序

我很高兴你看到了“线性代数引论第五版”，这是我在 MIT 开放课程(**ocw.mit.edu** 以及 **YouTube**)的视屏演讲的文字版。我希望这些演讲对你有所帮助(或许也是愉悦的！)

上百所学院与大学已经选用这本教科书作为基础线性代数课程的教材，一个较长的假期给了我机会来准备关于概率与统计与理解性材料的两个新章节，其中也包含了只有作者可知的上千处改善...以下是关于学生与读者的新增论述：

> 每个段落以简要的摘录作为开头来说明段落内容，当你读到一个新段落，以及当你重新复习个段落并且在心中整理时，这些条目是一个快速指引，而且可以协助记忆。

另一个较大的变化来自本书的网址 **math.mit.edu/linearalgebra**，这个网址现在包含了书中问题集的解答。在无限的空间中，可以比印出简短解答更具弹性，我们有三个关键网址：

ocw.mit.edu 这个开放课程网址包含来自成千上万学生与组织的关于线性代数的信息。18.06 与 18.06 SC 课程包含了一个完整课程学期的演讲视频，这些演讲提供基于本书整体主题的独立复习––教授的时间保持自由，学生时间固定在凌晨 2 点(读者完全不需要参与。) 全世界六百万的观众看过这些视频(惊人！)，我希望你觉得这些对你有帮助。

web.mit.edu/18.06 这个网址包含目前教授课程的作业与考试(附有解答)，可以回溯到 1996 年。网站还包含复习问题，Java 展示，教学代码，以及短论文(以及演讲视频)，我的目的是基于我们能够提供的课程材料，尽可能让本书对你有所帮助。

math.mit.edu/linearalgebra 这个网址目前非常活跃，包含了习题解答，而且有观念说明。同时也包含来自许多不同课程的习题作为复习––实际问题，教科书范例的发展，MATLAB 代码，Julia，Python，以及考试集(18.06 与其他)。

请参访线性代数网站，发送建议到 **linearalgebrabook@gmail.com**

第五版

本书封面展示目**四个基础子空间**--左侧是行空间与零空间，右侧是列空间与 A^{T} 的零空间，这样展示主题中心概念的方式并不多见！当你在第三章接触这四个空间时，你会了解为什么这张图是线性代数的中心思想。

本书把这些空间称为四个基础子空间，他们全部来自一个矩阵 A。A 的每一行都是 n 维空间中的向量，当矩阵具有 m 个行时，每一列都是 m 维空间中的向量。线性代数中最重要的运算是取得**列向量的线性组合**，这正是矩阵-向量乘积的结果，Ax 是 $\boldsymbol{A}$ **的列**的一个组合。

当我们取得列向量的全部组合 Ax 时，我们就得到**列空间**，如果这个空间包含向量 b，我们可以求解方程式 $Ax = b$。

我会给予段落 1.3 更多的关注，段落中较早引进两个特殊范例，没有要求你在一天内了解向量空间的每个细节！但是你会在本书看到重要的矩阵，以及他们的列空间图形，这里甚至存在**逆矩阵**与微积分的关联，你将以最佳与最有效的方式学习线性代数的语言：用就对了！

基础课程的每个段落都会以大量的复习问题集作为结束，要求你使用该段落的观念--列空间的维度，空间的一组基，以及 A 的秩，逆，行列式与特征向量。很多问题需要你针对小型矩阵进行手算，这些都是获得盛赞的问题集合。**挑战问题**会更进一步，有些会更加深入。让我举个例子：

段落 2.1：数读矩阵的哪些行交换可以产生另一个数读矩阵？

段落 2.7：若 P 是一个交换矩阵，为什么某些幂次方 P^k 会等于 I?

段落 3.4：若 $Ax = b$ 与 $Cx = b$，对于每个 b 具有相同的解，那么 A 会等于 C?

段落 4.1：四个向量 r, n, c, l 在什么条件下可以作为一个 2×2 矩阵的行空间，零空间，列空间与左零空间的基？

课程的开始

方程式 $Ax = b$ 立即使用线性代数语言，向量 Ax 是 $\boldsymbol{A}$ **的列**的一个组合，方程式的问题在于找出一个可以产生 b 的组合。解向量 x 来自三种层次，全部都很重要：

1. **直接解**，使用前向消元与反向代入求 x。
2. **矩阵解**，使用逆矩阵：$x = A^{-1}b$ (若 A 具有逆矩阵)。
3. **特解**(对应 $Ay = b$)加**零空间解**(对应 $Az = 0$)。

   本书封面展示了向量空间解 $x = y + z$。

直接消元是科学计算中最常使用的演算法，矩阵 A 变成三角形--快速得到解。我们也看到四个子空间的基，但是不要一直练习消元法... 好点子来了。

每个新的超级电脑都会在 $Ax = b$ 进行测试：这是纯线性代数。但是纵使是超级电脑，也不想计算逆矩阵：太慢了。逆矩阵得到最简易的公式 $x = A^{-1}b$，但是速度不是最快。每个人必须知道行列式的计算甚至更慢--任何的线性代数课程都不应该从 $n \times n$ 矩阵的行列式公式开始，这些公式有他们的地位，却不是第一位。

本书的结构

如同序言所述，你可以看到本书的形态与目的，这项目的是严肃的，用来说明数学中美丽与有用的部分，你会看到线性代数概念如何强化关键概念--每个层次自然而来，每个人都可以心领神会。

下列是本书关于学习与教学的 12 个重点：

1. 第一章从向量与点积开始，如果同学已经学过，就快速聚焦在线性组合。段落 1.3 提供三个无关向量，他们的组合形成全部的 3 维空间；也提出了在同一平面的三个相关向量。**这两个例子就是线性代数的开始**。

2. 第二章展示 $Ax = b$ 的行图与列图，线性代数的中心是 A 的行与 A 的列的关联：相同数目却是非常不同的图形。然后开始矩阵的代数：一个消元矩阵乘 A 得到一个零，目的是捕获全部的程序--从 A 开始，用 E's 乘，最终得到 U。

   从 $A = LU$ 可以看出消元法的美妙，**下三角** L 实现前向消元步骤，**上三角** U 实现反向代入。

3. 第三章是线性代数的最佳层次：**子空间**。列空间包含所有列的组合，重要问题是：**需要多少列**？答案告诉我们零空间的维度，以及 A 的关键信息。我们接触了**线性代数的基础定理**。

4. 当方程式比未知数多的时候，几乎可以确认 $Ax = b$ 无解。我们不能抛弃每个接近但不是确切的量测值！当我们使用**最小二乘法**(也称最小平方法)求解时，关键在于 $A^{\mathrm{T}}A$。当 A 是矩形时(非方阵)，这个美好的矩阵在应用数学领域随处可见。

5. **行列式**给出以前发明公式的解答--克拉默规则，逆矩阵，n 维中的体积。我们不想计算这些公式，他们太花时间。但是 $\det A = 0$ 告诉我们矩阵是奇异的：这是**特征值**的关键。

6. **段落 6.1 说明 2×2 矩阵的特征值**，许多课程希望早一点看到特征值。因为 2×2 矩阵的行列式比较简单，直接从第三章的结果得到特征值是完全合理，关键方程式是 $Ax = \lambda x$。

   特征值与特征向量是了解方形矩阵的惊奇方法，他们不是用在 $Ax = b$，而是用在动态方程式 $du/dt = Au$。观念都是一样的：追随**特征向量**。在这些特别的反向中，A 的作用好像是一个数(特征值λ)，变成一维的问题。

   第六章的高光处是**对角化一个对称矩阵**，当所有的特征值都是正数，矩阵称为“正定”。这项关键概念联系整个课程––正主元与行列式与特征值与能量。我很认真的说明这项重点并且使用范例来解释。

7. 第七章是新的内容，介绍了**奇异值**与**奇异向量**。他们把所有的矩阵分割成简单片段，依据重要性进行排序。你会看到压缩图像的方法，特别的是你可以分析充满资料的矩阵。

8. 第八章说明**线性变换**，这是没有轴的几何与没有坐标的代数。当我们选择了一组基，我们得到最佳的可能矩阵。

9. 第九章从实数与向量转移到复数向量与矩阵，傅里叶矩阵 F 是我们看到的最重要的复数矩阵，而且快速傅里叶变换(用 F 与 F^{-1} 快速乘)是革命性的。

10. 第十章充满着比任何单一课程都要多的应用：

    10.1 图形与网络––引入关于基尔霍夫法则的边-点矩阵。
    10.2 工程用矩阵––平行于矩阵方程式的微分方程式。
    10.3 马可夫矩阵––如同谷歌页序的演算法。
    10.4 线性规划––一个新需求 $x \geq 0$，极小化成本。
    10.5 傅里叶级数––函数的线性代数与数字信号处理。
    10.6 电脑图形––利用矩阵移动、转、压缩图像。
    10.7 密码学的线性代数––这个新段落很有趣，希尔密码不是很安全，它使用模数算术：从 0 到 $p - 1$ 的整数。乘法得到 $4\times5 \equiv 1 \pmod{19}$。对于解码而言得到 $4^{-1} \equiv 5$。

11. 线性代数课程如何包含计算？这个可以开启矩阵的新认识––每位个班级会找到新平衡。MATLAB 与 Maple 与 Mathematica 在许多方面非常强大，Julia 与 Python 是免费而且可以直接网络下载，这些新编程语言同样强大。

12. 第十二章讨论的概率与统计是新内容，当随机变量与我们所得的协方差矩阵相关时，会有非常重要的应用。幸运的是他们是对称正定，此时需要第六章的线性代数。

线性代数的多样性

微积分大多关注在一个特殊运算(导数)与它的逆运算(积分)，当然我承认微积分很重要.....但是很多数学应用是离散而非连续，数字而非模拟，资料的世纪已经开始！在我的网站你可以看到一篇轻松小论文“太多微积分”，**事实是向量与矩阵已经变成必知的语言**。

语言的部分是矩阵的美妙多样性，让我举出三个例子：

对称矩阵 正交矩阵 三角矩阵

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \qquad \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

关键目标是学习“读懂”矩阵，你需要看到数字的意义，这正是数学的中心—模式与意义。

我使用了斜体与**粗体**字标示每页的关键字，我知道有时候读者想要快速阅读，寻找重点。

请让我以关于教授的思维作为结束，你可能觉得方向是正确的，并且怀疑你的学生准备好了没有，**就是给他们一个机会**！不夸张的说有成千上万的学生写信给我，通常都是建议，令人惊喜的是感谢。他们知道这门课程的目的，因为教授与这本书就在他们身旁。线性代数是难以置信的主题，享受它吧！

本书的协助

最大的鼓励是你感觉到所做的事情在生命中是值得的，许多慷慨的读者发来本书中的范例与修正(以及喜爱的矩阵)，谢谢你们。

有一位协助本书的每个字词，他是 Ashley C. Fernandes，他处理本书的 LATEX 档案。经过他的书已经有六本，新写或重写，专注精确也专注生活，与好朋友共同工作是快乐的生活方式。

在 MIT 数学系内或外的朋友更是美妙，Alan Edelman 处理 Julia(并且更多)。Alex Townsend 处理 7.1 的旗帜范例，Peter Kempthorne 处理 7.3 的金融范例：这些都非常杰出，Don Spickler 关于编码的网站特别优秀。我要感谢 Jon Bloom, Jack Dongarra, Hilary Finucane,Pavel Grinfield, Randy LeVeque, David Vogan, Liang Wang，以及 Karen Willcox。7.3 的特征脸孔来自 Matthew Turk 与 Jeff Jauregui，奇异值的进阶依靠 Raj Rao's 在密歇根的伟大课程得以加速。

本书得力于在牛津的快乐假期，谢谢你，Nick Trefethen 与每个人，特别是你，读者！祝你工作愉快。

作者的背景

这是我的线性代数的第九本书，我还是犹豫要不要介绍自己，毕竟数学还有读者才是最重要的。下个段落有一些自我介绍，用来说明是人写的书！

我出生在芝加哥，在华盛顿与辛辛那提以及圣路易读书，我的学院是 MIT(我的线性代数课程超级抽象)。之后去了牛津与 UCLA，然后回到 MIT 非常长的时间。我不知道有多少万的学生参与了 18.06(超过 6 百万人访问了 ocw.mit.edu)。新鲜的尝试是正确的，因为惊奇的主题只是给数学家看的—**我们要向全世界展示线性代数**。

我感恩可以一辈子教数学，尽在不言中。

Gilbert Strang

PS　我希望下本书(2018?)引进“从资料学习”，这个主题快速成长，特别是“深度学习”。已知一个旧资料的训练集合，我们可以近似新资料，这项近似只使用简单的线性函数 $f(x) = \max(0, x)$。它是 n 次的矩阵乘法的最佳化，得到深度学习：$x_1 = f(A_1x + b_1)$，$x_2 = f(A_2x_1 + b_2)$，…，$x_n = f(A_nx_{n-1} + b_n)$。在输入 x 与输出 x_n 之间存在 $n - 1$ 隐藏层—可以近似训练集 $F(x)$。

矩阵表

A	任意矩阵	P	排列矩阵
B	基本矩阵	P	投影矩阵
C	余子式矩阵	Q	正交矩阵
D	对角矩阵	R	上三角矩阵
E	消元矩阵	R	简化阶梯矩阵
F	傅里叶矩阵	S	对称矩阵
H	哈达马德矩阵	T	线性变换
I	单位矩阵	U	上三角矩阵
J	若尔当矩阵	U	左奇异向量
K	刚度矩阵	V	右奇异向量
L	下三角矩阵	X	特征向量矩阵
M	马尔可夫矩阵	Λ	特征值矩阵
N	零空间矩阵	Σ	奇异值矩阵

第一章
向量引论

线性代数的中心是两个运算——都是针对向量。我们把向量相加得到 $\boldsymbol{v} + \boldsymbol{w}$，我们用数字 c 与 d 乘向量得到 $c\boldsymbol{v}$ 与 $d\boldsymbol{w}$，组合这两种运算($c\boldsymbol{v}$ 加到 $d\boldsymbol{w}$)得到**线性组合** $c\boldsymbol{v} + d\boldsymbol{w}$。

线性组合 $$c\boldsymbol{v} + d\boldsymbol{w} = c\begin{bmatrix}1\\1\end{bmatrix} + d\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}c+2d\\c+3d\end{bmatrix}$$

范例 $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix}1\\1\end{bmatrix} + \begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}3\\4\end{bmatrix}$ 是 $c = d = 1$ 的组合。

线性组合是这个主题中最重要的概念！有时候我们需要一个特定的组合，选择 $c = 2$ 与 $d = 1$ 产生 $c\boldsymbol{v} + d\boldsymbol{w} = (4, 5)$。其他时候我们想要 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 全部的组合(来自所有的 c 与 d)。

向量 $c\boldsymbol{v}$ 落在一条直线，当 $\boldsymbol{w}$ 不在这条线上时，**组合 $c\boldsymbol{v} + d\boldsymbol{w}$ 会形成整个二维平面**。如果从 4 维空间的 4 个向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z}$ 开始，他们的组合 $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} + f\boldsymbol{z}$ 好像会形成一个空间——但不是绝对，向量以及他们的组合可以形成一个平面或是一条直线。

第一章说明建立所有事物的中心观念，我们从二维与三维向量开始，这些向量比较容易制图，之后再讨论更高的维度。线性代数最深刻的特征在于如何平顺的将这些步骤扩展到 n 维空间。纵使不可能画出 10 维的向量，你内心的蓝图依然完全正确。

这是本书的目的(推进到 n 维空间)。第一步是段落 1.1 与 1.2 的运算，段落 1.3 介绍三个基本概念。

1.1　向量加法 $\boldsymbol{v} + \boldsymbol{w}$ 与线性组合 $c\boldsymbol{v} + d\boldsymbol{w}$。

1.2　两个向量的点积(dot product) $\boldsymbol{v} \cdot \boldsymbol{w}$ 与长度 $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ 。

1.3　矩阵 A，线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$，解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$。

请尊重版权与译者的劳动成果，侵权必究！

1.1 向量与线性组合

1 $3\boldsymbol{v}+5\boldsymbol{w}$ 是向量 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的**线性组合** $c\boldsymbol{v}+d\boldsymbol{w}$ 的一个典型。

2 当 $\boldsymbol{v}=\begin{bmatrix}1\\1\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}2\\3\end{bmatrix}$ 时，前述的组合是 $3\begin{bmatrix}1\\1\end{bmatrix}+5\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}3+10\\3+15\end{bmatrix}=\begin{bmatrix}13\\18\end{bmatrix}$。

3 在 xy 平面，向量 $\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}+\begin{bmatrix}0\\3\end{bmatrix}$，首先横跨到 $x=2$ 再往上到 $y=3$。

4 组合 $c\begin{bmatrix}1\\1\end{bmatrix}+d\begin{bmatrix}2\\3\end{bmatrix}$ 形成整个 xy 平面，他们产生每个 $\begin{bmatrix}x\\y\end{bmatrix}$。

5 组合 $c\begin{bmatrix}1\\1\\1\end{bmatrix}+d\begin{bmatrix}2\\3\\4\end{bmatrix}$ 形成 xyz 空间中的一个平面。$\begin{bmatrix}1\\1\\1\end{bmatrix}$，$\begin{bmatrix}3\\4\\5\end{bmatrix}$ 形成相同的平面。

6 但是 $\begin{aligned}c+2d&=1\\c+3d&=0\\c+4d&=0\end{aligned}$ 无解，这是因为它的右侧 $\begin{bmatrix}1\\0\\0\end{bmatrix}$ 不在那个平面上。

“你无法把苹果与橙子相加”，以奇特的方式来说，向量也是一样的理由。我们有两个个别的数字 v_1 与 v_2，这项配对产生**二维向量 $\boldsymbol{v}$**：

$$\textbf{列向量 }\boldsymbol{v}\qquad \boldsymbol{v}=\begin{bmatrix}v_1\\v_2\end{bmatrix}\qquad \begin{aligned}v_1&=\boldsymbol{v}\text{的第一分量}\\v_2&=\boldsymbol{v}\text{的第二分量}\end{aligned}$$

我们把 $\boldsymbol{v}$ 写成一列(column)，而不是一行(row)，重点在于单一的字母 $\boldsymbol{v}$(**粗斜体字**)代表这项配对数字 v_1 与 v_2 (浅色斜体字)。

我们不是把 v_1 与 v_2 相加，我们是把**向量相加**。$\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的第一分量与第二分量仍然分开：

向量加法 $\qquad \boldsymbol{v}=\begin{bmatrix}v_1\\v_2\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}w_1\\w_2\end{bmatrix}$ 相加得到 $\boldsymbol{v}+\boldsymbol{w}=\begin{bmatrix}v_1+w_1\\v_2+w_2\end{bmatrix}$

减法追随相同的概念：$\boldsymbol{v}-\boldsymbol{w}$ 的分量是 v_1-w_1 与 v_2-w_2。

另一个基础运算是纯量(scalar)乘法(**也称为数乘**)，向量可以用 2 或–1 或任意数 c 去乘。要计算 $2\boldsymbol{v}$，用 2 乘 $\boldsymbol{v}$ 的每个分量：

纯量乘法 $\qquad 2\boldsymbol{v}=\begin{bmatrix}2v_1\\2v_2\end{bmatrix}=\boldsymbol{v}+\boldsymbol{v}$，$-\boldsymbol{v}=\begin{bmatrix}-v_1\\-v_2\end{bmatrix}$。

$c\boldsymbol{v}$ 的分量是 cv_1 与 cv_2，数字 c 称为“纯量”。

注意$-\boldsymbol{v}$ 与 $\boldsymbol{v}$ 的总和(sum)是零向量，以粗体 $\boldsymbol{0}$ 表示，与一般的数字 0 不相同，向量 $\boldsymbol{0}$ 的分量是 0 与 0。请原谅我一直在反复谈论向量与分量的差别，线性代数就是建立在 $\boldsymbol{v}+\boldsymbol{w}$ 与 $c\boldsymbol{v}$ 与 $d\boldsymbol{w}$ 的运算——**向量加法与纯量乘法**。

线性组合

我们现在结合向量加法与纯量乘法产生 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的“**线性组合**”。用 c 乘 $\boldsymbol{v}$ 以及 d 乘 $\boldsymbol{w}$，然后相加得到 $c\boldsymbol{v}+d\boldsymbol{w}$。

$c\boldsymbol{v}$ 与 $d\boldsymbol{w}$ 的总和是	线性组合 $c\boldsymbol{v}+d\boldsymbol{w}$。

四种特殊的线性组合是：和，差，零，纯量乘积 $c\boldsymbol{v}$：

$1\boldsymbol{v}+1\boldsymbol{w}=$ 向量的和，如图 1.1a

$1\boldsymbol{v}-1\boldsymbol{w}=$ 向量的差，如图 1.1b

$0\boldsymbol{v}+0\boldsymbol{w}=$ **零向量**

$c\boldsymbol{v}+0\boldsymbol{w}=$ 沿着 $\boldsymbol{v}$ 方向的向量 $c\boldsymbol{v}$

零向量永远是可能的组合(它的系数为零)，每次我们看到向量的“空间”都包含零向量。从大局来看，取得 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 所有的线性组合就是线性代数的工作。

图形让你看到向量，对代数来说，我们只需要向量的分量(例如 4 与 2)。向量 $\boldsymbol{v}$ 由箭头表示，箭头往右横跨 $v_1=4$ 个单位，再往上走 $v_2=2$ 个单位，终点的 x, y 坐标等于 4, 2。这个点是向量的另外一种表示法——我们有三种方式来描述 $\boldsymbol{v}$：

向量 $\boldsymbol{v}$ 表示法	**两个数字**	**由(0, 0)出发的箭头**	**平面上的点**

我们用数字做加法，我们用箭头视觉化 $\boldsymbol{v}+\boldsymbol{w}$：

向量加法(头到尾)，$\boldsymbol{v}$ 的终点就是 $\boldsymbol{w}$ 的起点

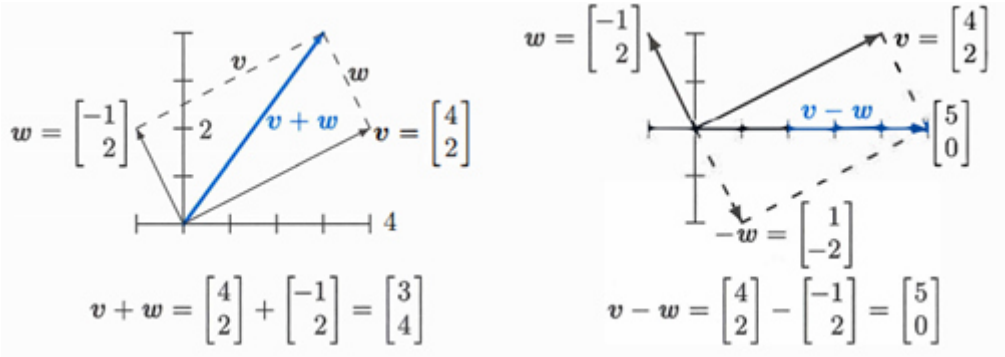


图 1.1：向量加法 $\boldsymbol{v}+\boldsymbol{w}=(3,4)$产生平行四边形的对角线，$\boldsymbol{w}$ 的反向是$-\boldsymbol{w}$，右侧的线性组合是 $\boldsymbol{v}-\boldsymbol{w}=(5,0)$。

我们先沿着 $\boldsymbol{v}$ 再沿着 $\boldsymbol{w}$ 前进，或者我们沿着 $\boldsymbol{v}+\boldsymbol{w}$ 走对角捷径；我们也可以先沿着 $\boldsymbol{w}$ 再沿着 $\boldsymbol{v}$。换言之，$\boldsymbol{w}+\boldsymbol{v}$ 与 $\boldsymbol{v}+\boldsymbol{w}$ **的答案相同**。沿着平行四边形(本例题是矩形)存在不同的前进方向。

请尊重版权与译者的劳动成果，侵权必究!

三维向量

具有两个分量的向量对应到 xy 平面上的一个点，$\boldsymbol{v}$ 的分量就是点的坐标: $x = v_1$ 与 $y = v_2$。向量从(0, 0)出发，箭头在点(v_1，v_2)结束。我们现在允许向量有三个分量 (v_1, v_2, v_3)。

xy 平面换成三维的 xyz 空间，以下是一些典型向量(依然是列向量，只是有 3 个分量)：

$$\boldsymbol{v}=\begin{bmatrix}1\\1\\-1\end{bmatrix} \quad 与 \quad \boldsymbol{w}=\begin{bmatrix}2\\3\\4\end{bmatrix} \quad 与 \quad \boldsymbol{v}+\boldsymbol{w}=\begin{bmatrix}3\\4\\3\end{bmatrix}$$

向量 $\boldsymbol{v}$ 对应到三维空间的一个箭头，通常箭头由**原点**出发，原点是 xyz 轴的交点，坐标为(0, 0, 0)，箭头终点的坐标是 v_1, v_2, v_3。**列向量，原点出发的箭头与箭头的终点**，这三种表示法之间有着完美的适配。

平面向量(x, y)与 3 维空间的$(x, y, 0)$不相同！

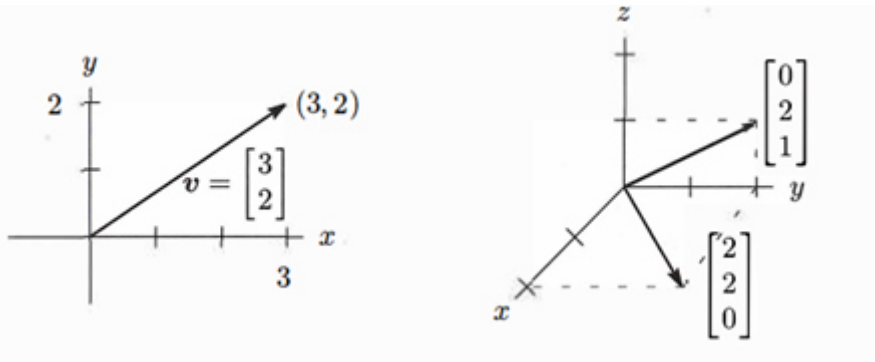


图 1.2：向量$\begin{bmatrix}x\\y\end{bmatrix}$与$\begin{bmatrix}x\\y\\z\end{bmatrix}$对应点$(x, y)$与$(x, y, z)$

从此以后 $\boldsymbol{v}=\begin{bmatrix}1\\1\\-1\end{bmatrix}$ 也可以写成 $\boldsymbol{v}=(1, 1, -1)$

写成**行形式**(在括号中)的理由是为了节省空间，但是 $\boldsymbol{v}=(1, 1, -1)$不是行向量！实际上还是列向量，只是暂时横躺而已。行向量[1　1　−1]是绝对不同的，虽然它有相同的三个分量。1×3 的行向量是 3×1 的列向量 $\boldsymbol{v}$ 的“转置”(transpose)。

三维空间中，$\boldsymbol{v}+\boldsymbol{w}$ 仍然是每次计算一个分量，向量总和的分量是 v_1+w_1 与 v_2+w_2 与 v_3+w_3，你现在了解如何在 4 或 5 或 n 维空间中相加向量。当 $\boldsymbol{w}$ 从 $\boldsymbol{v}$ 的终点出发，第三边是 $\boldsymbol{v}+\boldsymbol{w}$，平行四边形的另一个环绕方向是 $\boldsymbol{w}+\boldsymbol{v}$。问题：这四边是否在同一平面？答案：**是**。向量总和 $\boldsymbol{v}+\boldsymbol{w}-\boldsymbol{v}-\boldsymbol{w}$ 走完一圈产生____向量。

三维空间中典型的三个向量的线性组合是 $\boldsymbol{u}+4\boldsymbol{v}-2\boldsymbol{w}$：

分别用 1, 4, –2 乘再相加的线性组合 $\begin{bmatrix}1\\0\\3\end{bmatrix}+4\begin{bmatrix}1\\2\\1\end{bmatrix}-2\begin{bmatrix}2\\3\\-1\end{bmatrix}=\begin{bmatrix}1\\2\\9\end{bmatrix}$

重要问题

一个向量 $\boldsymbol{u}$，唯一的线性组合是倍数 $c\boldsymbol{u}$。对于两个向量，线性组合是 $c\boldsymbol{u}+d\boldsymbol{v}$。对于三个向量，线性组合是 $c\boldsymbol{u}+d\boldsymbol{v}+e\boldsymbol{w}$。你可以由一个组合跨一大步到**所有的组合**吗？允许每个 c 与 d 与 e，假设 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 是三维空间中的向量：

1. 所有 $c\boldsymbol{u}$ 的组合，图形是什么？
2. 所有 $c\boldsymbol{u}+d\boldsymbol{v}$ 的组合，图形是什么？
3. 所有 $c\boldsymbol{u}+d\boldsymbol{v}+e\boldsymbol{w}$ 的组合，图形是什么？

上述的答案与特定向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 有关，如果他们都是零向量(非常极端案例)，所有的线性组合都是零。如果他们是典型的非零向量(随机选定分量)，这里有三种答案，这是我们主题的关键：

1. 所有 $c\boldsymbol{u}$ 的组合形成一条**通过(0, 0, 0)的直线**。
2. 所有 $c\boldsymbol{u}+d\boldsymbol{v}$ 的组合形成一个**通过(0, 0, 0)的平面**。
3. 所有 $c\boldsymbol{u}+d\boldsymbol{v}+e\boldsymbol{w}$ 的组合形成**三维空间**。

因为 c 可以是 0，零向量(0, 0, 0)会在直线上；当 c 与 d 都是 0，零向量会在平面上。向量 $c\boldsymbol{u}$ 形成的直线是无限长(正向与反向)，我特别要求你去思考全部 $c\boldsymbol{u}+d\boldsymbol{v}$ 形成的平面(三维空间中两个向量的组合)。

一条直线上的所有 $c\boldsymbol{u}$ 加到另一条直线上的所有 $d\boldsymbol{v}$ 会形成图 1.3 的平面。

当我们引入第三向量 $\boldsymbol{w}$ 时，$e\boldsymbol{w}$ 的倍数得到第三条直线。假设第三条直线不在 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 形成的平面上，则 $e\boldsymbol{w}$ 与 $c\boldsymbol{u}+d\boldsymbol{v}$ 的组合可以形成整个三维空间。

这是典型的状况！线，平面，然后空间，但是还有其他可能性存在。当 $\boldsymbol{w}$ 恰好等于 $c\boldsymbol{u}+d\boldsymbol{v}$，第三向量 $\boldsymbol{w}$ 落在前两个向量形成的平面上，$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 的组合无法离开 $\boldsymbol{uv}$ 平面，我们没有得到整个三维空间。请思考问题 1 的特殊例子。

请尊重版权与译者的劳动成果，侵权必究!



图 1.3：(a) 穿过 $\boldsymbol{u}$ 的直线。(b) 包含穿过 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 直线的平面

主要观念复习

1. 二维空间的向量 $\boldsymbol{v}$ 具有两个分量 v_1 与 v_2。
2. $\boldsymbol{v}+\boldsymbol{w}=(v_1+w_1, v_2+w_2)$与 $c\boldsymbol{v}=(cv_1, cv_2)$，每次计算一个分量。
3. 三个向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 的线性组合是 $c\boldsymbol{u}+d\boldsymbol{v}+e\boldsymbol{w}$。
4. 选取所有 $\boldsymbol{u}$ 或是 $\boldsymbol{u}$，$\boldsymbol{v}$ 或是 $\boldsymbol{u}$，$\boldsymbol{v}$，$\boldsymbol{w}$ 的线性组合，在三维空间中，这些组合典型的形成一条直线或是一个平面或是整个空间 $\mathbf{R}^3$。

已解范例

1.1A $\boldsymbol{v}=(1, 1, 0)$与 $\boldsymbol{w}=(0, 1, 1)$的线性组合会形成 $\mathbf{R}^3$ 的一个平面，请描述这个平面，并且找出一个不是 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的线性组合的向量--不在该平面上。

解 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 形成的平面包含所有组合 $c\boldsymbol{v}+d\boldsymbol{w}$，平面上的向量允许任意的 c 与 d。图 1.3 的平面形成两条线之间的区域。

$$\text{组合 } c\boldsymbol{v}+d\boldsymbol{w}=c\begin{bmatrix}1\\1\\0\end{bmatrix}+d\begin{bmatrix}0\\1\\1\end{bmatrix}=\begin{bmatrix}c\\c+d\\d\end{bmatrix} \text{ 形成一个平面。}$$

这个平面的四个向量是(0, 0, 0)与(2, 3, 1)与(5, 7, 2)与$(\pi, 2\pi, \pi)$，第二分量 $c+d$ 永远是第一与第三分量的总和。大部分的向量，例如(1, 2, 3)就不在这个平面上，因为 $2\neq 1+3$。

通过(0, 0, 0)平面的另一种描述方法是知道 $\boldsymbol{n}=(1, -1, 1)$与平面**垂直**，段落 1.2 测试点积来确认 90°角：$\boldsymbol{v}\cdot\boldsymbol{n}=0$ 与 $\boldsymbol{w}\cdot\boldsymbol{n}=0$。垂直向量具有**零点积**。

1.1B $\boldsymbol{v} = (1, 0)$与 $\boldsymbol{w} = (0, 1)$，描述所有 $c\boldsymbol{v}$ 的点，当(1) 所有的数值 c。(2) 非负数值 $c \geq 0$。将 $c\boldsymbol{v}$ 加上所有 $d\boldsymbol{w}$，描述所有的 $c\boldsymbol{v} + d\boldsymbol{w}$。

解

(1) 当 c 是任意数时，向量 $c\boldsymbol{v} = (c, 0)$是沿着 x 轴($\boldsymbol{v}$ 的方向)的等距点，他们包含(–2, 0), (–1, 0), (0, 0), (1, 0), (2, 0)。

(2) 当 $c \geq 0$，向量 $c\boldsymbol{v}$ 形成一条**半线**，就是正 x 轴。这条半线从(0, 0)开始，此时 $c = 0$。它包含了点(100, 0)与(π, 0)，但不包含(–100, 0)。

(1′) 加上所有的向量 $d\boldsymbol{w} = (0, d)$，会在这些等距点 $c\boldsymbol{v}$ 上放置一条铅直(vertical)线，我们得到无限多来自(全部数值 c，任意数值 d)的平行线。

(2′) 加上所有的向量 $d\boldsymbol{w} = (0, d)$，会在半线上的每一个 $c\boldsymbol{v}$ 放置一条铅直线，现在我们有一个**半平面**，xy 平面的右半部分有任意的 $x \geq 0$ 与任意的 y。

1.1C 求出 c 与 d 的两个方程式，使得**线性组合** $c\boldsymbol{v} + d\boldsymbol{w}$ 等于 $\boldsymbol{b}$：

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

解 在应用数学中，很多问题都有两个部分：

1. 建模(modeling)部分：利用一些方程式来表示问题。
2. 计算部分：利用快速又正确的演绎法求解方程组。

在此我们只讨论第一部分(方程组)，第二章会介绍第二部分(求解)。我们的范例适配一个线性代数的基础模型：

$$\text{求 } n \text{ 个数值 } c_1, \ldots, c_n \text{ 使得 } c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n = \boldsymbol{b}$$

当 $n = 2$ 我们可以找到关于 c's 的公式，第二章介绍的“消元法”适用于远超过 $n = 1000$ 的系统。当 n 大于 10 亿时，参考第 11 章。此处 $n = 2$：

向量方程式 $c\boldsymbol{v} + d\boldsymbol{w} = \boldsymbol{b}$ $\quad c\begin{bmatrix} 2 \\ -1 \end{bmatrix} + d\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

c 与 d 所需要的方程组来自个别的两个分量：

两个一般方程式

$$\begin{aligned} 2c - d &= 1 \\ -c + 2d &= 0 \end{aligned}$$

每个方程式产生一条直线，两条直线相交于解 $c = 2/3, d = 1/3$。也可以把这个问题视为**矩阵方程式**，这就是我们想要往下介绍的：

2×2 矩阵 $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

请尊重版权与译者的劳动成果，侵权必究!

问题集 1.1

问题 1-9 有关向量加法与线性组合。

1 描述下列所有的线性组合的几何意义(线，面，或是整个$\mathbf{R}^3$)：

(a) $\begin{bmatrix}1\\2\\3\end{bmatrix}$ 与 $\begin{bmatrix}3\\6\\9\end{bmatrix}$　　(b) $\begin{bmatrix}1\\0\\0\end{bmatrix}$ 与 $\begin{bmatrix}0\\2\\3\end{bmatrix}$　　(c) $\begin{bmatrix}2\\0\\0\end{bmatrix}$ 与 $\begin{bmatrix}0\\2\\2\end{bmatrix}$ 与 $\begin{bmatrix}2\\2\\3\end{bmatrix}$

2 在 xy 平面上画出 $\boldsymbol{v}=\begin{bmatrix}4\\1\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}-2\\2\end{bmatrix}$ 以及 $\boldsymbol{v}+\boldsymbol{w}$ 与 $\boldsymbol{v}-\boldsymbol{w}$。

3 若 $\boldsymbol{v}+\boldsymbol{w}=\begin{bmatrix}5\\1\end{bmatrix}$ 与 $\boldsymbol{v}-\boldsymbol{w}=\begin{bmatrix}1\\5\end{bmatrix}$，计算并画出向量 $\boldsymbol{v}$ 与 $\boldsymbol{w}$。

4 从 $\boldsymbol{v}=\begin{bmatrix}2\\1\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}1\\2\end{bmatrix}$，计算 $3\boldsymbol{v}+\boldsymbol{w}$ 与 $c\boldsymbol{v}+d\boldsymbol{w}$ 的分量。

5 计算 $\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$ 以及 $2\boldsymbol{u}+2\boldsymbol{v}+\boldsymbol{w}$。你如何知道 $\boldsymbol{u}$，$\boldsymbol{v}$，$\boldsymbol{w}$ 在同一个平面上?

因为 $\boldsymbol{w}=c\boldsymbol{u}+d\boldsymbol{v}$
这些直线在同一平面
求 c 与 d

$\boldsymbol{u}=\begin{bmatrix}1\\2\\3\end{bmatrix}$，$\boldsymbol{v}=\begin{bmatrix}-3\\1\\-2\end{bmatrix}$，$\boldsymbol{w}=\begin{bmatrix}2\\-3\\-1\end{bmatrix}$

6 $\boldsymbol{v}=(1,-2,1)$与 $\boldsymbol{w}=(0,1,-1)$的每个组合的分量总和是 _____。求 c 与 d 使得 $c\boldsymbol{u}+d\boldsymbol{w}=(3,3,-6)$。为什么不可能是$(3,3,6)$?

7 在 xy 平面上标示出九个线性组合的点：

$$c\begin{bmatrix}2\\1\end{bmatrix}+d\begin{bmatrix}0\\1\end{bmatrix}，\text{其中 } c=0,1,2 \text{ 与 } d=0,1,2$$

8 图 1.1 的平行四边形的对角线是 $\boldsymbol{v}+\boldsymbol{w}$，另一个对角线为何？这两个对角线向量的总和是多少？画出向量的总和。

9 如果平行四边形的三个角点是$(1,1)$, $(4,2)$与$(1,3)$，则三个可能的第四点为何?画出其中两个点。

问题 10-14 有关图 1.4 的立方体与时钟的特殊向量。

10 立方体上的哪个点是 $\boldsymbol{i}+\boldsymbol{j}$? 哪个点是向量 $\boldsymbol{i}=(1,0,0)$与 $\boldsymbol{j}=(0,1,0)$与 $\boldsymbol{k}=(0,0,1)$的总和? 描述立方体上所有的点(x,y,z)。

11 单位立方体的四个角点是$(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$，其他四个角点为何？求出立方体的中心点坐标。六个面的中心点坐标是_____。立方体有几个边?

12 **复习问题**。在 xyz 空间中，$\boldsymbol{i}=(1,0,0)$与 $\boldsymbol{i}+\boldsymbol{j}=(1,1,0)$的所有线性组合形成的平面为何?

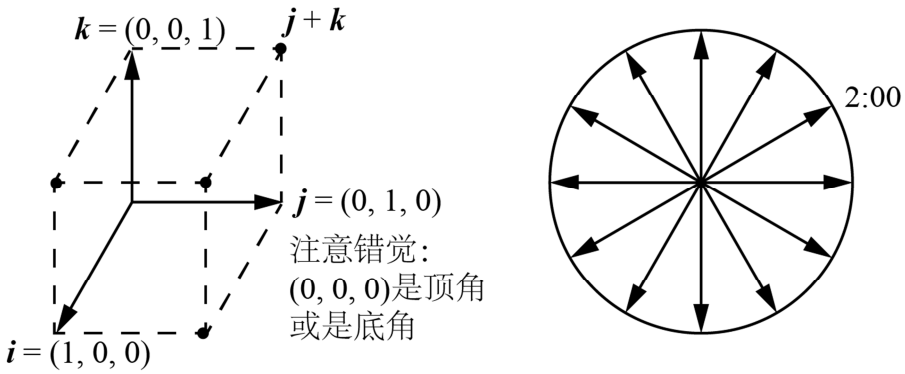


图 1.4：***i, j, k*** 构成的单位立方体与 12 个时钟向量

13 (a) 考虑 12 个由中心点指向时钟 1:00，2:00, ..., 12:00 的向量，其总和 $\boldsymbol{V}$ 为何？

(b) 如果把 2:00 的向量移走，为什么其余 11 个时钟向量的总和指向 8:00?

(c) 在 2:00 的向量 $\boldsymbol{v} = (\cos\theta, \sin\theta)$，求出 x 与 y 的分量。

14 假设 12 个向量是由底部的 6:00 出发而不是由中心的(0, 0)出发，指向 12:00 的向量变成两倍的(0, 2), 则全部 12 个新的时钟向量总和为何?

问题 15-19 深入讨论 *v* 与 *w* 的线性组合(图 1.5a)。

15 图 1.5a 显示 $\boldsymbol{v}/2 + \boldsymbol{w}/2$，请标示点 $3\boldsymbol{v}/4 + \boldsymbol{w}/4$ 与 $\boldsymbol{v}/4 + \boldsymbol{w}/4$ 与 $\boldsymbol{v} + \boldsymbol{w}$。

16 标示点 $-\boldsymbol{v} + 2\boldsymbol{w}$ 以及当 $c + d = 1$ 时所有可能的 $c\boldsymbol{v} + d\boldsymbol{w}$。当 $c + d = 1$ 时，画出所有组合的直线。

17 标示点 $\boldsymbol{v}/3 + \boldsymbol{w}/3$ 与 $2\boldsymbol{v}/3 + 2\boldsymbol{w}/3$。$c\boldsymbol{v} + c\boldsymbol{w}$ 会形成哪条直线?

18 限制 $0 \le c \le 1$ 且 $0 \le d \le 1$, 画出 $c\boldsymbol{v} + d\boldsymbol{w}$ 所有的组合形成的区域。

19 只限制 $c \ge 0$ 且 $d \ge 0$，画出 $c\boldsymbol{v} + d\boldsymbol{w}$ 所有的组合形成的“锥形(cone)”。

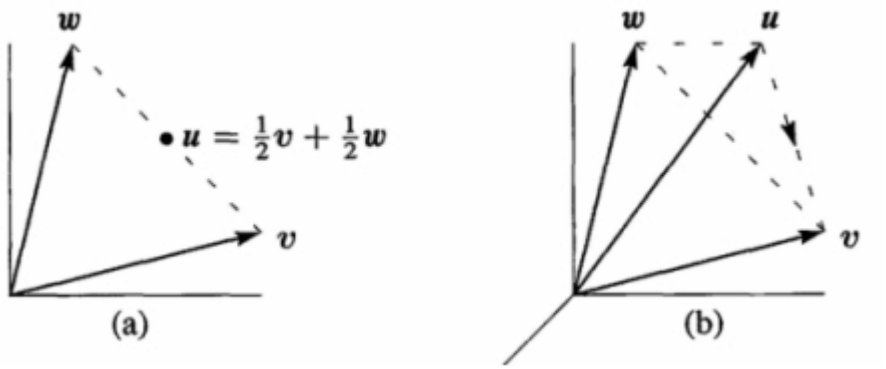


图 1.5：问题 **15-19** 的平面　　问题 **20-25** 的三维空间

请尊重版权与译者的劳动成果，侵权必究!

问题 20-25 处理三维空间向量 u, v, w。(见图 1.5b)。

20 标示图 1.5 中 $u/3 + v/3 + w/3$ 以及 $u/2 + w/2$ 的位置。挑战问题：c, d, e 在什么限制条件下，$cu + dv + ew$ 会形成图上的虚线三角形? 要想留在三角形内，一个要求是 $c \geq 0, d \geq 0, e \geq 0$。

21 虚线三角形的三个边是 $v - u, w - v, u - w$, 他们的总和是_____。画出环绕平面三角形的“头至尾”的加法：$(3, 1) + (-1, 1) + (-2, -2)$。

22 当 $c \geq 0, d \geq 0, e \geq 0$ 且 $c + d + e \leq 1$，画出 $cu + dv + ew$ 覆盖的金字塔(pyramid)区域。确认$(u + v + w)/2$ 是否在金字塔之内？

23 如果考虑 u, v, w 所有的线性组合，有没有向量无法由 $cu + dv + ew$ 得到? 如果 u, v, w 全部位于_____，会有不同的答案。

24 哪些向量是 u 与 v 的线性组合，同时也是 v 与 w 的线性组合?

25 画出向量 u, v, w，使得他们的组合 $cu + dv + ew$ 只形成一条直线。求出向量 u, v, w 使得组合 $cu + dv + ew$ 只形成一个平面。

26 什么样的组合 $c\begin{bmatrix}1\\2\end{bmatrix} + d\begin{bmatrix}3\\1\end{bmatrix}$ 得到 $\begin{bmatrix}14\\8\end{bmatrix}$？将方程组表示成线性组合的系数 c 与 d 的两个方程式。

挑战问题

27 在 4 维空间的立方体会有几个角点? 有几个 3 维的面? 有几个边? 一个典型的角点在(0, 0, 1, 0)，一个典型的边会走到(0, 1, 0, 0)。

28 求出向量 v 与 w，使得 $v + w = (4, 5, 6)$且 $v - w = (2, 5, 8)$。这是一个有____个未知数的方程组? 要有相同数量的方程式才能求出这些数值。

29 三个向量 $u = (1, 3), v = (2, 7), w = (1, 5)$，找出两种不同的线性组合得到 $b = (0, 1)$。稍微变动一下题目：如果我随意选出三个平面上的向量 u, v, w，是不是永远存在两种不同的线性组合可以得到 $b = (0, 1)$?

30 $v = (a, b)$与 $w = (c, d)$的线性组合可以得到一个平面，除非_____。找出四个 4 维空间的向量 u, v, w, z 使得他们的线性组合 $cu + dv + ew + fz$ 产生 4 维空间的所有向量(b_1, b_2, b_3, b_4)。

31 写下 c, d, e 的方程组，使得 $cu + dv + ew = b$。你可以找出满足 b 的 c, d, e 吗?

$$u = \begin{bmatrix}2\\-1\\0\end{bmatrix} \qquad v = \begin{bmatrix}-1\\2\\-1\end{bmatrix} \qquad w = \begin{bmatrix}0\\-1\\2\end{bmatrix} \qquad b = \begin{bmatrix}1\\0\\0\end{bmatrix}$$

1.2 长度与点积

1. $\boldsymbol{v}=\begin{bmatrix}1\\2\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}4\\5\end{bmatrix}$ 的“**点积**”是 $\boldsymbol{v}\cdot\boldsymbol{w}=(1)(4)+(2)(5)=4+10=\mathbf{14}$。

2. $\boldsymbol{v}\cdot\boldsymbol{w}$ 是零，所以 $\boldsymbol{v}=\begin{bmatrix}1\\3\\2\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}4\\-4\\4\end{bmatrix}$ 垂直，$(1)(4)+(3)(-4)+(2)(4)=\mathbf{0}$。

3. $\boldsymbol{v}=\begin{bmatrix}1\\3\\2\end{bmatrix}$ 的长度平方是 $\boldsymbol{v}\cdot\boldsymbol{v}=1+9+4=14$，**长度**是 $\|\boldsymbol{v}\|=\sqrt{14}$。

4. $\boldsymbol{u}=\dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|}=\dfrac{\boldsymbol{v}}{\sqrt{14}}=\dfrac{1}{\sqrt{14}}\begin{bmatrix}1\\3\\2\end{bmatrix}$ 有长度 $\|\boldsymbol{u}\|=1$，检验 $\dfrac{1}{14}+\dfrac{9}{14}+\dfrac{4}{14}=1$。

5. $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 之间的角度 θ 有 $\cos\theta=\dfrac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$。

6. $\begin{bmatrix}1\\0\end{bmatrix}$ 与 $\begin{bmatrix}1\\1\end{bmatrix}$ 之间的角度有 $\cos\theta=\dfrac{1}{(1)(\sqrt{2})}$，则角度 $\theta=45°$。

7. 所有的角都有 $|\cos\theta|\le 1$，所以所有的向量都有 $\boxed{|\boldsymbol{v}\cdot\boldsymbol{w}|\le\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$。

第一个段落回避了向量乘法的问题，我们现在定义 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的点积 (dot product)。这种乘法包含了个别的乘积 v_1w_1 与 v_2w_2，但是没有停在这里，这两个数字相加产生一个数字 $\boldsymbol{v}\cdot\boldsymbol{w}$。

这是几何的段落(向量的长度与两向量之间夹角的余弦)。

两个向量 $\boldsymbol{v}=(v_1,v_2)$ 与 $\boldsymbol{w}=(w_1,w_2)$ 的**点积**或**内积**是数字 $\boldsymbol{v}\cdot\boldsymbol{w}$：

$$\boldsymbol{v}\cdot\boldsymbol{w}=v_1w_1+v_2w_2 \tag{1}$$

范例 1 向量 $\boldsymbol{v}=(4,2)$ 与 $\boldsymbol{w}=(-1,2)$ 有零点积：

点积为零，垂直向量 $\qquad\begin{bmatrix}4\\2\end{bmatrix}\cdot\begin{bmatrix}-1\\2\end{bmatrix}=-4+4=0$

数学里面的 0 一直是特殊的数字。对于点积，它表示这两个向量垂直，夹角 90°。当我们在图 1.1 中作图，我们看到了矩形(不只是任意平行四边形)。垂直向量最明显的例子是沿着 x 轴的 $\boldsymbol{i}=(1,0)$ 与沿着 y 轴的 $\boldsymbol{j}=(0,1)$，再一次点积是 $\boldsymbol{i}\cdot\boldsymbol{j}=0+0=0$，这些向量 $\boldsymbol{i}$ 与 $\boldsymbol{j}$ 形成**直角**。

请尊重版权与译者的劳动成果，侵权必究！

向量 $\boldsymbol{v}=(1,2)$与 $\boldsymbol{w}=(3,1)$的点积是 5，很快的 $\boldsymbol{v}\cdot\boldsymbol{w}$ 就会显示 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 之间的夹角(不是 90°)。请验证 $\boldsymbol{w}\cdot\boldsymbol{v}$ 也是 5。

点积 $\boldsymbol{v}\cdot\boldsymbol{w}$ 与 $\boldsymbol{w}\cdot\boldsymbol{v}$ 相等，无关 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的顺序。

范例 2　放一个重量 4 的东西在点 $x=-1$(零的左边)，放一个重量 2 的东西在点 $x=2$(零的右边)，x 轴会在中心点取得平衡(好像跷跷板)，重量取得平衡是因为点积 $(4)(-1)+(2)(2)=0$。

这个例子是典型的科学工程，重量的向量是$(w_1,w_2)=(4,2)$，距离中心点的距离向量是$(v_1,v_2)=(-1,2)$。重量乘距离 v_1w_1 与 v_2w_2 得到“矩(moments)”，跷跷板的平衡方程式是 $v_1w_1+v_2w_2=0$。

范例 3　点积在经济与商业都会用到，比如我们要买卖 3 个商品，他们的单价分别是(p_1,p_2,p_3)--这是“价格向量”；我们买或卖的数量为(q_1,q_2,q_3)，卖的时候取正号，买的时候取负号。单价 p_1 的商品卖出 q_1 个得到 p_1q_1，全部收入(数量 q 乘价格 p)就是在**三维空间的点积 $\boldsymbol{q}\cdot\boldsymbol{p}$**：

$$\textbf{收入}=(q_1,q_2,q_3)\cdot(p_1,p_2,p_3)=q_1p_1+q_2p_2+q_3p_3=\textbf{点积}$$

零点积表示账目平衡。如果 $\boldsymbol{q}\cdot\boldsymbol{p}=0$，全部销售额等于全部买进额，$\boldsymbol{p}$ 垂直 $\boldsymbol{q}$(在三维空间)。一家超市有几千种货品，货物的维度会非常高。

小注释：电子表格在管理中非常重要，可以计算线性组合与点积，在屏幕上看到的是个矩阵。

重点　对于 $\boldsymbol{v}$ 与 $\boldsymbol{w}$，每个 v_i 乘 w_i，则 $\boldsymbol{v}\cdot\boldsymbol{w}=v_1w_1+\ldots+v_nw_n$。

长度与单位向量

有个重要的案例就是向量**自己与自己**的点积，此时 $\boldsymbol{v}=\boldsymbol{w}$。当向量 $\boldsymbol{v}=(1,2,3)$，自己与自己的点积为 $\boldsymbol{v}\cdot\boldsymbol{v}=\|\boldsymbol{v}\|^2=14$：

$$\textbf{点积 } \boldsymbol{v}\cdot\boldsymbol{v} \textbf{ 是长度的平方}\quad \|\boldsymbol{v}\|^2=\begin{bmatrix}1\\2\\3\end{bmatrix}\cdot\begin{bmatrix}1\\2\\3\end{bmatrix}=1+4+9=14。$$

现在向量之间的角度不是 90°而是 0°，$\boldsymbol{v}$ 与自己不垂直所以点积不是 0。点积 $\boldsymbol{v}\cdot\boldsymbol{v}$ 给出 $\boldsymbol{v}$ 的长度平方。

定义　向量 $\boldsymbol{v}$ 的长度 $\|\boldsymbol{v}\|$ 等于 $\boldsymbol{v}\cdot\boldsymbol{v}$ 的平方根：

$$\textbf{长度}=\|\boldsymbol{v}\|=\sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}=(v_1^{\,2}+v_2^{\,2}+\ldots+v_n^{\,2})^{1/2}$$

在二维的长度是$\sqrt{v_1^2+v_2^2}$，在三维的长度是$\sqrt{v_1^2+v_2^2+v_3^2}$，这样的计算得到 $\boldsymbol{v}=(1, 2, 3)$的长度是$\|\boldsymbol{v}\|=\sqrt{14}$。

此处$\|\boldsymbol{v}\|=\sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$是代表向量的箭头的一般长度。如果分量是 1 与 2，箭头就是图 1.6 所示直角三角形的第三边，毕氏定理 $a^2+b^2=c^2$ 得到三边的关系是 $1^2+2^2=\|\boldsymbol{v}\|^2$。

对于 $\boldsymbol{v}=(1, 2, 3)$的长度来说，我们使用直角三角形的公式两次。位于基底的向量(1, 2, 0)的长度是$\sqrt{5}$，基底向量与直線向上的向量(0, 0, 3)垂直，所以盒子的对角线长度$\|\boldsymbol{v}\|=\sqrt{5+9}=\sqrt{14}$。

四维向量的长度等于$\sqrt{v_1^2+v_2^2+v_3^2+v_4^2}$，于是向量(1, 1, 1, 1)的长度是$\sqrt{1^2+1^2+1^2+1^2}=2$，这是一个四维空间的单位立方体的对角线长度。$n$ 维空间立方体的对角长度是$\sqrt{n}$

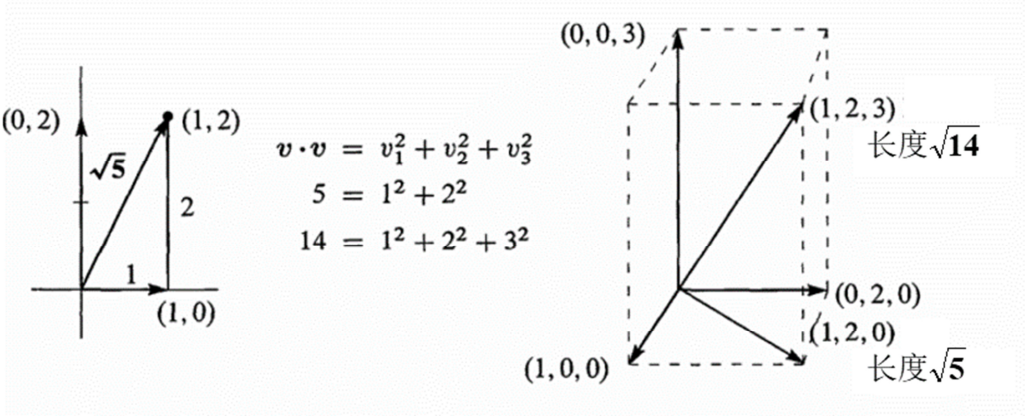


图 1.6：二维与三维向量的长度$\sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$

单位这个词通常用来表示某种测量值等于 1，单位价格是指一个物品的价格，单位立方体的边长为 1，单位圆的半径为 1。现在讨论“单位向量”的意义。

定义　　单位向量 $\boldsymbol{u}$ 是长度为 1 的向量，$\boldsymbol{u}\cdot\boldsymbol{u}=\mathbf{1}$。

举例来说四维的单位向量是 $\boldsymbol{u}=(1/2, 1/2, 1/2, 1/2)$，则 $\boldsymbol{u}\cdot\boldsymbol{u}=1/4+1/4+1/4+1/4=1$。向量 $\boldsymbol{v}=(1, 1, 1, 1)$，除以本身的长度$\|\boldsymbol{v}\|=2$ 得到单位向量。

范例 4　沿着 x 轴与 y 轴的标准单位向量写成 $\boldsymbol{i}$ 与 $\boldsymbol{j}$，在 xy 平面中，单位向量与 x 轴形成夹角 θ，这个单位向量就是$(\cos\theta, \sin\theta)$。

$$\textbf{单位向量}\qquad \boldsymbol{i}=\begin{bmatrix}1\\0\end{bmatrix},\ \boldsymbol{j}=\begin{bmatrix}0\\1\end{bmatrix},\ \boldsymbol{u}=\begin{bmatrix}\cos\theta\\\sin\theta\end{bmatrix}$$

当 $\theta=0$，水平向量 $\boldsymbol{u}$ 就是 $\boldsymbol{i}$；当 $\theta=90°$(或 $\pi/2$ 弧度(radian))，垂直向量就是 $\boldsymbol{j}$。因为 $\cos^2\theta+\sin^2\theta=1$，在任何角度下分量 $\cos\theta$ 与 $\sin\theta$ 得到 $\boldsymbol{u}\cdot\boldsymbol{u}=1$。这些向量往外

请尊重版权与译者的劳动成果，侵权必究!

伸展得到图 1.7 的单位圆，单位圆上角度为 θ 的点坐标是 $\cos\theta$ 与 $\sin\theta$。

由于(2, 2, 1)的长度是 3，向量(2/3, 2/3, 1/3)长度是 1，检验得到 $\boldsymbol{u}\cdot\boldsymbol{u} = 4/9 + 4/9 + 1/9 = 1$。任何非零向量 $\boldsymbol{v}$ 除以本身的长度$\|\boldsymbol{v}\|$就是单位向量。

单位向量 $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$ **是在 $\boldsymbol{v}$ 方向的单位向量。**

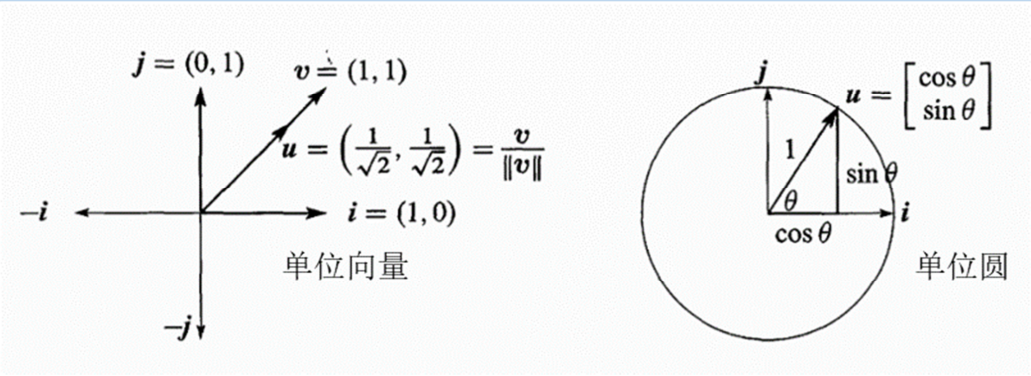


图 1.7：坐标向量 $\boldsymbol{i}$ 与 $\boldsymbol{j}$。位于 45°(左图)的单位向量 $\boldsymbol{u}$ 是 $\boldsymbol{v} = (1, 1)$除以本身长度$\|\boldsymbol{v}\| = \sqrt{2}$。单位向量 $\boldsymbol{u} = (\cos\theta, \sin\theta)$的角度是 θ。

两个向量之间的夹角

我们说垂直向量有 $\boldsymbol{v}\cdot\boldsymbol{w} = 0$，当角度是 90°时点积为 0。为了说明，我们将角度与点积进行关联，展示如何利用 $\boldsymbol{v}\cdot\boldsymbol{w}$ 求出两个非零向量 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的夹角。

直角 **当 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 垂直，点积 $\boldsymbol{v}\cdot\boldsymbol{w} = 0$。**

证明 当 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 垂直，他们形成直角的两个边，第三边为 $\boldsymbol{v} - \boldsymbol{w}$ (斜边如图 1.8)。毕氏定理说明直角三角形的边有 $a^2 + b^2 = c^2$：

$$\text{垂直向量} \qquad \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 \tag{2}$$

以二维的方式写出上述长度的公式，方程式是：

$$\text{毕氏定理} \qquad (v_1^2 + v_2^2) + (w_1^2 + w_2^2) = (v_1 - w_1)^2 + (v_2 - w_2)^2 \tag{3}$$

右侧从 $v_1^2 - 2v_1w_1 + w_1^2$ 开始，两侧都有 v_1^2 与 w_1^2 可以对消，剩下$-2v_1w_1$。同理 v_2^2 与 w_2^2 可以对消，剩下$-2v_2w_2$。(如果是三维向量，还会有$-2v_3w_3$)。现在两边同除–2 得到 $\boldsymbol{v}\cdot\boldsymbol{w} = 0$。 【原文写成 $\boldsymbol{v} - \boldsymbol{w} = 0$ 是错误】

$$0 = -2v_1w_1 - 2v_2w_2 \text{ 得到 } v_1w_1 + v_2w_2 = 0 \tag{4}$$

结论 直角產生 $\boldsymbol{v}\cdot\boldsymbol{w} = 0$。当角度 $\theta = 90°$，点积为 0，此时 $\cos\theta = 0$。因为 $\boldsymbol{0}\cdot\boldsymbol{w}$ 永远为 0，所以零向量 $\boldsymbol{0}$ 与所有向量 $\boldsymbol{w}$ 垂直。

现在假设 $\boldsymbol{v}\cdot\boldsymbol{w}$ 不是零，可以为正也可以为负，$\boldsymbol{v}\cdot\boldsymbol{w}$ 的正负号可以立即知道是小于还是大于直角。当 $\boldsymbol{v}\cdot\boldsymbol{w}$ 为正，角度小于 90°；当 $\boldsymbol{v}\cdot\boldsymbol{w}$ 为负，角度大于 90°。图 1.8 的右侧展示典型向量 $\boldsymbol{v}=(3,1)$，它与 $\boldsymbol{w}=(1,3)$的夹角小于 90°，这是因为 $\boldsymbol{v}\cdot\boldsymbol{w}=6$ 是正数。

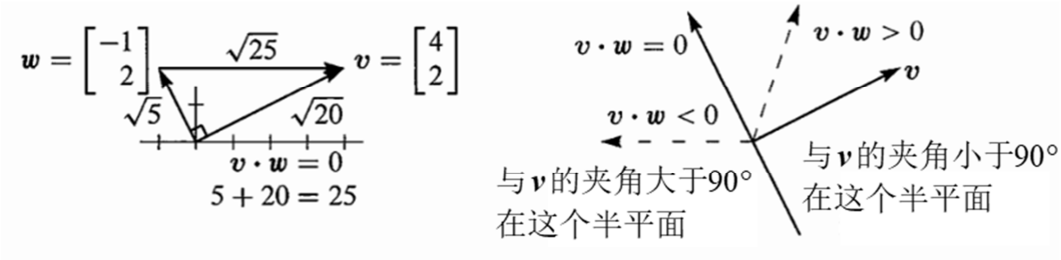


图 1.8：垂直向量有 $\boldsymbol{v}\cdot\boldsymbol{w}=0$，则 $\|\boldsymbol{v}\|^2+\|\boldsymbol{w}\|^2=\|\boldsymbol{v}-\boldsymbol{w}\|^2$

边界线是向量与 $\boldsymbol{v}$ 垂直的位置，(1, –3)位于正负之间的分界线，所以(1, –3)与(3, 1)垂直，点积为零。

点积可以算出真正的角度 θ。对于两个单位向量 $\boldsymbol{u}$ 与 $\boldsymbol{U}$ 来说，$\boldsymbol{u}\cdot\boldsymbol{U}$ 的符号可以决定 $\theta<90°$或是 $\theta>90°$。除此之外，点积 $\boldsymbol{u}\cdot\boldsymbol{U}$ 的值就是 $\cos\theta$。对于 n 维空间而言，前面的观念也是正确的。

两个单位向量 $\boldsymbol{u}$ 与 $\boldsymbol{U}$ 的夹角为 θ，则 $\boldsymbol{u}\cdot\boldsymbol{U}=\cos\theta$，当然 $|\boldsymbol{u}\cdot\boldsymbol{U}|\le 1$。

记得 $\cos\theta$ 不会大于 1，也不会小于–1，单位向量之间的点积会落在–1 与 1 之间，**$\boldsymbol{u}\cdot\boldsymbol{U}$ 的值就是 $\cos\theta$**。

图 1.9 清楚显示 $\boldsymbol{u}=(\cos\theta,\sin\theta)$与 $\boldsymbol{i}=(1,0)$，点积 $\boldsymbol{u}\cdot\boldsymbol{i}=\cos\theta$，这是两个向量夹角的余弦。

旋转任何角度 α 之后，他们仍然是单位向量。向量 $\boldsymbol{i}=(1,0)$旋转至$(\cos\alpha,\sin\alpha)$，向量 $\boldsymbol{u}$ 旋转至$(\cos\beta,\sin\beta)$，其中 $\beta=\alpha+\theta$。他们的点积是 $\cos\alpha\cos\beta+\sin\alpha\sin\beta$，由三角定理得到 $\cos(\beta-\alpha)=\cos\theta$。

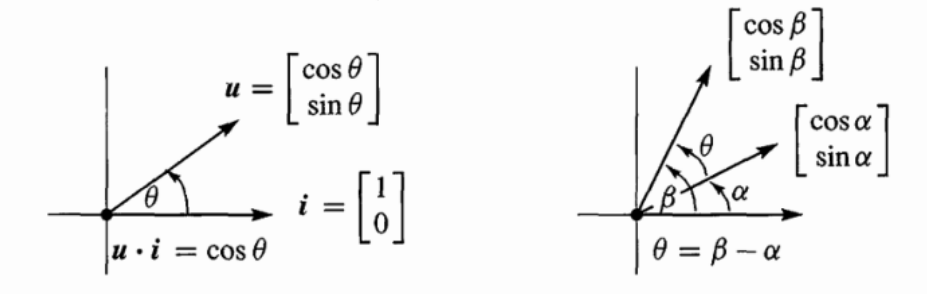


图 1.19：单位向量：$\boldsymbol{u}\cdot\boldsymbol{U}$ 等于 θ(夹角)的余弦

请尊重版权与译者的劳动成果，侵权必究!

如果 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 不是单位向量会怎么样？除以个别的长度得到 $\boldsymbol{u}=\boldsymbol{v}/\|\boldsymbol{v}\|$, $\boldsymbol{U}=\boldsymbol{w}/\|\boldsymbol{w}\|$, 则单位向量的点积得到 $\cos\theta$。

余弦公式 若 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 是非零向量，则 $$\frac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}=\cos\theta \tag{5}$$

无论什么角度，$\boldsymbol{u}=\boldsymbol{v}/\|\boldsymbol{v}\|$与 $\boldsymbol{U}=\boldsymbol{w}/\|\boldsymbol{w}\|$ 的点积不会超过 1，这就是“**苏瓦兹不等式**”：$|\boldsymbol{v}\cdot\boldsymbol{w}|\le\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$——更准确的说法是柯西-苏瓦兹-布尼亚克斯基不等式，分别在法国、德国、俄罗斯发表(也许有其他地方，这是数学上最重要的不等式)。由于$|\cos\theta|$不会超过 1，余弦公式得到两个伟大的不等式：

苏瓦兹不等式 $$|\boldsymbol{v}\cdot\boldsymbol{w}|\le\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$$

三角不等式 $$\|\boldsymbol{v}+\boldsymbol{w}\|\le\|\boldsymbol{v}\|+\|\boldsymbol{w}\|$$

范例 5 对于 $\boldsymbol{v}=\begin{bmatrix}2\\1\end{bmatrix}$ 与 $\boldsymbol{w}=\begin{bmatrix}1\\2\end{bmatrix}$ 求 $\cos\theta$，并且验证两个不等式。

解 点积 $\boldsymbol{v}\cdot\boldsymbol{w}=4$, 两者的长度都是$\sqrt{5}$,余弦是 4/5。

$$\cos\theta=\frac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}=\frac{4}{\sqrt{5}\sqrt{5}}=\frac{4}{5}$$

苏瓦兹不等式得到 $\boldsymbol{v}\cdot\boldsymbol{w}=4$ 小于$\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|=5$。使用三角不等式，第三边长是$\|\boldsymbol{v}+\boldsymbol{w}\|$会小于第一边长加第二边长。针对 $\boldsymbol{v}+\boldsymbol{w}=(3,3)$，三个边长是$\sqrt{18}<\sqrt{5}+\sqrt{5}$，两边平方得到 $18<20$。

范例 6 $\boldsymbol{v}=(a,b)$与 $\boldsymbol{w}=(b,a)$的点积是 $2ab$，两者的长度都是$\sqrt{a^2+b^2}$，苏瓦兹不等式 $\boldsymbol{v}\cdot\boldsymbol{w}\le\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$得到 $\boldsymbol{2ab\le a^2+b^2}$。

如果写成 $x=a^2$ 与 $y=b^2$, 会得到更著名的结果。“几何平均值”$\sqrt{xy}$ 不大于“算术(arithmetic)平均值”$=(x+y)/2$。

几何平均值 ≤ 算术平均值 $ab\le\dfrac{a^2+b^2}{2}$ 变成 $\sqrt{xy}\le\dfrac{x+y}{2}$

范例 5 的 $a=2$ 与 $b=1$, 所以 $x=4$ 与 $y=1$, 几何平均值$\sqrt{xy}=2$ 小于算术平均值$(1+4)/2=2.5$。

计算上的注解

MATLAB, Python, Julia 可以直接进行整个向量的计算，不用透过向量的分量。当 $\boldsymbol{v}$, $\boldsymbol{w}$ 定义完成后就可以直接得到 $\boldsymbol{v}+\boldsymbol{w}$。以行的方式输入 $\boldsymbol{v}$, $\boldsymbol{w}$, 利用符号 prime ′ 可以转置成列向量。$2\boldsymbol{v}+3\boldsymbol{w}$ 写成 2*$\boldsymbol{v}$ + 3*$\boldsymbol{w}$。除非结尾输入半分号符号“;”，否则结果会马上显示出来。

MATLAB　$v = [2 \quad 3 \quad 4]'$；　$w = [1 \quad 1 \quad 1]'$；　$u = 2*v + 3*w$

点积 $v \cdot w$ 是**行向量乘列向量**(使用*而不是 ·)。

点积通常写成 $\begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$　或　**v'*w**　而不是 $\begin{bmatrix}1\\2\end{bmatrix}\cdot\begin{bmatrix}3\\4\end{bmatrix}$。

在 MATLAB 中，v 的长度写成 norm(v)，也就是 sqrt(v' * v)，然后利用点积 v' * w 求出余弦，再求出对应此余弦的角(单位是弧度 radian)。

余弦公式	cosine = v' * w / (norm(v) * norm(w))
反余弦	angle = acos(cosine)

M-档案可以建立新的函数 cosine(v, w)。Python 与 Julia 都是开放源(open source)。

主要观念的复习

1. 点积 $v \cdot w$ 的计算：所有的 $v_i w_i$ 先乘好再全部相加。
2. 长度$\|v\|$是 $v \cdot v$ 的平方根，$u = v / \|v\|$是单位向量，长度等于 1。
3. 当 v 与 w 垂直，点积 $v \cdot w = 0$。
4. θ(任意两个非零向量 v 与 w 的夹角)的余弦不超过 1：

余弦 $\cos\theta = \dfrac{v \cdot w}{\|v\|\|w\|}$，　　苏瓦兹不等式　$|v \cdot w| \le \|v\|\|w\|$

已解范例

1.2A　向量 $v = (3, 4)$与 $w = (4, 3)$，针对 $v \cdot w$ 验证苏瓦兹不等式，以及针对$\|v + w\|$验证三角不等式。求出 v 与 w 之间角度的 cosθ。什么样的 v 与 w 得到等式 $|v \cdot w| = \|v\|\,\|w\|$ 与 $\|v + w\| = \|v\| + \|w\|$？

解　点积是 $v \cdot w = (3)(4) + (4)(3) = 24$，$v$ 的长度$\|v\| = \sqrt{9 + 16} = 5$，$\|w\|$也是 5。总和 $v + w = (7, 7)$的长度是 $7\sqrt{2} < 10$。

苏瓦兹不等式	$\|v \cdot w\| \le \|v\|\,\|w\|$　得到 $24 < 25$
三角不等式	$\|v + w\| \le \|v\| + \|w\|$　得到 $7\sqrt{2} < 5 + 5$
角度的余弦	cos $\theta = 24/25$，这是 $v = (3, 4)$到 $w = (4, 3)$的夹角

等式：一个向量是另一个向量的倍数，如同 $w = cv$，角度是 0°或是 180°。在此情形下，$|\cos\theta| = 1$ 且 $|v \cdot w| = \|v\|\,\|w\|$。如果角度是 0°，比如说 $w = 2v$，则$\|v + w\| = \|v\| + \|w\|$ (两侧都是 $3\|v\|$)，三边是 $v, 2v, 3v$ 的三角形是扁平的!

请尊重版权与译者的劳动成果，侵权必究!

1.2B 求出在 $\boldsymbol{v} = (3, 4)$方向的单位向量 $\boldsymbol{u}$。求出垂直于 $\boldsymbol{u}$ 的单位向量 $\boldsymbol{U}$，$\boldsymbol{U}$ 有多少种可能性?

解 将 $\boldsymbol{v}$ 除以本身的长度$\|\boldsymbol{v}\| = 5$ 得到单位向量 $\boldsymbol{u}$，选择垂直向量 $\boldsymbol{V} = (-4, 3)$，这是因为$(3)(-4) + (4)(3) = 0$。将 $\boldsymbol{V}$ 除以长度$\|\boldsymbol{V}\|$得到与 $\boldsymbol{u}$ 垂直的单位向量：

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \left(\frac{3}{5}, \frac{4}{5}\right) \qquad \boldsymbol{U} = \frac{\boldsymbol{V}}{\|\boldsymbol{V}\|} = \left(\frac{-4}{5}, \frac{3}{5}\right) \qquad \boldsymbol{u} \cdot \boldsymbol{U} = 0$$

另外还有一个垂直单位向量是$-\boldsymbol{U} = (4/5, -3/5)$。

1.2C 给定两个向量 $\boldsymbol{r} = (2, -1)$与 $\boldsymbol{s} = (-1, 2)$，求出向量 $\boldsymbol{x} = (c, d)$使得内积 $\boldsymbol{x} \cdot \boldsymbol{r} = 1$ 以及 $\boldsymbol{x} \cdot \boldsymbol{s} = 0$，

解 两个点积得到 c 与 d 的线性方程组，现在 $\boldsymbol{x} = (c, d)$

$\boldsymbol{x} \cdot \boldsymbol{r} = 1$ 则 $2c - d = 1$ **已解范例 1.1C**

$\boldsymbol{x} \cdot \boldsymbol{s} = 0$ 则 $-c + 2d = 0$ **的相同方程组**

关于 n 维空间的 $\boldsymbol{x} = (x_1, x_2, ..., x_n)$的 n 个方程式的注解

段落 1.1 从列 $\boldsymbol{v}_j$ 开始，目标是建立 $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + ... + x_n\boldsymbol{v}_n = \boldsymbol{b}$。本段落从行 $\boldsymbol{r}_i$ 开始，现在的目标是找到 $\boldsymbol{x}$ 使得 $\boldsymbol{x} \cdot \boldsymbol{r}_i = \boldsymbol{b}_i$。

很快的 $\boldsymbol{v}$'s 就是矩阵 A 的列，$\boldsymbol{r}$'s 就是 A 的行，(唯一的)问题就是求解 $A\boldsymbol{x} = \boldsymbol{b}$。

问题集 1.2

1 计算点积 $\boldsymbol{u} \cdot \boldsymbol{v}$，$\boldsymbol{u} \cdot \boldsymbol{w}$，$\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w})$以及 $\boldsymbol{w} \cdot \boldsymbol{v}$：

$$\boldsymbol{u} = \begin{bmatrix} -.6 \\ .8 \end{bmatrix} \quad \boldsymbol{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad \boldsymbol{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

2 计算问题 1 的向量长度$\|\boldsymbol{u}\|$与$\|\boldsymbol{v}\|$与$\|\boldsymbol{w}\|$，验证苏瓦兹不等式 $|\boldsymbol{u} \cdot \boldsymbol{v}| \le \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$以及 $|\boldsymbol{v} \cdot \boldsymbol{w}| \le \|\boldsymbol{v}\| \, \|\boldsymbol{w}\|$。

3 求出问题 1 在 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 方向的单位向量，求出夹角的余弦。选择向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 使得这些向量与 $\boldsymbol{w}$ 的夹角分别是 0°，90°，180°。

4 对于任意单位向量 $\boldsymbol{v}$ 与 $\boldsymbol{w}$，求出点积 (实际数字)

(a) $\boldsymbol{v}$ 与 $-\boldsymbol{v}$ (b) $\boldsymbol{v} + \boldsymbol{w}$ 与 $\boldsymbol{v} - \boldsymbol{w}$ (c) $\boldsymbol{v} - 2\boldsymbol{w}$ 与 $\boldsymbol{v} + 2\boldsymbol{w}$

5 分别求出在向量 $\boldsymbol{v} = (1, 3)$与 $\boldsymbol{w} = (2, 1, 2)$方向的单位向量 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$。再分别求出垂直 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$ 的单位向量 $\boldsymbol{U}_1$ 与 $\boldsymbol{U}_2$。

6 (a) 描述与向量 $\boldsymbol{v} = (2, -1)$垂直的每个向量 $\boldsymbol{w} = (w_1, w_2)$。

(b) 所有与向量 $\boldsymbol{V} = (1, 1, 1)$垂直的向量落在三维空间的一个_____。

(c) 同时与向量$(1, 1, 1)$与$(1, 2, 3)$垂直的向量落在一个_____。

7 求出两个向量之间夹角θ(来自它的余弦)

(a) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ $\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (b) $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ $\boldsymbol{w} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$

(c) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ $\boldsymbol{w} = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$ (d) $\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ $\boldsymbol{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$

8 是非题(如果正确，举出理由；如果错误，举出反例)：

(a) 如果 $\boldsymbol{u} = (1, 1, 1)$垂直 $\boldsymbol{v}$ 也垂直 $\boldsymbol{w}$，则 $\boldsymbol{v}$ 平行 $\boldsymbol{w}$。

(b) 如果 $\boldsymbol{u}$ 垂直 $\boldsymbol{v}$ 也垂直 $\boldsymbol{w}$，则 $\boldsymbol{u}$ 垂直 $\boldsymbol{v} + 2\boldsymbol{w}$。

(c) 如果 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 是互相垂直的单位向量，则$\|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{2}$？*Yes!*

9 从$(0, 0)$出发到(v_1, v_2)与(w_1, w_2)的斜率分别是 v_2/v_1 与 w_2/w_1，假设斜率的乘积 v_2w_2/v_1w_1 是-1，证明 $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ 且两个向量垂直。(直线 $y = 4x$ 与直线 $y = -x/4$ 垂直。)

10 分别画出从$(0, 0)$出发到点 $\boldsymbol{v} = (1, 2)$与点 $\boldsymbol{w} = (-2, 1)$的箭头，求出两个斜率的乘积。答案会是 $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ 的信号，而且两个箭头_____。

11 若 $\boldsymbol{v} \cdot \boldsymbol{w}$ 为负数，两者的夹角可以得到什么结论？画出一个三维向量 $\boldsymbol{v}$(箭头)，然后说明如何找到所有的 $\boldsymbol{w}$'s 使得 $\boldsymbol{v} \cdot \boldsymbol{w} < 0$。

12 若 $\boldsymbol{v} = (1, 1)$与 $\boldsymbol{w} = (1, 5)$，求出 c 使得 $\boldsymbol{w} - c\boldsymbol{v}$ 垂直于 $\boldsymbol{v}$。若 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 是任意的非零向量，找出 c 的公式。

13 互相垂直的非零向量 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 都与$(1, 0, 1)$垂直，求出 $\boldsymbol{v}$ 与 $\boldsymbol{w}$。

14 彼此垂直的非零向量 $\boldsymbol{u}, \boldsymbol{v}$ 与 $\boldsymbol{w}$，他们都与$(1, 1, 1, 1)$垂直，求出 $\boldsymbol{u}, \boldsymbol{v}$ 与 $\boldsymbol{w}$。

15 $x = 2$ 与 $y = 8$ 的几何平均值是$\sqrt{xy} = 4$，算术平均值会稍大：$(x + y)/2 =$ _____。这是来自范例 6 中有关 $\boldsymbol{v} = (\sqrt{2}, \sqrt{8})$与 $\boldsymbol{w} = (\sqrt{8}, \sqrt{2})$的苏瓦兹不等式，求 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的 $\cos\theta$ 。

16 9 维空间中的向量 $\boldsymbol{v} = (1, 1, ..., 1)$的长度为何？求 $\boldsymbol{v}$ 方向的单位向量 $\boldsymbol{u}$，再求一个单位向量 $\boldsymbol{w}$ 垂直 $\boldsymbol{v}$。

17 向量$(1, 0, -1)$与坐标轴单位向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 的夹角分别是α, β, θ，求出三个角的余弦值。验证公式 $\cos^2\alpha + \cos^2\beta + \cos^2\theta = 1$。

请尊重版权与译者的劳动成果，侵权必究!

问题 18-28 有关长度以及三角形的角的重要事实。

18 $\boldsymbol{v}=(4,2)$与 $\boldsymbol{w}=(-1,2)$组成的平行四边形是个矩形，验证毕氏定理 $a^2+b^2=c^2$ 只有在直角三角形成立：$(\boldsymbol{v}$ 的长度$)^2+(\boldsymbol{w}$ 的长度$)^2=(\boldsymbol{v}+\boldsymbol{w}$ 的长度$)^2$

19 (点积的规则) 下列方程式简单又好用：

(1) $\boldsymbol{v}\cdot\boldsymbol{w}=\boldsymbol{w}\cdot\boldsymbol{v}$ (2) $\boldsymbol{u}\cdot(\boldsymbol{v}+\boldsymbol{w})=\boldsymbol{u}\cdot\boldsymbol{v}+\boldsymbol{u}\cdot\boldsymbol{w}$ (3) $(c\boldsymbol{v})\cdot\boldsymbol{w}=c(\boldsymbol{v}\cdot\boldsymbol{w})$

利用(2)，当 $\boldsymbol{u}=\boldsymbol{v}+\boldsymbol{w}$，证明 $\|\boldsymbol{v}+\boldsymbol{w}\|^2=\boldsymbol{v}\cdot\boldsymbol{v}+2\boldsymbol{v}\cdot\boldsymbol{w}+\boldsymbol{w}\cdot\boldsymbol{w}$

20 余弦法则来自：$(\boldsymbol{v}-\boldsymbol{w})\cdot(\boldsymbol{v}-\boldsymbol{w})=\boldsymbol{v}\cdot\boldsymbol{v}-2\boldsymbol{v}\cdot\boldsymbol{w}+\boldsymbol{w}\cdot\boldsymbol{w}$：

余弦法则 $\|\boldsymbol{v}-\boldsymbol{w}\|^2=\|\boldsymbol{v}\|^2-2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta+\|\boldsymbol{w}\|^2$

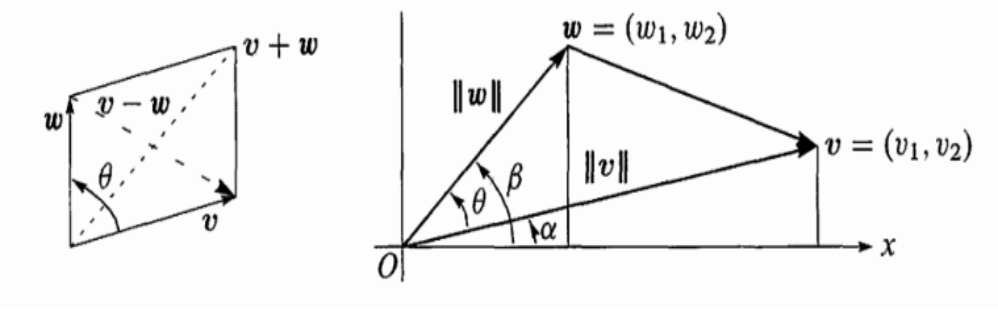
画出三边为 $\boldsymbol{v},\boldsymbol{w},\boldsymbol{v}-\boldsymbol{w}$ 的三角形，哪一个角是θ?

21 三角不等式说明：$(\boldsymbol{v}+\boldsymbol{w}$ 的长度$)\le(\boldsymbol{v}$ 的长度$)+(\boldsymbol{w}$ 的长度$)$，

问题 19 指出 $\|\boldsymbol{v}+\boldsymbol{w}\|^2=\|\boldsymbol{v}\|^2+2\boldsymbol{v}\cdot\boldsymbol{w}+\|\boldsymbol{w}\|^2$，将 $\boldsymbol{v}\cdot\boldsymbol{w}$ 变大成为$\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$，证明

$\|$边 3$\|$ 不能超过 $\|$边 1$\|+\|$边 2$\|$。

三角不等式 $\|\boldsymbol{v}+\boldsymbol{w}\|^2\le(\|\boldsymbol{v}\|+\|\boldsymbol{w}\|)^2$ 或者 $\|\boldsymbol{v}+\boldsymbol{w}\|\le\|\boldsymbol{v}\|+\|\boldsymbol{w}\|$



22 苏瓦兹不等式 $|\boldsymbol{v}\cdot\boldsymbol{w}|\le\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$，从代数的角度(不从三角学)来看：

(a) 两侧同乘本身得到 $(v_1w_1+v_2w_2)^2\le(v_1^2+v_2^2)(w_1^2+w_2^2)$。

(b) 证明上式左右两侧的差(difference)为$(v_1w_2-v_2w_1)^2$。

因为平方不能为负--所以不等式成立。

23 图形显示 $\cos\alpha=v_1/\|\boldsymbol{v}\|$，$\sin\alpha=v_2/\|\boldsymbol{v}\|$，同理 $\cos\beta=$_____，$\sin\beta=$_____。

角度$\theta=\beta-\alpha$，代入三角公式 $\cos\beta\cos\alpha+\sin\beta\sin\alpha=\cos(\beta-\alpha)$，可以得到 $\cos\theta=\boldsymbol{v}\cdot\boldsymbol{w}/\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$。

24 对于单位向量(u_1, u_2)与(U_1, U_2)，利用一个式子证明$|\boldsymbol{u} \cdot \boldsymbol{U}| \le 1$：

$$|\boldsymbol{u} \cdot \boldsymbol{v}| \le |u_1||U_1| + |u_2||U_2| \le \frac{{u_1}^2 + {U_1}^2}{2} + \frac{{u_2}^2 + {U_2}^2}{2} = 1$$

将$(u_1, u_2) = (0.6, 0.8)$，$(U_1, U_2) = (0.8, 0.6)$放在整条直线上，求 $\cos\theta$。

25 为什么一开始就说$|\cos\theta|$不会大于 1?

26 (推荐) 画出一个平行四边形。

27 平行四边形的两个边是 $\boldsymbol{v}$ 与 $\boldsymbol{w}$，证明两条对角线的平方和会等于四个边的平方和：$\|\boldsymbol{v}+\boldsymbol{w}\|^2 + \|\boldsymbol{v}-\boldsymbol{w}\|^2 = 2\|\boldsymbol{v}\|^2 + 2\|\boldsymbol{w}\|^2$。

28 若 $v = (1, 2)$，在 xy 平面画出使得 $\boldsymbol{v} \cdot \boldsymbol{w} = x + 2y = 5$ 的所有向量 $\boldsymbol{w} = (x, y)$，为什么答案会是一条直线？最短的 $\boldsymbol{w}$ 为何？

29 (推荐) 若$\|\boldsymbol{v}\| = 5, \|\boldsymbol{w}\| = 3$，则$\|\boldsymbol{v} - \boldsymbol{w}\|$的最小值与最大值为何？$\boldsymbol{v} \cdot \boldsymbol{w}$ 的最小值与最大值为何？

挑战问题

30 在 xy 平面会不会存在三个向量，使得 $\boldsymbol{u} \cdot \boldsymbol{v} < 0, \boldsymbol{v} \cdot \boldsymbol{w} < 0, \boldsymbol{u} \cdot \boldsymbol{w} < 0$？我不知道在 xyz 平面有多少这样的向量，使得所有的点积是负数。(在平面绝对不存在 4 个向量有这样的性质。)

31 任意选取数值使得 $x + y + z = 0$，求出向量 $\boldsymbol{v} = (x, y, z)$与 $\boldsymbol{w} = (z, x, y)$的夹角。挑战问题：说明为什么 $\boldsymbol{v} \cdot \boldsymbol{w} / \|\boldsymbol{v}\| \|\boldsymbol{w}\|$ 永远是–1/2。

32 你如何证明 $\sqrt[3]{xyz} \le \dfrac{x+y+z}{3}$ (几何平均值 ≤ 算术平均值)?

33 从向量$\left(\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}\right)$中，利用正负号选取 4 个互相垂直的单位向量。

34 利用 MATLAB 的 $\boldsymbol{v}$ = randn(3, 1)指令，制造一个随机的单位向量 $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$；利用 $\boldsymbol{V}$ = randn(3, 30)指令，制造 30 个以上的单位向量 $\boldsymbol{U}_j$。点积 $|\boldsymbol{u} \cdot \boldsymbol{U}_j|$ 的平均大小是多少？在微积分中，平均值是$\int_0^\pi |\cos\theta| \, d\theta / \pi = \dfrac{2}{\pi}$。

请尊重版权与译者的劳动成果，侵权必究!

1.3 矩阵

1 $A=\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}$ 是 3×2 矩阵：$m=3$ 行与 $n=2$ 列。

2 $A\boldsymbol{x}=\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$ 是**列的组合**，$A\boldsymbol{x}=x_1\begin{bmatrix}1\\3\\5\end{bmatrix}+x_2\begin{bmatrix}2\\4\\6\end{bmatrix}$。

3 $A\boldsymbol{x}$ 的 3 个分量是 A 的 3 个行与向量 $\boldsymbol{x}$ 的点积：

每次处理一行 $\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}\begin{bmatrix}7\\8\end{bmatrix}=\begin{bmatrix}1\cdot 7+2\cdot 8\\3\cdot 7+4\cdot 8\\5\cdot 7+6\cdot 8\end{bmatrix}=\begin{bmatrix}23\\53\\83\end{bmatrix}$

4 矩阵形式的方程式 $A\boldsymbol{x}=\boldsymbol{b}$：使用 $\begin{bmatrix}2&5\\3&7\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\end{bmatrix}$ 取代 $\begin{matrix}2x_1+5x_2=b_1\\3x_1+7x_2=b_2\end{matrix}$。

5 $A\boldsymbol{x}=\boldsymbol{b}$ 的解可以写成 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$，但是有些矩阵不允许 A^{-1}。

本段落从三个向量 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ 开始，我会使用矩阵来结合他们。

三个向量 $\boldsymbol{u}=\begin{bmatrix}1\\-1\\0\end{bmatrix}$ $\boldsymbol{v}=\begin{bmatrix}0\\1\\-1\end{bmatrix}$ $\boldsymbol{w}=\begin{bmatrix}0\\0\\1\end{bmatrix}$

在三维空间他们的线性组合是 $x_1\boldsymbol{u}+x_2\boldsymbol{v}+x_3\boldsymbol{w}$：

向量的组合
$$x_1\begin{bmatrix}1\\-1\\0\end{bmatrix}+x_2\begin{bmatrix}0\\1\\-1\end{bmatrix}+x_3\begin{bmatrix}0\\0\\1\end{bmatrix}=\begin{bmatrix}x_1\\x_2-x_1\\x_3-x_2\end{bmatrix}\tag{1}$$

重要事项：利用矩阵改写组合，$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ 变成矩阵 A 的列，矩阵乘向量(x_1,x_2,x_3)：

矩阵乘向量，列的组合
$$A\boldsymbol{x}=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}x_1\\x_2-x_1\\x_3-x_2\end{bmatrix}\tag{2}$$

数值 x_1,x_2,x_3 是向量 $\boldsymbol{x}$ 的分量，矩阵 A 乘向量 $\boldsymbol{x}$ 等同于方程式(1)的三个列的组合 $x_1\boldsymbol{u}+x_2\boldsymbol{v}+x_3\boldsymbol{w}$。

这不只是 $A\boldsymbol{x}$ 的定义，因为改写带来了重要的不同观点，刚开始是三个数字 x_1,x_2,x_3 是乘向量，现在是矩阵乘这些数字。矩阵 A 作用在向量 $\boldsymbol{x}$，输出 $A\boldsymbol{x}$ 是 A 的

列的组合 $\boldsymbol{b}$。

为了观察作用，我把 b_1, b_2, b_3 写成 $A\boldsymbol{x}$ 的分量：

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \boldsymbol{b} \tag{3}$$

输入是 $\boldsymbol{x}$，输出是 $\boldsymbol{b} = A\boldsymbol{x}$。这个 A 是一个差分矩阵(difference matrix)，因为 $\boldsymbol{b}$ 包含输入 $\boldsymbol{x}$ 的差，最上方的差是 $x_1 - x_0 = x_1 - 0$。

这是一个展示 $\boldsymbol{x} = (1, 4, 9)$的差的范例：平方数在 $\boldsymbol{x}$ 中，奇数在 $\boldsymbol{b}$ 中。

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 4 \\ 9 \end{bmatrix} = \text{平方数} \qquad A\boldsymbol{x} = \begin{bmatrix} 1-0 \\ 4-1 \\ 9-4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} = \boldsymbol{b} \tag{4}$$

这个模式可以延伸到 4×4 矩阵，下一个平方数是 16，下一个差是 $x_4 - x_3 = 16 - 9 = 7$(下个奇数)。这个矩阵同时将所有的差 1, 3, 5, 7 全部计算出来。

重要注解： 每次乘一个行。你已经研读矩阵与向量的乘法 $A\boldsymbol{x}$，也许可以换一种方式来解释，我们使用行而不使用列。一般情形是计算每一行与 $\boldsymbol{x}$ 的点积：

$A\boldsymbol{x}$同样也是行的点积

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} (1,\ 0,\ 0)\cdot(x_1,\ x_2,\ x_3) \\ (-1,\ 1,\ 0)\cdot(x_1,\ x_2,\ x_3) \\ (0,\ -1,\ 1)\cdot(x_1,\ x_2,\ x_3) \end{bmatrix} \tag{5}$$

这些点积与方程式(3)的结果 $x_1, x_2 - x_1, x_3 - x_2$ 完全相同。新的方法是每次处理 $A\boldsymbol{x}$ 的一个列，线性组合是线性代数的关键，输出 $A\boldsymbol{x}$ 是 $\boldsymbol{A}$ **的列**的线性组合。

如果是数字，你可以用行来乘 $A\boldsymbol{x}$；如果是文字，用列处理比较好。第二章会再重复这些矩阵乘法的规则，并且解释这些概念。

线性方程组

观念的改变至关重要。目前为止数字 x_1, x_2, x_3 是已知，右侧的 $\boldsymbol{b}$ 未知，我们利用 A 乘 $\boldsymbol{x}$ 得到差的向量。**现在我们考虑 $\boldsymbol{b}$ 已知，要求解 $\boldsymbol{x}$。**

老问题：计算线性组合 $x_1\boldsymbol{u} + x_2\boldsymbol{v} + x_3\boldsymbol{w}$ 求出 $\boldsymbol{b}$。

新问题：什么样的 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 的线性组合产生特定的 $\boldsymbol{b}$?

这是一个逆反问题--找出输入 $\boldsymbol{x}$ 得到想要的输出 $\boldsymbol{b} = A\boldsymbol{x}$。你已经见过这个系统，这是 x_1, x_2, x_3 的线性方程组。方程式右侧是 b_1, b_2, b_3，我现在要求解 $A\boldsymbol{x} = \boldsymbol{b}$ 找出 x_1, x_2, x_3：

请尊重版权与译者的劳动成果，侵权必究!

$$\text{方程式 } A\boldsymbol{x}=\boldsymbol{b} \quad \begin{matrix} x_1 & & & = b_1 \\ -x_1 & + x_2 & & = b_2 \\ & -x_2 & + x_3 & = b_3 \end{matrix} \quad \text{解 } \boldsymbol{x}=A^{-1}\boldsymbol{b} \quad \begin{matrix} x_1 = b_1 \\ x_2 = b_1 + b_2 \\ x_3 = b_1 + b_2 + b_3 \end{matrix} \qquad (6)$$

我现在立刻承认---大部分的线性系统都不容易求解。这个范例中，第一个方程式决定 $x_1 = b_1$，第二个方程式产生 $x_2 = b_1 + b_2$。因为 A 是三角矩阵，这些方程式可以有序的求出解答(顶端至底部)。

检视两个特殊案例，右侧的 b_1, b_2, b_3 设成 0, 0, 0 与 1, 3, 5：

$$\boldsymbol{b}=\begin{bmatrix}0\\0\\0\end{bmatrix} \text{得到 } \boldsymbol{x}=\begin{bmatrix}0\\0\\0\end{bmatrix}, \quad \boldsymbol{b}=\begin{bmatrix}1\\3\\5\end{bmatrix} \text{得到 } \boldsymbol{x}=\begin{bmatrix}1\\1+3\\1+3+5\end{bmatrix}=\begin{bmatrix}1\\4\\9\end{bmatrix}$$

第一个解(全部是 0)比看起来更重要。口头说法：如果输出 $\boldsymbol{b}=\boldsymbol{0}$，输入必须 $\boldsymbol{x}=\boldsymbol{0}$。对 A 来说，理论是成立的，但是对所有的矩阵而言却是不一定。第二个例子会说明(不同的矩阵 C)，当 $C\neq 0$ 且 $x\neq \boldsymbol{0}$ 时，如何才能得到 $C\boldsymbol{x}=\boldsymbol{0}$。

这个矩阵 A 是可逆，从 $\boldsymbol{b}$ 可以反推得到 $\boldsymbol{x}$，写成 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$。

逆矩阵

让我重复一次方程式(6)的解 $\boldsymbol{x}$，会出现一个总和矩阵！

$$\text{求解 } A\boldsymbol{x}=\boldsymbol{b} \qquad \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_1+b_2\\b_1+b_2+b_3\end{bmatrix}=\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad (7)$$

如果 $\boldsymbol{x}$ 之间的差是 $\boldsymbol{b}$，那么 $\boldsymbol{b}$ 之间的和就是 $\boldsymbol{x}$。对于奇数 $\boldsymbol{b}=(1, 3, 5)$与平方数 $\boldsymbol{x}=(1, 4, 9)$来说，这是正确的，对于所有的向量都是正确的。方程式(7)的**总和矩阵就是差分矩阵 A 的逆矩阵 A^{-1}**。

范例：$\boldsymbol{x}=(1, 2, 3)$的差是 $\boldsymbol{b}=(1, 1, 1)$，所以 $\boldsymbol{b}=A\boldsymbol{x}$ 且 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$:

$$A\boldsymbol{x}=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix}=\begin{bmatrix}1\\1\\1\end{bmatrix} \qquad A^{-1}\boldsymbol{b}=\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix}$$

方程式(7)的解 $\boldsymbol{x}=(x_1, x_2, x_3)$告诉我们两件重要的事实：

1. 对每一个 $\boldsymbol{b}$，存在一个 $A\boldsymbol{x}=\boldsymbol{b}$ 的解。 **2.** 矩阵 A^{-1} 得到 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$。

下一章会询问其他方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 有没有解？如何求解？

微积分的注解：将这些特殊矩阵与微积分产生关联，向量 $\boldsymbol{x}$ 变成函数 $x(t)$，差分 $A\boldsymbol{x}$ 变成导数 $dx/dt=b(t)$。反向来说，总和 $A^{-1}\boldsymbol{b}$ 变成 $b(t)$的积分。**差的和就如同导数的积分**。

微积分的基础定理告诉我们：**积分是微分的逆反**。

$$A\boldsymbol{x}=\boldsymbol{b} \text{ 且 } \boldsymbol{x}=A^{-1}\boldsymbol{b} \qquad \frac{dx}{dt}=b \text{ 与 } x(t)=\int_0^t b\,dt \tag{8}$$

平方数 0, 1, 4, 9 的差是 1, 3, 5，$x(t)=t^2$ 的微分是 $2t$。当 $t=1, 2, 3$ 时得到偶数 $b=2, 4, 6$，这会是一个完美的类比。但是差分与微分不同，我们的矩阵 A 得到的不是 $2t$, 而是 $2t-1$:

$$\textbf{反向} \qquad x(t)-x(t-1)=t^2-(t-1)^2=t^2-(t^2-2t+1)=2t-1 \tag{9}$$

问题集会继续说明前向差分(forward difference)得到 $2t+1$。最好的选择(在微积分课程中不常见)是中心差分(centered difference)，形式是 $x(t+1)-x(t-1)$。将 Δx 除以距离 Δt，其中 Δt 是由 $t-1$ 到 $t+1$，所以 $\Delta t=2$：

$$x(t)=t^2 \text{ 的中心差分} \qquad \frac{(t+1)^2-(t-1)^2}{2}=2t \quad \text{(恰好)} \tag{10}$$

差分矩阵很伟大，中心差分是最佳，我们的第二个范例是不可逆。

循环差分(cyclic difference)

这个范例保持同样的 $\boldsymbol{u}$ 与 $\boldsymbol{v}$，只是把 $\boldsymbol{w}$ 换成 $\boldsymbol{w}^*$：

$$\textbf{第二个范例} \qquad \boldsymbol{u}=\begin{bmatrix}1\\-1\\0\end{bmatrix} \quad \boldsymbol{v}=\begin{bmatrix}0\\1\\-1\end{bmatrix} \quad \boldsymbol{w}^*=\begin{bmatrix}-1\\0\\1\end{bmatrix}$$

现在 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$的线性组合得到**循环差分矩阵** C:

$$\textbf{循环} \qquad C\boldsymbol{x}=\begin{bmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}x_1-x_3\\x_2-x_1\\x_3-x_2\end{bmatrix}=\boldsymbol{b} \tag{11}$$

矩阵 C 不是三角形，给定 $\boldsymbol{b}$ 时不是那么简单求解 $\boldsymbol{x}$。事实上是不可能求得 $C\boldsymbol{x}=\boldsymbol{b}$ 的解，因为这三个方程式不是得到**无限多解**(偶尔)，就是**无解**(经常)：

$$C\boldsymbol{x}=\boldsymbol{0} \text{ 有无限多 } \boldsymbol{x} \qquad \begin{bmatrix}x_1-x_3\\x_2-x_1\\x_3-x_2\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \text{ 的解是所有的向量 } \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}c\\c\\c\end{bmatrix} \tag{12}$$

每一个常数向量，例如 $\boldsymbol{x}=(3, 3, 3)$在循环时的差都是 0。未定常数 c 就如同求取积分时的$+C$。循环差分使得第一分量变成 x_1-x_3，而不是从 $x_0=0$ 开始。

请尊重版权与译者的劳动成果，侵权必究!

$C\boldsymbol{x}=\boldsymbol{b}$ 更大的可能性是 $\boldsymbol{x}$ **完全无解**：

$$C\boldsymbol{x}=\boldsymbol{b} \qquad \begin{bmatrix} x_1-x_3 \\ x_2-x_1 \\ x_3-x_2 \end{bmatrix}=\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \qquad \begin{matrix} \text{左侧相加等于0} \\ \text{右侧相加等于9} \\ x_1,\ x_2,\ x_3 \text{ 无解} \end{matrix} \tag{13}$$

从几何的观点来看，不存在 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$的线性组合可以得到向量 $\boldsymbol{b}=(1, 3, 5)$，这样的线性组合无法形成全部的三维空间。右侧必须有 $b_1+b_2+b_3=0$，才能允许 $C\boldsymbol{x}=\boldsymbol{b}$ 有一个解，这是因为左侧的 $x_1-x_3, x_2-x_1, x_3-x_2$ 相加必定为 0。换言之：

所有的线性组合 $x_1\boldsymbol{u}+x_2\boldsymbol{v}+x_3\boldsymbol{w}^*$落在 $b_1+b_2+b_3=0$ **给定的平面**。

这个主题突然把代数与几何结合在一起，线性组合可以形成整个空间，也可以只形成一个平面。我们需要一个图形来展示 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$(第一个例子)与 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$(在同一个平面)之间的重要差别:

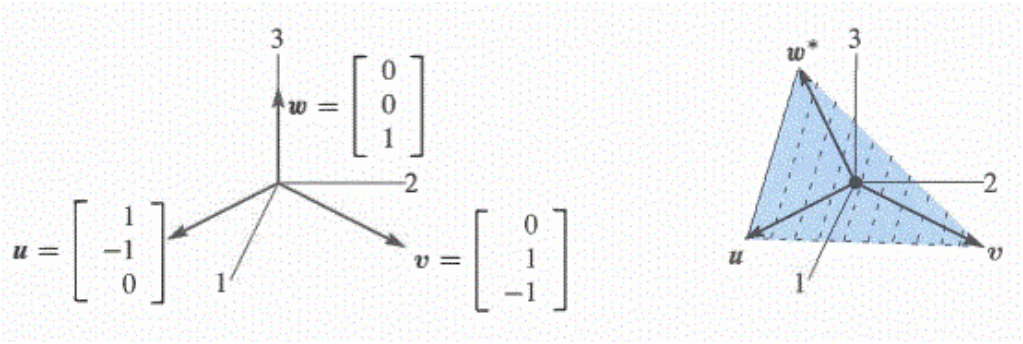


图 1.10：无关向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$；落在同一平面的相关向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$。

无关与相关

图 1.10 中第一个展示矩阵 A 的列向量，后面展示 C 的列向量。前两个列 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 在两个图形中完全相同，如果我们只检视这两个向量的组合，我们得到一个二维平面，**关键问题在于第三个向量是否在这个平面上**。

无关(independence) $\boldsymbol{w}$ 不在 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的平面上。

相关(dependence) $\boldsymbol{w}^*$在 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的平面上。

重点是新向量 $\boldsymbol{w}^*$是 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的线性组合：

$$\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}^*=\boldsymbol{0} \qquad \boldsymbol{w}^*=\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}=-\boldsymbol{u}-\boldsymbol{v} \tag{14}$$

这三个向量 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$的分量总和都是零，所有的线性组合会有 $b_1 + b_2 + b_3 = 0$(如同我们前面说的，把三个方程式相加)，这就是 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的线性组合形成平面的方程式。现在引进 $\boldsymbol{w}^*$，因为 $\boldsymbol{w}^*$已经在这个平面上，我们没有得到新的向量。

原始 $\boldsymbol{w} = (0, 0, 1)$不在平面上：$0 + 0 + 1 \neq 0$，$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 的线性组合形成整个三维空间。我们知道，因为方程式(6)的解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 给出正确的组合产生任意 $\boldsymbol{b}$。

前面两个矩阵 A 与 C，对应的第三个向量分别是 $\boldsymbol{w}$ 与 $\boldsymbol{w}^*$，允许我提出两个线性代数的关键字：无关与相关。本课程的前半段会更深入探讨这些观念--如果你在前面两个例子中已经学会，我会很快乐的：

$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 无关，除了 $0\boldsymbol{u} + 0\boldsymbol{v} + 0\boldsymbol{w} = \boldsymbol{0}$ 之外，没有其他线性组合得到 $\boldsymbol{b} = \boldsymbol{0}$。

$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}^*$相关，存在诸如 $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}^*$的其他组合得到 $\boldsymbol{b} = \boldsymbol{0}$。

你可以在三维空间作图，三个向量落在一个平面，或者不在同一个平面。第二章讨论 n 维空间中的 n 个向量，无关或相关是关键点，这些向量进入一个 $n \times n$ 矩阵的列:

无关列：$A\boldsymbol{x} = \boldsymbol{0}$ 有一个解，A 是**可逆矩阵**。

相关列：$C\boldsymbol{x} = \boldsymbol{0}$ 有很多解，A 是**奇异矩阵**。

最终我们会讨论 m 维空间中的 n 个向量，有 n 个列的矩阵 A 现在是矩形($m \times n$)。了解 $A\boldsymbol{x} = \boldsymbol{b}$ 是第三章的问题。

主要观念的复习

1. 矩阵乘向量：$A\boldsymbol{x} = \boldsymbol{A}$ **的列的组合**。
2. 当 A 是可逆矩阵，$A\boldsymbol{x} = \boldsymbol{b}$ 的解是 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$。
3. 循环矩阵 C 没有逆矩阵，它的三个列落在同一平面，这些相关的列相加得到零向量，$C\boldsymbol{x} = \boldsymbol{0}$ 有很多解。
4. 本段落是关键概念的超前理解，还没有完全说明。

已解范例

1.3A　把 A 的西南角单元 a_{31}(行 3，列 1)改成 $a_{31} = \boldsymbol{1}$：【entry 翻译成单元，有别于元素 element】

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ \boldsymbol{1} & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 + x_2 \\ \boldsymbol{x_1} - x_2 + x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

对任意 $\boldsymbol{b}$ 求解 $\boldsymbol{x}$。从 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 看出逆矩阵 A^{-1}。

请尊重版权与译者的劳动成果，侵权必究!

解　由上而下求解(线性三角形)系统 $A\boldsymbol{x}=\boldsymbol{b}$:

$$\begin{matrix}\text{首先 } x_1=b_1\\ \text{然后 } x_2=b_1+b_2\\ \text{然后 } x_3=b_2+b_3\end{matrix}\quad \text{意思是 } \boldsymbol{x}=A^{-1}\boldsymbol{b}=\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

这是很好的练习，可以看到逆矩阵的列乘 b_1, b_2, b_3。A^{-1} 的第一列是对应 $\boldsymbol{b}$ = (1, 0, 0)的解；第二列是对应 $\boldsymbol{b}$ = (0, 1, 0)的解；A^{-1} 的第三列 $\boldsymbol{x}$ 是对应 $A\boldsymbol{x}=\boldsymbol{b}$ = (0, 0, 1)的解。

矩阵 A 的三个列仍然无关，他们不在同一个平面，这三个列的线性组合，使用正确的加权 x_1, x_2, x_3，可以产生任意的三维向量 $\boldsymbol{b}=(b_1, b_2, b_3)$，这些加权来自 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$。

1.3B　E 是一个**消元(elimination)矩阵**，E 有一个减法，E^{-1} 有一个加法。

$$\boldsymbol{b}=E\boldsymbol{x}\qquad \begin{bmatrix}b_1\\b_2\end{bmatrix}=\begin{bmatrix}x_1\\x_2-lx_1\end{bmatrix}=\begin{bmatrix}1&0\\-l&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\qquad E=\begin{bmatrix}1&0\\-l&1\end{bmatrix}$$

第一个方程式是 $x_1=b_1$，第二个方程式是 $x_2-lx_1=b_2$。因为消元矩阵存在减法，逆矩阵会把 lb_1 加到 b_2：

$$\boldsymbol{x}=E^{-1}\boldsymbol{b}\qquad \begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\lb_1+b_2\end{bmatrix}=\begin{bmatrix}1&0\\l&1\end{bmatrix}\begin{bmatrix}b_1\\b_2\end{bmatrix}\qquad E^{-1}=\begin{bmatrix}1&0\\l&1\end{bmatrix}$$

1.3C　把 C 从循环差分变成**中心差分**产生 x_3-x_1:

$$C\boldsymbol{x}=\boldsymbol{b}\qquad \begin{bmatrix}0&1&0\\-1&0&1\\0&-1&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}x_2-0\\x_3-x_1\\0-x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}\tag{15}$$

$C\boldsymbol{x}=\boldsymbol{b}$ 只有在 $b_1+b_3=x_2-x_2=0$ 才有解，这是三维空间中向量 $\boldsymbol{b}$ 的一个平面。C 的每一列都在这个平面上，这个矩阵没有逆矩阵，所以平面包含了这些列的全部组合(就是所有的向量 $C\boldsymbol{x}$)。

我将 0 包含进去，所以你看到 C 产生“中心差分”，$C\boldsymbol{x}$ 的行 i 是 x_{i+1}(中心的右)减去 x_{i-1}(中心的左)。以下是 4×4 的例子：

$$\begin{matrix}C\boldsymbol{x}=\boldsymbol{b}\\ \text{中心差分}\end{matrix}\qquad \begin{bmatrix}0&1&0&0\\-1&0&1&0\\0&-1&0&1\\0&0&-1&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}x_1-0\\x_3-x_1\\x_4-x_2\\0-x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}\tag{16}$$

惊讶的是这个矩阵现在是可逆！第一行与最后一行告诉你 x_2 与 x_3，中间的行给出 x_1 与 x_4。可以继续往下写出逆矩阵 $\boldsymbol{C}^{-1}$，但是 5×5 矩阵又变成奇异(不可逆)...

问题集 1.3

1 求线性组合 $3\boldsymbol{s}_1 + 4\boldsymbol{s}_2 + 5\boldsymbol{s}_3 = \boldsymbol{b}$，然后将 $\boldsymbol{b}$ 写成矩阵-向量的乘积 $S\boldsymbol{x}$，其中 3, 4, 5 在 $\boldsymbol{x}$ 里面。再计算三个点积：(S 的行) $\cdot\, \boldsymbol{x}$：

$$\boldsymbol{s}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \quad \boldsymbol{s}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix} \quad \boldsymbol{s}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix} \quad \text{进入 } S \text{ 的列}$$

2 求解 $S\boldsymbol{y} = \boldsymbol{b}$，其中 S 的列是问题 1 的 $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$：

$$\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix} \text{ 以及 } \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}1\\4\\9\end{bmatrix}$$

S 是一个总和矩阵。前五个奇数的总和是_____。

3 求解以下三个方程式，使用 c_1, c_2, c_3 表示 y_1, y_2, y_3：

$$S\boldsymbol{y} = \boldsymbol{c} \qquad \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix} = \begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}$$

将 $\boldsymbol{y}$ 写成矩阵 $A = S^{-1}$ 乘向量 $\boldsymbol{c}$，请问 S 的列是无关或是相关？

4 当 $x_1 = 1$ 时，什么样的组合 $x_1\boldsymbol{w}_1 + x_2\boldsymbol{w}_2 + x_3\boldsymbol{w}_3$ 可以得到零向量：

$$\boldsymbol{w}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix} \quad \boldsymbol{w}_2 = \begin{bmatrix}4\\5\\6\end{bmatrix} \quad \boldsymbol{w}_3 = \begin{bmatrix}7\\8\\9\end{bmatrix}$$

这些向量是无关或是相关？这三个向量落在一个_____上。这三个列向量构成的矩阵 W 是不可逆。

5 矩阵 W 的行产生三个向量(我把他们写成列向量)：

$$\boldsymbol{r}_1 = \begin{bmatrix}1\\4\\7\end{bmatrix} \quad \boldsymbol{r}_2 = \begin{bmatrix}2\\5\\8\end{bmatrix} \quad \boldsymbol{r}_3 = \begin{bmatrix}3\\6\\9\end{bmatrix}$$

线性代数说这些向量必须落在同一平面，必定存在许多组合 $y_1\boldsymbol{r}_1 + y_2\boldsymbol{r}_2 + y_3\boldsymbol{r}_3 = \boldsymbol{0}$，求出两组的 y's。

6 什么样的 c 会得到相关的列，使得列的一个组合等于零。

$$\begin{bmatrix}1&1&0\\3&2&1\\7&4&c\end{bmatrix} \quad \begin{bmatrix}1&0&c\\1&1&0\\0&1&1\end{bmatrix} \quad \begin{bmatrix}c&c&c\\2&1&5\\3&3&6\end{bmatrix} \qquad \text{当 } c \neq 0, \text{ 可能永远无关吗?}$$

请尊重版权与译者的劳动成果，侵权必究!

7　如果列组合得到 $A\boldsymbol{x}=\boldsymbol{0}$，则每一行都有 $\boldsymbol{r}\cdot\boldsymbol{x}=0$:

$$\begin{bmatrix} & & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{以行来看} \quad \begin{bmatrix} \boldsymbol{r}_1\cdot\boldsymbol{x} \\ \boldsymbol{r}_2\cdot\boldsymbol{x} \\ \boldsymbol{r}_3\cdot\boldsymbol{x} \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

这三个行也落在同一平面，为什么这个平面与 $\boldsymbol{x}$ 垂直?

8　现在讨论 4×4 的差分方程式 $A\boldsymbol{x}=\boldsymbol{b}$，求出四个分量 x_1, x_2, x_3, x_4，再将解写成 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$，求出逆矩阵：

$$A\boldsymbol{x}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}=\boldsymbol{b}$$

9　4×4 的循环差分矩阵 C 为何? 它在每一行与每一列都会有 1 与–1。求出所有的解 $\boldsymbol{x}=(x_1, x_2, x_3, x_4)$使得 $C\boldsymbol{x}=\boldsymbol{0}$。$C$ 的四个列会落在四维空间的一个三维超平面上。

10　前向差分矩阵Δ是上三角形：

$$\Delta\boldsymbol{z}=\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}=\begin{bmatrix} z_2-z_1 \\ z_3-z_2 \\ 0-z_3 \end{bmatrix}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}=\boldsymbol{b}$$

利用 b_1, b_2, b_3 求 z_1, z_2, z_3。$\boldsymbol{z}=\Delta^{-1}\boldsymbol{b}$ 的逆矩阵为何？

11　证明前向差分$(t+1)^2-t^2$ 是 $2t+1=$ 奇数。如同微积分所述，$(t+1)^n-t^n$ 会从 t^n 的导数开始，就是_____。

12　**已解范例**的最后一行显示方程式(16)里面的 4×4 中心差分矩阵是可逆，求解 $C\boldsymbol{x}=(b_1, b_2, b_3, b_4)$，找出 $\boldsymbol{x}=\boldsymbol{C}^{-1}\boldsymbol{b}$ 的逆矩阵。

挑战问题

13　前述说明 5×5 中心差分矩阵不可逆。请写出 5 个 $C\boldsymbol{x}=\boldsymbol{b}$ 的方程式，求出左侧的线性组合得到零。b_1, b_2, b_3, b_4, b_5 的何种组合必须等于零？(这 5 列落在 5 维空间的 4 维超平面上，难以视觉化。)

14　若(a, b)是(c, d)的倍数且 $abcd\neq 0$，证明(a, c)是(b, d)的倍数。这个结果非常重要，这两列落在同一条直线上。你可以先用数字看看 a, b, c, d 之间的关系。方程组会得到：

$$\text{若}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\text{有相关的行，必定有相关的列。}$$

第二章

求解线性方程组

2.1 向量与线性方程组

1. $A\boldsymbol{x}=\boldsymbol{b}$ 的**列图**：(A 的 n 个列)的组合产生向量 $\boldsymbol{b}$。
2. 这是向量方程式 $A\boldsymbol{x}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+...+x_n\boldsymbol{a}_n=\boldsymbol{b}$：$A$ 的列是 $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n$。
3. 当 $\boldsymbol{b}=\boldsymbol{0}$，列的组合 $A\boldsymbol{x}$ 是零：有一种可能是 $\boldsymbol{x}=(0, 0, ..., 0)$。
4. $A\boldsymbol{x}=\boldsymbol{b}$ 的**行图**：来自 m 个行的 m 个方程式，产生 m 个在 $\boldsymbol{x}$ 交会的平面。
5. 点积得到每个平面的方程式：(**行 1**) $\cdot\,\boldsymbol{x}=b_1$，...，(**行 *m***) $\cdot\,\boldsymbol{x}=b_m$。
6. 当 $\boldsymbol{b}=\boldsymbol{0}$，所有的平面(**行 *i***) $\cdot\,\boldsymbol{x}=\boldsymbol{0}$ 会经过中心点 $\boldsymbol{x}=(0, 0, ..., 0)$。

线性代数的中心问题是求解一个方程组(或称方程式系统)，这些方程式是线性的，也就是说未知数只会乘数字—我们没有见过 x 乘 y。刚开始的线性系统比较小，但是你看到他们是可以继续延伸的：

两个方程式
两个未知数

$$\begin{aligned} x-2y&=1\\ 3x+2y&=11 \end{aligned} \tag{1}$$

我们一次处理一行，第一个方程式是 $x-2y=1$，在 xy 平面是一条直线，点 $x=1$, $y=0$ 在这条直线上，因为这个点坐标是方程组的解。点 $x=3, y=1$ 也在这条直线上，因为 $3-2=1$。如果选取 $x=101$，我们发现 $y=50$。

前述直线的斜率是 1/2，这是因为 x 改变 2 时，y 会增加 1。斜率在微积分课程很重要，只是我们现在学的是线性代数！

图 2.1 显示第一条直线 $x-2y=1$。在“行图”的第二条直线来自第二个方程式 $3x+2y=11$。你不要错过了两条直线的交会点 $x=3, y=1$，点(3, 1)会同时落在两条直线上，也是两个方程式的解。

请尊重版权与译者的劳动成果，侵权必究！

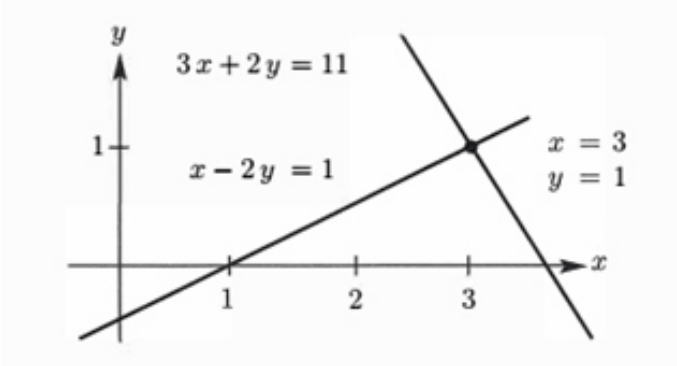


图 2.1：行图：点(3, 1)是两条直线的交点，也是两个方程式的解。

行　　行图显示两条直线交于一点(解)。

现在转到列图，我想要把相同的线性系统写成“**向量方程式**”，我们需要观察向量而不是数字。如果你把原始系统分割成列而不是行，你得到向量方程式:

$$\textbf{组合等于}\ \boldsymbol{b} \qquad x\begin{bmatrix}1\\3\end{bmatrix}+y\begin{bmatrix}-2\\2\end{bmatrix}=\begin{bmatrix}1\\11\end{bmatrix}=\boldsymbol{b} \tag{2}$$

左侧有两个列向量，问题在于找到这些向量的组合能够等于右侧的向量。我们对第一列乘 x，对第二列乘 y，然后相加。正确的选择是 $x = 3$ 与 $y = 1$(与前述数字相同),，得到 3(**列 1**) + 1(**列 2**) = $\boldsymbol{b}$。

列　列图组合左侧的列向量产生右侧的向量 $\boldsymbol{b}$。

图 2.2 是两个方程式两个未知数的**列图**，第一部分展示两个不同的列，然后第一列乘 3，纯量(数字)的乘法是线性代数中两个基本运算其中之一：

$$\textbf{纯量乘法} \qquad 3\begin{bmatrix}1\\3\end{bmatrix}=\begin{bmatrix}3\\9\end{bmatrix}$$

如果向量 $\boldsymbol{v}$ 的分量是 v_1 与 v_2，则 $c\boldsymbol{v}$ 的分量是 cv_1 与 cv_2。

另一个基础运算是向量加法，分别对第一分量与第二分量相加，向量总和是(1, 11)，也就是想要的 $\boldsymbol{b}$：

$$\textbf{向量加法} \qquad \begin{bmatrix}3\\9\end{bmatrix}+\begin{bmatrix}-2\\2\end{bmatrix}=\begin{bmatrix}1\\11\end{bmatrix}$$

图 2.2 的右侧显示这样的加法。两个向量以黑线表示，沿着对角线的总和就是线性方程式右侧的向量 $\boldsymbol{b} = (1, 11)$。

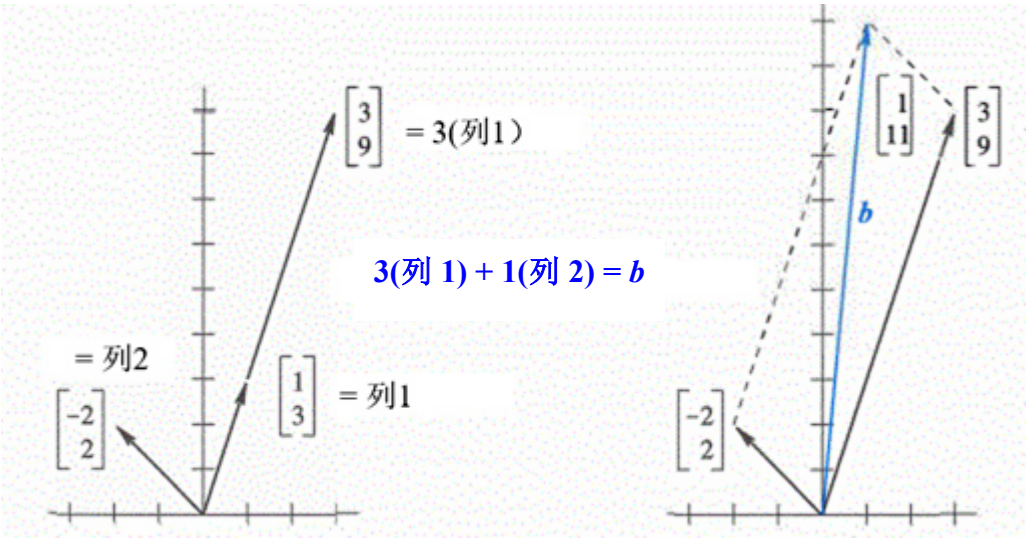


图 2.2：列图：列的组合产生右侧的(1, 11)。

重复说明：向量方程式的左侧是列的**线性组合**，问题在于找出正确的系数 $x = 3$ 与 $y = 1$。我们把纯量乘法及向量加法组合成一个步骤，这个步骤非常重要，因为它包含了两个基础运算：**乘 3 及 1，然后相加**。

$$\textbf{线性组合} \qquad 3\begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}-2\\2\end{bmatrix}=\begin{bmatrix}1\\11\end{bmatrix}$$

解 $x = 3, y = 1$ 必然与行图的解相同。我不知道你比较喜欢哪个图？我觉得一开始两条交会的直线会比较熟悉，你可能喜欢行图多一点，但是只能维持一天，我个人的偏好是组合列向量。在四维空间观察 4 个向量的组合会比起视觉化某些超平面交于一点来的更容易一些。(一个超平面就已经很难了…)

方程式左侧的**系数矩阵**是 2×2 矩阵 A：

$$\textbf{系数矩阵} \qquad A=\begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}$$

从行或是从列来观察矩阵，这是线性代数非常典型的做法。行可以得到行图，列可以得到列图，相同的数字，不同的图，相同的方程组。我们结合那些方程式成为矩阵问题 $A\boldsymbol{x} = \boldsymbol{b}$:

$$\textbf{矩阵方程式 } A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{bmatrix}1 & -2\\3 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1\\11\end{bmatrix}$$

译者注：在英文中 A multiplied by B 表示“A 被 B 乘 ＝B 乘 A”，如果 B 是纯量，就无所谓顺序。A multiplies B 或 A times B 表示 A 乘 B。原文的 Multiply A times B，意义就是执行“A 乘 B”。

请尊重版权与译者的劳动成果，侵权必究！

行图处理 A 的两个行，列图组合列。数值 $x=3$ 与 $y=1$ 进入 $\boldsymbol{x}$，以下是矩阵-向量乘法：

行的点积
列的组合

$$A\boldsymbol{x}=\boldsymbol{b} \qquad \begin{bmatrix}1 & -2\\ 3 & 2\end{bmatrix}\begin{bmatrix}3\\ 1\end{bmatrix}=\begin{bmatrix}1\\ 11\end{bmatrix}$$

展望未来：本章求解有 n 个未知数的 n 个方程式(任意 n)，我不想极速前进，因为较小的系统会让范例与图形得到完全的了解。只要**矩阵乘法与逆矩阵变得清晰**，你可以自由的走快一些。这两个观念将是逆矩阵的关键。

我可以列出四个步骤来了解使用矩阵执行消元法。

1. 借由一系列的矩阵步骤 E_{ij}，消元法把 A 变成三角 U。
2. 三角系统使用**反向代入法**求解：由底部往顶端工作。
3. 用矩阵语言来说，A 分解成 $LU=$(下三角)(上三角)。
4. 若 A 可逆，消元法才能成功。(但是可能需要交换行。)

计算机科学里面最常使用的演绎法都是使用这些步骤(MATLAB 称为 lu)，最快速的形式是反斜线(backslash)：$\boldsymbol{x}=A\backslash\boldsymbol{b}$。但是线性代数会超越方形逆矩阵！对 $m\times n$ 矩阵来说，$A\boldsymbol{x}=\boldsymbol{b}$ 可能会有很多解，这些解形成**向量空间**，A 的**秩(rank)**就是向量空间的**维度**。

这些会在第三章讨论，我不能急躁，但是我必须完成使命。

三个未知数的三个方程式

三个未知数是 x, y, z，我们有三个方程式：

$$A\boldsymbol{x}=\boldsymbol{b} \qquad \begin{aligned} x+2y+3z&=6\\ 2x+5y+2z&=4\\ 6x-3y+z&=2\end{aligned} \tag{3}$$

我们想要一次性找到 x, y, z 解答三个方程式，这些数值可能存在也可能不存在。本系统的解是存在的，当未知数的个数与方程式的个数相等时(本例题中 3 = 3)，**通常**会有一个解。

在求解之前，我们从两个方面来观察：

行　行图显示三个平面交会于一个点。

列　列图组合三个列产生 $\boldsymbol{b}=(6, 4, 2)$。

在行图中，每个方程式产生一个三维空间中的平面，图 2.3 的第一个平面来自第一个方程式 $x+2y+3z=6$，这个平面与 x 与 y 与 z 轴的交点是(6, 0, 0)与(0, 3, 0)与 (0, 0, 2)，这三个点满足方程组并且决定整个平面。

向量$(x, y, z) = (0, 0, 0)$不是 $x + 2y + 3z = 6$ 的解，因此这个平面不包含原点。平面 $x + 2y + 3z = 0$ 会通过原点，并且与 $x + 2y + 3z = 6$ 平行。当右侧从 0 增加到 6，平行平面会离开原点。

第二个平面由第二个方程式 $2x + 5y + 2z = 4$ 决定，它与第一个平面的交集是一条直线 $\boldsymbol{L}$，三个未知数两个方程式的解通常是一条直线 $\boldsymbol{L}$。(如果方程式是 $x + 2y + 3z = 6$ 以及 $x + 2y + 3z = 0$，交集就不是直线)。

第三个方程式得到第三个平面，它把直线 $\boldsymbol{L}$ 切在一个点，这个点落在全部三个平面上，就是三个方程式的解。画出三个平面交于一点比较困难，用想象的会比较容易。这三个平面会在解的地方交会(我们还没有找到)。**列形式可以马上说明为什么 $z = 2$。**

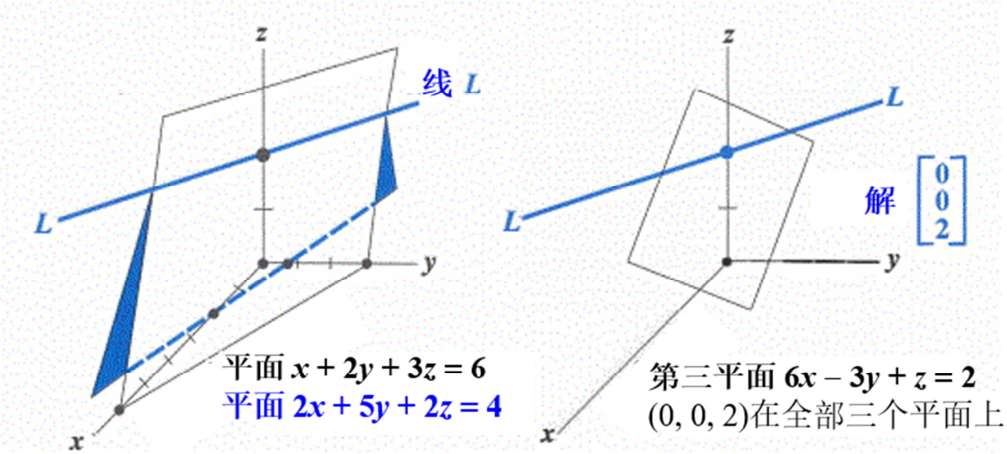


图 2.3：行图：两个平面交于直线 $\boldsymbol{L}$，三个平面交于一点

列图从方程式 $\boldsymbol{Ax = b}$ 的向量形式开始：

$$\textbf{组合列}\qquad x\begin{bmatrix}1\\2\\6\end{bmatrix}+y\begin{bmatrix}2\\5\\-3\end{bmatrix}+z\begin{bmatrix}3\\2\\1\end{bmatrix}=\begin{bmatrix}6\\4\\2\end{bmatrix}=\boldsymbol{b} \tag{4}$$

未知数是系数 x, y, z，我们想要对三个列向量乘正确数字 x, y, z 产生 $\boldsymbol{b} = (6, 4, 2)$。图 2.4 展示本题的列图，这些列向量的线性组合可以产生任意的 $\boldsymbol{b}$！产生 $\boldsymbol{b} = (6, 4, 2)$的组合就是第三列乘 2，我们需要的系数是 $x = 0, y = 0, z = 2$。

行图中的三个平面也交会在相同的解点(0, 0, 2):

$$\textbf{正确组合}(x, y, z) = \textbf{(0, 0, 2)}\qquad \mathbf{0}\begin{bmatrix}1\\2\\6\end{bmatrix}+\mathbf{0}\begin{bmatrix}2\\5\\-3\end{bmatrix}+\mathbf{2}\begin{bmatrix}3\\2\\1\end{bmatrix}=\begin{bmatrix}6\\4\\2\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

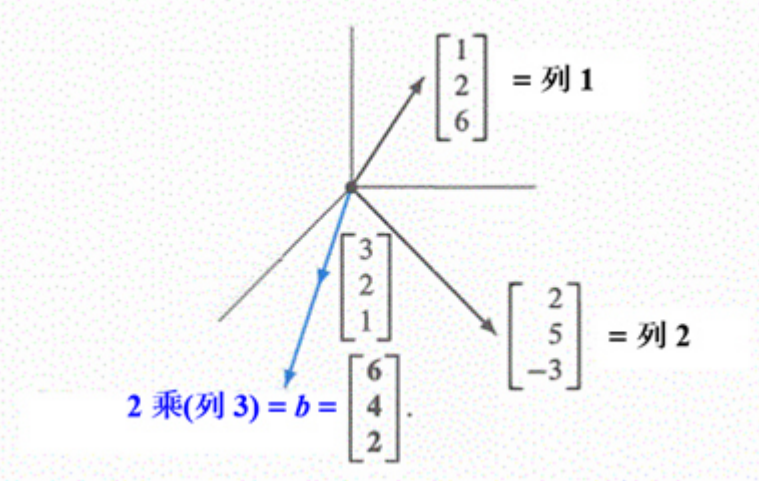


图 2.4：列图：组合列向量，其中$(x, y, z) = (0, 0, 2)$

方程组的矩阵形式

在行图我们有三个行，在列图有三个列(加上右侧)，这三行与三列包含 9 个数字，这 9 个数字形成一个 3×3 矩阵 A：

$$A\boldsymbol{x} = \boldsymbol{b} \text{ 的系数矩阵为} \qquad A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix}$$

大写字母 A 代表 9 个系数(形成方形阵列)，字母 $\boldsymbol{b}$ 代表列向量，其分量是 6, 4, 2。未知数 $\boldsymbol{x}$ 也是一个列向量，分量是 x, y, z。(因为是向量，我们用粗体字表示；因为它是未知，用 $\boldsymbol{x}$ 表示)。以行来看方程式是(3)，以列来看方程式是(4)，以矩阵来看方程式是(5)

$$A\boldsymbol{x} = \boldsymbol{b} \text{ 的矩阵方程式} \qquad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix} \tag{5}$$

基础问题：**A 乘 x 的意义是什么**？我们可以**被行乘**或**被列乘**，不管哪一种方法，$A\boldsymbol{x} = \boldsymbol{b}$ 必须是三个方程式的正确表述。你可以用两种方法做同样的 9 个乘法。

被行乘

$A\boldsymbol{x}$ 来自**点积**，每个行乘列 $\boldsymbol{x}$：

$$A\boldsymbol{x} = \begin{bmatrix} (\text{行 } 1) \cdot \boldsymbol{x} \\ (\text{行 } 2) \cdot \boldsymbol{x} \\ (\text{行 } 3) \cdot \boldsymbol{x} \end{bmatrix} \tag{6}$$

被列乘

$A\boldsymbol{x}$ 是**列向量的组合**：

$$A\boldsymbol{x} = x(\text{列}1) + y(\text{列}2) + z(\text{列}3) \tag{7}$$

当我们代入解 $\boldsymbol{x} = (0, 0, 2)$，乘法 $A\boldsymbol{x}$ 产生 $\boldsymbol{b}$：

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = 2\text{乘}(\text{列}3) = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

来自第一行的点积是$(1, 2, 3) \cdot (0, 0, 2) = 6$，其他行的点积是 4 与 2。**本书将 $A\boldsymbol{x}$ 视为 A 的列的组合**。

范例 1　给定 3×3 矩阵 A 与单位矩阵 I，I 有 3 个 1 与 6 个 0：

$$A\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \qquad I\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

如果你喜欢行，(1, 0, 0)与(4, 5, 6)的点积是 4。如果你喜欢列，线性组合 $A\boldsymbol{x}$ 是 4 乘第一列(1, 1, 1)，矩阵 A 的第二与第三列都是零向量。

另一个矩阵 I 很特别，**主对角线**都是 1，这个矩阵乘任何向量都不会改变原来的向量，如同乘上 1 一样，差别只是现在是矩阵与向量而已。本例题的特殊矩阵称为 3×3 **单位矩阵**：

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{永远存在乘法} \quad I\boldsymbol{x} = \boldsymbol{x}$$

矩阵表示法

一个 2×2 矩阵的第一行包含 a_{11} 与 a_{12}，第二行包含 a_{21} 与 a_{22}。第一个下标(index)表示行数，所以 a_{ij} 是第 i 行的单元。第二个下标 j 表示列数。但是这些下标在打字的时候不太方便，一般我们不写成 a_{ij} 而是写成 $A(i, j)$。**单元 $A(5, 7)$就在第 5 行第 7 列的位置**。

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} A(1,\ 1) & A(1,\ 2) \\ A(2,\ 1) & A(2,\ 2) \end{bmatrix}$$

对于 $m \times n$ 矩阵来说，行的下标 i 由 1 到 m，列的下标 j 最多到 n，总共有 mn 个单元 $a_{ij} = A(i, j)$，一个 n 阶(order)的方形矩阵有 n^2 个单元。

请尊重版权与译者的劳动成果，侵权必究！

MATLAB 中的乘法

我想要使用 MATLAB 的指令表达 A 与 $\boldsymbol{x}$ 以及乘积 $A\boldsymbol{x}$，这是学习 MATLAB 语言(其他软件也是)的第一步。我们从定义 A 与 $\boldsymbol{x}$ 开始，R^n 的向量 $\boldsymbol{x}$ 是一个 $n \times 1$ 的矩阵(本书也是这么描述)，输入矩阵时每次输入一行，用分号(;)代表一行的结束，或者以列输入，然后利用′代表转置:

$A = [1 \quad 2 \quad 3; \quad 2 \quad 5 \quad 2; \quad 6 \quad -3 \quad 1];$

$\boldsymbol{x} = [0 \quad 0 \quad 2]'$ 或 $\boldsymbol{x} = [0; 0; 2]$

在 MATLAB 有三种方式得到 $A\boldsymbol{x}$， 实际上 $A * \boldsymbol{x}$ 是很好的方法。MATLAB 是高阶语言，可以处理矩阵：

矩阵乘法 $\quad \boldsymbol{b} = A * \boldsymbol{x}$

我们可以选出 A 的第一行(视为较小的矩阵！)，这个 1×3 矩阵的表示法为 $A(1, :)$。此处的冒号: 代表第一行的全部列。 【原文 3×3 是错误】

一次处理一行 $\quad \boldsymbol{b} = [A(1, :) * \boldsymbol{x}; \quad A(2, :) * \boldsymbol{x}; \quad A(3, :) * \boldsymbol{x}]$

$\boldsymbol{b}$ 的每一个单元都是一个点积，行乘列，1×3 矩阵乘 3×1 矩阵。

另一种方式是使用 A 的列来乘，第一列是 3×1 子矩阵 $A(:, 1)$，现在冒号放在前面，意思是第一列全部的行。第一列乘 $x(1)$，其他两列乘 $x(2)$与 $x(3)$：

一次处理一列 $\quad \boldsymbol{b} = A(:, 1) * \boldsymbol{x}(1) + A(:, 2) * \boldsymbol{x}(2) + A(:, 3) * \boldsymbol{x}(3)$

我认为矩阵是以列来存放，一次乘一个列应该会比较快速，所以 $A * \boldsymbol{x}$ 实际上是用列来处理的。

数学与统计的程序语言

还有 5 种语言的指令可以处理 $A\boldsymbol{x}$:

Julia	$A * \boldsymbol{x}$	julialang,org
Python	dot(A, $\boldsymbol{x}$)	python,org
R	A %* %$\boldsymbol{x}$	r-projice.org
Mathematica	$A.\boldsymbol{x}$	wolfram.com/mathematica
Maple	$A * \boldsymbol{x}$	maplesoft.com

Julia, Python, R 是免费的开源语言，R 是特别为了统计的应用。其他统计软件(SAS, JMP, …)在维基百科里有统计软件的比较。

Mathematica 与 Maple 允许符号式的单元 a, b, x, …，不一定需要实际的数字。如同 MATLAB 的符号工具箱一样，可以处理例如 $x^2 - x$ 的符号表述。Mathemtica 的强大可以在 Wolfram Alpha 搜寻引擎中看到。

Julia 结合 SciPy 或 R 在科学计算的高性能，这是可以跟 *C* 或是 Fortran 相比拟的，也可以呼叫 Python 与 C/Fortran 的函数库，但是无法依赖向量式函数库来提升速度。Julia 是快速的。

我输入 juliabox.org, 然后借由谷歌 gmail 邮件账号点击“sign in”，点击右侧的 new，选择一个 Julia notebook。我选择 0.4.5 而不是选择开发中的项目，Julia 命令列就会马上显示出来。

对于新手而言，我想计算 1 + 1，键入 shift + enter 可以看到结果。我也想学习浮点计算 1.0 + 1.0，对于大型问题来说会快很多。

网址：math.mit.edu/linearalgebra 会看到 Julia, Python, R 的部分强大功能。

Python 是一个流行的通用程序语言，结合 Numpy 与 SciPy 的软件包，提供一个全方位的科学计算环境。NumPy 有具有基础的线性代数指令，可以从 https://www.continuum.io 下载 Anaconda Python(包含 Python 的预包装集以及大部分重要的数学库，是一个图形的安装器)。

R 是统计的计算与图形的免费软件，请至 https://r-projice.org 下载。指令的提示符号是 >, R 是一个脚本语言，主要是处理表列资料，可以扩展至矩阵与向量。

郑重的推荐 RStudio 在编辑与图形的功能，当你从 www.RStudio.com 下载，R 命令的视窗会出现，还有编辑、处理档案以及画图的视窗会出现。告诉 R 矩阵的形式以及内含的单元数字：

```
> A = matrix(c(1,2,3,2,5,2,6,-3,1), nrow=3, byrow = TRUE)
> x = matrix(c(0,0,2), nrow=3)
```

要查看 A 与 x，可以在>之后键入名字；乘法的表示：b = A% * % x, 转置以 t(A) 代表，利用 as.matrix 将向量变成矩阵。

MATLAB 与 Julia 在矩阵计算的语法上比 R 来的清楚，但是 R 已经非常流行广泛使用。本书的网址有一些软件的 demo，包含 MATLAB, Julia, Python 以及 R。

主要观念的复习

1. 向量的基础运算是纯量乘法 $c\boldsymbol{v}$ 与向量加法 $\boldsymbol{v}+\boldsymbol{w}$。
2. 上述两者结合得到线性组合 $c\boldsymbol{v}+d\boldsymbol{w}$。
3. 矩阵––向量的乘法 $A\boldsymbol{x}$ 可以由点积得到，一次处理一行。但是 $A\boldsymbol{x}$ 必须理解为 A 的列的线性组合。
4. 列图：$A\boldsymbol{x}=\boldsymbol{b}$ 在寻找列的组合产生 $\boldsymbol{b}$。
5. 行图：$A\boldsymbol{x}=\boldsymbol{b}$ 的每个方程式得到一条直线($n=2$)，或一个平面($n=3$)或是一个超平面($n>3$)。如果有解，交集会在一个解或是多个解。

请尊重版权与译者的劳动成果，侵权必究!

已解范例

2.1A 描述三个方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 的列图，仔细观察列(不用消元法)然后求解：

$$\begin{aligned} x+3y+2z&=-3\\ 2x+2y+2z&=-2\\ 3x+5y+6z&=-5 \end{aligned} \quad \text{即为} \quad \begin{bmatrix} 1 & \mathbf{3} & 2\\ 2 & \mathbf{2} & 2\\ 3 & \mathbf{5} & 6 \end{bmatrix}\begin{bmatrix} x\\ y\\ z \end{bmatrix}=\begin{bmatrix} -3\\ -2\\ -5 \end{bmatrix}$$

解 列图在于寻找 A 的三个列的线性组合，这些线性组合产生 $\boldsymbol{b}$。本题中 $\boldsymbol{b}$ 是第二列的负号，所以解是 $x=0, y=-1, z=0$。为了证明$(0,-1,0)$是唯一的解，我们必须确认 A 是可逆，三个列之间是无关，行列式不为 0。

这些名词还没有完全定义清楚，但是可以借由消元法得到证实。我们需要(对于本矩阵找到)一组三个非零**主元(pivot)**的集合。

假设右侧变成 $\boldsymbol{b}=(4,4,8)=$ 前两个列的和，正确的组合变成 $x=1, y=1, z=0$，新的解是 $\boldsymbol{x}=(1,1,0)$

2.1B 下列系统无解，行图中的平面没有交于一个点。不存在这三个列的线性组合产生 $\boldsymbol{b}$，如何证明?

$$\begin{aligned} x+3y+5z&=4\\ x+2y-3z&=5\\ 2x+5y+2z&=8 \end{aligned} \qquad \begin{bmatrix} 1 & 3 & 5\\ 1 & 2 & -3\\ 2 & 5 & 2 \end{bmatrix}\begin{bmatrix} x\\ y\\ z \end{bmatrix}=\begin{bmatrix} 4\\ 5\\ 8 \end{bmatrix}=\boldsymbol{b}$$

概念：执行(方程式 1) + (方程式 2) – (方程式 3)，结果是 **0 = 1**，这个系统不可能有解。我们可以说：向量$(1,1,-1)$与 A 的三个列全部正交，但是与 $\boldsymbol{b}$ 不正交。

(1) 三个平面中存在任意两个平行平面吗? 什么样的方程式会与平面 $x+3y+5z=4$ 平行?

(2) 计算 A 的每一列与 $\boldsymbol{y}=(1,1,-1)$的点积，也计算 $\boldsymbol{b}$ 与 $\boldsymbol{y}$ 的点积。这些点积如何说明不存在列的线性组合等于 $\boldsymbol{b}$?

(3) 求出右侧三个不同的向量 $\boldsymbol{b}^*, \boldsymbol{b}^{**}, \boldsymbol{b}^{***}$，使得方程式有解。

解

(1) 纵使没有两个平面平行，这些平面仍然没有交于一点。改变“4”就可以得到与 $x+3y+5z=\mathbf{4}$ 平行的平面，平面 $x+3y+5z=0$ 会经过原点$(0,0,0)$。方程式乘任意非零常数得到相同的平面，如同 $2x+6y+10z=8$。

(2) A 的每一列与 $\boldsymbol{y}=(1,1,-1)$的点积为 0，$\boldsymbol{y}\cdot\boldsymbol{b}=(1,1,-1)\cdot(4,5,8)=\mathbf{1}$ 不是零，$A\boldsymbol{x}=\boldsymbol{b}$ 变成 $0=1$：**无解**。

(3) 当 $\boldsymbol{b}$ 是列的线性组合时，存在一个解。三种 $\boldsymbol{b}$ 的选择对应解 $\boldsymbol{x}^*=(1,0,0)$, $\boldsymbol{x}^{**}=(1,1,1), \boldsymbol{x}^{***}=(0,0,0)$：

$$\boldsymbol{b}^*=\begin{bmatrix} 1\\ 1\\ 2 \end{bmatrix}=\text{第一列} \qquad \boldsymbol{b}^{**}=\begin{bmatrix} 9\\ 0\\ 9 \end{bmatrix}=\text{列的和} \qquad \boldsymbol{b}^{***}=\begin{bmatrix} 0\\ 0\\ 0 \end{bmatrix}$$

问题集 2.1

问题 1-8 有关 $Ax = b$ 的行图与列图。

1 当 $A = I$ (单位矩阵), 在行图中画出以下平面。一个盒子的三边交集在解 $\boldsymbol{x} = (x, y, z) = (2, 3, 4)$:

$$\begin{aligned} 1x + 0y + 0z &= 2 \\ 0x + 1y + 0z &= 3 \\ 0x + 0y + 1z &= 4 \end{aligned} \qquad \text{或是} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

画出列图中的向量。2 乘列 1 加 3 乘列 2 加 4 乘列 3 等于右侧的 $\boldsymbol{b}$。

2 若问题 1 的方程式分别乘以 2, 3, 4，他们变成 $D\boldsymbol{X} = \boldsymbol{B}$:

$$\begin{aligned} 2x + 0y + 0z &= 4 \\ 0x + 3y + 0z &= 9 \\ 0x + 0y + 4z &= 16 \end{aligned} \qquad \text{或是} \qquad D\boldsymbol{X} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \\ 16 \end{bmatrix} = \boldsymbol{B}$$

为什么行图保持不变? 解 $\boldsymbol{X}$ 与 $\boldsymbol{x}$ 是否相同? 列图有什么变化--这些列或是正确的组合会产生 $\boldsymbol{B}$ 吗?

3 若方程式 1 加到方程式 2，何者会改变: 行图中的平面? 列图中的向量? 系数矩阵? 方程组的解? 问题 1 的新方程组会变成 $x = 2, x + y = 5, z = 4$。

4 当 $z = 2$，找出两个平面 $x + y + 3z = 6$ 与 $x - y + z = 4$ 交线上的点。找出当 $z = 0$ 的交点，找出第三点是两个点的中点。

5 第一式加第二式等于第三式：

$$\begin{aligned} x + \ y + \ z &= 2 \\ x + 2y + \ z &= 3 \\ 2x + 3y + 2z &= 5 \end{aligned}$$

前两个平面交会在一条直线，第三个平面包含这条直线，这是因为如果 x, y, z 满足前两个方程式，他们也_____。方程式有无限多解(整条直线 **L**)。找出 **L** 上的三个解。

6 把问题 5 的第三个平面移到平行平面 $2x + 3y + 2z = 9$，现在这三个方程式变成无解--为什么? 前两个方程式的交集是直线 **L**，但是第三平面不_____那条直线。

7 问题 5 的列是(1, 1, 2), (1, 2, 3)与(1, 1, 2)，因为第三列是_____，所以是奇异案例。找出两个列组合得到 $\boldsymbol{b} = (2, 3, 5)$。若 c = ___，只有当 $\boldsymbol{b} = (4, 6, c)$时才有解。

请尊重版权与译者的劳动成果，侵权必究!

8　通常 4 维空间的 4 个平面会交会在一个_____。通常 4 维空间的 4 个列向量的组合可以得到 $\boldsymbol{b}$，请问(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)的什么组合产生 $\boldsymbol{b} = (3, 3, 3, 3)$？对应上述问题的 x, y, z, t 的方程式为何？

问题 9-14 有关矩阵与向量的乘法。

9　利用行与列的点积，计算下列的 $A\boldsymbol{x}$？

$$(a)\begin{bmatrix} 1 & 2 & 4 \\ -2 & 3 & 1 \\ -4 & 1 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} \qquad (b)\begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$$

10　$A\boldsymbol{x}$ 视为列的组合，计算问题 9 的答案：

9(a) 变成　$A\boldsymbol{x} = 2\begin{bmatrix} 1 \\ -2 \\ -4 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + 3\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \ \\ \ \\ \ \end{bmatrix}$

当矩阵是 3×3 时，计算 $A\boldsymbol{x}$ 需要几次乘法?

11　利用行或是列，求出 $A\boldsymbol{x}$ 的两个分量。

$$\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}\begin{bmatrix} 4 \\ 2 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$

12　计算 A 乘 $\boldsymbol{x}$ 求出 $A\boldsymbol{x}$ 的三个分量。

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad 与 \quad \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

13　(a)　m 行 n 列的矩阵乘有____个分量的向量，会得到有____个分量的向量。

(b)　由 $A\boldsymbol{x} = \boldsymbol{b}$ 的 m 个方程式构成的平面会落在____维空间。A 的列的组合会在____维空间。

14　将 $2x + 3y + z + 5t = 8$ 写成矩阵 A(有几行?)乘列向量 $\boldsymbol{x} = (x, y, z, t)$产生 $\boldsymbol{b}$。请问解 $\boldsymbol{x}$ 是 4 维空间中的平面还是超平面？平面是三维的，且没有 4D 的体积。

问题 15-22 有关矩阵对于向量的特殊操作。

15　(a) 2×2 的单位矩阵为何？　I 乘 $\begin{bmatrix} x \\ y \end{bmatrix}$ 等于 $\begin{bmatrix} x \\ y \end{bmatrix}$。

(b) 2×2 的交换矩阵为何？　P 乘 $\begin{bmatrix} x \\ y \end{bmatrix}$ 等于 $\begin{bmatrix} y \\ x \end{bmatrix}$。

16 (a) 什么样的 2×2 矩阵 R 会旋转每一个向量 90°? R 乘 $\begin{bmatrix} x \\ y \end{bmatrix}$ 等于 $\begin{bmatrix} y \\ -x \end{bmatrix}$。

(b) 什么样的 2×2 矩阵 R 会旋转每一个向量 180°?

17 求矩阵 P 乘上(x, y, z)得到(y, z, x)? 求矩阵 Q 乘上(y, z, x)回到(x, y, z)?

18 什么样的 2×2 矩阵 E 会从第二分量减去第一分量? 什么样的 3×3 矩阵 E 会做同样的事情?

$$E\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad 与 \qquad E\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix}$$

19 什么样的 3×3 矩阵 E 乘上(x, y, z)得到$(y, z, z + x)$? 什么样的矩阵 E^{-1} 乘上(x, y, z)得到$(y, z, z - x)$? 如果你用 E 乘(3, 4, 5)再用 E^{-1} 乘前述结果，这两个结果是_____与 _____。

20 什么样的 2×2 矩阵 P_1 把向量(x, y)投影在 x 轴得到$(x, 0)$? 什么样的 2×2 矩阵 P_2 把向量(x, y)投影在 y 轴得到$(0, y)$? 如果你用 P_1 乘(5, 7)再用 P^{-1} 乘前述结果，你会得到_____与 _____。

21 什么样的 2×2 矩阵 R 把每个向量旋转 45°? 向量(1, 0)走到$\left(\sqrt{2}/2, \sqrt{2}/2\right)$，向量(0, 1)走到$\left(-\sqrt{2}/2, \sqrt{2}/2\right)$，这些可以决定一个矩阵。在 xy 平面画出这些特殊向量并且求出 R。

22 利用矩阵乘法 $A\boldsymbol{x}$ 的形式写出(1, 4, 5)与(x, y, z)的点积，矩阵 A 有一个行。$A\boldsymbol{x} = \boldsymbol{0}$ 的解是一个_____垂直于向量_____。A 的列只会落在____维空间。

23 以 MATLAB 的方式，写出定义矩阵 A 与列向量 $\boldsymbol{x}$ 与 $\boldsymbol{b}$ 的指令。什么样的指令可以测试是否 $A\boldsymbol{x} = \boldsymbol{b}$?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 5 \\ -2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

24 MATLAB 的指令 A = eye(3) 以及 v = [3: 5]' 可以得到 3×3 单位矩阵以及向量(3, 4, 5)。请问 A*v 以及 v'*v 的输出为何? (不需要电脑！)，如果写成 v*A 会发生什么？

25 假设 4×4 矩阵 A 的单元全部是 1 的 A = ones(4)，列向量 v = ones(4, 1)，A*v 为何? (需要电脑！) 若 B = eye(4) + ones(4)且 w = zeros(4, 1) + 2*ones(4, 1)，则 B*w 为何?

请尊重版权与译者的劳动成果，侵权必究!

问题 26 – 28 复习 2, 3, 4 维的行图与列图。

26 对于方程组 $x - 2y = 0, x + y = 6$，画出行图与列图。

27 对于三个未知数 x, y, z 的两个方程式，行图会得到在(2 或 3)–维空间中的(2 或 3) (线或平面)。列图会位于(2 或 3) –维空间。方程组的解会落在_____上。

28 对于两个未知数 x, y 的四个方程式，行图会得到 4 个_____。列图会落在_____ -维空间。除非方程式右侧向量是_____的组合，否则方程组无解。

29 从向量 $\boldsymbol{u}_0 = (1, 0)$开始，一直在**左侧**乘相同的“马可夫(Markov)矩阵”A = [.8 .3; .2 .7]，它的下三个向量是 $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$：

$$\boldsymbol{u}_1 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .8 \\ .2 \end{bmatrix} \qquad \boldsymbol{u}_2 = A\boldsymbol{u}_1 = _____ \qquad \boldsymbol{u}_3 = A\boldsymbol{u}_2 = _____$$

你注意到这四个向量有什么性质?

挑战问题

30 继续问题 29，$\boldsymbol{u}_0 = (1, 0)$开始直到 $\boldsymbol{u}_7$，也从 $\boldsymbol{v}_0 = (0, 1)$开始直到 $\boldsymbol{v}_7$。你注意到 $\boldsymbol{u}_7$ 与 $\boldsymbol{v}_7$ 有什么特质？下面是两个 MATLAB 的代码，while 以及 for，可以画出 $\boldsymbol{u}_0$ 到 $\boldsymbol{u}_7$ 以及 $\boldsymbol{v}_0$ 到 $\boldsymbol{v}_7$。你也可以使用其他语言。

```
u = [1;0]; A = [.8 .3;.2 .7]
x = u; k=[0; 7];
while size(x, 2) <=7
    u = A*u; x = [x u];
end
plot(k, x)
```

```
v = [0;1]; A = [.8 .3;.2 .7]
x = v; k=[0; 7];
for j = 1 : 7
    v = A*v; x = [x v];
end
plot(k, x)
```

这些 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 会趋向一个稳定状态向量 $\boldsymbol{s}$，猜测一下 $\boldsymbol{s}$？并且检验 $A\boldsymbol{s} = \boldsymbol{s}$。如果从 $\boldsymbol{s}$ 出发，就停在 $\boldsymbol{s}$。

31 发明一个 3×3 **魔术(magic)矩阵** M_3，它的单元是 1, 2, …, 9。所有的行与列以及对角线加起来都是 15。第一行可以是 8, 3, 4，求 M_3 乘(1, 1, 1) = ？M_4 乘(1, 1, 1, 1) = ？其中 M_4 是 4×4 矩阵，它的单元是 1, 2, …, 16。

32 假设 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 是 3×3 矩阵 A 的前两列，什么样的第三列 $\boldsymbol{w}$ 会使得矩阵为奇异？在奇异情形下，请描述 $A\boldsymbol{x} = \boldsymbol{b}$ 典型的列图以及典型的行图(随机的 $\boldsymbol{b}$)。

33 被 A 乘可以视为一种**线性变换(linear transform)**，意思是：

若 $\boldsymbol{w}$ 是 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的组合，则 $A\boldsymbol{w}$ 也是 $A\boldsymbol{u}$ 与 $A\boldsymbol{v}$ 的相同组合。

$A\boldsymbol{w} = cA\boldsymbol{u} + dA\boldsymbol{v}$ 称为**线性**，也就是**线性代数**的由来。

问题：若 $\boldsymbol{u} = (1, 0)$与 $\boldsymbol{v} = (0, 1)$，则 $A\boldsymbol{u}$ 与 $A\boldsymbol{v}$ 是 A 的列。

组合 $\boldsymbol{w} = c\boldsymbol{u} + d\boldsymbol{v}$, 若 $\boldsymbol{w} = (5, 7)$，$A\boldsymbol{w}$ 与 $A\boldsymbol{u}$ 及 $A\boldsymbol{v}$ 的关系为何?

34 由四个方程式开始：$-x_{i+1} + 2x_i - x_{i-1} = i$ (其中 $i = 1, 2, 3, 4$ 且 $x_0 = x_5 = 0$), 将方程式写成 $A\boldsymbol{x} = \boldsymbol{b}$ 的矩阵形式，你能够求解 x_1, x_2, x_3, x_4 吗？

35 一个 9×9 的**数读(Sudoku)矩阵** S，在每行与每列都有 1 到 9 的数字，而且每个 3×3 的方块也有 1 到 9 的数字。若全 1 向量 $\boldsymbol{x} = (1, ..., 1)$，求 $S\boldsymbol{x}$？

更好的问题：哪些行的交换会产生另一个数读矩阵？同理，哪些行方块的交换会得到另一个数读矩阵？

段落 2.7 会检视所有可能的行的排列(permutation)情形(重新排序)，我可以看出前三行有六种顺序，这些都可以得到数读矩阵。同理，下三行也有六种排序，最后三行也是一样。那么六个方块行的方块排列？

请尊重版权与译者的劳动成果，侵权必究!

2.2 消元法的概念

1 对于 $m = n = 3$，$\boldsymbol{Ax} = \boldsymbol{b}$ 有三个方程式与三个未知数 x_1, x_2, x_3。
2 前两个方程式是 $a_{11}x_1 + \ldots = b_1$ 与 $a_{21}x_1 + \ldots = b_2$。
3 第一方程式乘 a_{21} / a_{11}，然后从第二式减去：x_1 被**消元**。
4 角落单元 a_{11} 是第一个**主元**(pivot)，比值 a_{21} / a_{11} 是第一个**乘数**。
5 从每个剩下的方程式 i 减去 a_{i1}/a_{11} 乘第一方程式，方程式 i 会消去 x_1。
6 现在最后 $n - 1$ 个方程式包含 $n - 1$ 个未知数 $x_2, \ldots, x_n$。重复步骤消去 x_2。
7 如果零出现在主元，消元法崩溃，交换两个方程式可能可以解救。

本章说明一个系统化的方法来求解线性方程组，称为**消元法(elimination)**，在我们的 2×2 范例中可以立即看到。消元之前，x 与 y 出现在两个方程式，消元之后，第一个未知数 x 会从第二式消失得到 $8y = 8$:

$$\text{之前} \quad \begin{matrix} x - 2y = 1 \\ 3x + 2y = 11 \end{matrix} \qquad \text{之后} \quad \begin{matrix} x - 2y = 1 & \text{(第一式乘 3)} \\ 8y = 8 & \text{(相减消去 } 3x\text{)} \end{matrix}$$

新方程式 $8y = 8$ 立刻得到 $y = 1$。将 $y = 1$ 反向代入第一式得到 $x - 2 = 1$，于是 $x = 3$，求得解 $(x, y) = (3, 1)$。

消元法得到一个**上三角形系统**——就是我们的目的。非零系数 1, –2, 8 形成三角形，系统由底部往上求解。首先 $y = 1$ 然后 $x = 3$。这项快速程序称为**反向代入法**，经由消元法得到三角形后，反向代入可以应用在任何大小(size)的上三角形。

【size 指的是行与列的数目，中文可以翻译成“大小”或“尺寸”】

重点：原始方程组有相同的解 $x = 3$ 与 $y = 1$，图 2.5 各使用一对直线来表示两个系统，两个系统都交于点(3, 1)。经过消元法，这些线还是交会于同样的点，每一个步骤都得到正确的方程式。

我们如何从第一对直线得到第二对直线？我们把第一式的 3 倍从第二式减去，从第二式消去 x 的步骤是本章的基础运算。我们会经常使用，所以我们仔细观察：

消去 x：从方程式 2 减去方程式 1 的倍数

3 倍的 $x - 2y = 1$ 是 $3x - 6y = 3$，从 $3x + 2y = 11$ 减去前述方程式，右侧得到 8。重点是 $3x$ 与 $3x$ 抵消，左侧剩下 $2y - (-6y) = 8y$，x 被消去，**系统变成三角形**。

问问自己如何得到乘数 $l = 3$？第一方程式包含 $1x$，所以第一个主元是 1(x 的系数)，第二方程式包含 $3x$，所以乘数是 3。减法 $3x - 3x$ 得到三角形的零。

如果把第一式变成 $4x - 8y = 4$(一样的直线，但是第一主元变成 4)，你就会看到乘数的规则。现在正确的乘数变成 $l = 3/4$。把要消去的系数 3 除以主元 4 就得到所需的乘数：

$$\begin{array}{ll} 4x - 8y = 4 & \text{第一式乘 } 3/4 \\ 3x + 2y = 11 & \text{从第二式减去} \end{array} \qquad \begin{array}{r} 4x - 8y = 4 \\ 8y = 8 \end{array}$$

最终系统是三角形而且最后的方程式得到 $y = 1$。反向代入得到 $4x - 8 = 4$ 与 $4x = 12$ 与 $x = 3$。我们改变了数字，但是没有改变直线，也没有改变解。**除以主元得到乘数 $l = 3/4$：**

主元 = 执行消去的行的第一个非零数
乘数 = (要消去的单元) / (主元) = 3/4

新的第二式从新的第二主元开始，这个数字是 8。如果还要从第三式(如果有的话)消去 y，就会用到这个主元 8。要求解 n 个方程式，我们需要 n 个主元，在完成消元之后，**这些主元会在三角形的对角线上**。

没有读到本书之前你可能已经会求解这些方程式，这是一个很简单的问题，但是我们想要多停留一会。纵使是 2 × 2 系统，消元法也可能崩溃(breakdown)。了解可能的崩溃(无法找到整组的主元)之后，你就会了解消元法的全部程序。

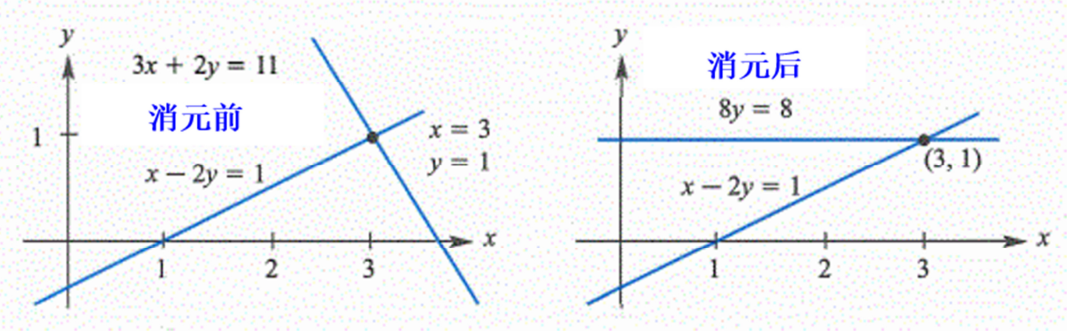


图 2.5：消去 x 使得第二条直线变成水平，则 $8y = 8, y = 1$。

消元法的崩溃

正常情况下消元法产生主元带领我们找到解，但是也有可能失败。有时候我们会遇到除以 0 的情况，这个无法执行，程序必须停止。可能有办法调整顺序再继续—或者是遇到无法避免的失败。

范例 1 因为 $0y = 8$，**无解**而失败。范例 2 因为 $0y = 0$，**太多解**而失败。范例 3 借由交换方程式而成功。

请尊重版权与译者的劳动成果，侵权必究！

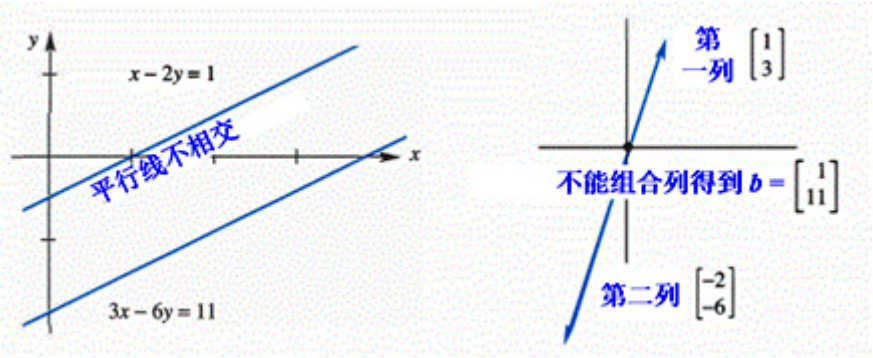


图 2.6：范例 1 的行图与列图：**无解**。

范例 1 **无解而永久失败**，消元之后变得清楚：

$$\begin{array}{ll} x-2y=1 & \text{从第二式减去} \\ 3x-6y=11 & 3\text{乘第一式} \end{array} \qquad \begin{array}{r} x-2y=1 \\ 0y=8 \end{array}$$

因为 $0y=8$ 所以无解。通常我们会让右侧的 8 除以第二个主元，但是这个系统没有第二个主元。(**零永远不允许成为主元**！) 行图与列图如图 2.6 所示，图形可以看出为什么无法避免失败。如果不存在解，消元法的过程就会出现如同 $0y=8$ 这样的方程式。

失败的行图显示两条平行线——永远无法交会，解必须同时落在两条直线上。因为没有交点，方程式无解。

列图显示两个列(1, 3)与(–2, –6)位于同一个方向，所有列的组合在一条直线上，但是右侧的列却是不同方向的(1, 11)，我们找不到列的组合可以得到正确的右侧——因此无解。

如果我们把右侧改成(1, 3)，会因为一整条直线都是解而失败。下个例子不是无解的类型，而是有无限多解。

范例 2 **无限多解而失败**，把 $\boldsymbol{b}$ 从(1, 11)变成(1, 3)。

$$\begin{array}{ll} x-2y=1 & \text{第二式减去} \\ 3x-6y=3 & 3\text{乘第一式} \end{array} \qquad \begin{array}{r} x-2y=1 \\ 0y=0 \end{array} \qquad \begin{array}{l} \text{仍然只有} \\ \text{一个主元} \end{array}$$

所有的 y 都满足 $0y=0$，实际上只有一个方程式 $x-2y=1$。未知数 y 是自由的，选取 y 之后就可以从 $x=1+2y$ 得到 x。

在行图中，平行线变成同一条线，线上的每一点都满足两个方程式，图 2.7 显示一整条线都是解的情形。

在列图中，现在 $\boldsymbol{b}=(1, 3)$与列 1 相同，所以可以选取 $x=1$ 及 $y=0$。我们也可以选取 $x=0$ 及 $y=-1/2$；列 2 乘–1/2 等于 $\boldsymbol{b}$。每一个行问题的解(x, y)，也会是列问题的解。

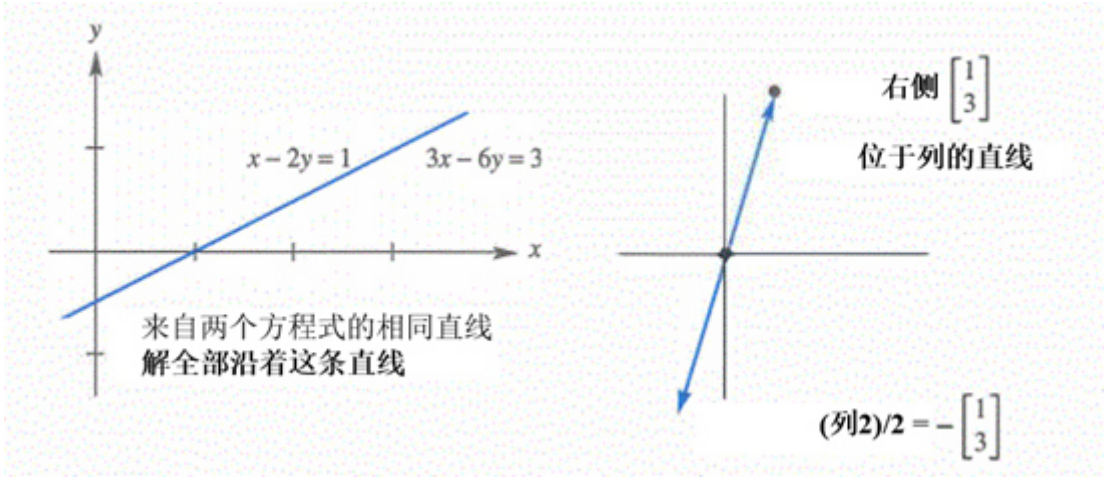


图 2.7：范例 2 的行图与列图：无限多解。

失败 n 个方程式找不到 n 个主元。

消元法得到方程式 $0 \neq 0$ (无解)，或是 $0 = 0$ (无限多解)。

得到 n 个主元代表成功，但是可能需要交换这 n 个方程式。

消元法失败的第三种形式是可以修正的。假设第一个主元位置是 0，我们拒绝允许 0 做为主元。当第一式没有 x 项时，我们可以把它与下方的方程式交换：

范例 3 **暂时性失败(主元出现 0)，交换行得到两个主元：**

$$\text{排列} \quad \begin{matrix} 0x + 2y = 4 \\ 3x - 2y = 5 \end{matrix} \quad \text{两个方程式交换} \quad \begin{matrix} 3x - 2y = 5 \\ 2y = 4 \end{matrix}$$

新系统已经是三角形，这个小例子已经可以执行反向代入。最后一个方程式得到 $y = 2$，从第一式可以得到 $x = 3$。行图是正常的(两条相交直线)，列图也是正常的(列向量不在同一个方向。) 主元 3 与 2 是正常的--只是需要一次的**行交换**。

范例 1 与 2 是**奇异**--没有第二个主元。范例 3 是**非奇异**--有完整的主元集合，以及确切一个解，奇异方程式无解或是无限多解。因为主元要当做除数，主元必须不是零。

三个方程式三个未知数

要了解高斯消元法，你必须要超越 2×2 系统，3×3 系统足够看出模式。现在方程式是方形--行数与列数相等。以下是一个特殊建构的 3×3 系统，所有的消元过程得到整数而不是分数：

请尊重版权与译者的劳动成果，侵权必究!

$$\begin{aligned} \mathbf{2}x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \tag{1}$$

步骤为何？第一主元是粗体字 **2**(左上角)，在这个主元下方要消去的是 4，第一个乘数就是比值 4/2 = 2。主元方程式乘 $l_{21} = 2$ 然后被减去，把 $4x$ 从第二式中消除：

步骤 1　从方程式 2 减去 2 乘方程式 1，得到 $y + z = 4$。

使用第一个主元也可以消去第三式的$-2x$，最快的方法是将第一式加到第三式，$2x$ 会抵消$-2x$。本书一直强调是先乘后减，而不是加，其实是因为乘数 $l_{31} = -2/2 = -1$，减去第一式的-1 倍就如同加法。

步骤 2　方程式 3 减去-1 乘方程式 1，得到 $y + 5z = 12$。新的两个方程式只与 y 及 z 有关，第二个主元(粗体字)是 1：

$$\textbf{x 已经被消去} \qquad \begin{aligned} \mathbf{1}y + 1z &= 4 \\ 1y + 5z &= 12 \end{aligned}$$

我们已经得到 2×2 系统，最后的步骤是消去 y 得到 1×1 系统：

步骤 3　从方程式 3_{new} 减去方程式 2_{new}，乘数是 1/1 = 1，得到 $4z = 8$。原始系统 $A\boldsymbol{x} = \boldsymbol{b}$ 已经变换成**上三角形** $U\boldsymbol{x} = \boldsymbol{c}$:

$$\begin{aligned} 2x + 4y - 2z &= 2 \\ 4x + 9y - 3z &= 8 \\ -2x - 3y + 7z &= 10 \end{aligned} \qquad \begin{matrix} A\boldsymbol{x} = \boldsymbol{b} \\ \text{已经变成} \\ U\boldsymbol{x} = \boldsymbol{c} \end{matrix} \qquad \begin{aligned} \mathbf{2}x + 4y - 2z &= 2 \\ \mathbf{1}y + 1z &= 4 \\ \mathbf{4}z &= 8 \end{aligned} \tag{2}$$

目的达成——从 A 到 U 完成前向消元。注意在 $\boldsymbol{U}$ **的对角线的主元 2, 1, 4**，主元 1 与 4 隐藏在原始系统，消元法使得他们出现。$U\boldsymbol{x} = \boldsymbol{c}$ 已经可以使用快速的**反向代入**法：

$(4z = 8$ 得到 $z = 2)$　　$(y + z = 4$ 得到 $y = 2)$　　(方程式 1 给出 $x = -1)$

系统的解是$(x, y, z) = (-1, 2, 2)$。行图有三个方程式构成的三个平面，这三个平面都会经过解。原始平面都是斜的，最后一个平面经过消元法之后变成水平。

列图显示列向量的组合 $A\boldsymbol{x}$ 产生右侧的 $\boldsymbol{b}$，组合的系数是$-1, 2, 2$ (解):

$$A\boldsymbol{x} = (-\mathbf{1})\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + \mathbf{2}\begin{bmatrix} 4 \\ 9 \\ -3 \end{bmatrix} + \mathbf{2}\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} \text{ 等于 } \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \boldsymbol{b} \tag{3}$$

在 $A\boldsymbol{x} = \boldsymbol{b}$ 中，数字 x, y, z 乘列 1, 2, 3，在三角 $U\boldsymbol{x} = \boldsymbol{c}$ 中也是如此。

从 A 到 U 的消元法

对于 4×4 的问题，或是 $n \times n$ 的问题，消元法的程序都是一样。以下是全部的概念，当高斯消元法成功时，一列接着一列的从 A 变成 U。

列 1.　利用第一式使得第一主元以下都变成 0。

列 2.　利用新的第二式使得第二主元以下都变成 0。

列 3 到 n.　持续进行，找到 n 个主元以及上三角矩阵 U。

$$\text{列 2 之后我们有 } \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{bmatrix}, \text{ 我们想要 } \begin{bmatrix} \boldsymbol{x} & x & x & x \\ & \boldsymbol{x} & x & x \\ & & \boldsymbol{x} & x \\ & & & \boldsymbol{x} \end{bmatrix} \tag{4}$$

前向消元法的结果是一个上三角系统，如果存在一整组 n 个主元(非零数字!)的集合，矩阵是非奇异。问题：因为主元已知，左侧的哪些 x 在消元过程中不会被改变？下面是最后一个展示原始系统 $A\boldsymbol{x} = \boldsymbol{b}$，三角系统 $U\boldsymbol{x} = \boldsymbol{c}$ 以及反向代入得到解 (x, y, z)的例子：

$$\begin{array}{rcrcl} x+y+z=6 & & x+y+z=6 & & \\ x+2y+2z=9 & \text{前向} & y+z=3 & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} & \text{反向} \\ x+2y+3z=10 & \text{前向} & z=1 & & \text{反向} \end{array}$$

所有的乘数都是 1，所有的主元都是 1，所有的平面交会于(3, 2, 1)。A 的列使用 3, 2, 1 的组合得到 $\boldsymbol{b} = (6, 9, 10)$，三角形显示 $U\boldsymbol{x} = \boldsymbol{c} = (6, 3, 1)$。

主要观念的复习

1. 线性系统($A\boldsymbol{x} = \boldsymbol{b}$)在消元之后变成**上三角**($U\boldsymbol{x} = \boldsymbol{c}$)。
2. 我们从方程式 i **减去** l_{ij} 乘方程式 j 使得(i, j)单元为 0。
3. **乘数** $l_{ij} = \dfrac{\text{第}i\text{行中要消去的单元}}{\text{第}j\text{行的枢轴}}$，**主元不能为 0**!
4. 当 0 出现在主元的位置，如果下方有非零单元，**交换行**。
5. 上三角 $U\boldsymbol{x} = \boldsymbol{c}$ 可以使用**反向代入**求解(从底部开始)。
6. 当**崩溃**一直出现，$A\boldsymbol{x} = \boldsymbol{b}$ 无解或是无限多解。

请尊重版权与译者的劳动成果，侵权必究!

