c v\,(t)\,dt + \int_c^b v\,(t)\,dt = \int_a^b v\,dt. \tag{3.5.4}$$

In this way we get an equation that coincides exactly with (3.5.2).

We can likewise consider different arrangements of numbers a, b, and c, points on the number axis (there are six such variants in all). The reader will find no difficulty in considering them and in convincing himself that formula (3.5.2) proves to be valid in all these cases, that is to say, irrespective of the mutual arrangement of the numbers (points) a, b, and c.

Actually, we have derived all these properties of definite integrals from the definition of an integral as the limit of a sum. These properties likewise follow from the expression for a definite integral in terms of an indefinite integral. Indeed, suppose that the indefinite integral $\int v\,(t)\,dt$ equals $z\,(t)$. Then

$$\int_a^b v\,(t)\,dt = z\,(b) - z\,(a) \text{ and}$$

$$\int_b^a v\,(t)\,dt = z\,(a) - z\,(b) = -\int_a^b v\,(t)\,dt.$$

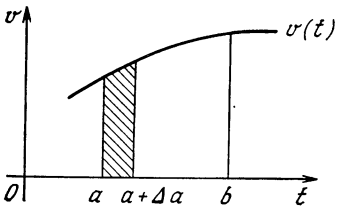


Figure 3.5.5

Similarly,

$$\int_a^b v\,(t)\,dt = z\,(b) - z\,(a), \quad \int_a^c v\,(t)\,dt = z\,(c) - z\,(a), \quad \int_c^b v\,(t)\,dt = z\,(b) - z\,(c),$$

whence

$$\int_a^b v\,(t)\,dt = \int_a^c v\,(t)\,dt + \int_a^b v\,(t)\,dt.$$

The fundamental law by which the derivative of an integral is equal to the integrand refers to the derivative with respect to the upper limit. If the definite integral is regarded as a function of the lower limit of integration, with the upper limit held constant, then we get the answer with opposite sign:

$$\frac{dz\,(a,\ b)}{da} = \frac{d}{da}\left(\int_a^b v\,(t)\,dt\right) = -v\,(a). \tag{3.5.5}$$

The minus sign in this formula is easy to understand if we regard the integral as an area: a positive increment Δa clearly diminishes the area under the curve (Figure 3.5.5).[3.14] We can formally obtain the same result by interchanging the limits of integration (this will introduce a minus sign) and by

[3.14] The area bounded by the vertical lines $t = a + \Delta a$ and $t = b$, the curve $v = v\,(t)$, and the t axis is *smaller* than the one bounded by the lines $t = a$ and $t = b$, the curve $v = v\,(t)$, and the t axis.

using the familiar law on the derivative with respect to the upper limit:

$$\frac{d}{da}\left(\int_a^b v(t)\,dt\right) = \frac{d}{da}\left(-\int_b^a v(t)\,dt\right)$$
$$= -v(a)$$

In connection with the question of the sign of the integral, we note an example that frequently disconcerts the beginner. Let us consider

$$\int \frac{dx}{x^2} = -\frac{1}{x}. \qquad (3.5.6)$$

This equation follows from the earlier found value of the derivative

$$\frac{d(1/x)}{dx} = -\frac{1}{x^2}.$$

Is the sign of the integral correct here? Can the integral of a positive function, $1/x^2$, be negative? Does not this sign contradict the assertions made above?

Any dismay is due to the fact that formula (3.5.6) is not written in proper fashion. If we write it as

$$\int \frac{dx}{x^2} = -\frac{1}{x} + C,$$

then we cannot say that the sign of the integral is always negative, since this also depends on the sign and value of quantity C.

Actually, all statements concerning the sign of the integral have referred to the *definite* integral. Let us take

$$\int_a^b \frac{1}{x^2}\,dx = \left(-\frac{1}{x}\right)_a^b = \left(-\frac{1}{b}\right) - \left(-\frac{1}{a}\right)$$
$$= \frac{1}{a} - \frac{1}{b} = \frac{(b-a)}{ab}.$$

When $b > a$ (and both a and b are positive), the integral is positive, as it should be, that is, formula (3.5.6) is correct and leads to a correct result for the definite integral.

Looking ahead, we may note that the integral $\int dx/x^2$ involves other, no longer fictitious but real, difficulties, which will be considered in Section 5.2.

3.6. Examples and Applications

In Chapter 2 and in the preceding sections of Chapter 3 we examined the relationship between distance and velocity and the relationship between the equation of a curve and the area under the curve. These relationships are the concrete problems on the basis of which differential and integral calculus took shape in the history of mathematics. But the derivative and the integral are, of course, applicable not only to the foregoing problems but to an extraordinarily broad range of phenomena in the most diverse fields of science, technology, and everyday life. Actually, the derivative, the integral, and the Newton-Leibniz theorem (which establishes the relationship between these two concepts) constitute the most suitable language describing the laws of nature.

Anyone beginning the study of a foreign language has to repeat all manner of simple phrases just to get used to the new medium: "there is a table in the room," "a glass is on the table," "there is a cat on the floor," "there is a mouse near the cat." The same goes for the student of higher mathematics. He, or she, has to repeat the relationships between the derivative and the integral in a multitude of similar examples. One must first learn a foreign language before attempting to express ideas in it. And that is our task now: we have to learn to express familiar relationships and to formulate problems in the language of higher mathematics before trying to solve problems and obtain new results.[3.15]

[3.15] Goethe put it this way, "Mathematicians are a species of Frenchmen; if something is said to them, they translate it into their own language and presto! it is something entirely different." A well-known statement in a similar vein was made by the distinguished American physicist and mathematician ***Josiah Willard Gibbs*** (1839-1903), father of modern thermodynamics. During a discussion of what should be given preference to in curricula—languages, especially Latin and Greek, or mathematics—Gibbs said that mathematics is a language too. That was his only public

Here are a few typical examples that illustrate the content of the present chapter. The majority of these will be developed further in subsequent chapters.

1. Acceleration and uniformly accelerated motion; the work performed by a force. Earlier, in Chapters 2 and 3, we constantly considered the displacement of a body and the velocity of motion (as the derivative of the body's coordinate with respect to time). But after the instantaneous velocity $v(t) = dz(t)/dt$ has been found and we know the relationship between the instantaneous velocity and time, we can ask how the velocity varies with time. In brief this question has been studied in Section 2.7.

The derivative of velocity with respect to time is called the ***acceleration*** and is denoted by a:

$$\frac{dv(t)}{dt} = a(t) \qquad (3.6.1)$$

Since the dimensions of velocity are cm/s or m/s, the dimensions of acceleration are cm/s^2 or m/s^2.

We write the velocity in the form of the derivative of the coordinate with respect to time, $v = dz/dt$, and substitute this into Eq. (3.6.1). The result is

$$a = \frac{d}{dt}\left(\frac{dz}{dt}\right) = \frac{d^2z}{dt^2}. \qquad (3.6.2)$$

Thus, *acceleration is the second derivative of the coordinate with respect to time.*

Note the exponents (the "twos") in the expression d^2z/dt^2: the dimensions of acceleration are those of z/t^2 (the symbol d is not a quantity).

If we know the acceleration as a function of time, the instantaneous velocity $v = v(t)$ may be written in the form of an integral of $a(t) = dv/dt$:

$$v(t) = v_0 + \int_{t_0}^{t} a(t)\,dt. \qquad (3.6.3)$$

(Here $v_0 = v(t_0)$ is the instantaneous velocity at the initial moment of time $t = t_0$; to verify this we need only substitute t_0 for t on both sides of Eq. (3.6.3).) For one thing, if the acceleration is constant (a = constant, the case of *uniformly accelerated* motion), from (3.6.3) it follows that

$$v(t) = v_0 + a\int_{t_0}^{t} dt = v_0 + a(t - t_0), \qquad (3.6.4)$$

or

$$v(t) = at + b, \qquad (3.6.4a)$$

where $b = v_0 - at_0$, that is, the velocity of uniformly accelerated motion is a *linear* function of time.

Knowing the velocity $v = v(t)$ of a motion, we can establish how the coordinate varies with time, that is, find the *law of motion* $z = z(t)$:

$$z(t) = z_0 + \int_{t_0}^{t} v(t)\,dt, \qquad (3.6.5)$$

or, in view of (3.6.3),

$$z(t) = z_0 + \int_{t_0}^{t}\left[v_0 + \int_{t_0}^{t} a(t)\,dt\right]dt$$

$$= z_0 + v_0(t - t_0) + \int_{t_0}^{t}\left[\int_{t_0}^{t} a(t)\,dt\right]dt, \qquad (3.6.5a)$$

where $z_0 = z(t_0)$ is the initial position (coordinate) at time t_0 (substitute

statement (all the more remarkable!) on the general aspects of teaching. Almost the same thing was stated much earlier by the famous ***Galileo Galilei*** (1564-1642), who declared that "the laws of Nature are recorded in the greatest of all books, which is always open to our gaze (I refer to the Universe), but you cannot understand it unless you first learn to understand the language in which it is written, and it is written in the language of mathematics." (And, rather surprisingly, he goes on to say that "its letters [i.e. the letters of mathematics—Ya.Z. and I.Ya.] are triangles, circles, and other geometric figures without which its words cannot be understood." Thus Galileo regarded geometry and not differential and integral calculus as the language of Nature. But we must bear in mind that in Galilei's time that is all there was to mathematics—the geometry of the ancient Greeks. Naturally, Galileo could not foresee the creation of higher mathematics by Leibniz and Newton, who came later.)

t_0 for t on both sides of (3.6.5)). Of course, in general form, formula (3.6.5a) looks unwieldy. But if a = constant, then the velocity $v = v(t)$ is given by a simple formula (3.6.4a) or (3.6.4), and Eq. (3.5.5a) enables us to describe the motion of a body in closed form:

$$z(t) = z_0 + \int_{t_0}^{t} (at+b)\,dt = z_0 + a\,\frac{t^2-t_0^2}{2} + b(t-t_0) = \frac{a}{2}t^2 + bt + c, \qquad (3.6.6)$$

where $c = -(a/2)\,t_0^2 - bt_0 + z_0$ (check to see whether the derivative of the right-hand side of (3.6.6) with respect to t coincides with the right-hand side of (3.6.4a)). Thus, the dependence of the coordinate z of a body in uniformly accelerated motion on time is *quadratic*. For example, in the case of motion under the action of gravity (free fall) we have $a = -g$, where $g \simeq 9.8$ m/s^2 (the minus sign being due to the fact that the upward direction is taken to be positive). Assuming $a = -g$ in (3.6.4), (3.6.4a), and (3.6.6), we get the well-known formulas

$$v(t) = v_0 - \int_{t_0}^{t} g\,dt = v_0 - g(t-t_0) = v_0 - gt_1,$$

where $t_1 = t - t_0$ is the new time variable reckoned from $t = t_0$, and

$$z(t) = z_0 + \int_{t_0}^{t} [v_0 - g(t-t_0)]\,dt = z_0 + \int_0^{t_1} (v_0 - gt_1)\,dt_1 = z_0 + v_0 t_1 - \frac{g}{2}t_1^2.$$

In connection with acceleration we recall that the *work* performed by a force is equal to force F times distance s (here we assume that the distance or displacement coincides with the direction in which the force acts; see

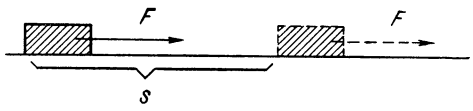


Figure 3.6.1

Figure 3.6.1). However, this fact can be employed only when the force is constant, a situation that almost always is unrealistic. If the force F varies in the course of motion, then the whole path, or distance, from $z = z_0$ to another (variable) position of the body, or coordinate, z has to be partitioned into small intervals of length Δz. The products $F\Delta z$, obtained on the assumption that the force in each interval remains constant (since the Δz are small, F does not have time to vary appreciably), must then be added. In this way we arrive at the "integral sum" $\sum F\Delta z$, in which in order to obtain an expression for A we must pass to the limit as all the Δz tend to zero. The last circumstance makes it possible to assume that $v = dz/dt$ is constant on each subinterval, with the result that the length Δz can be expressed by the product of v by the time Δt which it takes to traverse Δz. Hence, the sum $\sum F\Delta z$ is replaced with the sum $\sum Fv\Delta t$, which enables representing the work A either in the form of an integral with coordinate z as the variable of integration or in the form of an integral with time t as the variable of integration:

$$A = \int_{z_0}^{z} F\,dz = \int_{t_0}^{t} Fv\,dt. \qquad (3.6.7)$$

Now we recall that the force F is equal to the mass of the body, m, times the acceleration $a = dv/dt$ (Newton's second law). This enables us to rewrite (3.6.7) thus:

$$A = m\int_{t_0}^{t} \frac{dv}{dt}\,v\,dt. \qquad (3.6.8)$$

Since, obviously, the function $v\,(dv/dt)$ is the first derivative of $v^2/2$ (see Sec-

tion 2.3), we finally obtain

$$A = m\left(\frac{v^2}{2}\right)\Big|_{t_0}^{t} = \frac{mv^2}{2} - \frac{mv_0^2}{2}, \qquad (3.6.9)$$

where $v = v(z)$ and $v_0 = v(z_0)$: the *work* A done by a force on a body *is equal to the increment* $mv^2/2 - mv_0^2/2$ *of the kinetic energy* $mv^2/2$ *of the body* (of course, both the work A and the increment of the kinetic energy may prove to be negative simultaneously).

For a further detailed investigation of the questions touched on here, see Chapters 9 and 10.

2. Stretching of a wire. Consider a rod one meter long, 0.4 cm in diameter, and made of copper. This piece of wire is hung by one of its ends. The question is: how much will the rod stretch under its own weight? The stretching of the wire is governed by ***Hooke's law***,[3.16] which is valid for not too large forces (and elongations) and states that *the deformation* (and *stretching*) l *is directly proportional to the applied force* F *and the length of the wire,* L, *and is inversely proportional to the cross-sectional area* S:

$$l = \frac{1}{E}\frac{FL}{S}, \qquad (3.6.10)$$

where E is a constant factor depending on the material of the wire and known as *Young's modulus.*

Naturally, we cannot replace L with the length of the rod, 1 m (100 cm), and F with the weight of the rod, which can easily be found from the volume of the rod and the density of copper, since the forces acting on different portions of the rod differ from point to point: while for point A at which the rod is suspended this force is indeed equal to the weight of the entire rod, for the other (free) end B this force is simply zero, since this point is under no load (Figure 3.6.2). For this reason we must partition the entire wire (with coordinate z being reckoned from the free end B) into small sections of length Δz each; for each such section at a distance z from B, the force may be assumed constant and equal to the weight of the part of the rod from point B upward (the length of this part is z): $F = \rho g S z$, where ρ is the density of copper, g the acceleration of gravity, and $Sz = V$ the volume of the part of the rod under consideration. Thus, to this small section of the wire of length Δz there corresponds an elongation

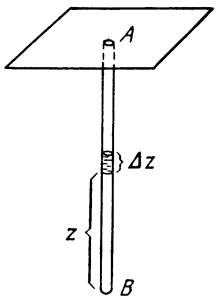


Figure 3.6.2

$$\Delta l = \frac{\rho g}{E}\frac{Sz\Delta z}{S} = \frac{\rho g}{E} z\Delta z, \qquad (3.6.11)$$

and the total elongation of the entire wire is obtained by summation (or integration) of all such "elementary" elongations:

$$l = \frac{\rho g}{E}\int_0^{100} z\,dz = \frac{\rho g}{E}\left(\frac{z^2}{2}\right)\Big|_0^{100}$$

$$= \frac{\rho g}{E}\frac{(100)^2}{2} \qquad (3.6.11a)$$

Substituting the numerical values $g = 980$ cm/s^2, $\rho = 8.9$ g/cm^3, $E = 9.8 \times 10^{13}$ g/cm·s^2, we obtain

$$l = \frac{980 \times 8.9 \times 10^4}{2 \times 9.8 \times 10^{13}} \simeq 4.5 \times 10^{-5} \text{ cm}$$

(note that the factor $100^2 = 10^4$ in the numerator has the dimensions of cm^2).

3. Water flow from a vessel. Picture a vessel of arbitrary shape (Figure 3.6.3) with water flowing out. The mass of liquid in the vessel at a given instant

[3.16] Robert ***Hooke*** (1635-1703), an outstanding English scientist, chemist, and mathematician and a contemporary (and in many cases a scientific opponent) of Sir Isaac Newton.

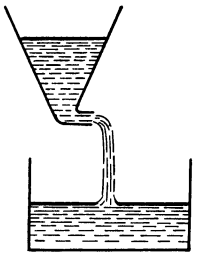


Figure 3.6.3

of time is equal to M, which is a function of time: $M = M(t)$. The liquid collects in another vessel; the mass of liquid here is $m(t)$. We denote the mass of liquid flowing out of the first vessel in unit time by $W = W(t)$; this quantity is known as the *flow rate* and has the dimensions of g/s. The quantities m, M, and W are connected by the following relations:

$$\frac{dM}{dt} = -W(t), \quad \frac{dm}{dt} = W(t). \qquad (3.6.12)$$

These differential, that is, involving derivatives, relations can be written in the form of integrals. Suppose that at a certain initial moment t_0 the mass of the liquid in the first vessel was $M(t_0) = M_0$, while the second vessel was empty, or $m(t_0) = 0$. Then

$$m(t_1) = \int_{t_0}^{t_1} W(t)\,dt,$$
$$M(t_1) = M(t_0) - \int_{t_0}^{t_1} W(t)\,dt. \qquad (3.6.13)$$

Thus, the mass of liquid at a definite time t_1 is expressed in terms of an integral in which the variable of integration t runs through all values from t_0 to t.

It will be convenient in (3.6.13) to replace t_1 with t and rename the variable of integration (using the fact that it is a dummy variable) and call it, say, tau (which is the Greek letter corresponding to the Latin t). We then have

$$m(t) = \int_{t_0}^{t} W(\tau)\,d\tau,$$
$$M(t) = M(t_0) - \int_{t_0}^{t} W(\tau)\,d\tau. \qquad (3.6.13a)$$

Ordinarily, this is simply written as

$$m(t) = \int_{t_0}^{t} W(t)\,dt,$$
$$M(t) = M(t_0) - \int_{t_0}^{t} W(t)\,dt, \qquad (3.6.13b)$$

but remember that the t under the integral sign has a different meaning from the independent variable t in $M(t)$ and $m(t)$, which coincides with the upper limit of integration. In this respect, the notation (3.6.13) and (3.6.13a) is more exact than (3.6.13b).

The formulas given above correspond to an experiment in which M and flow rate W are measured at distinct moments of time. Often, however, the problem is stated differently: the flow rate W depends in some known fashion on the pressure, that is, the height of a column of liquid, h. In turn, given a definite shape of vessel, the function h is a function of M. Thus, we know the flow rate W as a function of the quantity of liquid in the vessel, $W = W(M)$. Then (3.6.12) takes the form

$$\frac{dM}{dt} = -W(M). \qquad (3.6.14)$$

Equations of such a type, which connect the unknown function (in our case the function $M = M(t)$) with its derivative are called ***differential equations***; thus, Eq. (3.6.14) is a differential equa-

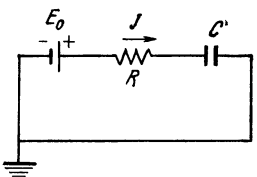


Figure 3.6.4

tion for the unknown function $M(t)$, that is, a function that will be found when we solve the equation. The topic of differential equations, as well as the solution to this specific problem, will be considered in Chapter 6.

4. Capacitor. We denote the charge accumulated on a capacitor C (Figure 3.6.4) (the quantity of electricity) by q. In the SI system of units (see Appendix 5), q is measured in coulombs (abbreviated C). The electric current j flowing in the wire is the quantity of electricity flowing in unit time and is measured in amperes (abbreviated A). One ampere is a current of one coulomb per second: 1 A = 1 C/s; thus, the quantity of electricity has the dimensions of A·s, since in the SI system of units the ampere is a base unit.

The charge on a capacitor[3.17] and the current are connected by the equation

$$\frac{dq}{dt} = j \qquad (3.6.15)$$

(the positive direction of the current is indicated by the arrow in Figure 3.6.4). If the variation in current flow with time, $j = j(t)$, is given or has been found via experiment, then we can write the integral relation

$$q(t_1) = q(t_0) + \int_{t_0}^{t_1} j(t)\,dt. \qquad (3.6.15a)$$

If the *capacitance* C of the capacitor is given (note that the capacitance and the capacitor are designated in Figure 3.6.4 by the same letter), then the voltage drop across the capacitor may be expressed in terms of q in the following manner: $\varphi_C = q/C$, and the voltage drop across the resistance R is $\varphi_R = E_0 - \varphi_C = E_0 - q/C$, where E_0 is the battery voltage. By Ohm's law, the current flowing through resistance R is $j = R^{-1}(E_0 - q/C)$. Using (3.6.15), we arrive at the *differential equation*

$$\frac{dq}{dt} = \frac{1}{R}\left(E_0 - \frac{q}{C}\right). \qquad (3.6.16)$$

[3.17] We use the term *charge on the capacitor* to mean the quantity of positive electricity on the left-hand plate of capacitor C in Figure 3.6.4 expressed in coulombs.

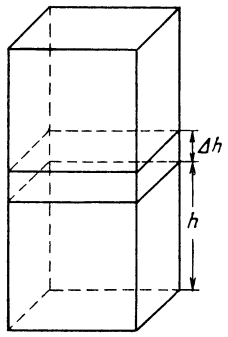


Figure 3.6.5

Problems involving capacitors are discussed in detail in Part 2 of this book.

5. Atmospheric pressure. Imagine a vertical column of air with constant cross-sectional area S cm^2. The density of the air is ρ g/cm^3 and depends on the altitude h above the earth's surface. The volume of a thin layer contained between h and $h + \Delta h$ is equal to $S\Delta h$ (Figure 3.6.5). Inside this thin layer, the density $\rho(h)$ may be regarded as constant, which is precisely why we took a thin layer. In the given instance, Δh may be pictured as 1 meter or 10 meters or even (with a slightly lower accuracy) as 100 meters, since air density varies roughly by 12 to 14% per kilometer of altitude.

Since the volume of the layer of air Δh thick is $S\Delta h$, the mass of air contained in this layer, by the definition of density, is $\Delta m = \rho S \Delta h$. To find the mass of air in a column extending from h_1 to h_2, we must construct the sum of expressions of the type $\rho S \Delta h$, with S constant and ρ varying with altitude h. The sum is extended over all layers Δh into which the column of air is partitioned. Decreasing all Δh without limit (that is, sending them to zero), we arrive at the integral

$$m(h_1, h_2) = S\int_{h_1}^{h_2} \rho(h)\,dh. \qquad (3.6.17)$$

The mass of air in a column from the earth's surface ($h = 0$) to an altitude

h is

$$m(0, h) = S\int_0^h \rho(h)\,dh. \qquad (3.6.17a)$$

The mass of air above a given altitude h is

$$m = S\int_h^\infty \rho(h)\,dh, \qquad (3.6.17b)$$

where the symbol ∞ in the upper limit takes the place of a very large number H such that any subsequent increase in this quantity does not substantially change the integral.

The pressure P at some altitude h multiplied by the area S is equal to the force with which the entire column of air above h is attracted to the earth. The force of gravity is equal to the mass multiplied by the acceleration of gravity g,[3.18] whence

$$P(h) = \int_h^\infty g\rho(h)\,dh. \qquad (3.6.18)$$

Using formula (3.5.5), we get the *differential equation*

$$\frac{dP}{dh} = -g\rho(h). \qquad (3.6.19)$$

This formula could have been written straight off by considering the equilibrium of a thin layer dh acted upon from below by the pressure $P(h)$ and from above by the pressure $P(h + dh)$; the resultant of these two forces balances the attraction to the earth of the mass of air in the layer Δh.

We will continue on this subject in Chapter 11.

6. Volume. Let us employ integral calculus to compute the volume of bodies (solids). Let us slice a solid of total volume V (Figure 3.6.6) by planes z = constant into thin layers. In this

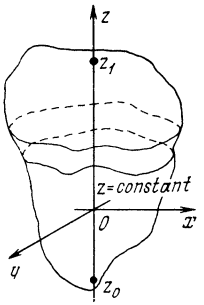


Figure 3.6.6

problem we will employ a system of coordinates in space specified by the x axis and the y axis in the horizontal plane xOy and by the vertical z axis (the third coordinate axis in space). The volume ΔV of a thin layer bounded by the planes corresponding to the values z and $z + \Delta z$ on the z axis and parallel to the xOy plane is approximately equal to the product of the cross-sectional area $S(z)$ by the thickness of the layer, Δz. Thus, if we know the area of a cross section of the solid cut by a horizontal plane, $S = S(z)$, then the volume of the solid can be computed by the formula

$$V = \int_{z_0}^{z_1} S(z)\,dz, \qquad (3.6.20)$$

where z_0 and z_1 are the smallest and greatest values of the z coordinate in our solid.

Let us apply this formula to a regular quadrangular pyramid with its

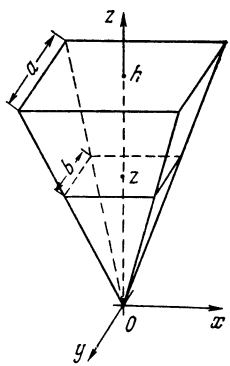


Figure 3.6.7

[3.18] In the majority of cases we can assume that the acceleration of gravity g does not alter over the altitudes h considered in this problem and that the earth has a flat surface, although both assumptions do not in any way change our arguments.

vertex at the origin of coordinates and with its axis of symmetry directed along the z axis (Figure 3.6.7). Let the altitude of the pyramid be h and the base (at the top of the figure) a square with side a. From geometry we recall that a cross section of a pyramid cut by a horizontal plane at an altitude z is a square, the side b of which is to a as z is to h, that is, $b = b(z) = a(z/h)$. Hence, the cross-sectional area is $S(z) = b^2 = (a^2/h^2)\,z^2$. The volume of the pyramid is

$$V = \int_0^h \frac{a^2}{h^2} z^2\,dz = \frac{a^2}{h^2}\int_0^h z^2\,dz.$$

Let us take advantage of the result of Section 3.4:

$$\int z^2\,dz = \frac{1}{3}z^3, \quad \int_0^h z^2\,dz = \frac{1}{3}h^3$$

We then get an expression for the volume of a pyramid:

$$V = \frac{a^2}{h^2}\,\frac{1}{3}\,h^3 = \frac{1}{3}\,a^2h$$

We found that *the volume of a pyramid is equal to one third the product of the area of the base by the altitude of the pyramid.* To derive this formula in solid geometry without the aid of integrals is a rather complicated job.

We will return to the question of the use of integral calculus for the computation of volumes of solids in Chapter 7.

Here is another example, and it is much more striking than the preceding examples. We wish to find the volume of a *cylindrical "hoof"*, that is, the part cut off from a (right circular) cylinder by a plane that intersects the cylinder along the diameter AB of the cylinder's base and forms an angle of 45° with the base (Figure 3.6.8). For the sake of simplicity we take the radius of the cylinder equal to unity (cf. Exercise 3.6.3). The cross section of the "hoof" cut by a plane perpendicular to diameter AB and distant $OM = x$ from the center O of the base is a right isosceles triangle (the right triangle MPQ with acute angle $\angle OMP = 45°$) in which the length of side MP is $\sqrt{OP^2 - OM^2} = \sqrt{1 - x^2}$. But then $PQ = PM = \sqrt{1 - x^2}$ and $S_{\triangle MPQ} =$

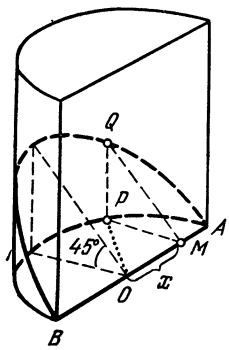


Figure 3.6.8

$S(x) = (1/2)\,MP \times PQ = (1/2)\,(1 - x^2)$. Whence, in view of formula (3.6.20), where now we must write x instead of z, the volume of the "hoof" is

$$V = \int_{-1}^{1} \frac{1}{2}(1 - x^2)\,dx = \frac{1}{2}\int_{-1}^{1}(1 - x^2)\,dx$$

$$= \frac{1}{2}\left[\int_{-1}^{1} dx - \int_{-1}^{1} x^2\,dx\right]$$

$$= \frac{1}{2}\left[x\,\Big|_{-1}^{+1} - \frac{x^3}{3}\,\Big|_{-1}^{+1}\right]$$

$$= \frac{1}{2}\left[(1-(-1)) - \left(\frac{1}{3} - \frac{-1}{3}\right)\right]$$

$$= \frac{1}{2}\left(2 - \frac{2}{3}\right) = \frac{2}{3}$$

(here we employed the obvious properties of the definite integral; $\int_a^b cf\,dx = c\int_a^b f\,dx$, where c is a number, and $\int_a^b (f+g)\,dx = \int_a^b f\,dx + \int_a^b g\,dx$).

Note that the volume of this "round" solid is a rational fraction and in no way is connected with the number π.

Exercises

3.6.1. Derive the formula for the volume of an arbitrary pyramid using the properties of parallel sections.

3.6.2. Find the volume of a (right circular) cone of altitude h and with a base that is a circle of radius R. Will the result change if

the (circular) cone is oblique, that is, if its vertex does not lie exactly over the center of the base.

3.6.3. Find the volume of a "hoof" cut off from a cylinder of radius R by a plane passing through the diameter of the cylinder's base and forming a known angle α with the base.

3.6.4. Use the formula for the volume of cylindrical "hoof" and evaluate the integral $\int_0^1 y\sqrt{1-y^2}\,dy$.

* * *

In Chapters 2 and 3 we considered the concepts of the derivative and the integral, some of their simpler properties, and the relationship between the integral and the derivative; in other words, we introduced the framework in which the concepts and theorems of the so-called higher mathematics must be developed. The techniques involved in practical computation of derivatives and integrals of various functions will be discussed in Chapters 4 and 5. Only the simplest examples were illustrated in Chapters 2 and 3.

A word of warning to the reader: do not get into the habit of measuring the significance (and even the difficulty) of any section of mathematics by the number of formulas, their complexity, and unwieldiness. Actually, the most important thing and the most difficult thing is the mathematical statement (or formulation) of a problem in the form of an algebraic equation or an integral or even a differential equation, and not the transformations that follow. That is where our attention should be focused.

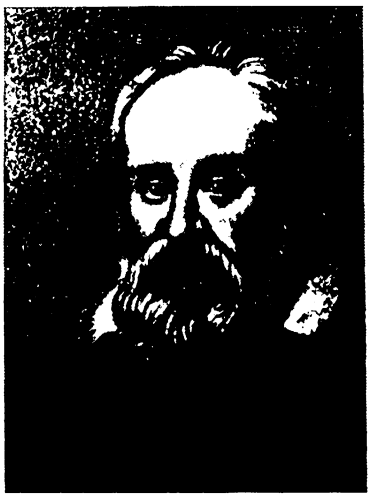
Almost every physicist knows that those pieces of research that he did not succeed in completing (and which other workers did complete) were left undone because he confined himself to a general reflection and did not find the courage to write down the equations and formulate the problem mathematically. The computational difficulties in a properly posed problem with a clear-cut physical content are always surmountable, at least via approximate procedures if not by precise methods.

If the last three sections of this chapter have appeared to be difficult, the reader will do well to reread Chapters 2 and 3.

The rise of "higher mathematics", that is, *differential and integral calculus*, was a turning point in the history of civilization: it gave man a powerful tool for analyzing processes of many different kinds, for developing a fundamental explanation of physical phenomena, and for constructing a scientific picture of the world. Actually, it was the thinkers of ancient Greece, primarily the genius of ***Archimedes*** of Syracuse (287-212 B.C.), who skillfully solved the first problems of differential and integral calculus. Archimedes successfully applied mathematics in designing machines and devices (the weapons of war which he invented struck terror into the Romans laying siege to his native Syracuse). Archimedes knew, for example, how to draw tangents to curves and how to calculate an area bounded by curves. He solved the problem of drawing a tangent to what came to be called the ***Archimedes spiral***, that is, the line described by a snail crawling steadily along the spoke of a wheel turning at a constant rate, and also the problem of ***squaring the parabola***, that is, finding the area of a segment of a parabola. Today problems of this kind can be solved without difficulty by any college student (or even a student of higher school), but in those remote times they were within the powers of only a giant like Archimedes. There was no general method for solving such problems. Each one required great effort.

For that matter, the primitive level of technology in ancient Greece did not require the solution of problems that called for great inventiveness but were devoid of applications outside mathematics: as a rule, the thinkers of ancient times regarded mathematics not as an effective method of solving practical problems but only as a theoretical science whose perfection reflected the profound harmony of the world, yet only explained the world in a purely philosophical sense. (Archimedes was practically the only exception in this respect because he closely linked mathematics with mechanics and physics. But he was a genius; in this, as in many other respects, he was far ahead of his time.)

The static nature of life in ancient Greece, where hardly any machines were known and life centered around city squares, palaces, and temples decorated with beautiful statues as immobile as the temples and city squares themselves, gave rise to the *metaphysical thinking* (rigidity) that was so characteristic of the mathematics of antiquity: it was not customary to regard processes in flux. They limited themselves to unchanging "states," a



Galileo Galilei



Johann Kepler

reflection or which were, for the famous geometer ***Euclid*** of Alexandria (early third century B.C.), merely chains of congruent triangles, which occurred frequently in his arguments.

With the flowering of Italian cities in the 15th and 16th centuries and the appearance of the first factories, which heralded the imminent rise of machine-based production, the static metaphysical thinking of the ancients became inconceivable. In this period it could only hinder much needed scientific progress. The great ***Galileo Galilei*** (1564-1642) was the first to proclaim publicly that mathematics provided the key to the mysteries of the Universe (see footnote 3.15). Under Galileo's influence, his pupils—Evangelista ***Torricelli*** (1608-1647) who discovered the principle of the barometer, and the geometrician Bonaventura ***Cavalieri*** (1598-1647) who continued the work of Archimedes, for whom their teacher had such great affection—solved many specific problems which today lie within the province of higher mathematics. For one, Cavalieri evolved a method of calculating volumes that is very similar to those described above. Galileo was the first to see that the problem *of determining the path of an object from its velocity* practically coincides with the problem that had so interested Archimedes, that of *determining the area of curved figures.* Continuing Galileo's work, Torricelli established that the inverse problem, that of *finding the velocity from the path*, was akin to the problem of *drawing a tangent to a curve.* But at that time there did not yet exist general methods of solving such problems; there was no common algorithm to enable one "to calculate without thinking."

Nor did Johann ***Kepler*** (1571-1630), that outstanding astronomer and mathematician whose discoveries played such a big role in building a scientific picture of the world, have any such method. Kepler was indisputably the leading master of integration in his day (though it was not called that yet). In 1614, when Kepler married a second time, he had occasion to buy a good deal of wine for the wedding reception, and he saw how difficult it was to estimate the cubical contents of wine casks of different shapes from the base radius and the height. The problem intrigued him. The following year, 1615, he published his *Nova Stereometria Doliorum Vinariorum* (The New Science of Measuring Volumes of Wine Casks), "with addenda to Archimedean stereometry," as he also informed the reader. Here he collected a large number of problems in determining the volume of bodies limited by curved surfaces and displayed great ingenuity in developing formulas for such volumes, which are today established by integral calculus (see Section 7.10).

René ***Descartes*** (1596-1650) was truly the forerunner of higher mathematics. Soldier, diplomat, natural scientist, and abstract thinker, he was the father of *analytic geometry* (the method of coordinates in geometry; see Chapter 1) and a profound philosopher who declared that the world was knowable. He stated the basic principles of dialectics and the place occupied in life by various processes whose study, he believed, constituted the chief aim of mathematics. Descartes substantially improved the symbols and language of mathematics, giving them their present-day appearance. This greatly stimulated further progress and democratization of mathematical knowledge.

Another French scientist, Pierre ***Fermat*** (1601-1665), a lawyer by profession and a pro-



René Descartes



Pierre Fermat

found amateur mathematician, was a contemporary and, in a way, a rival of Descartes. Working independently of Descartes, Fermat somewhat earlier elaborated (though it was published later) a system of using, in geometry, the methods of coordinates and algebraic computations now known as analytic geometry. Fermat's exposition of the subject was perhaps closer to the modern one than that of Descartes; and the fact that the writings of Descartes have come to hold a much more significant place in the history of science is connected primarily with his active teaching and his more advanced system of notation. (The situation was much the same in the development of differential and integral calculus.) Simultaneously, Fermat occupied himself with problems that are now part of mathematical analysis. He evolved the concept of the *differential* (see Section 4.1 below) and the overall idea that the *maxima and minima of a (smooth) function must lie at the points where the (first) derivative of the function vanishes*, and also carried out ingenious calculations of the values of some integrals (e.g., the method of deducing formula (5.2.1) indicated in Exercise 3.2.3). Descartes, too, made contributions to the field of analysis.

Christian ***Huygens*** (1629-1695) of the Netherlands (a junior contemporary of Descartes and Fermat) created the wave theory of light and was noted for his work in the application of mathematics to mechanics and physics (e.g., he produced the strict mathematical theory of pendulum clocks (see Section 7.9)). In his investigations, however, Huygens used the archaic methods of thinkers of antiquity, Archimedes for one, because he believed that a newer method would yield no advantages inasmuch as he could solve any problem "in the old way." (True, but you had to have the brains of Huygens to do that.)

Two outstanding thinkers of the 17th century, the Englishman Sir Isaac ***Newton*** (1643-1727) and the German Gottfried ***Leibniz*** (1646-1716), are rightly regarded as the real founders of higher mathematics. Both indisputably rank among the most profound scientists that the world has ever known. To them we owe a coherent exposition of the new calculus, a chain of formulas for finding the derivative of any given algebraic function without any difficulty and a complete understanding of the connection between the derivative and the integral and the significance of that connection, which provides a general algorithm for calculating integrals by turning to a list of derivatives (see Chapters 4 and 5).

Leibniz was a truly versatile scientist: he delved into philosophy, philology, history, psychology (in which he was one of the pioneers of penetration into the sphere of the subconscious), biology (he was one of the precursors of the theory of evolution), geology, mining, mathematics, and mechanics (he evolved the concept of "living force," that is, kinetic energy). At the same time he was active in politics and diplomacy (he strove to reconcile the German principalities because he foresaw their eventual union into a single state, and dreamed of uniting the Catholic and Protestant churches). He was also an organizer of scientific academies; for one, he was the founder and first president of the Prussian Academy of Sciences in Berlin, and he had several talks with Peter the Great of Russia about founding a Russian academy of sciences, which was later established in full conformity with his suggestions and wishes. The mathematical



Christian Huygens



Isaac Newton

analysis of Leibniz [to whom we owe, among other things, the modern terms "derivative" (*Ableitung* in German) and "integral,"[3.19]] was of a form much like that adopted in our book. Leibniz designated the (first) derivative of a function $y = f(x)$ by the symbol dy/dx; he understood the "differentials" dy and dx as the "limiting" values of the increments Δy and Δx at which we arrive by an unbounded reduction of Δx (neither the word "limit" nor its concept existed in the mathematics of Leibniz). This approach corresponded to introducing the concept of the derivative as the tangent of the angle formed by the tangent line to the graph of the function and the x axis: to a small Δx there corresponds a small triangle MNP (see Figure 7.9.1), where $\tan \alpha_1 = NP/PM = \Delta y/\Delta x$; when Δx is reduced without limit, the increments Δx and Δy are replaced by the differentials dx and dy, and the secant MN of length ds (the differential of the arc length; see Section 7.9) is replaced by the appropriate tangent line (Leibniz called such a triangle the *characteristic* triangle).

Leibniz emphasized in every possible way the algorithmic aspect of the new calculus and the system of rules that automatically guarantee a correct result when seeking derivatives. He also worked out a method of handling differentials (dealt with here in Chapter 4); use of this method provides a recipe for calculating derivatives. Leibniz used the modern symbol $\int f(x)\,dx$ to designate the integral of the function $y = f(x)$.

Leibniz was a tireless teacher, and that plus his felicitous system of notation and terms resulted in the universal acceptance of higher mathematics in the form he developed it. Leibniz is also considered a classic in the field of philosophy (here he did much to extend and to amplify the work done by Descartes). He also deserves credit for many profound ideas that partially became reality only in the 19th and 20th centuries: e.g., the idea of a "geometrical calculus," from which the modern vector calculus later arose; his attempts to "algorithmize thinking," in which, as Leibniz wrote, the two sides in a controversy would no longer have to conduct lengthy debates inasmuch as one of them could always say to the other, "Well, let us verify which of us is right, let us calculate, my dear sir." The rough outlines of this kind of "propositional calculus," found in the papers of Leibniz, closely resemble the mathematical logic of the 19th and 20th centuries.

Newton arrived at the same ideas as Leibniz quite independently and even somewhat earlier. He was perhaps more of a physicist and astronomer than a mathematician; his contributions to the birth of physics and to the rise of a new method in natural science cannot be exaggerated. Joseph Louis ***Lagrange*** (1736-1813), the distinguished 18th and 19th century mathematician and investigator in the field of mechanics (we will discuss his work later on) once said of Newton: "He is the most fortunate of men, for the system of the world can be discovered only once." The

[3.19] At first Leibniz simply spoke of the sum (and also "summational calculus" instead of "integral calculus"); he enthusiastically adopted the terms *integral* (from the Latin *integer*, meaning whole) and *integral calculus* proposed by his pupils, the brothers Jakob and Johann Bernoulli.



Gottfried Leibniz

same idea is conveyed in the famous "Epitaph for Sir Isaac Newton" written by his contemporary Alexander ***Pope*** (1688-1744):

> Nature, and Nature's laws lay hid in night: God said, *Let Newton be*! and all was light.

Newton, of course, identified the derivative with velocity: he regarded its properties as the physical properties of velocity. Yet the formal mathematical theory of derivatives (and also integrals), founded on a variation of the theory of limits, was not alien to him either. However, in the absence of a definition of the term limit, the theory did not make his constructions more flawless than the (logically not indisputable) manipulations of Leibniz with infinitesimals, but only made them more unwieldy; also the fact that Leibniz clearly exerted a greater influence on European mathematics than Newton may be connected with this.

Newton gave the name of *fluxion* to the derivative, and the initial function from which the derivative was calculated was called the *fluent* (from the Latin *fluere*, to flow), emphasizing that the quantities under consideration were variable; the fluxion arose as the rate of change of the fluent, while the fluent was restored from the fluxion as the path from the speed. Newton began his exposition of the analysis with two basic problems, to which all the others are reduced.

1. Knowing the length of the path traversed, find the speed of motion over a fixed time interval.

2. Knowing the speed of motion, find the length of the path traversed over a fixed time interval.

These are obviously the problems of finding the derivative from the given function, and of finding the function from its derivative, that is, the problem of calculating an indefinite integral. The fact that these two problems (in their geometrical presentation, the drawing of a tangent to a given curve and the calculation of the area bounded by the curve) are inverses of each other had evidently been discovered first by Newton's teacher, Isaac ***Barrow*** (1630-1677), who subsequently resigned his post as Lucasian Professor of Mathematics at Cambridge (a rare case in the history of science!) in favor of his brilliant pupil because he felt that Newton was more worthy of the post than he. However, it is not at all by accident that the connection between derivatives and integrals is named after Newton and Leibniz instead of after Barrow, for only these two great scientists realized the full extent of the mutually inverse nature of the operations of differentiation and integration, which had been discovered by Barrow (and also, independently, by Leibniz), and the possibility of making this fact the foundation for a broad calculation of derivatives and integrals. It is to the far-sightedness of Newton and Leibniz that mathematical analysis owes the rapid progress that began immediately after the first publications and statements by these two great scientists and which truly marked the dawn of a new era, the era of great scientific discoveries, the era of a sweeping assault on Nature's secrets to promote human welfare.

Priority in applying differential and integral calculus to fathom Nature's secrets undoubtedly goes to Isaac Newton, who put forward the general idea that the laws of Nature must have the form of *differential equations* linking the functions that describe the phenomenon under study. We owe to Newton the formulation of the basic equations of motion (for more details see Chapters 9 and 10), which he then applied to deduce the law of gravitation and to study the motion of celestial bodies and the fairly complicated problems of celestial mechanics (see Lagrange's above-cited assessment of Newton). In effect, Newton's work gave birth to a theoretical physics based on the use of a profound mathematical apparatus, which physics, in its substantive part, has gone far beyond the limits of the problems which Newton grappled with and, in its mathematical part, rests on the whole body of modern science, many times surpassing the differential and integral calculus created by Newton and Leibniz.

Newton designated a fluxion (derivative) by a dot placed above the letter symbolizing the initial function. For example, he wrote the derivative of the function $y = f(x)$ as $\dot{y}$. (This notation is retained today in mechanics, but we will not use it.) Newton proposed designating the operation of a transition from a fluxion to a fluent by a dot placed underneath:

thus, if $y = x^2$ and $y_1 = 2x$, then, in this system of notation, $\dot{y} = y_1$ and $\dot{y}_1 = y$. The symmetry of these designations makes them attractive. The theorem of the connection between the derivative and the integral can be written as follows: $\dot{y} = y$, where the expression $\dot{y}$ can be understood in two ways: as the fluxion of the function y and as the fluent of the function $\dot{y}$. But today this appealing notation has only historical significance (besides, Newton himself was not particularly consistent in using it). The designation y' for the derivative of a function was introduced by the French mathematician Augustin ***Cauchy*** (1789-1857), whom we will be hearing more about later.

One of the regrettable episodes in the history of science was the bitter controversy between Newton and Leibniz, conducted chiefly by their admirers and pupils and not by the two outstanding scientists themselves.[3.20] It was a controversy that brought neither honour nor advantage to either side. Leibniz, accused (quite groundlessly, as we know today) of directly borrowing his ideas from Newton, lost the patronage of the dukes of Hanover, who had long supported him, and died in poverty and oblivion. His death was not reported in a single German (to say nothing of an English) publication; and even the Berlin (Prussian) Academy, which he founded, took no notice of his death. Only the French Academy, of which Leibniz had been an active member, paid tribute to him in a eulogy. On the other hand, the refusal to recognize Leibniz (and also the refusal to recognize his differential and integral calculus) substantially hindered the progress of English science and completely separated English pure and applied mathematics from continental mathematics: English university graduates were not familiar with the terms "derivative" and "integral" or with the notation introduced by Leibniz, and hence were unable to read books and treatises by German or French scientists. It seems to us, however, that the disagreement between Newton and Leibniz was not accidental and deserves a more detailed examination of its causes.

The fact that Newton and Leibniz simultaneously and, beyond dispute, independently discovered differential and integral calculus best of all demonstrates the timeliness and inevitability of the great scientific revolution of the 17th century. Moreover, the different (even opposite) psychic make-up and scientific programs of these two outstanding scientists, which shaped the pathways by which they arrived at their discoveries, make a comparison of their investigations highly edifying.[3.21] Leibniz, the philosopher, was largely guided by what he called the "metaphysics of infinitesimals" (differentials; incidentally, this concept, which came from Fermat, appeared in Newton's works under another name before it did in those of Leibniz). Leibniz was carried onwards by the language he had created, by the developed symbols and terminology characteristic of the new calculus. (We remind our readers once again of Leibniz's great interest in the science of language: he was one of the founders of what is now known as comparative-historical linguistics.)

Moreover, the definition (later included in all textbooks) of infinitesimals as variables whose limit is equal to zero was undoubtedly alien to Leibniz. It corresponded much more closely to the scientific thinking of Newton. Leibniz regarded infinitesimals as "special numbers," the rules for using which completely differed from those governing operations car-

[3.20] The Royal Society of London appointed a special committee to discuss the controversy over priority (the committee sided with Newton). This committee, naturally, did not include scientists whose works it discussed. More than that, the first phrase of the committee's report (published under the title Commercium Epistolicium in 1712) stated that only the absence of the scientists involved in the controversy could ensure impartiality. But, alas, a copy of the draft of the report written in the hand so well known to historians of science shows with all certainty that the report (including the first phrase) was written completely by the then President of the Royal Society, Sir Isaac Newton. On the other hand, the behavior of Leibniz in the controversy with Newton can also hardly be considered irreproachable.

[3.21] These differences in approach and viewpoint make, we believe, the whole controversy over priority quite meaningless, since now it can be said that the inventions of Newton and Leibniz were entirely different (e.g., see A.R. Hall, *Philosophers at War: The Quarrel Between Newton and Leibniz*, Cambridge University Press, Cambridge, 1980).

Note that analytic geometry was created simultaneously and independently by Descartes and Fermat; moreover, these two scientists also belonged to different psychological types. In contrast to Descartes, who thought largely in terms of physics (and also geometry), Fermat can be regarded as an algorithmist similar to Leibniz. This is reflected in his far more systematic treatment of the subject of analytic geometry. (The "algorithmicity" of Fermat's thinking was most fully manifested in his outstanding achievements in the theory of numbers, a theory that did not interest Descartes in the least; incidentally, Fermat had a most profound grasp of physics—he discovered a remarkable principle that the path of a ray of light from one point to another, through one or more media, is such that the time taken is a minimum. Fermat's principle played an outstanding role in the further development of physics.

ried out with ordinary (real) numbers.[3.22] Newton, the physicist, however, proceeded from the substantive meaning of the new concepts and their role in natural science: he regarded the derivative as speed, and the reason why he did not carry through to the end the theory of limits which he had set forth in outline was because he saw no particular need for it. In his view, the physical meaning of all the concepts that he introduced fully justified them, and he felt no need for formal mathematical theories here.

We can assume that the profound mutual antipathy that arose between Newton and Leibniz occurred at the time of their first (and, apparently, only) meeting. This was when Leibniz came to England to demonstrate his version of a "mathematical" (to be more exact, arithmetical) machine [the first to put forward this idea was the outstanding 17th-century French mathematician and physicist Blaise ***Pascal*** (1623-1662)[3.23]: the very idea of such a machine ran counter to Newton's type of intellect]. Still more remote from Newton was Leibniz's dream (and prevision) of "logical machines" (actually, the prototypes of today's computers) for mechanizing mental processes. On the other hand, Leibniz completely rejected Newtonian mechanics (although he had an exceptionally high opinion of Newton as a mathematician), because the very idea of action-at-a-distance (Newtonian gravitation) contradicted his general scientific and religious views.

The subsequent evolution of the scientific ideas of Newton and Leibniz in European science is not without interest. Newton emerged the indisputable victor in the debate about priority, but the outward forms of differential and integral calculus have come down to us entirely from Leibniz. As for scientific ideology in the broad sense of the word, this was influenced for many centuries to a far greater degree by Newton than by Leibniz. This is illustrated not only by the above-cited lines from Pope and Lagrange. The entire history of European science in the 18th, 19th, and first half of the 20th centuries was inspired by Newtonian mechanics as a true model of a scientific theory in the finest meaning of the term. And our book, too, owes much more to the views of Newton than to the world-outlook of Leibniz.

But in the second half of the present century Leibniz's dream of "thinking machines" has suddenly come true. What is more, today's "computerization of knowledge" has once again put the algorithmic ideas of Leibniz in the forefront of scientific thinking, not to mention the significance of diverse particular achievements of the great German, from the elements of *mathematical logic* he created to his serious interest in *combinatorics*, something quite unusual for the 17th century.

Summing up, we can say that while Newton and Leibniz approached higher mathematics from different starting points, the subsequent history of science has confirmed the unquestionable value of *both* avenues of thought which they represent, the value of both Newton's approaches and the scientific ideas of Leibniz.

[3.22] Interestingly, these ideas of Leibniz received a fully logical substantiation only in 1960 in what is known as nonstandard analysis introduced by the American mathematician Abraham ***Robinson*** (1918-1974), which has already found definite application.

[3.23] In some respects Leibniz's machine surpassed Pascal's arithmetical machine: it not only could perform arithmetic operations on numbers but could, for instance, extract square roots.

Chapter 4 Calculation of Derivatives

Convenient and pictorial modes of notation for various relationships and simple rules that permit carrying out computations mechanically without errors are of great significance both for teaching and for the development of mathematics as such. The place occupied by Descartes and Leibniz in the history of mathematics is due not only to the mathematical discoveries they made but also to the new notation developed by these scholars; in particular, the language of the calculus of differentials developed by Leibniz.

In Chapter 2 we analyzed the meaning of the derivative concept and gave simple examples of how to calculate derivatives. In Chapter 4 we present the general rules for finding the derivatives of various functions: polynomials, rational functions involving ratios of polynomials, algebraic functions involving the unknown under the radical sign (fractional powers), the exponential function, the logarithmic function, trigonometric functions, and so forth. We will find the general rules for the derivatives of various combinations of function whose derivatives are known: a sum of functions, a product of functions, a ratio of functions, a composite function. A table of derivatives of a number of functions that summarizes the work carried out in Sections 4.1 to 4.13 is given in Appendix 1 at the end of the book.

4.1 The Differential

From the definition of a derivative we have the following rule (algorithm) for calculating the derivative: specify a value of x and its increment Δx, find $f(x)$ and $f(x + \Delta x)$, find the increment $\Delta f = f(x + \Delta x) - f(x)$, form the ratio $\Delta f / \Delta x$, and then pass to the limit as $\Delta x \to 0$. However, the formula which yields the general expression for $\Delta f / \Delta x$ for arbitrary Δx (not tending to zero) is, as a rule, more complicated than the formula for the *limit*, $\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \frac{df}{dx}$, that is, for the derivative.

For this reason we will frequently write formulas which are only valid in the limit, when the increments tend to zero. For small Δx these formulas will be "almost" valid, that is, exact equalities are replaced with approximate ones, and for large Δx they may prove to be totally incorrect. To emphasize this we will use the symbols dy and dx instead of Δy and Δx. We have to work out rules for handling the quantities dy and dx, which are called ***differentials***,[4.1] so that the basic equation

$$\frac{dy}{dx} = y'(x) \qquad (4.1.1)$$

holds true.

Note that this equation was already discussed in Chapter 2. There, however, it was not a true equation but only a symbolic way of writing the derivative; it meant that the notation $y'(x)$ and the notation dy/dx coincide. Now, however, we wish to consider (4.1.1) as a meaningful equation that connects the derivative $y'(x)$ and the differentials dy and dx of the function y and the independent variable x, namely, the derivative $y'(x)$ must be equal to the ratio dy/dx of the differentials.

Before, we wrote an approximate expression for the increment of the function:

$$y(x + \Delta x) - y(x) = \Delta y \simeq y' \Delta x. \qquad (4.1.2)$$

We can say that Eq. (4.1.2) becomes exact in the limit, as $\Delta x \to 0$. Here the words "becomes exact in the limit" do not mean, of course, only that at $\Delta x = 0$ the left- and right-hand sides of (4.1.2) coincide (are equal to zero)—they stress that for very small

[4.1] In Latin this word means "difference," or increment. In this way, Δx and Δy are increments, while dx and dy are differentials, that is, "almost" increments.

values of Δx the left- and right-hand sides of (4.1.2) are "almost" equal in the sense that the difference between them is much smaller than the members themselves. Indeed, the difference between the left- and right-hand sides of (4.1.2) for small Δx is of the order of $(\Delta x)^2$ or even higher (see the end of Section 2.4), so that it is indeed (for $y' \neq 0$, of course) much smaller than Δy and $y'\Delta x$, which are of the first order of smallness: even if we divide both sides of (4.1.2) by a very small Δx, we will nevertheless arrive at an "almost" exact equality. Let us now agree to replace (4.1.2) with the *exact* equation

$$dy = y'(x)\,dx. \qquad (4.1.2a)$$

Let us explain this further. We assume that the *differential of the independent variable*, dx, is simply the increment Δx (by definition). Then we define the *differential* dy *of the function* $y = y(x)$ by the exact equation (4.1.2a),[4.2] which can be augmented in the following manner:

$$dy = y'(x)\,dx = y'(x)\,\Delta x. \qquad (4.1.2b)$$

Thus, the differential dy of a function $y = y(x)$ differs from the increment Δy, since for Δy the equality (4.1.2) related to (4.1.2b) is satisfied approximately, $\Delta y \simeq y'(x)\,\Delta x$, while the exact equality for $\Delta y = f(x + \Delta x) - f(x)$ is of the form

$$\Delta y = y'(x)\,\Delta x + \sigma, \qquad (4.1.3)$$

where σ is a quantity of a higher order of smallness than Δx. It is this fact that we have in mind when we speak of (4.1.2), or $\Delta y/\Delta x \simeq y'(x)$, as being "exact in the limit": the ratio $\Delta y/\Delta x$ generally differs from $y'(x)$, but $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = y'(x)$, since for small values of Δx we can ignore the term σ on the right-hand side of (4.1.3). However, the words about the limit as $\Delta x \to 0$ and about ignoring a term in an equation become redundant if we replace the language of increments with the language of differentials; as a result Eqs. (4.1.1), (4.1.2a), and (4.1.2b) are common (exact) equalities.

All this has been discussed in great detail in Section 2.4 in connection with the question of approximately calculating the values of functions (p. 77 on). When we calculate, say, the values of the function $y = x^2$ or $y_1 = x^3$, we are justified to assume that if Δx is small, then $\Delta y \simeq 2x\,\Delta x$ or $\Delta y_1 \simeq 3x^2\,\Delta x$, since the error introduced by such an assumption is negligible. For instance, the approximate equality $\Delta y \simeq 2x\,\Delta x$ means that the increment Δy of the area ($y = x^2$) of the square $ABCD$ with a side $AB = x$ (Figure 4.1.1) caused by the increment of the length of the side, Δx, can be replaced with the sum of the areas of the two rectangles BB_1c_1C and DD_1c_2C, ignoring the area of the very small square $Cc_1C_1c_2$ equal to $(\Delta x)^2$ if $BB_1 = \Delta x$ is small. If we examine the function $y_1 = x^3$ and put, say, $x = 2$ and $y_1(x) = 8$, we can construct the following table of increments $\Delta y_1 = y_1(x + \Delta x) - y_1(x)$ and differentials $dy_1 = y_1'(x)\,\Delta x$ for this function and the errors $\sigma/\Delta y_1 = (\Delta y_1 - dy_1)/\Delta y_1$ introduced by the approximate formula (4.1.2):

Δx	1	0.5	0.2	0.1
Δy_1	19	7.625	2.648	1.261
dy_1	12	6	2.4	1.2
$\sigma/\Delta y_1$	37%	21%	9%	5%

Δx	0.05	0.01	0.001
Δy_1	0.615	0.1206	0.012006
dy_1	0.6	0.12	0.012
$\sigma/\Delta y_1$	2.5%	0.5%	0.05%

This table vividly demonstrates the validity of the statement that formula (4.1.2) is exact in the limit.

More information concerning the remainder σ in the approximate formula (4.1.2) can be found in Chapter 6. There we will prove that σ can be represented in the form of a series $a(\Delta x)^2 + b(\Delta x)^3 + \ldots$, whose coefficients are independent of Δx. But for small increments Δx all the powers, $(\Delta x)^2$, $(\Delta x)^3$, etc., are much smaller than Δx, whereby $|\sigma| = |a(\Delta x)^2 + b(\Delta x)^3 + \ldots|$ is much smaller than the differential $|dy| =$

[4.2] Note that while the approximate equation (4.1.2) is a *theorem*, that is, a statement that requires proof (i.e. we need to prove that σ on the right-hand side of (4.1.3) is of a higher order of smallness than Δy and Δx), Eqs. (4.1.2a) and (4.1.2b) are simply *definitions* of dy.

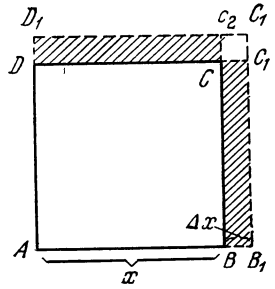


Figure 4.1.1

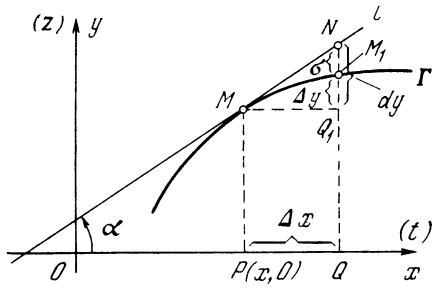


Figure 4.1.2

$|\, y'(x)\, \Delta x \,|$ (which is proportional to $|\, \Delta x \,|$). Of course, the last statement becomes invalid when $y'(x) = 0$, that is, when the differential dy vanishes and hence no small positive $|\, \sigma \,|$ can be smaller. However, even in this case we can employ (4.1.2) and ignore the small error σ. We advise the reader to ignore, for the time being, cases where a function $y = y(x)$ has no derivative (and differential) since these are really exotic (e.g. see Section 7.2).

The idea of a differential and the notation dy/dx for a derivative were introduced by Leibniz, who did not give much thought to the content of the idea of a differential, since he was more interested in the rules of operation on differentials, rules that constitute the essence of Leibniz's calculus of differentials. However, an exact meaning can easily be attributed to the idea of a differential; moreover, this idea can be illustrated by a drawing.

We can rewrite Eq. (4.1.3) thus:

$$\Delta y = dy + \sigma. \qquad (4.1.3a)$$

This leads us to the following statement: *the differential of a function is the principal linear part of the increment of the function.* In other words, the differential $dy = y'(x)\, \Delta x\, (= y'(x)\, dx)$ is a *linear* function of the increment Δx, while the increment Δy generally depends on Δx in a complicated manner (is not a linear function of Δx).[4.3] On the other hand, dy constitutes the principal linear part of the increment Δy of a function in the sense that for Δx small the second term on the right-hand side of (4.1.3a) is much smaller than the first.

The reader will recall that the derivative $y'(x)$ is the tangent of the angle α formed by the tangent line to the curve $y = y(x)$ and the x axis (see Section 2.5). Therefore, if $PQ = \Delta x$ (Figure 4.1.2), then $NQ_1 = \tan \alpha \times \Delta x = y'(x)\, \Delta x = dy$, while the increment of the function, $\Delta y = y(x + \Delta x) - y(x)$ is depicted by segment M_1Q_1. We see that for small values of Δx the differential dy differs little from the increment Δy of the function, while for Δx large this, in general, is not the case. However, the representation of the differential of a function in the form of a graph is needed for a correct understanding of the role that differentials play in approximate calculations, a role that has been described somewhat differently in Section 2.4 (without the use of the word "differential"). In this chapter we need only the technique for handling differentials, the symbolic calculus of differentials; we will nowhere use the exact meaning of the differential illustrated by Figure 4.1.2, and the reader may simply forget about it.

The concept of a differential can be given a simple and graphic mechanical interpretation. Let us suppose that the abscissa x in Figure 4.1.2 is the time variable (it is for this reason we have labeled the x axis by (t)), while the ordinate y (or z) is the distance covered. We are interested in how the distance covered varies with time,

[4.3] The differential $dy = y'(x)\, \Delta x$ of a function depends on *two* numbers (variables), x and Δx, or is a *function of two variables* (it is linear in variable Δx but is generally a complicated function of variable x).

$y = y(x)$, or in the customary notation of Chapters 2 and 3, $z = z(t)$. The idea of a velocity varying constantly under forces is not very simple; therefore, in studying motion in the neighborhood of a fixed moment in time (the position of a moving body at this moment is depicted by point M on the graph of motion), it is convenient to assume that starting from this moment the velocity ceases to change (this assumption is equivalent to the hypothesis that at that moment we "turn off" the forces acting on the body, with the result that subsequent motion is entirely by inertia, that is, with a constant velocity; see Chapter 9). Then, starting with that moment x, the velocity remains constant and equal to instantaneous velocity $v(x) = dy/dx$ at time x (or dz/dt at time t), and the distance covered during time Δx will be equal to $v(x)\,\Delta x = y'\,\Delta x = dy$. (In Figure 4.1.2 the uniform motion of the body corresponds to the straight line l, while the graph of the real (nonuniform) motion corresponds to the curve Γ.) For small Δx (or Δt), this hypothetical path NQ_1 (= dy or dz) differs from the true path M_1Q_1 (= Δy or Δz) by an extremely small quantity $NM_1 = \sigma$ whose order of smallness is higher than that of PQ (= Δx or Δt).[4.4]

The rules for using differentials must be such that the ratio of differentials is equal to the derivative, or (4.1.1) is valid. To achieve this, in formulas we have to drop all terms proportional to $(dx)^2$ and higher powers of dx.

[4.4] It was in this mechanical form that Newton's differential, which he named "fluent" appeared; Newton arrived at this concept earlier than Leibniz, who did not begin to speak about differentials until 1675, whereas Newton's variable quantities (or "fluents," as he preferred to say) were in evidence as early as 1666. Even earlier than that, approximately in the late 1630s, Fermat began to use differentials, but without giving them any special name or sign; it was on these concepts that he based the well-known theorem of maximum and minimum points. However, the credit for the extensive calculus of differentials undoubtedly goes to Leibniz.

Let us consider an elementary example and then compare the increment technique with the technique of differentials. We take the function $y = x^2$. Originally, we did as follows:

$$\Delta y = (x + \Delta x)^2 - x^2 = 2x\,\Delta x + (\Delta x)^2, \qquad (4.1.4)$$

whence

$$\frac{\Delta y}{\Delta x} = 2x + \Delta x \text{ and } y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 2x.$$

Using differentials, we write

$$dy = (x + dx)^2 - x^2 = 2x\,dx, \qquad (4.1.4a)$$

where the term $(dx)^2$ in the right member was dropped immediately.

The meaning of such operations with differentials is quite clear now. Of course, $\Delta y = (x + dx)^2 - x^2 = 2x\,dx + (dx)^2$ (see (4.1.4); we remind the reader that dx and Δx is the same thing). When we go over from Δy to dy, we leave only the principal linear part of Δy, that is, drop all terms that involve Δx (or dx) in second or higher powers and retain the terms that are proportional to the first power of dx. These rules of the calculus of differentials considerably simplify all calculations after one gets accustomed to them, and in subsequent sections we will widely employ them.

Finally, let us discuss a remarkable property of differentials, a property that Leibniz knew and rated highly. Let us take a function, say $y = x^2$. Then

$$y' = 2x \text{ and } dy = 2x\,\Delta x. \qquad (4.1.5)$$

Now suppose that x is not an independent variable but an auxiliary variable dependent on an independent variable t (this can be the time variable). Will (4.1.5) retain its meaning? Of course not, since if $x = t^2$, then the derivative of $y = x^2 = t^4$ is equal to $4t^3$ and not to $2x = 2t^2$. In the second formula in (4.1.5), Δx is now equal to $(t + \Delta t)^2 - t^2 = 2t\Delta t + (\Delta t)^2$, whence

$$2x\Delta x = 2t^2\,[2t\Delta t + (\Delta t)^2] = 4t^3\Delta t + 2t^2\,(\Delta t)^2,$$

which differs from the expression $dy = 4t^3\Delta t$ on the left-hand side of the second

formula in (4.1.5). But the main relationship (4.1.2a) retains its validity in the new conditions: here $dx = 2t\Delta t$, so that $2x\,dx = 2t^2\,(2t\Delta t) = 4t^3\Delta t = dy$.

Is this fact accidental? No, it is not. Formula (4.1.2a) retains its meaning in the case where x is the independent variable and in the case where x is an (intermediate) function of another independent variable, t (cf. Exercises 4.1.1 and 4.1.2).

In view of this it is often more convenient to work with differentials than with derivatives and increments, since there is no need to remember whether x is the independent variable or an auxiliary function. We will see many examples of this fact below.

The simplest way to substantiate the above-formulated property of differentials is to use formula (4.1.3), where σ is a quantity of higher order of smallness than that of dx (and $dy = A\,dx$). In view of this formula we have

$$\Delta y = B\,dt + \alpha, \tag{4.1.6}$$

where $B\,dt = dy$, and α is small. On the other hand,

$$\Delta x = C\,dt + \beta, \tag{4.1.6a}$$

where $C\,dt = dx$, and β is small, and if the derivative dy/dx at the given point (at $t = t_0$ or at $x = x\,(t_0) = x_0$) is equal to D, then

$$\Delta y = D\,\Delta x + \gamma, \tag{4.1.6b}$$

where γ is small. But from (4.1.6a) and (4.1.6b) it follows that $D\,dx$ $(= DC\,\Delta t)$ coincides with the differential of $y = y\,(t)$, since (a) it is linear in Δt and (b) the difference

$$\Delta y - D\,dx = (D\,\Delta x + \gamma) - D\,dx$$
$$= [D\,(dx + \beta) + \gamma] - D\,dx = D\beta + \gamma$$

is a small quantity (a quantity with an order of smallness higher than that of Δt) because β and γ are such quantities and $D\,dx$ is, of course, simply $(dy/dx)\,dx$.

The second derivative $y''\,(x)$ of a function $y = y\,(x)$ was denoted earlier (see Section 2.7) as

$$y''\,(x) = \frac{d^2y}{dx^2}. \tag{4.1.7}$$

Such notation can be justified by a line of reasoning similar to the one used in connection with the notation $y' = dy/dx$.

We call the quantity $y''\,(x)\,(\Delta x)^2$ $(= y''\,(x)\,(dx)^2)$ the ***second differential*** of $y\,(x)$ and denote it by d^2y. In this case the notation (4.1.7) for the second derivative $y''\,(x)$ can be

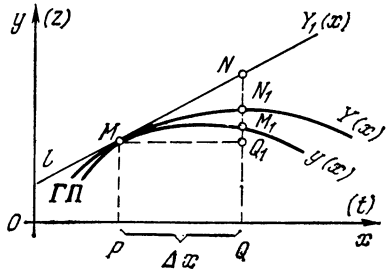


Figure 4.1.3

interpreted as the ratio $d^2y \div (dx)^2$ of the second differential d^2y of y to the square of the (first) differential (increment) of the independent variable x, or $(dx)^2$. The second differential d^2y of the function $y = y\,(x)$ may also be depicted on a graph, which offers possibilities for using the second differential in calculating values of a function. For the time being we will write x_0 instead of x and x instead of $x + \Delta x$. We wish to find the parabola II,

$$Y = a\,(x - x_0)^2 + b\,(x - x_0) + c, \tag{4.1.8}$$

that is the closest to the graph Γ of the function $y = y\,(x)$, that is, a parabola for which the difference $Y\,(x) - y\,(x)$ for small $x - x_0 = \Delta x$ is as small as possible.[4.5] The problem of finding such a parabola (the problem of the "best quadratic approximation" (4.1.8) of $y\,(x)$) coincides with one of the problems we will discuss in Chapter 6; there we will find that the coefficients in the equation of parabola II (4.1.8) are $a = (1/2)\,y''\,(x_0)$, $b = y'\,(x_0)$, and $c = y\,(x_0)$. In this case the difference $Y\,(x) - y_1(x)$ for small Δx is a quantity of third order of smallness (or even higher) with respect to $\Delta x = x - x_0$. If we depict the curve $y = y\,(x)$, the parabola $Y = Y\,(x)$, and the tangent line l,

$$Y_1 = kx + s, \tag{4.1.9}$$

to the curve Γ (where $k = y'\,(x_0)$ and $s = y\,(x_0) - kx_0$) in one drawing (Figure 4.1.3), then the straight line $x =$ constant will intersect Γ, l, and II at the points M_1, N, and N_1, with (see Figure 4.1.3)

$$Q_1M_1 = y\,(x) - y\,(x_0) = \Delta y,\ Q_1N = y'\,(x_0)\,\Delta x = dy,$$

$$Q_1N_1 = \frac{1}{2}\,y''\,(x_0)\,(\Delta x)^2 + y'\,(x_0)\,(\Delta x)$$
$$= \frac{1}{2}\,d^2y + dy,$$

[4.5] This parabola (the ***osculating parabola*** II of curve Γ; cf. Section 7.9) can be described as follows. We draw a parabola Π_1 through three close-lying points (x_0, y_0), (x_1, y_1), and (x_2, y_2) of curve Γ, that is, we select a, b, and c in (4.1.8) in such a manner that $Y\,(x_0) = y_0$, $Y\,(x_1) = y_1$, and $Y_1(x_2) = y_2$. Then, as x_1 and x_2 tend to x_0, the parabola will tend to II (so that II is the limit of Π_1 as x_1 and x_2 tend to x_0).

i.e.

$$Q_1N = dy \quad \text{and} \quad NN_1 = \frac{1}{2}\, d^2y$$

(in the case at hand d^2y is negative).

Figure 4.1.3 provides a persuasive illustration of the advantage of the approximation $Y = y_0 + dy + (1/2)\, d^2y\ (= y(x_0) + y'(x_0)\, \Delta x + (1/2)\, y''(x_0)\, \Delta x^2)$ for the function $y(x)$ over the approximation $Y_1 = y_0 + dy\ (= y(x_0) + y'(x_0)\, \Delta x)$: in the second case point M_1 in Figure 4.1.3 is replaced with point N, while in the first it is replaced with N_1, a point that lies closer to M_1 than N.

The second differential d^2y of a function $y = y(x)$ can be given, just as in the case with the first differential, a certain mechanical interpretation. We again assume that the axis of abscissas in Figure 4.1.3 is the time axis t and the axis of ordinates (the z axis) is the distance axis. Now suppose that starting from a certain time (corresponding to point M of Γ) the body (whose motion we are studying) moves with *uniform acceleration*; this assumption is equivalent to the hypothesis that the forces acting on the body *cease to vary* at that point in time, since a body is uniformly accelerated if the forces on it are *constant* (see Chapter 9). In this case parabola Π becomes the graph of motion and the (hypothetical) path traversed by the body in time Δx (or Δt) will be equal to $dy + (1/2)\, d^2y$ (or $dz + (1/2)\, d^2z$), which in Figure 4.1.3 is depicted by segment Q_1N_1.

Exercises

4.1.1. Prove that $\dfrac{dy}{dt} = \dfrac{dy}{dx}\dfrac{dx}{dt}$ if (a) $y = x^2$ and $x = \sqrt{t}$, and (b) $y = \sqrt{x}$ and $x = t^2$.

4.1.2. Suppose that $y = x^2$, with $x = \sqrt{t+1}$, and $t = u^2$. Prove that

$$dy = \frac{dy}{dx}\, dx = \frac{dy}{dt}\, dt = \frac{dy}{du}\, du\,.$$

4.2 Derivatives of a Sum and of a Product of Functions

As another example of how to use the language of differentials we take the *sum* of two functions, $f(x)$ and $g(x)$, taken with (arbitrary) constant coefficients A and B:

$$y = Af(x) + Bg(x).$$

Using differentials, we can write

$$\begin{aligned} dy &\simeq y(x + dx) - y(x) \\ &= [Af(x + dx) + Bg(x + dx)] \\ &\quad - [Af(x) + Bg(x)] \\ &= A[f(x + dx) - f(x)] \\ &\quad + B[g(x + dx) - g(x)] \\ &\simeq A\,df + B\,dg = Af'dx + Bg'dx, \end{aligned}$$

whence

$$y' = \frac{dy}{dx} = Af' + Bg'. \tag{4.2.1}$$

The reader can easily arrive at this formula if increments and limits are used instead of differentials.

In particular, for the sum and difference of two functions ($A = B = 1$ and $A = 1$, $B = -1$) we have

$$(f + g)' = f' + g', \quad (f - g)' = f' - g'. \tag{4.2.2}$$

Let us now find the derivative of a *product* of two functions, $g(x)$ and $h(x)$. We put $f(x) = g(x)\, h(x)$. Then

$$\begin{aligned} df(x) &\simeq f(x + dx) - f(x) \\ &= g(x + dx)\, h(x + dx) \\ &\quad - g(x)\, h(x). \end{aligned}$$

But

$$g(x + dx) \simeq g(x) + dg,$$
$$h(x + dx) \simeq h(x) + dh.$$

Whence

$$\begin{aligned} df &\simeq [g(x) + dg]\,[h(x) + dh] \\ &\quad - g(x)\, h(x) = g(x)\, dh \\ &\quad + h(x)\, dg + dg\,dh. \end{aligned}$$

Note that $dg = g'(x)\, dx$ and $dh = h'(x)\, dx$, whence $dh\,dg = g'(x)\, h'(x)\, (dx)^2$. The quantity $dh\,dg$ is proportional to $(dx)^2$ and so, according to the rules for handling differentials, we ignore the product $dh\,dg$ in the expression for df. Finally we get

$$df = g(x)\, dh + h(x)\, dg. \tag{4.2.3}$$

Dividing all the terms in (4.2.3) by dx, we get

$$\frac{df}{dx} = \frac{d(gh)}{dx} = g\frac{dh}{dx} + h\frac{dg}{dx}\,. \tag{4.2.4}$$

The expression is remembered in the following manner: the derivative of the product gh is equal to the sum of the derivative taken on the assumption that h is a function of x while g is held constant (the term $g\,(dh/dx)$) and the derivative taken on the assumption that h is held constant and only g depends on x (the term $h\,(dg/dx)$). Here, naturally, the constant value of g in the term $g\,(dh/dx)$ is taken for the x for which the value of the derivative is being sought. The same goes for h in the second term. Below we will see that a similar procedure can be applied in other situations,[4.6] for instance in the case of $y = f(x)^{g(x)}$.

How would we have handled this in the old way? Simple algebra yields the exact equation

$$\Delta f = (g + \Delta g)(h + \Delta h) - gh = g\Delta h + h\Delta g + \Delta h \Delta g.$$

Dividing both sides by Δx, we get

$$\frac{\Delta f}{\Delta x} = g(x)\frac{\Delta h}{\Delta x} + h(x)\frac{\Delta g}{\Delta x} + \frac{\Delta h}{\Delta x}\frac{\Delta g}{\Delta x}\Delta x. \tag{4.2.5}$$

Note that the last term for the sake of convenience we multiplied and divided by Δx.

Up to now all the equations are exact and hold true for all values of Δx. Now pass to the limit as $\Delta x \to 0$. Then

$$\lim_{\Delta x \to 0}\frac{\Delta f}{\Delta x} = f', \quad \lim_{\Delta x \to 0}\frac{\Delta h}{\Delta x} = h',$$

$$\lim_{\Delta x \to 0}\frac{\Delta g}{\Delta x} = g'$$

and, in view of (4.2.5),

$$f' = gh' + g'h. \tag{4.2.6}$$

In passing to the limit, the last term on the right-hand side of (4.2.5) vanished since the first two factors yield in the limit the product $h'g'$ and we sent Δx to zero.

Using increments and passage to the limit, we obtain the same result as that obtained with the aid of differentials, but it takes more time. This is not surprising since in the case of differentials we dropped $dhdg$ mechanically, on the basis of an earlier acquired rule according to which we have to reject terms involving $(dx)^2$, $(dx)^3$, etc., hence, any products of two, three, or more differentials. When carrying out the computations with the aid of increments, we actually, in the very process, *proved* this rule once again for the case of a product of functions.

The succession of operations using increments is needed to justify the rules and to understand them. But once they are understood, the use of differentials is faster and more efficient. It would be silly every time to start from rock bottom in solving a specific problem and to write out in full that the derivative is the limit of a ratio, and so on.

Example. Suppose that $f(x) = (3x^2 + 5)(2x - 4)$. Find $f'(x)$ and, in particular, $f'(0)$.

Here

$$g = 3x^2 + 5, \quad \frac{dg}{dx} = 3 \times 2x + 0 = 6x;$$

$$h = 2x - 4, \quad \frac{dh}{dx} = 2 - 0 = 2.$$

Therefore

$$\frac{df}{dx} = (3x^2 + 5)\,2 + (2x - 4)\,6x = 18x^{2x} - 24x + 10,$$

and in particular

$$f'(0) = \left.\frac{df}{dx}\right|_{x=0} = 10.$$

The rule for finding the derivative of a product generalizes to the case of *many* factors. For example, for the product of four functions $f(x)$, $g(x)$, $h(x)$, and $k(x)$ we get

$$\frac{d(fghk)}{dx} = fgh\frac{dk}{dx} + fgk\frac{dh}{dx} + fhk\frac{dg}{dx} + ghk\frac{df}{dx}. \tag{4.2.7}$$

[4.6] It is clear that the formula for the derivative of the sum of functions (if $F = f + g$, then $F' = f' + g'$) also obeys this rule, since if f is assumed constant, the derivative F' is simply g', while if g is constant, the derivative F' is f'.

Thus, here we once more encounter the same rule, according to which the derivative of a product of several functions can be found as the sum of the derivatives computed on the assumption that each time only one function varies while the other remains constant.

The formulas we have obtained enable making further conclusions. It is clear that if $f(x) = g(x) + h(x)$ and $f'(x) = g'(x) + h'(x)$, then

$$f''(x) = (f'(x))' = (g'(x) + h'(x))' = g''(x) + h''(x). \quad (4.2.8)$$

The formula for the second derivative of a *product* of functions is somewhat more complicated: if $f(x) = g(x)h(x)$, then $f'(x) = g(x)h'(x) + g'(x)h(x)$ and $f''(x) = (g(x)h'(x) + g'(x)h(x))' = (g(x)h''(x) + g'(x)h'(x)) + (g'(x) \times h'(x) + g''(x)h(x))$, that is,

$$f''(x) = g(x)h''(x) + 2g'(x)h'(x) + g''(x)h(x). \quad (4.2.9)$$

Exercises

4.2.1. Use the rules found above to calculate the derivatives of the following functions: (a) $y = x^4 (= x^2x^2)$, (b) $y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$, and (c) $y = \sqrt{x}$, using the identity $\sqrt{x}\sqrt{x} = x$. [*Hint.* Use the following equality: $y'y + yy' = x' = 1$.]

4.2.2. Express the third derivative $f'''(x) = (f''(x))'$ of the function $f(x) = g(x)h(x)$ in terms of the functions g and h and their derivatives.

4.2.3. Suppose that $f(x) = g(x)h(x)k(x)$. Find $f''(x)$.

4.3 The Composite Function. The Derivative of the Fraction of Two Functions

Let z be given as a function of y, say $z = 1/y$, and y as a function of x, say $y = x^2 + 5$. It is clear that to each x there corresponds a definite y and since to each y there corresponds a definite z, we see that each x is associated with a definite z, that is, z is a function of x. It is always possible, by substituting the expression of y in terms of x, to write directly formulas for $z(x)$; in the given example, $z = (x^2 + 5)^{-1}$.

Finding derivatives is simplified if we reduce all functions to combinations of the simplest possible functions: separately, each of the functions $z = 1/y$ and $y = x^2 + 5$ is simpler than $z = (x^2 + 5)^{-1}$. By reducing complicated functions to combinations of simpler functions, we are able to get by with the rules for finding the derivatives of these simple functions.

Let us find the differential of the composite function $z[y(x)]$. Regarding z as a function of y, we write

$$dz = \frac{dz}{dy}\,dy \quad (4.3.1)$$

(compare this with what was said in Section 4.1 about the differential of a composite function). But y is a function of x, whereby

$$dy = \frac{dy}{dx}\,dx. \quad (4.3.2)$$

Substituting (4.3.2) into (4.3.1), we obtain

$$dz = \frac{dz}{dy}\frac{dy}{dx}\,dx. \quad (4.3.3)$$

Dividing both sides by dx, we get *a rule* (known as ***the chain rule***) *for determining the derivative of a composite function*:[4.7]

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}. \quad (4.3.3a)$$

The form of the formula is in full accord with what was stated about the possibility of handling differentials as ordinary algebraic quantities: we can cancel out dy in the product $(dz/dy)(dy/dx)$.

Recall that z is given as a function of y, and so dz/dy is also a function of y. But since y itself is a function of x, it follows that by substituting $y = y(x)$ into the expression for dz/dy, we get dz/dy as a function of x and, hence, also dz/dx as a function of x.

[4.7] The prime notation z' instead of dz/dx can lead to confusion. When we write z', it is not clear whether we mean dz/dx or dz/dy. (However, we could have written z'_x or z'_y.)

Let us carry out computations for a case that will be needed later on. Suppose

$$z=\frac{1}{y(x)}\ . \qquad (4.3.4)$$

We know that if $z = 1/y$, then $dz/dy = -1/y^2$ and so

$$dz=-\frac{1}{y^2}dy=-\frac{1}{y^2}\frac{dy}{dx}dx$$

and

$$\frac{dz}{dx}=-\frac{1}{y^2}\frac{dy}{dx}\ . \qquad (4.3.5)$$

For example, if $y = x^2 + 5$ and $z = 1/y = (x^2 + 5)^{-1}$, then

$$\frac{dz}{dx}=-\frac{1}{(x^2+5)^2}\frac{d(x^2+5)}{dx}=-\frac{2x}{(x^2+5)^2}\ .$$

Here is another simple example: $y = x^{3/2} = \sqrt{x^3}$. This formula can be rewritten as $y = \sqrt{u}$, with $u = x^3$. According to (4.3.3a) we have

$$\frac{dy}{dx}=\frac{dy}{du}\frac{du}{dx}=\frac{1}{2\sqrt{u}}\,3x^2=\frac{3x^2}{2\sqrt{x^3}}$$
$$=\frac{3}{2}\sqrt{x}.$$

Earlier we arrived at the same result by a more complicated method (see Exercise 2.4.3).

The chain rule (4.3.3a) for forming the derivative of a composite function holds true for a more complicated relationship as well. Suppose that $z = z(y)$, $y = y(x)$, $x = x(t)$, and $t = t(w)$. Then

$$\frac{dz}{dw}=\frac{dz}{dy}\frac{dy}{dx}\frac{dx}{dt}\frac{dt}{dw}\ . \qquad (4.3.6)$$

Using (4.3.5), we can easily find the derivative of a *fraction* (quotient, ratio) of two functions, $f = h/g$. To do this, we write f in the form of a product: $f = h(1/g)$. Then

$$f'=h\left(\frac{1}{g}\right)'+h'\frac{1}{g}\ . \qquad (4.3.7)$$

But $(1/g)' = -(1/g^2)\,g'$ (see (4.3.5). Substituting this result into (4.3.7), we get the sought result

$$f'=\left(\frac{h}{g}\right)'=-\frac{h}{g^2}g'+h'\frac{1}{g}\ ,$$

or[4.8]

$$\left(\frac{h}{g}\right)'=\frac{h'g-hg'}{g^2}\ . \qquad (4.3.8)$$

Exercises

4.3.1. Find the derivative of $z = (ax + b)^2$ as that of the composite function $z = y^2$, with $y = ax + b$. Remove the parentheses in $(ax + b)^2$ and find the same derivative directly.

4.3.2. Find the derivatives of the following functions:

(a) $z=\dfrac{1}{ax+b}$, (b) $z=\dfrac{1}{(ax+b)^2}$,

(c) $z=\dfrac{1}{1+1/x}$, (d) $y=\dfrac{x^3+5x^2}{x+1}$,

(e) $y=\dfrac{x-1}{x^2+2}$.

4.3.3. Suppose that $y = f(x)/g(x)$. Express the second derivative d^2y/dx^2 of the function y in terms of the functions f and g and their first and second derivatives.

4.4 The Inverse Function. Parametric Representation of a Function

Specifying y as a function of x signifies that to each x there corresponds a definite value of y. Hence, conversely, we can often say that with each definite y there is associated an x. For instance, if $y = 8x^3 - 9$, then $x^3 = (y + 9)/8$ and $x=\sqrt[3]{\frac{y+9}{8}}=\sqrt[3]{\frac{y+9}{2}}$. Thus, the specification of $y(x)$ also yields the functional relation $x(y)$. This relationship is called the ***inverse function***, as we have already said in Section 1.6; on the other hand, $y(x)$ is the inverse of $x(y)$.[4.9]

In many cases the inverse function has a more simple form than the initial function; for instance, the function $y = \sqrt[3]{x-1}$ contains a cube root, while the inverse function $x = y^3 + 1$ is a polynomial, which is simpler

[4.8] It is clear that formula (4.3.8) corresponds to the general principle formulated in Section 4.2: $(h/g)' = h'/g + h(1/g)'$.

[4.9] In the general case the function $x = x(y)$ that is the inverse of the given function $y = y(x)$ is *multiple-valued* (see Sections 1.6 and 4.11).

than y (x). In this case, it is simpler and easier to find the derivative of the inverse function, dx/dy, than it is to find the derivative of the initial function, dy/dx. The problem now is to express the derivative of the initial function in terms of the derivative of the inverse. For y (x) we have

$$dy = \frac{dy}{dx}\,dx = y'(x)\,dx. \qquad (4.4.1)$$

This enables finding *the derivative* x' (y) *of the inverse function* x (y):

$$x'(y) = \frac{dx}{dy} = \frac{1}{y'(x)}. \qquad (4.4.2)$$

(Of course, (4.4.2) once more illustrates that differentials can be treated as numbers, or $dx/dy = 1/(dy/dx)$.) On the right is an expression in the form of a function of x. But if we know $x = x(y)$, this expression can be represented as a function of y, namely, as $1/y'(x(y))$.

A few examples will serve as illustrations of the foregoing. The first example (a linear function) is too simple. We start with the second example:

$$y = x^2, \quad \frac{dy}{dx} = y'(x) = 2x, \qquad (4.4.3)$$
$$\frac{dx}{dy} = \frac{1}{y'(x)} = \frac{1}{2x}.$$

Substituting into (4.4.3) the inverse function $x = \sqrt{y}$, we get

$$\frac{dx}{dy} = \frac{d\sqrt{y}}{dy} = \frac{1}{2x} = \frac{1}{2\sqrt{y}}. \qquad (4.4.4)$$

In Chapter 2 we obtained the same result (the derivative of $y = \sqrt{x}$) in a more roundabout fashion.

Here is a third example:

$$y = x^3 + 1, \quad \frac{dy}{dx} = y'(x) = 3x^2,$$
$$\frac{dx}{dy} = \frac{1}{y'(x)} = \frac{1}{3x^2},$$
$$\frac{dx}{dy} = \frac{d\left(\sqrt[3]{y-1}\right)}{dy} = \frac{1}{3x^2} = \frac{1}{3\sqrt[3]{(y-1)^2}}$$
$$= \frac{1}{3}(y-1)^{-2/3}$$

(cf. Section 4.5).

If a function is represented *parametrically* (see Section 1.8), this representation may be regarded as a special case of a composite function. Indeed, if it is given that

$$x = f(t), \quad y = g(t), \qquad (4.4.5)$$

then the first of these equations may be regarded as an equation whose solution yields $t = t(x)$. Substituting this $t(x)$ into the second equation, we get $y = g(t) = g(t(x))$. Hence,

$$\frac{dy}{dx} = \frac{dy}{dt}\,\frac{dt}{dx}.$$

But to use this formula one need not express t as a function of x (if we were to do this we would get rid of the parameter, but this is not always possible). It suffices to know $x = f(t)$, which is the inverse of $t(x)$. Thus,

$$\frac{dt}{dx} = \frac{1}{\frac{dx}{dt}},$$

and so

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}. \qquad (4.4.6)$$

This formula is yet another instance showing that we can handle differentials like ordinary algebraic quantities: the quantity dt in the right member of (4.4.6) is canceled out.

Here is an *example*. Suppose that

$$x = t^2 - t, \quad y = t^2 + t. \qquad (4.4.7)$$

Then

$$\frac{dx}{dt} = 2t - 1, \quad \frac{dy}{dt} = 2t + 1,$$
$$\frac{dy}{dx} = \frac{2t+1}{2t-1}.$$

In constructing the graph of a function represented parametrically, $x = x(t)$ and $y = y(t)$, we set up a table where we specify the values of x and y corresponding to the chosen values of t, this set of x and y specifying a point $M = M(x, y)$ on the graph of the function. It is also convenient to calculate for a given t the derivatives

dx/dt and dy/dt; then the ratio $(dy/dt)/(dx/dt)$ will fix the slope of the tangent to the graph at point $M(x, y)$.

Formulas (4.4.2) and (4.3.5) also make it possible to arrive at formulas for the second derivative of a function $x = x(y)$ that is the inverse of another function $y = y(x)$. Of course, following the line of reasoning that led us to (4.4.2), we would like to think that $x''(y) = 1/y''(x)$—but this is not so if only because of dimensionality considerations. It is clear that if x and y are expressed in units of e_1 and e_2, the derivatives $y'(x) = dy/dx$ and $x'(y) = dx/dy$ have the dimensions of e_2/e_1 and e_1/e_2, respectively, so that the expressions on both sides in (4.4.2) have the same dimensions. However, the second derivatives $y''(x) = d^2y/dx^2$ and $x''(y) = d^2x/dy^2$ have the dimensions of e_2/e_1^2 and e_1/e_2^2, in view of which $x''(y)$ (whose dimensions are those of e_1/e_2^2) can in no way be equal to $1/y''(x)$ (whose dimensions are e_1^2/e_2).

The true formula for the second derivative of $x(y)$ can be written as follows:

$$\frac{d^2x}{dy^2}=\frac{d}{dy}\left(\frac{dx}{dy}\right)=\frac{d}{dy}\left(\frac{1}{y'(x)}\right)$$

$$=-\frac{\frac{d}{dy}(y'(x))}{(y'(x))^2}=-\frac{\frac{d}{dx}(y'(x))\frac{dx}{dy}}{(y'(x))^2}$$

$$=-\frac{y''(x)\left[1\Big/\left(\frac{dy}{dx}\right)\right]}{(y'(x))^2}$$

$$=-\frac{y''(x)(1/y'(x))}{(y'(x))^2}.$$

Thus, finally,

$$x''(y)=-\frac{y''(x)}{(y'(x))^3}. \qquad (4.4.8)$$

(It is clear that if x and y are measured in units of e_1 and e_2, respectively, then the right-hand side of (4.4.8) has the dimensions of $(e_2/e_1^2)/(e_2/e_1)^3 = e_1/e_2^2$, that is, dimensions that coincide with those of the left-hand side of $x''(y)$.)

For instance, if $y = x^2$, $x = \sqrt{y}$, then $y' = 2x$, $y'' = 2$, and, hence,

$$x''(y)=\frac{d^2(\sqrt{y})}{dy^2}=-\frac{2}{(2x)^3}=-\frac{1}{4x^3}$$

$$=-\frac{1}{4y^{3/2}}.$$

Similarly it is easy to use (4.4.6) and derive a formula for the second derivative of a function $y = y(x)$ specified parametrically, $x = x(t)$ and $y = y(t)$:

$$\frac{d^2y}{dx^2}=\frac{d}{dx}\left(\frac{dy}{dx}\right), \quad \text{where} \quad x=x(t),$$

$$\frac{dy}{dx}=\frac{y'(t)}{x'(t)}.$$

Hence,

$$\frac{d^2y}{dx^2}=\frac{\frac{d}{dt}\left(\frac{y'(t)}{x'(t)}\right)}{x'(t)}=\frac{\frac{y''(t)x'(t)-y'(t)x''(t)}{(x'(t))^2}}{x'(t)},$$

that is,

$$\frac{d^2y}{dx^2}=\frac{y''(t)x'(t)-y'(t)x''(t)}{(x'(t))^3}. \qquad (4.4.9)$$

(If x, y, and t are measured in units of e_1, e_2, and e, then the derivatives dx/dt and dy/dt have the dimensions of e_1/e and e_2/e, so that the fraction on the right-hand side of (4.4.6) has the dimensions of $(e_2/e)/(e_1/e) = e_2/e_1$, that is, dimensions of the right-hand side of (4.4.6). On the other hand, the second derivatives $x''(t)$ and $y''(t)$ have the dimensions of e_1/e^2 and e_2/e^2, so that the products in the numerator of the right-hand side of (4.4.9) all have the same dimensions, those of $(e_2/e^2)(e_1/e) = (e_1/e^2)(e_2/e) = (e_1e_2)/e^3$, with the dimensions of the entire right-hand side being those of $(e_1e_2)/e^3 \div (e_1/e)^3 = e_2/e_1^2$, that is, coinciding with the dimensions of the second derivative d^2y/dx^2.)

For instance, in the case of function (4.4.7) we have

$$\frac{d^2x}{dt^2}=2, \quad \frac{d^2y}{dt^2}=2,$$

$$\frac{d^2y}{dx^2}=\frac{2(2t-1)-(2t+1)2}{(2t-1)^3}$$

$$=-\frac{4}{(2t-1)^3}.$$

Exercises

4.4.1. Find first derivatives of the following function: (a) $y=\sqrt[3]{2x+1}$, (b) $y=\sqrt{\frac{x-1}{x+1}}$, and (c) $y=1/\sqrt[4]{x}$.

4.4.2. $x=a\cos t$, $y=b\sin t$. Find the derivative dy/dx.

4.4.3. Find the second derivative d^2y/dx^2 of the following functions; (a) $y=\sqrt{\frac{x-1}{x+1}}$, (b) $y=1/\sqrt[4]{x}$, and (c) $x=a\cos t$, $y=b\sin t$.

4.5 The Power Function

Let us consider the power function,

$$y = x^n, \qquad (4.5.1)$$

where n is a constant number. For n a positive integer, x^n is the product of n identical factors, $y = \underbrace{x \cdot x \cdot x \ldots x}_{n \text{ factors}}$, and, hence (via formula (4.2.7)),

$$\frac{dy}{dx} = \underbrace{1 \cdot x^{n-1} + 1 \cdot x^{n-1} + \ldots + 1 \cdot x^{n-1}}_{n \text{ terms}},$$

whence

$$\frac{dy}{dx} = nx^{n-1}. \qquad (4.5.2)$$

We will show that this formula holds true for *arbitrary* n, whether fractional or negative. [4.10]

For fractional n we write $n = m/p$, where m and p are integers:

$$y = x^{m/p}, \qquad (4.5.3a)$$

or

$$y^p = x^m. \qquad (4.5.3b)$$

The expression y^p on the left-hand side of (4.5.3b) is a composite function of x, since y depends on x. Therefore, calculating the derivatives of both sides and employing (4.3.3a), we obtain

$$\frac{d}{dx}(y^p) = \frac{d}{dx}(x^m), \text{ or } py^{p-1}\frac{dy}{dx} = mx^{m-1},$$

whence

$$\frac{dy}{dx} = \frac{m}{p}\frac{x^{m-1}}{y^{p-1}} = \frac{m}{p}\frac{x^{m-1}}{(x^{m/p})^{p-1}} = \frac{m}{p}x^{\frac{m}{p}-1}.$$

Noting that $m/p = n$, we finally have

$$\frac{dy}{dx} = nx^{n-1}.$$

For a *negative* exponent we write $n = -k$, where k is a positive integer, so that $y = x^n = x^{-k} = 1/x^k$. Using again (4.3.3a) (to be more precise, (4.3.5); here $y = 1/f$ and $f = x^k$), we find that

$$\frac{dy}{dx} = -\frac{1}{f^2}\frac{df}{dx} = -\frac{1}{x^{2k}}kx^{k-1} = -kx^{-k-1}.$$

Recalling that $k = -n$, we find that for n negative

$$\frac{dy}{dx} = \frac{dx^n}{dx} = nx^{n-1},$$

which coincides with (4.5.2).

Thus, the formula for the derivative of a power is applicable to *any* rational exponent n. It can also be extended to the case of an *irrational* exponent. [4.11]

Formula (4.5.2) is of greatest importance. It implies, for one, that for r small in absolute value we can write

$$(1 + r)^n \simeq 1 + nr. \qquad (4.5.4)$$

Indeed, if $x = 1$, $\Delta x = r$, and $y = x^n$, then $\Delta y = (x + \Delta x)^n - x^n = (1 + r)^n - 1$. On the other hand, $dy = y'(x)\,\Delta x = n \cdot 1^{n-1} \cdot r = nr$. Formula (4.5.4) follows from the fact that (for $\Delta x = r$ small) the increment Δy is close to the differential dy.

It is also expedient to write (4.5.2) in another form:

$$\text{if } y = cx^n, \text{ then } \frac{dy}{dx} = n\frac{y}{x}. \qquad (4.5.5)$$

One has to get the feeling of this result. For positive n, the power function has the obvious property that for $x = 0$, y is also equal to 0. For a given $n > 0$, the curve $y = cx^n$ can be drawn through any point (x_0, y_0), just choose $c = y_0/x_0^n$. Let the curve pass through the origin and through the point (x_0, y_0). We wish to find the average value of the derivative over the section of the curve between the origin and point (x_0, y_0). According to the definition of an average (see Section 7.8),

$$\overline{y'} = \frac{\int_0^{x_0} y'(x)\,dx}{x_0 - 0} = \frac{1}{x_0}\int_0^{x_0} y'(x)\,dx,$$

[4.10] The reader will immediately see that it is valid at $n = 0$, when $y =$ constant and $y' = 0$.

[4.11] The function $y = x^n$, where n is an *irrational* number, is defined by the condition $x^n = \lim x^r$, where r is a rational number and $\lim r = n$, whereby if formula (4.5.2) is valid for all rational exponents n, it will necessarily be valid for irrational exponents.

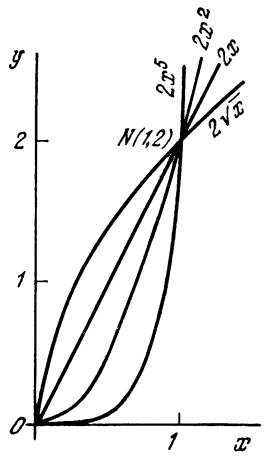


Figure 4.5.1

whence, using (3.4.9), we get

$$\overline{y'} = \frac{y(x_0) - y(0)}{x_0} = \frac{y(x_0)}{x_0} = \frac{y_0}{x_0}.$$

Indeed, as x varies from 0 to x_0, the value of y grows from 0 to y_0. Hence, the average rate of growth of y (i.e. the average derivative) is equal to y_0/x_0, a result obvious without integrals.

As evident from (4.5.5), the value of the derivative at point (x_0, y_0) is n times the average value of the derivative (n is the exponent). Figure 4.5.1 depicts a number of curves with different n ($n = 1/2, 1, 2, 5$) passing through one and the same point N (x_0, y_0) and, hence, having the same average derivative on the interval from 0 to x_0. It is clearly evident that the larger the n, the greater the derivative at point N (the more steeply the curve rises).

Let us return to formula (4.5.5), $dy/dx = n\,(y/x)$, whence $dy = n\,(y/x)dx$, and so for small increments Δx we have

$$\Delta y \simeq n \frac{y}{x} \Delta x. \qquad (4.5.6)$$

For $\Delta x = 0.01x$, that is, for a 1% variation in the independent variable, formula (4.5.6) can be assumed exact. Then

$$\Delta y \simeq n \frac{y}{x} \times 0.01x,$$

or $\Delta y \simeq n \times 0.01y$.
Thus, when the independent variable varies by 1%, the power function (4.5.1) varies by n%.

Exercises

4.5.1. Find the derivatives of the following functions: (a) $y = x^5 - 3x^4 + x^3 + 7x^2 - 2x + 5$, (b) $y = (x^3 + x + 1)^2$, (c) $y = (x^2 - x + 1)^4$, (d) $y = (3x^2 - 1)^{10}$, (e) $y = \sqrt{x^2 - 1}$, and (f) $y = \sqrt[5]{x^2}$.

4.5.2. Obtain formula (4.5.2) for n a positive integer with the aid of the binomial theorem. [*Hint.* Use the binomial theorem (Section 6.4) to calculate $\Delta x = (x + \Delta x)^n - x^n$.]

4.5.3. Find the values of y (9) and y (11) if y (10) = 5 and (a) $y \propto \sqrt{x}$, (b) $y \propto 1/x$, and (c) $y \propto x^2$, where $\propto$ stands for *proportional*. Solve the problem mentally without any computations. Compare the answers with the exact values.

4.6 Derivatives of Algebraic Functions

The collection of rules in Sections 4.1-4.5 enable finding the derivative of any function involving addition, subtraction, multiplication, division, and raising to an (arbitrary) power including fractional powers, i.e. extraction of roots, of the independent variable x.

An example will illustrate how to do this in the best practical fashion. Find the derivative of the function

$$f(x) = x\sqrt[3]{x^2 - 1}.$$

It is best to write the answer at once, that is, without introducing any new designations (such as $\sqrt[3]{x^2 - 1} = y$). The derivative is taken, as it were, separately with respect to each side involving x, and we say roughly the following (the letters "a," "b," "c," "d" show to what portions in the expression of the derivative the words refer): the derivative of (a) with respect to x in front of the radical sign plus (b) the derivative with respect to $\sqrt[3]{x^2 - 1}$ multiplied by (c) the derivative $\sqrt[3]{x^2 - 1}$ with respect to $x^2 - 1$ multiplied by (d) the derivative of $x^2 - 1$ with respect to x:

$$\frac{df}{dx} = \overset{\text{(a)}}{1} \times \sqrt[3]{x^2 - 1} + \overset{\text{(b)}}{x} \times \overset{\text{(c)}}{\frac{1}{3}\frac{\sqrt[3]{x^2 - 1}}{x^2 - 1}} \overset{\text{(d)}}{2x}. \qquad (4.6.1)$$

It is well to get used to this efficient approach (without indulging in a lot of writing) by applying the following principles:

(a) The rule for differentiating a composite function (Section 4.3, formulas (4.3.3a) and (4.3.6)).

(b) If the expression is made up of several functions, then its derivative is equal to the sum of the derivatives computed on the assumption that each time only one of the functions is taken to be variable and the rest are held constant (see Section 4.2, formulas (4.2.4), (4.2.7), and also (4.2.1), (4.2.2), and (4.3.8)).

The formula for the derivative of a power is conveniently used in the form

$$y = cx^n, \quad \frac{dy}{dx} = n\frac{y}{x},$$

as was done above (see portion "c" in (4.6.1)).

To acquire the necessary facility in handling these rules, a good 10 to 20 exercises devoted to the pure techniques (without regard for the physical or geometrical content of the problems involving finding derivatives) are definitely needed.

Exercises

Find the derivatives of the following functions: **4.6.1.** $y = x^3(x^2-1)^2$. **4.6.2.** $y = x^3\sqrt{x^2+x}$. **4.6.3.** $y = x^5\sqrt[3]{x^2-1}\,(x^3-2x)^{1/5}$. **4.6.4.** $y = \left(x+\frac{1}{\sqrt{x}}\right)\sqrt{x^3-2}$. **4.6.5.** $y = x^2\sqrt{\sqrt[3]{x}+x}$. **4.6.6.** $y = \left(\sqrt[3]{x}+\frac{1}{\sqrt[3]{x}}\right)^5$ **4.6.7.** $y = \frac{x}{1-x^2}$. **4.6.8.** $y = \frac{x^2+x+1}{x^2-x+1}$. **4.6.9.** $y = \frac{(x-1)(x+3)}{x+1}$. **4.6.10.** $y = \frac{3x-1}{x^5}\sqrt{x^3+2}$. **4.6.11.** $y = \frac{x}{\sqrt{x^2-1}}$. **4.6.12.** $y = \frac{x}{\sqrt[3]{x+1}}$. **4.6.13.** $y = \sqrt{x^2+x\sqrt{x}}$. **4.6.14.** $y = \sqrt{x^2+\sqrt[3]{x}}$. **4.6.15.** $y = \frac{x^2-2}{(1+x^2)\sqrt{1+x^2}}$. **4.6.16.** $y = x\sqrt[3]{(2x+3)^2}$. **4.6.17.** $y = (x^3-1)\sqrt{x-1} + x\sqrt[3]{x^2-1}$. **4.6.18.** $y = \frac{x\sqrt[3]{(2x-3)^2}}{(x-1)^2}$. **4.6.19.** $y = \sqrt{\frac{x-1}{x+1}}$. **4.6.20.** $y = \frac{x^2+x+1}{x-2}\sqrt[3]{x+1}$. **4.6.21.** $y = \frac{x\sqrt{x^2-1}}{x^3+1}$. **4.6.22.** $y = \sqrt[3]{\frac{x^2+x+1}{x+1}}$. **4.6.23.** $y = x\sqrt{x^2-1}\times\sqrt[3]{x+\sqrt{x}}$ **4.6.24.** $y = \left(x+\frac{1}{\sqrt{x}}\right)^{1/7}x^2$. **4.6.25.** $y = \frac{\sqrt[3]{x}-2x}{(1+x)^{1/3}}$.

4.7 The Exponential Function

Consider the function $y = a^x$, with $a > 1$. The graphs of this function for $a = 2$ and $a = 3$ are shown in Figure 4.7.1*a*. When $x = 0$, $y = 1$ for arbitrary a (all graphs pass through the point (0, 1)).

For all x, the function y is positive and grows with increasing x so that the

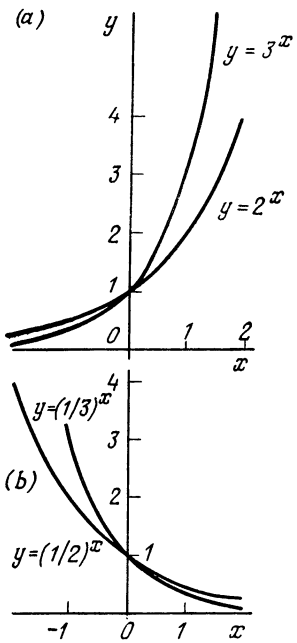


Figure 4.7.1

derivative is everywhere positive as well. When x is increased by a constant quantity c, we get $y(x + c) = a^{x+c} = a^c a^x = ba^x = by(x)$, with $b = a^c$, that is, the quantity y is multiplied by a constant quantity. Thus, if x is varied in successive fashion, by indentical steps (*in arithmetic progression*),

$$x = x_0,\ x_0 + c,\ x_0 + 2c,\ \ldots,\ x_0 + nc,\ \ldots,$$

then y will assume the values

$$y_0,\ by_0,\ b^2y_0,\ \ldots,\ b^ny_0,\ \ldots.$$

It will be recalled that such a law of growth is called a *geometric progression*.

Let us find the derivative of the exponential function for $a = 10$.[4.12] We can write

$$\frac{d(10^x)}{dx} = \frac{10^{x+dx} - 10^x}{dx} = 10^x \frac{10^{dx} - 1}{dx},$$

where dx (instead of Δx) emphasizes the passage to the limit as $\Delta x \to 0$.[4.13]

What is the quantity $(10^{dx} - 1)/dx$? It is the *limit* of the ratio $(10^{\Delta x} - 1)/\Delta x$ as $\Delta x \to 0$. Let us find this limit numerically, arithmetically. Using a four-place table of logarithms or a pocket calculator with a y^x key, we get

$$10^{0.1} \simeq 1.2589,\ \frac{10^{0.1} - 1}{0.1} \simeq 2.589;$$

$$10^{0.01} \simeq 1.0233,\ \frac{10^{0.01} - 1}{0.01} \simeq 2.33;$$

$$10^{0.001} \simeq 1.0023,\ \frac{10^{0.001} - 1}{0.001} \simeq 2.3.$$

Thus, we have

$$\lim_{\Delta x \to 0} \frac{10^{\Delta x} - 1}{\Delta x} \simeq 2.3$$

(where, of course, the constant on the right-hand side has been determined only approximately, albeit reliably). Hence,

$$\frac{d}{dx}(10^x) \simeq 10^x \times 2.3. \qquad (4.7.1)$$

We found the derivative of 10^x in experimental fashion, so to speak, by means of an arithmetic experiment. For any other exponential function it is now easy to reduce the problem to the preceding one. Using the concept of a logarithm, we write

$$a = 10^{\log a},\quad a^x = 10^{x \log a}. \qquad (4.7.2)$$

By the rule for finding a derivative of a composite function, we get

$$\frac{da^x}{dx} \simeq 10^{x \log a} \times 2.3 \log a$$
$$= a^x \times 2.3 \log a. \qquad (4.7.3)$$

The remarkable peculiarity of the exponential function is that its *derivative is directly proportional to the function itself*:

$$\frac{da^x}{dx} = ka^x, \qquad (4.7.4)$$

where the proportionality factor k depends on base a. Therein lies the chief property of a geometric progression: the greater the quantity, the faster the growth.[4.14]

If $0 < a < 1$, the graphs of the exponential function will be as shown in Figure 4.7.1*b*. When x increases in arithmetic progression, y *diminishes* in geometric progression. Formula (4.7.3) is still applicable, but now $\log a$ is negative (since $a < 1$) and, hence, the derivative, which is proportional to the function, is of opposite sign. In

---

[4.12] $a = 10$ is taken simply to facilitate computation involving a table of base-10 logarithms. The reader that owns a pocket calculator with a y^x key can write out the values of, say, $2^{0.1}, 2^{0.01}, \ldots$ or $3^{0.1}, 3^{0.01}, \ldots$ and the ratios corresponding to these values: $(2^{0.1} - 1)/0.1$, etc.

[4.13] It is this fact that makes us write $\frac{d(10^x)}{dx} = \frac{10^{x+dx} - 10^x}{dx}$ (an *exact* equation) instead of $\frac{d(10)^x}{dx} = \frac{10^{x+\Delta x} - 10^x}{\Delta x}$ (an *approximate* equation).

[4.14] This property of geometric progressions, their exceedingly fast rate of build-up, is a favorite topic in many popular-science books on mathematics, for instance, in Ya.I. Perelman's *Mathematics Can Be Fun* (Mir Publishers, Moscow, 1985).

Part 2 we will give some instances of a quantity diminishing in time in such a manner that the rate of decrease is proportional to the quantity at a given instant:

$$\frac{dy}{dx} = -cy, \quad c > 0. \tag{4.7.5}$$

From the foregoing it is evident that in this case the solution of the problem is the exponential function

$$y = y_0 a^x, \quad a < 1. \tag{4.7.6}$$

Exercises

Find the derivatives of the following functions:

4.7.1. $y = 2^x$. 4.7.2. $y = 5^{x+1}$. 4.7.3. $y = (1/2)^x$. 4.7.4. $y = 10^{\sqrt{x}}$. 4.7.5. $y = 2^{x^2}$. 4.7.6. $y = 2^{x+1/x}$.

4.8 The Number e

Let us find a base a for which the derivative of an exponential function $y = a^x$ is of the simplest form, namely, such that the coefficient in the expression for the derivative, k in (4.7.4), is equal to unity, so that it need not be written at all. We denote this base by e. Thus, by definition,

$$\frac{de^x}{dx} = e^x. \tag{4.8.1}$$

This number is easily found by formula (4.7.3):

$$2.3 \log e \simeq 1, \quad \log e \simeq 1/2.3 \simeq 0.43,$$

whence, referring to a table of logarithms, we get $e \simeq 2.7$. This practical approach, however, does not follow the historical development of mathematics and is fundamentally unsatisfactory. We made use of numbers taken from logarithmic tables and did not stop to think how they were computed.[4.15]

Let us find the number e solely on the basis of formula (4.8.1). By the general properties of exponential functions, $e^0 = 1$. Let us consider the function $y = e^x$. We have $y(0) = 1$ and $y'(0) = 1$ (see (4.8.1)).

We take a small $\Delta x = r$ and compute the increment of $y = e^x$ when passing from $x = 0$ to $x = r$. We know that we can write $\Delta y \simeq y'(x)\,\Delta x$, which means that $\Delta y \simeq 1 \times \Delta x = r$ and $y(x) = y(0) + \Delta y$, whence

$$e^r \simeq 1 + r. \tag{4.8.2}$$

This equation should be interpreted in the same way as we interpreted $y(x + \Delta x) \simeq y(x) + y'(x)\,\Delta x$, that is, in (4.8.2) all terms of the second and higher orders, or terms proportional to r^2, r^3, etc., are ignored, while only the first order term (proportional to r) is retained. What is important here is that the coefficient of r in (4.8.2) is exactly unity—this follows from the property of the number e formulated in (4.8.1) (this is simply the definition of e).

We write the small number r as a fraction with a large denominator, $r = 1/n$: if $r \ll 1$, then $n \gg 1$.[4.16] Then from (4.8.2) we get $e^{1/n} \simeq 1 + 1/n$, whence $e \simeq (1 + 1/n)^n$.

Since formula (4.8.2), which ignores terms proportional to r^2, r^3, etc., is the more exact the smaller r is, the last expression for e is the more exact the larger n is. We have thus arrived at another rigorous definition of e:

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \tag{4.8.3}$$

[4.15] The accuracy with which e can be determined in this manner is low; it is highly improbable that, using a four-place table of logarithms for determining the derivative of 10^x, you could find the correct value of the second digit after the decimal point in e. Actually, $\frac{10^{dx} - 1}{dx} \simeq 2.302585$, but of course this value was obtained without the use of logarithmic tables. Even less suitable are school logarithmic tables for finding a value of e that can easily be remembered (there is really no sense in this!): $e \simeq 2.718281828459045$ (15 decimal places!). Note that *e cannot be expressed by a periodic* decimal fraction.

On questions concerning the formulas (and methods) that enable finding e with a very high accuracy, see Sections 6.1 and 6.2.

[4.16] The notation $r \ll 1$ means that the number r is *very small compared to unity*; similarly, the notation $n \gg 1$ means that n is a very large number (*n is considerably greater than unity*).

(this is read: e is the limit of $(1 + 1/n)^n$ as n tends to infinity).[4.17]

Similarly, for a fixed k and a small r (such that $kr \ll 1$) we get $e^{kr} \simeq 1 + kr$, which yields, if we put $r = 1/n$ (where n is very large), $e^k \simeq (1 + k/n)^n$, or, to be more precise,

$$e^k = \lim_{n\to\infty} \left(1 + \frac{k}{n}\right)^n. \qquad (4.8.3a)$$

One should not fear such words as "limit" or "infinity". Actually $(1 + 1/100)^{100} \simeq 2.705$, which is only slightly different from the exact value of e. We advise the reader to find $(1 + 1/8)^8$ for himself.

We have found that $e^r \simeq 1 + r$ for small r, and this is the more exact the smaller r is. [4.18] Let us check this using numbers. We set up the following table:[4.19]

r	−0.05	−0.4	−0.3	−0.2	−0.1	−0.01	0
$1+r$	0.5	0.6	0.7	0.8	0.9	0.99	1
e^r	0.6065	0.6703	0.7408	0.8187	0.9048	0.9901	1

r	0.01	0.1	0.2	0.3	0.4	0.5
$1+r$	1.01	1.1	1.2	1.3	1.4	1.5
e^r	1.0101	1.1052	1.2214	1.3499	1.4918	1.6487

We see that even when $r = \pm 3$, the error introduced by (4.8.2) does not exceed 6%. A physicist or engineer must remember not only that $e \simeq 2.72$ but also that $e^2 \simeq 7.4$, $e^3 \simeq 20$, $e^4 \simeq 55$, $e^5 \simeq 150$. Values of e^x and e^{-x} are given in Table A4.1 at the end of the book.

The number e greatly simplifies the solution of problems involving geometric progressions and compound interest. Consider the following example. How many times will production increase over a period of 50 years if the annual growth rate is 2%? We have to compute 1.02^{50}. The use of the number e consists in our setting, approximately, $1.02 = e^{0.02}$ (see formula (4.8.2), whence $1.02^{50} \simeq e^{0.02\times 50} = e \simeq 2.72$. The general formula is

$$(1+r)^m \simeq e^{mr}, \quad r \ll 1. \qquad (4.8.4)$$

To apply this formula r must be small (there is no other requirement), while m and mr need not be small. If mr is also small, then $e^{mr} \simeq 1 + mr$, and we obtain the earlier formula

$$(1 + r)^m \simeq 1 + mr \qquad (4.8.5)$$

(see formula (4.5.4)). However, we cannot use formula (4.8.5) if mr is large, whereas expression (4.8.4) remains valid (provided that $r \ll 1$). For the example given above, the exact value of 1.02^{50} (with three decimal places) is 2.693, formula (4.8.4) yields $1.02^{50} \simeq e^1 \simeq 2.72$, and formula (4.8.5) yields $(1 + 0.2)^{50} \simeq 1 + 50 \times 0.02 = 2$. Computations using e yielded an error of about 1% whereas formula (4.8.5) yielded an error of about 25%.

In general form an estimate of the precision of formula (4.8.4) is given in Section 6.1 (see Exercise 6.1.4). It will be found there that the relative error introduced by this formula is of the order of mr^2, so that formula (4.8.4) can be employed when $mr^2 \ll 1$ and formula (4.8.5), as we already know, only when $mr \ll 1$. But since (alas) the condition $mr \ll 1$ for r small is much stronger than $mr^2 \ll 1$, there are values of m and r for which the approximation (4.8.4) is valid but (4.8.5) is not. For instance, in the above-considered case of $r = 0.02$ and $m = 50$ we have $mr = 1$ and $mr^2 = 0.02$, which means that mr^2 is small and mr is not. In the same manner, for $r = 0.001$ and $m = 10\,000$ we have $mr = 10$ and there is no way in which condition $mr \ll 1$ can be satisfied, while the condition $mr^2 \ll 1$ can be assumed valid (cf. Exercise 4.8.2).

In accordance with the original definition of the number e by formula (4.8.1), the derivatives of exponential functions are especially simple when the base is equal to e. These derivatives are conveniently expressed in terms of the

[4.17] Of course, from the viewpoint of a pedantic mathematician the statement that from $e^{1/n} \simeq 1 + 1/n$ follows $e \simeq (1 + 1/n)^n$ is not rigorous, since both sides of the approximate equality (valid only for $n \gg 1$) cannot always be raised to a large power n. Thus, our reasoning does not prove the definition (4.8.3), but just explains it. Nevertheless, we hope the reader will find the reasoning convincing.

[4.18] Detailed tables have been compiled for the function $y = e^x$.

[4.19] The values of e^x are taken from a four-place table.

function itself. Here are a number of formulas:

$$y = e^x, \quad \frac{dy}{dx} = e^x = y;$$

$$y = Ce^x, \frac{dy}{dx} = C\frac{de^x}{dx} = Ce^x = y;$$

$$y = Ce^{kx}, \quad \frac{dy}{dx} = C\frac{de^{kx}}{dx}$$

$$= Ce^{kx}\frac{d(kx)}{dx} = ky;$$

$$y = e^{m(x)}, \quad \frac{dy}{dx} = \frac{de^{m(x)}}{dx} = e^{m(x)}\frac{dm(x)}{dx}$$

$$= y\frac{dm(x)}{dx};$$

$$y = f(x)\,e^{m(x)}, \quad \frac{dy}{dx} = f'(x)\,e^{m(x)}$$

$$+ f(x)\,e^{m(x)}m'(x) = y\left(\frac{f'(x)}{f(x)} + m'(x)\right).$$

The exponential function of x to base e is commonly written e^x. Often, however, another designation for e^x is used: $y = \exp x$ (read: the exponential of x). The law e^x is called the ***exponential law*** and the function e^x is termed the ***exponential function*** (in the narrow sense). The designation with exp is especially convenient when x proves to be a complicated and unwieldy expression. For example, it is more convenient to write $\exp\left(\frac{7t^2+24t}{t^2+5}\right)^3$ than $e^{\left(\frac{7t^2+24t}{t^3+5}\right)^3}$.

The number e can also be defined geometrically. Two exponential functions, $y = a^x$ and $y_1 = b^x$, are connected via a simple relation:

$$b^x = a^{kx}, \text{ where } k = \log_a b \qquad (4.8.6)$$

(this fact has already been used in formula (4.7.2)). From (4.8.6) it follows that the graph of y_1 is obtained from the graph of y by shrinking the latter k-fold along the x axis (Figure 4.8.1*a*; cf. Section 1.7).

It is clear that all graphs of the function $y = a^x$ corresponding to different values of a pass through one point (0, 1) (since $a^0 = 1$ for arbitrary a); however, the graphs intersect the y axis at this point at different angles.

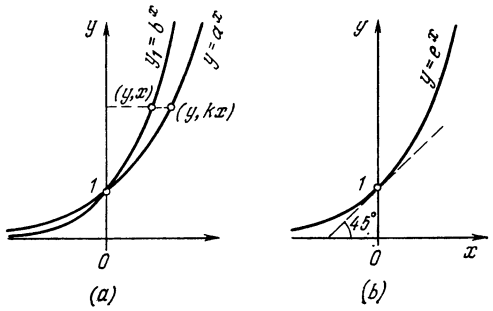


Figure 4.8.1

The simplest exponential function is the one whose curve intersects the y axis at (0, 1) at an angle of 45° (strictly speaking, the tangent to the curve at this point intersects the x axis at an angle of 45° (Figure 4.8.1*b*). From (4.8.1) it follows that this curve is the one corresponding to the function e^x (recall the geometric meaning of the derivative).

Conclusion. To summarize, we can give four distinct definitions of the number e: (1) from the condition $(e^x)' = e^x$, (2) from the condition $e^r \simeq 1 + r$ for $r \ll 1$, (3) as the limit of $(1 + 1/n)^n$ as $n \to \infty$, and (4) as the base in the exponential function $y = e^x$ (the graph of this function intersects the y axis at an angle of 45°).

The number e has other remarkable definitions and properties; some of these will be discussed in Chapter 6.

To fix this important material in his mind, the reader is advised to put aside the book and see how from each definition there follow the other three regarded as properties of number e.

Note, finally, that in the equation $y = e^x$ the quantities x and y are, of course, dimensionless numbers, since e^x with $x = 100$ m or $x = 5$ h has no meaning. Consequently, in the laws of science that involve the exponential law connecting two dimensional quantities, y and x, we always have $y = l_2 e^{x/l_1}$, where l_1 and l_2 are the units of measurement of x and y. The transition to a new unit for x, or $l_1' = cl_1$,

always leads to a transition from the old function $y = e^x$ to a new function $y = e^{cx'}$ (equivalent to the old one), where $x' = x/c$ is the same quantity x but measured in new units l_1'. Similarly, the substitution $l_2' = l_2/d$ for the unit for y leads to a transition from formula $y = e^x$ to $y' = de^x$ (where y' is measured in units of l_2'). Correspondingly, all functions of the type $y = ae^{kx}$, where a and k may be arbitrary, are called exponential functions, since in the majority of cases a transition to new a and k means only a change in the units of measurement, so that only the fact that the two quantities (x and y) are connected by an exponential law and the signs of a and k (k positive corresponds to an increasing function $y = e^{kx}$ and k negative to a decreasing function $y = e^{-|k|x}$) have physical meaning.

Exercises

4.8.1. Find the derivatives of the following functions: (a) $y = e^{-x}$, (b) $y = 5e^x - e^{3x}$, (c) $y = e^{x^2}$, (d) $y = e^{\sqrt{x}}$, and (e) $y = e^{x^3 - 3x + 1}$.

4.8.2. The interest rate on a sum on a bank account is 0.1% every month. How much will the sum in the bank account increase in a thousand years? [*Hint.* Use formula (4.8.4).]

4.9 Logarithms

By definition, the ***logarithm*** of a quantity h to a base a is the exponent f of the power to which a (the logarithmic base) must be raised in order to obtain the given number h:

$$f = \log_a h \text{ means that } h = a^f. \quad (4.9.1)$$

In this sense the logarithmic function $y = \log_a x$ is the inverse (see Section 1.6) of the exponential function $y = a^x$.

The curve representing the relationship $y = \log_a x$ (with $a > 1$) is depicted in Figure 4.9.1. Note that at $x = 1$ (irrespective of the value of a) we have $y = 0$, for $x > 1$ we have $y > 0$, and for $x < 1$ we have $y < 0$. The entire curve is located to the right of the axis of ordinates: since a positive number

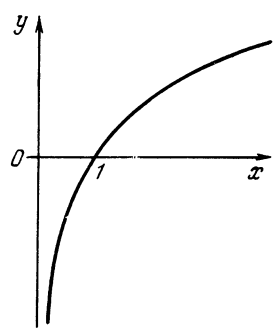


Figure 4.9.1

a raised to any power yields a positive number, there are no logarithms of negative numbers. [4.20]

As seen from Figure 4.9.1, the derivative of the function $y = \log_a x$ (for $a > 1$) is positive for all values of x, since y increases with x (for $a > 1$). Note also that the rate of growth of $y = \log_a x$ with x decreases, that is, the derivative decreases as x increases (below we will give a rigorous proof of this).

Let us derive a formula connecting the logarithms of one and the same number to different bases. Suppose that

$$f = \log_a h, \quad a^f = h. \quad (4.9.2)$$

Taking the logarithms of both sides of (4.9.2) to base b, we get

$$f \log_b a = \log_b h, \text{ whence } f = \frac{\log_b h}{\log_b a}.$$

Taking (4.9.2) into account, we get

$$\log_a h = \frac{\log_b h}{\log_b a}. \quad (4.9.3)$$

This can be rewritten as

$$\log_b x = k \log_a x, \quad \text{where} \quad k = \log_b a. \quad (4.9.3a)$$

The number $k = \log_b a$ is known as the *modulus of logarithms to base b with respect to logarithms to base a*. From (4.9.3a) it follows that the graph for $y = \log_b x$ is obtained from the graph for $y = \log_a x$ by a k-fold stretching along the y axis.

[4.20] See, however, Section 14.3.

Logarithms to base e are called ***natural logarithms*** and are denoted by $\ln x$. The "naturalness" of this system of logarithms is due to fact that in some respects this system appears to be the simplest; for instance, the rate of growth (the derivative) is the simplest in the case of logarithms if $y = \ln x$. These remarkable properties of natural logarithms, which we will repeatedly encounter below, led to a situation in which the two creators of the theory of logarithms, the Scottish amateur mathematician John ***Napier*** (1550-1617) and Swiss watch and instrument maker Joost ***Bürgi*** (1552-1632), independently and almost simultaneously discovered precisely this type of logarithms (or very similar system of logarithms). The ***base-10 logarithms*** are known as ***common logarithms*** and were first considered by Henry ***Briggs*** (1561-1630), an English astronomer, geometer, and numerical-table maker, prompted by Napier, who was his friend and whom he greatly admired.[4.21]

Natural logarithms can be characterized by the property that the graph of $y = \ln x$ intersects the axis of abscissas at point (1, 0) at an angle of 45°, since the curve is obtained from that of the inverse, $y = e^x$, by reflection with respect to the bisector of the coordinate angle (see Section 1.6), and the curve $y = e^x$ intersects the axis of ordinates at point (0, 1) at an angle of 45° (see Figure 4.8.1*b*).

Let us now find the derivative of a *natural* logarithm. We consider $d \ln x = \ln (x + dx) - \ln x$. Take advantage of the familiar formula $\ln a - \ln b = \ln (a/b)$. Then

$$d \ln x = \ln \frac{x+dx}{x} = \ln \left(1 + \frac{dx}{x}\right). \tag{4.9.4}$$

We already know (see formula (4.8.2)) that $e^r \simeq 1 + r$ for r small. Take the logarithms of both sides; since $\ln e^r = r$, we arrive at an extremely important formula:

$$\ln (1 + r) \simeq r \tag{4.9.5}$$

(the approximate equation has the same meaning as (4.8.2): it is valid for $r \ll 1$). Combining (4.9.4) and (4.9.5), we get

$$d \ln x = \ln \left(1 + \frac{dx}{x}\right) = \frac{dx}{x}$$

(employing dx instead of Δx makes it possible to assume that the last equation is exact), whence

$$\frac{d \ln x}{dx} = \frac{1}{x}. \tag{4.9.6}$$

When x varies in geometric progression, $\ln x$ varies in arithmetic progression, that is, if $x = a, ab, ab^2, ab^3, \ldots$, then $\ln x = \ln a$, $\ln a + c$, $\ln a + 2c$, $\ln a + 3c$, $\ldots$, where $c = \ln b$. For this reason, the larger x is, the slower $\ln x$ grows and the smaller is the derivative, which is reflected in formula (4.9.6).

The derivative of a natural logarithm can also be found by using the fact that the logarithmic function and the exponential function are inverse functions (with respect to each other). We can write

$$y = \ln x, \quad x = e^y, \quad x' = \frac{dx}{dy} = \frac{d(e^y)}{dy} = e^y,$$

$$\frac{dy}{dx} = \frac{1}{e^y} = \frac{1}{x}.$$

Now, using (4.9.3a), we can calculate the derivative of a logarithm to any base a. Suppose that $y = \log_a x$. Then (see (4.9.3) and (4.9.3a))

$$y = \frac{\ln x}{\ln a}, \quad \frac{dy}{dx} = \frac{1}{\ln a} \frac{d \ln x}{dx} = \frac{1}{\ln a} \frac{1}{x}. \tag{4.9.7}$$

Replacing a with e and b and h with a, we get $\ln a = (\log_a e)^{-1}$, we can rewrite (4.9.7) in the following manner:

$$\frac{d \log_a x}{dx} = \frac{\log_a e}{x}. \tag{4.9.7a}$$

[4.21] In the beginning, common logarithms were known as ***Briggs logarithms*** and natural logarithms as ***Napierian logarithms.***

The simplest one of the formulas (4.9.6), (4.9.7), and (4.9.7a) is (4.9.6), which is valid for logarithms to base e. For rough mental calculations, it is advisable to memorize the following facts: $\ln 2 \simeq 0.69$, $\ln 3 \simeq 1.1$, and $\ln 10 \simeq 2.3 \simeq 1/0.434$. A short table of natural logarithms is given in Table A4.2 at the end of the book.

If some function $f(x)$ is under the sign of the logarithm, the derivative is found by the rule for differentiating composite functions (see Section 4.3):

$$\frac{d \ln f(x)}{dx} = \frac{1}{f(x)} \frac{df(x)}{dx}. \qquad (4.9.8)$$

The derivative $(\ln f(x))'$ of the logarithm of a function $f(x)$ is often called the ***logarithmic derivative*** of $f(x)$; according to (4.9.8), it is equal to the ratio of the derivative $f'(x)$ to the function itself, $f(x)$. Here is an example that illustrates the usefulness of this concept.

Formula (4.9.8) enables one to find the derivatives of expressions of the form $f(x)^{h(x)}$, that is, such that contain the variable both in the base and in the exponent. Suppose that

$$y = f(x)^{h(x)}. \qquad (4.9.9)$$

Taking logs (the logarithms can be taken to any base; we choose natural logarithms), we obtain

$$\ln y = h(x) \ln f(x). \qquad (4.9.10)$$

Let us now take the derivatives of both sides in (4.9.10) and have regard for the fact that $\ln y$ is a composite function of x (just as $\ln f(x)$ is):

$$\frac{1}{y} y' = h'(x) \ln f(x) + h(x) \frac{f'(x)}{f(x)},$$

$$y' = y \left[h'(x) \ln f(x) + h(x) \frac{f'(x)}{f(x)} \right],$$

or, allowing for (4.9.9),

$$y' = f(x)^{h(x)} h'(x) \ln f(x) + h(x) f(x)^{h(x)-1} f'(x). \qquad (4.9.11)$$

Consider this formula. On the right we have a sum of two terms: the first term, $f(x)^{h(x)} h'(x) \ln f(x)$, is the derivative of f^h computed on the assumption that only h is a variable while f is held constant, and the second term, $h(x) f(x)^{h(x)-1} f'(x)$, is the derivative of f^h computed on the assumption that only f is a variable while h is held constant. This confirms the general principle expressed in Section 4.2.

Note, finally, that in $y = \log_a x$ (just as in $y = a^x$), the variables x and y and the constant a are dimensionless. If x is a dimensional quantity (say, length), then the number x is specified only to within a constant factor depending on the chosen unit for x. Correspondinly, the quantity $\log_a x$ (for any fixed base a) is determined only to within a constant term (to within an *additive constant*, as mathematicians usually say in such cases); in that respect it is similar to the indefinite integral.[4.22] If both x and y are dimensional quantities, then the right way to approach this problem is to write $y = l_2 \log_a (x/l_1)$, where the factors l_1 and l_2 are units of measurement for x and y, since a logarithm can be taken only of a dimensionless quantity, say of number 100 obtained as a result of taking the fraction, say (distance AB)/(unit l_1), where l_1 could be one meter. The value of a logarithmic function is also a dimensionless quantity, whereby the dimensional quantity must be equal to $l_2 \log_a (x/l_1)$, that is, $\log_a (x/l_1)$ units l_2. These simple considerations must be allowed for since the logarithmic function (just as the exponential function) is often encountered in the laws of science.

It must be also noted that, in view of (4.9.3) and (4.9.3a), the transition from one base a to another base b is reduced to multiplying all logarithms to the old base a by a constant, $k = \log_b a$, that is, to a change in the unit for $y = \log_a x$. This is why, in the majority of cases, the base of a logarithm does not play an

[4.22] As for the deep connection between logarithms and integrals (which partially explains this analogy), see formula (5.2.2) (which sometimes is used to define logarithms) and the subsequent discussion in Section 7.7.

important role, which makes it possible almost always to employ natural logarithms as being the simplest.

Exercises

4.9.1. What is the value of $\log_5 15$? [*Hint.* Use formula (4.9.3)].

4.9.2. Derive the formula for the derivative of (a) the product of two functions, using the formula $\log(uv) = \log u + \log v$, and (b) the quotient of two functions, using the formula $\log(u/v) = \log u - \log v$.

4.9.3. Find the derivative of each of the following functions: (a) $y = \ln(x+3)$, (b) $y = \ln 2x$, (c) $y = \ln(x^2+1)$, (d) $y = \ln(x+1/x)$, (e) $y = \ln(3x^3 - x + 1)$, (f) $y = \ln \dfrac{x-1}{x+1}$, (g) $y = \ln \dfrac{\sqrt{x}}{\sqrt[3]{x+1}}$, (h) $y = x \ln x$, (i) $y = x^3 \ln(x+1)$, (j) $y = x^x$, and (k) $y = \sqrt{x}\sqrt{x^2-1}$.

4.10 Trigonometric Functions

In this section we will find the derivatives of trigonometric functions.

Trigonometric functions are defined as ratios of line segments, ratios of the sides of right triangles or ratios of certain line segments in a circle (the line of sines, the line of cosines, etc.) to the circle's radius. For this reason trigonometric functions are dimensionless, and the independent variable in such functions is also dimensionless, an angle. The natural measure of angles in higher mathematics is the ***radian***, a central angle subtended in a circle by an arc whose length is equal to the radius of the circle. A central angle of t radians is subtended in a circle of radius 1 by an arc whose length is t. The importance of trigonometric functions to the description of physical processes will be discussed in Part 2 (see Chapter 10).

In what follows we will always consider a circle of unit radius. Then we will briefly say that the sine is equal to the length of the line of sines in that circle, the angle is equal to the arc length, and so on. The reader must bear in mind, however, that both trigonometric functions and angles are dimensionless quantities and are not measured

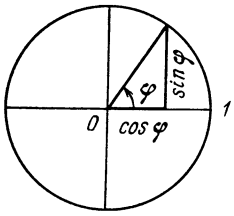


Figure 4.10.1

by any units of length (centimeters, inches, or meters). The sine is equal to the length of the line of sines (in centimeters) divided by the length of the radius (in centimeters,) and only when $r = 1$ cm is it numerically equal to the length of the line of sines. The lines of sines and cosines are shown in Figure 4.10.1.

Recall the form of the graphs of sine and cosine as functions of the angle (Figure 4.10.2). The period of the sine, like that of the cosine, is equal to $2\pi \simeq 6.28$ and corresponds to a complete revolution of the radius of the circle.

Let us find the derivatives of the sine and cosine geometrically. In Figure 4.10.3, the endpoint of the radius drawn at an angle φ is A, and the endpoint of the radius drawn at a close angle $\varphi + d\varphi$ is B. The length of arc AB is therefore equal to $d\varphi$. Draw from A a perpendicular AC to the line of sines BB' of the angle $\varphi + d\varphi$. As can be seen from Figure 4.10.3,

$$AA' = \sin\varphi,\ BB' = \sin(\varphi + d\varphi),$$
$$BC = \sin(\varphi + d\varphi) - \sin\varphi = d\sin\varphi$$

(here, of course, the last equation is "exact in the limit", as $d\varphi \to 0$). Moreover,

$$OA' = \cos\varphi,\ OB' = \cos(\varphi + d\varphi),$$
$$A'B' = AC = \cos\varphi - \cos(\varphi + d\varphi) = -d\cos\varphi.$$

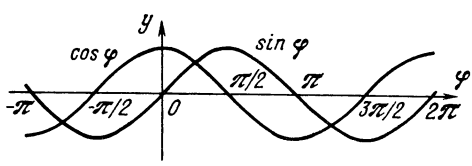


Figure 4.10.2

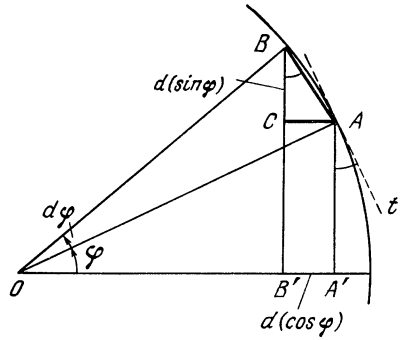


Figure 4.10.3

Since the angle $d\varphi$ is small, the arc length AB does not differ much from the length of the chord AB and the angle ABC formed by the chord AB and the vertical line BCB' is equal to φ.[4.23] From a consideration of triangle ABC we find that $BC = AB\cos\varphi$ and $AC = AB\sin\varphi$. Thus,

$$d\sin\varphi = \cos\varphi\, d\varphi,$$

$$-d\cos\varphi = \sin\varphi\, d\varphi$$

and, hence,

$$\frac{d\sin\varphi}{d\varphi} = \cos\varphi, \quad \frac{d\cos\varphi}{d\varphi} = -\sin\varphi. \tag{4.10.1}$$

Such a simple and pictorial way of calculating the derivative of the sine or cosine is quite convincing for a beginner. On the other hand, an experienced and knowledgeable reader (not the one for whom this book is intended) will notice the difficulties inherent in such an approach. For a small but finite (i.e. not infinitesimal) angle $\Delta\varphi$ the length of chord AB is less than the arc length: it can be proved that

$$AB = \Delta\varphi - \frac{(\Delta\varphi)^3}{12} + \cdots.$$

Furthermore, actually the angle CBA is $\varphi + \Delta\varphi/2$ instead of φ, so that

$$BC = AB\cos\left(\varphi + \frac{\Delta\varphi}{2}\right) = \left(\Delta\varphi - \frac{(\Delta\varphi)^3}{12} + \cdots\right)\left(\cos\varphi - \frac{\Delta\varphi}{2}\sin\varphi + \ldots\right).$$

[4.23] The tangent line t to the circle at point A forms with the vertical line AA' an angle $A'At$ equal to φ ($\angle A'At = \angle AOA'$ as angles with mutually perpendicular sides); $\angle ABC$ differs very little from $\angle A'At$ since chord AB has practically the same direction as tangent t.

Therefore, strictly speaking, the above reasoning is not rigorous and the relationships are incorrect. However, they are true to within infinitesimals of first order with respect to the small quantity $\Delta\varphi$. Hence, while for increments we have only the approximate equations $\Delta\sin\varphi \simeq \Delta\varphi \times \cos\varphi$ and $\Delta\cos\varphi \simeq -\Delta\varphi \times \sin\varphi$, the corresponding equations for differentials are exact, which means that Eqs. (4.10.1) are exact, too.

The derivation of the formulas for the derivatives of trigonometric functions we have just described is equivalent to the derivation based on mechanical analogies (for one, a derivation that incorporates the idea of a *vector*). We imagine a point $M = M(t)$ whose movement in the plane is described by two parametric equations:

$$x = \varphi(t), \quad y = \psi(t), \tag{4.10.2}$$

where t is the time variable (see Section 1.8). In this case the rate of variation in the coordinates x and y of point M will be given by the derivatives $x' = dx/dt$ and $y' = dy/dt$. The vector $\mathbf{MM_1}/\Delta t = \mathbf{v}_{av} = (\Delta x/\Delta t, \Delta y/\Delta t)$, where $\Delta x/\Delta t$ and $\Delta y/\Delta t$ are the coordinates of the vector (see Figure 4.10.4), characterizes the ***average velocity*** over the line segment MM_1 of the path (here $M_1 = M(t + \Delta t)$). Vector $\mathbf{v} = (dx/dt, dy/dt)$ characterizes the ***instantaneous velocity*** at point M of the path: it is directed along the tangent to the trajectory at point M and its length is equal to $\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt}$, where $\Delta s \simeq |\mathbf{MM_1}|$ is the increment of the distance s over time Δt.

Now suppose the point M moves uniformly in a circle of radius 1; it travels, say, 1 cm in 1 s. Then Eqs. (4.10.2) assume the form

$$x = \cos t, \quad y = \sin t, \tag{4.10.3}$$

and the velocity $\mathbf{v}$ of the motion is given by a vector with coordinates $d(\cos t)/dt$, $d(\sin t)/dt$. Clearly, in this case vector $\mathbf{v}$ will be of unit length (since by hypothesis the increment

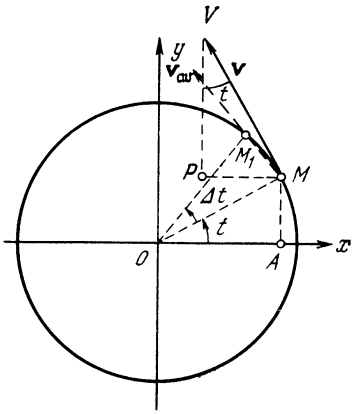


Figure 4.10.4

Δs of distance is equal to the increment Δt of time) and directed along the tangent to the circle (at every point; see Figure 4.10.4). The coordinates of vector **v**, which is perpendicular to the radius **OM** = **r** of the circle, are obviously $-\sin t$ and $\cos t$ (compare the right triangles AOM and VMP whose sides are mutually perpendicular in Figure 4.10.4; they have the same hypotenuses $OM = VM = 1$ and equal acute angles $\angle AOM = \angle MVP = t$; the minus sign in front of $\sin t$ here specifies the direction of the line segment **MP**). Thus, we again arrive at formula (4.10.1):

$$\frac{d\cos t}{dt} = -\sin t, \quad \frac{d\sin t}{dt} = \cos t.$$

Finally, we will give another method for calculating the derivatives of $\sin\varphi$ and $\cos\varphi$, which does not require a drawing and is more formal. According to the general formulas, $\Delta\sin\varphi = \sin(\varphi + \Delta\varphi) - \sin\varphi$. Recall the formula for the sine of the sum of two angles,

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta,$$

and apply it to $\sin(\varphi + \Delta\varphi)$ to get

$$\sin(\varphi + \Delta\varphi) = \sin\varphi\cos\Delta\varphi + \cos\varphi\sin\Delta\varphi,$$

whence

$$\Delta\sin\varphi = \sin\varphi\cos\Delta\varphi + \cos\varphi\sin\Delta\varphi - \sin\varphi = \cos\varphi\sin\Delta\varphi - \sin\varphi(1 - \cos\Delta\varphi).$$

Let us form the ratio of the increments as follows:

$$\frac{\Delta\sin\varphi}{\Delta\varphi} = \cos\varphi\frac{\sin\Delta\varphi}{\Delta\varphi} - \sin\varphi\frac{1-\cos\Delta\varphi}{\Delta\varphi}. \qquad (4.10.4)$$

Now we have to send $\Delta\varphi$ to zero and thus pass to the limit. We know that for angles α or $\Delta\varphi$ tending to zero, the sine is equal to the arc length, $\sin\alpha \simeq \alpha$, or $\sin\Delta\varphi \to \Delta\varphi$ as $\Delta\varphi \to 0$. In other words,

$$\lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi} = 1.$$

The second term on the right-hand side of (4.10.4) must first be transformed by using the familiar formula $1 - \cos 2\alpha = 2\sin^2\alpha$, whereby $1 - \cos\Delta\varphi = 2\sin^2(\Delta\varphi/2)$. In this formula, we substitute $\Delta\varphi/2$ for $\sin(\Delta\varphi/2)$ (this is justified since $\Delta\varphi$ is small). Hence,

$$\frac{1-\cos\Delta\varphi}{\Delta\varphi} \simeq \frac{2(\Delta\varphi/2)^2}{\Delta\varphi} = \frac{\Delta\varphi}{2}.$$

Hence, in the limit, as $\Delta\varphi \to 0$, the second term vanishes: $\lim_{\Delta\varphi\to 0}\frac{1-\cos\Delta\varphi}{\Delta\varphi} = 0$, whence $\lim_{\Delta\varphi\to 0}\frac{\Delta\sin\varphi}{\Delta\varphi} = \frac{d\sin\varphi}{d\varphi} = \cos\varphi$. Similarly,

$$\Delta\cos\varphi = \cos(\varphi + \Delta\varphi) - \cos\varphi = \cos\varphi\cos\Delta\varphi - \sin\varphi\sin\Delta\varphi - \cos\varphi = -\cos\varphi(1 - \cos\Delta\varphi) - \sin\varphi\sin\Delta\varphi$$

(here we have used the formula for the cosine of the sum of two angles), and therefore

$$\frac{\Delta\cos\varphi}{\Delta\varphi} = -\cos\varphi\frac{1-\cos\Delta\varphi}{\Delta\varphi} - \sin\varphi\frac{\sin\Delta\varphi}{\Delta\varphi}.$$

Since we already know that

$$\lim_{\Delta\varphi\to 0}\frac{1-\cos\Delta\varphi}{\Delta\varphi} = 0, \quad \lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi} = 1,$$

the final result is

$$\frac{d\cos\varphi}{d\varphi} = \lim_{\Delta\varphi\to 0}\frac{\Delta\cos\varphi}{\Delta\varphi} = -\cos\varphi\times 0 - \sin\varphi\times 1 = -\sin\varphi.$$

Equations (4.10.1) are valid for arbitrary angles and not only for those in the first quadrant. It is also useful to verify, glancing at the graphs of the functions $\sin x$ and $\cos x$, that Eqs. (4.10.1) give the proper signs of the derivatives for *any* x.

Let us check the validity of Eqs. (4.10.1) for small angles. For small φ, it is obvious geometrically that $\sin\varphi \simeq \varphi$ and $\cos\varphi \simeq 1$, where the approximate equality sign $\simeq$, just as in many other cases, means equality to

within terms of the first order of smallness. The first equation in (4.10.1) for small φ yields $d(\sin\varphi)/d\varphi \simeq 1$ and the second yields $d(\cos\varphi)/d\varphi \simeq -\varphi$, which means that $d(\cos\varphi)/d\varphi = 0$ at $\varphi = 0$. The fact that the derivative is zero signifies that the cosine has a maximum at $\varphi = 0$, while the fact that $d(\sin\varphi)/d\varphi = 1$ at $\varphi = 0$ signifies that for small φ the sine increases approximately like φ.

If we know the derivatives of the functions $y = \sin x$ and $y = \cos x$, it is easy to find the derivatives of all other trigonometric functions by using the formulas interrelating them. For instance, $\tan x = \frac{\sin x}{\cos x}$. By the formula for finding the derivative of a quotient we get

$$\frac{d\tan x}{dx} = \frac{d(\sin x/\cos x)}{dx} = \frac{\cos x\cos x - \sin x(-\sin x)}{\cos^2 x},$$

whence

$$\frac{d\tan x}{dx} = \frac{\cos^2 x+\sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}. \quad (4.10.5)$$

From Figure 4.10.5 (the graph of $\tan x$) we can see that the function $y = \tan x$ has a positive derivative for arbitrary x. Near the points of discontinuity ($x = \pi/2$, $x = 3\pi/2$, . . .) the derivative increases without bound, and both the function and its derivative become infinite at these points. Both of these conclusions are in full agreement with formula (4.10.5).

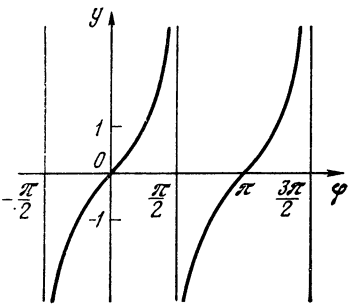


Figure 4.10.5

By a completely analogous device we find that

$$\frac{d\cot x}{dx} = -\frac{1}{\sin^2 x}. \quad (4.10.6)$$

The derivatives of a tangent and a cotangent can also be found directly. Note that

$$\tan\alpha-\tan\beta = \frac{\sin\alpha}{\cos\alpha}-\frac{\sin\beta}{\cos\beta} = \frac{\sin\alpha\cos\beta-\sin\beta\cos\alpha}{\cos\alpha\cos\beta} = \frac{\sin(\alpha-\beta)}{\cos\alpha\cos\beta},$$

whence

$$\Delta\tan\varphi = \tan(\varphi+\Delta\varphi)-\tan\varphi = \frac{\sin\Delta\varphi}{\cos(\varphi+\Delta\varphi)\cos\varphi}. \quad (4.10.7)$$

Bearing in mind that $\lim\limits_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi} = 1$, we get, from (4.10.7),

$$\frac{d\tan\varphi}{d\varphi} = \lim_{\Delta\varphi\to 0}\frac{\Delta\tan\varphi}{\Delta\varphi} = \lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi}\cdot\frac{1}{\cos(\varphi+\Delta\varphi)+\cos\varphi} = 1\times\frac{1}{\cos^2\varphi} = \frac{1}{\cos^2\varphi}.$$

Exercises

Find the derivatives of the following functions:

4.10.1. $y = \sin(2x+3)$. **4.10.2.** $y = \cos(x-1)$. **4.10.3.** $y = \cos(x^2 - x + 1)$. **4.10.4.** $y = \sin^2 x$. **4.10.5.** $y = \sin 3x\cos^2 x$. **4.10.6.** $y = (\sin 2x)^x$. **4.10.7.** $y = x\tan x$. **4.10.8.** $y = e^{\tan 2x}$. **4.10.9.** $y = \cot(x/2)$.

4.11 Inverse Trigonometric Functions

New and very interesting results are obtained when we consider the ***inverse trigonometric functions***. We remind the reader of how these functions are defined. The function

$$y = \operatorname{Arcsin} x \quad (4.11.1)$$

is an angle y whose sine is equal to x:

$$\sin y = x. \quad (4.11.2)$$

These two equations denote the same thing. Similarly, the function

$$y = \operatorname{Arctan} x \quad (4.11.3)$$

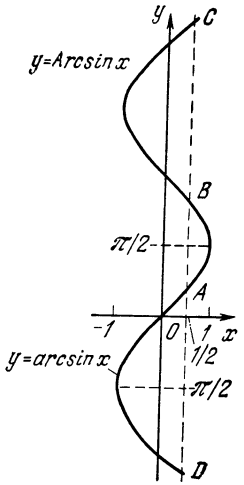


Figure 4.11.1

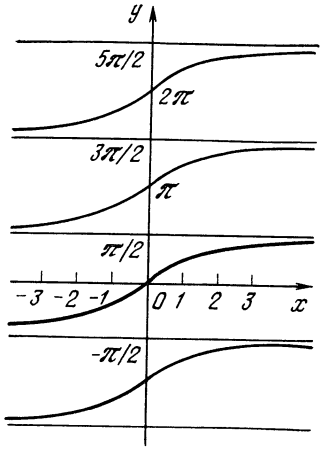


Figure 4.11.2

is an angle y whose tangent is equal to x:

$$\tan y = x. \qquad (4.11.4)$$

The definitions are similar for the functions $y = \text{Arccos } x$ (i.e. $x = \cos y$) and $y = \text{Arccot}$ (i.e. $x = \cot y$). Note that the functions $y = \text{Arcsin } x$ and $y = \text{Arccos } x$ are meaningful only for $-1 \leqslant x \leqslant 1$ (cf. (4.11.2)). The functions $y = \text{Arctan } x$ and $y = \text{Arccot } x$ are meaningful for *all* values of x.

Let us consider in more detail the function $y = \text{Arcsin } x$. For instance, let $x = 1/2$. Then $y = \text{Arcsin } (1/2)$. We can take $y = \pi/6$ since $\sin(\pi/6) = 1/2$; however, we can also take $y = 5\pi/6$, since $\sin(5\pi/6)$ is also equal to $1/2$. We can likewise take $y = 13\pi/6$, $y = 17\pi/6$, and so on. We see that one value of x is associated with an *infinitude* of values of y. All that this means is that $\text{Arcsin } x$ is *multiple-valued* (see p. 49) and even *infinitely valued.* These properties of $y = \text{Arcsin } x$ are seen in the graph of Figure 4.11.1.

Take the portion of the curve representing the function $x = \sin y$ on which the function is monotonic; usually the portion of the curve $y = \text{Arcsin } x$ on which this condition is satisfied is such that $-\pi/2 \leqslant y \leqslant \pi/2$. This part of the curve is called the ***principal value*** of the function $\text{Arcsin } x$ and is denoted by $y = \arcsin x$. If we confine ourselves to a consideration of $y = \arcsin x$, then to each x there corresponds only *one* value of y. Thus, the function $y = \arcsin x$ is *single-valued.*

The principal value of the arctangent function is defined in similar fashion:

$$-\frac{\pi}{2} < \arctan x < \frac{\pi}{2}$$

(since the function $x = \tan y$ is monotonic on the segment $-\pi/2 < y < \pi/2$, Figure 4.11.2).

Let us find the derivative of $y = \arcsin x$. We take advantage of the fact that the arcsine is the inverse of the sine, $y = \arcsin x$ and $x = \sin y$. We then have

$$x'(y) = \frac{dx}{dy} = \cos y,$$
$$y'(x) = \frac{dy}{dx} = \frac{1}{x'(y)} = \frac{1}{\cos y}. \qquad (4.11.5)$$

But we consider x the independent variable and not y, so dy/dx should be expressed in terms of x and not in terms of y, as in (4.11.5).

From $\sin^2 y + \cos^2 y = 1$ it follows that $\cos y = \pm\sqrt{1 - \sin^2 y}$. Since we are considering the principal value of the arcsine function, it follows that $-\pi/2 \leqslant y \leqslant \pi/2$, $\cos y \geqslant 0$, and so

we take the plus sign in front of the radical sign: $\cos y = \sqrt{1 - \sin^2 y}$. Since $\sin y = x$ by (4.11.2), it follows that $\cos y = \sqrt{1 - x^2}$. Substituting this into (4.11.5) yields

$$\frac{dy}{dx} = \frac{1}{\sqrt{1-x^2}}, \text{ or } \frac{d \arcsin x}{dx} = \frac{1}{\sqrt{1-x^2}}. \quad (4.11.6)$$

This formula may be used only for the principal value. If we want it to be valid for other values, we must choose the appropriate sign in front of the radical $\sqrt{1 - x^2}$ (this radical can be considered double-valued). Indeed, for one and the same value of x, the derivative has different signs: on various portions of the curve at points A and C of the graph of the function $y = \text{Arcsin } x$ (Figure 4.11.1) the derivative is positive and at points B and D it is negative.

Let us now find the derivative $\frac{d \arctan x}{dx}$. If $y = \arctan x$, then $x = \tan y$, whence, by the foregoing, we find that

$$x'(y) = \frac{dx}{dy} = \frac{1}{\cos^2 y}, \qquad y'(x) = \frac{dy}{dx} = \frac{1}{x'(y)} = \cos^2 y. \quad (4.11.7)$$

From trigonometry we have

$$\tan^2 y + 1 = \frac{1}{\cos^2 y},$$

and therefore

$$\frac{1}{\cos^2 y} = 1 + x^2.$$

Using formula (4.11.7), we finally get

$$\frac{dy}{dx} = \frac{d \arctan x}{dx} = \frac{1}{1+x^2}. \quad (4.11.8)$$

This formula holds true for any other branch of the arctangent (see Figure 4.11.2) since any other branch is obtained from the principal one by a parallel translation, and this does not affect the magnitude of the derivative.

Exercises

Find the derivatives of the following functions:

4.11.1. $y = \arccos x$ (this function is defined by the condition $0 \leqslant y \leqslant \pi$). **4.11.2.** $y = \text{arccot } x$ (here $0 < y < \pi$). **4.11.3.** $y = \text{arcsin } 2x$. **4.11.4.** $y = \arctan (3x + 1)$. **4.11.5** $y = \arctan (x^2 - x)$. **4.11.6.** $y = e^{\arctan \sqrt{x}}$.

4.12 Differentiating Functions Dependent on a Parameter and Functions of Several Variables. Partial Derivatives

Often in physical problems we are forced to find the derivative of a function $y(x)$ which, aside from depending on the independent variable x, depends on one or several parameters, or quantities that characterize the system being investigated, say, masses or linear dimensions, and that are assumed constant, fixed, for the system considered but in general may have different values. Clearly, if the function $y = f(x, t)$ depends on the independent variable x and the parameter t, its derivative $y' = dy/dx$ will also depend on the same parameter t. For instance, if $y = \sin(\lambda x)$ (λ is the parameter), then $y'(x) = dy/dx = \lambda \cos(\lambda x)$ and $y''(x) = d^2y/dx^2 = -\lambda^2 \sin(\lambda x)$; if $y = e^{-\alpha x^2}$ (α being the parameter), then $y'(x) = dy/dx = 2\alpha x e^{-\alpha x^2}$ and $y''(x) = d^2y/dx^2 = 2\alpha(2\alpha x^2 - 1) e^{-\alpha x^2}$ (verify this).

Note that since the function $y = f(x, t)$ and its derivatives $y' = dy/dx$, d^2y/dx^2, etc. depend not only on x but on parameter t, which may assume different values, it would be interesting to know to what extent y and y' and y'', all taken at a fixed value of x, depend on parameter t if the latter is slightly varied. The answer lies in the value of the rate of change of y with t, that is, in $(dy/dt)_{x=\text{constant}}$. For instance, in the case of the two functions we considered above, $y_1 = \sin(\lambda x)$ and $y_2 = e^{-\alpha x^2}$, the rates $dy_1/d\lambda$ and $dy_2/d\alpha$ are, respectively, $x \sin(\lambda x)$ and $-x^2 e^{-\alpha x^2}$ (x is fixed but λ and α vary).

Obviously, if we are dealing with a function $y = f(x, t)$, the notation y', y'', etc. is ambiguous since it does not allow us to know whether we are speaking of $(dy/dx)_{t=\text{constant}}$ or of an entirely different quantity $(dy/dt)_{x=\text{constant}}$. To explain the notation used here, we turn to the general case of a function of *two* (independent) *variables*, $z = F(x, y)$, since the above is a particular case of such a function. True, the function was denoted by y and the independent variables by x and t, but nothing prevents us from choosing other letters.

If we assume that $y = y_0$ is constant and x varies, we will arrive at a function $z = F(x, y_0) = \varphi(x)$ of one variable x. The derivative $\varphi'(x)$ of such a function can, naturally, be written as $(dF(x, y)/dx)_{y=y_0=\text{constant}}$. Since such notation is cumbersome and not too useful, we use $\partial F(x, y)/\partial y$, where the symbol ∂ (the "round dee") is used instead of d to emphasize the fact that the derivative is taken at y *constant*, that is, the function $F(x, y)$ is considered as a function of one variable x (it depends, however, also on parameter y). The derivative

$$\frac{\partial F(x, y)}{\partial x} = \frac{dF(x, y)}{dx}\Big|_{y=\text{constant}} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x, y) - F(x, y)}{\Delta x}$$

is said to be the ***partial derivative*** of $F(x, y)$ with respect to the independent variable x; similarly,

$$\frac{\partial F(x, y)}{\partial y} = \frac{dF(x, y)}{dy}\Big|_{x=\text{constant}} = \lim_{\Delta y \to 0} \frac{F(x, y+\Delta y) - F(x, y)}{\Delta y}$$

is the partial derivative of $F(x, y)$ with respect to y. For example, if $y = y(x, \lambda) = \sin(\lambda x)$, then $\partial y/\partial x = \lambda \cos(\lambda x)$ and $\partial y/\partial \lambda = x \cos(\lambda x)$, while if $y = y(x, \alpha) = e^{-\alpha x^2}$, we have $\partial y/\partial x = -2\alpha x e^{-\alpha x^2}$ and $\partial y/\partial \alpha = -x^2 e^{-\alpha x^2}$.

Of course, in calculating the values of the partial derivatives $\partial F(x, y)/\partial x$ and $\partial F(x, y)/\partial y$ of a function $z = F(x, y)$ we must fix *both* independent variables, $x = x_0$ and $y = y_0$, that is, $(\partial F(x, y)/\partial y)_{x=x_0,\ y=y_0}$ means that we have the partial derivative of $F(x, y)$ with respect to x calculated on the assumption that $y = y_0$ and taken at the point $x = x_0$ (in other words, the derivative of $\varphi(x) = F(x, y_0)$ with respect to x, which is equal to $\lim_{\Delta x \to 0} \frac{\varphi(x_0+\Delta x) - \varphi(x_0)}{\Delta x}$). Note also that if the value z of the function $F(x, y)$ is measured in units of E and the variables x and y are measured in units of e_1 and e_2, respectively, then the partial derivatives dz/dx and $\partial z/\partial y$ have *different* dimensions, $E/e_1 \neq E/e_2$.

The partial derivatives of a function $z = F(x, y)$ enable estimating the increment $\Delta z = F(x + \Delta x, y + \Delta y) - F(x, y)$ of the function caused by small variations in x and y, or increments Δx and Δy that are small. Obviously,

$$\begin{aligned}\Delta z &= F(x + \Delta x, y + \Delta y) \\ &= [F(x + \Delta x, y + \Delta y) \\ &\quad - F(x, y + \Delta y)] \\ &\quad + [F(x, y + \Delta y) - F(x, y)].\end{aligned}$$

But the difference $\Delta_1 z = F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y)$ is the increment of a function of a single variable, $F(x, y + \Delta y)\,|_{y+\Delta y=\text{constant}} = \varphi(x)$, due to an increase in the independent variable x by Δx; according to formula (4.1.2), $\Delta_1 z = \varphi(x + \Delta x) - \varphi(x) \simeq \varphi'(x)\,\Delta x = \frac{\partial F(x, y+\Delta y)}{\partial x}\,\Delta x$. Similarly, the difference $F(x, y + \Delta y) - F(x, y) = \Delta_2 z$ is the increment of the function $\psi(y) = F(x, y)\,|_{x=\text{constant}}$ of a single variable y due to an increase in the independent variable y by Δy: according to (4.1.2), $\Delta_2 z \simeq \psi'(y)\,\Delta y = \frac{\partial F(x, y)}{\partial y}\,\Delta y$. And since we can always assume that for Δy small we have $\frac{\partial F(x, y+\Delta y)}{\partial x} \simeq \frac{\partial F(x, y)}{\partial x}$, the final expression for the

increment $\Delta z = \Delta_1 z + \Delta_2 z$ of the function $z = F(x, y)$ is

$$\Delta z \simeq \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y. \qquad (4.12.1)$$

It is clear that the *accuracy of the approximation* (4.12.1) *is the higher, the smaller the increments* Δx *and* Δy *of the independent variables* x *and* y of function z. In other words, for small Δx and Δy (which here are assumed to be of the same order of smallness), the difference $\Delta z - \left(\frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y\right)$ will be of a higher order of smallness, and, as Δx and Δy tend to zero, the ratio of the left- to right-hand sides of (4.12.1) (where it is assumed, of course, that the right-hand side does not vanish) tends to unity. For instance, if $z = xy^3$, then, obviously $dz/\partial x = y^3$ and $\partial z/\partial y = 3xy^2$, and small increments Δx and Δy lead to an increase in z by approximately $y^3\Delta x + 3xy^2\Delta y \left(= \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y\right)$.

Although the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ of a function $z = F(x, y)$ are measured in different units, this in no way influences formula (4.12.1). If x, y, and z are measured in units of e_1, e_2, and E, then the derivatives $\partial z/\partial x$ and $\partial z/\partial y$ have the dimensions of E/e_1 and E/e_2, but the increments Δx and Δy have dimensions that coincide with those of x and y, that is, of e_1 and e_2, so that the two terms on the right-hand side of (4.12.1) have the same dimensions, of E, which coincide with the dimensions of the left-hand side, Δz. If we go over to a new unit for x, that is, $e_1' = ke_1$, the partial derivative $\partial z/\partial x$ will get multiplied by k, but the numerical value of Δx will become k times smaller, so that the product $(\partial z/\partial x)\Delta x$ will remain unchanged. But if we go over to a new unit for z, or $E' = KE$, then the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ will be divided by K and, therefore, the increment Δz given by (4.12.1) will be K times smaller (as was to be expected).

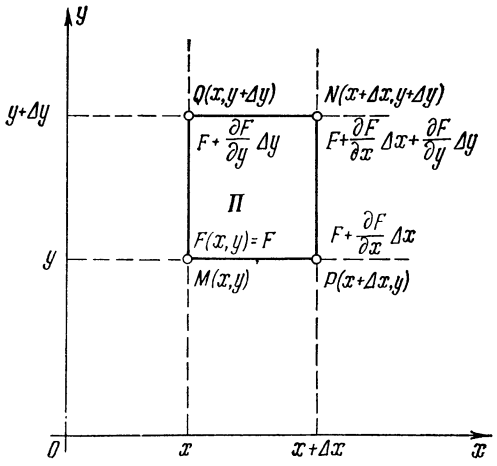


Figure 4.12.1

The meaning of (4.12.1) can be explained geometrically. A function $z = F(x, y)$ of two variables x and y is defined at all points of a *plane domain* Φ (the domain of the function) where each point is fixed by its two coordinates, x and y. Points $M(x, y)$ and $N(x + \Delta x, y + \Delta y)$ are the opposite vertices in the rectangle $\Pi = MPNQ$ depicted in Figure 4.12.1. Movement along the straight lines MP and QN changes only the variable x, so that if we confine ourselves to these two straight lines, the function z can be considered a function of only one variable x (these are the functions $\varphi(x) = F(x, y)|_{y=\text{constant}}$ and $\varphi_1(x) = F(x, y + \Delta y)|_{y+\Delta y=\text{constant}}$); the increment (due to the transition from M to P and from Q to N) can be found via (4.1.2). In the same fashion, if we move only along the sides MQ and PM of the rectangle, the function z transforms into a function of only one variable y (these are the functions $\psi(y) = F(x, y)|_{x=\text{constant}}$ and $\psi_1(y) = F(x + \Delta x, y)|_{x+\Delta x=\text{constant}}$. At each vertex of the rectangle in Figure 4.12.1 we have written the (approximate) values of the function z.

The situation with a function with a number of variables greater than two is similar. For instance, if $u = f(x, y, z, t)$,

then the increment of the function that is caused by the increments in all four independent variables, Δx, Δy, Δz, Δt, is

$$\Delta u \simeq \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \frac{\partial u}{\partial z}\Delta z + \frac{\partial u}{\partial t}\Delta t, \tag{4.12.1a}$$

where, say, $\frac{\partial u}{\partial x} = \left.\frac{df(x, y, z, t)}{dx}\right|_{\substack{y=\text{constant},\\ z=\text{constant},\\ t=\text{constant},}}$

etc. The accuracy of formula (4.12.1a) will be the greater, the smaller the quantities Δx, Δy, Δz, Δt are.

Note also that the function $z = F(x, y)$ actually connects three variables, x, y, and z, and each variable can be considered a function of the other two. For instance, y is a function of x and z, since by fixing the values $x = x_0$ and $z = z_0$ of the variables x and z we can, using the equation $z_0 = F(x_0, y)$, find the value of y (it could happen that there are more than one value, but this would only mean that $z = f(x, y)$ specifies y as a *multiple-valued* function of x and z). This fact enables one to determine such partial derivatives as $(\partial y/\partial z)_{x=\text{constant}}$ and $(\partial y/\partial x)_{z=\text{constant}}$ and the like. In physical problems the choice of variables on which a given quantity z depends can be very different. Two different functions $z = F(x, y)$ and $z = G(x, u)$ where y and u are different physical parameters on which z depends, enable finding two different values of the partial derivative $\partial z/\partial x$, which would be appropriately denoted by $(\partial z/\partial x)_{y=\text{constant}}$ and $(\partial z/\partial x)_{u=\text{constant}}$ $\left(\text{or, briefly, by } \left.\frac{\partial z}{\partial x}\right|_y \text{ and } \left.\frac{\partial z}{\partial x}\right|_u\right)$. It is clear that the two derivatives have the same dimensions, those of E/e, where E and e are the units of measurement of z and x, but numerically they are not necessarily the same. For instance, the energy W of a capacitor is $(1/2)C\varphi^2 = (1/2)q^2/C$, where C is the capacitance, φ the voltage across the capacitor, and q the quantity of electricity, or charge, on the capacitor (see formula (13.4.3)). Therefore, if we start with the function $W = W(C, \varphi)$, we get $\left.\frac{\partial W}{\partial C}\right|_\varphi = \frac{1}{2}\varphi^2 = \frac{W}{C}$, which is the rate of variation of the energy as the capacitance varies (and the voltage is kept constant), while if we start with the function $W = W(C, q)$, we will find that $\left.\frac{\partial W}{\partial C}\right|_q = -\frac{1}{2}\frac{q^2}{C} = -\frac{W}{C}$ (with the charge on the capacitor being constant); thus,

$$\left.\frac{\partial W}{\partial C}\right|_\varphi = -\left.\frac{\partial W}{\partial C}\right|_q.$$

It is clear that the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ of a function $z = F(x, y)$ (with respect to x and y) are defined for all x and y from the domain Φ of the function F, that is, are themselves functions of x and y. Hence, we can define the partial derivatives of partial derivatives, or what is known as ***partial derivatives of second order***:

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x^2}, \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x\,\partial y},$$

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y\,\partial x}, \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y^2}.$$

For instance, in the case of the function $z = xy^3$, which we have already met before, we have $\partial^2 z/\partial x^2 = \partial(y^3)/\partial x = 0$, $\partial^2 z/\partial x \partial y = \partial(y^3)/\partial y = 3y^2$, $\partial^2 z/\partial y \partial x = \partial(3xy^2)/\partial x = 3y^2$, and $\partial^2 z/\partial y^2 = \partial(3xy^2)/\partial y = 6xy$. Thus, we see that $\partial^2 z/\partial y \partial x = \partial^2 z/\partial x \partial y$.

We can easily see that this is not accidental but is a general law: $\partial^2 z/\partial x \partial y = \partial^2 z/\partial y \partial x$. Indeed, according to the definition of a derivative (or formula (4.1.2)),

$$\frac{\partial^2 z}{\partial x\,\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) \simeq \frac{\dfrac{\partial z(x, y+\Delta y)}{\partial x} - \dfrac{\partial z(x, y)}{\partial x}}{\Delta y}, \tag{4.12.2}$$

where the approximate equality may be as precise as desired (we need only select Δy small enough). On the other hand, it is obvious that

$$\frac{\partial z(x, y+\Delta y)}{\partial x} \simeq \frac{z(x+\Delta x, y+\Delta y) - z(x, y+\Delta y)}{\Delta x} \tag{4.12.3a}$$

$$\frac{\partial z(x, y)}{\partial x} \simeq \frac{z(x+\Delta x, y) - z(x, y)}{\Delta x}, \tag{4.12.3b}$$

where the approximate equalities become exact equalities as $\Delta x \to 0$. Substituting (4.12.3a) and (4.12.3b) into (4.12.2), we get

$$\frac{\partial^2 z}{\partial x\,\partial y} \simeq \frac{1}{\Delta y}\left[\frac{z(x+\Delta x, y+\Delta y)-z(x, y+\Delta y)}{\Delta x} - \frac{z(x+\Delta x, y)-z(x, y)}{\Delta x}\right]$$
$$= \frac{z(x+\Delta x, y+\Delta y)-z(x, y+\Delta y)}{\Delta x\,\Delta y} - \frac{z(x+\Delta x, y)-z(x, y)}{\Delta x\,\Delta y} \qquad (4.12.4a)$$

(note that in the numerator on the right-hand side we have the sum (with alternating signs) of the values of function F at the vertices of rectangle Π depicted in Figure 4.12.1, and in the denominator, the area of Π). Similarly

$$\frac{\partial^2 z}{\partial y\,\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) \simeq \frac{\dfrac{\partial z(x+\Delta x, y)}{\partial y} - \dfrac{\partial z(x, y)}{\partial y}}{\Delta x}$$

and

$$\frac{\partial z(x+\Delta x, y)}{\partial y} \simeq \frac{z(x+\Delta x, y+\Delta y)-z(x+\Delta x, y)}{\Delta y}$$
$$\frac{\partial z(x, y)}{\partial y} \simeq \frac{z(x, y+\Delta y)-z(x, y)}{\Delta y},$$

whence

$$\frac{\partial^2 z}{\partial y\,\partial x} \simeq \frac{1}{\Delta x}\left[\frac{z(x+\Delta x, y+\Delta y)-z(x+\Delta x, y)}{\Delta y} - \frac{z(x, y+\Delta y)-z(x, y)}{\Delta y}\right]$$
$$= \frac{z(x+\Delta x, y+\Delta y)-z(x+\Delta x, y)}{\Delta x\,\Delta y} - \frac{z(x, y+\Delta y)-z(x, y)}{\Delta x\,\Delta y}. \qquad (4.12.4b)$$

Since the approximate equalities (4.12.4a) and (4.12.4b) may be made as accurate as desired, we conclude that $\partial^2 z/\partial x\partial y = \partial^2 z/\partial y\partial x$.

Partial derivatives of higher orders are defined similarly; the value of the derivative is determined by how many times we differentiate with respect to each variable and not by the order in which this is done. It is quite easy to generalize the facts for the case of a function of three or more variables, but we will not do this here.

If in space we introduce a Cartesian system of coordinates (see Figure 4.12.2) consisting of three coordinate axes, x, y, and z, then the "graph" of the function $y = F(x, y)$ is a *surface* (Σ). This surface is formed by all the points $M(x, y, z)$ obtained as a result of the following process. On each straight line that is perpendicular to the xy-plane and erected at point $P(x, y)$ (the perpendiculars are parallel to the z axis) we lay off a segment PM of length z (this segment is laid off upward or downward depending on whether z is positive or negative). Suppose that the surface Σ is *smooth* in the neighborhood of a point $M(x_0, y_0, z_0)$, with $z_0 = F(x_0, y_0)$, that is, it has no salient points or discontinuities or "cusps" (like the vertex, or apex, of a conical surface) or similar "singular" points. In this case the tangents to all the curves that pass through point M and belong to Σ form a plane σ known as the *tangent plane* (the plane tangent to a surface at a certain point). Clearly, the tangent plane σ is completely defined by the tangents t_1 and t_2 to the curves along which the planes $y =$ constant $(=y_0)$ and $x =$ constant $(=x_0)$ cut Σ, and the slopes of t_1 and t_2 in relation to the xy-plane are given by the values of the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$. (Why?) It is easy to see that the equation of plane σ in the xyz coordinate system has the form

$$z - z_0 = \frac{\partial z}{\partial x}(x-x_0) + \frac{\partial z}{\partial y}(y-y_0), \qquad (4.12.5)$$

where, say, $\partial z/\partial x$ is $(\partial z/\partial x)_{x=x_0,\ y=y_0}$.

The right-hand side of (4.12.1) $(\partial z/\partial x)\,\Delta x + (\partial z/\partial y)\,\Delta y$, is called the ***total differential*** of function $z = F(x, y)$ and is denoted by dz. If $z = x$ (i.e. $F(x, y) = x$), then, obviously, the partial derivatives $\partial z/\partial x$ and $\partial z/\partial y$ of this simplest function of two variables (which actually depends only on x) are 1 and 0, respectively, and therefore the total differential of this function is $dx = 1 \times \Delta x + 0 \times \Delta y = \Delta x$. Similarly, it can be proved that $dy = \Delta y$, where on the left-hand side we have the total differential of the function $G(x, y) = y$ of two variables (actually of one variable) and on the right-hand side we have the increment of y. Since $\Delta x = dx$ and $\Delta y = dy$, we can rewrite the formula for the total differential of $z = F(x, y)$ as

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy. \qquad (4.12.6)$$

The equation of the plane σ tangent to surface Σ, (4.12.5), clarifies the geometrical meaning of the total differential. This equation implies that the Z-coordinate (below we will see why the letter "Z" is more convenient than the letter "z") of point $N_1(x, y, Z)$ in plane σ, a point corresponding to the value $x_0 + \Delta x$ of the x-coordinate and to the value $y_0 + \Delta y$ of the y-coordinate, is equal to $z_0 + dz$ (compare with (4.12.5), where the increments $x - x_0$ and $y - y_0$ of the x- and y-coordinates are now denoted by Δx and Δy). On the other hand, the z-coordinate of the respective point $N(x, y, z)$ on Σ is equal to $z_0 + \Delta z$, where Δz is the increment of z, or simply $z - z_0$. Thus, the approximate formula (4.12.1),

$$z \simeq z_0 + dz, \qquad (4.12.7)$$

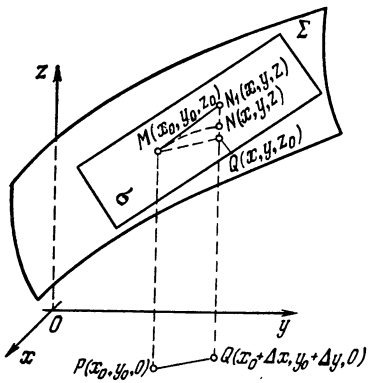


Figure 4.12.2

suggests that, for small Δx and Δy, the point $N\ (x,\ y,\ z)$ of Σ, where $x = x_0 + \Delta x$ and $y = y_0 + \Delta y$, will lie sufficiently close to point $N_1\ (x,\ y,\ z)$, where $z = z_0 + \Delta z$ and $Z = z_0 + dz$, that belongs to plane σ tangent to Σ at point $M\ (x_0,\ y_0,\ z_0)$ (see Figure 4.12.2). The error introduced by the approximate formula (4.12.7) is a quantity of second order of smallness if compared with dx and dy (or Δx and Δy), which are assumed to be quantities of equal order of smallness.

Quite similarly, for a function of, say, four variables, $u = f\ (x,\ y,\ z,\ t)$, the total differential du is the expression on the right-hand side of (4.12.1a) in which the increments Δx, Δy, Δz, Δt of the independent variables can be replaced with dx, dy, dz, dt. Thus, we can rewrite the approximate formula (3.12.1a) as

$$\Delta u \simeq du.$$

Exercises

Find the first partial derivatives of the following functions:

4.12.1. $z = x^2 + y^2$. **4.12.2.** $z = 1/(x^2 + y^2)$. **4.12.3.** $z = \sqrt{x^2 + y^2}$. **4.12.4.** $z = e^{-(x^2+y^2)}$. **4.12.5.** $z = x^2y^3$. **4.12.6.** $z = xe^y + ye^x$. **4.12.7.** $z = \ln(x + y)$. **4.12.8.** $z = \cos(xy)$.

4.12.9. Find the second partial derivatives of the functions in Exercises 4.12.1-4.12.8 and check whether $\partial^2 z/\partial x \partial y = \partial^2 z/\partial y \partial x$.

4.13 The Derivative of an Implicit Function

To define a function implicitly means to define it via an expression of the form

$$F\ (x,\ y) = 0. \qquad (4.13.1)$$

If the equation can be solved for x or y, then we revert to the ordinary representation of the function in explicit form. But sometimes such a solution leads to complicated formulas and at other times it cannot be found at all. For instance, the equation of a circle in the form

$$x^2 + y^2 - 1 = 0 \qquad (4.13.2)$$

is simpler than the following expression derived from it:

$$y\ (x) = \pm\sqrt{1 - x^2}. \qquad (4.13.3)$$

If the left-hand side of (4.13.1) is an arbitrary polynomial involving x and y to a power exceeding the fourth, then in general this equation cannot be solved for x or for y. Also, for example, unsolvable is the simple-looking equation

$$F\ (x,\ y) = x \sin x + y \sin y - \pi = 0. \qquad (4.13.4)$$

However, even in those cases where there is no solution in the form of a formula that specifies directly a procedure for computing y for a given x, it still remains a fact that y is a definite function of x. For every x it is possible, by solving the equation numerically, to find a corresponding y and to construct a curve (4.13.1) in the xy-plane. It may be that the curve will not exist for all x (in the case of a circle (4.13.2), for example, it exists only for x between -1 and $+1$) or that the function is multiple-valued, that is, for a given x there may be more than one value of y (in the case of the circle, for instance, there are two values in accord with the $\pm$ sign in front of the square root sign). However, these complications do not detract from the basic fact, which is that the equation $F\ (x,\ y) = 0$ defines y as a function of x.

How can we find the derivative dy/dx? And can this be done without solving the equation, that is, without expressing $y\ (x)$ in explicit fashion?

The procedure which enables finding y' from Eq. (4.13.1) was developed by

Newton. Let the variables x and y satisfy Eq. (4.13.1) and let $x + \Delta x$ and $y + \Delta y$ be adjacent values of the variables that still satisfy Eq. (4.13.1), where the word "adjacent" emphasizes the smallness of Δx and Δy. Since $z = F(x, y) = 0$ and $z + \Delta z = F(x + \Delta x, y + \Delta y) = 0$, we have $\Delta z = F(x + \Delta x, y + \Delta y) - F(x, y) = 0$. But, in view of the basic formula (4.12.1), we have

$$\Delta z \simeq \frac{\partial F(x, y)}{\partial x} \Delta x + \frac{\partial F(x, y)}{\partial y} \Delta y,$$

which means that

$$\frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y \simeq 0 \qquad (4.13.5)$$

and, so

$$\frac{\Delta y}{\Delta x} \simeq -\frac{\dfrac{\partial F(x, y)}{\partial x}}{\dfrac{\partial F(x, y)}{\partial y}}, \qquad (4.13.6)$$

where the approximate equality in (4.13.5) and, therefore, in (4.13.6), is the more exact the smaller the increments Δx and Δy are.

Passing to the limit as $\Delta x \to 0$ (this will also mean that $\Delta y \to 0$), we get the derivative on the left, and the approximate equality in (4.3.6) becomes an exact equality. Finally, we have

$$\frac{dy}{dx} = -\frac{\dfrac{\partial F(x, y)}{\partial x}}{\dfrac{\partial F(x, y)}{\partial y}}. \qquad (4.13.7)$$

Note the minus sign in (4.13.7) and also the fact that we cannot simply cancel the $\partial F(x, y)$ in the numerator and the denominator.

We will demonstrate the application of (4.13.7) using as an example Eq. (4.13.2). We have $F(x, y) = x^2 + y^2 - 1$. Then

$$\frac{\partial F(x, y)}{\partial x} = 2x, \quad \frac{\partial F(x, y)}{\partial y} = 2y,$$

$$\frac{\partial y}{\partial x} = -\frac{2x}{2y} = -\frac{x}{y}. \qquad (4.13.8)$$

It is easy to see that this result coincides with that obtained if we compute the derivative of (4.13.3).

Let us find the derivative in the case of (4.13.4):

$$\frac{\partial F(x, y)}{\partial x} = \sin x + x \cos x,$$

$$\frac{\partial F(x, y)}{\partial y} = \sin y + y \cos y,$$

$$\frac{\partial y}{\partial x} = -\frac{\sin x + x \cos x}{\sin y + y \cos y}.$$

Thus, the expression of the derivative of an implicit function involves both quantities, x and y. To find it numerically, we must find y numerically for a given x. But if we did not have formula (4.13.7), then to find the derivative would require finding, numerically, two values y_2 and y_1 for two adjacent values x_2 and x_1 and finding the ratio $(y_2 - y_1)/(x_2 - x_1)$ and the limit of this ratio.[4.24] Here, the closer x_2 is to x_1, the more exactly we would need to compute y_2 and y_1, and this is often very difficult to do. On the other hand, the use of (4.13.7) presents no difficulties.

If $F(x, y) = 0$ leads to a nonsingle-valued function $y(x)$, that is, when there are two or more values of y for one value of x (several branches of the curve), then (4.13.7) yields the values of the derivative at appropriate points of the curve $F(x, y) = 0$ when a given x and different y's are substituted. The reader is advised to verify this by using the equation of the circle, (4.13.2), for which the derivative is given by formula (4.13.8).

Let us assume that we have a function of two variables, $z = F(x, y)$, for which we do not require that it be zero. Instead we will consider this function along a curve γ in the xy-plane fixed parametrically by the equations $x = \varphi(t)$ and $y = \psi(t)$, where parameter t can be the time variable, for instance. The value of z will then depend

[4.24] Additional difficulties arise when Eq. (4.3.1) specifies a *multiple-valued* function $y = y(x)$, since then one must "match" the values $y_1(x_1)$ and $y_2(x_2)$; otherwise for a small difference $x_2 - x_1$ the difference $y_2 - y_1$ may prove to be large.

on t, too. Let us find the rate of variation of z, or the derivative dz/dt.

We introduce a small increment Δt into the variable t. The variables x and y will then increase by $\Delta x \simeq \varphi'(t)\,\Delta t = (dx/dt)\,\Delta t$ and $\Delta y \simeq \psi'(t)\,\Delta t = (dy/dt)\,\Delta t$. On the other hand, by (4.12.1), we have

$$\Delta z \simeq \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y \simeq \frac{\partial z}{\partial x}\frac{\partial x}{\partial t}\Delta t + \frac{\partial z}{\partial y}\frac{dy}{dt}\Delta t,$$

whence

$$\frac{\Delta z}{\Delta t} \simeq \frac{\partial z}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial z}{\partial y}\frac{\Delta y}{\Delta t}, \qquad (4.13.9)$$

and the approximation is the more exact the smaller the increment Δt is (and, hence, the smaller the Δx and Δy are). We now send Δt to zero in (4.13.9). The ratios $\Delta z/\Delta t$, $\Delta x/\Delta t$ and $\Delta y/\Delta t$ will transform into the respective derivatives and the equality in (4.13.9) will become exact. The final result is

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \qquad (4.13.10)$$

In particular, if the curve γ is specified explicitly, by an equation $y = f(x)$ (i. e. if parameter γ is the abscissa x), then (4.13.10) transforms into

$$\frac{dz}{dx} = \frac{\partial z}{\partial y}\frac{dy}{dx} + \frac{\partial z}{\partial x}. \qquad (4.13.11)$$

For instance, if $z = xy$, the rate of growth of z along the circle $x = a\cos t$, $y = a\sin t$ is $\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = y(-a\sin t) + x(a\cos t) = a^2\cos^2 t - a^2\sin^2 t = a^2\cos 2t$, while along the straight line $y = kx$ we have $\frac{dz}{dx} = \frac{\partial z}{\partial y}\frac{dy}{dx} + \frac{\partial z}{\partial x} = xk + y = 2kx$.

Formulas (4.13.10) and (4.13.11) specify the increment of the z-coordinate of a point M of the surface Σ (whose equation is $z = F(x, y)$; see Figure 4.12.2) caused by a shift along a curve Γ lying on Σ, a curve whose projection in the xy-plane is curve γ with equations $x = x(t)$ and $y = y(t)$ (or $y = f(x)$). The same formulas can be applied to a function with a higher number of variables. Suppose that T is the temperature at a point $M(x, y, z)$ in space and that the temperature varies with time t; we can say that the temperature is a function of four variables: $T = T(x, y, z, t)$. We wish to find the rate with which the temperature changes when we go over from point $M(x, y, z) = M(x_0, y_0, z_0)$ to a neighboring point $M_1(x + dx, y + dy, z + dz)$ (this transition occurs over a small time interval dt, we assume). Ignoring the changes in temperature in time, we arrive at a *stationary* (or steady-state or time-independent) temperature distribution $T = T(x, y, z)$ $(= T(x, y, z, t_0))$. Here the rate of temperature variation caused by transition from point M to point M_1, in accord with formula (4.13.9) modified in the appropriate manner, is

$$v_{\text{con}} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} + \frac{\partial T}{\partial z}\frac{dz}{dt}.$$

This change in temperature, as we have already said, is caused by a shift in space from one point, $M(x, y, z)$, to another point, $M_1(x + dx, y + dy, z + dz)$, along a certain curve $x = x(t)$, $y = y(t)$, $z = z(t)$, say, a transition from a point with a lower temperature to a point with a higher temperature; in accord with this rate of change of temperature, v_{con} associated with such a shift is called the ***convective rate***. Now let us turn to the study of the temperature variation in time at a certain point $M(x_0, y_0, z_0)$. Here the temperature depends only on time t, that is $T = T(t)$ $(= T(x_0, y_0, z_0, t))$. The rate of variation of temperature associated with this function (the fact that the temperature field also depends on time) is called the ***local rate***:

$$v_{\text{loc}} = \frac{\partial T}{\partial t}.$$

Finally (compare this with formula (4.13.11), the ***total rate*** of temperature variation can be found by adding the

local rate to the convective rate:

$$v_{\text{tot}} = v_{\text{con}} + v_{\text{loc}} = \frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt} + \frac{\partial T}{\partial z}\frac{dz}{dt} + \frac{\partial T}{\partial t}$$

(the derivative dT/dt is sometimes called the ***total derivative*** of $T = T(x, y, z, t)$, where $x = x(t), y = y(t), z = z(t)$).

In finding the derivative of an implicit function (or the derivative of a function $z = z(t)$ specified in the form of a composite function of two variables, $z = F(x, y)$, where $x = x(t)$ and $y = y(t)$) we had to introduce a new concept, that of the *partial* derivative (see Section 4.12). This notion is of great importance to the theory of functions of several variables, on which we have only briefly touched here. Actually, we have already, latently, made use of the concept of a partial derivative. For instance, the definite integral $I = \int_a^b f(x)dx = I(a, b)$ is a function of two variables: above we found the partial derivative $\partial I/\partial a$ $(= -f(a))$ and $\partial I/\partial b$ $(= f(b))$ with respect to one of the limits of integration (a or b) assuming the other limit of integration (b or a) fixed (or constant). Even in such elementary questions as finding the derivative of the product of several functions, $y = h(x)\, g(x)$, or as finding the derivative of $y = h(x)^{g(x)}$ (see Section 4.9), the concept of a partial derivative was actually present. Indeed, as was mentioned earlier, y' in these cases is composed of terms obtained when taking the derivatives with respect to x in $h(x)$ and with respect to x in $g(x)$, which clearly involves partial derivatives. The corresponding general rule can be written as follows: if $y = F[g(x), h(x)]$, then, in accord with (4.13.10),

$$y' = \frac{dy}{dx} = \frac{\partial F}{\partial g}\frac{dg}{dx} + \frac{\partial F}{\partial h}\frac{dh}{dx}.$$

Exercises

4.13.1. Find the derivative dy/dx of a function defined by Eq. (4.13.4) at the following points: (a) $z = \pi/2$, $y = \pi/2$, and (b) $x = -\pi/2$, $y = \pi/2$.

4.13.2. Find the derivative dy/dx at the point $x = y = 1$ of a function defined by the equation $x^3 + 3x + y^3 - 8 = 0$.

4.13.3. Let $z = \sqrt{x^2 + y^2}$. Find (a) dz/dt with $x = e^{kt}\cos(lt)$ and $y = e^{kt}\sin(lt)$, and (b) dz/dx with $y = x^2$.

Chapter 5 Integration Techniques

5.1 Statement of the Problem[5.1]

In Chapter 3 we introduced the concept of the *integral* and noted the close connection between two different (at first glance) problems. The problems are

(1) finding the sum of a large number of small summands when the terms can be represented as $v(t)\,\Delta t$ (or $v(t)\,dt$);

(2) finding the function $z(t)$ whose derivative is equal to a given function $v(t)$:

$$\frac{dz}{dt} = v(t). \qquad (5.1.1)$$

Most of the problems that arise in physics, engineering, chemistry, as well as in mathematics, are problems of type (1), that is, they involve finding the sum of a large number of small summands. This statement of the problem is more pictorial and suggests a simple, though approximate, way of computing the quantity of interest, namely, by directly adding the (small) summands of which we spoke earlier. However, this "direct" method of solving problems of type (1) does not make it possible to express the answer by any general formula, and higher mathematics appeared when the relationship was established between problems of types (1) and (2), which opened the way for general algorithms for solving problems of type (1).

The statement of problems of type (2) is more artificial, but it has its advantages. The finding of derivatives proved to be a very simple matter, which reduced to four or five formulas (the derivatives of x^n, e^x, $\ln x$, $\sin x$, $\cos x$) and two or three rules. It is therefore easy to find the derivatives of a large number of functions. Every time the derivative $dz/dt = v$ of some function is found, we can record the fact that for this v the integral z is known (see Section 5.2). In this way we can build up a range of particular cases in which it is possible to solve problems of type (2). For certain simple types of functions v, it has been possible, with the aid of algebraic identity transformations, to find the general rules of solving problems of type (2), say, when v is a sum of functions whose integrals are known (see Section 5.3).

However, this is not possible for *all* the elementary functions, so that integration is more difficult than differentiation (finding derivatives). Nevertheless, the formulas found for certain integrals in the second statement of the problem are very important. If for a given $v(t)$ it is possible to find the integral (the indefinite integral or the antiderivative) $z(t)$, then all problems in the first statement, all sums, that is, all definite integrals $\int_a^b v(t)\,dt$, are then expressed by simple formulas via the function $z(t)$:

$$\int_a^b v(t)\,dt = z(b) - z(a).$$

Such a result is more complete, more exact, and more valuable than the result of every separate numerical computation of a sum, that is, of the definite integral $\int_a^b v(t)\,dt$ between definite limits a and b. For this reason we aim primarily to solve the problem of integration in its second statement.

5.2 Elementary Integrals

Let us write down the formulas for derivatives that have been found in Chapter 4 and the corresponding (indefinite) integrals:

$$\frac{d}{dx}(x^n) = nx^{n-1}, \qquad n\int x^{n-1}\,dx = x^n + C;$$

$$\frac{d}{dx}(e^{kx}) = ke^{kx}, \qquad k\int e^{kx}\,dx = e^{kx} + C;$$

$$\frac{d}{dx}(\ln x) = \frac{1}{x}, \qquad \int \frac{1}{x}\,dx = \ln x + C;$$

[5.1] Before reading this chapter, the reader advised to reread Chapter 3.

$$\frac{d}{dx}(\sin kx) = k\cos kx,$$

$$k\int \cos kx\,dx = \sin kx + C;$$

$$\frac{d}{dx}(\cos kx) = -k\sin kx,$$

$$-k\int \sin kx\,dx = \cos kx + C;$$

$$\frac{d}{dx}(\tan x) = \frac{1}{\cos^2 x},$$

$$\int \frac{1}{\cos^2 x}\,dx = \tan x + C;$$

$$\frac{d}{dx}(\cot x) = -\frac{1}{\sin^2 x},$$

$$-\int \frac{1}{\sin^2 x}\,dx = \cot x + C;$$

$$\frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}},$$

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C;$$

$$\frac{d}{dx}(\arctan x) = \frac{1}{1+x^2},$$

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C.$$

Let us perform a few manipulations. In the first integral, we denote $n-1$ by m, that is, $n = m+1$, and rewrite it thus:

$$\int x^m\,dx = \frac{1}{m+1}x^{m+1} + C. \qquad (5.2.1)$$

It is clear that the formula is valid for all m except $m = -1$; for $m = -1$ the denominator vanishes and $x^{m+1} = x^0 = 1$, so that we arrive at an expression unsuitable for computations: $1/0 + C$. However, it is precisely in the case where $m = -1$, that is, for $\int (1/x)\,dx$ that we have the formula[5.2]

$$\int \frac{1}{x}\,dx = \ln x + C. \qquad (5.2.2)$$

This formula holds true only for positive values of x, since $\ln x$ is meaningful only for $x > 0$. For $x < 0$, $\ln x$ is meaningless, but $\ln(-x)$ is meaningful. Since

$$\frac{d\ln(-x)}{dx} = \frac{1}{-x}\times(-1) = \frac{1}{x},$$

for $x < 0$ we have $\int dx/x = \ln(-x) + C$. Both formulas for $\int dx/x$ can be combined into one:

$$\int \frac{dx}{x} = \ln|x| + C. \qquad (5.2.3)$$

This formula may be used for any domain of integration that does not contain $x = 0$.

The integral of the exponential function can be written thus:

$$\int e^{kx}\,dx = \frac{1}{k}e^{kx} + C. \qquad (5.2.4)$$

Similarly, for the sine and cosine we get

$$\int \sin kx\,dx = -\frac{1}{k}\cos kx + C, \qquad (5.2.5)$$

$$\int \cos kx\,dx = \frac{1}{k}\sin kx + C. \qquad (5.2.6)$$

We conclude with an example that shows the necessity of a careful examination of the function and the danger of a purely formal approach. We will evaluate the integral $I = \int_a^b dx/x^2$. The indefinite integral is given by the formula $\int dx/x^2 = -1/x + C$, whence

$$I = \int_a^b \frac{dx}{x^2} = -\frac{1}{x}\Big|_a^b = -\frac{1}{b} + \frac{1}{a} = \frac{b-a}{ab}. \qquad (5.2.7)$$

Since the integrand is positive, the result must be positive, too, provided that $b > a$. The answer is indeed positive if $b > a$ and both a and b have the same sign. However, for the integral $\int_{-1}^{1} dx/x^2$ formula (5.2.7) yields a clearly absurd result $I = -2$. The reason for this is that the integrand becomes infinitely large inside the domain of

[5.2] In Section 7.6 we will discuss the marked difference between formulas (5.2.1) and (5.2.2) for integrals of power functions.

integration (at $x = 0$), and at this point the $-1/x$ experiences an infinite jump (this function is the indefinite integral of $1/x^2$).

To find the true reason for all this, we must exclude a small neighbourhood of the "singular point" $x = 0$ from the domain of integration $-1 < x < 1$, that is, we must take the interval $\varepsilon_1 < x < \varepsilon_2$, where ε_1 and ε_2 are small positive numbers, and consider the sum

$$K = \int_{-1}^{-\varepsilon_1} \frac{dx}{x^2} + \int_{\varepsilon_2}^{1} \frac{dx}{x^2}.$$

It is natural to assume that integral I is obtained from K if we send ε_1 and ε_2 to zero. But formula (5.2.7) yields

$$K = \frac{1-\varepsilon_1}{\varepsilon_1} + \frac{1-\varepsilon_2}{\varepsilon_2} = -2 + \frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2}.$$

When ε_1 and ε_2 tend to zero, K tends to ∞.

In other cases an integral whose integrand becomes infinitely large within the domain of integration may assume a finite value. For instance,

$$\int_0^1 dx/\sqrt{x} = 2.$$

Indeed, the corresponding indefinite integral is given by the formula

$$\int \frac{dx}{\sqrt{x}} = \int x^{-1/2}\,dx = \frac{1}{1/2} x^{1/2} + C = 2\sqrt{x} + C,$$

whence

$$\int_{\varepsilon}^{1} \frac{dx}{\sqrt{x}} = 2\sqrt{x}\Big|_{\varepsilon}^{1} = 2 - 2\sqrt{\varepsilon},$$

which tends to 2 as $\varepsilon \to 0$.

Consideration of this kind must always be carried out if the integrand becomes infinitely large; here, however, we will not discuss this important topic any further (see Section 6.3).

Exercises

Evaluate the following definite integrals:

5.2.1. $\int_0^1 x^5\,dx$. 5.2.2. $\int_1^2 \frac{1}{x^2}\,dx$

5.2.3. $\int_0^{\pi} \sin x\,dx$. 5.2.4. $\int_0^1 \frac{dx}{x^2+1}$.

5.2.5. $\int_{-1}^{1} e^x\,dx$. 5.2.6. $\int_0^{\pi/4} \frac{dx}{\cos^2 x}$

5.3 General Properties of Integrals

In Sections 4.1 to 4.3 we established the formulas for the derivative of a sum (and difference) of functions, the derivative of a product (and ratio) of functions, and the derivatives of a composite function and of the inverse of a function. To each of these properties there corresponds a definite property referring to integrals. This section and Sections 5.4 and 5.5 are devoted to finding the "integral" analogs of the properties of derivatives considered.

For integrals we have the following formula:

$$\int [Af(x) + Bg(x)]\,dx = A\int f(x)\,dx + B\int g(x)\,dx. \qquad (5.3.1)$$

To prove it, we need only take the derivative of the expression on the right. If the formula is true, then we will arrive at the integrand. Indeed,

$$\left[A\int f(x)\,dx + B\int g(x)\,dx\right]' = A\left[\int f(x)\,dx\right]' + B\left[\int g(x)\,dx\right]' = Af(x) + Bg(x).$$

We have thus proved the validity of (5.3.1). It shows that *the integral of a sum of several terms splits up into the sum of the integrals of the separate summands*, and *any constant factors can be taken outside the integrrl sign.*

As an example to illustrate the above statement, we consider the integral of a *polynomial*:

$$\int (a_0x^k + a_1x^{k-1} + a_2x^{k-2} + \ldots + a_{k-2}x^2 + a_{k-1}x + a_k)\,dx = a_0 \int x^k\,dx + a_1 \int x^{k-1}\,dx + a_2 \int x^{k-2}\,dx + \ldots + a_{k-2} \int x^2\,dx + a_{k-1} \int x\,dx + a_k \int dx = \frac{a_0}{k+1} x^{k+1} + \frac{a_1}{k} x^k + \frac{a_2}{k-1} x^{k-1} + \ldots + \frac{a_{k-2}}{3} x^3 + \frac{a_{k-1}}{2} x^2 + a_k x + C. \quad (5.3.2)$$

Whence, for one thing,

$$\int_n^m (a_0x^k + a_1x^{k-1} + a_2x^{k-2} + \ldots + a_{k-2}x^2 + a_{k-1}x + a_k)\,dx = \frac{a_0}{k+1}(m^{k+1} - n^{k+1}) + \frac{a_1}{k}(m^k - n^k) + \frac{a_2}{k-1}(m^{k-1} - n^{k-1}) + \ldots + \frac{a_{k-2}}{3}(m^3 - n^3) + \frac{a_{k-1}}{2}(m^2 - n^2) + a_k(m-n).$$

For instance,

$$\int (2x^2 - 6x + 1)\,dx = \frac{2}{3}x^3 - 3x^2 + x + C,$$

$$\int_1^2 (2x^2 - 6x + 1)\,dx = \frac{2}{3}(8-1) - 3(4-1) + (2-1) = -3\frac{1}{3}.$$

It is possible, under the integral sign, to make a change of variable and pass to a new and more convenient variable. We will discuss this question in detail in Section 5.5, while here we examine a number of simple examples.

1. Find $\int (ax + b)^n dx$, with $n \neq -1$.

For the new variable, call it z, we take the expression in the parentheses:

$$ax + b = z. \quad (5.3.3)$$

In so doing, we also have to pass from differential dx to differential dz. From (5.3.3) we get $dz = a\,dx$, or $dx = a^{-1}dz$. Thus,

$$\int (ax+b)^n\,dx = \int z^n \frac{dz}{a} = \frac{1}{a}\int z^n\,dz = \frac{z^{n+1}}{a(n+1)} + C = \frac{(ax+b)^{n+1}}{a(n+1)} + C. \quad (5.3.4)$$

The correctness of the result is easily seen if we compute the derivative of the right member:

$$\frac{d}{dx}\left[\frac{(ax+b)^{n+1}}{a(n+1)} + C\right] = \frac{d}{dx}\left[\frac{(ax+b)^{n+1}}{a(n+1)}\right] = \frac{n+1}{a(n+1)}(ax+b)^n \frac{d}{dx}(ax+b) = \frac{(ax+b)^n}{a}\,a = (ax+b)^n.$$

2. In similar fashion, in the integral $\int \frac{dx}{ax+b}$ we can make the change of variable $z = ax + b$, $dx = a^{-1}\,dz$, and hence

$$\int \frac{dx}{ax+b} = \frac{1}{a}\int \frac{dz}{z} = \frac{1}{a}\ln|z| + C = \frac{\ln|ax+b|}{a} = C. \quad (5.3.5)$$

When we are dealing with such simple examples in practical situations, the transformations are ordinarily carried out without introducing separate designations for the new intermediate variables. For example, instead of (5.3.4) one writes

$$\int (ax+b)^n\,dx = \int (ax+b)^n \frac{1}{a}\,d(ax+b) = \frac{1}{(n+1)a}(ax+b)^{n+1} + C.$$

We now give a more complicated example of an integral that can be reduced to the integrals we already know via algebraic transformations.

Consider the integral $\int \frac{dx}{(x-a)(x-b)}$. Note that

$$\frac{1}{x-a} - \frac{1}{x-b} = \frac{a-b}{(x-a)(x-b)}, \quad (5.3.6)$$

from which it follows that

$$\frac{1}{(x-a)(x-b)} = \frac{1}{a-b}\left(\frac{1}{x-a} - \frac{1}{x-b}\right).$$

whence

$$\int \frac{dx}{(x-a)(x-b)} = \frac{1}{a-b}\int\left(\frac{1}{x-a}-\frac{1}{x-b}\right)dx$$
$$= \frac{1}{a-b}[\ln|x-a| - \ln|x-b|] + C$$
$$= \frac{1}{a-b}\ln\left|\frac{x-a}{x-b}\right| + C. \qquad (5.3.7)$$

(Formula (5.3.7) can be used in any domain of integration that *does not include* the points $x = a$ and $x = b$.) An integral of any algebraic fraction, that is, the ratio $P_n(x)/Q_m(x)$ of polynomials of degrees n and m, can be expressed as a combination of elementary functions (algebraic functions, logarithms, and arctangents). Several simple examples of this type are given in the exercises below.

Exercises

Find the following (indefinite) integrals:

5.3.1. $\int (3x^2-2x+1)\,dx$. 5.3.2. $\int (4x^4-3x^3+x^2-x)\,dx$. 5.3.3. $\int x(x-1)^2\,dx$.

5.3.4. $\int \frac{x^2+2x+3}{x}\,dx$. 5.3.5. $\int \frac{2x-1}{x-1}\,dx$.

5.3.6. $\int \frac{ax+b}{cx+d}\,dx$. [*Hint.* Employ the fact that $\frac{ax+b}{cx+d} = \frac{a}{c} + \frac{bc-ad}{c(cx+d)}$.] 5.3.7. $\int \frac{x\,dx}{(x-2)(x-3)}$. [*Hint.* Employ the identity $\frac{x}{(x-2)(x-3)} = \frac{A}{x-2} + \frac{B}{x-3}$; the numbers A and B are found via the method of undetermined coefficients, that is, by equating the coefficients of identical powers of x after clearing fractions.]

5.3.8. $\int \frac{dx}{(x-1)(x-2)}$. 5.3.9. $\int \frac{x+1}{x^2-3x+2}\,dx$.

5.3.10. $\int \frac{x+2}{x^3+x}\,dx$. [*Hint.* Use the identity $\frac{x+2}{x^3+x} = \frac{x+2}{x(x^2+1)} = \frac{A}{x} + \frac{B}{x^2+1} + \frac{Cx}{x^2+1}$, where the coefficients A, B, and C are found via the method of undetermined coefficients (see the hint to Exercise 5.3.7); to find $\int \frac{Cx\,dx}{x^2+1}$, it proves convenient to denote x^2+1 by z.]

5.4 Integration by Parts

Let $f(x)$ and $g(x)$ be two distinct functions of the variable x. The rule for finding the derivative of a product yields

$$\frac{d}{dx}(fg) = g\frac{df}{dx} + f\frac{dg}{dx}. \qquad (5.4.1)$$

This enables us to write

$$fg = \int\left(g\frac{df}{dx} + f\frac{dg}{dx}\right)dx$$
$$= \int f\frac{dg}{dx}\,dx + \int g\frac{df}{dx}\,dx. \qquad (5.4.2)$$

We can see the validity of (5.4.2) by taking the derivatives of the left and right members to get the true equation (5.4.1).