已解范例

2.2A 对矩阵 A 执行消元法，第一与第二主元为何？第一步骤的乘数 l_{21} 为何(行 2 减去 l_{21} 乘行 1)?

$$A=\begin{bmatrix}1&1&0\\1&2&1\\0&1&2\end{bmatrix}\to\begin{bmatrix}1&1&0\\0&1&1\\0&1&2\end{bmatrix}\to\begin{bmatrix}1&1&0\\0&1&1\\0&0&1\end{bmatrix}=U$$

在 2, 2 的位置上出现什么数字(不是原始的 2)会使得行 2 与行 3 必须交换?

为什么左下角的乘数 $l_{31}=0$，我们要从行 3 减去 0 乘行 1?

如果把角落单元 $a_{33}=2$ 变成 $a_{33}=1$，为什么消元法会失败?

解 第一个主元是 1，乘数 $l_{21}=1/1$【原文 1, 1 有误】。当从行 2 减去 1 乘行 1 时，第二主元也是 1。如果原始中间位置的单元不是 2 而是 1，就要处理行交换。

因为 $a_{31}=0$，乘数 l_{31} 是零，某一行的开始单元是零就不需要消元。现在得到 A 是一个**带状矩阵**(band matrix)，在**中心带**之外的单元都是零。

最后一个主元也是 1，如果原始的角落单元 $a_{33}=2$ 减少 1，消元法会产生 0，**没有第三主元，消元法失败**。

2.2B 假设 A 已经是**三角形矩阵** (上三角或是下三角)，主元会在哪些位置？当 $\boldsymbol{b}$ 为任意时，什么时候 $A\boldsymbol{x}=\boldsymbol{b}$ 会确切有一个解?

解 矩阵的对角线单元就是现成的主元，这些主元全部不是零，可以成功执行消元法。当 A 是上三角时，使用**反向**代入；当 A 是下三角时，使用**前向**代入。

2.2C 利用消元法得到上三角形矩阵 U，使用反向代入求解，或者解释为什么无法执行？主元(不是 0)为何？必要时可以交换方程式的顺序。范例中两个系统唯一的差别是最后一个方程式是$-x$。

$$\textbf{成功}\quad\begin{aligned}x+y+z&=7\\x+y-z&=5\\x-y+z&=3\end{aligned}\qquad\textbf{失败}\quad\begin{aligned}x+y+z&=7\\x+y-z&=5\\-x-y+z&=3\end{aligned}$$

解 对于系统 1，方程式 2 与方程式 3 分别减去方程式 1(乘数 $l_{21}=1, l_{31}=1$,)，单元 2, 2 变成 0，所以交换方程式 2 与 3：

$$\textbf{成功}\quad\begin{aligned}x+y+z&=7\\\mathbf{0}y-2z&=-2\\-2y+0z&=-4\end{aligned}\qquad\textbf{交换后变成}\quad\begin{aligned}x+y+z&=7\\-\mathbf{2}y+0z&=-4\\-2z&=-2\end{aligned}$$

反向代入得到 $z=1$ 与 $y=2$ 与 $x=4$。主元是 1, –2, –2。

对于系统 2，一样执行方程式 2 减去方程式 1，然后方程式 1 加到方程式 3，使得 2, 2 单元变成零，2, 2 单元的下方也是零：

$$\begin{array}{lll} & x+y+z=7 & \text{列 } \mathbf{2} \text{ 没有主元(列 2 原来就是列1)} \\ \textbf{失败} & \mathbf{0}y-2z=-2 & \text{进一步消元得到} \mathbf{0}z=\mathbf{8} \\ & \mathbf{0}y+2z=10 & \text{三个平面不相交} \end{array}$$

平面 1 与 2 交于一条直线，平面 1 与 3 的交线与前面的直线平行，**无解**。

如果我们把原来第三方程式的“3”换成“–5”，消元法得到 0 = 0，有无限多解！三个平面的交集是一整条直线。

把 3 换成–5 使得第三平面与前面两个平面相交，第二方程式得到 $z=1$，第一个方程式变成 $x+y=6$。**第二列不存在主元使得 *y* 变成自由**(自由变数可以是任意数)，则 $x=6-y$。

问题集 2.2

问题 1-10 有关 2×2 系统的消元。

1　从方程式 2 减去方程式 1 所需要的乘数 l_{21} 为何？

$$\begin{aligned} 2x+3y&=1 \\ 10x+9y&=11 \end{aligned}$$

消元之后，写出上三角矩阵并且圈出两个主元。数字 1 与 11 不会影响主元——使用他们完成反向代入。

2　求解问题 1 的三角形系统，反向代入先求 y 再求 x。验证 x 乘(2, 10)加上 y 乘(3, 9)得到(1, 11)。如果右侧变成(4, 44)，新的解为何?

3　从方程式 2 减去方程式 1 所需要的乘数为何？

$$\begin{aligned} 2x-4y&=6 \\ -x+5y&=0 \end{aligned}$$

消元步骤之后，求解三角系统。如果右侧变成(–6, 0)，新的解为何?

4　从方程式 2 减去方程式 1 所需要的乘数 l 为何才能消除 c?

$$\begin{aligned} ax+by&=f \\ cx+dy&=g \end{aligned}$$

假设 a 不为 0，第一个主元就是 a。消元法得到第二个主元的公式为何？当 $ad=bc$，第二个主元消失：奇异。

请尊重版权与译者的劳动成果，侵权必究!

5 下列方程式的空白处为何值使得系统无解？空白处为何值使得系统无限多解？写出两个解。

奇异系统
$$\begin{aligned} 3x+2y&=10\\ 6x+4y&= \end{aligned}$$

6 选择系数 b 使得系统成为奇异，再选择右侧的 g 使得系统有解，写出两个解。

$$\begin{aligned} 2x+by&=16\\ 4x+8y&=g \end{aligned}$$

7 哪些数值 a 使得消元法崩溃，(1) 永久崩溃 (2) 暂时崩溃。

$$\begin{aligned} ax+3y&=-3\\ 4x+6y&=6 \end{aligned}$$

借由行的交换修正消元法的崩溃，然后求解 x 与 y。

8 哪三个数值 k 使得消元法崩溃，哪一种情形可以利用行交换修正？三种情形下，解的个数分别是 0 或 1 或 ∞？

$$\begin{aligned} kx+3y&=6\\ 3x+ky&=-6 \end{aligned}$$

9 探索 b_1 与 b_2 的条件，使得下列方程式有解。

$$\begin{aligned} 3x-2y&=b_1\\ 6x-4y&=b_2 \end{aligned}$$

10 在 xy 平面中，画出直线 $x+y=5$ 与 $x+2y=6$，经过消元之后得到方程式 $y=$ _____？若 $c=$ _____，直线 $5x-4y=c$ 会通过前述方程式的解。

问题 11-20 研究 3×3 系统的消元(可能会失败)。

11 (推荐) 线性方程组不可能恰好有两个解，为什么？

(a) 若(x, y, z)与(X, Y, Z)是两个解，其他解为何？

(b) 若 25 个平面的交集在两个点，其他交集在哪儿？

12 利用两个行运算简化系统成为上三角。

$$\begin{aligned} 2x+3y+z &= 8 \\ 4x+7y+5z &= 20 \\ -2y+2z &= 0 \end{aligned}$$

圈出主元，利用反向代入求解 z, y, x。

13 利用消元法(圈出主元)以及反向代入求解

$$\begin{aligned} 2x-3y &= 3 \\ 4x-5y+z &= 7 \\ 2x-y-3z &= 5 \end{aligned}$$

列出三个行运算：从行____减去_____乘行_____。

14 什么数值 d 会强制一个行交换？此时三角系统(非奇异)为何？什么数值 d 会使得系统奇异(没有第三个主元)?

$$\begin{aligned} 2x+5y+z &= 0 \\ 4x+dy+z &= 2 \\ y-z &= 3 \end{aligned}$$

15 什么数值 b 会产生强制的行交换？什么数值 b 会使得主元不存在？在奇异情形下找出非零解 x, y, z。

$$\begin{aligned} x+by &= 0 \\ x-2y-z &= 0 \\ y+z &= 0 \end{aligned}$$

16 (a) 建立一个 3×3 系统，需要两次行交换得到三角矩阵以及一个解。

(b) 建立一个 3×3 系统，需要一次行交换才能继续执行，但是稍后会崩溃。

17 如果行 1 与行 2 相同，你可以使用消元法到哪个地步(允许行交换)？如果列 1 与列 2 相同，哪个主元会消失?

相等行			相等列
	$2x-y+z=0$	$2x+2y+z=0$	
	$2x-y+z=0$	$4x+4y+z=0$	
	$4x+y+z=2$	$6x+6y+z=2$	

18 建立一个 3×3 范例，左侧的 9 个系数都不相同，但是行 2 与行 3 在消元法之后都变成 0。当 $\boldsymbol{b} = (1, 10, 100)$时，你的系统有多少解？当 $\boldsymbol{b} = (0, 0, 0)$时，你的系统有多少解?

请尊重版权与译者的劳动成果，侵权必究!

19 什么数字 q 使得系统奇异？什么样的右侧 t 使得系统有无限多解？当 $z=1$ 时，求解？

$$\begin{aligned} x+4y-2z&=1\\ x+7y-6z&=6\\ 3y+qz&=t \end{aligned}$$

20 纵使没有平面是平行，三个平面也可能无法交于一点。若 A 的行 3 是前两行的_____，则系统为奇异。给定方程式 $x+y+z=0$ 与 $x-2y-z=1$，求出第三个方程式使得系统无解。

21 求出以下两个系统的主元与解($A\boldsymbol{x}=\boldsymbol{b}$ 与 $K\boldsymbol{x}=\boldsymbol{b}$):

$$\begin{aligned} 2x+\ \ y\ \ &=0\\ x+2y+\ \ z\ \ &=0\\ y+2z+\ \ t\ &=0\\ z+\ 2t\ &=5 \end{aligned}\qquad\qquad \begin{aligned} 2x-\ \ y\ \ &=0\\ -x+2y-\ \ z\ \ &=0\\ -y+2z-\ \ t\ &=0\\ -z+\ 2t\ &=5 \end{aligned}$$

22 如果你延伸问题 21 的 1, 2, 1 模式，或是–1, 2, –1 的模式，第五个主元是多少？第 n 个主元是多少？K 矩阵是我最喜欢的矩阵。

23 若消元法得到 $x+y=1$ 与 $2y=3$，求三个可能的原来问题。

24 a 的哪两个数值使得消元法崩溃？其中 $A=\begin{bmatrix} a & 2\\ a & a \end{bmatrix}$。

25 a 的哪三个数值使得消元法无法得到三个主元？

$A=\begin{bmatrix} a & 2 & 3\\ a & a & 4\\ a & a & a \end{bmatrix}$，$a$ 的三个数值使得系统奇异

26 找出一个矩阵使得行的总和是 4 与 8，列的总和是 2 与 s：

$$\text{矩阵}=\begin{bmatrix} a & b\\ c & d \end{bmatrix}\qquad \begin{aligned} a+b&=4\\ c+d&=8 \end{aligned}\qquad \begin{aligned} a+c&=2\\ b+d&=s \end{aligned}$$

只有 $s=$_____，四个方程式才有解。找出两个不同的矩阵满足正确的行总和与列总和。额外积分：写下 4×4 系统 $A\boldsymbol{x}=\boldsymbol{b}$，其中 $x=(a, b, c, d)$，在消元法之后 A 变成三角形。

27 使用正常顺序的消元法得到什么矩阵 U 并且求解下列的下三角系统。事实上我们做的是前向代入：

$$\begin{aligned} 3x\ \ &=3\\ 6x+2y+\ \ &=8\\ 9x-2y+z\ &=9 \end{aligned}$$

28 若矩阵 A 已知，建立 MATLAB 指令 A(2,:) = ... 得到新的行 2，从现存的行 2 减去 3 乘行 1。

挑战问题

29 假设 MATLAB 指令为[L, U] = lu(rand(3))，找出第一主元，第二主元，第三主元的平均大小。平均大小 abs(U(1, 1))大于 1/2，这是因为 lu 会选取列 1 最大可能的主元。此处 A = rand(3, 3)具有 0 至 1 的随机数。

30 若最右下角单元为 $A(5, 5) = 11$，A 的最后一个主元为 $U(5, 5) = 4$。当 $A(5, 5)$ 是多少时，系统是奇异？

31 假设消元法使得 A 到 U，而且没有任何的行交换。U 的行 j 是 A 的哪些行的线性组合？若 $A\boldsymbol{x} = \boldsymbol{0}$，$U\boldsymbol{x}$ 会等于 $\boldsymbol{0}$？若 $A\boldsymbol{x} = \boldsymbol{b}$，$U\boldsymbol{x}$ 会等于 $\boldsymbol{b}$？若 A 是下三角矩阵，则上三角矩阵 U = ?

32 若 $A\boldsymbol{x} = \boldsymbol{0}$ 是 100 个方程式 100 个未知数的系统，其中 $\boldsymbol{x} = (x_1, ..., x_{100})$。假设消元法使得第 100 个方程式变成 0 = 0，所有系统是奇异：

(a) 消元法取得行的线性组合，所以奇异系统有以下的奇异性质：这 100 行的某一个线性组合等于_____。

(b) 奇异系统 $A\boldsymbol{x} = \boldsymbol{0}$ 有无限多解，表示这 100 列的某一个线性组合等于_____。

(c) 找出一个没有单元是零的 100×100 奇异矩阵。

(d) 针对你的矩阵用文字描述 $A\boldsymbol{x} = \boldsymbol{0}$ 的行图与列图。不需要画出 100 维空间。

请尊重版权与译者的劳动成果，侵权必究！

2.3 使用矩阵消元

1 第一步用矩阵 E_{21} 乘 $A\boldsymbol{x}=\boldsymbol{b}$ 产生 $E_{21}A\boldsymbol{x}=E_{21}\boldsymbol{b}$。

2 矩阵 $E_{21}A$ 在行 2, 列 1 有一个零，这是因为方程式 2 的 x_1 被消去。

3 把**单位矩阵**(对角线都是 1's)在行 2, 列 1 减去乘数 a_{21}/a_{11} 得到 E_{21}。

4 矩阵-矩阵乘法包含 n 个矩阵-向量乘法：$EA=[E\boldsymbol{a}_1 \ldots E\boldsymbol{a}_n]$。

5 我们也必须乘 $E\boldsymbol{b}$。所以是 E 乘**增广矩阵** $[A \quad \boldsymbol{b}]=[\boldsymbol{a}_1 \ldots \boldsymbol{a}_n \quad \boldsymbol{b}]$。

6 消元矩阵乘 $A\boldsymbol{x}=\boldsymbol{b}$ 的顺序 $E_{21}, E_{31}, \ldots, E_{n1}$，然后 $E_{32}, E_{42}, \ldots, E_{n2}$，持续下去。

7 **行交换矩阵**不是 E_{ij} 而是 P_{ij}，将 I 的第 i 行与第 j 行交换得到 P_{ij}。

本段落介绍第一个**矩阵乘法**的例子，很自然的我们会从包含很多零的矩阵开始。我们的目标是想了解矩阵能够做什么？E 作用在向量 $\boldsymbol{b}$ 或是矩阵 A 得到新的向量 $E\boldsymbol{b}$ 或是新的矩阵 EA。

我们的第一个例子是**消元矩阵**，他们可以执行消元的步骤。从第 i 个方程式减去 l_{ij} 乘第 j 个方程式 (把 x_j 从方程式 i 消除)，我们需要很多个简单矩阵 E_{ij}，每一个对应一个主对角线下方要消除的非零数字。

很幸运的，在后面章节不会看到全部的 E_{ij}，他们是开始学习时很好的例子，但是数量太多了。他们可以结合成为一个整体矩阵 E，一次性完成所有步骤。最简洁的方法是结合所有的逆矩阵$(E_{ij})^{-1}$，变成一个整体矩阵 $L = E^{-1}$。下列是后续页面的内容。

1. 了解每个步骤都是一次矩阵乘法。

2. 组合所有的步骤 E_{ij} 变成一个消元矩阵 E。

3. 了解每个 E_{ij} 如何变成逆矩阵 E_{ij}^{-1}。

4. 组合所有的逆矩阵 E_{ij}^{-1}(右序)变成 L。

L 的特殊性质是所有的乘数 l_{ij} 都会井然有序，这些数字在 E 中弄得混乱不堪(从 A 到 U 的前向消元)，在 L 中(撤销消元，反向从 U 到 A)却是非常完美。反向会让这些步骤与矩阵 E_{ij}^{-1} 落在反向序列中，并且可以防止混乱。

本段落求取矩阵 E_{ij}，段落 2.4 介绍四种矩阵相乘的方法，段落 2.5 介绍逆反操作，(我们已经在这里看到消元矩阵以及 E_{ij}^{-1}。) 这些逆反操作进入 L。

矩阵乘向量与 $Ax = b$

前一个段落介绍 3×3 矩阵时，有这样的缩写 $\boldsymbol{Ax} = \boldsymbol{b}$：

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 2 \\ 4x_1 + 9x_2 - 3x_3 &= 8 \\ -2x_1 - 3x_2 + 7x_3 &= 10 \end{aligned} \quad \text{等同于} \quad \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} \tag{1}$$

左侧的九个数字组成矩阵 A，矩阵不是静坐在 $\boldsymbol{x}$ 旁边而已，矩阵 A 还要乘 $\boldsymbol{x}$，A 乘 $\boldsymbol{x}$ 的规则要能够得到三个方程式。

A 乘 x 的复习　矩阵乘向量得到向量，当方程式的数目(3 个)与未知数的数目(3 个)相等时，矩阵是方阵，我们的矩阵是 3×3。方阵一般表示成 $n \times n$，向量 $\boldsymbol{x}$ 在 n 维空间中。

$$\textbf{未知数是} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \textbf{解是} \quad \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$$

关键点：　$\boldsymbol{Ax} = \boldsymbol{b}$ 可以代表方程式的**行形式**，也可以是**列形式**：

$$\textbf{列形式} \quad \boldsymbol{Ax} = (-1)\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} + 2\begin{bmatrix} 4 \\ 9 \\ -3 \end{bmatrix} + 2\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix} = \boldsymbol{b} \tag{2}$$

$\boldsymbol{Ax}$ 是 A 的列的线性组合，要计算 $\boldsymbol{Ax}$ 的分量时，我们使用矩阵乘法的**行形式**，$\boldsymbol{Ax}$ 的分量就是 $\boldsymbol{x}$ 与(A 的行)的点积。与 $\boldsymbol{x}$ 点积的短公式使用“累加表示”：

$\boldsymbol{Ax}$ 的第一分量为 $(-1)\,(2) + (2)\,(4) + (2)\,(-2)$。

$\boldsymbol{Ax}$ 的第 i 分量为 $(\textbf{行 } \boldsymbol{i}) \cdot \boldsymbol{x} = a_{i1}\,x_1 + a_{i2}\,x_2 + \ldots + a_{in}\,x_n$。

有时候写成累加符号的形式：　$\sum_{j=1}^{n} a_{ij} x_j$

Σ的意思是累加，从 $j = 1$ 开始直到 $j = n$ 停止，总和从 $a_{i1}\,x_1$ 开始一直到 $a_{in}\,x_n$，可以得到点积(行 i) $\cdot\, \boldsymbol{x}$。

重复说明矩阵的表示法：行 1, 列 1 的单元(左上角)是 a_{11}，行 1, 列 3 的单元是 a_{13}，行 3, 列 1 的单元是 a_{31}，(行数目在前，列数目在后)。单元这个词相对于矩阵就如同分量相对于向量。一般规则：$\boldsymbol{a_{ij} = A(i, j)}$，**位置在行 $\boldsymbol{i}$，列 $\boldsymbol{j}$**。

范例 1　矩阵的 $a_{ij} = 2i + j$，则 $a_{11} = 3$，$a_{12} = 4$，$a_{21} = 5$。以下是由行得到 $\boldsymbol{Ax}$，分别以数字与字母表示：

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 4 \cdot 1 \\ 5 \cdot 2 + 6 \cdot 1 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$$

一行乘一列得到一个点积。

---

[1] 爱因斯坦删去Σ得到更简短的形式，$a_{ij}\,x_j$ 的重复 j 自动代表总和。他也写过 $a_i^j x_j$。我们不是爱因斯坦，我们使用Σ。

请尊重版权与译者的劳动成果，侵权必究!

一个消元步骤的矩阵形式

$A\boldsymbol{x}=\boldsymbol{b}$ 是原始方程式的便利形式，消元步骤又是如何？本范例中，从第二式减去 2 乘第一式，在右侧 $\boldsymbol{b}$ 的第二分量减去第一分量的 2 倍：

$$\textbf{第一步}\quad \boldsymbol{b}=\begin{bmatrix}2\\ \mathbf{8}\\ 10\end{bmatrix}\quad \text{变成}\quad \boldsymbol{b}_{\text{new}}=\begin{bmatrix}2\\ \mathbf{4}\\ 10\end{bmatrix}$$

我们想用矩阵完成刚才的动作！目的在找到**消元矩阵** E 乘 $\boldsymbol{b}$ 得到相同的效果 $\boldsymbol{b}_{\text{new}}=E\boldsymbol{b}$，也就是从 b_2 减去 $2b_1$：

$$\textbf{消元矩阵}\qquad \boldsymbol{E}=\begin{bmatrix}1&0&0\\ -2&1&0\\ 0&0&1\end{bmatrix}$$

用 E 乘的效果是从行 2 减去 2 乘行 1，行 1 与行 3 保持不变:

$$\begin{bmatrix}1&0&0\\ \mathbf{-2}&\mathbf{1}&0\\ 0&0&1\end{bmatrix}\begin{bmatrix}\mathbf{2}\\ \mathbf{8}\\ 10\end{bmatrix}=\begin{bmatrix}2\\ \mathbf{4}\\ 10\end{bmatrix}\qquad \begin{bmatrix}1&0&0\\ \mathbf{-2}&\mathbf{1}&0\\ 0&0&1\end{bmatrix}\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\\ b_3\end{bmatrix}=\begin{bmatrix}b_1\\ \boldsymbol{b}_2-2\boldsymbol{b}_1\\ b_3\end{bmatrix}$$

E 的第一与第三行来自单位矩阵 I，没有改变第一与第三个的数字(2 与 10)，新的第二分量在消元步骤之后变成 4，就是 b_2-2b_1。

很容易描述“基础矩阵”或是“消元矩阵”的 E，从单位矩阵 I 开始，把其中的一个 0 变成乘数 $-l$:

单位矩阵在对角线的单元都是 1，其余单元都是 0。对于所有的 $\boldsymbol{b}$ 存在 $I\boldsymbol{b}=\boldsymbol{b}$。“**基础矩阵**”或是“**消元矩阵**” E_{ij} 在位置 i,j 多了一个非零单元 $\boldsymbol{-l}$，则 E_{ij} 会得到从行 i 减去行 j 的 l 倍。

范例 2　矩阵 E_{31} 在位置 3, 1 是 $-l$：

$$\text{单位矩阵}\quad I=\begin{bmatrix}1&0&0\\ 0&1&0\\ 0&0&1\end{bmatrix}\qquad \text{消元矩阵}\quad E_{31}=\begin{bmatrix}1&0&0\\ 0&1&0\\ \boldsymbol{-l}&0&1\end{bmatrix}$$

当你用 I 乘 $\boldsymbol{b}$ 得到 $\boldsymbol{b}$，但是当 E_{31} 乘 $\boldsymbol{b}$ 得到第三分量减去 l 乘第一分量。当 $l=4$，这个例子给出 $9-4=5$：　　【下方原文的 E 是错误】

$$I\boldsymbol{b}=\begin{bmatrix}1&0&0\\ 0&1&0\\ 0&0&1\end{bmatrix}\begin{bmatrix}1\\ 3\\ 9\end{bmatrix}=\begin{bmatrix}1\\ 3\\ 9\end{bmatrix}\qquad \boldsymbol{E_{31}b}=\begin{bmatrix}1&0&0\\ 0&1&0\\ \mathbf{-4}&0&1\end{bmatrix}\begin{bmatrix}1\\ 3\\ \mathbf{9}\end{bmatrix}=\begin{bmatrix}1\\ 3\\ \mathbf{5}\end{bmatrix}$$

$A\boldsymbol{x} = \boldsymbol{b}$ 的左侧会怎么样？两侧都会被 E_{31} 乘，E_{31} 的**目的是在矩阵的(3, 1)位置产生一个 0。**

表示式满足它的目的。从 A 开始，应用数个 E's 使得在主元下方的位置产生零(第一个 E 是 E_{21})，最终得到三角形 U。现在我们关注详细的步骤。

从小处开始，向量 $\boldsymbol{x}$ 不变。解 $\boldsymbol{x}$ 不会因为消元法而改变(这就不是小处了！)，只有系数矩阵会改变。当我们从 $A\boldsymbol{x} = \boldsymbol{b}$ 开始，然后被 E 乘，结果是 $EA\boldsymbol{x} = E\boldsymbol{b}$，新的矩阵 EA 是 E 乘 A 的结果。

坦白　消元矩阵 E_{ij} 是伟大的例子，但是后面就见不到了。他们展示矩阵如何影响行。经过几次消元之后，我们了解如何乘矩阵(这些 E 矩阵的顺序变得非常重要。)**乘积与逆矩阵**对于 E 来说特别清晰。以下是本书会看到的两个观念。

矩阵乘法

最大的问题：**两个矩阵如何乘**？当第一个矩阵是 E，我们预期的是 EA，这个特殊的 E 使得 A 从行 2 减去 2 乘行 1，乘数为 2：

$$EA = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-2} & \mathbf{1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{4} & \mathbf{-2} \\ \mathbf{4} & \mathbf{9} & \mathbf{-3} \\ -2 & -3 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -2 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \\ -2 & -3 & 7 \end{bmatrix} \quad \text{(得到一个零)} \tag{3}$$

这个步骤没有影响 A 的行 1 与行 3，这些行在 EA 中保持不变，只有行 2 不同，行 2 已经减去行 1 的 2 倍。矩阵乘法与消元法一致——新系统变成 $EA\boldsymbol{x} = E\boldsymbol{b}$。

$EA\boldsymbol{x}$ 很简单但是有一个巧妙的概念。从 $A\boldsymbol{x} = \boldsymbol{b}$ 开始，两侧同时用 E 乘给出 $E(A\boldsymbol{x}) = E\boldsymbol{b}$。使用矩阵乘法，这也是 $(EA)\boldsymbol{x} = E\boldsymbol{b}$。

第一个是 E 乘 $A\boldsymbol{x}$，第二个是 EA 乘 $\boldsymbol{x}$，两者相同。

括号就不需要了，我们直接写成 $EA\boldsymbol{x}$。

这个规则扩展至具有数个列向量的矩阵 C，计算 EAC 时可以先做 AC，也可以先做 EA，这就是所谓的**结合律**，好像 $3\times(4\times5) = (3\times4)\times5$ 一样。3 乘 20 或者 12 乘 5，两者的答案都是 60。这个定律看起来很清楚，很难想象它也有可能错误。

交换律 $3\times4 = 4\times3$ 看起来更明显，但是 EA 通常不会等于 AE。当 E 在右乘，它作用的是 A 的列而不是行。AE 会使得列 1 减去 2 乘列 2，所以 $EA \neq AE$。

结合律为真	$\boldsymbol{A(BC) = (AB)C}$
交换律为假	**通常** $\boldsymbol{AB \neq BA}$

请尊重版权与译者的劳动成果，侵权必究！

矩阵乘法还有另一个要求，假设 B 只有一列(就是列 $\boldsymbol{b}$)，针对 EB 的矩阵-矩阵法则必须与矩阵-向量的法则一致，甚至我们能够使得矩阵乘法 EB 可以一次乘一个列：

若 B 具有数个列 b_1, b_2, b_3，则 EB 的列是 Eb_1, Eb_2, Eb_3。

$$\textbf{矩阵乘法} \qquad \boldsymbol{EB = E[b_1, b_2, b_3] = [Eb_1, Eb_2, Eb_3]} \tag{4}$$

【依据上下文关系，原文的 A 应该是 E 才合理】

对于(3)的矩阵乘法来说，这项性质也适用。如果你用 E 乘 A 的列 3，你会正确得到 EA 的列 3：

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 7 \end{bmatrix} \qquad E(A \text{ 的列 } j) = EA \text{ 的列 } j$$

当消元法应用到行时，这个要求在于**处理列**。下个段落会介绍每个乘积 AB 的每个单元，矩阵乘法的美在于三种处理方法(行，列，整个矩阵)都有正确的结果。

行交换矩阵 P_{ij}

我们使用 E_{ij} 从行 i 减去行 j。要交换或是排列这些行，我们使用另一种矩阵 P_{ij} (排列矩阵)。当零出现在主元位置时，我们要做交换行的动作。往下方看，主元这一列可能包含一个非零数，两行交换就有了主元，消元法可以继续执行。

什么样的矩阵 P_{23} 可以交换行 2 与行 3？我们可以把单位矩阵的行 2 与行 3 交换，就得到 P_{23}：

$$\textbf{排列矩阵} \qquad P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

这是**行交换矩阵**。P_{23} 乘任意列向量，会使得向量的第二分量与第三分量交换，因此也可以交换矩阵的行 2 与行 3：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ \mathbf{3} \\ \mathbf{5} \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{5} \\ \mathbf{3} \end{bmatrix} \qquad \text{与} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & 4 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{3} \\ 0 & 6 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 1 \\ 0 & 6 & 5 \\ \mathbf{0} & \mathbf{0} & \mathbf{3} \end{bmatrix}$$

在右侧，P_{23} 做了该做的事情。因为 0 出现在第二主元的位置而且下方有个“6”，交换后 6 出现在主元的位置。

矩阵是有作用的，不是枯坐在旁边。我们很快会看到其他排列矩阵，这些矩阵可以改变行的顺序。行 1, 2, 3 可以变成 3, 1, 2，刚才的 P_{23} 是一个特殊排列矩阵——它交换行 2 与 3。

行交换矩阵　单位矩阵的行 i 与 j 反序得到 P_{ij}。当**排列矩阵** P_{ij} 乘一个矩阵时，它交换行 i 与行 j。

$$\textbf{要交换方程式 1 与 3，左乘 } P_{13}=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}$$

通常行交换不太用得到，很好的机会下用消元矩阵 E_{ij} 已经足够。有时候为了把主元移动至对角线，P_{ij} 还是可以用的。

增广矩阵

本书最终会离开消元法很远，矩阵相乘有很多实际的应用，最佳的出发点是方阵 E 乘方阵 A。我们在介绍消元法的时候已经见过，我们也知道 EA 的预期答案是什么。下个步骤允许出现矩形(长方形)矩阵，他们还是一样来自原始方程式，只是现在多加了 $\boldsymbol{b}$。

关键观念：消元法对于 A 与 $\boldsymbol{b}$ 有相同的行运算，我们可以把 $\boldsymbol{b}$ 引入一个**额外列**，然后一样执行消元法。加入 $\boldsymbol{b}$ 之后矩阵 A 变大了，或是增广了(augmented)：

增广矩阵　$\begin{bmatrix}A&\boldsymbol{b}\end{bmatrix}=\begin{bmatrix}2&4&-2&\mathbf{2}\\4&9&-3&\mathbf{8}\\-2&-3&7&\mathbf{10}\end{bmatrix}$

消元法作用在矩阵的**整个行**，用 E 同时乘左侧与右侧，得到方程式 2 减去 2 乘方程式 1，消元步骤同时作用在$[A\quad \boldsymbol{b}]$：

$$\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&4&-2&\mathbf{2}\\4&9&-3&\mathbf{8}\\-2&-3&7&\mathbf{10}\end{bmatrix}=\begin{bmatrix}2&4&-2&\mathbf{2}\\0&1&1&\mathbf{4}\\-2&-3&7&\mathbf{10}\end{bmatrix}$$

新的第二行包含 0, 1, 1, 4，新的第二方程式是 $x_2+x_3=4$。行与列同时参与矩阵乘法的工作：

行：　E 的每一行作用在$[A\quad \boldsymbol{b}]$得到$[EA\quad E\boldsymbol{b}]$的一个行。

列：　E 作用在$[A\quad \boldsymbol{b}]$的每一列得到$[EA\quad E\boldsymbol{b}]$的列。

请尊重版权与译者的劳动成果，侵权必究!

再次注意这个字“作用(act)”，这个很重要，矩阵的确做了事情！矩阵 A 作用在 $\boldsymbol{x}$ 得到 $\boldsymbol{b}$，矩阵 E 作用在 A 得到 EA。消元的整体过程就是一系列的行运算，也就是矩阵乘法。A 走到 $E_{21}A$，然后是 $E_{31}E_{21}A$，最终 $E_{32}E_{31}E_{21}A$ 是个三角矩阵。

方程式右侧被引入到增广矩阵，最后的结果是方程式的三角系统。在写下所有的矩阵乘法规则前(包含方块计算)，我们将停止左乘 E 的练习。

主要观念的复习

1. $A\boldsymbol{x}=(x_1$ 乘列 1) + ... + (x_n 乘列 n)，$(A\boldsymbol{x})_i=\sum_{j=1}^{n}a_{ij}x_j$。
2. 单位矩阵= I，消元矩阵= E_{ij} 使用 l_{ij}，交换矩阵= P_{ij}。
3. 用 E_{21} 乘 $A\boldsymbol{x}=\boldsymbol{b}$，得到方程式 2 减去 l_{21} 乘方程式 1。数字$-l_{21}$ 是消元矩阵 E_{21} 在位置(2, 1)的单元。
4. 对于增广矩阵$[A\quad \boldsymbol{b}]$，消元步骤得到$[E_{21}A\quad E_{21}\boldsymbol{b}]$。
5. 当 A 乘任意矩阵 B，等于 A 分别乘 B 的每一列。

已解范例

2.3A 什么样的 3×3 矩阵 E_{21} 使得第二行减去 4 乘第一行？什么样的矩阵 P_{32} 使得第二行与第三行交换？如果对 A 进行右乘而不是原来的左乘，请描述 AE_{21} 与 AP_{32} 的结果？

解 对于单位矩阵 I 执行所需的运算，得到

$$E_{21}=\begin{bmatrix}1&0&0\\-4&1&0\\0&0&1\end{bmatrix}\qquad P_{32}=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}$$

E_{21} 乘在右侧得到从列 1 减去 4 乘列 2，P_{32} 乘在右侧得到列 2 与列 3 交换。

2.3B 引入额外的列，写下系统的增广矩阵$[A\quad \boldsymbol{b}]$:

$$\begin{aligned}x+2y+2z&=1\\4x+8y+9z&=3\\3y+2z&=1\end{aligned}$$

应用 E_{21}，再应用 P_{32} 可以得到三角系统。应用反向代入求解。组合矩阵 $P_{32}\ E_{21}$ 可以同时完成什么工作？

解 E_{21} 移除列 1 的 4，但是 0 出现在列 2：

$$[A\ \ \boldsymbol{b}]=\begin{bmatrix}1&2&2&1\\ \mathbf{4}&8&9&3\\ 0&3&2&1\end{bmatrix}\qquad E_{21}[A\ \ \boldsymbol{b}]=\begin{bmatrix}1&2&2&1\\ \mathbf{0}&\mathbf{0}&1&-1\\ 0&3&2&1\end{bmatrix}$$

现在 P_{32} 交换行 2 与 3，反向代入产生 z 然后 y 与 x。

$$P_{32}E_{21}[A\ \ \boldsymbol{b}]=\begin{bmatrix}1&2&2&1\\ 0&3&2&1\\ 0&0&1&-1\end{bmatrix}\qquad \begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}1\\ 1\\ -1\end{bmatrix}$$

组合矩阵 $P_{32}E_{21}$ 可以同时执行两个步骤，应用 P_{32} 到 E_{21}。

一个矩阵，两个步骤　　$P_{32}E_{21}=$交换 E_{21} 的行$=\begin{bmatrix}1&0&0\\ 0&0&1\\ -4&1&0\end{bmatrix}$

2.3C　矩阵乘法有两种方式，第一，A 的行乘 B 的列。第二，A 的列乘 B 的行，这个不寻常的方法产生两个矩阵，相加之后得到 AB。两个方法各需要多少次正常的乘法?

两种方法　　$AB=\begin{bmatrix}3&4\\ 1&5\\ 2&0\end{bmatrix}\begin{bmatrix}2&4\\ 1&1\end{bmatrix}=\begin{bmatrix}10&16\\ 7&9\\ 4&8\end{bmatrix}$

解　(A 的行)乘(B 的列)是向量的点积：

$$(\text{行 }1)\cdot(\text{列 }1)=\begin{bmatrix}3&4\end{bmatrix}\begin{bmatrix}2\\ 1\end{bmatrix}=10\quad \text{是 } AB \text{ 的(1, 1)单元}$$

$$(\text{行 }2)\cdot(\text{列 }1)=\begin{bmatrix}1&5\end{bmatrix}\begin{bmatrix}2\\ 1\end{bmatrix}=7\quad \text{是 } AB \text{ 的(2, 1)单元}$$

我们需要 6 个点积，每个有 2 个乘法，全部是(3·2·2) = 12。同样的 AB 来自 A 的列乘 B 的行，一个列乘一个行得到一个矩阵：

$$AB=\begin{bmatrix}3\\ 1\\ 2\end{bmatrix}\begin{bmatrix}2&4\end{bmatrix}+\begin{bmatrix}4\\ 5\\ 0\end{bmatrix}\begin{bmatrix}1&1\end{bmatrix}=\begin{bmatrix}6&12\\ 2&4\\ 4&8\end{bmatrix}+\begin{bmatrix}4&4\\ 5&5\\ 0&0\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

问题集 2.3

问题 1-15 有关消元矩阵。

1 写下产生下列消元步骤的 3×3 矩阵：

(a) E_{21} 从行 2 减去 5 乘行 1。

(b) E_{32} 从行 3 减去 -7 乘行 2。

(c) P 交换行 1 与 2，然后交换行 2 与 3。

2 问题 1 中，应用 E_{21} 然后 E_{32} 至 $\boldsymbol{b} = (1, 0, 0)$得到 $E_{32}E_{21}\,\boldsymbol{b}$ = _____。在 E_{21} 之前应用 E_{32} 得到 $E_{21}E_{32}\boldsymbol{b}$ = _____。当 E_{32} 先使用时，行___无法感受来自行_____的效应。

3 哪三个矩阵 E_{21}，E_{31}，E_{32} 使得 A 变成三角形 U?

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 4 & 6 & 1 \\ -2 & 2 & 0 \end{bmatrix} \quad \text{且} \quad E_{32}\,E_{31}\,E_{21}\,A = U$$

这些 E's 矩阵乘在一起得到整体消元矩阵 M：$MA = U$。

4 引入 $\boldsymbol{b} = (1, 0, 0)$作为问题 3 的第四列得到$[A \quad \boldsymbol{b}]$，应用消元步骤至增广矩阵，求解 $A\boldsymbol{x} = \boldsymbol{b}$。

5 假设 $a_{33} = 7$ 且第三主元是 5，如果你把 a_{33} 改成 11，第三主元变成____？如果你把 a_{33} 改成_____，就不存在第三主元。

6 如果 A 的每一列都是(1, 1, 1)的倍数，则 $A\boldsymbol{x}$ 永远是(1, 1, 1)的倍数。制造一个 3×3 范例，消元步骤会得到几个主元?

7 假设 E 从行 3 减去 7 乘行 1。

(a) 想要逆反这个步骤，你必须 7 乘行____，然后_____到行_____。

(b) 什么样的逆矩阵 E^{-1} 会执行这个逆反步骤(所以 $E^{-1}E = I$)？

(c) 如果先应用逆反步骤(然后才是 E)，证明 $EE^{-1} = I$ 。

8 $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的行列式是 $\det M = ad - bc$。从行 2 减去 l 乘行 1 得到新的 M^*，证明任意 l，$\det M^* = \det M$。当 $l = c/a$，主元乘积等于行列式：$(a)\,(d - lb)$等于 $ad - bc$ 。

9 (a) E_{21} 从行 2 减去行 1，然后 P_{23} 交换行 2 与 3。什么样的矩阵 $M = P_{23}\,E_{21}$ 同时执行这两个步骤?

(b) P_{23} 交换行 2 与 3，然后 E_{31} 从行 3 减去行 1，什么样的矩阵 $M = E_{31}P_{23}$ 同时执行这两个步骤 ？说明为什么 E's 不一样，M 却是相同?

10 (a) 什么样的 3×3 矩阵 E_{13} 会把行 3 加到行 1?

(b) 什么样的矩阵把行 1 加到行 3，同时行 3 加到行 1?

(c) 什么样的矩阵把行 1 加到行 3，然后行 3 加到行 1?

11 创建一个矩阵使得 $a_{11} = a_{22} = a_{33} = 1$，但是消元法在没有行交换的情形下，产生两个负的主元。(第一主元是 1。)

12 矩阵相乘：

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 4 & 0 \end{bmatrix}$$

13 解释下列事实。如果 B 的第三列全部是零，则 EB 的第三列也全部是零(任意的 E)。如果 B 的第三行全部是零，则 EB 的第三行可能不是零。

14 下列的 4×4 矩阵需要 E_{21} 与 E_{32} 与 E_{43} 的消元矩阵，这些矩阵为何?

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

15 写下 3×3 矩阵，其中 $a_{ij} = 2i - 3j$。矩阵的 $a_{32} = 0$，但是消元法还是需要 E_{32} 使得 3, 2 位置得到 0。前面哪个步骤破坏了原始的 0，且 E_{32} 为何？

问题 16-23 有关建立矩阵与乘矩阵。

16 将这些古老的问题写成 2×2 矩阵形式 $A\boldsymbol{x} = \boldsymbol{b}$ 并求解。

(a) X 的年纪是 Y 的两倍，两者的年纪相加等于 33。

(b) $(x, y) = (2, 5)$与$(3, 7)$在直线 $y = mx + c$ 上，求 m 与 c。

17 抛物线 $y = a + bx + cx^2$ 会通过$(x, y) = (1, 4)$与$(2, 8)$与$(3, 14)$。找出矩阵方程式并求解未知数(a, b, c)。

18 分别以顺序 EF 及 FE 执行矩阵乘法：

$$E = \begin{bmatrix} 0 & 0 & 1 \\ a & 1 & 0 \\ b & 0 & 0 \end{bmatrix} \qquad F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}$$

同时计算 $E^2 = EE$ 与 $F^3 = FFF$，你可以猜猜 F^{100}。

请尊重版权与译者的劳动成果，侵权必究!

19 分别以顺序 PQ 与 QP 与 P^2 计算行交换矩阵的乘积：

$$P=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\qquad Q=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}$$

找出另一个非对角矩阵 M，使得 $M^2=I$。

20 (a) 假设 B 的所有的列都相同，则 EB 的所有的列也相同。这是因为每个都是 E 乘_____。

(b) 假设 B 所有的行都是[1　2　4]，以例子说明 EB 所有的行都不是[1　2　4]。所有的行是_____是真实的。

21 如果 E 把行 1 加到行 2 以及 F 把行 2 加到行 1，请问 $EF=FE$？

22 A 的单元是 a_{ij}，$\boldsymbol{x}$ 的单元是 x_j，所以 $A\boldsymbol{x}$ 的第一分量是$\Sigma\, a_{1j}x_j=a_{11}x_1+\ldots+a_{1n}x_n$。若 E_{21} 从行 2 减去行 1，针对下列写出公式：

(a) $A\boldsymbol{x}$ 的第三分量。

(b) $E_{21}A$ 的(2, 1)单元。

(c) E_{21} ($E_{21}A$)的(2, 1)单元。

(d) $E_{21}A\boldsymbol{x}$ 的第一分量。

23 消元矩阵 $E=\begin{bmatrix}1&0\\-2&1\end{bmatrix}$从 A 的行 2 减去 2 乘 A 的行 1，结果是 EA。请问 $E(EA)$ 的效应是什么？对于反序 AE 来说，我们是从_____减去 2 乘 A 的_____(以例子说明。)

问题 24-27 列 $\boldsymbol{b}$ 引入增广矩阵[A　$\boldsymbol{b}$]。

24 应用消元法到 2×3 的增广矩阵[A　$\boldsymbol{b}$]，会得到什么样的三角系统 $U\boldsymbol{x}=\boldsymbol{c}$？解 $\boldsymbol{x}$ 为何？

$$A\boldsymbol{x}=\begin{bmatrix}2&3\\4&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\17\end{bmatrix}$$

25 应用消元法到 3×4 的增广矩阵[A　$\boldsymbol{b}$]，你如何知道系统无解？将最后的数字 6 改变使得存在一个解。

$$A\boldsymbol{x}=\begin{bmatrix}1&2&3\\2&3&4\\3&5&7\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}1\\2\\6\end{bmatrix}$$

26 方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 与 $A\boldsymbol{x}^*=\boldsymbol{b}^*$有相同的矩阵 A，如果想用消元法同时求解两个方程式，你会使用怎样的双重增广矩阵？在 2×4 矩阵上运算，同时求解下列方程式：

$$\begin{bmatrix}1&2\\4&7\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}\ \text{与}\ \begin{bmatrix}1&4\\2&7\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}$$

27 选择增广矩阵中的数字 a, b, c, d 使得(a) 无解 (b) 无限多解。

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & a \\ 0 & 4 & 5 & b \\ 0 & 0 & d & c \end{bmatrix}$$

a, b, c, d 中哪个数字不影响有没有解?

28 若 $AB = I$ 且 $BC = I$，利用结合律证明 $A = C$。

挑战问题

29 求出三角矩阵 E 使得**帕斯卡(Pascal)矩阵**简化到较小的帕斯卡：

对列 1 消元 $$E\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{bmatrix}$$

哪个矩阵 M (等于数个 E's 相乘)将帕斯卡一路简化到 I？帕斯卡三角矩阵是例外，所有的乘数都是 $l_{ij} = 1$。

30 $M = \begin{bmatrix} 3 & 4 \\ 5 & 7 \end{bmatrix}$ 是许多因子 $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ 与 $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ 的乘积。

(a) 什么矩阵 E 从行 2 减去行 1，使得 EM 的行 2 变小?

(b) 什么矩阵 F 从行 1 减去 EM 的行 2 来简化 FEM 的行 1?

(c) 持续 E's 与 F's，直到(许多 E's 与 F's)乘(M)等于(A 或 B)。

(d) E 与 F 是 A 与 B 的逆矩阵！把所有的 E's 与 F's 移动到右侧，可以得到想要的结果 ***M* = *A*'s 与 *B*'s 的乘积**。

对于整数矩阵 M 是可能的，其中 $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} > 0$ 而且 $ad - bc = 1$。

31 求消元矩阵 E_{21} 然后 E_{32} 然后 E_{43} 使得 K 变成 U：

$$E_{43}E_{32}E_{21}\begin{bmatrix} 1 & 0 & 0 & 0 \\ -a & 1 & 0 & 0 \\ 0 & -b & 1 & 0 \\ 0 & 0 & -c & 1 \end{bmatrix} = I$$

应用这三个步骤到单位矩阵 *I*，求 $E_{43}E_{32}E_{21}$ 的乘积。

请尊重版权与译者的劳动成果，侵权必究!

2.4 矩阵运算规则

1 有 n 个列的矩阵 A 乘有 n 个行的矩阵 B：$\boldsymbol{A_{m\times n}B_{n\times p}=C_{m\times p}}$。

2 $AB=C$ 的每个单元都是一个点积：$C_{ij}=(A$ 的行 $i)\cdot(B$ 的列 $j)$。

3 规则设定 AB 乘 C 等于 A 乘 BC，且$(AB)\boldsymbol{x}=A(B\boldsymbol{x})$。

4 更多计算 AB 的方法：(A 乘 $\boldsymbol{B}$ **的列**)，(***A*** **的行**乘 B)，(列乘行)。

5 $AB=BA$ 通常不会成立，大部分情形下 A 与 B 不能交换。

6 矩阵也可以方块相乘：$A=[A_1 \quad A_2]$乘 $B=\begin{bmatrix}B_1\\B_2\end{bmatrix}=A_1B_1+A_2B_2$。

我从基础事实开始，矩阵是数字或是单元的矩形阵列。当 A 有 m 行 n 列，它是一个 $m\times n$ 的矩阵。如果矩阵的形状一样，矩阵可以相加，矩阵可以乘上任意常数 c。以下是 3×2 矩阵的例子，包含 $A+B$，$2A$：

$$\begin{bmatrix}1&2\\3&4\\0&0\end{bmatrix}+\begin{bmatrix}2&2\\4&4\\9&9\end{bmatrix}=\begin{bmatrix}3&4\\7&8\\9&9\end{bmatrix} \quad \text{以及} \quad 2\begin{bmatrix}1&2\\3&4\\0&0\end{bmatrix}=\begin{bmatrix}2&4\\6&8\\0&0\end{bmatrix}$$

矩阵加法与向量加法相同，一次处理一个单元。我们也可以把列向量视为只有一列的矩阵($n=1$)。矩阵的负数($-A$)视为矩阵乘 $c=-1$(全部单元的符号都相反)。A 加上 $-A$ 得到零矩阵，全部单元都是 0。这些都只是常识而已。

行 *i* 列 *j* 的单元称为 $\boldsymbol{a_{ij}}$ 或 $\boldsymbol{A(i,j)}$，第一行的 n 个单元是 $a_{11}, a_{12}, \ldots, a_{1n}$。左下角的单元为 a_{m1}，右下角的单元为 a_{mn}。行的数字 i 从 1 到 m，列的数字 j 从 1 到 n。

矩阵加法很容易，比较严重的问题是乘法。当 A 与 B 可乘时，乘积 AB 是多少？本段落给定 4 种方法求 AB，但是当 A 与 B 都是 3×2 时，我们无法计算 AB，他们无法通过下列的测试：

***AB* 可乘：若 *A* 有 *n* 列，*B* 必须有 *n* 行**

当 A 是 3×2 时，B 可以是 2×1(向量)或是 2×2(方阵)或是 2×20，A 乘 $\boldsymbol{B}$ **的每一列**。我想从点积开始矩阵的乘法，再回到刚才的列方法：A 乘 B 的列。这两种方法都会遵守下列规则：

矩阵乘法的基础法则 AB 乘 C 等于 A 乘 BC (1)

括号在$(AB)C=A(BC)$可以安全的移动，线性代数依靠这个法则运行。

假设 A 是 $m\times n$，B 是 $n\times p$，我们可以乘，乘积 AB 是 $m\times p$。

$$(m\times n)\times(n\times p)=(m\times p)\qquad \begin{bmatrix} m\text{行} \\ n\text{列}\end{bmatrix}\begin{bmatrix} n\text{行} \\ p\text{列}\end{bmatrix}=\begin{bmatrix} m\text{行} \\ p\text{列}\end{bmatrix}$$

一个行乘一个列是极端的例子，也就是 $1\times n$ 乘 $n\times 1$，结果是 1×1，这个单一数字就是点积。

任何情况下 AB 都是点积，AB 的左上角(1, 1)单元是**(A 的行 1) · (B 的列 1)**，这是第一种常见计算矩阵乘法的方法，**计算 A 的每一行与 B 的每一列的点积**。

1. AB 的行 i 列 j 的单元是 (A 的行 i) · (B 的列 j)

图 2.8 选出 4×5 矩阵 A 的第二行($i=2$)，也选出 5×6 矩阵 B 的第三列($j=3$)，他们的点积位于 AB 的行 2 列 3。矩阵 AB 的行个数与 A(4 行)相同，列个数与 B 相同。

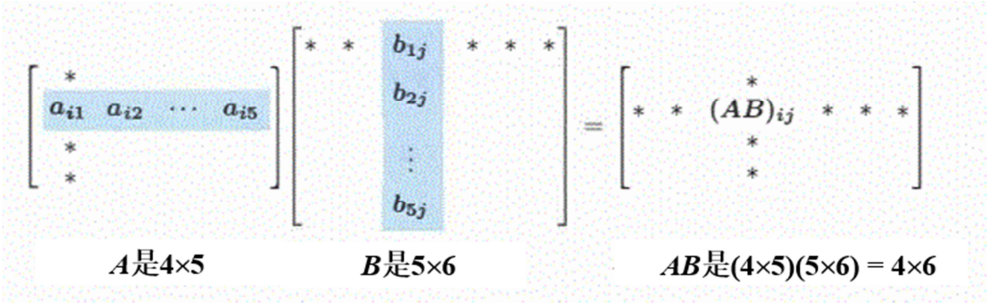


图 2.8：$i=2, j=3$，则$(AB)_{23}$ 等于**(行 2) · (列 3)** $=a_{2k}b_{k3}$ 的总和

范例 1 方形矩阵可以相乘，**当且仅当**(if and only if)他们有相同的大小:

$$\begin{bmatrix}1 & 1\\ 2 & -1\end{bmatrix}\begin{bmatrix}2 & 2\\ 3 & 4\end{bmatrix}=\begin{bmatrix}5 & 6\\ 1 & 0\end{bmatrix}$$

第一个点积是 $1\cdot 2+1\cdot 3=5$，另外三个点积是 6, 1, 0。每个点积需要两次乘法——总共是 8 次。

若 A 与 B 都是 $n\times n$，那么 AB 也是 $n\times n$，就会有 n^2 次点积，等于 A 的行数乘 B 的列数。每一个点积需要 n 个乘法，计算 AB 就需要 n^3 的乘法运算。当 $n=100$，我们需要乘 100 万次，当 $n=2$，需要 $n^3=8$ 次。

直到最近数学家还是认为绝对需要 8 次的乘法，有人找到只需要 7 次(额外有加法)的方法。把 $n\times n$ 的矩阵分割成 2×2 方块，这个想法也可以降低大型矩阵乘法的运算次数，现在乘法量 n^3 变成 $n^{2.376}$。那么有可能变成 n^2 吗？这个演算法有点难以处理，目前还是认为正常的科学计算需要 n^3。

【原文的 2 by 2 blocks 意思是每个方块都是 2×2 矩阵】

请尊重版权与译者的劳动成果，侵权必究!

范例 2 假设 A 是一个行向量(1×3)，B 是一个列向量(3×1)，则 AB 是 1×1(因为点积只有一个单元)。反过来说，BA(**一个列乘一个行**)会是一个完整的 3×3 矩阵，这样的相乘是允许的！

$$\underset{(n\times 1)\times(1\times n)=(n\times n)}{\text{列乘行}} \qquad \begin{bmatrix}0\\1\\2\end{bmatrix}\begin{bmatrix}1 & 2 & 3\end{bmatrix}=\begin{bmatrix}0 & 0 & 0\\1 & 2 & 3\\2 & 4 & 6\end{bmatrix}$$

一个行乘一个列是**内积**，这是点积的另外一个名称。一个列乘一个行是**外积**(outer product)，这些是矩阵乘法的极端例子。

第二与第三种方法：行与列

在大图中，A 乘(***B* 的每一列**)，结果是 AB 的一个列，在这个列我们组合 A 的列。(AB)的每个列都是 ***A* 的列**的线性组合，这就是矩阵乘法的列图：

2. 矩阵 *A* 乘(*B* 的每一列) $\quad A[\boldsymbol{b}_1 \ldots \boldsymbol{b}_p] = [A\boldsymbol{b}_1 \ldots A\boldsymbol{b}_p]$

行图恰好相反，是(A 的每一行)乘整个 B，结果就是 AB 的一个行。***AB* 的每一行都是(*B* 的行)的组合：**

3. (*A* 的每一行)乘矩阵 *B* $\quad [\boldsymbol{A}\text{ 的行 }\boldsymbol{i}]\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix}=[\boldsymbol{AB}\text{ 的行 }\boldsymbol{i}]$

我们在消元法看到行运算(E 乘 A)，很快的我们会看到 $AA^{-1}=I$ 的列。“行–列图”有行与列的点积，手算矩阵乘法时经常使用点积：mnp 次个别乘法/加法的步骤。

$$\boldsymbol{AB}=(\boldsymbol{m}\times\boldsymbol{n})\,(\boldsymbol{n}\times\boldsymbol{p})=(\boldsymbol{m}\times\boldsymbol{p}) \qquad \boldsymbol{mp}\text{ 次点积，每次点积需要 }\boldsymbol{n}\text{ 个步骤} \tag{2}$$

第四种方法：列乘行

矩阵乘法有第四个方法，没有很多人了解它的重要性。我觉得像魔术师一样在解说戏法，魔术师不会做这种事，但是数学家一直在尝试。第四个方法在本书前一版本有介绍，但是我的强调好像不够。

4. *A* 的列 1 至 *n*，乘 *B* 的行 1 至 *n*，全部矩阵相加。

A 的列 1 乘 B 的行 1，A 的列 2, 3 乘 B 的行 2, 3，然后相加：

$$\begin{bmatrix}\text{列}1 & \text{列}2 & \text{列}3\\ \bullet & \bullet & \bullet\\ \bullet & \bullet & \bullet\end{bmatrix}\begin{bmatrix}\text{行}1 & \bullet & \bullet & \bullet\\ \text{行}2 & \bullet & \bullet & \bullet\\ \text{行}3 & \bullet & \bullet & \bullet\end{bmatrix}=(\text{列}1)\,(\text{行}1)+(\text{列}2)\,(\text{行}2)+(\text{列}3)\,(\text{行}3)$$

如果我使用列-行的方法处理 2×2 矩阵的乘法，你会了解 AB 是正确的：

$$AB=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}E & F\\ G & H\end{bmatrix}=\begin{bmatrix}aE+bG & aF+bH\\ cE+dG & cF+dH\end{bmatrix}$$

A 的列乘 B 的行，相加

$$AB=\begin{bmatrix}a\\ c\end{bmatrix}\begin{bmatrix}E & F\end{bmatrix}+\begin{bmatrix}b\\ d\end{bmatrix}\begin{bmatrix}G & H\end{bmatrix} \tag{3}$$

A 的列 k 乘 B 的行 k 会得到一个矩阵(不是一个数)，令 $k=1, 2, ..., n$，然后全部相加可以得到 AB。

若 AB 是$(m\times n)$ $(n\times p)$，则 n 个矩阵会是(列) (行)，他们全部是 $m\times p$ 矩阵。这个方法同样使用与点积相同的 mnp 个步骤——但是顺序不同。

矩阵运算法则

我可以立案列出六个矩阵必须遵守的法则，却只是强调一个不遵守的法则吗？矩阵可以是方形或是矩形，与 $A+B$ 有关的法则都很单纯并且遵守，以下是三个加法法则：

$$A+B=B+A \qquad \text{(交换律)}$$
$$c(A+B)=cA+cB \qquad \text{(分配律)}$$
$$A+(B+C)=(A+B)+C \qquad \text{(结合律)}$$

另外三个法则对于乘法成立，但是 $AB=BA$ 通常不在其中：

$$AB\neq BA \qquad \text{(交换律通常不成立)}$$
$$A(B+C)=AB+AC \qquad \text{(左侧分配律)}$$
$$(A+B)C=AC+BC \qquad \text{(右侧分配律)}$$
$$A(BC)=(AB)C \qquad (ABC \text{ 的结合律}) \text{ (不需要括号)}$$

当 A 与 B 不是方形，AB 的大小与 BA 的大小不同，这些矩阵无法相等——纵使两个乘法都是被允许的。对于方阵而言，几乎任意例子都会显示 $AB\neq BA$：

$$AB=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & \mathbf{1}\end{bmatrix} \quad \text{但是} \quad BA=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}=\begin{bmatrix}\mathbf{1} & 0\\ 0 & 0\end{bmatrix}$$

$AI=IA$ 永远成立，所有的方形矩阵可以与 I 交换乘，也可以与 cI 交换乘，只有矩阵 cI 可以与其他所有矩阵交换乘。

法则 $A(B+C)=AB+AC$ 可以一次证明一个列，对于第一列从 $A(\boldsymbol{b}+\boldsymbol{c})=A\boldsymbol{b}+A\boldsymbol{c}$ 开始，这是所有事物的关键——线性。不再说了。

法则 $A(BC)=(AB)C$ 表示可以先乘 BC 也可以先乘 AB，直接的证明有点无趣(问题 37)，但是这个法则非常有用，我们特意加高亮，它是矩阵乘法的关键。

请尊重版权与译者的劳动成果，侵权必究!

检视特殊情况 $A = B = C$ =方形矩阵，(A 乘 A^2)等于(A^2 乘 A)，两者的乘积都是 A^3。矩阵的次方(power) A^p 遵循如同数字的规则:

$$A^p = AAA\ldots A \quad (p \text{ 个因子}) \qquad \boldsymbol{(A^p)\,(A^q) = A^{p+q}} \qquad \boldsymbol{(A^p)^q = A^{pq}}$$

这些是指数的一般法则，A^3 乘 $A^4 = A^7$(7 个因子)，但是 A^3 的 4 次方是 A^{12}(12 个 A)。假设 A 存在–1 次方—代表逆矩阵 A^{-1}，当 p 与 q 是零或是负数时，这些法则仍然成立。$A^0 = I$ 是单位矩阵，如同 $2^0 = 1$。

一个数 $a^{-1} = 1/a$，对矩阵来说，逆矩阵写成 A^{-1}。(不是 I/A，除了在 MATLAB。)除了 $a = 0$ 以外，所有的数都有倒数。【数的 inverse 翻译成“倒数”】要决定 A 是否有逆矩阵是线性代数的中心问题，段落 2.5 会讨论这个答案。本段落是矩阵的“人权法案”，只是在说明当 A 与 B 可以乘时，该如何乘。

方块矩阵与方块乘法

我们必须再多说一个关于矩阵的事情，他们可以分割成数个方块(blocks，比较小的矩阵)，这个经常发生。下列是一个 4×6 的矩阵，分割成 2×2 的方块，本例题中每个方块都是单位矩阵 I:

4×6 矩阵
2×2 方块得到
2×3的方块矩阵

$$A = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} I & I & I \\ I & I & I \end{bmatrix}$$

若 B 也是 4×6，且方块的大小适配，可以对 $A + B$ 一次执行一个方块相加。

在之前你就看了过方块矩阵，右侧的 $\boldsymbol{b}$ 放在 A 的旁边变成增广矩阵，那么[A $\boldsymbol{b}$]就有两个不同大小的方块，左乘一个消元矩阵得到[EA $E\boldsymbol{b}$]。左乘或是右乘方块都没有问题，只要形状合适即可。【2 by 3 block matrix 是 2×3 = 6 个方块】

方块乘法 若 A 的方块可以乘 B 的方块，则允许 AB 的方块乘法，A 的列切割必须适配 B 的行切割。

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix} \tag{4}$$

这个方程式与方块是数字(每个方块是 1×1)时是一样，我们注意所有的 A 要放在 B 之前，因为 AB 与 BA 不同。

重点：当矩阵分割成方块时，会比较容易看出他们是如何作用。上述由 I's 组成的方块矩阵就会比 4×6 矩阵 A 简洁许多。

范例 3 **(重要的特殊情形)** 令矩阵 A 的方块有 n 列，矩阵 B 的方块有 n 行，则 AB 的方块乘积等于列乘行的总和：

$$\begin{matrix}\text{列}\\ \text{乘}\\ \text{行}\end{matrix}\quad \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} — & \boldsymbol{b}_1 & — \\ & \cdots & \\ — & \boldsymbol{b}_n & — \end{bmatrix}=\left[\boldsymbol{a}_1\boldsymbol{b}_1+\ldots+\boldsymbol{a}_n\boldsymbol{b}_n\right] \tag{5}$$

这是矩阵乘法的规则 4。以下是数字的例子：

$$\begin{bmatrix}1 & 4\\ 1 & 5\end{bmatrix}\begin{bmatrix}3 & 2\\ 1 & 0\end{bmatrix}=\begin{bmatrix}1\\ 1\end{bmatrix}\begin{bmatrix}3 & 2\end{bmatrix}+\begin{bmatrix}4\\ 5\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix}=\begin{bmatrix}3 & 2\\ 3 & 2\end{bmatrix}+\begin{bmatrix}4 & 0\\ 5 & 0\end{bmatrix}=\begin{bmatrix}7 & 2\\ 8 & 2\end{bmatrix}$$

小结：通常会用行乘列求矩阵乘积，需要 4 次点积(8 个乘法)。新的方法用列乘行，得到两个完整的矩阵(一样是 8 次乘法)。

范例 4 **(利用方块执行消元法)** 假设 A 的第一列包含 1, 3, 4，要把 3, 4 变成 0, 0，需要减去主元行的 3 与 4 倍，这些行运算就是消元矩阵 E_{21} 与 E_{31}:

$$\text{一次一个动作}\quad E_{21}=\begin{bmatrix}1 & 0 & 0\\ -3 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\text{ 以及 }E_{31}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -4 & 0 & 1\end{bmatrix}$$

“方块概念”就是希望用一个矩阵 E 同时做两件事情，这个矩阵把第一列的主元 $a=1$ 下方的数字一次性变成 0：

$$E=\begin{bmatrix}\boldsymbol{1} & 0 & 0\\ \boldsymbol{-3} & 1 & 0\\ \boldsymbol{-4} & 0 & 1\end{bmatrix}\text{ 乘 }\begin{bmatrix}\boldsymbol{1} & x & x\\ \boldsymbol{3} & x & x\\ \boldsymbol{4} & x & x\end{bmatrix}\text{ 得到 }EA=\begin{bmatrix}\boldsymbol{1} & x & x\\ \boldsymbol{0} & y & y\\ \boldsymbol{0} & z & z\end{bmatrix}$$

利用逆矩阵，方块矩阵 E 可以对于整个(方块)列做消元。假设矩阵有 4 个方块 A, B, C, D，观察 E 如何借由方块消去 C：

$$\text{方块消元}\quad \begin{bmatrix} I & \boldsymbol{0} \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}=\begin{bmatrix} A & B \\ \boldsymbol{0} & \boldsymbol{D-CA^{-1}B} \end{bmatrix} \tag{6}$$

消元法对于第一行$[A \quad B]$左乘 CA^{-1}(之前是 c/a)，使得原来 C 的方块变成 0 方块，原来 D 的方块变成 $S=D-CA^{-1}B$。

这是一般的消元法，一次执行一列--现在是利用方块进行。此处主元方块是 A，最后的方块是 $D-CA^{-1}B$，如同 $d-cb/a$，称为**舒尔补充**(Schur complement)。

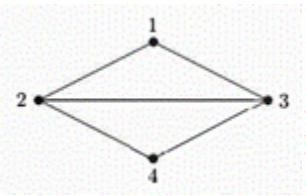
请尊重版权与译者的劳动成果，侵权必究!

主要观念的复习

1. AB 的(i,j)单元是(A 的行 i) · (B 的列 j)。
2. $m\times n$ 矩阵乘 $n\times p$ 矩阵会使用 mnp 次乘法。
3. A 乘 BC 等于 AB 乘 C (非常重要)。
4. AB 也是 n 个矩阵的总和：(A 的列 i) 乘 (B 的行 j) 的总和。
5. 当方块的形状适配时，方块乘法是允许的。
6. 方块消元法可以得到**舒尔补充** $D - CA^{-1}B$。

已解范例

2.4A 一个图形(或网络)有 n 个节点(node)，它的邻接矩阵(adjacency matrix)S 是 $n\times n$，这是一个非 0 即 1 的矩阵，当节点 i 与节点 j 有边(edge)相连时 $s_{ij}=1$。



$$S=\begin{bmatrix}0&1&1&0\\1&0&1&1\\1&1&0&1\\0&1&1&0\end{bmatrix}$$

对于无向图而言，邻接矩阵是方形而且对称，边的两个方向都可以行走。

矩阵 S^2 有一个很有用的解释，$(S^2)_{ij}$ 可以算出节点 i 与节点 j 之间长度 2 的路径个数。举例说明，上图中节点 2 与 3 之间长度 2 的路径有 2 个：经过 1 或经过 4。节点 1 到节点 1 的长度 2 路径也有 2 个：1-2-1 与 1-3-1。

$$S^2=\begin{bmatrix}2&1&1&2\\1&3&2&1\\1&2&3&1\\2&1&1&2\end{bmatrix}\qquad S^3=\begin{bmatrix}2&5&5&2\\5&4&5&5\\5&5&4&5\\2&5&5&2\end{bmatrix}$$

你可以找出 5 条节点 1 与 2 之间长度 3 的路径吗？【参考答案：1-2-1-2, 1-3-4-2, 1-2-3-2, 1-3-1-2, 1-3-4-2)】

真正的问题在于为什么 S^N 可以算出在两个节点之间长度 N 的路径数目？从 S^2 开始，检视点积得到的矩阵：

$$(S^2)_{ij}=(S\text{ 的行 }i)\cdot(S\text{ 的列 }j)=s_{i1}s_{1j}+s_{i2}s_{2j}+s_{i3}s_{3j}+s_{i4}s_{4j}\qquad(7)$$

若存在两步路径 $i\to 1\to j$，第一个乘法得到 $s_{i1}s_{1j}=(1)\,(1)=1$，如果 $i\to 1\to j$ 不是路径，那么不是 $i\to 1$ 丢失就是 $1\to j$ 丢失，这种情形会得到 $s_{i1}s_{1j}=0$。

$(S^2)_{ij}$ 会把所有的两步路径 $i \to k \to j$ 的 1 累加，得到总路径数。同样的方法，$S^{N-1}S$ 可以计算 N 步路径数，S^{N-1} 代表从 i 到 k 的$(N-1)$步的路径数目，接着 S 代表 k 到 j 的一步路径。矩阵乘法非常适合计算图形的路径，也可以视为公司内两个雇员之间的通信频道数。

2.4B 下列三个矩阵，何时会有 $AB = BA$？何时会有 $BC = CB$？何时会有 A 乘 BC $= AB$ 乘 C？给出矩阵单元 p, q, r, z 的条件：

$$A = \begin{bmatrix} p & 0 \\ q & r \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix}$$

若 $p, q, r, 1, z$ 是 4×4 的方块而不是数字，答案会改变吗？

解 首先 A 乘 BC 永远等于 AB 乘 C，在 $A(BC) = (AB)C = ABC$ 中括号不需要存在，但是矩阵的顺序必须保持：

通常来说 $AB \neq BA$ $\quad AB = \begin{bmatrix} p & p \\ q & q+r \end{bmatrix} \quad BA = \begin{bmatrix} p+q & r \\ q & r \end{bmatrix}$

碰巧会有 $BC = CB$ $\quad BC = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix} \quad CB = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix}$

B 与 C 只是碰巧可以交换，一部分的解释是 B 的对角线是 I，可以跟 2×2 矩阵交换。当 p, q, r, z 是 4×4 的方块，1 变成 I，所有的乘积维持正确，答案还是一样。

问题集 2.4

问题 1-16 有关矩阵乘法的法则。

1 A 是 3×5，B 是 5×3，C 是 5×1，D 是 3×1，所有的单元都是 1。下列的矩阵运算哪些是允许的？结果为何？

$BA \qquad AB \qquad ABD \qquad DC \qquad A(B+C)$

2 什么样的行或列或矩阵乘在一起，会得到

(a) AB 的第二列。

(b) AB 的第一行。

(c) AB 的行 3 列 5 单元。

(d) CDE 的行 1 列 1 单元。

3 把 AB 加到 AC，然后比较 $A(B+C)$:

$$A = \begin{bmatrix} 1 & 5 \\ 2 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 3 & 1 \\ 0 & 0 \end{bmatrix}$$

4 问题 3 中，计算 A 乘 BC，然后计算 AB 乘 C。

请尊重版权与译者的劳动成果，侵权必究！

5 计算 A^2 与 A^3，对 A^5 与 A^n 进行预测。

$$A=\begin{bmatrix}1 & b\\ 0 & 1\end{bmatrix} \text{ 与 } A=\begin{bmatrix}2 & 2\\ 0 & 0\end{bmatrix}$$

6 证明$(A+B)^2$ 与 $A^2+2AB+B^2$ 不同，当

$$A=\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1 & 0\\ 3 & 0\end{bmatrix}$$

写下正确的规则：$(A+B)\,(A+B)=A^2+____+B^2$。

7 是非题，如果是错误的，举出一个特定的例子。

(a) 若 B 的列 1 与列 3 相同，AB 的列 1 与列 3 也会相同。

(b) 若 B 的行 1 与行 3 相同，AB 的行 1 与行 3 也会相同。

(c) 若 A 的行 1 与行 3 相同，ABC 的行 1 与行 3 也会相同。

(d) $(AB)^2=A^2B^2$。

8 DA 与 EA 的每个行如何与 A 的行产生关联，当

$$D=\begin{bmatrix}3 & 0\\ 0 & 5\end{bmatrix} \text{ 与 } E=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix} \text{ 与 } A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$$

9 A 的行 1 加到行 2，得到的 EA 如下列所示。然后 EA 的列 1 加到列 2 得到$(EA)F$:

$$EA=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}\begin{bmatrix}a & b\\ c & d\end{bmatrix}=\begin{bmatrix}a & b\\ a+c & b+d\end{bmatrix}$$

$$\text{与 } (EA)F=(EA)\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}=\begin{bmatrix}a & a+b\\ a+c & a+c+b+d\end{bmatrix}$$

(a) 以反向顺序执行这些步骤，首先借由 AF 把 A 的列 1 加到列 2，然后借由 $E(AF)$把 AF 的行 1 加到行 2。

(b) 比较$(EA)F$ 与 $E(AF)$，结果会遵守矩阵乘法哪个法则?

10 A 的行 1 再次加到行 2 得到 EA，然后 F 把 EA 的行 2 加到行 1，结果是 $F(EA)$。

$$F(EA)=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}a & b\\ a+c & b+d\end{bmatrix}=\begin{bmatrix}2a+c & 2b+d\\ a+c & b+d\end{bmatrix}$$

(a) 以反向顺序执行这些步骤，首先借由 FA 把 A 的行 2 加到行 1，然后 FA 的行 1 加到行 2。

(b) 矩阵乘法的哪个法则有遵守或是没有遵守?

11 这个事实让我很惊奇，如果你对 A 执行一个行运算，然后再执行一个列运算，得到的结果与你先做列运算是相同的。(试试看) 为什么这是真的?

12 (3×3 矩阵) 选择唯一的 B 使得每一个矩阵 A 都有：

(a) $BA = 4A$。

(b) $BA = 4B$。

(c) BA 把 A 的行 1 与行 3 交换，且行 2 没变。

(d) BA 所有的行与 A 的行 1 相同。

13 假设 $AB = BA$ 且 $AC = CA$，其中两个特定矩阵 B 与 C：

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{与 } B, C \text{ 交换乘} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

证明 $a = d$ 且 $b = c = 0$，然后 A 是 I 的倍数。唯一可以与 B 及 C 及其他 2×2 矩阵进行交换乘的矩阵是 $A = I$ 的倍数。

14 哪些矩阵确定等于 $(A - B)^2$：　$A^2 - B^2$，$(B - A)^2$，$A^2 - 2AB + B^2$，

$A(A - B) - B(A - B)$，$A^2 - AB - BA + B^2$？

15 是非题：

(a) 若 A^2 有定义，则 A 必须是方形。

(b) 若 AB 与 BA 有定义，则 A 与 B 是方形。

(c) 若 AB 与 BA 有定义，则 AB 与 BA 是方形。

(d) 若 $AB = B$，则 $A = I$。

16 若 A 是 $m \times n$，下列步骤会牵涉到多少的个别乘法？

(a) A 乘具有 n 个分量的向量 $\boldsymbol{x}$。

(b) A 乘 $n \times p$ 矩阵 B。

(c) A 乘自己得到 A^2？　此处 $m = n$。

17 $A = \begin{bmatrix} 2 & -1 \\ 3 & -2 \end{bmatrix}$ 且 $B = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 0 & 6 \end{bmatrix}$，计算下列答案：

(a) AB 的列 2。

(b) AB 的行 2。

(c) $AA = A^2$ 的行 2。

(d) $AAA = A^3$ 的行 2。

问题 18-20 使用 a_{ij} 作为 A 的行 i 列 j 单元。

18 写出 3×3 矩阵 A，它的单元如下：

请尊重版权与译者的劳动成果，侵权必究！

(a) $a_{ij} = i$ 与 j 的较小者。

(b) $a_{ij} = (-1)^{i+j}$。

(c) $a_{ij} = i / j$。

19 你会用那些字来描述矩阵的分类？在每一类给定 3×3 的例子，哪个矩阵属于全部 4 类？

(a) $a_{ij} = 0$，若 $i \neq j$。

(b) $a_{ij} = 0$，若 $i < j$。

(c) $a_{ij} = a_{ji}$。

(d) $a_{ij} = a_{1j}$。

20 A 的单元是 a_{ij}，假设全部单元都不是 0，请问

(a) 第一主元 =？

(b) 从列 3 减去列 1 所需的乘数 l_{31} = ？

(c) 减去之后，取代原来 a_{32} 的新单元 = ？

(d) 第二主元 =？

问题 21-24 与 A 的次方有关。

21 计算 A^2，A^3，A^4 以及 $A\boldsymbol{v}$，$A^2\boldsymbol{v}$，$A^3\boldsymbol{v}$，$A^4\boldsymbol{v}$ ：

$$A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ 以及 } \boldsymbol{v} = \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}$$

22 利用尝试错误法求实际的 2×2 矩阵，使得

$A^2 = -I$　　$BC = 0$　　$DE = -ED$ (不允许 $DE = 0$)。

23 (a) 求非零矩阵 A 使得 $A^2 = 0$。

(b) 找出矩阵有 $A^2 \neq 0$ 但是 $A^3 = 0$

24 利用 $n = 2$ 与 $n = 3$ 的实验结果，对于下列矩阵预测 A^n ：

$$A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \text{ 与 } A_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ 与 } A_3 = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$$

问题 25-31 使用列-行乘法以及方块乘法。

25 使用 A(3×3)的列乘 I 的行，计算 A 乘 I。

26 使用列乘行，计算 AB：

$$AB=\begin{bmatrix}1&0\\2&4\\2&1\end{bmatrix}\begin{bmatrix}3&3&0\\1&2&1\end{bmatrix}=\begin{bmatrix}1\\2\\2\end{bmatrix}\begin{bmatrix}3&3&0\end{bmatrix}+_____=_____$$

27 证明上三角矩阵的乘积永远是上三角。

$$AB=\begin{bmatrix}x&x&x\\0&x&x\\0&0&x\end{bmatrix}\begin{bmatrix}x&x&x\\0&x&x\\0&0&x\end{bmatrix}=\begin{bmatrix}0&&\\0&0&\\&&\end{bmatrix}$$

利用点积(行乘列)证明　　(A 的行 2) · (B 的列 1) = 0

哪些其他的点积也等于零?

利用整个矩阵(列乘行)证明　　画出(A 的列 2)乘(B 的行 2)的 x's 与 0's，也展示(A 的列 3)乘(B 的行 3)。

28 画出 A(2×3)与 B(3×4)与 AB 中的切割，展示 4 种乘法规则每一种实际上都是方块乘法：

(1) (A 矩阵)乘(B 的列)　　AB 的列

(2) (A 的行)乘(B 矩阵)　　AB 的行

(3) (A 的行)乘(B 的列)　　内积(AB 的数字)

(4) (A 的列)乘(B 的行)　　外积(矩阵相加得到 AB)

29 哪个矩阵 E_{21} 与 E_{31} 产生 $E_{21}A$ 与 $E_{31}A$ 在位置(2, 1)与位置(3, 1)的零?

$$A=\begin{bmatrix}2&1&0\\-2&0&1\\8&5&3\end{bmatrix}$$

求出单一矩阵 $E=E_{21}E_{31}$ 同时产生两个 0，计算 EA。

30 方块乘法说：列 1 被消元，借由

$$EA=\begin{bmatrix}1&\mathbf{0}\\-\boldsymbol{c}/a&I\end{bmatrix}\begin{bmatrix}a&\boldsymbol{b}\\\boldsymbol{c}&D\end{bmatrix}=\begin{bmatrix}a&\boldsymbol{b}\\\mathbf{0}&D-\boldsymbol{cb}/a\end{bmatrix}$$

问题 29 中，什么数字进到 $\boldsymbol{c}$ 与 D? $D-\boldsymbol{cb}/a$ =? 【a: 数字，$\boldsymbol{b}$ 与 $\boldsymbol{c}$: 向量】

31 令 $i^2=-1$，则($A+iB$)与($\boldsymbol{x}+i\boldsymbol{y}$)的乘积是 $A\boldsymbol{x}+iB\boldsymbol{x}+iA\boldsymbol{y}-B\boldsymbol{y}$。利用方块把没有 i 的实部与有 i 的虚部进行分离：

$$\begin{bmatrix}A&-B\\?&?\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{bmatrix}=\begin{bmatrix}A\boldsymbol{x}-B\boldsymbol{y}\\?\end{bmatrix}\quad\begin{matrix}\text{实数部分}\\\text{虚数部分}\end{matrix}$$

32 (非常重要) 假设你针对三个特殊的右侧 $\boldsymbol{b}$ 求解方程式 $A\boldsymbol{x}=\boldsymbol{b}$：

请尊重版权与译者的劳动成果，侵权必究!

$$A\boldsymbol{x}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix} \qquad A\boldsymbol{x}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix} \qquad A\boldsymbol{x}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

如果三个解 $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ 是一个矩阵 X 的列，计算 A 乘 X?

33 若问题 32 的三个解是 $\boldsymbol{x}_1 = (1, 1, 1)$，$\boldsymbol{x}_2 = (0, 1, 1)$，$\boldsymbol{x}_3 = (0, 0, 1)$。当 $\boldsymbol{b} = (3, 5, 8)$ 时，求解 $A\boldsymbol{x} = \boldsymbol{b}$。挑战问题：$A$ 是什么？

34 求出满足 $A\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} = \begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}A$ 的所有矩阵 $A = \begin{bmatrix}a & b\\c & d\end{bmatrix}$。

35 假设一个“圆形图”有 4 个节点，点与点之间有边(双向)连接。参考已解范例 **2.4A**，求图形的邻接矩阵 S? 求 S^2? 找出由 S^2 预测的所有 2 步路径。

挑战问题

36 **实际问题：**假设 A 是 $m\times n$，B 是 $n\times p$，C 是 $p\times q$，则 AB 需要 mnp 的乘法个数，$(AB)C$ 需要 $mnp + mpq$ 个乘法；来自 A 乘(BC)的相同矩阵需要 $mnq + npq$ 个乘法。注意 BC 需要 npq 个乘法。

(a) 若 A 是 2×4，B 是 4×7，C 是 7×10，你会偏好$(AB)C$ 还是 $A(BC)$?

(b) 如果是 N-分量向量，你会选择$(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})\,\boldsymbol{w}^{\mathrm{T}}$ 还是 $\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{v}\boldsymbol{w}^{\mathrm{T}})$?

(c) 当 $n^{-1} + q^{-1} < m^{-1} + p^{-1}$ 时，借由除以 $mnpq$ 证明$(AB)C$ 会比较快速。

37 利用 B 的列 $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ 证明$(AB)C = A(BC)$。首先假设 C 只有一个列 $\boldsymbol{c}$，其中 $\boldsymbol{c}$ 的分量是 $c_1, \ldots, c_n$：

AB 的列是 $A\boldsymbol{b}_1, \ldots, A\boldsymbol{b}_n$，则$(AB)\boldsymbol{c}$ 等于 $c_1A\boldsymbol{b}_1 + \ldots + c_nA\boldsymbol{b}_n$ 。

$B\boldsymbol{c}$ 有一个列 $c_1\,\boldsymbol{b}_1 + \ldots + c_n\,\boldsymbol{b}_n$，则 $A(B\boldsymbol{c})$等于 $A\,(c_1\boldsymbol{b}_1 + \ldots + c_n\boldsymbol{b}_n)$。

线性的性质使得两个和(sum)相等，证明了$(AB)\boldsymbol{c}$ 等于 $A(B\boldsymbol{c})$。对于 C 的其他的_____也是一样，因此$(AB)C = A(BC)$。应用到逆反：若 $BA = I$ 且 $AC = I$，证明左逆反 B 等于右逆反 C。

38 (a) 假设 A 的行是 $\boldsymbol{a}_1^{\mathrm{T}}, \ldots, \boldsymbol{a}_m^{\mathrm{T}}$，为什么 $A^{\mathrm{T}}A$ 等于 $\boldsymbol{a}_1\boldsymbol{a}_1^{\mathrm{T}} + \ldots + \boldsymbol{a}_m\boldsymbol{a}_m^{\mathrm{T}}$?

(b) 假设 C 是对角矩阵且 $c_1, \ldots, c_m$ 在对角线，对于 $A^{\mathrm{T}}CA$，求出一个类似列乘行的总和。首先做一个 $m = n = 2$ 的例子。

2.5 逆矩阵

1 若矩阵 A 有逆矩阵，则 $A^{-1}A=I$ 且 $AA^{-1}=I$。

2 用来测试可逆性质的演绎法是消元法：A 必须具有 n 个(非零)主元。

3 用来测试可逆性质的代数法是 A 的行列式：$\det A \neq 0$。

4 用来测试可逆性质的方程式是 $A\boldsymbol{x}=\boldsymbol{0}$：$\boldsymbol{x}=\boldsymbol{0}$ 必须是唯一解。

5 若 A 与 B(相同大小)都是可逆，则 AB 也是可逆：$(AB)^{-1}=B^{-1}A^{-1}$。

6 $AA^{-1}=I$ 是 A^{-1} 的 ***n* 个列**的 ***n* 个方程式**，高斯-乔丹消元法使得$[A\ I]$到$[I\ A^{-1}]$。

7 本书的最后一页描述了方形矩阵 A 可逆的 14 种等效条件。

假设 A 是方形矩阵，我们寻找具有相同大小的**逆矩阵** A^{-1}，使得 A^{-1} 乘 A 等于 I。A 可以做的事情，A^{-1} 都会反着来，他们的乘积是单位矩阵——单位矩阵对于向量毫无影响，所以 $A^{-1}A\boldsymbol{x}=\boldsymbol{x}$，但是 A^{-1} 有可能不存在。

矩阵最常做的事情就是乘一个向量 $\boldsymbol{x}$，A^{-1} 乘$(A\boldsymbol{x}=\boldsymbol{b})$，得到 $A^{-1}A\boldsymbol{x}=A^{-1}\boldsymbol{b}$，所以 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$。乘积 $A^{-1}A$ 很像乘一个数再除一个数。若一个数不为 0，必然存在倒数—矩阵的情形比较复杂却更有趣，A^{-1} 称为 A 的逆矩阵。

定义 如果存在一个矩阵 A^{-1}“逆反”A，则矩阵 A 可逆：

$$\text{双边逆反} \qquad \boldsymbol{A^{-1}A=I}\ \text{且}\ \boldsymbol{AA^{-1}=I} \tag{1}$$

不是所有的矩阵都有逆矩阵，关于方形矩阵的第一个问题：A 是否可逆？我们不是说要马上计算 A^{-1}，在大部分的问题中，我们都没有去计算逆矩阵。以下有 6 个关于 A^{-1} 的**注释**。

注释 1 矩阵存在逆矩阵，当且仅当消元法**得到 *n* 个主元**(允许行交换)。消元法求解 $A\boldsymbol{x}=\boldsymbol{b}$ 时，不需要明确使用 A^{-1}。

注释 2 矩阵 A 不可能存在两个不同的逆矩阵。假设 $BA=I$ 且 $AC=I$，依据“括号证明法”可以得到 $B=C$：

$$B(AC)=(BA)C \ \text{得到}\ BI=IC \ \text{或是}\ B=C \tag{2}$$

上式证明了“左逆矩阵”B(从左乘)与“右逆矩阵”C(从右乘 A 得到 $AC=I$)必须是相同的矩阵。

注释 3 若矩阵 A 可逆，$A\boldsymbol{x}=\boldsymbol{b}$ 的唯一解是 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$：

$$A^{-1}\text{乘}\ A\boldsymbol{x}=\boldsymbol{b}\ \text{得到}\ \boldsymbol{x}=A^{-1}A\boldsymbol{x}=A^{-1}\boldsymbol{b}$$

请尊重版权与译者的劳动成果，侵权必究！

注释 4 (**重要**) 假设有一个非零向量 $\boldsymbol{x}$ 使得 $A\boldsymbol{x}=\boldsymbol{0}$，则 A 不存在逆矩阵，没有这样的矩阵可以从 $\boldsymbol{0}$ 反推到 $\boldsymbol{x}$。

若 A 可逆，则 $A\boldsymbol{x}=0$ 只存在零的解 $\boldsymbol{x}=A^{-1}\boldsymbol{0}=\boldsymbol{0}$。

注释 5 一个 2×2 矩阵可逆，当且仅当 $ad-bc\neq 0$：

$$\textbf{2×2 逆矩阵}\qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{3}$$

其中 $ad-bc$ 是 A 的行列式，若矩阵的行列式不为零(第五章)，则矩阵可逆。通常在行列式出现之前已经完成 n 个主元的测试。

注释 6 若对角线单元全部不是零，则对角线矩阵可逆：

$$\text{若}\quad A=\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}\quad \text{则}\quad A^{-1}=\begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_n \end{bmatrix}$$

范例 1 一个 2×2 矩阵 $A=\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ 不可逆，因为 $ad-bc=2-2=0$，注释 5 的测试失败。当 $\boldsymbol{x}=(2,-1)$时 $A\boldsymbol{x}=\boldsymbol{0}$，注释 3 的测试失败。找不到两个主元，注释 1 的测试失败。

消元法使得矩阵 A 的第二行变成零行。

乘积 AB 的逆矩阵

两个非零数 a 与 b，两者的总和不一定可逆。当 $a=3$ 与 $b=-3$，两者的倒数是 $a=1/3$ 与 $b=-1/3$，两者的总和是零，所以没有倒数。但是乘积 $ab=-9$ 存在倒数，倒数 1/3 乘(−1/3)。

对两个矩阵而言情形与上述差不多，我们很难去讨论 $A+B$ 的可逆性质。但是两者的乘积 AB 有逆矩阵，当且仅当两个矩阵 A 与 B 个别都是可逆(相同大小)。重点在于 A^{-1} 与 B^{-1} 呈现**反序**：

若 A 与 B 都是可逆，则 AB 也是可逆。AB 的逆矩阵是：

$$\boldsymbol{(AB)^{-1}=B^{-1}A^{-1}} \tag{4}$$

要了解为什么是相反顺序，计算 AB 乘 $B^{-1}A^{-1}$，中间有 $BB^{-1}=I$:

$\boldsymbol{AB}$ **的逆矩阵**　　$(AB)\,(B^{-1}A^{-1})=AIA^{-1}=AA^{-1}=I$

我们移去括号先求 BB^{-1}，同理 $B^{-1}A^{-1}$ 乘 AB 等于 I。

$B^{-1}A^{-1}$ 说明了一个数学的法则，逆矩阵是相反顺序。这也是常识：如果你先穿袜子再穿鞋子，反过来要先脱去_____？相同的反序性质可以应用到三个或更多的矩阵：

$$\text{反序} \qquad (ABC)^{-1} = C^{-1}B^{-1}A^{-1} \tag{5}$$

范例 2 **消元矩阵的逆矩阵**。若 E 从行 2 减 5 乘行 1，则 E^{-1} 会把 5 乘行 1 加到行 2：

E **减去**
E^{-1} **加回**

$$E = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{-5} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{5} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

EE^{-1} 得到单位矩阵 I，$E^{-1}E$ 也得到 I。我们是先加再减第一行的 5 倍，若 $AC = I$，则自动得到 $CA = I$。

对于方形矩阵而言，在一侧的逆矩阵会自动成为另一侧的逆矩阵。

范例 3 若 F 从行 3 减去 4 乘行 2，那么 F^{-1} 会反加回去：

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \qquad F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

现在用 F 乘范例 2 的 E，求 FE? 也计算 E^{-1} 乘 F^{-1} 求出 $(FE)^{-1}$。注意 FE 与 $E^{-1}F^{-1}$ 的顺序！

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ \boxed{20} & -4 & 1 \end{bmatrix} \quad \text{反矩阵得到} \quad E^{-1}F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{5} & 1 & 0 \\ \boxed{0} & \mathbf{4} & 1 \end{bmatrix} \tag{6}$$

得到的结果既美丽又正确，乘积 FE 包含 20，但是它的逆矩阵却没有。E 从行 2 减去 5 乘行 1，然后 F 从行 3 减去 4 乘新的行 2 (因为行 1 而变化)，依据这个顺序，行 3 会感受到行 1 带来的变化。

依据 $E^{-1}F^{-1}$ 的顺序，效应没有发生。首先 F^{-1} 使得 4 乘行 2 加到行 3，然后 E^{-1} 使得 5 乘行 1 加到行 2，这里就没有 20，因为行 3 没有再改变。在 $E^{-1}F^{-1}$ 的顺序下，行 3 没有感受到行 1 带来的效应。

这就是为什么下个段落要选取 $A = LU$，可以从三角矩阵 U 回到 A，乘数会很自然又完美的出现在下三角矩阵 L。

消元顺序先 E 后 F；反向顺序先 F^{-1} 再 E^{-1}。

$E^{-1}F^{-1}$ 是快速的，乘数 5, 4 落在对角线单元 1's 的下方。

请尊重版权与译者的劳动成果，侵权必究！

高斯-乔丹消元法计算 A^{-1}

我提示过 A^{-1} 不一定明确需要。方程式 $A\boldsymbol{x} = \boldsymbol{b}$ 的解是 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$，但是没必要也没效率去计算 A^{-1} 再乘 $\boldsymbol{b}$。消元法直接得到 $\boldsymbol{x}$，同时消元法也是计算 A^{-1} 的方法，如同我们展示的。高斯-乔丹的概念在于求解 $AA^{-1} = I$，找出 A^{-1} 的每一列。

A 乘上 A^{-1} 的第一列(称为 $\boldsymbol{x}_1$)，得到 I 的第一列(称为 $\boldsymbol{e}_1$)，我们的方程式是 $A\boldsymbol{x}_1 = \boldsymbol{e}_1 = (1, 0, 0)$，后面还有两个方程式。$A^{-1}$ 的每一列表示成 $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$，用 A 乘后产生 I 的列：

A^{-1} 的三个列

$$AA^{-1} = A[\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \boldsymbol{x}_3] = [\boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \boldsymbol{e}_3] = I \tag{7}$$

要逆反 3×3 矩阵，我们必须求解三个方程式：$A\boldsymbol{x}_1 = \boldsymbol{e}_1$ 与 $A\boldsymbol{x}_2 = \boldsymbol{e}_2 = (0, 1, 0)$与 $A\boldsymbol{x}_3 = \boldsymbol{e}_3 = (0, 0, 1)$，高斯-乔丹消元法利用这个方法求 A^{-1}。

高斯-乔丹消元法同时求解 n 个方程式来计算 A^{-1}，一般来说增广矩阵多了一列 $\boldsymbol{b}$，现在我们有三个右侧 $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$(当 A 是 3×3)，他们是 I 的列，所以增广矩阵就是方块矩阵$[A \quad I]$。我冒险来逆反我最爱的矩阵 K，它的对角线单元都是 2，然后 2 的旁边都是–1，其他单元都是 0：

$$[K \quad \boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \boldsymbol{e}_3] = \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad \text{开始对 } K \text{ 执行高斯-乔丹法}$$

$$\rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{-1} & \frac{1}{2} & \mathbf{1} & \mathbf{0} \\ 0 & -1 & 2 & 0 & 0 & 1 \end{bmatrix} \quad \tfrac{1}{2}(\text{行}\mathbf{1}) + \text{行}\,\mathbf{2}$$

$$\rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & \mathbf{1} \end{bmatrix} \quad \tfrac{2}{3}(\text{行}\,\mathbf{2}) + \text{行}\,\mathbf{3}$$

我们完成 K^{-1} 的一半，矩阵的前三列是 U(上三角)，主元是对角线的 2, 3/2, 4/3。高斯想用反向代入法，乔丹的贡献在于继续消元！他一路走到**简化阶梯形式 (reduced echelon form)** $R = I$，行被加到在上方的行，使得**主元上方都是零**。

$$\text{第三主元以上都是 0} \rightarrow \begin{bmatrix} 2 & -1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \frac{3}{2} & \mathbf{0} & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad \tfrac{3}{4}(\text{行}\,\mathbf{3}) + \text{行}\,\mathbf{2}$$

$$\text{第二主元以上都是 0} \rightarrow \begin{bmatrix} \mathbf{2} & \mathbf{0} & \mathbf{0} & \frac{3}{2} & \mathbf{1} & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \quad \tfrac{2}{3}(\text{行}\,\mathbf{2}) + \text{行}\,\mathbf{1}$$

高斯-乔丹的最后一个步骤是把每行除以本身的主元，新的主元全部是 1。

因为 K 可逆，矩阵的左半部变成 I，K^{-1} 的三个列就在$[I \quad K^{-1}]$的第二部分：

除以 2
除以 3/2
除以 4/3

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & \mathbf{3/4} & \mathbf{1/2} & \mathbf{1/4} \\ 0 & \mathbf{1} & 0 & \mathbf{1/2} & \mathbf{1} & \mathbf{1/2} \\ 0 & 0 & \mathbf{1} & \mathbf{1/4} & \mathbf{1/2} & \mathbf{3/4} \end{bmatrix} = [I \quad \boldsymbol{x}_1 \quad \boldsymbol{x}_1 \quad \boldsymbol{x}_3] = [I \quad \boldsymbol{K^{-1}}]$$

从 3×6 矩阵$[K \quad I]$开始，最终得到$[I \quad K^{-1}]$。对于任意可逆矩阵 A，以下是整个高斯-乔丹消元法的程序(以简单一行来描述):

高斯-乔丹 **A^{-1} 乘 $[A \quad I]$ 得到 $[I \quad A^{-1}]$**

消元步骤把 A 变成 I 得到逆矩阵，对于大型矩阵，我们大概一点都不想要 A^{-1}。但是对小矩阵来说，还是值得知道这个技巧。对于 K^{-1} 我们加上三个观察：**一个重要范例**。

1. K 是**对称**，则 K^{-1} 也是对称的。
2. K 是三对角(**tridiagonal**)矩阵(只有 3 个非零对角线)，但是 K^{-1} 是一个没有 0 的浓密(dense)矩阵，这是另一个我们不常计算逆矩阵的原因。带状矩阵的逆矩阵通常都是浓密矩阵。
3. 主元的乘积是 2 (3/2) (4/3) = 4，这个 4 就是 K 的行列式。

K^{-1} 与除以 K 的行列式有关

$$K^{-1} = \frac{1}{4}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} \tag{8}$$

这就是为什么可逆矩阵的行列式不能是零：因为要除以它。

范例 4 $A = \begin{bmatrix} 2 & 3 \\ 4 & 7 \end{bmatrix}$，应用高斯-乔丹消元法求 A^{-1}。

$$[A \quad I] = \begin{bmatrix} 2 & 3 & 1 & 0 \\ 4 & 7 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 2 & 3 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \quad \text{这是} \begin{bmatrix} U & L^{-1} \end{bmatrix}$$

$$\to \begin{bmatrix} 2 & 0 & 7 & -3 \\ 0 & 1 & -2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 7/2 & -3/2 \\ 0 & 1 & -2 & 1 \end{bmatrix} \quad \text{这是} \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

范例 5 若 A 可逆的上三角矩阵，则 A^{-1} 也是。从 $AA^{-1} = I$ 开始。

1. A 乘(A^{-1} 的列 j)等于 I 的列 j，这个列 j 末端有 $n - j$ 个零。
2. 反向代入，保持这些 $n - j$ 个零在 A^{-1} 的列 j 的末端。
3. 把这些列$[* \ldots * \ 0 \ldots 0]^{\mathrm{T}}$ 放进 A^{-1}，这个矩阵就是上三角矩阵！

请尊重版权与译者的劳动成果，侵权必究!

$$A^{-1}=\begin{bmatrix}1&-1&0\\0&1&-1\\0&0&1\end{bmatrix}^{-1}=\begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}$$

列 1 与列 2 的末端
有 $3-j=2$ 与 1 个 0

代码 X = inv(A)利用 rref: reduced row echelon form(简化行阶梯形式)，见第三章。

```
I = eye(n);              % 定义 n×n 单位矩阵
R = rref([A   I];        % 在矩阵[A   I]上执行消元法
X = R(:, n + 1: n + n);  % 在 R 后面 n 列就是 X = A^-1
```

A 必须可逆，否则消元法就不能将 A 简化至 I (R 的左半部。)

高斯-乔丹说明了为什么计算 A^{-1} 的成本很高，我们对于它的 n 个列求解 n 个方程式，然而所有的方程式都跟左侧的同一个矩阵 A 有关(最大的工作量)，A^{-1} 的整体成本为 n^3 次乘法与减法，要求解一个 $A\boldsymbol{x}=\boldsymbol{b}$ 的成本为 $n^3/3$(见下一个段落)。

不用 A^{-1} 去求解 $A\boldsymbol{x}=\boldsymbol{b}$，我们处理一个列 $\boldsymbol{b}$ 去找一个列 $\boldsymbol{x}$。

奇异对比可逆

我们回到中心问题，哪些矩阵有逆矩阵？本段落的开始介绍了主元测试：当 A 有一整组的主元(允许行交换)时，A^{-1} 存在。我们可以用高斯-乔丹消元法来证明：

1. 利用 n 个主元，消元法可以求解全部的方程式 $A\boldsymbol{x}_i=\boldsymbol{e}_i$，列 $\boldsymbol{x}_i$ 会进入 A^{-1}，则 $AA^{-1}=I$ 且 A^{-1} 至少是个“**右逆矩阵**”。
2. 消元法事实上就是一系列矩阵的乘法：E's，P's，及 D^{-1}。

左逆矩阵 $\boldsymbol{C}$　　$$CA=(D^{-1}\ldots E\ldots P\ldots E)\,A=I \tag{9}$$

D^{-1} 除以主元，矩阵 E 产生在主元上方以及下方的零，P 在需要的时候交换行(见段落 2.7)。方程式(9)里面的乘积很明显就是 A 的**左逆矩阵**，利用 n 个主元，我们得到 $A^{-1}A=I$。

右逆矩阵等于左逆矩阵，这就是本段落开头提到的注释 2。所以具有一整组主元的方形矩阵永远有双边逆矩阵。

现在反过来证明若 $AC=I$，则 A 必定有 n 个主元。

1. 若 A 没有 n 个主元，消元法会得到一个零行。
2. 这些消元步骤由可逆矩阵 M 取得，所以 MA 的一行是零。
3. 若 $AC=I$，则 $MAC=M$，MA 的零行乘 C 得到 M 本身的零行。
4. 可逆矩阵 M 不可能有零行！若 $AC=I$，则 A 必定有 n 个主元。

上述的理论需要 4 个步骤，但是结果很短很重要。C 就是 A^{-1}。

消元法提供方形矩阵可逆性的完整试验，**当 A 有 n 个主元时，A^{-1} 确实存在(高斯-乔丹可以找到)**，以上的论点更说明了：

$$\text{若 } AC=I \quad \text{则} \quad CA=I \text{ 且 } C=A^{-1} \tag{10}$$

范例 6 若 L 是下三角且对角线都是 1，则 L^{-1} 也是。

三角矩阵可逆 当且仅当 没有对角线单元是 0

此处 L 有单元 1's，所以 L^{-1} 也有 1's。利用高斯-乔丹方法从 E_{32}, E_{31}, E_{21} 创建 L^{-1}。注意 L^{-1} 是如何包含奇特的单元 11，11 来自 $3\times5-4$。

应用高斯-乔丹在三角 L

$$\begin{bmatrix}1&0&0&1&0&0\\3&1&0&0&1&0\\4&5&1&0&0&1\end{bmatrix}=\begin{bmatrix}L&I\end{bmatrix}$$

$$\rightarrow\begin{bmatrix}1&0&0&1&0&0\\0&1&0&-3&1&0\\0&5&1&-4&0&1\end{bmatrix}\begin{matrix}\text{行2减去3乘行1}\\\text{行3减去4乘行1}\\\text{行3减去5乘行2}\end{matrix}$$

逆矩阵仍然是三角形

$$\rightarrow\begin{bmatrix}1&0&0&1&0&0\\0&1&0&-3&1&0\\0&0&1&11&-5&1\end{bmatrix}=\begin{bmatrix}I&L^{-1}\end{bmatrix}$$

辨识逆矩阵

正常来说，需要一些功夫来决定矩阵是否可逆，常见的方法是在消元过程中找到一整组的非零主元。(非零行列式来自这些主元的乘积。) 但是对于某些矩阵，你可以快速看出是否可逆，因为每个对角线数字 a_{ii} 对于该行其他非对角线单元，具有支配(dominate)的效应。

对角式支配(diagonally dominant)矩阵为可逆。每个 a_{ii} 比行 i 其他单元的总和还要大很多，对于每一行：

$$|a_{ii}|>\sum_{j\neq i}|a_{ij}| \text{ 就是说 } \quad |a_{ii}|>|a_{i1}|+|a_{i2}|+\ldots\left(\text{跳过 }|a_{ii}|\right)+\ldots+|a_{in}| \tag{11}$$

范例 A 是对角式支配$(3>2)$，B 不是(但是仍然可逆)，C 是奇异。

$$A=\begin{bmatrix}3&1&1\\1&3&1\\1&1&3\end{bmatrix}\quad B=\begin{bmatrix}2&1&1\\1&2&1\\1&1&3\end{bmatrix}\quad C=\begin{bmatrix}1&1&1\\1&1&1\\1&1&3\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

推论 选取任意非零向量 $\boldsymbol{x}$，假设它的最大分量是$|x_i|$，当 A 是对角支配矩阵时，不可能有 $A\boldsymbol{x}=\boldsymbol{0}$。因为如果 $A\boldsymbol{x}$ 的行 i 为零，需要有

$$a_{i1}x_1+\ldots+a_{ii}x_i+\ldots+a_{in}x_n=0$$

这些项的总和不可能是零！$|a_{ii}x_i|$(某个特定项)的大小比其他组合项要大很多：

$$\text{所有}|x_j|<|x_i| \qquad \sum_{j\neq i}|a_{ij}x_j|\le\left(\sum_{j\neq i}|a_{ij}|\ |x_i|\right)<|a_{ii}|\ |x_i| \qquad \text{因为 } \boldsymbol{a_{ii}} \text{ 是支配项}$$

上述说明只有在 $\boldsymbol{x}=\boldsymbol{0}$ 时，才有可能 $A\boldsymbol{x}=\boldsymbol{0}$，所以 A 是可逆。范例 B 也是可逆，但不是对角式支配矩阵：2 没有大于 1 + 1。

主要观念的复习

1. 逆矩阵得到 $AA^{-1}=I$ 且 $A^{-1}A=I$。
2. 矩阵 A 有逆矩阵，当且仅当 A 有 n 个主元(允许交换行)。
3. **重要。**若有非零向量 $\boldsymbol{x}$ 使得 $A\boldsymbol{x}=\boldsymbol{0}$，则 A 没有逆矩阵。
4. AB 的逆矩阵是反序乘积 $B^{-1}A^{-1}$，且$(ABC)^{-1}=C^{-1}B^{-1}A^{-1}$。
5. 高斯-乔丹方法求解 $AA^{-1}=I$ 得到 A^{-1} 的 n 个列。增广矩阵$[\,A\quad I\,]$可以使用行简化得到$[\,I\quad A^{-1}]$。
6. 对角式支配矩阵是可逆矩阵，每个$|a_{ii}|$支配它所在的行。

已解范例

2.5A 三角形差分矩阵 A 的逆矩阵，是三角形总和矩阵 S：

$$[A\ \ I]=\left[\begin{array}{ccc|ccc}1&0&0&1&0&0\\-1&1&0&0&1&0\\0&-1&1&0&0&1\end{array}\right]\rightarrow\left[\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&-1&1&0&0&1\end{array}\right]$$

$$\rightarrow\left[\begin{array}{ccc|ccc}1&0&0&1&0&0\\0&1&0&1&1&0\\0&0&1&1&1&1\end{array}\right]=\left[\begin{array}{ccc|ccc}I&&&A^{-1}\end{array}\right]=\left[\begin{array}{ccc|ccc}I&&&\text{总和矩阵}\end{array}\right]$$

如果我把 a_{13} 变成–1，则 A 的所有的行加起来都是 0，方程式 $A\boldsymbol{x}=\boldsymbol{0}$ 现在会有非零解 $\boldsymbol{x}=(1,1,1)$。一个明显的信号：**新的矩阵 $\boldsymbol{A}$ 不可逆**。

2.5B 下列矩阵有三个可逆矩阵，三个不可逆矩阵。当逆矩阵存在时，求出逆矩阵。其他三个不可逆矩阵，请给出理由(零行列式，主元太少，$A\boldsymbol{x}=\boldsymbol{0}$ 有非零解。)下列矩阵的顺序是 A, B, C, D, S, E。

$$\begin{bmatrix}4&3\\8&6\end{bmatrix}\quad\begin{bmatrix}4&3\\8&7\end{bmatrix}\quad\begin{bmatrix}6&6\\6&0\end{bmatrix}\quad\begin{bmatrix}6&6\\6&6\end{bmatrix}\quad\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix}\quad\begin{bmatrix}1&1&1\\1&1&0\\1&1&1\end{bmatrix}$$

解

$$B^{-1}=\frac{1}{4}\begin{bmatrix}7&-3\\-8&4\end{bmatrix}\qquad C^{-1}=\frac{1}{36}\begin{bmatrix}0&6\\6&-6\end{bmatrix}\qquad S^{-1}=\begin{bmatrix}1&0&0\\-1&1&0\\1&-1&1\end{bmatrix}$$

A 不可逆，因为行列式值= $4\times6-3\times8=24-24=0$。$D$ 不可逆，因为只有一个主元；第二行减第一行变成零。E 有两个相等的行(且第二列减去第一列是零)，换言之，$E\boldsymbol{x}=\boldsymbol{0}$ 的解可以是 $\boldsymbol{x}=(-1,1,1)$。

当然全部三个不可逆的理由都可以应用到 A, D, E。

2.5C 应用高斯-乔丹方法求三角帕斯卡矩阵 L 的逆矩阵。你可以看到帕斯卡三角形--**每个单元加上自身左侧的单元**放在该单元的下方。L 的单元是二项式系数(binomial coefficients)，第五行会变成 1, 4, 6, 4, 1。

$$\textbf{三角帕斯卡矩阵}\ L=\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&2&1&0\\1&3&3&1\end{bmatrix}=\textbf{abs(pascal (4, 1))}$$

解 高斯-乔丹从$[\,L\quad I\,]$开始，借由减去行 1 得到想要的 0：

$$\left[\boldsymbol{L}\quad\boldsymbol{I}\right]=\left[\begin{array}{cccc|cccc}\boldsymbol{1}&0&0&0&1&0&0&0\\\boldsymbol{1}&\boldsymbol{1}&0&0&0&1&0&0\\\boldsymbol{1}&\boldsymbol{2}&\boldsymbol{1}&0&0&0&1&0\\\boldsymbol{1}&\boldsymbol{3}&\boldsymbol{3}&\boldsymbol{1}&0&0&0&1\end{array}\right]\rightarrow\left[\begin{array}{cccc|cccc}1&0&0&0&1&0&0&0\\\boldsymbol{0}&1&0&0&\boldsymbol{-1}&1&0&0\\\boldsymbol{0}&2&1&0&\boldsymbol{-1}&0&1&0\\\boldsymbol{0}&3&3&1&\boldsymbol{-1}&0&0&1\end{array}\right]$$

下个阶段产生第二主元下方的 0，使用的乘数是 2 与 3。最后一个阶段从新行 4 减去 3 乘新行 3：

$$\rightarrow\left[\begin{array}{cccc|cccc}1&0&0&0&1&0&0&0\\0&1&0&0&-1&1&0&0\\0&\boldsymbol{0}&1&0&\boldsymbol{1}&\boldsymbol{-2}&1&0\\0&\boldsymbol{0}&3&1&\boldsymbol{2}&\boldsymbol{-3}&0&1\end{array}\right]\rightarrow\left[\begin{array}{cccc|cccc}1&0&0&0&\boldsymbol{1}&0&0&0\\0&1&0&0&\boldsymbol{-1}&1&0&0\\0&0&1&0&\boldsymbol{1}&\boldsymbol{-2}&1&0\\0&0&0&1&\boldsymbol{-1}&\boldsymbol{3}&\boldsymbol{-3}&\boldsymbol{1}\end{array}\right]=\left[\boldsymbol{I}\quad\boldsymbol{L^{-1}}\right]$$

所有的主元都是 1！所以每个行不需要除以主元得到 I。逆矩阵 L^{-1} 跟 L 很像，除了奇数对角线有负号。

同样模式可以扩展到 $n\times n$ 的帕斯卡矩阵，L^{-1} 具有交替(alternative)对角。

请尊重版权与译者的劳动成果，侵权必究!

问题集 2.5

1 求出 A，B，C 的逆矩阵。(直接或用 2×2 公式)。

$$A=\begin{bmatrix}0&3\\4&0\end{bmatrix}\qquad B=\begin{bmatrix}2&0\\4&2\end{bmatrix}\qquad C=\begin{bmatrix}3&4\\5&7\end{bmatrix}$$

2 借由尝试错误法(利用 1's 与 0's)，求出**排列矩阵**的逆矩阵 P^{-1}：

$$P=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}\quad 与\quad P=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}$$

3 求解 A^{-1} 的第一列(x, y)与第二列(t, z)：

$$\begin{bmatrix}10&20\\20&50\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}\quad 与\quad \begin{bmatrix}10&20\\20&50\end{bmatrix}\begin{bmatrix}t\\z\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}$$

4 尝试求解 $AA^{-1}=I$ 得到 A^{-1} 的列 1，证明 $\begin{bmatrix}1&2\\3&6\end{bmatrix}$ 不可逆：

$$\begin{bmatrix}1&2\\3&6\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}\qquad \begin{pmatrix}如果是不同的 A，有没有可能\\A^{-1}\text{的列1有解，列2却无解}\boldsymbol{?}\end{pmatrix}$$

5 求一个上三角 U(不是对角矩阵)，使得 $U^2=I$，也就是 $U=U^{-1}$。

6 (a) 若 A 可逆且 $AB=BC$，快速证明 $B=C$。

(b) 若 $A=\begin{bmatrix}1&1\\1&1\end{bmatrix}$，找出两个不同的矩阵使得 $AB=AC$。

7 (重要) 若 A 有行 1 + 行 2 = 行 3，证明 A 不可逆。

(a) 解释为什么 $A\boldsymbol{x}=(0, 0, 1)$不可能有解。执行方程式 1+方程式 2。

(b) 哪些右侧(b_1, b_2, b_3)可能允许 $A\boldsymbol{x}=\boldsymbol{b}$ 有一个解?

(c) 在消元过程中，方程式 3 发生什么？

8 若 A 有列 1 + 列 2 = 列 3，证明 A 不可逆。

(a) 对于 $A\boldsymbol{x}=\boldsymbol{0}$ 找出一个非零解 $\boldsymbol{x}$，矩阵是 3×3。

(b) 消元法保持列 1 + 列 2 = 列 3，解释为什么没有第三主元?

9 假设 A 可逆，你可以交换前两个行得到 B，请问 B 可逆吗？你要如何从 A^{-1} 得到 B^{-1}?

10 求逆矩阵(任意合法的方法都可以):

$$A=\begin{bmatrix}0&0&0&2\\0&0&3&0\\0&4&0&0\\5&0&0&0\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}3&2&0&0\\4&3&0&0\\0&0&6&5\\0&0&7&6\end{bmatrix}$$

11 (a) 找出可逆矩阵 A 与 B，使得 $A+B$ 不可逆。

(b) 找出奇异矩阵 A 与 B，使得 $A+B$ 可逆。

12 若乘积 $C=AB$ 是可逆(A 与 B 是方形)，则 A 本身是可逆。求出 A^{-1} 的公式，这个公式与 C^{-1} 与 B 有关。

13 若乘积 $M=ABC$ 是可逆(A 与 B 与 C 是方形)，则 B 是可逆(A 与 C 也是)。找出 B^{-1} 的公式，这个公式与 M^{-1} 与 A 与 C 有关。

14 如果你把 A 的行 1 加到行 2 得到 B，你如何从 A^{-1} 求得 B^{-1}？

注意顺序。$B=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}A\end{bmatrix}$，则 B 的逆矩阵是_____。

15 证明有一个列都是 0 的矩阵没有逆矩阵。

16 计算$\begin{bmatrix}a&b\\c&d\end{bmatrix}$乘 $\begin{bmatrix}d&-b\\-c&a\end{bmatrix}$。若 $ad\neq bc$，求每个矩阵的逆矩阵。

17 (a) 什么样的矩阵 E 有相同的三个步骤的效应？从行 2 减去行 1，从行 3 减去行 1，然后从行 3 减去行 2。

(b) 什么样的单一矩阵 L 有相同的三个反向步骤的效应？把行 2 加到行 3，把行 1 加到行 3，然后把行 1 加到行 2。

18 若 B 是 A^2 的逆矩阵，证明 AB 是 A 的逆矩阵。

19 求数值 a 与 b 得到 5 * eye(4) – ones(4, 4) 的逆矩阵：

$$\begin{bmatrix}4&-1&-1&-1\\-1&4&-1&-1\\-1&-1&4&-1\\-1&-1&-1&4\end{bmatrix}^{-1}=\begin{bmatrix}a&b&b&b\\b&a&b&b\\b&b&a&b\\b&b&b&a\end{bmatrix}$$

求数值 a 与 b 得到 6 * eye(5) – ones(5, 5) 的逆矩阵？

20 证明 A = 4 * eye(4) – ones(4, 4) 不可逆。执行乘法 A * ones(4, 1)。

21 2×2 矩阵的单元可以是 1 或 0 的矩阵有 16 个，其中有几个矩阵是可逆？

请尊重版权与译者的劳动成果，侵权必究!

问题 22-28 有关计算 A^{-1} 的高斯-乔丹方法。

22 如同 A 简化至 I 一样，把 I 变成 A^{-1} (借由行运算)：

$$[A\ \ I]=\begin{bmatrix}1&3&1&0\\2&7&0&1\end{bmatrix} \text{ 与 } [A\ \ I]=\begin{bmatrix}1&4&1&0\\3&9&0&1\end{bmatrix}$$

23 追随教科书上的 3×3 例子，但是 A 的单元都变成正号。消去主元的上方与下方，简化[A　I]成为[I　A^{-1}]：

$$[A\ \ I]=\begin{bmatrix}2&1&0&1&0&0\\1&2&1&0&1&0\\0&1&2&0&0&1\end{bmatrix}$$

24 对[U　I]应用高斯-乔丹法，求出上三角 U^{-1}：

$$UU^{-1}=I \qquad \begin{bmatrix}1&a&b\\0&1&c\\0&0&1\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1&\boldsymbol{x}_2&\boldsymbol{x}_3\\&&\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

25 对[A　I]与[B　I]应用消元法，求出 A^{-1} 与 B^{-1} (如果存在)：

$$A=\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix} \text{ 与 } B=\begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix}$$

26 什么样的矩阵 E_{21} 与 E_{12} 与 D^{-1} 可以简化 $A=\begin{bmatrix}1&2\\2&6\end{bmatrix}$ 得到单位矩阵？计算 $D^{-1}E_{12}E_{21}$ 求 A^{-1}。

27 利用高斯-乔丹法至[A　I]，求出 A 的逆矩阵：

$$A=\begin{bmatrix}1&0&0\\2&1&3\\0&0&1\end{bmatrix} \text{ 与 } A=\begin{bmatrix}1&1&1\\1&2&2\\1&2&3\end{bmatrix}$$

28 交换行之后继续使用高斯-乔丹法，求出 A^{-1}：

$$[A\ \ I]=\begin{bmatrix}0&2&1&0\\2&2&0&1\end{bmatrix}$$

29 是非题 (如果错，举出一个反例；如果对，说明理由)：

(a) 一个 4×4 矩阵有一个零行，则不可逆。

(b) 主对角线单元都是 1 的矩阵必定可逆。

(c) 若 A 可逆，则 A^{-1} 及 A^2 都是可逆。

30 (推荐) 若 $a \neq 0$ 且 $a \neq b$，证明 A 是可逆(求主元或是 A^{-1})，然后找出 3 个数值 c 使得 C 不可逆。

$$A = \begin{bmatrix} a & b & b \\ a & a & b \\ a & a & a \end{bmatrix} \qquad C = \begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix}$$

31 这个矩阵有非凡的逆矩阵，应用消元法至$[A \quad I]$，求出 A^{-1}。延伸至 5×5 的“交替矩阵”，猜测逆矩阵；然后执行乘法进行确认。

$$\text{逆反} \quad A = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{以及求解 } A\boldsymbol{x} = (1, 1, 1, 1)。$$

32 假设矩阵 P 与 Q 有与单位矩阵相同的行，只是顺序不同，他们是“排列矩阵”。求解$(P - Q)\boldsymbol{x} = \boldsymbol{0}$ 证明 $P - Q$ 是奇异。

33 针对下列的方块矩阵，求出并验证逆矩阵(假设存在)：

$$\begin{bmatrix} I & \boldsymbol{0} \\ C & I \end{bmatrix} \quad \begin{bmatrix} A & \boldsymbol{0} \\ C & D \end{bmatrix} \quad \begin{bmatrix} \boldsymbol{0} & I \\ I & D \end{bmatrix}$$

34 如果一个 4×4 矩阵 A 的每一行都包含某种顺序的数字 0, 1, 2, 3，请问这个矩阵可逆吗？如果 B 的每一行都包含某种顺序的数字 0, 1, 2, –3，请问可逆吗?

35 在已解范例 **2.5C** 中，三角帕斯卡矩阵 L 有 $L^{-1} = DLD$，其中对角矩阵 D 有交替单元 1, –1, 1, –1，则 $LDLD = I$，求 LD = pascal(4, 1)的逆矩阵。

36 希尔伯特矩阵有 $H_{ij} = 1/(i + j - 1)$，从 MATLAB 求 6×6 的逆矩阵 invhilb(6)。然后计算 inv(hilb(6))，电脑是不会错误的，但是这两个为什么会有不同?

37 (a) 在 MATLAB 使用 inv(P)求 4×4 对称矩阵 P = pascal(4)的逆矩阵。

(b) 创建帕斯卡下三角 L = abs(pascal(4, 1))，并测试 $P = LL^{\mathrm{T}}$。

38 若 A = ones(4)且 $\boldsymbol{b}$ = rand(4, 1)，MATLAB 会怎么告诉你 $A\boldsymbol{x} = \boldsymbol{b}$ 无解？对于特殊的 $\boldsymbol{b}$ = ones(4, 1)，$A\boldsymbol{x} = \boldsymbol{b}$ 的哪个解可以由 $A\backslash\boldsymbol{b}$ 得到？

挑战问题

39 (推荐) A 是一个 4×4 矩阵，其中对角线都是 1，对角线上方是–a, –b, –c。求此**双对角**(bidiagonal)矩阵的逆矩阵 A^{-1}。

请尊重版权与译者的劳动成果，侵权必究!

40 假设 E_1，E_2，E_3 都是 4×4 单位矩阵，除了 E_1 的第一列有 a，b，c；E_2 的第二列有 d，e；E_3 的第三列有 f(数字都在 1's 的下方)。执行 $L = E_1E_2E_3$ 证明这些非零数字都会进入 L。

$E_1\,E_2\,E_3$ 是消元法的反向序列(E_3 先作用)，但是 $E_1\,E_2\,E_3 = L$ 是正确的顺序，可以逆反消元法以及回复 A。

41 如果第二差分矩阵从 $T_{11} = 1$(不是 $K_{11} = 2$)开始，他们会有美丽的逆矩阵。下列是 3×3 **“三对角矩阵”** T 与它的逆矩阵：

$$T = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

一个方法是对[T I]执行高斯-乔丹消元法，我偏好将 T 写成第一差分 L 乘 U 的乘积。在已解范例 **2.5A** 中，L 与 U 的逆矩阵就是总和矩阵，所以 $T = LU$ 以及 $T^{-1} = U^{-1}L^{-1}$ ：

$$T = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 & 0 \\ & 1 & -1 \\ & & 1 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix}$$

差　　差　　和　　和

问题　(4×4) T 的主元是什么？它的 4×4 逆矩阵是什么？反向顺序 UL 可以得到什么矩阵 T^*？矩阵 T^*的逆矩阵是什么？

42 此处还有两个差分矩阵，都很重要，但是他们可逆吗？

$$\textbf{循环}\ C = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \qquad \textbf{自由端}\ F = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

43 对**方块矩阵**执行消元法。当你从第二方块行[C　D]减去 CA^{-1} 乘第一方块行[A　B]时，**“舒尔补充”** S 变成：

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & S \end{bmatrix} \qquad \begin{matrix} A\text{与}D\text{是方形} \\ S = D - CA^{-1}B \end{matrix}$$

从方块列 2 减去方块列 1 **右乘** $A^{-1}B$：

$$\begin{bmatrix} A & B \\ 0 & S \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = ? \qquad \text{当}\begin{bmatrix} A & B \\ C & I \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 \\ 4 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}\text{，求 } S\text{。}$$

方块主元是 A 与 S。如果他们可逆，那么[A　B; C　D]也是。

44 **等式 $A(I + BA) = (I + AB)A$** 如何与 $I + BA$ 以及 $I + AB$ 产生关联？两个都是可逆或是两个都是奇异：不明显。

2.6 消元=分解：$A = LU$

1 每个消元步骤 E_{ij} 被 L_{ij} 逆反，非对角线的 $-l_{ij}$ 变成 $+l_{ij}$。

2 整体前向消元程序(没有行交换)被 L 逆反：

$$L = (L_{21}\, L_{31} \ldots L_{n1})\,(L_{32} \ldots L_{n2})\,(L_{43} \ldots L_{n3}) \ldots (L_{n\,n-1})$$

3 乘积矩阵 L 仍然是下三角，**每个乘数 l_{ij} 位于行 i 列 j。**

4 从 U 回复到原始矩阵 A：$A = LU =$ (下三角) (上三角)。

5 作用在 $A\boldsymbol{x} = \boldsymbol{b}$ 的消元法，得到 $U\boldsymbol{x} = \boldsymbol{c}$，使用反向代入求解 $U\boldsymbol{x} = \boldsymbol{c}$。

6 求解三角系统需要 $n^2/2$ 次乘法/减法，消元法得到 U 需要 $n^3/3$。

学生通常会说数学课程太理论了，好吧，本段落绝对不是，大部分都是纯应用，目标是以最有用的方法来描述高斯消元法。当你仔细观察线性代数的许多关键概念时，实际上就是矩阵的**分解**(factorization)，原始矩阵 A 变成两个或是三个特殊矩阵的乘积。第一个分解——也是实际上最重要的分解——来自消元法。**因子 L 与 U 都是三角矩阵，来自消元法的分解是 $A = LU$。**

我们已经知道 U 是上三角矩阵，主元位于对角线，消元步骤使得 A 变成 U。我们会说明如何利用下三角矩阵 L 来逆反这些步骤(U 回复到 A)，L 的单元恰好是乘数 l_{ij}——从行 i 减去乘数乘主元行 j。

从 2×2 的例子开始，矩阵 A 包含 2, 1, 6, 8，要消去的数字是 6。从行 2 减去 3 乘行 1，这个前向步骤就是乘数 $l_{21} = 3$ 的消元矩阵 E_{21}。从 U 回复到 A 的逆反步骤是 $L^{-1} = E_{21}^{-1}$ (变成加法，乘数为+3):

前向 A 到 U：$E_{21}A = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = U$

反向 U 到 A：$E_{21}^{-1}U = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = A$

第二行就是我们的分解 $LU = A$，我们写成 L 代替 E_{21}^{-1}。较大的矩阵需要更多的 E's，**L 包含他们所有的逆矩阵**。

A 到 U 的每个步骤就是左乘一个矩阵 E_{ij}，使得位置(i, j)变成 0。为了要保持清晰，我们采用最常见的案例——没有交换行的行为。若 A 是 3×3，我们左乘 E_{21}, E_{31}, E_{32}，乘数 l_{ij} 使得(2, 1), (3, 1), (3, 2)位置变成 0——都在对角线下方，消元法得到上三角 U 时停止。

请尊重版权与译者的劳动成果，侵权必究!

现在把 E's 移到另一边，变成逆矩阵乘 U：

$(E_{32}\,E_{31}E_{21})A = U$ **变成** $A = (E_{21}^{-1}\,E_{31}^{-1}\,E_{32}^{-1})U$ **得到** $A = LU$

逆反的顺序必须与前向相反，三个逆矩阵的乘积等于 L。我们得到 $\boldsymbol{A = LU}$，现在我们停止这个主题。

说明与范例

第一点：每个逆矩阵 E^{-1} 是下三角，非对角线单元是 l_{ij}，用来恢复 $-l_{ij}$ 的减法效应。E 与 E^{-1} 的主对角线都是 1，前面的例子 $l_{21} = 3$, $E = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ 与 $L = E^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$。

第二点：方程式(2)展示下三角矩阵(E_{ij} 的乘积)乘 A，也展示所有 E_{ij}^{-1} 乘 U 回复 A，逆矩阵乘积得到的下三角矩阵就是 L。【方程式(2)指的是上方蓝底的公式)】

处理这些逆矩阵的一个理由，是我们想要分解 A 而不是分解 U。“逆反形式”得到 $A = LU$。另一个理由是我们好像额外获得了一些超过我们应得的东西，这就是第三点原因，显示 L 非常完美(exactly right)。

第三点：每个 l_{ij} 都会落在位置 i, j，在逆矩阵的乘积 L 也是一样。通常矩阵乘法会把数字的位置弄乱，在这里却没有发生。逆矩阵的顺序是正确的，使得 l 不变。这个理由会在后面的方程式(2)说明。【这里指的是下一页的方程式(2)】

每个 E^{-1} 的对角线都是 1，最后的优点是 L 也具备这个的性质。

$\boldsymbol{A = LU}$ **这是没有交换行的消元法**，上三角 U 的主元在对角线上，下三角 L 的对角线都是 1，**乘数** l_{ij} 在 $\boldsymbol{L}$ **的对角线下方**。

范例 1 消元法从行 2 减去 1/2 乘行 1，最后的步骤是从行 3 减去 2/3 乘行 2。下三角 L 的 $l_{21} = 1/2$ 及 $l_{32} = 2/3$。LU 乘积产生 A：

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix} = LU$$

因为 $A(3, 1)$位置的单元是 0，所以(3, 1)位置的乘数是 0，不需要操作。

范例 2 把 A 的左上角单元从 2 变成 1 得到 B，所有的主元全部变成 1，所有的乘数都是 1。持续这样的模式，当 B 是 4×4 时：

特殊模式 $B = \begin{bmatrix} \mathbf{1} & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 0 & 1 & 1 & \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ & 1 & 1 & 0 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$

这些 LU 范例说明了额外的东西，在实际上非常有用。假设没有交换行，什么时候我们可以预测 L 与 U 中的 0?

当 A 的某一行从 0 开始，L 对应的那一行也是一样。

当 A 的某一列从 0 开始，U 对应的那一列也是一样。

若某一行从 0 开始，我们不需要消元步骤。L 有 0，可以节省电脑时间，同样的从 0 开始的列也在 U 发生。但是请了解：当消元法前向清除时，矩阵中间区域的 0 好像会被填满数字。我们现在说明为什么 L 有乘数 l_{ij} 在相应位置，不会混乱。

为什么 $A = LU$ 的关键理由：问问自己关于主元行被下方行减去时，他们是不是 A 的原始行？不是！消元的过程可能会改变他们。那么他们是 U 的行吗？是的！主元行不会再改变。计算 U 的第三行时，我们减去 U 的较早行的乘积(不是 A 的行):

$$U \text{ 的行 } 3 = (A \text{ 的行 } 3) - l_{31}(U \text{ 的行 } 1) - l_{32}(U \text{ 的行 } 2) \tag{1}$$

改写上述方程式，观察行$[l_{31} \quad l_{32} \quad 1]$如何乘 U：

$$A \text{ 的行 } 3 = l_{31}(U \text{ 的行 } 1) + l_{32}(U \text{ 的行 } 2) + 1(U \text{ 的行 } 3) \tag{2}$$

恰好就是$(A = LU)$的行 3。L 的行 3 也有分量 $l_{31}, l_{32}, 1$。不论 A 是任何大小，所有的行看起来都像这样。没有交换行，我们有 $A = LU$。

来自 LDU 较好的平衡　因为 U 有主元在对角线而 L 的对角线都是 1，所以 $A = LU$ “不是对称”，这个是容易改变的。U 除以一个对角矩阵 D，它的对角线包含所有 U 的主元。U 会变成新的三角矩阵，它的对角线都是 1's:

$$U\text{分割成} \begin{bmatrix} d_1 & & & \\ & d_1 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} 1 & u_{12}/d_1 & u_{13}/d_1 & \bullet \\ & 1 & u_{23}/d_2 & \bullet \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

新的三角矩阵使用相同的字母 U 很方便(有一些混淆)，它的 1's 都在对角线上(像 L)。与正常的 LU 不同，新形式有 D 在中间：**(下三角 L)乘(对角 D)乘(上三角 U)**。

三角分解可以写成　$A = LU$　或　$A = LDU$

你只要看到 LDU，就知道 U 的对角线都是 1，每一行都除以第一个非零单元——主元，然后 L 与 U 在 LDU 被同等对待。

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & 8 \\ 0 & \mathbf{5} \end{bmatrix} \quad \text{进一步分解成} \quad \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & \\ & \mathbf{5} \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} \tag{3}$$

主元 2 与 5 进入 D。把行除以 2 与 5，在新的 U 留下了$[1 \quad 4]$与$[0 \quad 1]$而且 U 的对角线都是 1。乘数 3 仍然在 L 里面。

请尊重版权与译者的劳动成果，侵权必究!

我自己的课程有时候到这里会直接跳到 2.7，下面的内容展示如何组织消元代码，以及需要多长时间。如果手边有 MATLAB(或其他软件)，你可以以秒计来测量所需的电脑时间。

一个方形系统 = 两个三角系统

矩阵 L 包含高斯消元法的记忆，它保存从下方行减去乘主元行的乘数，我们何时需要这项记录，又该如何在求解 $A\boldsymbol{x} = \boldsymbol{b}$ 时使用?

当存在右侧的 $\boldsymbol{b}$ 时，我们需要 L。因子 L 与 U 完全由左侧(矩阵 A)决定，在 $A\boldsymbol{x} = \boldsymbol{b}$ 的右侧，我们使用 L^{-1}，然后是 U^{-1}。这个**求解**步骤处理两个三角矩阵。

1 **分解** (对左侧的矩阵 A 执行消元法，得到 L 与 U)。

2 **求解** (利用 L 对 $\boldsymbol{b}$ 执行前向消元法，利用 U 反向代入得到 x)。

较早时我们同时处理 A 与 $\boldsymbol{x}$，这个没有问题——就是增广成矩阵$[A \quad \boldsymbol{b}]$。但是更多的电脑代码会把两侧分开。L 与 U 有消元法的记忆，只要我们想就可以处理 $\boldsymbol{b}$。LAPACK 的使用手册注明“这种情况很普通，节省非常重要，求解一个单一系统只需要一个子程序就够了，不需要其他供应。”

“求解”如何对 $\boldsymbol{b}$ 工作? 首先应用前向消元法至右侧(乘数储存在 L，现在可以使用)，会把原来的 $\boldsymbol{b}$ 变成新右侧 $\boldsymbol{c}$，现在是求解 $L\boldsymbol{c} = \boldsymbol{b}$，然后反向代入求解 $U\boldsymbol{x} = \boldsymbol{c}$，原始系统被分解成两个三角系统：

前向与反向 求解 $L\boldsymbol{c} = \boldsymbol{b}$，然后求解 $U\boldsymbol{x} = \boldsymbol{c}$ (4)

要检验 $\boldsymbol{x}$ 是正确的，令 L 乘 $U\boldsymbol{x} = \boldsymbol{c}$，$LU\boldsymbol{x} = L\boldsymbol{c}$ 就是 $A\boldsymbol{x} = \boldsymbol{b}$。

强调：这些步骤没有新的东西，这些只是我们一直在做的事情，前向消元时，我们就是在求解三角系统 $L\boldsymbol{x} = \boldsymbol{b}$，然后反向代入得到 $\boldsymbol{x}$。下面的例子说明我们实际上在做什么。

范例 3 对 $A\boldsymbol{x} = \boldsymbol{b}$ 执行前向消元(往下)，在 $U\boldsymbol{x} = \boldsymbol{c}$ 结束：

$$A\boldsymbol{x} = \boldsymbol{b} \quad \begin{matrix} u + 2v = 5 \\ 4u + 9v = 21 \end{matrix} \quad \text{变成} \quad \begin{matrix} u + 2v = 5 \\ v = 1 \end{matrix} \quad U\boldsymbol{x} = \boldsymbol{c}$$

乘数是 4，储存在 L，右侧使用这个 4 使得 21 变成 1：

$Lc = b$ 下三角系统 $\begin{bmatrix}1 & 0\\ 4 & 1\end{bmatrix}[c]=\begin{bmatrix}5\\ 21\end{bmatrix}$ 得到 $c=\begin{bmatrix}5\\ 1\end{bmatrix}$

$Ux = c$ 上三角系统 $\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix}[x]=\begin{bmatrix}5\\ 1\end{bmatrix}$ 得到 $x=\begin{bmatrix}3\\ 1\end{bmatrix}$

L 与 U 用去 n^2 的储存位置，这是原来 A 所用的(现在可以遗忘)。

消元法的成本

一个非常实际的问题是成本--或是计算时间，我们可以在个人电脑上求解 1000 个方程式，如果 n = 100,000 呢? (A 是稀疏还是浓密呢?) 科学计算常常会遇到大型系统，其中三维空间的问题很容易会有一百万个未知数。我们可以让计算不分昼夜的运行，但是我们没办法等待 100 年。

消元法的第一个阶段得到列 1 的第一主元下方的 0，要得到主元行下方的新单元值，需要一个乘法与一个减法，我们算出第一阶段需要 n^2 个乘法与 n^2 个减法，实际上会小一点，$n^2 - n$，因为行 1 不需要改变。

下个阶段会把第二列位于第二主元下方的单元清除，工作矩阵的大小变成 $n - 1$，估计这个阶段需要$(n-1)^2$ 个乘法与减法。得到 U 所需要的粗估计算将是一串平方的总和 $n^2 + (n-1)^2 + ... + 2^2 + 1^2$。

上面的平方和有一个确切的公式 $\frac{1}{3}n\left(n+\frac{1}{2}\right)(n+1)$，当 n 很大时，1/2 与 1 就不重要了，关键数值就是 $n^3/3$。平方和好像 x^2 的积分，积分由 0 到 n 等于 $n^3/3$：

对 A 执行消元法大概需要 $n^3/3$ 次乘法 与 $n^3/3$ 次减法

右侧的 $\boldsymbol{b}$ 又会如何? 继续往前看，我们从下方的分量 $b_2, ..., b_n$ 减去 b_1 的倍数，需要 $n - 1$ 步，因为 b_1 没有参与。所以第二阶段需要 $n - 2$ 步，前向消元的最后一个阶段需要 1 步。

现在开始反向代入，计算 x_n 需要 1 步(除以最后一个主元)，求得下个未知数需要 2 步。当我们求得 x_1 时需要 n 步($n - 1$ 个其他未知数的代入，以及除以第一主元)。右侧从 $\boldsymbol{b}$ 到 $\boldsymbol{c}$ 到 $\boldsymbol{x}$ 的全部计算次数--前向到底部，反向到顶部--恰好是 n^2。

$$[(n-1) + (n-2) + ... + 1] + [1 + 2 + ... + (n-1) + n] = n^2 \qquad (5)$$

要得到总和，把$(n-1)$与 1 配对，$(n-2)$与 2 配对，这样的配对全部有 n 个，每项都等于 n，得到 n^2。右侧的成本比左侧小很多。

求解 每个右侧需要 n^2 个乘法 与 n^2 个减法

一个带状矩阵 B 的主对角线的上方或下方只有ω个非零对角线，位于带状之外的 0 单元在消元过程中都保持 0(在 L 与 U 都是 0)。

请尊重版权与译者的劳动成果，侵权必究!

清除第一列需要 ω^2 次的乘法与减法(在主元下产生 ω 个 0，每个 0 使用一个长度为 ω 的主元行)。清除全部 n 个列得到 U，刚好需要 $n\omega^2$，可以节省很多时间。

带状矩阵	***A* 到 *U*** $n^3/3$ 简化至 $n\omega^2$	**求解** n^2 简化至 $2n\omega$

一个三对角矩阵(带宽 $\omega = 1$)允许非常快速的运算，不必须要储存 0！

本书的网址有“教学代码”，可以将 A 分解成 LU 以及求解 $A\boldsymbol{x} = \boldsymbol{b}$。职业等级的代码会往下观察每个列，找到最大可用的主元，完成行的交换，降低数值的舍去误差。

MATLAB 的反斜线指令 $x = A\backslash b$ 结合分解与求解得到 x。

求解 $Ax = b$ 需要多长时间？ 对于阶数为 $n = 1000$ 的随机矩阵而言，个人电脑上的典型时间大概是 1 秒，当 n 放大两倍，时间就要放大 8 倍。职业性代码请查看 netlib.org。

依据 n^3 的规则，矩阵放大 10 倍(阶数 10,000)，需要 1000 秒；阶数 100,000 的矩阵需要一百万秒，如果没有超级电脑，时间成本太高了，但是不要忘了这些矩阵是当做满矩阵来看，实际上大部分矩阵都是稀疏矩阵(有很多 0)，这种情形下 $A = LU$ 会快速许多。

主要观念的复习

1. 高斯消元法(没有交换行)分解 A 变成 L 乘 U。
2. 下三角 L 包含用来乘主元行的乘数 l_{ij}，这些乘数使得 A 变成 U。乘积 LU 反加这些行回复 A。
3. 在右侧，我们求解 $L\boldsymbol{c} = \boldsymbol{b}$(前向)，以及 $U\boldsymbol{x} = \boldsymbol{c}$(反向)。
4. 分解：在左侧，有 $(n^3 - n)/3$ 个乘法与减法。
5. 求解：在右侧，有 n^2 个乘法与减法。
6. 对于带状矩阵来说，会把 $n^3/3$ 简化至 $n\omega^2$ 以及 n^2 简化至 $2n\omega$。

已解范例

2.6A 下三角帕斯卡矩阵 L 包含著名的“帕斯卡三角形”。在范例 **2.5C** 中，高斯-乔丹将 L 逆反，在此我们分解帕斯卡。

对称帕斯卡矩阵 *P* 是三角帕斯卡矩阵 *L* 与 *U* 的乘积。对称 *P* 矩阵冠名帕斯卡三角，所以每个单元都是上方单元与左方单元的和。在 MATLAB 中，$n\times n$ 的对称 *P* 写成 pascal(n)。

问题： 建立惊奇的下-上三角分解 $P = LU$。

$$\text{pascal}(4)=\begin{bmatrix}1&1&1&1\\1&2&3&4\\1&3&6&10\\1&4&10&20\end{bmatrix}=\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&2&1&0\\1&3&3&1\end{bmatrix}\begin{bmatrix}1&1&1&1\\0&1&2&3\\0&0&1&3\\0&0&0&1\end{bmatrix}=LU$$

请预测并检验 5×5 的帕斯卡矩阵的下一行与下一列。

解 你可以计算 *LU* 得到 *P*，最好是从对称 *P* 开始，利用消元法得到上三角 *U*：

$$P=\begin{bmatrix}1&1&1&1\\1&2&3&4\\1&3&6&10\\1&4&10&20\end{bmatrix}\to\begin{bmatrix}1&1&1&1\\0&1&2&3\\0&2&5&9\\0&3&9&19\end{bmatrix}\to\begin{bmatrix}1&1&1&1\\0&1&2&3\\0&0&1&3\\0&0&3&10\end{bmatrix}\to\begin{bmatrix}1&1&1&1\\0&1&2&3\\0&0&1&3\\0&0&0&1\end{bmatrix}=U$$

前述步骤用到的乘数 l_{ij} 会完美的进入 *L*，则 *P*= *LU* 是一个特别整洁有序的例子。注意 *U* 对角线上的主元都是 1。

下个段落会说明对称性质如何在三角 *L* 与 *U* 之间产生特殊关系，对于帕斯卡来说，*U* 是 *L* 的转置。

你可以预期 MATLAB 的指令 lu(pascal(4)) 可以得到这些 *L* 与 *U*，可惜这个不会发生，因为 lu 子程序会在每一列选取最大的主元，因此第二个主元会从 1 变成 3，但是**乔里斯基(Cholesky)分解**不会产生行交换：*U* = chol(pascal)。

对于所有大小的帕斯卡矩阵来说，$P = LU$ 的完整证明是相当迷人的。论文“Pascal Matrices”在课程网页 web.mit.edu/18.06 可以查阅，在 MIT 的公开课程集–ocw.mit.edu 可以看到。这些帕斯卡矩阵具有如此非凡的性质––后面还会见到。

2.6B 问题：求解 $P\boldsymbol{x} = \boldsymbol{b} = (1, 0, 0, 0)$，方程式右侧 $= I$ 的列，表示 $\boldsymbol{x}$ 会是 P^{-1} 的第一列。这就是高斯-乔丹消元法，会适配 $PP^{-1} = I$ 的列。我们已经知道帕斯卡矩阵 *L* 与 *U* 是 *P* 的因子：

两个三角系统 $L\boldsymbol{c} = \boldsymbol{b}$ (前向) $U\boldsymbol{x} = \boldsymbol{c}$ (反向)

解： 下三角系统 $L\boldsymbol{c} = \boldsymbol{b}$ 由上至下求解：

$$\begin{array}{ll}c_1 & =1\\ c_1+c_2 & =0\\ c_1+2c_2+c_3 & =0\\ c_1+3c_2+3c_3+c_4 & =0\end{array}\quad\text{得到}\quad\begin{array}{l}c_1=+1\\ c_2=-1\\ c_3=+1\\ c_4=-1\end{array}$$

请尊重版权与译者的劳动成果，侵权必究!

左乘 L^{-1} 执行前向消元，得到上三角系统 $U\boldsymbol{x}=\boldsymbol{c}$，一直都是利用反向代入求解 $\boldsymbol{x}$，由底部至顶端：

$$\begin{aligned} x_1+x_2+x_3+x_4 &= 1 \\ x_2+2x_3+3x_4 &= -1 \\ x_3+3x_4 &= 1 \\ x_4 &= -1 \end{aligned} \quad \text{得到} \quad \begin{aligned} x_1 &= +\mathbf{4} \\ x_2 &= -\mathbf{6} \\ x_3 &= +\mathbf{4} \\ x_4 &= -\mathbf{1} \end{aligned}$$

我看到了 $\boldsymbol{x}$ 的模式，但是我不知道来自何处。试试 inv(pascal(4))。

问题集 2.6

问题 1-14 计算 $A=LU$ 的分解(也计算 $A=LDU$)。

1 (重要) 前向消元法把 $\begin{bmatrix}1&1\\2&1\end{bmatrix}\boldsymbol{x}=\boldsymbol{b}$ 变成三角 $\begin{bmatrix}1&1\\0&1\end{bmatrix}\boldsymbol{x}=\boldsymbol{c}$：

$$\begin{aligned} x+y &= \mathbf{5} \\ x+2y &= \mathbf{7} \end{aligned} \rightarrow \begin{aligned} x+y &= \mathbf{5} \\ y &= \mathbf{2} \end{aligned} \qquad \begin{bmatrix}1&1&\mathbf{5}\\1&2&\mathbf{7}\end{bmatrix} \rightarrow \begin{bmatrix}1&1&\mathbf{5}\\0&1&\mathbf{2}\end{bmatrix}$$

这个步骤从行 2 减去 $l_{21}=$ ____乘行 1，反向步骤把 l_{21} 乘行 1 加到行 2。反向步骤的矩阵是 $L=$_____。L 乘三角系统 $\begin{bmatrix}1&1\\0&1\end{bmatrix}\boldsymbol{x}=\begin{bmatrix}5\\2\end{bmatrix}$ 得到_____ = _____。以字母表示，L 乘 $U\boldsymbol{x}=\boldsymbol{c}$ 得到_____。

2 写下问题 1 中 2×2 三角系统 $L\boldsymbol{c}=\boldsymbol{b}$ 与 $U\boldsymbol{x}=\boldsymbol{c}$。检验 $\boldsymbol{c}=(5,2)$是第一方程式的解，再求第二方程式的解 $\boldsymbol{x}$。

3 (现在是 3×3) 前向消元法把 $A\boldsymbol{x}=\boldsymbol{b}$ 变成三角 $U\boldsymbol{x}=\boldsymbol{c}$：

$$\begin{aligned} x+y+z &= 5 \\ x+2y+3z &= 7 \\ x+3y+6z &= 11 \end{aligned} \qquad \begin{aligned} x+y+z &= 5 \\ y+2z &= 2 \\ 2y+5z &= 6 \end{aligned} \qquad \begin{aligned} x+y+z &= 5 \\ y+2z &= 2 \\ z &= 2 \end{aligned}$$

$U\boldsymbol{x}=\boldsymbol{c}$ 中的方程式 $z=2$ 来自原始系统 $A\boldsymbol{x}=\boldsymbol{b}$ 中的 $x+3y+6z=11$，步骤是减去 $l_{31}=$ ____乘方程式 1，以及 $l_{32}=$ ____乘最后的方程式 2。反向运算回复 A 与 $\boldsymbol{b}$ 的最后一行的[1 3 6 11]，以及来自 U 与 $\boldsymbol{c}$ 最终的[1 1 1 5]，[0 1 2 2]，[0 0 1 2]：

$$[A \quad \boldsymbol{b}]\text{的行 } 3 = [U \quad \boldsymbol{c}]\text{的 } (l_{31}\text{ 行 }\mathbf{1} + l_{32}\text{ 行 }\mathbf{2} + 1\text{ 行 }\mathbf{3})$$

写成矩阵形式就是被 L 乘，所以 $A=LU$ 以及 $\boldsymbol{b}=L\boldsymbol{c}$。

4 问题 3 的 3×3 三角系统 $L\boldsymbol{c}=\boldsymbol{b}$ 与 $U\boldsymbol{x}=\boldsymbol{c}$ 为何？检验 $\boldsymbol{c}=(5,2,2)$满足第一个方程式，再求解第二个的 $\boldsymbol{x}$。

5 什么样的矩阵 E 使得 A 变成三角形式 $EA = U$？左乘 $E^{-1} = L$ 把 A 分解成 LU：

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix}$$

6 哪两个矩阵 E_{21} 与 E_{32} 使得 A 变成上三角形式 $E_{32}E_{21}A = U$？左乘 $E_{32}{}^{-1}$ 与 $E_{21}{}^{-1}$ 把 A 分解成 $LU = E_{21}{}^{-1}\, E_{32}{}^{-1}U$：

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 5 \\ 0 & 4 & 0 \end{bmatrix}$$

7 哪三个矩阵 E_{21} 与 E_{31} 与 E_{32} 使得 A 变成上三角形式 $E_{32}E_{31}E_{21}A = U$？左乘 $E_{32}{}^{-1}$, $E_{31}{}^{-1}$, $E_{21}{}^{-1}$ 把 A 分解成 LU：

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix} \qquad L = E_{21}{}^{-1}\, E_{31}{}^{-1}\, E_{32}{}^{-1} \text{ 。}$$

8 这是展示逆矩阵 $E_{ij}{}^{-1}$ 如何相乘得到 L 的问题。当 A 已经是下三角且对角线都是 1 时，你看到了最佳案例，此时 $U = I$ ：

$$A = L = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}$$

消元矩阵 E_{21} 与 E_{31} 与 E_{32} 包含 $-a$, 然后 $-b$, 然后 $-c$。

(a) 执行乘法 $E_{32}E_{31}E_{21}$ 找出单一矩阵 E 使得 $EA = I$。

(b) 执行乘法 $E_{21}{}^{-1}\, E_{31}{}^{-1}\, E_{32}{}^{-1}$ 找回 L。

乘数 a, b, c 在 E 中是混合的，但是在 L 中是完美的。

9 当 0 出现在主元，$A = LU$ 就不可能出现。(我们需要 U 的非零主元。)直接证明为什么两组方程式都不可能：

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l & 1 \end{bmatrix} \begin{bmatrix} d & e \\ 0 & f \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ l & 1 & \\ m & n & 1 \end{bmatrix} \begin{bmatrix} d & e & g \\ & f & h \\ & & i \end{bmatrix}$$

这些矩阵需要行交换，可以利用交换矩阵 P。

10 哪个数字 c 使得第二主元位置变成零？需要一次的行交换，否则不可能得到 $A = LU$。哪个数字 c 使得第三主元变成零？此时行交换已经不能帮忙且消元法失败：【原文 and $A = LU$ ... 应该是 otherwise $A = LU$ ....】

$$A = \begin{bmatrix} 1 & c & 0 \\ 2 & 4 & 1 \\ 3 & 5 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

11 矩阵 A 的 L 与 D(对角主元矩阵)为何? $A = LU$ 的 U 为何? $A = LDU$ 的新 U 为何?

$$\textbf{已经是三角形} \quad A = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}$$

12 A 与 B 是横跨对角线的对称矩阵(因为 4 = 4)，求出他们的三重分解 LDU，并且说明 U 如何与 L 产生关联。

$$\textbf{对称} \quad A = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 4 & 8 \\ 4 & 12 & 4 \\ 0 & 4 & 0 \end{bmatrix}$$

13 (推荐) 计算对称矩阵 A 的 L 与 U：

$$A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix}$$

求 a, b, c, d 的 4 个条件，使得 $A = LU$ 有 4 个主元。

14 非对称矩阵会有与问题 13 一样的 L：

$$A = \begin{bmatrix} a & r & r & r \\ a & b & s & s \\ a & b & c & t \\ a & b & c & d \end{bmatrix} \quad \textbf{求 } \boldsymbol{A} \textbf{ 的 } \boldsymbol{L} \textbf{ 与 } \boldsymbol{U} \text{ 。}$$

求 a, b, c, d, r, s, t 的 4 个条件，使得 $A = LU$ 有 4 个主元。

问题 15-16 使用 L 与 U(不需要 A)求解 $Ax = b$。

15 求解三角系统 $Lc = b$ 得到 c，然后求解 $Ux = c$ 得到 x。

$$L = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \text{ 与 } U = \begin{bmatrix} 2 & 4 \\ 0 & 1 \end{bmatrix} \text{ 与 } \boldsymbol{b} = \begin{bmatrix} 2 \\ 11 \end{bmatrix}$$

安全起见，如同一般的计算 LU 以及求解 $Ax = b$。当你看到 c 时，圈起来。

16 求解 $Lc = b$ 的 c，然后求解 $Ux = c$ 的 x。A 是什么?

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ 与 } U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ 与 } \boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

17 (a) 当你使用正常的消元步骤至 L 时，你会得到什么矩阵?

$$L=\begin{bmatrix}1 & 0 & 0\\ l_{21} & 1 & 0\\ l_{31} & l_{32} & 1\end{bmatrix}$$

(b) 当你使用相同的步骤到 I 时，你会得到什么矩阵?

(c) 当你使用相同的步骤到 LU 时，你会得到什么矩阵?

18 若 $A=LDU$ 同时 $A=L_1D_1U_1$，所有的因子都是可逆，则 $L=L_1$ 与 $D=D_1$ 与 $U=U_1$，“这个三个因子是唯一的。”

推导方程式 $L_1^{-1}LD=D_1U_1U_1^{-1}$，方程式的两侧是三角形还是对角线?

推论：$L=L_1$ 且 $U=U_1$ (他们都有 1's 在对角线)，则 $D=D_1$。

19 **三对角矩阵**除了在主对角线以及两个相邻对角线有非零单元，其他位置都是 0。分解三对角矩阵得到 $A=LU$ 及 $A=LDL^{\mathrm{T}}$。

$$A=\begin{bmatrix}1 & 1 & 0\\ 1 & 2 & 1\\ 0 & 1 & 2\end{bmatrix} \quad 与 \quad A=\begin{bmatrix}a & a & 0\\ a & a+b & b\\ 0 & b & b+c\end{bmatrix}$$

20 当 T 是**三对角**，它的因子 L 与 U 只有两个非零的对角单元，你会如何利用已知 T 的零，作为高斯消元法的暗码。求 L 与 U。

$$\textbf{三对角}\quad T=\begin{bmatrix}1 & 2 & 0 & 0\\ 2 & 3 & 1 & 0\\ 0 & 1 & 2 & 3\\ 0 & 0 & 3 & 4\end{bmatrix}$$

21 若 A 与 B 在标示 x 的位置有非零单元，哪些标示 0 的单元在因子 L 与 U 中保持是 0?

$$A=\begin{bmatrix}x & x & x & x\\ x & x & x & 0\\ 0 & x & x & x\\ 0 & 0 & x & x\end{bmatrix} \qquad B=\begin{bmatrix}x & x & x & 0\\ x & x & 0 & x\\ x & 0 & x & x\\ 0 & x & x & x\end{bmatrix}$$

22 假设你往上消元(基本没听过)，使用最后一行产生最后一列的零(主元是 1)，然后使用第二行产生第二主元上方的零。求非正常顺序的因子 $A=UL$。

$$\textbf{上三角乘下三角}\quad A=\begin{bmatrix}5 & 3 & 1\\ 3 & 3 & 1\\ 1 & 1 & 1\end{bmatrix}$$

23 **简单却重要**。若 A 有主元 5, 9, 3 且没有行交换，左上方 2×2 子矩阵 A_2 (去除行 3 与列 3)的主元为何？

请尊重版权与译者的劳动成果，侵权必究!

挑战问题

24 哪些可逆矩阵允许 $A = LU$ (消元法没有行交换)？好问题！检视每个 A 的左上角的方形子矩阵 A_k。

所有左上角的 $k \times k$ 子矩阵 A_k 必须可逆 (大小 $k = 1, 2, \ldots, n$)。

解释这个答案：A_k 分解成_____，因为 $LU = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix}\begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix}$

25 对于 6×6 的第二差分**常数-对角**矩阵 K 而言，将主元与乘数引入 $K = LU$，(L 与 U 只有两个非零对角单元，因为 K 有三个。) 求出 L^{-1} 的 i, j 单元的公式，再利用好像 MATLAB 软件的 inv(L)或运用检视法得到良好的模式。

$-1, 2, -1$ 矩阵 $K = \begin{bmatrix} 2 & -1 & & & & \\ -1 & \ddots & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & -1 \\ & & & & -1 & 2 \end{bmatrix}$ = toeplitz([2 −1 0 0 0 0]

26 如果你印出 K^{-1}，它看起来没有那么好看(6×6)。但是如果你印出 $7K^{-1}$，矩阵看起来是多么美妙。依据下列的模式用手算写下 $7K^{-1}$：

1 行 1 与列 1 是(6, 5, 4, 3, 2, 1)。

2 主对角线及对角线之上，行 i 是 i 乘行 1。

3 主对角线及对角线之下，列 j 是 j 乘列 1。

执行 K 乘 $7K^{-1}$ 得到 $7I$，下列是 $n = 3$ 的 $4K^{-1}$：

3×3案例 K的行列式是4

$$(K)\,(4K^{-1}) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & & \\ & 4 & \\ & & 4 \end{bmatrix}$$

【原文写的不清楚，K 是对称，K^{-1} 也是对称，忽略第 3 点。以下是 $7K^{-1}$：】

$$\begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ & 2\cdot5 & 2\cdot4 & 2\cdot3 & 2\cdot2 & 2\cdot1 \\ & & 3\cdot4 & 3\cdot3 & 3\cdot2 & 3\cdot1 \\ & & & 4\cdot3 & 4\cdot2 & 4\cdot1 \\ & & & & 5\cdot2 & 5\cdot1 \\ & & & & & 6\cdot1 \end{bmatrix} = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ & 10 & 8 & 6 & 4 & 2 \\ & & 12 & 9 & 6 & 3 \\ & & & 12 & 8 & 4 \\ & & & & 10 & 5 \\ & & & & & 6 \end{bmatrix} = \begin{bmatrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 10 & 8 & 6 & 4 & 2 \\ 4 & 8 & 12 & 9 & 6 & 3 \\ 3 & 6 & 9 & 12 & 8 & 4 \\ 2 & 4 & 6 & 8 & 10 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

2.7 转置与排列

1 $A\boldsymbol{x}$ 与 AB 与 A^{-1} 的转置分别是 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ 与 $B^{\mathrm{T}}A^{\mathrm{T}}$ 与 $(A^{\mathrm{T}})^{-1}$。

2 点积(内积)为 $\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$，这是 $(1\times n)\,(n\times 1)=(1\times 1)$。

外积是 $\boldsymbol{x}\,\boldsymbol{y}^{\mathrm{T}}=$ 列乘行 $=(n\times 1)\,(1\times n)=n\times n$ 矩阵。

3 A^{T} 的潜在意义是 $A\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}\cdot A^{\mathrm{T}}\boldsymbol{y}$，因为 $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$。

4 对称矩阵有 $\boldsymbol{S}^{\mathrm{T}}=\boldsymbol{S}$，(乘积 $A^{\mathrm{T}}A$ 永远对称)。

5 正交矩阵有 $\boldsymbol{Q}^{\mathrm{T}}=\boldsymbol{Q}^{-1}$，$Q$ 的列是**正交单位**向量[1]。

6 排列矩阵有与 I 相同的行(任意顺序)，会有 $n!$ 的不同排序。

7 $P\boldsymbol{x}$ 会把分量 $x_1, x_2, \ldots, x_n$ 变成新的顺序，且 $P^{\mathrm{T}}=P^{-1}$。

我们还需要一个矩阵，幸运的是它比逆矩阵简单多了，它是 A 的**转置**，写成 A^{T}，A^{T} 的列就是 A 的行。

当 A 是 $m\times n$ 的矩阵，它的转置就是 $n\times m$：

转置 若 $A=\begin{bmatrix}1&2&3\\0&0&4\end{bmatrix}$ 则 $A^{\mathrm{T}}=\begin{bmatrix}1&0\\2&0\\3&4\end{bmatrix}$

你可以把 A 的行写成 A^{T} 的列，或者可以把 A 的列写成 A^{T} 的行。对称矩阵沿着主对角线“翻转”，A^{T} 的(行 i 列 j)的单元来自原始矩阵 A 的(行 j 列 i)的单元：

交换行与列 $(A^{\mathrm{T}})_{ij}=A_{ji}$

下三角矩阵的转置是上三角矩阵，(但是逆矩阵还是下三角。) A^{T} 的转置是 A。

注解：MATLAB 中 A 的转置符号是 A'，键入[1 2 3]得到行向量，列向量是 $\boldsymbol{v}=$[1 2 3]′。矩阵 M 的第二列是 $\boldsymbol{w}=$[4 5 6]′，可以定义 $M=[\boldsymbol{v}\ \ \boldsymbol{w}]$。也可以快速键入行，然后整个矩阵做转置：$M=$[1 2 3; 4 5 6]′。

转置的规则非常直接，我们可以转置 $A+B$ 得到 $(A+B)^{\mathrm{T}}$，或者我们可以分别转置 A 与 B，然后 $A^{\mathrm{T}}+B^{\mathrm{T}}$——答案相同。

[1] **译者注**：本书“正交矩阵”的定义已经是“正交**单位**矩阵”。

请尊重版权与译者的劳动成果，侵权必究!

比较严肃的问题是乘积 AB 以及逆矩阵 A^{-1} 的转置：

和 $(A+B)^{\mathrm{T}}=A^{\mathrm{T}}+B^{\mathrm{T}}$ (1)

乘积 $(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$ (2)

逆矩阵 $(A^{-1})^{\mathrm{T}}=(A^{\mathrm{T}})^{-1}$ (3)

特别注意 $B^{\mathrm{T}}A^{\mathrm{T}}$ 的顺序是如何反序。对于逆矩阵而言，反向顺序很快得到检验：$B^{-1}A^{-1}$ 乘 AB 得到 I。要了解$(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$，可以从$(A\boldsymbol{x})^{\mathrm{T}}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ 开始，此时 B 只是一个向量：

当 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$ 组合 A^{T} 的行，则有 $A\boldsymbol{x}$ 组合 A 的列

它是相同向量的相同组合！在 A 中是列，在 A^{T} 则是行，所以列 $A\boldsymbol{x}$ 的转置就是行 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$，满足我们的公式$(A\boldsymbol{x})^{\mathrm{T}}=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}$。现在我们可以证明公式$(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$，其中 B 有数个列。

若 $B=[\boldsymbol{x}_1,\boldsymbol{x}_2]$有两个列，应用相同的观念在每个列，$AB$ 的列就是 $A\boldsymbol{x}_1$ 与 $A\boldsymbol{x}_2$，他们的转置会正确的出现在 $B^{\mathrm{T}}A^{\mathrm{T}}$ 的行：

$$\text{转置 } AB=\begin{bmatrix} A\boldsymbol{x}_1 & A\boldsymbol{x}_2 & \cdots \\ & & \end{bmatrix} \text{ 得到 } \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{T}}A^{\mathrm{T}} \\ \boldsymbol{x}_2^{\mathrm{T}}A^{\mathrm{T}} \\ \vdots \end{bmatrix} \text{ 就是 } B^{\mathrm{T}}A^{\mathrm{T}} \tag{4}$$

正确的答案 $B^{\mathrm{T}}A^{\mathrm{T}}$ 一次出现一个行，下面是$(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$ 的数字：

$$AB=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}5&0\\4&1\end{bmatrix}=\begin{bmatrix}5&0\\9&1\end{bmatrix},\quad B^{\mathrm{T}}A^{\mathrm{T}}=\begin{bmatrix}5&4\\0&1\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}=\begin{bmatrix}5&9\\0&1\end{bmatrix}$$

反序规则可以扩展到三个或以上的乘积：$(ABC)^{\mathrm{T}}=C^{\mathrm{T}}B^{\mathrm{T}}A^{\mathrm{T}}$。

若 $A=LDU$ 则 $A^{\mathrm{T}}=U^{\mathrm{T}}D^{\mathrm{T}}L^{\mathrm{T}}$。　主元矩阵有 $D=D^{\mathrm{T}}$

现在应用上述的乘积规则来转置 $A^{-1}A=I$。在右侧 $I^{\mathrm{T}}=I$，我们确认这项规则：$(A^{-1})^{\mathrm{T}}$ 是 A^{T} 的逆矩阵，他们的乘积得到 I：

逆矩阵的转置　$A^{-1}A=I$　转置得到　$A^{\mathrm{T}}(A^{-1})^{\mathrm{T}}=I$ (5)

同理 $AA^{-1}=I$，得到$(A^{-1})^{\mathrm{T}}A^{\mathrm{T}}=I$。我们可以**对转置矩阵取逆矩阵**，也可以**转置逆矩阵**。**特别注意**：当 A 为可逆矩阵，A^{T} 也是可逆。

范例 1　$A=\begin{bmatrix}1&0\\6&1\end{bmatrix}$的逆矩阵是 $A^{-1}=\begin{bmatrix}1&0\\-6&1\end{bmatrix}$，且 $A^{\mathrm{T}}=\begin{bmatrix}1&6\\0&1\end{bmatrix}$。

$(A^{-1})^{\mathrm{T}}$ 与 $(A^{\mathrm{T}})^{-1}$ 两者都等于 $\begin{bmatrix}1&-6\\0&1\end{bmatrix}$

内积的意义

我们知道 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的点积(内积)，它是数字 $x_i y_i$ 的累加。现在我们有更好的方式来表示 $\boldsymbol{x}\cdot\boldsymbol{y}$，不再需要不职业的点符号，而是使用矩阵表示：

T 在内	点积或内积是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$	$(1\times n)\ (n\times 1)$
T 在外	秩一的乘积或是外积 $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$	$(n\times 1)\ (1\times n)$

$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 是一个数字，$\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$ 是一个矩阵。量子力学会写成$\langle \boldsymbol{x}|\boldsymbol{y}\rangle$ (内积)与$|\boldsymbol{x}\rangle\langle \boldsymbol{y}|$ (外积)。宇宙可能是被线性代数管理的，下面三个例子对于内积有不同的意义：

来自力学	功(work) = (位移) (力) = $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{f}$
来自电路	热损失 = (压降) (电流) = $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{y}$
来自经济	收入= (数量) (价格) = $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p}$

我们非常接近应用数学的中心了，还有一个重点要强调，内积与 A 的转置有很深的关联。

我们定义 A^{T} 是把矩阵横跨主对角线翻转，这个不是数学，有一个更好的方法来说明转置。A^{T} 是使得下列两个**内积**相等的矩阵，其中 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 是任意向量：

$$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y}) \qquad A\boldsymbol{x} \text{ 与 } \boldsymbol{y} \text{ 的内积} = \boldsymbol{x} \text{ 与 } A^{\mathrm{T}}\boldsymbol{y} \text{ 的内积}$$

从 $A=\begin{bmatrix}-1 & 1 & 0\\ 0 & -1 & 1\end{bmatrix}$，$\boldsymbol{x}=\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}$，$\boldsymbol{y}=\begin{bmatrix}y_1\\ y_2\end{bmatrix}$ 开始，

一侧是 $A\boldsymbol{x}$ 乘 $\boldsymbol{y}$：$(x_2-x_1)y_1+(x_3-x_2)y_2$，这个与 $x_1(-y_1)+x_2(y_1-y_2)+x_3(y_2)$相同，现在 $\boldsymbol{x}$ 乘 $A^{\mathrm{T}}\boldsymbol{y}$。

$$A^{\mathrm{T}}\boldsymbol{y} \text{ 必须是} \begin{bmatrix}-y_1\\ y_1-y_2\\ y_2\end{bmatrix} \text{ 可以如预期得到 } A^{\mathrm{T}}=\begin{bmatrix}-1 & 0\\ 1 & -1\\ 0 & 1\end{bmatrix}$$

对称矩阵

对称矩阵从 A 变成 A^{T} 没有变化，$A^{\mathrm{T}}=A$。它的**(j, i)单元**横跨对角线等于**(i, j)单元**。依据我的观点，这些是最重要的矩阵，我们给定特殊字母 S 表示对称矩阵。

定义 **对称矩阵**有 $S^{\mathrm{T}}=S$， 表示 $s_{ji}=s_{ij}$

请尊重版权与译者的劳动成果，侵权必究!

对称矩阵 $S=\begin{bmatrix}1&2\\2&5\end{bmatrix}=S^{\mathrm{T}}$ 且 $D=\begin{bmatrix}1&0\\0&10\end{bmatrix}=D^{\mathrm{T}}$

对称矩阵的逆矩阵也是对称，S^{-1} 的转置是$(S^{-1})^{\mathrm{T}}=(S^{\mathrm{T}})^{-1}=S^{-1}$，说明 S^{-1} 也是对称(当 S 可逆)：

对称逆矩阵 $S^{-1}=\begin{bmatrix}5&-2\\-2&1\end{bmatrix}$ 且 $D^{-1}=\begin{bmatrix}1&0\\0&0.1\end{bmatrix}$

现在我们用任意矩阵 A^{T} 乘 A 产生一个对称矩阵 S。

对称的乘积 $A^{\mathrm{T}}A$，AA^{T}，LDL^{T}

选取任意矩阵 A，有可能是矩形，乘积 $S=A^{\mathrm{T}}A$ 自动成为方形对称矩阵：

$$A^{\mathrm{T}}A \text{ 的转置是 } A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}}=A^{\mathrm{T}}A \tag{6}$$

这是 $A^{\mathrm{T}}A$ 的对称性的快速证明，我们可以检视 $A^{\mathrm{T}}A$ 的(i,j)单元，它是 A^{T} 的行 i(A 的列 i) 与 A 的列 j 的点积。$(j,\ i)$单元有相同的点积，也是列 j 与列 i，所以 $A^{\mathrm{T}}A$ 是对称。

矩阵 AA^{T} 也是对称(A 与 A^{T} 的形状允许乘法)，但是 AA^{T} 与 $A^{\mathrm{T}}A$ 不同。在我们的经验，大部分的科学问题会从矩形矩阵 A 开始，然后在 $A^{\mathrm{T}}A$ 或是 AA^{T} 或是两者都有时结束，如同最小二乘(least squares)。

范例 2 以两种顺序乘 $A=\begin{bmatrix}-1&1&0\\0&-1&1\end{bmatrix}$ 以及 $A^{\mathrm{T}}=\begin{bmatrix}-1&0\\1&-1\\0&1\end{bmatrix}$。

$$AA^{\mathrm{T}}=\begin{bmatrix}2&-1\\-1&2\end{bmatrix} \text{ 以及 } A^{\mathrm{T}}A=\begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&1\end{bmatrix} \text{ 两者都是对称矩阵。}$$

乘积 $A^{\mathrm{T}}A$ 是 $n\times n$，反序时 AA^{T} 是 $m\times m$，两者都是对称，其中对角线都是正数(为什么?)。纵使 $m=n$，也很可能会有 $A^{\mathrm{T}}A\neq AA^{\mathrm{T}}$，也许等号会成立，却是不正常的。

对称矩阵的消元 $S^{\mathrm{T}}=S$ 使得消元法更加快速，因为我们只需要处理半边的矩阵就可以(加上对角线)。上三角矩阵不太可能是对称的，对称性会出现在三重乘积 $S=LDU$。还记得主元对角矩阵 D 如何利用除法，使得 L 与 U 的对角线都是 1's：

$$\begin{bmatrix}1&2\\2&7\end{bmatrix}=\begin{bmatrix}1&0\\2&1\end{bmatrix}\begin{bmatrix}1&2\\0&3\end{bmatrix} \qquad LU \text{ 失去 } S \text{ 的对称性}$$

$$\begin{bmatrix}1&2\\2&7\end{bmatrix}=\begin{bmatrix}1&0\\2&1\end{bmatrix}\begin{bmatrix}1&0\\0&3\end{bmatrix}\begin{bmatrix}1&2\\0&1\end{bmatrix} \qquad LDL^{\mathrm{T}} \text{ 捕捉对称性}$$

现在 U 是 L 的转置

当 S 是对称，通常的形式 $A = LDU$ 变成 $\boldsymbol{S = LDL^{\mathrm{T}}}$。最后的 U(对角线都是 1)是 L 的转置(对角线也都是 1)。包含主元的对角矩阵 D，本身也是对称。

若 $S = S^{\mathrm{T}}$ 分解成 LDU，其中没有出现行交换，则 U 就是 L^{T}。

对称矩阵的对称分解是 $S = LDL^{\mathrm{T}}$

注意 LDL^{T} 的转置自动变成$(L^{\mathrm{T}})^{\mathrm{T}}D^{\mathrm{T}}L^{\mathrm{T}} = LDL^{\mathrm{T}}$。消元法的工作量减去一半，从 $n^3/3$ 乘法变成 $n^3/6$，储存位置也基本少了一半。我们只需要保存 L 与 D，不需要储存 U，因为 U 就是 L^{T}。

排列矩阵

转置对于排列矩阵来说扮演特殊的角色，矩阵 P 有单一的“1”在每一行与每一列，P^{T} 也是排列矩阵--可能与 P 相同也可能不同。任何乘积 $P_1 P_2$ 再次是个排列矩阵。

我们现在重新排列 I 的行的顺序，得到来自单位矩阵的每个 P。
最简单的排列矩阵是 $P = I$(无交换)，下一个最简单的排列矩阵应该是行交换 P_{ij}。这些矩阵可以通过交换 I 的两个行 i 与 j 而创建，其他排列矩阵重组更多的行的顺序。对于 I 执行所有可能的行交换，我们得到全部可能的排列矩阵。

定义　　排列矩阵 P 具有单位矩阵 I 的任意顺序的行

范例 3　下列是 6 个 3×3 排列矩阵，隐藏 0 后得到

$$I = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \quad P_{21} = \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} \quad \left| \begin{array}{l} P_{32}P_{21} = \begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix} \\ P_{21}P_{32} = \begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix} \end{array} \right.$$

$$P_{31} = \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \quad P_{32} = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}$$

阶数 n 会有 $n!$ 的排列矩阵，符号 $n!$ 表示 n 阶层，就是数字(1)(2) ... (n)，于是 3! = (1)(2)(3) = 6。若阶数 $n = 4$ 会有 24 种排列矩阵，阶数 5 会有 120 种排列矩阵。

阶数 2 时只有两种排列矩阵，即为$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$与$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$。

重要：P^{-1} 也是排列矩阵。前述 6 个 3×3 的 P 矩阵中，左边的 4 个矩阵就是自己的逆矩阵，右边的两个互为逆矩阵。所有单一的行交换都是自己的逆矩阵，如果再次重复相同的交换会回复到 I。但是对于 $P_{32}P_{21}$ 来说，逆矩阵总是变成相反的顺序，逆矩阵是 $P_{21}P_{32}$。

请尊重版权与译者的劳动成果，侵权必究!

更为重要：P^{-1} 永远等于 P^{T}。在右边的两个矩阵互为逆矩阵，也互为转置矩阵。当我们计算乘积 PP^{T} 时，P 的第一行里面的 1，会碰上 P^{T} 的第一列里面的 1(由于 P 的第一行是 P^{T} 的第一列)，其他列都不会有 1 存在，所以 $PP^{T}=I$。

$P^{T}=P^{-1}$ 的另一个证明，将 P 视为行交换矩阵的乘积，每个行交换都是自己的转置，也是自己的逆矩阵，P^{T} 与 P^{-1} 都是来自反序的行交换矩阵的乘积，所以 P^{T} 与 P^{-1} 相同。

排列(消元前先交换)得到 $PA=LU$

利用行交换的 $PA=LU$ 分解

我们当然希望你还记得 $A=LU$，从 $A=(E_{21}{}^{-1}\ldots E_{ij}{}^{-1}\ldots)U$ 开始，执行一个 E_{ij} 代表一个消元步骤，然后取逆矩阵 $E_{ij}{}^{-1}$。这些逆矩阵压缩后变成一个矩阵 L，下三角 L 的对角线都是 1，结果就是 $A=LU$。

这是一个伟大的分解，但是不一定能够行得通。有时候需要交换行得到主元，然后 $A=(E^{-1}\ldots P^{-1}\ldots E^{-1}\ldots P^{-1}\ldots)U$。每个行交换由 P_{ij} 执行，再取 P_{ij} 的逆矩阵。我们现在把这些行交换矩阵压缩变成一个单一的排列矩阵 P，可以得到每一个可逆矩阵 A 的分解——这是我们想要的。

主要的问题是去哪儿收集 P_{ij}。有两个良好的可能性——在消元之前做好所有的交换，或是在 E_{ij} 之后再做交换。第一个方法得到 $PA=LU$，第二个方法会在中间有一个交换矩阵 P。

1. 行交换可以事先执行。他们的乘积 P 把 A 的行放在正确的顺序，对于 PA 来说就不需要执行交换，则 $PA=LU$。
2. 如果在全部消元步骤执行完成后再交换行，主元行的排列会是非常奇怪的顺序。P_1 将主元行放在 U_1，而且呈现正确的三角顺序，则 $A=L_1P_1U_1$。

$PA=LU$ 在所有的计算中经常使用，我们会聚焦在这个形式。

分解 $A=L_1P_1U_1$ 可能可以更优雅，如果两者并论，之间的差别不是广为人知。你可能不会花长时间在这两个方法，请不要，最重要的情形是当 $A=LU$，而且不存在行交换时，$P=I$。

下列的矩阵 A 从 $a_{11}=0$ 开始，交换行 1 与行 2 使得第一主元进入惯常位置，然后持续对 PA 执行消元法：

$$\underset{A}{\begin{bmatrix}0&1&1\\1&2&1\\2&7&9\end{bmatrix}}\rightarrow\underset{PA}{\begin{bmatrix}1&2&1\\0&1&1\\2&7&9\end{bmatrix}}\rightarrow\underset{l_{31}=2}{\begin{bmatrix}1&2&1\\0&1&1\\0&3&7\end{bmatrix}}\rightarrow\underset{l_{32}=3}{\begin{bmatrix}1&2&1\\0&1&1\\0&0&4\end{bmatrix}}$$

矩阵 PA 的行有良好的顺序，如往常一般分解成 LU：

$$P=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\qquad PA=\begin{bmatrix}1&0&0\\0&1&0\\2&3&1\end{bmatrix}\begin{bmatrix}1&2&1\\0&1&1\\0&0&4\end{bmatrix}=LU \tag{7}$$

我们从 A 开始到 U 结束，唯一的要求是 A 的可逆性。

若 A 可逆，排列矩阵 P 会把 A 的行放在正确的顺序来分解 $PA = LU$。在行交换之后，必定存在一整组的主元使得 A 是可逆。

在 MATLAB 中，A([r k], :) = A([k r], :) 会交换行 k 以及在其下方的行 r(其中第 k 主元已经存在。) 则 lu 代码会更新 L 与 P 以及 P 的符号：

这是部分的[L, U, P] = lu(A)

```
A([r k], :) = A([k r], :);
L([r k], 1:k-1) = L([k r], 1:k-1);
P([r k], :) = P([k r], :);
sign = -sign
```

P 的“sign”可以分辨行交换的数目是否为偶数(sign = +1)，奇数的行交换会得到 sign = –1。开始的时候，P 是 I 且 sign = +1，出现一个行交换，sing 就会变换符号。sign 的最终值就是 P 的行列式，行列式与行交换的顺序无关。

对 PA 来说，我们回到熟悉的 LU。实际上诸如 lu(A)的代码通常不会用在第一个可用的主元上，从数学来说我们可以接受小的主元--除了 0 以外，所有良好的代码会往下寻找最大主元的列。

段落 11.1 说明为什么“部分主元(partial pivoting)”可以降低舍去误差，则 P 可以包含行交换，虽然这个不是代数上的必需，仍然有 $PA = LU$。

我们的建议是了解排列，却是让电脑去执行工作，$A = LU$ 的计算在没有 P 时可以手算。教学代码 splu(A) 会分解 $PA = LU$，splv(A, b)可以求解 $A\boldsymbol{x} = \boldsymbol{b}$，其中 A 是可逆。如果在列 k 找不到主元，网址上的程序 splu 会终止，A 就不是可逆。

主要观念的复习

1. 转置使得 A 的行变成 A^{T} 的列，则$(A^{\mathrm{T}})_{ij} = A_{ji}$ 。
2. AB 的转置是 $B^{\mathrm{T}}A^{\mathrm{T}}$，$A^{-1}$ 的转置是 A^{T} 的逆矩阵。
3. 点积是 $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$，$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}$ 等于点积 $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$。
4. 当 S 对称($S^{\mathrm{T}} = S$)，它的 LDU 分解是对称：$S = LDL^{\mathrm{T}}$。
5. 排列矩阵 P 有一个 1 在每一行以及每一列，$P^{\mathrm{T}} = P^{-1}$。
6. 阶数为 n 时会有 $n!$ 个排列矩阵，半偶，半奇。【参考习题 10】
7. 若 A 可逆，排列矩阵 P 会重新排列 A 的行，使得 $PA = LU$。

请尊重版权与译者的劳动成果，侵权必究!

已解范例

2.7A 应用排列矩阵 P 至 S 的行，会摧毁 S 的对称性。

$$P=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}\qquad S=\begin{bmatrix}1&4&5\\4&2&6\\5&6&3\end{bmatrix}\qquad PS=\begin{bmatrix}4&2&6\\5&6&3\\1&4&5\end{bmatrix}$$

什么样的排列 Q 应用至 PS 的列，会使得 PSQ 恢复对称性？数字 1, 2, 3 必须回到主对角线(不一定要原来顺序)。证明 $Q=P^{\mathrm{T}}$ 使得 PSP^{T} 恢复对称性。

解 要恢复对称性，2 要回到对角线，所以 PS 的列 2 要移动列 1。PS 的列 3(包含 3 的数字)必须移到列 2，然后“1”移到 3, 3 的位置。可以操作列交换的排列矩阵 Q 为

$$PS=\begin{bmatrix}4&2&6\\5&6&3\\1&4&5\end{bmatrix}\quad Q=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\quad PSQ=\begin{bmatrix}2&6&4\\6&3&5\\4&5&1\end{bmatrix}\text{ 是对称。}$$

矩阵 Q 就是 P^{T}，这个选择永远可以恢复对称性，因为 PSP^{T} 可以确保对称性。(它的转置依然是 PSP^{T}。) 矩阵 Q 同时也是 P^{-1}，因为每一个排列矩阵的逆矩阵是自己的转置。

若 D 是对角矩阵，得到的 PDP^{T} 也是对角矩阵。当 P 把行 1 往下移到行 3，在右侧的 P^{T} 会把列 1 移到列 3，(1, 1)单元就会往下移动到(3, 1)然后到(3, 3)。

2.7B 给定上述的矩阵 S，求 $S=LDL^{\mathrm{T}}$ 的对称分解。

解 要将 S 分解成为 LDL^{T}，我们也要使用消元法得到 U：

$$S=\begin{bmatrix}1&4&5\\4&2&6\\5&6&3\end{bmatrix}\to\begin{bmatrix}1&4&5\\0&-14&-14\\0&-14&-22\end{bmatrix}\to\begin{bmatrix}1&4&5\\0&-14&-14\\0&0&-8\end{bmatrix}=U$$

乘数是 $l_{21}=4$，$l_{31}=5$，$l_{32}=1$。主元 1, –14, –8 进到 D 里面。当我们把 U 的行除以主元，L^{T} 应该变成：

$$\text{当}S=S^{\mathrm{T}}\text{ 对称分解}\qquad S=LDL^{\mathrm{T}}=\begin{bmatrix}1&0&0\\4&1&0\\5&1&1\end{bmatrix}\begin{bmatrix}1&&\\&-14&\\&&-8\end{bmatrix}\begin{bmatrix}1&4&5\\0&1&1\\0&0&1\end{bmatrix}$$

S 有 3 个主元，S 可逆，它的逆矩阵是$(L^{\mathrm{T}})^{-1}D^{-1}L^{-1}$，且 S^{-1} 也是对称。数字 14 与 8 会出现在 S^{-1} 的分母，S 的行列式是主元的乘积(1)(–14)(–8) = 112。

2.7C A 是矩形矩阵，鞍点(saddle point)矩阵 S 是对称而且重要：

$$\textbf{来自最小二乘的方块矩阵}\quad S=\begin{bmatrix} I & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}=S^{\mathrm{T}}\quad \text{它的大小是 } m+n$$

应用方块消元法得到方块分解 $S=LDL^{\mathrm{T}}$，然后测试可逆性质：

$$\boldsymbol{S}\textbf{ 可逆}\quad\Leftrightarrow\quad \boldsymbol{A^{\mathrm{T}}A}\textbf{ 可逆}\quad\Leftrightarrow\quad \textbf{当 }\boldsymbol{x\neq 0}\textbf{ 则 }\boldsymbol{Ax\neq 0}$$

解 第一个方块主元是 I，从行 2 减去 A^{T} 乘行 1：

$$\textbf{方块消元}\quad S=\begin{bmatrix} I & A \\ A^{\mathrm{T}} & 0 \end{bmatrix}\text{ 变成 }\begin{bmatrix} I & A \\ 0 & -A^{\mathrm{T}}A \end{bmatrix}\quad\text{这是 } U$$

方块主元矩阵 D 包含 I 与 $-A^{\mathrm{T}}A$，则 L 与 L^{T} 包含 A^{T} 与 A：

$$\textbf{方块分解}\quad S=LDL^{\mathrm{T}}=\begin{bmatrix} I & 0 \\ A^{\mathrm{T}} & I \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & -A^{\mathrm{T}}A \end{bmatrix}\begin{bmatrix} I & A \\ 0 & I \end{bmatrix}$$

L 必然是可逆且对角线都是 1。中间矩阵的逆矩阵与 $(A^{\mathrm{T}}A)^{-1}$ 有关，段落 4.2 会回答有关矩阵 $A^{\mathrm{T}}A$ 的一个关键问题。

何时 $A^{\mathrm{T}}A$ 可逆？答案：A 必须有无关的列。
只有 $\boldsymbol{x}=0$ 时才有 $A\boldsymbol{x}=0$，否则 $A\boldsymbol{x}=0$ 会得到 $A^{\mathrm{T}}A\boldsymbol{x}=0$。

问题集 2.7

问题 1-7 有关转置矩阵的规则。

1 求 A^{T} 与 A^{-1} 与 $(A^{-1})^{\mathrm{T}}$ 与 $(A^{\mathrm{T}})^{-1}$：

$$A=\begin{bmatrix} 1 & 0 \\ 9 & 3 \end{bmatrix}\quad\text{以及}\quad A=\begin{bmatrix} 1 & c \\ c & 0 \end{bmatrix}$$

2 验证 $(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$，但是会与 $A^{\mathrm{T}}B^{\mathrm{T}}$ 不同。

$$A=\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\quad B=\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}\quad AB=\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}$$

证明 AA^{T} 与 $A^{\mathrm{T}}A$ 不同，但是两个矩阵是_____。

3 (a) 矩阵 $((AB)^{-1})^{\mathrm{T}}$ 来自 $(A^{-1})^{\mathrm{T}}$ 与 $(B^{-1})^{\mathrm{T}}$，顺序为何？

(b) 若 U 是上三角，则 $(U^{-1})^{\mathrm{T}}$ 是___三角。

4 证明 $A^2=0$ 是可能的，但是 $A^{\mathrm{T}}A=0$ 是不可能的(除非 A =零矩阵)。

请尊重版权与译者的劳动成果，侵权必究！

5 (a) 行向量 $\boldsymbol{x}^{\mathrm{T}}$ 乘 A 乘列向量 $\boldsymbol{y}$ 得到什么数字?

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{y}=\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{bmatrix}\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}=_____$$

(b) 这就是行 $\boldsymbol{x}^{\mathrm{T}}A=$_____乘列 $\boldsymbol{y}=(0,1,0)$。

(c) 这就是行 $\boldsymbol{x}^{\mathrm{T}}=[0\quad 1]$乘列 $A\boldsymbol{y}=$ _____。

6 方块矩阵 $M=\begin{bmatrix}A & B\\ C & D\end{bmatrix}$的转置是 $M^{\mathrm{T}}=$ _____。测试一个例子。A, B, C, D 在什么条件下，这个方块矩阵是对称的?

7 是非题：

(a) 方块矩阵$\begin{bmatrix}0 & A\\ A & 0\end{bmatrix}$自动成为对称矩阵。

(b) 若 A 与 B 都是对称，则他们的乘积 AB 也是对称。

(c) 若 A 不是对称，则 A^{-1} 不是对称。

(d) 当 A, B, C 都是对称，则 ABC 的转置是 CBA。

问题 8-15 有关排列矩阵。

8 矩阵的阶数为 n，为什么会有 $n!$ 个排列矩阵?

9 若 P_1 与 P_2 是排列矩阵，乘积 P_1P_2 也是排列矩阵，这个仍然有 I 的行依据某种顺序排列。举例说明 $P_1P_2\neq P_2P_1$ 以及 $P_3P_4=P_4P_3$。

10 (1, 2, 3, 4)存在 12 种“偶(even)”排序，其中交换次数是偶数。其中两种是(1, 2, 3, 4)没有交换以及(4, 3, 2, 1)交换两次，请列出其他 10 种，不需要写出每个 4×4 矩阵，只要数字排序即可。

11 何种排列矩阵使得 PA 是上三角? 何种排列矩阵使得 P_1AP_2 是下三角? 在 A 的右侧乘 P_2 会交换 A 的____。

$$A=\begin{bmatrix}0 & 0 & 6\\ 1 & 2 & 3\\ 0 & 4 & 5\end{bmatrix}$$

12 解释为什么 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的点积等于 $P\boldsymbol{x}$ 与 $P\boldsymbol{y}$ 的点积? 则$(P\boldsymbol{x})^{\mathrm{T}}(P\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 告诉我们对于任何排列矩阵都有 $P^{\mathrm{T}}P=I$。当 $\boldsymbol{x}=(1,2,3)$与 $\boldsymbol{y}=(1,4,2)$时，选取 P 证明$(P\boldsymbol{x})\cdot\boldsymbol{y}$ 不一定等于 $\boldsymbol{x}\cdot(P\boldsymbol{y})$。

13 (a) 求一个 3×3 的排列矩阵，其中 $P^3=I$ (但是 $P\neq I$)。

(b) 求一个 4×4 的排列矩阵$\hat{P}$，其中$\hat{P}^4\neq I$。

14 若 P 在反对角线从(1, n)到(n, 1)有 1's 存在，描述 PAP。注意 $P=P^{\mathrm{T}}$。

【对角线是左上到右下，反对角线(antidiagonal)是右上到左下】

15 所有的行交换矩阵都是对称：$P^{\mathrm{T}} = P$，则 $P^{\mathrm{T}}P = I$ 变成 $P^2 = I$。其他排列矩阵可能是也可能不是对称。

(a) 若 P 把行 1 送到行 4，则 P^{T} 把行____送到____。当 $P^{\mathrm{T}} = P$ 时，行交换会成对(pair)出现，而且没有重叠。

(b) 求一个 4×4 矩阵，其中 $P^{\mathrm{T}} = P$，会移动全部 4 个行。

问题 16-21 有关对称矩阵与分解。

16 若 $A^{\mathrm{T}} = A$ 与 $B^{\mathrm{T}} = B$，下列哪些矩阵一定是对称：

(a) $A^2 - B^2$　　(b) $(A + B)(A - B)$　　(c) ABA　　(d) $ABAB$

17 求一个 2×2 对称矩阵 $S = S^{\mathrm{T}}$，其中 S 分别具有下列 3 个性质。

(a) S 不可逆。

(b) S 可逆但是不能分解成 LU (需要行交换)。

(c) S 可以分解成 LDL^{T}，但是不能分解成 LL^{T} (因为负的 D)。

18 (a) 若 $S = S^{\mathrm{T}}$ 且为 5×5，S 有多少单元可以无关的选取?

(b) L 与 D(仍然是 5×5)如何在 LDL^{T} 中得到相同的选取数值？

(c) 若 A 是反对称矩阵，A 有多少单元可以选取? ($A^{\mathrm{T}} = -A$)。

19 假设 A 是矩形($m \times n$)，S 是对称($m \times m$)。

(a) 转置 $A^{\mathrm{T}}SA$ 展示对称性，矩阵的形状为何?

(b) 说明为什么 $A^{\mathrm{T}}A$ 对角线没有负数。

20 分解下列矩阵得到 $S = LDL^{\mathrm{T}}$，主元矩阵 D 是对角矩阵：

$$S = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \quad 与 \quad S = \begin{bmatrix} 1 & b \\ b & c \end{bmatrix} \quad 与 \quad S = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

21 在消元法清除列 1 在第一主元的下方后，求出现在右下角的 2×2 对称矩阵：

$$从\ S\ 出发 \quad S = \begin{bmatrix} 2 & 4 & 8 \\ 4 & 3 & 9 \\ 8 & 9 & 0 \end{bmatrix} \quad 与 \quad S = \begin{bmatrix} 1 & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

问题 22-24 有关分解 $PA = LU$ 与 $A = L_1P_1U_1$。

22 求出 $PA = LU$ 分解(并验证)：

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 3 & 4 \end{bmatrix} \quad 与 \quad A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

23 找出一个 4×4 排列矩阵(称为 A)，需要 3 次交换完成消元。对矩阵来说它的 P，L，U 为何？

24 分解下列矩阵得到 $PA = LU$，也分解成 $A = L_1P_1U_1$ (行 3 先不交换，直到完成从行 2 减去 3 乘行 1)：

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 3 & 8 \\ 2 & 1 & 1 \end{bmatrix}$$

25 证明单位矩阵不可能是 3 个行交换(或 5 个)的乘积，它可以是 2 个(或 4 个)行交换的乘积。

26 (a) 选择 E_{21} 移除第一主元下方的 3，然后执行乘法 $E_{21}SE_{21}^{\mathrm{T}}$ 消去两个 3：

$$S = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 11 & 4 \\ 0 & 4 & 9 \end{bmatrix} \quad 走向目标 \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(b) 选择 E_{32} 移除第二主元下方的 4，然后利用 $E_{32}E_{21}SE_{21}^{\mathrm{T}}E_{32}^{\mathrm{T}}$ 简化 S 成为 D，逆反 E's 求出 $S = LDL^{\mathrm{T}}$ 的 L。

27 若 4×4 矩阵的每一行都包含某种顺序的数字 0, 1, 2, 3，这个矩阵可能对称吗？

28 证明：不存在行的重排序与列的重排序可以转置一个典型矩阵。(注意对角线的单元)。

下面 3 个问题有关等式$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$的应用。

29 在波士顿，芝加哥，西雅图之间铺设电缆，这些城市所在的电压是 x_B, x_C, x_S。如果城市之间都是单位电阻，则城市之间的电流在 $\boldsymbol{y}$ 中：

$$\boldsymbol{y} = A\boldsymbol{x} \ 是 \ \begin{bmatrix} y_{BC} \\ y_{CS} \\ y_{BS} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_B \\ x_C \\ x_S \end{bmatrix}$$

(a) 求出离开三个城市的总电流 $A^{\mathrm{T}}\boldsymbol{y}$。

(b) 验证$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}$ 与 $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$具有一致性——两者都有 6 项。

30 制造 x_1 辆卡车与 x_2 架飞机需要 $x_1 + 50x_2$ 吨的钢材，$40x_1 + 1000x_2$ 磅的橡胶以及 $2x_1 + 50x_2$ 个月的人力。如果单位成本 y_1, y_2, y_3 是每吨\$700，每磅\$3，每个月\$3000，请问一辆卡车与一架飞机的价值各是多少？这些是 $A^{\mathrm{T}}\boldsymbol{y}$ 的分量。

31 $A\boldsymbol{x}$ 得到钢材，橡胶，人力的量，然后产生问题 31 的 $\boldsymbol{x}$，求出 A。当 $\boldsymbol{x} \cdot A^{\mathrm{T}}\boldsymbol{y}$ 是_____的价值时，$A\boldsymbol{x} \cdot \boldsymbol{y}$ 是输入的_____。

32 矩阵 P 乘(x, y, z)得到(z, x, y)也是一种旋转矩阵，求 P 与 P^3。旋转轴 $\boldsymbol{a} = (1, 1, 1)$不会移动，它等于 Pa。请问从 $\boldsymbol{v} = (2, 3, -5)$到 $P\boldsymbol{v} = (-5, 2, 3)$的旋转角为何？

33 将 $A = \begin{bmatrix} 1 & 2 \\ 4 & 9 \end{bmatrix}$ 写成 ES 的乘积，其中 E 是基础行运算矩阵，S 是对称矩阵。

34 此处是 A 的新分解得到 LS：三角矩阵(有 1's)乘对称矩阵：

从 $A = LDU$ 开始，则 A 等于 $L(U^{\mathrm{T}})^{-1}$ 乘 $(S = U^{\mathrm{T}}DU)$。

为什么 $L(U^{\mathrm{T}})^{-1}$ 是三角形？它的对角都是 1。为什么 $U^{\mathrm{T}}DU$ 是对称?

35 如果一个矩阵群组包含 A 与 B，则 AB 与 A^{-1} 也在这个矩阵群组中。“乘积与逆矩阵保留在群组”。下列哪些集合是矩阵群组?

对角线都是 1 的下三角矩阵 L，对称矩阵 S，正矩阵 M，对角可逆矩阵 D，排列矩阵 P，$Q^{\mathrm{T}} = Q^{-1}$ 的矩阵。再找出两个矩阵群组。

挑战问题

36 方形**西北矩阵** B 在反对角线(连接$(1, n)$与$(n, 1)$)下方的东南角单元都是 0，B^{T} 与 B^2 也是西北矩阵? B^{-1} 是西北或东南矩阵? $BC =$ **西北**乘**东南**的形状?

37 如果你对排列矩阵取幂次方，为什么某些 P^k 最终会等于 I？找出一个 5×5 排列矩阵 P，使得最小的幂次方等于 I 的数字是 6。

38 (a) 写下任意的 3×3 矩阵 M，分割 M 变成 $S + A$，其中 $S = S^{\mathrm{T}}$ 是对称，

$A = -A^{\mathrm{T}}$ 是反对称。

(b) 写出 S 与 A 的公式，这个公式与 M 及 M^{T} 有关。我们希望 $M = S + A$。

39 假设 $Q^{\mathrm{T}} = Q^{-1}$ (转置等于逆，所以 $Q^{\mathrm{T}}Q = I$)。

(a) 证明列 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ 都是单位矩阵：$\| \boldsymbol{q}_i \|^2 = 1$。

(b) 证明 Q 的任意两个列互相垂直：$\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_2 = 0$。

(c) 求出一个 2×2 的例子，其中第一个单元是 $q_{11} = \cos\theta$ 。

请尊重版权与译者的劳动成果，侵权必究!

导数的转置

你会允许我讲一些微积分吗？这个非常重要或者是我不想离开线性代数。(这真的是函数 $x(t)$的线性代数。) 矩阵变成微分，因此 $A = d/dt$。要求出这个不寻常 A 的转置，我们需要定义两个函数 $x(t)$与 $y(t)$的内积：

函数的内积 $$x^{\mathrm{T}}y = (x,\ y) = \int_{-\infty}^{\infty} x(t)y(t)dt$$

从内积来看，我们知道对于 A^{T} 的要求。当我们处理微分时，用“伴随(adjoint)”这个词语会比“转置”来的正确。

矩阵的转置有$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$，$A = d/dt$ 的伴随有

$$(Ax,\ y) = \int_{-\infty}^{\infty} \frac{dx}{dt}\, y(t)dt = \int_{-\infty}^{\infty} x(t)\left(\frac{-dy}{dt}\right)dt = \left(x,\ A^{\mathrm{T}}y\right)$$

我希望你了解部分积分(integration by parts)。微分从第一个函数 $x(t)$移到第二个函数 $y(t)$，在移动时出现了一个负号，这个告诉我们**微分的转置是微分加负号。**

微分是反对称，$\boldsymbol{A = d/dt}$ 且 $\boldsymbol{A^{\mathrm{T}} = -\,d/dt}$。对称矩阵有 $S^{\mathrm{T}} = S$，反对称矩阵有 $A^{\mathrm{T}} = -A$。第八章会介绍线性代数包含微分与积分，因为它们两者都是线性的。

微分的反对称也可以应用到中心差分矩阵：

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{转置成} \quad A^{\mathrm{T}} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = -A$$

前向差分矩阵的转置会变成反向差分矩阵，相当于乘–1。在微分方程式中，第二导数(加速度)是对称的，第一导数(正比于速度的阻尼)是反对称的。

第三章
向量空间与子空间

3.1 向量空间

1 标准 n 维空间$\mathbf{R}^n$包含全部具有 n 个分量的实数列向量。

2 若 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 在**向量空间** $\boldsymbol{S}$ 中，则所有的组合 $c\boldsymbol{v}+d\boldsymbol{w}$ 也必须在 $\boldsymbol{S}$ 中。

3 $\boldsymbol{S}$ 中的“向量”可以是矩阵或是 x 的函数。“一个点”空间 $\boldsymbol{Z}$ 由 $\boldsymbol{x}=\boldsymbol{0}$ 构成。

4 $\mathbf{R}^n$的**子空间**是$\mathbf{R}^n$里面的一个向量空间，例如：直线 $y=3x$ 在$\mathbf{R}^2$里面。

5 A 的**列空间**包含 $\boldsymbol{A}$ **的列**的所有组合：$\mathbf{R}^m$的子空间。

6 列空间包含所有的向量 $A\boldsymbol{x}$，所以当 $\boldsymbol{b}$ **在** $\boldsymbol{C}(A)$**中**，$A\boldsymbol{x}=\boldsymbol{b}$ 有解。

对新手而言，矩阵计算包含很多数字，对你来说，他们与向量有关。$A\boldsymbol{x}$ 与 AB 的列都是 n 个向量的线性组合--这些向量就是 A 的列。本章从数字与向量转移到第三层次的了解(最高层次)，我们检视向量的空间，而不仅仅看着个别的向量。如果不了解空间，特别是他们的子空间，你就无法了解 $A\boldsymbol{x}=\boldsymbol{b}$ 的全部面貌。

由于本章走得有点深入，可能会有一点难度，这个很正常。我们检视的是计算的内在，是在发现数学，作者的工作就是让他们清晰。本章在“**线性代数的基础定理**”结束。

我们从最重要的向量空间开始，他们标示成$\mathbf{R}^1$，$\mathbf{R}^2$，$\mathbf{R}^3$，$\mathbf{R}^4$，...，每个空间$\mathbf{R}^n$包含一整个向量的集合。$\mathbf{R}^5$包含全部具有 5 个分量的列向量，称为 5 维空间。

定义　空间$\mathbf{R}^n$包含全部具有 $\boldsymbol{n}$ 个分量的列向量 $\boldsymbol{v}$。

请尊重版权与译者的劳动成果，侵权必究!

向量 $\boldsymbol{v}$ 的分量是实数，这是使用字母 **R** 的理由。具有 n 个复数分量的向量会落在空间$\mathbf{C}^n$中。

向量空间$\mathbf{R}^2$就是平常的 xy 平面，每个$\mathbf{R}^2$里面的向量有两个分量。词语“空间”要我们思考全部的向量——整个平面，每个向量给出平面的一个点的 x 与 y 坐标：$\boldsymbol{v} = (x, y)$。

相似的情形，$\mathbf{R}^3$的向量对应三维空间的点(x, y, z)，一维空间$\mathbf{R}^1$是一条直线(如同 x 轴)。跟以前一样，我们把向量写成在括号中的列，或是在水平行用逗号与括号表示：

$$\begin{bmatrix}4\\ \pi\end{bmatrix} \text{在} \mathbf{R}^2，(1, 1, 0, 1, 1) \text{在} \mathbf{R}^5，\begin{bmatrix}1+i\\ 1-i\end{bmatrix} \text{在} \mathbf{C}^2$$

线性代数最伟大的地方就是它很容易处理 5 维空间，我们不必画出向量，我们只需要 5 个数字(或 n 个数字)。

$\boldsymbol{v}$ 乘 7 是把每个分量乘 7，此处的 7 是纯量。要在$\mathbf{R}^5$中计算向量加法，就是将对应分量相加。向量空间一样有两个重要的向量运算，他们产生了线性组合。

我们可以在$\mathbf{R}^n$中任意相加向量，也可以对任意向量 $\boldsymbol{v}$ 乘任意纯量 c。

“在向量空间中”表示“结果”还停留在空间里面。若 $\boldsymbol{v}$ 是$\mathbf{R}^4$的向量，分量是 1, 0, 0, 1，则 $2\boldsymbol{v}$ 是$\mathbf{R}^4$的向量，分量是 2, 0, 0, 2。(本范例中 2 是纯量。) 一整个系列性质可以在$\mathbf{R}^n$中验证，交换律是 $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$；分配律是 $c(\boldsymbol{v} + \boldsymbol{w}) = c\boldsymbol{v} + c\boldsymbol{w}$；存在唯一的“零向量”满足 $\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$；这些是问题集开头叙述的八个条件中的三个。

每个向量空间都要满足这八个条件，他们**是向量而不是列向量**，在$\mathbf{R}^n$之外还存在其他向量空间，所有的向量空间都必须遵守这八个规则。

一个实数向量空间是一个向量的集合，遵循实数向量加法与实数纯量乘法的规则，加法与乘法得到的向量还必须留在原来的空间中，同样要满足这八个条件(通常是没有问题)。下列是三个$\mathbf{R}^n$之外的向量空间：

M	所有**实数 2×2 矩阵**组成的向量空间
F	所有**实数函数** $f(x)$组成的向量空间
Z	只有一个**零向量**的向量空间

在 **M** 中的“向量”实际上是矩阵，在 **F** 中的向量是函数，在 **Z** 中唯一的加法是 $\mathbf{0} + \mathbf{0} = \mathbf{0}$。每种案例都可以加：矩阵加矩阵，函数加函数，零向量加零向量。我们可以用 4 乘矩阵，或用 4 乘函数，或用 4 乘零向量，得到的结果仍然在 **M** 或 **F** 或 **Z** 中，而且这八个条件很容易检验。

函数空间 **F** 是无限维度，比较小的空间是 **P** 或 $\mathbf{P_n}$，包含所有的 n 次多项式 (polynomial) $a_0 + a_1x + \ldots a_nx^n$ 。

空间 **Z** 是一个零维空间(合理的维度定义)，**Z** 是最小可能的向量空间，我们有点顾虑的称它为$\mathbf{R}^0$，表示没有分量——你可以想象成没有向量，向量空间 **Z** 确实具有一个向量(零)。每个向量都有自己的零向量——零矩阵，零函数，$\mathbf{R}^3$中的向量(0, 0, 0)。

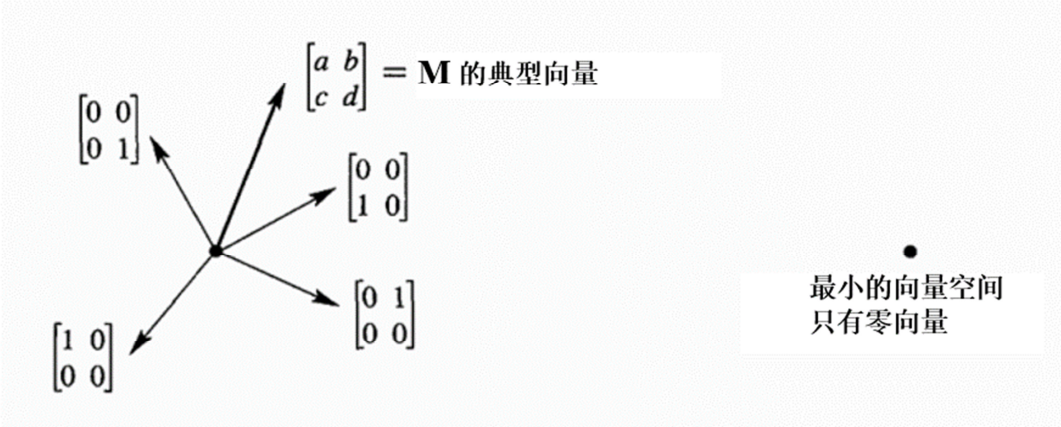


图 3.1：“四维”矩阵空间 **M**，“零维度空间”**Z**。

子空间

在不同的时间，我们要求你把矩阵与函数想象成向量，但是不管什么时候，我们最需要的还是正常的列向量，他们是具有 n 个分量的向量——但是可能不是具有 n 个分量的**全部**向量。在$\mathbf{R}^n$中存在重要的向量空间，他们是$\mathbf{R}^n$的子空间。

从平常的三维空间$\mathbf{R}^3$开始，选择一个通过原点(0, 0, 0)的平面，这个平面本身就是一个向量空间。如果我们把两个在平面上的向量相加，他们的总和也会在这个平面上。如果我们把平面上的向量乘 2 或–5，得到的向量还在这个平面上。三维空间中的平面不是$\mathbf{R}^2$(虽然看起来像$\mathbf{R}^2$)，三维向量具有三个分量，向量属于$\mathbf{R}^3$，平面是$\mathbf{R}^3$中的向量空间。

现在说明线性代数中最重要的基础概念其中之一：通过(0, 0, 0)的平面，是整个向量空间$\mathbf{R}^3$中的**子空间**。

定义　向量空间的**子空间**是一个向量的集合(包含零向量)，这些向量满足两个要求：若 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 是子空间中的向量且 c 是任意纯量，则

(i)　$\boldsymbol{v}+\boldsymbol{w}$ 在子空间中　　(ii)　$c\boldsymbol{v}$ 在子空间中

请尊重版权与译者的劳动成果，侵权必究!

换言之，向量集合中的加法 $\boldsymbol{v}+\boldsymbol{w}$ 及乘法 $c\boldsymbol{v}$(与 $d\boldsymbol{v}$)是封闭的(closed)，这些运算的结果仍然在子空间中。我们也可以执行减法，因为$-\boldsymbol{w}$ 在子空间中，且$-\boldsymbol{w}$ 与 $\boldsymbol{v}$ 的总和是 $\boldsymbol{v}-\boldsymbol{w}$。简单的说，**所有的线性组合仍然在子空间中**。

这些运算遵循**主空间**(host space)的规则，所以自动遵守八个必须条件，我们只需要检验子空间的线性组合的要求即可。

第一个事实：每个子空间都有零向量，$\mathbf{R}^3$中的平面必须通过(0, 0, 0)，我们会再个别讨论，再次额外强调，但是它直接遵循规则(ii)，选取 $c=0$，所需要的 $0\boldsymbol{v}$ 就会在子空间中

不包含原点的平面无法通过试验，这些平面不是子空间。

通过原点的直线也是子空间，当我们对直线上的向量乘 5 或是直线上的两个向量相加，仍然会留在直线上，但是直线必须通过(0, 0, 0)。

另一个子空间是全部的$\mathbf{R}^3$，整个空间是自己的子空间。下列是$\mathbf{R}^3$里面可能的子空间：

(L)	任意通过(0, 0, 0)的直线	**($\mathbf{R}^3$)**	整个空间
(P)	任意通过(0, 0, 0)的平面	**(Z)**	单一向量(0, 0, 0)

如果我们尝试只保留一条直线或是一个平面的一部分，此时就无法满足子空间的要求。检视下列$\mathbf{R}^2$的例子——他们不是子空间。

范例 1　只保留大或等于零的分量的向量(x, y) (这是一个四分之一平面)，包含向量(2, 3)，却不包含(–2, –3)。当我们尝试乘 $c=-1$ 时，会违反规则(ii)，所以四分之一平面不是子空间。

范例 2　范例 1 的条件再加上两个分量同时为负数的向量，现在有两个四分之一平面。规则(ii)满足，因为我们可以乘任意的 c。但是规则(i)失败，$\boldsymbol{v}=(2, 3)$与 $\boldsymbol{w}=(-3, -2)$的总和是(–1, 1)，不在我们的两个四分之一平面上，所以不是子空间。

规则(i)及(ii)与向量加法 $\boldsymbol{v}+\boldsymbol{w}$ 以及纯量 c 与 d 的乘法有关，这两个规则可以合并成为单一的要求——子空间的规则：

> 一个包含 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的子空间，必须包含所有的线性组合 $c\boldsymbol{v}+d\boldsymbol{w}$。

范例 3　若 **M** 是 2×2 矩阵形成的向量空间，下列是两个子空间：

(U) 所有的上三角矩阵 $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$　　**(D)** 所有的对角矩阵 $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$

任意两个 **U** 的矩阵相加，总和还在 **U** 中。对角矩阵相加，总和还是对角。此处的 **D** 也是 **U** 的子空间。当 a, b, d 都是 0，这些子空间当然都包含零矩阵。

单位矩阵的倍数也会形成一个子空间，$2I+3I$ 在子空间中，3 乘 $4I$ 也在子空间中。矩阵 cI 在 **M** 与 **U** 与 **D** 中形成“矩阵线(line of matrix)”。

单位矩阵 I 是自己的子空间吗？当然不是，只有零矩阵才是。你的心里应该想到更多 2×2 矩阵的子空间了——针对问题 5 写出答案。

A 的列空间

最重要的子空间直接与矩阵 A 有关。我们尝试求解 $A\boldsymbol{x}=\boldsymbol{b}$。若 A 不可逆，对于某些 $\boldsymbol{b}$，系统有解；对于某些 $\boldsymbol{b}$，系统却是无解。我们想要描述正确的右侧向量 $\boldsymbol{b}$——可以写成 A 乘某个向量 $\boldsymbol{x}$ 的向量，这些 $\boldsymbol{b}$'s 形成 A 的“列空间”。

记得 $A\boldsymbol{x}$ 是 $\boldsymbol{A}$ **的列**的组合，要想得到任何可能的 $\boldsymbol{b}$，我们使用任何有可能的 $\boldsymbol{x}$。从 A 的列开始，取得他们所有的线性组合，就会得到 A 的列空间，**这是由列向量组成的向量空间**。

$\boldsymbol{C}(A)$不仅包含 A 的 n 个列，而是包含所有列的组合 $A\boldsymbol{x}$。

定义　列空间包含列的所有线性组合，这些组合就就是所有可能的 $A\boldsymbol{x}$，他们构成**列空间** $\boldsymbol{C}(\boldsymbol{A})$。

列空间对于整本书来说非常重要，我们来说明理由。求解 $A\boldsymbol{x}=\boldsymbol{b}$ 就是把 $\boldsymbol{b}$ 表示成列的组合，正确的右侧 $\boldsymbol{b}$ 必须位于左侧 A 的列空间中，否则就是无解。

系统 $A\boldsymbol{x}=\boldsymbol{b}$ 有解　当且仅当　$\boldsymbol{b}$ 在 A 的列空间中。

当 $\boldsymbol{b}$ 在列空间中，它是列的组合，这个组合的系数得到系统 $A\boldsymbol{x}=\boldsymbol{b}$ 的解 $\boldsymbol{x}$。

假设 A 是 $m\times n$ 矩阵，它的列有 m 个分量(不是 n)，所以列属于$\mathbf{R}^m$，A 的列空间是$\mathbf{R}^m$(不是$\mathbf{R}^n$)的子空间。所有列的组合 $A\boldsymbol{x}$ 组成的集合，满足子空间的规则(i)与(ii)：当我们把两个线性组合相加，或乘纯量时，得到的结果仍然是列的线性组合。单词“子空间”是借由取得所有的线性组合得到验证。

假设 A 是 3×2 矩阵，它的列空间是$\mathbf{R}^3$的子空间，A 的列空间是如图 3.2 所示的平面。因为只有两个列，$\boldsymbol{C}(A)$不可能是全部的$\mathbf{R}^3$。

范例 4　假设

$$A\boldsymbol{x}\text{ 是 }\begin{bmatrix}1&0\\4&3\\2&3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}\quad\text{就是}\quad x_1\begin{bmatrix}1\\4\\2\end{bmatrix}+x_2\begin{bmatrix}0\\3\\3\end{bmatrix}$$

这两列的所有组合(就是列空间)形成$\mathbf{R}^3$的平面。我们画出一个特别的 $\boldsymbol{b}$ (列的组合)，$\boldsymbol{b}=A\boldsymbol{x}$ 落在这个平面。平面没有厚度，所以大部分$\mathbf{R}^3$中的右侧 $\boldsymbol{b}$ 不在这个列空间。因为是三个方程式与两个未知数，对于大部分的 $\boldsymbol{b}$ 来说系统都是无解。

请尊重版权与译者的劳动成果，侵权必究！

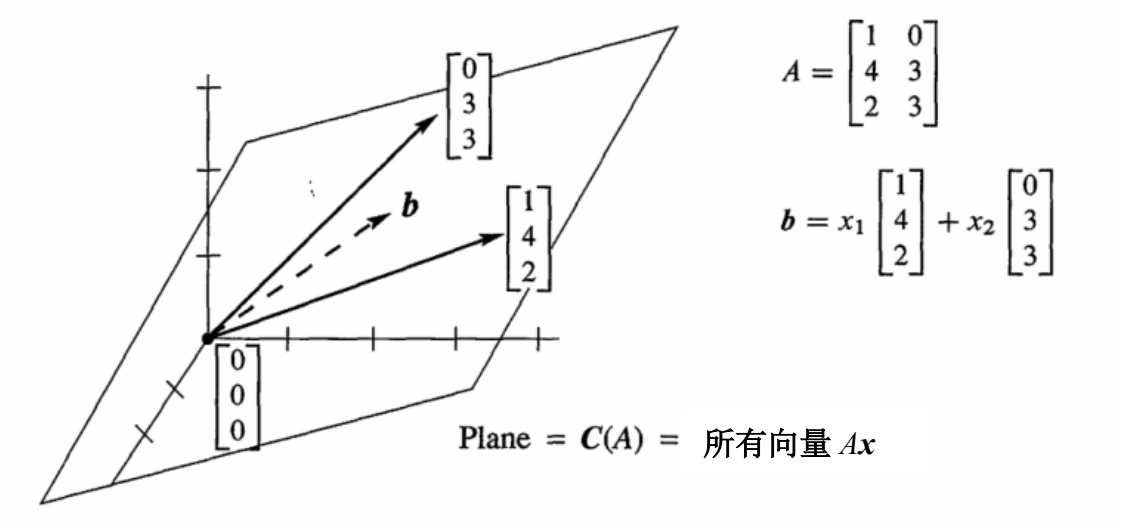


图 3.2：列空间 $\boldsymbol{C}(A)$是一个包含两个列的平面，当 $\boldsymbol{b}$ 落在这个平面时，$A\boldsymbol{x} = \boldsymbol{b}$ 有解，此时 $\boldsymbol{b}$ 是列的组合。

当然(0, 0, 0)会在列空间中，平面会通过原点，$A\boldsymbol{x} = \boldsymbol{0}$ 有一个当然解，这个解一直存在，就是 $\boldsymbol{x} =$ _____。

重复一次，有解的右侧 $\boldsymbol{b}$ 确实在列空间中。一种可能的解就是第一列本身，此时 $x_1 = 1, x_2 = 0$。另一种可能的解就是第二列，此时 $x_1 = 0, x_2 = 1$。认知的新高度在于看见所有的组合——整个子空间是由这两个列创建而成。

表示法 $\boldsymbol{C}(A)$代表 A 的列空间，从这些列开始取得所有列的线性组合，我们可以得到整个$\mathbf{R}^m$，或是一个子空间。

重要 我们不从$\mathbf{R}^m$的列开始，我们从向量空间 **V** 的任意向量集合 **S** 开始。要想得到 **V** 的子空间 **SS**，我们选取集合内向量的所有线性组合：

S = **V** 的向量集合 (可能不是子空间)

SS = **S** 的向量的所有组合 (一定是子空间)

SS = 所有的 $c_1\boldsymbol{v}_1 + \ldots + c_N\boldsymbol{v}_N$ = **S** 生成的(**V** 的子空间)

当 **S** 是列的集合，**SS** 就是列空间。当 **S** 只有一个非零向量 $\boldsymbol{v}$，子空间 **SS** 就是穿过 $\boldsymbol{v}$ 的直线。**SS** 永远是包含 **S** 的最小子空间，这是创建子空间的基础方法，我们还会再回来讨论。

子空间 SS 是 S 的生成(span)，包含 S 中向量的所有组合。

范例 5 描述下列矩阵的列空间(他们是$\mathbf{R}^2$的子空间)：

$$I=\begin{bmatrix}1&0\\0&1\end{bmatrix}\quad A=\begin{bmatrix}1&2\\2&4\end{bmatrix}\quad B=\begin{bmatrix}1&2&3\\0&0&4\end{bmatrix}$$

解 I 的列空间是整个空间$\mathbf{R}^2$，每一个向量都是 I 的列的组合，以向量空间语言来说，$\boldsymbol{C}(I)$是$\mathbf{R}^2$。

A 的列空间只是一条直线，第二列(2, 4)是第一列(1, 2)的倍数。向量虽然不一样，但是我们着重在向量空间。这个列空间包含(2, 4)与(1, 2)，以及其他向量$(c, 2c)$，这些向量都在同一条直线上。当 $\boldsymbol{b}$ 在这条直线时，$A\boldsymbol{x}=\boldsymbol{b}$ 才有解。

第三个矩阵(有三个列)的列空间 $\boldsymbol{C}(B)$是整个$\mathbf{R}^2$，每个向量 $\boldsymbol{b}$ 都有解。向量 $\boldsymbol{b}=(5, 4)$是列 2 加列 3，所以 $\boldsymbol{x}$ 可以是(0, 1, 1)。同样的向量(5, 4)也是 2(列 1) + 列 3，所以另一个可能的 $\boldsymbol{x}$ 是(2, 0, 1)。这个矩阵的列空间与 I 的列空间相同——任意 $\boldsymbol{b}$ 都是允许的，但是现在 $\boldsymbol{x}$ 有额外的分量，有更多的解——更多的组合可以得到 $\boldsymbol{b}$。

下个段落会创造一个向量空间 $\boldsymbol{N}(A)$，用来描述 $A\boldsymbol{x}=\boldsymbol{0}$ 的所有解。本段落创造了列空间 $\boldsymbol{C}(A)$，用来描述所有存在解的右侧 $\boldsymbol{b}$。

主要观念的复习

1. $\mathbf{R}^n$包含所有具有实数分量的列向量。
2. **M**(2×2 矩阵)，**F**(函数)，**Z**(单一零向量)都是向量空间。
3. 包含 $\boldsymbol{v}, \boldsymbol{w}$ 的子空间必须包含所有的线性组合 $c\boldsymbol{v}+d\boldsymbol{w}$。
4. $\boldsymbol{A}$ **的列**的组合形成列空间 $\boldsymbol{C}(A)$，列空间是列的**生成**。
5. 当 $\boldsymbol{b}$ 在 A 的列空间中，$A\boldsymbol{x}=\boldsymbol{b}$ 有解。

$$\boldsymbol{C}(A)=\text{列的所有组合}=\text{所有向量 }A\boldsymbol{x}$$

已解范例

3.1A 给定 3 个不同向量 $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$，创建一个矩阵使得方程式 $A\boldsymbol{x}=\boldsymbol{b}_1$ 与 $A\boldsymbol{x}=\boldsymbol{b}_2$ 有解，但是 $A\boldsymbol{x}=\boldsymbol{b}_3$ 无解。你如何判断这是可能的？你要如何创建 A？

请尊重版权与译者的劳动成果，侵权必究!

解 我们想要 $\boldsymbol{b}_1, \boldsymbol{b}_2$ 在 A 的列空间中，则 $A\boldsymbol{x} = \boldsymbol{b}_1$ 与 $A\boldsymbol{x} = \boldsymbol{b}_2$ 有解。最快的方法是让 $\boldsymbol{b}_1$ 与 $\boldsymbol{b}_2$ 变成 A 的两个列，解就是 $\boldsymbol{x} = (1, 0)$与 $\boldsymbol{x} = (0, 1)$。

同样的，我们不希望 $A\boldsymbol{x} = \boldsymbol{b}_3$ 有解，所以不再变大列空间，只留下列 $\boldsymbol{b}_1$ 与 $\boldsymbol{b}_2$，问题是：

$$A\boldsymbol{x} = \begin{bmatrix} & \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 \\ & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{b}_3 \quad \text{是否有解？} \quad \boldsymbol{b}_3 \text{是否} \boldsymbol{b}_1 \text{与} \boldsymbol{b}_2 \text{的组合?}$$

如果答案是“否”，就得到想要的矩阵 A。如果答案是“是”，就无法建立 A。当列空间包含 $\boldsymbol{b}_1$ 与 $\boldsymbol{b}_2$ 时，必然会包含他们所有的线性组合，如果 $\boldsymbol{b}_3$ 在这个列空间中，此时 $A\boldsymbol{x} = \boldsymbol{b}_3$ 必然有解。

3.1B 描述每个向量空间 **V** 的子空间 **S**，以及 **S** 的子空间 **SS**。

$\mathbf{V}_1 = (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)$的所有组合。

$\mathbf{V}_2 =$ 所有与 $\boldsymbol{u} = (1, 2, 1)$垂直的向量，所以 $\boldsymbol{u} \cdot \boldsymbol{v} = 0$。

$\mathbf{V}_3 =$ 所有的 2×2 对称矩阵(**M** 的子空间)。

$\mathbf{V}_4 =$ 方程式 $d^4y / dx^4 = 0$ 的解(**F** 的子空间)。

使用两个方式描述每个 **V**：“....的所有组合” “方程式....的所有解”

解 $\mathbf{V}_1$ 从三个向量开始，一个子空间 **S** 来自前两个向量(1, 1, 0, 0)与(1, 1, 1, 0)的所有组合。**S** 中存在一个子空间 **SS** 来自第一个向量的所有倍数$(c, c, 0, 0)$。有很多可能性。

$\mathbf{V}_2$ 的一个子空间是穿过(1, –1, 1)的直线，这条直线与 $\boldsymbol{u}$ 垂直。向量 $\boldsymbol{x} = (0, 0, 0)$ 在 **S** 中，并且所有的倍数 $c\boldsymbol{x}$ 得到最小子空间 **SS** = **Z**。

对角矩阵是对称矩阵的一个子空间 **S**，倍数 cI 是对角矩阵的一个子空间 **SS**。

$\mathbf{V}_4$ 包含所有的三次多项式 $y = a + bx + cx^2 + dx^3$，使得 $d^4y / dx^4 = 0$。二次多项式是一个子空间 **S**，线性多项式可以是 **SS** 的一个选择，常数可以是 **SSS**。

全部四个部分都可以选择 **S** = **V** 本身，以及 **SS** = 零子空间 **Z**。

描述每个 **V**： “....的所有组合” “方程式....的所有解”

$\mathbf{V}_1 = 3$ 个向量的所有组合	$\mathbf{V}_1 =$ 所有 $v_1 - v_2 = \mathbf{0}$ 的解。
$\mathbf{V}_2 = (1, 0, -1)$与$(1, -1, 1)$的所有组合	$\mathbf{V}_2 =$ 所有 $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ 的解。
$\mathbf{V}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$的所有组合	$\mathbf{V}_3 =$ 所有 $b = \mathrm{c}$ 的 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$解。
$\mathbf{V}_4 = 1, x, x^2, x^3$ 的所有组合	$\mathbf{V}_4 =$ 所有 $d^4y / dx^4 = 0$ 的解。

问题集 3.1

前 8 个问题 1-8 有关向量空间的一般性质，空间中的向量不需要一定是列向量。在向量空间的定义中，向量加法 $x+y$ 与纯量乘法 cx 必须遵守 8 个规则：

(1) $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$

(2) $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$

(3) 对于所有的 $\boldsymbol{x}$ 存在**唯一的零向量**使得 $\boldsymbol{x}+\boldsymbol{0}=\boldsymbol{x}$

(4) 对于每个 $\boldsymbol{x}$ 存在**唯一的向量** $-\boldsymbol{x}$ 使得 $\boldsymbol{x}+(-\boldsymbol{x})=\boldsymbol{0}$

(5) 1 乘 $\boldsymbol{x}=\boldsymbol{x}$

(6) $(c_1c_2)\boldsymbol{x}=c_1(c_2\boldsymbol{x})$

(7) $c(\boldsymbol{x}+\boldsymbol{y})=c\boldsymbol{x}+c\boldsymbol{y}$

(8) $(c_1+c_2)\boldsymbol{x}=c_1\boldsymbol{x}+c_2\boldsymbol{x}$

(1)到(4) 与 $\boldsymbol{x}+\boldsymbol{y}$ 有关

(5)到(6) 与 $c\boldsymbol{x}$ 有关

(7)到(8) 产生关联

1 假设$(x_1, x_2)+(y_1, y_2)$定义成(x_1+y_2, x_2+y_1)，伴随正常的纯量乘法 $c\boldsymbol{x}=(cx_1, cx_2)$，前面 8 个条件有哪些不满足?

2. 假设乘法 $c\boldsymbol{x}$ 定义成$(cx_1, 0)$而不是(cx_1, cx_2)，伴随正常的 $\mathbf{R}^2$ 加法，前面 8 个条件有哪些满足?

3 (a) 如果 $\mathbf{R}^1$ 只保留正数 $x>0$，前面 8 个条件有哪些崩溃？必须允许每个 c，这条半线不是空间。

(b) 对于正数来说，把 $\boldsymbol{x}+\boldsymbol{y}$ 与 $c\boldsymbol{x}$ 的定义改成一般的 xy 与 x^c，还是满足这 8 个条件。测试规则 7：当 $c=3, x=2, y=1$。(则 $\boldsymbol{x}+\boldsymbol{y}=2$，$c\boldsymbol{x}=8$)。哪个数字扮演零向量的角色？

4 矩阵 $A=\begin{bmatrix}2 & -2\\ 2 & -2\end{bmatrix}$是所有 2×2 矩阵形成的 M 空间的一个向量，写出此空间的零向量，向量 $A/2$，向量$-A$。哪些矩阵会在包含 A 的最小子空间中？

5 (a) 描述 M 的一个子空间包含 $A=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$，但不包含 $B=\begin{bmatrix}0 & 0\\ 0 & -1\end{bmatrix}$。

(b) 若 M 的子空间确实包含 A 与 B，就一定包含 I 吗？

(c) 描述 M 的一个子空间，这个子空间包含所有对角线都是 0 的矩阵。

6 函数 $\boldsymbol{f}(x)=x^2$ 与 $\boldsymbol{g}(x)=5x$ 是 $\mathbf{F}$ 中的“向量”，这是所有实数函数的空间。(函数定义在$-\infty<x<\infty$。) 组合 $3\boldsymbol{f}(x)-4\boldsymbol{g}(x)$就是函数 $\boldsymbol{h}(x)=$ _____。

请尊重版权与译者的劳动成果，侵权必究!

7 如果 $\boldsymbol{f}(x)$乘 c 得到 $\boldsymbol{f}(cx)$，哪些规则会被破坏？保持正常的加法 $\boldsymbol{f}(x)+\boldsymbol{g}(x)$。

8 如果向量 $\boldsymbol{f}(x)$与 $\boldsymbol{g}(x)$的总和定义成 $\boldsymbol{f}(\boldsymbol{g}(x))$，则零向量就是 $\boldsymbol{g}(x)=\boldsymbol{x}$。保持正常的纯量乘法 $c\boldsymbol{f}(x)$，并且找出两个崩溃的规则。

问题 9-18 有关“子空间的要求”：$\boldsymbol{x}+\boldsymbol{y}$ 与 $c\boldsymbol{x}$ (以及所有的线性组合 $c\boldsymbol{x}+d\boldsymbol{y}$)保持在空间中。

9 一个要求满足，另一个不能满足。找出例子证明：

(a) $\mathbf{R}^2$的一组向量集合，其中 $\boldsymbol{x}+\boldsymbol{y}$ 还在其中，但是 $\boldsymbol{x}/2$ 已经超出。

(b) $\mathbf{R}^2$的一组向量集合(不是两个四分之一平面)，其中 $c\boldsymbol{x}$ 还在其中，但是 $\boldsymbol{x}+\boldsymbol{y}$ 已经超出。

10 哪些$\mathbf{R}^3$的集合实际上是子空间?

(a) 向量(b_1, b_2, b_3)的平面，其中 $b_1=b_2$。

(b) 向量(b_1, b_2, b_3)的平面，其中 $b_1=1$。

(c) 向量(b_1, b_2, b_3)的平面，其中 $b_1\,b_2\,b_3=0$。

(d) $\boldsymbol{v}=(1, 4, 0)$与 $\boldsymbol{w}=(2, 2, 2)$的所有线性组合。

(e) 所有满足 $b_1+b_2+b_3=0$ 的向量。

(f) 所有满足 $b_1\le b_2\le b_3$ 的向量。

11 描述矩阵空间 M 的最小子空间，这个最小子空间包含：

(a) $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ 与 $\begin{bmatrix}0&1\\0&0\end{bmatrix}$ (b) $\begin{bmatrix}1&1\\0&0\end{bmatrix}$ (c) $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ 与 $\begin{bmatrix}1&0\\0&1\end{bmatrix}$

12 令 P 是$\mathbf{R}^3$的平面，方程式是 $x+2y-2z=4$，原点(0, 0, 0)不在 P 中。找出 P 的两个向量，验证他们的总和不在 P 中。

13 令 $\mathbf{P}_0$ 是通过(0, 0, 0)的平面，而且与前面的平面 P 平行。$\mathbf{P}_0$ 的方程式是什么？找出 $\mathbf{P}_0$ 的两个向量，验证他们的总和在 $\mathbf{P}_0$ 中。

14 $\mathbf{R}^3$的子空间是平面，直线，$\mathbf{R}^3$本身，或是 $\mathbf{Z}$ 只包含(0, 0, 0):

(a) 描述$\mathbf{R}^2$的子空间的三个形式。

(b) 给定 2×2 的对角矩阵形成的空间 $\mathbf{D}$，请描述 $\mathbf{D}$ 的所有子空间。

15 (a) 两个通过(0, 0, 0)的平面的交集可能是$\mathbf{R}^3$的_____，也可能是_____，只是不可能是 **Z**。

(b) 一个通过(0, 0, 0)的平面与一条通过(0, 0, 0)的直线，两者的交集可能是_____，也可能是_____。

(c) 若 **S** 与 **T** 是$\mathbf{R}^5$的子空间，证明他们的交集 **S**∩ **T** 是$\mathbf{R}^5$的子空间。此处的 **S**∩ **T** 包含同时在两个子空间中的向量。若 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 同时在两个空间中，验证 $\boldsymbol{x}+\boldsymbol{y}$ 与 $c\boldsymbol{x}$ 也在 **S**∩ **T** 中。

16 假设 **P** 是一个通过(0, 0, 0)的平面，**L** 是一条通过(0, 0, 0)的直线。同时包含 **P** 与 **L** 的最小的向量空间不是_____就是_____。

17 (a) 证明在 **M** 中的可逆矩阵的集合不是子空间。

(b) 证明在 **M** 中的奇异矩阵的集合不是子空间。

18 是非题 (利用一个例子检验每个案例的加法)

(a) 在 **M** 中的对称矩阵($A^{\mathrm{T}}=A$)形成一个子空间。

(b) 在 **M** 中的反对称矩阵($A^{\mathrm{T}}=-A$)形成一个子空间。

(c) 在 **M** 中的非对称矩阵($A^{\mathrm{T}}\neq A$)形成一个子空间。

问题 19-27 有关列空间 $C(A)$与方程式 $Ax=b$。

19 描述下列矩阵的列空间(直线或平面)：

$$A=\begin{bmatrix}1&2\\0&0\\0&0\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1&0\\0&2\\0&0\end{bmatrix} \text{ 与 } C=\begin{bmatrix}1&0\\2&0\\0&0\end{bmatrix}$$

20 哪些正确的右侧(找出 b_1, b_2, b_3 的条件)，使得系统有解？

$$\text{(a)}\begin{bmatrix}1&4&2\\2&8&4\\-1&-4&-2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \qquad \text{(b)}\begin{bmatrix}1&4\\2&9\\-1&-4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

21 A 的行 1 加到行 2 得到 B，列 1 加到列 2 得到 C。(B 或 C)的列的组合也是 A 的列的组合？哪两个矩阵有相同的列_____：

$$A=\begin{bmatrix}1&2\\2&4\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1&2\\3&6\end{bmatrix} \text{ 与 } C=\begin{bmatrix}1&3\\2&6\end{bmatrix}$$

22 哪些向量(b_1, b_2, b_3)使得下列系统有解?

$$\begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \text{ 与 } \begin{bmatrix}1&1&1\\0&1&1\\0&0&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix} \text{ 与 } \begin{bmatrix}1&1&1\\0&0&1\\0&0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

23 (推荐) 引进额外的列向量 $\boldsymbol{b}$ 进入矩阵 A，除非_____，列空间会变大。各举一个例子来说明列空间变大与列空间不变的情形。为什么只有当列空间不会变大—— A 的列空间与$[A \quad \boldsymbol{b}]$的列空间相同时，$A\boldsymbol{x}=\boldsymbol{b}$ 才有解？

24 AB 的列是 A 的列的组合，意思是：AB 的列空间被包含在(有可能相等)A 的列空间中。举例说明 A 的列空间与 AB 的列空间不相等的情形。

25 假设 $A\boldsymbol{x}=\boldsymbol{b}$ 与 $A\boldsymbol{y}=\boldsymbol{b}^*$都有解，则 $A\boldsymbol{z}=\boldsymbol{b}+\boldsymbol{b}^*$有解，求 $\boldsymbol{z}$？也可以表示成：若 $\boldsymbol{b}$ 与 $\boldsymbol{b}^*$在列空间 $\boldsymbol{C}(A)$中，则 $\boldsymbol{b}+\boldsymbol{b}^*$也在 $\boldsymbol{C}(A)$中。

26 若 A 是任意 5×5 的可逆矩阵，则它的列空间是_____，为什么？

27 是非题 (如果错误，举出反例):

(a) 不在列空间 $\boldsymbol{C}(A)$中的向量 $\boldsymbol{b}$ 形成一个子空间。

(b) 若 $\boldsymbol{C}(A)$只包含零向量，则 A 是一个零矩阵。

(c) $2A$ 的列空间等于 A 的列空间。

(d) $(A-I)$的列空间等于 A 的列空间(请测试)。

28 创建一个 3×3 矩阵，它的列空间包含(1, 1, 0)与(1, 0, 1)，但不包含(1, 1, 1)。创建一个 3×3 矩阵，它的列空间只是一条直线。

29 若一个 9×12 的系统，给定任意的 $\boldsymbol{b}$，$A\boldsymbol{x}=\boldsymbol{b}$ 都有解，则 $\boldsymbol{C}(A)=$ _____。

挑战问题

30 假设 **S** 与 **T** 是向量空间 **V** 的两个子空间。

(a) **定义**：总和 **S** + **T** 包含所有的 $\boldsymbol{s}+\boldsymbol{t}$，其中 $\boldsymbol{s}$ 来自 **S** 且 $\boldsymbol{t}$ 来自 **T**。证明 **S** + **T** 满足向量空间的要求(加法与纯量乘法)。

(b) 若 **S** 与 **T** 是$\mathbf{R}^5$中的线，则 **S** + **T** 与 **S** ∪ **T** 之间有何不同？联集(union)包含所有来自 **S** 或 **T** 或同时来自两者的向量。说明下列的叙述：**S** ∪ **T** 的生成是 **S** + **T**。(段落 3.5 再出现 span 的字词)。

31 若 **S** 是 A 的列空间，**T** 是 B 的列空间 $\boldsymbol{C}(B)$，则 **S** + **T** 是哪个矩阵 M 的列空间？A 与 B 与 M 的列空间都在$\mathbf{R}^m$中。(我不认为 $A+B$ 永远是正确的 M)。

32 证明矩阵 A 与$[A \quad AB]$ (多了额外的列)有相同的列空间。但是请找出一个方形矩阵，这个矩阵的 $\boldsymbol{C}(A^2)$比 $\boldsymbol{C}(A)$小。**重点：**

只有当 $n\times n$ 矩阵 A 是一个_____矩阵，才有 $\boldsymbol{C}(A)=\mathbf{R}^n$。

3.2 A 的零空间：求解 $Ax=0$ 与 $Rx=0$

1. $\mathbf{R}^n$中的**零空间** $N(A)$包含所有 $Ax=0$ 的解 x，也包含 $x=0$。
2. 消元法(从 A 到 U 到 R)不会改变零空间：$N(A)=N(U)=N(R)$。
3. **简化行阶梯形式** $R=$ **rref(**A**)**的所有主元都是 1，上方及下方都是 0。
4. 若 R 的列 j 是自由(没有主元)，则 $Ax=0$ 会有一个 $x_j=1$ 的**特殊解**[1]。
5. 主元的数目 $=R$ 的非零行的数目 $=$ **秩** r，会有 $n-r$ 个自由列。
6. 每个 $m<n$ 的矩阵，$Ax=0$ 的零空间中存在非零解。

本段落是有关包含所有 $Ax=0$ 的解的子空间。一个 $m\times n$ 矩阵 A 可以是方形也可以是矩形，右侧的 b 是零，一个立即的解就是 $x=0$。对于可逆矩阵而言，这是唯一解。其他不可逆矩阵就有 $Ax=0$ 的非零解，这些解都属于 A 的零空间。

消元法可以找到所有的解，并且鉴别这个非常重要的子空间。

零空间 $N(A)$包含所有 $Ax=0$ 的解，这些向量 x 都在$\mathbf{R}^n$中。

验证这些解可以形成子空间，假设 x 与 y 在零空间中(意思是 $Ax=0$, $Ay=0$)，矩阵乘法规则会得到 $A(x+y)=0+0$，而且 $A(cx)=c0$，右侧仍然都是零，因此 $x+y$ 与 cx 都在零空间 $N(A)$中。我们可以在不离开零空间的情形下，完成向量加法与纯量乘法，所以它是一个零空间。

重复说明：解 x 有 n 个分量，他们是$\mathbf{R}^n$中的向量，所以零空间是$\mathbf{R}^n$的一个子空间，列空间 $C(A)$是$\mathbf{R}^m$的一个子空间。

范例 1 描述 $A=\begin{bmatrix}1 & 2\\ 3 & 6\end{bmatrix}$的零空间，这个矩阵是奇异的：

解 应用消元法至线性方程式 $Ax=0$：

$$\begin{matrix}x_1+2x_2=0\\ 3x_1+6x_2=0\end{matrix}\quad\rightarrow\quad\begin{matrix}x_1+2x_2=0\\ 0=0\end{matrix}$$

实际上只有一个方程式，第二个方程式是第一个方程式的 3 倍。在行图中，直线 $x_1+2x_2=0$ 与 $3x_1+6x_2=0$ 完全相同，这条直线就是零空间 $N(A)$，它包含所有的解(x_1, x_2)。

要描述 $Ax=0$ 的解，有一个很有效率的方法。选取一个直线上的点(一个特殊解)，所有直线上的点都是这个点的倍数。选取第二个分量 $x_2=1$(特殊选择)，依据方程式 $x_1+2x_2=0$，第一分量必须是 $x_1=-2$，这个**特殊解**是 $s=(-2, 1)$。

---

[1] **译者注：**原文的 special solution 翻译成特殊解。下个段落出现的 particular solution 翻译成“**特解**”，用来区分“**特殊解**”。】

请尊重版权与译者的劳动成果，侵权必究!

特殊解 $A\boldsymbol{s}=\boldsymbol{0}$　$A=\begin{bmatrix}1&2\\3&6\end{bmatrix}$ 的零空间包含 $\boldsymbol{s}=\begin{bmatrix}-2\\1\end{bmatrix}$ 的所有倍数。

借由计算 $A\boldsymbol{x}=\boldsymbol{0}$ 的特殊解，这是描述零空间最好的方法。因为我们令自由变量 $x_2=1$，解是特殊解。

A 的零空间包含 $A\boldsymbol{x}=\boldsymbol{0}$ 的特殊解的所有组合。

范例 2　$x+2y+3z=0$ 来自 1×3 矩阵 $A=[1\quad 2\quad 3]$，$A\boldsymbol{x}=\boldsymbol{0}$ 产生一个平面，平面上所有的向量都与(1, 2, 3)垂直，这个平面是 A 的零空间。此处有两个自由变量 y 与 z：**设成 0 与 1**：

$$\begin{bmatrix}1&2&3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=0 \text{ 有两个特殊解 } \boldsymbol{s}_1=\begin{bmatrix}-2\\1\\0\end{bmatrix} \text{ 与 } \boldsymbol{s}_2=\begin{bmatrix}-3\\0\\1\end{bmatrix}$$

这些向量 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 位于平面 $x+2y+3z=0$，平面上所有的向量都是 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 的组合。

注意 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 有何特殊，最后两个分量是自由的，我们特意选成 1, 0 与 0, 1，则第一分量的–2 与–3 可以由方程式 $A\boldsymbol{x}=\boldsymbol{0}$ 求得。

$x+2y+3z=\boldsymbol{6}$ 的解也是一个平面，但是这个平面不是子空间，只有 $\boldsymbol{b}=\boldsymbol{0}$ 时向量 $\boldsymbol{x}=\boldsymbol{0}$ 才是解。段落 3.3 会说明 $A\boldsymbol{x}=\boldsymbol{b}$ 的解(如果有解)会如何借由一个特殊解而离开零。

本段落的两个关键步骤：
(1) 简化 A 得到行阶梯形式 R
(2) 求 $A\boldsymbol{x}=\boldsymbol{0}$ 的特殊解

138 页所示 4×5 矩阵 A 与 R，有 3 个主元

方程式 $A\boldsymbol{x}=0$ 以及 $R\boldsymbol{x}=0$ 都有 $5-3=2$ 个特殊解 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$。

主元列与自由列

$A=[1\quad 2\quad 3]$的第一列包含唯一的主元，所以 $\boldsymbol{x}$ 的第一分量不是自由，**自由分量是对应到没有主元的列**。特殊选择(1 或 0)只能应用在特殊解的自由变量。

范例 3　求 A, B, C 的零空间，以及 $C\boldsymbol{x}=\boldsymbol{0}$ 的两个特殊解。

$$A=\begin{bmatrix}1&2\\3&8\end{bmatrix}\qquad B=\begin{bmatrix}A\\2A\end{bmatrix}=\begin{bmatrix}1&2\\3&8\\2&4\\6&16\end{bmatrix}\qquad C=\begin{bmatrix}A&2A\end{bmatrix}=\begin{bmatrix}1&2&2&4\\3&8&6&16\end{bmatrix}$$

解 方程式 $A\boldsymbol{x}=\mathbf{0}$ 只有一个解 $\boldsymbol{x}=\mathbf{0}$，零空间就是 $\mathbf{Z}$，它只包含$\mathbf{R}^2$的单一点 $\boldsymbol{x}=\mathbf{0}$。这个事实来自消元法：

$$A\boldsymbol{x}=\begin{bmatrix}1&2\\3&8\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \text{ 得到 } \begin{bmatrix}1&2\\0&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \text{ 与 } \begin{bmatrix}x_1=0\\x_2=0\end{bmatrix}$$

A 是可逆，没有特殊解，矩阵的两个列都有主元。

矩形矩阵 B 有相同的零空间 $\mathbf{Z}$。$B\boldsymbol{x}=\mathbf{0}$ 的前两个方程式再次需要 $\boldsymbol{x}=\mathbf{0}$，最后两个方程式也会强制 $\boldsymbol{x}=\mathbf{0}$。当我们加入额外的方程式(给了额外的行)，零空间当然不能变大，多出来的行会增加零空间中的 $\boldsymbol{x}$ 更多的条件。

矩形矩阵 C 就不一样了，它有额外的列而不是额外的行，解 $\boldsymbol{x}$ 有 4 个分量。消元法会在 C 的前两列产生主元，但是 C 与 U 的后两列都是自由的，他们没有主元：

从 C 的行 2 减去 3 倍的行 1 $\quad C=\begin{bmatrix}1&2&2&4\\3&8&6&16\end{bmatrix}$ 变成 $U=\begin{bmatrix}1&2&2&4\\0&2&0&4\end{bmatrix}$

主元列　自由列

对自由变量 x_3 与 x_4，我们选取特殊的 1 与 0。首先 $x_3=1$ 与 $x_4=0$，第二个是 $x_3=0$ 与 $x_4=1$，主元变量 x_1 与 x_2 由方程式 $U\boldsymbol{x}=\mathbf{0}$(或 $C\boldsymbol{x}=\mathbf{0}$ 或等效的 $R\boldsymbol{x}=\mathbf{0}$)决定。我们得到 C 的零空间的两个特殊解，这个也是 U 的零空间：消元法不会改变解。

$C\boldsymbol{s}=\mathbf{0}$ 与 $U\boldsymbol{s}=\mathbf{0}$ 的特殊解 $\quad \boldsymbol{s}_1=\begin{bmatrix}-2\\0\\1\\0\end{bmatrix}$ $\boldsymbol{s}_1=\begin{bmatrix}0\\-2\\0\\1\end{bmatrix}$ $\begin{matrix}\leftarrow \text{主元}\\ \leftarrow \text{变量}\\ \leftarrow \text{自由}\\ \leftarrow \text{变量}\end{matrix}$

简化行阶梯形式 R

当 A 是矩形时，消元法不会在上三角 U 停止，我们可以继续使得矩阵更简化——有两个方法，这些步骤会得到最佳矩阵 R：

1. 在主元上方产生 0	**在 R 中，利用主元行往上消元**
2. 在主元产生 1	**整个主元行除以自己的主元**

这些步骤不会改变方程式右侧的零向量，零空间仍然保持相同：$\boldsymbol{N}(A)=\boldsymbol{N}(U)=\boldsymbol{N}(R)$。当我们得到**简化行阶梯形式 $\boldsymbol{R}=\mathbf{rref}(\boldsymbol{A})$**，零空间会变得更加容易观察，**$R$ 的主元列会包含 I**。

请尊重版权与译者的劳动成果，侵权必究!

简化形式 R　　$U=\begin{bmatrix}1&2&2&4\\0&2&0&4\end{bmatrix}$ 变成 $R=\begin{bmatrix}1&0&2&0\\0&1&0&2\end{bmatrix}$

我从 U 的行 1 减去行 2，然后对行 2 乘 1/2 得到主元 = 1。现在(自由列 3) = 2(主元列 1)，所以–2 会出现在 $\boldsymbol{s}_1=(-2,0,1,0)$。从简化系统 $R\boldsymbol{x}=\boldsymbol{0}$ 找出特殊解就容易多了。对于 R 的每个自由列，我会改变全部的符号得到 $\boldsymbol{s}$，第二个特殊解 $\boldsymbol{s}_2=(0,-2,0,1)$。

转移到 $m\times n$ 矩阵与他们的零空间 $\boldsymbol{N}(A)$以及特殊解之前，请允许我再重复一个注解。对于很多矩阵而言，$A\boldsymbol{x}=\boldsymbol{0}$ 唯一的解是 $\boldsymbol{x}=\boldsymbol{0}$，零空间 $\boldsymbol{N}(A)=\boldsymbol{Z}$(只有一个零向量)而且没有特殊解。唯一可以得到 $\boldsymbol{b}=\boldsymbol{0}$ 的列组合就是“零组合”，$A\boldsymbol{x}=\boldsymbol{0}$ 的解就是平凡(trivial)解(只有 $\boldsymbol{x}=\boldsymbol{0}$)，但是这个概念却不平凡。

零的零空间(zero nullspace) $\boldsymbol{Z}$ 的案例非常重要，它说明了 A 的列是无关。没有任何列的组合可以得到零向量(除了零组合)，所有的列都有主元，没有自由列。你会一再的看到**无关**的概念。

阶梯矩阵 R 的主元变量与自由变量

$$A=\begin{bmatrix}p&p&f&p&f\\|&|&|&|&|\\|&|&|&|&|\\|&|&|&|&|\\|&|&|&|&|\end{bmatrix}\quad R=\begin{bmatrix}\boldsymbol{1}&0&a&0&c\\0&\boldsymbol{1}&b&0&d\\0&0&0&\boldsymbol{1}&e\\0&0&0&0&0\end{bmatrix}\quad \boldsymbol{s}_1=\begin{bmatrix}-a\\-b\\\boldsymbol{1}\\0\\\boldsymbol{0}\end{bmatrix}\quad \boldsymbol{s}_2=\begin{bmatrix}-c\\-d\\\boldsymbol{0}\\-e\\\boldsymbol{1}\end{bmatrix}$$

3 个主元列 p	I 在主元列	特殊解 $R\boldsymbol{s}_1=\boldsymbol{0}$，$R\boldsymbol{s}_2=\boldsymbol{0}$
2 个自由列 f	F 在自由列	从 R 选取–a 到–e
由 R 显示	3 个主元：秩= 3	$R\boldsymbol{s}=\boldsymbol{0}$ 就是 $A\boldsymbol{s}=\boldsymbol{0}$

R 明确显示：列 3 = a(列 1) + b(列 2)，A 也必须有同样结果。

特殊解 $\boldsymbol{s}_1$ 重复前述组合，使得$(-a,-b,1,0,0)$有 $R\boldsymbol{s}_1=\boldsymbol{0}$。

A 的零空间 = R 的零空间 = $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 的所有组合。

下列是具有 3 个主元的 4×7 简化行阶梯矩阵 R 的步骤：

$$R=\begin{bmatrix}\boldsymbol{1}&\boldsymbol{0}&x&x&x&\boldsymbol{0}&x\\\boldsymbol{0}&\boldsymbol{1}&x&x&x&\boldsymbol{0}&x\\\boldsymbol{0}&\boldsymbol{0}&0&0&0&\boldsymbol{1}&x\\0&0&0&0&0&0&0\end{bmatrix}$$

3 个主元变数 x_1, x_2, x_6
4 个自由变数 x_3, x_4, x_5, x_7
4 个特殊解 s 在 $\boldsymbol{N}(R)$中
主元行与主元列包含 I

问题：上述矩阵 R 的列空间是什么？零空间是什么？

解 R 的列有 4 个分量，所以他们在$\mathbf{R}^4$中(不是$\mathbf{R}^3$！)。每个列的第四分量是 0，列的每个组合——列空间的每个向量——第四个分量都会是 0。列空间 $\boldsymbol{C}(R)$包含所有具有$(b_1, b_2, b_3, 0)$形式的向量，对于这些向量可以求解 $R\boldsymbol{x} = \boldsymbol{b}$。

零空间 $\boldsymbol{N}(R)$是$\mathbf{R}^7$的子空间，$R\boldsymbol{x} = \mathbf{0}$ 的解是 4 个特殊解的所有组合——每个特殊解对应一个自由变量：

1. 列 3, 4, 5, 7 没有主元，所以四个自由变量是 x_3, x_4, x_5, x_7。
2. 令一个自由变量为 1，再令其他三个自由变量为 0。
3. 求解 $R\boldsymbol{x} = \mathbf{0}$ 得到主元变量 x_1, x_2, x_6，就可以求得 $\boldsymbol{s}$。

计算主元的个数得到一个非常重要的定理。假设 A 的列数比行数多，当 $n > m$ 时，最少存在一个自由变量，系统 $A\boldsymbol{x} = \mathbf{0}$ 最少有一个特殊解，这个解不是零。

假设 $A\boldsymbol{x} = \mathbf{0}$ 的**未知数的个数比方程式的个数多**($n > m$，列数比行数多)，则至少存在一个自由列，$A\boldsymbol{x} = \mathbf{0}$ 有非零解。

一个矮宽矩阵($n > m$)在它的零空间中必然有非零向量，由于主元的数量不可能超过 m，所以最少会有 $n - m$ 个自由变量。(矩阵只有 m 个行，一个行不可能有 2 个主元。) 当然一个行可能没有主元——表示存在额外的自由变量，但是重点在这里：当存在自由变量时，它就可以被设成 1，则方程式 $A\boldsymbol{x} = \mathbf{0}$ 最少有一个非零解。

零空间是一个子空间，它的维度与自由变量的维度相等。这个中心思想——子空间的维度——本章会给与定义与说明。

矩阵的秩

数字 m 与 n 决定了矩阵的大小——但是不一定是线性系统真正的大小。比如 0 = 0 的方程式就不能算在里面。如果 A 有两个完全相同的行，那么第二个行在消元之后会消失。同理若行 3 是行 1 与行 2 的组合，在三角 U 中行 3 会变成都是 0，而且在简化阶梯形式 R 中也是相同情形。我们不想计算零的行，A 的真正大小是由**秩**(rank)来决定。

秩的定义 **A 的秩是主元的数量，数值为 r。**

这个定义是计算上的定义，我想要说更多有关秩 r 的事情，最后的矩阵 R 会有 r 个非零行。从一个 3×4 矩阵开始，它的秩 $r = 2$：

4 个列，2 个主元
$$A = \begin{bmatrix} 1 & 1 & 2 & 4 \\ 1 & 2 & 2 & 5 \\ 1 & 3 & 2 & 6 \end{bmatrix} \qquad R = \begin{bmatrix} \mathbf{1} & \mathbf{0} & 2 & 3 \\ \mathbf{0} & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A 的前两个列是(1, 1, 1)与(1, 2, 3)，分处不同的方向，这些是主元列(R 会显示)。第三列(2, 2, 2)是第一列的 2 倍，我们在第三列看不到主元。第四列(4, 5, 6)是前三

请尊重版权与译者的劳动成果，侵权必究！

列的总和，第四列也没有主元，所以 A 与 R 的秩就是 2。

每一个自由列都是较早的主元列的组合，特殊解 s 会告诉我们这些组合：

列 3 = **2**(列 1) + **0**(列 2)　　$\boldsymbol{s}_1 = (\mathbf{-2}, \mathbf{-0}, 1, 0)$

列 4 = **3**(列 1) + **1**(列 2)　　$\boldsymbol{s}_2 = (\mathbf{-3}, \mathbf{-1}, 0, 1)$

R 的第三列的数字 2 与 0 在 $\boldsymbol{s}_1$ 中出现(符号相反)；R 的第四列的数字 3 与 1 在 $\boldsymbol{s}_2$ 中出现(符号相反，变成–3 与–1)。

秩一

秩一的矩阵只有一个主元，当消元法在第一列产生零时，其他列也会得到零。每个行都是主元行的倍数，同时每个列都是主元列的倍数。

秩一矩阵　$A = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ \mathbf{2} & 6 & 20 \\ \mathbf{3} & 9 & 30 \end{bmatrix} \rightarrow R = \begin{bmatrix} \mathbf{1} & 3 & 10 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

秩一矩阵的列空间是“一维”，此处所有的列都会通过 $\boldsymbol{u} = (1, 2, 3)$的直线。$A$ 的列分别是 $\boldsymbol{u}, 3\boldsymbol{u}, 10\boldsymbol{u}$，把这些数字放在一个行 $\boldsymbol{v}^{\mathrm{T}} = [1 \quad 3 \quad 10]$，你就可以得到特殊的秩一形式 $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ ：

$$\boldsymbol{A} = \text{列乘行} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \qquad \begin{bmatrix} 1 & 3 & 10 \\ 2 & 6 & 20 \\ 3 & 9 & 30 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{3} & \mathbf{10} \end{bmatrix}$$

A 是秩一时，$A\boldsymbol{x} = \mathbf{0}$ 很容易了解，方程式 $\boldsymbol{u}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}) = \mathbf{0}$，让我们得到 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x} = \mathbf{0}$。所有零空间中的向量 $\boldsymbol{x}$，都必须与行空间中的 $\boldsymbol{v}$ 垂直。当 $r = 1$ 的几何意义是：行空间 = 直线，零空间 = 垂直平面。

范例 4　当所有的行都是主元行的倍数时，秩是 $r = 1$：

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix} \text{与} \begin{bmatrix} 0 & 3 \\ 0 & 5 \end{bmatrix} \text{与} \begin{bmatrix} 5 \\ 2 \end{bmatrix} \text{与} \begin{bmatrix} 6 \end{bmatrix} \quad \text{都是秩一矩阵}$$

对这些矩阵而言，简化行阶梯 R= rref(A)可以用眼睛检验：

$$R = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & 0 \end{bmatrix} \text{与} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{与} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{与} \begin{bmatrix} 1 \end{bmatrix} \quad \text{都只有一个主元}$$

我们关于秩的第二个定义会比较高阶，它处理整个行以及整个列—向量而且不只是数字。全部三个矩阵 A 与 U 与 R 都有 r 个无关行。

A 与 U 与 R 同样也有 r 个**无关列(主元列)**，段落 3.4 会说明无关行或是无关列的意义。

秩的第三个定义，在线性代数的顶级层次会牵涉到向量空间。秩 r 就是列空间的维度，也是行空间的维度。伟大的地方在于 $n-r$ 就是零空间的维度。

主要观念的复习

1. 零空间 $\boldsymbol{N}(A)$是$\mathbf{R}^n$的子空间，包含所有 $A\boldsymbol{x}=\mathbf{0}$ 的解。
2. 对 A 执行消元法得到一个行简化 R，具有主元列及自由列。
3. 每个自由列得到一个特殊解，令一个自由变量为 1，其他为 0。
4. A 的秩 r 是主元的数目，$R=\text{rref}(A)$的所有主元都是 1's。
5. $A\boldsymbol{x}=\mathbf{0}$ 的完整解是 $n-r$ 个特殊解的组合。
6. 若 $n>m$，A 总是有一个自由列，得到一个 $A\boldsymbol{x}=\mathbf{0}$ 的非零解。

已解范例

3.2A 若 $EA=R$ 且 E 是可逆，为什么 A 与 R 有相同的零空间?

解 若 $A\boldsymbol{x}=\mathbf{0}$ 则 $R\boldsymbol{x}=EA\boldsymbol{x}=E\mathbf{0}=\mathbf{0}$

若 $R\boldsymbol{x}=\mathbf{0}$ 则 $A\boldsymbol{x}=E^{-1}R\boldsymbol{x}=E^{-1}\mathbf{0}=\mathbf{0}$

A 与 R 有同样的行空间及同样的秩。

3.2B 建立一个 3×4 矩阵 R，且 $R\boldsymbol{x}=\mathbf{0}$ 的特殊解是 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$：

$$\boldsymbol{s}_1=\begin{bmatrix}-3\\1\\0\\0\end{bmatrix}\qquad \boldsymbol{s}_2=\begin{bmatrix}-2\\0\\-6\\1\end{bmatrix}$$

主元列 1 与 3，自由变量 x_2 与 x_4。

描述所有可能的矩阵 A，其中零空间 $\boldsymbol{N}(A)=\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 的所有组合。

解 简化矩阵 R 有主元 $=1$ 在列 1 与 3，没有第三个主元，所以 R 的行 3 全部是零。自由列 2 与 4 是主元列的组合：R 的 3, 0, 2, 6 来自 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$ 的–3, –0, –2, –6。每个 $A=ER$。

每一个 3×4 矩阵至少有一个特殊解，这个矩阵有两个特殊解：

$$R=\begin{bmatrix}\mathbf{1}&\mathbf{3}&0&\mathbf{2}\\0&0&\mathbf{1}&\mathbf{6}\\0&0&0&0\end{bmatrix}\quad \text{有 } R\boldsymbol{s}_1=\mathbf{0} \text{ 与 } R\boldsymbol{s}_2=\mathbf{0}$$

请尊重版权与译者的劳动成果，侵权必究!

3.2C　求行简化形式 R 以及 A 与 B 个别的秩(与 c 有关)。哪些是 A 的主元列？特殊解为何？

求特殊解　　$A=\begin{bmatrix}1&2&1\\3&6&3\\4&8&c\end{bmatrix}$　　$B=\begin{bmatrix}c&c\\c&c\end{bmatrix}$

解　矩阵 A 有行 2 = 3(行 1)，如果 c 不是 4，则 A 的秩就是 2。行 **3** – 4(行 1)的末端得到 $c-4$，此时主元在列 1 与列 3，第二个变量 x_2 是自由的。注意 R 的形式：行 3 已经移到行 2。　　【原文 4 是错误】

$$c\neq 4\quad R=\begin{bmatrix}\mathbf{1}&2&0\\0&0&\mathbf{1}\\0&0&0\end{bmatrix}\qquad c=4\quad R=\begin{bmatrix}1&2&1\\0&0&0\\0&0&0\end{bmatrix}$$

两个主元留下一个自由变量 x_2。但是当 $c=4$，只有列 1 有一个主元(秩 1)，第二与第三变量是自由，得到两个特殊解。

$c\neq 4$　特殊解 $(-2,1,0)$　　$c=4$　另一个特殊解 $(-1,0,1)$

2×2 矩阵 $B=\begin{bmatrix}c&c\\c&c\end{bmatrix}$，除了 $c=0$ 之外它的秩 $r=1$；如果 $c=0$，它的秩是 0。

$$c\neq 0\quad R=\begin{bmatrix}1&1\\0&0\end{bmatrix}\qquad c=0\quad R=\begin{bmatrix}0&0\\0&0\end{bmatrix}\text{ 且零空间 } = \mathbf{R}^2$$

问题集 3.2

1　简化 A 与 B 至他们的三角阶梯形式 U。哪个变量是自由?

(a)　$A=\begin{bmatrix}1&2&2&4&6\\1&2&3&6&9\\0&0&1&2&3\end{bmatrix}$　　(b)　$B=\begin{bmatrix}2&4&2\\0&4&4\\0&8&8\end{bmatrix}$

2　对于问题 1 的矩阵，找出每个自由变量的特殊解。(令自由变量为 1，其他自由变量为 0)。

3　对于问题 1 的 U 执行更深入的行运算，求出简化阶梯形式 R。回答对或错，并给出理由：R 的零空间等于 U 的零空间。

4　对于~~相同的~~ A 与 B，找出 $A\boldsymbol{x}=\mathbf{0}$ 与 $B\boldsymbol{x}=\mathbf{0}$ 的特殊解。对于 $m\times n$ 矩阵而言，主元变量的个数加上自由变量的个数等于_____。这就是“**计数定理**”：$r+(n-r)=n$。　【原文多了“same”】

(a)　$A=\begin{bmatrix}-1&3&5\\-2&6&10\end{bmatrix}$　　(b)　$B=\begin{bmatrix}-1&3&5\\-2&6&7\end{bmatrix}$

问题 5-14 有关自由变量与主元变量。

5　是非题 (如果正确，给出理由；如果错误，给出反例)

(a) 方形矩阵没有自由变量。

(b) 可逆矩阵没有自由变量。

(c) $m\times n$ 矩阵最多有 n 个主元变量。

(d) $m\times n$ 矩阵最多有 m 个主元变量。

6　在 4×7 的阶梯矩阵 U 中放置最多可能的 1，使得 U 的主元列分别是

(a) 2, 4, 5　　(b) 1, 3, 6, 7　　(c) 4 与 6

7　在 4×8 的简化阶梯矩阵 R 中放置最多可能的 1，使得 R 的自由列分别是

(a) 2, 4, 5, 6　　(b) 1, 3, 6, 7, 8

8　假设 3×5 矩阵的列 4 全部是 0，则 x_4 必然是_____变量。这个变量的特殊解是向量 $\boldsymbol{x}=$ _____。

9　假设 3×5 矩阵的第一列与最后一列相同(不是 0)，则_____是自由变量。求出这个变量的特殊解。

10　假设 $m\times n$ 矩阵有 r 个主元，特殊解的个数是____。当 $r=$ ____时，零空间只包含 $\boldsymbol{x}=\boldsymbol{0}$。当 $r=$ ____时，列空间是全部的$\mathbf{R}^m$。

11　当 5×5 矩阵的主元个数是_____，则零空间只包含 $\boldsymbol{x}=\boldsymbol{0}$。当主元个数是_____，列空间是全部的$\mathbf{R}^5$。解释为什么?

12　方程式 $x-3y-z=0$ 决定了$\mathbf{R}^3$的一个平面，这个方程式的矩阵 A 为何? 哪些是自由变量? 特殊解是_____与_____。

13　(推荐) 平面 $x-3y-z=12$ 与 $x-3y-z=0$ 平行，第一个平面有一个特解是(12, 0, 0)，所有平面上的点具有下列形式：　　【填空题】

$$\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}+y\begin{bmatrix}1\\0\end{bmatrix}+z\begin{bmatrix}0\\1\end{bmatrix}$$

14　假设 4×5 矩阵有 4 个主元，且列 1 + 列 3 + 列 5 = $\boldsymbol{0}$，哪个列没有主元? 特殊解为何? 描述 $\boldsymbol{N}(A)$。

问题 15-22 在特定性质下寻找矩阵(如果有的话)。

15　创建一个矩阵使得 $\boldsymbol{N}(A)=(2, 2, 1, 0)$与$(3, 1, 0, 1)$的所有组合。

16　创建 A 使得 $\boldsymbol{N}(A)=(4, 3, 2, 1)$的所有组合，它的秩是____。

请尊重版权与译者的劳动成果，侵权必究!

17 创建一个矩阵使得列空间包含(1, 1, 5)与(0, 3, 1)，而且零空间包含(1, 1, 2)。

18 创建一个矩阵使得列空间包含(1, 1, 0)与(0, 1, 1)，而且零空间包含(1, 0, 1)与(0, 0, 1)。

19 创建一个矩阵使得列空间包含(1, 1, 1)，零空间是(1, 1, 1, 1)的倍数的直线。

20 创建一个 2×2 矩阵使得零空间等于列空间。这是可能的。

21 为什么没有一个 3×3 矩阵的零空间等于列空间?

22 若 $AB = 0$，则 B 的列空间被 A 的_____包含。为什么？

23 一个随机选取单元的 3×3 矩阵的简化形式 R，几乎可以确定是_____。若随机的 A 是 4×3，什么样的 R 是几乎确定的？

24 举例证明这三个叙述大概率是错的：

(a) A 与 A^{T} 有相同的零空间。

(b) A 与 A^{T} 有相同的自由变量。

(c) 若 R 是简化形式 rref(A)，则 R^{T} 就是 rref(A^{T})。

25 若 $\boldsymbol{N}(A) =$ 所有 $\boldsymbol{x} = (2, 1, 0, 1)$的倍数，$R$ 是什么？秩是多少?

26 若 $R\boldsymbol{x} = \boldsymbol{0}$ 的特殊解在下列零空间矩阵 N 的列里面，反向回去求出简化矩阵 R 的非零行。

$$N = \begin{bmatrix} 2 & 3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ 与 } N = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ 与 } N = \begin{bmatrix} \ \\ \ \\ \ \end{bmatrix} \text{ (空的 3×1)}$$

27 (a) 写出 5 个 2×2 简化矩阵 R，它的单元全部是 0 或 1。

(b) 写出 8 个 1×3 矩阵，它的单元全部是 0 或 1。这 8 个矩阵全部是简化阶梯形式矩阵 R 吗？

28 解释为什么 A 与 $-A$ 有相同的简化阶梯形式 R。

29 若 A 是 4×4 且可逆，描述 4×8 矩阵[A　A]的零空间。

30 若 $C = \begin{bmatrix} A \\ B \end{bmatrix}$，则零空间 $\boldsymbol{N}(C)$与空间 $\boldsymbol{N}(A)$与 $\boldsymbol{N}(B)$的关系是如何？

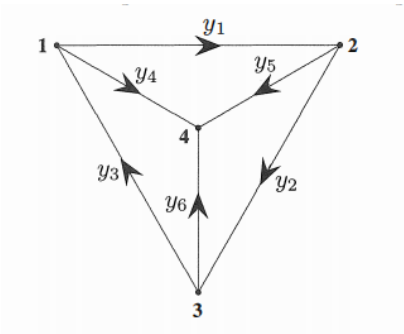
31 求出简化行阶梯形式 R 以及矩阵的秩:

(a) 3×4 矩阵，所有的单元都是 4。

(b) 3×4 矩阵，其中 $a_{ij} = i + j - 1$

(c) 3×4 矩阵，其中 $a_{ij} = (-1)^j$。

32 基尔霍夫电流定律 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 说明在每一个节点都有“流进的电流 = 流出的电流”。在点 1，这是 $y_3=y_1+y_4$。写出 4 个节点的 4 个基尔霍夫方程式(箭头的方向表示每个 y 的正方向)。简化 A^{T} 变成 R，然后在 A^{T}(4×6 矩阵)的零空间中找出三个特殊解



$$A=\begin{bmatrix}-1 & 1 & 0 & 0\\ 0 & -1 & 1 & 0\\ 1 & 0 & -1 & 0\\ -1 & 0 & 0 & 1\\ 0 & -1 & 0 & 1\\ 0 & 0 & -1 & 1\end{bmatrix}$$

33 下列哪些结果是 A 的秩的正确定义?

(a) R 的非零行的个数。

(b) 列的个数减去全部的行数。

(c) 列的个数减去自由列的个数。

(d) 矩阵 R 的 1's 的总数。

34 对每个(方块)矩阵求出简化 R：

$$A=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 3\\ 2 & 4 & 6\end{bmatrix}\qquad B=\begin{bmatrix}A & A\end{bmatrix}\qquad B=\begin{bmatrix}A & A\\ A & 0\end{bmatrix}$$

35 假设所有的主元变量不是一开始而是在最后才出现，描述简化阶梯形式中的所有 4 个方块(方块 B 必须是 $r\times r$)：

$$R=\begin{bmatrix}A & B\\ C & D\end{bmatrix}$$

36 (笨问题) 描述所有的 2×3 矩阵 A_1 与 A_2，他们的行阶梯形式是 R_1 与 R_2，使得 R_1+R_2 是 A_1+A_2 的行阶梯形式。在本案例中，$R_1=A_1$ 且 $R_2=A_2$ 正确吗？

R_1-R_2 等于 rref(A_1-A_2)吗？

37 若 A 有 r 个主元列，你如何知道 A^{T} 有 r 个主元列？举一个 3×3 例子，其中 A 与 A^{T} 的主元列码(pivcol，意思是所在列的编号)不相同。

38 $R\boldsymbol{x}=\boldsymbol{0}$ 以及 $\boldsymbol{y}^{\mathrm{T}}R=\boldsymbol{0}$ 的特殊解是什么？R 在下方：

$$R=\begin{bmatrix}1 & 0 & 2 & 3\\ 0 & 1 & 4 & 5\\ 0 & 0 & 0 & 0\end{bmatrix}\qquad R=\begin{bmatrix}0 & 1 & 2\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

39 空格填充使得他们的秩都是 1：

$$A=\begin{bmatrix}1&2&4\\2&&\\4&&\end{bmatrix}\quad B=\begin{bmatrix}&9&\\1&&\\2&6&-3\end{bmatrix}\quad M=\begin{bmatrix}a&b\\c&\end{bmatrix}$$

40 若 A 是 $m\times n$ 矩阵且 $r = 1$，它的列都是某一列的倍数，它的行都是某一行的倍数。列空间是$\mathbf{R}^m$的_____，零空间是$\mathbf{R}^n$的_____，零空间矩阵 N 的形状是_____。

41 选择向量 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 使得 $A=\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}=$ 列乘行：

$$A=\begin{bmatrix}3&6&6\\1&2&2\\4&8&8\end{bmatrix}\ \text{与}\ A=\begin{bmatrix}2&2&6&4\\-1&-1&-3&-2\end{bmatrix}$$

$A=\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 是每一个秩 $r=1$ 矩阵的自然形式。

42 若 A 是秩一矩阵，R 的第二行是_____。举个例子。

问题 43-45 有关在 A 中的 $r\times r$ 可逆矩阵。

43 若 A 有秩 r，则它有 $r\times r$ 的子矩阵 S，其中 S 可逆。移除 $m-r$ 个行及 $n-r$ 个列，求出分别在 A, B, C 中的可逆子矩阵 S。你可以保持主元行与主元列。

$$A=\begin{bmatrix}1&2&3\\1&2&4\end{bmatrix}\quad B=\begin{bmatrix}1&2&3\\2&4&6\end{bmatrix}\quad C=\begin{bmatrix}0&1&0\\0&0&0\\0&0&1\end{bmatrix}$$

44 假设 P 是 $m\times n$ 矩阵中只包含 r 个主元列的子矩阵，解释为什么这个 $m\times r$ 的子矩阵有秩 r。

45 对问题 44 的 P 取转置，求 P^{T}(这是 $r\times m$)的 r 个主元列。反转置回来，得到在 P 与 A 中的一个 $r\times r$ 可逆子矩阵 S：

当 $A=\begin{bmatrix}1&2&3\\2&4&6\\2&4&7\end{bmatrix}$，找出 P (3×2)以及可逆的 S (2×2)。

问题 46-51 展示 AB 的秩不大于 A 的秩，也不大于 B 的秩。

46 求 AB 与 AC 的秩(秩一矩阵乘秩一矩阵)：

$$A=\begin{bmatrix}1&2\\2&4\end{bmatrix}\quad B=\begin{bmatrix}2&1&4\\3&1.5&6\end{bmatrix}\quad C=\begin{bmatrix}1&b\\c&bc\end{bmatrix}$$

47 秩一矩阵 $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 乘秩一矩阵 $\boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$ 等于 $\boldsymbol{u}\boldsymbol{z}^{\mathrm{T}}$ 乘数字_____。除非_____ = 0，否则乘积 $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$ 也是秩一。

48 (a) 假设 B 的列 j 是前面列的组合，证明 AB 的列 j 也是 AB 前面列的相同组合。AB 不可能有新的主元，所以秩$(AB) \leq$ 秩(B)。

(b) 求 A_1 与 A_2 使得秩$(A_1B)=1$ 且秩$(A_2B)=0$，其中 $B=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$。

49 问题 48 证明了秩$(AB) \leq$ 秩(B)。同样的理由也得到秩$(B^{\mathrm{T}}A^{\mathrm{T}}) \leq$ 秩(A^{T})。你如何推论秩$(AB) \leq$ 秩(A)?

50 (重要) 假设 A 与 B 是 $n\times n$ 矩阵且 $AB=I$。由秩$(AB) \leq$ 秩(A)，证明 A 的秩是 n，所以 A 是可逆且 B 必须是它的双边逆矩阵(段落 2.5)，因此 $BA=I$。(这个不是显而易见的)

51 若 A 是 2×3，B 是 3×2，且 $AB=I$，由秩证明 $BA\neq I$。给出一个 A 与 B 的例子使得 $AB=I$。当 $m<n$，右侧逆矩阵不等于左侧逆矩阵。

52 假设 A 与 B 有相同的简化行阶梯形式 R。

(a) 证明 A 与 B 有相同的零空间以及相同的行空间。

(b) 我们知道 $E_1A=R$ 与 $E_2B=R$，所以 A 等于一个_____矩阵乘 B。

53 把 A 与 B 表示成两个秩一矩阵的总和 ：

$$\textbf{秩}=\mathbf{2} \qquad A=\begin{bmatrix}1 & 1 & 0\\ 1 & 1 & 4\\ 1 & 1 & 8\end{bmatrix} \qquad B=\begin{bmatrix}2 & 2\\ 2 & 3\end{bmatrix}$$

54 回答已解范例 **3.2C** 的相同问题：

$$A=\begin{bmatrix}1 & 1 & 2 & 2\\ 2 & 2 & 4 & 4\\ 1 & c & 2 & 2\end{bmatrix} \qquad B=\begin{bmatrix}1-c & 2\\ 0 & 2-c\end{bmatrix}$$

55 A，B，C 的零空间矩阵 N(包含特殊解)为何？

$$\text{方块矩阵} \qquad A=[I \quad I] \qquad B=\begin{bmatrix}I & I\\ 0 & 0\end{bmatrix} \qquad C=[I \quad I \quad I]$$

56 简洁的事实　每一个秩 r 的 $m\times n$ 矩阵可以简化成 $(m\times r)$ 乘 $(r\times n)$。

$$A=(A \text{ 的主元列})\,(R \text{ 的前 } r \text{ 行})=(\textbf{列})\,(\textbf{行})$$

写出一个全部单元都是 1 的 3×4 矩阵 A，且 A 等于(来自主元列的 3×1 矩阵)乘(来自 R 的 1×4 矩阵)的乘积。

请尊重版权与译者的劳动成果，侵权必究!

挑战问题

57 假设 A 是秩 r 的 $m\times n$ 矩阵，它的简化阶梯形式是 R。精确描述矩阵 Z (它的形状以及全部的单元)，其中 Z 来自简化阶梯形式矩阵的转置 R^{T}：

$$R=\mathrm{rref}(A)\qquad Z=(\mathrm{rref}(A^{\mathrm{T}}))^{\mathrm{T}}$$

58 (推荐) 假设 R 是秩 r 的 $m\times n$ 矩阵，其中主元列放在最前面：

$$R=\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

(a) 这 4 个方块的形状是什么？

(b) 若 $r=m$，求出右逆矩阵 B，其中 $RB=I$。零方块消失。

(c) 若 $r=n$，求出左逆矩阵 C，其中 $CR=I$。F 与 0 列消失。

(d) R^{T} 的简化阶梯形式为何？ (要有形状)

(e) $R^{\mathrm{T}}R$ 的简化阶梯形式为何？ (要有形状)

59 我想 $R^{\mathrm{T}}R$ 的简化阶梯形式永远是 R(除了额外的零行)，当 R 是 2×3，你可以举个例子吗？稍后会展示 $A^{\mathrm{T}}A$ 与 A 永远有相同的零空间 (重要的事实)。

60 假设你允许对 A 操作基础列运算以及基础行运算(都得到 R)，一个秩 r 的 $m\times n$ 矩阵，它的“行与列简化形式”为何？

消元法：大局观

本页在向量层次以及子空间层次说明消元法，其中 A 简化至 R。你知道这些步骤，我就不再重复。消元法从第一主元开始，一次移动一个列(左到右)，以及一次一个行(顶到底)。当它移动时，消元法回答了两个问题：

问题 1　这一列会是前面列的组合吗？

如果这一列具有主元，答案是 no，主元列与前面的列是无关的。若列 4 没有主元，它会是列 1, 2, 3 的组合。

问题 2　这一行会是前面行的组合吗？

如果这一行具有主元，答案是 no，主元行与前面的行是无关的。若行 3 最终没有主元，它是一个零行，会被移动到 R 的底部。

对我来说，通过矩阵的一个动作，可以同时回答这两个问题是惊奇的。事实上这个动作得到的是三角形阶梯矩阵 U，而不是简化阶梯矩阵 R。从 U 到 R 的简化是由底部往顶端前进，U 可以分辨哪些列是较早列的组合(丢失主元)，而 R 会告诉我们是什么样的组合。

换言之，R 告诉我们 $A\boldsymbol{x} = \boldsymbol{0}$ 的特殊解，我们可以借由不同的行交换以及消元步骤从 A 得到 R，但是得到的永远是同一个 R(因为特殊解是由 A 决定)。后面的术语会很快来到，R 显露了三个基础子空间的“基底(base)”：

A 的列空间 – 选择 A 的主元列作为基底。

A 的行空间 – 选择 R 的非零行作为基底。

A 的零空间 – 选择 $R\boldsymbol{x} = \boldsymbol{0}$(与 $A\boldsymbol{x} = \boldsymbol{0}$)的特殊解。

我们从消元法学到了单一个最重要的数值––秩(rank) r，这个数值就是主元列以及主元行的个数，则 $n - r$ 就是自由列与特殊解的个数。

我曾经提过简化$[A \quad I]$得到$[R \quad E]$会告诉你更多关于 A 的事情––事实上是几乎全部的事情(包含 $EA = R$)。矩阵 E 会保存从 A 到 R 的消元记录，否则会遗失。当 A 是方形而且可逆时，R 就是 I，E 就是 A^{-1}。

请尊重版权与译者的劳动成果，侵权必究！

3.3 $Ax = b$ 的完整解

1. $A\boldsymbol{x} = \boldsymbol{b}$ 的**完整解**：$\boldsymbol{x}$ = (一个特解 $\boldsymbol{x}_p$) + (零空间的任意 $\boldsymbol{x}_n$)。
2. $[A \quad \boldsymbol{b}]$通过消元法得到$[R \quad \boldsymbol{d}]$，则 $A\boldsymbol{x} = \boldsymbol{b}$ 等效于 $R\boldsymbol{x} = \boldsymbol{d}$。
3. $A\boldsymbol{x} = \boldsymbol{b}$ 与 $R\boldsymbol{x} = \boldsymbol{d}$ 有解的条件：R 的全零行在 $\boldsymbol{d}$ 也必须是 0。
4. 当 $R\boldsymbol{x} = \boldsymbol{d}$ 有解，存在一个自由变量全部为 0 的**特解**(particular solution)$\boldsymbol{x}_p$。
5. 当零空间 $\boldsymbol{N}(A)$ = 零向量时，A 具有**满列秩** $r = n$: 没有自由变量。
6. 当列空间 $\boldsymbol{C}(A) = \mathbf{R}^m$时，$A$ 具有**满行秩** $r = m$：$A\boldsymbol{x} = \boldsymbol{b}$ 永远有解。
7. 四种情形分别是 $r = m = n$ (A 可逆)， $r = m < n$ (每个 $A\boldsymbol{x} = \boldsymbol{b}$ 有解)，$r = n < m$ ($A\boldsymbol{x} = \boldsymbol{b}$ 有一个或零个解)，$r < m, r < n$ (0 或∞解)。

上个段落都在求解 $A\boldsymbol{x} = \mathbf{0}$，消元法把问题转换成 $R\boldsymbol{x} = \mathbf{0}$。自由变量给定特殊的值 (0 或 1)，再由反向代入求得主元变量。我们没有关注右侧的 $\boldsymbol{b}$，这是因为它一直停留在零，解 $\boldsymbol{x}$ 在 A 的零空间中。

现在 $\boldsymbol{b}$ 不是零，作用在左侧的行运算也要作用在右侧。$A\boldsymbol{x} = \boldsymbol{b}$ 简化成比较简单的系统 $R\boldsymbol{x} = \boldsymbol{d}$。一种方法是将 $\boldsymbol{b}$ 加入原始矩阵作为额外列，我会加入$(b_1, b_2, b_3) = (1, 6, 7)$去制造一个**增广矩阵**$[A \quad \boldsymbol{b}]$：

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ \mathbf{7} \end{bmatrix} \quad \text{得到增广矩阵} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 1 & 3 & 1 & 6 & \mathbf{7} \end{bmatrix} = \begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}$$

当我们对 A 执行正常的消元法得到 R 时，我们也对 $\boldsymbol{b}$ 执行同样的操作。

在这个范例中，我们从行 3 减去行 1，然后从行 3 减去行 2，这样在 R 得到一整行的 0，然后把 $\boldsymbol{b}$ 变成新的右侧 $\boldsymbol{d} = (1, 6, 0)$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ \mathbf{0} \end{bmatrix} \quad \text{得到增广矩阵} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$$

最后一个 0 非常重要，第三方程式变成 0 = 0，所以方程式有解。在原始的矩阵 A 中，第一行加第二行等于第三行，如果方程式是一致，方程式的右侧也要有这样的性质。右侧 $\boldsymbol{b}$ 最重要的性质就是 1 + 6 = 7，引导至 0 = 0。

下列是对于一般 $\boldsymbol{b}=(b_1, b_2, b_3)$的增广矩阵，其中 A 不变：

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix}=\begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 1 & 3 & 1 & 6 & b_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 0 & 2 & b_1 \\ 0 & 0 & 1 & 4 & b_2 \\ 0 & 0 & 0 & 0 & b_3-b_1-b_2 \end{bmatrix}=\begin{bmatrix} R & \boldsymbol{d} \end{bmatrix}$$

只有在 $b_3-b_1-b_2=0$ 时，第三方程式得到 $0=0$，就是 $b_1+b_2=b_3$。

一个特解 $A\boldsymbol{x}_p=\boldsymbol{b}$

要想得到一个简单的解 $\boldsymbol{x}_p$，我们选择自由变量为 0：$\boldsymbol{x}_2=\boldsymbol{x}_4=0$，则两个非零方程式得到两个主元变量 $\boldsymbol{x}_1=1$ 与 $\boldsymbol{x}_3=6$。对于 $A\boldsymbol{x}=\boldsymbol{b}$ (也是 $R\boldsymbol{x}=\boldsymbol{d}$)的一个特解就是 $\boldsymbol{x}_p=(1, 0, 6, 0)$。这个特解是我的最爱：自由变量= 0，主元变量来自 $\boldsymbol{d}$，这个方法永远有效。

若存在解，R 的零行在 $\boldsymbol{d}$ 也必须是 0。由于 I 在 R 的主元行与主元列，则 $\boldsymbol{x}_{特解}$ 的主元变量来自 $\boldsymbol{d}$：

$$R\boldsymbol{x}_p=\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix}=\begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \quad \begin{matrix} \text{主元变量} 1,\ 6 \\ \text{自由变量} 0,\ 0 \\ \text{解}\ \boldsymbol{x}_p=(1,\ 0,\ 6,\ 0) \end{matrix}$$

注意我们是如何选取自由变量(设成 0)与求解主元变量。行简化至 R 后，这些步骤就很快速了。当自由变量是 0，$\boldsymbol{x}_p$ 的主元变量已经可以在右侧的向量 $\boldsymbol{d}$ 中看见。

$\boldsymbol{x}_{特解}$	特解是求解	$A\boldsymbol{x}_p=\boldsymbol{b}$
$\boldsymbol{x}_{零空间}$	$n-r$ 个特殊解是求解	$A\boldsymbol{x}_n=\boldsymbol{0}$

特解是(1, 0, 6, 0)，$R\boldsymbol{x}=\boldsymbol{0}$ 的两个**特殊解**(零空间)来自 R 的两个自由列，借由反转 3, 2, 4 的符号得到解。**请注意我是如何把 $A\boldsymbol{x}=\boldsymbol{b}$ 的完整解写成 $\boldsymbol{x}_p+\boldsymbol{x}_n$：**

完整解 一个$\boldsymbol{x}_p$ 许多$\boldsymbol{x}_n$

$$\boldsymbol{x}=\boldsymbol{x}_p+\boldsymbol{x}_n=\begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix}+\boldsymbol{x}_2\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}+\boldsymbol{x}_4\begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

问题 假设 A 是方形可逆矩阵，$m=n=r$，$\boldsymbol{x}_p$ 与 $\boldsymbol{x}_n$ 为何？

答案 **特解**是 $\boldsymbol{x}_p=A^{-1}\boldsymbol{b}$ 的唯一解，不存在**特殊解**或是自由变量。$R=I$ 没有零行，零空间中只有一个向量 $\boldsymbol{x}_n=\boldsymbol{0}$。完整解就是 $\boldsymbol{x}_p+\boldsymbol{x}_n=A^{-1}\boldsymbol{b}+\boldsymbol{0}$。

请尊重版权与译者的劳动成果，侵权必究!

在第二章我们没有提到零空间，因为 A 是可逆，$\boldsymbol{N}(A)$只包含零向量。从$[A \quad \boldsymbol{b}]$到$[I \quad A^{-1}\boldsymbol{b}]$的简化，矩阵 A 利用各种方法简化到 I，$A\boldsymbol{x}=\boldsymbol{b}$ 就变成 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$ 就是 $\boldsymbol{d}$。这是一个特殊例子，但是方形矩阵是实际上最常看到的矩阵，所以他们在本书的开头拥有自己的章节。

对于小型的范例，我们进行$[A \quad \boldsymbol{b}]$到$[R \quad \boldsymbol{d}]$的简化，对于大型矩阵来说，MATLAB 做得更好。一个**特解**(不一定是我们想要的)就是 $\boldsymbol{x}=A\backslash\boldsymbol{b}$ 的反斜线而已。下列是满列秩的例子，两个列都有主元。

范例 1　求 b_1, b_2, b_3 的条件，使得 $A\boldsymbol{x}=\boldsymbol{b}$ 有解，若

$$A=\begin{bmatrix}1 & 1\\ 1 & 2\\ -2 & -3\end{bmatrix} \text{ 且 } \boldsymbol{b}=\begin{bmatrix}b_1\\ b_2\\ b_3\end{bmatrix}$$

条件就是 $\boldsymbol{b}$ 必须在 A 的列空间中，求完整的 $\boldsymbol{x}=\boldsymbol{x}_p+\boldsymbol{x}_n$。

解　加入额外的 $\boldsymbol{b}$ 得到增广矩阵，对$[A \quad \boldsymbol{b}]$执行从行 2 减去行 1，然后 2 倍的行 1 加到行 3，得到$[R \quad \boldsymbol{d}]$：

$$A=\begin{bmatrix}1 & 1 & b_1\\ 1 & 2 & b_2\\ -2 & -3 & b_3\end{bmatrix}\rightarrow\begin{bmatrix}1 & 1 & b_1\\ 0 & 1 & b_2-b_1\\ 0 & -1 & b_3+2b_1\end{bmatrix}\rightarrow\begin{bmatrix}1 & 0 & 2b_1-b_2\\ 0 & 1 & b_2-b_1\\ 0 & 0 & b_3+b_1+b_2\end{bmatrix}$$

如果 $b_1+b_2+b_3=0$，最后的方程式是 $0=0$，这就是 $\boldsymbol{b}$ 进入列空间的条件，则 $A\boldsymbol{x}=\boldsymbol{b}$ 有解。A 的三个行相加产生零行，这是为了一致性(他们是方程式！)，$\boldsymbol{b}$ 的分量相加也要是零。

因为 $n-r=2-2$，这个范例没有自由变量，所以没有**特殊解**，零空间的解是 $\boldsymbol{x}_n=\boldsymbol{0}$。$A\boldsymbol{x}=\boldsymbol{b}$ 与 $R\boldsymbol{x}=\boldsymbol{d}$ 的**特解**就是最后一列 $\boldsymbol{d}$ 的顶端：

$$A\boldsymbol{x}=\boldsymbol{b} \text{ 的唯一解} \qquad \boldsymbol{x}=\boldsymbol{x}_p+\boldsymbol{x}_n=\begin{bmatrix}2b_1-b_2\\ b_2-b_1\end{bmatrix}+\begin{bmatrix}0\\ 0\end{bmatrix}$$

如果 $b_1+b_2+b_3\neq 0$，则 $A\boldsymbol{x}=\boldsymbol{b}$ 无解($\boldsymbol{x}_p$ 与 $\boldsymbol{x}$ 不存在)。

这个例子是典型的重要案例：A 具有满列秩，每个列都有一个主元，秩 $r=n$，矩阵是高且瘦($m\geq n$)。当 A 简化至 R 且秩 $=n$ 时，行简化会把 I 放在最顶端：

$$R=\begin{bmatrix}I\\ 0\end{bmatrix}=\begin{bmatrix}n\times n\text{的单位矩阵}\\ m-n\text{个零行}\end{bmatrix} \tag{1}$$

此时没有自由列或是自由变量，零空间是 $\boldsymbol{Z}=\{\text{零向量}\}$。

我们会收集不同的方法来辨识矩阵的形式。

具有 **满列秩 $(r = n)$** 的矩阵 A 都有下列性质：

1. 所有的列都是主元列。
2. 没有自由变量或是特殊解。
3. 零空间 $N(A)$只包含零向量 $\boldsymbol{x} = \boldsymbol{0}$。
4. 若 $A\boldsymbol{x} = \boldsymbol{b}$ 有一个解(可能没有)，则只能有一个唯一解。

下个段落的重要语言：这个 A 有**无关列**，只有当 $\boldsymbol{x} = \boldsymbol{0}$ 才有 $A\boldsymbol{x} = \boldsymbol{0}$。在第四章我们会对上表再加入一个事实：若方形矩阵 $A^{\mathrm{T}}A$ 的秩等于 n，则为可逆矩阵。

在这个案例中，A(或 R)的零空间萎缩至零向量，$A\boldsymbol{x} = \boldsymbol{b}$ 的解是唯一解(如果存在的话)，在 R 会有 $m - n$ 个零行。为了在这些行得到 $0 = 0$，会存在 $m - n$ 个条件加在 $\boldsymbol{b}$ 上，此时 $\boldsymbol{b}$ 在列空间中。

满列秩 $r = n$ 时，$A\boldsymbol{x} = \boldsymbol{b}$ 有一个解或是无解 $(m > n$ 是过度决定)。

完整解

另一个极端例子是满行秩，现在 $A\boldsymbol{x} = \boldsymbol{b}$ 有一个解或是无限多解，这种情形下 A 必须是矮宽形$(m \le n)$。**若一个矩阵的 $\boldsymbol{r = m}$，则矩阵有满行秩。**“这些行是无关”，每个行有一个主元，以下是一个例子：

范例 2　系统 $A\boldsymbol{x} = \boldsymbol{b}$ 有 3 个未知数但是只有 2 个方程式：

$$\textbf{满行秩}\qquad \begin{matrix} x + y + z = 3 \\ x + 2y - z = 4 \end{matrix} \qquad (\text{秩 } r = m = 2)$$

这些是 xyz 空间的两个平面，两个平面不平行，所以交会于一条直线，这条直线就是解，经过消元法之后可以看到。**特解**会是直线上的一个点，再加上零空间向量 $\boldsymbol{x}_n$，使得点在图 3.3 的直线上移动，则 $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$ 得到整条线的解。

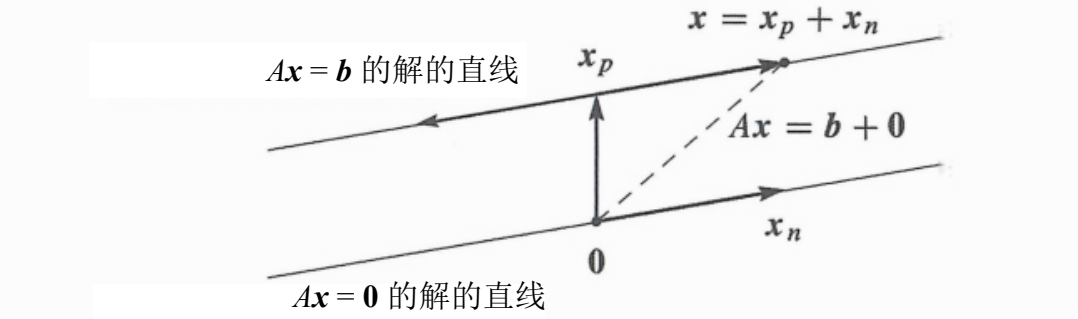


图 3.3：**完整解 = 一个特解 + 所有的零空间解**

请尊重版权与译者的劳动成果，侵权必究!

我们对[A　$\boldsymbol{b}$]执行消元法可以得到 $\boldsymbol{x}_p$ 与 $\boldsymbol{x}_n$，从行 2 减去行 1，然后再从行 1 减去行 2：

$$\begin{bmatrix}1 & 1 & 1 & 3\\ 1 & 2 & -1 & 4\end{bmatrix}\rightarrow\begin{bmatrix}1 & 1 & 1 & 3\\ 0 & 1 & -2 & 1\end{bmatrix}\rightarrow\begin{bmatrix}1 & 0 & 3 & 2\\ 0 & 1 & -2 & 1\end{bmatrix}=\begin{bmatrix}R & \boldsymbol{d}\end{bmatrix}$$

特解有自由变量 $x_3=0$，**特殊解**有 $x_3=1$：

$\boldsymbol{x}_{\text{特解}}$ 直接来自右侧的 $\boldsymbol{d}$：$\boldsymbol{x}_p=(2,1,0)$

$\boldsymbol{x}_{\text{特殊}}$ 来自 R 的第三列(自由列)：$\boldsymbol{s}=(-3,2,1)$

检验 $\boldsymbol{x}_p$ 与 $\boldsymbol{s}$ 是否满足原始方程式 $A\boldsymbol{x}_p=\boldsymbol{b}$ 以及 $A\boldsymbol{s}=\boldsymbol{0}$ 是聪明的：

$$2+1=3 \qquad -3+2+1=0$$
$$2+2=4 \qquad -3+4-1=0$$

零空间解 $\boldsymbol{x}_n$ 是 $\boldsymbol{s}$ 的任意倍数，沿着解的直线移动，起点就是 $\boldsymbol{x}_{\text{特解}}$。请再次注意如何写出答案：

完整解

$$\boldsymbol{x}=\boldsymbol{x}_p+\boldsymbol{x}_n=\begin{bmatrix}2\\1\\0\end{bmatrix}+x_3\begin{bmatrix}-3\\2\\1\end{bmatrix}$$

解的直线画在图 3.3 中，直线的任意点都可以选成**特定点**，我们选的点是 $x_3=0$。

特解不能乘任意常数，特殊解才需要常数，现在你知道为什么要得到所有零空间的 $\boldsymbol{x}_n$。

我们现在总结满行秩的矮宽矩阵，若 $m<n$，方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 是**欠决定**(很多解)。

具有 **满行秩 ($r=m$)** 的矩阵 A 有下列性质：

1. 所有的行都有主元，R 没有零行。
2. 对于所有的右侧 $\boldsymbol{b}$，$A\boldsymbol{x}=\boldsymbol{b}$ 一定有解。
3. 列空间是整个$\mathbf{R}^m$空间。
4. 在 A 的零空间中有 $n-r=n-m$ 个**特殊解**。

这个案例有 m 个主元，这些行是**线性无关**，所以 A^{T} 的列是线性无关，A^{T} 的零空间是零向量。

在总结 4 个可能性之后——与秩有关，我们已经准备好介绍**线性无关**的定义。请注意 m, r, n 如何成为重要数字：

线性方程式与秩 r 相关的 4 种可能性

$r = m$ 且 $r = n$	方形可逆	$A\boldsymbol{x} = \boldsymbol{b}$ 有 1 个解
$r = m$ 且 $r < n$	矮且宽	$A\boldsymbol{x} = \boldsymbol{b}$ 有无限多解
$r < m$ 且 $r = n$	高且瘦	$A\boldsymbol{x} = \boldsymbol{b}$ 有 0 个或 1 个解
$r < m$ 且 $r < n$	非满秩	$A\boldsymbol{x} = \boldsymbol{b}$ 有 0 个或无限多解

简化后的 R 与矩阵 A 是相同的种类，如果主元列碰巧在第一个出现，我们可以展示 R 的 4 个可能性。$R\boldsymbol{x} = \boldsymbol{d}$(以及原始的 $A\boldsymbol{x} = \boldsymbol{b}$)要有解，$\boldsymbol{d}$ 必须在尾端有 $m - r$ 个零。F 表示 R 的自由部分。

R的4种形式	$\begin{bmatrix} I \end{bmatrix}$	$\begin{bmatrix} I & F \end{bmatrix}$	$\begin{bmatrix} I \\ 0 \end{bmatrix}$	$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$
他们的秩	$r = m = n$	$r = m < n$	$r = n < m$	$r < m,\ r < n$

案例 1 与 2 有满行秩 $r = m$，案例 1 与 3 有满列秩 $r = n$，案例 4 在理论上是最常见的，但是在实际上却是少见。

主要观念的复习

1. 秩 r 是主元的数目，矩阵 R 有 $m - r$ 个零行。
2. $A\boldsymbol{x} = \boldsymbol{b}$ 有解，当且仅当最后 $m - r$ 个方程式可以简化成 $0 = 0$。
3. 一个**特解** $\boldsymbol{x}_p$ 的所有自由变量都是 0。
4. 主元变量是在选定自由变量后才决定。
5. 满列秩 $r = n$ 表示没有自由变量：一个或零个解。
6. 满行秩 $r = m$ 下，若 $m = n$ 表示有一个解；若 $m < n$ 表示有无限多解。

已解范例

3.3A　下面的问题把消元法(**主元列与反向代入**)与**列空间–零空间–秩–可解性**(较高层次的图) 产生关联。A 的秩是 2：

$$A\boldsymbol{x} = \boldsymbol{b} \quad \text{是} \quad \begin{aligned} x_1 + 2x_2 + 3x_3 + 5x_4 &= b_1 \\ 2x_1 + 4x_2 + 8x_3 + 12x_4 &= b_2 \\ 3x_1 + 6x_2 + 7x_3 + 13x_4 &= b_3 \end{aligned}$$

请尊重版权与译者的劳动成果，侵权必究!

1. 简化$[A \quad \boldsymbol{b}]$得到$[U \quad \boldsymbol{c}]$，使得 $A\boldsymbol{x}=\boldsymbol{b}$ 变成三角系统 $U\boldsymbol{x}=\boldsymbol{c}$。
2. 找出 b_1, b_2, b_3 的条件使得 $A\boldsymbol{x}=\boldsymbol{b}$ 有解。
3. 描述 A 的列空间，是$\mathbf{R}^3$的哪个平面?
4. 描述 A 的零空间，哪些是在$\mathbf{R}^4$的**特殊解**?
5. 简化$[U \quad \boldsymbol{c}]$得到$[R \quad \boldsymbol{d}]$：写出来自 R 的**特殊解**，来自 $\boldsymbol{d}$ 的**特解**。
6. 求出 $A\boldsymbol{x}=(0, 6, -6)$的一个**特解**，然后写出完整解。

解

1. 消元法的乘数是 2, 3, –1，会将$[A \quad \boldsymbol{b}]$变成$[U \quad \boldsymbol{c}]$。

$$\begin{bmatrix}1&2&3&5&b_1\\2&4&8&12&b_2\\3&6&7&13&b_3\end{bmatrix}\rightarrow\left[\begin{array}{cccc|c}1&2&3&5&b_1\\0&0&2&2&b_2-2b_1\\0&0&-2&-2&b_3-3b_1\end{array}\right]\rightarrow\left[\begin{array}{cccc|c}1&2&3&5&b_1\\0&0&2&2&b_2-2b_1\\0&0&0&0&b_3+b_2-5b_1\end{array}\right]$$

2. 最后的方程式显示有解的条件是 $b_3+b_2-5b_1=0$，则 $0=0$。
3. **第一描述：**列空间是包含主元列(1, 2, 3)与(3, 8, 7)所有的线性组合的平面，主元在列 1 与 3。**第二描述：**列空间包含所有 $b_3+b_2-5b_1=0$ 的向量，使得 $A\boldsymbol{x}=\boldsymbol{b}$ 有解，所以 $\boldsymbol{b}$ 在列空间中，这是第一描述中的平面方程式。
4. **特殊解**有自由变量 $x_2=1, x_4=0$ 然后是 $x_2=0, x_4=1$：

$A\boldsymbol{x}=0$的特殊解
在$U\boldsymbol{x}=0$反向代入
或改变 R 中2, 2, 1 的符号

$$\boldsymbol{s}_1=\begin{bmatrix}-2\\1\\0\\0\end{bmatrix}\qquad\boldsymbol{s}_2=\begin{bmatrix}-2\\0\\-1\\1\end{bmatrix}$$

在$\mathbf{R}^4$的零空间 $\boldsymbol{N}(A)$包含所有 $\boldsymbol{x}_n=c_1\boldsymbol{s}_1+c_2\boldsymbol{s}_2$。

5. U 的第三列是(3, 2, 0)，简化后的 R 的第三列变成(0, 1, 0)，右侧的 $\boldsymbol{c}=(0, 6, 0)$变成 $\boldsymbol{d}=(-9, 3, 0)$，说明了–9 与 3 在 $\boldsymbol{x}_p$ 中：

$$\begin{bmatrix}U & \boldsymbol{c}\end{bmatrix}=\begin{bmatrix}1&2&3&5&\mathbf{0}\\0&0&2&2&\mathbf{6}\\0&0&0&0&\mathbf{0}\end{bmatrix}\rightarrow\begin{bmatrix}R & \boldsymbol{d}\end{bmatrix}=\begin{bmatrix}1&2&0&2&\mathbf{-9}\\0&0&1&1&\mathbf{3}\\0&0&0&0&\mathbf{0}\end{bmatrix}$$

6. 一个特解 $\boldsymbol{x}_p$ 的自由变量都是 0，在 $U\boldsymbol{x}=\boldsymbol{c}$ 中反向代入：

$A\boldsymbol{x}_p=b$的特解
从向量d 得到–9与3
自由变量x_2 与x_4 为0

$$x_p=\begin{bmatrix}-9\\0\\3\\0\end{bmatrix}$$

$A\boldsymbol{x}=(0, 6, -6)$的完整解就是 $\boldsymbol{x}=\boldsymbol{x}_p+\boldsymbol{x}_n=\boldsymbol{x}_p+c_1\boldsymbol{s}_1+c_2\boldsymbol{s}_2$。

3.3B　假设你已经有了 $A\boldsymbol{x} = \boldsymbol{b}$ 的解的信息，其中 $\boldsymbol{b}$ 是个特定值。下列的描述会告诉你有关 m, n, r(以及 A 本身)的什么信息？以及可能关于 $\boldsymbol{b}$ 的信息?

1. 恰好有一个解。
2. $A\boldsymbol{x} = \boldsymbol{b}$ 所有的解具有 $\boldsymbol{x} = \begin{bmatrix}2\\1\end{bmatrix} + c\begin{bmatrix}1\\1\end{bmatrix}$ 的形式。
3. 无解。
4. $A\boldsymbol{x} = \boldsymbol{b}$ 所有的解具有 $\boldsymbol{x} = \begin{bmatrix}1\\1\\0\end{bmatrix} + c\begin{bmatrix}1\\0\\1\end{bmatrix}$ 的形式。
5. 有无限多解。

解　案例 1 恰好有一个解，A 必须有满列秩 $r = n$，A 的零空间只包含零向量。当然必须 $m \geq n$。

案例 2 中，A 必须有 $n = 2$ 个列(m 任意)，其中(1, 1)在 A 的零空间中，列 2 是列 1 的负号，此外 $A \neq 0$：秩等于 1。$\boldsymbol{x} = (2, 1)$是一个解，$\boldsymbol{b} = 2$(列 1) + (列 2)，我对于 $\boldsymbol{x}_p$ 的选择会是(1, 0)。

案例 3 中，我们只知道 $\boldsymbol{b}$ 不在 A 的列空间中，A 的秩必须小于 m。我猜我们知道 $\boldsymbol{b} \neq \boldsymbol{0}$，否则 $\boldsymbol{x} = \boldsymbol{0}$ 就会是一个解。

案例 4 中，A 必须有 $n = 3$ 个列，(1, 0, 1)在 A 的零空间中，列 3 就是列 1 的负号，列 2 不可以是列 1 的倍数，否则零空间会包含另外一个**特殊解**，因此 A 的秩就是 3 – 1 = 2。A 必须有 $n \geq 2$ 个行，右侧 $\boldsymbol{b}$ 是列 1+列 2。

案例 5 有无限多解，零空间必须包含非零向量，秩 r 必须小于 n (不是满列秩)，且 $\boldsymbol{b}$ 必须在 A 的列空间中。我们不知道是否每个 $\boldsymbol{b}$ 都在列空间中，所以我们不知道是否 $r = m$。

3.3B　针对$[A \quad \boldsymbol{b}]$执行前向消元法，求出完整解 $\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$：

$$\begin{bmatrix}1&2&1&0\\2&4&4&8\\4&8&6&8\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}4\\2\\10\end{bmatrix}$$

求出数值 y_1, y_2, y_3 使得 y_1(行 1) + y_2 (行 2) + y_3(行 3) = 零行。验证 $\boldsymbol{b} = (4, 2, 10)$满足条件 $y_1b_1 + y_2b_2 + y_3b_3 = 0$，为什么这个是方程式有解以及 $\boldsymbol{b}$ 在列空间的条件?

解　对于$[A \quad \boldsymbol{b}]$执行前向消元法，在$[U \quad \boldsymbol{c}]$中会得到一个零行，第三个方程式变成 0 = 0，方程式具有一致性(有解)：

$$\begin{bmatrix}1&2&1&0&\boldsymbol{4}\\2&4&4&8&\boldsymbol{2}\\4&8&6&8&\boldsymbol{10}\end{bmatrix} \to \begin{bmatrix}1&2&1&0&\boldsymbol{4}\\0&0&2&8&\boldsymbol{-6}\\0&0&2&8&\boldsymbol{-6}\end{bmatrix} \to \begin{bmatrix}1&2&1&0&\boldsymbol{4}\\0&0&2&8&\boldsymbol{-6}\\0&0&0&0&\boldsymbol{0}\end{bmatrix}$$

列 1 与 3 包含主元，x_2 与 x_4 是自由变量。如果把自由变量设成 0，我们可以利用反向代入求出**特解**，或者我们继续操作得到 R。

请尊重版权与译者的劳动成果，侵权必究!

令自由变量= 0，$R\boldsymbol{x} = \boldsymbol{d}$ 显示**特解**是 $\boldsymbol{x}_p = (7, 0, -3, 0)$:

$$\begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 2 & 8 & \mathbf{-6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 1 & 0 & \mathbf{4} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 0 & -4 & \mathbf{7} \\ 0 & 0 & 1 & 4 & \mathbf{-3} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}$$

对于 $\boldsymbol{b} = \mathbf{0}$ 的零空间部分 $\boldsymbol{x}_n$ 来说，设定自由变量 $\boldsymbol{x}_2$ 与 $\boldsymbol{x}_4$ 分别是 1, 0 与 0, 1:

特殊解　　$\boldsymbol{s}_1 = (-2, 1, 0, 0)$与 $\boldsymbol{s}_2 = (4, 0, -4, 1)$

则 $A\boldsymbol{x} = \boldsymbol{b}$ (以及 $R\boldsymbol{x} = \boldsymbol{d}$) 的完整解就是 $\boldsymbol{x}_{\text{完整}} = \boldsymbol{x}_p + c_1\boldsymbol{s}_1 + c_2\boldsymbol{s}_2$。

对 A 来说，2(行 1) + (行 2) – (行 3) = (0, 0, 0, 0) 得到 A 的零行，于是 $\boldsymbol{y} = (2, 1, -1)$。对 $\boldsymbol{b} = (4, 2, 10)$而言，就是 2(4) + 1(2) – 1(10) = 0。

如果行的组合(左侧)得到零行，则同样的组合也必须在右侧得到 0。当然，如果没有得到 0 就是无解。

稍后我们会再重复用不同的语句来说明: 若 A 的每一列都与 $\boldsymbol{y} = (2, 1, -1)$垂直，这些列的任意组合 $\boldsymbol{b}$ 也必须与 $\boldsymbol{y}$ 垂直，否则 $\boldsymbol{b}$ 就不在列空间中，$A\boldsymbol{x} = \boldsymbol{b}$ 就会无解。

再次重复：若 $\boldsymbol{y}$ 在 A^{T} 的零空间中，则 $\boldsymbol{y}$ 必须垂直于所有在 A 的列空间中的 $\boldsymbol{b}$。再往前看看....

问题集 3.3

1　(推荐) 执行已解范例 **3.3A** 的 6 个步骤，描述 A 的列空间与零空间，以及 $A\boldsymbol{x} = \boldsymbol{b}$ 的完整解：

$$A = \begin{bmatrix} 2 & 4 & 6 & 4 \\ 2 & 5 & 7 & 6 \\ 2 & 3 & 5 & 2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}$$

2　对秩一的矩阵 A 执行相同的 6 个步骤，你会发现 b_1, b_2, b_3 的两个条件使得 $A\boldsymbol{x} = \boldsymbol{b}$ 有解。这两个条件会把 $\boldsymbol{b}$ 带入_____空间(两个平面得到一条线)：

$$A = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 \\ 6 & 3 & 9 \\ 4 & 2 & 6 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 30 \\ 20 \end{bmatrix}$$

问题 3-15 有关 $A\boldsymbol{x} = \boldsymbol{b}$ 的解。依据书上的步骤得到 $\boldsymbol{x}_p$ 与 $\boldsymbol{x}_n$，从引入 $\boldsymbol{b}$ 作为最后一列的增广矩阵开始。

3　写出 $\boldsymbol{x}_p$ 加上零空间中 $\boldsymbol{s}$ 的任意倍数得到完整解：

$$\begin{aligned} x + 3y + 3z &= 1 \\ 2x + 6y + 9z &= 5 \\ -x - 3y + 3z &= 5 \end{aligned}$$

4　求出完整解(也称为通解)：

$$A=\begin{bmatrix}1&3&1&2\\2&6&4&8\\0&0&2&4\end{bmatrix}\begin{bmatrix}x\\y\\z\\t\end{bmatrix}=\begin{bmatrix}1\\3\\1\end{bmatrix}$$

5　在 b_1, b_2, b_3 的什么条件下系统有解？引进 $\boldsymbol{b}$ 作为第四列执行消元法，当满足条件时，求出所有的解。

$$\begin{aligned}x+2y-2z&=b_1\\2x+5y-4z&=b_2\\4x+9y-8z&=b_3\end{aligned}$$

6　在 b_1, b_2, b_3, b_4 的什么条件下系统有解？当满足条件时，求出所有的解 $\boldsymbol{x}$。

$$\begin{bmatrix}1&2\\2&4\\2&5\\3&9\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}\qquad\begin{bmatrix}1&2&3\\2&4&6\\2&5&7\\3&9&12\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}$$

7　利用消元法证明：若 $b_3-2b_2+4b_1=0$，则(b_1, b_2, b_3)在列空间中。

$$A=\begin{bmatrix}1&3&1\\3&8&2\\2&4&0\end{bmatrix}$$

A 的行的什么组合得到零行？

8　哪些向量(b_1, b_2, b_3)在 A 的列空间中？A 的行的什么组合得到零行？

$$\text{(a)}\quad A=\begin{bmatrix}1&2&1\\2&6&3\\0&2&5\end{bmatrix}\qquad\text{(b)}\quad A=\begin{bmatrix}1&1&1\\1&2&4\\2&4&8\end{bmatrix}$$

9　(a) 已解范例 **3.3A** 中从$[A\quad \boldsymbol{b}]$变成$[U\quad \boldsymbol{c}]$，把乘数放进 L 并验证 LU 等于 A，以及 $L\boldsymbol{c}=\boldsymbol{b}$。

(b) 组合 A 的主元列且数字–9 与 3 在特解 $\boldsymbol{x}_p$ 中，这是怎么样的线性组合？为什么？

10　建立一个 2×3 系统 $A\boldsymbol{x}=\boldsymbol{b}$ 且特解 $\boldsymbol{x}_p=(2, 4, 0)$与齐性(homogeneous)解 $\boldsymbol{x}_n=(1, 1, 1)$的任意倍数。

11　为什么一个 1×3 系统不能有特解 $\boldsymbol{x}_p=(2, 4, 0)$以及齐性解 $\boldsymbol{x}_n=(1, 1, 1)$的任意倍数。

请尊重版权与译者的劳动成果，侵权必究!

12 (a) 若 $A\boldsymbol{x}=\boldsymbol{b}$ 有两个解 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$，求 $A\boldsymbol{x}=\boldsymbol{0}$ 的两个解。

(b) 再求另一个 $A\boldsymbol{x}=\boldsymbol{0}$ 的解，以及另一个 $A\boldsymbol{x}=\boldsymbol{b}$ 的解。

13 解释为什么这些是错的：

(a) 完整解是 $\boldsymbol{x}_p$ 与 $\boldsymbol{x}_n$ 的任意线性组合。

(b) 系统 $A\boldsymbol{x}=\boldsymbol{b}$ 最多有一个特解。

(c) 解 $\boldsymbol{x}_p$ 设定全部自由变量为 0，是最短的解(极小长度$\|\boldsymbol{x}\|$)。求一个 2×2 的反例。

(d) 若 A 是可逆，则零空间中不存在 $\boldsymbol{x}_n$ 的解。

14 假设 U 的列 5 没有主元，则 x_5 是_____变量。零向量(是) (或不是) $A\boldsymbol{x}=\boldsymbol{0}$ 的唯一解。若 $A\boldsymbol{x}=\boldsymbol{b}$ 有一个解，则它是_____解。

15 假设 U 的行 3 没有主元，则此行是_____。方程式 $U\boldsymbol{x}=\boldsymbol{c}$ 只有在_____时，才有解。方程式 $A\boldsymbol{x}=\boldsymbol{b}$ (有解) (无解) (可能无解)。

问题 16-20 有关“满秩”矩阵 $r=m$ 或 $r=n$。

16 一个 3×5 矩阵最大可能的秩是_____？在 U 与 R 的_____存在主元。$A\boldsymbol{x}=\boldsymbol{b}$ 的解(永远存在) (唯一解)。A 的零空间是_____。一个例子是 $A=$_____。

17 一个 6×4 矩阵最大可能的秩是_____？在 U 与 R 的每一个_____存在一个主元。$A\boldsymbol{x}=\boldsymbol{b}$ 的解(永远存在) (唯一解)。A 的零空间是_____。一个例子是 $A=$_____。

18 利用消元法求 A 的秩以及 A^{T} 的秩：

$$A=\begin{bmatrix}1&4&0\\2&11&5\\-1&2&10\end{bmatrix}\quad 与\quad A=\begin{bmatrix}1&0&1\\1&1&2\\1&1&q\end{bmatrix}\quad 秩与\ q\ 有关$$

19 求 A 的秩，$A^{\mathrm{T}}A$ 的秩以及 AA^{T} 的秩：

$$A=\begin{bmatrix}1&1&5\\1&0&1\end{bmatrix}\quad 与\quad A=\begin{bmatrix}2&0\\1&1\\1&2\end{bmatrix}$$

20 简化 A 至阶梯形式 U，求三角 L 使得 $A=LU$。

$$A=\begin{bmatrix}3&4&1&0\\6&5&1&2\end{bmatrix}\quad 与\quad A=\begin{bmatrix}1&0&1&0\\2&2&0&3\\0&6&5&4\end{bmatrix}$$

21 对于下列满秩系统，求 $\boldsymbol{x}_p + \boldsymbol{x}_n$ 的完整解形式：

(a) $x + y + z = 4$　　(b) $\begin{matrix} x + y + z = 4 \\ x - y + z = 4 \end{matrix}$

22 若 $A\boldsymbol{x} = \boldsymbol{b}$ 有无限多解，为什么 $A\boldsymbol{x} = \boldsymbol{B}$(新的右侧)不可能只有一个解? $A\boldsymbol{x} = \boldsymbol{B}$ 可能是无解吗?

23 选择数字 q 使得(可能的话)秩是(a) 1, (b) 2, (c) 3:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ -3 & -2 & -1 \\ 9 & 6 & q \end{bmatrix} \quad 与 \quad B = \begin{bmatrix} 3 & 1 & 3 \\ q & 2 & q \end{bmatrix}$$

24 给出矩阵 A 的例子，使得 $A\boldsymbol{x} = \boldsymbol{b}$ 的解的个数为：

(a) 0 或 1，与 $\boldsymbol{b}$ 有关。

(b) ∞，与 $\boldsymbol{b}$ 无关。

(c) 0 或∞，与 $\boldsymbol{b}$ 有关。

(d) 1，与 $\boldsymbol{b}$ 无关。

25 若 $A\boldsymbol{x} = \boldsymbol{b}$ 有下列的解的情形，写下所有在 r 与 m 与 n 之间已知关系：

(a) 对某些 $\boldsymbol{b}$，无解。

(b) 对每个 $\boldsymbol{b}$，无限多解。

(c) 对某些 $\boldsymbol{b}$，恰好唯一解；对其他 $\boldsymbol{b}$ 无解。

(d) 对每个 $\boldsymbol{b}$，恰好唯一解。

问题 26-33 有关高斯-乔丹消元法(向上也向下)以及简化阶梯矩阵 $\boldsymbol{R}$。

26 从 U 到 R 继续使用消元法，每个行除以自身的主元使得新主元都是 1，然后产生主元上方的 0 得到 R：

$$U = \begin{bmatrix} 2 & 4 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{bmatrix} \quad 与 \quad U = \begin{bmatrix} 2 & 4 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$$

27 若 A 是三角矩阵，什么时候 $R = \text{rref}(A) = I$ ？

28 应用高斯-乔丹消元法至 $U\boldsymbol{x} = \boldsymbol{0}$ 以及 $U\boldsymbol{x} = \boldsymbol{c}$，得到 $R\boldsymbol{x} = \boldsymbol{0}$ 与 $R\boldsymbol{x} = \boldsymbol{d}$:

$$\begin{bmatrix} U & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 3 & \boldsymbol{0} \\ 0 & 0 & 4 & \boldsymbol{0} \end{bmatrix} \quad 与 \quad \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & \boldsymbol{5} \\ 0 & 0 & 4 & \boldsymbol{8} \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

29 应用高斯-乔丹消元法简化至 $R\boldsymbol{x}=\boldsymbol{0}$ 以及 $R\boldsymbol{x}=\boldsymbol{d}$:

$$\begin{bmatrix}U & 0\end{bmatrix}=\begin{bmatrix}3&0&6&\mathbf{0}\\0&0&2&\mathbf{0}\\0&0&0&\mathbf{0}\end{bmatrix} \quad 与 \quad \begin{bmatrix}U & \boldsymbol{c}\end{bmatrix}=\begin{bmatrix}3&0&6&\mathbf{9}\\0&0&2&\mathbf{4}\\0&0&0&\mathbf{5}\end{bmatrix}$$

求解 $U\boldsymbol{x}=\boldsymbol{0}$ 或 $R\boldsymbol{x}=\boldsymbol{0}$ 得到 $\boldsymbol{x}_n$(自由变量 $=1$)。$R\boldsymbol{x}=\boldsymbol{d}$ 的解为何?

30 简化至 $U\boldsymbol{x}=\boldsymbol{c}$ (高斯消元法),然后至 $R\boldsymbol{x}=\boldsymbol{d}$ (高斯-乔丹):

$$A\boldsymbol{x}=\begin{bmatrix}1&0&2&3\\1&3&2&0\\2&0&4&9\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}2\\5\\10\end{bmatrix}=\boldsymbol{b}$$

求出特解 $\boldsymbol{x}_p$ 以及所有的齐性解 $\boldsymbol{x}_n$。

31 给定下列的性质,求出矩阵 A 与 B。如果无法求解,给出理由:

(a) $A\boldsymbol{x}=\begin{bmatrix}1\\2\\3\end{bmatrix}$ 的唯一解是 $\boldsymbol{x}=\begin{bmatrix}0\\1\end{bmatrix}$。

(b) $B\boldsymbol{x}=\begin{bmatrix}0\\1\end{bmatrix}$ 的唯一解是 $\boldsymbol{x}=\begin{bmatrix}1\\2\\3\end{bmatrix}$。

32 求 A 的 LU 分解,以及 $A\boldsymbol{x}=\boldsymbol{b}$ 的完整解。

$$A=\begin{bmatrix}1&3&1\\1&2&3\\2&4&6\\1&1&5\end{bmatrix} \quad 与 \quad \boldsymbol{b}=\begin{bmatrix}1\\3\\6\\5\end{bmatrix} \quad 然后 \quad \boldsymbol{b}=\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

33 $A\boldsymbol{x}=\begin{bmatrix}1\\3\end{bmatrix}$ 的完整解是 $\boldsymbol{x}=\begin{bmatrix}1\\0\end{bmatrix}+c\begin{bmatrix}0\\1\end{bmatrix}$,求 A。

挑战问题

34 (推荐) 假设你知道 3×4 矩阵 A 有向量 $\boldsymbol{s} = (2, 3, 1, 0)$ 作为 $A\boldsymbol{x} = \boldsymbol{0}$ 的唯一特殊解：

(a) A 的秩为何？$A\boldsymbol{x} = \boldsymbol{0}$ 的完整解为何？

(b) A 的确切的行简化阶梯形式 R 为何？

(c) 你如何知道对于所有的 $\boldsymbol{b}$，$A\boldsymbol{x} = \boldsymbol{b}$ 有解？

35 假设 K 是 9×9 的第二差分矩阵(2's 在对角线，–1 在对角线的上方及下方)，求解方程式 $K\boldsymbol{x} = \boldsymbol{b} = (10, ..., 10)$。如果在 x 轴的点 1, ..., 9 的上方画出对应的点画出 $x_1, ..., x_9$，我想这 9 个点会是一个抛物线。

36 假设对于每一个 $\boldsymbol{b}$，$A\boldsymbol{x} = \boldsymbol{b}$ 与 $C\boldsymbol{x} = \boldsymbol{b}$ 都有相同(完整)的解，则 A 等于 C 是对的吗？

37 描述简化阶梯形式 R 的列空间。

请尊重版权与译者的劳动成果，侵权必究!

3.4 无关，基底与维度

1. A 的**无关列**：$A\boldsymbol{x}=\mathbf{0}$ 的唯一解是 $\boldsymbol{x}=\mathbf{0}$，零空间是 $\mathbf{Z}$。
2. 无关向量：组合 $c_1\boldsymbol{v}_1+\ldots+c_k\boldsymbol{v}_k=\mathbf{0}$ 时，必然所有的 c's 都是 0。
3. 矩阵的 $m<n$ 时，会有**相关的列**：最少 $n-m$ 个**自由变量/特殊解**。
4. 若 $\boldsymbol{S}$ 是所有向量 $\boldsymbol{v}$'s 的组合，则 $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$ **生成**空间 $\boldsymbol{S}$。
5. 若向量 $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k$ 是无关而且生成 $\boldsymbol{S}$，则他们是 $\boldsymbol{S}$ 的**一组基底**[1]。
6. **空间 $\boldsymbol{S}$ 的维度**是 $\boldsymbol{S}$ 的每一组基底的向量个数。
7. 若 A 是 4×4 的可逆矩阵，则 A 的列是$\mathbf{R}^4$的一组基底，$\mathbf{R}^4$的维度是 4。

这个重要段落是关于子空间的**真实大小**，$m\times n$ 的矩阵有 n 个列，但是列空间的真实维度不一定是 n。维度的量测是依据**无关列**的个数--我们必须说明它的意义，我们会看到列空间的真实维度就是秩 r。

在任意的向量空间中，都可以针对向量 $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_s$ 使用无关的概念。本段落大部分聚焦在我们知道且常用的子空间--特别是 A 的列空间与零空间。前个部分我们学习了“向量”不一定是列向量，他们可以是矩阵与函数，他们可以是线性无关(或是相关)。第一个关键范例使用列向量。

我们的目标是了解基底：生成空间的无关向量。

空间中的每一个向量都是基底向量唯一的组合。

我们位于主题的中心，没有基底就不能继续前进。本段落 (第一个实际意义的暗示) 4 个重要的观念是：

1. **无关向量** (没有额外向量)
2. **生成一个空间** (足够的向量去产生剩余的向量)
3. **空间的基底** (不太多也不太少)
4. **空间的维度** (基底中向量的个数)

线性无关

无关的第一个定义不是那么传统，但是你应该准备充分了。

定义 当 $A\boldsymbol{x}=\mathbf{0}$ 的唯一解是 $\boldsymbol{x}=\mathbf{0}$，**$A$ 的列是线性无关**。没有其他的列组合 $A\boldsymbol{x}$ 是零向量。

[1] **译者注：**基底会包含数个向量，所以用“组”当做基底的单位。

当零空间 $N(A)$只有一个零向量时，A 的列是线性无关。让我们用$\mathbf{R}^3$的 3 个向量来说明线性无关(与相关)：

1. 若 3 个向量不在同一个平面，他们是无关的。除了 $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3$ 得到零向量之外，不存在 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 的组合可以得到零向量，如图 3.4 所示。

2. 若 3 个向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ 在同一个平面，他们是相关的。

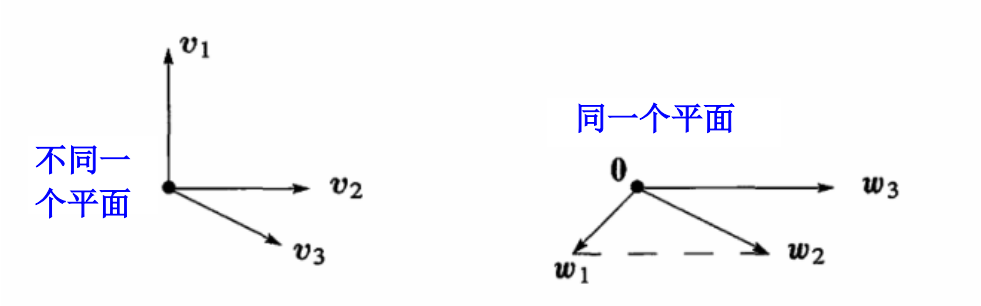


图 3.4：无关向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$，只有在 $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3$ 得到零向量。相关向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$，组合 $\boldsymbol{w}_1 - \boldsymbol{w}_2 + \boldsymbol{w}_3 = (0, 0, 0)$。

把无关的观念应用到 12 维空间的 7 个向量，如果他们是 A 的列，而且是无关，则零空间只包含零向量。7 个向量中没有任何一个是其他 6 个向量的组合。

现在我们用不同的词语来表达同样的概念，下列无关的定义会应用至任意向量空间中的任意向量序列。当向量是 A 的列，这两个定义完全相同。

定义　如果唯一得到零向量的组合是 $0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \ldots + 0\boldsymbol{v}_n$，则向量序列 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 是**线性无关**。

线性无关

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \ldots + x_n\boldsymbol{v}_n = \mathbf{0} \quad \text{只在 } x\text{'s 全是 0 的时候发生。} \tag{1}$$

当 x's 不全是 0，如果存在一个组合得到 $\mathbf{0}$，这些向量就是线性相关。

正确的语言：**“向量序列是线性无关。”** 可接受的短语：“向量是无关”。不可接受的：“矩阵是无关”。

一个向量序列不是相关就是无关，他们的组合可以得到零(x's 不全是 0)，也可能无法得到。所以关键问题是：什么样的向量组合可以得到零向量？我们从$\mathbf{R}^2$一些小范例开始：

(a) 向量(1, 0)与(0, 1)是无关。

(b) 向量(1, 0)与(1, 0.00001)是无关。

(c) 向量(1, 1)与(−1, −1)是相关。

(d) 向量(1, 1)与(0, 0)是相关，因为有零向量。

(e) 在$\mathbf{R}^2$中，任意 3 个向量(a, b)与(c, d)与(e, f)是相关。

请尊重版权与译者的劳动成果，侵权必究!

从几何来看，(1, 1)与(−1, −1)在一条通过原点的直线上，他们是相关。使用定义，求 x_1 与 x_2 使得 $x_1(1,1)+x_2(-1,-1)=(0,0)$，这个如同求解 $A\boldsymbol{x}=\boldsymbol{0}$：

$$\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix} \text{ 得到 } x_1=1,\ x_2=1$$

当零空间中有非零向量，这些列确实是相关。

如果其中一个 $\boldsymbol{v}$ 是零向量，就不可能是无关，为什么?

在$\mathbf{R}^2$中的 3 个向量不可能是无关，有一个方法可以看出：矩阵 A 有三个列必然会有一个自由变量，对于 $A\boldsymbol{x}=\boldsymbol{0}$ 就会有一个特殊解。另一种方法：如果前两个向量是无关，他们的某种组合会得到第三个向量，详情见下面第二个高亮处。

现在回到$\mathbf{R}^3$中的 3 个向量，若其中一个是另外一个的倍数，这些向量是相关。但是完整的测试是三个向量同时进行，我们把他们放到一个矩阵，然后试着求解 $A\boldsymbol{x}=\boldsymbol{0}$。

范例 1　A 的列是相关，$A\boldsymbol{x}=\boldsymbol{0}$ 有一个非零解：

$$A\boldsymbol{x}=\begin{bmatrix}1 & 0 & 3\\ 2 & 1 & 5\\ 1 & 0 & 3\end{bmatrix}\begin{bmatrix}-3\\ 1\\ 1\end{bmatrix} \quad \text{是} \quad -3\begin{bmatrix}1\\ 2\\ 1\end{bmatrix}+1\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}+1\begin{bmatrix}3\\ 5\\ 3\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

矩阵的秩只有 $r=2$。无关列会产生满列秩 $r=n=3$。

在这个矩阵中，行也是相关。行 1 减去行 3 是零行。对于方形矩阵而言，可以证明“相关的列”则有“相关的行”。

问题：如何求解 $A\boldsymbol{x}=\boldsymbol{0}$? 系统化的方法是消元法：

$$A=\begin{bmatrix}1 & 0 & 3\\ 2 & 1 & 5\\ 1 & 0 & 3\end{bmatrix} \quad \text{简化得到} \quad R=\begin{bmatrix}1 & 0 & 3\\ 0 & 1 & -1\\ 0 & 0 & 0\end{bmatrix}$$

解 $\boldsymbol{x}=(-3,1,1)$恰好就是**特殊解**，可以说明自由列(列 3)如何成为主元列的组合，这就杀死了无关。

满列秩　当 A 的秩恰好 $r=n$ 时，则 A 的列是无关：有 n 个主元，没有自由变数，只有 $\boldsymbol{x}=\boldsymbol{0}$ 在零空间中。

有一种情形特别重要，这是因为从开头它就很清楚。假设有 7 个列，每个列有 5 个分量($m=5$ 小于 $n=7$)，这些列必然是相关。$\mathbf{R}^5$中的 7 个向量是相关，A 的秩不可能大于 5，5 个行不可能有超过 5 个主元。$A\boldsymbol{x}=\boldsymbol{0}$ 最少有 $7-5=2$ 个自由变量，所以有非零解—这就表示列是相关。

若 $n>m$，则$\mathbf{R}^m$的任意 n 个向量必然是相关。

这种类型矩阵的列比行多--属于矮宽形。若 $n > m$，这些列必然是相关，这是因为 $A\boldsymbol{x} = \mathbf{0}$ 有非零解。

若 $n \le m$，这些列可能相关也可能无关。从消元法会得到 r 个主元列，这 r 个主元列是无关。

注意：另一种描述线性相关的方法是：“一个向量是其他向量的组合。”听起来很简洁，为什么我们不从这个方法开始？我们的定义比较长：“除了每个系数 x 都是零的平凡组合外，存在某个组合可以得到零向量。”我们必须排除简单方法可以得到零向量的可能，零的平凡组合让每位作者都很头痛。如果一个向量是其他向量的组合，这个向量的系数是 $x = 1$。

有一点要说明，我们的定义并没有把选出一个特定向量视为罪恶，A 的所有列都是等同对待。我们检视 $A\boldsymbol{x} = \mathbf{0}$，方程式有非零解也可能没有。在最后，这个会比询问最后一列(或第一列，或中间的某一列)是否为其他列的组合来的较好。

向量生成子空间

本书第一个子空间是列空间，从列 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 开始，子空间被所有的组合 $x_1\boldsymbol{v}_1 + \ldots + x_n\boldsymbol{v}_n$ 填满，列空间包含列的所有组合 $A\boldsymbol{x}$。我们现在介绍单一的单词 span(生成)，它的描述是：列空间是被列生成。

定义　若一个向量集合的线性组合填满一个空间，则向量集合**生成**这个空间。

矩阵的列生成它的列空间，他们可能是相关。

范例 2　$\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 与 $\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 生成整个二维空间 $\mathbf{R}^2$。

范例 3　$\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$，$\boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$，$\boldsymbol{v}_3 = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$ 也生成整个空间 $\mathbf{R}^2$。

范例 4　$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 与 $\boldsymbol{w}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ 只生成 $\mathbf{R}^2$ 的一条直线，$\boldsymbol{w}_1$ 自己就能做到。

考虑两个 3 维空间中由(0, 0, 0)出发的两个向量，一般来说会生成一个平面。你的心里会想可以由他们的线性组合得到这个平面。数学上来说，你还知道有其他可能性：两个向量可以生成一条直线，三个向量可以生成整个 $\mathbf{R}^3$，或者只生成一个平面，甚至有可能三个向量只生成一条直线，或是 10 个向量只生成一个平面。他们当然不是无关。

列生成列空间，有个新的子空间--由行生成。**行的组合得到“行空间”**。

请尊重版权与译者的劳动成果，侵权必究!

定义　矩阵的**行空间**是$\mathbf{R}^n$的子空间，**行生成行空间**。

A 的行空间是 $C(A^{\mathrm{T}})$，就是 A^{T} 的列空间。

$m\times n$ 矩阵的行有 n 个分量，他们是$\mathbf{R}^n$的向量--或者把他们直接写成列向量。有一个快速的方法可以做到：转置矩阵。不再关注 A 的行，而是检视 A^{T} 的列。同样的数字，但是在列空间 $\boldsymbol{C}(A^{\mathrm{T}})$中。$A$ 的行空间是$\mathbf{R}^n$的子空间。

范例 5　描述 A 的列空间与行空间。

$$A=\begin{bmatrix}1&4\\2&7\\3&5\end{bmatrix}\text{与 }A^{\mathrm{T}}=\begin{bmatrix}1&2&3\\4&7&5\end{bmatrix}\text{，此处 }m=3\text{，}n=2\text{。}$$

A 的列空间是$\mathbf{R}^3$的平面，由 A 的两个列生成。A 的行空间由 A 的三个行(A^{T} 的列)生成，行空间是全部的$\mathbf{R}^2$。记得：这些行在$\mathbf{R}^n$中生成了行空间，列在$\mathbf{R}^m$中，生成了列空间。同样的数字，不同的向量，不同的空间。

向量空间的基底

两个向量不能生成整个$\mathbf{R}^3$，纵使他们是无关。四个向量不可能是无关，纵使他们生成$\mathbf{R}^3$。我们想要有足够无关向量去生成空间(不能超过)，基底就是我们想要的。

定义　向量空间的一组基底是由一序列的向量组成，具有两个性质：

基底向量是线性无关，他们生成了空间。

这两个性质的组合是线性代数的基础，空间中的每一个向量 $\boldsymbol{v}$ 都是基底向量的组合，因为他们生成这个空间。除此之外，得到 $\boldsymbol{v}$ 的组合是唯一的，因为基底向量 $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$ 是无关。

存在一个而且只有一个方式将 $\boldsymbol{v}$ 写成基底向量的组合。

理由：假设 $\boldsymbol{v}=a_1\boldsymbol{v}_1+\ldots+a_n\boldsymbol{v}_n$，且 $\boldsymbol{v}=b_1\boldsymbol{v}_1+\ldots+b_n\boldsymbol{v}_n$，两式相减得到$(a_1-b_1)\boldsymbol{v}_1+\ldots+(a_n-b_n)\boldsymbol{v}_n=\boldsymbol{0}$。基于 $\boldsymbol{v}$ 的无关性，每个 $a_i-b_i=0$，所以 $a_i=b_i$，不存在两种方式得到 $\boldsymbol{v}$。

范例 6　$I=\begin{bmatrix}1&0\\0&1\end{bmatrix}$的列产生$\mathbf{R}^2$的**标准基底(standard basis)**。

$$\text{基底向量 } \boldsymbol{i}=\begin{bmatrix}1\\0\end{bmatrix} \text{ 与 } \boldsymbol{j}=\begin{bmatrix}0\\1\end{bmatrix} \text{是无关，他们生成} \mathbf{R}^2$$

每个人都会先想到这组基底，向量 $\boldsymbol{i}$ 是横跨移动，$\boldsymbol{j}$ 是铅直移动。3×3 单位矩阵的列就是标准基底 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$，$n\times n$ 单位矩阵的列就是$\mathbf{R}^n$的**标准基底**。

现在我们发现了其他很多基底(无限多)，基底不是唯一。

范例 7 **(重要)** 每一个可逆的 $n\times n$ 矩阵所有的列，是$\mathbf{R}^n$的一组基底：

$$\begin{array}{l}\text{可逆矩阵}\\ \text{无关的列}\\ \text{列空间是}\mathbf{R}^3\end{array} \quad A=\begin{bmatrix}1&0&0\\1&1&0\\1&1&1\end{bmatrix} \qquad \begin{array}{l}\text{奇异矩阵}\\ \text{相关的列}\\ \text{列空间}\neq\mathbf{R}^3\end{array} \quad B=\begin{bmatrix}1&0&1\\1&1&2\\1&1&2\end{bmatrix}$$

$A\boldsymbol{x}=\mathbf{0}$ 的唯一解是 $\boldsymbol{x}=A^{-1}\mathbf{0}=\mathbf{0}$，列是无关，他们生成整个空间$\mathbf{R}^n$——因为每个向量 $\boldsymbol{b}$ 都是列的组合，$A\boldsymbol{x}=\boldsymbol{b}$ 永远可以求解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$。你看出来每件事情都跟可逆矩阵有关吗？用一句话来说明：

当向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 恰好是一个 $n\times n$ 可逆矩阵的列时，这些向量是$\mathbf{R}^n$的一组基底，于是$\mathbf{R}^n$有无限多基底。

当列是相关，我们只保存主元列——上述 B 的前两个列具有两个主元。他们是无关且生成列空间。

A 的主元列是列空间的一组基底，A 的主元行是行空间的一组基底。阶梯形式 R 的主元行也是如此。

范例 8 一个不可逆矩阵，它的列不是任何空间的基底。

$$\begin{array}{l}\text{一个主元列}\\ \text{一个主元行}\quad (r=1)\end{array} \qquad A=\begin{bmatrix}2&4\\3&6\end{bmatrix} \text{ 简化成 } R=\begin{bmatrix}1&2\\0&0\end{bmatrix}$$

A 的列 1 是主元列，单独这个列是列空间的基底。A 的第二列可以是一个不同的基底，列 1 所有的非零倍数都是基底。我们不缺基底，一个肯定的选择是主元列。

注意 R 的主元列(1, 0)的尾端是 0，这是 R 的列空间的一组基底，但是它不属于 A 的列空间，A 与 R 的列空间不相同，他们的基底也不相同(他们的维度相同)。

A 的行空间与 R 的行空间相同，它包含(2, 4), (1, 2)以及这些向量全部的其他倍数，我们永远有无限多的基底可以选择，这些基底来自前述的向量。一个自然的选择是选取 R 的非零行(有主元的行)，所以这个秩一的矩阵 A 只有一个向量在基底中：

请尊重版权与译者的劳动成果，侵权必究!

列空间的基底 $\begin{bmatrix}2\\3\end{bmatrix}$　　行空间的基底 $\begin{bmatrix}1\\2\end{bmatrix}$

下一章会再探讨列空间与行空间的基底，我们很高兴遇到刚开始的范例，其中的情景很清晰(基底的概念还是新的)。下个例子比较大，但是一样很清晰。

范例 9　求出列空间与行空间的基底，其中的矩阵是秩二。

$$R=\begin{bmatrix}1&2&0&3\\0&0&1&4\\0&0&0&0\end{bmatrix}$$

列 1 与 3 是主元列，他们是 R 的列空间的一组基底。列空间中的向量全部具有 $\boldsymbol{b}=(x,y,0)$的形式，R 的列空间是整个三维 xyz 空间中的“xy”平面，这个平面不是$\mathbf{R}^2$，而是$\mathbf{R}^3$的子空间。列 2 与 3 也是相同列空间的一组基底，R 的哪两个列不是列空间的一组基底?

R 的行空间是$\mathbf{R}^4$的子空间，这个行空间的最简单基底就是 R 的两个非零行。第三行(零向量)也在行空间中，但它不是行空间的基底，基底向量必须无关。

问题：　给定$\mathbf{R}^7$的 5 个向量，如何求得这 5 个向量生成的空间的一组基底?

第一答案：把这些行变成矩阵 A 的行，利用消元法找到 R 的非零行。

第二答案：把这 5 个向量变成矩阵 A 的列，利用消元法找到 A 的主元列(不是 R 的)，这些主元列就是列空间的一组基底。

其他的基底有可能具有更多或是更少的向量吗？这是一个很重要的问题，它的答案是 no，向量空间所有的基底都具有相同的向量数目。

任何一组基底的向量的个数，就是空间的维度。

向量空间的基底

我们必须证明前面的说法。存在很多种方法来选取基底向量，但是每组基底向量的个数永远不会改变。

若 $\boldsymbol{v}_1,\dots,\boldsymbol{v}_m$ 与 $\boldsymbol{w}_1,\dots,\boldsymbol{w}_n$ 都是同一个向量空间的基底，则有 $\boldsymbol{m}=\boldsymbol{n}$。

证明：　假设 $\boldsymbol{w}$ 的数目比 $\boldsymbol{v}$ 多，$n>m$，我们希望得到矛盾。全部 $\boldsymbol{v}$ 向量构成一组基底，所以 $\boldsymbol{w}_1$ 必须是 $\boldsymbol{v}$ 向量的组合。

若 $\boldsymbol{w}_1 = a_{11}\boldsymbol{v}_1 + \ldots + a_{m1}\boldsymbol{v}_m$，这是矩阵乘法 VA 的第一列：

每个 $\boldsymbol{w}$ 是 $\boldsymbol{v}$ 的组合

$$W = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \ldots & \boldsymbol{w}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_m \end{bmatrix} \begin{bmatrix} a_{11} & & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{bmatrix} = VA$$

我们不知道每个 a_{ij}，但是我们知道 A 的形状(它是 $m \times n$)。第二个向量 $\boldsymbol{w}_2$ 也是向量 $\boldsymbol{v}$'s 的组合，组合的系数得到 A 的第二列。关键是 A 对于每个 $\boldsymbol{v}$ 有一行，对于每个 $\boldsymbol{w}$ 有一列。由于假设 $n > m$，所以 A 是一个矮宽矩阵，$A\boldsymbol{x} = \boldsymbol{0}$ 有一个非零解。

$A\boldsymbol{x} = \boldsymbol{0}$ 得到 $VA\boldsymbol{x} = \boldsymbol{0}$，也就是 $W\boldsymbol{x} = \boldsymbol{0}$，$\boldsymbol{w}$ 的组合得到零！所以这些 $\boldsymbol{w}$ 不能成为基底——我们的假设 $n > m$，对于两组基底来说是不可能的。

如果 $m > n$，我们交换 $\boldsymbol{v}$ 与 $\boldsymbol{w}$，再重复刚才的步骤。唯一避免矛盾的方法就是 $m = n$，我们完成了 $m = n$ 的证明。

基底向量的个数与空间有关——不是与特定的基底有关。每组基底的数字都会相同，这个数字就是空间的自由度，空间 $\mathbf{R}^n$ 的维度是 n。我们现在介绍一个重要的单词“维度”，也适用于其他的向量空间。

定义 **一个空间的维度就是每组基底的向量个数。**

这个适配我们的直觉，通过 $\boldsymbol{v} = (1, 5, 2)$ 的直线的维度是 1，它是一个子空间，这个子空间的基底是一个向量 $\boldsymbol{v}$。垂直这条直线的平面是 $x + 5y + 2z = 0$，这个平面的维度是 2。要证明前面的推论，我们知道一组基底是 $(-5, 1, 0)$ 与 $(-2, 0, 1)$，因为基底有两个向量，所以维度是 2。

这个平面是矩阵 $A = [1 \quad 5 \quad 2]$ 的零空间，其中有两个自由变量。我们的基底 $(-5, 1, 0)$ 与 $(-2, 0, 1)$ 是 $A\boldsymbol{x} = \boldsymbol{0}$ 的**特殊解**。下个段落会展示 $n - r$ 个**特殊解**永远可以得到零空间的一组基底。$\boldsymbol{C}(A)$ 有维度 r，零空间 $\boldsymbol{N}(A)$ 有维度 $n - r$。

线性代数语言的注解 我们没有说过“空间的秩”或是“基底的维度”或是“矩阵的基底”，这些术语是没有意义的。我们说“列空间的维度等于矩阵的秩”。

矩阵空间的基底与函数空间

这些单词“无关”“基底”“维度”并不局限在列向量。我们可以询问三个矩阵是否无关，这些矩阵是在所有 3×4 矩阵形成的空间中，某些组合可能会得到零矩阵。我们也可以询问所有 3×4 矩阵形成的矩阵空间的维度是多少？(答案是 12)。

微分方程式 $d^2y / dx^2 = y$ 有一个解空间，其中一组基底是 $y = e^x$ 与 $y = e^{-x}$，计算基底函数的个数，得到所有解组成的空间的维度是 2。(因为是二次导数，所以维度是 2)。

请尊重版权与译者的劳动成果，侵权必究！

矩阵空间与函数空间在$\mathbf{R}^n$之后看起来会有点小奇怪，但是在某个方面来说，直到你可以应用基底与维度到列向量之外的“向量”之前，你还没有完全了解基底与维度的概念。

矩阵空间　向量空间 **M** 包含所有的 2×2 矩阵，它的维度是 4。

$$\textbf{一组基底是}\quad A_1, A_2, A_3, A_4 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$$

这些矩阵是线性无关，我们检视的不是列向量，而是整个矩阵。这 4 个矩阵的组合可以产生 **M** 的任意矩阵，所以他们生成了空间：

$$\begin{matrix}\text{每个}A\text{都是基}\\ \text{底矩阵的组合}\end{matrix}\quad c_1A_1 + c_2A_2 + c_3A_3 + c_4A_4 = \begin{bmatrix}c_1 & c_2\\ c_3 & c_4\end{bmatrix} = A$$

只有在所有的 c 都是零，A 才会是零––这个证明了 A_1, A_2, A_3, A_4 的无关性。

三个矩阵 A_1, A_2, A_4 是子空间的一组基底––**上三角矩阵**，它的维度是 3。A_1 与 A_4 是对角矩阵的一组基底。请问对称矩阵的基底是什么？保持 A_1 与 A_4，剩下两个单元填入 $A_2 + A_3$。

更深入的说，考虑所有 $n\times n$ 矩阵形成的空间，一个可能的基底就是每个矩阵都只在一个单元有非零数(单元是 1)，有 n^2 个位置是 1，所以有 n^2 组基底矩阵。

整个 $n\times n$ 矩阵形成的空间，维度是 n^2。
上三角矩阵形成的子空间，维度是 $n^2/2 + n/2$。
对角矩阵形成的子空间，　维度是 n。
对称矩阵形成的子空间，　维度是 $n^2/2 + n/2$ (为什么?)