Let us rewrite (5.4.2) as

$$\int f\frac{dg}{dx}\,dx = fg - \int g\frac{df}{dx}\,dx,$$

or, which is the same,

$$\int f\,dg = fg - \int g\,df. \qquad (5.4.3)$$

What is the meaning of this formula? When evaluating an integral, there is unfortunately no rule that expresses the integral of a product of two functions in terms of the integrals of each of the factors. However, if in the product of two functions fw the integral of one of the factors is known, or

$$\int w\,dx = g,\quad w = \frac{dg}{dx}, \qquad (5.4.4)$$

then it becomes possible to express the integral $\int fw\,dx$ in terms of an integral involving the derivative df/dx. Using (5.4.4), we rewrite (5.4.3) in the form

$$\int fw\,dx = f\left(\int w\,dx\right) - \int\left(\int w\,dx\right)\frac{df}{dx}\,dx. \qquad (5.4.5)$$

Since $\int w\,dx = g$, it follows that the last integral in (5.4.5) is simply $\int g\,(df/dx)\,dx$. Sometimes it is simpler than the original integral $\int fw\,dx$ or reduces to a known integral. If f is a power function or a polynomial, df/dx is also a power function or a polynomial whose power is that of f *minus* unity.

Formula (5.4.3) or (5.4.5) is called the formula for ***integration by parts.*** From this formula follows a similar relationship for definite integrals:

$$\int_a^b f(x)\,dg(x) = f(x)\,g(x)\Big|_a^b - \int_a^b g(x)\,df(x)$$

$$= [f(b)\,g(b) - f(a)\,g(a)] - \int_a^b g(x)\,df(x). \qquad (5.4.3a)$$

Here are some examples illustrating the method of integration by parts.

1. Find $\int xe^x\,dx$. Put $f = x$. Then $w = dg/dx = e^x$ and $e^x\,dx = dg$, $g = \int e^x\,dx = e^x$, and $df = dx$. By formula (5.4.3)

$$\int xe^x\,dx = xe^x - \int e^x\,dx = xe^x - e^x + C$$

$$= e^x(x-1) + C.$$

2. Find $\int x^2e^x\,dx$. Put $f = x^2$. Then $w = dg/dx = e^x$, $e^x\,dx = dg$, $g = \int e^x\,dx = e^x$, and $df = 2x\,dx$. By formula (5.4.3),

$$\int x^2e^x\,dx = x^2e^x - 2\int xe^x\,dx,$$

whence, using the result of the first example, we get

$$\int x^2e^x\,dx = x^2e^x - 2xe^x + 2e^x + C$$

$$= (x^2 - 2x + 2)\,e^x + C.$$

To find $\int P_n(x)\,e^{kx}\,dx$, where $P_n(x)$ is a polynomial of degree n, we have to perform integration by parts n times. We then get $Q_n(x)\,e^{kx}$, where $Q_n(x)$ is a polynomial of degree n. Knowing this, we need not perform integration by parts n times, but can write down directly the coefficients of the polynomial $Q_n(x)$.

Let us take the same Example 2. Find $\int x^2e^x dx$. We write the equation $\int x^2e^x dx = Q_2(x)\,e^x + C$ with the (still) unknown coefficients of the polynomial $Q_2(x)$:

$$\int x^2e^x\,dx = (a_2x^2 + a_1x + a_0)\,e^x + C. \qquad (5.4.6)$$

Taking the derivatives of both sides of (5.4.6), we get

$$x^2e^x = (2a_2x + a_1)\,e^x + (a_2x^2 + a_1x + a_0)\,e^x.$$

or

$$x^2e^x = [x^2a_2 + x\,(2a_2 + a_1) + (a_1 + a_0)]\,e^x.$$

Equate the coefficients of identical powers of x in the polynomials on the right and left to get $a_2 = 1$, $2a_2 + a_1 = 0$, $a_1 + a_0 = 0$, whence $a_1 = -2$ and $a_0 = 2$. Finally as before, we get

$$\int x^2e^x\,dx = (x^2 - 2x + 2)\,e^x + C.$$

By a similar technique we can find the integrals of the functions $P_n(x)\cos kx$ and $P_n(x)\sin kx$, with $P_n(x)$ a polynomial. In both cases the answer is of the form $Q_n(x)\cos kx + R_n(x)\sin kx$, where $Q_n(x)$ and $R_n(x)$ are polynomials of degree n (or less than n). Examples of this kind are given in the exercises below.

Exercises

Find the following integrals:

5.4.1. $\int x\cos x\,dx$. **5.4.2.** $\int \ln x\,dx$.

5.4.3. $\int x^2\sin 2x\,dx$. **5.4.4.** $\int x^3e^{-x}dx$.

5.4.5. $\int (x^2+x+1)\cos x\,dx$. **5.4.6.** $\int (2x^2+1)\times \cos 3x\,dx$. [*Hint.* Exercise 5.4.6. is studied in detail in Hints, Answers, and Solutions at the end of the book; Examples 5.4.3 to 5.4.5 are handled similarly to Example 5.4.6]

5.4.7. $\int \arcsin x\,dx$. **5.4.8.** $\int \arctan x\,dx$.

5.4.9. $\int e^{2x}\sin 3x\,dx$. **5.4.10.** $\int e^x\cos 2x\,dx$.

5.5 Change of Variables in Integration

We have already mentioned the general method of *changing the variables* in integration, or *integration by substitution* (of a new unknown for the initial unknown). If in the integral $\int F(x)\,dx$ the integrand $F(x)$ can be represented in the form

$$F(x) = f(\varphi(x))\,\varphi'(x),$$

where $\varphi(x)$ is a function of variable x, we can denote $\varphi(x)$ by a single letter z and take z as the new variable. Since $\varphi'(x)\,dx = dz$, we arrive at the integral $\int f(z)\,dz$, which may prove to be tabulated (that is, it may be found among the integrals of Section 5.2) or even be simpler than the initial integral. This general method is widely used in integral calculus.

For instance, the integration of many functions involving radicals and trigonometric functions may be reduced, by means of an appropriate change of variable, to the integration of polynomials or algebraic fractions with integral powers. Let us consider several examples.

Find $\int x\sqrt{x+1}\,dx$. We make a change of variable: $z = \sqrt{x+1}$, $x+1 = z^2$, whence $x = z^2 - 1$ and $dx = 2z\,dz$. Passing to the new variable in the integral, we obtain

$$\int x\sqrt{x+1}\,dx = \int (z^2-1)\,z\,2z\,dz$$
$$= 2\int (z^4 - z^2)\,dz = 2\frac{z^5}{5} - 2\frac{z^3}{3} + C$$
$$= \frac{2}{5}\sqrt{(x+1)^5} - \frac{2}{3}\sqrt{(x+1)^3} + C.$$

In the integral $\int \frac{dx}{a^2+x^2}$ the appropriate change of variable is $x = at$, that is, $t = x/a$; then $dx = a\,dt$ and we have

$$\int \frac{dx}{a^2+x^2} = \int \frac{a\,dt}{a^2+a^2t^2} = \int \frac{a\,dt}{a^2(1+t^2)}$$
$$= \frac{1}{a}\int \frac{dt}{1+t^2} = \frac{1}{a}\arctan t + C$$
$$= \frac{1}{a}\arctan\frac{x}{a} + C.$$

In the integral $\int \frac{dx}{\sqrt{(a^2-x^2)^3}}$ the appropriate change of variable is $x = a\cos t$, that is, $t = \arccos(x/a)$; then $dx = -a\sin t\,dt$, $a^2 - x^2 = a^2(1 - \cos^2 t) = a^2\sin^2 t$, and we obtain

$$\int \frac{dx}{\sqrt{(a^2-x^2)^3}} = \int \frac{-a^2\sin t\,dt}{\sqrt{(a^2\sin^2 t)^3}}$$
$$= -\int \frac{a\sin t\,dt}{a^3\sin^3 t} = -\frac{1}{a^2}\int \frac{dt}{\sin^2 t}$$
$$= \frac{1}{a^2}\cot t + C = \frac{1}{a^2}\cot\left(\arccos\frac{x}{a}\right) + C.$$

This expression can be simplified still further (see Exercise 5.5.10).

An almost similar substitution can be applied to the "almost tabular" integral $\int \frac{dx}{\sqrt{a^2-x^2}}$ (which by the substitution $x = az$, $dx = a\,dz$ is reduced to the known result $\int \frac{dz}{\sqrt{1-z^2}} = \arcsin z + C$). Indeed, if $x = a\sin t$, $dx = a\cos t\,dt$, and $\sqrt{a^2-x^2} = a\cos t$, then

$$\int \frac{dx}{\sqrt{a^2-x^2}} = \int \frac{a\cos t\,dt}{a\cos t} = \int dt = t + C$$
$$= \arcsin\frac{x}{a} + C,$$

since $t = \arcsin(x/a)$.

The integral $\int \frac{dx}{\sqrt{a^2+x^2}}$ can be evaluated in a similar manner, only instead of trigonometric functions we must employ the so-called ***hyperbolic sine*** $\sinh t = (1/2)(e^t - e^{-t})$ and the ***hyperbolic cosine*** $\cosh t = (1/2)(e^t + e^{-t})$ (for more details on these functions see Section 14.4). We can easily see (verify this!) that $\cosh^2 t - \sinh^2 t = 1$, $(\sinh t)' = \cosh t$, and $(\cosh t)' = \sinh t$. Therefore, if we put $x = a\sinh t$, then $dx = a\cosh t\,dt$ and $\sqrt{a^2+x^2} = a\cosh t$. Thus, our integral takes the form

$$\int \frac{dx}{\sqrt{a^2+x^2}} = \int \frac{a\cosh t\,dt}{a\cosh t}$$
$$= \int dt = t + C = \operatorname{arsinh}\left(\frac{x}{a}\right) + C,$$

where arsinh z is the inverse hyperbolic sine defined by the condition: if $t = \text{arsinh } z$, then $z = \sinh t$.

We leave to the reader the task of bringing this example to the final result (see Exercise 5.5.11).

Finally, we give an example of an integral that *cannot* be represented in terms of a finite number of elementary functions. The integral is

$$f(x) = \int e^{-x^2}\,dx. \tag{5.5.1}$$

The proof that it cannot be expressed in terms of a finite number of elementary functions (power functions, including functions with fractional powers, the exponential function, the logarithm, the trigonometric functions, and the inverse trigonometric functions) is extremely complicated and we will not give it here.

The integral (5.5.1) constitutes a function whose properties can be studied. The definition of $f(x)$ implies that

$$\frac{df}{dx} = e^{-x^2}. \tag{5.5.2}$$

Since $f'(x) = e^{-x^2}$ is positive for arbitrary x, it follows that $f(x)$ is an increasing function. The derivative of $f(x)$ has a maximum at $x = 0$, where it is equal to unity, which means that $f(x)$ has a maximum angle of the tangent line with the x axis at $x = 0$ (the angle is 45°). For large absolute values of x (positive or negative), the derivative df/dx is very small, whereby the function is almost constant for such values. The graph of the function $\Phi(x) = \int_0^x e^{-x^2}\,dx$ is depicted in Figure 5.5.1 (for the sake of definiteness, the lower limit of integration has been chosen to be equal to zero). The fact that $\Phi(\infty) \simeq 0.885$ is equal to $\sqrt{\pi}/2$ is given without proof in this book.

Extensive tables have been compiled for the function $\Phi(x)$, and so computations involving integral (5.5.1), are

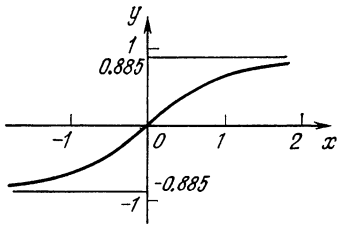


Figure 5.5.1

no more complicated than, say, those involving trigonometric functions.

Note that often, when using various techniques, one obtains distinct expressions for one and the same integral (see Exercise 5.5.6). This should not dismay the reader. If the computations are correct, such expressions should differ by a constant only. The results are identical when evaluating a definite integral (for any limits of integration).

We see that computing an integral is much more complicated than finding derivatives: here one must employ various ingenious techniques (integration by parts, change of variable), and often it is not clear at first glance what technique to apply (e.g. how to choose the new variable). Moreover, in some cases (as in the case with integral (5.5.1)) there is not a single technique that enables finding the integral in closed form, that is, expressing it in terms of a finite number of elementary functions, and there is not a single general method that could establish whether a given integral can be expressed in such a form.[5.3] Integration techniques can-

[5.3] Often we encounter the case where the formula for a definite integral $\int_a^b f(x)\,dx$ with known limits of integration a and b exists, but the general formula for the corresponding indefinite integral cannot be found. For instance, as noted earlier, $\int_0^\infty e^{-x^2}\,dx = \sqrt{\pi}/2$, where $\int_0^\infty e^{-x^2}dx = \lim_{M\to\infty} \int_0^M e^{-x^2}\,dx$,

not, in principle, be reduced to algorithms, while with differentiation techniques (i.e. finding the derivatives) this is possible. Often it is easier to use special tables when one has to find integrals of functions. Such tables of integrals contain lists of (indefinite) integrals classified according to the type of integrand and other characteristics that enables finding the sought integral easily, as well as certain types of definite integrals (with limits of integration 0 and ∞, 0 and π, $-\infty$ and ∞, etc.).[5.4] When an integral cannot be evaluated, one has to resort to (approximate) numerical integration.

Exercises

Find the following integrals:

5.5.1. $\int \cos(3x-5)\,dx$. 5.5.2. $\int \sin(2x+1)\,dx$. 5.5.3. $\int \sqrt{3x-2}\,dx$. 5.5.4. $\int \frac{x\,dx}{x+\sqrt{x}}$. [*Hint.* Make the substitution $\sqrt{x}=z$.]

5.5.5. $\int \frac{x\,dx}{\sqrt{x^2-5}}$. [*Hint.* Make the substitution $x^2-5=z$.] 5.5.6. $\int \sin^3 x \cos x\,dx$. [*Hint.* Make the substitution $\sin x=z$ or $\cos x=u$.]

5.5.7. $\int \frac{\cos^3 x}{\sin^4 x}\,dx$. [*Hint.* Make the substitution $\sin x=z$.] 5.5.8. $\int \tan x\,dx$.

5.5.9. $\int \frac{dx}{\sqrt{a^2-x^2}}$.

5.5.10. Prove that $\int \frac{dx}{\sqrt{(a^2-x^2)^3}} = \frac{x}{a^2\sqrt{a^2-x^2}}+C$.

5.5.11. Find the integral $\int \frac{dx}{\sqrt{a^2+x^2}}$.

5.5.12. Find the integral $\int \frac{dx}{(x^2+a^2)^2}$.

5.6 Change of Variable in a Definite Integral

We consider an example. Let it be required to calculate the definite integral $\int_p^q (ax+b)^2\,dx$. We can do as follows. First calculate the indefinite integral $\int (ax+b)^2\,dx$ and then form the difference of its values for $x=q$ and $x=p$.

To compute $\int (ax+b)^2\,dx$ we make a change of variable using the formula $z=ax+b$. Then $dz=a\,dx$ and

$$\int (ax+b)^2\,dx=\frac{1}{a}\int z^2\,dz=\frac{z^2}{3a}=\frac{(ax+b)^3}{3a}.$$

Therefore

$$\int_p^q (ax+b)^2\,dx=\frac{(ax+b)^3}{3a}\Big|_p^q$$
$$=\frac{(aq+b)^3-(ap+b)^3}{3a}.$$

However, it is possible to do otherwise. Let us determine how z will vary when x varies from p to q. Since z and x are related as $z=ax+b$, it follows that as x varies from p to q, the new variable z will vary from $ap+b$ to $aq+b$. Hence,

$$\int_p^q (ax+b)^2\,dx=\frac{1}{a}\int_{ap+b}^{aq+b} z^2\,dz$$
$$=\frac{z^3}{3a}\Big|_{ap+b}^{aq+b}=\frac{(aq+b)^3-(ap+b)^3}{3a}.$$

When evaluating integrals, it is convenient to do just that way, that is, when making a change of variable, to find the new limits of integration at the same time. This will obviate a return to the old variable in the expression for the indefinite integral.

---

while the corresponding indefinite integral cannot be calculated. (One method of evaluating a definite integral that does not rely on formulas for the corresponding indefinite integral is discussed in Section 17.3.)

[5.4] See, for instance, I.N. Bronsthein and K.A. Semendyayew, *A Guide-Book in Mathematics*, Deutsch, Frankfurt, 1971; H.B. Dwight, *Mathematical Tables*, McGraw-Hill, New York, 1941; and I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products*, 4th ed., Academic Press, New York, 1966.

Let us consider some examples.

1. Calculate $\int_0^1 \frac{dx}{(2-x)^3}$. Note from the start that the function $(2-x)^{-3}$ assumes positive values as x varies from 0 to 1, and therefore the integral is positive. At the same time, the denominator in this domain of integration does not vanish, so that the integrand is finite throughout the domain. Make the change of variable $2-x=y$, with $dx=-dy$. Then $y=2$ at $x=0$ and $y=1$ at $x=1$, so that

$$\int_0^1 \frac{dx}{(2-x)^3} = -\int_2^1 \frac{dy}{y^3}. \qquad (5.6.1)$$

On the right-hand side of (5.6.1) the limits of integration are given for y. The reader may wonder about the minus sign in the last equation. Indeed, on the right and left we have integrals of positive functions, so why is the right-hand side of (5.6.1) positive? The point is that the lower limit of integration on the right is greater than the upper limit. Since an integral changes sign upon interchanging the limits of integration, (5.6.1) may be written as follows:

$$\int_0^1 \frac{dx}{(2-x)^3} = \int_1^2 \frac{dy}{y^3}.$$

Now, in the right-hand integral, the upper limit exceeds the lower one and it is clear that the integral on the right is positive. The computation can now be completed with ease:

$$\int_1^2 \frac{dy}{y^3} = -\frac{1}{2y^2}\Big|_1^2 = -\frac{1}{8}+\frac{1}{2}=\frac{3}{8}.$$

2. In Section 5.5 we considered the function $\Phi(x)=\int_0^x e^{-x^2}dx$. One often has to deal with the function $\varphi(a)=\int_0^a e^{-kx^2}dx$, where k is a constant. We will show that there is a simple relationship between the functions Φ and φ.

In the expression for $\varphi(a)$ make the change of variable $kx^2=t^2$. From this we find that $\sqrt{k}x=t$, $x=t/\sqrt{k}$, and $dx=(1/\sqrt{k})\,dt$. It is clear that $t=0$ at $x=0$ and $t=a\sqrt{k}$ at $x=a$. And so we get

$$\varphi(a)=\int_0^{a\sqrt{k}} e^{-t^2}\frac{dt}{\sqrt{k}}=\frac{1}{\sqrt{k}}\int_0^{a\sqrt{k}} e^{-t^2}dt$$
$$=\frac{1}{\sqrt{k}}\Phi(a\sqrt{k}).$$

Consequently, for an arbitrary value of the independent variable x,

$$\varphi(x)=\frac{1}{\sqrt{k}}\Phi(x\sqrt{k}). \qquad (5.6.2)$$

If we have a table of values of the function $\Phi(x)$, it is possible to find the values of $\varphi(x)$ for any value of k, and vice versa, knowing the values of $\varphi(x)$ which correspond to a definite value of k, we can find, via (5.6.2), the value of $\Phi(z)$ for any z (in (5.6.2) we are forced to put $z=x\sqrt{k}$).[5.5]

3. In Chapter 3 we saw that the definite integral has dimensions if the integrand and the limits of integration have. It is often convenient, however, to reduce the integral to a *dimensionless* form by taking all factors having dimensions outside the integral sign. What remains to be done is to consider

[5.5] Note that more customary are not tables of the values of $\Phi(x)=\int_0^x e^{-x^2}dx$, of which we spoke at the end of Section 5.5, but those of the function $\varphi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dx$, which is closely related to $\Phi(x)$ (see formula (5.6.2), where in this case we must put $k=1/2$). The function $\varphi(x)$ plays a very important role in the more advanced section of mathematics, the *theory of probability*, of which we speak in brief in the Conclusion of this book.

(transform) a numerical (dimensionless) integral. Let us see how to transform an integral to dimensionless form.

Suppose we have an integral $\int_a^b f(x)\,dx$, where, for the sake of simplicity, we assume $f(x)$ to be non-negative. Denote by $f_{\max}$ the greatest value of the function $f(x)$ on the domain of integration and take it outside the integral sign:[5.6]

$$\int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{f_{\max}} f_{\max}\,dx = f_{\max}\int_a^b \frac{f(x)}{f_{\max}}\,dx. \qquad (5.6.3)$$

It is obvious that in the last integral the integrand $f(x)/f_{\max}$ is dimensionless, since $f(x)$ and $f_{\max}$ have the same dimensions. Let us pass to dimensionless limits of integration. To do this, make the change of variable

$$z = \frac{x-a}{b-a}, \text{ or } x = a + z(b-a). \qquad (5.6.4)$$

We see that z is dimensionless. Since $dx = (b-a)\,dz$ and $z = 0$ at $x = a$ and $z = 1$ at $x = b$, it follows that the integral on the right-hand side of (5.6.3) takes the form

$$\int_a^b \frac{f(x)}{f_{\max}}\,dx = (b-a)\int_0^1 \frac{f(a+z(b-a))}{f_{\max}}\,dz. \qquad (5.6.5)$$

Set $f(a + z(b-a))/f_{\max} = \varphi(z)$. Then from (5.6.5) we have

$$\int_a^b \frac{f(x)}{f_{\max}}\,dx = (b-a)\int_0^1 \varphi(z)\,dz.$$

and, finally,

$$\int_a^b f(x)\,dx = f_{\max}(b-a)\int_0^1 \varphi(z)\,dz, \qquad (5.6.6)$$

where $\int_0^1 \varphi(z)\,dz$ is a dimensionless quantity.

If $f(x)$ varies but slightly on the domain of integration, then $f(x)/f_{\max} \simeq 1$ (but, of course, $f(x)/f_{\max} \leqslant 1$, since $f(x) \leqslant f_{\max}$), and so $\varphi(z) \simeq 1$ (but $\varphi(z) \leqslant 1$) and $I = \int_0^1 \varphi(z)\,dz \simeq 1\int_0^1 dz = 1$ (but $I \leqslant 1$). Then the dimensionless factor $I = \int_0^1 \varphi(z)\,dz$ is of the order of unity and the value of integral (5.6.3) is determined mainly by the product $f_{\max}(b-a)$.

Let us consider a simple example: the free fall of a body in the course of time t_0. The distance of free fall is equal to $\int_0^{t_0} v(t)\,dt$. The acceleration of the falling body is g, and, therefore, the rate of fall, or velocity, is $v(t) = gt$ (we assume that $v(0) = 0$ and that the maximum velocity $v_{\max}$ is attained at time t_0, or $v_{\max} = gt_0$). Note that the maximum here is not due to any decrease in velocity after $t = t_0$ but simply to the fact that times exceeding t_0 and corresponding to velocities greater than gt_0 lie outside the domain of integration $0 \leqslant t \leqslant t_0$. We introduce the notation $z = t/t_0$, $\varphi(z) = v/v_{\max} = gt/gt_0 = t/t_0 = z$. The result is

$$s = \int_0^{t_0} v(t)\,dt = v_{\max}t_0\int_0^1 z\,dz = gt_0^2\int_0^1 z\,dz = \frac{1}{2}gt_0^2.$$

[5.6] We have assumed that $f(x) \geqslant 0$ and $f_{\max} \neq 0$. In the case of a function that changes sign or is negative, for $f_{\max}$ we must take the greatest of the absolute values of function f, or $|f|_{\max}$.

We see that in the given case the dimensionless factor $I = \int_0^1 \varphi(z)\,dz$ (whose order of magnitude is generally equal to unity but whose magnitude is always smaller than unity) is equal to 0.5, which agrees fairly well with our rough estimate.

5.7 Integrating Functions Dependent on a Parameter

In Section 4.12 we discussed briefly the problem of differentiating functions dependent on a parameter. Here we wish to investigate the question of integration of such functions. The definite integral $I = \int_a^b f(x)\,dx$ is a number, but if the integrand depends on a parameter s, that is, $f = f(x, s)$, the number will depend on this parameter, too, that is, we will have $I = I(s)$. For instance, it is clear that

$$\int_{-1}^{+1} (kx + k^{-1})\,dx = \left(\frac{1}{2}kx^2 + \frac{1}{k}x\right)\Big|_{-1}^{1} = \frac{2}{k}, \qquad (5.7.1)$$

$$\int_0^1 \sin \lambda x\,dx = \frac{1}{\lambda}\int_0^{\lambda} \sin y\,dy = \frac{1}{\lambda}(-\cos y)\Big|_0^{\lambda} = \frac{1-\cos\lambda}{\lambda}, \qquad (5.7.2)$$

$$\int_0^1 (s+1)\,x^s\,dx = (s+1)\,\frac{x^{s+1}}{s+1}\Big|_0^1 = 1$$

for $s > -1$. (5.7.3)

We see that integral (5.7.1) depends on parameter k, integral (5.7.2) depends on number λ, while integral (5.7.3) seems to depend on s, but this dependence is fictitious, since $I(s) \equiv 1$ is independent of s, that is, $I(s)$ is a function of s that equals unity identically. Example (5.7.3) is also instructive due to the condition that s must be greater than -1 (without this condition the result will be simply incorrect; we will return to this condition later).

The definition (3.2.4) of a definite integral reduces an integral to a sum of a large number of small summands. If the integrand $f(x) = f(x, s)$ (the function $v(t)$ in the notation of formula (3.2.4) or the function $y(x)$ in the notation of formula (3.2.6)) depends on a parameter s, all the summands will also depend on this parameter. Since the sum of a finite number of functions (in our case, functions of parameter s) can be differentiated and integrated term by term, the function

$$\int_a^b f(x, s)\,dx = I(s) \qquad (5.7.4)$$

can also be differentiated and integrated with respect to s under the integral sign:

$$I'(s) = \frac{d}{ds}\int_a^b f(x, s)\,dx = \int_a^b \frac{\partial f(x, s)}{\partial s}\,dx, \qquad (5.7.5a)$$

$$\int_\sigma^\tau I(s)\,ds = \int_\sigma^\tau \left[\int_a^b f(x, s)\,dx\right] ds = \int_a^b \left[\int_\sigma^\tau f(x, s)\,ds\right] dx, \qquad (5.7.5b)$$

where under the integral sign on the right-hand side of (5.7.5a) is the derivative of $f(x, s)$ with respect to s calculated on the assumption that x is kept constant, what is known as the partial derivative of $f(x, s)$ with respect to s (see Section 4.12). (To prove the validity of the rules (5.7.5a) and (5.7.5b) of differentiation and integration under the integral sign, it suffices to replace the integral $I(s)$ with the approximating "integral sum," of the type used in formulas (3.2.4) and (3.2.6), to differentiate or integrate this sum term by term, and then pass to

the limit as the number of summands grows without limit). For instance,

$$\frac{d}{dk}\int_{-1}^{1}(kx+k^{-1})\,dx=\int_{-1}^{1}\frac{\partial(kx+k^{-1})}{\partial k}\,dx$$

$$=\int_{-1}^{1}(x-k^{-2})\,dx$$

and

$$\int_{\alpha}^{\beta}\left(\int_{0}^{1}\sin\lambda x\,dx\right)d\lambda=\int_{0}^{1}\left(\int_{\alpha}^{\beta}\sin\lambda x\,d\lambda\right)dx$$

$$=\int_{0}^{1}\left(-\frac{\cos\lambda x}{\lambda}\right)_{\lambda=\alpha/x}^{\lambda=\beta/x}dx$$

$$=\int_{0}^{1}\left(\frac{\cos\alpha}{\alpha}-\frac{\cos\beta}{\beta}\right)x\,dx.$$

Difficulties may arise when the integral (5.7.4) is "improper," that is, when either the function $f(x)$ in the domain of integration becomes infinite or the domain itself is infinitely large (here, say, $\int_{a}^{\infty}f(x)\,dx=\lim_{M\to\infty}\int_{a}^{M}f(x)\,dx$; cf. (7.5.5)). For instance, for $s<0$ the integral (5.7.3) becomes an improper integral, since the function $(s+1)\,x^{s}$ for such values of x become infinite at $x=0$; for $0\geqslant s>-1$ this integral assumes a constant value equal to unity; and for $s<-1$ the integral itself becomes infinitely large. But if, say, $I(s)=\int_{a}^{\infty}f(x,\,s)\,dx$ is "normally convergent," that is, there exists a function $F(x)$ that is greater than $|f(x,\,s)|$ and such that $\int_{0}^{\infty}F(x)\,dx$ is finite, then we can deal with $I(s)$ as if it were an ordinary, "proper," integral (for instance, formulas (5.7.5a) and (5.7.5b) remain valid for this integral). For example, since $|(\sin\lambda x)/x^{2}|\leqslant 1/x^{2}$ and $\int_{1}^{\infty}dx/x^{2}=(-1/x)\,|_{1}^{\infty}=1<\infty$, we can state that $I(\lambda)=\int_{1}^{\infty}\frac{\sin\lambda x}{x^{2}}\,dx$ is a normally convergent integral and, say, $I'(\lambda)=\int_{1}^{\infty}\frac{\partial[(\sin\lambda x)/x^{2}]}{\partial\lambda}\,dx=\int_{1}^{\infty}\frac{\cos\lambda x}{x}\,dx$.

If integral (5.7.4) is not normally convergent, the direct application of rules (5.7.5a) and (5.7.5b) may lead to mistakes (here we may even encounter a case where the integral $I(s)$ is finite, while, say, its derivative, constructed according to (5.7.5a), $I'(s)=\int_{a}^{b}\frac{\partial f(x,\,s)}{\partial s}\,dx$, is infinite, or the other way round).

For instance, we are sure that $\int_{0}^{\infty}\frac{dx}{x+s}=\ln(x+s)\,\Big|_{x=0}^{x=\infty}=\infty$ (we assume that s is positive). But if we find the derivative of the integrand with respect to s, we arrive at the "finite" derivative $I'(s)$ of the "infinite" integral $I(s)$:

$$\frac{d}{ds}\int_{0}^{\infty}\frac{dx}{x+s}=\int_{0}^{\infty}\frac{\partial\,[1/(x+s)]}{\partial x}\,dx$$

$$=\int_{0}^{\infty}-\frac{1}{(x+s)^{2}}\,dx=\frac{1}{x+s}\,\Big|_{x=0}^{x=\infty}=-\frac{1}{s}.$$

To clarify the meaning of this paradoxical result, we proceed as follows. Let us "cut off" the singularity in the integral $I(s)=\int_{0}^{\infty}\frac{dx}{x+s}$, that is, replace the integral with the (proper) integral $I_{N}(s)=\int_{0}^{N}\frac{dx}{x+s}$, where N is large. We then study the asymptotic behavior of $I_{N}(s)$ as $N\to\infty$. Obvi-

ously,

$$I_N(s) = \int_0^N \frac{dx}{x+s} = \ln(x+s)\Big|_{x=0}^{x=N}$$
$$= \ln(N+s) - \ln s = \ln(1+N/s),$$

whence

$$I'_N(s) = \frac{1}{N+s} - \frac{1}{s},$$

which agrees with (5.7.4a) (since $I_N(s)$ is a proper integral):

$$I'_N(s) = \int_0^N \frac{\partial}{\partial s}\left(\frac{1}{x+s}\right) dx = \int_0^N \frac{-1}{(x+s)^2} dx$$
$$= \frac{1}{x+s}\Big|_{x=0}^{x=N} = \frac{1}{N+s} - \frac{1}{s}.$$

If we now send N to infinity, we will see that $I_N(s)$ tends to infinity and $I'_N(s)$ tends to $-1/s$, in complete agreement with the result we arrived at above.

The above-described procedure of cutting off the singularities often helps understand the meaning of results related to nonnormally convergent integrals or even divergent integrals, results that at first glance seem paradoxical. For instance, notwithstanding the fact that $I(s) = \int_0^\infty \frac{dx}{x+s} = \infty$, the difference between the integrals $I(s_2)$ and $I(s_1)$ proves to be finite:

$$I(s_2) - I(s_1) = \int_0^\infty \frac{dx}{x+s_2} - \int_0^\infty \frac{dx}{x+s_1}$$
$$= \int_0^\infty \left(\frac{1}{x+s_2} - \frac{1}{x+s_1}\right) dx$$
$$= \int_0^\infty \frac{s_1 - s_2}{x^2 + (s_1+s_2)x + s_1 s_2} dx$$
$$= \ln \frac{s_1}{s_2} \qquad (5.7.6)$$

(the last integral is normally convergent, since if, say, $0 > s_2 > s_1$, we have $\frac{s_1 - s_2}{x^2 + (s_1+s_2)x + s_1 s_2} > \frac{s_1 - s_2}{x^2}$, and $\int_1^\infty \frac{s_1 - s_2}{x^2} dx = \frac{s_1 - s_2}{x}\Big|_{x=1}^{x=\infty} = s_1 - s_2$). The result (5.7.6) can also be easily understood if we start by considering the defference $I_N(s_2) - I_N(s_1) = \int_0^\infty \frac{dx}{x+s_2} - \int_0^\infty \frac{dx}{x+s_1}$.

Here is one more striking example dealing with a nonnormally convergent integral dependent on a parameter. Take the quantity

$$I(\lambda) = \int_0^\infty \frac{\sin \lambda x}{\lambda x} dx. \qquad (5.7.7)$$

It can be proved that $I(1) = \int_0^\infty \frac{\sin x}{x} dx = \pi/2$; here we will not discuss the validity of this neat relationship. On the other hand, for $\lambda > 0$ we have

$$I(\lambda) = \int_0^\infty \frac{\sin \lambda x}{x} dx = \int_0^\infty \frac{\sin \lambda x}{\lambda x} d(\lambda x)$$
$$= \int_0^\infty \frac{\sin y}{y} dy = I(1) = \frac{\pi}{2},$$

while for $\lambda < 0$ we have

$$I(\lambda) = \int_0^\infty \frac{\sin \lambda x}{x} dx = \int_0^\infty \frac{\sin(-|\lambda| x)}{x} dx$$
$$= -\int_0^\infty \frac{\sin(|\lambda| x)}{x} dx = -I(|\lambda|)$$
$$= -\frac{\pi}{2}$$

(at $\lambda = 0$, obviously, $I(0) = 0$). Thus, although a small variation in λ changes the integrand but little, the integral (5.7.7) for some reason experiences a jump at $\lambda = 0$:

$$I(\lambda) = \begin{cases} \pi/2 & \text{if } \lambda > 0, \\ 0 & \text{if } \lambda = 0, \\ -\pi/2 & \text{if } \lambda < 0. \end{cases} \qquad (5.7.8)$$

This is due to the fact that, when we integrate over an infinitely large domain, even a small variation of the integrand $f(x)$ may result in an appreciable effect, since $\int f_1(x)\,dx - \int f(x)\,dx = \int [f_1(x) - f(x)]\,dx$ and here we add small differences $f_1(x) - f(x)$ "extended" over an infinitely large domain.

In this case, too, cutting off singularities in the integral (5.7.7) may clarify the situation. Of course, the integral $I_N(\lambda) = \int_0^N \frac{\sin \lambda x}{x}\,dx$ depends on λ in a continuous fashion, since this function, $I_N(\lambda)$, has a derivative: $I'_N(\lambda) = \int_0^N \frac{\partial}{\partial \lambda}\left(\frac{\sin \lambda x}{x}\right) dx = \int_0^N \cos \lambda x\,dx$ (while if we apply formula (5.7.5a) to $I(\lambda)$, we get $I'(\lambda) = \int_0^\infty \cos \lambda x\,dx$, whose right-hand side is not defined).[5.7] But it is clear (in view of the very definition of the improper integral (5.7.7)) that, as $N \to \infty$, the value of $I_N(\lambda)$ tends to the values defined in (5.7.8); for any N

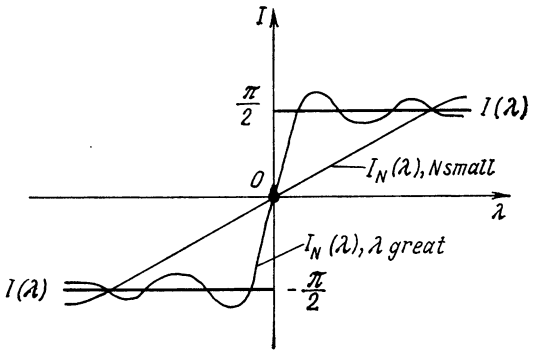


Figure 5.7.1

the value of $I_N(\lambda)$ is positive for $\lambda > 0$, negative for $\lambda < 0$, and zero at $\lambda = 0$. (Why?) If N is very great, the function $I_N(\lambda)$ very rapidly passes from "large" positive values (close to $\pi/2$) to values that are very close to $-\pi/2$, performing in the neighborhood of point $\lambda = 0$ a "rapid" jump (but, of course, remaining continuous at this point). But if N is relatively small, the passage of the function $I_N(\lambda)$ from positive values to negative ones is much smoother (Figure 5.7.1). In the limit, when $N \to \infty$, the transition from "essentially positive" to "essentially negative" values of $I_N(\lambda)$ occurs, so to say, over an infinitely small interval of values of λ, that is, here the function $I(\lambda) = \lim_{N\to\infty} I_N(\lambda)$ experiences a jump (cf. Chapter 16).

[5.7] Compare this with the procedure, described in Section 16.3, of finding the derivative of sgn x, a function that is closely related to $I(\lambda)$ (we can easily see that $I(\lambda) = (\pi/2)\,\mathrm{sgn}\,\lambda$).

Chapter 6 Series. Simple Differential Equations

6.1 A Series Representation of a Function

Suppose that a function $y\ (x)$ is defined exactly by a formula so complex that it is awkward for calculating the values of $y = y\ (x)$. We pose the problem of constructing a simple and convenient approximate expression for the function $y\ (x)$ over a small range of the independent variable x, say, for values of x close to a fixed number a.

The definition of a derivative given in Chapter 2 may be written as follows:

$$y'(a) = \lim_{x \to a} \frac{y\ (x) - y\ (a)}{x - a},$$

from which it follows that

$$y(x) \simeq y(a) + (x - a)\, y'(a), \qquad (6.1.1)$$

where the approximate equality is "exact in the limit," that is, the smaller the difference $|\ x - a\ |$ the greater the accuracy. This formula shows that for values of x close to a the variation of the independent variable by a (small) quantity $x - a$ corresponds to a variation in the value of the function equal to $y'(a)(x - a)$, that is, the formula fits the meaning of the derivative as the ***rate of change of the function*** at point $x = a$, or $(dy/dx)_{x=a} = y'(a)$.

For large values of $|\ x - a\ |$ the accuracy of (6.1.1) becomes poor: the greater the difference $|\ x - a\ |$ the poorer the accuracy. Indeed, if we put $\Delta y = y\ (x) - y(a) = y'(a)(x - a)$, we are assuming that the rate of change of the function everywhere between x and a is the same and equal to the rate $y'\ (a)$ of change of the function at point a, while actually the rate y' also *changes* in the interval between x and a. The exact formula that takes into account this change in $y'(x)$ has the form[6.1]

$$y\ (x) = y\ (a) + \int_a^x y'\ (t)\, dt. \qquad (6.1.2)$$

This formula is known as the *first*, or *linear*,[6.2] *approximation* for function $y\ (x)$.

Applying (6.1.1) to the derivative $y'(x)$, we get

$$y'(x) \simeq y'\ (a) + (x - a)\, y''\ (a). \qquad (6.1.3)$$

Before proceeding, let us recall that $y''(x) = dy'/dx$ is the *second derivative* of function $y\ (x)$ with respect to x, denoted also d^2y/dx^2. It is obtained from y' just as y' is obtained from y. The third derivative $y''' = dy''/dx$ is derived in a similar fashion. The fourth derivative is denoted y^{IV}, the fifth y^{V} or $y^{(5)}$, and so on. The *derivative of order n*, or the *n*th derivative, which is obtained by taking the derivative of the function $y\ (x)$ n times in succession, is denoted by $y^{(n)}$ or d^ny/dx^n (the second, third, and fourth derivatives are denoted by y'', y''', and y^{IV}, but not by $y^{(2)}$, $y^{(3)}$, and $y^{(4)}$. In the notation $y^{(n)}$, the n is enclosed in parentheses to distinguish it from an exponent.

Clearly, if x is measured in units of e_1 and y in units of e_2, then the dimensions of the nth derivative d^ny/dx^n is e_2/e_1^n. For instance, if y is distance (measured in centimeters) and x is time (measured in seconds), the second derivative (the acceleration) y'' has the dimensions of $\mathrm{cm/s^2}$; the dimensions of the increment $\Delta y''$ of the acceleration are also $\mathrm{cm/s^2}$, which means that the rate of change

[6.1] Formula (6.1.2) follows directly from (3.4.5a), since $\int_a^x y'\ (t)\, dt = y\ (t)\Big|_a^x = y\ (x) - y\ (a)$. Formula (6.1.5) is verified in the same manner.

[6.2] In the expression (6.1.1) for y, the quantity x appears only in the first power, in other words, y is a polynomial of the first degree in x. This relationship is termed *linear* because its graph is a straight line (see Section 1.3).

of the acceleration $y''' = \lim (\Delta y''/\Delta x)$ has the dimensions of cm/s^3, and so on.

Now let us return to the problem of the approximate expression of a function. Formula (6.1.3) for the derivative is nothing but formula (6.1.1) applied to y' instead of y. Now substitute the approximate expression (6.1.3) for the derivative y' into the exact formula (6.1.2). The result is

$$y(x) \simeq y(a) + \int_a^x [y'(a) + (t-a)\, y''(a)]\, dt$$

$$= y(a) + (x-a)\, y'(a) + \frac{(x-a)^2}{2} y''(a). \tag{6.1.4}$$

This formula is approximate but is more exact than (6.1.1). In deriving (6.1.4) we took into account that the derivative $y'(x)$ is not constant, but the variation in $y'(x)$ was considered only approximately: formula (6.1.3), which we made use of when deriving (6.1.4), assumes that $y''(x)$ is constant (and equal to $y''(a)$), which is what gives us the *linear* dependence of y' on x. For $y(x)$ the relationship is *quadratic*. Let us again check the dimensions in (6.1.4) by employing the standard distance-time example, namely, if y is measured in centimeters and x in seconds, then the dimensions of y' and y'' are, respectively, cm/s (velocity) and cm/s^2 (acceleration), while the quantities $y(x)$ and $y(a)$ have the dimensions of cm and the factors $(x-a)$ and $(1/2)(x-a)^2$ have, respectively, the dimensions of s and s^2, whereby the dimensions of the left- and right-hand sides of (6.1.4) coincide (cm).

Let us make formula (6.1.4) still more precise by taking into consideration that y'' is not constant. We take advantage of the formula

$$y'(x) = y'(a) + \int_a^x y''(t)\, dt, \tag{6.1.5}$$

which is obtained from (6.1.2) by substituting y' for y. Let us represent $y''(x)$ using a formula of the type (6.1.1) applied not to the initial function $y(x)$ but to the function $y''(x)$:

$$y''(x) \simeq y''(a) + (x-a)\, y'''(a) \tag{6.1.6}$$

and then substitute this expression into (6.1.5). From (6.1.5) and (6.1.6) we obtain

$$y'(x) \simeq y'(a) + \int_a^x [y''(a) + (t-a)\, y'''(a)]\, dt,$$

or

$$y'(x) \simeq y'(a) + (x-a)\, y''(a) + \frac{(x-a)^2}{2} y'''(a). \tag{6.1.7}$$

Note that (6.1.7) is a formula of the type (6.1.4) written for $y'(x)$.

Now let us substitute the expression for $y'(x)$ (6.1.7) into (6.1.2):

$$y(x) \simeq y(a) + \int_a^x \left[y'(a) + y''(a)(t-a) + y'''(a)\frac{(t-a)^2}{2}\right] dt$$

$$= y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \frac{y'''(a)}{6}(x-a)^3. \tag{6.1.8}$$

The general law becomes all the more obvious if we compare the expressions obtained above. In the crudest approximation we assume that for a small difference $|x-a|$,

$$y(x) \simeq y(a) \tag{6.1.9}$$

(this fact is obvious and does not require knowing higher mathematics). We call this equation the ***zeroth approximation***. Then expression (6.1.1) is called the ***first approximation***, expression (6.1.4) the ***second approximation***, and expression (6.1.8) the ***third approximation***. Listed, they are

$y(x) \simeq y(a)$ (zeroth appoximation)

$y(x) \simeq y(a) + (x-a)\, y'(a)$ (first approximation)

$$y(x) \simeq y(a) + (x-a)\,y'(a) + \frac{(x-a)^2}{2}\,y''(a)$$

(second approximation)

$$y(x) \simeq y(a) + (x-a)\,y'(a) + \frac{(x-a)^2}{2}\,y''(a) + \frac{(x-a)^3}{6}\,y'''(a)$$

(third approximation).

It is now easy to imagine the aspect of the approximate formulas for $y(x)$ if the approximation process is continued: if we take into account that y''' is not constant, then $y^{IV}(a)$ will be involved and the expression for $y(x)$ (the fourth approximation) will contain a term with $(x-a)^4$. Each subsequent step in approximating $y(x)$ yields an additional term with a higher power of $x-a$. One can expect that the more powers of $x-a$ that enter into the formula, the more exact the formula is. Of course, what we have said is valid only for small $|x-a|$ while for large $|x-a|$ all the above formulas may prove invalid (see Section 6.3).

The general formula for the nth approximation has the form

$$y(x) \simeq y(a) + c_1 y'(a)(x-a) + c_2 y''(a)(x-a)^2 + c_3 y'''(a)(x-a)^3 + \ldots + c_n y^{(n)}(a)(x-a)^n, \quad (6.1.10)$$

which follows from dimensional considerations. Indeed, if y is measured in units of e_2 and x in units of e_1, then $y^{(k)}(a) = (d^k y/dx^k)_{x=a}$ has the dimensions of e_2/e_1^k, so that in the approximate expression for $y(x)$ the term with $y^{(k)}(a)$ must contain a factor of dimensions of e_1^k, that is, the factor $(x-a)^k$. Of course, these ideas do not enable us to calculate the numerical values of the coefficients $c_1, c_2, c_3, \ldots$ in (6.1.10). (We know that $c_1 = 1$, $c_2 = 1/2$, $c_3 = 1/6$; but what are c_4 and all the other coefficients?) These coefficients can be calculated by the following method.

We denote the right-hand side of (6.1.10) (a polynomial of degree n in variable x) by $P_n(x)$:

$$P_n(x) = y(a) + c_1 y'(a)(x-a) + c_2 y''(a)(x-a)^2 + c_3 y'''(a)(x-a)^3 + \ldots + c_n y^{(n)}(a)(x-a)^n. \quad (6.1.11)$$

From this it readily follows that $P_n(a) = y(a)$. Compute the first derivative of (6.1.11):

$$P_n'(x) = c_1 y'(a) + 2c_2 y''(a)(x-a) + 3c_3 y'''(a)(x-a)^2 + \ldots + nc_n y^{(n)}(a)(x-a)^{n-1}. \quad (6.1.11a)$$

This means that $P_n'(a) = c_1 y'(a)$. Now let us require that not only the *value of the polynomial* (6.1.11) *at point* $x = 0$ *coincide with the value* $y(a)$ *of the function* $y(x)$ *at the same point, but also the rate* $P_n'(a)$ *of the variation at point* $x = a$ *be the same as the rate* $y'(a)$ *of the variation of function* $y(x)$. For this to happen, we must put $c_1 = 1$. Finding the first and higher-order derivatives of (6.1.11a) and replacing x with a in the resulting expressions yields

$$P_n''(a) = 2c_2 y''(a)$$
and $P_n''(a) = y''(a)$ at $c_2 = 1/2$,
$$P_n'''(a) = 6c_3 y'''(a)$$
and $P_n'''(a) = y'''(a)$ at $c_3 = 1/6$,
$$P_n^{IV}(a) = 24c_4 y^{IV}(a)$$
and $P_n^{IV}(a) = y^{IV}(a)$ at $c_4 = 1/24$,
. . . . . .
$$P_n^{(n)}(a) = 2\cdot 3\cdot 4 \ldots n\,y^{(n)}(a)$$
and $P_n^{(n)}(a) = y^{(n)}(a)$
at $c_n = 1/(2\cdot 3\cdot 4 \ldots n)$.

Thus, if we require that the *values of* n *successive derivatives of the polynomial* $P_n(x)$ *at* $x = a$ *coincide with the derivatives of the same order of the function* $y(x)$ *at the same point* (this requirement means a closeness between $P_n(x)$ and $y(x)$), this means that we must put

$c_1 = 1$, $c_2 = 1/2$, $c_3 = 1/(2\cdot 3)$, $c_4 = 1/(2\cdot 3\cdot 4)$, . . ., $c_n = 1/(2\cdot 3 \ldots n)$. Thus,

$$P_n(x) = y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \ldots + \frac{y^{(n)}(a)}{2\cdot 3\cdot 4 \ldots n}(x-a)^n. \quad (6.1.12)$$

We have arrived at the approximation formula

$$y(x) \simeq y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \frac{y'''(a)}{6}(x-a)^3 + \ldots + \frac{y^n(a)}{2\cdot 3 \ldots n}(x-a)^n. \quad (6.1.13)$$

We have a convenient designation for the product of a succession of natural numbers $1\cdot 2\cdot 3\ldots n$. It is $n!$ and is read *factorial n*. For example, $3! = 1\cdot 2\cdot 3 = 6$, $4! = 1\cdot 2\cdot 3\cdot 4 = 24$, $5! = 1\cdot 2\cdot 3\cdot 4\cdot 5 = 120$, etc., while $1! = 1$ and $2! = 2$. Using the factorial notation, we can rewrite (6.1.13) as

$$y(x) \simeq y(a) + \frac{y'(a)}{1!}(x-a) + \frac{y''(a)}{2!}(x-a)^2 + \ldots + \frac{y^n(a)}{n!}(x-a)^n. \quad (6.1.13a)$$

Here is a fuller proof of formulas (6.1.13) and (6.1.13a), which also leads to results that are stronger than the ones obtained above. Let us return to the exact equation (6.1.2) and substitute $d(t-x)$ for dt under the integral sign:

$$y(x) = y(a) + \int_a^x y'(t)\,dt = y(a) + \int_a^x y'(t)\,d(t-x).$$

This substitution is justified since the upper limit of integration x is regarded as a constant and therefore $d(t-x) = dt$.

Let us now integrate by parts (see formula (5.4.3a)) to get

$$y(x) = y(a) + \int_a^x y'(t)\,d(t-x) = y(a) + y'(t)(t-x)\Big|_a^x - \int_a^x (t-x)\,dy'(t) = y(a) + y'(a)(x-a) + \int_a^x (x-t)\,y''(t)\,dt. \quad (6.1.14)$$

If we again substitute $d(t-x)$ for dt and integrate by parts, we get

$$\int_a^x (x-t)\,y''(t)\,d(t-x) = -\int_a^x y''(t)\,d\frac{(x-t)^2}{2} = -y''(t)\frac{(x-t)^2}{2}\Big|_a^x + \int_a^x \frac{(x-t)^2}{2}\,dy''(t) = y''(a)\frac{(x-a)^2}{2} + \frac{1}{2}\int_a^x (x-t)^2\,y'''(t)\,dt,$$

whence

$$y(x) = y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \frac{1}{2}\int_a^x (x-t)^2\,y'''(t)\,dt \quad (6.1.14a)$$

Performing the integration by parts n times, we get an exact expression for $y(x)$ consisting of $n+2$ terms:

$$y(x) = y(a) + y'(a)(x-a) + \frac{y''(a)}{2!}(x-a)^2 + \frac{y'''(a)}{3!}(x-a)^3 + \ldots + \frac{y^n(a)}{n!}(x-a)^n + \frac{1}{n!}\int_a^x (x-t)^n y^{(n+1)}(t)\,dt. \quad (6.1.15)$$

This formula, in contrast to (6.1.13a), is exact, since it was obtained as a result of transformation of the exact formula (6.1.2). The last term, $r_n = \frac{1}{n!}\int_a^x (x-t)^n y^{(n+1)}(t)\,dt$, on the right-hand side of (6.1.15) is known as the ***remainder***; it is the difference between the $y(x)$ in (6.1.13) and the approximate expression for $y(x)$ expressed by the right-hand side of the same formula.

Formula (6.1.15) can be rewritten in a more convenient form that does not contain an integral:[6.3]

$$y(x)=y(a)+y'(a)(x-a)+\frac{y''(a)}{2!}(x-a)^2 + \ldots + \frac{y^{(n)}(a)}{n!}(x-a)^n+\frac{(x-a)^{n+1}}{(n+1)!}y^{(n+1)}(c), \tag{6.1.16}$$

where c is a number lying between x and a.

In the general case of an arbitrary function $y(x)$, no finite number of powers of $x-a$ with constant coefficients can yield an absolutely exact formula for the initial function.[6.4] In other words, the right-hand side of (6.1.13) or (6.1.13a) only *approximately* coincides with $y(x)$, but is not equal to $y(x)$ exactly. An exact formula can be obtained only by adding an infinite number of powers of $x-a$:

$$y(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \ldots + c_n(x-a)^n + \ldots. \tag{6.1.17}$$

The expression on the right-hand side of (6.1.17) is called an ***infinite series.*** Ordinarily, we drop the word "infinite" and simply say "series". Of course, we can also say that (6.1.17) is true "in the limit", but this means something different than it did above. The crux of the problem is that we cannot add directly an infinite number of terms; thus, in writing out the sum on the right, we are forced to restrict ourselves to a finite number n of the first terms in the series. The statement that for sufficiently small $|x-a|$ formula (6.1.17) is "exact" means that the sum on the right can be made *as close* to $y(x)$ *as desired* (what is required is a sufficiently large number n of terms). We can also say that $y(x)$ is the limit of the sum of the first n terms of the series as $n \to \infty$.[6.5].

The coefficients $c_0, c_1, \ldots, c_n, \ldots$ in (6.1.17) are different for different functions; they also depend on point a. As we have seen earlier, $c_0 = y(a)$, $c_1 = y'(a)$, $c_2 = y''(a)/2!$, $c_3 = y'''(a)/3!$, etc., that is,

$$y(x)=y(a)+y'(a)(x-a)+\frac{y''(a)}{2!}(x-a)^2 + \frac{y'''(a)}{3!}(x-a)^3+\ldots + \frac{y^n(a)}{n!}(x-a)^n+\ldots. \tag{6.1.18}$$

A particular case of (6.1.18) where $a=0$ is

$$y(x)=y(0)+y'(0)x+\frac{y''(0)}{2}x^2 + \frac{y'''(0)}{6}x^3+\ldots+\frac{y^n(0)}{n!}x^n+\ldots. \tag{6.1.19}$$

Using the summation sign, we can write formulas (6.1.18) and (6.1.19) in compact form as follows:

$$y(x)=y(a)+\sum_{n=1}^{n=\infty}\frac{y^n(a)}{n!}(x-a)^n, \tag{6.1.18a}$$

$$y(x)=y(0)+\sum_{n=1}^{n=\infty}\frac{y^n(0)}{n!}x^n. \tag{6.1.19a}$$

These two formulas (as well as (6.1.18) and (6.1.19)) yield the expansion of a function $y(x)$ in a series of integral powers of $x-a$ and x, respectively. The series on the right-hand side of

[6.3] As for the validity of the transition from (6.1.15) to (6.1.16), see the text in small print at the end of Section 7.8, in particular, formula (7.8.13), in which the integration variable x must be replaced with $x-t=z$ and the function $f(x)$ with $y^{(n+1)}(t)$ (which depends on the new variable $x-t=z$; we remind the reader that $d(x-t)=-dt$, since x is regarded as constant). We then get

$$\int_a^x (x-t)^n y^{(n+1)}(t)\,dt = \frac{(x-a)^{n+1}}{n+1}\,y^{(n+1)}(c),$$

where c is a value of the independent variable lying between x and a.

[6.4] Except for the case when $y(x)$ is a polynomial (see the beginning of Section 6.2).

[6.5] It may so happen that for different values of x the accuracy with which the sum of the first n terms approximates $y(x)$ depends on n. (We will not dwell on the general questions of the applicability of formulas of type (6.1.17) and of the functions for which the expansions (6.1.18) and (6.1.19) prove to be invalid.)

(6.1.18) or (6.1.18a) is called a ***Taylor's series***, while the particular case of a Taylor's series, present on the right-hand side of (6.1.19) or (6.1.19a), is known as ***Maclaurin's series***.

Examples of applying Taylor's and Maclaurin's series to concrete functions are discussed in Section 6.2, but a simple example of representing functions in the form of series will be considered here.

Suppose that $y = e^x$. Then

$$y' = e^x,\ y'' = e^x,\ \ldots,\ y^{(n)} = e^x,\ \ldots.$$

We take advantage of the formula (6.1.19) for a Maclaurin's series. In the case at hand,

$$y(0) = y'(0) = y''(0) = \ldots = 1.$$

Substituting into (6.1.19), we arrive at the expansion of the function $y = e^x$ in a Maclaurin's series:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots + \frac{x^n}{n!} + \ldots. \tag{6.1.20}$$

Let us now examine the formula obtained from the Taylor expansion (6.1.18) if we confine ourselves to, say, three terms in the series:

$$y(x) \simeq y(a) + (x-a)\,y'(a) + \frac{(x-a)^2}{2!}\,y''(a).$$

Removing brackets on the right-hand side and arranging the result in increasing powers of x, we get

$$y(x) \simeq \left[y(a) - ay'(a) + \frac{a^2}{2}\,y''(a)\right] + [y'(a) - ay''(a)]\,x + \frac{1}{2}\,y''(a)\,x^2. \tag{6.1.21}$$

On the right is a polynomial of degree two. Note that this polynomial does not coincide with what we would have if we took three terms in Maclaurin's series:

$$y(x) \simeq y(0) + y'(0)\,x + \frac{y''(0)}{2}\,x^2. \tag{6.1.21a}$$

This will become clear if we recall that formula (6.1.21) yields a good result if x is close to a, while formula (6.1.21a) is good when x is close to zero.

In Chapter 2 we gave a definition of a derivative as the limit of the ratio of the increment of the function to the increment of the independent variable. Now that we have expressed the function as a series, we state generally the law according to which the ratio $\Delta y/\Delta x$ approaches dy/dx as Δx tends to zero.

Let us take Taylor's series (6.1.18) and denote $x - a$ by Δx. We then shift $y(a)$ to the left-hand side and divide both sides by Δx, denoting $y(x) - y(a)$ by Δy. The result is

$$\frac{\Delta y}{\Delta x} = y'(a) + \frac{1}{2}\,y''(a)\,\Delta x + \frac{1}{6}\,y'''(a)\,(\Delta x)^2 + \ldots. \tag{6.1.22}$$

For small Δx, the second term on the right-hand side, containing Δx, is greater than the third. Discarding the latter, we conclude that the difference of the ratio $\Delta y/\Delta x$ from the value of the derivative $y'(a)$ at the endpoint a of the interval extending from $x = a$ to $x = a + \Delta x$ is proportional to the second derivative $y''(a)$ and the size of the interval, Δx:

$$\frac{\Delta y}{\Delta x} \simeq y'(a) + \frac{1}{2}\,y''(a)\,\Delta x.$$

Note that here we compare the ratio of the increment of the function on the range of variation of x considered in the problem to the increment of x with the value of the derivative $y'(a)$ at the *endpoint* a of this interval.

The derivative may be evaluated differently by assuming that $[f(x + \Delta x/2) - f(x - \Delta x/2)]/2$ is its approximate value rather than $[f(x + \Delta x) - f(x)]/2$. In other words, to calculate the derivative at point a, we take the increment of the function as x varies from $a - \Delta x/2$ to $a + \Delta x/2$ and divide it by Δx. Let us now compare this new ratio with the derivative $y'(a)$ evaluated at the *midpoint* of the interval. We get

$$\Delta y = f\left(a + \frac{\Delta x}{2}\right) - f\left(a - \frac{\Delta x}{2}\right),$$

$$f\left(a + \frac{\Delta x}{2}\right) \simeq f(a) + \frac{\Delta x}{2}\,f'(a) + \frac{1}{2}\left(\frac{\Delta x}{2}\right)^2 f''(a) + \frac{1}{6}\left(\frac{\Delta x}{2}\right)^3 f'''(a),$$

$$f\left(a - \frac{\Delta x}{2}\right) \simeq f(a) - \frac{\Delta x}{2}\,f'(a) + \frac{1}{2}\left(\frac{\Delta x}{2}\right)^2 f''(a) - \frac{1}{6}\left(\frac{\Delta x}{2}\right)^3 f'''(a),$$

$$\frac{\Delta y}{\Delta x} \simeq f'(a) + \frac{2}{6 \cdot 2}\left(\frac{\Delta x}{2}\right)^2 f'''(a) = f'(a) + \frac{(\Delta x)^2}{24}\,f'''(a). \tag{6.1.23}$$

We see that the new method of estimating $f'(a)$ is much more exact: the difference between the ratio of the increments and the derivative is proportional to $(\Delta x)^2$ and not to Δx and, what is more, contains the small coefficient 1/24. Thus, approximately,

$$f'(a) \simeq \frac{f(a+\Delta x)-f(a)}{\Delta x}, \qquad (6.1.24)$$

while the more exact formula follows from (6.1.22),

$$f'(a) \simeq \frac{f(a+\Delta x)-f(a)}{\Delta x} - \frac{1}{2} f''(a)\,\Delta x. \qquad (6.1.24a)$$

On the other hand,

$$f'(a) \simeq \frac{f(a+\Delta x/2)-f(a-\Delta x/2)}{\Delta x}, \qquad (6.1.25)$$

while the more exact formula follows from (6.1.23),

$$f'(a) \simeq \frac{f(a+\Delta x/2)-f(a-\Delta x/2)}{\Delta x} - \frac{1}{24} f'''(a)(\Delta x)^2. \qquad (6.1.25a)$$

All these equations, (6.1.24), (6.1.24a), (6.1.25), and (6.1.25a), are of course approximate, since we have ignored higher-order quantities in comparison to those that enter into these formulas (with respect to Δx).

In a similar manner we can obtain approximate formulas for calculating the second derivative. By definition,

$$f'' = \frac{df'}{dx} \simeq \frac{\Delta f'}{\Delta x}.$$

Replacing $\Delta f'$ with $f'(a+\Delta x) - f'(a)$ and estimating $f'(a+\Delta x)$ and $f'(a)$ via the ratios $[f(a+2\Delta x) - f(a+\Delta x)]/\Delta x$ and $[f(a+\Delta x) - f(a)]/\Delta x$, we arrive at an approximate expression for the second derivative of the function that incorporates $[f(a+2\Delta x) - 2f(a+\Delta x) + f(a)]/(\Delta x)^2$. To estimate this ratio, we again turn to Taylor's series:

$$f(a+2\Delta x) \simeq f(a) + \frac{1}{2} f'(a)(2\Delta x) + \frac{1}{6} f''(a)(2\Delta x)^2 + \frac{1}{24} f'''(a)(2\Delta x)^3,$$

$$f(a+\Delta x) \simeq f(a) + \frac{1}{2} f'(a)\Delta x + \frac{1}{6} f''(a)(\Delta x)^2 + \frac{1}{24} f'''(a)(\Delta x)^3.$$

Using the first three terms on the right-hand sides of these formulas, we obtain

$$\frac{f(a+2\Delta x)-2f(a+\Delta x)+f(a)}{(\Delta x)^2} \simeq \frac{1}{3} f''(a),$$

while the more exact formula (the one that allows for all four terms) is

$$f''(a) \simeq 3\frac{f(a+2\Delta x)-2f(a+\Delta x)+f(a)}{(\Delta x)^2} - \frac{3}{4} f'''(a)\,\Delta x. \qquad (6.1.26)$$

To obtain $f''(a)$, it is best to combine not the values $f(a)$, $f(a+\Delta x)$, and $f(a+2\Delta x)$ of the function $f(x)$ calculated on the interval between a and $a+\Delta x$ but the values $f(a-\Delta x)$, $f(a)$, and $f(a+\Delta x)$. (Here the value of the independent variable $x = a$ coincides with the *midpoint* of the range of variation of x.) Indeed, it can be easily shown that

$$f''(a) \simeq 3\frac{f(a+\Delta x)-2f(a)+f(a-\Delta x)}{(\Delta x)^2},$$

or, more precisely,

$$f''(a) \simeq 3\frac{f(a+\Delta x)-2f(a)+f(a-\Delta x)}{(\Delta x)^2} - \frac{1}{20} f^{IV}(a)(\Delta x)^4 \qquad (6.1.27)$$

(see Exercise 6.1.6).

Geometrically, the differences between the approximate formulas (6.1.24) and (6.1.25) for the derivative lie in the fact that in the first case we replaced the tangent t to the graph of the function $y = f(x)$ (Figure 6.1.1) at point $M(a, f(a))$ (the slope of the tangent line at this point is equal to $f'(a)$) with the straight line $MM_1 \equiv l$, where $M_1 = M_1(a+\Delta x, f(a+\Delta x))$, while in the second case t replaced with the straight line $N_1N_2 = l_1$, where $N_1 = N_1(a - \Delta x/2, f(a - \Delta x/2))$ and $N_2 = N_2(a+\Delta x/2, f(a+\Delta x/2))$. It is clear that the straight line l_1 is much closer in direction to the tangent line t than the straight line l (see Figure 6.1.1). In a similar manner, formula (6.1.26) is obtained on the assumption that the rate of change of derivative y' (the rate with which the tangent to the curve $y = f(x)$ rotates as point M moves along the curve) is estimated through the differences in variation of the derivative from point M to point M_1 and from point M_1 to point $M_2(x+2\Delta x, f(x+2\Delta x))$, while for formula (6.1.27)

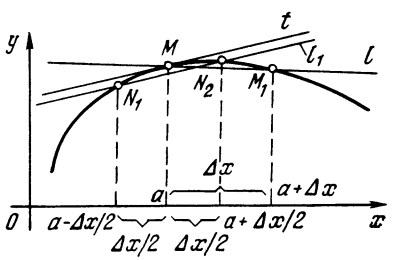


Figure 6.1.1

the rate of change of derivative y' is estimated through the differences in variation of the derivative from point M' $(x - \Delta x,\ f(x - \Delta x))$ to M and from M to M_1 (make a rough sketch).

Exercises

6.1.1. Expand the third-degree polynomial $y = ax^3 + bx^2 + cx + d$ in a series in powers of $x - x_0$, where x_0 is an arbitrary fixed number. Compare the sum of the first two, three, and four terms of the expansion obtained with the polynomial.

6.1.2. Expand the function $y = xe^x$ in a Maclaurin's series. Verify that the expansion can be obtained from the expansion of e^x multiplying the latter by x.

6.1.3. Expand the function e^x in a Taylor's series in powers of $x - 1$.

6.1.4. Determine the accuracy of the approximate formula $(1 + r)^m \simeq e^{mr}$ for $r \ll 1$ and $mr \ll 1$. To do this, write the left- and right-hand members approximately as the sum $a + br + cr^2$, obtained by ignoring the third and higher powers of r.

6.1.5. Given the interval $\Delta x = 1$, 1/2, 1/4, and 1/8 and using formulas (6.1.24) and (6.1.25) (or (6.1.26) and (6.1.27)), respectively, find the approximate values of (a) the derivative of e^x at $x = 0$ and (b) the second derivative of e^x at $x = 0$. Compare the results with the exact values.

6.1.6. Prove formula (6.1.27).

6.2 Computing the Values of Functions by Means of Series

Let us dwell briefly on the principles underlying the formulas of Section 6.1. When we began the study of the derivative, or the rate of change of a function, we assumed the function as known, that is, we proceed from the fact we could compute the value of the function for any value of the independent variable. This is why, when we considered derivatives, we found them directly, empirically so to say, by computing the values of the function for close-lying values of the independent variable. Later on we learned how to find derivatives by formulas and it turned out that setting up formulas for derivatives is a rather simple matter. And so finding the values of a function by means of a formula involving derivatives (all formulas of Section 6.1 were of this type) turns out to be even simpler than a direct computation of the function.

Since only in the case of a polynomial does a Taylor's series terminate, or contain a finite number of terms, it follows that any function different from a polynomial will be represented by an *infinite* series (this statement will be proved somewhat later). The practical value of such infinite series for computational purposes is due to the possibility of confining ourselves to a few (commonly two or three) terms of the series in order to obtain a sufficiently accurate result. This requires, of course, that the discarded terms of the series be small.

Let us consider a few very simple examples of series expansions of functions. We have already mentioned the (simple) example when the function

$$y = b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \quad (6\ 2\ 1)$$

is a polynomial of degree n. Its derivative $y'(x)$ is a polynomial of degree $n - 1$, the derivative $y''(x)$ is a polynomial of degree $n - 2$, and so on; finally, its derivative $y^{(n)}(x)$ is a constant, while $y^{(n+1)}(x)$ and all higher-order derivatives are identically zero. For this reason, for a polynomial the corresponding Taylor's series terminates; in other words, it contains a finite number of terms. We have arrived at the same polynomial, but in the form of a Taylor expansion in powers of $x - a$:

$$y = c_0 + c_1 (x - a) + c_2 (x - a)^2 + \ldots + c_n (x - a)^n. \quad (6.2.1a)$$

For polynomials of nth degree, the sum of the first $n + 1$ terms in the Taylor's series yields the *exact* expression valid for all values of x rather than only for values close to a, while formula (6.2.1) can be regarded as Maclaurin's series (6.1.19) for our function $y(x)$. (Why?)

In the preceding section we considered the exponential function $y = e^x$ and obtained formula (6.1.20),

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots + \frac{x^n}{n!} + \ldots. \quad (6.2.2)$$

TABLE 6.1

x	e^x	$1+x$	$1+x+\frac{x^2}{2}$	$1+x+\frac{x^2}{2}+\frac{x^3}{6}$	$1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}$
0.10	1.1052	1.10	1.1050	1.1052	1.1052
0.25	1.2840	1.25	1.2812	1.2838	1.2840
0.50	1.6487	1.50	1.6250	1.6458	1.6484
0.75	2.1170	1.75	2.0312	2.1015	2.1147
1.00	2.7183	2.00	2.5000	2.6667	2.7083
1.25	3.4903	2.25	3.0312	3.3568	3.4585
1.50	4.4817	2.50	3.6250	4.1876	4.3986
2.00	7.3891	3.00	5.0000	6.3333	7.0000

In particular, substituting $x = 1$ and $x = -1$, we get expressions for numbers e and e^{-1} in the form of series:

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \ldots + \frac{1}{n!} + \ldots, \tag{6.2.3}$$

$$\frac{1}{e} = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} + \ldots + \frac{1}{(2n)!} - \frac{1}{(2n+1)!} + \ldots. \tag{6.2.3a}$$

Formula (6.2.2) enables us to compute e^x rapidly and to a high degree of accuracy, as can be seen from Table 6.1

If at $x = 0.1$ we confine ourselves only to the first two terms, the error will not exceed 0.5% of the true value; at $x = 0.5$ the first three terms yield an error of 1.4%; finally, at $x = 1.0$ the first four terms yield an error of 1.8% of the function $y = e^x$.

Such a high accuracy is plainly due to the fact that the terms of the series (6.2.2) *fall off* rapidly. Each subsequent term in the series is less than the preceding one primarily because the denominator of the $(n + 1)$st term is n times the denominator of the preceding nth term. If $x < 1$, then in addition we have that x^n is the smaller the greater n is, and x^n rapidly decreases as n grows. But even if $x > 1$, the increase in the denominator in the distant terms of the series as n grows will inevitably overcome the increase in the numerator, since as we move from the nth term to the $(n + 1)$st term the numerator increases x-fold while the denominator increases n-fold, and all we have to do is to wait until the increase in the denominator overcomes the increase in the numerator (that is, until n becomes greater then x). As can be seen from Table 6.1, when $x = 2$, the sum of the first five terms in the series yields an error of 5%. But if we add the sixth term, $x^5/120$, then we get 7.3500, which yields an error of only 0.5%.

Let us construct formulas of the same type for trigonometric functions. For instance, for $y(x) = \sin x$ we have

$$y'(x) = \cos x, \quad y''(x) = -\sin x,$$
$$y'''(x) = -\cos x, \quad y^{IV}(x) = \sin x$$

The law for subsequent derivatives is obvious. Substituting $x = 0$, we get

$$y(0) = 0, \quad y'(0) = 1,$$
$$y''(0) = 0, \quad y'''(0) = -1, \ldots.$$

Consequently,

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \ldots. \tag{6.2.4}$$

In similar fashion we get the formula

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \ldots. \tag{6.2.5}$$

Figure 6.2.1 shows the graphs of the sine, cosine, and also the graphs of the polynomials if we take one, two, and three terms in the corresponding series. Accuracy improves visibly when we take more and more terms of the series.

Tables 6.2 and 6.3 list the values of the sine and cosine functions, respectively. In the tables both φ and x are

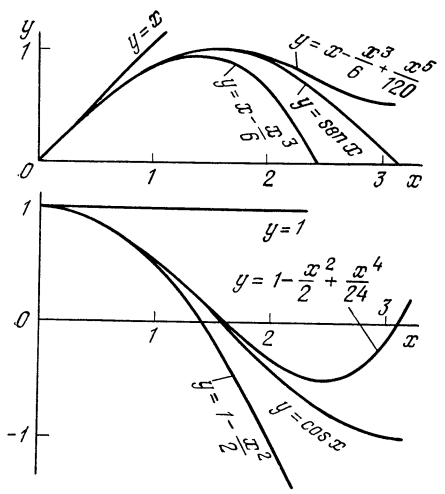


Figure 6.2.1

angles, but φ is expressed in degrees, while x is the corresponding angle expressed in radians. It is evident from the tables that two or three terms of the series suffice to obtain excellent accuracy in the interval from 0 to $\pi/4$. Thus, a power series offers a very convenient practical method for computing the values of trigonometric functions. Note that in absolute value the nonzero terms of the series for the sine and cosine are exactly equal to the corresponding terms of the series for the function e^x. For this reason, everything we have said pertaining to the falling off of terms with high powers of x in formula (6.2.2) for e^x refers also to the series (6.2.4) and (6.2.5), and these series, just like (6.2.2), enable calculating the value of the corresponding function for an arbitrary x.

It must be emphasized, however, that the fact that the terms in a series decrease, as $n \to \infty$, for an arbitrary x and, the more so, the sequence of "partial" (finite) sums tends to a definite

TABLE 6.2

x	$\varphi°$	$\sin x$	x	$x-\frac{x^3}{6}$	$x-\frac{x^3}{6}+\frac{x^5}{120}$
0	0	0.0000	0.0000	0.0000	0.0000
$\pi/20$	9	0.1564	0.1571	0.1564	0.1564
$\pi/10$	18	0.3090	0.3142	0.3090	0.3090
$3\pi/20$	27	0.4540	0.4712	0.4538	0.4540
$4\pi/20$	36	0.5878	0.6283	0.5869	0.5878
$5\pi/20$	45	0.7071	0.7854	0.7046	0.7071
$6\pi/20$	54	0.8090	0.9425	0.8029	0.8091
$7\pi/20$	63	0.8910	1.0996	0.8780	0.8914
$8\pi/20$	72	0.9510	1.2566	0.9258	0.9519
$9\pi/20$	81	0.9877	1.4137	0.9427	0.9898
$\pi/2$	90	1.0000	1.5708	0.9248	1.0045

TABLE 6.3

x	$\varphi°$	$\cos x$	$1-\frac{x^2}{2}$	$1-\frac{x^2}{2}+\frac{x^4}{24}$	$1-\frac{x^2}{2}+\frac{x^4}{24}-\frac{x^6}{720}$
0	0	1.0000	1.0000	1.0000	1.0000
$\pi/20$	9	0.9877	0.9877	0.9877	0.9877
$\pi/10$	18	0.9510	0.9506	0.9510	0.9510
$3\pi/20$	27	0.8910	0.8890	0.8911	0.8910
$4\pi/20$	36	0.8090	0.8026	0.8091	0.8090
$5\pi/20$	45	0.7071	0.6916	0.7075	0.7071
$6\pi/20$	54	0.5878	0.5558	0.5887	0.5877
$7\pi/20$	63	0.4540	0.3954	0.4563	0.4539
$8\pi/20$	72	0.3090	0.2105	0.3144	0.3089
$9\pi/20$	81	0.1564	0.0007	0.1672	0.1561
$\pi/2$	90	0.0000	−2.337	0.0200	−0.0009

number as $n \to \infty$, can in no way serve as a general rule: in certain respects the series (6.2.1), (6.2.1a), (6.2.2), (6.2.4), and (6.2.5) and more "safe" than the general series (6.1.18) and (6.1.19). Section 6.3 is devoted to this question.

Taylor's and Maclaurin's series find wide application in higher mathematics; some of these applications are discussed below (e.g. see Section 6.6). Here we wish to discuss the problem of employing series in the *integration* of functions. If the function $y = f(x)$ can be expanded in a power series,

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots, \tag{6.2.6}$$

then, generally speaking (with certain reservations discussed in Section 6.3; here we will not touch on these restrictions),

$$\begin{aligned}\int f(x)\,dx &= \int (a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots)\,dx \\ &= C + a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 \\ &+ \ldots + \frac{a_n}{n+1} x^{n+1} + \ldots.\end{aligned} \tag{6.2.7}$$

This result is valid regardless of whether the indefinite integral (antiderivative) $\int f(x)\,dx$ can be expressed by an algebraic formula or not. For instance, above we stated that the function $\int e^{-t^2}\,dt$ cannot be expressed by a (finite) formula, but an infinite series ("formula") can be employed to express it. Replacing x with $-t^2$ in (6.2.2), we get

$$\begin{aligned}e^{-t^2} &= 1 - t^2 + \frac{t^4}{2} - \frac{t^6}{6} \\ &+ \ldots + (-1)^n \frac{t^{2n}}{n!} + \ldots,\end{aligned} \tag{6.2.8}$$

whence

$$\begin{aligned}\int e^{-t^2}\,dt &= C + t - \frac{t^3}{3} + \frac{t^5}{10} - \frac{t^7}{42} \\ &+ \ldots + (-1)^n \frac{t^{2n+1}}{n!\,(2n+1)} + \ldots.\end{aligned} \tag{6.2.9}$$

In particular,[6.6]

$$\begin{aligned}\int_0^x e^{-t^2}\,dt &= x - \frac{x^3}{3} + \frac{x^5}{2 \cdot 5} - \frac{x^7}{6 \cdot 7} \\ &+ \ldots + (-1)^n \frac{x^{2n+1}}{n!\,(2n+1)} + \ldots.\end{aligned} \tag{6.2.10}$$

Exercises

6.2.1. Express the coefficients c_0, c_1, c_2, ... of the series (polynomial) (6.2.1a) in terms of the coefficients b_0, b_1, b_2, ... of polynomial (6.2.1).

6.2.2. It is known that the sine-integral function $\operatorname{si} x = \int \frac{\sin x}{x}\,dx$ cannot be expressed by a formula explicitly. Represent si x as an infinite series.

6.3 Cases Where Series Expansions Cannot be Applied. The Geometric Progression

In the preceding section we set up formulas in the form of series in powers of x with constant coefficients for four functions: the polynomial, e^x, $\sin x$, and $\cos x$. In all these cases it was found that for arbitrary x, each subsequent term in a series, with the possible exception of the first few terms, is less than the preceding one, and the greater the number-label of the term, the closer the term is to zero (while in the first, "safest," case all terms starting from a certain term vanish). In these examples, we can compute the value of the function for *any* x by means of a series if a large enough number of terms of the series is taken so that the discarded terms have practically no effect on the result.

To summarize, then, we began with the problem of approximating a function in a small range of the variable and constructed more and more exact formulas by taking into account the first,

[6.6] Formula (6.2.10) is applicable for any *finite* x; however, it is clear that the important result $\int_0^\infty e^{-t^2}\,dt = \sqrt{\pi}/2$ cannot be derived from (6.2.10)—quite different methods are required to do this.

second, third, and higher derivatives. The accuracy of each formula,

$$y(x) \simeq y(a), \tag{0}$$

$$y(x) \simeq y(a) + (x-a)\,y'(a), \tag{I}$$

$$y(x) \simeq y(a) + (x-a)\,y'(a) + \frac{(x-a)^2}{2}\,y''(a), \tag{II}$$

etc., is the greater, the smaller the quantity $|x-a|$. Moreover, for a given $x-a$ formula (I) is more accurate than formula (0), formula (II) is more accurate than formula (I), and so on. Hence, if we increase the number of terms of a series, this permits increasing the quantity $|x-a|$ while preserving a given accuracy.

The question now arises as to whether it is always possible to attain a given accuracy for *any* value of $|x-a|$ merely by increasing the number of terms in a series. We will use an important example to illustrate this point and will show that this is not so. A power series constructed so as to yield a good approximation in a small range of x, that is, for small $|x-a|$, can be useless in the region of large $|x-a|$. In other words, this series may have a natural limit of applicability, the limit of admissible increase in $|x-a|$ (the limit not depending on the number of terms taken). In the examples of the preceding section this was not evident because the series were specially chosen, but such good behavior is an exception rather than a general rule.

Consider the following simple function:

$$y = \frac{1}{1-x} = (1-x)^{-1}.$$

Taking the derivatives in succession yields

$$y' = \frac{1}{(1-x)^2}, \quad y'' = \frac{1\cdot 2}{(1-x)^3}, \ldots,$$

$$y^{(n)} = \frac{n!}{(1-x)^{n+1}}, \ldots.$$

Substituting $x = 0$, we get

$$y(0) = 1,\ y'(0) = 1,\ y''(0) = 2, \ldots,$$
$$y^{(n)}(0) = n! \ldots.$$

We thus arrive at the series

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots. \tag{6.3.1}$$

The example of the function $(1-x)^{-1}$ is remarkable not only due to the unusually simple form of the resulting power series (all the coefficients are equal to 1). Here it is easy to give an exact formula for the sum of the first n terms of the series (6.3.1):

$$1 + x + x^2 + \ldots + x^{n-1} = \frac{1-x^n}{1-x}. \tag{6.3.2}$$

The validity of this formula, the formula for the sum of the n terms of a *geometric progression*, is known to any secondary-school student. (Just multiply both sides of (6.3.2) into $1-x$.) We can rewrite formula (6.3.2) as follows:

$$1 + x + x^2 + \ldots + x^{n-1} = \frac{1}{1-x} - \frac{x^n}{1-x}. \tag{6.3.3}$$

Comparing (6.3.1) and (6.3.3), we see that $x^n/(1-x)$ is the remainder in the approximate equation

$$\frac{1}{1-x} \simeq 1 + x + x^2 + \ldots + x^{n-1}, \tag{6.3.4}$$

that is, the quantity we neglect when confining ourselves to the first n terms in the series

$$1 + x + x^2 + x^3 + \ldots + x^n + \ldots, \tag{6.3.1a}$$

which is the right-hand side of (6.3.1).

If x lies between -1 and 1, then the greater the n, the closer x^n is to zero and, consequently, if we take a sufficiently large number of terms, we discard only a small quantity (remainder). Series that possess this property are known as ***convergent***. The convergence of the series (6.3.1a) at $x = 1/2$ is illustrated by the table below; here $1/(1-x) = 2$. The quantity Δ in the table is the error (in %) introduced by the approximate formula (6.3.4), that is, the ratio of the difference between the left-

and right-hand sides of (6.3.4) to the left-hand side.

n	1	2	3
$1+x+x^2+\ldots+x^{n-1}$	1	1.5	1.75
Δ	50%	25%	12.5%
n	4	5	6
$1+x+x^2+\ldots+x^{n-1}$	1.875	1.9325	1.9637
Δ	6.25%	3.12%	1.56%

Note that the closer x is to 1, the more terms we have to take in order to obtain a given accuracy.

The whole picture changes if we take $|x| \geqslant 1$. For instance, if $x > 1$, each subsequent term in the series (6.3.1a) is greater than the preceding one. Formula (6.3.3) remains valid, but for $x > 1$ the quantity x^n increases without bound together with n and for this reason we cannot disregard the fraction $x^n/(1-x)$. Here, (6.3.1) does not hold true. There is not even any qualitative similarity between the sum of the positive terms of (6.3.1a) and the negative (since x is greater than 1) quantity $1/(1-x)$. For instance, at $x = 2$ formula (6.3.1) becomes absurd:

$$-1 = 1 + 2 + 4 + 8 + 16 + \ldots$$

For $x > 1$, the sum of the series (6.3.1a) increases without bound with n (this is clearly seen from (6.3.3)). And if $x \leqslant -1$, in the right-hand side of (6.3.1) we have an alternating sum, which is not represented in any way by the quantity $1/(1-x)$. For instance, at $x = -1$ formula (6.3.1) yields

$$1/2 = 1 - 1 + 1 - 1 + 1 - \ldots$$

Series similar to those into which (6.3.1a) transforms when $|x| \geqslant 1$ are called ***divergent.***

We have established that the sum of the terms of an infinite geometric progression (6.3.1a) is equal to $1/(1-x)$ if $|x| < 1$, but if $|x| \geqslant 1$, the infinite geometric progression has no finite sum (it diverges), in other words, the sum of the first n terms of such a series does not tend, as $n \to \infty$, to any finite limit.

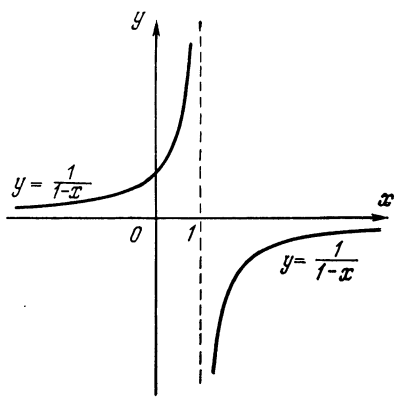


Figure 6.3.1

Also note that any periodic (decimal) fraction is a sum of terms of a geometric progression. For instance,

$$\begin{aligned} 1.(1) &= 1.1111\ldots \\ &= 1 + 0.1 + 0.01 + 0.001 + \ldots \\ &= 1 + 0.1 + (0.1)^2 + (0.1)^3 + \ldots \\ &= \frac{1}{1-0.1} = \frac{1}{0.9} = \frac{10}{9} = 1\frac{1}{9}. \end{aligned}$$

Thus, we have already encountered an elementary series (the geometric progression) in arithmetic and algebra.

The function $y = 1/(1-x)$ (Figure 6.3.1) has a discontinuity at $x = 1$; if x is close to 1 but greater than 1, then $1/(1-x)$ is a large (in absolute value) negative number, while if x is close to 1 but smaller than 1, then $1/(1-x)$ is a large positive number. Thus, when x passes through the value $x = 1$, the value of the function $1/(1-x)$ jumps from large positive values to large (in absolute value) negative numbers. A series cannot describe such behavior, and so it is of no use here.

We note yet another circumstance. When $x = 1$, the function $y = 1/(1-x)$ becomes infinite (the closer x is to 1, the greater y is in absolute value), and at $x = 1$ the terms in (6.3.1a) cease to decrease as $n \to \infty$. A series is suitable for computational purposes only if its terms tend to zero rapidly, that

is, decrease in absolute value.[6.7] At $x = 1$, the formula (6.3.1) is incorrect, since the terms of the series on the right-hand side do not decrease. This means that the series (6.3.1a) is unsuitable for computing the values of the function $y = 1/(1-x)$ at $x = -1$ and for $x < -1$ (the terms in (6.3.1a) do not diminish either) although the function itself does not have a discontinuity at $x = -1$ and is equal to $1/[1-(-1)] = 1/2$.

No matter how we choose the coefficients of a polynomial, the graph will always be a solid continuous line: a polynomial does not have discontinuities. Therefore, if some function $f(x)$ has a discontinuity at $x = x_0$ ($x_0 = 1$ in the example involving $1/(1-x)$), then for the value $x = x_0$ the series constructed for $f(x)$ is definitely unsuitable for computations. Since the greater the absolute value of x, the greater (in absolute value) each term $c_n x^n$ in the series $c_0 + c_1 x + c_2 x^2 + \ldots$ is, it follows that for an arbitrary x that is greater in absolute value than x_0, the series is likewise unsuitable for computation.

Thus, in the case of a discontinuity in a function $f(x)$ we can indicate beforehand an x_0 such that for all x exceeding x_0 in absolute value the series representing $f(x)$ will prove to be unsuitable for computational purposes.

Consider another example. We wish to construct Maclaurin's series for the function $y = \tan x$. Applying the general rule for finding derivatives, we obtain

$$y = \tan x = \frac{\sin x}{\cos x}, \quad y' = \frac{1}{\cos^2 x},$$

$$y''(x) = \frac{2\sin x}{\cos^3 x}, \quad y'''(x) = \frac{2+4\sin^2 x}{\cos^2 x},$$

$$y^{IV}(x) = \frac{16\sin x + 8\sin^3 x}{\cos^5 x},$$

$$y^{V}(x) = \frac{16 + 88\sin^2 x + 16\sin^4 x}{\cos^6 x}, \ldots.$$

This yields $y(0) = 0$, $y'(0) = 1$, $y''(0) = 0$, $y'''(0) = 2$, $y^{IV}(0) = 0$, $y^{V}(0) = 16$, and so on. Whence,

$$\tan x = 0 + 1\cdot x + 0\cdot x^2 + \frac{2}{3\cdot 2\cdot 1} x^3 + 0\cdot x^4 + \frac{16}{5\cdot 4\cdot 3\cdot 2\cdot 1} x^5 + \ldots.$$

Thus,

$$\tan x = x + \frac{1}{3} x^3 + \frac{2}{15} x^5 + \frac{17}{315} x^7 + \frac{62}{2835} x^9 + \ldots. \qquad (6.3.5)$$

(the coefficients of x^7 and x^9 can be obtained in the same manner as the coefficients of x, x^3, and x^5 were obtained).

What can be said of the range of applicability of the series (6.3.5)? The graph of the tangent function (Figure 4.10.5) tells us that this series is suitable for computational purposes only for $|x| < \pi/2$, since at $x = \pi/2$ the function's behavior is as bad as that of function $1/(1-x)$ was for $x = 1$.

But the appearance of the series (6.3.5) does not suggest the value of x for which the series cannot be employed, since the law by which the expansion coefficients are constructed is not simple, in contrast to the law for the coefficients in series (6.3.1a).[6.8]

Note that the presence of a discontinuity of a function is a *sufficient* condition for the series to cease to converge, but it is *not a necessary* condition. By way of an illustration let us consider the function $y = 1/(1+x)$. Applying formula (6.1.19) or simply replacing x with $-x$ in (6.3.1), we get

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \ldots. \qquad (6.3.6)$$

Take $x = 2$, for instance. Then

$$\left.\frac{1}{1+x}\right|_{x=2} = \frac{1}{1+2} = \frac{1}{3}.$$

[6.7] Of course, if one or two or several of the first terms increase, there is no harm done if the subsequent terms of the series fall off rapidly, see the example involving e^x for $x = 2$ (Table 6.1).

[6.8] Note that the ratios of the successive coefficients in the series (6.3.5), $1 \div 1/3 = 3$, $1/3 \div 2/15 = 2.5$, $2/15 \div 17/315 \simeq 2.4706$, $17/315 \div 62/2835 \simeq 2.4677$, rapidly converge to the value $(\pi/2)^2 \simeq 2.4674$. However, there is no way in which we can prove this remarkable fact here.

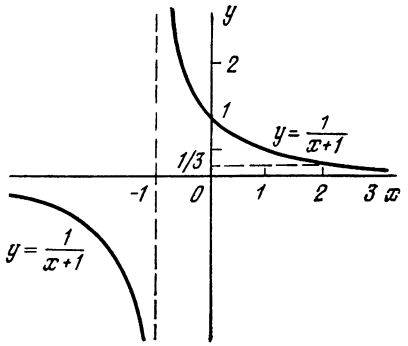


Figure 6.3.2

The sum of the terms of the series

$$1 - x + x^2 - x^3 + \ldots, \qquad (6.3.7)$$

however, oscillates rapidly as the number n of terms increases, and does not resemble 1/3 in any way:

n	1	2	3	4	5	6	7	...
Sum of n terms of series (6.3.7)	1	−1	3	−5	11	−21	43	...

The series is clearly unsuitable for calculating at $x = 2$ the value of the function $y = 1/(1 + x)$. Why does this occur, particularly since the function itself, $y = 1/(1 + x)$, has no discontinuity either for 2 or anywhere between 0 and $x = 2$ (within these limits the function is smooth, well-behaved, see Figure 6.3.2)?

The reason why the series (6.3.7) does not "work" at $x = 2$ is that $y = 1/(1 + x)$ has a discontinuity at $x = -1$. For this reason, at $x = -1$ the absolute values of the terms of the series (6.3.7) do not diminish as n grows. But the absolute values of the terms in (6.3.7) do not depend on the sign of x. Consequently, for $x = 1$ (and all the more so for $x > 1$) this series is not suitable for computation.

Thus, even if we are interested in the behavior of a series only for positive values of x, we still have to take into account all the values of x, including negative values as well, for which the function undergoing expansion has a discontinuity.

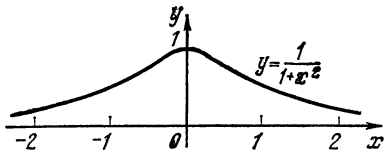


Figure 6.3.3

For the reader acquainted with *complex numbers* (see Chapters 14 and 15) we note that the convergence of a series for real-valued x is affected by the behavior of the corresponding function for *complex* values of the independent variable. Here is an example (see also Section 15.2). Replacing x with x^2 in (6.3.6), we get

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \ldots. \qquad (6.3.8)$$

The graph of the function $y = 1/(1 + x^2)$ (Figure 6.3.3) has no discontinuities for positive and negative x's and does not go to infinity, but the series (6.3.8) is suitable for computations only if $x^2 < 1$, that is, for $-1 < x < 1$. The reason for this is that when $x = \pm\sqrt{-1} = \pm i$, that is, at $x^2 = -1$, the function $y = 1/(1 + x^2)$ becomes infinite and therefore the terms of the series do not decrease in absolute value at $x^2 = -1$. Hence, neither do they decrease in absolute value at $x^2 = 1$ (see p. 481).

Other examples of functions that cannot be expanded in a Taylor's series, say, the function e^{-1/x^2} at $x = 0$ (see p. 198), can also be cited. Fractional powers of x and functions that contain fractional powers of the independent variable, say, the function $\sin\sqrt{x}$, cannot be expanded in a Maclaurin's series.

Remaining within the scope of this book for beginners, we find it rather difficult to go any further into the question of the applicability of Taylor's and Maclaurin's series, to do this, the reader must turn to a textbook on the theory of functions of a complex variable (cf. Chapter 15). However, we must say, to comfort the reader, that practically all functions that a physicist or engineer needs can be expanded in Taylor's series everywhere with the exception of a finite number of (singular) points. As a rule we encounter functions that can be expanded, at least in a properly chosen finite interval, into Taylor's series.

Exercises

6.3.1. Write Maclaurin's series for the functions (a) $y = (x + 1)/(1 - x)$, and (b) $y = \ln(1 + x)$.

6.3.2. Write the Taylor expansion of $y = \ln x$ in powers of $x - 1$.

What are the ranges of applicability of the series obtained in Exercises 6.3.1 and 6.3.2?

6.3.3. Suppose that $f(x)$ and $g(x)$ are known functions. Find the first three terms of the series expansion in powers of x of the function $f(x)\,g(x)$. Construct the same series by multiplying together the series for $f(x)$ and the series for $g(x)$. Compare the results.

6.4 The Binomial Theorem for Integral and Fractional Exponents

Let us form Maclaurin's series expansion of a binomial $a + x$ to an arbitrary power m, that is, $y = (a + x)^m$.

Using the general rule, let us first find the derivatives

$$y' = m(a + x)^{m-1},$$
$$y'' = m(m - 1)(a + x)^{m-2}, \ldots, \tag{6.4.1}$$
$$y^n = m(m - 1)\ldots$$
$$\ldots(m - n + 1)(a + x)^{m-n}, \ldots$$

and the values of the function and derivatives at $x = 0$:

$$y(0) = a^m,\ y'(0) = ma^{m-1},$$
$$y''(0) = m(m - 1)a^{m-2}, \ldots, \tag{6.4.2}$$
$$y^n(0) = m(m - 1)\ldots$$
$$\ldots(m - n + 1)a^{m-n}, \ldots.$$

From this we get Maclaurin's series

$$(a+x)^m = a^m + \frac{m}{1}a^{m-1}x$$
$$+ \frac{m(m-1)}{1\cdot 2}a^{m-2}x$$
$$+ \ldots + \frac{m(m-1)(m-2)\ldots(m-n+1)}{n!}$$
$$\times a^{m-n}x^n + \ldots. \tag{6.4.3}$$

If the exponent m is a positive integer, then $(a + x)^m$ is a polynomial of degree m, so that in this case the series (6.4.3) is finite: the $(m + 1)$st derivative of the function $(a + x)^m$ is zero, and so are all higher derivatives. The formulas (6.4.1) to (6.4.3) reflect this circumstance. Indeed, at $n = m + 1$ the factor $m - n + 1$ vanishes, for $n > m + 1$ there will be, some place in the sequence of the factors m, $m - 1$, . . ., a factor equal to zero and, the product will be equal to zero, too.

For a positive integer m, the product in the numerator can be written in a more convenient form. To this end we multiply and then divide the product $m(m - 1)\ldots(m - n + 1)$ by $(m - n)(m - n - 1)\ldots 3\cdot 2\cdot 1$. The resultis

$$m(m-1)\ldots(m-n+1) = \frac{m(m-1)\ldots 3\cdot 2\cdot 1}{(m-n)(m-n-1)\ldots 3\cdot 2\cdot 1} = \frac{m!}{(m-n)!}.$$

Thus, for positive integral m we finally have

$$(a+x)^m = a^m + \frac{m!}{1!\,(m-1)!}a^{m-1}x$$
$$+ \frac{m!}{2!\,(m-2)!}a^{m-2}x^2$$
$$+ \ldots + \frac{m!}{n!\,(m-n)!}a^{m-n}x^n$$
$$+ \ldots + \frac{m!}{(m-2)!\,2!}a^3x^{m-2}$$
$$+ \frac{m!}{(m-1)!\,1!}ax^{m-1} + x^m. \tag{6.4.4}$$

In formula (6.4.4) we have polynomials of degree m on the right and on the left. Thus, for the case of a positive integer m, we obtain an exact equation that is valid for arbitrary powers of x, that is, formula (6.4.4). This formula is symmetric with respect to x and a: the coefficients of the terms $a^{m-n}x^n$ and a^nx^{m-n} are the same. This is obvious since $(x + a)^m$ does not depend on the order of the summands in the parentheses: $(x + a)^m = (a + x)^m$

Formula (6.4.4) is called the ***binomial theorem*** (sometimes it is called ***Newton's binomial theorem***, see below) or the ***binomial expansion.*** It can be obtained without resorting to derivatives and series. We have to take the product $(a + x)(a + x)(a + x)\ldots(a + x)$

consisting of m cofactors, perform the multiplication, and collect like terms. However, when m is specified in the general form by a symbol and not a number, collecting like terms is rather difficult. On the whole, the derivation of the binomial expansion via methods of higher mathematics, that is, using Maclaurin's series, is simpler.

We note that Newton obtained the general formula (6.4.3), that is, the expansion of $(x + a)^m$ for the case of an arbitrary exponent m. It would therefore be more appropriate to call formula (6.4.3) Newton's binomial theorem instead of (6.4.4), which was known before Newton's time[6.9] and which is a simple particular case of the relationship connecting the coefficients in the representations (6.2.1) and (6.2.1a) of a polymial, see Exercise 6.2.1.

Let us return to the general formula (6.4.3). Suppose that m is not a positive integer. Since the powers n of the variable x are positive integers, this means that not a single factor m, $m-1$, ..., $m-n+1$ in the numerator of the coefficient of x^n will vanish and, hence (6.4.3) yields an *infinite* series. For instance, for $m = -1$ this series is of the form

$$\frac{1}{a+x} = \frac{1}{a} - \frac{x}{a^2} + \frac{x^2}{a^3} - \frac{x^3}{a^4} + \ldots . \tag{6.4.5}$$

Note that at $a = 1$ formula (6.4.5) passes into the familiar formula (6.3.6) for the sum of the terms of an (infinite) geometric progression.

From (6.4.5) we also find that

$$\frac{1}{a-x} = \frac{1}{a} + \frac{x}{a^2} + \frac{x^2}{a^3} + \frac{x^3}{a^4} + \ldots .$$

For $m = 1/2$ we have

$$\sqrt{a+x} = \sqrt{a} + \frac{1}{2}\frac{x}{\sqrt{a}} - \frac{1}{8}\frac{x^2}{a\sqrt{a}} + \frac{1}{16}\frac{x^3}{a^2\sqrt{a}} - \frac{5}{128}\frac{x^4}{a^3\sqrt{a}} + \frac{7}{256}\frac{x^5}{a^4\sqrt{a}} - \frac{21}{1024}\frac{x^6}{a^5\sqrt{a}} + \ldots . \tag{6.4.6}$$

[6.9] In Europe formula (6.4.4), where m is a positive integer, was first discovered by Niccolo ***Tartaglia*** (c. 1500-1557), but even before that the formula was known to Arabian mathematicians.

In the expansion of $(a + x)^m$ with arbitrary m, all the terms have the same sum of powers of a and x, each subsequent term differing from the preceding one by the factor x/a and the coefficient. A physicist would say that a and x in formula (6.4.3) must have the same dimensions, and so x/a is dimensionless. From the very beginning we could take a outside the brackets:

$$(a + x)^m = a^m (1 + x/a)^m,$$

and expand $(1 + x/a)^m$ in powers of x/a.

It turns out that for all m (fractional, positive, and negative) the series (6.4.3) is suitable only for $|x/a| < 1$, that is, for $|x| < |a|$. For $|x/a| \geqslant 1$ the series in (6.4.3) is divergent. The positive integers m are an exception because in that case formula (6.4.3) consists of a finite number of terms.

Putting $a = 1$ in (6.4.3), replacing x with r (where $|r| < 1$), and confining ourselves to the first two terms on the right-hand side of (6.4.3), we arrive at the following approximate formula

$$(1 + r)^m \simeq 1 + mr,$$

which was often used above, this formula is valid only for small $|r|$ (a more exact formula has the form $(1 + r)^m \simeq 1 + mr + m\frac{m(m-1)}{2}r^2\Big)$.

Formula (6.4.6) offers a good method for taking roots. Here, the smaller the ratio $|x/a|$ the fewer terms one has to take in (6.4.6) to attain a specified accuracy.

Exercises

6.4.1. Using a series expansion, find $\sqrt{1.1}$ and $\sqrt{1.5}$ as $\sqrt{1+x}$ at $x = 0.1$ and $x = 0.5$ retaining two, three, and four terms in the expansion. Compare the results with the tabular values.

6.4.2. Show that for $|x| < 1$ the approximate formulas (a) $\sqrt[n]{1+x} \simeq 1 + x/n$, and (b) $\sqrt[n]{1+x} \simeq 1 + \frac{x}{n} + \frac{n-1}{2n^2}x^2$ are valid and that the smaller the value of x, the more accurate formulas (a) and (b) are (formula (b) is more accurate than formula (a)).

6.4.3. Using the formulas of the preceding exercise, find $\sqrt[3]{1.2}$, $\sqrt[3]{1.1}$, and $\sqrt[3]{1.05}$, compare the values obtained with tabular values.

6.4.4. Find $\sqrt{6}$ to three decimal places. [*Hint.* Employ the fact that $6 = 4 + 2$ and $\sqrt{4} = 2$ and apply formula (6.4.6).]

6.4.5. Why is it impossible to expand $y = \sqrt{x}$ by Maclaurin's formula? Can the function $y = \sqrt[3]{x^{1000}}$ be expanded using this formula?

6.5 The Order of Increase and Decrease of Functions. L'Hospital's Rule

The series expansion of functions yields a general method for reducing different functions to the same form and enables one to compare the functions. This method of comparison is needed, for example when we consider the ratio of two functions, $f(x)/g(x)$, for a value of the independent variable x for which the values of the two functions are close to zero. In the computation of derivatives it was demonstrated that the ratio of two almost-zero quantities can be a quite definite number. Such ratios may prove to be not very small (but not very large, either). In certain cases, this ratio may be equal to zero or infinity (positive or negative), and such cases can be classified. A few examples will suffice. For the sake of simplicity of notation, we take examples in which the value of x that interests us is equal to zero.

For small values of x, the functions $\sin x$ and $\tan x$ are also small. The function e^x and $\cos x$ are close to unity and, hence, $e^x - 1$ and $1 - \cos x$ are small. Here the smaller the value of $|x|$ the closer are the values of the functions $\sin x$, $\tan x$, $e^x - 1$, and $1 - \cos x$ to zero.

Let us compare these functions with x. To do this, we write out their Maclaurin-series expansions:

$$\sin x = x - \frac{x^3}{6} + \dots,$$

$$\tan x = x + \frac{x^3}{3} + \dots,$$

$$1 - \cos x = \frac{x^2}{2} - \frac{x^4}{24} + \dots,$$

$$e^x - 1 = x + \frac{x^2}{2} + \dots. \qquad (6.5.1)$$

This for one, yields $(\sin x)/x = 1 - x^2/6 + \dots$. Hence, $(\sin x)/x \to 1$ as $x \to 0$, or $\lim_{x \to 0} \frac{\sin x}{x} = 1$. Similarly, from (6.5.1) we find that

$$\frac{\tan x}{x} = 1 + \frac{x^2}{3} + \dots \to 1 \text{ as } x \to 0,$$

$$\frac{1 - \cos x}{x^2} = \frac{1}{2} - \frac{x^2}{24} + \dots \to \frac{1}{2} \text{ as } x \to 0,$$

$$\frac{1 - \cos x}{x} = \frac{x}{2} - \frac{x^3}{24} + \dots \to 0 \text{ as } x \to 0,$$

$$\frac{e^x - 1}{x} = 1 + \frac{x}{2} + \dots \to 1 \quad \text{as} \quad x \to 0.$$

More complicated relations can also be found. For instance, from

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \dots,$$

$$\tan x = x + \frac{1}{3} x^3 + \frac{2}{15} x^5 + \dots$$

it follows that

$$\tan x - \sin x = \frac{1}{2} x^3 + \frac{1}{8} x^5 + \dots$$

and, hence,

$$\frac{\tan x - \sin x}{x^3} \to \frac{1}{2} \quad \text{as} \quad x \to 0.$$

A scale can be constructed of the order of decrease of various functions as x tends to zero. If $f(x)$ tends to zero as $x \to 0$, then we term the *order of decrease* (or *order of smallness*) of $f(x)$ with respect to x the power of x that decreases just as rapidly as $f(x)$. Precisely, if we say that $f(x)$ has kth order of decrease with respect to x, this means that it decreases as x^k, that is, the ratio $f(x)/x^k$ has for its limit, as $x \to 0$, a finite nonzero number.

Thus, $\sin x$, $\tan x$, $e^x - 1$ decrease by order one as $x \to 0$, while $1 - \cos x$ decreases by order two, and $\tan x - \sin x$ decreases by order three with respect to x.

In certain cases it is possible to determine the order of decrease without

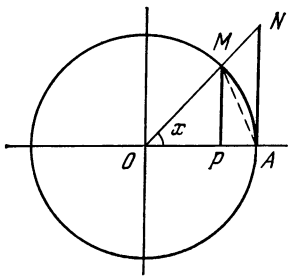


Figure 6.5.1

a series expansion. For instance, from the well-known Figure 6.5.1, where MP and NA are lines of the sine and the tangent corresponding to $\angle AOM = x$, we can easily see that $\tan x > x > \sin x$ (the reasoning is based on the fact that $S_{\Delta OMA} < S_{\text{sctr}OMA} < S_{\Delta ONA}$). Since $\sin x \simeq \tan x$ for small x's (because in this case $\sin x/\tan x = \cos x \simeq 1$), we conclude that $\sin x \simeq x$ and $\tan x \simeq x$ when x is small, that is, $\sin x$ and $\tan x$ have the first order of decrease. Next, since $1 - \cos x = 2\sin^2 (x/2)$, and $\sin (x/2)$ is of the first order of decrease, we conclude that $1 - \cos x$ must be of the second order of decrease. Finally, the function $\tan x - \sin x$ can be written as $\sin x/\cos x - \sin x = (\sin x/\cos x)(1 - \cos x)$. Since for small x's the function $\sin x$ is of the first order of decrease, $1 - \cos x$ is of the second order, and $\cos x \simeq 1$, it is clear that $\tan x - \sin x$ must be of third order of decrease. However, these concrete devices require a great deal of ingenuity and so precisely for this reason a general method that operates without failure is particularly useful.

Such a relationship between ingenious solutions of individual problems and general methods is in evidence everywhere: the properties of tangent lines to a parabola, the area of a circle, the volume of a pyramid, and the volume of a sphere were all familiar to the ancient Greeks, but only differential and integral calculus provided us with general and simple methods for solving *all* problems of that type, methods that any engineer, physicist, or student can master, while the ingenious methods of ancient scholars were suitable only for the best minds of the time.

Using series, it is possible not only to find the ratio of a function to a power of x, but also the ratio of one function to another. Here are some examples:

$$\frac{e^x - 1}{\sin x} = \frac{x + \frac{x^2}{2} + \frac{x^3}{6} + \dots}{x - \frac{x^3}{6} + \dots}$$

$$= \frac{1 + \frac{x}{2} + \dots}{1 - \frac{x^2}{6} + \dots} \to 1 \quad \text{as} \quad x \to 0,$$

$$\frac{e^x - 1}{1 - \cos x} = \frac{x + \frac{x^2}{2} + \dots}{\frac{x^2}{2} - \frac{x^4}{24} + \dots}$$

$$= \frac{1 + \frac{x}{2} + \dots}{\frac{x}{2} - \frac{x^3}{24} + \dots} \to \infty \quad \text{as} \quad x \to 0,$$

$$\frac{e^x - 1}{\sqrt{x}} = \frac{x + \frac{x^2}{2} + \dots}{\sqrt{x}} = \frac{\sqrt{x} + \frac{x^{3/2}}{2} + \dots}{1}$$

$\to 0$ as $x \to 0$.

The coefficients of Maclaurin's series are expressed in terms of derivatives. It is therefore possible to state the results obtained by means of series in the form of rules referring to derivatives. If $f(0) = g(0) = 0$, then

$$f(x) = f(0) + f'(0)\, x + \frac{1}{2} f''(0)\, x^2 + \dots,$$

$$g(x) = g(0) + g'(0)\, x + \frac{1}{2} g''(0)\, x^2 + \dots$$

can be simplified thus:

$$f(x) = f'(0)\, x + \frac{1}{2} f''(0)\, x^2 + \dots,$$

$$g(x) = g'(0)\, x + \frac{1}{2} g''(0)\, x^2 + \dots.$$

From this we have

$$\frac{f(x)}{g(x)} = \frac{f'(0)\,x + \frac{1}{2}f''(0)\,x^2 + \dots}{g'(0)\,x + \frac{1}{2}g''(0)\,x^2 + \dots}$$

$$= \frac{f'(0) + \frac{1}{2}f''(0)\,x + \dots}{g'(0) + \frac{1}{2}g''(0)\,x + \dots} \to \frac{f'(0)}{g'(0)}$$

as $x \to 0$

and, hence,

if $f(0) = g(0) = 0$,

then $\frac{f(x)}{g(x)} \to \frac{f'(0)}{g'(0)}$ as $x \to 0$ (6.5.2)

(if, in addition, $f'(0) = g'(0)$, then the ratio $f'(x)/g'(x)$ tends to the limit of the ratio $f''(x)/g''(x)$ as $x \to 0$[6.10].

Of course, rule (6.5.2) retains its meaning completely when $f(a) = g(a) = 0$, where a is an arbitrary number, and we are interested in the value of the ratio $f(x)/g(x)$ when x is close to a:

if $f(a) = g(a) = 0$, then

$$\frac{f(x)}{g(x)} \to \frac{f'(a)}{g'(a)} \quad \text{as} \quad x \to a \qquad (6.5.2a)$$

(if, in addition, $f'(a) = g'(a)$, then the limits of the ratios $f'(x)/g'(x)$ and $f(x)/g(x)$ coincide as $x \to a$, in view of which $f(x)/g(x) \to f''(a)/g''(a)$ as $x \to a$). Here, one must use Taylor's series instead of Maclaurin's series (see Exercise 6.5.2).

Thus, if the values of two functions are close to zero, that is, if the two functions vanish at a single value of the independent variable in the vicinity of which these functions are considered, then the ratio of the functions can be replaced with the ratio of their derivatives. This result is called ***L'Hospital's rule*** (actually discovered by Johann Bernoulli and given to L'Hospital in return for salary).[6.11]

After studying series, it is more convenient not to bother remembering some special rules for finding derivatives but, for small values of x, to use series in which the function is expanded in powers of x. Whenever there is a sum of different powers of x, we leave only the lowest-degree term when passing to small values of x.

Just as we considered, for small x, the order of decrease of functions equal to zero when $x = 0$, we can examine the behavior of functions when x *increases without bound*, that is, as $x \to \infty$. Here it is appropriate to expand the function in powers of $1/x$ (since $1/x$ is small by hypothesis) via substitution of t for $1/x$. If our function is a polynomial, then it is obvious that for large values of x only the highest-degree term in x is of importance, since all other terms are smaller: $ax^n + bx^{n-1} + \dots = ax^n\left(1 + \frac{a}{b}\frac{1}{x} + \dots\right)$; here the expansion in the parentheses tends to 1 as $x \to \infty$. We can speak of the ***order of increase*** of a function, that is, that a function increases as x, as x^2, as x^3, etc. It is said that a function $f(x)$ that increases without bound in absolute value as $x \to \infty$ has kth order of increase if the ratio $f(x)/x^k$ has for its limit, as $x \to \infty$, a finite nonzero number. It is also clear that any polynomial of degree n has an order of increase equal to n.

A fact of prime importance is that the *function e^x increases faster than any power x^n* for x increasing without bound, that is, e^x has an infinitely large order of increase, or $k = \infty$. To prove this, use the series expansion (6.2.2) of e^x which, as pointed out above, is valid

---

[6.10] Since here, in view of the same formula (6.5.2), $\lim_{x\to 0}\frac{f'(x)}{g'(x)} = \frac{f''(0)}{g''(0)}$, we arrive at $\lim_{x\to 0}\frac{f(x)}{g(x)} = \frac{f''(0)}{g''(0)}$. Similarly, if $f(0) = f'(0) = f''(0) = g(0) = g'(0) = g''(0) = 0$, then $\lim_{x\to 0}\frac{f(x)}{g(x)} = \frac{f'''(0)}{g'''(0)}$, and so on.

[6.11] The same term is used in similar situations, for instance, the one mentioned in footnote 6.10 and those examined in Exercises 6.5.3 and 6.5.4.

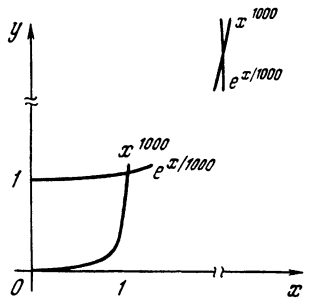


Figure 6.5.2

for any value of x. We have

$$\frac{e^x}{x^n} = \frac{1}{x^n} + \frac{1}{x^{n-1}} + \ldots + \frac{1}{n!} + \frac{x}{(n+1)!} + \frac{x^2}{(n+2)!} + \ldots \qquad (6.5.3)$$

For a given n and a sufficiently large x, the fraction e^x/x^n will become as large as desired due to terms with positive powers of x on the right-hand side of (6.5.3). Clearly, the same goes for the function e^{kx} with any positive value of k (or for the function $a^x = e^{x \ln a}$, with $a > 1$): setting $kx = y$, we find that

$$\frac{e^{kx}}{x^n} = k^n \frac{e^{kx}}{(kx)^n} = k^n \frac{e^y}{y^n} \to \infty \text{ as } y \to \infty,$$

i.e. as $x \to \infty$. (6.5.4)

Thus, any exponential function e^{kx}, where $k > 0$ (or a^x, where $a > 1$) grows faster, as $x \to \infty$, than any power function x^n (or cx^n, where both n and c are positive); although for small values of x the graph, say, of $y = x^{1000}$ slopes upward much faster than the graph of the function $y_1 = e^{0.001x}$ (Figure 6.5.2), with the growth of x the graph of function y_1 will finally intersect the graph of function y, since y_1 will overcome y.

As $x \to \infty$, the exponential function with a negative exponent *decreases faster than any negative-power function* x^{-n}. This assertion, for arbitrary $n > 0$, is written

$$f = \frac{e^{-x}}{x^{-n}} = x^n e^{-x} \to 0 \quad \text{as} \quad x \to \infty \qquad (6.5.5)$$

It is risky to use the series expansion of e^{-x} for large x to prove this because the expansion is an alternating expansion, so that the various terms compensate each other to a certain extent. We therefore consider the reciprocal:

$$1/f = x^{-n}/e^{-x} = e^x/x^n.$$

According to (6.5.4), for arbitrary n the quantity $f^{-1} = e^x/x^n$ tends to ∞ as $x \to \infty$ which simply means that $f \to 0$ as $x \to \infty$.

To summarize: in the limit, for large absolute values of the independent variable in the exponent, the exponential function e^x depends more strongly on x than any constant power of x, that is, for any integer n, the function e^x increases faster than x^n and e^{-x} decreases faster than x^{-n}. This is vividly demonstrated in the table for x^5 and e^x:

x	1	3	5	10
x^5	1	243	3125	10^5
e^x	2.72	20	150	2×10^4
$\frac{x^5}{e^x} = \frac{e^{-x}}{x^{-5}}$	0.37	12	21	5

x	20	50	100
x^5	3×10^6	3×10^8	10^{10}
e^x	4×10^8	5×10^{21}	10^{43}
$\frac{x^5}{e^x} = \frac{e^{-x}}{x^{-5}}$	0.01	10^{-13}	10^{-33}

What we have said about the exponential function can be applied to the logarithmic function $y = \log_a x$, namely, the *logarithmic function* $\log_a x$ (with an arbitrary base $a > 1$) *increases, as* $x \to \infty$, *slower than any* (however small) *power* x^n *of the independent variable*, while the function $e^{-1/x}$ for $x > 0$ and $x \to 0$ decreases faster than any positive power of x:

$$\frac{\log_a x}{x^n} \to 0 \text{ as } x \to \infty, \qquad (6.5.6)$$

$$\frac{e^{-1/x}}{x^n} \to 0 \text{ as } x \to 0 \text{ and } x > 0. \qquad (6.5.7)$$

To prove (6.5.6) it is sufficient to introduce the substitution $\log_a x = u$,

$x = a^u$, $x^n = a^{nu}$, which transforms the ratio of interest to us into $u/a^{nu} = 1 \div a^{nu}/u$, while to prove (6.5.7) it is sufficient to introduce the variable $v = 1/x$, which transforms the ratio into e^{-v}/v^{-n} (compare with (6.5.5) and also see Exercise 6.5.5).

If we turn to the function $w = e^{-1/x^2}$, which is naturally assumed equal to zero at $x = 0$ (since for small, in absolute value, x, that is, large values of $1/x^2$, $w = e^{-1/x^2}$ can be made as small as desired), we can state that the function decreases, as $x \to 0$, faster than any power of x, that is, $w/x^n \to 0$ for all $n > 0$. This implies, in turn, that it is impossible to expand this (seemingly well-behaved) function in a Maclaurin's series (see the graph of this function in Figure 15.1.1). Indeed, if the formula (6.1.19) were valid for this function, then a number of first coefficients on the right-hand side of this formula could vanish and the series expansion would be $w(x) = a_k x^k + a_{k+1}x^{k+1} + \ldots$, with $a_k \neq 0$, and the function would have, for small values of x, a finite order of decrease (k) with respect to x instead of an "infinite" order of decrease (cf. Section 15.1).

Exercises

6.5.1. Find the following limits:
(a) $\lim_{x\to 0} \frac{\ln(1+x)}{x}$, (b) $\lim_{x\to 0} \frac{\ln(1+x)-x}{x^2}$,
(c) $\lim_{x\to 0} \frac{\tan x - x}{x^2}$, (d) $\lim_{x\to 0} \frac{e^x - 1 - \tan x}{x^3}$,
(e) $\lim_{x\to 0} \frac{e^x - 1}{\sin x}$, (f) $\lim_{x\to 0} \frac{\sin x - x}{x - \tan x}$.

6.5.2. Prove the variant (6.5.2a) of L'Hospital's rule.

6.5.3. Prove that if $f(x) = f'(x) = f''(x) = \ldots = f^{n-1}(x) = 0$ at $x = 0$, and, similarly, $g(0) = g'(0) = g''(0) = \ldots = g^{(n-1)}(0) = 0$, then $\lim_{x\to 0} \frac{f(x)}{g(x)} = \frac{f^{(n)}(x)}{g^{(n)}(x)}$ (where, of course, we assume that the last ratio, which can vanish or become infinite, exists).

6.5.4. Prove that (a) if $f(x) \to 0$ as $x \to \infty$ and $g(x) \to 0$ as $x \to \infty$, the ratios $f(x)/g(x)$ and $f'(x)/g'(x)$ tend to the same limit as $x \to \infty$, and (b) if $f(x) \to \infty$ as $x \to a$ (where, possibly, $a = \infty$ and $g(x) \to \infty$ as $x \to a$, then $\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$.

6.5.5. Prove that the function $e^{-1/x}$ grows in absolute value, for $x < 0$ and $x \to 0$, faster than any (integral) negative power of x.

6.6 First-Order Differential Equations. The Case of Variables Separable

Above we encountered (e.g. see Section 3.6) the concept of a ***differential equation***, that is, an equation that connects the (unknown) function $y = y(x)$ and its derivatives:

$$F(x, y, y', y'', \ldots) = 0, \qquad (6.6.1)$$

where F is a known function depending on several variables. The laws of science and the operation of various devices or mechanisms can often be described using the language of differential equations, so that in many cases a substantial part of studying the phenomenon that interests us consists in analyzing and solving an appropriate equation. In Chapter 9 we will illustrate all that has just been said with examples taken from mechanics, a science whose basis is formed by *Newton's second law*, the differential equation (9.4.2).

The ***order*** of a differential equation is the order of the highest derivative which appears. Here we limit our discussion to *first-order* differential equations:

$$F(x, y, y') = 0, \qquad (6.6.2)$$

or, if we solve this equation for the derivative y',

$$y' = f(x, y). \qquad (6.6.3)$$

The simplest case of an equation of type (6.6.3) is when f is independent of y, that is, when the equation has the form

$$y' = f(x). \qquad (6.6.4)$$

This equation implies that y is the *antiderivative* of function $f(x)$, or the *indefinite integral* of $f(x)$:

$$y = \int f(x)\,dx. \qquad (6.6.5)$$

Chapters 3 and 5 were devoted to the solution of Eq. (6.6.4). Of course, this

equation is the simplest, and its solution follows from the very definition of the infinite integral, speaking of (general) differential equations, it is useful to bear in mind this first example that sheds light on the general situation. For instance, already in this example we see the ambiguity in the statement of the problem, namely, the presence of an *infinitude* of solutions to a given equation. For instance, in the case of Eq. (6.6.4) these solutions have the form

$$y = G(x) + C, \tag{6.6.5a}$$

where $G'(x) = f(x)$. The graphs of all the functional relations described by $y = y(x)$ are obtained from a single graph by parallel translation along the y axis (see Figure 6.6.1). To select a unique solution to the problem, we must specify a value $y = y_0$ for a given initial value $x = x_0$ of the independent variable, the so-called *initial condition* for differential equation (6.6.4); this initial condition determines the unique solution

$$y = y_0 + \int_{x_0}^{x} f(x)\,dx \tag{6.6.5b}$$

to Eq. (6.6.4), that is, it determines the unique integral curve $y = y(x)$ passing through the given point $M_0 = M_0(x_0, y_0)$.

Many textbooks are devoted to the theory of differential equations, and we do not intend to repeat their content here. We will limit our discussion to equations of type (6.6.3) that can be solved as simply as Eq. (6.6.4) for the antiderivative (or in the same complicated way as Eq. (6.6.4), since finding the antiderivative may constitute a complex problem). Precisely, if the function $f(x, y)$ is the product $g(x)\,h(y)$ of a function of x and a function of y, then the variables x and y in the equation

$$\frac{dy}{dx} = g(x)\,h(y) \tag{6.6.6}$$

can be *separated* by carrying all terms dependent on x over to one side of the equation and the terms dependent on y over to the other. Such equations are called ***differential equations with variables separable.***

Let us divide both sides of Eq. (6.6.6) by $h(y)$:

$$\frac{1}{h(y)}\frac{dy}{dx} = g(x), \tag{6.6.6a}$$

and then multiply both sides of (6.6.6a) by dx:

$$\frac{dy}{h(y)} = g(x)\,dx. \tag{6.6.6b}$$

Now, to find the solution to Eq. (6.6.6) we need only integrate the left- and right-hand sides of Eq. (6.6.6b):

$$\int \frac{dy}{h(y)} = \int g(x)\,dx. \tag{6.6.7}$$

If these integrals can be computed, then we have the function $y = y(x)$ in implicit form:

$$H(y) = G(x) + C, \tag{6.6.8}$$

where $H(y) = \int \frac{dy}{h(y)}$ and $G(x) = \int g(x)\,dx$ are definite forms of these integrals, and C is an arbitrary constant.

The simplest example of an equation with variables separable is Eq. (6.6.4), which is obtained if we put $g \equiv f$ and $h \equiv 1$ in Eq. (6.6.6). Substituting this into (6.6.7), we arrive at the solution (6.6.5) to Eq. (6.6.4). An equally simple case is the one where $f(x, y) = f(y)$ in (6.6.3) depends only on y, that is, where (6.6.3) becomes

$$y' = f(y), \tag{6.6.4a}$$

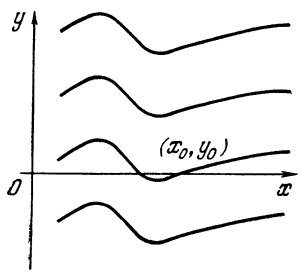


Figure 6.6.1

here, as we see, $g \equiv 1$ and $h \equiv f$. For one, the equation

$$y' = ky, \tag{6.6.9}$$

which we will often encounter below (e.g. see Chapter 8), also belongs to this type of differential equations of variables separable. The solution (6.6.7) to Eq. (6.6.9) assumes the form $\int dy/y = k \int dx$, or $\ln y = kx + C$, or, finally,

$$y = e^{kx+C} = Ae^{kx}, \tag{6.6.10}$$

where we have put $A = e^C$ (since C is an arbitrary number, we conclude that A is an arbitrary (positive) number). Thus, we see once more that the exponential function $y = Ae^{kx}$ is characterized by the proportionality (expressed by (6.6.9)) between the derivative of the function and the function proper.

Let us examine some simple examples of differential equations with variables separable.

1. Suppose that

$$y' = xy, \text{ or } \frac{dy}{dx} = xy. \tag{6.6.11}$$

Separation of variables yields $\frac{dy}{y} = x\,dx$, whence $\int y^{-1}\,dy = \int x\,dx$, or $\ln y = x^2/2 + C$, or, finally,

$$y = \exp\left(\frac{x^2}{2} + C\right) = A \exp\left(\frac{x^2}{2}\right), \tag{6.6.12}$$

where, as before, $A = e^C$. The graphs of the function (6.6.12) are depicted in Figure 6.6.2*a*.

2. Suppose that

$$y' = \frac{y}{x}, \text{ or } \frac{dy}{dx} = \frac{y}{x}. \tag{6.6.13}$$

Then $\int y^{-1}\,dy = \int x^{-1}\,dx$, or $\ln|y| = \ln|x| + C$, or, finally,

$$y = Ax \tag{6.6.14}$$

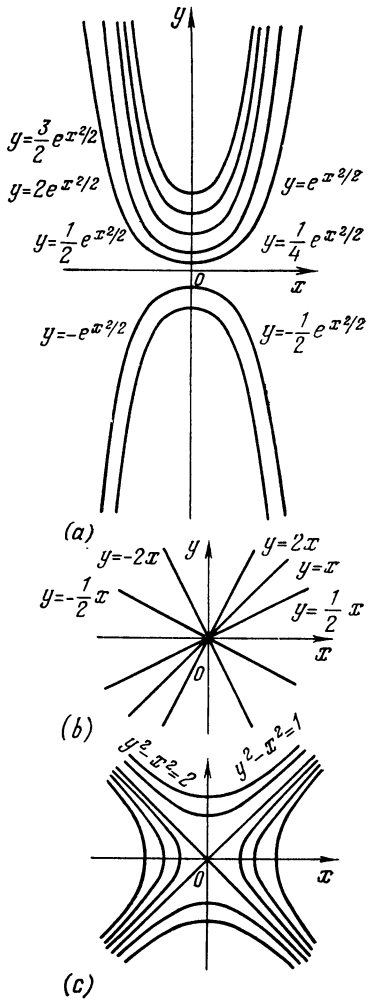


Figure 6.6.2

Figure 6.6.2*b*; here again $|A| = e^C$).

3. If

$$y' = \frac{x}{y}, \text{ or } \frac{dy}{dx} = \frac{x}{y}, \tag{6.6.15}$$

then we obtain $\int y\,dy = \int x\,dx$, or $y^2/2 = x^2/2 + C$, or, finally,

$$y^2 - x^2 = a, \tag{6.6.16}$$

where we have denoted $2C$ by a (Figure 6.6.2*c*).

Of course, all these examples were specially selected, since here not only the given equation (6.6.3) is that of type (6.6.6) but the

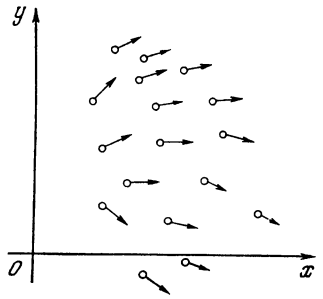


Figure 6.6.3

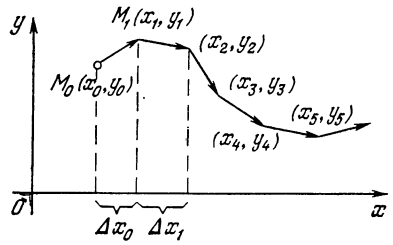


Figure 6.6.4

integrals in solution (6.6.7), too, can easily be calculated, so that the general solutions (6.6.12), (6.6.14), and (6.6.16) are easily found. But in real life the situation is not all that simple. In the general case, the right-hand side of Eq. (6.6.3), of course, need not separate into two factors each depending on x or on y; besides, if the variables do separate, this does not mean that the integrals in (6.6.7) can be expressed in terms of elementary functions. The only path here is to resort to the various methods of approximate (numerical) solution of differential equations.

Note that Eq. (6.6.3) specifies the dependence of the slope y' of the solution curve $y = y(x)$ for this equation on the position of point (x, y) in the plane; in other words, Eq. (6.6.3) specifies a *field of directions* in the xy-plane, that is, it specifies at each point $M = M(x, y)$ a direction with a given slope $k = y' = f(x, y)$ (Figure 6.6.3). This fact suggests a natural method for an approximate solution of Eq. (6.6.3).

Suppose that we wish to find the curve $y = f(x)$ that passes through a given point $M_0(x_0, y_0)$ in the xy-plane and such that the function $y(x)$ satisfies condition (6.6.3). Let us move from point M_0 along the direction specified at this point to an adjacent point $M_1(x_1, y_1)$, where $x_1 = x_0 + \Delta x_0$, $y_1 = y_0 + \Delta y_0$, and $\Delta y_0/\Delta x_0 = k_0 = f(x_0, y_0)$. Then from point M_1 we move along the direction specified at this point to another adjacent point $M_2(x_2, y_2)$, where $x_2 = x_1 + \Delta x_1$, $y_2 = y_1 + \Delta y_1$, and $\Delta y_1/\Delta x_1 = k_1 = f(x_1, y_1)$, and so on. Repeating this procedure many times, we obtain a broken curve (known as the ***Euler broken line***; see Figure 6.6.4), which (for small "steps" Δx_0, Δx_1, etc.) yields a rough picture of the behavior of the *integral curve* (i.e. solution) $y = y(x)$ to Eq. (6.6.3) (in a certain sense the Euler broken lines "converge" to the integral curve when the lengths of all of its "chains" tend to zero). Computer calculations carried out via this method prove to be simple, and the method on the whole is convincing and reliable.

Another method of solving differential equations is based on the use of power series (here, just as in the case with the Euler broken lines, we are speaking, of course, not only of equations with variables separable). Suppose that we have an equation of type (6.6.3) in which the dependence of $f(x, y)$ on x and y is quite simple (the meaning of this rather hazy statement will be elaborated on below) and wish to find the solution $y = y(x)$ to this equation that passes through a fixed point $M_0(x_0, y_0)$ in the xy-plane. We assume that the function $y(x)$ in the vicinity of point $x = x_0$, at which $y_0(x_0) = a_0$, is expandable in a Taylor's series

$$y(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \ldots \qquad (6.6.17)$$

Then[6.12]

$$y'(x) = a_1 + 2a_2(x - x_0) + 3a_3(x - x_0)^2 + \ldots \qquad (6.6.18)$$

Substituting (6.6.17) and (6.6.18) into the initial equation (6.6.3), we in many cases arrive at the condition of equality of two power series: the series (6.6.18) for y' and the series expressing the function $f(x, y)$, where y is given by the series (6.6.17) (it is at this point that it is desirable for the function f to be "simple", that is, such that $f(x, a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots)$ can be represented as a power series $b_0 + b_1(x - x_0) + b_2(x - x_0)^2 + \ldots$). Equating the coefficients of like powers of x of the two series, we can then often use the system $a_1 = b_0$, $2a_2 = b_1$, $3a_3 = b_2$, etc. to find numerically the coefficients a_1, a_2, ... of the series (6.6.17), that is, find the solution to Eq. (6.6.3) (since the coefficient a_0 is known beforehand).[6.13]

[6.12] The reader will have to take it for granted that a power series, like the one on the right-hand side of (6.6.17), can be differentiated term by term. We have not proved that this can be done, but the supposition seems quite natural.

[6.13] Note that expansion (6.6.17), as we know, is only valid in a certain neighbourhood of point $x = x_0$, whereby it often happens that, upon reaching a certain point $x = x_1$, we are forced to discard expansion (6.6.17) and look for the solution in the form of a series expansion in powers of $x - x_1$, and so on. The examples we consider below are selected so that this difficulty does not arise.

To illustrate this method, let us take two simple examples, the first being of a purely illustrative value since the respective equation can be solved without resorting to series, the second being typical.

1. Let us return to the equation considered above, (6.6.11), or $y' = xy$, and write

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots. \quad (6.6.19)$$

Then

$$y' = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \ldots. \quad (6.6.20)$$

Substituting these two series expansions into (6.6.11), we get

$$a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + 5a_5x^4 + 6a_6x^5 + \ldots = a_0x + a_1x^2 + a_2x^3 + a_3x^4 + a_4x^5 + a_5x^6 + \ldots. \quad (6.6.21)$$

Equating the coefficients of like powers of x, we get

$$a_1 = 0, \quad a_2 = \frac{a_0}{2}, \quad a_3 = \frac{a_1}{3} = 0,$$

$$a_4 = \frac{a_2}{4} = \frac{a_0}{2 \cdot 4}, \quad a_5 = \frac{a_3}{5} = 0,$$

$$a_6 = \frac{a_4}{6} = \frac{a_0}{2 \cdot 4 \cdot 6}, \quad \ldots,$$

where $a_0 = y_0 = y(0)$. In view of these equalities we can write expansion (6.6.19) in the following form:

$$y = a_0 \left(1 + \frac{x^2}{2} + \frac{1}{2}\frac{x^4}{4} + \frac{1}{2 \cdot 3}\frac{x^6}{8} + \ldots\right),$$

from which it follows that $y = a_0 e^{x^2/2}$ (substitute $x^2/2$ for x in the right-hand side of (6.2.2)).

2. Suppose that

$$y' = xy + 1. \quad (6.6.22)$$

This is not a differential equation with variables separable, and it can be proved that its solution cannot be written in the form of an explicit dependence of y on x. Nevertheless, it is easy to express the solution to Eq. (6.6.22) in the form of a power series.

Let us again employ expansions (6.6.19) and (6.6.20) and substitute them into (6.6.22). The result is

$$a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + 5a_5x^4 + 6a_6x^5 + \ldots = 1 + a_0x + a_1x^2 + a_2x^3 + a_3x^4 + a_4x^5 + \ldots,$$

which implies that

$$a_1 = 1, \quad a_2 = \frac{a_0}{2}, \quad a_3 = \frac{a_1}{3} = \frac{1}{1 \cdot 3},$$

$$a_4 = \frac{a_2}{4} = \frac{a_0}{2 \cdot 4}, \quad a_5 = \frac{a_3}{5} = \frac{1}{1 \cdot 3 \cdot 5},$$

$$a_6 = \frac{a_4}{6} = \frac{a_0}{2 \cdot 4 \cdot 6}, \quad \ldots.$$

This enables finding the general solution to Eq. (6.6.22):

$$y = \left(\frac{x}{1} + \frac{x^3}{1 \cdot 3} + \frac{x^5}{1 \cdot 3 \cdot 5} + \frac{x^7}{1 \cdot 3 \cdot 5 \cdot 7} + \ldots\right) + a_0 \left(1 + \frac{x^2}{2} + \frac{x^4}{2 \cdot 4} + \frac{x^6}{2 \cdot 4 \cdot 6} + \ldots\right), \quad (6.6.23)$$

where the value of a_0 can be determined from the *initial conditions*; for instance, if we know that $y(0) = 0$, then $a_0 = 0$ (see Exercise 6.6.4).

Exercises

6.6.1. Solve the following differential equations (with variables separable): (a) $y' = y^2$, (b) $y' = k(x/y)$, (c) $y' = k(y/x)$, (d) $y' = x^2y^3$, (e) $y' = \cos x \sin y$.

6.6.2. Prove that the straight lines (a) $y = 0$ and (b) $y = \pm x$ depicted in Figure 6.6.2*a* and Figure 6.6.2*b* are also integral curves of Eqs. (6.6.11) and (6.6.15), respectively.

6.6.3. Point out the "singular" solution to Eq. (6.6.9) similar to the "singular" solutions of Eqs. (6.6.11) and (6.6.15) considered in Exercise 6.6.2.

6.6.4. Prove that the integral curve (6.6.23) of Eq. (6.6.22) that passes through the origin can be written via the following formula:

$$y = e^{x^2/2}\Phi(x), \text{ where } \Phi(x) = \int_0^x e^{-t^2/2}\,dt.$$

6.6.5. Solve by the method of series expansion (a) Eq. (6.6.9), (b) Eq. (6.6.13), and (c) equation (a) in Exercise 6.6.1.

6.7* The Differential Equation for Water Flow from a Vessel

As one more example of a problem that requires solving a differential equation, we now consider the flow of water from a vessel (see Section 3.6). We will assume that the vessel has an opening near the bottom (see Figure 3.6.3) or that a small pipe is connected to the bottom, in view of which the water can flow out of the vessel, on

the other hand, we will assume that the vessel can also receive an inflow of water from some outside source. The statement of this problem is very simple and pictorial. At the same time, the mathematical methods required to describe the flow of water are also employed in more complicated and interesting problems.

So let us imagine a vessel with water flowing in or out. We denote the volume of water in the vessel by V. This volume varies with time, which means that V is a function of time. What meaning has the quantity dV/dt?

It is quite clear that $dV \simeq V(t + dt) - V(t)$ is the amount of water that has entered the vessel during time dt (if dV is negative, then it is the amount of water that has left the vessel during time dt). Therefore, $dV/dt = q(t)$ is the *rate of change of the amount of water* in the vessel. The quantity $q(t)$ has a special name, ***flow rate***. If $q > 0$, then water is entering the vessel; if $q < 0$, water is being discharged from the vessel and the amount (or mass) of water in the vessel diminishes. In the particular case where the flow rate is constant, that is, q does not depend on time, we have $V(t + 1) = V(t) + q$. Here q is the amount of water entering the vessel in unit time (the quantity may be negative as well). In general, in the course of a unit of time (which can be a second, an hour, or even a year) the value of q may change, whereby we define q as a derivative. The reader that has tackled the relationship between the average velocity and the instantaneous velocity, the ratio of increments and the derivative (see Chapter 2) needs no further explanations.

We know that

$$\frac{dV}{dt} = q(t). \tag{6.7.1}$$

If we know the dependence of the flow rate on time, or the function $q = q(t)$, the differential equation (6.7.1) in all respects is similar to Eq. (6.6.4), and the problem of finding V does not differ mathematically from the velocity-distance problem, which, as we know (see Chapter 3) is solved by computing a (definite) integral.

For this problem to have a definite solution, we must know the amount V_0 of water in the vessel at a definite initial time t_0. The condition that $V = V_0$ at $t = t_0$ is called the *initial condition*, which specifies a unique solution to Eq. (6.7.1) (cf. Section 6.6).

The amount (volume) of water received by, or flowing out of, the vessel during time t_0 to t_1 is $\int_{t_0}^{t_1} q(t)\,dt$. Whence the amount of water in the vessel at time t_1 is

$$V(t_1) = V_0 + \int_{t_0}^{t_1} q(t)\,dt. \tag{6.7.2}$$

This expression holds true for any time t_1 and, consequently, fully determines the desired dependence of V on t_1. At $t_1 = t_0$, the integral in (6.7.2) is zero and $V(t_0) = V_0$. Thus, solution (6.7.2) does indeed satisfy the stated condition relative to the amount of water at time t_0 (the initial condition).

Formula (6.7.2) can also be used for $t_1 < t_0$, but its meaning is different from that when $t_1 > t_0$. For $t_1 > t_0$ the quantity $V(t_1)$ is the amount of water that will be in the vessel at time t_1 if at time t_0 the amount of water was V_0 and the flow rate is given by the function $q = q(t)$. For $t_1 < t_0$ the quantity $V(t_1)$ is the amount of water that must be in the vessel at time t_1 so that at a later time, by time t_0, there should be an amount of water equal to V_0, with the flow rate fixed by the function $q = q(t)$.

Instead of writing t_1 in (6.7.2) we can simply write t, although such notation is not rigorous because t is the integration variable, but this is not really confusing. Formula (6.7.2) takes the form

$$V(t) = V_0 + \int_{t_0}^{t} q(t)\,dt. \tag{6.7.3}$$

Remember only that $q(t)$ here is not the value of q at the upper limit of integration but the (variable) function of the variable of integration, which runs through all values from t_0 to t.

Formula (6.7.3), which yields the solution to the water-discharge problem if the flow rate $q(t)$ is given and so is the amount $V_0 = V(t_0)$ at the initial time $t = t_0$, can also be obtained by somewhat different reasoning. From (6.7.1), by virtue of the definition of the indefinite integral, it follows that

$$V(t) = \int q(t)\,dt.$$

Suppose that the indefinite integral of the function $q(t)$ has been found in some way. Denote it by $I(t)$. Then

$$\int q(t)\,dt = I(t) + C,$$

where C is the constant of integration. From this,

$$V(t) = I(t) + C. \qquad (6.7.4)$$

To determine the constant of integration, let us take advantage of the initial condition, that is, let us require that at $t = t_0$ we have $V = V_0$. Substituting t_0 for t in (6.7.4), we get $V_0 = I(t_0) + C$, whence $C = V_0 - I(t_0)$. Substituting this value of C into (6.7.4) yields

$$V(t) = V_0 + I(t) - I(t_0),$$

which coincides with (6.7.3), since

$$\int_{t_0}^{t} q(t)\,dt = I(t)\,|_{t_0}^{t} = I(t) - I(t_0)$$

Formula (6.7.4) may be termed the *general solution* of Eq. (6.7.1). By choosing one or another value of C, we can, from formula (6.7.4), obtain a variety of *particular solutions* that correspond to different initial conditions.

Ordinarily, however, the flow rate is not known as a function of time. More often we know a physical law describing the flow rate as a function of the water head, that is, of the height

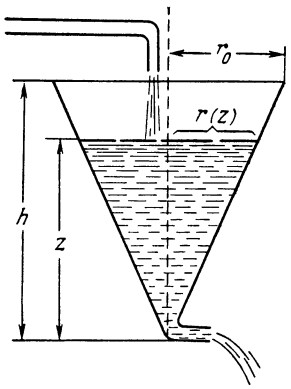


Figure 6.7.1

of water level z (Figure 6.7.1). For example, when water is flowing out of a long thin pipe, we can assume that

$$q = -kz, \qquad (6.7.5)$$

where coefficient k is a positive constant, the minus sign meaning that the water *is being discharged.* When the water flows out through an opening in a thin wall, the law is quite different:

$$q = -a\sqrt{z} \qquad (6.7.5a)$$

(this law was established by a pupil of Galileo Galilei, one Evangelista ***Torricelli*** (1608-1647), an Italian physicist and mathematician). Also possible is a combination of a constant influx of water from above (q_0) and a discharge of water by law (6.7.5) or (6.7.5a):

$$q = q_0 - kz \text{ or } q = q_0 - a\sqrt{z}. \qquad (6.7.6)$$

In each of these cases, until the problem has been solved, we do not know the dependence of water level on time, $z = z(t)$, and hence we do not know the flow rate. For this reason, the problem of determining V from the equation

$$\frac{dV}{dt} = q(z) \qquad (6.7.7)$$

cannot be reduced to the preceding problem.

Here we stated the problem in the general case for an arbitrary relation of

the flow rate $q = q(z)$ as a function of the level z. Equation (6.7.7) involves two unknown quantities: the amount (volume) of water V and the water level z. It is clear that these quantities are not independent. A definite water level is associated with a very definite amount of water so that V is a function of z, or $V = V(z)$.

It is clear that the form of this function is determined by the shape of the vessel. For a *cylindrical* vessel, for example, of radius r_0 we have $V(z) = \pi r_0^2 z$, where z is the water level reckoned from the base of the vessel. For a *conical* vessel depicted in Figure 6.7.1, the formula for the volume of a cone yields $V(z) = (1/3)\, S(z)\, z$, where $S(z) = \pi [r(z)]^2$ is the area of a cross section of the vessel at level z, and $r(z)$ is the radius of the cross section at level z. From the similarity of triangles, we find that $r(z) = r_0 z/h$, where r_0 is the radius of the upper base of the cone and h is the total height, so that $V(z) = (1/3)\,(\pi r_0^2/h^2)\, z^3$.

Substituting the function $V = V(z)$ into Eq. (6.7.7) yields

$$\frac{dV(z)}{dt} = \frac{dV(z)}{dz}\,\frac{dz}{dt} = q(z).$$

The derivative dV/dz of the volume with respect to the height of water level, z, is equal to the cross-sectional area $S(z)$ at height z (see formula (3.6.20)), or $dV(z)/dz = S(z)$. The final equation is

$$S(z)\,\frac{dz}{dt} = q(z). \tag{6.7.8}$$

Clearly, this is a differential equation with variables separable. Let us consider the solution of this equation for different functions $S = S(z)$ and $q = q(z)$.[6.14]

[6.14] Of course, other variants of water discharge from a vessel are also possible (say, a vessel with a shape that does not allow us to write out the functions $S = S(z)$ explicitly) when an exact solution of the respective equation is impossible and one is forced to resort to approximate methods (see also the text at the end of Section 6.6 in small print).

Let us rewrite Eq. (6.7.8) in the form $dz/dt = q(z)/S(z)$ and put $S(z)/q(z) = f(z)$. Then we have

$$\frac{dz}{dt} = \frac{1}{f(z)}. \tag{6.7.9}$$

We have already dealt with such equations (see Eq. (6.6.4a)). Let us rewrite Eq. (6.7.9) as

$$\frac{dt}{dz} = f(z). \tag{6.7.10}$$

This notation fits the fact that we temporarily regard t as a function of z, that is, we seek not the law $z = z(t)$ describing the change of water level with time but the inverse function $t = t(z)$, and after finding it we will express z in terms of t.

Let us multiply the left- and right-hand sides of (6.7.10) by dz and integrate:

$$\int_{t_0}^{t} dt = \int_{z_0}^{z} f(z)\, dz. \tag{6.7.11}$$

From this it follows that

$$t = t_0 + \int_{z_0}^{z} f(z)\, dz. \tag{6.7.12}$$

We have the solution to the problem: on the right is a function of z, on the left is the time t, so that t is expressed as a function of variable z (the reader will recall that t_0 is simply a number). Solution (6.7.12) also enables us, for each value of t, to find the corresponding value of z and satisfies the initial condition: $z = z_0$ at $t = t_0$ (at the initial time $t = t_0$ the level of the water in the vessel, z_0, is given).

Here are two concrete examples of the functions $f(z)$.

1. Water flowing out of a *cylindrical vessel* of radius r_0 through a *thin pipe*. Here $S(z) = \text{constant} = \pi r_0^2$ and $q = -kz$, so that Eq. (6.7.8) assumes the form

$$\pi r_0^2\,\frac{dz}{dt} = -kz, \tag{6.7.13}$$

or

$$-\pi r_0^2\,\frac{dz}{z} = k\, dt.$$

Integrating the left-hand side from z_0 to z and the right-hand side from t_0 to t, we get

$$-\pi r_0^2 (\ln z - \ln z_0) = -\pi r_0^2 \ln \frac{z}{z_0}$$

$$= \pi r_0^2 \ln \frac{z_0}{z} = k(t - t_0). \qquad (6.7.14)$$

In this example it is easy to express z in terms of t. Indeed,

$$\ln z = \ln z_0 - \frac{k}{\pi r_0^2}(t - t_0),$$

$$\text{i.e. } z = z_0 \exp\left[-\frac{k}{\pi r_0^2}(t - t_0)\right].$$

We consider two instants of time, t and $t + \Delta t$, and find the ratio $z(t + \Delta t)/z(t)$:

$$\frac{z(t+\Delta t)}{z(t)}$$

$$= \exp\left[-\frac{k}{\pi r_0^2}(t + \Delta t - t_0 - t + t_0)\right]$$

$$= \exp\left(-\frac{k\Delta t}{\pi r_0^2}\right).$$

We see that this ratio depends only on Δt and is independent of t. Therefore, the water level z falls in equal ratio during equal intervals of time; it constantly diminishes exponentially, but will never reach the value $z = 0$, that is, the water will never flow out of the vessel completely.[6.15]

The problem changes little if there is a constant influx of water into the vessel, or $q(z) = q_0 - kz$. Equation (6.7.13) then assumes the form

$$\pi r_0^2 \frac{dz}{dt} = q_0 - kz. \qquad (6.7.15)$$

Here it proves convenient to start by finding the so-called *steady-state mode*, that is, we look for a solution of this equation in the form $z = z_{st} = \text{constant}$. Since $z = z_{st}$ and $dz/dt = 0$, we get

$$q_0 - kz_{st} = 0, \quad z_{st} = q_0/k.$$

It is quite clear that if the process starts at the water level z being equal to z_{st}, then z will not change: the discharge of water from the vessel is completely compensated for by the influx of water, and the water level remains constant. Now suppose that the initial water level z_0 is higher than z_{st}, so that the discharge (proportional to z) exceeds the influx and the level z diminishes. We will look for the difference $z_1 = z - z_{st}$ between z and the steady-state level z_{st}. This difference, obviously, satisfies the equation obtained by substituting $z = z_1 + z_{st}$ and $q_0 = kz_{st}$:

$$\pi r_0^2 \frac{d(z_1 + z_{st})}{dt} = kz_{st} - k(z_1 + z_{st}),$$

$$\text{i.e. } \pi r_0^2 \frac{dz_1}{dt} = -kz_1.$$

But this equation differs from (6.7.13) only in notation, so that its solution has the form

$$z_1 = z_1^{(0)} \exp\left[-\frac{k}{\pi r_0^2}(t - t_0)\right],$$

where $z_1^{(0)} = z_0 - z_{st}$. The final result is

$$z = z_{st} + (z_0 - z_{st}) \exp\left[-\frac{k}{\pi r_0^2}(t - t_0)\right];$$

here z tends to the steady-state level z_{st} of the water but never reaches it.[6.16]

2. Water flowing out of a *conical vessel* (altitude h and radius of base r_0; see Figure 6.7.1) through a *thin pipe*. Here $S(z) = \pi r_0^2 (z/h)^2$ and $q = -kz$. In this case we have the equation

$$\frac{dz}{dt} = \frac{q(z)}{S(z)} = -\frac{kzh^2}{\pi r_0^2 z^2} = -\frac{kh^2}{\pi r_0^2}\frac{1}{z}, \qquad (6.7.16)$$

which can be rewritten thus:

$$-\frac{\pi r_0^2}{kh^2} z\, dz = dt.$$

[6.15] This paradox (the water will *never* flow out of the vessel) stems from the fact that Eq. (6.7.13) gives a relatively accurate picture of the process only when $z \gg \rho$, where ρ is the radius of the pipe through which the discharge occurs.

[6.16] The last statement remains valid for $z_0 < z_{st}$ (see Exercise 6.7.3), which indicates that the steady state $z = z_{st}$ of the water in the vessel is *stable* (see also Section 9.3).

Integration yields

$$-\frac{\pi r_0^2}{kh^2}\int_{z_0}^{z} z\,dz = -\frac{\pi r_0^2}{kh^2}\left(\frac{z^2}{2}-\frac{z_0^2}{2}\right)$$
$$= t - t_0.$$

Here we can easily express z in terms of t:

$$\frac{z^2}{2} = \frac{z_0^2}{2} - \frac{kh^2}{\pi r_0^2}(t-t_0),$$

or

$$z = \sqrt{z_0^2 - \frac{2kh^2}{\pi r_0^2}(t-t_0)}. \qquad (6.7.17)$$

This formula completely solves the problem. It is readily verified that

$$\frac{dz}{dt} = -\frac{kh^2}{\pi r_0^2}\left[z_0^2 - \frac{2kh^2}{\pi r_0^2}(t-t_0)\right]^{-1/2}$$
$$= -\frac{kh^2}{\pi r_0^2 z},$$

so that z does indeed satisfy the equation. It is also clear that $z = z_0$ at $t = t_0$. The expression (6.7.17) permits finding the time when the vessel is completely emptied:

$$z = 0 \text{ when } t = t_0 + \frac{\pi r_0^2}{kh^2}\frac{z_0^2}{2}. \qquad (6.7.18)$$

There is an interesting qualitative difference between these two examples. In the second case, the water completely flows out of the vessel during a finite time interval $T = (\pi r_0^2/kh^2)\, z_0^2/2$ (see (6.7.18), while in the first case the water level z approaches the zero level $z = 0$ asymptotically (it never reaches this level; in other words, an infinitely long interval of time is needed for the water to flow out completely).

Exercises

6.7.1. Consider the solution to Eq. (6.7.15) that satisfies the initial condition $z_0 < z_{st}$. How will the water flow out of the vessel in this case?

6.7.2. Consider the case where the water flows out of a conical vessel through a thin pipe under the condition of a constant influx of water. [*Hint.* The differential equation of this process is $\pi r_0^2 \frac{z^2}{h^2}\frac{dz}{dt} = q_0 - kz$.]

* * *

While in Chapters 2 and 3 we discussed the ideas underlying differential and integral calculus, Chapters 4 to 6 were devoted to the tools of higher mathematics, and each person who wants to master mathematical analysis to such an extent so as to have the possibility of employing the theorems and concepts in practical applications must become familiar with these tools. Of course, the main problem that confronts a scientist or engineer is the problem of adequately describing a problem of interest in mathematical terms (most often in the form of a differential equation); the analysis of the model that follows is purely mathematical (solution of a differential equation) and can be entrusted to a mathematician or passed over to a programmer for processing on a computer. (However, the formulation of the model cannot be reassigned to anyone; it must be carried out by a specialist who has a deep knowledge of the phenomenon in question.) On the basis of this the engineer may decide there is no need to study the mathematical tools, but he, or she, will be deeply mistaken. Understanding the essence of higher mathematics is closely connected with mastering certain mathematical tools: without these tools there can be no real understanding of the principles. To take an example from another field, reading special literature in a foreign language, even with the help of a dictionary (which plays the role of a programmer, so to say), is impossible without knowing the basic facts of the grammar of the language (the "theory" of the language) and without having a certain minimal vocabulary (the basic of "engineering" skills). Without knowing the grammar of the language we are helpless, since we do not know what words (and in what form) we must look up in the dictionary, while not possessing even a minimal vocabulary means that the dictionary is useless, since the totally unfamiliar text is a stumbling block. However, even a small vocabulary of

words and expressions and a rudimentary knowledge of the grammar make possible the full employment of a dictionary. In the same way, a physicist or engineer often (but not always) needs only a basic knowledge of higher mathematics and a smattering of technical skills.

Chapters 4 and 5 give an overall picture (rather sketchy, perforce) of differentiation (finding derivatives) and integration techniques. Finding derivatives is quite simple, and therefore it is desirable to develop this skill to a state when finding derivatives of simple functions is carried out automatically, so to say. Here there is no need to solve as many complicated problems as one can put his, or her, hands on, since a physicist or engineer seldom encounters complicated functions. It is much more important to know how to solve simple problems at first glance, so we recommend solving all the exercises on calculating derivatives collected in this book and not calculating derivatives of more complicated functions. If solving problems from this book proves to be insufficient for developing the necessary skills, the reader can turn to any problem book on differential calculus or even compile a list of arbitrary functions and try to find their derivatives. There is no need in trying to memorize the entire table of derivatives (see Appendix 1); it is much more advisable at the first stages in the reader's studies to copy the formulas onto a card and use the compiled table when solving problems. The formulas that are used more often will be memorized, and finally there will be no need of the card.

Integration techniques developed in Chapter 5 are much more complicated. We believe there is no need to devote too much time to these techniques. Of course, it is useful to know how to evaluate simple integrals, but it is much more important to know how to use tables of integrals freely. Of course, skills in integrating the simplest functions and the knowledge of the basic properties of integrals (say, a fluent use of the integration-by-substitution method) developed in the course of calculating integrals are necessary to every student of science and engineering, since most often the integral that you will encounter will not exactly coincide with an integral in a table but can almost always be reduced to such an integral via simple transformations, including the change of variables.

Chapter 6 begins with the topic of *series*. The technique by which complicated functions (for instance, obtained through experiments) can be reduced to simple functions is one of the basic contributions of higher mathematics to engineering practice—every physicist and engineer must be acquainted with it. The key to the technique lies in the theory of expansion of functions into power series (Sections 6.1 to 6.4); in other cases the expansion of functions into trigonometric series, that is, the representation of functions as linear combinations of trigonometric functions (see Section 10.9), constitutes a good addition to the technique.

In the present book we have restricted our discussion to elements of the corresponding theory, which is now undergoing rapid growth, for which there are several reasons; among them are the rapid development of the sections of pure mathematics that are widely used in the theory of function approximation (functional analysis, the theory of functions of a complex variable, and other fields), practical reasons connected with the constant need to simplify complicated functions, and the invention of the electronic computer, which opens up new possibilities in simplifying functions and at the same time poses requirements on the form of the functions used in computer calculations.

The second topic discussed in Chapter 6 is *differential equations*, which constitutes the basic tool of scientists and engineers. Indeed, a mathematical analysis of natural phenomena usually

starts with attempts to represent the laws of nature as differential equations — these connect the various (variable) quantities that describe the phenomenon of interest (such representation is usually "abstract", that is, it gives only an approximate picture of the real phenomenon). In Section 6.7 we studied this general scheme using the example of water flowing out of a vessel. Note that the content of Section 6.7 and the examples of differential equations discussed in Section 6.6 are of an illustrative value only; they can be freely replaced with any other examples and there is no need to memorize them (although the concept of a first-order differential equation with variables separable is so important that it is advisable to master it). In this respect it is appropriate to note the difference between Chapter 4 and the last sections of Chapter 6: while the reader must carefully read Chapter 4, Section 6.6 and especially Section 6.7 can just be browsed through, so as to understand the principal points.

Higher mathematics was created by Newton and Leibniz in the 17th century, and even in their works it was a highly developed discipline. Of course, its development did not stop there. Newton and Leibniz did not confine themselves to the basic definitions and initial theorems. Their contribution was much greater. Newton clearly realized the importance of *differential equations* in analyzing natural phenomena. Take his famous "Principia" (1687), which contains Newtonian mechanics (the Newtonian world system), and you will see it starts with the differential equation of motion (see Eq. (9.4.2)). This equation is taken as an axiom, whereas all subsequent propositions of mechanics are theorems derived from this axiom (and also from the law of gravity that follows from experimental findings (Kepler's laws) and axiom (9.4.2)). In his mathematical investigations Newton formulated "the main problems of mathematical analysis" as follows:

(1) from a given relationship between fluents (initial functions) to determine the relation between fluxions (derivatives);

(2) from a given equation containing fluxions to find the relationship between fluents.

The first of these problems is obviously a problem of the differentiation of known combinations of functions: thus, in Newton's notation, if $z = uv$, then $\dot{z} = \dot{u}v + u\dot{v}$, where u, v, and z are fluents and $\dot{u}$, $\dot{v}$, and $\dot{z}$ are fluxions. The second problem is a problem of the solution of differential equations.

Newton also devoted much attention to infinite *series* (discovery of the binomial series made a great impression on him[6.17]). The whole of Newton's theory of fluxions and fluents (derivatives and integrals) was obviously the fruit of his investigations into the theory of infinite series. Nor is it accidental that the basic theorems (see formulas (6.1.18) and (6.1.19)) of the theory of series are linked with the names of Newton's pupils and associates: Brook ***Taylor*** (1685-1731), who was Secretary to the Royal Society of London (the British academy of sciences) when Newton was its President, and Professor Colin ***Maclaurin*** (1698-1746) of Edinburgh University, who knew Newton personally and greatly admired him. Incidentally, both formulas, (6.1.18) and (6.1.19), were first recorded by Taylor. Maclaurin's contribution was that he posed the question of the sphere of applicability of the formulas and partially answered it. The exact formula (6.1.16), the one containing a finite number of terms, was established by Lagrange (we will return to him again later on).

Leibniz, too, made important contributions to the theory of series and to differential equations in the solution of various geometrical problems. Note that Leibniz was more interested in geometry than in mechanics: he associated the concept of the derivative with a tangent and not speed. Significantly, his fundamental memoir on the "new mathematics" had an involved title: "A New Method of Maxima and Minima, and Also of Tangents, for Which Method Neither Fractions Nor Irrational Quantities are an Obstacle, and a Special Kind of Calculus for This".

The excellent school of Leibniz, headed by the two brothers, ***Jakob Bernoulli*** (1654-1705) and ***Johann Bernoulli*** (1667-1748), made an outstanding contribution to the differential and integral calculus. They were the founders of the famous Swiss family of mathematicians, which in the 17th and 18th centuries produced eight first-class mathematicians. (The most prominent of the later members of the family was Johann's son ***Daniel Bernoulli*** (1700-1782), who worked in Basel and St. Petersburg and was one of the founders of hydrodynamics and the kinetic theory of gases. We will have occasion again to discuss his work.)

[6.17] For an account of how this outstanding discovery was evidently made, see, for example, G. Polya, *Mathematical Discovery (On Understanding, Learning and Teaching Problem Solving)*, Vol. 1, Wiley, New York, 1962, pp. 91-93.



Jakob Bernoulli



Johann Bernoulli

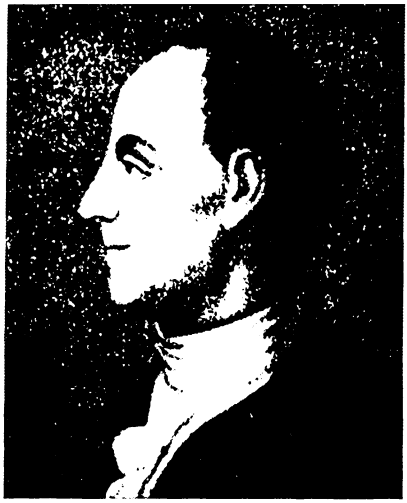
Jakob Bernoulli received a theological education, but his interest in mathematics prevailed. He studied the mathematical literature by himself and came upon the works of Leibniz, which made a profound impression on him. In fact, they so overwhelmed him that he gave up the secure life of a pastor and for a number of years held the little-respected and low-paid post of a home tutor. It was only thanks to the good offices of Leibniz that in 1683 he was appointed a professor (at first, of physics, and later of mathematics, too) at the University of Basel.

Johann Bernoulli's father wanted him to go into commerce, but thanks to his elder brother, who had taught him mathematics and physics, his clearly expressed scientific interests ran counter to his father's wishes. Since Switzerland in those years provided very little opportunity for the study of mathematics, Johann Bernoulli was compelled to obtain a medical education and for many years earned his living as a physician. His doctoral dissertation on the movement of muscles was interesting in that it was evidently the first attempt to apply the methods of higher mathematics to physiology. On the recommendation of the famous Christian Huygens, Johann was appointed to a professorship in physics at Groningen (the Netherlands) in 1695. Returning to Basel in 1705, he at first could obtain a university post only as a professor of Greek. It was only after the death of his brother Jakob that Johann Bernoulli was appointed professor of mathematics at the University of Basel and held this post to the end of his life.

The Bernoulli brothers maintained a lively correspondence with Leibniz, who time and again expressed his admiration for their successes. It was precisely in the course of this correspondence that mathematical analysis received its present form, basic symbolism, and terminology. For example, originally Leibniz had been inclined to speak of "differential calculus" and "summational calculus" as the two branches of the "new mathematics," but on a suggestion from Johann Bernoulli he finally chose the Latinized term "integral calculus" (instead of "summational"). Jakob and Johann Bernoulli greatly advanced the new calculus, obtaining, in particular, important results in the theory of differential equations (see Section 6.6) and laying the foundations of what is called *calculus of variations*, which is mentioned in passing in Section 7.2 (see also the text above Example 5 in Section 7.1).[6.18]

The first printed textbook of differential and integral calculus, entitled "Infinitesimal Analysis for the Study of Curves" (note how it echoes the famous memoirs of Leibniz) appeared in 1696. Its author was Guillaume F. A. ***L'Hospital*** (1661-1704), a French nobleman (Marquis de St. Mesme) who was a pupil of Leibniz and Johann Bernoulli. This excellent textbook went through many editions and was translated into many languages. The ideas L'Hospital set forth closely follow the lectures which he heard Johann Bernoulli deliver in Paris; they also follow Bernoulli's manuscript manual "Lectures on Differential Calculus," which was discovered in the library of the University of Basel in the 20th century and was published for the first time in 1922. Thus, while Bernoulli's text remained unknown, the lectures which were based on it and were attended by only one student,

[6.18] Variational calculus became an independent scientific discipline in the 18th century thanks to the works of Euler and Lagrange.



Leonard Euler



Jean Le Rond d'Alembert

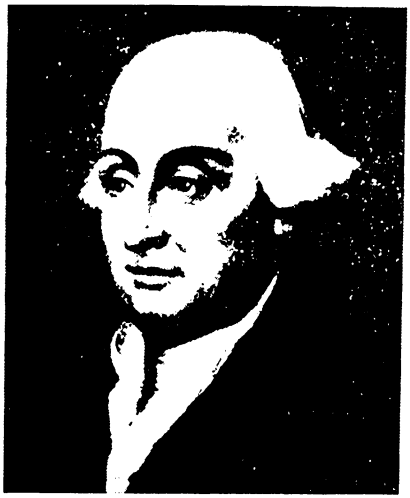
L'Hospital, exerted a tremendous influence on the entire subsequent development of higher mathematics (in particular, it spread the symbols and terminology of Leibniz). Among other things, L'Hospital's textbook saw the first publication of the method of calculating limits which Johann Bernoulli taught his pupil and which is now with so little justification simply called "L'Hospital's rule" (see Section 6.5).

Another of Johann Bernoulli's pupils was the Swiss mathematician Leonhard ***Euler*** (1707-1783). He was introduced to Bernoulli by his father, Pastor Paul Euler, who had once studied mathematics under Jakob Bernoulli. Pastor Euler wanted his son to be a pastor too, but Johann Bernoulli convinced Paul Euler that the boy had outstanding mathematical ability. On Johann Bernoulli's recommendation Leonhard Euler went to St. Petersburg in 1727, where an Academy of Sciences had recently been founded and where two of his teacher's sons (one of them was Daniel Bernoulli) were working. It was originally planned that Euler would take the vacant post of Professor of Physiology at the St. Petersburg Academy, and he made a thorough study of this science in order to follow his teacher's example and apply mathematical methods in it. But when he arrived in St. Petersburg, he found that the post of professor of mathematics was also vacant. So he gave up physiology and devoted himself to mathematics, physics, mechanics, and astronomy (for one, the movements of the moon).

Leonhard Euler spent a large part of his life in St. Petersburg. There was a break of 25 years when, on invitation from King Frederick II of Prussia, he moved from St. Petersburg, in those years a place not very conducive to calm scientific research, to Berlin.[6.19] Here he became head of the physics and mathematics department of the Berlin (Prussian) Academy of Sciences, but he kept up his close contacts with the St. Petersburg Academy. Euler was the most prominent and most productive mathematician of the 18th century. His work dealt with literally every sphere of mathematics and mathematical physics. We write about some of Euler's findings below (see Section 10.8 and Chapters 14 and 15). Euler's multivolume textbook of differential and integral calculus was translated into other languages.

Among Euler's contemporaries was Jean Le Rond ***d'Alembert*** (1717-1783), a distinguished French mathematician and author of treatises on mechanics. His exceptional scientific versatility is particularly evident in his collaboration with Denis ***Diderot*** (1713-1784). They produced the famous "Encyclopédie, ou Dictionaire Raisonné des Sciences, des Arts et des Métiers," in which he wrote nearly all the articles dealing with the natural sciences. D'Alembert was named after the church (Jean Le Rond; literally, Jean the Round) on the steps of which he was found as an abandoned infant. Notwithstanding this handicap he became a prominent man of science. In Section 10.8 we describe a most edifying and fruitful scientific debate in which Leonhard Euler, d'Alembert, and Daniel Bernoulli took part. Both Euler and d'Alembert strove for concrete

[6.19] This refers to the "Biron period" in Russian history, named after E. J. Biron, a disreputable favorite of Empress Anne (1693-1740), whom she brought from Courland. The reign of Empress Anne, from 1730 to 1740, was marked by arbitrary arrests and executions of innocent people and total disorganization of the machinery of government.



Joseph Louis Lagrange



Augustin Louis Cauchy

results, displaying outstanding intuition in mathematics and physics and a lack of interest in scholastic discussions of mathematical ideas. This is illustrated, for example, by d'Alembert's well-known call to the young: "Keep on working—complete understanding will come in time."

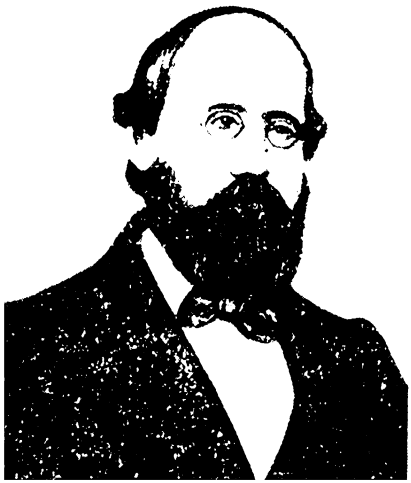
When Euler decided to return to St. Petersburg after having headed the physico-mathematical department of the Berlin Academy from 1741 to 1766, the question of his successor arose. Euler recommended the young mathematician Joseph Louis ***Lagrange*** (1736-1813), who was born into a French family that had moved to Italy. Lagrange was only 30 at the time. His candidature was enthusiastically supported by d'Alembert, who corresponded with King Frederick II. Lagrange was already a well-known scientist. He had begun to teach in the Turin Artillery School in 1726, when he was 20, and the following year he was one of the founding members of the Turin Academy of Sciences, in whose proceedings he published many papers. In 1787, after the death of Frederick II, Lagrange moved to France, where he played an outstanding role in the rise of the Parisian Polytechnical School (École Polytechnique), an institution of higher learning of a new type, which trained research engineers.[6.20] It was during the Parisian period that Lagrange compiled his two-volume "Analytical Mechanics" (1788), which greatly furthered mathematics and physics. Lagrange published an excellent, and in many respects revolutionary, textbook of mathematical analysis ("The Theory of Analytic Functions," 1797) based on lectures that he delivered at the Polytechnical School. He is also known for his fundamental studies in algebra. In the breadth and range of his mathematical interests and in the brilliance of his achievements, Lagrange was second perhaps only to Euler among the mathematicians of the 18th century.

One of the pupils and ardent admirers of Lagrange was Jean Baptiste ***Fourier*** (1768-1830), a professor at the Polytechnical School whose name is linked with outstanding achievements in the field of *partial differential equations* (also called *equations of mathematical physics*; see Section 10.8), for one, in the theory of the propagation of heat and in the theory of trigonometric series (see Section 10.9).

Another professor of the Polytechnical School was the famous Augustin Louis ***Cauchy*** (1789-1857), who created the modern theory of limits (incidentally, his main ideas were borrowed from d'Alembert), which helped to clarify all the concepts of higher mathematics. Cauchy is rightly regarded as the father of the *theory of functions of a complex variable* (see Chapters 14, 15, and 17). He shares this honor, incidentally, with Georg Friedrich Bernhard ***Riemann*** (1826-1866), the German mathematician and one of the most prominent scientists of the 19th century, who made outstanding contributions in literally every field of mathematics and mathematical physics.

While Cauchy was a representative of the Polytechnical School of Paris, Riemann was connected with another educational and scien-

[6.20] The example of the famous Polytechnical School of Gaspard ***Monge*** (1746-1818) and Lagrange was undoubtedly taken into account by the founders of the Moscow Physico-Technical Institute in the Soviet Union and the founders of the Massachusetts and California institutes of technology in the United States.



Georg Riemann

tific center that played a big role in the development of mathematics and physics in the 19th and 20th centuries—the University of Göttingen in Germany. Here he attended lectures by the famous mathematician, physicist, astronomer, and geodesist Karl Friedrich ***Gauss*** (1777-1855), who is usually acknowledged to be the first mathematician of the 19th century. At the University of Göttingen Riemann presented the first-ever course in the theory of functions of a complex variable, as well as a course in the theory of partial differential equations. His lectures in this second course, published after his death, were the first textbook on the equations of mathematical physics.

David ***Hilbert*** (1862-1943) belongs to an altogether different generation of Göttingen scientists. He is regarded as the father of *functional analysis*, which deals with the functions of functions, or functionals (see the text preceding Example 5 in Section 7.1). Today this branch of mathematics includes generalized functions (or distributions), such as Paul Dirac's delta function (see Chapters 16 and 17).

Today the ideas of mathematical analysis continue to undergo intensive development, and computers are giving them a new and powerful impetus.

Chapter 7 Investigation of Functions. Some Problems from Geometry

In Chapters 1 to 6 we introduced the main concepts of higher mathematics, the concepts of a *derivative* and an *integral*, and suggested some techniques for using them. The present chapter, which concludes Part 1, partially develops the material of the previous chapters. For instance, here we will dwell in detail on the methods used to find the maximum and minimum points of a function, of which we spoke briefly in Section 2.6. We will also tell about the basic geometrical applications of differential and integral calculus. Some themes in the chapter can be discussed in extracurricular studies and are written for a reader with an inquiring mind (e.g. see Sections 7.4, 7.6, and 7.7).

Since this chapter was written as an addition to Chapters 1 to 6, it is naturally of a patchy nature. Loosely, the material can be grouped as follows: Sections 7.1 to 7.3, Section 7.5, Sections 7.9 to 7.11, Section 7.12, Sections 7.4 and 7.8, Section 7.6, and Section 7.7.

7.1 Smooth Maxima and Minima

The problem of finding the value of x_0 for which a given function $y = f(x)$ attains a ***maximum*** or ***minimum*** is not solvable in general form by the tools of elementary algebra.

In Chapter 2 we established that at points *where a function has a maximum or a minimum the derivative is equal to zero.* It was also shown there how, using the derivative y', to establish exactly what the function has at the given point x_0 (where $f'(x_0) = 0$), a maximum, a minimum, or an inflection (neither a maximum nor a minimum). To do this we were forced to compute the values of y' for values of x close to x_0 on the right and on the left of x_0. For instance, we found that the conditions $y'(x) > 0$ for $x < x_0$, $y' = 0$ at $x = x_0$, and $y'(x) < 0$ for $x > x_0$ imply that at point x_0 the function $y = y(x)$ attains a maximum. In Chapter 2 we also investigated another method for finding the maxima or minima of a function, a method that invokes the second derivative y'' (only its value at point $x = x_0$ was required, however), namely, we found that if, say, $f'(x_0) = 0$ and $f''(x_0) < 0$, then at point $x = x_0$ the function $f(x)$ has a maximum.

The same conclusions can be drawn from the considerations involving the idea of ***convexity*** of the curve representing a function; we touched on this concept in Section 2.7 and will return to it in Section 7.4. Indeed, the condition $f'(x_0) = 0$ means that the tangent to the graph of the function at point $x = x_0$ is *horizontal.* From the fact that $f''(x_0) < 0$ it follows that point $x = x_0$ is a ***point of convexity***, that is, the curve representing $y = f(x)$ close to $x = x_0$ lies *under* the tangent (see Figure 2.7.1). But these two facts simply mean that the function $f(x)$ has a maximum at point $x = x_0$. Similar reasoning shows that if $f'(x_1) = 0$ and $f''(x_1) > 0$, at point $x = x_1$ the function $f(x)$ has a *minimum.*

These conclusions are also obtained by considering *Taylor's series*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \ldots \qquad (7.1.1)$$

for the function $f(x)$ at point $x = x_0$. If $f'(x_0) \neq 0$, then for x close to x_0 the quantities $(x - x_0)^2$, $(x - x_0)^3$, etc. may be neglected when compared with $x - x_0$. We therefore arrive at the approximate expression

$$f(x) \simeq f(x_0) + f'(x_0)(x - x_0),$$

or $f(x) - f(x_0) \simeq f'(x_0)(x - x_0)$.

From this we see that if, say, $f'(x_0) > 0$, then $f(x) - f(x_0) > 0$ (i.e. $f(x) > f(x_0)$) for $x > x_0$ and $f(x) - f(x_0) < 0$ (i.e. $f(x) < f(x_0)$) for $x < x_0$. This means that at $x = x_0$ the function has neither a maximum nor a minimum. Similarly, there is no maximum or minimum when $f'(x_0) < 0$, only in the first case the function is increasing at point x_0, while in the second it is decreasing.

But if $f'(x_0) = 0$, then the term with $(x - x_0)$ in (7.1.1) vanishes and we cannot ignore the term containing $(x - x_0)^2$. Ignoring terms with $(x - x_0)^3$, $(x - x_0)^4$, etc., small as compared with the term with $(x - x_0)^2$, we get, from (7.1.1),

$$f(x) \simeq f(x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 \quad (7.1.2)$$

(the approximate formula (7.1.2) is the more exact the smaller $|x - x_0|$ is). From this we see that if $f''(x_0) > 0$, then $f(x) > f(x_0)$, irrespective of whether $x < x_0$ or $x > x_0$, that is, $f(x_0)$ is *less* than any adjacent value of $f(x)$ and therefore $f(x_0)$ is a *minimum* of the function. If $f''(x_0) < 0$, then $f(x) < f(x_0)$ and $f(x_0)$ is a *maximum* of the function.

It may, however, happen that both $f'(x_0) = 0$ and $f''(x_0) = 0$. We then have to take the next terms in Taylor's series (7.1.1). If $f'''(x_0) \neq 0$, then we cannot neglect the term with $(x - x_0)^3$, but we can freely ignore the terms with $(x - x_0)^4$, $(x - x_0)^5$, etc., which are small if compared with the term with $(x - x_0)^3$. Then we have

$$f(x) \simeq f(x_0) + \frac{1}{6} f'''(x_0)(x - x_0)^3, \quad (7.1.2a)$$

and we have neither a maximum nor a minimum, just as we had neither a maximum nor a minimum at $x = 0$ for the function $y = x^3$ (see Figure 1.5.1; for this function $y'(0) = y''(0) = 0$). But if $f'(x_0) = f''(x_0) = f'''(x_0) = 0$ and $f^{IV}(x_0) \neq 0$, then in the vicinity of point x_0 we have

$$f(x) \simeq f(x_0) + \frac{1}{24} f^{IV}(x_0)(x - x_0)^4. \quad (7.1.2b)$$

Here the sign of $f(x) - f(x_0)$ is the same for $x > x_0$ and for $x < x_0$; it is determined by the sign of $f^{IV}(x_0)$. If $f^{IV}(x_0)$ is positive, we have a *minimum* (just as in the case of the function $y = x^4$; see Figure 1.5.1), and if $f^{IV}(x_0)$ is negative, we have a *maximum*.

The attentive reader has probably already guessed that if for $x = x_0$ the first nonzero derivative is of *odd* order (first, third, fifth, etc.), then the function has neither a maximum nor a minimum at this point, while if the first nonzero derivative is of *even* order (second, fourth, etc.), the function has either a maximum or a minimum depending on the sign of that derivative (a minimum if the derivative is positive and a maximum if it is negative). Here, as everywhere in this section, we consider only the maxima and minima that a function attains at points where it is represented by a *smooth curve*, that is, at points where the function *has a derivative* (or even several of them, $y'(x_0)$, $y''(x_0)$, etc.). Such maxima and minima are sometimes called *smooth*; other cases related to maxima and minima that are not smooth will be considered in Section 7.2.

Let us consider some examples.

Example 1. Suppose that

$$\text{(a) } y = A + B(x - a)^3, \quad \text{(b) } y = A + B(x - a)^4, \quad (7.1.3)$$

where $B \neq 0$. Find the maxima and minima of y.

The derivatives of (7.1.3) are

(a) $y' = 3B(x - a)^2$, $y'' = 6B(x - a)$, with $y'(a) = y''(a) = 0$, and $y''' = 6B \neq 0$.

(b) $y' = 4B(x - a)^3$, $y'' = 12B(x - a)^2$, $y''' = 24B(x - a)$, with $y'(a) = y''(a) = y'''(a) = 0$, and $y^{IV} = 24B \neq 0$.

In the case (a) the first nonzero derivative is the derivative of the third (*odd*) order, whereby here at point $x = a$ the function has neither a maximum nor a minimum (the function has in general no maxima or minima since $y'(x) = 0$ only at $x = a$), but it has an inflection point at $x = a$. In the case (b) the only nonzero derivative at point $x = a$ is that of fourth (*even*) order and the function has a minimum for $B > 0$ ($y^{IV}(a) > 0$) and a maximum for $B < 0$ ($y^{IV} < 0$), which also follows from the formula for the function. (Why?)

Example 2. It is required to construct an open-at-the-top box of maximum volume using a square sheet of tin with side $2a$ by cutting out equal

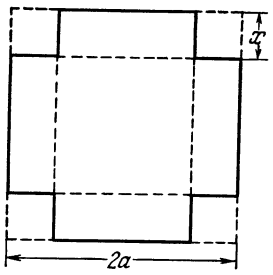


Figure 7.1.1

squares at all corners of the sheet and then bending the tin to form the sides of the box (Figure 7.1.1). What is the length of the side of the squares to be cut out?

Let the length of the side of the cut-out squares be x. The volume of the box will depend on what kind of square we cut out and therefore it is natural to denote it by $V(x)$. Let us compute this volume:

$$V(x) = (2a - 2x)^2 x = 4(a - x)^2 x, \qquad (7.1.4)$$

which is true because the base of the rectangular box is a square with side $2a - 2x$, while the height of the box is x.

Now find the derivative of (7.1.4):

$$V'(x) = -8(a - x)x + 4(a - x)^2.$$

Solve the equation $V'(x) = 0$:

$$-8(a - x)x + 4(a - x)^2 = 0,$$

or $(a - x)(a - 3x) = 0$,

whence $x_1 = a$ and $x_2 = a/3$.

We note at once that the value $x_1 = a$ is of no interest to us because then we wouldn't have a box by cutting the sheet in that fashion. There remains $x = a/3$. Then

$$V\left(\frac{a}{3}\right) = 4\frac{4a^2}{9}\frac{a}{3} = \frac{16a^3}{27} \simeq 0.593a^3.$$

Here $V'(a/3) = 0$, and since

$$V''(x) = 8x - 8(a - x) - 8(a - x) = 24x - 16a,$$

we have $V''(a/3) = -8a < 0$. Consequently, the function $V(x)$ has a maximum at $x = a/3$.

To summarize, the maximum value is obtained for $x = a/3$, that is, we have to cut out squares whose sides are 1/6 the side of the original square.

Let us compute $V(x)$ for several x close to $a/3$ and tabulate the results:

x	$0.25a$	$0.30a$	$0.33a$	$0.40a$	$0.45a$
$V(x)$	$0.562a^3$	$0.588a^3$	$0.592a^3$	$0.576a^3$	$0.540a^3$

We see that small variations in the value of x near $x = a/3$, that is, near the value of x to which corresponds the maximum of the function, bring about very small changes in V, which means that the function near the maximum varies very slowly.

This is also evident from Taylor's formula (7.1.1). If $f'(x) = 0$ at the point of maximum or minimum of the function, (7.1.1) takes the form

$$f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{6}f'''(x_0)(x - x_0)^3 + \ldots. \qquad (7.1.5)$$

This series does not contain a term with $x - x_0$: the smallest power of $x - x_0$ on the right-hand side of (7.1.5) is $(x - x_0)^2$, which is extremely small for x close to x_0. In our example, a change in x by 9% (from $0.33a$ to $0.30a$) causes a change in V by less than 1%, while a change in x by 24% (from $0.33a$ to $0.25a$) causes a change in V by 5%. Therefore, if we are interested in the maximal value of the function and if we make a small error in finding x_0 from the equation $V'(x) = 0$ (for example, if we solved this equation in an approximate fashion), then this has but slight effect on the maximal value of the function.

Note also that point $x = a/3$ in the problem we are discussing here is precisely the *maximum point*. Common sense leads to such a conclusion. The function that expresses the volume of

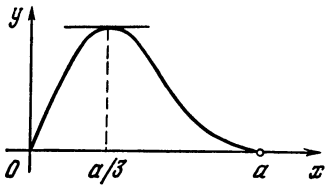


Figure 7.1.2

the box is, obviously, positive for all values of x between 0 and a; at the "endpoints" $x = 0$ and $x = a$ it vanishes (in the first case the box has zero height and in the second it has a zero base area). This implies that as x is changed from 0 to a, the quantity $V(x)$ first grows and then diminishes. This means that somewhere in the middle the function becomes maximal, and the only "suspect" is the point $x = a/3$ (since only at this point does $V'(x)$ vanish), which means that this point is the maximum point. To make this line of reasoning more graphic, we depict the curve $V = V(x)$ in Figure 7.1.2 (at point $x = a$ the graph touches the axis of abscissas since $V'(a) = 0$).

Example 3. Suppose we have an electric circuit that consists of an emf source (the emf equal to u; compare with Chapter 13), a resistor (the resistance of which is r, including the internal resistance of the emf source), and a variable resistor (resistance R) (Figure 7.1.3). What will be the value of R if the power output of this resistor, W, is maximal?

Since the total resistance of the circuit is $r + R$, the current j flowing through the circuit is, by Ohm's law, $j = u/(r + R)$. The power output $W = ju_R$, where u_R is the voltage across R, which by Ohm's law is equal to jR. Substituting this value of u_R into the

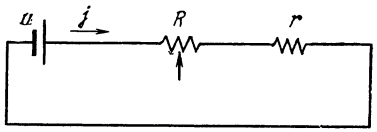


Figure 7.1.3

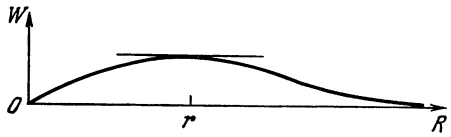


Figure 7.1.4

expression for the power output W, we get

$$W = j^2R = \frac{u^2R}{(R+r)^2}. \qquad (7.1.6)$$

We must find the maximum of (7.1.6) with R variable. We find the derivative dW/dR:

$$\frac{dW}{dR} = \frac{u^2}{(r+R)^2} - \frac{2u^2R}{(r+R)^3} = \frac{u^2(r-R)}{(r+R)^3}. \qquad (7.1.7)$$

This yields $dW/dR = 0$ at $R = r$.

It can be easily seen that $R = r$ corresponds to precisely a *maximum* of the function $W(R)$; this follows if only from the fact that for $R < r$ the derivative (7.1.7) is positive and for $R > r$ it is negative. Common sense also leads to the same result: for small R the energy output W is low because the resistance is low (in this case almost the entire energy output "resides" in r, while the presence of R changes almost nothing; if $R = 0$, then $W = 0$), while for R large, then, in view of Ohm's law, the current j will be low, which means W will be low, too. Somewhere between small R and large R lies the value of R for which the power output W is maximal, and this can happen only at $R = r$ (see the graph of $W = W(R)$ in Figure 7.1.4).

Example 4. Using available boards, we can build a fence of length l. How can we fence off a rectangular yard of maximal area using for one side the wall of an adjacent building (Figure 7.1.5)?

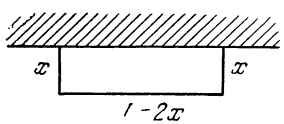


Figure 7.1.5

Let two sides have length x. Then the third side is $l - 2x$. The area of the yard is

$$S(x) = (l - 2x)\,x = -2x^2 + lx, \tag{7.1.8}$$

whence $S'(x) = -4x + l$. Clearly, $S'(x) = 0$ only at $x = l/4$; here $S''(x) = -4 < 0$, so that at $x = l/4$ the function $S(x)$ has a *maximum*.

This result can also be obtained without resorting to higher mathematics. Indeed, suppose we have to seek the maximum (or minimum) of a second-degree polynomial (7.1.8) of variable x. But for any second-degree polynomial

$$y = ax^2 + bx + c \tag{7.1.9}$$

we have

$$\begin{aligned} y &= a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) \\ &= a\left[x^2 + 2\frac{b}{2a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right] \\ &= a\left[\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a^2}\right] \\ &= a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a}. \end{aligned} \tag{7.1.9a}$$

Since $(x + b/2a)^2$ is nonnegative for all x's equality occurring only at $x = -b/2a$, we find that y has a *maximum* if a is negative and this maximum is at $x = -b/2a$, and y has a *minimum* if a is positive and this minimum is at $x = -b/2a$. In particular, the function (7.1.8), with $a = -2$, $b = l$, and $c = 0$, attains its maximum at $x = l/4$.

The example we have just considered is one of the variants of ***Dido's problem***. As legend has it, the mythical Elissa Dido, the daughter of the Tyrian king Mutton, after her husband was slain by her brother, fled to Cyprus, and thence to the coast of Africa, where she purchased from a local chieftain, Iarbas, a piece of land on which she built Carthage. Iarbas agreed to sell a piece of land on the mocking condition that it be no larger than the area covered by an oxhide. But the cunning Dido did not cover the tiny piece of land by the oxhide, as Iarbas expected she should; instead, she cut the oxhide into thin strips and "fenced" off a large piece of land, which was made even larger by using the coastline. If we assume the coastline to form a straight line and require that the fenced-off area be a rectangle, then the problem reduces to the one we have just solved. However, the situation complicates if we assume that the boundary of the area, the long strip made out of the oxhide, can have an arbitrary shape. In this case the area whose maximum we are seeking depends not on one variable x but on an arbitrary function that specifies the shape of the boundary. It is possible in this case to employ the methods of higher mathematics, too, by introducing an analog of the concept of a differential and by proving that the points of maxima and minima of the "function of functions" considered (in mathematics such objects are known as ***functionals***) correspond to the zeros of the "generalized" differential. With the aid of such methods (which cannot be discussed in a book as elementary as this) it can be proved that the general solution to Dido's problem is a semicircle (Figure 7.1.6).

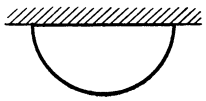


Figure 7.1.6

Example 5. A hiker walking from tent A to camp-fire B wants to gather water in river A_1B_1. How can he do this by traveling the shortest possible distance (Figure 7.1.7)?

We have $AA_1 = a$, $BB_1 = b$, and $A_1B_1 = c$; the values of a, b, and c are given and $AA_1 \parallel BB_1 \perp A_1B_1$. Let the broken line AMB be the path taken by the hiker. Our aim is to find out for what position of point M on the line A_1B_1 is this path the shortest. To determine the position of M it suffices to specify the distance from M to point A_1, the foot of the perpendicular dropped from A onto the straight line representing the river. Denote this distance A_1M by x. Then

$$AM = \sqrt{a^2 + x^2}, \quad MB = \sqrt{b^2 + (c - x)^2},$$

and the distance $S(x)$ traveled by the

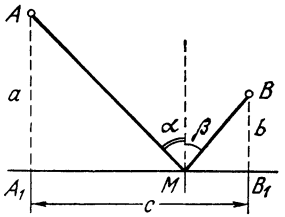


Figure 7.1.7

hiker will be

$$S(x) = AM + MB = \sqrt{a^2 + x^2} + \sqrt{b^2 + (c-x)^2}. \quad (7.1.10)$$

This yields

$$S'(x) = \frac{x}{\sqrt{a^2+x^2}} - \frac{c-x}{\sqrt{b^2+(c-x)^2}}.$$

Equating $S'(x)$ to zero yields

$$\frac{x}{\sqrt{a^2+x^2}} = \frac{c-x}{\sqrt{b^2+(c-x)^2}}. \quad (7.1.11)$$

It is easy to solve this equation. Squaring both sides, we get

$$\frac{x^2}{a^2+x^2} = \frac{(c-x)^2}{b^2+(c-x)^2},$$

or $x^2b^2 + x^2(c-x^2) = a^2(c-x)^2 + x^2(c-x)^2$,

that is,

$$x^2b^2 = a^2(c-x)^2, \quad \frac{x^2}{(c-x)^2} = \frac{a^2}{b^2},$$

whence

$$\frac{x}{c-x} = \pm\frac{a}{b}, \text{ or } x_1 = \frac{ac}{a+b},$$

$$x_2 = \frac{ac}{a-b}.$$

Substituting the values of x_1 and x_2 into the original equation (7.1.11), we see that the second root does not satisfy the equation. This is an extraneous root generated by the squaring process.[7.1] Thus, $x = ac/(a+b)$.

It is possible, however, to give a pictorial geometrical representation that will enable us to obtain the answer without solving the equation. Rewrite the condition (7.1.11) thus:

$$\frac{A_1M}{AM} = \frac{MB_1}{MB}. \quad (7.1.11a)$$

But $A_1M/AM = \cos \angle A_1MA = \sin \alpha$. Similarly $MB_1/MB = \cos \angle B_1MB = \sin \beta$. The condition (7.1.11a) means that $\sin \alpha = \sin \beta$, or, since both α and β are acute,

$$\alpha = \beta. \quad (7.1.12)$$

Thus, the hiker must take the path of a ray of light bounced off a mirror (*the angle of incidence is equal to the angle of reflection*).

For a complete solution to the problem it remains to demonstrate that for such a position of point M the distance is indeed *minimal* (and not maximal). This can be done by computing the second derivative of (7.1.10).

But it is also possible to reason differently. From the expression (7.1.10) for $S(x)$ we see that $S(x)$ is positive for all x's and increases without bound together with the absolute value of x, irrespective of whether $x > 0$ or $x < 0$. And since $S'(x)$ vanishes only for one value of x, it is clear that at this value of x the function $S(x)$ has a *minimum*. In general, if in the interval we are interested in the first derivative of a function has only one root, obvious considerations often permit dispensing with a formal investigation by means of the second derivative (see what was said in this connection in Example 2).

The problem in Example 5 can be solved in a purely geometrical manner without resorting to methods of higher mathematics. Referring to Figure 7.1.8, extend the segment AA_1 to A' ($A_1A' = AA_1$) and join A' with B. Then $AM = A'M$ since $\Delta AA_1M = \Delta A_1A'M$. Therefore $A'B = A'M + MB = AM + MB$. For any other point D on the segment A_1B_1 we will have $AD + DB = A'D + DB$ and $A'D + DB > A'B$, since a polygonal line is longer than any segment of

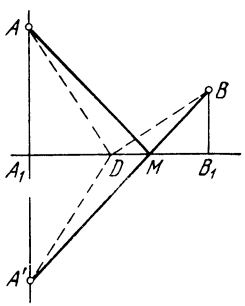


Figure 7.1.8

[7.1] The extraneous root x_2 of Eq. (7.1.11) can be dropped immediately because $x/(c-x)$ is positive and, hence, is not $-a/b$.

a straight line. Consequently, the desired point M is the point of intersection of the straight lines $A'B$ and A_1B_1, whence follows (7.1.12). (Why?)

Examples 4 and 5 show that certain problems involving the finding of maxima and minima may be solved by the tools of elementary mathematics, without resorting to differential calculus (see the text to the solutions to the corresponding problems printed in small type). But, first, not all problems can be tackled without appealing to higher mathematics and, second, the solution by elementary means requires a good deal of ingenuity, whereas higher mathematics offers a standard method of solution of such problems.

Do not get the idea that higher mathematics does not require ingenuity! It will now be used for still harder problems.

Let us now turn to the question of maxima and minima of functions of *two* variables.

The definition $dy = y'dx$ of the differential of a function $y = f(x)$ (see Section 4.1) shows that the points of (smooth) maxima or minima of the function are characterized by the fact that the differential vanishes at these points. This is closely related to the geometrical meaning of the differential (see Figure 4.1.2) and the fact that the tangent to the curve representing the function at the point where the function is at its maximum or minimum must be horizontal. It is equally easy to see that at the point of a maximum or a minimum of a function $z = F(x, y)$ of two variables, the differential of that function, $dz = \frac{\partial F}{\partial x}\,dx + \frac{\partial F}{\partial y}\,dy$ must vanish; in other words, $\partial F/\partial x = \partial F/\partial y = 0$. This is due to the fact that the plane tangent to the surface specified by the function $z = F(x, y)$ is horizontal at the point of maximum or minimum (Figure 7.1.9*a*),[7.2] as well as to the fact that at the point (x_0, y_0) where the function $z = F(x, y)$ attains a maximum the functions $f(x) = F(x, y_0)$ and $g(y) = F(x_0, y)$ of one variable (x in the first case and y in the second) attain a maximum, too, that is, $f'(x)$ $(=\partial F/\partial x)$ and $g'(y)$ $(=\partial F/\partial y)$ must vanish at this point. However, one must bear in mind that while the fact that $f'(x) = 0$ almost always means that f has a maximum or a minimum (since cases where the second derivative vanishes are exceptional), the conditions $\partial F(x_0, y_0)/\partial x = \partial F(x_0, y_0)/\partial y = 0$ (or $dz = 0$) might mean that one of the abovenoted functions of one variable, say $f(x)$, attains a maximum at (x_0, y_0) and the other, $g(y)$, a minimum. In this case the point (x_0, y_0) constitutes a saddle point of the surface $z = F(x, y)$ and at this point F attains neither a maximum nor a minimum (Figure 7.1.9*b*).

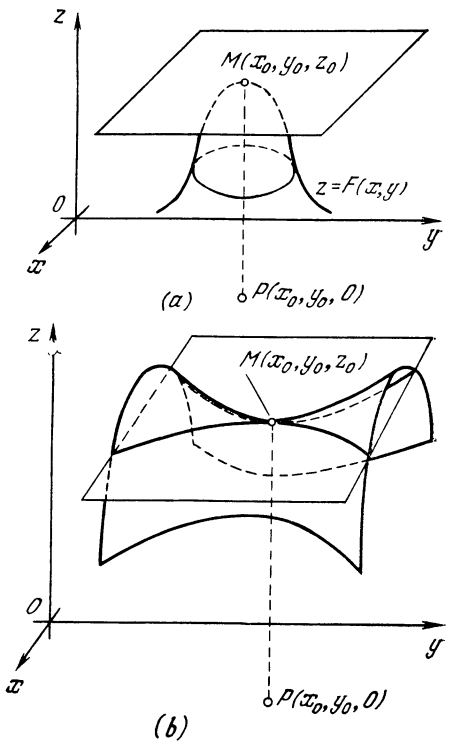


Figure 7.1.9

Exercises

7.1.1. We want to build a box out of a rectangular sheet of tin of sides a and b cutting out equal squares at the corners. What must the side of a square be so that the box is of maximum volume?

7.1.2. Inscribe in an acute-angled triangle with base a and altitude h a rectangle of the largest possible area, two vertices of which lie on the base of the triangle and the other two on the sides of the triangle.

7.1.3. Determine the greatest possible area of a rectangle inscribed in a circle of radius R.

7.1.4. For what radius of the base and for what altitude will a closed cylindrical can of a given volume V have a minimum total surface area?

7.1.5. Two bodies are moving along the sides of a right angle with constant speeds v_1 and v_2 (meters per second) in the direction of the vertex, from which, at the beginning, the first was at distance a meters and the second, b

[7.2] *The plane that is tangent to surface* Π (represented by the equation $z = F(x, y)$) at point $M = M(x_0, y_0)$ contains the tangent lines to the curves $y = y_0$, $z = F(x, y_0) = f(x)$ and $x = x_0$, $z = F(x_0, y) = g(y)$ (to the curves $y =$ constant and $x =$ constant); of all the planes that pass through M, the tangent plane lies closest to Π.

meters. How many seconds after they started will the distance between the bodies be at a minimum?

7.1.6. Prove that the product of two positive numbers whose sum is a constant is the greatest when the factors are equal.

7.1.7. A straight line l divides a plane into two parts (medium I and medium II). A body moves in medium I with a speed v_1 and in medium II at a speed v_2. What path must the point take so as to get, in minimum time, from a given point A of medium I to a given point B of medium II? (This problem poses the question of the *law of light refraction* when light passes from medium I to medium II, and the speed of light in these media is different; it is known that the path the light takes between points A and B is always such that the time taken is a minimum (***Fermat's principle***; see footnote 3.21)).

7.2 Other Types of Maxima and Minima. Salient Points and Discontinuities. The Left and Right Derivatives of a Function

Up to now we have said that maxima and minima of a function occur at values of x for which the first derivative vanishes. However, maxima and minima can also arise for values of the argument that *do not make* the first derivative vanish.

We revert to Example 3 of Section 7.1, where an emf source, resistance r, and variable resistance R are connected in series (Figure 7.1.3). For what value of R will the power output W of r be maximal?

In a manner quite similar to that of Example 3, we get, from (7.1.6)

$$W = j^2 r = \frac{u^2 r}{(r+R)^2}. \qquad (7.2.1)$$

The reader will recall that u and r are assumed known and R variable and unknown. Equation (7.2.1) implies that

$$\frac{dW}{dR} = -\frac{2u^2 r}{(r+R)^3},$$

with the result that the condition $dW/dR = 0$ becomes

$$\frac{u^2 r}{(r+R)^3} = 0.$$

which cannot be satisfied for a *single* R.

Does this mean that the power can increase without bound and that the problem of maximum power does not have a solution? From the physical essence of the problem it follows that the power output will be maximal at $R = 0$ (in this case $W = u^2/r$). Why did we not get the value $R = 0$ from the equation $dW/dR = 0$?

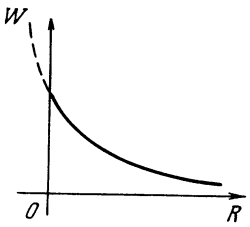


Figure 7.2.1

To see why, consider the graph of $W = W(R)$ (Figure 7.2.1). It is evident from the graph that if R could assume negative values, then for $R = 0$ there would be no maximum. But from physical considerations it follows that negative R have no meaning—every physical problem presupposes that R is nonnegative. Thus, W has a maximum at $R = 0$ because the range of the independent variable is bounded. This means that if the range of the independent variable (the domain of the function) is bounded, we must take into consideration the *boundary* values of the independent variable when testing for maxima and minima.

Let us touch once more on the example of specific heat capacities discussed in Section 2.6, only now we will take water instead of diamond. The quantity of heat (in joules) that is required to heat 1 kg of water (at atmospheric pressure) from 0 to T °C is given approximately by the formula $Q(T) = 4186.68T + 8373.36 \times 10^{-5}T^2 + 1256 \times 10^{-6}T^3$, which implies that the specific heat capacity of water, $c = c(T)$, at a temperature T is

$$c(T) = \frac{dQ}{dT} = 4186.68 + 16746.72 \times 10^{-5}T + 3768 \times 10^{-6}T^2$$

(see Exercise 2.2.2 and its solution). Clearly, for T positive $c(T)$ increases

with T. Does this mean that $c(T)$ for water may be as high as desired? Of course not, since water can exist in the liquid state only in the temperature interval from 0 to 100 °C, and the specific heat capacity is *maximal* at the right endpoint of the interval and *minimal* at the left endpoint.

Let us now return to the question of the maxima and minima that a function attains at its boundaries. When a maximum (minimum) is attained at an endpoint x_0 of the domain of a function $y = f(x)$, the series

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots$$

may begin with the term with $(x - x_0)$. Therefore, if a maximum (minimum) of the function is obtained when $x = x_0$ and we have departed somewhat from x_0 (an endpoint of the domain of the function), we may err considerably in determining y, since the error will be proportional to $x - x_0$ and not to $(x - x_0)^2$, as was the case in Section 7.1. Hence, even a slight error in determining the value of the independent variable that yields a maximum may lead to a sizable error in the value of the function.

In the above examples the function $f(x)$ existed for $x < x_0$ as well, but the values of the function for $x < x_0$ did not interest us because they were devoid of any physical or geometric meaning. It may happen, however, that $f(x)$ is simply meaningless for certain values of the independent variable. For example, there can be no meaning in an even-degree root, say a square root, of a negative number, with the result that the values of the independent variable that make the radicand vanish are the boundary values. In a test for maxima or minima such values of the independent variable must be considered separately. For instance, if $y = a - \sqrt{b - x}$, then $y' = 1/2\ \sqrt{b - x}$, and although y' does not vanish anywhere, y has a maximum. The maximum is attained at $x = b$; indeed, we see that $y = a$ nullifies the radicand, while x cannot be greater than b because otherwise the radicand would become negative. At $x = b$ we have $y = a$; for other values of x the value of y is obtained through subtraction of a positive number [7.3] $\sqrt{b - x}$ from a and is therefore smaller than a.

A maximum (or minimum) may also occur at interior points where the derivative does not vanish. This is the case when the curve has a *salient point* (*corner*). Such points occur, in particular, when the curve consists of two parts described by different formulas for $x > x_0$ and for $x < x_0$. Here is an instance of a physical problem of this nature.

Suppose a teakettle is being heated on an electric hot plate. Our problem is to determine the instant of time when the teakettle has the greatest amount of heat. For the sake of simplicity, we assume that the efficiency of the hot plate is 100%, which means that all the heat is delivered to the teakettle. We put the teakettle on to heat at time $t = 0$, at which time it had q joules of heat.[7.4] By the definition of the unit amount of heat Q (the joule), the amount of heat released by the hot plate is j^2Rt, where j is the current in amperes, R the resistance in ohms, t the time in seconds. Hence, the amount of heat in the teakettle by time t will be (in joules) $Q = q + j^2Rt$.

At the time $t = t_0$ the water in the teakettle begins to boil, and the amount of heat accumulated by the teakettle will be $Q_0 = q + j^2Rt_0$. When the water boils, it begins to turn into steam and boil away.[7.5] The formation of 1 g of steam requires approximately 2256.7

[7.3] $\sqrt{b - x}$ is understood to be the positive root.

[7.4] For zero we take the thermal energy of water at 0 °C.

[7.5] Formation of steam starts at less than 100 °C, that is, even before boiling starts, but we ignore this fact.

joules of heat. Denoting by dm the quantity of water that boils away in time dt, we get

$$dm \simeq (j^2R/2256.7)\,dt$$
$$\simeq 45 \times 10^{-5} j^2R\,dt.$$

And so in one second a total of $dm/dt \simeq 45 \times 10^{-5} j^2R$ grams of water boil away. This requires

$$\frac{dQ_1}{dt} \simeq 418.68 \frac{dm}{dt}$$
$$\simeq 418.68 \times 45 \times 10^{-5} j^2R$$
$$\simeq 0.1884 j^2R \text{ J/s}$$

of heat because 1 g of water contains approximately 418.68 joules of heat at the temperature of boiling. Therefore, by time $t > t_0$ the amount of heat spent on transforming the boiled-away water into steam will be $Q_1 \simeq 0.1884 j^2R\,(t - t_0)$ joules of heat.

We see that the amount of heat (in joules) in the teakettle is expressed by two different formulas: for $t < t_0$ (i.e. prior to boiling) it is $Q = q + j^2RT$, while for $t > t_0$ (after the water started boiling) it is

$$Q \simeq q + j^2Rt_0 - 0.1884 j^2R\,(t - t_0)$$
$$= q + j^2R\,(1.1884t_0 - 0.1884t).$$

The graph of $Q = Q(t)$ is shown in Figure 7.2.2*a*. It is clear from the drawing that $Q(t)$ has a maximum when $t = t_0$, although the derivative $Q'(t)$ does not vanish at this value of t, that is, the tangent to the graph at point $(t_0, Q(t_0))$ is not horizontal (the function $Q = Q(t)$ has no derivative at $t = t_0$ and the graph does not have a tangent line at this point).

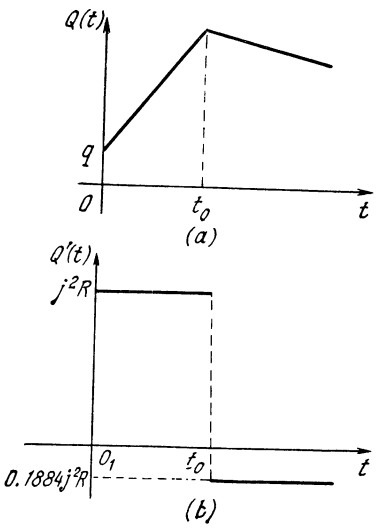


Figure 7.2.2

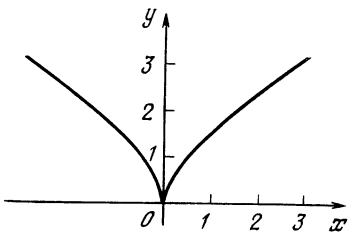


Figure 7.2.3

The derivative of the function $Q = Q(t)$ has a *discontinuity* at $t = t_0$. Indeed, if we consider only values of t less than t_0, we must assume that $Q'(t) = j^2R$, while for $t > t_0$ the derivative is $Q'(t) \simeq -0.1884 j^2R$. The graph of the derivative is shown in Figure 7.2.2*b*.

This example shows that a maximum (or minimum) may occur if the derivative is discontinuous, that is, if the curve representing the function has a *salient point* (*corner*).

From Figure 7.2.3 it is clear that a minimum (or maximum) may occur for those values x_0 of the independent variable at which the derivative has an infinite discontinuity (in the previous example the discontinuity was finite). (We have depicted in Figure 7.2.3 the graph of the function $y = x^{2/3} = \sqrt[3]{x^2}$; the inverse of this function, or $y = x^{3/2}$, or $y^2 = x^3$, is a ***semicubical parabola***; see Section 1.5 and, in particular, Figure 1.5.7.) A point of this type is called a ***cusp***; the function $y = x^{2/3}$ has a minimum at this point. The graph of the derivative of $y = x^{2/3}$ is shown in Figure 7.2.4. Here, as in the case of an ordinary minimum, $y' < 0$ for $x < x_0$ (in Figure 7.2.4, $x_0 = 0$); the function falls off as x approaches x_0 from the left. For $x > x_0$ we have $y' > 0$, and the func-

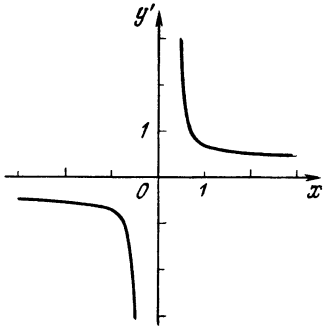


Figure 7.2.4

tion increases with x after the value $x = x_0$ has been passed. But at $x = x_0$ it becomes meaningless to speak of a derivative. The derivative becomes arbitrarily large for x close to x_0 and greater than x_0 and arbitrarily large in absolute value but negative for x close to x_0 and smaller than x_0.

The maxima and minima attained for values of the independent variables when the derivative is discontinuous are called ***cuspidal.*** Cuspidal maxima and minima and the maxima and minima attained at the endpoints (boundary points) of the domain of the function can also be said to be ***nonsmooth.***

In connection with this consideration of singular points on curves, primarily salient points (see Figure 7.2.2*a*), we can make precise our reasoning that led us to the concept of the derivative. In Chapter 2 we considered only *smooth* curves without specially stipulating this fact. The derivative $y'(t)$ taken at the point t is equal to the limit of the ratio

$$\frac{y(t_2)-y(t_1)}{t_2-t_1} \qquad (7.2.4)$$

as t_2 and t_1 tend to t (it is clear then that the difference $t_2 - t_1$ tends to zero). We have specially emphasized that this limit does not depend on how t_2 and t_1 are chosen: they can both be greater than t or both smaller than t or one greater and the other smaller than t or one equal to t and the other greater or smaller than t. Indeed, when we take Δt positive, the fraction $\frac{y(t+\Delta t)-y(t)}{\Delta t}$ corresponds to $t_1 = t$ and $t_2 = t + \Delta t > t$, whereas when we take the fraction $\frac{y(t)-y(t-\Delta t)}{\Delta t}$, we have $t_1 = t - \Delta t < t$ and $t_2 = t$.[7.6]

If $y = y(t)$ is a smooth function, all these fractions yield the same limit, which is equal to the derivative at the given point. The situation changes when we deal with a curve with a salient point (see Figure 7.2.2). If by t_0 we denote the value of t at which the salient point occurs, then, taking $\frac{y(t_0+\Delta t)-y(t_0)}{\Delta t}$, we get, for Δt *positive* and tending to zero, a definite quantity (in the example with the teakettle this quantity is equal to $-0.1884j^2R$) called the ***derivative on the right*** of the function $y(t)$ at point $t = t_0$. On the other hand, taking $\frac{y(t_0)-y(t_0-\Delta t)}{\Delta t}$, we get, for Δt *positive* and tending to zero, another limit (equal in the aforementioned example to j^2R) called the ***derivative on the left*** of the function.

Taking t_2 and t_1 on different sides of t_0, we can obtain *different* values of the ratio (7.2.4) as $t_2 \to t_0$ and as $t_1 \to t_0$; it is easy to see that if A is a corner or a cusp, then the chord BC, where B and C lie on different sides of A, may have different directions (imagine such a chord in Figures 7.2.2*a* and 7.2.3). This means that the limit of the chord, as B and C tend to A, may also be different—it depends on how precisely the points B and C tend to point A. Thus, at a salient point of a function the derivative of that function has no definite value, but the derivatives on the left and on the right can be found unambiguously.

7.6 For *smooth* curves the derivative was calculated as the limit of the ratio $\left[y\left(t+\frac{\Delta t}{2}\right)-y\left(t-\frac{\Delta t}{2}\right)\right]/\Delta t$ as $\Delta t \to 0$ (see the text in small print at the end of Section 6.1); here $t_1 = t - \Delta t/2 < t$ and $t_2 = t + \Delta t/2 > t$.

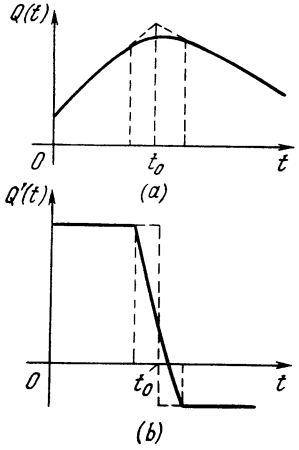


Figure 7.2.5

In Chapter 2, when we first began studying derivatives, we simplified matters by not assuming all the time that the derivative has a definite value irrespective of the mode of approach of Δt to zero (from the left or from the right) only for points at which the curve representing the function is *smooth*. As is evident from Figure 7.2.2*b*, the curve of the derivative $y'(t)$ has a *discontinuity* at the point where the curve $y(t)$ has a *salient point*. Now if we replace the salient point on the curve $y(t)$ by an arc of small radius that is tangent to the curve on the left and on the right (what draftsmen call ***conjugation***), the resulting smooth curve will correspond to a continuously varying derivative; however, on the range of t where the curve $y(t)$ is replaced by the arc the curve $y'(t)$ changes direction sharply (compare Figures 7.2.5 and 7.2.2).

If the curve $y(t)$ has a discontinuity at point t_0 (Figure 7.2.6*a*), then we can say that at t_0 the derivative $y'(t)$ is *infinite* (although actually the function has no derivative at this point). Indeed, if the discontinuity is replaced by a smooth variation of y from y_1 to y_2 over a small interval from $t_0 - \varepsilon$ to $t_0 + \varepsilon$, then on this interval the derivative is equal to $(y_2 - y_1)/2\varepsilon$, which is a very large quantity, increasing as ε decreases (Figure 7.2.6*b*).

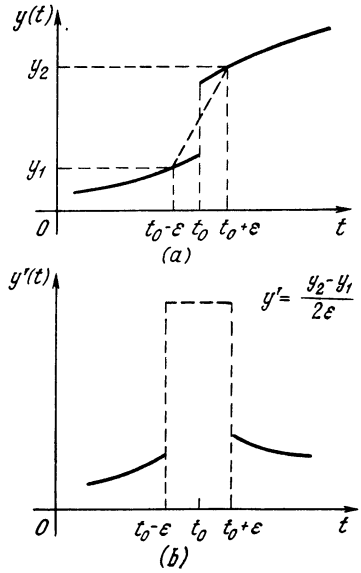


Figure 7.2.6

Now how does the *integral* $\int_a^b y(t)\,dt$ behave when the function $y(t)$ is not smooth? If the function has a salient point, then no new problems arise when we compute the area bounded by the curve $y(t)$. In Section 3.2 we split up the definite integral (the area under the curve) into rectangular strips with an area $y(t_n)(t_{n+1} - t_n)$ or $y(t_{n+1}) \times (t_{n+1} - t_n)$. In the limit, as the intervals, that is, the differences $t_{n+1} - t_n$, get smaller and smaller, it makes no difference whether one takes $y(t_n)$ or $y(t_{n+1})$ either in the case of a smooth curve or in the case of a curve with salient points.

If a curve $y(t)$ is discontinuous at a point $t = t_0$ but remains bounded, then for the interval that contains the discontinuity $(t_n < t_0 < t_{n+1})$ the quantities $y(t_n)$ and $y(t_{n+1})$ remain distinct no matter how t_n and t_{n+1} approach one another. To summarize, then, in the expression of the integral as a sum, the value of one of the summands in this case to a great extent depends on how the sum is taken: by formula (3.2.1) or by formula (3.2.2). However, as $t_{n+1} - t_n$ tends to zero,

the summand itself tends to zero, whereby the limit of the sum, the integral, has a definite value (independent of the way in which the sum was computed) also in the case where the integrand has a discontinuity in the domain of integration. Here, of course, we assume that the values of the function to the left and to the right of the discontinuity and, hence, the size of the discontinuity are finite, the function does not behave as $y = 1/x$ does in the neighborhood of point $x = 0$.

The relationship between the integral and the derivative is likewise preserved. Referring to Figure 7.2.2, let us take the function $Q'(t)$ (whose graph is shown in Figure 7.2.2*b*) and denote it by $f(t)$. Then the function $Q(t)$, whose graph is shown in Figure 7.2.2*a*, is the integral $Q(t) = \int f(t)\,dt$. This example shows that a discontinuity in the integrand function $f(t) = Q'(t)$ leads to a salient point in the integral $Q(t)$ of this function. The definite integral of a function with a discontinuity of this type can be found with the aid of the indefinite integral by the general rule

$$\int_a^b f(t)\,dt = Q(b) - Q(a).$$

We may continue: consider Figure 7.2.6. We can say that for a function tending to infinity on an interval tending to zero (Figure 7.2.6*b*),[7.7] the integral is a *discontinuous* function (Figure 7.2.6*a*). However, in this case we must make precise the law by which the function tends to infinity and the interval to zero. We will not dwell on that here. Examples of this kind (a fuller consideration of which requires additional refinement) lead to the concept of the delta function (see Chapter 16).

Thus, the final scheme for solving problems that involve finding the maxima and minima of a function goes as follows. First we must find the derivative of the function under investigation and determine the stationary points, that is, points at which the variation of the function is the smallest—points at which the derivative vanishes. But in addition to these points, "suspects" (i.e. points at which the function may be maximal or minimal) are the boundary points of the domain of the function, salient points, and points at which the derivative becomes infinite and changes its sign ("cusps"). We must then check how the function changes at all these points. If at a point where $f'(x) = 0$ the graph is represented by a smooth curve, information on the type of the point (maximum, minimum, inflection) can be obtained through the second derivative of the function (or higher-order derivatives if the first and second derivatives vanish at this point simultaneously). In other cases an idea about the behavior of the function in a point being considered may be given by the one-sided (i.e. left or right) derivatives: it is clear that at the boundary points of the domain of a function only one-sided derivatives can exist. Finally, if we are interested in the *absolute maximum* and/or *minimum* of a function over the range of variation of the function, that is, the largest of all the (local) maxima or the smallest of all the (local) minima, we must compare the values of the function at all points of local maxima (or minima).

Exercises

7.2.1. Find the smallest value of the function $y = x^2 - 2x + 3$ as x varies from 2 to 10.

7.2.2. A fisherman F is sitting on the bank of a river and fishing. Going down the river, at a distance h from the place where F is sitting and with speed v, there is a steamer S of length l, and at time $t = t_0$ the bow of the steamer is positioned exactly opposite F. (We ignore the fact that S has width and assume that S moves along a straight line parallel to the bank.) For the distance between F and S it is natural to take the distance from S to the point on S *closest* to F. How does this distance D vary with time? When will $D =$

[7.7] The expressions "an interval tending to zero" and "a function tending to infinity" point to the fact that in essense we are speaking not of a single function but of a *family* of functions with the following property: as we go over from one function in this family to another, the "growth interval" becomes more and more narrow and the "degree of growth" higher and higher. For more details see Chapter 16.

D (t) be minimal? Draw the graph of the function $D = D(t)$ for $h = 300$ m, $l = 60$ m, and $v = 5$ m/s.

7.2.3. Find the cuspidal maxima of the following functions: (a) $y = (x - 5)\sqrt[3]{x^2}$, and (b) $y = 1 - \sqrt[3]{x^2}$.

7.3 Investigating Maxima and Minima of Functions Depending on a Parameter

In Sections 7.1 and 7.2 we separated the cases of "internal" maxima and minima, which are attained by a function in interior points of the domain of the function (the range of the independent variable on which the function depends), and maxima and minima attained by the function at the endpoints of the domain of the function. However, in studying functions one may encounter problems in which both types of maxima (or minima) are present. Situations of this kind occur when the function we are interested in depends not only on the independent variable but also on a parameter, and a variation in this parameter within the limits allowed by the problem changes the nature of the maxima (or minima) or even the nature of the function itself. To illustrate what we have just said there are several examples.

Example 1. Select a point M on the diameter AC of a circle σ. What must the angle α be between the diameter and a chord BD of σ passing through point M so that the quadrangle $ABCD$ inscribed in σ is of maximal area (Figure 7.3.1)?

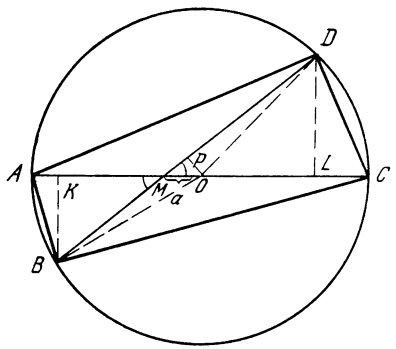


Figure 7.3.1

We take the radius of σ as the unit of length; the distance OM from the center of circle O to point M is denoted by a. Obviously, the altitudes BK and DL of $\triangle ABC$ and $\triangle ADC$ are, respectively, $MB \sin\alpha$ and $DM \sin\alpha$, which yields

$$S_{ABCD} = S = S_{\triangle ABC} + S_{\triangle ADC}$$
$$= \frac{1}{2} AC \cdot BM \sin\alpha + \frac{1}{2} AC \cdot DM \sin\alpha$$
$$= \frac{1}{2} AC\,(BM + DM) \sin\alpha$$
$$= \frac{1}{2} AC \cdot BD \sin\alpha = BD \sin\alpha,$$

since, obviously, $AC = 2$.

Moreover, the distance OP ($=b$) from the center of σ to the chord BD is $OM\sin\alpha = a\sin\alpha$, which yields $BP = PD = \sqrt{1 - a^2\sin^2\alpha}$ (since $OB = OD = 1$) and, hence, $BD = 2\sqrt{1 - a^2\sin^2\alpha}$, that is,

$$S = BD \sin\alpha = 2\sin\alpha\sqrt{1 - a^2\sin^2\alpha}. \tag{7.3.1}$$

The problem of finding the maximum of S becomes really simple if we go over from S to the square of the area $S^2 = 4\sin^2\alpha\,(1 - a^2\sin^2\alpha)$, since it is clear that S and S^2 attain their maxima (or minima) simultaneously, and introduce a new variable, $x = \sin^2\alpha$. Then

$$S^2 = 4x\,(1 - a^2x). \tag{7.3.2}$$

Since in the xy-plane ($y = S^2$) the graph of (7.3.2) is a *parabola*, we can do without finding the derivatives; however, since there is a general (and simple) method for finding the maxima and minima of a function through differentiation, we use this method. Clearly, if $F(x) = 4x - 4a^2x^2$, then $F'(x) = 4 - 8a^2x$ and $F'(x) = 0$ at $x = 1/2a^2$. And since $S(\alpha)$ (and, hence, $S^2(\alpha) = F(x)$) vanishes at $\alpha = 0$ and $\alpha = 180°$ and between these two values is positive, we conclude that it attains a maximum (see the text at the end of Example 2 in Section 7.1) somewhere in this interval; precisely, we should expect

that the maximum of $S(\alpha)$ and that of $F(x)$ is attained at $x = \sin^2\alpha = 1/2a^2$.

But it is clear that $\sin^2\alpha$ is equal to $1/2a^2$ only for $1/2a^2 \leqslant 1$, or $2a^2 \geqslant 1$, or $a \geqslant \sqrt{2}/2$, and in this case the solution of the problem is indeed supplied by the condition $\sin^2\alpha = 1/2a^2$, or $\sin\alpha = \pm\sqrt{2}/2a$; there are two chords, BD and B_1D_1, that are symmetric about AC and form the appropriate angle α with AC (at $a = \sqrt{2}/2$ merge into one chord $BD \perp AC$). For such a value of x the function $F(x)$ has an "internal" (*smooth*) maximum. For $a < \sqrt{2}/2$ the function $F(x)$ varies *monotonically* with x varying between 0 and 1 (since $x = \sin^2\alpha$), whereby the greatest value is attained at the *boundary point* $x = 1$, corresponding to $\alpha = 90°$ (since $x = \sin^2\alpha$): here only one chord, BD, which is perpendicular to AC, yields a maximum for S. We note also that in view of (7.3.2) the value $S = S_{\max}$ is expressed differently for the two cases we are considering: for $a \geqslant \sqrt{2}/2$ we have $F_{\max} = F(1/2a^2) = 4 \times (1/2a^2) \times (1 - 1/2) = 1/a^2$, which means that in this case $S_{\max} = \sqrt{F_{\max}} = 1/a$; but if $a \leqslant \sqrt{2}/2$, then $F_{\max} = F(1) = 4(1 - a^2)$ and, hence, $S_{\max} = \sqrt{F_{\max}} = 2\sqrt{1 - a^2}$. Thus, the graph for $S_{\max} = S(a)$ (in the aS-plane) is represented, on the segment $0 \leqslant a \leqslant 1$, by a "broken" curve consisting of two different arcs: the arc of the ellipse $a^2 + (S/2)^2 = 1$ on the segment $0 \leqslant a \leqslant \sqrt{2}/2$ and the arc of the hyperbola $S = 1/a$ on the segment $\sqrt{2}/2 \leqslant a \leqslant 1$ connected to the first arc.

Note also that the square T inscribed in σ for which AC is the midline cuts out of this midline a segment L_1L_2, with $OL_1 = OL_2 = \sqrt{2}/2$; thus, if point M lies inside T, then the sought chord BD is perpendicular to AC, but if M lies outside T, then the first case of the above two occurs. Here $b = OP = a\sin\alpha = a\sqrt{2}/2a = \sqrt{2}/2 =$ constant, and in this way the distance between BD and the center O of circle σ is constant and is equal to the distance from O to a side of square T; the chords BD and B_1D_1 that pass through point M, which is exterior with respect to T (lies outside T) and lies on AC, touch the circle σ_1 inscribed in T.

This result, which appears unexpected at first glance, becomes crystal clear if we draw the graph of the function $y = F(x) = 4x - 4a^2x^2 = -a^2(x - 1/2a^2)^2 + 1/4a^2$, a straight line at $a = 0$ and a parabola with its vertex at point $Q(1/2a^2, 1/4a^2)$ for $a \neq 0$ (Figure 7.3.2). For $a > \sqrt{2}/2$ the vertex Q lies inside the domain of $F(x)$, $0 \leqslant x \leqslant 1$, and the function attains its maximum at this vertex, while for $a < \sqrt{2}/2$ the point Q lies outside the domain, and the maximum is attained at the boundary point $x = 1$. Note also that for $a > \sqrt{2}/2$ the value $x = 1$ of the independent variable corresponds not to a *maximum* of $F(x)$ but, on the contrary, to a (local) *minimum* because for values of x close to unity but less than unity (for values of α close to a right angle) the magnitude of $F(x)$ (the area $S(\alpha)$) will be larger than at point $x = 1$ (at $\alpha = 90°$).

Of course, in solving this problem there was no need to go over from function $S(\alpha) = 2\sin\alpha\sqrt{1 - a^2\sin^2\alpha}$ to function $F(x)$. The derivative $S(\alpha)$ with respect to α can be found by the rules of differentiation. The result is

$$\frac{dS}{d\alpha} = \frac{2\cos\alpha \times (1 - 2a^2\sin^2\alpha)}{\sqrt{1 - a^2\sin^2\alpha}}. \qquad (7.3.3)$$

From (7.3.3) we see that the fact that $S'(\alpha)$ vanishes is equivalent to $\cos\alpha = 0$ or to

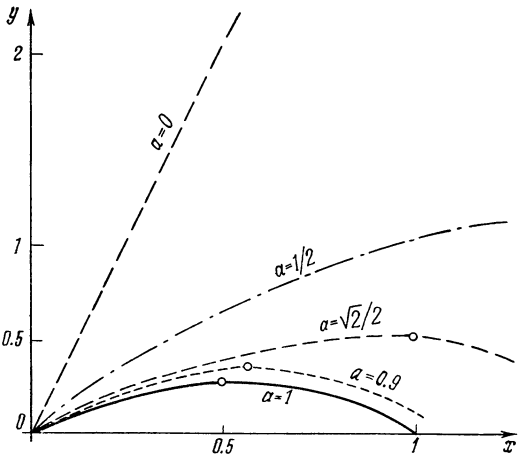


Figure 7.3.2

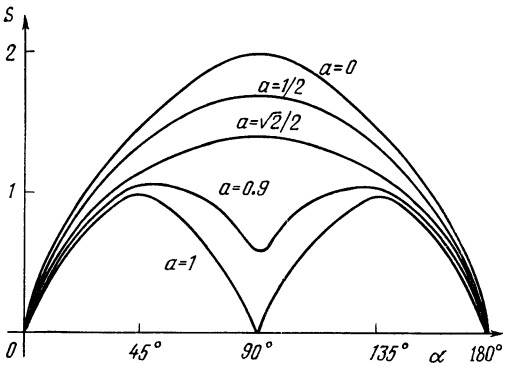


Figure 7.3.3

$\sin\alpha = \pm 1/a\sqrt{2}$. The first case corresponds to $\alpha = 90°$, that is, a chord $BD \perp AC$, while the second case is possible only if $a > \sqrt{2}/2$ ($\simeq 0.71$). Accordingly, we arrive at different solutions to this problem depending on whether $a > \sqrt{2}/2$ or $a < \sqrt{2}/2$. For $a > \sqrt{2}/2$, the function $S(\alpha)$ vanishes at a single point, $\alpha = 90°$, and since the function $S = S(\alpha)$ must have a maximum, there can be no doubt that $\alpha = 90°$ corresponds to precisely a maximum of the function. However, when a is less than $\sqrt{2}/2$, the derivative $S'(\alpha)$ vanishes at *three* points: at $\alpha = 90°$ and at $\sin\alpha = \pm 1/\sqrt{2}a$. If we check the sign of $S''(\alpha)$ (or check the sign of $S'(\alpha)$ in the neighborhood of points $\alpha = 90°$ and $\alpha = \pm\arcsin(1/\sqrt{2}a)$), we will see that in this case the *maxima* are attained precisely at points $\alpha = \pm\arcsin(1/\sqrt{2}a)$, while at point $\alpha = 90°$ the function has a local *minimum* (Figure 7.3.3).

Here is another aspect in which Figure 7.3.2 differs from Figure 7.3.3. In the former, point $x = 1$ is a *boundary* point of function $F(x)$, and this function has no geometrical meaning when x is greater than unity; the maximum (or minimum) of $F(x)$ at this point is a "boundary" one, since here $F'(1) \neq 0$. In the latter, on the other hand, point $\alpha = 90°$ is an *interior* point for $S(\alpha)$; the maximum (or local minimum) attained by $S(\alpha)$ at this point is a *smooth* one. This should not be a cause for surprise, since it is clear that the functions $S(\alpha)$ and $S^2(\alpha)$ ($= F(x)$) attain their maximum and minimum (since $S \geqslant 0$) in the same point, but the nature of the maximum (minimum) may change. For instance, all three functions $y = \sqrt{|x|}$, $y_1 = y^2 = |x|$, and $y_2 = y_1^2 = x^2$ (Figure 7.3.4) have the same minimum point $x = 0$; however, y has at this point a cuspidal minimum (both the left and right derivatives become infinite at this point), y_1 has a "corner" minimum (the left and right derivatives are nonzero and do not coincide), and y_2 has a smooth minimum at $x = 0$ (here $y_2'(0) = 0$). Indeed, if $Y = F(x) = y^2(x)$, then $dY/dx = 2yy'$, and the conditions $Y' = 0$ and $y' = 0$ are not equivalent. Another reason why the nature of the maximum or minimum may change is the change in the function itself (in addition to this, when we went over from S to F we simultaneously changed the independent variable from α to $x = \sin^2\alpha$): since if $y = f(x)$ and $x = \varphi(t)$, we have $dy/dt = (dy/dx)(dx/dt)$, and $dy/dx = 0$ is not equivalent to $dy/dt = 0$.

Figure 7.3.5 depicts the graph of the function $\alpha = \alpha(a)$, where α is defined by the condition $S_{ABCD} = S = S_{\max}$; in a certain sense this drawing summarizes the information we obtained during solving the problem. The reader can clearly see that if $a \leqslant \sqrt{2}/2$, the problem of finding the quadrangle $ABCD$ with the greatest possible area has a *unique* solution, while if $a > \sqrt{2}/2$, there are *two* solutions, and they correspond to two

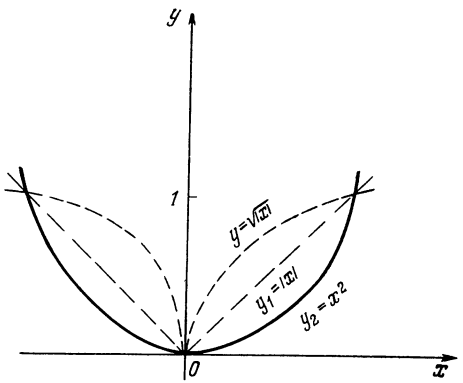


Figure 7.3.4

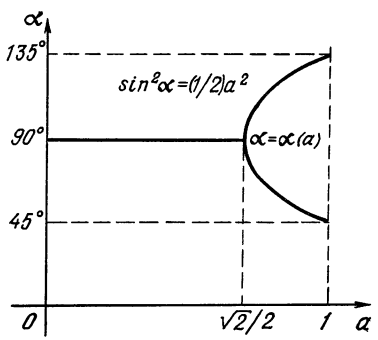


Figure 7.3.5

chords BD and B_1D_1 symmetric about AC. (Note that such a "splitting", or branching, at a critical point of one solution into two, an event corresponding to a change in the nature of the solution to the problem, is rather often encountered in mathematics.) Also, the very fact that at the critical point $a = \sqrt{2}/2$ the point of maximum $\alpha = 90°$ of the function $S = S(\alpha)$ transforms into a point of (local) minimum is inherent in many problems and not only in this, rather randomly chosen, problem.

Of course, this problem, too, can be solved without resorting to differential calculus, that is, by purely geometrical methods. It is obvious that

$$S_{ABCD} = S = S_{\triangle MAB} + S_{\triangle MBC} + S_{\triangle MCD} + S_{\triangle MDA}$$

$$= \frac{1}{2}\sin\alpha \times (MA \cdot MB + MB \cdot MC + MC \cdot MD + MD \cdot MA)$$

$$= \frac{1}{2}(AC \cdot BD)\sin\alpha = BD\sin\alpha;$$

on the other hand,

$$S_{\triangle OBD} = s = \frac{1}{2}BD \cdot OP = \frac{1}{2}BD \cdot a\sin\alpha$$

$$= \frac{1}{2}a \cdot BD\sin\alpha,$$

that is, $s = (a/2)S$, which means that s and S attain their maxima simultaneously. But since

$$s = \frac{1}{2}OB \cdot OD \sin \angle BOD = \frac{1}{2}\sin\beta$$

where $\beta = \angle BOD$, it remains to be established at what point $\sin\beta$ attains its maximum. Clearly, if β can be equal to 90° ($\sin\beta = 1$), this will correspond to the absolute maximum of s (and, hence of S, too). But $\beta = \angle OBD = 90°$ if the length of the chord BD of σ is equal to $\sqrt{2}$, the length of the side of the square T inscribed in σ, or if the distance OP from the center of σ to BD is equal to half the side of T (the half, of course, is $\sqrt{2}/2$ units long), that is, when BD touches the circle σ_1 inscribed in T, which is obviously impossible if point M belonging to AC lies inside the square T (with midline AC) inscribed in σ. Thus, when point M belonging to AC lies outside T, the problem is solved by drawing through M the straight line BD tangent to σ_1 (either BD or B_1D_1) for which $\beta = 90°$; but if M lies inside T, the situation changes.

In the latter case the distance OP from the center O to BD is equal, as we saw earlier, to $a\sin\alpha$, that is, it can vary from 0 (the case where $\alpha = 0$ and BD coincides with AC) to a (the case where $\alpha = 90°$, or $BD \perp AC$). Accordingly, the length $d = BD$ of chord BD can vary from 2 (when $BD \equiv AC$) to $2\sqrt{1-a^2}$ (when $BD \perp AC$); the angle β varies from 180° (when $BD \equiv AC$) to $2\arccos a$ (when $BD \perp AC$), that is, remains constantly obtuse (since a is less than $\sqrt{2}/2$, we conclude that $\arccos a$ is greater than 45°). The quantity $\sin\beta$ attains its maximum at $\beta = \beta_{\min} = 2\arccos a$, that is, at $BD \perp AC$, and it is this chord BD that ensures that the area of $ABCD$ is at its maximum.

Of course, the unnatural solution that we are discussing here is much more difficult to comprehend than the elegant solution involving finding the derivative of $S(\alpha)$ (or $F(x)$), since one must guess that the areas of S and s are proportional, and this is not an easy job.

Example 2. Let the product of the distances from a (variable) point M to two fixed points A and B be a. What is the greatest distance $d = MP = d_{\max}$ from M to the straight line AB, and how to specify the point M for which this distance $d_{\max}$ is realized?

In many respects this problem is similar to the previous one, and for this reason we go into less detail. The set of all points M for which $MA \cdot MB =$ constant $(= a)$ constitutes a curve Σ called the ***oval of Cassini*** (Figure 7.3.6) in honor of the Italian astronomer Giovanni Domenico ***Cassini*** (1625-1712). (Being a follower of the heliocentric theory of the solar system, Cassini tried to correct the faults of Ptolemy's system by substituting for Kepler's ellipses as the paths of the planets curves like the one just mentioned; such curves are also called ***Cassinians***). If we introduce a plane system of coordinates in such a manner that points A and B have coordinates $(-1, 0)$ and $(1, 0)$ (in this way the distance AB between the points is equal to two units of length), then, in view of the fact that for $M = M(x, y)$, obviously, $MA = \sqrt{(x+1)^2 + y^2}$ and $MB = \sqrt{(x-1)^2 + y^2}$, the equation of curve Σ has the form

$$\sqrt{(x+1)^2+y^2} \times \sqrt{(x-1)^2+y^2} = a,$$

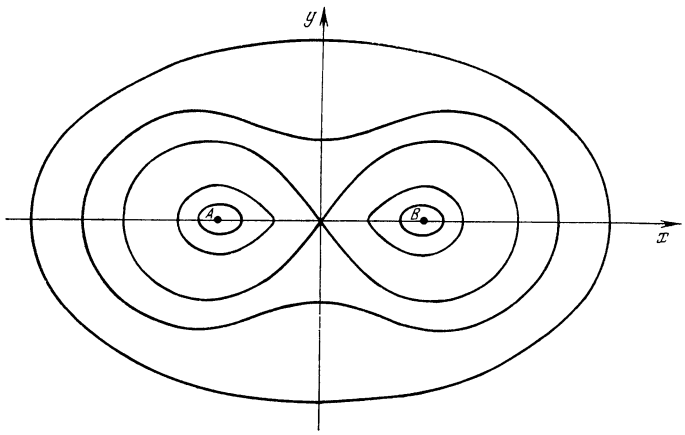


Figure 7.3.6

which can be simplified thus:

$$(x^2 + y^2)^2 - 2(x^2 - y^2) = a^2 - 1. \tag{7.3.4}$$

(Verify this.)

It is clear that the oval of Cassini degenerates into a pair of points, A and B (in which case our problem has no meaning), if $a = 0$. If a is very small, then either the distance MA or the distance MB is very small (otherwise the product $MA \cdot MB$ cannot be small), and in this case the oval of Cassini splits into two ovals surrounding points A and B. As a increases, the two ovals of Cassini increase in size and, at $a = 1$, finally touch at the origin O, so that the entire curve resembles an "eight" lying on its side. This curve (in view of (7.3.4), its equation is $(x^2 + y^2)^2 = 2(x^2 - y^2)$), was studied by Jakob Bernoulli and is called the ***lemniscate of Bernoulli.*** When a is greater than unity, the oval of Cassini is a single closed curve enveloping the two "poles" of the curve, A and B. Finally, when a is very large, the distances MA and MB are also large and will differ little; here the oval of Cassini resembles a circle of an extremely large radius $\sqrt{a}$ centered at O.

Let us now find the derivative dy/dx, where the dependence of y on x is given by the following equation: $F(x, y) = 0$, with $F(x, y) = (x^2 + y^2)^2 - 2(x^2 - y^2) - a^2 + 1$ (cf. (7.3.4)). In view of (4.13.7), we have

$$\frac{dy}{dx} = -\frac{\partial F/\partial x}{\partial F/\partial y} = -\frac{2(x^2+y^2)\,2x - 4x}{2(x^2+y^2)\,2y + 4y} = \frac{(x^2+y^2-1)\,x}{(x^2+y^2+1)\,y}, \tag{7.3.5}$$

whence $dy/dx = 0$ *if and only if* $x = 0$ or $x^2 + y^2 = 1$. But at $x = 0$ the equation of the oval of Cassini, (7.3.4), yields $y^4 + 2y^2 = a^2 - 1$, or $(y^2 + 1)^2 = a^2$, from which it becomes completely clear that for $a < 1$ the oval of Cassini does not intersect the straight line $x = 0$ (the y axis); here $dy/dx = 0$ only at points for which $x^2 + y^2 = 1$, or at points where the oval of Cassini intersects the unit circle S with diameter AB (since S has the equation $x^2 + y^2 = 1$). It is clear that there are four points at which Σ intersects S, and these points are pairwise symmetric about the x axis and the y axis. They necessarily yield the solution to our problem, since they lie farthest from the x axis.

The points of intersection of Σ and S can be found by solving the following system of equations:

$$x^2 + y^2 = 1 \quad \text{and}$$
$$2(x^2 - y^2) = (x^2 + y^2)^2 + 1 - a^2 = 2 - a^2,$$

or

$$x^2 + y^2 = 1 \quad \text{and}$$
$$x^2 - y^2 = 1 - a^2/2,$$

which immediately yields

$$x^2 = 1 - a^2/4 \quad \text{and} \quad y^2 = a^2/4.$$

This readily implies that these points exist only when a is no greater than 2, since x^2 is necessarily positive. Whence, for $a > 2$ the condition $dy/dx = 0$ is satisfied only at $x = 0$, which means that the points of the oval of Cassini that lie farthest from the x axis (such points are sure to exist) are those at which Σ intersects the y axis.

Finally, for $1 < a < 2$, the derivative dy/dx on the upper half of the oval of Cassini (in view of the fact that Σ is symmetric about the axis of abscissas we can always consider only the upper half of this curve) vanishes three times: at the point where Σ intersects the axis of ordinates and at the two points where Σ and S intersect. If we study how dy/dx changes sign near these points (or find the sine of d^2y/dx^2), we will see that y is at its maximum exactly at the points where Σ and S intersect, while $x = 0$ corresponds to a (local) minimum of y (see Figure 7.3.6). Thus, the set of all points of maxima of y for different values of a consists of circle S and the part of the axis of ordinates external to S (Figure 7.3.7): for $a \leqslant 2$ the point of maximum of y belongs to circle S, and the closer a is to 2, the closer is the pair of points above the x axis to each other (the same, of course, is true of the pair of points below the x axis); at $a = 2$ the two points with $y > 0$ (and the two points with $y < 0$) merge, and for a greater than 2 the points of maximum of y belong to the y axis and not to circle S. (The reader can easily convince himself that the graph of $y = y_{\max}(a)$ in the ay-plane consists of the segment of the straight line $y = a/2$ corresponding to $0 \leqslant a \leqslant 2$ and the arc of the parabola $y^2 = a - 1$ corresponding to the values of a greater than 2.)

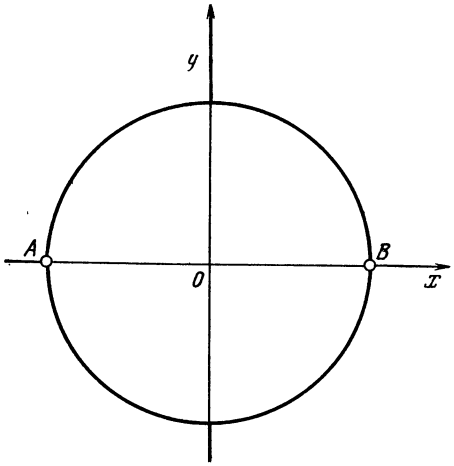


Figure 7.3.7

Of course, this problem can also be solved by the methods of elementary geometry, that is, without using differential calculus. It is clear that

$$S_{\Delta MAB} = \frac{1}{2} MA \cdot MB \sin\alpha = \frac{a}{2} \sin\alpha,$$

while on the other hand

$$S_{\Delta MAB} = \frac{1}{2} AB \cdot MP = MP,$$

where MP is the distance from M to AB, and $\alpha = \angle AMB$, whereby the maximum of MP corresponds to the maximum of $\sin\alpha$. Therefore, if our curve Σ contains points M such that $\angle AMB = 90°$ and $\sin\alpha = 1$, then these points yield the maximum of MP, since all points M of the circle S whose diameter is AB with $\angle AMB = 90°$ are indeed the points of maximum of MP (of course, different points on S correspond to different values of parameter a). Completing ΔMAB to parallelogram MAM_1B centered at O readily leads to

$$\begin{aligned} 4MO^2 &= 2MA^2 + 2MB^2 - AB^2 \\ &= 2(MA - MB)^2 + 4MA \cdot MB - 4 \\ &= 4(a - 1) + 2(MA - MB)^2 \\ &\geqslant 4(a - 1), \end{aligned}$$

where we still assume that $AB = 2$. Thus, if $a > 2$, then the equality $MO = 1$, which means that M belongs to Σ and to circle S with diameter AB and, hence, $\alpha = \angle AMB = 90°$, is simply impossible (while at $a = 2$ only the points of intersection of Σ with the axis of ordinates, points for which $MA - MB = 0$, satisfy the condition $MO = 1$). Hence, while for $a \leqslant 2$ the maximum of the distance $|y| = MP$ from point M on curve Σ

to the axis of abscissas is attained at the points where Σ and S intersect, for $a > 2$ the situation is different.

The reader can easily see that if $a > 2$, then for each point M belonging to Σ the angle $\alpha = \angle AMB$ is always acute (since point M lies outside circle S whose diameter is AB); therefore, we need only find a point M for which angle α (and, hence, $\sin\alpha$) is *maximal*. By the laws of cosines applied to $\angle MAB$,

$$2MA \cdot MB \cos\alpha$$
$$= 2a\cos\alpha = MA^2 + MB^2 - AB^2$$
$$= (MA - MB)^2 + 2MA \cdot MB - 4$$
$$= (MA - MB)^2 + 2a - 4,$$

which implies that the smaller the value of $\cos\alpha$ (and, hence, the greater the value of α), the smaller the quantity $(MA - MB)^2$, whence the smallest possible value of $\cos\alpha$ (and, hence, the greatest possible value of $\sin\alpha$) will be achieved at $MA - MB = 0$, that is, when M belongs to the y axis. Thus, for $a > 2$ the maximum distance MP from M to AB is attained for the points on the axis of ordinates. This completes the solution to the problem.

Example 3. Here is another example. It is more complicated than the previous two, has a serious physical meaning, and also deals with finding the maxima and minima of a function. It is well known that the relationship between the volume v, the pressure p, and the absolute temperature T (in kelvins) of an ideal gas is given by Boyle's law and Gay-Lussac's law, which can be united into the ideal gas law

$$pv = RT, \tag{7.3.6}$$

where R is the ***molar gas constant of an ideal gas***, the adjective "molar" meaning that when related to one mole of a gas, R is the same for all gases; for instance, if v is measured in m^3 and p in Pa, then $R \simeq 8.25$ N·m/K. However, formula (7.3.6) does not agree very well with the properties of real gases, and so a common way to correct it is to write it in the form

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT, \tag{7.3.7}$$

which is known as the ***van der Waals equation of state***.[7.8] The constants a and b are empirical (i.e. determined for each gas separately) positive constants (in the system of units used above to define R the dimensions of these constants are $N \cdot m^4$ and m^3, respectively). If the temperature of the gas under investigation is kept constant, the formulas (7.3.6) and (7.3.7) correspond to certain curves in the vp-plane known as ***isotherms*** (curves of equal temperature). Each value of T has its isotherm.

It is clear that ideal-gas isotherms given by formula (7.3.6) for different values of (parameter) T comprise a family of hyperbolas $pv = \text{constant}\ (= RT)$ with asymptotes $v = 0$ and $p = 0$; these isotherms have neither maxima nor minima. The situation complicates considerably when we go over to the van der Waals isotherms (7.3.7). We will now investigate such isotherms. The curve specified by (7.3.7) depends on *three* parameters: the coefficients a and b in the van der Waals equation of state (these coefficients are characteristics of the gas being investigated) and the temperature of the gas, T. Later we will see that the qualitative behavior of isotherm (7.3.7) is determined by the ratio of T to the fraction a/b.

Let us use (7.3.7) and find how p depends on v, or the function $p = p(v)$. The answer is

$$p = \frac{RT}{v - b} - \frac{a}{v^2}. \tag{7.3.8}$$

From this it follows that

$$p' = \frac{dp}{dv} = -\frac{RT}{(v-b)^2} + \frac{2a}{v^3}$$
$$= \frac{1}{(v-b)^2}\left[\frac{2a(v-b)^2}{v^3} - RT\right], \tag{7.3.9}$$

in view of which the possible maxima and minima of $p = p(v)$ (for a given temperature T) are determined from the condition

$$\frac{2a(v-b)^2}{v^3} - RT = 0, \tag{7.3.10}$$

which is equivalent to $p' = 0$. Unfortunately, however, (7.3.10) is a cubic equation in v, and the solution of this equation is not that simple.

Not being able to solve Eq. (7.3.7) directly, we will study it qualitatively. Consider the function $f(v) = 2a(v - b)^2/v^3 - RT$ (which depends on the same three parameters a, b, and T). We will try to establish how it varies, that is, find its maxima and minima. To this end we find the derivative of f:

$$\frac{df}{dv} = 2a\left[\frac{2(v-b)}{v^3} - \frac{3(v-b)^2}{v^4}\right]$$
$$= -\frac{2a(v-b)(v-3b)}{v^4}. \tag{7.3.11}$$

From the physics of the problem (see Eq. (7.3.7)) it follows that v is always greater than b (the absolute temperature T cannot be negative), whereby (7.3.11) may vanish only at $v = 3b$. Further, for $b < v < 3b$ the right-hand side of (7.3.11) is negative, which implies that $v = 3b$ corresponds to the *maximum* of $f(v)$, the greatest possible value, equal to $8a/27b - RT$. Now we must consider three different cases, corresponding to three temperature intervals.

[7.8] Johannes Diderik ***van der Waals*** (1837-1923), a Dutch physicist.

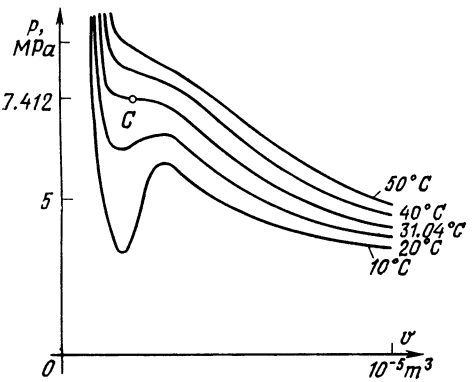


Figure 7.3.8

(a) $8a/27b - RT < 0$, that is $T > 8a/27bR$, or $T/(a/b) > 8/27R$. In this case $f(v) < 0$ for all v and this means that p' is negative for all values of v, whereby the function $p = p(v)$ is constantly decreasing as v grows (the upper two van der Waals isotherms for carbon dioxide in Figure 7.3.8; the temperature at each curve is given in degrees Celsius). The respective curves of p vs. v resemble hyperbolas, and the asymptotes are the straight lines $p = 0$ and $v = b$.

(b) $8a/27b - RT > 0$, that is $T < 8a/27bR$, or $T/(a/b) > 8/27R$ (see the lower two curves in Figure 7.3.8). Since at $v = b$ the function $f(v)$ is negative (it is unimportant that there is no such state of matter with $v = b$) and at $v = 3b$ we have $f(v) = 8a/27b - RT > 0$, between b and $3b$ there must be a value v_0 of v such that $f(v_0) = 0$ and, hence $p'(v_0) = 0$. For this value the function $p = p(v)$ is at a *minimum* (for smaller values of v the derivative p' is negative and for larger values it is positive). This is not all, however. As v becomes greater and greater ($v \to \infty$), the value of p diminishes (in view of (7.3.8)), and so somewhere at $v = v_1 > 3b$ the function $p = p(v)$ must attain its *maximum*. We see that while in case (a) the van der Waals isotherms have neither maxima nor minima and resemble hyperbolas (ideal gas isotherms), in case (b) their nature changes: here the function $f(v)$ $(=p)$ has a (local) minimum and a (local) maximum. The asymptotes of the p vs. v curves in case (b) are the same as in case (a).

(c) $8a/27b - RT = 0$, that is, $T = 8a/27bR$, or $T/(a/b) = 8/27R$. Since the greatest possible value of $f(v)$ here is zero, for all other values of v this function is negative, which means that $p' = f(v)/(v - b)^2$ is negative, too. Thus, the function $p = p(v)$ is everywhere a decreasing function, just as in case (a). At $v = 3b$ it has a *point of inflection* (since $p'(3b) = 0$).

In the case (a), obviously, to each value of pressure p there corresponds a single value of volume v, which corresponds to the possibility of only one state of matter occurring at $T > 8a/27bR$, and this is the gaseous state. On the other hand, if $T < 8a/27bR$, to each value of pressure p there correspond, as shown by Figure 7.3.8, three different values of volume v, so that here matter can exist in different states. The state with the smallest volume $v = v_{liq}$ (i.e. the greatest density) is the *liquid state*. The state with the greatest volume $v = v_{gas}$ (at the same temperature and pressure as the liquid state) is the *gaseous state*. Finally there is the intermediate state, which proves to be unstable. If the substance we are studying is placed in a vessel of a certain volume v (somewhere between v_{liq} and v_{gas}) and heated up to a temperature T lower than a certain temperature T_c (see below), part of the substance will be in the gaseous state and the remainder will be in the liquid state. A definite state with the intermediate volume v does not exist.

The intermediate case with $T = 8a/27bR$ plays an exceptional role: the temperature T in this case is called the ***critical temperature*** of the gas (it is usually denoted by T_c). The point of inflection C of the respective isotherm corresponds to the following values of volume and pressure: $v = 3b$ $(=v_c)$ and $p = a/27b^2$ $(=p_c)$; these volume and pressure are also called ***critical***. At $T = T_c$, if $p > p_c$ (or $v < v_c$), the substance is in the gaseous state, and if $p < p_c$ (or $v > v_c$), the substance is in the liquid state; thus, p_c is the highest pressure at which the substance can exist in the form of saturated vapor. Note also that point C of the critical isotherm (each substance has only one critical isotherm) corresponds to $p_c v_c = a/9b = (3/8)\,RT_c$, which is only three-eighths of the quantity predicted by the ideal gas law.

A fuller study of the van der Waals equation, its consequences, and further modifications can be found in textbooks on physics and engineering thermodynamics.

7.4* Convex Functions and Algebraic Inequalities

In Section 1.4 we defined the *convexity of a curve* $y = f(x)$ (or a function $y = f(x)$) in the following fashion: a function $y = f(x)$ is said to be ***convex*** on the interval $a \leqslant x \leqslant b$ of variation of x (or curve $y = f(x)$ is said to be ***convex upward***, or simply ***convex***) if any chord AB of this curve $(A = A(x_1, f(x_1))$ and $B = B(x_2, f(x_2))$, with $a \leqslant x_1 < x_2 \leqslant b$) lies *below* the corresponding arc AB of the curve $y = f(x)$ (Figure 7.4.1*a*). On the other hand, if a chord

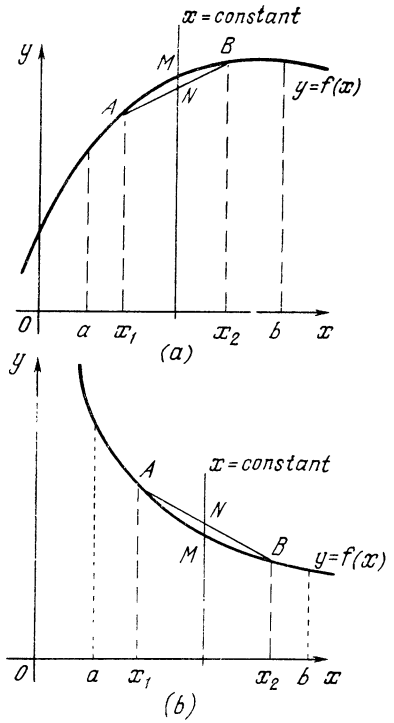


Figure 7.4.1

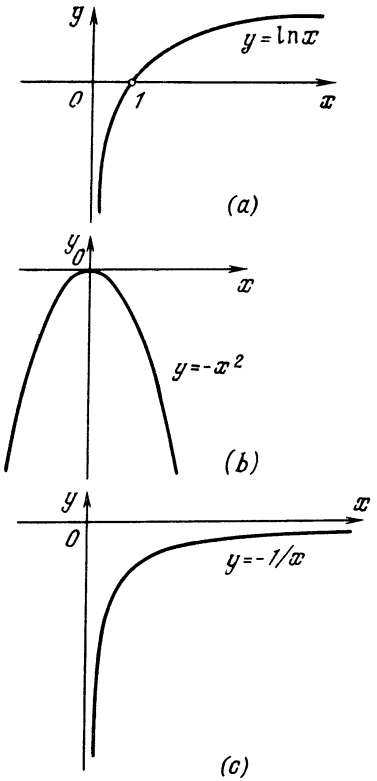


Figure 7.4.2

AB lies *above* the corresponding arc AB of the curve $y = f(x)$, the function f is said to be *concave* (or the curve $y = f(x)$ is said to be ***convex downward***, or simply ***concave***) (Figure 7.4.1*b*). In other words, the function $y = f(x)$ is ***convex*** on the interval $a \leqslant x \leqslant b$ if the point N at which the straight line $x = \text{constant}$ $(a \leqslant x_1 < x < x_2 \leqslant b)$ intersects the chord AB lies *below* the point M where the same straight line intersects the curve representing $y = f(x)$; on the other hand, if N lies *above* M, the function f is ***concave***. This definition already relates the concept of convexity with inequalities, since if the points M and N have coordinates (x, y) and (x, Y), the convexity of $y = f(x)$ implies $y > Y$, while the concavity of $y = f(x)$ implies $y < Y$.

In Section 2.7 we established a simple condition for the convexity of a function $y = f(x)$, namely, that a function is convex *if its second derivative*, $y'' = d^2y/dx^2$, *is negative*. For instance, from this convexity condition it immediately follows that the functions $y = \ln x$, $y = -x^2$, and $y = -1/x$ (for $x > 0$) are convex since their second derivatives are respectively, $y'' = (1/x)' = -1/x^2$, $y'' = (-2x)' = -2$, and $y'' = (1/x^2)' = -2/x^3$ (Figure 7.4.2; the reader will recall that the function $y = \ln x$ is defined only for $x > 0$).

There exists a simple but important

Theorem 1 *If $y = f(x)$ is a function that is convex on the interval from a to b and x_1 and x_2 are two values of the independent variable within the interval (i.e. two arbitrary numbers such that $a \leqslant x_1 < x_2 \leqslant b$), then*

$$\frac{f(x_1)+f(x_2)}{2} < f\left(\frac{x_1+x_2}{2}\right). \qquad (7.4.1)$$

Proof. In Figure 7.4.3, $OA = x_1$ and $OB = x_2$. Then $AM = f(x_1)$ and $BN = f(x_2)$. Moreover, if S is the middle of

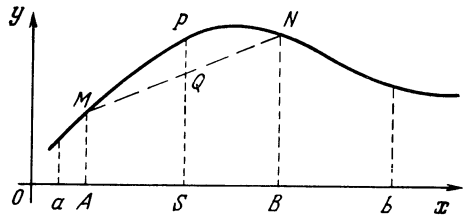


Figure 7.4.3

the line segment AB, then $OS = (x_1 + x_2)/2$ and, hence, $SP = f\left(\frac{x_1 + x_2}{2}\right)$. On the other hand, since the length of the midline SQ of the trapezoid $ABNM$ is equal to one-half the sum of the lengths of the bases AM and BN, we can write $SQ = [f(x_1) + f(x_2)]/2$. But according to the definition of a convex function, the midpoint Q of the chord MN lies *below* point P of the arc MN; hence

$$\frac{f(x_1)+f(x_2)}{2} < f\left(\frac{x_1+x_2}{2}\right),$$

which is what we set out to prove.[7.9]

Examples.

(a) $y = \ln x$. The convexity of this function implies that

$$\frac{\ln x_1 + \ln x_2}{2} < \ln \frac{x_1+x_2}{2},$$

i.e. $$\ln \sqrt{x_1 x_2} < \ln \frac{x_1+x_2}{2},$$

or, finally,

$$\sqrt{x_1 x_2} < \frac{x_1+x_2}{2}, \qquad (7.4.2)$$

which demonstrates that the *geometric mean of two distinct positive numbers is smaller than their arithmetic mean.*

(b) $y = -x^2$. Reasoning in the same way as in (a), we get

$$-\frac{x_1^2+x_2^2}{2} < -\left(\frac{x_1+x_2}{2}\right)^2,$$

or, in another form,

$$\frac{x_1^2+x_2^2}{2} > \left(\frac{x_1+x_2}{2}\right)^2,$$

$$\sqrt{\frac{x_1^2+x_2^2}{2}} > \frac{x_1+x_2}{2}$$

The expression $\sqrt{\frac{a_1^2+a_2^2+\ldots+a_k^2}{k}}$, which is the square root of the arithmetic mean of k squares of the numbers $a_1, a_2, \ldots, a_k$, is known as the ***root-mean-square*** of these numbers. Thus, the above result can be formulated as follows: *the root-mean-square of two distinct positive numbers is always greater than the arithmetic mean of these numbers.*

(c) $y = -1/x$. Theorem 1 implies that

$$-\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right) < -\frac{1}{(x_1+x_2)/2},$$

or $$\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right) > \frac{2}{x_1+x_2},$$

or, finally,

$$\frac{2}{1/x_1+1/x_2} < \frac{x_1+x_2}{2}. \qquad (7.4.3)$$

The quotient $1 \div \frac{1/a_1+1/a_2+\ldots+1/a_k}{k}$ $\left(= \frac{k}{1/a_1+1/a_2+\ldots+1/a_k}\right)$, which is simply the arithmetic mean of the numbers $1/a_1, 1/a_2, \ldots, 1/a_k$ (k reciprocals of the positive numbers $a_1, a_2, \ldots, a_k$), is called the ***harmonic mean*** of the numbers $a_1, a_2, \ldots, a_k$. Thus, the inequality (7.4.3) states that *the harmonic mean of two distinct positive numbers is smaller than the arithmetic mean of these numbers.*

Theorem 1 can be generalized as follows:

Theorem 2 *If a function $y = f(x)$ is convex in the interval from a to b and x_1 and x_2 are two values of the independent variable within the interval*

[7.9] In proving Theorem 1 (and all subsequent theorems in this section), we confine our discussion to the case where $f(x_1)$ and $f(x_2)$ are of the same sign. We advise the reader to consider the case of opposite signs on his own (instead of employing the property of the midline of a trapezoid, the reader should use the following theorem: the length of the segment of the midline of a trapezoid lying between the diagonals is equal to one-half the difference of the lengths of the bases).

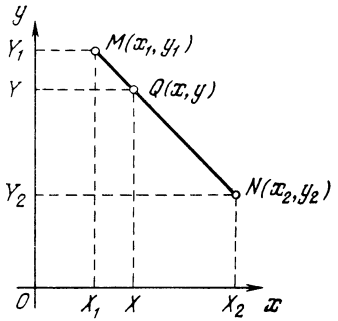


Figure 7.4.4

($a \leqslant x_1 < x_2 \leqslant b$) and p and q are two arbitrary positive numbers whose sum is equal to unity, then

$$pf(x_1) + qf(x_2) < f(px_1 + qx_2). \tag{7.4.4}$$

(For $p = q = 1/2$ this theorem transforms into Theorem 1.)

Proof. First of all, we note that if M and N are two points with coordinates (x_1, y_1) and (x_2, y_2) and Q is the point that divides segment MN in the ratio $MQ \div QN = q \div p$, with $p + q = 1$, then the coordinates of Q are $(px_1 + qx_2, py_1 + qy_2)$. Indeed, if we denote by X_1, X_2, and X and Y_1, Y_2, and Y the projections of points M, N, and Q on the x and y axes (Figure 7.4.4), then points X and Y divide the segments X_1X_2 and Y_1Y_2 in the ratio $q \div p$. This yields[7.10]

$$\begin{aligned} OX &= OX_1 + X_1X \\ &= x_1 + q(x_2 - x_1) \\ &= (1 - q)x_1 + qx_2 = px_1 + qx_2 \end{aligned}$$

and

$$\begin{aligned} OY &= OY_2 + Y_2Y \\ &= y_2 + p(y_1 - y_2) \\ &= (1 - p)y_2 + py_1 = py_1 + qy_2. \end{aligned}$$

We return now to the graph of our convex function $y = f(x)$ (Figure 7.4.5). Let us assume that $OA = x_1$, $OB = x_2$,

[7.10] In Figure 7.4.4 we depicted the case where all four numbers x_1, x_2, y_1, and y_2 are positive. The reader is advised to consider the alternative cases himself.

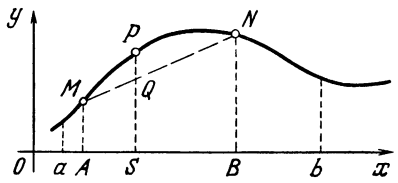


Figure 7.4.5

$AM = f(x_1)$, and $BN = f(x_2)$. According to what we have just proved, the coordinates of the point Q that divides MN in the ratio $MQ \div QN = q \div p$ are $(px_1 + qx_2, pf(x_1) + qf(x_2))$; thus, the length of SQ in Figure 7.4.5 is equal to $pf(x_1) + qf(x_2)$ and that of SP is $f(px_1 + qx_2)$. But since $y = f(x)$ is a convex function, point Q lies below point P, which means that $pf(x_1) + qf(x_2) < f(px_1 + qx_2)$, which is what we set out to prove.[7.11]

Examples.

(a) $y = \ln x$. In this case (7.4.4) yields

$$p \ln x_1 + q \ln x_2 < \ln(px_1 + qx_2),$$

whence

$$x_1^p x_2^q < px_1 + qx_2, \quad p > 0, \quad q > 0, \quad p + q = 1.$$

(b) $y = -x^2$. We have

$$-px_1^2 - qx_2^2 < -(px_1 + qx_2)^2,$$

$$\text{or } px_1^2 + qx_2^2 > (px_1 + qx_2)^2,$$

$$\text{or } \sqrt{px_1^2 + qx_2^2} > px_1 + qx_2,$$

where $p > 0$, $q > 0$, and $p + q = 1$.

(c) $y = -1/x$. Here we have

$$-\frac{p}{x_1} - \frac{q}{x_2} < -\frac{1}{px_1 + qx_2},$$

$$\frac{p}{x_1}\,\frac{q}{x_2} > \frac{1}{px_1 + qx_2},$$

[7.11] As the reader can easily see, the coordinates of *any* point belonging to the segment MN can be represented in the form $(px_1 + qx_2, py_1 + qy_2)$, where the numbers p and q (which are different for each such point) are such that $p > 0$, $q > 0$, and $p + q = 1$. Thus, the inequality (7.4.4) states that the entire chord MN lies below the curve $y = f(x)$, which is equivalent to the function being convex.

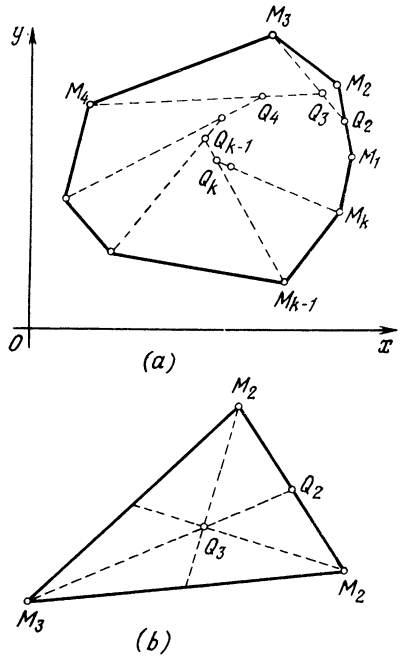


Figure 7.4.6

where $x_1 > 0$, $x_2 > 0$, $p > 0$, $q > 0$, and $p + q = 1$.

Theorem 1 can also be generalized as follows:

Theorem 3 *If $y = f(x)$ is a function convex on the interval from a to b and $x_1, x_2, \ldots, x_k$ is a set of k values of the independent variable within this interval such that some are distinct, then*

$$\frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k} < f\left(\frac{x_1+x_2+\ldots+x_k}{k}\right) \qquad (7.4.5)$$

(a particular case of ***Jensen's inequality***[7.12]). For $k = 2$ Theorem 3 transforms into Theorem 1.

Proof. We start by defining a concept that is often used in geometrical and analytical problems. Suppose that $M_1M_2M_3 \ldots M_k$ is an arbitrary k-gon (Figure 7.4.6*a*), Q_2 is the midpoint of side M_1M_2 of this k-gon ($M_1Q_2 \div Q_2M_2 = 1/2 \div 1/2$), Q_3 is the point that divides M_3Q_2 in the ratio $2 \div 1$ ($M_3Q_3 \div Q_3Q_2 = 2/3 \div 1/3$), Q_4 is the point that divides M_4Q_3 in the ratio $3 \div 1$ ($M_3Q_4 \div Q_4Q_3 = 3/4 \div 1/4$), ..., and, finally, Q_k is the point that divides M_kQ_{k-1} in the ratio $(k-1) \div 1$, that is, $M_kQ_k \div Q_kQ_{k-1} = (k-1)/k \div 1/k$.

Point Q_k is known as the ***centroid*** (the center of mass) of the k-gon $M_1M_2 \ldots M_k$. In the case of a triangle, $M_1M_2M_3$ (Figure 7.4.6*b*) the centroid Q_3 *coincides with the point of intersection of the medians* of the triangle; indeed, in this case Q_2 is the middle of side M_1M_2, the segment M_3Q_2 is a median, and the point Q_3 that divides M_3Q_2 in the ratio $M_3Q_3 \div Q_3Q_2 = 2 \div 1$ is the point of intersection of the medians of the triangle.

Let us prove that *if the coordinates of the vertices $M_1, M_2, \ldots, M_k$ of a k-gon are $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$, then the coordinates of the centroid Q_k are $((x_1 + x_2 + \ldots x_k)/k, (y_1 + y_2 + \ldots + y_k)/k)$.*[7.13]

Indeed, by the assumption we made at the beginning of the proof of Theorem 2, the points $Q_2, Q_3, Q_4, \ldots, Q_k$ have the following coordinates:

$$Q_2\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right);$$

$$Q_3\left(\frac{2}{3}\frac{x_1+x_2}{2}+\frac{1}{3}x_3, \frac{2}{3}\frac{y_1+y_2}{2}+\frac{1}{3}y_3\right),$$

$$\text{or } Q_3\left(\frac{x_1+x_2+x_3}{3}, \frac{y_1+y_2+y_3}{3}\right);$$

$$Q_4\left(\frac{3}{4}\frac{x_1+x_2+x_3}{3}+\frac{1}{4}x_4, \frac{3}{4}\frac{y_1+y_2+y_3}{3}+\frac{1}{4}y_4\right),$$

7.12 Johan Ludvig William Valdemar ***Jensen*** (1859-1925), a Danish analyst, algebraist, and engineer.

7.13 This implies, for one, that the centroid of a k-gon is determined solely by the k-gon and does not depend on the order in which the vertices of the k-gon are numbered (contrary to what we may assume from the definition of a centroid). In the case of a triangle this also follows from the fact that the centroid of a triangle coincides with the point of intersection of the medians.

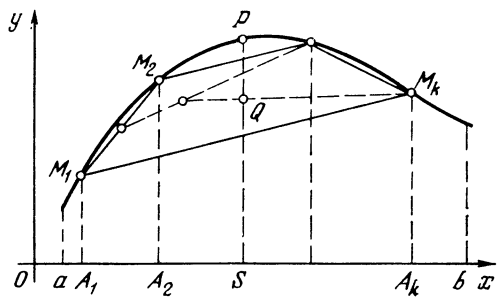


Figure 7.4.7

or

$$Q_4\left(\frac{x_1+x_2+x_3+x_4}{4}, \frac{y_1+y_2+y_3+y_4}{4}\right),$$

. . . . . .

$$Q_k\left(\frac{k-1}{k}\frac{x_1+x_2+\ldots+x_{k-1}}{k-1}+\frac{1}{k}x_k, \frac{k-1}{k}\frac{y_1+y_2+\ldots+y_{k-1}}{k-1}+\frac{1}{k}y_k\right),$$

or

$$Q_k\left(\frac{x_1+x_2+\ldots+x_{k-1}+x_k}{k}, \frac{y_1+y_2+\ldots+y_{k-1}+y_k}{k}\right).$$

Let us now return to the convex function $y = f(x)$. Suppose that M_1, M_2, . . ., M_k are k sequential points on the graph of this function taken within the interval from a to b (Figure 7.4.7). Due to the convexity of the function, the k-gon $M_1M_2 \ldots M_k$ is convex and lies entirely under the curve $y = f(x)$. If the abscissas of the points M_1, M_2, . . ., M_k are x_1, x_2, . . ., x_k, then their ordinates are, obviously, $f(x_1)$, $f(x_2)$, . . ., $f(x_k)$. Hence, the coordinates of the centroid Q of the k-gon $M_1M_2 \ldots M_k$ are $\left(\frac{x_1+x_2+\ldots+x_k}{k}, \frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k}\right)$ with the result that

$$OS = \frac{x_1+x_2+\ldots+x_k}{k},$$

$$SQ = \frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k},$$

and

$$SP = f\left(\frac{x_1+x_2+\ldots+x_k}{k}\right)$$

(see Figure 7.4.7). But the centroid of a convex k-gon lies *inside* the k-gon (this follows from the very definition of a centroid); hence, point Q lies below point P and

$$\frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k} < f\left(\frac{x_1+x_2+\ldots+x_k}{k}\right),$$

which is what we set out to prove.

This line of reasoning remains valid when some (but not all) points M_1, M_2, . . ., M_k coincide (not all of the numbers x_1, x_2, . . ., x_k are distinct) and the k-gon degenerates into a polygon with a smaller number of vertices.

Examples.

(a) $y = \ln x$. From Theorem 3 it follows that

$$\frac{\ln x_1+\ln x_2+\ldots+\ln x_k}{k} < \ln\frac{x_1+x_2+\ldots+x_k}{k},$$

or $\sqrt[k]{x_1x_2 \ldots x_k} < \frac{x_1+x_2+\ldots+x_k}{k}$,

that is, *the geometric mean of k positive numbers some of which are distinct is smaller than the respective arithmetic mean.* This is known as the ***theorem on geometric and arithmetic means.***

(b) $y = -x^2$. In this case we get

$$-\frac{x_1^2+x_2^2+\ldots+x_k^2}{k} < -\left(\frac{x_1+x_2+\ldots+x_k}{k}\right)^2,$$

or

$$\sqrt{\frac{x_1^2+x_2^2+\ldots+x_k^2}{k}} > \frac{x_1+x_2+\ldots+x_k}{k},$$

that is, *the root-mean-square of k positive numbers some of which are distinct is greater than the respective arithmetic mean.*

(c) $y = -1/x$. In this case Theorem 3 yields

$$-\frac{1}{k}\left(\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_k}\right) < -\frac{k}{x_1+x_2+\ldots+x_k},$$

that is,

$$\frac{1}{k}\left(\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_k}\right) > \frac{k}{x_1+x_2+\ldots+x_k},$$

whence

$$\frac{k}{1/x_1+1/x_2+\ldots+1/x_k} < \frac{x_1+x_2+\ldots+x_k}{k},$$

that is, *the harmonic mean of k positive numbers some of which are distinct is smaller than the respective arithmetic mean.*

Finally, we will prove a theorem that generalizes Theorems 2 and 3.

Theorem 4 *Suppose that* $y = f(x)$ *is a function convex in an interval from a to b and* $x_1, x_2, \ldots, x_k$ *are k values of the independent variable some of which are distinct and taken within the interval and* $p_1, p_2, \ldots, p_k$ *are k positive numbers whose sum is equal to unity. Then*

$$p_1 f(x_1) + p_2 f(x_2) + \ldots + p_k f(x_k) < f(p_1 x_1 + p_2 x_2 + \ldots + p_k x_k) \qquad (7.4.6)$$

(*the general case of Jensen's inequality*).

For $k = 2$ Theorem 4 transforms into Theorem 2, while for $p_1 = p_2 = \ldots = p_k = 1/k$ it transforms into Theorem 3.

Proof. Let us once more take the graph of a convex function $y = f(x)$ and a convex k-gon $M_1 M_2 \ldots M_k$ inscribed into this graph. The vertices of the k-gon have the following coordinates: $(x_1, f(x_1))$, $(x_2, f(x_2))$, $\ldots$, $(x_k, f(x_k))$ (Figure 7.4.8). Suppose that Q_2 is a point of the side M_1M_2 of this k-gon such that $M_1Q_2 \div Q_2M_2 = \frac{p_2}{p_1+p_2} \div \frac{p_1}{p_1+p_2}$, Q_3 is a point of the line segment M_3Q_2 such that

$$M_3Q_3 \div Q_3Q_2 = \frac{p_3}{p_1+p_2+p_3} \div \frac{p_1+p_2}{p_1+p_2+p_3},$$

Q_4 is a point of the line segment M_4Q_3 such that

$$M_4Q_4 \div Q_4Q_3 = \frac{p_4}{p_1+p_2+p_3+p_4} \div \frac{p_1+p_2+p_3}{p_1+p_2+p_3+p_4}, \ \ldots,$$

finally, $Q_k = Q$ is a point of the segment M_kQ_{k-1} such that $M_kQ \div QQ_{k-1} = p_k \div (p_1+p_2+\ldots+p_{k-1})$ (if $p_1 = p_2 = \ldots = p_k = 1/k$, then Q is the centroid of the k-gon $M_1M_2 \ldots M_k$). Using the proposition from which we started the proof of Theorem 2, we can find the coordinates of the points $Q_2, Q_3, Q_4, \ldots, Q$:

$$Q_2\left(\frac{p_1x_1+p_2x_2}{p_1+p_2}, \frac{p_1f(x_1)+p_2f(x_2)}{p_1+p_2}\right);$$

$$Q_3\left(\frac{p_1+p_2}{p_1+p_2+p_3}\frac{p_1x_1+p_2x_2}{p_1+p_2} + \frac{p_3}{p_1+p_2+p_3}x_3, \frac{p_1+p_2}{p_1+p_2+p_3}\frac{p_1f(x_1)+p_2f(x_2)}{p_1+p_2} + \frac{p_3}{p_1+p_2+p_3}f(x_3)\right),$$

or

$$Q_3\left(\frac{1}{p_1+p_2+p_3}(p_1x_1+p_2x_2+p_3x_3), \frac{1}{p_1+p_2+p_3}[p_1f(x_1)+p_2f(x_2)+p_3f(x_3)]\right);$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Q\left(\frac{p_1x_1+p_2x_2+\ldots+p_{k-1}x_{k-1}+p_kx_k}{p_1+p_2+\ldots+p_{k-1}+p_k}, \frac{p_1f(x_1)+p_2f(x_2)+\ldots+p_{k-1}f(x_{k-1})+p_kf(x_k)}{p_1+p_2+\ldots+p_{k-1}+p_k}\right)$$

or, in another form,

$$Q(p_1x_1+p_2x_2+\ldots+p_kx_k, \ p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k)),$$

since $p_1+p_2+\ldots+p_k = 1$. Thus, in Figure 7.4.8,

$$SQ = p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k),$$

$$OS = p_1x_1+p_2x_2+\ldots+p_kx_k,$$

$$SP = f(p_1x_1+p_2x_2+\ldots+p_kx_k).$$

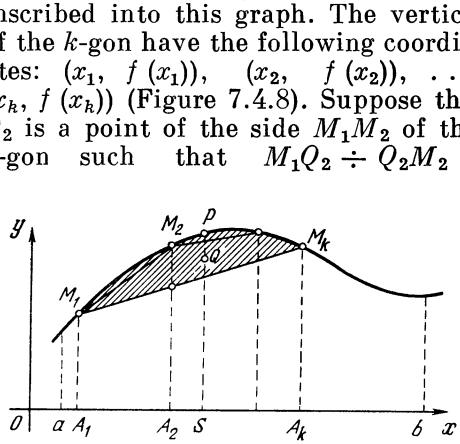


Figure 7.4.8

Since point Q lies below point P (the entire k-gon $M_1M_2 \ldots M_k$ lies *under* the curve $y = f(x)$, and Q is an interior point of this k-gon), we have

$$p_1 f(x_1) + p_2 f(x_2) + \ldots + p_k f(x_k) < f(p_1x_1 + p_2x_2 + \ldots + p_kx_k),$$

which is what we set out to prove.[7.14]

Examples.

(a) $y = \ln x$. In this case we get

$$p_1 \ln x_1 + p_2 \ln x_2 + \ldots + p_k \ln x_k < \ln(p_1x_1 + p_2x_2 + \ldots + p_kx_k),$$

whence, taking antilogs, we have

$$x_1^{p_1}x_2^{p_2} \ldots x_k^{p_k} < p_1x_1 + p_2x_2 + \ldots + p_kx_k,$$

where $p_1, p_2, \ldots, p_k \geqslant 0$ and $p_1 + p_2 + \ldots + p_k = 1$. (This is *the generalized theorem on geometric and arithmetic means.*)

(b) $y = -x^2$. We have

$$-p_1x_1^2 - p_2x_2^2 - \ldots - p_kx_k^2 < -(p_1x_1 + p_2x_2 + \ldots + p_kx_k)^2,$$

or

$$\sqrt{p_1x_1^2 + p_2x_2^2 + \ldots + p_kx_k^2} > p_1x_1 + p_2x_2 + \ldots + p_kx_k,$$

where $p_1, p_2, \ldots, p_k \geqslant 0$ and $p_1 + p_2 + \ldots + p_k = 1$. (This is *the generalized theorem on the root-mean-square and the arithmetic mean.*)

(c) $y = -1/x$. Here Theorem 4 yields

$$-\frac{p_1}{x_1} - \frac{p_2}{x_2} - \ldots - \frac{p_k}{x_k} < -\frac{1}{p_1x_1 + p_2x_2 + \ldots + p_kx_k},$$

whence we can easily obtain

$$\frac{1}{p_1/x_1 + p_2/x_2 + \ldots + p_k/x_k} < p_1x_1 + p_2x_2 + \ldots + p_kx_k,$$

where $p_1, p_2, \ldots, p_k \geqslant 0$ and $p_1 + p_2 + \ldots + p_k = 1$.

[7.14] The reader can easily see that the coordinates of any interior point of the k-gon $M_1M_2 \ldots M_k$ can be represented in the form $(p_1x_1 + p_2x_2 + \ldots + p_kx_k, p_1f(x_1) + p_2f(x_2) + \ldots + p_kf(x_k))$, where $p_1, p_2, \ldots, p_k$ are positive and $p_1 + p_2 \ldots + p_k = 1$. Thus, inequality (7.4.6) expresses the fact that a polygon inscribed in the graph of a convex function always lies below that graph.

Exercises

7.4.1. Prove that the following functions are convex:

(a) $y = -x^n$ for $n > 1$ and $x > 0$, (b) $y = x^m$ for $0 < m < 1$ and $x > 0$, (c) $y = -1/x^k$ for $k > 0$ and $x > 0$, (d) $y = -x \log x$ for $x > 0$, (e) $y = -x \log x - (1 - x) \log (1 - x)$ for $0 < x < 1$ (the base in (d) and (e) is any number greater than unity).

7.4.2. Write out the inequalities (7.4.1), (7.4.4), (7.4.5), and (7.4.6), where the function $f(x)$ is (a) the function of Exercise 7.4.1a, (b) the function of Exercise 7.4.1b, (c) the function of Exercise 7.4.1c, (d) the function of Exercise 7.4.1d.

7.4.3. (a) Prove that the geometric mean of two (positive) numbers is the geometric mean of their arithmetic and harmonic means. (b) Derive inequality (7.4.3) using the result of (a) and inequality (7.4.2).

7.5 Computing Areas

In Chapter 3 we showed that the value of a definite integral $\int_a^b f(x)\,dx$ yields the *area* of a figure bounded from above by the curve $y = f(x)$, from below by the x axis, and on the sides by the vertical lines $x = a$ and $x = b$, or the vertical bases of the trapezoid (see Figure 7.5.1; for the sake of simplicity we assume that $f(x) > 0$ and $a < b$). Thus, being able to evaluate definite

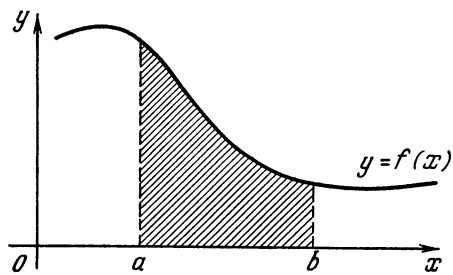


Figure 7.5.1

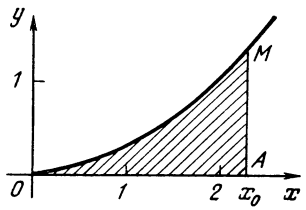


Figure 7.5.2

integrals enables us to use standard techniques in computing various areas, whereas elementary mathematics only allows for calculating the areas of rectilinear figures (polygons) and also of the circle and some of its parts (segment, sector).

Let us find the area of a figure bounded from above by the curve of the *power function* $y = c\,x^n$ ($c > 0$ and $n > 0$), from below by the x axis, and on the right by the vertical line $x = x_0$ (we assume that $x_0 > 0$; in Figure 7.5.2, $n = 2$ and $c = 0.25$):

$$S = \int_0^{x_0} cx^n\,dx = \frac{cx^{n+1}}{n+1}\Big|_0^{x_0} = \frac{cx_0^{n+1}}{n+1}. \quad (7.5.1)$$

Let us rewrite formula (7.5.1) as

$$S = \frac{1}{n+1}\,cx_0^n x_0,$$

or, since $cx_0^n = y(x_0) = y_0$,

$$S = \frac{1}{n+1}\,y_0 x_0. \quad (7.5.2)$$

Since the quantities y and x have the dimensions of length, from (7.5.2) it follows that S is indeed measured in units of area (the square of the unit of length). (We see that the area S is, as to order of magnitude, y_0x_0, and differs from this product solely in the factor $1/(n+1)$, which, as to order of magnitude, is close to unity for n not too large (compare with (5.6.6), where $y_0 = f_{\max}$, $x_0 = b - a$, and $1/(n+1)$ is a dimensionless factor of the order of unity).

In the next example, we find they are bounded from above by the *exponential curve*

$$y = ce^{-x/a}, \quad c > 0, \quad a > 0, \quad (7.5.3)$$

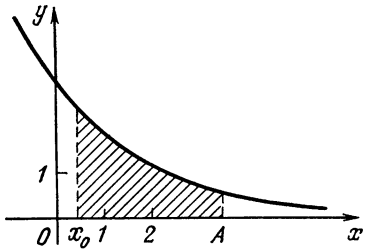


Figure 7.5.3

from below by the x axis, and on the left and right by the straight lines $x = x_0$ and $x = A$ (here $A > x_0$; Figure 7.5.3). This area is

$$S_A = \int_{x_0}^{A} ce^{-x/a}\,dx = -cae^{-x/a}\Big|_{x_0}^{A}$$
$$= ca\,(e^{-x_0/a} - e^{-A/a}). \quad (7.5.4)$$

If A is great compared with x_0, then $e^{-x_0/a} \gg e^{A/a}$. It will be seen from (7.5.4) that increasing A hardly at all changes S_A. As A increases without bound, the value of $e^{A/a}$ approaches zero without bound. And so we can speak of the area of the figure in Figure 7.5.3 as being unbounded on the right: the *unlimited* figure obtained from the figure in Figure 7.5.3 as $A \to \infty$ has a *finite* area

$$S_\infty = \int_{x_0}^{\infty} ce^{-x/a}\,dx = cae^{-x_0/a} = ay_0, \quad (7.5.5)$$

where $y_0 = y(x_0) = ce^{-x_0/a}$.

In formula (7.5.3), the exponent must be a dimensionless number, which means that the dimensions of a must coincide with those of x, that is, a has the dimensions of length. The dimensions of y and S are those of length and area respectively.

It turns out that the area under one arch of a *sine curve* (Figure 7.5.4) is

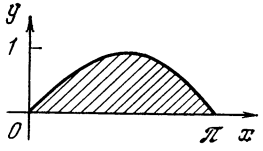


Figure 7.5.4

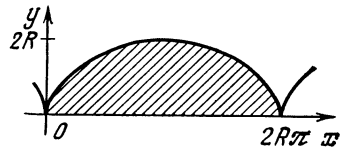


Figure 7.5.5

expressed very simply. Indeed, this area, bounded from below by the line segment of the x axis from 0 to π, is

$$S = \int_0^{\pi} \sin x\, dx = -\cos x \,|_0^{\pi} = 2. \quad (7.5.6)$$

It is not difficult to find the area under one arch of a *cycloid* (Figure 7.5.5; see also Section 1.8). If the curve is given parametrically, $x = \varphi(t)$ and $y = \psi(t)$ (see Section 1.8), the formula for the area of a curvilinear trapezoid assumes the form

$$S = \int_a^b y\, dx = \int_{\alpha}^{\beta} y(t) \frac{dx}{dt}\, dt,$$

where $a = x(\alpha)$ and $b = x(\beta)$ are the endpoints of the segment considered. At $x = R(t - \sin t)$ and $y = R(1 - \cos t)$, with $\alpha = 0$ and $\beta = 2\pi$, we have

$$\begin{aligned} S &= \int_0^{2\pi} R(1-\cos t)\, R(1-\cos t)\, dt \\ &= R^2 \int_0^{2\pi} (1-\cos t)^2\, dt \\ &= R^2 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t)\, dt \\ &= R^2 \int_0^{2\pi} dt - 2R^2 \int_0^{2\pi} \cos t\, dt \\ &\quad + R^2 \int_0^{2\pi} \frac{1+\cos 2t}{2}\, dt = R^2 t \,|_0^{2\pi} \\ &\quad - 2R^2 \sin t \,|_0^{2\pi} + \frac{R^2}{2}\left(t + \frac{1}{2}\sin 2t\right)\Big|_0^{2\pi} \\ &= R^2 \times 2\pi - 2R^2 \times 0 + \frac{R^2}{2} \times 2\pi \\ &\quad + \frac{R^2}{4} \times 0 = 3\pi R^2, \end{aligned} \quad (7.5.7)$$

which means that the area under one arch of a cycloid is three times the area of the circle that generates the cycloid.

The results (7.5.2) and (7.5.5)-(7.5.7) can also be formulated in a different manner. The area of the hatched curvilinear triangle OAM in Figure 7.5.2 is $(n+1)^{-1} OA \cdot AM = OA\,[(n+1)^{-1} AM]$, that is, coincides with the area of the rectangle with base CA and altitude $(n+1)^{-1} AM$. In view of this it is sometimes said that the ***effective altitude*** of our curvilinear traingle (or the curve $y = cx^n$ within the limits $x = 0$ to $x = x_0$) is one-$(n+1)$th of the real altitude AM. Here the effective altitude is understood to be the altitude of the rectangle with the same base as that of the curvilinear triangle and the same area. Similarly, the *effective altitudes* of the curvilinear trapezoids depicted in Figures 7.5.4 and 7.5.5 with bases π and $2R\pi$, respectively, are $2/\pi$ ($\simeq 0.6366$) and $1.5R$, while the real altitudes, or the values of $y_{\max}$, are 1 and $2R$, respectively. Finally, the ***effective length*** of the *infinite* figure depicted in Figure 7.5.3, a figure bounded by the exponential curve $y = ce^{-x/a}$, the x axis, and the straight line $x = x_0$, is equal to a, since the area of this figure is that of the rectangle with base a and altitude y_0.

The concept of effective length (or effective altitude) of a curve often proves to be useful. For instance, it can be shown that the (infinitely long) bell-shaped curve in Figure 7.5.6, $y = c \exp(-x^2/a^2)$ restricts a finite area $S = ca\sqrt{\pi}$; in other words, the ***effective width*** of this curve, whose altitude is $y_{\max} = y(0) = c$, is equal to $a\sqrt{\pi} \simeq 1.77a$ (see [15], Section 3.2).

Let us determine the area S of an ellipse with semiaxes a and b (Figure 7.5.7). Of course, this area can easily be found by methods of elementary geometry (see the text below printed in small type); however, we prefer using a more standard method (although in the given case a more cumbersome one)

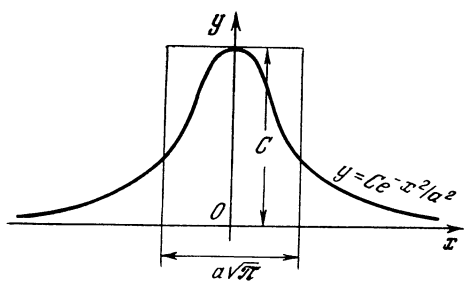


Figure 7.5.6

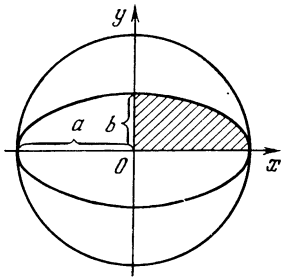


Figure 7.5.7

involving integral calculus. Note that by virtue of symmetry it suffices to find the area S_1 of that portion which lies in the first quadrant and then multiply it by four: $S = 4S_1$. To compute S_1 we find y from the equation of the ellipse $x^2/a^2 + y^2/b^2 = 1$ (see Section 1.7): $y = (b/a)\sqrt{a^2 - x^2}$ (here the square root is understood to be the positive root, since in the first quadrant y is positive). Thus,

$$S_1 = \frac{b}{a}\int_0^a \sqrt{a^2 - x^2}\,dx. \qquad (7.5.8)$$

The integral (7.5.8) can easily be found by making the change of variable $x = a \sin t$. This yields

$$\int_0^a \sqrt{a^2 - x^2}\,dx = \int_0^{\pi/2} a\sqrt{1 - \sin^2 t}\, a\cos t\,dt$$

$$= \int_0^{\pi/2} a^2 \cos^2 t\,dt = a^2 \int_0^{\pi/2} \frac{1 + \cos 2t}{2}\,dt$$

$$= a^2\left[\frac{t}{2} + \frac{\sin 2t}{4}\right]_0^{\pi/2} = \frac{\pi a^2}{4}. \qquad (7.5.9)$$

Using this, from (7.5.8) we get

$$S_1 = \frac{b}{a}\cdot\frac{\pi a^2}{4} = \frac{\pi ab}{4}.$$

The area of the entire ellipse is $S = \pi ab$. If $a = b = r$, then we have $S = \pi r^2$ (the area of a circle) in complete accord with the fact that for $a = b = r$ an ellipse becomes a circle.

An ellipse with semiaxes a and b ($a > b$) is obtained from a circle of radius a by shrinking the latter to the x axis with a ratio of $k = b/a$ (see Figure 7.5.7). It is easy to see that such shrinking transforms a figure F of area s into a figure F' of area $s' = ks$. Indeed, a grid of small squares with side δ and area δ^2 (the sides of the squares being parallel to the coordinate axes) is transformed by the shrinking transformation into a grid of rectangles with sides δ and $k\delta$ and area $k\delta^2$ each. But a grid of equal squares (a measuring grid) is used to measure areas: if a figure F is covered by N squares of the grid, its area s is approximately $N\delta^2$. On the other hand, the figure F', obtained through shrinking figure F in the above-described manner, will be covered by N rectangles of the transformed grid; therefore, $s' \simeq Nk\delta^2$, whereby $s' = ks$ (since δ can be chosen as small as desired, so that the approximate equalities $s \simeq N\delta^2$ and $s' \simeq Nk\delta^2$ can be assumed as precise as desired). And since the area of a circle of radius a is equal to πa^2, the area of the ellipse with semiaxes a and b (the ellipse being obtained by shrinking the circle with a ratio $k = b/a$) is $k\pi a^2 = (b/a)\,\pi a^2 = \pi ab$.

Note an important circumstance. In Chapter 3 we already pointed out that the area (the integral) can be either positive or negative. This calls for a certain amount of care when finding areas. Suppose we want to know the amount of paint needed to paint an area bounded by two arches of a sine curve, from $x = 0$ to $x = 2\pi$, and the x axis (see Figure 3.5.2) if unit area requires a grams of paint. As was shown above, one integral cannot be used to compute the entire area. We have to take separate integrals over the intervals from $x = 0$ to $x = \pi$ and from $x = \pi$ to $x = 2\pi$.

Generally, if the integrand $y = f(x)$ *changes sign*, then to solve the problem in paint consumption, so to say, we must split the interval of integration into parts in which $f(x)$ preserves sign, then evaluate the integral over the separate parts, and finally sum the absolute values of the resulting integrals.

Exercises

7.5.1. Find the area of a figure bounded by the x axis and a single arch of (a) $y = \sin^2 x$ and (b) $y = \cos^2 x$. [*Hint.* Draw the graphs

of both functions and take advantage of the formulas $\sin^2 x = 1/2 - (1/2)\cos 2x$ and $\cos^2 x = 1 - \sin^2 x$.]

7.5.2. Find the area of a figure bounded from above by the curve $y = x(1 - x)$ and from below by the x axis.

7.5.3. Find the areas into which the parabola $y = (1/2)x^2$ divides the circle $x^2 + y^2 = 8$.

7.5.4. Find the amount of paint needed to cover the area of a figure bounded by (a) the curve $y = x/(1 + x^2)$, the x axis, and the vertical lines $x = 1$ and $x = -1$, and (b) the curve $y = x^3 + 2x^2 - x - 2$ and the x axis. [*Hint.* First construct the graph of the function $y = x^3 + 2x^2 - x - 2$.]

7.5.5. Find the area of the ellipse $x^2/25 + y^2/4 = 1$.

7.6* Estimating Sums and Products

In Sections 3.1 and 3.2 we introduced the concept of a (definite) integral as the limit of a sequence of certain sums. By calculating these sums we can estimate the integral, or the area of the appropriate curvilinear trapezoid (see the *method of rectangles* and the *trapezoid method* described in Section 3.1). But the relationship that exists between integrals (or areas) and (finite) sums can be exploited in the opposite direction, for estimating sums by means of integrals. This simple idea (see also [15], Section 1.2) will be illustrated by several instructive examples.

We start with the *sum of powers of natural numbers*:

$$S_n^{(k)} = 1^k + 2^k + 3^k + \ldots + n^k, \qquad (7.6.1)$$

where $k > -1$. Let us consider the function $y = x^k$; Figure 7.6.1 depicts the case with k positive, while the reader is advised to consider the case

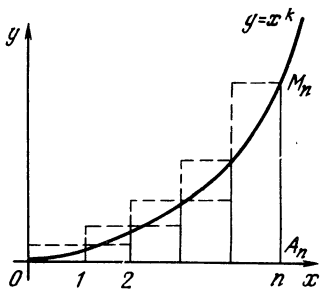


Figure 7.6.1

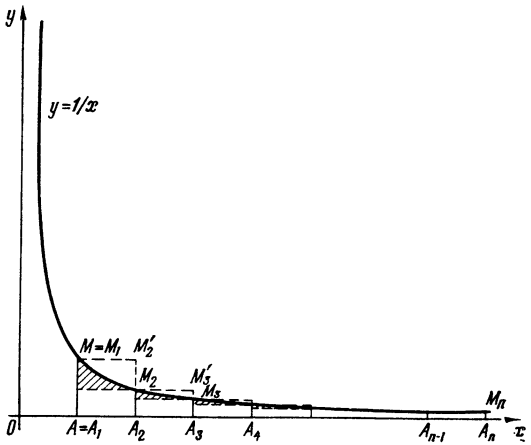


Figure 7.6.2

with $0 > k > -1$ as an exercise.[7.15] The area s of the curvilinear triangle OA_nM_n depicted in Figure 7.6.1 is, obviously,

$$s = \int_0^n x^k\,dx = \frac{x^{k+1}}{k+1}\Big|_0^n = \frac{n^{k+1}}{k+1}. \qquad (7.6.2)$$

On the other hand, the method of rectangles yields

$$1^k + 2^k + \ldots + n^k > s$$

and $0^k + 1^k + \ldots + (n-1)^k < s$,

that is,

$$1^k + 2^k + \ldots + n^k > \frac{n^{k+1}}{x+1} > 1^k + 2^k + \ldots + (n-1)^k,$$

or

$$\frac{n^{k+1}}{k+1} < S_n^{(k)} = 1^k + 2^k + \ldots + n^k < \frac{n^{k+1}}{k+1} + n^k. \qquad (7.6.3)$$

This implies that for large values of n we have

$$S_n^{(k)} = 1^k + 2^k + \ldots + n^k \simeq \frac{n^{k+1}}{k+1}, \qquad (7.6.4)$$

[7.15] For $k < 0$ we must consider not the curvilinear triangle OA_nM_n but the curvilinear trapezoid AA_nM_nM similar to the one depicted in Figure 7.6.2.

in the sense that for $n \gg 1$ the ratio $S_n^{(k)} \Big/ \frac{n^{k+1}}{k+1}$ differs but little from unity, since

$$1 \leqslant S_n^{(k)} \Big/ \frac{n^{k+1}}{k+1} < 1 + \frac{k+1}{n} \qquad (7.6.3a)$$

(these inequalities are obtained if we divide all members in (7.6.3) by $n^{k+1}/(k+1)$).

From (7.6.3) follows a more precise estimate of the sum $S_n^{(k)}$ than (7.6.4):

$$S_n^{(k)} = 1^k + 2^k + \ldots + n^k \simeq \frac{1}{k+1} n^{k+1} + cn^k,$$

with $0 < c < 1$. Indeed, it can be proved that for $k > 0$ we have

$$S_n^{(k)} \simeq \frac{1}{k+1} n^{k+1} + \frac{1}{2} n^k \qquad (7.6.4a)$$

in the sense that for large values of n the ratio $\left[S_n^{(k)} - \left(\frac{1}{k+1} n^{k+1} + \frac{1}{2} n^k\right)\right] \Big/ n^k$ is very small.

The behavior of the sum

$$S_n^{(-1)} = S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \qquad (7.6.5)$$

for large values of n is quite different. The graph of the function $y = 1/x$, to which we naturally turn, is the *hyperbola* (Figure 7.6.2). The area of the curvilinear trapezoid AA_nM_nM bounded by the x axis, the hyperbola $y = 1/x$, and the straight lines $x = 1$ and $x = n$, is given by the following formula:

$$\sigma = \int_1^n \frac{dx}{x} = \ln x \,|_1^n = \ln n, \qquad (7.6.6)$$

since $\ln 1 = 0$. On the other hand, by the method of rectangles (see Figure 7.6.2),

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} > \sigma > \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}, \qquad (7.6.7)$$

that is (compare with (7.6.5)), $\sigma < S_n - 1/n$ and $\sigma > S_n - 1$, whence

$$\sigma + \frac{1}{n} = \ln n + \frac{1}{n} < S_n < \ln n + 1 = \sigma + 1. \qquad (7.6.8)$$

Thus,

$$S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} = \ln n + \gamma_n, \qquad (7.6.9)$$

where the number γ_n lies within the interval

$$\frac{1}{n} < \gamma_n < 1. \qquad (7.6.9a)$$

Since, as it can easily be seen, γ_n is the difference between the area of the square $AA_2M_2'M_1$ (equal to unity) and the sum of the areas hatched in Figure 7.6.2, it monotonically decreases as $n \to \infty$ and tends to a limit that can be estimated if we find (approximately) the hatched area in Figure 7.6.2 for a fixed value of n that is not too large; this limit, $\gamma \simeq 0.577$, is known as ***Euler's constant***. Thus, for $n \gg 1$, we have the following approximate formula:

$$S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \simeq \ln n + \gamma \simeq \ln n + 0.577. \qquad (7.6.10)$$

The next example is the sum

$$T_n = \frac{\ln 2}{2} + \frac{\ln 3}{3} + \ldots + \frac{\ln n}{n}. \qquad (7.6.11)$$

The experience we have gained in the process of studying the sums (7.6.1) and (7.6.5) tells us that in this case it is expedient to start by estimating the area τ of the curvilinear triangle bounded by the curve $y = (\ln x)/x$, the axis of abscissas, and the straight line $x = n$ (Figure 7.6.3); we compare τ

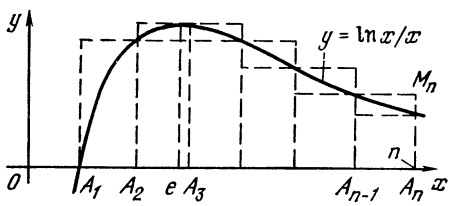


Figure 7.6.3

with the sum T_n. The area τ is given by the formula

$$\tau = \int_1^n \frac{\ln x}{x}\,dx = \int_1^n \ln x\,d(\ln x)$$
$$= \int_0^{\ln n} u\,du = \frac{u^2}{2}\Big|_0^{\ln n} = \frac{1}{2}(\ln n)^2 \quad (7.6.12)$$

(here we have introduced a new variable, $u = \ln x$).

Next we use a modified version of the method of rectangles. The area τ is replaced with the sum of areas of $n-1$ rectangles with bases A_1A_2, A_2A_3, . . ., $A_{n-1}A_n$ (see Figure 7.6.3), while the altitude of rectangle with the base $A_{i-1}A_i$ is in one case the greatest value of the function $y = (\ln x)/x$ attained in the interval $i-1 \leqslant x \leqslant i$ and in the other is the smallest value of the same function attained in the same interval. All this requires a detailed study of the curve $y = (\ln x)/x$. Obviously, $y(1) = 0$ and $y \to 0$ as $x \to \infty$ (cf. Section 6.5); let us find the maximum of the function $y(x)$. Since the equation

$$y'(x) = \frac{(\ln x)'}{x} + \ln x\left(\frac{1}{x}\right)'$$
$$= \frac{1}{x^2} - \frac{\ln x}{x^2} = \frac{1-\ln x}{x^2} = 0$$

has a unique solution, $x = e$ ($\ln x = 1$ only at $x = e$), and $y'(x) > 0$ for $x < e$ and $y'(x) < 0$ for $x > e$, the graph of the function $y = (\ln x)/x$ has the shape depicted in Figure 7.6.3.[7.16] At $x = e$ the function has its only maximum ($y_{\max} = e^{-1} \simeq 1/2.7$), from $x = 1$ to $x = e$ the function grows, and then decreases monotonically. Thus, the maximum of the function $y(x)$ on the interval from 2 to 3 is equal to $1/e$, while on all other intervals from $i-1$ to i, with i an integer no less than two, the maximum coincides with the value of the function at one of the endpoints of the specified interval. We arrive at the following estimate for the area τ:

$$\tau < \frac{\ln 2}{2} + \frac{1}{e} + \frac{\ln 3}{3} + \frac{\ln 4}{4}$$
$$+ \ldots + \frac{\ln(n-1)}{n-1}\left(= T_n + \frac{1}{e} - \frac{\ln n}{n}\right),$$

and

$$\tau > \frac{\ln 1}{1} + \frac{\ln 2}{2} + \frac{\ln 4}{4} + \ldots + \frac{\ln n}{n}$$
$$\left(= T_n - \frac{\ln 3}{3}\right),$$

that is,

$$\tau - \frac{1}{e} + \frac{\ln n}{n} < T_n < \tau + \frac{\ln 3}{3}.$$

Finally we get (see (7.6.12))

$$\frac{1}{2}(\ln n)^2 - \frac{1}{e} + \frac{\ln n}{n} < T_n$$
$$< \frac{1}{2}(\ln n)^2 + \frac{\ln 3}{3}; \quad (7.6.13)$$

in other words, T_n cannot differ considerably from $(1/2)(\ln n)^2$: for all values of n we have

$$T_n = \frac{1}{2}(\ln n)^2 + \delta_n,$$

where, in any case, $-0.368 \simeq -e^{-1} < \delta_n < (\ln 3)/3 \simeq 0.367$. Here, since the difference between τ and its lower bound determined by the method of rectangles, that is,

$$\tau - \left(T_n - \frac{\ln 3}{3}\right)\left(= \frac{\ln 3}{3} - \delta_n\right),$$

can only increase as n grows, $(\ln 3)/3 - \delta_n$ monotonically grows as $n \to \infty$, but always remains smaller than $(\ln 3)/3 - e^{-1} \simeq 0.735$. From this it follows that $(\ln 3)/3 - \delta_n$ tends to a certain limit as $n \to \infty$, which means that δ_n also tends to a certain limit as $n \to \infty$ (the limit lies between -0.37 and $+0.37$).

This method can be applied for estimating the value of $n! = 1 \cdot 2 \cdot 3 \ldots n$. We write the number $\ln(n!) = \ln(1 \cdot 2 \cdot 3 \ldots n)$ as

$$\ln(n!) = \ln 1 + \ln 2 + \ln 3$$
$$+ \ldots + \ln n. \quad (7.6.14)$$

[7.16] The fact that the value $x = e$ at which $y'(x) = 0$ corresponds precisely to the *maximum* of $y = (\ln x)/x$ can easily be established without analyzing the sign of $y'(x)$ for different values of x, that is, simply by common-sense reasoning (compare with what was said at the end of Example 2 in Section 7.1).

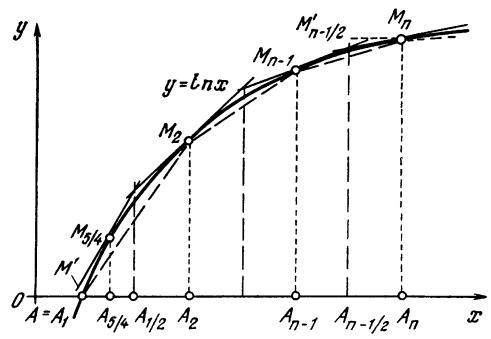


Figure 7.6.4

To estimate the sum on the right-hand side of (7.6.14), we consider the curvilinear triangle AA_nM_n bounded by the x axis, the curve $y = \ln x$, and the straight line $x = n$ (n is a natural number greater than unity) (Figure 7.6.4). The area σ_n of this triangle is, obviously,

$$\sigma_n = \int_1^n \ln x\,dx = x\ln x\,|_1^n - \int_1^n x\,d(\ln x)$$

$$= x\ln x\,|_1^n - \int_1^n x\frac{dx}{x} = x\ln x\,|_1^n - \int_1^n dx$$

$$= (x\ln x - x)\,|_1^n = n\ln n - n + 1$$

(see Section 5.4). If we inscribe in our curvilinear triangle $A_1A_nM_n$ the right triangle $A_1A_2M_2$ and $n-2$ trapezoids $A_2A_3M_3M_2$, $A_3A_4M_4M_3$, ... , $A_{n-1}A_nM_nM_{n-1}$ (here $A_1 = A, A_2, \ldots, A_n$ are points on the x axis with abscissas 1, 2, . . ., n, and $M_2, \ldots, M_n$ are the corresponding points on the curve $y = \ln x$), we will find that σ_n is greater than the sum of the areas of the right triangle and the trapezoids, the sum being equal to

$$s_n = \frac{1}{2}\ln 2 + \frac{1}{2}(\ln 2 + \ln 3)$$

$$+ \frac{1}{2}(\ln 3 + \ln 4)$$

$$+ \ldots + \frac{1}{2}[\ln(n-1) + \ln n]$$

$$= \ln 2 + \ln 3 + \ldots + \ln(n-1) + \frac{1}{2}\ln n$$

$$= (\ln 1 + \ln 2 + \ln 3 + \ldots + \ln n)$$

$$- \frac{1}{2}\ln n = \ln(n!) - \frac{1}{2}\ln n. \quad (7.6.15a)$$

On the other hand, the area σ_n of the curvilinear triangle is smaller than the sum S_n of the areas of $n-2$ trapezoids with midlines A_iM_i, $i = 2, 3, \ldots, n-1$, and altitudes of unit length (the altitudes are the line segments with endpoints $(i - 1/2, 0)$ and $(i + 1/2, 0)$, with i running through the same values; the upper lateral side is a segment of the line tangent to the curve $y = \ln x$ at point M_i) plus the area of the small trapezoid with the altitude $AA_{1/2}$ of length 1/2 and the midline $A_{5/4}M_{5/4}$ (here $A_{5/4} = (5/4, 0)$ and $M_{5/4}$ is the point on the curve $y = \ln x$ corresponding to $A_{5/4}$) plus the area of the rectangle $A_{n-1/2}A_nM_nM'_{n-1/2}$ (here $A_{n-1/2} = (n - 1/2, 0)$ and $M'_{n-1/2} = (n - 1/2, \ln n)$). It is clear that

$$S_n = \frac{1}{2}\ln\frac{5}{4} + 1\cdot\ln 2$$

$$+ 1\cdot\ln 3 + \ldots + 1\cdot\ln(n-1) + \frac{1}{2}\ln n$$

$$= \ln 2 + \ln 3 + \ldots + \ln(n-1)$$

$$+ \ln n + \frac{1}{2}\ln\frac{5}{4} - \frac{1}{2}\ln n$$

$$= \ln(n!) - \frac{1}{2}\ln n + \frac{1}{2}\ln\frac{5}{4}. \quad (7.6.15b)$$

Since $s_n < \sigma_n \leqslant S_n$, we can write

$$\ln(n!) - \frac{1}{2}\ln n < n\ln n - n + 1$$

$$< \ln(n!) - \frac{1}{2}\ln n + \frac{1}{2}\ln\frac{5}{4},$$

or

$$\ln(n!) > n\ln n + \frac{1}{2}\ln n - n + 1$$

$$- \frac{1}{2}\ln\frac{5}{4}$$

and $\ln(n!) < n\ln n + \frac{1}{2}\ln n - n + 1,$

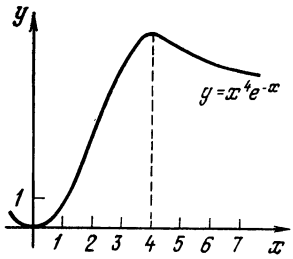


Figure 7.6.5

that is

$$\left(1-\frac{1}{2}\ln\frac{5}{4}\right)+\ln\left(n^n\sqrt{n}\,e^{-n}\right)<\ln(n!)$$
$$<1+\ln\left(n^n\sqrt{n}\,e^{-n}\right).$$

In other words,

$$\sqrt{\frac{4}{5}}\,en^n\sqrt{n}\,e^{-n}<n!<en^n\sqrt{n}\,e^{-n}. \tag{7.6.16}$$

This formula shows that for large n the value of $n!$ is close to $C\sqrt{n}n^ne^{-n}$, where the number C lies between $e\simeq 2.72$ and $\sqrt{4/5}\,e\simeq 2.43$. Here, from the fact that the difference $\sigma_n - s_n$ increases monotonically as $n\to\infty$, we can easily derive, as we did earlier, that $n!/\sqrt{n}n^ne^{-n}$ tends to a certain limit as $n\to\infty$ (this limit lies between $\sqrt{4/5e}$ and e, that is, between 2.43 and 2.72). It can be proved that this limit is equal to $\sqrt{2\pi}\simeq 2.507$. Thus, we arrive at ***Stirling's formula***[7.17]

$$n!\simeq\sqrt{2\pi n}\,n^ne^{-n}, \tag{7.6.17}$$

in which the approximate equality means that, as $n\to\infty$, the ratio $n!/\sqrt{2\pi n}n^ne^{-n}$ tends to unity (and for n large it is very close to unity).

Stirling's formula (7.6.17) can be substantiated in another manner. Let us evaluate the integral

$$I_n=\int_0^\infty x^ne^{-x}\,dx, \tag{7.6.18}$$

where n is a positive integer. The integral I_n can be thought of as the area of an (unlimited) figure bounded from above by the curve $y=x^ne^{-x}$ and from below by the x axis for $x>0$ (Figure 7.6.5; $n=4$).[7.18] To evaluate the integral, we use integration by parts setting $e^{-x}\,dx=dg$ and $x^n=f$, which means that $g=-e^{-x}$ and $df=nx^{n-1}\,dx$. Thus, we obtain

$$\int_0^\infty x^ne^{-x}\,dx=(-x^ne^{-x})\,|_0^\infty+\int_0^\infty nx^{n-1}e^{-x}\,dx.$$

In Section 6.5 it was established that $x^ne^{-x}=x^n/e^x\to 0$ as $x\to\infty$. Since $x^ne^{-x}=0$ for $x=0$, it follows that $-x^ne^{-x}\,|_0^\infty=0$; hence,

$$\int_0^\infty x^ne^{-x}\,dx=n\int_0^\infty x^{n-1}e^{-x}\,dx,\quad\text{i.e.}\quad I_n=nI_{n-1}.$$

Applying integration by parts to I_{n-1}, we likewise get $I_{n-1}=(n-1)I_{n-2}$, and so forth. For this reason

$$I_n=n(n-1)(n-2)\ldots 3\cdot 2\cdot 1\cdot I_0,$$

where $I_0=\int_0^\infty e^{-x}\,dx=-e^{-x}\,|_0^\infty=-0+1=1$.

Thus,

$$I_n=\int_0^\infty x^ne^{-x}\,dx$$
$$=n(n-1)(n-2)\ldots 3\cdot 2\cdot 1=n!. \tag{7.6.19}$$

For large values of n we can arrive at a good approximation for I_n if we find the value $x=x_{\max}$ at which the integrand x^ne^{-x} attains its maximum value. This value is found from the condition

$$(x^ne^{-x})'=nx^{n-1}e^{-x}+x^n(-e^{-x})$$
$$=(n-x)x^{n-1}e^{-x}=0,$$

where $x_{\max}=n$. The magnitude of this maximum, obviously, is $y_{\max}=(x^ne^{-x})_{\max}=n^ne^{-n}$. If we write our function in the form $y=e^{f(x)}$, where $f(x)=n\ln x-x$, and expand $f(x)$ in a Taylor's series (6.1.18) near the value $x=x_{\max}\,(=n)$, we can estimate the *effective width* of our (unlimited) figure, that is, the width of a rectangle whose height is $y_{\max}$ and whose area is that of the figure considered, or I_n. This width proves to be approximately equal to $\sqrt{2\pi n}$ (see [15], Sections 3.2 and 3.3). Hence, we arrive at Stirling's formula: $n!\simeq\sqrt{2\pi n}\,n^ne^{-n}$ for $n\gg 1$.

[7.17] James ***Stirling*** (1692-1770), a Scottish mathematician.

[7.18] Note that for any (arbitrarily large) value of n the quantity $y=x^ne^{-x}$ tends to zero as $x\to\infty$ (see Section 6.5; this fact will be used later); if this was not so, we could not speak of an area equal to I_n.

7.7* More on Natural Logarithms

Let us examine the problem of *natural logarithms*. In Chapter 3, the integration of power functions led us to the following results:

$$\int x^n\,dx = \frac{1}{n+1}\,x^{n+1} + C \quad \text{for } n \neq -1, \tag{7.7.1}$$

$$\int \frac{dx}{x} = \ln x + C, \tag{7.7.2}$$

where, of course, the constant C may be any number and not the same in Eqs. (7.7.1) and (7.7.2).

But why does the case (7.7.2) differ so strongly from (7.7.1)? What is the reason for this? At first glance there is a contradiction here. Imagine a sequence of power functions, or curves, $y = x^{-0.9}$, $y = x^{-0.99}$, $y = x^{-1}$, $y = x^{-1.01}$, and $y = x^{-1.1}$. All these curves lie close to each other, and the curve with $n = -1$ in no way differs qualitatively from the adjacent curves, for instance, from the curves with $n = -0.99$ and $n = -1.01$. Hence, the area under the curves $y = x^n$ (within fixed limits $a < x < b$) must be a smooth function of n, that is, it must change smoothly as we pass the value $n = -1$. But this area is expressed by a definite integral, whose form at $n = -1$ is quite different from that at $n = -0.99$ and $n = -1.01$.

To resolve this seeming contradiction, we must show that for n close to -1 the integral in (7.7.1) is close to the logarithmic expression in (7.7.2). Let us carry out the following calculations:

$$S = \int_a^b \frac{dx}{x} = \ln\frac{b}{a} \quad \text{for } n = -1$$

and $a,\ b > 0$; (7.7.3)

$$S = \int_a^b x^{-1+\varepsilon}\,dx = \frac{1}{\varepsilon}\,(b^\varepsilon - a^\varepsilon)$$

for $n = -1 + \varepsilon$. (7.7.4)

If the exponent n is close to -1, then ε is small and the expression (7.7.4) for S assumes a form that is inconvenient for calculations. Any number raised to the zeroth power is equal to unity, which means that if we substitute $\varepsilon = 0$ into (7.7.4), we will have $b^\varepsilon - a^\varepsilon = 1 - 1 = 0$. Thus, at $\varepsilon = 0$, both the numerator and the denominator in (7.7.4) vanish.

In this lies the answer to a question often posed by students that compare the indefinite integrals (7.7.1) and (7.7.2): the right-hand side of (7.7.1) contains $n+1$ in the denominator, and at first glance it seems to tend to infinity as $n \to -1$, while the logarithm is finite. But this infinity in (7.7.1) is fictitious, since it disappears when we pass to a definite integral and, hence, it can be eliminated by appropriately selecting the constant of integration C (e.g. $C = -(n+1)^{-1} = -1/\varepsilon$).

How does one go about simplifying the expression for an area that incorporates a^ε and b^ε? How should we calculate powers with small exponents? Let us write the following identities:

$$a = e^{\ln a}, \quad a^\varepsilon = e^{\varepsilon \ln a}.$$

We use the main property of the number e, that is, that for small $r \ll 1$ we have $e^r \simeq 1 + r$ (to within terms of the second and higher orders of smallness, that is, quantities of the order of r^2, r^3, etc.). If n is close to -1, that is, $n = -1 + \varepsilon$, where the term ε is extremely small, we can apply the formulas we have just written:

$$S = \frac{1}{\varepsilon}\,(b^\varepsilon - a^\varepsilon) = \frac{1}{\varepsilon}\,(e^{\varepsilon \ln b} - e^{\varepsilon \ln a})$$

$$\simeq \frac{1}{\varepsilon}\,[(1 + \varepsilon \ln b) - (1 + \varepsilon \ln a)]$$

$$= \ln b - \ln a = \ln\frac{b}{a}. \tag{7.7.5}$$

This resolves the contradiction. In the limit of small ε the quantity ε has canceled out, and we have found that formula (7.7.1) leads to the same result as formula (7.7.2). The area under the curve $y = x^n$ proves to be a smooth function of the exponent n even when n passes through the "singular" value $n = -1$.

We know how to calculate small powers x^ε for $\varepsilon \ll 1$. From school mathematics we know that $x^0 = 1$. Now we know how much x^ε differs from unity if ε is close to zero but still differs somewhat from zero.

We can approach the integral $J = \int x^{-1}\,dx$ from still another angle. Imagine a person that knows nothing about logarithms. However, we assume that he or she knows what integral calculus is and, confronted with an integral, would like to study its properties. Thus, let us consider the integral

$$J(a, b) = \int_a^b \frac{dx}{x}. \qquad (7.7.6)$$

A remarkable property of integral (7.7.6) is that *the value of* $J(a, b)$ *does not change if* a *and* b *are multiplied by a quantity* k (*the same for* a *and* b). As for the area under the curve $y = x^{-1} = 1/x$, we can say that if we replace the limits of integration a and b in (7.7.6) with ka and kb the length of the strip whose area we are calculating will increase k-fold, but simultaneously its length will decrease k-fold (here we assume that $k > 1$; see Figure 7.7.1).

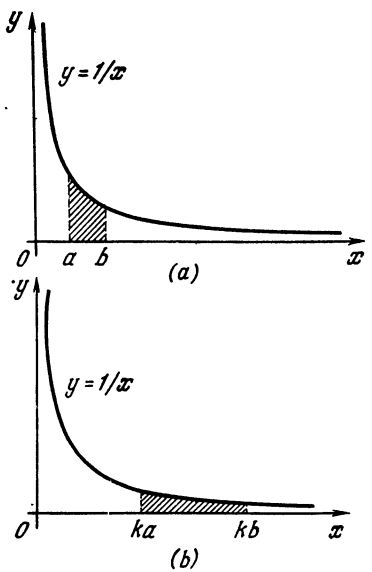


Figure 7.7.1

The area of the strip does not change in the process.

This result can easily be obtained without applying geometrical methods. Let us introduce the variable $z = kx$, $z = z/k$, $dx = (1/k)\,dz$. For the new variable the limits of integration will change, too; namely, $x = a$ becomes $z = ka$ and $x = b$ becomes $z = kb$. We have

$$J(a, b) = \int_a^b \frac{dx}{x} = \int_{ka}^{kb} \frac{\frac{1}{k}\,dz}{\frac{1}{k}\,z} = \int_{ka}^{kb} \frac{dz}{z}$$

$$= \int_{ka}^{kb} \frac{dx}{x} = J(ka, kb).$$

(At the end of the chain of formulas we employed the fact that the variable of integration is a dummy variable and can be denoted by any letter.)

If the function of two variables $J(a, b)$ does not change when we multiply both a and b by a constant k, then this function depends only on the *ratio* of the two variables:

$$J(a, b) = F(t), \text{ where } t = b/a. \qquad (7.7.7)$$

Indeed, suppose that $J(1, t) = F(t)$. Then, since $J(a, b) = J(ka, kb)$ for all values of k, we put $k = 1/a$ and obtain

$$J(a, b) = J\left(\frac{1}{a}a, \frac{1}{a}b\right) = J\left(1, \frac{b}{a}\right)$$

$$= J(1, t) = F(t).$$

But we know that J is an integral. Using the general properties of an integral, we obtain an important property of $F(t)$. Let us set $b = ac^m$, where m is an integer, and split the domain of integration into m parts. Then $J(a, b)$ is equal to a sum of integrals:

$$J(a, b) = J(a, ac) + J(ac, ac^2)$$

$$+ J(ac^2, ac^3) + \ldots + J(ac^{m-1}, b). \qquad (7.7.8)$$

But all the integrals on the right-hand side are the same, since for every one of them the ratio of the upper limit of integration to the lower limit is c

(we recall that $b = ac^m$). Thus,

$$J(a, b) = J(a, ac^m) = F\left(\frac{b}{a}\right) = F(c^m),$$

$$J(a, ac) = J(ac, ac^2) = \ldots = F(c),$$

and, hence,

$$F(c^m) = mF(c). \qquad (7.7.9)$$

We see that the function $F(t)$, which is defined as the integral $\int_1^t dx/x$ with the lower limit of integration equal to unity, possesses the following remarkable property: when the independent variable is raised to the mth power (i.e. $c \to c^m$), the function is multiplied by m. But this is the main property of the logarithm! If we denote the ratio $F(d)/F(c)$ (where c is fixed and d can vary) by $f(d)$, then

$$f(c^m) = \frac{F(c^m)}{F(c)} = m,$$

that is, if $d = c^m$, then $f(d) = m$; in other words, $f(d)$ is the power to which we must raise the constant fixed number c (the base) to obtain the number d. This means simply that $f(d)$ is the logarithm of number d (the reader will recall the common definition of a logarithm):[7.19]

$$f(d) = \log_c d.$$

Is there any need to consider the simple formula $\int x^{-1}\,dx = \ln + C$ in such detail? For many calculations this is unnecessary, and in the majority of textbooks this is not done. Of course, it is nice to know how to deal with the problem when the exponent in the integral is close to unity but is not exactly unity and what the error introduced is when we substitute -1 for $-1 + \varepsilon$; the first part of this section was devoted to these aspects. But besides the purely technical aspects, there is an aspect of principal importance here (and the remaining part of the section is devoted to this), namely, that a function (in the case at hand, the logarithmic function) can be *defined* by means of an integral!

In school we often encounter functions that are specified by means of algorithms, or rules for calculating their values. What does the notation $f(x) = x^3$ mean? Here we have the rule: multiply x by x and then by x once more, and you get $f(x)$. Starting with this rule, we can study the properties of function $f(x)$; for instance, it is clear that if x is increased by a factor of two, the values of the function increase by a factor of eight. But a function may be specified by an integral, and its properties can be studied even if we do not know how to find its values directly (i.e. not by integration but by means of various algebraic transformations).[7.20] For instance, the function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt \qquad (7.7.10)$$

plays an important role in the theory of probability, while the function

$$t = \text{constant} \times \int_0^x \frac{dx}{\sqrt{x_0^2 - x^2}} \qquad (7.7.11)$$

in the theory of vibrations specifies the dependence of time on the position of a vibrating particle. Luckily, the integral on the right-hand side of (7.7.11) can be evaluated in finite form,

$$t = k \arcsin \frac{x}{x_0},$$

and this dependence can be reversed, that is, we can pass from the function $t = t(x)$ to the function $x = x(t)$. The result is $x = x_0 \sin \omega t$ (with $\omega = 1/k$). However, in more complicated cases in the theory of vibrations, we arrive at the integral

$$t = \int \frac{dx}{\sqrt{ax^4 + bx^3 + cx^2 + fx + g}}. \qquad (7.7.12)$$

[7.19] See A.I. Markushevich *Areas and Logarithms*, Mir Publishers, Moscow, 1981.

[7.20] Note that at the beginning of the 20th century the German mathematician F. Klein suggested introducing in school mathematics the logarithm by means of the formula $\int_1^a x^{-1}dx = \ln a$.

This integral cannot be evaluated directly, that is, it cannot be expressed in terms of simple functions, but its properties can be studied and tables of its values can be compiled.

Such a generalization of the concept of a function and a study of the properties of the function specified by the integral $\int x^{-1}\,dx$ require of the reader the involved line of reasoning given above.

Exercise

7.7.1. Derive from the definition (7.7.6) and (7.7.7) of $F(t)$ the main property of this function: $F(ab) = F(a) + F(b)$ for all positive a and b.

7.8 Average, or Mean, Values

The concept of an integral can be used to give an exact definition of the *mean*, or *average*, of a quantity that is a function of another quantity.

If we have a quantity that assumes a set of several values, say the m values $v_1, v_2, v_3, \ldots, v_m$, then the natural way of defining the mean $\bar{v}$ of this quantity is to write

$$\bar{v} = \frac{v_1 + v_2 + v_3 + \ldots + v_m}{m}.$$

But how should we define the *mean value of a function* $v(t)$ of a variable t that assumes all values in the interval from a to b $(a < t < b)$?

Let us assume that $v(t)$ is the instantaneous velocity. How should we define the value $\bar{v}(a, b)$, that is, the *mean velocity* over the time interval from a to b? The average, or mean, velocity is defined as the ratio of the distance traveled to the time that has elapsed:

$$\bar{v}(a, b) = \frac{z(a, b)}{b-a} = \frac{\int_a^b v(t)\,dt}{b-a}. \qquad (7.8.1)$$

This definition of the mean is meaningful when the function is not only the rate of motion but any other function. For instance, suppose that $y = y(x)$

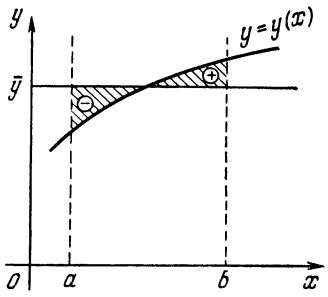


Figure 7.8.1

is the graph of a function in the xy-plane depicted in Figure 7.8.1. Then $\int_a^b y(x)\,dx$ is the area under the curve. The formula

$$\bar{y}(a, b) = \frac{\int_a^b y(x)\,dx}{b-a},$$

$$\text{or } (b-a)\,\bar{y} = \int_a^b y(x)\,dx, \qquad (7.8.1a)$$

means that $\bar{y}$ is the height of a rectangle with base $b - a$ whose area is equal to the area under the curve.[7.21] This means that the hatched area with the sign "+" above the straight line $y = \bar{y}$ (Figure 7.8.1) is exactly equal to the hatched area with the sign "—" under the straight line $y = \bar{y}$. The graph of the function $y = y(x)$ (if it is not a straight line parallel to the x axis) must lie partly above the line $y = \bar{y}$ and partly below, where $\bar{y}$ is the mean value determined by (7.8.1a). Hence $\bar{y}$ is greater than the smallest value of $y(x)$ and smaller than the greatest value of $y(x)$ on the interval $a < x < b$.

Let us consider the following examples.

[7.21] In Section 7.5 we called the quantity $\bar{y}$ (for $y(x) > 0$ on the interval $a < x < b$) the *effective altitude*, or height, of the curvilinear trapezoid with area $S = \int_a^b y\,dx$.

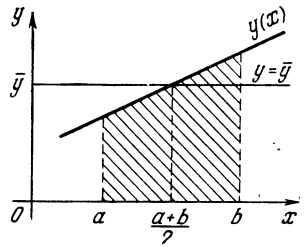


Figure 7.8.2

Suppose that $y\,(x)$ is the *linear* function $y = kx + m$. Then the integral $I\,(a, b)$, which is the area of the trapezoid in Figure 7.8.2 with altitude $b - a$, bases $y\,(a)$ and $y\,(b)$, and midline $y\left(\frac{a+b}{2}\right)$:

$$I\,(a, b) = \frac{y\,(a)+y\,(b)}{2}\,(b-a)$$

$$= y\,\frac{a+b}{2}\,(b-a). \qquad (7.8.2)$$

This formula can easily be obtained by resorting to geometrical methods:

$$I\,(a, b) = \int_a^b (kx+m)\,dx = \left(\frac{kx^2}{2}+mx\right)\Big|_a^b$$

$$= \frac{kb^2}{2}+mb-\left(\frac{ka^2}{2}+ma\right)$$

$$= (b-a)\left(\frac{kb}{2}+\frac{ka}{2}+m\right),$$

with $y\,(b) = kb + m$, $y\,(a) = ka + m$, and $y\left(\frac{a+b}{2}\right) = k\,(a+b)/2 + m$, from which readily follows (7.8.2). Thus, the *linear function*,

$$\bar{y} = \frac{y\,(a)+y\,(b)}{2} = y\left(\frac{a+b}{2}\right), \qquad (7.8.2a)$$

that is, for the linear function *the mean value over a given interval is the arithmetic mean of the values $y\,(a)$ and $y\,(b)$ at the endpoints of the interval.* Here is another way of stating this fact: *the mean value of the linear function over an interval is equal to the value of the function at the midpoint of the interval.*

An important example of the linear function is the time dependence of the velocity in *uniformly accelerated* or *decelerated motion*, that is, when a body moves under a *constant* force, say, the force of gravity, when $v = gt + v_0$. In calculating the distance traveled we use the properties of the mean of the linear function:

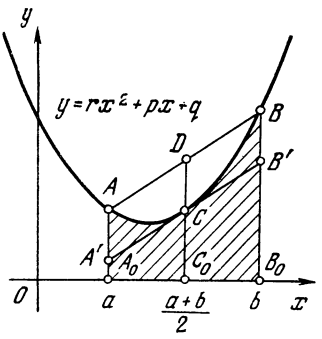


Figure 7.8.3

$$z\,(a,\ b) = (b-a)\left[\frac{v\,(b)+v\,(a)}{2}\right]$$

$$= (b-a)\left(\frac{gb+ga}{2}+v_0\right).$$

The reader must bear in mind, however, that for another, that is, nonlinear, dependence, the formula (7.8.2*a*) becomes invalid.

Let us consider the example of the *quadratic* function (parabola) $y = rx^2 + px + q$. Suppose, for the sake of definiteness, that r is positive, and let us consider an arch of the parabola, say with $a < x < b$. Figure 7.8.3 then shows that[7.22]

$$y\left(\frac{a+b}{2}\right) < \frac{y\,(a)+y\,(b)}{2}.$$

We now turn to the integral $\int_a^b y\,(x)\,dx$, that is, the area under the parabola. This area is smaller than that of the trapezoid with bases A_0A and B_0B. On the other hand, if we

[7.22] See Section 7.4. We recall that the parabola $y = rx^2$, $r > 0$, is convex downward, and the parabola $y = rx^2 + px + q$ with arbitrary p and q is derived from the parabola $y = rx^2$ via parallel translation (see Section 1.4).

draw through point $C\left(\frac{a+b}{2}, y\left(\frac{a+b}{2}\right)\right)$ the tangent to the parabola, it intersects the vertical lines $x=a$ and $x=b$ at points A' and B' and completes the trapezoid $A_0B_0B'A'$ with midline $C_0C=y\left(\frac{a+b}{2}\right)$; the area of this trapezoid is obviously smaller than the area under the curve. Thus, in the case of a parabola with $r>0$ we have

$$(b-a)\,y\left(\frac{a+b}{2}\right)<\int_a^b y(x)\,dx$$

$$<(b-a)\frac{y(a)+y(b)}{2}.$$

Correspondingly, we have the following inequalities for the mean value $\bar{y}$ over the interval from a to b:

$$y\left(\frac{a+b}{2}\right)<\bar{y}<\frac{y(a)+y(b)}{2}.$$

For the quadratic function there is also the exact formula, known as *Simpson's rule*[7.23] (we give it without derivation; see Exercise 7.8.2), which is valid for both signs of r:

$$\bar{y}=\frac{2}{3}y\left(\frac{a+b}{2}\right)+\frac{1}{3}\left[\frac{y(a)+y(b)}{2}\right]$$

$$=\frac{1}{6}y(a)+\frac{2}{3}y\left(\frac{a+b}{2}\right)+\frac{1}{6}y(b). \qquad (7.8.3)$$

This formula provides a good approximation scheme for calculating the area under any smooth curve (see Exercise 7.8.4*b*).

Here are two more simple facts pertaining to mean values.

1. *The mean value of a constant over any interval is that constant.* This is clear physically: if the instantaneous velocity does not change, then the mean, or average, velocity over an interval is equal to the constant value of the instantaneous velocity.

[7.23] Thomas ***Simpson*** (1710-1761), an English algebraist, analyst, geometer, and probabilist. Simpson's rule (7.8.3) is valid even for a *cubic* function $y=Ax^3+Bx^2+Cx+D$ (for arbitrary A, B, C, and D; if $A=0$, we have the case of a quadratic function; see Exercise 7.8.2c).

This is also very simply obtained from formula (7.8.1*a*):

$$\bar{C}(a,\,b)=\frac{\int_a^b C\,dx}{b-a}=\frac{C(b-a)}{b-a}=C.$$

2. *The mean value of the sum of two functions is equal to the sum of the mean values of the summands:*

$$\overline{y_1+y_2}=\bar{y}_1+\bar{y}_2.$$

Indeed,

$$\overline{y_1+y_2}=\frac{1}{b-a}\int_a^b[y_1(x)+y_2(x)]\,dx$$

$$=\frac{1}{b-a}\int_a^b y_1(x)\,dx$$

$$+\frac{1}{b-a}\int_a^b y_2(x)\,dx=\bar{y}_1+\bar{y}_2.$$

Here are some additional examples. Let us find the mean value of the function $y=\sin x$ on the interval from $x=0$ to $x=\pi$:

$$\bar{y}(0,\,\pi)=\frac{\int_0^\pi \sin x\,dx}{\pi-0}=\frac{2}{\pi}\simeq 0.637.$$

The mean value of the function $y=\sin x$ on the interval between $x=0$ and $x=b$ is

$$\bar{y}(0,\,b)=\frac{\int_0^b \sin x\,dx}{b-0}=\frac{1-\cos b}{b}. \qquad (7.8.4)$$

What will happen if we increase the number b without bound, that is, if we increase without bound the interval? In (7.8.4) the numerator does not exceed two for arbitrary b (it is equal to 2 if $\cos b=-1$, that is, for $b=\pi$, 3π, 5π, 7π, etc.). The denominator in (7.8.4) will increase without bound, and so we can say that over an *infinitely long* interval (we arrive at such an interval if we send b to ∞) the mean value of the sine is zero (namely, the larger the interval, the closer to zero is the mean value of $\sin x$). By applying

a similar method, we can show that the mean value of the function $y = \cos x$ over an infinite interval is also equal to zero. Indeed,

$$\bar{y}(0, b) = \frac{\int_0^b \cos x\, dx}{b-0} = \frac{\sin x \mid_0^b}{b} = \frac{\sin b}{b}, \tag{7.8.5}$$

and if we increase the number b without bound, the denominator of (7.8.5) will increase without bound and the numerator will not exceed unity. Hence, the entire fraction tends to zero: $\bar{y}(0, \infty) = 0$. Literally in the same way we find that the mean value of the function $y = \cos kx$ is also equal to zero over an infinitely long interval.

Let us find the mean value of the function $y = \sin^2 x$ over the infinite interval from $x = 0$ to $x = \infty$. Using a familiar formula of trigonometry, $\sin^2 x = (1/2)(1 - \cos 2x)$, we find that

$$\overline{\sin^2 x} = \frac{1}{2} - \overline{\frac{1}{2}\cos 2x} = \frac{1}{2} - 0 = \frac{1}{2}.$$

If we now take advantage of the formula $\sin^2 x + \cos^2 x = 1$, we get the mean value of $\cos^2 x$ over the same interval:

$$\overline{\cos^2 x} = \overline{1} - \overline{\sin^2 x} = 1 - 1/2 = 1/2.$$

Often it is convenient to use means instead of integrals. In essence, these quantities are equivalent: knowing the integral $I = \int_a^b y\, dx$, we can easily find the mean $\bar{y} = I/(b - a)$, while having calculated the mean $\bar{y}$ we can easily find $I = (b - a)\,\bar{y}$.

The convenience of the mean lies in the fact that $\bar{y}$ is a quantity with the same dimensions as those of y and, obviously, is of the same order of magnitude as y is on the interval under investigation (compare with what was said in connection with formula (5.6.6). Therefore, it is more difficult to miss an error that is ten times the value of $\bar{y}$ than to miss the same error in the value of the integral.

It is usually assumed that the student studying higher mathematics has mastered arithmetics and algebra and would never allow an error of ten times the quantity determined or an error in the sign. Experience has shown, however, that is not the case. For this reason, calculations must always be carried out in such a manner so as to reduce the probability of an undetected error.

We note also that the definition (7.8.1) of the mean $\bar{v}$ of the function $v(t)$ is closer, perhaps, to the concept of the weighted mean or weighted average than to the concept of the mean $\bar{v} = m^{-1}(v_1 + v_2 + \ldots + v_m)$ of a set of numbers understood as the arithmetic mean. Suppose that we have to determine the average mass $\bar{p}$ of one stone in a pile of stones (or the average price of a book in a bookstore). Let us first find the total mass of the stones (or the total price of the books) by setting up the sum $P = p_1 n_1 + p_2 n_2 + \ldots + p_k n_k$ (respectively, $Z = z_1 n_1 + z_2 n_2 + \ldots + z_k n_k$), where n_1 is the number of "like" stones (or books), that is, stones with the same mass p_i (or books with the same price z_i). Then we divided P (or Z) by the total number of stones in the heap (or books in the store). The result is

$$\bar{p} = \frac{p_1 n_1 + p_2 n_2 + \ldots + p_k n_k}{N}$$

or

$$\left(\bar{z} = \frac{z_1 n_1 + z_2 n_2 + \ldots + z_k n_k}{N}\right).$$

We do the same every time we wish to find the average of the quantities $y_1, y_2, \ldots, y_N$ when these quantities are separated into k groups of "like" objects, with $n_1, n_2, \ldots, n_k$ objects in each group (that is, n_1 objects in the first group, n_2 in the second, and so on; obviously, $n_1 + n_2 + \ldots + n_k = N$).

Let us now return to our example of a continuously changing velocity v. Suppose that we know the law $v = v(x)$ by which the (instantaneous) velocity v can be found at each point on

the path of travel. We wish to find the total time T it took an object to travel with velocity $v(x)$ from point $x = a$ to point $x = b$. The time Δt it takes the object to travel a small distance Δx will obviously be equal to the quotient $\Delta x/v$, where v is the velocity over Δx, which can be assumed constant since it has no time to change appreciably over Δx. (The more exact way of stating this is to write $\Delta t \simeq \Delta x/v$, since even over a small distance Δx the velocity changes somewhat). The total time of travel will be approximately expressed by the sum

$$T \simeq \frac{\Delta x_1}{v_1} + \frac{\Delta x_2}{v_2} + \ldots + \frac{\Delta x_k}{v_k}, \quad (7.8.6)$$

where $\Delta x_i = x_i - x_{i-1}$ (here $i = 1, 2, \ldots, k$, and $x_0 = a, x_1, x_2, \ldots, x_k = b$ are points that divide the entire distance from a to b into k small intervals), and where v_i can be taken as being either $v(x_i)$ or $v(x_{i-1})$ (cf. Section 3.1). The same value of T is provided by the following integral:

$$T = \int_a^b \frac{dx}{v(x)}, \quad (7.8.6a)$$

which is the limit of the weighted mean (7.8.6) in which the various Δx_i are taken with different weights $1/v_i$; the mean value of $1/v$ is given by the ratio of (7.8.6a) to the entire distance traveled, $b - a$.

There are several aspects that are related to the fact that in the formulas of integral calculus we often employ weighted means and that must not be ignored. Suppose we have a *chemical reaction* whose rate $F = F(\theta)$, that is, the quantity of substance entering the reaction per unit itme, depends on the temperature θ at which the reaction occurs; the temperature θ changes in the course of the reaction, or $\theta = \theta(t)$, where t is time. The average (or mean) temperature of the reaction lasting from time $t = t_1$ to time $t = t_2$ will, obviously, be

$$\bar{\theta} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \theta(t)\, dt, \quad (7.8.7)$$

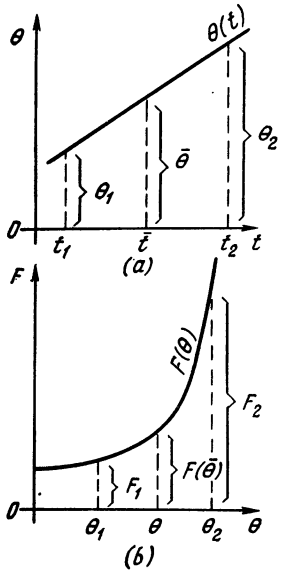


Figure 7.8.4

while the average reaction rate is expressed by the formula

$$\bar{F} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} F(\theta)\, dt. \quad (7.8.7a)$$

One must not think that the average reaction rate $\bar{F}$ coincides with the rate $F(\bar{\theta})$ taken at the average reaction temperature, $\bar{\theta}$. For instance, suppose that $\theta = \theta(t)$ is a linear function (Figure 7.8.4a). Then, obviously, $\bar{\theta} = (\theta_1 + \theta_2)/2 = \theta\left(\frac{t_1+t_2}{2}\right)$. As for the reaction rate F, we assume that it grows exponentially with temperature θ, that is, $F(\theta) = e^{k\theta}$. But because F grows so rapidly (see Figure 7.8.4b; note that since the temperature θ increases with time uniformly, $F(t)$ is also an exponential function), the value of the integral (7.8.7a) almost entirely is determined by the (large) values of F at the right end of the interval of variation of θ (or t) and depends but little on the other values of F, whereby we cannot write $\bar{F} = F(\bar{\theta}) = \frac{1}{2}[F(\theta_1) + F(\theta_2)]$. This fact completely explains the nonhomogeneity in the weights of $F(\theta)$

associated with different intervals $\Delta\theta$ of variation of θ.

In connection with the question of mean values we recall two theorems that refer to *continuous* and *smooth* functions.

1. Lagrange's theorem *The mean value of a function over an interval is equal to the value of the function at some point within the interval.* In the notation of formula (7.8.1a),

$$\overline{f}(a, b) = f(c), \quad a < c < b. \tag{7.8.8}$$

Let us draw a rectangle whose area is equal to the integral $\int_a^b f(x)\,dx$, which is the area of a curvilinear trapezoid bounded from above by the curve $f(x)$, from below by the x axis, and from the sides by the straight lines $x = a$ and $x = b$. The base of the rectangle is the line segment from a to b, while the height of the rectangle, or the effective altitude of the trapezoid, is equal to the mean value $\overline{f}(a, b)$ of the function. Obviously, the upper base of the rectangle is certain to intersect the arc of the curve $y = f(x)$ within the limits a and b at a certain point $(c, f(c))$ that lies somewhere between the endpoints of the arc (Figure 7.8.5a). Here $f(x)$ was considered *continuous*, since if $f(x)$ is discontinuous at some point within the interval, the above statement may prove to be invalid (Figure 7.8.5b).

Without looking at the figure, we can reason as follows. If M and m are the *greatest* and *smallest* values of $f(x)$ on the segment $a \leqslant x \leqslant b$, that is, if $M \geqslant f(x) \geqslant m$, then

$$M(b-a) = \int_a^b M\,dx \geqslant \int_a^b f(x)\,dx$$

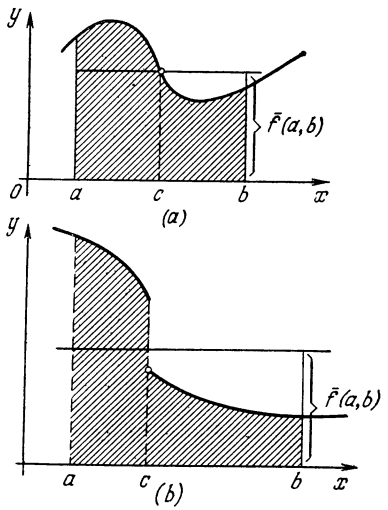


Figure 7.8.5

$$\geqslant \int_a^b m\,dx = m(b-a)$$

and, hence,

$$M \geqslant \frac{1}{b-a}\int_a^b f(x)\,dx = \overline{f(x)} \geqslant m.$$

But this leads us to (7.8.8), since a continuous function assumes on the interval from a to b (at a point c in this interval) a value A that is intermediate between the greatest and smallest values (i.e. such that $M \geqslant A \geqslant m$), in particular the value

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

In a similar manner we can establish a simple generalization of the above theorem: *if a function $f(x)$ on the line segment $a \leqslant x \leqslant b$ is continuous and another function $g(x)$ does not change its sign on this interval* (i.e. is either everywhere positive or everywhere negative), *then*

$$\int_a^b f(x)\,g(x)\,dx = f(c)\int_a^b g(x)\,dx, \tag{7.8.9}$$

where, as usual, $a < c < b$.

Indeed, suppose that, for the sake of definiteness, $g(x)$ is nonnegative for $a \leqslant x \leqslant b$. Then, obviously,

$$M\int_a^b g(x)\,dx = \int_a^b Mg(x)\,dx \geqslant \int_a^b f(x)\,g(x)\,dx$$

$$\geqslant \int_a^b mg(x)\,dx = m\int_a^b g(x)\,dx. \tag{7.8.10}$$

where again M and m are the greatest and smallest values of the function $f(x)$ on the interval $a \leqslant x \leqslant b$. Dividing both parts of (7.8.10) by $\int_a^b g(x)\,dx > 0$, we get[7.24]

[7.24] Clearly, if a function $g(x)$ that does not change its sign on an interval $a \leqslant x \leqslant b$ is not "pathological" (say, not like the function $f(x)$ described at the beginning of Section 14.1), then the equality $\int_a^b g(x)\,dx = 0$ for $b \neq a$ means that $g(x) \equiv 0$, and in this case the validity of (7.8.9) is obvious.

$$M \geqslant \int_a^b f(x)\, g(x)\, dx \int_a^b g(x)\, dx \geqslant m. \qquad (7.8.11)$$

But a continuous function $f(x)$ assumes on the line segment $a \leqslant x \leqslant b$ all the values that lie between the greatest and smallest values of the function on this interval; hence, each number that lies between M and m, which means the fraction in (7.8.11), too, is equal to the value $f(c)$ at an intermediate point c (lying between a and b):

$$\int_a^b f(x)\, g(x)\, dx \int_a^b g(x)\, dx = f(c). \qquad (7.8.12)$$

(It is clear that (7.8.12) is equivalent to (7.8.9)).

Example. If $f(x)$ is an arbitrary continuous function and a and b have the same sign (i.e. both are either nonnegative or nonpositive), then

$$\int_a^b x^n f(x)\, dx = f(c) \int_a^b x^n\, dx = \frac{b^{n+1} - a^{n+1}}{n+1} f(c), \qquad (7.9.13)$$

where number c lies between a and b.

2. *Let us take a function $y = y(x)$ such that $y(a) = y(b)$.* It is assumed that the function is smooth. The above hypothesis implies that the function cannot be monotone in the interval from a to b. But if the function is not monotone, it must have a maximum or minimum somewhere between a and b. *This means that within the interval at a certain point $x = c$, where $a < c < b$, the derivative of the function vanishes: $y'(c) = 0$.* This statement is known as ***Rolle's theorem.*** [7.25]

The geometric meaning of Rolle's theorem is clear from Figure 7.8.6, where we have depicted two types of curves with maxima and minima (the second curve has both a maximum and a minimum between a and b).

Lagrange's and Rolle's theorems are interconnected: Rolle's theorem can be obtained as a corollary of Lagrange's theorem. Indeed, let us write the following identity that follows from the relationship between a derivative and an integral (see Section 3.3):

$$y(b) = y(a) + \int_a^b y'(x)\, dx.$$

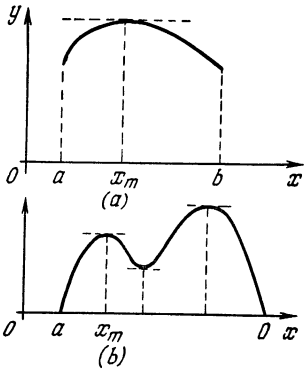


Figure 7.8.6

From this it follows that

$$\overline{y'}(a, b) = \frac{\int_a^b y'(x)\, dx}{b-a} = \frac{y(b) - y(a)}{b-a}.$$

But by the hypothesis of Rolle's theorem, $y(b) = y(a)$, whereby here $\overline{y'} = 0$. Now we only need to apply Lagrange's theorem to $y'(x)$: if the mean value of a function (in our case the derivative of $y(x)$) is zero on the interval from a to b, then there is a point c within the interval at which the function vanishes, $y'(c) = 0$.

Exercises

7.8.1. Find the mean value of the function $y = x^2$ on the interval from 0 to 2. Compare this mean value with the arithmetic mean of the values of the function at the endpoints of the interval and with the value in the middle of the interval.

7.8.2. Verify Simpson's rule (7.8.3) for (a) the function of Exercise 7.8.1, (b) the general quadratic function $y = rx^2 + px + q$, and (c) for the cubic function $y = Ax^3 + Bx^3 + Cx + D$.

7.8.3. The force of gravity diminishes as the distance from the earth's center grows according to the law $F = A/r^2$. Find the mean value of the force of gravity over a path that starts at the earth's surface (R the radius of the earth) and ends at a point that is R distant from the earth's surface (i.e. at a point that is $2R$ distant from the earth's center). [*Hint.* To find the mean value, use the "energy integral" $F(x)\, dx$, where F is the force and x the distance. This integral has an important physical meaning.]

[7.25] Michel ***Rolle*** (1652-1719), a French analyst, algebraist, and geometer.

7.8.4. Compare the mean value found in Exercise 7.8.3. with (a) the arithmetic mean of the values of the force of gravity at the endpoints of the interval (which coincides with the average when the quantity changes linearly with distance), and (b) the mean value calculated via Simpson's rule (7.8.3) (which coincides with the true mean value for a quadratic law of variation of the quantity with distance).

7.8.5. Find the average value of $y = x^n$ over the interval from $x = 0$ to $x = x_0$.

7.8.6. Find the mean value of the function $y = c\,e^{kx}$ over the interval in which y varies from $y = n$ to $y = m$; express this mean value in terms of n and m, eliminating c and k from the answer. Investigate the resulting expression when m is close to n, or $m = n + \nu$, $\nu \ll n$.

7.8.7. Find the mean values of the functions $y = \sin^2 x$ and $y = \cos^2 x$ over the following intervals: (a) from $x = 0$ to $x = \pi$, and (b) from $x = 0$ to $x = \pi/4$.

7.8.8. Determine the period of the function $y = \sin(\omega t + \alpha)$, where ω and α are constants. Find the mean value of y^2 over one period of y^2.

7.9 Arc Length

Let us pose the problem of finding the *arc length* s of a curve $y = f(x)$ from the point where $x = a$ to the point where $x = b$ (Figure 7.9.1). We replace the length Δs of a small arc MN of the curve $y = f(x)$ by the straight-line segment MN connecting M and N.[7.26] (We consider only curves that have no discontinuities or cusps.) By the Pythagorean theorem,

$$\Delta s \simeq \sqrt{(\Delta x)^2 + (\Delta y)^2} = \Delta x \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2},$$

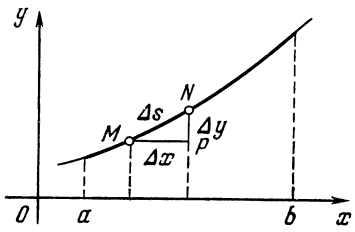


Figure 7.9.1

whence

$$\frac{\Delta s}{\Delta x} \simeq \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}. \qquad (7.9.1)$$

We pass to the limit in (7.9.1) as $\Delta x \to 0$. The ratio $\Delta y/\Delta x$ becomes the derivative $y' = f'(x)$. Thus, we get

$$ds = \sqrt{1 + (f'(x))^2}\,dx.$$

The entire length of the arc is

$$s = s(a, b) = \int_a^b \sqrt{1 + (f'(x))^2}\,dx. \qquad (7.9.2)$$

If the curve is specified parametrically, that is, $x = x(t)$ and $y = y(t)$ (see Section 1.8), then, dividing the (approximate) equations

$$\Delta s \simeq \sqrt{\Delta x^2 + \Delta y^2}$$

by Δt, passing to the limit as $\Delta t \to 0$, and taking into account the fact that $\lim\limits_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$ and $\lim\limits_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt}$, we obtain

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2},$$

or $ds = \sqrt{(x')^2 + (y')^2}\,dt$. Integrating this from $t = \alpha$ to $t = \beta$, we get

$$s(\alpha, \beta) = \int_\alpha^\beta \sqrt{(x')^2 + (y')^2}\,dt, \qquad (7.9.2a)$$

where $s(\alpha, \beta)$ is the length of the arc of the curve limited by the values α and β of parameter t.

Because of the radical under the integral sign in (7.9.2) and in (7.9.2a), it is rarely possible to evaluate the integral directly, that is, express $s(a, b)$

[7.26] The difference between the arc length and the length of the line segment (the chord) is of the order of $|\Delta x|^3$, and so can surely be neglected when we pass to the limit (to differentials). For instance, the length of the arc of a circle of radius r subtending a small angle Δt is equal to $r\Delta t$ (we use radians to measure angles), while the length of the curve corresponding to this arc is $2r \sin(\Delta t/2) = 2r[\Delta t/2 - (1/6)(\Delta t/2)^3 + \ldots] = r\Delta t - (1/24)\,r\,(\Delta t)^3 + \ldots$ (see Section 6.2), so that the difference between the two lengths is $(1/24)r\,(\Delta t)^3$, that is, is of the order of $(\Delta s)^3$, where Δs is a quantity of the order of the length of the arc (or chord).

and s (α, β) in terms of functions of the limits of integration a and b (or α and β) expressed explicitly.

Note also that from the same right triangle[7.27] with sides Δx, Δy and MN ($\simeq \Delta s$) which we used to express Δs in terms of Δx and Δy we conclude that

$$\Delta x \simeq \Delta s \cos\psi, \quad \Delta y \simeq \Delta s \sin\psi,$$

where ψ is the angle between the x axis and the line segment MN. As $\Delta x \to 0$, obviously $\psi \to \varphi$ where φ is the angle between the x axis and the *tangent* to the curve $y = f(x)$ at point M. Thus,

$$dx = ds\cos\varphi, \quad dy = ds\sin\varphi, \quad ds = \sqrt{dx^2 + dy^2}, \tag{7.9.2b}$$

where the fact that we have used differentials means that dx ($=\Delta x$), dy, and ds are small.

Some examples follow in which the computations can be carried out with relative ease.

1. *The circumference of a circle.* We seek the circumference of the circle $x^2 + y^2 = R^2$ or, to be precise, the length s of one-fourth of the circumference in the first quadrant and then multiply it by 4.

From the equation of the circle we have

$$y = \sqrt{R^2 - x^2}, \quad y' = -\frac{x}{\sqrt{R^2 - x^2}}.$$

By formula (7.9.2),

$$s = \int_0^R \sqrt{1 + \frac{x^2}{R^2 - x^2}}\, dx = \int_0^R \frac{R\,dx}{\sqrt{R^2 - x^2}}. \tag{7.9.3}$$

We introduce a new variable t thus: $x = R\sin t$. Then $dx = R\cos t\, dt$ and from (7.9.3) we get (see Section 5.6)

$$s = \int_0^{\pi/2} R\,dt = \frac{\pi R}{2},$$

[7.27] This triangle played an important role in the reasoning of Leibniz, who used it in introducing differentials.

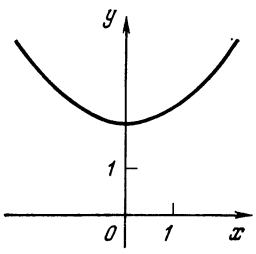


Figure 7.9.2

whence we obtain the well-known formula for the circumference $C = (\pi R/2) \times 4 = 2\pi R$.

2. *Catenary curve.* This is a curve whose equation is

$$y = \frac{a}{2}(e^{x/a} + e^{-x/a}), \tag{7.9.4}$$

where a is a constant.[7.28] The word "catenary" comes from the Latin *catena*, meaning "chain"; the curve has the form of a freely hanging heavy, flexible, and inextensible cable (or chain) suspended from two fixed points (the density of the cable is assumed the same at each section of the cable). The graph of the catenary curve is given in Figure 7.9.2 (for $a = 2$).

Let us find the length of arc of the catenary from $x = 0$ to $x = x_0$. From (7.9.4) we have $y' = \frac{e^{x/a} - e^{-x/a}}{2}$, and so

$$\sqrt{1 + (y')^2} = \sqrt{1 + \frac{e^{2x/a} - 2 + e^{-2x/a}}{4}} = \sqrt{\frac{(e^{x/a} + e^{-x/a})^2}{4}} = \frac{e^{x/a} + e^{-x/a}}{2}$$

and, hence

$$s = \int_0^{x_0} \frac{e^{x/a} + e^{-x/a}}{2}\, dx = \frac{a}{2}(e^{x/a} - e^{-x/a})\Big|_0^{x_0} = \frac{a}{2}(e^{x_0/a} - e^{-x_0/a}).$$

3. *Cycloid.* We know that the (parametric) equations of the cycloid are of the form

$$x = a(t - \sin t), \quad y = a(1 - \cos t),$$

[7.28] Equation (7.9.4) can also be written in the form $y = a\cosh(x/a)$, where by $\cosh t$ we have denoted the *hyperbolic cosine* of (see Section 14.4, in particular Figure 14.4.1).

where a is the radius of the circle generating the cycloid (see Section 1.8). Whence, by (7.9.2a),

$$ds = \sqrt{a^2(1-\cos t)^2 + (a\sin t)^2}\,dt$$
$$= a\sqrt{1-2\cos t+\cos^2 t+\sin^2 t}\,dt$$
$$= a\sqrt{2-2\cos t}\,dt = a\sqrt{4\sin^2\frac{t}{2}}\,dt$$
$$= 2a\sin\frac{t}{2}\,dt,$$

and the length of the arch of the cycloid lying between the points corresponding to the values α and β of parameter t is

$$s = 2a\int_\alpha^\beta \sin\frac{t}{2}\,dt = 4a\left(-\cos\frac{t}{2}\right)\Big|_\alpha^\beta$$
$$= 4a\left(\cos\frac{\alpha}{2}-\cos\frac{\beta}{2}\right).$$

Since two successive "cusps" of the cycloid (see Figure 1.8.3) correspond to values 0 and 2π of parameter t, the length of one arch of the cycloid is equal to $8a$, which is four times the diameter of the circle that generates the cycloid.

As pointed out earlier, in most cases it is difficult (or even impossible) to integrate the function $\sqrt{1+(y'(x))^2}$ in terms of elementary functions due to the radical. For this reason, *approximate* formulas for computing arc length are of particular interest.

Suppose that $(y'(x))^2$ is small compared with unity: $|y'(x)| \ll 1$. Then, neglecting $(y'(x))^2$ in (7.9.2), we get

$$s \simeq \int_a^b \sqrt{1}\,dx = b-a. \qquad (7.9.5)$$

The difference $b-a$ is the length of the horizontal line segment with endpoints $x=a$ and $x=b$. Formula (7.9.5) shows that if y' is small in absolute value (i.e. the curve differs little from the horizontal), then the length of the corresponding arc of this curve is also close to the length of the horizontal line segment (Figure 7.9.3a).

But if $(y'(x))^2 \gg 1$, then in (7.9.2) we can ignore unity in comparison with $(y'(x))^2$. The result is

$$s \simeq \int_a^b \sqrt{(y'(x))^2}\,dx = \int_a^b y'(x)\,dx = y(b)-y(a). \qquad (7.9.6)$$

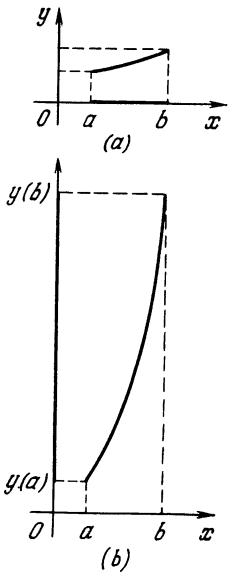


Figure 7.9.3

This formula shows us that in the given case the arc length of the curve is close to that of the vertical line segment with endpoints $y(a)$ and $y(b)$ (Figure 7.9.3b). Indeed, if the derivative y' is great, then the curve bends steeply upward and so is similar to a vertical straight line (for a vertical line the derivative is infinite).

The formulas (7.9.5) and (7.9.6) yield simple approximate formulas for the arc length. But these are very rough approximations, which can be obtained without appealing to (7.9.2). It is easy to obtain more exact formulas.

Let $|y'(x)| < 1$. Retaining two terms in the binomial expansion of $\sqrt{1+(y'(x))^2}$ (see Section 6.4) and discarding the other terms, we can write

$$\sqrt{1+(y'(x))^2} = [1+(y'(x))^2]^{1/2}$$
$$\simeq 1+\frac{1}{2}(y'(x))^2.$$

Formula (7.9.2) now yields

$$s \simeq \int_a^b \left[1+\frac{1}{2}(y'(x))^2\right]dx$$
$$= (b-a)+\frac{1}{2}\int_a^b (y'(x))^2\,dx.$$

But if $|y'(x)| > 1$, then

$$\sqrt{1+(y'(x))^2} = y'(x)\sqrt{1+\frac{1}{(y'(x))^2}}.$$

To the radical on the right-hand side we can apply the binomial theorem since $(y'(x))^{-2} < 1$. Retaining only two terms in the expansion, we get

$$y'(x)\sqrt{1+\frac{1}{(y'(x))^2}} \simeq y'(x)\left[1+\frac{1}{2(y'(x))^2}\right] = y'(x)+\frac{1}{2y'(x)}. \tag{7.9.7}$$

Substituting this into (7.9.2), we get

$$s \simeq \int_a^b \left[y'(x)+\frac{1}{2y'(x)}\right]dx$$

$$= \int_a^b y'(x)\,dx + \frac{1}{2}\int_a^b \frac{dx}{y'(x)}$$

$$= y(b)-y(a)+\frac{1}{2}\int_a^b \frac{dx}{y'(x)}.$$

We now have the following approximate formulas:

$$s \simeq (b-a)+\frac{1}{2}\int_a^b (y'(x))^2\,dx$$

if $|y'(x)|<1$,

$$s \simeq y(b)-y(a)+\frac{1}{2}\int_a^b \frac{dx}{y'(x)} \tag{7.9.8}$$

if $|y'(x)|>1$.

The integrals here are simpler than the integral in (7.9.2) and it is easier to perform the computations by these formulas than by (7.9.2). But these formulas are approximate, whereas (7.9.2) is exact.

What errors result from their use? The first formula in (7.9.8) is best for small $|y'|$ while the second is best for large $|y'|$. Both formulas yield the worst results at $|y'| = 1$. Therefore, to estimate the error here we will examine the worst case, which is $y'(x) = 1$. It is clear that if $y'(x) = 1$, then $y(x) = x + c$; the graph of this function is a straight line.

By the exact formula (7.9.2),

$$s = \int_a^b \sqrt{1+1}\,dx = \sqrt{2}(b-a) \simeq 1.415(b-a) \tag{7.9.9}$$

(this result immediately follows from the Pythagorean theorem or elementary trigonometry, since the straight line $y = x + c$ forms an angle of 45° with the axis of abscissas).

By the first of the formulas (7.9.8),

$$s \simeq (b-a)+\frac{1}{2}\int_a^b dx = \frac{3}{2}(b-a) = 1.5(b-a). \tag{7.9.10}$$

The second formula in (7.9.8) yields the same result: $s \simeq 1.5(b-a)$. Comparing (7.9.9) and (7.9.10), we see that the *maximum* error in the approximate formulas *is about* 6%.

When computing arc length, we must devide the curve into portions over which either $|y'| \leqslant 1$ everywhere or $y' \geqslant 1$ everywhere. Then the error will at least *not exceed* 6%. And since $(y'(x))^2$ assumes a value equal to unity only at certain points of the curve, a proper partition of the curve into portions will reduce the error *below* 6% (and usually substantially below 6%). Of course, there is no sense in finding the length of rectilinear portions (for one, portions over which $y'(x) \equiv 1$) by means of approximate formulas.

Let us take a look at some examples; almost always the computations are carried out to two decimal places.

1. Find the arc length of the *parabola* $y = x^2$ between the points with abscissas $x = 0$ and $x = 2$.

We find the derivative, $y' = 2x$. It is equal to 1 at $x = 0.5$, is greater than 1 for $x > 0.5$ and is less than unity for $x < 0.5$. Therefore, the arc length s_1 corresponding to a variation of x from 0 to 0.5 can be found from the first formula in (7.9.8) and the arc length s_2 corresponding to a variation of x from 0.5 to 2, by the second formula:

$$s_1 \simeq (0.5-0)+0.5\int_0^{0.5} 4x^2\,dx$$

$$= 0.5+2\frac{(0.5)^3}{3} \simeq 0.58,$$

$$s_2 \simeq 4-0.25+0.5\int_{0.5}^{2}\frac{dx}{2x}$$

$$= 3.75+0.25(\ln 2-\ln 0.5) \simeq 4.10,$$

whence, the sought-for arc length is

$$s = s_1+s_2 \simeq 0.58+4.10 = 4.68.$$

On the other hand, by (7.9.2),

$$s = \int_0^2 \sqrt{1+4x^2}\,dx,$$

whence, by making the change of variable $2x = z$ and employing formula 33 from Ap-

pendix 2 at the end of the book, we find the exact value of s:

$$\int \sqrt{1+4x^2}\,dx = \frac{1}{2}\left[x\sqrt{4x^2+1} + \frac{1}{2}\ln\left(2x+\sqrt{4x^2+1}\right)\right] + C, \qquad (7.9.11)$$

and, consequently,

$$s = \frac{1}{2}\left[2\sqrt{17} + \frac{1}{2}\ln\left(4+\sqrt{17}\right)\right] \simeq 4.65. \qquad (7.9.12)$$

(Any doubting reader can convince himself of the truth of the formula (7.9.11) by taking the derivative of the right-hand side of this formula.)

The error introduced by formulas (7.9.8) came out to about 0.7%

2. Find the arc length of the *exponential curve* $y = e^x$ between the points with abscissas $x = 0$ and $x = 1$.

In this case, $y' = e^x$, the derivative grows from 1 to e as x increases from 0 to 1. So we use the second formula in (7.9.8):

$$s \simeq e^1 - e^0 + 0.5\int_0^1 \frac{dx}{e^x}$$

$$\simeq 2.72 - 1 - 0.5e^{-x}\Big|_0^1 \simeq 2.04.$$

The exact formula for the arc length yields the value (see Exercises 7.9.1*b* and 7.9.2)

$$s = \sqrt{1+e^2} - \sqrt{2} + \frac{1}{2}\ln\frac{\sqrt{e^2+1}-1}{\sqrt{e^2+1}+1} - \frac{1}{2}\ln\frac{\sqrt{2}-1}{\sqrt{2}+1} \simeq 2.00.$$

The approximate formula introduces an error of 2%.

The arc length may also be approximated occasionally by a *series expansion* of the integrand in (7.9.2) in powers of x. By retaining an appropriate number of terms of the expansion, we can obtain the arc length to any degree of accuracy.

Let us consider an *example*. We wish to determine the circumference of a *circle* by seeking the arc length s of the circle corresponding to a central angle of 30°. The circumference is $C = 12s$. Quite obviously we will obtain the same kind of integral as in (7.9.3) but with a different upper limit:

$$s = \int_0^{R/2} \frac{R\,dx}{\sqrt{R^2-x^2}} \qquad (7.9.13)$$

(here we have used the fact that $R \sin 30° = R/2$). The integrand is transformed as follows:

$$\frac{R}{\sqrt{R^2-x^2}} = \frac{R}{R\sqrt{1-\left(\frac{x}{R}\right)^2}} = \frac{1}{\sqrt{1-\left(\frac{x}{R}\right)^2}} = \left[1-\left(\frac{x}{R}\right)^2\right]^{-1/2}. \qquad (7.9.14)$$

We put $(x/R)^2 = t$ and expand (7.9.14) in a binomial series (see formula (6.4.3)) to get

$$\left[1-\left(\frac{x}{R}\right)^2\right]^{-1/2} = (1-t)^{-1/2}$$

$$= 1 + \frac{1}{2}t + \frac{3}{8}t^2 + \frac{5}{16}t^3 + \frac{35}{128}t^4 + \ldots$$

$$= 1 + \frac{1}{2}\left(\frac{x}{R}\right)^2 + \frac{3}{8}\left(\frac{x}{R}\right)^4 + \frac{5}{16}\left(\frac{x}{R}\right)^6 + \frac{35}{128}\left(\frac{x}{R}\right)^8 + \ldots. \qquad (7.9.15)$$

Substituting (7.9.15) into (7.9.13) and integrating, we find that

$$s = R\left(\frac{1}{2} + \frac{1}{6\cdot 2^3} + \frac{3}{40\cdot 2^5} + \frac{5}{16\cdot 7\cdot 2^7} + \frac{35}{128\cdot 9\cdot 2^9} + \ldots\right). \qquad (7.9.16)$$

It is clear that the terms of the series (7.9.16) rapidly decrease and so we need only a few terms of the series to get s. Taking one term, we have $s \simeq R/2$, whence the circumference is $C \simeq 6R$. Taking two terms, we have $s \simeq 0.521R$, that is, $C \simeq 6.252R$. Three terms yield $s \simeq 0.523R$ and $C \simeq 6.276R$, and so on.

We know that the circumference of a circle is $C = 2\pi R$. Comparing this with the results we have obtained, we can approximate the value of the number π: 3, 3.126, 3.138, ... The more terms of the series (7.9.16) we take, the more exact the value of π we obtain. The value of π to four decimal places is $\pi \simeq 3.1416$.

Exercises

7.9.1. Write the arc length in the form of an integtal for (a) the parabola $y = x^2$ from point (0, 0) to point (1, 1), (b) the exponential curve $y = e^x$ from the point with $x = 0$ to the point with $x = 1$, and (c) the ellipse $x^2/a^2 + y^2/b^2 = 1$.

7.9.2. Complete Exercise 7.9.1*b* by making the change of variable $1 + e^{2x} = z^2$ in the integral.

7.9.3. Use the approximate formulas (7.9.8) to find the arc length of the catenary curve between points with $x = 0$ and $x = 2$

($a = 1$). Compare the result with exact value of the arc length.

7.9.4. Find approximately the arc length of the hyperbola $xy = -1$ between the points with $x = 0.5$ and $x = 1$. [*Hint.* In this case we cannot obtain an exact value because the integral in (7.9.2) cannot be expressed in terms of elementary functions.]

7.9.5. Obtain approximate values of the number π by computing the arc length of a circle with a central angle of 45° (retain three, four, and five terms in the series).

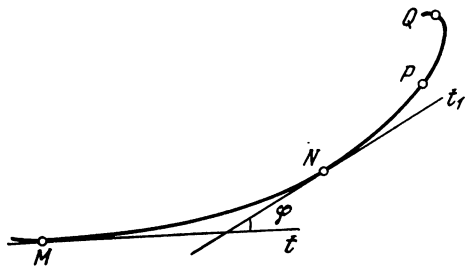


Figure 7.10.2

7.10 Curvature and the Osculating Circle

Related to the arc length is the problem of determining the ***curvature*** k and the ***radius of curvature*** R of a curve at a point. For a circle the radius of curvature R is simply the circle's radius, and the reciprocal of the radius is the curvature, $k = 1/R$. The smaller the radius R of a circle (i.e. the greater the curvature k), the tighter the circle (Figure 7.10.1). When R is very large, on the contrary, the curvature is hardly noticeable, that is, the circle is "almost" a straight line; for instance, the fact that the earth's radius is so large means that we do not notice the curvature of the earth's surface (to notice it, we would have to observe the earth from outer space, and this is not common practice).

Let us take an arbitrary curve Γ now. It is clear (Figure 7.10.2) that the curve is curved more on the portion PQ, where it sharply turns to the left, than on the portion MN, where its direction does not change so rapidly. But what is the exact meaning of the last statement? The explanation is as follows.

Consider the angle φ between the tangents of the curve at the endpoints M and N of the arc MN. If over this arc the tangent rotates only in one direction when we move from point m to point N, the angle φ shows the magnitude of the rotation of the tangent on arc MN (or the magnitude of the "rotation" of the curve, since the direction of the curve is specified by the direction of the tangent). To determine the extent to which the curve is curved we must also know the arc length s of MN: if s is large, then even a slow change in the direction of the tangent can lead to a large value of φ.

Thus, the ratio

$$k_{\text{av}} = \varphi/s \tag{7.10.1}$$

specifies the *average* (or *specific*) curvature of the curve over arc MN. *The curvature at point M of the curve* is defined then as the *limit* of the average curvature when the length of arc MN tends to zero:

$$k = \lim_{\Delta s \to 0} \frac{\Delta\alpha}{\Delta s},$$

where Δs is the length of a (small) arc MM_1 of curve Γ, and $\Delta\alpha$ is the increment of angle α corresponding to this arc (α is the angle between the tangent to the curve at point M and the x axis), or the angle of rotation of the tangent to the curve (the angle between the tangents t and t_1 to Γ at points M and M_1; Figure 7.10.3). But by virtue of the definition of a derivative (see Section 2.1), we can rewrite the last limit as

$$k = \frac{d\alpha}{ds}. \tag{7.10.2}$$

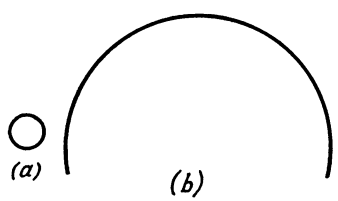


Figure 7.10.1

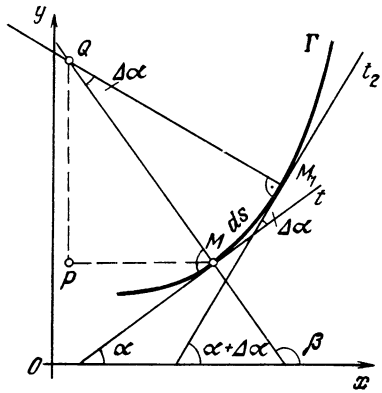


Figure 7.10.3

Thus, curvature k can be defined as the *rate of rotation of the tangent* to Γ as the point moves along the curve with a unit (linear) speed, that is, under conditions where point M travels the distance $MM_1 = ds$ in time $dt = ds$.

Now we need only recall that $\tan \alpha = y'$, so that $\alpha = \arctan y'$ and, hence,

$$d\alpha = d(\arctan y') = \frac{dy'}{1+(y')^2} = \frac{y''\,dx}{1+(y')^2};$$

moreover, $ds = \sqrt{1+(y')^2}\,dx$ (see Section 7.9). Thus, finally,

$$k = \frac{d\alpha}{ds} = \left[\frac{y''}{1+(y')^2}\,dx\right] \Big/ \left[\sqrt{1+(y')^2}\,dx\right] = \frac{y''}{[1+(y')^2]^{3/2}} \qquad (7.10.3)$$

and, hence,

$$R = \frac{1}{k} = \frac{[1+(y')^2]^{3/2}}{y''}. \qquad (7.10.3a)$$

If curve Γ is specified parametrically, $x = x(t)$ and $y = y(t)$ (see Section 1.8), then $dy/dx = y'/x'$ and $\frac{d^2y}{dx^2} = \frac{x'y''-x''y'}{(x')^3}$ (see formula (4.4.9)). Therefore, from (7.10.3) and (7.10.3a) it follows that

$$k = \left[\frac{x'y''-x''y'}{(x')^3}\right] \div \left[1+\frac{(y')^2}{(x')^2}\right]^{3/2} = \frac{x'y''-x''y'}{[(x')^2+(y')^2]^{3/2}},$$

$$R = \frac{1}{k} = \frac{[(x')^2+(y')^2]^{3/2}}{x'y''-x''y'}. \qquad (7.10.3b)$$

We note also that since α is a dimensionless quantity (we employ, as usual, radians for measuring angles) and s has the dimensions of length $[l]$, the curvature k of a curve has the dimensions $[l]^{-1}$ (say cm^{-1}) and the radius of curvature (as any radius), the dimensions of length $[l]$, say cm.

The notion of the radius of curvature (and hence the notion of curvature) of a curve may be introduced in a more geometrical manner. To this end we draw normals MQ and M_1Q (perpendicular lines) to the tangents to two close-lying points M and M_1 on Γ (by Q we have denoted the point of intersection of the normals; see Figure 7.10.3). The angle MQM_1 between the normals is equal to the angle $\Delta\alpha$ between the tangents to points M and M_1 (in accord with a familiar theorem of geometry on angles with mutually perpendicular sides). From this we can find the distance MQ from the curve to the point of intersection of the normals.

We regard the small portion of the curve as an arc of a circle. A normal to the circle is clearly a radius, and the point of intersection of normals is clearly the center of the circle. If the curve was a circle of radius R, then ds would be equal to $R\,d\alpha$, or $d\alpha/ds = 1/R$; this quantity, $d\alpha/ds$ (the curvature of the circle), is constant for any portion of an arc of the circle. Now we select the arc of a circle in such a manner that it is *as close as possible to* the arc MM_1 ("hugs" the arc). The words "as close as possible" are understood as a requirement that the first three terms in Taylor's expansion of the equation $y = y(x)$ of the curve,

$$y = y(a) + y'(a)(x-a) + \frac{1}{2}y''(a)(x-a)^2 + \frac{1}{6}y''(a)(x-a)^3 + \ldots \qquad (7.10.4)$$

(we assume that the point M on the curve corresponds to the value $x = a$ of the abscissa), coincide with the first

three terms in Taylor's expansion of the equation $y = Y(x)$ of the circle,

$$y = Y(a) + Y'(a)(x-a) + \frac{1}{2}Y''(a)(x-a)^2 + \frac{1}{6}Y''(a)(x-a)^3 + \ldots \qquad (7.10.4a)$$

(see formula (6.1.18)).[7.29] In other words, we require that not only the values $y(a)$ and $Y(a)$ coincide but so do the first two derivatives of $y(x)$ and $Y(x)$ at point $x = a$:

$$y(a) = Y(a), \quad y'(a) = Y'(a), \quad y''(a) = Y''(a). \qquad (7.10.5)$$

But by virtue of (7.10.3), all this guarantees that at point M the curvatures of the curve Γ with equation $y = y(x)$ and of the circle Σ with equation $y = Y(x)$ coincide. This means that the radius $R = MQ$ of circle Σ is expressed by formula (7.10.3a).

The circle Σ that passes through point M of the curve Γ and is the closest to Γ is known as the ***osculating circle*** of the curve.[7.30] The radius R of this circle is the ***radius of curvature*** of the curve, and the center Q of circle Σ is said to be the center of curvature of the curve (at point M). Thus, the radius of curvature of a curve can be defined as the radius of the osculating circle; the curvature of the curve is the reciprocal of the radius of the osculating circle.

We note also that since by virtue of (7.10.4), (7.10.4a), and (7.10.5), the difference $Y(x) - y(x) = (1/6) \times [Y'''(a) - y'''(a)]\,dx + \ldots$ changes sign as we pass through point M (i.e. as dx changes sign), the osculating circle Σ at point M of the curve Γ *intersects* the curve.[7.31] This makes it possible to describe the osculating circle thus.

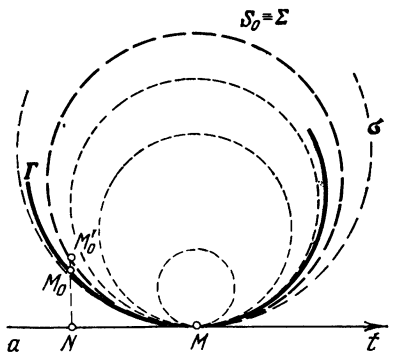


Figure 7.10.4

Let us consider various circles σ that touch point M of Γ and have at point M the same sense of convexity (the direction of bending), in other words, have the same tangent at point M as Γ and have centers that lie on the side of convexity of Γ (i.e. inside the curve) (Figure 7.10.4). Some of these circles lie completely inside Γ (i.e. on the side of convexity of Γ), and their curvature is *greater* than that of Γ. Others lie outside Γ, and their curvature is *smaller* than that of Γ. But there is only one circle that goes from one side of Γ to the other (this happens at point M), and this circle S_0 lies closest to Γ and is called the ***osculating circle*** Σ.

If conditions (7.10.5) are met for a straight line (the tangent t to Γ at point M), then the radius of curvature of Γ at this point (which is calculated via (7.10.3a)) becomes infinite and the curvature k vanishes. Such points M are sometimes called the ***rectifying points*** of curve Γ. It is clear that at

[7.29] A circle in a plane is defined by three numbers (parameters); these may be the coefficients a, b, and r in the equation of the circle $(x-a)^2 + (y-b)^2 - r^2 = 0$. To find these numbers (i.e. to define the circle) we must have *three* conditions, which may be the conditions in (7.10.5).

[7.30] An osculating circle Σ of a curve Γ can also be defined as follows. Consider the circle S that passes through three points M_1, M_2, and M_3 of curve Γ that lie close to point M. When all three points move toward M, circle S tends to the osculating circle Σ of curve Γ at point M.

[7.31] Exceptions in this case are only points where, in addition to conditions (7.10.5), we have $Y'''(a) = y'''(a)$. Examples are the vertices of an ellipse (points where the ellipse intersects the axes of coordinates); at such points, as symmetry considerations suggest, the osculating circle cannot interesct the ellipse (by analogy, all points of a curve Γ that possess this property are called the ***vertices*** of Γ).

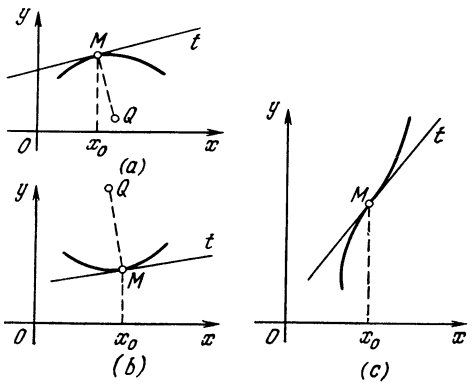


Figure 7.10.5

a rectifying point the curve Γ has no center of curvature Q (this center lies at infinity).

From (7.10.3) it follows that the signs of k and R (up till now we spoke only of the absolute values of k and R and ignored their signs) coincide with the sign of the second derivative y'', that is, characterize the sense of *convexity* of the curve (see Section 2.7). If the curvature is negative, that is, at point M we have a negative y'', the curve at this point lies below the tangent at point M, that is, the curve is *convex* (or *convex upward*; Figure 7.10.5*a*); the center of curvature Q then lies *below* point M. If the curvature is positive, that is, $y'' > 0$, the curve lies above the tangent and is said to be *concave* (*convex downward*; Figure 7.10.5*b*); in this case the center of curvature Q lies above point M. Finally, if the curvature at point M vanishes, that is, $y'' = 0$, then, as we already know, the curve at point M changes its sense of convexity; to one side of M the curve is convex upward and to the other it is convex downward (Figure 7.10.5*c*). At such a point the curve passes from one side of the tangent to the other, changes its sense of convexity (the direction of bending), and the point is called a ***point of inflection.*** At points of inflection, not only the first derivatives of the equation $y = y(x)$ of curve Γ and of the equation $y = Y(x)$ of tangent t coincide, but so do the second derivatives, $y'' = Y'' = 0$, that is, at such points the tangent t obeys the conditions (7.10.5); in other words, at a point of inflection M the *tangent* t acts as the osculating circle Σ, or M is a rectifying point of Γ.

For a straight line l, which of course can also be considered as a special case of a curve, the angle α between the tangent to l (which is simply l) and the x axis does not vary; this means that the curvature k of a straight line is identically zero (l is simply not curved in any way). For a circle S, the angle α changes, but the rate $k = d\alpha/ds$ at which α changes along the circle remains constant; in other words, we may say that circles (and straight lines, for that matter) are curves of constant curvature. It can be shown that these are the only types of such curves.

Suppose that $M(x, y, (x))$ is a point of a curve Γ. The tangent t to the curve at point M has a slope y' $(=\tan\alpha)$, and the normal n is perpendicular to t, that is, forms an angle $\beta = \alpha + 90°$ with the x axis (see Figure 7.10.3); the slope of the normal is $\tan\beta = -\cot\alpha = -1/y'$. Therefore, the (directed) projections MP and PQ of the axes of coordinates of the line segment MQ of length R are, respectively, $p = R\cos\beta = -R\sin\alpha$ and $q = R\sin\beta = R\cos\alpha$, which means that the coordinates of the center of curvature Q are

$$\begin{aligned} \xi &= x + p = x - R\sin\alpha, \\ \eta &= y + q = y + R\cos\alpha. \end{aligned} \tag{7.10.6}$$

If now we take into account formula (7.10.3a) for R and the obvious fact that

$$\cos\alpha = \frac{1}{\sqrt{1+\tan^2\alpha}} = \frac{1}{\sqrt{1+(y')^2}},$$

$$\sin\alpha = \frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}} = \frac{y'}{\sqrt{1+(y')^2}},$$

we finally get

$$\xi = x - \frac{\sqrt{[1+(y')^2]^3}}{y''}\,\frac{y'}{\sqrt{1+(y')^2}},$$

$$\eta = y + \frac{\sqrt{[1+(y')^2]^3}}{y''}\,\frac{1}{\sqrt{1+(y')^2}},$$

that is,

$$\xi = x - \frac{1+(y')^2}{y''}\,y', \quad \eta = y + \frac{1+(y')^2}{y''}. \tag{7.10.7}$$

(This agrees with the condition that $\eta > y$ for $k > 0$, or for $y'' > 0$, or the condition that $\eta < y$ for $k < 0$, or for $y'' < 0$.)

If the curve is specified parametrically, $x = x(t)$ and $y = y(t)$, we can write (see (7.10.3b))

$$\xi = x - \frac{(x')^2+(y')^2}{x'y''-x''y'}\,y', \qquad \eta = y + \frac{(x')^2+(y')^2}{x'y''-x''y'}\,x' \tag{7.10.7a}$$

(see Exercise 7.10.4).

The locus of the centers of curvature $Q(\xi, \eta)$ of a curve Γ constitutes a new curve γ known as the ***evolute*** of curve Γ. The curve Γ with respect to its own evolute γ is called the ***involute***.[7.32]

Example 1. The parabola $y = x^2$. Here $y' = 2x$ and $y'' = 2$ (a constant), whereby the curvature k and the radius of curvature R of the parabola at point $M(x, y)$ $(=M(x, x^2))$ are

$$k = \frac{2}{\sqrt{(1+4x^2)^3}}, \qquad R = \frac{\sqrt{(1+4x^2)^3}}{2}.$$

The center of curvature $Q(\xi, \eta)$ of the parabola with respect to point $M(x, x^2)$ has coordinates

$$\xi = x - \frac{1+4x^2}{2}\cdot 2x = -4x^3,$$

$$\eta = x^2 + \frac{1+4x^2}{2} = 3x^2 + \frac{1}{2},$$

that is,

$$\xi = -4\sqrt{\left(\frac{\eta-1/2}{3}\right)^3} = -\frac{4}{3\sqrt{3}}\sqrt{(\eta-1/2)^3}, \text{ or } \xi = c\eta_1^{3/2}, \tag{7.10.8}$$

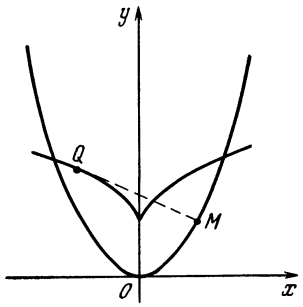


Figure 7.10.6

where $c = -4/3\sqrt{3}$ and $\eta_1 = \eta - 1/2$.

Thus, the evolute of $y = x^2$ is the *semicubical parabola* (7.10.8) whose cusp coincides with point (0, 1/2) and whose axis of symmetry is the axis of ordinates (Figure 7.10.6).

Example 2. The cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$. We have $x' = a(1 - \cos t)$, $y' = a \sin t$, $x'' = a \sin t$, and $y'' = a \cos t$ (the primes denote derivatives with respect to parameter t). Whence,

$$(x')^2 + (y')^2 = a^2(1-\cos t)^2 + a^2\sin^2 t$$
$$= a^2(2 - 2\cos t) = 4a^2\sin^2\frac{t}{2},$$
$$y''x' - x''y' = a\cos t \cdot a(1-\cos t) - (a\sin t)^2 = a^2(\cos t - 1)$$
$$= -2a^2\sin^2\frac{t}{2}.$$

Thus, in view of (7.10.7a) we have

$$\xi = a(t - \sin t) + 2a\sin t = a(t + \sin t), \tag{7.10.9}$$
$$\eta = a(1 - \cos t) - 2a(1 - \cos t) = -a + a\cos t.$$

To understand what curve is represented by these equations, we introduce a new parameter, $t_1 = t + \pi$ (or $t = t_1 - \pi$); since $\sin t = -\sin t$ and $\cos t = -\cos t_1$, we have

$$\xi = a(t_1 - \sin t_1) - a\pi, \qquad \eta = a(1 - \cos t_1) - 2a. \tag{7.10.10}$$

[7.32] The word "involute" is the Latin for "unwinding", while the word "evolute" is the Latin for "unwound," that is the curve that is being unwound; these names are related to the properties of evolutes and involutes discussed at the end of this section (see the text in small print).

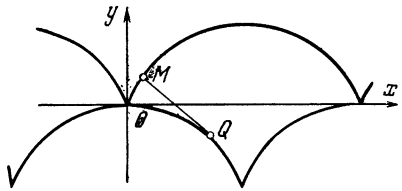


Figure 7.10.7

Thus, the evolute (7.10.9) (or (7.10.10)) of a cycloid is also a *cycloid*, $\xi_1 = a(t_1 - \sin t_1)$ and $\eta_1 = a(1 - \cos t_1)$, where $\xi_1 = \xi + a\pi$ and $\eta_1 = \eta + 2a$, only shifted in relation to the initial cycloid by $2a$ units downward ($2a$ is the diameter of the generating cirle) and πa units to be left (πa is the half-width of the arch of the initial cycloid) (Figure 7.10.7).

To discuss the properties of evolutes and involutes in detail, we must first differentiate in (7.10.6):

$$d\xi = dx - R\cos\alpha\, d\alpha - \sin\alpha\, dR,$$

$$d\eta = dy - R\sin\alpha\, d\alpha + \cos\alpha\, dR.$$

But since (see (7.9.2b))

$$dx = \cos\alpha\, ds = \cos\alpha \frac{ds}{d\alpha} d\alpha = R\cos\alpha\, d\alpha,$$

$$dy = \sin\alpha\, ds = \sin\alpha \frac{ds}{d\alpha} d\alpha = R\sin\alpha\, d\alpha,$$

we have

$$d\xi = -\sin\alpha\, dR, \quad d\eta = \cos\alpha\, dR. \tag{7.10.11}$$

Dividing one by the other, we get

$$\frac{d\eta}{d\xi} = -\cot\alpha = \tan(\alpha + 90°). \tag{7.10.12}$$

From this it follows that a tangent to evolute γ of a curve Γ (i.e. to the locus of the centers of curvature $Q(\xi, \eta)$) forms an angle $\beta = \alpha + 90°$ with the x axis, that is, coincides with the *normal* to Γ. And since point Q belongs to the normal n to curve Γ at point M, the respective tangent to γ coincides with n, that is, *tangents to the evolute γ of a curve Γ coincide with the normals to Γ* (Figure 7.10.8).[7.33]

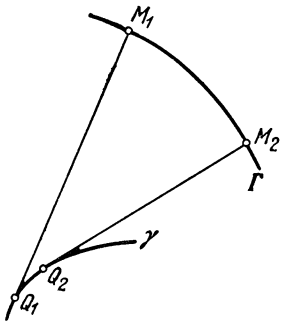


Figure 7.10.8

On the other hand, from (7.10.11) it follows that

$$d\xi^2 + d\eta^2 = \sin^2\alpha\, dR^2 + \cos^2\alpha\, dR^2 = dR^2,$$

$$\text{i.e.} \quad d\sigma = dR, \tag{7.10.12a}$$

where σ is the length of the arc of evolute γ of curve Γ (both $d\sigma$ and dR are assumed positive).

The last remark has the following meaning. Let us consider an arc M_1M_2 of curve Γ such that on it the derivative $R' = dR/ds$ retains its sign. We will agree to reckon the arc length along Γ in such a manner that the derivative of R is *positive*; in other words, we assume that the arc length is reckoned along Γ in the direction in which the radius of curvature R *grows*. Then the increment $M_2Q_2 - M_1Q_1 = R_2 - R_1 = \Delta R$ of the radius of curvature along arc M_1M_2 is equal to the length $\Delta\sigma$ of the evolute arc. And since the line segments M_1Q_1 ($=R_1$) and M_2Q_2 ($=R_2$) are tangents to evolute γ of curve Γ (see Figure 7.10.8), the process of reconstructing the involute Γ from its evolute γ resembles the unwinding under tension of a nonexpandable string from a spool shaped like γ; the end M of the string describes curve Γ. That is the origin of the names *evolute* and *involute* (see footnote 7.32).

By way of an example, let us construct the (parametric) equation of the *involute of a circle* of radius r and equation $x^2 + y^2 = r^2$, or $x = r\cos t$ and $y = r\sin t$, where t is the parameter (angle). Let us assume that the involute passes through point $A(r, 0)$ and that the unwinding of the string wound on the round spool occurs counterclockwise (Figure 7.10.9). The tangent MT to the circle at point $M(r\cos t,\ r\sin t)$ forms an angle $\alpha = t + 90°$ with the x axis, and the line segment MT of length rt along the tangent (rt is the length of arc AM of the circle) has projections $rt\cos\alpha = -rt\sin t$ and $rt\sin\alpha = rt\cos t$ on the axes of coordinates. This readily yields the following parametric expressions for the coordinates X, Y of point $T(X, Y)$ (i.e. the equation of the involute of the circle):

$$X = r\cos t + rt\sin t,$$

$$Y = r\sin t - rt\cos t. \tag{7.10.13}$$

The theory of evolutes and involutes was first developed by Christian Huygens in his

[7.33] In this connection it is often said that evolute γ of curve Γ is the *envelope of the normals* to Γ.

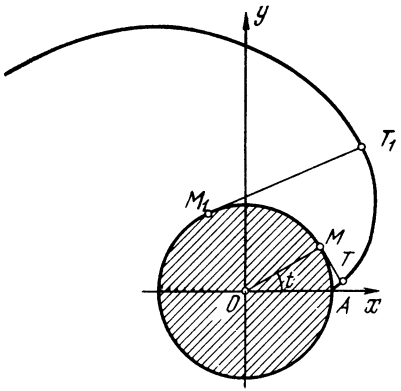


Figure 7.10.9

Horologium Oscillatorium (1673) in connection with the following problem. We wish to construct the equation of motion of a pendulum, that is, a material particle of mass m (the bob) moving along a curve γ:

$$m\frac{d^2s}{dt^2} = mg\sin\alpha, \qquad (7.10.14)$$

where s is the arc length of curve γ and t is time, so that d^2s/dt^2 is the (*tangenrtial*, i.e. directed along the tangent to γ) *acceleration* of the particle, g is the acceleration of gravity, α is formed by the tangent l to curve γ and the horizontal, and $mg\sin\alpha$ is the (vertical) projection of the force of gravity mg on l (Figure 7.10.10; see Chapters 9 and 10). The period of oscillations of the pendulum is, strictly speaking, dependent on the swing (i.e. on the uppermost point on curve γ at which the pendulum starts oscillating). Huygens posed the question of what shape should curve γ have so that the pendulum bob swinging along this curve would take exactly the *same* time to complete swings of large and of small amplitude. By analyzing the differential equation (7.10.14) (which was not a simple task because at that time there was no differential and integral calculus) Huygens found that γ must be a *cycloid*. But how do we make a pendulum move along a cycloid? It was in this connection that Huygens developed the general theory of involutes and found that the evolute of a cycloid is also a cycloid

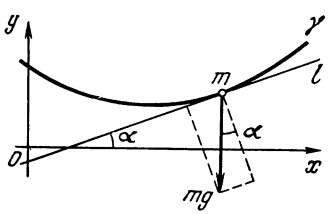


Figure 7.10.10

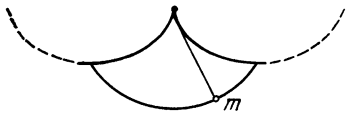


Figure 7.10.11

(see Figure 7.10.7 and Eq. (7.10.10)). He also demonstrated that if at the point of suspension of the string of the pendulum certain "lips" in the form of a cycloid are attached, so that in the process of oscillations the string winds about them, then the end of the string (with the bob) of constant length describes a cycloid (Figure 7.10.11). The first pendulum clock constructed by Huygens had just such "lips"; however, later he dropped this idea because he found that the "lips" had practically no effect on the precision of the clock, since in view of the smallness of angles α in Eq. (7.10.14) we can always replace $\sin\alpha$ with α and the equation $m(d^2s/dt^2) = mg\alpha$ implies that the period of oscillations of a pendulum of *constant length* l is constant (see Section 10.3).

Exercises

7.10.1. Calculate the curvature (in an arbitrary point of the curve) of (a) an ellipse $x = a\cos t$ and $y = b\sin t$, (b) a hyperbola $y = 1/x$, and (c) a fourth-order parabola $y = ax^4$.

7.10.2. Find the evolutes of (a) an ellipse $x = a\cos t$ and $y = b\sin t$, and (b) a hyperbola $y = 1/x$.

7.10.3. Prove that the points of maxima or minima in the curvature of a curve correspond to the cuspidal points of the evolute of this point (see Figures 7.10.6 and 7.10.7).

7.10.4. Prove the validity of formulas (7.10.7a).

7.11 Solid Geometry Applications of Integral Calculus

In Section 3.6 we obtained the formula

$$V = \int_a^b S(x)\,dx, \qquad (7.11.1)$$

where $S(x)$ is the area of a section of a solid by a plane $x =$ constant perpendicular to the x axis (we advise the reader to repeat the derivation of this formula). This formula was used to obtain an expression for the volume of a *pyramid*. The volume of a *cone* was obtained in exactly the same manner. Place the origin of the system of coordinates at the center of the circle of the

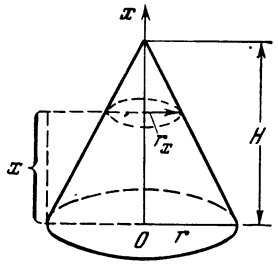


Figure 7.11.1

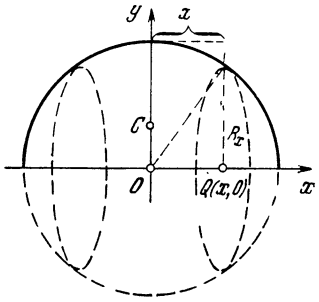


Figure 7.11.2

cone base and send the x axis along the altitude of the cone upward (Figure 7.11.1). Let $S(x)$ be the area of the section of the cone by a plane perpendicular to the altitude and distant x from the base. This section is a circle of radius r_x. From the similarity of triangles we have $r_x/r = (H - x)/H$, where r is the radius of the base and H is the altitude of the cone. Whence, $r_x = (r/H)(H - x)$ and $S(x) = \pi r_x^2 = \pi r^2 (H - x)^2/H^2$ and, consequently,

$$V = \int_0^H \pi \frac{r^2}{H^2} (H-x)^2\, dx$$

$$= -\frac{\pi r^2}{H^2} \frac{(H-x)^3}{3}\Big|_0^H = \frac{\pi r^2 H^3}{3H^2} = \frac{\pi r^2 H}{3}.$$

To obtain the volume of a *sphere* of radius R, we place the origin at the center of the sphere (Figure 7.11.2). The section cut by a plane perpendicular to the x axis and distant x from the origin is a circle of radius R_x. By the Pythagorean theorem $R_x = \sqrt{R^2 - x^2}$ and so $S(x) = \pi R_x^2 = \pi (R^2 - x^2)$, whence

$$V = \int_{-R}^{R} \pi (R^2 - x^2)\, dx$$

$$= \pi \left(R^2 x - \frac{x^3}{3}\right)\Big|_{-R}^{R} = \frac{4}{3}\pi R^3.$$

Formula (7.11.1) results in ***Cavalieri's theorem***,[7.34] which says that *if two solids have equal altitudes and if sections made by planes P and Q parallel to the bases and at equal distances to them always have a given ratio, the volumes of the two solids have this given ratio to each other.* Indeed, in calculating the volumes of such solids using formula (7.11.1), where the x axis is perpendicular to planes P and Q, we see that the ratio of the volumes is the same as the ratio of the integrands $S(x)$.

Let a solid be generated by revolving the curvilinear trapezoid depicted in Figure 7.11.3 about the x axis. In this case, the section is a circle of radius $y = f(x)$ and $S(x) = \pi y^2$. Using (7.11.1), we find the well-known formula for the *volume of a solid of revolution*:

$$V = \pi \int_a^b y^2\, dx. \qquad (7.11.2)$$

Let us find, say, the volume of a solid generated by the revolution of the upper half of the ellipse $x^2/a^2 + y^2/b^2 = 1$ about the x axis (make a drawing). This solid is called an ***ellipsoid of revolution***. Since for an ellipse $y =$

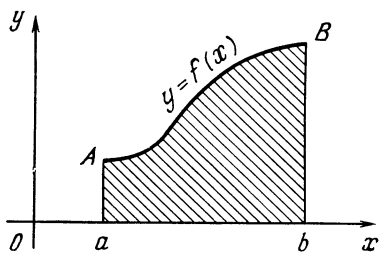


Figure 7.11.3

[7.34] This theorem played an important role in formulating the concept of the integral. Bonaventura ***Cavalieri*** (1598-1647), a pupil of Galileo, gave this theorem (without proof) in his *Geometria Indivisibilibus* (The Geometry of Indivisibles) (1635).

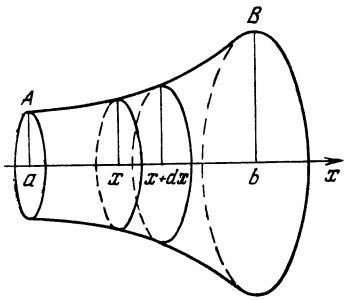


Figure 7.11.4

$(b/a)\sqrt{a^2-x^2}$, formula (7.11.2) yields

$$V=\pi\int_{-a}^{a}\frac{b^2}{a^2}(a^2-x^2)\,dx$$

$$=\frac{\pi b^2}{a^2}\left(a^2x-\frac{x^3}{3}\right)\Big|_{-a}^{a}=\frac{4}{3}\pi ab^2.$$

For $a=b=R$ we obtain, as expected the volume $(4/3)\pi R^3$ of a sphere of radius R.

Now let us derive the formula for the *surface area of a solid of revolution* (Figure 7.11.4). We consider a solid bounded by sections passing through the points x and $x+dx$. Denote by dF the lateral surface area of this solid. Regarding it as the frustrum of a cone, we get

$$dF\simeq\pi\,[y(x)+y(x+dx)]\,ds,$$

where ds is the length of the small portion of the curve by revolving which we get the surface of the solid; $ds=\sqrt{1+(y'(x))^2}\,dx$ (see Section 7.9). The sum $y(x)+y(x+dx)$ can be replaced with $2y(x)$, disregarding the quantity $y'(x)\,dx$ as compared with $y(x)$.[7.35] Therefore

$$dF\simeq 2\pi y(x)\sqrt{1+(y'(x))^2}\,dx.$$

The entire surface area of the initial solid of revolution is

$$F=2\pi\int_a^b y(x)\sqrt{1+(y'(x))^2}\,dx. \quad (7.11.3)$$

[7.35] Note that in the expression dF of the sum $y(x)+y(x+dx)$ is multiplied by ds, so that the quantity we ignore is of the order of $dx\,ds\simeq dx^2$.

The surface area of a *sphere* is readily found by means of this formula. Indeed, a sphere is generated by the revolution of the upper semicircle about the x axis. The equation of a circle is $x^2+y^2=R^2$, whence

$$y=\sqrt{R^2-x^2},\quad y'=-\frac{x}{\sqrt{R^2-x^2}}. \quad (7.11.4)$$

Substituting into (7.11.3) yields

$$F=2\pi\int_{-R}^{R}\sqrt{R^2-x^2}\,\frac{R}{\sqrt{R^2-x^2}}\,dx$$

$$=2\pi Rx\Big|_{-R}^{R}=4\pi R^2.$$

Moreover, from (7.11.3) and (7.11.4) we can easily find the surface of a *spherical layer* (*spherical zone*) cut out of a sphere by two parallel planes $x=a$ and $x=b>a$ (see Figure 7.11.2 if, say, $b=R$, the layer becomes a *spherical segment* or *spherical cap*). Indeed, we have

$$F=2\pi\int_a^b\sqrt{R^2-x^2}\,\frac{R\,dx}{\sqrt{R^2-x^2}}$$

$$=2\pi Rx\Big|_a^b=2\pi R(b-a).$$

Thus, the sought-for surface area depends only on the altitude $b-a$ of the layer (segment) and the radius R of the sphere, which is really a remarkable and beautiful result.

Formulas (7.11.2) and (7.11.3) can be written also in a different way. Suppose the object depicted in Figure 7.11.3 is a metal plate of constant thickness. In this case the mass M of this plate is

$$M=\mu\int_a^b y\,dx=\mu S, \quad (7.11.5)$$

where μ is the density of the material of the plate (density per unit surface area; see Section 9.13), and $S=\int_a^b y\,dx$ is the surface area of the plate. The distance y_C from the center

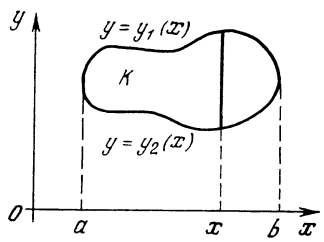


Figure 7.11.5

of gravity to the x axis is expressed by the formula

$$y_C = \frac{1}{2M}\int_a^b \mu y^2\,dx = \frac{1}{2S}\int_a^b y^2\,dx \qquad (7.11.6)$$

(see (9.13.12)). Comparing (7.11.2) and (7.11.6), we see that

$$V = 2\pi\, Sy_C = 2\pi y_C S, \qquad (7.11.7)$$

or that *the volume V of a solid revolution is equal to the area S of the figure whose rotation generates the solid multiplied by the circumference $2\pi y_C$ of the circle described in the process of rotation by the center of gravity of the figure.* (Here the center of gravity, or centroid, of a flat figure is assumed to coincide with the center of gravity of a homogeneous plate whose shape would be that of the figure.) This theorem is known as the ***(first) theorem of Pappus***.[7,36]

Formula (7.11.7) and the theorem of Pappus can be modified so that they incorporate the case of a solid generated by the rotation of a figure of arbitrary shape, and not only a curvilinear trapezoid.

Take, for example, the solid generated by the rotation about the x axis of the figure depicted in Figure 7.11.5. In this case the section of the solid by a plane perpendicular to the x axis is a circular ring (annulus) bounded by the circles of radii $y_1(x)$ and $y_2(x)$. Hence

[7] [36] ***Pappus*** of Alexandria (end of 3rd century A.D.) was the last of the great Greek mathematicians; in his works he set forth (usually without any proof) many results obtained by his predecessors and also by himself. The fact that the theorems of Pappus are often called the ***Pappus-Guldin theorems*** or even simply ***Guldin's theorems*** in the literature is completely groundless. Both theorems were formulated by Pappus, while their proofs, put forward by Paul ***Guldin*** (1577-1643), a Swiss monk and an amateur mathematician, were highly doubtful and substantially weaker than the proofs of the theorems given by B. Cavalieri and the famous Johann Kepler (1571-1630) (the latter proofs were harshly, but unjustly, criticized by Guldin).

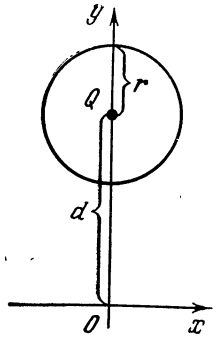


Figure 7.11.6

$$S(x) = \pi[y_1(x)]^2 - \pi[y_2(x)]^2 = \pi(y_1^2 - y_2^2)$$

and, therefore, the volume of the solid of revolution is

$$V = \pi\int_a^b (y_1^2 - y_2^2)\,dx. \qquad (7.11.8)$$

On the other hand, the distance y_C from the x axis to the center of gravity (centroid) C of a homogeneous plate shaped as K in Figure 7.11.5 is

$$y_C = \frac{1}{2S}\int_a^b (y_1^2 - y_2^2)\,dx = \frac{\int_a^b (y_1^2 - y_2^2)\,dx}{2\int_a^b (y_1 - y_2)\,dx} \qquad (7.11.9)$$

(see 9.13.11)). Hence here also $V = 2\pi y_C S$.

Formula 7.11.3) can be interpreted in a similar manner. Suppose that the surface of a solid of revolution is generated by the rotation of an arc AB (see Figure 7.11.4) and imagine that the arc is a homogeneous heavy string with a constant density σ (mass per unit length; see the text between formulas (9.12.1) and (9.12.2)). In this case the distance y_C from the center of gravity of the string C to the x axis is given by the following formula:

$$y_C = \frac{1}{M}\int_{s_0}^{s_1} \sigma y(x)\,ds = \frac{1}{L}\int_{s_0}^{s_1} y(x)\,ds$$

$$= \frac{1}{L}\int_a^b y(x)\sqrt{1+(y'(x))^2}\,dx; \qquad (7.11.10)$$

here $x(s_0) = a$, $x(s_1) = b$, ds is the element of length of arc AB, $L = \int_{s_0}^{s_1} ds = \int_a^b \sqrt{1+(y'(x))^2}\,dx$

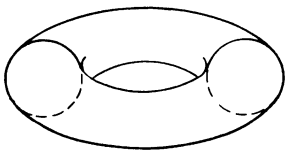


Figure 7.11.7

is the length of the string, and $M = \sigma L$ is the mass of string (see formulas (9.13.4) and (9.13.5b)).

Comparing (7.11.3) and (7.11.10), we get

$$F = 2\pi L y_C = 2\pi y_C L, \qquad (7.11.11)$$

which means that *the surface area of a solid of revolution is the product of the length L of the arc whose rotation generates the surface by the circumference $2\pi y_C$ of the circle described by the centroid of the curve in the process of rotation.* This result (known as the ***second theorem of Pappus***) can be generalized so as to incorporate the case of a curve of arbitrary shape (say the one that is the boundary of figure K in Figure 7.11.5) and not only arc AB (see Figure re 7.11.3), which is the boundary of a curvilinear trapezoid.

The theorems of Pappus (7.11.7) and (7.11.11) often prove very useful. For instance, let us find the *volume* and *surface area of a torus (anchor ring)*, which is generated by the rotation about the x axis of a circle of radius r whose center is d distant from the axis of rotation (Figure 7.11.6). Here, obviously, the center of gravity C of the circle (and of the circumference of the circle) coincides with the (geometrical) center Q of the circle. Hence, the volume of the torus in Figure 7.11.7 is

$$V = 2\pi d \times \pi r^2 = 2\pi^2 r^2 d, \qquad (7.11.12)$$

and the surface area is

$$F = 2\pi d \times 2\pi r = 4\pi^2 r d. \qquad (7.11.13)$$

The theorems of Pappus can be used in the reverse direction, so to say. The sphere in Figure 7.11.2 can be obtained by rotating a semicircle about the diameter that is part of the boundary of this semicircle. Then, if the radius of the sphere is R, the volume will be

$$V = 2\pi y_C \frac{\pi R^2}{2} = \pi^2 R^2 y_C.$$

Since we know that the volume V of a sphere is $(4/3)\,\pi R^3$, we can find the distance y_C from the *center of gravity of a semicircle* to the diameter:

$$y_C = \frac{4}{3}\pi R^3/(\pi^2 R^2) = \frac{4R}{3\pi} \simeq 0.4R.$$

Similarly, using the second theorem of Pappus, we can find the *center of gravity of a semicircumference* (see Exercise 7.11.3).

Exercises

7.11.1. Find the volume of a cone using the fact that the cone is a solid generated by rotating a right triangle about one of its legs.

7.11.2. Find the volume of a solid generated by rotating a figure bounded from above by the curve $y = \sqrt{x}$, from below by the x axis, and on the right by the vertical $x = 2$ about the x axis.

7.11.3. Using the second theorem of Pappus, find the center of gravity of a semicircumference.

7.12 Curve Sketching

The most primitive method of constructing the graph of a function $f(x)$ is to compute the value of $f(x_n)$ for a large number of points x_n. The usual procedure is to choose points x_n in the form $x_n = x_0 + na$, with $n = 0, \pm 1, \pm 2, \ldots$. This, clearly, is quite an extravagant method. In order to see the variation of the function on an interval Δx, we have to choose a step a much less than Δx, or $a \ll \Delta x$. And if the step (subinterval) is small, a very large number of points are required to embrace the whole range we are interested in. And yet we encounter the problem of constructing curves very often, since knowing the graph of a function provides extensive information about the properties of the function; for one, it immediately gives the number of real roots of the equation $f(x) = 0$, gives the intervals within which these roots lie, shows the maxima and minima of the function.

The techniques considered in Sections 7.1 and 7.2 make it possible to construct graphs much faster and more reliably and to gain a general picture of the shape of the curve. This requires, first of all, that we find the characteristic points of the graph—maxima, minima, discontinuities, salient points, points of inflection, etc.