函数空间　方程式 $d^2y/dx^2 = 0$ 与 $d^2y/dx^2 = -y$ 与 $d^2y/dx^2 = y$ 与二次导数有关，利用微积分的技巧我们求解函数 $y(x)$：

$y'' = 0$，　　解是任意函数 $y = cx + d$
$y'' = -y$，　　解是任意组合 $y = c\sin x + d\cos x$
$y'' = y$，　　解是任意组合 $y = ce^x + de^{-x}$

$y'' = -y$ 的解空间有两个基底函数 $\sin x$ 与 $\cos x$；$y'' = 0$ 的解空间有 x 与 1，这个是二次导数的零空间！每个案例的维度都是 2 (这些是二阶方程式)。

$y'' = 2$ 的解无法形成子空间––右侧的 $b = 2$ 不是 0，存在一个特定 $y(x) = x^2$，完整解是 $x^2 + cx + d$，这样的函数都满足 $y'' = 2$。注意：**特解**加上零空间中任意的函数 $cx + d$。线性微分方程式就好像矩阵方程式 $A\boldsymbol{x} = \boldsymbol{b}$，只是我们利用微积分求解，而不是线性代数。

我们以只包含零向量的空间 **Z** 作为结束，这个空间的维度是 0，**空集合**(没有任何向量)是 **Z** 的基底。因为会失去线性无关性，我们不再允许基底包含零向量。

主要观念的复习

1. 若 $\boldsymbol{x}=\mathbf{0}$ 是 $A\boldsymbol{x}=\mathbf{0}$ 的唯一解，则 A 的列是线性无关。
2. 若向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ 的组合填满一个空间，则这些向量生成了空间。
3. 基底包含线性无关的向量，这些向量生成 $\boldsymbol{S}$，空间中的每个向量都是基底向量的唯一组合。
4. 空间的所有基底有相同的向量个数，基底的向量个数就是空间的维度。
5. 主元列是列空间的一组基底，维度是 r。

已解范例

3.4A　从向量 $\boldsymbol{v}_1=(1, 2, 0)$与 $\boldsymbol{v}_2=(2, 3, 0)$开始，(a) 是否是线性无关？ (b) 是否是任意空间的基底？ (c) 会生成什么空间 **V**？ (d) 空间 **V** 的维度 = ？ (e) 哪个矩阵 A 的列空间是 **V**？ (f) 哪个矩阵 A 的零空间是 **V**？ (g) 描述所有的向量 $\boldsymbol{v}_3$，使得 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 可以是$\mathbf{R}^3$的基底。

解

(a) $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 是线性无关--得到 $\mathbf{0}$ 的唯一组合是 $0\boldsymbol{v}_1+0\boldsymbol{v}_2=\mathbf{0}$。

(b) Yes，他们是生成空间的一组基底。

(c) 空间 **V** 包含所有向量$(x, y, 0)$，这是$\mathbf{R}^3$中的 xy 平面。

(d) 空间 **V** 的维度是 2，因为基底有两个向量。

(e) 若 A 的每个列都是 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 的组合，则空间 **V** 是任意 $3\times n$ 矩阵 A 的列空间，A 的秩为 2。特定的 A 可以只有 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 作为列。

(f) 若 B 的每个行都是$(0, 0, 1)$的倍数，则空间 **V** 是任意 $m\times 3$ 矩阵 B 的零空间，B 的秩是 1。特定的 B 选成 $B=[0 \quad 0 \quad 1]$，则有 $B\boldsymbol{v}_1=\mathbf{0}$ 与 $B\boldsymbol{v}_2=\mathbf{0}$。

(g) 加入任意第三个向量 $\boldsymbol{v}_3=(a, b, c)$，其中 $c\neq 0$，可以组成$\mathbf{R}^3$的基底。

请尊重版权与译者的劳动成果，侵权必究！

3.4B 从三个无关向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ 开始，这些向量的组合得到 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$，组合的矩阵形式可以写成 $V = WB$:

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{w}_1 + \boldsymbol{w}_2 \\ \boldsymbol{v}_2 &= \boldsymbol{w}_1 + 2\boldsymbol{w}_2 + \boldsymbol{w}_3 \\ \boldsymbol{v}_3 &= \boldsymbol{w}_2 + c\boldsymbol{w}_3 \end{aligned} \quad \text{就是} \quad \begin{bmatrix} & & \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3 \\ & & \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & c \end{bmatrix}$$

针对 B 进行什么样的测试，可以验证 $V = WB$ 具有无关的列？若 $c \neq 1$，证明 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是线性无关。若 $c = 1$，证明 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是线性相关。

解 可以利用第一个定义测试 V 的列的无关性：V 的零空间必须只包含零向量，则 $\boldsymbol{x} = \boldsymbol{0}$ 是列的唯一组合使得 $V\boldsymbol{x} =$ 零向量。

若问题中的 $c = 1$，我们可以从两个方面看出相关性：首先 $\boldsymbol{v}_1 + \boldsymbol{v}_3$ 会等于 $\boldsymbol{v}_2$ (如果你把 $\boldsymbol{w}_1 + \boldsymbol{w}_2$ 与 $\boldsymbol{w}_2 + \boldsymbol{w}_3$ 相加会得到 $\boldsymbol{w}_1 + 2\boldsymbol{w}_2 + \boldsymbol{w}_3 = \boldsymbol{v}_2$，换言之，$\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3 = \boldsymbol{0}$——说明所有的 $\boldsymbol{v}$ 向量不是线性无关。

另一个方法是检视 B 的零空间。若 $c = 1$，则 $\boldsymbol{x} = (1, -1, 1)$在零空间中，且 $B\boldsymbol{x} = \boldsymbol{0}$，当然 $WB\boldsymbol{x} = \boldsymbol{0}$，也就是 $V\boldsymbol{x} = \boldsymbol{0}$，因此所有的 $\boldsymbol{v}$ 向量是相关。来自零空间的特定 $\boldsymbol{x} = (1, -1, 1)$再次告诉我们 $\boldsymbol{v}_1 - \boldsymbol{v}_2 + \boldsymbol{v}_3 = \boldsymbol{0}$。

现在假设 $c \neq 1$，矩阵 B 是可逆的，如果 $\boldsymbol{x}$ 是任意非零向量，我们知道 $B\boldsymbol{x}$ 不是零。由于 $V = WB$，就是说 $\boldsymbol{x}$ 不在 V 的零空间中，换言之，$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是无关。

通用的规则就是“当 B 可逆时，来自无关 $\boldsymbol{w}$ 的 $\boldsymbol{v}$ 也是无关。”如果这些向量是在$\mathbf{R}^3$中，他们不只是无关，还是$\mathbf{R}^3$的一组基底。

“当基底转换矩阵 B 是可逆，来自基底 $\boldsymbol{w}$ 的 $\boldsymbol{v}$ 也是基底。”

3.4C (重要范例) 假设 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 是$\mathbf{R}^n$的一组基底，且 $n \times n$ 矩阵 A 可逆。证明 $A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_n$ 也是$\mathbf{R}^n$的一组基底。

解 以矩阵语言来说：把基底向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 放在一个可逆矩阵 V 的列，则 $A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_n$ 就是 AV 的列。由于 A 是可逆，因此 AV 也是可逆，AV 的列就是一组基底。

以向量语言来说：假设 $c_1A\boldsymbol{v}_1 + \ldots + c_nA\boldsymbol{v}_n = \boldsymbol{0}$，可以写成 $A\boldsymbol{v} = \boldsymbol{0}$，其中 $\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n$。左乘 A^{-1} 使得 $\boldsymbol{v} = \boldsymbol{0}$，由于所有 $\boldsymbol{v}$'s 向量的无关性，所有的 c_i 必然是 0，我们证明了所有 $A\boldsymbol{v}$'s 向量是无关。

现在证明所有的 $A\boldsymbol{v}$'s 向量生成$\mathbf{R}^n$，求解 $c_1A\boldsymbol{v}_1 + \ldots + c_nA\boldsymbol{v}_n = \boldsymbol{b}$，等效于求解 $c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n = A^{-1}\boldsymbol{b}$，由于 $\boldsymbol{v}$'s 向量是一组基底，前式必然有解。

问题集 3.4

问题 1-10 有关线性无关与线性相关。

1 证明 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是线性无关，但是 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ 是线性相关：

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

求解 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + c_4\boldsymbol{v}_4 = \boldsymbol{0}$ 或 $A\boldsymbol{x} = \boldsymbol{0}$，把 $\boldsymbol{v}$'s 引入 A 的列。

2. (推荐) 求出下列向量极大可能的无关向量个数。

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v}_5 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_6 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

3 证明：若 $a = 0$ 或 $d = 0$ 或 $f = 0$ (3 种情形)，U 的列是相关：

$$U = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix}$$

4 若问题 3 的 a, d, f 全部不是 0，证明 $U\boldsymbol{x} = \boldsymbol{0}$ 的唯一解是 $\boldsymbol{x} = \boldsymbol{0}$，且上三角 U 有无关的列。

5 决定无关或相关：

(a) 向量(1, 3, 2), (2, 1, 3), (3, 2, 1)

(b) 向量(1, –3, 2), (2, 1, –3), (–3, 2, 1)

6 选择 U 的 3 个无关的列，然后做出另外两种选择，对 A 做同样的事情：

$$U = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & 6 & 7 & 0 \\ 0 & 0 & 0 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{与} \quad A = \begin{bmatrix} 2 & 3 & 4 & 1 \\ 0 & 6 & 7 & 0 \\ 0 & 0 & 0 & 9 \\ 4 & 6 & 8 & 2 \end{bmatrix}$$

7 若 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ 是无关向量，证明差 $\boldsymbol{v}_1 = \boldsymbol{w}_2 - \boldsymbol{w}_3, \boldsymbol{v}_2 = \boldsymbol{w}_1 - \boldsymbol{w}_3, \boldsymbol{v}_3 = \boldsymbol{w}_1 - \boldsymbol{w}_2$ 是相关。找出 $\boldsymbol{v}$'s 得到 $\boldsymbol{0}$ 的组合。前式写成 $[\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3] = [\boldsymbol{w}_1 \quad \boldsymbol{w}_2 \quad \boldsymbol{w}_3]\, A$，请问 A 是奇异吗？

8 若 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ 是无关向量，证明和 $\boldsymbol{v}_1 = \boldsymbol{w}_2 + \boldsymbol{w}_3, \boldsymbol{v}_2 = \boldsymbol{w}_1 + \boldsymbol{w}_3, \boldsymbol{v}_3 = \boldsymbol{w}_1 + \boldsymbol{w}_2$ 是无关。(将 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{0}$ 写成 $\boldsymbol{w}$'s 的项，求解方程式得到 c's，证明都是 0)

请尊重版权与译者的劳动成果，侵权必究!

9 假设 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$ 是$\mathbf{R}^3$的向量。

(a) 这 4 个向量是无关，因为_____。

(b) 向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 是无关，如果_____。

(c) 向量 $\boldsymbol{v}_1$ 与(0, 0, 0)是相关，因为_____。

10 在$\mathbf{R}^4$的平面 $x + 2y - 3z - t = 0$ 中，找出两个无关向量，再找出 3 个无关向量。为什么找不到 4 个？这个平面是哪个矩阵的零空间?

问题 1-14 有关一组向量集合生成空间，选取向量的所有线性组合。

11 描述下列向量生成的$\mathbf{R}^3$的子空间(线或平面或$\mathbf{R}^3$？)

(a) 两个向量(1, 1, −1)与(−1, −1, 1)

(b) 三个向量(0, 1, 1)与(1, 1, 0)与(0, 0, 0)

(c) 所有$\mathbf{R}^3$中分量是整个数值的向量。

(d) 所有$\mathbf{R}^3$中分量是正数的向量。

12 当_____有解时，向量 $\boldsymbol{b}$ 在 A 的列生成的子空间中。当_____有解时，向量 $\boldsymbol{c}$ 在 A 的行空间中。

是非题：若零向量在这个行空间中，行是相关。

13 求出 4 个空间的维度，哪两个空间是相同的？ (a) A 的列空间 (b) U 的列空间 (c) A 的行空间 (d) U 的行空间：

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 1 \\ 3 & 1 & -1 \end{bmatrix} \quad 与 \quad U = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

14 $\boldsymbol{v} + \boldsymbol{w}$ 与 $\boldsymbol{v} - \boldsymbol{w}$ 是 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的组合，将 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 写成 $\boldsymbol{v} + \boldsymbol{w}$ 与 $\boldsymbol{v} - \boldsymbol{w}$ 的组合。这两对向量_____相同的空间，在什么条件下，他们是这个相同空间的基底?

问题 15-25 有关基底的要求。

15 如果 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ 是线性无关，他们生成的空间的维度是_____。这些向量是空间的一组_____。如果这些向量是一个 $m \times n$ 矩阵的列，则 m ____ n。如果 $m = n$，矩阵是_____。

16 求出每个$\mathbf{R}^4$的子空间的一组基底：

(a) 所有分量相等的向量。

(b) 所有分量相加等于 0 的向量。

(c) 所有垂直于(1, 1, 0, 0)与(1, 0, 1, 1)的向量。

(d) I (4×4)的列空间与零空间。

17 对于 $U=\begin{bmatrix}1&0&1&0&1\\0&1&0&1&0\end{bmatrix}$ 的列空间，求出 3 组不同基底，然后求出 U 的行空间的两组不同基底。

18 假设 $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_6$ 是 $\mathbf{R}^4$ 的 6 个向量。

(a) 这些向量(会) (不会) (可能不会)生成 $\mathbf{R}^4$。

(b) 这些向量(是) (不是) (可能是)线性无关。

(c) 其中任意 4 个向量(是) (不是) (可能是) $\mathbf{R}^4$ 的一组基底。

19 A 的列是来自 $\mathbf{R}^m$ 的 n 个向量，如果他们是线性无关，A 的秩为何？如果他们生成 $\mathbf{R}^m$，秩是多少？如果他们是 $\mathbf{R}^m$ 的一组基底，秩是多少？往前看：秩 r 是_____列的个数。

20 求出 $\mathbf{R}^3$ 的平面 $x - 2y + 3z = 0$ 的一组基底，然后求出这个平面与 xy 平面的交集，然后求出所有垂直于这个平面的向量。

21 假设 5×5 矩阵 A 的列是 $\mathbf{R}^5$ 的一组基底。

(a) 因为_____，方程式 $A\boldsymbol{x} = \mathbf{0}$ 只有唯一解 $\boldsymbol{x} = \mathbf{0}$。

(b) 若 $\boldsymbol{b}$ 在 $\mathbf{R}^5$ 中，因为基底向量_____ $\mathbf{R}^5$，方程式 $A\boldsymbol{x} = \boldsymbol{b}$ 有解。

注意：A 是可逆，它的秩是 5，它的行也是 $\mathbf{R}^5$ 的一组基底。

22 假设 $\boldsymbol{S}$ 是 $\mathbf{R}^6$ 中的 5 维子空间，是非题(如果错误，举出例子)。

(a) 额外加入一个向量，$\boldsymbol{S}$ 的每一组基底都可以延伸成为 $\mathbf{R}^6$ 的一组基底。

(b) 移除一个向量，$\mathbf{R}^6$ 的每一组基底都可以简化成为 $\boldsymbol{S}$ 的一组基底。

23 把 A 的行 3 减去行 1 得到 U：

$$A=\begin{bmatrix}1&3&2\\0&1&1\\1&3&2\end{bmatrix} \quad 与 \quad U=\begin{bmatrix}1&3&2\\0&1&1\\0&0&0\end{bmatrix}$$

求出两个列空间的基底，求出两个行空间的基底，求出两个零空间的基底，哪个空间在消元法过程中保持固定。

24 是非题(给出理由)。

(a) 如果矩阵的列是无关，行也会是无关。

(b) 2×2 矩阵的列空间与它的行空间相同。

(c) 2×2 矩阵的列空间与行空间有相同的维度。

(d) 矩阵的列是它的列空间的一组基底。

请尊重版权与译者的劳动成果，侵权必究!

25 哪些数值 c 与 d 使得矩阵的秩为 2?

$$A=\begin{bmatrix}1&2&5&0&5\\0&0&c&2&2\\0&0&0&d&2\end{bmatrix} \text{ 与 } \quad B=\begin{bmatrix}c&d\\d&c\end{bmatrix}$$

问题 26-30 有关“向量”是矩阵的空间。

26 求出下列 3×3 矩阵的子空间的基底与维度。

(a) 所有的对角矩阵。

(b) 所有的对称矩阵($A^{\mathrm{T}}=A$)。

(c) 所有的反对称矩阵($A^{\mathrm{T}}=-A$)。

27 写出 6 个线性无关的 3×3 阶梯矩阵 $U_1, ..., U_6$。

28 所有每个列的总和都是 0 的 2×3 矩阵写成一个空间，求出它的一组基底。如果改成每个行的总和是 0，求出它的一组基底。

29 下列集合生成的 3×3 矩阵的子空间是什么？

(a) 可逆矩阵。

(b) 秩一矩阵。

(c) 单位矩阵。

30 2×3 矩阵的零空间包含(2, 1, 1)，求出此类矩阵形成的空间的一组基底。

问题 31-35 有关“向量”是函数的空间。

31 (a) 求出满足 $dy/dx = 0$ 的所有函数。

(b) 选择满足 $dy/dx = 3$ 的一个特解。

(c) 求出满足 $dy/dx = 3$ 的所有函数。

32 余弦空间 $\mathbf{F}_3$ 包含所有的组合 $y(x) = A\cos x + B\cos 2x + C\cos 3x$，求出子空间的一组基底，其中 $y(0) = 0$。

33 求出满足下列方程式的函数空间的一组基底。

(a) $dy/dx - 2y = 0$

(b) $dy/dx - y/x = 0$

34 假设 $y_1(x), y_2(x), y_3(x)$是 x 的 3 个不同函数，他们生成的空间的维度可能是 1，2 或 3，给出 y_1, y_2, y_3 的例子来说明这些可能性。

35 求出次数 ≤ 3 的多项式 $p(x)$形成的空间的一组基底。如果子空间的 $p(1) = 0$，求出一组基底。

36 对于 $a + c + d = 0$ 的向量(a, b, c, d)形成的空间 **S**，找出它的一组基底。同时对 $a + b = 0$ 与 $c = 2d$ 的向量(a, b, c, d)形成的空间 **T**，找出它的一组基底。两者交集 **S** ∩ **T** 的维度是多少？

网址：linear.neocities.org 邮箱：876152660@qq.com

37 若平移矩阵 S 有 $AS=SA$，证明 A 必须有下列的特殊形式：

若 $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$，则 $A=\begin{bmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{bmatrix}$

“与平移矩阵 S 可以交换的矩阵形成的子空间有维度_____。”

38 下列何者是$\mathbf{R}^3$的基底：

(a) (1, 2, 0)与(0, 1, –1)

(b) (1, 1, –1)，(2, 3, 4)，(4, 1, –1)，(0, 1, –1)

(c) (1, 2, 2)，(–1, 2, 1)，(0, 8, 0)

(d) (1, 2, 2)，(–1, 2, 1)，(0, 8, 6)

39 假设 A 是 5×4 且秩 4，证明当 5×5 矩阵[A $\boldsymbol{b}$]可逆时，$A\boldsymbol{x}=\boldsymbol{b}$ 无解。证明当 [A $\boldsymbol{b}$]奇异时，$A\boldsymbol{x}=\boldsymbol{b}$ 有解。

40 (a) 求所有$\dfrac{d^4y}{dx^4}=y(x)$的解的一组基底。

(b) 求$\dfrac{d^4y}{dx^4}=y(x)+1$的一个特解，求出完整解。

挑战问题

41 将 3×3 单位矩阵写成其他 5 个排列矩阵的组合，然后证明这 5 个矩阵是线性无关。(假设某个组合得到 $c_1P_1+...+c_5P_5=$ 零矩阵，检验矩阵的单元证明 c_1 至 c_5 都必须是 0。) 考虑行与列的总和都相等的 3×3 矩阵形成的子空间，前述的 5 个排列矩阵就是上述子空间的一组基底。

42 选择$\mathbf{R}^4$的 $\boldsymbol{x}=(x_1, x_2, x_3, x_4)$，它有 24 种排列顺序，比如$(x_2, x_1, x_3, x_4)$与$(x_4, x_3, x_1, x_2)$。这 24 个向量，包含 $\boldsymbol{x}$ 本身，生成一个子空间 $\mathbf{S}$，求出特定的向量 $\boldsymbol{x}$ 使得 $\mathbf{S}$ 的维度是(a) 0 (b) 1 (c) 3 (d) 4

43 交集与总和有 dim(**V**) + dim(**W**) = dim(**V** ∩ **W**) + dim(**V** + **W**)。从 **V** ∩ **W** 的一组基底 $\boldsymbol{u}_1, ..., \boldsymbol{u}_r$ 开始，$\boldsymbol{u}$'s 加上 $\boldsymbol{v}_1, ..., \boldsymbol{v}_s$ 形成 **V** 的一组基底，以及 $\boldsymbol{u}$'s 加上 $\boldsymbol{w}_1, ..., \boldsymbol{w}_t$ 形成 **W** 的一组基底。证明 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 与 $\boldsymbol{w}$'s 放在一起是线性无关，维度就如预期的$(r+s)+(r+t)=(r)+(r+s+t)$。

44 Mike Artin 提出一个关于问题 43 的简洁高层次证明。从 **V** 的所有输入 $\boldsymbol{v}$ 与 **W** 的 $\boldsymbol{w}$，“总和转换”得到 $\boldsymbol{v}+\boldsymbol{w}$，这些输出得到 **V** + **W**，零空间包含所有向量配对 $\boldsymbol{v}=\boldsymbol{u}, \boldsymbol{w}=-\boldsymbol{u}$，其中向量 $\boldsymbol{u}$ 在 **V** ∩ **W** 中。(则 $\boldsymbol{v}+\boldsymbol{w}=\boldsymbol{u}-\boldsymbol{u}=\mathbf{0}$)。依据计数定理，得到 dim(**V** + **W**) + dim(**V** ∩ **W**) = dim(**V**) + dim(**W**) (来自 **V** 与 **W** 的输入维度)。

输出的维度 + 零空间的维度 = 输入的维度

问题：对于秩 r 的 $m\times n$ 矩阵，这 3 个维度为何？输出=列空间。这个问题在段落 3.5 会再说明，你现在会做吗？

请尊重版权与译者的劳动成果，侵权必究!

45 在$\mathbf{R}^n$，假设维度(**V**) + 维度(**W**) > n，证明某些非零向量同时在 **V** 与 **W** 中。

46 假设 A 是 10×10 且 $A^2 = 0$(零矩阵)，所以 A 乘(A 的每个列)得到零向量。这就表示 A 的列空间在_____中。若 A 的秩 r，这些子空间的维度 $r \le 10 - r$，所以秩 $r \le 5$。

3.5 四个子空间的维度

1. 列空间 $\boldsymbol{C}(A)$与行空间 $\boldsymbol{C}(A^{\mathrm{T}})$两者都有维度 r (A 的秩)。
2. 零空间 $\boldsymbol{N}(A)$的维度是 $n-r$，左零空间 $\boldsymbol{N}(A^{\mathrm{T}})$的维度是 $m-r$。
3. 消元法可以得到 A 的行空间与零空间的基底：对 R 而言是相同的。
4. 消元法通常会改变列空间以及左零空间(但是维度不变)。
5. **秩一矩阵**：$A=\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}=$ 列乘行：$\boldsymbol{C}(A)$的基底是 $\boldsymbol{u}$，$\boldsymbol{C}(A^{\mathrm{T}})$的基底是 $\boldsymbol{v}$。

本章主要的定理会联结**秩与维度**，矩阵的秩是主元的个数，子空间的维度是基底的向量个数，我们计算主元或是基底向量的个数，A 的秩会显示全部 4 个基础子空间的维度。以下就是这些子空间，包含一个新的子空间。

两个子空间直接来自 A，其余两个来自 A^{T}：

4 个基础子空间

1. 行空间是 $\boldsymbol{C}(A^{\mathrm{T}})$，$\mathbf{R}^n$的一个子空间。
2. 列空间是 $\boldsymbol{C}(A)$，$\mathbf{R}^m$的一个子空间。
3. 零空间是 $\boldsymbol{N}(A)$，$\mathbf{R}^n$的一个子空间。
4. 左零空间(left nullspace)是 $\boldsymbol{N}(A^{\mathrm{T}})$，$\mathbf{R}^m$的一个子空间，这是我们的新空间。

在本书先看到列空间与零空间，我们已经非常了解 $\boldsymbol{C}(A)$与 $\boldsymbol{N}(A)$，现在另外两个子空间来了。行空间包含所有行的组合，A 的行空间就是 A^{T} 的列空间。

对于左零空间来说，我们求解 $A^{\mathrm{T}}\boldsymbol{y}=\mathbf{0}$——这个系统是 $n\times m$，这是 A^{T} 的零空间。当我们把方程式改写成 $\boldsymbol{y}^{\mathrm{T}}A=\mathbf{0}^{\mathrm{T}}$ 时，向量 $\boldsymbol{y}$ 会出现在 A 的左侧。矩阵 A 与 A^{T} 通常是不一样的，所以他们的列空间与零空间也会不一样，但是这些空间会以一种绝对美丽的方式结合在一起。

基础定理的第一个部分要找出这 4 个子空间的维度，有一个突出的事实：行空间与列空间具有相同的维度 r，数字 r 是矩阵的秩。另一个重要的事实与两个零空间有关。

$\boldsymbol{N}(A)$与 $\boldsymbol{N}(A^{\mathrm{T}})$的维度是 $n-r$ 与 $m-r$，用来加总全部的 n 与 m。

基础定理的第二个部分会描述如何将这 4 个子空间融合在一起(两个在$\mathbf{R}^n$，两个在$\mathbf{R}^m$)，让我们找到了解每一个 $A\boldsymbol{x}=\boldsymbol{b}$ 的“正确方法”。跟它在一起——你正在做真正的数学。

请尊重版权与译者的劳动成果，侵权必究!

R 的 4 个子空间

假设 A 简化为行阶梯形式 R，这个特殊的形状可以容易的鉴别这 4 个子空间，我们会找到每个子空间的基底以及维度。我们回过头看 A，并且检视这些子空间如何变化(其中有两个不变)，重点是 A 与 R 有相同的 4 个维度。

对于特定的 3×5 范例，从阶梯形式 R 检视这 4 个子空间。

$$\begin{matrix} m=3 \\ n=5 \\ r=2 \end{matrix} \qquad R=\begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{matrix} \text{主元行 1 与 2} \\ \\ \text{主元列 1 与 4} \end{matrix}$$

矩阵的秩 $r=2$(两个主元)，依据顺序处理这 4 个子空间。

1. R 的**行空间**的维度是 2，与秩相同。

理由： 前两行作为基底，行空间包含这三个行的所有组合，但是第三行(零行)没有引进新东西，所以行 1 与 2 生成行空间 $\boldsymbol{C}(R^{\mathrm{T}})$。

主元行 1 与 2 是无关，对于本例来说很明显，而且永远是正确的。如果我们只看主元列，可以看到 $r\times r$ 的单位矩阵，我们没办法组合这些行得到零行(除了系数为 0 的组合)，所以 r 个主元行是行空间的一组基底。

行空间的维度是秩 r，R 的非零行形成一组基底。

2. R 的**列空间**维度也是 $r=2$。

理由： 主元列是列 1 与 4 形成 $\boldsymbol{C}(R)$的一组基底，因为他们以 $r\times r$ 单位矩阵做开头，所以是无关。这些主元列的组合无法得到零列(除了系数为 0 的组合)，他们也生成了列空间。其他的每个自由列都是主元列的组合，实际上这些组合需要的是下列三个**特殊解**：

列 2 是 3(列 1)	特殊解是(–3, 1, 0, 0, 0)
列 3 是 5(列 1)	特殊解是(–5, 0, 1, 0, 0)
列 5 是 7(列 1) + 2(列 4)	解是(–7, 0, 0, –2, 1)

主元列是无关，又生成了 $\boldsymbol{C}(R)$，所以是 $\boldsymbol{C}(R)$的一组基底。

列空间的维度是秩 r，主元列形成一组基底。

3. R 的**零空间**的维度是 $n-r=5-2$，存在 $n-r=3$ 个自由变量。此处 x_2, x_3, x_5 是自由(这些列没有主元)，他们提供三个 $R\boldsymbol{x}=\boldsymbol{0}$ 的特殊解。令一个自由变量为 1，求解 x_1 与 x_4。

$$\boldsymbol{s}_2=\begin{bmatrix}-3\\1\\0\\0\\0\end{bmatrix}\quad \boldsymbol{s}_3=\begin{bmatrix}-5\\0\\1\\0\\0\end{bmatrix}\quad \boldsymbol{s}_5=\begin{bmatrix}-7\\0\\0\\-2\\1\end{bmatrix}$$

$R\boldsymbol{x}=\boldsymbol{0}$的完整解：
$\boldsymbol{x}=x_2\boldsymbol{s}_2+x_3\boldsymbol{s}_3+x_5\boldsymbol{s}_5$
零空间的维度是3

理由：这是对于每个自由变量的特殊解。如果有 n 个变量及 r 个主元，可以得到 $n-r$ 个自由变量与特殊解。这几个特殊解是无关，因为他们在行 2, 3, 5 包含单位矩阵，所以 $\boldsymbol{N}(R)$的维度是 $n-r$。

零空间的维度是 $n-r$，特殊解形成一组基底。

4. R^{T} 的**零空间**(R 的左零空间)的维度是 $m-r=3-2$。

理由：方程式 $R^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 在寻找 R^{T} 的列(R 的行)的组合得到零向量，方程式 $R^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 或者方程式 $\boldsymbol{y}^{\mathrm{T}}R=\boldsymbol{0}^{\mathrm{T}}$:

左零空间行的组合为零

$$\begin{array}{r} y_1\begin{bmatrix}1, & 3, & 5, & 0, & 7\end{bmatrix}\\ +y_2\begin{bmatrix}0, & 0, & 0, & 1, & 2\end{bmatrix}\\ +y_3\begin{bmatrix}0, & 0, & 0, & 0, & 0\end{bmatrix}\\ \hline \begin{bmatrix}0, & 0, & 0, & 0, & 0\end{bmatrix}\end{array}\qquad (1)$$

解 y_1, y_2, y_3 非常清楚，我们需要 $y_1=0$ 与 $y_2=0$，变量 y_3 是自由(可以是任意数)，**$\boldsymbol{R}^{\mathrm{T}}$ 的零空间包含所有的向量 $\boldsymbol{y}=(0, 0, y_3)$**。

所有的情形下，R 的尾端会有 $m-r$ 个零行，这些 $m-r$ 个行的每一个组合都是零，这些行是 R 中可能得到零组合的行，因为主元行是线性无关，所以左零空间中的 $\boldsymbol{y}$ 有 $y_1=0, \ldots, y_r=0$。

若 A 是 $m\times n$ 矩阵且秩为 r，它的左零空间的维度是 $m-r$。

为什么叫做“左零空间”？理由是 $R^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 可以转置成为 $\boldsymbol{y}^{\mathrm{T}}R=\boldsymbol{0}^{\mathrm{T}}$。现在 $\boldsymbol{y}^{\mathrm{T}}$ 是 R 的“左侧”的行向量，你在方程式(1)中看到这些 y's 乘上行。这个子空间是第四个介绍的，有些线性代数的书会删除它—但是这样会错失整个主题的美感。

在$\mathbf{R}^n$，行空间与零空间的维度分别是 r 与 $n-r$。(加起来得到 n)。
在$\mathbf{R}^m$，列空间与左零空间的维度分别是 r 与 $m-r$。(总和是 m)。

请尊重版权与译者的劳动成果，侵权必究!

A 的 4 个子空间

我们还有一个工作要做，A 的子空间的维度与 R 的维度完全相同，接下来的工作要解释为什么？现在 A 是可以简化得到 $R=\text{rref}(A)$的任意矩阵。

A 简化得到 R $\quad A=\begin{bmatrix}1&3&5&0&7\\0&0&0&1&2\\1&3&5&1&9\end{bmatrix}$ $\quad$ 注意 $\boldsymbol{C}(A)\neq\boldsymbol{C}(R)$ (2)

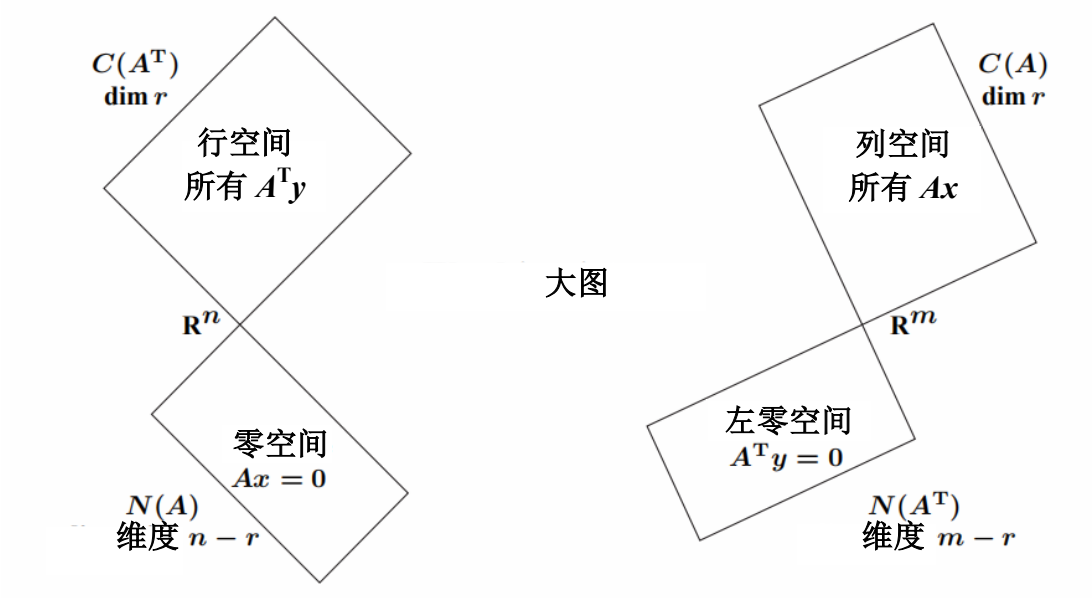


图 3.5：4 个基础子空间的维度(R 与 A)。

1　A 与 R 有相同的行空间。相同的维度 r 与相同的基底。

理由： A 的每个行是 R 的行的组合，R 的每个行也是 A 的行的组合。消元法会改变行，但不会改变行空间。

由于 A 与 R 有相同的行空间，我们可以选择 R 的前 r 个行作为一组基底，或是我们也可以选择原始 A 的 r 个合适的行，他们可能不是 A 的前 r 个行，因为 A 的前 r 个行可能是相关。A 适当的 r 个行最终是 R 的那些主元行。

2　A 的列空间的维度是 r，列的秩等于行的秩。

秩定理：无关列的个数 = 无关行的个数

错误理由： “A 与 R 有相同的列空间。”这是错误的。R 的列通常是以 0 结束，A 的列不常以 0 结束，则 $\boldsymbol{C}(A)\neq\boldsymbol{C}(R)$。

正确理由：相同的列组合对 A 与 R 都是零(或非零)。在 A 中相关 $\Leftrightarrow$ 在 R 中相关。另一种说法：只有当 $R\boldsymbol{x}=\boldsymbol{0}$ 时，才有 $A\boldsymbol{x}=\boldsymbol{0}$。他们的列空间不相同，但是维度相同——等于 r。

结论：A 的 r 个主元列是 A 的列空间 $C(A)$的一组基底。

3　A 与 R 有相同的零空间。相同的维度 $n-r$ 以及相同的基底。

理由： 消元法不会改变解，特殊解是零空间(我们已知)的一组基底，存在 $n-r$ 个自由变量，所以零空间的维度就是 $n-r$。这就是“**计数定理**(counting theorem)”：$r+(n-r)=n$。

(列空间的维度) + (零空间的维度) = $\mathbf{R}^n$的维度。

4　A 的左零空间 (A^{T} 的零空间) 的维度是 $m-r$。

理由： A^{T} 与 A 两者的本质相同，当我们知道每一个 A 的维度，我们也会知道 A^{T} 的维度，它的列空间已经证明具有维度 r。由于 A^{T} 是 $n\times m$，“整个空间”现在是 $\mathbf{R}^m$。A 的计数规则是 $r+(n-r)=n$，A^{T} 的计数规则就是 $r+(m-r)=m$。我们现在有了大定理的全部细节：

线性代数的基础定理，第一部分

列空间与行空间都有维度 r
零空间有维度 $n-r$ 与 $m-r$

专注在向量的空间，而不是个别的数字或是向量，我们得到一些清楚的规则。你很快视为理所当然——最终他们开始明显。但是如果你写下一个 11×17 的矩阵，其中有 187 个非零单元，我不觉得大部分人会了解为什么这些事实是真实的：

两个关键事实　　$\boldsymbol{C}(A)$的维度 $=\boldsymbol{C}(A^{\mathrm{T}})$的维度 $=A$ 的秩

$\boldsymbol{C}(A)$的维度 $+\boldsymbol{N}(A^{\mathrm{T}})$的维度 $=17$

范例 1　$A=[1\quad 2\quad 3]$的 $m=1$，$n=3$，秩 $r=1$。

行空间是$\mathbf{R}^3$的一条直线，零空间是平面 $A\boldsymbol{x}=x_1+2x_2+3x_3=0$，这个平面的维度是 2(=3−1)，维度相加是 1 + 2 = 3。

1×3 矩阵的列在$\mathbf{R}^1$中，列空间就是整个$\mathbf{R}^1$。左零空间只包含零向量，$A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 的唯一解是 $\boldsymbol{y}=\boldsymbol{0}$，没有其他[1　2　3]的倍数得到零行。于是 $\boldsymbol{N}(A^{\mathrm{T}})$是 $\mathbf{Z}$，零空间的维度是 0(=$m-r$)。在$\mathbf{R}^m$中，$\boldsymbol{C}(A)$与 $\boldsymbol{N}(A)$的维度相加得到 1 + 0 = 1。

请尊重版权与译者的劳动成果，侵权必究!

范例 2 $A=\begin{bmatrix}1&2&3\\2&4&6\end{bmatrix}$的 $m=2$，$n=3$，秩 $r=1$。

行空间是穿过(1, 2, 3)的同样直线，零空间必须是同样的平面 $A\boldsymbol{x}=x_1+2x_2+3x_3=0$，直线与平面的维度相加仍然是 $1+2=3$。

所有的列都是第一列(1, 2)的倍数。2 乘第一行减去第二行等于零，因此 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 有一个解是 $\boldsymbol{y}=(2,-1)$。列空间与左零空间是$\mathbf{R}^2$的垂直线，维度是 $1+1=2$。

$$\text{列空间}\ =\ \text{穿过}\begin{bmatrix}1\\2\end{bmatrix}\text{的直线}\qquad \text{左零空间}\ =\ \text{穿过}\begin{bmatrix}2\\-1\end{bmatrix}\text{的直线}$$

若 A 有三个相等的行，它的秩是_____？左零空间中的两个 $\boldsymbol{y}$ 是什么？

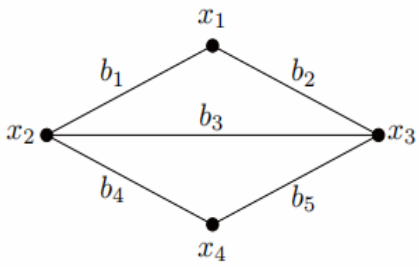
左零空间中的 y's 与行的组合得到零行

范例 3 构成方程式后，你已经接近完成三章的学习，但是无法永远继续。此处是一个较好的 5 个方程式的范例(图 3.6 的每个边对应一个方程式)。这 5 个方程式有 4 个未知数(每个节点对应一个未知数)，$A\boldsymbol{x}=\boldsymbol{b}$ 的矩阵是一个入射(incidence)矩阵，矩阵 A 的每一行有一个 1 与一个–1。

差矩阵 $A\boldsymbol{x}=\boldsymbol{b}$
在节点1, 2, 3, 4之间
存在横跨边1, 2, 3, 4, 5

$$\begin{aligned}
-x_1+x_2 &= b_1\\
-x_1+x_3 &= b_2\\
-x_2+x_3 &= b_3\\
-x_2+x_4 &= b_4\\
-x_3+x_4 &= b_5
\end{aligned}\tag{3}$$

如果你了解矩阵的 4 个基础子空间(A 与 A^{T} 的列空间与零空间)，你已经捕捉到线性代数的中心概念。



$$A=\begin{bmatrix}-1&1&&\\-1&&1&\\&-1&1&\\&-1&&1\\&&-1&1\end{bmatrix}\qquad\begin{matrix}\text{边}\\1\\2\\3\\4\\5\end{matrix}$$

图 3.6：一个具有 5 个边 4 个节点的“图”，A 是 5×4 的入射矩阵

零空间 $N(A)$ 令 $\boldsymbol{b}=\boldsymbol{0}$ 求出零空间，第一个方程式得到 $x_1=x_2$，第二个方程式是 $x_3=x_1$，方程式 4 是 $x_2=x_4$。这 4 个未知数 x_1, x_2, x_3, x_4 有相同的值 c，则向量 $\boldsymbol{x}=(c, c, c, c)$填满 A 的零空间。

零空间是$\mathbf{R}^4$的一条直线，特殊解 $\boldsymbol{x}=(1, 1, 1, 1)$是 $N(A)$的一组基底，$N(A)$的维度是 1(基底只有一个向量)。由于 $n-r=4-3=1$，A 的秩必须是 3。我们知道所有 4 个子空间的维度。

列空间 $C(A)$　必定有 $r=3$ 个无关列。快速的方法是检视前三个列，系统化的方法是求出 $R=\text{rref}(A)$。

$$A \text{ 的列 } 1,2,3 \quad \begin{matrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{matrix} \qquad R=\text{简化行阶梯形式} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

从 R 我们再度看到特殊解 $\boldsymbol{x}=(1,1,1,1)$，前三列是基本，第四列是自由。要得到 $\boldsymbol{C}(A)$的一组基底(不是 $\boldsymbol{C}(R)$的)，我们回到 A 的列 1, 2, 3，列空间的维度 $r=3$。

行空间 $C(A^{\mathrm{T}})$　维度再次必须是 3，但是 A 的前三行不是无关：行 3 = 行 2 − 行 1，所以行 3 被消元法变成零，行 3 与行 4 交换。前三个无关行是行 1, 2, 4，这三个行是行空间的一组基底(一组可能的基底)。

我注意到图中的边 1, 2, 3 形成一个回圈：相关行 1, 2, 3。边 1, 2, 4 形成一个树(tree)，树没有回路，无关行 1, 2, 4。

左零空间 $N(A^{\mathrm{T}})$　现在我们求解 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$，找到行的组合得到零。我们已经注意到行 3 = 行 2 − 行 1，所以一个解是 $\boldsymbol{y}=(1,-1,1,0,0)$。我会说：这个 $\boldsymbol{y}$ 是来自图中上方的回圈；另一个 $\boldsymbol{y}$ 来自下方的回圈且 $\boldsymbol{y}=(0,0,-1,1,1)$: 行 3 = 行 4 + 行 5。这两个 y's 是无关，他们是 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 的解，且 $\boldsymbol{N}(A^{\mathrm{T}})$的维度是 $m-r=5-3=2$，所以我们找到了左零空间的一组基底。

你可能会问“回圈”与“树”如何进入问题当中，这件事情没有必要发生。我们可以利用消元法求解 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$，则 4×5 矩阵 A^{T} 会有三个主元列 1, 2, 4 以及两个自由列 3, 5。存在两个特殊解且 A^{T} 的零空间的维度是 2：$m-r=5-3=2$。但是回圈与树可以用一种美丽的方法鉴别相关行与无关行，我们在段落 10.1 会针对每一个好像 A 的入射矩阵使用这个方法。

方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 在 4 个节点上得到“电压(voltage)”x_1, x_2, x_3, x_4；方程式 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 在 5 个边得到“电流(current)”y_1, y_2, y_3, y_4, y_5。这两个方程式就是“基尔霍夫电压定律与基尔霍夫电流定律”，这些是用在电网络的术语。但是隐藏在术语背后的是可以应用在所有的工程、科学、经济以及商务。

图形是离散(discrete)应用数学中最重要的模型，你可以在任何地方看到图形：比如说道路、管线、血液流、大脑、网络、国家或是世界经济，我们可以了解他们的矩阵 A 与 A^{T}。

请尊重版权与译者的劳动成果，侵权必究!

秩一矩阵(复习)

假设每个行都是第一行的倍数，典型的例子：

$$\begin{bmatrix} 2 & 3 & 7 & 8 \\ 2a & 3a & 7a & 8a \\ 2b & 3b & 7b & 8b \end{bmatrix} = \begin{bmatrix} 1 \\ a \\ b \end{bmatrix} \begin{bmatrix} 2 & 3 & 7 & 8 \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$$

左侧的矩阵有三行，但是它的行空间的维度只有 1。行向量 $\boldsymbol{v}^{\mathrm{T}} = [2 \quad 3 \quad 7 \quad 8]$是行空间的一组基底，行的秩是 1。

现在检视列，“列的秩等于行的秩等于 1。”矩阵所有的列必须都是某一列的倍数，你有没有看出这个线性代数的关键规则是否正确？列向量 $\boldsymbol{u} = (1, a, b)$分别乘了 2, 3, 7, 8，这个非零列向量是列空间的一组基底，列秩也是 1。

每一个秩一的矩阵是一列乘一行 $\boldsymbol{A} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 。

秩二矩阵 = 秩一加秩一

有一个矩阵 A 的秩 $r = 2$，我们无法从 A 立即看出 r，所以我们使用行运算把矩阵简化成 $R = \mathrm{rref}(A)$，有一些消元矩阵 E 把 A 简化成 $EA = R$，则逆矩阵 $C = E^{-1}$ 把 R 回复至 $A = CR$。

你已经知道重点：$\boldsymbol{R}$ 与 $\boldsymbol{A}$ 有相同的行空间。

秩二 $$A = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 1 & 7 \\ 4 & 2 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} = CR \tag{4}$$

R 的行空间明显有两个基底向量 $\boldsymbol{v}_1^{\mathrm{T}} = [1 \quad 0 \quad 3]$与 $\boldsymbol{v}_2^{\mathrm{T}} = [0 \quad 1 \quad 4]$，所以 A 的行空间也有同样的基底：行秩= 2。C 乘 R 可以得到 A 的行 3 是 $4\boldsymbol{v}_1^{\mathrm{T}} + 2\boldsymbol{v}_2^{\mathrm{T}}$。

现在检视列。R 的主元列明显是(1, 0, 0)与(0, 1, 0)，则 A 的主元列也是在列 1 与 2：$\boldsymbol{u}_1 = (1, 1, 4)$与 $\boldsymbol{u}_2 = (0, 1, 2)$。

当你将字母放在列与行时，你会看到**秩二 = 秩一 + 秩一**。

矩阵 $\boldsymbol{A}$ 的秩是二 $$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \text{零行} \end{bmatrix} = \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} = \text{秩一} + \text{秩一}$$

你有看到最后一个步骤吗？我是对矩阵使用列乘行，对于这个问题非常完美。每个秩 r 的矩阵，是 r 个秩一矩阵的总和：A 的主元列乘 R 的非零行，行[0 0 0]简单的消失了。

主元列 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$ 是列空间的一组基底，你懂的！

主要观念的复习

1. R 的 r 个主元行是 R 与 A 的行空间(相同空间)的一组基底。
2. A 的 r 个主元列是 A 的列空间 $\boldsymbol{C}(A)$的一组基底。
3. $n-r$ 个特殊解是 A 与 R 的零空间(相同空间)的一组基底。
4. 若 $EA=R$，E 最后的 $m-r$ 个行是 A 的左零空间的一组基底。

关于 4 个子空间的注释：这个基础定理看起来很像纯数学，但是它有非常重要的应用。我最喜欢的是在第十章介绍的网络(我通常在下一次上课时直接讲 10.1)，在左零空间中的 $\boldsymbol{y}$ 的方程式就是 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$:

流进一个节点等于流出，基尔霍夫电流定律就是“平衡方程式”

这个必须是应用数学中最重要的方程式，所有科学、工程与经济的模型都与平衡有关--包含力、热流、电荷、动量或是金钱的平衡。平衡方程式，加上胡克定律、欧姆定律或是与位能到流量相关的定律，给出了应用数学的明显框架。

我的“计算科学与工程”这本教科书开发了这样的框架，连同演绎法去求解方程式：有限差，有限元素，频谱方法，回归方法以及多网格法。

已解范例

3.5A　把 4 个 1 放进一个 5×6 的零矩阵中，让行空间的维度尽可能最小。描述所有使得列空间的维度尽可能最小的方法。描述所有使得零空间的维度尽可能最小的方法。如何使得 4 个子空间的维度总和最小?

解　如果 4 个 1 放在同一行，或者同一列，矩阵的秩是 1。他们也可以放在两行以及两列(所以 $a_{ii}=a_{ij}=a_{ji}=a_{jj}=1$)。由于列空间与行空间的维度相等，已经回答了前两个问题：维度 1。

当矩阵的秩是 $r=4$，零空间最小可能的维度是 $6-4=2$。要达成秩 4，4 个 1 必须放在 4 个不同行及 4 个不同列。

关于这个总和 $r+(n-r)+r+(m-r)=n+m$, 你是无计可施的。无论 4 个 1 怎么放置，总和永远是 $6+5=11$。纵使矩阵里面没有任何的 1，总和还是 11。

如果 A 的其他单元都是 2 而不是 0，这些答案会怎么改变?

请尊重版权与译者的劳动成果，侵权必究!

3.5B 事实：AB 所有的行都是 B 的行的组合，所以 B 的行空间会包含 AB 的行空间(有可能相等)，秩(AB) ≤ 秩(B)。

AB 所有的列是 A 的列的组合，所以 A 的列空间会包含 AB 的列空间(有可能相等)，秩(AB) ≤ 秩(A)。

如果我们左乘一个可逆矩阵，秩不会改变。秩不会掉落是因为当我们左乘逆矩阵时，秩不会跳回来。

问题集 3.5

1 (a) 若一个 7×9 矩阵的秩是 5，它的 4 个子空间的维度是多少？这 4 个维度总和是多少？

(b) 若一个 3×4 矩阵的秩是 3，则它的列空间与左零空间为何？

2 求出 A 与 B 的 4 个子空间的基底与维度：

$$A=\begin{bmatrix}1&2&4\\2&4&8\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1&2&4\\2&5&8\end{bmatrix}$$

3 求出 A 的 4 个子空间的基底：

$$A=\begin{bmatrix}0&1&2&3&4\\0&1&2&4&6\\0&0&0&1&2\end{bmatrix}=\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}0&1&2&3&4\\0&0&0&1&2\\0&0&0&0&0\end{bmatrix}$$

4 写出满足下列要求性质的矩阵或是说明为什么不可能：

(a) 列空间包含(1, 1, 0)与(0, 0, 1)，行空间包含(1, 2)与(2, 5)。

(b) 列空间的基底(1, 1, 3)，零空间的基底(3, 1, 1)。

(c) 零空间的维度 = 1 + 左零空间的维度。

(d) 零空间包含(1, 3)，列空间包含(3, 1)。

(e) 行空间 = 列空间，零空间 ≠ 左零空间。

5 若 **V** 是由(1, 1, 1)与(2, 1, 0)生成的子空间，求矩阵 A 使得 **V** 是它的行空间。求矩阵 B 使得 **V** 是它的零空间。求出乘积 AB。

6 不使用消元法，求出下列矩阵的 4 个子空间的维度与基底：

$$A=\begin{bmatrix}0&3&3&3\\0&0&0&0\\0&1&0&1\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1\\4\\5\end{bmatrix}$$

7 假设 3×3 矩阵 A 可逆，写出 A 的 4 个子空间的基底。令 3×6 矩阵 $B=[A \quad A]$，求出 B 的 4 个子空间的基底。(**Z** 的基底是空集合。)

8 若 I 是 3×3 的单位矩阵，0 是 3×2 的零矩阵，求 A, B, C 的 4 个子空间的维度：

$$A=\begin{bmatrix} I & 0 \end{bmatrix} \text{ 与 } B=\begin{bmatrix} I & I \\ 0^{\mathrm{T}} & 0^{\mathrm{T}} \end{bmatrix} \text{ 与 } C=\begin{bmatrix} 0 \end{bmatrix}$$

9 下列矩阵的大小都不相同，哪些子空间却是相同？

$$(a)\begin{bmatrix} A \end{bmatrix} \text{ 与 } \begin{bmatrix} A \\ A \end{bmatrix} \quad (b)\begin{bmatrix} A \\ A \end{bmatrix} \text{ 与 } \begin{bmatrix} A & A \\ A & A \end{bmatrix}$$

证明这些矩阵有相同的秩 r。

10 如果 3×3 矩阵的单元在 0 至 1 之间随意选取，4 个子空间最有可能的维度是多少？如果随机矩阵是 3×5？

11 (重要) A 是秩 r 的 $m\times n$ 矩阵，假设存在右侧的 $\boldsymbol{b}$ 使得 $A\boldsymbol{x}=\boldsymbol{b}$ 无解。

(a) 在 r，m，n 之间的不等号(< 或 ≤)必须如何安排才正确？

(b) 你如何知道 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 有除了 $\boldsymbol{y}=\boldsymbol{0}$ 的解？

12 以(1, 0, 1)与(1, 2, 0)作为基底的行空间的矩阵为何？它的列空间为何？为什么这组基底不能成为行空间与零空间的基底?

13 是非题 (给出理由或是反例)：

(a) 若 $m=n$ 则 A 的行空间等于 A 的列空间。

(b) 矩阵 A 与 $-A$ 共享相同的 4 个子空间。

(c) 若 A 与 B 共享相同的 4 个子空间，则 A 是 B 的倍数。

14 不计算 A，求出它的 4 个基础子空间：

$$A=\begin{bmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ 9 & 8 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

15 如果你交换 A 的前两行，4 个子空间中哪些保持不变？若 $\boldsymbol{v}$ = (1, 2, 3, 4)在 A 的左零空间中，写出一个在行交换后新矩阵的左零空间中的向量。

16 说明为什么 $\boldsymbol{v}$ = (1, 0, −1)不可能既是 A 的一个行又在零空间中?

17 描述下列$\mathbf{R}^3$中的矩阵的 4 个子空间：

$$A=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ 与 } I+A=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

18 (左零空间) 引入额外的列 $\boldsymbol{b}$，简化 A 至阶梯形式：

$$\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 4 & 5 & 6 & b_2 \\ 7 & 8 & 9 & b_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 0 & -3 & -6 & b_2 - 4b_1 \\ 0 & 0 & 0 & b_3 - 2b_2 + b_1 \end{bmatrix}$$

A 的行的组合得到零行，这是什么样的组合？(检视右侧的 $b_3 - 2b_2 + b_1$)，哪些向量在 A^{T} 的零空间中？哪些向量在 A 的零空间中？

19 依据问题 18 的方法，简化 A 至阶梯形式然后检视零行，$\boldsymbol{b}$ 列会告诉我们你用了什么样的行组合：

(a) $\begin{bmatrix} 1 & 2 & b_1 \\ 3 & 4 & b_2 \\ 4 & 6 & b_3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & b_1 \\ 2 & 3 & b_2 \\ 2 & 4 & b_3 \\ 2 & 5 & b_4 \end{bmatrix}$

从消元后的 $\boldsymbol{b}$ 列，会得到在左零空间的 $m - r$ 个基底向量，这些 $\boldsymbol{y}$'s 是行的组合，这些行得到阶梯形式的零行。

20 (a) 检验 $A\boldsymbol{x} = \boldsymbol{0}$ 的解垂直于 A 的行：

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} = ER$$

(b) $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$ 有几个无关的解？为什么 $\boldsymbol{y}^{\mathrm{T}} = E^{-1}$ 的行 3？

21 假设 A 是两个秩一矩阵的总和：$A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$。

(a) 哪些向量生成 A 的列空间？

(b) 哪些向量生成 A 的行空间？

(c) 如果_____或如果_____，则秩小于 2。

(d) 若 $\boldsymbol{u} = \boldsymbol{z} = (1, 0, 0)$且 $\boldsymbol{v} = \boldsymbol{w} = (0, 0, 1)$，计算 A 及它的秩。

22 创建 $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{w}\boldsymbol{z}^{\mathrm{T}}$，$A$ 的列空间有基底(1, 2, 4)与(2, 2, 1)，行空间有基底(1, 0)与(1, 1)。把 A 写成(3×2)乘(2×2)。

23 不计算矩阵乘积，求 A 的行空间与列空间的基底：

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} 3 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}$$

从这些形状，你如何知道 A 不可能是可逆？

24 (重要) 当 $\boldsymbol{d}$ 在 4 个子空间的哪个空间时，$A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{d}$ 有解？当_____只包含零向量时，有唯一解 $\boldsymbol{y}$。

25 是非题 (理由或反例)：

(a) A 与 A^{T} 有相同的主元个数。

(b) A 与 A^{T} 有相同的左零空间。

(c) 如果行空间等于列空间则，$A^{\mathrm{T}}=A$。

(d) 如果 $A^{\mathrm{T}}=-A$，则 A 的行空间等于列空间。

26 给定 a, b, c，其中 $a\neq 0$，你如何选择 d 使得 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 有秩 1。

27 求出 8×8 棋盘格矩阵 B 以及西洋棋矩阵 C 的秩：

$$B=\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \qquad C=\begin{bmatrix} r & n & b & q & k & b & n & r \\ p & p & p & p & p & p & p & p \\ & & & 4\text{个} & \text{零行} & & & \\ p & p & p & p & p & p & p & p \\ r & n & b & q & k & b & n & r \end{bmatrix}$$

r, n, b, q, k, p 是完全不同的数字，求出 B 与 C 的行空间与左零空间的基底。

挑战问题：求出 C 的零空间的一组基底。

28 完成井字游戏矩阵(A 有 5 个 1 与 4 个 0)之后的 rank(A) = 2，但是两边都没有走出致胜棋步，可能吗？【译者自己下了一盘才明白题目的意思】

29 若 $A=\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 是秩 1 的 2×2 矩阵，重画图 3.5 更清楚显示这 4 个子空间。若 B 有相同的 4 个子空间，B 与 A 的确切关系为何？

30 **M** 是 3×3 矩阵的空间，X 是 **M** 的任意矩阵，乘积 AX：

$$A=\begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad \text{注意：} A\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(a) 哪些矩阵 X 使得 AX=零矩阵。【求 A 的零空间，就是下面的(a)】

(b) 对矩阵 X，哪些矩阵具有 AX 的形式？【求 A 的列空间，就是下面的(b)】

(a) 求 AX 的零空间。 (b) 求列空间。**M** 的这两个子空间的维度为何？为什么维度加起来是 $(n-r)+r=9$。

31 假设 $m\times n$ 的矩阵 A 与 B 有相同的 4 个子空间，如果他们都已经是行简化阶梯形式，证明 $F=G$：

$$A=\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} \qquad B=\begin{bmatrix} I & G \\ 0 & 0 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

第四章

正交性质

4.1 四个子空间的正交性质

1 正交向量有 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$，则 $\|\boldsymbol{v}\|^2+\|\boldsymbol{w}\|^2=\|\boldsymbol{v}+\boldsymbol{w}\|^2=\|\boldsymbol{v}-\boldsymbol{w}\|^2$。

2 若子空间 $\boldsymbol{V}$ 中的每个 $\boldsymbol{v}$ 与子空间 $\boldsymbol{W}$ 中的每个 $\boldsymbol{w}$ 都存在 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$，则子空间 $\boldsymbol{V}$ 与子空间 $\boldsymbol{W}$ 正交。

3 A 的**行空间与零空间正交**，**列空间与** $\boldsymbol{N}(\boldsymbol{A}^{\mathrm{T}})$**正交**。

4 一组维度配对相加得到 $r+(n-r)=n$，另一组相加是 $r+(m-r)=m$。

5 行空间与零空间是“正交补充”：$\mathbf{R}^n$的每个 $\boldsymbol{x}$ 分成 $\boldsymbol{x}_{\mathrm{row}}+\boldsymbol{x}_{\mathrm{null}}$。

6 假设空间 $\boldsymbol{S}$ 的维度是 d，则 $\boldsymbol{S}$ 的每组基底包含 d 个向量。

7 若空间 $\boldsymbol{S}$ 的 d 个向量无关，这些向量生成 $\boldsymbol{S}$。若 d 个向量生成 $\boldsymbol{S}$，则这些向量无关。

两个向量的点积为零，则两个向量正交：$\boldsymbol{v}\cdot\boldsymbol{w}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$。本章转移到**正交子空间**与**正交基底**以及**正交矩阵**。在两个子空间中的向量，在一组基底的向量以及 Q 中的列向量，前述所有的向量配对都是正交。想起直角三角形有 $a^2+b^2=c^2$，其中三角形的边是 $\boldsymbol{v}$ 与 $\boldsymbol{w}$。

正交向量 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$ 且 $\|\boldsymbol{v}\|^2+\|\boldsymbol{w}\|^2=\|\boldsymbol{v}+\boldsymbol{w}\|^2$

当 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{v}=0$，直角边$(\boldsymbol{v}+\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{v}+\boldsymbol{w})$等于 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$。

第三章的子空间在阐述 $A\boldsymbol{x}=\boldsymbol{b}$，现在我们需要列空间与零空间，聚光转到 A^{T}，揭开另外两个子空间，这 4 个子空间显示了矩阵实际在做什么。

矩阵乘向量：A 乘 $\boldsymbol{x}$。第一层次只有数字，第二层次 $A\boldsymbol{x}$ 是列向量的组合，第三层次展示了子空间。但是我不认为你已经看到全貌，直到你研读了图 4.2。

子空间适配在一起显示 A 乘 $\boldsymbol{x}$ 的潜在真实性，两个子空间之间的 90°角是新的主题--我们现在可以说直角代表什么意义。

行空间与零空间垂直，A 的每个行与 $A\boldsymbol{x}=\boldsymbol{0}$ 的每个解垂直，得到图形左侧的 90°角。这个子空间的垂直性是线性代数基础定理的第二部分。

列空间与 A^{T} 的零空间垂直，当 $\boldsymbol{b}$ 在列空间之外--当我们想要求解 $A\boldsymbol{x}=\boldsymbol{b}$ 却做不到时--此时 A^{T} 的零空间就会显示出独特的优势，它包含“最小二乘”解的误差 $\boldsymbol{e}=\boldsymbol{b}-A\boldsymbol{x}$，最小二乘是线性代数在本章中的关键应用。

基础定理的第一部分给定子空间的维度，行与列空间有相同的维度 r (他们吸取相同大小)，两个零空间有剩下的维度 $n-r$ 与 $m-r$。现在我们证明**行空间与零空间是 $\mathbf{R}^n$ 中的正交子空间**。

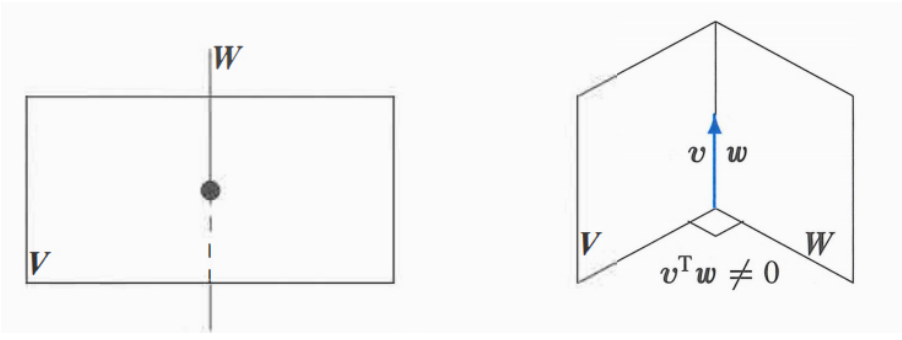
定义　向量空间的两个子空间 $\boldsymbol{V}$ 与 $\boldsymbol{W}$，如果 $\boldsymbol{V}$ 中的每个向量 $\boldsymbol{v}$ 都与 $\boldsymbol{W}$ 中的每个向量 $\boldsymbol{w}$ 垂直，则 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ **正交**：

正交子空间	$\boldsymbol{V}$ 中的所有 $\boldsymbol{v}$ 与 $\boldsymbol{W}$ 中的所有 $\boldsymbol{w}$ 都有 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=\boldsymbol{0}$。

范例 1　你房间的地板(延申至无限)是一个子空间 $\boldsymbol{V}$，两面墙的交线是子空间 $\boldsymbol{W}$ (1 维)，这两个子空间正交。墙与墙的交线上的向量与地板的每个向量垂直。

范例 2　两面墙看起来是垂直，但是这两个子空间没有正交！交线同时在 $\boldsymbol{V}$ 与 $\boldsymbol{W}$--这条直线与本身并没有垂直。两个平面(在 $\mathbf{R}^3$ 中，维度是 2 与 2)不可能是正交子空间。

当一个向量同时在两个正交子空间中，它必须是零，它垂直于本身。它是 $\boldsymbol{v}$ 也是 $\boldsymbol{w}$，所以 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=0$，这个只能是零向量。



正交平面 $\boldsymbol{V}$ 与直线 $\boldsymbol{W}$　　非正交平面

图 4.1：当 dim $\boldsymbol{V}$ + dim $\boldsymbol{W}$ > dim (整个空间)，不可能正交。

线性代数重要的范例来自 4 个基础子空间，零是零空间与行空间唯一的交点，此外，**$\boldsymbol{A}$ 的零空间与行空间是 90°交会**。关键的事实直接来自 $A\boldsymbol{x}=\boldsymbol{0}$：

请尊重版权与译者的劳动成果，侵权必究!

$A\boldsymbol{x}=\boldsymbol{0}$，$A$ 的零空间中的每个向量 $\boldsymbol{x}$ 与 A 的每一行垂直。**零空间 $N(A)$与行空间 $C(A^{\mathrm{T}})$是$\mathbf{R}^n$中的正交子空间**。

要了解为什么 $\boldsymbol{x}$ 与这些行正交，检视 $A\boldsymbol{x}=\boldsymbol{0}$，每个行乘 $\boldsymbol{x}$：

$$A\boldsymbol{x}=\begin{bmatrix}\text{行}\,1\\ \vdots\\ \text{行}\,m\end{bmatrix}\begin{bmatrix}\\ \boldsymbol{x}\\ \\ \end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}\qquad\begin{matrix}\leftarrow(\text{行}\,1)\cdot\boldsymbol{x}\text{是零}\\ \\ \leftarrow(\text{行}\,m)\cdot\boldsymbol{x}\text{是零}\end{matrix}\tag{1}$$

方程式 1 说明**行 1** 与 $\boldsymbol{x}$ 垂**直**，最后方程式说明**行 $\boldsymbol{m}$** 与 $\boldsymbol{x}$ 垂直。每一行与 $\boldsymbol{x}$ 的点积都是零，$\boldsymbol{x}$ 也与行的每个组合垂直。整个行空间 $\boldsymbol{C}(A^{\mathrm{T}})$与零空间 $\boldsymbol{N}(A)$是正交。

此处提供喜欢矩阵缩写的读者第二种证明，行空间的向量是行的组合 $A^{\mathrm{T}}\boldsymbol{y}$，计算 $A^{\mathrm{T}}\boldsymbol{y}$ 与零空间的 $\boldsymbol{x}$ 的点积，这些向量互相垂直：

零空间与行空间正交　　$\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})=(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}^{\mathrm{T}}\boldsymbol{y}=0$　　(2)

我们喜欢第一个证明，你可以从方程式(1)看到 A 的这些行乘 $\boldsymbol{x}$ 得到零，第二个证明展示了为什么 A 与 A^{T} 都在基础定理中。

范例 3　矩阵 A 的行与零空间中的 $\boldsymbol{x}=(1,1,-1)$垂直：

$$A\boldsymbol{x}=\begin{bmatrix}1&3&4\\5&2&7\end{bmatrix}\begin{bmatrix}1\\1\\-1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}\quad\text{得到点积}\quad\begin{matrix}1+3-4=0\\5+2-7=0\end{matrix}$$

现在回到另外两个子空间。在本例中，列空间是全部的$\mathbf{R}^2$，A^{T} 的零空间只有零向量(与所有向量正交)，A 的列空间与 A^{T} 的零空间永远是正交子空间。

A^{T} 的零空间中的每个向量 $\boldsymbol{y}$ 与 A 的每一列垂直，**左零空间 $N(A^{\mathrm{T}})$与列空间 $C(A)$在$\mathbf{R}^m$中正交**。

对 A^{T} 应用原始的证明，A^{T} 的零空间与 A^{T} 的行空间正交---A^{T} 的行空间就是 A 的列空间。**证明完毕**。

视觉化的证明：检视 $A^{\mathrm{T}}\boldsymbol{y}=0$，$A$ 的每一列乘 $\boldsymbol{y}$ 得到 0：

$$\boldsymbol{C}(A)\perp\boldsymbol{N}(A^{\mathrm{T}})\qquad A^{\mathrm{T}}\boldsymbol{y}=\begin{bmatrix}(\text{列}\,1)^{\mathrm{T}}\\ \cdots\\ (\text{列}\,m)^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\\ \boldsymbol{y}\\ \\ \end{bmatrix}=\begin{bmatrix}0\\ \cdot\\ 0\end{bmatrix}\tag{3}$$

$\boldsymbol{y}$ 与 A 的每一列的点积是零，则左零空间中的 $\boldsymbol{y}$ 与 A 的每一列垂直---也就是与整个列空间垂直。

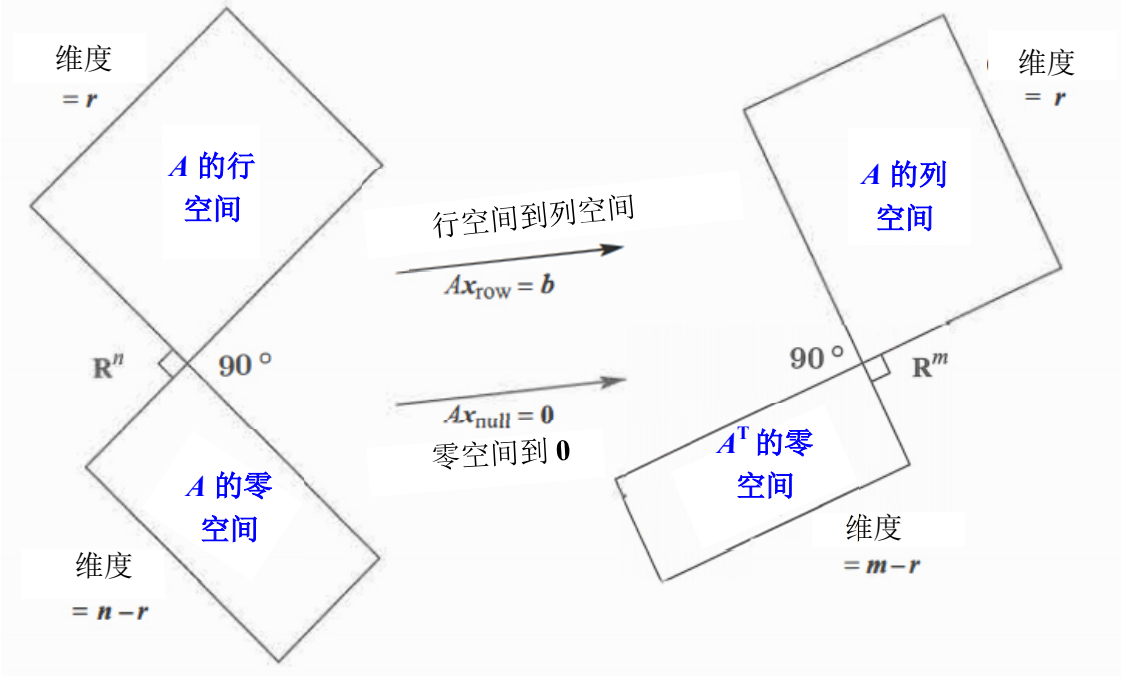


图 4.2：两对(pairs)正交子空间，维度总和是 n 与 m。**这是大图**——两个子空间在 $\mathbf{R}^n$ 中，两个子空间在 $\mathbf{R}^m$ 中。

正交补充

重要　基础子空间不止是正交(成对)而已，他们的维度也是恰当的。$\mathbf{R}^3$ 中的两条直线可能垂直，**但是这些直线不可能是 3×3 矩阵的行空间与零空间**。两条直线的维度是 1 与 1，加起来是 2，但是正确的维度 r 与 $n-r$ 加起来必须等于 3。

3×3 矩阵的基础子空间的维度有 2 与 1，或是 3 与 0。这些子空间配对不只是正交，他们还是**正交补充**。

定义　子空间 $\boldsymbol{V}$ 的**正交补充**包含每个与 $\boldsymbol{V}$ 垂直的向量，这个正交子空间写成 $\boldsymbol{V}^{\perp}$ (读成 $\boldsymbol{V}$ perp)。

基于这个定义，零空间是行空间的正交补充，每个垂直行的 $\boldsymbol{x}$ 满足 $A\boldsymbol{x} = \mathbf{0}$，并且落在零空间中。

反向也是成立的，**如果 $\boldsymbol{v}$ 与零空间正交，它必须在行空间中**，否则我们可以把 $\boldsymbol{v}$ 加入矩阵作为一个额外的行，而没有改变它的零空间。行空间会变大，破坏 $r + (n-r) = n$ 的法则。我们的结论是零空间补充 $\boldsymbol{N}(A)^{\perp}$ 确切是行空间 $\boldsymbol{C}(A^{\mathrm{T}})$。

同样的方式，左零空间与列空间在 $\mathbf{R}^m$ 中正交，他们是正交补充，他们的维度 r 与 $m-r$ 相加得到满维度 m。

请尊重版权与译者的劳动成果，侵权必究！

线性代数基础定理，第二部分

$N(A)$是行空间 $C(A^{\mathrm{T}})$的正交补充(在$\mathbf{R}^n$)

$N(A^{\mathrm{T}})$是列空间 $C(A)$的正交补充(在$\mathbf{R}^m$)

第一部分给定子空间的维度，第二部分给定他们之间的 90°角。“补充”的重点在于每一个 $\boldsymbol{x}$ 可以分成一个行空间分量 $\boldsymbol{x}_r$ 与一个零空间分量 $\boldsymbol{x}_n$。图 4.3 显示 $A\boldsymbol{x} = A\boldsymbol{x}_r + A\boldsymbol{x}_n$ 发生了什么:

零空间分量走到零：$A\boldsymbol{x}_n = \mathbf{0}$。

行空间分量走到列空间：$A\boldsymbol{x}_r = A\boldsymbol{x}$。

每个向量都走到列空间！左乘 A 不能做其他事情，除此之外：列空间中的每个向量 $\boldsymbol{b}$ 来自行空间中的一个而且是唯一的向量 $\boldsymbol{x}_r$。证明：若 $A\boldsymbol{x}_r = A\boldsymbol{x}_r'$，两者的差 $\boldsymbol{x}_r - \boldsymbol{x}_r'$ 会在零空间中，它也会在行空间中，因为 $\boldsymbol{x}_r$ 与 $\boldsymbol{x}_r'$都来自行空间。两者的差必须为零向量，这是因为零空间与行空间互相垂直，因此 $\boldsymbol{x}_r = \boldsymbol{x}_r'$。

如果我们抛开两个零空间，在 A 中隐藏一个 $r\times r$ 的可逆矩阵。**从行空间到列空间，A 是可逆**。“伪逆反(pseudoinverse)”会逆反段落 7.4 中 A 的那部分。

范例 4　每个秩 r 的矩阵有一个 $r\times r$ 的可逆子矩阵：

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ 包含子矩阵 } \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$$

其他 11 个 0 负责零空间。B 的秩也是 $r = 2$：

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix} \text{ 包含 } \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \text{ 在主元行与列}$$

当我们选择了正确的$\mathbf{R}^n$与$\mathbf{R}^m$的基底，每个矩阵都可以对角化，这个“**奇异值分解(singular value decomposition)**”在应用上已经变得非常重要。

让我重复一个清晰的事实，A 的行不可能在 A 的零空间中 (除了零行之外)，两个正交子空间中都存在的向量只有零向量。

如果一个向量 $\boldsymbol{v}$ 与本身正交，则 $\boldsymbol{v}$ 一定是零向量。

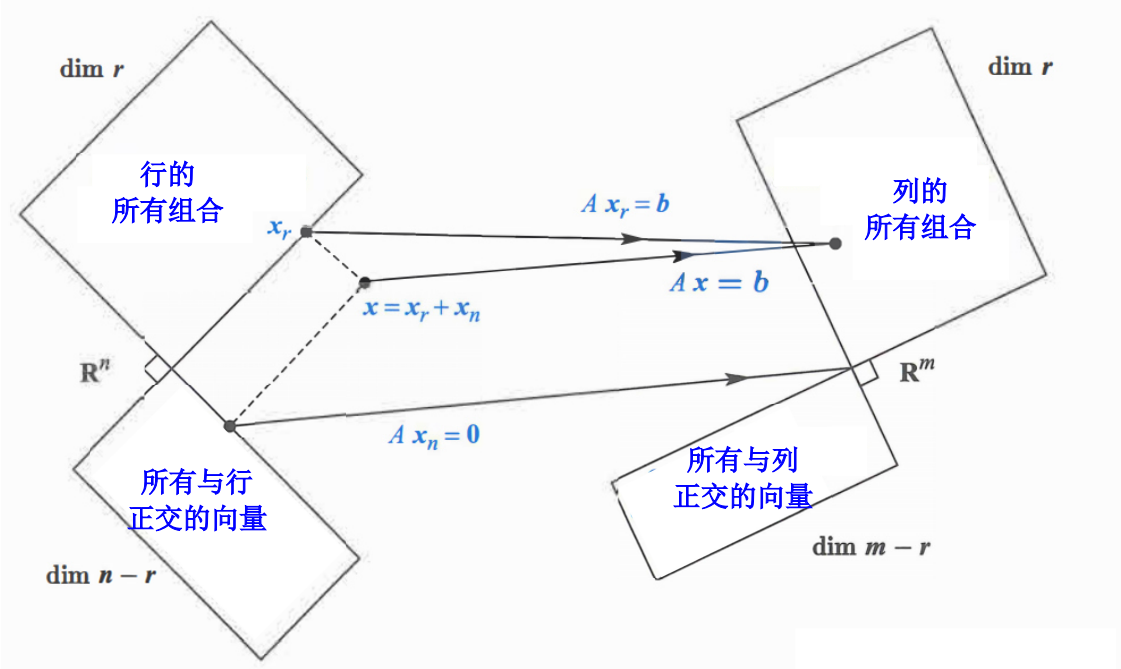


图 4.3：这是图 4.2 的更新图，显示 A 对于 $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$ 的真实作用。行空间向量 $\boldsymbol{x}_r$ 到列空间，零空间向量 $\boldsymbol{x}_n$ 到零。

画出大图

我不知道画出在图 4.2 与 4.3 的 4 个子空间的最佳方法，这张大图必须显示这些子空间的正交性质。我能够看到一个可能的方法去做这件事情，就是一条线与一个平面的交会——可能图 4.4 也显示了这些空间是无限的，比图 4.3 的矩形要更清晰。但是我该如何在$\mathbf{R}^4$画出一对 2 维的子空间，去展示他们彼此之间是正交？欢迎提供好点子。

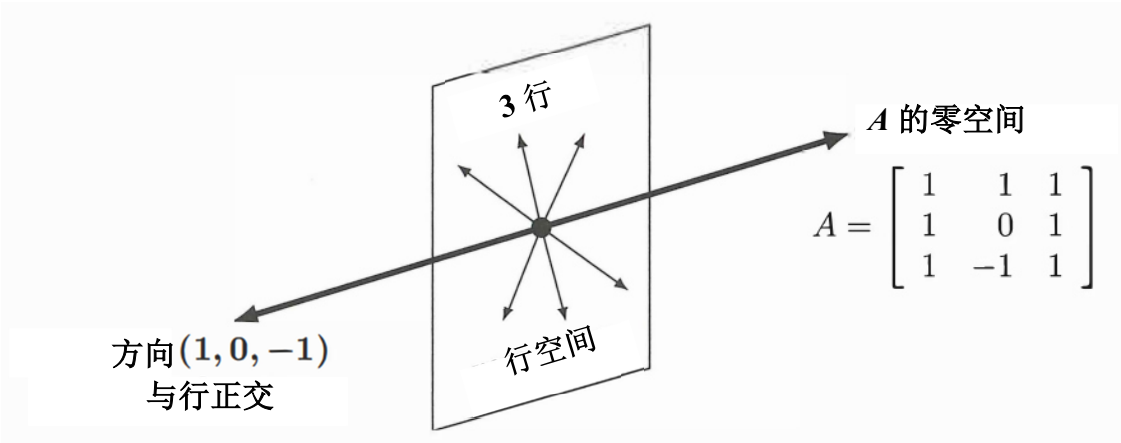


图 4.4：A 的行空间 = 平面，零空间 = 正交直线，维度 2 + 1= 3。

请尊重版权与译者的劳动成果，侵权必究！

从子空间组合基底

接下来是一些关于基底的有价值事实，他们一直储存到现在--当我准备好使用他们。一个星期以后你对于基底是什么(线性无关的向量生成空间)，就会有一个清晰的观念。正常情形下要检验两个性质，当计数是正确的，一个性质就可以推论至另一个性质：

> $\mathbf{R}^n$中任何 n 个无关向量必然生成$\mathbf{R}^n$，所以他们是一组基底。
> 生成$\mathbf{R}^n$的任何 n 个向量必然无关，所以他们是一组基底。

从向量的正确个数开始，基底的一个性质产生另一个性质。对于任意的向量空间来说这是真实的，但是我们会关注$\mathbf{R}^n$更多。当这些向量进入 $n\times n$ 方形矩阵 A 的列时，有两个相同的事实：

> 若 A 的 n 个列是无关，他们生成$\mathbf{R}^n$，所以 $A\boldsymbol{x}=\boldsymbol{b}$ 有解。
> 若这 n 个列生成$\mathbf{R}^n$，他们是无关，所以 $A\boldsymbol{x}=\boldsymbol{b}$ 有唯一解。

唯一性推论到存在性而且存在性也推论到唯一性，则 A 是可逆的。如果不存在自由变量，解 $\boldsymbol{x}$ 是唯一，此时必须有 n 个主元，然后利用反向代入法求解 $A\boldsymbol{x}=\boldsymbol{b}$ (存在解)。

从反向开始，假设对于每个 $\boldsymbol{b}$，$A\boldsymbol{x}=\boldsymbol{b}$ 都有解(存在解)，消元法不会得到零行，有 n 个主元，没有自由变量，零空间只包含 $\boldsymbol{x}=\mathbf{0}$ (唯一性)。

对于行空间与零空间的基底来说，我们有 $r+(n-r)=n$ 个向量，这是正确的数字，这 n 个向量是无关[2]，因此他们生成$\mathbf{R}^n$。

每个 $\boldsymbol{x}$ 是 $\boldsymbol{x}_r+\boldsymbol{x}_n$ 的总和，其中 $\boldsymbol{x}_r$ 来自行空间，$\boldsymbol{x}_n$ 来自零空间。

范例 5 $A=\begin{bmatrix}1&2\\3&6\end{bmatrix}$ 把 $\boldsymbol{x}=\begin{bmatrix}4\\3\end{bmatrix}$ 分成 $\boldsymbol{x}_r+\boldsymbol{x}_n=\begin{bmatrix}2\\4\end{bmatrix}+\begin{bmatrix}2\\-1\end{bmatrix}$。

向量(2, 4)在行空间中，正交向量(2, – 1)在零空间中。下个段落会计算对于任意 A 与 $\boldsymbol{x}$ 的分割，使用投影法(projection)。 【原文没有 footnote 1】

[2] 若全部 n 个向量的组合得到 $\boldsymbol{x}_r+\boldsymbol{x}_n=\mathbf{0}$，则 $\boldsymbol{x}_r=-\boldsymbol{x}_n$ 同时在两个空间，所以 $\boldsymbol{x}_r=-\boldsymbol{x}_n=\mathbf{0}$。行空间与零空间的基底的所有系数必须为零，证明了这 n 个向量在一起是无关。

主要观念的复习

1. 如果 $\boldsymbol{V}$ 中的每个 $\boldsymbol{v}$ 都与 $\boldsymbol{W}$ 中的每个 $\boldsymbol{w}$ 正交，则子空间 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 正交。
2. 如果 $\boldsymbol{W}$ 包含所有与 $\boldsymbol{V}$ 垂直的向量(反之亦然)，则 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 是“正交补充”。
3. 零空间 $\boldsymbol{N}(A)$与行空间 $\boldsymbol{C}(A^{\mathrm{T}})$是正交补充，维度是 $(n-r)+r=n$。相似的，$\boldsymbol{N}(A^{\mathrm{T}})$与 $\boldsymbol{C}(A)$是正交补充，维度是 $(m-r)+r=m$。
4. $\mathbf{R}^n$中任何 n 个无关向量生成$\mathbf{R}^n$，任何 n 个生成向量必然无关。

已解范例

4.1A 假设 $\boldsymbol{S}$ 是 9 维空间$\mathbf{R}^9$中的 6 维子空间。

(a) 与 $\boldsymbol{S}$ 正交的子空间的可能维度是多少？

(b) $\boldsymbol{S}$ 的正交补充 $\boldsymbol{S}^{\perp}$ 的可能维度是多少？

(c) 矩阵 A 的行空间是 $\boldsymbol{S}$，则 A 可能的最小大小(smallest size)为何？

(d) 矩阵 B 的零空间是 $\boldsymbol{S}^{\perp}$，则 B 可能的最小大小为何？

解

(a) 若 $\boldsymbol{S}$ 是$\mathbf{R}^9$的 6 维子空间，与 $\boldsymbol{S}$ 正交的子空间的可能维度是 0, 1, 2, 3。

(b) 正交补充 $\boldsymbol{S}^{\perp}$是最大的正交子空间，维度是 3。

(c) 最小的矩阵 A 是 6×9 (它的 6 个行是 $\boldsymbol{S}$ 的一组基底) 。

(d) 答案与(c)相同。

如果 B 的新行 7 是 A 的 6 个行的组合，则 B 与 A 有相同的行空间，也有相同的零空间。$A\boldsymbol{x}=\mathbf{0}$ 的特殊解 $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ 与 $B\boldsymbol{x}=\mathbf{0}$ 的特殊解相同。消元法会把 B 的第 7 行变成全部是零。

4.1B 方程式 $x-3y-4z=0$ 描述了$\mathbf{R}^3$的一个平面 P (实际是个子空间)。

(a) 平面 P 是哪个 1×3 矩阵 A 的零空间 $\boldsymbol{N}(A)$？答案：$A=[1 \quad -3 \quad -4]$。

(b) 找出 $x-3y-4z=0$ 的特殊解的基底 $\boldsymbol{s}_1$ 与 $\boldsymbol{s}_2$。(这些是零空间矩阵 N 的列)。
答案：$\boldsymbol{s}_1=(3, 1, 0)$与 $\boldsymbol{s}_2=(4, 0, 1)$。

(c) 找出垂直 P 的直线 $P^{\perp}$的一组基底。答案：$(1, -3, -4)$。

请尊重版权与译者的劳动成果，侵权必究!

问题集 4.1

问题 1-12 图 4.2 与 4.3 的四个子空间的发展。

1 写出任意 2×3 的秩 1 矩阵，复制图 4.2，在每个子空间中放置一个向量(零空间中放两个。) 哪些向量是正交?

2 对 3×2 秩 $r=2$ 的矩阵重画图 4.3，哪个子空间是 **Z**(只有零向量)? $\mathbf{R}^2$中任意向量 $\boldsymbol{x}$ 的零空间部分 $\boldsymbol{x}_n$ = _____。

3 写出一个满足要求的矩阵，或是说明为何不可能?

(a) 列空间包含 $\begin{bmatrix}1\\2\\-3\end{bmatrix}$ 与 $\begin{bmatrix}2\\-3\\5\end{bmatrix}$，零空间包含 $\begin{bmatrix}1\\1\\1\end{bmatrix}$

(b) 行空间包含 $\begin{bmatrix}1\\2\\-3\end{bmatrix}$ 与 $\begin{bmatrix}2\\-3\\5\end{bmatrix}$，零空间包含 $\begin{bmatrix}1\\1\\1\end{bmatrix}$

(c) $A\boldsymbol{x}=\begin{bmatrix}1\\1\\1\end{bmatrix}$有一个解且 $A^{\mathrm{T}}\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$。

(d) 每一行与每一列正交 (A 不是零矩阵)

(e) 全部列相加得到一列的零，全部行相加得到一行的 1's。

4 若 $AB=0$，则 B 的列在 A 的_____中，A 的行在 B 的_____中。$AB=0$，为什么 A 与 B 不可能是秩 2 的 3×3 矩阵。

5 (a) 若 $A\boldsymbol{x}=\boldsymbol{b}$ 有一个解且 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$，则$(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}=0)$或是$(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}=0)$?

(b) 若 $A^{\mathrm{T}}\boldsymbol{y}=(1,1,1)$有一个解且 $A\boldsymbol{x}=\boldsymbol{0}$，则_____。

6 方程式系统 $A\boldsymbol{x}=\boldsymbol{b}$ 无解(他们得到 0 = 1)：

$$\begin{aligned}x+2y+2z&=5\\2x+2y+3z&=5\\3x+4y+5z&=9\end{aligned}$$

求出数字 y_1, y_2, y_3 分别乘方程式相加后得到 0 = 1，你会发现向量 $\boldsymbol{y}$ 在哪个子空间？它的点积 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$ 是 1，所以 $\boldsymbol{x}$ 无解。

7 每个无解的系统就像问题 6 一样，存在数字 $y_1, \ldots, y_m$ 分别乘 m 个方程式，相加后得到 0 = 1。这个称为“Fredholm's 替代(alternative)”：

下列问题中恰好只有一个问题有一个解

$A\boldsymbol{x}=\boldsymbol{b}$ 或 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 且 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$ 是 1

若 $\boldsymbol{b}$ 不在 A 的列空间中，它与 A^{T} 的零空间就不正交。分别用数字 y_1, y_2, y_3 乘方程式 $x_1-x_2=1$ 与 $x_2-x_3=1$ 与 $x_1-x_3=1$，选取适当的 y's 使得方程式相加得到 0 = 1。

8 在图 4.3 中，我们如何知道 $A\boldsymbol{x}_r$ 等于 $A\boldsymbol{x}$？我们如何知道这个向量在列空间中？

若 $A=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$ 与 $\boldsymbol{x}=\begin{bmatrix}1\\ 0\end{bmatrix}$，求 $\boldsymbol{x}_r$。

9 若 $A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}$ 则 $A\boldsymbol{x}=\boldsymbol{0}$。理由：$A\boldsymbol{x}$ 在 A^{T} 的零空间中，也在 A 的_____中且这些空间是_____。结论：$A^{\mathrm{T}}A$ 与 A 有相同的零空间。下个段落会重复这个关键事实。

10 假设 A 是对称矩阵($A^{\mathrm{T}}=A$)。

(a) 为什么它的列空间与它的零空间垂直？

(b) 若 $A\boldsymbol{x}=\boldsymbol{0}$ 与 $A\boldsymbol{z}=5\boldsymbol{z}$，哪个子空间包含“固有向量”$\boldsymbol{x}$ 与 $\boldsymbol{z}$？

对称矩阵有垂直的固有向量 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}=0$。

11 (推荐) 画出图 4.2，正确的展示每个子空间：

$$A=\begin{bmatrix}1 & 2\\ 3 & 6\end{bmatrix} \text{ 与 } B=\begin{bmatrix}1 & 0\\ 3 & 0\end{bmatrix}$$

12 找出片段 $\boldsymbol{x}_r$ 与 $\boldsymbol{x}_n$ 并且正确画出图 4.3，如果

$$A=\begin{bmatrix}1 & -1\\ 0 & 0\\ 0 & 0\end{bmatrix} \text{ 与 } \boldsymbol{x}=\begin{bmatrix}2\\ 0\end{bmatrix}$$

问题 13-23 有关正交子空间。

13 把子空间 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 的基底放进矩阵 V 与 W 的列，说明为什么正交子空间的测试可以写成 $V^{\mathrm{T}}W=0$ 矩阵？这也符合正交向量的 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}=0$。

14 地板 $\boldsymbol{V}$ 与墙 $\boldsymbol{W}$ 不是正交子空间，因为他们共享一个相同的非零向量(沿着交线的向量)，没有$\mathbf{R}^3$中的平面 $\boldsymbol{V}$ 与平面 $\boldsymbol{W}$ 可以正交！找出下列两个矩阵的列空间的一个向量：

$$A=\begin{bmatrix}1 & 2\\ 1 & 3\\ 1 & 2\end{bmatrix} \text{ 与 } B=\begin{bmatrix}5 & 4\\ 6 & 3\\ 5 & 1\end{bmatrix}$$

这会是一个向量 $A\boldsymbol{x}$ 以及 $B\hat{\boldsymbol{x}}$。想象 3×4 的矩阵[A　B]。

15 延申问题 14 至$\mathbf{R}^n$的一个 p 维的子空间 $\boldsymbol{V}$ 与一个 q 维的子空间 $\boldsymbol{W}$，什么样的 $p+q$ 的不等式确保 $\boldsymbol{V}$ 与 $\boldsymbol{W}$ 的交集是非零向量？这些子空间不可能正交。

16 利用方程式(2)的矩阵缩写，证明 $\boldsymbol{N}(A^{\mathrm{T}})$中每一个 $\boldsymbol{y}$ 与列空间中每一个 $A\boldsymbol{x}$ 垂直。从 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 开始。

请尊重版权与译者的劳动成果，侵权必究!

17 若$\mathbf{R}^3$的子空间 $\boldsymbol{S}$ 只包含零向量，$\boldsymbol{S}^{\perp}$为何？若 $\boldsymbol{S}$ 由(1, 1, 1)生成，$\boldsymbol{S}^{\perp}$为何？若 $\boldsymbol{S}$ 由(1, 1, 1)与(1, 1, –1)生成，$\boldsymbol{S}^{\perp}$的一组基底为何？

18 假设 $\boldsymbol{S}$ 只包含两个向量(1, 5, 1)与(2, 2, 2) (不是子空间)，则 $\boldsymbol{S}^{\perp}$是矩阵 $A =$ _____ 的零空间。尽管 $\boldsymbol{S}$ 不是子空间，但是 $\boldsymbol{S}^{\perp}$是子空间。

19 假设 $\boldsymbol{L}$ 是$\mathbf{R}^3$的一维子空间(一条直线)，它的正交补充 $\boldsymbol{L}^{\perp}$是垂直 $\boldsymbol{L}$ 的_____。则$(\boldsymbol{L}^{\perp})^{\perp}$是垂直 $\boldsymbol{L}^{\perp}$的_____。事实上$(\boldsymbol{L}^{\perp})^{\perp}$与_____相同。

20 假设 $\boldsymbol{V}$ 是整个空间$\mathbf{R}^4$，$\boldsymbol{V}^{\perp}$只包含_____向量，则$(\boldsymbol{V}^{\perp})^{\perp}$是_____，所以$(\boldsymbol{V}^{\perp})^{\perp}$与_____相同。

21 假设 $\boldsymbol{S}$ 由向量(1, 2, 2, 3)与(1, 3, 3, 2)生成，找出两个向量生成 $\boldsymbol{S}^{\perp}$，这相当于求解 $A\boldsymbol{x} = \mathbf{0}$，求 A？

22 假设$\mathbf{R}^4$的平面 $\boldsymbol{P}$ 满足 $x_1 + x_2 + x_3 + x_4 = 0$，写出 $\boldsymbol{P}^{\perp}$的一组基底，写出把 P 作为零空间的矩阵。

23 假设子空间 $\boldsymbol{V}$ 包含子空间 $\boldsymbol{S}$，证明 $\boldsymbol{S}^{\perp}$包含 $\boldsymbol{V}^{\perp}$。

问题 24-30 有关垂直列与行。

24 假设一个 $n \times n$ 矩阵可逆：$A\,A^{-1} = I$，则 A^{-1} 的第一列与 A 的哪些行生成的空间正交？

25 假设 A 的列是单位向量且全部互相垂直，求 $A^{\mathrm{T}}A$。

26 建立 3×3 矩阵 A，其中 A 没有 0 单元且它的列互相垂直，计算 $A^{\mathrm{T}}A$。为什么它是对角矩阵?

27 直线 $3x + y = b_1$ 与 $6x + 2y = b_2$ 是_____。如果_____，他们是同一条线，这个情况下(b_1, b_2)与向量_____垂直。矩阵的零空间是直线 $3x + y =$ ____，这个零空间中的一个特定向量是_____。

28 为什么下列的叙述是错误的?

(a) (1, 1, 1)与(1, 1, –2)垂直，所以 $x + y + z = 0$ 与 $x + y - 2z = 0$ 是正交子空间。

(b) (1, 1, 0, 0, 0)与(0, 0, 0, 1, 1)生成的子空间，是(1, –1, 0, 0, 0)与(2, –2, 3, 4, –4)生成的子空间的正交补充。

(c) 交集只有零向量的两个子空间正交。

29 求出一个矩阵使得 $\boldsymbol{v}$ (1, 2, 3)同时在矩阵的行空间与列空间中。找出另一个矩阵使得 $\boldsymbol{v}$ 同时在零空间与列空间中。$\boldsymbol{v}$ 不能同时在哪些子空间的配对?

挑战问题

30 假设 A 是 3×4 与 B 是 4×5 且 $AB = 0$，所以 $\boldsymbol{N}(A)$包含 $\boldsymbol{C}(B)$。由 $\boldsymbol{N}(A)$与 $\boldsymbol{C}(B)$的维度证明 $\text{rank}(A) + \text{rank}\,(B) \le 4$。

31 指令 $N = \text{null}(A)$ 会得到 A 的零空间的一组基底。则指令 $B = \text{null}(N')$会得到 A 的_____的一组基底。

32 假设我给你 4 个$\mathbf{R}^2$的非零向量 $\boldsymbol{r}, \boldsymbol{n}, \boldsymbol{c}, \boldsymbol{l}$：

(a) 要想成为 2×2 矩阵的 4 个基础子空间 $\boldsymbol{C}(A^{\mathrm{T}}), \boldsymbol{N}(A), \boldsymbol{C}(A), \boldsymbol{N}(A^{\mathrm{T}})$的基底，这些向量要有什么条件？

(b) 可能的矩阵 A 为何？

33 假设我给你 8 个$\mathbf{R}^4$的向量 $\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{l}_1, \boldsymbol{l}_2$：

(a) 要想成为 4×4 矩阵的 4 个基础子空间的基底，这些向量要有什么条件？

(b) 可能的矩阵 A 为何？

请尊重版权与译者的劳动成果，侵权必究！

4.2 投影

1　向量 $\boldsymbol{b}$ 在通过 $\boldsymbol{a}$ 的直线的投影是直线与 $\boldsymbol{b}$ 的最近点 $\boldsymbol{p}=\boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$。

2　误差 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 与 $\boldsymbol{a}$ 垂直：直角 $\boldsymbol{b}\,\boldsymbol{p}\,\boldsymbol{e}$ 有 $\|\boldsymbol{p}\|^2+\|\boldsymbol{e}\|^2=\|\boldsymbol{b}\|^2$。

3　$\boldsymbol{b}$ 在子空间 $\boldsymbol{S}$ 的投影是 $\boldsymbol{S}$ 中与 $\boldsymbol{b}$ 的最近点 $\boldsymbol{p}$；$\boldsymbol{b}-\boldsymbol{p}$ 与 $\boldsymbol{S}$ 正交。

4　只有在 A 有无关列时，$A^{\mathrm{T}}A$ 可逆(且对称)：$\boldsymbol{N}(A^{\mathrm{T}}A)=\boldsymbol{N}(A)$。

5　$\boldsymbol{b}$ 在 A 的列空间的投影是向量 $\boldsymbol{p}=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$。

6　投影到 $\boldsymbol{C}(A)$的**投影矩阵**是 $\boxed{P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}},}$ 其中 $\boldsymbol{p}=P\boldsymbol{b}$ 且 $P^2=P=P^{\mathrm{T}}$。

本段落开始前我可以提出两个问题吗？(那个问题再加一。) 第一个问题的目的在于展示投影是很容易视觉化的，第二个问题是有关“投影矩阵(projection matrix)”——对称矩阵且 $P^2=P$。$\boldsymbol{b}$ 的投影是 $P\boldsymbol{b}$。

1　$\boldsymbol{b}=(2,3,4)$在 z 轴与在 xy 平面的投影为何？

2　什么样的矩阵 P_1 与 P_2 可以产生在直线与平面的投影？

当 $\boldsymbol{b}$ 投影到一条直线，它的投影 $\boldsymbol{p}$ 是 $\boldsymbol{b}$ 沿着这条直线的部分。如果 $\boldsymbol{b}$ 投影到一个平面，$\boldsymbol{p}$ 是 $\boldsymbol{b}$ 在平面的部分。投影 $\boldsymbol{p}$ 就是 $P\boldsymbol{b}$。

投影矩阵 P 乘 $\boldsymbol{b}$ 得到 $\boldsymbol{p}$，本段落求 $\boldsymbol{p}$ 也求 P。

我们把在 z 轴的投影称为 $\boldsymbol{p}_1$，第二个投影直接下降至 xy 平面，你心中的图形应该是图 4.5。从 $\boldsymbol{b}=(2,3,4)$开始，横向投影得到 $\boldsymbol{p}_1=(0,0,4)$，下降投影得到 $\boldsymbol{p}_2=(2,3,0)$，这些是 $\boldsymbol{b}$ 沿着 z 轴与在 xy 平面的部分。

投影矩阵 P_1 与 P_2 是 3×3，他们乘有 3 个分量的 $\boldsymbol{b}$ 得到有 3 个分量的 $\boldsymbol{p}$。在直线的投影来自一个秩 1 矩阵，在平面的投影来自一个秩 2 矩阵。

投影矩阵　到 z 轴：$P_1=\begin{bmatrix}0&0&0\\0&0&0\\0&0&\mathbf{1}\end{bmatrix}$　　到 xy 平面：$P_2=\begin{bmatrix}\mathbf{1}&0&0\\0&\mathbf{1}&0\\0&0&0\end{bmatrix}$

P_1 筛选出每个向量的 z 分量，P_2 筛选出 x 与 y 的分量。要求 $\boldsymbol{b}$ 的投影 $\boldsymbol{p}_1$ 与 $\boldsymbol{p}_2$，使用 P_1 与 P_2 乘 $\boldsymbol{b}$ (小 $\boldsymbol{p}$ 是向量，大写 P 是产生 $\boldsymbol{p}$ 的矩阵)：

$$\boldsymbol{p}_1=P_1\boldsymbol{b}=\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\\mathbf{0}\\\boldsymbol{z}\end{bmatrix}\qquad \boldsymbol{p}_2=P_2\boldsymbol{b}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}\boldsymbol{x}\\\boldsymbol{y}\\\mathbf{0}\end{bmatrix}$$

这个情形下投影 $\boldsymbol{p}_1$ 与 $\boldsymbol{p}_2$ 垂直，xy 平面与 z 轴是**正交子空间**，就好像房间的地板与两面墙的交线一样。

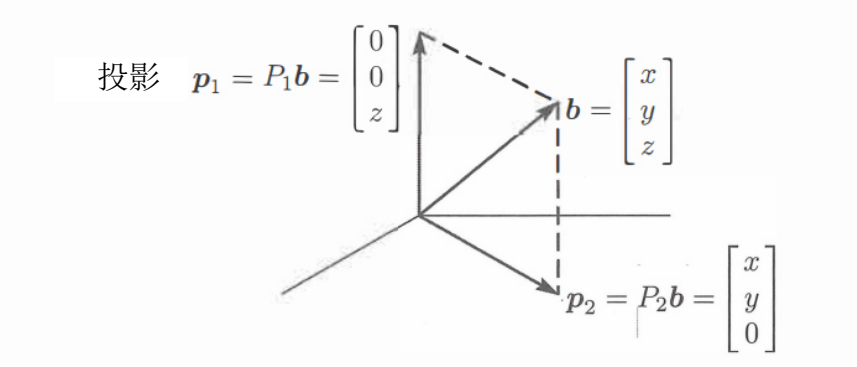


图 4.5：在 z 轴与 xy 平面的投影 $\boldsymbol{p}_1 = P_1\boldsymbol{b}$ 与 $\boldsymbol{p}_2 = P_2\boldsymbol{b}$。

不只是正交，直线与平面是**正交补充**，他们的维度加起来是 1 + 2 = 3。整个空间的每个向量 $\boldsymbol{b}$ 都是它在这两个子空间的部分的总和，投影 $\boldsymbol{p}_1$ 与 $\boldsymbol{p}_2$ 恰好就是 $\boldsymbol{b}$ 的这两个部分：

$$\text{向量得到 } \boldsymbol{p}_1 + \boldsymbol{p}_2 = \boldsymbol{b} \quad \text{矩阵得到 } P_1 + P_2 = I \tag{1}$$

这是完美的，我们达成了目标--在这个例子，对于任何直线与任何平面以及任何 n 维子空间，我们都有相同的目标。我们的目的是在每个子空间找到 $\boldsymbol{p}$ 的部分，以及可以得到那个部分的矩阵 P 使得 $\boldsymbol{p} = P\boldsymbol{b}$。$\mathbf{R}^m$的每一个子空间有自己的 $m\times m$ 投影矩阵，要计算 P，我们绝对需要投影的目标子空间的良好描述。

子空间的最佳描述就是基底，我们把基底放进 A 的列，现在我们是**对于 $\boldsymbol{A}$ 的列空间在投影**！当然 z 轴是 3×1 矩阵 A_1 的列空间，xy 平面是矩阵 A_2 的列空间，这个平面也是 A_3 的列空间(一个子空间有很多基底)，所以 $\boldsymbol{p}_2 = \boldsymbol{p}_3$ 且 $P_2 = P_3$。

$$A_1 = \begin{bmatrix}0\\0\\1\end{bmatrix} \quad \text{与} \quad A_2 = \begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix} \quad \text{与} \quad A_3 = \begin{bmatrix}1&2\\2&3\\0&0\end{bmatrix}$$

我们的问题是把**任意的 $\boldsymbol{b}$ 投影至任何 $\boldsymbol{m\times n}$ 矩阵的列空间**。从一条直线开始(维度 $n = 1$)，矩阵 A 只有一列，称为 $\boldsymbol{a}$。

在直线的投影

一条通过原点且方向是 $\boldsymbol{a} = (a_1, ..., a_m)$的直线，沿着这条直线，我们希望点 $\boldsymbol{p}$ 最靠近 $\boldsymbol{b} = (b_1, ..., b_m)$。投影的关键在于正交：**由 $\boldsymbol{b}$ 到 $\boldsymbol{p}$ 的直线与向量 $\boldsymbol{a}$ 垂直**，这是图 4.6 左侧标示 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ 的点线，其中 $\boldsymbol{e}$ 代表误差。我们现在用代数计算 $\boldsymbol{p}$。

请尊重版权与译者的劳动成果，侵权必究！

投影 $\boldsymbol{p}$ 会是 $\boldsymbol{a}$ 的某个倍数，写成 $\boldsymbol{p}=\hat{x}\boldsymbol{a}=$“$x$ hat”乘 $\boldsymbol{a}$。计算数字$\hat{x}$可以得到向量 $\boldsymbol{p}$，从 $\boldsymbol{p}$ 的公式我们得到投影矩阵 P。这三个步骤可以求得所有的投影矩阵：找出 $\hat{x}$，找出向量 $\boldsymbol{p}$，然后找出矩阵 P。

点线 $\boldsymbol{b}-\boldsymbol{p}$ 是“误差”$\boldsymbol{e}=\boldsymbol{b}-\hat{x}\boldsymbol{a}$，它与 $\boldsymbol{a}$ 垂直--这个可以决定 $\hat{x}$。当他们的点积为 0，使用 $\boldsymbol{b}-\hat{x}\boldsymbol{a}$ 与 $\boldsymbol{a}$ 垂直的事实：

$$\begin{array}{l}\boldsymbol{b}\text{在}\boldsymbol{a}\text{的投影有 } \boldsymbol{e}=\boldsymbol{b}-\hat{x}\boldsymbol{a}\\ \boldsymbol{a}\cdot(\boldsymbol{b}-\hat{x}\boldsymbol{a})=0 \text{ 或 } \boldsymbol{a}\cdot\boldsymbol{b}-\hat{x}\boldsymbol{a}\cdot\boldsymbol{a}=0\end{array} \qquad \hat{x}=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\boldsymbol{a}\cdot\boldsymbol{a}}=\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \tag{2}$$

乘法 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ 与 $\boldsymbol{a}\cdot\boldsymbol{b}$ 相同，使用转置会好一点，这是因为它也可以应用在矩阵。我们的公式 $\hat{x}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ 得到投影 $\boldsymbol{p}=\hat{x}\boldsymbol{a}$。

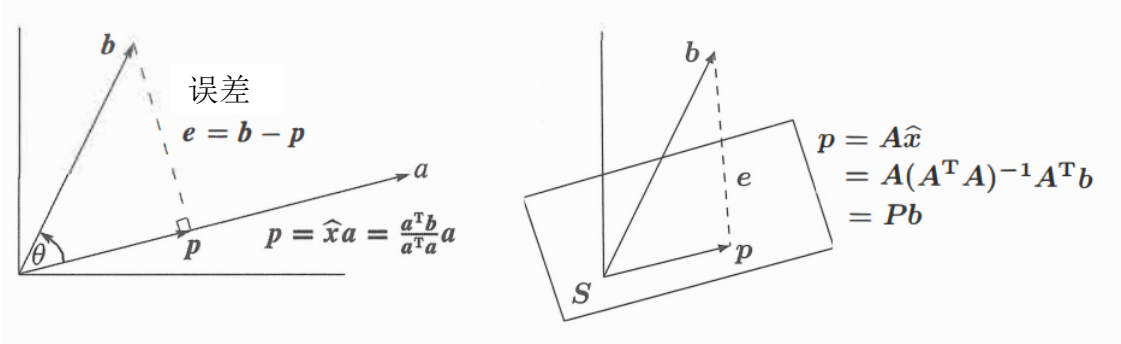


图 4.6：$\boldsymbol{b}$ 在一条直线的投影 $\boldsymbol{p}$ 以及 $\boldsymbol{b}$ 在 $\boldsymbol{S}=A$ 的列空间的投影

向量 $\boldsymbol{b}$ 在通过 $\boldsymbol{a}$ 的直线的投影是向量 $\boldsymbol{p}=\hat{x}\boldsymbol{a}=\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a}$

特殊案例 1：若 $\boldsymbol{b}=\boldsymbol{a}$ 则 $\hat{x}=1$，$\boldsymbol{a}$ 在 $\boldsymbol{a}$ 的投影就是自己，$P\boldsymbol{a}=\boldsymbol{a}$。

特殊案例 2：若 $\boldsymbol{b}$ 与 $\boldsymbol{a}$ 垂直则 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=0$，投影是 $\boldsymbol{p}=\boldsymbol{0}$。

范例 1　把 $\boldsymbol{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$ 投影在 $\boldsymbol{a}=\begin{bmatrix}1\\2\\2\end{bmatrix}$，求图 4.6 的 $\boldsymbol{p}=\hat{x}\boldsymbol{a}$。

解　数字 $\hat{x}$ 是 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}=5$ 与 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}=9$ 的比值，所以投影 $\boldsymbol{p}$ 是 $\dfrac{5}{9}\boldsymbol{a}$。

$\boldsymbol{b}$ 与 $\boldsymbol{p}$ 之间的误差向量是 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$，向量 $\boldsymbol{p}$ 与 $\boldsymbol{e}$ 相加得到 $\boldsymbol{b}=(1,1,1)$:

$$\boldsymbol{p}=\frac{5}{9}\boldsymbol{a}=\left(\frac{5}{9},\frac{10}{9},\frac{10}{9}\right) \quad 与 \quad \boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}=\left(\frac{4}{9},-\frac{1}{9},-\frac{1}{9}\right)$$

误差 $\boldsymbol{e}$ 会垂直 $\boldsymbol{a}=(1,2,2)$，就是 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{a}=\frac{4}{9}-\frac{2}{9}-\frac{2}{9}=0$。

检视直角三角形 $\boldsymbol{b}$，$\boldsymbol{p}$ 与 $\boldsymbol{e}$，向量 $\boldsymbol{b}$ 分成两个部分——沿着直线的分量是 $\boldsymbol{p}$，垂直的部分是 $\boldsymbol{e}$。这两个边 $\boldsymbol{p}$ 与 $\boldsymbol{e}$ 的长度是 $\|\boldsymbol{p}\|=\|\boldsymbol{b}\|\cos\theta$ 与 $\|\boldsymbol{e}\|=\|\boldsymbol{b}\|\sin\theta$。三角学与点积是适配的：

$$\boldsymbol{p}=\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a} \text{ 的长度 } \|\boldsymbol{p}\| = \frac{\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta}{\|\boldsymbol{a}\|^2}\|\boldsymbol{a}\| = \|\boldsymbol{b}\|\cos\theta \tag{3}$$

点积会比掺杂 $\cos\theta$ 与 $\boldsymbol{b}$ 的长度来得更简洁，这个例子在 $\cos\theta=5/3\sqrt{3}$ 与 $\|\boldsymbol{b}\|=\sqrt{3}$ 出现平方根，投影 $\boldsymbol{p}=5\boldsymbol{a}/9$ 却没有平方根出现，获得 5/9 的好方法是 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。

现在来到**投影矩阵**，在 $\boldsymbol{p}$ 的公式中，要用什么矩阵乘 $\boldsymbol{b}$？如果数字 $\hat{x}$ 在 $\boldsymbol{a}$ 的右侧，你就看得出来矩阵的优点：

投影矩阵 $\boldsymbol{P}$　当矩阵是 $P=\dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$ 则 $\boldsymbol{p}=\boldsymbol{a}\hat{x}=\boldsymbol{a}\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}=P\boldsymbol{b}$

P 是一列乘一行！列是 $\boldsymbol{a}$，行是 $\boldsymbol{a}^{\mathrm{T}}$，再除以数字 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。投影矩阵 P 是 $m\times m$，但是它的**秩是 1**。我们在对 1 维子空间做投影，直线通过 $\boldsymbol{a}$，这条线就是 P 的列空间。

范例 2　求出在通过 $\boldsymbol{a}=\begin{bmatrix}1\\2\\2\end{bmatrix}$ 的直线的投影矩阵 $P=\dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$。

解　执行乘法：(列 $\boldsymbol{a}$)乘(行 $\boldsymbol{a}^{\mathrm{T}}$)，再除以 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}=9$：

$$\text{投影矩阵} \qquad P=\frac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}=\frac{1}{9}\begin{bmatrix}1\\2\\2\end{bmatrix}\begin{bmatrix}1&2&2\end{bmatrix}=\frac{1}{9}\begin{bmatrix}1&2&2\\2&4&4\\2&4&4\end{bmatrix}$$

这个矩阵可以把任何 $\boldsymbol{b}$ 投影至 $\boldsymbol{a}$，检验范例 1 的 $\boldsymbol{p}=P\boldsymbol{b}$，其中 $\boldsymbol{b}=(1,1,1)$：

$$\boldsymbol{p}=P\boldsymbol{b}=\frac{1}{9}\begin{bmatrix}1&2&2\\2&4&4\\2&4&4\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}=\frac{1}{9}\begin{bmatrix}5\\10\\10\end{bmatrix} \quad 这是正确的。$$

如果向量 $\boldsymbol{a}$ 加倍，矩阵 P 保持相同！它仍然投影在相同的直线。如果矩阵变成平方，P^2 等于 P，第二次的投影不会改变任何东西，所以 $P^2=P$。P 的对角单元全部相加得到 $(1+4+4)/9=1$。

请尊重版权与译者的劳动成果，侵权必究！

矩阵 $I-P$ 必然也是投影矩阵，它可以得到三角形的另一边 $\boldsymbol{e}$ ——也就是 $\boldsymbol{b}$ 的垂直部分。注意$(I-P)\boldsymbol{b}$ 等于 $\boldsymbol{b}-\boldsymbol{p}$，这就是左零空间中的 $\boldsymbol{e}$。

当 P 投影在一个子空间，$I-P$ 会投影在垂直的子空间，此处的 $I-P$ 投影在垂直 $\boldsymbol{a}$ 的平面。

现在我们超越在一条直线的投影，投影至$\mathbf{R}^m$的 n 维子空间需要更多付出，重要的公式会收集在方程式(5)–(6)–(7)，基本上你需要记住这三个方程式。

投影在子空间

从$\mathbf{R}^m$的 n 个向量 $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$ 开始，假设这些 $\boldsymbol{a}$'s 是线性无关。

问题：找出组合 $\boldsymbol{p}=\hat{x}_1\,\boldsymbol{a}_1+...+\hat{x}_n\boldsymbol{a}_n$，使得 $\boldsymbol{p}$ 最靠近给定的向量 $\boldsymbol{b}$。

我们是把$\mathbf{R}^m$的每个 $\boldsymbol{b}$ 投影至 $\boldsymbol{a}$'s 生成的子空间。

当 $n=1$(一个向量 $\boldsymbol{a}_1$)，就是投影至一条直线，这条直线是 A 的列空间，A 只有一个列，一般来说 A 有 n 个列 $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$。

$\mathbf{R}^m$中的组合是列空间中的向量 $A\boldsymbol{x}$，我们在寻找最靠近 $\boldsymbol{b}$ 的特解 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$ (投影)，$\boldsymbol{x}$ 上方的帽子指出这是最佳选择$\hat{\boldsymbol{x}}$，产生列空间中最靠近的向量。当 $n=1$，这个选择是 $\hat{x}=\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$。当 $n>1$，我们现在要找出最佳 $\hat{\boldsymbol{x}}=(\hat{x}_1, ..., \hat{x}_n)$。

我们像以前一样用三个步骤来计算在 n 维子空间的投影：

找出向量 $\hat{\boldsymbol{x}}$，找出投影 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$，求出投影矩阵 P。

关键在几何！图 4.6 的点线从 $\boldsymbol{b}$ 到子空间中的最近点 $A\hat{\boldsymbol{x}}$，误差向量 $\boldsymbol{b}-A\hat{\boldsymbol{x}}$与子空间垂直。误差 $\boldsymbol{b}-A\hat{\boldsymbol{x}}$ 与基底的所有向量 $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$ 的角度都是直角，这 n 个直角得到 n 个 $\hat{\boldsymbol{x}}$ 的方程式：

$$\begin{matrix}\boldsymbol{a}_1^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=0\\ \vdots\\ \boldsymbol{a}_n^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=0\end{matrix}\quad\text{或}\quad\begin{bmatrix}- & \boldsymbol{a}_1^{\mathrm{T}} & -\\ & \vdots & \\ - & \boldsymbol{a}_n^{\mathrm{T}} & -\end{bmatrix}\begin{bmatrix}\\ \boldsymbol{b}-A\hat{\boldsymbol{x}}\\ \\ \end{bmatrix}=\begin{bmatrix}0\end{bmatrix}\tag{4}$$

具有行 $\boldsymbol{a}_i^{\mathrm{T}}$ 的矩阵是 A^{T}，这 n 个方程式就是 $A^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=\boldsymbol{0}$。

改写 $A^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=\boldsymbol{0}$ 成为著名的形式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$，这是 $\hat{\boldsymbol{x}}$ 的方程式，系数矩阵是 $A^{\mathrm{T}}A$。我们现在可以依照顺序求出 $\hat{\boldsymbol{x}}$ 与 $\boldsymbol{p}$ 与 P。

组合 $\boldsymbol{p}=\hat{x}_1\boldsymbol{a}_1+\ldots+\hat{x}_n\boldsymbol{a}_n=A\hat{\boldsymbol{x}}$，其中 $\boldsymbol{p}$ 是从 $\hat{\boldsymbol{x}}$ 到 $\boldsymbol{b}$ 的最近点：

$$\textbf{求 } \hat{\boldsymbol{x}} \ (n\times 1) \qquad A^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=\boldsymbol{0} \quad \text{或} \quad A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b} \tag{5}$$

对称矩阵 $A^{\mathrm{T}}A$ 是 $n\times n$，如果 $\boldsymbol{a}$'s 无关，$A^{\mathrm{T}}A$ 是可逆，解是$\hat{\boldsymbol{x}}=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$。$\boldsymbol{b}$ 在子空间的投影是 $\boldsymbol{p}$：

$$\textbf{求 } \boldsymbol{p} \ (m\times 1) \qquad \boldsymbol{p}=A\hat{\boldsymbol{x}}=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b} \tag{6}$$

下个公式筛选出投影矩阵，它就是(6)中乘 $\boldsymbol{b}$ 的部分：

$$\textbf{求 } P \ (m\times m) \qquad P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} \tag{7}$$

比较在直线的投影，当 A 只有一列时：$A^{\mathrm{T}}A=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。

$$\textbf{当 } \boldsymbol{n}=\boldsymbol{1}, \qquad \hat{\boldsymbol{x}}=\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \quad \text{与} \quad \boldsymbol{p}=\boldsymbol{a}\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \quad \text{与} \quad P=\frac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$$

这些公式与(5)与(6)与(7)完全相同，数字 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ 变成矩阵 $A^{\mathrm{T}}A$。当它是个数字，我们除以它，当它是个矩阵，我们逆反它，新的公式包含$(A^{\mathrm{T}}A)^{-1}$ 而不是 $1/(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$。这些列 $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$ 的线性无关确保逆矩阵的存在。

关键的步骤在于 $A^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=\boldsymbol{0}$，我们使用几何学($\boldsymbol{e}$ 与每个 $\boldsymbol{a}$ 正交)，线性代数也给出“正态方程式(normal equation)”，这是非常快速又美丽的方法：

1. 我们的子空间是 A 的列空间。
2. 向量 $\boldsymbol{b}-A\hat{\boldsymbol{x}}$ 与列空间垂直。
3. 因此 $\boldsymbol{b}-A\hat{\boldsymbol{x}}$ 在 A^{T} 的零空间中！意思就是 $A^{\mathrm{T}}(\boldsymbol{b}-A\hat{\boldsymbol{x}})=\boldsymbol{0}$。

左零空间在投影非常重要，A^{T} 的零空间包含误差向量 $\boldsymbol{e}=\boldsymbol{b}-A\hat{\boldsymbol{x}}$，向量 $\boldsymbol{b}$ 分成投影 $\boldsymbol{p}$ 与误差 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$。投影产生直角三角形的三个边个 $\boldsymbol{p},\boldsymbol{e}$ 与 $\boldsymbol{b}$。

范例 3 若 $A=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}$ 与 $\boldsymbol{b}=\begin{bmatrix}6\\0\\0\end{bmatrix}$，求出 $\hat{\boldsymbol{x}}$，$\boldsymbol{p}$ 与 P。

解 计算方形矩阵 $A^{\mathrm{T}}A$，也计算向量 $A^{\mathrm{T}}\boldsymbol{b}$：

$$A^{\mathrm{T}}A=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}=\begin{bmatrix}3&3\\3&5\end{bmatrix} \quad \text{与} \quad A^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}1&1&1\\0&1&2\end{bmatrix}\begin{bmatrix}6\\0\\0\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

现在求解正态方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\hat{\boldsymbol{x}}$：

$$\begin{bmatrix}3&3\\3&5\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}6\\0\end{bmatrix}\quad\text{得到}\quad\hat{\boldsymbol{x}}=\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}5\\-3\end{bmatrix}\tag{8}$$

组合 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$ 是 $\boldsymbol{b}$ 在 A 的列空间的投影：

$$\boldsymbol{p}=5\begin{bmatrix}1\\1\\1\end{bmatrix}-3\begin{bmatrix}0\\1\\2\end{bmatrix}=\begin{bmatrix}5\\2\\-1\end{bmatrix},\ \text{误差}\ \boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}=\begin{bmatrix}1\\-2\\1\end{bmatrix}\tag{9}$$

两个计算上的检查。第一个，误差 $\boldsymbol{e}=(1,-2,1)$与两个列$(1,1,1)$与$(0,1,2)$都垂直。第二个，矩阵 P 乘 $\boldsymbol{b}=(6,0,0)$正确的得到 $\boldsymbol{p}=(5,2,-1)$。在我们得到 P 时，立即可以对特定的 $\boldsymbol{b}$ 求解问题。

投影矩阵是 $P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$，$A^{\mathrm{T}}A$ 的行列式是 $15-9=6$；所以$(A^{\mathrm{T}}A)^{-1}$ 容易解决。A 乘$(A^{\mathrm{T}}A)^{-1}$ 乘 A^{T} 得到 P：

$$(A^{\mathrm{T}}A)^{-1}=\frac{1}{6}\begin{bmatrix}5&-3\\-3&3\end{bmatrix}\text{与}\ P=\frac{1}{6}\begin{bmatrix}5&2&-1\\2&2&2\\-1&2&5\end{bmatrix}\tag{10}$$

我们必须有 $P^2=P$，因为第二次的投影不会改变第一次的投影。

警告： 矩阵 $P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 具有误导性，你可能会尝试分开$(A^{\mathrm{T}}A)^{-1}$ 变成 A^{-1} 乘$(A^{\mathrm{T}})^{-1}$。如果你犯了这个错误继续代入 P，你会发现 $P=A\,A^{-1}(A^{\mathrm{T}})^{-1}A^{\mathrm{T}}$。很显然全部都消除了，看起来就是 $P=I$，单位矩阵，我们要来说明为什么是错误。

矩阵 A 是矩形，没有逆矩阵。因为没有最前方的 A^{-1}，我们不能分开$(A^{\mathrm{T}}A)^{-1}$ 变成 A^{-1} 乘$(A^{\mathrm{T}})^{-1}$。

在我们的经验看来，牵涉到矩形矩阵的问题大部分都会得到 $A^{\mathrm{T}}A$。当 A 有无关的列，$A^{\mathrm{T}}A$ 是可逆。这个事实非常重要，所以我们要清楚的阐述并且证明。

$A^{\mathrm{T}}A$ 是可逆 当且仅当 A 有线性无关的列。

证明 $A^{\mathrm{T}}A$ 是方形矩阵$(n\times n)$，对于每一个矩阵 A，我们现在展示 $A^{\mathrm{T}}A$ 与 A 有相同的零空间。当 A 的列是线性无关时，它的零空间只包含零向量，则有相同零空间的 $A^{\mathrm{T}}A$ 是可逆。

令 A 是任意矩阵，若 $\boldsymbol{x}$ 在它的零空间中则 $A\boldsymbol{x}=\boldsymbol{0}$，左乘 A^{T} 得到 $A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}$，所以 $\boldsymbol{x}$ 也在 $A^{\mathrm{T}}A$ 的零空间。

现在从 $A^{\mathrm{T}}A$ 的零空间开始，由 $A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}$，我们必须证明 $A\boldsymbol{x}=\boldsymbol{0}$。我们不能左乘$(A^{\mathrm{T}})^{-1}$，这个通常都不存在，只能乘 $\boldsymbol{x}^{\mathrm{T}}$：

$$(\boldsymbol{x}^{\mathrm{T}})A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}\ \text{ 或 }\ (A\boldsymbol{x})^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}\ \text{ 或 }\ \|A\boldsymbol{x}\|^2=0\tag{11}$$

我们已经展示：若 $A^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{0}$ 则 $A\boldsymbol{x}$ 有零长度，因此 $A\boldsymbol{x}=\boldsymbol{0}$。在一个零空间中的每个向量 $\boldsymbol{x}$，也会在另一个零空间中。若 $A^{\mathrm{T}}A$ 有相关的列，A 也是。若 $A^{\mathrm{T}}A$ 有无关的列，A 也是。这就是最好的情况：$A^{\mathrm{T}}A$ 是可逆。

当 A 有无关的列，$A^{\mathrm{T}}A$ 是方形，对称，可逆。

再次强调：$A^{\mathrm{T}}A$ 是$(n\times m)$乘$(m\times n)$，$A^{\mathrm{T}}A$ 是方形$(n\times n)$。它是对称，因为它的转置是 $(A^{\mathrm{T}}A)^{\mathrm{T}}=A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}}$ 等于 $A^{\mathrm{T}}A$。我们刚才证明了 $A^{\mathrm{T}}A$ 可逆--假设 A 有无关的列。注意无关列与相关列的差别：

$$\begin{matrix} A^{\mathrm{T}} & A & A^{\mathrm{T}}A \end{matrix}\qquad \begin{bmatrix}1&1&0\\2&2&0\end{bmatrix}\begin{bmatrix}1&2\\1&2\\0&0\end{bmatrix}=\begin{bmatrix}2&4\\4&8\end{bmatrix}$$

相关　　奇异

$$\begin{matrix} A^{\mathrm{T}} & A & A^{\mathrm{T}}A \end{matrix}\qquad \begin{bmatrix}1&1&0\\2&2&1\end{bmatrix}\begin{bmatrix}1&2\\1&2\\0&1\end{bmatrix}=\begin{bmatrix}2&4\\4&9\end{bmatrix}$$

无关　　可逆

非常简单的摘要　要找出投影 $\boldsymbol{p}=\hat{x}_1\boldsymbol{a}_1+\ldots+\hat{x}_n\boldsymbol{a}_n$，求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$，得到 $\hat{\boldsymbol{x}}$，投影是 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$ 以及误差 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}=\boldsymbol{b}-A\hat{\boldsymbol{x}}$，投影矩阵 $P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 得到 $\boldsymbol{p}=P\boldsymbol{b}$。

矩阵满足 $P^2=P$，从 $\boldsymbol{b}$ 到子空间 $C(A)$的距离是 $\|\boldsymbol{e}\|$。

主要观念的复习

1. $\boldsymbol{b}$ 在通过 $\boldsymbol{a}$ 的直线的投影是 $\boldsymbol{p}=\boldsymbol{a}\hat{x}=\boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$。
2. 秩一投影矩阵 $P=\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$，$P$ 乘 $\boldsymbol{b}$ 得到 $\boldsymbol{p}$。
3. 把 $\boldsymbol{b}$ 投影在子空间出现 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 与子空间垂直。
4. 当 A 有满秩 n，方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$，得到 $\hat{\boldsymbol{x}}$ 与 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$。
5. 投影矩阵 $P=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$，且 $P^{\mathrm{T}}=P$，$P^2=P$，$P\boldsymbol{b}=\boldsymbol{p}$。

已解范例

4.2A　把 $\boldsymbol{b}=(3,4,4)$投影在通过 $\boldsymbol{a}=(2,2,1)$的直线，然后投影在同时包含 $\boldsymbol{a}^*=(1,0,0)$的平面。检验第一个误差向量 $\boldsymbol{b}-\boldsymbol{p}$ 与 $\boldsymbol{a}$ 垂直，第二个误差向量 $\boldsymbol{e}^*=\boldsymbol{b}-\boldsymbol{p}^*$也与 $\boldsymbol{a}$ 垂直。

求出 3×3 投影矩阵 P 会把向量投影在包含 $\boldsymbol{a}$ 与 $\boldsymbol{a}^*$的平面。找出一个向量在上述平面的投影是零向量，为什么它就是误差 $\boldsymbol{e}^*$？

请尊重版权与译者的劳动成果，侵权必究！

解 $\boldsymbol{b} = (3, 4, 4)$在通过 $\boldsymbol{a} = (2, 2, 1)$的直线的投影是 $\boldsymbol{p} = 2\boldsymbol{a}$：

投影在直线 $$\boldsymbol{p} = (\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} / \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})\,\boldsymbol{a} = \frac{18}{9}(2, 2, 1) = (4, 4, 2) = 2\boldsymbol{a}$$

误差向量 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1, 0, 2)$与 $\boldsymbol{a} = (2, 2, 1)$垂直，所以 $\boldsymbol{p}$ 是正确。

包含 $\boldsymbol{a} = (2, 2, 1)$与 $\boldsymbol{a}^* = (1, 0, 0)$的平面是 $A = [\boldsymbol{a} \quad \boldsymbol{a}^*]$的列空间：

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 0 \\ 1 & 0 \end{bmatrix} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 2 \\ 2 & 1 \end{bmatrix} \quad (A^{\mathrm{T}}A)^{-1} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 9 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & .8 & .4 \\ 0 & .4 & .2 \end{bmatrix}$$

现在 $\boldsymbol{p}^* = P\boldsymbol{b} = (3, 4.8, 2.4)$，误差 $\boldsymbol{e}^* = \boldsymbol{b} - \boldsymbol{p}^* = (0, -.8, 1.6)$与 $\boldsymbol{a}$ 及 $\boldsymbol{a}^*$垂直。这个 $\boldsymbol{e}^*$ 在 P 的零空间中且它的投影是零！注意 $P^2 = P = P^{\mathrm{T}}$。

4.2B 假设你的心跳量测每分钟是 $x = 70$，然后是 $x = 80$，再然后是 $x = 120$。这三个方程式 $A\boldsymbol{x} = \boldsymbol{b}$ 只有一个未知数，且 $A = [1 \quad 1 \quad 1]$，$\boldsymbol{b} = (70, 80, 120)$。最佳$\hat{x}$是 70, 80, 120 的_____。使用微积分与投影：

1. 求解 $dE / dx = 0$，极小化 $E = (x - 70)^2 + (x - 80)^2 + (x - 120)^2$。

2. 把 $\boldsymbol{b}$ 投影在 $\boldsymbol{a} = (1, 1, 1)$，找出 $\hat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} / \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。

解 最接近高度 70, 80, 120 的水平线是平均值 $\hat{x} = 90$：

$$dE / dx = 2(x - 70) + 2(x - 80) + 2(x - 120) \text{ 得到} \hat{x} = \frac{70 + 80 + 120}{3} = 90 \text{ 。}$$

利用投影：$$\hat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} / \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = \frac{(1,\ 1,\ 1)^{\mathrm{T}}(70,\ 80,\ 120)}{(1,\ 1,\ 1)^{\mathrm{T}}(1,\ 1,\ 1)} = \frac{70 + 80 + 120}{3} = 90$$

在递归最小二乘，第 4 个量测值 130 把平均值 $\hat{x}_{\mathrm{OLD}} = 90$ 变成 $\hat{x}_{\mathrm{NEW}} = 100$。验证更新公式 $\hat{x}_{\mathrm{NEW}} = \hat{x}_{\mathrm{OLD}} + (130 - \hat{x}_{\mathrm{OLD}})/4$。当我们得到一个新的量测值，我们不需要把旧的量测值再度平均！

问题集 4.2

问题 1-9 询问投影 $\boldsymbol{p}$ 至直线，误差 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$，矩阵 P。

1 把 $\boldsymbol{b}$ 投影在通过 $\boldsymbol{a}$ 的直线，验证误差 $\boldsymbol{e}$ 与 $\boldsymbol{a}$ 垂直:

(a) $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ (b) $\boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} -1 \\ -3 \\ -1 \end{bmatrix}$

2　画出 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 的投影，也计算 $\boldsymbol{p} = \hat{x}\boldsymbol{a}$:

(a) $\boldsymbol{b} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$　　(b) $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

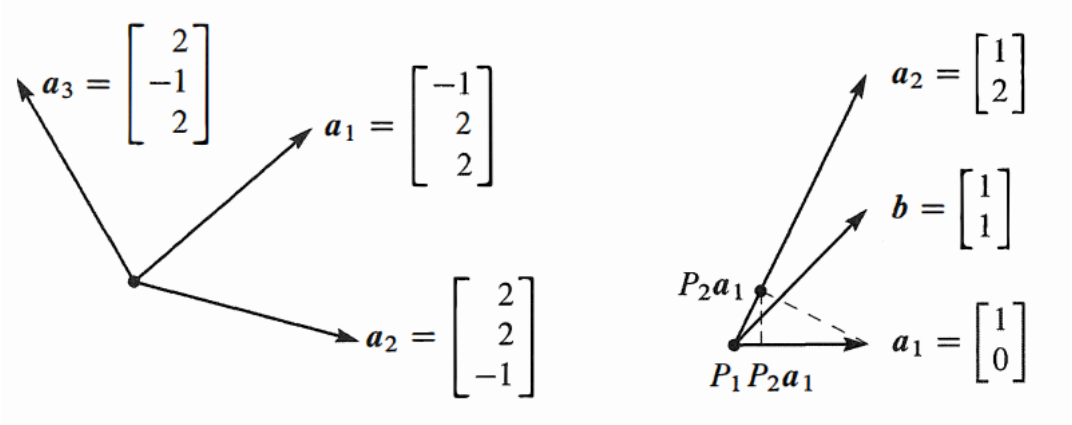
3　问题 1 中，求投影矩阵 $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ 在通过每个 $\boldsymbol{a}$ 的直线的投影。验证两个情形下都有 $P^2 = P$。依据两种情形下的 $P\boldsymbol{b}$ 计算投影 $\boldsymbol{p}$。

4　建立问题 2 中在通过 $\boldsymbol{a}$ 直线的投影矩阵 P_1 与 P_2。$(P_1 + P_2)^2 = P_1 + P_2$ 正确吗? 如果 $P_1P_2 = 0$ 前式才会成立。

5　分别计算在通过 $\boldsymbol{a}_1 = (-1, 2, 2)$的直线以及 $\boldsymbol{a}_2 = (2, 2, -1)$的直线的投影矩阵 $\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。把投影矩阵相乘，并解释为什么他们的乘积 P_1P_2 会是这样的结果?

6　把 $\boldsymbol{b} = (1, 0, 0)$投影在通过问题 5 中的 $\boldsymbol{a}_1$ 与 $\boldsymbol{a}_2$ 的直线，也投影在 $\boldsymbol{a}_3 = (2, -1, 2)$。把三个投影相加 $\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3$。

7　继续问题 5-6，求出在 $\boldsymbol{a}_3 = (2, -1, 2)$的投影矩阵 P_3，验证 $P_1 + P_2 + P_3 = I$。这是因为基底 $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 是正交。



问题 5–6–7：正交　　问题 8–9–10：不正交

8　把 $\boldsymbol{b} = (1, 1)$分别投影在通过 $\boldsymbol{a}_1 = (1, 0)$与 $\boldsymbol{a}_2 = (1, 2)$的直线，画出投影 $\boldsymbol{p}_1$ 与 $\boldsymbol{p}_2$，再计算 $\boldsymbol{p}_1 + \boldsymbol{p}_2$。因为 $\boldsymbol{a}$'s 不正交，投影相加不会等于 $\boldsymbol{b}$。

9　问题 8 中，$\boldsymbol{b}$ 在 $\boldsymbol{a}_1$ 与 $\boldsymbol{a}_2$ 形成的平面的投影会等于 $\boldsymbol{b}$。求 $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$，其中 $A = [a_1 \quad a_2] = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$= 可逆矩阵。

10　把 $\boldsymbol{a}_1 = (1, 0)$投影在 $\boldsymbol{a}_2 = (1, 2)$，然后把得到的结果再投影回到 $\boldsymbol{a}_1$。画出这些投影然后计算投影矩阵的乘积 P_1P_2：这是投影吗?

请尊重版权与译者的劳动成果，侵权必究!

问题 11-20 询问投影，投影矩阵，在子空间。

11 求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\boldsymbol{b}$ 在 A 的列空间的投影 $\boldsymbol{p} = A\hat{\boldsymbol{x}}$。

(a) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ 与 $B = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$　　(b) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$ 与 $B = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

求 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$，它应该与 A 的列垂直。

12 计算问题 11 的两个在列空间的投影矩阵 P_1 与 P_2。验证 $P_1\boldsymbol{b}$ 得到第一个投影 $\boldsymbol{p}_1$，也验证 $P_2^2 = P_2$。

13 (快速与推荐) 假设 A 是把 4×4 单位矩阵移除最后一列的矩阵，所以 A 是 4×3。把 $\boldsymbol{b} = (1, 2, 3, 4)$投影在 A 的列空间，投影矩阵 P 的形状为何？$P =$ ？

14 假设 $\boldsymbol{b}$ 是 2 乘 A 的第一列，$\boldsymbol{b}$ 在 A 的列空间的投影为何？本案例是否 $P = I$？当 $\boldsymbol{b} = (0, 2, 4)$以及 A 的列是$(0, 1, 2)$与$(1, 2, 0)$，计算 $\boldsymbol{p}$ 与 P。

15 如果 A 变成 2 倍，则 $P = 2A(4A^{\mathrm{T}}A)^{-1}\,2A^{\mathrm{T}}$，与 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 相同。$2A$ 的列空间与_____相同，A 与 $2A$ 的 $\hat{\boldsymbol{x}}$ 是否相同？

16 $(1, 2, -1)$与$(1, 0, 1)$的哪个组合最接近 $\boldsymbol{b} = (2, 1, 1)$？

17 (重要) 若 $P^2 = P$ 证明$(I - P)^2 = I - P$。当 P 投影至 A 的列空间时，$I - P$ 会投影至_____。

18 (a) 若 P 是在通过$(1, 1)$的直线的 2×2 投影矩阵，则 $I - P$ 是在_____的投影矩阵。

(b) 若 P 是在通过$(1, 1, 1)$的直线的 3×3 投影矩阵，则 $I - P$ 是在_____的投影矩阵。

19 要求出在平面 $x - y - 2z = 0$ 的投影矩阵，选择平面的两个向量使得他们是 A 的列，这个平面就是 A 的列空间！计算 $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$。

20 要求出在平面 $x - y - 2z = 0$ 的投影矩阵，写出一个与平面垂直的向量 $\boldsymbol{e}$，计算投影 $Q = \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}/\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}$，则 $P = I - Q$。

问题 21-26 展示投影矩阵满足 $\boldsymbol{P^2 = P}$ 与 $\boldsymbol{P^{\mathrm{T}} = P}$。

21 矩阵 $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 乘以自己，消除之后证明 $P^2 = P$。解释为什么 $P(P\boldsymbol{b})$永远等于 $P\boldsymbol{b}$：向量 $P\boldsymbol{b}$ 在 A 的列空间中，所以对列空间的投影是_____。

22 计算 P^{T} 证明 $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 是对称，记得对称矩阵的逆矩阵还是对称。

23 若 A 是方形可逆，对于分割$(A^{\mathrm{T}}A)^{-1}$ 的警告就不适用了，$AA^{-1}\ (A^{\mathrm{T}})^{-1}\ A^{\mathrm{T}} = I$ 是真实的。当 A 可逆，为什么 $P = I$？误差 $\boldsymbol{e}$ 为何？

24 A^{T} 的零空间与列空间 $\boldsymbol{C}(A)$是_____，所以若 $A^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{0}$，$\boldsymbol{b}$ 在 $\boldsymbol{C}(A)$的投影应该是_____。检查 $P = A\ (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 可以得到答案。

25 在 $\mathbf{R}^m$ 的 n 维子空间的投影矩阵的秩 $r = n$。理由：投影 $P\boldsymbol{b}$ 填满子空间 $\boldsymbol{S}$，所以 $\boldsymbol{S}$ 是 P 的_____。

26 若 $m \times m$ 矩阵有 $A^2 = A$ 且秩是 m，证明 $A = I$。

27 结束本段落的重要事实：若 $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$，则 $A\boldsymbol{x} = \boldsymbol{0}$。新的证明：向量 $A\boldsymbol{x}$ 在____的零空间，$A\boldsymbol{x}$ 永远在_____的列空间。要同时在两个垂直的子空间中，$A\boldsymbol{x}$ 必须是零。

28 利用 $P^{\mathrm{T}} = P$ 且 $P^2 = P$，证明列 2 长度的平方永远等于对角单元 P_{22}。这个数字是$\dfrac{2}{6} = \dfrac{4}{36} + \dfrac{4}{36} + \dfrac{4}{36}$：

$$P = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

29 若 B 的秩是 m(满行秩，无关的行)，证明 BB^{T} 是可逆。

挑战问题

30 (a) 求出在 A 的列空间的投影矩阵 P_C (仔细观察矩阵之后！)：

$$A = \begin{bmatrix} 3 & 6 & 6 \\ 4 & 8 & 8 \end{bmatrix}$$

(b) 求出在 A 的行空间的 3×3 投影矩阵 P_R。执行 $B = P_C A P_R$，你的答案 B 会有一点小惊奇–你能解释吗？

31 在$\mathbf{R}^m$，假设我给你 $\boldsymbol{b}$ 以及 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 的组合 $\boldsymbol{p}$，你如何测试 $\boldsymbol{p}$ 是不是 $\boldsymbol{b}$ 在 $\boldsymbol{a}$'s 生成子空间的投影？

32 假设 P_1 是在 A 的第一列生成的 1 维子空间的投影矩阵，假设 P_2 是在 A 的 2 维列空间的投影矩阵。稍微想想，计算 $P_2\,P_1$。

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

33 假设你知道 $b_1, b_2, ..., b_{999}$ 的平均值$\hat{x}_{\text{old}}$，当第 100 个到来，检验新的平均值是$\hat{x}_{\text{old}}$与错位(mismatch)$b_{1000} - \hat{x}_{\text{old}}$ 的组合：

$$\hat{x}_{\text{new}} = \frac{b_1 + \ldots + b_{1000}}{1000} = \frac{b_1 + \ldots + b_{999}}{999} + \frac{1}{1000}\left(b_{1000} - \frac{b_1 + \ldots + b_{999}}{999}\right)$$

这是“卡尔曼滤波器” $\hat{x}_{\text{new}} = \hat{x}_{\text{old}} + (b_{1000} - \hat{x}_{\text{old}})/1000$，其中增益矩阵 1/1000。本书的最后一页把卡尔曼滤波器延伸至矩阵的更新。

34 (2017) 假设 P_1 与 P_2 是投影矩阵 $(P_i^2 = P_i = P_i^{\text{T}})$。证明这个事实：

$$P_1P_2 \text{ 是投影矩阵 当且仅当 } P_1P_2 = P_2P_1$$

4.3 最小二乘近似

1　求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\boldsymbol{b}$ 在 A 的列空间的投影 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$。

2　当 $A\boldsymbol{x}=\boldsymbol{b}$ 无解，$\hat{\boldsymbol{x}}$是“最小二乘解”：$\|\boldsymbol{b}-A\hat{\boldsymbol{x}}\|^2=$ 极小值。

3　令 $\boldsymbol{E}=\|\boldsymbol{Ax}-\boldsymbol{b}\|^2$ 的部分导数为 0 $(\frac{\partial E}{\partial x_i}=0)$也会产生 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$。

4　利用一条直线去适配点$(t_1, b_1), ..., (t_m, b_m)$，$A$ 有$(1, ..., 1)$与$(t_1, ..., t_m)$的列。

5　在上述的情形下，$A^{\mathrm{T}}A$ 是 2×2 矩阵$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}$，$A^{\mathrm{T}}\boldsymbol{b}$ 是向量$\begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$。

$A\boldsymbol{x}=\boldsymbol{b}$ 经常是无解，一般的理由是方程式太多。矩阵 A 的行比列多，方程式的个数大于未知数的个数(m 大于 n)。这 n 个列只生成了 m 维空间的一小部分，除非所有的量测值都很完美，否则 $\boldsymbol{b}$ 会落在 A 的列空间之外。消元法得到不可能存在的方程式然后停止，但是我们不能停止，因为量测值必然包含杂讯。

重复：我们不可能一直把误差 $\boldsymbol{e}=\boldsymbol{b}-A\boldsymbol{x}$ 降为零，当 $\boldsymbol{e}$ 是零，$\boldsymbol{x}$ 是 $A\boldsymbol{x}=\boldsymbol{b}$ 的确切解。当 $\boldsymbol{e}$ 的长度尽可能的小，$\hat{\boldsymbol{x}}$ 就是“最小二乘解(least squares solution)”。我们在本段落的目标是计算并且使用 $\hat{\boldsymbol{x}}$，还有很多实际的问题需要解答。

前一个段落强调 $\boldsymbol{p}$(投影)，本段落强调$\hat{\boldsymbol{x}}$ (最小二乘解)，他们之间借由 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$ 得到关联，基础方程式还是 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$。此处有一个非正式的方法得到“正态方程式”：

当 $A\boldsymbol{x}=\boldsymbol{b}$ 无解，左乘 A^{T} 然后求解　$A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$。

范例 1　最小二乘的一个重要应用是对于 m 个点适配一条直线，从 3 个点开始：求最接近点(0, 6), (1, 0), (2, 0)的直线。

没有直线 $b=C+Dt$ 通过这三个点，我们现在要求两个数字 C 与 D 满足三个方程式：$n=2$ 且 $m=3$。此处的三个方程式在 $t=0, 1, 2$ 要适配给定的值 $b=6, 0, 0$：

$t=0$　　若 $C+D\cdot 0=6$ 则第一个点在直线 $b=C+Dt$

$t=1$　　若 $C+D\cdot 1=0$ 则第二个点在直线 $b=C+Dt$

$t=2$　　若 $C+D\cdot 2=0$ 则第三个点在直线 $b=C+Dt$

请尊重版权与译者的劳动成果，侵权必究!

这个 3×2 的系统无解：$\boldsymbol{b} = (6, 0, 0)$不是列$(1, 1, 1)$与$(0, 1, 2)$的一个组合。从这些方程式得到 A，$\boldsymbol{x}$ 与 $\boldsymbol{b}$ ：

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \qquad A\boldsymbol{x} = \boldsymbol{b} \text{ 无解}$$

同样的数字在前个段落范例 3 出现过，我们计算$\hat{\boldsymbol{x}} = (5, -3)$。这些数字是最佳的 C 与 D，所以 $5 - 3t$ 是这三个点的最佳直线。我们必须让投影与最小二乘产生关联，这就要说明为什么 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

实际的问题上，可能轻易有 100 个点而不是 3 个点，他们没有办法适配任何的直线 $C + Dt$。我们的数字 6, 0, 0 夸大了误差，你可以在图 4.6 看到 e_1, e_2 与 e_3。

极小化误差

我们如何使得误差 $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$ 尽可能的小？这是一个有着美丽答案的重要问题。最佳 $\boldsymbol{x}$(称为$\hat{\boldsymbol{x}}$)可以从**几何学**(误差 $\boldsymbol{e}$ 与 A 的列空间的夹角是 90°)。也可以从**代数**求得：$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。**微积分**也有相同的答案$\hat{\boldsymbol{x}}$：误差$\| A\boldsymbol{x} - \boldsymbol{b} \|^2$ 的导数在$\hat{\boldsymbol{x}}$为 0。

从几何学　每个 $A\boldsymbol{x}$ 都落在列$(1, 1, 1)$与$(0, 1, 2)$的平面，在这个平面上，我们寻找最接近 $\boldsymbol{b}$ 的点，这个“最近点”就是投影 $\boldsymbol{p}$。

$A\hat{\boldsymbol{x}}$ 最佳选择是 $\boldsymbol{p}$，最小可能的误差是 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$，与这些列垂直。这三个在高度$(p_1, p_2, p_3)$的点确实在同一个平面，因为 $\boldsymbol{p}$ 在 A 的列空间中。当我们适配一条直线时，$\hat{\boldsymbol{x}}$是(C, D)的最佳选择。

从代数　每个向量 $\boldsymbol{b}$ 分成两部分，在列空间的部分是 $\boldsymbol{p}$，垂直部分是 $\boldsymbol{e}$。我们无法求解这个方程式$(A\boldsymbol{x} = \boldsymbol{b})$，但是我们可以也确实求解一个方程式 $A\hat{\boldsymbol{x}} = \boldsymbol{p}$ (移除 $\boldsymbol{e}$ 然后求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$)：

$$A\boldsymbol{x} = \boldsymbol{b} = \boldsymbol{p} + \boldsymbol{e} \quad \text{不可能求解；} \quad A\hat{\boldsymbol{x}} = \boldsymbol{p} \text{ 有解：} \quad \hat{\boldsymbol{x}} \text{ 是} (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b} \tag{1}$$

$A\hat{\boldsymbol{x}} = \boldsymbol{p}$ 的解得到最小可能的误差(就是 $\boldsymbol{e}$)：

任意 $\boldsymbol{x}$ 的长度平方　　$$\| A\boldsymbol{x} - \boldsymbol{b} \|^2 = \| A\boldsymbol{x} - \boldsymbol{p} \|^2 + \| \boldsymbol{e} \|^2 \tag{2}$$

这是直角三角形公式 $c^2 = a^2 + b^2$，列空间的向量 $A\boldsymbol{x} - \boldsymbol{p}$ 与左零空间中的 $\boldsymbol{e}$ 垂直。选择 $\boldsymbol{x} = \hat{\boldsymbol{x}}$，我们把 $A\boldsymbol{x} - \boldsymbol{p}$ 降低至 0，这就留下了最小可能误差 $\boldsymbol{e} = (e_1, e_2, e_3)$，这是我们不能降低的。

注意“最小”的意义是什么，　$A\boldsymbol{x} - \boldsymbol{p}$ 的长度平方被极小化。

最小二乘解 $\hat{\boldsymbol{x}}$使得 $\boldsymbol{E} = \| A\boldsymbol{x} - \boldsymbol{b} \|^2$ 尽可能最小。

图 4.6a 显示最近线，它的错失距离是 $e_1, e_2, e_3 = 1, -2, 1$，这些都是**铅垂线**距离。这条最小二乘线极小化 $E = e_1^2 + e_2^2 + e_3^2$。

图 4.6b 显示在三维空间($\boldsymbol{b}$, $\boldsymbol{p}$, $\boldsymbol{e}$ 空间)相同的问题，向量 $\boldsymbol{b}$ 不在 A 的列空间，这是为什么不能求解 $A\boldsymbol{x} = \boldsymbol{b}$ 的原因，没有直线穿过这三个点。最小可能误差是垂直向量 $\boldsymbol{e}$，这是 $\boldsymbol{e} = \boldsymbol{b} - A\hat{\boldsymbol{x}}$，误差向量(1, –2, 1)在这三个方程式中，就是与最佳直线的距离。潜藏在两个图形背后的是基础方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

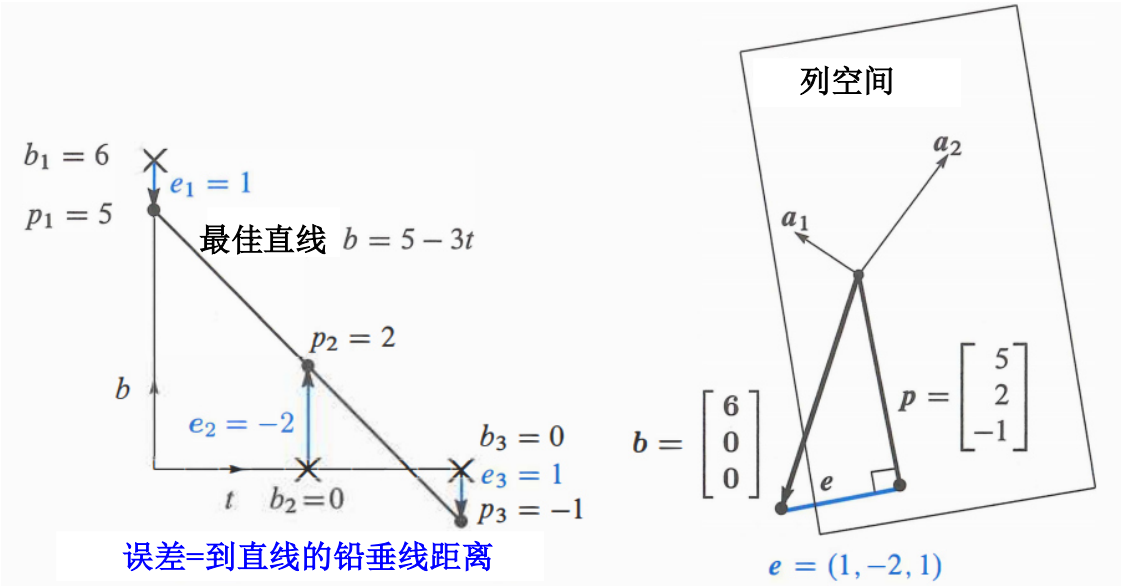


图 4.6：**最佳直线与投影：两个图形，同样问题。**直线有高度 $\boldsymbol{p} = (5, 2, -1)$且误差 $\boldsymbol{e} = (1, -2, 1)$。方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\hat{\boldsymbol{x}} = (5, -3)$，相同答案！最佳直线是 $b = 5 - 3t$ 且最近点是 $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\,\boldsymbol{a}_2$。

注意误差 1, –2, 1 加起来等于 0。理由：误差 $\boldsymbol{e} = (e_1, e_2, e_3)$与 A 的第一列(1, 1, 1)垂直，点积得到 $e_1 + e_2 + e_3 = 0$。

从微积分　大部分函数的极小化都是利用微积分！图形降到最低点而且在每一个方向的导数为零。此处要极小化的误差函数 E 是平方的总和 $e_1^2 + e_2^2 + e_3^2$ (每个方程式中误差的平方)：

$$E = \| A\boldsymbol{x} - \boldsymbol{b} \|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1)^2 + (C + D \cdot 2)^2 \qquad (3)$$

未知数是 C 与 D，两个未知数配上两个导数–在极小值时都是 0。他们是“部分导数”，因为$\partial E/\partial C$ 把 D 视为常数，$\partial E/\partial D$ 把 C 视为常数：

$$\partial E/\partial C = 2(C + D \cdot 0 - 6) \quad + 2(C + D \cdot 1) \quad + 2(C + D \cdot 2) \quad = 0$$
$$\partial E/\partial D = 2(C + D \cdot 0 - 6)(0) + 2(C + D \cdot 1)\,(1) + 2(C + D \cdot 2)(2) = 0$$

$\partial E/\partial D$ 包含来自链条法则的额外因子 0, 1, 2。(最后一个$(C + D \cdot 2)^2$ 的导数是 2 乘 $C + 2D$ 乘额外的 2。) 在$\partial E/\partial C$ 时，这些因子只是 1, 1, 1 而已。

请尊重版权与译者的劳动成果，侵权必究!

毫无意外的$\| A\boldsymbol{x}-\boldsymbol{b} \|^2$的导数的因子 1, 1, 1 与 0, 1, 2 是 A 的列。现在消去每一项的 2 然后收集全部的 C's 与全部的 D's：

$$\begin{matrix} C\text{导数是零：}3C+3D=6 \\ D\text{导数是零：}3C+5D=0 \end{matrix} \qquad \text{矩阵}\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}\text{是}A^{\mathrm{T}}A \tag{4}$$

这些方程式与 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 完全相同，最佳的 C 与 D 是$\hat{\boldsymbol{x}}$ 分量。来自微积分的方程式与来自线性代数的正态方程式完全相同，这些是最小二乘的关键方程式：

> 当 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$，$\| A\boldsymbol{x}-\boldsymbol{b} \|^2$ 的部分导数是零。

解是 $C=5$ 与 $D=-3$，因此 $b=5-3t$ 是最佳直线--它与这三个点最接近。在 $t=0$, 1, 2 时，这条线通过 $\boldsymbol{p}=5, 2, -1$，它不会通过 $\boldsymbol{b}=6, 0, 0$。误差是 1, –2, 1，这是向量 $\boldsymbol{e}$。

最小二乘的大图

本书的关键图形显示 4 个子空间与矩阵的真实作用，图 4.3 左侧的向量 $\boldsymbol{x}$ 走到右侧的 $\boldsymbol{b}=A\boldsymbol{x}$，在这个图形中 $\boldsymbol{x}$ 分割成 $\boldsymbol{x}_r+\boldsymbol{x}_n$，存在很多 $A\boldsymbol{x}=\boldsymbol{b}$ 的解。

本段落的情况刚好相反，$A\boldsymbol{x}=\boldsymbol{b}$ 无解。我们不分割 $\boldsymbol{x}$，而是分割 $\boldsymbol{b}=\boldsymbol{p}+\boldsymbol{e}$。图 4.7 显示最小二乘的大图，我们求解 $A\hat{\boldsymbol{x}}=\boldsymbol{p}$ 而不是求解 $A\boldsymbol{x}=\boldsymbol{b}$，误差 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 是无法避免的。

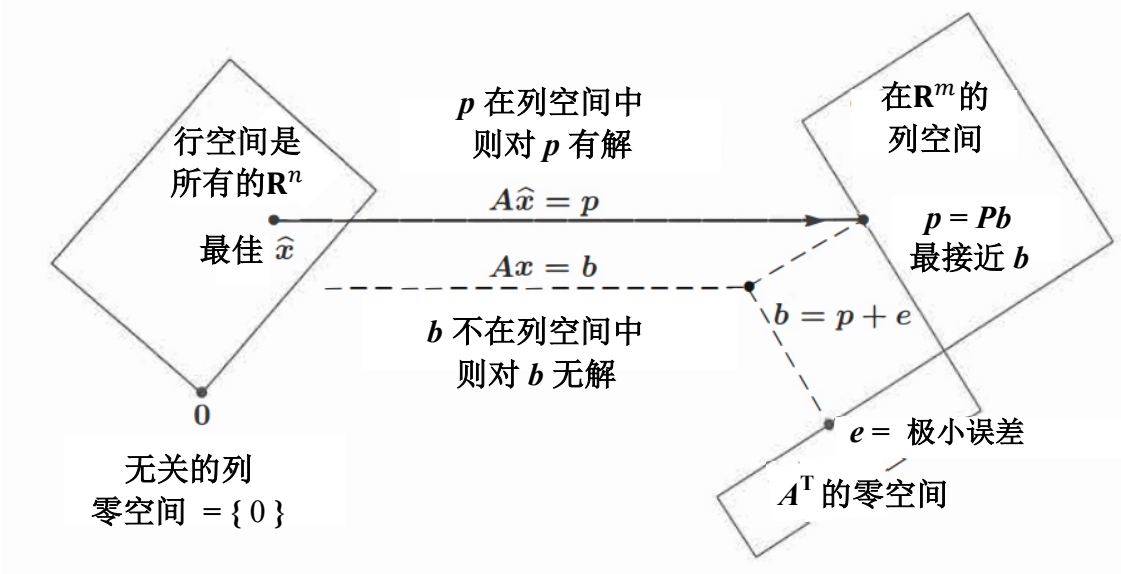


图 4.7：投影 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$ 接近 $\boldsymbol{b}$，所以 $\hat{\boldsymbol{x}}$ 小化 $E=\| A\boldsymbol{x}-\boldsymbol{b} \|^2$。

注意零空间 $\boldsymbol{N}(A)$是如何的小--只有一个点，由于无关的列，$A\boldsymbol{x}=\boldsymbol{0}$ 的唯一解是 $\boldsymbol{x}=\boldsymbol{0}$，方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}$ $A^{\mathrm{T}}\boldsymbol{b}$ 完全决定了最佳向量$\hat{\boldsymbol{x}}$。误差有 $A^{\mathrm{T}}\boldsymbol{e}=\boldsymbol{0}$。

第七章会有完整的图形–包含所有的 4 个子空间。每个向量 $\boldsymbol{x}$ 分割成 $\boldsymbol{x}_r + \boldsymbol{x}_n$，每个 $\boldsymbol{b}$ 分割成 $\boldsymbol{p} + \boldsymbol{e}$，最佳解是在行空间的$\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_r$。我们不能帮助 $\boldsymbol{e}$，我们不要来自零空间的 $\boldsymbol{x}_n$ —— 这就留下了 $A\hat{\boldsymbol{x}} = \boldsymbol{p}$。

适配一条直线

适配一条线是最小二乘最清晰的应用，从 $m > 2$ 个点开始，希望能够接近一条直线。在时间点 $t_1, ..., t_m$，那 m 个点的高度在 $b_1, ..., b_m$。最佳直线 $C + Dt$ 错失这些点的垂直距离是 $e_1, ..., e_m$。没有直线是完美的，最小二乘线极小化 $E = e_1^2 + ... + e_m^2$。

第一个例子是本段落图 4.6 所示的三个点问题，现在我们允许 m 个点(m 可以是很大)，$\hat{\boldsymbol{x}}$ 的两个分量仍然是 C 与 D。

当我们确切求解 $A\boldsymbol{x} = \boldsymbol{b}$ 时，这条直线会通过这 m 个点，一般情况下我们无法做到。两个未知数 C 与 D 决定一条线，所以 A 只有 $n = 2$ 个列。要想适配 m 个点，我们尝试求解 m 个方程式(我们只有两个未知数！)：

$$A\boldsymbol{x} = \boldsymbol{b} \text{ 是 } \begin{matrix} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \vdots \\ C + Dt_m = b_m \end{matrix} \text{ 其中 } A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \tag{5}$$

列空间是如此的薄，使得大部分的 $\boldsymbol{b}$ 都落在列空间之外。当 $\boldsymbol{b}$ 碰巧在列空间中，点也碰巧落在直线上，此时 $\boldsymbol{b} = \boldsymbol{p}$，则 $A\boldsymbol{x} = \boldsymbol{b}$ 有解，误差是 $\boldsymbol{e} = (0, ..., 0)$。

最近线 $C + Dt$ 的高度 $p_1, ..., p_m$，误差是 $e_1, ..., e_m$。

求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 得到 $\hat{\boldsymbol{x}} = (C, D)$，误差是 $e_i = b_i - C - Dt_i$。

直线上的适配点非常重要，我们一劳永逸的给定两个方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。$A$ 的两个列是无关的(除非所有的时间点 t_i 都相同)，所以我们转移到最小二乘并且求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

$$\text{点积矩阵 } A^{\mathrm{T}}A = \begin{bmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \tag{6}$$

正态方程式的右侧是 2×1 向量 $A^{\mathrm{T}}\boldsymbol{b}$：

$$A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{7}$$

在特定问题中，这些数字是给定的。最佳$\hat{\boldsymbol{x}} = (C, D) = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$。

当 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 时，直线 $C + Dt$ 极小化 $e_1^2 + ... + e_m^2 = \| A\boldsymbol{x} - \boldsymbol{b} \|^2$：

请尊重版权与译者的劳动成果，侵权必究！

$$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \qquad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix} \tag{8}$$

直线上 m 个点的垂直误差是 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ 的分量，这个误差向量(残余)$\boldsymbol{b} - A\hat{\boldsymbol{x}}$ 与 A 的列垂直(几何学)，误差在 A^{T} 的左零空间中(线性代数)。最佳$\hat{\boldsymbol{x}} = (C, D)$极小化总误差 E，E 是平方的和(微积分)：

$$E(\boldsymbol{x}) = \| A\boldsymbol{x} - \boldsymbol{b} \|^2 = (C + Dt_1 - b_1)^2 + \ldots + (C + Dt_m - b_m)^2$$

微积分令$\partial E/\partial C$ 与$\partial E/\partial D$ 等于 0，以及 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

另一个最小方程式问题的未知数比 2 来得多，适配最佳抛物线要有 $n = 3$ 个系数 C, D, E(见下方)。一般来说，我们用 n 个参数 $x_1, \ldots, x_n$ 适配 m 个资料点，矩阵 A 有 n 个列且 $n < m$。$\| A\boldsymbol{x} - \boldsymbol{b} \|^2$ 的导数得到 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$，平方的导数是线性的，这是为什么最小二乘会流行的原因。

范例 2　当量测时间点 t_i 的总和为零时，假设在时间点 $t = -2, 0, 2$，$b = 1, 2, 4$，这些时间点的总和是零。A 的列有零的点积：(1, 1, 1)与(–2, 0, 2)正交：

$$\begin{aligned} C + D(-2) &= 1 \\ C + D(0) &= 2 \\ C + D(2) &= 4 \end{aligned} \quad \text{或} \quad A\boldsymbol{x} = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

当 A 有正交的列，$A^{\mathrm{T}}A$ 会是一个对角矩阵：

$$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \text{ 是 } \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \tag{9}$$

重点：由于 $A^{\mathrm{T}}A$ 是对角，可以分开求解 $C = 7/3$ 与 $D = 6/8$，在 $A^{\mathrm{T}}A$ 中的零是 A 的垂直列的点积。当单元数 $m = 3$ 且 $t_1^2 + t_2^2 + t_3^2 = 8$，对角矩阵 $A^{\mathrm{T}}A$ 基本上与单位矩阵有差不多的性质。

正交矩阵非常有帮助，所以值得我们借由减去时间平均值 $\hat{t} = (t_1 + \ldots + t_m)/m$ 完成时间点的平移。如果原始时间点是 1, 3, 5，则平均值是 $\hat{t} = 3$，平移时间是 $T = t - \hat{t} = t - 3$，总和变成零！

$$\begin{aligned} T_1 &= 1 - 3 = -2 \\ T_2 &= 3 - 3 = 0 \\ T_3 &= 5 - 3 = 2 \end{aligned} \qquad A_{\text{new}} = \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ 1 & T_3 \end{bmatrix} \qquad A^{\mathrm{T}}{}_{\text{new}}\, A_{\text{new}} = \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}$$

现在 C 与 D 直接来自简单的方程式(9)，则最佳直线 $C + DT$ 就是 $C + D(t - \hat{t}) = C + D(t - 3)$。问题 30 甚至给定一个 C 与 D 的公式。

那是一个“格莱姆-施密特概念”的完美例子，下个段落会介绍：事先使得列变成正交，则 $A^{\mathrm{T}}_{\mathrm{new}}A_{\mathrm{new}}$ 是对角且$\hat{\boldsymbol{x}}_{\mathrm{new}}$变得简单。

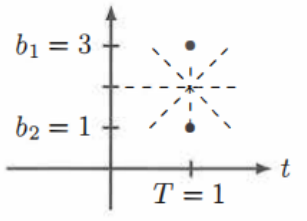
A 的相关列：$\hat{x}$ 是什么?

本章从头开始一直假设 A 的列是无关，$A^{\mathrm{T}}A$ 是可逆，$A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 得到 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解。

如果 A 的列是相关，什么样的$\hat{\boldsymbol{x}}$是最佳？下面是个特殊的例子。

$$\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}=\boldsymbol{b} \qquad A\boldsymbol{x}=\boldsymbol{b}$$

$$\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}1\\2\end{bmatrix}=\boldsymbol{p} \qquad A\hat{\boldsymbol{x}}=\boldsymbol{p}$$



量测值 $b_1=3$ 与 $b_2=1$ 在相同的时间点 T！直线 $C+Dt$ 不可能同时通过两个点，我想我们把 $\boldsymbol{b}=(3,1)$投影到 A 的列空间的 $\boldsymbol{p}=(2,2)$是正确的，这样把方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 变成 $A\hat{\boldsymbol{x}}=\boldsymbol{p}$，一个无解的方程式变成一个有无限多解的方程式，问题在于 A 有相关的列以及$(1,-1)$在它的零空间中。

我们应该选择哪个解$\hat{\boldsymbol{x}}$？图中所有的虚线在时间点 T 有相同的两个误差 1 与 -1，误差$(1,-1)=\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 要尽可能的小，但是没有告诉我们哪条虚线是最佳。

我的直觉是选择在高度 2 的水平线，如果最佳直线的方程式是 $b=C+Dt$，我选择$\hat{x}_1=C=2$ 与$\hat{x}_2=D=0$。但是如果直线写成 $b=ct+d$ 呢？这也是同样正确的(只是反转 C 与 D)。现在水平线有$\hat{x}_1=c=0$ 与 $\hat{x}_2=d=2$。我看不到其他出路。

在段落 7.4，A 的“伪逆矩阵(pseudoinverse)” 会选出 $A\hat{\boldsymbol{x}}=\boldsymbol{p}$ 的最短长度解。此处最短解是 $\boldsymbol{x}^{+}=(1,1)$，这是在 A 的行空间中的特解，且 $\boldsymbol{x}^{+}$的长度是$\sqrt{2}$。(两个解$\hat{\boldsymbol{x}}=(2,0)$与$(0,2)$的长度都是 2。) 我们是任意选取解 $\boldsymbol{x}^{+}$的零空间分量为零。

当 A 有无关的列，零空间只包含零向量且伪逆矩阵是我们正常的左逆矩阵 $L=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$。当我写成这个形式，伪逆矩阵听起来好像是选取 $\boldsymbol{x}$ 的最佳方式。

注解 MATLAB 的奇异矩阵试验不是得到 **Inf** 就是 **NaN**(不是一个数)或是 $\mathbf{10^{16}}$(不好的数)，每个情况都会有警告出现！我相信 **Inf** 与 **NaN** 与 $\mathbf{10^{16}}$ 都来自可能的 $\mathbf{0}\boldsymbol{x}=\boldsymbol{b}$ 与 $\mathbf{0}\boldsymbol{x}=\mathbf{0}$ 与 $\mathbf{10^{-16}}\boldsymbol{x}=\mathbf{1}$。

那些是 3 个小范例来自 3 个大困难：奇异而无解，奇异而无限多解以及非常非常接近奇异。

请尊重版权与译者的劳动成果，侵权必究!

适配一条抛物线

如果我们丢一颗球，想要用一条直线适配球的路径那是疯狂的行为，一条抛物线 $b = C + Dt + Et^2$ 允许球向上再向下(b 是在时间点的高度)。实际的路径不是完美的抛物线，但是投影的全部定理从近似开始。

当伽利略从比萨斜塔丢下一块石头，它会加速，距离包含一个二次项 $gt^2/2$。(伽利略的观点是与石头的重量无关。) 缺少 t^2 项我们无法将卫星送进它的轨道。但是纵使出现 t^2 的非线性函数，未知数 C, D 与 E 仍然是线性的出现！最佳抛物线的适配点仍然是线性代数的一个问题。

问题 在时间点 $t_1, ..., t_m$ 的高度是点 $b_1, ..., b_m$，利用抛物线适配这些点。

解 有 $m > 3$ 个点，确切适配的 m 个方程式一般是无解：

$$\begin{matrix} C + Dt_1 + Et_1^2 = b_1 \\ \vdots \\ C + Dt_m + Et_m^2 = b_m \end{matrix} \quad \text{是 } A\boldsymbol{x} = \boldsymbol{b}\text{，矩阵是 } m\times 3 \quad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix} \tag{10}$$

最小二乘 最接近的抛物线 $C + Dt + Et^2$ 选取 $\hat{\boldsymbol{x}} = (C, D, E)$ 来满足三个正态方程式。$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

我可以问你如何把这个变成投影的问题吗？A 的列空间的维度是_____，$\boldsymbol{b}$ 的投影是 $\boldsymbol{p} = A\hat{\boldsymbol{x}}$，其中 $\boldsymbol{p}$ 是用系数 C, D, E 来组合三个列。第一个资料点的误差是 $e_1 = b_1 - C - Dt - Et^2$，误差平方的总和 $e_1^2 +$ _____。如果你喜欢用微积分的极小化，就计算 E 相对于____，____，____的部分导数，当 $\hat{\boldsymbol{x}} = (C, D, E)$ 是 3×3 方程式系统 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 的解时，这三个导数等于零。

段落 10.5 有更多最小二乘的应用，最大的应用是傅里叶级数--**近似函数**而不是近似向量。要极小化的函数从平方总和 $e_1^2 + ... + e_m^2$ 变成误差平方的积分。

范例 3 一个抛物线 $b = C + Dt + Et^2$，在时间点 $t = 0, 1, 2$ 时，通过三个高度 $b = 6, 0, 0$。C, D, E 的方程式:

$$\begin{aligned} C + D\cdot 0 + E\cdot 0^2 &= 6 \\ C + D\cdot 1 + E\cdot 1^2 &= 0 \\ C + D\cdot 2 + E\cdot 2^2 &= 0 \end{aligned} \tag{11}$$

这是 $A\boldsymbol{x} = \boldsymbol{b}$，我们可以确切的求解。这 3 个资料点得到 3 个方程式与 1 个方形矩阵，解是 $\boldsymbol{x} = (C, D, E) = (6, -9, 3)$。通过这三点的抛物线如图 4.8a 所示是 $b = 6 - 9t + 3t^2$。

上述的结果对投影来说表示什么？矩阵有 3 个列，生成整个$\mathbf{R}^3$空间，投影矩阵是单位矩阵，$\boldsymbol{b}$ 的投影是 $\boldsymbol{b}$，误差是零。我们不需要$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$，因为我们可以求解 $A\boldsymbol{x} = \boldsymbol{b}$。当然我可以左乘 A^{T}，但是没有道理。

图 4.8 页显示了在时间点 t_4 的第四点 b_4，如果那个点落在抛物线上，新的 $A\boldsymbol{x} = \boldsymbol{b}$(4 个方程式)还是有解。当第四个点不在抛物线上，我们转移到$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。那么最小二乘抛物线会保持一样，使得所有的误差都落在第四点吗？不一定哦！

最小二乘会平衡四个误差得到三个 C, D, E 的方程式。

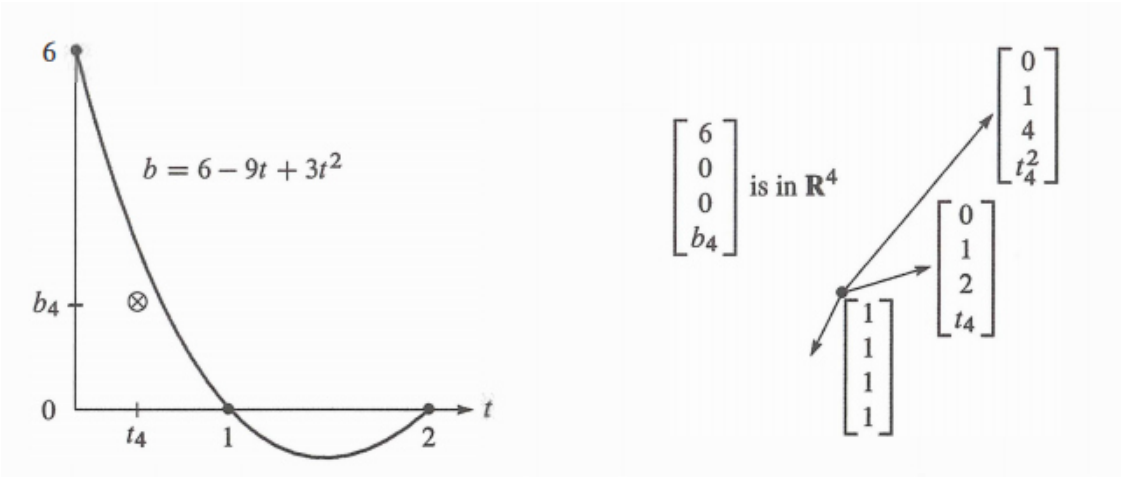


图 4.8：抛物线在 $t = 0, 1, 2$ 确切的适配表示 $\boldsymbol{p} = \boldsymbol{b}$ 且 $\boldsymbol{e} = \boldsymbol{0}$。第四个点⊗ 离开抛物线使得 $m > n$ 以及我们需要最小二乘：把 $\boldsymbol{b}$ 投影到 $\boldsymbol{C}(A)$。右侧的图形显示 $\boldsymbol{b}$––不是 A 的 3 个列的组合。

主要观念的复习

1. 最小二乘解$\hat{\boldsymbol{x}}$ 会极小化$\| A\boldsymbol{x} - \boldsymbol{b} \|^2 = \boldsymbol{x}^{\mathrm{T}} A^{\mathrm{T}}A\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$。

   这是 E，m 个方程式的误差的平方总和$(m > n)$。

2. 最佳$\hat{\boldsymbol{x}}$ 来自正态方程式$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$，$E$ 是极小值。

3. 利用直线 $b = C + Dt$ 适配 m 个点，正态方程式可以得到 C 与 D。

4. 最佳直线的高度是 $\boldsymbol{p} = (p_1, ..., p_m)$，与资料点的垂直距离是误差 $\boldsymbol{e} = (e_1, ..., e_m)$，关键方程式是 $A^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{0}$。

5. 如果我们尝试借由 n 个函数的组合去适配 m 个点$(n < m)$，这 m 个方程式 $A\boldsymbol{x} = \boldsymbol{b}$ 通常都是无解。这 n 个方程式$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 得到最小二乘解––最小 MSE(均方值，mean square value)的组合。

请尊重版权与译者的劳动成果，侵权必究！

已解范例

4.3A 从 9 个时间点 $t = 1, ..., 9$ 开始，这 9 个量测值 b_1 到 b_9 全部是零，第 10 个量测值 $b_{10} = 40$ 是一个局外者。请找出最佳水平线 $y = C$ 可以适配这 10 个点(1, 0), (2, 0), ..., (9, 0), (10, 40)，其中对于误差 E 使用三种选项：

(1) 最小二乘 $E_2 = e_1^2 + ... + e_{10}^2$ (C 的正态方程式是线性)

(2) 最小极大误差 $E_\infty = |e_{\max}|$ (3) 最小误差和 $E_1 = |e_1| + ... + |e_{10}|$

解 (1) 借由水平线 $C = 4$ 得到最小二乘适配 0, 0, ..., 0, 40：

$$A = 1\text{'s 的列} \qquad A^{\mathrm{T}}A = b_i \text{ 的总和 } = 40 \qquad \text{所以 } 10C = 40$$

(2) 最小极大误差(least maximum error)需要 $C = 20$，在 0 与 40 的半程。

(3) 最小误差和需要 $C = 0$。若 C 从零往上增，误差和是 $9|C| + |40 - C|$会增加。

最小和来自中位(median)量测(0, ..., 0, 40 的中位数)，许多统计学家觉得最小二乘解会被局外者严重影响，比如 $b_{10} = 40$，他们比较偏好最小和，但是方程式变成非线性。

现在找出最小二乘直线 $C + Dt$ 通过这 10 个点(1, 0)到(10, 40):

$$A^{\mathrm{T}}A = \begin{bmatrix} 10 & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} = \begin{bmatrix} 10 & 55 \\ 55 & 385 \end{bmatrix} \qquad A^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix} = \begin{bmatrix} 40 \\ 400 \end{bmatrix}$$

这些来自方程式(8)，则$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 得到 $C = -8$ 且 $D = 24/11$。

如果你对 $\boldsymbol{b} = (0, 0, ..., 40)$乘 3 后再加 30 得到 $\boldsymbol{b}_{\text{new}} = (30, 30, ..., 150)$，此时对于 C 与 D 会发生什么？线性允许我们重设 $\boldsymbol{b}$ 的比例，3 乘 $\boldsymbol{b}$ 会得到 3 乘 C 与 3 乘 D。加 30 到所有的 b_i 相当于 C 加 30。

4.3B 在时间点 $t = -2, -1, 0, 1, 2$ 得到值 $\boldsymbol{b} = (0, 0, 1, 0, 0)$，求出抛物线 $b = C + Dt + Et^2$ 来自前述量测的最近值(最小二乘误差)。首先写下 5 个方程式 $A\boldsymbol{x} = \boldsymbol{b}$，其中有 3 个未知数 $\boldsymbol{x} = (C, D, E)$，目的是希望抛物线通过这 5 个点。这个必然无解，因为没有这样的抛物线存在，只能求解$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

我会预测 $D = 0$。为什么最佳抛物线是会环绕 $t = 0$ 对称？在$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 中针对 D 的方程式 2 必须与方程式 1 与 3 解耦。

解 $A\boldsymbol{x} = \boldsymbol{b}$ 的 5 个方程式具有矩形的范德蒙德矩阵 A：

$$\begin{aligned} C + D\cdot(-2) + E\cdot(-2)^2 &= 0 \\ C + D\cdot(-1) + E\cdot(-1)^2 &= 0 \\ C + D\cdot(0) + E\cdot(0)^2 &= 1 \\ C + D\cdot(1) + E\cdot(1)^2 &= 0 \\ C + D\cdot(2) + E\cdot(2)^2 &= 0 \end{aligned} \qquad A = \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$$

那些 $A^{\mathrm{T}}A$ 的零表示 A 的列 2 与列 1 及列 3 垂直，我们直接在 A 可以看到(时间点–2, –1, 0, 1, 2 是对称)。抛物线 $C + Dt + Et^2$ 的最佳 C, D, E 来自 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$，且 D 与 C 及 E 是解耦的：

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{得到} \quad \begin{array}{l} C = 34/70 \\ D = 0，\text{如同预测} \\ E = -10/70 \end{array}$$

问题集 4.3

问题 1-11 利用 4 个资料点 $\boldsymbol{b} = (0, 8, 8, 20)$带出关键概念。

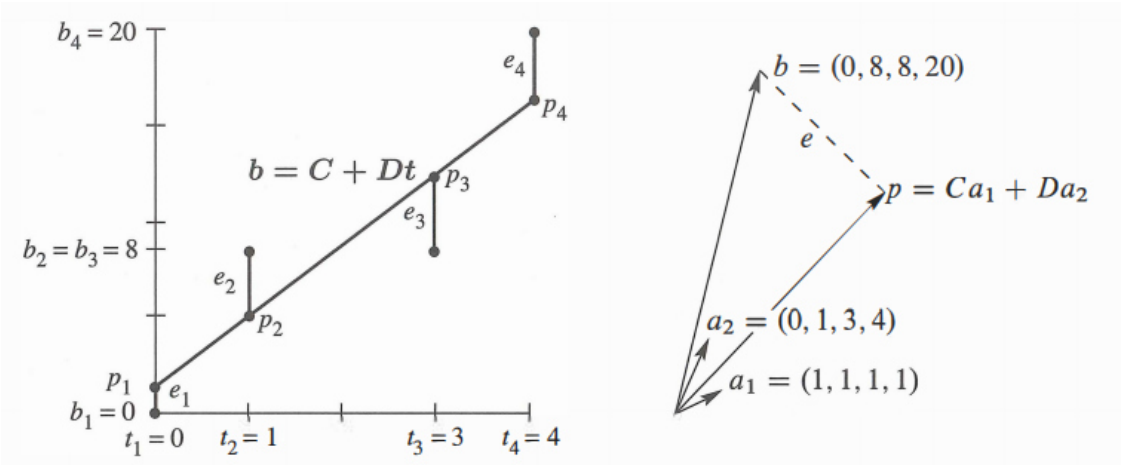


图 4.9：**问题 1-11**：最近线 $C + Dt$ 适配 $\mathbf{R}^4$ 中的 $C\boldsymbol{a}_1 + D\boldsymbol{a}_2$

1 在 $t = 0, 1, 3, 4$ 有 $b = 0, 8, 8, 0$，设定并求解正态方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。对于图 4.9a 的直线求出 4 个高度 p_i 与 4 个误差 e_i。极小值 $E = e_1^2 + e_2^2 + e_3^2 + e_4^2$ =?