Let us illustrate this fact by using the example of the graph of a third-degree *polynomial*,

$$y = ax^3 + bx^2 + cx + d. \qquad (7.12.1)$$

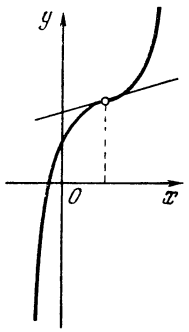


Figure 7.12.1

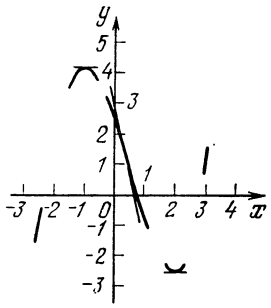


Figure 7.12.2

To be specific, we wish to construct the graph of the function

$$y = 0.5x^3 + 0.75x^2 - 3x + 2.5. \qquad (7.12.2)$$

First we find the *maxima* and *minima*. Equating to zero the derivative of (7.12.2), we have

$$y' = 1.5x^2 - 1.5x - 3 = 0, \qquad (7.12.3)$$

whence we find the two roots of Eq. (7.12.3): $x_1 = -1$ and $x_2 = 2$.

Let us investigate each of these values separately. To do this, we find y''. We see that $y'' = 3x - 1.5$, and $y''(-1) = -4.5 < 0$ and $y''(2) = 6 - 1.5 = 4.5 > 0$. This means that at $x = -1$ the function has a maximum,

$$y_{\max} = -0.5 - 0.75 + 3 + 2.5 = 4.25,$$

while at $x = 2$ the function has a *minimum*, $y_{\min} = -2.5$.

Now let us see how the polynomial behaves for *very large* x (in absolute value). Note that for very large x the term containing x^3 will appreciably exceed the other terms in absolute value. Therefore, the sign of the polynomial (7.12.2) is determined by the sign of $0.5x^3$ (for large absolute values of x): for $x \gg 0$ we have $y \gg 0$, that is, the right-hand branch of the curve goes up, while for $x \ll 0$ we have $y \ll 0$, that is, the left branch goes down. (It is clear that if a in (7.12.1) is negative, the left branch goes up and the right branch goes down.)

Let us find the *points of inflection*, that is, points, where $f''(x) = 0$ (see Section 7.10). From (7.12.3) it follows that $y'' = 3x - 1.5$, that is, $y'' = 0$ only at $x = 0.5$. Note that the graph of a third-degree polynomial always has a point of inflection, and it is unique (at this point the tangent to the graph intersects the graph; see Figure 7.12.1). Indeed, if y is a third-degree polynomial, then the equation $y'' = 0$ is a first-degree equation. It always has a unique root, x_0. At this point the difference $y - \tilde{y}$, where $\tilde{y}$ is the ordinate of the tangent to the curve, has the form $A(x - x_0)^3$, that is, changes sign when x passes through the value x_0. This means that at this point the tangent is sure to intersect the curve.

We return to the construction of the graph and compute the ordinate y of the point of inflection to get $y = 0.875$. Let us also determine the direction of the tangent to the curve at the point of inflection. Using (7.12.3), we get $\tan \alpha = y'(0.5) = 3.375$. Using all the foregoing arguments, we get the shape of the curve (7.12.2) depicted in Figure 7.12.2.

Of course, if we do not compute any other values of the function, the resulting graph will give only a very rough qualitative idea of the behavior of the function, but even such a graph enables us to count the number of roots (i.e. the number of points of intersection of the graph with the x axis) and to draw certain conclusions about their

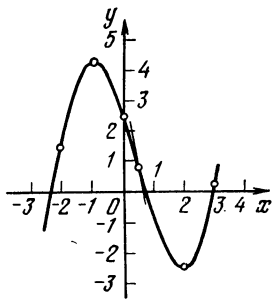


Figure 7.12.3

values. In our example, Figure 7.12.2, we see that there are three roots, that one of the roots lies somewhere between 0.5 and 2, that the second root must definitely be positive (it even exceeds 2), while the third root is negative (it is less than —1).

The graph may be improved by computing a few more values of the function for certain values of x. For example, let us compute three more values of the function in this example. For $x = 0$ we have $y = 2.5$. This permits us to get a better picture of the variation of the curve between maximum and minimum. For $x = 3$ we have $y = 0.25$. We computed this value so as to get an idea of the rate of climb of the right branch of the curve. Similarly, to get an idea of the rate of fall of the left-hand branch of the curve we take $x = -2$ and get $y = 1.5$. Using these values, we obtain the curve shown in Figure 7.12.3.

With this graph we can draw more accurate conclusions concerning the roots: one root lies between $x = 0.5$ and $x = 1$, the second between $x = 2$ and $x = 3$ (closer to 3), and the third is less than $x = -2$ (its value is most likely close to $x = -2.5$).

It may happen that after equating the derivative to zero we will not obtain any real roots. This will mean that the polynomial does not have a maximum or a minimum. Since all that has been said about the behavior of the polynomial for very large absolute values of x remains valid, the graph will intersect the x axis only at one point (the polynomial has one real root; see Figure 7.12.1).

Finally, the derivative of the polynomial may have only one (double) root x_0. Then the derivative will be of the form

$$y' = A(x - x_0)^2, \tag{7.12.4}$$

whence, integrating, we get

$$y = \frac{A}{3}(x - x_0)^3 + C. \tag{7.12.5}$$

From (7.12.5) we see that in this case the polynomial differs from a perfect cube only in a constant summand. It is clear that y has neither a maximum nor a minimum (see Example 1a in Section 7.1). The graph of this function intersects the x axis at one point, which can be found by equating y to zero:

$$\frac{A}{3}(x - x_0)^3 + C = 0,$$

that is,

$$(x - x_0)^3 = -\frac{3C}{A}, \quad x = x_0 - \sqrt[3]{\frac{3C}{A}}. \tag{7.12.6}$$

Finding the maximum and minimum of a third-degree polynomial (7.12.1) and, hence, the investigation of its graph can always be completed because by equating the derivative to zero we get a *quadratic* equation whose roots are not hard to find. (In general, however, the situation is not all that simple.) A polynomial (of any degree) is preferable over all other functions in that we can always find its value for any value of x and its graph has no discontinuities or corners; when x increases without bound in absolute value, the absolute value of y also increases without bound (and the higher the degree of the polynomial, the greater the rate of this growth). But for an arbitrary function (even if it is algebraic) the situation may be quite different.

In the general case, the construction of a graph requires studying the behavior of the function at infinity, that is, the behavior of the function as $x \to -\infty$

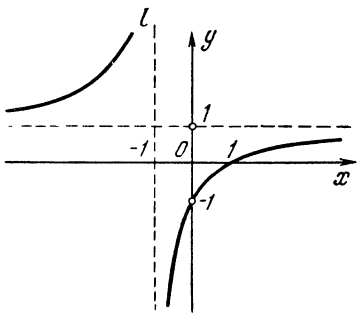


Figure 7.12.4

and $x \to \infty$. If, say, the function comes closer and closer (in value) to a fixed number C as x increases without bound, the function has a *horizontal asymptote* $y = C$, that is, as x grows, the graph comes closer and closer to the straight line $y = C$. Similarly, if at $x = a$ the function undergoes a discontinuity and, say, grows without bound as x approaches a, the straight line $x = a$ is a *vertical asymptote* of the graph of the function $y = f(x)$, that is, the graph moves closer and closer to the straight line $x = a$ as x approaches a. It is also clear that the construction of a graph must begin by determining the domain of the function (the range of the independent variable); in addition, we must determine whether or not the function is *even* or *odd* (it may be neither) (see Section 1.7), since in the case, say, of an even function we need only construct the branch of the graph corresponding to positive values of x and then apply symmetry considerations.

Let us illustrate what we have just said by several simple examples. Take the function

$$y = \frac{x-1}{x+1}. \tag{7.12.7}$$

It is clear that this function is defined for all values of x except $x = -1$, where the denominator vanishes and the function undergoes a discontinuity. Since as x approaches -1 the absolute values of y increase without bound, the straight line l whose equation is $x = -1$ is the *vertical asymptote* of the graph of the function. It is also easy to see that near $x = -1$ the value of y is positive to the left of l and negative to the right of l (Figure 7.12.4).

Further details of the behavior of the function are related to the fact that the values of

$$y = 1 - \frac{2}{x+1} \tag{7.12.8}$$

for large absolute values of x will get as close to unity as desired (since the fraction $2/(x+1)$ becomes very small in this case). Therefore, the straight line $y = 1$ serves as the *horizontal asymptote* of the graph of the function. Finally, by virtue of (7.12.8),

$$y' = \frac{2}{(x+1)^2}, \quad y'' = -\frac{4}{(x+1)^3},$$

that is, y' is positive for all values of x from the domain of the function, while y'' is negative for $x > -1$ and positive for $x < -1$. Thus, for all values of x (except $x = -1$, for which the function is not defined) our function *increases*; for $x > -1$ the graph is *convex upward*, while for $x < -1$ it is *convex downward*; the graph has no points of inflection (i.e. points where y'' vanishes). It is also important that $y = 0$ only at $x = 1$, which means that the graph intersects the x axis at point (1, 0), while $x = 0$ at $y = -1$, which means that the graph intersects the y axis at point (0, -1) (Figure 7.12.4).

Here are two somewhat more difficult examples. If

$$y = \frac{x^2-2}{x^2-1} \quad \left(= 1 - \frac{1}{x^2-1}\right), \tag{7.12.9}$$

the investigation is simplified by the fact that the function is *even*: the graph is symmetric about the y axis. Our function is defined for all values of x except $x = 1$ and $x = -1$; at these points the function becomes infinite. Thus, the straight lines $x = 1$ and $x = -1$ are the *vertical asymptotes* of the graph of the function. On the other hand, the straight line $y = 1$ is the *horizontal asymptote* (see the expression within the parentheses in

(7.12.9)). It is also clear that the equation $y(x) = 0$ has two roots: $x = \pm\sqrt{2}$, that is, $x \simeq 1.41$ and $x \simeq -1.41$.

If we employ the second form of (7.12.9), we readily see that

$$y' = \frac{2x}{(x^2-1)^2}, \quad y'' = \frac{2}{(x^2-1)^2} - \frac{2\cdot 2x\cdot 2x}{(x^2-1)^3} = -\frac{6x^2+2}{(x^2-1)^3}.$$

Thus, the derivative y' vanishes only at $x = 0$; since $y''(0) = 2 > 0$, the value $x = 0$ corresponds to a (local) minimum in the function, $y_{\min} = y(0) = 2$. For negative x the function decreases and for positive x it increases; it has no points of inflection because $y''(x)$ is never zero; at $x^2 < 1$, that is, for $-1 < x < 1$, the second derivative is positive, or the function is convex downward, while for $|x| > 1$ the function is convex upward. The graph of the function is shown in Figure 7.12.5.

Finally, we take the function

$$y = \sqrt{\frac{x^2-2}{x^2-1}} = \left(\frac{x^2-2}{x^2-1}\right)^{1/2} \left(= \left(1 - \frac{1}{x^2-1}\right)^{1/2}\right), \qquad (7.12.10)$$

whose domain coincides with the domain of positivity of function (7.12.9), that is, the function is defined only for

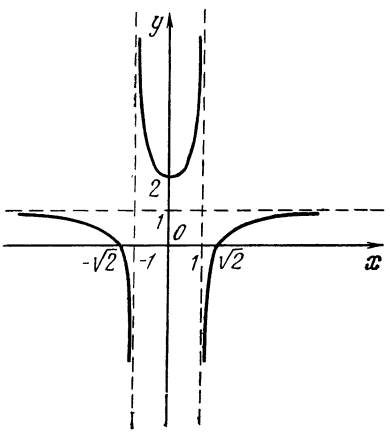


Figure 7.12.5

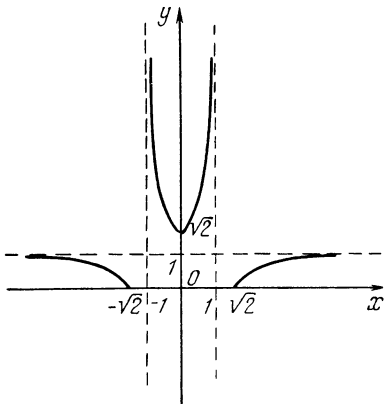


Figure 7.12.6

$|x| < 1$ and $|x| > \sqrt{2} \simeq 1.4$. Since the values of this function can be found by extracting the square root of the values corresponding to the function depicted in Figure 7.12.5, we can immediately construct the graph (Figure 7.12.6). A fuller idea of the behavior of the curve can be obtained by calculating several values of $y(x)$ corresponding to certain points on the curve or the values of $y'(x)$, which give the slopes of the tangents to the curve at various points $(x, y(x))$ (for instance, for the curves in Figures 7.12.5 and 7.12.6 it is advisable to find the slope $y'(x)$ of the tangent to point (1.41, 0)).

Of course, the details with which a specific curve is built depend on the reasons that prompted us to investigate the curve in the first place; in some cases it is quite sufficient to obtain a rough qualitative picture of this behavior, while in others we may wish to obtain a fuller idea of the behavior.

What we have discussed here in connection with curve sketching pertains, of course, to "school" methods of constructing curves that are specified by equations. Quite different possibilities have been opened up by the use of electronic computers. It is possible to program a computer in such a way that it plots the curve of a function on the monitor screen when you key in the various values of the function and those of the independent variable.

Exercises

7.12.1. Find the maxima and minima of the following functions and plot the graphs of these functions: (a) $y = x^3 - 3x^2 + 2$, (b) $y = x^3 - 3x^2 + 3x - 15$, and (c) $y = x^3 - 3x^2 + 6x + 3$.

7.12.2. Determine the number of real roots of the following equations: (a) $2x^3 - 3x^2 - 12x + 15 = 0$, (b) $4x^3 + 15x^2 - 18x - 2 = 0$, (c) $2x^3 - x^2 - 4x + 3 = 0$, and (d) $x^3 - x^2 + 2 = 0$.

7.12.3. Construct the graphs of the following functions:

(a) $y = \frac{x^3}{x-1}$, (b) $y^2 = x^3(1-x)$, (c) $y^2 = x^4(1+x)$, and (d) $y^3 = x^2(1-x)^2$.

Part 2

Higher Math Applied to Problems of Physics and Engineering

Chapter 8 Radioactive Decay and Nuclear Fission

8.1 The Basic Characteristics of Radioactive Decay

The ***basic law of radioactive decay***, which was established in experiments, states that the ratio of the number of atoms disintegrated in unit time to the total number of atoms is a constant that depends only on the species of atom. It is understood that the total number of atoms is extremely large.

This ratio is called the ***probabilily of disintegration***. Denote by $N(t)$ the quantity of atoms that have not disintegrated by time t. At time $t + dt$ there will be $N(t + dt)$ untransformed atoms. For this reason, during time dt (from t to $t + dt$) there will be $N(t) - N(t + dt) \simeq -dN$ disintegrations of atoms. If we divide the ratio of the number $-dN$ of atoms that have disintegrated during the time interval dt to the total number of atoms N (this ratio is the fraction of disintegrated atoms) by dt, we get the probability of disintegration $\omega = -dN/N\,dt$, from which it follows that

$$\frac{dN}{dt} = -\omega N \qquad (8.1.1)$$

From this relation, recalling that the dimensions of dN/dt are the same as those of the ratio N/t, we see that the dimensions of the probability of disintegration ω are $1/s$.[8.1]

The ***initial condition*** consists in specifying the number of atoms at the initial time: $N = N_0$ at $t = t_0$.

Solving Eq. (8.1.1) by the method given in Section 6.6 (see Eqs. (6.6.9) and (6.6.10)) and using the initial condition, we find that

$$N(t) = N_0 e^{-\omega t} \qquad (8.1.2)$$

(we advise the reader to perform all the computations). However, when the derivative is proportional to the desired function, a simpler solution to the equation can be offered.

In Chapters 4 and 6 we discovered that the derivative of the exponential

[8.1] Consequently, probability here is not to be understood in the sense of the assertion that, as in coin tossing, the probability is one half that the coin will fall heads. The definition of the probability of disintegration as the ratio of the number of disintegrations per unit time to the initial number of atoms holds true only for the case where the number of disintegrations per unit time (say, per second) constitutes a small fraction of the total number of atoms. The exact definition of probability of disintegration is given by the formula $\omega = -(1/N)\,dN/dt$, that is, the probability of disintegration is equal, *by definition*, to the ratio of the number of disintegrations during a small time interval to the total number of a atoms and to the magnitude of the time interval.

function is proportional to the function itself:

$$\frac{d(a^x)}{dx} = \text{const} \times a^x.$$

In particular,

$$\frac{d(Ce^{kx})}{dx} = Cke^{kx}$$

if C and k are constants. Recalling this property of the exponential function, let us suppose that the solution to Eq. (8.1.1) is of the form

$$N = Ce^{kt}, \qquad (8.1.3)$$

and let us try to choose C and k such that both the equation and the initial condition are satisfied. Differentiating (8.1.3) we get

$$\frac{dN}{dt} = Cke^{kt} = kN.$$

Substituting into Eq. (8.1.1) yields $kN = -\omega N$, whence $k = -\omega$. Assuming $t = 0$ in (8.1.3) and using the initial condition, we get $C = N_0$. And so $N = N_0 e^{-\omega t}$, where the quantity $-\omega t$ in the exponent is dimensionless, as it should be.

Radioactive atoms are characterized by their ***half-life*** T, which is the time during which the number of atoms N diminishes via disintegration by ***one half*** the original amount. Let us determine the half-life T. From (8.1.2), $N(T) = N_0 e^{-\omega T}$. On the other hand, $N(T) = N_0/2$, by definition. Therefore $N_0 e^{-\omega T} = N_0/2$, or $e^{-\omega T} = 1/2$, that is,

$$-\omega T = -\ln 2, \quad T = \frac{\ln 2}{\omega} \simeq \frac{0.69}{\omega}, \qquad (8.1.4)$$

Hence, the half-life is inversely proportional to the probability of disintegration.

Prior to disintegration, every atom exists a certain period of time, which is called the ***lifetime*** of the atom.

We wish to find the mean lifetime $\bar{t}$ of an atom of a given radioactive element. Suppose that at the initial time $t = 0$, say when the radioactive element was created, there were N_0 atoms. During the time interval from t to $t + dt$ the quantity of atoms that disintegrated was approximately

$$-dN = \omega N\, dt.$$

All the atoms of this group lived roughly the same lifetime t. Among the atoms taken at the initial time $t = 0$ there are groups of atoms that will have different lifetimes: from the common-to-all-atoms time of creation to the distinct-for-various-atoms time of disintegration. To find the mean lifetime, we must multiply the lifetime of each group by the number of atoms in the group, add these quantities for all groups, and divide the result by the total number of atoms in all groups.

Since we have to add a very large number of very small terms, the sum can be replaced by an integral, and therefore (compare with Section 7.8)

$$\bar{t} = \int_0^\infty t\omega N\, dt \Big/ \int_0^\infty \omega N\, dt. \qquad (8.1.5)$$

We substitute the expression for N from (8.1.2). The denominator on the right-hand side of (8.1.5) is

$$\int_0^\infty \omega N\, dt = \int_0^\infty \omega N_0 e^{-\omega t} dt$$

$$= \omega N_0 \int_0^\infty e^{-\omega t}\, dt = -\omega N_0 \frac{e^{-\omega t}}{\omega}\Big|_0^\infty = N_0,$$

as was to be expected, since the integral in the denominator yields the total number of all disintegrated atoms, which, clearly, is equal to the number of atoms existing at the initial time.

We integrate the integral in the numerator on the right-hand side of (8.1.5) by parts, setting $t = f$, $e^{-\omega t}\, dt = dg$, and $-(1/\omega)\, e^{-\omega t} = g$. As a result we have

$$\omega N_0 \int_0^\infty te^{-\omega t}\, dt$$

$$= \omega N_0 \left(-\frac{1}{\omega} te^{-\omega t} + \int \frac{1}{\omega} e^{-\omega t}\, dt \right)\Big|_0^\infty$$

$$= \omega N_0 \left(-\frac{1}{\omega} te^{-\omega t} - \frac{1}{\omega^2} e^{-\omega t} \right) \Big|_0^\infty = \frac{N_0}{\omega}.$$

From (8.1.5) we now obtain

$$\bar{t} = \frac{N_0}{\omega N_0} = \frac{1}{\omega}, \qquad (8.1.6)$$

or that the mean lifetime of an atom is exactly the inverse of the disintegration probability. Using this fact, we can write the basic equation (8.1.1) and its solution (8.1.2) as

$$\frac{dN}{dt} = -\frac{N}{\bar{t}}, \qquad (8.1.7)$$

$$N = N_0 e^{-t/\bar{t}}. \qquad (8.1.8)$$

We must not forget that the time t is the independent variable; the number of atoms depends on t. On the other hand, $\bar{t}$ is a constant that describes the given species of radioactive atom.

From (8.1.8) it is evident that during time $\bar{t}$ the number of atoms diminishes from N_0 to $N_0 e^{-1} = N_0/e$, by a factor of e, which is roughly equal to 2.72.

By formula (8.1.7), the initial rate of disintegration is such that if the number of atoms decaying per unit time did not fall off, all the atoms would disintegrate in time $\bar{t}$. Indeed, at $t = 0$ there were N_0 atoms and the rate of disintegration was $(dN/dt)_{t=0} = -N_0/\bar{t}$. At that rate, complete disintegration requires a time equal to $\bar{t}$. From (8.1.4) it follows that $\omega = (\ln 2)/T$, and so $\bar{t} = T/\ln 2 \simeq 1.45T$. Computationally, the quantity $\bar{t}$ is more convenient than the half-life T.

Exercises

8.1.1. The mean lifetime of radium is 2400 years. Determine the half-life of radium.

8.1.2. We start with 200 grams of radium. How much radium will be left in 300 years?

8.1.3. Ten grams of radium disintegrated in 500 years. How much was there at the beginning?

8.1.4. Determine how much time will elapse for 1%, 10%, 90%, and 99% of an original supply of radium to disintegrate.

8.1.5. The amount of radium in the earth in various rocks (the ratio of the number of radium atoms to the number of atoms of rock) comes out to about one part in 10^{12}. What was the content of radium in the rocks 10 000 years ago, 10^6 years ago, 5×10^9 years ago (5×10^9 is the age of the earth)?

8.2 Measuring the Mean Lifetime of Radioactive Atoms

The mean lifetime $\bar{t}$ of various radioactive atoms is extremely diversified. To illustrate, let us take the several known isotopes of *uranium*. One, with atomic weight 238 (U^{238}), has a mean lifetime of 7×10^9 years. Another (U^{235}) has a mean lifetime of 10^9 years (the fission of uranium-235 in nuclear power plants is the main source of atomic energy). The mean lifetime of *radium* is 2400 years.[8.2]

However, it would be wrong to think that the mean lifetimes of all radioactive atoms are measured in thousands of years. Among radioactive substances that occur in nature and were studied by Marie and Pierre Curie and Ernest Rutherford we find *polonium* with a mean lifetime of about 200 days, *radium* A with a mean lifetime of 4 minutes, and *radium* C′ with a mean lifetime of 2×10^{-4} second.

During recent decades, a huge number (over 1000) of different radioactive substances with a vast range of mean lifetimes have been discovered in connection with the development of nuclear physics and the use of atomic energy.

If at time t there are $N(t)$ untransformed atoms, then there will be $n(t) = \omega N(t) = -dN/dt$ disintegrations (of atoms) per unit time. The quantity $n(t)$ ($= n_0 e^{-\omega t}$, with $n_0 = -(dN/dt)_{t=t_0}$; compare with (8.1.2)) is the ***rate of disintegration*** of the atoms (see Section 8.1).

Suppose we observe the decay of uranium for ten years. During this time about 4×10^{12} atoms in one gram of uranium will have disintegrated. It would be extremely hard to detect that the original amount of 2.5×10^{21} atoms

[8.2] Reference books frequently give the half-life $T \approx 0.69\bar{t}$ instead of the mean lifetime $\bar{t}$; see Section 8.1.

would be diminished by 4×10^{12} atoms. Measuring the ***number of desintegrations*** often proves to be easier than measuring the number of atoms that have not disintegrated.

However, by experiments involving radioactive substances with relatively short mean lifetimes (from a few minutes to a few days) it has been possible to verify formula $n = n_0 e^{-\omega t}$ (which follows from the fact that n is proportional to N) with great accuracy and, thus, to corroborate formulas (8.1.1) and (8.1.2). To do this, let us calculate the number of disintegrations over small periods of time. Dividing the number of disintegrations by the time interval, we get the disintegration rate at various instants of time.

We construct the graph of the disintegration rate as a function of time. How can we be sure that this curve is the graph of the exponential function? We can compute the logarithms of the resulting values of the rate of disintegration and, on this basis, construct a graph of $\ln n\,(t)$ as a function of time t. The result should be a straight line, which can be checked roughly by eye. Numerous experiments do indeed yield a straight line.

Thus, $\ln n\,(t)$ is a linear function of time, or

$$\ln n\,(t) = a + bt, \qquad (8.2.1)$$

which means that $n\,(t) = e^{a+bt} = e^a e^{bt} = ce^{bt}$. The quantity b turns out to be negative on the graph, $b = -\omega$, where ω is the probability of disintegration. Thus, experiments confirm the basic result of the preceding section and enables us to determine ω by computing the tangent of the angle of inclination of the straight line (8.2.1) to the time axis.

Actually, this is a very remarkable result.[8.3] Imagine N_0 radioactive atoms "manufactured" simultaneously at time $t = 0$. They are all prepared in the same fashion and at the same time. We know that radioactive atoms are unstable and are capable of disintegrating. We can suppose that the disintegration of the atoms requires a definite time. Imagine that after the atoms are ready, a certain time must elapse before they are mature enough to disintegrate. But then we should expect all the atoms to mature in the same period of time and, at the expiration of that time, to disintegrate simultaneously. Imagine, further, that we have models of guns with stretched springs and gears (or clockwork time devices) that release their shells when the gears (or clocks) reach a particular position (time). Firing will be regarded as disintegration of the model. If all models are the same and manufactured at the same time, the shells should be fired after the lapse of an identical time interval.

But this picture of model disintegration (firing of the shells) has nothing whatsoever in common with the actual behavior of radioactive atoms. Though created at the same time, they disintegrate at all imaginable times. Let us try to find out what percentage disintegrates during a time less than the mean lifetime. From (8.1.2) we find that the rate of disintegration (the number of atoms that decay in unit time) is $dN/dt = -\omega N_0 e^{-\omega t}$ (it is clear that dN is negative). During time dt there will be

$$\frac{dN}{dt}\,dt = dN = -\omega N_0 e^{-\omega t}\,dt$$

atomic disintegrations, while during the time from $t = 0$ to $t = \bar{t}$ the following number of atoms will disintegrate:

$$M = \int_0^{\bar{t}} \omega N_0 e^{-\omega t}\,dt = -N_0 e^{-\omega t}\,\Big|_0^{\bar{t}} =$$

$$= N_0\,(1 - e^{-\omega t}) \text{ atoms.}$$

[8.3] Niels Bohr, speaking on radioactive transformations, said in this connection: "The meaning of the discussions on the mean lifetimes of atoms without any indication of a definite instant of time lies in the fact that the atoms, so to say, do not grow old until they begin to decay; this means that the probability of disintegration is the same for them at any time."

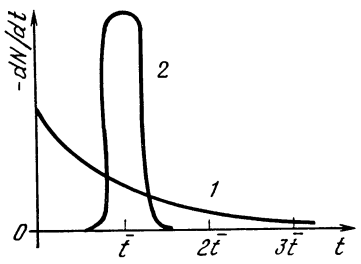


Figure 8.2.1

Since $\omega = 1/\bar{t}$, it follows that

$$M = N_0 \left(1 - \frac{1}{e}\right) \simeq 0.63 N_0.$$

Thus, 63% of the atoms will disintegrate during a time less than $\bar{t}$. Similarly, we compute that during the time from $\bar{t}$ to $2\bar{t}$ 23 % of the atoms disintegrate, and during the time exceeding $2\bar{t}$ only 14% of the atoms (the remainder).

In Figure 8.2.1 we see two curves: one for the number of disintegrations in unit time for radioactive atoms (curve *1*) and for the gun models (curve *2*). The gun-model curve has a certain width here. We can figure, perhaps, that the models were not quite exact and therefore did not fire off quite at the same time. The more precise the models, the narrower curve *2* in Figure 8.2.1.

The area under the curve represents the total number of disintegrated atoms for one curve and the total number of all models for the other curve. We can take the number of models to be equal to the number of atoms. Then both curves will have the same area. The abscissa of the center of gravity of both curves is also the same.[8.4] This means that we consider models in which the mean lifetime (prior to firing) is the same as the mean lifetime of the radioactive atoms.

We have thus done everything in our power to make the curves similar: we took as many models and with mechanism such that the total number of models and atoms and the mean lifetimes of the models and atoms are the same. And yet the curves are so strikingly different! Experimentation with radioactive nuclei irrefutably rejects the type of curve obtained for the models. The more accurate the experiment, the more precisely is the law (8.1.2) confirmed.

We examined this system with models so that the reader would not accept as ordinary and natural the relationship (8.1.2) for radioactive decay and would have cause for surprise and curiosity: "Indeed, why does radioactive disintegration proceed in this fashion?"

What is the physical meaning of the probability of disintegration? A long time ago, at the beginning of the century, it was suggested that radioactive decay requires some kind of external action, say the entry of a particle from outside. Then we would imagine that one atom disintegrated earlier since it was hit by an incoming particle, while some other atom remained untouched. But this hypothesis did not fit the facts which stated that radioactive disintegration proceeds at the same rate under all manner of conditions, irrespective of temperature, collisions of atoms among themselves, the action of cosmic radiation. Also, energy is strictly conserved in radioactive disintegration, which likewise rejects the idea of some kind of outside influence.

A second possible hypothesis is that at the initial time of creation of the radioactive atoms they were actually not quite alike and for this reason decayed at different times. This is in keeping with the clock-driven models with clocks set for different times. This hypothesis presumes that an exact knowledge of the state of every atom completely determines the whole subsequent history of the atom and, in particular, determines with exactitude when the given atom will disintegrate. If atoms disintegrate at different times after their creation, this means that the whole business was foreordained:

[8.4] As will be demonstrated in Section 9.12, this follows from Eq. (8.1.5).

when created, the different atoms of one and the same radioactive substance were not exactly the same and the diverse decay times were predetermined in the creation stage.

This view does not hold water either. For each specific mode of generation of atoms of a radioactive element we should have a definite relationship between the rate of disintegration and time. Experiment refutes this supposition.

One and the same species of radioactive atom can often be obtained in a variety of ways: say, atoms of Mo^{99} (molybdenium with atomic weight 99) are produced in nuclear reactors in the process of fission of uranium atoms. These same atoms were earlier obtained under the action of the nuclei of heavy hydrogen (deuterium) on the atoms of ordinary, naturally occurring, nonradioactive molybdenium. Experiments have shown that, irrespective of the mode of production of the atoms, the disintegration rate is given by formula (8.1.2) with a constant value of ω, which characterizes the given species of atom. Consequently, it is precisely the basic equation $dN/dt = -\omega N$ that all experiments confirm, with the characteristic ω of atoms of a given species quite independent of external conditions.

This equation is pregnant with meaning: all radioactive atoms of a single species are identical. The probability of disintegration does not depend on how and when the atoms were obtained. One hundred freshly produced atoms disintegrate in exactly the same fashion as in the case where 10^6 atoms are generated, a time interval elapses such that 100 atoms are left, and we consider the fate of these 100 remaining atoms.[8.5]

What is so remarkable in the fact that 100 atoms with a given atomic weight and a given number of electrons are the same? If these were nonradioactive atoms, there would indeed be no cause for surprise. But for radioactive atoms there certainly is cause enough when we recall that out of the 100 atoms 63 disintegrate in time $\bar{t}$ and the other 37 disintegrate after $\bar{t}$. What is strange here is that the disintegration time is different although the atoms are the same.

It is not fruitless to wonder in this fashion. In the phenomenon of radioactive decay we already perceive certain peculiarities in the laws of motion of atomic and nuclear particles that differ from the laws of motion of the bodies we are accustomed to in classical mechanics and ordinary life. These peculiarities are studied in quantum mechanics. All this is of course outside the scope of our book. Our aim is modest enough. It is to show that the necessity for elaborating radically new conceptions that differ drastically from those of ordinary mechanics stems from the very simple facts about radioactivity that can be comprehended by any school child. To realize that the old conceptions were insufficient, it was necessary to doubt, to wonder, to be surprised.

In his autobiography, Albert Einstein—the greatest physicist of the twentieth century—notes the surprise and wonder that he experienced when he first saw a compass and perceived the mysterious action of a magnetic force that passes through paper, wood, the earth and acts on the compass needle without any direct contact. He wrote that this wonderment served as a tremendous impetus for a further search. He wrote of curiosity, which he claims "the modern methods of teaching have

[8.5] Characteristic of the exponential function is that any portion of the curve is similar to the whole curve. Indeed, let us begin reckoning time anew from time t_1. Denote by τ the time reckoned from this instant: $\tau = t - t_1$, $t = t_1 + \tau$. Then $N = N_0 e^{-\omega t} = N_0 e^{-\omega(t_1+\tau)} = N_0 e^{-\omega t_1} e^{-\omega\tau} = N_1 e^{-\omega\tau}$, where $N_1 = N_0 e^{-\omega t_1}$ is the number of atoms at time t_1. Thus, the law of disintegration for the number of particles $N_1 = N_0 e^{\omega t_1}$ remaining after the previous decay is exactly the same as the law of disintegration of N_1 freshly obtained particles. This is precisely what is asserted in the text.

all but stifled." Einstein himself evinced an extraordinary capability for wonderment and he was able to derive inspiration and impetus for the creation of theories out of the most mundane facts of everyday life. For instance, underlying the brilliant general theory of relativity is the simple fact that caused Einstein to wonder why different bodies fall with the same acceleration.

Quite naturally, to wonder is not enough, and to merely pose a problem does not suffice. Einstein combined the ability to pose a problem and to solve it, which means mastering the requisite mathematical techniques. And yet, among a galaxy of outstanding scientists, Einstein is the celebrated physicist of the twentieth century because of his capacity to wonder and pose a problem where others were not able to see anything out of the ordinary.

Perhaps this analysis of radioactive decay will help the reader to see what depths of content are hidden behind simple facts and formulas.

To conclude this section, we will illustrate the curves of radioactive decay obtained experimentally in 1955 by Glenn Seaborg and his associates in the United States (Figure 8.2.2). They were first to observe Element No. 101 of the periodic table, to which they gave the name *mendelevium* (symbol Md) in honor of the great Russian chemist Dmitri Mendeleyev.

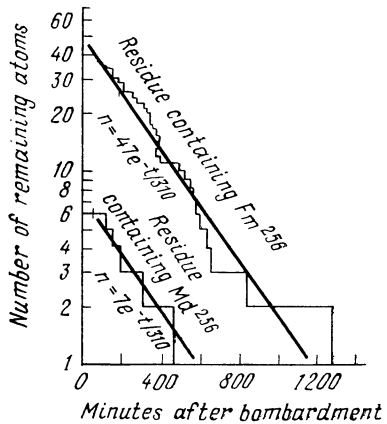


Figure 8.2.2

In this study the 98th element, *californium*, with atomic weight 252 was irradiated with neutrons. A neutron is captured by the nucleus of a Ca^{252} atom, which transforms into a Ca^{253} atom. Californium-253 ejects an electron and turns into Element No. 99, which is called *einsteinium* (symbol Es) and has the same atomic weight 253.

About 10^9 atoms of einsteinium (this is equivalent to 4×10^{-13} gram) were deposited on a gold plate and subjected to alpha-particle bombardment (α particles are nuclei of helium) in a cyclotron. This generates Element No. 100, *fermium* (symbol Fm), in accordance with the reaction

$$_{99}Es^{253} + {}_2He^4 = {}_{100}Fm^{256} + {}_1p^1,$$

and *mendelevium* via the reaction

$$_{99}Es^{253} + {}_2He^4 = {}_{101}Md^{256} + {}_0n^1.$$

In this notation, each chemical symbol has a subscript indicating the number in the periodic table (that is, the number of protons in the nucleus). The superscript indicates the atomic weight (rounded to a whole number, which is the number of neutrons and protons in the nucleus). The *helium* nucleus, or the α particle, is denoted by $_2He^4$, the nucleus of a hydrogen atom, or simply the proton, by $_1p^1$, and the neutron by $_0n^1$. In a nuclear reaction the sum of the subscripts on the left-hand side of the reaction equals that on the right-hand side, and the same is true of the superscripts, since in nuclear reactions all we have is an exchange of neutrons and protons between the nuclei.

After the α particle bombardment, the gold plate together with the newly formed fermium and mendelevium was dissolved in acid, and then the fermium and mendelevium were extracted chemically. As Seaborg writes, it was the periodic table of Mendeleyev that permitted foreseeing the chemical properties of an element that had never before existed in nature and had never been studied. After chemical separation, measurements were made of the

radioactive decay characteristics. Fermium-256 (with atomic weight of 256) disintegrates radioactively with a half-life of about 3.5 hours. It breaks up (that is, fissions) spontaneously into two nuclear fragments of roughly the same mass (the details of the fission process are given in Section 8.3).

The upper curve in Figure 8.2.2 shows the number of nuclei of fermium as a function of time in this experiment. Plotted on the horizontal axis is the time in minutes. Plotted on the vertical axis is the number of atoms available at a given time. [8.6] The scale on this axis is not uniform: the distance from the horizontal axis is proportional to the *logarithm* of the number of atoms. In particular, the horizontal axis ($y = 0$) corresponds to one remaining atom ($\ln 1 = 0$), while for zero atoms we have $-\infty$ on the vertical axis. The disintegration of each separate atom changes the number of atoms by 1; in between disintegrations the number of atoms is constant. In the experimental set-up in which each separate disintegration is recorded, we have a polygonal (step-like) curve instead of a smooth curve. On the step-like curve, each disintegration is associated with a vertical line connecting two steps. The upper straight line in Figure 8.2.2 corresponds to the decay law $n = n_0 e^{-t/\tau}$, where $\tau = T/\ln 2 \simeq 5$ hours, $T \simeq 3.5$ hours. It will be seen from Figure 8.2.2 that, in all, there were recorded 40 disintegrations of fermium. The more atoms there are, the closer is the polygonal line to a straight line. When there are fewer than five atoms left, then quite naturally the probabilistic nature of radioactive decay leads to appreciable deviations from the exponential law, which holds true for *large* number of atoms.

After chemical separation, the nucleus of mendelevium rapidly (in half an hour) captures an atomic electron and transforms into a nucleus of fermium. And so the precipitate containing mendelevium also yields (when the radioactivity is measured) a disintegration of atoms into two fragments with a half-life of 3.5 hours. The decay curve for fermium obtained from mendelevium lies in the lower left-hand corner of Figure 8.2.2 Six disintegrations were experimentally observed. Special experiments demonstrated that these six atoms could not have appeared as a fermium impurity in the precipitate being measured but most definitely had formed from the mendelevium.

Seaborg and his associates observed a total of 17 atoms of mendelevium in that series of experiments.

The foregoing example is not very good as an illustration of how exactly the exponential law (8.1.2) holds true in radioactive decay (see also (8.2.1)). Experiments demonstrating the validity of the exponential law were successfully carried out with more common radioactive substances. On the other hand, the example of mendelevium and fermium shows what peaks of experimental technique modern physicists have attained in synthesizing new elements and recording the disintegration of each separate atom.

In Seaborg's experiments, the counter recording mendelevium disintegrations was hooked up to an amplifier in the loudspeaker system of the institute and every disintegration event was heard by workers in the various laboratories on different floors so they could celebrate the birth (actually the recorded death) of every atom of the new element created by man (true, before the work was over, the local fire department got interested in these goings-on and the disintegration news broadcast was stopped).

In the Soviet Union the synthesis and investigation of the heaviest elements are being successfully conducted

[8.6] It is not possible to count the number of atoms available at a given instant. What is recorded experimentally are the disintegration events of the atoms. The number of atoms N at time t is computed after the experiment, when all N atoms have disintegrated.

by Academician G. N. Flyorov and his associates at the Joint Institute for Nuclear Research in Dubna, a city near Moscow.

8.3 Series Disintegration (Radioactive Family)

In a number of cases, radioactive decay produces atoms that again decay radioactively, so what we have is a chain of disintegrations: an atom of element A transforms into an atom of element B, which in turn disintegrates into an atom of element C, and so on. Let us consider the mathematical problem of determining the dependence on time of quantities of elements A, B, C and ways of solving the problem. We denote the quantities of substances (elements) A, B, C that have not yet decayed by time t by the italic letters A, B, C; thus, $A = A(t)$, $B = B(t)$, and $C = C(t)$.

Let the *probabilities of disintegration* of A, B, C be equal to ω, v, u, respectively. Then

$$\frac{dA}{dt} = -\omega A \qquad (8.3.1)$$

(here, A is termed the ***parent element***). We write the equation for element B (the ***daughter element***). In unit time a total of vB atoms of element B disintegrate. On the other hand, during the time we have ωA disintegrations of element A, and since each disintegration of an atom of A given rise to one atom of B, there are ωA, atoms of B formed in unit time. Therefore

$$\frac{dB}{dt} = -vB + \omega A. \qquad (8.3.2)$$

Similar reasoning yields

$$\frac{dC}{dt} = -uC + vB. \qquad (8.3.3)$$

Equations (8.3.1)-(8.3.3) form *a system of differential equations.* In the given instance, we can solve these equations one by one, having to deal each time only with one equation in one unknown. Indeed, neither B nor C enter into Eq. (8.3.1). From it we therefore find that $A(t) = A_0 e^{-\omega t}$, where A_0 is the number of atoms of element A at the initial time $t = 0$. (We assume that $B = C = 0$ at $t = 0$).

Substituting the expression for $A(t)$ into (8.3.2), we get an equation involving only one unknown function $B(t)$:

$$\frac{dB}{dt} = -vB + \omega A(t), \qquad (8.3.4)$$

How does one go about solving this equation? We can find a solution if we first consider the fate of the group of atoms of B that have formed during the same interval of time, from τ to $\tau + \Delta\tau$. We will consider the number of atoms of this group that are still "alive", ΔB (that is to say, atoms that have not disintegrated by time t), as a function of time t. So as to avoid confusion about the time t when we measure the number of atoms and the time of formation of the group, let us denote these times by different letters, t and τ, respectively. At time τ, the rate of formation of atoms of element B was $\omega A(\tau)$. During the small time interval $\Delta\tau$, a total of $\Delta B_0 = \omega A(\tau)\,\Delta\tau$ atoms of element B were formed.

How does the number of atoms in the group at hand depend on time t? For $t < \tau$ is it equal to zero: the atoms of interest have not yet formed since the group itself is still nonexistent, $\Delta B = 0$. Let $t > \tau$. Observe that a time $t - \tau$ has already passed since the group began to form. The decay probability of element B is v. Therefore, after the lapse of time $t - \tau$ from the time of formation of the group the number of untransformed atoms will be

$$\Delta B(\tau) = \Delta B_0 e^{-v(t-\tau)} = \omega A(\tau)\, e^{-v(t-\tau)}\,\Delta\tau.$$

To find the total number of atoms of element B at time t, we have to add the number of atoms in all groups that formed prior to t. If we take $\Delta\tau$ (and hence ΔB as well) very small, then the sum

turns into the integral

$$B(t) = \int_0^t \frac{\Delta B(\tau)}{\Delta \tau} d\tau$$

$$= \int_0^t \omega A(\tau) e^{-v(t-\tau)} d\tau.$$

Observe that here the variable of integration is denoted by τ. The argument t, upon which B depends, enters into the integral twice: as the upper limit and in the integrand. When integrating with respect to τ the quantity t is to be regarded as a constant. We can therefore write

$$e^{-v(t-\tau)} = e^{-vt} e^{v\tau}$$

and take ωe^{-vt} out from under the integral sign as a factor that is independent of τ. We then get

$$B(t) = \omega e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau. \qquad (8.3.5)$$

It is easy to verify, without evaluating the integral, that the solution (8.3.5) satisfies the original equation (8.3.4) for any function $A(\tau)$. Indeed, let us find the derivative $dB(t)/dt$. By the rule of differentiating a product we get

$$\frac{dB(t)}{dt} = -\omega v e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau$$

$$+ \omega e^{-vt} \frac{d}{dt} \left(\int_0^t A(\tau) e^{v\tau} d\tau \right).$$

Since, by the property of the derivative of an integral (see Section 3.3),

$$\frac{d}{dt} \left(\int_0^t A(\tau) e^{v\tau} d\tau \right) = A(t) e^{vt},$$

it follows that

$$\frac{dB}{dt} = -\omega v e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau$$

$$+ \omega A(t) = -vB + \omega A.$$

If we set $A(\tau) = A_0 e^{-\omega\tau}$, we get the concrete solution

$$B(t) = \frac{A_0 \omega}{v - \omega} (e^{-\omega t} - e^{-vt}). \qquad (8.3.6)$$

The solution could also have been found without resorting to a consideration of the separate groups of atoms. Now that the solution has been found, it is already a simple matter to guess the mathematical technique that will lead us to our goal. The solution (8.3.5) is of the form

$$B(t) = e^{-vt} I(t), \qquad (8.3.7)$$

where $I(t)$ stands for an integral that depends on t. We will seek the solution in the form of a product of e^{-vt} by the unknown function $I(t)$ and will set up an equation for $I(t)$:

$$\frac{dB}{dt} = \frac{d}{dt}(e^{-vt} I) = -ve^{-vt} I + e^{-vt} \frac{dI}{dt}. \qquad (8.3.8)$$

Substituting (8.3.8) and (8.3.7) into Eq. (8.3.4) yields

$$e^{-vt} \frac{dI}{dt} = \omega A(t),$$

or

$$\frac{dI}{dt} = \omega e^{vt} A(t). \qquad (8.3.9)$$

By hypothesis, at the initial time $t = 0$ we have $B = 0$, and hence $I = 0$ at $t = 0$. With this initial condition, the solution to Eq. (8.3.9) has the form

$$I(t) = \int_0^t \omega e^{v\tau} A(\tau) d\tau.$$

Thus, the final result is

$$B(t) = e^{-vt} I(t) = e^{-vt} \int_0^t \omega A(\tau) e^{v\tau} d\tau. \qquad (8.3.10)$$

In this formula it is essential, so as to avoid confusion, to retain strict designations and not to denote the variable of integration τ by the same letter we use for the upper limit of integration, t.

8.4 Investigating the Solution for a Radioactive Family (Series)

In the preceding section we brought to completion the solution of the problem in the case of *two* radioactive substances (elements). Let us now investigate this solution for two particular cases:

(1) a short-lived parent element A and long-lived daughter element B,

(2) a long-lived parent element A and short-lived daughter element B.

Below, in addition to the decay probabilities ω and v, we will make use of the mean lifetimes $\bar{t}_A = 1/\omega$ and $\bar{t}_B = 1/v$. In the first case, where $\bar{t}_A \ll \bar{t}_B$, the nature of the solution can be readily grasped without calculations and formulas. The entire process breaks down into two stages. First, when t is of the order of $\bar{t}_A$ (here, by hypothesis, $\bar{t}_A \ll \bar{t}_B$ and so also $t \ll \bar{t}_B$ in the first stage), element A is transformed into element B, while during this time there is hardly any disintegration of element B. During this period the amount of B is equal to the difference between the original amount A_0 and the amount A remaining at time t:

$$B(t) = A_0 - A(t) = A_0 - A_0 e^{-\omega t}$$
$$= A_0 (1 - e^{-\omega t}), \quad t \ll \bar{t}_B.$$

By the end of this period, practically the *whole* of element A has been converted into B, and the quantity of B becomes equal to the original amount of the parent element, A_0. The quantity of element A becomes zero. Then follows a slow and protracted disintegration of B:

$$B(t) = A_0 e^{-vt}, \quad t \gg \bar{t}_A.$$

We will show how these qualitative ideas follow from the exact formula. For the case of two radiactive elements A and B, we obtained in the preceding section the formula

$$B(t) = A_0 \frac{\omega}{v - \omega} (e^{-\omega t} - e^{-vt}).$$

In our case $\bar{t}_A \ll \bar{t}_B$ and $\omega \gg v$, and so it is more convenient to interchange the signs so as to be dealing with positive quantities in the parentheses and in the denominator of the fraction. Then

$$B(t) = A_0 \frac{\omega}{\omega - v} (e^{-vt} - e^{-\omega t}). \qquad (8.4.1)$$

Since $v \ll \omega$, it follows that $\omega/(\omega - v) \simeq \omega/\omega = 1$.

We consider the expression $e^{-vt} - e^{-\omega t}$ for two successive stages. First, when $t \ll \bar{t}_B = 1/v$, it will be true that $vt \ll 1$. Then $e^{-vt} \simeq 1$. Since t can be a quantity of the order of $\bar{t}_A$ and ωt, consequently, of the order of unity, it follows that $e^{-\omega t}$ must be calculated exactly. From formula (8.4.1) we get

$$B(t) \simeq A_0 (1 - e^{-\omega t}). \qquad (8.4.2)$$

In the second stage, when $t \gg \bar{t}_A = 1/\omega$, it will be true that $\omega t \gg 1$. We can disregard $e^{-\omega t}$ in this stage, since $e^{-\omega t}$ is small not only with respect to unity but also in comparison with e^{-vt}, because $v \ll \omega$. We get

$$B(t) = A_0 e^{-vt}. \qquad (8.4.3)$$

Thus, the exact formula does indeed yield the same results as those obtained from simple qualitative reasoning.

Now let us take up the second case, that of the long-lived parent element A and short-lived daughter element B:

$$\bar{t}_A \gg \bar{t}_B, \quad \omega \ll v.$$

We consider the period when a time t considerably exceeding $\bar{t}_B$ has passed since the onset of the process. In that case, element B that was formed at the beginning of the process has already fully disintegrated by time t. Since B disintegrates rapidly and in a short time, at every given instant of time there is only availabbe the quantity of element B that has recently formed. What we have here is a ***steady state*** (also known as a ***stationary state***): element B is formed from A and straightway disintegrates; element B does not accumulate because it decays rapidly, but does not disappear completely because A is producing new quantities of B all the time. In a steady state system, there

are just as many atoms of B disintegrating in unit time as there are atoms of B being formed from A, so that B here changes but little. Mathematically, this condition is written as $v\,B \simeq \omega A$, whence

$$B(t) \simeq \frac{\omega}{v} A(t) = \frac{\bar{t}_B}{\bar{t}_A} A(t). \qquad (8.4.4)$$

In the steady state the instantaneous quantity of B is proportional to the quantity of A and always represents some small fraction of A. This fraction is small because in the case at hand (the second case) $\bar{t}_B \ll \bar{t}_A$ and hence $\bar{t}_B/\bar{t}_A \ll 1$, for otherwise there would be no steady state.

How do we obtain the steady-state equation from the exact differential equation $dB/dt = -vB + \omega A$? Evidently, if we take it that dB/dt is small in comparison with each of the two terms on the right, replacing dB/dt by 0 we approximately get

$$0 = -vB + \omega A, \text{ or } vB = \omega A.$$

Let us now examine the beginning of the process. At $t = 0$ we have $A = A_0$ and $B = 0$. This means that at the beginning we do not have a steady state, since by steady-state formulas we should first have

$$B_{st} = \frac{\omega}{v} A_0$$

(the subscript on B denotes a steady, or stationary, state). At $t = 0$, element B is forming at the rate $dB/dt = \omega A_0$, while there is no disintegration of B at all at the initial time since $B = 0$.

It is possible to determine the time t_1 during which, for the initial (constant) rate of buildup of B, the quantity B_{st} will be attained. Indeed, if the rate of formation of B remains constant, equal to $(dB/dt)_{t=0}$, then

$$B = t\left(\frac{dB}{dt}\right)_{t=0}.$$

Putting $B_{st} = (\omega/v) A_0$ and $(dB/dt)_{t=0} = \omega A_0$, we get the desired time

$$t_1 = \frac{\omega A_0}{v \omega A_0} = \frac{1}{v} = \bar{t}_B.$$

Thus, the steady state is attained in a time roughly equal to the mean lifetime of element B (we recall that our assumption that the rate at which substance B is formed is valid only approximately). From the condition $\bar{t}_A \ll \bar{t}_B$ it is evident that the quantity of element A changes but little during this time.

On the whole, the approximate examination in the case of a short-lived daughter element yields the following:

$$B(t) = \left(\frac{dB}{dt}\right)_{t=0} t = \omega A_0 t \quad \text{if} \quad t < \bar{t}_B, \qquad (8.4.5)$$

$$B(t) = B_{st} = \frac{\omega}{v} A_0 e^{-\omega t} = \omega \bar{t}_B A_0 e^{-\omega t} \quad \text{if}$$

$$t > \bar{t}_B.$$

We get the function $B = B(t)$ in the form of two lines: first the straight line *1* (Figure 8.4.1), or a linear function, then an exponential curve. Figure 8.4.1 was constructed on the assumption that $t_A = 10\bar{t}_B$; instead of the exponential curve we depicted the straight line *2*. It is easy to verify that for $t = \bar{t}_B$ the two formulas in (8.4.5) yield almost the same result.

Let us see what the exact solution (8.4.1) to Eq. (8.3.4) gives us in the case at hand, where $v \gg \omega$ and $\bar{t}_B \ll \bar{t}_A$. In the denominator we neglect ω compared with v. For $vt \gg 1$ we also neglect e^{-vt} in the parentheses. This yields

$$B \simeq A_0 \frac{\omega}{v} e^{-\omega t}, \qquad (8.4.6)$$

which is precisely the steady-state solution.

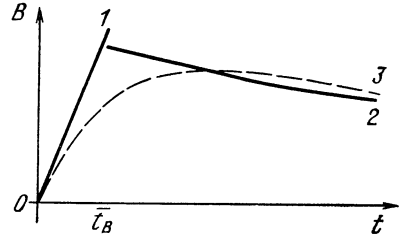


Figure 8.4.1

We determine how close the solution is to a steady state by how fast e^{-vt} falls off. For very small values of t (when $vt < 1$, so that ωt is sure to be small), we get the following by expanding e^{-vt} and $e^{-\omega t}$ in a series and confining ourselves to the first two terms:

$$B \simeq A_0 \frac{\omega}{v-\omega}(1-\omega t-1+vt) = A_0 \omega t, \tag{8.4.7}$$

which also coincides with the approximate result. Actually, however, the exact formula yields a single smooth curve without discontinuities and salient points (the dashed curve *3* in Figure 8.4.1). The approach of this curve to the steady-state solution depends on how fast e^{-vt} diminishes. For instance, for e^{-vt} to make a correction of the order of 10%, we must have $vt \simeq 2.3$, or $t \simeq 2.3/v = 2.3\bar{t}_B$. Here, owing to the smallness of ωt, we assume that $e^{-\omega t} \simeq 1$. Thus, true enough, the transition from the stage of initial buildup to the stage where the solution is equal, with sufficient precision, to the steady-state solution takes place in a time of the order of the time of disintegration, $\bar{t}_B$.

The example of a radioactive family (series) is highly instructive in the sense that obtaining a general exact solution does not in the least signify that the work is at an end. The construction of approximate theories for various limiting cases is an absolutely necessary part of the work and the existence of an exact formula does not at all take the place of an approximate theory. Approximate, yet clear-cut and pictorial conceptions serve as a check on an exact formula.

Also, approximate theories give us important new qualitative concepts such as that of the steady state. These can more easily be remembered and they possess a broader range of application than do the exact formulas. For instance, in the case of a radioactive family consisting of *several* generations, A → B → C → D, the exact formula is extremely unwieldy. But if $\bar{t}_A$ is greater than all other times, $\bar{t}_B$, $\bar{t}_C$, and $\bar{t}_D$, then all the results referring to the steady state are obtained just as easily as in the case of two elements, A and B.

Sometimes the easiest way consists in obtaining an exact solution valid for *arbitrary* v and ω (in our case), from which we then (for $v \ll \omega$ or $v \gg \omega$) obtain, via mathematical manipulations, some simple approximate formulas for the two extreme cases. But this is not yet all! If a simple approximate formula has been obtained in a simple yet long-winded manner via the general solution, then alongside this there should be another, simple, way of obtaining the approximate formula. One should always attempt to find simple pathways because there will invariably appear problems in which the approach to an exact solution is insuperably complicated and only a simple approximate approach makes it possible to advance in the solution.

In practical situations, exact formulas come up just as rarely as algebraic, trigonometric, or other equations with solutions in whole numbers, although most of the textbook problems lead to exact formulas, just as problem books for junior classes abound in equations that can always be solved in whole numbers.

Observe that the conceptions of radioactive families account for the strange result of exercise 8.1.5 about the amount of radium in the past: radium is a descendent (true, not direct, but via a number of intermediate substances) of uranium-238. It is therefore not correct to regard the present-day supply of radium as the result of the decay of primordial radium. Actually, radium is in a steady state with uranium. From the equation

$$B = \frac{\omega}{v} A$$

we find that the quantity of radium $B = 10^{-12}$ corresponds to the uranium

content

$$A = \frac{\bar{t}_A}{\bar{t}_B} B = 3 \times 10^6 B = 3 \times 10^{-6}.$$

We have approximately found the present-day amount of uranium-238 in rocks. The original abundance, 5×10^9 years ago, was twice as much, of the order of 6×10^{-6}. These magnitudes are quite reasonable, unlike the results of the exercises in Section 8.1.

8.5 The Chain Reaction in the Fission of Uranium

In 1938, Otto Hahn and Fritz Strassmann in Germany and Irène and Frédéric Joliot-Curie in France demonstrated that when a neutron enters a nucleus of uranium, fission occurs in which the nucleus breaks up into two large fragments with the simultaneous emission of two or three new neutrons. Uranium with an atomic weight of 235 (uranium-235 for short) is very active in this respect. Naturally occurring uranium contains about 0.7% of uranium-235 atoms and 99.3% of uranium-238 atoms.[8.7] The fission fragments of uranium-235 are medium atomic-weight nuclei from 75 to 160. The charge of these nuclei lies within the range from 35 to 57, the sum of the charges of two fragments always being equal to the charge on the nucleus of uranium, that is, 92 elementary charges. The sum of the atomic weights of the two fragments is equal to $235 + 1 - \nu$, where 235 is the atomic weight of uranium-235, 1 is the atomic weight of the neutron that caused the fission, and ν is the sum of the weights of the neutrons generated in the act of fission. An enormous energy of 6×10^{10} J/g (per gram of fissioned uranium) is released in the fission process. Thanks to this great energy, right after the fission process the fragments rush apart at speeds of about 10^9 cm/s; then they decelerate and their kinetic energy is transformed into heat.

The source of this energy is the electric repulsion of two like-charged fragments. Before the nucleus is separated into two parts, the nuclear forces between the particles that make up the nucleus balance the electric repulsive forces. But as soon as the nucleus has broken up into two separate fragments, the repulsion of these two fragments is not countered in any way and so they fly apart at high speed. The fragments are very quickly brought to rest in a dense substance. Their time of flight is between 10^{-13} and 10^{-12} s. In this time they traverse distances from 10^{-4} to 10^{-3} cm. The kinetic energy of the fragments is converted into heat. The neutrons produced in fission have velocities of about the same order as the fragments (about 2×10^9 cm/s).

Of crucial importance for the practical utilization of the energy of nuclear fission is the fact that a fission event caused by *one* neutron gives rise to *more than one* neutron. It is quite clear that if the neutrons do not leave the system, their number will increase in geometric progression with time, that is, in accordance with the law of the exponential function. The rate of energy release will build up by the same law, in proportion to the number of neutrons. And even if at the onset of the process

[8.7] This was followed in 1939, in the laboratory of I.V. Kurchatov in Leningrad, by the demonstration (carried out by the Soviet scientists G.N. Flyorov and K.A. Petrzhak) that uranium-238 is capable of undergoing *spontaneous* fission without the entry of any neutron, althought the probability of this event is extremely low. The probability of radioactive decay (with the emission of an α particle) of uranium-238 corresponding to a half-life of 4.5×10^9 years is $\omega = 5 \times 10^{-18}$ s^{-1}, while the probability of the spontaneous fission of uranium-238 is lower by a factor of 10^6 that is, it is equal to about 5×10^{-24} s^{-1}. Thus, in one second in one kilogram of uranium (which is about 2.5×10^{25} atoms) there occur roughly 10^7 radioactive disintegrations and only 10 events of spontaneous fission. On the other hand, in the very heaviest elements, spontaneous fission becomes the most probable decay process (see the end of Section 8.2, in particular Figure 8.2.2, where the decay curve of mendelevium is given). The problem of the chain reaction that we consider below does not involve spontaneous fission at all.

there were few neutrons, their number builds up so fast that the energy will be released at a rate convenient for practical use (for instance, as a source of energy for a nuclear power plant), and in just a short additional space of time the energy release will build up to such an extent that an atomic explosion will take place. In reality, part of the neutrons leave the system, some are captured by other nuclei without causing fission. We can utilize this to control the number of neutrons and, in a particular case, attain a steady-state system in which the number of newly formed neutrons in unit time is equal to the number of used up neutrons, so that the number of neutrons in the system remains the same in the course of time, and the energy can be released at a constant rate. That precisely is the regime we need if atomic energy is to be used for peaceful purposes.

Our immediate task is to set up and investigate the equation describing the number of neutrons is a system as a function of time. This will be done in the sections below.

8.6 Multiplication of Neutrons in a Large System

Let us first derive an equation for the variation in the number of neutrons with time in a very large system (say, in a large chunk of uranium-235), when loss of neutrons to the outside can be neglected.[8.8] The neutrons can all be regarded as having the same speed; we denote it by v.

Fission of a nucleus occurs in roughly half of all cases when a neutron enters a nucleus of uranium-235. In the other half, the neutron emerges leaving the nucleus in the same state, with the number of neutrons remaining unchanged. The uranium nucleus is a sphere of radius R of the order of 10^{-12} cm.

[8.8] We will consider the simplest case of metallic uranium-235 without graphite moderator.

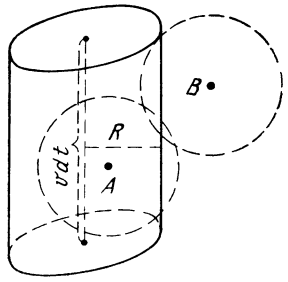


Figure 8.6.1

How often will a neutron in flight inside the metal hit a nucleus of uranium?

In a small time interval dt a neutron traverses a distance of vdt. Let us picture a cylinder whose axis is the route covered by the neutron; the radius of the cylinder is equal to the radius R of the uranium nucleus. The neutron collides with the nuclei whose centers lie inside the cylinder. If this is so, the path of the neutron will pass at a distance less than R from the center of the nucleus, and so the neutron will hit the nucleus and enter it. The volume of the cylinder is equal to $\pi R^2\, vdt$.

Figure 8.6.1 clarifies these ideas. The path of a neutron is shown by a dashed line along the cylinder's axis. In the case A the center of the nucleus lies inside the cylinder, with the result that the neutron hits the nucleus, while in the case B the center lies outside the cylinder and the neutron does not hit the nucleus.

There are N atoms in a unit volume of metallic uranium and hence N nuclei (the dimensions of N are cm^{-3}). Therefore, in the volume $\pi R^2 vdt$ that interests us there are $N\pi R^2 vdt$ nuclei. There will be just as many events of a neutron hitting a nucleus during the small time interval dt. Not every neutron hit makes a nucleus fission. Let α be the portion of cases a neutron hitting a nucleus causes fission ($\alpha \simeq 1/2$ in the case of uranium-235). Then the number of fissions during time dt is equal to $N\alpha\pi R^2\, vdt$.

The quantity $\alpha\pi R^2$, which has the dimensions of area since α and π

are dimensionless, is called the ***cross section*** of fission and is denoted by σ_f (the subscript f on the Greek letter sigma stands for "fission").

If there are n neutrons in the bulk of metallic uranium, the number of fissions in time dt is equal to

$$nN\sigma_f\, v dt. \tag{8.6.1}$$

Each act of fission produces ν neutrons, but this involves the absorption of one neutron, so the number of neutrons in every fission event increases by $\nu - 1$. Associated with the number of fissions (8.6.1) is the variation in the number of neutrons

$$dn = nN\,(\nu - 1)\,\sigma_f\, v dt. \tag{8.6.2}$$

From this equation we get

$$\frac{dn}{dt} = nN\,(\nu - 1)\,\sigma_f v.$$

Set

$$N\,(\nu - 1)\,\sigma_f v = a. \tag{8.6.3}$$

Then

$$\frac{dn}{dt} = an.$$

We already know that the solution to this equation is

$$n\,(t) = n_0 e^{at}, \tag{8.6.4}$$

where n_0 is the number of neutrons in the system at $t = 0$.

To summarize, then, if the number of neutrons in a system varies solely because of fission, then the number of neutrons increases in *geometric progression*, while time increases in *arithmetic progression*.

Indeed, if we take a number of equally spaced intervals of time,

$$t_1,\ t_1 + \Delta t,\ t_1 + 2\Delta t,\ t_1 + 3\Delta t,\ \ldots, \tag{8.6.5}$$

then the corresponding number of neutrons is

$$n_1 = n_0 e^{at_1},\ fn_1,\ f^2 n_1,\ f^3 n_1,\ \ldots, \tag{8.6.5a}$$

where $f = e^{a\Delta t}$.

This way of describing the process, growth in the number of neutrons in geometric progression, is common in the popular literature (see footnote 4.14). Physicists and engineers speak rather of an *exponential* growth (in accord with the law of exponential increase) and rarely use (8.6.5) and (8.6.5a). The exponential law (8.6.4) is characterized by the ***growth rate*** a.

Let us find the dimensions of a. In (8.6.4) at is a dimensionless quantity and, consequently, the dimensions of a are s^{-1}. The same result can be obtained if we recall that

$$a = N\,(\nu - 1)\,\sigma_f v,$$

where N is measured in cm^{-3}, σ_f in cm^2, and v in cm/s.

Let us find the approximate value of the constant a. The density of uranium is roughly equal to 18 g/cm³. The number of nuclei per cubic centimeter, N, can be calculated by recalling Avogadro's number, which is equal to 6×10^{23} atoms in one gram-atom of any substance. Hence, 235 grams of uranium-235 contain 6×10^{23} atoms, or 6×10^{23} nuclei. One cubic centimeter of uranium-235 contains $(18/235) \times 6 \times 10^{23} \simeq 4 \times 10^{22}$ nuclei, that is, $N \simeq 4 \times 10^{22}$ $1/cm^3$. Substituting the mean value $\nu \simeq 2.5$, $v \simeq 2 \times 10^9$ cm/s, and $\sigma_f = (1/2)\,\pi\,(10^{-12})^2 \simeq 1.6 \times 10^{-24}$ cm^2 into the expression for a, we get

$$a \simeq 4 \times 10^{22} \times 1.5 \times 1.6 \times 10^{-24} \times 2 \times 10^9 \simeq 2 \times 10^8\ s^{-1},$$

or

$$a^{-1} \simeq 5 \times 10^{-9}\ s.$$

To summarize, if the neutrons do not leave the system, their number increases by a factor of e in approximately 5×10^{-9} second. At this rate of build-up, in one microsecond, or 10^{-6} s, the number of neutrons has increased by a factor of $e^{2\times 1.08\times 10^{-6}} = e^{200}$, that is, approximately, $10^{0.43\times 200} = 10^{86}$.

One metric ton of uranium-235 contains roughly 2.5×10^{27} nuclei. If the neutrons do not leave the system, this quantity of uranium will fission in less than one microsecond. This process is an explosion of enormous force.

Such a rate of buildup is not permissible if we want to use the fission process for generating electric power. It is necessary that neutrons leave the system and thus reduce the rate of neutron buildup.

8.7 Escape of Neutrons

Picture a mass of uranium-235 in the form a sphere of radius r. We have to set up an equation for the variation of the number n of neutrons inside the sphere. Assume for the sake of simplicity that the sphere is fixed to a thin support so that it is surrounded by a complete void and a neutron that has left the sphere will never enter it again.

How can we determine the neutron flux (the number of neutrons leaving the sphere in unit time)? We make a rough calculation. Consider a small time interval dt. During this time each neutron covers a distance of vdt. Where are the neutrons that leave the sphere in time dt? Evidently, they will have to be inside the sphere in a thin layer adjacent to the surface of the sphere but at a distance not exceeding vdt from the surface, otherwise during time dt they will not reach the surface, cross it, and leave it for good. But neither will all those neutrons that are inside the layer of thickness vdt be able to leave in time dt since not all neutrons inside the layer have velocity directed outward along the radius. In our very rough calculation we will ignore this latter circumstance.

How is it possible to find the number of neutrons in the layer? There are a total of n neutrons in the whole sphere. The volume of the sphere is $V = (4/3)\,\pi r^3$, while the volume of the thin layer that interests us near the surface is approximately equal to $Svdt$ if vdt is small. Here, $S = 4\pi r^2$ is the surface area of the sphere.

The mean ***density*** of neutrons (the number per unit volume) is $C = n/V$. Suppose that the density near the surface in the thin layer does not differ from the mean density. Then the number of neutrons in this layer is

$$CSv\,dt = \frac{nS}{V}\,v\,dt.$$

Therefore the ***flux*** (the number of neutrons leaving in unit time) is

$$q = \frac{nS}{V}\,v = \frac{n4\pi r^2}{(4/3)\,\pi r^3}\,v = \frac{3v}{r}\,n.$$

Actually, the neutron density near the surface is less than the mean density and, what is more (this was noted above), the neutron velocities have different directions and not all the neutrons leave the sphere. The neutron flux is thus less than we obtained:

$$q = \frac{3kv}{r}\,n, \tag{8.7.1}$$

where k is a numerical factor less than unity. Later on, in Section 8.10, we will compare our results with experiment and find that k is close to 0.3. If nuclear fission does not occur inside the sphere and no new neutrons are generated, then for the number of neutrons inside the sphere we get the equation $dn/dt = -q$ or, using (8.7.1),

$$\frac{dn}{dt} = -\frac{3kv}{r}\,n.$$

Setting

$$\frac{3kv}{r} = b, \tag{8.7.2}$$

we obtain

$$\frac{dn}{dt} = -bn.$$

The solution to this equation is familiar:

$$n = n_0 e^{-bt} \tag{8.7.3}$$

The mean residence time of neutrons inside the sphere is, by (8.7.3),

$$\bar{t} = \frac{1}{b} = \frac{r}{3kv}. \tag{8.7.4}$$

If $k = 0.3$, then (8.7.4) yields $\bar{t} \simeq r/v$. Therefore, the mean residence time is roughly equal to the time during which a neutron moving at a speed of v trav-

els a distance equal to the radius r of the sphere.

An exact consideration of the escape of neutrons requires extraordinarily laborious computations. It is very important from the start of one's studies to get used to approximate calculations of all quantities of interest. Exact calculations are frequently very involved and require quite a different range of knowledge, at times even the collective efforts of many workers and the use of electronic computers, and so on. But does this mean that a student engaged in self-instruction should give up the desire to consider a problem? There always exist simple, even though rough, methods (similar to the one just considered) for an approximate approach to a problem. To stop short of an approximate solution because the exact computations are complicated is merely to hide one's lack of courage. Very often, just such hesitancy is distructive of the first steps of a scientist or inventor!

8.8 Critical Mass

Up to now we have considered separately two processes: the multiplication of neutrons without regard for their escape and the escape of neutrons without regard for their multiplication.

Let us now consider a system in which neutrons multiply and can escape. As we know, in unit time, a times n neutrons are formed and b times n neutrons escape from the system. Since the variation of the number of neutrons in unit time is dn/dt, it follows that

$$\frac{dn}{dt} = an - bn,$$

or

$$\frac{dn}{dt} = cn, \qquad (8.8.1)$$

with $c = a - b$. For a given initial number of neutrons n_0, Eq. (8.8.1) has the solution

$$n = n_0 e^{ct}. \qquad (8.8.2)$$

This solution leads to quite different results depending on whether c is positive or negative. Indeed, from (8.8.2) it is evident that when c is negative the number of neutrons n falls off with increasing t, which means that *n tends to zero as $t \to \infty$*. But if c is positive, then n increases with t, that is, n grows *without limit* in the course of time. Only the effect of new physical factors not accounted for in Eq. (8.8.1) can halt the growth of n.

Thus the value $c = 0$ is a critical value, for it separates the distinct types of solution with increasing and decreasing number of neutrons. Since $c = a - b$, for a given a we can speak of the critical value $b_{cr} = a$, since $c = a - b > 0$ for $b < b_{cr}$ and $c = a - b < 0$ for $b > b_{cr}$. The quantity a is determined by the properties of the fissionable substance: according to (8.6.2), $a = Nv\sigma_f(\nu - 1)$. The quantity b depends on the amount of fissionable substance taken:

$$b = \frac{3kv}{r}.$$

The concept is therefore introduced of the ***critical value*** of the radius r_{cr} for which $b = b_{cr} = a$. From (8.6.2) and (8.7.2) it follows that $3kv/r_{cr} = Nv\sigma_f(\nu - 1)$, whence

$$r_{cr} = \frac{3k}{N\sigma_f(\nu - 1)}.$$

The mass of the sphere of radius r_{cr} is called the ***critical mass***, m_{cr}. It is clear that

$$m_{cr} = \frac{4}{3}\pi r_{cr}^3 \rho, \qquad (8.8.3)$$

with ρ the density of the fissionable substance. [8.9]

For $r > r_{cr}$ (this is the same as $m > m_{cr}$), $c > 0$ and we have a multiplication of neutrons. For $r < r_{cr}$ ($m < m_{cr}$), $c < 0$ and the original quantity of neutrons (exponentially) diminishes. Suppose we have a sphere

[8.9] As before, we consider the mass of a fissionable material, say, uranium, in the form of a sphere.

of radius r. Its surface area S and volume V are

$$S = 4\pi r^2, \quad V = \frac{4}{3}\pi r^3,$$

whence

$$\frac{S}{V} = \frac{4\pi r^2}{(4/3)\pi r^3} = \frac{3}{r}.$$

If r is small, this ratio is great. But if r is great, the ratio is small. No wonder that when the radius is small, that is, when the ratio of surface area to volume is great, the neutron escape increases and the conditions for neutron multiplication deteriorate. It is surprising how sharply the number of neutrons varies with b: if $b > b_{cr}$, in a short time the number of neutrons becomes practically zero, irrespective of whether $b = 1.01b_{cr}$ or $b = 2b_{cr}$. If $b < b_{cr}$, the number of neutrons increases without limit both for $b = 0.99b_{cr}$ and for $b = 0.5b_{cr}$, although the rate differs. This is precisely why one speaks of the critical value of b, the critical value of r, or the critical value of mass. When above critical, the mass is said to be ***supercritical***, when less than critical, it is called a ***subcritical*** mass.

In Figure 8.8.1 are given the curves $n = n_0 e^{(a-b)t}$ for a number of values of b. Let us construct the curves of n as a function of b for a few definite values of time t. In the computations, a is taken equal to 2×10^8 s^{-1}. Figure 8.8.2 shows the n versus b curves for $t = 5 \times 10^{-9}$ s, $t = 15 \times 10^{-9}$ s, and $t = 30 \times 10^{-9}$ s.

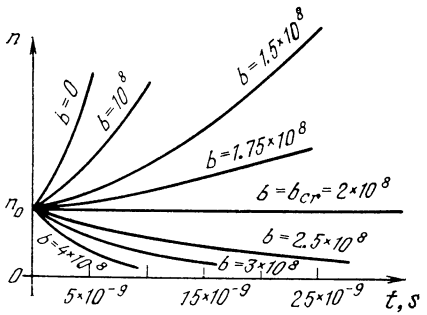


Figure 8.8.1

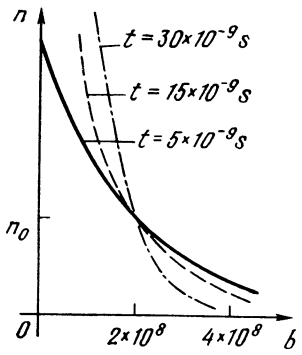


Figure 8.8.2

The curves corresponding to $t = 15 \times 10^{-9}$ s and $t = 30 \times 10^{-9}$ s intersect with the vertical axis ($b = 0$) at $n = 20n_0$ and $n = 400n_0$, respectively.

As is seen in Figures 8.8.1 and 8.8.2, the greater the time t, the more divergent are the n versus t curves in Figure 8.8.1, the steeper are the n versus b curves in Figure 8.8.2, and the more sharply is the criticality of the value $b = 2 \times 10^8$ s^{-1} (in this example) manifested.

If we take $t > 10^{-6}$ s, the n versus b curve cannot be distinguished from the vertical line $b = b_{cr} = 2 \times 10^8$ s^{-1}; $n = 0$ for $b > b_{cr}$ and $n = \infty$ for $b < b_{cr}$.

8.9 Subcritical and Supercritical Mass for a Constant Source of Neutrons

In the preceding section we considered the problem of the variation with time of the number of neutrons for a given initial number n_0 of neutrons. We now pose a somewhat different problem. Suppose at time zero ($t = 0$) the number of neutrons is zero and a *neutron source* is switched on at this instant of time, emitting q_0 neutrons per unit time. This problem leads to the equation

$$\frac{dn}{dt} = cn + q_0, \tag{8.9.1}$$

with $c = a - b$. We seek the solution to this equation with the initial condition $n = 0$ at $t = 0$.

A method of solution was given in Section 8.3 for a similar problem. We give a brief review of the reasoning there.

We seek the number of neutrons at time $t \neq 0$. The entire time interval from 0 to t is partitioned into subintervals $\Delta\tau$. We consider one such subinterval from τ to $\tau + \Delta\tau$. During this time the source emitted $q_0\Delta\tau$ neutrons. If the source operated only during one subinterval of time $\Delta\tau$, then we would be dealing with the problem of the preceding section with the initial number of neutrons $n_0 = q_0\Delta\tau$, the only difference being that these neutrons are emitted at time $t = \tau$ and not at time $t = 0$. Therefore, instead of the solution $n = n_0e^{ct}$ we would have the solution $n = n_0e^{c(t-\tau)} = q_0\Delta\tau e^{c(t-\tau)}$ (this solution refers to $t > \tau$; for $t < \tau$ we have $n = 0$), since clearly it is precisely on the time that elapsed after the initial number of neutrons was fixed that the number of neutrons depends, that is, in the given case, on the quantity $t - \tau$.

Actually, however, the neutron source is in constant operation during the whole time from 0 to t, and so we have to add the contributions of all neutrons emitted by the source in the various subintervals of time $\Delta\tau$, the sum of these intervals covering the entire time interval from 0 to t. Such a sum, given small subintervals $\Delta\tau$, is an integral, and so

$$n(t) = \int_0^t q_0e^{c(t-\tau)}\,d\tau.$$

It is easy to evaluate this integral:

$$n(t) = q_0e^{ct}\int_0^t e^{-c\tau}\,d\tau = q_0e^{ct}\,\frac{-1}{c}\,e^{-c\tau}\Big|_0^t$$

$$= q_0e^{ct}\,\frac{-1}{c}\,(e^{-ct}-1) = \frac{q_0}{c}\,(e^{ct}-1). \tag{8.9.2}$$

It is readily seen that this solution satisfies the equation

$$\frac{dn}{dt} = \frac{d}{dt}\left[\frac{q_0}{c}\,(e^{ct}-1)\right] = q_0e^{ct} = cn + q_0$$

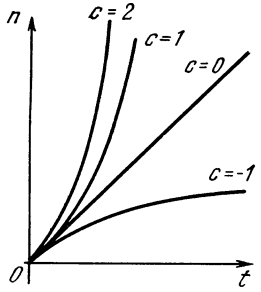


Figure 8.9.1

and the initial condition $n = 0$ at $t = 0$.

The same formula (8.9.2) yields a solution for a positive or negative value of c, though the shape of the n versus t curve is essentially different. For $c > 0$, that is, for $a > b$, the exponent ct is positive, so that e^{ct} quickly exceeds unity with increasing t. For large t and positive c we have

$$n \simeq \frac{q_0}{c}\,e^{ct}.$$

For $c < 0$ we have $ct < 0$, and so e^{ct} becomes much less than unity with increasing t, and the values of n approach the number $-q_0/c$ (this number is positive since $c < 0$). The n versus t curves are shown in Figure 8.9.1.

Note the curious case of $c = 0$. If $c = 0$, formula (8.9.2) cannot be used directly. Expand e^{ct} in a series:

$$e^{ct} = 1 + ct + \frac{(ct)^2}{2} + \ldots.$$

Substituting this into (8.9.1) yields

$$n(t) = \frac{q_0}{c}\left[1 + ct + \frac{(ct)^2}{2} + \ldots - 1\right]$$

$$= q_0\left[t + \frac{1}{2}\,ct^2 + \ldots\right].$$

This formula is valid for $c = 0$, too. We then have ($c = 0$)

$$n(t) = q_0t. \tag{8.9.3}$$

This result is also readily obtainable from (8.9.1). Indeed, Eq. (8.9.1) at $c = 0$ has the form $dn/dt = q_0$, whence $n(t) = q_0t + A$, where A is the constant of integration. For $t = 0$ it must

be true that $n = 0$, whence $A = 0$ and we arrive at (8.9.3).

As was shown above, when $c < 0$ the concentration of neutrons in the course of time attains a constant value q_0/c or, what is the same thing, $q_0/|c|$. The smaller the value of $|c|$ (the closer we are to the critical state), the greater this constant value. Thus, for a very weak source (small q_0), a mass close to critical can yield an arbitrarily large number of neutrons, a large number of fissions, and a great quantity of energy. Such in principle is the mode of operation of nuclear reactors.

The maintenance of such a regime is no easy task, since small variations in b and c drastically alter the magnitude of q_0/c when c is close to zero, and operation at c close to zero is necessary if we wish to obtain a big power output for small q_0. However, this engineering problem can be solved by means of automatic control: when n gets out of bounds, the control system changes a or b. Besides, there are also natural factors that facilitate control: for instance, when n increases, the temperature of the active material rises and then it turns out, c diminishes, so that to a certain extent the system is self-regulating.

8.10 The Critical Mass

We now know how sensitive the properties of a system are depending on whether we have a supercritical or subcritical mass. Let us examine in more detail the condition of criticality

$$r_{cr} = \frac{3k}{N\sigma_f(\nu - 1)}.$$

Substituting the numbers for uranium-235, $\sigma_f \simeq 1.6 \times 10^{-24}$ cm^2, $\nu \simeq 2.5$, $N \simeq 4 \times 10^{22}$ 1/cm^3, we get (in centimeters)

$$r_{cr} \simeq k \frac{3}{4 \times 10^{22} \times 1.6 \times 10^{-24} \times 1.5} \simeq 30k.$$

We do not know how to determine the coefficient k, all we know is that it is less than unity. Let us find this coefficient by comparing the formula with experiment. Experiments show that the critical mass of uranium-235 is about 50 kg. A uranium sphere weighing 50 kilograms has a radius of about 8.5 cm, so, in the given case,

$$k \simeq \frac{8.5}{30} \simeq 0.3.$$

Let us examine the physical significance of the formula for the critical radius. The neutron velocities cancelled out in the expression of r_{cr}, which means that the formula for r_{cr} can be obtained without regarding the course of the process in time and without examining the rate of neutron multiplication and the rate of neutron escape from the system.

If we disregard the dimensionless factor $3k$ (it is of the order of unity), the formula for the critical radius becomes

$$r_{cr} N \sigma_f \simeq \frac{1}{\nu - 1}. \qquad (8.10.1)$$

What is the quantity on the left? The volume of a cylinder of height equal to the radius and with area of the base equal to σ_f is $r_{cr}\,\sigma_f$. Recall that if a neutron is in motion along the axis of such a cylinder, then it causes fission of those nuclei of uranium-235 whose centers lie inside the cylinder. Since N is the number of nuclei in unit volume, we conclude that $N r_{cr}\,\sigma_f$ is the mean number of nuclei in the volume of the cylinder.

We can now give a different statement of the criticality condition. Earlier, we learned that the mean path, inside a fissionable material, of a neutron born inside the material (via fission) is of the order of the radius r. After a neutron has traversed a distance of about r, it leaves the fissionable material and is lost to the process. The criticality condition means that, on the average, prior to leaving the system, a neutron should produce one neutron over this distance. In fission, $\nu - 1$ new neutrons are generated. Hence,

it is necessary that the neutron, prior to escape, produce approximately $1/(\nu - 1)$ fissions, that is, that there be roughly $(1/(\nu - 1)$ nuclei in the volume of the cylinder of volume $r\sigma_f$. This is the condition that leads to formula (8.10.1).

Quite naturally, these arguments are not rigorous, but they are necessary for an understanding of the physical essence of the matter and cannot be replaced by any kind of calculations, even the most precise ones performed on modern electronic computers. Computer executed computations do not replace but merely supplement a clear-cut grasp of the qualitative physical aspect of the matter. In particular, the reader should pay special attention to the principle expressed at the beginning of the section: if some quantity (v) enters into the derivation of a formula but is cancelled out in the final result, this means that there is a derivation of this formula that dispenses altogether with that quantity. And one should always find that simpler derivation because a different derivation of a formula is tantamount to a fresh view of the process being investigated.

Chapter 9 Mechanics

9.1 Force, Work, and Power

The relations existing between the most important quantities of mechanics admit of exact formulations only by means of integrals and derivatives. In Chapter 2 we examined the relationship between the distance covered by, or the position of, a body, z, and its velocity, v, and also between the velocity v and the acceleration, a, precisely, $v = dz/dt$ and $a = dv/dt$. We now go on to examine the relationships between quantities such as force, work, energy, and power.

Let us consider the rectilinear motion of a body along the x axis. Suppose a force F acting on the body is also directed along the x axis. The work A performed by this force is defined as the product of the force F by the distance traversed by the body, $l = b - a$, where a is the initial position of the body and b is the terminal position:

$$A = Fl = F(b - a).$$

Obviously, the situation is the same as in the case of the relationship between velocity and distance, that is, the simple formula—work is equal to the product of force by distance—is valid only for the case where the force is *constant*. Now if the force *varies* during the process of the body's translation, the whole process has to be partitioned into separate small intervals (subintervals) so that over every subinterval the force may be taken to be constant (for this to be true each subinterval must correspond to a small increment of time or distance). Then for a small segment Δx_i of the translation corresponding to the *ith* subinterval (from position x_i of the body to position x_{i+1}) the work is

$$\Delta A_i = F_i \Delta x_i = F_i (x_{i+1} - x_i).$$

This means that in the general case of a variable force $F = F(x)$, the work is expressed not as a product but as an *integral*:

$$A = \int_a^b F\,dx.$$

We assume as known the motion of a body given by a known function $x = x(t)$. The translation dx of the body during a small time interval dt is equal to the product of the (instantaneous) velocity v by the time dt:

$$dx = v\,dt = \frac{dx}{dt}\,dt.$$

Therefore, the expression for work can be written thus:

$$A = \int_\alpha^\beta F \frac{dx}{dt}\,dt = \int_\alpha^\beta Fv\,dt, \qquad (9.1.1)$$

where the moments $t = \alpha$ and $t = \beta$ correspond to beginning and end of the body's motion.

The product Fv in this formula is the work performed in unit time and is called the ***power***. Indeed, in the case of constant velocity and force, the distance is equal to $x = vt$, the work is $A = Fx = Fvt$, and the ratio of work to the time elapsed (that is, the work performed in unit time, or power) is $A/t = Fv$. Denoting power A/t by W, we can write

$$A = \int_\alpha^\beta W\,dt. \qquad (9.1.1a)$$

Recall that in the SI system of units the unit of velocity is measured in m/s and the unit of acceleration is measured in m/s^2. The unit of force has a special name, the ***newton*** (denoted N), which is the force that imparts an acceleration of 1 m/s^2 to a mass of 1 kg. Clearly, the unit of energy or work is 1 N·m = 1 kg·m^2/s^2; it is called the ***joule*** (denoted J). Finally, the unit of power is 1 N·m/s = 1 kg·m^2/s^3 and is known as the ***watt*** (denoted W).

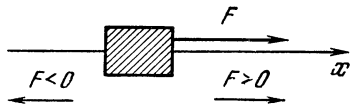


Figure 9.1.1.

A body can be acted upon by *several* forces simultaneously, say, F_1 and F_2. Then we can speak of the work performed by the first force, A_1, and that performed by the second force, A_2, during the time that the body was translated from the initial position a to the terminal position b. Regarding forces F_1 and F_2 as constant, we obtain

$$A_1 = (b - a)\,F_1, \quad A_2 = (b - a)\,F_2.$$

Note the signs of the quantities in these expressions. A force is taken to be *positive* when it acts in the direction of increasing x (Figure 9.1.1), while a force acting in the opposite direction (to the left) is regarded as *negative*. If the body is translated in the direction of the acting force, the work of that force is positive. But if the body is translated in the direction opposite that of the force, so that F_1 and $b - a$ have different signs, the work A of the force is negative. Now picture two forces acting on a body (Figure 9.1.2*a*): the force F_1 of a stretched spring and the force F_2 of the tension of a rope which you (the reader) hold in your hand. F_1 acts leftwards, $F_1 < 0$, while you are pulling rightwards, $F_2 > 0$. If you pull with more strength (which means that the absolute value of the force with which you pull rightwards is greater than the absolute value of the force with which the spring pulls the body leftwards, or $|F_2| > |F_1|$), the body, which initially was in the state of rest, will move from left to right. Figure 9.1.2*a* shows the initial position of the body and Figure 9.1.2*b* the terminal position: $(b - a) > 0$ and $F_1 < 0$. The work A_1 performed upon the body by the tension force of the spring, or more briefly, the work of the spring, in this translation is negative, while the work which you have performed is positive, $A_2 > 0$. The total work $A = A_2 + A_1$ is also positive, but $A < A_2$ since $A_1 < 0$. This means that only part (A) of the work performed by you (A_2) was received by the body, the other part ($|A_1|$) having gone into stretching the spring. Observe that in all cases the force of friction against a stationary surface is directed *against* the velocity of motion of the body, and so the work of the friction force against a fixed surface is always *negative*, irrespective of the direction of the motion of the body.

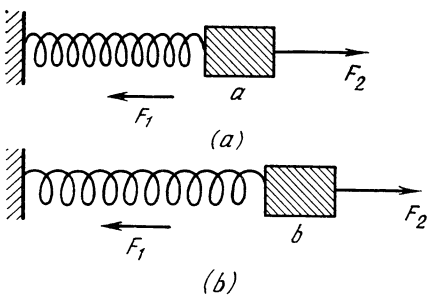


Figure 9.1.2

The force F_1 with which a spring, one end of which is fixed, acts on a body differs in one very important way: it depends exclusively on the position of the body. Not all forces, by any means, have this property. For example, the force of friction between a moving body and a fixed surface always retards the motion of the body: it is directed leftwards if the body is in motion rightwards, and it is directed rightwards if the body is in motion leftwards. Thus, the direction of the force of friction depends on the direction of motion of the body. Besides, the force of friction can depend on the magnitude of the velocity of the body. Thus, the force of friction depends on the magnitude and direction of the body's velocity and not only on the position of the body (in fact, it may even be independent of the position of the body).

The force F_2 with which you pull the rope in the example of Figure 9.1.2 can vary in any fashion, at your pleasure. The body can, say, move to the right and then to the left. In so doing

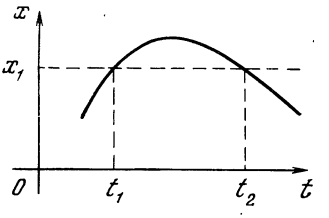


Figure 9.1.3

it will twice pass through the same position: the first time in the rightward movement at t_1, the second time on the return route at time t_2.

A possible graph of the motion of the body (the dependence of the x coordinate on the time t) is shown for this case in Figure 9.1.3. We can, at our will, at time t_1 pull the body to the right, $F_2(t_1) > 0$, and at time t_2 let go of the rope so that $F_2(t_2) = 0$ or even push the body leftwards so that $F_2(t_2) < 0$. But $x(t_1) = x(t_2) = x_1$ and so, speaking generally, an arbitrary force F_2 cannot be regarded as a function of the x coordinate.

The foregoing examples of the force of friction and the force applied by a person acting on his own free will serve to demonstrate that the dependence of force solely on the position of a body, $F_1 = F_1(x)$, which is characteristic of the force F_1 with which a spring acts on a body, is not a general property of all forces, but is a particular property associated with the elasticity of a spring.

To find the work A_i performed by a given force F_i (where the subscript i shows that we are speaking of one of the several forces, F_1, F_2, ..., acting on the body) one must use one of the formulas

$$A_i = \int_a^b F_i\,dx \quad \text{or} \quad A_i = \int_\alpha^\beta F_i v\,dt.$$

We have to know two things: (a) what the motion of the body was, that is, the dependence of the x coordinate of the body on time t, or $x = x(t)$, and (b) the expression for the force $F_i = F_i(x, t, v)$, which in general depends on x, t, and v.

Knowing the functions $x(t)$ and $v(t)$ and substituting them into the expression for $F_i(x, t, v)$, we arrive at an expression for F_i as a function of time and we can describe the work as an integral with respect to time. Note that other forces may also be acting on the body, with the result that each separate force F_i cannot be expressed in terms of the acceleration and mass of the body.

Example. Suppose we have a force $F(x) = -kx$ acting on a body and let the motion of the body be given by the equation $x = B\sin\omega t$, that is, $F(x, t) = -kB\sin\omega t$ and $v = dx/dt = B\omega\cos\omega t$; we do not require that the force F be equal to the mass of the body multiplied by the acceleration: it is assumed that other forces Q act on the body, which together with F ensure that the motion of the body obeys the given law $x = x(t)$. The work performed solely by force F can easily be found:

$$A = -B^2k\omega\int_{t_1}^{t_2}\sin\omega t\cos\omega t\,dt$$

$$= \frac{B^2k}{2}\int_{t_1}^{t_2}\sin 2\omega t\,dt = \frac{B^2k}{4}\cos 2\omega t\Big|_{t_1}^{t_2}$$

$$= \frac{B^2k}{4}(\cos 2\omega t_2 - \cos 2\omega t_1) \qquad (9.1.2)$$

(verify this).

In this case, where the force depends solely on the coordinate, it is much easier and convenient to take advantage of the expression of work as an integral with respect to x:

$$A = \int_a^b F(x)\,dx - k\int_a^b x\,dx$$

$$= \frac{ka^2}{2} - \frac{kb^2}{2}.$$

Substituting $x = B\sin\omega t$, we can also easily obtain an expression for work over a specified interval of time from t_1 to t_2:

$$A = \frac{kB^2 \sin^2 \omega t_1}{2} - \frac{kB^2 \sin^2 \omega t_2}{2}. \quad (9.1.3)$$

It is easy to see that this expression coincides exactly with the preceding one since

$$\cos 2\omega t = \cos^2 \omega t - \sin^2 \omega t$$
$$= 1 - 2\sin^2 \omega t,$$

whence

$$\cos 2\omega t_2 - \cos 2\omega t_1$$
$$= 1 - 2\sin^2 \omega t_2 - (1 - 2\sin^2 \omega t_1)$$
$$= 2(\sin^2 \omega t_1 - \sin^2 \omega t_2).$$

Substituting this identity into (9.1.2), we arrive at (9.1.3).

A good deal of caution is required when using the expression for work as an integral with respect to the x coordinate of a force $F(x, v, t)$ depending, generally, on x, v, and t. Indeed, in principle, if the motion $x = x(t)$ is given, this equation can be solved for t and we can determine $t(x)$. But one must bear in mind that t may *not be a single-valued* function of x, that is, one and the same position x may correspond to *two* distinct instants of time, which means that one and the same value of x is associated with two distinct values of t (see Figure 9.1.3). Then the overall motion has to be divided into separate periods during which the velocity does not change sign and t is a *single-valued* function of x. But for different periods, t is expressed by unlike functions of x.

For example, let a body be moving via the law $x = B\sin\omega t$, as in the preceding example, but the force be given as a function of time, $F = f\cos\omega t$. The force is then not a single-valued function of position x. Indeed, let $t = 0$, then $x = 0$ and $F = f$. But if we put $t = \pi/\omega$, again $x = 0$ but $F = -f$, so that the body will be in the same position $x = 0$ at different times ($t = 0$ and $t = \pi/\omega$) though the force will not be the same. This difficulty can be avoided when integrating with respect to time, since to every instant of time t there corresponds one definite value of the x coordinate, of the force F, and of all other quantities.

It is easy to find the work by integrating with respect to time:

$$A = \int_{t_1}^{t_2} Fv\,dt = \int_{t_1}^{t_2} f\cos\omega t\, B\omega\cos\omega t\,dt$$
$$= fB\omega\int_{t_1}^{t_2}\cos^2\omega t\,dt.$$

Let us take advantage of the above trigonometric formula $\cos 2\varphi = 2\cos^2\varphi - 1$ or, which is the same, $\cos^2\varphi = 1/2 + (\cos 2\varphi)/2$. Then

$$A = fB\omega\int_{t_1}^{t_2}\left(\frac{1}{2} + \frac{\cos 2\omega t}{2}\right)dt$$
$$= \frac{1}{2}fB\omega(t_2 - t_1) + \frac{1}{4}fB(\sin 2\omega t_2$$
$$- \sin 2\omega t_1). \quad (9.1.4)$$

Motion in accordance with the law $x = B\sin\omega t$ represents oscillations of the body (see Chapter 10). As evident from (9.1.4), the work increases without bound with the passage of time, that is, as t_2 increases. This is due to ***resonance*** that occurs between the force and the oscillations of the body (resonance will be examined in detail also in Chapter 10).

Consider the work performed by the force during one half-period, choosing for the initial time $t_1 = 0$, $x_1 = 0$, and the terminal time $t_2 = \pi/\omega$, $\sin\omega t_2 = \sin\pi = 0$, $x_2 = 0$. Then, in (9.1.4), $\sin 2\omega t_2 = \sin 2\omega t_1 = 0$, and the work is

$$A = \frac{1}{2}fB\omega\frac{\pi}{\omega} = \frac{\pi}{2}fB. \quad (9.1.5)$$

The body has returned to its initial state, while the work performed by the force is not equal to zero but has a definite magnitude. How is this result to be understood from the viewpoint

of the first formula $A = \int_{x_1}^{x_2} F\,dx$? At first glance, if we substitute $x_1 = x_2 = 0$, we get

$$A = \int_0^0 F\,dx = 0.$$

Actually, however, we have to consider separately the process of buildup of x from 0 to $x_{\max} = B$ and the process of decline of x from $x_{\max} = B$ to 0. During buildup, each value of x is associated with a definite value of the force F, which we denote by F_1:

$$F_1 = f\cos\omega t = f\sqrt{1-\sin^2\omega t} = f\sqrt{1-\left(\frac{x}{B}\right)^2} > 0.$$

During decline of x, the same positive values of x are associated with a (*negative*) value of force. [9.1] We denote this negative force by F_2:

$$F_2(x) = -f\sqrt{1-\left(\frac{x}{B}\right)^2}.$$

Thus the integral with coordinate x for the variable of integration breaks up into two,

$$A = \int_0^B F_1(x)\,dx + \int_B^0 F_2(x)\,dx. \qquad (9.1.6)$$

These two integrals cannot be combined by the formula

$$\int_a^b \varphi\,dx + \int_b^c \varphi\,dx = \int_a^c \varphi\,dx,$$

since the integrands in the two integrals on the right-hand side of (9.1.6) are expressed by different formulas, although their meaning is the same (force). This is due to the fact that F is given as a function of t, while t is expressed in terms of x by different formulas for increase of x from 0 to B and for decrease of x from B to 0. In this case, $F_2(x) = -F_1(x)$. Substituting the expressions for $F_1(x)$ and $F_2(x)$ into (9.1.6), we obtain

$$A = f\int_0^B \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx - f\int_B^0 \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx.$$

In the second integral we can interchange the limits of integration (this changes the sign of the integral) to get

$$A = 2f\int_0^B \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx. \qquad (9.1.7)$$

If we put $z = x/B$, then we have $dx = B\,dz$ and

$$A = 2Bf\int_0^1 \sqrt{1-z^2}\,dz.$$

But the integral

$$I = \int_0^1 \sqrt{1-z^2}\,dz = \frac{\pi}{4}$$

(the area of a quadrant of a circle of radius equal to unity), and so from (9.1.7) we get

$$A = 2Bf\,\frac{\pi}{4} = \frac{\pi}{2}Bf,$$

which coincides with formula (9.1.5) obtained by integrating with respect to time.

Thus, in the case of a force that is dependent on time and can assume different values for the same value of x, the work A is not a single-valued function of x either. In the foregoing case of oscillatory motion, $F = f\cos\omega t$, $x = B\sin\omega t$, the quantity x again and again passes through the same values in the course of time, and the work performed by the force (for positive f) continues to increase all the time.

If the force is a function of velocity (as is the case of friction), the situation

[9.1] The equation $\cos^2\omega t + \sin^2\omega t = 1$ is true for *all* values of ωt. From this it follows that $\cos\omega t = \pm(1-\sin^2\omega t)^{1/2}$, while the sign depends on the value of ωt. It is easy to see that for $-\pi/2 < \omega t < \pi/2$ one has to take the plus sign and for $\pi/2 < \omega t < 3\pi/2$ the minus sign, which was done above.

will be similar to the one discussed above: the body can return to its original position, but the work of the force will not be zero. In the case of friction the force is negative (see the exercises below).

Exercises

9.1.1. Find an expression in the form of an integral for the work of friction, the force of friction being proportional to the velocity of the body and in the opposite direction, $F = -hv$, with h positive. Demonstrate that the work is negative.

9.1.2. Suppose the force of friction is constant in magnitude and opposite in direction to the velocity, that is, $F = -h$ for $v > 0$ and $F = +h$ for $v < 0$. Suppose the body moves in accordance with the law $x = B \sin \omega t$. Find the work of the force of friction during the time interval from $t = 0$ to $t = \pi/\omega$.

9.1.3. The force acting on a body is given by the formula $F = f_0 \sin \omega_0 t$, with f_0 constant. Since the body is also acted upon by other forces, it moves according to the law $x = B \sin \omega_1 t$. Determine the work performed by force F during the time interval from $t = 0$ to $t = T$. Consider the case $\omega_0 = \omega_1$.

9.1.4. A body is falling according to the law $x = gt^2/2$ (the x axis is directed downwards). Find the formula for the work resulting from the air-resistance force $F = -aS\rho v^2/2$, with a being a constant of proportionality dependent on the shape of the body (see Sections 9.14 and 9.15), S the cross-sectional area of the body (in cm^2), ρ the air density (about 1.3×10^{-3} g/cm^3), and v the rate of fall (in cm/s). Also find the formula for the work done by the force of gravity $F = mg$, where m is the mass of the body.

Perform the computations and compare the results for a wooden ball of diameter 1 cm, $a = 0.8$, and for a steel bullet of length 3 cm, diameter 0.7 cm, $a = 0.2$, for $t = 1$ s, 10 s, 100 s.

Remark. The idea behind the calculation is that we assume the force of air resistance to be small compared with the force of gravity and not noticeably affecting the law of free fall. Computing the work done by the air resistance and comparing it with the work performed by gravity, we verify the correctness of our starting assumption concerning the small role of the force of air resistance. In Section 14 we give an exact solution of the problem of free fall with air resistance.

9.1.5. A wind blowing with a speed v_0 acts on the sail of a boat with a force F equal to $aS\rho (v_0 - v)^2/2$ for $v < v_0$ and $-aS\rho (v_0 - v)^2/2$ for $v > v_0$, is the speed of the boat, S the area of the sail, ρ the air density, and a a dimensionless coefficient (a is approximately unity for the sail put up perpendicular to the wind). Find the work done by the force of the wind in moving the boat b meters and the power of the wind force. (It can be assumed that the boat is in uniform motion, that is, the speed v is constant). Determine the work and power as functions of v. Finally, find the maximum power for $v_0 = 30$ m/s, $a = 1$, $S = 100$ m^2, and express the result in watts.

9.1.6. A body is moving according to the law $x = B \cos (\omega t + \alpha)$ under several forces, including a force that is time-dependent, $F = f \cos \omega t$. Find the work performed by the force during the time interval from $t = t_1$ to $t = t_2$, in particular, during one period of operation (from $t = 0$ to $t = 2\pi/\omega$). Determine the average power of the force.

9.2 Energy

We consider the case of a force that depends solely on the position (coordinate) of the body, $F = F(x)$. As we have already noted, an instance of this kind of force is the force with which a spring acts on a body, the other end of the spring being fixed. [9.2] In that case the expression $A = \int_{x_1}^{x_2} F\,dx$ may be applied without any complications (compare with Section 9.1). In particular, in this case if the body first moves in one direction from x_1 to $x_{\max} = X$ and then in the opposite direction returning to the initial position, we have $x_2 = x_1$, and the total work done by the force is actually equal to zero:

$$A = \int_{x_1}^{x_2 = x_1} F(x)\,dx = 0.$$

Dividing the path length into sections only corroborates this conclusion:

$$A = \int_{x_1}^{X} F\,dx + \int_{X}^{x_2} F\,dx$$
$$= \int_{x_1}^{X} F\,dx - \int_{x_2}^{X} F\,dx,$$

and $A = 0$ at $x_1 = x_2$.

[9.2] If the second end of the spring is allowed to move at random, the force acting on the body will depend not only on the position of the body but also on the position of the second end of the spring and this does not satisfy the stated condition.

In mechanics, ***potential energy*** is defined as the capacity to do work. A spring possesses a definite reserve of potential energy depending on how compressed or stretched it is. If one end is fixed in position, the potential energy of the spring depends on the position of the body to which the free end of the spring is attached. Thus, the potential energy $u = u(x)$ is a function of the x coordinate. If in the initial position $x = x_1$ the potential energy is $u(x_1)$, after the body has been displaced to position $x = x_2$, when the spring has performed work A equal to

$$A = \int_{x_1}^{x_2} F(x)\,dx,$$

the remaining potential energy is equal to $u(x_1) - A$. Thus,

$$u(x_2) = u(x_1) - A = u(x_1) - \int_{x_1}^{x_2} F(x)\,dx. \tag{9.2.1}$$

Get a good feeling of the sign affixed to A in this expression: if the spring does (positive) work, the reserve capacity of the spring to do work will diminish! The work performed by the spring is taken from the reserve of potential energy. For this reason the work done (that given up by the spring) is equal to the difference between the initial and final energy of the spring:

$$A = u(x_1) - u(x_2).$$

All formulas involve the *difference* of potential energy in two positions of a body. Therefore, if we replace $u(x)$ by $u(x) + C$, where C is an arbitrary constant, this will in no way affect the physical results. Indeed,

$$[u(x_1) + C] - [u(x_2) + C] = u(x_1) - u(x_2).$$

The value of $u(x)$ at some given point, call it x_0, can be chosen quite arbitrarily. Denote it by u_0. Then at some other point x the value of the function $u(x)$ is determined from the formula (9.2.1) if in it we put $x_1 = x_0$ and $x_2 = x$:

$$u(x) = u_0 - \int_{x_0}^{x} F(x)\,dx. \tag{9.2.2}$$

That is how the problem of *determining the potential energy from a given force* is solved.

We can pose the converse problem, namely, *knowing the potential energy as a function of* x, *or* $u = u(x)$, *find the force* $F(x)$. To solve this problem, take the derivative of both sides of (9.2.2). The derivative of the integral (with respect to the upper limit) is equal to the integrand, so that

$$\frac{du(x)}{dx} = -F(x). \tag{9.2.3}$$

The minus sign here is essential. The force is positive (in the direction of increasing x) if du/dx is negative, that is, as x increases, the potential energy u decreases. The force is negative (in the direction of decreasing x) if du/dx is positive, that is, as x increases, the energy u increases, too. In the latter case, obviously, as x decreases, the energy u also decreases. This means that the force is always in the direction of *diminishing* potential energy.

Let us examine in more detail the example of the *spring*. Let the body be at the origin when the spring is not under tension (Figure 9.2.1). When the body is pulled to the right, the force of tension grows in direct proportion to the displacement of the body and is directed leftwards:

$$F = -kx, \quad k > 0. \tag{9.2.4}$$

Assume that $u_0 = 0$, at $x = 0$, that is, regard the potential energy for the

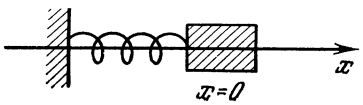


Figure 9.2.1

nontense spring as zero. This yields

$$u(x) = -\int_0^x F\,dx = k\int_0^x x\,dx = k\frac{x^2}{2}. \tag{9.2.5}$$

It is easy to see that this $u(x)$ is associated, via formula (9.2.3), with the force (9.2.4).

We consider a second example, the *force of gravity*. Send the z axis upward. The force of gravity acts downward and is equal to $-mg$, where g is the acceleration of gravity. It is independent of the height z, but a constant quantity is merely a special case of a function whose values are the same. The important thing is that the force of gravity does not depend on time or velocity. We can therefore make use of the formulas derived above. We take as zero the potential energy of the body at the earth's surface, where $z = 0$. Then

$$u(z) = -\int_0^z F\,dz = -\int_0^z (-mg)\,dz = mgz. \tag{9.2.6}$$

The potential energy grows linearly with increasing height of the body above the earth's surface.

In the latter example we assumed that the distance z is small compared with the radius of the earth. Let us now examine the attractive force on the assumption that the distances can be arbitrarily large. By Newton's law of gravitation, the attractive force is inversely proportional to the square of the distance between the bodies. We know that for a body above the earth's surface the force of gravitation towards the entire globe is equal to the force of attraction to a mass equal to the earth's mass and concentrated at the earth's center. [9.3] It is therefore convenient to reckon distances from the center of the earth and denote them by r. Then the force of attraction acting on a body is directed towards the earth's center and is C/r^2 in magnitude, where the constant C is positive.

The constant C can readily be determined from the condition that the force acting at the surface of the earth ($r = r_0 \simeq 6400$ km $= 6.4 \times 10^6$ m) is known: $F(r_0) = mg = C/r_0$, that is, $C = mgr_0^2$, where g is the acceleration of gravity at the earth's surface equal approximately to 9.81 m/s^2). We finally have

$$|F| = \frac{mgr_0^2}{r^2}. \tag{9.2.7}$$

For zero we again take the potential energy of the body at the earth's surface. Allowing for the fact that as the distance r from the earth's center increases the work performed by force F in the process is negative, we get

$$u = -\int_{r_0}^r F\,dr = mgr_0^2\int_{r_0}^r \frac{dr}{r^2}$$
$$= mgr_0^2\left(-\frac{1}{r}\Big|_{r_0}^r\right) = mgr_0^2\left(-\frac{1}{r}+\frac{1}{r_0}\right)$$
$$= mg\frac{r_0}{r}(r - r_0). \tag{9.2.8}$$

At a small height $z = r - r_0 \ll r_0$, the ratio r_0/r differs but slightly from unity, whence

$$u(r) \simeq mg(r - r_0) = mgz, \tag{9.2.9}$$

which coincides with formula (9.2.6) obtained earlier. But, as can be seen from (9.2.8), the potential energy does not increase without limit as r increases, as would have been the case in accordance with the approximate formula (9.2.6), but tends to a definite limit

$$u(\infty) = mgr_0. \tag{9.2.10}$$

Thus, making allowance for the decrease of gravity with distance, we can say that the energy of a body at an infinite distance from the earth is the same as, by the approximate formula, at a distance of r_0 from the earth's surface, or at a distance of $2r_0$ from the earth's center.

[9.3] This does not hold true for a body inside the earth, in which case we must allow only for the portion of the earth's mass between the earth's center and the body.

In this problem we encountered a physical situation involving *infinite* distance. In this respect we must note that in any physical problem we are always interested in *finite* quantities, or in our case finite distances. For instance, if we consider the motion of a body and the energy of the body as dependent on the earth's gravity, then we can be interested in attaining the moon, Mars, or other planets, or even stars. All these objects involve distances that are very very great relative to that of the earth's radius, but they are finite! And, obviously, for a physicist the expression $r = \infty$ means only that $r \gg r_0$.

Suppose we consider the problem of launching a rocket to a great height, to a considerable distance from the earth. We are interested in the energy required and the time of flight. Here are two cases.

(a) A space vehicle is to traverse a distance of $R = 10r_0$, where r_0 is the earth's radius.

(b) A space vehicle is to traverse $R = 100r_0$.

The work needed to tear away from the earth and go a distance R from the earth's center is

$$A = mgr_0^2 \left(\frac{1}{r_0} - \frac{1}{R}\right). \qquad (9.2.11)$$

Recalling that $r_0 \simeq 6.4 \times 10^6$ m, we get $A_1 \simeq mg \times 5.76 \times 10^6$ in the first case and $A_2 \simeq mg \times 6.34 \times 10^6$ in the second.

A ten-fold change in distance caused a relatively small change in the energy required. If we were to replace R by infinity, we would get $A_\infty \simeq mg \times 6.4 \times 10^6$ (note that ∞ is not a number). A_1 differs from A_∞ by 10%, and A_2 differs by 1%. That is why, when computing work, R may be replaced by infinity. But a change in R has a strong effect on the *time* of flight, and for this reason when considering the time of flight we must never replace R by infinity.

To summarise, then, one and the same quantity R in one and the same problem can either be replaced by infinity or not, depending on the aspect considered. The possibility of such a substitution depends not only on the quantity R itself (and its comparison with other quantities of the same dimensions entering into the formulas, r_0 in the given case) but also on the structure of the formula in which it occurs.

Returning to the question of potential energy of a body attracted to the earth, let us find the numerical value of $u(\infty)$ per unit mass—it is equal to $gr_0 \simeq 9.81 \times 6.4 \times 10^6 \simeq 6.28 \times 10^4$ J/g, which is 30 times the heat of evaporation of water and 10 times the chemical energy of explosives.

In problems of celestial mechanics and in physics it is advisable to choose for zero the potential energy of a body located at an *infinite* distance from the mass attracting it. Then for the potential energy of a body at a distance r we have

$$u(r) = u(\infty) - \int_\infty^r F(r)\, dr = -\frac{C}{r}, \qquad (9.2.12)$$

where C is the constant in the expression for the force, $F = -C/r^2$, and can be determined from the formula $C = mgr_0^2$ if we know the acceleration of gravity g at the earth's surface and the radius of the earth, r_0.

We can obtain a different expression for C. By Newton's law of gravitation, $F = -GmM/r^2$, where m is the mass of a body attracted to the earth, M is the mass of the earth, r the distance to the center of the earth, and G the gravitational constant equal to approximately 6.7×10^{-11} $\mathrm{N \cdot m^2/kg^2} = 6.7 \times 10^{-11}$ $\mathrm{m^3/kg \cdot s^2}$. Therefore $C = GmM$. Using this formula, we can easily determine C if we know G and M.

Indeed, by measuring the attraction of two heavy balls of known masses we can find G. Only after this, by measuring the attraction of a body to the earth can we find the earth's mass. The form of the function expressing Newton's law of gravitation, that is,

the fact that the force is inversely proportional to the square of the distance, is proved by comparing the attraction to the earth of a body at the earth's surface with the attraction of a remote body, the moon, as well as by comparing the attraction to the sun of planets that orbit the sun at different distances—from Mercury to Pluto.

The problem of potential energy of *two electric charges* e_1 and e_2 is quite similar to the preceding one. The interaction force between the charges is equal to

$$F = k\frac{e_1 e_2}{r^2}. \qquad (9.2.13)$$

Here, if the charges are expressed in electrostatic units (the unit of charge is $(1/3) \times 10^{-9}$ C (coulomb)) and the force in dynes, (1 dyne $= 10^{-5}$ N) k is equal to unity; in SI, where charge is measured in coulombs and force in newtons, $k = (1/9) \times 10^{-13}$. There is no minus sign in (9.2.13) that we see in the expression for the gravitational force. Indeed, if e_1 and e_2 are like charges (both positive or both negative), the product $e_1 e_2$ is positive. But like charges repulse one another, that is, the force F is positive.

Again defining $u(r)$ so that $u(\infty) = 0$, we obtain

$$u(r) = k\frac{e_1 e_2}{r}. \qquad (9.2.14)$$

The potential energy of two like charges separated by a finite distance is positive: they repulse and, moving from r to ∞ can perform work equal to

$$u(r) - u(\infty) = u(r).$$

The potential energy of two unlike charges is negative. Indeed, $e_1 e_2 < 0$ if, say, $e_1 > 0$ and $e_2 < 0$. This is clear physically: since unlike charges attract each other, energy must be expended to pull them apart to infinity.

Note that thanks to the law of conservation of energy, the potential energy may be defined not only as the capacity to do work but also as the work required to bring a system to a given state. A stretched spring can do a definite amount of work in returning to the unstretched state. That, clearly, was the work that had to be done to stretch the spring in the first place. Similar assertions may be made in the case of a body raised to a definite height above the earth or for a system of two charges.

9.3 Equilibrium and Stability

We consider a body that can move without friction along a straight line, which we take for the x axis. Let the body be acted upon by a force directed along this axis and dependent on the x coordinate. We can again picture the *spring*. Below we will examine other examples as well.

The *equilibrium* position of a body is defined as that position for which the force is zero and the body is at rest. Denote the point of equilibrium by x_0. Then $f(x_0) = 0$. Expanding the function $F = F(x)$ in a Taylor series and ignoring all powers of $x - x_0$ except the first, we see that two versions of the function $F(x)$ are possible in the neighbourhood of point x_0 (provided $F(x_0) = 0$), namely, $F(x) \simeq k_1(x - x_0)$ and $F(x) \simeq -k_2(x - x_0)$; in both formulas it is assumed that k_1 and k_2 are positive quantities. The first case is shown in Figure 9.3.1*a*, the second in Figure 9.3.1*b*.

These two cases are associated with an entirely different character of equilibrium. In the case of Figure 9.3.1*a*, if the body is somewhat to the right of

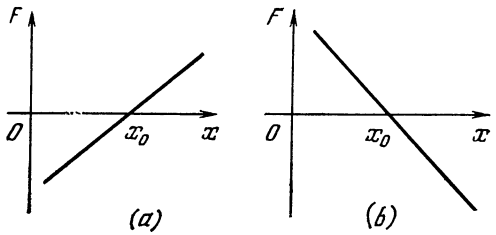


Figure 9.3.1

point $x = x_0$, then it is acted upon by a positive force, that is, a force which pulls it farther rightwards. Thus, the equilibrium at point $x = x_0$ in Figure 9.3.1*a* is ***unstable***. A slight deviation of the body (whether to the right or left makes no difference) suffices for a force to begin to act on the body, and this force will *increase* the deviation. On the other hand, in the case of Figure 9.3.1*b* the force is negative (pulls leftwards) when the body deviates to the right. Deviation of the body from the equilibrium position gives rise to a force that tends to *return* the body to the position of equilibrium. Here we have to do with ***stable*** equilibrium. It is easy to see that for a body attached to a spring the second case is realized.

In accordance with the above expressions for force, we find the expressions of potential energy via (9.2.2). In the case of unstable equilibrium,

$$u(x) \simeq u(x_0) - \frac{1}{2} k_1 (x - x_0)^2.$$

In the case of stable equilibrium,

$$u(x) \simeq u(x_0) + \frac{1}{2} k_2 (x - x_0)^2.$$

The appropriate curves are shown in Figure 9.3.2*a* and *b*.

Thus, in the case of unstable equilibrium, the potential energy has a *maximum*, while in the case of stable equilibrium it has a *minimum*. In both cases the force is zero at the point of maximum or minimum, that is

$$F = -(du/dx)_{x=x_0} = 0.$$

This result is quite natural. If a body is in the state of maximum potential energy, energy is released during displacements in both directions. This energy can be used to overcome inertia and is converted into kinetic energy. But if the body is in the state of minimum energy, energy from an outside source is required to move it to any other position. This energy will go to increase the potential energy. A small expenditure of energy will displace the body only a small distance. These properties of a body in a position of minimum potential energy fully accord with the concept of stable equilibrium.

Take the case of gravity near the earth's surface. The potential energy is mgz, where z is the height above the surface. The curves depicting the function $u(x)$ can be visualized as curves indicating the altitude z of a body as a function of the horizontal x coordinate. We have to imagine a body in motion along a curve like a bead on a stiff wire. The u versus x curve corresponds to the shape of the wire if the plane of the drawing is vertical. Then it is clear that the maximum of $u(x)$ corresponds to (see Figure 9.3.2*a*) a point on the wire from which the bead slides downward at the slightest touch, moving to the right or to the left away from the maximum, and the minimum of $u(x)$ (Figure 9.3.2*b*) corresponds to the lowest point, at which the bead is in a stable position, and any other beads on the wire would strive to take up that position.

Thus, the graph of $u(x)$ gives a pictorial visualization of the direction of forces and character of equilibrium.

Let us examine a few examples.

Example 1. Let a charged body be in motion along a straight line (which we take for the x axis) on which two identical charges e_1 are fixed symmetrically about the origin at a separation of $2a$ (Figure 9.3.3).

It is quite clear that at the origin the body is in a state of equilibrium.

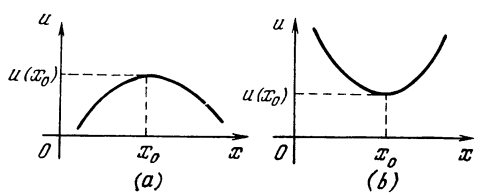


Figure 9.3.2

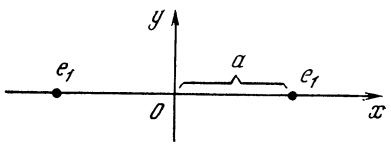


Figure 9.3.3

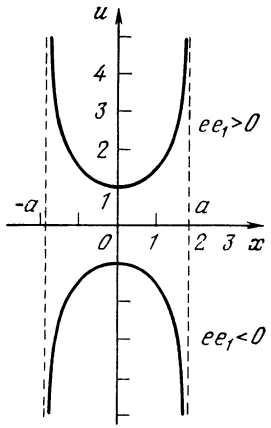


Figure 9.3.4

Indeed, the forces acting on the body from the fixed charges are equal in magnitude and opposite in direction so that they balance, which means their resultant is zero.

The potential energy $u(x)$ of a body with a charge e is

$$u(x) = \frac{e_1 e}{r'} + \frac{e_1 e}{r''},$$

where r' is the distance to the left-hand charge, and r'' the distance to the right-hand charge, that is, $r' = x + a$ and $r'' = a - x$. Whence

$$u(x) = e_1 e \left(\frac{1}{a+x} + \frac{1}{a-x}\right). \qquad (9.3.1)$$

The appropriate curves are shown in Figure 9.3.4. The upper curve corresponds to $e_1 e > 0$, which is the case of like charges of body and fixed charges, while the lower curve corresponds to $e_1 e < 0$, which means that the body has a charge opposite to the fixed charges.

In the case $e_1 e < 0$, equilibrium at the origin is *unstable*. Indeed, the body is attracted both by the left and the right charge and at the origin the forces of attraction balance. But if the body is displaced the slightest bit in any direction, say to the right, the attraction on the right will exert a stronger effect and will continue to pull it rightward. Similar reasoning shows

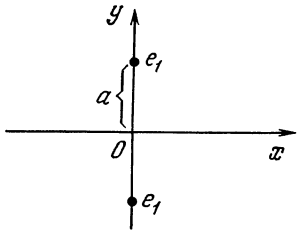


Figure 9.3.5

that in the case $e_1 e > 0$, the equilibrium is *stable*. Here, if the body is displaced rightward, the (greater) repulsion on the right will tend to move it into the initial position.

Let us find $(d^2u/dx^2)_{x=0}$. Using (9.3.1), we get

$$\frac{d^2u}{dx^2} = 2e_1 e\left[\frac{1}{(a+x)^3} + \frac{1}{(a-x)^3}\right]. \qquad (9.3.2)$$

Putting $x = 0$ in (9.3.2), we find that

$$\left.\frac{d^2u}{dx^2}\right|_{x=0} = \frac{4e_1 e}{a^3}.$$

Hence, if $e_1 e$ is positive, then $d^2u/dx^2 > 0$ at $x = 0$, the function $u(x)$ has a *minimum* at $x = 0$, and the equilibrium is *stable*. But if $e_1 e$ is negative, $d^2u/dx^2 < 0$ at $x = 0$, the function $u(x)$ has a *maximum* at $x = 0$, and the equilibrium is *unstable*.

Example 2. We consider a situation in which the charges are spaced in the same way from the origin, but along a straight line perpendicular to the line (the x axis) along which the charged body is in motion (Figure 9.3.5). In this the potential energy is

$$u(x) = 2\frac{e_1 e}{\sqrt{a^2+x^2}}.$$

The graph of the function $u(x)$ at $a = 1$ and $|e_1 e| = 1$ is shown in Figure 9.3.6. Here equilibrium at the origin is *unstable* for $e_1 e > 0$. If the charge e of the body is opposite to the fixed charges e_1 ($e_1 e < 0$), the equilibrium is *stable*.

This is easy to establish if we examine the force acting on a moving charge (Figure 9.3.7). Let $e_1 e$ be positive. Displace the body rightward from the

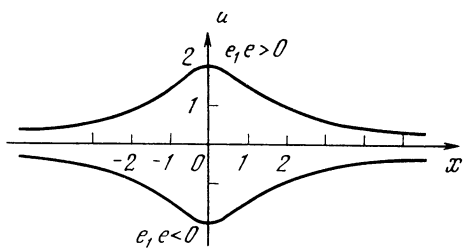


Figure 9.3.6

position of equilibrium. The resultant force of repulsion is also directed to the right, further *increasing* the deviation. The equilibrium is *unstable*. If e_1e is negative, the resultant force is in the direction of *decreasing* deviation and the equilibrium is *stable*.

These results are also readily arrived at by considering d^2u/dx^2 at $x = 0$ (do this yourself).

Note that for $e_1e > 0$, when stability occurred in Example 1 (Figure 9.3.3), in Example 2 (Figure 9.3.5) we had unstable equilibrium. For $e_1e < 0$ (unlike charges) the situation was reversed: the equilibrium is unstable for the arrangement of charges as given in Figure 9.3.5 and is stable for their arrangement shown in Figure 9.3.5.

Turning Figure 9.3.5 through 90°, we note that actually it refers to the same initial distribution of charges in the equilibrium position as in Figure 9.3.3. We can say that Figures 9.3.3 and 9.3.5 refer to the same initial distribution of charges, but the directions of motion under consideration differ. Then the equilibrium will *always* (for any signs of charges) be *unstable* in one direction or in the other.

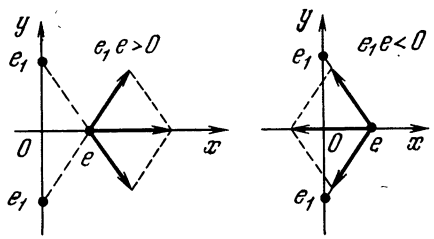


Figure 9.3.7

In electrostatics it is proved that this result is general: there is no point of equilibrium in the space between external fixed charges such that equilibrium is stable relative to displacements in *any* direction.

The general proof of this fact given below may appear too hard for the reader and he can skip it without any loss of continuity in the book.

For proof in the general form, note that the potential energy of a charge e at point (x, y, z) as a function of the charge's distance r from a fixed charge e_1 located at point (x_1, y_1, z_1) is

$$u = \frac{e_1e}{r} = \frac{e_1e}{[(x-x_1)^2+(y-y_1)^2+(z-z_1)^2]^{1/2}}.$$

Consider the motion along the x axis and find d^2u/dx^2 for y and z constant (this quantity is known as the second partial derivative of u with respect to x and is denoted by $\partial^2u/\partial x^2$). Then in a similar manner we find $\partial^2u/\partial y^2$ and $\partial^2u/\partial z^2$, which refer to motion along the y and x axes, respectively. It turns out [9.4] that for arbitrary x, y, z, x_1, y_1, and z_1 *the sum of the second derivatives along the three perpendicular axes is zero*:[9.5]

$$\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}+\frac{\partial^2u}{\partial z^2}=0. \qquad (9.3.3)$$

Obviously, this property will be preserved for the sum of any number of terms of the form e_ke/r_k, where e_k is the fixed charge at point (x_k, y_k, z_k), and r_k is the distance of charge e from this point. Consequently, formula (9.3.3) is valid for any distribution of fixed charges in space.

[9.4] The reader should convince himself of this. Note that the point (x_1, y_1, z_1) at which the fixed charge is positioned is special in that the function and its derivatives become infinite; we will ignore this point in our discussion. Various ways in which such points can be considered are discussed in Chapter 16.

[9.5] Functions that satisfy this condition are known as ***harmonic functions***; they are widely used in many fields of physics (see Chapters 15 and 17).

In particular, this formula is valid at the point at which charge e is in *equilibrium*. But for equilibrium it is necessary that the sum of the projections of the forces on each of the axes at this point be equal to zero. For this we must have

$$\frac{\partial u}{\partial x}=0, \quad \frac{\partial u}{\partial y}=0, \quad \frac{\partial u}{\partial z}=0,$$

since if the projections of a force on three perpendicular axes are zero, so is the force (a vector quantity), that is, the projection of the force on any direction is zero. [9.6]

For equilibrium to be stable relative to motion along all three perpendicular axes, it must be true that

$$\frac{\partial^2 u}{\partial x^2}>0, \quad \frac{\partial^2 u}{\partial y^2}>0, \quad \frac{\partial^2 u}{\partial z^2}>0. \tag{9.3.4}$$

But this contradicts (9.3.3) since the sum of three positive quantities cannot be zero.

Exercises

9.3.1. A charge e moves along a straight line on which are fixed two positive charges e_1 and $e_2 = 4e_1$ at a separation of $2a$. Find the point on the straight line at which equilibrium of the charge e is possible and determine the type of equilibrium. Consider two cases, $e > 0$ and $e < 0$.

9.3.2. Solve exercise 9.3.1 when the sign of e_2 is opposite to that of e_1.

9.4 Newton's Second Law

Newton's second law states that *the product of the mass by the acceleration is equal to the force applied.*[9.7] Acceleration a is the derivative of velocity v with respect to time; in turn, velocity is the derivative of the coordinate of the body with respect to time. Thus

$$ma = m\frac{dv}{dt} = F, \tag{9.4.1}$$

or

$$m\frac{d^2x}{dt^2} = F. \tag{9.4.2}$$

We begin with the case where the force is given as a function of time, $F = F(t)$. This means that the derivative d^2x/dt^2 is given as a function of time. Using Newton's law (9.4.1), we can easily find the velocity at any given instant of time. Besides the applied force we also have to specify the velocity at some time t_0. Then

$$v(t) = v(t_0) + \frac{1}{m}\int_{t_0}^{t} F(t)\,dt. \tag{9.4.3}$$

Knowing the velocity as a function of time, $v = v(t)$, and the initial position $x(t_0)$ of a body, we can find the position of the body at any given moment in time:

$$x(t) = x(t_0) + \int_{t_0}^{t} v(t)\,dt, \tag{9.4.4}$$

where $v(t)$ is given by (9.4.3).

The relationship between velocity and distance is considered in detail together with examples in Chapter 2.

On the whole, formulas (9.4.3) and (9.4.4) solve the problem of finding $x(t)$ from Eq. (9.4.2), which is a *second-order ordinary differential equation*, involving the second derivative of the unknown function $x(t)$. The answer includes not only the given function $F(t)$ but also two constants defined from the initial conditions, the position and the velocity of the body at a given time t_0.

If the law of motion of the body is given or has been experimentally established, that is, we know the function $x(t)$, it is easy to find the force applied to the body: to do this we must find the second derivative of the function $x(t)$ and multiply it by m (formula (9.4.2)).

[9.6] If we have a nonzero force F (a vector) acting in some direction, then there will be a force acting along each axis equal to the projection of the force F on the axis.

[9.7] ***Newton's first law, the law of inertia,*** states that *any body on which no forces act moves translationally and uniformly.* This means that the acceleration is equal to zero for a force equal to zero. Thus, the first law is contained in the second as a particular case.

Exercises

9.4.1. Find the law of motion of a body acted upon by a constant force F if at time $t = 0$ the body is at rest at the origin $x = 0$.

9.4.2. The same provided that (a) $x = 0$ and $v = v_0$ at $t = 0$, and (b) $x = x_0$ and $v = v_0$ at $t = 0$.

9.4.3. A body of mass 20 kg begins to move under a force of 1 N from the origin without an initial velocity. What distance will it cover in ten seconds?

9.4.4. A ball falls from a height of 100 meters (initial velocity zero). How long will it take the ball to reach the ground? (Disregard air resistance.)

9.4.5. Under the conditions of the preceding problem, the ball starts falling at a velocity $v_0 = 10$ m/s. Examine two cases: (a) the initial velocity of the ball is directed downward and (b) the initial velocity is directed upward. Determine the time required to reach the ground and the velocity the ball will have at the time of impact. Verify that in cases (a) and (b) the velocity of impact is the same.

9.4.6. A body is acted upon by a force proportional to the time that elapses from the beginning of motion (the constant of proportionality is equal to k). Find the law of motion of the body if it is known that the body begins moving from point $x = 0$ at an initial velocity v_0.

9.4.7. A body is acted upon by a force periodically varying in time, $F = f \cos \omega t$, where f and ω are constants.

(a) Find the law of motion of the body provided that $x = 0$ and $v = 0$ at $t = 0$. Establish that this is *oscillatory* motion. Determine the period of oscillation, the maximum value of $x(t)$, and the greatest value of the velocity.

(b) The same for a force $F = f \sin \omega t$ and $x = 0$ and $v = 0$ at $t = 0$.

9.4.8. A body is in motion under a constant force F. At time $t = t_0$ the body is at point $x = x_0$. Find the velocity the body must have at $t = t_0$ so that at $t = t_1$ the body will reach point $x = x_1$.

9.5 Impulse

The problem of finding the law of motion of a body for a given dependence of the force on the time was, in principle, solved in the preceding section. Here we will examine the properties of the solution and certain new concepts associated with the solution.

The product $P = mv$ of mass by velocity is called the ***quantity of motion***, or ***momentum***, while the quantity

$$I(t_0, t) = \int_{t_0}^{t} F(t)\, dt \qquad (9.5.1)$$

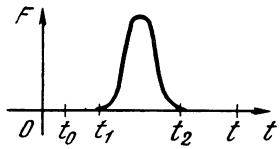


Figure 9.5.1

is known as the ***impulse of the force*** during the time interval between t_0 and t. Formula (9.4.3) may be written as follows:

$$P(t) - P(t_0) = \int_{t_0}^{t} F\, dt \; (= I(t_0, t)). \qquad (9.5.2)$$

In other words, impulse equals change in momentum.

There are forces that act during very *brief* time intervals (an instance is the blow of a hammer and the rebound after striking a body). Both prior and following the blow, the force is equal to zero. It is clear that in the absence of other forces (other than the brief blow) the body prior to the blow moves with a constant velocity and after the blow with another, also constant, velocity.

Let $F(t)$ differ from zero only during the interval between t_1 and t_2 (Figure 9.5.1). We consider the integral

$$I = \int_{t_1}^{t_2} F(t)\, dt. \qquad (9.5.3)$$

It may be called the *total* impulse in the sense that the integral is taken over the entire interval of time during which the force acts.

The expression (9.5.1) involves an integral from t_0 to t. If $t_0 < t_1$ and $t > t_2$, then

$$I(t_0, t) = \int_{t_0}^{t} F\, dt = I.$$

Indeed, we write

$$\int_{t_0}^{t} F\, dt = \int_{t_0}^{t_1} F\, dt + \int_{t_1}^{t_2} F\, dt + \int_{t_2}^{t} F\, dt.$$

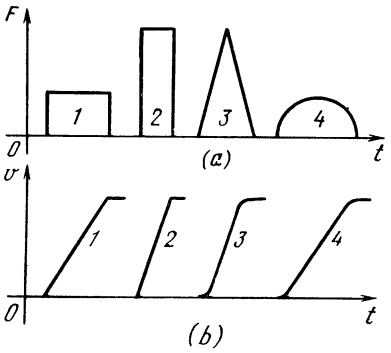


Figure 9.5.2

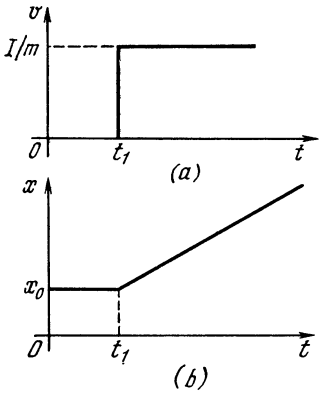


Figure 9.5.3

The first and third integrals on the right-hand side are equal to zero, since $F = 0$ over the respective intervals, and the second (middle) integral is I. Thus, from (9.5.2) and (9.5.3) we get $P(t) = P(t_0) + I$ if $t_0 < t_1$ and $t > t_2$.

From formula (9.4.3) we see that the velocity following the blow depends solely on the *impulse* of the force, that is, on the integral of the force, and not on the particular type of the function $F = F(t)$. For example, several different curves of $F(t)$ shown in Figure 9.5.2*a* all yield the same impulse, which is to say, they all change the velocity of a body by the same amount. It is not difficult to draw the appropriate graph of velocity, $v = v(t)$, for each of these curves representing $F(t)$. Figure 9.5.2*b* depicts these graphs under the general assumption that the initial velocity is equal to zero. The common element of all the curves in Figure 9.5.2*b* is the finite value of velocity: all the curves go into a horizontal straight line on the right at a height $v = I/m$.

Each of the curves representing $F(t)$ in Figure 9.5.2*a* may be compressed along the axis of time and proportionately stretched along the axis of force. The area under the curve of F, that is, $\int F\,dt$, the total impulse, does not change in the process. That is precisely how, say, curve *2* in Figure 9.5.2*a* was obtained from curve *1*.

The shorter the time of action of a force, the shorter the time interval during which the velocity of a body changes from the initial value $v_0 = 0$ to the final value $v = I/m$ (Figure 9.5.2*b*). Thus, in the limit of an extremely great force acting over an extremely short time interval, the graph of velocity takes the shape of a step (compare Figure 9.5.3*a* with curve *2* in Figure 9.5.2*b*). It is not essential here which of the curves in Figure 9.5.2*a* we compressed—the step is characterized by only one quantity, $v = I/m$, and this quantity is the same for all the curves.

If prior to the application of the force the body was at rest at point x_0, after a brief application of a great force the body begins to move with a constant velocity equal to I/m. If the force acted at time t_1 (we consider the interval between t_1 and t_2 to be small and therefore do not distinguish between t_2 and t_1) the position of the body as a function of time is given by the formulas

$$\begin{aligned} &x = x_0 \quad \text{if} \quad t < t_1, \\ &x = x_0 + \frac{I}{m}(t - t_1) \quad \text{if} \quad t > t_1. \end{aligned} \tag{9.5.4}$$

The appropriate graph is shown in Figure 9.5.3*b*. Note that $x = x(t)$ satisfies the equation

$$m\frac{d^2x}{dt^2} = F(t).$$

We recall that on the graph of $x(t)$ the first derivative dx/dt is connected with the *slope* of the tangent to the curve, while the second derivative d^2x/dt^2 describes the rate of change of the first derivative, that is, the second derivative is associated with the *curvature* of the curve $x = x(t)$ (see Section 7.10).

In Figure 9.5.3*b* the curve $x = x(t)$ has a *salient* point at $t = t_1$, $x = x_0$. The salient point can be regarded as a point at which the curvature is *infinite*, so that the existence of such a point corresponds to consideration of a very large force (infinite in the limit). However, both before and after the salient point the derivative dx/dt is finite. This means that a very large force acted over a very short time interval, so that the impulse is finite. [9.8] The impulse can readily be found from the graph (see Figure 9.5.3*b*) by computing the velocity after the application of the force and employing formula (9.5.2).

The law that we have found concerning the motion of a body that up to time $t = \tau$ was at rest and at that moment received an impulse I will help us to refine the formulas (9.4.3) and (9.4.4). For this we need a special case of the formula (9.5.4) where the body is at the origin $x_0 = 0$ at $t = \tau$. We introduce the notation

$$x_1(t, \tau) = \begin{cases} 0 & \text{if } t < \tau, \\ \dfrac{I}{m}(t-\tau) & \text{if } t > \tau. \end{cases} \tag{9.5.5}$$

If we substitute $v(t)$ from (9.4.3) into (9.4.4) and use more accurate designations (so that the upper limit and the variable of integration have different letters), we get an expression that at first glance is rather unwieldy:

$$x(t) = x(t_0) + (t-t_0)\,v(t_0) + \frac{1}{m}\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} F(t_2)\,dt_2. \tag{9.5.6}$$

[9.8] For greater details involving the various mathematical constructions associated with this example see Chapter 16.

The third term on the right-hand side can be transformed by the formal rules for handling iterated (sometimes called multiple) integrals. However, since we have not mentioned such rules anywhere, we get the transformed expression (in the form of a single integral) by using the law (9.5.5) of the motion of a body under the action of a single impulse of force.

The action of force $F(t)$ during the time interval $\Delta\tau$ from τ to $\tau + \Delta\tau$ can be approximately replaced by the (instant) action of the impulse $\Delta I = F(\tau)\,\Delta\tau$. We already know the motion of a body under the action of such an impulse—see formula (9.5.5) in which I is to be replaced by $\Delta I(\tau)$.

It then only remains to combine the contributions of all intervals Δt_i to coordinate $x(t)$ to get

$$x(t) \simeq \sum x_1(t, \tau) = \sum \frac{1}{m}(t-\tau)\,F(\tau)\,\Delta\tau \simeq \frac{1}{m}\int_{t_0}^{t} F(\tau)\,(t-\tau)\,\tau\,d\tau. \tag{9.5.7}$$

Here, as usual, we replaced the sum of a large number of terms corresponding to small intervals $\Delta\tau$ by the integral (the right- and left-hand sides of this formula can now be connected by a strict equality, ignoring the intermediate terms). Formula (9.5.7) does not take into account the initial coordinate $x(t_0)$ and the motion with the initial velocity, $(t - t_0)\,v(t_0)$, since in (9.5.7) we assumed that such motion is absent ($v(t_0) = 0$ and $x(t_0) = 0$). A more general expression can be obtained if we add these terms to the right-hand side of (9.5.7):

$$x(t) = x(t_0) + (t-t_0)\,v(t_0) + \frac{1}{m}\int_{t_0}^{t} F(\tau)\,(t-\tau)\,d\tau. \tag{9.5.8}$$

The advantage of formula (9.5.8) over (9.5.6) is that in (9.5.8) we have to integrate only once. We did not state

why we can merely add the terms corresponding to individual impulses (involving the initial velocity and the initial coordinate). This is examined in more detail in Section 17.4. Here it suffices for us that we can directly verify the values of $x(t_0)$, $(dx/dt)_{t=0}$, and d^2x/dt^2 using formula (9.5.8). To this end we must differentiate $x(t)$ in (9.5.8) in a manner similar to what was done in the verification of formula (8.3.5).

The reader will recall that by Newton's third law, in the interaction of two bodies the force with which the second body acts on the first, F_1, is equal in magnitude and opposite in direction to the force with which the first body acts on the second, F_2:[9.9]

$$F_2(t) = -F_1(t).$$

As applied to the first body and force F_1, formula (9.5.2) yields

$$P_1(t) - P_1(t_0) = \int_{t_0}^{t} F_1\,dt. \qquad (9.5.9)$$

The same formula applied to the second body and force F_2 yields

$$P_2(t) - P_2(t_0) = \int_{t_0}^{t} F_2\,dt. \qquad (9.5.10)$$

Since $F_2 = -F_1$ by Newton's third law, it follows that

$$\int_{t_0}^{t} F_2\,dt = -\int_{t_0}^{t} F_1\,dt.$$

That is why (9.5.10) takes the form

$$P_2(t) - P_2(t_0) = -\int_{t_0}^{t} F_1\,dt. \qquad (9.5.11)$$

Comparing (9.5.9) and (9.5.11), we find that

$$P_1(t) - P_1(t_0) = P_2(t_0) - P_2(t),$$

whence

$$P_1(t) + P_2(t) = P_1(t_0) + P_2(t_0).$$

[9.9] The subscripton F indicates the body acted upon by force F; the subscript on P also denotes the number of the body to which the momentum refers.

This formula shows that the *action of one body on the other does not change the sum of the momenta of the bodies.*

9.6 Kinetic Energy

Let us consider a body moving under the action of a known force $F(t)$ and find the relationship between the work done by the force and the velocity of the body.

Multiplying both sides of the basic equation $m\,(dv/dt) = F(t)$ (Newton's second law) by velocity v, we obtain

$$mv\frac{dv}{dt} = F(t)\,v. \qquad (9.6.1)$$

But according to the rule for calculating the derivatives of composite functions (see Section 4.3), the identity

$$v\frac{dv}{dt} = \frac{d}{dt}\left(\frac{v^2}{2}\right)$$

is valid no matter what the function $v(t)$. Using this fact, we can rewrite (9.6.1) as

$$m\frac{d}{dt}\left(\frac{v^2}{2}\right) = F(t)\,v,$$

or, since m is constant,

$$\frac{d}{dt}\left(\frac{mv^2}{2}\right) = F(t)\,v.$$

Introducing the notation

$$\frac{mv^2}{2} = K, \qquad (9.6.2)$$

we finally obtain

$$\frac{dK}{dt} = F(t)\,v. \qquad (9.6.3)$$

Recalling the expression (9.1.1) for work, we can write

$$A = \int_{t_0}^{t_1} F(t)\,v\,dt = \int_{t_0}^{t_1} \frac{dK}{dt}\,dt,$$

whence

$$A = K(t_1) - K(t_0). \qquad (9.6.4)$$

The quantity K is the ***kinetic energy*** of the body. Formula (9.6.4) expresses the ***conservation of energy:*** *the change in the kinetic energy of a body is equal to*

the work done by a force. Formula (9.6.3) expresses the law that *the rate of change in kinetic energy is equal to the power developed by a force.*

When the force is given by a definite function of time, the impulse and, hence, the change in momentum caused by the given force are dependent neither on the mass of the body nor on its initial velocity, since the impulse and the change in momentum are $\int_{t_0}^{t_1} F\,dt$. On the contrary, the work done by a force and the change in the kinetic energy of a body under the action of this force are essentially dependent, as may be seen from (9.6.2)-(9.6.4), not only on the force itself but also on the mass of the body and the initial velocity. Indeed, by acting with a given force over a specified time interval on a heavy body at rest at the start of motion we impart only a small velocity that will result in only a small displacement, and the work done by the force will likewise be small. A light body will take up appreciable work and will acquire a large energy. If prior to the action of the force the body was in motion in the opposite direction to the force, the force can reduce the energy of the body.

9.7 Inertial and Noninertial Reference Frames

Picture a body participating in two motions at once, say, a man walking in the cabin of a ship in motion, or a ball dropped in the cabin. Suppose that one of these motions (that of the ship, in our case) is uniform. The question then arises whether it is possible, by observing the ball falling in the cabin or the motion of some other body under the action of an applied force, to establish whether the ship is moving or not. To put it differently, does the uniform motion of the ship influence the character of motion of objects on the ship? No, it does not affect such motion in any way. Experiments have demonstrated that the absence of any influence of uniform motion on physical phenomena holds true not only for mechanics but also for the propagation of light and electric and magnetic phenomena. From this fact Einstein drew conclusions of tremendous importance in developing his theory of relativity (we do not explain the theory of relativity in this book). Newton, in formulating his laws of motion, assumed that there is ***absolute time***, which flows in the same way in the entire Universe, and ***absolute space***, in which all processes in the world take place. In the foregoing we tacitly assumed all these concepts to be valid. One can imagine, for instance, that somewhere in space there is an origin of coordinates O, given, say, by the point at which three metal rods placed at right angles to each other and having markings on them (or three axes of coordinates, as is commonly said) intersect. The same point has a watch attached to it. Then everything is simple: the coordinates of a body are defined as the projections of the body's position on the metal rods (the axes of coordinates); the body is at rest if its coordinates remain constant.[9.10] In the general case, knowing the dependence of the x, y, and z coordinates on time t, one can find its velocity and acceleration (more precisely, the three components dx/dt, dy/dt, and dz/dt of its velocity and the three components d^2x/dt^2, d^2y/dt^2, and d^2z/dt^2 of its acceleration). Experience (in this case astronomical observations) has shown that Newton's laws are valid when the origin of coordinates lies at the center of gravity of the solar system, one axis points at the North Star, the second points at an appropriate star in the equatorial plane, and the third is perpendicular to the first two and points at an-

[9.10] For the sake of simplicity we always assume that the body is very small (that is, we are actually talking of a material particle). The theory that studies the motion of finite bodies, that is, bodies that can deform and rotate, is too complicated for this book. If we speak of the motion of a finite body, we always mean the position and motion of its *center of gravity.*

other star (in other words, this star lies on the axis that is perpendicular to the first two axes). The sun's mass is 2×10^{30} kg, Jupiter's mass is 1.9×10^{27} kg, and the distance between Jupiter and the sun is 777.8×10^9 m; the mass of the rest of the planets is much less than that of Jupiter. All this positions the center of gravity of the solar system at a distance of about 0.78×10^9 m from the side of the sun facing Jupiter. The sun's radius is 7×10^8 m, so that the center of gravity of the solar system and, hence, the origin of coordinates used in astronomical calculations lies outside the sun's surface.

A system of coordinates can also be constructed using, say, a historical landmark. We could place the origin of coordinates at Trafalgar Square in London, sending one axis "up" Nelson's Column, the second horizontally to the North Pole, and the third along the latitude line in the west-east direction that passes through the chosen origin. Newton's laws operate in such a system of coordinates with less accuracy than in the above-noted astronomical system; the reasons for this will be discussed later.

Even without conducting any special experiments, Newton's laws themselves imply that the system in which these laws operate is not unique. Indeed, Newton's second law is

$$m \frac{d^2x}{dt^2} = F_x. \qquad (9.7.1)$$

If to x we add a linear function of time, the acceleration does not change, since if $x_1(t) = x(t) + a + bt$, we have

$$\frac{dx_1}{dt} = \frac{dx}{dt} + b, \quad \frac{d^2x_1}{dt^2} = \frac{d^2x}{dt^2} = F_x.$$

But if in the old system of coordinates the position of the body was characterized by the function $x = x(t)$, what system should we take so that the position is characterized by $x_1(t) = x(t) + a + bt$? If at time $t = 0$ we shift the origin to point O_1 such that $x(O_1, t = 0) = -a$, then in the new system of coordinates at $t = 0$ we will have $x_1 = x + a$. If, in addition, the new origin *is moving* leftward with a velocity (or speed) $-b$, then at time $t \neq 0$ its position is $x(O_1) = -a - bt$. The position of the body reckoned from this origin is

$$x_1 = x(t) + a + bt, \qquad (9.7.2)$$

which is just the right value for Newton's second law to be valid (see above). The new system is moving, therefore, with a *constant* velocity $-b$ with respect to the old system, that is, it moves by inertia in the same manner as a body on which *no forces* act. For this reason it is said that Newton's laws are valid in ***inertial systems of coordinates***, that is, in systems that move by inertia with respect to one another without any forces acting on them.

Even before Newton's time the principle by which all inertial reference frames are equivalent was formulated by Galileo, who wrote that in the cabin of a ship sailing steadily in quiet water all phenomena occur in the same manner as if the ship were at rest. Thus, among the great variety of systems of coordinates, which can be associated with a star, planet, or comet, there emerges a narrower class of *inertial* systems, or systems in which Newton's laws hold.

In a system of coordinates *rotating* with respect to an inertial system, on the contrary, there appear additional terms in the equations of motion (centrifugal and Coriolis's forces), which cannot be considered in a book of this scope. Rotation of the earth introduces certain correction terms into the equations written in terms of coordinates associated with the earth. Fortunately, if we set the axes by far-off stars, we introduce, with a great accuracy, a practically nonrotating inertial system of coordinates (which is simply one of the infinitude of inertial reference frames). There was a time when some scientists concluded from this fact that the very property of inertia

depends on the presence of far-off stars; now, however, this view is considered archaic. One should not discuss the obvious property of inertia; rather, one should discuss the properties (precisely, the positions) of the stars. The stars closest to us are separated from the sun by distances of about 10^{16} to 10^{17} m and move (in relation to the sun) with speeds ranging from 10 km/s to 50 km/s. Although these speeds are not low, the angular displacements of the stars are infinitesimal; for this reason all systems of coordinates associated with these stars can be considered inertial and in which the sun is at rest.

But can we do without inertial systems of coordinates? One interesting case is a system in which the orientation of the axes does not change, the axes do not rotate with the passage of time, but the origin moves along one of the axes of an inertial system of coordinates with constant *acceleration*.

Thus, we take an inertial system of coordinates (sometimes called a reference frame if a clock is attached to the origin) and specify the law by which the origin of a new system of coordinates, $O_1(x)$, moves in relation to the old origin (the motion is along the x axis): $x(t) = -f(t)$. Then $x_1(t) = x(t) + f(t)$ and, hence,

$$\frac{d^2x_1}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2f}{dt^2} = -\frac{1}{m}F_x + \frac{d^2f}{dt^2}.$$

Measuring $x_1(t)$ and applying Newton's second law to this quantity, the observer will conclude that there is a force $m(d^2f/dt^2)$ acting on the body, in addition to the force F_x. If the observer compares the behavior of various bodies in the new system, he or she will arrive at the conclusion that this additional force is proportional to a body's mass, since only in this case the additional acceleration d^2f/dt^2 (which occurs in the system) will not depend on the mass of a body.

The force we have just studied resembles the force of gravity. In other words, the phenomena which take place in a noninertial system of coordinates moving with a certain acceleration are similar to those that occur under the force of gravity. This extremely simple line of reasoning led Einstein to his basic assumption when he constructed the modern theory of gravity, or the ***general theory of relativity.***

Let us examine the famous thought experiment involving a very special noninertial system of coordinates, or noninertial reference frame. Consider an observer in an elevator. As long as the elevator is at rest, the observer feels only the force of gravity. For instance, an observer standing still presses on the floor with a force Mg, where g is the acceleration of gravity, and M is the observer's mass.

Now suppose that at time $t = 0$ the elevator is not at the ground floor and is cut free of the cable. The elevator is therefore in free fall:

$$\frac{d^2Z}{dt^2} = -g, \quad Z = z_0 - \frac{gt^2}{2},$$

where Z is the altitude at which the elevator is at time t above the earth's level. The observer inside the elevator reckons his position z_1 from the elevator's floor. The position of the observer with respect to the earth's surface, z, is related to z_1 in the following manner:

$$z_1 = z - Z = z - z_0 + \frac{gt^2}{2}.$$

Let us write the equation for coordinate z_1 of a body (material particle) acted on by the force of gravity and other forces F (elastic forces of springs and the like). The force of gravity is $-gm$, where m is the mass of the particle. In the system of coordinates associated with the earth the equation of motion is

$$m\frac{d^2z}{dt^2} = F - gm.$$

Using this equation, we can find the equation governing the motion of the particle in the system of coordinates z_1 associated with the elevator:

$$m\frac{d^2z_1}{dt^2} = m\left(\frac{d^2z}{dt^2} + g\right) = F.$$

We see that there is no force of gravity in this system of coordinates. In a freely falling elevator the g-force is zero. This state is referred to as ***zero g***. For all bodies on which no outer forces act we have $d^2z_1/dt^2 = 0$, which yields $z_1 =$ constant, or the state of rest in relation to the elevator. (The reader will recall that from the viewpoint of an observer on the earth, the elevator and the particle inside it fall with the same velocity and the same acceleration g.)

The state of zero g is terminated only when the elevator hits the ground (even if the impact demolishes the elevator). Up to this moment there is no way in which experiments conducted inside the elevator can determine the presence of the earth's gravity.

The work on an earlier version of this book was started before man went on a space mission. Most of you have seen TV programs beamed from space stations and satellites. As you could see, the astronauts experience the state of zero g. But a spaceship is usually only some 200 to 300 km from the earth, where the force of gravity diminishes (according to Newton's law of gravitation) only by several percent:

$$\frac{\Delta g}{g} = -2\frac{\Delta R}{R} = -2\frac{300}{6400} \simeq -0.09$$

$$\simeq -9\%.$$

This leads us to the conclusion that the total weightlessness (zero g) existing in the spaceship is due mainly not to the fact that the force of gravity gets weaker as one moves away from the earth but to the fact that with the rockets of the spaceship switched off the latter is in a state of free fall, just as the falling elevator is.

We can now approach the problem of inertiality of the reference frame associated with the sun from a different angle. We do not insist that both the gravitational potential and the force of gravity acting on the sun and other planets (and, in fact, all bodies in the solar system) because of the attraction of other stars in our Galaxy and other galaxies are negligibly small. Since the density of matter is evenly distributed (on the average), both the potential and the force of gravity are expressed by integrals that assume infinite values (this is known as the ***gravitational paradox***). But the absolute value of the gravitational potential never appears in the theory, and the derivative of the potential, or the force due to the attraction of other bodies, is also completely compensated for by the free fall of the solar system regardless of whether this force is finite or infinite. This justifies a study of the solar system irrespective of the rest of the universe and removes the gravitational paradox.

9.8* The Galilean Transformations. Energy in a Moving Reference Frame

Let us go back to *inertial* reference frames and consider in greater detail the relationships between the description of one and the same phenomenon in different inertial reference frames. Suppose that an object is in motion in the coach of a train that is moving with a constant speed v_0 in a direction fixed by the rails of the track, the direction being that of the x axis. We will assume that the train is moving in the direction in which x increases. Then the coordinate x_1 of the object with respect to an observer who is at rest in relation to the track, say, an observer standing on the platform of the railroad station which the train has left, will be related to the coordinate x of the object in a reference frame fixed within the coach (the origin of the coordinate system may be fixed at the back wall of the coach, for instance) as follows:

$$x_1 = x + v_0 t + a,$$

where a depends on the choice of the origin of coordinates x_1; if this origin coincides with the station which the train has left, then a is the x_1-coordinate of the back wall of the coach at time $t = 0$ (cf. formula (9.7.2)). If, in

addition, we assume that the initial moment of the "train time" t (the initial moment may be assumed to be the time of departure of the train from the station) differs from the initial moment of the "absolute time" t_1 (not connected with any train), say, we can assume that $t_1 = 0$ coincides with 00:00 GMT at the station from which the train departed, [9.11] then we must write the additional formula $t_1 = t + b$, where b is the "absolute" time t_1 at $t = 0$ (the moment of departure). The transformations

$$x_1 = x + v_0 t + a, \quad t_1 = t + b \tag{9.8.1}$$

determine the relationship between the coordinates (x_1, t_1) and (x, t) of one and the same object in two inertial reference frames, one of which (the train) moves uniformly with respect to the other (the station) with a velocity (or speed in our one-dimensional case) v_0, and are called the ***Galilean transformations.***

Let us now assume that the object is moving within the coach (in the direction of increasing x's) with a certain speed v. Then

$$x = x_0 + vt. \tag{9.8.2}$$

We then have

$$x_1 = x + v_0 t + a = (x_0 + vt) + v_0 t + a = (x_0 + a) + (v_0 + v)\, t. \tag{9.8.3}$$

(If we substitute $t_1 - b$ for t, we can rewrite the last formula as $x_1 = [(x_0 + a) - (v + v_0)\, b] + (v + v_0)\, t_1$, but for the sake of simplicity we will not distinguish between t_1 and t.) From (9.8.3) it follows that in relation to the observer on the platform the object is moving with a speed $v_1 = v + v_0$. This speed will, of course, be different for the observer traveling in the coach. However, the accelerations with respect to the two observers will be the same:

$$a_1 = \frac{dv_1}{dt} = \frac{d}{dt}(v + v_0) = \frac{dv}{dt} + \frac{dv_0}{dt} = \frac{dv}{dt} = a,$$

since the constant term $v_0 =$ constant in the expression for the speed cannot, of course, change the acceleration. Therefore, a force acting on the object is the same for the two observers, or $F = ma_1 = ma$.

The difference in speeds before and after the force is also the same for the observer on the platform and the observer in the coach. Indeed, let the speed of the body in relation to the observer on the platform be v' prior to the force and v'' after the force, while for the observer on the platform these speeds are v_1' and v_1'', respectively. Then $v_1' = v' + v_0$ and $v_1'' = v'' + v_0$; whence

$$v_1'' - v_1' = v'' + v_0 - v' - v_0 = v'' - v'.$$

The situation is more complicated with *kinetic energy.* Not only the kinetic energy itself but even the differences of kinetic energies are *distinct* for different observers. For the observer standing on the platform,

$$K_1'' - K_1' = \frac{m(v_1'')^2}{2} - \frac{m(v_1')^2}{2} = \frac{m(v''+v_0)^2}{2} - \frac{m(v'+v_0)^2}{2} = \frac{m(v'')^2}{2} - \frac{m(v')^2}{2} + mv_0v'' - mv_0v' = K'' - K' + mv_0(v'' - v').$$

In this formula K_1'' and K_1' are the final and initial kinetic energies calculated by the observer on the platform, and K'' and K' are, respectively, the kinetic energies calculated by the observer in the coach.

The work done by a force and the power are also different for different observers, since although the force is the same, the distances and the velocities are different for the observer standing on the platform and for the observer in the coach. However, the law of equality of change in kinetic energy and work is valid for any observer, although each of these quantities taken separately differs for different observers (see the exercises to this section for examples corroborating this fact).

[9.11] It goes without saying that on a long journey the moment 00:00 GMT does not change but a new time zone may force us to set our watch back or ahead, which makes it especially appropriate to distinguish between "train time" t and "station time" t_1.

Note the following remarkable formula valid for a body moving under the action of only one given force $F(t)$:

$$A=\int_{t_0}^{t_1} F(t)\,v(t)\,dt=\frac{mv_1^2}{2}-\frac{mv_0^2}{2}$$

$$=\frac{m}{2}(v_1+v_0)(v_1-v_0)$$

$$=\frac{v_1+v_0}{2}(mv_1-mv_0)=\frac{v_1+v_0}{2}\int_{t_0}^{t_1} F(t)\,dt.$$

We see that in this case the velocity (speed) $v(t)$ may be taken out from under the integral sign and replaced by the arithmetic mean of the initial and terminal velocities.

This conclusion holds true only for the case where $v(t)$ is the velocity acquired by an object under the action of only *one* force $F(t)$. If the body is acted upon by *a number* of forces, say, F_1, F_2, and F_3, then the work performed by all these forces is equal to the product of the mean velocity by the sum of the impulses of all forces:

$$A=\frac{v_1+v_0}{2}\int_{t_0}^{t_1}(F_1+F_2+F_3)\,dt$$

$$=\frac{v_1+v_0}{2}\int_{t_0}^{t_1}F_1\,dt+\frac{v_1+v_0}{2}\int_{t_0}^{t_1}F_2\,dt$$

$$+\frac{v_1+v_0}{2}\int_{t_0}^{t_1}F_3\,dt. \qquad (9.8.4)$$

However, the work done by each of these forces (say, F_2) separately is not equal to the corresponding summand $\frac{(v_0+v_1)}{2}\int_{t_0}^{t_1}F_2dt$ in (9.8.4), since the force F_2 acting separately would impart a velocity to the object that differs from $v(t)$ (see Exercise 9.8.6 below).

Above we saw that two inertial reference frames, (x, t) and (x', t'), specified on the x axis are related through the following formulas:

$$x'=x+vt+a,\quad t'=t+b, \qquad (9.8.5)$$

where v is the velocity of the reference frame (x, t) in relation to the reference frame (x', t'), so that the velocity of (x', t') in relation to (x, t) is $-v$. The Galilean transformations (9.8.5) or (9.8.1) play in mechanics a role similar to that played in geometry by *motions* (cf. Section 1.9). The transformation from one inertial reference frame to another inertial reference frame has no effect on physical phenomena; hence, the only physically meaningful facts and quantities are those which do not change under transformations (9.8.5). For instance, the *linear* equation

$$x=vt+a \qquad (9.8.6)$$

expresses the fact of *uniform* motion of a point along the x axis (the very straight line along which we consider the motion), but the specific value of v depends on the choice of the concrete reference frame and has no physical meaning. (It is for this reason that Newton's laws contain only the *acceleration* and not the velocity of a material particle.) But if we are dealing with two uniform motions, expressed via (9.8.6) and $x=v_1t+a_1$, respectively, the difference v_1-v has a concrete physical meaning: it is the ***relative velocity*** of one motion with respect to the other motion and does not depend on the choice of (inertial) reference frame. There is another physical quantity with a definite meaning: the interval $\tau=t_2-t_1$ between two events, (x_1, t_1) and (x_2, t_2); if the two events occur simultaneously, $t_1=t_2$, we can speak of the spatial distance (or space interval) $d=x_2-x_1$ between these events (the "distance" τ between the events is measured in units of time, say, seconds, while the second "distance," d, which has a meaning only at $\tau=0$, is measured in entirely different units, say, meters).

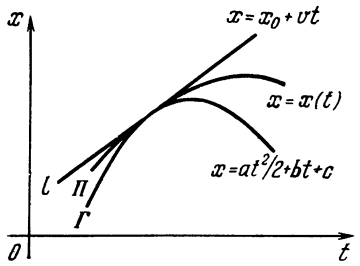


Figure 9.8.1

An arbitrary curve $x=x(t)$ in the xt-plane (Figure 9.8.1) fixes the law of motion of a (material) point along a straight line. The slope $k=x'\ (=dx/dt)$ of the tangent to the curve at point (or event) (x, t), which it would be more appropriate to denote by v, gives the *velocity* of motion at this point. If we substitute for curve Γ (equation $x=x(t)$) the tangent l at the given point (this tangent constitutes a good approximation of Γ in the neighborhood of this point), we transform the arbitrary motion into the "approximating" *uniform* motion with a velocity equal to the instantaneous velocity at the given moment in time. The second derivative $x''\ (=d^2x/dt^2)$ plays the role of the *curvature* of the curve and also expresses the *acceleration* of motion. If

$$\frac{d^2x}{dt^2}=a=\text{constant},$$

we are dealing with the law of motion

$$x = \frac{at^2}{2} + bt + c, \qquad (9.8.7)$$

with b and c arbitrary constants, which law corresponds to *uniformly accelerated* (if a is positive) or *uniformly decelerated* (if a is negative) motion and geometrically corresponds to a *parabola* (9.8.7) in the xt-plane. Replacing an arbitrary curve $x = x(t)$ by the approximating parabola Π (9.8.7) with the same values of derivatives x' $(=v)$ and x'' $(=a)$ means replacing an arbitrary motion by a *uniformly accelerated* (or decelerated) motion with the same (instantaneous) velocity and acceleration. Geometrically this replacement means that for curve $x = x(t)$ in the xt-plane we have substituted the *osculating parabola* (cf. Section 7.10). Physically, the transition from an arbitrary curve to its tangent (at a certain point, of course) means eliminating all forces, since a straight line in the xt-plane represents uniform motion, which occurs in the absence of all forces (motion by inertia), while the transition from a curve to a parabola (9.8.7) is equivalent to the assumption that all forces are constant.[9.12]

Exercises

9.8.1. Find the formula for the kinetic energy of a body moving under a constant force F (with zero velocity at the initial time) as a function of time and also as a function of the distance traveled.

9.8.2. A body is in motion under a force $F = f \cos \omega t$, and $v = 0$ at $t = 0$. Find the expression for the kinetic energy of the body, and determine the maximum of kinetic energy.

9.8.3. A body is moving in accordance with the law $x = x(t) = A \cos(\omega t + \alpha)$, with A, ω, and α constants. Determine the average kinetic energy provided that t increases without bound from $t = 0$.

9.8.4. A ball of mass m falls from a height H from a state of rest. Demonstrate that the kinetic energy of the ball, K, is equal to $mg(H - h)$, where h is the height of the ball above the ground at a given instant of time.

9.8.5. A train with a mass of 500 metric tons started out from a station, and in 3 minutes developed a speed of 45 km/h traveling 1.5 km. Determine (a) the work and the average power of the locomotive on the assumption that friction on the rails is absent, and (b) the same but having regard for friction (the coefficient of friction k is equal to 0.004; the force of friction is equal to the force of attraction of the train to the earth, or the train's weight, multiplied by k).

9.8.6. A body is under two forces: $F_1 = at$ and $F_2 = a(\theta - t)$. The impulses of these forces over the interval from 0 to θ are the same. At time $t = 0$ the body has a velocity v_0. Find the work done by each force during the time interval from 0 to θ and compare it with the product of the impulse by the average velocity.

9.8.7. A man standing still on the ground acts on a given mass m with a force during the time interval t. As a result the mass, which was originally at rest, acquires a velocity $v_1 = Ft/m$ and a kinetic energy $mv_1^2/2$ equal to the work performed by the man.

Consider the same experiment done in a train traveling at a velocity v_0. The mass m had a velocity v_0 prior to the experiment and $v_0 + v_1$ after the experiment. Find the change in the kinetic energy of the mass m. What work was done by the man? Assuming that the man rests firmly against the wall of the railway car and v_0 does not change, find the work of the force done by the train (locomotive) during the experiment.

9.8.8. A man of mass M standing in skates on ice (friction between skates and ice is neglected) acts with a force F on a mass m during time t. What kinetic energy will be imparted to the mass m? What kinetic energy will the man acquire? What is the total work done by the force acting on mass m and on the man? Why is it greater than in Exercise 9.8.7?

9.8.9. The same experiment as in Exercise 9.8.8, but the man has an initial velocity v_0 and moves together with mass m. The velocity of mass m is, after the action of the force, $v_0 + Ft/m$, and the velocity of the man is, respectively, $v_0 - Ft/M$. Find the change in kinetic energy of mass m and the man as a result of the action of the force. Find the work done by the force, which work is equal to the change in the total kinetic energy, and compare it with the result of the preceding exercise.

9.8.10. Prove that under the Galilean transformations (9.8.5) the following quantities retain their values: (a) the temporal interval τ between two events, (b) the spatial distance d between two simultaneous events, and (c) the relative velocity $v_1 - v$ of one uniform motion relative to another.

9.8.11. Verify that a transformation to a new (inertial) reference frame via the Galilean transformations (9.8.5) does not change the acceleration a of (uniformly accelerated or decelerated) motion in (9.8.7).

9.9* The Path of a Projectile. The Safety Parabola

Let us consider the problem of the flight path of a projectile (shell) fired from a gun with initial velocity v_0.

[9.12] See I. M. Yaglom, *A Simple Non-Euclidean Geometry and Its Physical Basis*, Springer, Berlin, 1969.

We take the point of ejection of the shell from the barrel of the gun for the origin and send the y axis vertically upward, while the x axis is assumed horizontal (the motion of the projectile occurs, therefore, in the xy-plane). For the sake of simplicity we disregard air resistance because this would introduce considerable complications.

By Newton's second law,

$$m\frac{dv}{dt} = F. \tag{9.9.1}$$

We applied this law earlier only for rectilinear motion. However, in the flight-path problem the direction of v varies with time (the velocity is always directed along a tangent to the path of the shell). For this reason we are forced to assume that the quantities v and F in Eq. (9.9.1) are ***vectors***, or $\mathbf{v} = (v_x, v_y)$ and $\mathbf{F} = (F_x, F_y)$.

Any motion in the xy-plane may be regarded as the result of combining two motions: one occurring along the x axis under force F_x with velocity v_x, and the other along the y axis under force F_y with velocity v_y. Applying Newton's second law to each of these motions separately yields

$$m\frac{dv_x}{dt} = F_x, \quad m\frac{dv_y}{dt} = F_y \tag{9.9.1a}$$

(see Figure 9.9.1). In each of these equations the force and the velocity are directed along a single straight line (the x axis in the first equation and the y axis in the second).

Denote by φ the angle which the barrel of the gun makes with the horizontal; call φ the ***angle of departure.*** Since we are considering the most elementary case in which the shell in flight is acted upon solely by the force of gravity directed earthwards, it follows that $F_x = 0$ and $F_y = -mg$. And so the equations in (9.9.1a) have the form

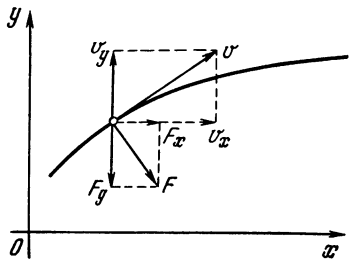


Figure 9.9.1

$$m\frac{dv_x}{dt} = 0, \quad m\frac{dv_y}{dt} = -mg. \tag{9.9.2}$$

What is remarkable (and important) here is that our equations have separated: the first contains only one unknown function v_x, and the second contains the unknown function v_y.

Let us find the initial conditions for the functions $v_x(t)$ and $v_y(t)$. At the time of emergence of the shell from the barrel, $t = 0$,

$$v_x(0) = v_0 \cos\varphi, \quad v_y(0) = v_0 \sin\varphi,$$

where v_0 ($= |\mathbf{v}_0|$) is the initial (or muzzle) velocity of the shell. The first equation in (9.9.2) yields $dv_x/dt = 0$, from which it follows that v_x is constant, and hence

$$v_x(t) = v_x(0) = v_0 \cos\varphi. \tag{9.9.3}$$

The second equation in (9.9.2) yields $dv_y/dt = -g$, whence, integrating both parts from 0 to t, we get $v_y(t) - v_y(0) = -gt$, or

$$v_y(t) = -gt + v_0 \sin\varphi. \tag{9.9.4}$$

To determine the displacements x and y along the coordinate axes, we take advantage of the fact that $\mathbf{v} = d\mathbf{r}/dt$, where $\mathbf{r} = (x, y)$ is the radius vector of the shell. This vector equation splits into two scalar equations:

$$\frac{dx}{dt} = v_x, \quad \frac{dy}{dt} = v_y, \tag{9.9.5}$$

or, in view of (9.9.3) and (9.9.4),

$$\frac{dx}{dt} = v_0 \cos\varphi, \quad \frac{dy}{dt} = -gt + v_0 \sin\varphi. \tag{9.9.6}$$

The equations are again separated: the first deals only with the unknown function $x(t)$, and second deals with the unknown function $y(t)$.

At the initial instant of time, the shell was at the origin, with the result that

$$x = 0 \text{ and } y = 0 \text{ at } t = 0. \qquad (9.9.7)$$

Integrating Eqs. (9.9.6) from 0 to t and using the initial conditions (9.9.7), we find that

$$x = v_0 t \cos \varphi, \quad y = v_0 t \sin \varphi - \frac{gt^2}{2}. \qquad (9.9.8)$$

These formulas enable us to determine the position of the shell at *any* instant of time t.

Taking various values of t, we can find the position of the shell from formulas (9.9.8) at different times and plot a graph of the path of the shell. Thus, Eqs. (9.9.8) yield a curve in the xy-plane. They determine a parametric representation of the path of the shell, with t the parameter (see Section 1.8).

From Eqs. (9.9.8) we can easily exclude t and obtain an equation of the path in the ordinary form, as a function of y in x. Indeed, the first equation in (9.9.8) yields $t = x/v_0 \cos \varphi$; then from the second we get

$$y = x \tan \varphi - x^2 \frac{g}{2v_0^2 \cos^2 \varphi}. \qquad (9.9.9)$$

From this we see that y is a second-degree polynomial in x, the graph of which is a *parabola*. Consequently, disregarding air resistance, we can say that the path of a shell is in the shape of a parabola. Figure 9.9.2 depicts the path specified by (9.9.9) for the case where $v_0 = 80$ m/s and $\varphi = 45°$.

From (9.9.9) it is evident that for one and the same v_0, the shape of the trajectory depends on the angle of departure φ. We will find the *maximum altitude of ascent* of the shell and the range of fire for given φ and v_0. To determine the coordinate x_1 of the maximum altitude of the shell, $y_{\max}$, we set up the equation $dy/dx = 0$ to get

$$\tan \varphi - x_1 \frac{g}{v_0^2 \cos^2 \varphi} = 0,$$

whence

$$x_1 = \frac{v_0^2 \sin \varphi \cos \varphi}{g} = v_0^2 \frac{\sin 2\varphi}{2g}.$$

For this value of x, the height y is at its maximum (it is physically clear that this is precisely the maximum; this can be verified easily by finding the sign of d^2y/dx^2). Substituting this value of x into (9.9.9) yields

$$y_{\max} = y(x_1) = \frac{v_0^2 \sin^2 \varphi}{2g}.$$

To determine the range of the shell, it suffices to find the value x_2 for which $y = 0$ (see Figure 9.9.2):

$$x_2 \tan \varphi - x_2^2 \frac{g}{2v_0^2 \cos^2 \varphi} = 0.$$

Disregarding the solution $x_2 = x_0 = 0$ which does not interest us, we find that

$$x_2 = \frac{v_0^2 \sin 2\varphi}{g}. \qquad (9.9.10)$$

The range of fire depends on the initial velocity and the angle of departure.

For what angle of departure (initial velocity v_0 unchanged) is the range of fire the greatest? Formula (9.9.10) shows that this happens when $\sin 2\varphi = 1$, or $\varphi = 45°$, and the range in this case is v_0^2/g. Also, since $\sin 2\varphi = \sin 2(90° - \varphi)$ and in view of (9.9.10), for angles φ_0 and $90° - \varphi_0$ (say, at angles of departure $\varphi = 30°$ and $\varphi = 60°$) the range is the same and shells hit the same target; artillerymen call fire with shells having angles of departure less than 45° *grazing*, and fire with angles of departure greater than 45° *plunging*.

Let us determine the time t_1 during which the shell rises upward. All we need to do is solve the equation $dy(t_1)/dt = 0$ because at time $t = t_1$, when y attains its maximum value, the shell will cease to rise and will begin to fall.

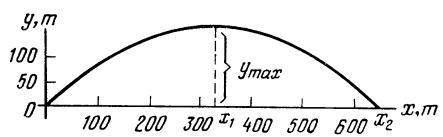


Figure 9.9.2

The condition $dy/dt = 0$ yields $v_0 \sin \varphi - gt = 0$, whence

$$t_1 = \frac{v_0 \sin \varphi}{g} . \qquad (9.9.11)$$

The total flight time t_2 can be found from the fact that the flight ceases when $x = x_2$. Combining (9.9.8) and (9.9.10), we find that

$$v_0 t_2 \cos \varphi = \frac{v_0^2 \sin 2\varphi}{g} ,$$

whence

$$t_2 = \frac{2v_0 \sin \varphi}{g} . \qquad (9.9.12)$$

Comparing (9.9.12) with (9.9.11), we see that the total flight time t_2 is twice the time of ascent t_1 for any angle of departure, or that the ascent time of a shell is equal to the descent time.

Now let us turn from the problem of firing at targets on the ground to *antiaircraft* fire. If for a given value of the shell's initial velocity (this velocity depends on the type of gun) we change the angle of departure, we obtain a whole family of trajectories (Figure 9.9.3), where the plunging-fire trajectories, that is, trajectories with $45° < \varphi < 90°$, touch a single curve

$$y = \frac{v_0^2}{2g} - x^2 \frac{g}{2v_0^2} \qquad (9.9.13)$$

(the grazing-fire trajectories lie inside the "dome" (9.9.13) and do not touch it). Indeed, solving Eqs. (9.9.9) and (9.9.13) simultaneously, we find that the curves (parabolas) have *only one* common point:

$$x = \frac{v_0^2}{g} \cot \varphi, \quad y = \frac{v_0^2}{2g} (1 - \cos^2 \varphi) \qquad (9.9.14)$$

(this point lies in the upper half-plane only for $\varphi \geqslant 45°$, in view of which only plunging-fire trajectories in (9.9.9) touch the curve specified by (9.9.13)). The fact that the appropriate parabolas only *touch* (that is, do not intersect) follows from the sole fact that the common point of (9.9.9) and (9.9.13) is unique;

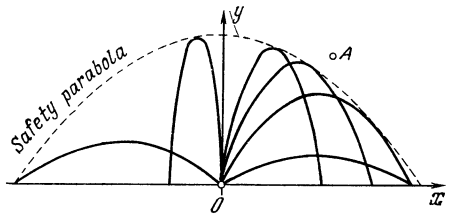


Figure 9.9.3

this can be verified by directly calculating the slope of the tangent to both curves at point (9.9.14). Parabola (9.9.13) can be called a ***safety parabola***— if a target A (say, an aircraft) is outside the "dome", there is no angle of departure at which it can be hit. The target can be hit only by using another type of gun.

Note, in conclusion, that the actual flight paths of shells are not exact parabolas. In reality the shell experiences air drag, which we ignored; the acceleration of gravity g depends on the distance from the surface of the earth; rotation of the earth also introduces certain distortions into the fall of the shell; and, finally, Newton's laws of motion of a material point cannot be applied without certain reservations to the motion of a projectile (although this factor is not so important). This, however, is not a unique situation—we encounter it whenever we apply mathematics to real-life problems; we always ignore certain factors and are content with a rough picture, which enables us to use mathematical tools we are accustomed to. The actual range of fire, the altitude of flight of a shell, the time of flight, and so on depend on the mass of the shell, the shell's shape, and the air density. For instance, when the initial velocity of a shell is low (see Figure 9.9.2 that refers to the case with $v_0 =$ 80 m/s), the role or air drag is indeed insignificant, while for a great v_0 there is no way in which the drag can be ignored; for a 305-mm caliber gun with a muzzle velocity of $v_0 =$ 800 m/s and $\varphi = 55°$, the air drag that acts on the flying shell diminishes

the flight range from 61 km (which value is given by formula (9.9.10) to 22.2 km.

Exercises

9.9.1. A shell leaves the gun with a velocity of 80 m/s. Determine the range of fire and the maximum height reached by the shell if the angle of departure $\varphi = 30°$, 45°, and 60°.

9.9.2. Determine the maximum altitude at which a shell with initial velocity $v_0 = 80$ m/s can hit a target located 500 meters from the gun.

9.10 The Motion of a Body in Outer Space

Let us now turn to the problems of the motion of objects (artificial satellites and spaceships) in outer space. For a body launched from the earth to become a satellite of the earth, that is, for it to orbit around the earth, the centrifugal force on the body must be balanced by the gravitational force of the earth. The respective velocity u_1 is commonly known as the ***satellite*** (or ***orbital***) ***velocity.*** [9.13] To find u_1, we set up the equation

$$m\frac{u_1^2}{R} = mg, \qquad (9.10.1)$$

where g is the acceleration of gravity, m the mass of the body in question, and R the orbit's radius (the left-hand side of Eq. (9.10.1) is the centrifugal force, and the right-hand side is the force of gravity of the earth). If the orbit is low, that is, the body flies close to the earth's surface, R is close to r_0, the earth's radius, whereby on the right-hand side of (9.10.1) we can put the force of gravity at the earth's surface. Formula (9.10.1) then yields

$$u_1 = \sqrt{gR} \simeq \sqrt{gr_0} \simeq 8 \text{ km/s}. \qquad (9.10.2)$$

For a satellite that is orbiting the earth at a distance R from the earth's center that exceeds r_0 considerably, we must take into account the variation of g with altitude; the value of g at the earth's surface is denoted by g_0 (it is this value that was substituted into formulas (9.10.1) and (9.10.2)). Indeed, by Newton's law of gravitation, a body placed at a distance R from the center of the earth is attracted to the earth with a force $F = GmM/R^2$, where m is the body's mass, and M the earth's mass. In addition, by Newton's second law, $F = mg$, where $g = g(R)$ is the acceleration of gravity at a distance R from the earth's center. Comparing the two expressions for F, we can write $g = GM/R^2$. If $R = r_0$, then $g = g_0$, whence $g_0 = GM/r_0^2$, which yields $G = g_0 r_0^2/M$ and, therefore,

$$g(R) = g_0 r_0^2/R^2.$$

The condition (9.10.1) for the centrifugal force being equal to the force of gravity then assumes the form

$$mu_1^2/R = mg_0 r_0^2/R^2,$$

which means that the satellite velocity u_1 is given by the following formula:

$$u_1 = (g_0 r_0^2/R)^{1/2}.$$

The greater the distance R, the lower the velocity u_1 necessary for a satellite to remain in orbit. However, this does not at all mean that it is easier to launch a satellite to a higher orbit than to a lower orbit, since to launch a satellite to a high orbit we must spend a lot of energy in overcoming the force of gravity on the trip from the earth's surface to the orbit. Of practical importance is the motion of satellites on a stationary, or 24-hour, orbit, which lies approximately at an altitude of 36 000 km above the earth's equator; a satellite launched into such an orbit will "hang" over a spot on the equator (see Exercise 9.10.1).

Let us now consider a more complicated problem. For a body to leave the bounds of earth's gravity, its initial kinetic energy must be greater than the difference between the potential energy of the body at a far-off point and that at the earth's surface. This difference was found in Section 9.2 (formula

[9.13] We are interested here in steady-state motion with a velocity u_1 (and later with velocities u_2 and u_3) and ignore the transient launching stage.

(9.2.8)). Here it is assumed that the body acquires its speed rapidly over a path that is short compared to the earth's radius, so that the variation in potential energy over this path can be ignored. In other words, we must assume that the thrust of the reactive force operating while the rockets launching the satellite are burning (see Section 9.11) is very high and the force of gravity can be ignored. [9.14]

The minimal velocity that a moving body (as a rocket) must have to escape from the gravitational field of the earth or of a celestial body and move outward into space is called the ***escape velocity***. Let us find this velocity for the earth (we denote it by u_2). By (9.2.8), the initial energy required to reach a point in space lying at a distance R from the earth's center, if prior to motion it was at a distance r_0 (on the earth's surface), is given by the formula $K_0 = mg(r_0/R) \times (R - r_0)$. In our case R is much larger than r_0, whence $R - r_0 \simeq R$, which yields $K_0 \simeq mgr_0$. We assume this to be equal to the kinetic energy of the rocket:

$$mu_2^2/2 = mgr_0.$$

This yields

$$u_2 = \sqrt{2gr_0} \simeq 11.2 \text{ km/s}. \qquad (9.10.3)$$

Combining (9.10.2) with (9.10.3), we find that

$$u_2 = \sqrt{2}u_1 \simeq 1.4u_1, \quad \text{and} \quad u_2^2 = 2u_1^2.$$

For the sun, this velocity (denoted u_3), when imparted to a body, takes the body outside the solar system. We will find it using the fact that the speed with which the earth orbits the sun is known: $v_1 \simeq 30$ km/s.

By Newton's law of gravitation, the force with which a body of mass m is attracted to the sun is $F = -kM_1m/r^2$, where M_1 is the sun's mass ($M_1 \simeq 2 \times 10^{30}$ kg), r the distance from the body to the sun's center, and k a constant. The potential energy of the body separated by a distance r from the sun's center is

$$u(r) = -\frac{kM_1}{r}m. \qquad (9.10.4)$$

Here the zero of potential energy is taken as its value at infinity (cf. Section 9.2).

The value of the potential energy of a body which orbits the sun together with the earth can easily be expressed in terms of the speed with which the earth orbits the sun. Indeed, on the earth's orbit the force with which the earth is attracted to the sun is balanced by the centrifugal force, that is,

$$M_2\frac{v_1^2}{r_1} = k\frac{M_1M_2}{r_1^2}, \qquad (9.10.5)$$

where v_1 we already know, r_1 is the radius of the earth's orbit ($r_1 \simeq 150 \times 10^6$ km $= 1.5 \times 10^{11}$ m), and M_2 is the earth's mass (this quantity is cancelled out from Eq. (9.10.5)). Whence, $kM_1 = v_1^2r_1$, and (9.10.4) assumes the form

$$u(r_1) = -v_1^2m.$$

For a body at a distance r_1 from the sun to leave the bounds of the sun's gravitation, the sum of the body's kinetic and potential energies at this distance must be nonnegative. This leads to the following inequality:

$$m\frac{v_2^2}{2} + u(r_1) = m\frac{v_2^2}{2} - mv_1^2 \geqslant 0, \qquad (9.10.6)$$

where v_2 is the velocity that the body must have to leave the solar system, and v_1 is the known orbital velocity of the earth.

We have already encountered a similar situation when we considered the motion of a body in the earth's gravitational field: the velocity necessary

[9.14] It has been found that it is more advantageous to burn the rocket fuel fast (less fuel is required) than to extend the burning process over the entire time necessary to travel a distance of the order of the earth's radius. Only in the layer where the density of the atmosphere is high, and so is the air drag, high velocities are not advantageous, but since the thickness of this layer is much smaller than the earth's radius (see Chapter 11), we will not take this layer into account.

for a body to leave this field corresponds to a kinetic energy that is *twice* the kinetic energy corresponding to the velocity that ensures that the body is an (artificial) satellite of the earth.

Inequality (9.10.6) yields the minimal velocity v_2 that a body must have to leave the solar system:

$$v_2 = \sqrt{2}v_1 \simeq 1.4v_1 \simeq 42 \text{ km/s}.$$

Thus, to leave the solar system, a body on the earth's orbit must have an initial velocity (with respect to the sun) that is at least 42 km/s. If the velocity is greater than 42 km/s, the body will leave the solar system, irrespective of the direction of this velocity, that is, irrespective of whether the body moves along the radius away from the sun (path *1* in Figure 9.10.1) or along a tangent to the earth's orbit (paths *2* and *3*) or even in the direction of the sun (but at a certain angle so as not to land on the sun's surface, path *4*). Only the shape of the trajectory depends on the direction of the initial velocity (see Figure 9.10.1).

It is clear that the most advantageous trajectory for launching a rocket from the earth's surface is *2*: the earth moves with a speed of 30 km/s, so that only 12 km/s is required to obtain 42 km/s if we launch the rocket in the direction of the earth's orbital motion. This speed v_2' the rocket must have after it has left the earth's gravitational field, that is, after it has traveled a distance large compared to the radius of the earth's orbit.

What initial velocity u_3 must the rocket have at the earth's surface for this to happen (it is this velocity that is called the escape velocity for the sun)? It can be found from the obvious relationship

$$m\frac{u_3^2}{2} = mgr_0 + m\frac{(v_2')^2}{2}. \qquad (9.10.7)$$

Here the first term on the right-hand side is the energy necessary for overcoming the attraction of the earth, and the second term is the energy that the rocket (or spaceship) of mass m must have after it has overcome the earth's gravity for its speed v_2 to be sufficient (together with the 30 km/s of the earth's motion in orbit) for the rocket to leave the solar system. Formula (9.10.7) then yields

$$u_3^2 = 2gr_0 + (v_2')^2 = u_2^2 + (v_2')^2, \qquad (9.10.8)$$

whence

$$u_3 = (u_2^2 + (v_2')^2)^{1/2} \simeq (11.2^2 + 12^2)^{1/2} \simeq 16.4 \text{ km/s}. \qquad (9.10.9)$$

Note that for the rocket to get closer to the sun or reach, say, Mars or Venus, u_2 is insufficient. Indeed, having this velocity, the rocket will leave the earth and move along the earth's orbit with the velocity of the earth, or 30 km/s.

Although the potential energy does diminish as the rocket moves toward the sun, but the centrifugal force that acts on the rocket moving along the orbit prevents the rocket from doing this. To get closer to the sun, the rocket must diminish its speed, which is as difficult as increasing the speed. For instance, to hit the sun, the rocket must be stopped, that is, it must move with a velocity of 30 km/s with respect to the earth (after it leaves the earth's gravitational field). For this the rocket must have an initial velocity

$$u_4 \simeq (30^2 + 11.2^2)^{1/2} \simeq 32 \text{ km/s} \qquad (9.10.10)$$

at the earth's surface.

We see that it is more difficult to hit the sun than to escape from its gravitational bounds. More "profitable"

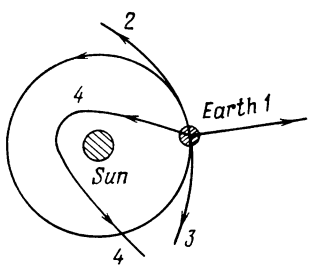


Figure 9.10.1

variants of space travel can be obtained by using the gravitational pull of other planets. However, we will not discuss such questions in this book.

Exercise

9.10.1. Find the radius of an orbit in which a satellite circles the earth every 24 hours. (If a satellite is launched into such an orbit in the equatorial plane, it will "hang" constantly over a single point on earth.)

9.11 Jet Propulsion and Tsiolkovsky's Formula

In the case of motion in airless space, the only method of flight control, which is widely used and is highly effective, that is, a method that enables changing speed and direction, consists in ejecting a portion of the mass of the flying object itself, which means applying the reaction (jet) principle.

The Russian scientist K. E. Tsiolkovsky (1857-1935) was the first to fully realize the significance of the jet principle and to investigate the fundamental regularities of ***reaction***, or ***jet propulsion***. He was also the first to suggest the possibility of conquering outer space by means of rockets. From him, via his pupils and followers—Soviet scientists and engineers—stems the scientific tradition that saw final embodiment in artificial earth satellites, space probes, and space stations with astronauts on board.

Let us derive the basic equation of the rectilinear motion of a rocket. The propellant, whether gunpowder or a mixture of fuel (alcohol or gasoline) and oxidizer (oxygen or nitric acid), possesses a definite supply Q of chemical energy per unit mass (of the order of 5×10^6 J/kg for smokeless powder and 10×10^6 J/kg for a gasoline (or petrol) and oxygen mixture[9.15]). In burning, this chemical energy is converted into the thermal energy of the products of combustion, which stream out of the nozzle, the thermal energy turning into the kinetic energy of motion.

When a reaction (rocket) engine is fixed on a test bed, the combustion products are exhausted at a definite velocity u_0. The kinetic energy they have per unit mass constitutes a definite portion of the chemical energy of the propellant:

$$\frac{1}{2} u_0^2 = \alpha Q, \qquad (9.11.1)$$

where α is a dimensionless number, the efficiency of the processes of combustion and ejection of gases.[9.16] From now on we will consider the exhaust velocity u_0 to be a given known quantity. It is roughly 2 km/s for powder and about 3 km/s for liquid propellant. It is easy to see that these quantities are associated with the values of $\alpha \simeq 0.5$ (an efficiency of the order of 50%).

Prior to combustion, the propellant was at rest. Suppose a mass dm of propellant is burnt and exits from the nozzle. In so doing, it acquires a momentum equal to $u_0 dm$. Clearly, the impulse dI of the force with which the rocket acts on this mass is equal to the momentum acquired by the mass,[9.17] or

$$dI = F\,dt = u_0\,dm.$$

By Newton's third law—every action has an equal and opposite reaction—the impulse of the force with which the mass dm of the combustion products acts on the rocket vehicle is equal to the same quantity with sign reversed. Suppose, for instance, that the exhaust velocity u_0 is in the direction of decreasing x. Then u_0 is negative, or $u_0 = -|u_0|$. For the impulse of the force acting on the rocket we have

$$dI_r = F_r dt = -u_0 dm = |u_0|\,dm. \qquad (9.11.2)$$

[9.15] The heating value of gasoline is about 50×10^6 J/kg, but burning 1 kg of gasoline requires 3.4 kg of oxygen. In a rocket launched into airless space, the oxygen has to be carried along and the energy must be referred to the sum of the masses of fuel and oxidizer.

[9.16] If in (9.11.1) Q is expressed in J/kg, then u_0 will be expressed in m/s.

[9.17] The designation dI is due to the fact that we consider a small mass dm.

The quantity

$$I' = \frac{dI}{dm} = |u_0| \qquad (9.11.3)$$

is the impulse per unit mass, what is called the ***unit impulse***. This quantity is equal to the exhaust speed of gases from a rocket at rest.

Let us check the dimensions in formula (9.11.3). The force F has the dimensions of kg·m/s², or N, and the impulse I is the product of force by time, so its dimensions are N·s (or kg·m/s). The dimensions of dI/dm are (kg·m/s)/kg = m/s, which are the dimensions of velocity. For powder gases, $u_0 = 2 \times 10^3$ m/s = 2 km/s, while for liquid fuel $u_0 = 3$ km/s.

The force acting on the rocket is, by formula (9.11.2),

$$F_r = |u_0| \frac{dm}{dt} .$$

It is proportional to the quantity of gases exhausted in unit time.

Now let us examine the derivation of the formula for the velocity of the rocket vehicle. If the rocket is itself in motion with a velocity u, then the exhaust velocity of the gases differs from u_0 and is equal to $u + u_0 = u - |u_0|$ (recall that when the vehicle is at rest, the exhaust velocity of gases is equal to $-|u_0|$). It is obvious that such quantities as the *difference* between the velocity of powder prior to combustion and the velocity of the exhausted powder gases and as the *force* with which the powder gases act on the rocket are independent of whether the rocket vehicle is in motion or at rest (for the sake of definiteness we speak of powder as the propellant).

Let us denote the initial mass of the rocket together with the powder by M_0 and the mass of exhausted powder gases by m. The quantity m is a function of time, or $m = m(t)$. The mass of the rocket with powder at time t is equal to

$$M = M(t) = M_0 - m(t). \qquad (9.11.4)$$

The equation of motion, or Newton's second law, is

$$M \frac{du}{dt} = F = |u_0| \frac{dm}{dt} ,$$

which can be rewritten as $M\,du = |u_0|\,dm$, or, using (9.11.4),

$$(M_0 - m) \frac{du}{dm} = |u_0|. \qquad (9.11.5)$$

The possibility of cancelling out dt has the physical meaning that, in the absence of other forces acting on the rocket, the velocity of the rocket depends only on the amount of exhausted powder gases (for a fixed value of u_0). By the time a given amount m of powder gases is exhausted through the nozzle, the rocket has acquired a definite velocity u, irrespective of the time during which the given amount of powder gases was released.

It is easy to solve the differential equation (9.11.5). At $t = 0$, $u = 0$ and $m = 0$. We thus have

$$u = |u_0| \int_0^{m_1} \frac{dm}{M_0 - m}$$

$$= -|u_0| \ln (M_0 - m)\Big|_0^{m_1}$$

$$= |u_0| [-\ln (M_0 - m_1) + \ln M_0]$$

$$= |u_0| \ln \frac{M_0}{M_0 - m_1} = |u_0| \ln \frac{M_0}{M} ,$$

where m_1 is the total mass of powder gases exhausted by a given time, and $M = M_0 - m_1$ is the mass of the rocket at this moment in time. We have therefore arrived at ***Tsiolkovsky's formula***

$$u = |u_0| \ln \frac{M_0}{M} . \qquad (9.11.6)$$

If we are interested in the terminal velocity u_{ter} at burnout, in formula (9.11.6) we must substitute M_{ter} (terminal mass of the rocket after all the fuel has burnt out) for M, that is, $M_{ter} = M_0 - m_{tot}$, with m_{tot} the total mass of the propellant. We get

$$u_{ter} = |u_0| \ln \frac{M_0}{M_{ter}} . \qquad (9.11.7)$$

This formula can easily be used to solve the inverse problem: what initial mass M_0 of the rocket vehicle must be taken so that to a given terminal mass M_{ter} is imparted a definite velocity u_{ter}:

$$\ln \frac{M_0}{M_{ter}} = \frac{u_{ter}}{|u_0|},$$

whence

$$M_0 = M_{ter} \exp \frac{u_{ter}}{|u_0|}. \qquad (9.11.8)$$

Everywhere in our analysis we ignored the (transient) process of "slow" combustion, when both the force of gravity and the reaction force act on the body (rocket). Of course, in real life this situation is always present—we only hoped that by separating these two forces we do not introduce a large error into the results; the forces of gravity were studied in Section 9.10, while jet propulsion has been studied in the present section. [9.18] We will now try to unite the results of these two sections by using formulas (9.11.6)-(9.11.8) to estimate the values of M_0/M_{ter} necessary for attaining the satellite velocity u_1 and the escape velocities u_2 and u_3. We will again assume that the propellant is gun powder, for which $|u_0| = 2$ km/s. Using formula (9.11.8), we get $M_0/M_{ter\,1} = e^4 \simeq 54$ for $u_1 = 8$ km/s, $M_0/M_{ter\,2} = e^{5.6} \simeq 270$ for $u_2 = 11.2$ km/s, and $M_0/M_{ter\,3} = e^{8.2} \simeq 3640$ for $u_3 = 16.4$ km/s. In the case of liquid propellant, $|u_0| = 3$ km/s, and similar calculations yield $M_0/M_{ter\,1} \simeq 14.5$, $M_0/M_{ter\,2} \simeq 42$, and $M_0/M_{ter\,3} \simeq 245$. From the foregoing we see that the magnitude of M_0/M_{ter} is strongly dependent on the exhaust velocity of the gases, u_0. To get an idea of the difficulty of the problem of launching a rocket, one should bear in mind that M_{ter} includes the mass of the fuel tanks, etc.

Let us find the efficiency of a rocket as a whole. We define this quantity as the ratio of the kinetic energy of the rocket at burnout, $M_{ter}u_{ter}^2/2$, to the chemical energy of the burnt fuel, $mQ = (M_0 - M_{ter})Q$. The efficiency is

$$\eta = \frac{M_{ter}u_{ter}^2}{2Q(M_0 - M_{ter})}. \qquad (9.11.9)$$

Substituting the expression for u_{ter} from (9.11.7) and expressing u_0^2 from (9.11.1), we finally obtain

$$\eta = \alpha \frac{M_{ter}}{M_0 - M_{ter}} \left(\ln \frac{M_0}{M_{ter}}\right)^2.$$

The efficiency proves to be the product of the "internal efficiency" α (which characterizes the completeness of burning of the fuel and the conversion of thermal energy into the kinetic energy of the gases) and a second factor that depends solely on the choice of the ratio between the mass of propellant, m, and the mass M_{ter} of the payload. We denote m/M_{ter} by z. Then $M_0 = M_{ter} + m = M_{ter}(1 + z)$, and

$$\eta = \alpha \frac{M_{ter}}{m} \left(\ln \frac{M_{ter} + m}{M_{ter}}\right)^2 = \frac{\alpha}{z} [\ln (1 + z)]^2. \qquad (9.11.10)$$

At first glance it might appear that due to the fraction $1/z$ the efficiency is very great for small z's. In reality, for small z's we have $\ln(1 + z) \simeq z$ (see Section 4.9) and so $\eta \simeq (\alpha/z)z^2 = \alpha z$ for $z \ll 1$. The efficiency is proportional to z and, hence, is small for small z's, with the result that the rocket moves slowly and almost all the energy is carried away by the gases. For very great z's, the efficiency again falls because of diminished payload mass. [9.19]

[9.18] Note that such a "coarse" approach to a real process was undertaken in Section 9.10, when we separated into two stages the launching of an object into outer space: in the first stage we assumed the acceleration of gravity g to be constant and equal to g_0, the acceleration of gravity at the earth's surface, while in the second stage the object moved in a mode in which g diminished as the distance to the earth's center increased.

[9.19] For large z's, the quantity $[\ln(1 + z)]^2$ grows more slowly than z. Indeed, denoting $y = \ln(1 + z)$, we get $z = e^y - 1$ and the function e^y grows faster than *any* power y (see Section 6.5).

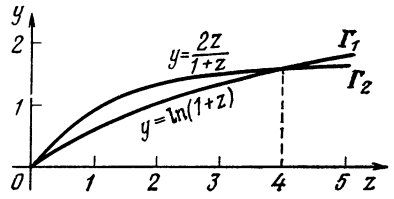


Figure 9.11.1

Since the terminal velocity of the rocket is also dependent solely on z, we can say that the efficiency of the rocket is determined by the requisite velocity. At small velocities the efficiency of the rocket vehicle is low and so it is disadvantageous to employ jet propulsion in automobiles and other cases of relatively slow motion. At high velocities, the energy efficiency of the rocket again diminishes, but the use of the rocket is nevertheless justified since we do not possess any other means of accelerating bodies to high velocities.

In conclusion let us find the value of $z = m/M_{\text{ter}}$ which yields maximum efficiency η and the magnitude of this maximum. In view of (9.11.10), the problem reduces to determining the maximum of the function $F(z) = z^{-1}[\ln(1+z)]^2$, that is, to solving the equation

$$F'(z) = \frac{2[\ln(1+z)]}{(1+z)z} - \frac{[\ln(1+z)]^2}{z^2} = 0,$$

or

$$\ln(1+z) = \frac{2z}{1+z}. \qquad (9.11.11)$$

The simplest way of solving this equation is by graph. To do this, we determine the point of intersection of the curves $f_1 = \ln(1+z)$ and $f_2(z) = 2z/(1+z)$ (curves Γ_1 and Γ_2 in Figure 9.11.1). This yields $z \simeq 4$, whence $F(z) \simeq 0.65$ and $\eta_{\max} \simeq 0.65\alpha$ (it is easy to verify that $z \simeq 4$ corresponds to a *maximum* of $F(z)$ and not to a minimum).

Exercise

9.11.1. Prove that the value $z \simeq 4$ corresponds to the maximum of the function $F(z) = z^{-1}[\ln(1+z)]^2$.

9.12 The Mass, Center of Gravity, and Moment of Inertia of a Rod

We consider a thin rod. The x axis will lie along the rod. Denote by σ the mass per unit length of rod. Thus, on a portion dx between x and $x + dx$ there will be a mass

$$dm = \sigma\, dx. \qquad (9.12.1)$$

The quantity σ (g/cm) is the product of the volume density d (g/cm^3) of the rod's material and the cross-sectional area S (cm^2) of the rod, or $\sigma = Sd$. The rod may have a cross-sectional area and a density that depend on x, so that σ is a function of x, or $\sigma = \sigma(x)$. The quantity σ should be called the ***linear density*** or the ***density per unit length***, but since the real density d (volume density) does not enter into the subsequent computations, we will call σ, for short, the ***density***. We consider the thickness of the rod to be small and depict it merely as a straight line, a line segment of the x axis. The total mass of the rod is clearly

$$m = \int_a^b \sigma(x)\, dx, \qquad (9.12.2)$$

where a and b are the coordinates of the ends of the rod.

Let the rod be fixed on the x axis, which is assumed to be a rigid rod so thin and of so small a mass that it can be considered a massless line; the x axis is horizontal, while the y axis is directed vertically upward. The force of gravity acts on the rod (as indicated in Figure 9.12.1) tending to pull the rod downward.

Imagine that the x axis is a weight lever. Indicated schematically in the drawing is a prism supporting the x axis at the origin. The x axis can thus

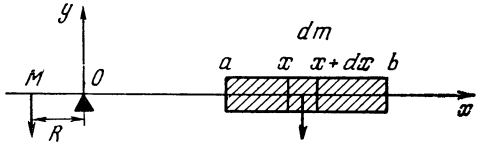


Figure 9.12.1

rotate about the axis perpendicular to the plane of the drawing. Let us find the load M on the left at a distance R along the x axis that is needed to balance the rod on the right.

By the laws of a lever, the element of mass dm distant x to the right of the support O is balanced by element of mass dM to the left if the *masses are inversely proportional to the distances*, that is, if

$$\frac{dM}{dm} = \frac{x}{R}, \quad \text{or} \quad R\,dM = x\,dm. \tag{9.12.3}$$

The element of mass dm is equal (as we learned above) to $\sigma\,dx$. To balance the entire rod we need a mass M that satisfies the equation

$$RM = \int_a^b x\sigma(x)\,dx. \tag{9.12.4}$$

This equation is the result of integrating the left and right members of (9.12.3). To the right of the axis, different elements of mass dm are located at different distances x from the support. This is why the quantity $x\sigma(x)$ appears under the integral sign. (In this way the integral determining the mass M that balances the rod differs from the integral determining the rod's mass m.) To the left of support O, all elements of mass dM (which balance the distinct elements dm of the rod) are collected together at the same distance R from the support. R is a constant and so $\int R\,dM = R\int dM = RM$.

The question now is: if we concentrate the entire mass of the rod, m, at one point C, then at what distance x_C must this point be from the support (from the origin, that is) in order to balance the mass M at distance R that is balanced by the rod (Figure 9.12.2)? We find that

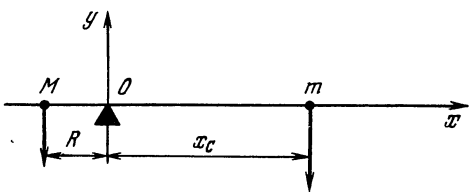


Figure 9.12.2

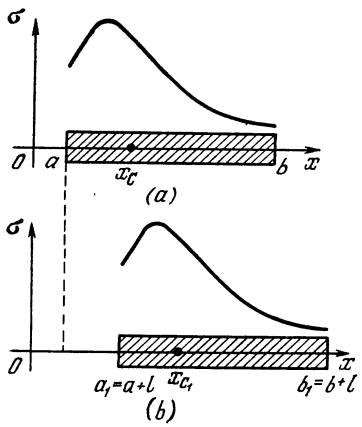


Figure 9.12.3

$$RM = x_C m = \int_a^b x\sigma\,dx, \tag{9.12.5}$$

whence

$$x_C = \frac{1}{m}\int_a^b x\sigma\,dx = \int_a^b x\sigma\,dx \Big/ \int_a^b \sigma\,dx. \tag{9.12.6}$$

The quantity x_C is the coordinate of the ***center of gravity*** or, as it is also called, the ***center of mass*** of the rod, C. It is highly important that point C is indeed a definite point of the rod: if we displace the rod as a whole along the x axis, say, a distance l to the right (Figure 9.12.3), x_C will also increase by the same quantity l, so that the point C with coordinate $x = x_C$ is always (for a given rod) at a very definite distance from the endpoints of the rod. We will prove this.

Let us consider a rod displaced a distance l to the right from the original position (Figure 9.12.3*b*); the center of gravity of the shifted rod will be denoted by C_1. Then

$$x_C = \frac{\int_a^b x\sigma(x)\,dx}{\int_a^b \sigma(x)\,dx}, \tag{9.12.7}$$

$$x_{C_1} = \frac{\int_a^b (x+l)\,\sigma\,dx}{\int_a^b \sigma\,dx} = \frac{\int_a^b x\sigma\,dx + \int_a^b l\sigma\,dx}{\int_a^b \sigma\,dx}$$

$$= \frac{\int_a^b x\sigma\,dx}{\int_a^b \sigma\,dx} + \frac{l\int_a^b \sigma\,dx}{\int_a^b \sigma\,dx} = x_C + l,$$

which constitutes a result that is obvious from the start.

The most convenient thing is to choose the system of coordinates with origin at the center of gravity of the rod (Figure 9.12.4). The quantities in this coordinate system will be denoted by a zero subscript. It is clear that

$$\int_{a_0}^{b_0} \sigma_0(x)\,dx = m.$$

The coordinate of the center of gravity x_{C_0} is zero in this system and so

$$\int_{a_0}^{b_0} x\sigma_0(x)\,dx = 0. \qquad (9.12.8)$$

In other words, if we wish to balance a rod supported at the center of gravity by a load M situated at a fixed distance R from the support, we will find that the mass required is zero (cf. (9.12.4) and (9.12.8)): if the point of support coincides with the center of gravity of a rod, the rod is balanced without applying any additional load.

We will now show that for any position of the rod its *potential energy* in

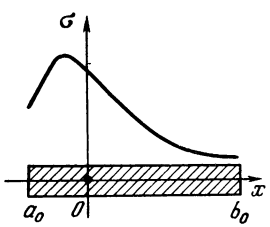


Figure 9.12.4

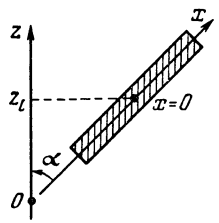


Figure 9.12.5

the field of gravity is equal to the potential energy of its entire mass concentrated at the center of gravity of the rod. We consider the position of the rod as indicated in Figure 9.12.5. The potential energy du of an element of rod with mass dm is equal to $gz\,dm$, where z is the altitude and g is the acceleration of gravity. The potential energy u of the whole rod is found by integrating. For the variable of integration we choose a length reckoned along the rod from its center of gravity; the density at point x is denoted by $\sigma_0(x)$. We express the altitude z in terms of x. As is evident from Figure 9.12.5, $z(x) = z_C + x\cos\alpha$, where z_C is the height of the center of gravity of the rod. We get

$$\int_{a_0}^{b_0} gz\,\sigma_0(x)\,dx = g\int_{a_0}^{b_0} (z_C + x\cos\alpha)\,\sigma_0(x)\,dx$$

$$= gz_C \int_{a_0}^{b_0} \sigma_0(x)\,dx + g\cos\alpha \int_{a_0}^{b_0} x\sigma_0(x)\,dx$$

$$= gx_C m,$$

since the second integral is equal to zero by formula (9.12.8). Thus, the potential energy depends only on the mass of the rod and the height of the rod's center of gravity but does not depend on the angle between the rod and the horizontal: rotation of the rod about its center of gravity requires no energy; a rod fixed by a hinge at the center of gravity is capable of freely rotating about the support without release or expenditure of energy.

Now let us investigate the concept of the *moment of inertia* of a rod. This con-

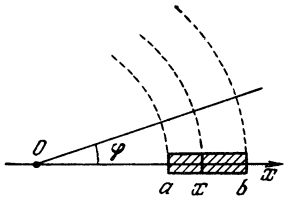


Figure 9.12.6

cept comes up in a consideration of the rotational motion of a rod. Let a rod be in rotation about an axis perpendicular to the plane of the drawing and passing through the origin. Then each point of the rod will describe a circle centered at the origin and having a radius equal to the abscissa x of the given point in the initial (horizontal) position of the rod (Figure 9.12.6). Denote by ω the angular velocity of rotation expressed in radians per second. This means that during time dt the rod rotates through the angle $d\varphi = \omega dt$. The arc length traversed by an arbitrarily chosen point with abscissa x is $dl = xd\varphi = x\omega\ dt$; hence the linear velocity of motion of a point on the rod with abscissa x is $v(x) = dl/dt = \omega x$.

Let us find the *kinetic energy* of rotation of the whole rod. An element of mass dm distant x from the center of rotation (i.e. from the origin; here we are speaking of a segment of the rod of length dx with endpoints at x and $x + dx$) has kinetic energy

$$\frac{v^2}{2}\,dm = \frac{\omega^2 x^2}{2}\,dm = \frac{\omega^2 x^2}{2}\,\sigma(x)\,dx.$$

Hence, the kinetic energy of the whole rod is

$$E = \frac{\omega^2}{2}\int_a^b x^2\sigma(x)\,dx.$$

The integral in this formula is called the ***moment of inertia*** of the rod about the axis passing through the origin and is symbolized by I, or

$$I = \int_a^b x^2\sigma(x)\,dx. \qquad (9.12.9)$$

Thus, $E = I\omega^2/2$, or the kinetic energy of rotation is expressed in terms of the moment of inertia and the angular velocity in exactly the same way that the kinetic energy of simple translation is expressed in terms of mass and linear velocity, $E = mv^2/2$. Note also that the moment of inertia (9.12.9) is always positive (as the mass m is).

If we are dealing with a system of point masses $m_1, m_2, \ldots, m_k$ lying at points $M_1, M_2, \ldots, M_k$ in a plane, the *moment of inertia* of this system about a straight line l is

$$\sum m_i d_i^2 = m_1 d_1^2 + m_2 d_2^2 + \ldots + m_k d_k^2 \qquad (9.12.10)$$

of the products of mass m_i by the square of the distance from point M_i to l, or d_i^2 (here $i = 1, 2, \ldots, k$). If we have a body (a plate or a rod, as in our case) instead of a system of point-like masses, we are dealing with a *continuous* distribution of mass and the moment of inertia is approximately equal to the moment of inertia of a system of point masses obtained through partitioning the body into many little parts and replacing each part by a point mass equal to the mass of the part. It is clear that this definition, where it is assumed that the number of these parts tends to infinity while the size of each part tends to zero, leads to a formula for the moment of inertia in the form of an *integral* (cf. formula (9.12.9)).[9.20] A sum similar to (9.12.10) but with distances d_i between points M_i and the straight line l substituted for the squares of such distances, d_i^2, is known as the ***static moment*** of a system of masses about l:

$$\sum m_i d_i = m_1 d_1 + m_2 d_2 + \ldots m_k d_k. \qquad (9.12.10a)$$

Here, however, we must allow for the sign of d_i, depending on what side of l the particular point M_i lies. It is clear, that the transition from a system of point-like masses to a continuous body leads in this case, too, to an integral replacing the sum; for instance, the static moment $\int_a^b x\sigma(x)\,dx$ of a rod about point O is present in formula (9.12.6) for the center of mass of a rod.

We now take up the evaluation of I. For a rod whose center of gravity lies at the origin O (the center of rotation),

9.20 In the case of a flat plate or a solid we have to deal with a *multiple integral* (*double* or *triple*) over the area of the plate or the volume of the solid. In this book we will not consider multiple integrals, however.

the moment of inertia assumes the value I_0:

$$I_0 = \int_{a_0}^{b_0} x^2 \gamma_0(x)\, dx. \qquad (9.12.11)$$

Let us determine the moment of inertia of a rod for the case where the center of gravity C_1 is distant l rightward from the origin, so that $x_{C_1} = l$. In this case $a = a_0 + l$, $b = b_0 + l$, $\sigma(x) = \sigma_0(x - l)$, and $I = \int_a^b x^2 \sigma(x)\, dx$. If we set $z = x - l$, then $x = z + l$ and $dx = dz$. When x varies from a to b, the quantity z varies from a_0 to b_0. Therefore,

$$I = \int_{a_0}^{b_0} (z+l)^2 \sigma_0(z)\, dz = l^2 \int_{a_0}^{b_0} \sigma_0(z)\, dz + 2l \int_{a_0}^{b_0} \sigma_0(z)\, z\, dz + \int_{a_0}^{b_0} \sigma_0(z)\, z^2\, dz. \qquad (9.12.12)$$

Note that $\int_{a_0}^{b_0} \sigma_0(z)\, dz = m$, while the second integral on the right-hand side of (9.12.12) is zero by formula (9.12.8). Finally, the third integral in (9.12.12) is I_0 by (9.12.11).

Thus, formula (9.12.12) assumes the form

$$I = ml^2 + I_0. \qquad (9.12.13)$$

The quantity ml^2 is clearly the moment of inertia of a point mass m distant l from the axis of rotation (from the origin). Thus, the moment of inertia of a rod about an arbitrary axis perpendicular to the rod is equal to the sum of the moment of inertia of the rod about the parallel axis passing through the center of gravity plus the moment of inertia of a mass equal to the rod mass about an axis parallel to the first two axes and separated from this mass by a distance equal to the distance of the center of gravity of the rod from the arbitrary axis.

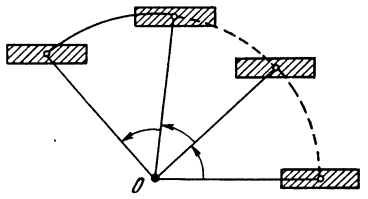


Figure 9.12.7

We can picture a rod hinged at the center of gravity. Rotation of the axis and the hinge need not be accompanied by rotation of the rod, that is, we can visualize a motion with successive stages as indicated in Figure 9.12.7. The kinetic energy E' of such motion is $mv_{C_1}^2/2$, with v_{C_1} the velocity of the center of gravity of the rod. But $v_{C_1} = \omega l$, so that $E' = (\omega^2/2)\, ml^2$.

The motion we considered earlier (see Figure 9.12.6) differs from that of Figure 9.12.7 in that in the former case the rod itself was in rotation with an angular velocity ω about its center of gravity. For this reason, the kinetic energy of rotation in Figure 9.12.6 proves to be equal to the sum of the energy of rotation of the type in Figure 9.12.7 and of the energy of rotation about the center of gravity, which is equal to $I_0\omega^2/2$.

It is evident from the derivation of the formula that such a simple addition of energies in the combination of two motions only results when we consider the motion of the center of gravity; only then do we find the integral (9.12.8) to be equal to zero.

We can approach the concept of the center of gravity of a rod from another angle. From elementary physics it is well known that the resultant of two parallel (and pointing in the same direction) forces f_1 and f_2 applied at points M_1 and M_2, respectively, is a force $f_1 + f_2$ applied at the point C that divides the segment M_1M_2 in the ratio $M_1C : CM_2 = f_2 : f_1$ (Figure 9.12.8*a*). If f_1 and f_2 are forces of gravity acting on masses m_1 and m_2, then point C (such that $M_1C : CM_2 = f_2 : f_1 = m_2 : m_1$) is called the ***center of gravity*** of the two masses; if, in addition, we take the straight line M_1M_2 as the abscissa axis and assume the abscissas of points M_1 and M_2 to be x_1

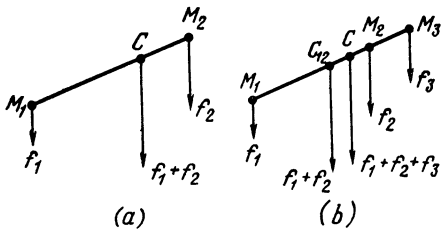


Figure 9.12.8

and x_2, then abscissa x_C of point C is $(m_1x_1 + m_2x_2)/(m_1 + m_2)$. Indeed, then

$$M_1C \div CM_2$$

$$= \left| x_1 - \frac{m_1x_1+m_2x_2}{m_1+m_2} \right| \div \left| \frac{m_1x_1+m_2x_2}{m_1+m_2} - x_2 \right|$$

$$= \left| \frac{m_2(x_1-x_2)}{m_1+m_2} \right| \div \left| \frac{m_1(x_1-x_2)}{m_1+m_2} \right| = m_2 \div m_1.$$

Similarly, if we have three masses m_1, m_2, and m_3 lying on a single straight line at points M_1 (x_1), M_2 (x_2), and M_3 (x_3) (Figure 9.12.8*b*), the resultant of the forces of gravity applied to the first two masses is the force applied to a mass m_1+m_2 lying at point $C_{12}\left(\frac{m_1x_1+m_2x_2}{m_1+m_2}\right)$. The resultant of all three forces will be applied to the center of gravity of the masses $m_1 + m_2$ and m_3, that is, to the point C with the abscissa

$$x_C = \left[(m_1+m_2) \frac{m_1x_1+m_2x_2}{m_1+m_2} + m_3x_3 \right] \Big/ [(m_1 + m_2)+m_3] = \frac{m_1x_1+m_2x_2+m_3x_3}{m_1+m_2+m_3}.$$

Quite similarly we can prove that the resultant of the forces of gravity of a system of masses $m_1, m_2, \ldots, m_k$ lying at points M_1 (x_1), M_2 (x_2), . . ., M_k (x_k) is applied at the point

$$C\left(\frac{m_1x_1+m_2x_2+\ldots+m_kx_k}{m_1+m_2+\ldots+m_k}\right) = C\left(\frac{\sum\limits_i m_ix_i}{\sum\limits_i m_i}\right) \tag{9.12.14}$$

(see Exercise 9.12.4a).

Reasoning along the same lines, we can establish that if k masses m_1, m_2, . . ., m_k lie in space at points M_1 (x_1, y_1, z_1), M_2 (x_2, y_2, z_2), . . ., M_k (x_k, y_k, z_k), the resultant of all the forces of gravity is a force applied to the mass $m_1 + m_2 + \ldots + m_k$ at the center of gravity of the system, a point C (x_C, y_C, z_C) with coordinates

$$x_C = \frac{\sum\limits_i m_ix_i}{\sum\limits_i m_i}, \quad y_C = \frac{\sum\limits_i m_iy_i}{\sum\limits_i m_i}, \quad z_C = \frac{\sum\limits_i m_iz_i}{\sum\limits_i m_i} \tag{9.12.15}$$

(see Exercise 9.12.4b).

Now let us take a rod with a linear density $\sigma = \sigma(x)$. If we partition the rod by points $x_0 = a$, x_1, x_2, . . ., $x_n = b$ into n small parts with masses $m_i = \sigma(x_i)\,\Delta x_i$, where $\Delta x_i = x_i - x_{i-1}$ and $i = 1, 2, \ldots, n$, and then consider n point-like masses m_i corresponding to the n parts of the rod (each part is assumed homogeneous since within each Δx_i the density changes little) and concentrated at the endpoints of each part, $x = x_i$, the resultant force of gravity for these point masses will be applied to the center of gravity C_1 (x_{C_1}), where $x_{C_1} = \sum\limits_i \sigma(x_i)x_i\,\Delta x_i \Big/ \sum\limits_i \sigma(x_i)\,\Delta x_i$. Sending n to infinity and all the Δx_i to zero and replacing the sums with integrals, we arrive at formula (9.12.6) for the center of gravity of the rod.

Exercises

9.12.1. Find the moment of inertia about the center of gravity of a rod of length l with a uniform distribution of mass.

9.12.2. A rod is made up of two pieces: one piece of length l_1 has a constant density σ_1, and the other one of length l_2 has a constant but different density σ_2. Find the position of the center of gravity of the rod.

9.12.3. Find the position of the center of gravity and the magnitude of the moment of inertia about the center of gravity of a rod in the form of a thin triangle of length L. Express these quantities in terms of the length L and the mass m of the rod. [*Hint.* If the x axis lies along a median and the origin is chosen at the respective vertex of the triangle, then $\sigma(x) = ax$, with a a constant.]

9.12.4. Prove (a) formula (9.12.14) and (b) formula (9.12.15).

9.13* Centers of Gravity of a String and of a Plate

In Section 7.11 we encountered the notions of the *center of gravity* of a flat *plate* and of a *string* that was bent in an arbitrary manner. Suppose AB, with $A = A(a_1, b_1)$ and $B = B(a_2, b_2)$, is a string of arbitrary shape of linear density $\sigma(x, y)$ (Figure 9.13.1). In other words, we assume that a small sec-

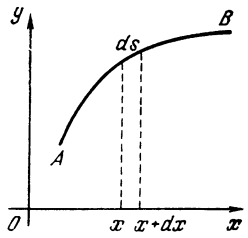


Figure 9.13.1

tion of the string of length $ds = \sqrt{(x')^2 + (y')^2}\,dt$, with $x = x(t)$ and $y = y(t)$ the parametric equations of AB, has a mass $dm = \sigma(x, y)\,ds$, where (x, y) is a point within the considered section, say, its end. If we replace the continuous distribution of mass, or our string, with a system of n point-like masses $dm_i = \sigma(x_i, y_i)\,ds_i$ $(i = 1, 2, \ldots, n)$, each corresponding to the section ds_i of the string with endpoints $A_{i-1}(x_{i-1}, y_{i-1})$ and $A_i(x_i, y_i)$, where $A_0 = A$, A_1, A_2, . . ., $A_n = B$ are the points that partition the string into n small parts, we conclude that the center of gravity C_1 of these point-like masses has the following coordinates:

$$x_{C_1} = \frac{\sum\limits_i x_i \sigma(x_i, y_i)\,ds_i}{\sum\limits_i \sigma(x_i, y_i)\,ds_i}, \qquad y_{C_1} = \frac{\sum\limits_i y_i \sigma(x_i, y_i)\,ds_i}{\sum\limits_i \sigma(x_i, y_i)\,ds_i} \tag{9.13.1}$$

(cf. (9.12.14) and (9.12.15)). Going over to the limit $n \to \infty$, $ds_i \to 0$, we arrive at the following expressions for the coordinates x_C and y_C of the center of gravity C of the string:

$$x_C = \frac{\int\limits_A^B x\sigma(x, y)\,ds}{\int\limits_A^B \sigma(x, y)\,ds}, \qquad y_C = \frac{\int\limits_A^B y\sigma(x, y)\,ds}{\int\limits_A^B \sigma(x, y)\,ds}; \tag{9.13.2}$$

here $\int\limits_A^B$ stands for integration along the curve AB. If $x = x(t)$ and $y = y(t)$ are the parametric equations of AB, and t_1 and t_2 correspond to the endpoints A and B, then along this string we have $\sigma = \sigma(x(t), y(t)) = \sigma(t)$, the total mass of the string is

$$m = \int\limits_{t_1}^{t_2} \sigma(t) \sqrt{(x')^2 + (y')^2}\,dt,$$

and the coordinates of the center of gravity are

$$x_C = \frac{1}{m} \int\limits_{t_1}^{t_2} x(t)\,\sigma(t) \sqrt{(x')^2 + (y')^2}\,dt,$$

$$y_C = \frac{1}{m} \int\limits_{t_1}^{t_2} y(t)\,\sigma(t) \sqrt{(x')^2 + (y')^2}\,dt. \tag{9.13.3a}$$

If the form of the string AB is given explicitly,[9.21] $y = y(x)$, with $a \leqslant x \leqslant b$, then $\sigma = \sigma(x, y(x)) = \sigma(x)$, the mass is

$$m = \int\limits_a^b \sigma(x) \sqrt{1 + (y')^2}\,dx,$$

and

$$x_C = \frac{1}{m} \int\limits_a^b x\sigma(x) \sqrt{1 + (y')^2}\,dx,$$

$$y_C = \frac{1}{m} \int\limits_a^b y(x)\,\sigma(x) \sqrt{1 + (y')^2}\,dx. \tag{9.13.3b}$$

In the particular case of a *uniform* string of constant density, $\sigma = \text{constant}$, we find that $m = \sigma S$, where $S = \int\limits_a^b ds$ is the length of the string, whereby formulas (9.13.2), (9.13.3a),

[9.21] Here we assume that each value of x corresponds to a *single* point of the string; if this is not so, we can simply break up the string into several sections in each of which the above condition is satisfied.

and (9.13.3b) can be rewritten thus:

$$x_C = \frac{1}{S}\int_A^B x\,ds,\quad y_C = \frac{1}{S}\int_A^B y\,ds, \tag{9.13.4}$$

or

$$x_C = \frac{1}{S}\int_{t_1}^{t_2} x(t)\sqrt{(x')^2+(y')^2}\,dt,$$

$$y_C = \frac{1}{S}\int_{t_1}^{t_2} y(t)\sqrt{(x')^2+(y')^2}\,dt, \tag{9.13.5a}$$

and

$$x_C = \frac{1}{S}\int_a^b x\sqrt{1+(y')^2}\,dx,$$

$$y_C = \frac{1}{S}\int_a^b y(x)\sqrt{1+(y')^2}\,dx. \tag{9.13.5b}$$

Note that all formulas for the centers of gravity contain the *static moments* of the string about axes Ox and Oy (see p. 340).

The situation is somewhat more complicated when we deal with a plate D (Figure 9.13.2; the thickness of the plate can be assumed so small that it can be ignored). Actually, the problem of finding the center of gravity of a plate is reduced to evaluating certain double integrals over the area of the plate, which integrals are not discussed (or even formulated) in this book. However, there is a certain method that makes it possible to overcome this difficulty.

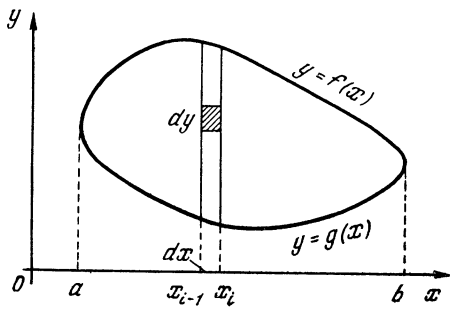


Figure 9.13.2

We denote by $\rho = \rho(x, y)$ the density of the plate per unit surface area; in other words, we will assume that the mass of a small rectangle with sides dx and dy, with dx and dy very small (see Figure 9.13.2; the sides of the rectangle are taken parallel to the coordinate axes), is equal to $\rho(x_0, y_0)\,dxdy$, where $dxdy$ is the area of the rectangle, obviously, and $\rho(x_0, y_0)$ is the density at a point (x_0, y_0) of the rectangle, say, at one of its vertices (since the rectangle is small, the density changes but little within its limits). Next we "cut" the plate into n vertical strips by the straight lines $x = x_0 = a$, $x = x_1$, $x = x_2$, ... , $x = x_n = b$, the width of each strip being $\Delta x_i = x_i - x_{i-1}$ $(i = 1, 2, \ldots, n)$; here a and b are the smallest and greatest of the abscissas within plate D (Figure 9.13.2). We may also assume that the ith strip (or the ith rod) has a linear density that depends on y,

$$\sigma_i(y) = \rho(x_i, y)\,\Delta x_i, \tag{9.13.6}$$

and a mass

$$\Delta m_i = \int_{g(x_i)}^{f(x_i)} [\rho(x_i, y)\,\Delta x_i]\,dy$$

$$= \left[\int_{g(x_i)}^{f(x_i)} \rho(x_i, y)\,dy\right]\Delta x_i, \tag{9.13.7}$$

where $y = g(x_i)$ and $y = f(x_i)$ are the smallest and greatest values of the ordinates y of the points of the ith rod; precisely, $(x_i, g(x_i))$ and $(x_i, f(x_i))$ are points of intersection of the straight line $x = x_i$ with the boundary [9.22] of plate D. (Note that the integrand in the term on the right-hand side of (9.13.7) is a function of a *single* variable y, since the value of x_i is fixed.) This mass Δm_i distributed "over" the rod can be replaced with a point-like mass Δm_i concentrated at the center of gravity

[9.22] For the sake of simplicity we assume that plate D has the "oval" shape as depicted in Figure 9.13.2, that is, is restricted by two simple curves $y = g(x)$ and $y = f(x)$ taken within the limits $x = a$ and $x = b$.

of the rod at point $C_i(x_{C_i}, y_{C_i})$ with coordinates

$$x_{C_i} = x_i,$$

$$y_{C_i} = \int_{g(x_i)}^{f(x_i)} y\rho(x_i, y)\,dy \Big/ \int_{g(x_i)}^{f(x_i)} \rho(x_i, y)\,dy \tag{9.13.8}$$

(cf. (9.13.6) and (9.12.6)); [9.23] note that the factor Δx_i in the density (9.13.6) cancels out in the fraction in the expression for the ordinate y_{C_i} of the center of gravity C_i.

Thus, we have replaced the force of gravity acting on the plate with n forces acting on the masses Δm_1, Δm_2, . . ., Δm_n (cf. (9.13.7)) at points $C_1, C_2, \ldots, C_n$ (see (9.13.8)). What remains to be done is to find the resultant of these forces applied to the center of gravity of the system of rods considered here and then (as usual) to pass from sums to integrals via the limiting process $n \to \infty$, $\Delta x_i \to 0$ (for all i's from 1 to n).

The derivation of the corresponding formulas poses no difficulties (see Exercise 9.13.2), but still it lies outside the scope of this book, since it involves evaluating *iterated* integrals, that is, integrals that contain functions expressed in terms of other integrals. For this reason we will not clutter our exposition with such formulas. Instead, we consider the case of a *homogeneous* plate with constant density ρ = constant, which of course can be set at unity: $\rho(x, y) \equiv 1$. Then a rod of width Δx_i possesses a mass

$$\Delta m_i = \int_{g(x_i)}^{f(x_i)} dy\,\Delta x_i = [f(x_i) - g(x_i)]\,\Delta x_i, \tag{9.13.9}$$

which coincides with the area of the rod (the product of its height $f(x_i) - g(x_i)$ by the width Δx_i), and the center of gravity C_i lies, naturally, in the midpoint:

$$x_{C_i} = x_i, \quad y_{C_i} = \frac{f(x_i) + g(x_i)}{2}. \tag{9.13.10}$$

The sum of all masses (9.13.9) is

$$m = \sum_i \Delta m_i = \sum_i [f(x_i) - g(x_i)]\,\Delta x_i.$$

Going over to the limit $n \to \infty$ $\Delta x_i \to 0$, we get

$$M = \int_a^b [f(x) - g(x)]\,dx,$$

which is obviously the area S of plate D (cf. Section 7.5). The coordinates of the center of gravity C' of the strips (rods) with the masses (9.13.9) concentrated at points (9.3.10) are

$$x_{C'} = \sum_i x_i [f(x_i) - g(x_i)]\,\Delta x_i / S,$$

$$y_{C'} = \sum_i \frac{f(x_i) + g(x_i)}{2} [f(x_i) - g(x_i)]\,\Delta x_i / S$$

(cf. (9.12.14) and (9.12.15)), or

$$x_{C'} = \sum_i x_i [f(x_i) - g(x_i)]\,\Delta x_i / S,$$

$$y_{C'} = \frac{1}{2} \sum_i \{[f(x_i)]^2 - [g(x_i)]^2\}\,\Delta x_i / S.$$

The same limiting process $n \to \infty$, $\Delta x \to 0$ leads to the following (exact) formulas for the coordinates x_C and y_C of the center of gravity C of the ho-

[9.23] Formula (9.12.6) refers to the case of a *horizontal* rod, but already from the definition of the center of gravity of a rod as a point C such that when the rod is fixed at this point, it is in equilibrium (see p. 339), follows the equal status of the x and y axes (one is transformed into the other if we rotate the rod through 90° about point C), whereby the second formula in (9.13.8) follows directly from (9.12.6).

mogeneous plate D with surface area S:

$$x_C = \frac{1}{S}\int_a^b x\,[f(x) - g(x)]\,dx,$$

$$y_C = \frac{1}{2S}\int_a^b \{[f(x)]^2 - [g(x)]^2\}\,dx, \qquad (9.13.11)$$

which we will now write out in full for the special case of a curvilinear trapezoid $D \equiv ABCD$, which is formed by the x axis, the straight lines $x = a$ and $x = b$, and the curve $y = f(x)$, that is, for the case with $g(x) \equiv 0$:

$$x_C = \frac{1}{S}\int_a^b xf(x)\,dx,$$

$$y_C = \frac{1}{2S}\int_a^b [f(x)]^2\,dx, \qquad (9.13.12)$$

where, as we already know, $S = \int_a^b f(x)\,dx$.

Exercises

9.13.1. Find the center of gravity of a homogeneous plate that is (a) a triangle, (b) a trapezoid, and (c) a half-disk.

9.13.2. How can we find the coordinates x_C and y_C of the center of gravity C of a plate D of density $\rho = \rho(x, y)$ limited by the curves $y = g(x)$ and $y = f(x)$ (with $a \leqslant x \leqslant b$)?

9.14 The Motion of a Body in a Medium that Resists this Motion with a Force Dependent Solely on the Velocity

When in motion, every body experiences a counteraction from the medium in which the motion is taking place. If the resistance is slight, say, in the motion in air with a low velocity, it can often be neglected. However, in some cases this approach is not satisfactory and the resistance has to be taken into account.

It has been established in experiments that if a body is moving in a liquid or a gas and the speed is low and the body is small, the force of resistance is ***proportional to the speed***:

$$F(t) = -kv(t). \qquad (9.14.1)$$

Here the coefficient of proportionality, k, is positive, and the minus sign shows that the force of resistance is in opposition to the velocity of the body. The number k depends on the properties of the medium and is proportional to the ***viscosity*** of the medium.[9.24] In addition, k is dependent on the shape and dimensions of the body. For example, if the body is a ball of radius R, formula (9.14.1) assumes the form of the ***Stokes law***:[9.25]

$$F = -6\pi R\eta v(t), \qquad (9.14.2)$$

where η is the viscosity of the medium. For air, $\eta = 1.8 \times 10^{-4}$ g/cm·s, while for water at 20 °C, $\eta = 0.01$ g/cm·s.

We consider the problem of deceleration of a body. Suppose that some force has imparted a velocity to a body and at time $t = t_0$ has ceased to act. The body continues to move and is acted upon by the force of resistance alone.

[9.24] Viscosity η can be defined as follows. Let a liquid (or gas) be in motion along the x axis, and the velocities of the various particles be distinct and dependent on the y coordinate. It is clear that a solid could not move in that fashion and would be destroyed. In a liquid or a gas, in this case there arises, between adjacent layers, a force of friction proportional to the difference in the velocities of the adjacent layers, that is, to the derivative dv/dy. The proportionality constant in the expression for force f per square centimeter of horizontal surface area is called the ***viscosity*** $f = \eta\,(dv/dy)$ (here the force f is expressed in N/cm^2 rather than in N, which implies that the dimensions of viscosity are g/cm·s).

[9.25] Sir George Gabriel ***Stokes*** (1819-1903) was a British mathematician and physicist. Formula (9.14.2) is valid for $vR\rho/\eta < 5$, where ρ is the density of the medium. The reader can easily see that the quantity $vR\rho/\eta$ is *dimensionless*. It is called the ***Reynolds number*** (abbreviated: N_{Re}) after the British physicist and engineer Osborn ***Reynolds*** (1842-1912).

By Newton's second law,

$m\,(dv/dt) = -kv.$

Dividing both sides by m and setting $k/m = \alpha$, $(\alpha > 0)$, we get

$$\frac{dv}{dt} = -\alpha v. \qquad (9.14.3)$$

The solution to this equation, as we already know (e.g. see Chapters 4 and 8) is

$$v(t) = v_0 \exp(-\alpha(t - t_0)), \qquad (9.14.4)$$

where $v_0 (= v(t_0))$ is the value of the velocity at time $t = t_0$. Since α is positive, it follows that for $t > t_0$ the exponent in (9.14.4) is negative, $\exp(-\alpha(t - t_0)) < 1$, and, hence, $v(t) < v_0$, that is, the velocity falls off with the passage of time—the medium *retards* the motion of the body, as expected.

Let us find an expression for the distance traveled by the body. From (9.14.4) it follows that

$$\frac{dx}{dt} = v_0 \exp(-\alpha(t - t_0)), \text{ or}$$

$$dx = v_0 \exp(-\alpha(t - t_0))\,dt. \qquad (9.14.5)$$

Suppose that at $t = t_0$ (initial time) the body is at the origin: $x(t_0) = 0$. Integrating (9.14.5) we find that

$$x(t) = v_0 \int_{t_0}^{t} e^{-\alpha(t - t_0)}\,dt,$$

that is,

$$x(t) = \frac{v_0}{\alpha}[1 - e^{-\alpha(t - t_0)}]. \qquad (9.14.6)$$

Using formula (9.14.6), we can find the entire distance covered by the body after time t_0, which is after the force ceases to act, up to the moment when the body stops. To this end, we note that for very large values of t, the quantity $\exp(-\alpha(t - t_0))$ is extremely small ($\exp(-\alpha(t - t_0)) \ll 1$) and can be neglected. Therefore, the body travels a total distance of v_0/α.

Let us now examine the fall of a body in air (allowing for air drag). We send the x axis downward toward the ground, and place the origin at a height H above the ground (i.e. $x = H$ on the ground). Let the motion begin at $t = 0$ with a velocity v_0. Then $x(0) = 0$ and $v(0) = v_0$. The body is acted upon by two forces: the force of gravity (assists the downward motion) and the force of air resistance (inhibits the motion).

By Newton's second law,

$$m\frac{dv}{dt} = mg - kv. \qquad (9.14.7)$$

Dividing all terms of this equation by m, we get $dv/dt = g - \alpha v$ (since $k/m = \alpha$), or

$$\frac{dv}{dt} = \alpha\left(\frac{g}{\alpha} - v\right). \qquad (9.14.8)$$

We establish the dimensions of g/α. Since $\alpha = k/m$ and $k = -F/m$, it follows that α has the dimensions of s^{-1}. The dimensions of g/α are then $cm \cdot s/s^2 = cm/s$, which means that g/α has the dimensions of *velocity*. [9.26]

Set $g/\alpha = v_1$. Equation (9.14.8) takes the form

$$\frac{dv}{dt} = \alpha(v_1 - v). \qquad (9.14.9)$$

Suppose that v_0 is less than v_1. Then the right member of (9.14.9) is positive at the start of the motion and, hence, the left member is positive too, $dv/dt > 0$; therefore, the velocity $v(t)$ *increases*. And the closer v is to v_1, the closer dv/dt is to zero and, consequently, the slower v grows. If at some time t_1 it is true that $v(t_1) = v_1$, after this v remains constant, since $v = v_1$ is the solution to Eq. (9.14.9) with initial condition $v(t_1) = v_1$. Similarly, if at the start of motion $v_0 > v_1$, velocity v approaches v_1, too, but in this case v decreases. For this reason, a certain time after the start of motion the body begins to fall at practically a con-

[9.26] The calculation performed here of the dimensions of g/α is a check. The dimensions of g/α are evident from formula (9.14.8). Since only quantities having the same dimensions can be subtracted, it follows that g/α must have the dimensions of velocity, v.

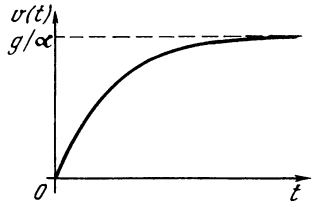


Figure 9.14.1

stant velocity of $v_1 = g\ (\alpha)$, irrespective of the velocity it had at the start of fall. The graph of velocity for $v_0 = 0$ is shown in Figure 9.14.1, where $v = g/\alpha$ is the horizontal asymptote to this graph.

The foregoing examination shows that a number of properties of $v\ (t)$ can be detected even without solving Eq. (9.14.9). Now let us solve this equation. Put $v_1 - v = z$. Then $dz/dt = -dv/dt$, and Eq. (9.14.9) can be rewritten as $dz/dt = -\alpha z$. At $t = 0$ it must be true that $z = v_1 - v_0$. The desired solution is $z\ (t) = (v_1 - v_0)\ e^{-\alpha t}$. Passing to the function $v\ (t)$, we get

$$v_1 - v\ (t) = (v_1 - v_0)\ e^{-\alpha t}, \text{ or}$$

$$v\ (t) = v_1 + (v_0 - v_1)\ e^{-\alpha t}. \quad (9.14.10)$$

Considering this equation, it is easy to see that we can draw the same conclusions as were evident from our rough analysis of Eq. (9.14.9). If $v_0 > v_1$ then $v\ (t) > v_1$, since $(v_0 - v_1)^{-\alpha t} e > 0$; but if $v_0 < v_1$, then $(v_0 - v_1)e^{-\alpha t} < 0$ and, therefore, $v\ (t) < v_1$. Secondly, no matter what v_0 is, the quantity $e^{-\alpha t}$ is small for sufficiently large t, and, practically speaking, $v\ (t) = v_1$.

Using (9.4.10), we find that

$$\frac{dx}{dt} = v_1 + (v_0 - v_1)\ e^{-\alpha t} \quad (9.14.11)$$

(the reader will recall that $v = dx/dt$), whence, knowing that $x\ (0) = 0$, we have

$$x\ (t) = v_1 t + \frac{v_0 - v_1}{\alpha}\ (1 - e^{-\alpha t}). \quad (9.14.12)$$

If the body has a large velocity and considerable dimensions, the force of resistance is *proportional to the square of the velocity*. It has been experimentally established that in this case

$$F = -kS\rho\frac{v^2}{2}, \quad (9.14.13)$$

where S is the cross-sectional area of the body, and ρ is the density of the medium. [9.27] We see that the force of resistance in this case is practically independent of the viscosity of the medium. The coefficient k in this formula is a dimensionless number, and its magnitude depends on the shape of the body (for streamlined bodies, k can drop to 0.03-0.05, while for bodies with poor streamline characteristics, k can reach 1.0-1.5). Denoting $kS\rho/2$ by $\varkappa$, we obtain

$$F = -\varkappa v^2\ (t). \quad (9.14.14)$$

It is clear that $\varkappa$ has the dimensions of g/cm.

Let us solve the problem of deceleration for the force of resistance (9.14.14). The appropriate equation is of the form

$$m\ (dv/dt) = -\varkappa v^2.$$

Dividing both sides by m and setting $\varkappa/m = \beta$, with $\beta > 0$, we get $dv/dt = -\beta v^2$, whence $dv/v^2 = -\beta\ dt$. Integrating, we get $-\frac{1}{v}\Big|_{v_0}^{v} = -\beta t\Big|_{t_0}^{t}$, where v_0 is the velocity of the body at time $t = t_0$. Therefore, $-v^{-1} + v_0^{-1} = -\beta\ (t - t_0)$, whence

$$v = \frac{v_0}{1 + \beta v_0\ (t - t_0)}. \quad (9.14.15)$$

From the formula $\beta = \varkappa/m$ we find that β has the dimensions of cm^{-1}. (Note that for $t - t_0 \gg 1/\beta v_0$, the ve-

[9.27] This formula is valid for Reynolds numbers $Rv\rho/\eta > 100$. The meaning of the formula given in the text is that for the motion of a large body, the energy expended on overcoming the resistance of the medium is not spent on the friction of layers of the liquid but on increasing the kinetic energy of the liquid, which is forced to move in order to let the body pass through it. The reader is advised to derive the formula for the force (cf. Section 9.15).

locity practically coincides with the initial velocity: $v \simeq \frac{1}{\beta(t-t_0)}$.)

Let us find the formula for the distance. Using (9.14.15), we get

$$dx = \frac{v_0}{1+\beta v_0 (t-t_0)}\, dt,$$

whence

$$x(t) = x(t_0) + \int_{t_0}^{t} \frac{v_0}{1+\beta v_0 (t-t_0)}\, dt. \qquad (9.14.16)$$

Assuming that the body begins to move from the origin, $x(t_0) = 0$, we get, from (9.14.16),

$$x(t) = \frac{1}{\beta} \ln [1+\beta v_0 (t-t_0)]. \qquad (9.14.17)$$

It is easy to see that this formula corresponds to the velocity being an *exponential* function of the distance traveled: $v = v_0 e^{-\beta x}$. If we now wish to find the *entire* distance traveled by the body after the force that imparted the velocity ceases to act, we will see that this distance (formula (9.14.17)) increases without limit with time. (Note that by formula (9.14.17) $x \to \infty$ as $t \to \infty$.)

Actually, this is not so. The point is that when the velocity of the body is small, the relation (9.14.14) no longer holds true. We have to resort to (9.14.1) and, for the distance, to formula (9.14.6).

Let us consider the problem of a body falling in air for the case of air drag being proportional to the *square* of the velocity. At start, let the body fall from the origin with an initial velocity v_0. In analogous fashion to the case where the resistance is proportional to the velocity, we get the equation

$$\frac{dv}{dt} = g - \beta v^2, \quad \text{or} \quad \frac{dv}{dt} = \beta \left(\frac{g}{\beta} - v^2\right). \qquad (9.14.18)$$

It is easy to establish that $(g/\beta)^{1/2}$ has the dimensions of velocity. Set $(g/\beta)^{1/2} = v_1$. Then $g/\beta = v_1^2$, and Eq. (9.14.18) assumes the form

$$\frac{dv}{dt} = \beta (v_1^2 - v^2). \qquad (9.14.19)$$

An exact solution to Eq. (9.14.19) is given in the answers to Exercise 9.14.1. Let us consider the general properties of the solution. Reasoning in a manner similar to the way we did for Eq. (9.14.9), we can show that in this case a velocity of $v_1 = (g/\beta)^{1/2}$ should set in. We will show that after a long enough time lapse following the start of fall, the formula

$$v - v_1 = C \exp(-2\beta v_1 t), \qquad (9.14.20)$$

with C a constant, will hold true. Equation (9.14.19) can be rewritten thus:

$$\frac{dv}{dt} = \beta (v_1 + v)(v_1 - v). \qquad (9.14.21)$$

For t large, $v \simeq v_1$, and so in this equation we replace $v_1 + v$ by $2v_1$. But if we replace v by v_1 in $v_1 - v$, we get $dv/dt = 0$, which yields $v =$ constant $= v_1$. Since we are interested precisely in the small difference between v and v_1 (the law by which v approaches v_1), it is not permissible to ignore the difference $v - v_1$. Thus, we replace (9.14.21) with

$$\frac{dv}{dt} = 2\beta v_1 (v_1 - v). \qquad (9.14.22)$$

Put $v_1 - v = z$. Then $dz/dt = -dv/dt$, and Eq. (9.14.22) assumes the form

$$\frac{dz}{dt} = -2\beta v_1 z. \qquad (9.14.23)$$

The solution to this equation is

$$z = C \exp(-2\beta v_1 t), \qquad (9.14.24)$$

which coincides with (9.14.20).

The value of C in (9.14.24) cannot be determined from the initial condition $v(0) = v_0$ (or $z(0) = v_1 - v_0$) because Eq. (9.14.22) holds true only for sufficiently large t's (near $t = 0$ we cannot replace $v + v_1$ by $2v_1$).

Also observe that the formula $F(t) = -\varkappa v^2(t)$ is valid only when $v > 0$. Indeed, if $v < 0$, then it must

be true that $F(t) = \varkappa v^2(t)$, since the force of resistance is in opposition to the velocity and, hence, is positive if the velocity is negative. Both cases ($v > 0$ and $v < 0$) are embraced by the formula

$$F(t) = -\varkappa v(t) \mid v(t) \mid.$$

Exercises

9.14.1. Find the expression for the velocity as a function of time from the equation $dv/dt = \beta(v_1^2 - v^2)$ with the initial condition $v(0) = v_0$. Using the formula for v, show that a velocity of $v_1 = (g/\beta)^{1/2}$ sets in. Show that C in (9.14.24) (or (9.14.20)) is equal to $2v_1(v_0 - v_1)/(v_1 + v_0)$.

9.14.2. In the problem of falling bodies (the resistance is proportional to the velocity) have regard for the fact that the body is acted upon by an expulsive force in accord with the Archimedean law.

9.14.3. Applying the result of the preceding exercise to a ball and noting that for a sphere $k = 6\pi R\eta$, where R is the ball's radius, and η is the viscosity of the medium, demonstrate that for large values of t the velocity of the ball, v, is equal to $2R^2g(\rho - \rho')/9\eta$, where ρ is the density of the material of the ball, and ρ' is the density of the medium.

9.15* The Motion of a Body in a Fluid

Let us now discuss in somewhat greater detail the physical nature and the regularities that characterize the forces that act on a body moving in a continuos medium, say, in air or in water. In view of the great importance of this question to technology, especially to aviation, it has been studied in great detail and constitutes the subject of a special branch of science, ***fluid dynamics***. The particular cases discussed in Section 9.14 correspond to extreme simplification of the true laws of motion of fluids.

An assumption that is more exact than the one formulated in Section 9.14 is that the force of resistance of the medium on a given body can be written in the following form:

$$F = -\frac{kS\rho v^2}{2}, \qquad (9.15.1)$$

where S is the cross-sectional area of the body (cm^2), ρ is density of the medium (g/cm^3), and k a dimensionless quantity (cf. (9.14.14)). If we do not restrict our discussion to slow motion, the coefficient k is a function of two dimensionless quantities, namely, the *Reynolds number* N_{Re} (see footnote 9.25) and the ***Mach***[9.28] ***number*** N_{Ma}, which is the ratio of the velocity v of a body with respect to the surrounding fluid to the speed of sound c in the medium, or $N_{Ma} = v/c$. Hence

$$k = k(N_{Re}, N_{Ma}). \qquad (9.15.2)$$

In the particular case of a body moving with a velocity that is low compared to the speed of sound, or $N_{Ma} \ll 1$, we have

$$k(N_{Re}, N_{Ma} = 0) = k(N_{Re}). \qquad (9.15.3)$$

The behavior of this function depends on the shape of the body, on the orientation of the body with respect to the direction of the velocity, and, finally, on what quantity is chosen as the characteristic size in the definition of the Reynolds number. Only in the simplest case of a ball do all these three factors play no role.

For low Reynolds numbers, $N_{Re} < 1$, the asymptotic behavior of (9.15.3) is

$$k = \frac{\text{constant}}{N_{Re}} + \cdots, \qquad (9.15.4)$$

where we have left out the terms that are small compared to the leading term. Substituting into (9.15.1) the value of k obtained through disregarding such terms in (9.15.4) and allowing for the fact that $N_{Re} = vR\rho/\eta$ (by definition), we get

$$F = -\frac{\text{constant} \times \eta}{R v \rho}\frac{\rho v^2}{2}S$$
$$= \text{constant} \times \eta R v, \qquad (9.15.5)$$

where the constants in the middle and right members of (9.15.5) are, of course, *different*.

We have again arrived at the Stokes law (see Eq. (9.14.2)), but what is im-

---

9.28 Ernst ***Mach*** (1838-1916), an Austrian physicist and philosopher.

portant here is that in this approximation the force of resistance is independent of the density of the fluid. The last transformation in (9.15.5) is based on the fact that the cross-sectional area S of the body is proportional to the square of the linear dimensions of the body, R^2 (here we consider geometrically similar bodies), and it is for this reason that only the first power of R remains in the expression for the force. One must bear in mind, however, that for any finite value of N_{Re} formula (9.15.4) is only approximate—it contains correction terms that are infinitesimal compared to the leading term only for very low values of N_{Re}.

In the other limiting case, $N_{\mathrm{Re}} \to \infty$, the situation proves to be more complicated. Here k can be assumed constant, as was done earlier; although the assumption that k is constant is rather crude, but in a book for beginners it is quite appropriate, we believe. In many textbooks on mechanics and exterior ballistics, the equations of motion are integrable explicitly on the assumption that k is constant, and F = constant $\times v^2$. But the aerodynamicist will say that k depends, even if weakly, on N_{Re} for large values of the Reynolds value. For instance, for a ball at $N_{\mathrm{Re}} = 100$, $k \simeq 1.2$, while at $N_{\mathrm{Re}} = 2 \times 10^5$, $k \simeq 0.4$. Then, in a relatively narrow region of Reynolds numbers from 2×10^5 to 5×10^5, the force of resistance diminishes by a factor of approximately three, after which, up to $N_{\mathrm{Re}} = 10^8$, it may be assumed that $k = 0.12$. On the whole, we see that when N_{Re} changes by a factor of a million ($10^8 \div 10^2 = 10^6$), k changes by a factor of 10 (from 1.2 to 0.12), that is, over a broad range of values of N_{Re} the following interpolation holds true: $k \propto N_{\mathrm{Re}}^{-1/6}$. In elementary (or "crude") calculations we can always substitute a constant for the extremely slowly varying function $N_{\mathrm{Re}}^{-1/6}$ (provided that the velocity varies within an interval that is not very broad).

Let us now discuss the physical picture of the motion of a liquid around a body and the origin of the forces acting on the body. Here it is more convenient to transform to a system of coordinates in which the body is fixed, that is, picture a *fixed* body in a flowing liquid.

Prior to contact with the body, that is, far from the body, in the direction opposite to that of the velocity of the liquid, the entire liquid moves with the same velocity (both in direction and magnitude). Near the body the flow changes its behavior. Clearly, the component of the velocity perpendicular to the surface of body is zero—no liquid flows in, or out of, the solid, that is, no liquid crosses the solid boundary of the body. But if the liquid has viscosity, it can be stated that the other component, the one tangent to the surface of the body, is zero too—in this case the velocity of the liquid at the surface of the body is zero.

Two forces act at the surface of the body: the force of pressure normal to the surface and the viscous force tangent to the surface and directed in the same direction as the velocity of the liquid at a small distance from the surface. The viscous force is equal to $\eta\,(\partial v_t/\partial \mathbf{n})_S$, where v_t is the tangential component of the velocity, and $\partial/\partial \mathbf{n}$ stands for the derivative along the *outward normal* $\mathbf{n}$ to the surface of the body; at the surface $v_t = 0$, but the derivative $\partial v_t/\partial \mathbf{n}$ differs from zero.

If it is known how the liquid moves at each point near the surface of a body, the calculation of the pressure and the viscous force poses no difficulty. However, the calculation of the motion is extremely difficult, and in general form is inaccessible at pressent—the calculation has been carried out only for a few simple cases.

In slow motion, the viscous forces play the principal role. The velocity distribution in various fluids with various viscosities will be similar when the fluids flow around geometrically similar bodies of different dimensions; for this reason the derivative $\partial v_t/\partial \mathbf{n}$ at a given point on the surface of the body

is proportional to the velocity v divided by the size R of the body, which means that the viscous force per unit surface area is of the order of $F_1 = \eta v/R$. The total force is $F = SF_1$, where the surface area S for geometrically similar bodies is proportional to the square of the characteristic linear dimension: $S = \text{constant} \times R^2$. Thus,

$$F = \text{constant} \times \eta v R. \qquad (9.15.6)$$

It can be demonstrated that, in the case of slow motion, allowing for the pressure distribution over the surface changes only the constant. The Stokes law (9.14.2) for a ball of radius R, where the constant is equal to 6π (here the characteristic size of the body is the ball's radius), is a particular case of the general formula (9.15.6). The product 6π in (9.14.2) was obtained through calculations and agrees well with experimental data. However, for an arbitrary nonsymmetric body the calculation becomes very complicated even in the limiting case $N_{\text{Re}} \to 0$.

Even more complicated is the case where a large body is in the stream of a liquid with large Reynolds numbers. The natural way to approach this problem is to allow for the inertia of the liquid's motion; a change in direction and magnitude of the velocity depends then on the nonuniformity in the pressure distribution.

When the stream (or jet) is decelerated, there appears a pressure of the order of ρv^2. Since the viscous force is proportional to the first power of the velocity, it can be ignored in rapid motion with large Reynolds numbers. It would seem that the result is simply $F_1 = \rho v^2$, or $F = \text{constant} \times \rho v^2 S$, with a k that is constant.

However, reality proved to be more complicated than this. In the 18th century the French mathematician, philosopher, and physicist Jean Le Rond ***d'Alembert*** found that for a liquid whose viscosity is zero there is a solution to the equations of motion in which the pressure at different points on the surface is different but the resultant of the forces of pressure is *identically zero*. In reality this situation never occurs since even the smallest viscosity modifies the flow considerably. A new phenomena emerges: although the oncoming flow is time independent, near the surface of the body and behind it the motion of the liquid changes in a random manner (the picture of the motion is like that of a flame of a bonfire). The phenomenon is known as ***turbulence***, or ***turbulent flow***.

The rise in pressure at the leading side of the body (on which the stream hits the body) is not compensated for by the pressure behind the body. The modification of the flow brought on by the viscosity of the liquid proves to be significant and results in a nonzero force of resistance, or drag. As noted earlier, the numerical value of this resistance depends very little on the Reynolds number, that is, depends but weakly on the viscosity.

In the system of coordinates where the body is in motion and the liquid is initially, that is, in the absence of the body, at rest, the cylinder of liquid through which the body travels during time t (the mass of this cylinder is $M_1 = \rho S v t$) will part to let the body through and will acquire a velocity of the order of v. Hence, the energy of this cylinder is $E = M_1 v^2/2 = t\rho S v^3/2$. This energy has to be taken away from the body; in other words, the body performs work. The work done by the body over time t is force times the distance: $A = Fvt$. Bearing in mind that $Fvt = S\rho v^3/2$, we arrive at the following formula:

$$F = \frac{S\rho v^2}{2}, \qquad (9.15.7)$$

which correctly yields the order of magnitude of this quantity. The liquid needs a certain time to "quiet down" after the body has passed through it, but its energy transforms into heat instead of returning to the body.

The way in which a liquid flows around a body depends essentially on the body's shape. Certain shapes have been

found that have k's that are smaller than the k's for a ball and disk by a factor of several tens (it is assumed that the k of the disk is measured when the disk is perpendicular to the flow).

The general idea is that the liquid, after it has let the body through, smoothly "closes ranks" behind the body. If this is so, the contribution of the difference in pressures diminishes and becomes 10 to 15% of the entire force of resistance. Even at high Reynolds numbers, the greatest part of the force of resistance (85 to 90%) directly acting on the body's surface is provided by the viscous force. But this does not mean that the total resistance force and k are proportional to the viscosity for streamlined bodies for high values of N_{Re}.

The viscosity, it appears at first glance, enters the formula for k (and F) in the first power. However, the viscous force is proportional to $\eta\,(\partial v_t/\partial \mathbf{n})$ and, all other things remaining the same (the dimension of the body, the velocity of the liquid at infinity, and the like), the transient layer within which the velocity varies gets thinner as η decreases, that is, the derivative $\partial v_t/\partial \mathbf{n}$ grows. Even for streamlined bodies and high Reynolds numbers, k is proportional to an extremely low power of η, precisely, it changes like $\eta^{1/6}$. (Some researchers prefer formulas of the type $k = (a + b \ln N_{\mathrm{Re}})^{-1}$.)

This is also the situation in the extreme case of a liquid flowing around a thin plate lying along the flow. It would seem that in the limit of small viscosities the liquid should slip along the plate, since the pressure contributes nothing and the friction force is proportional to the viscosity and is small. But in reality turbulence sets in (at $N_{\mathrm{Re}} > 10^6$) and the drag grows. [9.29] If the shape of the body or its position in in relation to the flow is nonsymmetric, for instance, the plate is at an angle to the flow, there appears, in addition to the force parallel to the velocity of the liquid, a perpendicular component—the hydrodynamic lift (or aerodynamic lift if the fluid is air), which can exceed the drag by a factor greater than 20. The flight of airplanes and gliders is based on this fact.

[9.29] In calculating the Reynolds number we substitute the length L of the plate for the radius R of the ball: $N_{\mathrm{Re}} = \rho v L/\eta$.

Chapter 10 **Oscillations**

10.1 Motion Under the Action of an Elastic Force

Let us examine the case where the force acting on a body depends solely on the position of the body, $F = F(x)$. Above (see Section 9.2) we considered in detail the work done by such a force and found out that in this case the system has a definite ***potential energy*** $u(x)$ with which the force is connected by the relation

$$F(x) = -\frac{du(x)}{dx}.$$

Let us now turn to the problem on the motion of a body under the action of such a force. The basic equation of motion is of the form

$$m\frac{dv}{dt} = F(x). \qquad (10.1.1)$$

This equation cannot be solved directly since it involves the derivative with respect to time, while the force is given as a function of the x coordinate. It is therefore natural to seek the quantities of interest as functions of the x coordinates. For one thing, we will seek the velocity as a function of the coordinate, that is, $v = v(x)$. We will then represent the derivative with respect to time, dv/dt, as the derivative of a composite function, since the x coordinate itself depends on the time:

$$\frac{dv[x(t)]}{dt} = \frac{dv[x(t)]}{dx}\frac{dx(t)}{dt}.$$

But dx/dt is nothing other than the velocity $v(x)$. Thus, $dv/dt = v(dv/dx)$, whence

$$m\frac{dv}{dt} = mv\frac{dv}{dx} = \frac{d}{dx}\left(\frac{mv^2}{2}\right). \qquad (10.1.1a)$$

Substituting this into the equation of motion (10.1.1), we get

$$\frac{d}{dx}\left(\frac{mv^2}{2}\right) = F(x). \qquad (10.1.2)$$

Integrating, we find that

$$\int_{v_0}^{v_1}\frac{d}{dx}\left(\frac{mv^2}{2}\right)dx = \int_{x_0}^{x_1}F(x)\,dx,$$

$$\text{or}\quad \frac{mv_1^2}{2} - \frac{mv_0^2}{2} = \int_{x_0}^{x_1}F(x)\,dx.$$

The physical meaning of this expression is quite clear: *the change in kinetic energy is equal to the work done by the force.*

Using potential energy $u = u(x)$, we can write Eq. (10.1.2) as

$$\frac{d}{dx}\left(\frac{mv^2}{2}\right) = -\frac{du}{dx},$$

$$\text{or}\quad \frac{d}{dx}\left(\frac{mv^2}{2} + u(x)\right) = 0.$$

If the derivative of an expression is identically zero, the expression itself is a constant, therefore

$$\frac{mv^2}{2} + u(x) = \text{constant}. \qquad (10.1.3)$$

In this form, formula (10.1.3) expresses the ***law of conservation of energy***, namely, when a body is in motion due to the action of a force depending exclusively on the position of the body (its coordinate), the sum of the kinetic energy of the body, $mv^2/2$, and its potential energy $u(x)$ remains constant. This sum is called the ***total energy*** of the body.

The purpose of these lengthy transformations here and in Section 9.14 was to show that the law of conservation of energy (as applied to mechanics) is a consequence of Newton's second law. Also a corollary to Newton's second law is the fact that the kinetic energy of a body is precisely $mv^2/2$ and not some other function of the velocity of the body. Of course, the value of the law of energy conservation is not limited by the simple examples considered in this book. This law remains valid within a much broader class of phenomena, even such phenomena whose essence has yet to be clarified.

How do we continue the solution of the problem of the motion of the body? Using the value of the velocity v_0 and the x coordinate x_0 of the body at the initial time ($t = t_0$), we find the total energy E of the body, which is a quantity that remains constant throughout the motion: $E = mv_0^2/2 + u(x_0)$. With the aid of formula (10.1.3) and knowing E, we can find the velocity of the body as a function of x:

$$v(x) = \sqrt{\frac{2}{m}[E - u(x)]}. \qquad (10.1.4)$$

It remains to find the relationship between x and t. From (10.1.4) we get

$$\frac{dx}{dt} = \sqrt{\frac{2}{m}[E - u(x)]},$$

whence

$$\frac{dx}{\sqrt{\frac{2}{m}[E - u(x)]}} = dt,$$

$$t = t_0 + \int_{x_0}^{x} \frac{dx}{\sqrt{\frac{2}{m}[E - u(x)]}}. \qquad (10.1.5)$$

Thus, the time t is expressed as a function of the x coordinate: $t = t(x)$. This function is given by an integral. We can find $x = x(t)$ by solving (10.1.5) for x. Since v is expressed as a function of x by means of a square root (Eq. (10.1.4)), even a simple expression for $u = u(x)$ frequently leads to extremely awkward expressions for $t = t(x)$.

To get a general idea of the nature of the motion, it will be useful to draw the curve of $u(x)$. Suppose the graph of $u = u(x)$ has the form depicted in Figure 10.1.1. If on the same graph we plot the horizontal line at height E (the straight line $u = E$), the picture is vivid in the extreme. Intuitively, it is clear that here the body will move between two extreme points, or that the motion is ***oscillatory***. The velocity is proportional to the square root of the difference $E - u(x)$. For instance, when $x = x_A$, the velocity is proportional to the square root of the length of the line segment AB (see Figure 10.1.1). To the right of point C and to the left of point D is a region where $E < u(x)$, that is, a region that a body with a given total energy E cannot reach for lack of energy (this is reflected in the fact that in (10.1.5) there appears a root of a negative quantity). In the region where $E > u(x)$, the square root yields two possible values for $v(x)$ in accord with two signs of the radical:

$$v = \pm\sqrt{\frac{2}{m}[E - u(x)]}.$$

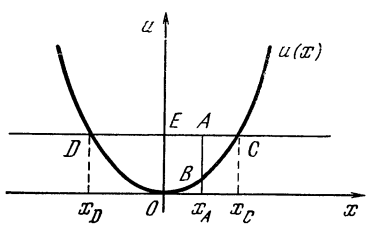


Figure 10.1.1

At the initial time, both the quantity and the sign of v_0 are determined by the initial conditions. From then on, the motion occurs in the direction specified by the sign of the initial velocity v_0. Evidently, at $v \neq 0$ the sign of the velocity cannot change suddenly. For instance, if at the initial time the body was located at the point $x = x_0$ (lying somewhere between points x_D and x_C) and $v_0 > 0$, the body will reach the extreme admissible point x_C.

At point x_C, where the body's velocity vanishes, the formula $v = \sqrt{(2/m)[E - u(x)]}$ is replaced by $v = -\sqrt{(2/m)[E - u(x)]}$. Since $v = 0$ at this point, the change in sign takes place without a jump (discontinuity) in the velocity. The situation is similar at the point $x = x_D$. Thus, in the case depicted in Figure 10.1.1, the motion of the body will be in the form of oscillations between the two extreme positions, x_C and x_D.

Let us consider another example. Suppose the potential energy of a body is given by the function $u = ax$, with a positive. Find the law of motion of the body.

We denote the values of the coordinate x and the velocity v at the initial instant of time, t_0, by $x = x_0$ and $v = v_0$. Then the total energy E is equal to $mv_0^2/2 + ax_0$. Using (10.1.4) yields

$$v(x) = \sqrt{\frac{2}{m}\left(\frac{mv_0^2}{2} + ax_0 - ax\right)}$$

$$= \sqrt{v_0^2 + \gamma(x_0 - x)},$$

with $\gamma = 2a/m$. Knowing $v(x)$, we find the time

$$t = t_0 + \int_{x_0}^{x} \frac{dx}{\sqrt{v_0^2 + \gamma(x_0 - x)}}.$$

Since under the integral sign we have $-\gamma^{-1}u^{-1/2}du$, where $u = v_0^2 + \gamma(x_0 - x)$, we can write

$$t = t_0 - \frac{2}{\gamma}\sqrt{v_0^2 + \gamma(x_0 - x)}\Big|_{x_0}^{x}$$

$$= t_0 - \frac{2}{\gamma}\left[\sqrt{v_0^2 + \gamma(x_0 - x)} - v_0\right].$$

From this we find x:

$$\frac{\gamma}{2}(t - t_0) = -\sqrt{v_0^2 + \gamma(x_0 - x)} + v_0.$$

Transposing v_0 to the left, squaring, and canceling γ, we obtain

$$x = -\frac{\gamma}{4}(t - t_0)^2 + v_0(t - t_0) + x_0. \qquad (10.1.6)$$

Finding d^2x/dt^2 we are convinced that what we have is a case of ***uniformly decelerated motion.*** This was to be expected since $F = -du/dx = -a$, that is, the force is constant and negative, which means that the motion is uniformly decelerated. In this elementary case where the force was actually independent of x, there was of course no reason for employing such a complicated computational procedure.

In the next example (in this connection see also Chapter 16) we will consider the potential energy whose graph is in the form of a *step* (Figure 10.1.2*a*). Such a function $u(x)$ is associated with the graph of force given in Figure 10.1.2*b* (to convince oneself that this is so, the reader should recall that $F = -du/dx$): the force is extremely great and negative, that is, in the direction of decreasing x. The steeper the curve representing $u(x)$ is (the curve

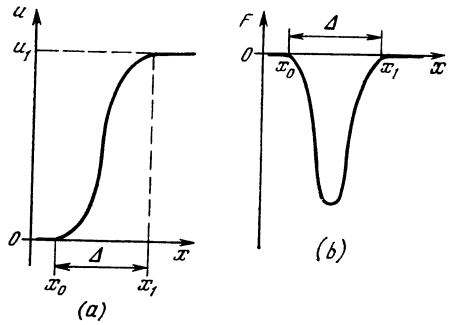


Figure 10.1.2

in Figure 10.1.2*a*), that is, the shorter the interval $\Delta = x_1 - x_0$ over which the rise in $u(x)$ occurs, the greater is the force in absolute value. The force is zero where the function $u = u(x)$ is constant (to the left of x_0 and to the right of x_1).

Let a body start moving from x_0 (see Figure 10.1.2*a*) with a velocity v_0. Suppose the total energy of the body is equal to E. For what values of E can the body reach point x_1? Since $u(x_0) = 0$, it follows that $E = mv_0^2/2$. On the other hand, $E = mv_1^2/2 + u_1$, where v_1 is the body's velocity at point x_1, and u_1 is the potential energy for $x = x_1$. Therefore

$$\frac{mv_1^2}{2} = E - u_1. \qquad (10.1.7)$$

From this formula it is evident that if E is lower than u_1, the body cannot reach x_1 because then v_1^2 becomes negative, which is impossible. For this reason, the body can reach $x = x_1$ only if $E \geqslant u_1$.

For this case we determine the work done by the force F in displacing the body from x_0 to x_1:

$$A = \frac{mv_1^2}{2} - \frac{mv_0^2}{2} = \frac{mv_1^2}{2} - E.$$

Using (10.1.7), we find that

$A = E - u_1 - E = -u_1$.

The force F does not perform any work in further motion of the body to the right of point x_1, since $F = 0$ for $x > x_1$.

Exercises

10.1.1. The potential energy is given by the formula $u = kx^2/2$, where k is positive. Construct a graph and show that the motion is oscillatory.

10.1.2. The potential energy $u(x)$ is zero for $x \leqslant 0$, equal to $2x$ for $0 \leqslant x \leqslant 1$, and equal to 2 for $x \geqslant 1$. At the initial time t_0 a body of mass 1 gram leaves the origin and moves rightward with a velocity v_0 (cm/s): (a) $v_0 = 1$, (b) $v_0 = 1.9$, and (c) $v_0 = 2.1$. In each case investigate whether the body can continue moving indefinitely to the right.

If it cannot, find the point at which it will stop.

10.1.3. $u(x) = -x^3 + 4x^2$. At the initial instant of time a body of mass 2 grams starts out from a point x_0 at a velocity v_0 (cm/s): (a) $x_0 = 1$, $v_0 = 1$, (b) $x_0 = -2$, $v_0 = 1$, and (c) $x_0 = -2$, $v_0 = -1$. In each case investigate the nature of the body's motion (points at which the body stops, regions which the body cannot reach). In the case of stopping points, give at least a rough idea of their coordinates.

10.1.4. The same requirements relative to $u(x) = x^2/(1 + x^2)$ as in Exercise 10.1.3, with $m = 2$ grams: (a) $x_0 = 0$, $v_0 = 2$ and (b) $x_0 = 1/2$, $v_0 = 1/2$. Express the time t as a function of x in terms of an integral.

10.2 The Case of a Force Proportional to Deviation. Harmonic Oscillations

Consider a body acted upon by a force $F = -kx$. As we know, to this force there corresponds a potential energy $u = kx^2/2$. The origin is the position of stable equilibrium. The curve of potential energy (parabola) has the shape shown in Figure 10.1.1.

The motion of a body under the action of such a force constitutes ***oscillations*** to the left and to the right of the position of equilibrium. We can imagine a ball rolling from one branch of the parabola, building up speed and, by inertia, rolling up the other branch, rolling down again, and so forth. [10.1] By Newton's second law, the equation of these oscillations is

$$m\frac{d^2x}{dt^2} = -kx. \qquad (10.2.1)$$

We will not solve it by the general and rather involved procedure given in Section 10.1. Instead we will guess the type of solution and concentrate our attention on investigating the properties of the solution.

Thus, we assume that

$$x = a\cos\omega t. \qquad (10.2.2)$$

This form of the solution is chosen because the cosine is one of the simplest periodic solutions, and the process we are interested in ("oscillations") is without doubt a periodic one. Substituting (10.2.2) into the basic equation (10.2.1) yields

$$-ma\omega^2\cos\omega t = -ka\cot\omega t, \qquad (10.2.3)$$

since

$$v = \frac{dx}{dt} = -a\omega\sin\omega t,$$

$$\frac{d^2x}{dt^2} = -a\omega^2\cos\omega t.$$

Formula (10.2.3) is valid for any t if $m\omega^2 = k$. Therefore, the function (10.2.2) does indeed satisfy Eq. (10.2.1) if $m\omega^2 = k$, whence

$$\omega = \sqrt{\frac{k}{m}}. \qquad (10.2.4)$$

Thus

$$x = a\cos\left(t\sqrt{\frac{k}{m}}\right). \qquad (10.2.5)$$

Observe that the square root in the expression for ω does not lead to two solutions since $\cos\omega t = \cos(-\omega t)$.

We sought the solution to Eq. (10.2.1) in the form (10.2.2), but it is quite clear that the sine is no worse than the cosine. [10.2] Equation (10.2.1) then has another solution:

$$x(t) = b\sin\omega t. \qquad (10.2.2a)$$

[10.1] Of course, this picture does not constitute an exact description of the process we are interested in, since the velocity of the body rolling from one branch of the parabola is directed along a tangent to the parabola and can be decomposed into two components, the horizontal v_x and the vertical v_z. The law of energy conservation includes, naturally, both quantities, v_x and v_z, in view of which the force on the body (more precisely, the horizotal component of this force) will not be exactly equal to $-kx$ (think of the reasons for this). But if the parabola is so gently sloping that v_z is moderate, the "rolling from the parabola" can be described as oscillatory motion obeying formula (10.2.1).

[10.2] Both functions, $x_1 = a\sin\omega t$ and $x_2 = b\cos\omega t$, are characterized by the property that their second derivative is proportional to the respective function (the proportionality factors are negative: $x_1'' = -\omega^2 x_1$ and $x_2'' = -\omega^2 x_2$). The difference lies only in the initial conditions (at $t = 0$): while $x_1(0) = 0$, for x_2 the initial rate of its change is zero ($x_2'(0) = 0$) instead of the function itself.

Indeed, $d^2 (\sin \omega t)/dt^2 = -\omega^2 \sin \omega t$. Substituting this function $x(t)$ and its second derivative into Eq. (10.2.1) and canceling $\sin \omega t$, we get $\omega = \sqrt{k/m}$, which coincides with (10.2.4).[10.3] Whence

$$x(t) = b \sin \left(t \sqrt{\frac{k}{m}} \right). \qquad (10.2.5a)$$

The reader can easily see that the sum

$$x = a \cos \omega t + b \sin \omega t \qquad (10.2.6)$$

of the solutions (10.2.2) and (10.2.2a) to Eq. (10.2.1) is also a solution. (To verify this, find the second derivative of (10.2.6) and substitute it into Eq. (10.2.1), a and b can be chosen absolutely arbitrarily.) Thus, we have the solution (10.2.6) to Eq. (10.2.1) containing two arbitrary constants, a and b.

The solution (10.2.6) to Eq. (10.2.1) can be written in a somewhat different way. Let us consider a solution to Eq. (10.2.1) of the form (10.2.2) shifted by a phase angle φ:

$$x = C \cos (\omega t + \varphi), \qquad (10.2.7)$$

where the numerical factor of the trigonometric function is now denoted by C. Since here, too, $x' = -\omega C \sin (\omega t + \varphi)$ and, hence, $x'' = -\omega^2 C \times \cos (\omega t + \varphi) = -\omega^2 x$, the function (10.2.7) is also a solution to Eq. (10.2.1) provided condition (10.2.4) is met.[10.4] But since $\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$, we can rewrite (10.7.7) thus:

$$x = C \cos \varphi \cos \omega t - C \sin \varphi \sin \omega t,$$

which means that (10.2.7) and (10.2.6) represent the same solution only if

$$a = C \cos \varphi, \quad b = -C \sin \varphi, \qquad (10.2.8)$$

where C and φ can be expressed in terms of a and b: [10.5]

$$C = \sqrt{a^2 + b^2}, \quad \tan \varphi = -b/a, \quad \varphi = \arctan (-b/a). \qquad (10.2.8a)$$

Let us find the ***period*** of oscillation, T, that is, the time during which the body returns to the initial velocity (see formulas (10.2.10) and (10.2.11)). The function $\cos \alpha$ (or $\sin \alpha$ or $\cos (\alpha + \varphi)$) returns to its initial value when the angle makes a complete revolution (changes by 2π). Thus, in the expression $a \cos \omega t$ quantity ω should vary through 2π in one period T. Therefore, T is found from the condition

$$\omega (t + T) = \omega t + 2\pi,$$

whence

$$\omega T = 2\pi, \quad T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{m}{k}}. \qquad (10.2.9)$$

The quantity $\nu = 1/T$ gives the number of oscillations per unit time and is called the ***frequency***. It has the dimensions $s^{-1} = 1/s$ (reciprocal second). The unit of frequency—one cycle, or oscillation, per second—goes by the special name ***hertz*** (abbreviated: Hz), in honor of the German physicist Heinrich ***Hertz*** (1857-1894). It is evident, from formula (10.2.9), that $\nu = \omega/2\pi$, but it is more convenient in all formulas to deal with ω and not with ν, otherwise the coefficients 2π and $4\pi^2$ will appear throughout. The quantity $\omega = 2\pi/T$ is called the ***circular frequency*** [10.6], and φ in (10.2.7) is called the ***initial phase angle.***

[10.3] The frequency $\omega = -\sqrt{k/m}$ does not yield a new solution since $b \sin (-\sqrt{k/m}\, t) = -b \sin (\sqrt{k/m}\, t)$. Thus, the value $\omega = -\sqrt{k/m}$ yields the same solution (10.2.5a), but with a different coefficient b.

[10.4] Clearly, a *shift* in the independent variable of a trigonometric function does not affect the differential properties of this function if $x = \cos (t + \alpha) = \cos \tau$, where $\tau = t + \alpha$ with α constant, then $dx/dt = (dx/d\tau)(d\tau/dt) = (dx/d\tau) \times 1 = dx/d\tau$. (Going from (10.2.2) to (10.2.7) is the same as a shift in time: $t \to t + t_0$, with $t_0 = \varphi/\omega$.)

[10.5] Of course, the simple formula $\varphi = \arctan(-b/a)$ is not sufficient for determining φ completely—we must take into account the signs of $\cos \varphi$ and $\sin \varphi$ determined from (10.2.8).

[10.6] To grasp the origin of this term, consider a line segment of length a revolving counterclockwise. The similarity between rotation and oscillation is quite apparent, since a revolving hand of a clock returns to its original position after each revolution just as an oscillating body returns to its original position after one period. Here the x coordi-

The constant a in (10.2.2) cannot be determined from Eq. (10.2.1) because the equation is satisfied for *arbitrary* a, which can be cancelled from both members of (10.2.3). The same is true for the constant b in (10.2.2a) and for both constants a and b in (10.2.6), the constants C and φ in (10.2.7) are also arbitrary (check this statement!).

If the motion of the body is described by (10.2.2), the velocity is

$$v = \frac{dx}{dt} = -a\omega \sin \omega t. \qquad (10.2.10)$$

From the relation $\cos^2 \omega t + \sin^2 \omega t = 1$ it follows that for $\cos \omega t = \pm 1$ it will be true that $\sin \omega t = 0$. Consequently, the velocity v is equal to zero at times of maximum deviation of the body in one direction or the other ($x = a$ or $x = -a$). Imagine that at $t < 0$ the body is placed at the point $x = a$ and is held at rest in this point with the aid of some external force (say a hook) until time $t = 0$, and then the hook is disengaged. At this instant the body is at rest, and oscillations are initiated under the action of a force $F = -kx$. In this case, the dependence of the coordinate x of the body upon time t is given by the formula $x = a \cos \omega t$. Since the absolute value of $\cos \omega t$ does not exceed unity, it follows that a is the largest value of x, or *the maximum deviation of the body from the position of equilibrium*. The constants b in formula (10.2.2a) and C in (10.2.7) have the same meaning. The number a (or b or C, respectively) is called the ***amplitude of the oscillations***. From (10.2.2) and (10.2.10) it follows that the amplitude is equal to the original deviation of the body if at the onset of oscillations the body was at rest.[10.7]

Let us note in passing that, generally speaking, if $A(t) = L \cos \omega t$ (or $A(t) = L \sin \omega t$), then L is the greatest value of the quantity $A(t)$ and is termed the ***amplitude*** of $A(t)$. We note also that the oscillation frequency ω is independent of the amplitude a.

Let $x = x_1(t)$ be a solution of Eq. (10.2.1), that is, let it be true that $m(d^2x_1/dt^2) = -kx_1$. We consider the function $x_2(t) = cx_1(t)$, where c is a constant. Substituting into Eq. (10.2.1) the quantities x_2 and d^2x_2/dt^2, we get $mc(d^2x_1/dt^2) = -kcx_1(t)$, or, cancelling out c,

$$m\frac{d^2x_1}{dt^2} = -kx_1.$$

Thus, if $x = x_1(t)$ satisfies Eq. (10.2.1), so does $x_2(t) = cx_1(t)$. Similarly, if $x_1(t)$ and $x_2(t)$ are two solutions to Eq. (10.2.1), $x(t) = x_1(t) + x_2(t)$ is also a solution. It is in this manner that we constructed the general solution (10.2.6) from the solutions (10.2.2) and (10.2.2a). The velocity corresponding to solution (10.2.6) is

$$v = \frac{dx}{dt} = -a\omega \sin \omega t + b\omega \cos \omega t. \qquad (10.2.11)$$

Suppose we have taken $x = a \cos \omega t$ as the solution to Eq. (10.2.1). Putting $t = 0$, we get $x_0 = a$, where $x_0 = x(0)$ is the initial position of the body, hence $x = x_0 \cos \omega t$. But then $v = dx/dt = -x_0 \omega \sin \omega t$, so that at $t = 0$ we have $v = 0$. Therefore, using the solution $x = a \cos \omega t$, all we can solve is the problem with *zero initial velocity*.

Let us try the solution $x = b \sin \omega t$. Here $v = dx/dt = b\omega \cos \omega t$, and at

---

nate of the end of the clock hand varies according to the law $x = a \cos \omega t$ if the hand revolves with an angular velocity ω. If T is the period of one revolution, then $\nu = 1/T$ is the number of revolutions per unit time and $\omega = 2\pi\nu$ is the angular velocity of rotation expressed in radians per second. Since the radian is a dimensionless quantity (see Appendix 5), ω has the dimensions of 1/s. In view of the simple meaning of ω in the case of circular motion, this quantity in problems involving oscillations is termed the ***circular frequency***.

[10.7] We defined the amplitude as *one half* of the full span of one oscillation. Between the extreme left-hand point $x = -a$ and the extreme right-hand point $x = a$, the body travels a distance of $2a$, which is *twice* the amplitude.

$t = 0$ we get $v_0 = b\omega$, where $v_0 = v(0)$ is the initial velocity, and

$$b = v_0/\omega, \qquad (10.2.12)$$

whence $x = (v_0/\omega) \sin \omega t$. But at $t = 0$ we have only $x = 0$. Again, we are unable to solve the general problem with the aid of this solution—we must assume that the oscillations started at the point of equilibrium $x = 0$.

The relation (10.2.12) offers a practical and convenient device for measuring impulse and velocity that is widely used in mechanics and goes by the name ***ballistic pendulum***. If a body is suspended in the form of a pendulum or is held in a position of equilibrium by springs and its frequency is known, the initial velocity following a blow may be determined from the amplitude of oscillations caused by the blow.

We will now show how (10.2.12) can be approximately obtained via some general elementary reasoning. The dimensions of amplitude are cm, those of velocity cm/s, and those of the oscillation period are s. Therefore, on dimensional grounds, the amplitude must be of the same order as the product of the initial velocity into some portion of the period. Since motion caused by a blow lasts a quarter period up to maximum deviation and $v < v_0$ (because the motion is retarded), it follows that $b < v_0 T/4$. If the motion has a *constant deceleration*, the average velocity would be equal to *half* the initial value and, consequently, $b \simeq v_0 T/8$. Actually, however, as follows from formulas (10.2.9) and (10.2.12),

$$b = \frac{v_0}{2} = \frac{v_0 T}{2\pi} = \frac{v_0 T}{6.28},$$

which is quite close to the result obtained via the approximate approach. The important thing here is that because the period is independent of the amplitude, the latter is directly proportional to the initial velocity.

We have verified that two different functions, (10.2.2) and (10.2.2a), satisfy the equation $m(d^2x/dt^2) = -kx$. Suppose we want to solve the problem of the motion of a body having a given initial position and a given initial velocity, that is, $x = x_0$ and $v = v_0$ at $t = 0$ (and the values of x_0 and v_0 are arbitrary, each may not be equal to zero). We will call this the ***general problem***. Up to now, in contrast to the general problem, we have only considered ***particular problems*** in one of which $x = x_0$ and $v = 0$ at $t = 0$ and in another $v = v_0$ at $t = 0$ but $x = 0$ (at $t = 0$).

With the help of (10.2.6) we can solve the general problem of the motion of a body with an arbitrary position and an arbitrary velocity at the initial time: $x = x_0$ and $v = v_0$ at $t = 0$. Using the initial conditions, we find from (10.2.6) and (10.2.11) that $a = x_0$ and $b = v_0/\omega$. Whence

$$x = x_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t. \qquad (10.2.13)$$

The solution (10.2.6) is termed the ***general solution*** to Eq. (10.2.1), while (10.2.2) and (10.2.2a) are ***particular solutions*** to Eq. (10.2.1).

The motion described by (10.2.6) or (10.2.7) is known as ***harmonic oscillations***.

If the solution is written in the form (10.2.7), then it is clear that the amplitude is equal to C. Hence, if the solution is of the form (10.2.6), $C = \sqrt{a^2 + b^2}$. Suppose $x = x_0$ and $v = v_0$ at $t = 0$. Then $a = x_0$ and $b = v_0/\omega$ and so the amplitude is

$$C = \sqrt{x_0^2 + \frac{v_0^2}{\omega^2}}. \qquad (10.2.14)$$

Exercise

10.2.1. A body is in oscillation according to the law $d^2x/dt^2 = -x$. Find the function $x = x(t)$ and determine the period for the following cases: (a) $x = 0$ and $v = 2$ cm/s at $t = 0$, (b) $x = 1$ and $v = 0$ at $t = 0$, and (c) $x = 1$ and $v = 2$ cm/s at $t = 0$. For case (c) write the solution in the form of (10.2.6) and as (10.2.7).

10.3 Pendulums

A model problem closely associated with the entire theory of harmonic oscillations is the ***pendulum*** problem. Imagine a small (point-like) mass m at the end of a weightless cord of length l, with the other end suspended from a

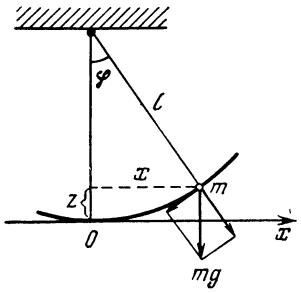


Figure 10.3.1

fixed point (Figure 10.3.1). The system is called a ***simple pendulum***, and the mass is called a bob. Suppose the bob (and the cord) is deflected through an angle φ from the vertical. One of the forces acting on the bob is mg exerted by gravity, where g is the acceleration of gravity. This force is directed downward. The *normal component* of this force, that is, the one perpendicular to the path of the bob and directed outward from the fixed point, is balanced by the force exerted by the cord, while the *tangential component* of the force of gravity, that is, the one tangent to the bob's path, is equal to $-mg \sin \varphi$; it is the latter component that drives the pendulum. (The minus sign in the expression for the tangential component shows that the component tends to *diminish* the angle φ). Since the path s that the bob travels from the equilibrium position O is $l\varphi$, the bob's velocity is $l\,(d\varphi/dt)$ and the acceleration is $l\,(d^2\varphi/dt^2)$. Newton's second law for this case is

$$ml\frac{d^2\varphi}{dt^2} = -mg\sin\varphi, \qquad (10.3.1)$$

$$\text{or } \frac{d^2\varphi}{dt^2} = -\frac{g}{l}\sin\varphi.$$

The exact solution of Eq. (10.3.1) is in no way simple (see Eq. (7.10.14) and the material that follows). But if φ is small, we can replace $\sin\varphi$ with the first term in the Taylor expansion of $\sin\varphi$ (see Section 6.2). Equation (10.3.1) is then replaced with the following approximate equation:

$$\frac{d^2\varphi}{dt^2} = -\frac{g}{l}\varphi. \qquad (10.3.1a)$$

Clearly, Eq. (10.3.1a) is simply Eq. (10.2.1), which describes harmonic oscillations, with g/l playing the role of k/m (if we divide both sides of Eq. (10.2.1) by m). Therefore, the solution to (10.3.1a) is given by (10.2.7), with $\omega = \sqrt{g/l}$. The oscillation period of a simple pendulum in this approximation is given by the formula

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}}. \qquad (10.3.2)$$

For instance, if we require that $T = 2$ s, or that it takes one second for the pendulum to swing in one direction, then we get (putting $g = 9.8$ m/s^2) $l = gt^2/4\pi^2 \simeq 0.993$ m, or $l \simeq 1$ m.[10.8]

Here is a somewhat different approach to the same simple pendulum problem, which will be more convenient when we turn to the problem of a physical pendulum. Suppose the simple pendulum of Figure 10.3.1 at $x = 0$ (the equilibrium position) has potential energy u_0, its kinetic energy at this point is zero. We deflect the pendulum through an angle φ; its horizontal deflection is then x (see Figure 10.3.1). In this position the pendulum potential energy is $u_1 = u_0 + mgz$, where $z = l - \sqrt{l^2 - x^2}$, and the kinetic energy is $(m/2)\,(dx/dt)^2$ (see Section 9.6). The sum of potential and kinetic

[10.8] In the days of the French Revolution, when the metric system was being adopted in France, two measures of length were candidates, so to say, for the basic unit of length. One was the length of 1/10 000 000 of the distance from the equator to the North pole, measured along the circle (called a meridian circle) passing through both poles and Paris (this length was called a ***meter*** and formed the base of the new metric system), while the other was the length of a "one-second" simple pendulum (the latter definition depends, obviously, on location because g varies in different parts of the world). Both units are of the same order of magnitude and agree with the characteristic dimensions of the human body.

energies will not change in the oscillation process, whence

$$u_0 + mg(l - \sqrt{l^2 - x^2}) + \frac{m}{2}\left(\frac{dx}{dt}\right)^2 = u_0,$$

$$\text{or } \left(\frac{dx}{dt}\right)^2 = -2g(l - \sqrt{l^2 - x^2}).$$

Now we employ the fact that $x \ll l$ (the deflection angle φ is samall), $x/l \ll 1$. This enables representing $\sqrt{l^2 - x^2}$ in the form

$$\sqrt{l^2 - x^2} = l\sqrt{1 - \left(\frac{x}{l}\right)^2}$$

$$\simeq l\left[1 - \frac{1}{2}\left(\frac{x}{l}\right)^2\right] = l - \frac{x^2}{2l}$$

(here we retain only the first two terms in the expansion of $\sqrt{1 - (x/l)^2} = [1 - (x/l)^2]^{1/2}$ in powers of x/l via Maclaurin's formula; see Chapter 6), after which the equation of pendulum oscillations assumes the form

$$\left(\frac{dx}{dt}\right)^2 = -g\frac{x^2}{l}.$$

Differentiating both sides of the above equation with respect to t twice, we get

$$2\frac{dx}{dt}\frac{d^2x}{dt^2} = -2g\frac{x}{l}\frac{dx}{dt},$$

whence

$$\frac{d^2x}{dt^2} = -\frac{g}{l}x. \qquad (10.3.1b)$$

This constitutes a different form of the equation of small oscillations of a (simple) pendulum; the reason why it is similar to Eq. (10.3.1*a*) is that for small x and φ we have $\sin\varphi = x/l \simeq \varphi$, that is $x \simeq l$.

The ideas developed here can be applied to the study of the oscillatory motion of a *solid* fixed at a definite "point of suspension". For the sake of simplicity we will speak of a rod suspended at a point A (Figure 10.3.2). Let the center of gravity be below the point of suspension, the distance between the point of suspension and the center of gravity being l. Let us determine the period of oscillation of the rod (the ***physical pendulum***).

If the pendulum is deflected from its

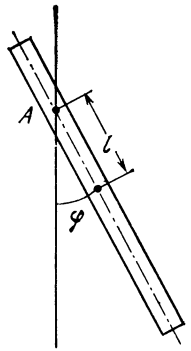


Figure 10.3.2

position of equilibrium through a small angle φ, the difference between the current height z of the center of gravity of the rod and the lowest position of this center, corresponding to the position of equilibrium, is $l - l\cos\varphi$ (cf. the aforesaid). The potential energy u of the rod can be written as

$$u = mgz = mg(l - l\cos\varphi)$$

$$= mgl(1 - \cos\varphi). \qquad (10.3.3)$$

We expand $\cos\varphi$ in a series in powers of φ and, since φ is small, confine ourselves to the first two terms: $\cos\varphi \simeq 1 - \varphi^2/2$. Therefore approximately we have $u \simeq mgl\,\varphi^2/2$. Clearly, this relationship implies that the potential energy grows with angle φ, that is, as the distance from the position of the center of gravity to the position of equilibrium (where $\varphi = 0$, $z = 0$, and $u = 0$) increases.

The kinetic energy of rotation of the rod about point A is

$$K = I\frac{\omega^2}{2} = I \times \frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2, \qquad (10.3.4)$$

where I is the moment of inertia of the rod (about the point of suspension). But according to (9.12.13), $I = ml^2 + I_0$, whence $K = \frac{1}{2}(ml^2 + I_0)\left(\frac{d\varphi}{dt}\right)^2$, where I_0 is the moment of inertia of the rod about the center of gravity.

Suppose that the rod performs harmonic oscillations, that is, $\varphi = a\cos\omega t$. By the law of conservation of energy, $u + K =$ constant and $u_{\max} = K_{\max}$, which means that $K_{\max}$ is attained at

the position of equilibrium, where $u = 0$, while u_{max} is attained at the position where $K = 0$. Since $d\varphi/dt = -a\omega \sin \omega t$, we have

$$K_{max} = \frac{1}{2}(ml^2 + I_0)\, a^2\omega^2,$$

$$u_{max} = mgl\frac{a^2}{2},$$

whence

$$mgl\frac{a^2}{2} = \frac{1}{2}(ml^2 + I_0)\, a^2\omega^2,$$

that is,

$$\omega = \sqrt{\frac{mgl}{ml^2 + I_0}}. \qquad (10.3.5)$$

The oscillation period is

$$T = 2\pi/\omega. \qquad (10.3.5a)$$

Formulas (10.3.5) and (10.3.5a) imply that the greater the moment of inertia I_0 of the rod, the lower the frequency of oscillations of the physical pendulum and, hence, the greater the period T.

If the entire mass of the rod is concentrated at the center of gravity, then $I_0 = 0$. In this case we obtain the well-known formulas (10.3.2) for the oscillation period (and frequency) of a simple pendulum.

If $I_0 \neq 0$, there is a definite position of the point of suspension for which the frequency is at a *maximum*. Since the position of the suspension point is characterized by l, to find the position we desire let us solve the equation $d\omega/dl = 0$. After simple algebraic manipulations we arrive at an equation for l:

$$mg\,(ml^2 + I_0) - mg\,l \times 2ml = 0, \qquad (10.3.6)$$

whence we have $l_{max} = \sqrt{I_0/m}$.

For a rod of length L with *uniform* distribution of mass, $I_0 = mL^2/12$ (see Exercise 9.12.1) and therefore $l_{max} = L/\sqrt{12} \simeq 0.3L$.

Note that the *minimum* frequency, $\omega = 0$, is attained, obviously, at $l = 0$, that is, when the point of suspension coincides with the center of gravity of the rod, or when no oscillatory motion takes place.[10.9] But $l = 0$ is not a root of Eq. (10.3.6), that is, $d\omega/dl$ does not vanish at this point. Here we are dealing with a so-called cuspidal minimum attained at the boundary of the region over which l varies (see Section 7.2).

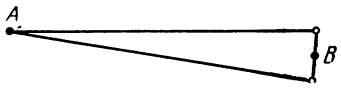


Figure 10.3.3

Exercise

10.3.1. A pendulum is in the form of a triangular lamina (tin or cardboard) (Figure 10.3.3). Determine the period of oscillation if the pendulum is suspended (a) from the acute end A, and (b) from the middle of the base, or point B. In both cases, indicate how the point of suspension is to be displaced so as to obtain a minimum oscillation period.

10.4 Oscillation Energy. Damped Oscillations

Let us write down the general solution to Eq. (10.2.1) (the general harmonic oscillation) as $x = C\cos(\omega t + \alpha)$. The potential energy of the oscillating body is equal, at every instant, to

$$u(x(t)) = \frac{kx^2(t)}{2} = \frac{kC^2}{2}\cos^2(\omega t + \alpha),$$

and the kinetic energy is

$$K(t) = \frac{mv^2}{2} = \frac{m}{2}[-C\omega \sin(\omega t + \alpha)]^2.$$

$$= \frac{mC^2\omega^2}{2}\sin^2(\omega t + \alpha).$$

The oscillation frequency, as we already know, is determined by the formula $\omega^2 = k/m$. Substituting ω^2 into the expression for the kinetic energy, we get

$$K(t) = \frac{kC^2}{2}\sin^2(\omega t + \alpha).$$

[10.9] The value $\omega = 0$, is, of course, purely nominal; it means that $\omega \to 0$ as $l \to 0$ (the closer the suspension point is to the center of gravity of the rod, the greater the period T).

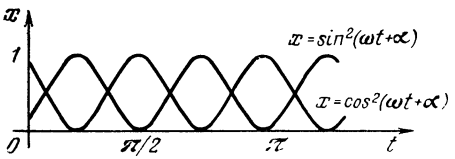


Figure 10.4.1

Thus, the factor in front of the trigonometric function in the expression for potential energy and in the expression for kinetic energy is the same. The functions themselves, $\cos^2(\omega t + \alpha)$ and $\sin^2(\omega t + \alpha)$, are very much alike, one being derivable from the other by a displacement along the time axis amounting to $\Delta t = \pi/2\omega$ (Figure 10.4.1). Each of the quantities u and K oscillates between maximum and zero: when one is at maximum, the other is at zero. Observe that the functions $\cos^2(\omega t + \alpha)$ and $\sin^2(\omega t + \alpha)$ describe oscillations about the mean (average) value equal to half the maximum. This is clearly evident from Figure 10.4.1 and also from the familiar formulas

$$\cos^2\beta = \frac{1}{2}(1+\cos 2\beta) = \frac{1}{2} + \frac{1}{2}\cos 2\beta,$$
$$\sin^2\beta = \frac{1}{2}(1-\cos 2\beta) = \frac{1}{2} - \frac{1}{2}\cos 2\beta. \qquad (10.4.1)$$

It is clear here that the quantity $(1/2)\cos 2\beta$ oscillates, taking on positive and negative values alternately, and 1/2 is the mean value of $\cos^2\beta$. The sum of the potential and kinetic energies (that is, the total energy of the system).

$$E = K + u = \frac{kC^2}{2}[\cos^2(\omega t+\alpha) + \sin^2(\omega t+\alpha)] = \frac{kC^2}{2}, \qquad (10.4.2)$$

is constant, as was to be expected.

Note that if we were to specify the motion with a frequency that did not satisfy the formula $\omega = \sqrt{k/m}$, the sum of the potential and kinetic energies would not be a constant and the maximum kinetic energy would not be equal to the maximum potential energy. This is not surprising since oscillations with a frequency different from $\omega = \sqrt{k/m}$ do not satisfy the basic equation of motion. Hence, for such oscillations to occur, it is necessary that the body be driven by some kind of other, external, forces besides the force $F = -kx$ (this force is associated with the potential $u(x) = kx^2/2$). Because of the work of external forces, the total energy $mv^2/2 + kx^2/2$ will no longer be conserved.

Let us now investigate the problem of *damping* of oscillations. Let a body be acted upon by the force of friction in addition to the force $F = -kx$ of a sping. Suppose the friction is so small over one period that the work of the friction force is small compared with the oscillation energy. We can then assume approximately that the oscillations occur as in the case of no friction: $x(t) = C\cos(\omega t + \varphi)$ (see Eq. (10.2.7)). The oscillation energy is $kC^2/2$. In the case of friction, the energy of the oscillations diminishes with the passage of time. Thus, friction will cause the amplitude C in (10.2.7) to decrease slowly instead of remaining constant. The law of decrease of C is defined by the condition that the decrease in total energy E is equal to the work of the force of friction F_1.

With respect to unit time, these quantities are related thus:

$$\frac{dE}{dt} = \frac{d(kC^2/2)}{dt} = kC\frac{dC}{dt} = F_1 v = W_1, \qquad (10.4.3)$$

where v is the velocity of the body, and W_1 is the power of the friction force. In the oscillation process, both v and F_1 vary periodically. The friction force F_1 is always dependent on velocity v and is directed *opposite* to v, so that the product $F_1 v$ always remains negative. In the case at hand of *small* friction, which results in a slow damping of the oscillations, we can assume that the change in amplitude $C(t)$ is small over several oscillation periods. The product $F_1 v$ may be understood as the average value of the product over one period.

Formula (10.4.3) holds true only for time intervals exceeding one oscillation period.

By way of an illustration, let us examine the effect of a force of friction *proportional to the first power of the velocity* of the body:

$$F_1 = -hv, \quad F_1 v = -hv^2. \tag{10.4.4}$$

If $v = dx/dt = -C\omega \sin(\omega t + \varphi)$, then

$$F_1 v = -hC^2\omega^2 \sin^2(\omega t + \varphi). \tag{10.4.5}$$

Note that the average value of $\sin^2(\omega t + \varphi)$ over one period is 1/2 (see formula (10.4.1) and Section 7.8). Using (10.4.3) and (10.4.4), we finally get

$$\frac{d}{dt}\left(\frac{kC^2}{2}\right) = kC\frac{dC}{dt} = -hC^2\omega^2 \times \frac{1}{2},$$

whence

$$\frac{dC}{dt} = -\frac{h\omega^2}{2k}C,$$

that is (cf. (6.6.10)),

$$C = C_0 e^{-\gamma t}, \tag{10.4.6}$$

where

$$\gamma = h\omega^2/2k. \tag{10.4.6a}$$

But since $\omega^2 = k/m$, formula (10.4.6a) can be simplified:

$$\gamma = h/2m. \tag{10.4.6b}$$

In formula (10.4.6), C_0 is determined by the initial conditions. Multiplying both sides of (10.4.6) by $\cos(\omega t + \varphi)$ and employing (10.2.7), we get

$$x(t) = C_0 e^{-\gamma t}\cos(\omega t + \varphi), \tag{10.4.7}$$

where $\omega = \sqrt{k/m}$ (see a characteristic graph of damped oscillations in Figure 10.4.2). This is an approximate formula obtained on the assumption that the friction is low and is proportional to the first power of velocity.

In all cases where the friction force is proportional to the first power of velocity, the problem has an exact solution. The body is acted on by two

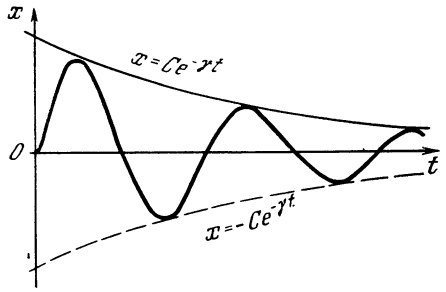


Figure 10.4.2

forces, the elastic force $-kx$ and the force of friction $-h(dx/dt)$. By Newton's second law,

$$m\frac{d^2x}{dt^2} = -kx - h\frac{dx}{dt}. \tag{10.4.8}$$

We will seek the solution $x(t)$ in the same form as was obtained for a small friction:

$$x(t) = C_0 e^{-\gamma t}\cos(\omega_1 t + \alpha). \tag{10.4.9}$$

This solution depends on four parameters: C_0, γ, ω_1, and α. The values of γ and ω_1 are determined from Eq. (10.4.8), while C_0 and α remain arbitrary, that is, Eq. (10.4.8) does not determine them, and require specifying the initial conditions. Indeed, from (10.4.9) we obtain

$$\frac{dx}{dt} = -\gamma C_0 e^{-\gamma t}\cos(\omega_1 t + \alpha) - C_0\omega_1 e^{-\gamma t}\sin(\omega_1 t + \alpha),$$

$$\frac{dx^2}{dt^2} = \gamma^2 C_0 e^{-\gamma t}\cos(\omega_1 t + \alpha) + \gamma C_0\omega_1 e^{-\gamma t}\sin(\omega_1 t + \alpha) + C_0\omega_1\gamma e^{-\gamma t}\sin(\omega_1 t + \alpha) - C_0\omega_1^2 e^{-\gamma t}\cos(\omega_1 t + \alpha).$$

Substituting into (10.4.8) the expressions for x, dx/dt, and d^2x/dt^2 and canceling out $C_0 e^{-\gamma t}$, we get

$$m\gamma^2\cos(\omega_1 t + \alpha) + m\gamma\omega_1\sin(\omega_1 t + \alpha) + m\gamma\omega_1\sin(\omega_1 t + \alpha) - m\omega_1^2\cos(\omega_1 t + \alpha) = -k\cos(\omega_1 t + \alpha) + h\gamma\cos(\omega_1 t + \alpha) + h\omega_1\sin(\omega_1 t + \alpha),$$

or

$$(m\gamma^2 - m\omega_1^2)\cos(\omega_1 t + \alpha) + 2m\gamma\omega_1 \sin(\omega_1 t + \alpha) = (-k + h\gamma)\cos(\omega_1 t + \alpha) + h\omega_1 \sin(\omega_1 t + \alpha). \quad (10.4.10)$$

The last equation holds true for arbitrary t if

$$m\gamma^2 - m\omega_1^2 = -k + h\gamma,$$
$$2m\gamma\omega_1 = h\omega_1. \quad (10.4.11)$$

Canceling ω_1 out of the second equation, we obtain $\gamma = h/2m$. Then from the first we can find ω_1:

$$\omega_1^2 = \frac{k}{m} - \frac{h^2}{4m^2}. \quad (10.4.12)$$

Consequently,

$$x(t) = C_0 e^{-\frac{h}{2m}t} \times \cos\left(\sqrt{\frac{k}{m} - \frac{h^2}{4m^2}}\, t + \alpha\right). \quad (10.4.13)$$

When friction is low, that is, h is small in comparison with k, we can ignore $h^2/4m^2$ under the radical sign as compared with k/m. Then (10.4.13) becomes (10.4.7). Thus, in the approximate consideration we correctly obtained the law of decrease of the oscillation amplitude but did not notice the small variation in frequency due to friction. Note also that of the two laws that govern the decrease in amplitude, (10.4.6), (10.4.6a) and (10.4.6), (10.4.6b), only the second holds true for arbitrary friction (not too large, that is, such that $h^2 < 4km$ and the root in (10.4.13) is meaningful), while the first coincides with the second at small k's but in the general case is invalid because it contains ω (which is not generally a constant).

If the friction is so high that $h^2 > 4km$, the radicand in (10.4.13) becomes negative and the formula becomes meaningless. This means that motion involving appreciable friction is no longer oscillatory. In this case the solution to Eq. (10.4.8) is to be sought in the form $x = Ce^{-\gamma t}$. Substituting this into Eq. (10.4.8), we arrive at quadratic equation for γ with two solutions: $\gamma = \gamma_1$ and $\gamma = \gamma_2$. The sum $C_1 e^{-\gamma_1 t} + C_2 e^{-\gamma_2 t}$ of the two solutions corresponding to γ_1 and γ_2 will yield the general solution to Eq. (10.4.8) and will enable solving the problem for arbitrary initial data. This case of large h is considered in detail in Section 13.10 in connection with electrical oscillations.

Exercises

10.4.1. Find the law of damping of oscillations for friction proportional to the second power of velocity (this friction is characteristic of rapid motion of a body in a low-viscosity fluid). Show that after the lapse of a large time interval, the amplitude $C(t) = 1/bt$, where b is a constant independent of C_0, which is the amplitude at the initial time.

10.4.2. Find the law of damping of oscillations for friction that is independent of velocity (this is characteristic of the friction between hard dry surfaces). Determine the time interval after which the oscillations cease.

10.5 Forced Oscillations and Resonance

Consider a body acted upon by an elastic force $F = -kx$. We have established that this force causes the body to oscillate with a definite frequency $\omega = \sqrt{k/m}$, which is known as the ***natural*** (or free) ***frequency.*** From now on we will denote the natural frequency by ω_0 so that $\omega_0 = \sqrt{k/m}$.

Now let the body be acted upon by the elastic force and a periodic *external* force with frequency ω, which generally may differ from ω_0. It then turns out that the amplitude of the oscillations brought about by the external force is very strongly dependent on how close the frequency ω of the external force is to the natural frequency ω_0, that is, to the frequency of natural oscillations occurring in the absence of external forces. The phenomenon in which the oscillation amplitude increases as ω approaches ω_0 is called ***resonance*** and has a wide range of applications. It refers to any systems that admit oscillations and vibrations. In

mechanical systems (machine tools or motors) such vibrations can result in deformation and destruction of the equipment. On the other hand, resonance can be useful, at times it is purposely used to produce, via a small force, vibrations of the operating tool with great amplitude.

In electric systems, resonance enables us, using several periodic forces with different frequencies (say, a number of radio transmitters), to achieve a situation in which the oscillations in our system depend solely on *one* of the periodic forces (the one whose frequency is close to the natural frequency of the system). This allows for tuning a radio set to a definite station.

Let us set up the oscillation equation:

$$m\frac{d^2x}{dt^2} = -kx - h\frac{dx}{dt} + f\cos\omega t, \tag{10.5.1}$$

where the terms $-kx$ and $-h\,(dx/dt)$ correspond to the elastic force and the friction force (see Eq. (10.4.8)), and $f\cos\omega t$ represents the external force. Divide both sides of (10.5.1) by m and set $k/m = \omega_0^2$ in accordance with the fact that (in the absence of friction) ω_0 is the natural frequency of oscillations of the body. Denote the ratio h/m by 2γ (see formula (10.4.6b)). Then Eq. (10.5.1) assumes the form

$$\frac{d^2x}{dt^2} = -\omega_0^2 x - 2\gamma\frac{dx}{dt} + \frac{f}{m}\cos\omega t. \tag{10.5.1a}$$

It is natural to expect that under the action of a force having a frequency ω the body will oscillate with that frequency. We therefore seek the solution to Eq. (10.5.1) or (10.5.1a) in the form

$$x = a\cos\omega t + b\sin\omega t. \tag{10.5.2}$$

Substituting the expressions for x and its derivatives into Eq. (10.5.1a), we obtain

$$-a\omega^2\cos\omega t - b\omega^2\sin\omega t = -a\omega_0^2\cos\omega t - b\omega_0^2\sin\omega t + 2\gamma a\omega\sin\omega t - 2\gamma b\omega\cos\omega t + \frac{f}{m}\cos\omega t.$$

For this equation to be true for *all* t's, the separate terms involving $\cos\omega t$ and $\sin\omega t$ in the right- and left-hand sides must be equal. Thus, we arrive at the following system of equations:

$$\begin{aligned} -a\omega^2 &= -a\omega_0^2 - 2\gamma b\omega + \frac{f}{m}, \\ -b\omega^2 &= -b\omega_0^2 + 2\gamma a\omega \end{aligned} \tag{10.5.3}$$

(cf. Eq. (10.4.11)). The second equation yields

$$b = \frac{2\gamma\omega}{\omega_0^2 - \omega^2}a.$$

Substituting this into the first equation in (10.5.3) yields

$$a = \frac{f}{m}\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + (2\gamma\omega)^2}. \tag{10.5.4a}$$

Then

$$b = \frac{f}{m}\frac{2\gamma\omega}{(\omega_0^2 - \omega^2)^2 + (2\gamma\omega)^2}. \tag{10.5.4b}$$

Going over to the form $x = C\cos(\omega t + \varphi)$ for oscillations and recalling that $C = \sqrt{a^2 + b^2}$ (see (10.2.8a)), we obtain the amplitude of oscillations produced by the external force:

$$C = \frac{f}{m}\frac{1}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\gamma\omega)^2}}. \tag{10.5.5}$$

We see that C is the greater the closer ω is to ω_0. The curve of C as a function of ω for a given ω_0 is shown in Figure 10.5.1 for two values of γ (with $f/m = 1$ and $\omega_0 = 1$). The less the friction h, that is, the less the quantity $\gamma = h/2m$, the sharper the rise in amplitude for the frequency of the external force equal to the natural frequency.

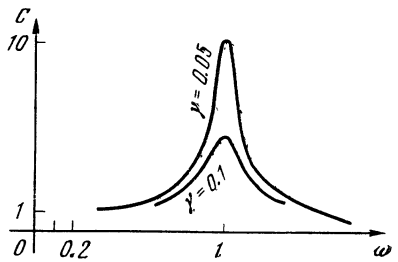


Figure 10.5.1

It is not hard to see that the sum of the solution (10.4.9) to Eq. (10.4.8) and the general solution (10.5.2) to Eq. (10.5.1),

$$x = a \cos \omega t + b \sin \omega t + C_0 e^{-\gamma t} \cos (\omega_1 t + \alpha), \qquad (10.5.6)$$

where a and b are given by (10.5.4a) and (10.5.4b), and ω_1 and γ by (10.4.12) and (10.4.6b), is also a solution to Eq. (10.5.1). Formula (10.5.6) provides a solution that satisfies *any* initial data by choosing C_0 and α. Indeed, suppose that $x = x_0$ and $v = v_0$ at $t = 0$. Then, using (10.5.6), we find that

$$x_0 = a + C_0 \cos \alpha,$$
$$v_0 = b\omega - C_0 (\gamma \cos \alpha + \omega_1 \sin \alpha). \qquad (10.5.7)$$

From this system of equations we can determine C_0 and α (see Exercise 10.5.1).

Thus, (10.5.6) is the *general solution* of the problem involving oscillations of a body under the action of an elastic force and a periodic external force. This general solution confirms the assumption, made at the beginning of this section, that under the protracted action of an external force with frequency ω a body will oscillate with the same frequency ω. This is true because no matter what the initial conditions, they only affect the values of C_0 and α, which is to say, only the last summand in the solution (10.5.6). However, in the course of time, this term, which has frequency ω_1, becomes very close to zero, due to the factor $e^{-\gamma t}$, which tends to zero as $t \to \infty$, and we can neglect it for large t. The remaining terms describe oscillations with frequency ω, which do not decay with the passage of time since they are maintained by the action of the external force, whose amplitude is constant by assumption.

Exercises

10.5.1. Determine C_0 and α from the system (10.5.7)

10.5.2. Because of friction, the maximum amplitude C is obtained at ω^2_{max} somewhat different from ω_0^2. Find the deviation of $\omega^2_{max}/\omega_0^2$ from unity as a function of γ.

Hint. Test the radicand in (10.5.5) for a minimum, denoting $\omega^2 = z$.

10.6 On Exact and Approximate Solutions of Physical Problems

In the preceding section we had the luck to find with comparative ease the *exact* solution of the problem involving oscillations of a body under the action of a periodic external force in the case of an elastic force that tends to move the body to the position of equilibrium, $-kx$, and friction, $-h\,(dx/dt)$. With the exact solution at our disposal, we can easily find a number of important limiting cases.

(1) The frequency ω of the external force is *extremely low* compared with ω_0, where $\omega_0^2 = k/m$. Neglecting ω in (10.5.5) in comparison with ω_0 we get

$$C = f/m\omega_0^2 = f/k. \qquad (10.6.1)$$

(2) The frequency ω of the external force is *much higher* than ω_0. Then we can neglect ω_0 in (10.5.5) and assume that

$$C = \frac{f}{m} \frac{1}{\sqrt{\omega^4 + 4\gamma^2\omega^2}}. \qquad (10.6.2)$$

But $\gamma^2\omega^2 \ll \omega^4$ (friction is not very great, since otherwise there would be no oscillations); neglecting the term $4\gamma^2\omega^2$, we obtain

$$C \approx \frac{f}{m\omega^2}. \qquad (10.6.2a)$$

(3) The *friction force is small*, that is, h *is small*. Then in (10.5.5) we can neglect the term containing γ. As a result we have

$$C \approx \frac{f}{m} \frac{1}{|\omega_0^2 - \omega^2|}. \qquad (10.6.3)$$

(The appearance of the absolute value of the difference in (10.6.3) is due to the fact that in (10.5.5) we take the positive value of the root, since it is clear that C and f, that is, the oscillation amplitude and the external force amplitude, can be only positive.)

(4) The phenomenon of *exact resonance*: the frequency ω of the external force is exactly equal to the natural frequency ω_0. Then (10.5.5) is replaced with

$$C = \frac{f}{m}\frac{1}{2\gamma\omega} = \frac{f}{h\omega} = \frac{f}{h\omega_0}. \qquad (10.6.4)$$

These limiting cases actually make up over 90% of the content of all the results obtained. When one obtains a general result, it is always advisable to simplify it by considering various limiting cases, as we have just done. The simple formulas relating to the limiting cases are more easily remembered and more frequently used in practical situations. Only once in a while does one have to resort to general formulas. Although limiting cases do not cover everything but we know *almost everything* that is contained in the more complicated exact formula.

The question arises quite naturally as to the possibility of obtaining limiting formulas directly via simplifications in the equation itself rather than in its solution. To solve an involved equation in exact fashion and then simplify the solution is just as senseless as to use intricate machinery to package goods elegantly and then immediately tear open the package to get them since each piece is to be used separately. Therefore it is important to learn to extract from the general solution of the problem the particular (and limiting) cases without solving the equation itself.

To obtain limiting (approximate) expressions directly is particularly important for the added reason that an exact solution is very sensitive to the slightest variations in the statement of the problem. A slight complication in the problem and one finds it impossible to get an exact solution. An approximate solution is rougher but more stable with respect to variations in the problem.

Of particular importance to students are cases where it is possible to obtain and compare both solutions, exact and approximate. It is precisely in such cases that one can acquire some experience in the proper choice of approximations and be sure of the results, since the correctness of the limiting cases that follow from the general formulas can serve as a global check on the entire process of solving the general equation, both the setting up of the equation that describes the process of interest to us and the solving of this equation.

Let us now return to the first case: the frequency ω of the external force is low. This is clearly a *slow* motion. Therefore, in the original equation (10.5.1).

$$m\frac{d^2x}{dt^2} = -kx - h\frac{dx}{dt} + f\cos\omega t,$$

we can drop terms $m\,(d^2x/dt^2)$ and $h\,(dx/dt)$ involving acceleration and velocity, since if $x \simeq C\cos(\omega t + \varphi)$, then dx/dt is proportional to ω and d^2x/dt^2 to ω^2, and for low frequencies these quantities are small. In this case we get $0 \simeq -kx + f\cos\omega t$, whence

$$x \simeq k^{-1} f\cos\omega t = C\cos\omega t,$$
$$\text{with } C = f/k. \qquad (10.6.5)$$

Therefore, for low frequencies of the external force, at each moment the applied external force $f\cos\omega t$ is balanced by the elastic force kx. This result is clearly very general, for it refers to *any* motion with a low frequency. This limiting case is called ***a static case***. For one, the elastic force $G\,(x)$ may be any function of the coordinate and not only a linear function of the coordinate like $G = -kx$, and the external force $F\,(t)$ may be any function of time. The oscillation equation takes the form

$$m\frac{d^2x}{dt^2} = G(x) - h\frac{dx}{dt} + F(t). \qquad (10.6.6)$$

It is not always possible to obtain the exact solution of this equation, but the approximate approach is preserved. Indeed, neglecting in the case of slow motion the terms in (10.6.6) that are proportional to velocity dx/dt and accel-

eration d^2x/dt^2, we get $G(x) + F(t) \simeq 0$, or $G(x) \simeq -F(x)$. From this we find an approximate relationship between x and t, or $x \simeq x(t)$. Substituting this $x(t)$ into the exact equation (10.6.6), we can find the order of the error introduced by neglecting the terms $h(dx/dt)$ and $m(d^2x/dt^2)$.

Let us take a look at the second limiting case, very high frequency ω. The time of action of the external force and, hence, the impulse during each half-cycle (while the force is acting in one direction) are small because of the short duration of the half-cycle in this case. Thus, for a given amplitude f of the external force, the higher the frequency ω, the smaller the velocity that the body can acquire and the smaller the displacement of the body. Neglecting the terms kx and $h(dx/dt)$ in Eq. (10.5.1), we arrive at an equation of motion of a free body with no forces acting except the external force:

$$m\frac{d^2x}{dt^2} \simeq f\cos\omega t. \qquad (10.6.7)$$

We seek the solution to this equation in the form $x = B\cos\omega t$. Then $d^2x/dt^2 = -B\omega^2\cos\omega t$, and whence Eq. (10.6.7) takes the form $-Bm\omega^2 \times \cos\omega t \simeq f\cos\omega t$, from which it follows that $B \simeq -f/m\omega^2$, and hence

$$x \simeq -\frac{f}{m\omega^2}\cos\omega t. \qquad (10.6.8)$$

In the standard form $x \simeq C\cos(\omega t + \varphi)$ the solution (10.6.8) can be written as follows (with C positive):

$$x \simeq \frac{f}{m\omega^2}\cos(\omega t + \pi). \qquad (10.6.8a)$$

Here, the elastic force is

$$-kx \simeq \frac{kf}{m\omega^2}\cos\omega t = \frac{\omega_0^2}{\omega^2}f\cos\omega t, \qquad (10.6.9)$$

and the force of friction is

$$-h\frac{dx}{dt} \simeq -\frac{hf}{m\omega}\sin\omega t. \qquad (10.6.9a)$$

When comparing forces that depend periodically on time, one should compare not their instantaneous values but their amplitudes. The ratio of the external force $f\cos\omega t$ to the elastic force (10.6.9) (the ratio of the amplitudes) is $f \div (\omega_0^2/\omega^2)f = \omega^2/\omega_0^2$. The ratio is the higher the greater ω is. Similarly, the ratio of the external force to that of friction, $f \div (hf/m\omega) = (m/h)\omega$, grows without limit with ω. For this reason, given high ω, the external force appreciably exceeds both the elastic force and the force of friction. This supports the possibility of an approximate consideration of motion under the action of an external force alone.[10.10]

The third limiting case, neglect of friction, presents no difficulties at all. Here Eq. (10.5.1) takes the form

$$m\frac{d^2x}{dt^2} \simeq -kx + f\cos\omega t$$
$$= -m\omega_0^2 x + f\cos\omega t, \qquad (10.6.10)$$

since $\omega_0^2 = k/m$ and, hence, $k = m\omega_0^2$. We seek the solution to this equation in the form $x = C\cos(\omega t + \varphi)$. Substituting into this equation the expressions for x and d^2x/dt^2, we get $\varphi = 0$, $\varphi = \pi$, and $m|\omega_0^2 - \omega^2| \times C\cos\omega t \simeq f\cos\omega t$, whence

$$C \simeq \frac{f}{m|\omega_0^2 - \omega^2|}. \qquad (10.6.11)$$

Comparison of this formula with the exact formula (10.5.5) enables estimating the conditions under which this approximation works.

Clearly, at $\omega = \omega_0$ approximation (10.6.11) is not valid and the fourth limiting case (the frequency of the external force is exactly equal to the natural frequency of oscillations, or resonance) must be treated separately. Here we will not ignore friction, however.) We seek the solution to (10.5.1)

[10.10] It is essential that the friction force considered above be closer to zero the closer the velocity is to zero. In ***dry friction*** (force of friction independent of velocity), an external force less than that of friction will not cause oscillations at *any* frequency.

in the standard form $x = C\cos(\omega_0 t + \varphi)$. Then

$$m\frac{d^2x}{dt^2} = -mC\omega_0^2\cos(\omega_0 t + \varphi),$$

and, since $\omega_0^2 = k/m$,

$$m\frac{d^2x}{dt^2} = -kC\cos(\omega_0 t + \varphi) = -kx.$$

Substituting into (10.5.1) the expressions for x and the time derivatives yields

$$hC\omega_0 \sin(\omega_0 t + \varphi) + f\cos\omega_0 t = 0.$$

This equation will hold true for any t if $C = f/h\omega_0$ and $\varphi = -\pi/2$. Hence, the solution of our equation is

$$x = \frac{f}{h\omega_0}\cos\left(\omega_0 t - \frac{\pi}{2}\right). \qquad (10.6.12)$$

The oscillation amplitude at resonance is $C = f/h\,\omega_0$, and we clearly see that there is no way in which we can ignore the friction, or h, at resonance.

Referring to Figure 10.6.1, let us look at the relationship between C and ω provided by the approximate formulas (10.6.3) and (10.6.4). Formula (10.6.3) yields two branches of the curve $C = C(\omega)$, branches that go off to infinity as $\omega \to \omega_0$, while formula (10.6.4) yields a finite value $C = C_0\ (= f/h\omega_0)$ at $\omega = \omega_0$. If we construct the curves of (10.6.3) and place the point $A = A(\omega_0, C_0)$ that corresponds to formula (10.6.4), it is then easy to draw free-hand a smooth curve (dashed in Figure 10.6.1) which, far from resonance, coinsides with the curves of (10.6.3) and has a maximum C_0 at $\omega = \omega_0$.

In the case of resonance, the amplitude C can be determined by means of

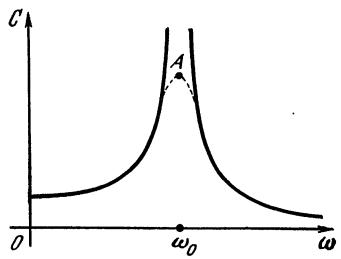


Figure 10.6.1

energy considerations, whose value consists in the fact that they also enable solving approximately certain problems that have no exact solutions. The ***power*** developed by an external force $f\cos\omega t$ in the case of motion specified by the expression $x = C\cos(\omega t + \varphi)$ is

$$W_{ex} = f\cos\omega t\frac{dx}{dt} = -fC\omega\cos\omega t\sin(\omega t + \varphi).$$

Let us determine the ***average power*** of the external force during a large (more precisely, infinite) interval of time:

$$\overline{W}_{ex} = -fC\omega\,\overline{\cos\omega t\sin(\omega t + \varphi)},$$

here the bar stands for averaging (see Section 7.8). Note that

$$\cos\omega t\sin(\omega t + \varphi) = \frac{1}{2}\sin 2\omega t\cos\varphi + \cos^2\omega t\sin\varphi,$$

and so

$$\overline{\cos\omega t\sin(\omega t + \varphi)} = \frac{1}{2}\overline{\sin 2\omega t}\cos\varphi + \overline{\cos^2\omega t}\sin\varphi = \frac{1}{2}\sin\varphi,$$

since $\overline{\sin 2\omega t} = 0$ and $\overline{\cos^2\omega t} = 1/2$. Consequently,

$$\overline{W}_{ex} = -\frac{fC\omega}{2}\sin\varphi,$$

which can be written as

$$\overline{W}_{ex} = \frac{fC\omega}{2}\cos\left(\varphi + \frac{\pi}{2}\right). \qquad (10.6.13)$$

Now let us determine the average power of the force of friction. Since $F_{fr} = -hv$ it follows that

$$\overline{W}_{ex} = -\overline{hv^2}. \qquad (10.6.14)$$

But

$$\overline{v^2} = \overline{\left(\frac{dx}{dt}\right)^2} = C^2\omega^2\,\overline{\sin^2(\omega t + \varphi)} = \frac{C^2\omega^2}{2}$$

see (10.4.1)). Therefore (10.6.14) yields

$$\overline{W}_{ex} = -hC^2\omega^2/2.$$

Since the work of the external force goes to overcome friction, the average

powers of the external force and the force of friction must be equal:

$$\overline{W}_{\text{fr}} = \overline{W}_{\text{ex}}, \qquad (10.6.15)$$

that is,

$$\frac{fC\omega}{2}\cos\left(\varphi+\frac{\pi}{2}\right) = -h\frac{C^2\omega^2}{2},$$

or $f\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right| = hC\omega,$

whence

$$C = \frac{f}{h\omega}\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right|. \qquad (10.6.16)$$

The maximum possible amplitude (at resonance) results, as may be seen from (10.6.16), at $\cos(\varphi + \pi/2) = 1$, that is, at $\varphi = -\pi/2$, and this determines the initial phase angle φ. Here $\omega = \omega_0$ and $C = f/h\omega_0$. Hence, the solution of the equation of motion in the case of resonance is

$$x = \frac{f}{h\omega_0}\cos\left(\omega_0 t-\frac{\pi}{2}\right).$$

We again have the formula (10.6.12).

Let us return to formula (10.6.13). From this formula we see that at resonance $\overline{W}_{\text{ex}}$ has the largest value since at resonance $\cos(\varphi + \pi/2) = 1$. For this reason, in the case of resonance an external force develops maximum average power and, consequently, performs maximum work.

These arguments of an energy nature make it possible to determine the amplitude at resonance also in the case of a more complicated dependence of the force of friction on velocity. Let the force of friction be proportional to a (fixed) power of the absolute value of velocity:

$$F_{\text{fr}} = -hv\,|v|^{n-1}, \qquad (10.6.17)$$

where h is positive, and therefore the first factor ensures that the signs of v and F_{fr} are different. At $v > 0$ this formula yields $F_{\text{fr}} = -hv^n$, while at $v < 0$ we get $F_{\text{fr}} = h\,|v|^n$, that is, in all cases the force of friction is in opposition to velocity. Since we assume, as before, that the motion is on the whole determined by the external periodic force $F = f\cos\omega t$, that is, is of an oscillatory nature with frequency ω, or $x = C\cos(\omega t + \varphi)$, the average power of the external force is given by formula (10.6.13) for this case, too.

Let us determine $\overline{W}_{\text{fr}}$. The instantaneous value is $W_{\text{fr}} = vF_{\text{fr}} = -hv^2\,|v|^{n-1}$; since $v^2 = |v|^2$, we can write $W_{\text{fr}} = -h\,|v|^{n+1}$. Substituting here the value of v, we find that

$$W_{\text{fr}} = -hC^{n+1}\omega_0^{n+1}\,|\sin(\omega_0 t+\varphi)|^{n+1}. \qquad (10.6.18)$$

Using this, we get [10.11]

$$\overline{W}_{\text{fr}} = -hC^{n+1}\omega_0^{n+1}A, \quad A = \overline{|\sin(\omega_0 t+\varphi)|^{n+1}}.$$

The condition (10.6.15) yields $hC^{n+1}\omega_0^{n+1}A = \frac{1}{2}fC\omega_0\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right|$, whence $C = \sqrt[n]{\frac{f}{2hA\omega_0^n}\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right|}$. The maximum amplitude attained at resonance is

$$C = \frac{1}{\omega_0}\sqrt[n]{\frac{f}{2hA}}. \qquad (10.6.19)$$

A particular case of formula (10.6.19) for $n = 1$ (friction proportional to the first power of velocity) is the earlier-found formula $C = f/h\omega_0$. Note that at $n \neq 1$ the equation of motion becomes *nonlinear* and we cannot write a general solution to this equation that would be similar to the one found in Section 10.5—there is no way in which an exact solution can be found via elementary methods this case.

10.7 Combining Oscillations. Beats

Quite common is the situation when a body performs *several* oscillations simultaneously. Visible light, for instance, is the sum effect of many vibrational processes related to the various electron transitions in the atoms of the body emitting the light (see Chapter 12). Similarly, the sound vibrations that we perceive when listening to a symphony orchestra are the sum of vibrations of the air generated by each musical instrument in the orchestra—and the conductor's goal is to control the various instruments and hence all these vibrations to produce an effect in which all the vibrations merge into a beautiful melody. (The effect of purely random mixing of many sound vibrations is known to anybody who has listened to an orchestra tuning up before

[10.11] For reference we give values of A for a few n: $n \to 0$, $A \to 2/\pi \simeq 0.64$; $n = 1$, $A = 0.5$; $n = 2$, $A \simeq 4/3\pi \simeq 0.42$; and $n = 3$, $A = 3/8 = 0.375$.

a concert, each musician tuning his instrument without regard for the other instruments.) There are also many cases (sometimes important) when two or more harmonic oscillations are superimposed; more than that, almost any motion can be represented as a sum of many harmonic oscillations (see Section 10.9).

We start with the addition of *two* harmonic oscillations. The simplest case here is when the oscillations are of the *same* frequency, for instance, when we are considering the sound produced by two violins playing in unison. The result is a harmonic oscillation of the same frequency. Indeed, if the two oscillatory processes are described by the formulas

$$x_1 = A \cos \omega t + B \sin \omega t,$$
$$x_2 = C \cos \omega t + D \sin \omega t, \qquad (10.7.1)$$

then the sum $x = x_1 + x_2$ has the form

$$x = (A + C) \cos \omega t + (B + D) \sin \omega t. \qquad (10.7.2)$$

The relationship between oscillation (10.7.2) and oscillations (10.7.1) are most conveniently described as follows: if to oscillations (10.7.1) we relate points M and N with coordinates (A, B) and (C, D) (or vectors $\overrightarrow{OM}$ and $\overrightarrow{ON}$), then oscillation (10.7.2) has corresponding to it the vertex P of the parallelogram $OMPN$ (or the vector $\overrightarrow{OP} = \overrightarrow{OM} + \overrightarrow{ON}$; see Figure 10.7.1). The energy of oscillations (10.7.1) is then fixed by the quantities $(OM)^2 = A^2 + B^2$ and $(ON)^2 = C^2 + D^2$ (cf. (10.4.2) and (10.2.8a)), while the energy of the combined oscillation (10.7.2) is proportional to $(OP)^2$. If the initial phases in (10.7.1) are the same, the segments OM and ON point in the same direction and the length of the segment OP is equal to the *sum* of the lengths of segments OM and ON; the energy of oscillation (10.7.2) then proves to be much higher than the sum of the energies of oscillations (10.7.1), since the energy is specified by the *square* of the amplitude of oscillation (OM, ON, and OP); if the oscillations in (10.7.1) are the same, the energy of their sum will be four times the energy of each oscillation. If the initial phases in (10.7.1) are opposite (that is, differ by π), the segments OM and ON are in opposition; here the amplitude OP of the sum oscillation is equal to the *difference* of amplitudes OM and ON of the component oscillations. The oscillations may even "cancel out" if their amplitudes (or energies) are the same. But if the energies differ only by a small quantity and the phases of x_1 and x_2 differ by π, the sum oscillation will have an extremely low energy (see Exercise 10.7.1).

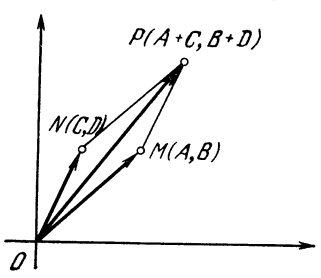


Figure 10.7.1

The case of addition of oscillations with *different* frequencies, when the sum oscillation is not harmonic, is more complicated. Let us consider, by way of an example, a mechanism known as the ***crankgear***, which transforms translational motion into rotation, and vice versa. Figure 10.7.2 shows such a mechanism. The crank OA is connected with the slide block B (a piston moving backward and forward inside a closed cylinder) by means of the connecting rod AB; the length of the crank will be denoted by R and the length of the connecting rod by L. Let us assume that the crank is in uniform rotation about the axis O with a frequency ω, so that the projection C of point A on the OB axis of the cylinder in which the slide block moves performs a harmonic oscillation about point O. The slide block B will also perform oscillations about

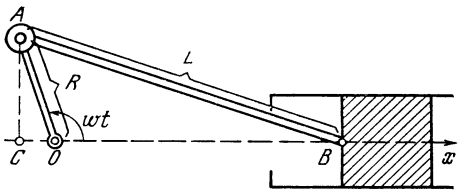


Figure 10.7.2

its midpoint, but these oscillations will not be harmonic.

Indeed, if $\angle AOB = \omega t$, then $OC = R\cos\omega t$, $AC = R\sin\omega t$, and, hence,

$$BC = \sqrt{L^2 - R^2\sin^2\omega t}$$
$$= L\sqrt{1-(R/L)^2\sin^2\omega t}.$$

Thus, we have

$$x = OB = R\cos\omega t + L\sqrt{1-(R/L)^2\sin^2\omega t}. \qquad (10.7.3)$$

But in the real mechanism R is usually much less than L, so that the ratio R/L is small (often of the order of 1/5, so that $(R/L)^2 \simeq 1/25$); for a small z we can replace $\sqrt{1-z}$ with $1 - z/2$ (see Section 6.4). For this reason the comparatively complicated formula (10.7.3) can be simplified:

$$x \simeq R\cos\omega t + L\left(1 - \frac{R^2}{2L^2}\sin^2\omega t\right)$$
$$= R\cos\omega t + L\left[1 - \frac{R^2}{2L^2}\frac{(1-\cos 2\omega t)}{2}\right]$$
$$= \left(L - \frac{R^2}{4L}\right) + R\cos\omega t + \frac{R^2}{4L}\cos 2\omega t. \qquad (10.7.4)$$

The first term $L - R^2/4L$ specifies the midpoint in the position of slide block B (the slide block oscillates about this midpoint), participating in two harmonic oscillations simultaneously, that is, in the oscillation with frequency ω and amplitude R fixed by the motion of point C, and in the oscillation with double frequency 2ω and amplitude $R^2/4L$ fixed by the change in altitude CA of point A (Figure 10.7.3).

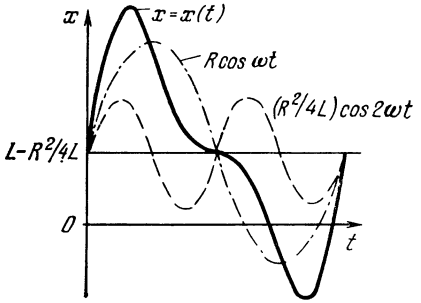


Figure 10.7.3

A peculiar situation arises when two oscillations with *close* frequencies ω_1 and ω_2 are combined. If, for istance, $x_1 = A\cos\omega_1 t$ and $x_2 = A\cos\omega_2 t$ (note that the amplitudes of both oscillations coincide), then

$$x = x_1 + x_2 = A\cos\omega_1 t + A\cos\omega_2 t$$
$$= A(\cos\omega_1 t + \cos\omega_2 t)$$
$$= A \times 2\cos\frac{\omega_1-\omega_2}{2}t\cos\frac{\omega_1+\omega_2}{2}t. \qquad (10.7.5)$$

Here, obviously, the factor $2A\times\cos\frac{\omega_1+\omega_2}{2}t$ changes very slowly, and the factor $\cos\frac{\omega_1+\omega_2}{2}t$ corresponds to oscillations with a frequency $\frac{\omega_1+\omega_2}{2} \simeq \omega$. Oscillations specified by (10.7.5) are known as ***beats*** (see Figure 10.7.4*a*); often they are described (not quite

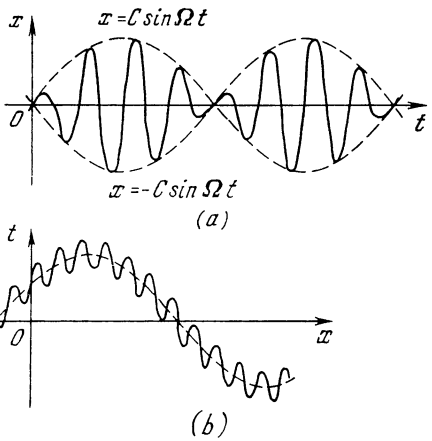


Figure 10.7.4

correctly) as harmonic oscillations with frequency $\omega \left(\simeq \frac{\omega_1 - \omega_2}{2}\right)$ and a slowly varying amplitude $A(t) = 2A \times \cos \frac{\omega_1 - \omega_2}{2} t$ $(= C \cos \Omega t)$. A similar situation arises when the phases of the oscillations x_1 and x_2 differ—only in this case the phase of the sum oscillation x also experiences slow oscillations (see Exercise 10.7.2a).

What happens if the frequencies ω and Ω of the component oscillations *differ considerably*? Suppose, $x_1 = A \times \cos \omega t$ and $x_2 = B \cos \Omega t$, with $\Omega \gg \omega$. Then the sum oscillation $x = x_1 + x_2$ can be represented in the form of a harmonic oscillation with amplitude B and frequency Ω about the slowly varying position of equilibrium, $a = a(t) = A \cos \omega t$ (see Figure 10.7.4*b*, where $A \gg B$).

We encounter the phenomenon of beats when we study ocean tides: the water level in oceans changes due to the attraction from the moon and the sun; in view of the earth's rotation about its axis, the level oscillates with a period of 12 hours. But since the moon orbits the earth, the period of the level oscillations caused by the moon's attraction differs somewhat from the period of the level oscillations caused by the sun's attraction: while the period of the "sun" tides is exactly 12 hours, the period of the "moon" tides (which are stronger) is close to $12\frac{3}{4}$ h. Since these oscillations have different amplitudes, the amplitude of the sum oscillation is never zero, that is, the sum oscillation does not coincide exactly with the beats depicted in Figure 10.7.4*a*; however, here also we can speak of a slow variation of the oscillation amplitude of the water level, the greatest height of the tide exceeding the lowest height by a factor of almost three (see Exercise 10.7.2b). Oscillations with a varying amplitude are also used in radiobroadcasting; for instance, the electromagnetic waves in the SW band with a wavelength of about 20 meters have a frequency of 1.5×10^7 Hz (or 15 MHz), while the amplitude of these oscillations varies within the audio-frequency range, that is, the range of frequencies to which the human ear is sensitive, approximately 15 to 20 000 Hz.

Exercises

10.7.1. What are the (a) energy and (b) amplitude of the oscillations obtained through combining the oscillations x_1 and x_2 with the same frequency: $x_1 = A_1 \cos(\omega t + \varphi_1)$ and $x_2 = A_2 \cos(\omega t + \varphi_2)$?

10.7.2. (a) What type of oscillations is the sum of two harmonic oscillations, $x_1 = a \cos(\omega_1 t + \varphi_1)$ and $x_2 = a \cos(\omega_2 t + \varphi_2)$, if ω_1 and ω_2 are close? (b) Answer the same question for oscillations x_1 and x_2 with different amplitudes, a_1 and a_2.

10.8 The Vibrations of a String

In Section 10.7 we studied the function $x = f(t)$ that was the sum of two harmonic oscillations (with the same or distinct frequencies). Next we will investigate the extremely important question of representing any function (continuous or even discontinuous) in the form of a sum of harmonic oscillations. Physicists and mathematicians arrived at this question in their studies of a wide range of physical phenomena. One of these deals with the well-known ***problem of a vibrating string.***

Let us consider in the xy-plane a string fixed at points $x = 0$ and $x = l$; the string is understood to be a thin thread that can bend freely and along which there is a constant tensile force acting. Let us disturb the equilibrium (a state in which the string lies along the x axis), that is, we shape it in a certain way specified by a function $y = f(x)$ (Figure 10.8.1). We then let the string go. The question is: how do the

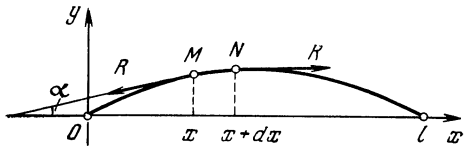


Figure 10.8.1

tensile forces change the shape of the string? We will restrict our discussion to *small* deformations, or strains, of the string, that is, we assume that y is small (the deflection of the string from the x axis is small) and dy/dx is small (the deformed string still closely resembles the undistorted string, so that the angle $\alpha = \arctan(\partial y/\partial x)$ between a tangent to the string and the x axis is small). We write $\partial y/\partial x$ instead of dy/dx because the deflection y of the string changes in time too, or $y = y(x, t)$ is a function of *two* variables, the abscissa x and the time t. We will now derive the differential equation that governs the behavior of $y(x, t)$.

Let us take a small segment of the string, MN, restricted by points $M(x, y)$ and $N(x + dx, y + dy)$. The length of this segment is obviously

$$ds = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2}$$

$$\simeq dx\left[1 + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right] \simeq dx$$

(cf. Sections 7.9 and 6.4), since we assume the derivative $\partial y/\partial x$ to be small and neglect terms of the order of $(\partial y/\partial x)^2$ and higher. Since we are considering string vibrations solely in the direction perpendicular to the x axis, the velocity v of point M is $\partial y/\partial t$ and the acceleration a coincides with the second (partial) derivative $\partial^2 y/\partial t^2$, while the mass of the segment MN of the string is $\sigma ds \simeq \sigma dx$, where σ is the (constant) linear density of the string (which is assumed homogeneous). The equation of motion of the segment MN of the string can be obtained if we equate the product $ma = \sigma\, dx\, (\partial^2 y/\partial t^2)$ to the force that acts on this segment.

The only type of forces that act on the string is a tensile force; in other words, here we are dealing only with *free* vibrations of the string and disregard all external forces that might act on it. On each end of the segment MN there acts a tensile force R that stretches the segment along the respective tangent. The components that are perpendicular to the x axis (and hence cause the string to vibrate in the transverse direction) are, respectively, $R \sin\alpha$ and $R \times \sin(\alpha + d\alpha)$, where α and $\alpha + d\alpha$ are the angles between the tangents to the string at points $M(x, y)$ and $N(x + dx, y + dy)$ and the x axis (see Figure 10.8.1). The resultant of these components is

$$R\sin(\alpha + d\alpha) - R\sin\alpha \simeq Rd((\sin\alpha)$$

(cf. Section 4.1). But we already know that $\tan\alpha = dy/dx$ and $\sin\alpha \simeq \tan\alpha \simeq \alpha$, so that we can write $\sin\alpha \simeq \tan\alpha \simeq \partial y/\partial x$ (here we again ignore quantities of the order of α^2 (or $\partial y/\partial x)^2$) or higher). For this reason, $d\sin\alpha \simeq d(\partial y/\partial x) = (\partial^2 y/\partial x^2)\, dx$, since for the time being t in our formulas is assumed to be fixed.

Thus, the force that causes the segment MN of the string to vibrate is close to $R(\partial^2 y/\partial x^2)\, dx$, and the equation of string vibrations can be written thus:

$$\sigma\frac{\partial^2 y}{\partial t^2}\, dx = R\frac{\partial^2 y}{\partial x^2}\, dx, \quad \text{or} \quad \frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}, \tag{10.8.1}$$

where $c^2 = R/\sigma$ (already a positive quantity). The solution to this equation must satisfy two *boundary conditions*

$$y(0, t) = 0 \text{ and } y(l, t) = 0 \tag{10.8.2a}$$

for all values of t (the endpoints of the string are fixed) and two *initial conditions*

$$y(x, 0) = f(x) \text{ and } \frac{\partial y(x, 0)}{\partial t} = 0 \tag{10.8.2b}$$

for all values of x (the last initial condition means that at $t = 0$ the string is motionless: we pull the string away from the position of equilibrium, and this is specified by the function $f(x)$, and then let it go [10.12]).

[10.12] Instead of (10.8.2b) we could require that the velocity $v(x, 0) = (\partial y/\partial t)_{t=0}$ of the string for all x's at initial time $t = 0$ be a given function, that is, $v(x, 0) = g(x)$, where $g(x)$ is a known function (this constitutes the *general* problem for a fixed string; see Exercise 10.8.2).

The vibrating string equation was first formulated and solved by Jean Le Rond d'Alembert in 1747 (see Exercise 10.8.3). But of more importance was the somewhat later solution of this equation carried out by Daniel Bernoulli in 1753, who reasoned as follows. Let us seek the particular solution to Eq. (10.8.1) that satisfies the boundary conditions (10.8.2a) and the second initial condition $(\partial y/\partial t) = 0$ at $t = 0$ in the form of a product $y(x, t) = X(x)\,T(t)$ of a function $X(x)$ that depends only on x and a function $T(t)$ that depends only on t. Substituting this product into Eq. (10.8.1) yields

$$XT'' = c^2X''T, \quad \text{or} \quad \frac{1}{c^2}\frac{T''}{T} = \frac{X''}{X}, \tag{10.8.3}$$

where, obviously, $X'' = d^2X/dx^2$ and $T'' = d^2T/dt^2$, since both X and T are functions of a single variable, x and t, respectively. But since $c^{-2}T''/T$ depends only on t and X''/X only on x, the fact that these two ratios are equal means that they are independent of x and t, that is,

$$\frac{X''}{X} = \frac{1}{c^2}\frac{T''}{T} = \text{constant}. \tag{10.8.4}$$

Clearly, if the ratios in (10.8.4) are *equal to zero*, then $X'' \equiv 0$ and $X = ax + b$, which is a *linear* function in variable x. But since, in view of (10.8.2a), $X(0) = X(l) = 0$ (because $y = X(x)\,T(t)$, where, of course, $T(t) \not\equiv 0$, since otherwise the particular solution is of no interest at all), the fact that $X(x) = ax + b$ means that $b = 0$ (since $X(0) = 0$) and $al = 0$, or $a = 0$ (since $X(l) = 0$). Thus, we have arrived at a zero solution to Eq. (10.8.1), which is of no interest to anyone.

If the ratios in (10.8.4) are *positive*, say, equal to k^2, then

$$\frac{X''}{X} = k^2, \quad \text{that is,} \quad X'' = k^2X. \tag{10.8.5}$$

This equation is in many ways similar to the equation of harmonic oscillations (10.2.1). Just as with Eq. (10.2.1), its particular solutions can easily be guessed: $X_1 = e^{kx}$ and $X_2 = e^{-kx}$. These solutions yield $X_1' = ke^{kx}$, $X_1'' = k^2e^{kx} = k^2X_1$, $X_2' = -ke^{-kx}$, and $X_2'' = (-k)^2e^{-kx} = k^2X_2$ (cf. Section 13.10). An arbitrary linear combination of these solutions,

$$X = aX_1 + bX_2 = ae^{kx} + be^{-kx},$$

is also a (general) solution to Eq. (10.8.5) that satisfies any boundary conditions: $x_0 = X(0) = A$, $X_0' = X'(0) = B$. (Why?) But the requirement that conditions (10.8.2a) be satisfied leads to the following:

$$X(0) = a + b = 0, \; X(l) = ae^{kl} + be^{-kl} = 0$$

that is, $b = -a$ and $a(e^{kl} - e^{-kl}) = 0$, which are satisfied only if $a = b = 0$. Thus, even the assumption that the ratios in (10.8.4) are positive does not lead to a solution to Eq. (10.8.1) that is not identically zero: $X(x) \equiv 0$, and this means that $y(x, t) \equiv 0$.

Finally, suppose the ratios we are considering here are *negative*, that is, $X''/X = c^{-2}T''/T = -k^2 < 0$, or

$$\frac{X''}{X} = -k^2, \quad X'' = -k^2X, \tag{10.8.6a}$$

$$\frac{1}{c^2}\frac{T''}{T} = -k^2, \quad T'' = -c^2k^2T. \tag{10.8.6b}$$

In contrast to the two previous cases, here we have meaningful solutions to (10.8.1).

Clearly, both (10.8.6a) and (10.8.6b) are equations of the (10.2.1) type of harmonic oscillations for the functions $X = X(x)$ and $T = T(t)$ (the equations differ from (10.2.1) only in notation). From (10.8.6a) it follows that

$$X = A\cos kx + B\sin kx \tag{10.8.7}$$

(cf. Section 10.2). Since $X(0) = 0$ in view of the first boundary condition in (10.8.2a), the coefficient A on the right-hand side of (10.8.7) is zero: $X = B \times \sin kt$. Substituting this X into the second boundary condition $X(l) = 0$ yields $B\sin(kl) = 0$, which for $B \neq 0$

(the case $A = B = 0$ is, of course, of no interest to us) is possible only if

$$kl = n\pi, \quad \text{or} \quad k = \frac{\pi}{l} n, \tag{10.8.8}$$

with n an integer. Thus, the final solution (10.8.7) to Eq. (10.8.6a) assumes the form

$$X = B \sin\left(\frac{n\pi}{l} x\right), \tag{10.8.9}$$

where B is arbitrary, and n is an arbitrary integer.

Similarly, the second equation of harmonic oscillations, (10.8.6b) (with k defined in (10.8.8)) admits of the solution

$$T = C \cos(ckt) + D \sin(ckt)$$
$$= C \cos\left(\frac{cn\pi}{l} t\right) + D \sin\left(\frac{cn\pi}{l} t\right). \tag{10.8.10}$$

Since

$$T' = \frac{dT}{dt} = -\frac{cn\pi}{l} C \sin\left(\frac{cn\pi}{l} t\right)$$
$$+ \frac{cn\pi}{l} D \cos\left(\frac{cn\pi}{l} t\right),$$

the second initial condition in (10.8.2b), $(\partial y/\partial t)_{t=0} = 0$, that is, $T'(0) = 0$, which we assumed to be satisfied, yields $D = 0$. Thus,

$$T = C \cos(cn\pi t/l)$$

and, hence,

$$y = XT = b_n \sin\left(\frac{n\pi}{l} x\right) \cos\left(\frac{cn\pi}{l} t\right) \tag{10.8.11}$$

where $b_n = BC$ is an arbitrary number, and n is a natural number (which must be selected to fix a particular solution). [10.13]

We have thus arrived at a whole family (10.8.11) of solutions to Eq. (10.8.1), since the natural number n in (10.8.11) can assume any value. All these solutions are harmonic oscillations: each fixed point of the string (a fixed value $x = x_0$) oscillates according to the simple law

$$y(x_0) = d_n \cos(\omega_n t) \tag{10.8.12}$$

with an arbitrary amplitude $d_n = d_n(x_0) = b_n \sin(n\pi x_0/l)$ but with a fixed frequency $\omega_n = (cn/l)\pi$. This value ω_n of the frequency depends not only on the physical characteristics of the string (its length l, the tensile force R, and the linear density σ, through the ratio $R/\sigma = c^2$) but also on an arbitrary (but fixed for a particular solution) natural number n: the various allowed frequencies

$$\omega = \omega_1 = \frac{c\pi}{l}, \quad \omega_2 = \frac{2c\pi}{l} = 2\omega,$$
$$\omega_3 = 3\omega, \ \ldots \tag{10.8.13}$$

form an arithmetic progression—all frequencies are multiples of the minimum frequency ω. The law of oscillations (10.8.11) states that all the segments of the string vibrate in unison: the frequencies ω_n do not depend on x, and the various segments differ only in the (smoothly varying) amplitudes $d_n = d_n(x) = b_n \sin(n\pi x/l)$.

We have therefore found the solution (10.8.11) (more precisely, a family of solutions (10.8.11)) to Eq. (10.8.1) satisfying the boundary conditions (10.8.2a) and the second initial condition in (10.8.2b). But how about the first initial condition? It is clear that, generally speaking, we have no control over this condition: although we have the arbitrary constant b_n which can be chosen at will, it is obvious that the initial shape of the string, $y(x, 0) = b_n \sin(n\pi x/l)$, is a sinusoid rather than an arbitrary curve (with a half-period that is a multiple of l; see Figure 10.8.2, where $n = 3$).

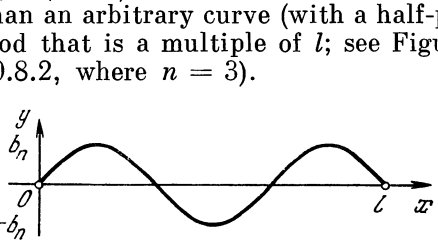


Figure 10.8.2

[10.13] The reader will recall that previously n was assumed to be an integer and not only a natural number (i.e. a positive integer). It is clear, however, that a change of sign of n in (10.8.11) leads only to a change of sign of b_n and yields no new solutions to Eq. (10.8.1), while $n = 0$ corresponds to the trivial zero solution $y \equiv 0$ of the equation.

So how should we proceed if the initial shape of the string, $y(x, 0) = f(x)$, is not a sinusoid? The brilliant idea of D. Bernoulli (which was developed in the 19th century by J.B.J. Fourier into a powerful method for solving various problems in mathematical physics; see Section 10.9) consisted in reducing an *arbitrary* vibration (a seemingly complicated function of x and t) of a string to a system of simple harmonic oscillations (or "pendulums") (10.8.11), namely, D. Bernoulli suggested employing the ***superposition principle***, that is, the fact that any linear combination of solutions to Eq. (10.8.1) is also a solution.

Let us set up the sum

$$y = b_1 \sin\left(\frac{\pi x}{l}\right) \cos\left(\frac{c\pi t}{l}\right) + b_2 \sin\left(\frac{2\pi x}{l}\right) \cos\left(\frac{2c\pi t}{l}\right) + \ldots \tag{10.8.14}$$

of the particular solutions (10.8.11) to Eq. (10.8.1). Clearly, this sum also satisfies Eq. (10.8.1), both boundary conditions in (10.8.2a) and the second initial condition in (10.8.2b), since each *term* in this sum satisfies the equation and the specified initial and boundary conditions.

Substituting $t = 0$ into (10.8.14) and allowing for the first initial condition in (10.8.2b), we get

$$y(x, 0) = f(x) = b_1 \sin\left(\frac{\pi x}{l}\right) + b_2 \sin\left(\frac{2\pi x}{l}\right) + b_3 \sin\left(\frac{3\pi x}{l}\right) + \ldots, \tag{10.8.15}$$

thus reducing the initial problem to determining from (10.8.15) the coefficients $b_1, b_2, b_3, \ldots$ in the representation (10.8.14) of a solution to Eq. (10.8.1). If this were to be done, it would be a remarkable achievement, since it would mean that a "continuous" vibration $y = y(x, t)$ of a string can be constructed from a discrete set (10.8.11) of separate harmonic oscillations ("pendulums") and that the general equation

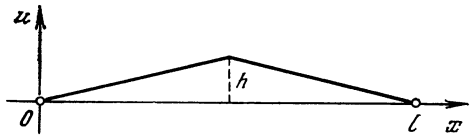


Figure 10.8.3

(10.8.1) can be reduced, via a method of separation of variables x and t, to ordinary (i.e. not containing partial derivatives) and simple (i.e. represented by elementary functions) equations (10.8.6a) and (10.8.6b). Of course, the representation (10.8.14) of the solution of Eq. (10.8.1) is of value only if it reduces to a finite (and preferably few) number of terms on the right-hand side of (10.8.14) and yields a complete description of the sought function $y = y(x, t)$.

Bernoulli put a lot of effort into the study of the theory of vibrations and was well-acquainted with the process of expansion of an arbitrary vibration into separate harmonics (i.e. simple harmonic oscillations): $f(t) = \sum_k c_k \cos(\omega_k t + \varphi_k)$. Using his experience as a base, he assumed that *any* function $f(x)$ can be represented in the form of a sum (10.8.15) of harmonics, that is, that formula (10.8.14) yields the *general solution* to the equation of string vibrations (10.8.1); the coefficients $b_1, b_2, b_3, \ldots$ in expansion (10.8.14) are found from condition (10.8.15). This statement was severely criticized by L. Euler (see p. 000). Euler argued that in the case of a simple initial function $f(x)$, say, the one depicted in Figure 10.8.3, where we pull the string away from the x axis at the string's middle, $x = l/2$, to a fixed (but small) height h and then let it go, the function $f(x)$ is given by *different* formulas in the two halves of the string:

$$f(x) = \begin{cases} \dfrac{2hx}{l} & \text{if } x \leqslant l/2, \\ -\dfrac{2hx}{l} + 2h & \text{if } x \geqslant l/2, \end{cases} \tag{10.8.16}$$

while formula (10.8.15) gives a single expression for $f(x)$ on the entire range from 0 to l. D'Alembert, who agreed with Euler, also thought it impossible to represent an arbitrary function by a single formula of the (10.8.15) type (the example (10.8.16) of a case that seems to contradict Bernoulli's statement was first suggested by d'Alembert). But Bernoulli did not yield a bit: his experience in mechanical phenomena told him that any function $y(x, 0)$ could be represented in the form (10.8.15). This debate played an extremely important role in clarifying the very notion of a function and in representing functions by trigonometric series, that is, sums of harmonics (or harmonic oscillations) given by formula (10.8.15) (see Section 10.9).

Exercises

10.8.1. Suppose $l = \pi$, $c = 1$, and $f(x) = \sin^3 x$. Find the solution (10.8.14) to Eq. (10.8.1) that satisfies the conditions (10.8.2a) and (10.8.2b).

10.8.2. How will D. Bernoulli's solution of Eq. (10.8.1) change if the initial conditions of Eq. (10.8.1) have the form $y(x, 0) = f(x)$ and $(\partial y(x, t)/\partial t)_{t=0} = g(x)$ (the initial shape of the string, $y(x, 0) = f(x)$, and the initial velocity, $v(x, 0) = (\partial y(x, t)/\partial t)_{t=0} = g(x)$, are specified, and then the string is released)?

10.8.3. (d'Alembert's solution of Eq. (10.8.1)). Prove that every function $y(x, t) = \varphi(x + ct) + \psi(x - ct)$, where φ and ψ are arbitrary functions of one variable, satisfies Eq. (10.8.1). How must φ and ψ be chosen so that the solution $y(x, t)$ satisfies the boundary conditions (10.8.2a) and the initial conditions (10.8.2b)? What happens if the boundary conditions are those specified in (10.8.2a) but the initial conditions are $y(x, 0) = f(x)$ and $v(x, 0) = g(x)$ (cf. Exercise 10.8.2)?

10.9 Harmonic Analysis. Fourier Series

Suppose that we have an arbitrary function fixed on a finite interval; it will be convenient to assume that this function, $y = f(x)$, is fixed on the interval $-\pi \leqslant x \leqslant \pi$ (see, however, p. 383 and Exercise 10.9.2). The problem consists in finding the coefficients in the expansion of $f(x)$ in a *trigonometric series*, that is, representing $f(x)$ in the form of a sum of harmonics 1, $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, $\cos 3x$, $\sin 3x$, etc., with appropriate numerical coefficients. To this end we write

$$f(x) = \frac{a_0}{2} + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + a_3 \cos 3x + b_3 \sin 3x + \ldots, \tag{10.9.1}$$

where the numbers $a_0/2$, a_1, b_1, a_2, b_2, ... have yet to be found (the term in (10.9.1) that is independent of x is written in the form $a_0/2$, instead of a_0, only for the sake of convenience).

First let us note the property of *orthogonality* of trigonometric functions:

$$\int_{-\pi}^{\pi} \cos mx \cos nx \, dx = \int_{-\pi}^{\pi} \sin mx \sin nx \, dx = 0$$

for all natural numbers m and n, with $m \neq n$;

$$\int_{-\pi}^{\pi} \cos mx \sin nx \, dx = 0 \tag{10.9.2}$$

for all natural numbers m and n; and

$$\int_{-\pi}^{\pi} \cos^2 mx \, dx = \int_{-\pi}^{\pi} \sin^2 mx \, dx = \pi$$

for all natural numbers m. These formulas follow from the well-known fact that

$$\cos mx \cos nx = \frac{1}{2}[\cos(m-n)x + \cos(m+n)x],$$

$$\sin mx \sin nx = \frac{1}{2}[\cos(m-n)x - \cos(m+n)x], \tag{10.9.3}$$

$$\cos mx \sin nx = \frac{1}{2}[\sin(m+n)x - \sin(m-n)x];$$

particular cases of the first two formulas in (10.9.3) corresponding to $m = n$ are the well-known relationships

$$\cos^2 mx = \frac{1}{2}(1 + \cos 2mx),$$
$$\sin^2 mx = \frac{1}{2}(1 - \cos 2mx) \qquad (10.9.3a)$$

(since $\cos 0 = 1$). To verify (10.9.2), we need only integrate from $-\pi$ to π the right-hand sides of all formulas (10.9.3) and (10.9.3a) employing the rules discussed in Chapter 5.

Suppose that we already have the representation of $y = f(x)$ in the form (10.9.1). The function $f(x)$ is assumed known; we wish to find the coefficients a_0, a_1, b_1, a_2, b_2, etc. in the trigonometric series. To find a_m, we multiply both sides of (10.9.1) by $\cos mx$ and integrate the products from $-\pi$ to π. On the left we have $\int_{-\pi}^{\pi} f(x) \cos mx\, dx$; on the right all terms will be equal to zero by virtue of (10.9.2) except

$$\int_{-\pi}^{\pi} a_m \cos^2 mx\, dx = a_m \int_{-\pi}^{\pi} \cos^2 mx\, dx = a_m \pi, \qquad (10.9.4)$$

since all the other terms in the right-hand side of (10.9.1) after integration transform into

$$\int_{-\pi}^{\pi} a_n \cos nx \cos mx\, dx = a_n \int_{-\pi}^{\pi} \cos nx \cos mx\, dx = 0,$$
$$\int_{-\pi}^{\pi} b_n \sin nx \cos mx\, dx = b_n \int_{-\pi}^{\pi} \sin nx \cos mx\, dx = 0 \qquad (10.9.5)$$

(at $n = 0$ the first formula in (10.9.5) becomes $\int_{-\pi}^{\pi} \frac{a_0}{2} \cos mx\, dx = \frac{a_0}{2} \times \int_{-\pi}^{\pi} \cos mx\, dx = 0$). The final result is

$$\int_{-\pi}^{\pi} f(x) \cos mx\, dx = a_m \pi, \text{ or}$$
$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos mx\, dx,$$
$$m = 1, 2, 3, \ldots \qquad (10.9.6)$$

Similarly, multiplying both sides of (10.9.1) by $\sin mx$ and integrating from $-\pi$ to π, we get

$$\int_{-\pi}^{\pi} f(x) \sin mx\, dx = b_m \pi,$$

or

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin mx\, dx,$$
$$m = 1, 2, 3, \ldots \qquad (10.9.6a)$$

Finally, simply integrating both sides of (10.9.1) from $-\pi$ to π and employing the fact that

$$\int_{-\pi}^{\pi} \cos nx\, dx = \int_{-\pi}^{\pi} \sin nx\, dx = 0$$

and $\int_{-\pi}^{\pi} dx = 2\pi$

yields

$$\int_{-\pi}^{\pi} f(x)\, dx = \int_{-\pi}^{\pi} \frac{a_0}{2}\, dx = \frac{a_0}{2} \times 2\pi,$$

or

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, dx, \qquad (10.9.7)$$

which is a particular case of formula (10.9.6) with $m = 0$.[10.14]

The formulas for the coefficients of a trigonometric series simplify when the function in question is even or odd. For an *even* function, $f(-x) = f(x)$,

$$\int_{-\pi}^{\pi} f(x) \sin mx\, dx = \int_{-\pi}^{0} f(x) \sin mx\, dx$$

[10.14] It is precisely for this reason we wrote $a_0/2$ in the right-hand side of (10.9.1) instead of a_0.

$$+\int_0^{\pi} f(x)\sin mx\,dx$$

$$=\int_{-\pi}^{0} f(-x)\sin(-mx)\,d(-x)$$

$$+\int_0^{\pi} f(x)\sin mx\,dx=\int_{\pi}^{0} f(x)\sin mx\,dx$$

$$+\int_0^{\pi} f(x)\sin mx\,dx=0,$$

since $f(-x)=f(x)$, $\sin(-mx)=-\sin x$, and $d(-x)=-dx$. Therefore, for an even function we have

$$f(x)=\frac{a_0}{2}+a_1\cos x+a_2\cos 2x+a_3\cos 3x+\ldots, \qquad (10.9.8)$$

where

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx=\frac{2}{\pi}\int_0^{\pi} f(x)\,dx,$$

$$a_m=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos mx\,dx$$

$$=\frac{1}{\pi}\left(\int_{-\pi}^{0} f(x)\cos mx\,dx+\int_0^{\pi} f(x)\cos mx\,dx\right)$$

$$=\frac{2}{\pi}\int_0^{\pi} f(x)\cos mx\,dx. \qquad (10.9.9)$$

In a similar manner, if we are dealing with an *odd* function, $f(-x)=-f(x)$, then

$$a_m=0 \quad\text{and}\quad b_m=\frac{2}{\pi}\int_0^{\pi} f(x)\sin mx\,dx. \qquad (10.9.10)$$

(Prove this!). The result is

$$f(x)=b_1\sin x+b_2\sin 2x+b_3\sin 3x+\ldots \qquad (10.9.11)$$

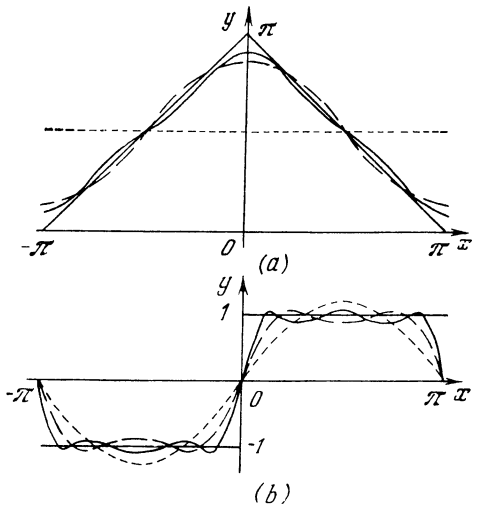


Figure 10.9.1

Thus, even functions can be expanded in a trigonometric series (10.9.1) containing only cosines, while odd functions can be expanded in a trigonometric series containing only sines.

Example 1. Suppose that

$$f(x)=\begin{cases}\pi+x & \text{for } -\pi\leqslant x\leqslant 0,\\ \pi-x & \text{for } 0\leqslant x\leqslant\pi\end{cases}$$

(Figure 10.9.1*a*; cf. Figure 10.8.3). This function is *even*, and therefore can be expanded in a trigonometric series containing only cosines, with coefficients

$$a_0=\frac{2}{\pi}\int_0^{\pi} f(x)\,dx=\frac{2}{\pi}\int_0^{\pi}(\pi-x)\,dx$$

$$=\frac{2}{\pi}\left(\pi x-\frac{x^2}{2}\right)\Big|_0^{\pi}=\frac{2}{\pi}\left(\pi^2-\frac{\pi^2}{2}\right)=\pi,$$

$$a_m=\frac{2}{\pi}\int_0^{\pi}(\pi-x)\cos mx\,dx$$

$$=\frac{2}{\pi}\left(\pi\int_0^{\pi}\cos mx\,dx-\int_0^{\pi} x\cos mx\,dx\right)$$

$$=\frac{2}{\pi}\left[\pi\int_0^{\pi}\cos mx\,dx\right.$$

$$-\frac{1}{m}\int_0^{\pi} x\,d(\sin mx)\Big] = \frac{2}{\pi}\Big[\pi\int_0^{\pi}\cos mx\,dx - \frac{1}{m}(x\sin m)\Big|_0^{\pi} + \frac{1}{m}\int_0^{\pi}\sin mx\,dx\Big] = \frac{2}{\pi}\left(\frac{\pi}{m}\sin mx\Big|_0^{\pi} - \frac{x}{m}\sin mx\Big|_0^{\pi} - \frac{1}{m^2}\cos mx\Big|_0^{\pi}\right)$$

$$= \begin{cases} 0 & \text{for } m \text{ even}, \\ \frac{4}{\pi m^2} & \text{for } m \text{ odd}. \end{cases}$$

The final result is

$$f(x) = \frac{\pi}{2} + \frac{4}{\pi}\cos x + \frac{4}{9\pi}\cos 3x + \frac{4}{25\pi}\cos 5x + \ldots . \qquad (10.9.12)$$

Figure 10.9.1*a* presents three functions that successively approximate $f(x)$, which is depicted by a heavy solid line: the function $\varphi_0(x) = \pi/2$ (the horizontal dotted line), $\varphi_1(x) = \pi/2 + (4/\pi)\cos x$ (the dashed curve), and $\varphi_2(x) = \pi/2 + (4/\pi)\cos x + (4/9\pi)\times \cos 3x$ (the thin solid curve).

Example 2. Suppose that

$$g(x) = \begin{cases} 1 \text{ for } & 0 < x \leqslant \pi, \\ -1 \text{ for } & -\pi \leqslant x < 0 \end{cases}$$

(Figure 10.9.1*b*). This is an *odd* function, and therefore it can be expanded into a trigonometric function containing only sines, with coefficients

$$b_m = \frac{2}{\pi}\int_0^{\pi}\sin mx\,dx = \frac{2}{m\pi}(-\cos mx)\Big|_0^{\pi} = \begin{cases} \frac{4}{m\pi} & \text{for } m \text{ odd}, \\ 0 & \text{for } m \text{ even}, \end{cases}$$

so that

$$g(x) = \frac{4}{\pi}\left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \ldots\right). \qquad (10.9.13)$$

Figure 10.9.1*b* presents three functions that successively approximate $g(x)$, which is depicted by two heavy solid lines: $\psi_1(x) = (4/\pi)\sin x$ (the dotted curve), $\psi_2(x) = (4/\pi)\times(\sin x + (\sin 3x)/3)$ (the dashed curve), and $\psi_3(x) = (4/\pi)(\sin x + (\sin 3x)/3 + (\sin 5x)/5)$ (the thin solid curve).

Up till now we spoke only of functions defined on the interval $-\pi \leqslant x \leqslant \pi$. But the case of a function defined on an arbitrary interval $-l \leqslant x \leqslant l$ introduces nothing new, since it is obvious that we can always introduce a new variable in such a way that the new interval will be reduced to the interval from $-\pi$ to π (cf. Section 1.7).

Let us introduce a new variable, $x_1 = (\pi/l)\,x$, or $x = (l/\pi)\,x_1$. Then $f(x)$ can be rewritten thus: $f(x_1 l/\pi) = f_1(x_1)$, where the function f_1 of the new independent variable $x_1 = (\pi/l)\,x$ is defined on the interval $-\pi \leqslant x_1 \leqslant \pi$. The expansion (10.9.1), in which x and $f(x)$ are replaced with x_1 and $f(x_1)$, generates the following expansion of the initial function $y = f(x)$:

$$y = \frac{a_0}{2} + a_1\cos\left(\frac{\pi}{l}x\right) + b_1\sin\left(\frac{\pi}{l}x\right) + a_2\cos\left(\frac{2\pi}{l}x\right) + b_2\sin\left(\frac{2\pi}{l}x\right) + a_3\cos\left(\frac{3\pi}{l}x\right) + b_3\sin\left(\frac{3\pi}{l}x\right) + \ldots, \qquad (10.9.14)$$

with

$$a_k = \frac{1}{l}\int_{-l}^{l} f_1(x)\cos\left(\frac{k\pi x}{l}\right)dx, \qquad b_k = \frac{1}{l}\int_{-l}^{l} f(x)\sin\left(\frac{k\pi x}{l}\right)dx \qquad (10.9.15)$$

(see Exercise 10.9.1).

Note also that a function $y = f(x)$ defined, say, in the interval $0 \leqslant x \leqslant l$ can always be continued onto negative values of x by setting $f(-x) = f(x)$, which results in the following representation for $f(x)$:

$$f(x) = \frac{a_0}{2} + a_1\cos\left(\frac{\pi}{l}x\right) + a_2\cos\left(\frac{2\pi}{l}x\right) + a_3\cos\left(\frac{3\pi}{l}x\right) + \ldots, \qquad (10.9.16)$$

where, obviously,

$$a_k = \frac{2}{l}\int_0^l f(x)\cos\left(\frac{k\pi x}{l}\right)dx,$$

$$k = 0,\ 1,\ 2,\ \ldots\ . \qquad (10.9.17)$$

(Why? See Exercise 10.9.1.) The function $f(x)$ can also be continued onto negative values of x by assuming that $f(-x) = -f(x)$. This leads to a sine expansion of $f(x)$:

$$f(x) = b_1 \sin\left(\frac{\pi}{l}x\right) + b_2 \sin\left(\frac{2\pi}{l}x\right) + b_3 \sin\left(\frac{3\pi}{l}x\right) + \ldots, \qquad (10.9.18)$$

where

$$b_k = \frac{2}{l}\int_0^l f(x)\sin\left(\frac{k\pi x}{l}\right)dx,$$

$$k = 1,\ 2,\ 3,\ \ldots\ . \qquad (10.9.19)$$

We encountered formula (10.9.18) when we studied string vibrations in Section 10.8 (see formula (10.8.15) for the function $y(x, 0) = f(x)$).

Note that the right-hand side of Eq. (10.9.1) can be understood as a function defined not only on the interval from $-\pi$ to π but on the *entire* number axis, since all trigonometric functions on the right-hand side of (10.9.1), $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, $\cos 3x$, $\sin 3x$, etc., are *periodic* with the least common period 2π: if $\varphi(x)$ is any one of the functions considered, then $\varphi(x + 2k\pi) = \varphi(x)$ for every integral k. For this reason the sum of the series on the right-hand side of (10.9.1) (we will, as usual, write this sum as $f(x)$) is a periodic function with a period of 2π, or $f(x + 2k\pi) = f(x)$ for all integrals k, since all terms on the right-hand side of (10.9.1) retain their values if we replace x with $x + 2k\pi$. Similarly, the right-hand side of the more general formula (10.9.14) defines a periodic function with a period of $2l$. It is even sometimes said that the formulas (10.9.1) and (10.9.14) *define* the expansion of *periodic* functions with periods 2π and $2l$, respectively. For instance, in Figure 10.9.2 we see the ("complete") functions $f(x)$ and $g(x)$ of Examples 1 and 2.

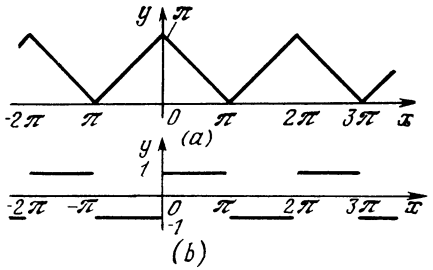


Figure 10.9.2

Of course, all the representations of functions in the form of infinite sums of trigonometric functions, that is, formulas (10.9.1), (10.9.8), (10.9.11), (10.9.14), (10.9.16), and (10.9.18), or, to use a term employed many times, in the form of *trigonometric series*, are valuable only if the respective series *converge*, if one is to use the terminology of Section 6.3. All this means that a sum of a finite number of terms (preferably, a small number of terms) in, say, the series (10.9.1),

$$s_n = s_n(x) = \frac{a_0}{2} + a_1\cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots + a_n \cos nx + b_n \sin nx, \qquad (10.9.20)$$

provides a fairly good approximation to the right-hand side of (10.9.1) and that, as $n \to \infty$, the sum s_n *tends* to $f(x)$, so that the greater the n, the more exact is the approximate equality

$$f(x) \simeq s_n(x). \qquad (10.9.21)$$

In the debate of D. Bernoulli with L. Euler and J. d'Alembert it was Bernoulli who was right: it has been established that practically any function (periodic or defined on a finite interval) can be expanded in a trigonometric series. [10.15] This assertion means

[10.15] For it to be impossible to expand a function $f(x)$ in a trigonometric series, $f(x)$ must have on the interval on which it is defined (or over one period if $f(x)$ is periodic) an infinite number of discontinuities (or jumps) or an infinite number of maxima and minima.

that, say, each function $f(x)$ that is defined on the interval from $-\pi$ to π (or, which is the same, each periodic function $f(x)$ with a period of 2π) can be represented in the form (10.9.1), that is, the approximate equality (10.9.21), with the right-hand side defined by (10.9.20), is valid. At points where the function $f(x)$ is continuous, the sums s_n are close to $f(x)$ (obviously, the greater the number n, the closer s_n is to $f(x)$), while at points where $f(x)$ has a discontinuity, or where it experiences a jump, so that its value on the left, $f(x-0)$, differs from the value on the right, $f(x+0)$, the sum s_n approaches, as $n \to \infty$, the *arithmetic mean* $(1/2)[f(x-0)+f(x+0)]$ of these values.[10.16] The rate of convergence of the sums $s_n(x)$ to the function $f(x)$ depends strongly on the behavior of the function: it is the highest in the case of smooth functions; the presence of discontinuities in the derivative of the function (see Example 1) reduces the rate of convergence, while if the function has discontinuities (as in Example 2), the rate is reduced still further, which requires using a large number of terms in the trigonometric series to represent the function with a given accuracy.

Let us explain, without laying claims to any mathematical rigor whatsoever, the reasons why the "degree of smoothness" of a function $f(x)$ affects the rate of convergence of the trigonometric series representing the function. It is clear that to be able to discard in the right-hand side of representation (10.9.1) all terms except a known (small) number we must ensure that the discarded terms rapidly decrease, so that even an infinite number of such terms has little effect on the result. And since the sine and cosine are bounded functions (neither can exceed unity in absolute value), the coefficients defined by (10.9.6) and (10.9.6a) must decrease rapidly as $m \to \infty$. But for smooth functions the decrease of these coefficients follows readily from the oscillatory (with alternating signs) nature of the functions $\cos mx$ and $\sin mx$ present in the integrands in the formulas for a_m and b_m.

Indeed, if m is high, the period of, say, $\sin mx$ proves to be so small that the (smooth) function $f(x)$ has no time to vary appreciably over the period. This means that the fraction of $\int f(x) \sin mx\, dx$ (which yields b_m) corresponding to one period of $\sin mx$ will be practically zero, since $f(x)$ may be assumed constant and the values of the integrals corresponding to the two half-periods of $\sin mx$ will almost completely cancel out (since they have opposite signs and their absolute values are practically the same).

A more exact calculation shows that, if m is high, the sum of all the fractions of the integrals in the right-hand sides of (10.9.6) and (10.9.6a) corresponding to separate waves of the cosine $y = \cos mx$ and the sine $y = \sin mx$ proves to be extremely small; this ensures the smallness of the expansion coefficients in (10.9.1) and, therefore, the convergence of the series in (10.9.1). If $f(x)$ is not smooth, that is, $f(x)$ suddenly changes its direction at certain points (see Figure 10.9.2*a*), the decrease in the expansion coefficients proves to be not so rapid, compared to the case where the function is smooth everywhere. The convergence is even worse if the function experiences a jump (see Figure 10.9.2*b*), in the vicinity of which the function $f(x)$ in no way can be considered a constant. Indeed, we see that the coefficients b_m corresponding to the discontinuous function $g(x)$ defined by (10.9.13) decrease only like $1/m$ as m increases, while in the case of the continuous function of Example 1 the coefficients a_m in (10.9.12) decrease, roughly speaking, like $1/m^2$ as $m \to \infty$; for a smooth function the rate of decrease of the coefficients in the trigonometric series corresponding to this function is even higher.

Let us take up once more the question of the famous debate between d'Alembert and Euler, on the one hand, and D. Bernoulli, on the other, concerning the vibrations of a string whose two ends are fixed. Particular solutions (10.8.11) of this problem were obtained in 1713-1715 by Brook Taylor (already mentioned earlier in this book), who assumed that no other solutions were possible (this assumption, however, lacked foundation). D. Bernoulli's achievement lay in the fact that he assumed that separate vibrations, (10.8.11), can mix, or that a string can generate tones of various frequencies simultaneously. The ***fundamental tone*** corresponds to

[10.16] For the function considered in Example 2 we must assume that $g(0-0) = -1$ and $g(0+0) = 1$; the sums $s_n(x)$ at $x = 0$ converge, as $n \to \infty$, to the value $(1/2)[(-1)+1] = 0$.

the first frequency ω, while the other frequencies, 2ω, 3ω, etc. (see 10.8.13)), correspond to ***overtones***, or higher harmonics.[10.17] It was the richness of the physical content of the string vibration problem provided by the representation (10.8.14) of the function $y(x, t)$ and the complete agreement of the results with the appropriate experimental data that made Bernoulli so sure of his results and enabled him to withstand the criticism of his friend Euler (whom he deeply respected) and of d'Alembert.

Further developments of Bernoulli's method are connected primarily with the name of J.B.J. Fourier, who was the first to give a coherent general theory for expanding functions in trigonometric series, a theory based on the simple formulas for the expansion coefficient obtained above. He also provided many examples of expansions of concrete functions. But even more important were his applications of such expansions to specific problems of mathematical physics, say, the problem of heat propagation. These applications were based on the same idea as the one put forward by Bernoulli: the initial conditions in the problem, conditions specified by an arbitrary function $f(x)$, were first assumed by Fourier to be sinusoidal, that is, he dealt only with functions of the sin mx or cos mx type. With such simple initial conditions solving the problem was relatively easy, while the general solution was then found via the superposition principle and the possibility of expanding $f(x)$ in a trigonometric series. And it appears quite natural that although many scientists before Fourier (Euler, D. Bernoulli, Gauss, and others) gave examples of expanding functions into trigonometric series, all such series are today called ***Fourier series***.

The procedure by which a function $f(x)$ is expanded in a Fourier series (or in a trigonometric series) is known as ***Fourier analysis***. For example, the expansion (10.9.12) of the function $f(x)$ of Example 1 shows that this function consists of a constant term $\pi/2$ and harmonics $\cos x$, $\cos 3x$, $\cos 5x$, etc. taken with coefficients $4/\pi$, $4/9\pi$, $4/25\pi$, etc., respectively, while the function $g(x)$ of Example 2 consists of harmonics $(4/\pi)\sin x$, $(4/3\pi)\sin 3x$, $(4/5\pi)\sin 5x$, etc. (see the expansion (10.9.13)).

Fourier analysis of (periodic) functions plays an extremely important role in mathematics and applications (for one, in the study of various oscillatory processes); it is often performed automatically via so-called ***Fourier analyzers***. Fourier analysis is sometimes called spectrum analysis, which of course stems from the fact that ordinary light consists of waves with definite wavelengths (or frequencies or periods), and a definite wavelength corresponds to a definite color in the spectrum. Generally, the expansion of an arbitrary (nonperiodic) function contains harmonics of the $a(\omega)\cos(\omega t + \varphi(\omega))$ type corresponding to the *entire range* of frequencies ω, so that we are forced to replace the Fourier series with the ***Fourier integral*** over this range (the case of a function having a continuous spectrum). Here, however, we will not dwell on this more complicated question.[10.18]

Exercises

10.9.1. Prove formulas (10.9.10), (10.9.15), (10.9.17), and (10.9.19).

10.9.2. Expand a function $y = f(x)$ defined in an arbitrary interval $a \leqslant x \leqslant b$ in a trigonometric series. [*Hint.* Substitution $x = \frac{b-a}{2\pi} x_1 + \frac{a+b}{2}$ transforms $f(x)$ into a function $f_1(x_1)$ defined on the interval $-\pi \leqslant x_1 \leqslant \pi$.]

10.9.3. Expand the following functions in trigonometric series: (a) $y = x^2$, $-\pi \leqslant x \leqslant \pi$; (b) $y = c_1 x$ for $-l \leqslant x \leqslant 0$ and $y = c_2 x$ for $0 \leqslant x \leqslant l$; and (c) $y = \sin x$ for $0 \leqslant x \leqslant \pi$ and $y = 0$ for $-\pi \leqslant x \leqslant 0$.

[10.17] Already in the fifth century B.C. the Pythagoreans of ancient Greece studied the relationships between the fundamental tone and the overtones of a vibrating string, which relationships reduce to simple ratios of the lengths of the periods of the corresponding sinusoidal waves, and the relation of these integral ratios to the euphony of the sounds produced by the string. The results obtained by the Pythagoreans played an extremely important role in the formation of their belief in the simplicity and cognoscibility of the world. They also considered mathematics the key to studying nature.

[10.18] The interested reader is referred to the book [15].

Chapter 11 The Thermal Motion of Molecules. The Distribution of Air Density in the Atmosphere

11.1 The Condition for Equilibrium in the Atmosphere

Let us consider the question of the law of distribution of air density in the atmosphere in altitude. It is common knowledge that at high altitudes the air is less dense and the air pressure is lower than at sea level. The reason for the dependence of pressure upon altitude is obvious. Let us mentally select a cylindrical volume (altitude Δh, base area S, volume $S\Delta h$). The air in this volume (mean density $\bar{\rho}$, mass $\bar{\rho}S\Delta h$) is attracted to the earth, that is, experiences the force of gravity directed downward and equal to $mg = \bar{\rho}Sg\Delta h$. However, this volume does not fall and is in a state of rest for the reason that at altitude h it is acted upon from below by a pressure $p\ (h)$ that exceeds the pressure from above at altitude $h + \Delta h$, which pressure is equal to $p\ (h + \Delta h)$ (Figure 11.1.1). The pressure on the lower base of the cylinder is $Sp\ (h)$; it balances the sum of the pressure on the upper base and the force of gravity:

$$Sp\ (h) = Sp\ (h + \Delta h) + \bar{\rho}Sg\Delta h. \tag{11.1.1}$$

Formula (11.1.1) can be rewritten as

$$p\ (h) - p\ (h + \Delta h) = \bar{\rho}g\Delta h. \tag{11.1.2}$$

We will assume that Δh is very small.

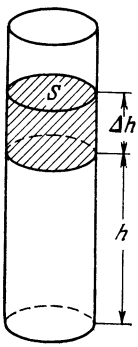


Figure 11.1.1

Then instead of $\bar{\rho}$ we can simply speak of the density ρ, since the altitudes h and $h + \Delta h$ are almost the same and $\bar{\rho}$ differs very little from the density $\rho\ (h)$ at altitude h. Therefore, assuming that on the left of (11.1.2) we have $-dp$ and replacing Δh with dh, we obtain

$$\frac{dp}{dh} = -g\rho. \tag{11.1.3}$$

We have thus obtained a *differential equation* for the dependence of pressure $p = p\ (h)$ on height. This equation also involves the air density ρ.

We will assume that the temperature of the atmosphere is the same at all altitudes. Actually, the air temperature depends on the heat flux from the sun and the removal of heat due mainly to heat radiation by the air into outer space or, to be more exact, by the water vapor and carbon dioxide in the air. A small portion of the solar radiation is absorbed by the upper rarefied layers of the air, while the larger portion reaches the earth, heats the ground, and then the ground heats the air. This explains the actually rather complicated distribution of temperature in the atmosphere: at ground level the temperature is known to fluctuate roughly between −40 °C to +40 °C depending on the geographical location and time of year; at an altitude of about 15 km the temperature is minimal (about −80 °C) and is approximately the same both in summer and winter round the globe. At considerable altitudes the air temperature increases, reaching +60 °C to +75 °C at altitudes between 50 and 60 km.

Recent measurements by means of artificial earth satellites show that at altitudes of 300 to 1000 km the air density is low but still is greater than was earlier thought. As we will see later on, high air density indicates a very high temperature of the air in these upper

strata. What is more, a substantial portion of the molecules of oxygen and nitrogen break up at these altitudes into atoms, ions, and electrons.

If there were no influx of heat from without or any removal of heat, that is, if we were to consider a heat-insulated column of air, then the temperature throughout the column would eventually even out. Below we will consider precisely this kind of an idealized case of total equilibrium, both thermal and mechanical. Thermal equilibrium means that the temperature is everywhere the same and so there are no fluxes of heat (if the temperature differed at distinct points in the air column, the heat would move from the hotter points to the colder ones, with a resultant flow of heat). Mechanical equilibrium consists in the resultant of all the forces acting on any volume of air chosen in the atmosphere being equal to zero. Here we have to consider the force of gravity on the air in the volume and the pressure on the entire surface bounding the given volume.

For the pressure distribution that satisfies Eq. (11.1.3), the atmosphere can be in a state of rest.

Since we consider altitudes h small by comparison with the earth's radius, g (the acceleration of gravity) can be regarded as constant.

11.2 The Relationship Between Density and Pressure

Equation (11.1.3) contains two unknown quantities: the pressure p and the density of air, ρ. We must therefore start by establishing a relation between them.

By ***Boyle's law*** *the product of the pressure of a gas and the volume occupied by this gas is constant* for a given mass m_0 of the gas and for a given temperature: $pV = a$, where a is a constant. If the gas density is ρ, then $m_0 = V\rho$. Hence, $V = m_0/\rho$, and since $p = a/V$, we can write

$$p = b\rho, \qquad (11.2.1)$$

with $b = a/m_0$. Thus *the gas pressure is directly proportional to the density.*

It is easy to find the constant of proportionality for air at room temperature. We know that the air pressure at sea level, p_0, is roughly equal to 10^5 N/m^2 (= 10^5 Pa). The air density ρ_0 at pressure p_0 is equal to 1.3 kg/m^3, approximately.[11.1] Substituting p_0 and ρ_0 into (11.2.1), we get $p_0 = b\rho_0$, whence [11.2]

$$b \simeq 10^5/1.3 \simeq 7.7 \times 10^4 \text{ m}^2/\text{s}^2.$$

Later on we will need not only the numerical value of b for air at room temperature but also the general expression of the constant b for any gas and any temperature. To this end we take advantage of the ***ideal gas law*** (sometimes called the ***Clapeyron ideal gas law***)

$$pV = RT, \qquad (11.2.2)$$

where V is the volume occupied by one gram-molecule (or simply one mole) of gas, T is the absolute temperature (reckoned from absolute zero, -273.16 K)[11.3], and R is the ***molar gas constant*** (sometimes called the ***universal gas constant***). We know that at 0 °C (equal to approximately 273 K on the absolute scale) and at atmospheric pressure at sea level, $p_0 = 10^5$ Pa, one mole of gas occupies a volume equal to 22.4 liters, or 2.24×10^{-2} m^3 (***Avogadro's law***), whence $10^5 \times 2.24 \times 10^{-2} = 273\ R$, or

$$R \simeq 8.3 \text{ (N/m}^2\text{)} \cdot \text{m}^3/\text{mol} \cdot \text{K}$$
$$= 8.3 \text{ J/mol} \cdot \text{K}.$$

[11.1] This quantity can easily be found experimentally by weighing. A hermetically sealed vessel of known volume is weighed with and without air (a vacuum pump is used to evacuate the vessel).

[11.2] Observe that b has the dimensions of the *square of velocity*. Actually this quantity is closely connected with the velocity of molecules and sound: the square of the speed of sound is equal to $1.4b$ (we will not derive this relation).

[11.3] Ordinarily, temperatures on the absolute scale are measured in kelvins (symbolized K), after the English scientist Lord Kelvin: 20 °C = 293 K (read: "20 degrees Celsius is equal to 293 kelvins").

We denote the fractional molecular weight of the gas by M. For hydrogen we have $M = 2$, for helium $M = 4$, for nitrogen $M = 28$, and for air the mean value of M is 29.4. By definition, V contains M grams, or $M \times 10^{-3}$ kilograms, of substance. This means that the density ρ is connected with V by the relation

$$\rho = 10^{-3}\, M/V, \text{ or } V = 10^{-3}\, M/\rho.$$

Here density ρ is expressed in kg/m^3 and volume V in m^3. Substituting this expression for V into (9.2.2), we get

$$p = 10^3 \rho \frac{RT}{M}. \qquad (11.2.3)$$

Comparing this with (11.2.1), we find that

$$b = 10^3 \frac{RT}{M}. \qquad (11.2.4)$$

Finally, let us express the pressure in terms of the number of molecules n contained in a unit volume of gas. We know that one mole of any substance contains about 6×10^{23} molecules. This quantity is known as ***Avogadro's number*** and is denoted by A. Thus, the mass of one molecule (in kilograms) is

$$m = \frac{M}{A} \times 10^{-3} \simeq \frac{1}{6 \times 10^{26}} M. \qquad (11.2.5)$$

If one mole of gas occupies volume V, then the number of molecules per unit volume is $n = A/V$. The gas density $\rho = nm$. The ideal gas law (11.2.2) yields

$$p = n \frac{RT}{A} = nkT,$$

where k is the ***Boltzmann constant***:

$$k = \frac{R}{A} \simeq \frac{8.3}{6 \times 10^{23}} = 1.38 \times 10^{-23} \text{ J/K}.$$

The quantity R refers to a conventionally chosen amount of substance, one mole, and so the dimensions of R involve the mole. The quantity k refers to one molecule, and so k has the dimensions of J/K. The quantity kT has the dimensions of energy (J). In Section 11.4 it will be shown that in the atmosphere the quantity kT is equal to the mean potential energy of one molecule in the field of gravity at temperature T. The mean kinetic energy of translational motion of one molecule is (3/2) kT.

11.3 Density Distribution

From formula (11.2.1) we find $\rho = p/b$. Putting this ratio into the differential equation (11.1.3) for the air pressure, we get

$$\frac{dp}{dh} = -\frac{g}{b} p.$$

The solution to the equation is $p = Ce^{-(g/b)h}$, where C must be determined from the initial condition. Let $p = p_0$ at $h = 0$. Then

$$p = p_0 e^{-(g/b)h}. \qquad (11.3.1)$$

Dividing both sides of (11.3.1) by b, we get

$$\rho = \rho_0 e^{-(g/b)h}, \qquad (11.3.2)$$

where ρ_0 is the air density at $h = 0$ (at sea level). From formula (11.3.1) it is evident that at altitude $H = b/g$ above sea level the air pressure diminishes by a factor of e. Using H, we can rewrite (11.3.1) and (11.3.2) thus:

$$p = p_0 e^{-h/H}, \quad \rho = \rho_0 e^{-h/H}. \qquad (11.3.3)$$

Let us compute H using the formula $H = b/g$:

$$H \simeq (7.7 \times 10^4)/10$$
$$= 7.7 \times 10^3 \text{ (m}^2\text{/s}^2\text{)/(m/s}^2\text{)} = 7.7 \text{ km}.$$

(This value of H corresponds to a mean temperature around 20 °C)

If the altitude is increased in arithmetic progression, the pressure and density fall in geometric progression: $p = p_0$ and $\rho = \rho_0$ at $h = 0$; if $h = H$, then $p = p_0/e \simeq 0.368\ p_0$ and $\rho \simeq 0.368\rho_0$; if $h = 2H$, then $p = p_0/e^2 \simeq 0.135 p_0$ and $\rho \simeq 0.135\rho_0$; if $h = 3H$, then $p = p_0/e^3 \simeq 0.05\ p_0$ and $\rho \simeq 0.05\rho_0$; and so on.

We obtain a formula relating H and kT. We know that $H = b/g$. Using

(11.2.4) and (11.2.5), we arrive at

$$H = 10^3 \frac{RT}{Mg} = \frac{RT}{Amg}, \text{ or } H = \frac{kT}{mg}. \quad (11.3.4)$$

Knowing the dependence of density on altitude, we can express the total mass of air, m_a, in a column with base area of 1 m^2. Indeed,

$$m_a = \int_0^\infty \rho\, dh = \int_0^\infty \rho_0 e^{-h/H}\, dh.$$

Make the change of variable $z = h/H$. Then $dz = (1/H)\, dh$ and

$$m_a = \rho_0 H \int_0^\infty e^{-z}\, dz = -\rho_0 H e^{-z}\Big|_0^\infty = \rho_0 H.$$

Using the relation $m_a = \rho_0 H$ we can compute H once again (by way of a check). Since the atmospheric pressure at sea level is approximately 10^5 Pa = 10^5 N/m^2, the mass of air in a column with base area of 1 m^2 presses against the earth with a force of 10^5 N, that is, this mass is approximately $10^5/g \simeq 10^4$ kg. Thus, $m_a = 10^4$ kg/m^2. Knowing that $\rho_0 \simeq 1.3$ kg/m^3, we get

$$H = \frac{m_a}{\rho_0} = \frac{10^4}{1.3} \simeq 7.7 \times 10^3 \text{ m} = 7.7 \text{ km},$$

in accord with the earlier computation.

Finally, let us find the *mean altitude* $\bar{h}$ at which the air is located, that is, the *altitude of the center of gravity* of a vertical cylindrical column of air. So as not to introduce extra quantities, we consider a column of air with base area of 1 m^2, although it is clear that the altitude of the center of gravity is independent of the base area of the cylinder. At a height between h and $h + dh$ is a mass of $dm = \rho dh$. The mean altitude is

$$\bar{h} = \frac{1}{m_a} \int_0^\infty h\, dm = \int_0^\infty h\rho(h)\, dh \Big/ \int_0^\infty \rho(h)\, dh.$$

Let us find the integral in the numerator. Using the second formula in (11.3.3) and substituting z for h/H, that is $dh = Hdz$, we get

$$\int_0^\infty h\rho(h)\, dh = \int_0^\infty h\rho_0 e^{-h/H}\, dh$$

$$= \rho_0 H^2 \int_0^\infty z e^{-z}\, dz = \rho_0 H^2$$

(since via the formula of integration by parts, $\int_0^\infty z e^{-z}\, dz = -ze^{-z}\Big|_0^\infty + \int_0^\infty e^{-z}\, dz = (-ze^{-z} - e^{-z})\Big|_0^\infty = 1$; compare with formula (7.6.19)). The final result is

$$\bar{h} = \frac{\rho_0 H^2}{\rho_0 H} = H. \quad (11.3.5)$$

Thus, the altitude H at which the density and the pressure of the air diminish e-fold is at the same time the mean altitude at which the air is located.

A similar result was obtained earlier when we considered radioactive decay (Section 8.1): if the probability of decay is ω, with $dn/dt = -\omega n$ and $n = n_0 e^{-\omega t}$, then during time $\tau = 1/\omega$ the amount of radioactive substance decreases by a factor of e, and the mean lifetime of a radioactive atom is equal to the same quantity: $\bar{t} = \tau = 1/\omega$.

Remember that the simple dependence of density and pressure on altitude, (11.3.3,) refers to the case of a constant temperature. Actually, the distribution of density and pressure departs somewhat from formula (11.3.3) and depends on the time of year and other factors.

Exercises

11.3.1. Find the air pressure in a mine at the following depths: 1 km, 3 km, and 10 km.

11.3.2. Find the dependence of air pressure on altitude for air temperatures equal to −40 °C and +40 °C.

11.3.3. Suppose the air temperature varies with altitude by the law $dT/dh = -\alpha T_0$, where T_0 is the air temperature at the earth's surface and α is a constant coefficient. Find the air pressure as a function of altitude.

11.3.4. We know that under the conditions of exercise 11.3.3 the constant α is approximately equal to 0.037×10^{-5} cm^{-1}. Using the result of exercise 11.3.3, determine the air pressure in a mine at depths of 1 km, 3 km, and 10 km. The temperature at ground level is taken to be 0 °C. Compare the results with those of exercise 11.3.1.

11.4 The Molecular Kinetic Theory of Density Distribution

In the preceding sections we found the distribution of air density in altitude under the action of gravity in a state of equilibrium. We regarded the air as a continuous medium with a given dependence of pressure on density.

Now let us take that result and approach it from another angle, namely, the viewpoint of molecular theory. We will consider the separate molecules and their motion. The idea that matter consists of individual atoms was first expressed in ancient Greece. However, the motion of molecules and its connection with heat was first examined by the great Russian scholar M.V. Lomonosov, who is thus the founder of the molecular kinetic theory.

The gaseous state differs from the liquid and solid states in that in a gas the molecules may be regarded as independent and noninteracting. The motion of molecules in a gas is that of free flight by inertia. From time to time the molecules collide. Under ordinary conditions, such collisions occur with extreme frequency and the path lengths which molecules traverse between collisions are extremely small.

At atmospheric pressure and a temperature of 0 °C, 22.4 liters of gas comprise one mole of a substance, or 6×10^{23} molecules, while 1 m^3 of gas contains $n \simeq 6 \times 10^{23}/2.24 \times 10^{-2} \simeq 2.7 \times 10^{25}$ molecules.

For our crude purposes we will regard molecules as spheres of radius about 2×10^{-10} m.[11.4] Then for a collision between two molecules to take place it is necessary that the trajectory of the center of one molecule hit a target of radius 4×10^{-10} m about the center of the other molecule. The area of such a target is $\sigma = \pi r^2 \simeq 5 \times 10^{-19}$ m^2. This means that over a path length of 1 m given molecule will collide with all molecules whose centers lie in a cylinder with base area 5×10^{-19} m^2 and altitude 1 m. The volume of such a cylinder is equal to σm^3, and the number of molecules in it is $n\sigma$, where n is the number of molecules in 1 m^3.

Thus, a molecule experiences $n\sigma$ collisions over a path of 1 m. Therefore, the mean distance of free flight between collisions is

$$l = \frac{1}{n\sigma} \simeq 7.5 \times 10^{-8} \text{ m}.$$

This quantity is known as the ***mean free path.***

Because of collisions, a molecule moves in a polygonal line, but the volume of a cylinder formed from polygonal lines differs but little from that of a right cylinder and so our computations remain valid.

Actually, one has also to consider the motion of those molecules that are hit in collisions. It can be proved that this circumstance changes but slightly the mean free path of a molecule, reducing it by a factor of only 1.5, so that in what follows we will assume that $l \simeq 5 \times 10^{-8}$ m.

Molecules have velocities v of the order of 300 to 500 m/s. Hence, the mean free path time, that is, the mean time between collisions, is of the order of

$$\tau \simeq \frac{5 \times 10^{-8}}{4 \times 10^2} \simeq 10^{-10} \text{ s}.$$

At first glance, the quantities $l \simeq 5 \times 10^{-8}$ m and $\tau \simeq 10^{-10}$ s are extremely small. But they have to be compared with the size of a molecule, whose radius is $r \simeq 2 \times 10^{-10}$ m, and with the duration of the collision itself, which is less than $r/v \simeq 10^{-13}$ s. If we do that, it will be apparent that the

[11.4] In reality, diatomic molecules, say of oxygen or nitrogen, are more like pairs of merged spheres, something reminiscent of peanuts (two nuts to a shell).

molecules of a gas collide very rarely: at atmospheric pressure, the molecules of air spend 99.9% of the time in free flight and only 0.1% of the time in a state of collision.

Molecular collisions in a gas do not affect the pressure of the gas and do not influence the law of distribution of density of the gas in the atmosphere. Confirmation of this fact lies in Boyle's law and the ideal gas law. In Section 11.2 these laws are written as $p = nkT$.

The gas pressure depends on the number of molecules per unit volume, but the radius r of the molecules and their cross section σ do not enter into the formula. This means that the quantities r and σ cannot enter into the formula for the density distribution in altitude.

Let us rewrite the formula for density distribution (11.3.2) by expressing b in terms of molecular quantities. Since $b = 10^3\ RT/M = AkT/Am = kT/m$, we can write

$$\rho = \rho_0 e^{-gh/b} = \rho_0 e^{-mgh/kT} \qquad (11.4.1)$$

(compare with (11.3.3) and (11.3.4)).

Divide both sides of (11.4.1) by m, where m denotes the mass of one molecule. Note that $\rho/m = n$ is the number of molecules in unit volume at altitude h, while $\rho_0/m = n_0$ is the number of molecules in unit volume at sea level. Formula (11.4.1) assumes the form

$$n = n_0 e^{-mgh/kT}. \qquad (11.4.2)$$

The quantity mgh is the potential energy of a molecule of mass m located at altitude h if for zero we take the potential energy of a molecule at sea level, since the potential energy of a molecule at sea level, $u\ (0)$, can be chosen arbitrarily (see Section 9.2). Then $u\ (h) = u\ (0) + mgh$, whence $mgh = u\ (h) - u\ (0)$. Formula (11.4.2) can be rewritten thus:

$$n\ (h) = n\ (0)\ e^{-[u(h)-u(0)]/kT}.$$

This is the law of distribution in altitude of the number of molecules. We can write it like this:

$$n\ (h) = Be^{-u(h)/kT},$$

where B is a constant defined by the value of density at sea level ($h = 0$), $n\ (0) = Be^{-u(0)/kT}$.

A remarkable fact is that the density of molecules at a certain altitude is only dependent on the potential energy of the molecules at the given site: the mass m of a molecule, the acceleration g of gravity, and the altitude h entered into formula (11.4.2) in exactly the same combination (mgh) as they entered into the expression for the potential energy u.

Let us find the *mean value* of the potential energy of a molecule, or

$$\bar{u} = \overline{mgh} = mg\bar{h} = mgH$$

(see formula (11.3.5)). Using formula (11.3.4), we get

$$\bar{u} = mgH = mg\frac{kT}{mg} = kT.$$

Thus, the mean potential energy of one molecule is kT.

We have established that the distribution of air molecules in the atmosphere depends on the temperature and on the potential energy of the molecules. But for a given mean potential energy $\bar{u}$ equal to kT different molecules have different potential energies. In other words, to a given value of $\bar{u}$ there corresponds a definite *distribution* of molecules in potential energy. Part of the molecules, those below altitude H, have a potential energy less than kT. Let us find the ratio of the number of such molecules to the total number of molecules. This ratio is

$$\int_0^H n\,dh \Big/ \int_0^\infty n\,dh$$

$$= n_0 \int_0^H e^{-mgh/kT}\,dh \Big/ n_0 \int_0^\infty e^{-mgh/kT}\,dh.$$

Let us evaluate these integrals:

$$\int_0^H e^{-mgh/kT}\,dh = -\frac{kT}{mg} e^{-mgh/kT}\Big|_0^H$$

$$= \frac{kT}{mg}(1 - e^{-mgh/kT}) = \frac{kT}{mg}(1 - e^{-1}),$$

$$\int_0^\infty e^{-mgh/kT}\,dh = -\frac{kT}{mg}e^{-mgh/kT}\Big|_0^\infty = \frac{kT}{mg}.$$

Therefore

$$\int_0^H n\,dh \Big/ \int_0^\infty n\,dh = \frac{kT}{mg}(1 - e^{-1}) \Big/ \frac{kT}{mg}$$

$$= 1 - e^{-1} \simeq 0.63.$$

To summarize, then, 63% of all the molecules have a potential energy less than the mean value, while 37% have a potential energy exceeding the mean value. It is now easy to calculate that 14% of all molecules have a potential energy exceeding $2kT$, 5% of all molecules exceeding $3kT$, and so on. Generally speaking, the portion of molecules whose potential energy is greater than a given value u equals $e^{-u/kT}$. This conclusion is very general and refers to the kinetic energy of molecules, too, which we will study in the section below.

11.5 The Brownian Movement and Kinetic-Energy Distribution of Molecules

Over a hundred years ago, an English botanist Robert Brown observed the random movement of microscopic particles suspended in water or other liquid. Einstein advanced the idea that this movement of particles is due to their thermal agitation. From this the conclusion was drawn that, for one thing, the particles would not all lie on the bottom of a vessel but would be distributed in height by the same law as the distribution of molecules.

If a suspended particle has the shape of a sphere of diameter $d = 5 \times 10^{-7}$ m, then its volume is $\pi d^3/6 \simeq 6.5 \times 10^{-20}$ m^3, and with a density $\rho = 2 \times 10^3$ kg/m^3 the mass of a particle is close to 1.3×10^{-16} kg. We must also take into account the Archimedean principle. The mass of the water (density 10^3 kg/m^3) displaced by the particle is half the mass of the particle. Taking this fact into account, we can say that the force of gravity "acts" only on half of the particle's mass, or approximately 6.5×10^{-17} kg. At room temperature, $T = 17$ °C $= 290$ K, such particles are distributed in height (by formula (11.4.2)) in accordance with the law

$$n \simeq n_0 \exp\left(-\frac{6.5 \times 10^{-17} \times 9.8}{290 \times 1.38 \times 10^{-23}}\,h\right),$$

or $n = n_0 \exp(-1.6 \times 10^5\,h)$. Thus, the number of particles per unit volume decreases by a factor of e when the altitude is increased by $(1.6 \times 10^5)^{-1}$ m $\simeq 0.62 \times 10^{-5}$ m.

By observing the distribution, in altitude, of particles of known size and density, it is possible to obtain the Boltzmann constant k. On the other hand, the ideal gas law yields the magnitude of $R = kA$, after which we can find Avogadro's number. This work was carried out by Einstein and Perrin in 1903-1907 and ser ved as a crucial exper imental corroboration of the entire atomic-molecular theory and played a tremendous part in the development of physics.

There is a constant transformation of energy taking place when molecules move under the force of gravity: if a molecule is moving downward at a given time, potential energy is being converted into kinetic energy, while if a molecule is in motion upward, kinetic energy is being converted into potential energy. When a gas is in a state of equilibrium, that is, the pressure of the gas is balanced by gravity, the gas molecules are actually moving at random with high speeds. However, if we picture to ourselves a horizontal plane in the gas, the number of molecules passing through it in unit time upward is equal to the number of molecules passing through the plane in the downward direction, so that, on the average, the gas is at rest. In the equilibrium state, the transition of kinetic energy into potential energy and the transition of potential energy into kinetic energy

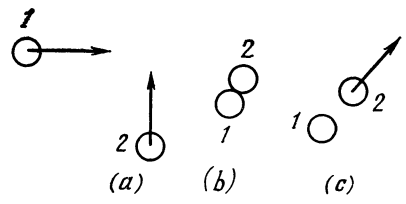


Figure 11.5.1

balance, since the number of molecules moving up equals the number moving down.

It is to be noted that in random motion the individual (identical) molecules have different velocities, or different kinetic energies. This is true because if two balls having identical speeds collide at an angle, the velocities of the balls after the collision may differ. Figure 11.5.1 illustrates a collision after which one of the balls (on the left) is brought to a halt, while the other one, moving upward, shoots off with double energy (Figure 11.5.1*a* depicts the balls prior to collision, Figure 11.5.1*b* in collision, and Figure 11.5.1*c* after collision). Note the positions of the balls at the instant of collision: if the second ball were, at collision, located below the first one, then it would stop and give up all its energy to the first ball.

Since in molecular motion there is a mutual conversion of kinetic and potential energy, it is natural to suppose that the distribution of the molecules as to kinetic energy is similar to that with respect to potential energy.

We give without proof the results of computations carried out at the end of the 19th century by Maxwell and Boltzmann. The number of molecules having velocity components along the x axis between v_x and $v_x + dv_x$, along the y axis between v_y and $v_y + dv_y$, and along the z axis between v_z and $v_z + dv_z$ is equal to

$$dn = \frac{n_0}{(2\pi kT/m)^{3/2}} \times \exp\left[-\frac{m(v_x^2+v_y^2+v_z^2)}{2kT}\right] dv_x\, dv_y\, dv_z, \qquad (11.5.1)$$

where n_0 is the total number of molecules, and m is the mass of one molecule. Observe that $v_x^2 + v_y^2 + v_z^2 = v^2$, where v is the speed of a molecule. Therefore, (11.5.1) has, in the exponent, the quantity $(mv^2/2)/kT$, which is the ratio of kinetic energy to potential energy. The mean kinetic energy calculated on the basis of (11.5.1) turned out to equal $(3/2)kT$. For the number of molecules n whose kinetic energy exceeds the given value E we have a rather unwieldy relationship. True, this complicated relationship can approximately be described by the simple formula

$$n \simeq n_0 e^{-E/kT}. \qquad (11.5.2)$$

This law yields an incorrect value for the mean kinetic energy of the molecules:

$$\overline{E}_{\text{kin}} = \frac{1}{n_0}\int_0^\infty n\, dE = \int_0^\infty e^{-E/kT}\, dE = kT$$

instead of $(3/2)\,kT$. It gives perceptible departures from the true value if E is of the order of kT. However, when $E \gg kT$, the divergence between the exact and the approximate law is not essential.

It will be noted that for the same temperature, molecules with different masses have the same mean kinetic energies and have the same distribution as to the magnitude of kinetic energy, since the mean velocity of a molecule is proportional to $1/\sqrt{m}$, where m is the mass of a molecule.

Considering the collisions of molecules against the walls of the containing vessel, we can find the gas pressure to be

$$p = \frac{2}{3} n_0 \overline{E}_{\text{kin}}.$$

Putting $\overline{E}_{\text{kin}} = (3/2)\,kT$, we get the ideal gas law

$$p = n_0 kT.$$

Mutual collisions of molecules give rise not only to an exchange of kinetic energy between molecules but also to a conversion of the kinetic energy

of motion of the molecules into the energy of rotation of a molecule and into the energy of vibrations of the atoms of the molecule, which is to say, into the internal energy of the molecule. The converse is also possible: in a collision, part of the internal energy of molecules is transformed into kinetic energy. It is therefore natural that the distribution of molecules as to their internal energy W also obeys the law of proportionality to the quantity $e^{-W/kT}$. The fact that the number of particles with a given energy is an exponential function of the energy is a universal law of nature.

11.6 Rates of Chemical Reactions

Of what use is the law of distribution of molecules as to kinetic energy? Such important characteristics of a gas as the pressure it exerts on the walls of the containing vessel, its heat capacity, and the total reserve of energy in the volume of the gas are defined by mean quantities, which is to say, they are defined by the bulk of the molecules whose energy is close to the mean value. For example, why do we have to know that a minute portion (of the order of 0.00001 %) of the molecules have kinetic energy exceeding 17 kT? These separate molecules with very large energies have practically no perceptible effect on the pressure and the general supply of energy of the gas.

However, the picture changes drastically if we consider *chemical reactions*. It turns out that precisely these rare molecules with high energy completely determine the course of chemical reactions. The mystery of chemical reactions stems from the fact that molecules entering into a reaction collide every 10^{-10} second, whereas a reaction frequently requires several minutes (sometimes hours). Which means that only an extremely small portion of all collisions result in a chemical reaction.

The idea was advanced that molecules have a certain very small "sensitive spot" that must be touched in order for a reaction to occur. This is reminiscent of the Greek hero Achilles who was vulnerable only in the heel.

A proper explanation was finally given at the end of the 19th century by the Swedish scientist Svanté ***Arrhenius***. It is this: reactions are initiated only by collisions of molecules whose energy exceeds a definite value, the so-called ***activation energy*** E_a.