2 (直线 $C + Dt$ 确实通过 p's) 在 $t = 0, 1, 3, 4$ 有 $b = 0, 8, 8, 0$，写下 4 个方程式 $A\boldsymbol{x} = \boldsymbol{b}$ (无解)。变更量测值成为 $p = 1, 5, 13, 17$，求 $A\hat{\boldsymbol{x}} = \boldsymbol{p}$ 的确切解。

3 检验 $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1, 3, -5, 3)$与同样 A 的两个列垂直。从 $\boldsymbol{b}$ 到 A 的列空间的最短距离 $\|\boldsymbol{e}\|$ 为何？

4 (使用微积分)写下 $E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ 作为 4 个平方的总和–前一个是 $(C + 4D - 20)^2$。求出导数方程式 $\partial E/\partial C = 0$ 及 $\partial E/\partial D = 0$。除 2 得到正态方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。

5 求出最佳水平线的高度 C 适配 $\boldsymbol{b} = (0, 8, 8, 20)$。确切的适配要求解无解方程式 $C = 0$, $C = 8$, $C = 8$, $C = 20$。找出这些方程式的 4×1 矩阵 A，求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。画出在高度 $\hat{\boldsymbol{x}} = C$ 的水平线以及 4 个误差 $\boldsymbol{e}$。

请尊重版权与译者的劳动成果，侵权必究!

6 把 $\boldsymbol{b}=(0, 8, 8, 20)$投影在通过 $\boldsymbol{a}=(1, 1, 1, 1)$的直线，求出$\hat{x}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ 与投影 $\boldsymbol{p}=\hat{x}\boldsymbol{a}$。检验 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 与 $\boldsymbol{a}$ 垂直，找出从 $\boldsymbol{b}$ 到通过 $\boldsymbol{a}$ 的直线的最短距离$\|\boldsymbol{e}\|$。

7 找出通过原点且与同样 4 个点(问题 6)最接近的直线 $b=Dt$，确切的适配要求解 $D\cdot 0=0$，$D\cdot 1=8$，$D\cdot 3=8$，$D\cdot 4=20$。找出 4×1 矩阵 A 并求解 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$。重画图 4.9a 所示的最佳直线 $b=Dt$ 以及 $\boldsymbol{e}$'s。

8 把 $\boldsymbol{b}=(0, 8, 8, 20)$投影在通过 $\boldsymbol{a}=(0, 1, 3, 4)$的直线，求出$\hat{x}=D$ 与投影 $\boldsymbol{p}=\hat{x}\boldsymbol{a}$。问题 5-6 的最佳 C 与问题 7-8 的最佳 D 与问题 1-4 的最佳(C, D)并不一致。这是因为(1, 1, 1, 1)与(0, 1, 3, 4)是_____垂直。

9 针对同样 4 个点的最接近抛物线 $C+Dt+Et^2$，写出不可解的方程式 $A\boldsymbol{x}=\boldsymbol{b}$，其中 3 个未知数是 $\boldsymbol{x}=(C, D, E)$，设定这 3 个正态方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ (不需要求解)。图 4.9a 中，你对 4 个点适配一条抛物线–在图 4.9b 中发生了什么？

10 针对同样 4 个点的最接近立方线(cubic) $C+Dt+Et^2+Ft^3$，写出 4 个方程式的 $A\boldsymbol{x}=\boldsymbol{b}$，利用消元法求解。在图 4.9a 中，这条立方线确实通过这 4 个点。求 p 与 e。

11 4 个时间点的平均是$\hat{t}=(0+1+3+4)/4=2$，4 个 b's 的平均是$\hat{b}=(0+8+8+20)/4=9$。

(a) 验证最佳直线通过中心点$(\hat{t},\ \hat{b})=(2, 9)$。

(b) 说明为什么 $C+D\hat{t}=\hat{b}$是来自 $A^{\mathrm{T}}A\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 的第一个方程式。

问题 2-16 介绍统计的基本概念–最小二乘的基础

12 (推荐) 这个问题把 $\boldsymbol{b}=(b_1, ..., b_m)$投影在通过 $\boldsymbol{a}=(1, ..., 1)$的直线，我们求解 n 个方程式 $\boldsymbol{a}\boldsymbol{x}=\boldsymbol{b}$，只有一个未知数(利用最小二乘)。

(a) 求解 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}\hat{\boldsymbol{x}}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$ 证明$\hat{\boldsymbol{x}}$ 是 $\boldsymbol{b}$'s 的平均值。

(b) 求出 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{a}\hat{\boldsymbol{x}}$ 与方差 $=\|\boldsymbol{e}\|^2$ 以及标准差$\|\boldsymbol{e}\|$。

(c) 水平线$\hat{b}=3$ 与 $\boldsymbol{b}=(1, 2, 6)$最接近，检验 $\boldsymbol{p}=(3, 3, 3)$与 $\boldsymbol{e}$ 垂直并求出 3×3 矩阵 P。

13 最小二乘背后的第一个假设：$A\boldsymbol{x}=\boldsymbol{b}-$(平均值为 0 的杂讯)。把误差向量 $\boldsymbol{e}=\boldsymbol{b}-A\boldsymbol{x}$ 左乘$(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$，得到在右侧的$\hat{\boldsymbol{x}}-\boldsymbol{x}$。这项估测误差$\hat{\boldsymbol{x}}-\boldsymbol{x}$ 的平均值也是零，估测值$\hat{\boldsymbol{x}}$是无偏的(unbiased)。

14 最小二乘背后的第二个假设：这 m 个误差 e_i 是无关且方差为σ^2，所以$(\boldsymbol{b}-A\boldsymbol{x})(\boldsymbol{b}-A\boldsymbol{x})^{\mathrm{T}}$的平均值是$\sigma^2 I$。把$(\boldsymbol{b}-A\boldsymbol{x})(\boldsymbol{b}-A\boldsymbol{x})^{\mathrm{T}}$左乘$(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$再右乘$A(A^{\mathrm{T}}A)^{-1}$，证明平均值矩阵$(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^{\mathrm{T}}$是$\sigma^2(A^{\mathrm{T}}A)^{-1}$，这就是段落 10.2 的协方差矩阵 W (covariance matrix)。

网址：linear.neocities.org 邮箱：876152660@qq.com

15 医生读取 4 次你的心率值，$x = b_1, ..., x = b_4$ 的最佳解是 $b_1, ..., b_4$ 的平均值$\hat{\boldsymbol{x}}$，矩阵 A 是一列的 1's。问题 14 得到期望误差$(x - \hat{x})^2$ 是$\sigma^2(A^{\mathrm{T}}A)^{-1}$ = _____。取平均值，方差从σ^2 掉到$\sigma^2/4$。

16 如果你知道了 9 个数值 $b_1, ..., b_9$ 的平均值$\hat{x}$，现在增多了一个数字 b_{10}，你如何快速的求得平均值$\hat{x}_{10}$？递归最小二乘的概念在于避免 10 个数字相加，计算 $\hat{x}_{10}$ 时，$\hat{x}_9$ 会乘什么数？

$\hat{x}_{10} = b_{10}/10 +$ _____$\hat{x}_9 = (b_1 + ... + b_{10})/10$　如同已解范例 4.2C。

问题 17-24 更多在的练习 $\hat{\boldsymbol{x}}$与 $\boldsymbol{p}$ 与 $\boldsymbol{e}$。

17 写下直线 $b = C + Dt$ 的三个方程式，其中 $t = -1$ 时 $b = 7$，$t = 1$ 时 $b = 7$，$t = 2$ 时 $b = 21$。找出最小二乘解 $\hat{\boldsymbol{x}} = (C, D)$并画出最接近的直线。

18 找出问题 17 的投影 $\boldsymbol{p} = A\hat{\boldsymbol{x}}$，会在最接近线得到 3 个高度。证明误差向量是 $\boldsymbol{e}$ = (2, –6, 4)。为什么 $P\boldsymbol{e} = \boldsymbol{0}$？

19 假设问题 18 的量测值，在 $t = -1, 1, 2$ 时，误差是 2, –6, 4。计算 $\hat{\boldsymbol{x}}$ 并画出最接近量测值的直线。说明答案：$\boldsymbol{b} = (2, -6, 4)$与_____垂直，所以投影是 $\boldsymbol{p} = \boldsymbol{0}$。

20 假设量测值在 $t = -1, 1, 2$ 时，$\boldsymbol{b} = (5, 13, 17)$。计算 $\hat{\boldsymbol{x}}$，接近的直线与 $\boldsymbol{e}$。因为 $\boldsymbol{b}$ 是_____，所以误差 $\boldsymbol{e} = \boldsymbol{0}$。

21 4 个子空间的哪一个包含误差向量 $\boldsymbol{e}$？哪个包含 $\boldsymbol{p}$？哪个包含 $\hat{\boldsymbol{x}}$？A 的零空间为何？

22 量测值在 $t = -2, -1, 0, 1, 2$ 时，$b = (4, 2, -1, 0, 0)$，求出最佳直线 $C + Dt$。因

23 误差向量 $\boldsymbol{e}$ 与 $\boldsymbol{b}$ 正交？与 $\boldsymbol{p}$ 正交？与 $\boldsymbol{e}$ 正交？与 $\hat{\boldsymbol{x}}$正交？证明 $\| \boldsymbol{e} \|^2$ 等于 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{b}$，也就是 $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{p}^{\mathrm{T}}\boldsymbol{b}$，这是最小的整体误差 E。

24 $\|A\boldsymbol{x}\|^2$ 相对于 $x_1, ..., x_n$ 的部分导数是向量 $2A^{\mathrm{T}}A\boldsymbol{x}$。$2\boldsymbol{b}^{\mathrm{T}}A\boldsymbol{x}$ 的导数是向量 $2A^{\mathrm{T}}\boldsymbol{b}$，所以当_____时，$\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ 的导数是零。

挑战问题

25 在什么条件下，3 个点$(t_1, b_1), (t_2, b_2), (t_3, b_3)$会在一条直线上？列空间的答案：$(b_1, b_2, b_3)$必须是(1, 1, 1)与$(t_1, t_2, t_3)$的组合。尝试求出关联 t's 与 b's 的方程式，我已经很快考虑到这个方程式。

请尊重版权与译者的劳动成果，侵权必究!

26 方形 4 个角的坐标(1, 0), (0, 1), (–1, 0), (0, –1)对应 $\boldsymbol{b} = (0, 1, 3, 4)$，找出一个最佳适配的平面。$C + Dx + Ey = b$ 在 4 个点的方程式是 $A\boldsymbol{x} = \boldsymbol{b}$，其中有 3 个未知数 $\boldsymbol{x} = (C, D, E)$。$A$ 为何？在方形的中心点(0, 0)，证明 $C + Dx + Ey = b$'s 的平均值。

27 (两条线的距离) 点 $P = (x, x, x)$与 $Q = (y, 3y, -1)$在空间中的两条不相交的直线上，选择 x 与 y 极小化距离平方$\|P - Q\|^2$，连接最接近 P 与 Q 的直线与_____垂直。

28 假设 A 的列是相关，你如何找出矩阵 B 使得 $P = B(B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}$ 确实得到在 A 的列空间的投影。(一般的公式在 $A^{\mathrm{T}}A$ 不可逆时会失败。)

29 在$\mathbf{R}^n$中，通常会存在唯一的超平面包含 n 个给定点 $\boldsymbol{x} = \mathbf{0}, \boldsymbol{a}_1, ..., \boldsymbol{a}_{n-1}$。(例如 $n = 3$：除非_____，存在一个平面包含 $\mathbf{0}, \boldsymbol{a}_1, \boldsymbol{a}_2$。如何测试在$\mathbf{R}^n$中恰好只有一个平面?

30 范例 2 平移时间点 t_i 使得总和是零，我们减去平均时间 $\hat{t} = (t_1 + ... + t_m)/m$ 得到 $T_i = t_i - \hat{t}$，这些 T_i 的总和是零。

现在列(1, ..., 1)与$(T_1, ..., T_m)$正交，$A^{\mathrm{T}}A$ 是对角，它的单元是 m 与 $T_1^2 + ... + T_m^2$。证明最佳的 C 与 D 有直接的公式：

$$T = t - \hat{t} \qquad C = \frac{b_1 + ... + b_m}{m} \qquad D = \frac{b_1T_1 + ... + b_mT_m)}{T_1^2 + ... + T_m^2}$$

最佳直线是 $C + Dt$ 或 $C + D(t - \hat{t})$。时间平移使得 $A^{\mathrm{T}}A$ 变成对角是格莱姆-施密特程序的一个范例：事先正交化 A 的列。

4.4 正交单位基底与格莱姆-施密特

1 若列 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases}$，则 $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n$ 是正交单位向量，$Q^{\mathrm{T}}Q=I$。

2 若 Q 也是方形，则 $QQ^{\mathrm{T}}=I$ 且 $Q^{\mathrm{T}}=Q^{-1}$，Q 是“**正交矩阵**”。

3 $Q\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解是 $\hat{\boldsymbol{x}}=Q^{\mathrm{T}}\boldsymbol{b}$，$\boldsymbol{b}$ 的投影：$\boldsymbol{p}=QQ^{\mathrm{T}}\boldsymbol{b}=P\boldsymbol{b}$。

4 **格莱姆-施密特**程序使得无关的 $\boldsymbol{a}_i$ 成为正交单位的 $\boldsymbol{q}_i$，从 $\boldsymbol{q}_1=\boldsymbol{a}_1/\|\boldsymbol{a}_1\|$开始。

5 $\boldsymbol{q}_i$ 是($\boldsymbol{a}_i$–投影 $\boldsymbol{p}_i$)/$\|\boldsymbol{a}_i-\boldsymbol{p}_i\|$；投影 $\boldsymbol{p}_i=(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_1)\boldsymbol{q}_1+\ldots+(\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{q}_{i-1})\boldsymbol{q}_{i-1}$。

6 每个 $\boldsymbol{a}_i$ 是 $\boldsymbol{q}_1$ 至 $\boldsymbol{q}_i$ 的组合，则 $A=QR$：正交 Q 与三角 R。

本段落有两个目的，**为什么**与**如何**，第一个目标是了解为什么正交性很棒。点积是零所以 $A^{\mathrm{T}}A$ 是对角，在求取$\hat{\boldsymbol{x}}$与 $\boldsymbol{p}=A\hat{\boldsymbol{x}}$时会变得如此简单。第二个目标是创建正交向量，你会看到格莱姆-施密特选取原始基底向量的组合获得直角。这些原始向量是 A 的列，可能不是正交，**正交单位基底向量是新矩阵 $\boldsymbol{Q}$ 的列**。

从第三章，一组基底包含生成空间的无关向量，基底向量可以交会在任何角度(除了 0°与 180°)，但是每次我们看到坐标轴，他们都是垂直的。在我们的想象中，坐标轴实际上永远正交，简化图形以及大大简化计算。

当他们的点积$\boldsymbol{q}_i\cdot\boldsymbol{q}_j$是零，向量 $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n$ 是**正交**，更准确的说只要 $i\neq j$ 则 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j=0$。再多一步，只要除以自己的长度--向量变成正交单位向量。他们的长度都是 1(正常)，此时基底称为**正交单位(orthonormal)**。

定义 向量 $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n$ 是**正交单位**，若

$$\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j=\begin{cases}0, & i\neq j \quad \text{正交向量}\\ 1, & i=j \quad \text{单位向量：}\|\boldsymbol{q}_i\|=1\end{cases}$$

一个有正交列的矩阵都指定特殊字母 $\boldsymbol{Q}$。

矩阵 Q 很容易处理，因为 $\boldsymbol{Q^{\mathrm{T}}Q=I}$。在矩阵语言中再次说明列 $\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n$ 是正交单位，**$\boldsymbol{Q}$ 不需要是方形**。

请尊重版权与译者的劳动成果，侵权必究!

具有正交单位列的矩阵 Q 满足 $Q^{\mathrm{T}}Q=I$：

$$Q^{\mathrm{T}}Q=\begin{bmatrix}— & \boldsymbol{q}_1^{\mathrm{T}} & — \\ — & \boldsymbol{q}_2^{\mathrm{T}} & — \\ & \vdots & \\ — & \boldsymbol{q}_n^{\mathrm{T}} & —\end{bmatrix}\begin{bmatrix}| & | & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_n \\ | & | & |\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1\end{bmatrix}=I \tag{1}$$

Q^{T} 的行 i 乘 Q 的列 j，点积是 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j$。非对角线($i\neq j$)因为正交使得点积为零，在对角线($i=j$)单位向量得到 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i=\|\boldsymbol{q}_i\|^2=1$。通常 Q 是矩形($m>n$)，有时候 $m=n$。

当 Q 是方形，$Q^{\mathrm{T}}Q=I$，表示 $Q^{\mathrm{T}}=Q^{-1}$：转置 = 逆矩阵

如果列只是正交(不是单位向量)，点积还是得到对角矩阵(不是单位矩阵)，这个对角矩阵与 I 近乎完全相同。重要的是正交性—它很容易创建单位向量。

重复说明：尽管 Q 是矩形，还是有 $Q^{\mathrm{T}}Q=I$，这种情形下 Q^{T} 只是左逆矩阵。对于方形矩阵我们也有 $QQ^{\mathrm{T}}=I$，所有 Q^{T} 是 Q 的双边逆矩阵。方形 Q 的行好像列一样也是正交，逆矩阵就是转置，我们称方形的案例 Q 是正交矩阵[1]。

下列有 3 个正交矩阵例子—旋转与排列与反射，最快的测试是检查 $Q^{\mathrm{T}}Q=I$。

范例 1 **(旋转)** Q 把平面的每个向量旋转 θ 角：

$$Q=\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} \quad \text{与} \quad Q^{\mathrm{T}}=Q^{-1}=\begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}$$

Q 的列是正交(取他们的点积)，因为 $\cos^2\theta+\sin^2\theta=1$，他们是单位向量，这些列是平面 $\mathbf{R}^2$ 的一组正交单位基底。

标准基底向量 $\boldsymbol{i}$ 与 $\boldsymbol{j}$ 旋转 θ 角(见图 4.10a)，Q^{-1} 把向量旋转 $-\theta$ 角回复原状。它与 Q^{T} 一致，因为 $-\theta$ 的余弦等于 θ 的余弦，$\sin(-\theta)=-\sin\theta$，我们有 $Q^{\mathrm{T}}Q=I$ 与 $QQ^{\mathrm{T}}=I$。

范例 2 **(排列)** 这些矩阵改变向量分量的顺序成为(y,z,x)与(y,x)：

$$\begin{bmatrix}0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0\end{bmatrix}\begin{bmatrix}x \\ y \\ z\end{bmatrix}=\begin{bmatrix}y \\ z \\ x\end{bmatrix} \quad \text{与} \quad \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix}\begin{bmatrix}x \\ y\end{bmatrix}=\begin{bmatrix}y \\ x\end{bmatrix}$$

这些 Q 的列都是单位向量(他们的长度明显是 1)，他们也是正交(1's 出现在不同位置)。排列矩阵的逆矩阵是它的转置：$Q^{-1}=Q^{\mathrm{T}}$，逆矩阵使得分量回复原始顺序：

[1] "正交单位矩阵"会是 Q 比较好的名字，但是没有使用它。任何具有正交单位列的矩阵就有字母 Q，但是当它是方形时，我们只称为**正交矩阵**。

$$\text{逆反 = 转置}\quad \begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}y\\z\\x\end{bmatrix}=\begin{bmatrix}x\\y\\z\end{bmatrix} \text{ 与 } \begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}y\\x\end{bmatrix}=\begin{bmatrix}x\\y\end{bmatrix}$$

每个排列矩阵是一个正交矩阵。

范例 3 **(反射)** 若 $\boldsymbol{u}$ 是任意的单位向量，令 $Q=I-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$。注意 $\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 是一个矩阵，而 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}$ 是一个数字 $\|\boldsymbol{u}\|^2=1$。则 Q^{T} 与 Q^{-1} 都等于 Q：

$$Q^{\mathrm{T}}=I-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}=Q \quad \text{且} \quad Q^{\mathrm{T}}Q=I-4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}+4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}=I \tag{2}$$

反射矩阵 $I-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 对称而且正交。如果把它平方你会得到单位矩阵：$Q^2=Q^{\mathrm{T}}Q=I$。通过镜子反射两次会回到原始状态，好像$(-1)^2=1$。注意方程式(2)的 $4\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 中间的 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1$。

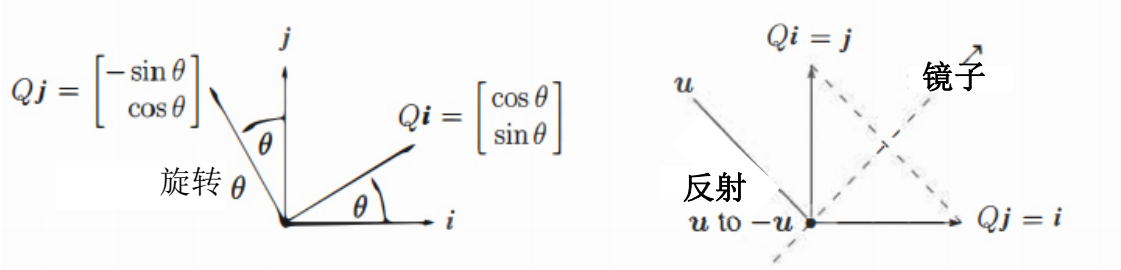


图 4.10：借由 $Q=\begin{bmatrix}\boldsymbol{c}&-\boldsymbol{s}\\\boldsymbol{s}&\boldsymbol{c}\end{bmatrix}$ 旋转以及借由 $Q=\begin{bmatrix}0&1\\1&0\end{bmatrix}$ 横跨 45°反射。

如果选择方向 $\boldsymbol{u}=(-1/\sqrt{2},\,1/\sqrt{2})$，计算 $2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$(列乘行)然后从 I 减去，可以得到在方向 $\boldsymbol{u}$ 的反射矩阵 Q：

$$\text{反射}\quad Q=I-2\begin{bmatrix}.5&-.5\\-.5&.5\end{bmatrix}=\begin{bmatrix}0&1\\1&0\end{bmatrix} \quad \text{与} \quad \begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}y\\x\end{bmatrix}$$

当(x,y)变成(y,x)，例如$(3,3)$的向量并没有移动，它在反射线上。

旋转会保有每个向量的长度，反射也是，排列也是，用任何正交矩阵去乘也是--**长度与角度没有改变**。

证明 $\|Q\boldsymbol{x}\|^2$ 等于 $\|\boldsymbol{x}\|^2$，因为$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}I\,\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$

若 Q 有正交单位列($Q^{\mathrm{T}}Q=I$)，它会保持长度不变：

$Q\boldsymbol{x}$ 有相同长度 **对于每个向量 $\boldsymbol{x}$ 都有** $\|Q\boldsymbol{x}\|=\|\boldsymbol{x}\|$ (3)

Q 也会保有点积：$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$，只应用了 $Q^{\mathrm{T}}Q=I$!

请尊重版权与译者的劳动成果，侵权必究!

投影使用正交单位基底：Q 代替 A

正交矩阵非常适合计算--当向量长度固定时，数字不会变的太大，稳定的电脑代码尽可能使用 Q's。

对于投影在子空间来说，所有的公式都跟 $A^{\mathrm{T}}A$ 有关，$A^{\mathrm{T}}A$ 的单元是基底向量 $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$ 的点积 $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j$。

假设基底向量是真实的正交单位，$\boldsymbol{a}$'s 变成 $\boldsymbol{q}$'s，则 $A^{\mathrm{T}}A$ 简化成 $Q^{\mathrm{T}}Q = I$。检视 $\hat{\boldsymbol{x}}$ 与 $\boldsymbol{p}$ 与 P 的改善，不用 $Q^{\mathrm{T}}Q$，我们打印一个空格代表单位矩阵：

$$_____\ \hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad 与 \quad \boldsymbol{p} = Q\hat{\boldsymbol{x}} \quad 与 \quad P = Q____Q^{\mathrm{T}} \tag{4}$$

$Q\boldsymbol{x} = \boldsymbol{b}$ 的最小二乘解是$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$，投影矩阵是 QQ^{T}。

没有矩阵需要逆转，这是正交单位基底的关键点。最佳$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ 只有 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ 与 $\boldsymbol{b}$ 的点积，我们有 1 维的投影！“耦合矩阵”或是“关联矩阵” $A^{\mathrm{T}}A$ 现在变成 $Q^{\mathrm{T}}Q = I$，不存在耦合。当 A 是 Q，具有正交单位的列，得到 $\boldsymbol{p} = Q\hat{\boldsymbol{x}} = QQ^{\mathrm{T}}\boldsymbol{b}$：

在 $\boldsymbol{q}$'s 的投影

$$\boldsymbol{p} = \begin{bmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + ... + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}) \tag{5}$$

重要案例： 当 Q 是方形且 $m = n$，子空间是整个空间，则有 $Q^{\mathrm{T}} = Q^{-1}$ 而且$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$ 与 $\boldsymbol{x} = Q^{-1}\boldsymbol{b}$ 相同，这个解是确切的！$\boldsymbol{b}$ 在整个空间的投影就是 $\boldsymbol{b}$ 自己，此时 $\boldsymbol{p} = \boldsymbol{b}$ 且 $P = Q^{\mathrm{T}}Q = I$。

你可以认为在整个空间的投影不值一提，但是当 $\boldsymbol{p} = \boldsymbol{b}$，我们的公式把 $\boldsymbol{b}$ 从它的一维投影中聚集起来。若 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ 是整个空间的一组正交单位基底，则 Q 是方形，每个 $\boldsymbol{b} = QQ^{\mathrm{T}}\boldsymbol{b}$ 是它沿着 $\boldsymbol{q}$'s 的分量的总和：

$$\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + ... + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}) \tag{6}$$

转换 $QQ^{\mathrm{T}} = I$ 是傅里叶级数的基础，也是所有应用数学中伟大“转换”的基础。他们把 $\boldsymbol{b}$ 或函数 $f(x)$打碎成为垂直的片段，借由(6)中的片段，逆转换把 $\boldsymbol{b}$ 与 $f(x)$ 再次一起恢复。

范例 4 正交 Q 的列是正交单位向量 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$：

$$m = n = 3 \qquad Q = \frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \text{ 有 } \boldsymbol{Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I}$$

$\boldsymbol{b} = (0, 0, 1)$分别在 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ 的投影是 $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3$：

$$\boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) = \frac{2}{3}\boldsymbol{q}_1 \quad 与 \quad \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b}) = \frac{2}{3}\boldsymbol{q}_2 \quad 与 \quad \boldsymbol{q}_3(\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{b}) = -\frac{1}{3}\boldsymbol{q}_3$$

前两项的总和是 $\boldsymbol{b}$ 在 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$ 的平面的投影，全部三项的总和是 $\boldsymbol{b}$ 在整个空间的投影——就是 $\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3 = \boldsymbol{b}$ 自己：

$$\text{重建 } \boldsymbol{b} = \boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_3 \qquad \frac{2}{3}\boldsymbol{q}_1 + \frac{2}{3}\boldsymbol{q}_2 - \frac{1}{3}\boldsymbol{q}_3 = \frac{1}{9}\begin{bmatrix} -2+4-2 \\ 4-2-2 \\ 4+4+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{b}$$

格莱姆-施密特程序

本段落的重点在于“正交就是好”，投影与最小二乘都与 $A^{\mathrm{T}}A$ 有关，当矩阵变成 $Q^{\mathrm{T}}Q = I$，逆矩阵就不是问题。一维的投影是非耦合的，最佳 $\hat{\boldsymbol{x}}$ 是 $Q^{\mathrm{T}}\boldsymbol{b}$(只是 n 个分别的点积。) 为了让此成真，我们必须说“**若向量是正交单位**”。现在我们说明**格莱姆-施密特方法来创建正交单位向量**。

从三个无关向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 开始，我们希望创建 3 个正交向量 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$，然后(在最后会最容易) 我们除以 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 的长度，就会得到 3 个正交单位向量 $\boldsymbol{q}_1 = \boldsymbol{A}/\|\boldsymbol{A}\|$，$\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$, $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$。

格莱姆-施密特 从选择 $\boldsymbol{A} = \boldsymbol{a}$ 开始，自然的接受第一个方向，下个方向 $\boldsymbol{B}$ 必须与 $\boldsymbol{A}$ 垂直。从 $\boldsymbol{b}$ 减去在 $\boldsymbol{A}$ 的投影开始，留下垂直的部分，这个就是正交向量 $\boldsymbol{B}$：

格莱姆-施密特的第一步

$$\boldsymbol{B} = \boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} \tag{7}$$

图 4.11 的 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 正交，使用 $\boldsymbol{A}^{\mathrm{T}}$ 左乘方程式(7)验证 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{0}$。向量 $\boldsymbol{B}$ 是我们以前所谓的误差向量 $\boldsymbol{e}$，与 $\boldsymbol{A}$ 垂直。注意方程式(7)中的 $\boldsymbol{B}$ 不是零(否则 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 相关)，方向 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 设定完成。

第三个方向从 $\boldsymbol{c}$ 开始，它不是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的组合(因为 $\boldsymbol{c}$ 不是 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的组合)，但是 $\boldsymbol{c}$ 不太可能与 $\boldsymbol{A}$ 及 $\boldsymbol{B}$ 垂直，所以减去 $\boldsymbol{c}$ 在两个方向的分量得到垂直方向 $\boldsymbol{C}$：

格莱姆-施密特的下一步

$$\boldsymbol{C} = \boldsymbol{c} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} - \frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B} \tag{8}$$

这是格莱姆-施密特程序唯一的概念，从每个新向量减去在已经存在的方向的投影，这个概念在每个步骤重复 [2]。如果我们有第四个向量，我们会减去三个在 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 的投影得到 $\boldsymbol{D}$。

[2] 我认为格莱姆有这个概念，我真的不知道施密特是哪来的。

请尊重版权与译者的劳动成果，侵权必究！

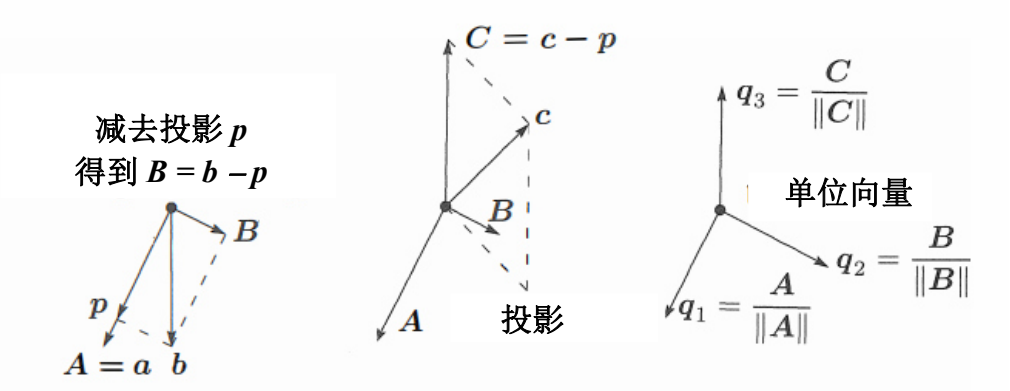


图 4.11：首先把 $\boldsymbol{b}$ 投影在通过 $\boldsymbol{a}$ 的直线，求出正交 $\boldsymbol{B}$ 是 $\boldsymbol{b}-\boldsymbol{p}$。把 $\boldsymbol{c}$ 投影在 $\boldsymbol{AB}$ 平面，求出 $\boldsymbol{C}=\boldsymbol{c}-\boldsymbol{p}$。除以 $\|\boldsymbol{A}\|, \|\boldsymbol{B}\|, \|\boldsymbol{C}\|$。

在最后或是每项都找到以后，$\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ 正交向量分别除以各自的长度，得到的向量 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3, \boldsymbol{q}_4$ 就是正交单位。

格莱姆-施密特的范例　假设无关但非正交向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$:

$$\boldsymbol{a}=\begin{bmatrix}1\\-1\\0\end{bmatrix} \text{ 与 } \boldsymbol{b}=\begin{bmatrix}2\\0\\-2\end{bmatrix} \text{ 与 } \boldsymbol{c}=\begin{bmatrix}3\\-3\\3\end{bmatrix}$$

则 $\boldsymbol{A}=\boldsymbol{a}$ 有 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=2$ 且 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}=2$，从 $\boldsymbol{b}$ 减去沿着 $\boldsymbol{A}$ 的投影 $\boldsymbol{p}$:

第一步
$$\boldsymbol{B}=\boldsymbol{b}-\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A}=\boldsymbol{b}-\frac{2}{2}\boldsymbol{A}=\begin{bmatrix}1\\1\\-2\end{bmatrix}$$

检验：$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}=0$ 满足要求。现在减去 $\boldsymbol{c}$ 在 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的投影得到 $\boldsymbol{C}$:

下一步
$$\boldsymbol{C}=\boldsymbol{c}-\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A}-\frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B}=\boldsymbol{c}-\frac{6}{2}\boldsymbol{A}-\frac{6}{6}\boldsymbol{B}=\begin{bmatrix}1\\1\\1\end{bmatrix}$$

检验：$\boldsymbol{C}$ = (1, 1, 1)与 $\boldsymbol{A}$ 及 $\boldsymbol{B}$ 都垂直。最终把 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 转成单位向量(长度 1，正交)。$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 的长度是 $\sqrt{2}$ 与 $\sqrt{6}$ 与 $\sqrt{3}$，除以这些长度得到正交单位基底：

$$\boldsymbol{q}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix} \text{ 与 } \boldsymbol{q}_2=\frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix} \text{ 与 } \boldsymbol{q}_3=\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

通常 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 都包含分数，大部分的 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ 都会有平方根。

分解 $A = QR$

我们从一个矩阵 A 开始，它的列是 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$。我们用矩阵 Q 结束，它的列是 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$。这些矩阵的关系为何？由于向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 是 $\boldsymbol{q}$'s 的组合(以此类推)，必然存在第三个矩阵关联 A 与 Q，第三个矩阵是 $A = QR$ 中的三角 R。

第一步是 $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ (其他向量不参与)，第二步是方程式(7)，其中 $\boldsymbol{b}$ 是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的组合，在这个步骤 $\boldsymbol{C}$ 与 $\boldsymbol{q}_3$ 没有参与。下一个向量不参与正是格莱姆-施密特的关键点：

- 向量 $\boldsymbol{a}$ 与 $\boldsymbol{A}$ 与 $\boldsymbol{q}_1$ 沿着单一直线。
- 向量 $\boldsymbol{a}, \boldsymbol{b}$ 与 $\boldsymbol{A}, \boldsymbol{B}$ 与 $\boldsymbol{q}_1, \boldsymbol{q}_2$ 在同一个平面。
- 向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 与 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 与 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ 在同一个子空间(维度 3)。

在每一步 $\boldsymbol{a}_1, ..., \boldsymbol{a}_k$ 是 $\boldsymbol{q}_1, ..., \boldsymbol{q}_k$ 的组合，后面的 $\boldsymbol{q}$'s 没有参与。关联矩阵 R 是三角形，我们有 $A = QR$：

$$\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix} \quad \text{或} \quad \boldsymbol{A = QR} \tag{9}$$

$A = QR$ 是格莱姆-施密特的概括，左乘 Q^{T} 确认上式的 $R = Q^{\mathrm{T}}A$：

(格莱姆-施密特) 从无关向量 $\boldsymbol{a}_1, ..., \boldsymbol{a}_n$，格莱姆-施密特创建正交单位向量 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$。这些列构成的矩阵满足 $\boldsymbol{A = QR}$，则 $R = Q^{\mathrm{T}}A$ 是上三角，因为后面的 $\boldsymbol{q}$'s 与较早的 $\boldsymbol{a}$'s 正交。

下列是上面例子的原始 $\boldsymbol{a}$'s 与最终的 $\boldsymbol{q}$'s，$R = Q^{\mathrm{T}}A$ 的 i, j 单元是 Q^{T} 的行 i 乘 A 的列 j，点积 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{a}_j$ 进到 R，则 $A = QR$：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix} = QR$$

近看 Q 与 R，$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 的长度是 R 的对角线 $\sqrt{2}, \sqrt{6}, \sqrt{3}$，$Q$ 的列是正交单位。因为有平方根，QR 看起来比 LU 困难，这两种分解是线性代数计算的绝对中心。

任何具有无关列的 $m \times n$ 矩阵 A 可以分解成为 $A = QR$，$m \times n$ 矩阵 Q 有正交单位的列，方形矩阵 R 是上三角形：$A^{\mathrm{T}}A = (QR)^{\mathrm{T}}QR = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}R$。最小二乘的方程式 $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ 简化成 $R^{\mathrm{T}}R\hat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$，最终得到 $R\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$：赞。

请尊重版权与译者的劳动成果，侵权必究！

最小二乘 $R^{\mathrm{T}}R\hat{\boldsymbol{x}}=R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$ 或 $R\hat{\boldsymbol{x}}=Q^{\mathrm{T}}\boldsymbol{b}$ 或 $\hat{\boldsymbol{x}}=R^{-1}Q^{\mathrm{T}}\boldsymbol{b}$ (10)

不是求解 $A\boldsymbol{x}=\boldsymbol{b}$，这是不可能的，我们利用反向代入法求解 $R\hat{\boldsymbol{x}}=Q^{\mathrm{T}}\boldsymbol{b}$——这个非常快速。格莱姆-施密特程序实际的成本是 mn^2 次乘法，这些是用来创建正交 Q 与三角 R 得到 $A=QR$。

下列是非正式的代码，它对于 $j=1$，$j=2$，最终到 $j=n$ 执行方程式(11)，重要的 4-5 行从 $\boldsymbol{v}=\boldsymbol{a}_j$ 减去在每个 $\boldsymbol{q}_j$ 的投影，$i<j$。代码的最后一行正态化 $\boldsymbol{v}$(除以 $r_{jj}=\|\boldsymbol{v}\|$)得到单位向量 $\boldsymbol{q}_j$：

$$r_{kj}=\sum_{i=1}^{m}q_{ik}v_{ij}\ \text{与}\ v_{ij}=v_{ij}-q_{ik}r_{kj}\ \text{与}\ r_{jj}=\left(\sum_{i=1}^{m}v_{ij}{}^{2}\right)^{1/2}\ \text{与}\ q_{ij}=\frac{v_{ij}}{r_{jj}} \tag{11}$$

从 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}=\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 开始，这个代码会创建 $\boldsymbol{q}_1$，然后 $\boldsymbol{B},\boldsymbol{q}_2$ 然后 $\boldsymbol{C},\boldsymbol{q}_3$：

$$\boldsymbol{q}_1=\boldsymbol{a}/\|\boldsymbol{a}\| \qquad \boldsymbol{B}=\boldsymbol{a}_2-\left(\boldsymbol{q}_1{}^{\mathrm{T}}\boldsymbol{a}_2\right)\boldsymbol{q}_1 \qquad \boldsymbol{q}_2=\boldsymbol{B}/\|\boldsymbol{B}\|$$
$$\boldsymbol{C}^{*}=\boldsymbol{a}_3-\left(\boldsymbol{q}_1{}^{\mathrm{T}}\boldsymbol{a}_3\right)\boldsymbol{q}_1 \qquad \boldsymbol{C}=\boldsymbol{C}^{*}-\left(\boldsymbol{q}_2{}^{\mathrm{T}}\boldsymbol{C}^{*}\right)\boldsymbol{q}_2 \qquad \boldsymbol{q}_3=\boldsymbol{C}/\|\boldsymbol{C}\|$$

方程式(11)一次减去一个投影，如同 $\boldsymbol{C}^{*}$ 与 $\boldsymbol{C}$，这项改变称为“修正格莱姆-施密特”。这个代码在数值上比方程式(8)来得稳定，其中(8)一次性减去所有的投影。

```
for j = 1 to n
    v = A(:, j);
    for i = 1 : j – 1
        R(i, j) = Q(i, j)' * v;
        v = v - R(i, j) * Q(i, j);
    end
    R(i, j) = norm(v);
    Q(:, j) = v/ R(j, j);
end
```

```
% 修正格莱姆-施密特
% 向量 v 从原始 A 的列 j 开始
% 列 q1 到 qj–1 已经在 Q 安顿
% 计算 Rjj = qi^T aj 就是 qi^T v
% 减去投影(qi^T v) qi
% 现在 v 与所有的 q1 到 qj–1 垂直
% 矩阵 Rjj 的对角单元是长度
% 向量 v 除以它的长度得到 qj
% 回圈 for j = 1 : n 得到所有的 qj
```

要恢复 A 的列 j，撤销最后一个步骤以及中间步骤的代码：

$$R(j,j)\boldsymbol{q}_j=(\boldsymbol{v}\text{减它的投影})=(A\text{的列 }j)-\sum_{i=1}^{j-1}R(j,j)\boldsymbol{q}_i \tag{12}$$

把总和移动到最左边，列 j 出现在乘积 $\boldsymbol{QR}=\boldsymbol{A}$ 中。

告白 好的软件，比如 LAPACK，应用在好的系统，比如 MATLAB，Julia 与 Python，都没有使用格莱姆-施密特代码，现在有更好的方法。“豪斯霍尔德反射”作用在 A 得到上三角 R，每次处理一个列，就如同消元法得到 LU 的 U 时相同的方法。

这些反射矩阵 $I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 在第十一章有关数值线性代数中会描述。如果 A 是三角，我们甚至可以简化至 2×2 的旋转，结果永远是 $A = QR$ 且 MATLAB 正交化 A 的指令是[Q，R] = qr(A)。我相信格莱姆-施密特仍然是很好的程序，纵使反射或是旋转会得到更完美的 Q。

主要观念的复习

1. 若正交单位向量 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ 是 Q 的列，则 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = 0$，$\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = 1$ 转换成矩阵乘法 $Q^{\mathrm{T}}Q = I$。
2. 若 Q 是方形(正交矩阵)则 $Q^{\mathrm{T}} = Q^{-1}$：转置 = 逆反。
3. $Q\boldsymbol{x}$ 的长度等于 $\boldsymbol{x}$ 的长度：$\| Q\boldsymbol{x} \| = \| \boldsymbol{x} \|$。
4. 在 Q 的列空间的投影是 $P = QQ^{\mathrm{T}}$，其中 Q 由 $\boldsymbol{q}$'s 生成。
5. 若 Q 是方形则 $P = QQ^{\mathrm{T}} = I$ 以及每个 $\boldsymbol{b} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + ... + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b})$
6. 格莱姆-施密特程序从 $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ 得到正交单位向量 $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$。矩阵形式是分解 $A = QR$ = (正交 Q) (三角 R)。

已解范例

4.4A 额外加入两个单元不是 1 就是–1 的列，使得 4×4“哈达玛矩阵”是正交。你如何把 H_4 变成正交矩阵 Q?

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad H_4 = \begin{bmatrix} 1 & 1 & x & x \\ 1 & -1 & x & x \\ 1 & 1 & x & x \\ 1 & -1 & x & x \end{bmatrix} \qquad Q_4 = \begin{bmatrix} & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}$$

方块矩阵 $H_8 = \begin{bmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{bmatrix}$ 是下一个矩阵单元都是 1 与–1 的哈达玛 d 矩阵，乘积 $H_8^{\mathrm{T}} H_8$ =? $\boldsymbol{b} = (6, 0, 0, 2)$在 H_4 的第一列的投影是 $\boldsymbol{p}_1 = (2, 2, 2, 2)$，求 $\boldsymbol{b}$ 在前两列生成的 2 维子空间的投影 $\boldsymbol{p}_{1,2}$ 为何?

请尊重版权与译者的劳动成果，侵权必究!

解 H_4 可以从 H_2 得到，如同 H_8 可以从 H_4 得到：

$$H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \text{ 有正交的列}$$

则 $Q = H/2$ 有正交单位的列，除以 2 得到 Q 的单位向量。5×5 的哈达玛矩阵是不可能的，这是因为列的点积有 5 个 1's 且/或–1's，总和不可能是零。H_8 有长度 $\sqrt{8}$ 的正交列。

$$H_8{}^{\mathrm{T}} H_8 = \begin{bmatrix} H^{\mathrm{T}} & H^{\mathrm{T}} \\ H^{\mathrm{T}} & -H^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} H^{\mathrm{T}} & H^{\mathrm{T}} \\ H^{\mathrm{T}} & -H^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 2H^{\mathrm{T}}H & 0 \\ 0 & 2H^{\mathrm{T}}H \end{bmatrix} = \begin{bmatrix} 8I & 0 \\ 0 & 8I \end{bmatrix}, \; Q_8 = \frac{8I}{\sqrt{8}}$$

4.4B 正交列的关键点是什么？答案：$A^{\mathrm{T}}A$ 是对角且容易逆反，我们可以投影到直线然后相加，轴是正交的。

加上 $\boldsymbol{p}$'s 在平面的投影 $\boldsymbol{p}_{1,2}$ 等于在正交直线的投影 $\boldsymbol{p}_1 + \boldsymbol{p}_2$。

问题集 4.4

问题 1-12 有关正交向量与正交矩阵。

1 下列的向量配对是正交单位？或只是正交？ 或只是无关？

(a) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 与 $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ (b) $\begin{bmatrix} .6 \\ .8 \end{bmatrix}$ 与 $\begin{bmatrix} .4 \\ -.3 \end{bmatrix}$ (c) $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ 与 $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$

有需要时改变第二个向量得到正交单位向量。

2 向量(2, 2, –1)与(–1, 2, 2)正交，除以自己的长度求出正交单位向量 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$，放进 Q 的列，求 $Q^{\mathrm{T}}Q$ 与 QQ^{T}。

3 (a) 若 A 有 3 个长度都是 4 的正交列，求 $A^{\mathrm{T}}A$。

(b) 若 A 有 3 个长度分别是 1, 2, 3 的正交列，求 $A^{\mathrm{T}}A$。

4 依据下列情形给定一个范例：

(a) 矩阵 Q 有正交单位的列，但是 $QQ^{\mathrm{T}} \neq I$。

(b) 两个向量正交，却不是线性无关。

(c) $\mathbf{R}^3$的一组正交单位基底，包含向量 $\boldsymbol{q}_1 = (1, 1, 1)/\sqrt{3}$。

5 找出平面 $x + y + 2z = 0$ 的两个正交向量，再使得他们之间。

6 若 Q_1 与 Q_2 是正交矩阵，证明乘积 Q_1Q_2 也是正交矩阵。(使用 $Q^{\mathrm{T}}Q = I$)。

网址：linear.neocities.org 邮箱：876152660@qq.com

7 若 Q 有正交单位的列，$Q\boldsymbol{x} = \boldsymbol{b}$ 的最小二乘解 $\hat{\boldsymbol{x}}$ 为何？

8 若 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$ 是$\mathbf{R}^5$的正交单位向量，最接近给定向量 $\boldsymbol{b}$ 是___$\boldsymbol{q}_1$ + ___$\boldsymbol{q}_2$ 的组合。

9 (a) 当 $\boldsymbol{q}_1 = (.8, .6, 0)$与 $\boldsymbol{q}_2 = (-.6, .8, 0)$，计算 $P = QQ^{\mathrm{T}}$，验证 $P^2 = P$。

(b) 使用 $Q^{\mathrm{T}}Q = I$，证明$(QQ^{\mathrm{T}})^2 = QQ^{\mathrm{T}}$，则 $P = QQ^{\mathrm{T}}$ 是在 Q 的列空间的投影矩阵。

10 正交单位向量自动是线性无关。

(a) 向量证明：当 $c_1\boldsymbol{q}_1 + c_2\boldsymbol{q}_2 + c_3\boldsymbol{q}_3 = \boldsymbol{0}$，什么样的点积得到 $c_1 = 0$？相似的 $c_2 = 0$ 与 $c_3 = 0$，于是 $\boldsymbol{q}$'s 是无关。

(b) 矩阵证明：证明 $Q\boldsymbol{x} = \boldsymbol{0}$ 得到 $\boldsymbol{x} = \boldsymbol{0}$。由于 Q 可以是矩形，你可以使用 Q^{T} 但不能使用 Q^{-1}。

11 (a) 格莱姆-施密特：找出由 $\boldsymbol{a} = (1, 3, 4, 5, 7)$与 $\boldsymbol{b} = (-6, 6, 8, 0, 8)$生成空间的正交单位向量 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$。

(b) 平面中的哪个向量最接近(1, 0, 0, 0, 0)？

12 若 $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 是$\mathbf{R}^3$的一组基底，任意向量 $\boldsymbol{b}$ 可以写成

$$\boldsymbol{b} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 \text{ 或 } \begin{bmatrix} & & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \boldsymbol{b}$$

(a) 假设 $\boldsymbol{a}$'s 正交单位，证明 $x_1 = \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}$。

(b) 假设 $\boldsymbol{a}$'s 正交，证明 $x_1 = \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b} / \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1$。

(c) 假设 $\boldsymbol{a}$'s 无关，x_1 是____乘 $\boldsymbol{b}$ 的第一分量。

问题 13-25 有关格莱姆-施密特程序与 $A = QR$。

13 $\boldsymbol{b} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$ 要减去多少乘 $\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 使得结果 $\boldsymbol{B}$ 与 $\boldsymbol{a}$ 垂直？画出图形展示 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{B}$。

14 计算问题 13 的格莱姆-施密特程序，令 $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ 且 $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$，然后分解成 QR：

$$\begin{bmatrix} 1 & 4 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix} \begin{bmatrix} \|\boldsymbol{a}\| & ? \\ 0 & \|\boldsymbol{B}\| \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

15 (a) 找出正交单位向量 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ 使得 $\boldsymbol{q}_1, \boldsymbol{q}_2$ 生成 A 的列空间：

$$A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$$

(b) 4 个子空间中哪个子空间包含 $\boldsymbol{q}_3$?

(c) 利用最小二乘法求解 $A\boldsymbol{x} = (1, 2, 7)$。

16 多少乘 $\boldsymbol{a} = (4, 5, 2, 2)$最接近 $\boldsymbol{b} = (1, 2, 0, 0)$? 在 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的平面找出正交单位向量 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$。

17 求出 $\boldsymbol{b}$ 在通过 $\boldsymbol{a}$ 的直线的投影：

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \qquad \boldsymbol{p} = ? \qquad \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = ?$$

计算正交单位向量 $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ 与 $\boldsymbol{q}_2 = \boldsymbol{e}/\|\boldsymbol{e}\|$。

18 (推荐) 利用格莱姆-施密特程序，从 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 找出正交向量 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$:

$$\boldsymbol{a} = (1, -1, 0, 0) \qquad \boldsymbol{b} = (0, 1, -1, 0) \qquad \boldsymbol{c} = (0, 0, 1, -1)$$

$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 与 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 是垂直于 $\boldsymbol{d} = (1, 1, 1, 1)$的向量的基底。

19 若 $A = QR$ 则 $A^{\mathrm{T}}A = R^{\mathrm{T}}R =$ ____三角乘____三角。格莱姆-施密特对 A 相当于消元法对 $A^{\mathrm{T}}A$，其中 $A^{\mathrm{T}}A$ 的主元必须是 R 的对角线单元的平方。针对 A 利用格莱姆-施密特求出 Q 与 R：

$$A = \begin{bmatrix} -1 & 1 \\ 2 & 1 \\ 2 & 4 \end{bmatrix} \quad \text{与} \quad A^{\mathrm{T}}A = \begin{bmatrix} 9 & 9 \\ 9 & 18 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 9 & \\ & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

20 是非题(每个案例给出一个例子)：

(a) 当 Q 是正交矩阵则 Q^{-1} 是正交矩阵。

(b) 若 $Q(3\times2)$有正交单位的列，则$\|Q\boldsymbol{x}\|$永远等于$\|\boldsymbol{x}\|$。

21 求出 A 的列空间的一组正交单位基底：

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{与} \quad \boldsymbol{b} = \begin{bmatrix} -4 \\ -3 \\ 3 \\ 0 \end{bmatrix}$$

然后计算 $\boldsymbol{b}$ 在列空间的投影。

22 利用格莱姆-施密特求出正交向量 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$：

$$\boldsymbol{a}=\begin{bmatrix}1\\1\\2\end{bmatrix} \quad \boldsymbol{b}=\begin{bmatrix}1\\-1\\0\end{bmatrix} \quad \boldsymbol{c}=\begin{bmatrix}1\\0\\4\end{bmatrix}$$

23 找出 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$(无关的列)的组合 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$，然后 A 写成 QR：

$$A=\begin{bmatrix}1&2&4\\0&0&5\\0&3&6\end{bmatrix}$$

24 (a) 在$\mathbf{R}^4$中，下列方程式的解生成子空间 $\boldsymbol{S}$，求出它的一组基底：

$$x_1+x_2+x_3+x_4=0$$

(b) 找出正交补充 $\boldsymbol{S}^{\perp}$的一组基底。

(c) 在 $\boldsymbol{S}$ 找出 $\boldsymbol{b}_1$，在 $\boldsymbol{S}^{\perp}$找出 $\boldsymbol{b}_2$ 使得 $\boldsymbol{b}_1+\boldsymbol{b}_2=\boldsymbol{b}=(1,1,1,1)$。

25 若 $ad-bc>0$，$A=QR$ 的单元是：

$$\begin{bmatrix}a&b\\c&d\end{bmatrix}=\frac{\begin{bmatrix}a&-c\\c&a\end{bmatrix}}{\sqrt{a^2+c^2}}\frac{\begin{bmatrix}a^2+c^2&ab+cd\\0&ad-bc\end{bmatrix}}{\sqrt{a^2+c^2}}$$

当 $a, b, c, d = 2, 1, 1, 1$ 以及 $1, 1, 1, 1$ 时写出 $A=QR$。当列是相关且格莱姆-施密特崩溃时，R 的哪个单元变成零？

问题 26-29 使用方程式(11)的 QR 代码，它执行格莱姆-施密特。

26 说明为什么 C(在方程式(11)之后，利用 C^*得到)等于方程式(8)的 C？

27 方程式(8)从 $\boldsymbol{c}$ 减去它沿着 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的分量，为什么不从 $\boldsymbol{c}$ 减去它沿着 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的分量？

28 方程式(11)中的 mn^2 次乘法是哪儿来的？

29 应用 MATLAB 的 qr 代码到 $\boldsymbol{a}=(2,2,-1)$, $\boldsymbol{b}=(0,-3,3)$, $\boldsymbol{c}=(1,0,0)$。$\boldsymbol{q}$ 为何？

问题 30-35 有关特殊的正交矩阵。

30 前 4 个小波位于小波(wavelet)矩阵 W 的列：

$$W=\frac{1}{2}\begin{bmatrix}1&1&\sqrt{2}&0\\1&1&-\sqrt{2}&0\\1&-1&0&\sqrt{2}\\1&-1&0&-\sqrt{2}\end{bmatrix}$$

这些列有什么特殊？求出逆反小波转换 W^{-1}。

请尊重版权与译者的劳动成果，侵权必究!

31 (a) 选择 c 使得 Q 是正交矩阵：

$$Q = c\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

把 $\boldsymbol{b} = (1, 1, 1, 1)$投影至第一列，然后把 $\boldsymbol{b}$ 投影到前两列的平面。

32 若 $\boldsymbol{u}$ 是单位向量，$Q = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 是一个反射矩阵(范例 3)。从 $\boldsymbol{u} = (0, 1)$求出 Q_1，从 $\boldsymbol{u} = (0, \sqrt{2}/2, \sqrt{2}/2)$求出 Q_2。当 Q_1 与 Q_2 乘向量$(1, 2)$与$(1, 1, 1)$，画出反射图形。

33 找出所有既是正交又是下三角的矩阵。

34 当 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$，$Q = I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 是一个反射矩阵，两次反射得到 $Q^2 = I$。

(a) 证明 $Q\boldsymbol{u} = -\boldsymbol{u}$，镜子与 $\boldsymbol{u}$ 垂直。

(b) 当 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0$ 求 $Q\boldsymbol{v}$。它会反射到自己。

挑战问题

35 (MATLAB) 因子[Q, R] = qr(A)，其中 A = eye(4) − diag([1 1 1, −1])。你在正交化 A 的列$(1, -1, 0, 0)$与$(0, 1, -1, 0)$与$(0, 0, 1, -1)$与$(0, 0, 0, 1)$。你能够重设 Q 的正交列的比例得到漂亮的整数分量吗？

36 若 A 是 $m \times n$ 且秩 n，qr(A)得到方形 Q 与 R 下方的零：

来自 MATLAB 的因子是$(m \times m)$ $(m \times n)$ $\qquad A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R \\ 0 \end{bmatrix}$。

Q_1 的 n 个列是哪个基础子空间的一组正交单位矩阵?

Q_2 的 $m - n$ 个列是哪个基础子空间的一组正交单位矩阵?

37 我们知道 $P = QQ^{\mathrm{T}}$ 是在 $Q(m \times n)$的列空间的投影，现在加入一个额外的列 $\boldsymbol{a}$ 得到 $A = [Q \quad \boldsymbol{a}]$。格莱姆-施密特会用什么向量 $\boldsymbol{q}$ 取代 $\boldsymbol{a}$？从 $\boldsymbol{a}$ 开始，减去_____，然后除以_____得到 $\boldsymbol{q}$。

第五章

行列式

1　$A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的**行列式**是 $\boldsymbol{ad-bc}$。奇异矩阵 $A=\begin{bmatrix} a & xa \\ c & xc \end{bmatrix}$ 有 $\det=0$。

2　**行交换反转符号**：$PA=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}=\begin{bmatrix} c & d \\ a & b \end{bmatrix}$ 有 $\det PA=bc-ad=-\det \boldsymbol{A}$。

3　$\begin{bmatrix} xa+yA & xb+yB \\ c & d \end{bmatrix}$ 的行列式是 $x(ad-bc)+y(Ad-Bc)$。**行列式在行 1 本身是线性。**

4　**消元法** $EA=\begin{bmatrix} a & b \\ 0 & d-\dfrac{c}{a}b \end{bmatrix}$，$\det \boldsymbol{EA}=a\left(d-\dfrac{c}{a}b\right)=$ **主元乘积** $=\det \boldsymbol{A}$。

5　若 A 是 $n\times n$ 则 1, 2, 3, 4 仍然成立：当 A 是奇异，$\det=0$；**交换行会反转符号**；$\det$ **在行 1 本身是线性**；$\det=$ **主元的乘积**。永远有：$\det \boldsymbol{BA}=(\det \boldsymbol{B})(\det \boldsymbol{A})$以及 $\det \boldsymbol{A}^{\mathrm{T}}=\det \boldsymbol{A}$。这是一个惊奇的数字。

5.1　行列式的性质

方形矩阵的行列式是单一的数字，这个数字包含一个关于矩阵的惊奇信息量，它能够立即分辨矩阵是否可逆。**当矩阵没有逆矩阵时，行列式为零**。当 A 可逆，A^{-1} 的行列式是 $1/(\det A)$。若 $\det A=2$ 则 $\det A^{-1}=1/2$。事实上行列式对 A^{-1} 的每个单元都能导出一个公式。

这是行列式的一个用途——求出逆矩阵与主元与解 $A^{-1}\boldsymbol{b}$ 的公式。对大型矩阵来说，我们很少使用这些公式，因为消元法很快速。一个 2×2 矩阵，它的单元是 a, b, c, d，它的行列式 $ad-bc$ 显示 A^{-1} 是如何随着 A 改变而改变。注意除以行列式！

$$A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}\ \text{有逆矩阵}\qquad A^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{1}$$

请尊重版权与译者的劳动成果，侵权必究！

矩阵乘逆矩阵得到 I。当行列式是 $ad-bc=0$，我们被要求除以零，这是不行的——此时 A 没有逆矩阵。(当 $a/c=b/d$，矩阵的两个行是平行，得到 $ad=bc$ 以及 $\det A=0$。) 相关的行永远得到 $\det A=0$。

行列式也与主元有关联， 2×2 矩阵的主元是 a 与 $d-(c/a)b$，主元的乘积是行列式：

$$\textbf{主元乘积}\qquad a\left(d-\frac{c}{a}b\right)=ad-bc\quad \textbf{就是}\quad \det A$$

在一次行交换之后，主元变成 c 与 $b-(a/c)d$，这些新的主元乘积是 $bc-ad$。针对 $\begin{bmatrix}a & b\\ c & d\end{bmatrix}$ 的行交换会**反转**行列式的符号。

往前看　$n\times n$ 矩阵的行列式有 3 种方法求得：

1　n 个主元的乘积(乘 1 或–1)	这是**主元公式**
2　$n!$ 项的总和(乘 1 或–1)	这是**“大公式”**
3　组合 n 个较小的行列式(乘 1 或–1)	这是**余子式公式**

你看到加号或减号——在 1 与–1 之间的决策——是行列式的重大部分，它来自以下针对 $n\times n$ 矩阵的规则：

当矩阵的两个行(两个列)交换，行列式改变符号。

单位矩阵的行列式是+1，交换两个行则 $\det P=-1$，再交换两个行则新的排列有 $\det P=+1$。半数的排列是偶数($\det P=1$)，半数的排列是奇数($\det P=-1$)。从 I 开始，半数的 P 是偶数次数的交换，半数需要奇数。在 2×2 的例子中，ad 有一个加号，bc 有减号——来自行交换：

$$\det\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=1 \text{ 与 } \det\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}=-1$$

另一个重要规则是线性——但是首先会有一个警告。线性不代表 $\det(A+B)=\det A+\det B$，这是绝对的错误。纵使 $A=I$ 且 $B=I$，这种线性还是错误。错误的规则会得到 $\det(I+I)=1+1=2$，正确的规则是 $\det 2I=2^n$。当矩阵乘 2，行列式要乘2^n(不是只有 2)。

我们不想用公式定义行列式，从它的性质开始会比较好——符号反转与线性。这些性质很简单(段落 5.1)，他们是为公式做准备(段落 5.2)，然后是应用，包含这三种：

(1) 行列式产生 A^{-1} 与 $A^{-1}\boldsymbol{b}$ (这个公式称为**克拉默规则**)。

(2) 当盒子的边是 A 的行，**体积**是 $|\det A|$ 。

(3) 存在 n 个特殊数字λ称为**特征值**(eigenvalue)，$A-\lambda I$ 的行列式是 0。这是非常重要的应用。特征值构成第六章。

行列式的性质

行列式有 3 个基本性质(规则 1, 2, 3)，使用这些规则我们可以计算任何方形矩阵的行列式。**这个数字写成两种形式，det A 或 $|A|$**。注意：中括号对矩阵，直线对行列式。当 A 是 2×2 矩阵，规则 1, 2, 3 得到我们想要的答案：

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ 的行列式是 } \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

从规则 1-3 可以延伸至规则 4-10，最后两个规则是 $\det AB = (\det A)(\det B)$与 $\det A^{\mathrm{T}} = \det A$。我们会以 2×2 矩阵来检查所有的规则，但是不要忘记：规则适用于任何 $n \times n$ 矩阵 A。

规则 1(最简单)，$\det I = 1$ 适配单位立方体的体积= 1。

1　$n \times n$ 单位矩阵的行列式是 1。

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \quad \text{与} \quad \begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1$$

2　交换两个行，行列式改变符号(符号反转)。

$$\text{检验：} \begin{vmatrix} c & d \\ a & b \end{vmatrix} = -\begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad (\text{两侧都等于 } bc - ad)$$

因为这个规则，我们可以求出任何排列矩阵 P 的行列式。只要对 I 持续交换行，直到你得到 P，则 $\det P = +1$ 对应**偶数次**的交换行，以及 $\det P = -1$ 对应**奇数次**。

第三个规则必须对所有矩阵的行列式做了大跳跃。

3　行列式对于个别的行是线性函数(其他行保持固定)。

如果用 t 乘第一行，t 也会乘行列式。如果第一行是相加，行列式也是相加。只有当其他行保持不变时，这个规则才能适用。注意 c 与 d 如何保持不变：

任意数 t 乘行 1，t 乘行列式

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

A 的行 1 与 A'的行 1 相加，行列式相加：

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

第一种情形，两侧都是 $tad - tbc$，取出因子 t。第二种情形，两侧都是 $ad + a'd - bc - b'c$。这个规则同样适用 A 是 $n \times n$，以及一次的行交换。

$$A = \begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = 4\begin{vmatrix} 1 & 2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} \quad \text{与} \quad \begin{vmatrix} 4 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 4 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} + \begin{vmatrix} 0 & 8 & 8 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix}$$

依据规则自身，规则 3 没有说这些行列式是多少(det $A = 4$)。

请尊重版权与译者的劳动成果，侵权必究!

组合乘法与加法，我们在一个行得到任意的线性组合。针对交换行的规则 2 可以先把那一行放在第一行，之后再换回来。

这个规则的意思不是 $\det 2I = 2\det I$，要得到 $2I$，我们必须用 2 乘两个行，因子 2 会提出来两次：

$$\begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 2^2 = 4 \quad 与 \quad \begin{vmatrix} t & 0 \\ 0 & t \end{vmatrix} = t^2$$

这就好像面积与体积一样，把矩形变成 2 倍，面积变成 4 倍。扩展一个 n 维的盒子 t 倍，体积会扩展 t^n 倍。这个关联不是偶然--我们会看到**行列式如何等于体积**。

给予规则 1-3 特殊的关注，他们完全决定了数字 $\det A$。我们可以停在这里然后寻找 $n \times n$ 行列式的公式(有点复杂)，我们喜欢缓步前进，因为规则 4-10 会使得行列式更容易处理。

4　若 A 有两个行相等，则 $\det A = 0$。

相等的行　　检查 2×2：$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0$

规则 4 来自规则 2，(记得我们必须使用这个规则，而不是用 2×2 的公式。)交换两个相等的行，行列式 D 应该要改变符号，但是 D 又保持相同，因为矩阵没有改变，唯一会有 $-D = D$ 的数字只有 $D = 0$--这必须是行列式。(注意：这个理由在**布尔**代数中不成立，因为 $-1 = 1$，则 D 是以规则 1, 3, 4 做定义。)

有两个相等的行的矩阵没有逆矩阵，规则 4 使得 $\det A = 0$，但是矩阵可以在没有相等行的情况下，可以是奇异，行列式也可以是零！规则 5 会是关键，我们可以做行运算(好像消元法)而不会改变 $\det A$。

5　从另一行减去乘数乘某一行，$\det A$ 不变。

从行 2 减去 l 乘行 1　　$$\begin{vmatrix} a & b \\ c - la & d - lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

规则 3(线性)把左侧分割成右侧加上另一项 $-l\begin{vmatrix} a & b \\ a & b \end{vmatrix}$，依据规则 4 这个额外项是零：相等的行。因此规则 5 正确(不只是 2×2)。

结论　正常从 A 到 U 的消元步骤不会改变行列式，所以 $\det A = \det U$。如果我们可以从三角矩阵 U 得到行列式，就可以得到所有矩阵 A 的行列式。每个列交换会反转符号，所以一定有 $\det A = \pm \det U$。规则 5 把问题缩小到三角矩阵。

6　有一整行都是零的矩阵，则 $\det A = 0$。

零的行　　$\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = 0$　与　$\begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = 0$

简单证明，把另一行加到零行，行列式不变(规则 5)，但是矩阵现在有两个相等的行，所以依据规则 4，$\det A = 0$。

7　若 A 是三角则 $\det A = a_{11}\, a_{22} \ldots a_{nn}$ = 对角单元的乘积。

三角　$\begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad$　与　$\begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad$

假设所有的对角单元都不是零，利用消元法移除非对角单元！(若 A 是下三角，从较低的行减去每一行的乘数倍；若 A 是上三角，从较高的行)。依据规则 5 行列式不变--现在矩阵是对角矩阵：

对角矩阵　$$\det \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{bmatrix} = (a_{11})(a_{22}) \ldots (a_{nn})$$

依据规则 3，因子 a_{11} 来自第一行，因子 a_{22} 来自第二行，最终因子 a_{nn} 来自最后一行，行列式就是 a_{11} 乘 a_{22} 乘 ... 乘 a_{nn} 乘 $\det I$。最后使用规则 1，$\det I = 1$。

如果对角单元 a_{ii} 是零该如何？则三角 A 是奇异，消元法得到零行；依据规则 5，行列式不变；依据规则 6，零行表示 $\det A = 0$，我们得到测试奇异或可逆矩阵的伟大方法。

8　若 A 奇异则 $\det A = 0$，若 A 可逆则 $\det A \neq 0$。

奇异　$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 当且仅当 $ad - bc = 0$

证明　从 A 到 U 执行消元法，若 A 奇异则 U 有一个零行，规则得到 $\det A = \det U = 0$。若 A 可逆则 U 有主元在对角线，非零主元的乘积(依据规则 7)得到非零的行列式。

乘主元　$\det A = \pm \det U = \pm$(主元的乘积)　(2)

2×2 矩阵的主元是 a (若 $a \neq 0$)与 $d - (c/a)b$：

行列式　$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & d - (c/a)b \end{vmatrix} = ad - bc$$

这是行列式的第一个公式，MATLAB 利用主元乘积求得 $\det A$。$\pm \det U$ 中的符号与行交换的次数是偶数或是奇数有关：+1 或–1 是交换行的排列矩阵 P 的行列式。

没有交换行，$P = I$ 且 $\det A = \det U =$ 主元的乘积且 $\det L = 1$：

若 $PA = LU$ 则 $\det P \det A = \det L \det U$ 且 $\det A = \pm \det U$　(3)

请尊重版权与译者的劳动成果，侵权必究！

9　*AB* 的行列式是 det *A* 乘 det *B*：$|AB| = |A||B|$。

乘积规则　$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & q \\ r & s \end{vmatrix} = \begin{vmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{vmatrix}$

当矩阵 B 是 A^{-1}，这个规则得到 A^{-1} 的行列式是 1/det A:

***A* 乘 A^{-1}**　$AA^{-1} = I$ 所以　$(\det A)\,(\det A^{-1}) = \det I = 1$

这个乘积规则是目前为止最错综复杂的，甚至 2×2 的案例也要一些代数：

$$|A||B| = (ad-bc)(ps-qr) = (ap+br)(cq+ds) - (aq+bs)(cp+dr) = |AB|$$

对于 $n \times n$ 的案例来说，此处有一个简洁的方法可以证明$|AB| = |A||B|$。当$|B| \neq 0$，考虑比值 $D(A) = |AB| / |B|$。检查比值 $D(A)$具有性质 1, 2, 3, 则 $D(A)$必须是行列式且我们有$|AB| / |B| = |A|$。赞。

性质 1　(I 的行列式)　若 $A = I$ 则比值 $D(A)$变成$|B| / |B| = 1$。

性质 2　(符号反转)　当 A 的两个行交换，AB 同样的两个行也会交换，因此$|AB|$会改变符号，比值$|AB| / |B|$也会。

性质 3　(线性)　当 A 的行 1 左乘 t，AB 的行 1 也是，行列式变成$|AB|$左乘 t，所以比值$|AB| / |B|$也会左乘 t。

把 A 的行 1 加到 A'的行 1，则 AB 的行 1 加到 $A'B$ 的行 1，依据规则 3，行列式也会相加，除以$|B|$之后，比值也相加—如同预期。

结论　比值$|AB|/|B|$有 3 个与定义 A 的相同性质，因此它等于$|A|$，这就证明了乘积规则$|AB| = |A||B|$。$|B| = 0$ 的案例分开处理但是很容易，因为当 B 奇异则 AB 奇异，则$|AB| = |A||B|$ 是 $0 = 0$。

10　转置 A^{T} 与 *A* 与相同的行列式。

转置　$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$　两侧都等于 $ad - bc$。

当 A 奇异时，方程式$|A^{\mathrm{T}}| = |A|$变成 $0 = 0$ (我们知道 A^{T} 也是奇异)，否则 A 会有正常的分解 $PA = LU$，两侧同时转置得到 $A^{\mathrm{T}}P^{\mathrm{T}} = U^{\mathrm{T}}L^{\mathrm{T}}$。依据有关乘积的规则 9 来证明 $|A| = |A^{\mathrm{T}}|$：

比较　$\det P \det A = \det L \det U$　且　$\det A^{\mathrm{T}} \det P^{\mathrm{T}} = \det U^{\mathrm{T}} \det L^{\mathrm{T}}$

第一，$\det L = \det L^{\mathrm{T}} = 1$(两者都有 1 在对角线)。第二，$\det U = \det U^{\mathrm{T}}$(这些三角矩阵有相同的对角线)。第三，$\det P = \det P^{\mathrm{T}}$(排列矩阵有 $P^{\mathrm{T}}P = I$，依据规则 9 得到 $|P^{\mathrm{T}}|\,|P| = 1$；于是$|P^{\mathrm{T}}|$与$|P|$同时等于 1 或同时等于–1)。所以 L, U, P 与 $L^{\mathrm{T}}, U^{\mathrm{T}}, P^{\mathrm{T}}$ 有相同的行列式，我们得到 $\det A = \det A^{\mathrm{T}}$　。

列的重要注解 对于行的每条规则都适用于列(利用转置，因为$|A| = |A^{\mathrm{T}}|$)，当两个列交换，行列式改变符号；一个零列或是两个相等的列使得行列式为零；如果一个列乘 t，行列式也会；行列式是个别列的线性函数。

是时候停止了，性质表已经够长了，接下来我们会求出并使用行列式的公式。

主要观念的复习

1. 行列式的定义：$\det I = 1$，符号反转，以及个别行的线性。
2. 消元法之后 $\det A = \pm$(主元的乘积)
3. A 不可逆时，行列式为零。
4. 两个重要性质：$\det AB = (\det A)\,(\det B)$与 $\det A^{\mathrm{T}} = \det A$。

已解范例

5.1A 对 A 应用下列的运算，求出 M_1, M_2, M_3, M_4 的行列式：

在 M_1，每个 a_{ij} 乘$(-1)^{i+j}$ 得到**检核版**符号模式。

在 M_2，从 A 的行 2, 3, 1 减去 A 的行 1, 2, 3。

在 M_3，把 A 的行 1, 2, 3 加到 A 的行 2, 3, 1。

M_1, M_2, M_3 的行列式与 A 的行列式有什么关联？ 【本题没有 M_4，原文错误】

$$\begin{bmatrix} a_{11} & -a_{12} & a_{13} \\ -a_{21} & a_{22} & -a_{23} \\ a_{31} & -a_{32} & a_{33} \end{bmatrix} \quad \begin{bmatrix} \text{行 1} - \text{行 3} \\ \text{行 2} - \text{行 1} \\ \text{行 3} - \text{行 2} \end{bmatrix} \quad \begin{bmatrix} \text{行 1} + \text{行 3} \\ \text{行 2} + \text{行 1} \\ \text{行 3} + \text{行 2} \end{bmatrix}$$

解 三个行列式是 $\det A$，0，$2\det A$，理由是：

$$M_1 = \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix} \quad \text{所以 } \det M_1 = (-1)\,(\det A)\,(-1)$$

M_2 是奇异，因为行的总和是零，所以行列式是零。

依据规则 3(个别行的线性)，M_3 可以分割成 8 个矩阵：【圆圈处原文 3，错误】

$$\begin{vmatrix} \text{行 1} + \text{行 3} \\ \text{行 2} + \text{行 1} \\ \text{行 3} + \text{行 ②} \end{vmatrix} = \begin{vmatrix} \text{行 1} \\ \text{行 2} \\ \text{行 3} \end{vmatrix} + \begin{vmatrix} \text{行 3} \\ \text{行 2} \\ \text{行 3} \end{vmatrix} + \begin{vmatrix} \text{行 1} \\ \text{行 1} \\ \text{行 3} \end{vmatrix} + \ldots + \begin{vmatrix} \text{行 3} \\ \text{行 1} \\ \text{行 2} \end{vmatrix}$$

除了第一与最后之外，其余的都有重复的行与零行列式。第一个是 A 以及最后一个有两个行交换，所以 $\det M_3 = \det A + \det A = 2\det A$。(尝试 $A = I$。)

请尊重版权与译者的劳动成果，侵权必究!

5.1B 说明如何利用行运算得到下列的行列式：

$$\det\begin{bmatrix}1-a & 1 & 1\\ 1 & 1-a & 1\\ 1 & 1 & 1-a\end{bmatrix} = a^2(3-a) \tag{4}$$

解 从行 1 减去行 3，然后从行 2 减去行 1，得到

$$\det\begin{bmatrix}-a & 0 & a\\ 0 & -a & a\\ 1 & 1 & 1-a\end{bmatrix}$$

现在把列 1 加到列 3，列 2 也加到列 3，得到一个下三角矩阵，其中对角线是 $-a, -a, 3-a$：$\det = (-a)\,(-a)\,(3-a)$。

若 $a=0$ 或 $a=3$ 则行列式是零。如果 $a=0$ 我们得到全 1 矩阵，必然奇异；如果 $a=3$，每个行的总和都是零——再度奇异。这些数字 0 与 3 是全 1 矩阵的“特征值(eigenvalue)”，这个例子有揭示性并且重要，直指第六章。

问题集 5.1

问题 1-12 有关行列式的规则。

1 若 4×4 矩阵有 $\det A = 1/2$，求 $\det(2A)$与 $\det(-A)$与 $\det(A^2)$与 $\det(A^{-1})$。

2 若 3×3 矩阵有 $\det A = -1$，求 $\det(\frac{1}{2}A)$与 $\det(-A)$与 $\det(A^2)$与 $\det(A^{-1})$。

3 是非题，如果对，给出理由；如果错，举出反例：

(a) $I + A$ 的行列式是 $1 + \det A$。

(b) ABC 的行列式是$|A|\,|B|\,|C|$。

(c) $4A$ 的行列式是 $4|A|$。

(d) $AB - BA$ 的行列式是零。尝试 $A = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$。

4 哪些行交换会使得这些“反单位矩阵” J_3 与 J_4 有$|J_3| = -1$ 与$|J_4| = +1$？

$$\det\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix} = -1 \quad \text{但是} \quad \det\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\end{bmatrix} = +1$$

5 当 $n = 5, 6, 7$，计算行交换的次数使得反单位矩阵 J_n 变成单位矩阵 I_n。提出对于任意大小 n 的规则，并且预测 J_{101} 的行列式是+1 或是−1。

6　说明规则 6 (若一行都是零，行列式是零)如何来自规则 3。

7　求出旋转与镜射的行列式：

$$Q=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} \text{ 与 } Q=\begin{bmatrix}1-2\cos^2\theta & -2\cos\theta\sin\theta\\ -2\cos\theta\sin\theta & 1-2\sin^2\theta\end{bmatrix}$$

8　证明每个正交矩阵($Q^{\mathrm{T}}Q=I$)有行列式+1 或−1。

(a) 使用乘积规则$|AB|=|A|\,|B|$与转置规则$|Q|=|Q^{\mathrm{T}}|$。

(b) 只使用乘积规则。若$|\det Q|>1$ 则 $\det Q^n=(\det Q)^n$ 会爆炸。你如何知道这个不可能发生在 Q^n？

9　下列矩阵的行列式是 0, 1, 2 或 3？

$$A=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\quad B=\begin{bmatrix}0&1&1\\1&0&1\\1&1&0\end{bmatrix}\quad C=\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$$

10　若 A 的每一行的总和是零，求解 $A\boldsymbol{x}=\boldsymbol{0}$ 证明 $\det A=0$。若这些单元的总和是 1，证明 $\det(A-I)=0$。这个代表 $\det A=1$ 吗？

11　假设 $CD=-DC$，找出下列推论的错误：取行列式得到$|C|\,|D|=-|D|\,|C|$，因此$|C|=0$ 或是$|D|=0$，其中一个或是两者必须是奇异。(这是错误的。)

12　2×2 矩阵的逆矩阵似乎有行列式 $=1$：

$$\det A^{-1}=\det\frac{1}{ad-bc}\begin{bmatrix}d&-b\\-c&a\end{bmatrix}=\frac{ad-bc}{ad-bc}=1$$

这个计算哪儿错了？正确的 $\det A^{-1}$ 为何？

问题 13-27 使用规则计算特定行列式。

13　简化 A 到 U 然后求出 $\det A=$ 主元乘积：

$$A=\begin{bmatrix}1&1&1\\1&2&2\\1&2&3\end{bmatrix}\qquad A=\begin{bmatrix}1&2&3\\2&2&3\\3&3&3\end{bmatrix}$$

14　应用行运算得到上三角 U，计算

$$\det\begin{bmatrix}1&2&3&0\\2&6&6&1\\-1&0&0&3\\0&2&0&7\end{bmatrix}\quad\text{与}\quad\det\begin{bmatrix}2&-1&0&0\\-1&2&-1&0\\0&-1&2&-1\\0&0&-1&2\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

15 应用行运算简化并计算下列矩阵的行列式：

$$\det\begin{bmatrix}101 & 201 & 301\\ 102 & 202 & 302\\ 103 & 203 & 303\end{bmatrix} \quad 与 \quad \det\begin{bmatrix}1 & t & t^2\\ t & 1 & t\\ t^2 & t & 1\end{bmatrix}$$

16 求秩一矩阵与反对称矩阵的行列式：

$$A=\begin{bmatrix}1\\2\\3\end{bmatrix}\begin{bmatrix}1 & -4 & 5\end{bmatrix} \quad 与 \quad A=\begin{bmatrix}0 & 1 & 3\\ -1 & 0 & 4\\ -3 & -4 & 0\end{bmatrix}$$

17 反对称矩阵有 $A^{\mathrm{T}}=-A$，把问题 16 中的 1, 3, 4 换成 a, b, c 然后证明 $|A|=0$。写下一个 4×4 的例子 $|A|=1$。

18 应用行运算证明 3×3 的“**范德蒙德行列式**”是：

$$\det\begin{bmatrix}1 & a & a^2\\ 1 & b & b^2\\ 1 & c & c^2\end{bmatrix}=(b-a)(c-a)(c-b)$$

19 求出 U 与 U^{-1} 与 U^2 的行列式：

$$U=\begin{bmatrix}1 & 4 & 6\\ 0 & 2 & 5\\ 0 & 0 & 3\end{bmatrix} \quad 与 \quad U=\begin{bmatrix}a & b\\ 0 & d\end{bmatrix}$$

20 假设你同时做了两次行运算，

$$从 \begin{bmatrix}a & b\\ c & d\end{bmatrix} 到 \begin{bmatrix}a-Lc & b-Ld\\ c-la & d-lb\end{bmatrix}$$

求第二个的行列式，它会等于 $ad-bc$ 吗？

21 行交换：A 的行 1 加到行 2，然后从行 1 减去行 2，然后行 1 加到行 2，然后 –1 乘行 1，最终得到 B。哪些规则说明

$$\det B=\begin{vmatrix}c & d\\ a & b\end{vmatrix} 等于 -\det A=-\begin{vmatrix}a & b\\ c & d\end{vmatrix} ?$$

这些规则可以取代规则 2 作为行列式的定义。

22 从 $ad-bc$，求出 A 与 A^{-1} 与 $A-\lambda I$ 的行列式：

$$A=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \quad 与 \quad A^{-1}=\frac{1}{3}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix} \quad 与 \quad A-\lambda I=\begin{bmatrix}2-\lambda & 1\\ 1 & 2-\lambda\end{bmatrix}$$

哪两个数字 λ 使得 $\det(A-\lambda I)=0$？对于这些数字 λ，写下矩阵 $A-\lambda I$ ——它应该是不可逆。

23 从 $A = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix}$ 求出 A^2 与 A^{-1} 与 $A - \lambda I$ 以及他们的行列式。哪两个数字λ使得 $\det(A - \lambda I) = 0$？

24 使用消元法简化 A 到 U 然后 $A = LU$：

$$A = \begin{bmatrix} 3 & 3 & 4 \\ 6 & 8 & 7 \\ -3 & 5 & -9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & -1 \end{bmatrix} = LU$$

求出下列的行列式：L, U, A, U^{-1}, L^{-1} 以及 $U^{-1}L^{-1}A$。

25 若 A 的 i,j 单元是 i 乘 j，证明 $\det A = 0$ (除了 $A = [1]$)。

26 若 A 的 i,j 单元是 $i + j$，证明 $\det A = 0$ (除了 $n = 1$ 或 2)。

27 应用行运算计算矩阵的行列式：

$$A = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & b \\ c & 0 & 0 \end{bmatrix} \quad 与 \quad B = \begin{bmatrix} 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \\ d & 0 & 0 & 0 \end{bmatrix} \quad 与 \quad C = \begin{bmatrix} a & a & a \\ a & b & b \\ a & b & c \end{bmatrix}$$

28 是非题 (正确则给定理由，错误则举出 2×2 的例子)：

(a) 若 A 不可逆则 AB 不可逆。

(b) A 的行列式永远是主元的乘积。

(c) $A - B$ 的行列式等于 $\det A - \det B$。

(d) AB 与 BA 有相同的行列式。

29 下列是投影矩阵的行列式 $\det P = 1$ 的证明，哪儿出错了？

$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} \quad 所以 \quad |P| = |A| \frac{1}{|A^{\mathrm{T}}||A|} |A^{\mathrm{T}}| = 1$$

30 (微积分问题) 证明 $\ln(\det A)$的部分导数得到 A^{-1}！

$$f(a, b, c, d) = \ln(ad - bc) \quad 得到 \quad \begin{bmatrix} \frac{\partial f}{\partial a} & \frac{\partial f}{\partial c} \\ \frac{\partial f}{\partial b} & \frac{\partial f}{\partial d} \end{bmatrix} = A^{-1}$$

31 (MATLAB) 希尔伯特矩阵 hilb(n)的 i, j 单元等于 $1/(i + j - 1)$，求出 hilb(1), hilb(2)，..., hilb(10)的行列式。希尔伯特矩阵不好处理！hilb(5)的主元为何？

32 (MATLAB) 对于 $n = 50, 100, 200, 400$ 的 rand(n)与 randn(n)来说，典型(实验性)的行列式为何？(MATLAB 的“Inf”是什么意思？)

33 (MATLAB) 6×6 矩阵的全部单元不是 1's 就是–1's，找出最大的行列式。

34 如果你知道 $\det A = 6$，B 的行列式为何？

$$从 \det A = \begin{vmatrix} 行\,1 \\ 行\,2 \\ 行\,3 \end{vmatrix} = 6 \quad 求 \quad \det B = \begin{vmatrix} 行\,3 + 行\,2 + 行\,1 \\ 行\,2 + 行\,1 \\ 行\,1 \end{vmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

5.2 排列与余子式

1 2×2：$ad-bc$ 有 2!个带±号的项。$n\times n$：$\det A$ 加起来有 **$n!$个带±号的项**。

2 $n=3$，$\det A$ 有 $3!=6$ 个项。两个项是$+a_{12}\,a_{23}\,a_{31}$ 与$-a_{13}\,a_{22}\,a_{31}$。**行 1, 2, 3 与列 1, 2, 3 在每个项出现一次**。

3 因为顺序 3, 2, 1 需要一次交换才能恢复顺序 1, 2, 3，所以产生负号。

4 这六个项包含 $+a_{11}\,a_{22}\,a_{33}-a_{11}\,a_{23}\,a_{32}=a_{11}(\boldsymbol{a_{22}\,a_{33}-a_{23}\,a_{32}})=a_{11}$(**余子式 $\boldsymbol{C_{11}}$**)。

5 永远有 $\det A=a_{11}\,C_{11}+a_{12}\,C_{12}+\ldots+a_{1n}\,C_{1n}$，余子式是大小 $n-1$ 的行列式。

计算机利用主元计算行列式，本段落介绍另外两种方法。存在一个“大公式”用到全部 $n!$ 的排列，存在一个“**余子式(cofactor)公式**”用到大小 $n-1$ 的行列式。最佳范例是我最喜欢的 4×4 矩阵：

$$A=\begin{bmatrix}2&-1&0&0\\-1&2&-1&0\\0&-1&2&-1\\0&0&-1&2\end{bmatrix}\quad\text{有}\quad\det A=5$$

我们可以用 3 种方法求出行列式：**主元，大公式，余子式**。

1.主元的乘积是 $2\cdot 3/2\cdot 4/3\cdot 5/4$，消除之后得到 5。

2.方程式(8)的**大公式**有 $4!=24$ 项，只有 5 项不是零：【方程式(8)在 261 页】

$$\det A=16-4-4-4+1=5$$

16 来自 A 的对角线 $2\cdot 2\cdot 2\cdot 2$。–4 与+1 来自哪儿？当你能够找出这 5 项，你就已经了解公式(8)。

3. 第一行的数字 2, –1, 0, 0 乘上自己来自其他行的余子式 4, 3, 2, 1，得到 $2\cdot 4-1\cdot 3=5$。这些余子式是 3×3 的行列式，余子式使用的行与列，是第一行的单元未使用的部分。

行列式的每一项使用每个行与列一次。

主元公式

当消元法得到 $A=LU$，主元 $d_1,\ldots,d_n$ 在上三角 U 的对角线。如果没有牵涉到行交换，这些**主元乘积**就是行列式：

$$\det A=(\det L)\,(\det U)=(1)\,(d_1\,d_2\ldots d_n)\qquad(1)$$

这个 $\det A$ 的公式出现在段落 5.1，后面有可能出现行交换，排列进入时 $PA=LU$，P 的行列式是–1 或+1。

(det P) (det A) = (det L) (det U) 得到 $\det A = \pm (d_1\, d_2 \ldots d_n)$ (2)

范例 1 一次行交换得到主元 4, 2, 1 以及重要的负号：

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad PA = \begin{bmatrix} 4 & 5 & 6 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} \qquad \det A = -(4)\,(2)\,(1) = -8$$

奇数次的行交换(一次)代表 $\det P = -1$。

下一个例子没有行交换。它可能是我们分解得到 LU(当时是 3×3)的第一个矩阵，重要的是我们可以直接跳到 $n \times n$。主元得到行列式，我们也会看到行列式如何给出主元。

范例 2 三角矩阵 A 的前几个主元是 2, 3/2, 4/3，后面是 5/4 与 6/5 最终是 $(n+1)/n$。分解这个 $n \times n$ 可以看出行列式：

$$\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \bullet & \\ & & \bullet & \bullet & -1 \\ & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -\frac{1}{2} & 1 & & & \\ & -\frac{2}{3} & 1 & & \\ & & \bullet & \bullet & \\ & & & -\frac{n-1}{n} & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & & & \\ & \frac{3}{2} & -1 & & \\ & & \frac{4}{3} & -1 & \\ & & & \bullet & \bullet \\ & & & & \frac{n+1}{n} \end{bmatrix}$$

主元在 U(最后一个矩阵)的对角线。当 2 与 3/2 与 4/3 与 5/4 相乘，分数对消，这个 4×4 矩阵的行列式是 5。3×3 的行列式是 4，$n \times n$ 的行列式是 $n+1$：

–1, 2, –1 矩阵 $\det A = (2)\,(3/2)\,(4/3) \ldots ((n+1)/n) = n+1$

重点：前面的主元与原始矩阵的左上角有关，这是没有行交换的矩阵都适用的规则。

前 k 个主元来自 A 的左上角的 $k \times k$ 矩阵 A_k 。

角落子矩阵 A_k 的行列式是 $d_1\, d_2 \ldots d_k$ (前 k 个主元) 。

1×1 矩阵 A_1 只包含第一个主元 d_1，这是 $\det A_1$。角落的 2×2 矩阵有 $\det A_2 = d_1 d_2$，最终 $n \times n$ 矩阵是所有的主元乘积。

当我们开始处理整个矩阵，消元法针对左上角矩阵 A_k，我们假设没有行交换——则 $A = LU$ 且 $A_k = L_k U_k$。一个行列式除以前一个行列式($\det A_k$ 除以 $\det A_{k-1}$)会全部抵消只留下最近的主元 d_k。**每个主元都是行列式的比值。**

来自行列式的主元　第 k 个主元是 $d_k = \dfrac{d_1 d_2 \ldots d_k}{d_1 d_2 \ldots d_{k-1}} = \dfrac{\det A_k}{\det A_{k-1}}$ (3)

当所有的左上子矩阵有 $\det A_k \neq 0$，我们就不需要行交换。

请尊重版权与译者的劳动成果，侵权必究!

主元很容易计算，他们聚集了很多信息——足够求取行列式，但是它很难与原始的 a_{ij} 产生联结。如果我们回到规则 1-2-3，这个部分就比较清晰，规则 1-2-3 就是线性与符号反转与 $\det I = 1$。我们希望推导出单一又明确的行列式公式，直接来自单元 a_{ij}。

这个公式有 $n!$ 个项，它的大小会快速增加，这是因为 $n! = 1, 2, 6, 24, 120, \ldots$。当 $n = 11$ 大概会有百万个项。$n = 2$，这两项是 ad 与 bc。半数的项是负号(如同 $-bc$)，另外半数是正号(如同 ad)。$n = 3$ 有 $3! = (3)(2)(1)$ 个项。下列是六个项：

$$3\times3 \text{ 行列式 } \begin{vmatrix} a_{11} & a_{12} & \boldsymbol{a_{13}} \\ \boldsymbol{a_{21}} & a_{22} & a_{23} \\ a_{31} & \boldsymbol{a_{32}} & a_{33} \end{vmatrix} = \begin{matrix} +a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + \boldsymbol{a_{13}a_{21}a_{32}} \\ -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{matrix} \tag{4}$$

注意模式。每个乘积，比如 $a_{11}\,a_{23}\,a_{32}$ 各有**来自每一行的一个单元，也有来自每一列的一个单元**。列顺序 1, 3, 2 代表这个特定项有一个负号，$a_{13}\,a_{21}\,a_{32}$ 的列顺序 3, 1, 2 有一个正号(粗体字)，正是**排列**告诉我们符号。

下个步骤($n = 4$)带来 $4! = 24$ 项，有 24 种方式从每个行每个列选取一个单元。沿着主对角线，$a_{11}\,a_{22}\,a_{33}\,a_{44}$ 的列顺序是 1, 2, 3, 4，有一个正号，这是单位排列。

我从 $n = 2$ 开始推导大公式，目标是以系统化的方式得到 $ad - bc$。把每一行分成两个比较简单的行：$[a \quad b] = [a \quad 0] + [0 \quad b]$ 且 $[c \quad d] = [c \quad 0] + [0 \quad d]$

现在应用线性，首先在行 1(行 2 固定)然后在行 2(行 1 固定)：

$$\begin{aligned} \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \qquad (\text{分割行1}) \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} \qquad (\text{分割行2}) \end{aligned} \tag{5}$$

最后一行有 $2^2 = 4$ 个行列式。第一个与第四个是零，因为一个行是另一个行的倍数。我们剩下 $2! = 2$ 个行列式要计算：

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = ad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + bc \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = ad - bc$$

这项分割得到排列矩阵，他们的行列式提供一个正号或是负号。排列可以看出列的序列，本案例是(1, 2)或(2, 1)。

现在尝试 $n = 3$，每个行分割成 3 个简单行，比如$[a_{11} \quad 0 \quad 0]$。利用每个行的线性，$\det A$ 分割成 $3^3 = 27$ 个简单的行列式。如果一个列的选项重复了——如果我们也选取了行$[a_{21} \quad 0 \quad 0]$——则这个简单的行列式是零。

我们只关注单元 a_{ij} 是来自不同的列，例如(3, 1, 2):

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & \boldsymbol{a_{13}} \\ \boldsymbol{a_{21}} & & \\ & \boldsymbol{a_{32}} & \end{vmatrix}$$

$$+ \begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}$$

六个项

列的排序有 3! = 6 个方式，所以有 6 个行列式。(1, 2, 3)的 6 种排列包含来自 $P = I$ 的单位排列(1, 2, 3)。

$$\textbf{列的数字} = (1,2,3),\ (2,3,1),\ (\mathbf{3},\mathbf{1},\mathbf{2}),\ (1,3,2),\ (2,1,3),\ (3,2,1) \tag{6}$$

后 3 个是奇排列(一次交换)，前 3 个是偶排列(0 或 2 次交换)。当列的顺序是(**3**, **1**, **2**)，我们选取单元 $a_{13}\ a_{21}\ a_{32}$ ––这项特定的列顺序带来一个正号(2 次交换)，A 的行列式现在分割成 6 个简单项，取出因子 a_{ij}:

$$\begin{aligned} \det A = a_{11}a_{22}a_{33}\begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31}\begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{\mathbf{13}}a_{\mathbf{21}}a_{\mathbf{32}}\begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix} \\ + a_{11}a_{23}a_{32}\begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33}\begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31}\begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix} \end{aligned} \tag{7}$$

前 3 个(偶排列)有 $\det P = +1$，后 3 个(奇排列)有 $\det P = -1$，我们已经用系统化的方式证明 3×3 的公式。

现在你可以看到 $n \times n$ 公式，有 $n!$个列的顺序。列(1, 2, ..., n)进到每一个可能的顺序($\alpha, \beta, ..., \omega$)，从行 1 选取 $a_{1\alpha}$，从行 2 选取 $a_{2\beta}$，最终从行 n 选取 $a_{n\omega}$，行列式包含 $a_{1\alpha}\, a_{2\beta} \ldots a_{n\omega}$ 的乘积再乘上+1 或–1。半数顺序的符号是–1。

A 的行列式是这 $n!$简单行列式的总和，再乘上+1 或–1，简单行列式 $a_{1\alpha}\, a_{2\beta} \ldots a_{n\omega}$ 选取来自每个行每个列的一个单元。对 5×5 而言，$a_{15}\ a_{22}\ a_{33}\ a_{44}\ a_{51}$ 会有 $\det P = -1$，因为交换了 5 与 1。

$$\det A = n!\ \text{个列的排列}\ P = (\alpha,\ \beta, \ldots, \omega)\ \text{的总和}$$

$$\sum (\det P)\, a_{1\alpha}\, a_{2\beta} \ldots a_{n\omega} \ = \ \textbf{大公式} \tag{8}$$

2×2 的情形是 $+a_{11}\,a_{22} - a_{12}\,a_{21}$ (就是 $ad - bc$)，此处 P 是(1, 2)或(2, 1)。

请尊重版权与译者的劳动成果，侵权必究!

3×3 的情形有 3 个“右斜下”乘积(见问题 28)与 3 个“左斜下”乘积。警告：很多人相信这个模式应该也可以在 4×4 的案例使用，他们只选取了 8 项—但是我们需要 24 项。

范例 3 (U 的行列式) 当 U 是上三角，$n!$ 个乘积中只有一个不是零，这一项来自对角线：$\det U = +u_{11}u_{22}\ldots u_{nn}$ 。其他的列顺序至少选取一个在对角线下方的单元，U 的这些地方都是零。很快的我们选取一个数比如 $u_{21}=0$，在方程式(8)的那一项确定是零。

当然 $\det I = 1$，唯一的非零项是来自对角线的 $+(1)(1)\ldots(1)$。

范例 4 假设 Z 是除了列 3 不一样的单位矩阵，则

$$Z \text{ 的行列式 } = \begin{vmatrix} 1 & 0 & a & 0 \\ 0 & 1 & b & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & d & 1 \end{vmatrix} \text{ 是 } c \tag{9}$$

符号为正的(1)(1)(c)(1)这一项来自主对角线，有 $4! = 24$ 个乘积(选取一个来自每个行每个列的因子)，但是其他 23 个乘积都是零。理由：如果我们从列 3 选取 a, b 或 d，这一列就用尽了，来自列 3 唯一可用的选择是零。

此处有一个相同答案却是不同的理由，若 $c=0$，则 Z 有一整行的零，$\det Z = c = 0$ 是正确的。若 c 不是零，使用消元法，从其他行减去乘数乘行 3，把 a, b, d 去除，剩下一个对角矩阵，$\det Z = c$。

这个例子很快会用在“克拉默规则”，如果我们把 a, b, c, d 移进 Z 的第一列，行列式是 $\det Z = a$ (为什么？) 改变 I 的一个列使得 Z 有容易的行列式，来自主对角线。

范例 5 假设 A 有 1's 在主对角线的上方与下方，此处 $n=4$：

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ 与 } P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ 的行列式 } = 1$$

第一行唯一的非零选择在列 2，行 4 唯一的非零选择在列 3，则行 2 与 3 必须选择列 1 与 4。换言之，$\det P = \det A$。P 行列式是+1(两次交换得到 2, 4, 1, 3)，因此 $\det A = +1$。

余子式求行列式

方程式(8)是行列式的直接定义，一次性把全部都给你—但是你必须消化领悟它。毕竟 $n!$ 个项都必须满足规则 1-2-3 (然后其他性质 4-10 跟上)。最简单的是 $\det I = 1$，已经验证完毕。

当你分离来自第一行的因子 a_{11} 或 a_{12} 或 $a_{1\alpha}$，你会看到线性。对 3×3 而言，行列式的 6 个项可以分成 3 组配对：

$$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \qquad (8)$$

括号中的 3 个项称为“**余子式**”，他们是来自行 2 与 3 的 2×2 的行列式。第一行贡献了因子 a_{11}, a_{12}, a_{13}，下方的行贡献了**余子式** C_{11}, C_{12}, C_{13}。当然行列式 $a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$ 与 a_{11}, a_{12}, a_{13} 有线性的相关——这是规则 3。

a_{11} 的余子式是 $C_{11} = a_{22}a_{33} - a_{23}a_{32}$，你可以在下方的分割图看出来：

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & a_{23} \\ a_{31} & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & a_{22} & \\ a_{31} & a_{32} & \end{vmatrix}$$

我们仍然选择一个来自每一行每一列的单元，由于 a_{11} 用尽了行 1 与列 1，剩下一个 2×2 行列式作为它的余子式。

你一直都需要注意符号，这个跟随 a_{21} 的 2×2 行列式好像是 $a_{21}a_{33} - a_{23}a_{31}$，但是在余子式 C_{12} 中，它的符号是相反的，$a_{12}C_{12}$ 是正确的 3×3 行列式，沿着第一行的余子式的符号模式是正–负–正–负。你画掉行 1 与列 1 得到一个大小是 $n-1$ 的子矩阵 M_{1j}，它的行列式乘上符号 $(-1)^{1+j}$ 得到余子式：

$$\text{行 1 的余子式是 } C_{1j} = (-1)^{1+j} \det M_{1j}$$

$$\textbf{余子式扩展}\text{是 } \det A = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n} \qquad (11)$$

在大公式(8)中，a_{11} 乘上的一些项组合在一起得到 $C_{11} = \det M_{11}$，符号是 $(-1)^{1+1}$，意思是正号。方程式(11)是方程式(8)与方程式(10)的另一种形式，来自行 1 的因子乘上只用到其他行的余子式。

注解 行 1 可以做的，行 i 一样可以。在某一行的单元 a_{ij} 同样有余子式 C_{ij}，他们是阶数 $n-1$ 的行列式乘上 $(-1)^{i+j}$。由于 a_{ij} 代表行 i 列 j，子矩阵 M_{ij} 移除行 i 列 j。下表展示 a_{43} 与 M_{43} (移除行 4 与列 3)，符号 $(-1)^{4+3}$ 乘 M_{43} 的行列式得到 C_{43}。符號矩阵显示±的模式：

$$A = \begin{bmatrix} \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ & & a_{43} & \end{bmatrix} \quad \text{符号} (-1)^{i+j} = \begin{bmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

行列式是 A 的任意行 i 与使用其他行得到的余子式的点积：

余子式公式 $\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$ (12)

每个余子式 C_{ij} (阶数 $n-1$，移除行 i 与列 j)包含它的正确符号

余子式 $C_{ij} = (-1)^{i+j} \det M_{ij}$

阶数 n 的行列式是阶数 $n-1$ 的行列式的组合，递归的工作还要持续下去，每个子行列式再分成阶数 $n-2$ 的行列式。我们可以依靠方程式(12)定义所有的行列式，规则是从 n 到 $n-1$ 到 $n-2$，最终到阶数 1。定义 1×1 行列式为$|a|$等于数字 a，则余子式方法完成。

我们偏好从性质创建 $\det A$ (线性，符号反转，$\det I = 1$)，大公式(8)与余子式公式(10)–(12)都是遵循这些规则。最后一个来自规则的公式是 $\det A = \det A^{\mathrm{T}}$，我们可以采用余子式扩展，沿着一个列而不是沿着一个行。沿着列 j 的单元是 a_{1j} 到 a_{nj}，余子式是 C_{1j} 到 C_{nj}，行列式是下列的点积：

沿着列 j 的余子式 $\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$ (13)

当矩阵有很多零时，余子式非常有用——如同下个例子。

范例 6 –1, 2, –1 矩阵在第一行只有两个非零数，所以只有两个余子式 C_{11} 与 C_{12} 与行列式有关。我会高亮 C_{12}：

$$A = \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} = 2\begin{vmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{vmatrix} - (-1)\begin{vmatrix} \mathbf{-1} & \mathbf{-1} & \\ & \mathbf{2} & \mathbf{-1} \\ & \mathbf{-1} & \mathbf{2} \end{vmatrix} \quad (14)$$

在右侧你首先看到 2 乘 C_{11}，C_{11} 要移除行 1 与列 1，余子式 C_{11} 与原始 A 有同样的–1, 2, –1 模式——但是大小变小了。

要计算粗体字的 C_{12}，沿着第一列使用余子式，唯一的非零数在顶端，贡献了另一个–1(所以我们回过头要加负号)。它的余子式是–1, 2, –1 的行列式，大小是 2×2，比原始的 A 小了 2 阶。

概要：每个阶数 n 的行列式 D_n 来自 D_{n-1} 与 D_{n-2}：

$$D_4 = 2D_3 - D_2 \quad \text{一般式是} \quad D_n = 2D_{n-1} - D_{n-2} \quad (15)$$

直接计算得到 $D_2 = 3$ 及 $D_3 = 4$，方程式(14)得到 $D_4 = 2(4) - 3 = 5$，行列式 3, 4, 5 满足公式 $D_n = n + 1$，则 D_n 等于 $2n - (n-1)$。“**特殊三对角答案**”同样来自范例 2 的主元乘积。

范例 7 这是相同的矩阵，除了第一单元(左上角)现在是 1：

$$B_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

矩阵所有的主元都变成 1，所以行列式是 1。行列式是如何来自余子式？在行 1 扩展，余子式与范例 6 相同，只是 $a_{11} = 2$ 变成 $b_{11} = 1$：

$$\det B_4 = D_3 - D_2 \quad \text{而不是} \quad \det A_4 = 2D_3 - D_2$$

B_4 的行列式是 $4 - 3 = 1$，每个 B_n 的行列式是 $n - (n-1) = 1$。

如果你也把最后一个 2 变成 1，为什么行列式是 0？

主要观念的复习

1. 没有行交换，$\det A =$ (主元乘积)。在 A 的左上角，$\det A_k =$ (前 k 个主元的乘积)。
2. 大公式(8)的每一项用到每一行与每一列。$n!$ 的半数有正号(当 $\det P = +1$)，半数有负号。
3. 余子式 C_{ij} 是 $(-1)^{i+j}$ 乘较小的行列式，这些行列式会把行 i 列与 j 移除(因为 a_{ij} 用了那一行与列)。
4. 行列式是任意行与它的余子式的行的点积，当 A 的某一行有很多的零，我们只需要一些余子式。

已解范例

5.2A 一个海森堡(Hessenberg)矩阵是一个三角矩阵，这个矩阵多了一个对角线。使用行 1 的余子式证明 4×4 的行列式满足斐波那契规则：$|H_4| = |H_3| + |H_2|$。相同的规则适用于所有的大小，$|H_n| = |H_{n-1}| + |H_{n-2}|$。哪个斐波那契数字等于 $|H_n|$？

$$H_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad H_3 = \begin{bmatrix} 2 & 1 & \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad H_4 = \begin{bmatrix} 2 & 1 & & \\ 1 & 2 & 1 & \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

解 H_4 的余子式 C_{11} 是行列式$|H_3|$，我们也需要 C_{12}(粗体字)：

$$C_{12} = -\begin{vmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{2} \end{vmatrix} = -\begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix} + \begin{vmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{vmatrix}$$

行 2 与 3 维持相同然后使用行 1 的线性。右侧的两个行列式是$-|H_3|$与$+|H_2|$，则 4×4 的行列式是

$$|H_4| = 2C_{11} + 1C_{12} = 2|H_3| - |H_3| + |H_2| = |H_3| + |H_2|$$

实际数字是$|H_2| = 3$ 与$|H_3| = 5$ (当然$|H_1| = 2$)。由于$|H_n| = 2, 3, 5, 8$ 遵循斐波那契规则$|H_n| = |H_{n-1}| + |H_{n-2}|$，它必须是$|H_n| = F_{n+2}$。

5.2B 这些问题使用±符号(偶与奇 *P*'s)在 det *A* 的大公式中：

1. 若 *A* 是 10×10 的全 1 矩阵，大公式如何得到 det *A* = 0？
2. 如果你把所有的 *n*! 排列乘在一起得到单一的 *P*，*P* 是奇或是偶？
3. 如果你把每个 a_{ij} 乘上分数 *i* / *j*，为什么 det *A* 不变？

解 在问题 1 中，所有的 $a_{ij} = 1$，大公式(8)的所有乘积都是 1，半数带正号，半数带负号，互相消除得到 det *A* = 0。(当然全 1 矩阵是奇异，我假设 $n > 1$)。

问题 2 中，乘积$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$得到奇排列，同样对于 3×3，3 个奇排列乘一起得到奇(无关顺序)。但是对于 $n > 3$，所有排列的乘积是偶数。因为 *n*!包含 4 的因子，所以 *n*!/2 的奇排列还是偶数。

问题 3 中，每个 a_{ij} 乘 *i* / *j*，所以大公式的每个乘积 $a_{1\alpha}a_{2\beta}\ldots a_{n\omega}$会乘上所有的行数字 $i = 1, 2, \ldots, n$ 以及除以所有的列数字 $j = 1, 2, \ldots, n$。(列来自相同的的排列顺序!)，每个乘积都不变所以 det *A* 维持原状。

问题 3 的另一方法：当**行 *i*** 乘 *i*，如同使用对角线矩阵 *D* = diag(1 : n)乘矩阵 *A*；当列 *j* 除 *j*，如同右乘 D^{-1}。根据乘积规则，DAD^{-1} 的行列式与 det *A* 相同。

问题集 5.2

问题 1-10 使用有 *n*！项的大公式：$|A| = \Sigma \pm a_{1\alpha}a_{2\beta}\ldots a_{n\omega}$，每一项使用每个行与每个列。

1. 从六个项计算 *A*, *B*, *C* 的行列式，他们的行是无关吗？

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 4 & 4 \\ 5 & 6 & 7 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

2　计算 A, B, C, D 的行列式，他们的列是无关吗？

$$A=\begin{bmatrix}1&1&0\\1&0&1\\0&1&1\end{bmatrix}\qquad B=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}\qquad C=\begin{bmatrix}A&0\\0&A\end{bmatrix}\qquad D=\begin{bmatrix}A&0\\0&B\end{bmatrix}$$

3　证明 $\det A=0$，而且与 5 个标示 x 的非零数无关：

$$A=\begin{bmatrix}x&x&x\\0&0&x\\0&0&x\end{bmatrix}$$

行1的余因子为何？
A的秩为何？
$\det A$的六项为何？

4　从 4 个不同的行与列，找出两种方法选择非零数：

$$A=\begin{bmatrix}1&0&0&1\\0&1&1&1\\1&1&0&1\\1&0&0&1\end{bmatrix}\qquad B=\begin{bmatrix}1&0&0&2\\0&3&4&5\\5&4&0&3\\2&0&0&1\end{bmatrix}$$ (B 与 A 有相同的的零)

$\det A = 1+1$ 或 $1-1$ 或 $-1-1$？　$\det B =$ ？

5　4×4 矩阵中，如何放置最少个数的零确保 $\det A = 0$。如何放置最多个数的零确保 $\det A \neq 0$。

6　(a) 若 $a_{11}=a_{22}=a_{33}=0$，$\det A$ 的六项有多少项为零？

(b) 若 $a_{11}=a_{22}=a_{33}=a_{44}=0$，$\det A$ 的 24 项乘积 $a_{1j}a_{2k}a_{3l}a_{4m}$ 有多少项为零？

7　5×5 排列矩阵有多少个 $\det P=+1$？这些是偶排列，求出需要 4 次交换得到单位矩阵的排列矩阵。

8　若 $\det A$ 不是零，大公式(8)的 $n!$项至少有一个不是零。从大公式推导 A 的行的某种顺序使得对角线不存在零。(不使用来自消元法的 P；PA 可以有零在对角线。)

9　单元不是 1's 就是–1's 的 3×3 矩阵，证明最大的行列式是 4。

10　(1, 2, 3, 4)有多少偶排列？他们是哪些？额外积分：求 4×4 的矩阵 $I+P_{\text{even}}$ 的所有可能的行列式。

问题 11-22 使用余子式 $C_{ij}=(-1)^{i+j}\det M_{ij}$，移除行 i 与列 j。

11　求出所有的余子式并放进余子式矩阵 C 与 D，求 AC 与 $\det B$。

$$A=\begin{bmatrix}a&b\\c&d\end{bmatrix}\qquad B=\begin{bmatrix}1&2&3\\4&5&6\\7&0&0\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

12 求余子式矩阵 C 与 AC^{T}，比较 AC^{T} 与 A^{-1}：

$$A=\begin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}\qquad A^{-1}=\frac{1}{4}\begin{bmatrix}3&2&1\\2&4&2\\1&2&3\end{bmatrix}$$

13 $n\times n$ 行列式 C_n 有 1's 在主对角线的上方与下方：

$$C_1=\begin{vmatrix}0\end{vmatrix}\qquad C_2=\begin{vmatrix}0&1\\1&0\end{vmatrix}\qquad C_3=\begin{vmatrix}0&1&0\\1&0&1\\0&1&0\end{vmatrix}\qquad C_4=\begin{vmatrix}0&1&0&0\\1&0&1&0\\0&1&0&1\\0&0&1&0\end{vmatrix}$$

(a) C_1, C_2, C_3, C_4 的行列式为何？

(b) 利用余子式求出 C_n 与 C_{n-1} 与 C_{n-2} 的关系。求 C_{10}。

14 问题 13 的矩阵有 1's 在主对角线的上方与下方，矩阵继续往下，哪个列顺序(如果有的话)得到全部的 1's？说明为什么当 $n=4, 8, 12, ...$的排列是偶，当 $n=2, 6, 10, ...$的排列是奇。然后

$$C_n=0\ (\text{奇数 } n)\qquad C_n=1\ (n=4, 8, ...)\qquad C_n=-1\ (n=2, 6, ...)$$

15 阶数 n 的**三对角** 1, 1, 1 矩阵的行列式 E_n：

$$E_1=\begin{vmatrix}1\end{vmatrix}\qquad E_2=\begin{vmatrix}1&1\\1&1\end{vmatrix}\qquad E_3=\begin{vmatrix}1&1&0\\1&1&1\\0&1&1\end{vmatrix}\qquad E_4=\begin{vmatrix}1&1&0&0\\1&1&1&0\\0&1&1&1\\0&0&1&0\end{vmatrix}$$

(a) 利用余子式证明 $E_n=E_{n-1}-E_{n-2}$。

(b) 从 $E_1=1$ 与 $E_2=0$ 开始，求 $E_3, E_4, ..., E_8$。

(c) 注意前式最终的重复情形，求 E_{100}。

16 阶数 n 的 1, 1, –1 **三对角**矩阵的行列式是 F_n：

$$F_2=\begin{vmatrix}1&-1\\1&1\end{vmatrix}=2\qquad F_3=\begin{vmatrix}1&-1&0\\1&1&-1\\0&1&1\end{vmatrix}=3\qquad F_4=\begin{vmatrix}1&-1&0&0\\1&1&-1&0\\0&1&1&-1\\0&0&1&0\end{vmatrix}\neq 4$$

使用余子式扩展，证明 $F_n=F_{n-1}+F_{n-2}$。这些行列式是斐波那契数字 1, 2, 3, 5, 8, 13, ...。这个序列通常从 1, 1, 2, 3 开始(里面有两个 1)，所以我们的 F_n 是通常的 F_{n+1}。

网址：linear.neocities.org 邮箱：876152660@qq.com

17 矩阵 B_n 是–1, 2, –1 矩阵 A_n 把 $a_{11}=2$ 变成 $b_{11}=1$ 的矩阵，利用 B_4 的最后一行扩展余子式，证明 $|B_4|=2|B_3|-|B_2|=1$。

$$B_4=\begin{bmatrix}1&-1&&\\-1&2&-1&\\&-1&2&-1\\&&-1&2\end{bmatrix}\quad B_3=\begin{bmatrix}1&-1&\\-1&2&-1\\&-1&2\end{bmatrix}\quad B_2=\begin{bmatrix}1&-1\\-1&2\end{bmatrix}$$

当每个 $|B_n|=1$ 时，满足递归关系 $|B_n|=2|B_{n-1}|-|B_{n-2}|$。这个递归与范例 6 的递归相同，差别在于大小为 $n=1, 2, 3$ 的起始值是 1, 1, 1。

18 回到问题 17 的 B_n，除了 $b_{11}=1$ 之外，它与 A_n 相同。所以使用第一行的线性，其中[1 –1 0] = [2 –1 0] – [1 0 0]：

$$|B_n|=\begin{vmatrix}1&-1&0\\-1&&\\&A_{n-1}&\\0&&\end{vmatrix}=\begin{vmatrix}2&-1&0\\-1&&\\&A_{n-1}&\\0&&\end{vmatrix}-\begin{vmatrix}1&0&0\\-1&&\\&A_{n-1}&\\0&&\end{vmatrix}$$

线性使得 $|B_n|=2|A_n|-|A_{n-1}|=$ _____。

19 说明为什么下列的范德蒙德行列式包含 x^3 而不是 x^4 或 x^5：

$$V_4=\begin{bmatrix}1&a&a^2&a^3\\1&b&b^2&b^3\\1&c&c^2&c^3\\1&x&x^2&x^3\end{bmatrix}$$

在 $x=$ ____, ____与____，行列式是零。x^3 的余子式是 $V_3=(b-a)(c-a)(c-b)$，则 $V_4=(b-a)\,(c-a)\,(c-b)\,(x-a)\,(x-b)\,(x-c)$。

20 求出 G_2 与 G_3，然后使用消元法求 G_4，你能够预测 G_n 吗？

$$G_2=\begin{vmatrix}0&1\\1&0\end{vmatrix}\quad G_3=\begin{vmatrix}0&1&1\\1&0&1\\1&1&0\end{vmatrix}\quad G_4=\begin{vmatrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{vmatrix}$$

21 对于下列 1, 3, 1 矩阵，计算 S_1, S_2 与 S_3，利用斐波那契猜测并检验 S_4。

$$S_1=|3|\quad S_2=\begin{vmatrix}3&1\\1&3\end{vmatrix}\quad S_3=\begin{vmatrix}3&1&0\\1&3&1\\0&1&3\end{vmatrix}$$

22 问题 21 中，把矩阵左上角的 3 改成 2，为什么这样会有从行列式 S_n 减去 S_{n-1}？证明新矩阵的行列式变成斐波那契数字 2, 5, 13 (永远是 F_{2n+1})。

请尊重版权与译者的劳动成果，侵权必究!

问题 23-26 有关方块矩阵与方块行列式。

23 4×4 矩阵中有 2×2 的方块，你不可以一直使用方块行列式：

$$\begin{vmatrix} A & B \\ 0 & D \end{vmatrix} = |A||D| \quad \text{但是} \quad \begin{vmatrix} A & B \\ C & D \end{vmatrix} \neq |A||D| - |B||C|$$

(a) 为什么第一个叙述是正确的? B 好像没有参与。

(b) 举例说明当 C 参与时，等式失败(如所示)。

(c) 举例说明 det $(AD - CB)$的答案也是错误的。

24 使用方块乘法，$A = LU$ 有左上角的 $A_k = L_k U_k$ ：

$$A = \begin{bmatrix} A_k & * \\ * & * \end{bmatrix} = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix} \begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix}$$

(a) 假设 A 的前 3 个主元是 2, 3, –1，求出下列的行列式：L_1, L_2, L_3(对角线都是 1's)，U_1, U_2, U_3 以及 A_1, A_2, A_3。

(b) 若 A_1, A_2, A_3 的行列式是 5, 6, 7，从方程式(3)求出 3 个主元。

25 方块消元法从第二行$[C \quad D]$减去 CA^{-1} 乘第一行$[A \quad B]$，会在右下角得到舒尔补充 $D - CA^{-1}B$：

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}$$

对上述方块矩阵取行列式，若 A^{-1} 存在，证明下列的正确规则：

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \left| D - CA^{-1}B \right| = |AD - CB| \quad \text{假设 } AC = CA$$

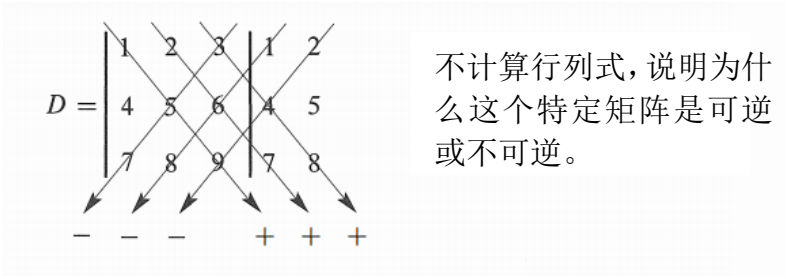
26 若 A 是 $m \times n$ 与 B 是 $n \times m$，方块乘法得到 det M = det AB ：

$$M = \begin{bmatrix} 0 & A \\ -B & I \end{bmatrix} = \begin{bmatrix} AB & A \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -B & I \end{bmatrix}$$

若 A 是单一行且 B 是单一列，求 det M ？若 A 是单一列且 B 是单一行，求 det M ？两者各做一个 3×3 的例子。

27 (微积分问题) 证明 det A 相对于 a_{11} 的导数是余子式 C_{11}。其他的单元固定--我们只改变 a_{11}。

28 3×3 行列式有 3 个“右斜下”乘积以及 3 个带负号的“左斜下”乘积，计算比如(1)(5)(9) = 45 的六项求 D。



不计算行列式，说明为什么这个特定矩阵是可逆或不可逆。

29 对问题 15 的 E_4 来说，大公式(8)的 4！= 24 项中的 5 项是非零数，找出这 5 项证明 $E_4 = -1$。

30 对 4×4 三对角第二差分矩阵(单元−1, 2, −1)，找出大公式中的 5 项使得 $\det A = 16 - 4 - 4 - 4 + 1$。

31 利用行 1 的余子式求出循环 P 的行列式，然后再使用大公式。4, 1, 2, 3 需要多少次交换才能变成 1, 2, 3, 4？$|P^2| = 1$ 或 -1？

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad P^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

挑战问题

32 问题 21 的 1, 3, 1 矩阵的余子式得到递归关系 $S_n = 3S_{n-1} - S_{n-2}$，惊奇的发现这项递归得到每隔一个的斐波那契数字。以下是挑战：

利用 $F_{2n+2} = 3F_{2n} - F_{2n-2}$ 的证明来展示 S_n 是斐波那契数字 F_{2n+2}。持续使用斐波那契规则：从 $k = 2n + 2$ 开始，$F_k = F_{k-1} + F_{k-2}$。

33 对称帕斯卡矩阵的行列式是 1。如果我把 n, n 单元减 1，为什么行列式变成 0？(使用规则 3 或余子式。)

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix} = 1 \text{ (已知)} \qquad \det \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 19 \end{bmatrix} = 0 \text{ (需要说明)}$$

请尊重版权与译者的劳动成果，侵权必究！

34 这个问题有两个方法证明 $\det A = 0$ (x's 是任意数)：

$$A = \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \end{bmatrix}$$

(a) 你如何知道行是线性相关？

(b) 说明为什么 A 的行列式的大公式的全部 120 项都是零。

35 若$|\det A| > 1$，证明 A 的次方 A^n 不可能有界。但是若$|\det A| \le 1$，证明 A^n 的某些单元仍然可能增长太高。特征值会提供稳定度的良好试验，行列式只是告诉我们一个数字而已。

5.3 克拉默规则，逆矩阵与体积

1 $\boldsymbol{A^{-1} = C^{\mathrm{T}}/\det A}$，则$(A^{-1})_{ij} =$ 余子式 $\boldsymbol{C_{ji}}$ 除以 A 的行列式。

2 克拉默规则：从 $x_j = \det$ (A 的列 j 换成 $\boldsymbol{b}$) / $\det A$，计算 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$。

3 若平行四边形的 4 个角是$(0, 0), (a, b), (c, d)$与$(a + c, b + d)$，则面积 $= |ad - bc|$。

4 若 A 的行(或 A 的列)是盒子的边，则盒子的体积 $= |\det A|$。

5 **叉积** $\boldsymbol{w} = \boldsymbol{u} \times \boldsymbol{v}$ 是 $\det\begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$，

注意 $\boldsymbol{v} \times \boldsymbol{u} = -(\boldsymbol{u} \times \boldsymbol{v})$。
w_1, w_2, w_3 是行 1 的余子式。
注意 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u} = 0$ 且 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{v} = 0$。

本段落求解 $A\boldsymbol{x} = \boldsymbol{b}$ 也计算 A^{-1}——利用代数而不是消元法。所有你看到的公式都有除以 $\det A$，A^{-1} 与 $A^{-1}\boldsymbol{b}$ 的每个单元都是一个行列式除以 A 的行列式。让我们从克拉默(Cramer's Rule)规则开始。

克拉默规则求解 $A\boldsymbol{x} = \boldsymbol{b}$，一个简洁的公式得到第一分量 x_1。把 $\boldsymbol{x}$ 取代 I 的第一列得到有行列式 x_1 的矩阵。当你左乘 A，第一列变成 $A\boldsymbol{x}$ 就是 $\boldsymbol{b}$，B_1 的其他列从 A 复制：

关键概念
$$\begin{bmatrix} & A & \end{bmatrix}\begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1 \tag{1}$$

我们一次乘一个列，三个矩阵取行列式求得 x_1：

乘积规则 $(\det A)(x_1) = \det B_1$ 或是 $x_1 = \boldsymbol{(\det B_1)/(\det A)}$。 (2)

这是克拉默规则的 $\boldsymbol{x}$ 的第一分量！改变 A 的一个列得到 B_1。现在要求 x_2 与 B_2，把向量 $\boldsymbol{x}$ 与 $\boldsymbol{b}$ 放在 I 与 A 的第二列：

同样概念
$$\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix}\begin{bmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{b} & \boldsymbol{a}_3 \end{bmatrix} = B_2 \tag{3}$$

取行列式发现$(\det A)(x_2) = \det B_2$，这个得到 $x_2 = \boldsymbol{(\det B_2)/(\det A)}$。

范例 1 求解 $3x_1 + 4x_2 = \boldsymbol{2}$ 与 $5x_1 + 6x_2 = \boldsymbol{4}$ 需要 3 个行列式：

$$\det A = \begin{vmatrix} 3 & 4 \\ 5 & 6 \end{vmatrix} \quad \det B_1 = \begin{vmatrix} \boldsymbol{2} & 4 \\ \boldsymbol{4} & 6 \end{vmatrix} \quad \det B_2 = \begin{vmatrix} 3 & \boldsymbol{2} \\ 5 & \boldsymbol{4} \end{vmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

矩阵 A, B_1, B_2 的行列式是–2, –4 与 2，所有的比值都是除以 $\det A = 2$：

克拉默规则 若 $\det A$ 不是零，$A\boldsymbol{x} = \boldsymbol{b}$ 利用行列式求解：

$$x_1 = \frac{\det B_1}{\det A} \quad x_2 = \frac{\det B_2}{\det A} \quad \cdots\cdots \quad x_n = \frac{\det B_n}{\det A} \tag{4}$$

矩阵 $\boldsymbol{B_j}$ 是把 $\boldsymbol{A}$ 的第 $\boldsymbol{j}$ 列换成向量 $\boldsymbol{b}$。

要求解 $n \times n$ 系统，克拉默规则计算 $n + 1$ 个行列式(A 以及 n 个不同的 B's)。当每个行列式都是 $n!$项的总和——使用所有排列的大公式——得到总数$(n + 1)!$项，如果这样求解方程式会发疯的，但是我们最终确实得到**求解 $\boldsymbol{x}$ 的明确公式**。

范例 2 克拉默规则对数字没有效率，但是很适合字母。当 $n = 2$，求解 $AA^{-1} = I$ 得到 $A^{-1} = [\boldsymbol{x} \quad \boldsymbol{y}]$的列：

$$\boldsymbol{A^{-1}}\textbf{ 的列是 }\boldsymbol{x}\textbf{ 与 }\boldsymbol{y} \qquad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}$$

他们共享相同的矩阵 A。我们需要$|A|$以及 4 个针对 x_1, x_2, y_1, y_2 的行列式：

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ 与 } \begin{vmatrix} \mathbf{1} & b \\ \mathbf{0} & d \end{vmatrix} \quad \begin{vmatrix} a & \mathbf{1} \\ c & \mathbf{0} \end{vmatrix} \quad \begin{vmatrix} \mathbf{0} & b \\ \mathbf{1} & d \end{vmatrix} \quad \begin{vmatrix} a & \mathbf{0} \\ c & \mathbf{1} \end{vmatrix}$$

最后 4 个行列式是 $d, -c, -b$ 与 a。(他们都是余子式!) 下列是 A^{-1}：

$$x_1 = \frac{d}{|A|},\ x_2 = \frac{-c}{|A|},\ y_1 = \frac{-b}{|A|},\ y_2 = \frac{a}{|A|} \text{ 以及} A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

我选择 2×2 矩阵所以重点会很清楚，新的概念是：A^{-1} 与余子式有关。当右侧是单位矩阵的一个列，如同 $AA^{-1} = I$ 一样，**克拉默规则中的每个 $\boldsymbol{B_j}$ 的行列式就是 $\boldsymbol{A}$ 的一个余子式**。

当 $n = 3$，你可以看到这些余子式。求解 $A\boldsymbol{x} = (1, 0, 0)$得到 A^{-1} 的列 1：

$$\begin{matrix} B\text{'s 的行列式} \\ = A\text{的余因子} \end{matrix} \qquad \begin{vmatrix} \mathbf{1} & a_{12} & a_{13} \\ \mathbf{0} & a_{22} & a_{23} \\ \mathbf{0} & a_{32} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & \mathbf{1} & a_{13} \\ a_{21} & \mathbf{0} & a_{23} \\ a_{31} & \mathbf{0} & a_{33} \end{vmatrix} \quad \begin{vmatrix} a_{11} & a_{12} & \mathbf{1} \\ a_{21} & a_{22} & \mathbf{0} \\ a_{31} & a_{32} & \mathbf{0} \end{vmatrix} \tag{5}$$

第一个行列式$|B_1|$是余子式 $C_{11} = a_{22}\,a_{33} - a_{23}\,a_{32}$，$|B_2|$是余子式 C_{12}。注意正确的负号出现在$-(a_{21}\,a_{33} - a_{23}\,a_{31})$。这个余子式 C_{12} 进入 A^{-1} 的列 1，当我们除以 $\det A$，我们得到了逆矩阵。

A^{-1} 的 i, j 单元是余子式 $\boldsymbol{C_{ji}}$ **(不是** $\boldsymbol{C_{ij}}$**)**除以 $\det A$：

A^{-1} 的公式

$$(A^{-1})_{ij} = \frac{C_{ji}}{\det A} \quad \text{与} \quad A^{-1} = \frac{C^{\mathrm{T}}}{\det A} \tag{6}$$

余子式 C_{ij} 进入“**余子式矩阵**”C，C 的转置得到 A^{-1}。要计算 A^{-1} 的 i, j 单元，移除 A 的**行 j 与列 i**，行列式乘上$(-1)^{i+j}$得到余子式 C_{ji} 然后除以 $\det A$。

对 A^{-1} 的 3, 1 单元检验这项规则。对于列 1 我们求解 $A\boldsymbol{x} = (1, 0, 0)$，第三个分量 x_3 需要方程式(5)的第三个行列式除以 $\det A$。这个行列式确实就是余子式 $C_{13} = a_{21}\,a_{32} - a_{22}\,a_{31}$，所以$(A^{-1})_{31} = C_{13} \,/\, \det A$。

概要 求解 $A^{-1}A = I$ 时，I 的每个列得到 A^{-1} 的一个列，A^{-1} 的每个单元是一个比值：大小 $n-1$ 的行列式 / 大小 n 的行列式。

公式 $\boldsymbol{A^{-1} = C^{\mathrm{T}} \,/\, \det A}$ 的直接证明 这个表示 $\boldsymbol{AC^{\mathrm{T}} = (\det A)I}$ ：

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix} \tag{7}$$

(A 的行 1)乘(C^{T} 的列 1)得到右侧的第一个 $\det A$。

$$a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = \det A \qquad \text{这个确切是余子式规则！}$$

同理 A 的行 2 乘 C^{T} 的列 2 (注意转置)也得到 $\det A$，单元 a_{2j} 应该乘上余子式 C_{2j} 得到行列式。

如何解释方程式(7)的主对角线之外的零？A 的行乘上来自不同行的余子式，为什么答案是零?

A 的行 2，C 的行 1 $\qquad a_{21}C_{11} + a_{22}C_{12} + a_{23}C_{13} = 0 \tag{8}$

答案：当 A 的第二行复制到自己的第一行，这是新矩阵的余子式规则。新矩阵 A^*有两个相等的行，所以方程式(8)的 $\det A^* = 0$。注意 A^*与 A 有相同的余子式 C_{11}, C_{12}, C_{13}——因为在第一行之后所有的行都一致，于是重要的乘法(7)是正确的：

$$AC^{\mathrm{T}} = (\det A)I \quad \text{或} \quad A^{-1} = \frac{C^{\mathrm{T}}}{\det A}$$

请尊重版权与译者的劳动成果，侵权必究!

范例 3 “总和矩阵”A 的行列式是 1，A^{-1} 只包含余子式：

$$A=\begin{bmatrix}1&0&0&0\\1&1&0&0\\1&1&1&0\\1&1&1&1\end{bmatrix}\quad\text{的逆矩阵}\quad A^{-1}=\frac{C^{\mathrm{T}}}{1}=\begin{bmatrix}1&0&0&0\\-1&1&0&0\\1&-1&1&0\\1&1&-1&1\end{bmatrix}$$

去除 A 的行 1 与列 1 得到 3×3 的余子式 $C_{11}=1$，现在去除行 1 与列 2 得到 C_{12}，这个 3×3 子矩阵仍然是三角形且行列式是 1，但是余子式 C_{12} 是–1，因为符号$(-1)^{1+2}$，这个数字–1 进入 A^{-1} 的(2, 1)单元——不要忘记转置 C。

三角矩阵的逆矩阵是三角形，余子式给出了理由。

范例 4 若所有的余子式都不是零，A 一定可逆吗？不可能。

三角形的面积

每个人都知道矩形的面积——底乘高，三角形的面积是**一半的底乘高**。但是这里有一个那些公式无法回答的问题，若我们已知三角形的三个角(x_1, y_1)与(x_2, y_2)与(x_3, y_3)，面积是多少？利用这些角求出底与高**不是**计算面积的好方法。

行列式是求面积的最佳方法，三角形的面积是一个 3×3 行列式的一半，底与高的平方根在优质公式中会抵消。如果一个角在原点，比如$(x_3, y_3)=(0, 0)$，行列式只有 2×2。

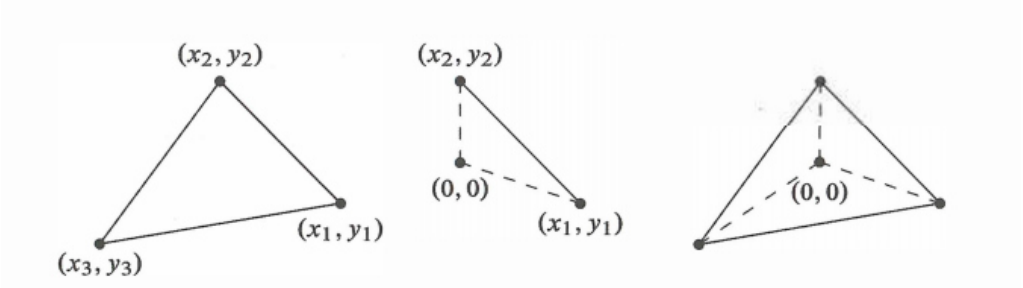


图 5.1：一般三角形；来自(0, 0)的特殊三角形；来自三个特定点的一般形

三角形的三个角在(x_1, y_1)与(x_2, y_2)与(x_3, y_3)，**面积=行列式 / 2**：

三角形面积 $\dfrac{1}{2}\begin{vmatrix}x_1&y_1&1\\x_2&y_2&1\\x_3&y_3&1\end{vmatrix}$

当$(x_3, y_3)=(0, 0)$，面积 $=\dfrac{1}{2}\begin{vmatrix}x_1&y_1\\x_2&y_2\end{vmatrix}$

当你在 3×3 行列式中令 $x_3 = y_3 = 0$，你得到 2×2 行列式，这些公式没有平方根——他们很容易记忆。3×3 行列式分成 3 个 2×2(余子式)的总和，如同图 5.1 所示的第三个三角形，分成 3 个来自(0, 0)的三角形：

$$\text{面积} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{matrix} +\frac{1}{2}(x_1y_2 - x_2y_1) \\ +\frac{1}{2}(x_2y_3 - x_3y_2) \\ +\frac{1}{2}(x_3y_1 - x_1y_3) \end{matrix} \tag{9}$$

如果(0, 0)在三角形之外，其中两个特殊面积可能是负的——但是总和仍然是正确的。实际的问题是说明来自(0, 0)的三角形面积。

为什么$|x_1y_2 - x_2y_1|/2$ 是三角形的面积？我们移除 1/2 的因子当成平行四边形(两倍大，因为平行四边形包含两个相等的三角形)，我们现在证明平行四边形面积是行列式 $x_1y_2 - x_2y_1$，在图 5.2 的面积是 11，因此三角形面积是 11/2。

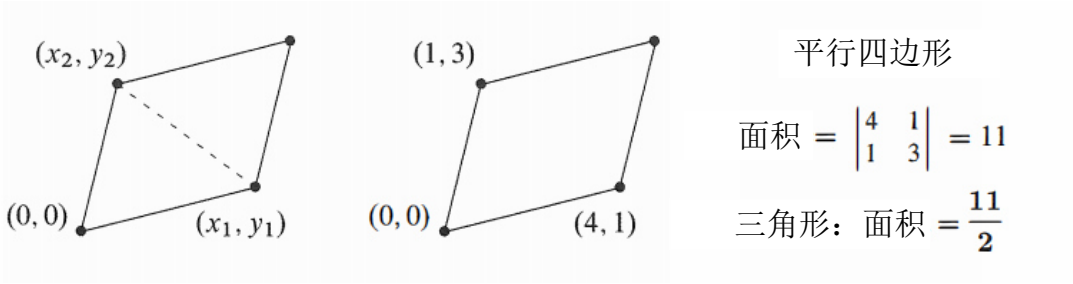


图 5.2：三角形是平行四边形的一半，面积是行列式的一半。

证明：从(0, 0)开始的平行四边形的面积 = 2×2 行列式

有很多种证明的方法，但是这一个最适合本书，我们证明面积与行列式一样的性质 1-2-3，则面积=行列式！记得这 3 个定义行列式的规则然后得到全部其他性质。

1　当 $A = I$，平行四边形变成单位方形，面积是 $\det I = 1$。

2　当出现行的交换，行列式反转符号，绝对值(正的面积)保持不变——相同的平行四边形。

3　若行 1 乘 t，图 5.3a 显示面积也乘 t。假设一个新行(x'_1, y'_1)加到(x, y) (行 2 保持固定)，图 5.3b 显示实线平行四边形面积加到点线平行四边形面积(因为这两个点线完成的三角形相同)。

当我们应用了平面几何，这是一个奇异的证明，但是这个证明有一个主要的吸引力——它适用在 n 维。原点出发的 n 个边是 $n \times n$ 矩阵的行，盒子会有更多的边，就好像平行四边形一样。

请尊重版权与译者的劳动成果，侵权必究！

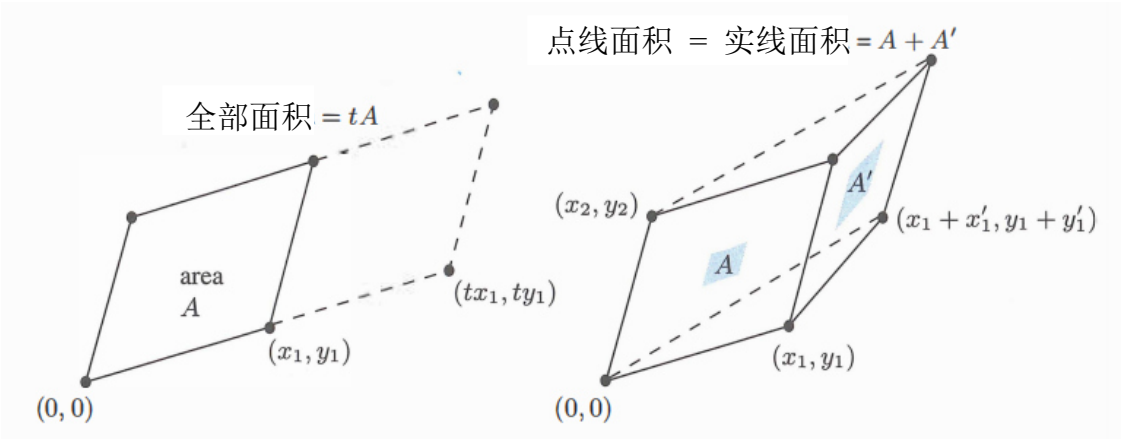


图 5.3：面积在边 1 遵守线性规则(保持边(x_2, y_2)不变)。

图 5.4 显示一个三维盒子——边都不是直角，体积等于 det A 的绝对值。我们的证明再度检查行列式的规则 1-3 是否也被体积遵守。当一个边被拉伸一个因子 t，体积也乘 t。当边 1 加到边 1′，体积是两个原始体积的总和。这是图 5.3b 提升到 3 维或是 n 维，我想画出这些图，但是纸张只有二维。

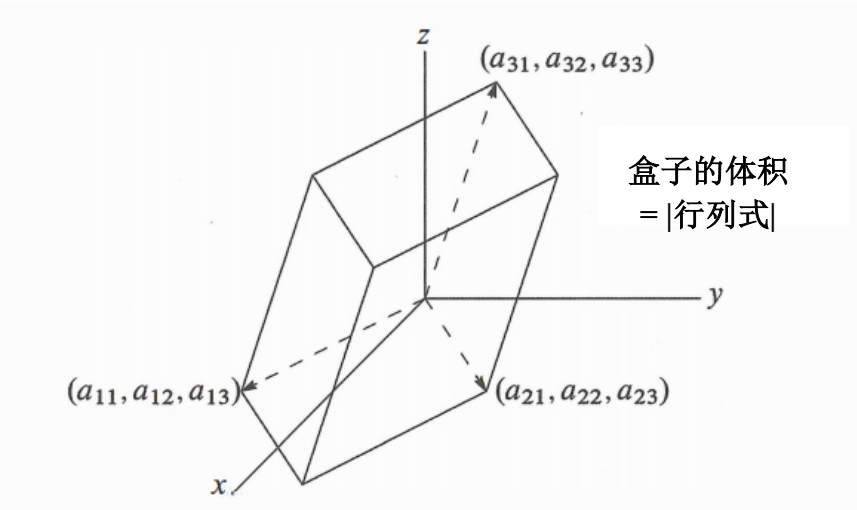


图 5.4：来自 A 的三个行形成的三维盒子。

单位盒子的体积是 1，就是 det I。行交换或是边交换得到相同的盒子以及相同的绝对值体积。行列式改变符号会指出边是“右手三重(right-handed triple)”(det $A > 0$)或是“左手三重(left-handed triple)”(det $A < 0$)。盒子的体积遵守行列式的规则，所以 det A 的体积 = 绝对值。

范例 5 假设矩形盒子(90°角)的边长是 r, s 与 t，它的体积是 r 乘 s 乘 t。对角矩阵 A 的对角线单元 r, s 与 t 产生这三个边，则 det A 也等于体积 $r\,s\,t$。

范例 6 微积分中，盒子当做超级无限小！要针对一个圆积分，我们要把 x 与 y 变成 r 与 θ，这些是极坐标转换：$x = r\cos\theta$，$y = r\sin\theta$。“极盒(polar box)”的面积是 J 的行列式乘 $dr\,d\theta$：

微积分中的面积：$r\,dr\,d\theta$
$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

在小面积 $dA = r\,dr\,d\theta$ 的行列式是 r，拉伸因子 J 进入双重积分，就好像 dx/du 进入正常积分 $\int dx = \int (dx/du)\,du$。在三重积分中，有 9 个导数的雅克比矩阵是 3×3。

叉积

叉积(cross product)是额外(可选择的)应用，特别是对于三维。从向量 $\boldsymbol{u} = (u_1, u_2, u_3)$ 与 $\boldsymbol{v} = (v_1, v_2, v_3)$开始，不像点积是一个数字，**叉积是一个向量**——也是在三维，写成 $\boldsymbol{u} \times \boldsymbol{v}$，读作 $\boldsymbol{u}$ cross $\boldsymbol{v}$。叉积的分量是 2×2 的余子式，我们会解释一些性质，这些性质使得 $\boldsymbol{u} \times \boldsymbol{v}$ 在几何学与物理学非常有用。

现在我们硬着头皮在性质之前写下公式。

定义 $\boldsymbol{u} = (u_1, u_2, u_3)$与 $\boldsymbol{v} = (v_1, v_2, v_3)$的叉积是一个向量：

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2v_3 - u_3v_2)\boldsymbol{i} + (u_3v_1 - u_1v_3)\boldsymbol{j} + (u_1v_2 - u_2v_1)\boldsymbol{k} \tag{10}$$

向量 $\boldsymbol{u} \times \boldsymbol{v}$ 与 $\boldsymbol{u}$ 及 $\boldsymbol{v}$ 垂直，叉积 $\boldsymbol{v} \times \boldsymbol{u} = -(\boldsymbol{u} \times \boldsymbol{v})$。

注解 3×3 行列式是记忆 $\boldsymbol{u} \times \boldsymbol{v}$ 最容易的方法，它不是特别合法，因为第一行包含向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$，而其他行包含数字。在行列式中，向量 $\boldsymbol{i} = (1, 0, 0)$乘 u_2v_3 与$-u_3v_2$，结果是$(u_2v_3 - u_3v_2, 0, 0)$，这个就是叉积的第一分量。

注意下标的循环模式：2 与 3 得到 $\boldsymbol{u} \times \boldsymbol{v}$ 的分量 1，然后 3 与 1 得到分量 2，再然后 1 与 2 得到分量 3，这样完成了叉积的定义。现在我们列表叉积的性质：

性质 1 $\boldsymbol{v} \times \boldsymbol{u}$ 反转行列式的行 2 与 3，所以等于$-(\boldsymbol{u} \times \boldsymbol{v})$。

性质 2 叉积 $\boldsymbol{u} \times \boldsymbol{v}$ 与 $\boldsymbol{u}$ 垂直(与 $\boldsymbol{v}$ 也垂直)，直接的证明是观察**项的消除**，得到零点积：

$$\boldsymbol{u} \cdot (\boldsymbol{u} \times \boldsymbol{v}) = u_1(u_2v_3 - u_3v_2) + u_2(u_3v_1 - u_1v_3) + u_3(u_1v_2 - u_2v_1) = 0 \tag{11}$$

$\boldsymbol{u} \cdot (\boldsymbol{u} \times \boldsymbol{v})$的行列式有三个行：$\boldsymbol{u}, \boldsymbol{u}$ 与 $\boldsymbol{v}$ (两个相等的行)，所以是零。

请尊重版权与译者的劳动成果，侵权必究!

性质 3　任意向量与自己的叉积(两个相等的向量)是 $\boldsymbol{u}\times\boldsymbol{u}=\boldsymbol{0}$。

当 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 平行，叉积是零。当 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 垂直，点积是零。一个与 $\sin\theta$ 有关，另一个与 $\cos\theta$ 有关：

$$\|\boldsymbol{u}\times\boldsymbol{v}\|=\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\sin\theta| \quad 与 \quad |\boldsymbol{u}\cdot\boldsymbol{v}|=\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\cos\theta| \tag{12}$$

范例 7　$\boldsymbol{u}=(3,2,0)$与 $\boldsymbol{v}=(1,4,0)$在 xy 平面，$\boldsymbol{u}\times\boldsymbol{v}$ 会在 z 轴上：

$$\boldsymbol{u}\times\boldsymbol{v}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\3&2&0\\1&4&0\end{vmatrix}=10\boldsymbol{k}$$

叉积是 $\boldsymbol{u}\times\boldsymbol{v}=(0,0,10)$。

$\boldsymbol{u}\times\boldsymbol{v}$ 的长度等于两边是 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的平行四边形面积，这个很重要：本例面积是 10。

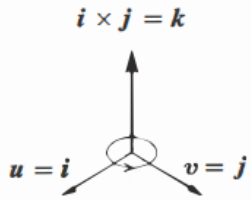
范例 8　$\boldsymbol{u}=(1,1,1)$与 $\boldsymbol{v}=(1,1,2)$的叉积是$(1,-1,0)$：

$$\boldsymbol{u}\times\boldsymbol{v}=\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\1&1&1\\1&1&2\end{vmatrix}=\boldsymbol{i}\begin{vmatrix}1&1\\1&2\end{vmatrix}-\boldsymbol{j}\begin{vmatrix}1&1\\1&2\end{vmatrix}+\boldsymbol{k}\begin{vmatrix}1&1\\1&1\end{vmatrix}=\boldsymbol{i}-\boldsymbol{j}$$

向量$(1,-1,0)$与$(1,1,1)$及$(1,1,2)$如预测一般的垂直。面积是$\sqrt{2}$。

范例 9　$\boldsymbol{i}=(1,0,0)$与 $\boldsymbol{j}=(0,1,0)$的叉积遵守“右手规则”，叉积 $\boldsymbol{k}=\boldsymbol{i}\times\boldsymbol{j}$ 往上不是往下：

$$\begin{vmatrix}\boldsymbol{i}&\boldsymbol{j}&\boldsymbol{k}\\1&0&0\\0&1&0\end{vmatrix}=\boldsymbol{k}$$



规则　当手指头从 $\boldsymbol{u}$ 到 $\boldsymbol{v}$ 卷曲，$\boldsymbol{u}\times\boldsymbol{v}$ 会指向你的右拇指方向。

于是 $\boldsymbol{i}\times\boldsymbol{j}=\boldsymbol{k}$。右手规则也会得到 $\boldsymbol{j}\times\boldsymbol{k}=\boldsymbol{i}$ 与 $\boldsymbol{k}\times\boldsymbol{i}=\boldsymbol{j}$，注意循环的顺序。在相反顺序(反循环)的情形下，拇指反向且叉积走到另一个方向：$\boldsymbol{k}\times\boldsymbol{j}=-\boldsymbol{i}$ 与 $\boldsymbol{i}\times\boldsymbol{k}=-\boldsymbol{j}$ 与 $\boldsymbol{j}\times\boldsymbol{i}=-\boldsymbol{k}$。你可以从 3×3 行列式看到 3 个正号与 3 个负号。

$\boldsymbol{u}\times\boldsymbol{v}$ 可以基于向量定义，而不是他们的分量：

定义　叉积是长度$\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\,|\sin\theta|$的向量，它的方向与 $\boldsymbol{u}$ 及 $\boldsymbol{v}$ 垂直。依据右手规则决定指上或指下。

这个定义很吸引物理学家，他们讨厌轴与坐标。他们看(u_1,u_2,u_3)就是质量的位置，看(F_x,F_y,F_z)就是作用在质量的力。

如果 $\boldsymbol{F}$ 与 $\boldsymbol{u}$ 平行，则 $\boldsymbol{u} \times \boldsymbol{F} = \boldsymbol{0}$——没有转动。叉积 $\boldsymbol{u} \times \boldsymbol{F}$ 是转动力或是转矩(torque)，它的指向是沿着转动轴(与 $\boldsymbol{u}$ 及 $\boldsymbol{F}$ 垂直)，它的长度 $\| \boldsymbol{u} \| \| \boldsymbol{F} \| \sin\theta$ 是产生转动的“矩”的量测值。

三重积 = 行列式 = 体积

由于 $\boldsymbol{u} \times \boldsymbol{v}$ 是一个向量，可以把它与第三向量 $\boldsymbol{w}$ 做点积，得到三重积 $(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$。它被称为**“纯量”三重积**(triple product)，因为它是一个数字。事实上它是一个行列式——它给出 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 盒子的体积：

$$\textbf{三重积} \qquad (\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \tag{13}$$

我们可以把 $\boldsymbol{w}$ 放在顶端或是底部，这两个行列式相等，这是因为从一个到另一个_____的行交换。注意当行列式为零时：

$(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w} = 0$　当 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 确实在同一个平面上。

第一个理由　$\boldsymbol{u} \times \boldsymbol{v}$ 与那个平面垂直，所以它与 $\boldsymbol{w}$ 的点积为零。

第二个理由　平面上的三个向量是相关，矩阵是奇异($\det = 0$)。

第三个理由　当 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 的盒子展开成平面，体积是零。

$(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w}$ 等于 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 作为边的盒子的体积，这个性质非常重要，这个 3×3 行列式承载了大量信息。比如 2×2 矩阵的 $ad - bc$，它分离了可逆与奇异。第六章会探索奇异的性质。

主要观念的复习

1. 克拉默规则利用比值求解 $A\boldsymbol{x} = \boldsymbol{b}$，就像 $x_1 = |B_1| / |A| = |\boldsymbol{b}\ \boldsymbol{a}_2 \ldots \boldsymbol{a}_n| / |A|$。
2. 当 C 是 A 的余子式，逆矩阵 $A^{-1} = C^{\mathrm{T}} / \det A$。
3. 当盒子的边是 A 的行时，盒子的体积是 $|\det A|$。
4. 在双重与三重积分中，面积与体积需要变换变数。
5. 在 $\mathbf{R}^3$ 中，叉积 $\boldsymbol{u} \times \boldsymbol{v}$ 与 $\boldsymbol{u}$ 及 $\boldsymbol{v}$ 垂直。注意 $\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}$。

请尊重版权与译者的劳动成果，侵权必究！

已解范例

5.3A 若 A 是奇异，方程式 $AC^{\mathrm{T}} = (\det A)I$ 变成 $\boldsymbol{AC^{\mathrm{T}}}$ **= 零矩阵**，则 C^{T} 的每一列在 A 的零空间中。这些列包含沿着 $\boldsymbol{A}$ **的行**的余子式，所以余子式快速的求出秩 2 的 3×3 矩阵的的零空间。我很抱歉这个来的太晚！

对于秩 2 的奇异矩阵，利用 $\boldsymbol{x}$ 等于沿着一个行的余子式来求解 $A\boldsymbol{x} = \boldsymbol{0}$：

余子式得到零空间 $\quad A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 3 & 9 \\ 2 & 2 & 8 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

解 第一个矩阵沿着第一行的余子式(注意每个负号)是：

$$\begin{vmatrix} 3 & 9 \\ 2 & 8 \end{vmatrix} = 6 \qquad -\begin{vmatrix} 2 & 9 \\ 2 & 8 \end{vmatrix} = 2 \qquad \begin{vmatrix} 2 & 3 \\ 2 & 2 \end{vmatrix} = -2$$

则 $\boldsymbol{x} = (6, 2, -2)$是 $A\boldsymbol{x} = \boldsymbol{0}$ 的解。沿着第二行的余子式是$(-18, -6, 6)$刚好是$-3\boldsymbol{x}$，这也是 A 的一维零空间。

第二个矩阵沿着第一行都是零余子式，零空间 $\boldsymbol{x} = (0, 0, 0)$不够吸引力。行 2 的余子式得到 $\boldsymbol{x} = (1, -1, 0)$是 $A\boldsymbol{x} = \boldsymbol{0}$ 的解。

从问题 3.3.12 知道每个秩 $n - 1$ 的 $n \times n$ 矩阵，最少有一个非零的余子式，但是对秩 $n - 2$ 来说所有的余子式都是零，我们只能找到 $\boldsymbol{x} = \boldsymbol{0}$。

5.3B 利用克拉默规则的比值 $\det B_1 / \det A$ 求解 $A\boldsymbol{x} = \boldsymbol{b}$，也求出 $A^{-1} = C^{\mathrm{T}} / \det A$。当 $\boldsymbol{b} = (0, 0, 1)$，解 $\boldsymbol{x}$ 是 A^{-1} 的列 3。计算列 $\boldsymbol{x} = (x, y, z)$时，哪些余子式有参与？

$\boldsymbol{A^{-1}}$ 的列 3 $\quad \begin{bmatrix} 2 & 6 & 2 \\ 1 & 4 & 2 \\ 5 & 9 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

求出两个盒子的体积：边是 A 的列与以及边是 A^{-1} 的行。

解 B_j 的行列式(右侧 $\boldsymbol{b}$ 放在列 j)是

$$|B_1| = \begin{bmatrix} \boldsymbol{0} & 6 & 2 \\ \boldsymbol{0} & 4 & 2 \\ \boldsymbol{1} & 9 & 0 \end{bmatrix} = 4 \qquad |B_2| = \begin{bmatrix} 2 & \boldsymbol{0} & 2 \\ 1 & \boldsymbol{0} & 2 \\ 5 & \boldsymbol{1} & 0 \end{bmatrix} = -2 \qquad |B_3| = \begin{bmatrix} 2 & 6 & \boldsymbol{0} \\ 1 & 4 & \boldsymbol{0} \\ 5 & 9 & \boldsymbol{1} \end{bmatrix} = 2$$

这些是行 3 的余子式 C_{31}, C_{32}, C_{33}，他们与行 3 的点积是 $\det A = 2$。

$$\det A = a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} = (5, 9, 0) \cdot (4, -2, 2) = 2$$

三个比值 $\det B_j / \det A$ 得到 $\boldsymbol{x} = (2, -1, 1)$的三个分量，这个 $\boldsymbol{x}$ 是 A^{-1} 的第三列，因为 $\boldsymbol{b} = (0, 0, 1)$是 I 的第三列。

沿着 A 的其他行的余子式，除以 $\det A$，得到 A^{-1} 的其他列：

$$\boldsymbol{A^{-1}} = \frac{C^{\mathrm{T}}}{\det A} = \frac{1}{2}\begin{bmatrix} -18 & 18 & 4 \\ 10 & -10 & -2 \\ -11 & 12 & 2 \end{bmatrix} \qquad \text{计算验证} \quad AA^{-1} = I$$

来自 A 的列的盒子有体积 $= \det A = 2$，由于$|A^{\mathrm{T}}| = |A|$，来自行的盒子也有体积 2。来自 A^{-1} 的行的盒子有体积 $1/|A| = 1/2$。

问题集 5.3

问题 1-5 有关 $\boldsymbol{x = A^{-1}b}$ 的克拉默规则。

1　利用克拉默规则 $x_j = \det B_j / \det A$ 求解：

(a) $\begin{aligned} 2x_1 + 5x_2 &= 1 \\ x_1 + 4x_2 &= 2 \end{aligned}$　　(b) $\begin{aligned} 2x_1 + x_2 &= 1 \\ x_1 + 2x_2 + x_3 &= 0 \\ x_2 + 2x_3 &= 0 \end{aligned}$

2　利用克拉默规则只求 y，称呼 3×3 行列式为 D：

(a) $\begin{aligned} ax + by &= 1 \\ cx + dy &= 0 \end{aligned}$　　(b) $\begin{aligned} ax + by + cz &= 1 \\ dx + ey + fz &= 0 \\ gx + hy + iz &= 0 \end{aligned}$

3　当 $\det A = 0$，克拉默规则崩溃。例题(a)无解，例题(b)无限多解。两个例子中的比值 $x_j = \det B_j / \det A$ 为何？

(a) $\begin{aligned} 2x_1 + 3x_2 &= 1 \\ 4x_1 + 6x_2 &= 1 \end{aligned}$ (平行线)　　(b) $\begin{aligned} 2x_1 + 3x_2 &= 1 \\ 4x_1 + 6x_2 &= 2 \end{aligned}$ (同一条线)

4　克拉默规则的快速证明。行列式是列 1 的线性函数，如果两个列相同，行列式是零。当 $\boldsymbol{b} = A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3$ 进入 A 的第一列，矩阵 B_1 的行列式是

$$|\,\boldsymbol{b} \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3\,| = |\,x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3\,| = x_1|\,\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3\,| = x_1 \det A$$

(a) 左侧等于右侧可以得到 x_1 的什么公式？

(b) 哪个步骤得到中间的方程式？

5　若右侧 $\boldsymbol{b}$ 是 A 的第一列，求解 3×3 系统 $A\boldsymbol{x} = \boldsymbol{b}$。克拉默规则中的每个行列式如何得到解 $\boldsymbol{x}$？

请尊重版权与译者的劳动成果，侵权必究！

问题 6-15 有关 $A^{-1} = C^{\mathrm{T}}/\det A$，记得转置 C。

6 从余子式公式 $C^{\mathrm{T}}/\det A$ 求出 A^{-1} 。利用(b)的对称。

(a) $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 7 & 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

7 如果所有的余子式都是零，你如何知道 A 没有逆矩阵？如果所有的余子式都不是零，A 一定可逆吗?

8 求出 A 的余子式，计算 AC^{T} 求出 $\det A$：

(a) $A = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 5 \end{bmatrix}$ 与 $C = \begin{bmatrix} 6 & -3 & 0 \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{bmatrix}$ 以及 $AC^{\mathrm{T}} =$ _____。

如果你把 4 改成 100，为什么 $\det A$ 不变？

9 假设 $\det A = 1$，你也知道 C 中的所有余子式，如何求 A？

10 从公式 $AC^{\mathrm{T}} = (\det A)I$ 证明 $\det C = (\det A)^{n-1}$。

11 如果 A 的全部单元都是整数，且 $\det A = 1$ 或–1，证明 A^{-1} 的全部单元都是整数。给出一个全部单元都不是零的 2×2 范例。

12 如果 A 与 A^{-1} 的全部单元都是整数，证明 $\det A = 1$ 或–1。
提示：$\det A$ 乘 $\det A^{-1}$ 等于什么？

13 利用余子式完成范例 5 中 A^{-1} 的计算。

14 L 是下三角，S 是对称，假设都是可逆：

逆反
三角矩阵 L $L = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$ $S = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}$
对称矩阵 S

(a) L 的哪三个余子式是零？L^{-1} 也是下三角矩阵。

(b) S 的哪三对(pair)余子式相等？S^{-1} 也是对称矩阵。

(c) 正交 Q 的余子式矩阵 C 会是_____。为什么？

15 当 $n = 5$，矩阵 C 有___个余子式。每个 4×4 余子式有___项，每项有___次乘法。段落 2.4 中计算 A^{-1} 的高斯-乔丹法需要 $5^3 = 125$，比较两者的差异。

问题 16-26 有关利用行列式求面积与体积。

16 (a) 平行四边形的边是 $\boldsymbol{v} = (3, 2)$与 $\boldsymbol{w} = (1, 4)$，求面积。

(b) 三角形的边是 $\boldsymbol{v}, \boldsymbol{w}$ 与 $\boldsymbol{v} + \boldsymbol{w}$，求面积并画出图形。

(c) 三角形的边是 $\boldsymbol{v}, \boldsymbol{w}$ 与 $\boldsymbol{w} - \boldsymbol{v}$，求面积并画出图形。

网址：linear.neocities.org 邮箱：876152660@qq.com

17 一个盒子的三边分别从(0, 0, 0)到(3, 1, 1)与(1, 3, 1)与(1, 1, 3)，求盒子的体积。利用$\|\boldsymbol{u} \times \boldsymbol{v}\|$也求出每个平行四边形的面积。

18 (a) 三角形的角是(2, 1)与(3, 4)与(0, 5)，求面积。
(b) 在(−1, 0)加入一个角得到倾斜区域(4 个边)，求面积。

19 平行四边形的边是(2, 1)与(2, 3)，与边是(2, 2)与(1, 3)的平行四边形有相同的面积，利用 2×2 行列式求出面积，并且说明为什么两个图形的面积必须相等？(我从图形看不出相等，如果你可以，麻烦写信告诉我。)

20 哈达玛矩阵 H 有正交的行，盒子是超立方体(hypercube)！

$$|H| = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{vmatrix} = ?$$ 这是 $\mathbf{R}^4$ 的一个超立方体的体积。

21 如果 4×4 矩阵的列有长度 L_1, L_2, L_3, L_4，则最大可能的行列式(基于体积)值是多少？如果矩阵的单元都是 1 或−1，这些列的长度是多少？最大行列式是多少？

22 利用图形证明：面积 x_1y_2 的矩形减去面积 x_2y_1 的矩形，会得到相同面积的平行四边形。

23 当边向量 $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ 互相垂直，盒子的体积是$\|\boldsymbol{a}\|$乘$\|\boldsymbol{b}\|$乘$\|\boldsymbol{c}\|$。矩阵 $A^{\mathrm{T}}A$ 是_____，求 $\det A^{\mathrm{T}}A$ 与 $\det A$。

24 盒子的边分别是 $\boldsymbol{i}$ 与 $\boldsymbol{j}$ 与 $\boldsymbol{w} = 2\boldsymbol{i} + 3\boldsymbol{j} + 4\boldsymbol{k}$，盒子的高度是_____。体积是多少？具有此行列式的矩阵为何？$\boldsymbol{i} \times \boldsymbol{j}$ 为何？？$\boldsymbol{i} \times \boldsymbol{j}$ 与 $\boldsymbol{w}$ 的点积为何？？

25 n 维立方体有几个角？几个边？几个$(n-1)$-维的面？$\mathbf{R}^n$ 的一个立方体的边是 $2I$ 的行，立方体的体积是_____。超立方体电脑在角有平行处理器，有边连接这些角。

26 三角形的角在(0, 0), (1, 0), (0, 1)，面积是 1/2。$\mathbf{R}^3$ 的锥体有四个角(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)，体积是多少？$\mathbf{R}^4$ 的一个锥体，它的 5 个角在(0, 0, 0, 0)以及 I 的行，体积是多少？

问题 27-30 有关微积分的面积 *dA* 与体积 *dV*。

27 极坐标满足 $x = r\cos\theta$ 与 $y = r\sin\theta$，极面积是 $J\,dr\,d\theta$：

$$J = \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}$$

两个列是正交，他们的长度是_____，于是 J = _____。

请尊重版权与译者的劳动成果，侵权必究！

28 球坐标ρ, ϕ, θ 满足 $x = \rho\sin\phi\cos\theta$ 与 $y = \rho\sin\phi\sin\theta$ 与 $z = \rho\cos\phi$。求出 3×3 部分导数矩阵：在行 1 有 $\partial x/\partial\rho$，$\partial x/\partial\phi$，$\partial x/\partial\theta$。简化行列式得到 $J = \rho^2\sin\phi$，则球坐标的 dV 是 $\rho^2\sin\phi\, d\rho\, d\phi\, d\theta$，这是一个“无限小坐标盒”的体积。

29 问题 27 展示连接 r, θ 到 x, y 的矩阵，求 2×2 矩阵的逆矩阵：

$$J^{-1} = \begin{vmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\theta/\partial x & \partial\theta/\partial y \end{vmatrix} = \begin{vmatrix} \cos\theta & ? \\ ? & ? \end{vmatrix} = ?$$

很讶异的 $\partial r/\partial x = \partial x/\partial r$ (微积分，Gilbert Strang，p. 501)， J 与 J^{-1} 的矩阵乘积得到链条规则 $\dfrac{\partial x}{\partial x} = \dfrac{\partial x}{\partial r}\dfrac{\partial r}{\partial x} + \dfrac{\partial x}{\partial\theta}\dfrac{\partial\theta}{\partial x} = 1$。

30 三角形的角在(0, 0)与(6, 0)与(1, 4)，面积是_____。当你把三角形旋转 $\theta = 60°$，面积是_____。旋转矩阵的行列式是

$$J = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \begin{vmatrix} 1/2 & ? \\ ? & ? \end{vmatrix} = ?$$

问题 31-38 有关在三维的三重积 $(\boldsymbol{u}\times\boldsymbol{v})\cdot\boldsymbol{w}$。

31 盒子的底面积是 $\|\boldsymbol{u}\times\boldsymbol{v}\|$，它的垂直高度是 $\|\boldsymbol{w}\|\cos\theta$，盒子的**底面积乘高**=体积= $\|\boldsymbol{u}\times\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta$，这就是 $(\boldsymbol{u}\times\boldsymbol{v})\cdot\boldsymbol{w}$。当 $\boldsymbol{u} = (2, 4, 0)$, $\boldsymbol{v} = (-1, 3, 0)$, $\boldsymbol{w} = (1, 2, 2)$，计算底面积，高与体积。

32 同样盒子的体积可以更直接从 3×3 行列式求得，计算这个行列式。

33 利用 u_1, u_2, u_3 的行的余子式，扩展方程式(13)的 3×3 行列式，这个扩展是 $\boldsymbol{u}$ 与向量_____的点积。

34 $(\boldsymbol{u}\times\boldsymbol{w})\cdot\boldsymbol{v}$ 与 $(\boldsymbol{w}\times\boldsymbol{u})\cdot\boldsymbol{v}$ 与 $(\boldsymbol{v}\times\boldsymbol{w})\cdot\boldsymbol{u}$ 中，哪个与 $(\boldsymbol{u}\times\boldsymbol{v})\cdot\boldsymbol{w}$ 相同？什么样的行 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ 顺序，可以得到正确的行列式？

35 令 $P = (1, 0, -1)$与 $Q = (1, 1, 1)$与 $R = (2, 2, 1)$，选择 S 使得 $PQRS$ 是一个平行四边形，计算它的面积。选择 T, U, V 使得 $OPQRSTUV$ 是一个倾斜盒，计算它的体积。

36 假设(x, y, z)与(1, 1, 0)与(1, 2, 1)落在一个通过原点的平面，什么样的行列式是零？什么样的方程式可以得到这个平面？

37 假设(x, y, z)是(2, 3, 1)与(1, 2, 3)的一个线性组合，什么样的行列式是零？什么样的方程式可以产生所有组合形成的平面？

38 (a) 对于 $n\times n$ 矩阵，借由体积解释为什么 $\det 2A = 2^n \det A$？

(b) 什么样大小的矩阵会使得错的叙述 $\det A + \det A = \det(A + A)$变成正确？

39 如果你知道 4×4 可逆矩阵 A 全部 16 项的余子式，你如何求出 A?

40 假设 A 是 5×5 矩阵，它的行 1 的单元乘来自行 2-5 的行列式(余子式)得到行列式。“雅克比公式”把来自行 1 与 2 的 2×2 行列式，乘上来自行 3-5 的 3×3 行列式得到 $\det A$。你能猜想“雅克比公式”吗？

41 2×2 矩阵 AB = (2×3) (3×2)有一个求取 $\det AB$ 的“Cauchy-Binet 公式”：

$$\det AB = (A \text{ 的 } 2\times2 \text{ 行列式})\text{乘}(B \text{ 的 } 2\times2 \text{ 行列式})\text{的总和}$$

(a) 猜想会用到哪些来自 A 与 B 的 2×2 行列式。

(b) 当 A 的行是[1, 2, 3]与[1, 4, 7]且 $B = A^{\mathrm{T}}$，测试你的公式。

42 大公式有 $n!$ 项，但是如果 A 有一个单元是零，$(n-1)!$ 项会消失。如果 A 只有 3 条对角线单元，会留下多少项？

当 n = 1, 2, 3, 4 时，**三对角**行列式有 1, 2, 3, 5 项，这些是段落 6.2 的斐波那契数字！证明为什么**三对角** 5×5 行列式有 5 + 3 = 8 个非零项(再次是斐波那契数字)，利用 a_{11} 与 a_{12} 的余子式。

请尊重版权与译者的劳动成果，侵权必究!

第六章
特征值与特征向量

6.1 特征值引论

1 **特征向量** $\boldsymbol{x}$ 与 $A\boldsymbol{x}$ 在同一条直线上：$A\boldsymbol{x} = \lambda\boldsymbol{x}$，$\lambda$是**特征值**。

2 若 $A\boldsymbol{x} = \lambda\boldsymbol{x}$，则 $A^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$ 与 $A^{-1}\boldsymbol{x} = \lambda^{-1}\boldsymbol{x}$ 与$(A + cI)\boldsymbol{x} = (\lambda + c)\boldsymbol{x}$：相同 $\boldsymbol{x}$。

3 若 $A\boldsymbol{x} = \lambda\boldsymbol{x}$，则$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ 且 $A - \lambda I$ 奇异，$\det(A - \lambda I) = 0$。n 个特征值。

4 检验λ's：$\det A = (\lambda_1)(\lambda_2) \ldots (\lambda_n)$与对角总和 $a_{11} + a_{22} + \ldots + a_{nn} = \lambda$'s 总和。

5 投影有$\lambda = 1$ 与 0，镜射有 1 与–1，旋转有$e^{i\theta}$与 $e^{-i\theta}$：复数。

本章加入线性代数的新内容，第一部分是有关 $A\boldsymbol{x} = \boldsymbol{b}$：平衡与均衡与稳定状态。现在第二部分是有关**改变**，时间进入蓝图––连续时间的微分方程式 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 或是离散方程式的 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$，这些方程式无法由消元求解。

关键概念在于避免矩阵 A 带来复杂性，假设解向量 $\boldsymbol{u}(t)$保持在固定向量 $\boldsymbol{x}$ 的方向，我们只需要找出乘 $\boldsymbol{x}$ 的数字(随时间变化)，一个数字比一个向量容易。**我们希望特征向量**(eigenvector)$\boldsymbol{x}$ **在被** $\boldsymbol{A}$ **乘时不会改变方向**。

一个好的模式来自矩阵的次方 A, A^2, A^3, ...，假设你需要 100 次方 A^{100}，它的列会非常接近特征向量(.6, .4):

$$A, A^2, A^3 = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .70 & .45 \\ .30 & .55 \end{bmatrix} \begin{bmatrix} .650 & .525 \\ .350 & .475 \end{bmatrix} \qquad A^{100} \approx \begin{bmatrix} .6000 & .6000 \\ .4000 & .4000 \end{bmatrix}$$

A^{100} 是利用 A 的**特征值**(eigenvalue)求得，不是矩阵乘 100 次。这些特征值(此处是 $\lambda = 1$ 与 1/2)是看见矩阵核心的新方法。

要说明特征值，我们首先说明特征向量。大部分的向量被 A 乘之后都会改变方向，某些例外向量 $\boldsymbol{x}$ 与 $A\boldsymbol{x}$ 在同一个方向，他们是**特征向量**。A 乘一个特征向量，得到向量 $A\boldsymbol{x}$ 等于一个数λ乘原始的 $\boldsymbol{x}$。

基本方程式是 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，数字$\lambda$是 A 的一个特征值。

特征值λ会告诉我们这个特殊向量$\boldsymbol{x}$是被拉伸或收缩或反向或保持不变--当$\boldsymbol{x}$被A乘。我们可以发现$\lambda=2$ 或 1/2 或–1 或 1，特征值λ也可以是零！于是 $A\boldsymbol{x}=0\boldsymbol{x}$ 表示特征向量 $\boldsymbol{x}$ 在零空间中。

若 A 是单位矩阵，每个向量都有 $A\boldsymbol{x}=\boldsymbol{x}$，所有的向量都是 I 的特征向量。所有的特征值是$\lambda=1$，这是不寻常的。大部分 2×2 矩阵有两个特征向量的方向以及两个特征值，我们会展示 $\det(A-\lambda I)=0$。

这个段落说明如何计算 $\boldsymbol{x}$'s 与λ's，这个可以来的这么早是因为我们我们只需要 2×2 矩阵的行列式。让我用 $\det(A-\lambda I)=0$ 找出第一个范例的特征值，然后在方程式(3)正确的推导。

范例 1　矩阵 A 有两个特征值$\lambda=1$ 与$\lambda=1/2$，检视 $\det(A-\lambda I)$：

$$A=\begin{bmatrix}.8 & .3\\ .2 & .7\end{bmatrix}\qquad \det\begin{bmatrix}.8-\lambda & .3\\ .2 & .7-\lambda\end{bmatrix}=\lambda^2-\frac{3}{2}\lambda+\frac{1}{2}=(\lambda-1)\left(\lambda-\frac{1}{2}\right)$$

我把二次式分解成$\lambda-1$ 乘$\lambda-1/2$，得到两个特征值是$\lambda=1$ 与$\lambda=1/2$。这些数字使得矩阵 $A-\lambda I$ 变成奇异(零行列式)，特征向量 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 就在 $A-I$ 与 $A-\frac{1}{2}I$ 的零空间中。

$(A-I)\boldsymbol{x}_1=\boldsymbol{0}$ 得到 $A\boldsymbol{x}_1=\boldsymbol{x}_1$ 以及第一个特征向量是**(.6, .4)**。

$(A-I/2)\boldsymbol{x}_2=\boldsymbol{0}$ 得到 $A\boldsymbol{x}_2=(1/2)\,\boldsymbol{x}_2$ 以及第二个特征向量是**(1, –1)**。

$$\boldsymbol{x}_1=\begin{bmatrix}.6\\ .4\end{bmatrix}\quad 与\quad A\boldsymbol{x}_1=\begin{bmatrix}.8 & .3\\ .2 & .7\end{bmatrix}\begin{bmatrix}.6\\ .4\end{bmatrix}=\boldsymbol{x}_1\qquad (A\boldsymbol{x}=\boldsymbol{x}\text{ 表示}\lambda_1=1)$$

$$\boldsymbol{x}_2=\begin{bmatrix}1\\ -1\end{bmatrix}\quad 与\quad A\boldsymbol{x}_2=\begin{bmatrix}.8 & .3\\ .2 & .7\end{bmatrix}\begin{bmatrix}1\\ -1\end{bmatrix}=\begin{bmatrix}.5\\ -.5\end{bmatrix}\qquad (\text{这是 }\boldsymbol{x}_2/2\text{ 所以}\lambda_2=1/2)$$

如果 $\boldsymbol{x}_1$ 再次被 A 乘，我们仍然得到 $\boldsymbol{x}_1$，A 的每个次方都得到 $A^n\boldsymbol{x}_1=\boldsymbol{x}_1$。$\boldsymbol{x}_2$ 被 A 乘得到$\frac{1}{2}\boldsymbol{x}_2$，如果再次被 A 乘我们得到$(\frac{1}{2})^2$ 乘 $\boldsymbol{x}_2$。

当 A 变成平方，特征向量保持不变，特征值变成平方。

因为特征向量维持自己的方向(如图 6.1)而且不会被混淆，模式持续下去，A^{100} 的特征向量是相同的 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$。A^{100} 的特征值是 $1^{100}=1$ 与$(\frac{1}{2})^{100}$=非常小的数字。

其他向量确实会改变方向，但是其他的向量都是这两个特征向量的组合。A 的第一列是 $\boldsymbol{x}_1+(.2)\boldsymbol{x}_2$ 的组合：

分开特征向量，然后被 A 乘
$$\begin{bmatrix}.8\\ .2\end{bmatrix}=\boldsymbol{x}_1+(.2)\,\boldsymbol{x}_2=\begin{bmatrix}.6\\ .4\end{bmatrix}+\begin{bmatrix}.2\\ -.2\end{bmatrix}\qquad (1)$$

请尊重版权与译者的劳动成果，侵权必究!

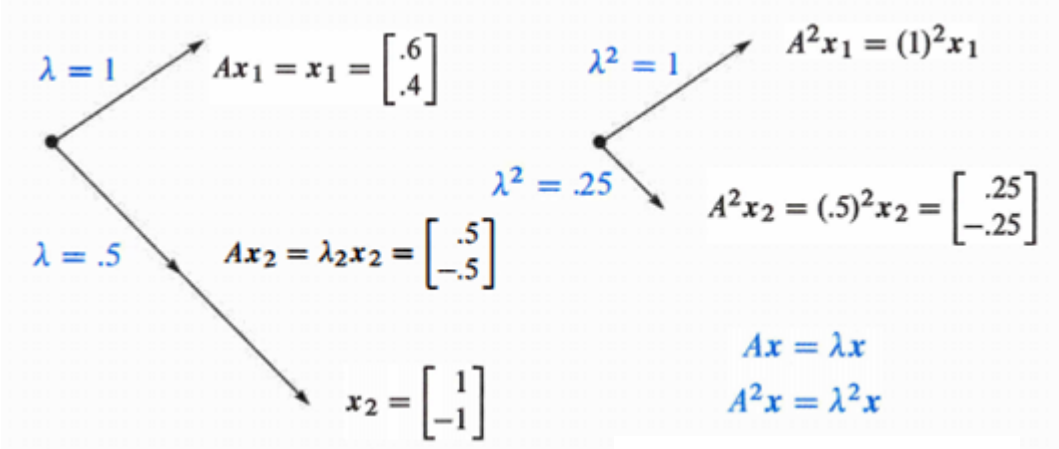


图 6.1：特征向量维持方向。$A^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$，其中$\lambda^2 = 1^2$与$(0.5)^2$。

当我们分开乘 $\boldsymbol{x}_1$ 与$(.2)\boldsymbol{x}_2$，A 乘 $\boldsymbol{x}_2$ 就是它的特征值 $\frac{1}{2}$ 乘 $\boldsymbol{x}_2$：

$$\boldsymbol{\lambda_i}\textbf{乘每个}\ \boldsymbol{x_i} \qquad A\begin{bmatrix} .8 \\ .2 \end{bmatrix} \quad \text{是} \quad \boldsymbol{x}_1 + \frac{1}{2}(.2)\,\boldsymbol{x}_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} .1 \\ -.1 \end{bmatrix} = \begin{bmatrix} .7 \\ .3 \end{bmatrix}$$

当我们左乘 A 时，每个特征向量被自己的特征值乘。在每一个步骤 $\boldsymbol{x}_1$ 都没有改变而 $\boldsymbol{x}_2$ 被 $\frac{1}{2}$ 乘，所以 99 步之后得到很小的数$(\frac{1}{2})^{99}$:

$$A^{99}\begin{bmatrix} .8 \\ .2 \end{bmatrix} \quad \text{就是} \quad \boldsymbol{x}_1 + (.2)(\frac{1}{2})^{99}\boldsymbol{x}_2 = \begin{bmatrix} .6 \\ .4 \end{bmatrix} + \begin{bmatrix} \text{非常小} \\ \text{的向量} \end{bmatrix}$$

这是 A^{100} 的第一列，我们原始写的.6000 不是很精确，我们省略了$(.2)\ (\frac{1}{2})^{99}$，这个数字在小数点 30 位数还不会出现。

特征向量 $\boldsymbol{x}_1$ 是一个不会改变的“稳定状态”(因为$\lambda_1 = 1$)，特征向量 $\boldsymbol{x}_2$ 是一个几乎消失的“衰减模式”(因为$\lambda_2 = .5$)。A 的次方越高，它的列趋近稳定状态。

这个特殊的 A 是一个“**马尔可夫**矩阵”，最大特征值是$\lambda = 1$，它的特征向量 $\boldsymbol{x}_1 = (.6, .4)$是稳定状态--$A^k$ 的所有的列都会趋近它。段落 10.3 说明马尔可夫矩阵如何在谷歌搜寻时出现。

对于投影矩阵，我们可以看到 $P\boldsymbol{x}$ 什么时候平行 $\boldsymbol{x}$。对应$\lambda = 1$ 与$\lambda = 0$ 的特征向量形成列空间与零空间。列空间没有移动($P\boldsymbol{x} = \boldsymbol{x}$)，零空间移到零($P\boldsymbol{x} = 0\boldsymbol{x}$)。

范例 2 **投影矩阵** $P=\begin{bmatrix}.5 & .5\\ .5 & .5\end{bmatrix}$ 有特征值 $\lambda=1$ 与 $\lambda=0$。

它的特征向量是 $\boldsymbol{x}_1=(1,1)$与 $\boldsymbol{x}_2=(1,-1)$。对于这些向量，$P\boldsymbol{x}_1=\boldsymbol{x}_1$(稳定状态)与 $P\boldsymbol{x}_2=\boldsymbol{0}$ (零空间)。这个例子阐述了马尔可夫矩阵与奇异矩阵与(最重要)对称矩阵，这些矩阵都有特殊的λ's 与 $\boldsymbol{x}$'s：

1. 马尔可夫矩阵：P 的每个列总和是 1，所以$\lambda=1$ 是一个是特征值。
2. P 是奇异，所以$\lambda=0$ 是一个是特征值。
3. P 是对称，所以它的特征向量(1, 1)与(1, −1)垂直。

投影矩阵的特征值只有 0 与 1，对应$\lambda=0$ 的特征向量(表示 $P\boldsymbol{x}=0\boldsymbol{x}$)得到零空间，对应$\lambda=1$ 的特征向量(表示 $P\boldsymbol{x}=\boldsymbol{x}$)得到列空间。零空间投影到零，列空间投影到自己，投影维持列空间而摧毁零空间。

每个部分的投影 $\boldsymbol{v}=\begin{bmatrix}1\\-1\end{bmatrix}+\begin{bmatrix}2\\2\end{bmatrix}$ 投影到 $P\boldsymbol{v}=\begin{bmatrix}\boldsymbol{0}\\\boldsymbol{0}\end{bmatrix}+\begin{bmatrix}\boldsymbol{2}\\\boldsymbol{2}\end{bmatrix}$

投影有$\lambda=0$ 与 1，排列的所有$|\lambda|=1$。下个矩阵 R 是镜射同时也是排列，R 也有特殊的特征值。

范例 3 **镜射矩阵** $R=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ 有特征值 1 与−1。

R 不会改变特征向量(1, 1)，第二个特征向量是(1, −1)--R 会反转它的符号。一个没有负数单元的矩阵仍然可以有负数的特征值！R 的特征向量与 P 的特征向量相同，因为**镜射= 2(投影) − $\boldsymbol{I}$** ：

$$\boldsymbol{R=2P-I}\qquad \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}=2\begin{bmatrix}.5 & .5\\ .5 & .5\end{bmatrix}-\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \tag{2}$$

当一个矩阵被 $\boldsymbol{I}$ 平移，每个λ平移 1，特征向量没有改变。

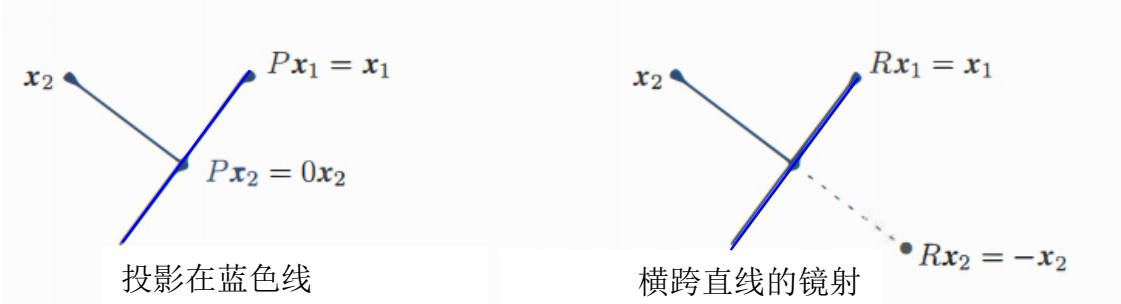


图 6.2：投影 P 有特征值 1 与 0，镜射 R 有特征值$\lambda=1$ 与−1。典型的 $\boldsymbol{x}$ 改变方向，但是特征向量保持沿着同一直线。

请尊重版权与译者的劳动成果，侵权必究!

特征值的方程式

我们对投影矩阵利用几何方法找出λ's 与 $\boldsymbol{x}$'s ：$P\boldsymbol{x}=\boldsymbol{x}$ 与 $P\boldsymbol{x}=\boldsymbol{0}$。对其他矩阵我们使用行列式与线性代数，这是本章的关键计算--几乎所有的应用都是从求解 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 开始。

首先把$\lambda\boldsymbol{x}$ 移到左侧，把方程式 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 写成$(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$，矩阵 $A-\lambda I$ 乘特征向量 $\boldsymbol{x}$ 得到零向量，特征向量构成 $A-\lambda I$ 的零空间。当我们知道一个特征值λ，我们求解$(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$ 得到一个特征向量。

首先是特征值。如果$(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$ 有非零解，$A-\lambda I$ 不可逆，$A-\lambda I$ 的行列式必须等于 0，这是求出特征值λ的方法：

特征值 数字λ是 A 的一个特征值，当且仅当 $A-\lambda I$ 奇异。

特征值方程式 $\det(A-\lambda I)=0$ (3)

“特性多项式”$\det(A-\lambda I)$只跟λ有关，与 $\boldsymbol{x}$ 无关。当 A 是 $n\times n$，方程式(3)的次数是 n，A 有 n 个特征值(可能有重复!)，每个λ得到 $\boldsymbol{x}$：

对于每个特征值 λ 求解$(A-\lambda I)\,\boldsymbol{x}=0$ 或 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 得到一个特征向量 $\boldsymbol{x}$。

范例 4 $A=\begin{bmatrix}\boldsymbol{1}&\boldsymbol{2}\\ \boldsymbol{2}&\boldsymbol{4}\end{bmatrix}$ 已经是奇异(行列式是零)，求出它的λ's 与 $\boldsymbol{x}$'s。

当 A 奇异，$\lambda=0$ 是一个特征值，方程式 $A\boldsymbol{x}=0\boldsymbol{x}$ 有解，他们是对应$\lambda=0$ 的特征向量。但是 $\det(A-\lambda I)=0$ 是找出所有λ's 与 $\boldsymbol{x}$'s 的方法，都是从 A 减去λI：

从对角线减去λ得到 $$A-\lambda I=\begin{bmatrix}1-\lambda&2\\ 2&4-\lambda\end{bmatrix} \tag{4}$$

计算 2×2 矩阵的行列式“$ad-bc$”，“ad”的部分是 $1-\lambda$乘 $4-\lambda$等于$\lambda^2-5\lambda+4$，不包含λ的部分“bc”是 2 乘 2。

$$\det\begin{bmatrix}1-\lambda&2\\ 2&4-\lambda\end{bmatrix}=(1-\lambda)(4-\lambda)-(2)(2)=\lambda^2-5\lambda \tag{5}$$

令行列式$\lambda^2-5\lambda=0$，一个解是$\lambda=0$(如预期，A 是奇异)，分解得到λ乘$\lambda-5$，另一个根是$\lambda=5$：

$\det(A-\lambda I)=\lambda^2-5\lambda=0$ 得到 $\lambda_1=0$ 与 $\lambda_2=5$

现在找出特征向量。分开$\lambda_1 = 0$ 与$\lambda_2 = 5$ 求解$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$：

$\lambda_1 = 0$ 时，$(A - 0I)\boldsymbol{x} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 得到一个特征向量 $\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$。

$\lambda_2 = 5$ 时，$(A - 5I)\boldsymbol{x} = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 得到一个特征向量 $\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$。

矩阵 $A - 0I$ 与 $A - 5I$ 都是奇异(因为 0 与 5 是特征值)，特征向量(2, –1)与(1, 2)在零空间中：$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ 就是 $A\boldsymbol{x} = \lambda\boldsymbol{x}$。

我们需要强调：$\lambda = 0$ 并不是特殊情形，好像其他数字一样，零可能是特征值也可能不是。如果 A 是奇异，$\lambda = 0$ 对应的特征向量产生零空间：$A\boldsymbol{x} = 0\boldsymbol{x} = \boldsymbol{0}$。如果 A 可逆，零就不是特征值。我们把 A 平移 I 的倍数使得 A 奇异。

这个例子中，平移矩阵 $A - 5I$ 是奇异，5 是另一个特征值。

摘要 对 $n \times n$ 矩阵求解特征值问题，依循下列步骤：

1. 计算 $A - \lambda I$ 的行列式。对角线减去λ，这个行列式从λ^n 或$-\lambda^n$ 开始，它是λ的 n 次多项式。
2. 找出多项式的根。求解 $\det(A - \lambda I) = 0$，这 n 个根是 A 的 n 个特征值，他们使得 $A - \lambda I$ 奇异。
3. 对每个特征值λ，求解$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ 得到一个特征向量 $\boldsymbol{x}$。

一个 2×2 矩阵特征向量的注解。当 $A - \lambda I$ 奇异，两个行都是一个向量(a, b)的倍数，特征向量是$(b, -a)$的任意倍数。这个例子有：

$\lambda = 0$：$A - 0I$ 的行在方向(1, 2)；特征向量的方向是(2, –1)。

$\lambda = 5$：$A - 5I$ 的行在方向(–4, 2)；特征向量的方向是(2, 4)。

在前面我们把最后一个特征向量写成(1, 2)，两个向量(1, 2)与(2, 4)都是正确的，一整条线都是特征向量–– $\boldsymbol{x}$ 的任意非零倍数与 $\boldsymbol{x}$ 有同样性质。MATLAB 的 eig(A) 会除以自己的长度使得特征向量变成单位向量。

我们必须加上一个警告，某些 2×2 矩阵只有一条直线的特征向量，这个只会发生在两个特征值相等的情形下。(另一方面 $A = I$ 有相等的特征值以及非常多的特征向量。) 缺乏一整组的特征向量，我们就没有一组基底，我们不可能把每个向量 $\boldsymbol{v}$ 写成特征向量的组合。下一章的内容说明在没有 n 个无关的特征向量时，我们无法对角化一个矩阵。

请尊重版权与译者的劳动成果，侵权必究!

行列式与迹

首先是坏消息：如果你把 A 的某一行加到另一行，或是交换行，通常都会改变特征值，消元法不能维持λ's。三角 U 在对角线上有自己的特征值--他们是主元，但是他们不是 A 的特征值！当行 1 加到行 2，特征值会改变：

$$U=\begin{bmatrix}1&3\\0&0\end{bmatrix}\text{有}\lambda=0\text{ 与}\lambda=1;\quad A=\begin{bmatrix}1&3\\2&6\end{bmatrix}\text{有}\lambda=0\text{ 与}\lambda=7。$$

好的新消息：从矩阵可以快速的找出λ_1 乘λ_2 的乘积与$\lambda_1+\lambda_2$ 的总和，对 A 来说乘积是 0 乘 7，这个与行列式一致(行列式是 0)。特征值的总和是 0 + 7，这个与主对角线的总和一致(**迹(trace)**是 1 + 6)，这些快速检验都是可用的：

n 个特征值的乘积等于行列式。
n 个特征值的总和等于 n 个对角线单元的总和。

主对角线单元的总和称为 A 的**迹**：

$$\lambda_1+\lambda_2+\ldots+\lambda_n=\text{迹}=a_{11}+a_{22}+\ldots+a_{nn}\tag{6}$$

这些检验非常有用，他们的证明在问题 16-17，下个段落也会再出现。他们没有消除计算λ's 的痛苦，但是计算错误时，他们会如此告诉我们。要计算正确的λ's，回到 $\det(A-\lambda I)=0$。

当矩阵是 2×2 时，迹与行列式确实告诉我们每件事情，我们从来都不想弄错！此处迹= 3 与 det = 2，所以特征值是λ = 1 与 2：

$$A=\begin{bmatrix}1&9\\2&0\end{bmatrix}\quad\text{或}\quad\begin{bmatrix}3&1\\-2&0\end{bmatrix}\quad\text{或}\quad\begin{bmatrix}7&-3\\10&-4\end{bmatrix}\tag{7}$$

下列是最佳矩阵找出特征值的问题：三角矩阵。

为什么三角矩阵的特征值都在对角线上?

虚数特征值

再来一点消息(不是太可怕)，特征值可能不是实数。

范例 5 旋转 90° 的 $Q=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$ **没有实数特征值**，它的特征值是$\lambda_1=i$ 与 $\lambda_2=-i$，则$\lambda_1+\lambda_2=$ 迹 $=0$ 且$\lambda_1\lambda_2=$ **行列式** $=1$。

旋转之后，没有实数向量 Qx 留在与 x 相同的方向($x=0$ 没有用)，除非我们使用**虚数**，否则就不存在特征向量，我们会这么做的。

看看 $i=\sqrt{-1}$ 有什么帮助，检视 $Q^2=-I$，如果 Q 旋转 90°，Q^2 就旋转 180°，它的特征值是–1 与–1。(当然 $-I\boldsymbol{x}=-1\boldsymbol{x}$)。对 Q 平方也会对 λ 平方，所以我们必须有 $\lambda^2=-1$，因此 90°旋转矩阵 Q 的特征值是 $+i$ 与 $-i$，因为 $i^2=-1$。

这些 i's 也来自 $\det(Q-\lambda I)=0$，方程式得到 $\lambda^2+1=0$，它的根是 i 与 $-i$。我们在特征向量也会看到虚数：

复数特征向量
$$\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}1\\i\end{bmatrix}=-i\begin{bmatrix}1\\i\end{bmatrix}\quad\text{与}\quad\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}i\\1\end{bmatrix}=i\begin{bmatrix}i\\1\end{bmatrix}$$

也许这些复数向量 $\boldsymbol{x}_1=(1,i)$ 与 $\boldsymbol{x}_2=(i,1)$ 在旋转之后仍然维持原来的方向，不要问我怎么回事。这个例子指出最重要的一点，就是实数矩阵很容易有复数特征值与特征向量。这些特殊的特征值 i 与 $-i$ 也说明了 Q 的两个特殊性质：

1. Q 是正交矩阵所以每个 λ 的绝对值 $|\lambda|=1$。
2. Q 是反对称矩阵所以每个 λ 是纯虚数。

对称矩阵($S^{\mathrm{T}}=S$)可以比喻成实数，反对称矩阵($A^{\mathrm{T}}=-A$)可以比喻成虚数，正交矩阵($Q^{\mathrm{T}}Q=I$)对应 $|\lambda|=1$ 的复数。对于 S 与 A 与 Q 的特征值来说，他们不只是类比——段落 6.4 证明了他们是事实。

所有这些特殊矩阵的特征向量都互相垂直，$(i,1)$ 与 $(1,i)$ 也是一种垂直(第九章说明复数的点积)。

AB 与 *A*+*B* 的特征值

第一个有关 AB 的特征值的猜想是错误的，A 的一个特征值 λ 乘 B 的一个特征值 β 通常不会得到 AB 的一个特征值：

错误的证明
$$AB\boldsymbol{x}=A\beta\boldsymbol{x}=\beta A\boldsymbol{x}=\beta\lambda\boldsymbol{x}\tag{8}$$

似乎 β 乘 λ 是一个特征值，当 $\boldsymbol{x}$ 同时是 A 与 B 的特征向量，这个证明是正确的。这个错误在于期待 A 与 B 自动地共享同样的特征向量 $\boldsymbol{x}$，一般来说这是不可能的，A 的特征向量通常不会是 B 的特征向量。当 1 是 AB 的一个特征值，A 与 B 会有全零的特征值：

$$A=\begin{bmatrix}0&1\\0&0\end{bmatrix}\quad\text{与}\quad B=\begin{bmatrix}0&0\\1&0\end{bmatrix};\ \text{则}\ AB=\begin{bmatrix}1&0\\0&0\end{bmatrix}\quad\text{与}\quad A+B=\begin{bmatrix}0&1\\1&0\end{bmatrix}$$

相同的理由，$A+B$ 的特征值通常不是 $\lambda+\beta$。当 $A+B$ 有特征值 1 与–1 时(至少他们相加得到零)，此时 $\lambda+\beta=0$。

请尊重版权与译者的劳动成果，侵权必究!

前面错误的证明也提供了一些信息，假设 $\boldsymbol{x}$ 真的是同时 A 与 B 的特征向量，我们确实有 $AB\boldsymbol{x} = \lambda\beta\boldsymbol{x}$ 与 $BA\boldsymbol{x} = \lambda\beta\boldsymbol{x}$，当所有的特征向量都共享，我们可以乘特征值。共享特征向量的测试 $AB = BA$ 在量子力学非常重要--说明线性代数应用的时间到了：

> A 与 B 共享同样 n 个无关的特征向量 当且仅当 $\boldsymbol{AB = BA}$。

海森堡的不确定原理 在量子力学中，位置矩阵 P 与动量矩阵 Q 不能交换，事实上 $QP - PQ = I$ (这些是无限矩阵)。要同时有 $P\boldsymbol{x} = \boldsymbol{0}$ 且 $Q\boldsymbol{x} = \boldsymbol{0}$，必然需要 $\boldsymbol{x} = I\boldsymbol{x} = \boldsymbol{0}$。如果我们确切的知道位置，我们就不可能确切的知道动量。问题 36 推导海森堡不确定原理 $\|P\boldsymbol{x}\|\,\|Q\boldsymbol{x}\| \geq \frac{1}{2}\|\boldsymbol{x}\|^2$。

主要观念的复习

1. $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 表示 $\boldsymbol{x}$ 被 A 乘时，$\boldsymbol{x}$ 维持同样的方向。
2. $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 也表示 $\det(A - \lambda I) = 0$，这个决定了 n 个特征值。
3. A^2 与 A^{-1} 的特征值是 λ^2 与 λ^{-1}，而且有相同的特征向量。
4. λ's 的总和等于 A 主对角线单元的总和(迹)，λ's 的乘积等于 A 的行列式。
5. 投影 P，镜射 R，90°旋转 Q 有特殊的特征值 1 与 0 与–1 与 i 与–i。奇异矩阵 $\lambda = 0$，三角矩阵有 λ's 在对角线。
6. 矩阵的特殊性质得到特殊特征值与特殊特征向量，这是本章的主题(在最后制成一张表) 。

已解范例

6.1A 求出 A 与 A^2 与 A^{-1} 与 $A + 4I$ 的特征值与特征向量：

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \text{与} \quad A^2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$$

检验迹 $\lambda_1 + \lambda_2 = 4$ 与行列式 $\lambda_1\lambda_2 = 3$。

解 A 的特征值来自 $\det(A - \lambda I) = 0$：

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \det A = \begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3 = 0$$

分解得到 $(\lambda - 1)(\lambda - 3) = 0$ 所以 A 的特征值是 $\lambda_1 = 1$ 与 $\lambda_2 = 3$。对迹来说，总和 2 + 2 与 1 + 3 一致，行列式 3 与 $\lambda_1\lambda_2$ 的乘积一致。

特征向量分别来自求解$(A-\lambda I)\boldsymbol{x}=\mathbf{0}$也就是$A\boldsymbol{x}=\lambda\boldsymbol{x}$：

$\lambda=1$：$(A-I)\boldsymbol{x}=\begin{bmatrix}1&-1\\-1&1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$ 得到特征向量 $\boldsymbol{x}_1=\begin{bmatrix}1\\1\end{bmatrix}$。

$\lambda=3$：$(A-3I)\boldsymbol{x}=\begin{bmatrix}-1&-1\\-1&-1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$ 得到特征向量 $\boldsymbol{x}_2=\begin{bmatrix}1\\-1\end{bmatrix}$。

A^2 与 A^{-1} 与 $A+4I$ 维持与 A 相同的特征向量，特征值是λ^2 与λ^{-1} 与$\lambda+4$：

A^2 有特征值 $1^2=1$ 与 $3^2=9$　　A^{-1} 有$\frac{1}{1}$与$\frac{1}{3}$　　$A+4I$ 有 $\begin{matrix}1+4=5\\3+4=7\end{matrix}$

后面段落的注解：A 有正交的特征向量(段落 6.4 的对称矩阵)。由于$\lambda_1\neq\lambda_2$，A 可以对角化(段落 6.2)。与任何特征值是 1 与 3 的 2×2 矩阵相似的 A(段落 6.2)。由于 $A=A^{\mathrm{T}}$ 且λ's 都是正数，A 是正定(positive definite)矩阵(段落 6.5)。

6.1B　你如何估测任意矩阵 A 的特征值？Gershgorin 给了答案。

A 的每一个特征值必须与最少一个的对角线单元 a_{ii} 很“接近”。λ很“接近 a_{ii}”表示$|a_{ii}-\lambda|$不大于 R_i，其中 **R_i 是矩阵行 i 其他的$|a_{ij}|$的总和**，则 $R_i=\Sigma_{j\neq i}|a_{ij}|$是以 a_{ii} 为圆心的圆半径。

每一个λ落在环绕一个或更多个对角单元 a_{ii} 的圆：$|a_{ii}-\lambda|\leq R_i$。

下列是推理过程：如果λ是特征值，则 $A-\lambda I$ 不可逆，那么 $A-\lambda I$ 就不可能是对角支配(见段落 2.5)，所以最少有一个对角单元 $a_{ii}-\lambda$不大于 R_i，其中 R_i 是行 i 中其他单元$|a_{ij}|$(我们取绝对值！)的总和。

范例 1　A 的每个特征值λ落在一个或是同时落在两个 Gershgorin 圆中，圆心是 a 与 d，半径是 $R_1=|b|$与 $R_2=|c|$。

$$A=\begin{bmatrix}a&b\\c&d\end{bmatrix}\qquad\begin{matrix}\text{第一个圆：}&|\lambda-a|\leq|b|\\\text{第二个圆：}&|\lambda-d|\leq|c|\end{matrix}$$

这些是复数平面的圆，因为λ当然可能是复数。

范例 2　A 的所有特征值λ落在一个半径 $R=3$ 的圆中，这个圆环绕一个或是更多的对角线单元 d_1,d_2,d_3：

$$A=\begin{bmatrix}d_1&1&2\\2&d_2&1\\-1&2&d_3\end{bmatrix}\qquad\begin{matrix}|\lambda-d_1|\leq1+2=R_1\\|\lambda-d_2|\leq2+1=R_2\\|\lambda-d_3|\leq1+2=R_3\end{matrix}$$

在这个例子中，你看到的“接近”表示距离 d_1 或 d_2 或 d_3 不超过 3。

请尊重版权与译者的劳动成果，侵权必究！

6.1C 求出对称 3×3 矩阵 S 的特征值与特征向量：

对称矩阵
奇异矩阵
迹 1 + 2 + 1 = 4

$$S=\begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&1\end{bmatrix}$$

解 S 的所有行的总和都是零，向量 $\boldsymbol{x}=(1,1,1)$得到 $S\boldsymbol{x}=\boldsymbol{0}$，这是$\lambda=0$ 对应的一个特征向量。要求得λ_2 与λ_3，我会计算 3×3 的行列式：

$$\det(S-\lambda I)=\begin{vmatrix}1-\lambda&-1&0\\-1&2-\lambda&-1\\0&-1&1-\lambda\end{vmatrix}\begin{matrix}=(1-\lambda)(2-\lambda)(1-\lambda)-2(1-\lambda)\\=(1-\lambda)[(2-\lambda)(1-\lambda)-2]\\=(1-\lambda)(-\lambda)(3-\lambda)\end{matrix}$$

这三个因子得到$\lambda=0,1,3$，每个特征值对应一个特征向量(或是特征向量的直线)：

$$\boldsymbol{x}_1=\begin{bmatrix}1\\1\\1\end{bmatrix}\quad S\boldsymbol{x}_1=0\boldsymbol{x}_1\qquad \boldsymbol{x}_2=\begin{bmatrix}1\\0\\-1\end{bmatrix}\quad S\boldsymbol{x}_2=1\boldsymbol{x}_2\qquad \boldsymbol{x}_3=\begin{bmatrix}1\\-2\\1\end{bmatrix}\quad S\boldsymbol{x}_3=3\boldsymbol{x}_3$$

我再次注意到当 S 对称时特征向量互相垂直。我们很幸运求出$\lambda=0,1,3$，对于大型矩阵我会用 eig(A)，从来不碰行列式。

完整的指令 [X, E] = eig(A)在 X 的列会得到单位特征向量。

问题集 6.1

1 本章开头的例子有 A 的次方：

$$A=\begin{bmatrix}.8&.3\\.2&.7\end{bmatrix}\quad 与\quad A^2=\begin{bmatrix}.70&.45\\.30&.55\end{bmatrix}\quad 与\quad A^\infty=\begin{bmatrix}.6&.6\\.4&.4\end{bmatrix}$$

求出这些矩阵的特征值，所有的次方有相同的**特征向量**。

(a) 从 A 说明行交换或如何得到不同的特征值。

(b) 为什么零特征值在消元步骤下不会改变?

2 求出这两个矩阵的特征值与特征向量：

$$A=\begin{bmatrix}1&4\\2&3\end{bmatrix}\quad 与\quad A+I=\begin{bmatrix}2&4\\2&4\end{bmatrix}$$

$A+I$ 有与 A_____的特征向量，它的特征值是_____多 1。

3 计算 A 与 A^{-1} 的特征值与特征向量，检验迹！

$$A=\begin{bmatrix}0&2\\1&1\end{bmatrix}\quad 与\quad A^{-1}=\begin{bmatrix}-1/2&1\\1/2&0\end{bmatrix}$$

A^{-1} 有与 A___的特征向量，当 A 有特征值λ_1 与λ_2，逆矩阵有特征值____。

4 计算 A 与 A^2 的特征值与特征向量：

$$A=\begin{bmatrix}-1 & 3\\ 2 & 0\end{bmatrix} \quad 与 \quad A^2=\begin{bmatrix}7 & -3\\ -2 & 6\end{bmatrix}$$

A^2 有与 A____的特征向量，当 A 有特征值λ_1与λ_2，则 A^2 有特征值_____。本例子中，为什么 $\lambda_1^2+\lambda_2^2=13$？

5 计算 A 与 B(三角矩阵很容易)与 $A+B$ 的特征值：

$$A=\begin{bmatrix}3 & 0\\ 1 & 1\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}1 & 1\\ 0 & 3\end{bmatrix} \quad 与 \quad A+B=\begin{bmatrix}4 & 1\\ 1 & 4\end{bmatrix}$$

$A+B$ 的特征值(等于) (不等于) A 的特征值加 B 的特征值。

6 计算 A 与 B 与 AB 与 BA 的特征值：

$$A=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix} \quad 与 \quad AB=\begin{bmatrix}1 & 2\\ 1 & 3\end{bmatrix} \quad 与 \quad BA=\begin{bmatrix}3 & 2\\ 1 & 1\end{bmatrix}$$

(a) AB 的特征值等于 A 的特征值乘 B 的特征值吗？

(b) AB 的特征值等于 BA 的特征值吗？

7 消元法得到 $A = LU$，U 的特征值在对角线上，他们是_____。L 的特征值在对角线上，他们全部是_____。A 的特征值与_____不相等。

8 (a) 若你知道 $\boldsymbol{x}$ 是特征向量，找出λ的方法是_____。

(b) 若你知道λ是特征值，找出 $\boldsymbol{x}$ 的方法是_____。

9 为了证明(a), (b)与(c)，你会对 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 做什么动作？

(a) λ^2 是 A^2 的特征值，如同问题 4。

(b) λ^{-1} 是 A^{-1} 的特征值，如同问题 3。

(c) $\lambda + 1$ 是 $A + I$ 的特征值，如同问题 2。

10 计算马尔可夫矩阵 A 与 A^∞ 的特征值与特征向量，从答案说明为什么 A^{100} 很接近 A^∞ ？

$$A=\begin{bmatrix}.6 & .2\\ .4 & .8\end{bmatrix} \quad 与 \quad A^\infty=\begin{bmatrix}1/3 & 1/3\\ 2/3 & 2/3\end{bmatrix}$$

11 2×2 矩阵的特征值$\lambda_1 \neq \lambda_2$时，有一个奇怪的事实：$A - \lambda_1 I$ 的列是特征向量 $\boldsymbol{x}_2$ 的倍数，有什么想法为什么它会这样？

请尊重版权与译者的劳动成果，侵权必究!

12 找出矩阵 P(投影矩阵有$\lambda = 1$ 与 0)的 3 个特征向量：

$$\textbf{投影矩阵}\quad P = \begin{bmatrix} .2 & .4 & 0 \\ .4 & .8 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

如果两个特征向量共享同一个λ，他们所有的线性组合也是。找出 P 的一个不含零分量的特征向量。

13 单位向量 $\boldsymbol{u} = (\frac{1}{6}, \frac{1}{6}, \frac{3}{6}, \frac{5}{6})$，利用 $\boldsymbol{u}$ 创建秩一投影矩阵 $P = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$。因为 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$，这个矩阵有 $P^2 = P$。

(a) $P\boldsymbol{u} = \boldsymbol{u}$ 来自$(\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})\boldsymbol{u} = \boldsymbol{u}$(_____)，则 $\boldsymbol{u}$ 是对应$\lambda = 1$ 的特征向量。

(b) 若 $\boldsymbol{v}$ 与 $\boldsymbol{u}$ 垂直，证明 $P\boldsymbol{v} = \boldsymbol{0}$，则$\lambda = 0$。

(c) 找出对应$\lambda = 0$ 的 P 的三个无关特征向量。

14 利用二次公式求解 $\det(Q - \lambda I) = 0$，得到$\lambda = \cos\theta \pm i\sin\theta$ ：

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$ 旋转 xy 平面角度θ，没有实数λ's。

求解$(Q - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ 找出 Q 的特征向量，使用 $i^2 = -1$。

15 每个排列矩阵使得 $\boldsymbol{x} = (1, 1, ..., 1)$保持不变，则$\lambda = 1$。利用 $\det(P - \lambda I) = 0$ 找出下列排列矩阵的另外两个λ's(可能是复数)：

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad 与 \quad P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

16 **A 的行列式等于乘积$\lambda_1\lambda_2 \ldots \lambda_n$**，从多项式 $\det(A - \lambda I)$开始，分解它的 n 个因子(一定可以)，然后令$\lambda = 0$：

$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda)$ 所以 $\det A =$ _____。

检查范例 1 的规则，其中马尔可夫矩阵有$\lambda = 1$ 与 1/2。

17 对角线单元的总和(迹)等于特征值的总和：

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ 有 } \det(A - \lambda I) = \lambda^2 - (a + d)\lambda + ad - bc = 0$$

二次公式得到特征值$\lambda =$ $(a + d + \sqrt{\quad})/2$ 与$\lambda =$ _____，他们的总和是_____。

如果 A 有$\lambda_1 = 3$ 与$\lambda_2 = 4$，则 $\det(A - \lambda I) =$ _____。

18 如果 A 有$\lambda_1 = 4$ 与$\lambda_2 = 5$，则 $\det(A - \lambda I) = (\lambda - 4)(\lambda - 5) = \lambda^2 - 9\lambda + 20$。找出三个矩阵使得迹 $a + d = 9$ 与行列式 20 与$\lambda = 4, 5$。

19 一个 3×3 矩阵 B 已知有特征值 0, 1, 2，这个信息足够找出下列哪三个(可能的话给出答案)：

(a) B 的秩。

(b) $B^{\mathrm{T}}B$ 的行列式。

(c) $B^{\mathrm{T}}B$ 的特征值。

(d) $(B^2 + I)^{-1}$ 的特征值。

20 选择 A 与 C 的最后一行，使得特征值是 4, 7 与 1, 2, 3:

伴随(companion)矩阵 $A = \begin{bmatrix} 0 & 1 \\ * & * \end{bmatrix}$ $C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ * & * & * \end{bmatrix}$

21 **A 的特征值与 A^{T} 的特征值相同**，这是因为 $\det(A - \lambda I) = \det(A^{\mathrm{T}} - \lambda I)$。因为 _____，这是成立的。举出例子说明 A 的特征向量与 A^{T} 的特征向量不相同。

22 创建任意的 3×3 马尔可夫矩阵：正的单元且每个列的总和是 1。证明 $M^{\mathrm{T}}(1, 1, 1) = (1, 1, 1)$。使用问题 21，$\lambda = 1$ 也是 M^{T} 的特征值。挑战：一个 3×3 奇异马尔可夫矩阵的迹是 1/2，它有什么样的特征值λ's?

23 找出三个 2×2 矩阵使得$\lambda_1 = \lambda_2 = 0$，迹是 0，行列式是 0。$A$ 可能不是零矩阵，但是请检验 $A^2 = 0$。

24 矩阵是奇异且秩一，找出三个λ's 与三个特征向量：

$$A = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 4 & 2 & 4 \\ 2 & 1 & 2 \end{bmatrix}$$

25 假设 A 与 B 有相同的特征值$\lambda_1, \ldots, \lambda_n$，而且有相同对应的特征向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$，则 $A = B$。理由：任意向量 $\boldsymbol{x}$ 是 $c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$ 的组合，则 $A\boldsymbol{x} =$ ？ $B\boldsymbol{x} =$ ？

26 方块 B 有特征值 1, 2 且 C 有特征值 3, 4 且 D 有特征值 5, 7。求出 4×4 矩阵 A 的特征值。

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 & 0 \\ -2 & 3 & 0 & 4 \\ 0 & 0 & 6 & 1 \\ 0 & 0 & 1 & 6 \end{bmatrix}$$

27 求出 A 与 C 的秩以及四个特征值：

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{与} \quad C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

28 从 I 减去前面的 A，求出λ's 与行列式：

$$B = A - I = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad 与 \quad C = I - A = \begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$$

29 (复习) 求出 A, B, C 的特征值:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \quad 与 \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 0 & 0 \end{bmatrix} \quad 与 \quad C = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

30 当 $a + b = c + d$ 证明(1, 1)是一个特征向量并找出两个特征值：

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

31 如果我们交换行 1 与 2，交换列 1 与 2，特征值不会改变。若$\lambda = 11$ 求出 A 与 B 的特征向量。秩一得到$\lambda_2 = \lambda_3 = 0$。

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 3 \\ 4 & 8 & 4 \end{bmatrix} \quad 与 \quad B = PAP^{\mathrm{T}} = \begin{bmatrix} 6 & 3 & 3 \\ 2 & 1 & 1 \\ 8 & 4 & 4 \end{bmatrix}$$

32 假设 A 有特征值 0, 3, 5，对应的无关特征向量是 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$。

(a) 求出零空间的一组基底与列空间的一组基底。

(b) 求出 $A\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{w}$ 的一个特定解，求出全部的解。

(c) $A\boldsymbol{x} = \boldsymbol{u}$ 无解。如果有解，_____会在列空间。

挑战问题

33 证明 $\boldsymbol{u}$ 是秩一的 2×2 矩阵 $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 的一个特征向量，找出 A 的两个特征值，检验$\lambda_1 + \lambda_2$ 与迹 $u_1v_1 + u_2v_2$ 一致。

34 从 $\det(P - \lambda I) = 0$ 求出排列矩阵 P 的特征值。哪些向量不因排列而改变？他们是对应$\lambda = 1$ 的特征向量。你能够找出三个以上的特征向量吗？

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

35 有 6 个 3×3 的排列矩阵 P，P 的行列式可以是什么数字？主元可以是什么数字？P 的迹可以是什么数字？哪 4 个数字可以作为 P 的特征值，如同问题 15?

36 (海森堡不确定原理) 无限矩阵会有 $AB-BA=I$，其中 $A=A^{\mathrm{T}}$ 与 $B=-B^{\mathrm{T}}$，则

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}AB\boldsymbol{x}-\boldsymbol{x}^{\mathrm{T}}BA\boldsymbol{x}\le 2\|A\boldsymbol{x}\|\|B\boldsymbol{x}\|$$

说明最后一个步骤使用了苏瓦兹不等式 $|\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}|\le\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$，则海森堡不等式说明了 $\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$ 乘 $\|B\boldsymbol{x}\|/\|\boldsymbol{x}\|$ 最小是 1/2，所以不可能同时得到位置误差与动量误差都很小。

37 求出一个 2×2 旋转矩阵 A(不是 I)使得 $A^3=I$，它的特征值必须满足 $\lambda^3=1$，他们可以是 $e^{2\pi i/3}$ 与 $e^{-2\pi i/3}$，矩阵的迹与行列式为何？

38 (a) 找出 A 的特征值与特征向量，他们与 c 有关：

$$A=\begin{bmatrix}.4 & 1-c\\ .6 & c\end{bmatrix}$$

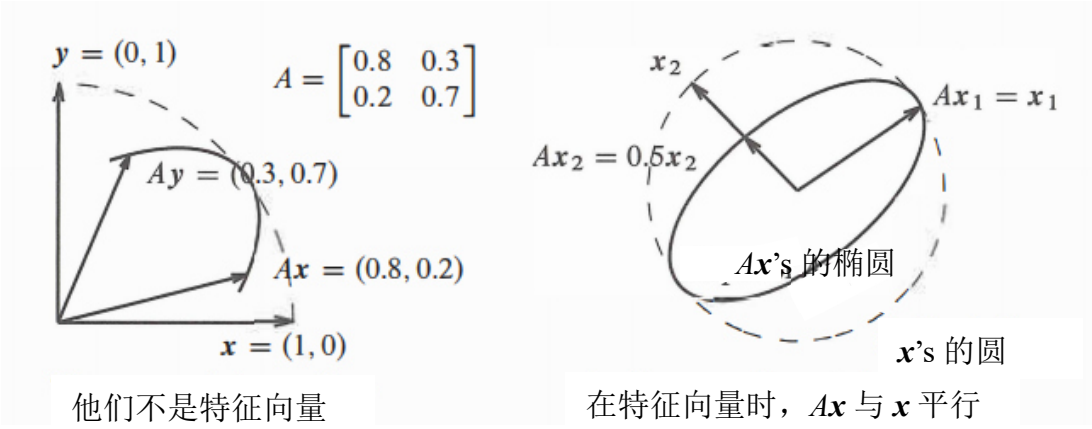
(b) 当 $c=1.6$，证明 A 只有一条线的特征向量。

(c) 当 $c=.8$ 这是一个马尔可夫矩阵，则 A^n 会趋近哪个矩阵 A^∞ ？

MATLAB 的 eigshow

MATLAB 有一个展示(键入 **eigshow**)，会显示 2×2 矩阵的特征值问题。它从单位向量 $\boldsymbol{x}$ 开始，鼠标使得向量环绕单位圆，同时屏幕显示 $A\boldsymbol{x}$，不同颜色也在移动。$A\boldsymbol{x}$ 可能超前 $\boldsymbol{x}$ 也可能落后 $\boldsymbol{x}$，有时候 $A\boldsymbol{x}$ 与 $\boldsymbol{x}$ 平行。

在平行的时候，$A\boldsymbol{x}=\lambda\boldsymbol{x}$(在第二图的 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$)。



他们不是特征向量　　在特征向量时，$A\boldsymbol{x}$ 与 $\boldsymbol{x}$ 平行

当单位向量 $\boldsymbol{x}$ 共线时，特征值 λ 是 $A\boldsymbol{x}$ 的长度。A 的内建选择有三种可能性：当 $\boldsymbol{x}$ 与 $A\boldsymbol{x}$ 交会时有 0, 1 或 2 个实数向量。

椭圆的轴是段落 7.4 的**奇异向量**——若 $A^{\mathrm{T}}=A$，轴也是**特征向量**。

请尊重版权与译者的劳动成果，侵权必究!

6.2 对角化矩阵

1 $AX=X\Lambda$的列是 $A\boldsymbol{x}_k=\lambda_k\boldsymbol{x}_k$。特征值矩阵$\Lambda$是对角矩阵。

2 X 的 n 个无关的特征向量可以对角化 A：$\boldsymbol{A=X\Lambda X^{-1}}$ 与 $\boldsymbol{\Lambda=X^{-1}AX}$。

3 特征向量矩阵 X 同样可以对角化所有的次方 A^k：$\boldsymbol{A^k=X\Lambda^kX^{-1}}$。

4 利用 $\boldsymbol{u}_k=A^k\boldsymbol{u}_0=X\Lambda^kX^{-1}\boldsymbol{u}_0=c_1(\lambda_1)^k\boldsymbol{x}_1+\ldots+c_n(\lambda_n)^k\boldsymbol{x}_n$，求解 $\boldsymbol{u}_{k+1}=A\boldsymbol{u}_k$。

5 **没有相等的特征向量** $\Rightarrow X$ 是可逆且 A 可以对角化。
有相等的特征向量 $\Rightarrow A$ 的无关特征向量可能不足，则 X^{-1} 失败。

6 每个矩阵 $C=B^{-1}AB$ 与 A 有相同的特征值，这些 C 与 A “**相似**”。

当 $\boldsymbol{x}$ 是一个特征向量，被 A 乘只是乘一个数字λ：$A\boldsymbol{x}=\lambda\boldsymbol{x}$。矩阵的所有困难一扫而空，我们可以分开处理特征向量，而不是处理一个互联系统。如同对角矩阵一样--没有非对角线的互联，对角矩阵的 100 次方非常容易。

本段落的重点非常直接，我们使用特征向量的性质，**使得矩阵 A 变成对角矩阵Λ**，这个矩阵形式是我们的关键概念。我们立即开始一个重要的计算，下一页会解释为什么 $AX=X\Lambda$。

对角化 假设 $n\times n$ 矩阵 A 有 n 个线性无关的特征向量 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$，把他们放在**特征向量矩阵 X 的列，则 $X^{-1}AX$ 就是对角矩阵Λ**：

固有向量矩阵 X
固有值矩阵 Λ

$$X^{-1}AX=\Lambda=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix} \tag{1}$$

矩阵 A 被“对角化”，我是使用大写的Λ代表特征值矩阵，是因为小写的λ's(特征值)用在对角线上。

范例 1 A 是三角矩阵，所以它的特征值在对角线上：$\lambda=1$ 与$\lambda=6$。

特征向量进入 X $\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}$

$$\underset{X^{-1}}{\begin{bmatrix}1&-1\\0&1\end{bmatrix}}\quad\underset{A}{\begin{bmatrix}\mathbf{1}&\mathbf{5}\\0&\mathbf{6}\end{bmatrix}}\quad\underset{X}{\begin{bmatrix}1&1\\0&1\end{bmatrix}}=\underset{\Lambda}{\begin{bmatrix}\mathbf{1}&0\\0&\mathbf{6}\end{bmatrix}}$$

换言之，$A=X\Lambda X^{-1}$，检视 $A^2=X\Lambda X^{-1}\,X\Lambda X^{-1}$，所以 $\boldsymbol{A^2}$ **是** $\boldsymbol{X\Lambda^2X^{-1}}$。

A^2 有跟 X 中的相同特征向量，以及在Λ^2 的特征值平方。

为什么 $AX=X\Lambda$？ A 乘自己的特征向量，这些特征向量是 X 的列，AX 的第一列是 $A\boldsymbol{x}_1$，就是$\lambda_1\boldsymbol{x}_1$。$X$ 的每一列被**自己的特征值**乘：

A 乘 X
$$AX=A\begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}=\begin{bmatrix}\lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n\end{bmatrix}$$

技巧在于把矩阵 AX 分割成 $X\Lambda$：

X 乘Λ
$$\begin{bmatrix}\lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=X\Lambda$$

这些矩阵保持右序！则λ_1 乘第一列 $\boldsymbol{x}_1$，如所示。现在完成对角化，我们可以把 $AX=X\Lambda$写成两个方式：

$$AX=X\Lambda \quad \text{是} \quad \boldsymbol{X^{-1}AX=\Lambda} \quad \text{或} \quad \boldsymbol{A=X\Lambda X^{-1}} \tag{2}$$

矩阵 X 可逆，这是因为它的列(A 的特征向量)假设是线性无关，没有 n 个无关的特征向量，我们不能对角化。

A 与Λ有相同的特征值$\lambda_1, \ldots, \lambda_n$，特征向量则不同。原始特征向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 的工作是对角化 A，X 中的特征向量产生 $A=X\Lambda X^{-1}$。你会很快看到它们的简洁性，重要性以及意义。第 k 个次方是 $A^k=X\Lambda^kX^{-1}$，这个很容易计算：

$$\boldsymbol{A^k=(X\Lambda X^{-1})\,(X\Lambda X^{-1})\ldots(X\Lambda X^{-1})=X\Lambda^kX^{-1}}$$

范例 1 的 A 的次方
$$\begin{bmatrix}1 & 5\\ 0 & 6\end{bmatrix}^k=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & \\ & 6^k\end{bmatrix}\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{1} & \boldsymbol{6^k-1}\\ 0 & \boldsymbol{6^k}\end{bmatrix}=A^k$$

$k=1$ 我们得到 A；$k=0$ 我们得到 $A^0=I(\lambda^0=1)$；$k=-1$ 我们得到 A^{-1}。你可以看到 $A^2=$[1　35; 0　36]是如何适配 $k=2$ 的公式。

在范例 2 再次使用Λ之前，下列是 4 个小注解。

注解 1　假设特征值$\lambda_1, \ldots, \lambda_n$ 都不相同，则特征向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 自动是无关，特征向量矩阵可逆。**任何没有重复特征值的矩阵都可以对角化。**

注解 2　我们可以使用任意非零乘数乘特征向量，$A(c\boldsymbol{x})=\lambda(c\boldsymbol{x})$仍然成立。范例 1 中，我们可以把 $\boldsymbol{x}=(1, 1)$除以 $\sqrt{2}$ 产生单位向量。

MATLAB 以及其他代码都是得到长度$\|\boldsymbol{x}\|=1$ 的特征向量。

请尊重版权与译者的劳动成果，侵权必究!

注解 3　X 特征向量的顺序与Λ特征值的顺序相同，要反转Λ的顺序，把 X 的特征向量(1, 1)放在(1, 0)之前：

新顺序 6, 1　$$\begin{bmatrix}0 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 5\\ 0 & 6\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix} = \begin{bmatrix}6 & 0\\ 0 & 1\end{bmatrix} = \Lambda_{new}$$

要对角化 A，我们必须使用特征向量矩阵，从 $X^{-1}AX = \Lambda$我们知道 $AX = X\Lambda$。假设 X 的第一列是 $\boldsymbol{x}$，则 AX 与 $X\Lambda$的第一列是 $A\boldsymbol{x}$ 与$\lambda_1\boldsymbol{x}$，这些要相等，$\boldsymbol{x}$ 必须是特征向量。

注解 4　(重复特征值的**重复警告**) 有些矩阵的特征向量不足，这些矩阵不能对角化，下列是两个例子：

不能对角化　$$A = \begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix} \quad 与 \quad B = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$

他们的特征值刚好是 0 与 0，$\lambda = 0$ 没有什么特殊，问题是重复的λ。第一个矩阵的所有特征向量是(1, 1)的倍数：

只有一条特征向量直线　$A\boldsymbol{x} = 0\boldsymbol{x}$ 表示 $\begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$ 且 $\boldsymbol{x} = c\begin{bmatrix}1\\ 1\end{bmatrix}$

没有第二个特征向量，所以这个不寻常的矩阵 A 不能对角化。

这些矩阵是检验特征向量叙述的最佳例子，在许多是非题中，不能对角化的矩阵会导致测试失败。

请记住**可逆与可对角化之间没有关联**：

- **可逆**　与特征值有关($\lambda = 0$ 或$\lambda \neq 0$)。
- **可对角化**　与特征向量有关(X 的向量不足或足够)。

每个特征值对应最少一个特征向量！$A - \lambda I$ 是奇异，如果$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ 引导你得到 $\boldsymbol{x} = \boldsymbol{0}$，使得$\lambda$不是特征值，这是求解 $\det(A - \lambda I) = 0$ 的一个错误。

对应 *n* 个不同λ's 的特征向量是无关，则我们可以对角化 *A*。

对应不同 λ 的无关 $\boldsymbol{x}$　对应相异(全部不同)特征值的特征向量 $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ 必然是**线性无关**；一个具有不同特征值(**没有重复的λ's**)的 $n \times n$ 矩阵必然可以对角化。

证明　假设 $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}$，被 A 乘得到 $c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}$，被$\lambda_2$ 乘得到 $c_1\lambda_2\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}$，现在两式相减：

相减得到　$(\lambda_1 - \lambda_2)c_1\boldsymbol{x}_1 = \boldsymbol{0}$，因此 $c_1 = 0$

由于λ's 不相同且 $\boldsymbol{x}_1 \neq \boldsymbol{0}$，我们强迫得到 $c_1 = 0$ 的结论。同理 $c_2 = 0$，只有 $c_1 = c_2 = 0$ 的组合才能得到 $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}$，所以特征向量 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 必须是无关。

这个证明可以直接延伸至 j 个特征向量，假设 $c_1\boldsymbol{x}_1 + \ldots + c_j\boldsymbol{x}_j = \boldsymbol{0}$，左乘 A，左乘λ_j，然后相减，得到$(\lambda_j - \lambda_j) = 0$ 乘 $\boldsymbol{x}_j$，于是 $\boldsymbol{x}_j$ 消失。现在左乘 A，左乘λ_{j-1}，然后相减，移除 $\boldsymbol{x}_{j-1}$。最终只剩下 $\boldsymbol{x}_1$：

我们得到 $(\lambda_1 - \lambda_2) \ldots (\lambda_1 - \lambda_j)\, c_1\boldsymbol{x}_1 = \boldsymbol{0}$ 强制使得 $c_1 = 0$ (3)

同理，每个 $c_i = 0$。当所有的λ都不相同，特征向量是**无关**。一整个特征向量的集合可以进入特征向量矩阵 X 的列。

范例 2 ***A* 的次方** 前一个段落的马尔可夫矩阵 $A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$ 有特征值$\lambda_1 = 1$ 与$\lambda_2 = .5$，此时 $A = X\Lambda X^{-1}$ 使得特征值在Λ的对角线上：

马尔可夫范例 $$\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & .5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix} = X\Lambda X^{-1}$$

特征向量(.6, .4)与(1, −1)在 X 的列，他们也是 A^2 的特征向量。观察 A^2 如何有相同的 X，且 A^2 的**特征值矩阵**是Λ^2：

A^2 有相同的 X $$A^2 = X\Lambda X^{-1}\, X\Lambda X^{-1} = X\Lambda^2 X^{-1} \quad (4)$$

继续前进，你会看到为什么高次方 A^k 趋近一个“稳定状态”：

***A* 的次方** $$A^k = X\Lambda^k X^{-1} = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix} \begin{bmatrix} 1^k & 0 \\ 0 & (.5)^k \end{bmatrix} \begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix}$$

当 k 变大，$(.5)^k$ 变小，它在极限时完全消失，这个极限是 A^∞：

极限 $k \to \infty$ $$A^\infty = \begin{bmatrix} .6 & 1 \\ .4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ .4 & -.6 \end{bmatrix} = \begin{bmatrix} .6 & .6 \\ .4 & .4 \end{bmatrix}$$

这个极限的两个列是特征向量 $\boldsymbol{x}_1$，我们在第六章的第一页已经看过 A^∞，现在我们知道它来自类似的 $A^{100} = X\Lambda^{100}X^{-1}$。

问题 **何时 $A^k \to$ 零矩阵？** **答案** **所有的 $|\lambda| < 1$**

相似矩阵：相同特征值

假设特征值矩阵Λ固定，当我们改变特征向量矩阵 X，我们得到不同矩阵 $A = X\Lambda X^{-1}$ 构成的一整个家族——全部矩阵都有与Λ相同的特征值。所有的这些矩阵 A(对应相同的Λ)称为：“**相似(similar)**”。

这个概念可以延伸至不能对角化的矩阵，我们再次选择一个常数矩阵(不需要是Λ)，我们观察整个家族矩阵 $\boldsymbol{A = BCB^{-1}}$，允许所有的可逆矩阵 B，这些矩阵 A 与 C 也是**相似**。

请尊重版权与译者的劳动成果，侵权必究！

我们没有使用Λ而使用 C 是因为 C 可能不是对角，我们没有使用 X 而使用 B 是因为 B 的列可能不是特征向量。我们只需要 B 的可逆性质--它的列包含 $\mathbf{R}^n$ 的任意一组基底。相似矩阵的关键事实仍然成立。**相似矩阵 A 与 C 有相同特征值。**

所有矩阵 $A = BCB^{-1}$ 是“相似”，他们全部共享 C 的特征值。

证明 假设 $Cx = \lambda x$，则 BCB^{-1} 有相同的特征值对应新的特征向量 Bx：

相同λ $$(BCB^{-1})(Bx) = BCx = B\lambda x = \lambda(Bx) \tag{5}$$

一个固定的矩阵 C 产生一个家族的相似矩阵 BCB^{-1}，允许所有的 B。当 C 是单位矩阵，这个家族非常小，唯一的成员是 $BIB^{-1} = I$。单位矩阵是特征值$\lambda = 1$ 唯一可以对角化的矩阵。

当$\lambda = 1$ 与 1 而且只对应一个特征向量(不可对角化)，这个家族变大一点。最简单的 C 是**若尔当形式**(Jordan form)--段落 8.3 会推导。所有的相似 A's 有两个参数 r 与 s，不同时为零：永远是行列式= 1 与迹= 2。

$$C = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} \end{bmatrix} = \text{若尔当形式得到 } A = BCB^{-1} = \begin{bmatrix} 1-rs & r^2 \\ -s^2 & 1+rs \end{bmatrix} \tag{6}$$

我会在一个重要例子中选择特征值$\lambda = 1$ 与 0(不重复!)，现在整个家族有相同的特征值矩阵Λ，而且可以对角化。我们取得每一个有特征值 1 与 0 的 2×2 矩阵，迹是 1 且行列式是零：

所有的相似 $$\Lambda = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{0} \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \text{ 或 } A = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} \text{ 或任意 } A = \frac{\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}$$

这个家族包含所有 $A^2 = A$ 的矩阵，包含当 $B = I$ 时 $A = \Lambda$。当 A 是对称时，这些也是投影矩阵，特征值 1 与 0 使得生活容易多了。

斐波那契数字(Fibonacci number)

我们演示一个著名的例子，其中特征值会告诉我们斐波那契数字如何的快速增长。**每个新的斐波那契数字是前两个斐波那契数字的总和。**

序列 **0, 1, 1, 2, 3, 5, 8, 13, ...** **来自** $\boldsymbol{F_{k+2} = F_{k+1} + F_k}$

这些数字以神奇多样的方式出现，植物与树木以螺旋模式成长，一棵梨树每 3 转有 8 个成长。如果是柳树，这些数字是 13 与 5。冠军则是 Daniel O'connell 的向日葵，在 144 圈有 233 个种子。这是斐波那契数字的 F_{13} 与 F_{12}，我们的问题比较基本。

问题：找出斐波那契数字 F_{100} 最慢的方法是应用规则 $F_{k+2}=F_{k+1}+F_k$ 一步一步计算，把 $F_6=8$ 加上 $F_7=13$ 得到 $F_8=21$，最终我们来到 F_{100}。线性代数有好方法。

关键从矩阵方程式 $\boldsymbol{u}_{k+1}=A\boldsymbol{u}_k$ 开始，这是向量的“一步”规则，斐波那契数字是纯量的两步规则。两个斐波那契数字放在一个向量来适配这些规则，你会看到矩阵 A。

$$\text{令 } \boldsymbol{u}_k=\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix},\ \text{规则}\quad \begin{matrix}F_{k+2}=F_{k+1}+F_k\\F_{k+1}=F_{k+1}\end{matrix}\quad \text{是}\quad \boldsymbol{u}_{k+1}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\boldsymbol{u}_k \tag{7}$$

每个步骤左乘 $A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$，100 步之后我们得到 $\boldsymbol{u}_{100}=A^{100}\boldsymbol{u}_0$：

$$\boldsymbol{u}_0=\begin{bmatrix}1\\0\end{bmatrix},\quad \boldsymbol{u}_1=\begin{bmatrix}1\\1\end{bmatrix},\quad \boldsymbol{u}_2=\begin{bmatrix}2\\1\end{bmatrix},\quad \boldsymbol{u}_3=\begin{bmatrix}3\\2\end{bmatrix},\ \ldots,\ \boldsymbol{u}_{100}=\begin{bmatrix}F_{101}\\F_{100}\end{bmatrix}$$

这个问题就是特征值问题，从 A 的对角线减去 λ：

$$A-\lambda I=\begin{bmatrix}1-\lambda&1\\1&-\lambda\end{bmatrix}\ \text{得到}\quad \det(A-\lambda I)=\lambda^2-\lambda-1$$

求解二次公式得到方程式 $\lambda^2-\lambda-1=0$ 的解是 $\left(-b\pm\sqrt{b^2-4ac}\right)/2a$：

$$\textbf{特征值}\quad \lambda_1=\frac{1+\sqrt{5}}{2}\approx 1.618\quad \text{与}\quad \lambda_2=\frac{1-\sqrt{5}}{2}\approx -.618$$

这些特征值得到特征向量 $\boldsymbol{x}_1=(\lambda_1,1)$ 与 $\boldsymbol{x}_2=(\lambda_2,1)$。步骤 2 找出这些特征向量的组合得到 $\boldsymbol{u}_0=(1,0)$：

$$\begin{bmatrix}1\\0\end{bmatrix}=\frac{1}{\lambda_1-\lambda_2}\left(\begin{bmatrix}\lambda_1\\1\end{bmatrix}-\begin{bmatrix}\lambda_2\\1\end{bmatrix}\right)\quad \text{或}\quad \boldsymbol{u}_0=\frac{\boldsymbol{x}_1-\boldsymbol{x}_2}{\lambda_1-\lambda_2} \tag{8}$$

步骤 3 把 A^{100} 乘 $\boldsymbol{u}_0$ 得到 $\boldsymbol{u}_{100}$。特征向量 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 还是分开，分别乘 $(\lambda_1)^{100}$ 与 $(\lambda_2)^{100}$：

$$\textbf{来自 } \boldsymbol{u}_0 \textbf{ 的第 100 步}\qquad \boldsymbol{u}_{100}=\frac{(\lambda_1)^{100}\boldsymbol{x}_1-(\lambda_2)^{100}\boldsymbol{x}_2}{\lambda_1-\lambda_2} \tag{9}$$

我们要的 F_{100} 是 $\boldsymbol{u}_{100}$ 的第二分量，$\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 的第二分量是 1，$\lambda_1=(1+\sqrt{5})/2$ 与 $\lambda_2=(1-\sqrt{5})/2$ 之间的差是 $\sqrt{5}$，且 $\lambda_2^{100}\approx 0$。

$$\text{第 100 个斐波那契数字}\ =\ \frac{(\lambda_1)^{100}-(\lambda_2)^{100}}{\lambda_1-\lambda_2}=\text{最接近}\ \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{100}\ \text{的整数} \tag{10}$$

每个 F_k 都是一整个数字，F_{101}/F_{100} 的比值必须非常接近极限比值 $(1+\sqrt{5})/2$，希腊人称此为“**黄金分割(golden mean)**”。基于某些原因，矩形的边是 1.618 与 1 看起来特别优雅。

请尊重版权与译者的劳动成果，侵权必究！

矩阵次方 A^k

斐波那契的例子是典型的差分方程式 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$，每个步骤左乘 A，解是 $\boldsymbol{u}_k = A^k\boldsymbol{u}_0$。我们想要弄清楚对角化矩阵如何得到快速计算 A^k 与 $\boldsymbol{u}_k$ 的方法，而且是在三个步骤之内完成。

特征向量矩阵 X 得到 $A = X\Lambda X^{-1}$，这是矩阵的一种分解，好像 $A = LU$ 或是 $A = QR$。这个新分解完美的适配计算幂次方，因为每次 X^{-1} 乘 X 我们得到 I：

A 的次方 $\quad A^k\boldsymbol{u}_0 = (X\Lambda X^{-1})\ \ldots\ (X\Lambda X^{-1})\boldsymbol{u}_0 = X\Lambda^k X^{-1}\boldsymbol{u}_0$

我会把 $X\Lambda^k X^{-1}\boldsymbol{u}_0$ 分割成三个步骤，展示特征值是如何工作的：

1. 把 $\boldsymbol{u}_0$ 写成特征向量的组合 $c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$，则 $\boldsymbol{c} = X^{-1}\boldsymbol{u}_0$。
2. 把每个特征向量 $\boldsymbol{x}_i$ 左乘 $(\lambda_i)^k$，现在我们有 $\Lambda^k X^{-1}\boldsymbol{u}_0$
3. 累加所有的片段 $c_i(\lambda_i)^k\boldsymbol{x}_i$，求出解 $\boldsymbol{u}_k = A^k\boldsymbol{u}_0$。这是 $X\Lambda^k X^{-1}\boldsymbol{u}_0$

$\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$ 的解 $\quad \boldsymbol{u}_k = A^k\boldsymbol{u}_0 = c_1(\lambda_1)^k\boldsymbol{x}_1 + \ldots + c_n(\lambda_n)^k\boldsymbol{x}_n$ $\qquad (11)$

以矩阵语言来说，A^k 等于 $(X\Lambda X^{-1})^k$，也就是 X 乘 Λ^k 乘 X^{-1}。在步骤 1，X 的特征向量得到组合 $\boldsymbol{u}_0 = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$ 的 c's：

步骤 1 $\quad \boldsymbol{u}_0 = \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 表示 $\boldsymbol{u}_0 = X\boldsymbol{c}$ $\qquad (12)$

步骤 1 的系数是 $\boldsymbol{c} = X^{-1}\boldsymbol{u}_0$。然后步骤 2 左乘 Λ^k，步骤 3 的最终结果 $\boldsymbol{u}_k = \Sigma\, c_i(\lambda_i)^k\boldsymbol{x}_i$ 是 X 与 Λ^k 与 $X^{-1}\boldsymbol{u}_0$ 的乘积：

$$A^k\boldsymbol{u}_0 = X\Lambda^k X^{-1}\boldsymbol{u}_0 = X\Lambda^k\boldsymbol{c} = \begin{bmatrix} \boldsymbol{x}_1 & \ldots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} (\lambda_1)^k & & \\ & \ddots & \\ & & (\lambda_1)^k \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \qquad (13)$$

这个结果恰好是 $\boldsymbol{u}_k = A^k\boldsymbol{u}_0 = c_1(\lambda_1)^k\boldsymbol{x}_1 + \ldots + c_n(\lambda_n)^k\boldsymbol{x}_n$，是 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$ 的解。

范例 3 从 $\boldsymbol{u}_0 = (1, 0)$开始，计算快速斐波那契数字的 $A^k\boldsymbol{u}_0$：

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \text{ 有 } \lambda_1 = 2 \text{ 与 } \boldsymbol{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \lambda_2 = -1 \text{ 与 } \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

除了规则改成 $F_{k+2} = F_{k+1} + 2F_k$ 之外，这个矩阵很像斐波那契数字，新的数字从 0, 1, 1, 3 开始。因为 $\lambda = 2$，所以他们成长的比较快速。

三个步骤求出 $\boldsymbol{u_k = A^k u_0}$　　$\boldsymbol{u_0} = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2$ 且 $\boldsymbol{u_k} = A^k\boldsymbol{u_0} = c_1(\lambda_1)^k\,\boldsymbol{x}_1 + c_2(\lambda_2)^k\,\boldsymbol{x}_2$

步骤 1　　$\boldsymbol{u}_0 = \begin{bmatrix}1\\0\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\-1\end{bmatrix}$　所以　$c_1 = c_2 = 1/3$

步骤 2　　分别左乘 $(\lambda_1)^k = 2^k$ 与 $(\lambda_2)^k = (-1)^k$

步骤 3　　组合特征向量 $c_1(\lambda_1)^k\boldsymbol{x}_1$ 与 $c_2(\lambda_2)^k\boldsymbol{x}_2$ 进入 $\boldsymbol{u_k}$ ：

$$\boldsymbol{u_k = A^k u_0} \qquad \boldsymbol{u_k} = \frac{1}{3}2^k\begin{bmatrix}2\\1\end{bmatrix} + \frac{1}{3}(-1)^k\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}$$

新的数字是 $\boldsymbol{F_k = (2^k - (-1)^k)/3}$，在 0, 1, 1, 3 之后是 $F_4 = 15/3 = 5$。

在这些数字范例背后有一个基础概念：**追随特征向量**。在段落 6.3，这是从线性代数到微分方程式(λ^k 变成$e^{\lambda t}$)的重要连接。第八章会看到“转换至特征向量基底”的相同概念，其中的最佳范例是“**傅里叶级数**(Fourier series)”，创建自 d/dx 的特征向量 e^{ikx}。

不可对角化矩阵(选读)

假设λ是 A 的特征值，我们从两个方面发现这个事实：

1. **特征向量**(几何的)　$A\boldsymbol{x} = \lambda\boldsymbol{x}$ 有非零解。
2. **特征值**(代数的)　$A - \lambda I$ 的行列式是零。

数字λ可能是单一特征值或是多重特征值，我们想要知道它的**倍数值**(multiplicity)。大部分特征值的倍数值是 $M = 1$(单一特征值)，有一条特征向量直线，$\det(A - \lambda I)$ 没有双重因子(double factor)。

有一些例外的矩阵，特征值可以重复，有两个不同的方式计算它的倍数值，对于每个λ，永远是 GM ≤ AM：

1. **(几何倍数值= GM)**　GM 计算λ对应的无关特征向量的个数，GM 是 $A - \lambda I$ 的零空间的维度。
2. **(代数倍数值= AM)**　AM 计算λ在特征值的重复次数，检视 $\det(A - \lambda I) = 0$ 的 n 个根。

若 A 有$\lambda = 4, 4, 4$，则这个特征值有 AM = 3 与 GM = 1, 2 或 3。

下列的矩阵 A 是标准的麻烦范例，它的特征值有重复的 0，它是一个双重特征值(AM = 2)，却只有一个特征向量(GM = 1)。

$$\begin{matrix}\textbf{AM} = 2\\ \textbf{GM} = 1\end{matrix} \qquad A = \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} \text{ 有 } \det A = \begin{vmatrix}-\lambda & 1\\0 & -\lambda\end{vmatrix} = \lambda^2 \qquad \begin{matrix}\boldsymbol{\lambda = 0,\ 0}\text{ 但是}\\ \textbf{1}\text{ 个固有向量}\end{matrix}$$

请尊重版权与译者的劳动成果，侵权必究!

因为$\lambda^2=0$有双重根，理论上应该有两个特征向量，双重因子λ^2使得 AM = 2，但是只有一个特征向量$\boldsymbol{x}=(1,0)$与 GM = 1。当 GM 小于 AM 时，特征向量的不足使得 A 不能对角化。

下列三个矩阵的特征向量都不足，重根是$\lambda=5$，迹是 10，行列式是 25：

$$A=\begin{bmatrix}5&1\\0&5\end{bmatrix} \quad 与 \quad A=\begin{bmatrix}6&-1\\1&4\end{bmatrix} \quad 与 \quad A=\begin{bmatrix}7&2\\-2&3\end{bmatrix}$$

他们都有 $\det(A-\lambda I)=(\lambda-5)^2$，代数倍数值是 AM = 2。但是每个 $A-5I$ 的秩是 1，几何倍数值是 GM = 1，对应$\lambda=5$只有一个特征向量直线，这个三个矩阵都是不可对角化。

主要观念的复习

1. 若 A 有 n 个无关的特征向量 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$，他们进入 X 的列。

   A 被 X 对角化 $\quad X^{-1}AX=\Lambda \quad 与 \quad A=X\Lambda X^{-1}$

2. A 的次方是 $A^k=X\Lambda^kX^{-1}$，X 的特征向量保持不变。

3. A^k 的特征值是矩阵Λ^k中的$(\lambda_1)^k,\ldots,(\lambda_n)^k$。

4. 从 $\boldsymbol{u}_0$ 开始，$\boldsymbol{u}_{k+1}=A\boldsymbol{u}_k$ 的解是 $\boldsymbol{u}_k=A^k\boldsymbol{u}_0=X\Lambda^kX^{-1}\boldsymbol{u}_0$：

   假设 $\boldsymbol{u}_0=c_1\boldsymbol{x}_1+\ldots+c_n\boldsymbol{x}_n$ 则 $\boldsymbol{u}_k=c_1(\lambda_1)^k\boldsymbol{x}_1+\ldots+c_n(\lambda_n)^k\boldsymbol{x}_n$

   这些是步骤 1, 2, 3 (c's 来自 $X^{-1}\boldsymbol{u}_0$，λ^k 来自Λ^k，$\boldsymbol{x}$'s 来自 X)

5. 如果每个特征值都足够的特征向量(GM = AM)，则 A 可以对角化。

已解范例

6.2A 除了从 $L_1=1$ 与 $L_2=3$ 开始之外，**卢卡斯数字**很像斐波那契数字。使用相同的规则 $L_{k+2}=L_{k+1}+L_k$，后面卢卡斯数字是 4, 7, 11, 18。证明卢卡斯数字 L_{100} 是 $\lambda_1^{100}+\lambda_2^{100}$。

解 $\boldsymbol{u}_{k+1}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\boldsymbol{u}_k$ 与斐波那契相同，因为 $L_{k+2}=L_{k+1}+L_k$ 是相同的规则（只是不同的起始值），方程式变成 2×2 系统：

令 $\boldsymbol{u}_k=\begin{bmatrix}L_{k+1}\\L_k\end{bmatrix}$，规则 $\begin{matrix}L_{k+2}=L_{k+1}+L_k\\L_{k+1}=L_{k+1}\end{matrix}$ 是 $\boldsymbol{u}_{k+1}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\boldsymbol{u}_k$

$A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$ 的特征值与特征向量仍然来自 $\lambda^2=\lambda+1$：

$$\lambda_1=\frac{1+\sqrt{5}}{2}\ \text{且}\ \boldsymbol{x}_1=\begin{bmatrix}\lambda_1\\1\end{bmatrix}\qquad \lambda_2=\frac{1-\sqrt{5}}{2}\ \text{且}\ \boldsymbol{x}_2=\begin{bmatrix}\lambda_2\\1\end{bmatrix}$$

现在求解 $c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2=\boldsymbol{u}_1=(3,1)$，解是 $c_1=\lambda_1$ 与 $c_2=\lambda_2$。检查：

$$\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2=\begin{bmatrix}\lambda_1^2+\lambda_2^2\\\lambda_1+\lambda_2\end{bmatrix}=\begin{bmatrix}A^2\text{的迹}\\A\text{的迹}\end{bmatrix}=\begin{bmatrix}1\\3\end{bmatrix}=\boldsymbol{u}_1$$

$\boldsymbol{u}_{100}=A^{99}\boldsymbol{u}_1$ 告知我们卢卡斯数字(L_{101},L_{100})。$\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 的第二分量是 1，所以 $\boldsymbol{u}_{100}$ 的第二分量就是我们想要的答案：

卢卡斯数字 $\boldsymbol{L_{100}=c_1\lambda_1^{99}+c_2\lambda_2^{99}=\lambda_1^{100}+\lambda_2^{100}}$

卢卡斯开始比斐波那契快，最终大约比斐波那契大$\sqrt{5}$倍。

6.2B 求出矩阵 A 的逆矩阵与特征值以及行列式：

$$A=5*\text{eye}(4)-\text{ones}(4)=\begin{bmatrix}4&-1&-1&-1\\-1&4&-1&-1\\-1&-1&4&-1\\-1&-1&-1&4\end{bmatrix}$$

描述一个特征向量矩阵 X 使得 $X^{-1}AX=\Lambda$。

解 全 1 矩阵的特征值为何？它的秩当然是 1，所以 3 个特征值是$\lambda=0,0,0$。它的迹是 4，所以另一个特征值是$\lambda=4$。从 $5I$ 减去全 1 矩阵得到我们的矩阵 A：

从 5, 5, 5, 5 减去特征值 4, 0, 0, 0，得到 A 的特征值是 1, 5, 5, 5

A 的行列式是 125，是 4 个特征值的乘积。对应$\lambda=1$ 的特征向量是 $\boldsymbol{x}=(1,1,1,1)$ 或(c,c,c,c)，另一个特征向量与 $\boldsymbol{x}$ 垂直(因为 A 对称)，最棒的特征向量矩阵 X 是对称正交**哈达玛矩阵 $\boldsymbol{H}$**。因子 1/2 得到单位列向量。

正交单位特征向量 $X=H=\frac{1}{2}\begin{bmatrix}1&1&1&1\\1&-1&1&-1\\1&1&-1&-1\\1&-1&-1&1\end{bmatrix}=H^{\mathrm{T}}=H^{-1}$

请尊重版权与译者的劳动成果，侵权必究！

逆矩阵 A^{-1} 的特征值是 1, 1/5, 1/5, 1/5, 特征向量不变所以 $A^{-1} = H\Lambda H^{-1}$，逆矩阵是令人意外的整洁有序：

$$A^{-1} = (\text{eye}(4) + \text{ones}(4))/5 = \frac{1}{5}\begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

A 是从 $5I$ 变化而来的秩一矩阵，所以 A^{-1} 是从 $I/5$ 变化而来的秩一矩阵。

在具有 5 个节点的图形中，行列式 125 是“生成树(spanning tree)”(接触所有节点的树)的个数，树没有回圈(图形与树在段落 10.1)。

如果是 6 个节点，矩阵 6*eye(5) – ones(5)有 5 个特征值 1, 6, 6, 6, 6。

问题集 6.2

问题 1-7 有关特征值矩阵与特征向量矩阵Λ与 X。

1 (a) 分解下列矩阵变成 $A = X\Lambda X^{-1}$：

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \qquad \text{与} \qquad A = \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$$

(b) 若 $A = X\Lambda X^{-1}$，则 $A^3 = (\quad)(\quad)(\quad)$且 $A^{-1} = (\quad)(\quad)(\quad)$。

2 若 A 有 $\lambda_1 = 2$ 对应特征向量 $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$，与 $\lambda_2 = 5$ 对应特征向量 $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$，使用 $X\Lambda X^{-1}$ 求 A。没有其他矩阵有相同的 λ's 与 $\boldsymbol{x}$'s。

3 假设 $A = X\Lambda X^{-1}$，则 $A + 2I$ 的特征值矩阵为何？特征向量矩阵为何？检查 $A + 2I = (\quad)(\quad)(\quad)^{-1}$。

4 是非题：若 X 的列(A 的特征向量)是线性无关，则

(a) A 可逆　　(b) A 可对角化

(c) X 可逆　　(d) X 可对角化

5 若 A 的特征向量是 I 的列，则 A 是一个_____矩阵。若特征向量矩阵 X 是三角形，则 X^{-1} 是三角形，证明 A 也是三角形。

6 描述能够对角化矩阵 A 的所有矩阵 X(求出所有特征向量)：

$$A = \begin{bmatrix} 4 & 0 \\ 1 & 2 \end{bmatrix}$$

然后描述能够对角化 A^{-1} 的所有矩阵。

7 写下具有特征向量 $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ 与 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 的最平常的矩阵。

问题 8-10 有关斐波那契与 Gibonaccis 数字。

8　计算 X^{-1} 然后对角化斐波那契矩阵：

$$\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}=\begin{bmatrix}\lambda_1 & \lambda_2\\ 1 & 1\end{bmatrix}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix} & \\ & \end{bmatrix}$$

计算 $X\Lambda^k X^{-1}\begin{bmatrix}1\\ 0\end{bmatrix}$ 找出第二分量，这是第 k 个斐波那契数字 $F_k=(\lambda_1{}^k-\lambda_2{}^k)/(\lambda_1-\lambda_2)$。

9　假设 G_{k+2} 是前两个数字 G_{k+1} 与 G_k 的平均：

$$\begin{aligned}G_{k+2}&=\frac{1}{2}G_{k+1}+\frac{1}{2}G_k\\ G_{k+1}&=G_{k+1}\end{aligned}\quad \text{是}\quad \begin{bmatrix}G_{k+2}\\ G_{k+1}\end{bmatrix}=\begin{bmatrix}A\end{bmatrix}\begin{bmatrix}G_{k+1}\\ G_k\end{bmatrix}$$

(a) 求出 A 的特征值与特征向量。

(b) 求出 $A^n=X\Lambda^n X^{-1}$ 在 $n\to\infty$ 的极限。

(c) 若 $G_0=0$ 与 $G_1=1$，证明 Gibonaccis 数字趋近 2/3。

10　证明斐波那契数字 0, 1, 1, 2, 3, ...，每三个数字一定是偶数。

问题 11-14 有关可对角化。

11　是非题：若 A 的特征值是 2, 2, 5，则矩阵必然是

(a) 可逆　(b) 可对角化　(c) 不可对角化

12　是非题：若 A 的特征向量只有(1, 4)的倍数，则 A

(a) 不可逆　(b) 重复的特征值　(c) 不可对角化，没有 $X\Lambda X^{-1}$。

13　完成下列矩阵使得 $\det A=25$，检查$\lambda=5$ 是重复特征值--迹是 10 所以 $A-\lambda I$ 的行列式是$(\lambda-5)^2$。找出 $A\boldsymbol{x}=5\boldsymbol{x}$ 的特征向量。这些矩阵不可对角化，因为没有第二条特征向量直线。

$$A=\begin{bmatrix}8 & \\ & 2\end{bmatrix}\quad 与\quad A=\begin{bmatrix}9 & 4\\ & 1\end{bmatrix}\quad 与\quad A=\begin{bmatrix}10 & 5\\ -5 & \end{bmatrix}$$

14　矩阵 $A=\begin{bmatrix}3 & 1\\ 0 & 3\end{bmatrix}$ 不可对角化，因为 $A-3I$ 的秩是____。改变一个单元使得 A 可以对角化。你会改变哪个单元?

问题 15-19 有关矩阵的次方。

15　$A^k=X\Lambda^k X^{-1}$ 在 $k\to\infty$ 趋近零矩阵，当且仅当每个λ的绝对值比____小。下列矩阵哪个有 $A^k\to 0$?

$$A_1=\begin{bmatrix}.6 & .9\\ .4 & .1\end{bmatrix}\quad 与\quad A_2=\begin{bmatrix}.6 & .9\\ .1 & .6\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

16 (推荐) 求Λ与 X 对角化问题 15 的 A_1。Λ^k 在 $k \to \infty$ 的极限为何？$X\Lambda^k X^{-1}$ 的极限为何？在这个极限矩阵的列，你看到了什么？

17 求Λ与 X 对角化问题 15 的 A_2。$(A_2)^{10}\boldsymbol{u}_0$ 为何？$\boldsymbol{u}_0$ 如下：

$$\boldsymbol{u}_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad 与 \quad \boldsymbol{u}_0 = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad 与 \quad \boldsymbol{u}_0 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

18 对角化 A 并且计算 $X\Lambda^k X^{-1}$，证明 A^k 的公式：

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ 有 } A^k = \frac{1}{2}\begin{bmatrix} 1+3^k & 1-3^k \\ 1-3^k & 1+3^k \end{bmatrix}$$

19 对角化 B 并且计算 $X\Lambda^k X^{-1}$，证明 B^k 的公式：

$$B = \begin{bmatrix} 5 & 1 \\ 0 & 4 \end{bmatrix} \text{ 有 } B^k = \begin{bmatrix} 5^k & 5^k - 4^k \\ 0 & 4^k \end{bmatrix}$$

20 假设 $A = X\Lambda X^{-1}$，取行列式证明 $\det A = \det \Lambda = \lambda_1\lambda_2 \ldots \lambda_n$。这个快速证明只能在 A 是_____才有效。

21 把 XY 的对角线单元加到 YX 的对角线单元，证明 XY 的迹等于 YX 的迹：

$$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ 与 } Y = \begin{bmatrix} q & r \\ s & t \end{bmatrix}$$

现在选择 Y 等于ΛX^{-1}，则 $X\Lambda X^{-1}$ 与$\Lambda X^{-1}X = \Lambda$有相同的迹，这就证明了 A 的迹 = Λ的迹 = 特征值的总和。

22 不可能存在 $AB - BA = I$，因为左侧的迹 = _____，但是找出一个消元矩阵使得 $A = E$ 与 $B = E^{\mathrm{T}}$ 得到

$$AB - BA = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}，\text{迹等于零。}$$

23 假设 $A = X\Lambda X^{-1}$，对角化方块矩阵 $B = \begin{bmatrix} A & 0 \\ 0 & 2A \end{bmatrix}$，求出特征值与特征向量(方块)矩阵。

24 考虑可以被相同固定的特征向量矩阵 X 对角化的所有 4×4 矩阵 A，证明 A's 形成一个子空间(cA 与 $A_1 + A_2$ 有相同的 X)。当 $X = I$，这个子空间为何？维度是多少？

25 假设 $A^2 = A$，在左侧是 A 乘 A 的每个列，我们 4 个子空间的哪一个包含对应$\lambda = 1$ 的特征向量？哪个子空间包含对应$\lambda = 0$ 的特征向量？从这些子空间的维度来看，A 有一整组的无关特征向量，所以 $A^2 = A$ 的矩阵可以对角化。

26 (推荐) 假设 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，若$\lambda=0$ 则 $\boldsymbol{x}$ 在零空间中。若$\lambda\neq 0$ 则 $\boldsymbol{x}$ 在列空间中。这些空间有维度$(n-r)+r=n$，所以为什么不是每个方形矩阵都有 n 个线性无关的特征向量？

27 A 的特征值是 1 与 9，B 的特征值是–1 与 9：

$$A=\begin{bmatrix}5 & 4\\ 4 & 5\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}4 & 5\\ 5 & 4\end{bmatrix}$$

从 $R=X\sqrt{\Lambda}\,X^{-1}$，求出 A 的平方根矩阵。为什么 B 没有实数的平方根矩阵？

28 若 A 与 B 有相同的λ's 与相同的特征向量，他们分解成_____是相同的，所以 $A=B$。

29 假设相同的 X 同时对角化 A 与 B，在 $A=X\Lambda_1X^{-1}$ 与 $B=X\Lambda_2X^{-1}$ 有相同的特征向量。证明 $AB=BA$。

30 (a) 若 $A=\begin{bmatrix}a & b\\ 0 & d\end{bmatrix}$，则 $A-\lambda I$ 的行列式是$(\lambda-a)(\lambda-d)$。检查“凯利-汉密尔顿定理”$(A-aI)(A-dI)$ =零矩阵。

(b) 对斐波那契的 $A=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$测试凯利-汉密尔顿定理，定理预测 $A^2-A-I=0$，因为多项式 $\det(A-\lambda I)$是$\lambda^2-\lambda-1$。

31 把 $A=X\Lambda_1X^{-1}$ 代入乘积$(A-\lambda_1I)(A-\lambda_2I)\ldots(A-\lambda_nI)$，并且说明为什么这个乘积是零矩阵。我们是把矩阵 A 代入多项式 $p(\lambda)=\det(A-\lambda I)$的数字$\lambda$，**凯利-汉密尔顿**定理指出这个乘积永远是 $p(A)=$ 零矩阵，纵使 A 不可对角化。

32 若 $A=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$且 $AB=BA$，证明 $B=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$也是对角矩阵，$B$ 与 A 有相同的特征_____，但是有不同的特征_____。这些对角矩阵 B 形成一个二维矩阵空间的子空间，$AB-BA=0$ 得到 4 个有关 a, b, c, d 的方程式--求出这个 4×4 矩阵的秩。

33 如果所有的$|\lambda_i|<1$，次方 A^k 趋近零；如果任意的$|\lambda_i|>1$，次方 A^k 会爆炸。**彼得雷克斯**在他的线性代数的书中给出反例：

$$A=\begin{bmatrix}3 & 2\\ 1 & 4\end{bmatrix} \quad B=\begin{bmatrix}3 & 2\\ -5 & -3\end{bmatrix} \quad C=\begin{bmatrix}5 & 7\\ -3 & -4\end{bmatrix} \quad D=\begin{bmatrix}5 & 6.9\\ -3 & -4\end{bmatrix}$$

$$\|\boldsymbol{A}^{1024}\|>\boldsymbol{10}^{700} \qquad \boldsymbol{B}^{1024}=\boldsymbol{I} \qquad \boldsymbol{C}^{1024}=-\boldsymbol{C} \qquad \|\boldsymbol{D}^{1024}\|<\boldsymbol{10}^{-78}$$

求出 B 与 C 的特征值$\lambda=e^{i\theta}$，证明 $B^4=I$ 与 $C^3=-I$。

请尊重版权与译者的劳动成果，侵权必究!

挑战问题

34 旋转θ角的 n 次方相当于旋转 $n\theta$:

$$A^n=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}^n=\begin{bmatrix}\cos n\theta & -\sin n\theta\\ \sin n\theta & \cos n\theta\end{bmatrix}$$

利用对角化 $A=X\Lambda X^{-1}$ 证明这个简洁公式。特征向量(X 的列)是$(1, i)$与$(i, 1)$。你需要知道欧拉公式 $e^{i\theta}=\cos\theta+i\sin\theta$ 。

35 $A=X\Lambda X^{-1}$ 的转置是 $A^{\mathrm{T}}=(X^{-1})^{\mathrm{T}}\Lambda X^{\mathrm{T}}$，$A^{\mathrm{T}}\boldsymbol{y}=\lambda\boldsymbol{y}$ 的特征向量是$(X^{-1})^{\mathrm{T}}$的列，他们通常称为 A 的**左特征向量**，因为 $\boldsymbol{y}^{\mathrm{T}}A=\lambda\boldsymbol{y}^{\mathrm{T}}$。要如何矩阵相乘找出 A 的公式？

秩一矩阵的总和 $\quad A=X\Lambda X^{-1}=\lambda_1\boldsymbol{x}_1\boldsymbol{y}_1{}^{\mathrm{T}}+\ldots+\lambda_n\boldsymbol{x}_n\boldsymbol{y}_n{}^{\mathrm{T}}$

36 A = eye(n) + ones(n)的逆矩阵是 A^{-1} = eye(n) + C*ones(n)。执行 AA^{-1} 求出数字 C(与 n 有关)。

37 假设 A_1 与 A_2 是 $n\times n$ 可逆矩阵，什么矩阵使得 $A_2A_1=B(A_1A_2)B^{-1}$？则 A_2A_1 与 A_1A_2 相似：相同的特征值。

38 什么时候矩阵 A 与它的特征值矩阵Λ相似？

A 与Λ永远有相同的特征值，但是相似性需要一个矩阵 B 使得 $A=B\Lambda B^{-1}$，则 B 是_____矩阵且 A 必须有 n 个无关的_____。

39 (Pavel Grinfeld) 不写出任何计算，你能够找出下列矩阵的特征值吗？你能够找出下列矩阵的第 2017 次方 A^{2017} 吗？

$$A^n=\begin{bmatrix}110 & 55 & -164\\ 42 & 21 & -62\\ 88 & 44 & -131\end{bmatrix}$$

若 A 是 $m\times n$ 且 B 是 $n\times m$，则 AB 与 BA 有相同的非零特征值。

证明 从两个方形矩阵的等式开始(容易检验)，第一矩阵是第三矩阵的逆矩阵。“大小矩阵”显示所有方块的形状。

$$\begin{bmatrix}I & -A\\ 0 & I\end{bmatrix}\begin{bmatrix}AB & 0\\ B & 0\end{bmatrix}\begin{bmatrix}I & A\\ 0 & I\end{bmatrix}=\begin{bmatrix}0 & 0\\ B & BA\end{bmatrix}\qquad\begin{bmatrix}m\times m & m\times n\\ n\times m & n\times n\end{bmatrix}$$

方程式 $D^{-1}ED=F$ 说明 ***F* 与 *E* 相似**—他们有相同的 $m+n$ 个特征值。

$E=\begin{bmatrix}AB & 0\\ B & 0\end{bmatrix}$ 有 AB 的 m 个特征值，加上 n 个零

$F=\begin{bmatrix}0 & 0\\ B & BA\end{bmatrix}$ 有 BA 的 n 个特征值，加上 m 个零

所以除了$|n-m|$个零之外，AB 与 BA 有**相同的特征值**。哇！

若 $A=[1, 1]$且 $B=A^{\mathrm{T}}$，则 $A^{\mathrm{T}}A=\begin{bmatrix}\mathbf{1} & \mathbf{1}\\ \mathbf{1} & \mathbf{1}\end{bmatrix}$ (注意$\lambda=2$ 与 0) 以及 $AA^{\mathrm{T}}=[\,2\,]$。

6.3 微分方程式系统

1 若 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 则 $\boldsymbol{u}(t)=e^{\lambda t}\boldsymbol{x}$ 是 $\dfrac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}$ 的解，每个λ与 $\boldsymbol{x}$ 得到一个解$e^{\lambda t}\boldsymbol{x}$。

2 若 $A=X\Lambda X^{-1}$ 则 $\boxed{\boldsymbol{u}(t)=\ e^{At}\boldsymbol{u}(0)=Xe^{\Lambda t}X^{-1}\boldsymbol{u}(0)=\ c_1e^{\lambda_1 t}\boldsymbol{x}_1+\ldots+c_ne^{\lambda_n t}\boldsymbol{x}_n}$。

3 当 A 的所有特征值的实部 <0，则 A 是**稳定**且 $\boldsymbol{u}(t)\to\boldsymbol{0}$ 以及$e^{At}\ \to 0$。

4 若 A 可以对角化，则**矩阵指数** $e^{At}=I+At+\ldots+(At)^n/n!+\ldots=Xe^{\Lambda t}X^{-1}$

5 **二阶方程式 一阶方程式系统** $u''+Bu'+Cu=0$ 等效于 $\begin{bmatrix}u\\u'\end{bmatrix}'=\begin{bmatrix}0&1\\-C&-B\end{bmatrix}\begin{bmatrix}u\\u'\end{bmatrix}$。

特征值与特征向量与 $A=X\Lambda X^{-1}$ 对于矩阵次方 A^k 非常完美，他们对于微分方程式 $\dfrac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}$ 一样完美。本段落大部分是线性代数，但是你需要一个微积分的事实才能研读：$e^{\lambda t}$**的导数是**$\lambda e^{\lambda t}$。本章的重点在于：**把常数系数微分方程式转换成线性代数。**

常微分方程式 $\dfrac{du}{dt}=u$ 与 $\dfrac{du}{dt}=\lambda u$ 的解是指数形式：

$$\frac{du}{dt}=u\text{ 得到 }u(t)=\boldsymbol{C}e^{t}\qquad\frac{du}{dt}=\lambda u\text{ 得到 }u(t)=Ce^{\lambda t}\tag{1}$$

在 $t=0$ 这些解包含 $e^0=1$，所以他们简化成 $u(0)=C$，这个“初始值”告诉我们 C 的正确选择。在 $t=0$ 的数字 $u(0)$开始的解是 $u(t)=u(0)e^{t}$ 与 $u(t)=u(0)e^{\lambda t}$ 。

我们只是求解了 1×1 的问题，线性代数要移到 $n\times n$。未知数是一个向量 $\boldsymbol{u}$(现在是粗体字)，它从给定初始值 $u(0)$开始。这 n 个方程式包含了一个方形矩阵 A，我们期待从 n 个λ's 得到在 $\boldsymbol{u}(t)$的 n 个指数$e^{\lambda t}$：

n 个方程式的系统 $\dfrac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}$ 在 $t=0$ 从向量 $\boldsymbol{u}(0)=\begin{bmatrix}u_1(0)\\\ldots.\\u_n(0)\end{bmatrix}$ 开始 (2)

这些微分方程式是线性，若 $\boldsymbol{u}(t)$与 $\boldsymbol{v}(t)$是解，则 $C\boldsymbol{u}(t)+D\boldsymbol{v}(t)$也是解，我们需要类似 C 与 D 的常数来适配 $\boldsymbol{u}(0)$的 n 个分量。我们第一个工作是利用 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 找出 n 个“**纯指数解**” $\boldsymbol{u}=e^{\lambda t}\boldsymbol{x}$。

注意 A 是常数矩阵，在其他线性方程式 A 随着 t 变化。在非线性方程式中，A 随着 $\boldsymbol{u}$ 变化。我们没有那些困难，$d\boldsymbol{u}/dt=A\boldsymbol{u}$ 是线性而且系数是常数。这样的微分方程式，我们直接转换成线性代数，以下是关键：

当 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，我们使用指数$e^{\lambda t}\boldsymbol{x}$ 求解线性常数系数方程式。

请尊重版权与译者的劳动成果，侵权必究!

我们的单纯指数解是$e^{\lambda t}$乘一个固定向量 $\boldsymbol{x}$，你可以猜测λ是 A 的特征值，以及 $\boldsymbol{x}$ 是特征向量。把 $\boldsymbol{u}(t)=e^{\lambda t}\boldsymbol{x}$ 代入方程式 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 来证明你是正确的。因子$e^{\lambda t}$会被消除得到$\lambda\boldsymbol{x}=A\boldsymbol{x}$：

$$\text{当 } A\boldsymbol{x}=\lambda\boldsymbol{x} \quad \text{选择 } \boldsymbol{u}=e^{\lambda t}\boldsymbol{x} \qquad \frac{d\boldsymbol{u}}{dt}=\lambda e^{\lambda t}\boldsymbol{x} \qquad \text{与} \quad A\boldsymbol{u}=Ae^{\lambda t}\boldsymbol{x} \quad \text{一致} \tag{3}$$

这个特殊解 $\boldsymbol{u}=e^{\lambda t}\boldsymbol{x}$ 的所有分量共享相同的$e^{\lambda t}$，当$\lambda>0$ 解会**成长**；当$\lambda<0$，解会**衰减**。若λ是一个复数，它的**实数部分决定成长或衰减**，虚数部分ω 产生震荡$e^{i\omega t}$，如同正弦波。

范例 1 求解 $\dfrac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\boldsymbol{u}$ 从 $\boldsymbol{u}(0)=\begin{bmatrix}4\\2\end{bmatrix}$开始。

这是一个 $\boldsymbol{u}$ 的向量方程式，它包含分量 y 与 z 的两个纯量方程式。因为矩阵 A 不是对角，他们“耦合”在一起：

$$\frac{d\boldsymbol{u}}{dt}=A\boldsymbol{u} \quad \frac{d}{dt}\begin{bmatrix}y\\z\end{bmatrix}=\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}y\\z\end{bmatrix} \text{ 表示 } \frac{dy}{dt}=z \text{ 与 } \frac{dz}{dt}=y$$

特征向量的概念是以某种方法结合这些方程式，然后反推得到 1×1 的问题。组合 $y+z$ 与 $y-z$ 可以做到这件事情，把两个方程式既加又减：

$$\frac{d}{dt}(y+z)=z+y \quad \text{与} \quad \frac{d}{dt}(y-z)=-(y-z)$$

组合 $y+z$ 如同e^{t}成长，因为它的$\lambda=1$。组合 $y-z$ 如同e^{-t}衰减，因为它的$\lambda=-1$。这是重点：我们没必要去评断原始方程式 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 寻找这些特殊组合，A 的特征向量与特征值会帮我们做好这件事情。

矩阵 A 的特征值是 1 与–1，特征向量 $\boldsymbol{x}$ 是(1, 1)与(1, –1)，纯指数解 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$ 具有 $e^{\lambda t}\boldsymbol{x}$ 的形式，其中$\lambda_1=1$ 与$\lambda_2=-1$：

$$\boldsymbol{u}_1(t)=e^{\lambda_1 t}\boldsymbol{x}_1=e^{t}\begin{bmatrix}1\\1\end{bmatrix} \qquad \text{与} \qquad \boldsymbol{u}_2(t)=e^{\lambda_2 t}\boldsymbol{x}_2=e^{-t}\begin{bmatrix}1\\-1\end{bmatrix} \tag{4}$$

注意这些 $\boldsymbol{u}$'s 满足 $A\boldsymbol{u}_1=\boldsymbol{u}_1$ 与 $A\boldsymbol{u}_2=\boldsymbol{u}_2$，就好像 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 一样。因子e^{t}与e^{-t}随时间变化，这些因子得到 $d\boldsymbol{u}_1/dt=\boldsymbol{u}_1=A\boldsymbol{u}_1$ 与 $d\boldsymbol{u}_2/dt=-\boldsymbol{u}_2=A\boldsymbol{u}_2$，对于 $d\boldsymbol{u}/dt=A\boldsymbol{u}$ 我们有两个解。要求出其他所有的解，把常数 C 与 D 乘上这些特殊解然后相加：

完整解 $$\boldsymbol{u}(t)=Ce^{t}\begin{bmatrix}1\\1\end{bmatrix}+De^{-t}\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}Ce^{t}+De^{-t}\\Ce^{t}-De^{-t}\end{bmatrix} \tag{5}$$

$du/dt = Au$ 的解

常数 C 与 D 可以适配任何的初始向量 $\boldsymbol{u}(0) = (u_1(0), u_2(0))$，令 $t = 0$ 且 $e^0 = 1$，范例 1 要求初始值是 $\boldsymbol{u}(0) = (4, 2)$：

$$\boldsymbol{u}(0)\text{决定 } C \text{ 与 } D \qquad C\begin{bmatrix}1\\1\end{bmatrix} + D\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}4\\2\end{bmatrix} \text{ 得到 } C = 3 \text{ 与 } D = 1$$

方程式(5)的 $C = 3$ 与 $D = 5$ 完全解答了初始值问题。

求解 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$ 的相同三个步骤现在求解 $d\boldsymbol{u}/dt = A\boldsymbol{u}$：

1. 把 $\boldsymbol{u}(0)$写成 A 的特征向量的组合 $c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$。
2. 每个特征向量左乘它的成长因子 $e^{\lambda_i t}$。
3. 方程式的解与单纯解 $e^{\lambda t}\boldsymbol{x}$ 有相同的组合：

$$\frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u} \qquad \boldsymbol{u}(t) = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + \ldots + c_n e^{\lambda_n t}\boldsymbol{x}_n \tag{6}$$

不包含：若两个λ's 相等，而且只有一个特征向量，我们还需要另外一个解(它会是 $te^{\lambda t}\boldsymbol{x}$)。步骤 1 需要对角化 $A = X\Lambda X^{-1}$：n 个特征向量形成一组基底。

范例 2 求解 $\dfrac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$，已知 A 的特征值$\lambda = 1, 2, 3$：

典型范例
$\boldsymbol{u}$的方程式
初始值 $\boldsymbol{u}(0)$

$$\frac{d\boldsymbol{u}}{dt} = \begin{bmatrix}1 & 1 & 1\\0 & 2 & 1\\0 & 0 & 3\end{bmatrix}\boldsymbol{u} \quad \text{初始值 } \boldsymbol{u}(0) = \begin{bmatrix}9\\7\\4\end{bmatrix}$$

特征向量是 $\boldsymbol{x}_1 = (1, 0, 0)$与 $\boldsymbol{x}_2 = (1, 1, 0)$与 $\boldsymbol{x}_3 = (1, 1, 1)$。

步骤 1. 向量 $\boldsymbol{u}(0) = (9, 7, 4)$是 $2\boldsymbol{x}_1 + 3\boldsymbol{x}_2 + 4\boldsymbol{x}_3$，于是$(c_1, c_2, c_3) = (2, 3, 4)$。
步骤 2. 因子 $e^{\lambda t}$ 产生指数解 $e^t\boldsymbol{x}_1$ 与 $e^{2t}\boldsymbol{x}_2$ 与 $e^{3t}\boldsymbol{x}_3$。
步骤 3. 从 $\boldsymbol{u}(0)$出发的组合是 $\boldsymbol{u}(t) = 2e^t\boldsymbol{x}_1 + 3e^{2t}\boldsymbol{x}_2 + 4e^{3t}\boldsymbol{x}_3$。

系数 2, 3, 4 来自求解线性方程式 $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + c_3\boldsymbol{x}_3 = \boldsymbol{u}(0)$：

$$\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}1 & 1 & 1\\0 & 1 & 1\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}2\\3\\4\end{bmatrix} = \begin{bmatrix}9\\7\\4\end{bmatrix} \quad \text{就是 } X\boldsymbol{c} = \boldsymbol{u}(0) \tag{7}$$

你现在已经有了基本概念--如何求解 $d\boldsymbol{u}/dt = A\boldsymbol{u}$。本段落剩下的内容更深入，我们求解包含第二导数的方程式，他们在应用中经常出现。我们也要决定 $\boldsymbol{u}(t)$趋近零或是爆炸还是震荡。

最后探讨矩阵指数 e^{At}，短公式 $e^{At}\boldsymbol{u}(0)$解答了 $d\boldsymbol{u}/dt = A\boldsymbol{u}$，如同 $A^k\boldsymbol{u}_0$ 解答了 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$ 一般。范例 3 会展示“**差分方程式**”如何**协助求解微分方程式**。

请尊重版权与译者的劳动成果，侵权必究!

这些步骤都用到了λ's 与 $\boldsymbol{x}$'s，这个段落求解**转成线性代数**的常数系数问题，它阐述了最简单却最重要的微分方程式--解是完全基于成长因子 $e^{\lambda t}$。

二阶方程式

力学最重要的方程式是 $my'' + by' + ky = 0$，第一项是质量 m 乘加速度 $a = y''$，ma 项平衡了力 F(牛顿定律)，这个力引入了阻尼$(-by')$以及正比位移的弹性力$(-ky)$。这是一个二阶方程式，因为它包含第二导数 $y'' = d^2y/dt^2$，它仍然是常数系数 m, b, k 的线性系统。

在微分方程式的课程中，求解的方法是把 $y = e^{\lambda t}$代入，每个 y 的导数提出一个因子λ，我们希望 $y = e^{\lambda t}$解答下列的方程式：

$$m\frac{d^2y}{dt^2} + b\frac{dy}{dt} + ky = 0 \text{ 变成 } \left(m\lambda^2 + b\lambda + k\right)e^{\lambda t} = 0 \tag{8}$$

一切都与 $m\lambda^2 + b\lambda + k = 0$ 有关，这个λ的方程式有两个根λ_1 与λ_2，则这个 y 的方程式有两个单纯解 $y_1 = e^{\lambda_1 t}$ 与 $y_2 = e^{\lambda_2 t}$。除非$\lambda_1 = \lambda_2$，否则单纯解的组合 $c_1y_1 + c_2y_2$ 产生完整解。

线性代数课程中我们期待矩阵与特征值，因此我们把纯量方程式(包含 y'')转换成一个包含 y 与 y'的向量方程式：**只有第一导数**。假设质量 $m = 1$，两个 $\boldsymbol{u} = (y, y')$的方程式得到 $d\boldsymbol{u}/dt = A\boldsymbol{u}$：

$$\begin{aligned} \frac{dy}{dt} &= y' \\ \frac{dy'}{dt} &= -ky - by' \end{aligned} \quad \text{转换成} \quad \frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -k & -b \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix} = A\boldsymbol{u} \tag{9}$$

第一方程式 $dy/dt = y'$是平凡方程式(成立的)，第二方程式是联结 y''与 y'及 y 的方程式(8)，也联结了 $\boldsymbol{u}'$与 $\boldsymbol{u}$。所以我们使用 A 的特征值求解 $\boldsymbol{u}' = A\boldsymbol{u}$：

$$A - \lambda I = \begin{bmatrix} -\lambda & 1 \\ -k & -b-\lambda \end{bmatrix} \text{ 有行列式 } \lambda^2 + b\lambda + k = 0$$

λ的方程式与(8)相同！由于 $m = 1$，方程式还是 $\lambda^2 + b\lambda + k = 0$，方程式的根$\lambda_1$ 与λ_2 现在是 A 的特征值，所以特征向量以及解是

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} \quad \boldsymbol{u}(t) = c_1e^{\lambda_1 t}\begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} + c_2e^{\lambda_2 t}\begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$$

$\boldsymbol{u}(t)$的第一分量有 $y = c_1e^{\lambda_1 t} + c_2e^{\lambda_2 t}$--与前面的解相同，不可能是其他解。你看到的 $\boldsymbol{u}(t)$的第二分量是速度 dy/dt，速度问题与纯量问题完全一致。2×2 矩阵 A 称为**伴随矩阵**--一个有 y''的二阶方程式的伴随。

范例 3　环绕圆的运动方程式 $y''+y=0$ 与 $y=\cos t$。

这是我们的主方程式，其中 $m=1$ 与刚度系数 $k=1$ 与 $d=0$：没有阻尼。把 $y=e^{\lambda t}$ 代入 $y''+y=0$ 得到 $\boldsymbol{\lambda^2+1=0}$，根是 $\lambda=i$ 与 $-i$，$e^{it}+e^{-it}$ 的一半得到解 $y=\cos t$。

作为一阶系统，初始值 $y(0)=1$，$y'(0)=0$ 进入 $\boldsymbol{u}(0)=(1,0)$：

$$\textbf{使用 } \boldsymbol{y''=-y} \qquad \frac{d\boldsymbol{u}}{dt}=\frac{d}{dt}\begin{bmatrix}y\\y'\end{bmatrix}=\begin{bmatrix}0&1\\-1&0\end{bmatrix}\begin{bmatrix}y\\y'\end{bmatrix}=A\boldsymbol{u} \tag{10}$$

A 的特征值还是同样的 $\lambda=i$ 与 $\lambda=-i$（没有讶异），A 是反对称矩阵且特征向量 $\boldsymbol{x}_1=(1,i)$ 与 $\boldsymbol{x}_2=(1,-i)$，适配 $\boldsymbol{u}(0)=(1,0)$ 的组合是 $(\boldsymbol{x}_1+\boldsymbol{x}_2)/2$。步骤 2 把 e^{it} 与 e^{-it} 分别乘 $\boldsymbol{x}$'s，步骤 3 组合单纯震荡函数进入 $\boldsymbol{u}(t)$，如预期般得到 $y=\cos t$：

$$\boldsymbol{u}(t)=\frac{1}{2}e^{it}\begin{bmatrix}1\\i\end{bmatrix}+\frac{1}{2}e^{-it}\begin{bmatrix}1\\-i\end{bmatrix}=\begin{bmatrix}\cos t\\-\sin t\end{bmatrix}\ \text{就是}\ \begin{bmatrix}y(t)\\y'(t)\end{bmatrix}$$

都很棒。这个向量 $\boldsymbol{u}=(\cos t,-\sin t)$ 环绕一个圆（图 6.3），因为 $\cos^2 t+\sin^2 t=1$，得到半径是 1。

差分方程式(选读)

要在屏幕上显示一个圆，利用**差分方程式**取代 $y''=-y$。使用 $Y(t+\Delta t)-2Y(t)+Y(t-\Delta t)$ 有三种选择，除以 $(\Delta t)^2$ 来近似 y''。

F	**从 $n-1$ 前向**	$\dfrac{Y_{n+1}-2Y_n+Y_{n-1}}{(\Delta t)^2}=\begin{matrix}-\boldsymbol{Y}_{n-1}\\-\boldsymbol{Y}_n\\-\boldsymbol{Y}_{n+1}\end{matrix}$	(11**F**)
C	**中心在时间 n**		(11**C**)
B	**从 $n+1$ 反向**		(11**B**)

图 6.3 显示 $y(t)=\cos t$ 确实在 $t=2\pi$ 完成一个圆。在一个时间步阶(time step)长度 $\Delta t=2\pi/32$ 的情况下，这**三种差分方法**在 32 个时间步阶不能完成一个完美的圆。我们用特征值来说明这些图形：

前向 $|\lambda|>1$ (螺旋往外)　　**中心 $|\lambda|=1$ (最佳)**　　**反向 $|\lambda|<1$ (螺旋往内)**

2-步方程式(11)简化成 1-步系统 $\boldsymbol{U}_{n+1}=A\boldsymbol{U}_n$，我们不用 $\boldsymbol{u}=(y,y')$，而是使用离散未知数 $\boldsymbol{U}_n=(Y_n,Z_n)$。从 $\boldsymbol{U}(0)$ 开始，我们取 n 个时间步阶 Δt：

前向(11F)　$\begin{matrix}Y_{n+1}=Y_n+\Delta t\,Z_n\\Z_{n+1}=Z_n-\Delta t\,Y_n\end{matrix}$ 变成 $\boldsymbol{U}_{n+1}=\begin{bmatrix}1&\Delta t\\-\Delta t&1\end{bmatrix}\begin{bmatrix}Y_n\\Z_n\end{bmatrix}=A\boldsymbol{U}_n$　(12)

这些很像 $Y'=Z$ 与 $Z'=-Y$，他们是与时间 n 与 $n+1$ 有关的一阶方程式。消去 Z 会回复“前向”二阶方程式(11**F**)。

请尊重版权与译者的劳动成果，侵权必究！

我的问题很简单，点(Y_n, Z_n)会停留在圆 $Y^2 + Z^2 = 1$ 吗？不会，他们一直朝着如图 6.3 的无限大成长。我们取的是 A^n 而不是 e^{At}，所以我们测试的是大小$|\lambda|$而不是特征值的实数部分。

A 的特征值 $\lambda = 1 \pm i\Delta t$ 则 $|\lambda| > 1$ 且 (Y_n, Z_n) 螺旋往外

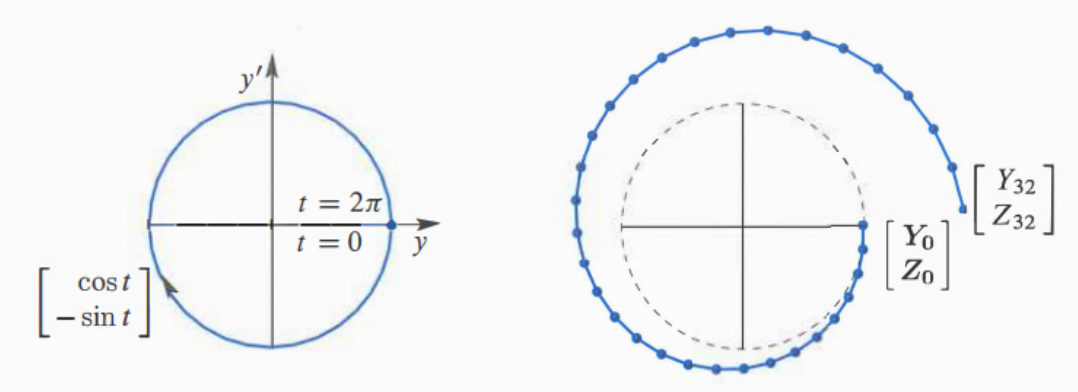


图 6.3：精确的 $\boldsymbol{u} = (\cos t, \sin t)$在一个圆上，**前向欧拉螺旋往外**(32 步)

(11B)的反向选择会做出如图 6.4 相反的结果，注意新的 A：

$$\textbf{反向(11B)} \quad \begin{matrix} Y_{n+1} = Y_n + \Delta t\, Z_{n+1} \\ Z_{n+1} = Z_n - \Delta t\, Y_{n+1} \end{matrix} \quad \textbf{是} \quad \begin{bmatrix} 1 & -\Delta t \\ \Delta t & 1 \end{bmatrix} \begin{bmatrix} Y_{n+1} \\ Z_{n+1} \end{bmatrix} = \begin{bmatrix} Y_n \\ Z_n \end{bmatrix} = \boldsymbol{U}_n \tag{13}$$

这个矩阵有特征值 $1 \pm i\Delta t$，但是我们要反过来从 $\boldsymbol{U}_n$ 得到 $\boldsymbol{U}_{n+1}$，则$|\lambda| < 1$ 可以说明为什么反向差的解是螺旋往内到(0, 0)。

在图 6.4 的右侧你看到中心选择的 32 步，如果$\Delta t < 1$，解维持与圆很接近的状态(问题 28)，这是经常使用的“跳背法(leapfrog)”，第二差 $Y_{n+1} - 2Y_n + Y_{n-1}$ 跳过(11)的中心值 Y_n。

这是化学家掌握分子运动的方法(分子动力学导致巨大的计算)，计算科学一直生气勃勃，这是因为一维方程式可以被许多**差分方程式**取代--有些不稳定，有些稳定，有些持平。问题 30 有第四种(非常好的)方法可以精确维持在圆上。

实际的工程与实际的物理都在处理系统(不止是在一个点的单一质量)，未知的 $\boldsymbol{y}$ 是一个向量；y''的系数是有 n 个物体的质量矩阵 M；y 的系数是一个刚度矩阵 K，而不是一个数字 k；y'的系数是可能为零的阻尼矩阵。

向量方程式 $M\boldsymbol{y}'' + K\boldsymbol{y} = \boldsymbol{f}$ 是计算力学的重要部分，它是由 $K\boldsymbol{x} = \lambda M\boldsymbol{x}$ 的 $M^{-1}K$ 的特征值所控制。 #

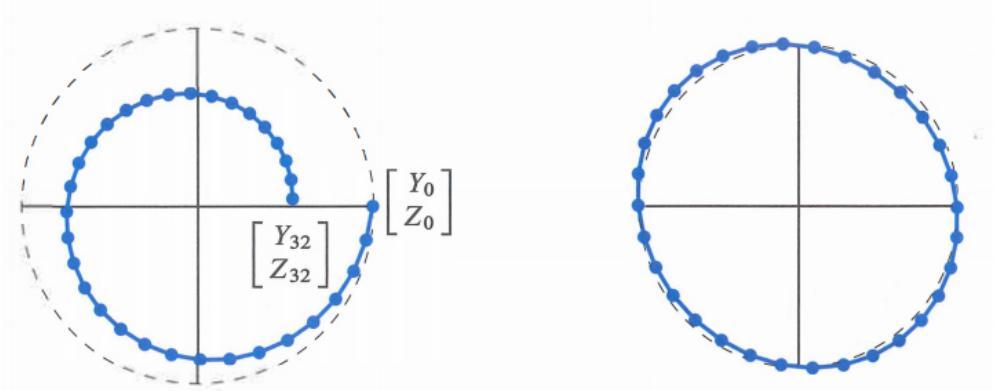


图 6.4：反向差分螺旋往内，跳背维持与正确圆接近

2×2 矩阵的稳定

对于 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的解有一个基础问题，在 $t \to \infty$时这个解趋近 $\boldsymbol{u} = \boldsymbol{0}$ 吗？这个问题依靠消耗能量而稳定？包含 e^t 的解不稳定，系统的解与 A 的特征值有关。

$\boldsymbol{u}(t)$的完整解来自单纯解$e^{\lambda t}\boldsymbol{x}$，如果特征值$\lambda$是实数，我们确切知道$e^{\lambda t}$什么时候趋近零：数字$\lambda$必须是负数。如果特征值是复数$\lambda = r + is$，实数部分 r 必须是负数。当$e^{\lambda t}$分成$e^{rt}e^{ist}$，因子e^{ist}的绝对值固定在 1：

$$e^{ist} = \cos st + i \sin st \quad \text{有} \quad |e^{ist}|^2 = \cos^2 st + \sin^2 st = 1$$

λ的实数部分控制成长$(r > 0)$或是衰减$(r < 0)$。

问题是：哪些矩阵有负的特征值？更准确的说，什么时候λ's 的实数部分全部是负数？2×2 矩阵允许一个清楚的答案。

稳定　当所有的特征值λ有**负的实数部分**，A 是**稳定**且 $\boldsymbol{u}(t) \to \boldsymbol{0}$。

2×2 矩阵 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$必须通过两个测试：

$\lambda_1 + \lambda_2 < 0$　　迹 $T = a + d$ 必须负数

$\lambda_1 \lambda_2 > 0$　　行列式 $D = ad - bc$ 必须正数

理由　若λ's 是实数且是负数，他们的总和是负数，这是迹 T。他们的乘积是正数，这是行列式 D。这个论点也可以反向说明，若 $D = \lambda_1\lambda_2$ 是正数，则λ_1 与λ_2 有相同的符号。如果 $T = \lambda_1 + \lambda_2$ 是负数，符号就是负。我们可以测试 T 与 D。

如果λ's 是复数，他们必然有这样形式 $r + is$ 与 $r - is$，否则 T 与 D 不会是实数。因为$(r + is)(r - is) = r^2 + s^2$，行列式 D 自动是正数。迹 T 是 $r + is + r - is = 2r$，所以负数的迹表示实数部分 r 是负数且矩阵是稳定，**证明完毕**。

请尊重版权与译者的劳动成果，侵权必究！

图 6.5 显示抛物线 $T^2 = 4D$ 分隔复数λ's 与实数λ's，求解$\lambda^2 - T\lambda + D = 0$ 与平方根$\sqrt{T^2 - 4D}$ 有关，这是在抛物线下方的实数与抛物线上方的虚数。稳定区域是图形的左上四分之一 —— 有负数的迹 T 与正数的行列式 D。

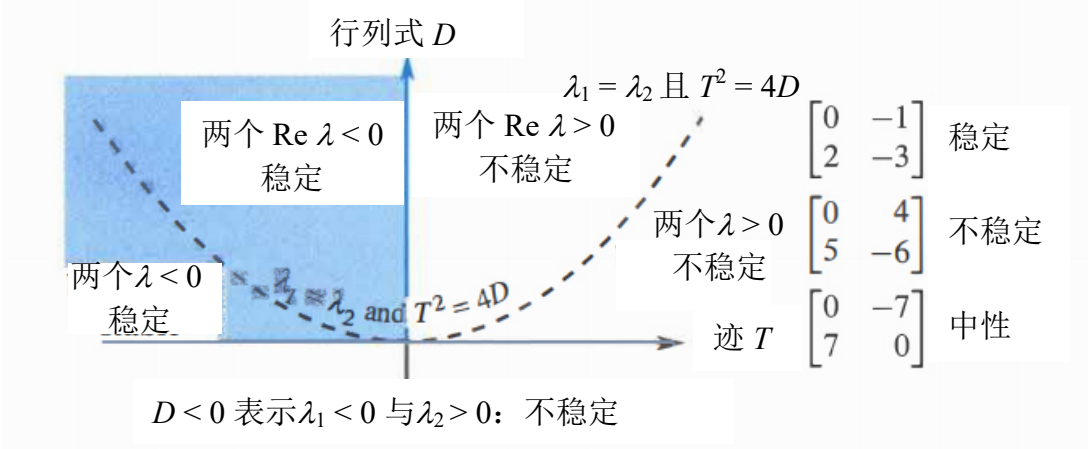


图 6.5：当**迹** < 0 与 **det** > 0，2×2 矩阵稳定($\boldsymbol{u}(t) \to \boldsymbol{0}$)

矩阵的指数

我们想要把解 $\boldsymbol{u}(t)$写成新的形式$e^{At}\boldsymbol{u}(0)$，首先我们要说一个矩阵当做指数e^{At}代表什么，要定义矩阵e^{At}，我们复制数字的e^x。

e^x直接的定义是无限级数$1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \ldots$，当你把 x 变成矩阵 At，下列的级数定义矩阵指数e^{At}：

矩阵指数 e^{At}	$e^{At} = I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 + \ldots$	(14)
它的 t 导数是 Ae^{At}	$A + A^2t + \frac{1}{2}A^3t^2 + \ldots = Ae^{At}$	
它的特征值是$e^{\lambda t}$	$\left(I + At + \frac{1}{2}(At)^2 + \ldots\right)\boldsymbol{x} = \left(1 + \lambda t + \frac{1}{2}(\lambda t)^2 + \ldots\right)\boldsymbol{x}$	

$(At)^n$ 除的数字是“**n 阶层**”，就是 $n! = (1)(2) \ldots (n-1)(n)$，在 1, 2, 6 之后是 $4! = 24$ 与 $5! = 120$，他们成长很快速。这个级数e^{At}永远收敛且它的导数永远都是Ae^{At}，因此$e^{At}\boldsymbol{u}(0)$利用一个快速公式解答了微分方程式--纵使没有用到特征向量。我会在范例 4 用到这个级数，看它在没有特征向量时怎么工作，它会产生 $te^{\lambda t}$。首先让我在良好(可对角化)情况下完成 $Xe^{\Lambda t}X^{-1}$。

本章强调如何利用对角化得到 $\boldsymbol{u}(t) = e^{At}\boldsymbol{u}(0)$，假设 A 确实有 n 个无关的特征向量，所以它可以对角化。把 $A = X\Lambda X^{-1}$ 代入e^{At}的级数，只要发生 $X\Lambda X^{-1}\ X\Lambda X^{-1}$，中间的 $X^{-1}X$ 可以消除：

使用级数
$$e^{At} = I + X\Lambda X^{-1}t + \frac{1}{2}\left(X\Lambda X^{-1}t\right)\left(X\Lambda X^{-1}t\right) + \ldots$$

提出 X 与 X^{-1}
$$= X\left[I + \Lambda t + \frac{1}{2}(\Lambda t)^2 + \ldots\right]X^{-1} \tag{15}$$

e^{At} 完成对角化
$$\boxed{e^{At} = Xe^{\Lambda t}X^{-1}}$$

e^{At}与 A 有相同的特征向量矩阵 X，Λ是一个对角矩阵，$e^{\Lambda t}$也是对角，数字 $e^{\lambda_i t}$ 在对角线上，乘积 $Xe^{\Lambda t}X^{-1}\boldsymbol{u}(0)$得到 $\boldsymbol{u}(t)$：

$$e^{At} = Xe^{\Lambda t}X^{-1} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}\begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \tag{16}$$

解 $e^{At}\boldsymbol{u}(0)$与方程式(6)的三步骤答案相同：

1. $\boldsymbol{u}(0) = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n = X\boldsymbol{c}$，此处我们需要 n 个无关的特征向量。
2. 每个 $\boldsymbol{x}_i$ 左乘成长因子 $e^{\lambda_i t}$，$\boldsymbol{x}_i$ 追随成长因子随时间前进。
3. $e^{At}\boldsymbol{u}(0)$的最佳形式是 $\boldsymbol{u}(t) = c_1e^{\lambda_1 t}\boldsymbol{x}_1 + \ldots + c_ne^{\lambda_n t}\boldsymbol{x}_n$ (17)

范例 4　当你把$e^{\lambda t}$代入 $y'' - 2y' + y = 0$，你得到有重根的方程式$\lambda^2 - 2\lambda + 1 = 0$，其中$\lambda = 1, 1$。微分方程式的课程会建议$e^t$与 te^t作为两个无关解，此处我们探讨为什么：

线性代数简化 $y'' - 2y' + y = 0$ 变成 $\boldsymbol{u} = (y, y')$的向量方程式：

$$\frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} y' \\ 2y' - y \end{bmatrix} \quad 是 \quad \frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}\boldsymbol{u} \tag{18}$$

A 有重根$\lambda = 1, 1$(迹= 2 与 $\det A = 1$)，只有一个特征向量是 $\boldsymbol{x} = (1, 1)$的倍数，不可以对角化。A 只有一条特征向量直线，所以我们从定义计算e^{At}的级数：

短级数
$$e^{At} = e^{It}\,e^{(A-I)t} = e^t\left[I + (A - I)t\right] \tag{19}$$

$e^{(A-I)t}$的“无限级数”很快结束，这是因为$(A - I)^2$ 是零矩阵！你可以在方程式(19)看到 te^t。$e^{At}\boldsymbol{u}(0)$的第一分量就是我们的答案 $y(t)$：

$$\begin{bmatrix} y \\ y' \end{bmatrix} = e^t\left[I + \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}t\right]\begin{bmatrix} y(0) \\ y'(0) \end{bmatrix} \qquad y(t) = e^t y(0) - te^t y(0) + te^t y'(0)$$

请尊重版权与译者的劳动成果，侵权必究！

范例 5 使用无限级数求 $A=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$ 的 te^{At}。注意 $A^4=I$：

$$A=\begin{bmatrix}&1\\-1&\end{bmatrix}\quad A^2=\begin{bmatrix}-1&\\&-1\end{bmatrix}\quad A^3=\begin{bmatrix}&-1\\1&\end{bmatrix}\quad A^4=\begin{bmatrix}1&\\&1\end{bmatrix}$$

A^5, A^6, A^7, A^8 是 A, A^2, A^3, A^4 的重复。A 的次方矩阵的右上角单元一直是 1, 0, –1, 0 的重复方式，所以e^{At}的无限级数的右上单元从 $t-t^3/6$ 开始，左上单元从 $1-t^2/2$ 开始：

$$e^{At}=I+At+\frac{1}{2}(At)^2+\frac{1}{6}(At)^3+\ldots=\begin{bmatrix}1-\frac{1}{2}t^2+\ldots & t-\frac{1}{6}t^3+\\ -t+\frac{1}{6}t^3-\ldots & 1-\frac{1}{2}t^2+\ldots\end{bmatrix}$$

矩阵e^{At}的上方行显示了余弦与正弦的无限级数！

$$A=\begin{bmatrix}0&1\\-1&0\end{bmatrix}\qquad e^{At}=\begin{bmatrix}\cos t&\sin t\\-\sin t&\cos t\end{bmatrix}\tag{20}$$

A 是反对称矩阵($A^{\mathrm{T}}=-A$)，它的指数e^{At}是一个正交矩阵，A 的特征值是 i 与$-i$，e^{At} 的特征值是 e^{it} 与 e^{-it}。三个规则：

1. **e^{At} 永远有逆矩阵 e^{-At}。**
2. **e^{At} 的特征值永远有是$e^{\lambda t}$。**
3. **当 A 是反对称，e^{At} 是正交，逆矩阵 = 转置 $=e^{-At}$。**

Antisymmetric 与 skew-symmetric 相同，这些矩阵都有纯虚数的特征值，例如 i 与 $-i$，则e^{At}有如同 e^{it} 与 e^{-it} 的特征值，他们的绝对值是 1：中性稳定，纯震荡，能量守恒，所以$\|\boldsymbol{u}(t)\|=\|\boldsymbol{u}(0)\|$。

我们的最后一个范例是一个三角矩阵 A，则特征向量矩阵 X 也是三角形，X^{-1} 也是，e^{At}也是。你会看到解的两个形式：特征向量的组合以及短形式$e^{At}\boldsymbol{u}(0)$。

范例 6 求解 $\dfrac{d\boldsymbol{u}}{dt}=A\boldsymbol{u}=\begin{bmatrix}1&1\\0&2\end{bmatrix}\boldsymbol{u}$，初始值在 $t=0$ 时 $\boldsymbol{u}(0)=\begin{bmatrix}2\\1\end{bmatrix}$。

解 特征值 1 与 2 在 A 的对角线上(因为 A 是三角形)，特征向量是(1, 0)与(1, 1)。开始的 $\boldsymbol{u}(0)$是 $\boldsymbol{x}_1+\boldsymbol{x}_2$ 所以 $c_1=c_2=1$，则 $\boldsymbol{u}(t)$是单纯指数(当$\lambda=1$ 与 2，没有 $te^{\lambda t}$)的相同组合：

$$\boldsymbol{u}'=A\boldsymbol{u}\text{ 的解}\qquad \boldsymbol{u}(t)=e^t\begin{bmatrix}1\\0\end{bmatrix}+e^{2t}\begin{bmatrix}1\\1\end{bmatrix}$$

这是最清楚的形式，但是矩阵形式的e^{At}可以从任意的 $\boldsymbol{u}(0)$得到 $\boldsymbol{u}(t)$：

$$\boldsymbol{u}(t)=Xe^{\Lambda t}X^{-1}\boldsymbol{u}(0)\text{ 是 }\begin{bmatrix}1&1\\0&1\end{bmatrix}\begin{bmatrix}e^t&\\&e^{2t}\end{bmatrix}\begin{bmatrix}1&-1\\0&1\end{bmatrix}\boldsymbol{u}(0)=\begin{bmatrix}e^t&e^{2t}+e^t\\0&e^{2t}\end{bmatrix}\boldsymbol{u}(0)$$

最后的矩阵是 e^{At}，它很棒，因为 A 是三角形，这个情况与 $A\boldsymbol{x}=\boldsymbol{b}$ 与逆矩阵相同。我们不需要 A^{-1} 求解 $\boldsymbol{x}$，也不需要e^{At}求解 $d\boldsymbol{u}/dt=A\boldsymbol{u}$，但是作为解答的快速公式，$A^{-1}\boldsymbol{b}$ 与$e^{At}\boldsymbol{u}(0)$是不可击败的。

主要观念的复习

1. 方程式 $\boldsymbol{u}' = A\boldsymbol{u}$ 是线性且 A 的系数是常数，从 $\boldsymbol{u}(0)$开始。
2. 它的解通常是指数的组合，与每个λ及 $\boldsymbol{x}$ 有关：

$$\text{无关的特征向量} \qquad \boldsymbol{u}(t) = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + \ldots + c_n e^{\lambda_n t}\boldsymbol{x}_n$$

3. 常数 $c_1, \ldots, c_n$ 由 $\boldsymbol{u}(0) = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n = X\boldsymbol{c}$ 决定。
4. 若每个λ都有负数实数部分，则 $\boldsymbol{u}(t)$趋近零(稳定)：所有$e^{\lambda t} \to 0$
5. 解有短形式$e^{At}\boldsymbol{u}(0)$，其中e^{At}是矩阵指数。
6. 把 y 与 y'组合成向量 $\boldsymbol{u}$，则 y''的方程式简化成 $\boldsymbol{u}' = A\boldsymbol{u}$。

已解范例

6.3A 把$e^{\lambda t}$代入 $y'' + 4y' + 3y = 0$ 求解，也使用线性代数求解。

解 把 $y = e^{\lambda t}$代入，得到$(\lambda^2 + 4\lambda + 3)e^{\lambda t} = 0$。分解二次式得到$\lambda^2 + 4\lambda + 3 = (\lambda + 1)(\lambda + 3) = 0$，因此$\lambda_1 = -1$ 与$\lambda_2 = -3$，单纯解是 $y_1 = e^{-t}$与 $y_2 = e^{-3t}$，完整解 $y = c_1y_1 + c_2y_2$ 趋近零。

使用线性代数，我们令 $\boldsymbol{u} = (y, y')$，则向量方程式是 $\boldsymbol{u}' = A\boldsymbol{u}$：

$$\begin{matrix} dy/dt = y' \\ dy'/dt = -3y - 4y' \end{matrix} \quad \text{转换成} \quad \frac{d\boldsymbol{u}}{dt} = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}\boldsymbol{u}$$

这个 A 是伴随矩阵，它的特征值再次是–1 与–3：

$$\text{同样的二次式} \qquad \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -3 & -4-\lambda \end{vmatrix} = \lambda^2 + 4\lambda + 3 = 0$$

A 的特征向量是$(1, \lambda_1)$与$(1, \lambda_2)$，$y(t)$的衰减不是来自e^{-t}就是来自 e^{-3t}。因为是常数系数，微积分变成线性代数 $A\boldsymbol{x} = \lambda\boldsymbol{x}$。

注解 在线性代数中，严重的危险是特征向量的不足，如果$\lambda_1 = \lambda_2$，我们的特征向量$(1, \lambda_1)$与$(1, \lambda_2)$相同，我们无法对角化 A，这种情况下我们无法得到 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的两个无关解。

在微分方程式中，危险也是重复的λ。在 $y = e^{\lambda t}$之后，还需要第二个解，它会变成 $y = te^{\lambda t}$。这个“不单纯”的解(有额外的 t)出现在矩阵指数e^{At}中，范例 4 会展示如何出现。

请尊重版权与译者的劳动成果，侵权必究!

6.3B 求出 A 的特征值与特征向量，把 $\boldsymbol{u}(0)=(0,2\sqrt{2},0)$写成特征向量的组合，求解两个方程式 $\boldsymbol{u}'=A\boldsymbol{u}$ 与 $\boldsymbol{u}''=A\boldsymbol{u}$：

$$\frac{d\boldsymbol{u}}{dt}=\begin{bmatrix}-2&1&0\\1&-2&1\\0&1&-2\end{bmatrix}\boldsymbol{u} \quad 与 \quad \frac{d^2\boldsymbol{u}}{dt^2}=\begin{bmatrix}-2&1&0\\1&-2&1\\0&1&-2\end{bmatrix}\boldsymbol{u}，其中\frac{d\boldsymbol{u}}{dt}(0)=0$$

$\boldsymbol{u}'=A\boldsymbol{u}$ 好像热方程式 $\partial u/\partial t=\partial^2 u/\partial x^2$。

它的解 $u(t)$会衰减(A 有负数特征值)。

$\boldsymbol{u}''=A\boldsymbol{u}$：好像波方程式 $\partial^2 u/\partial t^2=\partial^2 u/\partial x^2$。

它的解会震荡(λ的平方根是虚数)。

解 特征值与特征向量来自 $\det(A-\lambda I)=0$：

$$\det(A-\lambda I)=\begin{bmatrix}-2-\lambda&1&0\\1&-2-\lambda&1\\0&1&-2-\lambda\end{bmatrix}=(-2-\lambda)[(-2-\lambda)^2-2]=0$$

当$-2-\lambda=0$，一个特征值是$\lambda=-2$。另一个因子是$\lambda^2+4\lambda+2$，所以其他的特征值(也是负数实数)是$\lambda=-2\pm\sqrt{2}$。找出特征向量：

$$\lambda=-2 \qquad (A+2I)\boldsymbol{x}=\begin{bmatrix}0&1&0\\1&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \quad 得到\ \boldsymbol{x}_1=\begin{bmatrix}1\\0\\-1\end{bmatrix}$$

$$\lambda=-2-\sqrt{2} \qquad (A-\lambda I)\boldsymbol{x}=\begin{bmatrix}\sqrt{2}&1&0\\1&\sqrt{2}&1\\0&1&\sqrt{2}\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \quad 得到\ \boldsymbol{x}_2=\begin{bmatrix}1\\-\sqrt{2}\\1\end{bmatrix}$$

$$\lambda=-2+\sqrt{2} \qquad (A-\lambda I)\boldsymbol{x}=\begin{bmatrix}-\sqrt{2}&1&0\\1&-\sqrt{2}&1\\0&1&-\sqrt{2}\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix} \quad 得到\ \boldsymbol{x}_3=\begin{bmatrix}1\\\sqrt{2}\\1\end{bmatrix}$$

特征向量是正交(在段落 6.4 会证明所有对称矩阵都是这样)，全部三个λ_i 都是负数，这个 A 是负定(negative definite)且 e^{At} 衰减到零(稳定)。

起始值 $\boldsymbol{u}(0)=(0,2\sqrt{2},0)$是 $\boldsymbol{x}_3-\boldsymbol{x}_2$，解是 $\boldsymbol{u}(t)=e^{\lambda_3 t}\boldsymbol{x}_3-e^{\lambda_2 t}\boldsymbol{x}_2$。

热方程式 在图 6.6a，中心点的温度从$2\sqrt{2}$开始，热扩散至相邻的盒子然后到盒子外部(在 0°冰冻)，在盒子之间的热流率是温度差。从盒子 2，热流向左与右的流率是 u_1-u_2 与 u_3-u_2，所以流出是 $A\boldsymbol{u}$ 的第二行 $u_1-2u_2+u_3$。

波方程式 $d^2\boldsymbol{u}/dt^2=A\boldsymbol{u}$ 有相同的特征向量 $\boldsymbol{x}$，但是现在特征值λ产生震荡$e^{i\omega t}\boldsymbol{x}$与$e^{-i\omega t}\boldsymbol{x}$，频率来自$\omega^2=-\lambda$：

$$\frac{d^2}{dt^2}\left(e^{i\omega t}\boldsymbol{x}\right)=A\left(e^{i\omega t}\boldsymbol{x}\right) \quad 变成 \quad (i\omega)^2 e^{i\omega t}\boldsymbol{x}=\lambda e^{i\omega t}\boldsymbol{x} \quad 且 \quad \omega^2=-\lambda$$

$-\lambda$有两个平方根，所以我们有$e^{i\omega t}\boldsymbol{x}$ 与$e^{-i\omega t}\boldsymbol{x}$。三个特征向量得到 $\boldsymbol{u}''=A\boldsymbol{u}$ 的六个解，某个组合会适配 $\boldsymbol{u}(0)$与 $\boldsymbol{u}'(0)$的六个分量。由于本问题的 $\boldsymbol{u}'(0)=\boldsymbol{0}$，$e^{i\omega t}\boldsymbol{x}$ 与 $e^{-i\omega t}\boldsymbol{x}$ 得到 $2\cos\omega t\,\boldsymbol{x}$。 【原文少了 (0)】

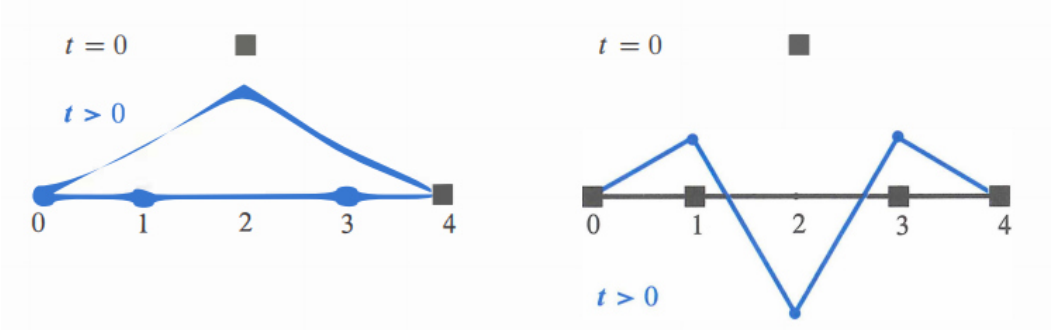


图 6.6：热从盒子 2 往外扩散(左)，波从盒子 2 传播(右)

6.3C　依据顺序求解 4 个方程式 $da/dt = 0$，$db/dt = a$，$dc/dt = 2b$，$dz/dt = 3c$，从 $\boldsymbol{u}(0) = (a(0), b(0), c(0), z(0))$出发。利用矩阵指数 $\boldsymbol{u}(t) = e^{At}\boldsymbol{u}(0)$求解同样的方程式。

4 个方程式
$\lambda = 0,\ 0,\ 0,\ 0$
固有值在对角线

$$\frac{d}{dt}\begin{bmatrix} a \\ b \\ c \\ z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \\ z \end{bmatrix} \quad \text{是} \quad \frac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$$

首先求出 A^2, A^3, A^4 与 $e^{At} = I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3$，为什么级数停止了？为什么 $(e^A)(e^A) = (e^{2A})$是正确的？**e^{As} 乘 e^{At} 永远是 $e^{A(s+t)}$。**

解 1　积分 $da/dt = 0$ 然后 $db/dt = a$ 然后 $dc/dt = 2b$ 然后 $dz/dt = 3c$：

$a(t) = \quad a(0)$
$b(t) = t\,a(0) + \quad b(0)$
$c(t) = t^2\,a(0) + 2tb(0) + \quad c(0)$
$z(t) = t^3\,a(0) + 3t^2b(0) + 3tc(0) + z(0)$

4×4 矩阵乘
$a(0), b(0), c(0), \boldsymbol{z}(0)$
产生 $a(t), b(t), c(t), \boldsymbol{z}(t)$
必须是与下方相同的 e^{At}

解 2　A 的次方(严格三角形)在 A^3 以后都是零。　【上方原文 z 写成 d，错误】

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & \mathbf{3} & 0 \end{bmatrix} \quad A^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathbf{2} & 0 & 0 & 0 \\ 0 & \mathbf{6} & 0 & 0 \end{bmatrix} \quad A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathbf{6} & 0 & 0 & 0 \end{bmatrix} \quad A^4 = 0$$

每一个步骤对角线往下移，所以e^{At}的级数在 4 项后停止：

与解 1 相同的 e^{At}

$$e^{At} = I + At + \frac{1}{2}(At)^2 + \frac{1}{6}(At)^3 = \begin{bmatrix} 1 & & & \\ t & 1 & & \\ t^2 & 2t & 1 & \\ t^3 & 3t^2 & 3t & 1 \end{bmatrix}$$

e^A 的平方是 e^{2A}，但是 e^Ae^B 与 e^Be^A 与 e^{A+B} 可以是完全不同。

请尊重版权与译者的劳动成果，侵权必究!

问题集 6.3

1 求出两个λ's 与 $\boldsymbol{x}$'s 使得$e^{\lambda t}\boldsymbol{x}$ 解答

$$\frac{d\boldsymbol{u}}{dt}=\begin{bmatrix}4 & 3\\ 0 & 1\end{bmatrix}\boldsymbol{u}$$

从 $\boldsymbol{u}(0)=(5,-2)$出发，会有什么样的组合 $\boldsymbol{u}(t)=c_1e^{\lambda_1 t}\boldsymbol{x}_1+c_2e^{\lambda_2 t}\boldsymbol{x}_2$ ？

2 利用反向代入法，先 z 后 y 求解问题 1 的 $\boldsymbol{u}=(y,z)$：

从 $z(0)=-2$，求解 $\dfrac{dz}{dt}=z$，然后从 $y(0)=5$，求解 $\dfrac{dy}{dt}=4y+3z$ 。

y 的解是 e^{4t} 与 e^{t} 的组合，λ's 是 4 与 1。

3 (a) 若 A 的每一列相加等于零，为什么$\lambda=0$ 是一个特征值？

(b) 负数对角线单元与正数非对角线单元相加等于零，$\boldsymbol{u}'=A\boldsymbol{u}$ 是“连续”的马尔可夫方程式，求出特征值与特征向量，以及 $t\to\infty$的稳定状态。

$$\text{从 } \boldsymbol{u}(0)=(4,1)\text{，求解 } \frac{d\boldsymbol{u}}{dt}=\begin{bmatrix}-2 & 3\\ 2 & -3\end{bmatrix}\boldsymbol{u}\text{，}\boldsymbol{u}(\infty)\text{为何？}$$

4 一扇门在两个房间之间开启，保持 $v(0)=30$ 人与 $w(0)=10$ 人。两个房间之间的移动正比于 $v-w$：

$$\frac{dv}{dt}=w-v \quad \text{与} \quad \frac{dw}{dt}=v-w$$

证明总数 $v+w$ 保持常数(40 人)，求出 $d\boldsymbol{u}/dt=A\boldsymbol{u}$ 的矩阵以及它的特征值与特征向量。在 $t=1$ 与 $t=\infty$时，v 与 w 为何？

5 反转问题 4 的人潮扩散问题变成 $d\boldsymbol{u}/dt=-A\boldsymbol{u}$：

$$\frac{dv}{dt}=v-w \quad \text{与} \quad \frac{dw}{dt}=w-v$$

总数 $v+w$ 仍然保持常数，现在 A 变成$-A$，λ's 会如何变化？证明 $v(t)$从 $v(0)=30$ 成长至无限大。

6 A 有实数特征值但是 B 有复数特征值：

$$A=\begin{bmatrix}a & 1\\ 1 & a\end{bmatrix} \qquad B=\begin{bmatrix}b & -1\\ 1 & b\end{bmatrix} \qquad (a \text{ 与 } b \text{ 是实数})$$

求出 a 与 b 的条件使得在 $t\to\infty$时，$d\boldsymbol{u}/dt=A\boldsymbol{u}$ 与 $d\boldsymbol{v}/dt=B\boldsymbol{v}$ 全部的解都趋近零：所有的特征值 $\mathbf{Re}\ \lambda<0$。

网址：linear.neocities.org 邮箱：876152660@qq.com

7 假设 P 是在 $\mathbf{R}^2$ 的 45°直线 $y = x$ 的投影矩阵，它的特征值为何？如果 $d\boldsymbol{u}/dt = -P\boldsymbol{u}$(注意负号)，你可以找出从 $\boldsymbol{u}(0) = (3, 1)$出发的 $\boldsymbol{u}(t)$在 $t \to \infty$时的极限吗？

8 兔子的总数显示成长快速(来自 $6r$)，但是被狼吃掉的损失(来自$-2w$)。狼的总数一直以这种模式成长($-w^2$ 会控制狼)：

$$\frac{dr}{dt} = 6r - 2w \quad 与 \quad \frac{dw}{dt} = 2r + w$$

求出特征值与特征向量。若 $r(0) = w(0) = 30$，在时间 t 时，总数各是多少？在很长时间之后，兔子对狼的比值是多少？

9 (a) 把(4, 0)写成 A 的两个特征向量的组合 $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2$：

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ i \end{bmatrix} = i\begin{bmatrix} 1 \\ i \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ -i \end{bmatrix} = -i\begin{bmatrix} 1 \\ -i \end{bmatrix}$$

(b) 从(4, 0)开始，$d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的解是 $\boldsymbol{u}(t) = c_1e^{it}\boldsymbol{x}_1 + c_2e^{-it}\boldsymbol{x}_2$，代入 $e^{it} = \cos t + i\sin t$ 与 $e^{-it} = \cos t - i\sin t$，求出 $\boldsymbol{u}(t)$。

问题 10-13 简化二阶方程式至(y, y')的一阶系统。

10 求 A 使得纯量方程式 $y'' = 5y' + 4y$ 变成 $\boldsymbol{u} = (y, y')$的向量方程式：

$$\frac{d\boldsymbol{u}}{dt} = \begin{bmatrix} y' \\ y'' \end{bmatrix} = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix} = A\boldsymbol{u}$$

A 的特征值为何？也利用 $y = e^{\lambda t}$代入 $y'' = 5y' + 4y$ 求出特征值。

11 $y'' = 0$ 的解是一条直线 $y = C + Dt$，转换成矩阵方程式：

$$\frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix} \quad 的解是 \quad \begin{bmatrix} y \\ y' \end{bmatrix} = e^{At}\begin{bmatrix} y(0) \\ y'(0) \end{bmatrix}$$

矩阵 A 的$\lambda = 0, 0$ 且不能对角化。求出 A^2 并计算$e^{At} = I + At + \frac{1}{2}A^2t^2 + \ldots$，执行$e^{At}$乘$(y(0), y'(0))$，检验直线 $y(t) = y(0) + y'(0)t$。

12 把 $y = e^{\lambda t}$代入 $y'' = 6y' - 9y$，证明$\lambda = 3$ 是重根。这是一个麻烦；在 e^{3t} 之后我们需要第二个解。矩阵方程式是：

$$\frac{d}{dt}\begin{bmatrix} y \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -9 & 6 \end{bmatrix}\begin{bmatrix} y \\ y' \end{bmatrix}$$

证明矩阵的$\lambda = 3, 3$ 且只有一条特征向量直线，也是一个麻烦。证明 $y'' = 6y' - 9y$ 的第二个解是 $y = te^{3t}$ 。

请尊重版权与译者的劳动成果，侵权必究!

13 (a) 写下两个熟悉的函数可以求解 $d^2y/dt^2 = -9y$，哪一个解是从 $y(0) = 3$ 与 $y'(0) = 0$ 出发？

(b) 二阶方程式 $y'' = -9y$ 产生向量方程式 $\boldsymbol{u}' = A\boldsymbol{u}$：

$$\boldsymbol{u} = \begin{bmatrix} y \\ y' \end{bmatrix} \qquad \frac{d\boldsymbol{u}}{dt} = \begin{bmatrix} y' \\ y'' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -9 & 0 \end{bmatrix} \begin{bmatrix} y \\ y' \end{bmatrix} = A\boldsymbol{u}$$

利用 A 的特征值与特征向量求出 $\boldsymbol{u}(t)$：$\boldsymbol{u}(0) = (3, 0)$。

14 方程式中的矩阵是反对称($A^{\mathrm{T}} = -A$)：

$$\frac{d\boldsymbol{u}}{dt} = \begin{bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{bmatrix} \boldsymbol{u} \quad \text{或} \quad \begin{aligned} u_1' &= cu_2 - bu_3 \\ u_2' &= au_3 - cu_1 \\ u_3' &= bu_1 - au_2 \end{aligned}$$

(a) $\| \boldsymbol{u}(t) \|^2 = u_1^2 + u_2^2 + u_3^2$ 的导数是 $2u_1u_1' + 2u_2u_2' + 2u_3u_3'$，代入 u_1', u_2', u_3' 得到零，则 $\| \boldsymbol{u}(t) \|^2$ 维持与 $\| \boldsymbol{u}(0) \|^2$ 相等。

(b) 当 A 是反对称矩阵，$Q = e^{At}$ 是正交。从 $Q = e^{At}$ 的级数证明 $Q^{\mathrm{T}} = e^{-At}$，则 $Q^{\mathrm{T}}Q = I$。

15 若 A 可逆，则 $d\boldsymbol{u}/dt = A\boldsymbol{u} - \boldsymbol{b}$ 的一个特定解是 $\boldsymbol{u}_p = A^{-1}\boldsymbol{b}$。$d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的一般解得到 $\boldsymbol{u}_n$，求出完整解 $\boldsymbol{u} = \boldsymbol{u}_p + \boldsymbol{u}_n$：

(a) $\dfrac{d\boldsymbol{u}}{dt} = \boldsymbol{u} - 4$ (b) $\dfrac{d\boldsymbol{u}}{dt} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \boldsymbol{u} - \begin{bmatrix} 4 \\ 6 \end{bmatrix}$

16 若 c 不是 A 的特征值，代入 $\boldsymbol{u} = e^{ct}\boldsymbol{v}$，求出 $d\boldsymbol{u}/dt = A\boldsymbol{u} - e^{ct}\boldsymbol{b}$ 的一个特定解。当 c 是 A 的一个特征值，它是如何崩溃的？$d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的“零空间”包含平常解 $e^{\lambda_i t}\boldsymbol{x}_i$。

17 求出矩阵 A 说明图 6.5 的每个不稳定区域：

(a) $\lambda_1 < 0$ 与 $\lambda_2 > 0$ (b) $\lambda_1 > 0$ 与 $\lambda_2 > 0$ (c) $\lambda = a \pm ib$，$a > 0$

问题 18-27 有关矩阵的指数 e^{At}。

18 写出 e^{At} 的无限级数的前 5 项。对每一项取 t 的导数。说明你有 4 项的 Ae^{At}。结论：$e^{At}\boldsymbol{u}(0)$可以解答 $\boldsymbol{u}' = A\boldsymbol{u}$。

19 矩阵 $B = \begin{bmatrix} 0 & -4 \\ 0 & 0 \end{bmatrix}$ 有 $B^2 = 0$，从(短)无限级数求出 e^{Bt}。检验 e^{Bt} 的导数是 Be^{Bt}。

20 从 $\boldsymbol{u}(0)$出发，在时间 T 的解是$e^{AT}\boldsymbol{u}(0)$，再走额外的时间 t 得到$e^{At}e^{AT}\boldsymbol{u}(0)$。在时间 $t + T$ 的解也可以写成_____。结论：e^{At} 乘e^{AT} = _____。

21 把 $A = \begin{bmatrix} 1 & 4 \\ 0 & 0 \end{bmatrix}$ 写成 $X\Lambda X^{-1}$ 的形式，从 $Xe^{\Lambda t}X^{-1}$ 求出e^{At}。

网址：linear.neocities.org 邮箱：876152660@qq.com

22 若 $A^2 = A$，证明无限级数 $e^{At} = I + \left(e^t - 1\right)A$。当 $A = \begin{bmatrix} 1 & 4 \\ 0 & 0 \end{bmatrix}$(问题 21)，求 e^{At}。

23 通常 $e^A e^B \neq e^B e^A$，他们都与 e^{A+B} 不同。利用问题 21-22 与 19，检验这个结果。(若 $AB = BA$，这三者都相同)。

$$A = \begin{bmatrix} 1 & 4 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & -4 \\ 0 & 0 \end{bmatrix} \quad A + B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

24 把 $A = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix}$ 写成 $X\Lambda X^{-1}$，执行 $Xe^{\Lambda t}X^{-1}$ 求出矩阵指数 e^{At}。在 $t = 0$ 时，检查 e^{At} 与 e^{At} 的导数。

25 把 $A = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$ 代入无限级数求 e^{At}，先计算 A^2 然后 A^n：

$$e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} t & 3t \\ 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} & \\ & \end{bmatrix} + \ldots = \begin{bmatrix} e^t & \\ 0 & \end{bmatrix}$$

26 (推荐) 给出两个理由，为什么矩阵指数 e^{At} 绝对不是奇异？

(a) 写出逆矩阵。

(b) 为什么特征值不是零？若 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 则 $e^{At}\boldsymbol{x} =$ _____$\boldsymbol{x}$。

27 找出解 $x(t), y(t)$ 在 $t \to \infty$ 时解变得很大。为了避免不稳定，科学家交换两个方程式：

$$\begin{matrix} dx/dt = 0x - 4y \\ dy/dt = -2x + 2y \end{matrix} \quad \text{变成} \quad \begin{matrix} dy/dt = -2x + 2y \\ dx/dt = 0x - 4y \end{matrix}$$

现在矩阵变成 $\begin{bmatrix} -2 & 2 \\ 0 & -4 \end{bmatrix}$ 是稳定，它的特征值是负数，这是怎么做到的？

挑战问题

28 范例 3 的中心化 $y'' = -y$ 会产生 $Y_{n+1} - 2Y_n + Y_{n-1} = -(\Delta t)^2 Y_n$，这个可以改写成 1-步差分方程式 $\boldsymbol{U} = (Y, Z)$：

$$\begin{matrix} Y_{n+1} = Y_n + \Delta t\, Z_n \\ Z_{n+1} = Z_n - \Delta t\, Y_{n+1} \end{matrix} \qquad \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix}\begin{bmatrix} Y_{n+1} \\ Z_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} Y_n \\ Z_n \end{bmatrix}$$

逆反左侧的矩阵写成 $\boldsymbol{U}_{n+1} = A\boldsymbol{U}_n$，证明 $\det A = 1$。选择大的时间步阶 $\Delta t = 1$，找出 A 的特征值 λ_1 与 $\lambda_2 = \overline{\lambda_1}$：

$$\boldsymbol{A = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}} \text{ 有 } \boldsymbol{|\lambda_1| = |\lambda_2| = 1}\text{，证明 } \boldsymbol{A^6} \text{ 恰好是 } \boldsymbol{I}\text{。}$$

在 6 步之后的 $t = 6$，$\boldsymbol{U}_6 = \boldsymbol{U}_0$。在 $t = 2\pi$，确切的 $y = \cos t$ 回到 1。

请尊重版权与译者的劳动成果，侵权必究!

29 问题 28 的中心选择(跳背方法)对于小的时间步阶非常有用，对于$\Delta t=\sqrt{2}$ 与 2 求特征值：

$$A=\begin{bmatrix}1 & \sqrt{2}\\ -\sqrt{2} & -1\end{bmatrix} \quad 与 \quad A=\begin{bmatrix}1 & 2\\ -2 & -3\end{bmatrix}$$

两个矩阵都有$|\lambda|=1$。两个情况下计算 A^4 然后找出 A 的特征向量。第二个$\Delta t=2$ 落在不稳定的边境，任何时间步阶$\Delta t>2$ 导致$|\lambda|>1$，所以 $\boldsymbol{U}_n=A^n\boldsymbol{U}_0$ 的次方会爆炸。

注解　你可能说没有人会在$\Delta t>2$ 时计算，但是如果一个原子的震荡是 $y''=-1000000y$，则$\Delta t>.0002$ 会导致不稳定。跳背有个非常严格的稳定极限，$Y_{n+1}=Y_n+3Z_n$ 与 $Z_{n+1}=Z_n-3Y_{n+1}$ 会爆炸的原因是$\Delta t=3$ 太大，这个矩阵有$|\lambda|>1$。

30 另一个关于 $y''=-y$ 的好概念是梯形方法(半前向/半反向)，**这个可能是确切维持(Y_n, Z_n)在一个圆上的最佳方法。**

梯形
$$\begin{bmatrix}1 & -\Delta t/2\\ \Delta t/2 & 1\end{bmatrix}\begin{bmatrix}Y_{n+1}\\ Z_{n+1}\end{bmatrix}=\begin{bmatrix}1 & \Delta t/2\\ -\Delta t/2 & 1\end{bmatrix}\begin{bmatrix}Y_n\\ Z_n\end{bmatrix}$$

(a) 逆反左侧的矩阵，把方程式写成 $\boldsymbol{U}_{n+1}=A\boldsymbol{U}_n$，证明 A 是一个正交矩阵：$A^{\mathrm{T}}A=I$。**这些点 $\boldsymbol{U}_n$ 不会离开这个圆**。若 $B^{\mathrm{T}}=-B$，则 $A=(I-B)^{-1}(I+B)$永远是正交矩阵。

(b) (选读 MATLAB) 从 $\boldsymbol{U}_0=(1, 0)$取 32 个时间步阶到 $\boldsymbol{U}_{32}$，其中$\Delta t=2\pi/32$。$\boldsymbol{U}_{32}$ 等于 $\boldsymbol{U}_0$ 吗？我想会有一个小误差。

31 矩阵的余弦的定义如同e^A，利用复制 $\cos t$ 的级数：

$$\cos t=1-\frac{1}{2!}t^2+\frac{1}{4!}t^4-\ldots \qquad \cos A=I-\frac{1}{2!}A^2+\frac{1}{4!}A^4-\ldots.$$

(a) 若 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，每项乘 $\boldsymbol{x}$ 求 $\cos A$ 的特征值。

(b) 求 $A=\begin{bmatrix}\pi & \pi\\ \pi & \pi\end{bmatrix}$的特征值，对应的特征向量是(1, 1)与(1, −1)。

(c) $\cos(At)$的二次导数是$-A^2\cos(At)$。

$$\boldsymbol{u}(t)=\cos(At)\boldsymbol{u}(0)是\frac{d^2\boldsymbol{u}}{dt}=-A^2\boldsymbol{u}\text{ 的解，从 }\boldsymbol{u}'(0)=0\text{ 开始。}$$

利用针对特定 A 的三步骤，创建 $\boldsymbol{u}(t)=\cos(At)\boldsymbol{u}(0)$：

1. 扩展成特征向量组合 $\boldsymbol{u}(0)=(4, 2)=c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2$。
2. 使用_____与_____左乘这些特征向量(不用$e^{\lambda t}$)。
3. 把解相加得到 $\boldsymbol{u}(t)=c_1$____$\boldsymbol{x}_1+c_2$____$\boldsymbol{x}_2$。

网址：linear.neocities.org　邮箱：876152660@qq.com

32 下列三个都是 e^A 的平方等于 e^{2A} 的证明，选择其中一个说明。

1. 从 $t = 0$ 到 1 然后 1 到 2 求解 e^A，结果与从 0 到 2 的 e^{2A} 相同。
2. 级数的平方 $\left(I + A + \frac{1}{2!}A^2 + \ldots\right)^2$ 适配 $I + 2A + \frac{(2A)^2}{2!} + \ldots = e^{2A}$。
3. 若 A 可以对角化，则 $(Xe^{\Lambda}X^{-1})\,(X\,e^{\Lambda}X^{-1}) = Xe^{2\Lambda}\,X^{-1}$。

微分方程式课程的注解

常数系数线性方程式当然最容易求解，本书的段落 6.3 介绍了微分方程式课程的一部分，但是还有更多：

1. 二阶方程式 $mu'' + bu' + ku = 0$ 在应用上非常重要。解 $u = e^{\lambda t}$ 的指数 λ 是 $m\lambda^2 + b\lambda + k = 0$ 的解，阻尼系数非常重要：

   欠(under)阻尼 $b^2 < 4bk$　临界(critical)阻尼 $b^2 = 4bk$　过(over)阻尼 $b^2 > 4bk$

   这个决定了 λ_1 与 λ_2 是实数根，或重根，或复数根。如果 λ's 是复数，解 $u(t)$ 会边震荡边衰减。

2. 我们的方程式没有外力项 $f(t)$，我们是在找“零空间的解”。除了 $u_n(t)$ 之外，我们需要加上一个平衡外力 $f(t)$ 的特定解 $u_p(t)$：

   在时间 s 输入 $f(s)$
   成长因子 $e^{A(t-s)}$　　$u_{par} = \int_0^t e^{A(t-s)} f(s)\,ds$
   累加得到时间 t 的输出

   我们也可以使用拉普拉斯转换得到并研读这个解--这是著名的从线性微分方程式转换成线性代数的方法。

   在实际应用上，非线性微分方程式采用数值方法求解。有一个标准而且精确的方法是“Runge-Kutta”--以发现者命名。分析可以发现 $du/dt = f(u)$ 的常数解，这些是 $u(t) = Y$ 且 $f(Y) = 0$ 及 $du/dt = 0$ 的解：没有移动。我们也可以了解靠近 $u = Y$ 是稳定或不稳定。远离 Y，让电脑接管。

   这项基础课程是我的“微分方程式与线性代数”教科书(伴随这本)的一个主题：math.mit.edu/dela。

请尊重版权与译者的劳动成果，侵权必究!

6.4 对称矩阵

1 一个对称矩阵 S 有 n 个实数特征值λ_i与 n 个正交单位特征向量 $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$。

2 每一个实数对称 S 可以被对角化：$S = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$ 。

3 S 的正数特征值的个数与正数主元的个数相等。

4 反对称矩阵 $A = -A^{\mathrm{T}}$ 有虚数λ's 与正交单位(复数)$\boldsymbol{q}$'s。

5 段落 9.2 说明为什么 $S = S^{\mathrm{T}}$ 的测试，在复数矩阵时变成 $S = \overline{S}^{\mathrm{T}}$ 。

$$S = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix} = \overline{S}^{\mathrm{T}} \text{ 有实数} \lambda = 1, -1, \quad A = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} = -\overline{A}^{\mathrm{T}} \text{ 有} \lambda = i, -i \text{。}$$

不夸张的说对称矩阵 S 是世界上能看到的最重要的矩阵——在线性代数原理也在实际的应用。我们马上提出对称的关键问题，不只是问题，也是两部分的答案。

当 S 是对称，$Sx = \lambda x$ 有什么特殊？

当 $S = S^{\mathrm{T}}$，我们寻找特征值λ与特征向量 $\boldsymbol{x}$ 的特殊性质，对角化 $S = X\Lambda X^{-1}$ 会反映 S 的对称性。我们从转置 $S^{\mathrm{T}} = (X^{-1})^{\mathrm{T}}\Lambda X^{\mathrm{T}}$ 获得一些提示，因为 $S = S^{\mathrm{T}}$ 所以这些是相同的，第一形式的 X^{-1} 可能等于第二形式的 X^{T} 吗？这样就会有 $X^{\mathrm{T}}X = I$。当 $S = S^{\mathrm{T}}$，这会使得 X 的每个特征向量与其他特征向量垂直。下列是关键事实：

1. 对称矩阵只有实数特征值。
2. 特征向量可以选成正交单位。

这 n 个正交单位特征向量 $\boldsymbol{x}_1$ 进入 X 的列，每个对称矩阵都可以对角化，它的特征向量矩阵 X 变成正交矩阵 Q。正交矩阵有 $Q^{-1} = Q^{\mathrm{T}}$——我们怀疑的特征向量矩阵的性质是正确的。要记忆它，当我们“选择”正交单位特征向量时，我们不写 X 而是写成 Q。

为什么我们要用“选择”这个词？因为特征向量不需要是单位向量，他们的长度随我们处理，我们会选择单位向量——长度 1 的特征向量，不只是正交还是正交单位，则 $A = X\Lambda X^{-1}$ 是自己的独特形式，$S = Q\Lambda Q^{\mathrm{T}}$ 则是关于对称矩阵的特定形式。

(光谱定理) 每个对称矩阵的分解是 $S = Q\Lambda Q^{\mathrm{T}}$，且在$\Lambda$中有**实数特征值**，在 Q 的列中有**正交单位特征向量**：

对称对角化 $S = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$，且 $Q^{-1} = Q^{\mathrm{T}}$ (1)

很容易看出 $Q\Lambda Q^{\mathrm{T}}$ 是对称，取转置，你得到$(Q^{\mathrm{T}})^{\mathrm{T}}\Lambda^{\mathrm{T}}Q^{\mathrm{T}}$，再次是 $Q\Lambda Q^{\mathrm{T}}$。比较困难的部分是证明每一个对称矩阵都有实数λ's 与正交单位 $\boldsymbol{x}$'s，这就是数学的“**光谱**(spectral)定理”以及几何学与物理学的“**主轴**(principal axis)定理”。我们必须证明它，没有选择！我会用三个步骤来阐述这个证明：

1. 找一个例子，显示在Λ中的实数λ's 以及在 Q 中的正交单位 $\boldsymbol{x}$。
2. 当特征值没有重复时，证明这些事实。
3. 证明允许有重复的特征值(在本段落的结束)。

范例 1 当 $S=\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}$ 与 $S-\lambda I=\begin{bmatrix}1-\lambda & 2\\ 2 & 4-\lambda\end{bmatrix}$，求$\lambda$'s 与 $\boldsymbol{x}$'s。

解 $S-\lambda I$ 的行列式是$\lambda^2-5\lambda$，特征值是 0 与 5(都是实数)。我们可以直接看出：$\lambda=0$ 是一个特征值，因为 S 是奇异，以及$\lambda=5$ 适配 S 对角线的迹：$0+5$ 与 $1+4$ 一致。

两个特征向量是(2, –1)与(1, 2)––正交但还不是正交单位。$\lambda=0$ 对应的特征向量在 A 的零空间中，$\lambda = 5$ 对应的特征向量在列空间中。我们问问自己，为什么零空间与列空间垂直？基础定理告诉我们零空间与行空间垂直––而不是列空间，但是我们的矩阵是对称！它的**行空间与列空间相同**，它的特征向量(2, –1)与(1, 2)必须(也的确是)垂直。

这些特征向量的长度是$\sqrt{5}$，把他们除以$\sqrt{5}$ 得到单位向量。把这些单位特征向量放进 Q 的列，则 $Q^{-1}SQ$ 是Λ而且 $Q^{-1}=Q^{\mathrm{T}}$。

$$Q^{-1}SQ=\frac{1}{\sqrt{5}}\begin{bmatrix}2 & -1\\ 1 & 2\end{bmatrix}\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix}2 & 1\\ -1 & 2\end{bmatrix}=\begin{bmatrix}0 & 0\\ 0 & 5\end{bmatrix}=\Lambda$$

现在讨论 $n\times n$ 的情形，当 $S=S^{\mathrm{T}}$ 且 $S\boldsymbol{x}=\lambda\boldsymbol{x}$，$\lambda$'s 是实数。

实数特征值 一个实数对称矩阵的所有特征值都是实数。

证明 假设 $S\boldsymbol{x}=\lambda\boldsymbol{x}$，直到我们确认之前，$\lambda$可能是一个复数 $a+ib$(a 与 b 实数)，它的共轭(conjugate)是 $\overline{\lambda}=a-ib$。相似的，$\boldsymbol{x}$ 的分量也可能是复数，把虚数部分反转符号得到 $\overline{\boldsymbol{x}}$。

请尊重版权与译者的劳动成果，侵权必究!

好消息是 $\bar{\lambda}$ 乘 $\bar{\boldsymbol{x}}$ 永远是λ乘 $\boldsymbol{x}$ 的共轭，所以我们可以取 $S\boldsymbol{x}=\lambda\boldsymbol{x}$ 的共轭，记得 S 是实数：

$$S\boldsymbol{x}=\lambda\boldsymbol{x} \quad \text{导致} \quad S\bar{\boldsymbol{x}}=\bar{\lambda}\bar{\boldsymbol{x}} \quad \text{转置得到} \quad \bar{\boldsymbol{x}}^{\mathrm{T}}S=\bar{\boldsymbol{x}}^{\mathrm{T}}\bar{\lambda}$$

现在取第一方程式与 $\bar{\boldsymbol{x}}$ 的点积，以及最后方程式与 $\boldsymbol{x}$ 的点积：

$$\bar{\boldsymbol{x}}^{\mathrm{T}}S\,\boldsymbol{x}=\bar{\boldsymbol{x}}^{\mathrm{T}}\lambda\,\boldsymbol{x} \quad \text{以及} \quad \bar{\boldsymbol{x}}^{\mathrm{T}}S\,\boldsymbol{x}=\bar{\boldsymbol{x}}^{\mathrm{T}}\bar{\lambda}\,\boldsymbol{x} \tag{2}$$

左侧相同所以右侧相等，一个方程式有λ，另一个有$\bar{\lambda}$。$\bar{\boldsymbol{x}}^{\mathrm{T}}\bar{\boldsymbol{x}}=|x_1|^2+|x_2|^2+\ldots=$ 长度平方且不是零，因此λ必须等于 $\bar{\lambda}$，且 $a+ib=a-ib$，所以 $b=0$ 且$\lambda=a=$ 实数，**证明完毕**。

特征向量来自求解实数方程式$(S-\lambda I)\boldsymbol{x}=\boldsymbol{0}$，所以 $\boldsymbol{x}$'s 也是实数，重要的事实是他们互相垂直。

正交特征向量 一个实数对称矩阵的特征向量(对应不同的λ's)永远垂直。

证明 假设 $S\boldsymbol{x}=\lambda_1\boldsymbol{x}$ 与 $S\boldsymbol{y}=\lambda_2\boldsymbol{y}$，我们假设此处的$\lambda_1\neq\lambda_2$。取第一方程式与 $\boldsymbol{y}$ 的点积，以及第二式与 $\boldsymbol{x}$ 的点积：

利用 $S=S^{\mathrm{T}}$ $$(\lambda_1\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=(S\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\bar{\boldsymbol{x}}^{\mathrm{T}}S\,\boldsymbol{y}=\bar{\boldsymbol{x}}^{\mathrm{T}}\lambda_2\boldsymbol{y} \tag{3}$$

左侧是$\boldsymbol{x}^{\mathrm{T}}\lambda_1\boldsymbol{y}$，右侧是$\boldsymbol{x}^{\mathrm{T}}\lambda_2\boldsymbol{y}$，由于$\lambda_1\neq\lambda_2$，证明了 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0$，特征向量 $\boldsymbol{x}$(对应λ_1)与特征向量 $\boldsymbol{y}$(对应λ_2)垂直。

范例 2 一个 2×2 对称矩阵的特征向量有一个特殊形式：

不是广为人知 $$S=\begin{bmatrix}a & b\\ b & c\end{bmatrix} \text{ 有 } \boldsymbol{x}_1=\begin{bmatrix}b\\ \lambda_1-a\end{bmatrix} \text{ 与 } \boldsymbol{x}_2=\begin{bmatrix}\lambda_2-c\\ b\end{bmatrix} \tag{4}$$

这个在问题集里面。重点是 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 垂直：

$$\boldsymbol{x}_1{}^{\mathrm{T}}\boldsymbol{x}_2=b(\lambda_2-c)+(\lambda_1-a)b=b(\lambda_1+\lambda_2-a-c)=0$$

因为$\lambda_1+\lambda_2$ 等于迹 $a+c$，所以上式是零，于是 $\boldsymbol{x}_1{}^{\mathrm{T}}\boldsymbol{x}_2=0$。鹰眼可能会注意到特殊情形 $S=I$，此时 b 与λ_1-a 与λ_2-c 与 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 都是零，$\lambda_1=\lambda_2=1$ 是重复。但是当然 $S=I$ 有垂直的特征向量。

对称矩阵 S 有正交特征向量矩阵 Q，再次检视：

对称 $S=X\Lambda X^{-1}$ 变成 $S=Q\Lambda Q^{\mathrm{T}}$ 且 $Q^{\mathrm{T}}Q=I$

表示每一个 2×2 对称矩阵是 **(旋转) (拉伸) (反旋转)**：

$$S=Q\Lambda Q^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2\end{bmatrix}\begin{bmatrix}\lambda_1 & \\ & \lambda_2\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_1{}^{\mathrm{T}}\\ \boldsymbol{q}_2{}^{\mathrm{T}}\end{bmatrix} \tag{5}$$

列 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$ 乘行$\lambda_1\boldsymbol{q}_1{}^{\mathrm{T}}$ 与$\lambda_2\boldsymbol{q}_2{}^{\mathrm{T}}$ 得到 $\boldsymbol{S}=\lambda_1\,\boldsymbol{q}_1\boldsymbol{q}_1{}^{\mathrm{T}}+\lambda_2\,\boldsymbol{q}_2\boldsymbol{q}_2{}^{\mathrm{T}}$。

每个对称矩阵 $S = Q\Lambda Q^{\mathrm{T}} = \lambda_1 q_1 q_1^{\mathrm{T}} + \ldots + \lambda_n q_n q_n^{\mathrm{T}}$。 (6)

记住这些伟大结果的步骤(光谱定理)。

段落 6.2　写成 $Ax_i = \lambda_i x_i$ 的矩阵形式　$AX = \Lambda X$　或　$A = X\Lambda X^{-1}$

段落 6.4　正交单位 $x_i = q_i$　得到　$X = Q$　$S = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{T}}$

方程式(6)的 $Q\Lambda Q^{\mathrm{T}}$ 有 $Q\Lambda$的列乘 Q^{T} 的行，下列是直接的证明。

S 有正确的固有向量
这些 q's 是正交

$$Sq_i = (\lambda_1 q_1 q_1^{\mathrm{T}} + \ldots + \lambda_n q_n q_n^{\mathrm{T}})\, q_i = \lambda_i q_i \tag{7}$$

实数矩阵的复数特征值

任意实数矩阵有 $Sx = \lambda x$ 得到 $S\bar{x} = \bar{\lambda}\bar{x}$，如果是对称矩阵，$\lambda$与 x 变成实数，这两个方程式变成相同。但是非对称矩阵很容易有复数的λ与 x，则 $A\bar{x} = \bar{\lambda}\bar{x}$ 成立但是与 $Ax = \lambda x$ 就不相同。我们得到另一个复数特征值(就是$\bar{\lambda}$)以及一个新的特征向量(就是 $\bar{x}$)：

实数矩阵的复数λ's 与 x's 都是“共轭配对”一起出现。

$$\lambda = a + ib \qquad \bar{\lambda} = a - ib$$

若 $Ax = \lambda x$　则　$A\bar{x} = \bar{\lambda}\bar{x}$ (8)

范例 3　$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ 有 $\lambda_1 = \cos\theta + i\sin\theta$ 与 $\lambda_2 = \cos\theta - i\sin\theta$。

这些特征值是彼此的共轭，他们是λ与$\bar{\lambda}$。特征向量必须是 x 与 $\bar{x}$，因为 A 是实数：

$$\text{这是 } \lambda x \qquad Ax = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = (\cos\theta + i\sin\theta) \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

$$\text{这是 } \bar{\lambda}\bar{x} \qquad A\bar{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = (\cos\theta - i\sin\theta) \begin{bmatrix} 1 \\ i \end{bmatrix} \tag{9}$$

因为 A 是实数，这些特征向量$(1, -i)$与$(1, i)$是共轭复数。

这个旋转矩阵的特征值的绝对值$|\lambda| = 1$，因为$= \cos^2\theta + \sin^2\theta = 1$。**每一个正交矩阵 Q 的特征值都有这个事实$|\lambda| = 1$。**

我们很抱歉不小心插进来复数，纵使矩阵是实数也无法避免。第九章会超越复数λ与复数 x 来到复数矩阵，到时候你就可以一窥全貌。

我们以两个选读的讨论作为结束。

请尊重版权与译者的劳动成果，侵权必究!