For instance, when molecules of hydrogen and iodine collide, they form two molecules of hydrogen iodine HI, the energy of the colliding molecules must exceed 3×10^{-19} J. Compare this with the fact that at 0 °C the magnitude of kT is $1.38 \times 10^{-23} \times 273 \simeq 3.8 \times 10^{-21}$ J. This means that at room temperature (or at 0 °C, which is of the same order of magnitude) only a minute fraction of the molecules possess the needed energy, $\alpha = e^{-\nu}$, where $\nu = 3 \times 10^{-19}/3.8 \times 10^{-21} \simeq 80$, whence $\alpha \simeq 1/10^{35}$.

We get the reaction time by multiplying the time between collisions (it is of the order of 10^{-10} s) by the mean number of the collisions among which there will be one collision involving the required energy. This mean number of collisions is of the order of $1/\alpha \simeq 10^{35}$. We obtain the reaction time at 0 °C of the order of 10^{25} s $\simeq 3 \times 10^{17}$ years. This result accords with the fact that at 0 °C the reaction $H_2 + I_2 = 2HI$ is practically unobservable.

From the reasoning given above it follows that, depending on the temperature, the reaction time is expressed by the formula

$$t = \tau e^{E_a/kT},$$

where τ is the time between two collisions, and E_a is the activation energy. This formula gives a true picture of the dependence of the rate of chemical reactions on the temperature. A characteristic feature of this formula is the extremely sharp decrease in reaction time and increase in reaction rate for slight variations in temperature.

However, it frequently happens that chemical reactions are much more involved because they may proceed via diverse intermediate stages. By way of an illustration, let us examine the reaction

$$H_2 + Cl_2 = 2HCl.$$

This reaction proceeds not via collisions of molecules of hydrogen and a molecule of chlorine but by the scheme

$$H_2 + Cl_2 = H_2 + Cl + Cl,$$
$$Cl + H_2 = HCl + H,$$
$$H + Cl_2 = HCl + Cl.$$

As a result the actually observed reaction rate involves complicated relationships. However, for each separate reaction, say for

$Cl + H_2 = HCl + H$ (in the reaction $H_2 + Cl_2 = 2HCl$),

the Arrhenius law holds true, and the reaction rate is proportional to $e^{-E_a/kT}$, with the activation energy E_a having different values for each reaction.

The Soviet scientist Academician N.N. Semyonov made a thorough investigation of complex (chain) chemical reactions and elucidated the laws governing the course of such reactions and the general causes that lead to such complicated reaction schemes.

11.7 Evaporation. The Emission Current of a Cathode

The idea of Svanté Arrhenius concerning the role of a small number of molecules whose energy greatly exceeds the mean value of energy is helpful in analyzing not only chemical reactions but also a series of other phenomena including the ***evaporation of a liquid.***

Evaporation requires the expenditure of a considerable amount of energy. For example, the evaporation of 1 gram of water at 100°C requires the consumption of about 2260 J.[11.5] Per molecule, this comes out to[11.6] $Q = 18 \times 2260/6 \times 10^{23} \simeq 7 \times 10^{-20}$ J. But at $T = 0\ °C = 273$ K we have $kT \simeq 273 \times 1.38 \times 10^{-23}$ J $\simeq 3.8 \times 10^{-21}$ J, whence $Q/kT \simeq 20$. Only those molecules can tear away from the surface and evaporate whose energy exceeds the evaporation heat Q. The fraction of such molecules is equal to $e^{-Q/kT}$. Therefore, the rate of evaporation is also proportional to $e^{-Q/kT}$. For computational convenience it is common practice to multiply the numerator and the denominator of the expression Q/kT by Avogadro's number A:

$$\frac{Q}{kT} = \frac{QA}{kAT} = \frac{QA}{RT}.$$

The quantity QA is the evaporation heat of 6×10^{23} molecules, which is to say the evaporation heat of one mole of substance. The quantity $kA = R$ is the molar gas constant: $R \simeq 8.3$ J/mol·K (see Section 11.2). The evaporation heat of one mole of water is equal to $Q_m \simeq 18 \times 2260 \simeq 40\,000$ J/mol. Thus, the rate of the evaporation of water is proportional to $e^{-40000/8.3T} \simeq e^{-5000/T}$.

Let us consider the saturated vapor above a water surface. If the vapor is saturated, the number of molecules of water evaporating per unit time is equal to the number of molecules in the vapor and adhering to the surface of the water (condensing) in unit time. The rate of evaporation is $Ce^{-Q_m/RT}$, where C is a constant proportional to the area of the water surface. The rate of condensation is proportional to the pressure of water vapor and to the surface area. Hence, in the case of saturated vapor, when the rates of evaporation and condensation are equal,

$$Dp = Ce^{-Q_m/RT},$$

where D and C are quantities proportional to the surface area and only slightly dependent on the temperature

[11.5] The evaporation heat is but slightly dependent on temperature; for water $Q =$ 2260 J/g at 100 °C and 2500 J/g at 0 °C. We henceforth disregard this dependence.

[11.6] Water has a molecular weight of 18 and Avogadro's number is equal to 6×10^{23}.

and totally independent of the pressure, whence

$$p = Fe^{-Q_m/RT},$$

where the constant F does not depend on the surface area of the water. Thus a relationship is established between the pressure of saturated vapor and evaporation heat.

Let us consider yet another process similar to evaporation, that of the ***emission of electrons*** from a heated surface. This process occurs on the cathode of an electron tube. A cold cathode in vacuum does not emit electrons.[11.7] But at high temperatures the cathode does emit electrons. Then if the anode (also called plate) has a sufficiently high positive potential, it will attract the electrons and each electron torn out of the surface of the cathode will fall onto the anode. The electric current flowing in a circuit through an electron tube is equal to the product of the number of electrons released by the cathode in unit time into the magnitude of the charge of a single electron.

Experiments show that in these conditions the following relationship exists between current j and temperature T:

$$j = ge^{-Q/kT}.$$

The value of Q differs for different cathodes. For instance, for a cathode made of pure tungsten, $Q/k = 55\,000$ °C while for barium oxide, $Q/k = 30\,000$ °C and, hence, such a cathode can operate at lower temperatures. Using the dependence of j on T, we can determine Q/k. Here the quantity Q, which enters into the latter formula, coincides with the energy necessary for tearing an electron out of the cathode (this electron-ejection energy can be determined by other methods).

An electron tube offers a marvelous method for measuring the distribution of electrons leaving the surface of a cathode in accordance with their speeds at a given temperature. When the cathode is heated, we will apply a small negative potential φ to the anode. With this potential, the anode will repulse the electrons ejected by the cathode. For this reason, a large portion of the electrons will not reach the anode and will fall back onto the cathode. However, there will be some electrons that will reach the anode over the repulsive force. For this to occur, the kinetic energy of the electron ejected from the cathode must exceed the difference in potential energy of the anode and cathode, that is the quantity $\varepsilon\varphi$, where the Greek letter ε (epsilon) stands for the electron charge (unfortunately, the letter e stands here for the base of the natural logarithms). The fraction of such electrons is equal to $e^{-\varepsilon\varphi/kT}$. Thus, for a negative potenial φ across the anode, the current is equal to $j = j_0 e^{-\varepsilon\varphi/kT}$, where j_0 is the current for a positive potential. In this experiment, it is necessary that the distance between the cathode and anode be small so that the number of electrons between them should not be great and the mutual repulsion of electrons should not affect the result of the experiment.

The Soviet scientist Academician A.F. Ioffe proposed using this phenomenon for direct conversion of thermal energy into electric energy. If electrons go from the cathode to a negatively charged anode, such a system is a source of voltage: the current in an external circuit between the negatively charged anode and the positive cathode is in a direction such that it performs work. This method of obtaining electric current is remarkable in that there are no moving parts and the circuit is fundamentally simple. In this respect it resembles the generation of electric power by means of thermoelectric cells, which was also proposed by Academi-

[11.7] Here we do not consider the case of a very strong electric field (10^6 V/cm and more) capable of tearnig electrons even out of a cold cathode. Neither do we discuss the ejection of electrons from a cathode by the action of light or bombardment of a cathode by electrons, ions, or other particles.

cian Ioffe. At present semiconductors, which contain free electrons already at room temperature, have successfully replaced electron tubes in radio and TV sets and in electronic computers. But the heated cathode, which emits electrons according to the above-described law, is still widely used in cathode-ray tubes.

Chapter 12 Absorption and Emission of Light. Lasers

12.1 Absorption of Light: Statement of the Problem and a Rough Estimate

Let us consider the absorption of light in air containing black particles of soot. Suppose a unit volume contains N particles. The area of a section of one particle by a plane perpendicular to the ray of light is denoted by σ. For short, we call σ the ***cross section.*** For example, for a particle in the shape of a sphere of radius r, σ is the area of a cross section passing through the center of the sphere, or $\sigma = \pi r^2$. [12.1]

We will assume that the light incident on the surface of the particle of soot is completely absorbed. The problem consists in determining the portion of absorbed light and the portion of transmitted light as a function of the quantities N, σ, and the path length x that a light ray traverses through air containing the soot.

We begin with the roughest kind of estimate of the distance over which an appreciable portion of light is absorbed. We denote this distance by L. Just what the pregnant expression "appreciable portion of light" means will be examined later on in the sections that follow. Let us not be upset by the clumsy statement of the problem, since such a situation is quite common in physical problems where time and again refinement of a problem is postponed and is carried out in the process of solution.

Consider a cylinder with base area S and height L. We require that *the sum of the cross sections of all particles in this cylinder be equal to* S.

What is the physical meaning of the condition thus posed? If it were possible to arrange the particles so that the areas covered by various particles do not overlap, then using the particles in the cylinder of height L and base area S it would be possible to cover the *whole* area of the cylinder and achieve a complete absorption of all the light. For $x < L$, total absorption of the light is clearly impossible: no matter how the particles of soot are placed, the total area of their cross sections does not suffice to cover the whole base of the cylinder. In other words, there is not enough soot to ensure total absorption of light.

It is clear that at $x = L$ or even for $x > L$ there will not really be complete absorption. For a random arrangement of soot particles and for *arbitrary* x, there will remain certain directions along which there will not be a single particle in the path of the light, which will then pass through.

In the volume SL of the cylinder there are NSL particles, the sum of the cross sections of which is σNSL, and so we require that $\sigma NSL = S$, or

$$L = 1/\sigma N. \qquad (12.1.1)$$

Let us verify the dimensions in (12.1.1): σ is the area, so its dimensions are m^2, and N is the number of particles per unit volume and has the dimensions of m^{-3}. Consequently $[L] = m^{-2} \times m^{+3} = m$, as required.

The energy transmitted through an area in 1 second is called ***radiant flux.*** Let I be the radiant flux through an area of 1 square meter. This is called the ***energy flux***, or ***irradiance***, and its dimensions are $J \cdot m^{-2}s^{-1}$, or W/m^2. Below we consider the energy flux of light $I(x)$ as a function of the thickness x of a layer. Clearly,

$$I(x) = I_0 f(x), \qquad (12.1.2)$$

[12.1] An exact definition of the cross section σ for a particle of intricate shape is this: σ is the *mean* (*average*) area of the shadow cast by a particle on a surface perpendicular to a ray of light. The mean area is determined from measurements of the shadow area at different orientations of the soot particles in relation to the light ray, and the plane on which the shadow from the particle falls is assumed to be perpendicular to the ray in all measurements.

where I_0 is the energy flux of the incident light, and $f(x)$ is the desired function that characterizes the attenuation of the light.

What can we say about the properties of the function $f(x)$? If $x = 0$, there is no attenuation of light, $I(0) = I_0$, and so $f(0) = 1$. If $x > 0$, then the light is attenuated, $I(x) < I_0$, and therefore $f(x) < 1$. Clearly, $f(x)$ decreases with increasing x and approaches zero, that is, $f(x)$ is a *decreasing* function. Thus, its derivative is negative, $df/dx < 0$.

We have already said that there will not be complete absorption either for $x = L$ or for $x > L$, and so we do not expect $f(x)$ to vanish when $x = L$. However, we may assume that the value $x = L$ is a characteristic length. This means that when light is transmitted over a path $x \ll L$, the fraction of absorbed light is extremely small when compared with the fraction of transmitted light. Over a path $x \simeq L$ a perceptible portion of the light is absorbed, and over a path $x \gg L$, most of the light is absorbed and only a very small portion is transmitted.

As may be seen from formula (12.1.2), the function $f(x)$ is dimensionless. We can assume that if a dimensionless variable x/L is introduced, then the function $f(x/L)$ will always be the same for any kind of particles, for any N and σ. These suppositions will be corroborated and made precise in the sections that follow.

12.2 The Absorption Equation and Its Solution

We conduct all calculations for a column of air in the form of a cylinder with base area 1 m^2 (in the preceding section, when we considered a cylinder with base area S m^2, the quantity S canceled out anyway in (12.1.1)). We consider a thin layer of air between x and $x + dx$ in the cylinder. A beam of light consists of parallel rays and is characterized by the energy flux I. If no light were absorbed by the soot particles, I would be constant.

The layer under consideration contains $N\,dx$ particles covering an area of $\sigma N\,dx$ of the total area of 1 m^2 of the base of the layer (we assume that the shadows from these particle do not overlap, which is quite reasonable if dx is small). Hence, the layer absorbs a fraction $\sigma N\,dx$ of the energy incident on the layer: $dQ = IN\sigma dx$. When light passes through the layer dx, the radiant flux is attenuated by an amount equal to the quantity of absorbed energy dQ. Prior to entry into the layer, the energy flux was $I(x)$, after emergency from the layer, it became $I(x + dx)$, and so

$$I(x) - I(x + dx) = I\sigma N\,dx. \quad (12.2.1)$$

But $I(x + dx) - I(x) \simeq dI$, whence (12.1.1) yields

$$\frac{dI}{dx} = -IN\sigma. \quad (12.2.2)$$

The solution of this differential equation, as we know, is

$$I = I_0 e^{-\sigma N x} \quad (12.2.3)$$

(compare with formulas (6.6.10) and (6.6.9) and see Chapter 8). Here $I_0 = I(0)$ is the value of the energy flux at $x = 0$.

If the layer thickness is increased in arithmetic progression, $x_1 = a$, $x_2 = 2a$, $x_3 = 3a$, etc., the energy flux decreases in geometric progression. Indeed, if we introduce the notation $\alpha = e^{-\sigma N a}$ (where, of course, $0 < \alpha < 1$), we find using (12.2.3), that $I(x_1) = I_0\alpha$, $I(x_2) = I_0\alpha^2$, $I(x_3) = I_0\alpha^3$, etc.

12.3 The Relationship Between Exact and Approximate Absorption Calculations

It will be highly instructive to compare the exact solution (Section 12.2) and the rough estimate (Section 12.1). Such a comparison will help us to make use of rough estimates in complicated problems where an exact solution is hard to find. Also such a comparison

helps one to understand the range of applicability of a rough solution (see Section 10.6).

In the rough solution we found the distance over which appreciable absorption takes place, $L = 1/N\sigma$. With the aid of the characteristic length L, the exact solution (12.2.3) can be expressed as follows:

$$I = I_0 e^{-x/L}. \qquad (12.3.1)$$

Thus the supposition that the quantity L, found in a crude chain of reasoning, enters into the exact solution is fully corroborated.

The exact solution is indeed of the form

$$I = I_0 f(x/L),$$

where (cf. (12.3.1)) $f(x/L) = e^{-x/L}$.

We consider the distance $x = L$. Approximate reasoning gave complete absorption of light over this distance. Actually, from the exact solution (12.3.1), putting $x = L$, we find that $I = I_0 e^{-1} \simeq 0.37\, I_0$, which means that 37% of the light is transmitted through a cylinder of length L and, hence, the absorption is 63%. For small x/L we express $e^{-x/L}$ by means of the approximate formula $e^{-u} \simeq 1 - u$ (see Section 4.8), confining ourselves to two terms of a series expansion of e^{-u} (cf. Section 6.1) to get

$$e^{-x/L} \simeq 1 - x/L. \qquad (12.3.2)$$

Geometrically, this is tantamount to replacing the curve $y = e^{-x/L}$ by the tangent to the curve at point $x = 0$ (Figure 12.3.1). As can be seen from (12.3.2), the tangent line intersects the x axis at $x = L$. Therefore, if absorption were to occur at the same rate, that is, so that the same amount of light is absorbed on every unit of length (precisely, the amount absorbed on the initial path), all the light would be absorbed over the distance $x = L$.

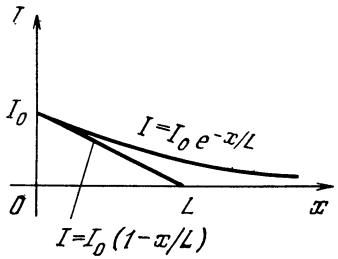


Figure 12.3.1

To summarize, then, the quantity L, which was obtained via rough considerations, is indeed of extreme importance in the exact solution as well.

Note once more the importance in the practical work of a physicist or engineer of knowing how to find and apply approximate, rough, solutions, and one should take every opportunity to develop skill in finding and understanding such solutions. This is far more important and fruitful than malicious snickering over the drawbacks of rough solutions. We will be pleased that a very rough solution yields 100% absorption where the exact solution yields 63%, the error is only by a factor of 1.5. The rough solution, for $x = L$, yields 0% light transmission instead of the exact value of 37%, that is, the approximate approach diminishes the transmission of light by a factor of ∞, but that is not so bad either because from the very start it was evident that we could not expect good accuracy from a rough solution and the 0% transmission of light must be understood only as a hint that absorption is considerable.

If it has been established that a problem does not have an exact solution in the form of an explicit formula, one should not be deterred in the least. Seek only for a very rough solution of the problem. But when using it, be sure to remember that the solution is a rough one, an approximate one, and by no means an exact one.

Let us again dwell for a moment on the question of *dimensions*. We have verified the dimensions of $L = 1/N\sigma$ and have established that this is *length*. It is often possible to find an approximate expression for the quantity that interests us when all we know are its dimensions and the dimensions of the

initial quantities given in the statement of the problem.[12.2] In the case at hand, however, this is not possible. Indeed, a quantity having the dimensions of length can be constructed by proceeding solely from the concentration N (m^{-3}), namely, $l_1 = N^{-1/3}$. (The quantity l_1 is the *mean distance* between the particles.) A quantity having the dimensions of length can also be constructed out of the cross section σ (m^2), namely, $l_2 = \sigma^{1/2}$. (The quantity l_2 characterizes the *mean* particle *size*.) It is then obvious that the quantity $l_\alpha = l_1^\alpha l_2^{1-\alpha}$ for *any* value of the exponent α, also has the dimensions of length. For instance, the L that interests us is obtained at $\alpha = 3$.

Thus, in the problem at hand, dimensional analysis does not give a definite answer. To find L, that is, the quantity with dimensions of length entering into the exact solution, it is precisely the rough solution to the problem that one, it turns out, has to find. But even when dimensional reasoning yields a *unique* answer, it is also desirable to get a rough solution to the problem so as to obtain a clearer picture of the phenomenon.

Finally, we note that the calculations carried out in Section 12.2 were based on the assumption that the soot particles are distributed randomly for a given mean concentration N. We assume that at the entrance ($x=0$) the energy flux q_0 is uniform over the entire cross section. But after the light has traveled a certain distance x through the air containing the particles, there appear, strictly speaking, certain parts of the cross section (of total area S_1) covered by shadows from the particles (and there $q = 0$), while other parts remain unshadowed (of total area S_2) where $q = q_0$ (here $S_1 + S_2 = S$, since the total cross-sectional area S remains constant). The quantity we are calculating here is the energy flux averaged over the entire cross-sectional area:

$$q(x) = \frac{q_0 S_2}{S_1 + S_2} = \frac{q_0 S_2}{S}.$$

In calculating the variation of q over the length it is essential to assume that the particles occurring within the layer between x and $x + dx$ are distributed in this layer *randomly*, that is, do not prefer illuminated or shadowed areas.

The reader is advised to investigate the situation in which the particles form an ordered array in the air like atoms in a crystal and how this influences $q(x)$.

12.4 The Effective Cross Section

In the problem of attenuation of light passing through dusty air, the quantity σ has a simple geometric meaning of the area of the shadow cast by a single dust particle. The law of attenuation of light (12.2.2) is the same for light of different wavelengths (that is, different colors), since σ is independent of the wavelength.

Contrary to the above situation, in the absorption of light by separate molecules and atoms there is observed a strong dependence of the law of attenuation of light upon the wavelength of the light. For example, in clean air at atmospheric pressure, visible light is hardly at all attenuated (attenuation is less than 1% per kilometer of path length; accordingly, the attenuation is by a factor of e over a distance of about 100 km). Ultraviolet rays of wavelength 1.8×10^{-7} m = 1800 Å (Å stands for Angström, 1 Å = 10^{-10} m) are attenuated by a factor of e over a distance of $L = 0.001$ m = 0.1 cm. Still shorter ultraviolet rays of wavelength 1.1×10^{-5} cm = 1100 Å are attenuated e times over a path length of $L = 0.01$ cm.

Consequently, the absorption of light by air is not like the absorption of light by a black dust particle, which absorbs light of any wavelength to the same degree.

[12.2] This constitutes the ***method of dimensions*** (***dimensional analysis***). The approach has wide application in physics, mechanics, and engineering.

The light energy absorbed by a single atom in unit time, q, is proportional to the energy flux I of the light at the point where the atom is located:

$$q = \sigma I. \tag{12.4.1}$$

Here σ is the constant of proportionality. Let us determine the dimensions of σ. The dimensions of q are J/s. The dimensions of I are J/m^2·s. Hence, the dimensions of σ are m^2. The quantity σ is called the ***effective cross section.*** For a black dust particle, the constant of proportionality coincides with the geometric area of the shadow. For molecules and atoms, σ is strongly dependent on the wavelength of the light.

In rough fashion, we can picture the cause of this dependence as follows. The amount of energy absorbed by an atom when acted upon by light proves to be particularly great when the frequency of the light oscillations coincides with the frequency of motion of the electrons in the atom. This is *resonance*: the electron oscillates intensively and absorbs a particularly large amount of light energy. Such a resonance is attained, for instance, in the absorption by sodium atoms (in the vapor state) of yellow light of wavelength 5890 Å $= 5.89 \times 10^{-7}$ m. The very same yellow light is emitted by sodium atoms at higher temperatures, when electron oscillations are caused by energetic collisions of atoms among themselves.

At resonance, σ reaches 10^{-14} m^2, that is, is of the order of magnitude of the square of the wavelength. Atoms and molecules are of size 10^{-10} to 10^{-9} m, which corresponds to a cross section of the order of 10^{-20} to 10^{-18} m^2. Thus, the maximum effective cross sections are many times greater than the true cross-sectional areas of atoms and molecules. On the other hand, for the light, whose frequency does not correspond to the natural frequency of the atom, the effective cross section is small, much less than the cross-sectional area of the atom.

We note, finally, that in a highly rarefied gas an atom excited by light, as a rule, re-emits the light in an arbitrary direction, while in a dense gas the excited atom transmits the energy to other atoms in collisions.

Thus, in a rarefied gas we have light scattering while in a dense gas we have true absorption. However, the "fate" of the energy taken away from the atom is unimportant in calculations of the decrease in intensity of a *directed* light ray.

12.5 Attenuation of a Charged-Particle Flux of Alpha and Beta Rays

The exponential law of diminution of particle flux as a function of distance,

$$I = I_0 e^{-x/L}, \tag{12.5.1}$$

is based on a very general assumption that the attenuation of the flux over a small distance dx is proportional to the intensity of the flux:

$$\frac{dI}{dx} = -\frac{1}{L} I, \tag{12.5.2}$$

where the constant of proportionality $1/L$ is dependent solely on the type of particle. Since (12.5.1) is the solution to the differential equation (12.5.2), it is obvious that formulas (12.5.1) and (12.5.2) are equivalent.

Experiments have shown that in certain cases the exponential law (12.5.1) is quite exact, but sometimes deviations from the law are observed. Let us consider carefully the reasons that can give rise to deviations from formula (12.5.1) or (which is the same thing) from (12.5.2).

It is easy to answer the question about the meaning of deviations from formula (12.5.2). Formula (12.5.2) presumes that when x and I vary, the light (or other radiation) under consideration does not vary qualitatively, otherwise the number L would change. We rewrite Eq. (12.5.2) as

$$\frac{1}{I}\frac{dI}{dx} = -\frac{1}{L}.$$

From this we see that the quantity $I^{-1}(dI/dx)$ is constant. If it turns out that at different points in space this quantity is different, this means that at such points not only is the intensity of radiation different but also its physical characteristics (say a different color light, that is, having a different mean wavelength).

When considering problems of protection against radioactive radiation and questions of the passage of α-, β-, and γ-rays and neutrons through various substances, there is a different reason for departure from the simple law (12.5.2).

As applied to the process of light absorption, the law (12.5.2) signifies the following: if the light encounters a dust particle, some passes by the particle without any change while the other portion of light is completely absorbed by the dust particle. The situation is more complicated in the case of radioactive radiations: an α-particle is a nucleus of the helium atom flying out of the radioactive parent nucleus at high speed (of the order of 0.07 c, where c is the speed of light, that is, at a speed of about 2×10^7 m/s). In passing through an atom, the α-particle gives up a small part of its energy to the electrons. After roughly 50 000 collisions with atoms, the α-particle will have lost half of its energy. It will not cease to exist, but its energy and speed will have changed. After 100 000 collisions the α-particle comes to a halt, ceases to collide with atoms and to knock out electrons. The stopped particle removes two electrons from the surrounding atoms and becomes a neutral helium atom at rest (with the exception of thermal motion). When Rutherford found that incidence of an α-particle into a vessel with gas is accompanied by helium buildup in the vessel, this fact signified the end of a very important stage in radioactive studies.

The number of collisions necessary to stop an α-particle is such that it takes only several centimeters of air

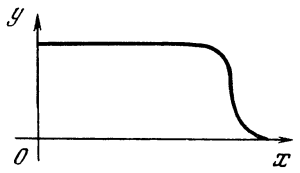


Figure 12.5.1

to do so. Actually, different α-particles (having the same initial energy) experience *different* numbers of collisions—they are not necessarily equal exactly to 100 000. However, since 100 000 is a big number, over a given path length the departures of the number of collisions of separate α-particles from the mean (100 000) are but slight (of the order of 300, which is about 0.3% of the total number of collisions).[12.3] For this reason, α-particles of the same energy always lose all their energy over roughly the same distance. This path length depends on the initial energy of the α-particle. If a flux of α-particles of the same energy is directed along the x axis, the relationship between the intensity of the flux and the path length x is shown in Figure 12.5.1. The curve is quite different from the graph of the exponential function. Over a considerable portion of the path length, the intensity of the flux of particles does not change: the same number of α-particles fly through an area of 1 cm^2 in the same intervals of time. Then the intensity falls off sharply. This drastic fall was prepared over the section where the intensity remained constant, because over this portion the enegy of the α-particles diminished with increasing path length. The sharp drop in the flux intensity

[12.3] Here the ***law of large numbers*** operates. This law deals with a *multiple repetition of random events* of the same type, the events being such that their outcome cannot be predicted. This law states that in a large number of trials the fraction of cases with a definite outcome differs very little from the expected (mean) number of such outcomes (this fraction necessarily tends to the mean number as the number N of repetitions of the event tends to infinity).

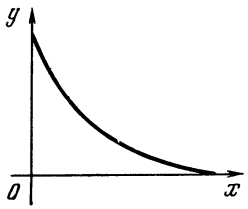


Figure 12.5.2

occurs where the energy of the α-particles becomes extremely small.

The picture is similar in the case of fast electrons (β-particles emitted when a neutron is converted into a proton in the nucleus of an atom). Here the situation is complicated by the fact that in radioactive decay electrons with different velocities are emitted; what is more, the electrons give up part of their energy to the atom near which they fly and also experience a considerable lateral deviation.

The curve for the flux intensity of β-particles as a function of path length is of the shape shown in Figure 12.5.2. We see that it differs greatly from the one in Figure 12.5.1. Even for small x's some of the electrons fall out of the beam. These are mostly electrons which had low initial velocities. Therefore, near $x = 0$ the behavior of the curve is similar to that of the exponential function. Further on, however, the curve reaches the x axis, the intensity I becomes zero for a very definite value of x corresponding to the maximum energy of the electrons generated in the given type of radioactive decay.

The most important practical problems (they are also the most difficult ones) are those connected with protection against γ-rays (gamma rays) emitted by radioactive substances and against neutrons produced in the fission of nuclei in nuclear reactors and nuclear explosions. The situation here is confused and complicated in the extreme by the fact that γ-rays and neutrons give up energy in large portions and are strongly deflected from their original directions in the process. Even in a thick layer of air (100 to 200 meters) there is a considerable probability (of the order of 37%) of the passage of unaltered γ-rays and neutrons. This is why they require thick shielding. A flux of γ-rays and neutrons does not become zero for a definite thickness of the protective layer, as was the case for α- and β-particles. As experiments and complicated calculations show, for a given large thickness of the protective layer, a flux of γ-rays and neutrons falls off in rough accord with the exponential law, with the exponent often differing from the one corresponding to the initial section of the curve.

12.6* Absorption and Emission of Light by a Hot Gas

Let us imagine a hot gas through which a beam of light is passing. We wish to study in greater detail the effect that light propagation has on the atoms of the gas, say, sodium atoms. More precisely, we wish to know what happens to the atoms when they absorb or emit light and when they collide with other atoms and molecules, bearing in mind that the temperature of the gas is high.

First we must note that the assumption that an atom is like a small ball on a spring, or an oscillator (a device with a definite natural frequency), in many cases is not sufficiently accurate. An atom is a system consisting of a nucleus and electrons that obeys the laws of *quantum mechanics*. These laws cannot be discussed here; suffice it to say that a quantum mechanical system of this type may be only in certain states with definite (or, as the mathematician would say ***discrete***) energy values. The lowest state, which is characterized by the *lowest* energy possible for the system, is called the ***ground state*** (we will sometimes call it the unexcited, or zeroth, state and label the quantities referring to this state with a subscript "0": the energy of the atom in this state is E_0, the number of atoms in the ground state per unit volume is n_0, and so on). In the same way we will speak of the first state

(after the ground state) as the first excited state and label the respective quantities with a subscript "1": the energy E_1, the concentration of atoms n_1, etc. The second excited state has energy E_2, a concentration of atoms n_2, and so on. We will always assume that $E_0 < E_1 < E_2 \ldots$.

At a low temperature all the atoms in a system have the lowest possible energy, that is, are in the ground state: the energy of each atom is E_0, and the concentration n_0 of atoms is equal to the concentration n of the entire collection of atoms (in all possible states).[12.4]

An atom in its ground state can absorb light, going over in the process to another state, say, the first excited. What does quantum mechanics say about this? In the language of quantum mechanics, light consists of separate particles, called sometimes ***light quanta*** and sometimes ***photons***, and the energy of each such particle (or quantum of light or photon) is $h\nu$, where $h = 6.63 \times 10^{-34}$ J·s is ***Planck's constant***, and ν is the ***frequency*** of the light (number of oscillations per second). A peculiar feature of quantum theory is that while speaking of particles of light we do not give up such notions as the frequency of the light and wavelength, which strictly speaking belong only to the concept of light as an electromagnetic wave.

But let us go no deeper into the philosophy of quantum mechanics.

Energy conservation states that when an atom absorbs light (a single photon) and goes over to its first excited state, then

$$E_0 + h\nu = E_1. \qquad (12.6.1)$$

The same holds when an excited atom emits a photon, going over from the first excited state to the ground state. This means that the "resonance" frequency ν of the absorbed or emitted light depends on the *difference* in the energies of the two states of the atom; this frequency (which can be measured using a spectrometer) characterizes the *transition* from one state of an atom to another. The quantity ν, more precisely, $\nu_{1,0}$, is not the frequency of electron vibrations inside the atom in a definite state (zeroth or first excited)—it depends on *both* states. Spectroscopists (scientists dealing with the theory and interpretation of interactions between matter and radiation) noticed this pattern long ago, since it is evident from tables of spectral lines in the spectra of various atoms. Indeed, it was found that

$$\nu_{4,1} + \nu_{1,0} = \nu_{4,3} + \nu_{3,0} = \ldots.$$

(Why? Verify this.) The validity of such equalities can be verified without knowing anything about atoms.

The famous Danish physicist Niels Bohr (1885-1962) was the first to understand the true meaning of, and reason for, such a pattern. He formulated the following "quantum postulates":

(1) there is a set of states (0, 1, 2, . . .), called stationary states, in which the atom does not emit light;

(2) emission of light occurs during the *transition* of the atom between two stationary states, and the frequency of the radiation emitted equals the difference in the energy of the states divided by Planck's constant.

Similar statements can be formulated for light absorption by atoms.

Let us now return to the main question of this section—how a *hot* gas absorbs and emits light. The reader will recall that at a *low* temperature all the atoms in a gas are in their ground state and that light of resonance frequency (we will again write ν instead of $\nu_{1,0}$) is absorbed, so that

$$\frac{dI}{dx} = -\sigma n_0 I, \qquad (12.6.2)$$

where I is the light flux and σ is the so-called cross section (see formula (12.2.2)). What will change if the gas is *hot*?

[12.4] Strictly speaking, $n = n_0 + n_1 + n_2 + \ldots$, but at a low temperature $n_1 \ll n_0$, $n_2 \ll n_0$, etc., and $n_1 + n_2 + \ldots \ll n_0$, so that $n \simeq n_0$.

In a hot gas, as a result of the thermal motion of the atoms and collisions of the atoms with atoms and molecules of other gases (if we are dealing with a mixture of gases), a fraction of the atoms will transfer to states with higher energies. In other words, we will have excited atoms in addition to those in the ground state, that is, we can no longer ignore n_1, n_2, ... According to the general law of thermal motion (see Section 11.4) at a given temperature T, we have

$$n_0 \div n_1 \div n_2 \div \ldots = e^{-E_0/kT} \div e^{-E_1/kT} \div e^{-E_2/kT} \div \ldots . \qquad (12.6.3)$$

If the total number of atoms per unit volume, n, is fixed, then, combining (12.6.3) and the fact that $n = n_0 + n_1 + n_2 + \ldots$, we get

$$n_0 = n \frac{e^{-E_0/kT}}{e^{-E_0/kT} + e^{-E_1/kT} + \ldots} = \frac{n}{1 + e^{-(E_1 - E_0)/kT} + \ldots}, \qquad (12.6.4)$$

$$n_1 = \frac{ne^{-(E_1 - E_0)/kT}}{1 + e^{-(E_1 - E_0)/kT} + \ldots}, \ldots .$$

We see that for a fixed n the absorption drops as the temperature rises, since $n_0 < n$. But in addition to this—and this is the most important aspect—there appear excited atoms in the gas, atoms that are capable of emitting light. The frequency of the light they emit is the same as that of the light they absorbed earlier.

It may be assumed that in the presence of excited atoms the equation for the intensity of the light is

$$? \frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1 .? \qquad (12.6.5)$$

Here the coefficient $\varkappa$ must have the meaning of the product of the probability w of an excited atom going over in one second to the ground (zeroth) state and emitting a photon with the probability α of the emitted photon joining the light flux. Since I is expressed in units of J/m^2·s, we must multiply $w\alpha$ by the energy $h\nu$ of one photon. The reasoning behind all this is as follows: in a layer dx cm thick and a cross-sectional area of one square centimeter there are $n_1 dx$ excited atoms, with the rest of the reasoning done on dimensional grounds, namely, w has the dimensions of 1/s = s^{-1}, α is dimensionless, and n_1 has the dimensions of m^{-3}, so that if $\varkappa = w\alpha h\nu$, the dimensions of $\varkappa n_1$ are J/m^3·s, which is the same as the dimensions of dI/dx. We see that on dimensional grounds formula (12.6.5) is correct. And yet it is not correct!

The question marks on both sides of Eq. (12.6.5) are to remind the reader that this formula is incorrect. We will see how this formula must be modified when we consider the *thermodynamic equilibrium* of light fluxes.

12.7* Radiation in Thermodynamic Equilibrium

Suppose that the vessel containing the gas has ideally reflecting walls. The temperature of the gas is fixed and is maintained constant.

It is natural to assume that if initially there was no light in the vessel, it will appear as a result of emission by excited atoms, while if initially the intensity of the light was too high, the surplus of light will be absorbed by the atoms in the ground state.

It can be expected that, regardless of the concrete initial state, a certain intensity of light will set in in the vessel as a result of absorption and emission of light by the gas with a fixed temperature. More refined reasoning shows that the intensity of light of a given frequency depends *only* on the temperature of the gas. Indeed, imagine two vessels filled with two different gases but having the *same* temperature. Next connect these vessels with a light guide containing an optical filter that transmits light only of a definite frequency in both directions. If the light intensities were different, then, in spite of having the same temperature, one vessel would give off more energy than it receives, while the other would receive

more energy than it gives off, so that the first vessel would cool off while the other would heat up, which is clearly impossible.

Thus, we conclude that there is a certain ***equilibrium intensity*** of light, an intensity at which there is equilibrium between emission and absorption. This equilibrium intensity at a given frequency depends only on the temperature of the gas and not on the properties of the gas. It is natural then to try to find the equilibrium intensity of light starting directly from the general principles of thermal motion.

Let us consider radiation (or light) of a certain frequency and direction in a vessel. The energy of this radiation cannot assume arbitrary values: according to quantum theory (in our case, the theory of photons), the energy may be zero (not a single photon of a given wavelength) or $h\nu$ (one photon) or $2h\nu$ (two photons), and so on.

From general principles, the ratios of the probabilities p_0, p_1, p_2, . . . of the corresponding states are

$$p_0 \div p_1 \div p_2 \div \ldots = 1 \div e^{-h\nu/kT} \div e^{-2h\nu/kT} \div \ldots, \quad (12.7.1)$$

in accordance with the fact that the energies are 0, $h\nu$, $2h\nu$, . . . and the probabilities are proportional to $e^{-E/kT}$ (since $e^0 = 1$, the sequence on the right-hand side of (12.7.1) starts with unity). But the sum of all these probabilities must be equal to unity because we have exhausted all the possibilities—the given light ray may be characterized by one, two, three, etc., or by the absence of photons. This means that we can normalize, so to say, the distribution (12.7.1) by adding to these ratios of probabilities the following:

$$p_0 + p_1 + p_2 + \ldots = 1, \quad (12.7.2)$$

which shows that the sum of the probabilities is unity. From (12.7.1) it clearly follows that

$$p_n = ae^{-nh\nu/kT}, \quad n = 0, 1, 2, \ldots \quad (12.7.3)$$

(i.e. $p_0 = a$, $p_1 = ae^{-h\nu/kT}$, etc.), where factor a has yet to be found. To find a, we employ (12.7.2):

$$p_0 + p_1 + p_2 + \ldots = a\,(1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \ldots) = \frac{a}{1 - e^{-h\nu/kT}}, \quad (12.7.4)$$

and, hence,

$$a = 1 - e^{-h\nu/kT} \quad (12.7.4a)$$

note that inside the parentheses in (12.7.4) we have the sum of the terms of a *geometric progression*).

Now we can find the *average number of photons* φ and the *average energy* (more precisely, the *energy density*) ε in a given type of rays:

$$\varphi = 1 \times p_1 + 2 \times p_2 + \ldots, \quad (12.7.5)$$

$$\varepsilon = \varphi h\nu \quad (12.7.5a)$$

(since φ is the average number of photons per unit volume, the dimensions of this quantity are m^{-3}, while the dimensions of ε are J/m^3). Substituting (12.7.3) and (12.7.4a) into (12.7.5) and then the result into (12.7.5a), we get

$$\varphi = \frac{e^{-h\nu/kT}}{1 - e^{-h\nu/kT}} = \frac{1}{e^{h\nu/kT} - 1}, \quad (12.7.6)$$

$$\varepsilon = \frac{h\nu}{e^{h\nu/kT} - 1} \quad (12.7.6a)$$

(here we have left out the simple calculations and have given the final result[12.5]). A given frequency ν corresponds to a certain period $1/\nu$ of the oscillations and a certain wavelength $\lambda = c/\nu$ of the wave.

We can now find the number of waves in a volume V with frequencies ranging from ν to $\nu + d\nu$. One must bear in mind that electromagnetic waves have a certain polarization: the electric field in a wave is at right angles to the direction of propagation of the wave (in three-dimensional space, however, there are numerous mutually perpendicular directions that are at right angles

[12.5] To derive (12.7.6) from (12.7.3), (12.7.4a), and (12.7.5), it is sufficient to note that $\varphi - e^{-h\nu/kT}\,\varphi = p_1 + p_2 + p_3 + \ldots = 1 - p_0\ (= 1 - a)$. (Why?)

to a given direction). We must also allow for the conditions of wave reflection from the walls of the vessel, but it has been found that the number of independent types of oscillations with frequencies within a given range depends, to the first approximation, only on the volume V and is proportional to this volume. Thus, we can speak of a certain number dK of oscillations within a frequency range $d\nu$ per unit volume; this number proves to be

$$dK = \frac{8\pi\nu^2\, d\nu}{c^3}, \qquad (12.7.7)$$

where c is the speed of light. Since the frequency ν (and, hence, $d\nu$) has the dimensions of s^{-1} and c the dimensions of m/s, the dimensions of dK are m^{-3}.

A full derivation of (12.7.7) is too tedious to include in a book for beginners, but it is easy to see that (12.7.7) is reasonable and can easily be understood. Note that $\nu/c = 1/\lambda$, whence $\nu = c/\lambda$ and $|\, d\nu\,| = (c/\lambda^2)\, d\lambda$. Then, according to (12.7.7), the number of oscillations with wavelengths ranging from λ to $\lambda + d\lambda$ is

$$dK = \frac{8\pi}{\lambda^4}\, d\lambda. \qquad (12.7.7a)$$

This means that the number of oscillations with wavelengths ranging, say, from λ_0 to $2\lambda_0$ is

$$K_{\lambda_0,\, 2\lambda_0} = 8\pi \int_{\lambda_0}^{2\lambda_0} \frac{d\lambda}{\lambda^4} = \frac{8\pi}{3}\left(\frac{1}{\lambda_0^3} - \frac{1}{2^3\lambda_0^3}\right) = \frac{7\pi}{3}\,\frac{1}{\lambda_0^3}. \qquad (12.7.8)$$

If we abstract ourselves from the numerical factor, which differs but little, in order of magnitude, from unity, we can say that the number of oscillations with wavelengths close to λ_0 in a unit volume is approximately equal to the number of oscillations that fit inside a cube with an edge λ_0. On the average, each oscillation occupies a volume of λ_0^3, which constitutes a result that not only can easily be remembered but is also quite natural.

Now we can easily write an expression for the total *energy of radiation* per unit volume, Q. To this end we take the energy of each oscillation with a definite frequency, multiply it by the number of oscillations of the given type, and add the products (that is, integrate them):

$$Q = \int \varepsilon\, dK = \int_0^\infty \frac{h\nu}{e^{h\nu/kT} - 1}\,\frac{8\pi\nu^2\, d\nu}{c^3} \qquad (12.7.9)$$

(Q has the dimensions of energy density, J/m^3). It is convenient to introduce the notation $h\nu/kT = x$ and isolate the dimensionless integral thus:

$$Q = \frac{8\pi h}{c^3}\left(\frac{kT}{h}\right)^4 \int_0^\infty \frac{x^3\, dx}{e^x - 1}. \qquad (12.7.9a)$$

The integral on the right-hand side of (12.7.9a) cannot be expressed in terms of elementary functions. Approximately, the integral

$$I = \int_0^\infty \frac{x^3\, dx}{e^x - 1} \qquad (12.7.10)$$

can be evaluated if we note that

$$\frac{1}{e^x - 1} = e^{-x}\,\frac{1}{1 - e^{-x}} = e^{-x}(1 + e^{-x} + e^{-2x} + \ldots) = e^{-x} + e^{-2x} + \ldots,$$

and, hence,

$$\int \frac{x^3\, dx}{e^x - 1} = \int x^3 e^{-x}\, dx + \int x^3 e^{-2x}\, dx + \ldots. \qquad (12.7.11)$$

But (see (7.6.19))

$$\int_0^\infty x^3 e^{-x}\, dx = 3! = 6, \qquad \int_0^\infty x^3 e^{-2x}\, dx = \frac{1}{2^4}\int_0^\infty y^3 e^{-y}\, dy = \frac{1}{16}\times 6, \text{ etc.},$$

so that

$$\int_0^\infty \frac{x^3\, dx}{e^x - 1} = 6\left(1 + \frac{1}{16} + \frac{1}{81} + \ldots\right). \qquad (12.7.11a)$$

The series on the right-hand side of (12.7.11a) is convergent; if we keep only five terms, we get 6.482 for the approximate value of integral (12.7.10).

The theory of functions of a complex variable (see Section 17.2) yields the *exact* value of (12.7.10), which is $\pi^4/15$ or approximately 6.494; thus, the error introduced by limiting the sum in (12.7.11a) to five terms is smaller than 0.2%.

Thus, the exact expression for the equilibrium radiation density (in J/m^3) at a given temperature T (in kelvins) is

$$Q = \frac{8\pi^5}{15} \frac{k^4 T^4}{c^3 h^3} \simeq 7.5 \times 10^{-16} T^4. \quad (12.7.12)$$

Suppose that the vessel in which such a radiation density has set in has a hole in one of its walls with an area of S square centimeters through which energy flows out. The energy flux will be equal to RS, where R is measured in units of $J/m^2 \cdot s$, with

$$R = \frac{c}{4} Q \simeq 5.67 \times 10^{-8} T^4. \quad (12.7.13)$$

Electromagnetic energy "flows," or propagates, with the speed of light c, but at each given moment of time only a half of the photons are moving toward the hole, with the other half moving in the opposite direction. In addition to this, even the photons that do move toward the hole are flying at different angles to the wall with the hole, which diminishes the flux by a factor of two. Exact allowance for these factors gives the coefficient 1/4 in (12.7.13).

Now let us close the hole in our box (vessel with mirror walls) with a black material heated to the same temperature as that of the box and the radiation inside it. When we speak of the material as *black*, this means that any photon with any frequency falling on the material is absorbed. Our "stopper" should heat up, since it absorbs radiation. However, since the stopper has the same temperature as the radiation, it obviously cannot heat up to a higher temperature. Hence, the stopper will emit as much radiation as it absorbs: we conclude, therefore, that a black body with a temperature T emits per unit area per unit time $5.67 \times 10^{-8} T^4$ joules of energy (T in kelvins). This quantity characterizes what is called ***blackbody radiation.***

The order in which we have discussed all these questions is opposite to the order in which these concepts developed historically. First the law (12.7.12) was discovered in experiments, that is, it was established that the total energy radiated by a black body is proportional to the fourth power of the temperature of the body (1888; the ***Stefan-Boltzmann law*** or the ***fourth-power law*** or ***Stefan's law of radiation***), with the proportionality coefficient equal to $5.67 \times 10^{-8} J/m^2 \cdot s \cdot K^4$ (the ***Stefan-Boltzmann constant***).

In 1899 formula (12.7.6a), known as the ***Planck radiation formula***, was discovered. ***Planck's constant*** $h = 6.63 \times 10^{-34}$ J·s for the first time appeared in this formula.

Planck assumed that energy associated with electromagnetic radiation is emitted and absorbed in discrete amounts proportional to the frequency of the radiation, with h being the coefficient of proportionality, but that quantum theory has nothing to do with the propagation of the radiation. Later, in 1905, Einstein established that light consists of certain portions with the energy of each portion equaling $h\nu$ and momentum equaling $h\nu/c$. Finally, in 1924, the Indian physicist Satyendra Nath ***Bose*** (1894-1974) gave a derivation of the Planck radiation formula based on the calculation of the probabilities p_0, p_1, p_2, . . ., with which we started our discussion.

12.8* Emission Probability and the Conditions for Thermodynamic Equilibrium

Let us return to the formula for the intensity of a beam of light passing through a hot gas.

From Eq. (12.6.5) (the reader will recall that this equation is incorrect),

$$? \frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1, ?$$

and the condition $dI/dx = 0$ we obtain the equilibrium intensity I_{eq}:

$$? I_{eq} = \frac{\varkappa n_1}{\sigma n_0}. ? \qquad (12.8.1)$$

The ratio n_1/n_0 of the concentration of excited atoms, n_1, to the concentration of atoms in the ground state, n_0, is equal to $\exp[-(E_1 - E_0)/kT]$ (see (12.6.3) or (12.6.4)). According to the quantum condition (12.6.1), $E_1 - E_0 = h\nu$, which means that $n_1/n_0 = e^{-h\nu/kT}$, so that

$$? I_{eq} = \frac{\varkappa}{\sigma} e^{-h\nu/kT}. ? \qquad (12.8.1a)$$

Formula (12.8.1a) for I_{eq} is similar to the correct formula (12.7.6a) obtained by Planck, but does not coincide with it.

We will start with the similar aspects. The equilibrium radiation density I_{eq} does not depend on the concentration of the gas whose atoms absorb (in the ground state) and emit (in an excited state) radiation; it depends only on the frequency ν of the light and the temperature T of the gas. In the limit $h\nu/kT \gg 1$, the dependence of I_{eq} on these two quantities is the same as in the Planck radiation formula. Indeed, if $h\nu/kT \gg 1$, then $e^{h\nu/kT} \gg 1$, and in the Planck radiation formula (12.7.6a),

$$\varepsilon = \frac{h\nu}{e^{h\nu/kT} - 1},$$

we can ignore the one in the denominator in comparison to the large term $e^{h\nu/kT}$: for $h\nu \gg kT$ we have

$$\varepsilon \simeq \frac{h\nu}{e^{h\nu/kT}} = h\nu e^{-h\nu/kT},$$

from which it follows that

$$I_{eq} \simeq ch\nu e^{-h\nu/kT}, \quad \frac{h\nu}{kT} \gg 1, \qquad (12.8.2)$$

since I is the energy flux, and c is the speed of light. (Note that since the energy density ε has the dimensions of J/m^3 and c has the dimensions of m/s, the product $c\varepsilon$ has the dimensions of $J/m^2 \cdot s$, which are the dimensions of I.)

Thus, the formula (12.8.1a), which was obtained via naive assumptions, expresses a dependence of I_{eq} on ν and T that does not differ significantly from (12.8.2). For the similarity to be complete, we must require that the ratio of the energy given off by an excited atom to a single wave in the form of radiation (this energy is characterized by coefficient $\varkappa$) to the cross section of absorption of light by the atom in the ground state, σ, obeys the following expression:

$$\frac{\varkappa}{\sigma} = ch\nu. \qquad (12.8.3)$$

From the point of view of the theory of resonance (and from the point of view of quantum mechanics, as well), this formula looks quite natural. If the absorption cross section is small (e.g., because the charge on the vibrating ball is small), so is the probability of emission (which also depends on the charge on the ball).

Now we wish to go from the probability of emission of a single wave to the total probability of energy emission by an excited atom. If we divide the energy emitted per unit time into a single wave by the energy $h\nu$ of one photon, we obtain the probability of emission into a single wave. The next step is to add all these probabilities, that is, to integrate over the frequency range:[12.6]

$$w = \int \frac{\varkappa}{h\nu} \frac{8\pi\nu^2 \, d\nu}{c^3} = \int \frac{\sigma 8\pi\nu^2 \, d\nu}{c^2} \qquad (12.8.4)$$

(the dimensions of w are $1/s = s^{-1}$). Since w is the emission probability (precisely, the probability of emission per unit time, or emission rate), we

[12.6] It is unimportant that this range includes frequencies which the given excited atom will not emit, since for such frequencies $\varkappa = 0$ and the factor $\varkappa(\nu)$ in the integrand will cut off such frequencies.

have the following expression for the concentration of excited atoms in the absence of external radiation:

$$\frac{dn_1}{dt} = -wn_1. \qquad (12.8.5)$$

The solution to this equation is obvious:

$$n_1(t) = n_1(t_0) \exp[-w(t - t_0)] \qquad (12.8.6)$$

Accordingly, $\tau = w^{-1}$ is the mean lifetime of the excited atom (cf. Section 8.2).

The reader will note that above only the *first* excited state was considered, a state that can decay in only one manner, that is, go over to the ground (or unexcited) state. As a rule, for a given ground state and a fixed excited state (with a given spatial orientation of the atom), the absorption cross section σ depends only on the direction of propagation and the polarization of the incident wave. The exact formula in this case assumes the form

$$w = \int (\sigma_1 + \sigma_2) \frac{\nu^2 d\nu}{c^2} d\Omega, \qquad (12.8.7)$$

where σ_1 and σ_2 refer to the two polarizations of the wave, and $d\Omega$ is the solid-angle element. When $\sigma_1 = \sigma_2 = \sigma$, that is, the cross sections are independent of direction, we have $\sigma_1 + \sigma_2 = 2\sigma$ and $\int d\Omega = 4\pi$, hence the factor 8π in the simplified formula (12.8.4).

Note the dimensions of the various quantities: σ has the dimensions of m^2 and $c/\nu = \lambda$ the dimensions of wavelength, m, so that the quantity σ/λ^2 in the integral

$$w = \int \frac{8\pi\sigma}{\lambda^2} d\nu \qquad (12.8.8)$$

is dimensionless. What is important is that σ depends strongly on the frequency—light absorption is a resonance process. The reader will recall the results obtained in the theory of oscillations: the absorption of energy by a damped oscillator depends on frequency via the following formula:

$$A \propto \frac{1}{(\nu - \nu_0)^2 + \gamma^2}. \qquad (12.8.9)$$

It has been found that in the case of an atom at rest the absorption of light obeys a similar formula:[12.7]

$$\sigma = \frac{\lambda^2}{2\pi} \frac{\gamma^2}{(\nu - \nu_0)^2 + \gamma^2}. \qquad (12.8.9a)$$

We see that the maximum cross section of absorption of light is of the order of the square of the light's wavelength. For visible light this is many times the size of the atom: the yellow line of sodium has a wavelength of $\lambda = 5400 \times 10^{-10}$ m, while the radius of an electron orbit is of the order of 10^{-10} m, which means that the resonance cross section of absorption of a photon is greater than the area limited by the electron orbit by a factor of several millions. Here a photon manifests itself not as a point-like particle flying along a straight line and either hitting the target or missing it. According to the laws of quantum mechanics, the movement of a photon is determined by the wave equations of the electromagnetic field, and in this lies the possibility for an atom to capture a photon over a distance of the order of the light's wavelength.

Up till now we emphasized the similarity between the approximate formula (12.8.1a) for the equilibrium radiation density and the exact Planck radiation formula (12.8.2) (note that $h\nu \gg kT$). But in what respects does the exact formula differ from the approximate; in other words, how should we modify the expression for dI/dx so that the equilibrium radiation density corresponds to the Planck radiation formula? The answer to this question was given by Einstein in 1917. As often happens in physics, the answer proved to be easier than formulating the question correctly.

[12.7] Substituting (12.8.9a) into the integral on the right-hand side of (12.8.8), we get $w = 4\pi\gamma$.

Einstein's answer was as follows. We replace Eq. (12.6.5),

$$\frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1,$$

with the following:

$$\frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1 (1+\varphi), \qquad (12.8.10)$$

where according to what was established earlier $\varkappa = ch\nu\sigma$. If φ is the average number of photons in a given ray (see Section 12.7), then

$$I = ch\nu\varphi, \quad \frac{dI}{dx} = ch\nu \frac{d\varphi}{dx}. \qquad (12.8.11)$$

Substituting this into Eq. (12.8.10) yields

$$\frac{dI}{dx} = ch\nu \frac{d\varphi}{dx} = ch\nu\sigma [-n_0\varphi + n_1 (1+\varphi)]. \qquad (12.8.12)$$

First let us see whether we achieved our goal. Indeed, the condition $I = I_{eq}$, that is, $dI/dx = 0$, leads to the following formula for the equilibrium number of photons:

$$\frac{\varphi_{eq}}{1+\varphi_{eq}} = \frac{n_1}{n_0} = e^{-h\nu/kT},$$

that is, $-n_0\varphi_{eq} + n_1 (1+\varphi_{eq}) = 0$, or

$$\varphi_{eq} = \frac{e^{-h\nu/kT}}{1-e^{-h\nu/kT}} = \frac{1}{e^{h\nu/kT}-1}. \qquad (12.8.13)$$

We have thus arrived at an expression for the probability of absorption and emission of light which leads to the Planck formula. Now let us stop and evaluate the price we paid for this result.

In formula (12.8.10) we added to the term $\varkappa n_1$ another term, $\varkappa n_1\varphi$, equal to $\sigma n_1 I$. Thus we (following Einstein) maintain that there exists a process of photon emission induced by the presence of the light beam and proportional to the number of photons, or radiation intensity, in the already existing beam. This process is known as ***stimulated emission*** (the word "stimulated" refers to the fact that the emission is caused by the presence of a light beam in the hot gas).

Stimulated emission appears alongside ***spontaneous emission;*** the two terms in the parentheses in (12.8.10) correspond to the two types of emission, $\varkappa n_1$ and $\sigma n_1 I$. Accordingly, the equation for the variation of the number of excited atoms n_1 is

$$\frac{dn_1}{dt} = -wn_1 - \frac{\sigma n_1 I}{h\nu} + \frac{\sigma n_0 I}{h\nu}. \qquad (12.8.14)$$

In the absence of radiation we have only one term, $-wn_1$, as we wrote earlier. However, in the presence of light, an atom can not only "jump" from the ground state to an excited (or upper) state (this fact is represented by the term $\sigma n_0 I/h\nu$ on the right-hand side of (12.8.14)) but can emit stimulated radiation by going over from an excited state to the (lower) ground state (the term $-\sigma n_1 I/h\nu$ in (12.8.14)). We also note that the cross section σ in the terms describing absorption ($\sigma n_0 I/h\nu$) and stimulated emission ($-\sigma n_1 I/h\nu$) is the same. If the intensity I of the beam is high and we can neglect the term $-wn_1$ describing spontaneous emission in comparison to the terms responsible for absorption and stimulated emission, the equation for I simplifies:

$$\frac{dI}{dx} = -\sigma (n_0 - n_1) I. \qquad (12.8.15)$$

Obviously, the condition for this is the "strong" inequality $I \gg I_{eq}$, where I_{eq} is the equilibrium radiation intensity at a given temperature of the gas.

Let us substitute into (12.8.15) the equilibrium concentration of excited atoms, $n_1 = n_0 e^{-h\nu/kT}$. The result is a formula for the effective absorption cross section:

$$\frac{dI}{dx} = -\sigma (1 - e^{-h\nu/kT}) n_0 I. \qquad (12.8.16)$$

When the gas is heated, the effective cross section, that is, the coefficient of I, decreases not only due to the decrease in n_0 but also due to the countereffect produced by the stimulated emission, which yields an additional factor inside the parentheses on the right-hand side

of (12.8.16). If $h\nu \ll kT$, then

$$e^{-h\nu/kT} \simeq 1 - \frac{h\nu}{kT} \quad \text{and}$$

$$1 - e^{-h\nu/kT} \simeq \frac{h\nu}{kT} \ll 1,$$

and the decrease in absorption is considerable. If there are only two levels, 0 and 1, so that the total number of atoms is $n = n_0 + n_1$, then n_0 drops by a factor not greater than two as the temperature T is raised from 0 to ∞.[12.8]

For example, in radio astronomy a marked effect is produced by the transition of hydrogen atoms from a low-lying excited state to the ground state accompanied by emission of radiowaves with $\lambda = 21$ cm (the hydrogen line). The energy of the respective photons is $h\nu = hc/\lambda \simeq 6.5 \times 10^{-34} \times 3 \times 10^8/0.21 \simeq 10^{-24}$ J. If the temperature of the hydrogen cloud is 100 K, then $kT \simeq 1.37 \times 10^{-23} \times 100 \simeq 10^{-21}$ J, and the value of the factor inside the parentheses in (12.8.16) is $(1 - e^{-h\nu/kT}) \simeq h\nu/kT \simeq 10^{-3}$ (cf. Section 4.8).

What are the qualitative features of the radiation produced in stimulated emission? The excited atom in this process emits radiation that amplifies the forward wave (the wave that stimulates the atom, so to say), since otherwise there would be no mechanism for compensation of the energy absorbed by the atoms in the ground (unexcited) state. In this respect, stimulated emission differs from spontaneous emission, in which the excited atom emits radiation in all directions and only partially in the direction of the forward wave.

Using the language of classical mechanics, we could say that radiation absorption is similar to resonance build-up of oscillations (cf. Section 10.5), while stimulated emission is analogous to damping of an oscillator, since to damp oscillations we need a force that is in resonance with the oscillations, that is, has the same frequency. The difference between a driving force and a damping force lies in the different ratios of the phase of the force to the phase of the oscillator.

However, there is no classical picture that can completely explain the phenomenon of stimulated emission. An exact and complete theory of stimulated emission is possible only if we employ quantum mechanics and the quantum theory of electromagnetic radiation (quantum field theory). The theory of stimulated emission enters as an integral part into the quantum theory of radiation, whose bases were developed by P. Dirac in 1927.

This fact stresses even more the remarkable intuition of Einstein who in 1917 was able to clarify the main features of stimulated emission proceeding only from purely thermodynamic considerations, approximately in the same manner as we have done. Note that the very notion of light as a stream of particles, photons, was first introduced into science by Einstein in 1905, and it was for this discovery that he was awarded the Nobel Prize.

12.9* Lasers

Stimulated emission opens up a remarkable possibility for generating electromagnetic oscillations (light, for one thing).

Suppose that we have a medium in which the concentration n_1 of atoms in an excited state is *higher* than n_0, the concentration of atoms in the ground state. Under ordinary conditions of thermal equilibrium, with $n_1/n_0 = e^{-h\nu/kT}$, such a situation is impossible, since here $n_1/n_0 < 1$. But there are several ways in which the reverse can be achieved.

One method of making n_1 greater than n_0 will be discussed later. Now, without going into details, we will simply assume that $n_1 > n_0$, so that

[12.8] Of course, it would be more correct to say that the temperature is raised from 0 to $T \gg (E_1 - E_0)/k$, since from the standpoint of physics the passage to the limit $T \to \infty$ is meaningless—the atoms would simply decay.

the exact equation (12.8.10) can be written in the following form:

$$\frac{dI}{dx} = bI + \varkappa n_1, \qquad (12.9.1)$$

with $b = \sigma (n_1 - n_0)$ positive. The reversal of sign of the coefficient of I in the expression for dI/dx changes drastically the behavior of the solution. The changes in the properties of the solution can be compared to those that occur in the solution for the concentration of neutrons in radioactive decay when we go over from the subcritical mass of the fissionable material (uranium-235 or plutonium) to the supercritical mass (Section 8.9).

Indeed, if b in the equation for dI/dx is positive, the light flux will increase exponentially as it passes through the medium. Neglecting the second term on the right-hand side of (12.8.1), we find that

$$I = I_0 e^{bx} > I_0, \qquad (12.9.2)$$

where I_0 is the intensity of the light at the origin of coordinates, that is, at $x = 0$ (at the entrance to the medium). A definite volume of the gas (the substance) with $b > 0$, say, a pipe of length L with the gas in it, operates like an *amplifier*: with a flux I_0 at the entrance, we have a flux $I_L = I_0 e^{bL} > I_0$ at the exit which is more powerful than the incident flux.

Now suppose that at both ends of the pipe we have mirrors with reflection coefficients β and δ. If at the left end ($x = 0$) we have an influx of radiation I_0, then at the right end of the pipe we have, as we know, $I_L = I_0 e^{bL}$. This flux will be reflected by the right mirror (reflection coefficient β), and the reflected flux

$$|I'_L| = \beta I_L = I_0 \beta e^{bL} \qquad (12.9.3)$$

enters the pipe from the right (this flux must be considered *negative* since the energy is transferred by this flux in the direction opposite to the positive direction of the x axis). So as not to have to think about the right sign, we put a prime on the respective flux

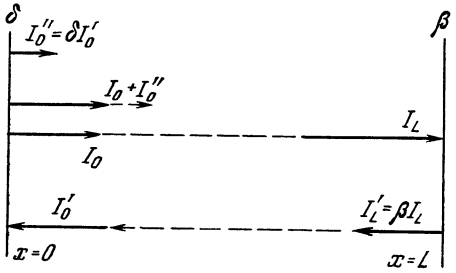


Figure 12.9.1

(we will write I') and consider only its absolute value.

The equation for I' has the form

$$\frac{d\,|\,I'\,|}{dx} = -bx. \qquad (12.9.4)$$

We integrate this equation "from right to left," with the "initial condition" being the quantity $|\,I'_L\,| = I_0 \beta e^{bL}$ corresponding to the value $x = L$ (see (12.9.3)), and find the flux value I'_0 corresponding to $x = 0$:

$$|I'_0| = |I'_L| \exp\left(-b \int_L^0 dx\right) = |I'_L|\, e^{bL} = I_0 \beta e^{2bL}. \qquad (12.9.5)$$

After reflection from the left mirror (reflection coefficient δ) we have a flux (see Figure 12.9.1) equal to

$$I''_0 = \beta\delta e^{2bL} I_0. \qquad (12.9.6)$$

Now we must allow for the fact that the total flux from left to right is

$$I_{\text{tot}} = I_0 + I''_0,$$

where I_0 is the intensity of the incident light, and I''_0 is the intensity of the twice reflected light. On the right-hand side of the expression (12.9.6) for I''_0 we must put I_{tot} instead of I_0 (above we did not distinguish between I_0 (the incident intensity) and I_{tot} because double reflection was not considered a possibility). We then have

$$I_{\text{tot}} = I_0 + I''_0 = I_0 + \beta\delta e^{2bL} I_{\text{tot}},$$

whence

$$I_{\text{tot}} = \frac{I_0}{1 - \beta\delta e^{2bL}}. \qquad (12.9.7)$$

From the standpoint of physics we can say that at $\beta\delta e^{2bL} = 1$ the system consisting of a pipe with the active gas and two mirrors (with reflection coefficients β and δ) is in a "critical" state: a stationary state is possible for any value of I_0. If $\beta\delta e^{2bL}$ is less than unity, the system is "subcritical" (cf. Section 8.9). If we ignore spontaneous emission, there is only one solution, $I_0 = 0$, while if we allow for spontaneous emission (that is, the term $\varkappa n_1$ in Eq. (12.9.1)), radiation with an intensity proportional to $\varkappa n_1 L$ sets in in the system. Finally, if we allow for stimulated emission, the amplification (or ***gain***, as it is called) is even higher: it is proportional to the ratio $1/(1 - \beta\delta e^{2bL})$.

From the standpoint of technology such a situation is of little interest. A remarkable situation emerges when $\beta\delta e^{2bL}$ is greater than unity. By analogy with a chain nuclear reaction, here we can speak of the system being in the "supercritical" state. In this case, even for a very small initial intensity and low spontaneous emission there will be an exponential growth in the radiation intensity with time, approximately proportional to $e^{\gamma t}$, where $\gamma \simeq (c/2L)(\beta\delta e^{2bL} - 1)^{-1}$. The radiation intensity will grow, but so will the number of excited atoms used up; n_1 and b will decrease as long as the system remains noncritical; the system will become critical only at a high value of I.

Thus, in a system with b positive, radiation of high intensity can be achieved. The radiation is strictly monochromatic (it is represented by a single oscillation wave), with the result that it can be focused into a point. A device that converts input power into a very narrow, intense beam of coherent visible or infrared light is called a ***laser*** (for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation) or ***maser*** (for *m*icrowave *a*mplification by *s*timulated *e*mission of *r*adiation).

The applications of a laser beam have been described in numerous articles, booklets, and books. Here, in a textbook on mathematics, we will only briefly point out two methods for obtaining inverse population, or an active medium—this is the name that physicists use to describe the situation with $n_1 > n_0$, that is, $b > 0$, which is required for operation of a laser.

One method (historically the first), which was employed by the Nobel Prize winners N. G. Basov, A. M. Prokhorov (both USSR), and Ch. Townes (USA), consists in sending molecules that are in states 1 and 0 into a capacitor placed in a vacuum, in which they move without colliding with each other under an electric field. The electrical properties, namely, the polarizabilities, of molecules in states 0 and 1 are different. Using this fact, the inventors of the laser were able to set up conditions in which the molecules in the ground state were deflected away from a vessel, while the molecules in state 1 gathered in the vessel. Thus, in the vessel, n_1 was greater than n_0, which is just the situation required for amplification of light.

The other method, which is more convenient and does not require a high vacuum but which historically was developed later than the first, consists in employing a system with several levels: 0, 1, 2, 3. The system "works," that is, emits radiation, using the transition onto one of the intermediate levels, say, the transition $3 \to 2$ with the frequency $\nu_{3,2} = (E_3 - E_2)/h$. For this to happen, we must excite state 3, that is, create a high concentration n_3. Here, however, n_3 remains lower than n_0, since it is impossible to create an inverse population, that is, to make n_3 greater than n_0, by irradiating the gas with light of frequency $\nu_{3,0}$ because of stimulated emission of light. (Why?)

But if molecules go over very rapidly from level 2 to low-lying levels, $2 \to 1$ or $2 \to 0$, there is a high probability that $n_2 \ll n_0$ and $n_2 < n_3$, that is, the system becomes supercritical with respect to the emission of light of frequency $\nu_{3,2}$.

There are also other methods for

creating an active medium. A common feature of all lasers is the transformation of input power into a highly energetic, plane, monochromatic electromagnetic wave. If such energy transfer is achieved, the energy can easily be focused into an extremely small volume. With high energy concentration it becomes possible to achieve ultrahigh temperatures and pressures, which could be used in controllable thermonuclear fusion. Today lasers are being widely used in surgery, in processing of materials, and in chemical reactions, but all this, as we have already said, belongs to books on other topics.

Chapter 13 Electric Circuits and Oscillatory Phenomena in Them

13.1 Basic Concepts and Units of Measurement

In this chapter we consider phenomena that occur in ***electric circuits***. The principal elements of an electric circuit are ***resistance***, ***capacitance***, ***inductance***, and ***sources of current (voltage)***. Our exposition is not designed to take the place of a standard physics textbook but rather to supplement, develop, and refine some of the knowledge contained in a school textbook. We will therefore confine ourselves to a brief review of the definitions of resistance, capacitance, and so forth and to their units, on the assumption that the reader is sufficiently acquainted with the basic notions from the school physics course.

The ***quantity of electricity*** is determined as the net charge, or the difference between the positive charge and the negative charge. We denote it by q. The unit for the quantity of electricity is the ***coulomb*** (abbreviated: C). The elementary charge, the charge of the proton, is $e_p = 1.6 \times 10^{-19}$ C, and the electron charge is $e_e = -1.6 \times 10^{-19}$ C.

Electric current is defined as the quantity of electricity flowing in unit time through a cross section of a conductor. We will denote current by j. The unit of current is the ***ampere*** (abbreviated: A); this is a current in which 1 coulomb of electricity passes through a cross section of a conductor in one second. Thus, 1 A = 1 C/s, but it is more convenient to say that 1 C = 1 A·s, since in the SI system of units the ampere is a base unit and the coulomb is a derived unit (see Appendix 5).

For the (positive) ***direction of current*** we take the direction in which *positive* charges would have to move in order to produce a given current. Actually, in metallic conductors positive charges are stationary, and the current flows due to the motion of electrons. As a rule, a positively charged body is one which lost part of its electrons (only in rare cases is a positive charge the result of a body acquiring positive charges). A negatively charged body is one which has acquired a surplus of electrons. The direction of current is *opposite* to that in which the electrons move in a conductor.

The ***electric potential*** of a given point is the potential energy that a positive charge of 1 C possesses when placed at the given point. The electric potential of the ground (earth) is taken to be *zero*. Hence, the point of a circuit connected to the ground by a metal conductor (we say it is grounded or earthed) has potential zero. The unit of potential is the ***volt*** (abbreviated: V). The potential of a point is equal to 1 volt (1 V) if a charge of 1 coulomb placed at this point has a potential energy of 1 joule. (The reader will recall that the joule, the unit of energy and work, is defined as the work performed by a force of 1 N over a distance of 1 m, or 1 J = 1 N·m = 1 kg·m^2/s^2.) The potential energy u of a charge q placed at a point where the potential is equal to φ is

$$u\ (\text{J}) = q\ (\text{C}) \cdot \varphi\ (\text{V}). \qquad (13.1.1)$$

We have to imagine here that q is small, because if a large charge (say 1 C) is placed at the given point, the potential φ will change. For this reason, it is better to say that the potential is the coefficient of q in (13.1.1), or the factor that connects the potential energy of a charge and the quantity of electricity.

The work A performed by a field in transferring a charge from a point where the potential is equal to φ_1 to a point where the potential is φ_2 is

$$A = u_1 - u_2 = q\ (\varphi_1 - \varphi_2).$$

Just as in mechanics only the *difference* of potential energies enters into all physical results, so in electricity the formulas always involve a *difference of potentials* but never the potential ener-

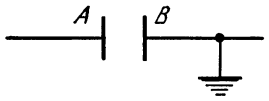


Figure 13.1.1

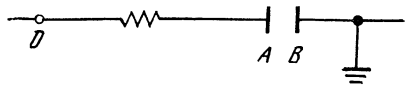


Figure 13.1.2

gies themselves. There will be no change in the potential difference if to all potentials at all points we add an identical summand. It is therefore possible to choose the potential of any point in a circuit or a piece of equipment in arbitrary fashion, say, set it equal to zero. However, after this has been done, the potentials of all other points become quite definite quantities. It is precisely for this reason that we can take the ground potential as zero.

Let us consider a ***capacitor*** (Figure 13.1.1) consisting of two parallel plates, A and B. One of the plates (say A) can be connected to some source of voltage. The quantity of electricity on this plate is directly proportional to the potential difference across the plates of the capacitor,

$$q_A = C\varphi_C,$$

where φ_C is the difference of potentials defined as the potential of plate A minus the potential of plate B. Since in Figure 13.1.1 plate B is grounded, it follows that φ_C in this case is equal to the potential of plate A.

The coefficient of proportionality C is called the ***capacitance*** of the capacitor. The unit of capacitance is the ***farad*** (abbreviated: F). It is the capacitance of a capacitor in which the potential difference across the plates is 1 volt for a charge of 1 coulomb (10^{-6} F is a microfarad (μF), 10^{-9} F a nanofarad (nF), and 10^{-12} F a picofarad (pF)).

An equal quantity of electricity (but opposite in sign) accumulates (builds up) on plate B:

$$q_B = -q_A = -C\varphi_C.$$

The electric charge of a closed system is a conserved quantity, that is, electric charges of the same sign never appear or disappear in any process whatsoever.[13.1] A change in the charge on plate A of the capacitor is due to the fact that a portion of the charge left the plate and moved to some other site, say, point D in Figure 13.1.2.

If a current j is flowing from D to A (from left to right), then in time dt a quantity of electricity $j\,dt$ will flow through the cross section of the conductor, whence

$$dq_A = j\,dt, \text{ or } \frac{dq_A}{dt} = j.$$

Let us now find out what a current flowing in a conductor depends on. By Ohm's law the current is proportional to the potential difference across the terminals of the conductor, the current flowing from higher to lower potential. Thus,

$$j = k(\varphi_D - \varphi_A) = \frac{1}{R}(\varphi_D - \varphi_A). \quad (13.1.2)$$

The positive quantity k is called the ***conductance***. The reciprocal, $1/k$, is called the ***resistance*** of a conductor and is denoted by R. The unit of resistance is the ***ohm*** (abbreviated: Ω), which is the resistance of a conductor through which a current of 1 ampere is flowing when a potential difference of 1 volt is impressed across the terminals, so that 1 Ω = 1 V/A.

We denote the quantity $\varphi_D - \varphi_A$ by φ_R. This is the potential across the resistance R. The value of φ_R is defined (just as like that of φ_C) as the left-hand potential minus the right-hand potential (the current flows from left to right). Ohm's law (13.1.2) can then be written as

$$j = \frac{\varphi_R}{R}, \text{ or } \varphi_R = Rj. \quad (13.1.3)$$

[13.1] The net electric charge of a system remains unchanged when two particles of equal and opposite charge appear or disappear.

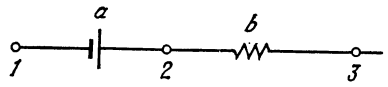


Figure 13.1.3

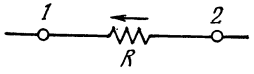


Figure 13.1.4

As a ***source of voltage*** in a circuit we can take a voltaic cell. There is a definite potential difference across the terminals of the cell. We can assume, roughly, that the potential difference is independent of the current flowing through the cell. In particular, in a cell the current can flow from a low potential to a high potential. Through a resistance, the current always flows from a high potential to a low potential, like water in a tube connecting two vessels flows from high level to low level.

A voltaic cell is like a pump that can take in water in a low-level vessel and pump it up to a high-level vessel, that is to say, make the water move uphill. To operate the pump we need some kind of external source of energy. The same applies to the cell. When the current flows from low to high potential, chemical reactions take place in the cell. The energy of these chemical reactions in the cell is transformed into electric energy.

The potential difference which the cell yields is termed the ***electromotive force***, which we abbreviate to emf.

The potential difference across the cell taken as the left-hand potential minus the right-hand potential (Figure 13.1.3) is equal to minus the electromotive force of the cell:

$$\varphi_1 - \varphi_2 = -E.$$

In reality, the emf is slightly dependent on the current flowing through the cell. When the current is flowing (from left to right in Figure 13.1.3) in the direction from low potential to high potential (which is the normal operating conditions of the cell when it is generating electric energy), the emf E diminishes with increasing current flow. Approximately, we can take it that the emf is constant, but more exactly we must allow for a linear dependence of E on j:

$$E = a - bj, \qquad (13.1.4)$$

where the (positive) coefficients a and b are characteristics of the cell. We call a cell whose emf is *independent* of the current j, that is, $b = 0$ in (13.1.4) an ***ideal cell***.

Let us consider a series connection of an ideal cell with an emf equal to a and a resistance b (see Figure 13.1.4). Then

$$\varphi_E = \varphi_1 - \varphi_2 = -a,$$

$$\varphi_b = \varphi_2 - \varphi_3 = bj;$$

whence

$$\begin{aligned}\varphi_1 - \varphi_3 &= (\varphi_1 - \varphi_2) + (\varphi_2 - \varphi_3)\\ &= -a + bj = -(a - bj)\\ &= -E.\end{aligned}$$

We have again arrived at formula (13.1.4) for the emf E, whereby b is often called the ***internal resistance*** of the cell. The real cell, with which we usually deal, which has an emf that satisfies formula (13.1.4), yields the same dependence of E on j as the series connection of an ideal cell and a resistance b. The name for a remains the same, the emf of a real cell, bearing in mind that $E = a$ when $j = 0$, and the emf drop for $j \neq 0$ is characterized by b.[13.2]

In the sequel, when considering electric circuits involving current sources, for example a cell, and various resistances, we can imagine that we are dealing with an ideal cell with constant emf independent of the current, while the internal resistance b may be combined with the external resistance R. Thus, a real cell with internal resistance b connected in series with a resistance R

[13.2] Since the current is zero for an open circuit, the emf may be defined as the potential difference across a disconnected cell.

is equivalent to an ideal cell connected in series with a resistance $R_1 = R + b$. The current flowing in the circuit is then given by the formula

$$j = \frac{E}{R+b}.$$

It is worth once again paying special attention to the difference between resistance and source of voltage. If in a circuit there is a potential difference across a resistance, such that $\varphi_2 > \varphi_1$ (Figure 13.1.4), then, by our definition, $\varphi_R = \varphi_1 - \varphi_2 < 0$, that is, φ_R is negative. Hence, by formula (13.1.3), the current is negative, too, which means that the current flows from right to left, from point *2* to point *1*. Now suppose that there is a potential difference of the same sign across the terminals of the voltage source, and the dependence of E on j is given by formula (13.1.4) (see Figure 13.1.3). Here, let $\varphi_3 > \varphi_1$ but $\varphi_3 - \varphi_1 < a$. Then $bj = \varphi_1 - \varphi_3 + a = a - (\varphi_3 - \varphi_1) > 0$, that is, $j > 0$. Therefore, the current flows from left to right despite the fact that the potential φ_1 on the left is less than the potential φ_3 on the right. Thus, the voltage source is capable of overcoming the potential difference and yielding a positive current (from left to right) for a negative potential difference ($\varphi_1 - \varphi_3 < 0$), provided that this negative potential difference does not exceed in absolute value the emf of the source. Yet for a negative potential difference, the resistance always yields a negative current. In any circuit containing only resistance, that is, circuits without emf, the current can only be identically zero, while if there are cells, that is, emf's (Figure 13.1.5), a nonzero solution (current) is possible.

Now let us consider ***inductance***. The phenomenon of inductance is connected

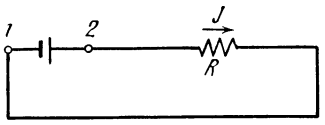


Figure 13.1.5

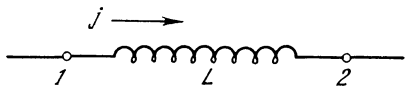


Figure 13.1.6

with the magnetic field that is produced in the space surrounding a conductor that carries a current. The magnetic field is particularly high if the conductor has the form of a coil with a large number of turns. The field is further increased if the coil is wound on an iron core.

The magnetic field, in turn, gives rise to electric phenomena (if the magnetic field varies). As we know, given a varying magnetic field, each turn (even every portion of a turn) of the coil becomes a source of voltage, something like a voltaic cell. In a coil in which the turns are wound so that the current traverses the core of the coil in the same direction throughout the length of the coil, all these voltage sources are connected in series so that the overall voltage builds up (the voltages are additive in a series connection).

On the whole, a coil is equivalent to a voltage source with a potential difference proportional to the rate of change of the magnetic field. But the magnetic field in a coil is proportional to the current flowing in the coil.[13.3] For this reason, the rate of change of the magnetic field is proportional to the *rate* of change of current flow, that is, to the *derivative* dj/dt. Referring to Figure 13.1.6, we find that in the coil

$$\varphi_L = \varphi_1 - \varphi_2 = L\frac{dj}{dt}, \qquad (13.1.5)$$

and the positive direction of the current is taken to be from point *1* to point *2* inside the coil, while the quantity φ_L is the potential difference across the

[13.3] We will not discuss the case of two coils wound on one core, which is a transformer connecting two electric circuits that carry different currents. Neither do we consider cases of a more complicated dependence of the magnetic field on the current when an iron core is inserted in the coil and the current is so great that the iron is saturated.

coil (it is defined as the potential φ_1 on the left minus the potential φ_2 on the right). When considering in detail the direction of the magnetic field and the emf induced by its variation, it is found that the coefficient L (the ***inductance***) is always positive.

From formula (13.1.5) it follows that if dj/dt is negative, then $\varphi_1 - \varphi_2$ is negative, too, that is, $\varphi_2 > \varphi_1$. Thus, if the current is positive (flows from *1* to *2*) and decreases in magnitude, the coil plays the role of a cell sustaining a positive current in the circuit, despite the fact that φ_L is negative. But if the current is positive and increasing, dj/dt is positive, and so φ_L is positive, too. In this case, the coil plays the part of an additional resistance, since the potential difference across the coil is positive for a positive current (cf. (13.1.3)).

A coil differs essentially from a voltage source and from a resistance in that the quantity φ_L depends not on the current intensity j but on the rate of change of the current, dj/dt.

The coefficient L in Eq. (13.1.5) bears the name "coil inductance" (also self-inductance).[13.4] The unit of inductance is the ***henry*** (abbreviated: H). If the inductance of a coil is equal to 1 henry, this means that when the current is changing at a rate of 1 ampere per second, a potential difference of 1 volt is induced in the coil. We obtain the dimensions of inductance from formula (13.1.5):

$$1\ \text{H} = 1\ \text{V}\cdot\text{s/A}.$$

From the foregoing it is clear that inductance influences the current in a circuit just like an inert mass (flywheel) affects velocity: inductance impedes any change in the current, and a mass (by Newton's second law) tends to impede any change in velocity. This similarity will be discussed in more detail in Section 13.8.

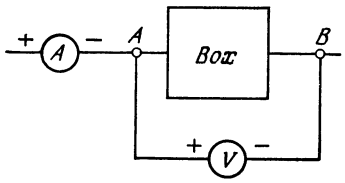


Figure 13.1.7

From the standpoint of subsequent computations, capacitance, resistance, emf, and inductance have one thing in common, they all require two terminals for connection in a circuit (unlike, say, a transformer, which requires four leads, or a transistor, which has three leads: collector, base, emitter). Devices with circuit connections involving two leads are called ***two-terminal networks***, while if there are four leads, they are ***two-terminal pair networks*** (or four-pole networks). Each circuit element—capacitance, resistance, emf, and inductance—is characterized at any given time by a specific current passing through it and a definite potential difference at input and output.

We can imagine a closed box (labelled "Box" in Figure 13.1.7) with two leads, A and B, sticking out of it. The interior of the box may contain anything: R, E, L, and C. Connect an ammeter A and a voltmeter V. With the circuit connections as shown in Figure 13.1.7 (the "+" and "—" signs correspond to the labels at the terminals of the ammeter and voltmeter), the ammeter records the current j passing in the direction from point A to point B, the voltmeter indicates the potential difference $\varphi_{\text{Box}} = \varphi_A - \varphi_B$. The relationship between φ_{Box} and j depends on what is inside the box:

in the case of a resistance R,

$$\varphi_{\text{Box}} = Rj, \tag{13.1.6}$$

in the case of an emf[13.5] E_0

$$\varphi_{\text{Box}} = -E_0, \tag{13.1.7}$$

[13.4] Often instead of saying "a coil with inductance L" we simply say "inductance L", just as we say "capacitance C" instead of "a capacitor with capacitance C", or emf E instead of mentioning a voltaic cell or a voltage source.

[13.5] The internal resistance of the emf source is disregarded.

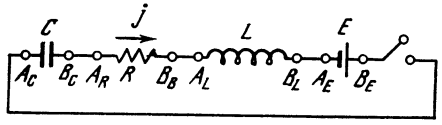


Figure 13.1.8

in the case of an inductance, L,

$$\varphi_{\text{Box}} = L\frac{dj}{dt}, \qquad (13.1.8)$$

in the case of a capacitance[13.6] C,

$$\varphi_{\text{Box}} = (\varphi_{\text{Box}})_0 + \frac{1}{C}\int_{t_0}^{t} j\,dt, \qquad (13.1.9)$$

or

$$\frac{d\varphi_{\text{Box}}}{dt} = \frac{1}{C}j. \qquad (13.1.9a)$$

There are of course cases in which more complicated relationships are involved. For example, a ***rectifier*** (a vacuum-tube diode or a semiconductor diode) does not fit any of the formulas (13.1.6)-(13.1.9). However, in a large number of important problems we can confine ourselves to considering the circuit elements for which these formulas are valid to a high degree of accuracy. These are the circuits that we will investigate (with the exception of Section 17, where we give special consideration to the properties of a circuit with a device that exhibits a complicated relationship between current and potential difference).

Let us study the circuit shown in Figure 13.1.8. We will first write down the voltage drops on the separate elements of the circuit:

$$\varphi_C = \varphi_{A_C} - \varphi_{B_C}, \quad \varphi_R = \varphi_{A_R} - \varphi_{B_R}, \qquad (13.1.10)$$

$$\varphi_L = \varphi_{AL} - \varphi_{BL}, \quad \varphi_E = \varphi_{AE} - \varphi_{BE}.$$

And observe that $\varphi_{BC} = \varphi_{AR}$, $\varphi_{BR} = \varphi_{AL}$, and $\varphi_{BL} = \varphi_{AE}$. Whence, adding all the equations in (13.1.10) termise, we obtain

$$\varphi_C + \varphi_R + \varphi_L + \varphi_E = \varphi_{AC} - \varphi_{BE}.$$

If the circuit in Figure 13.1.8 is closed, then $\varphi_{AC} = \varphi_{BE}$. In this case, consequently,

$$\varphi_C + \varphi_R + \varphi_L + \varphi_E = 0. \qquad (13.1.11)$$

This general equation, together with the expressions (13.1.6) to (13.1.9), fully describes all the processes that occur in the circuit. Below we will use this equation to examine a variety of circuits beginning with the very simplest which consists of only *two* elements.

13.2 Discharge of a Capacitor Through a Resistor

Let us examine the process in a circuit with capacitance C and resistance R (Figure 13.2.1). We denote by φ the potential of point A (plate B of the capacitor will be grounded). To begin with, let $\varphi = \varphi_0$. The corresponding quantity of electricity on plate A is $q_A = C\varphi_0$.

Can we speak of a current flowing through a capacitor? A capacitor consists of two plates separated by an insulator (say, air) so that in reality an electron cannot go through the capacitor, say, from A to B. However, if a positive charge is impressed on plate A, then plate B will have a negative charge, and a positive charge will flow out of plate B along the wire (the current also goes from left to right). Two ammeters, A_1 and A_2, one of which measures current in the wire connected to plate A, the other in the wire connected to plate B, will have identical readings. What precisely is it that flows

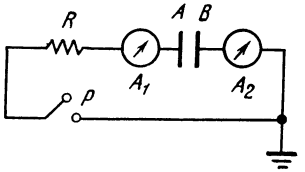


Figure 13.2.1

[13.6] Here $q_A = C\varphi_{\text{Box}}$ and $dq_A/dt = j$, whence $d\varphi_{\text{Box}}/dt = j/C$. If at the initial time $t = t_0$ we have $\varphi_{\text{Box}} = (\varphi_{\text{Box}})_0$, then

$$\varphi_{\text{Box}} = (\varphi_{\text{Box}})_0 + C^{-1}\int_{t_0}^{t} j\,dt.$$

(positive charges or electrons) through different portions of the electric circuit does not interest us, just as we are not interested in whether the same electrons pass through A_2 that have passed through A_1 or not. For this reason we will henceforth only speak of *the current passing through the capacitor* and will have in mind the current flowing in the conductors connected to the plates of the capacitor. We can speak of current flowing in an electric circuit through a capacitance in the same way that we speak of current flowing through a resistance or an inductance. The difference lies in the different type of relationship between current and potential difference as expressed by the formulas (13.1.9) and (13.1.9a).

When we close the switch P (see Figure 13.2.1), a current

$$j = \frac{1}{R}\varphi_R$$

will flow through resistance R. By Eq. (13.1.11), $\varphi + \varphi_R = 0$, whence $\varphi_R = -\varphi$ and so

$$j = -\frac{1}{R}\varphi. \qquad (13.2.1)$$

Since current flowing from left to right is taken to be positive, it follows from (13.2.1) that for $\varphi > 0$ the current is negative, it flows from right to left, and the capacitor becomes discharged.[13.7] Recalling that $j = dq/dt$ (current flowing through a capacitance) and $q = C\varphi$, we find that

$$j = C\frac{d\varphi}{dt}. \qquad (13.2.2)$$

Comparing (13.2.1) and (13.2.2), we obtain

$$\frac{d\varphi}{dt} = -\frac{1}{RC}\varphi. \qquad (13.2.3)$$

[13.7] Observe that in all circuits having the form of a rectangle (see Figure 13.1.8 and subsequent figures), we speak of the direction of current in the *upper* side of the rectangle; the current flow in the bottom side that closes the circuit will clearly be in the *opposite* direction.

We solved a differential equation like this in connection with the problem of radioactive decay (see Chapter 8) and in many other sections of this book. If $\varphi = \varphi_0$ when $t = 0$, then

$$\varphi(t) = \varphi_0 e^{-t/RC}, \qquad (13.2.4)$$

whence

$$j(t) = -\frac{\varphi_0}{R}e^{-t/RC}.$$

It can be seen from formula (13.2.3) that the quantity RC has the dimensions of *time*. Let us verify this:

$$[R] = \Omega = \text{V/A} = \text{V}\cdot\text{s/C}, \quad [C] = \text{C/V},$$

whence

$$[RC] = \frac{\text{V}\cdot\text{s}}{\text{C}}\,\frac{\text{C}}{\text{V}} = \text{s}.$$

During time $t = RC$ the charge q on the capacitor and the current j diminish by a factor of e.

The discharging process in a capacitor can easily be observed experimentally. Buy a capacitor with capacitance $C = 20\ \mu\text{F} = 20 \times 10^{-6}$ F and a resistor with resistance $R = 20\ \text{M}\Omega = 20 \times 10^6\ \Omega$. For an RC circuit of this type we get $RC = 400$ s, which is a very convenient time for observational purposes.

The quantity RC is known as the ***time constant*** of a circuit consisting of a capacitance and a resistance (recall that in the case of radioactive decay the ***mean lifetime*** was an analogous quantity).

We will consider the problem of ***charging*** a capacitor through a resistor. The circuit diagram is shown in Figure 13.2.2. If switch P is closed, then, by (13.1.11), $\varphi_E + \varphi_R + \varphi = 0$, where φ is the potential of the nongrounded plate of the capacitor. Since $\varphi_E = -E_0$ and $\varphi_R = Rj$, it follows that $-E_0 +$

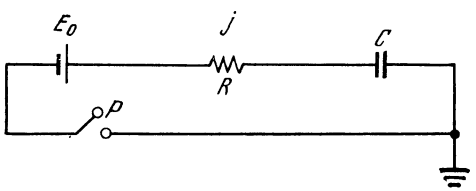


Figure 13.2.2

$Rj + \varphi = 0$. The current flowing through the capacitance is $j = dq/dt = C\,(d\varphi/dt)$, whereby

$$-E_0 + RC\frac{d\varphi}{dt} + \varphi = 0, \text{ or}$$

$$\frac{d\varphi}{dt} = -\frac{1}{RC}(\varphi - E_0). \qquad (13.2.5)$$

To find out how φ varies with time, it will be convenient to make the change of variable, $z = \varphi - E_0$, with $dz = d\varphi$.

Equation (13.2.5) can then be rewritten thus:

$$\frac{dz}{dt} = -\frac{z}{RC}.$$

Its solution is

$$z = z_0 e^{-t/RC}, \qquad (13.2.6)$$

where z_0 is the value of z at the initial time.

Let us find the solution for the case where at the initial time the capacitor is not charged: $\varphi = 0$ at $t = 0$. Then $z_0 = -E_0$. From (13.2.6) we get $z = -E_0 e^{-t/RC}$, or

$$\varphi = z + E_0 = -E_0 e^{-t/RC} + E_0$$
$$= E_0\,(1 - e^{-t/RC}). \qquad (13.2.7)$$

The graph of φ as a function of t is given in Figure 13.2.3. The curve corresponds to the formula (13.2.7), while the dashed horizontal line represents the value $\varphi = E_0$ which the solution approaches with the passage of time. The quantity z has the geometric meaning of vertical distance from the curve to the dashed line. This distance diminishes exponentially with the passage of time.

During a time equal to RC, the charge of the capacitor reaches 63% of its final value, during time $2RC$ it reaches 86%, and during time $3RC$ it reaches 95% of its final value.

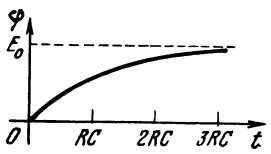


Figure 13.2.3

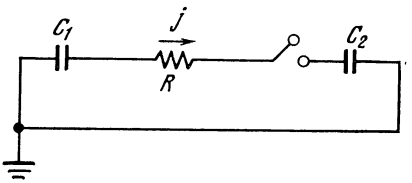


Figure 13.2.4

From formulas (13.2.4) and (13.2.7) it is evident that charging and discharging the capacitor is the faster the smaller the resistance R and the smaller the capacitance C.

Exercises

13.2.1. Referring to Figure 13.2.1, $C = 10^{-6}$ F and $R = 10^7\ \Omega$, $R = 10^8\ \Omega$, $R = 10^9\ \Omega$. For each of these cases, determine the time lapse during which the current flowing through the capacitor at the initial time falls off by 10%; decreases by a factor of 2.

13.2.2. Consider the process of equalizing the potential across a resistance R in series with *two* capacitors, C_1 and C_2, one of which at time $t = 0$ is charged to a potential difference $\varphi_{C_1}(0) = a$, while the other is not charged at all, that is, $\varphi_{C_2}(0) = 0$ (Figure 13.2.4).

13.2.3. Determine the variation of the time constant of the circuit depicted in Figure 13.2.1 if all linear dimensions of the circuit diagram are increased n-fold (for the case of a plane-parallel capacitor). (The condition of the problem is to be understood in this way: the dimensions of the capacitor and the resistance are increased but the materials of which they are made are not changed.) [*Hint.* The formula for the capacitance of a plane-parallel capacitor is known from physics: $C = \varepsilon\, S/4\pi d$, where S is the area of a plate of the capacitor, d the distance between the plates, and ε a constant characterizing the material between the plates (the ***dielectric constant***). The resistance of a wire resistor is found from the formula $R = \rho l/\sigma$, where l is the length of the wire, σ the cross-sectional area of the wire, and ρ a constant characterizing the material of the wire.]

13.3 Oscillations in a Capacitance Circuit with Spark Gap

A typical circuit diagram involving a capacitance is shown in Figure 13.3.1. The circuit includes a voltage source with emf E and resistance R (the role of R may be played by the internal resistance of the voltage source) Under-

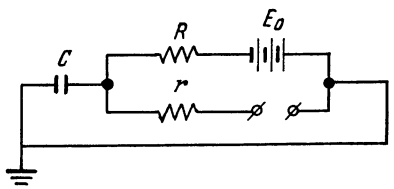


Figure 13.3.1

neath is a ***spark gap***. For a potential difference less than a certain value φ_1, the spark gap is an insulator. At $\varphi = \varphi_1$ a spark jumps the gap and the air between the wires heats up and becomes a good conductor. We denote the total resistance of the leads and the incandescent air by r. The quantity r is small and remains small as long as a current flows maintaining a high temperature of the air. For a definite small value of current j_2 the air cools and the spark gap again becomes an insulator. This current value is associated with the potential difference $\varphi_2 = j_2 r$. Here $\varphi_1 > \varphi_2$: a higher voltage is needed to initiate a spark than to keep it burning.

Figure 13.3.2 shows the dependence of φ on t for such a circuit. The capacitor is charged over the section OA, with no current flowing through the spark gap. In this case the solution (13.2.6) is valid:

$$\varphi = E\,(1 - e^{-t/RC}). \qquad (13.3.1)$$

The potential difference at point A at time $t = t_A$ reaches the value φ_1, the spark gap begins to conduct current, and the capacitor discharges. Since in this case $R \gg r$, the current from the voltage source can be ignored as compared to the current passing

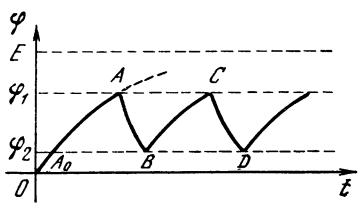


Figure 13.3.2

through the spark gap. Therefore, for φ we get the equation

$$\frac{d\varphi}{dt} = -\frac{\varphi}{rC},$$

with $\varphi = \varphi_1$ at $t = t_A$, whence we obtain

$$\varphi = \varphi_1 e^{-(t-t_A)/rC}. \qquad (13.3.2)$$

At time $t = t_B$ (point B), $\varphi = \varphi_2$, the spark gap again becomes an insulator, the charging process is initiated (section BC), and so on.

Let us determine the time $t_B - t_A$ during which the capacitor discharges. To do this, we take advantage of the fact that $\varphi = \varphi_2$ at $t = t_B$. Putting $\varphi = \varphi_2$ and $t = t_B$ in (13.3.2), we get

$$\varphi_2 = \varphi_1 e^{-(t_B - t_A)/rC},$$

whence

$$t_B - t_A = rC \ln (\varphi_1/\varphi_2).$$

Over segment BC (charging) the relation (13.3.1) displaced in time by the amount τ holds true (in Figure 13.3.2, τ is depicted by the line segment A_0B). For this reason

$$\varphi = E\,(1 - e^{-(t-\tau)/RC}).$$

Putting $t = t_B$, we get

$$\varphi_2 = E\,(1 - e^{-(t_B - \tau)/RC}).$$

Similarly, setting $t = t_C$, we find that

$$\varphi_1 = E\,(1 - e^{-(t_C - \tau)/RC}).$$

Combining the last two formulas, we get

$$\frac{E - \varphi_2}{E - \varphi_1} = e^{-(t_C - t_B)/RC}, \quad \text{or}$$

$$t_C - t_B = RC \ln \frac{E - \varphi_2}{E - \varphi_1}.$$

The complete period (the charge-discharge cycle) is

$$T = t_C - t_A = (t_C - t_B) + (t_B - t_A)$$
$$= RC \ln \frac{E - \varphi_2}{E - \varphi_1} + rC \ln \frac{\varphi_1}{\varphi_2}.$$

Ordinarily, the resistance R in the circuit of the voltage source is many times greater than that of the spark gap, and for this reason the charging

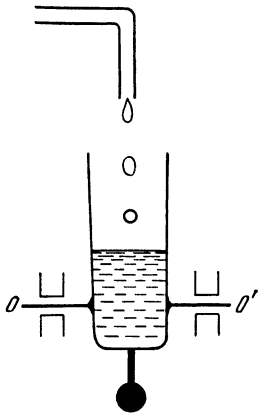


Figure 13.3.3

time is much longer than the time of discharge. On the other hand, the discharge current is many times greater than the charging current, greater than the maximum current obtainable from the voltage source (with an internal resistance of R_1 the voltage source does not produce a current exceeding E/R_1). The circuit shown in Figure 13.3.1 transforms a long-term small current generated by the voltage source into a strong current, which, however, is of short duration (it is customary to speak of "short pulses" of current).

This circuit operates like a system in which a tiny flow of water gradually fills a vessel (Figure 13.3.3). The vessel is fixed in such a manner that when a sufficient amount of water has accumulated, the vessel turns over and the water pours out. The vessel then rights itself and the process begins anew. In the figure, the vessel is fixed on a horizontal axis OO' below the midpoint. A weight is attached at the bottom of the vessel so that the center of gravity of the empty vessel lies below the axis. But when the vessel fills up with water, the center of gravity of the full vessel lies above this axis and the vessel tips over.

Let us return to the circuit diagrams in Figures 13.2.1 and 13.2.2. In these circuits, which consist of capacitances, resistances, and emf sources, the potentials even out in the course of time. Indeed, in Figure 13.2.1 the value $\varphi = 0$ sets in, and in Figure 13.2.2, $\varphi = E_0$ is the steady-state value (see formulas (13.2.4) and (13.2.7)).[13.8] The situation is quite different in the case of a spark-gap circuit. Here we have undamped oscillations of φ (true, they are very different from those that we studied earlier). These oscillations are connected with certain specific properties of the spark gap, in particular with the fact that until a definite potential is reached (the ***breakdown potential*** φ_1), no current flows through the spark gap.

Many books have been written about the properties of discharge through air in a spark gap. All we have given here is a smattering of information—only enough to understand the operation of the circuit shown in Figure 13.3.1. This information does not even suffice to answer the simple question: What will happen if we connect the spark gap to a voltage source without a capacitor?

Indeed, if no current flows, the voltage across the spark gap will be E_0. Since E_0 is greater than φ_1, breakdown should occur. But if this occurred, the resistance of the spark gap would become small, equal to r. Then a potential difference equal to $E_0 r/(r + R)$ would appear across the spark gap and the current would be $j = E_0/(r + R)$. If R is great, the current j is small, less than j_2, and the potential difference across the spark gap is small, less than φ_2. But then the air will not heat up and the resistance of the spark gap will not become the small quantity r, which means the potential difference will be great and equal to E. We have a contradiction.

Actually, under these conditions we have an electric discharge of a different type, the ***glow discharge*** (small current without heating of the air), instead of the spark with incandescent air.

[13.8] Below we will see that the potential tends to a steady-state value as $t \to \infty$ in a circuit with an inductance, too.

13.4 The Energy of a Capacitor

A charged capacitor has a definite supply of energy, which can be given up very quickly if the capacitor is discharged through a small resistance.

Let us find the supply of energy of a capacitor of capacitance C, one plate of which is grounded and the other has a potential φ_0. Then the quantity of electricity $q_0 = C\varphi_0$.

It would appear at first glance that the energy is equal to the product $q_0\varphi_0$ $(= C\varphi_0^2)$. In reality, this expression is not exact, though it is correct as to order of magnitude: it differs from the true value by a factor of 2.

Let us consider the charging of the capacitor. When the capacitor's potential is φ and the charge is q, the addition of a small quantity of electricity dq increases the energy by

$$dW = \varphi\, dq. \qquad (13.4.1)$$

The essential thing is that during charging the potential φ changes, since $\varphi = q/C$. Substituting this value of φ into (13.4.1), we get

$$dW = \frac{1}{C} q\, dq. \qquad (13.4.2)$$

Integrating (13.4.2) from $q = 0$ (uncharged capacitor) to $q = q_0$, we get

$$W(q_0) = \frac{1}{C}\int_0^{q_0} q\, dq = \frac{1}{2}\frac{q_0^2}{C}$$

$$= \frac{1}{2}\varphi_0 q_0 = \frac{1}{2} C\varphi_0^2. \qquad (13.4.3)$$

Thus an exact evaluation yields the coefficient 1/2.

Now let us examine the charging of a capacitor from a voltage source through a resistance (see Figure 13.2.2). The voltage source has a constant emf, E_0. Therefore, when a quantity of electricity dq flows, the voltage source does work $E_0\, dq$ (this work is performed at the expense of the chemical energy of the voltage source, that is, the chemical energy diminishes). Hence, the total work done by the voltage source is equal to $E_0 q_0$, where q_0 is the total quantity of electricity that has flowed. When the capacitor is being charged, the process stops at $\varphi = E_0$. In this process, the voltage source will have performed work

$$E_0 q_0 = E_0 C E_0 = C E_0^2.$$

What supply of energy will the capacitor possess? This can readily be computed from formula (13.4.3):

$$W = \frac{1}{2} C E_0^2.$$

Where has half the work performed by the source gone? We will show that it went to heat up the resistance R. Recall that if a quantity dq of electricity flows through a resistance, the energy released will be

$$dA = \varphi_R\, dq, \qquad (13.4.4)$$

where φ_R is the potential difference across the resistance. Using the fact that $dq = j\, dt$ and $j = \varphi_R/R$, we can transform (13.4.4) to the familiar form

$$dA = \frac{\varphi_R^2}{R}\, dt = j^2 R\, dt.$$

The quantity $j^2R = \varphi_R^2/R$ is the amount of energy released on the resistance in unit time, which is to say, it is the ***power output*** of the resistance. It is measured, as expected, in watts:

$$[j^2R] = \mathrm{A}^2\cdot\Omega = \mathrm{A}^2\cdot(\mathrm{V/A}) = \mathrm{A}\cdot\mathrm{V}$$
$$= \mathrm{A}\cdot(\mathrm{W/A}) = \mathrm{W}.$$

The time dependence of j in the case of the charging of a capacitor through a resistance was found in Section 13.2:

$$j(t) = \frac{E_0}{R} e^{-t/RC}.$$

Therefore $dA = (E_0^2/R)\, e^{-2t/RC} dt$.

The energy released during time T is

$$A(T) = \frac{E_0^2}{R}\int_0^T e^{-2t/RC}\, dt,$$

whence

$$A(T) = -\frac{CE_0^2}{2} e^{-2t/RC}\Big|_0^T$$

$$= \frac{1}{2} CE_0^2 (1 - e^{-2T/RC}). \qquad (13.4.5)$$

We know that as the time interval T increases without limit, the potential φ approaches the value E_0. As can be seen from (13.4.5), A then approaches $CE_0^2/2$. Therefore, the total energy released on the resistance is

$$A = \frac{1}{2} CE_0^2. \qquad (13.4.6)$$

Thus, calculations confirm the fact that in the charging of a capacitor half the energy is lost on the resistance. The efficiency of charging is only 50%. Note that if we directly connect the voltage source to the capacitor, nothing will change, the efficiency remains at 50%, the role of resistance being taken by the internal resistance of the voltage source, which will then heat up. From formula (13.4.6) it is evident that the energy lost uselessly on the resistance in charging the capacitor is independent of the magnitude of R and hence does not depend on how fast the charging takes place.

Since R did not appear in (13.4.6), this formula may be obtained without introducing R into the intermediate transformations. Indeed, for the circuit of Figure 13.2.2, $\varphi_E + \varphi_R + \varphi_C = 0$, whence $\varphi_R = -\varphi_E - \varphi_C = E_0 - q/C$. Therefore $dA = (E_0 - q/C)\, dq$. Integrating this expression from $q = 0$ to $q = q_0 = E_0 C$ yields

$$A = \frac{1}{2} CE_0^2.$$

The last derivation holds true also for the case where the resistance R varies with time, while the previous derivation held true only for R constant, since only in this case could the formulas of Section 13.2 be applied.

In order to reduce losses in charging a capacitor, we should have done as follows: first take a voltage source with small emf E_1 and charge the capacitor to the potential E_1, then disconnect the first voltage source and connect a second source with greater emf E_2. Having charged the capacitor to potential E_2, disconnect the second source and connect a third with emf E_3,

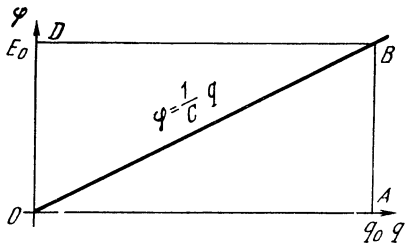


Figure 13.4.1

and so on. The gain is clearly seen in the graph: lay off the charge q of the capacitor on the axis of abscissas and the potential φ on the axis of ordinates. The two are connected by the relation $\varphi = q/C$, which represents a *straight* line (Figure 13.4.1). The energy of a capacitor is equal to the area of the triangle OAB. The work done by the voltage source is equal to the area of the rectangle $OABD$. The energy lost on the resistance is equal to the area of the triangle ODB.

If the capacitor is charged in stages, the sum of the works performed by all voltage sources is equal to the shaded area in Figure 13.4.2. We leave it to the reader to find the efficiency for the case where the charging process is divided into n stages:

$$E_1 = \frac{\varphi}{n}, \quad E_2 = \frac{2\varphi}{n},$$

$$E_3 = \frac{3\varphi}{n}, \quad \ldots, \quad E_n = \varphi.$$

In the foregoing, one plate of the capacitor was grounded, that is, its potential was $\varphi_1 = 0$. Here, the energy of the capacitor depends on the po-

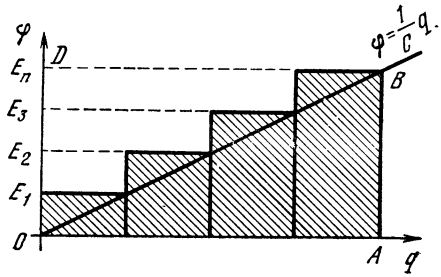


Figure 13.4.2

tential of the other plate, φ_2, that is,

$$W = \frac{1}{2} C\varphi_2^2.$$

If neither plate is grounded, the energy of the capacitor depends on the potential difference across the plates, φ_C, that is, $W = \frac{1}{2} C\varphi_C^2$. Indeed, we know that the charge q on each plate of the capacitor depends on the potential difference, the charges on the plates being equal in magnitude but opposite in sign:

$$q_A = C\varphi_C, \quad q_B = -C\varphi_C = -q_A,$$
$$dq_A = -dq_B.$$

When computing the variation of energy in the charging process, one has to take into account the variation of charge on both plates. Let the potential of plate A be φ_1, the potential of plate B be φ_2, and $\varphi_1 - \varphi_2 = \varphi_C$. Then

$$dW = \varphi_1\, dq_A + \varphi_2\, dq_B = \varphi_1\, dq_A$$
$$- \varphi_2\, dq_A = (\varphi_1 - \varphi_2)\, dq_A$$
$$= \varphi_C\, dq_A.$$

But since $\varphi_C = q_A/C$, we can write

$$dW = \frac{1}{C} q_A\, dq_A. \qquad (13.4.7)$$

Integrating (13.4.7) from 0 to q_A yields

$$W = \frac{q_A^2}{2C} = \frac{1}{2} C\varphi_C^2.$$

Knowing the expression for the energy of a charged capacitor as a function of the capacitance, we can find the mechanical forces acting between the plates of the capacitor. Imagine the plates to be connected mechanically with some kind of lever and suppose that the capacitance C depends on the position of the lever. If the position of the lever is characterized by the value of the x coordinate, the capacitance is a function of x, or $C = C(x)$. At a definite position x_0 of the lever the capacitance is $C(x_0) = C_0$. If in this position the capacitor is charged to the potential φ_0, the charge on the plates is $q_0 = C_0\varphi_0$ and the energy of the capacitor is

$$W = \frac{C_0\varphi_0^2}{2} = \frac{q_0^2}{2C_0}.$$

Let us disconnect the capacitor from the voltage source and move the lever. The charge will then remain constant (the potential varies in inverse proportion to the capacitance), and the energy will change:

$$W(x) = \frac{q_0^2}{2C(x)}.$$

The electric energy of a charged capacitor is similar to the elastic energy of a spring: $W(x)$ increases if an external force applied to the lever does work. Then the external force overcomes the forces with which the plates of the capacitor act on the lever. Contrariwise, if $W(x)$ decreases, the lever is displaced and does work in opposition to the external applied forces. We may conclude that the force with which the plates act on the lever is

$$F = -\frac{dW}{dx} = -\frac{d}{dx}\frac{q_0^2}{2C(x)}$$
$$= \frac{q_0^2}{2[C(x)]^2}\frac{dC(x)}{dx} = \frac{\varphi^2(x)}{2}\frac{dC(x)}{dx}. \qquad (13.4.8)$$

The force is directed toward increasing capacitance. For instance, if the capacitor consists, say, of two identical parallel plates, the capacitance is in inverse proportion to the distance between the plates. This means the capacitance increases when the plates are brought closer together. Quite true, because when a capacitor is charged, the charges on the plates are of opposite sign and so the plates attract with more force if they are close together.

Formula (13.4.8) enables us to find the force in more complicated situations, as, for example, in the case of a variable capacitor, in which one plate can move in and out between two fixed plates.

It is important to note that we took the derivative dW/dx for a given con-

stant charge q. However, it is not permissible, when seeking the force by the formula $F = -dW/dx$, to take the derivative of $W = C(x)\,\varphi^2/2$ assuming φ constant and having regard solely for the fact that C depends on x. We would then obtain an incorrect sign for the force. Indeed, if the capacitor is disconnected from the voltage source, φ is not constant: $\varphi = q/C$ and $C = C(x)$. If the capacitor is connected to a voltage source, then φ remains constant as the capacitance varies. But then the charge q varies, which means that a current is flowing through the voltage source, that is, the voltage source is doing work equal to $\varphi\,dq$ (as C increases). Hence, when applying the law of conservation of energy for constant φ and variable capacitance, one has to take into consideration not only the variation in the energy of the capacitor and the work of the force but also the work done by the voltage source.

13.5 Inductance Circuit

Let us consider a circuit consisting of *resistance* R and *inductance* L (Figure 13.5.1). By formula (13.1.11),

$$\varphi_R + \varphi_L = 0. \tag{13.5.1}$$

Since $\varphi_R = Rj$ and $\varphi_L = L\,(dj/dt)$, using (13.5.1) we find that

$$Rj + L\frac{dj}{dt} = 0.$$

Thus, the current in the circuit of Figure 13.5.1 satisfies the equation

$$\frac{dj}{dt} = -\frac{R}{L}\,j. \tag{13.5.2}$$

The solution to this equation is

$$j(t) = j_0 e^{-(R/L)t}. \tag{13.5.3}$$

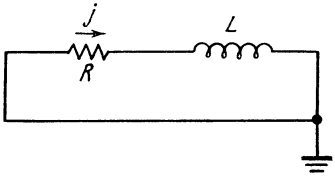


Figure 13.5.1

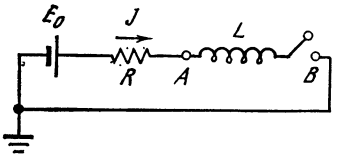


Figure 13.5.2

We see that the current in the circuit of Figure 13.5.1 falls off exponentially. The current diminishes e-fold during a time interval $T = L/R$.

Let us verify the dimensions of L/R. L is measured in henrys, that is, V·s/A and R in ohms, or V/A. Therefore,

$$[L/R] = (\text{V}\cdot\text{s/A})/(\text{V/A}) = \text{s},$$

so that L/R indeed has the dimensions of *time*. We will call L/R the ***decay time***. In the circuit diagram shown in Figure 13.5.1, in which there is no voltage source, the current tends to zero with the passage of time. How to set up an initial current of j_0 will be discussed later on.

For the present, let us consider a circuit consisting of a source having an emf E_0, of a resistance R, and of an inductance L (Figure 13.5.2). From the condition

$$\varphi_E + \varphi_R + \varphi_L = 0,$$

recalling that $\varphi_E = -E_0$, we find that

$$-E_0 + Rj + L\frac{dj}{dt} = 0. \tag{13.5.4}$$

We rewrite this equation as

$$\frac{dj}{dt} = \frac{R}{L}\left(\frac{E_0}{R} - j\right).$$

This equation is similar to Eq. (13.2.5) and is solved by the very same procedure. We get

$$j(T) = \frac{E_0}{R} + Ae^{-(R/L)t}, \tag{13.5.5}$$

where A is determined by the initial condition. Suppose the switch is closed at $t = 0$. Then $j(0) = 0$ because there was no current flowing in the circuit when the switch was open. Given this

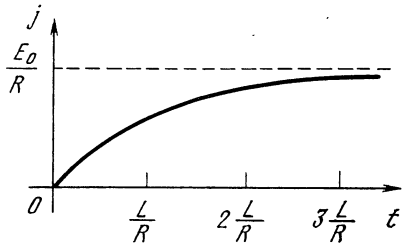


Figure 13.5.3

condition, we find that $A = -E_0/R$, and (13.5.5) assumes the form

$$j = \frac{E_0}{R}(1 - e^{-(R/L)t}). \qquad (13.5.6)$$

In the course of time the current approaches the value

$$j(\infty) = \frac{E_0}{R}. \qquad (13.5.7)$$

This value is independent of the inductance L and is simply obtained from Ohm's law in a circuit with an emf E_0 and a resistance R. However, this current sets in not at once but gradually (Figure 13.5.3), and the time required for $j(\infty)$ to set in depends on L: in time L/R the current becomes $0.63j(\infty)$, in time 2 (L/R) the current becomes $0.86j(\infty)$, in time 3 (L/R) the current becomes $0.95j(\infty)$, and so forth. Thus, the ratio L/R can also be called the ***buildup time for the current.***

According to the basic equation (13.5.4), the sum of the difference of potential across the resistance, Rj, and across the inductance, $L\,(dj/dt)$, is equal to the emf E_0. It is interesting to follow each term separately. They are shown in Figure 13.5.4. At the initial time, $j = 0$, $Rj = 0$, and $E_0 = L\,(dj/dt)$. We

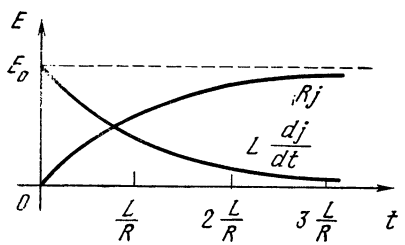


Figure 13.5.4

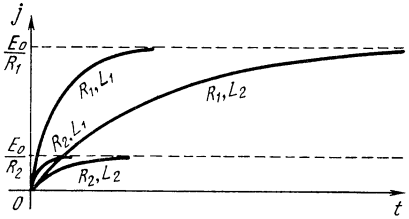


Figure 13.5.5

say that the voltage is absorbed entirely by the inductance. As time passes, the current approaches the constant value E_0/R, dj/dt tends to zero, and the voltage is absorbed by the resistance.

It is interesting to compare the solutions that formula (13.5.6) yields for the same E_0 but different values of R and L. Let R_1 be small, R_2 great, L_1 small, and L_2 great. For different combinations of R and L we get four curves of current as functions of time (see Figure 13.5.5). The final current $j(\infty)$ depends on R alone; it is the same for the pairs, R_1, L_1 and R_1, L_2; $j(\infty)$ is also the same for the pairs R_2, L_1 and R_2, L_2. The initial rate of buildup of current depends only on the inductance L and does not depend on the resistance: the curves specified by the pairs R_1, L_1 and R_2, L_1 and R_1, L_2 and R_2, L_2 all merge at the point O (0, 0).

On dimensional grounds, it is clear that the steady-state current is proportional to the initial rate of current buildup and the time of buildup. For our definition of buildup time, the formula is correct without any additional coefficients. Indeed, the initial rate of current buildup, $(dj/dt)_{t=0}$, is equal to E_0/L, and the buildup time, T, is equal to L/R, whence the established current is

$$j(\infty) = T\,(dj/dt)_{t=0} = \frac{L}{R}\,\frac{E_0}{L} = \frac{E_0}{R}.$$

Now comes the question we posed at the beginning of our discussion of setting up the initial current j_0 in the circuit in Figure 13.5.1. We can take the circuit diagrams of Figure 13.5.6. We start by closing switch A and leav-

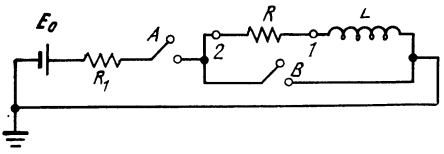


Figure 13.5.6

ing switch B open. Then a current will flow and soon attain the value $E_0/(R + R_1)$ in accordance with formula (13.5.7). We choose E_0 in such a manner that $E_0/(R + R_1) = j_0$. We wait for the steady state to set in when the current is equal to j_0 with A closed and B open. Then in this state we close B and open A. The result is the circuit shown in Figure 13.5.1. At the initial instant of time (when closing B), a current j_0 flows. The potential at point *1* (see Figure 13.5.5), prior to closing the switch, is $\varphi_1 = 0$, since in the steady state, when j_0 is constant, the voltage drop on L is zero. Prior to closing the switch, the potential at point *2* is equal to $\varphi_2 = Rj_0$. When switch B is closed, point *2* becomes grounded and so the potential at point *2* is $\varphi_2 = 0$. There is then a corresponding readjustment of potentials at all other points of the circuit. In particular, the potential at point *1* is now $\varphi_1 = -Rj$.

13.6 Breaking an Inductance Circuit

Above we considered the process of a steady-state current setting in in the circuit shown in Figure 13.5.2, which consisted of a voltage source, a resistance R, an inductance L, and a switch. Figure 13.5.3 shows the curve of current buildup when the switch is closed at time $t = 0$. In time, the current reaches the value $j_0 = E_0/R$. What will happen now if we suddenly open the switch B? If the current ceases to flow in a very short time τ, the derivative of the current with respect to time can be represented as follows:

$$\frac{dj}{dt} \simeq \frac{j(t+\tau)-j(t)}{\tau} = -\frac{j_0}{\tau},$$

which is to say, that the derivative will be very great in absolute value if τ is very small. And at point A there will appear a very large (in absolute value) negative potential:

$$\varphi_A = L\frac{dj}{dt} \simeq -L\frac{j_0}{\tau}.$$

The potential difference across the resistance R (it is equal to Rj) and the emf of the voltage source change but slightly when the switch is opened. For this reason, the great potential difference that appears across the inductance L when the switch is opened falls entirely on the switch; by this we mean that the potential difference across the open plates of the switch becomes very great, of the order of Lj_0/τ, and can exceed many times over the emf of the voltage source, E_0. If the potential difference is great, the air gap between the open contacts will break down and a spark will jump across.

The problem of current variation in a circuit when a switch is opened proves to be very complicated; this is due to the involved nature of the laws of electric discharge in air between plates. Indeed, prior to breakdown, when $\varphi < \varphi_b$, there was no current; but when breakdown occurs, the resistance of the spark falls drastically, a big current flows at a potential difference considerably less than φ_b. Here we only note the basic fact: large potential differences arise in inductance circuits in circuit breaking. When such a circuit is closed, the potential difference does not exceed E_0 (the emf of the source) anywhere.

The two circuits shown in Figure 13.6.1 give a quantitative idea of the phenomenon of sudden brief increases in potential difference. These circuits differ from that of Figure 13.5.2 in that current can also flow in inductance L when switch B is open, so that circuit breaking occurs without any spark. However, if the resistance R is much greater than the resistance r, a large potential difference appears across the inductance at break.

As an example, let us consider the circuit shown in Figure 13.6.1*a*. We

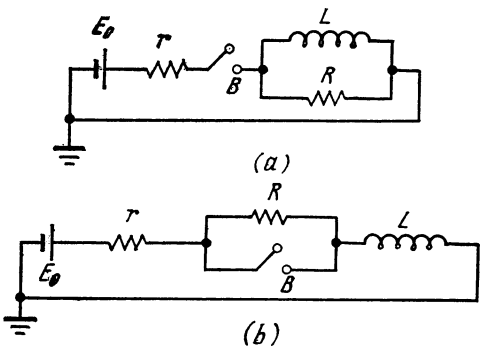


Figure 13.6.1

assume $R \gg r$. If the switch is closed, at an arbitrary time the current in the left portion of the circuit (r, E) is equal to the sum of the currents in the parallel connection RL:

$$j_r = j_E = j_R + j_L.$$

It is then always true that $\varphi_L = \varphi_R$.

Let the switch be closed at time $t = 0$. At this instant all the current will go through the resistance R, so that $j_{r0} = j_{R0} = E_0/(R + r)$, by Ohm's law. Then

$$\varphi_{r0} = E_0 \frac{r}{r+R}, \quad \varphi_{r0} = E_0 \frac{R}{R+r},$$

hence

$$\frac{dj_L}{dt}\Big|_{t=0} = \frac{\varphi_{R0}}{L} = \frac{E_0}{L} \frac{R}{R+r}.$$

After a sufficient time lapse after making the circuit, a constant current will flow. In the steady state, the entire current will go through the inductance. Indeed, if the current j does not vary with time, then $dj/dt = 0$, therefore $\varphi_L = 0$, and so $\varphi_R = 0$, whence $j_R = 0$.

In the steady state, $\varphi_{r\infty} = E_0$ and $j_{r\infty} = j_{L\infty} = E_0/r$, whence it is easy to obtain the order of the time τ_1 during which the steady-state current sets in:

$$j_{L\infty} - j_{L0} \simeq \frac{dj_L}{dt}\Big|_{t=0} \tau_1 \text{ or}$$

$$\frac{E_0}{r} \simeq \frac{E_0}{L} \frac{R}{R+r} \tau_1,$$

whence

$$\tau_1 \simeq \frac{L(R+r)}{rR} \simeq \frac{L}{R}.$$

Now let us examine breaking the circuit after a time lapse of $t \gg \tau_1$ after the circuit is closed, that is to say, after a constant current $j_\infty = E_0/r$ has been set up in the circuit. When the circuit is broken $j_r = j_E = 0$ and $j_R + j_L = 0$, whence $j_R = -j_L$. This means that the entire current passing through L must pass through R in the reverse direction. As before, of course, $\varphi_L = \varphi_R$. Therefore, $\varphi_R = Rj_R = -Rj_L$, or $\varphi_L = -Rj_L$. But since $\varphi_L = L\,(dj_L/dt)$, it follows that $L\,(dj_L/dt) = -Rj_L$.

We have thus arrived at Eq. (13.5.2), which is quite natural since the right-hand part of the diagram in Figure 13.6.1*a* (after the circuit is broken) does not differ from the circuit diagram in Figure 13.5.1.

The current diminishes by a factor of e during time $\tau_2 = L/R$. Here, $\tau_2 \ll \tau_1$, since $R \gg r$. At the time of break, the current has the value $j_{L\infty} = E_0/r$. After the break has taken place, but prior to the current falling off perceptibly, that is, for break time $t < \tau_2$, we get

$$\varphi_R = \varphi_L \simeq -Rj_{L\infty} = -E_0 \frac{R}{r}.$$

Thus, at break we can obtain a potential difference that is many times greater than the emf of the voltage source. This fact is extensively employed in engineering, in particular in the ignition systems of internal combustion engines. Observe that this large potential difference occurs over an extremely small time interval.

The foregoing is a rough consideration of the problem without the use of derivatives and other tools of higher mathematics. An exact consideration of the problem of closing a switch in the circuit of Figure 13.6.1*a* yields the following. Proceeding from the relations

$$\varphi_E + \varphi_r + \varphi_L = 0, \; j = j_R + j_L,$$

$$\varphi_R = \varphi_L,$$

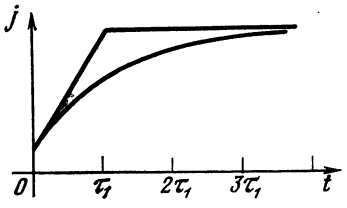


Figure 13.6.2

we get the differential equation

$$\frac{dj_L}{dt}+\frac{rR}{(R+r)L}j_L=\frac{E_0R}{(R+r)L}.$$

At the initial time, $t = 0$, the current flowing through the inductance is zero: $j_L = 0$ at $t = 0$. Therefore,

$$j_L=\frac{E_0}{r}\left\{1-\exp\left[-\frac{rR}{(R+r)L}t\right]\right\}$$
$$=\frac{E_0}{r}(1-e^{-t/\tau_1}). \qquad (13.6.1)$$

The current in the circuit is

$$j=j_L+j_R=\frac{E_0}{r}(1-e^{-t/\tau_1})$$
$$+\frac{E_0}{R+r}e^{-t/\tau_1}. \qquad (13.6.2)$$

In Figure 13.6.2 the approximate solution is shown by the broken line and the exact solution (13.6.2) by the smooth curve.

We advise the reader to examine the process of variation of current and potential difference at make and break in the circuit of Figure 13.6.1*b*. It is useful to solve the problem twice: once by setting up the differential equation and seeking its solution in the form of an exponential function, and the second time in the approximate fashion, as we did for the circuit depicted in Figure 13.6.1*a*.

13.7 The Energy of Inductance

We have seen that in a circuit consisting only of an inductance L and a resistance R, current continues to flow after the voltage source has been disconnected. The current gradually falls off in time. In the process, a quantity of heat $Q = Rj^2$ is released on the resistance in unit time.

What is the source of the electric energy that is converted into heat in the resistance? The energy is given up by the inductance, which has a certain supply of energy.

Let us find this supply of energy by considering the elementary circuit diagram shown in Figure 13.5.1 and let us calculate the entire thermal energy released on R. Suppose at the initial time, $t = 0$, this circuit has a current j_0. The current will then decay in time in accordance with the law

$$j(t)=j_0e^{-Rt/L}.$$

The quantity of energy released on the resistance R in unit time, that is, the ***rate of energy release*** (dimensions: W/s), is the instantaneous ***power output*** h. Using the time dependence of j given above, we find that

$$h=Rj^2=Rj_0^2e^{-2Rt/L}. \qquad (13.7.1)$$

Knowing h, it is easy to find the total amount of heat released from time $t = 0$ to infinity ($t = \infty$), that is, to complete decay of the current. To do this, it suffices to integrate (13.7.1) from $t = 0$ to $t = \infty$. This yields

$$Q=\int_0^\infty Rj_0^2e^{-2Rt/L}\,dt=Rj_0^2\int_0^\infty e^{-2Rt/L}dt$$
$$=\frac{Rj_0^2L}{2R}=\frac{Lj_0^2}{2}. \qquad (13.7.2)$$

This heat is equal to the supply of energy of the inductance through which the current j_0 flows. This supply of energy does not depend on the magnitude of the resistance R. An inductance L with current j_0 has a definite supply of energy which, ultimately, is transformed completely into heat, irrespective of the magnitude of the resistance R, which only affects the rate of transformation of energy into heat but not the total amount of energy.

Formula (13.7.2) can also be obtained by considering the process of current buildup in an inductance. Indeed, the ***power*** of the current (that is, work per

unit time) is equal to φj. This work is done by external sources of voltage and goes to increase the inductance energy W:

$$h = \frac{dW}{dt} = \varphi j. \qquad (13.7.3)$$

Using the fact that $\varphi = L\,(dj/dt)$, we find, from (13.7.3), that

$$\frac{dW}{dt} = Lj\frac{dj}{dt} = \frac{1}{2}L\frac{d(j^2)}{dt}. \qquad (13.7.4)$$

We will assume that $j = 0$ and $W = 0$ at $t = 0$, and $j = j_0$ and $W = W_0$ at $t = t_0$. Then, integrating (13.7.4) from $t = 0$ to $t = t_0$, we obtain

$$W_0 = \frac{1}{2}Lj_0^2.$$

To be specific, imagine the circuit in Figure 13.5.2 ($\varphi = \varphi_A$) and carry out the detailed computations for buildup of the energy of inductance. In the steady-state mode, when the current has reached a constant value j_0, the potential φ_A is zero and the energy of inductance does not vary, but the source of emf needed to maintain the constant current j_0 must continue to generate energy, which is released as heat on resistance R.

The energy W of inductance is proportional to the square of the current, which is to say, it is proportional to the square of the rate of motion of the electrons. Therefore, externally, W resembles kinetic energy. But is W the kinetic energy of the electrons?

Let us consider the orders of magnitude of W and the electron energy. Using a copper wire of length 100 m and diameter 0.35 mm (cross-sectional area equal to approximately 10^{-7} m^2), we can wind a coil having an inductance of 0.02 henry. A current of 1 A flowing in this coil will release $W = 0.02 \times 1^2 \times 0.5 = 10^{-2}$ J. We will now find the kinetic energy of the electrons.

We will assume that for each atom of copper there is one electron carrying current (the conduction electron). The atomic weight of copper is about 63, so 63 grams of copper contain 6×10^{23} conduction electrons, or roughly 10^{22} electrons per gram, or 10^{25} electrons per kilogram. Copper has a density of about 8 g/cm^3 = 8 $\times$ 10^3 kg/m^3, and so one cubic meter of copper wire contains roughly $n = 8 \times 10^{28}$ conduction electrons. Imagine a piece of copper wire of length $v\,dt$ and a cross-sectional area S to the left of cross section O (Figure 13.7.1). If the electron velocity is v, then through S there will flow $Snv\,dt$ electrons in time dt.[13.9] In time dt the electrons at cross section A will move to cross section O, which means that during this time all the electrons in the volume between O and A will pass through O. That is, they will pass through a cylinder of altitude $v\,dt$ and base S.

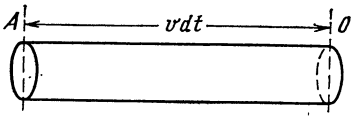


Figure 13.7.1

Denote by e the electron charge in coulombs, $e = -1.6 \times 10^{-19}$ C. The quantity of electricity which the $Snv\,dt$ electrons transfer in time dt is equal to the current in amperes multiplied by the time dt. Therefore, $Snve\,dt = j\,dt$, whence $j = Snve$, or $v = j/Sne$. Substituting $j = 1$ A, $S = 10^{-7}$ m^2, $n = 8 \times 10^{28}$ m^{-3}, $e = -1.6 \times 10^{-19}$ C, we obtain

$$v = \frac{1}{10^{-7} \times 8 \times 10^{28} \times 1.6 \times 10^{-19}} \simeq 8 \times 10^{-4}.$$

The dimensions are that of *velocity*:

$$[v] = \text{A/m}^2 \cdot \text{m}^{-3} \cdot \text{C} = \text{A} \cdot \text{m/A} \cdot \text{s} = \text{m/s}.$$

Now we have to calculate the *kinetic energy* of the (conduction) electrons. The electron mass m is 9×10^{-31} kg. The total number of electrons moving in the wire ($S = 10^{-7}$ m^2 and $l = 10^2$ m) is $10^2 \times 10^{-7} \times 8 \times 10^{28} \simeq$

[13.9] We have in view the mean velocity of their motion in the *direction of current flow* and not the velocity of random thermal motion.

10^{24}. The kinetic energy of one electron is

$$\frac{1}{2}mv^2 \simeq \frac{1}{2} \times 9 \times 10^{-31} \times 64 \times 10^{-8} \simeq 3 \times 10^{-37}\ \text{J},$$

while the total kinetic energy of the electrons moving in the wire is

$$T = n\frac{mv^2}{2} \simeq 10^{24} \times 3 \times 10^{-37} = 3 \times 10^{-13}\ \text{J}.$$

Thus, the kinetic energy of the electrons constitutes a minute fraction of the inductance, although it depends on the current via the same law (it is proportional to j^2) as the inductance energy. Physically, the inductance energy is the energy of the *magnetic field* which appears in the coil when current flows through it.

Let us point out some similarities and differences between *capacitance* and *inductance*. Both can serve as reservoirs of energy, both can be used to accumulate electric energy from a weak primary source and then release it quickly at the required place and time.

A capacitor can be charged with a small current j_1 during a long time t_1; then rapidly discharging it through a small resistance during a short time t_2, we can obtain a large current $j_2 \simeq j_1 t_1/t_2$, and the potential difference across the capacitor does not exceed the emf of the primary source. In other words, a capacitor enables us to increase the current but not the voltage.

We can send a large current through an inductance under a small voltage (small emf) E_0 of the primary source. The only requirement here is that the resistances of the inductance and the primary source be small. Then it takes a comparatively long time t_3 for a large current to build up in the inductance. When the inductance is shunted by a high resistance, we can obtain a large potential difference φ for a short time t_4, with $\varphi \simeq E_0 t_3/t_4$. An inductance enables us to increase the voltage but not the current.

The essential practical difference between a capacitance and an inductance is that a capacitor disconnected from a voltage source can retain its supply of energy for a very long time—hours or even days. The discharge time of a capacitor is equal to RC, where C is the capacitance and R the ***leakage resistance***. Using good insulators, we can obtain enormous values of R, that is, long discharge times. An inductance in the form of a coil and short-circuited (minimal resistance) retains its energy (if current is flowing) for only a fraction of a second.

The decay time of the current in an inductance is of the order of L/R, but even with the best conductors (copper, silver) it is impossible to make L/R greater than a few seconds for an ordinary laboratory-type coil. It will be noted that if we increase the number of turns in the coil for a given volume by using thinner wire, L will increase, but so will R, their ratio, however, to within order of magnitude, does not change. Therefore, under laboratory conditions, inductance is conveniently used for increasing voltage but not for long-term storage of energy.

Circuits involving capacitances and inductances can be used to accumulate energy from, say, a flashlight battery which with an internal resistance of several ohms yields a few volts so that the maximum power output is of the order of one to two watts. Using circuits of this kind, we are able to obtain powers up to hundreds of kilowatts. But a power output of this kind lasts for a time interval of the order of 10^{-6} s.

It has been noted that the electric energy in an inductance is quickly converted into heat due to resistance. This assertion holds true for coils of the ordinary laboratory kind and at ordinary (normal) temperatures. In two extreme cases, however, this does not hold true.

(1) At *very low temperatures* of the order of —260 °C down to absolute zero (—273 °C), many metals (for instance, lead, mercury, but not copper) pass in-

to what is known as the ***superconducting state***. Their specific resistance (resistivity) becomes exactly equal to zero.

The Dutch scientist Heike Kamerlingh Onnes (1853-1926), who discovered this phenomenon in 1911, observed a constant current in a ring circuit of superconducting material that lasted many days without any decrease in intensity. The presence of current in such a ring circuit is detected via the magnetic field of the current.

The practical application of superconductors is limited not only by the difficulty of generating low temperatures. A strong magnetic field converts a superconductor to the normal state (with *finite* resistance). This is why large currents cannot be transmitted through a superconductor.[13.10]

(2) The relationship between inductance and resistance and the conditions of current decay vary drastically *when all the dimensions of the coil are increased*, particularly when passing to astronomical phenomena (on the astronomical scale).[13.11]

Picture two geometrically similar coils, one of which is n times the other in size, the number of turns in both being the same. In the large coil, the diameter of the coil is n times greater, but so is the height of the coil and the diameter of the wire. Suppose the coils are made of the same material. Quantities relating to the small coil will be labeled with the subscript 1, those referring to the large coil will have the subscript 2. Let us calculate the relation between the resistances of the coils:

$$R_1 = \rho l_1/S_1, \quad R_2 = \rho l_2/S_2,$$

where ρ is the resistivity (specific resistance) of the coil material, l is the length of the wire, and S the cross-sectional area of the wire. Geometrically, it is clear that $l_2 = nl_1$ and $S_2 = n^2S_1$ and, hence, that $R_2 = R_1/n$. The resistance is inversely proportional to n, that is, to the dimensions.

It can be proved that the inductance of the large coil is exactly n times the inductance of the small coil, $L_2 = nL_1$, which means that increasing the linear dimensions of the coil n times increases the inductance n times, too. The decay time of the current, τ, is of the order of L/R; consequently, $\tau_1 = L_1/R_1$ and $\tau_2 = L_2/R_2 = n^2L_1/R_1 = n^2\tau_1$.

Thus, the decay time of the current is proportional to the *square* of the dimensions. If the earth were to consist of copper, the decay time of a current flowing in it would be of the order of 10^{15} to 10^{16} s, or approximately 10^8 years.

The conductivity of ionized gases is of the same order of magnitude as that of copper. For this reason, the decay time of a current in astronomical phenomena is enormous. This means that resistance and Ohm's law do not play any role whatsoever in these phenomena. Recall Figure 13.5.5: the current on the initial portion of any curve depends on L alone but not on R, and in astronomy we are always on the "initial portion."

Terrestrial magnetism is due to the magnetic field of the current flowing in the viscous molten mass of the central core of the earth. The slow motions of this molten mass in the magnetic field sustain the currents, just as the motion of the armature in a dynamo in a magnetic field sustains the current in the armature winding and in the winding of the electromagnet. The same is true of the magnetic field of the sun. The theory successfully predicts

---

13.10 In 1961 an alloy was discovered of the rare element niobium and tin in which a current up to 100 000 A/cm^2 and a magnetic field up to 25 T (that is, 250 000 gauss) were not able to destroy superconductivity. At present superconducting magnets are widely used in science and technology, for instance, in particle accelerators. There is reason to believe that in the near future superconductivity will be employed in electric transmission lines.

13.11 Compare with Exercise 13.2.3; for a circuit consisting of a capacitance and a resistance the discharge time does not change when all dimensions are altered.

the periodic changes in the direction of the sun's field (the half-period is 11 years).

13.8 The Oscillatory Circuit

Let us consider a circuit consisting of a capacitance C and an inductance L (Figure 13.8.1). Let point B of the circuit be grounded. By formula (13.1.11), $\varphi_C + \varphi_L = 0$, where $\varphi_L = L\ (dj/dt)$ and $\varphi_C = q/C$. The voltage drop on the capacitance, φ_C, will be denoted simply φ. Then

$$\varphi + L\frac{dj}{dt} = 0. \tag{13.8.1}$$

Note that $j = dq/dt$ and so $dj/dt = d^2q/dt^2$. But since $d^2q/dt^2 = C\ (d^2\varphi/dt^2)$, it follows that, using formula (13.8.1), we find that

$$\varphi + LC\frac{d^2\varphi}{dt^2} = 0, \text{ or}$$
$$\frac{d^2\varphi}{dt^2} = -\frac{1}{LC}\varphi. \tag{13.8.2}$$

We considered a similar equation in Chapter 10 in the study of mechanical vibrations. It was established that the functions $\varphi = A \sin \omega t$ and $\varphi = B \cos \omega t$ are solutions to Eq. (13.8.2) for arbitrary A and B and a suitably chosen ω. Let us verify this, say, for $\varphi = A \sin \omega t$ and in passing we will define ω. Substituting φ and $d^2\varphi/dt^2$ into (13.8.2), we get

$$-ALC\omega^2 \sin \omega t = -A \sin \omega t,$$

or, canceling out $-A \sin \omega t$, we get $LC\omega^2 = 1$, whence

$$\omega = \frac{1}{\sqrt{LC}}. \tag{13.8.3}$$

Consequently, for a solution to Eq. (13.8.2) we have functions that describe oscillations with a circular frequency $1/\sqrt{LC}$. The oscillation period is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{LC}. \tag{13.8.4}$$

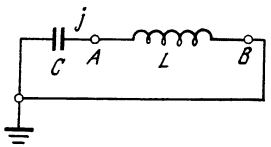


Figure 13.8.1

Let us check the dimensions in (13.8.4):

$$[C] = \text{F} = \text{C/V} = \text{A}\cdot\text{s/V},$$
$$[L] = \text{H} = \text{V/(A/s)} = \text{V}\cdot\text{s/A},$$

so that $[LC] = \text{s}^2$ and $T \propto \sqrt{LC}$ does indeed have the dimensions of time, s.

Let us examine in detail the solution to Eq. (13.8.2). The solutions $\varphi = A \sin \omega t$ and $\varphi = B \cos \omega t$ are actually indistinguishable since the sine curve is obtained from the cosine curve by a shift along the t axis; the two can be combined into a single expression:

$$\varphi = B \cos (\omega t + \alpha) \tag{13.8.5}$$

(cf. Section 10.2). The amplitude B of the oscillations and the initial phase α (in Section 10.2 we denoted this quantity by φ) may be arbitrary. For a given $\varphi\ (t)$ we find the dependence of current on time:

$$j = C\frac{d\varphi}{dt} = -CB\omega \sin (\omega t + \alpha). \tag{13.8.6}$$

Let us find the energy of capacitance and the energy of inductance:

$$W_C = \frac{C\varphi^2}{2} = \frac{CB^2}{2}\cos^2 (\omega t + \alpha),$$
$$W_L = \frac{Lj^2}{2} = \frac{LC^2B^2\omega^2}{2}\sin^2 (\omega t + \alpha).$$

Substituting the expression (13.8.3) for ω yields

$$W_L = \frac{CB^2}{2}\sin^2 (\omega t + \alpha).$$

The total energy is independent of time, as was to be expected. Indeed,

$$P \equiv W_C + W_L = \frac{CB^2}{2}[\cos^2 (\omega t + \alpha)$$
$$+ \sin^2 (\omega t + \alpha)] = \frac{CB^2}{2}.$$

To summarize, the motion of charges in an LC circuit is similar to the motion of a mass attached to a spring.

The energy of a charged capacitor may be likened to the elastic energy of a spring, which is at a maximum when the mass is in the extreme position of maximum separation from the equilibrium position. The energy of an inductance may be likened to the kinetic energy of a moving mass. When the charge on a capacitance is equal to zero, the current reaches its maximum (in absolute value), at this instant the capacitance energy is zero and the inductance energy is equal to the total energy ($\cos^2(\omega t + \alpha) = 0$ and $\sin^2(\omega t + \alpha) = 1$). This is exactly what happens in the oscillations of a mass on a spring: when the mass passes through the position of equilibrium, the potential energy is zero and the kinetic energy is equal to the total energy of the oscillations.

Let us use the term ***general problem*** for the problem of finding the potential in a circuit provided that at the initial time $t = 0$ we have $j = j_0$ and $\varphi = \varphi_0$. Neither the particular solution $\varphi = A \sin \omega t$ nor the particular solution $\varphi = B \cos \omega t$ enables us to solve the general problem. To solve the general problem we will need the general solution (13.8.5), with the two (indeterminate) parameters B and α at our disposal.

Setting $t = 0$ in (13.8.5) and (13.8.6) yields

$$\varphi(0) = B \cos \alpha = \varphi_0,$$

$$j(0) = -CB\omega \sin \alpha = j_0,$$

whence we find the solution with given φ_0 and j_0:

$$B = \sqrt{\varphi_0^2 + j_0^2/(C\omega)^2},$$

$$\tan \alpha = -j_0/C\omega\varphi_0. \qquad (13.8.7)$$

We already know that for such oscillations the energy is conserved and, hence, is equal to the initial energy $C\varphi_0^2/2 + Lj_0^2/2$. The expression for the energy may also be written as $CB^2/2$, which enables writing the first formula in (13.8.7) as

$$B = \varphi_{\max} = \sqrt{\varphi_0^2 + (L/C) j_0^2}, \qquad (13.8.8)$$

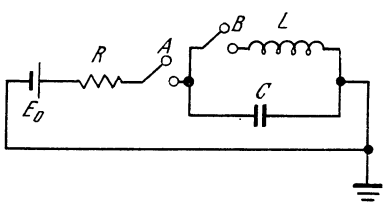


Figure 13.8.2

whence (as well as from (13.8.6)) we also obtain

$$j_{\max} = \sqrt{j_0^2 + (C/L)\varphi_0^2} \qquad (13.8.9)$$

(note that $\omega = \sqrt{1/LC}$, in view of (13.8.4)).

The circuit diagram for generating oscillations is shown in Figure 13.8.2. Here we have a source of voltage with emf E_0. If we close A and leave B open, after a lapse of time $\tau \gg RC$ after closing the circuit the capacitance will be charged to potential E_0. Open A and at time $t = 0$ close B. Then oscillations in the LC circuit will set in with $\varphi = \varphi_0 = E_0$ and $j = 0$ at $t = 0$. Note that with these oscillations, the potential difference across the electrodes of the open switch A will vary periodically from 0 to $2E_0$.

There is another way of setting up oscillations in the circuit of Figure 13.8.2. First close both switches A and B. The current flowing in the circuit will be $j_0 = E_0/R$. At $t = 0$ open A. Then oscillations in the LC circuit will set in with $\varphi_0 = 0$ and $j_0 = E_0/R$ at $t = 0$. For these oscillations, the maximum amplitude of the potential will reach

$$\varphi_{\max} = j_0 \sqrt{L/C} = E_0 \frac{1}{R} \sqrt{L/C}.$$

It will be recalled that in a circuit without capacitance, when we break the circuit containing inductance L, the potential difference developed on the switch is the larger the greater the resistance of the air gap between the electrodes of the switch. When such a circuit (with no capacitance) is broken, there is always a discharge in the air gap between the contacts of the

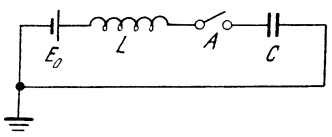


Figure 13.8.3

switch. In the case of a capacitance, the maximum potential difference between the electrodes of switch A does not exceed a definite value, $E_0 + \varphi_{max}$. If this value is less than what is required to initiate the discharge in the air gap of the switch, there will be no discharge. We say that the capacitance C extinguishes the discharge when an inductance circuit is broken. Note that the quantity $R^{-1}\sqrt{L/C}$ may be greater than unity. Then by opening switch B a quarter-period after opening A, we obtain a potential on capacitance C that is higher than the potential on the voltage source, E_0.

Exercises

13.8.1. Consider the variation of potential with time in the circuit shown in Figure 13.8.3. Determine the greatest value of φ and the time required to attain it. Assume that switch A is closed at time $t = 0$.

13.8.2. In the preceding problem, find the energy of capacitance and the energy released by the voltage source when φ is at a maximum.

13.9 Damped Oscillations

Let us consider a circuit with a resistance R in series with an inductance L (Figure 13.9.1). We assume R to be small. If R is not taken into account at all, we get the circuit diagram of Figure 13.8.1. If $\varphi = \varphi_0$ and $j = 0$ at $t = 0$, then, by (13.8.7), we have

$$\varphi = \varphi_0 \cos \omega t, \quad j = j_m \sin(\omega t + \pi) = -j_m \sin \omega t, \tag{13.9.1}$$

with

$$j_m = j_{max} = C\varphi_0\omega, \quad \omega = \frac{1}{\sqrt{LC}}. \tag{13.9.2}$$

The total energy P is then $C\varphi_0^2/2$ or, using (13.9.2),

$$P = \frac{Lj_m^2}{2}, \tag{13.9.3}$$

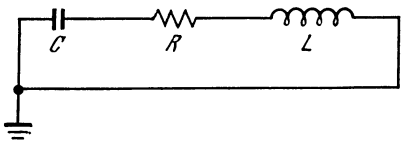


Figure 13.9.1

When a resistance is present in the circuit, electric energy is converted into thermal energy. The ***power output*** of the resistance, h, is given by the following formula:

$$h = Rj^2 = Rj_m^2 \sin^2 \omega t = \frac{Rj_m^2}{2}(1 - \cos 2\omega t). \tag{13.9.4}$$

The power output in the case of electric oscillations does not remain constant: over each period (cycle), h twice reaches a maximum and twice becomes zero (the sign of course does not change, that is, the power output is always positive). Let us find the *average value* of h over one period. From (13.9.4) we find that

$$\bar{h} = \frac{Rj_m^2}{2}(1 - \overline{\cos 2\omega t}).$$

Recalling that the average value of the cosine over one period is zero, we obtain

$$\bar{h} = \frac{Rj_m^2}{2}.$$

Heat release on resistance R can occur only as a result of reduction in the electric energy P. Therefore

$$\frac{dP}{dt} = -h. \tag{13.9.5}$$

We assumed that R was small, and so h is small. The oscillation energy falls off slowly and an appreciable change in energy becomes noticeable only after several cycles. Considering time intervals that are large compared with the oscillation period T, we replace h by $\bar{h}$ in the right-hand side of (13.9.5):

$$\frac{dP}{dt} \simeq -\bar{h} = -\frac{Rj_m^2}{2}. \tag{13.9.6}$$

Since the energy P varies slowly, from (13.9.3) we see that j_m too is a slowly

varying quantity. Expressing j_m from (13.9.3), we get

$$j_m = \sqrt{2P/L}. \qquad (13.9.7)$$

Using this, from (13.9.6) we get $dP/dt = -(R/L)P$ (here $\simeq$ was replaced with $=$). The solution to this equation is

$$P = P_0 e^{-(R/L)t},$$

where P_0 is the value of P at $t = 0$. Therefore, according to (13.9.7), $j_m = \sqrt{2P_0/L}\, e^{-Rt/2L}$. Then

$$j = \sqrt{2P_0/L}\, e^{-Rt/2L} \sin(\omega t + \pi). \qquad (13.9.8)$$

Recalling that $\varphi = \varphi_0 \cos \omega t$ and $\varphi_0 = j_m/C\omega$, we get

$$\varphi = \frac{j_m}{C\omega} \cos \omega t$$

$$= \frac{1}{C\omega} \sqrt{\frac{2P_0}{L}}\, e^{-Rt/2L} \cos \omega t. \qquad (13.9.9)$$

Formulas (13.9.8) and (13.9.9) show that if a small resistance is present, the electric oscillations *damp out* via an exponential law.

The solution given above was obtained by means of an approximate calculation. Note that in this approximate solution the relation $j = C(d\varphi/dt)$ is not satisfied, although it holds the more exactly the smaller R is. Let us now try to solve the problem exactly. For the circuit shown in Figure 13.9.1 we have $\varphi + \varphi_R + \varphi_L = 0$, whence

$$\varphi + Rj + L\frac{dj}{dt} = 0, \qquad (13.9.10)$$

and $j = C(d\varphi/dt)$. Substituting the expression for j and dj/dt into (13.9.10), we find that

$$LC\frac{d^2\varphi}{dt^2} = -\varphi - RC\frac{d\varphi}{dt}. \qquad (13.9.11)$$

We will seek the solution to this equation in the form obtained in the approximate consideration, or

$$\varphi = Ae^{-\lambda t} \cos \omega t, \qquad (13.9.12)$$

where λ, ω, and A are constants that we have to determine. The procedure that follows agrees completely with the one encountered in the theory of oscillations (see Section 10.4, and compare the solution (13.9.12) of Eq. (13.9.11) with formula (10.4.7)), so that here we could simply refer to the results obtained in Section 10.4. However, we will not complicate matters by referring the reader to Chapter 10 to compare the different notations, and will briefly repeat the required derivation.

We put the expression (13.9.12) for φ and the expressions for the first and second derivatives of φ that follow from (13.9.12) into Eq. (13.9.11) and cancel the common factor $Ae^{-\lambda t}$ out of all terms to get

$$LC\lambda^2 \cos \omega t + 2LC\lambda\omega \sin \omega t$$
$$- LC\omega^2 \cos \omega t = -\cos \omega t$$
$$+ RC\lambda \cos \omega t + RC\omega \sin \omega t.$$

For this equation to be valid for *arbitrary* t, it is necessary that the coefficients of $\cos \omega t$ and $\sin \omega t$ be equal separately on the right and on the left:

$$LC\lambda^2 - LC\omega^2 = RC\lambda - 1, \qquad (13.9.13)$$

$$2LC\lambda\omega = RC\omega. \qquad (13.9.14)$$

The condition (13.9.14) yields the value of λ, and by substituting this value into (13.9.13) we get ω:

$$\lambda = \frac{R}{2L}, \quad \omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}. \qquad (13.9.15)$$

The constant A in (13.9.12) has yet to be determined. To do this we must specify the initial condition (say, $\varphi = \varphi_0$ at $t = 0$). Finally, knowing $\varphi(t)$, we can easily find $j = C(d\varphi/dt)$. We then have

$$j = -CAe^{-\lambda t}(\omega \sin \omega t + \lambda \cos \omega t). \qquad (13.9.16)$$

Comparing the exact solution with the approximate, we note the following: (1) in the approximate consideration of the problem we correctly determined the number λ, which describes the rate of decay of the oscillations (however, the approximate solution does not yield the dependence of frequency ω on resistance R), and (2) the formula for current is somewhat differ-

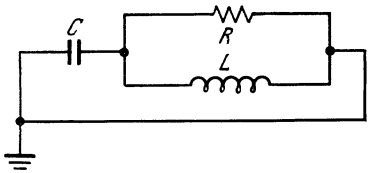


Figure 13.9.2

ent from the one that was obtained in approximate fashion.

In exactly the same manner we can show that Eq. (13.9.11) has yet another solution,

$$\varphi = Be^{-\lambda t} \sin \omega t, \qquad (13.9.17)$$

where ω and λ are the same as in (13.9.15). The corresponding current is

$$j = CBe^{-\lambda t} (\omega \cos \omega t - \lambda \sin \omega t). \qquad (13.9.18)$$

The sum $\varphi = e^{-\lambda t} (A \cos \omega t + B \sin \omega t)$ of the solutions (13.9.12) and (13.9.17) is also a solution to Eq. (13.9.11). It is only with the aid of this sum that we can solve the *general* problem: to find the solution to Eq. (13.9.11) with the initial conditions $\varphi = \varphi_0$ and $j = j_0$ at $t = 0$. Indeed, for the coefficients A and B we then have the equations $\varphi_0 = A$ and $j_0 = -CA\lambda + CB\omega$, whence $A = \varphi_0$ and $B = (C\lambda\varphi_0 + j_0)/C\omega$.

Exercises

13.9.1. Find $j(t)$ for the circuit of Figure 13.9.1 if $C = 1$, $L = 1$, and $R = 0.1$, 0.5, 1. Assume that $\varphi = 1$, and $j = 0$ at $t = 0$.

13.9.2. The same question if $\varphi = 0$ and $j = 0$ at $t = 0$.

13.9.3. Using the approximate method, find the rate of decay λ of oscillations in the circuit shown in Figure 13.9.2 on the assumption that R is very great.

13.10* The Case of a Large Resistance

The case considered here of a large resistance is mainly of mathematical interest and is not connected with the sequel. It may therefore be skipped in a first reading.

The solution of Eq. (13.9.11) obtained in the preceding section is valid only for R that are not too large. Indeed, from (13.9.15) it is evident that if R is greater than $2\sqrt{L/C}$, the formula we have for ω is meaningless since the radicand is negative. In that case Eq. (13.9.11) has a different kind of solution. We will seek the solution in the form $\varphi = Ae^{-\beta t}$ (and accordingly, $j = -AC\beta e^{-\beta t}$). Substituting into (13.9.11) the expressions for φ and its derivatives and canceling $Ae^{-\beta t}$ out of all terms, we get $LC\beta^2 = -1 + RC\beta$. This is a quadratic equation in β. Solving it, we find that

$$\beta = \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}. \qquad (13.10.1)$$

The radicand in (13.10.1) differs in sign from the radicand in (13.9.15). Hence, in those cases where it is impossible to find ω we can find β, and vice versa. Formula (13.10.1) yields two distinct values of β, and so we can set up two solutions to Eq. (13.9.11), $\varphi = Ae^{-\beta_1 t}$ and $\varphi = Be^{-\beta_2 t}$. Their sum is also a solution:

$$\varphi = Ae^{-\beta_1 t} Be^{-\beta_2 t}, \qquad (13.10.2)$$

Accordingly

$$j = -AC\beta_1 e^{-\beta_1 t} - BC\beta_2 e^{-\beta_2 t}. \qquad (13.10.3)$$

If $\varphi = \varphi_0$ and $j = j_0$ at $t = 0$, then, assuming that $t = 0$ in (13.10.2) and (13.10.3), we get $A + B = \varphi_0$ and $-AC\beta_1 - BC\beta_2 = j_0$. We can find A and B from this system of equations.

Let us consider in more detail the expression for β. Let $R \gg 2\sqrt{L/C}$. Then

$$\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} = \frac{R}{2L} \sqrt{1 - \frac{4L}{R^2C}}$$

can be expanded by the binomial theorem (see Section 6.4). We confine ourselves to two terms:

$$\frac{R}{2L} \sqrt{1 - \frac{4L}{R^2C}} \simeq \frac{R}{2L} \left(1 - \frac{1}{2} \frac{4L}{R^2C}\right) = \frac{R}{2L} - \frac{1}{RC}.$$

Therefore

$$\beta_1 \simeq \frac{R}{2L} + \frac{R}{2L} - \frac{1}{RC} = \frac{R}{L} - \frac{1}{RC} \simeq \frac{R}{L}$$

since R is great, similarly,

$$\beta_2 \simeq \frac{R}{2L} - \frac{R}{2L} + \frac{1}{RC} = \frac{1}{RC}.$$

These values of β_1 and β_2 are familiar from Sections 13.1 to 13.5. Indeed, β_1 corresponds to current decay by the law $e^{-(R/L)t}$, which means that this is an RL circuit (see Section 13.5). The second root β_2 corresponds to current decay by the law $e^{-t/RC}$, which means that this is an RC circuit (see Section 13.2).

Of mathematical interest is the particular case where the radicand in (13.10.1) is exactly zero:

$$\frac{R^2}{4L^2} = \frac{1}{LC},$$

so that both roots, β_1 and β_2, coincide. We obtain only one solution to Eq. (13.9.11). But in order to solve the problem with initial condition $\varphi = \varphi_0$ and $j = j_0$ at $t = 0$ we need *two* solutions.

How can we find the second solution? Suppose that $\beta_1 \neq \beta_2$ but the difference $\beta_1 - \beta_2$ is small. Then we have two solutions: $e^{-\beta_1 t}$ and $e^{-\beta_2 t}$. Their difference is also a solution. We write this solution as

$$e^{-\beta_1 t} - e^{-\beta_2 t} = e^{-\beta_2 t}\,[e^{(\beta_2 - \beta_1)t} - 1].$$

Since $\beta_2 - \beta_1$ is small, it follows that $e^{(\beta_2 - \beta_1)t} \simeq 1 + (\beta_2 - \beta_1)t$ (in the Taylor series only two terms are retained), whence

$$e^{-\beta_1 t} - e^{-\beta_2 t} \simeq e^{-\beta_2 t} t\,(\beta_2 - \beta_1).$$

The last expression suggests that if $\beta_2 = \beta_1 = \beta$, the second solution must be taken in the form $\varphi = Bte^{-\beta t}$. Substituting this φ into Eq. (13.9.11) and noting that $\beta = R/2L$, we see that the equation is indeed satisfied. Thus, when $\beta_1 = \beta_2 = \beta$, we must take φ in the form

$$\varphi = Ae^{-\beta t} + Bte^{-\beta t}.$$

This φ (and the corresponding j) permits solving the problem with arbitrary initial φ_0 and j_0.

Exercises

13.10.1. Find $\varphi(t)$ for $L = 1$, $C = 1$, and $R = 2$, 6, 10. Assume that $\varphi_0 = 1$ and $j_0 = 0$ at $t = 0$.

13.10.2. Find $\varphi(t)$ for $L = 1$, $C = 1$, and $R = 2$, 4, provided that $\varphi_0 = 1$ and $j_0 = 1$ at $t = 0$.

13.11 Alternating Current

We will now examine circuits in which the voltage source has an emf that *varies periodically* with time with a definite given frequency ω. These problems are very important in radio-circuit work. The frequency of alternating current exerts quite a different effect on the passage of current through an inductance and a capacitance. The higher the frequency, the faster the current varies and the "harder" it is for the current to pass through an inductance and the greater the potential difference that a current of a given intensity can set up. Contrariwise, the potential difference of the plates of a capacitor is the smaller the greater the frequency. When the frequency is increased, the period diminishes and, consequently, the time interval decreases during which the current flowing in one direction can charge up the capacitor. Therefore, as the frequency increases, the charge on the capacitor decreases and so does the potential difference on the plates of the capacitor.

We have already pointed out (see Section 13.8) that the movement of charges in an LC circuit may be likened to the oscillations of a body suspended from a spring: whereas in the case of vibrating body the distance of the body from the origin and also its velocity vary periodically with time, in a circuit the potential and current vary periodically.

The frequency with which a body vibrates under the action of the elastic force of the spring (in the absence of any other forces) is called the natural frequency. Similarly, the frequency of oscillation of potential in an LC circuit is termed ***the natural frequency of the circuit.***

Developing this analogy further, we can assume that if the circuit is connected to an alternating current network (which means the potential impressed

on the circuit will vary periodically), we will have what is known as ***resonance***. What resonance means is that the amplitude of oscillation is at a maximum when the frequency ω of the current is equal to the natural frequency ω_0 of the circuit. The amplitude increases sharply when the difference $\omega - \omega_0$ approaches zero. Resonance actually does take place and we will consider it in Section 13.13.

For every *two-terminal network* (see Figure 13.1.7) connected to an alternating-current circuit, there is a definite relationship between the potential difference and the current. Let us find this relationship first for the simplest case of separated elements R, L and C and then, in Sections 13.13 and 13.14 for more complicated circuit.

We will consider alternating current of a definite frequency ω; as before, ω is connected with the period through the relation $\omega = 2\pi/T$. For example, in the USSR the standard current is 50 cycles per second (or 50 Hz), that is, $T = 1/50$ s and $\omega = 2\pi \times 50 \simeq 314\ \text{s}^{-1}$.

We refer to the circuit diagram in Figure 13.11.1, which contains an ammeter A that indicates current j and a voltmeter V measuring the voltage (difference of potentials). Suppose that the ammeter and voltmeter are so inertialess (high-speed) that they permit measuring the *instantaneous* value of current at each instant and, hence, their readings vary with a period equal to that of the current. This experiment is usually accomplished with the aid of an

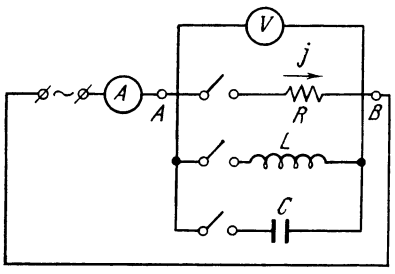


Figure 13.11.1

oscillograph (a so-called ***loop oscillograph*** with two loops or a ***cathode-ray oscillograph*** with two beams). The positive direction of current is shown by an arrow. The voltmeter V measures $\varphi = \varphi_A - \varphi_B$. By closing one or another of the switches we can investigate the current flowing through the resistance, inductance, or capacitance.

Suppose the current varies with time in accordance with the law

$$j = j_0 \cos(\omega t + \alpha). \tag{13.11.1}$$

If this current flows through resistance R, by Ohm's law we have

$$\varphi_R = Rj = Rj_0 \cos(\omega t + \alpha). \tag{13.11.2}$$

For the sake of generality, let us write this as

$$\varphi_R(t) = \varphi_1 \cos(\omega t + \alpha_1),$$

where $\varphi_1 = Rj_0$ and $\alpha_1 = \alpha$.

Let the current (13.11.1) flow through inductance L. Then

$$\varphi_L = L\frac{dj}{dt} = -L\omega j_0 \sin(\omega t + \alpha).$$

Set

$$\varphi_L = \varphi_2 \cos(\omega t + \alpha_2). \tag{13.11.3}$$

Then $\varphi_2 = L\omega j_0$ and $\alpha_2 = \alpha + \pi/2$. Indeed, we know that $\cos(\beta + \pi/2) = -\sin\beta$ for arbitrary β and so

$$\cos(\omega t + \alpha + \pi/2) = -\sin(\omega t + \alpha).$$

Thus, in the case of alternating current the relationship between the amplitude of current, j_0, and the amplitude of voltage, φ_2, in the inductance is the same as in a resistance equal to $R_2 = L\omega$. If L is expressed in henrys (H) and ω in reciprocal seconds (s^{-1}), then R_2 will be expressed in ohms (Ω).

Inductance differs from resistance in that the curve of the voltage is displaced a quarter-period from the current curve (Figure 13.11.2). This is quite evident from the formula

$$-\sin(\omega t + \alpha) = \cos(\omega t + \alpha + \pi/2)$$
$$= \cos[\omega(t + \pi/2\omega) + \alpha].$$

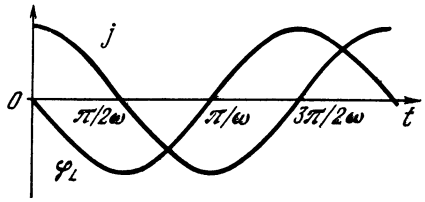


Figure 13.11.2

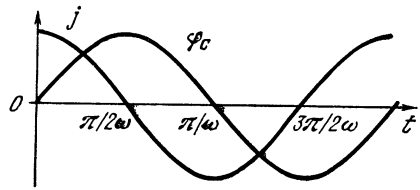


Figure 13.11.3

Let the function $\cos(\omega t + \alpha)$, to which the current is proportional, reach a definite value at time t_1:

$$\cos(\omega t_1 + \alpha) = a.$$

The function $\cos(\omega t + \alpha + \pi/2)$, to which the voltage on the inductance is proportional, reaches the same value at a different time t_2, so that

$$\cos(\omega t_2 + \alpha + \pi/2) = a$$
$$= \cos(\omega t_1 + \alpha).$$

Therefore, $\omega t_2 + \pi/2 = \omega t_1$, whence $t_2 = t_1 - \pi/2\omega = t_1 - T/4$, which means the voltage leads the current by *one quarter of a period.*

Quite naturally, we can add any integral number of periods to t_1 and write $t_2 = t_1 - T/4 + T = t_1 + (3/4)\,T$ or $t_2 = t_1 + (7/4)\,T$. The formula indicates the smallest (in absolute value) time shift that carries the current curve into the voltage curve.

Let us consider the case of *capacitance.* Here, $j = C\,(d\varphi_C/dt)$ and so

$$\varphi_C = \frac{1}{C}\int j\,dt = \frac{1}{C}\int j_0 \cos(\omega t + \alpha)\,dt$$
$$= \frac{j_0}{C\omega}\sin(\omega t + \alpha) \qquad (13.11.4)$$

(the constant of integration is equal to $\overline{\varphi}_C$, but for alternating current it is always true that $\overline{\varphi}_C = 0$). Writing φ_C as $\varphi_C = \varphi_3 \cos(\omega t + \alpha_3)$, we get $\varphi_3 = (1/C\omega)\,j_0$ and $\alpha_3 = \alpha - \pi/2$.

Thus, in an alternating-current circuit the relationship between the amplitude of current and the amplitude of voltage is the same on a capacitance as on a resistance equal to $R_3 = 1/C\omega$. Expressing capacitance in farads (F) and frequency in reciprocal seconds, we obtain R_3 in ohms.

In a capacitance, the voltage curve is shifted *forward* with respect to the current curve by a quarter-period (Figure 13.11.3). Thus, the curve of voltage φ_C in a capacitance is shifted in the opposite direction to the curve of voltage φ_L in an inductance.

For a given current, φ_L and φ_C are of opposite sign. If the curves of φ_L and φ_C are brought to coincidence, we will see that the current flowing through the capacitance and the current flowing through the inductance have opposite signs. Indeed, all formulas expressing φ as a function of t can readily be transformed into formulas expressing j as a function of t:

$$\varphi = \varphi_0 \cos(\omega t + \alpha),$$
$$\varphi_R = Rj_0 \cos(\omega t + \alpha),$$
$$\varphi_L = L\omega j_0 \cos(\omega t + \alpha + \pi/2)$$
$$= -L\omega j_0 \sin(\omega t + \alpha),$$
$$\varphi_C = \frac{1}{C\omega} j_0 \cos\left(\omega t + \alpha - \frac{\pi}{2}\right)$$
$$= \frac{1}{C\omega} j_0 \sin(\omega t + \alpha)$$

and

$$j = j_0 \cos(\omega t + \alpha),\quad j_R = \frac{\varphi_0}{R}\cos(\omega t + \alpha),$$
$$j_L = \frac{\varphi_0}{L\omega}\cos\left(\omega t + \alpha - \frac{\pi}{2}\right)$$
$$= \frac{\varphi_0}{L\omega}\sin(\omega t + \alpha),$$
$$j_C = \varphi_0 C\omega \cos\left(\omega t + \alpha + \frac{\pi}{2}\right)$$
$$= -\varphi_0 C\omega \sin(\omega t + \alpha).$$

The opposite phase shift and the opposite signs in the formulas referring to

inductance and capacitance are of crucial importance when considering L and C connected in one circuit.

In alternating-current experiments, one frequently makes use of a *single-beam cathode-ray oscillograph*. A voltage proportional to the current is impressed on one pair of deflection plates (deflection along the x axis) and a voltage proportional to φ is applied to the other pair of deflection plates (deflection along the y axis). The beam moves along a line whose equation has the form $x = aj$ and $y = b\varphi$, where the coefficients a and b depend on the sensitivity of the oscillograph. Since j and φ are periodic functions of time, the beam sweeps out the same curve on the screen all the time. At 50 cycles per second, the human eye cannot detect any motion of the ray and sees a solid luminous curve.

If the potential difference from the *resistance*, φ_R, is impressed on the vertical deflection plates of the oscillograph, the ray describes a *straight line*. True enough, for

$$x = aj = aj_0 \cos(\omega t + \alpha),$$
$$y = b\varphi_R = bRj_0 \cos(\omega t + \alpha).$$

Eliminating t, we find that $y = (bR/a)\,x$. If to these plates we apply the potential difference from the *capacitance*, φ_C, the result is an *ellipse*:

$$x = aj_0 \cos(\omega t + \alpha),$$
$$y = b\frac{1}{C\omega} j_0 \sin(\omega t + \alpha),$$

whence

$$\left(\frac{x}{aj_0}\right)^2 + \left(\frac{y}{bj_0/C\omega}\right)^2 = \cos^2(\omega t + \alpha) + \sin^2(\omega t + \alpha) = 1.$$

Also, an *ellipse* results from φ_L (*inductance*). If we connect $\varphi_R + \varphi_L$ or $\varphi_R + \varphi_C$, the axes of symmetry of the ellipse no longer coincide with the x and y axes. Thus, the shape of an oscillogram tells us what the circuit is made up of (C, L, or R), what the "innards of the box" consist of (see Figure 13.1.7).

13.12 Average Quantities. Power and Phase Shift

In the preceding section, the current and voltage were regarded as functions of time. However, in many cases it suffices to know the *average (constant) values* of these quantities (cf. Section 7.8).

As an elementary case, let us consider a heating device with resistance R. We know that in a direct-current (dc) circuit the power output (that is, the quantity of energy released per unit of time) is $h = \varphi j = Rj^2 = \varphi^2/R$. In an alternating-current (ac) circuit the power output varies, that is, the (instantaneous) power output at time t is

$$h(t) = \varphi(t)\, j(t) = Rj^2(t) = \varphi^2(t)/R.$$

Therefore (see (13.11.1)),

$$h(t) = Rj_0^2 \cos^2(\omega t + \alpha). \qquad (13.12.1)$$

Over one period, $h(t)$ becomes zero twice and reaches a maximum (equal to Rj_0^2) twice. When considering electric heaters, however, we are usually interested in the amount of heat generated during a time interval t that is many times greater than the period T of alternating current (in the USSR this period is usually 0.02 s). It is therefore sufficient to know the average value of the power output over a large time interval t. In view of (13.12.1),

$$\overline{h} = \overline{Rj_0^2 \cos^2(\omega t + \alpha)} = Rj_0^2 \overline{\cos^2(\omega t + \alpha)} \simeq \frac{Rj_0^2}{2}, \qquad (13.12.2)$$

since, as we repeatedly noted in this book, $\overline{\cos^2(\omega t + \alpha)} = 1/2$ (cf. Section 7.8 or 10.4). Equation (13.12.2) is *approximate*, being the more exact the greater t is.

The ***average value of an alternating current***, $\overline{j}$, is ordinarily defined as the *intensity of a direct current that generates an equivalent power output on a re-*

sistance R:[13.12]

$$R\bar{j}^2 = Rj_0^2/2, \qquad (13.12.3)$$

whence

$$\bar{j} = \frac{1}{\sqrt{2}} j_0 \simeq 0.71\, j_0 \qquad (13.12.4)$$

In the same way, the average value of the voltage, $\bar{\varphi}$, is determined from the condition

$$h = \frac{\bar{\varphi}^2}{R} = \frac{\varphi_0^2}{2R},$$

whence

$$\bar{\varphi} = \sqrt{\bar{\varphi}^2} = \frac{1}{\sqrt{2}} \varphi_0 \simeq 0.71\, \varphi_0. \qquad (13.12.5)$$

Instruments that measure alternating current (ammeters and voltmeters) are calibrated so that they give the *average value* of current or voltage, $\bar{j}$ or $\bar{\varphi}$.

From formulas (13.12.4) and (13.12.5) it follows that the maximum values of current and voltage attained in an ac circuit exceed the average values by a factor of $\sqrt{2} \simeq 1.41$. For example, in a circuit with an average voltage of 220 volts the maximum instantaneous voltage (peak voltage) reaches ± 310 volts.

From the relations (13.12.2), (13.12.4), (13.12.5) and the formula $\varphi_0 = Rj_0$ it follows that $\bar{\varphi} = R\bar{j}$ and $\bar{h} = \bar{\varphi}\bar{j}$, so that Ohm's law and the relationship between power, current, and voltage on a resistance hold true for mean values.

When we considered alternating current flowing through a capacitance and an inductance, we saw that the current and voltage vary along curves that are shifted with respect to one another, although the frequency is the same. Let us consider the power output in the general case of an arbitrary phase shift of j with respect to φ (that is, an arbitrary shift of one curve with respect to another). Let

$$j = j_0 \cos(\omega t + \beta),$$
$$\varphi = \varphi_0 \cos(\omega t + \beta + \alpha).$$

Then

$$h(t) = j_0\varphi_0 \cos(\omega t + \beta) \cos(\omega t + \beta + \alpha).$$

But

$$\cos(\omega t + \beta + \alpha) = \cos(\omega t + \beta) \times \cos\alpha - \sin(\omega t + \beta)\sin\alpha,$$

so that

$$\cos(\omega t+\beta)\cos(\omega t+\beta+\alpha)$$
$$= \cos\alpha\cos^2(\omega t+\beta)$$
$$- \sin\alpha\cos(\omega t+\beta)\sin(\omega t+\beta)$$
$$= \cos\alpha\cos^2(\omega t+\beta)$$
$$- \sin\alpha\,\frac{1}{2}\sin(2\omega t+2\beta).$$

And since

$$\overline{\cos^2(\omega t+\beta)} = \frac{1}{2}, \quad \overline{\sin(2\omega t+2\beta)} = 0,$$

we have

$$\bar{h} = j_0\varphi_0\cos\alpha \times \frac{1}{2} = \bar{j}\bar{\varphi}\cos\alpha.$$

Thus, the average value of the power output in the general case where there is a phase shift α is proportional to $\cos\alpha$. In the particular case of a resistance, $\alpha = 0$, $\cos\alpha = 1$, and we return to formula (13.12.2).

In the case of a capacitance, $\alpha = -\pi/2$ and $\cos\alpha = 0$, while in the case of an inductance, $\alpha = +\pi/2$ and $\cos\alpha = 0$. Hence, in both cases *the average power output is equal to zero.* This result is quite understandable from the standpoint of physics. In a capacitance and an inductance, electric energy is not transformed into heat, it is

[13.12] Clearly, if j is given by (13.11.1), the average value $\bar{j}$ defined by the formulas of Section 7.8 is zero. For this reason it is advisable to define the average value (also called the mean value) as we do here: $\bar{j} \equiv (\overline{j^2})^{1/2}$ (in mathematics this quantity is known as the ***root-mean-square*** of j). The square root is required so that $\bar{j}$ should have the dimensions of current, while the operation of squaring makes it possible to avoid the mutual compensation of positive and negative current values. Finally, the sign $\equiv$ here means "equal by definition," that is, we introduce a new definition of $\bar{j}$, while $\overline{j^2}$ is defined in the usual manner.

merely stored up. In an ac circuit, a capacitance during one half of a period takes electric energy from the circuit and stores it, only to release it back into the circuit during the other half. The same goes for inductance in an ac circuit.

An ordinary transformer without any load is actually a pure inductance (if we ignore the slight losses in the wires). A current flows through the transformer with an amplitude $j = \varphi/L\omega$. However, as already stated, no power is taken from the circuit because $\alpha = \pi/2$ and $\cos \alpha = 0$. It is an interesting fact that electric meters are designed to measure just the quantity $j\varphi \cos \alpha$. For this reason, a nonloaded transformer will hardly add anything to your electric bill, it will only increase the total current flowing in the wires.

If a large number of inductances (transformers, unloaded electric motors, and the like) are connected in parallel, the total current can become large, and then the losses in the electric wiring will be rather noticeable. This effect is an important factor relative to the electric networks of a whole city. The laws that govern the flow of current through circuits with inductances and capacitances show how such losses can be minimized.

13.13. An Alternating-Current Oscillatory Circuit. Series Resonance

Let us now consider an ac circuit comprising resistance, inductance, and capacitance in series (Figure 13.13.1). It is obvious that in this system the current flowing through R, L, and C is the same. We write the current in the form

$$j = j_0 \cos (\omega t + \alpha). \qquad (13.13.1)$$

The potential difference in the series is

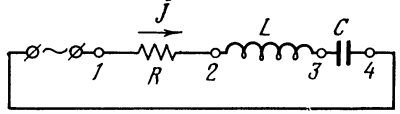


Figure 13.13.1

$$\varphi = \varphi_1 - \varphi_4 = \varphi_R + \varphi_L + \varphi_C.$$ Recalling the formulas (13.11.2) to (13.11.4), we get

$$\varphi = Rj_0 \cos (\omega t + \alpha) - L\omega j_0 \sin (\omega t + \alpha)$$

$$+ \frac{j_0}{C\omega} \sin (\omega t + \alpha) = Rj_0 \cos (\omega t + \alpha)$$

$$+ j_0 \left(\frac{1}{C\omega} - L\omega\right) \sin (\omega t + \alpha). \qquad (13.13.2)$$

We see from this formula that the potential difference on the inductance and that on the capacitance have different signs, and so the coefficient of $\sin (\omega t + \alpha)$ is the difference of two terms. We write φ as

$$\varphi = b \cos (\omega t + \alpha + \beta), \qquad (13.13.3)$$

where b is the amplitude of the potential difference, which is to say, the maximum value of potential difference (the peak voltage). To find b, we rewrite (13.13.3) as follows:

$$\varphi = b \cos \beta \cos (\omega t + \alpha)$$

$$- b \sin \beta \sin (\omega t + \alpha). \qquad (13.13.3a)$$

Comparing this expression with (13.13.2), we find that

$$b \cos \beta = Rj_0,$$

$$b \sin \beta = j_0 \left(L\omega - \frac{1}{C\omega}\right). \qquad (13.13.4)$$

Squaring both equations in (13.13.4), adding, and taking the square root of the sum yields

$$b = j_0 \sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2}. \qquad (13.13.5)$$

This formula shows that for a given amplitude of the current, j_0, the peak voltage b is at a minimum when

$$L\omega = \frac{1}{C\omega}. \qquad (13.13.6)$$

Writing (13.13.5) as

$$j_0 = b \Big/ \sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2},$$

we see that for a given peak voltage, b, the peak current, j_0, is at a maximum if condition (13.13.6) is met. This condition may be written thus: $\omega =$

$1/\sqrt{LC}$, which is nothing other than the natural frequency of an LC circuit. Therefore, (13.13.6) is the condition of ***resonance***, the condition of coincidence of the natural frequency of the circuit and the driving frequency of the alternating current. Observe that at resonance the circuit voltage is

$$\varphi = Rj_0 \cos(\omega t + \alpha). \qquad (13.13.7)$$

Using (13.13.1), we find that at resonance

$$\varphi = Rj. \qquad (13.13.8)$$

Let us now pass to *average values*. We determine the average values of current and potential difference in accord with formulas (13.12.4) and (13.12.5). From $\varphi_L = L\omega j_0 \sin(\omega t + \alpha)$ we obtain (see footnote 13.12)

$$\bar{\varphi}_L = L\omega \bar{j} = \frac{L\omega}{R}\bar{\varphi}. \qquad (13.13.9)$$

Similarly, from $\varphi_C = (1/C\omega)\, j_0 \times \sin(\omega t + \alpha)$ we get

$$\bar{\varphi}_C = \frac{1}{C\omega}\,\bar{j} = \frac{\bar{\varphi}}{C\omega R}. \qquad (13.13.10)$$

Formulas (13.13.7) to (13.13.10) are valid only in the case of resonance, that is, when $\omega_0 = 1/\sqrt{LC}$. Putting the value of ω into (13.13.9) and (13.13.10), we get

$$\bar{\varphi}_L = \bar{\varphi}_C = \frac{1}{R}\sqrt{\frac{L}{C}}\,\bar{\varphi}.$$

For this reason, when we have resonance, the voltage on the inductance and the capacitance is the greater the smaller the resistance R, and $\bar{\varphi}_L$ and $\bar{\varphi}_C$ may exceed the ac source voltage $\bar{\varphi}$ many times over.

In a series circuit the resistances are additive. But the "resistances" of the capacitance and the inductance are of opposite sign and are different functions of the frequency. At resonance frequency they are equal in absolute value and hence cancel each other.

Thus, at resonance, ($\omega = \omega_0$) an RLC series circuit has a minimum "resistance" and carries a maximum current for a given peak voltage as compared with that at any nonresonance frequency $\omega \neq \omega_0$.

It is of interest to investigate in detail how the peak voltage and the peak current vary in the case of departure from exact resonance, that is, when $\omega \neq 1/\sqrt{LC}$. To do this, let us take advantage of formula (13.13.3). We find that $\bar{\varphi} = b/\sqrt{2}$, whence $b = \sqrt{2}\bar{\varphi}$. Substituting this into (13.13.5), we get

$$\bar{\varphi} = \frac{j_0}{\sqrt{2}}\sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2}.$$

But $j_0/\sqrt{2} = \bar{j}$, and so

$$\bar{\varphi} = \bar{j}\sqrt{R^2 + (L\omega - 1/C\omega)^2}.$$

Finding $\bar{j}$ from this formula and putting it into (13.13.9) and (13.13.10), we obtain

$$\bar{\varphi}_L = \frac{L\omega}{\sqrt{R^2 + (L\omega - 1/C\omega)^2}}\,\bar{\varphi},$$

$$\bar{\varphi}_C = \frac{1}{C\omega}\,\frac{1}{\sqrt{R^2 + (L\omega - 1/C\omega)^2}}\,\bar{\varphi}.$$

If we denote by ω_0 the natural frequency of the circuit, then $\omega_0^2 = 1/LC$. These formulas can now be written thus:

$$\bar{\varphi}_L = \frac{\omega^2\bar{\varphi}}{\sqrt{\frac{R^2\omega^2}{L^2} + (\omega_0^2 - \omega^2)^2}},$$

$$\bar{\varphi}_C = \frac{\omega_0^2\bar{\varphi}}{\sqrt{\frac{R^2\omega^2}{L^2} + (\omega_0^2 - \omega^2)^2}}. \qquad (13.13.11)$$

In this form it is quite evident how the ratio $\bar{\varphi}_L/\bar{\varphi}$ or $\bar{\varphi}_C/\bar{\varphi}$ depends on the closeness of the natural frequency of the circuit, ω_0, to the driving frequency ω.

The ratio $\bar{\varphi}_L/\bar{\varphi}$ as a function of ω near $\omega = \omega_0$ is shown in Figure 13.13.2. The graph is constructed for the case of $R/L\omega_0 = 0.05$. It is illustrative of the typical ***resonance curve***.

If $R/L\omega \ll 1$, the dependence of $\bar{\varphi}_L/\bar{\varphi}$ or $\bar{\varphi}_C/\bar{\varphi}$ on ω is mainly determined by the second term of the radicand, $(\omega_0^2 - \omega^2)^2$. When $\omega = \omega_0$, this term

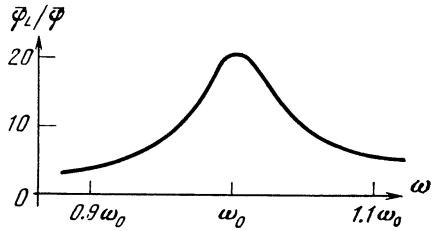


Figure 13.13.2

vanishes, and under the assumption that $R/L\omega \ll 1$ the denominator has a minimum and the ratio $\bar{\varphi}_L/\bar{\varphi}$ or $\bar{\varphi}_C/\bar{\varphi}$ a maximum. The ratio constitutes 70% of the maximum value when $(\omega_0^2 - \omega^2)^2 = R^2\omega^2/L^2$, that is, at $\omega_0^2 - \omega^2 = \pm R\omega/L$, whence

$$\omega_0 - \omega = \pm \frac{R}{L} \frac{\omega}{\omega_0 + \omega} \simeq \pm \frac{R}{2L}.$$

The variation in frequency for which the square of the ratio $\bar{\varphi}_L/\bar{\varphi}$ or $\bar{\varphi}_C/\bar{\varphi}$ falls to one half of the square of the maximum value is called the ***half-width of the resonance curve***. If the ratio is 70% (0.7) of the maximum, its square comes to $0.7^2 \simeq 0.5$ of the maximum value of the square. Therefore, the half-width of the resonance curve, $\omega - \omega_0$, constitutes $R/2L$, which means that the half-width (also called the ***half-width at half-maximum***) is equal to the quantity that characterizes the rate of decay of oscillations in that circuit (see Section 13.9).

Consequently, the smaller the resistance R, the smaller the half-width of the resonance curve and the steeper the curve near $\omega = \omega_0$. From formulas (13.13.11) we see that the smaller the resistance R the greater the maximum of the ratio $\bar{\varphi}_L/\bar{\varphi}$ or $\bar{\varphi}_C/\bar{\varphi}$. That is why the phenomenon of resonance is particularly evident if R is small.

13.14 Inductance and Capacitance in Parallel. Parallel Resonance

Consider the circuit in Figure 13.14.1, which differs from that in Figure 13.13.1 in that L and C are in parallel. We take the resistance of the circuit to be

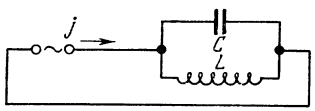


Figure 13.14.1

extremely small (in other words, we neglect it). Then φ_C and φ_L coincide and are equal to the voltage φ in the circuit (that is, of the ac source), while the current j is made up of the current j_C flowing through C and the current j_L flowing through L. Let $\varphi_L = \varphi_C = \varphi = \varphi_0 \cos(\omega t + \alpha)$. Using the formulas of Section 13.11, we find that

$$j_C = -C\omega\varphi_0 \sin(\omega t + \alpha),$$

$$j_L = \frac{\varphi_0}{L\omega} \sin(\omega t + \alpha).$$

Therefore

$$j = j_C + j_L$$
$$= \varphi_0 \left(\frac{1}{L\omega} - C\omega\right) \sin(\omega t + \alpha),$$

whence, assuming that $\omega_0 = 1/\sqrt{LC}$,

$$\bar{j} = \bar{\varphi} \left(\frac{1}{L\omega} - C\omega\right),$$

$$\bar{\varphi} = \frac{\bar{j}}{1/L\omega - C\omega} = \frac{\omega\bar{j}}{C(\omega_0^2 - \omega^2)}.$$

In this case too we see a typical resonance relationship: for a given current $\bar{j}$, the voltage $\bar{\varphi}$ is the greater the closer ω is to ω_0. It is easy to see that when ω is close to ω_0, $\bar{j}_L$ and $\bar{j}_C$ are much greater than the current j in the circuit, that is, an ac circuit containing L and C experiences strong oscillations. A small external current suffices to sustain much stronger currents in the circuit.

It will be recalled that in a parallel circuit, *conductances* (which are the reciprocal of resistances) are additive:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots.$$

The "conductances" (that is, the ratios of current to potential difference) of a capacitance and an inductance have opposite signs and depend differently

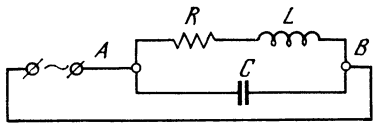


Figure 13.14.2

on the frequency. At resonance ($\omega = \omega_0$) they cancel each other and the total "conductance" is at a minimum, which is to say, the current is the smallest for a given potential difference and, hence, the potential difference φ_{AB} is at a maximum for a given current in the external circuit.

In a simplified circuit without resistance, the amplitude of oscillations grows without limit as ω approaches ω_0. In reality, the resistance in the circuit makes the amplitude finite when $\omega = \omega_0$.

If R is connected in parallel with L and C, all calculations become very similar to those of the preceding section. But this case is rarely encountered in practical situations. Ordinarily, the inductance has a perceptible "resistance" and therefore the typical circuit diagram is that shown in Figure 13.14.2. In this case the calculations are somewhat longer than in the preceding section and we will not carry them through in detail. The result for ω close to ω_0 and for small ratios $R/L\omega$ is

$$\frac{\bar{j}_L}{\bar{j}} \simeq \frac{\bar{j}_C}{\bar{j}} = \frac{\omega_0^2}{\sqrt{(R\omega/L)^2 + (\omega_0^2 - \omega^2)^2}}.$$

It thus turns out that current amplification at resonance in a parallel ac circuit obeys the same law as voltage amplification in a series circuit discussed in Section 13.13.

13.15 General Properties of Resonance in a Linear System

Note in our reasoning that when we added oscillations caused by a force at different moments of time, we assumed all along that the system without that force was *linear*. The behavior of the system is described by a linear differential equation, and the unknown quantity appears in all terms in the first power. Therefore, the superposition principle is valid, that is, the sum of two or several particular solutions of the equation is also a solution.

In Section 13.13 we obtained a formula for the half-width of a resonance curve, $\omega - \omega_0 = R/2L$, which shows that the slower the decay of oscillations the smaller the half-width. This does not occur only with respect to electric oscillations. We can consider any system capable of oscillation. Let an external force give rise to oscillations in such a system and then cease to operate. The system is now on its own. The oscillations begin to decay. If the amplitude of oscillations decays like $e^{-\gamma t}$, then γ characterizes the rate of decay (it has the dimensions s^{-1}). During time $\tau = 1/\gamma$ the amplitude diminishes by a factor of e, or by 63%.

Let us now consider the resonant stepup of such a system by a periodic external force. The amplitude of oscillations at a given time is the sum of the amplitudes acquired during the time of stepup. In the presence of damping, an amplitude acquired a long time before will have time to decay and will not play any part or make any contribution to the amplitude of oscillation at the given time.

The decay time is clearly $1/\gamma$. In this time the amplitude of free oscillations will have diminished e times, or by 63%. Hence, even when the stepup force is constantly operating from $t = -\infty$, the oscillation amplitude will still be determined solely by the time interval from $t - 1/\gamma$ to t, where t is the time of observation. The action of the force at earlier times will have already decayed.

For the difference between two periodic forces with somewhat distinct periods, $F_0 \sin \omega_0 t$ and $F_0 \sin \omega t$, to manifest itself conspicuously, a time T of observation is needed during which their phases will have separated by approximately π units: $\omega T = \omega_0 T \pm \pi$, so that $|\omega - \omega_0| = \pi/T$. Consequently,

if the oscillations of a system "remember" only the action of the force during time $T = 1/\gamma$, then in such a system a difference $\pi/T = \pi\gamma$ in the frequencies of the exciting force hardly affects the amplitude. From this we see that the half-width of the resonance curve is proportional to the decay factor γ.

On the other hand, since the system "remembers" and accumulates the action of a force during time $1/\gamma$, the oscillation amplitude at resonance (hence also the height of the resonance peak) is inversely proportional to γ. Calculations confirm this reasoning.

We would like to remind the reader once more that the system we are considering is, in the absence of an external force, *linear*. The behavior of such a system is described by a linear differential equation, thus making the ***superposition principle*** formulated above quite applicable. For *nonlinear* systems, the superposition principle fails, and the theory of such systems proves much more complex. Also note that a linear system does not generate oscillations even if a constant emf source (an ideal source) is present in the circuit.

13.16* Displacement Current and the Electromagnetic Theory of Light

Up to now we have almost everywhere considered current flow through a capacitor without any reservations. Indeed, if we connect a capacitor in an ac circuit in series with an ammeter, the ammeter will indicate a definite current $\bar{j} = C\omega\bar{\varphi}$. On the other hand, no current flows through a capacitor because the plates of the capacitor are separated by an insulator (air or a material or even a vacuum), and so the individual current carriers (electrons) in the left-hand conductor and plate will never get over to the right-hand plate and conductor. Consequently, no charged particles are in motion in the space between the plates, that is, there is no electric current in the sense that we have spoken of current up to now. All there is in this space is an ***electric field*** that varies when the charge on the plates varies, which is to say, when a current flows in the right- and left-hand conductors. We can now do one of two things:

(1) either beg the reader's pardon and explain that whatever we have spoken of current flowing through a capacitor (capacitance) this was not so—actually there was no current, the only current flow being in the conductors on the right and left;

(2) or regard the varying electric field in the space between the plates on a par with ordinary current (the motion of charged particles). Maxwell, who suggested this view, was able to draw conclusions of tremendous significance.

It had long since been known that *electric current* (the motion of charged particles) gives rise to a *magnetic field.* But if a varying electric field is similar to an electric current, an electric field varying in a vacuum should also set up a magnetic field. This hypothesis of Maxwell led to a remarkable symmetry between electric and magnetic fields. Faraday experimentally discovered ***induction***, that is, the fact that any variation of a magnetic field gives rise to an electric field. Maxwell, in strictly theoretical fashion, hypothesized the existence of a similar phenomenon in which any variation of an electric field gives rise to a magnetic field. Only then did the theory of electric and magnetic fields acquire its modern form.

The mathematical theory of Maxwell is written in the form of differential equations that are too complicated for this book and so we do not give them.[13.13]

[13.13] The characteristics of electric and magnetic fields in space are functions of *several* variables—they depend on the three coordinates of a point in space and on time. Accordingly, the derivatives that are present in the Maxwell equations are the *partial derivatives* of functions of several variables, and the equations themselves belong to the class of ***partial differential equations*** (in the Conclusion we touch on such equations). Moreover, these characteristics are *vector* quantities. Finally, note that complex numbers and the theory of analytic functions of a complex variable (see Chapter 17) fit elegantly into the theory of electromagnetism.

The solutions to these equations describe the propagation of electric and magnetic fields in empty space. Both fields must be present at all times: a variation in the electric field gives rise to a magnetic field, and any change in the magnetic field generates an electric field.

At the time when Maxwell worked, Faraday's experiments had already been completed, that is, the relationship between a varying magnetic field and an emf induced by it was known. Also known was the magnetic field of a current. Finally, the relationship between the charge on a capacitor and the electric field between the plates was likewise known. These findings sufficed for writing down the equations for fields in empty space.

Maxwell found the rate of propagation of the fields in vacuo. This velocity proved to be equal to that of light! From this it was natural to conclude that light is nothing other than electromagnetic oscillations. Furthermore, the theory predicted the possibility of the existence of electromagnetic oscillations of any wavelength including X rays (whose wavelength is thousands of times shorter than that of visible light) and radio waves with very large wavelengths. It was thus that the investigations of Faraday and Maxwell began the work that culminated in the discovery of radio waves by Hertz, the invention of radio as a means of communication by the Russian A. S. Popov and the Italian Marchese G. Marconi, the discovery of X rays by Wilhelm K. Röntgen, and the creation of a complete theory of the electromagnetic field and the interaction of this field with matter.

13.17* Nonlinear Resistance and the Tunnel Diode

Let us consider a two-terminal network (a "box") that is similar to a resistance in the sense that current flowing through the "box" depends solely on the instantaneous value of the potential difference. In this respect, the "box" is not like an inductance, where φ depends on dj/dt, neither is it like a capacitance, where φ depends on $\int j\,dt$. However, the "box" differs from an ordinary resistance in that the function $j(\varphi)$ differs from Ohm's law $j = \varphi/R$. The "box" has a more involved function $j(\varphi)$. This function is called the ***characteristic curve*** of the "box".

The only general assertion that can be made with respect to $j(\varphi)$ is that φ and j cannot have different signs if batteries or some sources of energy are not hidden in the box. If φ and j are of the same sign, energy *is absorbed* inside the box during current flow and the box takes up electric energy from the circuit to which it is connected. In the box this electric energy is converted into heat and is dissipated. Since φ is constant and negative when $j < 0$ and positive when $j > 0$, it follows that $\varphi = 0$ at $j = 0$. In all other respects the functional relationship between j and φ can be of any kind. For example, for current rectifiers we have boxes whose characteristic curve is shown in Figure 13.17.1: the current flows easily in one direction for a small potential difference and hardly at all in the other direction. As can be seen from the graph, the current is small even for a large negative potential difference. Such are the properties possessed by ***diodes*** made of two semiconductors. In 1958 the Japanese devised a "box" made up of specially chosen semiconductors, the ***tunnel diode*** (in reality, this "box" is in the form of a minute cylinder just a few millimeters in diameter and altitude), which has an unusual curve of $j(\varphi)$ with a minimum (see

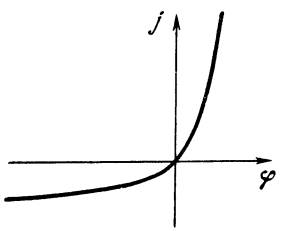


Figure 13.17.1

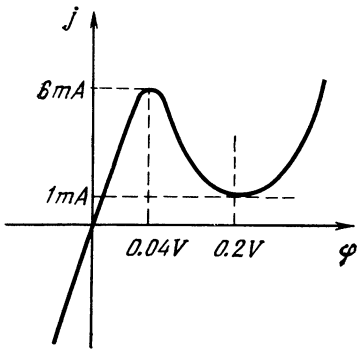


Figure 13.17.2

Figure 13.17.2 in which typical values of φ and j are indicated).[13.14] This curve does not contradict the principle expressed above: the sign of φ is the same as that of j everywhere, which means the "box" only absorbs energy. We will not go into the physical reasons for such a strange curve, but we will examine the consequences for a circuit involving a tunnel diode. For the sake of brevity we will continue to call it a box.

We start with the simplest type of circuit consisting of three parts: a battery with emf E, a resistance R (the ordinary kind that obeys Ohm's law), and the box (Figure 13.17.3). We include the internal resistance of the battery in R.

The equation defining the current and distribution of potential in the circuit is of the form $-E + Rj(\varphi) + \varphi = 0$, where φ is the potential difference across the box and $j(\varphi)$ is the function defined by the properties of the box (see Figure 13.17.2). The current through R is equal to the current through the box, and so $\varphi_R = Rj = Rj(\varphi)$. This equation is conveniently solved by graph. We write it down as

$$\varphi = E - Rj(\varphi) \qquad (13.17.1)$$

and we construct in the φj-plane the straight line $\varphi = E - Rj$. This line may be called the load curve of the battery-resistance system. The solution to the problem is given by the intersection of the straight line (13.17.1) (that is, the straight line $j = (E - \varphi)/R$) with the $j(\varphi)$ curve, which is the characteristic curve of the box. In Figure 13.17.4 we have a graphic solution of the problem involving one battery and three different resistances: R_1 (small), R_2 (medium), and R_3 (large).

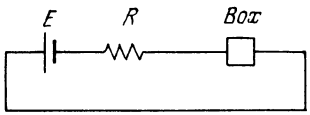


Figure 13.17.3

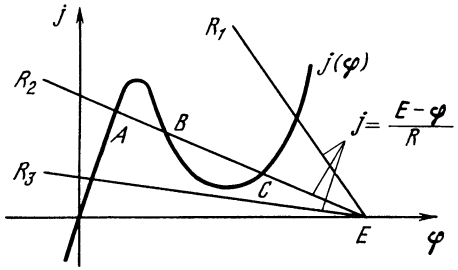


Figure 3.17.4

From the graph we can see that for a sufficiently large E we can choose an R such that it will not be too small or too large and there will be *three* points of intersection, A, B, and C, and thus three solutions to Eq. (13.17.1).

For three solutions to exist, the $j(\varphi)$ curve must have a descending portion. It is clear that the line on which $dj/d\varphi$ is positive everywhere can only once intersect the load curve, no matter what E and $R > 0$.

Now let us consider a somewhat more complicated circuit diagram involving capacitance in parallel with the box (Figure 13.17.5). For this circuit we

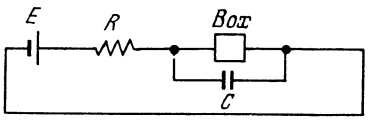


Figure 13.17.5

[13.14] By the end of the 1970s both physicists and designers had greatly advanced the miniaturization of all parts of electronic devices. The modern diode (or capacitor or resistor) now comes in the form of a tiny spot a dwindling fraction of millimeter in diameter and consists of multiple layers with the specified properties, hundredths or even thousandths of a millimeter thick.

find that the current through the box, $j(\varphi)$, and the current flowing through the capacitance, $C(d\varphi/dt)$, together equal the current flowing through the resistance,

$$j(\varphi) + C\frac{d\varphi}{dt} = \frac{E-\varphi}{R},$$

whence

$$\frac{d\varphi}{dt} = \frac{1}{C}\left[\frac{E-\varphi}{R} - j(\varphi)\right].$$

The intersection points of the characteristic curve $j(\varphi)$ of the box with the load curve $(E-\varphi)/R$ correspond to the solutions φ = constant, $d\varphi/dt = 0$. Let us examine the sign of $d\varphi/dt$ near these points. A glance at Figure 13.17.4 shows us that $d\varphi/dt$ is positive if $\varphi < \varphi_A$ or $\varphi_B < \varphi < \varphi_C$ and negative if $\varphi_A < \varphi < \varphi_B$ or $\varphi > \varphi_C$.

The arrows in Figure 13.17.6 indicate the direction of variation of φ with time. We can see that the intermediate solution at point B is *unstable*: all we need to do is depart slightly to the left or to the right, and $d\varphi/dt$ acquires a sign such that the deviation of φ from φ_B increases. On the other hand, the points A and C in Figure 13.17.4 characterize two *stable* solutions, which correspond to stable states of the system.

The existence of two stable states permits using these boxes in mathematical machines as memory cells. By making a lot of such circuits and transferring (by an external means) some into the A state and others into the C state, we can record ("remember") any desired number or other information. Using such systems, we record information in coded form as $AACACACCC\ldots$, where each letter A or C indicates a state of the appropriate system (the first in A, the second in A, the third in C, the fourth in A, etc.).

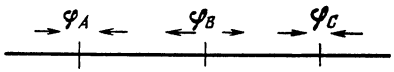


Figure 13.17.6

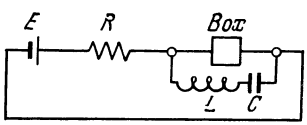


Figure 13.17.7

We now consider a system (Figure 13.17.7) consisting of an inductance and a capacitance connected in parallel with the box. We again denote by φ the potential difference across the box, $j = j(\varphi)$ the characteristic curve of the box, and j_1 the current flowing through L and C:

$$j(\varphi) + j_1 = \frac{E-\varphi}{R}, \quad C\frac{d\varphi_C}{dt} = j_1,$$

$$\varphi - \varphi_C = \varphi_L = L\frac{dj_1}{dt}. \qquad (13.17.2)$$

We consider the process in the circuit when the current and the potential in the box are close to the intermediate point B of intersection. Let

$$\varphi = \varphi_B + f, \quad \varphi_C = \varphi_B + g,$$

$$j \simeq j(\varphi_B) + f\left.\frac{dj}{d\varphi}\right|_{\varphi=\varphi_B} = j(\varphi_B) + kf,$$

where k stands for the derivative $dj/d\varphi$ at $\varphi = \varphi_B$. We used the first two terms of the Taylor series to represent current; the intermediate potential of the capacitance is φ_B, and φ_B and $j(\varphi_B)$ satisfy the condition $j(\varphi_B) = (E - \varphi_B)/R$. Substituting these expressions into Eq. (13.17.2) and simplifying, we get

$$kf + j_1 = -\frac{f}{R}, \quad C\frac{dg}{dt} = j_1,$$

$$f - g = L\frac{dj_1}{dt}. \qquad (13.17.3)$$

From these equations it follows that

$$f = -\frac{1}{1/R+k}j_1 = -rj_1, \qquad (13.17.4)$$

where $r^{-1} = R^{-1} + k = R^{-1} + (dj/d\varphi)_{\varphi=\varphi_B}$. Equations (13.17.3) and (13.17.4) also yield

$$L\frac{d^2j_1}{dt^2} = \frac{df}{dt} - \frac{dg}{dt} = -r\frac{dj_1}{dt} - \frac{1}{C}j_1,$$

and finally

$$\frac{d^2 j_1}{dt^2} + \frac{r}{L}\frac{dj_1}{dt} + \frac{1}{LC} j_1 = 0, \qquad (13.17.5)$$

which is the ordinary equation describing an oscillatory circuit with capacitance C, inductance L, and resistance r. The resistance r reflects the fact that the capacitance and inductance are in parallel to two circuits, the circuit involving the battery and resistance R and the circuit with box and resistance $k^{-1} = d\varphi/dj$. Since the two circuits are connected in parallel, the conductances (reciprocals of resistances) are additive, whence follows the expression for r.

The currents and potentials in the system break up into two terms: the constant term (φ_B, $j(\varphi_B)$) and the oscillatory term ($j_1(t)$, $f(t)$, $g(t)$). Here, for the oscillatory term the role of resistance of the box is played by the derivative $d\varphi/dj$ taken along the characteristic curve. If the box consisted of an ordinary (ohmic) resistance, $\varphi = Rj$, the derivative would be equal to the resistance, $d\varphi/dj = R$.

What does the unusual characteristic curve $j(\varphi)$ of the box (tunnel diode) of Figure 13.17.4 lead to? At point B the derivative $dj/d\varphi$ is negative, which means that with respect to oscillations the box has a negative resistance! What is more, from Figure 13.17.4 it is evident that at point B we have $|k| = |dj/d\varphi| > R^{-1}$, since R^{-1} is precisely the slope of the load curve $j = (E - \varphi)/R$ intersecting the characteristic curve at the point B. Consequently, the total resistance of the oscillatory circuit, $r = (R^{-1} + k)^{-1}$, is negative.

The equation for an oscillatory circuit involving L, C, and r, with r positive, yielded damped oscillations. For $r < 0$ this equation will yield stepup oscillations, which grow with time. Thus, a tunnel diode is capable of generating oscillations in a circuit.

The capability of generating oscillations is a consequence of the instability of the solution at point B. The oscillation energy is taken from the battery. The oscillation amplitude increases with time by an exponential law only so long as it may be considered small and we can employ the Taylor series expansion of the characteristic curve $j(\varphi)$ about the point B. Roughly speaking, the maximum amplitude is limited by the points A and C in Figure 13.17.4.

Already by 1961 oscillators with efficiencies up to 25% and power outputs of 0.5 MW at 7500 MHz (at a 4-cm wavelength) had been tested. For us, tunnel-diode circuits are interesting from the standpoint of a mathematical consideration of a nonlinear problem, questions of power output, and the representation of currents in a system as a superposition of a constant solution and oscillations.