特征值对比主元

矩阵 A 的特征值与主元非常不同，我们求解 $\det(A-\lambda I)=0$ 得到特征值，我们使用消元法得到主元，到目前为止唯一的关联是：

主元的乘积 = 行列式 = 特征值的乘积

我们假设一整组主元 $d_1, \ldots, d_n$，还有 n 个实数特征值 $\lambda_1, \ldots, \lambda_n$，这些 d's 与 λ's 不同，但是他们来自相同的对称矩阵。这些 d's 与 λ's 有隐藏的关联，**对称矩阵的主元与特征值有相同的符号：**

$S=S^{\mathrm{T}}$ 的正数特征值的个数等于正数主元的个数。

特殊情形：S 所有的 $\lambda_i>0$ 当且仅当所有的主元是正数。

这个特殊情形是段落 6.5 的“**正定矩阵**(positive definite)”的最重要事实。

范例 4 下列对称矩阵有一个正数特征值与一个正数主元：

$$\text{适配符号}\qquad S=\begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}\quad\begin{matrix}\text{主元1与}-8\\ \text{特征值4与}-2\end{matrix}$$

主元的符号与特征值的符号是适配的，一个正与一个负。如果矩阵不对称，这个结论可能是错的：

$$\text{相反符号}\qquad B=\begin{bmatrix}1 & 6\\ -1 & -4\end{bmatrix}\quad\begin{matrix}\text{主元1与2}\\ \text{特征值}-1\text{与}-2\end{matrix}$$

对角线单元是第三组数字，只是我们没有什么可以说的。

当 $S=S^{\mathrm{T}}$ 时，下列是主元与特征值适配符号的证明。

当主元从 U 的行用除法提出来是最佳视觉，则 $S=LDL^{\mathrm{T}}$。对角主元矩阵 D 在三角矩阵 L 与 L^{T} 中间：

$$\begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ \mathbf{3} & 1\end{bmatrix}\begin{bmatrix}1 & \\ & -8\end{bmatrix}\begin{bmatrix}1 & \mathbf{3}\\ 0 & 1\end{bmatrix}$$

这是 $S=LDL^{\mathrm{T}}$，它是**对称**

当 L 变成 I，观察 LDL^{T} 的特征值，此时 S 变成 D

LDL^{T} 的特征值是 4 与 -2，IDI^{T} 的特征值是 1 与 -8(主元！)，当 L 里面的“3”移动到 0，特征值改变了。但是要改变符号，实数的特征值必须跨越零，在这个时候矩阵是奇异的。我们的变化矩阵永远有主元 1 与 -8，所以它不会是奇异，当 λ's 移动到 d's，符号不可能改变。

我们针对任意的 $S=LDL^{\mathrm{T}}$ 重复证明。借由非对角线单元移动到零，我们把 L 朝着 I 移动，主元没有改变且不是零，LDL^{T} 的特征值 λ 变成 IDI^{T} 的特征值 d。由于特征值在移动到主元的过程中不能跨越零，他们的符号没有改变，**λ's 与 d's 的符号相同。**

这个结论把应用线性代数的两个半边联结在一起—主元与特征值。

所有对称矩阵可对角化

当 A 没有重复的特征值，特征向量确定是无关，则 A 可以对角化。但是重复的特征值使得特征向量不足，有时候会发生在非对称矩阵。但是对称矩阵一定不会，永远有足够的特征向量对角化 $S = S^{\mathrm{T}}$。

这是证明的一个概念，利用对角矩阵 **diag**(c, $2c$, ..., nc)稍微改变 S，如果 c 很小，新的对称矩阵没有重复的特征值，我们知道它有一整组的正交单位特征向量。当 $c \to 0$，我们得到原始矩阵 S 的 n 个正交单位特征向量--纵使 S 的某些特征值是重复的。

每个数学家都知道这样的论点是不完全的，我们如何确保小的对角矩阵会分隔特征值？(我确信这是真的。)

有一个不同的证明来自一种有效的分解法，可以应用到所有的方形矩阵 A，不论对称与否。当 S 是任意实数对称矩阵，这个新的的分解很快的得到 $S = Q\Lambda Q^{\mathrm{T}}$，而且有一整组的实数正交单位特征向量。

每个方形 A 分解成 QTQ^{-1}，其中 T 是上三角且 $\overline{Q}^{\mathrm{T}} = Q^{-1}$。

如果 A 有实数特征值，则 Q 与 T 可以选成实数：$Q^{\mathrm{T}}Q = I$。

这是舒尔定理，它的证明在网址 math.mit.edu/linearalgebra。

此处我会证明当 S 对称时，T 如何成为对角($T = \Lambda$)，则 S 是 $Q\Lambda Q^{\mathrm{T}}$。

我知道每个对称 S 有实数特征值，而且舒尔允许重复的λ's：

舒尔的 $S = QTQ^{-1}$ 表示 $T = Q^{\mathrm{T}}SQ$，转置之后还是 $Q^{\mathrm{T}}SQ$。

当 $S^{\mathrm{T}} = S$，三角 T 是对称，则 T 必须是对角矩阵且 $T = \Lambda$。

这个证明了 $S = Q\Lambda Q^{-1}$，对称 S 有 n 个在 Q 中的正交单位特征向量。

注解 我在本书段落 7.2 加了另一个证明，这个证明说明如何同时描述特征值λ。最大的λ_1 是 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，如果我们只允许与第一个特征向量垂直的向量 $\boldsymbol{x}$，则λ_2(第二大)还是这个相同的极大值。第三个λ_3 来自 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{q}_1 = 0$ 与 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{q}_2 = 0$ ....

这个证明在第七章，这是因为相同的“同时”概念也可以应用在任意矩阵 A 的奇异值(singular value)。**奇异值来自 $A^{\mathrm{T}}A$ 与 AA^{T}。**

主要观念的复习

1. 每个对称矩阵 S 有实数特征值与垂直特征向量。
2. 对角化变成 $S = Q\Lambda Q^{\mathrm{T}}$，$Q$ 是正交特征向量矩阵。
3. 所有的对称矩阵都可以对角化，纵使是重复的特征值。
4. 当 $S = S^{\mathrm{T}}$，特征值的符号适配主元的符号。
5. 借由 $A = QTQ^{-1}$，每个方形矩阵可以“三角化”。若 $A = S$ 则 $T = \Lambda$。

请尊重版权与译者的劳动成果，侵权必究!

已解范例

6.4A 什么矩阵 A 有特征值 $\lambda = 1$ 与 -1 以及特征向量 $\boldsymbol{x}_1 = (\cos\theta, \sin\theta)$ 与 $\boldsymbol{x}_2 = (-\sin\theta, \cos\theta)$？哪些性质可以事先预测？

$A = A^{\mathrm{T}}$　　$A^2 = A$　　$\det A = -1$　　主元是+与−　　$A^{-1} = A$

解 所有的性质都可以预测！实数特征值 1, −1 以及正交单位的 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$，矩阵 $A = QTQ^{\mathrm{T}}$ 必须是对称。特征值 1 与 −1 告诉我们 $A^2 = I$ (因为 $\lambda^2 = 1$) 与 $A^{-1} = A$ (相同意思)，以及 $\det A = -1$。两个主元必须跟特征值一样是一正一负，因为 A 对称。

矩阵是镜射矩阵，在 $\boldsymbol{x}_1$ 方向的向量被 A 乘时保持不变(因为 $\lambda = 1$)，在垂直方向的向量被反向(因为 $\lambda = -1$)。镜射矩阵 $A = Q\Lambda Q^{\mathrm{T}}$ 横跨“θ–线”，用 c 代替 $\cos\theta$ 与 s 代替 $\sin\theta$：

$$A = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} c & s \\ -s & c \end{bmatrix} = \begin{bmatrix} c^2 - s^2 & 2cs \\ 2cs & s^2 - c^2 \end{bmatrix} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$$

注意 $\boldsymbol{x} = (1, 0)$ 得到在 2θ–线的 $A\boldsymbol{x} = (\cos 2\theta, \sin 2\theta)$，而且 $(\cos 2\theta, \sin 2\theta)$ 横跨 θ–线回到 $\boldsymbol{x} = (1, 0)$。

6.4B 求出 A_3 与 B_4 的有特征值与特征向量(离散正弦与余弦)：

$$A_3 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad B_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 1 \end{bmatrix}$$

两个矩阵的 −1, 2, −1 模式是“第二差分”，很像第二导数。$A\boldsymbol{x} = \lambda\boldsymbol{x}$ 与 $B\boldsymbol{x} = \lambda\boldsymbol{x}$ 很像 $d^2x/dt^2 = \lambda x$，会有特征向量 $x = \sin kt$ 与 $x = \cos kt$，这些是傅里叶级数的基底。

A_n 与 B_n 得到“离散正弦”与“离散余弦”，这些是离散傅里叶转换的基底。DFT 绝对是数字信号处理各个领域的中心，图像处理中最受欢迎的选择 JPEG 就是大小 $n = 8$ 的 B_8。

解 A_3 的特征值是 $\lambda = 2 - \sqrt{2}$ 与 2 与 $2 + \sqrt{2}$ (见 **6.3B**)，他们的总和是 6 (A_3 的迹)，他们的乘积是 4 (行列式)。特征向量矩阵得到“**离散正弦转换**”，而且这些特征向量落在正弦波上：

$$\text{正弦} = \begin{bmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{bmatrix} \qquad \text{余弦} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \sqrt{2}-1 & -1 & 1-\sqrt{2} \\ 1 & 1-\sqrt{2} & -1 & \sqrt{2}-1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

正弦矩阵 $= A_3$ 的特征向量　　**余弦矩阵 $= B_4$ 的特征向量**

B_4 的特征值是 $\lambda = 2 - \sqrt{2}$ 与 2 与 $2 + \sqrt{2}$ 与 0 (与 A_3 相同，加一个零特征值)，迹还是 6，但是行列式现在是零。特征向量矩阵得到 4 点的“**离散余弦转换**”，网络上展示前两个特征向量如何落在余弦波上。(B 的所有特征向量都是。) 这些特征向量适配在半途点 $\pi/8, 3\pi/8, 5\pi/8, 7\pi/8$ 的余弦。

问题集 6.4

1 哪些 ASB 矩阵是对称且特征值是 1 与–1？

$$\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}1&\\&-1\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}\quad\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}1&\\&-1\end{bmatrix}\begin{bmatrix}1&0\\-1&1\end{bmatrix}\quad\begin{bmatrix}0&-1\\1&0\end{bmatrix}\begin{bmatrix}1&\\&-1\end{bmatrix}\begin{bmatrix}0&1\\-1&0\end{bmatrix}$$

$B = A^{\mathrm{T}}$ 做不到。$B = A^{-1}$ 做不到。$B =$ _____ = _____会成功，所以 B 必须是一个_____矩阵。

2 假设 $S = S^{\mathrm{T}}$，什么时候 ASB 也是对称且与 S 有相同的特征值？

(a) 转置 ASB 并检视当 $B =$ _____时，它维持对称。

(b) 当 $B =$ _____时，ASB 与 S 相似(相同特征值)。

把(a)与(b)放在一起，与 S 相似的对称矩阵看起来好像(___)S(___)。

3 把 A 写成 $S + N$，对称矩阵 S 加反对称矩阵 N：

$$A=\begin{bmatrix}1&2&4\\4&3&0\\8&6&5\end{bmatrix}=S+N\quad(S^{\mathrm{T}}=S\text{ 与 }N^{\mathrm{T}}=-N)$$

对于任意方形矩阵，$S = (A + A^{\mathrm{T}})/2$ 与 $N =$ _____，两者相加等于 A。

4 若 C 对称，证明 $A^{\mathrm{T}}CA$ 也是对称。(取转置) 当 A 是 6×3，C 与 $A^{\mathrm{T}}CA$ 的形状为何？

5 求出特征值与单位特征向量：

$$S=\begin{bmatrix}2&2&2\\2&0&0\\2&0&0\end{bmatrix}$$

6 求出一个可以对角化 $S=\begin{bmatrix}-2&6\\6&7\end{bmatrix}$ 的正交矩阵 Q。Λ是什么？

7 求出一个可以对角化下列对称矩阵的正交矩阵 Q：

$$S=\begin{bmatrix}1&0&2\\0&-1&-2\\2&-2&0\end{bmatrix}$$

8 求出所有可以对角化 $S=\begin{bmatrix}9&12\\12&16\end{bmatrix}$ 的正交矩阵。

9 (a) 找出一个有负数特征值的对称矩阵 $\begin{bmatrix}1&b\\b&1\end{bmatrix}$

(b) 你怎么知道它一定有一个负数主元？

(c) 你怎么知道它不能有两个负数主元？

请尊重版权与译者的劳动成果，侵权必究！

10 若 $A^3=0$ 则 A 的特征值必须是_____。给出一个例子且 $A\neq 0$。但是如果 A 是对称，把它对角化，证明 A 必须是个零矩阵。

11 若$\lambda=a+ib$ 是实数矩阵 A 的特征值。则它的共轭 $\overline{\lambda}$ 也是一个特征值。(若 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 则 $A\overline{\boldsymbol{x}}=\overline{\lambda}\overline{\boldsymbol{x}}$；共轭对$\lambda$与$\overline{\lambda}$)。说明为什么每个 3×3 矩阵最少有一个实数特征值。

12 此处是一个关于每个实数矩阵的特征值都是实数的快速“证明”：

错误的证明 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 得到 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}=\lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 所以 $\lambda=\dfrac{\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}=\dfrac{\text{实数}}{\text{实数}}$

找出这项推理的错误--有一个隐藏的假设没有被证实。你可以在 90°旋转矩阵[0 –1; 1 0]测试这些步骤，其中旋转矩阵的$\lambda=i$ 与 $\boldsymbol{x}=(i,1)$。

13 把 S 与 B 写成光谱定理 $Q\Lambda Q^{\mathrm{T}}$ 的$\lambda_1\boldsymbol{x}_1\boldsymbol{x}_1{}^{\mathrm{T}}+\lambda_2\boldsymbol{x}_2\boldsymbol{x}_2{}^{\mathrm{T}}$ 的形式。

$$S=\begin{bmatrix}3&1\\1&3\end{bmatrix}\quad B=\begin{bmatrix}9&12\\12&16\end{bmatrix}\quad (\text{保持}\|\boldsymbol{x}_1\|=\|\boldsymbol{x}_2\|=1)。$$

14 每个 2×2 对称矩阵是$\lambda_1\boldsymbol{x}_1\boldsymbol{x}_1{}^{\mathrm{T}}+\lambda_2\boldsymbol{x}_2\boldsymbol{x}_2{}^{\mathrm{T}}=\lambda_1P_1+\lambda_2P_2$，说明 $P_1+P_2=\boldsymbol{x}_1\boldsymbol{x}_1{}^{\mathrm{T}}+\boldsymbol{x}_2\boldsymbol{x}_2{}^{\mathrm{T}}=I$，这是 Q 的列乘行。为什么 $P_1P_2=0$?

15 $A=\begin{bmatrix}0&b\\-b&0\end{bmatrix}$的特征值为何？创建一个 4×4 反对称矩阵($A^{\mathrm{T}}=-A$)，并验证它所有的特征值都是虚数。

16 (推荐) 下列矩阵 M 是反对称，也是_____，则所有的特征值是纯虚数，而且有$|\lambda|=1$。(对于每个 $\boldsymbol{x}$ 有$\|M\boldsymbol{x}\|=\|\boldsymbol{x}\|$，所以对特征向量有$\|\lambda\boldsymbol{x}\|=\|\boldsymbol{x}\|$。) 从 M 的迹求出全部 4 个特征值：

$$M=\frac{1}{\sqrt{3}}\begin{bmatrix}0&1&1&1\\-1&0&-1&1\\-1&1&0&-1\\-1&-1&1&0\end{bmatrix}\quad \text{只能有特征值 } i \text{ 与} -i$$

17 证明 A(对称但是复数)只有一条特征向量直线：

$$A=\begin{bmatrix}i&1\\1&-i\end{bmatrix}\ \text{甚至是不可对角化：特征值}\lambda=0,0$$

对于复数矩阵而言，$A^{\mathrm{T}}=A$ 不是什么特殊性质，好的性质是 $\overline{A}^{\mathrm{T}}=A$(段落 9.2)，则所有的$\lambda$'s 是实数且特征向量是正交。

18 纵使 A 是矩形，方块矩阵 $S=\begin{bmatrix}0&A\\A^{\mathrm{T}}&0\end{bmatrix}$是对称：

$$S\boldsymbol{x}=\lambda\boldsymbol{x}\quad \text{是}\quad \begin{bmatrix}0&A\\A^{\mathrm{T}}&0\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\lambda\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\quad \text{就是}\quad \begin{matrix}A\boldsymbol{z}=\lambda\boldsymbol{y}\\A^{\mathrm{T}}\boldsymbol{y}=\lambda\boldsymbol{z}\end{matrix}$$

(a) 证明$-\lambda$也是特征值，对应特征向量$(\boldsymbol{y}, -\boldsymbol{z})$。

(b) 证明 $A^{\mathrm{T}}A\boldsymbol{z} = \lambda^2\boldsymbol{z}$，所以$\lambda^2$ 是 $A^{\mathrm{T}}A$ 的特征值。

(c) 若 $A = I$ (2×2)，求出 S 的 4 个特征值与特征向量。

19 若问题 18 的 $A = \begin{bmatrix}1\\1\end{bmatrix}$，求出 S 的所有特征值与特征向量。

20 当 $S = S^{\mathrm{T}}$，特征向量互相垂直的另一个证明，两个步骤：

1. 假设 $S\boldsymbol{x} = \lambda\boldsymbol{x}$ 与 $S\boldsymbol{y} = 0\boldsymbol{y}$ 且$\lambda \neq 0$，则 $\boldsymbol{y}$ 在零空间中且 $\boldsymbol{x}$ 在列空间中，因为 _____，所以互相垂直。仔细研究--为什么这些子空间正交？

2. 若 $S\boldsymbol{y} = \beta\boldsymbol{y}$，应用前面的论点至 $S - \beta I$，$S - \beta I$ 的一个特征值会移到零，特征向量 $\boldsymbol{x}, \boldsymbol{y}$ 保持不变--所以他们互相垂直。

21 找出 S 的特征向量矩阵 Q，与 B 的特征向量矩阵 X。证明纵使$\lambda = 1$ 重复，在 $d = 1$ 时，X 不会崩溃。这些特征向量互相垂直吗？

$$S = \begin{bmatrix}0 & d & 0\\ d & 0 & 0\\ 0 & 0 & 1\end{bmatrix} \qquad B = \begin{bmatrix}-d & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & d\end{bmatrix} \qquad \text{有}\lambda = 1, d, -d$$

22 写成一个 2×2 复数矩阵使得 $\overline{S}^{\mathrm{T}} = S$ (“厄米矩阵”)，求出你的矩阵的λ_1 与λ_2。检查 $\overline{\boldsymbol{x}}_1^{\mathrm{T}}\boldsymbol{x}_2 = 0$ (这是复数正交性。)

23 是(说明理由)非(举出例子)题。

(a) 有实数特征值以及 n 个特征向量的矩阵是对称。

(b) 有实数特征值以及 n 个正交单位特征向量的矩阵是对称。

(c) 可逆对称矩阵的逆矩阵是对称。

(d) 对称矩阵的特征向量矩阵是对称。

24 (讲师的悖论) 若 $AA^{\mathrm{T}} = A^{\mathrm{T}}A$，则 A 与 A^{T} 共享相同的特征向量(正确)，A 与 A^{T} 共享相同的特征值。找出这项结论的错误：A 与 A^{T} 必须有相同的 X 与相同的 Λ，因此 A 等于 A^{T}。

25 (推荐) A 与 B 属于：可逆，正交，投影，排列，可对角化，马尔可夫矩阵？

$$A = \begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix} \qquad B = \frac{1}{3}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

A 与 B 可能的分解是：LU，QR，$X\Lambda X^{-1}$，$Q\Lambda Q^{\mathrm{T}}$ ？

26 矩阵 $A = \begin{bmatrix}2 & b\\ 1 & 0\end{bmatrix}$中什么样的 b 使得 $A = Q\Lambda Q^{\mathrm{T}}$ 是可能？什么数字使得 A 不能对角化？什么数字使得 A 奇异？

请尊重版权与译者的劳动成果，侵权必究!

27 找出所有正交且对称的 2×2 矩阵。哪两个数字可以是这两个矩阵的特征值？

28 这个 A 是近似对称，但是它的特征向量远离正交：

$$A=\begin{bmatrix}1 & 10^{-15}\\ 0 & 1+10^{-15}\end{bmatrix} \text{ 有固有向量 } \begin{bmatrix}1\\0\end{bmatrix} \text{ 与 } [\,?\,]$$

特征向量的夹角是多少？

29 (MATLAB) 选取两个有不同特征向量的对称矩阵，例如 $A=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$ 与 $B=\begin{bmatrix}8 & 1\\ 1 & 0\end{bmatrix}$。画出特征值$\lambda_1(A+tB)$与$\lambda_2(A+tB)$的图形，其中$-8<t<8$。Peter Lax 在线性代数 113 页说在某个时间 t，λ_1 与λ_2 会出现可能的碰撞，“在最后一分钟他们会偏离。”λ_1 与λ_2 会有多么接近？

挑战问题

30 对于复数矩阵，对称性 $S=S^{\mathrm{T}}$ 得到实数特征值，在段落 9.2 必须改成 $\overline{S}^{\mathrm{T}}=S$。从 $\det(S-\lambda I)=0$，找出 2×2 的厄米矩阵 $S=[4 \quad 2+i\,;\,2-i \quad 0]=\overline{S}^{\mathrm{T}}$ 的特征值.

31 正态矩阵有 $\overline{N}^{\mathrm{T}}N=N\overline{N}^{\mathrm{T}}$。对实数矩阵来说，这是 $\overline{N}^{\mathrm{T}}N=N\overline{N}^{\mathrm{T}}$，正态包含对称，反对称与正交(实数$\lambda$，虚数$\lambda$，与$|\lambda|=1$)，其他正态矩阵可以有任意的复数特征值。

关键点：正态矩阵有 n 个正交单位特征向量，这些向量 $\boldsymbol{x}_i$ 可能有复数分量。在复数情况下(第九章)，正交性表示 $\overline{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{x}_j=0$，内积(点积)$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 变成 $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{y}$。

Q 的 n 个正交单位列的测试变成 $\overline{Q}^{\mathrm{T}}Q=I$，而不是 $Q^{\mathrm{T}}Q=I$。

N 有 n 个正交单位特征向量($N=Q\Lambda\overline{Q}^{\mathrm{T}}$)当且仅当 N 是正态。

(a) 从 $N=Q\Lambda\overline{Q}^{\mathrm{T}}$ 开始，且 $\overline{Q}^{\mathrm{T}}Q=I$，证明 $\overline{N}^{\mathrm{T}}N=N\overline{N}^{\mathrm{T}}$：$N$ 是正态。
(b) 现在从 $\overline{N}^{\mathrm{T}}N=N\overline{N}^{\mathrm{T}}$ 开始，舒尔发现对每个矩阵 A，$A=QT\overline{Q}^{\mathrm{T}}$，其中 T 是三角形。对于正态矩阵 $A=N$，我们必须证明(三个步骤)三角形矩阵 T 实际上是对角矩阵，则 $T=\Lambda$。

步骤 1. 把 $N=Q\Lambda\overline{Q}^{\mathrm{T}}$ 放进 $\overline{N}^{\mathrm{T}}N=N\overline{N}^{\mathrm{T}}$ 找出 $\overline{T}^{\mathrm{T}}T=T\overline{T}^{\mathrm{T}}$。

步骤 2. 假设 $T=\begin{bmatrix}a & b\\ 0 & d\end{bmatrix}$ 有 $\overline{T}^{\mathrm{T}}T=T\overline{T}^{\mathrm{T}}$，证明 $b=0$。

步骤 3. 延伸步骤 2 到大小为 n，任意正态三角 T 必须是对角矩阵。

32 若λ_{max}是对称矩阵 S 的最大特征值，没有对角单元比λ_{max}大。$S = Q\Lambda Q^{T}$ 的第一个单元 a_{11} 为何？说明为什么 $a_{11} \le \lambda_{max}$？

33 假设 $A^{T} = -A$(实数反对称矩阵)，说明 A 的这些事实：

(a) 对于每个实数向量 $\boldsymbol{x}$，都有 $\boldsymbol{x}^{T}A\boldsymbol{x} = 0$。

(b) A 的特征值是纯虚数。

(c) A 的行列式是正或零(非负)。

对(a)，实际计算一个 $\boldsymbol{x}^{T}A\boldsymbol{x}$，观察项的消除，或是逆反 $\boldsymbol{x}^{T}(A\boldsymbol{x})$变成$-(A\boldsymbol{x})^{T}\boldsymbol{x}$。对(b)，$A\boldsymbol{z} = \lambda\boldsymbol{z}$ 得到 $\bar{\boldsymbol{z}}^{T}A\boldsymbol{z} = \lambda\,\bar{\boldsymbol{z}}^{T}\boldsymbol{z} = \lambda\|\boldsymbol{z}\|^{2}$。(a) 显示 $\bar{\boldsymbol{z}}^{T}A\boldsymbol{z} = (\boldsymbol{x} - i\boldsymbol{y})^{T}A(\boldsymbol{x} + i\boldsymbol{y})^{T}$ 有零的实数部分，则(b)可以帮助(c)。

34 若 S 是对称且所有的特征值是$\lambda = 2$，你如何知道 S 必定是 $2I$？关键点：对称确保 $S = Q\Lambda Q^{T}$，Λ为何？

35 哪些对称矩阵 S 也是正交？则 $S^{T} = S^{-1}$。

(a) 说明对称与正交得到 $S^{2} = I$。

(b) S 可能的特征值为何？

(c) 矩阵Λ可能的特征值为何？对于这些Λ与任意正交 Q，S 必须是 $Q\Lambda Q^{T}$。

36 若 S 是对称，证明 $A^{T}SA$ 也是对称(取 $A^{T}SA$ 的转置)。此处的 A 是 $m\times n$ 且 S 是 $m\times m$，S 的特征值 $= A^{T}SA$ 的特征值吗？

如果 A 是方形可逆，$A^{T}SA$ 称为“全等于(congruent to)”S。他们有相同的正特征值的个数，负特征值的个数与零特征值的个数：**惯性法则(law of inertia)。**

37 有一个方法证明 a 在 S 的特征值λ_1与λ_2之间：

$$S = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \qquad \det(S - \lambda I) = \lambda^{2} - a\lambda - c\lambda + ac - b^{2}$$

是开口朝上的抛物线(因为 λ^{2})

证明在$\lambda = a$ 时，$\det(S - \lambda I)$是负数，所以抛物线在$\lambda = a$ 的左侧与右侧跨越轴，跨越点是 S 的两个特征值，所以他们必须包住 a。

A 的 $n-1$ 个特征值永远落在 $S = \begin{bmatrix} A & \boldsymbol{b} \\ \boldsymbol{b}^{T} & c \end{bmatrix}$ 的 n 个特征值之间。

请尊重版权与译者的劳动成果，侵权必究!

6.5 正定矩阵

1 对称 S：所有特征值 >0 ⇔ 所有主元 >0 ⇔ 所有的**左上行列式** >0。

2 矩阵 S 是**正定**(positive definite)，能量测试是对所有的 $\boldsymbol{x}\neq 0$ 都有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$。

3 正定的另一个测试：$S=A^{\mathrm{T}}A$，且 A 有无关的列。

4 正半定 S 允许 $\lambda=0$，主元$=0$，行列式$=0$，能量 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=0$。

5 当 S 是对称正定，方程式 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=1$ 得到$\mathbf{R}^n$的一个椭圆。

本段落聚焦在具有**正数特征值**的对称矩阵，如果对称性使得矩阵重要，这个额外的性质(全部$\lambda>0$)使得矩阵更特殊。我们说特殊并不代表稀有，有正数特征值的对称矩阵是所有应用的中心，他们称为**正定**。

第一个问题是确认正定矩阵，你可以说就是找出所有的特征值然后测试$\lambda>0$，这正是我们想要避免的。计算特征值很费劲，当我们需要λ's 时我们可以计算他们，但是如果我们只是想要知道全部的λ's 是正数，有很多快速的方法。下列是本段落的两个目标：

- 找出对于对称矩阵确保正数更重要的快速测试法。
- 说明正定性的重要应用。

因为矩阵是对称，每个特征值都是实数。

从 2×2 开始，什么时候 $S=\begin{bmatrix}a & b\\ b & c\end{bmatrix}$ 有$\lambda_1>0$ 且$\lambda_2>0$ ？

测试：S 的特征值是正数当且仅当 $a>0$ 且 $ac-b^2>0$。

$S_1=\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$不是正定，因为 $ac-b^2=1-4<0$。

$S_2=\begin{bmatrix}1 & -2\\ -2 & 6\end{bmatrix}$是正定，因为 $a=1$ 且 $ac-b^2=6-4<0$。

$S_3=\begin{bmatrix}-1 & 2\\ 2 & -6\end{bmatrix}$不是正定(纵使 $\det A=+2$)，因为 $a=-1$。

S_1 的特征值 3 与−1 确认 S_1 不是正定，正数迹 $3-1=2$，但是负数行列式$(3)(-1)=-3$。$S_3=-S_2$ 是负定，S_2 有两个正数特征值，S_3 有两个负数特征值。

当$\lambda_1>0$ 且$\lambda_2>0$，证明可以通过 2×2 的测试。他们的乘积$\lambda_1\lambda_2$是行列式所以 $ac-b^2>0$，他们的总和$\lambda_1+\lambda_2$是迹，所以 $a+c>0$。

则 a 与 c 同时是正数(若 a 或 c 不是正数，$ac-b^2>0$ 会失败)。问题 1 逆反推理证明 $a>0$ 与 $ac>b^2$ 的测试，可以确保$\lambda_1>0$ 且$\lambda_2>0$。

这项测试使用了 1×1 行列式 a 与 2×2 行列式 $ac-b^2$，当 S 是 3×3，$\det S>0$ 是测试的第三部分。

测试：S 的特征值是正数当且仅当主元是正数：

$$a>0 \text{ 且 } \frac{ac-b^2}{a}>0$$

$a>0$ 在两个测试中都是必要的，所以对于行列式测试以及现在的主元测试，$ac>b^2$ 也是必要的。重点是确认 S 的第二主元的比值：

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \begin{matrix} \text{第一枢轴是 } a \\ \rightarrow\rightarrow\rightarrow \\ \text{乘数是 } b/a \end{matrix} \quad \begin{bmatrix} a & b \\ 0 & c-\frac{b}{a}b \end{bmatrix} \quad \begin{matrix} \text{第二枢轴是} \\ c-\frac{b^2}{a}=\frac{ac-b^2}{a} \end{matrix}$$

这个联结了线性代数的两大部分，正数特征值代表正数主元，反之亦然。每个主元是左上角行列式的比值，对于$\lambda>0$，主元提供了一个快速测试，他们比计算特征值要快很多。看到主元与行列式与特征值在课程中走在一起，心情非常满足。

3×3 范例 $S=\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ 是正定 $\quad \begin{matrix} \text{固有值 1 与 1 与 4} \\ \text{行列式 2 与 3 与 4} \\ \text{枢轴 2 与 3/2 与 4/3} \end{matrix}$

$S-I$ 会是正半定：特征值 0, 0, 3。$S-2I$ 是未定(indefinite)因为$\lambda=-1,-1,2$。

现在用不同的方式检视具有正数特征值的对称矩阵。

能量基定义

从 $S\boldsymbol{x}=\lambda\boldsymbol{x}$，左乘 $\boldsymbol{x}^{\mathrm{T}}$ 得到 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，右侧是正数$\lambda$乘一个正数 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\|\boldsymbol{x}\|^2$，所以对于任意的特征向量，左侧的 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是正数。

重点： 新的概念是对于所有非零向量 $\boldsymbol{x}$，都有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是正数，不只是特征向量。在很多应用中，数字 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$(或 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}/2$)是系统的能量(energet)。正能量的必要性给出正定矩阵的另一个定义，我想这个能量基(energy-based)的定义是基本的方式。

特征值与主元是测试新要求 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$ 的两个等效方法。

请尊重版权与译者的劳动成果，侵权必究!

定义 若对于所有非零向量 $\boldsymbol{x}$ 都有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$，则 S 是正定：

2×2
$$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\begin{bmatrix}x & y\end{bmatrix}\begin{bmatrix}a & b\\ b & c\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=ax^2+2bxy+cy^2>0 \tag{1}$$

4 个单元 a, b, b, c 给出 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 的 4 个部分，从 a 与 c 得到纯平方 ax^2 与 cy^2，从非对角线的 b 与 b 得到交叉项 bxy 与 byx(相同)。把 4 个部分相加得到 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$，这个能量基定义导出一个基本事实：

若 S 与 T 是对称正定，$S+T$ 也是。

理由： $\boldsymbol{x}^{\mathrm{T}}(S+T)\boldsymbol{x}$ 就是 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}T\boldsymbol{x}$，这两项是正数(当 $\boldsymbol{x}\neq\boldsymbol{0}$)所以 $S+T$ 也是正定。当矩阵相加时，主元与特征值不容易追踪，但是能量只需要相加。

$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 也可以联结我们确认正定矩阵的 4 种方法，从任意矩阵 A 开始，可能是矩形。我们知道 $S=A^{\mathrm{T}}A$ 是方形且对称，不止如此，当 A 有无关列时 S 是正定：

测试：若 A 的列是无关，则 $S=A^{\mathrm{T}}A$ 是正定。

再次提出特征值与主元并不容易，但是数字 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 与 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}$ 相同，$\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}$ 就是 $(A\boldsymbol{x})^{\mathrm{T}}(A\boldsymbol{x})=\|A\boldsymbol{x}\|^2$——另一个利用括号的重要证明！当 $\boldsymbol{x}\neq\boldsymbol{0}$，向量 $A\boldsymbol{x}$ 不是零(这是无关列的意义)，则 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是正数$\|A\boldsymbol{x}\|^2$ 且矩阵 S 是正定。

让我把原理收集起来，变成关于正定的 5 个等效叙述。你会看到关键概念如何联结线性代数的整个主题：主元，行列式，特征值，以及最小平方(从 $A^{\mathrm{T}}A$)，然后是应用。

当一个对称矩阵 S 有下列 5 种性质之一，就会有全部的性质：

1. S 的所有 n **个主元**都是正数。
2. 所有 n **个左上行列式**都是正数。
3. S 的所有 n **个特征值**都是正数。
4. 除了 $\boldsymbol{x}=\boldsymbol{0}$ 之外，$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是正数，这是**能量基**的定义。
5. $S=A^{\mathrm{T}}A$，其中矩阵 A 有**无关的列**。

“左上行列式”是 1×1，2×2，...，$n\times n$，最后一个是全部的矩阵 S 的行列式。这个定理把全部线性代数课程都结合在一起。

范例 1 测试下列对称矩阵 S 与 T 的正定性：

$$S=\begin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}\qquad 与\qquad T=\begin{bmatrix}2&-1&b\\-1&2&-1\\b&-1&2\end{bmatrix}$$

解 S 的主元是 2, 3/2, 4/3 都是正数。它的左上行列式是 2, 3, 4 都是正数。S 的特征值是 $2-\sqrt{2}$，2，$2+\sqrt{2}$ 都是正数。完成了测试 1, 2, 3，任何一种都有决定性！

我有三个候选矩阵 A_1, A_2 与 A_3 作为 $S = A^{\mathrm{T}}A$ 的建议，他们都显示 S 是正定。A_1 是第一差分矩阵，4×3，产生 S 的–1, 2, –1：

$$S=A_1{}^{\mathrm{T}}A_1\qquad \begin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}=\begin{bmatrix}1&-1&0&0\\0&1&-1&0\\0&0&1&-1\end{bmatrix}\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\\0&0&-1\end{bmatrix}$$

A_1 的三个列是无关，所以 S 正定。

A_2 来自 $S = LDL^{\mathrm{T}}$ ($S = LU$ 的对称版本)，消元法得到主元是 D 的 2, 3/2, 4/3，以及乘数是 L 的–1/2, 0, –2/3。**就让 $A_2=(L\sqrt{D})^{\mathrm{T}}$**。【原文少了“( )^{T}”，错误】

$$LDL^{\mathrm{T}}=\begin{bmatrix}1&&\\-\frac{1}{2}&1&\\0&-\frac{2}{3}&1\end{bmatrix}\begin{bmatrix}2&&\\&\frac{3}{2}&\\&&\frac{4}{3}\end{bmatrix}\begin{bmatrix}1&-\frac{1}{2}&0\\&1&-\frac{2}{3}\\&&1\end{bmatrix}=(L\sqrt{D})(L\sqrt{D})^{\mathrm{T}}=A_2{}^{\mathrm{T}}A_2$$

A_2 是 S 的乔列斯基因子

A 的三角选择出现平方根(不美丽)，它是 S 的**乔列斯基**因子且 MATLAB 的指令是 A = chol(S)。在应用中，矩形 A_1 是我们如何建立 S，**乔列斯基** A_2 是我们如何把它分离。

特征值给出对称选择 $A_3=Q\sqrt{\Lambda}Q^{\mathrm{T}}$ 也是可以成功，其中 $A_3{}^{\mathrm{T}}A_3=Q\Lambda Q^{\mathrm{T}}=S$。所有的测试显示–1, 2, –1 矩阵 S 是正定。

要看出能量 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是正数，我们改写成平方的总和。三个选择 A_1, A_2 与 A_3 产生三种不同的方法来分割 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$：

$$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=2x_1{}^2-2x_1x_2+2x_2{}^2-2x_2x_3+2x_3{}^2 \qquad \textbf{改写成平方}$$

$$\|A_1\boldsymbol{x}\|^2=x_1{}^2+(x_2-x_1)^2+(x_3-x_2)^2+x_3{}^2 \qquad \textbf{使用 } \boldsymbol{A_1} \textbf{ 的差}$$

$$\|A_2\boldsymbol{x}\|^2=2\left(x_1-\tfrac{1}{2}x_2\right)^2+\tfrac{3}{2}\left(x_2-\tfrac{2}{3}x_3\right)^2+\tfrac{4}{3}x_3{}^2 \qquad \textbf{使用 } \boldsymbol{S=LDL^{\mathrm{T}}}$$

$$\|A_3\boldsymbol{x}\|^2=\lambda_1\left(\boldsymbol{q}_1{}^{\mathrm{T}}\boldsymbol{x}\right)^2+\lambda_2\left(\boldsymbol{q}_2{}^{\mathrm{T}}\boldsymbol{x}\right)^2+\lambda_3\left(\boldsymbol{q}_3{}^{\mathrm{T}}\boldsymbol{x}\right)^2 \qquad \textbf{使用 } \boldsymbol{S=Q\Lambda Q^{\mathrm{T}}}$$

现在转到 T (本页的上方)，(1, 3)与(3, 1)单元从 0 变成 b，b 必须不能太大！行列式的测试很容易。1×1 行列式是 2，2×2 行列式 ~~T~~还是 3。3×3 行列式与 b 有关：

【原文不该出现“T”】

对 T 的测试 $\det T=4+2b-2b^2=(1+b)(4-2b)$ 必须是正数

在 b = –1 与 b = 2 我们得到 det T = 0。在 b = –1 与 b = 2 之间，矩阵 T 是正定。S 矩阵的角落单元 b = 0 在–1 与 2 之间，所以是安全的。

请尊重版权与译者的劳动成果，侵权必究！

正半定矩阵

通常我们会落在正定的边缘，行列式是零，最小的特征值是 0，它的特征向量的能量是 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}0\boldsymbol{x}=0$。这些在边缘的矩阵称作正半定(semidefinite)，下列是两个例子(不可逆) ：

$$S=\begin{bmatrix}1&2\\2&4\end{bmatrix}\quad 与\quad T=\begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix}\ 是正半定。$$

S 有特征值 5 与 0，它的左上行列式是 1 与 0，它的秩只有 1。这个矩阵 S 分解成 $A^{\mathrm{T}}A$，其中 A 有相关的列：

$$\begin{matrix}A\text{ 的相关列}\\ \text{正半定 }S\end{matrix}\qquad \begin{bmatrix}1&2\\2&4\end{bmatrix}=\begin{bmatrix}1&0\\2&0\end{bmatrix}\begin{bmatrix}1&2\\0&0\end{bmatrix}=A^{\mathrm{T}}A$$

如果 4 稍微变大一点点，矩阵 S 变成正定。

循环 T 也有零行列式 (当 $b=-1$，前面计算过)，T 是奇异。特征向量 $\boldsymbol{x}=(1,1,1)$有 $T\boldsymbol{x}=\boldsymbol{0}$ 以及能量 $\boldsymbol{x}^{\mathrm{T}}T\boldsymbol{x}=0$，向量 $\boldsymbol{x}$ 在其他方向确实有正的能量。这个 T 有很多方式可以改写成 $A^{\mathrm{T}}A$，但是 A 一定会有相关的列，因为(1, 1, 1)在它的零空间：

$$\begin{matrix}\text{第二差 }T\\ \text{来自第一差 }A\\ \text{循环 }T\text{ 来自循环 }A\end{matrix}\qquad \begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix}=\begin{bmatrix}1&-1&0\\0&1&-1\\-1&0&1\end{bmatrix}\begin{bmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{bmatrix}$$

正半定矩阵的所有 $\lambda\ge 0$ 且所有的 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}\ge 0$，这些弱不等式($\ge$ 而不是 $>$)包含正定 S 与边缘的奇异矩阵。

椭圆 $ax^2+2bxy+cy^2=1$

思考如图 6.7 所示倾斜的椭圆 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=1$，它的中心是(0, 0)。把椭圆旋转与坐标轴对齐 (X 与 Y 轴)，这是图 6.7b。这两个图显示 $S=Q\Lambda Q^{-1}=Q\Lambda Q^{\mathrm{T}}$ 的分解背后的几何意义：

1. 标题的椭圆对应 S，它的方程式是 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=1$。
2. 对齐的椭圆对应Λ，它的方程式是 $X^{\mathrm{T}}\Lambda X=1$
3. 使得椭圆对齐的旋转矩阵是特征向量矩阵 Q。

范例 2　求出倾斜椭圆 $5x^2+8xy+5y^2=1$ 的轴。

解　从适配这个方程式的正定矩阵开始：

$$\text{方程式是 }\begin{bmatrix}x&y\end{bmatrix}\begin{bmatrix}5&4\\4&5\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=1\qquad \text{矩阵是 }S=\begin{bmatrix}5&4\\4&5\end{bmatrix}$$

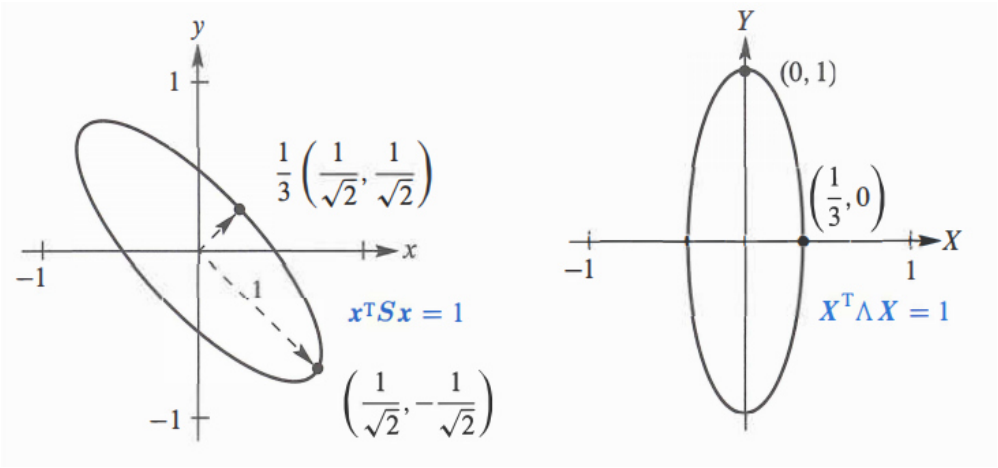


图 6.7：标题椭圆 $5x^2 + 8xy + 5y^2 = 1$，对齐后是 $9X^2 + Y^2 = 1$。

特征向量是$\begin{bmatrix}1\\1\end{bmatrix}$与$\begin{bmatrix}1\\-1\end{bmatrix}$，除以$\sqrt{2}$得到单位向量，则 $S = Q\Lambda Q^T$:

Q的固有向量
固有值9与1

$$\begin{bmatrix}5 & 4\\4 & 5\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}\begin{bmatrix}9 & 0\\0 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}$$

现在左乘$[x \quad y]$再右乘$\begin{bmatrix}x\\y\end{bmatrix}$得到 $\boldsymbol{x}^T S\boldsymbol{x} = (\boldsymbol{x}^T Q)\Lambda(Q^T\boldsymbol{x})$ ：

$$\boldsymbol{x}^T S\boldsymbol{x} \text{=平方的总和} \qquad 5x^2 + 8xy + 5y^2 = 9\left(\frac{x+y}{\sqrt{2}}\right)^2 + 1\left(\frac{x-y}{\sqrt{2}}\right)^2 \tag{2}$$

系数不是来自 D 的主元 5 与 9/5，他们是来自Λ的特征值 9 与 1。在平方里面的是特征向量 $\boldsymbol{q}_1 = (1, 1)/\sqrt{2}$ 与 $\boldsymbol{q}_2 = (1, -1)/\sqrt{2}$。

倾斜椭圆的轴指向特征向量。这个解释了为什么 $S = Q\Lambda Q^T$ 称为“主轴定理”——它显示了轴，不只是轴方向(来自特征向量)，还有轴长度(来自特征值)。要看见全貌，使用大写字母作为对齐椭圆的新坐标：

$$\textbf{对齐} \qquad \frac{x+y}{\sqrt{2}} = X \quad \text{与} \quad \frac{x-y}{\sqrt{2}} = Y \quad \text{与} \quad 9\boldsymbol{X}^2 + \boldsymbol{Y}^2 = \boldsymbol{1}$$

X^2 的最大值是 1/9，较短轴的端点有 $X = 1/3$ 与 $Y = 0$。

注意： 较大的特征值λ_1得到短轴，一半长度是 $1/\sqrt{\lambda_1} = 1/3$。较小特征值$\lambda_2 = 1$ 得到较大的长度 $1/\sqrt{\lambda_2} = 1$。

在 xy 系统中，轴是沿着 S 的特征向量。在 XY 系统，**轴是沿着Λ的特征向量**——坐标轴。全部都来自 $S = Q\Lambda Q^T$。

请尊重版权与译者的劳动成果，侵权必究!

当所有的$\lambda_i > 0$，$S = Q\Lambda Q^{\mathrm{T}}$ 是正定，$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = 1$ 的图形是个椭圆：

$$\textbf{椭圆} \quad [x \quad y]\, Q\Lambda Q^{\mathrm{T}}\begin{bmatrix} x \\ y \end{bmatrix} = [X \quad Y]\,\Lambda\begin{bmatrix} X \\ Y \end{bmatrix} = \boldsymbol{\lambda_1 X^2 + \lambda_2 Y^2 = 1} \tag{3}$$

轴指向 S 的特征向量，椭圆的半长是 $1/\sqrt{\lambda_1}$ 与 $1/\sqrt{\lambda_2}$ 。

$S = I$ 得到圆 $x^2 + y^2 = 1$。如果一个特征值是负数(交换 S 的 4's 与 5's)，椭圆变成双曲线，平方总和变成平方的差：$9X^2 - Y^2 = 1$。例如 $S = -I$ 这种负定矩阵，两个λ's 都是负数，$-x^2 - y^2 = 1$ 的图形就是一个点都没有。

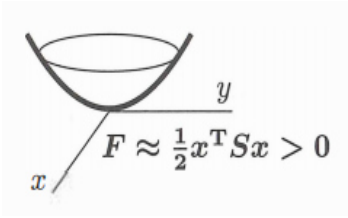
若 S 是 $n\times n$，$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = 1$ 是$\mathbf{R}^n$的一个椭球面，它的轴是 S 的特征向量。

重要应用：极小值测试

若在点$(x, y) = (0, 0)$时，$\partial F / \partial x = 0$ 且 $\partial F / \partial y = 0$，$F(x, y)$有极小值吗？

对 $f(x)$来说，极小值的测试来自微积分：df/dx 是零且 $d^2f/dx^2 > 0$。$F(x, y)$的两个变数产生一个对称矩阵 S，它包含 4 个二次导数，**正数 d^2f/dx^2 变成正定 S：**

二次导数 $$S = \begin{bmatrix} \partial^2 F/\partial x^2 & \partial^2 F/\partial x\partial y \\ \partial^2 F/\partial y\partial x & \partial^2 F/\partial y^2 \end{bmatrix}$$



若 $\partial F / \partial x = \partial F / \partial y = 0$ 且 S 是正定，则 $F(x, y)$有一个极小值。

理由： S 显示靠近$(x, y) = (0, 0)$的最重要项 $ax^2 + 2bxy + cy^2$，F 的二次导数是 $2a$, $2b$, $2b$, $2c$。对于 $F(x, y, z)$的 S 矩阵会是 3×3。

主要观念的复习

1. 正定矩阵有正数特征值与正数主元。
2. 左上行列式是一个快速测试：$a > 0$ 且 $ac - b^2 > 0$。
3. 能量 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 是一个从 $\boldsymbol{x} = \mathbf{0}$ 朝上的“碗(bowl)”

   除了在$(x, y) = (0, 0)$之外，$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = ax^2 + 2bxy + cy^2$ 是一个正数。
4. 若 A 有无关的列，则 $S = A^{\mathrm{T}}A$ 自动是正定。
5. 椭圆 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = 1$ 的轴沿着 S 的特征向量，长度 $1/\sqrt{\lambda}$。
6. 若 $\partial F / \partial x = \partial F / \partial y = 0$ 且二次导数矩阵是正定，则 $F(x, y)$有一个极小值。

已解范例

6.5A 对称矩阵的伟大分解是来自主元与乘数的 $S = LDL^{\mathrm{T}}$，以及来自特征值与特征向量的 $S = Q\Lambda Q^{\mathrm{T}}$。尝试对于 $n \times n$ 的测试，例如 pascal(6), ones(6), hilb(6)以及其他 MATLAB 展览馆的矩阵。

pascal(6) 是正定，因为所有的主元都是 1(已解范例 **2.6A**)。

ones(6) 是正半定，因为它的特征值是 0, 0, 0, 0, 0, 6。

H = hilb(6) 是正定，纵使 eig(H)显示特征值非常靠近零。

希尔伯特矩阵 $\boldsymbol{x}^{\mathrm{T}}H\boldsymbol{x} = \int_0^1 \left(x_1 + x_2 s + \ldots + x_6 s^5\right)^2 ds > 0$，$H_{ij} = 1/(i + j - 1)$。

rand(6) + rand(6)′ 可以是正定也可以不是，试验显示 20000 次只有 2 次是正定。

n = 20000; p = 0; for k = 1:n, A = rand(6); p = p + all(eig(A + A') > 0); end, p/n

6.5B 什么时候方块矩阵 $M = \begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix}$ 是正定？

解 从 M 的第二行减去 $B^{\mathrm{T}}A^{-1}$ 乘第一行得到方块零，舒尔补充 $S = C - B^{\mathrm{T}}A^{-1}B$ 会出现在角落：

$$\begin{bmatrix} I & 0 \\ -B^{\mathrm{T}}A^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & C - B^{\mathrm{T}}A^{-1}B \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & S \end{bmatrix} \tag{4}$$

这两个方块 A 与 S 必须是正定，他们的主元是 M 的主元。

6.5C 求出−1, 2, −1 三对角 $n \times n$ 矩阵 S (我的最爱) 的特征值。

解 最佳方法是猜测λ与 $\boldsymbol{x}$，然后检验 $S\boldsymbol{x} = \lambda\boldsymbol{x}$。对大部分矩阵来说可能猜不了，但是特殊案例是数学的一大部分(纯数学与应用)。

关键隐藏在一个微分方程式，第二差分矩阵 S 好像是二次导数，这些特征值更容易了解：

$$\begin{array}{ll} \text{固有值 } \lambda_1,\ \lambda_2, \ldots. & \\ \text{固有函数 } y_1,\ y_2, \ldots. & \end{array} \qquad \frac{d^2y}{dx^2} = \lambda y(x) \quad \text{且} \quad \begin{array}{l} y(0) = 0 \\ y(1) = 0 \end{array} \tag{5}$$

尝试 $y = \sin cx$，它的二次导数是 $y'' = -c^2 \sin cx$。假设 $y = \sin cx$ 满足端点条件 $y(0) = 0 = y(1)$，方程式(5)中的特征值会是$\lambda = -c^2$，

当然 sin 0 = 0(其中余弦被消除)，另一个端点 $x = 1$，我们需要 $y(1) = \sin c = 0$，数字 c 必须是 $k\pi$，一个π的倍数。则λ是$-k^2\pi^2$：

$$\begin{array}{ll} \text{固有值 } \lambda = -k^2\pi^2 & \\ \text{固有函数 } y = \sin k\pi x & \end{array} \qquad \frac{d^2}{dx^2} \sin k\pi x = -k^2\pi^2 \sin k\pi x \tag{6}$$

现在回到矩阵 S 并且猜测它的特征值，在 n 个点 $x = h, 2h, \ldots, nh$，$\sin k\pi x$ 是 0 与 1 之间的相等等空间点，距离Δx 是 $h = 1/(n + 1)$，所以第 $n + 1$ 个点是$(n + 1)h = 1$。把 S 乘正弦向量 $\boldsymbol{x}$：

$$\begin{array}{ll} S \text{的固有值是正数} & S\boldsymbol{x} = \lambda_k \boldsymbol{x} = \left(2 - 2\cos k\pi h\right)\boldsymbol{x} \\ S \text{的固有向量是正弦向量} & \boldsymbol{x} = \left(\sin k\pi h, \ldots, \sin nk\pi h\right) \end{array} \tag{7}$$

请尊重版权与译者的劳动成果，侵权必究!

问题集 6.5

问题 1-13 有关正定的测试。

1 假设通过 2×2 测试 $a>0$ 与 $ac-b^2>0$，则 $c>b^2/a>0$.

(i) λ_1 与 λ_2 有相同的符号，因为乘积 $\lambda_1\lambda_2$ 等于____。

(ii) 上面的符号是正，因为 $\lambda_1+\lambda_2$ 等于____。

结论：测试 $a>0$ 与 $ac-b^2>0$ 确保正数特征值 λ_1 与 λ_2。

2 S_1, S_2, S_3, S_4 哪一个矩阵有两个正数特征值？ 使用测试法，不要计算 λ's。同时求出一个 $\boldsymbol{x}$ 使得 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}<0$，所以 S_1 不是正定。

$$S_1=\begin{bmatrix}5&6\\6&7\end{bmatrix}\quad S_2=\begin{bmatrix}-1&-2\\-2&-5\end{bmatrix}\quad S_3=\begin{bmatrix}1&10\\10&100\end{bmatrix}\quad S_4=\begin{bmatrix}1&10\\10&101\end{bmatrix}$$

3 什么数字 b 与 c 使得下列矩阵是正定？

$$S=\begin{bmatrix}1&b\\b&9\end{bmatrix}\quad S=\begin{bmatrix}2&4\\4&c\end{bmatrix}\quad S=\begin{bmatrix}c&b\\b&c\end{bmatrix}$$

把主元放在 D 与乘数放在 L，把每个 A 分解成 LDL^{T}。

4 下列矩阵的函数 $f=ax^2+2bxy+cy^2$ 为何？完成平方项使得每个 f 写成一个或两个平方的总和 $f=d_1(\quad)^2+d_2(\quad)^2$。

$$S_1=\begin{bmatrix}1&2\\2&9\end{bmatrix}\quad S_2=\begin{bmatrix}1&3\\3&9\end{bmatrix}\qquad f=[x\ \ y]\,[\,S\,]\begin{bmatrix}x\\y\end{bmatrix}$$

5 把 $f(x,y)=x^2+4xy+3y^2$ 写成平方的差，并求出一个点 (x,y) 使得 f 是负数。纵使 f 有正数系数，在 $(0,0)$ 还是没有极小值。

6 函数 $f(x,y)=2xy$ 确定有一个鞍点(saddle point)并且在 $(0,0)$ 没有极小值。什么样的矩阵 S 得到这个 f？特征值为何？

7 对于下列矩阵测试 $A^{\mathrm{T}}A$ 是否正定？A 需要无关的列。

$$A=\begin{bmatrix}1&2\\0&3\end{bmatrix}\quad 与\quad A=\begin{bmatrix}1&1\\1&2\\2&1\end{bmatrix}\quad 与\quad A=\begin{bmatrix}1&1&2\\1&2&1\end{bmatrix}$$

8 函数 $f(x,y)=3(x+2y)^2+4y^2$，除了在 $(0,0)$ 之外都是正数。$f=[x\quad y]\,S\,[x\quad y]^{\mathrm{T}}$ 的矩阵为何？检查 A 的主元是 3 与 4。

9　求出 3×3 矩阵 S 以及它的主元，秩，特征值，行列式：

$$[x_1 \quad x_2 \quad x_3]\,[\,S\,]\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=4(x_1-x_2+2x_3)^2$$

10　哪个 3×3 对称矩阵 S 与 T 产生下列二次式：

$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}=2(x_1^2+x_2^2+x_3^2-x_1x_2-x_2x_3)$，为什么 S 是正定？

$\boldsymbol{x}^{\mathrm{T}}T\boldsymbol{x}=2(x_1^2+x_2^2+x_3^2-x_1x_2-x_1x_3-x_2x_3)$，为什么 T 是正半定？

11　计算 S 的三个左上行列式建立正定性。验证他们的比值得到第二与第三主元。

主元 = 行列式的比值　　$S=\begin{bmatrix}2&2&0\\2&5&3\\0&3&8\end{bmatrix}$

12　哪些数字 c 与 d 使得 S 与 T 正定？测试他们的三个行列式：

$$S=\begin{bmatrix}c&1&1\\1&c&1\\1&1&c\end{bmatrix}\quad 与 \quad T=\begin{bmatrix}1&2&3\\2&d&4\\3&4&5\end{bmatrix}$$

13　给定 $a>0$ 与 $c>0$ 与 $a+c>2b$，找出有一个负数特征值的矩阵。

问题 14-20 有关测试的应用。

14　若 S 是正定则 S^{-1} 是正定。最佳证明：S^{-1} 的特征值是正数，因为_____。

第二种证明(只能是 2×2)：

$S^{-1}=\dfrac{1}{ac-b^2}\begin{bmatrix}c&-b\\-b&a\end{bmatrix}$ 的单元通过行列式测试_____。

15　若 S 与 T 是正定，他们的总和 $S+T$ 也是正定，主元与特征值对 $S+T$ 不方便，最好使用 $\boldsymbol{x}^{\mathrm{T}}(S+T)\,\boldsymbol{x}>0$。同时 $S=A^{\mathrm{T}}A$ 与 $T=B^{\mathrm{T}}B$ 得到 $S+T=[A\ B]^{\mathrm{T}}\begin{bmatrix}A\\B\end{bmatrix}$ 具有无关列。

16　正定矩阵不能有一个零(更糟的是有一个负数)在对角线，证明这种矩阵不可能有 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$:

当$(x_1,x_2,x_3)=(\quad,\quad,\quad)$，则$[x_1 \quad x_2 \quad x_3]\begin{bmatrix}4&1&1\\1&0&2\\1&2&5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$不是正定。

17　一个对称矩阵的某个对角线单元 s_{jj} 不会比所有的λ's 小，如果它是，则 $S-s_{jj}I$ 会有_____特征值并且是正定。但是 $S-s_{jj}I$ 会有一个____在对角线上。

请尊重版权与译者的劳动成果，侵权必究!

18 若 $S\boldsymbol{x}=\lambda\boldsymbol{x}$ 则 $\boldsymbol{x}^{\mathrm{T}}S\,\boldsymbol{x}=$ ____。当$\lambda>0$，为什么这个数字是正数？

19 逆反问题 18，证明若所有的$\lambda>0$ 则 $\boldsymbol{x}^{\mathrm{T}}S\,\boldsymbol{x}>0$。我们必须对每一个非零 $\boldsymbol{x}$ 做这件事情，不只是特征向量，所以把 $\boldsymbol{x}$ 写成特征向量的组合然后说明为什么所有的的交叉项 $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j=0$。则 $\boldsymbol{x}^{\mathrm{T}}S\,\boldsymbol{x}$ 是

$$(c_1\boldsymbol{x}_1+\ldots+c_n\boldsymbol{x}_n)^{\mathrm{T}}\,(c_1\lambda_1\boldsymbol{x}_1+\ldots+c_n\lambda_n\boldsymbol{x}_n)=c_1{}^2\lambda_1\boldsymbol{x}_1{}^{\mathrm{T}}\boldsymbol{x}_1+\ldots+c_n{}^2\lambda_n\boldsymbol{x}_n{}^{\mathrm{T}}\boldsymbol{x}_n>0$$

20 给出快速的理由说明为什么下列叙述都是正确的：

(a) 每一个正定矩阵是可逆。

(b) 唯一的正定投影矩阵是 $P=I$。

(c) 对角线单元都是正数的对矩阵是正定。

(d) 正数行列式的对称矩阵不一定是正定。

问题 21-24 使用特征值；问题 25-27 基于主元。

21 什么样的 s 与 t 使得 S 与 T 的全部$\lambda>0$ (所以是正定)?

$$S=\begin{bmatrix}s&-4&-4\\-4&s&-4\\-4&-4&s\end{bmatrix}\quad\text{与}\quad T=\begin{bmatrix}t&3&0\\3&t&4\\0&4&t\end{bmatrix}$$

22 从 $S=Q\Lambda Q^{\mathrm{T}}$，计算下列每个矩阵的正定对称平方根 $Q\sqrt{\Lambda}Q^{\mathrm{T}}$。验证这个平方根得到 $A^{\mathrm{T}}A=S$。

$$S=\begin{bmatrix}5&4\\4&5\end{bmatrix}\quad\text{与}\quad S=\begin{bmatrix}10&6\\6&10\end{bmatrix}$$

23 你可能看过椭圆方程式 $x^2/a^2+y^2/b^2=1$，当方程式写成$\lambda_1x^2+\lambda_2y^2=1$ 时，a 与 b 为何？椭圆 $9x^2+4y^2=1$ 的轴的半长是 $a=$ ____与 $b=$ ____。

24 画出倾斜椭圆 $x^2+xy+y^2=1$ 的图形，从对应矩阵 S 的特征值求出轴的半长。

25 D 有正数主元，$S=LDL^{\mathrm{T}}$ 的分解变成 $L\sqrt{D}\sqrt{D}L^{\mathrm{T}}$ (主元的平方根得到 $D=\sqrt{D}\sqrt{D}$)，则 $C=\sqrt{D}L^{\mathrm{T}}$ 得到**乔列斯基**分解 $A=C^{\mathrm{T}}C$，称为“对称化 LU”：

$$\text{从 }C=\begin{bmatrix}3&1\\0&2\end{bmatrix}\text{求 }S。\qquad\text{从 }S=\begin{bmatrix}4&8\\8&25\end{bmatrix}\text{求 }C=\mathrm{chol}(S)$$

26 Cholesky 分解 $S=C^{\mathrm{T}}C$，其中 $C=\sqrt{D}L^{\mathrm{T}}$，主元的平方根在 C 的对角线。求 C(上三角)：

$$S=\begin{bmatrix}9&0&0\\0&1&2\\0&2&8\end{bmatrix}\quad\text{与}\quad S=\begin{bmatrix}1&1&1\\1&2&2\\1&2&7\end{bmatrix}$$

27 对称分解 $S = LDL^{\mathrm{T}}$ 表示 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}LDL^{\mathrm{T}}\boldsymbol{x}$：

$$[x \quad y]\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = [x \quad y]\begin{bmatrix} 1 & 0 \\ b/a & 1 \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & \left(ac - b^2\right)/a \end{bmatrix}\begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

左侧是 $ax^2 + 2bxy + cy^2$，右侧是 $a(x + by/a)^2 +$ ___ y^2。第二主元完成平方项！使用 $a = 2, b = 4, c = 10$ 进行测试。

28 不要乘开 $S = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$，求出

(a) S 的行列式　　(b) S 的特征值

(c) S 的特征向量　(d) 一个理由：为什么 S 是对称正定？

29 $F_1(x, y) = x^4/4 + x^2y + y^2$ 与 $F_2(x, y) = x^3 + xy - x$，求出二次导数矩阵 S_1 与 S_2：

极小值测试　　$S = \begin{bmatrix} \partial^2 F/\partial x^2 & \partial^2 F/\partial x\partial y \\ \partial^2 F/\partial y\partial x & \partial^2 F/\partial y^2 \end{bmatrix}$ 是正定

S_1 是正定，所以是凹向上(=凸函数)，求 F_1 的极小点。求 F_2 的鞍点(只寻找一次导数是零的地方)。

30 $z = x^2 + y^2$ 的图形是一个开口朝上的碗，$z = x^2 - y^2$ 的图形是一个鞍形(saddle)，$z = -x^2 - y^2$ 的图形是一个开口朝下的碗。针对 $z = ax^2 + 2bxy + cy^2$ 的图形，如果在 $(x, y) = (0, 0)$ 有一个鞍点，求 a, b, c 的条件？

31 针对 $z = 4x^2 + 12xy + cy^2$ 的图形，什么样的 c 得到一个碗以及什么样的 c 得到一个鞍点？描述图形在边界线的 c 值。

函数 $F(x, y, z)$ 的极小值

对于极小点你会期待什么样的测试？首先是零斜率：

在极小点的一次导数是零　$\dfrac{\partial F}{\partial x} = \dfrac{\partial F}{\partial y} = \dfrac{\partial F}{\partial z} = 0$

然后是常用微积分测试 $d^2f/dx^2 > 0$ 的线性代数版本:

二次导数矩阵 S 是正定　　$S = \begin{bmatrix} F_{xx} & F_{xy} & F_{xz} \\ F_{yx} & F_{yy} & F_{yz} \\ F_{zx} & F_{zy} & F_{zz} \end{bmatrix}$

此处 $\boldsymbol{F_{xy}} = \dfrac{\partial}{\partial x}\left(\dfrac{\partial F}{\partial y}\right) = \dfrac{\partial}{\partial y}\left(\dfrac{\partial F}{\partial x}\right) = \boldsymbol{F_{yx}}$，是“混合”二次导数。

请尊重版权与译者的劳动成果，侵权必究！

挑战问题

32 一群非线性矩阵如果包含 A 与 B，则包含 AB 与 A^{-1}。“乘积与逆矩阵保持在群内”，下列哪些是群(如同 2.7.37)?

对于两个群发明一个“子群”(不是 I 本身 = 最小的群)。

(a) 正定对称矩阵 S。

(b) 正交矩阵 Q。

(c) 一个固定矩阵 A 的所有指数 e^{tA}。

(d) 有正数特征值的矩阵 P。

(e) 行列式是 1 的矩阵 D。

33 当 S 与 T 是对称正定，ST 甚至可能不是对称，但是它的特征值仍然是正数。从 $ST\boldsymbol{x} = \lambda\boldsymbol{x}$ 开始，然后计算与 $T\boldsymbol{x}$ 的点积，证明 $\lambda > 0$。

34 从已解范例 **6.5C** 写下 5×5 正弦矩阵 Q，包含当 $n = 5$ 与 $h = 1/6$ 的 S 的特征向量。执行 SQ 查看这 5 个 λ's。

λ's 的总和应该等于迹 10，他们的乘积应该是 $\det S = 6$。

35 假设 C 是正定(只要 $\boldsymbol{y} \neq \boldsymbol{0}$，$\boldsymbol{y}^{\mathrm{T}}C\boldsymbol{y} > 0$)以及 A 有无关的列(只要 $\boldsymbol{x} \neq \boldsymbol{0}$，$A\boldsymbol{x} \neq \boldsymbol{0}$)。对于 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}CA\boldsymbol{x}$ 应用能量测试，证明 $S = A^{\mathrm{T}}CA$ 是正定：工程的重要矩阵。

36 **重要!** 假设 S 是正定且有特征值 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$。

(a) 矩阵 $\lambda_1 I - S$ 的特征值为何？它是正半定吗？

(b) 如何推论：对每一个 $\boldsymbol{x}$ 都有 $\lambda_1\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \geq \boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$。

(c) 得出结论：$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}/\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 的极大值是____。

37 什么样的 a 与 c 使得矩阵是正定？什么样的 a 与 c 使得它是正半定(包含正定)?

$$S = \begin{bmatrix} a & a & a \\ a & a+c & a-c \\ a & a-c & a+c \end{bmatrix}$$

所有五个测试都可能能量 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ 等于 $a(x_1+x_2+x_3)^2 + c(x_2-x_3)^2$

特征值与特征向量表

反映在特征值与特征向量的矩阵性质有哪些？这个问题贯穿了第六章的根本，如果有一个表来组织这些关键事实会有很大的帮助。下列是特征值λ_i与特征向量$\boldsymbol{x}_i$的特殊性质。

对称：$S^{\mathrm{T}}=S=Q\Lambda Q^{\mathrm{T}}$	实数特征值	正交 $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j=0$
正交：$Q^{\mathrm{T}}=Q^{-1}$	所有的$\lvert\lambda\rvert=1$	正交 $\overline{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{x}_j=0$
反对称：$A^{\mathrm{T}}=-A$	虚数λ's	正交 $\overline{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{x}_j=0$
复数厄米：$\overline{S}^{\mathrm{T}}=S$	实数λ's	正交 $\overline{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{x}_j=0$
正定：$\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}>0$	所有的$\lambda>0$	由于 $S^{\mathrm{T}}=S$，正交
马尔可夫：$m_{ij}>0$ $\sum_{i=1}^{n}m_{ij}=1$	$\lambda_{\max}=1$	稳定状态 $\boldsymbol{x}>0$
相似：$A=BCB^{-1}$	$\lambda(A)=\lambda(C)$	B 乘 C 的特征向量
投影：$P=P^2=P^{\mathrm{T}}$	$\lambda=1;0$	列空间；零空间
平面旋转：余弦-正弦	$e^{i\theta}$ 与 $e^{-i\theta}$	$\boldsymbol{x}=(1,i)$与$(1,-i)$
镜射：$I-2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$	$\lambda=-1,1,\ldots,1$	$\boldsymbol{u}$；整个平面 $\boldsymbol{u}^{\perp}$
秩一：$\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$	$\lambda=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u};0,\ldots,0$	$\boldsymbol{u}$；整个平面 $\boldsymbol{u}^{\perp}$
逆矩阵：A^{-1}	$1/\lambda(A)$	维持 A 的特征向量
平移：$A+cI$	$\lambda(A)+c$	维持 A 的特征向量
稳定次方：$A^n\to 0$	全部$\lvert\lambda\rvert<1$	任意特征向量
稳定指数：$e^{At}\to 0$	全部 $\mathrm{Re}\lambda<0$	任意特征向量
循环排列：$P_{i,i+1}=1,P_{n1}=1$	$\lambda_k=e^{\frac{2\pi ik}{n}}=$(1 的根)	$\boldsymbol{x}_k=(1,\lambda_k,\ldots,\lambda_k^{n-1})$
循环行列式：$c_0I+c_1P+\ldots$	$\lambda_k=c_0+c_1e^{2\pi k/n}+\ldots$	$\boldsymbol{x}_k=(1,\lambda_k,\ldots,\lambda_k^{n-1})$
三对角：对角线是$-1,2,-1$	$\lambda_k=2-2\cos\frac{k\pi}{n+1}$	$\boldsymbol{x}_k=\left(\sin\frac{k\pi}{n+1},\cos\frac{k\pi}{n+1},\ldots\right)$
可对角化：$A=X\Lambda X^{-1}$	Λ的对角线	X 的列是无关
舒尔：$A=QTQ^{-1}$	三角 T 的对角线	若 $A^{\mathrm{T}}A=AA^{\mathrm{T}}$, Q 的列
若尔当：$A=BJB^{-1}$	J 的对角线	每个方块一个特征向量
SVD：$A=U\Sigma V^{\mathrm{T}}$	在Σ有 r 个奇异值	在 V,U 的 $A^{\mathrm{T}}A$, AA^{T} 的特征向量

请尊重版权与译者的劳动成果，侵权必究！

第七章
奇异值分解(SVD)

7.1 线性代数做图像处理

1 一个图像是一个灰度值的大型矩阵，每个对应一个像素与颜色。

2 当邻近像素相互关联(不是随机)时，可以压缩图像。

3 SVD 可以分解任意矩阵 A 成为秩一片段 $\boldsymbol{uv}^{\mathrm{T}}$ = (列) (行)。

4 列与行是对称矩阵 AA^{T} 与 $A^{\mathrm{T}}A$ 的特征向量。

A 的奇异值定理是 $A^{\mathrm{T}}A$ 与 AA^{T} 的特征值定理。

这是你在本章看到内容的快速预习。A 有两个奇异向量的集合($A^{\mathrm{T}}A$ 与 AA^{T} 的特征向量)，有一个正数奇异值的集合(因为 $A^{\mathrm{T}}A$ 与 AA^{T} 有相同的正数特征值)。A 通常是矩形，但是 $A^{\mathrm{T}}A$ 与 AA^{T} 是方形，而且是正半定。

奇异值分解(SVD)可以分解任意矩阵成为简单的片段(pieces)。

每个片段是一个列向量乘一个行向量。一个 $m\times n$ 的矩阵有 m 乘 n 个单元(当矩阵代表图像时是个很大的数字)，但是一个列与一个行只有 $m+n$ 个分量，远远小于 m 乘 n。这些(列)(行)片段是可以用超快速度处理的满大小(full size)矩阵——他们只需要 $m+n$ 个数字。

通常来说，SVD 的图像处理应用在它依靠的线性代数之前来到，我会从只包含一个或两个片段的简单图像开始。现在我把一个图像想象成一个大型的矩形矩阵，单元 a_{ij} 告诉我们图像的全部像素的灰度。把像素想象成一个小方形，从左下角开始横跨 i 步以及往上 j 步，它的灰度是一个数字(通常在 $0 \le a_{ij} < 256 = 2^8$ 的范围内)。一个全白的像素有 $a_{ij} = 255 = 11111111$，电脑把 255 写成 2 进位表示法时，这个数字有 8 个 1's。

你已经知道一个图像为什么有 m 乘 n 个像素，每个像素用 8 比特(0 或 1)表示它的灰度，变成一个 $m \times n$ 的矩阵，其中每个单元有 256 种的可能值。

简单的说，一个图像是一个大型矩阵，要完美复制它，我们需要 8$(m)(n)$比特的信息。高分辨率电视通常有 $m = 1080$ 与 $n = 1920$，每秒通常有 24 帧，而且你可能喜欢看彩色画面(3 个颜色度)，总共需要传输(3) (8) (48,470,400)比特/秒。这个太昂贵也不可行，传输器跟不上表演的速度。

当压缩做的很好时，你是看不出来与原始的差别。图像的边(edge)(灰度的突然改变)是最难压缩的部分。

如果每个 a_{ij} 都是无关的随机数字，压缩就不可能成功。我们完全依靠这个事实：邻近像素通常都有相似的灰度。当你跨越边的时候，它会产生一个突跳。实际世界的图像到处是边，所以动画更容易压缩。

对视频来说，数字 a_{ij} 在帧之间不太会改变，我们只传输小变化，这是 H.264 视频压缩标准(本书的网站)的不同编码方法，我们用线性代数压缩每个变化矩阵(也用非线性“量子化”变成电脑最有效率的整数) 。

我们每天看到的自然图像绝对可以而且准备好被压缩--但是没有变得容易。

低秩图像(范例)

最简单压缩的图像是全黑或全白或是常数的灰度 g，矩阵 A 的每个单元都是相同的 g：$a_{ij} = g$。当 $g = 1$ 且 $m = n = 6$，下列是一个图像处理的极端例子，这个例子是中央 SVD 教义派：

范例 1　**不要传送**　$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$　**传送这个** $A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [1\ 1\ 1\ 1\ 1\ 1]$

36 个数字变成 12 个数字，300×300 个像素，90000 个数字变成 600 个。如果我们事先定义好全 1 向量 $\boldsymbol{x}$，我们只需要传送一个数字，这个数字就是常数灰度 g 使得 $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ 得到这个矩阵。

当然第一个是极端的例子，但是它给出一个重点，如果有可以事先仔细定义的特殊向量，例如 $\boldsymbol{x}$ = ones，图像处理就可以超级快速。需要战斗是在由图像决定的事先选定基底(preselected bases)(傅里叶基底允许从 FFT 加速)与适应性基底(adaptive bases)之间做选择。SVD 产生来自图像本身的基底--这是适应性的，可能不便宜。

我不是说 SVD 永远或是通常给出实际上最有效率的演绎方法，下面例子的目的是指引而不是产生。

请尊重版权与译者的劳动成果，侵权必究!

范例 2
“王牌旗”
法国国旗 A
意大利国旗 A
德国国旗 A^{T}

不要传送 $A=\begin{bmatrix} a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \end{bmatrix}$ **传送** $A=\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}[a\ a\ c\ c\ e\ e]$

国旗有 3 个颜色但是秩还是 1，我们仍然有一个列乘一个行。假设这 36 个单元保持秩 1 的模式 $A=\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$，他们甚至可以全部不相同。但是当秩变成 $r=2$，我们需要 $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$，下列是一种选择：

范例 3
嵌入方形

$A=\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ 等于 $A=\begin{bmatrix} 1 \\ 1 \end{bmatrix}[1\ \ 1]-\begin{bmatrix} 1 \\ 0 \end{bmatrix}[0\ \ 1]$

A 的 1's 与 0 可以是 1's 的方块与 0 的方块，我们仍然有秩 2，我们仍然只需要两个项 $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$。一个 6×6 图像可以压缩成 24 个数字，一个 $N\times N$ 图像(N^2 个数字)可以压缩成来自 4 个向量 $\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2$ 的 $4N$ 个数字。

我有做出 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 的最佳选择了吗？这个不是来自 SVD 的选择！我注意到 $\boldsymbol{u}_1=(1,1)$与 $\boldsymbol{u}_2=(1,0)$没有正交，且 $\boldsymbol{v}_1=(1,1)$与 $\boldsymbol{v}_2=(0,1)$也没有正交。原理告诉我们正交性会产生较小的第二片段 $c_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$。(**SVD 按重要顺序选择秩 1 片段**。)

若 A 的秩远大于 2，如同我们预期实际图像一般，A 会是很多秩 1 片段的总和。我们希望小东西是真正的小——小到可以被丢弃而无损视觉品质。图像压缩变成有损，但是好的图像压缩实际上是无法被人类的视觉系统侦测。

问题变成：**来自 SVD 的正交选择是什么**？

SVD 的特征向量

我想要介绍特征向量的用途，但是大部分图像的特征向量不是正交。此外，特征向量 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 只给出一个向量集合，而我们需要两个集合($\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s)。这两个困难的答案正是 SVD 的概念：

使用 AA^{T} 的特征向量 $\boldsymbol{u}$ 与 $A^{\mathrm{T}}A$ 的特征向量 $\boldsymbol{v}$。

由于 AA^{T} 与 $A^{\mathrm{T}}A$ 自动是对称(但是通常不相等！)，$\boldsymbol{u}$'s 会是一个正交集合，而且 $\boldsymbol{v}$'s 会是**另一个正交集合**。我们可以也会让他们变成单位向量：$\|\boldsymbol{u}_i\|=1$ 且$\|\boldsymbol{v}_i\|=1$，则我们的秩 2 矩阵会是 $A=\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$。这些数字$\sigma_1$ 与σ_2 的大小会决定他们在压缩时能不能被忽略。我们保存大的σ's，我们丢弃小的σ's。

SVD 的 $\boldsymbol{u}$'s 称为**左奇异向量**(AA^{T} 的单位特征向量)，$\boldsymbol{v}$'s 称为**右奇异向量**($A^{\mathrm{T}}A$ 的单位特征向量)，σ's 是奇异值，就是 AA^{T} 与 $A^{\mathrm{T}}A$ 的**相同特征值**的平方根：

来自 SVD 的选择 $\quad \boldsymbol{AA^{\mathrm{T}}u_i=\sigma_i^2u_i} \quad \boldsymbol{A^{\mathrm{T}}Av_i=\sigma_i^2v_i} \quad \boldsymbol{Av_i=\sigma_iu_i}$ (1)

范例 3 中(嵌入方形)，下列是对称矩阵 AA^{T} 与 $A^{\mathrm{T}}A$：

$$AA^{\mathrm{T}}=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}1&1\\0&1\end{bmatrix}=\begin{bmatrix}1&1\\1&2\end{bmatrix} \qquad A^{\mathrm{T}}A=\begin{bmatrix}1&1\\0&1\end{bmatrix}\begin{bmatrix}1&0\\1&1\end{bmatrix}=\begin{bmatrix}2&1\\1&1\end{bmatrix}$$

他们的行列式是 1，所以$\lambda_1\lambda_2=1$，他们的迹(对角线的总和)是 3：

$$\det\begin{bmatrix}1-\lambda&1\\1&2-\lambda\end{bmatrix}=\lambda^2-3\lambda+1=0 \text{ 得到 } \lambda_1=\frac{3+\sqrt{5}}{2}\text{与}\lambda_2=\frac{3-\sqrt{5}}{2}$$

λ_1 与λ_2 的平方根是$\sigma_1=\dfrac{\sqrt{5}+1}{2}$ 与 $\sigma_2=\dfrac{\sqrt{5}-1}{2}$ 且 $\sigma_1\sigma_2=1$

最接近 A 的秩 1 矩阵是$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$，误差只有$\sigma_2\approx 0.6=$ 最佳可能。

AA^{T} 与 $A^{\mathrm{T}}A$ 的正交单位特征向量是

$$\boldsymbol{u}_1=\begin{bmatrix}1\\\sigma_1\end{bmatrix}\quad \boldsymbol{u}_2=\begin{bmatrix}\sigma_1\\-1\end{bmatrix}\quad \boldsymbol{v}_1=\begin{bmatrix}\sigma_1\\1\end{bmatrix}\quad \boldsymbol{v}_2=\begin{bmatrix}1\\-\sigma_1\end{bmatrix} \text{ 全部除以 } \sqrt{1+\sigma_1^{\,2}} \qquad (2)$$

每位读者都知道在实际生活中这些计算是由计算机完成！(当然不是依赖不可靠的教授，我都是用 MATLAB 的 svd(A)来更正自己。) 借由$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$，我们可以检验正确的回复到矩阵 A：

$$A=\begin{bmatrix}\boldsymbol{u}_1&\boldsymbol{u}_2\end{bmatrix}\begin{bmatrix}\sigma_1&\\&\sigma_2\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^{\mathrm{T}}\\\boldsymbol{v}_2^{\mathrm{T}}\end{bmatrix} \text{ 或更简单的 } A\begin{bmatrix}\boldsymbol{v}_1&\boldsymbol{v}_2\end{bmatrix}=\begin{bmatrix}\sigma_1\boldsymbol{u}_1&\sigma_2\boldsymbol{u}_2\end{bmatrix} \qquad (3)$$

重要　关键点不是图像趋向有低的秩，不是的，图像通常有满秩，但是他们确实有低效率的(low effective)秩，这就表示：很多奇异值很小，可以被设定成零。我们传输的是低秩的近似。

范例 4　假设旗子有两个不同颜色的三角形，左下的三角形有 1's，右上的三角形有 0's，主对角线都是 1's。下列是 $n=4$ 的图像矩阵，它有满秩 $r=4$ 所以可逆：

三角形旗子矩阵 $\quad A=\begin{bmatrix}\mathbf{1}&0&0&0\\\mathbf{1}&\mathbf{1}&0&0\\\mathbf{1}&\mathbf{1}&\mathbf{1}&0\\\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{1}\end{bmatrix}$ 与 $A^{-1}=\begin{bmatrix}1&0&0&0\\-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix}$

请尊重版权与译者的劳动成果，侵权必究！

矩阵有满秩，所以 A 有一整组 n 个奇异值σ(都是正数)，SVD 会产生 n 个秩一的片段$\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$，完美重建需要 n 个片段。

压缩时我们丢弃小的σ's 不会严重损失图像的品质，我们希望了解并且画出 $n = 4$ 以及更大 n 的σ's 的图形。注意范例 3 是三角形范例 4 的特殊例子 $n = 2$。

用手算，我们从 AA^{T} 开始(电脑会有不同的处理)：

$$AA^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \quad \text{且} \quad (AA^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}} A^{-1} = \begin{bmatrix} \mathbf{2} & -1 & 0 & 0 \\ -1 & \mathbf{2} & -1 & 0 \\ 0 & -1 & \mathbf{2} & -1 \\ 0 & 0 & -1 & \mathbf{1} \end{bmatrix} \tag{4}$$

因为这个–1, 2, –1 逆矩阵的特征值都有 $2 - 2\cos\theta$ 的形式，所以它被引进，我们就可以知道 AA^{T} 的λ's 与 A 的σ's：

$$\lambda = \frac{1}{2 - 2\cos\theta} = \frac{1}{4\sin^2(\theta/2)} \quad \text{得到} \quad \sigma = \sqrt{\lambda} = \frac{1}{2\sin(\theta/2)} \tag{5}$$

n 个不同角度θ是等距离，使得这个例子如此特殊：

$$\theta = \frac{\pi}{2n+1}, \frac{3\pi}{2n+1}, \ldots, \frac{(2n-1)\pi}{2n+1} \quad \left(n = 4 \text{ 包含 } \theta = \frac{3\pi}{9} \text{ 且 } 2\sin\frac{\theta}{2} = 1 \right)$$

这个特殊例子得到 AA^{T} 在 $n = 4$ 的特征值$\lambda = 1$，所以$\sigma = \sqrt{\lambda} = 1$ 是 A 的一个奇异值。你可以检查向量 $\boldsymbol{u} = (1, 1, 0, -1)$时，$AA^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{u}$ (真正的特例)。

重点是画出 A 的 n 个奇异值的图形，这些数字逐渐减少(不像 A 的特征值全部是 1)，但是不是很陡的减少，所以 SVD 只是适度的压缩这个三角形旗子。希尔伯特伟大的压缩。

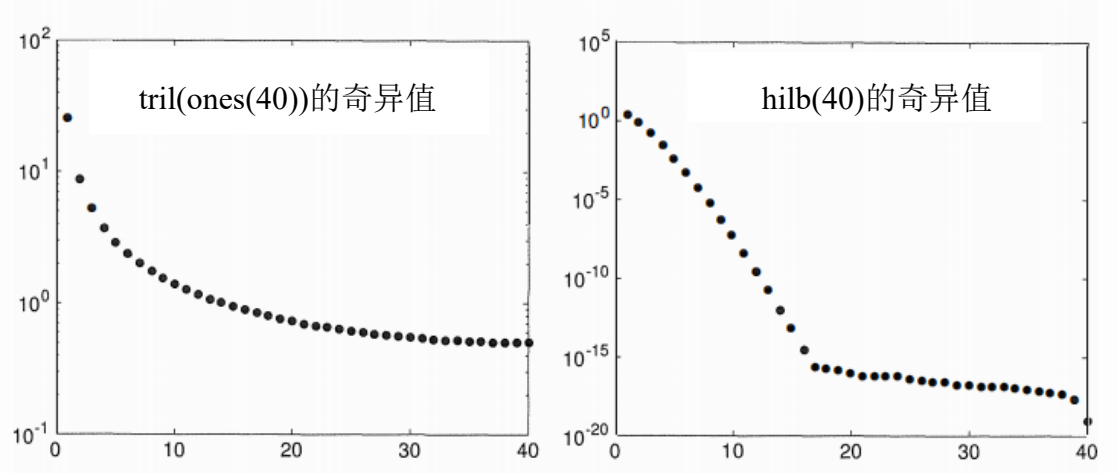


图 7.1：范例 3-4 中三角 1 矩阵的奇异值(不可压缩)与段落 8.3 的邪恶希尔伯特矩阵 $H(i, j) = (i + j - 1)^{-1}$：压缩后使用。

你忠实的作者一直从事国旗秩的研究，有很多是基于水平或是垂直条纹，这些秩 1--所有的行或列都是 1 向量(1, 1, .., 1)的倍数。亚美尼亚，奥地利，比利时，保加利亚，乍得，哥伦比亚，爱尔兰，马达加斯加，马里，荷兰，尼日利亚，罗马尼亚，俄罗斯(还有更多)都有 3 个条纹。印尼与波兰有两条！利比亚在卡扎菲时代 1977 到 2011 是最极端的案例(正面国旗都是绿色)。

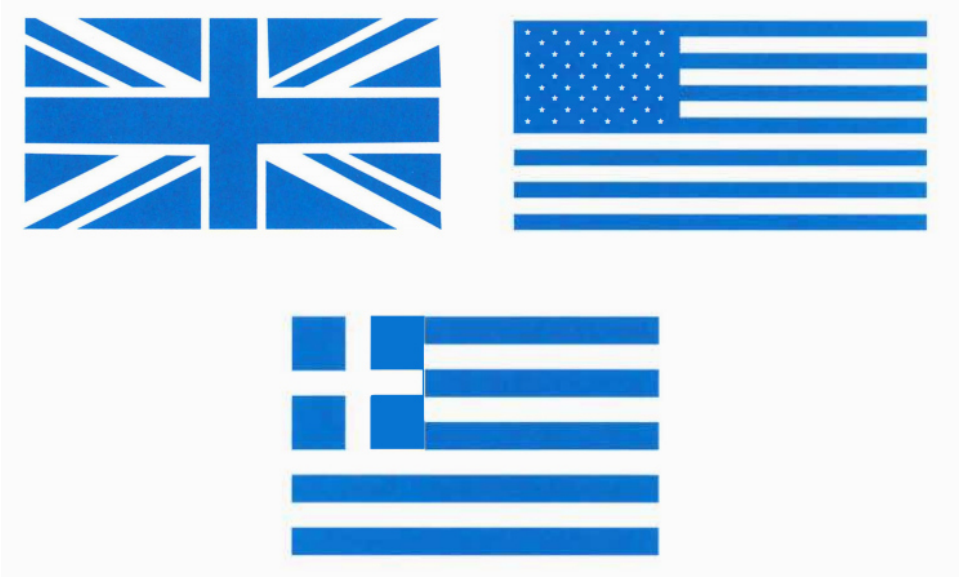
另一种极端是很多国旗含有对角线，好像英国国旗的长对角线，或是具有来自星星的边的短对角线--例如美国国旗。本书的三角形 1 的例子显示了国旗矩阵如何拥有很大的秩，秩增加至无限大，像素变得非常小。

其他国旗有圆形或月牙形或是不同的曲线形状，他们的秩很大也会一直增加至无限大，但是他们仍然可以压缩！压缩的图像不完美，但是我们的眼睛看不出差别(来自 SVD 足够的项$\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$)，这些例子真正引出图像压缩的主要目的：

很大的秩简化仍然可以维护视觉品质。

为了好玩，我们回头看看有限秩的国旗，他们可以有条纹也可以有交叉线--假设横跨边是水平或是垂直。某些国旗有薄框环绕交叉线，这些艺术触感会增加秩。现在我的冠军是下列所示的希腊国旗，有交叉线也与条纹，它的秩是 3×(我的计数，三个不同的列数)。我没有看到有限秩的美国国旗！

读者可以去 GOOGLE“national flags”看看不同的设计与颜色，我会很高兴知道任何 $n > 3$ 的有限秩的例子。各种类型的好例子可以到本书的网站 math.mit/linearalgebra (有全颜色的国旗)。



请尊重版权与译者的劳动成果，侵权必究！

问题集 7.1

1 求出单元 i, j 是 i 乘 j 与 $i+j$ 的矩阵的秩。把 A 与 B 写成 r 个片段 $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 的总和，不需要 $\boldsymbol{u}_1{}^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} = \boldsymbol{0}$。

$$A=\begin{bmatrix}1&2&3&4\\2&4&6&8\\3&6&9&12\\4&8&12&16\end{bmatrix} \qquad B=\begin{bmatrix}2&3&4&5\\3&4&5&6\\4&5&6&7\\5&6&7&8\end{bmatrix}$$

2 我们通常把单位矩阵 I 想成是最简单的，但是为什么 I 完全不可压缩？画出有一条交叉线而且是秩 5 的国旗。

3 这些国旗的秩 2，把 A 与 B 写成 $\boldsymbol{u}_1\boldsymbol{v}_1{}^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2{}^{\mathrm{T}}$。

$$A_{\text{瑞典}} = A_{\text{芬兰}} = \begin{bmatrix}1&2&1&1\\2&2&2&2\\1&2&1&1\end{bmatrix} \qquad B_{\text{贝宁}} = \begin{bmatrix}1&2&2\\1&3&3\end{bmatrix}$$

4 求出问题 3 中 BB^{T} 与 $B^{\mathrm{T}}B$ 的迹与行列式，B 的奇异值接近 $\sigma_1{}^2 = 28 - 1/14$ 与 $\sigma_2{}^2 = 1/14$，B 可以压缩吗？

5 使用 $[U, S, V] = \mathrm{svd}(A)$ 求出 $A_{\text{瑞典}}$ 的两个正交片段 $\sigma\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$。

6. 求出 2×2 矩阵 A 的特征值与奇异值。

$$A=\begin{bmatrix}2&1\\4&2\end{bmatrix} \quad \text{且} \quad A^{\mathrm{T}}A=\begin{bmatrix}20&10\\10&5\end{bmatrix} \qquad \text{与} \quad AA^{\mathrm{T}}=\begin{bmatrix}5&10\\10&20\end{bmatrix}$$

A 的特征向量(1, 2)与(1, –2)没有正交，你如何知道 $A^{\mathrm{T}}A$ 的特征向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 是正交？注意 $A^{\mathrm{T}}A$ 与 AA^{T} 有相同的的特征值(25 与 0)。

7 方程式(3)的第二形式 $\boldsymbol{AV} = \boldsymbol{U\Sigma}$ 是如何来自第一形式 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$？这是 SVD 最著名的形式。

8 $AV = U\Sigma$ 的两个列是 $A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$ 与 $A\boldsymbol{v}_2 = \sigma_2\boldsymbol{u}_2$，所以我们希望

$$A\boldsymbol{v}_1=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}\sigma_1\\1\end{bmatrix}=\sigma_1\begin{bmatrix}1\\\sigma_1\end{bmatrix} \quad \text{与} \quad A\boldsymbol{v}_2=\begin{bmatrix}1&0\\1&1\end{bmatrix}\begin{bmatrix}1\\-\sigma_1\end{bmatrix}=\sigma_2\begin{bmatrix}\sigma_1\\-1\end{bmatrix}$$

第一需要 $\sigma_1 + 1 = \sigma_1{}^2$ 且第二需要 $1 - \sigma_1 = -\sigma_2$，这些是正确吗？

9 MATLAB 的指令 $A = \mathrm{rand}(20, 40)$ 与 $B = \mathrm{randn}(20, 40)$ 产生 20×40 的随机矩阵，A 的单元在 0 与 1 之间是均匀概率，B 的单元有一个正态的“钟形”概率分布。使用 svd 指令，求出并画出他们的奇异值 σ_1 到 σ_{20} 的图形。为什么他们有 20 个 σ's。

7.2 SVD 的基底与矩阵

1 SVD 产生 4 个基础子空间的 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 的**正交单位基底**。

2 使用这些基底，A 变成对角矩阵Σ且 $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ ： $\sigma_i =$ **奇异值**。。

3 双基底对角化 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$ 通常拥有比 $A = X\Lambda X^{-1}$ 更多的信息。

4 $U\Sigma V^{\mathrm{T}}$ 把 A 分成秩 1 矩阵$\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \ldots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$， $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ 是最大！

奇异值分解是线性代数的一个亮点，A 的任意 $m\times n$ 矩阵，方形或矩形，它的秩是 r。我们会对角化 A，但不是用 $X^{-1}AX$，X 里面的特征向量有三个大问题：他们通常不是正交，不一定有足够的特征向量，而且 $A\boldsymbol{x} = \lambda \boldsymbol{x}$ 要求 A 是方形。A 的奇异向量以完美的方式解决了全部的问题。

让我描述我们想要从 SVD 得到什么：**4 个子空间的直角基底**，然后我会依据重要顺序写下求出基底向量的步骤。

我们付出的代价是两组奇异向量 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 的集合，$\boldsymbol{u}$'s 在$\mathbf{R}^m$而且 $\boldsymbol{v}$'s 在$\mathbf{R}^n$，他们会是 $m\times m$ 矩阵 U 以及 $n\times n$ 矩阵 V 的列。我首先会利用这些基底来描述 SVD，然后我也会用正交矩阵 U 与 V 来描述 SVD。

(使用向量) $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 产生 4 个基础子空间的基底：

$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$	是**列空间**的一组正交单位基底
$\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m$	是**左零空间** $\boldsymbol{N}(A^{\mathrm{T}})$的一组正交单位基底
$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$	是**行空间**的一组正交单位基底
$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$	是**零空间** $\boldsymbol{N}(A)$的的一组正交单位基底

不只是正交，这些基底向量可以对角化 A：

“对角化 $\boldsymbol{A}$” $A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1$ $A\boldsymbol{v}_2 = \sigma_2 \boldsymbol{u}_2$ ... $A\boldsymbol{v}_r = \sigma_r \boldsymbol{u}_r$ (1)

这些**奇异值σ_1 到σ_r**是正数：σ_i 是 $A\boldsymbol{v}_i$ 的长度。这些σ's 进入一个对角矩阵，这个矩阵其他单元都是零，就是矩阵Σ。

(使用矩阵) 由于 $\boldsymbol{u}$'s 是正交单位，有 r 个列的矩阵 U_r 会有 $U_r^{\mathrm{T}}U_r = I$。由于 $\boldsymbol{v}$'s 是正交单位，矩阵 V_r 会有 $V_r^{\mathrm{T}}V_r = I$。则方程式 $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ 会一列一列的告诉我们 $\boldsymbol{AV_r = U_r \Sigma_r}$ ：

$$\begin{matrix}(m\times n)(n\times r)\\ \boldsymbol{AV_r = U_r \Sigma_r}\\ (m\times r)(r\times r)\end{matrix} \qquad A\begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r\end{bmatrix}\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix} \tag{2}$$

请尊重版权与译者的劳动成果，侵权必究！

这是 SVD 的心脏，但是还有更多。这些 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 负责 A 的行空间与列空间，我们还有 $n-r$ 个 $\boldsymbol{v}$'s 与 $m-r$ 个 $\boldsymbol{u}$'s，分别来自零空间 $\boldsymbol{N}(A)$与左零空间 $\boldsymbol{N}(A^{\mathrm{T}})$。他们自动与前面 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 正交(因为整个零空间是正交)。我们现在引入 V 与 U 的所有 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s，所以矩阵变成方形，我们仍然有 $\boldsymbol{AV=U\Sigma}$。

$$\begin{matrix}(m\times n)(n\times n)\\ \boldsymbol{AV=U\Sigma}\\ (m\times m)(m\times n)\end{matrix}\qquad A\begin{bmatrix}\boldsymbol{v}_1\cdots\boldsymbol{v}_r\cdots\boldsymbol{v}_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{u}_1\cdots\boldsymbol{u}_r\cdots\boldsymbol{u}_n\end{bmatrix}\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & \end{bmatrix}\tag{3}$$

新的Σ是 $m\times n$，它就是方程式(2)的 $r\times r$ 矩阵加上 $m-r$ 个额外零行以及 $n-r$ 个零列，真正的改变在于 U 与 V 的形状。这些是方形矩阵且 $V^{-1}=V^{\mathrm{T}}$，所以 $AV=U\Sigma$ 变成 $\boldsymbol{A=U\Sigma V^{\mathrm{T}}}$，这就是奇异值分解。我可以把 V^{T} 的行乘来自 $U\Sigma$的列 $\boldsymbol{u}_i\sigma_i$：

SVD　　$$\boldsymbol{A=U\Sigma V^{\mathrm{T}}=\sigma_1 u_1 v_1^{\mathrm{T}}+\ldots+\sigma_r u_r v_r^{\mathrm{T}}}\tag{4}$$

方程式(2)是一个简化 SVD，具有行空间与列空间的基底。方程式(3)是包含零空间的完整 SVD，他们两者都把 A 分解成相同的 r 个秩 1 的$\sigma_i u_i v_i^{\mathrm{T}}$矩阵。列乘行是第四种矩阵相乘的方法。

我们看到每个σ_i^2 是 $A^{\mathrm{T}}A$ 的特征值，也是 AA^{T} 的。当我们把奇异值以降序的方式排列，$\sigma_1\geq\sigma_2\geq\ldots\geq\sigma_r>0$，方程式(4)的分割得到 A 的 r 个**依据重要顺序**排序的秩 1 片段，这个很重要。

范例 1　什么时候 $A=U\Sigma V^{\mathrm{T}}$(奇异值)与 $X\Lambda X^{-1}$(特征值)相等？

解　A 需要正交特征向量才能允许 $X=U=V$。如果$\Lambda=\Sigma$，A 也需要特征值$\lambda\geq 0$，所以 A 必须是一个**正半定(或是正定)对称矩阵**。只有在这个时候 $A=X\Lambda X^{-1}$，也就是 $Q\Lambda Q^{\mathrm{T}}$，才能与 $A=U\Sigma V^{\mathrm{T}}$ 一致。

范例 2　若 $A=\boldsymbol{xy}^{\mathrm{T}}$(秩 1)有单位向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$，A 的 SVD 为何？

解　方程式(2)的简化 SVD 就是 $\boldsymbol{xy}^{\mathrm{T}}$，其中秩 $r=1$，它有 $\boldsymbol{u}_1=\boldsymbol{x}$ 与 $\boldsymbol{v}_1=\boldsymbol{y}$ 与$\sigma_1=1$。对于完整 SVD 来说，让 $\boldsymbol{u}_1=\boldsymbol{x}$ 作为 $\boldsymbol{u}$'s 的一组正交单位基底，让 $\boldsymbol{v}_1=\boldsymbol{y}$ 作为 $\boldsymbol{v}$'s 的一组正交单位基底，不存在新的σ's，只有$\sigma_1=1$。

SVD 的证明

我们需要展示这些惊人的 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 是如何创建的，$\boldsymbol{v}$'s 会是 $A^{\mathrm{T}}A$ 的正交单位特征向量，这个必须成立，因为我以下列方程式为目标

$$\boldsymbol{A^{\mathrm{T}}A}=(U\Sigma V^{\mathrm{T}})^{\mathrm{T}}(U\Sigma V^{\mathrm{T}})=V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}}=\boldsymbol{V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}}\tag{5}$$

在右侧你会看到对称正定(正半定)矩阵 $A^{\mathrm{T}}A$ 的特征向量矩阵 V，以及($\Sigma^{\mathrm{T}}\Sigma$)必须是($A^{\mathrm{T}}A$)的特征值矩阵：每个$\sigma^2$ 是$\lambda(A^{\mathrm{T}}A)$。

现在 $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ 告诉我们单位向量 $\boldsymbol{u}_1$ 到 $\boldsymbol{u}_r$，这是关键方程式(1)。重要的部分——SVD 能够成功的理由——就是这些 $\boldsymbol{u}_1$ 到 $\boldsymbol{u}_r$ 自动是彼此正交(因为 $\boldsymbol{v}$'s 正交)：

$$\textbf{关键步骤 } i \neq j \qquad \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{u}_j = \left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right)^{\mathrm{T}} \left(\frac{A\boldsymbol{v}_j}{\sigma_j}\right) = \frac{\boldsymbol{v}_i^{\mathrm{T}} A^{\mathrm{T}} A \boldsymbol{v}_j}{\sigma_i \sigma_j} = \frac{\sigma_j^{\,2}}{\sigma_i \sigma_j} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{v}_j = 0 \qquad (6)$$

$\boldsymbol{v}$'s 是 $A^{\mathrm{T}}A$(对称)的特征向量，他们是正交而且现在 $\boldsymbol{u}$'s 也是正交，实际上这些 $\boldsymbol{u}$'s 会是 AA^{T} 的特征向量。

最终我们完成了 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 至 n 个 $\boldsymbol{v}$'s 与 m 个 $\boldsymbol{u}$'s 具有任何 $\boldsymbol{N}(A)$与 $\boldsymbol{N}(A^{\mathrm{T}})$的正交单位基底，我们求出了 $A = U\Sigma V^{\mathrm{T}}$ 的 V 与Σ与 U。

SVD 的一个例子

下列是一个展示计算 $A = U\Sigma V^{\mathrm{T}}$ 的三个矩阵的例子。

范例 3 $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$，秩是 2，找出 U，Σ，V。

秩是 2，这个 A 有正数奇异值σ_1 与σ_2，我们会看到σ_1 比$\lambda_{\max} = 5$ 大，σ_2 比$\lambda_{\min} = 3$ 小。从 $A^{\mathrm{T}}A$ 与 $A^{\mathrm{T}}A^{\mathrm{T}}$ 开始：

$$A^{\mathrm{T}}A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \qquad A^{\mathrm{T}}A^{\mathrm{T}} = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

这些矩阵有相同的迹(50)与相同的特征值$\sigma_1^{\,2} = 45$ 与$\sigma_2^{\,2} = 5$，平方根是$\sigma_1 = \sqrt{45}$ 与$\sigma_2 = \sqrt{5}$ 则$\sigma_1\sigma_2 = 15$ 而且这是 A 的行列式。

一个关键步骤是求出 $A^{\mathrm{T}}A$ 的特征向量(对应特征值 45 与 5)：

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

则 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 是重设比例成为长度 1 的正交向量，除以$\sqrt{2}$。

$$\textbf{右奇异向量 } \boldsymbol{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \textbf{左奇异向量 } \boldsymbol{u}_i = \frac{A\boldsymbol{v}_i}{\sigma_i}$$

现在计算 $A\boldsymbol{v}_1$ 与 $A\boldsymbol{v}_2$，这些会等于$\sigma_1\boldsymbol{u}_1 = \sqrt{45}\,\boldsymbol{u}_1$ 与$\sigma_2\boldsymbol{u}_2 = \sqrt{5}\,\boldsymbol{u}_2$：

$$A\boldsymbol{v}_1 = \frac{3}{\sqrt{2}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sqrt{45}\, \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1 \boldsymbol{u}_1$$

$$A\boldsymbol{v}_2 = \frac{3}{\sqrt{2}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sqrt{5}\, \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \sigma_2 \boldsymbol{u}_2$$

除以$\sqrt{10}$使得 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$ 正交单位，则如同预期的$\sigma_1 = \sqrt{45}$ 与$\sigma_2 = \sqrt{5}$，A 的奇异值分解是 U 乘Σ乘 V^{T}。

请尊重版权与译者的劳动成果，侵权必究!

$$U=\frac{1}{\sqrt{10}}\begin{bmatrix}1 & -3\\ 3 & 1\end{bmatrix} \qquad \Sigma=\begin{bmatrix}\sqrt{45} & \\ & \sqrt{5}\end{bmatrix} \qquad V=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix} \tag{7}$$

U 与 V 包含列空间与行空间的正交单位基底(两个空间都是 $\mathbf{R}^2$)，实际的成果是这两组基底可以对角化 A：$AV=U\Sigma$。矩阵 A 分割成两个秩 1 矩阵的组合，列乘行：

$$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}=\frac{\sqrt{45}}{\sqrt{20}}\begin{bmatrix}1 & 1\\ 3 & 3\end{bmatrix}+\frac{\sqrt{5}}{\sqrt{20}}\begin{bmatrix}3 & -3\\ -1 & 1\end{bmatrix}=\begin{bmatrix}3 & 0\\ 4 & 5\end{bmatrix}=A$$

一个极端矩阵

下列是一个较大系统，当 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 恰好是单位矩阵的列，所以计算容易，但是要注意列的顺序。矩阵 A 是极度不平衡(严格三角形)，所有的特征值都是零。AA^{T} 与 $A^{\mathrm{T}}A$ 不接近，矩阵 U 与 V 会是正确修正那些问题的排列。

$$A=\begin{bmatrix}0 & \mathbf{1} & 0 & 0\\ 0 & 0 & \mathbf{2} & 0\\ 0 & 0 & 0 & \mathbf{3}\\ 0 & 0 & 0 & 0\end{bmatrix}$$

固有值 $\lambda=0,\ 0,\ 0,\ 0$ 全部是 0
只有一个固有向量 $(1,\ 0,\ 0,\ 0)$
奇异值 $\sigma=\mathbf{3},\ \mathbf{2},\ \mathbf{1}$
奇异向量是 I 的列

$A^{\mathrm{T}}A$ 与 AA^{T} 是对角(有简单的特征向量，但是不同顺序)：

$$A^{\mathrm{T}}A=\begin{bmatrix}\mathbf{0} & 0 & 0 & 0\\ 0 & \mathbf{1} & 0 & 0\\ 0 & 0 & \mathbf{4} & 0\\ 0 & 0 & 0 & \mathbf{9}\end{bmatrix} \qquad AA^{\mathrm{T}}=\begin{bmatrix}\mathbf{1} & 0 & 0 & 0\\ 0 & \mathbf{4} & 0 & 0\\ 0 & 0 & \mathbf{9} & 0\\ 0 & 0 & 0 & \mathbf{0}\end{bmatrix}$$

他们的特征向量($\boldsymbol{u}$'s 对应 AA^{T} 与 $\boldsymbol{v}$'s 对应 $A^{\mathrm{T}}A$)依据特征值的降序排列 $\sigma_1^2>\sigma_2^2>\sigma_3^2$，这些特征值是 $\sigma^2=9,4,1$。

$$U=\begin{bmatrix}0 & 0 & \mathbf{1} & 0\\ 0 & \mathbf{1} & 0 & 0\\ \mathbf{1} & 0 & 0 & 0\\ 0 & 0 & 0 & \mathbf{1}\end{bmatrix} \qquad \Sigma=\begin{bmatrix}\mathbf{3} & & & \\ & \mathbf{2} & & \\ & & \mathbf{1} & \\ & & & \mathbf{0}\end{bmatrix} \qquad V=\begin{bmatrix}0 & 0 & 0 & \mathbf{1}\\ 0 & 0 & \mathbf{1} & 0\\ 0 & \mathbf{1} & 0 & 0\\ \mathbf{1} & 0 & 0 & 0\end{bmatrix}$$

U 与 V 的第一列 $\boldsymbol{u}_1$ 与 $\boldsymbol{v}_1$ 有 1's 在位置 3 与 4，则 $\boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}}$ 选出原始矩阵 A 的最大数字 $A_{34}=3$。SVD 的三个秩 1 矩阵(针对这个极端矩阵)确实来自 A 中的 3, 2, 1。

$$\boldsymbol{A=U\Sigma V^{\mathrm{T}}=3u_1v_1^{\mathrm{T}}+2u_2v_2^{\mathrm{T}}+1u_3v_3^{\mathrm{T}}}$$

注解　假设我移除 A 的最后一行(都是零)，则 A 是一个 3×4 矩阵且 AA^{T} 是 3×3—它的第四个行与列会消失，我们在 $A^{\mathrm{T}}A$ 与 AA^{T} 是对角中仍然有特征值 $\lambda=1,4,9$，在 Σ 中得到相同的奇异值 $\sigma=3,2,1$。

移除 A 的零行(现在 3×4)就是移除了Σ的最后一行也是 U 的最后一行与最后一列，则(3×4) = $U\Sigma V^{\mathrm{T}}$ = (3×3) (3×4) (4×4)，SVD 完全可以适应矩形矩阵。

好消息，因为资料矩阵 A 的行与列通常有完全不同的意义(好像电子计算表)，如果我们有全部课程的成绩，就会有一列对应每个学生，一行对应每个课程：单元 a_{ij} 就是成绩，则$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ 会有 $\boldsymbol{u}_1$ = 课程的组合与 $\boldsymbol{v}_1$ = 学生的组合，且σ_1 会是这些组合的成绩：最高的成绩。

矩阵 A 可以计数在期刊的关键字的频率：一个不同文章对应 A 的每个列，一个不同的字对应每个行，全部文章使用矩阵做索引，最重要的信息放在$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$，则$\sigma_1$ 是对应在伪文章 $\boldsymbol{v}_1$ 的一个伪字(字的组合 $\boldsymbol{u}_1$)的频率。

段落 7.3 会应用 SVD 至金融与遗传与搜寻引擎。

奇异值稳定对比特征值不稳定

4×4 范例 A 提供了一个特征值不稳定的例子(一个极端案例)，假设 4, 1 单元从零轻微的变成 1/60000，秩还是 4。

$$A=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \dfrac{1}{60000} & 0 & 0 & 0 \end{bmatrix}$$

只有1/60000的变化
产生固有值很大的跳动
$\lambda = 0,\ 0,\ 0,\ 0$ 变成 $\lambda = \dfrac{1}{10},\ \dfrac{i}{10},\ \dfrac{-1}{10},\ \dfrac{-i}{10}$

4个特征值会从零变成落在环绕零的圆上，当新单元只是1/60000，圆半径是1/10。当 AA^{T} 远离 $A^{\mathrm{T}}A$，这个情形显示了特征值严重的不稳定。在另一个极端，若 $A^{\mathrm{T}}A = AA^{\mathrm{T}}$(一个“正态矩阵”)，$A$ 的特征向量是正交而且 A 的特征值全部稳定。

相对来说，**任意矩阵的奇异值都是稳定**，他们的改变不会比矩阵的改变多。在这个例子中，新的奇异值是 3, 2, 1 与 1/60000，矩阵 U 与 V 维持不变，A 的新的第 4 片段是$\sigma_4\boldsymbol{u}_4\boldsymbol{v}_4^{\mathrm{T}}$，有 15 个零与小单元$\sigma_4$ = 1/60000。

A 的奇异向量与 $S = A^{\mathrm{T}}A$ 的特征向量

方程式(5-6)一口气“证明”了 SVD，奇异向量 $\boldsymbol{v}_i$ 是 $S = A^{\mathrm{T}}A$ 的特征向量 $\boldsymbol{q}_i$，S 的特征值λ_i 与对应 A 的σ_i^2 相同，S 的秩等于 A 的秩。特征向量与奇异向量的扩展是完美对应。

对称 S	$S = Q\Lambda Q^{\mathrm{T}} = \lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}} + \lambda_2\boldsymbol{q}_2\boldsymbol{q}_2^{\mathrm{T}} + \ldots + \lambda_r\boldsymbol{q}_r\boldsymbol{q}_r^{\mathrm{T}}$
任意矩阵 A	$A = U\Sigma V^{\mathrm{T}} = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \ldots + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{T}}$

请尊重版权与译者的劳动成果，侵权必究!

$\boldsymbol{q}$'s 是正交单位，$\boldsymbol{u}$'s 是正交单位，$\boldsymbol{v}$'s 是正交单位。美丽极了。

但是我想要再看一次，有两个理由。一个是修正一个在特征值部分的弱点，这个在第六章没有完全处理好。如果λ是 S 的一个双重根，我们可以也必须找出两个正交单位的特征向量。另一个理由是看看 SVD 如何在$\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$之前选中最大项$\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$。我们想要一次一个的了解这些特征值$\lambda$ (S 的)与奇异值σ (A 的)，而不是同时进行。

从 S 最大的特征值λ_1开始，它解答了下列问题：

$$\boldsymbol{\lambda_1 =}\ \textbf{极大比值}\ \frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\text{，致胜向量是 } \boldsymbol{x}=\boldsymbol{q}_1 \text{ 且 } S\boldsymbol{q}_1=\lambda_1\boldsymbol{q}_1 \tag{8}$$

比较 A 的最大奇异值σ_1，它解答了下列问题：

$$\boldsymbol{\sigma_1 =}\ \textbf{极大比值}\ \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\text{，致胜向量是 } \boldsymbol{x}=\boldsymbol{v}_1 \text{ 且 } A\boldsymbol{v}_1=\sigma_1\boldsymbol{u}_1 \tag{9}$$

这个“一次一个的方法”也应用在λ_2与σ_2，但是不是全部的 $\boldsymbol{x}$ 都是允许的：

$$\boldsymbol{\lambda_2 =}\ \textbf{极大比值}\ \frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\text{，其中 } \boldsymbol{x} \text{ 满足 } \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{x}=0\text{，}\boldsymbol{x}=\boldsymbol{q}_2 \text{ 会获胜} \tag{10}$$

$$\boldsymbol{\sigma_2 =}\ \textbf{极大比值}\ \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\text{，其中 } \boldsymbol{x} \text{ 满足 } \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{x}=0\text{，}\boldsymbol{x}=\boldsymbol{v}_2 \text{ 会获胜} \tag{11}$$

当 $S=A^{\mathrm{T}}A$ 我们发现$\lambda_1=\sigma_1^2$与$\lambda_2=\sigma_2^2$。为什么这个方法会成功？

从比值 $r(\boldsymbol{x})=\boldsymbol{x}S^{\mathrm{T}}\boldsymbol{x}/\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 开始，这个称为“拉里商(Rayleigh quotient)”。要想极大化 $r(\boldsymbol{x})$，令它的部分导数为零：$\partial r/\partial x_i=0$，其中 $i=1,\ldots,n$。这些导数很凌乱，下面是它的结果：针对致胜 $\boldsymbol{x}$ 的一个向量方程式：

$$\text{当 } S\boldsymbol{x}=r(\boldsymbol{x})\boldsymbol{x}\text{，则 } \boldsymbol{r(x)=}\frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}\ \textbf{的导数是零} \tag{12}$$

所以致胜 $\boldsymbol{x}$ 是 S 的一个特征向量，最大比值 $r(\boldsymbol{x})$是 S 的最大特征值λ_1，一切都很好。现在回到 A---注意与 $S=A^{\mathrm{T}}A$ 的关联！

$$\text{极大化}\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\text{ 也会极大化}\left(\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\right)^2=\frac{\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}=\frac{\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$$

所以(9)的致胜 $\boldsymbol{x}=\boldsymbol{v}_1$ 与(8)的 $S=A^{\mathrm{T}}A$ 的最高特征向量 $\boldsymbol{q}_1$ 相同。

现在我必须解释为什么 $\boldsymbol{q}_2$ 与 $\boldsymbol{v}_2$ 是(10)与(11)的致胜向量，我们知道他们与 $\boldsymbol{q}_2$ 及 $\boldsymbol{v}_2$ 正交，所以允许他们加入竞争。如果读者只想专注在 SVD 的动作(段落 7.3)，这些章节可以作为选读。

从任意正交矩阵 Q_1 开始，其中 $\boldsymbol{q}_1$ 是 Q_1 的第一列，其他 $n-1$ 个正交单位列都与 $\boldsymbol{q}_1$ 正交，然后使用 $S\boldsymbol{q}_1 = \lambda_1 \boldsymbol{q}_1$：

$$SQ_1 = S\,[\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \ldots \boldsymbol{q}_n] = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \ldots \boldsymbol{q}_n]\begin{bmatrix} \lambda_1 & \omega^{\mathrm{T}} \\ \mathbf{0} & S_{n-1} \end{bmatrix} = Q_1 \begin{bmatrix} \lambda_1 & \omega^{\mathrm{T}} \\ \mathbf{0} & S_{n-1} \end{bmatrix} \tag{13}$$

左乘 Q_1^{T}，记得 $Q_1^{\mathrm{T}}Q_1 = I$，确认 $Q_1^{\mathrm{T}} S\, Q_1$ 如同 S 是对称：

$$Q_1^{\mathrm{T}} S\, Q_1 = \begin{bmatrix} \lambda_1 & \boldsymbol{\omega}^{\mathrm{T}} \\ \mathbf{0} & S_{n-1} \end{bmatrix} \text{的对称性强迫 } \boldsymbol{w} = \mathbf{0} \text{ 且 } S_{n-1}{}^{\mathrm{T}} = S_{n-1}。$$

$\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{x} = 0$ 的要求会简化极大值问题(10)变成大小 $n-1$，S_{n-1} 的最大特征值会是 S 的第二大，它是λ_2。(10)的致胜下列会是特征向量 $\boldsymbol{q}_2$，其中 $S\boldsymbol{q}_2 = \lambda_2 \boldsymbol{q}_2$。

我们继续前进––或是使用神奇的字词“归纳(induction)”––产生所有的特征向量 $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ 与对应的特征值$\lambda_1, \ldots, \lambda_n$。纵使有重复特征值，光谱定理 $S = Q\Lambda Q^{\mathrm{T}}$ 仍然成立，所有的对称矩阵可以对角化。

相似的，从(9)与(11)以及后续方程式，SVD 是一次一个步骤求得，段落 7.4 会展示几何––我们在找的椭圆的轴。这里我问一个不同的问题：实际上λ's 与σ's 是如何计算的？

计算 S 的特征值与 A 的奇异值

A 的奇异值σ_i 是 $S = A^{\mathrm{T}}A$ 的特征值λ_i 的平方根，这个会把 SVD 与对称特征值问题产生关联(很棒)，但是最终我们不想计算 $A^{\mathrm{T}}A$(平方很花时间：不好)。

第一个概念是在没有改变任何σ's 与λ's 的情况下，产生在 A 与 S 的零，奇异向量与特征向量都会改变––没有问题。相似矩阵 $Q^{-1}SQ$ 与 S 有相同的λ's，如果 Q 是正交，这个矩阵是 $Q^{\mathrm{T}}SQ$ 而且仍然是对称。

段落 11.3 会展示如何从 2×2 旋转建立 Q，使得 $Q^{\mathrm{T}}SQ$ 是对称而且是三对角(很多零)。但是旋转无法一路产生对角矩阵，要得到 S 的全部特征值需要一个新概念以及更多的工作。

针对 SVD，什么会与 $Q^{\mathrm{T}}SQ$ 对应？现在我们不想改变 A 的任和奇异值。自然答案：你可以用两个不同的正交矩阵 Q_1 与 Q_2 去乘 A，利用他们产生 $Q_1^{\mathrm{T}}AQ_2$ 的零，σ's 没有改变：

$$(Q_1^{\mathrm{T}}AQ_2)^{\mathrm{T}}(Q_1^{\mathrm{T}}AQ_2) = Q_2^{\mathrm{T}}A^{\mathrm{T}}AQ_2 = Q_2^{\mathrm{T}}SQ_2 \text{ 得到相同的} \sigma(A) \text{与} \lambda(S)。$$

两个 Q's 的自由度允许我们得到 $Q_1^{\mathrm{T}}AQ_2$ =双对角矩阵(bidiagonal) (2 条对角线)，这就完美的达成了 $Q^{\mathrm{T}}SQ = 3$ 对角。注意两者之间的关联是很棒的事情：

(双对角)^{T}(双对角) = 三对角。

对角Λ与对角Σ的最终步骤需要更多点子，这个问题不简单，因为表面上我们在求解 $\det(S - \lambda I) = 0$，实际上多项式的次数是 $n = 100$ 或 1000 或更多，我们的人不会使用这些多项式。

请尊重版权与译者的劳动成果，侵权必究!

在 LAPACK 求取λ's 与σ's 最爱的方法，是使用简单的正交矩阵得到 $Q^{\mathrm{T}}SQ = \Lambda$ 以及 $U^{\mathrm{T}}AV = \Sigma$。当非常接近Λ与Σ时，我们就停止。

这个两步骤方法(零在前)收录到 eig(S)与 svd(A)。

主要观念的复习

1. SVD 把 A 分解成 $U\Sigma V^{\mathrm{T}}$，其中有 r 个奇异值$\sigma_1 \geq ... \geq \sigma_r > 0$。
2. 数字$\sigma_1^2, ..., \sigma_r^2$ 是 AA^{T} 与 $A^{\mathrm{T}}A$ 的非零特征值。
3. U 与 V 的正交单位列是 AA^{T} 与 $A^{\mathrm{T}}A$ 的特征向量。
4. 这些列支撑了 A 的 4 个基础子空间的正交基底。
5. 这些基底使得矩阵对角化：$A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$，其中 $i \leq r$，这就是 $AV = U\Sigma$。
6. $A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + ... + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{T}}$ 且σ_1 是比值$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$的极大值。

已解范例

7.2A 依据名字鉴别下列使得 A 变成列乘行的总和的分解：

1.	正交列	$\boldsymbol{u}_1\sigma_1, ..., \boldsymbol{u}_r\sigma_r$	乘	正交单位行	$\boldsymbol{v}_1^{\mathrm{T}}, ..., \boldsymbol{v}_r^{\mathrm{T}}$
2.	正交单位列	$\boldsymbol{q}_1, ..., \boldsymbol{q}_r$	乘	三角行	$\boldsymbol{r}_1^{\mathrm{T}}, ..., \boldsymbol{r}_r^{\mathrm{T}}$
3.	三角列	$\boldsymbol{l}_1, ..., \boldsymbol{l}_r$	乘	三角行	$\boldsymbol{u}_1^{\mathrm{T}}, ..., \boldsymbol{u}_r^{\mathrm{T}}$

A 的秩与枢轴与奇异值落在图形的哪儿？

解 这三种分解是线性代数的基础，纯数学或是应用数学：

1. 奇异值分解 $A = U\Sigma V^{\mathrm{T}}$
2. 格莱姆-施密特正交化分解 $A = QR$
3. 高斯消元法 $A = LU$

你可能偏好分离奇异值σ_i与高度 h_i与枢轴 d_i：

1. $A = U\Sigma V^{\mathrm{T}}$ 在 U 与 V 有单位向量，r 个奇异值σ_i在Σ中。
2. $A = QHR$ 在 Q 有单位向量，在 R 有对角线的 1's，r 个高度 h_i 在 H 中。
3. $A = LDU$ 在 L 与 U 对角线的 1's，r 个枢轴 d_i 在 D 中。

每个 h_i 告知列 i 在列 1 至列 $i-1$ 的平面上方的高度，完整 n 维盒子的体积($r = m = n$)来自 $A = U\Sigma V^{\mathrm{T}} = LDU = QHR$：

$$|\det A| = |\ \sigma\text{'s 的乘积}\ | = |\ d\text{'s 的乘积}\ | = |\ h\text{'s 的乘积}\ |$$

7.2B 证明$\sigma_1 \geq |\lambda|_{\max}$。最大的奇异值支配所有的特征值。

解 从 $A = U\Sigma V^{\mathrm{T}}$ 开始，记得左乘一个正交矩阵不会改变长度：$\| Q\boldsymbol{x} \| = \| \boldsymbol{x} \|$，因为$\| Q\boldsymbol{x} \|^2 = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \| \boldsymbol{x} \|^2$。这个性质应用到 $Q = U$ 与 $Q = V^{\mathrm{T}}$，在中间的是对角矩阵Σ：

$$\| A\boldsymbol{x} \| = \| U\Sigma V^{\mathrm{T}}\boldsymbol{x} \| = \| \Sigma V^{\mathrm{T}}\boldsymbol{x} \| \leq \sigma_1 \| V^{\mathrm{T}}\boldsymbol{x} \| \leq \sigma_1 \| \boldsymbol{x} \| \tag{14}$$

特征向量有$\| A\boldsymbol{x} \| = |\lambda| \| \boldsymbol{x} \|$，所以(14)说明$| \lambda | \| \boldsymbol{x} \| \leq \sigma_1 \| \boldsymbol{x} \|$，则$|\lambda| \leq \sigma_1$。

同样应用到单位向量 $\boldsymbol{x} = (1, 0, ..., 0)$，现在 $A\boldsymbol{x}$ 是 A 的第一列，依据(14)，这个列的长度$\leq \sigma_1$，每个单元必须有$|a_{ij}| \leq \sigma_1$。

方程式(14)再次展示$\| A\boldsymbol{x} \| / |\lambda| \| \boldsymbol{x} \|$的极大值等于$\sigma_1$。

段落 11.2 会说明在求解 $A\boldsymbol{x} = \boldsymbol{b}$ 时，比值$\sigma_{\max}/\sigma_{\min}$ 掌管舍去误差。如果这个“条件数字”太大，MATLAB 会警告你，这个 $\boldsymbol{x}$ 就不可靠。

问题集 7.2

1 求出下列矩阵的特征值，再从 $A^{\mathrm{T}}A$ 求出奇异值：

$$A = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix} \qquad A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$$

对每个 A，从 $A^{\mathrm{T}}A$ 的特征向量创建 V，从 AA^{T} 的特征向量创建 U，检验 $A = U\Sigma V^{\mathrm{T}}$。

2 求出 $A^{\mathrm{T}}A$ 与 V 与Σ与 $\boldsymbol{u}_i = A\boldsymbol{v}_i/\sigma_i$ 以及全部的 SVD：

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = U\Sigma V^{\mathrm{T}}$$

3 在问题 2，展示 AA^{T} 是对角，它的特征向量 $\boldsymbol{u}_1, \boldsymbol{u}_2$ 是_____，它的特征值σ_1^2, σ_2^2 是_____。A 的行是正交但是他们不是_____，所以 A 的列不是正交。

4 计算 $A^{\mathrm{T}}A$ 与 AA^{T} 以及他们对应 U 与 V 的特征值及单位特征向量。

矩形矩阵 $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

检验 $AV = U\Sigma$(这个决定 U 中的±号)，Σ有与 A 相同的形状：2×3。

5 (a) $A = \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$的行空间是一个 1-维，求出在行空间的 $\boldsymbol{v}_1$ 与在列空间的 $\boldsymbol{u}_1$。σ_1 为何？为什么没有σ_2。

请尊重版权与译者的劳动成果，侵权必究!

(b) 选择 V 与 U 中的 $\boldsymbol{v}_2$ 与 $\boldsymbol{u}_2$，则 $A = U\Sigma V^{\mathrm{T}} = \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1{}^{\mathrm{T}}$(只有一项)。

【原文是 U and V，换过来比较合适】

6 把 A 与 A^{T} 的 SVD 代入，展示 $A^{\mathrm{T}}A$ 的特征值在$\Sigma^{\mathrm{T}}\Sigma$中，$AA^{\mathrm{T}}$ 的特征值在$\Sigma\Sigma^{\mathrm{T}}$ 中。对角矩阵$\Sigma^{\mathrm{T}}\Sigma$与$\Sigma\Sigma^{\mathrm{T}}$ 有相同的非零数字，我们再次看到 $A^{\mathrm{T}}A$ 与 AA^{T} 有相同的非零特征值。

7 若$(A^{\mathrm{T}}A)\boldsymbol{v} = \sigma^2\boldsymbol{v}$，左乘 A，移动括号得到$(AA^{\mathrm{T}})A\boldsymbol{v} = \sigma^2(A\boldsymbol{v})$。若 $\boldsymbol{v}$ 是 $A^{\mathrm{T}}A$ 的特征向量，则_____是 AA^{T} 的特征向量。

8 求出 $A^{\mathrm{T}}A$ 的特征值与单位特征向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$，然后求出 $\boldsymbol{u}_1 = A\boldsymbol{v}_1/\sigma_1$：

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \qquad A^{\mathrm{T}}A = \begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix} \qquad AA^{\mathrm{T}} = \begin{bmatrix} 5 & 15 \\ 15 & 45 \end{bmatrix}$$

验证 $\boldsymbol{u}_1$ 是 AA^{T} 的单位特征向量，计算矩阵 U, Σ, V。

$$\text{SVD} \qquad A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}^{\mathrm{T}}$$

9 写下这个 A 的 4 个基础子空间的正交单位基底。

10 (a) 为什么 $A^{\mathrm{T}}A$ 的迹等于所有的 $a_{ij}{}^2$ 的总和？在范例 3 它是 50。

(b) 对于每个秩 1 矩阵，为什么$\sigma_1{}^2$ 等于所有的 $a_{ij}{}^2$ 的总和？

11 求出 $A^{\mathrm{T}}A$ 与 AA^{T} 的特征值与单位特征向量，使得每个 $A\boldsymbol{v} = \sigma\boldsymbol{u}$。然后创建奇异值分解并且验证 A 等于 $U\Sigma V^{\mathrm{T}}$。

$$\textbf{费氏矩阵} \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

12 使用 MATLAB 展示 eigshow 的 SVD 部分求出 $\boldsymbol{v}$'s 的图形。

13 若 $A = U\Sigma V^{\mathrm{T}}$ 是方形可逆矩阵，则 A^{-1} = _____ _____ _____。检查 $A^{-1}A$。这个展示了 A^{-1} 的奇异值是 $1/\sigma_i$。

注解：A^{-1} 的最大奇异值是 $1/\sigma_{\min}(A)$，最大的特征值$|\lambda(A^{-1})|_{\max}$ 是 $1/|\lambda(A^{-1})|_{\min}$，则方程式(14)说明了$\sigma_{\min}(A) \le |\lambda(A^{-1})|_{\min}$。

14 假设 $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ 与 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ 是$\mathbf{R}^n$ 的正交单位基底，创建矩阵 $A = U\Sigma V^{\mathrm{T}}$，把每个 $\boldsymbol{v}_j$ 变换成 $\boldsymbol{u}_j$ 得到 $A\boldsymbol{v}_1 = \boldsymbol{u}_1, ..., A\boldsymbol{v}_n = \boldsymbol{u}_n$。

15 创建秩 1 矩阵使得 $A\boldsymbol{v} = 12\boldsymbol{u}$，其中 $\boldsymbol{v} = \frac{1}{2}(1, 1, 1, 1)$与 $\boldsymbol{u} = \frac{1}{3}(2, 2, 1)$。它的唯一奇异值是$\sigma_1$ = ____。

16 假设 A 有正交列 $\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_n$，长度分别是$\sigma_1, \sigma_2, ..., \sigma_n$。SVD 中的 U, Σ, V 为何？

17 假设 A 是 2×2 对称矩阵，它的单位特征向量是 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$。如果它的特征值是$\lambda_1 = 3$ 与$\lambda_2 = -2$，则 SVD 中的 $U, \Sigma, V^{\mathrm{T}}$ 为何？

18 若 $A = QR$，Q 是正交矩阵，A 的 SVD 几乎与 R 的 SVD 相同。因为 Q，这三个矩阵 U, Σ, V 何者产生变化？

19 假设 A 是可逆($\sigma_1 > \sigma_2 > 0$)，对 A 进行尽可能小的改变得到奇异矩阵 A_0。提示：U 与 V 不变。

$$\text{从} \quad A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}^{\mathrm{T}} \quad \text{求最接近的 } A_0$$

20 从指令 svd(A)或是手算，求出 A 的奇异值。

$$A = \begin{bmatrix} 1 & 0 \\ 100 & 0 \end{bmatrix} \quad \text{为什么这个矩阵有} \sigma_2 = 1/\sigma_1 ?$$

21 为什么 $A + I$ 的 SVD 直接使用 $\Sigma + I$?

22 若 $A = U\Sigma V^{\mathrm{T}}$ 则 $Q_1 A Q_2^{\mathrm{T}} = (Q_1 U)^{\mathrm{T}}\ \Sigma(\ Q_2 V)^{\mathrm{T}}$。为什么任意的正交矩阵 Q_1 与 Q_2 会使得 $Q_1 U =$ 正交矩阵且 $Q_2 V =$ 正交矩阵？然后 Σ 在奇异值中看不到变化：$Q_1 A Q_2^{\mathrm{T}}$ 与 A 有相同的 σ's。

23 若 Q 是正交矩阵，为什么它的所有奇异值都是 1？

24 (a) 求出 $\dfrac{\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} = \dfrac{3x_1^{\,2} + 2x_1 x_2 + 3x_2^{\,2}}{x_1^{\,2} + x_2^{\,2}}$ 的极大值，矩阵 S 为何？

(b) 求出 $\dfrac{(x_1 + 4x_2)^2}{x_1^{\,2} + x_2^{\,2}}$ 的极大值，这个 $\dfrac{\|A\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}$ 中的矩阵 A 为何？

25 比值 $\dfrac{\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}}$ 与 $\dfrac{\|A\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}$ 的极小值为何？我们应该选择 S 的哪一个特征向量？$\boldsymbol{x}$ 永远是 A 的特征向量吗？

26 每个矩阵 $A = U\Sigma V^{\mathrm{T}}$ 使得圆变成椭圆。$AV = U\Sigma$ 说明圆的半径向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 来到椭圆的半轴(semi-axes) $\sigma_1 \boldsymbol{u}_1$ 与 $\sigma_1 \boldsymbol{u}_2$。画出这个圆以及 $\theta = 30°$ 的椭圆：

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

段落 7.4 会从关于 2×2 矩阵的一个重要的 SVD 图形开始：

A = (旋转) (拉伸) (旋转)，且对称 S = (旋转) (拉伸) (旋转回去)

27 这个问题在寻找所有的矩阵 A，其中 A 给定一个在 $\mathbf{R}^m$ 的列空间以及给定一个在 $\mathbf{R}^n$ 的行空间。假设 $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_r$ 与 $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$ 是这两个空间的基底，然后分别进入 C 与 B 的列，我们的目的是展示 A 有下列形式：

$A = CMB^{\mathrm{T}}$，其中 M 是 $r \times r$ 的可逆矩阵。提示：从 $A = U\Sigma V^{\mathrm{T}}$ 开始。

U 与 V 的前 r 个列必须借由可逆矩阵与 C 及 B 产生关联，因为他们包含对应相同列空间(在 U)与行空间(在 V)的基底。 &

请尊重版权与译者的劳动成果，侵权必究!

7.3 主要分量分析(SVD 做 PCA)

1 资料通常写成矩阵：n 个样本，每个样本有 m 个量测值。

2 减去来自每个量测值的平均使得 A 的每个行中心化。

3 SVD 求出包含最多信息的资料的组合。

4 最大的奇异值$\sigma_1 \leftrightarrow$ 最大的方差 $\leftrightarrow \boldsymbol{u}_1$ 中最多的信息。

这个段落说明 SVD 在统计与资料分析的主要应用。我们的例子会来自人类遗传学与辨识与金融，问题在于了解一个大型的资料(=量测值)矩阵。我们对 n 个样本的每一个样本量测 m 个变数，资料矩阵 A_0 有 n 列 m 行。

从图形来看，A_0 的列是$\mathbf{R}^m$的 n 个点，在减去每一行的平均值之后得到 A，这 n 个点通常会沿着一条直线聚集，或是接近一个平面(或是其他$\mathbf{R}^m$的低维子空间)。这些直线或平面或子空间为何？

让我从图形而不是数字开始，当 $m = 2$ 个变数，例如年纪与身高，这 n 个点落在 $\mathbf{R}^2$ 的平面。减去平均年纪与平均身高使得资料中心化，如果 n 个重置中心点沿着直线聚集，线性代数要如何找出这条直线？

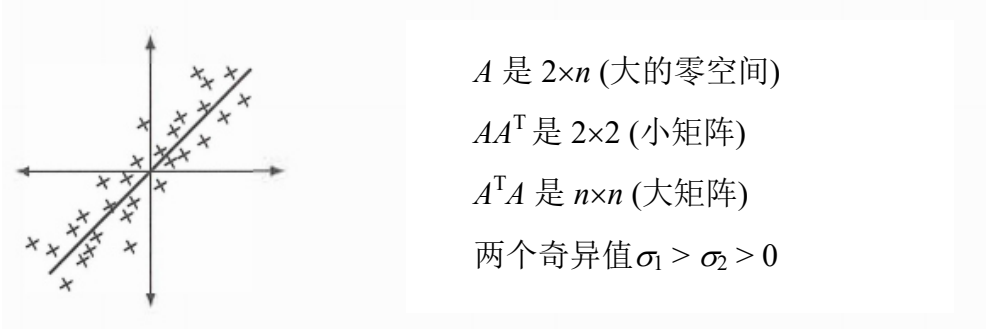


图 7.2：A 中的资料点通常靠近 $\mathbf{R}^2$ 的直线或是 $\mathbf{R}^m$ 的子空间。

让我仔细一点来创建资料矩阵。从 A_0 的量测值开始，找出每一行的平均值(平均数 mean)$\mu_1, \mu_2, \ldots, \mu_m$，从行 i 减去每个平均数μ_i使得资料中心化，现在中心化矩阵 A 的每个行的平均值变成零，所以图 7.2 的点(0, 0)现在是 n 个点的真正中心。

样本协方差矩阵(covariance matrix)的定义是 $\boldsymbol{S = \dfrac{AA^{\mathrm{T}}}{n-1}}$

A 显示来自每个量测值与行平均值μ_i的距离 $a_{ij} - \mu_i$。

$(AA^{\mathrm{T}})_{11}$ 与$(AA^{\mathrm{T}})_{22}$ 显示距离平方的总和(样本方差 s_1^2, s_2^2)。

$(AA^{\mathrm{T}})_{12}$ 显示样本协方差 $s_{12} = (A$ 的行 1$) \cdot (A$ 的行 2)。

方差是贯穿统计学的关键数字，一个考试的平均成绩$\mu = 85$ 告诉你这是一个相当不错的考试。一个 $s^2 = 25$(标准差 $s = 5$)的方差表示大部分的成绩都在 80 左右：密集堆积。一个样本的方差 $s^2 = 225$($s = 15$)表示分数是宽广分散。第十二章会说明方差。

数学考试与历史考试的协方差(covariance)是 A 的那些行的点积，这里事先把平均成绩减去。协方差小于零表示：一个科目强另一个科目弱。高的协方差表示：两者都强或是都弱。

我们除以 $n-1$ 而不是除以 n，是基于已知对统计学最好的理由。他们告诉我平均值用掉了一个自由度，剩下 $n-1$ 个。(我想最好的计划是同意他们)，任何情况下 n 应该是一个大的数字来得到可靠的统计。既然 A 的行有 n 个单元，AA^{T} 的大小成长的像 n，除以 $n-1$ 维持他们的稳定性。

范例 1　六个数学与历史的成绩(注意每个行的零平均)

$$A=\begin{bmatrix}3 & -4 & 7 & 1 & -4 & -3\\ 7 & -6 & 8 & -1 & -1 & -7\end{bmatrix}\text{有样本协方差 } S=\frac{AA^{\mathrm{T}}}{5}=\begin{bmatrix}20 & 25\\ 25 & 40\end{bmatrix}$$

A 的两个行是高度关联：$s_{12}=25$，高于数学平均会有低于历史平均，改变行 2 所有的符号得到负数协方差 $s=-25$。注意 S 有正数迹与正数行列式；AA^{T} 是正定。

S 的特征值接近 57 与 3，所以第一个秩 1 片段 $\sqrt{57}\,\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ 比第二个片段 $\sqrt{3}\,\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$ 大很多。你在图 7.2 的聚集图形看到的是带头特征向量 $\boldsymbol{u}_1$ 的方向，那个特征向量接近 $\boldsymbol{u}_1=(.6,.8)$，而且图形里面的方向近似得到 6 – 8 – 10 或是 3 – 4 – 5 的直角三角形。

A(中心化资料)的 SVD 展示在聚集图的支配方向。

第二奇异向量 $\boldsymbol{u}_2$ 与 $\boldsymbol{u}_1$ 垂直，第二个奇异值 $\sigma_2\approx\sqrt{3}$ 在量测横跨支配线的散布(spread)情况。如果 A 中的资料点确切落在直线上($\boldsymbol{u}_1$ 的方向)，则 σ_2 会是零，实际上只有一个 σ_1。

主要分量分析(PCA)的重要性

PCA 给你一个方法去了解在维度 m =量测变数的个数(此处是年纪与身高)的资料图，减去平均年纪与平均身高(n 个样本，每个样本 $m=2$)得到中心化的 $m\times n$ 矩阵 A。这个与线性代数的重要关联在于 A 的奇异值与奇异向量，他们来自特征值 $\lambda=\sigma^2$ 以及样本协方差矩阵 $S=AA^{\mathrm{T}}/(n-1)$ 的特征向量 $\boldsymbol{u}$。

- 资料的**总方差**(total variance)是所有特征值的总和，也是样本方差 s^2 的总和：

  总方差 $T=\sigma_1^2+\ldots+\sigma_m^2=s_1^2+\ldots+s_m^2=$ 迹(对角线的总和)

- S 的第一个特征向量 $\boldsymbol{u}_1$ 指向资料最显著的方向，这个方向代表(或是说明)总方差的一个分数 σ_1^2/T。
- 下一个特征向量 $\boldsymbol{u}_2$ (与 $\boldsymbol{u}_1$ 正交)代表较小的分数 σ_2^2/T。
- 当这些分数很小时我们就停止，你有 R 个方来向说明大部分的资料。这 n 个资料点非常接近 R-维子空间，子空间有基底 $\boldsymbol{u}_1$ 到 $\boldsymbol{u}_r$。这些 $\boldsymbol{u}$'s 是在 m-维空间的**主要分量**(principal component)。
- R 是 A 的“有效秩”，真正的秩 r 可能是 m 或 n：满秩矩阵。

请尊重版权与译者的劳动成果，侵权必究!

垂直最小平方

有件事情并不广为人知，图 7.2 的最佳直线(直线在 $\boldsymbol{u}_1$ 的方向)也可以求解垂直最小平方的问题(=正交回归(regression))：

点到线的距离平方的总合是个极小值。

证明：把每个列 a_j 分成沿着 $\boldsymbol{u}_1$ 线分量与沿着 $\boldsymbol{u}_2$ 线的分量：

$$\text{直角三角形} \qquad \sum_{j=1}^{n} \| \boldsymbol{a}_j \|^2 = \sum_{j=1}^{n} \left| \boldsymbol{a}_j{}^{\mathrm{T}} \boldsymbol{u}_1 \right|^2 + \sum_{j=1}^{n} \left| \boldsymbol{a}_j{}^{\mathrm{T}} \boldsymbol{u}_2 \right|^2 \qquad (1)$$

左侧来自资料点 $\boldsymbol{a}_j$(A 的列)的总和是固定的，右侧的第一个总和是 $\boldsymbol{u}_1{}^{\mathrm{T}} A A^{\mathrm{T}} \boldsymbol{u}_1$，所以当我们依据选取特征向量 $\boldsymbol{u}_1$ 来极大化 PCA 中的总和时，我们要极小化第二个总和，第二个总和(从资料点到最佳线的距离平方)就是垂直最小平方的极小值。

第四章的正常最小平方得到一个线性方程式 $A^{\mathrm{T}} A \hat{\boldsymbol{x}} = A^{\mathrm{T}} \boldsymbol{b}$，使用的是点到最佳线的铅锤线(vertical)距离。PCA 产生一个对应 $\boldsymbol{u}_1$ 的特征值问题，使用的是垂直(perpendicular)距离。“总最小平方”会允许 A 的误差，如同 $\boldsymbol{b}$ 的误差一般。

【vertical 是铅锤线距离】

样本关联矩阵

资料分析通常是在处理 A(中心化资料)，但是 A 的量测值可逆有不同的单位，例如吋，磅，年，元。改变一组单位(吋到米或年到秒)对于 A 与 S 的行会有巨大影响。如果调整比例是个问题，我们把协方差矩阵 S 改成关联(correlation)矩阵 C：

一个对角矩阵 D 调整 A 的比例，DA 的每一行的长度是 $\sqrt{n-1}$ 。

样本关联矩阵 $\boldsymbol{C = DAA^{\mathrm{T}}D/n - 1}$ 有 1's 在对角线上。

第十二章的概率与统计会介绍期望(expected)协方差矩阵 V 以及期望关联矩阵(1's 在对角线)，这些不是使用实际量测值而是使用概率。协方差矩阵预测未来量测值环绕平均值的分布，此时 A 与样本协方差 S 以及重调比例的关联矩阵 $C = DSD$ 都是使用实际的资料。全部都很重要--在统计与正定矩阵的线性代数与 SVD 是一个很大的联结。 【SNP：单核苷酸多态性】

遗传变异

我们可以检视染色体(genomes)来追踪人口数量的变异，要管理这么大量的资料件，有一个好方法是从 SNP's 观察遗传变异(genetic variation)。不寻常的对偶基因(alleles)(基于来自父亲与母亲的 A/C/T/G 配对)由 SNP 计数：

SNP = 0　没有来自人口数量普通基准的改变：正常基因型。

SNP = 1　基准配对显示有一个来自正常配对的改变。

SNP = 2　两种基准都比普通基因型少。

未中心化的矩阵 A_0 有一列对应每个人，有一行对应每个基准配对，矩阵的单元大部分是 0，很多 1，不多的 2。我们不测试全部 300 万个配对。从 A_0 减去每行的平均值之后，AA^{T} 的特征向量是极端发人深思的。在图 7.4，A 的第一个奇异向量几乎重建了整个欧洲地图。

这个表示：来自法国与德国与意大利的 SNP's 有很大的不同，就算是来自法国与德国与瑞士的意大利部分，这些“snips”都不一样！只有西班牙与葡萄牙是惊讶的混淆并且难以分开。通常来说，每个个人的 DNA 会显示他在 300 公里或 200 英里内的出生地。祖父辈的混合通常安置孙子辈在他们的原始出身之间。

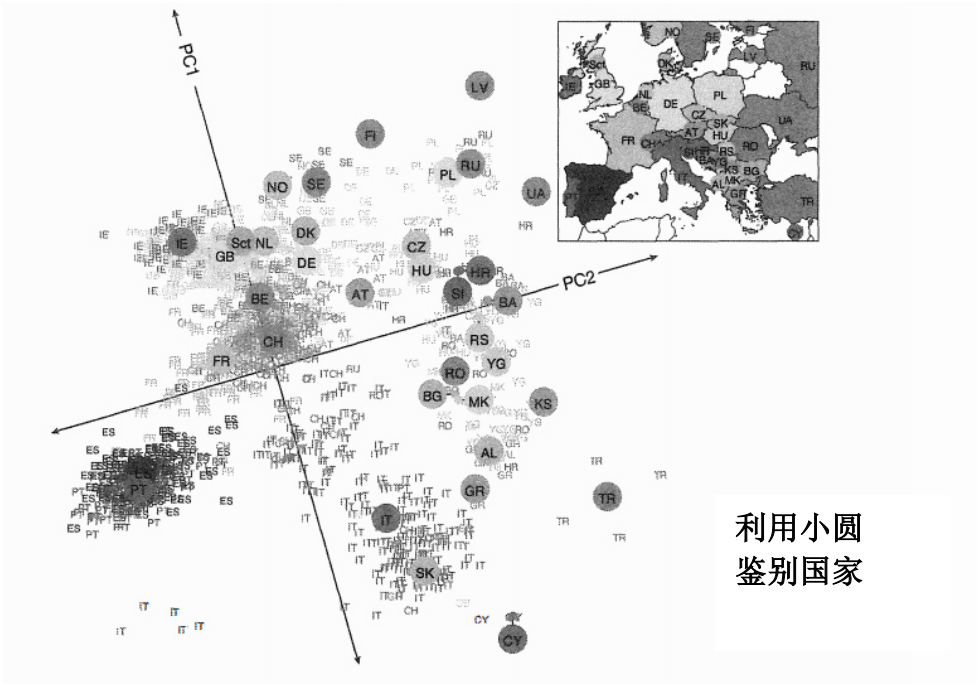


图 7.3：自然(2008) Novembre et al: vol. 456 pp.98-101/doc:1038/nature07331

什么是最显著的信息？如果我们测试染色体组取去了解他们与疾病的关联，我们不能忘记他们的空间变异。没有进行地理校正之下，那些看起来医疗上的特殊意义会有误导性。在医疗基因上混淆是一个严重的问题，PCA 与人口基因可以帮助求解––移除来自没有重要医疗的地理性。

事实上“空间统计”是难以应付的世界。例子：每个有三个对角 1, *C*, 1 的矩阵显示一个不令人惊讶的影响，其中的影响来自下一个隔门的邻居(来自 1's)。但是它的奇异向量在交叉地图上有正弦与余弦震荡，与 *C* 无关。你可能想到这些是真实的“像波”震动，但是他们可能是没有意义的。

也许统计学比数学产生更多的理论？简化大资料成为小信号“*P* 值”具有指导性或是极端的误导性。这个 *P* 值表示法出现在许多矩阵，*P* 代表与零假设(null hypothesis，纯机会)一致的概率。如果你在一个行看到 5 个人头的概率是 *P* = 1/32，这个来自公平铜板的机会(或是 2/32，如果你的观察是连续 5 个人头或是 5 个尾巴)。通常 *P* 值小于 0.05 使得零假设是可疑的––可能一个骗子在反转铜板。在这里，*P* 值不是统计学中最可靠的指引––但是他们非常方便。

请尊重版权与译者的劳动成果，侵权必究!

人脸辨识乍看之下--第一印象--与线性代数无关，但是 SVD 一个早期并且公开的应用就是**人脸辨识**。我们不压缩图像，我们辨识它。

计划是从“训练集合”A_0 开始，其中 A_0 有 n 个变化剧烈的图像。把所有的像素灰度堆积在一个列，每个图像变成一个非常长的向量，然后 A_0 必须中心化：减去 A_0 的**每一列**的平均值得到 A。

这个 A 的奇异向量 $\boldsymbol{v}_1$ 告诉我们已知人脸的组合，这些组合是新的人脸辨识的最佳，然后 $\boldsymbol{v}_2$ 告诉我们下一个最佳组合。

我们可能会使用 R 个最佳向量 $\boldsymbol{v}_1, ..., \boldsymbol{v}_R$，具有 A 的最大的奇异值 $\sigma_1 \geq ... \geq \sigma_R$，这些在辨识新的人脸会比其他 R 个向量更精确。也许 $R = 100$ 个**固有人脸**(eigenface)$A\boldsymbol{v}$ 会捕捉在训练集合中几乎所有的变异，这些 R 个**固有人脸**生成“**人脸空间**”。

这个攻击计划是由 Matthew Turk 与 Alex Pentland 所提出，它开发了 Sirovich 与 Kirby 使用 PCA 在压缩人脸图像的成果。我在网上学习了很多来自 Jeff Jauregui's 的描述，他的注解是：**PCA 提供透过代数平均值的一个机制去辨识几何的/光度的相似**。他组合了第一主要分量(第一奇异向量)进入第一固有人脸，当然每一列的平均值要加回去否则你看不到一张脸。

注解 在一封给 Nature 的信中，PCA 与 NMF 相提并论(Lee 与 Seung, vol. 401, 21 Oct. 1999)。非负矩阵分解不允许永远出现在奇异向量 $\boldsymbol{v}$ 的负数单元，所以每个东西都相加--需要更多的向量但是他们通常更有意义。

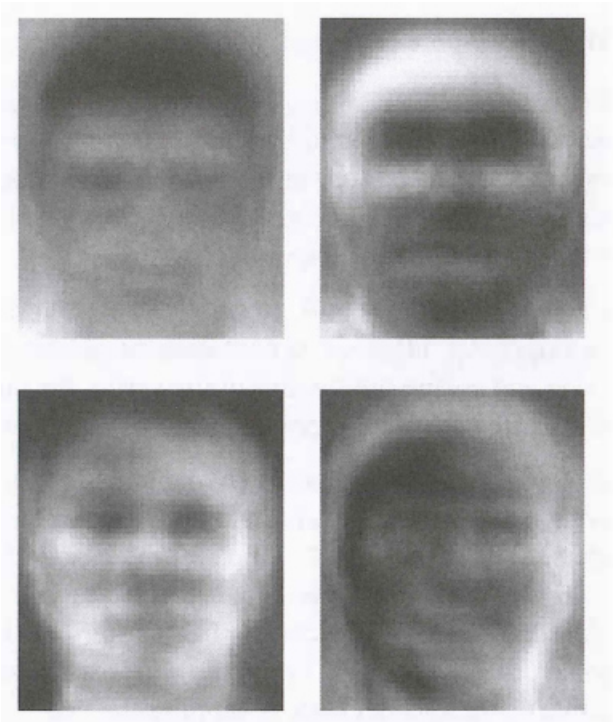


图 7.4：固有人脸筛选发际线与嘴与眼睛与形状。

固有人脸的应用

PCA 人脸辨识的第一个商业用途是法律执行与安全，在 Tampa 举行的超级碗 35 的初期测试得到来自群众非常负面的反应！球迷并不知道这项测试，报纸开始称它是“窥探(snooper)碗”。我不认为原始固有人脸概念还在商业使用(甚至是保密)。

SVD 方法的新应用引来新的鉴定问题：固有声音，固有步伐，固有眼睛，固有表情。我从 Matthew Turk(现在 Santa Barbara，原来是 MIT 的研究生，他告诉我他曾经上过我的课。)学习了这个，在它的论文里面固有人脸遇到面部图像的旋转，比例与光线的问题。但是关键概念还在。

最终，人脸空间是非线性的，所以我们还是需要非线性 PCA。

模型降阶

对于大型规模动态系统，计算成本会变得无法管理，“动态”表示解 $\boldsymbol{u}(t)$随着时间变化。流体，化学反应，波传导，生物成长，电子系统，这些问题无所不在。**简化的模型试着去鉴别系统的重要状态**，从简化的问题我们计算所需信息的成本大大的降低。

模式简化在计算方法非常重要，很多优质概念被提出来降低原始大型规模的问题，一个简单又有用的概念是取得流体的“快照(snapshot)”把他们放进矩阵 A，找出主要分量(A 的左奇异向量)，然后在小很多的子空间中工作：

快照是一个列向量，用来描述系统的状态。

它可以是典型真实状态 $\boldsymbol{u}(t^*)$的近似

从 n 个快照，建立一个矩阵 A，它的列生成状态的有用值域

现在求出 A 的前 R 个左奇异向量 $\boldsymbol{u}_1$ 到 $\boldsymbol{u}_R$，他们是“本征(proper)正交分解”的一组基底(POD 基底)。实际上我们选择 R 使得

方差≈能量 $\sigma_1^2 + ... + \sigma_R^2$ 是 99%或 99.9%的 $\sigma_1^2 + ... + \sigma_n^2$

这些向量是重建 A 的快照的最佳基底，如果这些快照选择良好，$\boldsymbol{u}_1$ 到 $\boldsymbol{u}_R$ 的组合在预想的时间 t 与参数 p 会很接近确切解 $\boldsymbol{u}(t)$。

很多都与快照相关！SIAM Review 2015 包含由 Beiner, Gugercin 与 Willcox 提出的非常棒的总论。SVD 压缩资料也压缩图像。

网络搜寻

我们相信谷歌建立一个追踪网页链接的等级系统，当路径经常到一个网址，等级就高。访问的频率给出“网页矩阵”的引导特征向量($\lambda = 1$)--曾经求解的最大特征值问题。

马可夫矩阵有超过 30 亿的行与列，来自 30 亿个网址。

请尊重版权与译者的劳动成果，侵权必究!

谷歌持有很多重要科技，他们可能从早期的特征向量开始作为近似值，然后快速进行随机路径。为了得到较高等级，你希望从重要网址得到很多链接。

这是 SVD 在搜寻引擎的一个应用，当你搜寻一个字词，依据重要性你得到一个网址列表，你可以试着键入“4 个子空间”。

HITS 演绎法是较早得到等级列表的方法，它从关键字的索引的大约 200 个网站开始，之后我们只观察页与页之间的链接，搜寻引擎是链接基而不是内容基。

从 200 个网站开始与所有链接他们以及他们链接的网站，这是我们义序放置的列表，可以使用连入与连出来衡量重要性。

1. 网站可能是权威：有很多进入的链接，特别是从核心来的链接。
2. 网站可能是核心：有很多列表中外出的链接，特别是到权威。

我们希望数字 $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$ 去衡量权威的等级以及 $\boldsymbol{y}_1, ..., \boldsymbol{y}_N$ 去衡量核心的等级。从一个简单的计数开始：$\boldsymbol{x}_i^0$ 与 $\boldsymbol{y}_i^0$ 计数网站 i 的连入与连出。

此处是重点：一个优质权威有从重要网站来的链接(比如核心)，从大学来的链接比来自朋友的链接比重大一点。一个优质核心要链接到重要网站(比如权威)。链接到 amazon.com 不幸的比 welleseleycambridge.com 来的重要。初步计数 $\boldsymbol{x}^0$ 与 $\boldsymbol{y}^0$ 借由优质链接更新成 $\boldsymbol{x}^1$ 与 $\boldsymbol{y}^1$(利用 $\boldsymbol{x}^0$ 与 $\boldsymbol{y}^0$ 量测他们的品质)：

$$\textbf{权威/核心} \qquad \boldsymbol{x}_i^1/\boldsymbol{y}_i^1 = \boldsymbol{y}_j^0/\boldsymbol{x}_j^0 \text{ 的累加，对所有的}\textbf{连入 } \boldsymbol{i}/\textbf{从 } \boldsymbol{i} \textbf{ 连出} \tag{2}$$

用矩阵语言来说这些是 $\boldsymbol{x}^1 = A^{\mathrm{T}}\boldsymbol{y}^0$ 与 $\boldsymbol{y}^1 = A\boldsymbol{x}^0$，矩阵 A 包含 1's 与 0's，其中 $a_{ij} = 1$ 当 i 连到 j。用图形的语言来说，A 是一个网络的“相邻矩阵”(巨大的矩阵)。新的 $\boldsymbol{x}^1$ 与 $\boldsymbol{y}^1$ 给出较好的等级，但不是最佳。再走像(2)的一步，从 $A^{\mathrm{T}}A\boldsymbol{x}^0$ 与 $AA^{\mathrm{T}}\boldsymbol{y}^0$ 到达 $\boldsymbol{x}^2$ 与 $\boldsymbol{y}^2$：

$$\textbf{权威} \qquad \boldsymbol{x}^2 = A^{\mathrm{T}}\boldsymbol{y}^1 = A^{\mathrm{T}}A\boldsymbol{x}^0 \quad \textbf{核心} \qquad \boldsymbol{y}^2 = A\boldsymbol{x}^1 = AA^{\mathrm{T}}\boldsymbol{y} \tag{3}$$

在两步内我们用 $A^{\mathrm{T}}A$ 与 AA^{T} 去乘，20 步用$(A^{\mathrm{T}}A)^{10}$ 与$(AA^{\mathrm{T}})^{10}$ 去乘。当我们取次方时，最大的特征值σ_1^2 开始支配，向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 会跟 $A^{\mathrm{T}}A$ 与 AA^{T} 的带头特征向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{u}_1$ 对齐。我们在计算 SVD 的顶端项，利用会在段落 11.3 讨论的次方方法，它是线性代数协助了解网络的奇妙方法，

这个HITS演绎法在1996科学美国(Jan 16)有描述，但是我不认为有提到SVD。最佳的书籍是 Langville 与 Meyer 写的”谷歌页等级与之上”，说明了搜寻引擎的详细科学。

金融中的 PCA：利率的动力学

金融数学持续使用线性代数与 PCA，我们选择一种应用：**国库债券的收益曲线**。收益是指国债，票据与债券的利率支付。利率与到期时间有关，长期国债(3 年到 20 年)的利率随长度增加，美联储调整短期收益来减慢或刺激经济。这是收益曲线，被风险管理师与证券商与投资者广泛使用。

这是 2001 年前 6 个营业日的资料--每一列是在某个特定日的投资收益曲线，到期时间是“期限”，在左边的 6 列是利率，每天都在变化。右侧的 5 列是天与天之间的利率差分，其中每行已经减去各行的平均差值。**这是中心化矩阵 A，它的行的总和是零**。一个实际世界的应用可能从 252 个营业日开始，而不是只有 5 或 6 天(一年而不是一星期)。

表 1　美国国债收益：6 天与 5 个中心化每日差

	2001 年 美国国债收益						基底点(0.01%)的矩阵 A				
期限	Jan 3	Jan 4	Jan 5	Jan 6	Jan 7	Jan 10	Jan 4	Jan 5	Jan 6	Jan 7	Jan 10
3 MO	5.87	5.69	5.37	5.12	5.19	5.24	−5.4	−19.4	−12.4	19.6	17.6
6 MO	5.58	5.44	5.20	4.98	5.03	5.11	−4.6	−14.6	−12.6	14.4	17.4
1 YR	5.11	5.04	4.82	4.60	4.61	4.71	1.0	−14.0	−14.0	9.0	18.0
2 YR	4.87	4.92	4.77	4.56	4.54	4.64	9.6	−10.4	−16.4	2.6	14.0
3 YR	4.82	4.92	4.78	4.57	4.55	4.65	13.4	−10.6	−17.6	1.4	13.4
5 YR	4.76	4.94	4.82	4.66	4.65	4.73	18.6	−11.4	−15.4	−0.4	8.6
7 YR	4.97	5.18	5.07	4.93	4.94	4.98	20.8	−11.2	−14.2	0.8	3.8
10 YR	4.92	5.14	5.03	4.93	4.94	4.98	20.8	−12.2	−11.2	−0.2	2.8
20 YR	5.46	5.62	5.56	5.50	5.52	5.53	14.6	−7.4	−7.4	0.6	−0.4

MO：月；　YR：年；　Jan：一月

从这 5 列我们可以预期 5 个奇异值，但是这 5 个列向量的总和是零向量(由于 A 的每一行在中心化之后的总和是零行)，所以 $S = AA^{\mathrm{T}}/(5-1)$有 4 个非零特征值 $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \sigma_4^2$。下列是奇异值$\sigma_i$与它们的平方$\sigma_i^2$以及全部方差 $T = \sigma_1^2 + \ldots + \sigma_4^2 = S$ 的迹的比例，由每个主要分量(S 的每个特征向量 $\boldsymbol{u}_i$)来解释：

	σ_i	σ_i^2	σ_i^2/T
主要分量 u_1	36.39	1323.9	.7536
主要分量 u_2	19.93	397.2	.2261
主要分量 u_3	5.85	34.2	.0195
主要分量 u_4	1.19	1.4	.0008
主要分量 u_5	0.00	0.0	.0000
		$T = 1756.7$	1.0000

“陡坡图”画出这些比例σ_i^2/T很快掉落到零，在大问题中你通常会看到在底部的平坦部分之后快速的衰减(接近$\sigma^2 = 0$)。在这两部分之间标示肘部(显著与不显著 PC’s)非常重要。

请尊重版权与译者的劳动成果，侵权必究!