Part 3

Some Additional Topics

Chapter 14 Complex Numbers

14.1 Basic Properties of Complex Numbers

The use of ***complex numbers*** can shed additional light on some topics discussed in this book. Originally, the number (natural) characterized the number of objects in any (finite) their set, for instance the number of children in a family, of boats on a river, or of the fingers of the hand. Numbers can be added: if we unite two groups of people consisting of a and b persons, we will get $a + b$ people altogether. Numbers can also be multiplied: a bunches of b flowers each will yield ab flowers. But sometimes natural numbers cannot be subtracted or divided. Manipulating with just natural numbers we cannot subtract 5 from 8 or divide 3 by 2.

The quest for carrying out operations of subtraction and division led to the appearance of ***negative*** and ***fractional*** (***rational***) ***numbers***: it was accepted to designate the difference $5 - 8$ as -3 and the dividend $3 \div 2$ as $3/2$. New numbers had quite a different meaning since neither group of people can contain -3 or $3/2$ persons. (Not without reason, Samuel Marshak, an outstanding Soviet poet and translator, wrote about a schoolboy who on solving the problem obtained $2^2/_3$ diggers.) Operations performed on such numbers are also defined in a new way; indeed it is senseless, say, to unite -3 boats and $3/2$ boats in a fleet. We have to assume, for instance, that the sum of the fractions a/b and c/d is equal to $(ad + bc)/bd$, and the product of $(-a)$ by $(-b)$ is $+ab$.[14.1]

That numbers can be multiplied without limit leads to the operation of raising a number to a (integral positive) power: $a^2 = a \cdot a$, $a^3 = a \cdot a \cdot a$, and generally $a^n = \underbrace{a \cdot a \cdot a \ldots a}_{n \text{ times}}$. But the inverse operation, that is, taking roots of ordinary numbers, can not always be performed: there is no (positive or negative) number x whose square would be equal to, say, -7, that is, there is no number $x = \sqrt{-7}$. In what follows we will call ordinary numbers, that is, positive, negative, and zero, ***real numbers***.[14.2] Attempts at taking (square) root of any real number lead also to the notion of complex numbers.

[14.1] These conventions only make it possible to apply all well-known operations performed on "old" numbers to "new" numbers (when extending the notion of number we have just to refresh our knowledge and not to start from the very beginning; so, according to our rules $a + b = a/1 + b/1 = (a \times 1 + 1 \times b)/1 \times 1 = (a + b)/1 = a + b$). All the general rules of operation are also valid (for instance, such as $a + b = b + a$ or $(a + b)\, c = ac + bc$). The same reasoning also holds in considering the rules of operation on *complex* numbers we are dealing with in this chapter.

[14.2] To make it possible to take square root of any positive number, we have to include in the real numbers not only all (positive and nonpositive) fractions (rational numbers), but also *irrational* numbers, such as $\sqrt{2} = 1.41 \ldots$. (In this connection see, for instance, I. Niven, *Numbers: Rational and Irrational*, New York, 1961, which is intended for a wide circle of readers.)

We introduce a new "number" $\sqrt{-1}$ defined by the condition that its square equals the number -1. Clearly, none of real numbers—either positive or negative—possesses such a square, and therefore we call it the ***imaginary unit***; in mathematics it is usually denoted i. Expressions of the form $c = a + ib$ or $z = x + iy$, where a and b (or x and y) are real, are named ***complex numbers***. Numbers of the form ib $(= 0 + ib)$ or iy are sometimes called ***pure imaginary*** (or simply ***imaginary***).

Four arithmetical operations on complex numbers $a = a_1 + ia_2$ and $b = b_1 + ib_2$ directly follow from the definition of the imaginary unit. They are introduced as the natural generalization of ordinary operations on real numbers: if $a + b = f$, $a - b = g$, and $ab = h$, then

$$f = f_1 + if_2 = (a_1 + ia_2) + (b_1 + ib_2) = (a_1 + b_1) + i(a_2 + b_2);$$

$$g = g_1 + ig_2 = (a_1 + ia_2) - (b_1 + ib_2) = (a_1 - b_1) + i(a_2 - b_2);$$

$$h = h_1 + ih_2 = (a_1 + ia_2)(b_1 + ib_2) = a_1b_1 + ia_1b_2 + ia_2b_1 + ia_2ib_2 = (a_1b_1 - a_2b_2) + i(a_1b_2 + a_2b_1)$$

(in calculating h we take into consideration that $i^2 = 1$); therefore

$$f_1 = a_1 + b_1, \quad f_2 = a_2 + b_2;$$

$$g_1 = a_1 - b_1, \quad g_2 = a_2 - b_2;$$

$$h_1 = a_1b_1 - a_2b_2, \quad h_2 = a_1b_2 + a_2b_1.$$

Similarly, if $k = a/b = k_1 + ik_2$, then $a = bk$, that is,

$$(b_1 + ib_2)(k_1 + ik_2) = (b_1k_1 - b_2k_2) + i(b_1k_2 + b_2k_1) = a_1 + ia_2.$$

For determining k_1 and k_2 we have now two linear equations,

$b_1k_1 - b_2k_2 = a_1$ and $b_2k_1 + b_1k_2 = a_2$,

which can be solved in all cases where $b \neq 0$ (that is, when b_1 and b_2 are not zero simultaneously):

$$k_1 = \frac{a_1b_1 + a_2b_2}{b_1^2 + b_2^2}, \quad k_2 = \frac{a_2b_1 - a_1b_2}{b_1^2 + b_2^2}.$$

When finding the quotient $k = a/b$, we can use the fact that at any (complex) number $b = b_1 + ib_2$ the product of b by the number $b^* = b_1 - ib_2$ (the number b^* is often denoted $\bar{b}$ and called the ***conjugate to*** b) is always *real*:

$$bb^* = (b_1 + ib_2)(b_1 - ib_2) = b_1^2 - (ib_2)^2 = b_1^2 - i^2b_2^2 = b_1^2 + b_2^2. \tag{14.1.1}$$

(If $b^* = b$, the number b is real; the square root of the product $\sqrt{bb^*} = \sqrt{b_1^2 + b_2^2}$ is denoted $|b|$ and called the ***absolute value***, or ***modulus***, of the complex number b.) Therefore

$$k = \frac{a}{b} = \frac{a_1 + ia_2}{b_1 + ib_2} = \frac{(a_1 + ia_2)(b_1 - ib_2)}{(b_1 + ib_2)(b_1 - ib_2)} = \frac{(a_1b_1 + a_2b_2) + i(a_2b_1 - a_1b_2)}{b_1^2 + b_2^2} = \frac{a_1b_1 + a_2b_2}{|b|^2} + i\frac{a_2b_1 - a_1b_2}{|b|^2}.$$

Our rules permit representing any rational function of the complex variable $z = x + iy$ (for instance, $f = z/(1 + z^2)$) as the sum $f = u + iv$, where u and v are two real-valued functions of two real variables x and y (see Exercise 14.1.1).

The problem of taking the square root of a complex number can be solved in a similar way. Indeed, if $r = r_1 + ir_2 = \sqrt{a} = \sqrt{a_1 + ia_2}$, then $(r_1 + ir_2)^2 = (r_1^2 - r_2^2) + i2r_1r_2 = a = a_1 + ia_2$, which results in the following system of equations for determining the unknowns r_1 and r_2:

$$r_1^2 - r_2^2 = a_1, \quad 2r_1r_2 = a_2.$$

This system is readily solvable. Since $r_1^2 + (-r_2^2) = a_1$ and $r_1^2(-r_2^2) = -(a_2^2/4)$, r_1^2 and $-r_2^2$ are roots of the quadratic equation $X^2 - a_1X - (a_2^2/4) = 0$, whence

$$r_1 = \sqrt{\frac{|a| + a_1}{2}}, \quad r_2 = \sqrt{\frac{|a| - a_1}{2}},$$

where, as usual, $|a| = \sqrt{a_1^2 + a_2^2}$. (Note that both integrands are positive.)

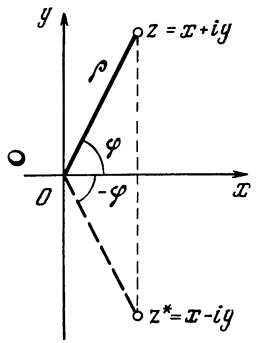


Figure 14.1.1

Finally we have

$$\sqrt{a} = \sqrt{a_1 + ia_2} = \pm\left(\sqrt{\frac{|a|+a_1}{2}} + i\sqrt{\frac{|a|-a_1}{2}}\right).$$

Thus, introducing the new, imaginary, number i in order to make solvable the problem of taking the square root of -1, we unexpectedly were able to take square roots of all numbers—"old", real, and "new", complex.

Before going further we will consider the *geometric representation* of complex numbers. The complex number $z = x + iy$ is usually considered as a point of the so-called ***complex plane*** (or the ***plane of the complex variables***) with (Cartesian) coordinates x and y (Figure 14.1.1). Here the x-axis is called the ***real axis*** (its points correspond to the real numbers $x + i0 = x$) and the y-axis, the ***imaginary axis***. To each complex number there corresponds a single point of the plane and vice versa; to conjugate complex numbers there correspond points which are symmetrical about the real axis. The absolute value $\sqrt{zz^*} = |z| = \sqrt{x^2 + y^2}$ of the complex number $z = x + iy$ is equal to the distance Oz of point z to the origin (corresponding to the number 0).

The straight line $b_1x + b_2y + c = 0$ of the plane of the complex variable $z = x + iy$ can be given by the "complex equation"

$$bz + b^*z^* + c = 0, \qquad (14.1.2)$$

where c is real, that is,

$$c^* = c, \quad \text{and} \quad b = \frac{1}{2}(b_1 - ib_2).$$

Indeed, if $z = x + iy$, then the left-hand side of (14.1.2), as is readily seen, equals $b_1x + b_2y + c$. The equation $(x^2 + y^2) + b_1x + b_2y + c = 0$ of the *perimeter* of the plane of complex variables can be rewritten in the form similar to (14.1.2):

$$zz^* + bz + b^*z^* + c = 0, \qquad (14.1.2a)$$

with $c^* = c$ and $b = (1/2)(b_1 - ib_2)$, since $zz^* = x^2 + y^2$ and $bz + b^*z^* = b_1x + b_2y$.

Equations (14.1.2) and (14.1.2a) can be combined as follows

$$azz^* + bz + b^*z^* + c = 0, \qquad (14.1.2b)$$

where a and c are real ($a^* = a$, $c^* = c$). Equation (14.1.2b) describes a circle if $a \neq 0$, and a straight line if $a = 0$.

The position of point z on a complex plane can be also characterized by its *polar coordinates*, the distance ρ of point z to the origin O (which is the absolute value $|z|$ of number z) and the angle φ between the ray Oz and the positive ray of the real axis (angle φ is called the ***argument*** (***amplitude*** or ***phase***) of number z; sometimes it is denoted Arg z). Here (see Figure 14.1.1)

$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi,$$

$$\rho = \sqrt{x^2 + y^2} = |z|,$$

$$\tan\varphi = \frac{y}{x}, \quad \cos\varphi = \frac{x}{|z|}, \quad \sin\varphi = \frac{y}{|z|}.$$

Thus

$$z = \rho(\cos\varphi + i\sin\varphi) \qquad (14.1.3)$$

(which is the *trigonometric form* of complex number).

The form (14.1.3) is especially convenient for multiplication and division of complex numbers

$$z_1 = \rho_1(\cos\varphi_1 + i\sin\varphi_1) \quad \text{and}$$
$$z_2 = \rho_2(\cos\varphi_2 + i\sin\varphi_2).$$

Indeed, it is clear that

$$z_1z_2 = \rho_1(\cos\varphi_1 + i\sin\varphi_1)\,\rho_2(\cos\varphi_2 + i\sin\varphi_2) = \rho_1\rho_2[(\cos\varphi_1 + i\sin\varphi_1) \times (\cos\varphi_2 + i\sin\varphi_2)]$$
$$= \rho_1\rho_2[(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + i(\cos\varphi_1\sin\varphi_2 + \sin\varphi_1\cos\varphi_2)]$$
$$= \rho_1\rho_2[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)], \qquad (14.1.4)$$

or in words: *when multiplying complex numbers their absolute values are multiplied and the arguments are added.* This leads to the following. If $z_1/z_2 = z_3 = \rho_3 (\cos \varphi_3 + i \sin \varphi_3)$, then $z_1 = z_2 z_3$, that is, $\rho_1 = \rho_2 \rho_3$, $\varphi_1 = \varphi_2 + \varphi_3$, whence $\rho_3 = \rho_1/\rho_2$, $\varphi_3 = \varphi_1 - \varphi_2$. Thus,

$$\frac{z_1}{z_2} = \frac{\rho_1}{\rho_2} [\cos (\varphi_1 - \varphi_2) + i \sin (\varphi_1 - \varphi_2)], \tag{14.1.4a}$$

when dividing complex numbers their absolute values are divided and the arguments are subtracted.

Note that the rules

$$|z_1 z_2| = |z_1| \, |z_2|, \quad \left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|} \tag{14.1.5}$$

completely coincide with those for real numbers; while the rules for the arguments of complex numbers:

$$\begin{aligned} \operatorname{Arg} (z_1 z_2) &= \operatorname{Arg} z_1 + \operatorname{Arg} z_2, \\ \operatorname{Arg} \left(\frac{z_1}{z_2} \right) &= \operatorname{Arg} z_1 - \operatorname{Arg} z_2, \end{aligned} \tag{14.1.5a}$$

are new to us. In form, Eqs. (14.1.5a) resemble the rules of operation on logarithms, but it is not yet quite clear what relates the *angle* $\operatorname{Arg} z = \varphi$ to logarithms (do complex numbers have really logarithms? We will answer this question later on).

Multiplying, according to Eq. (14.1.4), the number $z = \rho (\cos \varphi + i \sin \varphi)$ by itself n times we obtain the ***Moivre formula***

$$z^n = \rho^n (\cos n\varphi + i \sin n\varphi). \tag{14.1.6}$$

This formula is also valid for negative and fractional n. For example, noting that $1 = 1 (\cos 0 + i \sin 0)$ we get

$$\begin{aligned} z^{-1} &= \frac{1}{z} = \frac{1}{\rho} [\cos (-\varphi) + i \sin (-\varphi)] \\ &= \frac{1}{\rho} (\cos \varphi - i \sin \varphi) \end{aligned} \tag{14.1.6a}$$

and

$$\begin{aligned} z^{1/3} &= \sqrt[3]{r (\cos \alpha + i \sin \alpha)} \\ &= \sqrt[3]{r} \left(\cos \frac{\alpha}{3} + i \sin \frac{\alpha}{3} \right) \end{aligned} \tag{14.1.6b}$$

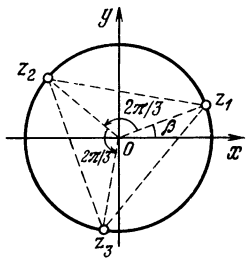


Figure 14.1.2

(see Exercise 14.1.2). Note that formula (14.1.6b) yields *three* values of the cube root of number z: since instead of α the argument z can be written as $\alpha + 2\pi$ or $\alpha + 4\pi$, we have

$$\begin{aligned} z_1 &= \sqrt[3]{r} (\cos \beta + i \sin \beta), \\ z_2 &= \sqrt[3]{r} \left[\cos \left(\beta + \frac{2\pi}{3} \right) \right. \\ &\left. + i \sin \left(\beta + \frac{2\pi}{3} \right) \right], \\ z_3 &= \sqrt[3]{r} \left[\cos \left(\beta + \frac{4\pi}{3} \right) \right. \\ &\left. + i \sin \left(\beta + \frac{4\pi}{3} \right) \right], \end{aligned}$$

with $\beta = \alpha/3$ (Figure 14.1.2).

Thus, the use of complex numbers makes it possible to find three roots z_1, z_2, and z_3 of the cubic equation $z^3 - c = 0$, where $c = r (\cos \alpha + i \sin \alpha)$ is an arbitrary complex number. Moreover, we can show that each *algebraic equation of nth degree*

$$\begin{aligned} P(z) &= a_0 z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots \\ &\ldots + a_{n-1} z + a_n = 0, \end{aligned}$$

where $a_0 \neq 0$, a_1, a_2, . . ., a_n are arbitrary complex numbers, *has* n (and only n), generally speaking, *complex roots* z_1, z_2, . . ., z_n which are not necessarily distinct (some time ago this important proposition was called the ***fundamental theorem of algebra***). Here the polynomial $P(z)$ is the product of n linear factors:

$$\begin{aligned} P(z) &= a_0 (z - z_1) (z - z_2) \ldots \\ &\ldots (z - z_n). \end{aligned}$$

For example, the quadratic equation

$$Q(z) = z^2 + pz + q = 0 \qquad (14.1.7)$$

has two roots

$$z_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}, \qquad (14.1.7a)$$

with

$$Q(z) = (z - z_1)(z - z_2).$$

The roots (14.1.7a) of Eq. (14.1.7) will be real for real p, q and distinct for $p^2/4 - q > 0$ (recall that the real number $x = x + i0$ is a particular case of a complex number); the roots will be the same (equal to $-p/2$ for $p^2/4 - q = 0$; and complex-conjugate (equal to $-p/2 \pm i\sqrt{q - p^2/4}$, the radicand being now positive) for $p^2/4 - q < 0$. (When we say that for real p, q and $p^2/4 < q$ the quadratic equation (14.1.7) *has no* solution, we are not mistaken and just want to say that manipulating only with real numbers we have to consider in this case that Eq. (14.1.7) has no roots.) Equation

$$z^5 - 3z^4 + 7z^3 - 13z^2 + 12z - 4 = (z - 1)^3(z^2 + 4) = 0 \qquad (14.1.8)$$

has five roots, of which only three are distinct:

$$z_1 = z_2 = z_3 = 1,\ z_4 = 2i, \text{ and } z_5 = -2i.$$

If all the coefficients of the equation (of any order) $a_0z^n + a_1z^{n-1} + \ldots + a_{n-1}z + a_n = 0$ are real, then to each "essentially complex" (i.e., not real) root $z = x + iy$ of the equation there corresponds a second root $z^* = x - iy$, which is *complex-conjugate* to the first root. This follows from the rules of operation on complex numbers, according to which

$$(z_1 + z_2)^* = z_1^* + z_2^*,\ (z_1 - z_2)^* = z_1^* - z_2^*,$$
$$(z_1 z_2)^* = z_1^* z_2^*,$$
$$\left(\frac{z_1}{z_2}\right)^* = \frac{z_1^*}{z_2^*} \qquad (14.1.9)$$

(check it). By virtue of (14.1.9), if, for instance, $z_0 = x_0 + iy_0$ is the root of the equation $Q(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 = 0$ with real coefficients, that is,

$$Q(z_0) = a_0z_0^4 + a_1z_0^3 + a_2z_0^2 + a_3z_0 + a_4 = 0,$$

then

$$a_0^*(z_0^*)^4 + a_1^*(z_0^*)^3 + a_2^*(z_0^*)^2 + a_3^*(z_0^*) + a_4^* = 0^*.$$

But $a_0^* = a_0$, $a_1^* = a_1$, ..., $a_4^* = a_4$ and naturally $0^* = 0$, therefore

$$Q(z_0^*) = a_0(z_0^*)^4 + a_1(z_0^*)^3 + a_2(z_0^*)^2 + a_3(z_0^*) + a_4 = 0,$$

that is, $z_0^* = x_0 - iy_0$ is also a root of the same equation (distinct from z_0 for $y_0 \neq 0$).

From the foregoing it follows that each *real polynomial* (i.e., such that all its coefficients are real) *can be expanded into (real) linear and quadratic factors* (e.g., see the expansion (14.1.8)). Indeed, the polynomial $P(z) = a_0z^n + a_1z^{n-1} + \ldots + a_{n-1}z + a_n$ has n (real or complex) roots $z_1, z_2, \ldots, z_n$, which nullifies it: $P(z_1) = P(z_2) = \ldots = P(z_n) = 0$. Here $P(z)$ can be expanded into a product

$$P(z) = a_0(z - z_1)(z - z_2)\ldots(z - z_n).$$

But if $z_0 = x_0 + iy_0$, where $y_1 \neq 0$ is a complex root of the polynomial $P(z)$ and $z_0^* = x_0 - iy_0$ is a root complex-conjugate to z_0, then

$$\begin{aligned}(z - z_0)(z - z_0^*) &= (z - x_0 - iy_0)(z - x_0 + iy_0)\\ &= [(z - x_0) - iy_0][(z - x_0) + iy_0]\\ &= (z - x_0)^2 - (iy_0)^2 = (z - x_0)^2 + y_0^2\\ &= z^2 - 2x_0z + (x_0^2 + y_0^2).\end{aligned}$$

Thus, to each *pair of complex-conjugate roots* z_0, z_0^* of the polynomial $P(z)$ there corresponds the *quadratic* factor $z^2 - 2x_0z + (x_0^2 + y_0^2)$ in its expansion, and to a *real root* $z_1 = x_1 + i0 = x_1$ (if only such a root exists) there corresponds a *linear* factor $z - z_1 = z - x_1$ of the expansion of $P(z)$. This reasoning proves the required assertion. For example, the inexperienced reader may consider the expansion

$$\begin{aligned}x^4 + a^4 &= x^4 + 2a^2x^2 + a^4 - 2a^2x^2\\ &= (x^2 + a^2)^2 - (\sqrt{2}\,ax)^2\\ &= [(x^2 + a^2) + \sqrt{2}\,ax][(x^2 + a^2) - \sqrt{2}\,ax]\\ &= (x^2 + \sqrt{2}\,ax + a^2)(x^2 - \sqrt{2}\,ax + a^2)\end{aligned}$$

to be accidental and artificial (how could we know that $2a^2x^2$ should have been added to and $2a^2x^2$ subtracted from the expression $x^4 + a^4$?). But the knowlegeable reader will understand that this expansion follows with necessity from the general fact formulated above of expandability of real polynomials and from the formula

$$x = \sqrt[4]{-a^4} = a\sqrt[4]{-1}$$

for the roots of the polynomial $x^4 + a^4$ (see Exercise 14.1.4).

The fact that in the region of complex numbers the square root can be taken of *any* (positive or negative, real or complex) number and that each quadratic equation has here two roots undoubtedly rejoices us since it proves that the set of complex numbers is, in a certain sense, "constructed more simply" than the set of real numbers. However, at least we could have expected this result; indeed, the imaginary unit i was especially introduced to make it possible for us to take the square root of -1 (or to solve the quadratic equation $x^2 + 1 = 0$). Of course, here we have obtained more than we had reason to expect (that is, the existence of not only one quadratic root $\sqrt{-1}$, previously nonexistent, but of all possible quadratic roots of *all* complex numbers; the solvability of not only the especially chosen equation $x^2 + 1 = 0$, but of *all* quadratic equations), but this just indicated how successful the introduction of $\sqrt{-1} = i$ was. However, the fact that application of complex numbers $z = x + iy$ permits taking *any* roots of all (both "old" and "new") numbers and solving algebraic equations of the third, fourth, fifth, and *any other* orders so that to make *all* equations solvable we have no need to introduce any new numbers and can manage with the complex numbers we used to solve quadratic equations, this fact is certainly a surprise. All this suggests that "there is something" in complex numbers, that they mean more than just the taking of roots of negative numbers, that complex numbers are not at all accidental and can be useful in many areas of mathematics and mathematized natural sciences.

These hopes have materialized.

Exercises

14.1.1. Let $z = x + iy$ and $f(z) = u + iv$. Find u and v as functions of x and y if the function $f(z)$ has the form: (a) $f = z^3 - 3z + 1$, (b) $f = z/(1 + z^2)$, (c) $f = z^n$.

14.1.2. Prove formulas (14.1.6a) and (14.1.6b) and the validity of de Moivre's formula (14.1.6) for any (real) n.

14.1.3. Prove that n values of the nth root of an arbitrary complex number c (n roots of an nth order equation $z^n - c = 0$) are represented geometrically by n points of the complex plane, which are the vertices of a regular polygon of n sides.

14.1.4. Write all the values of $\sqrt[4]{-1}$. Using the above formulas factorize the polynomial $z^4 + a^4$ (use real factors).

14.2 Raising a Number to an Imaginary Power and the Number e

We pose the problem of determining an *imaginary power*. How can we find the number 10^i? Can we multiply 10 into itself i times, that is, $\sqrt{-1}$ times? At first glance this question seems senseless. However, the question of the result of multiplication of 10 into itself -3 times or 1/2 time seems at first also senseless, since literal application of the rule of raising a number to an integral positive power ($a^n = \underbrace{aa \ldots a}_{n \text{ times}}$) is impossible here; yet the numbers 10^{-3} ($= 0.001$) and $10^{1/2}$ ($= \sqrt{10} \simeq 3.16$) undoubtedly exist.

Let us recall the logical sequence of determination of negative and fractional powers of numbers, the example being inspiring. First, we only define raising a number to an integral positive power n:

$$a^n = \underbrace{aaa \ldots a}_{n \text{ times}}.$$

From this definition—for integral positive powers as before—two rules follow:

$$a^{n+m} = \overbrace{aa \ldots a}^{n+m \text{ times}} = \overbrace{a \ldots a}^{n \text{ times}} \overbrace{a \ldots a}^{m \text{ times}} = a^n a^m, \tag{14.2.1}$$

$$(a^n)^m = \underbrace{\overbrace{a \ldots a}^{n \text{ times}} \overbrace{a \ldots a}^{n \text{ times}} \ldots \overbrace{a \ldots a}^{n \text{ times}}}_{m \text{ times}} = a^{nm} \tag{14.2.1a}$$

(cf. definitions of addition and multiplication of natural numbers on p. 458). Generalization consists in that we require the rules (14.2.1) and (14.2.1a)

be also fulfilled for negative and fractional powers.

Then the rule (14.2.1) yields

$$a^4 a^{-3} = a^{4+(-3)} = a^1 = a,$$

and therefore

$$a^{-3} = \frac{a}{a^4} = \frac{1}{a^3},$$

and similarly for any natural m

$$a^{-m} = \frac{1}{a^m}.$$

In a similar way, by virtue of (14.2.1a),

$$(a^{1/3})^3 = a^{\frac{1}{3}\cdot 3} = a^1 = a,$$

and therefore

$$a^{1/3} = \sqrt{a} \text{ and generally } a^{1/q} = \sqrt[q]{a}.$$

Now for any fraction $n = p/q$ we have

$$a^{p/q} = (a^{1/q})^p = (\sqrt[q]{a})^p = \sqrt[q]{a^p}. \qquad (14.2.2)$$

Since for any real (irrational) number we can find an arbitrarily close rational fraction, the problem of raising a number to *any* real power is completely solved. However, the techniques we used to find negative and fractional powers are themselves insufficient to determine the *imaginary* power of a^{ik} (a and k being real numbers); for this purpose we have to use the data we have gained in studying the derivative of the exponential function.

Let us use the fact that

$$e^r \simeq 1 + r \text{ for } |r| \ll 1 \qquad (14.2.3)$$

by the definition of the number e (cf. Section 4.8, formula (4.8.2)), the approximate equality (14.2.3) being the more exact the smaller $|r|$[14.3]. Let us agree now that Eq. (14.2.3) is also valid for complex r's which are small

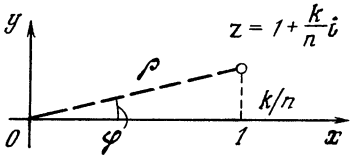


Figure 14.2.1

in absolute value, in particular also for pure imaginary numbers ri whose absolute value $|ri| = r$ is small. In order to determine the number e^{ki}, with k being any (not small) real number, we use the fact that, by virtue of (14.2.3),

$$e^{ki/n} \simeq 1 + \frac{k}{n} i \text{ for } n \gg 1, \qquad (14.2.3a)$$

where n, as follows from the notation $n \gg 1$, is understood to be a *very large* number (we will consider it to be *natural*) so that $k/n \ll 1$.

The point $z = 1 + ki/n$ ($= \rho(\cos\varphi + i \sin\varphi)$) of the complex plane has Cartesian coordinates $(1, k/n)$ and polar coordinates (ρ, φ) (the absolute value and argument of the complex number z), where

$$\rho = \sqrt{1 + \left(\frac{k}{n}\right)^2} = \left(1 + \frac{k^2}{n^2}\right)^{1/2}, \qquad (14.2.4)$$

$$\sin\varphi = \frac{k/n}{\rho}, \quad \tan\varphi = \frac{k}{n}$$

(Figure 14.2.1). For $n \gg 1$, that is when $k/n \ll 1$, the (exact) equalities (14.2.4) can be replaced by the (approximate) relations

$$\rho \simeq 1, \quad \varphi \simeq \frac{k}{n}, \qquad (14.2.5)$$

the approximate nature of (14.2.5) indicates that terms of the $(k/n)^2$th and higher orders of smallness are neglected (see the text below printed in small type). Since, by virtue of de Moivre's formula (14.1.6),

$$e^{ki} \simeq \left(1 + \frac{k}{n} i\right)^n = [\rho(\cos\varphi + i \sin\varphi)]^n$$
$$= \rho^n (\cos n\varphi + i \sin n\varphi),$$

[14.3] The difference $\sigma = e^r - (1 + r)$ is of the second order of smallness with respect to r: for small $|r|$ the ratio σ/r^2 remains "finite", that is, not too small and not too large. (Instead of referring to Eq. (4.8.2), we could have, which is the same, proceeded from the definition (4.8.3a) of an exponential function, $e^k = \lim_{n\to\infty} (1 + k/n)^n$, assuming that the latter formula is also valid for complex (pure imaginary, in particular) powers k.)

we have

$$\left(1+\frac{k}{n}i\right)^n \simeq 1^n\left[\cos\left(n\frac{k}{n}\right)+i\sin\left(n\frac{k}{n}\right)\right]=\cos k+i\sin k \qquad (14.2.3b)$$

and consequently

$$e^{ki}=\cos k+i\sin k. \qquad (14.2.6)$$

This relation can be assumed to be the *definition* of the exponential function of an imaginary argument; it follows from the supposition (14.2.3a). Formula (14.2.6) is known as the ***Euler formula.***

The Euler formula (14.2.6) helps find the number 10^i

$$10^i=(e^{\ln 10})^i=e^{\ln 10\cdot i}\simeq e^{2.3i}$$
$$=\cos 2.3+i\sin 2.3\simeq -0.67+0.77i.$$

Another method for deriving the Euler formula is connected with the expansion of functions into series. As we know (see Chapter 6)

$$e^x\simeq 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\ldots, \qquad (14.2.7)$$

this series being convergent for *all* x's, which gives us courage to consider the number x in the series to be complex as well. Suppose, by definition,

$$e^{ix}=1+\frac{ix}{1!}+\frac{(ix)^2}{2!}+\frac{(ix)^3}{3!}+\frac{(ix)^4}{4!}+\frac{(ix)^5}{5!}+\ldots$$
$$=\left[1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots\right]+i\left[\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-\ldots\right]. \qquad (14.2.8)$$

But since (see Chapter 6)

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots, \qquad (14.2.9)$$

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\ldots \qquad (14.2.9a)$$

(the series (14.2.9) and (14.2.9a) are also convergent at *all* x's), we again obtain the Euler formula

$$e^{ix}=\cos x+i\sin x.$$

Further we can naturally assume that if $z=x+iy$, then

$$e^z=e^{x+iy}=e^xe^{iy}=e^x(\cos y+i\sin y) \qquad (14.2.10)$$

(cf. Exercise (14.2.1)).

Formula (14.2.10) can, of course, be immediately taken as the *definition* of the exponential function e^z of the complex argument $z=x+iy$ (compare with the above reasoning for the Euler formula (14.2.6)). This definition is justified by that if $z=x+i0=x$ then e^z becomes the exponential function e^x of the *real* variable x, and also by the fact that the function (14.2.10) retains the basic property (14.2.1) of exponential function:

if $z=z_1+z_2$, where $z_1=x_1+iy_1$ and $z_2=x_2+iy_2$, then

$$e^z=\exp[(x_1+x_2)+i(y_1+y_2)]$$
$$=e^{(x_1+x_2)}[\cos(y_1+y_2)+i\sin(y_1+y_2)]$$
$$=(e^{x_1}e^{x_2})[(\cos y_1+i\sin y_1)(\cos y_2+i\sin y_2)]=[e^{x_1}(\cos y_1+i\sin y_1)]\times[e^{x_2}(\cos y_2+i\sin y_2)]=e^{z_1}e^{z_2}.$$

By virtue of formulas (14.2.4) and (6.4.6),

$$\rho=\left(1+\frac{k^2}{n^2}\right)^{1/2}=1+\frac{1}{2}\frac{k^2}{n^2}+\ldots$$
$$=1+\frac{k^2}{2n^2}+\ldots,$$

where the dots indicate terms of order higher than the term $k^2/2n^2$; in accordance with the (more general) formula (6.4.3), we have

$$\rho^n=\left(1+\frac{k^2}{2n^2}+\ldots\right)^n$$
$$=1+n\frac{k^2}{2n^2}+\ldots=1+\frac{k^2}{2n^2}+\ldots,$$

whence, owing to the smallness of the second term $k^2/2n$ for $n\ll 1$, it follows that the absolute value of the number $(1+ki/n)^n$ ($\simeq e^{ki}$) can be assumed to (approximately) equal 1. Similarly, we can also verify that it is not a big mistake if we assume that the argument of the number $(1+ki/n)$ equals k/n (see Exercise (14.2.2)).

Exercises

14.2.1. Prove that the relation $e^{x+iy} \simeq \left(1+\frac{x+iy}{n}\right)^n$ for $n \gg 1$ (that is, $e^{x+iy} = \lim_{n\to\infty}\left(1+\frac{x+iy}{n}\right)^n$) means the same as formula (14.2.10).

14.2.2. Prove that in the formula $1 + ki/n = \rho(\cos\varphi + i\sin\varphi)$ for $n \gg 1$ we have $\varphi = k/n + 1 \ldots$, where the points indicate terms of the $1/n^2$ and higher orders of smallness.

14.2.3. Prove that for any complex (and not only for pure imaginary) number Δz small in absolute value, we have $e^{\Delta z} = 1 + \Delta z + \ldots$, the points indicating the terms whose absolute values are of order $|\Delta z|^2$ and higher.

14.3 Trigonometric Functions and the Logarithm

The Euler formula (14.2.6) reveals a deep inner relation of periodic functions $\cos x$ and $\sin x$ to the exponential function. At first glance the exponential function a^x has nothing in common with periodicity: for $a > 1$ the value of a^x monotonically increases with increasing x, while for $a < 1$ it always decreases; there is no periodicity at all. But if $a = -1$, then the powers of the number a assume the following values:

$$-1,\ +1,\ -1,\ +1,\ -1,\ +1,\ \ldots, \qquad (14.3.1)$$

that is, they change periodically [14.4]. If we write

$$-1 = \cos\pi + i\sin\pi = e^{i\pi},$$

then we can consider the power n in the notation $(-1)^n$ to be any *real* number:

$$(-1)^n = (e^{i\pi})^n = e^{in\pi} = \cos(n\pi) + i\sin(n\pi).$$

Here (generally speaking, complex) quantity $(-1)^n$ assumes *various* values whose modulus is 1 and changes periodically with n. The value of a^x changes in the same manner, where $a = \cos\varphi + i\sin\varphi = e^{i\varphi}$ is an arbitrary complex number with unit modulus: in this case

$$a^x = (e^{i\varphi})^x = e^{i\varphi x} = \cos(\varphi x) + i\sin(\varphi x).$$

Now, if $a = \rho(\cos\varphi + i\sin\varphi) = \rho e^{i\varphi}$ is a complex number whose modulus ρ is not unity ($\rho > 0$ and $\rho \neq 1$), then

$$a^x = (\rho e^{i\varphi})^x = \rho^x e^{i\varphi x} = \rho^x[\cos(\varphi x) + i\sin(\varphi x)] \qquad (14.3.2)$$

can be expanded into a product of two cofactors: one of them (which determines the absolute value of the number a^x) will monotonically increase or monotonically decrease with increasing x, while the other equal to $e^{i\varphi x}$ (it is responsible for the argument of the number a^x) will change periodically. If $\rho < 1$, then the real part and the coefficient of the imaginary part of a^x, that is, the expressions $\rho^x\cos(\varphi x)$ and $\rho^x\sin(\varphi x)$, will simulate *damped oscillations* (see Section 10.4 and, in particular, Figure 10.4.2). Later (see Section 17.1) we will see that the relation between the function a^x and damped oscillations is by no means accidental.

The Euler formula (14.2.6) enables constructing many other important functions of a real, imaginary, or complex variable. Let us begin with rewriting this formula and expressing trigonometric functions via an exponential function, rather than an exponential function via trigonometric ones. From

$$e^{i\varphi} = \cos\varphi + i\sin\varphi, \qquad (14.3.3)$$

it also follows that

$$e^{-i\varphi} = \cos(-\varphi) + i\sin(-\varphi) = \cos\varphi - i\sin\varphi. \qquad (14.3.3a)$$

Adding (14.3.3) and (14.3.3a) term-by-term and then subtracting (14.3.3a) from (14.3.3) we readily obtain

$$\cos\varphi = \frac{e^{i\varphi}+e^{-i\varphi}}{2}, \quad \sin\varphi = \frac{e^{i\varphi}-e^{-i\varphi}}{2i}. \qquad (14.3.4)$$

[14.4] We have just limited ourselves to *natural* (integral positive) values of the power. If we let n also assume *integral negative* values and vanish, we will infinitely continue to the left the periodic sequence (14.3.1) of numbers $(-1)^n$.

The corollaries (14.3.4) of (14.3.3) are also often called *Euler formulas.*

Note that for real φ the terms on the right-hand sides of (14.3.4) are certainly real (they are the cosine and the sine of an angle): indeed, substituting the expansion (14.2.8) and a similar one for $e^{i\varphi}$ and $e^{-i\varphi}$ in these expressions we again obtain formulas (14.2.9) and (14.2.9a). There is another approach to readily convince ourselves that the expressions obtained are real. Let us replace i by $-i$ in the expression $z = P + iQ$ for a complex number z (P and Q being real); we obtain a complex-conjugate number $z^* = P - iQ$. If z did not change here, that is, $z^* = z$, then $Q = 0$ and z is real. But we can see that (for real φ) the right-hand sides of (14.3.4) do not change on substituting $-i$ for i, and consequently they are real.

Further, using expressions (14.3.4) for $\cos\varphi$ and $\sin\varphi$ we can also determine trigonometric functions of the imaginary argument via exponents. We replace in (14.3.4) $\varphi = i\psi$, assuming ψ real, and so φ is *pure imaginary.* We get

$$\cos(i\psi) = \frac{e^{-\psi}+e^{\psi}}{2}, \quad \sin(i\psi) = \frac{e^{-\psi}-e^{\psi}}{2i} = i\,\frac{e^{\psi}-e^{-\psi}}{2}. \tag{14.3.5}$$

Thus we are back again to the exponential functions of the real argument ψ and more accurately to certain combinations of exponents we are going to discuss in the next section of this chapter.

In discussing the functions $\cos(i\psi)$ and $\sin(i\psi)$ we can solve still another problem. For example, assume $\psi = 1$. Simple calculations yield

$$\cos i = \frac{e^1+e^{-1}}{2} \simeq \frac{2.72+0.36}{2} = 1.54 \tag{14.3.6}$$

$$\left(\text{and } \sin i \simeq i\,\frac{2.72-0.36}{2} = 1.18i\right).$$

We pose the question: What is the angle whose cosine equals 1.5? Before we could have answered that there is no such an angle since the cosine of any angle does not exceed unity. Here it is tacitly implied that the angle is a "real" angle, that which can be drawn or measured with an instrument. But now that we have introduced imaginary angles we can obtain a definite numerical solution: the equation

$$\cos\varphi = 1.5 \tag{14.3.7}$$

has the (approximate) solution $\varphi \simeq i$. Obviously, the value $\varphi \simeq -i$ is also a solution of Eq. (14.3.7); moreover, there are infinitely many solutions which differ by multiples of 2π. Thus, before we had no solutions of Eq. (14.3.7) but now we can consider them to be *infinitely many*:

$$\varphi = \pm i + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots. \tag{14.3.8}$$

The situation here is quite similar to that we encountered in the case of algebraic equations. When we used only real numbers, the quadratic equation $x^2 - n = 0$, with $n > 0$, had two solutions (roots), $x = \pm\sqrt{n}$, while the equation $x^2 + n = 0$ had no solution. But when we used complex numbers, the second equation also had two solutions, $x = \pm i\sqrt{n}$. Similarly, the equation $\cos\varphi = n$, for $n \leqslant 1$, had infinitely many solutions (for example, to the value $n = 1/2$ there correspond angles $\varphi = \pm\pi/3 + 2k\pi$, k being any integral number), while for $n > 1$ the equation had no solution; but when φ is complex, the equation has infinitely many solutions.

The problem of solving equations can be formulated as that of determining an ***inverse function***. Let us consider a ("direct") function $t = x^2$, the determination of its inverse reduces to solving the equation $x^2 = t$, where t is given and x must be found. Here for real variables the original function exists for all values of the argument x, while the inverse function only exists for the argument $t \geqslant 0$; in the complex plane the inverse function is also defined for *all* t. Similarly, the function

$\varphi = \arccos s$, the inverse of $s = \cos \varphi$, also exists for real variables only in the interval $-1 \leqslant s \leqslant 1$ (cf. Section 4.11); while for complex numbers the function is defined for *all* s.

The situation proved to be the same with the ***logarithmic*** function $z = \ln w$, which is defined as the inverse of the exponential function $w = e^z$. We know that real numbers do not always have logarithms, for example, negative numbers have no logarithms. Logarithms of complex numbers, as we will see later, are always defined. Note, finally, that all the three inverse functions, $x = \sqrt{t}$, $\varphi = \arccos s$, and $z = \ln w$, are not single-valued, the nature of this being different; the reasons for this difference will be discussed later.

We rewrite formula (14.2.10) thus

$$e^{x+iy} = \rho (\cos y + i \sin y) = \rho e^{iy}, \quad (14.3.9)$$

where $\rho = e^x$, that is, $x = \ln \rho$. Since the equation $w = e^x$ can be written as $z = \ln w$, from (14.3.9) it follows that if $w = \rho (\cos \varphi + i \sin \varphi)$, then

$$z = \ln w = \ln \rho + i\varphi$$
$$= \ln |w| + i \operatorname{Arg} w. \quad (14.3.10)$$

This relation can, of course, be considered as the *definition* (unknown to us before) of the number $\ln w$: we "derived" (14.3.10) from (14.3.9) on the assertion that $e^z = w$ means the same as $z = \ln w$, which is only valid for real z and w and not for complex (or even for negative real) w for which $\ln w$ had no meaning before. The expediency of the definition (14.3.10) is connected with that for $\varphi = \operatorname{Arg} z = 0$, that is, when z is a *real positive number*, formula (14.3.10) leads to the usual notion of logarithm (when introducing logarithms of complex numbers we have just to refresh our knowledge and not to start from the very beginning). Besides, by (14.3.9) here we always (for any w) have $e^{\ln w} = w$.

Expressing the numbers ρ (that is, $|w|$) and φ (that is, $\operatorname{Arg} w$) via the real part and the coefficient of the imaginary part of the complex number $w = u + iv$: $\rho = |w| = \sqrt{u^2 + v^2}$, $\varphi = \operatorname{Arg} w = \arctan (v/u)$, we write (14.3.10) thus

$$\ln (u + iv) = \ln \sqrt{u^2 + v^2} + i \arctan (v/u). \quad (14.3.11)$$

We have finally arrived at a new notion of the logarithm of a number, a notion wider than we learned at school. Formula (14.3.10) (or (14.3.11)) enables us to find, for example, the logarithm of a *negative* number. Indeed, we can write, say,

$$-5 = 5 \times (-1) = 5 (\cos \pi + i \sin \pi)$$
$$= 5e^{i\pi},$$

whence it follows that

$$\ln (-5) = \ln 5 + i\pi$$
$$\simeq 1.6 + 3.14i. \quad (14.3.12)$$

Since (cf. Section 4.9) $\log_{10} x = \ln x / \ln 10 - \ln x / 2.3$, we have

$$\log_{10} (-5) \simeq (1/2.3) \ln (-5)$$
$$\simeq 0.7 + 1.37\, i.$$

If the modulus ρ of the complex number z is unity, that is, if $z = \cos \varphi + i \sin \varphi$, then $\ln z = \varphi i = i \operatorname{Arg} z$. This relation of logarithms of complex numbers and their arguments is responsible for that the properties of the arguments of complex numbers are close to those of logarithms: for any two (complex) numbers u and v we have

$$\log (uv) = \log u + \log v,$$

$$\log \left(\frac{u}{v}\right) = \log u - \log v,$$

$$\operatorname{Arg} (uv) = \operatorname{Arg} u + \operatorname{Arg} v,$$

$$\operatorname{Arg} \left(\frac{u}{v}\right) = \operatorname{Arg} u - \operatorname{Arg} v,$$

that is, when multiplying complex numbers their arguments (and logarithms) are added, and when dividing they are subtracted. (Of course, here the symbol $\log u$ can also mean the logarithm to *any* base, since all logarithms are in proportion.)

Note that the angle φ, which is the argument of a complex number and enters into the formulas, can be replaced by the angle $\varphi + 2k\pi$, where k is any integer. Therefore, the logarithm of any number is not single-valued, but *infinitely many-valued*. For example, the complete solution of (14.3.12) is

$$\ln(-5) = \ln 5 + i(\pi + 2k\pi),$$
$$k = 0, \pm 1, \pm 2, \ldots.$$

The same refers to the logarithm of any positive number. For example,

$$\ln 5 = 1.6 + i \times 2k\pi,$$
$$k = 0, \pm 1, \pm 2, \ldots,$$

or

$$\ln 1 = i \times 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots,$$

since

$$1 = \cos 0 + i \sin 0 = \cos 2\pi + i \sin 2\pi = \ldots = e^0 = e^{2\pi i} = e^{4\pi i} = \ldots.$$

Later we will return to the question as to why the logarithm is not single-valued, why every number has not one, but infinitely many logarithms.

Note, finally, that logarithms enable us to raise to any (complex) power of any (complex) number. For example, let us find the number i^i (what will we obtain on multiplying the number i into itself i times?). Obviously,

$$i = \cos\frac{\pi}{2} + i \sin\frac{\pi}{2} = e^{\pi i/2}$$

and, consequently,

$$i^i = (e^{\pi i/2})^i = e^{(\pi i/2)\cdot i} = e^{-\pi/2} \simeq 2.7^{-1.57} = \frac{1}{2.7^{1.57}} \simeq 0.21$$

(the number is even found to be real).

Exercises

14.3.1. Let $u = \rho(\cos\varphi + i \sin\varphi)$ and $z = x + iy$. Write all the values of u^z.

14.3.2. (the jocular problem). What is larger: e^e or i^{i^i}? e^i or i^e?

14.4* Trigonometric Functions of a Purely Imaginary Independent Variable. Hyperbolic Functions

Let us return to formulas (14.3.5) for trigonometric functions of the *imaginary argument*:

$$\cos(i\psi) = \frac{e^{\psi} + e^{-\psi}}{2}, \qquad \sin(i\varphi) = i\frac{e^{\psi} - e^{-\psi}}{2}. \tag{14.4.1}$$

The combinations of exponents e^{ψ} and $e^{-\psi}$ contained in the formulas have special names and notation. The former is called the ***hyperbolic cosine*** and the latter the ***hyperbolic sine*** of the argument ψ, and denoted by

$$\cosh\psi = \frac{e^{\psi} + e^{-\psi}}{2}, \quad \sinh\psi = \frac{e^{\psi} - e^{-\psi}}{2}. \tag{14.4.2}$$

(The letter h reminds us of hyperbola.) Thus,

$$\cos(i\psi) = \cosh\psi, \qquad \sin(i\psi) = i\sinh\psi. \tag{14.4.3}$$

Some properties of hyperbolic functions copy the properties of (usual or circular) trigonometric functions, reproducing them sometimes in a somewhat distorted form. For example, from formulas (14.4.2) it follows that the function $\cosh\psi$ is even: $\cosh(-\psi) = (e^{-\psi} + e^{\psi})/2 = \cosh\psi$, while the function $\sinh\psi$ is odd: $\sinh(-\psi) =$

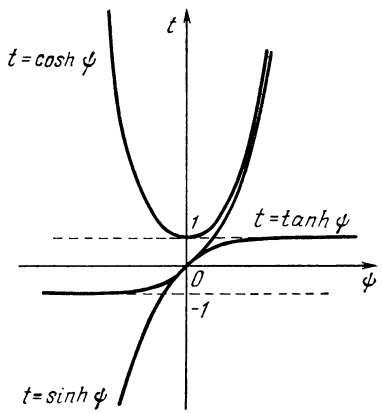


Figure 14.4.1

$(e^{-\psi} - e^{\psi})/2 = -\sinh\psi$. On the other hand it is obvious that $\cosh(0) = 1$ and $\sinh(0) = 0$ (compare the curves of the hyperbolic sine and cosine represented in Figure 14.4.1). Figure 14.4.1 also presents the curves of the ***hyperbolic tangent***:

$$\tanh\psi = \frac{\sinh\psi}{\cosh\psi} = \frac{e^{\psi}-e^{-\psi}}{e^{\psi}+e^{-\psi}};$$

clearly, the function $\tanh\psi$ is odd ($\tanh(-\psi) = -\tanh\psi$) and $\tanh(0) = 0$.[14.5] Note also that to the well-known formula $\cos^2\varphi + \sin^2\varphi = 1$ there corresponds the following formula of "hyperbolic trigonometry":

$$\cosh^2\psi - \sinh^2\psi = 1. \qquad (14.4.4)$$

Indeed,

$$\cosh^2\psi - \sinh^2\psi = \left(\frac{e^{\psi}+e^{-\psi}}{2}\right)^2 - \left(\frac{e^{\psi}-e^{-\psi}}{2}\right)^2 = \frac{e^{2\psi}+2+e^{-2\psi}}{4} - \frac{e^{2\psi}-2+e^{-2\psi}}{4} = 1.$$

The expansions into series

$$e^{\psi} = 1 + \frac{\psi}{1} + \frac{\psi^2}{2!} + \frac{\psi^3}{3!} + \frac{\psi^4}{4!} + \ldots,$$

$$e^{-\psi} = 1 - \frac{\psi}{1} + \frac{\psi^2}{2!} - \frac{\psi^3}{3!} + \frac{\psi^4}{4!} + \ldots$$

and formulas (14.4.2) yield

$$\cosh\psi = 1 + \frac{\psi^2}{2!} + \frac{\psi^4}{4!} + \frac{\psi^6}{6!} + \ldots, \qquad (14.4.5)$$

$$\sinh\psi = \psi + \frac{\psi^3}{3!} + \frac{\psi^5}{5!} + \frac{\psi^7}{7!} + \ldots$$

(compare with the expansion (14.2.9) and (14.2.9a) into a series of trigonometric functions), whence it follows, in particular, that

$$\sinh\psi \simeq \psi, \quad \cosh\psi \simeq 1 + \frac{\psi^2}{2}, \quad \tanh\psi \simeq \psi \qquad (14.4.6)$$

for small ψ (we have here neglected terms of the third and higher orders of smallness with respect to ψ). Further, it is obvious that

$$(\sinh\psi)' = \left(\frac{e^{\psi}-e^{-\psi}}{2}\right)' = \frac{e^{\psi}+e^{-\psi}}{2} = \cosh\psi,$$

$$(\cosh\psi)' = \left(\frac{e^{\psi}+e^{-\psi}}{2}\right)' = \frac{e^{\psi}-e^{-\psi}}{2} = \sinh\psi, \qquad (14.4.7)$$

whence, with account taken of (14.4.4), we have

$$(\tanh\psi)' = \left(\frac{\sinh\psi}{\cosh\psi}\right)' = \frac{(\sinh\psi)'\cosh\psi - \sinh\psi(\cosh\psi)'}{\cosh^2\psi} = \frac{\cosh^2\psi - \sinh^2\psi}{\cosh^2\psi} = \frac{1}{\cosh^2\psi}$$

(compare with the formulas for derivatives of trigonometric functions in Section 4.10).

Of no less surprise is the resemblance of the well-known ("school") formulas of addition for trigonometric functions:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta,$$

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta,$$

$$\tan(\alpha+\beta) = \frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}$$

and similar formulas for hyperbolic functions

$$\sinh(u+v) = \sinh u\cosh v + \cosh u\sinh v,$$

$$\cosh(u+v) = \cosh u\cosh v + \sinh u\sinh v,$$

$$\tanh(u+v) = \frac{\sinh(u+v)}{\cosh(u+v)} = \frac{\sinh u\cosh v + \cosh u\sinh v}{\cosh u\cosh v + \sinh u\sinh v} \qquad (14.4.8)$$

[14.5] We know that the (ordinary) cosine and sine (of a real variable) are bounded: $|\cos\varphi| \leqslant 1$, $|\sin\varphi| \leqslant 1$, while the tangent can assume arbitrarily large values. But the hyperbolic cosine and sine are not bounded since as is readily seen $\cosh\psi \to \infty$ as $\psi \to \infty$, $\sinh\psi \to \infty$ as $\psi \to \infty$; while $\tanh\psi$ is bounded: $|\tanh\psi| < 1$ (and obviously $\lim_{\psi\to\infty} \tanh\psi = 1$).

$$= \frac{\sinh u/\cosh u + \sinh v/\cosh v}{1 + (\sinh u \sinh v)/(\cosh u \cosh v)}$$

$$= \frac{\tanh u + \tanh v}{1 + \tanh u \tanh v}$$

(the last of these formulas was obtained by dividing the numerator and denominator of the fraction $\sinh(u+v)/\cosh(u+v)$ by $\cosh u \cosh v$; as to the first and the second formula see Exercise 14.4.1). From Eq. (14.4.8) it follows that

$$\sinh 2u = 2 \sinh u \cosh u,$$

$$\cosh 2u = \cosh^2 u + \sinh^2 u, \qquad (14.4.9)$$

$$\tanh 2u = \frac{2 \tanh u}{1 + \tanh^2 u}$$

(we have obtained these formulas by setting $u = v$ in (14.4.8); all formulas (14.4.4)-(14.4.9) resemble very much those you already know, don't they?

The names of *hyperbolic* functions $\cosh \psi$ and $\sinh \psi$ are connected with the following. Ordinary (circular) cosine and sine can be defined as the abscissa and the ordinate of a point M of a unit *circle* $x^2 + y^2 = 1$ (Figure 14.4.2*a*); the argument here is the angle $\angle AOM = \varphi$, which can be considered as a doubled area of sector S_{AOM} of the circle (our circle has the radius 1). Similarly, the hyperbolic cosine and sine can be described as being the abscissa and the ordinate of a variable point M of a unit hyperbola $x^2 - y^2 = 1$ (Figure 14.4.2*b*), where the argument is the *hyperbolic* angle ψ, which is the doubled area of sector AOM of the hyperbola (see Exercise 14.4.5). Figure 14.4.2 also presents geometrically the circular and hyperbolic tangents:

$$\tan \varphi = \frac{\sin \varphi}{\cos \varphi} = \frac{MP}{OP} = \frac{AT}{OA} = AT$$

(see Figure 14.4.2*a*)

$$\tanh \psi = \frac{\sinh \psi}{\cosh \psi} = \frac{MP}{OP} = \frac{AT}{OA} = AT$$

(see Figure 14.4.2*b*)

(note that in both cases $OA = 1$); the segment AT is called the *tangent line*, and the segments MP and OP the *sine* and *cosine lines*. Many properties of hyperbolic functions follow from the foregoing: formula (14.4.4) (the relations $\cos^2 \varphi + \sin^2 \varphi = 1$ and $\cosh^2 \psi - \sinh^2 \psi = 1$ are no other than the equations $x^2 + y^2 = 1$ and $x^2 - y^2 = 1$ of the circle and the hyperbola); the even nature of the function $\cosh \psi$ and the odd nature of the functions $\sinh \psi$ and $\tanh \psi$ (from Figure 14.4.2*b* it is readily seen that $\cosh(-\psi) = \cosh \psi$, $\sinh(-\psi) = -\sinh \psi$, and $\tanh(-\psi) = -\tanh \psi$); the equalities $\cosh(0) = 1$, $\sinh(0) = \tanh(0) = 0$; and some others.[14.6]

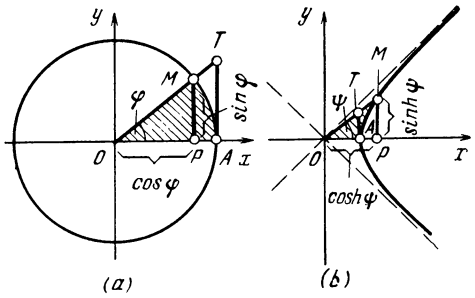


Figure 14.4.2

We can now look from a new angle at the relation (14.4.3) of circular and hyperbolic functions. Previously, the sine and the cosine of angle φ were defined as the segments in a circle of radius 1 (see Figure 14.4.2*a*) or as the ratios of the sides of a right triangle with acute angle φ. But these definitions tell us nothing of, say, what $\sin i$ or $\cos i$ is. What is the strange angle of magnitude i? How can we turn through such an angle? How can we construct a right triangle with such an angle? The last questions, of course, remain without any answer, but the values of $\sin i$ and $\cos i$ can quite possibly be found with the help of Euler's formulas. Note that the series (14.2.9) for the cosine, $\cos \varphi = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \ldots$, contains only *even* powers of angle φ; therefore, on substituting $\varphi = i$ (or, generally, $\varphi = \psi i$, where ψ is a real number) all the terms of the series remain real: the cosine of a pure imaginary angle is a real number. The series (14.2.9a) for the sine, $\sin \varphi = \varphi - \frac{\varphi^3}{3!} + \ldots$, on substituting $\varphi = i\psi$

[14.6] Note that formulas (14.4.8) for addition and the definitions (14.4.2) for hyperbolic functions can also be derived using the geometric representation of hyperbolic functions presented in Figure 14.4.2*b* (see, for example, V. G. Shervatov, *Hyperbolic Functions*, Heath, Boston, 1963).

yields a pure imaginary expression and whence the coefficient i appears in formula (14.4.3) relating sin $(i\psi)$ and the real function sinh ψ.

The analogy between trigonometric and hyperbolic functions can also be seen when we discuss an important topic on differential equations. We know that the differential equation

$$\frac{d^2x}{dt^2} = -kx \qquad (14.4.10)$$

for $k > 0$ admits of a general solution

$$x = A \cos \omega t + B \sin \omega t,$$
$$\omega = \sqrt{k} \qquad (14.4.11)$$

(see Section 10.2). On the other hand, the equation

$$\frac{d^2x}{dt^2} = kx, \qquad (14.4.10a)$$

where k is also greater than zero, as we have already seen (see Section 13.10), has particular solutions $x = e^{\sqrt{k}t}$ and $x = e^{-\sqrt{k}t}$, so that its general solution can be written in the form

$$x = Ce^{\sqrt{k}t} + De^{-\sqrt{k}t},$$

where the constants C and D are arbitrary. If we write this solution as

$$x = (C+D)\frac{e^{\sqrt{k}t}+e^{-\sqrt{k}t}}{2} + (C-D)\frac{e^{\sqrt{k}t}-e^{-\sqrt{k}t}}{2}$$

and denote $C + D = A$, $C - D = B$, we, by virtue of definitions (14.4.2), obtain a general solution of (14.4.10a), which is similar to the general solution (14.4.11) of (14.4.10):

$$x = A \cosh \omega t + B \sinh \omega t, \text{ where } \omega = \sqrt{k}. \qquad (14.4.11a)$$

The relation of trigonometric and exponential (or hyperbolic) functions sheds new light on the initiation, in similar conditions, of periodic motions (oscillations) and aperiodic motions. Let us consider a ball lying on a circular top of a hill (Figure 14.4.3*a*). On the very top the force of gravity is completely balanced by the reaction

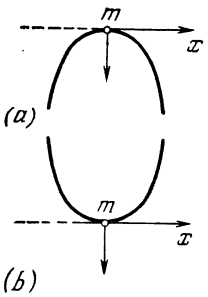


Figure 14.4.3

of the base of the hill, that is, the component of the force of gravity, which is tangent to the surface of the hill and tends to move the ball, is zero. However, once the ball moves even slightly, say, to the right, a force F begins to act on it in the direction which increases the motion, the force growing in proportion to the increasing displacement x (we can consider the force to be proportional to the displacement: $F = kx$ for $k > 0$). By Newton's second law (see Section 9.4) the equation of motion of the ball is

$$m\frac{d^2x}{dt^2} = F = kx. \qquad (14.4.12)$$

The solution to this equation (cf. Eq. (13.4.10a)) has the form

$$x = ae^{\sqrt{\frac{k}{m}}t} + be^{-\sqrt{\frac{k}{m}}t}, \qquad (14.4.13)$$

or, which is the same,
$x = A \cosh \omega t + B \sinh \omega t$, where

$$\omega = \sqrt{\frac{k}{m}}. \qquad (14.4.14)$$

The existence of a solution that grows exponentially[14.7] $y_1 = e^{\sqrt{k/m}t}$ indicates that the position of the ball on the top of the hill is *unstable*. We can expect the initial conditions to be generally such that in (14.4.13) $a \neq 0$

[14.7] Or the existence of the solutions $Y_1 = \cosh \omega t$, $Y_2 = \sinh \omega t$ (growing exponentially as well). It is readily understood that as ψ grows the functions $u = \cosh \psi$ and $v = \sinh \psi$ grow by the exponential law:

$$\lim_{\psi\to\infty} \frac{\cosh \psi}{e^{\psi}} = \lim_{\psi\to\infty} \frac{\sinh \psi}{e^{\psi}} = \frac{1}{2}.$$

and $b \neq 0$ (it is quite unlikely that the initial conditions ensure the strict observation of the equality $a = 0$). Consequently, in a certain period of time the first term on the right-hand side of (14.4.13) will necessarily become rather large and the ball will go down.

Now we consider another (quite opposite in a certain sense) situation. Let the ball lie at the bottom of a well which has oval walls (Figure 14.4.3b). When the ball is displaced a distance x from the equilibrium position, a force F appears which is directed toward the bottom of the well, the "returning" force. (Such direction of the force indicates that the equilibrium is *stable* in the case of the well as distinct from the equilibrium of the ball on the hill.) Suppose that the force F is proportional to the displacement x in magnitude, so that $F = -kx$, where k is positive and the minus sign in the formula indicates that F tends to return the ball to its initial position. Then the equation of motion of the ball will be

$$m\frac{d^2x}{dt^2} = -kx. \qquad (14.4.12a)$$

Formally, this equation looks like the previous one (Eq. (14.4.12)). Therefore, we can use the solution of Eq. (14.4.12) by substituting $-k$ for k, since this substitution transforms Eq. (14.4.12) into Eq. (14.4.12a). Here we obtain the solution to the new problem

$$x = ae^{\sqrt{-\frac{k}{m}}\,t} + be^{-\sqrt{-\frac{k}{m}}\,t} = ae^{i\omega t} + be^{-i\omega t}, \quad \omega = \sqrt{\frac{k}{m}}, \qquad (14.4.13a)$$

or, with the change of variables,

$$y = A\cos\omega t + B\sin\omega t, \quad \omega = \sqrt{\frac{k}{m}} \qquad (14.4.14a)$$

(cf. Eq. (14.4.11); expressions (14.4.13a) and (14.4.14a) coincide if we set $A = (a + b)/2$, $B = (a - b)\,i/2$). Here the (real) solution (14.4.14a) of Eq. (14.4.12a) is bounded for all t, which is closely connected with the steady state of the ball in the well.

Imaginary numbers (such as the square root of a negative number) were first introduced (although in a rather formal manner) by the prominent Italian Renaissance mathematician Girolomo ***Cardano*** (1501-1576). It was by no means accidental that complex numbers appeared when algebra was making such strides or that Cardano was the man who introduced them. (In ancient times, geometry was the center of attention; in the Renaissance, it was algebra that paved the way for the rise of mathematical analysis in the 17th century.) Cardano won himself a place of honor in the history of mathematics chiefly for his pioneering publication of the formula for solving an arbitrary[14.8] cubic equation[14.9] if

$$x^3 + px + q = 0, \qquad (14.4.15)$$

then

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \qquad (14.4.16)$$

But the seemingly simple Cardano formula (14.4.16) is tricky. For instance, if (14.4.15) has the form $x^3 - 3x = 0$, that is, if $p = -3$ and $q = 0$, the equation has the following simple roots: $x_1 = 0$, $x_2 = \sqrt{3} \simeq 1.73$, $x_3 = -\sqrt{3} \simeq -1.73$, then (14.4.16) yields the remarkable result $x = \sqrt[3]{\sqrt{-1}} + \sqrt[3]{-\sqrt{-1}} = \sqrt[3]{i} + \sqrt[3]{-i}$; that is, we suddenly have square roots of -1.

The procedure by which square and cube roots can be extracted from complex numbers was developed by one of the Cardano's followers, Raphael ***Bombelli*** (c. 1530-1572), who was able to explain how formula (14.4.16) guarantees in all cases a correct way of finding the roots (three roots, though not all necessarily distinct) of Eq. (14.4.15) (see Exercise 14.4.4).

The most influential French mathemati-

[14.8] Cardano did not discover it, however; he borrowed the formula—(14.4.16)—from Niccolò ***Tartaglia*** (whose real name was Niccolò Fontana) (c. 1500-1557) (e.g. see M. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972, p. 263).

[14.9] On reducing the arbitrary cubic equation $ax^3 + bx^2 + cx + d = 0$ to (14.4.15), see formulas (1.5.6) to (1.5.6a).

cian of the 16th century, François ***Viete*** (1540-1603), showed that when $(q/2)^2 + (p/3)^3$ is less than zero, that is, when the cube roots in Cardano's formula (14.4.16) are extracted from complex numbers, it is possible to reduce the right-hand side of (14.4.16) to a simple combination of trigonometric functions. This paved the way to using trigonometric functions in the theory of raising complex numbers to a power (even negative and fractional powers), that is, to the general formula (14.1.6). But in complete form, formula (14.1.6) was first expressed by the French mathematician Abraham de ***Moivre*** (1667-1754). L. Euler, however, gave it its modern form. (De Moivre fled from his native France to England to escape the persecution of the Huguenots, and he made a big contribution to the progress of mathematics in England.)

The famous Leonhard ***Euler*** made wide use of complex numbers. He was the one to introduce the notation i for $\sqrt{-1}$, and he also found most of the results of this chapter. Euler was convinced of the validity of the fundamental theorem of algebra, which states that every polynomial equation of degree n (with real or complex-valued coefficients) has exactly n roots (in general, complex-valued), and he made many attemps to prove it, but this requires a full understanding of the algebraic and the geometric meaning of complex numbers (see Figure 14.1.1), an understanding that the mathematicians of those days did not yet have. Jean ***d'Alembert***, the second outstanding mathematician of the 18th century, came closer to proving the fundamental theorem of algebra, but he failed to carry it through to the end.

We owe the first proof (proofs, to be more exact, for there were several) of the fundamental theorem of algebra to Carl Friedrich ***Gauss*** (1777-1855). He also provided the modern idea of the plane of the complex variable $z = x + iy$, thus freeing complex numbers from the last traces of mystery, even of mysticism,[14.10] and produced many other profound results in the field of "complex algebra and analysis." For one, Gauss created a nontrivial "arithmetic of whole complex numbers", numbers like $a + ib$, where a and b are whole numbers, or integers. In this unusual arithmetic, 3 is a *prime* but 5 *is not*, since $5 = (1 + 2i)(1 - 2i)$. The theory of complex numbers can be said to have become a full-fledged branch of mathematics only beginning with Gauss. Before Gauss, a geometric interpretation of complex numbers (see Figure 14.1.1) was presented by Caspar ***Wessel*** (1745-1818), a self-taught Norwegian-born surveyor, and the Swiss Jean-Robert ***Argand*** (1768-1822), who was also a self-educated mathematician and a bookkeeper. However, the publications of these two practically unknown scientists failed to attract any attention at the time, and Gauss arrived at his ideas independently.

[14.10] Yet even the much simpler *negative* numbers were first introduced only by Cardano and Viete, and even in Viete's time mathematicians viewed them with great suspicion.

Exercises

14.4.1. Prove formulas (14.4.8) for adding hyperbolic functions.

14.4.2. Prove that $\sinh t = \dfrac{2\tanh(t/2)}{1-\tanh^2(t/2)}$, $\cosh t = \dfrac{1+\tanh^2(t/2)}{1-\tanh^2(t/2)}$, and $\tanh t = \dfrac{2\tanh(t/2)}{1+\tanh^2(t/2)}$.

14.4.3. What curves are given by the following parametric equations:

(a) $x = \dfrac{1-t^2}{1+t^2}$, $y = \dfrac{2t}{1+t^2}$;

(b) $x = \dfrac{1+t^2}{1-t^2}$, $y = \dfrac{2t}{1-t^2}$?

What geometric meaning does parameter t have in these formulas?

14.4.4. How does formula (14.4.16) lead to the values 0 and $\pm\sqrt{3}$ of the roots of the cubic equation $x^3 - 3x = 0$ considered in the text?

14.4.5. Prove that $\cosh\psi = OP$ and $\sinh\psi = MP$ in the notation of Figure 14.4.2*a*, where ψ is twice the area of the hatched hyperbolic sector.

Chapter 15 Functions the Physicist Needs

15.1 Analytic Functions of a Real Variable

The idea of a function and of a functional relationship have undergone much change through the history of mathematics. Different approaches to the notion of a function in various periods sometimes led to stormy discussions,[15.1] and at present are reflected in the different ways in which a function can be defined.

The broadest possible definition of a function states that any relationship between the element of two sets can be called a function. If Mary is wearing a white dress, Ann red, and Dorothy blue, we can say that the color of the dress is a function of the name of the girl (or that the dress is a function of a certain person). Here the domain of definition of the function (or "input," as mathematicians today say) consists of three names (or three girls)—Mary, Ann, and Dorothy—while the range of values of this function (or "output") consists of three colors—white, red, and blue (or three dresses). Experimentally observed dependences of one type of physical quantities on another type, for instance, electrical resistance of a wire on the wire's temperature, are also functions. Finally, one can write many formulas of the type $y = ax^2 + bx + c$ and $y = e^{-x^2/2}$, and all these are functions, too.

But is it advisable to gather all these different ideas under one name, function? In some respects it is, since this stresses the enormous generality of the concept of function. This generality, however, is attained at a certain price.

A function given purely verbally (color of dress as function of name of girl) cannot be expressed by a formula, so one cannot find the derivative or integral of such a function, with the result that all the tools developed over the centuries by mathematicians prove to be useless, since these tools cannot be applied to such functions.

Functions obtained as a result of experiments in the form of lists of experimental data always prove to be given for a certain set of values of the independent variable within a certain accuracy.[15.2] Can the derivative or integral of such a function be found? By the very meaning of a derivative (or integral) such calculations involve finding the values of the function in a small neighborhood of a fixed value of the function and the independent variable. But it is impossible to obtain such values from a table or experiments directly, especially if one takes into account the errors of measurements. One must assume that there exists a smooth functional relationship between the quantities that are measured, since only then can we use the tools of higher mathematics; the experimental data or tables are needed to construct such a relationship.

Of tremendous importance is the fact that a fundamental theory always leads to quite natural mathematical formulas. This is also true when we are forced to reverse certain relationships (say, solve algebraic equations) and arrive at solutions that are discontinuous, that is, a gap between the roots of an equation. A physical theory is always constructed in such a way that it preserves the possibility of using that precious technique for studying phenomena, differential and integral calculus.

Thus, functions given by *formulas* possess enormous advantages. It is clear, firstly, that specifying a formula

---

15.1 Best known of all is the controversy on this subject (mentioned in Section 10.8) involving three outstanding mathematicians of the 18th century—Leonhard Euler, Jean d'Alembert, and Daniel Bernoulli. (Actually they never gave up their original opinions.) The real question was: can a graph with a break in it or a salient point be considered to represent a single function?

15.2 There are inevitable errors when the graph of a function is the only source of information about the function (say, plotted by a self-recorder).

enables calculating a function to any accuracy and for any value of the independent variable. In contrast to the case where the function is given by a table or lists of experimental data, repeated calculations by a formula always yield the same result—the accuracy can always be preassigned. Another convenient feature of a formula is that calculating the values of the function for *different* values of the independent variable is reduced to a sequence of similar, *single-type*, operations; this feature is especially important when computers are employed. A formula also enables calculating, to any accuracy, the difference of two values of the function at different (but closely lying) values of the independent variable, which makes possible the numerical calculation of the derivative. Adding the values of a function at adjacent points, we can calculate the integral of this function, and thus both the derivative and the integral can be calculated to *any degree of accuracy*. Today, in the computer age, such direct computation of derivatives and integrals is quite simple.

However, even among functions for which one can state an exact method for calculating $f(x)$ from the value of x there are monstrosities. Take the following function: $f(x) = 1$ if x is an irreducible fraction with an even denominator and $f(x) = 0$ for all the other values of x:[15.3]

$$f(x) = \begin{cases} 1 \text{ if } x = m/n, \text{ where } m \text{ and } n \text{ are relatively prime and } n \text{ is even,} \\ 0 \text{ otherwise.} \end{cases}$$

Just try and construct such a function, and you will see that this is impossible. You would have to build a dense fence of vertical segments of unit length each erected at points of the x axis that correspond to numbers of the form $m/n = m/2n_1$, where m and n_1 are relatively prime positive integers and m is odd; there is an infinitude of such numbers on any given finite segment of the x axis. The techniques developed in this book are not suitable, of course, for studying such functions.

On the other hand, we have, say, the "ideal" function $y = 1 - x$, whose graph is a smooth (even straight) line and which at each point has a tangent line (which coincides with the graph of the function itself), that is, the function has a derivative everywhere. The inverse of this function can also be easily found (it coincides with the initial function); the same is true of the function $f(f(x))$ (which is simply $F(x) = x$); it is easy to find the integral of this function:

$$\int_a^b f(x)\,dx = \left(x - \frac{x^2}{2}\right)\Big|_a^b$$

$$= (a-b)\left(\frac{a+b}{2} - 1\right);$$

and so on.

Intermediate cases are also possible, that is, functions that are not as "bad" (or, as mathematicians say, pathological) as the "fence" we described above and yet not as "good" as the function we have just described. So how is one to classify the "good" and "bad" functions? What functions are considered "good"? One approach that clarifies this matter considerably is the transition to complex numbers and (complex-valued) functions of a complex variable—but first we will try to answer these questions while remaining in the real domain for the time being.

One of the main results of Chapter 6 is the possibility of representing functions in the form of series:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots \qquad (15.1.1)$$

[15.3] Even such an exotic function (but still mathematically cerrect) can be expressed by a formula, albeit a rather complicated one.

(Maclaurin's series) or

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \ldots \quad (15.1.2)$$

(Taylor's series, which generalizes Maclaurin's series). These series make it possible, if we know the value of a function at one point and the derivative at the same point, to reconstruct the function (with as high an accuracy as desired) in the vicinity of the point and continue it to another point. If, say, formula (15.1.1) holds true, then, differentiating it term by term, we find the series expansion for the derivative

$$f'(x) = f'(0) + \frac{f''(0)}{1!}x + \frac{f'''(0)}{2!}x^2 + \ldots \quad (15.1.3)$$

and, if required, for higher-order derivatives. Similarly, integrating (15.1.1), we can write

$$\int f(x)\,dx = C + f(0)\,x + \frac{f'(0)}{2!}x^2 + \frac{f''(0)}{3!}x^3 + \frac{f'''(0)}{4!}x^4 + \ldots . \quad (15.1.4)$$

It is to functions of this type that the techniques developed in this book can be applied most successfully. Functions that can be represented in the form (15.1.1) or (15.1.2) are known as ***analytic functions***.[15.4]

[15.4] Of course, formulas (15.1.1) and (15.1.2) can be applied usually only in a limited range of variation of x (that is, in a limited vicinity of point $x = 0$ or $x = a$), a situation discussed in detail in Section 6.3. In accordance with this, only the concept of a ***function analytic at a point*** (i.e. such that it can be expanded in a power series (15.1.1) or (15.1.2) in the vicinity of the point under consideration) has a rigorous meaning, a ***function*** that is ***analytic on an interval*** is analytic at *each* point of that interval. Here we will not dwell on the possibility of term-by-term differentiation and integration of formulas (15.1.1) and (15.1.2); this possibility usually materializes for the majority of functions that a scientist or engineer needs.

It is clear that an integral rational function (a polynomial)

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \quad (15.1.5)$$

is analytic: here $f(0) = a_0$, $f'(0) = a_1$, $f''(0) = 2!a_2$, $f'''(0) = 3!a_3$, ... , $f^{(n)}(0) = n!a_n$, and $f^{(m)}(0) = 0$ for $m > n$ (thus, representation (15.1.1) of such a function coincides with (15.1.5)). We can say that analytic functions, such as $e^x = 1 + x + x^2/2! + x^3/3! + \ldots$ or $\cos x = 1 - x^2/2! + x^4/4! - \ldots$ or $\ln x = 1 + x + x^2/2 + x^3/3 + \ldots$ (see Chapter 6), are the most natural generalizations of polynomials: a general analytic function (15.1.1) or (15.1.2) is, so to say, a polynomial of infinite degree.

Of course, a function that is analytic at a point $x = a$ is "infinitely smooth" at this point, that is, has derivatives of *all* orders at this point, since the derivatives are present in representation (15.1.2), the existence of which is equivalent to the function being analytic. However, there is *no rule* by which the existence of all the derivatives of a function leads to the analyticity of the function; in other words, analyticity is a *stronger* requirement than the condition that the function have an infinitude of derivatives. Indeed, the fact that a function has all the derivatives implies only that we can write the series on the right-hand side of (15.1.2), while for the function $f(x)$ to be analytic it must be represented by the series, which means that the series must converge (the sum exists) for all x close to a. Take, for example, the function $y = e^{-1/x^2}$ (Figure 15.1.1). At $x = 0$ the expression e^{-1/x^2} has no meaning, but since for small (in absolute value) x the number $1/x^2$ is extremely large and, hence, e^{-1/x^2} is very small, it is natural to assume that $y(0) = 0$. The reader can easily see than all the derivatives also exist in this case: $y^{(n)}(x) = d^ny/dx^n$, with $n = 1, 2, 3, \ldots$, which can easily be found by the formulas of Chapter 4 (say, $dy/dx = e^{-1/x^2}[d(-1/x^2)/dx] = (2/x^3)e^{-1/x^2}$), but are all zero at $x = 0$ (see Exercise 15.1.1), which means that the right-hand side of Maclaurin's series (15.1.1) re-

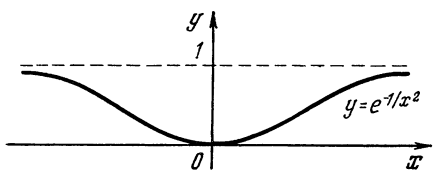


Figure 15.1.1

presenting our function is identically zero, while the function proper does not vanish, of course, at $x \neq 0$. Fortunately, functions like e^{-1/x^2} are very rare and are almost never encountered in physical and technical applications.[15.5]

The requirement that a function be analytic is really very strong. For instance, an analytic function (within the range of applicability of formula (15.1.1) or (15.1.2)) may be reconstructed from its values on an infinitely small segment (arc) of the graph of the function, since knowing $f(a)$ and the value of $f(x)$ for values of x that are arbitrarily close to a is quite sufficient for calculating all the derivatives of the function at point a. Therefore, analytic functions stand rather far from the definition of a function as an arbitrary correspondence between the values of x and the values of $f(x)$. However, almost all functions given by simple formulas are analytic; functions for which the analyticity condition is violated at separate points (like the function e^{-1/x^2} considered above) can be considered as pathological.

Of course, in physical problems we may also encounter functions that are not analytic at some points (say, functions that have discontinuities or salient points), and Chapter 16 is devoted to even more remarkable functions (it is questionable whether the name "function" can even be applied). In the majority of cases, however, the functions that a scientist or engineer needs are "good," or analytic; this statement is especially evident when we go over to functions of a complex variable (see below).

Exercise

15.1.1. Suppose $f(x) = e^{-1/x^2}$ for $x \neq 0$ and $f(0) = 0$. What is the derivative of such a function at $x = 0$? The second derivative $f''(0)$? The nth derivative $f^{(n)}(0)$?

[15.5] The reasons for such strange behavior of this function are clarified if we go into the complex domain, that is, if we continue the function in such a manner that x can take on *any complex* values (in this connection see Section 15.2).

15.2 The Derivative of a Function of a Complex Variable

Let us take a (complex-valued) function w of a complex variable z, that is, $w = f(z)$, where $w = u + iv$ and $z = x + iy$, with the numbers x and y and the functions $u = u(x, y)$ and $v = v(x, y)$ being real-valued.

If a function $f(z)$ can be expressed by a formula, then its derivative $df/dz = \lim_{\Delta z \to 0} [f(z + \Delta z) - f(z)]/\Delta z$ can be found by the general rules of Chapter 4 for (real-valued) functions of a real variable. Here are some examples: if $w = z^2$, then $dw/dz = 2z$; if $w = e^{kz}$, then $dw/dz = ke^{kz}$ (here k may be complex-valued); if $w = \ln z$, then $dw/dz = 1/z$; and so on. There are numerous examples of this type; for instance, we can form various combinations, such as sums (say, $z^2 + e^z$), products (say, z^2e^z), and functions of functions (say, e^{z^2}), with $d(z^2 + e^z)/dz = 2z + e^z$, $d(z^2e^z)/dz = 2ze^z + z^2e^z$, and $de^{z^2}/dz = 2ze^{z^2}$. It is quite natural, therefore, that all the rules referring to functions of a real variable remain valid when we go over to functions of a complex variable, since these rules are based on the general laws of algebra, say, the distributive property $(a + b)c = ac + bc$, which shows how to remove parentheses in calculations, and similar properties. These laws, or properties, do not depend on the nature of the quantities z, Δz, w, Δw. The rules for finding the derivative of $f(z) = z^2$ or of the product of two functions, w_1w_2, are based on the following relationships:

$$(z + \Delta z)^2 = z^2 + 2z\Delta z + (\Delta z)^2,$$

$$\Delta(z^2) = (z + \Delta z)^2 - z^2 = 2z\Delta z + (\Delta z)^2,$$

and, respectively,

$$(w_1 + \Delta w_1)(w_2 + \Delta w_2) = w_1w_2 + w_1\Delta w_2 + w_2\Delta w_1 + \Delta w_1\Delta w_2,$$

$$\Delta(w_1w_2) = (w_1 + \Delta w_1)(w_2 + \Delta w_2) - w_1w_2 = w_1\Delta w_2 + w_2\Delta w_1 + \Delta w_1\Delta w_2.$$

The raising to an imaginary or complex-valued power was defined by us

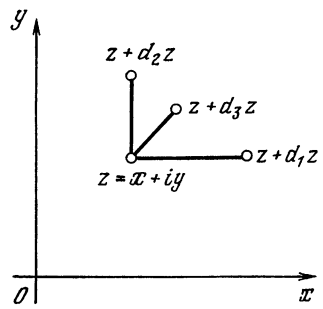


Figure 15.2.1

in such a way that the (approximate) equality $e^{\Delta z} \simeq 1 + \Delta z$, where we have discarded terms whose order of smallness is greater than that of Δz, remains valid (see Section 14.2), that is, we employed a formula that lies at the base of calculations of derivatives of the exponential function of a real variable, e^x. We can therefore be sure that all the rules for finding derivatives are applicable to functions of a complex variable: after we have defined that $e^{\Delta z} \simeq 1 + \Delta z$, all subsequent reasoning is on a stable foundation.

Note that the variation of the independent variable z of the function $f(z)$ can occur in different ways: we can vary only the real part of z, that is, $d_1 z = d_1 x$ and $d_1 y = 0$, or only the imaginary part of z, that is, $d_2 z = i d_2 y$ and $d_2 x = 0$, or the two parts simultaneously (Figure 15.2.1). The function w will change differently depending on the way we change z, but if dz is small, the ratio dw/dz, or the derivative, will remain the same in all cases.

Let us verify this using the function $w = z^2$ as an example. We have

$$w = (x + iy)^2 = x^2 - y^2 + 2ixy = u + iv, \qquad (15.2.1)$$

where $u = x^2 - y^2$ and $v = 2xy$; here

$$\frac{\partial u}{\partial x} = 2x, \quad \frac{\partial u}{\partial y} = -2y, \quad \frac{\partial v}{\partial x} = 2y, \quad \frac{\partial v}{\partial y} = 2x. \qquad (15.2.2)$$

Suppose $dy = 0$, that is, $dz = dx$. Then

$$dw \simeq w(z + dx) - w(z) = [u(x + dx, y) + iv(x + dx, y)] - [u(x, y) + iv(x, y)] \simeq \frac{\partial u}{\partial x} dx + i \frac{\partial v}{\partial x} dx.$$

Therefore, the derivative in this case is equal to

$$w'(z) = \frac{dw}{dz} = \frac{dw}{dx} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = 2x + i2y. \qquad (15.2.3)$$

Now assume that $dx = 0$ and $dz = idy$. Then $w + dw = w[x + (y + dy)i]$ and, hence,

$$dw = \frac{\partial u}{\partial y} dy + i \frac{\partial v}{\partial y} dy,$$

$$\frac{dw}{dz} = \frac{dw}{i\,dy} = -i \frac{dw}{dy} = -\frac{i\,\partial u}{\partial y} + \frac{\partial v}{\partial y} = -i(-2y) + 2x = 2x + i2y. \qquad (15.2.4)$$

In transformations (15.2.3) and (15.2.4) we took into account formulas (15.2.1) and (15.2.2) only in the last stages, while all the other reasoning retains its value for *all* functions, that is, if the derivative of the function $w = w(z) = u(x, y) + iv(x, y)$ does not depend on the choice of the increment dz of the independent variable, we have

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \qquad (15.2.5)$$

Since both u and v are real-valued functions, the above relation yields two formulas, the famous ***Cauchy-Riemann equations***:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \qquad (15.2.6)$$

These equations are satisfied automatically if w is given by a formula. Here are some examples:

(1) If $w = z^2 = (x + iy)^2$, then $u = x^2 - y^2$, $v = 2xy$, $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = 2x$, and $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} = 2y$ (compare with (15.2.2)).

(2) If $w=(2+3i)\,z^2=(2+3i)\,[(x^2-y^2)+i2xy] = [2(x^2-y^2)-6xy]+i[3(x^2-y^2)+4xy]$, then $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=4x-6y$ and $\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}=-6x-4y$.

(3) If $w=e^z=e^x\cos y+ie^x\sin y$, that is, $u=e^x\cos y$ and $v=e^x\sin y$, then $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=e^x\cos y$ and $\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}=-e^x\sin y$.

Note that the Cauchy-Riemann equations not only connect $u(x, y)$ and $v(x, y)$ but also impose certain conditions on *each* of these (real-valued) functions. Indeed, in view of (15.2.6),

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right)=\frac{\partial}{\partial x}\left(\frac{\partial v}{\partial y}\right),\ \text{i.e.}\ \frac{\partial^2 u}{\partial x^2}=\frac{\partial^2 v}{\partial x\,\partial y};\ \frac{\partial}{\partial y}\left(\frac{\partial v}{\partial x}\right)=\frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial y}\right),$$

i.e. $\frac{\partial^2 v}{\partial x\,\partial y}=-\frac{\partial^2 u}{\partial y^2}$.

From this we get $\frac{\partial^2 u}{\partial x^2}=-\frac{\partial^2 u}{\partial y^2}$ and, similarly, $\frac{\partial^2 v}{\partial x^2}=-\frac{\partial^2 v}{\partial y^2}$; in other words,

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,\quad \frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=0. \tag{15.2.7}$$

Any function $\varphi=\varphi(x, y)$ of two (real-valued) variables x and y that satisfies $\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}=0$ is known as *harmonic* (see footnote 9.5). Thus, if $w=f(z)=u(x, y)+iv(x, y)$ is a differentiable function of the complex variable $z=x+iy$, the real and imaginary parts of w (i.e. functions $u(x, y)$ and $v(x, y)$, respectively) are harmonic functions.

At this point it is advisable to return to the question of how functions are defined or classified, that is, the question posed at the beginning of the chapter. If we are speaking of functions of a complex variable, the most general approach would be as follows. The complex variable z consists of two parts, the real numbers x and y or, which is the same, it corresponds to a point in the plane (with coordinates x and y). The function $w=f(z)$ also consists of the real part u and the imaginary part v, whereby to each value of this function there corresponds a point in another plane (with coordinates u and v). To each z, that is, to each pair (x, y), or to each point in the xy-plane, we assign a pair of values of u and v, that is, a specific value of w (which can also be understood as a point in the uv-plane, and it is best to distinguish between the *two* planes so as to avoid confusion, the complex z plane and the complex w plane). If we assume that $u(x, y)$ and $v(x, y)$ are smooth functions, we can even obtain the (partial) derivatives $\partial u/\partial x$, $\partial v/\partial x$, $\partial u/\partial y$, $\partial v/\partial y$, $\partial w/\partial x$, and $\partial w/\partial y$. However, even then we cannot be sure, generally speaking, that the derivative $\frac{dw}{dz}=\frac{du+idv}{dx+idy}$ has a definite value, that is, a value independent of the choice of dx and dy. Thus, the general correspondence under which to each point $z=x+iy$ there corresponds (or is assigned) another point $w=u(x, y)+iv(x, y)$ cannot be used to determine dw/dz.

If we write a formula that expresses w as a function of z, such as z^2, e^z, or $z^3e^{z^2}$, the real and imaginary parts of w, that is, the functions $u(x, y)$ and $v(x, y)$, prove to be interconnected via the Cauchy-Riemann equations (15.2.6) and each obeys the harmonicity conditions (15.2.7). Thus, a formula that connects z and w restricts considerably the choice of the functions $u(x, y)$ and $v(x, y)$, but in return it enables finding the derivative dw/dz; moreover, the fact that w and z are related through an analytic function results in other important and neat corollaries. In this case one usually says $w(z)$ is *an analytic function of complex variable* z. All simple functions, such as z^2 and e^z, are analytic.

One must understand that analytic functions of a complex variable emerge when we specify the dependence between w and z by a formula; this does not mean, however, that all functions (that is to say, not all maps of

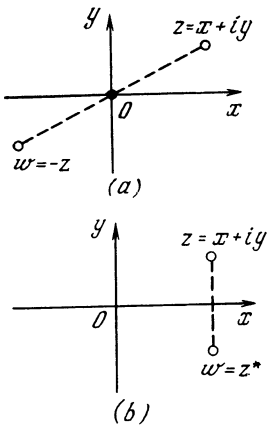


Figure 15.2.2

the complex z plane into the complex w plane) are necessarily analytic. For instance, it is clear that the function $w = -z$, which maps each point in the complex z plane into a point w (z) in the complex w plane that is *symmetric to z about the origin* (Figure 15.2.2*a*), is analytic; here $u = -x$ and $v = -y$, and, of course,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0$$

$$\left(\text{since in this case } \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} = \frac{\partial^2 v}{\partial y^2} = 0\right),$$

$$\frac{\partial u}{\partial x} = -1 = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = 0 = -\frac{\partial v}{\partial x}.$$

On the other hand, if we take the function $w = f(z) = z^*$, which maps point z of the complex z plane into point w (z) *symmetric to z about the real axis* (Figure 15.2.2*b*), we see that this is not an analytic function. Indeed, here $u = x$ and $v = -y$, and although, of course,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0,$$

we have

$$1 = \frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y} = -1, \text{ but } \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = 0$$

(for a function to be nonanalytic, *at least one* of the Cauchy-Riemann equations must not be valid). Similarly, the function $w = |z|^2 = zz^*$ is nonanalytic, too. (But is there an algebraic formula that can express the dependence of $|z|^2$ on z without involving z^*? Of course not!) The same is true of the function $w = f(z) = |z| = \sqrt{zz^*}$. Examples of such functions abound. Take the function $w = |z|^2 = x^2 + y^2$. Obviously, $u = x^2 + y^2$ and $v = 0$, or $\partial u/\partial x = 2x$, $\partial u/\partial y = 2y$, and $\partial v/\partial x = \partial v/\partial y = 0$, which implies that the Cauchy-Riemann equations are not satisfied in this case. In general, there is not a single function of a complex variable involving z^* that is analytic—not one of such functions can be written in the form of an algebraic (or analytical) formula.[15.6] The same is true of functions of variable z that involve $|z|$ or (the more so) Arg z.

The properties of analytic functions of a complex variable sometimes shed light on the properties of ordinary (i.e. real-valued) functions of a real variable, properties that often seem mysterious if one does not resort to complex variables. As an example, we can take the function $y = (1 + x^2)^{-1}$ discussed in Section 6.3 (the graph of this function is shown in Figure 6.3.3). The function can be expanded in a Maclaurin's series (formula (6.3.8)), which unexpectedly converges only in a limited range, $|x| < 1$, although this function possesses no singularities in the real domain. The reason for such behavior is that the (analytic) function $w = (1 + z^2)^{-1}$ of the complex variable z becomes infinite at points $z = \pm i$, which lie on the circle $|z| = 1$, and it is these singularities that make it impossible to continue Maclaurin's series for this function outside the circle $|z| < 1$. Of similar origin is the singularity at point $x = 0$ of the function $y = e^{-1/x^2}$ discussed in Section 15.1 (the graph of this function is shown in Figure 15.1.1). For the function e^{-1/z^2} of the complex variable z the point $z = 0$ is, as mathematicians usually say, an ***essential singular point***; at such a point the function cannot have *any* value at all. Indeed, when we send z to zero along the real axis, that is, $z = x$, with the (real) number x being small in absolute value, we find that $-z^{-2} = -x^{-2}$ is a very large (in absolute value) *negative* number and $e^{-1/z^2} = e^{-1/x^2}$ is extremely small. But if we send z to zero along the imaginary axis, that is, $z = iy$, where y is very small in absolute value, then $-z^{-2} = -(iy)^{-2} = y^{-2}$ is a very large *positive* number and, hence, $e^{-1/z^2} = e^{1/y^2} \to \infty$ as $y \to 0$. If we send z to zero in another direction, the function e^{-1/z^2} can tend to *any* arbitrary value. This means that our function can in no way be considered analytic at point $z = 0$ (even no value can be assigned to it at $z = 0$), and it is also clear that in the neighborhood of this point the function cannot be expanded in a power series (however, in the neighborhood of any other point the function can be expanded in a power series).

Note that, say, the function $w = f(z)$ such that $w = 0$ for $x < 0$ and $w = 1$

[15.6] This, of course, is not true of such artificially constructed functions as $w = z^{**}$, or $w = (z^*)^* = z$, where the asterisk has a purely formal function.

for $x > 0$ is not analytic, and the reason for this is not so much the jump in w at $x = 0$ as the fact that the quantities x and y, that is, the real and imaginary parts of z, are not of equal status, so to say, in relation to function w. For similar reasons the function $w = z^* = x - iy$ is not analytic either, since in constructing z^* we approach x and y from different "angles," with the real part of z remaining the same and the imaginary part of z changing its sign.

On the other hand, and this is important, if a function incorporates z and not x or y separately or z^* or $|z|$, it is analytic even if it cannot be expressed in the form of a finite combination of elementary functions. For instance, the function given by a *series*, say

$$w(z) = \sum_n \frac{z^{3n}}{(3n)!}, \qquad (15.2.8)$$

is analytic, and so is the function given by a definite *integral*, say

$$w(z) = \int_0^z e^{-p^2} dp \qquad (15.2.9)$$

(on the meaning of (15.2.9), where z and p are complex numbers, is touched on below), and the function that is a solution of a differential equation, say

$$\frac{dw}{dz} = z^3 - w^3, \quad \text{where} \quad w = 0 \text{ at } z = 0,$$

although in all three cases w cannot be expressed in terms of z via elementary functions. But the example of the Cauchy-Riemann equations (or the conditions of harmonicity for functions $u(x, y)$ and $v(x, y)$) shows that we can make certain statements concerning an analytic function even if we only know that it is analytic and know nothing more about its concrete form.

More than that, the general theory of functions of a complex variable states that *every* analytic function can be expressed in the form of a *series* or *integral*. While for an arbitrary (real-valued) function of a real variable the presence of the first derivative does not necessarily mean that the function has derivatives of higher orders, for analytic functions of a complex variable this is not so (other functions in the complex domain are of no interest to us). The fact is that a function $w = f(z)$ that has a first derivative $w' = f'(z)$ automatically has a second derivative $f''(z) = dw'(z)/dz$ and, in general, all higher-order derivatives (see Exercise 15.2.2). In addition, in the neighborhood of a point $z = z_0$ where the function $w = f(z)$ has a derivative we have

$$f(z) = f(z_0) + \frac{f'(z_0)}{1!}(z - z_0) + \frac{f''(z_0)}{2!}(z - z_0)^2 + \frac{f'''(z_0)}{3!}(z - z_0)^3 + \ldots . \qquad (15.2.10)$$

This implies that, knowing the function in an *arbitrary small* neighborhood of point z_0 (this is sufficient for finding all the derivatives of the function at point $z = z_0$, that is, all the expansion coefficients in (15.2.10)), we can continue the function $w = f(z)$ outside the neighborhood by defining it via series (15.2.10). Moreover, finding the value of the function $w = f(z)$ at point $z = z_1$ in this manner (and in neighboring points), we can write an expansion at point $z = z_1$ similar to (15.2.10) and, with the help of this expansion, find the values of w at new points. This process can be continued indefinitely. The very possibility of expanding analytic functions of a complex variable in a Taylor's series (15.2.10) justifies the use of the term "analytic" for functions of both real and complex variables. It also suggests that the idea of a "reasonable," or analytic, function of a complex variable stands rather far from the general idea of an "arbitrary" function, the definition of which requires specifying *all* its values, that is, specifying the entire range of the independent variable.

The properties of analytic functions of a complex variable are widely used in theoretical

physics. As yet there is no complete theory of elementary particles. But if we assume that such a theory can be constructed, that there exist formulas (unknown to us) expressing the main relationships of this theory, we can arrive at meaningful assumptions about the theory. It is natural to suppose that the formulas expressing the functional relationships are *analytic*. Then, even if we do not know their exact form, we can make some statements about the properties of elementary particles. For instance, the mass of a particle and that of the corresponding antiparticle must coincide. As for the properties of unstable particles, we can say that the lifetime of such a particle and that of the corresponding antiparticle coincide, too, although the decay products may be quite different. There is no way of explaining how scientists arrive at these far-reaching conclusions. However, it must be stressed that the assumption about the analyticity of the formulas that have still to be discovered plays the main role here.

The first attempts to apply the operations of differentiation and integration to complex quantities were made by Gottfried ***Leibniz*** and Johann ***Bernoulli***. For instance, Bernoulli found the simple formula $\int \frac{a\,dx}{b^2+x^2} = \frac{a}{b}\arctan\frac{x}{b}$ in 1712 by using manipulations connected with differentiation and integration of complex-valued expressions, manipulations which, although absolutely correct, had not yet been sufficiently substantiated. In the same work Bernoulli obtained, in the process of his transformations, the beautiful formula $\left(\frac{\tan\alpha - i}{\tan\alpha + i}\right)^n = \frac{\tan n\alpha - i}{\tan n\alpha + i}$ allied to the then still unknown de Moivre formula (14.1.6). However, neither Leibniz nor Bernoulli had fully understood the situation, as is testified by the debate between the two on the meaning of the logarithms of negative numbers, a debate connected with the subject of complex-valued functions and their analysis. Leibniz regarded these logarithms as imaginary, while Bernoulli saw them as real quantities. The publication in 1745 of the correspondence between Leibniz and Bernoulli immediately attracted the attention of the most prominent mathematicians of the period, Leonhard Euler and Jean d'Alembert, with Euler taking the side of Leibniz and d'Alembert supporting Bernoulli. Euler's article entitled "De la controverse entre Mrs. [Messers] Leibnitz et Bernoulli sur les logarithmes négatifs et imaginaires" (On the Dispute Between Leibniz and Bernoulli Concerning Logarithms of Negative and Imaginary Numbers), published in 1749, contained the modern theory of logarithms of complex numbers set forth in Section 14.3, and Euler emphasized the fact that the logarithmic function is *many-valued*, something which Leibniz and Bernoulli had not even suspected. Another feature characteristic of Euler was that he paid special attention to analytic functions of a real variable, which he called "continuous"[15.7]. Euler was also inclined to believe that these were the only functions worthy of the attention of mathematicians. Euler's work abounds in results that today belong to the theory of analytic functions of a complex variable, including many examples of the expansion of such functions into series and of the application of these series. However, Euler did not advance very far in studying the properties of analytic functions, both real-valued and complex-valued. For instance, he thought that knowing the values of a real-valued analytic function $y = f(x)$ on an arbitrarily small interval over which the independent variable x varies permits reconstructing the function on the entire real x axis, which is, of course, not the case.

D'Alembert made a fundamental step forward in the theory of functions when he clearly stipulated that both the independent variable and the value of the function can be both real-valued and complex-valued. On this point Euler's investigations followed d'Alembert's. The analytic function (that is, one fixed by a formula) $w = f(z)$, where $z = x + iy$, was written by d'Alembert as $u(x, y) + iv(x, y)$. For instance, by differentiating z^k, with k the constant and z the variable (both complex-valued), he found the correct formula (in the form of $u + iv$) for raising a complex number to a complex power. He was also the first to arrive at equations that are now known as the Cauchy-Riemann equations, (15.2.6), in his Essay on a New Theory of Resistance of Fluids (1752), where he considered the motion of a body through a homogeneous, weightless, ideal fluid. (The relationship between this problem of mechanics and the theory of analytic functions of a complex variable will be discussed below in Section 17.3.) In 1761 d'Alembert derived Eqs. (15.2.7) from the Cauchy-Riemann equations (15.2.6). Joseph Louis Lagrange did much to advance all this work on the mechanics of liquids and on the theory of analytic functions of a complex variable.

Continuing the work of d'Alembert in hydromechanics, Euler, in 1755, also obtained (independently of d'Alembert) Eqs. (15.2.6), to which he attached great importance and from which he obtained a number of corollaries (including results of a geometric nature connected with what are now called ***conformal transformations***[15.8]). In 1776/1777 Euler (who

[15.7] At present this term has quite a different meaning (notice the way in which Riemann uses it in his doctoral dissertation, from which we quote later on).

[15.8] A conformal transformation (of a plane or space) is one that preserves angles between curves and that, in the neighborhood of a point, possesses the properties of a simila-

was then about 70) launched on a pioneering study (and also application) of the integrals of functions of a complex variable (see Section 17.2). However, it was not until 1811 that Carl Friedrich ***Gauss***, in a letter to the well-known German mathematician and astronomer Friedrich Wilhelm ***Bessel*** (1784-1846), gave the first clear-cut definition of the integral of an analytic function of a complex variable and an idea of the basic properties of such integrals. Incidentally, this letter was published only in 1880, when all the results set forth in it were already known from investigations by Augustin ***Cauchy***, who had worked on this theme from 1813 onwards.

In some textbooks on the theory of functions of a complex variable Eqs. (15.2.6) are called "d'Alembert-Euler equations." It seems to us, however, that the traditional name, "Cauchy-Riemann equations," is more justified. The fact is that neither d'Alembert nor Euler created the theory of functions of a complex variable. For instance, their works contain no full definitions of the notions of the derivative and the integral of an analytic function; neither did they establish the relationship between Eqs. (15.2.6) and the very fact of the existence of the derivative of a function of a complex variable,

$$\frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{w(z+\Delta z) - w(z)}{\Delta z},$$

a derivative that is independent of the way in which Δz tends to zero. It was Cauchy who, in a whole series of lengthy papers (the most important of these was a paper on "complex integration" submitted for publication in 1825), gave a consistent theory of (differentiable, that is, analytic) functions of a complex variable. Riemann's brilliant doctoral dissertation of 1851 entitled "Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse" (Fundamentals of the General Theory of the Functions of a Complex Variable) played what was perhaps a greater role in advancing the theory of analytic functions of a complex variable. This was the first consistent and full exposition of the new theory and it laid the foundation for what we today call "the geometric theory of analytic functions." (The dissertation produced a whole course of lectures that Riemann delivered in Göttingen in the winter of 1855/1856 and the summer of 1856, in which he presented an amazing wealth of material and many profound ideas.[15.9])

An altogether different line in the theory of functions of a complex variable was begun by Karl Theodor Wilhelm ***Weierstrass*** (1815-1897) of Berlin, a man who was indisputably one of the most distinguished mathematicians of the 19th century. Unlike Riemann's reasoning, the Weierstrassian theory of functions could quite aptly be called "algebraic." Weierstrass, to whom geometric concepts were completely alien and whose thinking followed a strictly formal direction, harshly criticized Riemann (for whom, at the same time, he had profound respect and whose investigations he held in the highest esteem) for drawing on geometric pictorialness and for a somewhat casual attitude toward logic; the modern (very strict) criteria of "mathematical rigor" owe their origin largely to Weierstrass. In the theory of functions Weierstrass's main tools were power series and operations performed on such series; he greatly advanced the "formal algebraic" theory of analytic functions of a complex variable. For one thing, following an altogether different approach, he obtained a large number of results that Riemann had first stated but had not backed up with flawless proofs. The rivalry between Riemann and Weierstrass can be compared with the relations between Leibniz and Newton. Here Weierstrass followed Leibniz, and Riemann followed Newton.

Riemann was a most profound thinker in the realm of physics and to a certain extent anticipated the general idea of Einstein's theory of gravity and created a consistent mathematical formalism without which the general theory of relativity would have been impossible. The mutual respect in which these two scholars held each other, although their thinking was completely different (they did not even understand each other very well!) shows that differences in scientific ideology do not necessarily lead to personal enmity. The example of how Riemann and Weierstrass, proceeding "from different angles," built the theory of analytic functions of a complex variable, serves as another reminder of the fruitfulness of different approaches in the process of scientific investigation.

We conclude this chapter with an extensive quotation from the above-mentioned disser-

---

rity transformation. It is easy to see that the transformation, or map, $z \to w(z)$, of the plane of the complex variable z on itself specified by the analytic function $w = f(z)$ with derivative $f'(z)$ is a conformal transformation: from the fact that $w - w_0 \simeq f'(z_0)(z - z_0)$, where $w_0 = f(z_0)$, in the neighborhood of a point $z = z_0$ and from the rules of operation on complex numbers (see Chapter 14) it follows that near z_0 the transformation consists of a translation of the complex z plane to the point w_0 (by the vector $\overrightarrow{z_0 w_0}$), a ρ-fold dilation, where $\rho = |f'(z_0)|$, and a rotation through the angle φ, with $\varphi = \operatorname{Arg} f'(z_0)$.

15.9 This course of lectures, which played a great role in the history of mathematics, was attended by only three persons. (By the way, on p. 210 we mentioned a most important lecture course by Johann Bernoulli which was attended by only one student.)

tation by Riemann.[15.10] It embraces the main points of everything we have said above and expressively demonstrates how it all began:

Let z denote a variable that can change without limit and assumes all possible real values. If to each of its values there corresponds only one value of another variable, w, then w will be called a function of z. If, furthermore, while variable z changes continuously and runs through all the values contained between two constant limiting points, the magnitude of w also changes continuously, then within the aforesaid interval this function is called continuous. It is clear that this definition in no way establishes a connection between separate values of the function: if the behavior of the function is indicated only in a certain definite interval, beyond this interval it can be continued quite arbitrarily.

The relationship between variable w and variable z can be fixed by a mathematical formula in such a way that the value of w corresponding to each value of z is obtained through certain mathematical operations conducted on the numerical value of z....

Suppose that the range of variable z is not restricted to real values but spreads to complex numbers $x + yi$, with $i = \sqrt{-1}$, as well. Let $x + yi$ and $(x + dx) + (y + dy)\,i$ be two infinitely close values of z to which there correspond two values of variable w, namely $u + vi$ and $(u + du) + (v + dv)\,i$. In this case, if the dependence of variable w on variable z is assumed to be quite arbitrary, the ratio $(du + dvi)/(dx + dyi)$, generally speaking, changes with dx and dy. If we assume that $dx + dyi = \varepsilon e^{\varphi i}$, we clearly see that

$$\frac{du+dvi}{dx+dyi} = \frac{1}{2}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right) + \frac{1}{2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right) i + \frac{1}{2}\left[\left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right) + \left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right) i\right]\frac{dx-dyi}{dx+dyi}$$

$$= \frac{1}{2}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1}{2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right) i + \frac{1}{2}\left[\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right) i\right] e^{-2\varphi i}.$$

But each time the function w of variable z is defined by a formula containing simple mathematical operations, it is found that the value of the derivative dw/dz is independent of the value of the differential dz.[15.11] Thus, it becomes clear that not every dependence of the complex variable w on the complex variable z can be expressed in the above manner.

This characteristic property of all functions that can be defined in terms of elementary mathematical operations will be put at the basis of our further investigation..., namely, we will proceed from the following definition...:

A complex variable w is said to be a function of another complex variable z if both vary in such a manner that the value of the derivative dw/dz does not depend on the value of dz.

...If we represent the ratio $(du + dvi)/(dx + dyi)$ in the form

$$\frac{\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial x}\,i\right)dx + \left(\frac{\partial v}{\partial y}-\frac{\partial u}{\partial y}\right)dyi}{dx+dyi},$$

it becomes clear that the ratio has one and only one value for all values of dx and dy provided that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

Thus, these conditions are necessary and sufficient for $w = u + vi$ to be a function of $z = x + yi$. This leads us to the following conditions imposed on the separate terms of the function:

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2} = 0,$$

which serve as a basis for studying the properties of each term of such functions when it is considered separately from the other term.

Exercises

15.2.1. Check the validity of the Cauchy-Riemann equations for the functions of a complex variable that were encountered in Exercise 14.1.1 and for the function $w = \ln z$.

15.2.2. Suppose that $w = f(z) = u(x, y) + iv(x, y)$ is an analytic function of the complex variable z and $f'(z) = dw/dz = u_1(x, y) + iv_1(x, y)$ is its derivative. Prove that the Cauchy-Riemann equations for $f(z)$ automatically imply the validity of the Cauchy-Riemann equations for $f'(z)$.

15.10 G. F. B. Riemann, *Gesammelte mathematische Werke*, 2nd ed., Dover (reprint), New York, 1953, pp. 3-43.

15.11 This statement is obviously valid in all cases where $\partial w/\partial x$ can be obtained from the expression of w in terms of z via the ordinary rules of differentiation.... (Riemann's footnote.)

Chapter 16 Dirac's Remarkable Delta Function

16.1 Various Ways of Defining a Function

The functions we have been studying up to now have ordinarily been defined by formulas. This means that a procedure was always indicated for computing the values of the function for any given value of the independent variable. We could call this an *algorithmic* representation (algorithm meaning a procedure for computing something). To illustrate, take the function $y = f(x) = 2x + 3x^2$. It actually amounts to this: "take x, multiply by 2, square x, multiply by 3, and then add the two numbers to get the value of y for the given value of x." Trigonometric functions were defined differently, by means of geometric concepts, by measuring arcs and line segments in a circle. However, here too we can speak of an algorithm for calculating their values, which amounts to turning to *trigonometric tables* or to a pocket calculator with the appropriate keys.

Up to now we have studied the properties of functions specified in this fashion, the laws of increase and decrease of functions, the laws for finding the maxima and minima of functions, etc. The study of these properties leads to new ways of defining functions. For instance, the function $f = e^x$ may be defined as a function whose derivative is equal to the function itself: $df/dx = f$, with the supplementary condition that $f(0) = 1$. The sine, the function $\varphi = \sin x$, and the cosine, or $\psi = \cos x$, may be defined as functions that satisfy one and the same equation

$$\frac{d^2\varphi}{dx^2} = -\varphi, \quad \frac{d^2\psi}{dx^2} = -\psi$$

under different initial conditions

$$\varphi(0) = 0, \quad \left.\frac{d\varphi}{dx}\right|_0 = 1;$$

$$\psi(0) = 1, \quad \left.\frac{d\psi}{dx}\right|_0 = 0.$$

These definitions are in many respects more to the point, so to say, and more closely related to the applications of the exponential function and of trigonometric functions to many problems of physics, say, to problems involving oscillations, than are the definitions

$$e^x = \left[\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n\right]^x$$

$$\left(\text{or } e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n\right),$$

and of the sine and cosine as certain segments in a circle or as ratios of sides of a right triangle.

It is curious to note that the definitions of e^x, $\sin x$, and $\cos x$ via differential equations prove to be convenient for electronic computers. If in a calculation the investigator has to substitute into a formula the values of e^x for different values of x, he copies them out of a table. When operating a computer, it is more convenient and faster to have the machine compute e^x step by step via the approximate formula $e^{(x+\Delta x)} \simeq e^x(1 + \Delta x)$ in accordance with the equation $dy/dx = y$ for the function e^x (or via more exact formulas based on the same equation) than it is to refer to a table of values. The same goes for the functions $\sin x$ and $\cos x$. It is easier to compute them every time simultaneously:[16.1]

$$\sin(x + \Delta x) \simeq \sin x - \Delta x \cos x,$$

$$\cos(x + \Delta x) \simeq \cos x - \Delta x \sin x.$$

Thus, one general approach to the concept of a function lies in specifying a procedure for computing it and in subsequently investigating it. There is also another approach. We can seek a function with definite general properties, the aim being later to attempt (on the basis of these properties) to find the formula that describes the function at hand. Such is the usual procedure when handling experimental

[16.1] Where do you think these formulas come from? Check to see whether they accord with the equation for the second derivative of the sine and cosine.

data and finding empirical formulas by trial. Our object here is to construct in this way a remarkable function that is useful and important both in mathematics and in its applications.

16.2 Dirac and His Function

Paul Adrien Maurice ***Dirac*** (1902-1983), the celebrated English theoretical physicist, came to fame in 1929. He had elaborated a theory capable of describing the motions of electrons in electric and magnetic fields with arbitrary velocities almost up to the velocity of light. This was the quantum theory which also accounts for the fact that the electrons in an atom move only in specific orbits with definite energy values. Dirac knew that an electron possesses a definite rotational moment, or is similar to a spinning top, and he took this into account in building his theory. When the theory was constructed, it turned out that a conclusion could be drawn that Dirac had not foreseen, namely, the existence of particles with mass the same as the electron mass but with opposite (positive) charge. For two years it was thought that Dirac's theory was good for describing electron motion but that the conclusion concerning particles with positive electron charge was erroneous and that as soon as he got rid of it the theory would be a very good one.

But in 1932 a positively charged particle with the electron mass, called the ***positron*** (also called the antiparticle of the electron), was discovered. The big drawback of Dirac's theory became its triumph, its principal contribution: Dirac's discovery was the first instance of a new particle being discovered "at the tip of a pencil." This is an instructive example from the standpoint of the relationships of theory and experiment. Theory rests on the findings of experiment, but a consistent, logical and mathematical, development of a theory takes the investigator beyond the confines of the material used as its foundation and leads to fresh predictions.



Paul Dirac

Dirac was not only one of the best theoretical physicists in the world, he was a marvelous mathematician. In his classical *Principles of Quantum Mechanics* Dirac introduced and made wide use of a new function, which he denoted by $\delta(x)$. It is known as ***Dirac's delta function*** or, simply, the ***delta function***.

The delta function can be defined as follows: $\delta(x) = 0$ for any $x \neq 0$, that is, for $x < 0$ and $x > 0$, and $\delta(0) = \infty$. In addition, the delta function must satisfy the following condition:

$$\int_{-\infty}^{+\infty} \delta(x)\, dx = 1. \qquad (16.2.1)$$

Figure 16.2.1 gives a pictorial view of a function similar to the delta function. The narrower we make the strip between the left and right branches, the higher must the strip (the integral, that is) be to retain the given value of 1. As the strip becomes ever narrower, we approach the condition $\delta(x) = 0$ for $x \neq 0$, and the function approaches the delta function. In one of the following sections these arguments will be utilized in the construction of formulas

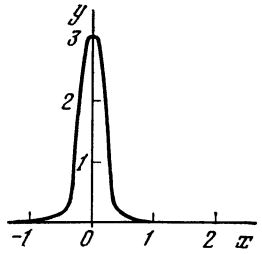


Figure 16.2.1

that yield the delta function. Here we continue the study of its general properties.

It must be stressed, of course, that the "equations" defining the delta function,

$$\delta(x) = \begin{cases} 0 & \text{if } x \neq 0, \\ \infty & \text{if } x = 0, \end{cases} \qquad \int_{-\infty}^{+\infty} \delta(x)\, dx = 1, \tag{16.2.2}$$

must not be taken too literally, since from the standpoint of classical mathematics (that is, the mathematics we are accustomed to) these conditions are meaningless and contradictory. Obviously, the above-described process of stretching (and contraction) of the strips clarifies the meaning of the delta function, the "intuitive" definition (16.2.2) of the delta function can only stimulate our imagination in relation to this process.

It is clear that by defining the delta function we introduce a completely new object, so unlike any other object discussed earlier (there is no way in which we can say that our new function is analytic in the sense of Chapter 15). However, the description of this new object is fairly simple and must not put us off. For example, the number $\sqrt{2}$, which is constantly used in mathematics, is not a "real" number from the standpoint of arithmetic: it is not given by its exact value but by the sequence 1, 1.4, 1.41, 1.414, 1.4142, . . . consisting of approximate values of $\sqrt{2}$ (which, incidentally, characterizes $\sqrt{2}$ sufficiently in order to freely use it). Similarly, $\delta(x)$ is given not by its exact values but by approximations (tall and narrow "steps") which completely define the function.[16.2]

Dirac and other physicists have for many years used the delta function freely and productively, without any strict definitions existing for such strange objects, just like mathematicians for centuries used irrational numbers until, at the end of the 19th century, the first theories of the real variable appeared. Mathematical justification for such remarkable entities as the delta function ("generalized functions", or distributions, as they are usually called) was given by the Soviet mathematicians S. L. Sobolev and I. M. Gel'fand, the French mathematician Laurent Schwartz, and others; this topic is discussed in many books.[16.3]

Possibly, the following important formula gives a better description of the Dirac delta function than (16.2.2):

$$\int_{-\infty}^{+\infty} f(x)\,\delta(x)\,dx = f(0). \tag{16.2.3}$$

Indeed, since $\delta(x) = 0$ for $x \neq 0$, we conclude that the value of the integral on the left-hand side of (16.2.3) cannot depend on the values of $f(x)$ no matter what $x \neq 0$. The only essential value of $f(x)$ is the one where $\delta(x) \neq 0$, that is, for $x = 0$. This means that in the narrow region where $\delta(x) \neq 0$ (in the limit the width of this region tends to zero; see Figure 16.2.1) $\delta(x)$ is multiplied by $f(0)$. Hence, formula (16.2.3) follows from the conditions (16.2.2).

16.2 For those interested, here are the "decimal" approximations of the delta function: $y = 0$ for $|x| > 1$, $y = 0.5$ for $|x| < 1$; $y = 0$ for $|x| > 0.1$. $y = 5$ for $|x| < 0.1$; $y = 0$ for $|x| > 0.01$, $y = 50$ for $|x| < 0.01$; . . . (it is more convenient to "smooth out" these steps, so that all functions are made continuous); below we will give other (and still more convenient) definitions of the delta function.

16.3 Possibly, the simplest of these are two small books by J. Mikusinski and R. Sikorski, *The Elementary Theory of Distributions*, published in English in Warsaw in 1957 and 1961. For instance, in the first book the authors introduce generalized functions as limits of sequences consisting of ordinary functions, say, the functions $\varphi_n(x)$, where $\varphi_n(x) = 0$ for $|x| > 1/10^n$, $\varphi_n(x) = 10^n/2$ for $|x| < 1/10^n$; see footnote 16.2. (It goes without saying that the concept of the limit of a function requires a rigorous definition.)

We can also argue in reverse. We can say that $\delta(x)$ is a function such that no matter what form the auxiliary function $f(x)$ has, we always have formula (16.2.3). This condition alone brings us to all the conclusions concerning the form of $\delta(x)$ that were earlier employed in the definition of $\delta(x)$. From formula (16.2.3) it follows that $\delta(x)=0$ at $x\neq 0$, that $\int \delta(x)\,dx=1$, and that $\delta(0)=\infty$.[16.4]

Note that, of course, to grasp the meaning of formula (16.2.3) requires, like the definition (16.2.2) of the delta function, a certain effort of thought, since the integrand on the left-hand side of (16.2.3) contains a thing that can only loosely be called a function. However, a "rough" treatment of the delta function makes it possible to justify both the doubtful definition (16.2.2) and formula (16.2.3), that is to say, both formulas are sufficient for applying this new concept and must always be taken into account.

Let us carry through a few more obvious consequences of the definition of $\delta(x)$. By the general rule of a change of variables (discussed in detail in Section 1.7), the function $\delta(x-a)$ is displaced a units to the right in relation to $\delta(x)$, that is, $\delta(x-a)=\infty$ at $x=a$ and is zero otherwise. Accordingly,

$$\int_{-\infty}^{+\infty} f(x)\,\delta(x-a)\,dx=f(a). \qquad (16.2.3a)$$

Now it is easy to see, if we consider a curve of the form depicted in Figure 16.2.1, that $b\delta(x)$ is b times higher than $\delta(x)$ and that $\delta(cx)$ is $|c|$ times narrower than $\delta(x)$, so that the area under the curve $\delta(cx)$ is $|c|$ times smaller than the area under $\delta(x)$. Therefore,

$$\int f(x)\,b\delta(x)\,dx=bf(0), \qquad (16.2.4)$$

$$\int f(x)\,\delta(cx)\,dx=\frac{1}{|c|}f(0). \qquad (16.2.5)$$

A comparison of (16.2.4) and (16.2.5) enables us to say that

$$\delta(cx)=\frac{1}{|c|}\delta(x).$$

Formula (16.2.4) is quite obvious, but we can also hope that the reader who has gone through the trials and tribulations of the preceding chapters of this book will be able to grasp formula (16.2.5) as well. Formally, it is readily obtained by the change of variable

$$u=|c|\,x, \quad dx=\frac{1}{|c|}\,du.$$

Here we also make use of the fact that the function $\delta(x)$ defined by formula (16.2.2) is an *even* function of its argument: $\delta(-x)=\delta(x)$.

Exercises

16.2.1. Show that for a function $\varphi(x)$ having a unique zero x_0, so that $\varphi(x_0)=0$ and $\varphi(x)\neq 0$ at $x\neq x_0$, we have the formula $\delta(\varphi(x))=\frac{1}{|\varphi'(x_0)|}\delta(x-x_0)$. What is the function $\delta(\psi(x))$ if the function $\psi(x)$ vanishes at several values of x?

16.2.2. Evaluate the integral

$$\int_{-\infty}^{+\infty} \psi(x)\,\delta(\sin x)\,dx.$$

16.3 Discontinuous Functions and Their Derivatives

Let us consider the integral of $\delta(x)$ as a function of the upper limit, that is,

$$\theta(x)=\int_{-\infty}^{x} \delta(z)\,dz. \qquad (16.3.1)$$

It is easy to see that the graph of this function has the form of a step (Figure 16.3.1). Indeed, as long as $x<0$, the domain of integration in (16.3.1) is wholly located where $\delta(x)=0$,

[16.4] Here the integral sign without any limits of integration will always be understood as being over the entire range of the variable of integration, from $-\infty$ to $+\infty$.

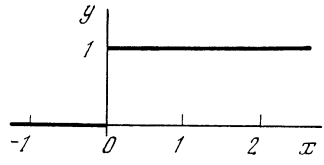


Figure 16.3.1

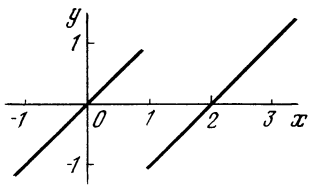


Figure 16.3.2

which means that $\theta(x) = 0$ when $x < 0$. But if $x > 0$, the domain of integration involves the neighborhood of the origin, where $\delta(0) = \infty$. On the other hand, since $\delta(x) = 0$ when $x > 0$, the value of the integral does not change when the upper limit changes from 0.1 (or even 0.000001) to 1 or to 10 or to ∞. Hence, for any $x > 0$ we have

$$\theta(x) = \int_{-\infty}^{x} \delta(z)\,dz = \int_{-\infty}^{+\infty} \delta(z)\,dz = 1,$$

as is shown in Figure 16.3.1.

Thus, with the aid of the delta function we have constructed the simplest discontinuous function $\theta(x)$ such that when $x < 0$ then $\theta(x) = 0$, and $\theta(x) = 1$ in the domain $x > 0$.[16.5] These simple considerations enable us to approach the problem of the derivative of a function having discontinuities in a more consistent fashion, without apparent exceptions and extensive reservations.

If we did not know about the delta function, we would have to say that derivatives cannot be found at points where a function is discontinuous. But we have just constructed a discontinuous function, $\theta(x)$. The general rule of the relationship between an integral and the derivative (see Section 3.3) is:

$$\text{if } F(x) = \int_{x_0}^{x} y(z)\,dz, \text{ then } y(x) = \frac{dF}{dx}.$$

If we apply it to the expression (16.3.1), we obtain

$$\frac{d\theta(x)}{dx} = \delta(x). \qquad (16.3.2)$$

Thus, we do not need to make an exception for the derivative of a discontinuous function. We merely say that at the point of discontinuity the derivative is equal to a "singular" function, the delta function.

We have learned to handle the derivative of an elementary discontinuous function and can now very simply find derivatives in more complicated situations. Here are some examples. Let

$$y = x \text{ for } x < 1 \text{ and}$$
$$y = x - 2 \text{ for } x > 1. \qquad (16.3.3)$$

We refer to the graph of this function in Figure 16.3.2. The jump occurs at $x = 1$, and the magnitude of the jump is $y(1 + 0) - y(1 - 0) = -2$. (Here we use the notation $y(1 + 0)$ to denote the limiting value of y as x approaches 1 from the *right*, that is, from the direction of $x > 1$, and $y(1 - 0)$ denotes the same on the left; see Figure 16.3.2.) From this we formally get

$$\frac{dy}{dx} = 1 - 2\delta(x - 1). \qquad (16.3.4)$$

This notation is better than the dreary statement that $dy/dx = 1$ everywhere except at point $x = 1$, where the function has a discontinuity and does not have a derivative.

The delta function is a typical brainchild of the twentieth century. The nineteenth century had a passion for investing all its arguments—true, not quite true, and simply false—in the form of "impossibilities." It is impos-

[16.5] At point $x = 0$ the function $\theta(x)$ undergoes a jump, and there is no need to define the value of $\theta(0)$ any further; suffice it to say that at this point the value of $\theta(x)$ changes from 0 to 1.

sible to invent a perpetual motion machine, transformations of chemical elements are impossible, it is impossible to change the total mass of a substance, it is impossible to prove the existence of atoms, it is impossible to determine the composition of the stars, it is impossible to find the derivative of a discontinuous function. The twentieth century has found numerous constructive solutions to what appeared to be impossible in the nineteenth century, say, to the problem of element transformations or to the reduction of the total mass of a substance by transforming mass into energy. To take an example, the delta function resolves the problem of the derivative at a point of discontinuity (at any rate for a discontinuity in the form of a finite jump). Indeed, the notation of (16.3.4) contains in one line the fact of discontinuity of the function $y(x)$ (since dy/dx involves the delta function), the nature of the discontinuity (a jump), the site of the discontinuity ($x = 1$), and the magnitude of the jump (the coefficient -2 of the delta function).

Integrating (16.3.4) with the condition $x = 0$, $y = 0$, we can restore the graph of $y(x)$ in its entirety. True, in the case of the function (16.3.3) we had it easy in the sense that a simple case was especially chosen where the derivatives on the left and on the right are expressed by a single formula. This of course is not obligatory. Why should the derivative be continuous if the function itself suffers a discontinuity?

Let us consider a more complicated example: $y = -x^2$ for $x < 1$, and $y = x^2$ for $x > 1$ (Figure 16.3.3). For $x < 1$ we have $y' = -2x$, while for $x > 1$ we have $y' = +2x$. The discontinuity is associated with $y' = 4\delta(x - 1)$. We can now, at our pleasure, adjoin the point $x = 1$ to the left-hand region and then write $y' = -2x + 4\delta(x - 1)$ for $x \leqslant 1$ and $y' = 2x$ for $x > 1$. Or, another version (quite similar to the first), we can adjoin $x = 1$ to the right-hand region and then, with the same

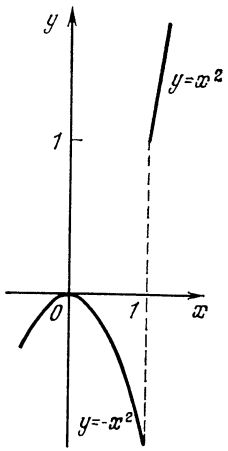


Figure 16.3.3

full justification, write $y' = -2x$ for $x < 1$ and $y' = 2x + 4\delta(x - 1)$ for $x \geqslant 1$. (Note how the signs $<$ (less than) and $\leqslant$ (less than or equal to), $>$ (greater than), and $\geqslant$ (greater than or equal to) are placed in the formulas. Be careful not to write the delta function twice.[16.6]) We can also write

$$y' = \varphi(x) + 4\delta(x - 1),$$

$$\text{where } \varphi(x) = \begin{cases} -2x & \text{if } x < 1, \\ 2x & \text{if } x > 1. \end{cases}$$

To verify the notation, integrate the expression of the derivative. You will obtain the original discontinuous function.

Sometimes in mathematics use is made of the so-called ***signum function*** sgn x or Sgn x.[16.7] It is defined thus: sgn $x = -1$ for $x < 0$ and sgn $x = +1$

[16.6] Thus, the derivative of a discontinuous function $f(x)$ at the point of discontinuity $x = x_0$ has a definite "value" $\delta(x - x_0)$. As for the discontinuous function $f(x)$ itself, there is no sense in asking for its value at the actual point of discontinuity. At any rate, this question is meaningless in nearly all applied problems.

[16.7] Sgn x is often read as "signum x" or "the sign of x," since the Latin for sign is signum. Sometimes it is assumed that sgn $0 = 0$, but we will never use this physically meaningless notation.

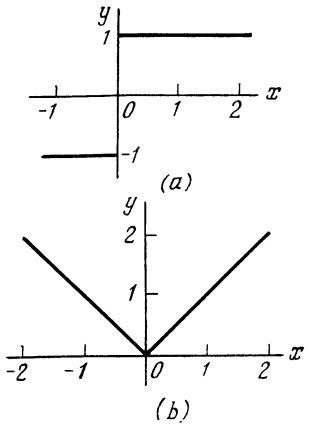


Figure 16.3.4

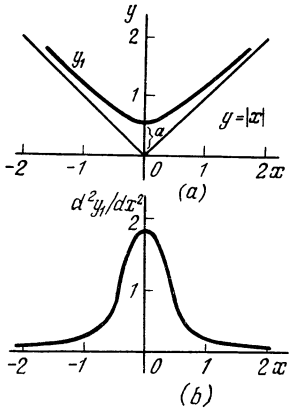


Figure 16.3.5

for $x > 0$. We can also write $\operatorname{sgn} x = x/|x|$, where $|x|$ is the modulus (absolute value) of x. The curve of the signum function, $y = \operatorname{sgn} x$, is shown in Figure 16.3.4*a*. It is easy to see that

$$\operatorname{sgn} x = -1 + 2\theta(x) = -1 + 2\int_{-\infty}^{x} \delta(x)\,dx. \qquad (16.3.5)$$

Now let us examine the function $|x|$ itself. Its graph is depicted in Figure 16.3.4*b*. We find the derivatives to be $d|x|/dx = -1$ for $x < 1$ and $d|x|/dx = +1$ for $x > 1$ or, briefly, with the aid of the new function,

$$\frac{d|x|}{dx} = \operatorname{sgn} x. \qquad (16.3.6)$$

From formula (16.3.5) it then follows that

$$\frac{d^2|x|}{dx^2} = 2\delta(x). \qquad (16.3.7)$$

This formula is so important that one would do well to get a good feeling of it. We know that the second derivative of a function is connected with the curvature of the curve on a graph (see Section 7.9); the curvature of a straight line is zero since the second derivative of a linear function is zero. It would appear then that if the graph of $|x|$ consists of two straight lines, on each of which $d^2|x|/dx^2 = 0$, then why not simply say that the second derivative is equal to zero everywhere in this case? Of course, the crux of the matter lies in the salient point at $x = 0$, where the two straight lines meet. But how are we to be sure that it is at the salient point that the value of the second derivative is $2\delta(x)$? To assure ourselves of this, let us round off the salient point: take the function

$$y_1 = \sqrt{x^2 + a^2}. \qquad (16.3.8)$$

The smaller the value of a, the closer this function is to the broken line $y = |x|$ (Figure 16.3.5*a*). Clearly, at $a = 0$ formula (16.3.8) yields $y_1 = |x|$. The function specified by Eq. (16.3.8) is smooth and differentiable everywhere. Employing the rules of Chapter 4, we can easily find that

$$\frac{dy_1}{dx} = \frac{x}{\sqrt{x^2+a^2}}, \quad \frac{d^2y_1}{dx^2} = \frac{a^2}{(x^2+a^2)^{3/2}}. \qquad (16.3.9)$$

Figure 16.3.5*b* depicts the graph of d^2y_1/dx^2 (both Figure 16.3.5*a* and Figure 16.3.5*b* are constructed for $a = 0.5$). It is easy to see that this curve becomes higher and narrower as a decreases; the integral $\int (d^2y_1/dx^2)\,dx$ is equal to 2 for all values of a (see Exercise 16.3.3); whence, in the limit when $a = 0$ we obtain the expression $d^2|x|/dx^2 = 2\delta(x)$ given above.

Exercises

16.3.1. Draw the graphs of the following (discontinuous) functions and write down the (first) derivatives of the same functions: (a) $y = x$ if $x < 1$ and $y = x - 1$ if $x > 1$; (b) $y = e^{1/x}/(1 + e^{1/x})$.

16.3.2. Find the curvature of the curve $y_1 = y_1(x)$ representing function (16.3.8). How does the curvature change when we send a to zero?

16.3.3. Prove that $\int_{-\infty}^{\infty} y_1''(x)\,dx = 2$ for the function in (16.3.8).

16.4 Representing the Delta Function by Formulas

At the end of the previous section we inadvertently obtained an expression

$$y = \frac{a^2}{2(x^2+a^2)^{3/2}}, \qquad (16.4.1)$$

which approaches $\delta(x)$ in the limit as a tends to zero.[16.8] Let us examine in detail the problem of an analytic expression for the delta function.

Let us take a function $\varphi(x)$ of which we only demand that it vanish for $x = \pm\infty$ (that is, we require that $\varphi(x) \to 0$ as $x \to \pm\infty$) and that the integral $I = \int \varphi(x)\,dx$ be nonzero. It is always possible to make this integral equal to unity by multiplying φ by an appropriate constant. Suppose that this has already been done, so that $\int \varphi(x)\,dx = 1$. It is clear that $\varphi(x)$ has a maximum somewhere between $-\infty$ and $+\infty$. The simplest examples are functions that are everywhere positive and even, that is, symmetric about the y axis (see Section 1.7). Here are some concrete instances:

$$\varphi_1(x) = \frac{1}{2(1+x^2)^{3/2}}, \quad \varphi_2(x) = \frac{1}{\pi}\frac{1}{1+x^2},$$

$$\varphi_3(x) = \frac{1}{\sqrt{\pi}}e^{-x^2}.$$

The graph of each of these functions is bell-shaped (Figure 16.4.1). If these

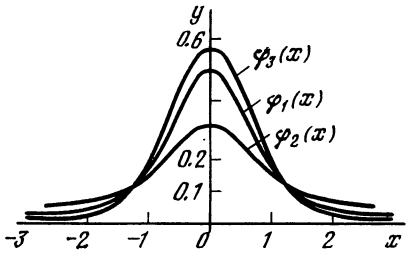


Figure 16.4.1

graphs are brought to the same height (see below), it is hard to distinguish them at a glance. Incidentally, the last (exponential) function is much closer than the others to the axis of abscissas for large values of x far away from the maximum.[16.9] Now recall what needs to be done to increase the height of the bell n-fold and decrease the width m-fold: take $n\varphi(mx)$. If the area under the bell is to be preserved choose $n = m$. To summarize, then, the function $n\varphi(nx) \to \delta(x)$ as $n \to \infty$ or, to put it otherwise, $\delta(x) = \lim_{n\to\infty} n\varphi(nx)$. It is also easy to formally verify that

$$\int n\varphi(nx)\,dx = \int \varphi(z)\,dz = \int \varphi(x)\,dx = 1$$

via the substitution $z = nx$.

Thus, to the three variants of $\varphi(x)$ correspond the following three representations of the delta function:

$$\delta(x) = \lim_{n\to\infty} \frac{n}{2(1+n^2x^2)^{3/2}},$$

$$\delta(x) = \lim_{n\to\infty} \frac{n}{\pi(1+n^2x^2)}, \qquad (16.4.2)$$

$$\delta(x) = \lim_{n\to\infty} \frac{n}{\sqrt{\pi}}e^{-n^2x^2}.$$

Let us verify that the procedure that was proposed earlier,

$$\delta(x) = \lim_{a\to 0} \frac{a^2}{2(x^2+a^2)^{3/2}},$$

[16.8] Note the "2" in the denominator of (16.4.1); without it we would have $2\delta(x)$.

[16.9] The intersection of all three curves φ_1, φ_2, and φ_3 at just about the same points in Figure 16.4.1 is of course purely accidental.

fits this definition. To do this, we write

$$\frac{a^2}{2(x^2+a^2)^{3/2}} = \frac{a^2}{2a^3(x^2/a^2+1)^{3/2}} = \frac{1}{2a(x^2/a^2+1)^{3/2}}$$

and set $n = 1/a$ to get the first representation of the delta function in accord with (16.4.2).

We conclude that there is no single simple formula that can yield $\delta(x)$. Clearly, the fact that $\delta(0) = \infty$ is not enough. To define $\delta(x)$, it still remains to be demonstrated that this is precisely the infinity that is needed. However, $\delta(x)$ can be obtained as the result of passage to the limit ($n \to \infty$) from well-behaved (well-defined) functions of x that involve the auxiliary quantity n as a parameter. We must stress particularly here that $\delta(x)$ may be obtained by such a limit process from *different* functions φ. As long as n is finite, the functions $n\varphi(nx)$ differ from one another and, in particular,

$$I = \int f(x)\, n\varphi(nx)\, dx \neq f(0). \qquad (16.4.3)$$

And only in the limit, as $n \to \infty$, do all the distinct functions $n\varphi(nx)$ tend to a single limit $\delta(x)$ and the corresponding integrals[16.10] (16.4.3) tend to $f(0)$:

$$\lim_{n\to\infty} I = f(0). \qquad (16.4.4)$$

The arbitrariness that is evident in the choice of the original function $\varphi(x)$ from which we obtain $\delta(x)$ is in full accord with the essence of the matter. In Section 17.4, we will consider examples of the application of $\delta(x)$ to physics. The description of some kind of action, that is to say, some kind of finite function $\psi(x)$, with the aid of the delta function is possible and desirable precisely when the detailed form of the action (which is to say, the true dependene of it on x) is inessential, the important thing being only the integral.

The examples given above do not exhaust by any means the diverse $\varphi(x)$ from which we can "manufacture" $\delta(x)$. We can give up the symmetry of $\varphi(x)$: as we pass to $n\varphi(nx)$ and increase n, the distance of the maximum from $x = 0$ diminishes, that is, even an asymmetric function approaches $\delta(x)$. Here is an example: the function $\pi^{-1/2}e^{-(x-1)^2}$ passes into the function $\pi^{-1/2}ne^{-(nx-1)^2}$ whose maximum lies at $x = 1/n$.

We can give up the notation of $\varphi(x)$ by means of a simple unified formula that ensures the smoothness of $\varphi(x)$. For instance, we can take the function $\varphi(x)$ to be discontinuous, say,

$$\varphi(x) = \begin{cases} 1/2 & \text{for } -1 < x < 1, \\ 0 & \text{for } x < -1 \text{ and } x > 1. \end{cases} \qquad (16.4.5)$$

The limit process consists in our taking

$\varphi_n(x) = n/2$ for $-1 < nx < 1$,
i.e., $-1/n < x < 1/n$,

$\varphi_n(x) = 0$ for $x < -1/n$ and $x > 1/n$ (16.4.6)

and sending n to infinity. (Sketch the graph of $\varphi(x)$ according to (16.4.5) and also $\varphi_n(x)$ according to (16.4.6) for $n = 3$ and $n = 10$.)

Finally, we can reject the condition that $\varphi(x)$ be positive. A curious and important example is

$$R(x) = \frac{1}{2\pi}\int_{-\omega}^{+\omega} \cos\xi x\, d\xi = \frac{1}{\pi}\,\frac{\sin \omega x}{x}. \qquad (16.4.7)$$

The graph of $R(x)$ for a given ω is shown in Figure 16.4.2. The value of $R(0)$ is equal to ω/π (the indeterminate form involved in the vanishing, simultaneously, of the numerator and the denominator at $x = 0$ is evaluated

[16.10] Strictly speaking, if $f(x)$ is discontinuous at certain values of x and tends to infinity at these points (or $f(x) \to \infty$ as $x \to \infty$ or as $x \to -\infty$), not all $\varphi(x)$ can be used to obtain (16.4.4). Furthermore, the function $f(x)$ must not be discontinuous or at least its discontinuities must not fall on the point $x = 0$ where $\delta(x) = \infty$, since otherwise we will have those meaningless questions about the value of a function at a point of discontinuity.

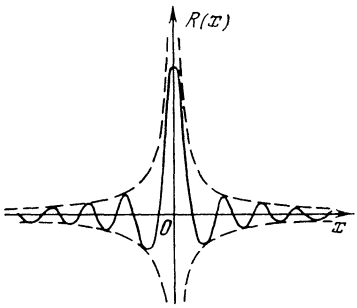


Figure 16.4.2

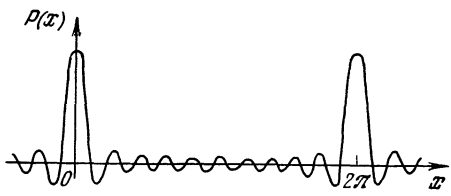


Figure 16.4.3

in elementary fashion). The function $R(x)$ passes through zero and changes sign at $x = \pm\pi/\omega, \pm 2\pi/\omega, \pm 3\pi/\omega, \ldots$. The oscillations of $R(x)$ damp out as they recede from $x = 0$ due to the denominator. The curve does not go beyond the curves $y = \pm 1/\pi \mid x \mid$ shown dashed in Figure 16.4.2. It can be verified that $\int_{-\infty}^{+\infty} R(x)\,dx = 1$ for all values of ω.

It turns out that we can regard $R(x)$ as a delta function if we send ω to infinity. This is likely since as ω is increased, the altitude ω/π of the principal maximum on the R axis grows, while the width of the half-wave $-\pi/\omega < x < \pi/\omega$ decreases. But how are we to deal with the fact that the amplitude of oscillations does not decrease when ω increases; as before, R attains $\pm 1/\pi \mid x \mid$ and the dashed lines do not move toward each other? Let us consider the integral $\int f(x)\,R(x)\,dx$. The greater the value of ω, the more frequent the oscillations and the more exactly the positive and negative half-waves compensate each other, yet the contribution of the first half-wave and the ones closest to it are all the time the same. It is for this reason (we do not give the proof) that $\lim_{\omega\to\infty} \int f(x)\,R(x)\,dx = f(0)$, but this means that $\lim_{\omega\to\infty} R(x)$ has the properties of the delta function.

Of interest is a similar function:

$$P(x) = \frac{1}{\pi}\left(\frac{1}{2} + \sum_{k=1}^{k=q} \cos kx\right). \qquad (16.4.8)$$

Using the formulas of elementary trigonometry, we can obtain the expression

$$P(x) = \frac{1}{2\pi}\,\frac{\sin\left(q+\frac{1}{2}\right)x}{\sin\frac{x}{2}} \qquad (16.4.8a)$$

(see Exercise 16.4.1). The graph of $P(x)$ for $q = 10$ is shown in Figure 16.4.3. As $q \to \infty$, the function $P(x)$ behaves near $x = 0$ just like $R(x)$ does as $\omega \to \infty$, but differs from $R(x)$ in that its high maxima repeat periodically at $x = 0, \pm 2\pi, \pm 4\pi, \ldots$. In other words, $P(x)$ represents, as $q \to \infty$, a *sum* of delta functions:

$$P(x) = \delta(x) + \delta(x - 2\pi) + \delta(x + 2\pi) + \delta(x - 4\pi) + \delta(x + 4\pi) + \ldots.$$

The functions R and P and their connection with the delta function are not mathematical oddities. Recall how R and P were constructed: R is an integral of the cosine and P is a sum of cosines. If from cosines it is possible by addition (integration is a kind of addition) to construct the delta function, then $\delta(x - a)$ can be constructed from $\cos\omega(x - a) = \cos\omega x \cos\omega a + \sin\omega x \sin\omega a$, that is, from cosines and sines with constant coefficients. But then any function $f(x)$ can be represented as a sum of cosines and sines. Any function can be replaced by a series of steps $f(x_i)\,\Delta x_i$, and each step is actually $\delta(x - x_i)\,f(x_i)\,\Delta x_i$. Thus, with the aid of R and P, that is, essentially via delta functions, the possibility is proved of expanding functions in a ***Fourier series*** (if the function is periodic; cf. Section 10.9) and into a ***Fourier integral*** (if the function is nonperiodic).

Reread Sections 10.8 and 10.9 after finishing with this section. Ordinarily, mathematics textbooks do not mention the delta function. Many mathematicians prefer to keep such "physical heresy" away from the student as long as possible. The realization that actually the delta function is being used in the proofs will help you to grasp the meaning of these proofs.

Exercise

16.4.1. Prove that the right-hand sides of (16.4.8) and of (16.4.8a) coincide.

Chapter 17 Applying Functions of a Complex Variable and the Delta Function

17.1 Complex Numbers and Mechanical Oscillations

The third part of the book is just intended to excite the reader's curiosity in the functions that seem to be unlike the ordinary functions, such that the functions of the complex variable and the delta function, and to show the possibility, in principle, of extending the techniques of higher mathematics to these unusual objects as well. This last chapter of the book outlines some possible ways of applying functions described in Chapters 14, 15, and 16 to some problems of physics and engineering.

We will show how calculations can be simplified by using complex variables using an example of *forced oscillations* of a heavy body (see Chapter 10, especially Section 10.5). We start with the equation of such oscillations. Suppose a body of mass m (the material point of mass m since we neglect the dimensions of the body) is acted on by an elastic force $-kx$ which is proportional to deviation x (to make it more clear imagine a spring trying to return the body to equilibrium), a friction force $-h\,(dx/dt)$ which is proportional to velocity dx/dt (and directed in opposition to the velocity), and an external force $F = F(t)$. Here equation (9.4.2) for the body's motion (Newton's second law, see Section 9.4) assumes the form

$$m\frac{d^2x}{dt^2} = -kx - h\frac{dx}{dt} + F,$$

or

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + kx = F. \qquad (17.1.1)$$

We confine ourselves mainly to the case of a periodic external force $F = f\cos\omega t$ (compare this with what is said below at the end of this section). For a periodic force (with *frequency* ω and *amplitude* f) equation (17.1.1) takes the form

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + kx = f\cos\omega t \qquad (17.1.2)$$

(see (10.5.1)); here m, k, h, f, and ω are constants.[17.1]

We have considered all the quantities in (17.1.1) and (17.1.2) to be real-valued. Let us allow for the time being (since in the final solution we have to eliminate this assumption again) *complex-valued* solutions x to our equation; and let all the remaining quantities, which are the coefficients in (17.1.2), to be as before real-valued. It is more convenient now to replace the expression $f\cos\omega t$ for the external force with

$$fe^{i\omega t}\ (= f\cos\omega t + if\sin\omega t) \qquad (17.1.3)$$

adding to the right-hand side of (17.1.2) the imaginary term $if\sin\omega t$.

We find the solution to the equation

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + kx = fe^{i\omega t} \qquad (17.1.2a)$$

in the form

$$x = ae^{i\omega t}. \qquad (17.1.4)$$

Such a form of the solution is suggested by the fact that all the derivatives of an exponential function are as we know proportional to that function; therefore after substituting this expression for x in the left- and right-hand sides of (17.1.2a) both its sides will contain the same factor $e^{i\omega t}$ which we can simply cancel. Indeed, since by (17.1.4) we

---

[17.1] Since m has the dimensions of kg and x, dx/dt, and d^2x/dt^2 the dimensions of m, m/s, and m/s^2 respectively ($\cos\omega t$ is of course dimensionless), it is clear that the coefficients k, h, and f in equation (17.1.2) have the dimensions of kg/s^2, kg/s, and kg·m/s^2 (i.e. N, since f is the force).

have $dx/dt = ia\omega e^{i\omega t}$ and $d^2x/dt^2 = -a\omega^2 e^{i\omega t}$, equation (17.1.2a) yields

$$-am\omega^2 e^{i\omega t} + iha\omega e^{i\omega t} + ake^{i\omega t} = fe^{i\omega t},$$

or

$$a(-m\omega^2 + k + ih\omega) = f,$$

that is,

$$a = \frac{f}{-m\omega^2 + k + ih\omega}. \qquad (17.1.5)$$

Thus we have found that (complex) amplitude a of oscillations $x = ae^{i\omega t}$ ($= a(\cos \omega t + i \sin \omega t)$) is proportional to the applied force f, which is quite natural for a linear system (for the solution of the linear equation (17.1.2a)). When the friction force is not large (i.e. when h is not large), the maximum of the amplitude (the amplitude's modulus, to be more exact) is attained at a frequency ω such that $m\omega^2 \simeq k$, that is, at a frequency of the external force close to the natural frequency $\omega_0 = \sqrt{k/m}$ of the system (cf. Section 10.2); thus, the phenomenon of *resonance* occurs here (Section 10.5).

Expression (17.1.5) for the complex-valued amplitude a can be rewritten as

$$a = fre^{i\varphi}, \qquad (17.1.5a)$$

where

$$re^{i\varphi} = \frac{1}{-m\omega^2 + k + i\omega h},$$

that is,

$$r = \frac{1}{\sqrt{(-m\omega^2 + k)^2 + h^2\omega^2}}, \qquad \varphi = \arctan\left(\frac{h\omega}{m\omega^2 - k}\right). \qquad (17.1.5b)$$

Then,

$$x = ae^{i\omega t} = fre^{i(\omega t + \varphi)}, \qquad (17.1.6)$$

where r and φ are given by formula (17.1.5b). Thus, the concise complex notation (17.1.4) of the real solution to equation (17.1.2) gives both the amplitude $|x| = fr$ and the frequency ω and phase φ of the solution, that is, the phase shift of the oscillation with respect to the phase of the force assumed to be zero (see (17.1.5b)).

Let us now return to real variables. Since the initial equation (17.1.1) is linear, the *principle of superposition* is valid for the mechanical system which the equation describes, that is, the motion caused by a composite force $F_1 + F_2$ can be obtained by adding the motions produced separately by forces F_1 and F_2 (we have frequently used this principle before). In our case the force (17.1.3) is composed of the real component $f \cos \omega t$ and imaginary component $if \sin \omega t$. But all the coefficients in (17.1.1), except for the free term F (the term which does not contain the unknown function x), are real; therefore, to a real force F there must correspond only real value x and to a (pure) imaginary force F an imaginary x. In other words, our linear system generates a real response to a real external force, and an imaginary response to an imaginary force. Therefore we can be sure that to the external force $F = f \cos \omega t$ there corresponds the solution

$$x = fr \cos(\omega t + \varphi) \qquad (17.1.7)$$

of equation (17.1.2), where r and φ are given by (17.1.5b), and to the force $F = if \sin \omega t$ there corresponds the solution $x = ifr \sin(\omega t + \varphi)$ (i.e. to the force $F = f \sin \omega t$ of equation (17.1.1) there corresponds the solution $x = fr \sin(\omega t + \varphi)$, which, by the way, supplies us with no new information, since this solution can be obtained from (17.1.7) by substituting $t + \pi/2\omega$ for t).

The same result can be obtained somewhat differently. By Euler's formulas (14.3.4) the real force $F = f \cos \omega t$ is the sum of (complex-conjugate) forces $F_1 = (1/2) fe^{i\omega t}$ and $F_2 = (1/2) fe^{-i\omega t}$; here due to the fact that equation (17.1.1) is linear (by the superposition principle) the motion brought about by the sum $F_1 + F_2$ of the forces is the sum of the motions produced by each of them separately. But we have already seen that the force F_1 generates the motion $x = (1/2) fre^{i(\omega t + \varphi)}$, where r and φ are given by (17.1.5b). Similarly,

we can find the solution corresponding to the force F_2 and then the sum of the two solutions (see Exercise 17.1.1).

Note finally some properties of the solution obtained which are even retained in more general cases where the oscillatory system at hand consists of a number of masses, springs ("returning forces"), and friction surfaces (a number of friction forces).

1. The imaginary part of the "complex amplitude" (17.1.5a) is negative and equals $-((f\omega h)i)/r^2$. (Why?) This corresponds to a negative angle φ (or to the angle φ within the limits $\pi < \varphi < 2\pi$).

The negative nature of the imaginary part a is connected with the work performed by the force F on the average during many oscillations or exactly in one cycle. If an (external) force is given by the equation $F = f \cos \omega t$ and the solution has the form (17.1.7), then the work of the force is

$$\int F\,dx = \int F \frac{dx}{dt}\,dt$$

$$= \int f \cos(\omega t)\,[-\omega f r \sin(\omega t + \varphi)]\,dt$$

$$= -f^2 r\omega \int \sin(\omega t + \varphi)\cos(\omega t)\,dt$$

$$= -\frac{1}{2} f^2 r\omega \int [\sin(2\omega t + \varphi) + \sin\varphi]\,dt$$

$$= \frac{1}{2} f^2 r\omega \left[-\int \sin\varphi\,dt \right.$$

$$\left. - \int \sin(2\omega t + \varphi)\,dt \right]. \qquad (17.1.8)$$

Clearly, the second integral on the right-hand side of (17.1.8) extended over the time τ, which corresponds to one cycle of oscillations ($\tau = T = 2\pi/\omega$), vanishes, so that the work is equal to $-(1/2)\, f^2\, \omega\, (\sin\varphi)\, \tau$. But this work must be *positive* since during the considered time the system returns to the initial position corresponding to the same values of both the kinetic and the potential energy; therefore the work of the force F will not be expended on the displacement of the system and will be spent to equalize the friction force, which is always negative: the friction force is opposite to velocity dx/dt and hence to the displacement dx.

Note that during the time differing from the full cycle of oscillations, for instance, during the short time which is much less than T, the work of force F can be both positive and negative, since during this short period the energy of the body (the sum of its potential and kinetic energies) can decrease. We can be sure of the sign of work only if we consider a very long period of time of functioning of the system (or the time of one cycle—a large period of time is made up of many similar periods (cycles)).

2. The quantity a (see (17.1.5)-(17.1.5b)) can be considered as a function of the *complex-valued* frequency ω. Here a becomes infinity for $\omega = \omega_r$, where ω_r is the resonance frequency and equals $\lambda + i\mu$ with the imaginary part necessarily positive, i.e. $\mu > 0$. That a is infinite, i.e. to be more exact that the ratio a/f is infinite, means that $f/a = 0$, i.e. that at the frequency ω_r *free* oscillations may set up without the action of any external force F. But in this case the solution $x(t) = ae^{i\omega_r t}$ satisfies the equation

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + kx = 0, \qquad (17.1.2b)$$

which is obtained from (17.1.1) by substituting $F = 0$, whence it follows

$$-m\omega_r^2 + ih\omega_r + k = 0 \qquad (17.1.9)$$

(cf. the denominator in (17.1.5) for a). Clearly, such free oscillations (when friction is present) can only be *damped*. Therefore in the solution $e^{i\omega_r t} = e^{i(\lambda+i\mu)t} = e^{(-\mu+i\lambda)t}$ the quantity $-\mu$ must be negative, i.e. μ must be positive.

In our case the quadratic equation (17.1.9) yields

$$\omega_r = \frac{ih}{2m} \pm \sqrt{\left(\frac{ih}{2m}\right)^2 + \frac{k}{m}}$$

$$= \pm \sqrt{\frac{k}{m} - \left(\frac{h}{2m}\right)^2} + \frac{ih}{2m}.$$

If here $k/m - h^2/4m^2 = \varkappa^2 > 0$, then we have two solutions $\omega_{r1} = \varkappa + ih/2m$ and $\omega_{r2} = -\varkappa + ih/2m$ with the same (positive) imaginary part $ih/2m$. (These solutions are almost complex-conjugate, more exactly $\omega_{r1} = -\omega_{r2}^*$ and $\omega_{r2} = -\omega_{r1}^*$.) If $k/m - h^2/4m^2 < 0$, i.e. $h^2/4m^2 - k/m = \varkappa_1^2 > 0$, then the solutions have the form $\omega_{r1} = (h/2m + \varkappa_1)\, i$ and $\omega_{r2} = (h/2m - \varkappa_1)\, i$ which are two pure imaginary numbers having positive coefficients of i, since $\varkappa_1 < h/2m$.

The same techniques based on using complex variable x can also be applied in the more general case of an arbitrary, and not only sinusoidal, force F. In the case of an arbitrary periodic force F we can just expand the force in a *Fourier series* (see Section 10.9) and then use the superposition principle considering separately the effect produced by each sinusoidal force (17.1.3); here the complex form of a Fourier series, that is, representation of an arbitrary periodic function $F(t)$ with pe-

riod T in the form of a linear set of functions $e^{ni\omega t}$, where $\omega = 2\pi/T$ (fundamental frequency; cf. (10.2.5)) and $n = \ldots, -2, -1, 0, 1, 2,$ is an integer, that is, in the form of the sum of functions $e^{in\omega t}$ taken with some coefficients f_n. In the case of a *nonperiodic* force F it must be expanded not in a Fourier series but in the so-called *Fourier integral* extended over all possible values of frequency ω (and not over frequencies $\omega_n = n\omega$ that are multiples of the fundamental frequency ω as in the case of the Fourier series); here the most convenient form of the Fourier integral is also its complex form.

Without dwelling on the details of the corresponding constructions (see, however, Exercise 17.1.2) we note that another approach is possible to the problem of oscillations generated by an arbitrary force $F = F(t)$: by decomposing force F into separate delta-like components which correspond to a narrow strip of the graph of function F, that is, the replacement of this function with the sum of functions $F_\tau(t)$, where

$$F_\tau(t) = \begin{cases} F(t) & \text{for } \tau \leqslant t \leqslant \tau + d\tau, \\ 0 & \text{for all different } t, \end{cases} \tag{17.1.10}$$

and then using the principle of superposition. (The solution corresponding to the delta-like force $F_\tau(t)$ is termed the ***Green function*** of the respective differential equation; in this connection, see Section 17.4.) Here we arrive at the solution of equation (17.1.1) having, in a more simple case when friction forces are absent (when $h = 0$), the following form

$$x(t) = A \int_{-\infty}^{t} e^{i\omega(t-\tau)} F(\tau)\, d\tau, \tag{17.1.11}$$

where A and ω are defined by the coefficients m and k of equation (17.1.1), or in the "real" form

$$x(t) = A \int_{-\infty}^{t} \sin[\omega(t-\tau)]\, F(\tau)\, d\tau, \tag{17.1.11a}$$

where $\omega = \sqrt{k/m}$ (cf. Section 10.2) and $A = \sqrt{km}$.

For the derivation of formula (17.1.11) or (17.1.11a) see Section 17.4; in particular, compare (17.1.11a) with (17.4.5a). Without considering the reasoning used in obtaining formula (17.1.11a) it is also easy to verify directly that it gives the correct solution to equation (17.1.1) for the case $h = 0$. Indeed, the function on the right-hand side of (17.1.11a) depends on t in two ways: first, t is the upper limit of the integral and, second, the integrand depends on t (in the expression $\sin[\omega(t-\tau)]$). Hence the derivative dx/dt must involve two terms: the first term is the result of differentiation of the integral with respect to the upper limit, and the second term is obtained in differentiating the integrand with respect to t.[17.2] But the derivative of the integral with respect to the upper limit is equal, by the Newton-Leibniz theorem, to the value of the integrand at the upper limit (see Section 3.3); thus we have

$$\frac{dx}{dt} = A \sin 0 \cdot F(t) + A \int_{-\infty}^{t} \omega \cos[\omega(t-\tau)]\, F(\tau)\, d\tau$$

$$= A\omega \int_{-\infty}^{t} \cos[\omega(t-\tau)]\, F(\tau)\, d\tau \tag{17.1.12}$$

and consequently

[17.2] It follows from the fact that if $y = F(x) = \int_a^x f(x, \tau)\, d\tau$, then

$$\Delta y = F(x+\Delta x) - F(x)$$

$$= \int_a^{x+\Delta x} f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x, \tau)\, d\tau$$

$$= \left[\int_a^{x+\Delta x} f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x+\Delta x, \tau)\, d\tau\right]$$

$$+ \left[\int_a^x f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x, \tau)\, d\tau\right].$$

Thus, the ratio $\Delta y/\Delta x$ decomposes into the sum of two fractions, the limit of the first one is the derivative of the integral $\int_a^x f(x+\Delta x, \tau)\, d\tau$ with respect to the upper limit and equals the function $f(x+\Delta x, \tau)$ at $\tau = x$

$$\frac{d^2x}{dt^2} = A\omega \cos 0 \cdot F(t) + A\omega \int_{-\infty}^{t} -\omega \sin[\omega(t-\tau)] F(\tau)\, d\tau$$

$$= A\omega F(t) - A\omega^2 \int_{-\infty}^{t} \sin[\omega(t-\tau)] F(\tau)\, d\tau$$

$$= A\omega F(t) - \omega^2 x(t). \qquad (17.1.13)$$

Therefore, if $\omega^2 = k/m$, i.e. $\omega = \sqrt{k/m}$, and $A\omega = 1/m$, i.e. $A = 1/\omega m = 1/\sqrt{km}$, then

$$-\frac{d^2x}{dt^2} = -\frac{k}{m}x + \frac{1}{m}F(t),$$

which is equation (17.1.1) at $h = 0$.

Finally we note that the same methods of "complex analysis" can be applied to more complicated problems, say, to those related to the establishment of oscillations, where $x = 0$ at $t < t_0$ and $x \neq 0$ only at $t > t_0$, or to the case of several oscillating bodies when in the system there act a number of "returning forces" ("springs") and there are many friction surfaces (many forces of friction are present which are proportional to derivatives dx/dt). In the theory of electric circuits (see Chapter 13) we also frequently encounter oscillatory processes—in equations related to equation (17.1.1) the role of coefficients m, h, and k is played here by inductance L, resistance R, and the quantity $1/C$, the reciprocal of capacitance C, and the role of the "external force" F is played by the voltage supplied to the circuit (emf). Here as well methods similar to those outlined above (consideration of complex solutions of real differential equations) are used rather widely, e.g. the concept of "imaginary current" is familiar to all those acquainted with electrical engineering.

We omit the discussion of all these questions.

(which transforms to $f(x, x)$ as $\Delta x \to 0$); the second fraction is equal to

$$\frac{1}{\Delta x}\int_a^x [f(x+\Delta x, \tau) - f(x, \tau)]\, d\tau$$

$$= \int_a^x \frac{f(x+\Delta x, \tau) - f(x, \tau)}{\Delta x}\, d\tau$$

and under ordinary conditions its limit is

$$\int_a^x \frac{\partial f(x, \tau)}{\partial x}\, d\tau.$$

Exercises

17.1.1. Find the solution (17.1.7) to equation (17.1.2) as the sum of two solutions to equation (17.1.1) corresponding to the values $F_1 = (1/2) fe^{i\omega t}$ and $F_2 = (1/2) fe^{-i\omega t}$ of force F.

17.1.2. Prove that the (real) function $f(x)$ defined in the interval from $-\pi$ to π can be represented as a sum (the *complex form* of the Fourier series):

$$f(x) = \sum_{-\infty}^{\infty} c_k e^{ikx} \quad (= \ldots + c_{-2}e^{-2ix} + c_{-1}e^{-ix} + c_0 + c_1 e^{ix} + c_2 e^{2ix} + \ldots),$$

with $c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\, e^{-ikt}\, dt$ (here $k = \ldots, -2, -1, 0, 1, 2, \ldots$) and $c_{-k} = c_k^*$; if the values of $f(x)$ are complex numbers (while x is still real), then the same formulas are still valid, with the exception that c_{-k} is not necessarily equal to c_k^*.

17.2 Integrals in the Complex Plane

We consider the definite integral

$$w(z) = \int_a^z f(z)\, dz, \qquad (17.2.1)$$

where the lower limit is a complex number a, while the upper limit z is variable. The integral here is understood in the same way as when we dealt with real variables (see Chapter 3), that is, the right-hand side of (17.2.1) is approximately equal to the sum

$$f(z)_0 \Delta z_0 + f(z_1)\Delta z_1 + \ldots + f(z_{n-1})\Delta z_{n-1} = \sum_{i=0}^{n-1} f(z_i)\Delta z_i,$$

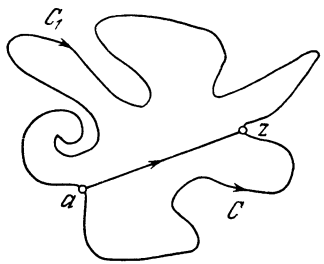


Figure 17.2.1

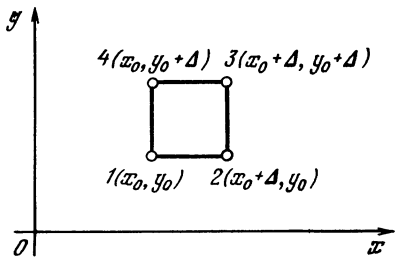


Figure 17.2.2

where $z_0 = a$, z_1, z_2, . . ., z_{n-1}, $z_n = z$ are the points which divide the path C (connecting a and z) of integration into rather small portions, and $\Delta z_i = z_{i+1} - z_i$, $i = 0, 1, \ldots, n-1$.

Let us find the integration path C. Figure 17.2.1 shows three such paths as an example. We can prove, however, that *if within the region bounded by two paths C and C_1 the function has no singularities*, that is, it has sense everywhere and nowhere equals infinity, *then integral* (17.2.1) *taken over these paths has one and the same value.* Of course, when we mention the derivative $f'(z)$ of the function $f(z)$ of a complex variable we mean that the function is considered *analytic.*

The property of integrals (17.2.1) given above admits of another formulation equivalent to the original one. Let us denote by $\oint f(z)\,dz$ the integral extended over a *closed* contour, say, the contour which obtains if we go from point a to point z along path C_1 and then return from point z to point a along path $-C_1$ (path C_1 joins a and z, but we go along the same path in the opposite direction; see Figure 17.2.1). Since (cf. Section 3.2) $\int_{-C_1} f(z)\,dz = -\int_{C_1} f(z)\,dz$, we have

$$\oint f(z)\,dz = \int_C f(z)\,dz + \int_{-C_1} f(z)\,dz = \int_C f(z)\,dz - \int_{C_1} f(z)\,dz,$$

and if $\int_{C_1} f(z)\,dz = \int_C f(z)\,dz$, then $\oint f(z)\,dz = 0$: *the integral* $\oint f(z)\,dz$ *of an analytic function over a closed contour is always zero if inside the contour the function has no singularities.* Conversely, if the integral is equal to zero over any closed contour, it is independent of the path of integration.

In order to prove the above statement, let us first consider a small square with vertices *1* (x_0, y_0), *2* $(x_0 + \Delta, y_0)$, *3* $(x_0 + \Delta, y_0 + \Delta)$, *4* $(x_0, y_0 + \Delta)$ (Figure 17.2.2). We denote by $\oint f(z)\,dz$ the integtal over a closed contour joining consecutively points *1*, *2*, *3*, *4*, and again *1*; we evaluate this integral.

Let $f(z) = f(x + iy) = u(x, y) + iv(x, y)$; we set Δ so small that we can assume that

$$\oint f(z)\,dz = \int_1^2 f(z)\,dz + \int_2^3 f(z)\,dz + \int_3^4 f(z)\,dz + \int_4^1 f(z)\,dz \simeq f(z_1)\,\Delta z_1 + f(z_2)\,\Delta z_2 + f(z_3)\,\Delta z_3 + f(z_4)\,\Delta z_4,$$

where z_j is understood as point j and $\Delta z_j = z_{j+1} - z_j$; here $j = 1, 2, 3, 4$ and $z_{4+1} = z_1$ is again point *1*. Since $\Delta z_1 = z_2 - z_1 = \Delta = -\Delta z_3 = z_3 - z_4$ and $\Delta z_2 = z_3 - z_2 = i\Delta = -\Delta z_4 = z_4 - z_1$, we have

$$\oint f(z)\,dz \simeq \{[u(x_0, y_0) + iv(x_0, y_0)] - [u(x_0 + \Delta, y_0 + \Delta) + iv(x_0 + \Delta, y_0 + \Delta)]\}\,\Delta + \{(u(x_0 + \Delta, y_0) + iv(x_0$$

$$+ \Delta, y_0)] - [u(x_0, y_0 + \Delta) + iv(x_0, y + \Delta)]\} i\Delta = -\{[u(x_0 + \Delta, y_0 + \Delta) - u(x_0, y_0)] + [v(x_0 + \Delta, y_0) - v(x_0, y_0 + \Delta)]\}\Delta - \{v(x_0 + \Delta, y_0 + \Delta) - v(x_0, y_0) + [u(x_0, y_0 + \Delta) - u(x_0 + \Delta, y_0)]\} i\Delta = -A \cdot \Delta - B \cdot i\Delta,$$

where A and B denote the expressions in the braces. But

$$u(x_0 + \Delta, y_0 + \Delta) - u(x_0, y_0) = [u(x_0 + \Delta, y_0 + \Delta) - u(x_0, y_0 + \Delta)] + [u(x_0, y_0 + \Delta) + u(x_0, y_0)] \simeq \left[\frac{\partial u(x_0, y_0 + \Delta)}{\partial x} + \frac{\partial u(x_0, y_0)}{\partial y}\right]\Delta$$

and

$$v(x_0 + \Delta, y_0) - v(x_0, y_0 + \Delta) = [v(x_0 + \Delta, y_0) - v(x_0 + \Delta, y_0 + \Delta)] + [v(x_0 + \Delta, y_0 + \Delta) - v(x_0, y_0 + \Delta)] \simeq \left[-\frac{\partial v(x_0 + \Delta, y_0)}{\partial y} + \frac{\partial v(x_0, y_0 + \Delta)}{\partial x}\right]\Delta,$$

where we, as before, use the approximate formula: $f(x + \Delta) - f(x) \simeq f'(x)\Delta$. Thus,

$$A \simeq \left\{\left[\frac{\partial u(x_0, y_0 + \Delta)}{\partial x} - \frac{\partial v(x_0 + \Delta, y_0)}{\partial y}\right] + \left[\frac{\partial u(x_0, y_0)}{\partial y} + \frac{\partial v(x_0, y_0 + \Delta)}{\partial x}\right]\right\}\Delta \simeq \left\{\left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right] + \left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right]\right\}\Delta = 0$$

by the Cauchy-Riemann relations. (In the last expressions the arguments of the functions $\partial u/\partial x$, $\partial v/\partial y$, $\partial u/\partial y$, $\partial v/\partial x$ can be equated to x_0, y_0, that is, we can consider that the partial derivatives are calculated at the vertices of the square shown in Figure 17.2.2; here the error is of the order of Δ.) Similarly, we can show that B is also equal to zero.

Of course, we have really only established that the integral is *almost* zero, that A and B differ from zero by a quantity of order higher than Δ, and the integral by a quantity of order of smallness higher than the area Δ^2 of the small square (of Δ^3 or Δ^4 or still lower order). And even this implies that the integral is equal to zero.

Indeed, let us subdivide the initial square with side Δ into n^2 squares with side Δ/n. The integral over the contour of the original square is, as can be easily understood, equal to the sum of n^2 similar integrals over the contours of all the small squares (we go along all the contours in the same direction, say, counterclockwise): in fact, when summing up all the integrals we go along each internal line segment two times in two different directions and the sum of these summands will cancel. But while the value of the total integral is of the order of Δ^3, the value of each simple square is of the order of $(\Delta/n)^3$ and the sum of these integrals is of the order of $n^2 (\Delta/n)^3 = \Delta^3/n$. Since this estimate is valid for *any* n, our integral must necessarily vanish.

Similarly, that is, by dividing the region inside a closed contour into small squares we can prove that the integral taken along the contour of this region equals zero, thereby proving that integral (17.2.1) is independent of the path of integration.

Example. Consider the integral

$$I = \int_0^{1+i} z^2\, dz. \qquad (17.2.2)$$

Suppose path A goes from point $z_0 = 0$ to point $z = 1 + i$ first along the real axis (from $z_0 = 0$ to $z_1 = 1$) and then along a straight line which is parallel to the imaginary axis (from $z_1 = 1$ to $z = 1 + i$). Here we have

$$I = \int_A z^2\, dz = \int_0^1 (x^2 - y^2 + 2ixy)|_{y=0}\, dx + \int_0^1 (x^2 - y^2 + 2ixy)|_{x=1}\, i\, dy = -\frac{2}{3} + \frac{2}{3}i.$$

The numerical result is the same if we choose another path B first along the imaginary axis from point $z_0 = 0$ to point $z_2 = i$ and then along the straight line which is parallel to the real axis (from point $z_2 = i$ to point $z = 1 + i$. Verify this). Finally, we can take integrals along diagonal C of the rectangle

(the square here) having vertices at points z_0, z_1, z, z_2: put $z = (1 + i)\,t$, with $0 \leqslant t \leqslant 1$; then $dz = (1 + i)\,dt$ and

$$I = \int_0^{1+i} z^2\,dz = \int_0^1 [(1+i)\,t]^2\,(1+i)\,dt$$

$$= (1+i)^3 \int_0^1 t^2\,dt = \frac{1}{3}(1+i)^3$$

$$= -\frac{2}{3} + \frac{2}{3}\,i.$$

Integral (17.2.2) can be evaluated without resorting to all these simple arithmetic calculations. To this end, as in the case of integration of real-valued functions, it is sufficient to find the *antiderivative* $F(z)$ of function z^2. Let $F(z)$ be the function (of course, analytic) of the complex variable such that

$$\frac{dF}{dz} = f(z). \qquad (17.2.3)$$

Then, by the Newton-Leibniz theorem,

$$\int_a^z f(z)\,dz = F(z) + C,$$

$$I(a,\ b) = \int_a^b f(z)\,dz = F(b) - F(a), \qquad (17.2.4)$$

where a, b, C, $F(a)$, $F(b)$ are complex numbers. Equations (17.2.4) can be proved as in the case of functions of a real variable: we find the change in the integral $I(a,\ z) = \int_a^z f(z)\,dz$ at a small change in the upper limit of integration:

$$I(a,\ z+\Delta z) - I(a,\ z) = \int_z^{z+\Delta z} f(z)\,dz = f(z)\,\Delta z + \ldots \qquad (17.2.5)$$

(we have omitted here the summands of $|\Delta z|^2$ and even smaller orders of magnitude), whence in the limit at small (in absolute value) Δz (irrespective of the increment Δz, that is, of the ratio of its real part Δu to the coefficient Δv of its imaginary part) we obtain

$$\frac{\Delta F(z)}{\Delta z} \simeq f(z). \qquad (17.2.6)$$

From (17.2.6) it follows that the value of integral $I(a,\ z)$ can only differ from the function $F(z)$ by a constant. The value of the constant can be found from the condition $I(a,\ z) = 0$ at $z = a$, whence it follows that $I(a,\ z) = F(z) - F(a)$ in accordance with (17.2.4). But since (see Section 15.2) for complex numbers we also have $(z^3)' = 3z^2$, $\int z^2\,dz = z^3/3 + C$ from which the value of integral (17.2.2) obtained above follows immediately.

We have presented all the reasoning so briefly because there is no difference here between (smooth) functions of the real variable and (analytic) functions of the complex variable—all the results of Chapter 3 are applied to the case of functions of a complex variable.

Now let us consider specific properties of integrals of functions of complex variables which have no analogs in the real region. Consider the integral

$$\int \frac{1}{z}\,dz \quad (= \ln z). \qquad (17.2.7)$$

We take a circle of fixed radius ρ with center at point $z = 0$ as the path of integtation, that is, we set $z = \rho e^{i\varphi}$, where ρ = constant and $0 \leqslant \varphi \leqslant 2\pi$; then $dz = i\rho e^{i\varphi}\,d\varphi$ and $dz/z = id\varphi$ (equality $d(e^{i\varphi}) = ie^{i\varphi}\,d\varphi$ follows from the fact that by the definition of the quantity $e^{i\varphi}$ for small $\Delta\varphi$ we have $e^{i\Delta\varphi} \simeq 1 + i\Delta\varphi$). Therefore, integral (17.2.7) taken along our closed contour is found to be equal to

$$\oint \frac{1}{z}\,dz = \int_0^{2\pi} id\varphi = 2\pi i, \qquad (17.2.8)$$

that is, its value is far from being zero. The value (17.2.8) of the integral is independent of radius ρ of the circle; moreover, it is also independent of the shape of the contour enveloping point 0; we will obtain the same result, $2\pi i$, when integrating along any other closed contour containing 0 inside, say, along the boundary of a square with vertices $\pm 1 \pm i$ (see Exercise 17.2.1).

Why is the main result concerning the equality to zero of integral $\oint f(z)\,dz$ found to be invalid? The reason is that in the given case the integrand $f(z) = 1/z$ has a singularity inside the contour of integration: its value is infinite at point $z = 0$. No matter how small the subregions we take inside the contour, the value $f(z) = \infty$ is inevitably inside one of the subregions; it will make its contribution—murder (and the infinity inside a contour) will out!

Let us examine in detail our integral (17.2.7). This integral taken along the contour surrounding point 0 is equal to $2\pi i$. But when integrating the function $1/z$ along any other path, we cannot also ignore the fact that the integrand has a singular point.

Let

$$I = \int_1^2 \frac{dz}{z} = \ln z \Big|_1^2. \qquad (17.2.9)$$

The value of this integral taken along the shortest path A going from point $z = 1$ to $z = 2$ along the real axis (Figure 17.2.3) is equal to $\ln 2 \simeq 0.69$. But we can also go from point $z = 1$ to point $z = 2$ by another path. For example, let us go from point $z = 1$ along circle C of radius 1 with center at point 0 and only then go along the segment A of the real axis from point 1 to point 2. Here we obtain a different value of integral (17.2.9):

$$\int_1^2 \frac{dz}{z} = \int_C \frac{dz}{z} + \int_A \frac{dz}{z}$$

$$= 2\pi i + \ln 2 \simeq 0.69 + 6.28i.$$

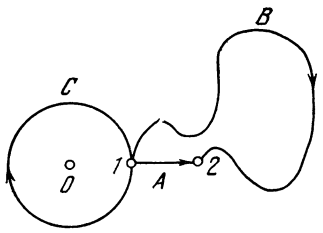


Figure 17.2.3

We can go along circle C not once but n times and only then go along the segment A of the real axis; here the value of integral (17.2.9) will be

$$\int_1^2 \frac{dz}{z} = \int_{nC} \frac{dz}{z} + \int_A \frac{dz}{z}$$

$$= \int_A \frac{dz}{z} + n \int_C \frac{dz}{z} = \ln 2 + 2n\pi i,$$

where nC denotes the contour obtained in an n-fold rotation along circle C in the same direction (counterclockwise). On the other hand we can also go along the same circle C in the opposite direction, clockwise; if we go along this path m times we obtain the value of integral (17.2.9) equal to $\ln 2 - 2m\pi i$. Here the value $\ln 2 \simeq 0.69$ of integral (17.2.9) can be obtained over a multitude of integration paths: all paths which do not go around point 0, like path B shown in Figure 17.2.3; the result $I = \int_1^2 dz/z = \ln 2 - 4\pi i$ can also be obtained over many (other) paths and so on.

Thus, the final solution is

$$I = \int_1^2 \frac{dz}{z} = \ln 2 + i2\pi k,$$

where $k = 0, \pm 1, \pm 2, \ldots$. We may say that this is the most general determination of the natural logarithm of 2, while the number 0.69 is only one of many values of $\ln 2$, this is its prime value. Indeed $e^I = e^{\ln 2 + i2\pi k} = 2e^{i2\pi k} = 2$.

Thus, we have kept our word given in Section 14.3 and explained in a new way the reason for the logarithm $w = \ln z$ being many-valued. The many-valued nature is connected, on the one hand, with the fact that the function $e^{i\varphi}$ is periodic, that is, that the function equals 1 for $\varphi = 2\pi k$. On the other hand, the logarithm can be defined as the integral $\int dz/z$; in such an approach, as well, the fact that the logarithm is many-valued is connected with the behavior of the function $1/z$ at $z = 0$ and that in integrating we can avoid the point $z = 0$ at which the function $f(z) = 1/z$ is infinite (such a point is termed the ***pole*** of the function of the complex variable).

Note that it is the function $1/z$ that, upon integration along the contour which does not surround point $z = 0$, yields a nonzero result, $2\pi i$ (while similar integration of function c/z, where c is a fixed, generally speaking, complex number, gives the result $c \cdot 2\pi i$). Clearly, the integral along a closed curve of function $f(z) = a$, where a is a constant number, or of function $f(z) = az^n$, where $n > 0$ (for example, of function $2z^2$ or $-iz^4$), is zero—in fact, these functions have no singular points (poles). But quite unexpectedly integration of functions that are negative (and greater than unity in absolute value) powers of z, say $1/z^2$ or $1/z^3$, along a closed contour surrounding the pole $z = 0$, in which the function is infinite, also yields zero: for such functions $f(z)$ we again have $\oint f(z)\, dz = 0$. Indeed, substitute into integral $\int dz/z^2$ the values $z = \rho e^{i\varphi}$, $dz = i\rho e^{i\varphi}\, d\varphi$, where ρ = constant and $0 \leqslant \varphi \leqslant 2\pi$; we get

$$\oint \frac{dz}{z^2} = i \int_0^{2\pi} \frac{1}{\rho} e^{-i\varphi}\, d\varphi$$

$$= \frac{i}{\rho(-i)} e^{-i\varphi}\Big|_0^{2\pi} = 0.$$

In the same simple way we can prove the equality $\oint dz/z^n = 0$, where the integer n is greater than unity.

To sum up, the value of the integral along a closed contour depends not just on whether the function is infinite inside the contour, but on the presence or absence in the expression for the function $f(z)$ (in its expansion in series extended over both positive and negative powers of $z - a$, where a is the singular point of the function in which $f(z) = \infty$) of the term which is proportional to $1/(z - a)$ and on the coefficient of this term.[17.3] In view of the fact that the coefficient of $(z - a)^{-1}$ in the expansion of $f(z)$ in powers of $(z - a)$ is so important, it bears a special name, the ***residue*** of $f(z)$ at point $z = a$. Clearly, at any point where the function is analytic, that is, where it assumes a finite value and can be expanded in a Taylor's series (in the neighborhood of such a point), the residue is zero. If we take the function $f(z) = 1/z^2$, which becomes infinite at $z = 0$, it has a zero residue at the same point. In contrast, the function

$$\varphi(z) = \frac{1}{z^2+1} = \frac{1}{(z+i)(z-i)}$$

$$= \frac{1}{2i}\frac{1}{z-i} - \frac{1}{2i}\frac{1}{z+i} \qquad (17.2.10)$$

has two singular points: $z = i$ and $z = -i$. The residue of $\varphi(z)$ at $z = i$ is equal to $1/2i = -i/2$, and the residue at point $z = -i$ is equal to $-1/2i = i/2$. Indeed, if, say, we take the expansion of $(-1/2i)(z + i)^{-1}$, it has no terms with negative powers of $(z - i)$, so that the total coefficient of $(z - i)^{-1}$ here is equal to $1/2i$.

There exists a remarkable application of integrals of functions of a complex variable to the evaluation of integrals of a real variable. Here is a typical and yet important example:

$$I = \int_{-\infty}^{\infty} \frac{\cos \omega x}{1+x^2}\, dx. \qquad (17.2.11)$$

[17.3] We do not consider more complicated cases, where the integrand itself is many-valued (like, say, the function $f(z) = \ln z$).

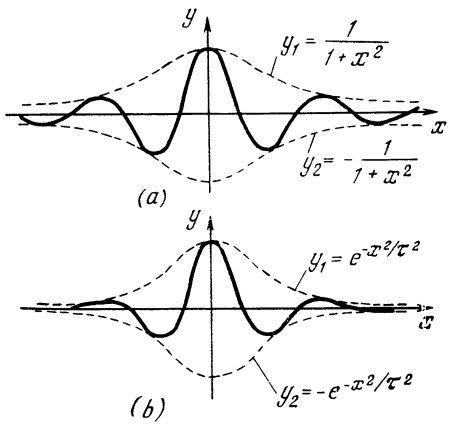


Figure 17.2.4

This is the integral of a product of a rapidly varying function $\cos \omega x$ (where it will be assumed that ω is much higher than unity) by a slowly varying function $y_1 = (1 + x^2)^{-1}$. The curve representing the integrand is depicted in Figure 17.2.4*a*, with the dashed curves being the graphs of the functions $y_1 = (1 + x^2)^{-1}$ and $y_2 = -(1 + x^2)^{-1}$.

According to Euler's formulas (14.3.4),

$$\cos \omega x = \frac{1}{2} e^{i\omega x} + \frac{1}{2} e^{-i\omega x}. \qquad (17.2.12)$$

Substituting this into the right-hand side of (17.2.11), we split I into two integrals, I_+ and I_-, where

$$I_+ = \frac{1}{2} \int_{-\infty}^{\infty} \frac{e^{i\omega z}}{1+z^2}\, dz. \qquad (17.2.13)$$

The second integral I_- can be written in a similar manner, only $e^{i\omega z}$ must be replaced with $e^{-i\omega z}$. The (dummy) variable of integration in (17.2.11) is denoted here not by x but by z, which suggests that it assumes not only real values but also complex values (this idea is central to our discussion).

Suppose that z runs along the semicircle γ: $|z| = R$ in the *upper* half-plane ($z = x + iy$, $y > 0$), with R being very large. Since

$$|f(z)| = \frac{1}{2} \left| \frac{e^{i\omega z}}{1+z^2} \right| = \frac{1}{2} \frac{|e^{i\omega z}|}{|1+z^2|}$$

and along γ we have

$$|e^{i\omega z}| = |e^{i\omega(x+iy)}| = |e^{-\omega y}|\, |e^{i\omega x}|$$
$$= |e^{-\omega y}| \times 1 < 1$$

(since in the upper half-plane y is greater than zero and thus $e^{-\omega y} < 1$), while

$$\left| \frac{1}{1+z^2} \right| \simeq \left| \frac{1}{z^2} \right| = \frac{1}{R^2},$$

we conclude that along γ the absolute value of the function $f(z) = e^{i\omega z}/2 \times (1 + z^2)$ is extremely small; therefore, the integral[17.4] $\int_\gamma f(z)\, dz$ along γ is small, too. This implies that for R large, the integral

$$\frac{1}{2} \oint_\Gamma \frac{e^{i\omega z}}{1+z^2}\, dz, \qquad (17.2.14)$$

where Γ is a closed contour consisting of the line segment $-R \leqslant x \leqslant R$ of the x axis, with R very large, and the semicircle γ in the upper half-plane (Figure 17.2.5), is very close to I_+ along the real axis. But the integral of an analytic function over a closed contour depends, as we know, only on the singular points of $f(z)$ lying inside this contour. But does the function $f(z) = e^{i\omega z}/(1 + z^2) = e^{i\omega z} \varphi(z)$ have any singular points inside our contour?

It is clear that inside Γ the function $e^{i\omega z}$ never becomes infinite and has no

[17.4] We see that it can be assumed that $|f(z)| < R^{-2}$ along γ; this together with the inequality $\left| \int f(z)\, dz \right| \leqslant \int |f(z)|\, |dz|$ (see Exercise 17.2.2), which follows from the very definition of an integral, and the fact that the length of the semicircle γ increases like the first power of R, imply that the integral $\int_\gamma f(z)\, dz$ decreases, as R grows, in any case no slower than R^{-1}.

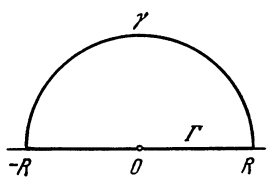


Figure 7.2.5

singular points. The function $\varphi(z) = 1/(1 + z^2)$, as noted earlier, has inside Γ the singular point $z = i$, and the residue of $\varphi(z)$ at this point is $-i/2$. This implies that the function $f(z) = (1/2)\, e^{i\omega z}\varphi(z)$ at its singular point $z = i$ has a residue $-(i/4)\, e^{i\omega i} = -(i/4)\, e^{-\omega}$. By the general properties of analytic functions, the integral (17.2.14) over the huge contour Γ is equal to $\int_\sigma e^{i\omega z}(1 + z^2)^{-1}\, dz$, where σ is a small circle with its center at point $z = i$; this integral is equal to

$$-\frac{ie^{-\omega}}{4}\oint_\sigma \frac{dz}{z-i} = -\frac{i}{4}\, e^{-\omega} 2\pi i$$

$$= \frac{1}{2}\pi e^{-\omega}. \qquad (17.2.15)$$

The same value has the integral (17.2.14), which, hence, is independent of R. For this reason integral (17.2.13) is also equal to $(1/2)\, \pi e^{-\omega}$.

In a similar manner we find that $I_- = (1/2)\, \pi e^{-\omega}$, which is the second term in the integral I. Thus, the final result is

$$I = \pi e^{-\omega}. \qquad (17.2.16)$$

It is extremely difficult to arrive at this formula without using functions of a complex variable. Numerical integration is also very involved since for $\omega \gg 1$ the expression on the right-hand side of (17.2.16) or (17.2.15) is very small. Separate positive and negative half-waves of the periodic function $\cos \omega x$ will not be especially small: their amplitudes, if we are speaking of waves corresponding to moderate values of x, are of the order of unity, which implies that the contribution of a half-wave (of the type we are considering here) to the integral I is of the order of $\pm 1/\omega$. (Why?) On the other hand, for $\omega \gg 1$, the total value of I given by (17.2.15) is much smaller than $1/\omega$. Hence, positive and negative half-waves cancel each other with a high degree of accuracy, which is related to the analyticity, or smoothness, of the "amplitude function" $y_1 = (1 + x^2)^{-1}$.

Integrals similar to (17.2.11) play an important role in the theory of vibrations and in theoretical physics in general (compare with the integral in (17.1.11a)). A slowly varying pulse (whose shape is defined by the function of time $f(t) = (1 + t^2)^{-1}$) does not excite a high-frequency oscillating system because at high frequencies ($\omega \gg 1$) the integral $\int_{-\infty}^{\infty} \frac{\cos \omega t}{1 + t^2}\, dt$ is small. Only the theory of functions of a complex variable makes it possible to establish that this integral falls off exponentially as $\omega \to \infty$.

Exercises

17.2.1. Prove by direct calculation that $\oint_K dz/z = 2\pi i$, where the integration contour K coincides with the boundary of the square whose vertices lie at the points $\pm 1 \pm i$.

17.2.2. Prove the inequality $\left|\int_\Lambda f(z)\, dz\right| \leqslant \int_\Lambda |f(z)|\, |dz|$, where Λ is an arbitrary arc in the complex plane.

17.2.3. Prove that $\int_{-\infty}^{\infty} e^{-t^2/\tau^2} \cos \omega t\, dt = \tau \sqrt{\pi}\, e^{-\omega^2\tau^2}$. (This result is physically important: the so-called Gaussian profile e^{-x^2/τ^2} of a slow pulse (see Figure 17.2.4*b*) has little influence (does not excite) high-frequency oscillating systems, corresponding to high frequencies ω.)

17.3 Analytic Functions of a Complex Variable and Liquid Flow

In this section we conclude the large topic of complex numbers and functions of a complex variable. One can only marvel (we already mentioned this before) at how the introduction of an "imaginary unit" $i = \sqrt{-1}$ not only resolved all the difficulties that appear in solving the various quadratic

equations but also introduced harmony into the entire theory of algebraic equations (or simply algebra) as well as the theory of functions and mathematical analysis. All functions can naturally be generalized to functions of a complex variable: not only polynomials and algebraic functions (only here do the questions concerning the number of zeros of polynomials and the power functions with fractional exponents gain completeness), but also the exponential function, the logarithms, and trigonometric functions. We replace the general definition of a function as correspondence between two sets of variables (even in the real domain we are not able to preserve this definition of a function in all generality) by the definition of an *analytic* function. By restricting ourselves in this way, we gain colossal information about the functions: the entire apparatus of higher mathematics (i.e. the concepts of a derivative, an integral, a differential equation, a power series) can be applied to analytic functions of a complex variable; however, many statements that are true for analytic functions prove to be invalid for functions in general. Of great importance to analytic functions are the points (the values of the variables) where the functions become infinite. It has been established that the behavior of analytic functions at such points determines a broad spectrum of properties of these functions; for instance, the presence of such points enables evaluating integrals of functions over various contours without actually calculating them.

It has been found that by specifying a function of a complex variable on a small segment (as small as desired) we can, at least in principle, determine the function in the entire complex plane, over the entire range of the independent variable. It turns out that analyticity is determined not by the fact that there exists a simple formula for expressing the function in terms of the independent variable z: an analytic function may be expressed, say, by an *integral* that cannot be evaluated in terms of elementary functions (of the type $\int_a^z e^{-z^2/2}\,dz$) or by the *solution of differential equation* that cannot be solved in the traditional sense (i.e. we cannot express the dependence of the unknown function on the independent variable by means of a formula) or by an *infinite series*. What is important is only that the value $w = f(z)$ of the function be dependent on $z = x + iy$ and not on x or y separately; in other words, the function must not break up the natural bond between x and y in their combination $x + iy$.

Functions of a complex variable prove to be a powerful tool for solving mathematical, physical, and engineering problems. In such problems, naturally, the data is given by real quantities. The answer must also be expressed in real values, since "the square root of minus one piece of chalk" cannot be allowed into everyday life. Nevertheless, at some intermediate stage in a problem, somewhere between the formulation of the problem and the answer, it often proves to be useful to introduce complex numbers or quantities and taking (but only temporarily) a complex number of pieces of chalk, so to say (compare with Section 17.1).

For instance, in Section 17.2 we saw that to evaluate $\int_{-\infty}^{\infty} \frac{\cos \omega x}{1+x^2}\,dx$ (which is a purely real integral) it is advisable to write $\cos \omega t$ as $(1/2)\, e^{i\omega x} + (1/2)\, e^{-i\omega x}$ and thus introduce what seem to be quite unnecessary imaginary units. Even if we do not know the precise formulas that express a law of nature, often only the assumption that there exists an analytic function of a complex variable that expresses the law (even a function that cannot be expressed by a formula) is fruitful and informative. The very fact that such a function exists leads to certain relationships generated by the analyticity of the function (a function unknown to us). Such an ap-

proach is widely used at the forefront of modern theoretical physics, say, in what is known as *dispersion relations* (unfortunately, nothing more can be said about these relations here). The greatest miracle is that the imaginary unit and complex numbers have proved to be necessary for quantum theory, the physical theory of the microworld. The place of complex numbers in quantum theory is so important that today in teaching analytic geometry to students of physics and chemistry it is often thought necessary almost at the start to speak of a complex plane and a space where points are specified by complex-valued coordinates, although of course there is no way in which such a space can be visualized.[17.5]

Apparently, the first serious applications of the theory of analytic functions of a complex variable (the very creation of this theory was due probably to these applications; see p. 483) were in *fluid mechanics*, which is the theory of the movement of gases and liquids. Suppose we have a *plane-parallel* flow of an (incompressible) liquid of constant density. The velocity vector $\mathbf{v}$ of any particle of the liquid is at all moments of time directed parallel to the (horizontal) xy-plane and the entire column of the liquid corresponding to given x and y coordinates and different altitudes z moves as a whole (in "one piece"). In this case we can ignore the third dimension and consider the motion as occurring in the xy-plane. Suppose that $\mathbf{v}$ (a vector) of the flow at point $M = M(x, y)$ has components v_x and v_y (Figure 17.3.1); it is clear that $v_x = v_x(x, y)$ and $v_y = v_y(x, y)$ depend on point M, that is, are functions of this point, or functions of the coordinates x and y.[17.6]

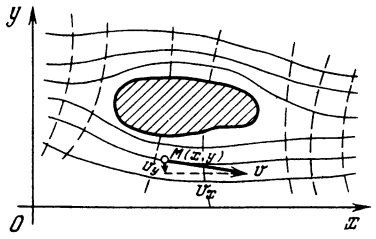


Figure 17.3.1

In many cases it is admissible to assume that there exist functions $\varphi(x, y)$ and $\psi(x, y)$ such that[17.7]

$$v_x = -\frac{\partial \varphi(x, y)}{\partial x}, \quad v_y = -\frac{\partial \varphi(x, y)}{\partial y}, \tag{17.3.1}$$

$$v_x = -\frac{\partial \psi(x, y)}{\partial y}, \quad v_y = \frac{\partial \psi(x, y)}{\partial x}. \tag{17.3.1a}$$

The function φ is known as the ***velocity potential***, and the motion of liquids that obeys the above assumption is called ***potential flow***. It can be said that the flow of a liquid in the vicinity of point $M = M(x, y)$ is potential if and only if it is reduced to the transfer of a mass of liquid in the direction of $\mathbf{v}$ and in the deformation of a small volume of the liquid (which is admissible in view of the fluidity of the liquid) but does not include rotation of the liquid about point M (such rotation is often observed as whirlpools in rivers or even in a bathtub). In accord with this, potential flow is sometimes called ***irrotational*** (or ***vortex-free***) and the points (generally, isolated) where this condition breaks

17.5 E.g. see I.M. Yaglom, *Complex Numbers in Geometry*, Academic Press, New York, 1968.

17.6 For the sake of simplicity we have restricted our discussion to the case of *steady-state* flow, when the velocity does not change in time and the $v_x(x, y)$ and $v_y(x, y)$ are independent of time.

17.7 Since the functions φ and ψ are defined only through the values of their derivatives in (17.3.1) and (17.3.1a), they are similar to potentials, that is, are defined only to within arbitrary summands C_1 and C_2, which enables choosing the zero values of φ and ψ at our will. Only the *difference* of values of φ (or ψ) at two different points but not the values themselves has any physical meaning. (Note also that since v_x and v_y have the dimensions of velocity, m/s, and the coordinates x and y have the dimensions of length, m, the functions φ and ψ defined by (17.3.1) and (17.3.1a) have the dimensions of m^2/s.)

down are known as ***vortexes***. The difference

$$\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}, \tag{17.3.2}$$

which is zero if the flow is potential (why?) and, hence, is nonzero only at vortexes, is known as the strength of the particular vortex, or simply ***vorticity***.

The function $\psi = \psi(x, y)$ is known as the ***stream function***. It can be proved that if at a point $M(x, y)$ there exists a stream function ψ, the amount of liquid flowing into a small contour surrounding point M will be equal to the amount of liquid flowing out of the contour every second (the liquid is assumed incompressible). Thus, the existence of a stream function ψ at a given point $M = M(x, y)$ guarantees that point M is neither a source nor a sink of liquid. The sum

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} \tag{17.3.2a}$$

(which is identically zero if at point $M = M(x, y)$ there exists a stream function ψ) characterizes the *strength* of the source or sink, while the sign of this sum characterizes the nature of the "singular point" M: if (17.3.2a) is positive, we have a *source* at point M (i.e. more liquid flows out of the contour surrounding M than in), while if (17.3.2a) is negative, we have a *sink*. (The sign of the difference (17.3.2) characterizes the *direction in which the liquid rotates* at the vortex.)

The reader can easily see (Exercise 17.3.1) that the curves $\psi = \text{constant}$ are simply the *streamlines*, that is, the curves along which the particles of the liquid flow: along each streamline (the tangent to each such line at each point M has the direction of the velocity vector $\mathbf{v}$) the magnitude of ψ remains the same. *Equipotential curves*, which are specified by the condition $\varphi(x, y) = \text{constant}$, on the other hand, are perpendicular at each point to the velocity vector (see Exercise 17.3.2): the liquid moves across these curves in the direction in which the potential decreases. (In Figure 17.3.1 the equipotential curves are depicted by dashed curves.) Thus, the mesh of curves $\varphi = \text{constant}$ and $\psi = \text{constant}$ provides a complete picture of flow of the liquid as a whole.

Now let us assume that in the neighborhood of a given point $M(x, y)$ in a liquid there are neither vortexes nor sources and sinks; in other words, both functions, φ and ψ, exist. Obviously, from (17.3.1) and (17.3.1a) it follows that

$$\begin{aligned} -v_x &= \frac{\partial \varphi}{\partial x} = \frac{\partial \psi}{\partial y}, \\ -v_y &= \frac{\partial \psi}{\partial y} = -\frac{\partial \psi}{\partial x}. \end{aligned} \tag{17.3.3}$$

Equations (17.3.3) are close in form to the Cauchy-Riemann equations (15.2.6), which connect the real and imaginary parts of an analytic function $w = w(z) = u(x, y) + iv(x, y)$ of a complex variable $z = x + iy$. Thus, if we assume the xy-plane to be the complex z plane, then we naturally arrive at the analytic function

$$w = w(z) = \varphi + i\psi \tag{17.3.4}$$

of complex variable z, a function that is closely linked to liquid flow.

The function (17.3.4) is called the ***complex potential of liquid flow***. It completely characterizes the flow of a liquid, so that the study of various plane-parallel flows is reduced to the study of analytic functions $w(z)$. The derivative of the complex potential,

$$v = \frac{dw}{dz} \quad \left(= \frac{\partial \varphi}{\partial x} + i\,\frac{\partial \psi}{\partial x} = -v_x + iv_y \right) \tag{17.3.5}$$

(see (15.2.3)), is also an analytic function of complex variable z. This function is called the ***complex velocity of flow***. This name suggests a relationship between the complex number v and the velocity $\mathbf{v}$; obviously,

$$\begin{aligned} |v|^2 &= (-v_x)^2 + (v_y)^2 = v_x^2 + v_y^2 \\ &= |\mathbf{v}|^2, \end{aligned}$$

that is, the absolute value of v is equal to the length of $\mathbf{v}$; on the other hand,

Arg v characterizes the direction of vector $\mathbf{v}$ (see Exercise 17.3.3). Flows specified by the complex potentials $w = \varphi + i\psi$ and $iw = -\psi + i\varphi$ can be called ***conjugate***, because the streamlines of one flow coincide with the equipotential curves of the other, and vice versa.

Quite close to the above scheme of liquid flow is the use of analytic functions in the classical *theory of the electromagnetic field.*[17.8] Let us consider a flat electric field. We will assume that within the plane region we are considering there are no free electric charges or variable magnetic fields (the presence of such fields is equivalent to the presence of charges, in accord with the law of induction). Suppose that $\Phi = \Phi(x, y)$ is the *potential* of the electric field at a point $M = M(x, y)$. Then the *electric field strength* at this point is fixed by vector $\mathbf{E} = (E_x, E_y)$ with coordinates

$$E_x = -\frac{\partial \Phi}{\partial x}, \quad E_y = -\frac{\partial \Phi}{\partial y}. \qquad (17.3.6)$$

In addition to potential Φ we can introduce the *streamline of the field*, $\Psi = \Psi(x, y)$, such that[17.9]

$$E_x = -\frac{\partial \Psi}{\partial y}, \quad E_y = \frac{\partial \Psi}{\partial x}. \qquad (17.3.6a)$$

The curves $\Psi =$ constant are the ***lines of force*** of our electric field, while the curves $\Phi =$ constant are the ***equipotential curves.*** By virtue of (17.3.6) and (17.3.6a), the function

$$W = \Phi + i\Psi \qquad (17.3.7)$$

is an analytic function of the complex variable $z = x + iy$; it is called the ***complex potential of the electric field.*** The derivative

$$E = \frac{dW}{dz} \left(= \frac{\partial \Phi}{\partial x} + i \frac{\partial \Psi}{\partial x} = -E_x + iE_y \right) \qquad (17.3.8)$$

of the complex potential can be called the ***complex electric field strength***, since this complex quantity is very close in meaning to the electric field vector $\mathbf{E}$; in particular, $|E| = |\mathbf{E}|$. The electric fields with $W = \Phi + i\Psi$ and $iW = -\Psi + i\Phi$ are sometimes called *conjugate*, the streamlines of one coincide with the equipotential curves of the other, and vice versa.

We will not dwell any further on the various ideas of the mathematical theories of liquid flow and electric field, which have to do with functions of two variables x and y and, respectively, with differential equations involving the partial derivatives of these functions, which are simply known as *partial differential equations* (see Section 10.8). We note only that the theory of analytic functions of a complex variable plays an extremely important role in these purely applied problems.

Exercises

17.3.1. Prove that (a) the curve $\psi =$ constant is a steamline of liquid flow; in other words, that the tangent to this curve at each point $M(x, y)$ of this point coincides in direction with the vector $\mathbf{v}(x, y)$, and (b) if AB is an arbitrary arc (in the plane of liquid flow) with endpoints $A = A(x_1, y_1)$ and $B = B(x_2, y_2)$, then liquid flow across the arc per unit time is equal to the difference $\psi(x_2, y_2) - \psi(x_1, y_1)$ (the increment of the streamline along arc AB).

17.3.2. Prove that the equipotential curves $\varphi =$ constant and the streamlines $\psi =$ constant are perpendicular at each point in the liquid flow (precisely, the tangents at point $M(x, y)$ to the streamline and equipotential curve passing through this point are perpendicular to each other).

17.3.3. Suppose $V = V(x, y)$ is a complex number such that $\overrightarrow{OV} = \mathbf{v}$ (where $\overrightarrow{OV}$ is a vector with the initial point at $O(0, 0)$ and the terminal point at V, and $\mathbf{v}$ is the velocity vector at point $M(x, y)$). Prove that points v and V in the complex plane, where v

---

[17.8] Precisely, here we can speak only of *electrostatics* and *magnetostatics*, since two coordinates (x and y) prove to be insufficient for the general problems involving electromagnetic phenomena—three spatial and one time variable are required.

[17.9] Since Φ and Ψ have the dimensions of voltage, V, and the coordinates x and y have the dimensions of length, m, it is clear that the dimensions of E_x and E_y are those of V/m (we can also say that V/m are the dimensions of $\mathbf{E}$).

is the complex velocity of flow, are symmetric with respect to the imaginary axis (and, hence, Arg $v = \pi - \text{Arg } V$).

17.3.4. Describe (plane-parallel) flow of an (incompressible) liquid for the following complex potentials $W = \varphi + i\psi$: (a) $w = az$, where a is a real or purely imaginary number, (b) $w = az^2$ (a is a real or purely imaginary number), (c) $w = a/z$ (a is a real or purely imaginary number), and (d) $w = \ln z$ and $w = i \ln z$.

17.4 Application of the Delta Function

If Sections 17.1 to 17.3 logically belonged to the material of Chapters 14 and 15, since in these sections we dealt with complex numbers and analytic functions of a complex variable, the present section is a continuation of Chapter 16, which must be looked through before going on. We note also that the first example, which starts this section, is closely related to the content of Section 9.12, while the second example, which deals with Newton's second law (17.4.1), continues the discussion in Section 9.5 about the impulse and the motion of a particle moving under a short impulse, say, a blow. Therefore, before studying these examples it is advisable to go over the appropriate sections of Chapter 9.

First of all, we will show you how the delta function permits abridging and making the writing of the conditions in many problems more convenient.

Let us consider a rod with a variable cross section to which a number of separate point loads of masses m_1, m_2, etc. are attached (Figure 17.4.1). Let the mass per unit length of the rod be expressed by the function $\sigma(x)$. The mass of the rod without the loads is $\int_a^b \sigma(x)\,dx$, while with the loads it is

$$M = \int_a^b \sigma(x)\,dx + \sum_i m_i,$$

where the summation is extended over all the loads attached to the rod. The position of the center of gravity is

$$X = \frac{1}{M}\left(\int_a^b x\sigma(x)\,dx + \sum_i x_i m_i\right).$$

The moment of inertia with respect to the origin is

$$I = \int_a^b x^2\sigma(x)\,dx + \sum_i x_i^2 m_i.$$

But with the aid of the delta function it is possible to include the separate masses in a generalized density function, which we denote by $\eta(x)$. It is defined by the formula

$$\eta(x) = \sigma(x) + \sum_i m_i \delta(x - x_i).$$

Indeed, if we consider the general distribution of mass along the rod, we can say that at the points where the loads are attached the density exhibits infinite jumps. With the aid of the new function $\eta(x)$ all the quantities can be written in uniform fashion and more succinctly:

$$M = \int_a^b \eta(x)\,dx, \quad X = \frac{1}{M}\int_a^b x\eta(x)\,dx,$$

$$I = \int_a^b x^2\eta(x)\,dx.$$

The concept of the delta function permits combining continuously distributed masses and point masses in a single general expression.

Another example of the use of the delta function refers to the motion of a mass point (or material particle). The

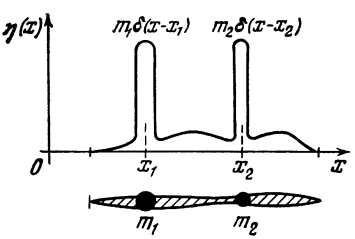


Figure 17.4.1

basic equation, it will be recalled, expresses Newton's second law:

$$m\frac{d^2x}{dt^2} = F(t) \qquad (17.4.1)$$

(see Section 9.4). Recall the arguments in Section 9.5 to the effect that the action of the impulse is independent of the law of variation of the force, provided that the force is sufficiently brief. These considerations are similar to what was said in Section 16.4 to the effect that the delta function can be constructed out of a *variety* of functions $\varphi(x)$ and concerning the conditions when it is possible to replace a finite function $\varphi(x)$ with the generalized, singular function $\delta(x)$.

If the concrete form of the function of the force is not essential in the problem about a blow, this means that $F(t)$ may be replaced by the delta function, $F(t) \to J\delta(t-\tau)$, where τ is the instant of the blow, and $J = \int F(t)\,dt$ is the impulse. We will carry out the integration of the equation of the motion under a unit delta force formally and according to all the rules. Let the particle, prior to the blow, be at rest at the origin: $x = 0$ and $v = dx/dt = 0$ at $t = -\infty$. We will also assume that the (unimportant) factor J in the expression for the generalized force is of the order of unity. Then Eq. (17.4.1) takes the form

$$m\frac{d^2x}{dt^2} = m\frac{dv}{dt} = \delta(t-\tau). \qquad (17.4.2)$$

Integrating (17.4.2) and bearing in mind that $d^2x/dt^2 = dv/dt$, with $v(t)$ the velocity of the particle, we get

$$v(t) = \frac{1}{m}\int_{-\infty}^{t}\delta(t-\tau)\,dt$$

$$= \frac{1}{m}\theta(t-\tau), \qquad (17.4.3)$$

where the function $\theta(x)$ is expressed by formula (16.3.1) (see the graph of this function in Figure 16.3.1). Thus, the velocity is expressed by a step-like

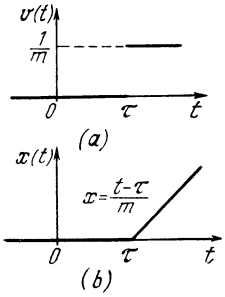


Figure 17.4.2

function of time t (Figure 17.4.2*a*): $v = 0$ for $t < \tau$ and $v = 1/m$ for $t > \tau$.

The next step consists in determining the path. From the fact that $v = dx/dt$ we get, by integrating both sides of Eq. (17.4.3),

$$x = \begin{cases} 0 & \text{for } t < \tau, \\ \frac{1}{m}(t-\tau) & \text{for } t > \tau. \end{cases} \qquad (17.4.4)$$

(Why?) The graph of the path is shown in Figure 17.4.2*b*.

Characteristic of the curve $x(t)$ of the path is the salient point at $t = \tau$. Here again we are convinced that the second derivative of a function having a salient point contains the delta function: the function $x(t)$ has a salient point. According to the equation of motion, the force is proportional to d^2x/dt^2, and $x(t)$ with a salient point was obtained precisely for a force that was proportional to $\delta(t-\tau)$, so that in the case of a salient point, d^2x/dt^2 contains a delta function, which is what we set out to prove.

Now let us take the next step. The problem of the motion of a body under a given force is *linear*. This means that if there are two solutions, $x_1(t)$ and $x_2(t)$, corresponding to two distinct forces, $F_1(t)$ and $F_2(t)$, the sum of the solutions, $x(t) = x_1(t) + x_2(t)$, is a solution that corresponds to the action of the sum of the forces, $F(t) = F_1(t) + F_2(t)$. This property (the ***superposition principle***) is a consequence of the simple fact that the second derivative of a sum of functions is the sum

of the second derivatives of the functions:

$$\frac{d^2x}{dt^2} = \frac{d^2(x_1+x_2)}{dt^2} = \frac{d^2x_1}{dt^2} + \frac{d^2x_2}{dt^2}.$$

Taking into account that $d^2x_1/dt^2 = F_1(t)/m$ and $d^2x_2/dt^2 = F_2(t)/m$, we get

$$\frac{d^2x}{dt^2} = \frac{F_1(t)}{m} + \frac{F_2(t)}{m} = \frac{F(t)}{m},$$

which is what we set out to prove—that the sum of the solutions, $x_1 + x_2$, describes the motion produced by the sum of the forces.

Only one reservation is in order: solution of the equation of motion depends not only on the law of force but also on the initial conditions, that is, on the initial position and velocity of the given mass. If we choose these conditions thus: $x_1 = 0$ and $dx_1/dt = 0$ at $t = -\infty$, and $x_2 = 0$ and $dx_2/dt = 0$ at $t = -\infty$, then the sum of the solutions, x, will also satisfy the same condition: $x = 0$ and $dx/dt = 0$ at $t = -\infty$.

Let us now combine the reasoning concerning linearity and the familiar solution for the delta function so as to obtain a general solution for a force that depends on time in an *arbitrary* fashion. We partition the graph of the force $F(t)$ into strips of width $\Delta\tau$ (Figure 17.4.3). What does a separate strip located between τ and $\tau + \Delta\tau$ of time t represent? Let us change the designations, leaving t for "current" time varying from $-\infty$ to $+\infty$, whereas τ will refer to the given chosen strip. The height of the strip is $F(\tau)$, the width is $\Delta\tau$, and the area (i.e., the impulse) is $F(\tau)\,\Delta\tau$. Since the strip is located at $t = \tau$, it is obvious that it can be replaced by the delta function with a coefficient equal to the impulse:

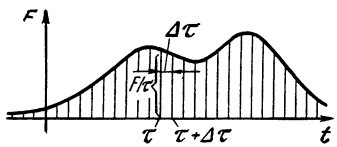


Figure 17.4.3

$F(\tau)\,\Delta\tau\,\delta(t-\tau)$. We already know the solution of the equation of motion for the delta function. We denote it by $x_1(t, \tau)$. The solution as the function of time t depends on the instant τ of application of the force. Recall that

$$x_1(t, \tau) = \begin{cases} 0 & \text{for } t < \tau, \\ \dfrac{t-\tau}{m} & \text{for } t > \tau. \end{cases} \qquad (17.4.4a)$$

One of the strips into which the force has been decomposed, from τ to $\tau + \Delta\tau$, is $\delta(t-\tau)$ with the coefficient $F(\tau)\,\Delta\tau$. Thanks to the linearity of the equation, the solution for a force in the form of such a strip is obtainable by multiplying x_1 into that coefficient: $F(\tau)\,\Delta\tau\,x_1(t, \tau)$. This is the solution for the action of a single strip.

Now let us take advantage of linearity and write out the solution for the function $F(t)$, which we consider as the sum of the strips. It is clear here that the summation should actually be replaced by integration. The result is

$$x(t) = \int_{-\infty}^{\infty} x_1(t, \tau)\,F(\tau)\,d\tau. \qquad (17.4.5)$$

At first glance this formula is rather strange: the x-coordinate at time t is expressed by an integral from $-\infty$ to $+\infty$ with respect to τ, that is, the force enters into this expression at *all* instants of time. Yet it is clear that the law of force subsequent to time t does not affect the preceding motion. However, there is no error in the expression for $x(t)$. The properties of the function $x_1(t, \tau)$ ensure reasonable properties of the solution. Indeed, $x_1(t, \tau)$, is zero for $t < \tau$. Hence, when integrating with respect to τ we actually do not need to take $\tau > t$, since the integrand is identically zero due to the factor $x_1(t, \tau)$ being equal to zero. Recalling the expression (17.4.4a) for $x_1(t, \tau)$, we obtain

$$x(t) = \frac{1}{m}\int_{-\infty}^{t} (t-\tau)\,F(\tau)\,d\tau. \qquad (17.4.5a)$$

This method of obtaining a solution has very great general significance. To summarize then: if for a linear system we know a solution referring to the action of the delta function, the solution referring to the action of an arbitrary function ($F(t)$ in our example) is obtained by simple summation (integration, to be more precise).

The ideas of linearity and addition (the real term is *superposition*) of solutions apply not only to such simple problems as the motion of a point; they hold true in vast areas of mathematics, physics, and the natural sciences. It sometimes happens that a system is very complicated and it is impossible to solve the equations even for the most simple action of the delta function. A solution corresponding to the delta function can occasionally be obtained experimentally. In other cases such a solution can be obtained from physical reasoning (see Exercise 17.4.1). Then linearity comes into play and we obtain the answer for an *arbitrary* acting function. The solution that corresponds to the delta function ($x_1(t, \tau)$ in our example above) is so important that it has a special name, the ***Green function*** of the problem. Curiously enough, the English mathematician George ***Green*** (1793-1841), for which the function was named, lived in the 19th century and quite naturally knew nothing about the delta function.[17.10] But it was only

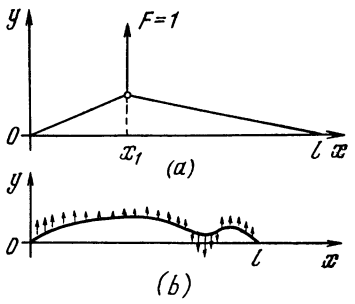


Figure 17.4.4

the introduction of the delta function that clearly and succinctly explained the essence of the Green function.

Examples of this nature abound in mathematics, for we know numerous results pertaining to tangent lines, areas, and volumes that were obtained before the invention of derivatives and integrals. The advance of science lies not only in the attainment of new heights and new results but also in popularizing and simplifying earlier derivations. The aim of this book is precisely that: to simplify the understanding of our classical heritage, the fundamentals of higher mathematics.

Exercises

17.4.1. Consider a string held taut by a force k and with ends fixed at points $x = 0$ and $x = l$. Regarding the deviation as small, determine by the parallelogram-of-forces law the shape of the string under a unit load at point $x = x_1$ (Figure 17.4.4*a*). Obtain the formula for deviation of the string under a force distributed along the length of the string via an arbitrary law $f(x)$ (Figure 17.4.4*b*).

17.4.2. Find the motion of a pendulum under a force expressed by the delta function, that is, solve the equation $m(d^2x/dt^2) = -kx + \delta(t - \tau)$ provided that $x = 0$ and $dx/dt = 0$ at $t = -\infty$. Using this solution, find the motion of the pendulum under a force dependent on time via an *arbitrary* law.

[17.10] The author of excellent papers in mathematics and mathematical physics, Green came from a family so poor that he did not receive any education at all; it was not until comparatively late in life that he became acquainted, through his own efforts, with mathematics and physics. At the end of his life, though, chiefly thanks to support from the influential William Thomson (Baron Kelvin), Green was offered the post of professor of mathematics at Cambridge University.

Conclusion. What Next?

Higher mathematics, or, to be more exact, differential and integral calculus, makes it possible to solve a large class of problems that do not lend themselves to solution by the methods of elementary mathematics (arithmetic, algebra, and geometry). Of tremendous importance is the very formulation of such new concepts as instantaneous velocity, acceleration, impulse. These notions (and numerous others in diverse fields) are formulated exactly only in the language of derivatives and integrals.

The knowledge you have gained in reading this book constitutes but a small part of the whole science of mathematics, to be exact, a small part of those divisions of mathematics that find application in the natural sciences and technology. In the Preface we said that a cultivated person should have a general picture of mathematics, irrespective of his or her field of interest. The three parts of our book provide enough general information for that purpose. But if your field is connected with engineering, chemistry, or (especially) physics, you will need more. Here we wish to outline in brief the fields of physics and the associated divisions of mathematics that you will most likely study in the future.

Note that up till now the exposition has been that of a textbook, and if you have put your mind to the matter at hand, you will have mastered the material in all its details. What now follows is a very short outline of the difficult problems that lie ahead. The style is no longer that of a textbook. We do not hope to explain the content of complex mathematical theories on so few pages, of course. We wish merely to give the reader a general impression of the problems of these theories to arouse interest in them. Some of the books listed in the Selected Readings will help you to obtain a broader view of mathematics and of mathematical physics.

For a better understanding of what is to follow, let us briefly state a common feature of all the problems we have dealt with up to now. These were problems involving the motion of a single particle in mechanics or the flow of current in a circuit. We dealt with functions of *one* variable (time). The number of functions was one (current as a function of time) or two (the position of a body, $x = x(t)$, and the velocity of the body, $v = v(t)$).

Quite naturally there follows the purely quantitative generalization: problems involving the motion of two bodies, three bodies, etc., will ultimately lead to problems of the motion of a gas or a liquid. But one gram of hydrogen consists of 3×10^{23} molecules, hence 3×10^{23} separate bodies. It must be clear at this point that the old method is now useless and new methods are needed. Not only is it impossible to solve 3×10^{23} equations, there is neither paper nor time enough to write them down. Even writing them down is out of the question because we can never know either the exact number of molecules or their initial positions, which would then yield the initial conditions imposed on the equations.

To our aid come new sciences that arose much later than differential and integral calculus. These are hydrodynamics and gasdynamics, which did not come into being until the 19th century. Both differ substantially from the mechanics of a single point in the formulation of problems and in the methods used to solve them (compare Sections 9.14 and 9.15 and Section 17.3). Here we are not interested in the actual number of molecules of a gas contained in a volume. Instead, such notions are introduced as the ***distribution of the density*** of the gas in space, $\rho = \rho(x, y, z)$ (the mass of the gas per unit volume), [C.1]

[C.1] Strictly speaking, the introduction of the notion of density at a point in space, $\rho(x, y, z)$, requires an operation similar to differentiation. This quantity is defined as the *limit* of the mean density of the gas

the ***pressure*** of the gas, $p = p(x, y, z)$, and the ***velocity*** of the gas at different points in space. What is more, all these quantities depend on time, t, so it is natural to write $p = p(x, y, z, t)$. Thus, from problems of several functions of one variable we pass to functions of several independent variables.

Accordingly, in setting up the equations that describe the motion of a gas and other characteristics, we encounter *partial derivatives* with respect to time and the spatial coordinates, for instance, $\partial\rho/\partial t$, $\partial\rho/\partial x$, $\partial\rho/\partial y$, and $\partial\rho/\partial z$, where, say,

$$\frac{\partial\rho}{\partial y} = \lim_{\Delta y \to 0} \frac{\rho(x, y+\Delta y, z, t) - \rho(x, y, z, t)}{\Delta y}.$$

An extremely important division of mathematical physics is the investigation of *partial differential equations*; we gave a very rough outline of them in Section 10.8. The equations describe the motion of liquids, gases, and solids, the propagation of heat in various media, the diffusion of atoms and molecules. They are so important that they are often called ***equations of mathematical physics.***

Another point that must be mentioned is that velocity is a *vector* quantity, $\mathbf{v} = \mathbf{v}(x, y, z)$ or even $\mathbf{v} = \mathbf{v}(x, y, z, t)$, which means that at every point in a gas (or liquid) the magnitude and direction of the velocity are specified. In other words, we can say that three components of the vector $\mathbf{v}$, namely, v_x, v_y, and v_z, are given and that these components depend on the point in space and on time. This leads to more complications (or simplifications) in problems of hydrodynamics and gasdynamics; however, we will not go into them here.

In all these theories it is possible, at least in principle, to continue regarding the system as consisting of separate particles and as being described by many functions of one variable (time).

contained in a small volume (the mass-to-volume ratio) as the volume becomes ever smaller. Changes in the density due to one or two molecules leaving or entering the volume do not interest us here.

But there are other physical theories, primarily the ***theory of electromagnetism***, where such an approach is impossible.

Consider two point charges at rest. The force acting between them depends on their position (the distance r between them). This would seem to be a problem involving six functions (the coordinates x_1, y_1, and z_1 of one charge and the coordinates x_2, y_2, and z_2 of the other) of one variable (time). To a first approximation, the motion of the charges changes but little—one merely has to take into account the magnetic interaction that appears between the charges, an interaction that depends on the velocities of the particles.

An extremely important factor—one demanding a fundamentally new approach—is the existence of a ***lag in the interaction*** due to the propagation of the interaction with a finite speed (the speed of light). The action of one charge on the other at time t depends on the position (and velocity) of the first charge at some *earlier* time $t - \tau$, which lags behind t by a finite amount $\tau = r/c$, where c is the speed of light. What takes place in this interval of time?

In a vacuum the space between the charges contains an electric field and a magnetic field. At each point in space we specify two vectors, $\mathbf{E}$ and $\mathbf{H}$, called the ***electric field strength*** and the ***magnetic field strength***, respectively. The magnitude and direction of each vector depend on the presence and motion of charges. Small ("test") charges placed at certain points in space provide the possibility of measuring the fields $\mathbf{E}$ and $\mathbf{H}$ at the given points. But there is still more to the theory of electromagnetism. In the vacuum, where there are no charges, the electric and magnetic fields act on each other: a change (over time) in one field is related, via Maxwell's equations, to the spatial derivatives of the other field. The time derivative of the magnetic field generates a circular electric field, while the time derivative of the electric field plays the same role as electric current and gener-

ates a magnetic field. The result is a neat pattern: charges generate fields at the points where they exist, and the interaction of the fields carries the information concerning the position and motion of the charges throughout the entire space.

Decisive in the development of this theory is a consideration of the two vectors **E** and **H**, or the six quantities E_x, E_y, E_z, H_x, H_y, and H_z, as functions of four quantities: the three coordinates x, y, and z and the time t. But specifying these functions at a certain time amounts to the fixing of an infinitude of their values at all points of space. The theory is more difficult than the theory of the motion of one or several particles, but the results are richer.

Mathematically, the theory of the electromagnetic field is a theory of partial differential equations and in this respect is similar to the theory of elasticity, acoustics, and gasdynamics. The only difference is that in the latter case the equations are arrived at via an idealization: in speaking of the density of a gas we abstract ourselves from the separate molecules and only in this approximate sense can a gas be considered as a continuous medium characterized by a continuous function $\rho(x, y, z, t)$ (the density of the gas). An electric (or, more precisely, an electromagnetic) field is actually a continuous function of the spatial coordinates and time.

Hydrodynamics, the foundations of which were laid in the 18th century by the works of D. Bernoulli, J. d'Alembert, and L. Euler, prepared the mathematical tools for the electromagnetic theory worked out in the second half of the 19th century by James Clerk ***Maxwell*** (1831-1879). No wonder, then, that at first attempts were made to transfer the ideas of mechanics to the electromagnetic theory. A special substance, called the ***ether***, was hypothesized as being responsible for electric and magnetic phenomena. We know that the mathematical analogy between hydrodynamics and electromagnetism, and also the similarity of the equations of the two theories, remains; however, the physical meaning of the electromagnetic theory has proved to be different and does not reduce to mechanics. After prolonged and intense debates scientists completely abandoned the idea of any ether.

When speaking of a mathematical theory, one must not only speak about the statement of the problem and the initial equations but also about the nature of the results.

We can name two types of solutions for partial differential equations. One is characteristic of phenomena that develop in a *limited* volume: these are natural oscillations with definite frequencies. A body of a given shape has a certain set of frequencies.

Recall the pendulum and its definite frequency of oscillation. If an external (periodic) force F acts on the pendulum, we get the characteristic phenomenon of ***resonance*** when the frequency of F is close to the (natural) frequency of the pendulum. All these physical ideas fit the theory of ordinary differential equations: $m\,(d^2x/dt^2) = -kx + F(t)$. In the theory of partial differential equations it appears that a body has *many* frequencies and behaves like a set, or collection, of many pendulums with distinct frequencies, whence there are many resonances. [C.2] You can verify this at once if you have a piano at home. Depress one of the keys slowly and soundlessly, so as to release the string without striking it with the hammer. Now strike the other key sharply and listen to the response of the free string.

The other type of solution of partial differential equations has to do with matter or fields that fill *all* space (propagation of waves). These are the waves of radio and light (in the electromagnetic theory) and sound waves in elastic media. Waves have the remarkable property of being able to carry infor-

[C.2] If you wish to understand this aspect in detail, go back to Section 10.8.

mation: pressure or an electric field at one point (near the receiver) as a function of time proves to be similar to the curve of that same source quantity (near the transmitter) as a function of time.

It is possible to construct the solutions of equations describing the directed beam of a searchlight or laser. The beam of a searchlight and the jet of water from a hose are strikingly similar. Knowledge of the properties of solutions of different kinds of problems and analogies between phenomena described by similar types of equations have always been extremely important in the development of physics.

The method of ***mathematical modeling*** consists in finding a mathematical scheme so close to the phenomenon under consideration that an examination of the scheme can replace a study of the phenomenon. The scheme might be either an ordinary differential equation or a system of such equations. Sometimes it is a partial differential equation with initial and boundary conditions of one kind or another (see Section 10.8). In other (and more sophisticated) mathematical constructions we have to deal with the notion of ***probability***, on which we will dwell below. It is essential that the model should present a sufficiently accurate description of the given real-life phenomenon.[C.3] If, in so doing, we come to an earlier-studied mathematical scheme, we can immediately transfer to the new case all the findings pertaining to it. For example, from the fact that the mechanical vibrations (see Chapter 10) and electromagnetic oscillations (see Chapter 13) are described by differential equations of the same type it follows that the theory of electric circuits has a number of attributes (resonance, for instance) that are characteristic of mechanical vibrations.

The power of mathematics is connected largely with the repetition of one and the same mathematical scheme applied to a large number of phenomena of different kinds. Mathematics might be described as a rather limited set of "keys," each of which unexpectedly opens up numerous doors that appear to be different (compare, for instance, Sections 2.1, 2.2, and 2.5). This circumstance is connected with the fact that nature's arguments are simple—there are no special intricacies—and there are not so many simple mathematical constructions. The fantastic generality of mathematical methods has always amazed scientists.[C.4]

For a long time atomic spectra were a mystery to physicists. It was not so much the specific laws and numerical values of the frequencies but the very fact that one and the same atom emits or absorbs (via resonance) the oscillations of several distinct but quite definite frequencies. The similarity to the oscillations of elastic bodies suggested the formulation of the equations of ***quantum mechanics.*** Likewise, the similarity between a *stream of particles* and solutions of equations particular to *waves* found its application in quantum mechanics, where it gave rise to particle-wave dualism. For instance, light can be regarded as a wave (the viewpoint of Christian Huygens) and at the same time as a stream of particles (or corpuscles) of light, or ***photons*** (Newton's concept; see Section 12.6 and subsequent sections). In pre-20th-century physics these two approaches seemed to contradict each other, and a

[C.3] The words "sufficiently accurate" emphasize the idealization (that is, making a rough approximation of the phenomenon under study by disregarding details that are secondary and do not interest us at the moment) that we invariably encounter when passing from physical reality to the mathematical scheme describing it.

[C.4] See, for instance, the following articles by several distinguished physicists: E.P. Wigner, "The unreasonable effectiveness of mathematics in the natural sciences" in *Symmetries and Reflections*, Indiana University Press, Bloomington, 1967 (reprinted by Ox Bow Press, Woodbridge, Conn., 1979), pp. 222-237, and C.N. Yang, "Einstein's impact on theoretical physics" in *Physics Today*, 33, No. 6, pp. 42-49, 1980.

choice had to be made. Modern quantum mechanics shows how and in what sense both points of view are correct and supplement each other.

We have already presented several examples pertaining to the theory of equations of different types.

An example of a different kind is offered by *complex numbers* and *functions of a complex variable* (see Chapters 14, 15, and 17). Complex numbers, which originally arose from the solution of algebraic equations, were for a long time regarded with much mistrust. However, progress in hydrodynamics and gasmechanics turned out to be closely connected with these unusual "numbers." The theory of functions of a complex variable was evolved by Augustin Cauchy and Georg Riemann and further developed by Karl ***Weierstrass*** (1815-1897) just when it was needed to make further scientific and technical advances possible. The rise of aeronautics at the beginning of the 20th century was based in large measure on calculations of fluid flow by the methods of "complex analysis." One of the pioneers in this field was Nikolai ***Zhukovsky*** (1847-1921), the distinguished Russian investigator in the field of mechanics.

Geometry offers another marvelous example of the development of mathematics.

Ordinary experience teaches us that in space it is convenient to introduce three coordinates: x, y, and z. Any further complications would seem to be superfluous, "a trick of the devil." Yet, coordinates ξ, η, and ζ can be introduced in a different way so that the condition ξ = constant corresponds to some ***curved surface*** (whereas x = constant for arbitrary y and z is the equation of a plane perpendicular to the x axis). For instance, it is possible to introduce into space the three coordinates ρ, φ, and ϑ, where $\rho = OM$ is the distance between the variable point M and the origin O, while φ and ϑ are the geographical coordinates (longitude and latitude) on the ***sphere*** ρ = constant. (In this case the surfaces φ = constant are ***half-planes*** and the surfaces ϑ = constant are ***cones*** (why?).) Coordinates can also be introduced by countless other methods.

To summarize, then, we can introduce ***curvilinear coordinates*** ξ, η, and ζ, and with a lot of effort, agonizingly, learn to compute point-to-point distances and other quantities with the aid of these new coordinates.

At first glance this is a dull and totally useless effort. One must possess a peculiar bent of mind to see beauty in overcoming difficulties and in developing a clumsy theory with arbitrary coordinates of the most general nature. C.5

Elaboration of the theory of curvilinear coordinates in Euclidean space might appear to be purely an achievement of method. However, the generalized coordinates ξ, η, and ζ are equally convenient (or equally inconvenient) in describing ordinary space (in which Euclidean geometry holds true) and "curved" space. The rectangular coordinates x, y, and z are convenient for ordinary space but are absolutely unsuitable for a description of curved space. These coordinates do not give even a hint of the possibility of the existence of any other kinds of space.

The transition to curvilinear coordinates, which seemed to be such a needless complication, actually prepared us for a vast range of spaces whose very existence was totally unknown to us. Then it turned out that the force of universal gravitation is linked up with

---

C.5 Such was the mind of the great Gauss, who worked out a "geometry in curvilinear coordinates." Displaying a profound understanding of the applied significance of mathematics (for example, Gauss arrived at the idea of curvilinear coordinates from his work in geodesy), he also possessed an insatiable curiosity concerning the secrets of the world of mathematics, which prompted him to devote hours and days to arduous calculations in the hope of some new results (say, the conversion of common fractions to decimal fractions with hundreds of figures in the period of the fraction).

the fact that space is somewhat curved. True, this "somewhat" had to do with the conditions here on the earth and in the solar system. In certain phenomena of a larger scale (catastrophic explosions of stars, evolutionary processes in the Universe), space may turn out to be highly curved, and here one must inevitably go over from conventional Euclidean geometry to non-Euclidean. Metaphorically speaking, the creation of non-Euclidean geometry, a purely mathematical achievement,[C.6] was a flash of lightning, and it was followed by a clap of thunder—the creation of the general theory of relativity, or the geometric theory of gravitation. Today physics is being invaded by completely new kinds of geometry, which at first glance seem quite removed from any reality.

Back in the 18th century, d'Alembert drew attention, in his article "Dimensionality" (in the famous *Encyclopédie* written together with Denis ***Diderot*** (1713-1784)), to the fact that our world is essentially four-dimensional—besides the three spatial coordinates x, y, and z of a point, it is also necessary, in order to distinguish events taking place in the world, to know the time, t. Moreover, the connection between the coordinates x, y, z, and t of the four-dimensional physical world later proved to be more complicated than Newton and d'Alembert had originally imagined; in particular, in this "world" it is impossible to distinguish unambiguously between the spatial coordinates x, y, and z and the temporal coordinate t (this is the essence of ***Einstein's special theory of relativity***). Later (beginning, essentially, with the works of Lagrange, a junior contemporary of d'Alembert) the idea of ***multidimensional spaces*** entered into theoretical physics. For instance, ***phase space***—for the simplest case of a moving point, the phase space is *six-dimensional*, with coordinates x, y, z, x', y', and z' ($x' = dx/dt$, $y' = dy/dt$, $z' = dz/dt$), which are the spatial coordinates and the projections of the velocity vector of the point.[C.7] Today, too, in theoretical physics we often have to deal with multidimensional worlds that are simply impossible to imagine; finally, we have the ultimate monstrosity—an infinite-dimensional world each point of which has an infinitude of coordinates (expressed as numbers—and we are lucky if the numbers are real and not complex). Instances of such infinite-dimensional spaces are found in quantum mechanics. It was most fortunate for physicists that the famous German mathematician David ***Hilbert*** (1862-1943) evolved a theory of infinite-dimensional spaces just before the appearance of quantum mechanics.

In recent years ***topology***, one of the newest and most abstract sections of geometry, which took its start in the works of Henri Poincaré, has found unexpected applications. Topology deals exclusively with the "roughest" properties of geometric figures. For instance, in topology a sphere does not differ from a cube, but a doughnut (torus) does because it has a hole. Topology analyzes these "rough" geometric properties with great thoroughness and depth. All of a sudden this not-so-trivial analysis aroused the interest of physicists.[C.8]

C.6 Again, how amazing that the non-Euclidean spaces that were so necessary for progress in physics came into being at just the right time in the works of mathematicians (Nikolai ***Lobachevsky*** (1792-1856) of Kazan, Russia, John ***Bolyai*** (1802-1860) of Hungary, and Carl Gauss; then Georg Riemann and, later, Felix ***Klein*** (1849-1925) of Germany and Henri ***Poincaré*** (1854-1912) of France).

C.7 The representation of a physical process in the phase space of a (mechanical) system is known as the "***phase portrait***" of the process. For instance, the phase portrait of the uniform motion of a point along the z axis, $z = at + b$, is the *straight line* $z' = a$ in the phase plane (z, z') of the moving point (here $z' = dz/dt$ is the velocity of the point). The phase portrait of harmonic oscillations, that is, oscillations characterized by a constant total energy $E = kx^2/2 + m(x')^2/2$, is described by a point in the phase plane (x, x') moving along an *ellipse* $E =$ constant.

C.8 We would like to draw the reader's attention to the excellent, though by no means easy, textbook by B.A. Dubrovin,

And no one can say what other esoteric approaches to space will be concocted by physicists of the future.[C.9]

At the intersection of (multidimensional) mathematical analysis and topology there lies a new theory, the theory of catastrophes. (Catastrophe theory is the invention primarily of the French mathematician René Thom (b. 1923).) It immediately gained (perhaps excessively wide) recognition in connection with the many different possibilities of its use in the natural, humanitarian, and social sciences.[C.10] The theory poses the question of the conditions under which smooth and generally well-behaved analytic functions can describe phenomena similar to explosions, that is, a discontinuity in the final result. Simple examples of this kind were known long before the name "catastrophe theory" appeared.

Let a quantity a be given as a smooth function of another quantity, b, or $a = f(b)$, and suppose that a can be fixed and then b can be observed or measured. The smoothness of the function $a = f(b)$ by no means guarantees the smoothness of the inverse function $b = \varphi(a)$. For instance, if $a = f(b)$ has a maximum $a = a_{\max}$ at $b = b_m$, then for $a < a_{\max}$ there are two values of b near b_m, whereas for $a > a_{\max}$ there is not a single value of b. Hence, b experiences a discontinuity as a varies smoothly.

We often meet phenomena of this kind in the theory of combustion and explosion, where the conditions imposed on the chemical reactions are determined via algebraic equations. The external conditions enter the problem as parameters, but the solution of the equations may be a discontinuous function of the parameters. Discontinuities are therefore specific features of the solutions. Catastrophe theory shows that these features can be classified by dividing them into several "rough" (topological) types corresponding to the characteristic phenomena described by a given equation.

In the study of nature, a diversified approach is desired: the overcoming of mathematical difficulties, the mastering of the mathematical apparatus, physical intuition, boldness of conception, experimentation, and mathematical modeling. All these combined make it possible to advance science.

Modern physics would be absolutely inconceivable without the concept of ***probability***, which is quite different from anything discussed in this book.

All the laws of physics mentioned above—Newton's second law $mx'' = F$ and its particular case, the law of inertia (see Section 9.4), Boyle's law $pV =$ constant and its generalization, van der Waals's law (see Section 7.3)—were of a *dynamical* nature, that is, they enabled us to make a precise prediction of the phenomenon: when the volume occupied by a gas is halved, its pressure doubles (Boyle's law); if the motion of a body takes place in the absence of any forces acting on it, its velocity in future will be exactly the same as it was at the start; and so on. The laws of the falling of, say, a randomly tossed coin are altogether different. Here all one can say in advance is that when a coin is tossed into the air many times, it will fall heads up in *approximately half of the cases*, that is, the *probability* of its falling heads up is equal to 0.5 (this result is guaranteed by the condition that the coin is true, that is, both sides are absolutely equal). The

A.T. Fomenko, and S. P. Novikov, *Modern Geometry: Methods and Applications*, 3 parts, Berlin, Springer, 1984, which is intended largely for theoretical physicists.

C.9 An interesting but not simple booklet which deals with the overall question of the relationships between physics and mathematics from the present-day viewpoint is Yu.I.Manin's *Mathematics and Physics*, Birkhäuser, Boston, 1981. The two papers mentioned in footnote C.4 deal with the same question.

C.10 See, for example, the article by E.C. Zeeman "Catastrophe theory" in *Scientific American*, 234, No. 4, 1976, pp. 65-83, and V.I. Arnold's book *Catastrophe Theory*, Springer, Berlin, 1984. A thorough treatment of the subject is presented in *Catastrophe Theory and Its Applications*, Pitman, London, 1978, by T. Poston and I. Stuwart, which also contains many expressive examples.

physical laws by which one cannot predict the exact outcome of an experiment but only the probability of one outcome or another, that is, the frequency with which the outcome is repeated when an experiment is carried out many times under the same conditions, are called ***statistical laws***, and the branch of mathematics that deals with them is known as the ***theory of probability***. C.11

The example of a coin being tossed many times is the simplest, almost a trivial, instance of a ***probabilistic process*** characterized by the accidental nature of the outcome of its separate stages (in this case, separate tossings).

Here is a still more typical example of the same kind, used by physicists as a simple model of the process of the diffusion of gases (the model of Paul and Tatiana Ehrenfest[C.12]). Let us take a vessel divided into two parts, A and B, by a porous membrane. It contains N particles (gas molecules, say) distributed in some way between A and B: at each moment of time, one of the particles is chosen at random and transferred from its part of the vessel to the other. We can pose the question of how many times, on the average, a transfer must be made so that the initial distribution, 100 particles in A and 0 particles in B, becomes 50 particles in A and 50 particles in B. It turns out that about 140 transfers are needed. The reverse process is also possible (at least in principle): with an initial distribution of 50 particles in A and 50 in B, random transfers can lead to 100 particles in A and 0 in B, but this requires $2^{100} \simeq 10^{30}$ transfers. This example illustrates the concepts of thermodynamic irreversibility and fluctuations.[C.13]

The case of two vessels is an idealization, of course. Typical problems deal with the movement of particles in space, where we seek the law that governs changes in the particle distribution depending on the coordinates and time. When it is a question of large particles, visible under a microscope, we speak about Brownian motion. It is characteristic that the physicists who studied Brownian motion (A. Einstein, M. Smoluchowski, A. Fokker, and M. Planck) turned to probabilistic processes long before the problem attracted the attention of mathematicians (A. Kolmogorov, A. Khinchin, W. Feller, Paul P. Lévy, and others). For this reason the main equation of (continuous) processes allied to the "toy" process of transferring particles we have examined is called the ***Einstein-Smoluchowski equation*** or the ***Fokker-Planck equation*** or the ***Kolmogorov equation.***

It is typical that present-day physics pays careful attention to the statistical laws of nature and, hence, makes extensive use of the mathematical tools of the theory of probability. For instance, an aspect of the transition from the classical mechanics of Galileo and Newton to quantum mechanics is the replacement of the "dynamical" world of the rationalists in the 17th to 19th centuries by the random world of modern physics of the microcosm, in which one cannot in principle speak of the exact trajectory of a particle (say, an electron) but only of the probability of finding the particle at one or another point in space at a certain time.

Finally, another relatively young branch of mathematics that has acquired primary importance for physics is the ***theory of groups***. This theory studies the "degree" of symmetry of physical or other objects. It is clear, for example, that a square (Figure C.1*a*) is more symmetrical than an isosceles trapezoid (Figure C.1*b*) and that the latter is more symmetrical than a trapezium (Figure C.1*c*). The exact meaning of these statements is as follows. It is clear that

C.11 The most elementary concepts of the theory of probability are explained for beginners in B.V. Gnedenko and A.Ya. Khinchin, *An Elementary Introduction to the Theory of Probability*, W.H. Freeman, San Francisco, 1961, in K.L. Chung, *Elementary Probability Theory with Stochastic Processes*, 3rd ed., Springer, New York, 1979, and in F. Mosteller, R.E.K. Rourke, and G.B. Thomas, Jr., *Probability with Statistical Applications*, 2nd ed., Addison-Wesley, Reading, Mass., 1970. Two other good books are: W. Feller, *An Introduction to Probability Theory and Its Applications*, 2 vols., Wiley, New York, 1968, 1971, and J. Neyman, *First Course in Probability and Statistics*, Holt, Rinehart, and Winston, New York, 1951.

C.12 Paul (or in full, in Russian, Pavel Sigizmundovich) ***Ehrenfest*** (1880-1933) and Tatiana Alexeyevna ***Ehrenfest-Afanasjewa*** (1876-1964) were prominent 20th-century physicists who worked in Russia (and, later, in the USSR) and the Netherlands.

C.13 For further details and proofs of the results we have just stated see, for instance, J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, 3rd printing, Springer, New York, 1983, Section 7.3.

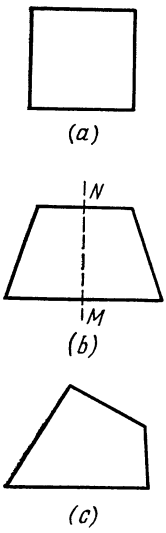


Figure C.1

the ***symmetry group*** of the square, a group consisting of all the rotations that transform the square into itself, is broader than the symmetry group of an isosceles trapezoid or that of a trapezium. The symmetry group of a square contains symmetries with respect to all the diagonals and the midlines and also with respect to rotations about the square's center through 90°, 180°, and 270°, while the symmetry group of an isosceles trapezoid is limited only to symmetry with respect to the midline *MN* (see Figure C.1*b*); the symmetry group of a trapezium (Figure C.1*c*) is even more limited, consisting merely of a single rotation through 360°, which leaves all the points in place and does not contain a single true movement!

The theory of groups was created by a brilliant young French revolutionary named Evariste ***Galois*** (1811-1832), who was killed in a duel when he was only 20.[C.14] (The duel had evidently been provoked by the police.) Galois developed the ideas of this theory as applied to algebraic equations in the early 1830s. He suggested classifying equations in accordance with their inherent "symmetry groups" and had no idea whatever of the possible physical applications of the concepts he had evolved. What is more, at the beginning of the 20th century the prominent English astrophysicist James Hopwood ***Jeans*** (1877-1946) energetically protested against the inclusion of elements of group theory in the course of mathematics for students of physics because he thought physicists would never have any use for it. Today, when the theory of groups has acquired such great significance in physics, Jeans's statement is often quoted as an example of a prediction that failed.

When scientists speak of the role that group theory plays in physics today, what they have in mind mainly is its application to the fundamental questions of the structure of matter, first and foremost to the theory of elementary particles. But the first practical applications of the ideas of symmetry were in *crystallography*. After all, crystals are highly symmetric bodies. It was found that crystals could be classified on the basis of their symmetry, and that such a classification sheds substantial light on all their properties. A big event at the end of the 19th century was the enumeration of all possible systems of crystal symmetry—there proved to be 230. These are the ***Fedorov groups***, so named in honor of Evgraf ***Fedorov*** (1853-1919) who, along with A. M. Schönflies, a German, and V. Barlow, an Englishman, discovered the crystallographic groups.

Physicists also found use for the ***Shubnikov groups*** named after the Soviet crystallographer Aleksei ***Shubnikov*** (1887-1970), who made a study of the symmetry of colored (for instance, black and white) ornaments. His theory also takes into account the colors of separate sections. The point is that other physical properties of objects (for example, electric charge) can play the role of color (black and white).

An enumeration of the Fedorov (and Shubnikov) groups required a thorough analysis of the very concept of "a sys-

[C.14] Read Leopold Infeld's book *Whom the Gods Love: The Story of Evariste Galois*, McGraw-Hill, New York, 1948.

tem of symmetry," and an understanding of the meaning of the operation of composition of rotations, which substitutes one rotation of a crystal for an equivalent of several rotations.

It is clear that each movement of a crystal or any other body that transforms it into itself constitutes a certain transformation in space; we can call a composition, that is, a sequence of two transformations α and β (first β and then α), a product: $\gamma = \alpha\beta$. For example, the composition of two reflections σ with respect to the straight line $0x$ or of four rotations π through 90° constitutes an identical transformation ε, which returns each point to its place: $\sigma^2 = \varepsilon$ and $\pi^4 = \varepsilon$ (Figure C.2). Thus we arrive at the "arithmetic" (or "algebra") of symmetries and the possibility of utilizing the precise tools of mathematics in a "symmetry calculus." This calculus of transformations (or symmetries) is the subject of the theory of groups. A peculiar situation arises here in that $\alpha\beta$ may not be equal to $\beta\alpha$, or that operations in symmetry calculus may be *noncommutative*.[C.15] For instance, the product $\sigma\pi$ of a rotation through 90° and a reflection with respect to the x axis (first the rotation, then the reflection) is equivalent to a single reflection with respect to the straight line $x + y = 0$ (Figure C.3*a*),

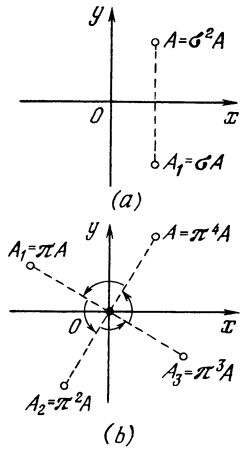


Figure C.2

whereas the product $\pi\sigma$ (first the reflection, then the rotation) is equivalent to an entirely different operation—a reflection with respect to the straight line $x - y = 0$ (Figure C.3*b*).

All these considerations played a substantial role in the study of the symmetry of crystals. Recall what was said in Section 14.1 about the origin of complex numbers. Counting began with the *natural* numbers 1, 2, 3, ... Later there was a need for *negative* numbers, —1, —2, —3, and so on, and also *zero*; at the same time, *fractions* came into

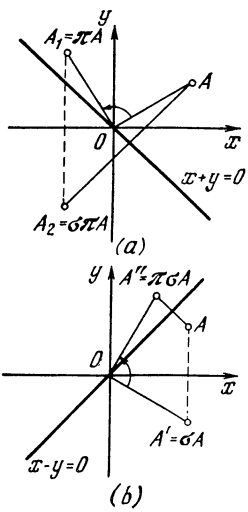


Figure C.3

[C.15] The first system of noncommutative numbers, called ***quaternions***, was discovered by the famous Irish algebraist, astronomer, and physicist William Rowan ***Hamilton*** (1805-1865). His quaternions differ from the common complex numbers $x + iy = z$ in that they contain *three* imaginary units, i, j, and k, so that the general quaternion is $w = u + xi + yi + zk$, where u, x, y and z are real numbers. The square of each imaginary unit is —1, but they are noncommutative, so $ij = k$, but $ji = -k$. Hamilton's calculus of quaternions was the first version of ***vector calculus*** (he called a quaternion without imaginary units, $w_0 = u$, a *scalar* and used the term *vector* for a purely imaginary quaternion, $w_1 = xi + yj + zk$). The renowned *Treatise on Electricity and Magnetism* (1873) of James Clerk Maxwell was written in the language of quaternions, to which Maxwell was accustomed, and not in the language of vectors so familiar to us.

being and also *irrational* numbers, such as $\sqrt{2}$. A further generalization of the number concept led to *complex* numbers; only after the introduction of these unusual numbers was the theory of equations complete (in particular, the elementary theory of quadratic equations). In Chapters 14, 15, and 17 we showed how fruitful the generalization of the number concept was, especially when combined with the derivative and the integral.

However, mathematicians could not bring themselves to challenge the principle of commutativity in numbers—for all newly introduced numbers it was always $a + b = b + a$ and $ab = ba$. What Galois accomplished should, therefore, be regarded as a most fundamental advance: he was the first to consider groups whose elements are noncommutative; in general, in the "arithmetic of Galois" $ab \neq ba$. In everything else, the theory of crystallographic groups is close to ordinary arithmetic: it has a unit element such that the product of this element with any member of the group, in either order, is that same member (in the "calculus of symmetries" the role of the unit element is played by the "identical transformation" ε), the operation of division is introduced as the inverse of multiplication ($a \div b = ab^{-1} = c$ if $a = cb$), and so on. However, the fact that multiplication may be noncommutative later played a most important role in physics.

Although the theory of groups was put to active use in the theory of crystals in the 19th century, the problems that arose seemed very remote from the profound problems of the main "building blocks of the Universe"—elementary particles. In 1932, however, the famous German physicist Werner Heisenberg established that replacing a (positively charged) ***proton*** in a nucleus by an (electrically neutral) ***neutron*** is similar to a rotation in mathematics. This discovery was a brilliant page in the history of physics. Actually, it was from this moment that groups and noncommutative operations and the theory of symmetry, regarded broadly, entered theoretical physics. And as physics brought to light more and ever newer particles, the idea of symmetry in the properties of particles and the application of the theory of groups became increasingly significant. The theory gradually became more complex as rotations were examined in an imaginary multidimensional space and account was taken of particle charge and spin. Today the approach to physical objects from the standpoint of their inherent symmetry is perhaps the prime method that physicists use as they try to analyze the great variety of elementary particles brought to light in the numerous experiments and theoretical works of scientists in the second half of the 20th century.

Modern physics makes extensive use not only of the ***theory of discrete (crystallographic) groups***, which, in a way, copies elementary arithmetic (noncommutative arithmetic!) but also the ***theory of continuous groups***, whose origins lie in algebra and mathematical analysis. Along with the question of the group of operations that transform the square in Figure C.1*a* (or a crystal) into itself, one can also pose the question of the symmetry group for the entire plane or for three-and multidimensional spaces. For example, the symmetry group of a Euclidean plane consists of all its motions

$$x' = x\cos\varphi + y\cos\varphi + a,$$
$$y' = -x\sin\varphi + y\cos\varphi + b, \qquad \text{(C.1)}$$

where angle φ and the numbers a and b are arbitrary (see formulas (1.9.6)). The fact that the angle φ of rotation and the quantities a and b characterizing the translation of a figure in the plane may be arbitrary permits us to speak of an *infinitely small* rotation (through an infinitely small angle $\Delta\varphi$ or $d\varphi$, with $d\varphi \to 0$) or an infinitely small translation. This in turn makes it possible to introduce the concept of a "group trajectory" (defined, say, by the fact that a and b are constant and φ varies at our choice) and also the concepts

of the differential, derivative, and integral in the group space. Here we can also define the main functions, such as, for instance, the exponential e^x (given, say, by formula (6.2.2), with x the "group variable"). The noncommutativity of the quantities that we meet in this theory creates absolutely new problems, which have no analogies either in the theory of functions of a real variable (or several real variables) or in the theory of complex variables. Continuous groups, like the group (C.1) of motions of a plane, were first studied in detail by the Norwegian mathematician Sophus ***Lie*** (1842-1899) and are now called ***Lie groups.*** During the past two decades, Lie group theory has acquired enormous significance for the classification of elementary particles in modern physics and for an understanding of particle interactions.

The outstanding importance of the theory of groups and, in particular, Lie groups, for modern physics prompts us to dwell briefly on the history of the origin of these departments of mathematics. C.[16] We have already spoken of the pioneering work done by the brilliant young Frenchman Evariste Galois in the theory of equations (his predecessors were the famous Lagrange, the Italian Paolo ***Ruffini*** (1765-1822), and the distinguished Norwegian mathematician Niels Henrik ***Abel*** (1802-1829)). C.[17] Galois's work was not appreciated during his lifetime; what is more, it even failed to gain notice. Galois wrote letters to the famous French mathematicians Cauchy and Simeon-Denis ***Poisson*** (1781-1840). Poisson was unable to grasp the importance of Galois's ideas, which were far ahead of his time, while Cauchy evidently did not even read the letter sent to him by an unknown youth.

Cauchy died in 1857, and in the 1860s the French Academy of Sciences decided to publish his complete works. The prominent mathematician Camille ***Jordan*** (1838-1922) was put in charge of the project, and in connection with this he made a close study of all of Cauchy's personal papers. Among them he discovered the letter from Galois, which, 30 years after it had been written, made a tremendous impression on him. Jordan painstakingly studied all the published and unpublished writings of Galois. He gradually arrived at the conviction that he must write a book about his young compatriot's ideas. Jordan's monograph, entitled *Traité des substitutions et des équations algébriques* (A Treatise on Substitutions and Algebraic Equations), which was published in 1870, was the first textbook on the theory of groups to appear in the mathematical literature.

In the period when Jordan was working on his treatise there were among his students two capable young friends, the mathematicians Felix Klein of Germany and Sophus Lie of Norway. They too took a keen interest in the ideas of Galois and the theory of groups. Jordan was the first to note the existence of two different types of groups: *discrete* (*crystallographic*) and *continuous*. The two pupils of his divided these two branches of mathematics between themselves, so to say. Klein's main achievements were in the sphere of discrete groups: a special type of discrete groups which he distinguished, now known as ***Klein groups***, has been arousing great interest lately. Along with that, in the dissertation which he submitted on the occasion of his admission to the faculty of the University of Erlangen and which later came to be called the Erlangen Programm he suggested a system of symmetry (the symmetry group) of any geometric manifold as a principle to be used in classifying the separate branches of mathematics. According to Klein, the Lobachevsky space differs from the conventional Euclidean space not only because its parallel lines have different properties (a secondary criterion which does not explain much) but also because it has a totally different symmetry group. Later this principle was applied in physics, and today we are inclined to believe that, for instance, Einstein's special theory of relativity differs from Newton's classical mechanics in that the former has a different symmetry-group structure in four-dimensional space-time; moreover, in "Einstein's world," this group is arranged so that the difference between the spatial and temporal coordinates (see Section 9.8 and the book mentioned in footnote 9.12) is somewhat diminished. Klein's principle of symmetry is of great importance to the whole of modern physics.

Sophus Lie was that rare type of scholar who devoted his entire life and all of his prodigious research and writings (six very large books and a multitude of articles which were

---

C.[16] For more details see I. M. Yaglom, *Felix Klein and Sophus Lie*, Znanie, Moscow, 1977 (an enlarged English version of this book will be published by Birkhäuser-Verlag, Boston, New York).

C.[17] The difference between the work done by Galois and that of his predecessors consisted not only, and not so much, in that Galois was the first to introduce the term "group" and to define this concept rigorously, as in that Lagrange, Ruffini, and Abel worked exclusively with ***commutative*** (or ***Abel***) ***groups***, where $ab = ba$ for any two elements a and b, whereas Galois introduced general-type (noncommutative) groups.

later collected in his multivolume *Collected Works*) to the theory of continuous groups (***Lie groups***). He worked out a wide-ranging theory of continuous groups. He transferred to differential equations the results that Galois had obtained for algebraic equations: the main difference here was that the symmetry group of an algebraic equation (the ***Galois group***) is finite, whereas the symmetry group of a differential equation (the ***Lie group***) is continuous. Lie's theory of the groups of differential equations (in particular, the classification of differential equations via their symmetry group) has attracted much attention of mathematicians and physicists in recent decades.

Lie did not live to the 20th century and, unfortunately, could not see the rapid flowering of his theory; for one thing, he did not know that the theory of continuous groups would be applied in physics. The group-theory period in the history of theoretical physics began with in-depth investigations by Nobel Prize winners Max Born and his pupil W. Heisenberg, both of Germany, and P. A. M. Dirac of Britain. Herman Weyl, [C.18] a pupil of D. Hilbert [C.19] (already mentioned on many occasions), both of them the most distinguished mathematicians of the 20th century, and the physicist E. Wigner, were also among the trailblazers of that period.

The theory of symmetry and the theory of groups play a very important role in the most fundamental and difficult part of physics, the theory of elementary particles.

For a long time this theory developed in one direction: physicists "split" complicated objects and discovered the simpler parts that comprised them. It thus appeared that molecules consist of atoms, atoms consist of nuclei and electrons, and nuclei consist of protons and neutrons. Today we would add that protons and neutrons consist of quarks. The evolution of physics was reminiscent of the way a child takes apart a nest of Russian painted wooden dolls and discovers a smaller doll inside each one it opens up. In each new layer, physicists saw a simpler picture: all the countless types of molecules proved to be combinations of approximately 100 types of atoms having approximately 1000 different nuclei. However, all these many nuclei were reduced to only two types of elementary particles, that is, protons and neutrons. But by about the middle of the 20th century an opposite trend appeared in theoretical physics—one toward an ever more complicated picture of the Universe. A very large number of elementary particles, more than 100, have been discovered in cosmic rays and also with the aid of accelerators. All these particles are just as "elementary" as the proton or neutron. Here the nest-of-dolls model came into operation again: many particles are now described as consisting of simpler particles, quarks, which have not yet been observed. At present, however, we are not only, and not so much, preoccupied with the question of whether "still more elementary particles" exist. The establishment of symmetry—that is, a resemblance between various particles—leads to extremely important conclusions about their properties. The interaction of particles with one another, the possibility of the transformation of certain particles into other particles, and other properties are very closely connected with symmetry. The secrets of quarks, electrons, and neutrino can perhaps be revealed by a deeper study of their symmetry. And today the theory of symmetry plays a cardinal role in all of the investigations that are being conducted. And so physicists naturally take a great interest in all aspects of the theory of symmetry and, first and foremost, the theory of groups (discrete and continuous).

Not long ago, in a serious article, we came across the following: "Since the theory of groups is now taught in high school, persons who recently finished school can skip the next section. However, older physicists should give it their attention in order to be able to discuss it with their children and to teach their students." We have not yet decided to include elements of

---

[C.18] See M. H. A. Newman, "Herman Weyl...," in *Biographical Memoirs of Fellows of the Royal Society*, 1957.

[C.19] See the informative book by C. Reid, *Hilbert*, Springer, Berlin, 1970.

the theory of groups in our mathematics textbook for students who are beginners in physics, but we may yet have to do so.

An absolutely new situation has arisen in mathematics and mathematical physics with the appearance of computers. They range from pocket calculators to computer complexes, each unit of which handles a separate stage of the operations. The functioning of these complexes can be compared in a way to the higher nervous activity of man, in which the left hemisphere of the brain (which controls speech) and the right hemisphere (which handles visual impressions) perform different functions. Computers have significantly changed the approach to problems in higher mathematics in that they often make it simpler to apply an approximate brute-force calculation than to use complicated mathematical models. Furthermore, computers have given rise to absolutely new branches of mathematics.

One occasionally hears the remark that "mathematics is a mill that grinds up only what is put into it." In this way poor results are explained by the fact that the original premises were faulty. In reality, the mill quite often turns out much more than is put in and the results are sometimes totally unexpected!

Fundamentally, mathematics can be regarded as a variety of refined logic. The remarkable thing is that having set up the rules of the logic and learned them, man has at his disposal a more powerful tool than ordinary "common sense," which is based on traditional logic.

Using his hands, man makes simple tools with the aid of which he makes machine tools; with the aid of these he constructs more complicated devices, and using these more complicated devices he does things which he could not do with his hands alone. Mathematics is very much like that. It develops more and more complicated theories, introduces fresh notions and enables us to comprehend and master the most unusual phenomena of nature.

At the end of the course of mathematical physics we can again start a new chapter entitled "What Next?", but do not lose heart, for not far off is the boundary line that marks the end of study and the beginning of creativity and the development of new theories.

We have tried to picture the reader who in a few moments will close this book with a sign of relief. Most likely, you are finishing school or are in your first year at college. May mathematics always remain for you an exact and beautiful language, a means of expressing ideas, a way of thought. May mathematics be more than merely another subject that must be "passed" at an examination and then left behind without a trace. Love mathematics—and mathematics will love you and help you in your work.

Selected Readings

Texts on Higher Math for Beginners

1. M. K. Potapov, V. V. Aleksandrov, and P. I. Pasichenko, *Algebra and Analysis of Elementary Functions*, Mir Publishers, Moscow, 1987.
2. L. S. Pontrjagin, *Learning Higher Mathematics*, Springer, Berlin, 1984.
3. S. M. Nikolsky, *Elements of Mathematical Analysis*, Mir Publishers, Moscow, 1983.
4. S. Lang, *A First Course in Calculus*, 4th ed., Addison-Wesley, Reading, Mass., 1978.
5. I. Niven, Calculus: *An Introductory Approach*, Van Nostrand, London, 1961.
6. L. D. Hoffmann, *Applied Calculus: A Year Course*, McGraw-Hill, New York, 1983.
7. M. Vygodsky, *Mathematical Handbook: Higher Mathematics*, Mir Publishers, Moscow, 1971; reprinted in 1975, 1978, and 1984.
8. R. Courant and H. Robbins, *What is Mathematics?*, Oxford University Press, London, 1941; reissued 1978, Chapter 8.

Advanced Texts (Higher Mathematics)

9. A. D. Myškis, *Introductory Mathematics for Engineers: Lectures in Higher Mathematics*, Mir Publishers, Moscow, 1972; 2nd ed. 1975, reprinted 1979.
10. I. P. Natanson, *Theory of Functions of a Real Variable*, 2 vols., Ungar, New York, 1955, 1959.
11. V. I. Smirnov, *A Course of Higher Mathematics*, vol. 1: *Elementary Calculus*. Addison-Wesley, Reading, Mass., 1964.
12. L. Bers, *Calculus*, Holt, Rinehart, and Winston, New York, 1969.
13. G. M. Fikhtengol'ts, *Fundamentals of Mathematical Analysis*, 2 vols., Pergamon Press, Oxford, 1965.
14. J. Marsden and A. Weinstein, *Calculus*, 3 vols., 2nd ed., Springer, New York, 1985.

Additional Reading (Higher Mathematics)

15. Ya. B. Zeldovich and A. D. Myškis, *Elements of Applied Mathematics*, Mir Publishers, Moscow, 1976.
16. R. Courant, Differential and Integral Calculus, 2 vols., Wiley, New York, 1937, 1936.
17. A. D. Myškis, *Advanced Mathematics for Engineers: Special Courses*, Mir Publishers, Moscow, 1975; reprinted 1979.
18. V. I. Smirnov, *A Course of Higher Mathematics*, Addison-Wesley, Reading, Mass. 1964: vol. 2, *Advanced Calculus*; vol. 3, Part 1, *Linear Algebra*; vol. 3, Part 2, *Complex Variables, Special Functions*; vol. 4, *Boundary Value Problems, Integral Equations, and Partial Differential Equations*; vol. 5, *Integration and Functional Analysis*.
19. "Mathematics in the Modern World," *Scientific American*, Sept. 1964.
20. P. Lax, S. Burstein, and A. Lax, *Calculus with Applications and Computing*, vol. 1, Springer, New York, 1976; 2nd printing 1984.
21. R. Courant and H. Robbins, *What is Mathematics?*, Oxford University Press, London, 1941; reissued 1978: Chapters 1 to 7.

Texts on Physics for Beginners

22. G. S. Landsberg, *Textbook in Elementary Physics*, 3 vols., Mir Publishers, Moscow, 1988.
23. J. B. Marion, *Physics and the Physical Universe*, Wiley, New York, 1971.
24. L. D. Landau and A. I. Kitaigorodsky, *Physics for Everyone*, 4 vols., Mir Publishers, Moscow, 1980, 1980, 1981, 1981; reprinted 1983, 1984, 1986 and 1987.
25. J. Orear, *Fundamental Physics*, 2nd ed., Wiley, New York, 1967.
26. F. J. Bueche, *Introduction to Physics for Scientists and Engineers*, 2nd ed., McGraw-Hill, New York, 1975.
27. F. A. Kaempffer, *The Elements of Physics: A New Approach*, Braisdall, Waltham, Mass., 1968.
28. *Physics, the PSSC text*, D. C. Heath, Boston, 1960.
29. F. W. Sears, M. W. Zemansky, and H. D. Young, *College Physics*, 5th ed., Addison-Wesley, Reading, Mass., 1980.
30. L. N. Cooper, *An Introduction to the Meaning and Structure of Physics*, Harper and Row, New York, 1968.
31. E. M. Rogers, *Physics for the Inquiring Mind: The Methods, Nature, and Philosophy of Physical Science*, Princeton University Press, Princeton, N. J., 1960.

Advanced Texts (Physics)

32. D. Halliday and R. Resnick, *Physics*, 3rd ed., 2 parts, Wiley, New York, 1978.
33. R. M. Eisberg and L. S. Lerner, *Physics: Foundations and Applications*, McGraw-Hill, New York, 1981.
34. R. P. Feynman, R. B. Leighton, and M. L. Sands, *Feynman Lectures on Physics*, 3 vols., Addison-Wesley, Reading, Mass., 1963, 1964, 1965.
35. R. P. Feynman, *The Character of Physical Law*, Cox and Wyman, London, 1965.

36. *Berkeley Physics Course*, McGraw-Hill, New York: vol. 1, *Mechanics*, 2nd ed., 1973; vol. 2, *Electricity and Magnetism*, 1965; vol. 3, *Waves*, 1968; vol. 4, *Quantum Physics*, 1971; vol. 5, *Statistical Physics*, 1967.
37. G. J. Kopylov, *Elementary Kinematics of Elementary Particles*, Mir Publishers, Moscow, 1983.

Readings in the History of Mathematics

38. D. J. Struik, *A Concise History of Mathematics*, 3rd. ed., Dover, New York, 1976.
39. E. T. Bell, *The Development of Mathematics*, 2nd ed., McGraw Hill, New York, 1945.

Appendices

Appendix 1. Derivatives

1. $y=c,\ \frac{dy}{dx}=0.$

2. $y=x,\ \frac{dy}{dx}=1.$

3. $y=x^a,\ \frac{dy}{dx}=ax^{a-1}=\frac{ay}{x}.$

4. $y=e^x,\ \frac{dy}{dx}=e^x.$

5. $y=a^x,\ \frac{dy}{dx}=a^x\ln a\simeq 2.3a^x\log_{10}a.$

6. $y=\ln x,\ \frac{dy}{dx}=\frac{1}{x}.$

7. $y=\log_a x,$

$$\frac{dy}{dx}=\frac{1}{x\ln a}\simeq\frac{0.434}{\log_{10}a}\frac{1}{x}.$$

8. $y=\sin x,\ \frac{dy}{dx}=\cos x.$

9. $y=\cos x,\ \frac{dy}{dx}=-\sin x.$

10. $y=\tan x,\ \frac{dy}{dx}=\frac{1}{\cos^2 x}.$

11. $y=\cot x,\ \frac{dy}{dx}=-\frac{1}{\sin^2 x}.$

12. $y=\arcsin x,\ \frac{dy}{dx}=\frac{1}{\sqrt{1-x^2}}.$

13. $y=\arccos x,\ \frac{dy}{dx}=-\frac{1}{\sqrt{1-x^2}}.$

14. $y=\operatorname{arc\,cot} x,\ \frac{dy}{dx}=-\frac{1}{1+x^2}.$

15. $y=\operatorname{arc\,cot} x,\ \frac{dy}{dx}=-\frac{1}{1+x^2}.$

Appendix 2. Integrals of Some Functions

1. $\int dx=x+C.$

2. $\int x^a\,dx=\frac{x^{a+1}}{a+1}+C\quad(a\neq-1).$

3. $\int\frac{dx}{x}=\ln|x|+C.$

4. $\int\frac{dx}{ax+b}=\frac{1}{a}\ln|ax+b|+C.$

5. $\int a^x\,dx=\frac{a^x}{\ln a}+C.$

6. $\int e^{kx}\,dx=\frac{1}{k}e^{kx}+C.$

7. $\int x^n e^{kx}\,dx=\frac{1}{k}x^n e^{kx}-\frac{n}{k}\int x^{n-1}d^{kx}\,dx.$

8. $\int\frac{dx}{1+e^{kx}}=\frac{1}{k}\ln\frac{e^{kx}}{1+e^{kx}}+C.$

9. $\int e^{kx}\sin ax\,dx$

$$=\frac{e^{kx}}{k^2+a^2}(k\sin ax-a\cos ax)+C.$$

10. $\int e^{kx}\cos ax\,dx$

$$=\frac{e^{kx}}{k^2+a^2}(k\cos ax+a\sin ax)+C.$$

11. $\int\sin kx\,dx=-\frac{1}{k}\cos kx+C.$

12. $\int\cos kx\,dx=\frac{1}{k}\sin kx+C.$

13. $\int\frac{dx}{\sin^2 kx}=-\frac{1}{k}\cot kx+C.$

14. $\int\frac{dx}{\cos^2 kx}=\frac{1}{k}\tan kx+C.$

15. $\int\sin^2 kx\,dx=\frac{x}{2}-\frac{1}{4k}\sin 2kx+C.$

16. $\int\cos^2 kx\,dx=\frac{x}{2}+\frac{1}{4k}\sin 2kx+C.$

17. $\int x^n\sin kx\,dx=-\frac{x^n}{k}\cos kx$

$$+\frac{n}{k}\int x^{n-1}\cos kx\,dx.$$

18. $\int x^n\cos kx\,dx=\frac{x^n}{k}\sin kx$

$$-\frac{n}{k}\int x^{n-1}\sin kx\,dx.$$

19. $\int\sin kx\sin lx\,dx=\frac{\sin(k-l)x}{2(k-l)}$

$-\frac{\sin(k+l)x}{2(k+l)}+C$ if $|k|\neq|l|$ (if $|k|$ $=|l|$, see No. 15).

20. $\int\cos kx\cos lx\,dx=\frac{\sin(k-l)x}{2(k-l)}$

$+\frac{\sin(k+l)x}{2(k+l)}+C$ if $|k|\neq|l|$ (if $|k|$ $=|l|$, see No. 16).

21. $\int \sin kx \cos lx\, dx = -\frac{\cos(k+l)x}{2(k+l)} - \frac{\cos(k-l)}{2(k-l)} + C$ if $|k| \neq |l|$.

22. $\int \tan kx\, dx = -\frac{1}{k} \ln|\cos kx| + C$.

23. $\int \cot kx\, dx = \frac{1}{k} \ln|\sin kx| + C$.

24. $\int \sqrt{ax+b}\, dx = \frac{2}{3a}(ax+b)^{3/2} + C$.

25. $\int \frac{dx}{\sqrt{ax+b}} = \frac{2\sqrt{ax+b}}{a} + C$.

26. $\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C$.

27. $\int x\sqrt{ax+b}\, dx = \frac{2(3ax-2b)(ax+b)^{3/2}}{15a^2} + C$.

28. $\int \sqrt{a^2-x^2}\, dx = \frac{1}{2}\left[x\sqrt{a^2-x^2} + a^2 \arcsin\frac{x}{a}\right] + C$.

29. $\int \frac{\sqrt{a^2-x^2}}{x}\, dx = \sqrt{a^2-x^2} - a \ln\left|\frac{a+\sqrt{a^2-x^2}}{x}\right| + C$.

30. $\int x\sqrt{x^2+m}\, dx = \frac{1}{3}(x^2+m)^{3/2} + C$.

31. $\int \frac{dx}{\sqrt{x^2+m}} = \ln|x+\sqrt{x^2+m}| + C$.

32. $\int \frac{\sqrt{a^2+x^2}}{x}\, dx = \sqrt{a^2+x^2} - a \ln\left|\frac{a+\sqrt{a^2+x^2}}{x}\right| + C$.

33. $\int \sqrt{x^2+m}\, dx = \frac{1}{2}[x\sqrt{x^2+m} + m \ln|x+\sqrt{x^2+m}|] + C$.

34. $\int \frac{\sqrt{x^2-a^2}}{x}\, dx = \sqrt{x^2-a^2} - a \arccos\frac{a}{x} + C$.

35. $\int \frac{dx}{x^2-a^2} = \frac{1}{2a} \ln\left|\frac{x-a}{x+a}\right| + C$.

36. $\int \frac{dx}{x^2+a^2} = \frac{1}{a} \arctan\frac{x}{a} + C$.

37. $\int \frac{dx}{ax^2+bx+c} = \frac{2}{\sqrt{4ac-b^2}} \times \arctan\frac{2ax+b}{\sqrt{4ac-b^2}} + C$ if $4ac - b^2 > 0$,

$\int \frac{dx}{ax^2+bx+c} = \frac{1}{\sqrt{b^2-4ac}} \times \ln\left|\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right| + C$ if $4ac - b^2 < 0$,

$\int \frac{dx}{ax^2+bx+c} = -\frac{2}{2ax+b} + C$ if $4ac - b^2 = 0$, i.e., $c = b^2/4a$.

38. $\int \frac{dx}{(ax^2+bx+c)^n} = \frac{2ax+b}{(n-1)(4ac-b^2)(ax^2+bx+c)^{n-1}} + \frac{(2n-3)2a}{(n-1)(4ac-b^2)} \int \frac{dx}{(ax^2+bx+c)^{n-1}}$ for $n \geqslant 2$ and $4ac - b^2 \neq 0$.

39. $\int \frac{x\,dx}{ax^2+bx+c} = \frac{1}{2a} \ln|ax^2+bx+c| - \frac{b}{2a} \int \frac{dx}{ax^2+bx+c}$ (see No. 37).

40. $\int \frac{dx}{x(ax^2+bx+c)} = \frac{1}{2c} \ln\frac{x^2}{|ax^2+bx+c|} - \frac{b}{2c} \int \frac{dx}{ax^2+bx+c}$ (see No. 37)

41. $\int \frac{dx}{x^m(ax^2+bx+c)^n} = -\frac{1}{(m-1)cx^{m-1}(ax^2+bx+c)^{n-1}} - \frac{(2n+m-3)a}{(m-1)c} \int \frac{dx}{x^{m-2}(ax^2+bx+c)^n} - \frac{(n+m-2)b}{(m-1)c} \int \frac{dx}{x^{m-1}(ax^2+bx+c)^n}$ (for $m > 1$).

42. $\int \sqrt[n]{ax+b}\, dx = \frac{n(ax+b)}{a(n+1)} \sqrt[n]{ax+b} + C$.

43. $\int \frac{dx}{\sqrt[n]{ax+b}} = \frac{n(ax+b)}{(n-1)a} \frac{1}{\sqrt[n]{ax+b}} + C$.

44. $\int \ln x\, dx = x \ln x - x + C$.

45. $\int (\ln x)^n\,dx$
$= x(\ln x)^n - n\int (\ln x)^{n-1}\,dx.$

46. $\int \arcsin\frac{x}{a}\,dx$
$= x\arcsin\frac{x}{a} + \sqrt{a^2-x^2} + C.$

47. $\int \arccos\frac{x}{a}\,dx$
$= x\arccos\frac{x}{a} - \sqrt{a^2-x^2} + C.$

48. $\int \arctan\frac{x}{a}\,dx$
$= x\arctan\frac{x}{a} - \frac{a}{2}\ln(a^2+x^2) + C.$

49. $\int \operatorname{arccot}\frac{x}{a}\,dx$
$= x\operatorname{arccot}\frac{x}{a} + \frac{a}{2}\ln(a^2+x^2) + C.$

Appendix 3. Series Expansions

1. $(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \ldots$
$(-1<x<1$ if m is not a positive integer).

2. $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$
(any x).

3. $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$
(any x).

4. $\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + \ldots \left(-\frac{\pi}{2} < x < \frac{\pi}{2}\right).$

5. $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots$
(any x).

6. $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$
$(-1<x\leqslant 1).$

7. $\arcsin x = x + \frac{x^3}{2\cdot 3} + \frac{1\cdot 3}{2\cdot 4\cdot 5}x^5 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}x^7 + \ldots \ (-1\leqslant x\leqslant 1).$

8. $\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots$
$(-1<x<1).$

Appendix 4. Numerical Tables

TABLE A4.1

x	e^x	e^{-x}	x	e^x	e^{-x}	x	e^x	e^{-x}
0	1.000	1.000	1.5	4.482	0.223	3.8	44.701	0.0224
0.1	1.105	0.905	1.6	4.953	0.202	4.0	54.598	0.0183
0.2	1.221	0.819	1.7	5.474	0.183	4.5	90.017	0.0111
0.3	1.350	0.741	1.8	6.050	0.165	5.0	148.41	0.00674
0.4	0.492	0.670	1.9	6.686	0.150	5.5	244.69	0.00409
0.5	1.649	0.607	2.0	7.389	0.135	6.0	403.43	0.00248
0.6	1.822	0.549	2.2	9.025	0.1108	6.5	665.14	0.00150
0.7	2.014	0.497	2.4	11.023	0.0907	7.0	1 096.6	0.000912
0.8	2.226	0.449	2.6	13.464	0.0743	7.5	1 808.0	0.000553
0.9	2.460	0.407	2.8	16.445	0.0608	8.0	2 981.0	0.000335
1.0	2.718	0.368	3.0	20.086	0.0498	8.5	4 914.8	0.000203
1.1	3.004	0.333	3.2	24.533	0.0408	9.0	8 103.1	0.000123
1.2	3.320	0.301	3.4	29.964	0.0334	9.5	13 360	0.000075
1.3	3.669	0.273	3.6	36.598	0.0273	10.0	22 026	0.000045
1.4	4.055	0.247						

TABLE A4.2

x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$
1.0	0	1.5	0.405	2.2	0.788	3.2	1.163	5.0	1.609	7.5	2.015
1.1	0.0953	1.6	0.470	2.4	0.875	3.4	1.224	5.5	1.705	8.0	2.079
1.2	0.182	1.7	0.531	2.6	0.956	3.6	1.281	6.0	1.792	8.5	2.140
1.3	0.262	1.8	0.588	2.8	1.030	3.8	1.335	6.5	1.872	9.0	2.197
1.4	0.336	2.0	0.693	3.0	1.099	4.5	1.504	7.0	1.946	10.0	2.303

TABLE A4.3

x	$\sin x$	$\cos x$	$\tan x$	x	$\sin x$	$\cos x$	$\tan x$
0	0.000	1.000	0.000	3.2	−0.0584	−0.998	0.0585
0.1	0.0998	0.995	0.100	3.3	−0.158	−0.987	0.160
0.2	0.199	0.980	0.203	3.4	−0.256	−0.967	0.264
0.3	0.296	0.955	0.309	3.5	−0.361	−0.936	0.375
0.4	0.389	0.921	0.423	3.6	−0.443	−0.897	0.493
0.5	0.479	0.878	0.546	3.7	−0.530	−0.848	0.625
0.6	0.565	0.825	0.684	3.8	−0.612	−0.791	0.774
0.7	0.644	0.765	0.842	3.9	−0.688	−0.726	0.947
0.8	0.717	0.697	1.030	4.0	−0.757	−0.654	1.158
0.9	0.783	0.622	1.260	4.1	−0.818	−0.575	1.424
1.0	0.841	0.540	1.557	4.2	−0.872	−0.490	1.778
1.1	0.891	0.454	1.965	4.3	−0.916	−0.401	2.286
1.2	0.932	0.362	2.572	4.4	−0.952	−0.307	3.096
1.3	0.964	0.268	3.602	4.5	−0.978	−0.211	4.637
1.4	0.985	0.170	5.798	4.6	−0.994	−0.112	8.860
1.5	0.997	0.0707	14.101	4.7	−1.000	−0.0124	80.713
1.6	0.9996	−0.0292	−34.233	4.8	−0.996	0.0875	−11.385
1.7	0.992	−0.129	−7.697	4.9	−0.982	0.187	−5.267
1.8	0.974	−0.227	−4.286	5.0	−0.969	0.284	−3.881
1.9	0.946	−0.323	−2.927	5.1	−0.926	0.378	−2.449
2.0	0.909	−0.416	−2.185	5.2	−0.883	0.469	−1.886
2.1	0.863	−0.505	−1.710	5.3	−0.832	0.554	−1.501
2.2	0.808	−0.589	−1.374	5.4	−0.773	0.635	−1.218
2.3	0.746	−0.666	−1.119	5.5	−0.706	0.709	−0.996
2.4	0.675	−0.737	−0.916	5.6	−0.631	0.776	−0.814
2.5	0.598	−0.801	−0.747	5.7	−0.551	0.835	−0.660
2.6	0.516	−0.857	−0.602	5.8	−0.465	0.886	−0.525
2.7	0.427	−0.904	−0.473	5.9	−0.374	0.927	−0.403
2.8	0.335	−0.942	−0.356	6.0	−0.279	0.960	−0.291
2.9	0.239	−0.971	−0.246	6.1	−0.182	0.983	−0.185
3.0	0.141	−0.990	−0.143	6.2	−0.0831	0.997	−0.0834
3.1	0.0416	−0.999	−0.0416	6.3	−0.0168	1.000	0.0168

Appendix 5. The International System, or SI

Quantity		SI unit	
Name	Dimensions	Name of unit	Unit symbol
	Base units		
length	m	meter	m
mass	kg	kilogram	kg
time	s	second	s
electric current	A	ampere	A
temperature	K	kelvin	K
luminous intensity	cd	candela	cd
amount of substance	mol	mole	mol
	Additional units		
plane angle	—	radian	rad
solid angle	—	steradian	sr, sterad
	Derived units		
velocity	m/s	meter per second	m/s
acceleration	m/s^2	meter per second squared	m/s^2
angular velocity	1/s	radian per second	rad/s
angular acceleration	$1/s^2$	radian per second squared	rad/s^2
force	$kg \cdot m/s^2$	newton	N
pressure	$kg/m \cdot s^2$	pascal	Pa
work, torque, energy, quantity of heat	$kg \cdot m^2/s^2$	joule	J
power, heat flux	$kg \cdot m^2/s^3$	watt	W
heat capacity (specific)	$m^2/s^2 \cdot K$	joule per kilogram kelvin	$J/kg \cdot K$
momentum	$kg \cdot m/s$	kilogram-meter per second	$kg \cdot m/s$
impulse	$kg \cdot m/s$	newton-second	$N \cdot s$
frequency	1/s	hertz	Hz
quantity of electricity	$A \cdot s$	coulomb	C
electromotive force, potential difference	$kg \cdot m^2/A \cdot s^3$	volt	V
electric field strength	$kg \cdot m/A \cdot s^3$	volt per meter	V/m
electric resistance	$kg \cdot m^2/A^2 \cdot s^3$	ohm	Ω
electric capacitance	$A^2 \cdot s^4/kg \cdot m^2$	farad	F
inductance	$kg \cdot m^2/A^2 \cdot s^2$	henry, weber per ampere	H, Wb/A
magnetic flux	$kg \cdot m^2/A \cdot s^2$	weber	Wb
magnetic field strength	A/m	ampere per meter	A/m
magnetic flux density, magnetic induction	$kg/A \cdot s^2$	tesla, weber per square meter	T, Wb/m^2
magnetomotive force	A	ampere	A

Appendix 6. Greek Alphabet

A α	Alpha	I ι	Iota	P ρ	Rho
B β	Beta	K κ	Kappa	Σ σ	Sigma
Γ γ	Gamma	Λ λ	Lambda	T τ	Tau
Δ δ	Delta	M μ	Mu	Υ υ	Upsilon
E ε	Epsilon	N ν	Nu	Φ φ	Phi
Z ζ	Zeta	Ξ ξ	Xi	X χ	Chi
H η	Eta	O o	Omicron	Ψ ψ	Psi
Θ θ	Theta	Π π	Pi	Ω ω	Omega

Hints, Answers, and Solutions

Part 1

Chapter 1

1.2.4. $r = \sqrt{2}$, $\alpha = 45°$; $r = 2\sqrt{2}$, $\alpha = -45°$; $r = 3\sqrt{2}$, $\alpha = -135°$; $r = 4\sqrt{2}$, $\alpha = 135°$. **1.2.5.** 2; $2\sqrt{2}$; $2\sqrt{2}$; $2\sqrt{2}$. **1.2.6.** $(0, a/\sqrt{2})$; $(a/\sqrt{2}, 0)$; $(0, -a/\sqrt{2})$; $(-a/\sqrt{2}, 0)$ (make a drawing). **1.2.7.** $(a, 0)$; $(a/2, a\sqrt{3}/2)$; $(-a/2, a\sqrt{3}/2)$; $(-a, 0)$; $(-a/2, -a\sqrt{3}/2)$; $(a/2, -a\sqrt{3}/2)$ (make a drawing). **1.2.8.** (a) Two cases: $(-a/2, 0)$, $(a/2, 0)$, $(0, \pm a\sqrt{3}/2)$. (b) Four cases: $(0, 0)$, $(a, 0)$, $(a/2, \pm a\sqrt{3}/2)$ and $(0, 0)$ $(-a, 0)$ $(-a/2, \pm a\sqrt{3}/2)$ (make a drawing). **1.2.9.** $A_2(x_1, -y_1)$, $A_3(-x_1, y_1)$, $A_4(-x_1, -y_1)$ (make a drawing).

1.3.3. (a) $y = -x$; (b) $y = 2x - 1$. **1.3.4.** If the straight line intersects the coordinate axes at the points M and N, then a and b taken with the appropriate signs are the lengths of the line segments OM and ON.

1.4.2. Use the fact that $x^2 - 2x + 2 = (x-1)^2 + 1$; $2x^2 + 4x = 2(x+1)^2 - 2$. **1.4.3.** The stretching from the origin (homothety) with a coefficient $\sqrt{k}$ transforms point $M(x, y)$ to point $M'(x', y')$, with $x' = \sqrt{k}\,x$, $y' = \sqrt{k}\,y$. Whence, $x'y' = kxy = k$. (b) The stretching along the x axis with the coefficient k transforms point $M(x, y)$ to point $M_1(x_1, y_1)$, with $x_1 = kx$, $y_1 = y$ (whence $x = \frac{1}{k}x_1$, $y = y_1$). **1.4.4.** We must prove that for $k, x_1, x_2 > 0$ the following inequality holds true $\frac{1}{2}\left(\frac{k}{x_1} + \frac{k}{x_2}\right) > \frac{k}{(x_1+x_2)/2}$, or $\frac{x_1+x_2}{2x_1x_2} > \frac{2}{x_1+x_2}$.

1.5.1. (c) The graph of the function $y = x^4 + x^2$ can be obtained by the "addition" of the graphs we have known of the functions $y_1 = x^4$ and $y_2 = x^2$; in shape it resembles these graphs. **1.5.3.** See Figure H. 1*a-d* corresponding to the limiting cases $n \to \infty$; for large n's the graphs are close to those given in Figure H.1.

1.6.1. (a) $y = \frac{x}{2} - 2$; (b) $y = \pm\sqrt{x+2}$; (c) $y = \sqrt[3]{x+1} - 1$; (d) $y = \pm\sqrt{\sqrt{x+1} - 1}$.

1.6.3. (a) $y = \frac{x}{a} - \frac{b}{a}$; (b) $y = \pm\sqrt{\frac{x - (c - b^2/4a)}{a}} - \frac{b}{2a}$; (c) $y = \frac{dx-b}{-cx+a}$.

1.7.8. Here we can easily prove in a purely formal manner that x and y appearing in the conditions for the substitution of the variable preserve the sign of aB (see the solution to Exercise 1.7.9). from which it immediately follows that the curves (1.7.10a), (1.7.10b). and (1.7.10c) cannot approach one another. A more interesting ("geometrical") reasoning is also possible here. It is easy to see that neither the translation of the coordinate origins, nor the change in scales along the x and y axes (which is equivalent to uniform stretching of the curve along the axes) affect the presence or absence of (local) maxima or minima of the curve $y = f(x)$ and also the presence or absence of salient points at which the tangent is horizontal. From this and from the fact that the curve (1.7.10c) has two maxima (the monotonically increasing curves (1.7.10a) and (1.7.10b) have no maxima) and the curve (1.7.10a) has the salient point (the tangent at which is horizontal)—the origin—the required assertion follows immediately. **1.7.9.** At $B = c - \frac{b^2}{3a} = 0$ the relation (1.7.9) reduces to the form (1.7.10a); if $B \neq 0$, this relation reduces to the form (1.7.10b) for $aB = ac - \frac{b^2}{3} > 0$, and to the form (1.7.10c) for $ac - \frac{b^2}{3} < 0$.

1.8.1. (a) The curve is represented in Figure H.2. **1.8.2.** The circle $x^2 + y^2 = 1$. **1.8.3.** An ellipse (Figure H.3). **1.8.4.** The line segment of the straight line $y = x$ between the points $(-1, -1)$ and $(1, 1)$. **1.8.5.** $x = r\arccos\frac{r-y}{r} - \sqrt{2ry - y^2}$. **1.8.6.** (a) The

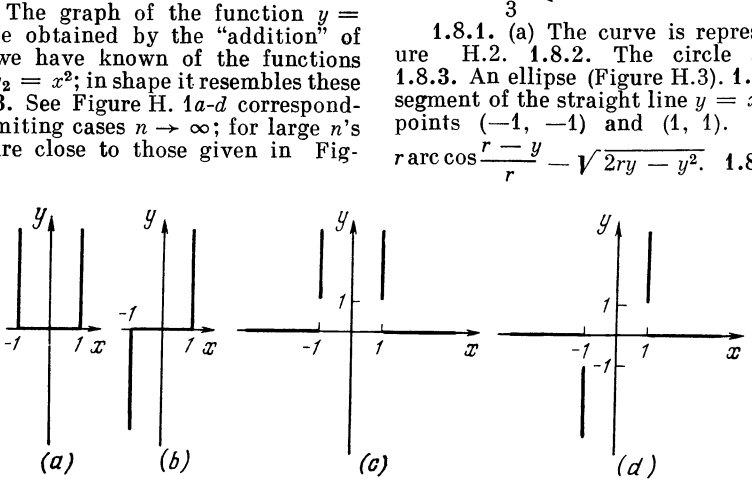


Figure H.1

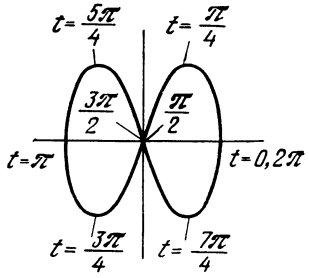


Figure H.2

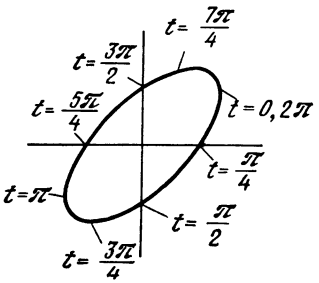


Figure H.3

equations of the epicycloid $x = (R + r)\cos t - r\cos\frac{R+r}{r}t$, $y = (R + r)\sin t - r\sin\frac{R+r}{r}t$. (b) The equations of the hypocycloid $x = (R - r)\cos t + r\cos\frac{R-r}{r}t$, $y = (R - r)\times \sin t - r\sin\frac{R-r}{r}t$. Figure H.4*a-d* shows respectively: the epicycloid for $R = r$ (cardioid); the hypocycloid for $r = (1/3)R$ (the Steiner curve); the hypocycloid for $r = (1/4)R$ (astroid); the hypocycloid for $r = (1/2)R$ (the diameter of a circle of radius R).

1.9.1. (a) $y = -x$; (b) $x = 0$; (c) $y = x + 2$; (d) $y = (-4/3)x$. **1.9.2.** For the first triple of points: $\angle A_2A_1A_3 = 90°$; $S_{\Delta A_1A_2A_3} = 25$; $h_{A_1A_2A_3} = 2\sqrt{5}$. **1.9.3.** (a) $2\sqrt{2}$. **1.9.4.** (a) $3\sqrt{2}/2$. **1.9.7.** If $y = kx + s$ is the equation of the straight line, then in the new coordinates x', y' related to the old coordinates x, y by formulas (1.9.6) the equation of the same straight line will have the form $y' = k'x' + s'$, where

$$k' = \frac{k\cos\alpha' + \sin\alpha'}{\cos\alpha' - k\sin\alpha'}, \quad s' = \frac{s + ka' - b'}{\cos\alpha' - k\sin\alpha'};$$

here $\alpha' = -\alpha = \angle x'Ox$ is the angle between the $O'x'$ and Ox axes, and a' and b' are the coordinates of the new origin O' in the old system xOy.

Chapter 2

2.1.1. (a) $v_{av} = 3t^2 + 1 + 3t\Delta t + (\Delta t)^2$; $v_{inst} = 3t^2 + 1$. **2.1.2.** $v_{av} = v_{inst}$.

2.2.1. $c_{av} = 0.3965 + 4.162 \times 10^{-3}T - 15.072 \times 10^{-7}T^2 + (2.081 \times 10^{-3} - 15.072 \times 10^{-7}T)\Delta T - 5.024 \times 10^{-7}(\Delta T)^2$; $c_{inst} = 0.3965 + 4.162 \times 10^{-3}T - 15.072 \times 10^{-7}T^2$; $c(0) = 0.3965$; $c(100) = 0.7977$; $c(500) = 2.1007$. **2.2.2.** $c_{inst} = 4186.68 + 16746.72 \times 10^{-5}T + 3768 \times 10^{-6}T^2$.

2.4.1. (a) $y' = 4x^3$; (b) $y' = 12x^2 - 6x + 2$; (c) $y' = 8x + 4$; (d) $y' = -2/x^3$; (e) $y' = a(1 - 1/x^2)$; (f) $y' = 2ax - 2b/x^3$. **2.4.4.** $(1.2)^2 = 1.44$; if $y(x) = x^2$, $x = 1$, $\Delta x = 0.2$, then $y(x) + y'(x)\Delta x = 1^2 + 2 \times 0.2 = 1.4$; the error is 3%. Similarly, $(1.1)^2 = 1.21$ and $(1.1)^2 \simeq 1^2 + 2 \times 0.1 = 1.2$ (the error is 1%) and so on.

2.5.2. The tangent is horizontal at the points $x = 0$ and $x = 2$. **2.5.3.** See Figure H.5; the tangent is horizontal at $x = \pm 1/\sqrt{3}$ **2.5.4.** See Figure H.6. **2.5.5.** See Figure H.7. **2.5.6.** (a) $y = (3/4)x - 1/4$, $(1/3, 0)$, $(0, -1/4)$; (b) $y = 3x - 2$, $(2/3, 0)$, $(0, -2)$. **2.5.7.** The tangent to the parabola (a) at the point (x_0, y_0) intersects the coordinate axes at the points $(x_0/2, 0)$ and $(0, -y_0)$; the tangent to the cubical parabola (b) intersects the coordinate axes at the points $((2/3)x_0, 0)$, $(0, -2y_0)$.

2.6.1. (a) $x = 0$; the minimum at $a > 0$, the maximum at $a < 0$; (b) $x = -1$, the maximum; $x = 1$, the minimum; (c) $x = -\sqrt{b/a}$ $(b/a > 0)$, the maximum at $a > 0$, the minimum at $a < 0$, $x = \sqrt{b/a}$ $(b/a > 0)$, the maximum at $a < 0$, the minimum at $a > 0$. There is no maximum or minimum at $b/a \leqslant 0$;

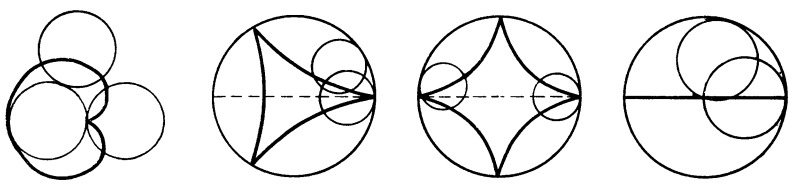


Figure H.4

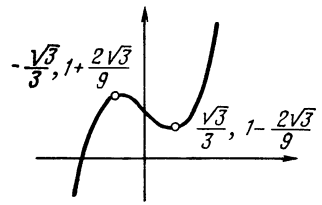


Figure H.5

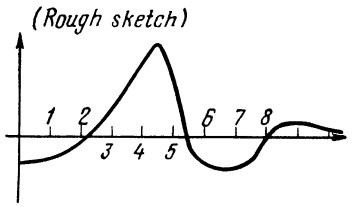


Figure H.6

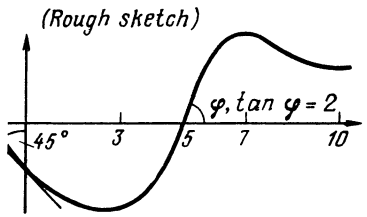


Figure H.7

(d) $x=-1$, the maximum; $x=1$, the minimum; (e) $a>0$, $x=0$, the minimum; $a=0$, $x=0$, the minimum; $a<0$, $x=0$, the maximum, $x=-\sqrt{-a/2}$, the minimum, $x=\sqrt{-a/2}$, the minimum. **2.6.2.** $t=4°$. **2.6.3.** (*a*) $t\simeq 4.08°$, (b) $t\simeq 3.92°$.

2.7.1. $2; 6x; 12x^2; 2a$. **2.7.2.** The acceleration equals $2a$, it is constant. **2.7.3.** (a), (c) The curve is convex at $x<0$, concave at $x>0$; (b) convex at $x<-p/3$, concave at $x>-p/3$.

Chapter 3

3.2.2. Below are given the values of the sums when the interval is partitioned into m parts ($m=10, 20, 50, \infty$):

m	10	20	50	∞
Δt	0.1	0.05	0.02	0
$\Sigma t_{l-1}^2\times\Delta t$	2.18	2.26	2.30	2.33
$\Sigma t_l^2\times\Delta t$	2.49	2.41	2.37	2.33

As seen, even at $m=50$ both sums differ but little from the limiting value at $m=\infty$.

For an exact evaluation of the integral (the limiting sum corresponding to the value $m=\infty$) we write

$$\sum_l t_l^2\frac{1}{m}=\sum_{l=1}^{m}\left(\frac{l}{m}\right)^2\frac{1}{m}$$

$$=(1^2+2^2+\ldots+m^2)\frac{1}{m^3}=\frac{m(m+1)(2m+1)}{6m^3}$$

$$=\frac{2m^2+3m+1}{6m^2}=\frac{1}{3}+\frac{1}{2}\frac{1}{m}+\frac{1}{6}\frac{1}{m^2},$$

whence it follows immediately that $\int_1^2 t^2\,dt=\frac{1}{3}$ ($\simeq 0.33$). **3.2.3.** Let $q=\sqrt[m]{2}$; we denote $x_0=1$, $x_1=q$, $x_2=q^2$, ..., $x_m=q^m=2$. Further $\sum_l x_l^3\Delta x_l=\sum_l (x_l)^3(x_l-x_{l-1})=\sum_l (q^l)^3\times(q^l-q^{l-1})=(q-1)\sum_{l=1}^{m} q^{4l-1}=(q-1)(q^3+q^7+q^{11}+\ldots+q^{4m-1})=(q-1)\frac{q^{4m+3}-q^3}{q^4-1}=\frac{q^3[(q^m)^4-1]}{q^3+q^2+q+1}=\frac{q^3(2^4-1)}{q^3+q^2+q+1}$; sending $m\to\infty$ (and $q\to 1$), we get $\int_1^2 x^3\,dx=\frac{15}{4}=3\frac{3}{4}=3.75$.

3.4.1. (a) 1/3; (b) $\frac{1}{3}(1.1^3-1)\simeq 0.11033$; (c) 1/2; (d) $2(\sqrt{3}-1)\simeq 1.464$. **3.4.2.** (a) $S=\int_0^b y(x)\,dx$, with $y=y(x)=\frac{h}{b}x$ the equation of the hypotenuse. Whence $S=\int_0^b \frac{h}{b}x\,dx=\frac{h}{b}\int_0^b x\,dx=\frac{h}{b}\frac{x^2}{2}\Big|_0^b=\frac{1}{2}bh$. **3.4.3.** (a) $S=\frac{1}{3}x_0y_0$; (b) $S=\frac{2}{3}x_0y_0$. **3.4.4.** $S=\int_{-r}^{r}\sqrt{r^2-x^2}\,dx$. **3.4.5.** 0.7837 for $n=5$; 0.7850 for $n=10$. **3.4.6.** See Figure H.8. **3.4.7.** See Figure H.9*a-c*.

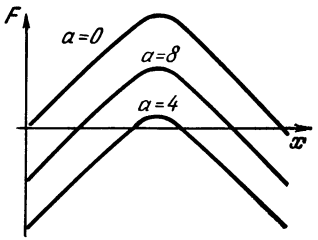


Figure H.8

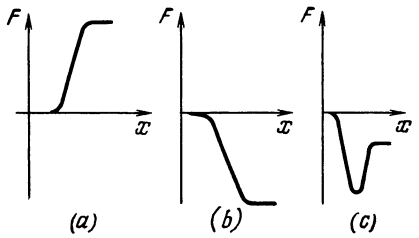


Figure H.9

3.6.2. It does not change. **3.6.3.** $(2/3)\,R^3\times\tan\alpha$. **3.6.4.** $\int_0^1 y\sqrt{1-y^2}\,dy = 1/3$; note that the area of the cross section of the "hoof" (Figure 3.6.8) with the plane, which is perpendicular to the plane ABP, parallel to AB, and distant by y from AB, is equal to $2y\sqrt{1-y^2}$.

Chapter 4

4.2.1. (a) $y' = (x^2)'\,x^2 + x^2\,(x^2)' = 2\times 2x\times x^2 = 4x^3$; (b) $y' = 5ax^4 + 4bx^3 + 3cx^2 + 2dx + e$; (c) $(5x + 3/2)\sqrt{x}$. **4.2.2.** $f'''(x) = g(x)\,h'''(x) + 3g'(x)\,h''(x) + 3g''(x)\,h'(x) + g'''(x)\,h(x)$. **4.2.3.** $f''(x) = ghk'' + gh''k + g''hk + 2gh'k' + 2g'hk' + 2g'h'k$.

4.3.1. $\dfrac{dz}{dx} = \dfrac{dz}{dy}\dfrac{dy}{dx} = 2ya = 2a\,(ax + b)$; $z' = (ax^2 + 2abx + b^2)' = 2a^2x + 2ab$.

4.3.2. (a) $z' = -\dfrac{a}{(ax+b)^2}$; (b) $z' = -\dfrac{2a}{(ax+b)^3}$; (c) $z' = \dfrac{1}{(x+1)^2}$; (d) $y' = \dfrac{2x\,(x^2+4x+5)}{(x+1)^2}$; (e) $y' = -\dfrac{x^2-2x-2}{(x^2+2)^2}$. **4.3.3.** $y'' = \dfrac{f''g^2 - 2f'gg' + 2f\,(g')^2 - fgg''}{g^3}$.

4.4.1. (a) $y' = \dfrac{2}{3}\dfrac{1}{\sqrt[3]{(2x+1)^2}}$; (b) $y' = \dfrac{1}{(x+1)\sqrt{x^2-1}}$; (c) $y' = -\dfrac{1}{4}\dfrac{1}{x\sqrt[4]{x}}$. **4.4.2.** $\dfrac{dy}{dx} = -\dfrac{b}{a}\cot t$. **4.4.3.** (b) $\dfrac{d^2y}{dx^2} = \dfrac{5}{16}\dfrac{1}{x^2\sqrt[4]{x}}$; (c) $\dfrac{d^2y}{dx^2} = -\dfrac{b}{a^2}\dfrac{1}{\sin^3 t}$.

4.5.1. (a) $y' = 5x^4 - 12x^3 + 3x^2 + 14x - 2$; (b) $y' = 2\,(x^3+x+1)\,(3x^2+1)$; (c) $y' = 4\,(x^2 - x + 1)^3\,(2x - 1)$; (d) $y' = 10\,(3x^2 - 1)^9\,6x$; (e) $y' = \dfrac{1}{\sqrt{x^2-1}}$; (f) $y = x^{2/5}$, $y' = \dfrac{2}{5}\times\dfrac{1}{\sqrt[5]{x^3}}$. **4.5.3.** (a) If x changes by 1%, then $\Delta y = n\times 0.01y$; therefore when x changes by k%, we have $\Delta y = n\times 0.01yk$. In the given case x changes by 10% an $\Delta y = n\times 0.1y$. Since $n = 1/2$, $\Delta y = \dfrac{1}{2}\,0.1y = 0.05y$. Therefore $y\,(11) \simeq y\,(10) + 0.05\times 5 = 5.25$; $y\,(9) \simeq y\,(10) - 0.05\times 5 = 4.75$. Let us now obtain the exact solution. We denote the proportionality factor by k; then $y = k\sqrt{x}$. Since at $x = 10$, $y = 5$, we have $5 = k\sqrt{10}$, whence $k \simeq 1.58$ (the calculations are carried out to two decimal places). Therefore $y = 1.58\sqrt{x}$; $y\,(11) = 1.58\sqrt{11} \simeq 5.24$; $y\,(9) \simeq 4.74$. (b) The approximate values are $y\,(11) \simeq 4.50$; $y\,(9) \simeq 5.50$. The exact values are $y\,(11) \simeq 4.54$; $y\,(9) \simeq 5.56$. (c) The approximate values are $y\,(11) \simeq 6.00$; $y\,(9) \simeq 4.00$. The exact values are $y\,(11) \simeq 6.05$; $y\,(9) \simeq 4.05$.

4.6.1. $y' = x^2\,(x^2-1)\,(7x^2-3)$. **4.6.2.** $y' = 3x^2\sqrt{x^2+x} + \dfrac{(2x+1)\,x^3}{2\sqrt{x^2+x}}\left(= \dfrac{(8x+7)\,x^3}{2\sqrt{x^2+x}}\right)$. **4.6.3.** $y' = 5x^4\sqrt[3]{x^2-1}\,(x^3 - 2x)^{1/5} + \dfrac{1}{3}\times\dfrac{\sqrt[3]{x^2-1}}{x^2-1}\cdot 2x x^5\,(x^3-2x)^{1/5} + \dfrac{1}{5}\dfrac{(x^3-2x)^{1/5}}{x^3-2x}\times(3x^2-2)\,x^5\sqrt[3]{x^2-1}$ (simplify). **4.6.4.** $y' = \left(1 - \dfrac{1}{2\sqrt{x^3}}\right)\sqrt{x^3-2} + \left(x + \dfrac{1}{\sqrt{x}}\right)\times\dfrac{3x^2}{2\sqrt{x^3-2}}$. **4.6.5.** $y' = 2x\sqrt{\sqrt[3]{x}+x} + x^2\times\dfrac{1/3\sqrt[3]{x^2}+1}{2\sqrt{\sqrt[3]{x}+x}}$ (simplify). **4.6.6.** $y' = 5\left(\sqrt[3]{x} + \dfrac{1}{\sqrt[3]{x}}\right)^4\left(\dfrac{1}{3\sqrt[3]{x^2}} - \dfrac{1}{3\sqrt[3]{x^4}}\right)x + \left(\sqrt[3]{x} + \dfrac{1}{\sqrt[3]{x}}\right)^5$. **4.6.7.** $y' = \dfrac{x^2+1}{(x^2-1)^2}$. **4.6.8.** $y' = 2\dfrac{1-x^2}{(x^2-x+1)^2}$. **4.6.9.** $y' = \dfrac{x^2+2x+5}{(x+1)^2}$. **4.6.10.** $y' = \dfrac{-12x+5}{x^6}\sqrt{x^3+2} + \dfrac{3\,(3x-1)}{2x^2\sqrt{x^3+2}}$.

4.6.11. $y' = -\frac{1}{(x^2-1)^{3/2}}$. **4.6.12.** $y' = \frac{2x+3}{3(x+1)^{4/3}}$. **4.6.13.** $y' = \frac{4x+3\sqrt{x}}{4\sqrt{x^2+x\sqrt{x}}}$. **4.6.14.** $y' = \frac{6x\sqrt[3]{x^2}+1}{6\sqrt[3]{x^2}\sqrt{x^2+\sqrt[3]{x}}}$. **4.6.15.** $y' = \frac{(8x-x^3)\sqrt{1+x^2}}{(1+x^2)^3}$. **4.6.16.** $y' = \sqrt[3]{(2x+3)^2} + x\frac{4}{3\sqrt[3]{2x+3}}$. **4.6.17.** $y' = 3x^2\sqrt{x-1} + \frac{x^3-1}{2\sqrt{x-1}} + \sqrt[3]{x^2-1} + \frac{2x^2}{3\sqrt[3]{(x^2-1)^2}}$. **4.6.18.** $y' = \frac{-2x^2-x+9}{3(x-1)^3\sqrt[3]{2x-3}}$. **4.6.19.** $y' = \sqrt{\frac{x+1}{x-1}} \times \frac{1}{(x+1)^2}$. **4.6.20.** $y' = \frac{4x^3-10x^2-22x-11}{3(x-2)^2\sqrt[3]{(x+1)^2}}$. **4.6.21.** $y' = \frac{-x^5+2x^3+2x^2-1}{(x^3+1)^2\sqrt{x^2-1}}$. **4.6.22.** $y' = \frac{1}{3}\left(\frac{x+1}{x^2+x+1}\right)^{2/3}\frac{x^2+2x}{(x+1)^2}$. **4.6.23.** $y' = \sqrt{x^2-1}\sqrt[3]{x+\sqrt{x}} + \frac{x^2}{\sqrt{x^2-1}} \times \sqrt[3]{x+\sqrt{x}} + \frac{(1+2\sqrt{x})x\sqrt{x^2-1}}{6\sqrt{x}\sqrt[3]{(x+\sqrt{x})^2}}$. **4.6.24.** $y' = \frac{1}{7}\left(x+\frac{1}{\sqrt{x}}\right)^{-6/7}\left(1-\frac{1}{2\sqrt{x^3}}\right)x^2 + 2x\left(x+\frac{1}{\sqrt{x}}\right)^{1/7}$. **4.6.25.** $y' = \frac{1-6\sqrt[3]{x^2}-4x\sqrt[3]{x^2}}{3(x+1)\sqrt[3]{x^2(x+1)}}$.

4.7.1. $y' \simeq 2.3\log 2 \times 2^x$. **4.7.2.** $y' \simeq 2.3\log 5 \times 5^{x+1}$. **4.7.3.** $y' \simeq -2.3\log 2\left(\frac{1}{2}\right)^x$. **4.7.4.** $y' \simeq 2.3\times 10^{\sqrt{x}}\frac{1}{2\sqrt{x}}$. **4.7.5.** $y' \simeq 2x \times 2.3\log 2 \times 2^{x^2}$. **4.7.6.** $y' \simeq \left(1-\frac{1}{x^2}\right) \times 2\,3\log 2 \times 2^{x+1/x}$.

4.8.1. (a) $y' = -e^{-x}$; (b) $y' = 5e^x - 3e^{3x}$; (c) $y' = 2xe^{x^2}$; (d) $y' = \frac{1}{2\sqrt{x}}e^x$; (e) $y' = 3(x^2-1)e^{x^3-3x+1}$. **4.8.2.** More than by a factor of 150 000.

4.9.1. $\log_5 15 = \frac{\log 15}{\log 5} \simeq 1.6825$. **4.9.2.** (a) Differentiating both sides of the equation (the logarithms are natural) we find $\frac{(uv)'}{uv} = \frac{u'}{u} + \frac{v'}{v}$, whence $(uv)' = u'v + uv'$. (b) Differentiate term-by-term the equation $\ln(u/v) = \ln u - \ln v$. **4.9.3.** (a) $y' = \frac{1}{x+3}$; (b) $y' = \frac{1}{2x}(2x)' = \frac{1}{x}$, or $y = \ln 2x = \ln 2 + \ln x$, therefore $y' = (\ln 2)' + (\ln x)' = \frac{1}{x}$; (c) $y' = \frac{2x}{x^2+1}$; (d) $y' = \frac{x^2-1}{x(x^2+1)}$; (e) $y' = \frac{6x-1}{3x^2-x+1}$; (f) $y' = \frac{2}{x^2-1}$; (g) $y' = \frac{x+3}{6x(x+1)}$; (h) $y' = \ln x + 1$; (i) $y' = 3x^2\ln(x+1) + \frac{x^3}{x+1}$; (j) since $\ln y = x\ln x$, differentiating we find $\frac{1}{y}y' = \ln x + 1$, whence $y' = y(\ln x + 1)$, or $y' = x^x(\ln x + 1)$; (k) $y' = \frac{1}{2}(\sqrt{x})^{\sqrt{x^2-1}} \times \left(\frac{x\ln x}{\sqrt{x^2-1}} + \frac{\sqrt{x^2-1}}{x}\right)$ (see the hint to (j)).

4.10.1. $y' = 2\cos(2x+3)$. **4.10.2.** $y' = -\sin(x-1)$. **4.10.3.** $y' = -(2x-1)\sin(x^2-x+1)$. **4.10.4.** $y' = 2\sin x\cos x$. **4.10.5.** $y' = 3\cos 3x\cos^2 x - 2\cos x\sin x\sin 3x$. **4.10.6.** $y' = (\sin 2x)^x\left(\ln\sin 2x + \frac{2x\cos 2x}{\sin 2x}\right)$ (cf. Exercise 4.9.3j). **4.10.7.** $y' = \tan x + \frac{x}{\cos^2 x}$. **4.10.8.** $y' = \frac{2}{\cos^2 2x}e^{\tan 2x}$. **4.10.9.** $y' = -\frac{1}{2}\frac{1}{\sin^2(x/2)}$.

4.11.1. $y' = -\frac{1}{\sqrt{1-x^2}}$. **4.11.2.** $y' = -\frac{1}{1+x^2}$. **4.11.3** $y' = \frac{2}{\sqrt{1-4x^2}}$. **4.11.4.** $y' = \frac{3}{9x^2+6x+2}$. **4.11.5.** $y' = \frac{2x-1}{x^4-2x^3+x^2+1}$. **4.11.6.** $y' = \frac{1}{2\sqrt{x}(x+1)}e^{\operatorname{arc\,tan}\sqrt{x}}$.

4.13.1. -1; 1. **4.13.2.** -1.

Chapter 5

5.2.1. 1/6. **5.2.2.** 1/2. **5.2.3.** 2. **5.2.4.** $\frac{\pi}{4}\left(=\operatorname{arc\,tan}1 - \operatorname{arc\,tan}0 = \frac{\pi}{4} - 0\right)$. **5.2.5.** $e - \frac{1}{e}$. **5.2.6.** 1.

5.3.1. $x^3 - x^2 + x + C$. **5.3.2.** $\frac{4}{5}x^5 - \frac{3}{4}x^4 + \frac{x^3}{3} - \frac{x^2}{2} + C$. **5.3.3.** $\frac{1}{4}x^4 - \frac{2}{3}x^3 + \frac{1}{2}x^2 + C$. **5.3.4.** $\frac{1}{2}x^2 + 2x + 3\ln|x| + C$. **5.3.5.** $2x + \ln|x-1| + C$. **5.3.6.** $\frac{a}{c}x + \frac{bc-ad}{c^2}\ln|cx+d| + C$. **5.3.7.** If $\frac{x}{(x-2)(x-3)} = \frac{A}{x-2} + \frac{B}{x-3}$, then $A(x-3)+B(x-2)=x$, i.e. $(A+B)x-(3A+2B)=x$ and $A+B=1$, $3A+2B=0$; therefore $A=-2$, $B=3$. *Answer.* $-2\ln|x-2| + 3\ln|x-3| + C$. **5.3.8.** $-\ln|x-1| + \ln|x-2| + C$. **5.3.9.** If $\frac{x+1}{x^2-3x+3} = \frac{x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}$, then $A(x-2)+B(x-1)=x+1$, whence $A=-2$, $B=3$ (in the previous equality first assume that $x=1$ and then $x=2$). *Answer.* $-2\ln|x-1|+3\ln|x-2| + C$. **5.3.10.** If $\frac{x+2}{x(x^2+1)} = \frac{A}{x} + \frac{B}{x^2+1} + \frac{Cx}{x^2+1}$, then $A(x^2+1)+Bx+Cx^2=x+2$, i.e. $A=2$, $B=1$, $C=-2$. *Answer.* $2\ln|x| + \arctan x - \ln(x^2+1) + C$.

5.4.1. $\cos x + x\sin x + C$. **5.4.2.** $x(\ln|x|-1)+C$. (*Hint.* In formula (5.4.3) put $f=\ln x$, $dg=dx$). **5.4.3.** $\frac{1}{2}x\sin 2x - \left(\frac{1}{2}x^2 - \frac{1}{4}\right)\times\cos 2x + C$. **5.4.4.** $(-x^3-3x^2-6x-6)e^{-x}+C$. **5.4.5.** $(2x+1)\cos x + (x^2+x-1)\sin x + C$. **5.4.6.** Let $(2x^2+1)\cos 3x\,dx = (a_1x^2+b_1x+c_1)\cos 3x + (a_2x^2+b_2x+c_2)\sin 3x$. Differentiating both sides of the equation we get $(2x^2+1)\cos 3x = (2a_1x+b_1)\cos 3x - (3a_1x^2+3b_1x+3c_1)\sin 3x + (2a_2x+b_2)\sin 3x + (3a_2x^2+3b_2x+3c_2)\cos 3x$, i.e. $(2x^2+1)\cos 3x = (-3a_1x^2-3b_1x-3c_1+2a_2x+b_2)\sin 3x + (3a_2x^2+3b_2x+3c_2+2a_1x+b_1)\cos 3x$. Thus we have $2x^2+1 = 3a_2x^2+3b_2x+3c_2+2a_1x+b_1$, $0=-3a_1x^2-3b_1x-3c_1+2a_2x+b_2$, whence $3a_2=2$, $3b_2+2a_1=0$, $3c_2+b_1=1$, $-3a_1=0$, $-3b_1+2a_2=0$, $-3c_1+b_2=0$. From this system of equations we find: $a_1=0$, $b_2=0$, $c_1=0$, $a_2=\frac{2}{3}$, $b_1=\frac{4}{9}$, $c_2=\frac{5}{27}$. *Answer.* $\frac{4}{9}x\cos 3x + \left(\frac{2}{3}x^2 + \frac{5}{27}\right)\sin 3x + C$. **5.4.7.** $x\arcsin x + \sqrt{1-x^2} + C$. **5.4.8.** $x\arctan x - \frac{1}{2}\ln(x^2+1)+C$. **5.4.9.** Set $f=\sin 3x$, $dg=e^{2x}dx$; then integrating by parts yields $\int e^{2x}\sin 3x\,dx = \frac{1}{2}e^{2x}\sin 3x - \frac{3}{2}\int e^{2x}\cos 3x dx$. In the last integral we put $f=\cos 3x$, $dg=e^{2x}dx$ and again integrate by parts to obtain $\int e^{2x}\sin 3x dx = \frac{1}{2}e^{2x}\sin 3x - \frac{3}{2}\left(\frac{1}{2}e^{2x}\cos 3x + \frac{3}{2}\int e^{2x}\sin 3x\,dx\right)$. Considering the last expression as the equation with the unknown $\int e^{2x}\sin 3x dx$ we find $\int e^{2x}\sin 3x dx = \frac{e^{2x}(2\sin 3x - 3\cos 3x)}{13} + C$. **5.4.10.** $\frac{e^x(\cos 2x + 2\sin 2x)}{5} + C$.

5.5.1. $\frac{1}{3}\sin(3x-5)+C$ $\left(\text{put } 3x-5=t,\ dx=\frac{dt}{3}\right)$. **5.5.2.** $-\frac{1}{2}\cos(2x+1)+C$. **5.5.3.** $\frac{2}{9}\sqrt{(3x-2)^3}+C$. **5.5.4.** Since at $\sqrt{x}=z$ we have $x=z^2$ and $dx=2z\,dz$, then $\int\frac{x\,dx}{x+\sqrt{x}} = \int\frac{z^2\cdot 2z\,dz}{z^2+z} = 2\int\frac{z^2\,dz}{1+z} = 2\int\frac{z^2-1+1}{1+z}dz = 2\int\left[(z-1)+\frac{1}{1+z}\right]dz = z^2-2z+2\ln(1+z)+C = x-2\sqrt{x}+2\ln(1+\sqrt{x})+C$. **5.5.5.** $\sqrt{x^2-5}+C$. **5.5.6.** $\frac{\sin^4 x}{4} + C$, or $-\frac{1}{2}\cos^2 x + \frac{1}{4}\cos^4 x + C$. **5.5.7.** $-\frac{1}{3\sin^3 x} + \frac{1}{\sin x} + C$. **5.5.8.** $-\ln|\cos x| + C$ (put $\cos x = t$). **5.5.9.** $\arcsin\frac{x}{a} + C$. **5.5.11.** $\ln|x+\sqrt{a^2+x^2}|+C$. (In addition to the method indicated in p. 168, we can make here an artificial substitution $x=a\tan t$ (i.e. $t=\arctan(x/a)$), $dx=a\,dt/\cos^2 t$ and $a^2+x^2=a^2/\cos^2 t$ or the substitution $\sqrt{a^2+x^2}=z-x$, whence, squaring both sides of this last equation we get $a^2=z^2-2zx$ (the term with x^2 cancels out in both sides of the equation) and thus $x=\frac{z^2-a^2}{2z}$, $\sqrt{a^2+x^2}=z-x=z-\frac{z^2-a^2}{2z}=\frac{z^2+a^2}{2z}$, $dx=\frac{1}{2}\cdot\frac{z2z-(z^2-a^2)}{z^2}dz=\frac{z^2+a^2}{2z^2}dz$, whence $\int\frac{dx}{\sqrt{a^2+x^2}} =$

$\int \left[\frac{(z^2+a^2)}{2z^2} \div \frac{z^2+a^2}{2z}\right] dz = \int \frac{dz}{z}$.)

5.5.12. Set $z = a \tan t$, then $x^2+a^2 = \frac{a^2}{\cos^2 t}$, $dx = \frac{a\,dt}{\cos^2 t}$, so that $\int \frac{dx}{(x^2+a^2)^2} = \frac{1}{a^3} \times \int \cos^2 t\,dt = \frac{1}{2a^3} \int (1+\cos 2t)\,dt = \frac{1}{2a^3}\left(t + \frac{1}{2}\sin 2t\right) + C$. Since $t = \arctan \frac{x}{a}$ and $\sin 2t = \frac{2\tan t}{1+\tan^2 t} = \frac{2x/a}{1+(x/a)^2} = \frac{2ax}{x^2+a^2}$, we finally have $\int \frac{dx}{(x^2+a^2)^2} = \frac{1}{2a^3} \times \arctan \frac{x}{a} + \frac{1}{2a^2}\frac{x}{x^2+a^2} + C$.

Chapter 6

6.1.1. $y = (ax_0^3 + bx_0^2 + cx_0 + d) + (3ax_0^2 + 2bx_0 + c)(x - x_0) + (3ax_0 + b) \times (x - x_0)^2 + a(x - x_0)^3$. All the subsequent terms equal zero; the sum of the above four terms equal the polynomial. **6.1.2.** Since $y(0) = 0$, $y'(0) = 1$, $y''(0) = 2, \ldots, y^{(n)}(0) = n, \ldots$, $y = x + x^2 + \frac{x^3}{2!} + \ldots = x\left(1 + x + \frac{x^2}{2!} + \ldots\right)$. **6.1.3.** $y = e\,[1 + (x-1) + \frac{1}{2!}(x-1)^2 + \frac{1}{3!}(x-1)^3 + \ldots]$. **6.1.4.** It is easy to verify that $(1+r)^m \simeq 1 + mr + \frac{m(m-1)}{2}r^2$ (cf. Section 6.4), $e^{mr} \simeq 1 + mr + \frac{m^2r^2}{2}$. Thus, for small r the quantity e^{mr} differs from $(1+r)^m$ by the term $(1/2)mr^2$. For example, if $m = 50$, $r = 0.02$ (see the example in p. 143), then $(1/2)mr^2 = 0.01$, i.e. the error is less than 0.5%. (When mr^2 is small, m can be large; here mr^3, mr^4, ... will of course be small.) **6.1.5.** (a) By formula (6.1.24), with $y(0) = 1$ and $\Delta y = y(\Delta x) - 1$, we have

Δx	1	1/2	1/4	1/8	$\Delta x \to 0$
$y(\Delta x)$	2.7183	1.6487	1.2840	1.1331	1
$\frac{\Delta y}{\Delta x}$	1.718	1.297	1.136	1.065	1

By formula (6.1.25), setting $\Delta y = y\left(\frac{\Delta x}{2}\right) - y\left(-\frac{\Delta x}{2}\right)$ we get

Δx	1	1/2	1/4	1/8	$\Delta x \to 0$
$y\left(\frac{\Delta x}{2}\right)$	1.6487	1.2840	1.1331	1.0645	1
$y\left(-\frac{\Delta x}{2}\right)$	0.6065	0.7788	0.8825	0.9394	1
Δy	1.0422	0.5052	0.2506	0.1251	Δx
$\frac{\Delta y}{\Delta x}$	1.042	1.010	1.002	1.006	1

6.1.6. Substitute into the numerator on the right-hand side of (6.1.27) expression (6.1.13a) for $f(a + \Delta x)$ and $f(a - \Delta x)$ (in (6.1.13a) set $x = a + \Delta x$ and $x = a - \Delta x$ respectively).

6.2.1. Apply Taylor's formula (6.1.18) to the function (6.1.1) (or Maclaurin's formula (6.1.19) to the function (6.1.1a)). **6.2.2.** si $x = C + x - \frac{x^3}{18} + \frac{x^5}{600} - \ldots$, where $C = 0$ if we assume that si $0 = 0$.

6.3.1. (a) $y = \frac{x+1}{1-x} = 1 + 2x + 2x^2 + 2x^3 + 2x^4 + \ldots$; (b) $y = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$. **6.3.2.** $y = \ln x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \ldots$. In Exercise 6.3.1 the series can be used for computations for $|x| < 1$; in Exercise 6.3.2, for $0 < x < 2$. **6.3.3.** $f(x)\,g(x) = f(0)\,g(0) + [f'(0)\,g(0) + g'(0)\,f(0)]\,x + \frac{1}{2}[f''(0)\,g(0) + 2f'(0)\,g'(0) + g''(0)\,f(0)]\,x^2 + \ldots$.

6.4.2. These are the formulas (6.4.3) for $a = 1$ and $m = \frac{1}{n}$ in which we retained two first (a) and three first (b) terms respectively. **6.4.3.** $\sqrt[3]{1.2} \simeq 1.067$ and $\simeq 1.062$ (the table gives 1.063); $\sqrt[3]{1.1} \simeq 1.033$ and $\simeq 1.032$ (the table gives 1.032). **6.4.5.** For $x = 0$ even $(\sqrt{x})'$ does not exist (becomes infinite).

6.5.1. (a) 1; (b) $-1/2$; (c) 1/3; (d) ∞; (e) 1; (f) 1/2. **6.5.4.** (a) Substitute $t = \frac{1}{x}$ and consider the ratio $f\left(\frac{1}{t}\right) \div g\left(\frac{1}{t}\right)$; use the rule (6.5.2). (b) Prove that $\frac{f(x)-f(x_0)}{g(x)-g(x_0)} = \frac{f'(c)}{g'(c)}$, where c is a number intermediate between x and x_0 (why?); further assuming that x and x_0 are close to a (and are on the same side of a), first let $x \to a$ with x_0 constant (here $\frac{f(x)-f(x_0)}{g(x)-g(x_0)} \to$

$\frac{f(x)}{g(x)}$, where by condition $x \to a$) and then let $x_0 \to a$.

6.6.1. (a) $y = -\frac{1}{x+C}$; (b) $y^2 - kx^2 = a$; (c) $y = Ax^k$; (d) $y = \frac{1}{\sqrt{-2/3\,x^3 + C}}$; (e) $\tan\frac{y}{2} = Ae^{\sin x}$. **6.6.3.** $y = 0$. **6.6.4.** Expand $e^{x^2/2}\Phi(x)$ in a series in powers of x (to this end we must expand $e^{-t^2/2}$ in series in powers of t and then integrate the obtained equation termwise); prove that for an appropriate x_0 the product $e^{x^2/2}\Phi(x)$ can be represented by the series (6.6.23) (where $a_0 = 0$).

6.7.1. $z = z_{st} - (z_{st} - z_0)\exp\left[-\frac{k}{\pi r_0^2}(t - t_0)\right]$; the water level grows, tending asymptotically to the value $z = z_{st}$. **6.7.2.** Find the steady-state value $z = z_{st} = \text{constant}\ (= q_0/k)$ which satisfies the differential equation; then denote $z = z_{st} + z_1$ and form the differential equation for the function $z_1 = z_1(t)$.

Chapter 7

7.1.1. $x = \frac{a+b-\sqrt{a^2+b^2-ab}}{6}$. **7.1.2.** Let x be the side of the rectangle which belongs to the base (equal to a) of the triangle, then the area of the rectangle is $S(x) = xh\left(1 - \frac{x}{a}\right) = hx - \frac{h}{a}x^2$. Solving the equation $S'(x) = 0$ we find $x = \frac{a}{2}$; then the adjacent side of the rectangle equals $\frac{h}{2}$ and the area $S(x) = \frac{ah}{4}$. **7.1.3.** $S = 2R^2$ (the desired rectangle is a square). **7.1.4.** The radius of the base $r = \sqrt[3]{\frac{V}{2\pi}}$, the altitude $h = 2r$. **7.1.5.** $t = \frac{av_1 + bv_2}{v_1^2 + v_2^2}$. **7.1.7.** The time of motion $T = \frac{1}{v_1}\sqrt{a^2+x^2} + \frac{1}{v^2}\sqrt{b^2+(c-x)^2}$, where c is the magnitude of the projection of AB on the line l of interface, a and b are the distance of points A and B to l. The condition $\frac{dT}{dx} = 0$ yields $\frac{x}{v_1\sqrt{a^2+x^2}} = \frac{c-x}{v^2\sqrt{b^2+(c+x)^2}}$ or since $\frac{x}{\sqrt{a^2+x^2}} = \sin\alpha$, $\frac{c-x}{\sqrt{b^2+(c-x)^2}} = \sin\beta$, where α and β are the angles formed by the lines of the body's motion in the I and II media and the perpendicular to the line of the interface, we have $\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}$.

(This is the ***Snellius law***, that is, the point must move in the same way as the light ray does intersecting the boundary between two media with different velocities in the media.) To prove that we have obtained the minimum of T it is sufficient to find $\frac{d^2T}{dx^2}$; it is easy to ascertain that $\frac{d^2T}{dx^2} > 0$ for all x.

7.2.1. $y_{min} = 3$. **7.2.2.** At the moment of time t the steamer S is represented by the line segment l whose endpoints are projected on the line, which corresponds to the bank where the fisherman F sits (we take it for the x axis) at the points with the abscissas $x_1 = v(t - t_0)$, $x_2 = v(t - t_0) - l$. The distance from F to the point M of the steamer S is $\Delta(x) = \sqrt{x^2 + h^2}$, where $x_1 \leqslant x \leqslant x_2$; we must find $D = \min\Delta(x)$. For $t_0 \leqslant t \leqslant t_0 + l/v$ this minimum equals h; it is attained at the point $x = 0$, which corresponds to the condition $\frac{d\Delta}{dx} = 0$; for $t < t_0$ and $t > t_0 + l/v$ there are boundary minima $\sqrt{v^2(t - t_0)^2 + h^2}$ and $\sqrt{[v(t - t_0) - l]^2 + h^2}$ respectively. (This problem can be solved geometrically without resorting to differential calculus.) **7.2.3.** (a) $y_{max} = 0$ at $x = 0$; (b) $y_{max} = 1$ at $x = 0$.

7.4.1. Prove that $y'' < 0$. **7.4.2.** (a) $\left(\frac{x_1^n + x_2^n}{2}\right)^{1/n} > \frac{x_1 + x_2}{2}$ at $n > 1$, $x_1 \neq x_2$; (b) $\left(\frac{x_1^m + x_2^m}{2}\right)^{1/m} < \frac{x_1 + x_2}{2}$ for $0 < m < 1$, $x_1 \neq x_2$; (c) $1 \div \left[\frac{1}{2}\left(\frac{1}{x_1^k} + \frac{1}{x_2^k}\right)\right]^{1/k} > \left(\frac{x_1 + x_2}{2}\right)$ at $k > 0$ and $x_1 \neq x_2$ (and $x_1, x_2 > 0$); (d) $\frac{1}{2}x_1\log x_1 + \frac{1}{2}x_2\log x_2 > \frac{1}{2}(x_1 + x_2)\log[(x_1 + x_2)/2]$ (in all these cases we only give an inequality, which is a particular case of the inequality (7.3.1)).

7.5.1. $\pi/2$. **7.5.2.** $1/6$. **7.5.3.** $2\pi + 4/3$ and $6\pi - 4/3$. **7.5.4.** (a) $a\ln 2$; (b) $\frac{37}{12}a$ (here a is the amount of paint needed for a unit surface area). **7.5.5.** 10π.

7.7.1. $F(a) + F(b) = \int_1^a \frac{dx}{x} + \int_1^b \frac{dx}{x} =$

$\int_1^a \frac{dx}{x} + \int_a^{ab} \frac{dx}{x} = \int_1^{ab} \frac{dx}{x} F(ab)$.

7.8.1. $\bar{y} = \overline{x^2} = \frac{1}{2}\int_0^2 x^2\,dx = \frac{4}{3}$. Here $y(1) = 1 < \bar{y} \simeq 1.33 < \frac{y(0)+y(2)}{2} = 2$. **7.8.2.** (a) $\bar{y} = \frac{1}{6}y(0) + \frac{2}{3}y(1) + \frac{1}{6}y(2) = \frac{2}{3}\times 1 + \frac{1}{6}\times 4 = \frac{4}{3}$; (b) $\int_a^b y\,dx = \left(\frac{1}{3}rx^3 + \frac{1}{2}px^2 + qx\right)\Big|_a^b = \frac{1}{3}r(b^3-a^3) + \frac{1}{2}p(b^2-a^2) + q(b-a) = (b-a)\left[\frac{1}{3}r(b^2+ab+a^2) + \frac{1}{2}p(b+a) + q\right]$. According to Simpson's rule (7.8.3) we have $\int_a^b y\,dx = (b-a)\bar{y} = (b-a)\left[\frac{1}{6}y(a) + \frac{2}{3}y\left(\frac{a+b}{2}\right) + \frac{1}{6}y(b)\right]$. Substituting the values $y(a)$, $y\left(\frac{a+b}{2}\right)$, and $y(b)$ and comparing the result with the above expression we see that they are identical. **7.8.3.** $\bar{F} = \frac{1}{2R-R}\int_R^{2R} \frac{A}{r^2}dr = \frac{1}{R}\left(-\frac{A}{r}\right)\Big|_R^{2R} = \frac{1}{R}\left(-\frac{A}{2R} + \frac{A}{R}\right) = 0.5\frac{A}{R^2}$. The average value of the force of gravity at this section is half as great as that at the earth's surface, $\bar{F} = 0.5F_0$. **7.8.4.** (a) $F(R) = F_0$, $F(2R) = \frac{1}{4}F_0$; $\frac{F(R)+F(2R)}{2} = 0.625F_0 > 0.5F_0$ (the error is of the order of 10%); (b) $F\left(\frac{R+2R}{2}\right) = F\left(\frac{3}{2}R\right) = \frac{4}{9}F_0$; therefore $\frac{1}{6}F_0 + \frac{2}{3}\frac{4}{9}F_0 + \frac{1}{6\times 4}F_0 = \frac{109}{216}F_0 \simeq 0.505F_0$ (the error is 1%). **7.8.5.** $\frac{x_0^n}{n+1}$. **7.8.6.** $\frac{m-n}{\ln m - \ln n}$. For $m = n+\nu$ with $\nu \ll n$, we have $\ln m = \ln(n+\nu) = \ln n + \ln\left(1+\frac{\nu}{n}\right) = \ln n + \frac{\nu}{n} - \frac{\nu^2}{2n^2} + \ldots$. **7.8.7.** (a) Both mean values are equal to $\frac{1}{2}$; (b) $\frac{1}{2} - \frac{1}{\pi}$ and $\frac{1}{2} + \frac{1}{\pi}$. **7.8.8.** If T is the period of the function y, then we have $\sin[\omega(t+T)+\alpha] = \sin(\omega t+\alpha)$, whence $\omega T = 2\pi$, $T = \frac{2\pi}{\omega}$. The period of the function y^2 is $\frac{T}{2} = \frac{\pi}{\omega}$; consequently, we must find the mean value of the function $y = \sin^2(\omega t+\alpha)$ in the interval from $t=0$ to $t=\frac{\pi}{\omega}$. We get

$$\bar{y} = \int_0^{\pi/\omega} \frac{\sin^2(\omega t+\alpha)\,dt}{\pi/\omega} = \frac{\omega}{\pi}\int_0^{\pi/\omega}\left[\frac{1}{2} - \frac{1}{2}\cos 2(\omega t+\alpha)\right]dt = \frac{1}{2}.$$

7.9.1. (a) $s = \int_0^1 \sqrt{1+4x^2}\,dx$; (b) $s = \int_0^1 \sqrt{1+e^{2x}}\,dx$; (c) $s = \frac{4}{a}\int_0^a \sqrt{a^2 + \frac{b^2x^2}{a^2-x^2}}\,dx$. **7.9.2.** On changing the variable we get $s = \int_{\sqrt{2}}^{\sqrt{1+e^2}} \frac{z^2}{z^2-1}\,dz$. But $\int\left(1+\frac{1}{z^2-1}\right)dz = z + \int\frac{dz}{z^2-1} = z + \int\left(\frac{1/2}{z-1} - \frac{1/2}{z+1}\right)dz = z + \frac{1}{2}\ln\frac{z-1}{z+1}\,(+C)$. Therefore, $s = \left(z + \frac{1}{2}\ln\frac{z-1}{z+1}\right)\Big|_{\sqrt{2}}^{\sqrt{1+e^2}} = \sqrt{1+e^2} - \sqrt{2} + \frac{1}{2}\ln\frac{\sqrt{1+e^2}-1}{\sqrt{1+e^2}+1} - \frac{1}{2}\ln\frac{\sqrt{2}-1}{\sqrt{2}+1}$. **7.9.3.** Partitioning the arc of the catenary curve from $x=0$ to $x=0.9$ and from $x=0.9$ to $x=2$ we find $s_1 \simeq 1.043$, $s_2 = \frac{e^2+e^{-2}}{2} - \frac{e^{0.9}+e^{-0.9}}{2} + \frac{1}{2}\int_{0.9}^2 \frac{2}{e^x - e^{-x}}\,dx$. In the last integral we can set $e^x = t$, then $\int\frac{2dx}{e^x-e^{-x}} = \int\frac{2e^x dx}{(e^x)^2-1} = \int\frac{2dt}{t^2-1} = \int\left(\frac{1}{t-1} - \frac{1}{t+1}\right)dt$. Finally we obtain $s_2 \simeq 2.624$ and $s = s_1 + s_2 \simeq 3.667$. From the exact formula we get $s = 3.627$. The error is approximately 1%. **7.9.4.** $s \simeq 1.146$. **7.9.5.** The length of the circle's arc being considered

$S = \int_0^{R/\sqrt{2}} \frac{R\,dx}{\sqrt{R^2-x^2}} = \int_0^{R/\sqrt{2}} \left[1 + \frac{1}{2}\left(\frac{x}{R}\right)^2 + \frac{3}{8}\left(\frac{x}{R}\right)^4 + \frac{5}{16}\left(\frac{x}{R}\right)^6 + \frac{35}{128}\left(\frac{x}{R}\right)^8 + \ldots\right] dx = R\frac{1}{\sqrt{2}}\left(1 + \frac{1}{12} + \frac{3}{160} + \frac{5}{896} + \frac{35}{18\,432} + \ldots\right)$.

Accordingly, we obtain the following estimates of the number π: (1) $\pi \simeq \frac{4}{\sqrt{2}}\left(1 + \frac{1}{12} + \frac{3}{160}\right) \simeq 3.117$; (2) $\pi \simeq \frac{4}{\sqrt{2}}\left(1 + \frac{1}{12} + \frac{3}{160} + \frac{5}{896}\right) \simeq 3.133$; (3) $\pi \simeq \frac{4}{\sqrt{2}}\left(1 + \frac{1}{12} + \frac{3}{160} + \frac{5}{896} + \frac{35}{18\,432}\right) \simeq 3.138$.

7.10.1. (a) $k = \frac{ab}{\sqrt{(a^2\sin^2 t + b^2\cos^2 t)^3}}$;

(b) $k = \frac{2x^3}{\sqrt{(1+x^4)^3}}$;

(c) $k = \frac{12ax^2}{\sqrt{(1+16a^2x^6)^3}}$.

7.11.2. $V = 2\pi$. **7.11.3.** $x_C = \frac{2}{\pi}R$.

7.12.1. (a) $y_{\max} = 2$ at $x = 0$; $y_{\min} = -2$ at $x = 2$; (b) there are no maxima or minima; the curve intersects the x axis at the point $x = 1 + \sqrt[3]{14} \simeq 3\,4$ and the y axis at point $y = -15$; (c) there are no maxima or minima; the curve intersects the x axis between the points $x = 0$ and $x = -1$ and the y axis at the point $y = 3$. **7.12.2.** (a) There are three roots; (b) three roots; (c) two roots; (d) one root.

Part 2

Chapter 8

8.1.1. $T \simeq 1600$ years. **8.1.2.** 176.5 g. **8.1.3.** 53.3 g. **8.1.4.** We know that $N(t) = N_0 e^{-t/\bar{t}}$, where N_0 is the quantity of substance at the initial time $t = 0$. We are interested in the time t_1 by which there will be (100 − 1) % = 99% untransformed substance: $N(t_1) = \frac{99}{100}N_0$. Therefore, $\frac{99}{100}N_0 = N_0 e^{-t/\bar{t}}$, whence $t_1 = \bar{t}\ln\frac{99}{100}$. For radium $\bar{t} = 2400$ years and therefore $t_1 = 2400 \ln\frac{99}{100} = 24$ (years). Similarly for the three remaining cases we find $t_2 \simeq 250$ years; $t_3 \simeq 5500$ years; $t_4 \simeq 11\,000$ years. **8.1.5.** Suppose that at the initial time $t = 0$ the number of radium atoms in 10^{12} atoms of rock is N_0; at the time $t = 10\,000$ it equals 1. Therefore we have $1 = N_0 e^{-\frac{10\,0000}{2\,400}}$, whence $N_0 \simeq e^4 \simeq 65$. Similarly, we find that 10^6 years ago $N_0 \simeq e^{417} \simeq 10^{181}$. Clearly, the result is preposterious: 10^{12} atoms of rock contain 10^{181} radium atoms! No less absurd is the result of calculations of the amount of radium 5×10^9 years ago. All this proves that the initial assumption is incorrect on that the present amount of radium can be regarded as the result of disintegration of the original supply of radium (when the earth originated) (cf. Section 8.4).

Chapter 9

9.1.1. $A(t) = -h\int_{t_0}^{t} v^2\,dt$; A is negative since $\int_{t_0}^{t} v^2\,dt \geqslant 0$ for $t > t_0$. **9.1.2.** The motion of the body is periodic with the period $T = \frac{2\pi}{\omega}$; we have to find the work for a half period. During the first quarter of the period the velocity is positive, and therefore $F = -h$; during the second quarter of the period the velocity is negative and $F = h$. During these time intervals the force is positive, and the work equals the product of the force by the distance covered by the body; for the first and second quarters of the period we have $A_1 = -hB$ and $A_2 = h(-B) = -hB$; the work during the half period is $A = A_1 + A_2 = -2hB$. **9.1.3.** $A = Bf_0\omega_1\int_0^T \sin\omega_0 t\cos\omega_1 t\,dt$. This integral can be easily evaluated if we recall that $\sin\omega_0 t\cos\omega_1 t = \frac{1}{2}[\sin(\omega_0+\omega_1)t + \sin(\omega_0-\omega_1)t]$. Therefore $A = \frac{Bf_0\omega_1}{2}\int_0^T \times [\sin(\omega_0+\omega_1)t + \sin(\omega_0-\omega_1)t]\,dt = \frac{Bf_0\omega_1}{2}\left[\frac{1}{\omega_0+\omega_1} + \frac{1}{\omega_0-\omega_1} - \frac{\cos(\omega_0+\omega_1)T}{\omega_0+\omega_1} - \frac{\cos(\omega_0-\omega_1)T}{\omega_0-\omega_1}\right]$. In the case where $\omega_1 = \omega_0$ we cannot use the last formula. In this case, however, $\sin\omega_0 t\cos\omega_1 t = \frac{1}{2}\sin 2\omega_0 t$, whence $A = \frac{Bf_0\omega_0}{2}\int_0^T \sin 2\omega_0 t\,dt = \frac{Bf_0}{4}(1 - \cos\omega_0 T)$. **9.1.4.** The work of the air resis-

tance is $A_1(t) = -\frac{aS\rho g^3}{8} t^4$. The work of the force of gravity is $A_2(t) = \frac{mg^2}{2} t^2$. For a ball: $A_1(1) = -0.00965$, $A_1(10) = -96.5$, $A_1(100) = -965 \times 10^3$, $A_2(1) = 0.177$, $A_2(10) = 1.77$, $A_2(100) = 177$. For a bullet: $A_1(1) = -1.18 \times 10^{-3}$, $A_1(10) = -11.8$, $A_1(100) = -118 \times 10^3$; $A_2(1) = 0.435$, $A_2(10) = 43.5$, $A_2(100) = 4350$ (the dimensions of A are in joules—J). **9.1.5.** $A = \frac{aS\rho (v-v_0)^2 b}{2}$. We find the power W using the formula $W = Fv$, $W = \frac{aS\rho (v_0-v)^2 v}{2}$. Let us find the velocity v (for a given v_0) at which the power is the greatest, i.e. when $\frac{dW}{dv} = 0$; we get $v_1 = v_0$ and $v_2 = v_0/3$. Clearly, we are not interested in the value $v = v_0$ at which the power is zero. Of interest here is the value $v = v_0/3$ (the reader can make a complete analysis based on taking into account the sign of d^2W/dv^2). For $v_0 = 30$ m/s and $v = 10$ m/s we have $W_{\max} \simeq 25.75 \times 10^5$ W. **9.1.6.** The work of the force for one period is $A = B\pi f \sin\alpha$; the power is $W = \frac{Bf\omega}{2} \sin\alpha$.

9.3.1. We take as the x axis the straight line on which the charges are located; let the charge e_1 coincide with the origin and the charge e_2 be at the point $x = 2a$. The equilibrium of the charge at x is only possible if $F = -\frac{du}{dx} = 0$. But $F = \frac{e_1 e}{x^2} - \frac{4e_1 e}{(2a-x)^2}$, if $0 < x < 2a$, if the charge e lies between e_1 and e_2; $F = -\frac{e_1 e}{x^2} - \frac{4e_1 e}{(2a-x)^2}$ if $x < 0$, and $F = \frac{e_1 e}{x^2} + \frac{4e_1 e}{(2a-x)^2}$, if $x > 2a$. In the first case the condition $F = 0$ yields $x_1 = 2a/3$, $x_2 = -2a$ (the second root x_2 is neglected since we must have $x > 0$); in the cases where $x < 0$ and $x > 2a$ the equation $F = 0$ has no solution at all. Consequently, there is one position of equilibrium $x_1 = 2a/3$. Then we calculate $\frac{d^2u}{dx^2}$ at the point $x = x_1 = 2a/3$, we find that if $e > 0$, then $\frac{d^2u}{dx^2} > 0$, i.e. the equilibrium is *stable*, but if $e < 0$, the equilibrium is *unstable*. **9.3.2.** There is one position of equilibrium x_1 outside the charges. If the coordinate system is chosen in the same way as that in the solution to Exercise 9.3.1, then $x_1 = -2a$. If $ee_1 > 0$, the equilibrium is *stable*, but if $ee_1 < 0$, the equilibrium is *unstable*.

9.4.1. The equation of motion is $m\frac{dv}{dt} = F$; using the fact that $v = 0$ at $t = 0$, we find $v = \frac{F}{m} t$. Therefore $\frac{dx}{dt} = \frac{F}{m} t$; $dx = \frac{F}{m} t\, dt$, whence $\int_0^x dx = \frac{F}{m} \int_0^t t\, dt$, since $x = 0$ at $t = 0$. Finally $x = \frac{F}{2m} t^2$. **9.4.2.** (a) $x = v_0 t + \frac{F}{2m} t^2$; (b) $x = x_0 + v_0 t + \frac{F}{2m} t^2$. **9.4.3.** 2.5 m. **9.4.4.** The equation of motion is $m\frac{dv}{dt} = mg$, whence we find $x = \frac{gt^2}{2}$, $t = \sqrt{\frac{2x}{g}} \simeq 4.5$ s for $x = 100$ m. **9.4.5.** (a) $t = 3.6$ s; (b) $t = 5.6$ s. Further, for the case (a) $v = gt + v_0$. Let v_1 be the velocity at the moment of impact; then $v_1 = gt_1 + v_0$, with t_1 the landing time. From the equation $v = gt + v_0$ we find $x = v_0 t + \frac{gt^2}{2}$. Let the ball fall from the height H; then $x = H$ at $t = t_1$; therefore $2H = 2v_0 t_1 + gt_1^2$, whence $t_1 = \frac{-v_0 + \sqrt{v_0^2 + 2gH}}{g}$ and $v_1 = gt_1 + v_0 = \sqrt{v_0^2 + 2gH}$. In the case (b) $v = gt - v_0$ and the remaining is the same; the final velocity obtained is the same as in (a). **9.4.6.** $x = \frac{k}{6m} t^3 + v_0 t$. **9.4.7.** (a) $x = -\frac{f}{m\omega^2} \cos\omega t$, $T = \frac{2\pi}{\omega}$, $x_{\max} = \frac{f}{m\omega^2}$, $v_{\max} = \frac{f}{m\omega}$; (b) $x = \frac{f}{m\omega} t - \frac{f}{m\omega^2} \sin\omega t$. **9.4.8.** Let the desired velocity be v_0. Then the law of body's motion is $x = x_0 + v_0 (t - t_0) + \frac{F}{2m} (t - t_0)^2$. Since $x = x_1$ at $t = t_1$, we have $x_1 = x_0 + v_0 (t_1 - t_0) + \frac{F}{2m} (t_1 - t_0)^2$, whence $v_0 = \frac{x_1 - x_0}{t_1 - t_0} - \frac{F}{2m} (t_1 - t_0)$.

9.8.1. $K = \frac{F^2}{2m} t^2 = F(x - x_0)$. **9.8.2.** $K = \frac{f^2}{2m\omega^2} \sin^2\omega t$; $K_{\max} = \frac{f^2}{2m\omega^2}$. **9.8.3.** $\overline{K} = \frac{mA^2\omega^2}{2} \overline{\sin^2(\omega t + \alpha)} = \frac{mA^2\omega^2}{4}$. **9.8.5.** (a) $A \simeq 4 \times 10^7$ J, $W \simeq 2.2 \times 10^5$ W; (b) $A \simeq 7 \times 10^7$ J, $W \simeq 38 \times 10^4$ W. **9.8.6.** We find the work of each separate force. First we determine the velocity of the body. From the equation $m\frac{dv}{dt} = at + a(\theta - t) = a\theta$ we find $v = v_0 + \frac{a\theta}{m} t$. The work of force F_1 is $A_1 = \int_0^\theta at \left(v_0 + \frac{a\theta}{m} t\right) dt = \frac{av_0\theta^2}{2} + \frac{a^2\theta^4}{3m}$. Simi-

larly, the work of force F_2 is $A_2=\frac{av_0\theta^2}{2}+\frac{a^2\theta^4}{6m}$. We now obtain the product of the impulse by the average velocity $I_1=\int_0^\theta at\,dt=\frac{a\theta^2}{2}$, $I_2=\int_0^\theta a(\theta-t)\,dt=\frac{a\theta^2}{2}$, $\bar{v}=\frac{\int_0^\theta\left(v_0+\frac{a\theta}{m}t\right)dt}{\theta}=v_0+\frac{a}{2m}\theta^2$, and therefore $I_1\bar{v}=\frac{a\theta^2}{2}\left(v_0+\frac{a}{2m}\theta^2\right)$, $I_2\bar{v}=\frac{a\theta^2}{2}\left(v_0+\frac{a}{2m}\theta^2\right)$. We can see that $(I_1+I_2)\bar{v}=A_1+A_2$ though $I_1\bar{v}\neq A_1$, $I_2\bar{v}\neq A_2$ (cf. p. 326). 9.8.7. At the beginning of the experiment the mass m had a velocity v_0 (it was moving together with the train) and a kinetic energy $K_1=\frac{mv_0^2}{2}$. After the action of the man the velocity of the mass became v_0+v_1, $K_2=\frac{m(v_0+v_1)^2}{2}$, with $v_1=\frac{Ft}{m}$. The change in the kinetic energy $\Delta K=K_2-K_1=m\frac{(v_0+v_1)^2}{2}-\frac{mv_0^2}{2}$. This is the work performed on the mass by the train and the man together. In order to find the work performed on the mass by the man, note that the velocity of the mass with respect to the man moving in the same train was zero before the experiment and became v_1 after the experiment. Therefore the work performed by the man is $A_1=\frac{mv_1^2}{2}-0=\frac{mv_1^2}{2}=\frac{F^2t^2}{2m}$. It is easy now to find the work A_2 of the locomotive $A_2=\Delta K-A_1=mv_0v_1=v_0Ft$. (The last result can also be obtained in a different way. Indeed, the locomotive's work is $A_2=\int_0^t vF\,dt$, since the velocity $v=v_0$ is constant, we have $A_2=v_0\int_0^t F\,dt=v_0Ft$.) 9.8.8. Before the experiment the velocity of mass m and that of the man are zero. After the experiment the mass m acquired the velocity v_1 and the man v_2. We find these velocities from the equations $m\frac{dv_1}{dt}=F$ and $M\frac{dv_2}{dt}=-F$, since if the man acts on the mass m with the force F, the mass m acts on the man with the force $-F$ (Newton's third law of motion). We find $v_1=\frac{Ft}{m}$, $v_2=-\frac{Ft}{M}$. The work performed by the force F on the mass m and the man is $A=K_1+K_2$, where $K_1=\frac{mv_1^2}{2}=\frac{F^2t^2}{2m}$ is the change in the kinetic energy of the mass, $K_2=\frac{Mv_2^2}{2}=\frac{F^2t^2}{2M}$ is the change in the kinetic energy of the man. Whence $A=\frac{F^2t^2(M+m)}{2Mm}$. 9.8.9. The change in the kinetic energy of mass m is $\Delta K_m=mv_0v_1+\frac{mv_1^2}{2}$, where $v_1=\frac{F}{m}t$. The change in the kinetic energy of the man is $\Delta K_M=\frac{Mv_2^2}{2}+Mv_0v_2$, where $v_2=-\frac{F}{M}t$, $A=\frac{F^2t^2(M+m)}{2Mm}$. 9.8.10. (a) If $t_1'=t_1+b$ and $t_2'=t_2+b$, then $\tau'=t_2'-t_1'=t_2-t_1=\tau$. (b) If $x_1'=x_1+vt_1+a$ and $x_2'=x_2+vt_2+a$, with $t_1=t_2$, then $d'=x_2'-x_1'=x_2-x_1=d$.

9.9.1. For $\varphi=30°$ we have $x_2=565$ m, $y_{max}=81.5$ m; for $\varphi=45°$ we have $x_2=650$ m, $y_{max}=163$ m; for $\varphi=60°$, we have $x_2=565$ m, $y_{max}=244$ m. 9.9.2. The equation of the trajectory is $y=x\tan\varphi-x^2\frac{g}{2v_0^2\cos^2\varphi}$. At a given $x=500$ m we seek φ for which $y=y_{max}$, that is, we solve the equation $\frac{dy}{d\varphi}=0$. It yields $\tan\varphi=\frac{v_0^2}{gx}$, i.e. $\frac{1}{\cos^2\varphi}=\frac{v_0^4+g^2x^2}{g^2x^2}$. Using this fact and the equation of the trajectory we find $y_{max}=\frac{v_0^2}{2g}-\frac{x^2g}{2v_0^2}$. Setting $v_0=80$ m/s, $x=500$ m we determine $y_{max}\simeq 135$ m.

9.10.1. $r=30\,000$ km.

9.11.1. Find the second derivative $\frac{d^2y}{dz^2}$ of the function $y=F(z)$ (compare it with what is said in pp. 337 and 216, note that $F(z)\to 0$ as $z\to 0$ and $z\to\infty$).

9.12.1. Setting the origin of coordinates at the center of gravity of the rod we get $I_0=\int_{-l/2}^{l/2}x^2\sigma\,\partial x=\sigma\int_{-l/2}^{l/2}x^2\,dx=\sigma\frac{l^3}{12}$; since $m=\sigma l$, the last result can be written as $I_0=m\frac{l^2}{12}$. 9.12.2. Set the origin at the joint of the pieces having different densities so that for the first piece it is σ_1 at $x<0$ and for the second piece it equals σ_2 at $x>0$. Then $x_c=\frac{1}{2}\frac{\sigma_2l_2^2-\sigma_1l_1^2}{l_1\sigma_1+l_2\sigma_2}$. 9.12.3. On choosing the coordinate system as indicated

in the hint we obtain $x_c = \frac{2}{3}L$; the moment of inertia with respect to the origin is $I = \int_0^L x^2\sigma(x)\,dx = \frac{aL^4}{4}$. Since $m = \frac{aL^2}{2}$, $a = \frac{2m}{L^2}$; therefore $I = \frac{mL^2}{2}$. In view of (9.12.13) $I_0 = I - mL_1^2$, with $L_1 = \frac{2}{3}L$; whence $I_0 = \frac{mL^2}{18}$.

9.13.1. (a) Compare with the hint to Exercise 9.12.3. *Answer.* The center of gravity belongs to the median of the triangle and is 2/3 of the distance from the vertex along the median (it coincides with the median point). (b) Let a and b be the lengths of the bases of the trapezoid, with $a > b$, the center of gravity lies on the straight line connecting the middle points of the bases at the distance $\frac{1}{3}h\frac{a+2b}{a+b}$ from the lower base, with h the altitude of the trapezoid. (c) The center of gravity lies on the symmetry axis of the half-disk at the distance $\frac{4}{3\pi}R$ from the diameter which bounds the half-disk, with R the radius of the half-disk.

9.14.1. We write the equation thus $\frac{dt}{dv} = \frac{1}{\beta(v_1^2 - v^2)}$, whence accounting for the initial condition $v(0) = v_0$ we find $t = \frac{1}{\beta}\int_{v_0}^{v}\frac{dv}{v_1^2 - v^2}$. (In order to evaluate the integral, it is sufficient to write the integrand in the form $\frac{1}{v_1^2 - v^2} = \frac{a}{v_1 - v} + \frac{b}{v_1 + v}$ and find a and b by the method of undetermined coefficients; see exercises to Section 5.3.) Thus we have $\ln\frac{v_1+v}{v_1-v} = \ln A + 2\beta v_1 t$, where $A = \frac{v_1+v_0}{v_1-v_0}$, or $\frac{v_1+v}{v_1-v} = Ae^{2\beta v_1 t}$, i.e. $v = v_1\frac{Ae^{2\beta v_1 t}-1}{Ae^{2\beta v_1 t}+1}$. For very large t we have $e^{2\beta v_1 t} \gg 1$ and therefore $v \simeq v_1$. We write the solution to the equation in the form $v = v_1\frac{A - e^{-2\beta v_1 t}}{A + e^{-2\beta v_1 t}}$, or $v - v_1 = -2v_1\frac{e^{-2\beta v_1 t}}{A + e^{-2\beta v_1 t}}$. For rather large t we can neglect $e^{-2\beta v_1 t}$ in the denominator since it is small compared to A; therefore in this case we have $v - v_1 \simeq -2v_1\frac{e^{-2\beta v_1 t}}{A}$, or $v - v_1 \simeq \frac{2v_1(v_0 - v_1)}{v_1 + v_0}e^{-2\beta v_1 t}$; comparing this with (9.14.24) we find $C = \frac{2v_1(v_0 - v_1)}{v_1 + v_0}$.

9.14.2. In the case where the resistance is proportional to the velocity the equation of motion has the form $\frac{dv}{dt} = g - \frac{k}{m}v - \frac{A}{m}$, where A is the buoyancy. In this case the velocity $v_1 = \frac{g}{\alpha} - \frac{A}{\alpha m}$ sets in. By the Archimedean law $A = V\rho' g$, where V is the volume of the body and $\rho' = \rho_l$ is the density of the liquid. Since $m = V\rho$ ($\rho = \rho_b$ is the density of the body), we have $A = \frac{mg\rho'}{\rho}$ and $v_1 = \frac{g}{\alpha}\left(1 - \frac{\rho'}{\rho}\right)$. For $\rho' > \rho$ the body floats up ($v_1 < 0$), while for $\rho' < \rho$ it sinks ($v_1 > 0$).

Chapter 10

10.1.2. (a) The body stops at $x_0 = 1/4$. (b) The stopping point is $x_0 \simeq 0.9$. (c) There are no points of stop. **10.1.3.** The position of maximum of $u(x)$ is $u_{\max} \simeq 9.5$ at $x = 8/3$. Noting that the left-hand branch of the graph rises and the right-hand branch descends we can construct a rough graph of $u(x)$, which is, however, sufficient for solving the problem. (a) From the fact that the sum of the kinetic and potential energies is constant it follows that two stopping points (at which the kinetic energy is zero) are the values of the least roots of the equation $-x^3 + 4x^2 = 6$. This equation can be solved either graphically or by a numerical method. We get $x_1 \simeq -1.09$, $x_2 \simeq 1.57$. The body oscillates between the points x_1 and x_2. (b) One stopping point is $x \simeq -2.04$. However, this point is to the left of the point from which the body starts at the initial moment. Since the initial velocity is directed to the right, the body will move to the right without reaching the point of stop. (c) One stopping point is $x = -2.04$. In this case the body will move to the left of the stopping point and then travel to the right. **10.1.4.** There are no points of stop. The body will travel to the right. (b) There are two stopping points: $x_1 = +\sqrt{9/11}$, $x_2 = -\sqrt{9/11}$. The motion of the body is the oscillation between the points x_1 and x_2. In the case (a) $t = t_0 + \int_0^x \sqrt{\frac{1+x^2}{4+3x^2}}\,dx$. In the case (b) $t = t_0 + \int_{0.5}^x \sqrt{\frac{20+20x^2}{9-11x^2}}\,dx$ at $x < x_1$ and $t = t_1 - \int_{x_1}^x \sqrt{\frac{20+20x^2}{9-11x^2}}\,dx$ at $x_1 < x <$

$< x_2$, where t_1 is the time when body reached x_1. (Note that the integrals remain finite though at the stop point the integrand is infinitely large.)

10.2.1. (a) $x = 2\sin t$, (b) $x = \cos t$, (c) $x = \cos t + 2\sin t$. This solution can be written in the form $x = C\cos(t+\alpha)$, where $C = \sqrt{5}$, $\alpha = \arctan(-2) \simeq -1.11$, i.e. $x = \sqrt{5}\cos(t-1.11)$. In all these cases $T = 2\pi$.

10.3.1. (a) Let L be the length of the pendulum. We know that $\omega = \sqrt{\frac{mgl}{I}}$, and in our case $l = \frac{2}{3}L$, $I = \frac{mL^2}{2}$ (see Exercise 9.12.3). Therefore $\omega = \sqrt{\frac{4g}{3L}} = \sqrt{\frac{8g}{9l}}$, $T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{3L}{4g}} \simeq 5.43\sqrt{\frac{L}{g}}$. The value of l corresponding to the maximum frequency is determined from the formula $l_{\max} = \sqrt{\frac{I_0}{m}}$. Since in our case $I_0 = \frac{mL^2}{18}$, we have $l_{\max} = \frac{L}{\sqrt{18}}$, therefore $\omega_{\max} = \sqrt{\frac{g\sqrt{18}}{2L}}$. Knowing $\omega_{\max}$ we find $T_{\min} = \frac{2\pi}{\omega_{\max}}$. (b) Here $l = \frac{L}{3}$, $I = I_0 + ml^2 = \frac{mL^2}{18} + \frac{mL^2}{9} = \frac{mL^2}{6}$, $\omega = \sqrt{\frac{2g}{L}}$. The minimum period coincides here with that obtained in the case (a), but the point of suspension is different (however, $l_{\max}$ is as before).

10.4.1. Let the oscillations be given by the formula $x = C\cos(\omega t + \alpha)$, then $v = -C\omega\sin(\omega t + \alpha)$. We use the relation $kC\frac{dC}{dt} = F_1 v$, where $F_1 = -hv\,|\,v\,|$. $F_1 v = -hv^2\,|\,v\,| = -hC^3\omega^3\,|\sin^3(\omega t+\alpha)|$, and therefore the law of the change of the amplitude C can be written as $kC\frac{dC}{dt} = -hC^3\omega^3 A$, where $A = \overline{|\sin^3(\omega t+\alpha)|}$. Note that $\sin(\omega t+\alpha)$ retains its sign when t varies from $t_1 = -\frac{\alpha}{\omega}$ to $t_2 = \frac{\pi-\alpha}{\omega}$, with $t_2 - t_1 = \frac{T}{2}$. Therefore $A = \frac{\int_{t_1}^{t_2}\sin^3(\omega t+\alpha)\,dt}{t_2-t_1} = \frac{\omega}{\pi}\int_{-\frac{\alpha}{\omega}}^{\frac{\pi-\alpha}{\omega}}\sin^3(\omega t+\alpha)\,dt$. Changing in the last integral the variable $\cos(\omega t+\alpha) = x$ we obtain $A = -\frac{1}{\pi}\int_1^{-1}(1-x^2)\,dx = \frac{4}{3\pi}$, whence $\frac{dC}{dt} = -\frac{hC^2\omega^3\times 4}{3\pi k}$. Denote $\frac{4h\omega^3}{3\pi k} = b$; then $\frac{dC}{dt} = -bC^2$ and consequently $\frac{dt}{dC} = -\frac{1}{bC^2}$. The solution to this equation is $t = -\frac{1}{b}\frac{C-C_0}{CC_0}$, whence $C = \frac{C_0}{1+C_0 bt}$; here C_0 is the value of the amplitude at the initial time $t = 0$ which can be determined from the initial conditions. (Note that the same law we have obtained for the velocity of the rectilinear motion in the case of resistance proportional to the squared velocity (see (9.14.15)).) **10.4.2.** In this case the work for a quarter of the period is $-fC$, therefore the average power is $-fC\frac{4}{T} = -fC\frac{2\omega}{\pi}$. We get the equation $kC\frac{dC}{dt} = -\frac{2fC\omega}{\pi}$, whence $\frac{dC}{dt} = -\frac{2f\omega}{k\pi}$ and therefore $C = C_0 - \frac{2f\omega}{k\pi}t$. The oscillations will cease at time t_1 when $C = 0$, that is, when $t_1 = \frac{C_0 k\pi}{2f\omega}$ (we suppose that $t_1 \gg T$).

10.5.1. From the first equation of the system we get $C_0\cos\varphi = x_0 - a$. Using this expression we readily obtain from the second equation $C_0\sin\varphi = b\frac{\omega}{\omega_1} - \frac{v_0}{\omega_1} - \gamma\frac{x_0-a}{\omega_1}$. Squaring these two equations and adding the result yields $C_0^2 = \left(b\frac{\omega}{\omega_1} - \frac{v_0}{\omega_1} - \gamma\frac{x_0-a}{\omega_1}\right)^2 + (x_0-a)^2$; taking square roots of both sides of the equation we find C_0 and then find $\tan\varphi = \frac{b\frac{\omega}{\omega_1} - \frac{v_0}{\omega_1} - \gamma\frac{x_0-a}{\omega_1}}{x_0-a}$.

10.7.2. Beats (cf. equation (10.7.5)).

10.8.1. Use the fact that $\sin^3 x = \sin x\sin^2 x = \frac{\sin x(1-\cos 2x)}{2} = \frac{\sin x}{2} - \frac{\sin x\cos 2x}{2} = \frac{\sin x}{2} - \frac{\sin 3x - \sin x}{4} = \frac{3}{4}\sin x - \frac{1}{4}\sin 3x$ (representation of $f(x)$ in the form (10.8.15)). *Answer.* $y(x,\ t) = \frac{3}{4}\cos t\sin x - \frac{1}{4}\cos 3t\sin 3x$. **10.8.2.** We are forced to abandon the condition $D = 0$ so that instead of (10.8.14) we get $y(x,\ t) = \sum_k b_k\sin\frac{k\pi x}{l}\cos\frac{kc\pi t}{l} + d_k\sin\frac{k\pi x}{l}\times\sin\frac{kc\pi t}{l}$. **10.8.3.** Conditions (10.8.2b) yield

$\varphi(x)+\psi(x)=f(x)$; $\varphi'(x)-\psi'(x)=0$, whence $y(x, t)=\frac{1}{2}[f(x+ct)+f(x-ct)]$, where we must still redefine $f(x)$ outside the interval $(0, l)$ assuming that the function is odd and periodic with period $2l$.

10.9.3. (a) $y=\frac{\pi^2}{3}-4\left(\cos x-\frac{\cos 2x}{2^2}+\frac{\cos 3x}{3^2}+\ldots\right)$; (b) $y=\left(\frac{C_1-C_2}{4}\right)\pi-\frac{2(C_1-C_2)}{\pi}\sum_{k=1}^{\infty}\frac{\cos[(2k-1)\pi x/l]}{2k-1}+(C_1+C_2)\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin(n\pi x/l)}{n}$.

Chapter 11

11.3.1. $p=1.13p_0$; $p=1.48p_0$; $p=3.67p_0$, where p_0 is the air pressure at the earth's surface. **11.3.2.** The dependence of the pressure on altitude is of the form $p=p_0e^{-gh/b}$, where $b=10^3\frac{RT}{M}$. For the temperature -40 °C we have $T=273-40=233$ K, $b=\frac{8.3\times10^3\times233}{29.4}\simeq6.6\times10^4$ cm²/s²; in this case $H=\frac{b}{g}=6.6$ km. For the temperature 40 °C we have $T=313$ K, $b=8.8\times10^4$ cm²/s², $H=8.8$ km. **11.3.3.** From the equation $\frac{dT}{dh}=-\alpha T_0$ it follows that $T=-\alpha T_0h+C$, where the constant C is determined from the condition $T=T_0$ at $h=0$, whence $C=T_0$; therefore $T=T_0(1-\alpha h)$. The basic equation for determining the density is $\frac{dp}{dh}=-g\rho$. We use Clapeyron's equation $p=10^3\rho\frac{RT}{M}$ and substitute the expression for T to find $p=10^3\rho\frac{RT_0(1-\alpha h)}{M}$. Set $10^3\frac{RT_0}{M}=b_0$, then $p=\rho b_0(1-\alpha h)$, whence $\rho=\frac{p}{b_0(1-\alpha h)}$. The differential equation for p assumes the form $\frac{dp}{dh}=-g\frac{p}{b_0(1-\alpha h)}$, or $\frac{dp}{p}=\frac{-g\,dh}{b_0(1-\alpha h)}$; integrating we get $\ln p=\frac{g}{b_0\alpha}\ln(1-\alpha h)+C$, whence $p=A(1-\alpha h)^{g/b_0\alpha}$, with $A=e^C$. Since $p=p_0$ at $h=0$, $A=p_0$; therefore $p=p_0(1-\alpha h)^{g/b_0\alpha}$. **11.3.4.** $p=p_0(1-0.037\times10^{-5}h)^{3.46}$; $p=1.13p_0$; $p=1.44p_0$; $p=2.97p_0$.

Chapter 13

13.2.1. The current decreases according to the law $j=j_0e^{-t/RC}$. At the time t_1, which interests us, we have $j=\frac{9}{10}j_0$; therefore $\frac{9}{10}j_0=j_0e^{-t_1/RC}$, whence $\frac{9}{10}=e^{-t_1/RC}$. Taking logarithms yields $\ln\frac{9}{10}=-\frac{t_1}{RC}$, whence $t_1=-RC\ln\frac{9}{10}\simeq0.105RC$. In veiw of the last formula we have $t_1\simeq1$ s for $R=10^7\ \Omega$; $t_1\simeq10.5$ s for $R=10^8\ \Omega$; $t_1\simeq105$ s for $R=10^9\ \Omega$. The time t_2, when the current decreases by a factor of 2, is determined in a similar way: $0.5j_0=j_0e^{-t_2/RC}$, whence $t_2\simeq0.693RC$. Here $t_2\simeq6.93$ s for $R=10^7\ \Omega$; $t_2\simeq69.3$ s for $R=10^8\ \Omega$; $t_2\simeq693$ s for $R=10^9\ \Omega$. **13.2.2.** We use formula (13.1.11) and find $\varphi_{C_1}+\varphi_R+\varphi_{C_2}=0$, whence $\varphi_R=-(\varphi_{C_1}+\varphi_{C_2})$. The current in the circuit is the same everywhere; therefore $j=C_1\frac{d\varphi_{C_1}}{dt}=C_2\frac{d\varphi_{C_2}}{dt}=\frac{\varphi_R}{R}=-\frac{\varphi_{C_1}+\varphi_{C_2}}{R}$. We obtain the equations $\frac{d\varphi_{C_1}}{dt}=-\frac{1}{RC_1}(\varphi_{C_1}+\varphi_{C_2})$, $\frac{d\varphi_{C_2}}{dt}=-\frac{1}{RC_2}(\varphi_{C_1}+\varphi_{C_2})$; adding them yields $\frac{d(\varphi_{C_1}+\varphi_{C_2})}{dt}=-\frac{1}{RC}\frac{\varphi_{C_1}+\varphi_{C_2}}{dt}$, where we set $C=\frac{C_1C_2}{C_1+C_2}$ (C is the total capacitance of two capacitors connected in series, C_1 and C_2). Since $\varphi_{C_1}+\varphi_{C_2}=a$ at $t=0$, the last equation yields $\varphi_{C_1}+\varphi_{C_2}=ae^{-t/RC}$. Clearly, $C_1\frac{d\varphi_{C_1}}{dt}-C_2\frac{d\varphi_{C_2}}{dt}=0$, or $\frac{d}{dt}(C_1\varphi_{C_1}-C_2\varphi_{C_2})=0$. Therefore $C_1\varphi_{C_1}-C_2\varphi_{C_2}=A$, where A is a constant. Using the initial conditions $\varphi_{C_1}=a$, $\varphi_{C_2}=0$ at $t=0$, we find $A=C_1a$. Thus, $\varphi_{C_1}+\varphi_{C_2}=ae^{-t/RC}$, $C_1\varphi_{C_1}-C_2\varphi_{C_2}=C_1a$. Whence $\varphi_{C_1}=a\frac{C_1}{C_1+C_2}\times\left(1+\frac{C_2}{C_1}e^{-t/RC}\right)$, $\varphi_{C_2}=a\frac{C_1}{C_1+C_2}(-1+e^{-t/RC})$. **13.2.3.** We label all the quantities belonging to the circuit before all linear dimensions are increased with the index 1 and after the increase with the index 2. Then $T_1=R_1C_1$, $T_2=R_2C_2$; $C_2=\frac{\varepsilon S_2}{4\pi d_2}=\frac{\varepsilon n^2S_1}{4\pi d_1n}=nC_1$; $R_2=\rho\frac{l_2}{\sigma_2}=\rho\frac{nl_1}{n^2\sigma_1}=\frac{R_1}{n}$. Therefore

$T_2=\frac{R_1}{n}nC_1=R_1C_1=T_1$. The time constant has not varied.

13.8.1. Let the potential difference be φ. For the circuit shown in Figure 13.8.3 we have $\varphi_E+\varphi_L+\varphi_C=0$, or $L\frac{dj}{dt}+\varphi=E_0$ (since $\varphi_E=-E_0$). As $j=C\frac{d\varphi}{dt}$, $LC\frac{d^2\varphi}{dt^2}+\varphi=E_0$, i.e. $LC\frac{d^2\varphi}{dt^2}=-(\varphi-E_0)$ or $LC\times\frac{d^2z}{dt^2}=-z$ with $z=\varphi-E_0$. The solution to the last equation has the form $z=B\cos(\omega t+\alpha)$, with $\omega=1/\sqrt{LC}$; therefore $\varphi=B\cos(\omega t+\alpha)+E_0$. But $\varphi=0$, $j=0$ at $t=0$; using this fact we get $B=-E_0$, $\alpha=0$. Finally we have $\varphi=E_0(1-\cos\omega t)$. The value $\varphi_{max}=2E_0$ corresponds to the equation $\cos\omega t=-1$, i.e. $t=\pi/\omega=T/2$; it is attained at half period of oscillations. **13.8.2.** The energy of capacitance $W=C\varphi^2/2=4CE_0^2/2=2CE_0^2$; the energy released by the voltage source $P=qE_0=C\varphi E_0=2CE_0^2$.

13.9.1. $j(t)=-\frac{\varphi_0}{L\omega}e^{-\lambda t}\sin\omega t$, with $\lambda=\frac{R}{2L}$, $\omega^2=\frac{1}{LC}-\lambda^2$. For the given three cases we have $j(t)=-1.0025e^{-0.05t}\sin t$; $j(t)=-1.031e^{-0.25t}\sin 0.97t$; $j(t)=-1.15e^{-0.5t}\sin 0.87t$. **13.9.2.** $j(t)=j_0\left(\cos\omega t-\frac{\lambda}{\omega}\sin\omega t\right)e^{-\lambda t}$; formulas for λ and ω see in Exercise 13.9.1. For the given cases we have $j(t)=e^{-0.05t}(-\cos t+0.05\sin t)$; $j(t)=e^{-0.25t}(-\cos 0.97t+0.26\sin 0.97t)$; $j(t)=e^{-0.5t}(-\cos 0.86t+0.58\sin 0.86t)$. **13.9.3.** If R is very great, then the current flowing through the resistance is small, i.e. the current flows mainly through the inductance. Therefore the greater the R the more close this circuit is to that shown in Figure 13.8.1, where $\varphi=\varphi_0\cos(\omega t+\alpha)$. If R is great, we can assume that φ in the circuit of Figure 13.9.2 has the same form, while φ_0 varies slowly with time. Here $\frac{dP}{dt}=-\bar{h}$; but $h=Rj_1^2$, where j_1 is the current flowing through the resistance R, $j_1=\frac{\varphi}{R}$; therefore $h=\frac{\varphi^2}{R}=\frac{\varphi_0^2\cos^2(\omega t+\alpha)}{R}$ and $\bar{h}=\frac{\varphi_0^2}{2R}$. Thus, $\frac{dP}{dt}=-\frac{\varphi_0^2}{2R}$. Noting that $P=\frac{C\varphi_0^2}{2}$, we find $\frac{dP}{dt}=C\varphi_0\frac{d\varphi_0}{dt}=-\frac{\varphi_0^2}{2R}$. Whence $\frac{d\varphi_0}{dt}=-\frac{1}{2RC}\varphi_0$. Therefore $\varphi_0=Ae^{-\frac{t}{2RC}}$; $\lambda=\frac{1}{2RC}$.

13.10.1. $\varphi(t)=e^{-t}+te^{-t}$; $\varphi(t)=-0.03e^{-5.83t}+1.03e^{-0.17t}$; $\varphi(t)=-0.01e^{-9.9t}+1.01e^{-0.1t}$. **13.10.2.** $\varphi(t)=e^{-t}+2te^{-t}$; $\varphi(t)=-0.37e^{-3.73t}+1.37e^{-0.27t}$.

Part 3

Chapter 14

14.1.1. (a) $u=x^3-3xy^2-3x+1$, $v=3x^2y-y^3-3y$; (b) $u=\frac{x^3+x+xy^2}{(1+x^2-y^2)^2+4x^2y^2}$, $v=\frac{y-x^2y-y^3}{(1+x^2-y^2)^2+4x^2y^2}$; (c) $u=x^n-\binom{n}{2}x^{n-2}y^2+\binom{n}{4}x^{n-4}y^4+\ldots$, $v=\binom{n}{1}\times x^{n-1}y-\binom{n}{3}x^{n-3}y^3+\binom{n}{5}x^{n-5}y^5-\ldots$, where $\binom{n}{k}=\frac{n!}{k!\,(n-k)!}$ is the number of combinations of n elements taken k at a time. **14.1.2.** Formulas (14.1.6a) and (14.1.6b) can be proved proceeding from the equations $z^{-1}z=1$ and $(z^{1/3})^3=z$; the general formula (14.1.6) follows from the fact that it is valid for any *rational* $n=p/q=(1/q)\,p$, where the cases of p being positive or negative must be considered separately. **14.1.3.** Use the fact that if $c=r(\cos\alpha+i\sin\alpha)$ then $\sqrt[n]{c}=\rho(\cos\beta+i\sin\beta)$, with $\rho=\sqrt[n]{r}$, and $\beta=\beta_0+\frac{2k\pi}{n}$, with $\beta_0=\frac{\alpha}{n}$, $k=0, 1, 2, \ldots, n-1$. **14.1.4.** In accordance with the results of Exercise 14.1.3 we have $\sqrt[4]{-1}=\cos\beta+i\sin\beta$ with $\beta=\pi/4$, $3\pi/4$, $5\pi/4\,(=-3\pi/4)$, $7\pi/4\,(=-\pi/4)$; note also that in all cases $\cos\beta$ and $\sin\beta$ equal $\pm\sqrt{2}/2$.

14.3.1. If $\ln u=w$, then $u=e^w$ and $u^z=e^{wz}$; use this fact. **14.3.2.** Clearly $e^e\simeq(2.7)^{2.7}>1>i^i$; while numbers $e^i\,(=\cos 1+i\sin 1$, where the angles are measured in radians) and $i^e\,(=e^{e\ln i}=e^{(e\pi/2)i})$ cannot be compared (we can compare in magnitude only real numbers and not arbitrary complex numbers).

14.4.3. (a) The circle; (b) the hyperbola. We have for (a) and (b) $t=\tan\frac{\varphi}{2}$ and $t=\tanh\frac{\psi}{2}$ respectively if at the same time the curves are also given by the "standard" parametric equations; (a) $x=\cos\varphi$, $y=\sin\varphi$; (b) $x=\cosh\psi$, $y=\sinh\psi$.

Chapter 15

15.1.1. All derivatives have the form $P\left(\frac{1}{x}\right)e^{-1/x^2}$, where $P(z)$ is the polynomial (in variable z); at $z=0$ they are equal to zero

due to a high rate of decrease of the function $y=e^{-1/x^2}$ as $x\to 0$ (cf. Section 6.5). For example, $f'(x)=2/x^3e^{-1/x^2}$ but e^t, with $t=-1/x^2$, decreases more rapidly as $t\to-\infty$ than the function $2\,|\,t\,|^{3/2}(=|\,2/x^3\,|)$ increases as $t\to-\infty$.

15.2.1. If $w=\ln z$, then $u=\ln\sqrt{x^2+y^2}$ and $v=\arctan\frac{y}{x}$, whence $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=\frac{x}{x^2+y^2}$, $\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}=\frac{y}{x^2+y^2}$. **15.2.2.** Use the fact that $u_1=\frac{\partial u}{\partial x}$, $v_1=\frac{\partial v}{\partial x}$.

Chapter 16

16.2.1. Near the point $x=x_0$ we expand the function $\varphi(x)$ in a series retaining only two first terms $\varphi(x)\simeq\varphi(x_0)+\varphi'(x_0)(x-x_0)=\varphi'(x_0)(x-x_0)$. Thus denoting $\varphi'(x_0)=c$, $x-x_0=y$, we get $\delta(\varphi(x))=\delta(cy)=\frac{1}{|\,c\,|}\delta(y)=\frac{1}{|\,\varphi'(x_0)\,|}\delta(x-x_0)$. In a more general case the function $\delta(\psi(x))$ is the sum of similar expressions valid for the zero's of ψ. **16.2.2.** The function $\varphi(x)=\sin x$ vanishes at $x_0=k\pi$, $k=0,\ \pm1,\ \pm2,\ \ldots$, $|\,\varphi'(x_0)\,|=|\,\cos x_0\,|=|\,\cos k\pi\,|=1$; therefore $\delta(\sin x)=\sum\limits_{k=-\infty}^{+\infty}\delta(x-k\pi)$ and $\int\limits_{-\infty}^{+\infty}\psi(x)\times\delta(\sin x)\,dx=\sum\limits_{k=-\infty}^{+\infty}\psi(k\pi)$.

16.3.1. (a) $y'(x)=1-\delta(x-1)$. (b) Since $y(+0)=1$, $y(-0)=0$, we have $\Delta y=1$ at $x=0$. For $x\neq0$ we have $y'(x)=-\frac{e^{1/x}}{x^2(1+e^{1/x})^2}$; and finally $y'(x)=\delta(x)-\frac{e^{1/x}}{x^2(1+e^{1/x})^2}$.

16.4.1. Multiply the right-hand side of (16.4.8) by $\sin\frac{x}{2}$ and use the fact that $\cos kx\sin\frac{x}{2}=\frac{1}{2}\left[\sin\left(k+\frac{1}{2}\right)x-\sin\left(k-\frac{1}{2}\right)x\right]$.

Chapter 17

17.3.1. If Γ is the streamline, then by the definition of the function ψ the tangent vector $\mathbf{t}=\left(-\frac{\partial\psi}{\partial y},\ \frac{\partial\psi}{\partial x}\right)$ to Γ (i.e. to the line $\psi=$ constant) coincides with the vector $\mathbf{v}$ of the flow rate which means that along the streamline of the liquid whose direction at each point is $\mathbf{v}=\mathbf{v}(x,\ y)$ $d\psi$ is equal to zero. (b) Let Γ be a small arc with the endpoints $P=P(x,\ y)$, $Q=Q(x+dx,\ y+dy)$ and $R=R(x+dx,\ y)$; consider the flow along the broken line PQR assuming that the vector of the flow rate is $\mathbf{v}(x,\ y)$ and prove that (accurate to infinitesimals of the first order) this flow equals $d\psi=\psi(x+dx,\ y+dy)-\psi(x,\ y)$. **17.3.2** Use the fact that the tangent vectors to the lines $\varphi=$ constant and $\psi=$ constant at point $M(x,\ y)$ are $\mathbf{v}_\varphi=\left(-\frac{\partial\varphi}{\partial y},\ \frac{\partial\varphi}{\partial x}\right)$ and $\mathbf{v}_\psi=\left(-\frac{\partial\psi}{\partial y},\ \frac{\partial\psi}{\partial x}\right)$, i.e. $\mathbf{v}_\varphi=(v_y,\ -v_x)$ and $\mathbf{v}_\psi=(v_x,\ v_y)$, and then take advantage of the condition of two vectors being perpendicular to each other. **17.3.4.** (a) When a is real, the streamlines are straight lines parallel to the x axis; equipotential lines are parallel to the y axis; when a is pure imaginary, the flow is conjugate to that described. (b) When a is real, the streamlines are hyperbolas $xy=$ constant; equipotential lines are hyperbolas $x^2-y^2=$ constant; the flow rate at point M is proportional to the distance OM. (c) When a is real, the streamlines are the circles tangent to the Ox axis at point O; equipotential lines are circles tangent to Oy axis at point O. (d) For $w=\ln z$ the streamlines are straight lines emanating from point O; equipotential lines are circles with center at O.

17.4.1. Equating to zero (equilibrium) the sum of the projections of forces onto the direction of the y axis we get (for the case of small deviations, $y\ll l$) $1-k\frac{y_1}{x_1}-k\frac{y_1}{l-x_1}=0$. Whence

$$y_1(x_1)=1\Big/\left(\frac{k}{x_1}+\frac{k}{l-x_1}\right)=x_1\frac{(l-x_1)}{kl},$$

$$y(x,\ x_1)=\begin{cases}\left(\frac{x}{x_1}\right)\Big/\left(\frac{k}{x_1}+\frac{k}{l-x_1}\right)\\ =\frac{x(l-x_1)}{kl},\quad 0\leqslant x\leqslant x_1,\\ \left(\frac{l-x}{l-x_1}\right)\Big/\left(\frac{k}{x_1}+\frac{k}{l-x_1}\right)\\ =\frac{x_1(l-x)}{kl},\quad x_1\leqslant x\leqslant l.\end{cases}$$

The function $y(x,\ x_1)$ is called *Green's function* in the problem on the string. For an arbitrarily distributed force $f(x)$ the deviation of the string is given by the formula $y(x)=\int\limits_0^l f(x_1)\,y(x,\ x_1)\,dx_1$. Note that Green's function helped us find the solution to the function $y(x)$ when we did not even know the

equation of its behavior (this equation has the form $\frac{d^2y}{dx^2}=-\frac{f(x)}{k}$, $y(0)=0$, $y(l)=0$).

17.4.2. The general solution to the equation without the force is $x(t)=C_1\sin\omega t+C_2\cos\omega t$, $\omega=\sqrt{k/m}$, C_1 and C_2 are arbitrary constants. Since the delta function differs from zero only at $t=\tau$, the solution to the equation with the delta-like force and the state of rest at $t=-\infty$ has the form

$$x(t)=\begin{cases}0 \\ \text{for } -\infty<t<\tau, \\ C_1\sin\omega t+C_2\cos\omega t \\ \text{for } \tau<t<+\infty;\end{cases} \qquad (*)$$

the delta-like force imparts a unit impulse to the body, therefore after the action of the delta force the body at rest acquires an initial velocity $v_0=\frac{\Delta p}{m}=\frac{1}{m}$, the initial position remaining zero. The solution to the equation of oscillations with such initial conditions at $t=\tau$ has the form

$$x(t,\ \tau)=\begin{cases}0 & \text{for } -\infty<t<\tau, \\ \frac{1}{m\omega}\sin\omega(t-\tau) & \text{for } \tau<t<+\infty.\end{cases}$$

In other words, in formula (*) $C_1=\frac{\cos\omega\tau}{m\omega}$, $C_2=-\frac{\sin\omega\tau}{m\omega}$.

The solution to the problem with an *arbitrary* force $f(t)$ is

$$x(t)=\int_{-\infty}^{+\infty} f(\tau)\,x(t,\ \tau)\,d\tau$$

$$=\int_{-\infty}^{t} f(\tau)\,\frac{1}{m\omega}\sin[\omega(t-\tau)]\,d\tau.$$

Name Index

Abel, N.H. 527
Aleksandrov, V.V. 530
Archimedes 120
Argand, J.R. 474
Arnold, V.I. 522
Arrhenius, S. 17, 395, 396

Barlow, V. 524
Barrow, I. 124
Basov, N.G. 416
Bell, E.T. 531
Bernoulli, D. 209, 211, 378, 380, 384, 385, 386, 475, 518
Bernoulli, Jacob 209, 210, 231
Bernoulli, Johann 209, 210, 211, 483, 484
Bers, K. 530
Bessel, F.W. 484
Bohr, N. 284, 406
Boltzmann, L. 394
Bolyai, J. 521
Bombelli, R. 474
Born, M. 528
Bose, S.N. 410
Briggs, H. 146
Bronsthein, I.N. 170
Brown, R. 393
Bürgi, J. 146
Burstein, S. 530
Bueche, F.J. 530

Cardano, G. 473,474
Cassini, G.D. 230
Cauchy, A.L. 125, 212, 484, 520, 527
Cavalieri, B. 272, 274
Chung, K.L. 523
Cooper, L.N. 530
Courant, R. 530
Curie, M. 283
Curie, P. 283

d'Alembert, J. 211, 212, 352, 379, 380, 384, 385, 386, 474, 475, 483, 518, 521
Descartes, R. 121
Diderot, D. 211, 521
Dirac, P.A.M. 414, 487, 528
Downes, Jr., F.L. 13
Dubrovin, B.A. 521
Dwight, H.W. 170

Ehrenfest, P.S. 523
Ehrenfest-Afanasjewa, T.A. 523
Einstein, A. 14, 16, 286, 287, 321, 323, 393, 410, 412, 414, 523
Eisberg, R.M. 530
Euclid, 121
Euler, L. 210, 211, 212, 379, 384, 385, 386, 474, 475, 483, 518

Faraday, M. 453, 454
Fedorov, E.S. 524
Feller, W. 523
Fermat, P. 121, 122, 130
Feynman, R.P. 530
Fikhtengol'ts, G.M. 530
Flyorov, G.N. 289, 294
Fokker, A. 523
Fomenko, A.T. 522
Fourier, J.B. 212, 379, 386

Galilei, G. 113, 121, 322
Galois, E. 524, 527
Gauss, K.F. 213, 380, 474, 484, 521
Gel'fand, I.M. 488
Gibbs, W. 112
Gnedenko, B.V. 523
Goethe, J.W. von 112
Gradshteyn, I.S. 170
Green, G. 515
Grenville, V. 13
Guldin, P. 274

Hahn, O. 294
Hall, A.R. 125
Halliday, D. 530
Hamilton, W.R. 525
Heisenberg, W. 526, 528
Hermite, Ch. 15
Hertz, H. 358, 454
Helbert, D. 213, 521, 528
Hoffmann, L.D. 530
Hooke, R. 115
Huygens, Ch. 122, 123, 210, 270

Infeld, L. 524
Ioffe, A.F. 397, 398
Jeans, J.H. 524
Jensen, J.L.W.V. 238
Joliot-Curie, F. 294
Joliot-Curie, I. 294
Jordan, C. 527

Kaempffer, F.A. 530
Kamerlingh-Onnes, H. 438
Kemery, J.G. 523
Kepler, J. 121, 274
Khinchin, A.Ya. 523
Kitaigorodsky, A.I. 530
Klein, F. 521, 527
Kline, M. 473
Kolmogorov, A. 523
Kopylov, G.I. 531
Krylov, A.N. 16
Kurchatov, I.V. 294

Lagrange, J.L. 123, 124, 210, 212, 521, 527
Landau, L.D. 530
Landsberg, G.S. 530
Lang, S. 530
Lax, A. 530
Lax P. 530
Leibniz, G. 10, 113, 122, 123, 130, 209, 210, 211, 261, 483, 484
Leighton, R.B. 530
Lerner, L.S. 530
Lévy, P.P. 523
L'Hospital, G.F.A. 210, 211
Lie, S. 527, 528
Lobachevsky, N. 521
Lomonosov, M.V. 391
Lorentz, H.A. 17
Luzin, N.N. 13

Mach, E. 350
Maclaurin, C. 209
Mandelbrot, B. 15
Manin, Yu.I. 522
Marconi, M.G. 454
Marion, J.B. 530
Markushevich, A.I. 252
Marsden, J. 530
Maxwell, J.C. 394, 453, 454, 518, 525
Mendeleyev, G. 287
Mikusinski, J. 488
Moise, E.E. 13
Moivre, A. de 474
Monge, G. 212
Mosteller, F. 523
Myškis, A.D. 530

Napier, J. 146
Natanson, I.P. 530
Nernst, W. 17
Newman, M.H.A. 528
Newton, I. 10, 16, 113, 122, 123, 130, 209, 321, 484
Neyman, J. 523

Nikolsky, S.M. 530
Niven, I. 458, 530
Novikov, S.P. 522

Orear, J. 530

Pappus of Alexandria 274
Pascal, B. 126
Pasichenko, P.I. 530
Perelman, Ya.I. 141
Perrin, J.B. 393
Petrzhak, K.A. 294
Planck, M. 410, 523
Poincaré, H. 521
Poisson, S.-D. 527
Polya, G. 17, 209
Pontrjagin, L.S. 530
Pope, A. 124
Popov, A.S. 454
Poston, T. 522
Potapov, M.K. 530
Prokhorov, A.M. 416

Reid, C. 528
Resnick, R. 530
Reynolds, O. 346
Riemann, G.F.B. 212, 213, 483, 484, 485, 520, 521
Robbins, H. 530
Robinson, A. 126
Rogers, E.M. 530
Rolle, M. 259
Röntgen, W.K. 454
Rourke, R.E.K. 523
Ruffini, P. 527
Rutherford, E. 283
Ryzhik, I.M. 170

Sands, M.L. 530
Schönflies, A.M. 524
Schwartz, L. 488
Seaborg, G. 287, 288
Sears, F.W. 530
Semendyayew, K.A. 170
Semyonov, N.N. 396
Shervatov, V.G. 472
Shubnikov, A.V. 524
Sikorski, R. 488
Simpson, T. 255
Smirnov, V.I. 530
Smoluchowski, M. 523
Snell, J.L. 523
Sobolev, S.L. 488
Steiner, J. 61
Stieltjes, T.J. 15
Stirling, J. 249
Stokes, G.G. 346
Strassmann, F. 294
Struik, D.J. 531
Stuwart, I. 522
Tartaglia, N. 193, 473
Taylor, B. 209, 385
Thom, R. 522
Thomas, G.B. 523
Thomson, W. (Lord Kelvin) 15, 388, 515
Torricelli, E. 121, 204
Townes, Ch. 416
Tsiolkovsky, K.E. 334

Viete, F. 474
Vygodsky, M. 530

Weierstrass, K.T.W. 484, 520
Weinstein, A. 530
Wessel, C. 474
Weyl, H. 528
Wigner, E. 519, 528

Yaglom, A.M. 58
Yaglom, I.M. 58, 327, 509, 527
Yang, C.N. 519
Young, H.D. 530

Zeeman, E.C. 522
Zeldovich, Ya.B. 13, 530
Zemansky, M.W. 530
Zhukovsky, N.E.

Subject Index

Abel groups 527
abscissa 26
absolute maxima (minima) of function 226
absolute space 321
absolute time 321
absolute value of complex number 459
acceleration 69, 89, 113
activation energy 395
algebra, fundamental theorem of 461
algebraic equation of nth degree 461
algebraic function 139
alternating current 444
amplitude of complex number 460
amplitude of oscillations 359
analytic functions 477
  of one complex variable 480
analytic geometry 121
angle 27f, 61f
angle of departure 328
antiderivative 106, 198
approximations of functions 178
Archimedes spiral 120
arc length 260
area bounded by curve 95
area of triangle 62
argument of complex number 460
argument of function 23
arithmetic mean 95, 236
arithmetic progression 141
astroid 61
asymptote(s), horizontal 278
  of a hyperbola 36
  vertical 278
atmospheric pressure 117
average (or specific) curvature 265
average velocity 68, 149
Avogadro's law 388
Avogadro's number 389
axis, polar 28
axis of abscissas 26
axis of ordinates 26

ballistic pendulum 360
base-10 logarithms 146
beats 374
binary logarithm 58
bit 58
Boltzmann constant 389
boundary conditions for differential equations 376
boundary value of independent variable 194
Boyle's law 388
breakdown potential 427

Briggs logarithms 146
Brownian movement 393
bulk expansion coefficient 72

calculus, differential and integral 101, 120
capacitance 117, 419
Cardano formula 473
cardioid 61
Cartesian coordinates 26
catastrophe theory 522
catenary curve 261
  length of 261
cathode-ray oscillograph 445
Cauchy-Riemann equations 479
Cavalieri's theorem 272
center of curvature 267
center of gravity 274, 341
center of symmetry 43
centroid 238, 274
chain reaction 294ff
chain rule 134
characteristic curve 454
circle, involute of 270
  length of circumference of 261, 264
  osculating 267
circular frequency of oscillations 358
Clapeyron ideal gas law 388
coefficient of linear expansion 71
coefficient of volume expansion 72
common logarithms 146
commutative groups 527
complex electric field strength 511
complex number(s) 459
  absolute value of 459
  amplitude of 460
  argument of 460
  geometric representation of 460
  modulus of 459
  phase of 460
  trigonometric form of 460
complex plane 460
complex potential 510
  of electric field 511
complex velocity of flow 510
composite function 134
  derivative of 134
concave curve 235
concave downward curve 91, 235
concave function 235
conductance 419
cone, volume of 271f
conformal transformations 483
conjugate flows 511
conjugate numbers 459
conjugation of curves 225
constant, Euler's 246
continuous spectrum 386
convective rate of temperature variation 160
convergent series 188
convex curve 234
convex function 234
convexity 214
  of curve 41
  point of 214
convex upward curve 91, 234
coordinates, Cartesian 26
  polar 28
corner point 222
cosine, hyperbolic 168, 261
cosine line 420
crankgear 373
critical mass 298ff
critical point 88
critical temperature 234
cross section 399
  effective 403
cubic equation 44
cubic function 57
curvature of a curve 265
curvature at a point 265
curve(s), average (or specific) curvature of 265
  catenary 261
  center of curvature of 267
  concave 235
  concave downward 91
  of constant curvature 268
  convex downward 235
  convexity of 41
  convex upward 91, 234
  effective width of 243
  equipotential 510
  integral 201
  parametric representation of 60
  radius of curvature of 267
  rectifying points of 267
  shrinking along x axis 54
  shrinking to x axis 54
  smooth 215
  Steiner 61
  stretching along x axis 53
  stretching away from x axis 53
  translation of 50
  vertices of 267
curvilinear coordinates 520
curvilinear trapezoid 99
cusp 223
cuspidal maxima (minima) of a function 224
cycloid 61
  length of 261f

damped oscillations 441, 466
daughter element 289
decay time 431
decrease of function 85
definite integral 96
deformation 115
delta function 487
density, linear 337
  distribution-of-gas 516
dependence, empirical 26
derivative 73
  of composite function 134
  of implicit function 158f
  of inverse function 136
  on the left 224
  logarithmic 147
  of order n 177
  partial 154
  on the right 224
  second 89, 176
  total 161
deuterium 286
Dido's problem 218
dielectric constant 425
differential 127
differential equation(s) 116, 198
  general solution to 204
  initial condition for 199
  order of 198
  partial 453
  partial solution to 204
  second 131
  total 157
  with variables separable 199
differentiation of composite function 134
differentiation, rules of 132-142
diodes 454
Dirac's delta function 487
direction of current 418
directions, field of 201
direct proportionality 34
disintegration, probability 281, 289
distance 27ff, 67
distance from point to straight line 63
distribution, steady-state 160
divergent series 189
domain of integration 98
dummy variable 100

e (the number) 142
effective altitude of figure 243
effective cross section 403
effective length of figure 243
effective width of curve 243
einstenium 287
Einstein-Smoluchowski equation 523
electric circuit(s) 418
electric current 418
electric field strength 517
electricity, quantity of 117
electric potential 418
electromagnetic field theory 511
electromagnetic theory of light 453
electromotive force 420
elementary function 169

ellipse 53
ellipsoid of revolution, volume of 273
emission of electrons 397
empirical dependence 26
empirical formula 26
energy, activation 395
kinetic 115, 320
law of conservation of 354
total 354
energy density 408
energy flux (or irradiance) 399
energy of inductance 435
energy of radiation 409
epicycloid 61
equation(s), algebraic of *n*th degree 461
Cauchy-Riemann 479
cubic 44
differential 116, 198
Einstein-Smoluchowski 523
Fokker-Planck 523
Kolmogorov 523
of mathematical physics 212, 213, 517
quadratic 39
of straight line 31
equation of state, van der Waals 233
equilateral hyperbola 35
equilibrium, stable 313
thermodynamic 407
unstable 313
equilibrium intensity of light 408
equipotential curve 510
escape velocity 332
essential singular point 481
ether 518
Euler broken line 201
Euler formula(s) 465, 466
Euler's constant 246
evaporation of liquids 396
even function 56
evolute 269
of a cycloid 270
of a parabola 269
exponential curve, arc length of 264
exponential function 144
extrapolation 26

Fedorov group 524
Fermat's principle 125, 221
fermium 287
field, electric 517
magnetic 453
field of directions 201
first approximation for functions 177
first theorem of Pappus 274
flow, complex velocity of 510
irrotational (or votex-free) 509
plane-parallel 509
potential 509
rate of 203
fluent 124
fluid dynamics 350
fluid mechanics 509
fluxion 124
Fokker-Planck equation 523
forced oscillations 366, 497f
force, impulse of 317
work performed by 114
force of gravity 310
formula, empirical 26
Fourier analysis 386
Fourier analyzer 386
Fourier integral 386, 495
Fourier series 386, 495, 498
fourth-power law 410
fractional (rational) numbers 458
frequency 55, 358
natural 366, 444
function(s) 23
algebraic 139
analytic 477
function(s)
approximations of 178
argument of 23
of a complex variable 212, 213
composite 134
concave 235
convex 234
cubic 57
cuspidal maxima (minima) of 224
decrease of 85
dependent on a parameter 173
elementary 169
even 56
exponential 144
first approximation for 177
Green 499
growth of 85
harmonic 315, 480
hyperbolic 469ff
implicit 158
increase of 85
infinitely valued 152
inverse 47, 135, 467
inverse trigonometric 151
linear 30
linear approximation for 177
linear-fractional 51, 57
locally monotonic 50
longarithmic 145, 468
maximum of 38, 83, 226
mean value of 253
minimum of 38, 83, 226
monotonic 49
multiple-valued 152, 159
nonsmooth maxima (minima) of 224
odd 56
order of increase of 196
parametric representation of 136
period of 55
periodic 55
pole of 505
power 46, 138
primitive 106
quadratic 37
rate of change of 177
rate of growth of 86
second derivative of 89
signum 491
sine integral 187
stream 510
trigonometric 148, 466
two-valued 50
of two variables 154
functional 218
functional analysis 213
functional relationship 23
fundamental theorem of algebra 461
fundamental theorem of higher mathematics 101
fundamental tone 385

Galilean transformations 325
Galois group 528
gaseous state of matter 234
gas pressure 388
general solution to differential equations 204
general theory of relativity 323
geometric mean 236
geometric progression 141, 188
glow discharge 427
graphs, transformation of 50
graph of function 30
gravitational paradox 325
gravity, force of 310
grazing fire 329
Green function 499, 515
group(s), Abel (or commutative) 527
continuous 526
discrete 526
Fedorov 524
Galois 528
Klein 527
Lie 527
Shubnikov 524
symmetry 524
growth of function 85
Guldin theorems 274

half-life of radioactive atoms 282
half-width at half-maximum 451
half-width of resonance curve 451
harmonic analysis 380
harmonic functions 315, 480
harmonic mean 236
harmonic oscillations 360
hertz 358
higher mathematics, fundamental theorem of 101
homothetic transformation 42
Hooke's law 115
horizontal asymptote 278
hyperbola 35
of degree *n* 45
equilateral 35
of order $n + 1$ 45
hyperbolic angle 471
hyperbolic cosine 168, 261, 469
hyperbolic functions 469ff
hyperbolic sine 168, 469
hyperbolic tangent 470
hypocycloid 61

ideal-gas isotherms 233
ideal gas law 233, 388
imaginary axis 460
imaginary power 463
imaginary unit 459
implicit functions 158
derivative of 158f
improper integral 174
impulse of force 317
unit 335
increase of function 85
indefinite integral 104, 198
independent variable 23
index 29
inductance 421f
energy of 435
induction 453
inequality, Jensen's 238
inertial systems of coordinates 322
infinitely valued functions 152
infinite series 181
infinite slope 33
information, theory of 58
initial conditions for differential equations 199, 376
initial condition for radioactive decay 281
initial ordinate of straight line 31
initial phase angle 359
initial velocity 69
instantaneous velocity 69, 149
integral(s), definite 96
Fourier 386, 495
improper 174
indefinite 104, 198
tabulated 168
integral curve 201
integrand 98
integrating functions dependent on a parameter 173
integration, changing the variables in 168
domain of 98
limits of 97
by parts 167
by substitution 168
variable of 98

internal resistance 420
interpolation 26
inverse function(s) 47, 135, 467
  derivative of 136
inverse proportionality 35
inverse trigonometric functions 151
involute 269
  of a circle 270
irrational numbers 458
irrotational (or vortex-free) flow 509
isotherms 233
  ideal gas 233
  van der Waals 233

Jensen's inequality 238
  general 240

kinetic energy 115, 320
Klein groups 527
Kolmogorov equation 523

Lagrange's theorem 258
laser 416
law(s), Avogadro's 388
  Boyle's 388
  Clapeyron ideal gas 388
  of conservation of energy 354
fourth-power 410
  Hooke's 115
  ideal gas 388
  of large numbers 404
  of light refraction 221
  of motion 113
  Newton's first 316
  Newton's second 316
  Newton's third 320
  Ohm s 24, 36
  of radioactive decay 281
  statistical 523
  Stefan-Boltzmann 410
  Torricelli's 204
law of radiation, Stefan's 410
leakage resistance 437
Leibniz and discovery of calculus 125
leminscate of Bernoulli 231
L'Hospital's rule 196
Lie group 527
lifetime of radioactive atom 282
light, electromagnetic heory of 453
  equilibrium intensity of 408
  spontaneous emission of 413
  stimulated emission of 413
light absorption 406f
light emission 406f
light quanta 406
limit 72
limits of integration 97
linear approximation for functions 177
linear density 337
linear expansion coefficient 71
linear-fractional function 51, 57
linear function 30
linear relationship 30
lines of force 511
liquid state of matter 234
locally monotonic function 50
local maximum of function 44, 87
local minimum of function 87
local rate of temperature variation 160
logarithm(s) 145f
  base-10, 146
  binary 58
  Briggs 146
  common 146
  modulus of base a with respect to base b 56
  Napierian 146
  natural 129
logarithmic derivative 147
logarithmic function 145, 468
loop oscillograph 445

Mach number 350
Maclaurin's series 182
magnetic field 453, 517
maser 416
mathematical analysis 101
mathematical modeling 519
maximum of function 38, 83, 214, 215
  absolute 226
  cuspidal 224
  at endpoint 222
  local 44, 87
  nonsmooth 224
  relative 44
maximum of function
  smooth 215
mean, arithmetic 95, 236
  geometric 236
  harmonic 236
mean free path 391
mean value of function 253
mendelevium 287
method, rectangular 96
  trapezoid 95
minimum of function 83, 214, 215
  absolute 226
  cuspidal 224
  at endpoint 222
  local 87
  nonsmooth 224
  smooth 215
modulus, of base a with respect to base b 56, 145
  of complex number 459
  Young's 115
Moivre formula 461
molar gas constant 233, 388
moment of inertia 339
momentum 317
monotonic function 49
  locally 50
motion, law of 113
  uniform 67f
  uniformly accelerated 69, 113
multiple-valued function 49, 152, 159

Napierian logarithms 146
natural frequency 366
natural frequency of circuit 444
natural logarithm 146
natural numbers, sum of powers of 245
negative numbers 458
neutron 526
neutron flux 297
Newton and discovery of calculus 124
Newton and Leibniz, controversy over priority in discovery of calculus 125f
Newton-Leibniz theorem 101
Newton's first law 316
Newton's second law 316
Newton's third law 320
nonsmooth maxima (minima) of a function 224
nth derivative 177
number(s), complex 459
  fractional (rational) 458
  irrational 458
  negative 458
  pure imaginary 459
  real 458
number of disintegrations 284

odd function 56
Ohm's law 24, 36
order of differential equation 198
order of increase of function 196
order of smallness 80
ordinate 26
origin of coordinates 26
osculating circle 267
osculating parabola 131
oscillation(s) 357
  damped 466
  forced 366, 497f
  period of 358
oscillatory circuit 439
oscillograph, cathode-ray 445
  loop 445
  single-beam cathode-ray 447
overtones 385

Pappus-Guldin theorems 274
Pappus theorem, first 274
  second 275
parabola 37
  arc length of 263
  of order n 42
  osculating 131
  semicubical 45
  squaring of 120
parallel lines 65
parameter 60
parametric representation of curve 60
parametric representation offunction 136
parent element 289
partial derivative(s) 154
  of second order 156
partial differential equations 453
partial solutions to differential equations 204
pendulum, ballistic 360
  physical 362
  simple 361
period of function 55
periodic function 55
period of oscillations 358
perpendicularity 62
pi (the number) 264
phase of complex number 460
phase space 521
photons 406
physical pendulum 362
Planck's constant 406, 410
Planck's radiation formula 410
plane-parallel flow 509
plunging fire 329
point of convexity 214
point of inflection 91, 268
polar axis 28
polar coordinates 28, 460
pole of function 505
pole of polar system of coordinates 28
positrons 487
potential difference 419
potential energy 309, 354
potential flow 509
polonium 283
power 303
power function 46, 138
power output 428, 435
pressure, atmospheric 117
primitive function 106
principal value of Arc sin x 152
principle, Fermat's 125, 221
probabilistic process 523
probability 519, 522, 523
  of disintegration 281, 289
problem of vibrating string 375
progression, arithmetic 141
  geometric 141, 188
proportionality factor 34
protons 526
pure imaginary numbers 459

quadratic equation 39
quadratic function 37
quantity of electricity 117, 418
quantum mechanics 405, 519
quaternions 525

radian 148
radiant flux 399

radioactive atom(s), half-life of 282
lifetime of 282
rate of disintegration of 245
radioactive decay, initial conditions for 281
law of 281
radioactive family 289
radium 283
radius of curvature 265, 267
random event 405
rate of change of function 177
rate of disintegration 283
rate of energy release 435
rate of growth of function 86
rates of temperature variation 160
reaction (or jet) propulsion 334
real axis 460
real numbers 458
rectangular method 96
rectifier 423
rectifying points of a curve 267
relationship, functional 23
linear 30
relative maximum 44
remainder 180
residue 505
resistance 419
leakage 437
resonance 306, 366, 445, 450, 452, 497, 518
Reynolds number 346, 350
Rolle's theorem 259
root-mean-square 236
rule, L'Hospital's 196

safety parabola 330
salient point 222
satellite (or orbital) velocity 331
secant line 80
second derivative of function 89, 176
second differential 131
second-order partial derivatives 156
second theorem of Pappus 275
semicubical parabola 45
series, convergent 188
divergent 189
Fourier 386, 495, 498
infinite 181
Maclaurin's 182
Taylor's 182, 214
series expansion 264
Shubnikov group 524
signum function 491
simple pendulum 361
Simpson's rule 255
sine, hyperbolic 168
sine-integral function 187
sine line 471
single-beam cathode-ray oscillograph 447
sink 510
slope, infinite 33
slope of line 32
slope of tangent 81
smooth curve 215
smooth maxima of function 215
smooth minima of function 215
solid of revolution, surface area of 273
volume of 272
source 510
source of voltage 420
space, absolute 321
spark gap 426
special theory of relativity 521
specific elongation 71
specific heat capacity 70
sphere, volume of 272
spherical layer (or zone) 273
surface area of 273
spherical segment (or cap) 273
surface area of 273
spontaneous emission of light 413
spring 309
squaring the parabola 103
stable equilibrium 313
static moment 341
statistical laws 523
stationary state 291
steady state 291
steady-state distribution 160
steady-state mode 206
Stefan-Boltzmann law 410
Stefan's law of radiation 410
Steiner curve 61
stimulated emission of light 413
Stirling's formula 249
Stokes law 346
Straight line, equation of 31
initial ordinate of 31
stream function 510
streamline 510
stretching of wire 115
string 375
subcritical mass 299
sum of powers of natural numbers 245
superconducting state 438
superconductivity 438
supercritical mass 299
superposition principle 379, 453, 497, 513
surface 157
surface area of solid of revolution 273
surface area of spherical layer 273
surface area of spherical segment 273
surface area of torus 275
symmetry, center of 43
about axis 43
symmetry axis 37, 43
symmetry group 524

tabulated integral 168
tangent line 80, 471
tangent plane 157
Taylor's series 182, 214
theorem(s), Cavalieri's 272
Guldin 274
Lagrange's 258
Newton-Leibniz 101
Pappus-Guldin 274
Rolle's 259
theorem of algebra, fundamental 461
theorem on geometric and arithmetic means 239, 241
theorem of higher mathematics, fundamental 101
theorem of Pappus, first 274
second 275
theory of continuous groups 526
theory of discrete groups 526
theory of electromagnetic field 511
theory of electromagnetism 517
theory of groups 523
theory of information 58
theory of probability 523
thermal expansion 71f
thermodynamic equilibrium 407
time, absolute 321
time constant of a current 424
time lag in interaction 517
topology 521
Torricelli's law 204
torus 275
surface area of 275
volume of 275
total derivative 161
total differential 157
total energy 354
total rate of temperature variation 160
trajectory 59
transformation, homothetic 42
transformation of groups 50
translation of curve 50
trapezoid, curvilinear 99
trapezoid method 95
triangle, area of 62
trigonometric functions 148, 466
of imaginary argument 469
inverse 151
trigonometric series 380
Tsiolkovsky's formula 335
turbulence 352
two-terminal network 422
two-terminal pair network 422
two-valued function 50

uniformly accelerated motion 69, 113
uniform motion 67f
unit impulse 335
universal gas constant 388
unlimited figure 242
unstable equilibrium 313
uranium 283

van der Waals equation of state 233
van der Waals isotherms 233
variable, dummy 100
independent 23
of integration 98
vector 328
vector calculus 525
velocity 67
velocity, average 68, 149
escape 332
initial 69
instantaneous 69, 149
satellite (or orbital) 331
velocity potential 509
vertical asymptote 278
vertices of a curve 267
vibrations of a string 375
viscosity 346
volume 118
of cone 271f
of ellipsoid of revolution 273
of pyramid 118f
of solid of revolution 272
of sphere 272
of torus 275
volume expansion coefficient 72
vortex 510
vorticity 510

wire, stretching of 115
work performed by force 114

Young's modulus 115