我们的目标在了解每个主要分量，这些 A 的奇异向量 $\boldsymbol{u}_i$ 是 S 的特征向量，这些向量的单元是“负载”。下列是对应范例收益曲线的 $\boldsymbol{u}_1$ 到 $\boldsymbol{u}_5(S\boldsymbol{u}_5 = \mathbf{0})$。

	$\boldsymbol{u}_1$	$\boldsymbol{u}_2$	$\boldsymbol{u}_3$	$\boldsymbol{u}_4$	$\boldsymbol{u}_5$
3 MO	0.383	0.529	−0.478	0.060	0.084
6 MO	0.336	0.436	−0.046	0.210	−0.263
1 YR	0.358	0.263	0.225	−0.491	0.237
2 YR	0.352	−0.028	0.460	0.096	0.242
3 YR	0.371	−0.131	0.430	0.258	−0.555
5 YR	0.349	−0.293	0.117	−0.188	0.446
7 YR	0.323	−0.365	−0.228	0.459	0.081
10 YR	0.297	−0.378	−0.351	−0.579	−0.470
20 YR	0.184	−0.280	−0.361	0.227	0.268

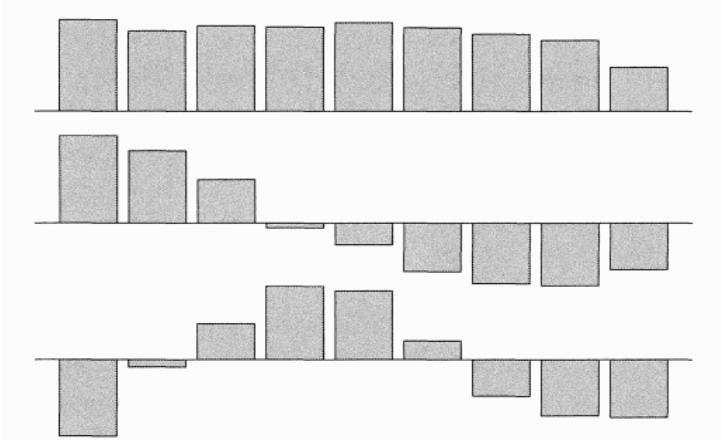
这 5 个 $\boldsymbol{u}$'s 是正交，他们给出 A 的 4 维列空间的一组基底以及 A^{T} 的 1 维零空间的基底。他们有什么样的金融意义？

$\boldsymbol{u}_1$　量测 9 个收益每天变化的权重平均

$\boldsymbol{u}_2$　量测长期与短期债券之间的收益分布的每天变化

$\boldsymbol{u}_3$　显示曲线的每天变化(短期债券对中期)

这些图形展示 $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ 的 9 个负载，从 3 个月到 20 年。



来自典型指令(以 R 撰写)的输出会包含额外两个表—这些我们的书本网站。一个会展示 A 的右奇异向量 $\boldsymbol{v}_i$，这些是 $A^{\mathrm{T}}A$ 的特征向量，他们与向量 $A^{\mathrm{T}}\boldsymbol{u}$ 成正比，他们有 5 个分量，他们显示收益的移动以及本周的短-长分布。

全部方差 $T = 1756.7$(S 的迹 $\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2$)也是 S 对角线单元的总和，这些是 A 的行的样本方差，他们是：$\boldsymbol{s_1^2} + \ldots + \boldsymbol{s_9^2} = 313.3+225.8+199.5+172.3+195.8+196.8+193.7+178.7+80.8=\mathbf{1756.7}$。每个 s^2 小于 σ_1^2，且 1756.7 也是 $A^{\mathrm{T}}A/(n-1)$的迹：列方差。

注意段落 7.3 的 PCA 是处理 A 的中心化行，在某些应用(例如金融)，通常会转置矩阵以及中心化列，然后样本协方差矩阵 S 使用 $A^{\mathrm{T}}A$，以及 $\boldsymbol{v}$'s 是更重要的主要分量。实用说明的线性代数告诉我们这么多。

问题集 7.3

1 假设 A_0 保持 5 个样本的 2 个量测值：

$$A_0 = \begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ -1 & 1 & 0 & 1 & -1 \end{bmatrix}$$

找出每行的平均值然后减去得到中心化矩阵 A，计算样本的协方差矩阵 $S = AA^{\mathrm{T}}/(n-1)$以及它的特征值$\lambda_1$ 与λ_2。哪条通过原点的直线与 A 的列的 5 个样本最接近？

2 针对下列矩阵 A_0，使用问题 1 的步骤：

$$A_0 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 3 & 2 & 1 \end{bmatrix}$$

3 样本方差 $s_1{}^2$ 与 $s_2{}^2$ 以及协方差 s_{12} 是 S 的单元，当 $A_0 = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 2 & 2 \end{bmatrix}$，求出 S (减去平均值之后)，σ_1 为何？

4 从 $S = AA^{\mathrm{T}}$ 的特征向量，求出与三维空间中的这 4 个点最接近的直线(通过中心点的 $\boldsymbol{u}_1$ 方向)以及平面($\boldsymbol{u}_1$，$\boldsymbol{u}_2$ 的方向)：

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix}$$

5 从样本的协方差矩阵 S，找出关联矩阵 DSD，其中有 1's 沿着主对角线。D 是产生这些 1's 的正数对角矩阵。

$$S = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 4 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

6 选择对角矩阵 D 产生 DSD，求出关联 c_{ij}：

$$S = \begin{bmatrix} s_1{}^2 & s_{12} & s_{13} \\ s_{12} & s_2{}^2 & s_{23} \\ s_{13} & s_{23} & s_3{}^2 \end{bmatrix} \qquad DSD = \begin{bmatrix} 1 & c_{12} & c_{13} \\ c_{12} & 1 & c_{23} \\ c_{13} & c_{23} & 1 \end{bmatrix}$$

7 假设 A_0 是一个 5×10 的矩阵，其中包含 5 个课程 10 年的平均成绩，你如何建立中心化矩阵 A 以及样本协方差矩阵 S？当你发现 S 的带头特征向量，它会告诉你什么？

请尊重版权与译者的劳动成果，侵权必究！

7.4 SVD 的几何

1 典型方形矩阵 $A = U\Sigma V^{\mathrm{T}}$ 分解成为(旋转) (拉伸) (旋转)。

2 几何显示 A 如何把圆上的向量 $\boldsymbol{x}$ 变换到椭圆上的 $A\boldsymbol{x}$。

3 A 的**范数**(norm)是$\|A\| = \sigma_1$，这个奇异值是它的最大的**成长因子**$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$。

4 **极分解**把 A 分解成 QS；旋转 $Q = UV^{\mathrm{T}}$ 乘拉伸 $S = V\Sigma V^{\mathrm{T}}$。

5 **伪逆**(pseudoinverse) $A^{+} = V\Sigma^{+}U^{\mathrm{T}}$ 把列空间的 $A\boldsymbol{x}$ 带回行空间的 $\boldsymbol{x}$。

SVD 把一个矩阵分成三个步骤：**(正交)×(对角)×(正交)**。潜藏的几何可以用平常的字词表示：**(旋转)×(拉伸)×(旋转)**。$U\Sigma V^{\mathrm{T}}\boldsymbol{x}$ 从旋转开始得到 $V^{\mathrm{T}}\boldsymbol{x}$，然后Σ拉伸向量来到$\Sigma V^{\mathrm{T}}\boldsymbol{x}$，然后 U 旋转到 $A\boldsymbol{x} = U\Sigma V^{\mathrm{T}}\boldsymbol{x}$。以下是图形。

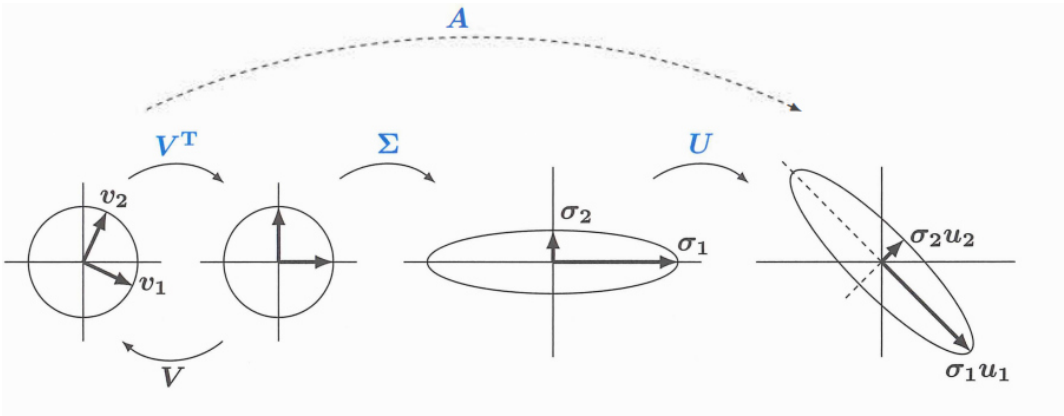


图 7.5：U 与 V 是旋转与可能的反射。Σ把圆拉伸成椭圆。

不可否认，这个图是应用在 2×2 矩阵，而且表示每一个 2×2 矩阵，因为 U 与 V 不允许反射——全部三个矩阵都有行列式 > 0。A 必须是可逆，因为所示的三个步骤都是可逆：

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = U\Sigma V^{\mathrm{T}} \tag{1}$$

矩阵 A 的 4 个数字 a, b, c, d 得到在 SVD 的 4 个数字θ, σ_1, σ_2, ϕ。

图形会引导我们领悟在矩阵代数的三个简洁概念：

1 **矩阵的范数**$\|A\|$ ——它的最大的成长因子。

2 **极分解** $A = Qs$ ——正交 Q 乘正定 S。

3 **伪逆** A^{+} ——当矩阵 A 不可逆时的最佳逆矩阵。

如果让我在图形中选择一个重要数字，它会是σ_1，这个数字是对于任意向量 $\boldsymbol{x}$ 的最大成长因子。如果你追踪在左侧的向量 $\boldsymbol{v}_1$，你会看到它旋转到(1, 0)然后拉伸到(σ_1, 0)，最终旋转到$\sigma_1\boldsymbol{u}_1$，这个叙述 $A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$ 就是 SVD 方程式。这个最大的奇异值σ_1 是矩阵 A 的范数(norm)。

范数$\|A\|$是比值 $\dfrac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ 的最大值　　$$\|A\| = \max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sigma_1 \tag{2}$$

MATLAB 使用 norm(x)求向量长度以及相同的词 norm(A)求矩阵的范数。数学符号有两个竖棒：$\|\boldsymbol{x}\|$与$\|A\|$，此处$\|\boldsymbol{x}\|$表示向量的标准长度$\|\boldsymbol{x}\|^2 = |x_1|^2 + \ldots + |x_n|^2$。当 $\boldsymbol{x} = \boldsymbol{v}_1$ 且 $A\boldsymbol{x} = \sigma_1\boldsymbol{u}_1$ 且$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\| = \sigma_1 =$ 最大比值 $= \|A\|$时，矩阵范数来自向量范数。

数字 norm(A)有两个直接来自定义的重要性质：

三角不等式 $\|A + B\| \le \|A\| + \|B\|$　　**乘积不等式** $\|AB\| \le \|A\|\,\|B\|$　　(3)

定义(2)说明对于每个 $\boldsymbol{x}$，$\|A\boldsymbol{x}\| \le \|A\|\,\|\boldsymbol{x}\|$，这是我们知道的！向量的三角不等式得到矩阵的三角不等式：

对于向量　$\|(A + B)\boldsymbol{x}\| \le \|A\boldsymbol{x}\| + \|B\boldsymbol{x}\| \le \|A\|\,\|\boldsymbol{x}\| + \|B\|\,\|\boldsymbol{x}\|$

除以$\|\boldsymbol{x}\|$，取对于所有 $\boldsymbol{x}$ 的极大值，则$\|A + B\| \le \|A\| + \|B\|$。

乘积不等式快速来自　$\|AB\boldsymbol{x}\| \le \|A\|\,\|B\boldsymbol{x}\| \le \|A\|\,\|B\|\,\|\boldsymbol{x}\|$

再次除以$\|\boldsymbol{x}\|$，取对于所有 $\boldsymbol{x}$ 的极大值，结果是$\|AB\| \le \|A\|\,\|B\|$。

范例 1　最基本可以得到的秩 1 矩阵是 $A = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$，它有一个非零特征值$\lambda_1$ 与一个非零奇异值σ_1。简单的说，特征向量是 $\boldsymbol{u}$ 且它的奇异向量(左与右)是 $\boldsymbol{u}$ 与 $\boldsymbol{v}$。

特征向量　$A\boldsymbol{u} = (\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}) = \lambda_1\boldsymbol{u}$　　所以　$\lambda_1 = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$

奇异向量　$A^{\mathrm{T}}A\boldsymbol{v} = (\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}})\,(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}) = \sigma_1{}^2\boldsymbol{v}$　　所以　$\sigma_1 = \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$

它使得你感觉良好，因为 $|\lambda_1| \le \sigma_1$ 就是**施瓦茨**不等式 $|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}| \le \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$。

们如何知道$|\lambda_1| \le \sigma_1$？$A\boldsymbol{x} = \lambda_1\boldsymbol{x}$ 的特征向量会得到比值$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\| = \|\lambda_1\boldsymbol{x}\|/\|\boldsymbol{x}\|$ 就是$|\lambda_1|$。最大的比值σ_1 不可能比$|\lambda_1|$小。

$|\lambda_2| \le \sigma_2$ 也是正确吗？**不是！**这是完全错误。事实上一个 2×2 矩阵会有$|\det A| = |\lambda_1\lambda_2| = \sigma_1\sigma_2$。在$|\lambda_1| \le \sigma_1$ 的情况下，必然强迫$|\lambda_2| \ge \sigma_2$。

请尊重版权与译者的劳动成果，侵权必究！

最接近 A 的秩 k 矩阵是 $A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \ldots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}$。

这是矩阵近似的关键事实：Eckart-Young-Mirsky 定理说明

$$\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}，\text{对所有的秩 } k \text{ 矩阵 } B$$

对我来说这个完成了线性代数的基础定理，$\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 给出了 4 个子空间的正交单位基底，前 k 个 $\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 与σ's 给出了近似 A 的最佳矩阵。

极分解 $A = QS$

每个复数 $x + iy$ 有极形式 $re^{i\theta}$，一个数字 $r \geq 0$ 乘一个在单位圆的数字$e^{i\theta}$，我们有 $x + iy = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta) = re^{i\theta}$。把这些数字当成 1×1 矩阵，则$e^{i\theta}$是一个正交矩阵 Q 以及 $r \geq 0$ 是一个正半定矩阵(称为 S)。极分解(polar decomposition)延伸这个概念到 $n \times n$ 矩阵：正交乘正半定，$\boldsymbol{A = QS}$。

> 每个实数方形矩阵可以分解成 $\boldsymbol{A = QS}$，其中 $\boldsymbol{Q}$ **是正交**且 $\boldsymbol{S}$ **是对称正半定**。若 A 是可逆，则 S 是正定。

为了证明我们在 SVD 的中间插入 $V^{\mathrm{T}}V = I$：

$$\text{极分解} \qquad A = U\Sigma V^{\mathrm{T}} = (UV^{\mathrm{T}})(V\Sigma V^{\mathrm{T}}) = (Q)(S) \tag{4}$$

第一个因子 UV^{T} 是 Q，正交矩阵的乘积是正交。第二个因子是 $V\Sigma V^{\mathrm{T}}$ 是 S，它是正半定，因为它的特征值在Σ中。

若 A 可逆，则Σ与 S 也是可逆，S 是 $A^{\mathrm{T}}A$ 的对称正定平方根，这是因为 $S^2 = V\Sigma^2 V^{\mathrm{T}} = A^{\mathrm{T}}A$，所以 S 的特征值是 A 的奇异值，S 的特征向量是 A 的奇异向量。

还有一种相反顺序的极分解 $A = KQ$，Q 相同，但是现在 $K = U\Sigma U^{\mathrm{T}}$，则 K 是 AA^{T} 的对称正定平方根。

范例 2 段落 7.2 的 SVD 范例是 $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = U\Sigma V^{\mathrm{T}}$。求出极分解 $A = QS$ 的因子 Q 与 S(旋转与拉伸)。

解 我只是从段落 7,2 复制 U 与Σ与 V：

$$Q = UV^{\mathrm{T}} = \frac{1}{\sqrt{20}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{20}}\begin{bmatrix} 4 & -2 \\ 2 & 4 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

$$S = V\Sigma V^{\mathrm{T}} = \frac{\sqrt{5}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & \\ & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \sqrt{5}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{则 } \boldsymbol{A = QS}$$

在力学中，极分解把旋转(在 Q)与拉伸(在 S)分开，S 的特征值给出拉伸因子如图 7.5 所示，S 的特征向量给出拉伸方向(椭圆的主轴)。正交矩阵 Q 包含两个旋转 U 与 V^{T}。

此处有个关于旋转的事实。$Q = UV^{\mathrm{T}}$ 是 A 的**最近(nearest)正交矩阵**，Q 使得范数$\|Q - A\|$尽可能最小，对应的事实是$e^{i\theta}$是在单位圆上与 $re^{i\theta}$ 最接近的数字。

SVD 告诉我们一个有关最近奇异矩阵更为重要的事实：

A 的最近奇异矩阵 A_0 等于把最小的 $\sigma_{\min}$ 变成零

所以 $\sigma_{\min}$ 在于量测从 A 到奇异的距离，等于范例 2 的矩阵，这个距离 $\sigma_{\min}=\sqrt{5}$。如果我把 $\sigma_{\min}$ 改成零，这个排除了在 $A=\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}+\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$ 中最后(最小)一个片段，这个最小的改变有范数 $\sigma_2=\sqrt{5}$ (小于 3)。

在计算实践时我们通常会排除一个非常小的 σ，处理奇异矩阵要比与零太接近而未注意好得多。

伪逆 A^+

为了选择优质基底，A 乘行空间中的 $\boldsymbol{v}_i$ 得到列空间中的 $\sigma_i\boldsymbol{u}_i$，A^{-1} 必须做相反的事情！若 $A\boldsymbol{v}=\sigma\boldsymbol{u}$ 则 $A^{-1}\boldsymbol{u}=\boldsymbol{v}/\sigma$，$A^{-1}$ 的奇异值是 $1/\sigma$，就如同 A^{-1} 的特征值是 $1/\lambda$。基底逆反了，$\boldsymbol{u}$'s 在 A^{-1} 的行空间，$\boldsymbol{v}$'s 在列空间。

直到现在我们还没有加上“若 A^{-1} 存在”，现在还不要！乘 $\boldsymbol{u}_i$ 得到 $\boldsymbol{v}_i/\sigma_i$ 的矩阵确实存在，它是伪逆矩阵 A^+：

A 的伪逆： $\boldsymbol{A^+=V\Sigma^+U^{\mathrm{T}}}=[\,\boldsymbol{v}_1\ \dots\ \boldsymbol{v}_r\ \dots\ \boldsymbol{v}_n\,]\begin{bmatrix}\sigma_1^{-1}&&\\&\ddots&\\&&\sigma_r^{-1}\end{bmatrix}[\,\boldsymbol{u}_1\ \dots\ \boldsymbol{u}_r\ \dots\ \boldsymbol{u}_m\,]^{\mathrm{T}}$

$n\times n$　　$n\times m$　　$m\times m$

伪逆 $\boldsymbol{A^+}$是一个 $n\times m$ 的矩阵，若 A^{-1} 存在(再说一次)，则 A^+ 与 A^{-1} 相同。在这个情况下 $m=n=r$，我们是在逆反 $U\Sigma V^{\mathrm{T}}$ 得到 $V\Sigma^{-1}U^{\mathrm{T}}$。新符号 A^+ 在 $r<m$ 或 $r<n$ 才需要，则 A 没有双边逆矩阵，但是它有一个伪逆矩阵 A^+，秩也一样是 r：

$$A^+\boldsymbol{u}_i=\boldsymbol{v}_i/\sigma_i,\ \text{其中 } i\le r\quad \text{且}\quad A^+\boldsymbol{u}_i=\boldsymbol{0},\ \text{其中 } i>r$$

在 A 的列空间中的向量 $\boldsymbol{u}_1,\dots,\boldsymbol{u}_r$ 回到行空间中的 $\boldsymbol{v}_1,\dots,\boldsymbol{v}_r$。其他向量 $\boldsymbol{u}_{r+1},\dots,\boldsymbol{u}_m$ 在左零空间中，而且 A^+ 把他们变成零。当我们知道这些基底发生什么事情，我们就了解 A^+。

注意对角矩阵Σ的伪逆，Σ的每个σ 被Σ^+的σ^{-1} 取代，乘积$\Sigma^+\Sigma$是尽可能的接近单位矩阵，它是一个投影矩阵，$\Sigma^+\Sigma$是一定程度上的 I 而且其他是零。我们可以逆反σ's，但是我们不会对零行与零列做任何事情。这个例子有 $\sigma_1=2$ 与 $\sigma_2=3$：

$$\Sigma^+\Sigma=\begin{bmatrix}\boldsymbol{1/2}&0&0\\0&\boldsymbol{1/3}&0\\0&0&\boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{2}&0&0\\0&\boldsymbol{3}&0\\0&0&\boldsymbol{0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{1}&0&0\\0&\boldsymbol{1}&0\\0&0&\boldsymbol{0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}&0\\0&\boldsymbol{0}\end{bmatrix}$$

伪逆 A^+ 是 $n\times n$ 矩阵使得 AA^+ 与 A^+A 成为投影。

请尊重版权与译者的劳动成果，侵权必究！

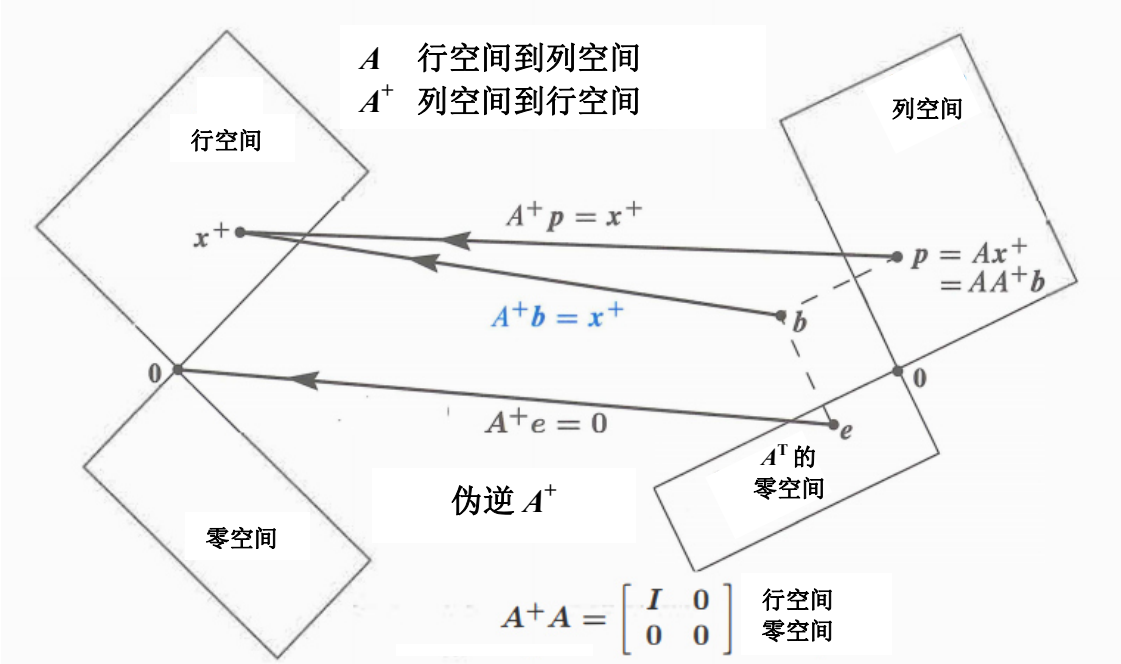


图 7.6：列空间中的 $A\boldsymbol{x}^+$ 回到行空间中的 $A^+A\boldsymbol{x}^+ = \boldsymbol{x}^+$。

尝试 $AA^{-1} = A^{-1}A = I$	AA^+ = 在 A 的列空间的投影矩阵 A^+A = 在 A 的行空间的投影矩阵

范例 3 每个秩 1 矩阵是一个列乘一个行，在 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 是单位向量的情况下就是 $A = \sigma\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$，它的伪逆是 $A^+ = \boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}/\sigma$，乘积 $AA^+ = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$，就是在通过 $\boldsymbol{u}$ 的直线的投影。乘积 $A^+A = \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$。

范例 4 求出 $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 的伪逆矩阵，这个矩阵不可逆，秩是 1，唯一的奇异值是 $\sigma_1 = 2$，在 Σ^+(也是秩 1)中变成 1/2。

$$A^+ = V\Sigma^+U^{\mathrm{T}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{\mathbf{1}}{\mathbf{4}}\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}$$

A^+也是秩 1，它的列空间永远是 A 的行空间。

相关列的最小平方

段落 4.3 关于最小平方有一个有 4 个 1's 的矩阵 A，它打破了无关列的要求。当我们同时在 $t = 1$ 制作两个量测值是，这个矩阵出现了。最接近的直线是沿着量测值 3 与 1 之间的直线，但是没有办法决定最佳线的斜率。

在矩阵语言来说，$A^{\mathrm{T}}A$ 是奇异，方程式 $A^{\mathrm{T}}A\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$ 有无限多解，伪逆矩阵给出一个方法去选择一个“最佳解” $\boldsymbol{x}^+ = A^+\boldsymbol{b}$。

让我重复无解的 $A\boldsymbol{x}=\boldsymbol{b}$ 于 1 无限多解的 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$：

$$A\boldsymbol{x}=\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}=\boldsymbol{b}\qquad A^{\mathrm{T}}A\hat{\boldsymbol{x}}=\begin{bmatrix}2&2\\2&2\end{bmatrix}\begin{bmatrix}\hat{x}_1\\\hat{x}_2\end{bmatrix}=\begin{bmatrix}4\\4\end{bmatrix}=A^{\mathrm{T}}\boldsymbol{b}$$

任意向量 $\hat{\boldsymbol{x}}=(1+c,1-c)$会是这些正态方程式 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 的解。伪逆矩阵的目的是选择一个解 $\hat{\boldsymbol{x}}=\boldsymbol{x}^{+}$。

$\boldsymbol{x}^{+}=A^{+}\boldsymbol{b}=(\mathbf{1},\mathbf{1})$ 是 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 与 $A\hat{\boldsymbol{x}}=\boldsymbol{p}$ 的**最短(shortest)解**。

你可以看到 $\boldsymbol{x}^{+}=(1,1)$比其他解 $\hat{\boldsymbol{x}}=(1+c,1-c)$都要短，$\hat{\boldsymbol{x}}$ 的长度平方是$(1+c)^2+(1-c)^2=2+2c^2$，最短的选择是 $c=0$，得到解 $\boldsymbol{x}^{+}=(1,1)$在 A 的行空间中。

几何告诉我们 A^{+}应该做的事情：把 A 的列空间带回到行空间，两个空间都有维度 r，把误差向量 $\boldsymbol{e}$ 扼杀在 A 的行空间。

伪逆 A^{+}与最佳解 $\boldsymbol{x}^{+}$在统计学非常重要，因为实验通常有相关列的矩阵，也会有相关行。

主要观念的复习

1. 向量 $A\boldsymbol{x}$ 的椭圆有轴沿着奇异向量 $\boldsymbol{u}_i$。
2. 矩阵范数$\|A\|=\sigma_1$ 来自向量长度：极大化 $\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$。
3. 可逆矩阵 $=$(正交矩阵) (正定矩阵)：$A=QS$。
4. 每个 $A=U\Sigma V^{\mathrm{T}}$ 有一个伪逆 $A^{+}=V\Sigma^{+}U^{\mathrm{T}}$，会把 $\boldsymbol{N}(A^{\mathrm{T}})$送到 $\mathbf{Z}$。

已解范例

7.4A 若 A 有秩 n (满列秩)，则它有一个**左逆矩阵** $L=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$，矩阵 L 给出 $LA=I$，说明为什么这个情况下伪逆是 $\boldsymbol{A^{+}=L}$。

若 A 有秩 m (满行秩)，则它有一个**右逆矩阵** $R=A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$，矩阵 R 给出 $AR=I$，说明为什么这个情况下伪逆是 $\boldsymbol{A^{+}=R}$。

求出对应 A_1 的 L 与对应 A_2 的 R。求出所有 3 个矩阵 A_1,A_2,A_3 的 $\boldsymbol{A^{+}}$：

$$A_1=\begin{bmatrix}\mathbf{2}\\\mathbf{2}\end{bmatrix}\qquad A_2=\begin{bmatrix}\mathbf{2}&\mathbf{2}\end{bmatrix}\qquad A_3=\begin{bmatrix}\mathbf{2}&\mathbf{2}\\\mathbf{1}&\mathbf{1}\end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

解 若 A 有无关列，则 $A^{\mathrm{T}}A$ 可逆--这是段落 4.2 的重点。当然 $L=(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 乘 A 得到 $LA=I$：一个左逆矩阵。

$AL=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 是在列空间的投影矩阵(段落 4.2)，所以 L 满足 A^{+}的要求：LA 与 AL 是在 $\boldsymbol{C}(A)$与 $\boldsymbol{C}(A^{\mathrm{T}})$的投影。

若 A 有秩 m (满行秩)，则 AA^{T} 可逆，当然 A 乘 $R=A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$ 得到 $AR=I$。相反的顺序之下，$RA=A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A$ 是在行空间(A^{T} 的列空间)的投影矩阵，所以 R 等于伪逆矩阵 A^{+}。

例子 A_1 有满列秩(对应 L)以及 A_2 有满行秩(对应 R)：

$$A_1^{+}=(A_1^{\mathrm{T}}A_1)^{-1}A_1^{\mathrm{T}}=\frac{1}{\sqrt{8}}\begin{bmatrix}2 & 2\end{bmatrix}\qquad A_2^{+}=A_2^{\mathrm{T}}(A_2A_2^{\mathrm{T}})^{-1}=\frac{1}{\sqrt{8}}\begin{bmatrix}2\\2\end{bmatrix}$$

注意 $A_1^{+}A_1=[1]$以及 $A_2A_2^{+}=[1]$，但是 A_3 没有左或右逆矩阵，**它不是满秩，它的伪逆把 $\boldsymbol{A_3}$ 的列空间带回到行空间：**

$$A_3^{+}=\begin{bmatrix}2 & 2\\1 & 1\end{bmatrix}^{+}=\frac{\boldsymbol{v}_1\boldsymbol{u}_1^{\mathrm{T}}}{\sigma_1}=\frac{1}{10}\begin{bmatrix}\mathbf{2} & \mathbf{1}\\\mathbf{2} & \mathbf{1}\end{bmatrix}$$

问题集 7.4

问题 1-4 计算并使用特定矩阵(不是可逆)的 SVD。

1 (a) 计算 $A^{\mathrm{T}}A$ 以及特征值与单位特征向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$。求出σ_1。

秩 1 矩阵 $\qquad A=\begin{bmatrix}1 & 2\\3 & 6\end{bmatrix}$

(b) 计算 AA^{T} 以及特征值与单位特征向量 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$。

(c) 验证 $A\boldsymbol{v}_1=\sigma_1\boldsymbol{u}_1$。把数字放进 $\boldsymbol{A=U\Sigma V^{\mathrm{T}}}$ (这是 SVD)。

2 (a) 从问题 1 的 $\boldsymbol{u}$'s 与 $\boldsymbol{v}$'s 写出矩阵 A 的 4 个基础子空间的正交单位基底。

(b) 描述拥有相同 4 个子空间的所有矩，是 A 的倍数吗？

3 从问题 1 的 U, V 与Σ，求出正交矩阵 $Q=UV^{\mathrm{T}}$ 与对称矩阵 $S=V\Sigma V^{\mathrm{T}}$。验证极分解 $A=QS$，S 只有正半定，这是因为_____。测试 $S^2=A$。

4 计算伪逆 $A^{+}=V\Sigma^{+}U^{\mathrm{T}}$，对角矩阵$\Sigma^{+}$包含 $1/\sigma_1$。重新命名图 7.6 的 4 个(A 的)子空间成为 A^{+}的 4 个子空间。计算在 A 的行空间与列空间的投影 $A^{+}A$ 与 AA^{+}。

问题 5-9 有关可逆矩阵的 SVD。

5　计算 $A^{\mathrm{T}}A$ 以及特征值与单位特征向量 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$。求出矩阵 A 的奇异值σ_1 与σ_2：

$$A=\begin{bmatrix}3 & 3\\ -1 & 1\end{bmatrix}$$

6　AA^{T} 与 $A^{\mathrm{T}}A$ 有相同的特征值σ_1^2 与σ_2^2，求出 AA^{T} 的单位特征向量 $\boldsymbol{u}_1$ 与 $\boldsymbol{u}_2$。把数字放进 SVD：

$$A=\begin{bmatrix}3 & 3\\ -1 & 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2\end{bmatrix}\begin{bmatrix}\sigma_1 & \\ & \sigma_2\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}^{\mathrm{T}}$$

7　在问题 6，计算列乘行展示 $A = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$。由 $A = U\Sigma V^{\mathrm{T}}$ 证明每个秩 r 的矩阵是 r 个秩 1 矩阵的总和。

8　从 U, V 与Σ，求出正交矩阵 $Q = UV^{\mathrm{T}}$ 与对称矩阵 $K = U\Sigma U^{\mathrm{T}}$。验证相反顺序的极分解 $A = KQ$。

9　A 的伪逆 A^{+}与_____相同，这是因为_____。

问题 10-11 计算并使用 1×3 矩形矩阵的 SVD。

10　当矩阵 $A = [3 \quad 4 \quad 0]$，计算 $A^{\mathrm{T}}A$ 与 AA^{T} 以及他们的特征值与单位特征向量。A 的奇异值为何？

11　把数字放进 A 的奇异值分解：

$$A = [3 \quad 4 \quad 0] = [\boldsymbol{u}_1]\,[\sigma_1 \quad 0 \quad 0\,][\,\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3\,]^{\mathrm{T}}$$

把数字放进 A 的伪逆 $V\Sigma^{+}U^{\mathrm{T}}$，计算 AA^{+}与 $A^{+}A$：

$$\textbf{伪逆}\qquad A^{+}=\begin{bmatrix} & \\ & \\ & \end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3\\ & & \\ & & \end{bmatrix}\begin{bmatrix}1/\sigma_1 & \\ 0 & \\ 0 & \end{bmatrix}\begin{bmatrix}\boldsymbol{u}_1\end{bmatrix}^{\mathrm{T}}$$

12　唯一没有枢轴且没有奇异值的 2×3 矩阵为何？这个矩阵的Σ为何？A^{+}是零矩阵，但是它的形状为何？

13　若 $\det A = 0$，为什么 $\det A^{+} = 0$？若 A 有秩 r，为什么 A^{+}有秩 r？

14　在单位圆$\|\boldsymbol{x}\| = 1$ 的向量，在椭圆的向量 $\boldsymbol{y} = A\boldsymbol{x}$ 会有$\|A^{-1}\boldsymbol{y}\| = 1$，这个椭圆有沿着奇异向量的轴，其中长度是$\sigma_1, \ldots, \sigma_r$ (如图 7.5)。对于 $A = [2 \quad 1; 1 \quad 2]$，扩展$\|A^{-1}\boldsymbol{y}\|^2 = 1$。

请尊重版权与译者的劳动成果，侵权必究！

问题 15-18 带出 A^+的重要性质以及 $x^+ = A^+b$。

15 本题的所有矩阵都有秩 1，向量 $\boldsymbol{b}$ 是(b_1, b_2)。

$$A=\begin{bmatrix}2&2\\1&1\end{bmatrix}\quad AA^{\mathrm{T}}=\begin{bmatrix}8&4\\4&2\end{bmatrix}\quad A^{\mathrm{T}}A=\begin{bmatrix}5&5\\5&5\end{bmatrix}\quad A^{+}=\begin{bmatrix}.2&.2\\.1&.1\end{bmatrix}$$

(a) 方程式 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 有很多解，这是因为_____。

(b) 验证 $x^+ = A^+b = (.2b_1 + .1b_2，\ .2b_1 + .1b_2)$是 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 的解。

(c) 把 x^+加上$(1, -1)$得到另一个 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ 的解，证明$\|\hat{\boldsymbol{x}}\|^2 = \|x^+\|^2 + 2$ 并且 x^+比较短。

16 向量 $\boldsymbol{x^+ = A^+b}$ 是 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}}=A^{\mathrm{T}}\boldsymbol{b}$ **最短的可能解**。理由：$\hat{\boldsymbol{x}} - \boldsymbol{x}^+$的差在 $A^{\mathrm{T}}A$ 的零空间中，这也是 A 的零空间，与 $\boldsymbol{x}^+$正交。说明如何推论$\|\hat{\boldsymbol{x}}\|^2 = \|x^+\|^2 + \|\hat{\boldsymbol{x}} - \boldsymbol{x}^+\|^2$。

17 $\mathbf{R}^m$ 中的每个 $\boldsymbol{b}$ 是 $\boldsymbol{p} + \boldsymbol{e}$，这是列空间的部分加上左零空间的部分。$\mathbf{R}^n$ 中的每个 $\boldsymbol{x}$ 是 $\boldsymbol{x}^+ + \boldsymbol{x}_n$，这是行空间的部分加上零空间的部分。则

$$AA^+\boldsymbol{p} = _____\qquad AA^+\boldsymbol{e} = _____\qquad A^+A\boldsymbol{x}^+ = _____\qquad A^+A\boldsymbol{x}_n = _____$$

18 对于矩阵 $A = U\Sigma V^{\mathrm{T}}$ 与下列的 $\boldsymbol{b}$，求出 A^+与 A^+A 与 AA^+与 $\boldsymbol{x}^+$：

$$A=\begin{bmatrix}3\\4\end{bmatrix}=\begin{bmatrix}.6&-.8\\.8&.6\end{bmatrix}\begin{bmatrix}5\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}\quad \boldsymbol{b}=\begin{bmatrix}3\\4\end{bmatrix}\text{与}\boldsymbol{b}=\begin{bmatrix}-4\\3\end{bmatrix}$$

19 一个普通的 2×2 矩阵 A 由 4 个数字决定。如果是三角形，由 3 个数字决定。如果是对角，由 2 个决定。如果是旋转，由 1 个决定。如果是单位特征向量，也是由 1 个决定。检验 A 的每个分解，所有的计数都是 4 个：

4 个数字在　　LU　　LDU　　QR　　$U\Sigma V^{\mathrm{T}}$　　$X\Lambda X^{-1}$

20 追踪问题 18，检验 LDL^{T} 与 $Q\Lambda Q^{\mathrm{T}}$ 是由 3 个数字决定。这是正确的，因为矩阵现在是_____。

21 由 A 与 A^+，证明 A^+A 是正确并且$(A^+A)^2 = A^+A =$ 投影。

$$A=\sum_1^r \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}\qquad A^+=\sum_1^r\frac{\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}}}{\sigma_i}\qquad A^+A=\sum_1^r\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\qquad AA^+=\sum_1^r\boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}$$

22 奇异向量 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的每个配对都有 $A\boldsymbol{v} = \sigma\boldsymbol{u}$ 与 $A^{\mathrm{T}}\boldsymbol{u} = \sigma\boldsymbol{v}$。证明双重向量$\begin{bmatrix}\boldsymbol{v}\\\boldsymbol{u}\end{bmatrix}$是对称方块矩阵 $M=\begin{bmatrix}0&A^{\mathrm{T}}\\A&0\end{bmatrix}$的特征向量。$A$ 的 SVD 等效于这个对称矩阵 M 的对角化。

第八章
线性变换

8.1 线性变换的概念

1 线性变换 T 使得每个向量 $\boldsymbol{v}$ 变成 $T(\boldsymbol{v})$，线性需要 $T(c\boldsymbol{v}+d\boldsymbol{w})=cT(\boldsymbol{v})+dT(\boldsymbol{w})$。注意 $T(\boldsymbol{0})=\boldsymbol{0}$，所以 $T(\boldsymbol{v})=\boldsymbol{v}+\boldsymbol{u}_0$ 不是线性。

2 输入向量 $\boldsymbol{v}$ 与输出 $T(\boldsymbol{v})$可以在 $\mathbf{R}^n$ 或是矩阵空间或是函数空间。

3 若 A 是 $m\times n$，$T(\boldsymbol{x})=A\boldsymbol{x}$ 是从输入空间 $\mathbf{R}^n$ 到输出空间的 $\mathbf{R}^m$ 的线性。

4 导数 $T(f)=df/dx$ 是线性，积分 $T^{+}(f)=\int_0^x f(t)dt$ 是它的“伪逆”。

5 两个线性变换的乘积 ST 仍然是线性：$(ST)(\boldsymbol{v})=S(T(\boldsymbol{v}))$。

当一个矩阵 A 乘一个向量 $\boldsymbol{v}$，矩阵把 $\boldsymbol{v}$ 变换成另一个向量 $A\boldsymbol{v}$，输入 $\boldsymbol{v}$ 输出 $T(\boldsymbol{v})=A\boldsymbol{v}$。变换 T 遵循函数的概念，输入一个数字 x，输出 $f(x)$。对于一个向量 $\boldsymbol{v}$ 或是一个数字 x，我们左乘矩阵或是计算函数。更深的目标是同时看到所有的向量 $\boldsymbol{v}$，当我们把 A 乘上每个 $\boldsymbol{v}$，我们在变换整个空间 $\mathbf{V}$。

再次从矩阵 A 开始，它把 $\boldsymbol{v}$ 变换成 $A\boldsymbol{v}$，把 $\boldsymbol{w}$ 变换成 $A\boldsymbol{w}$，我们知道对于 $\boldsymbol{u}=\boldsymbol{v}+\boldsymbol{w}$ 会发生什么，毫无疑问是 $A\boldsymbol{u}$，它必须等于 $A\boldsymbol{v}+A\boldsymbol{w}$。矩阵乘法 $T(\boldsymbol{v})=A\boldsymbol{v}$ 得到线性变换：

一个变换 T 指定一个输出 $T(\boldsymbol{v})$给 $\mathbf{V}$ 中的每个向量 $\boldsymbol{v}$。如果它满足下列对于所有 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的要求，则这个变换是**线性**：

(a) $T(\boldsymbol{v}+\boldsymbol{w})=T(\boldsymbol{v})+T(\boldsymbol{w})$。 (b) $T(c\boldsymbol{v})=cT(\boldsymbol{v})$，对所有的 c

请尊重版权与译者的劳动成果，侵权必究！

如果输入 $\boldsymbol{v}=\mathbf{0}$，输出必须是 $T(\boldsymbol{v})=\mathbf{0}$。我们结合(a)与(b)得到一项要求：

线性变换　　$T(c\boldsymbol{v}+d\boldsymbol{w})$　必须等于　$cT(\boldsymbol{v})+dT(\boldsymbol{w})$

我再次测试矩阵乘法的线性：$A(c\boldsymbol{v}+d\boldsymbol{w})=cA\boldsymbol{v}+dA\boldsymbol{w}$ 成立。

线性变换非常严格，假设 T 把 $\boldsymbol{u}_0$ 加到每个向量，则 $T(\boldsymbol{v})=\boldsymbol{v}+\boldsymbol{u}_0$ 且 $T(\boldsymbol{w})=\boldsymbol{w}+\boldsymbol{u}_0$。这个不好，或至少它不是线性。把 T 应用在 $\boldsymbol{v}+\boldsymbol{w}$ 得到 $\boldsymbol{v}+\boldsymbol{w}+\boldsymbol{u}_0$，这个与 $T(\boldsymbol{v})+T(\boldsymbol{w})$不相同：

平移不是线性　　$\boldsymbol{v}+\boldsymbol{w}+\boldsymbol{u}_0$ 不是 $T(\boldsymbol{v})+T(\boldsymbol{w})=(\boldsymbol{v}+\boldsymbol{u}_0)+(\boldsymbol{w}+\boldsymbol{u}_0)$

当 $\boldsymbol{u}_0=\mathbf{0}$ 会是例外，变换简化成 $T(\boldsymbol{v})=\boldsymbol{v}$，这是**单位变换**(没有改变，好像左乘单位矩阵)，这个当然是线性。在此情形下，输入空间 $\mathbf{V}$ 与输出空间 $\mathbf{W}$ 相同。

线性加平移变换 $T(\boldsymbol{v})=A\boldsymbol{v}+\boldsymbol{u}_0$ 称为“仿射(affine)”，直线维持直线，虽然 T 不是线性。段落 10.6 介绍电脑利用仿射变换处理图形。

范例 1　选择一个固定向量 $\boldsymbol{a}=(1,3,4)$，令 $T(\boldsymbol{v})$是点积 $\boldsymbol{a}\cdot\boldsymbol{v}$：

输入是　$\boldsymbol{v}=(v_1,v_2,v_3)$　　输出是　$T(\boldsymbol{v})=\boldsymbol{a}\cdot\boldsymbol{v}=v_1+3v_2+4v_3$

点积是线性，输入 $\boldsymbol{v}$ 来自三维空间，所以 $\mathbf{V}=\mathbf{R}^3$，输出只是个数字，所以输出空间是 $\mathbf{W}=\mathbf{R}^1$。我们是左乘行矩阵 $A=[1\quad 3\quad 4]$，则 $T(\boldsymbol{v})=A\boldsymbol{v}$。

你将会精通辨识哪些变换是线性，如果变换包含平方或乘积或长度，v_1^2 或 v_1v_2 或$\|\boldsymbol{v}\|$，则 T 不是线性。

范例 2　长度 $T(\boldsymbol{v})=\|\boldsymbol{v}\|$不是线性，线性的要求(a)是$\|\boldsymbol{v}+\boldsymbol{w}\|=\|\boldsymbol{v}\|+\|\boldsymbol{w}\|$，要求(b)是$\|c\boldsymbol{v}\|=c\|\boldsymbol{v}\|$，两者都失败。

不满足(a)：　三角形的边长满足不等式 $\|\boldsymbol{v}+\boldsymbol{w}\|<\|\boldsymbol{v}\|+\|\boldsymbol{w}\|$。

不满足(b)：　长度$\|-\boldsymbol{v}\|=\|\boldsymbol{v}\|$不是$-\|\boldsymbol{v}\|$，对于负数 c，失败。

范例 3　(旋转) T 是把每个向量旋转 30°的变换，T 的“定义域(domain)”是 xy 平面(所有的输入向量 $\boldsymbol{v}$)，T 的“值域(range)”也是 xy 平面(所有旋转后的向量 $T(\boldsymbol{v})$)。我们没有用矩阵来描述 T：把平面旋转 30°。

旋转是线性吗？是的。我们可以旋转两个向量，以及把结果相加。旋转的总和 $T(\boldsymbol{v})+T(\boldsymbol{w})$与总和的旋转 $T(\boldsymbol{v}+\boldsymbol{w})$相同。整个平面在这个线性变换中一起转动。

直线到直线，三角到三角，基底说明一切

图 8.1 显示输入空间中从 $\boldsymbol{v}$ 到 $\boldsymbol{w}$ 的直线，它也显示了输出空间中从 $T(\boldsymbol{v})$到 $T(\boldsymbol{w})$的直线。线性告诉我们：输入直线的每个点映射到输出线，更多的是：等空间点映射到等空间点。中点 $\boldsymbol{u} = \frac{1}{2}\boldsymbol{v} + \frac{1}{2}\boldsymbol{w}$ 映射到中点 $T(\boldsymbol{u}) = \frac{1}{2}T(\boldsymbol{v}) + \frac{1}{2}T(\boldsymbol{w})$。

第二个图形提升了一个维度，现在我们有三个角点 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$，这些输入有三个输出 $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), T(\boldsymbol{v}_3)$。输入三角形映射到输出三角形，等空间点仍然是等空间点(沿着边，在两个边之间)，中点 $\boldsymbol{u} = \frac{1}{3}(\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3)$映射到中点 $T(\boldsymbol{u}) = \frac{1}{3}(T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2) + T(\boldsymbol{v}_3))$。

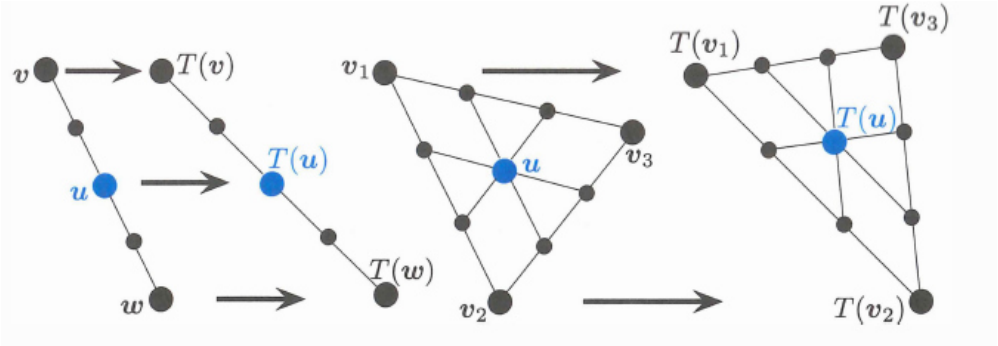


图 8.1：直线到直线，等空间到等空间，$\boldsymbol{u} = \boldsymbol{0}$ 到 $T(\boldsymbol{u}) = \boldsymbol{0}$。

线性的规则延伸到三个向量或是 n 个向量：

线性 $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n$ **必须变换到**

$$T(\boldsymbol{u}) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \ldots + c_nT(\boldsymbol{v}_n) \tag{1}$$

两个向量的规则开始三个向量的证明：$T(c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}) = T(c\boldsymbol{u}) + T(d\boldsymbol{v} + e\boldsymbol{w})$，同时对两侧使用线性规则，得到 $cT(\boldsymbol{u}) + dT(\boldsymbol{v}) + eT(\boldsymbol{w})$。

n-向量规则(1)得到线性变换最重要的事实：

假设你知道一组基底的向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 的变换 $T(\boldsymbol{v})$
则你会知道在空间中的每个 $\boldsymbol{u}$ 的 $T(\boldsymbol{u})$。

你看到理由：空间中的每个 $\boldsymbol{u}$ 是基底向量 $\boldsymbol{v}_j$ 的一个线性组合，线性告诉我们 $T(\boldsymbol{u})$ 是输出 $T(\boldsymbol{v}_j)$的相同组合。

请尊重版权与译者的劳动成果，侵权必究!

范例 4　变换 T 取得输入的导数：$T(u) = du/dx$。

你如何求得 $u = 6 - 4x + 3x^2$ 的导数？你从 $1, x$ 与 x^2 的导数开始，这些是基底向量。他们的导数是 $0, 1, 2x$，然后你使用线性求出这些导数的任意组合：

$$\frac{du}{dx} = 6(1\text{ 的导数}) - 4(x\text{ 的导数}) + 3(x^2\text{ 的导数}) = -4 + 6x$$

所有的微积分都依靠线性！初级微积分有一些对于 x^n 与 $\sin x$ 与 $\cos x$ 与 e^x 的关键导数，然后应用线性到他们所有的组合。

我会说微积分唯一比较特殊的规则是链条规则(chain rule)，产生链条函数 $f(g(x))$的导数。

$T(u) = du/dx$ 的**零空间**。我们求解 $T(u) = 0$ 得到零空间，当 u 是常数时它的导数是零，所以一维零空间是函数空间中的一条直线--这个特殊函数 $u = 1$ 的所有倍数。

$T(u) = du/dx$ 的**列空间**。我们的例子中，输入空间包含所有的二次式 $a + bx + cx^2$，输出(列空间)是所有的线性函数 $b + 2cx$。注意：“**计数定理**”仍然成立：$\boldsymbol{r + (n - r) = n}$。

$$\text{维度(列空间)} + \text{维度(零空间)} = \mathbf{2 + 1 = 3} = \text{维度(输入空间)}$$

d/dx 的矩阵为何？我不能离开导数而不提矩阵，我们有一个线性变换 $T = d/dx$，我们知道 T 对基底函数做了什么：

$$v_1, v_2, v_3 = 1, x, x^2 \qquad \frac{dv_1}{dx} = 0 \qquad \frac{dv_2}{dx} = 1 = v_1 \qquad \frac{dv_2}{dx} = 2x = 2v_2 \tag{2}$$

三维输入空间 **V**(=二次式)变换到二维输出空间 **W**(=线性函数)，如果 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是向量，我会知道这个矩阵。

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \text{导数}\left(T = \frac{d}{dx} \right)\text{的矩阵形式} \tag{3}$$

线性变换 du/dx 完美复制矩阵乘法 $A\boldsymbol{u}$。

$$\textbf{输入 } \boldsymbol{u}\text{：} a + bx + cx^2 \qquad \textbf{乘法 } A\boldsymbol{u} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ 2c \end{bmatrix} \quad \text{输出} \frac{du}{dx} = b + 2cx$$

从 T 到 A 的关联(我们会把每个变换关联到一个矩阵)与选择基底有关，输入基底 $1, x, x^2$ 与输出基底 $1, x$。

我们现在检视积分，他们产生导数的伪逆 $\boldsymbol{T^+}$！当 1 的导数是 0，我不能写成 T^{-1}，也不能说“T 的逆”。

范例 5 积分 T^+也是线性：$\int_0^x (D+Ex)dx = Dx + \frac{1}{2}Ex^2$。

现在输入基底是 $1, x$，输出基底是 $1, x, x^2$，T^+的矩阵 A^+是 3×2：

$$\textbf{输入 } \boldsymbol{v}\text{：}\ D+Ex \qquad \textbf{乘法}\ A^+\boldsymbol{v} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ D \\ \frac{1}{2}E \end{bmatrix} \quad \begin{matrix} \text{输出} = v\text{的积分} \\ T^+(v) = Dx + \frac{1}{2}Ex^2 \end{matrix}$$

微积分基础定理告诉我们积分是微分的“伪逆”，对于线性代数来说，矩阵 A^+ 是矩阵 A 的伪逆矩阵：

$$A^+A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{bmatrix} \quad \text{与} \quad AA^+ = \begin{bmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{bmatrix} \tag{4}$$

常数函数的导数是零，零在 A^+A 的对角线，没有 $T = d/dx$ 的一维零空间，微积分就不再是微积分。

变换的例子(大多是线性)

范例 6 把每个三维向量投影在水平面 $z = 1$ 上，向量 $\boldsymbol{v} = (x, y, z)$变换到 $T(\boldsymbol{v}) = (x, y, 1)$，这个变换不是线性。为什么不是？它甚至都没有把 $\boldsymbol{v} = \mathbf{0}$ 变换到 $T(\boldsymbol{v}) = \mathbf{0}$。

范例 7 假设 A 是可逆矩阵，当然 $T(\boldsymbol{v} + \boldsymbol{w}) = A\boldsymbol{v} + A\boldsymbol{w} = T(\boldsymbol{v}) + T(\boldsymbol{w})$。另一个线性变换是左乘 A^{-1}，得到逆变换 T^{-1}，这个变换把每个向量 $T(\boldsymbol{v})$带回 $\boldsymbol{v}$：

$$T^{-1}(T(\boldsymbol{v})) = \boldsymbol{v} \qquad \text{适配矩阵乘法} \quad A^{-1}(A\boldsymbol{v}) = \boldsymbol{v}$$

若 $T(\boldsymbol{v}) = A\boldsymbol{v}$ 与 $S(\boldsymbol{u}) = B\boldsymbol{u}$，则乘积 $T(S(\boldsymbol{u}))$适配乘积 $AB\boldsymbol{u}$。

我们遇到一个不可避免的问题，从 $\mathbf{V} = \mathbf{R}^n$ 到 $\mathbf{W} = \mathbf{R}^m$ 所有的线性变换都可以利用矩阵达成吗？当一个线性 T 描述成：“旋转”或“投影”或“...”，一定存在矩阵 A 隐藏在 T 背后吗？$T(\boldsymbol{v})$永远是 $A\boldsymbol{v}$ 吗？

答案是“对的”。这是线性代数里面一种不从矩阵开始的方法，我们仍然在结束的时候用到矩阵——在我们选择了输入基底与输出基底之后。

注解 变换有自己的语言。对于矩阵，列空间包含所有的输出 $A\boldsymbol{v}$，零空间包含所有使得 $A\boldsymbol{v} = \mathbf{0}$ 的输入。这些语言翻译后变成：“值域(range)”与“核(kernel)”：

***T* 的值域** = 所有输出 $T(\boldsymbol{v})$的集合，值域对应到列空间。

***T* 的核** = 所有使得 $T(\boldsymbol{v}) = \mathbf{0}$ 的输入的集合，核对应到零空间。

值域在输出空间 $\mathbf{W}$ 中，核在输入空间 $\mathbf{V}$ 中。当 T 是左乘一个矩阵，$T(\boldsymbol{v}) = A\boldsymbol{v}$，则值域是列空间以及核是零空间。

请尊重版权与译者的劳动成果，侵权必究！

平面的线性变换

了解一个变换要比定义它有趣多了，当一个 2×2 矩阵 A 乘所有 $\mathbf{R}^2$ 的向量，我们可以看出它是如何动作。从一个有 11 个端点的“房子”开始，我们变换 11 的向量 $\boldsymbol{v}$ 到 11 个向量 $A\boldsymbol{v}$，$\boldsymbol{v}$'s 之间的直线变成变换后向量 $A\boldsymbol{v}$ 之间的直线。(从房子到房子的变换是线性！) 应用 A 至一间标准房子得到一间新房子--可能拉伸或旋转或其他不适合居住的动作。

本书这个部分是视觉化的，不是理论的，我们会展示四间房子以及产生他们的矩阵。H 的列是第一间房子的 11 个角，(H 是 2×12，所以问题 25 的 plot2d 会关联第 11 个与第 1 个角。) A 乘房子矩阵 H 的 11 个点得到其他房子的角 AH。

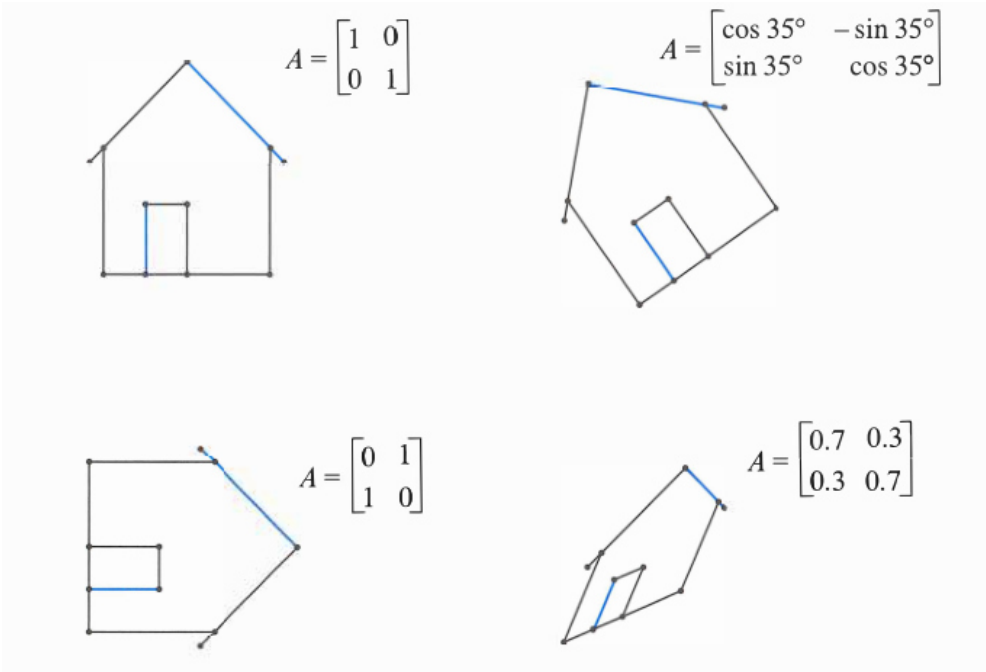


图 8.2：一间房子的线性变换，利用 plot2d(A* H)作图。

主要观念的复习

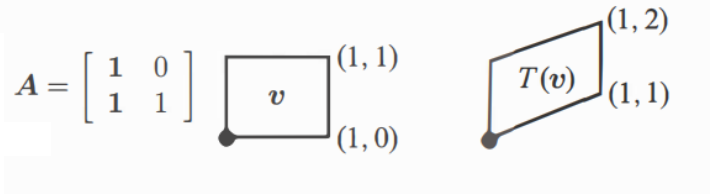
1. 变换 T 把输入空间的每个 $\boldsymbol{v}$ 带到输出空间的 $T(\boldsymbol{v})$。
2. 若 $T(\boldsymbol{v}+\boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$且 $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$，则 T 是线性：直线到直线。
3. 组合到组合：$T(c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n) = c_1T(\boldsymbol{v}_1) + \ldots + c_nT(\boldsymbol{v}_n)$。
4. $T =$ 导数与 $T^+ =$ 积分是线性，从 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的 $T(\boldsymbol{v}) = A\boldsymbol{v}$ 也是线性。

已解范例

8.1A 消元矩阵$\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$产生一个剪切(shearing)变换，从$(x, y)$到$T(x, y) = (x, x + y)$。如果填满一个方形，画出变换之后的方形。

解 在x轴的点$(1, 0)$与$(2, 0)$被T变换到45°直线上的$(1, 1)$与$(2, 2)$，在y轴的点不会移动：$T(0, y) = (0, y)$ =特征向量$(\lambda = 1)$。

垂直线往上斜
这是剪切
平面变成平行四边形



8.1B 若输出空间的每个$\boldsymbol{b}$来自输入空间中唯一的$\boldsymbol{x}$：$T(\boldsymbol{x}) = \boldsymbol{b}$永远只有一个解，则非线性变换$T$是可逆。下列哪些变换(对于实数$x$)是可逆，$T^{-1}$为何？没有线性的变换，就算是$T_3$也不是。当你求解$T(\boldsymbol{x}) = \boldsymbol{b}$，你是在逆反$T$。

$$T_1(x) = x^2 \quad T_2(x) = x^3 \quad T_3(x) = x + 9 \quad T_4(x) = e^x \quad T_5(x) = 1/x,\ x \neq 0$$

解 T_1不可逆：$x^2 = 1$有两个解且$x^2 = -1$无解。

T_4不可逆，因为$e^x = -1$无解。(如果输出空间变成正数b's，则$e^x = b$的逆反是$x = \ln b$。)

注意T_5^2=单位。但是$T_3^2(x) = x + 18$，$T_3^2(x)$与$T_4^2(x)$为何？

T_2, T_3, T_5可逆：$x^3 = b$，$x + 9 = b$，$1/x = b$只有一个解x？

$$x = T_2^{-1}(b) = x^{1/3} \quad x = T_3^{-1}(b) = b - 9 \quad x = T_5^{-1}(b) = 1/b$$

问题集 8.1

1 一个线性变换必须保持零向量固定：$T(\boldsymbol{0}) = \boldsymbol{0}$。选择$\boldsymbol{w}$ = ____，从$T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$证明$T(\boldsymbol{0}) = \boldsymbol{0}$ (完成证明)。也选择$\boldsymbol{v}$ = ____，从$T(c\boldsymbol{v}) = cT(\boldsymbol{v})$证明。

2 要求(b)得到$T(c\boldsymbol{v}) = cT(\boldsymbol{v})$以及$T(d\boldsymbol{w}) = dT(d\boldsymbol{w})$，利用加法，要求(a)得到$T(\quad) =$ (　)。$T(c\boldsymbol{v} + d\boldsymbol{w} + e\boldsymbol{u})$ = ？

3 下列哪些变换不是线性？输入是$\boldsymbol{v} = (v_1, v_2)$:

(a) $T(\boldsymbol{v}) = (v_2, v_1)$　(b) $T(\boldsymbol{v}) = (v_1, v_1)$　(c) $T(\boldsymbol{v}) = (0, v_1)$

(d) $T(\boldsymbol{v}) = (0, 1)$　(e) $T(\boldsymbol{v}) = v_1 - v_2$　(f) $T(\boldsymbol{v}) = v_1v_2$

请尊重版权与译者的劳动成果，侵权必究!

4 如果 S 与 T 是线性变换，$T(S(\boldsymbol{v}))$是线性还是二次式？

(a) (特殊案例) 若 $S(\boldsymbol{v}) = \boldsymbol{v}$ 且 $T(\boldsymbol{v}) = \boldsymbol{v}$，则 $T(S(\boldsymbol{v})) = \boldsymbol{v}$ 或 v^2？

(b) (一般案例) $S(\boldsymbol{v}_1 + \boldsymbol{v}_2) = S(\boldsymbol{v}_1) + S(\boldsymbol{v}_2)$且 $T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = T(\boldsymbol{v}_1) + T(\boldsymbol{v}_2)$组合成

$T(S(\boldsymbol{v}_1 + \boldsymbol{v}_2)) = T(_____) = _____ + _____$。

5 假设除了 $T(0, \boldsymbol{v}_2) = (0, 0)$之外 $T(\boldsymbol{v}) = \boldsymbol{v}$，证明这个变换满足 $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$，但是不满足 $T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$。

6 下列哪些变换满足 $T(\boldsymbol{v} + \boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$？哪些变换满足 $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$？

(a) $T(\boldsymbol{v}) = \boldsymbol{v}/\|\boldsymbol{v}\|$ (b) $T(\boldsymbol{v}) = v_1 + v_2 + v_3$ (c) $T(\boldsymbol{v}) = (v_1, 2v_2, 3v_3)$

(d) $T(\boldsymbol{v}) = \boldsymbol{v}$ 的最大分量

7 对下列 $\mathbf{V} = \mathbf{R}^2$ 到 $\mathbf{W} = \mathbf{R}^2$ 的变换，求出 $T(T(\boldsymbol{v}))$。证明当 $T(\boldsymbol{v})$是线性，则 $T(T(\boldsymbol{v}))$也是线性。

(a) $T(\boldsymbol{v}) = -\boldsymbol{v}$ (b) $T(\boldsymbol{v}) = \boldsymbol{v} + (1, 1)$

(c) $T(\boldsymbol{v}) = 90°$旋转 $= (-v_2, v_1)$

(d) $T(\boldsymbol{v}) =$ 投影 $= ½(v_1 + v_2, v_1 + v_2)$

8 求出 T 的值域与核(如同列空间与零空间)：

(a) $T(v_1, v_2) = (v_1 - v_2, 0)$ (b) $T(v_1, v_2, v_3) = (v_1, v_2)$

(c) $T(v_1, v_2) = (0, 0)$ (d) $T(v_1, v_2) = (v_1, v_1)$

9 “循环”变换 T 的定义是 $T(v_1, v_2, v_3) = (v_2, v_3, v_1)$，$T(T(\boldsymbol{v})) =$？ $T^3(\boldsymbol{v}) =$？ $T^{100}(\boldsymbol{v}) =$？对 $\boldsymbol{v}$ 使用 T 一百次。

10 当线性变换的值域是全部的 $\mathbf{W}$ 且核只有 $\boldsymbol{v} = \mathbf{0}$ 时，从 $\mathbf{V}$ 到 $\mathbf{W}$ 的线性变换转存在从 $\mathbf{W}$ 到 $\mathbf{V}$ 的逆变换，则对于每一个 $\mathbf{W}$ 中的 $\boldsymbol{w}$，$T(\boldsymbol{v}) = \boldsymbol{w}$ 有一个解 $\boldsymbol{v}$。为什么下列的 T 不可逆？

(a) $T(v_1, v_2) = (v_2, v_2)$ $\mathbf{W} = \mathbf{R}^2$

(b) $T(v_1, v_2) = (v_1, v_2, v_1 + v_2)$ $\mathbf{W} = \mathbf{R}^3$

(c) $T(v_1, v_2) = v_1$ $\mathbf{W} = \mathbf{R}^1$

11 如果 $T(\boldsymbol{v}) = A\boldsymbol{v}$ 且 A 是 $m \times n$，则 T 是“左乘 A”

(a) 输入与输出空间 $\mathbf{V}$ 与 $\mathbf{W}$ 为何？

(b) T 的值域 $= A$ 的列空间 $=$ ？

(c) T 的核 $= A$ 的零空间 $=$ ？

12 假设线性变换 T 把(1, 1)转到(2, 2)以及把(2, 0)转到(0, 0)，求 $T(\boldsymbol{v})$：

(a) $\boldsymbol{v} = (2, 2)$ (b) $\boldsymbol{v} = (3, 1)$ (c) $\boldsymbol{v} = (-1, 1)$ (d) $\boldsymbol{v} = (a, b)$

问题 13-19 会困难一些，输入空间包含所有的 2×2 矩阵 M。

13 M 是任意 2×2 矩阵且 $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$，变换 T 的定义是 $T(M) = AM$。什么样的矩阵乘法证明 T 是线性。

14 假设 $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$，证明 T 的值域是整个矩阵空间 $\mathbf{V}$ 且核是零矩阵：

(1) 若 $AM = 0$ 证明 M 必须是零矩阵。

(2) 对于任意 2×2 矩阵 B，求出一个 $AM = B$ 的解。

15 假设 $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$，证明单位矩阵 I 不在 T 的值域中。求出一个非零矩阵 M 使得 $T(M) = AM$ 是零。

16 假设 T 把每个 2×2 矩阵 M 转置，尝试找出一个矩阵 A 使得 $AM = M^{\mathrm{T}}$，证明没有矩阵 A 做得到。给教授：这个线性变换不是来自一个矩阵？这个矩阵应该是 4×4！

17 把每个 2×2 矩阵转置的变换 T 肯定是线性，下列哪些额外的性质是正确的？

(a) T^2 是单位变换。

(b) T 的核是零矩阵。

(c) 每个 2×2 矩阵在 T 的值域中。

(d) $T(M) = -M$ 是不可能。

18 假设 $T(M) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} [M] \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$，求出一个矩阵使得 $T(M) \neq 0$。描述所有使得 $T(M) = 0$ 的矩阵(核)以及所有的输出矩阵 $T(M)$ (值域)。

19 如果 A 与 B 可逆且 $T(M) = AMB$，求出(___) M (___)形式的 $T^{-1}(M)$。

问题 20-26 有关房子变换，输出是 $\boldsymbol{T(H) = AH}$。

20 你如何从 T(房子)的图形分辨 A 是

(a) 对角矩阵？

(b) 秩 1 矩阵？

(c) 下三角矩阵？

21 对于下列矩阵画出 T(房子)的图形：

$$D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad 与 \quad A = \begin{bmatrix} .7 & .7 \\ .3 & .3 \end{bmatrix} \quad 与 \quad U = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

22 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 在什么条件下使得 T(房子)会

(a) 坐直朝上？

(b) 所有方向都扩大 3 倍？

(c) 不改变形状的旋转？

23 当 $T(\boldsymbol{v}) = -\boldsymbol{v} + (1, 0)$，描述 T(房子)，这个 T 是“仿射”。

24 改变房子矩阵 H 多出一个烟囱。

25 标准房子可以用 plot2d(H)来画图，o 代表圆，–代表线：

```
x = H(1, :)' ; y = H(2, :)';
axis([-10 10 -10 10]), axis('square')  【原文错误，少了空格】
plot(x, y, 'o', x, y, '-');
```

测试 plot2d(A'*H)与 plot2d(A'*A*H)，房子是图 8.1 的矩阵。

26 对下列矩阵 A，不用电脑画出房子 A*H 的图形：

$$\begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} \quad 与 \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

27 这个代码建立一个 θ 分成 50 个角度的向量，它画出单位圆然后画出 T(圆) =椭圆，$T(\boldsymbol{v}) = A\boldsymbol{v}$ 把圆变换成椭圆。

```
A = [2 1;1 2]     % 你可以改变 A
theta = [0:2 * pi/50:2 * pi];
circle = [cos(theta); sin(theta)];
ellipse = A * circle;
axis([–4 4 –4 4]); axis('square')
plot(circle(1,:), circle(2,:), ellipse(1,:), ellipse(2,:))
```

28 把问题 27 的圆加上两个眼睛以及一个微笑，(如果一个眼睛深色另一个就用浅色，当脸横跨 y 轴产生反射时，你才能分辨。) 左乘矩阵 A 得到新的脸。

29 什么样的条件 $\det A = ad - bc$，可以确保输出房子 AH 会

(a) 挤压成一条线？

(b) 顺时针方向保持它的端点(不是反射)？

(a) 与原始房子有相同的面积？

30 为什么每个从 $\mathbf{R}^2$ 到 $\mathbf{R}^2$ 的线性变换 T，会把方形变成平行四边形？矩形也是变成平行四边形(如果 T 不可逆，会被挤压。)

8.2 线性变换的矩阵

1 给定一组输入基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$，如果我们知道 $T(\boldsymbol{v}_1), ..., T(\boldsymbol{v}_n)$，则我们知道所有的 $T(\boldsymbol{v})$：使用**线性**。

2 “对应 T 的矩阵”的列 j 来自应用 T 到输入基底向量 $\boldsymbol{v}_j$。

3 使用 $\boldsymbol{w}$'s 的输出基底把变换写成 $T(\boldsymbol{v}_j) = a_{1j}\boldsymbol{w}_1 + ... + a_{mj}\boldsymbol{w}_m$，这些 a_{ij} 进入列 j。

4 若输入与输出基底 $= I_{n\times n}$ 的列与 $I_{m\times m}$ 的列，则对应 $T(\boldsymbol{x}) = A\boldsymbol{x}$ 的矩阵是 A。

5 当基底变成 $\boldsymbol{v}$'s 与 $\boldsymbol{w}$'s，相同 T 的对应矩阵从 A 变成 $W^{-1}AV$。

6 最佳基底：$V = W =$ 特征向量以及 V, W 等于得到对角Λ与Σ的奇异向量。

下一页会指定一个矩阵给每一个变换 T。对于一般的列向量，输入 $\boldsymbol{v}$ 在 $\mathbf{V} = \mathbf{R}^n$ 中且输出 $T(\boldsymbol{v})$在 $\mathbf{W} = \mathbf{R}^m$ 中，这个变换的矩阵 A 是 $m\times n$，我们在 $\mathbf{V}$ 与 $\mathbf{W}$ 中选择的基底会决定 A。

$\mathbf{R}^n$ 与 $\mathbf{R}^m$ 的标准基底是 I 的列，这个选择得到一个标准矩阵，则 $T(\boldsymbol{v}) = A\boldsymbol{v}$ 在正常的方式。但是这些空间会有其他基底，所以其他矩阵也可以代表相同的变换 T。线性代数的主题是选择基底得到对应 T 的最佳矩阵(对角矩阵)。

所有的向量空间 $\mathbf{V}$ 与 $\mathbf{W}$ 有很多基底，每个基底的选择产生一个对应 T 的矩阵。当输入基底与输出基底不同时，对应 $T(\boldsymbol{v}) = \boldsymbol{v}$ 的矩阵就不是单位 I，它会是“基底矩阵的变化”。下列是关键概念：

假设我们知道输入基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ 的 $T(\boldsymbol{v})$，
矩阵的列 1 到列 n 会包含这些输出 $T(\boldsymbol{v}_1)$到 $T(\boldsymbol{v}_n)$。
A 乘 $\boldsymbol{c}$ = 矩阵乘向量 = 这 n 个列的组合
$A\boldsymbol{c}$ 是正确组合 $c_1T(\boldsymbol{v}_1) + ... + c_nT(\boldsymbol{v}_n) = T(\boldsymbol{v})$。

理由 每个 $\boldsymbol{v}$ 是基底向量 $\boldsymbol{v}_j$ 的唯一组合 $c_1\boldsymbol{v}_1 + ... + c_n\boldsymbol{v}_n$，由于 T 是线性变换(线性的时刻来临)，$T(\boldsymbol{v})$必须有与在列中的输出 $T(\boldsymbol{v}_j)$的相同组合 $c_1T(\boldsymbol{v}_1) + ... + c_nT(\boldsymbol{v}_n)$。

我们的第一个例子给出对应在 $\mathbf{R}^2$ 与 $\mathbf{R}^3$ 的标准基底的矩阵 A。

范例 1 假设 T 把 $\boldsymbol{v}_1 = (1, 0)$变换到 $T(\boldsymbol{v}_1) = (2, 3, 4)$，假设第二个基底向量 $\boldsymbol{v}_2 = (0, 1)$ 变成 $T(\boldsymbol{v}_2) = (5, 5, 5)$。若 T 是从 $\mathbf{R}^2$ 到 $\mathbf{R}^3$ 的线性变换，它的“标准矩阵”是 3×2，这些输出 $T(\boldsymbol{v}_1)$与 $T(\boldsymbol{v}_2)$进入 A 的列：

$$A = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix}, \quad c_1 = 1 \text{ 且 } c_2 = 1 \text{ 得到 } T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

范例 2 假设输入空间 $\mathbf{V}=\mathbf{R}^2$，输出空间 $\mathbf{W}$ 也是 $\mathbf{R}^2$。假设 $T(\boldsymbol{v})=\boldsymbol{v}$ 是单位变换，你会预期这个矩阵是 I，但是这个只会发生在输入基底与输出基底相同的时候。我会选择不同的基底来看看矩阵如何创建。

对于这个特例 $T(\boldsymbol{v})=\boldsymbol{v}$，我会把矩阵称为 B 而不是 A，我们只需要从 $\boldsymbol{v}$'s 到 $\boldsymbol{w}$'s 做基底变换，每个 $\boldsymbol{v}$ 是 $\boldsymbol{w}_1$ 与 $\boldsymbol{w}_2$ 的组合。

$$\text{输入基底}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}=\begin{bmatrix}3 & 6\\ 3 & 8\end{bmatrix}\quad \text{输出基底}\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{bmatrix}=\begin{bmatrix}3 & 0\\ 1 & 2\end{bmatrix}$$

$$\begin{matrix}\text{基底} & \boldsymbol{v}_1=1\boldsymbol{w}_1+1\boldsymbol{w}_2\\ \text{变换} & \boldsymbol{v}_2=2\boldsymbol{w}_1+3\boldsymbol{w}_2\end{matrix}$$

请注意！我利用输出基底 $\boldsymbol{w}_1$ 与 $\boldsymbol{w}_2$ 来表示输入基底 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$，这是因为我们的关键规则。我们把单位变换 T 应用到每一个输入基底向量：$T(\boldsymbol{v}_1)=\boldsymbol{v}_1$ 与 $T(\boldsymbol{v}_2)=\boldsymbol{v}_2$，然后我们利用输出基底 $\boldsymbol{w}_1$ 与 $\boldsymbol{w}_2$ 来表示这些输出 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$。这些粗体数字 **1**, **1**, **2**, **3** 告诉我们矩阵 B 的列 1 与列 2(基底矩阵的变换)：$WB=V$ 所以 $\boldsymbol{B}=\boldsymbol{W}^{-1}\boldsymbol{V}$。

基底变换的矩阵 $\boldsymbol{B}$
$$\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{bmatrix}\begin{bmatrix}B\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}\ \text{是}\ \begin{bmatrix}3 & 0\\ 1 & 2\end{bmatrix}\begin{bmatrix}\mathbf{1} & \mathbf{2}\\ \mathbf{1} & \mathbf{3}\end{bmatrix}=\begin{bmatrix}3 & 6\\ 3 & 8\end{bmatrix}\qquad (1)$$

> 当输入基底在矩阵 $\boldsymbol{V}$ 的列，输出基底在矩阵 $\boldsymbol{W}$ 的列，对应 $\boldsymbol{T}=\boldsymbol{I}$ 的基底变换矩阵是 $\boldsymbol{B}=\boldsymbol{W}^{-1}\boldsymbol{V}$。

关键 我看到一个了解 $B=W^{-1}V$ 的清晰方法，假设相同的向量 $\boldsymbol{u}$ 写成输入基底 $\boldsymbol{v}$'s 的组合以及输出基底 $\boldsymbol{w}$'s 的组合，我会用三种方式来进行：

$$\begin{matrix}\boldsymbol{u}=c_1\boldsymbol{v}_1+\ldots+c_n\boldsymbol{v}_n\\ \boldsymbol{u}=d_1\boldsymbol{w}_1+\ldots+d_n\boldsymbol{w}_n\end{matrix}\quad \text{是}\quad \begin{bmatrix}\boldsymbol{v}_1 & \ldots & \boldsymbol{v}_n\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{w}_1 & \ldots & \boldsymbol{w}_n\end{bmatrix}\begin{bmatrix}d_1\\ \vdots\\ d_n\end{bmatrix}\ \text{且}\ \boldsymbol{Vc}=\boldsymbol{Wd}$$

$\boldsymbol{w}$'s 新基底的系数 $\boldsymbol{d}$ 是 $\boldsymbol{d}=W^{-1}V\boldsymbol{c}$，则 B 是 $W^{-1}V$ (2)

这个公式 $B=W^{-1}V$ 产生世界上最伟大的奥秘之一：当标准基底 $V=I$ 变换成不同基底 W，基底变换矩阵不是 W 而是 $B=W^{-1}$。较大的基底向量有较小的系数！标准基底中的 $\begin{bmatrix}x\\ y\end{bmatrix}$，在 $\boldsymbol{w}_1$ 与 $\boldsymbol{w}_2$ 的基底中有系数 $\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{bmatrix}^{-1}\begin{bmatrix}x\\ y\end{bmatrix}$。

矩阵的建立

现在我们对任意的线性变换创建矩阵。假设 T 把空间 $\mathbf{V}(n$ 维)变换到空间 $\mathbf{W}(m$ 维)，我们选择 $\mathbf{V}$ 的一组基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ 以及 $\mathbf{W}$ 的一组基底 $\boldsymbol{w}_1, ..., \boldsymbol{w}_m$，矩阵 A 会是 $m \times n$。要求得 A 的第一列，应用 T 到第一个基底向量 $\boldsymbol{v}_1$，输出 $T(\boldsymbol{v}_1)$在 $\mathbf{W}$ 中。

$T(\boldsymbol{v}_1)$ 是 $\mathbf{W}$ 中的输出基底的一个组合 $a_{11}\boldsymbol{w}_1 + ... + a_{m1}\boldsymbol{w}_m$。

这些数字 $a_{11}, ..., a_{m1}$ 进入 A 的第一列，从 $\boldsymbol{v}_1$ 变换到 $T(\boldsymbol{v}_1)$适配 A 乘$(1, 0, ..., 0)$，它得到矩阵的第一列。当 T 是导数以及第一个基底向量是 1，它的导数是 $T(\boldsymbol{v}_1) = 0$，所以对于下列的导数矩阵，A 的第一列全部是零。

范例 3 输入 $\boldsymbol{v}$ 的基底是 $1, x, x^2, x^3$，输出 $\boldsymbol{w}$ 的基底是 $1, x, x^2$。

$$T \text{ 取导数：} T(v) = dv/dx \text{ 以及 } A = \text{“导数矩阵”}$$

若 $c_1 + c_2x + c_3x^2 + c_4x^3$
则 $\dfrac{dv}{dx} = \mathbf{1}c_2 + \mathbf{2}c_3x + \mathbf{3}c_4x^2$

$$A\boldsymbol{c} = \begin{bmatrix} 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & \mathbf{3} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} c_2 \\ \mathbf{2}c_3 \\ \mathbf{3}c_4 \end{bmatrix}$$

关键规则： A 的第 j 列是应用 T 到第 j 个基底向量 $\boldsymbol{v}_j$

$$T(\boldsymbol{v}_j) = \text{输出基底向量的组合} = a_{11}\boldsymbol{w}_1 + ... + a_{m1}\boldsymbol{w}_m \tag{3}$$

这些数字进入 A，矩阵的创建要得到正确的基底向量，然后线性得到其他正确的向量。每个 $\boldsymbol{v}$ 是一个组合 $c_1\boldsymbol{v}_1 + ... + c_n\boldsymbol{v}_n$，而且 $T(\boldsymbol{v})$是 $\boldsymbol{w}$'s 的一个组合。当 A 乘在 $\boldsymbol{v}$ 组合的向量 $\boldsymbol{c} = (c_1, ..., c_n)$，$A\boldsymbol{c}$ 得到在 $T(\boldsymbol{v})$组合的系数。这是因为矩阵乘法(组合列)跟 T 一样是线性。

矩阵 A 告诉我们 T 在做什么，每个从 $\mathbf{V}$ 到 $\mathbf{W}$ 的线性变换可以变换成一个矩阵，这个矩阵与基底有关。

范例 4 对于积分 $T^+(\boldsymbol{v})$，第一个基底函数还是 1，它的积分是第二个基底函数 x，所以“积分矩阵” A^+的第一列是$(0, 1, 0, 0)$。

$d_1 + d_2x + d_3x^2$ 的积分是
$d_1x + \dfrac{1}{2}d_2x^2 + \dfrac{1}{3}d_3x^3$

$$A^+\boldsymbol{d} = \begin{bmatrix} 0 & 0 & 0 \\ \mathbf{1} & 0 & 0 \\ 0 & \frac{\mathbf{1}}{\mathbf{2}} & 0 \\ 0 & 0 & \frac{\mathbf{1}}{\mathbf{3}} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ d_1 \\ \frac{\mathbf{1}}{\mathbf{2}}d_2 \\ \frac{\mathbf{1}}{\mathbf{3}}d_3 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

如果你积分一个函数再微分，你得到原来的函数，所以 $AA^+ = I$。但是如果你先微分再积分，常数项丢失了，所以 A^+A 不是 I。1 的导数的积分是零：

$$T^+T(1) = \text{零函数的积分} = 0$$

这个结果适配 A^+A，它的第一列全部是零。导数 T 有一个核(常数函数)，它的矩阵 A 有一个零空间。再说一次主要概念 $A\boldsymbol{v}$ 复制了 $T(\boldsymbol{v})$。

导数与积分的范例说明了三点，第一，线性变换 T 无所不在——在微积分，微分方程式与线性代数。第二，$\mathbf{R}^n$ 以外的空间很重要——我们有函数在 $\mathbf{V}$ 与 $\mathbf{W}$。第三，如果我们先微分再积分，我们可以执行矩阵相乘 A^+A。

矩阵乘积 AB 适配变换 TS

现在有些东西很重要——矩阵乘法规则的真正理由，最终我们发现了为什么！两个线性变换 T 与 S 由两个矩阵 A 与 B 代表，现在比较 TS 与乘积 AB：

当我们应用 T 去变换来自 S 的输出，我们根据下列规则得到 TS：

$(TS)(\boldsymbol{u})$的定义是 $T(S(\boldsymbol{u}))$，输出 $S(\boldsymbol{u})$变成 T 的输入。

当我们矩阵 A 到来自 B 的输出，我们根据下列规则得到乘积 AB：

$(AB)(\boldsymbol{x})$的定义是 $A(B\boldsymbol{x})$，输出 $B\boldsymbol{x}$ 变成 A 的输入。

矩阵乘法得到正确的矩阵 AB 来代表 TS。

变换 S 是从空间 U 到 V，它的矩阵 B 使用 U 的一组基底 $\boldsymbol{u}_1, ..., \boldsymbol{u}_p$ 以及 V 的一组基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$。变换 T 仍然是从 $\mathbf{V}$ 到 $\mathbf{W}$，它的矩阵 A 必须使用相同的 $\mathbf{V}$ 的基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$——这是 S 的输出空间，也是 T 的输入空间。则**矩阵 AB 适配 TS**。

乘法 线性变换 TS 从 $\mathbf{U}$ 中任意的向量 $\boldsymbol{u}$ 开始，来到 $\mathbf{V}$ 中的 $S(\boldsymbol{u})$，然后到 $\mathbf{W}$ 中的 $T(S(\boldsymbol{u}))$。矩阵 AB 从 $\mathbf{R}^p$ 任意的 $\boldsymbol{x}$ 开始，来到 $\mathbf{R}^n$ 的 $B\boldsymbol{x}$，然后到 $\mathbf{R}^m$ 的 $AB\boldsymbol{x}$。**矩阵 AB 正确的代表 TS：**

TS：$\mathbf{U} \to \mathbf{V} \to \mathbf{W}$　　AB：$(m\times n)\,(n\times p) = (m\times p)$

输入是 $\boldsymbol{u} = x_1\boldsymbol{u}_1 + ... + x_p\boldsymbol{u}_p$，输出 $T(S(\boldsymbol{u}))$适配输出 $AB\boldsymbol{x}$。**变换 TS 的乘积适配矩阵的乘积 AB**。

最重要的情况是当空间 $\mathbf{U}, \mathbf{V}, \mathbf{W}$ 完全相同，他们的基底也完全相同，此时 $m = n = p$ 我们有方形矩阵可以相乘。

范例 5　S 把平面旋转 θ，T 也把平面旋转 θ，则 TS 旋转 2θ。变换 T^2 对应旋转矩阵 A^2，角度是 2θ ：

$$T = S \quad A = B \quad T^2 = \text{旋转 } 2\theta \quad A^2 = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \tag{4}$$

把(变换)^2 适配(矩阵)^2，我们选出 $\cos 2\theta$ 与 $\sin 2\theta$ 的公式，执行 A 乘 A：

$$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}=\begin{bmatrix}\cos^2\theta-\sin^2\theta & -2\sin\theta\cos\theta\\ 2\sin\theta\cos\theta & \cos^2\theta-\sin^2\theta\end{bmatrix} \tag{5}$$

比较(4)与(5)得到 $\cos 2\theta=\cos^2\theta-\sin^2\theta$ 与 $\sin 2\theta=2\sin\theta\cos\theta$。三角函数(二倍角规则)来自线性代数。

范例 6 S 旋转 θ，T 旋转 $-\theta$，则 $TS=I$ 产生 $AB=I$。在这个情况下 $T(S(\boldsymbol{u}))$ 是 $\boldsymbol{u}$，我们往前旋转又往后旋转。矩阵要适配，$AB\boldsymbol{x}$ 必须等于 $\boldsymbol{x}$，这两个矩阵是逆矩阵。把 $\cos(-\theta)=\cos\theta$ 以及 $\sin(-\theta)=-\sin\theta$ 代入反向旋转矩阵 A：

$$AB=\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}=\begin{bmatrix}\cos^2\theta+\sin^2\theta & 0\\ 0 & \cos^2\theta+\sin^2\theta\end{bmatrix}=I$$

选择最佳基底

现在来到本书这个段落的最后一个步骤，选择基底对角化矩阵。有了标准基底(I 的列)我们的变换 T 产生某些矩阵 A——可能不是对角矩阵，当我们选择不同的基底，不同的矩阵可以代表相同的 T，这两个伟大的选择是特征向量与奇异向量。

> **特征向量** 若 T 是从 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 的变换，它的矩阵是方形。但是使用标准基底时，矩阵 A 可能不是对角矩阵。如果存在 n 个无关特征向量，选择这些向量作为输入与输出的基底，在这个好基底之下，**对应 T 的矩阵是对角特征值矩阵Λ。**

范例 7 投影矩阵 T 把 $\mathbf{R}^2$ 的每个向量 $\boldsymbol{v}=(x,y)$ 投影到直线 $y=-x$，使用标准基底，$\boldsymbol{v}_1=(1,0)$ 投影到 $T(\boldsymbol{v}_1)=(\frac{1}{2},-\frac{1}{2})$，$\boldsymbol{v}_2=(0,1)$ 投影到 $T(\boldsymbol{v}_2)=(-\frac{1}{2},\frac{1}{2})$。这些是 A 的列：

投影矩阵
标准基底
非对角矩阵

$$A=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\text{ 有 }A^{\mathrm{T}}=A\text{ 以及 }A^2=A$$

现在特征向量的重点来了，让他们作为基底向量！对角化！

请尊重版权与译者的劳动成果，侵权必究！

当基底向量是特征向量时，矩阵变成对角矩阵。

$\boldsymbol{v_1} = \boldsymbol{w}_1 = (1, -1)$ 投影到本身：$T(\boldsymbol{v}_1) = \boldsymbol{v}_1$ 与 $\lambda = 1$

$\boldsymbol{v_2} = \boldsymbol{w}_1 = (1, 1)$ 投影到零：$T(\boldsymbol{v}_2) = \boldsymbol{0}$ 与 $\lambda = 0$

固有向量基底 对角矩阵 新矩阵是 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \Lambda$ (6)

特征向量是最佳基底向量，他们产生特征值矩阵Λ。

其他输入基底 = 输出基底的选择又如何？把这些基底向量放进 B 的列，我们在上面看到的基底矩阵变换(在标准基底与新基底之间)是 $B_{\text{in}} = B$ 与 $B_{\text{out}} = B^{-1}$，T 的新矩阵与 A **相似**(similar)：

在新基底 *b*'s 的 $A_{\text{new}} = B^{-1}AB$ 与标准基底的 A 相似：

$$\boldsymbol{A}_{b\text{'s 到 }b\text{'s}} = \boldsymbol{B}^{-1}{}_{\text{标准到 }b\text{'s}} \quad \boldsymbol{A}_{\text{标准}} \quad \boldsymbol{B}_{b\text{'s 到标准}} \tag{7}$$

对于变换 ITI 我使用乘法规则，对应 I, T, I 的矩阵是 B^{-1}, A, B，矩阵 B 包含标准基底的输入向量 $\boldsymbol{b}$。

最终我们允许不同的空间 $\mathbf{V}$ 与 $\mathbf{W}$，以及不同的基底 $\boldsymbol{v}$'s 与 $\boldsymbol{w}$'s。当我们知道 T 以及选择的基底，我们得到一个矩阵 A，A 可能不是对称，甚至不是方形，但是我们永远可以选择 $\boldsymbol{v}$'s 与 $\boldsymbol{w}$'s 得到对角矩阵，这个会是在 $A = U\Sigma V^{\mathrm{T}}$ 的奇异值矩阵 Σ=对角线$(\sigma_1, ..., \sigma_r)$。

奇异向量 SVD 说明了 $U^{-1}AV = \Sigma$，右奇异向量 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ 会是输入基底，左奇异向量 $\boldsymbol{u}_1, ..., \boldsymbol{u}_m$ 会是输出基底。基于矩阵乘法规则，在新基底中对应相同变换的矩阵是 $B^{-1}{}_{\text{out}}\, AB_{\text{in}} = U^{-1}AV = \Sigma$。

我不能说Σ与 A 相似，我们现在处理两个基底，输入与输出，两者都是正交单位基底，而且他们维持了向量的长度。遵循 David Vogan 的好建议，我建议我们说：**Σ与 A 是“等距映射(isometric)”。**

定义 若 Q_1 与 Q_2 是正交矩阵，则 $C = Q_1{}^{-1}AQ_2$ 与 A 是等距映射。

范例 8 针对变换 $T = \frac{d}{dx}$ 创建矩阵 A，我们选择输入基底 $1, x, x^2, x^3$ 以及输出基底 $1, x, x^2$。矩阵 A 很简单，不幸的是它不是对角矩阵，但是我们可以反序基底。现在输入基底是 $x^3, x^2, x, 1$ 以及输出基底是 $x^2, x, 1$，基底变换矩阵 B_{in} 与 B_{out} 是排列矩阵。对应 $T(\boldsymbol{u}) = d\boldsymbol{u}/dx$ 的矩阵在新基底之下是

对角的奇异值矩阵 $\boldsymbol{B}^{-1}{}_{\text{out}}\, \boldsymbol{AB}_{\text{in}} = \boldsymbol{\Sigma}$，其中 σ's = 3, 2, 1

$$\boldsymbol{B}^{-1}{}_{\text{out}}\, \boldsymbol{AB}_{\text{in}} = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix} \begin{bmatrix} 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 \\ 0 & 0 & 0 & \mathbf{3} \end{bmatrix} \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix} = \begin{bmatrix} \mathbf{3} & 0 & 0 & 0 \\ 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \end{bmatrix} \tag{8}$$

嗯，这是一个艰难的段落，我们发现 $x^3, x^2, x, 1$ 的导数是 $3x^2, 2x, 1, 0$。

主要观念的复习

1. 如果我们知道基底的 $T(\boldsymbol{v}_1), ..., T(\boldsymbol{v}_n)$，线性会决定所有的 $T(\boldsymbol{v})$。

2. $\left\{\begin{matrix}\text{线性转换 } T \\ \text{输入基底 } \boldsymbol{v}_1,...,\boldsymbol{v}_n \\ \text{输出基底 } \boldsymbol{w}_1,...,\boldsymbol{w}_n\end{matrix}\right\} \rightarrow \begin{matrix}\text{矩阵 } A\,(m\times n) \\ \text{在这些基底} \\ \text{代表 } T\end{matrix}$

3. 基底变换矩阵 $B = W^{-1}V = B^{-1}{}_{\text{out}}\, B_{\text{in}}$，代表单位变换 $T(\boldsymbol{v}) = \boldsymbol{v}$。

4. 若 A 与 B 代表 T 与 S，对应 S 的输出基底是 T 的输入基底，则矩阵 AB 代表变换 $T(S(\boldsymbol{u}))$。

5. 最佳输入-输出基底是 A 的特征向量与/或奇异向量，则

$$B^{-1}AB = \Lambda = \text{ 特征值} \qquad B^{-1}{}_{\text{out}}\, AB_{\text{in}} = \Sigma = \text{ 奇异值}$$

已解范例

8.2A 2×2 矩阵的空间有 4 个“向量”作为一组基底：

$$\boldsymbol{v}_1 = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}$$

T 是转置每个 2×2 矩阵的线性变换，在此基底(输出基底=输入基底)下，什么矩阵代表 T？逆矩阵 A^{-1} = ？逆反转置运算的变换 T^{-1} = ？

解 转置这 4 个“基底矩阵”再交换 $\boldsymbol{v}_2$ 与 $\boldsymbol{v}_3$：

$$\begin{matrix}T(\boldsymbol{v}_1) = \boldsymbol{v}_1 \\ T(\boldsymbol{v}_2) = \boldsymbol{v}_3 \\ T(\boldsymbol{v}_3) = \boldsymbol{v}_2 \\ T(\boldsymbol{v}_4) = \boldsymbol{v}_4\end{matrix} \qquad \text{得到 } A \text{ 的 4 个列} \qquad A = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

逆矩阵 A^{-1} 与 A 相同，逆变换 T^{-1} 与 T 相同。如果我们转置再转置，最终的矩阵等于原始的矩阵。

注意 2×2 矩阵的空间是 4 维，所以矩阵 A(对应转置 T)是 4×4。A 的零空间是 **Z** 以及 T 的核是零矩阵––唯一转置后得到零的矩阵。A 的特征值是 1, 1, 1, –1。

哪条矩阵线有 $T(A) = A^{\mathrm{T}} = -A$，其中特征值 $\lambda = -1$？

请尊重版权与译者的劳动成果，侵权必究!

问题集 8.2

问题 1-4 延伸第一导数的范例到高阶导数。

1 变换 S 取得二次(第二)导数，维持 $1, x, x^2, x^3$ 作为输入基底 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$，也作为输出基底 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{w}_4$。使用 $\boldsymbol{w}$'s 写出 $S(\boldsymbol{v}_1), S(\boldsymbol{v}_2), S(\boldsymbol{v}_3), S(\boldsymbol{v}_4)$，求出对应 S 的 4×4 矩阵 A_2。

2 什么函数有 $S(\boldsymbol{v}) = \boldsymbol{0}$？他们落在二次导数 S 的核。什么向量落在问题 1 中 S 的矩阵 A_2 的零空间？

3 给定一次导数矩阵 A_1 与二次导数矩阵 A_2，其中 A_2 不是矩形矩阵 A_1 的平方：

$$A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}，\text{不允许 } {A_1}^2 = A_2$$

加入一个零行 4 到 A_1，使得输出空间=输入空间，比较 ${A_1}^2$ 与 A_2。结论：我们希望输出基底=_____基底，则 $m = n$。

4 (a) 一次与二次导数的乘积 TS 产生三次导数，加入零得到 4×4 矩阵，然后计算 $A_1A_2 = A_3$。

(b) 矩阵 ${A_2}^2$ 对应 $S^2 = 4$ 次导数，为什么这个是零？

问题 5-9 有关特定变换 T 与它的矩阵 A。

5 基底 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 与 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$，假设 $T(\boldsymbol{v}_1) = \boldsymbol{w}_2$ 与 $T(\boldsymbol{v}_2) = T(\boldsymbol{v}_3) = \boldsymbol{w}_1 + \boldsymbol{w}_3$，$T$ 是一个线性变换。求矩阵 A 然后乘向量(1, 1, 1)。当输入是 $\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3$ 时，来自 T 的输出为何？

6 由于 $T(\boldsymbol{v}_2) = T(\boldsymbol{v}_3)$，$T(\boldsymbol{v}) = \boldsymbol{0}$ 的解是 $\boldsymbol{v}$ = _____。什么向量在 A 的零空间中？求出 $T(\boldsymbol{v}) = \boldsymbol{w}_2$ 的全部解。

7 找出一个不在 A 的列空间中的向量。求出一个不在变换 T 的值域的 $\boldsymbol{w}$'s 组合。

8 你没有足够的信息来决定 T^2，为什么它的矩阵不一定是 A^2？你还需要什么更多的信息？

9 求出 A 的秩。这个秩不是整个输出空间 $\mathbf{W}$ 的维度，它是 T 的____的维度。

问题 10-13 有关特定变换 T 与它的矩阵 A。

10 假设 $T(\boldsymbol{v}_1) = \boldsymbol{w}_1 + \boldsymbol{w}_2 + \boldsymbol{w}_3$ 与 $T(\boldsymbol{v}_2) = \boldsymbol{w}_2 + \boldsymbol{w}_3$ 与 $T(\boldsymbol{v}_3) = \boldsymbol{w}_3$，利用这些基底向量求出对应 T 的矩阵 A。什么样的输入向量 $\boldsymbol{v}$ 使得 $T(\boldsymbol{v}) = \boldsymbol{w}_1$？

11 逆反问题 10 的矩阵 A，同时逆反变换 T--$T^{-1}(\boldsymbol{w}_1)$与 $T^{-1}(\boldsymbol{w}_2)$与 $T^{-1}(\boldsymbol{w}_3)$ = ？

12 下列哪些是正确？为什么其中一个是荒谬的?

(a) $T^{-1}T = I$　　(b) $T^{-1}T(\boldsymbol{v}_1) = \boldsymbol{v}_1$　　(c) $T^{-1}T(\boldsymbol{w}_1) = \boldsymbol{w}_1$

13 假设空间 **V** 与 **W** 有相同的基底 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$，

(a) 描述一个变换到本身的变换 T(不是 I)。

(b) 描述一个 $T = T^2$ 的变换(不是 I)。

(c) 为什么相同的 T 不能同时用在(a)与(b)。

问题 14-19 有关基底变换。

14 (a) 什么矩阵 B 把(1, 0)变换到(2, 5)以及把(0, 1)变换到(1, 3)?

(b) 什么矩阵 C 把(2, 5)变换到(1, 0)以及把(1, 3)变换到(0, 1)?

(c) 为什么没有矩阵把(2, 6)变换到(1, 0)以及把(1, 3)变换到(0, 1)?

15 (a) 什么矩阵 M 把(1, 0)变换到(0, 1)以及把(r, t)变换到(s, u)?

(b) 什么矩阵 N 把(a, c)变换到(b, d)以及把(1, 0)变换到(0, 1)?

(c) 什么样的 a, b, c, d 的条件使得(b)不可能？

16 (a) 问题 15 中，M 与 N 然后产生一个矩阵可以把(a, c)变换到(r, t)以及把(b, d)变换到(s, u)?

(b) 什么矩阵把(2, 5)变换到(1, 1)以及把(1, 3)变换到(0, 2)?17

17 如果你维持相同的基底但是放置成不同顺序，基底变换矩阵 B 是一个____矩阵。如果你维持基底向量在原来顺序却改变长度，B 是一个____矩阵。

18 把轴向量(1, 0)与(0, 1)旋转一个角度θ的矩阵是 Q，在新的(旋转后)轴之下，原始(1, 0)的坐标为何？如何逆反是个棘手的问题，画出图形或是求解 a 与 b：

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + b \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

19 把(1, 0)与(0, 1)变换到(1, 4)与(1, 5)的矩阵 B 是_____。组合 a(1, 4) + b(1, 5)等于(1, 0)，求(a, b) = (　　,　　)。(1, 0)的新坐标与 B 或 B^{-1} 的关联为何？

问题 20-23 有关二次多项式 $y = A + Bx + Cx^2$ 的空间。

20 在 $x = 1$ 时抛物线 $w_1 = ½(x^2 + x) = 1$，在 $x = 0$ 与 $x = –1$ 时抛物线是零。求出抛物线 w_2 与 w_3，然后使用线性求出 $y(x)$。

(a) $x = 0$ 时 $w_2 = 1$；$x = 1$ 与 $x = –1$ 时 $w_2 = 0$。

(b) $x = –1$ 时 $w_3 = 0$；$x = 0$ 与 $x = 1$ 时 $w_3 = 0$。

(c) $x = 1$ 时 $y(x) = 4$；$x = 0$ 时 $y(x) = 5$，$x = –1$ 时 $y(x) = 6$。使用 w_1, w_2, w_3。

21 二次多项式的一组基底是 $v_1 = 1$ 与 $v_2 = x$ 与 $v_3 = x^2$，另一组基底来自问题 20 的 w_1 与 w_2 与 w_3。求出两个基底变换矩阵，分别从 w's 到 v's 以及 v's 到 w's。

请尊重版权与译者的劳动成果，侵权必究!

22 如果抛物线 $y = A + Bx + Cx^2$ 在 $x = a$ 时抛物线= 4，在 $x = b$ 时抛物线= 5，在 $x = c$ 时抛物线= 6，则 A, B, C 的三个方程式为何？求出 3×3 矩阵的行列式，这个矩阵把 4, 5, 6 的数值变换到抛物线—还是有其他方式？

23 在 $m_1, m_2, ..., m_9$ 的什么条件下，下列三个抛物线构成所有抛物线 $a + bx + cx^2$ 空间的一组基底？

$\boldsymbol{v}_1 = m_1 + m_2x + m_3x^2$，$\boldsymbol{v}_2 = m_4 + m_5x + m_6x^2$，$\boldsymbol{v}_3 = m_7 + m_8x + m_9x^2$

24 Gram-Schmidt 程序把 $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 变成正交基底 $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$，这些是 $A = QR$ 的列。证明 R 是从 $\boldsymbol{a}$'s 到 $\boldsymbol{q}$'s 的基底变换矩阵(当 $A = QR$，$\boldsymbol{a}_2$ 是 $\boldsymbol{q}$'s 的什么组合？)

25 消元法使用 $A = LU$ 把 A 的行变成 U 的行，A 的行 2 是 U 的行的什么组合？写成 $A^{\mathrm{T}} = U^{\mathrm{T}}L^{\mathrm{T}}$ 处理列，基底变换矩阵是 $B = L^{\mathrm{T}}$，如果矩阵是_____，我们就有了基底。

26 假设 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ 是对应 T 的特征向量，这个表示 $T(\boldsymbol{v}_i) = \lambda_i\boldsymbol{v}_i$，$i = 1, 2, 3$。当输入与输出基底都是 $\boldsymbol{v}$'s 时，对应 T 的矩阵是什么？

27 每个可逆的线性变换可以用 I 作为它的矩阵！选择任意输入基底 $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$，选择 $\boldsymbol{w}_i = T(\boldsymbol{v}_i)$作为输出基底。为什么 T 必须是可逆？

28 利用 $\boldsymbol{v}_1 = \boldsymbol{w}_1$ 与 $\boldsymbol{v}_2 = \boldsymbol{w}_2$，求出下列 T's 的标准矩阵：

(a) $T(\boldsymbol{v}_1) = \boldsymbol{0}$ 与 $T(\boldsymbol{v}_2) = 3\boldsymbol{v}_1$　　(a) $T(\boldsymbol{v}_1) = \boldsymbol{v}_1$ 与 $T(\boldsymbol{v}_1 + \boldsymbol{v}_2) = \boldsymbol{v}_1$

29 假设 T 把 xy 平面横跨 x 轴做反射，S 把 xy 平面横跨 y 轴做反射。若 $\boldsymbol{v} = (x, y)$，则 $S(T(\boldsymbol{v}))$为何？找出对于乘积 ST 的简单描述。

30 假设 T 是跨越 45°直线的反射，S 是横跨 y 轴的反射。若 $\boldsymbol{v} = (2, 1)$则 $T(\boldsymbol{v}) = (1, 2)$。求 $S(T(\boldsymbol{v}))$与 $T(S(\boldsymbol{v}))$，通常 $ST \neq TS$。

31 **两个反射的乘积是一个旋转**。把反射矩阵相乘求出旋转矩阵：

$$\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix}$$

32 假设 A 是秩 $r = 2$ 的 3×4 矩阵且 $T(\boldsymbol{v}) = A\boldsymbol{v}$，从 A 的行空间选择 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_2$ 以及从零空间选择 $\boldsymbol{v}_3$ 与 $\boldsymbol{v}_4$ 作为输入基底。选择列空间的 $\boldsymbol{w}_1 = A\boldsymbol{v}_1$，$\boldsymbol{w}_2 = A\boldsymbol{v}_2$ 以及来自 A^{T} 的零空间的 $\boldsymbol{w}_3$ 作为输出基底向量。在这些基底下，什么样特别简单的矩阵可以代表 T?

33 已解范例 **8.2A** 中，2×2 矩阵的空间 **M** 有基底 $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4$。假设 T 是用 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 去乘每个基底矩阵，当 $\boldsymbol{w}$'s 等于 $\boldsymbol{v}$'s 时，什么样的 4×4 矩阵 A 在矩阵空间中可以代表变换 T?

34 是非题：如果我们知道 $\mathbf{R}^n$ 中 n 个相异非零向量的 $T(\boldsymbol{v})$，则我们知道 $\mathbf{R}^n$ 中每个向量 $\boldsymbol{v}$ 的 $T(\boldsymbol{v})$。

网址：linear.neocities.org　邮箱：876152660@qq.com

8.3 搜寻好基底

1 在新的输入基底 B_{in} 与新的输出基底 B_{out} 下，每个矩阵 A 变成 $\boldsymbol{B_{out}^{-1}AB_{in}}$。

2 $B_{in}=B_{out}=$ “$\boldsymbol{A}$ **的广义特征向量**”产生**乔丹形式** $J=B^{-1}AB$。

3 **傅里叶矩阵** $\boldsymbol{F}=B_{in}=B_{out}$ 对角化每个循环矩阵(使用 **FFT**)。

4 正弦与余弦，勒让德与契比雪夫：这些是**函数空间**的伟大基底。

这是本书的重要段落，我害怕大部分读者会跳过去--或是没有精读，这是说明基底概念的第一章。第六章介绍特征向量 $\boldsymbol{x}$ 与第七章找出奇异向量 $\boldsymbol{v}$ 与 $\boldsymbol{u}$，这是两位胜利者，但是很多其他的选择是非常有价值的。

首先是段落 8.2 的纯代数，然后是好基底，**输入基底向量**是 B_{in} 的列，**输出基底向量**是 B_{out} 的列。B_{in} 与 B_{out} 永远是可逆--基底向量无关!

纯代数 若 A 是在标准基底下对应变换 T 的矩阵，则

$$\boldsymbol{B_{out}^{-1}AB_{in}} \text{ 是新基底下的矩阵} \tag{1}$$

标准基底向量是单位矩阵的列：$\boldsymbol{B_{in}=I_{n\times n}}$ 与 $\boldsymbol{B_{out}=I_{m\times m}}$。
我们现在选择特殊的基底使得矩阵比 A 更清楚更简单，当 $B_{in}=B_{out}=B$，方形矩阵 $B^{-1}AB$ 与 A **相似**。

应用代数 应用都与选择优质基底有关，下列是 4 个重要的向量选择以及 3 个函数的选择。段落 8.2 说明特征向量与奇异向量产生Λ与Σ，乔丹形式是新的：

1 $\boldsymbol{B_{in}=B_{out}=}$ **特征向量矩阵** $\boldsymbol{X}$，**则** $\boldsymbol{X^{-1}AX=}$ **特征值在**Λ。

这个选择需要 A 是方形矩阵且有 n 个无关的特征向量，“A 必须可以对角化”，当 $B_{in}=B_{out}$ 是特征向量矩阵 X 时，我们得到Λ。

2 $\boldsymbol{B_{in}=V}$ **与** $\boldsymbol{B_{out}=U}$：$\boldsymbol{A}$ **的奇异向量，则** $\boldsymbol{U^{-1}AV=}$ **对角**Σ。

当 $B_{in}=B_{out}$ 是奇异向量矩阵 V 与 U 时，Σ是奇异值矩阵(有 $\sigma_1,\ldots,\sigma_r$ 在对角线)。回顾这些 B_{in} 与 B_{out} 的列是 A^TA 与 AA^T 的正交单位特征向量，则 $A=U\Sigma V^T$ 得到 $\Sigma=U^{-1}AV$。

3 $\boldsymbol{B_{in}=B_{out}=}$ “$\boldsymbol{A}$ **的广义特征向量**”，**则** $\boldsymbol{B^{-1}AB=}$ **乔丹形式** $\boldsymbol{J}$。

A 是方形矩阵但是它可能只有 s 个无关的特征向量，(若 $s=n$，则 B 是 X 而且 J 是Λ。) 所有情形下乔丹创建 $n-s$ 个额外的“**广义**(generalized)”特征向量，目标在使得乔丹形式 J 尽可能是对角：

i) 沿着 J 的对角线有 n 个方形的方块。

ii) 每个方块有一个特征值λ与一个特征向量以及在对角线上方的 1's。

好案例：有 n 个 1×1 的方块，每个方块有一个特征值，则 J 就是Λ(对角)。

请尊重版权与译者的劳动成果，侵权必究!

范例 1 乔丹矩阵 J 有特征值$\lambda = 2, 2, 3, 3$(两个双重特征值)，因为 J 是三角形。这些特征值落在对角线。对于$\lambda = 2$ 有两个无关特征向量，但是对应$\lambda = 3$ 只有一条特征向量直线。对于每个矩阵 $C = BJB^{-1}$ 与 J 相似是成立的。

$$\textbf{乔丹矩阵}\quad J = \begin{bmatrix} 2 & & & \\ & 2 & & \\ & & \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \end{bmatrix}$$

两个1×1方块
一个2×2方块
三个固有向量
固有值2, 2, 3, 3

对应$\lambda = 2$ 的特征向量是 $\boldsymbol{x}_1 = (1, 0, 0, 0)$与 $\boldsymbol{x}_2 = (0, 1, 0, 0)$，对应$\lambda = 3$ 的一个特征向量是 $\boldsymbol{x}_3 = (0, 0, 1, 0)$。这个乔丹形式的“广义特征向量”是第四个标准基底 $\boldsymbol{x}_4 = (0, 0, 0, 1)$，$J$ 的特征向量(正常与广义)刚好是单位矩阵 I 的列 $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_4$。

注意$\boldsymbol{(J - 3I)x_4 = x_3}$，**广义向量 $\boldsymbol{x}_4$ 与真正的特征向量 $\boldsymbol{x}_3$ 产生关联**，真正的 $\boldsymbol{x}_4$ 会有$(J - 3I)\boldsymbol{x}_4 = \boldsymbol{0}$，但是这里没有出现。

每个与 J 相似的矩阵 $C = BJB^{-1}$ 会有真正的特征向量 $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ 在 B 的前三列，B 的第四列会是一个 C 的广义特征向量 $\boldsymbol{b}_4$，并且紧扣真正的 $\boldsymbol{b}_3$。下列是一个快速证明：利用 $B\boldsymbol{x}_3 = \boldsymbol{b}_3$ 与 $B\boldsymbol{x}_4 = \boldsymbol{b}_4$ 说明第四列 $\boldsymbol{b}_4$ 借由$(C - 3I)\boldsymbol{b}_4 = \boldsymbol{b}_3$ 紧扣 $\boldsymbol{b}_3$。

$$(BJB^{-1} - 3I)\boldsymbol{b}_4 = BJ\boldsymbol{x}_4 - 3B\boldsymbol{x}_4 = B(J - 3I)\boldsymbol{x}_4 = B\boldsymbol{x}_3 = \boldsymbol{b}_3 \tag{2}$$

乔丹定理的重点是每个方形矩阵 A 有一整组的特征向量与广义特征向量，当这些向量进入 B 的列，矩阵 $B^{-1}AB = J$ 是乔丹形式。基于范例 1，下列是 J 的描述。

乔丹形式

对于每个 A，我们想要选择 B 使得 $B^{-1}AB$ 是**尽可能对角**。当 A 有一整组的 n 个特征向量，他们进入 B 的列，则 $B = X$，矩阵 $X^{-1}AX$ 是对角矩阵，**句号**。这是 A 的乔丹形式--当 A 可以对角化。在广义情况下，特征向量缺失，无法得到Λ。

假设 A 有 s 个无关的特征向量，则它与有 s 个方块的乔丹矩阵相似，每个方块有一个特征值在对角线，而且有 1's 在特征值的上方。这个方块对应 A 的确切的一个特征值，则 B 既有正常的特征向量也有广义特征向量。

当有 n 个特征向量时，全部 n 个方块都是 1×1，这个情况下 $J = \Lambda$。

对于任意方形矩阵 $A = BJB^{-1}$，乔丹形式是微分方程式 $d\boldsymbol{u}/dt = A\boldsymbol{u}$ 的解。其中解 $e^{At}\boldsymbol{u}(0)$变成 $\boldsymbol{u}(t) = Be^{Jt}B^{-1}\boldsymbol{u}(0)$，$J$ 是三角形，它的矩阵指数 e^{Jt} 包含$e^{\lambda t}$ 乘次方 $1, t, \ldots, t^{s-1}$。

(乔丹形式) 如果 A 有 s 个无关的特征向量，则它与有 s 个乔丹方块在对角线的乔丹矩阵**相似**，某个矩阵 B 把 A 放进乔丹形式：

$$\textbf{乔丹形式}\qquad B^{-1}AB=\begin{bmatrix}J_1 & & \\ & \ddots & \\ & & J_s\end{bmatrix}=J \tag{3}$$

每个方块 J_i 有一个特征值 λ_i 与一个特征向量与位于**对角线上方**的 1's：

$$\textbf{乔丹方块}\qquad J_i=\begin{bmatrix}\lambda_i & 1 & & \\ & \bullet & \bullet & \\ & & \bullet & 1 \\ & & & \lambda_i\end{bmatrix} \tag{4}$$

如果他们共享相同的乔丹形式 J，则矩阵相似—否则就不相似。

乔丹形式对于每个丢失的特征向量有一个非对角线的 1 (1's 就在特征值旁边)，在每个相似矩阵的家族中，我们选择一个杰出的成员称为 J，它是近乎对角(可能的话是完全对角)，我们可以快速求解 $d\boldsymbol{u}/dt=J\boldsymbol{u}$ 然后取次方 J^k。这个家族其他的矩阵都有 BJB^{-1} 的形式。

乔丹定理在我的教科书“线性代数与应用”中有证明，请参考那本书(或是更进阶的书)的证明。这项推理相当错综复杂，乔丹形式在实际计算上也不是那么流行——它的计算不稳定，A 的一点轻微变化就会分离重复的特征值，并且移除非对角线的 1's ——乔丹变回对角 Λ。

无论证明与否，你已经捕捉到相似性的中心概念——在保留重要性质下，使得 A 尽可能简单。最佳基底 B 得到 $B^{-1}AB=J$。

问题 若 $A^2=$ 零矩阵，求出特征值以及所有可能的乔丹形式。

答案 特征值必须全部是零，因为 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 得到 $A^2\boldsymbol{x}=\lambda^2\boldsymbol{x}=0\boldsymbol{x}$。$A$ 的乔丹形式有 $J^2=0$，因为 $J^2=(B^{-1}AB)\,(B^{-1}AB)=B^{-1}A^2B=0$。$J$ 的每个方块有 $\lambda=0$ 在对角线，检视方块大小是 1, 2, 3 的 J_k^2：

$$[\,0\,]^2=[\,0\,]\qquad \begin{bmatrix}0&1\\0&0\end{bmatrix}^2=\begin{bmatrix}0&0\\0&0\end{bmatrix}\qquad \begin{bmatrix}0&1&0\\0&0&1\\0&0&0\end{bmatrix}^2=\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix}$$

结论：若 $J^2=0$ 则所有的方块大小必须是 1 或 2，J^2 在 3×3 时不是零。

J 的秩(与 A)会是 1's 的全部个数，**最大的秩是 $n/2$**，这个发生在有 $n/2$ 个方块，每个方块的大小是 2 而且秩是 1。

请尊重版权与译者的劳动成果，侵权必究!

现在应用数学最伟大的基底来了，他们的离散形式是 $\mathbf{R}^n$ 中的向量，连续形式是函数空间中的函数。由于他们是一劳永逸的选择，不需要知道矩阵 A，这些基底 $B_{\text{in}} = B_{\text{out}}$ 可能无法把 A 对角化，但是对应用数学的许多重要矩阵来说，矩阵 $B^{-1}AB$ 已经接近对角。【下方的 4 是接 421 页的 3】

4　$B_{\text{in}} = B_{\text{out}}$ = 傅里叶矩阵 F，则 Fx 是 x 的离散傅里叶变换。

这些字词告诉我们：方程式(6)的傅里叶矩阵有$(1, \lambda, \lambda^2, \lambda^3)$的列是非常重要，这些是容易处理的优质基底。

我们会问：哪些矩阵被 F 对角化？这次我们从特征向量$(1, \lambda, \lambda^2, \lambda^3)$开始，然后找出拥有这些特征向量的矩阵：

$$\text{若}\lambda^4 = 1 \text{ 则} \quad P\boldsymbol{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda \boldsymbol{x} \tag{5}$$

P 是排列矩阵，方程式 $P\boldsymbol{x} = \lambda\boldsymbol{x}$ 说明 $\boldsymbol{x}$ 是特征向量而且λ是 P 的特征值。注意向量方程式的第 4 行是如何变成 $1 = \lambda^4$，这个λ的规则使得所有事情有效。

这个给出了 4 个不同的特征值λ? 是的！这 4 个数字$\lambda = 1, i, -1, -i$ 全部都满足 $\lambda^4 = 1$。(你知道 $i^2 = -1$，两边平方得到 $i^4 = 1$。) 所以这 4 个数是 P 的特征值，每个特征值对应它的特征向量 $\boldsymbol{x} = (1, \lambda, \lambda^2, \lambda^3)$。**特征向量矩阵 F 对角化排列矩阵 P：**

$$\textbf{特征值矩阵}\Lambda \quad \begin{bmatrix} 1 & & & \\ & i & & \\ & & -1 & \\ & & & -i \end{bmatrix} \qquad \textbf{固有向量矩阵是傅里叶矩阵 } F \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & i^2 & 1 & (-i)^2 \\ 1 & i^3 & -1 & (-i)^3 \end{bmatrix} \tag{6}$$

这些 F 的列是正交，因为他们是 P 的特征向量(一个正交矩阵)，不幸的是傅里叶矩阵 F 是复数(它是世界上最重要的复数矩阵)，利用快速傅里叶变换，可以很快速的执行乘法 $F\boldsymbol{x}$ 百万次。FFT 在段落 9.3。

关键问题：除了 P 之外的其他矩阵有相同的特征向量矩阵 F 吗？我们知道 P^2 与 P^3 与 P^4 有与 P 相同的特征值，相同的的矩阵 F 可以对角化所有 P 的次方，并且 P^2 与 P^3 与 P^4 的特征值是数字λ^2 与λ^3 与λ^4。例如 $P^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$：

$$P^2\boldsymbol{x} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda^2 \begin{bmatrix} 1 \\ \lambda \\ \lambda^2 \\ \lambda^3 \end{bmatrix} = \lambda^2\boldsymbol{x}, \text{ 当 } \lambda^4 = 1$$

4 次方是特殊情况，因为 $P^4 = I$。当我们进行 4 次的“循环排列”，$P^4\boldsymbol{x}$ 与我们开始的 $\boldsymbol{x}$ 相同，$P^4 = I$ 的特征值就是 1, 1, 1, 1，而且这个数字 1 与 P 的所有特征值的 4 次方是一致的：$1^4 = 1$ 且 $i^4 = 1$ 且$(-1)^4 = 1$ 且$(-i)^4 = 1$。

多一个步骤带来更多的矩阵，若 P 与 P^2 与 P^3 与 $P^4 = I$ 有相同的特征向量矩阵 F，则任意的组合 $C = c_1P + c_2P^2 + c_3P^3 + c_0I$ 也有相同的 F：

$$\textbf{循环矩阵 } \boldsymbol{C} = \begin{bmatrix} c_0 & \boldsymbol{c_1} & c_2 & c_3 \\ c_3 & c_0 & \boldsymbol{c_1} & c_2 \\ c_2 & c_3 & c_0 & \boldsymbol{c_1} \\ \boldsymbol{c_1} & c_2 & c_3 & c_0 \end{bmatrix} \quad \begin{array}{l} \text{有固有向量在傅里叶矩阵 } F \\ \text{有 4 个固有值 } c_0 + c_1\lambda + c_2\lambda^2 + c_3\lambda^3 \\ \text{来自 4 个数字 } \lambda = 1,\ i,\ -1,\ -i \\ \text{来自 } \lambda = 1 \text{ 的固有值是 } c_0 + c_1 + c_2 + c_3 \end{array}$$

这是一大步，我们发现所有的矩阵(循环矩阵 C)的特征向量是 F 中的傅里叶向量，我们也知道 C 的 4 个特征值，但是直到现在我们还没有给出一个公式或是一个名称：

$$\begin{array}{l} \boldsymbol{C} \text{的四个} \\ \text{固有值来自} \\ \text{傅里叶转换 } \boldsymbol{Fc} \end{array} \qquad \boldsymbol{Fc} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{array}{l} c_0 + c_1 + c_2 + c_3 \\ c_0 + ic_1 - c_2 - ic_3 \\ c_0 - c_1 + c_2 - c_3 \\ c_0 - ic_1 - c_2 + ic_3 \end{array}$$

范例 2　相同的概念一样适用于傅里叶矩阵 F 与一个任意大小的循环矩阵 C，2×2 的矩阵看起来很平凡但是他们很有用。现在 P 的特征值是$\lambda^2 = 1$ 而不是$\lambda^4 = 1$，而且不需要复数 i：$\lambda = \pm 1$。

$$\begin{array}{l} \text{傅里叶矩阵 } F \text{ 来自} \\ P \text{ 与 } C \text{ 的固有向量} \end{array} \quad F = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{array}{c} \text{循环} \\ c_0I + c_1P \end{array} \quad C = \begin{bmatrix} c_0 & c_1 \\ c_1 & c_0 \end{bmatrix}$$

C 的特征值是 $c_0 + c_1$ 与 $c_0 - c_1$，当向量 $\boldsymbol{c}$ 是(c_0, c_1)，这些特征值来自傅里叶变换 $\boldsymbol{Fc}$，变换 $\boldsymbol{Fc}$ 给出任意大小的 C 的特征值。

注意**循环矩阵有常数对角线**，相同的数字 c_0 沿着主对角线，数字 c_1 在对角线上方，而且包围或是环绕对角线直到 C 的东南角，这说明了“循环”的名字，它可以指出矩阵是周期性或是循环性。甚至λ的次方也是循环的，因为$\lambda^4 = 1$ 得到λ^5, λ^6, λ^7, $\lambda^8 = \lambda$, λ^2, λ^3, λ^4。

常数沿着对角线是 C 的一个主要性质，它对应微分方程式的常数系数，这正是傅里叶的完美工作！

方程式　　$d^2u/dt^2 = -u$ 的解是 $u = c_0 \cos t + c_1 \sin t$

方程式　　$d^2u/dt^2 = tu$ 无法由基础函数求解

这些方程式是线性，第一个是简单弹簧的震荡方程式，它是牛顿定律 $f = ma$，其中质量 $m = 1$，$a = d^2u/dt^2$，以及力 $f = -u$。常数系数产生你实际可以求解的微分方程式。

请尊重版权与译者的劳动成果，侵权必究！

方程式 $u'' = tu$ 有一个变数系数 t，这是物理学与光学的“埃里(Airy)”方程式(它的推导可以解释彩虹)，当 t 通过零，解会完全改变，这些解需要无限级数。在此我们不讨论。

重点是常数系数的方程式有简单解 $e^{\lambda t}$，你把 $e^{\lambda t}$ 代入微分方程式求出λ，这个数字λ就好像特征值。对于 $u = \cos t$ 与 $u = \sin t$，这个数字就是$\lambda = i$。欧拉伟大的公式 $e^{it} = \cos t + i \sin t$ 引入复数，就好像我们在 P 与 C 看到的特征值一样。

函数空间的基底

对于 x 的函数，第一个基底我会想到包含次方的 $1, x, x^2, x^3, \ldots$，不幸的是，这是一组很糟糕的基底。这些函数 x^n 只是刚好无关，x^{10} 几乎就是其他基底向量 $1, x, \ldots, x^9$ 的组合，实际上不可能使用这些劣质的“病态”基底去计算。

如果我们有的是向量而不是函数，对于优质基底的测试要检视 $B^{\mathrm{T}}B$，这个矩阵包含基底向量(B 的列)之间的全部内积，当 $B^{\mathrm{T}}B = I$，基底是正交单位，这是最好的可能。但是基底 $1, x, x^2, \ldots$产生邪恶的**希尔伯特矩阵**：$B^{\mathrm{T}}B$ 的最大特征值与最小特征值之间的比值超级巨大，一个巨大的**条件数**(condition number)表示一个不快乐的基底选择。

注解　现在 B 的列是函数而不是向量，我们仍然使用 $B^{\mathrm{T}}B$ 去测试无关性质，所以我们需要知道两个函数的点积(内积是比较好的名字)--这些是 $B^{\mathrm{T}}B$ 中的数字。

向量的点积就是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = x_1y_1 + \ldots + x_ny_n$，函数的内积使用积分而不是相加，但是概念是完全适用：

$$\text{内积}(\boldsymbol{f}, \boldsymbol{g}) \quad = \quad \int f(x)\, g(x)\, dx$$

$$\text{复数内积}(\boldsymbol{f}, \boldsymbol{g}) \quad = \quad \int \overline{f(x)}\, g(x)\, dx\,, \quad \overline{f} \ = \text{共轭复数}$$

$$\text{加权内积}(\boldsymbol{f}, \boldsymbol{g})_w \quad = \quad \int w(x)\overline{f(x)}\, g(x)\, dx\,, \quad w = \text{加权函数}$$

当积分从 $x = 0$ 到 $x = 1$，x^i 与 x^j 的内积是

$$\int_0^1 x^i\, x^j dx = \frac{x^{i+j+1}}{i+j+1}\bigg|_{x=0}^{x=1} = \frac{1}{i+j+1} = \text{希尔伯特矩阵 } B^{\mathrm{T}}B \text{ 的单元}$$

借由变成 $x = -1$ 到 $x = 1$ 对称的区间，我们立刻得到全部偶函数与全部奇函数的正交性：

$$\textbf{区间 [–1, 1]} \qquad \int_{-1}^{1} x^2\, x^5 dx = 0 \qquad \int_{-1}^{1} \mathbf{even}(x)\,\mathbf{odd}(x)\, dx = 0$$

这项改变使得一半的基底函数与另一半的基底函数正交。继续使用对称区间–1 到 1 (或是$-\pi$ 到π)是如此简单，但是我们想要一组比次方项 x^n 更好的基底--希望是正交基底。

函数空间的正交基底

下列是针对理论与计算的三个领先的偶-奇基底：

5. 傅里叶基底	$1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots.$
6. 勒让德基底	$1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x, \ldots$
7. 契比雪夫基底	$1, x, 2x^2 - 1, 4x^3 - 3x, \ldots$

傅里叶基底函数(正弦与余弦)全部都是周期性，他们每隔 2π 的区间一再重复，这是因为 $\cos(x + 2\pi) = \cos x$ 与 $\sin(x + 2\pi) = \sin x$。所以这些基底对于函数 $f(x)$特别的好，使得他们本身也是周期性：$f(x + 2\pi) = f(x)$。

这组基底也是正交，每个正弦与余弦跟其他的正弦与余弦都是正交，我们当然不期待基底函数 $\cos nx$ 与本身正交。

最重要的是**正弦-余弦**基底对于**近似**也是特别的棒，如果我们有一个平滑的周期函数，则一些正弦与余弦(低频率)是我们全部需要的。$f(x)$的突跳以及信号中的杂讯会在较高频率看到(较大的 n)，我们希望并且期待信号不会被杂讯淹没。

傅里叶变换把 $f(x)$与自己傅里叶级数中的系数 a_k 与 b_k 产生关联：

傅里叶级数 $f(x) = a_0 + b_1 \sin x + a_1 \cos x + b_2 \sin 2x + a_2 \cos 2x + \ldots$

我们看到**函数空间是无限维度**，它利用无限多的基底函数去完美捕捉一个典型函数 $f(x)$，但是每个系数的公式(例如 a_3)就像向量 $\boldsymbol{b}$ 在通过 $\boldsymbol{a}$ 的直线的投影一样的 $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{a}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$。

现在我们把函数 $f(x)$投影在函数空间中通过 $\cos 3x$ 的直线：

$$\textbf{傅里叶系数}\quad a_3 = \frac{(f(x),\ \cos 3x)}{(\cos 3x,\ \cos 3x)} = \frac{\int f(x)\cos 3x\, dx}{\int \cos 3x \cos 3x\, dx} \tag{7}$$

范例 3　三角学里面的二倍角公式是 $\cos 2x = 2\cos^2 x - 1$，这个告诉我们 $\cos^2 x = ½ + ½ \cos 2x$，一个非常短的傅里叶级数，所以 $\sin^2 x = ½ - ½ \cos 2x$。

傅里叶级数就是函数空间的线性代数。我会在第十章有关应用的章节作为亮点，好好的说明。

请尊重版权与译者的劳动成果，侵权必究！

勒让德多项式与契比雪夫多项式

勒让德多项式是应用格莱姆-施密特概念的结果(段落 4.4)，计划是正交化这些次方项 1, x, x^2, ...。首先在–1 到 1 的区间，奇函数 x 已经与偶函数 1 正交，他们的乘积$(x)(1) = x$ 的积分是零。但是 x^2 与 1 的内积是 $\int x^2 dx = 2/3$：

$$\frac{(x^2,\ 1)}{(1,\ 1)} = \frac{\int x^2\,dx}{\int 1\,dx} = \frac{2/3}{2} = \frac{1}{3} \qquad \text{格莱姆-施密特得到 } \boldsymbol{x^2 - \frac{1}{3} =} \text{ 勒让德}$$

同理奇数次方 x^3 在奇函数 x 的方向上有一个分量 $3x/5$:

$$\frac{(x^3,\ x)}{(x,\ x)} = \frac{\int x^4\,dx}{\int x^2\,dx} = \frac{2/5}{2/3} = \frac{3}{5} \qquad \text{格莱姆-施密特得到 } \boldsymbol{x^3 - \frac{3}{5}x =} \text{ 勒让德}$$

继续对 $x^4, x^5, \ldots$ 使用格莱姆-施密特，得到每个勒让德函数––一组优质基底。

最终我们转移到契比雪夫多项式 $1, x, 2x^2 - 1, 4x^3 - 3x$, 他们不是来自格莱姆-施密特，而是与 $1, \cos\theta, \cos 2\theta, \cos 3\theta$ 有关，这个给出了巨大的计算优势––我们可以使用**快速傅里叶变换**。当我们令 $\boldsymbol{x = \cos\theta}$, 会出现契比雪夫与傅里叶的关联：

契比雪夫	$\boldsymbol{2x^2 - 1 = 2(\cos\theta)^2 - 1 = \cos 2\theta}$
到傅里叶	$\boldsymbol{4x^3 - 3x = 4(\cos\theta)^3 - 3(\cos\theta) = \cos 3\theta}$

第 n^{th} 次的契比雪夫多项式 $\boldsymbol{T_n(x)}$变换成傅里叶的 $\boldsymbol{\cos n\theta = T_n(\cos\theta)}$。

注解　这些多项式是一个大型软件计划“**chebfun**”的基础，每个函数 $f(x)$被一个超级精确的**契比雪夫近似**取代，你可以积分 $f(x)$，可以求解 $f(x) = 0$，可以求出极大值与极小值。不止如此，你可以求解关于 $f(x)$的微分方程式––快速又高度精确。

当 **chebfun** 利用多项式取代了 $f(x)$，你已经准备好解决问题了。

主要观念的复习

1. 若矩阵 B 是良好条件(well-conditioned)，就是好基底。正交基底是最佳。
2. 若$\Lambda = B^{-1}AB$ 是对角矩阵，也是好基底，但是乔丹形式 J 非常不稳定。
3. 傅里叶矩阵可以对角化常数系数周期方程式：完美。
4. 基底 $1, x, x^2, \ldots$ 得到 $B^{\mathrm{T}}B =$ 希尔伯特矩阵：糟糕的计算。
5. 勒让德与契比雪夫多项式是函数空间的极佳基底。

问题集 8.3

1 范例 1 中，$J-3I$ 的秩是多少？它的零空间的维度是多少？这个维度给出了 $\lambda=3$ 对应的无关特征向量的个数。

代数倍数值是 2，因为 $\det(J-\lambda I)$ 有重复因子 $(\lambda-3)^2$。几何倍数值是 1，因为只有一个无关的特征向量。

2 这些矩阵 A_1 及 A_2 与 J 相似，求解 $A_1B_1=B_1J$ 与 $A_2B_2=B_2J$，找出基底矩阵 B_1 与 B_2，其中 $J=B_1^{-1}A_1B_1$ 与 $J=B_2^{-1}A_2B_2$。

3 方块转置 J^{T} 与 J 有相同的三重特征值 2(只有一个特征向量)，求出基底变换矩阵 B 使得 $J=B^{-1}J^{\mathrm{T}}B$(表示 $BJ=J^{\mathrm{T}}B$)：

$$J=\begin{bmatrix}2&1&0\\0&2&1\\0&0&2\end{bmatrix}\qquad J^{\mathrm{T}}=\begin{bmatrix}2&0&0\\1&2&0\\0&1&2\end{bmatrix}$$

4 J 与 K 是有相同的零特征值与相同的秩 2 的乔丹形式，但是请证明不存在可逆的 B 可以求解 $BK=JB$，所以 K 与 J 不相似：

$$J=\begin{bmatrix}0&1&0&0\\&0&0&0\\&&0&1\\&&&0\end{bmatrix}\qquad K=\begin{bmatrix}0&1&0&0\\&0&1&0\\&&0&0\\&&&0\end{bmatrix}$$

5 若 $A^3=0$ 显示所有的 $\lambda=0$，以及所有的乔丹方块有 $J^3=0$，且大小是 1 或 2 或 3，所以 $\operatorname{rank}(A)\le 2n/3$。若 $A^n=0$，为什么 $\operatorname{rank}(A)\le n$。

6 证明 $\boldsymbol{u}(t)=\begin{bmatrix}te^{\lambda t}\\e^{\lambda t}\end{bmatrix}$ 可以求解 $\dfrac{d\boldsymbol{u}}{dt}=J\boldsymbol{u}$，其中 $J=\begin{bmatrix}\lambda&1\\0&\lambda\end{bmatrix}$ 且 $\boldsymbol{u}(0)=\begin{bmatrix}0\\1\end{bmatrix}$

7 证明差方程式 $v_{k+2}-2\lambda v_{k+1}+\lambda^2 v_k=0$ 的解是 $v_k=\lambda^k$ 以及 $v_k=k\lambda^k$。这些对应到问题 6 的 $e^{\lambda t}$ 与 $te^{\lambda t}$。

8 $\lambda^3=1$ 的 3 个解是什么？他们是复数 $\lambda=\cos\theta+i\sin\theta=e^{i\theta}$。则 $\lambda^3=e^{3i\theta}=1$，其中角度 3θ 是 0 或 2π 或 4π。写出 3×3 傅里叶矩阵，其中列是 $(1,\lambda,\lambda^2)$。

9 从问题 8，确认任意 3×3 的循环 C 矩阵都有特征向量 $(1,\lambda,\lambda^2)$。若你的矩阵 C 的对角线包含 c_0,c_1,c_2，则它的特征值在 $F\boldsymbol{c}$ 中。

10 使用公式(7)求出 $f(x)=\begin{cases}1, & -L\le x\le L\\0, & L\le|x|\le 2\pi\end{cases}$ 的傅里叶级数中的 $a_3\cos 3x$。

请尊重版权与译者的劳动成果，侵权必究！

第九章
复数向量与矩阵

实数对比复数

$\mathbf{R}=$ 所有实数的线 $-\infty<x<\infty$ $\leftrightarrow$ $\mathbf{C}=$ 所有复数 $z=x+iy$ 的平面

$|x|=x$ 的绝对值 $\leftrightarrow$ $|z|=\sqrt{x^2+y^2}=r=(z$ 的绝对值$)$ (或模量)

1 与 -1 是 $x^2=1$ 的解 $\leftrightarrow$ $z=1, w, ..., w^{n-1}$ 是 $z^n=1$ 的解，其中 $w=e^{2\pi i/n}$

$z=x+iy$ 的**共轭复数**是 $\bar{z}=\boldsymbol{x-iy}$，$|z|^2=x^2+y^2=z\bar{z}$ 且 $\dfrac{1}{z}=\dfrac{\bar{z}}{|z|^2}$

$z=x+iy$ 的**极形式**是 $|z|e^{i\theta}=re^{i\theta}=r\cos\theta+ir\sin\theta$，角度是 $\tan\theta=y/x$

$\mathbf{R}^n$：向量有 n 个实数分量	$\leftrightarrow$	$\mathbf{C}^n$：向量有 n 个复数分量
长度：$\|\boldsymbol{x}\|^2=x_1{}^2+...+x_n{}^2$	$\leftrightarrow$	长度：$\|\boldsymbol{z}\|^2=\|z_1\|^2+...+\|z_n\|^2$
转置：$(A^{\mathrm{T}})_{ij}=A_{ji}$	$\leftrightarrow$	共轭转置：$(A^{\mathrm{H}})_{ij}=\overline{A_{ji}}$
点积：$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=x_1y_1+...+x_ny_n$	$\leftrightarrow$	内积：$\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}=\bar{u}_1v_1+...+\bar{u}_nv_n$
A^{T} 的思考：$(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$	$\leftrightarrow$	A^{H} 的思考：$(A\boldsymbol{u})^{\mathrm{H}}\boldsymbol{v}=\boldsymbol{u}^{\mathrm{H}}(A^{\mathrm{H}}\boldsymbol{v})$
正交性：$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0$	$\leftrightarrow$	正交性：$\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}=0$
对称矩阵：$S=S^{\mathrm{T}}$	$\leftrightarrow$	埃尔米特矩阵：$S=S^{\mathrm{H}}$
$\boldsymbol{S=Q\Lambda Q^{-1}=Q\Lambda Q^{\mathrm{T}}}$ (实数Λ)	$\leftrightarrow$	$\boldsymbol{S=U\Lambda U^{-1}=U\Lambda U^{\mathrm{H}}}$ (实数Λ)
反对称矩阵：$K^{\mathrm{T}}=-K$	$\leftrightarrow$	反埃尔米特矩阵：$K^{\mathrm{H}}=-K$
正交矩阵：$Q^{\mathrm{T}}=Q^{-1}$	$\leftrightarrow$	酉矩阵：$U^{\mathrm{H}}=U^{-1}$
正交单位列：$Q^{\mathrm{T}}Q=I$	$\leftrightarrow$	正交单位列：$U^{\mathrm{H}}U=I$
$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 且 $\|Q\boldsymbol{x}\|=\|\boldsymbol{x}\|$	$\leftrightarrow$	$(U\boldsymbol{x})^{\mathrm{H}}(U\boldsymbol{y})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$ 且 $\|U\boldsymbol{z}\|=\|\boldsymbol{z}\|$

线性代数完整的表述必须包含复数 $z=x+iy$，纵使矩阵是实数，**特征值与特征向量通常是复数**。例子：一个 2×2 的旋转矩阵具有复数特征向量 $\boldsymbol{x}=(1, i)$与 $\bar{\boldsymbol{x}}=(1, -i)$。我会在段落 9.1 与 9.2 整理这些难忘的字词：当你转置一个向量或是一个矩阵 A，取每个单元的共轭(***i* 变成–*i***)。段落 9.3 是有关复数最重要的矩阵——**傅里叶矩阵 *F***。

9.1 复数

从虚数 i 开始。每个人都知道 $x^2 = -1$ 没有实数解。当你平方一个实数，答案不可能是负数，所以这个世界同意一个解称为 i。(除了电气工程师称它 j。) 虚数遵守正常加法与乘法的规则，存在一个不同处，**用–1 取代 i^2**。

这个段落给出复数的主要结果，对某些学生来说是复习，对所有人是一种参考。每件事情都来自 $i^2 = -1$ 与 $e^{2\pi i} = 1$。

一个复数(比如 $3 + 2i$)**是一个实数**(3)**加一个虚数**($2i$)。加法时实数部分(**以后简称实部**)与虚数部分(**以后简称虚部**)分开，乘法使用 $i^2 = -1$：

$$\textbf{加：}(3 + 2i) + (3 + 2i) = 6 + 4i$$

$$\textbf{乘：}(3 + 2i)(1 - i) = 3 + 2i - 3i - 2i^2 = 5 - i$$

如果我把 $3 + i$ 加上 $1 - i$，答案是 4，实数 3 + 1 与虚数 $i - i$ 分开计算，我们是把向量(3, 1)与(1, –1)相加得到(4, 0)。

数字$(1 + i)^2$ 是$(1 + i)$乘$(1 + i)$，运算规则得到惊讶的答案 $2i$：

$$(1 + i)(1 + i) = 1 + i + i + i^2 = 2i$$

在复数平面上，$1 + i$ 在 45°角，它很像向量(1, 1)，当我们把$(1 + i)$平方得到 $2i$，角度变成 2 倍来到 90°，如果我们再平方，答案是$(2i)^2 = -4$，两倍 90°角来到 180°，这个方向是负数实数。

一个实数就是一个虚部为零的复数 $z = a + bi$：$b = 0$。

实部是 $a = \mathrm{Re}(a + bi)$　　**虚部**是 $b = \mathrm{Im}(a + bi)$

复数平面

复数对应平面的一个点，实数沿着 x 轴，纯虚数在 y 轴，**复数 $3 + 2i$ 是坐标(3, 2)的点**。数字零，就是 $0 + 0i$，在原点。

加减复数就好像加减平面的向量一样，实数分量与虚数分量维持分开，向量还是平常的**头至尾**。除了复数相乘之外，复数平面$\mathbf{C}^1$就好像平常的二维平面$\mathbf{R}^2$，只是我们不做向量相乘。

重要概念：**$3 + 2i$ 的共轭复数是 $3 - 2i$**，$z = 1 - i$ 的共轭复数是 $\bar{z} = 1 + i$。一般来说 $z = a + bi$ 的共轭复数是 $\bar{z} = a - bi$。(有些作者在数字上加横杠，有些作者使用星号*：$\bar{z} = z^*$。) z 与 $\bar{z}$ 的虚部是相反符号，在复数平面上，$\bar{z}$ 是 z 在实数轴的另一侧的影像。

请尊重版权与译者的劳动成果，侵权必究!

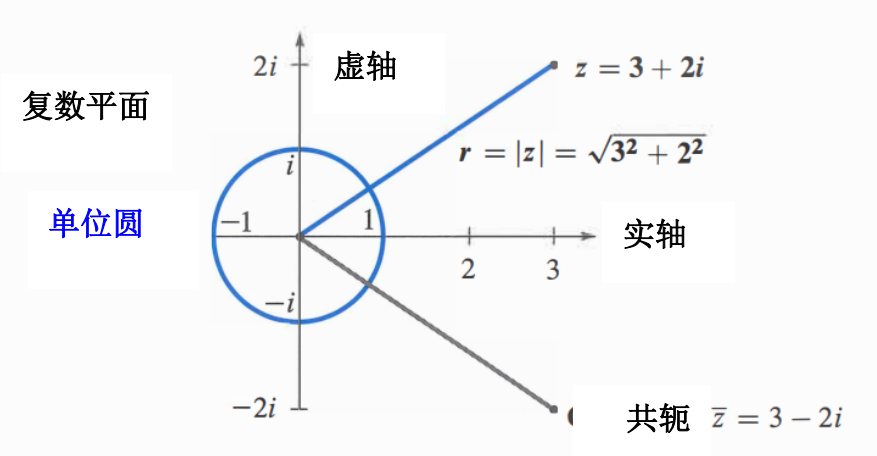


图 9.1：数字 $z = a + bi$ 对应到点(a, b)以及向量$\begin{bmatrix} a \\ b \end{bmatrix}$。

两个有用的事实。**当我们把共轭 $\bar{z}_1$ 与 $\bar{z}_2$ 相乘，我们得到 z_1z_2 的共轭**。并且当我们把 $\bar{z}_1$ 与 $\bar{z}_2$ **相加**，我们得到 $z_1 + z_2$ 的共轭：

$\bar{z}_1 + \bar{z}_2 = (3 - 2i) + (1 + i) = \mathbf{4 - i}$，这是 $z_1 + z_2 = \mathbf{4 + i}$ 的共轭。

$\bar{z}_1 \times \bar{z}_2 = (3 - 2i) \times (1 + i) = \mathbf{5 + i}$，这是 $z_1 \times z_2 = \mathbf{5 - i}$ 的共轭。

加法与乘法是线性代数的必需。当 A 是实数，借由选取 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 的共轭，我们得到另一个特征值 $\bar{\lambda}$ 以及它的特征向量 $\bar{\boldsymbol{x}}$：

$$\text{特征值}\quad \lambda \text{ 与 } \bar{\lambda} \qquad \text{若 } A\boldsymbol{x} = \lambda\boldsymbol{x} \text{ 且 } A \text{ 是实数，则 } A\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}} \tag{1}$$

当 $z = 3 + 2i$ 与自己的共轭$\bar{z} = 3 - 2i$ 结合时会产生特殊的事情，两者相加 $z + \bar{z}$ 或相乘 $z\bar{z}$ 的结果永远是实数：

$z + \bar{z} =$ **实数**　　$(3 + 2i) + (3 - 2i) = 6$　(实数)

$z\bar{z} =$ **实数**　　$(3 + 2i) \times (3 - 2i) = 9 + 6i - 6i - 4i^2 = 13$ (实数)

$z = a + bi$ 与自己的共轭$\bar{z} = a - bi$ 的总和是实数 $2a$，z 乘 $\bar{z}$ 的乘积是实数 $a^2 + b^2$：

$$z \text{ 乘 } \bar{z} \text{ 得到} |z|^2 = r^2 \qquad (a + bi)(a - bi) = a^2 + b^2 \tag{2}$$

下个步骤的复数是 $1/z$。如何除以 $a + ib$？最佳的概念是先乘$\bar{z}/\bar{z} = 1$，在分母产生 $z\bar{z}$，这是 $a^2 + b^2$：

$$\frac{1}{a+ib} = \frac{1}{a+ib}\frac{a-ib}{a-ib} = \frac{a-ib}{a^2+b^2} \qquad \frac{1}{3+2i} = \frac{1}{3+2i}\frac{3-2i}{3-2i} = \frac{3-2i}{13}$$

如果 $a^2 + b^2 = 1$，就是说$(a + ib)^{-1}$ 是 $a - ib$。**在单位圆上，$1/z = \bar{z}$**。稍后我们会说：$1/e^{i\theta}$是$e^{-i\theta}$。使用距离 r 与角度θ去乘与除。

a^2+b^2 的平方根是$|z|$，这是数字 $z=a+ib$ 的**绝对值**(或是**模量**)，平方根$|z|$也写成 r，因为它是从 0 到 z 的距离。**极形式的实数 r 给出复数 z 的大小：**

$z=a+ib$ 的绝对值 是 $|z|=\sqrt{a^2+b^2}$， 称为 r

$z=3+2i$ 的绝对值 是 $|z|=\sqrt{3^2+2^2}$， 这是 $r=\sqrt{13}$

极形式的另一个部分是角度θ，$z=5$ 的角度是$\theta=0$(因为 z 是正的实数)，$z=3i$ 的角度是$\pi/2$ 弧度(radian)，负数 $z=-9$ 的角度是π弧度，**数字取平方时角度变成两倍**。极形式在复数乘法非常棒(加法就不好了)。

当距离是 r 且角度是θ，三角学给出三角形的另外两个边。实部(沿着底部)是 $a=r\cos\theta$，虚部(上或下)是 $b=r\sin\theta$，两者放在一起，矩形形式变成极形式 $re^{i\theta}$。

数字 $z=a+ib$ 也是 $z=r\cos\theta+ir\sin\theta$，这是 $re^{i\theta}$。

注解：**$\cos\theta+i\sin\theta$ 有绝对值 $r=1$，因为 $\cos^2\theta+\sin^2\theta=1$**。于是 $\cos\theta+i\sin\theta$ 落在半径 1 的圆上——单位圆。

范例 1 求 $z=1+i$ 与 $\bar{z}=1-i$ 的 r 与θ 。

解 z 与 $\bar{z}$ 有相同的绝对值，它是 $r=\sqrt{1+1}=\sqrt{2}$：

$$|z|^2=1^2+1^2=2 \quad \text{同时} \quad |\bar{z}|^2=1^2+(-1)^2=2$$

从圆心出发的距离是 $r=\sqrt{2}$。角度θ 又是多少？数字 $1+i$ 是复数平面的点(1, 1)，到这个点的角度是$\pi/4$ 或是 45°，余弦是 $1/\sqrt{2}$，正弦也是 $1/\sqrt{2}$，结合 r 与θ 可以带回 $z=1+i$：

$$r\cos\theta+ir\sin\theta=\sqrt{2}\left(\frac{1}{\sqrt{2}}\right)+i\sqrt{2}\left(\frac{1}{\sqrt{2}}\right)=1+i$$

到达共轭 $1-i$ 的角度可以是正也可以是负，我们可以走 $7\pi/4$ 这是 315°，或者我们可以反向走负的角度$-\pi/4$ 弧度或是–45°。**如果 z 在角度θ，它的共轭 $\bar{z}$ 就会在 $2\pi-\theta$，同时也是$-\theta$。**

我们可以自由的加 2π 或 4π 或-2π 到任意的角!这些角度走完整个圆所以终点是相同的，说明了为什么θ 有无限多的选择。通常我们选择角度在 0 与 2π 之间，但是$-\theta$ 对于共轭 $\bar{z}$ 非常有用，而且 $1=e^0=e^{2\pi i}$。

请尊重版权与译者的劳动成果，侵权必究!

次方与乘积：极形式

计算$(1+i)^2$与$(1+i)^8$在极形式很快速，这个形式有$r=\sqrt{2}$与$\theta=\pi/4$ (或 45°)。如果我们把绝对值平方得到$r^2=2$，角度变成两倍得到$2\theta=\pi/2$ (或 90°)，我们得到$(1+i)^2$。对于 8 次方我们需要r^8与8θ：

$$\mathbf{(1+i)^8} \qquad r^8=2\cdot 2\cdot 2\cdot 2=16 \text{ 以及 } 8\theta=2\pi$$

这表示：$(1+i)^8$有绝对值 16 以及角度2π，所以$(1+i)^8=16$。

次方在极形式很简单，复数乘法也是一样：

$z=r(\cos\theta+i\sin\theta)$ 的 **n 次方**是 $\boldsymbol{z^n=r^n(\cos n\theta+i\sin n\theta)}$ (3)

上述情况是z乘自己。要得到z乘z'，**各自的r相乘以及角度相加**：

$$r(\cos\theta+i\sin\theta) \text{ 乘 } r'(\cos\theta'+i\sin\theta')=\boldsymbol{rr'}(\cos(\boldsymbol{\theta+\theta'})+i\sin(\boldsymbol{\theta+\theta'})) \qquad (4)$$

有一个方法是利用三角学来了解，为什么对于z^2我们得到两倍角2θ？

$$(\cos\theta+i\sin\theta)\times(\cos\theta+i\sin\theta)=\cos^2\theta+i^2\sin^2\theta+2i\sin\theta\cos\theta$$

实部$\cos^2\theta-\sin^2\theta$是$\cos 2\theta$，虚部$2\sin\theta\cos\theta$是$\sin 2\theta$，这些是二倍角公式，他们显示$z$的$\theta$变成$z^2$的$2\theta$。

还有第二种了解z^n规则的方法，它使用了本段落最惊艳的公式。记得$\cos\theta+i\sin\theta$的绝对值是 1，余弦是由偶数次方构成，从$1-\frac{1}{2}\theta^2$开始。正弦是由奇数次方构成，从$\theta-\frac{1}{6}\theta^3$开始。美丽的事实是$e^{i\theta}$结合两个级数变成$\cos\theta+i\sin\theta$：

$$e^x=1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\ldots \qquad \text{变成} \qquad e^{i\theta}=1+i\theta+\frac{1}{2}i^2\theta^2+\frac{1}{6}i^3\theta^3+\ldots$$

把i^2写成–1 看到$1-\frac{1}{2}\theta^2$，**复数$e^{i\theta}$是$\cos\theta+i\sin\theta$**：

欧拉(Euler)公式 $\boldsymbol{e^{i\theta}=\cos\theta+i\sin\theta}$ 得到 $\boldsymbol{z=r\cos\theta+ir\sin\theta=re^{i\theta}}$ (5)

特殊的选择$\theta=2\pi$得到$\cos 2\pi+i\sin 2\pi=1$，无限级数$e^{2\pi i}=1+2\pi i+\frac{1}{2}(2\pi i)^2+\ldots$的总和是 1。

现在计算$e^{i\theta}$乘$e^{i\theta'}$，因为指数相加所以角度也相加：

e^2乘e^3是e^5 $e^{i\theta}$乘$e^{i\theta}$是$e^{2i\theta}$ $e^{i\theta}$乘$e^{i\theta'}$是$e^{i(\theta+\theta')}$

次方$(re^{i\theta})^n$等于$r^n e^{in\theta}$，当 $r = 1$ 以及 $r^n = 1$ 时，他们在单位圆上。我们可以找到 n 个不同的数字，他们的 n 次方都会等于 1：

令 $w = e^{2\pi i/n}$，则 $1, w, w^2, ..., w^{n-1}$ 的 n 次方全部等于 1。

这些是“**1 的 n 次方根**”，他们是方程式 $z^n = 1$ 的解，他们是图 9.2b 单位圆上的等距点，其中整个 2π 除以 n。他们的角度乘 n 得到 n 次方的角，就是 $w^n = e^{2\pi i}$，这就是 1。同时$(w^2)^n = e^{4\pi i} = 1$。上述的每个数字取 n 次方都会环绕圆来到 1。

1 的 n 次方根是信号处理的关键，离散傅里叶转换使用 $w = e^{2\pi i/n}$以及它的次方。段落 9.3 展示如何利用快速傅里叶转换分解一个向量(信号)成为 n 个频率。

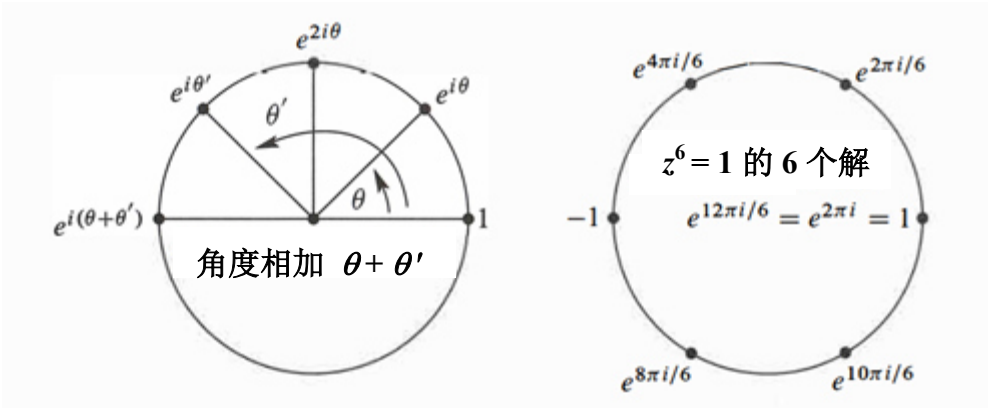


图 9.2：(a) $e^{i\theta}$乘 $e^{i\theta'}$ 是$e^{i(\theta+\theta')}$ (b) $e^{2\pi i/n}$的 n 次方是 $e^{2\pi i} = 1$。

主要观念的复习

1. 把 $a + ib$ 加到 $c + id$ 就好像$(a, b) + (c, d)$。使用 $i^2 = -1$ 去乘。
2. $z = a + bi = re^{i\theta}$的共轭是 $\bar{z} = z^* = a - bi = re^{-i\theta}$。
3. z 乘 $\bar{z}$ 是 $re^{i\theta}$乘 $re^{-i\theta}$，就是 $r^2 = |z|^2 = a^2 + b^2$ (实数)。
4. 次方与乘积在极形式 $z = re^{i\theta}$非常容易。乘 r's 以及加θ's。

请尊重版权与译者的劳动成果，侵权必究!

问题集 9.1

问题 1-8 有关复数的运算。

1 对下列配对的复数乘并且加：

(a) $2+i$，$2-i$　(b) $-1+i$，$-1+i$　(c) $\cos\theta+i\sin\theta$，$\cos\theta-i\sin\theta$

2 在复数平面标示位置，有需要的话做简化：

(a) $2+i$　(b) $(2+i)^2$　(c) $\dfrac{1}{2+i}$　(d) $|2+i|$

3 求出下列 4 个数的绝对值 $r=|z|$，若 θ 是 $6-8i$ 的角，其他 3 个数的角度为何？

(a) $6-8i$　(b) $(6-8i)^2$　(c) $\dfrac{1}{6-8i}$　(d) $(6+8i)^2$

4 若 $|z|=2$ 与 $|w|=3$，则 $|z\times w|=$ ___且 $|z+w|\le$ ____且 $|z/w|=$____且 $|z-w|\le$ ___。

5 单位圆上角度是 30°, 60°, 90°, 120°的数字，求出他们的 $a+ib$？若 w 在 30°角，检查 w^2 在 60°角。w 的多少次方等于 1？

6 若 $z=r\cos\theta+ir\sin\theta$，则 $1/z$ 的绝对值是____且角度是____。它的极形式是____。执行 $z\times(1/z)$得到 1。

7 复数乘法 $M=(a+bi)(c+di)$是一个 2×2 的实数乘法

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} c \\ d \end{bmatrix}=\begin{bmatrix} \ \\ \ \end{bmatrix}$$

右侧包含 M 的实部与虚部，测试 $M=(1+3i)(1-3i)$。

8 $A=A_1+iA_2$ 是一个复数的 $n\times n$ 矩阵及 $\boldsymbol{b}=\boldsymbol{b}_1+i\boldsymbol{b}_2$ 是一个复数向量，$A\boldsymbol{x}=\boldsymbol{b}$ 的解是 $\boldsymbol{x}_1+i\boldsymbol{x}_2$。把 $A\boldsymbol{x}=\boldsymbol{b}$ 写成大小 $2n$ 的实数系统：

$$\begin{matrix} \text{复数 } n\times n \\ \text{实数 } 2n\times 2n \end{matrix}\quad\begin{bmatrix} & \\ & \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}=\begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix}$$

问题 9-16 有关共轭 $\bar{z}=a-ib=re^{-i\theta}=z^*$。

9 把 i 改成 $-i$，写下每个数的共轭复数：

(a) $2-i$　(b) $(2-i)(1-i)$　(c) $e^{i\pi/2}$ (就是 i)

(d) $e^{i\pi}=-1$　(e) $\dfrac{1+i}{1-i}$ (也是 i)　(f) $i^{103}=$_____

10 $z+\bar{z}$ 的总和永远是____，$z-\bar{z}$ 的差永远是____。假设 $z\ne 0$，$z\times\bar{z}$ 的乘积永远是____，比值 $z/\bar{z}$ 的绝对值是____。

11 一个实数矩阵，$A\boldsymbol{x}=\lambda\boldsymbol{x}$ 的共轭是 $A\bar{\boldsymbol{x}}=\bar{\lambda}\bar{\boldsymbol{x}}$，这个证明两件事情：$\bar{\lambda}$ 是另一个特征值以及 $\bar{\boldsymbol{x}}$ 是它的特征向量。若 $A=[a \quad b\,;\,-b \quad a]$，求出特征值 $\lambda,\bar{\lambda}$ 与特征向量 $\boldsymbol{x},\bar{\boldsymbol{x}}$。

12 一个实数 2×2 矩阵的特征值来自二次公式：

$$\det\begin{bmatrix}a-\lambda & b\\ c & d-\lambda\end{bmatrix}=\lambda^2-(a+d)\lambda+(ad-bc)=0$$

得到两个特征值 $\lambda=\left[a+d\pm\sqrt{(a+d)^2-4(ad-bc)}\right]/2$

(a) 若 $a=b=d=1$，当 $c=$_____，则特征值是复数。

(b) 当 $ad=bc$，特征值为何？

13 当问题 12 的(迹)$^2=(a+d)^2$ 比_____小，则特征值不是实数。证明当 $bc>0$ 时，λ's 是实数。

14 一个实数反对称矩阵($A^{\mathrm{T}}=-A$) 有纯虚数的特征值。首先证明：若 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 则方块乘法得到

$$\begin{bmatrix}0 & A\\ -A & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ i\boldsymbol{x}\end{bmatrix}=i\lambda\begin{bmatrix}\boldsymbol{x}\\ i\boldsymbol{x}\end{bmatrix}$$

这个方块矩阵是对称，它的特征值必须是_____！所以 λ 是_____。

问题 15-22 有关复数 $r\cos\theta+ir\sin\theta$ 的 $re^{i\theta}$ 的形式。

15 把下列数字写成欧拉形式 $re^{i\theta}$，然后取平方：

(a) $1+\sqrt{3}\,i$　　(b) $\cos2\theta+i\sin2\theta$　　(c) $-7i$　　(d) $5-5i$

16 (最爱) 求出 $z=\sin\theta+i\cos\theta$(小心)的绝对值与角度，在复数平面标示这个 z。把 $\cos\theta+i\sin\theta$ 乘 z 得到_____。

17 在复数平面画出 $z^8=1$ 的 8 个解，这些根 $z=\bar{w}=\exp(-2\pi i/8)$的矩形形式 $a+ib$ 为何？

18 在复数平面标示 1 的立方根，标示–1 的立方根，放在一起是___的 6 次方根。

19 比较 $e^{3i\theta}=\cos3\theta+i\sin3\theta$ 与 $(e^{i\theta})^3=(\cos\theta+i\sin\theta)^3$，利用 $\cos\theta$ 与 $\sin\theta$ 求出三倍角公式 $\cos3\theta$ 与 $\sin3\theta$。

20 假设共轭 $\bar{z}$ 等于倒数 $1/z$，这些可能的 z's 为何？

21 (a) 为什么 e^i 与 i^e 都有绝对值 1？

(b) 在复数平面上靠近 e^i 与 i^e 的地方放上*号。

(c) 数字 i^e 可以是 $(e^{i\pi/2})^e$ 或 $(e^{5i\pi/2})^e$，两者相等吗？

22 从 $t=0$ 到 $t=2\pi$，画出下列数字在复数平面的路径：

(a) e^{it}　　(b) $e^{(-1+i)t}=e^{-t}e^{it}$　　(c) $(-1)^t=e^{t\pi i}$

请尊重版权与译者的劳动成果，侵权必究！

9.2 埃尔米特与酉矩阵

本段落的重要内容可以用一句话来说明：**当你转置一个复数向量 z 或是矩阵 A，也要取他们的共轭复数**。不要停在 z^{T} 或是 A^{T}，反转所有**虚部**的符号。一个列向量有 $z_j = a_j + ib_j$，优质行向量 $\bar{z}^{\mathrm{T}}$是**共轭转置**，它的分量是 $a_j - ib_j$：

$$\textbf{共轭转置}\quad \bar{z}^{\mathrm{T}} = [\bar{z}_1 \quad \dots \quad \bar{z}_n] = [a_1 - ib_1 \quad \dots \quad a_n - ib_n] \tag{1}$$

有一个理由用到 $\bar{z}$。实数向量的长度平方是 $x_1^2 + \dots + x_n^2$，复数向量的长度平方**不是** $z_1^2 + \dots + z_n^2$。错误的定义使得$(1, i)$的长度是 $1^2 + i^2 = 0$，一个非零向量有零的长度--**不好**。其他的向量要有复数长度，我们不要$(a + bi)^2$，而是想要绝对值的平方 $a^2 + b^2$--这个是$(a + bi)$乘$(a - bi)$。

对每个分量，我们想要 z_j 乘$\bar{z}_j$，这是$|z_j|^2 = a_j^2 + b_j^2$，这个会来自 z 分量乘 $\bar{z}$ 的分量：

$$\textbf{长度平方}\qquad \begin{bmatrix}\bar{z}_1 & \dots & \bar{z}_n\end{bmatrix}\begin{bmatrix}z_1\\ \vdots \\ z_n\end{bmatrix} = |z_1|^2 + \dots + |z_n|^2 \qquad \text{这是}\bar{z}^{\mathrm{T}}z = \|z\|^2 \tag{2}$$

现在$(1, i)$的长度平方是 $1^2 + |i|^2 = 2$，长度是$\sqrt{2}$，$(1 + i, 1 - i)$的长度平方是 4。唯一有零长度的向量是零向量。

长度$\|z\|$ 是 $\bar{z}^{\mathrm{T}}z = z^{\mathrm{H}}z = |z_1|^2 + \dots + |z_n|^2$ 的**平方根**。

在深入探讨之前，我们用一个符号取代两个符号。不用横杠代表共轭以及 T 代表转置，我们只使用一个上标 **H**，于是$\bar{z}^{\mathrm{T}} = z^{\mathrm{H}}$。这是“$z$ 埃尔米特(Hermitian)”，z 的共轭转置，这个新字念做“Hermeeshan”。这个新符号也应用在矩阵：一个矩阵 A 的共轭转置是 A^{H}。

另一个流行的记号是 A^*，MATLAB 的转置指令“′”自动取得共轭复数 (z'是$z^{\mathrm{H}} = \bar{z}^{\mathrm{T}}$以及 A'是 $A^{\mathrm{H}} = \bar{A}^{\mathrm{T}}$)。

$$\textbf{A^{H} 是“A 埃尔米特”}\qquad \text{若 } A = \begin{bmatrix}1 & i\\ 0 & 1+i\end{bmatrix} \text{ 则 } A^{\mathrm{H}} = \begin{bmatrix}1 & 0\\ -i & 1-i\end{bmatrix}$$

复数内积

实数向量的长度平方是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$--这是 $\boldsymbol{x}$ 与自己的内积。复数向量的长度平方是 $z^{\mathrm{H}}z$。如果 $z^{\mathrm{H}}z$ 是 z 与自己的内积，这就是我们想要的。为了实现这件事情，复数内积应该使用**共轭转置**(不只是转置)，这个对于实数向量没有影响。

定义 实数或是复数向量 $\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的内积是 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}$：

$$\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v} = \begin{bmatrix}\overline{u}_1 & \dots & \overline{u}_n\end{bmatrix}\begin{bmatrix}v_1\\ \vdots \\ v_n\end{bmatrix} = \overline{u}_1 v_1 + \dots + \overline{u}_n v_n \tag{3}$$

复数向量的 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}$ 与 $\boldsymbol{v}^{\mathrm{H}}\boldsymbol{u}$ 不同，现在向量的顺序很重要，事实上 $\boldsymbol{v}^{\mathrm{H}}\boldsymbol{u} = \overline{v}_1 u_1 + \dots + \overline{v}_n u_n$ 是 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}$ 的共轭复数。我们需要为了好东西而忍受一些不方便。

范例 1 $\boldsymbol{u} = \begin{bmatrix}1\\ i\end{bmatrix}$ 与 $\boldsymbol{v} = \begin{bmatrix}i\\ 1\end{bmatrix}$ 的内积是 $\begin{bmatrix}1 & -i\end{bmatrix}\begin{bmatrix}i\\ 1\end{bmatrix} = 0$。

范例 1 让人惊讶，向量$(1, i)$与$(i, 1)$看起来不垂直，但是他们是垂直。**零内积仍然表示(复数)向量正交**。同理，向量$(1, i)$与$(1, -i)$是正交，他们的内积是 $1 - 1$。我们正确的得到零内积--我们也可能忘记取共轭，使得$(1, i)$的长度不正确的变成零。

注解 我们选择第一个向量 $\boldsymbol{u}$ 的共轭，有些作者选择第二个向量 $\boldsymbol{v}$，他们的复数内积就是 $\boldsymbol{u}^{\mathrm{T}}\overline{\boldsymbol{v}}$，我想这是自由选择。

$A\boldsymbol{u}$ 与 $\boldsymbol{v}$ 的内积等于 $\boldsymbol{u}$ 与 $A^{\mathrm{H}}\boldsymbol{v}$ 的内积：

A^{H} 也称为 A 的“伴随(adjoint)” $(A\boldsymbol{u})^{\mathrm{H}}\boldsymbol{v} = \boldsymbol{u}^{\mathrm{H}}(A^{\mathrm{H}}\boldsymbol{v})$ (4)

$A\boldsymbol{u}$ 的共轭是 $\overline{A\boldsymbol{u}}$，转置 $\overline{A\boldsymbol{u}}$ 如同平常得到 $\overline{\boldsymbol{u}}^{\mathrm{T}}\overline{A}^{\mathrm{T}}$，这是 $\boldsymbol{u}^{\mathrm{H}}A^{\mathrm{H}}$。每件事情应该有效也真的有效，^{H} 的规则来自 ^{T} 的规则。

我们一直使用$(a - ib)\ (c - id)$是$(a + ib)\ (c + id)$的共轭的事实。

AB 的共轭转置是 $(AB)^{\mathrm{H}} = B^{\mathrm{H}}A^{\mathrm{H}}$

埃尔米特矩阵 $S = S^{\mathrm{H}}$

在实数矩阵中，对称矩阵形成最重要的特殊类别：$S = S^{\mathrm{T}}$，他们有实数特征值以及在正交矩阵 Q 中的正交特征向量。每个实数对称矩阵可以写成 $S = Q\Lambda Q^{-1}$，也可以是 $S = Q\Lambda Q^{\mathrm{T}}$(因为 $Q^{-1} = Q^{\mathrm{T}}$)。当 S 是实数，这些全部来自 $S^{\mathrm{T}} = S$。

请尊重版权与译者的劳动成果，侵权必究!

在复数矩阵中，这个特殊类别包含**埃尔米特矩阵**：$S = S^{\mathrm{H}}$，矩阵单元的条件是 $s_{ij} = \bar{s}_{ji}$，这种情况下我们称 S 是埃尔米特矩阵。每个实数对称矩阵都是埃尔米特矩阵，因为取共轭没有影响。下个矩阵也是埃尔米特，$S = S^{\mathrm{H}}$：

范例 2 $S = \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix}$ 主对角线必须是实数，因为 $s_{ii} = \bar{s}_{ii}$
横跨对角线是共轭 $3+3i$ 与 $3-3i$。

这个例子会说明所有埃尔米特矩阵的 3 个重要性质。

若 $S = S^{\mathrm{H}}$ 且 z 是任意实数或复数列向量，则数字 $z^{\mathrm{H}}Sz$ 是实数。

快速证明：$z^{\mathrm{H}}Sz$ 当然是 1×1，取它的共轭转置：

$$(z^{\mathrm{H}}Sz)^{\mathrm{H}} = z^{\mathrm{H}}S^{\mathrm{H}}(z^{\mathrm{H}})^{\mathrm{H}} \quad \text{再次是 } z^{\mathrm{H}}Sz。$$

所以数字 $z^{\mathrm{H}}Sz$ 等于自己的共轭，必须是实数。$z^{\mathrm{H}}Sz$ 是一种“能量”：

$$\begin{bmatrix} \bar{z}_1 & \bar{z}_2 \end{bmatrix} \begin{bmatrix} 2 & 3-3i \\ 3+3i & 5 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \underset{\text{对角线}}{2\bar{z}_1 z_1 + 5\bar{z}_2 z_2} + \underset{\text{非对角线}}{(3-3i)\bar{z}_1 z_2 + (3+3i) z_1 \bar{z}_2}$$

$2|z_1|^2$ 与 $5|z_2|^2$ 这两项来自对角线而且都是实数，非对角线项是彼此的共轭——所以总和是实数(相加时虚部抵消)。整个表述 $z^{\mathrm{H}}Sz$ 是实数，这个会使得λ是实数。

埃尔米特矩阵的每个特征值都是实数。

证明 假设 $Sz = \lambda z$，同时用 z^{H} 乘两侧，得到 $z^{\mathrm{H}}Sz = \lambda z^{\mathrm{H}}z$。左侧 $z^{\mathrm{H}}Sz$ 是实数，右侧 $z^{\mathrm{H}}z$ 是长度的平方，是正数实数，所以比值$\lambda = z^{\mathrm{H}}Sz/z^{\mathrm{H}}z$ 是实数，**证明完毕**。

上述的例子有实数特征值$\lambda = 8$ 与$\lambda = -1$，因为 $S = S^{\mathrm{H}}$：

$$\begin{vmatrix} 2-\lambda & 3-3i \\ 3+3i & 5-\lambda \end{vmatrix} = \lambda^2 - 7\lambda + 10 - |3+3i|^2$$

$$= \lambda^2 - 7\lambda + 10 - 18 = (\lambda - 8)(\lambda + 1)$$

埃尔米特矩阵的特征向量是正交(当他们对应不同的特征值时)：

若 $Sz = \lambda z$ 且 $Sy = \beta y$，则 $y^{\mathrm{H}}z = 0$。

证明 在 $Sz = \lambda z$ 的左侧乘 y^{H}，在 $y^{\mathrm{H}}S^{\mathrm{H}} = \beta y^{\mathrm{H}}$ 的右侧乘 z：

$$y^{\mathrm{H}}Sz = \lambda y^{\mathrm{H}}z \quad \text{与} \quad y^{\mathrm{H}}S^{\mathrm{H}}z = \beta y^{\mathrm{H}}z \tag{5}$$

两式的左侧相等，所以$\lambda y^{\mathrm{H}}z = \beta y^{\mathrm{H}}z$，则 $y^{\mathrm{H}}z$ **必须是零**。

在我们的例子中，对应$\lambda=8$与$\beta=-1$的特征向量是正交：

$$(S-8I)\boldsymbol{z}=\begin{bmatrix}-6 & 3-3i\\ 3+3i & -3\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}\qquad \boldsymbol{z}=\begin{bmatrix}1\\ 1+i\end{bmatrix}$$

$$(S+I)\boldsymbol{y}=\begin{bmatrix}3 & 3-3i\\ 3+3i & 6\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}\qquad \boldsymbol{y}=\begin{bmatrix}1-i\\ -1\end{bmatrix}$$

正交特征向量 $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{z}=\begin{bmatrix}1+i & -1\end{bmatrix}\begin{bmatrix}1\\ 1+i\end{bmatrix}=0$。

这些特征向量的长度平方是$1^2+1^2+1^2=3$，除以$\sqrt{3}$之后变成单位向量。原来是正交现在是**正交单位**，他们进入特征向量矩阵X的列，可以对角化S。

当S是实数对称，X是Q ——正交矩阵，现在S是复数而且埃尔米特，它的特征向量是复数并且是正交单位。**特征向量矩阵$\boldsymbol{X}$很像$\boldsymbol{Q}$，但是复数：$\boldsymbol{QQ=I}$**。我们指定Q一个新名字“**酉(unitary)**”但是仍然叫做Q。

酉矩阵

一个**酉矩阵**Q是一个(复数)方形矩阵具有**正交单位的列**。

酉矩阵可以对角化S： $$Q=\frac{1}{\sqrt{3}}\begin{bmatrix}1 & 1-i\\ 1+i & -1\end{bmatrix}$$

这个Q也是一个埃尔米特矩阵，我不期待这个！这个例子太完美了，我们会看到这个Q的特征值必须是1与–1。

实数正交单位的矩阵测试方式是$Q^{\mathrm{T}}Q=I$，零内积出现在非对角线。在复数情况下Q^{T}变成Q^{H}，当Q^{H}乘Q，这些列显示本身是正交单位。内积填满$\boldsymbol{Q^{\mathrm{H}}Q=I}$：

每个具有正交单位列的矩阵$\boldsymbol{Q}$都有$\boldsymbol{Q^{\mathrm{H}}Q=I}$。

若$\boldsymbol{Q}$是方形，它是一个酉矩阵，则 $\boldsymbol{Q^{\mathrm{H}}=Q^{-1}}$。

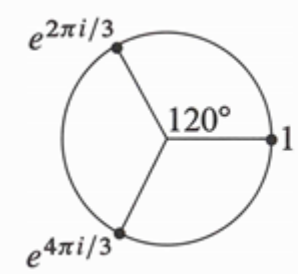
假设Q(有正交单位列)乘任意的$\boldsymbol{z}$，向量长度会维持不变，因为$\boldsymbol{z}^{\mathrm{H}}Q^{\mathrm{H}}Q\boldsymbol{z}=\boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}$。如果$\boldsymbol{z}$是$Q$的一个特征向量，我们可以学到更多：**酉(与正交)矩阵$\boldsymbol{Q}$的特征值全部都有$|\lambda|=1$**。

若$\boldsymbol{Q}$是酉，则 $\|\boldsymbol{Qz}\|=\|\boldsymbol{z}\|$。 因此$\boldsymbol{Qz}=\lambda\boldsymbol{z}$得到 $|\lambda|=1$。

我们的2×2范例同时是埃尔米特($Q=Q^{\mathrm{H}}$)与酉($Q^{-1}=Q^{\mathrm{H}}$)，表示有实数特征值以及$|\lambda|=1$。实数而且$|\lambda|=1$只有两种可能性：特征值是1或–1。Q的迹是0，所以$\lambda=1$与$\lambda=-1$.

请尊重版权与译者的劳动成果，侵权必究!

范例 3　图 9.3 显示 3×3 的**傅里叶矩阵**，它是埃尔米特吗？是酉吗？F_3 当然是对称，它等于自己的转置，但是它不等于自己的共轭转置——不是埃尔米特。如果你把 i 变成 $-i$，你得到不同的矩阵。



傅里叶矩阵　$$F = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{2\pi i/3} & e^{4\pi i/3} \\ 1 & e^{4\pi i/3} & e^{2\pi i/3} \end{bmatrix}.$$

图 9.3：1 的立方根进入傅里叶矩阵 $F = F_3$。

F 是酉？是的。每个列的长度平方是$(1 + 1 + 1)/3$ (单位向量)，第一列与第二列正交，因为 $1 + e^{2\pi i/3} + e^{4\pi i/3} = 0$，这是图 9.3 标示的 3 个数的总和。

注意图形的对称，如果把它旋转 120°，这三个点还在相同的位置，因此他们的总和 S 仍然在相同位置！旋转 120°之后，在相同方向唯一可能的总和是 $S = 0$。

F 的列 2 与列 3 正交吗？他们的点积看起来像

$$(1 + e^{6\pi i/3} + e^{6\pi i/3})/3 = (1 + 1 + 1)/3$$

这不是零，因为我们忘了取共轭复数，答案是错误。复数内积要用 ^{H} 而不是 ^{T}：

$$(\text{列 } 2)^{\mathrm{H}}(\text{列 } 3) = (1 \cdot 1 + e^{-2\pi i/3}e^{4\pi i/3} + e^{-4\pi i/3}e^{2\pi i/3})/3$$

$$= (1 + e^{2\pi i/3} + e^{-2\pi i/3})/3 = 0$$

所以我们得到正交性，**结论：F 是一个酉矩阵。**

下个段落会学习 $n \times n$ 的傅里叶矩阵，在所有的复数酉矩阵中，他们是最重要的。当我们用 F 乘一个向量，我们在计算它的**离散傅里叶转换**。当我们用 F^{-1} 乘一个向量，我们在计算它的**逆转换**。酉矩阵的特殊性质是 $F^{-1} = F^{\mathrm{H}}$，逆转换的差别只在于把 i 变成 $-i$：

把 i 变成 $-i$：　$$F^{-1} = F^{\mathrm{H}} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-2\pi i/3} & e^{-4\pi i/3} \\ 1 & e^{-4\pi i/3} & e^{-2\pi i/3} \end{bmatrix}$$

每个处理 F 的人都知道它的价值，本章最后一个段落会把傅里叶分析与复数与线性代数做结合。

问题集 9.2

1 求出 $\boldsymbol{u} = (1+i, 1-i, 1+2i)$与 $\boldsymbol{v} = (i, i, i)$的长度。求出 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v}$ 与 $\boldsymbol{v}^{\mathrm{H}}\boldsymbol{u}$.

2 计算 $A^{\mathrm{H}}A$ 与 AA^{H}，两者都是_____矩阵：

$$A = \begin{bmatrix} i & 1 & i \\ 1 & i & i \end{bmatrix}$$

3 对于问题 2 的 A，求解 $A\boldsymbol{z} = \boldsymbol{0}$，找出 A 的零空间的向量 $\boldsymbol{z}$。证明 $\boldsymbol{z}$ 与 A^{H} 的列正交，证明 $\boldsymbol{z}$ 与 A^{T} 的列不正交。**优质行空间不再是 $\boldsymbol{C}(\boldsymbol{A}^{\mathrm{T}})$，现在是 $\boldsymbol{C}(\boldsymbol{A}^{\mathrm{H}})$**。

4 问题 3 指出 4 个基础子空间是 $\boldsymbol{C}(A^{\mathrm{T}})$与 $\boldsymbol{N}(A)$与_____与_____。他们的维度仍然是 r 与 $n-r$ 与 r 与 $m-r$，他们仍然是正交子空间。代号 ^{H} 取代 ^{T}。

5 (a) 证明 $A^{\mathrm{H}}A$ 永远是埃尔米特矩阵。

(b) 若 $A\boldsymbol{z} = \boldsymbol{0}$ 则 $A^{\mathrm{H}}A\boldsymbol{z} = \boldsymbol{0}$。若 $A^{\mathrm{H}}A\boldsymbol{z} = \boldsymbol{0}$，左乘 $\boldsymbol{z}^{\mathrm{H}}$ 证明 $A\boldsymbol{z} = \boldsymbol{0}$。$A$ 与 $A^{\mathrm{H}}A$ 的零空间是_____，因此当 A 的零空间只包含 $\boldsymbol{z} = \boldsymbol{0}$ 时，$A^{\mathrm{H}}A$ 是一个可逆埃尔米特矩阵。

6 是非题(理由或反例)：

(a) 若 A 是实数矩阵则 $A + iI$ 是可逆。

(b) 若 S 是埃尔米特矩阵则 $S + iI$ 是可逆。

(c) 若 Q 是酉矩阵则 $Q + iI$ 是可逆。

7 当你对埃尔米特矩阵 S 乘上一个实数 c，cS 还是埃尔米特吗？当 S 是埃尔米特矩阵，证明 iS 是一个反埃尔米特矩阵。假设“纯量”是实数，则 3×3 埃尔米特矩阵是一个子空间。

8 下列的 P 属于哪一类矩阵：可逆，埃尔米特，酉？

$$P = \begin{bmatrix} 0 & i & 0 \\ 0 & 0 & i \\ i & 0 & 0 \end{bmatrix}$$

计算 P^2, P^3, P^{100}。P 的特征值为何？

9 求出问题 8 的 P 的单位特征向量，然后把他们放进一个酉矩阵 Q 的列，P 的什么性质使得这些特征向量正交？

10 写出 3×3 循环矩阵 $C = 2I + 5P$。它与问题 8 的 P 有相同的特征向量，求出特征值。

11 若 Q 与 U 都是酉矩阵，证明 Q^{-1} 是酉，QU 也是酉。从 $Q^{\mathrm{H}}Q = I$ 与 $U^{\mathrm{H}}U = I$ 开始。

请尊重版权与译者的劳动成果，侵权必究!

12 你如何知道每个埃尔米特矩阵的行列式是实数？

13 当 A 的列是无关，矩阵 $A^{\mathrm{H}}A$ 不只是埃尔米特也是正定。证明：若 $\boldsymbol{z}$ 不是零则 $\boldsymbol{z}^{\mathrm{H}}A^{\mathrm{H}}A\boldsymbol{z}$ 是正数，这是因为_____。

14 对角化下列埃尔米特矩阵得到 $S = Q\Lambda Q^{\mathrm{H}}$：

$$S=\begin{bmatrix} 0 & 1-i \\ i+1 & 1 \end{bmatrix} \qquad \text{与} \qquad S=\begin{bmatrix} 2 & 1+i \\ i-1 & 3 \end{bmatrix}$$

15 对角化下列反埃尔米特矩阵得到 $K = Q\Lambda Q^{\mathrm{H}}$，所有的$\lambda$'s 是_____：

$$K=\begin{bmatrix} 0 & -1+i \\ 1+i & i \end{bmatrix}$$

16 对角化下列正交矩阵得到 $U = Q\Lambda Q^{\mathrm{H}}$，现在所有的$\lambda$'s 是_____：

$$U=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

17 对角化下列酉矩阵得到 $U = Q\Lambda Q^{\mathrm{H}}$，再次所有的$\lambda$'s 是_____：

$$U=\frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}$$

18 若 $\boldsymbol{v}_1, ..., v_n$ 是 $\mathbf{C}^n$ 的一组正交单位基底，这些向量当做列的矩阵是_____矩阵。证明任意向量 $\boldsymbol{z}$ 等于$(\boldsymbol{v}_1{}^{\mathrm{H}}\boldsymbol{z})\boldsymbol{v}_1 + ... + (\boldsymbol{v}_n{}^{\mathrm{H}}\boldsymbol{z})\boldsymbol{v}_n$。

19 $\boldsymbol{v} = (1, i, 1)$，$\boldsymbol{w} = (i, 1, 0)$与 $\boldsymbol{z}$ = _____是_____的一组正交基底。

20 若 $S = A + iB$ 是埃尔米特矩阵，它的实部对称吗？虚部对称吗？

21 $\mathbf{C}^n$ 的(复数)维度是____。求出 $\mathbf{C}^n$ 的一个非实数基底。

22 描述所有的 1×1 与 2×2 的埃尔米特矩阵与酉矩阵。

23 A^{H} 的特征值与方形矩阵 A 的特征值有什么关联？

24 若 $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u} = 1$，证明 $I - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$ 是埃尔米特也是酉。秩 1 矩阵 $\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$ 是在 $\mathbf{C}^n$ 的哪条直线的投影？

25 若 $A + iB$ 是酉矩阵(A 与 B 是实数)，证明$\begin{bmatrix} A & -B \\ B & A \end{bmatrix}$是正交矩阵。

26 若 $A + iB$ 是埃尔米特矩阵(A 与 B 是实数)，证明$\begin{bmatrix} A & -B \\ B & A \end{bmatrix}$是对称。

27 证明一个埃尔米特矩阵的逆矩阵还是埃尔米特(转置 $S^{-1}S = I$)。

28 一个具有正交单位特征向量的矩阵 A 有这样的形式 $N = Q\Lambda Q^{-1} = Q\Lambda Q^{\mathrm{H}}$，证明 $NN^{\mathrm{H}} = N^{\mathrm{H}}N$。这些 N 恰好是**正态矩阵**，范例是埃尔米特，反埃尔米特与酉矩阵。利用Λ中的复数特征值，从 $Q\Lambda Q^{\mathrm{H}}$ 创建一个 2×2 的正态矩阵。

9.3 快速傅里叶转换

线性代数的许多应用需要时间发展，不容易在一个小时内说明清楚，教师与作者必须在完整原理与加入新应用之间做选择。通常原理获胜，但是本段落是例外，它说明了上个世纪最重要的数值演算法。

我们想要快速的左乘 F 与 F^{-1}，他们是傅里叶矩阵与逆矩阵，这可以利用快速傅里叶转换达成。一个正常的乘积 Fc 使用 n^2 次乘法(F 有 n^2 个单元)，FFT 只需要 n 乘½ $\log_2 n$。我们在后面会看到。

FFT 有革命性的信号处理技术，整个工业因为这个概念得到加速，电气工程师是第一个知道差别的--当他们看到你(如果你是一个函数)，就取傅里叶转换。傅里叶的的概念是把 f 表示成泛音(harmonics) $c_k e^{ikx}$ 的总和。通过这些系数 c_k 在**频率空间**中看到函数，而不是通过它的数值 $f(x)$ 在**物理空间**中看到。在 c's 与 f 's 之间往返的通道就是使用傅里叶转换，快速通道使用 FFT。

1 的根与傅里叶矩阵

二次方程式有两个根(或是一个重根)，n 次方程式有 n 个根(重复也计算在内)，这是线性代数的基础定理，为了让它成立我们必须允许复数根。本段落是关于非常特殊的方程式 $z^n = 1$，它的解 z 是“1 的 n 次方根”，他们是复数平面上 n 个环绕单位圆的平均分布点。

图 9.4 显示 $z^8 = 1$ 的 8 个根，他们的距离是(360°)/8 = 45°。第一个根在 45°或是 $2\pi/8$ 弧度，也就是复数 $w = e^{i\theta} = e^{i2\pi/8}$。我们称这个数 w_8 来强调它是 8 次方根，你可以写成 $\cos(2\pi/8)$与 $\sin(2\pi/8)$，但是不要这么做。其他 7 个 8 次方根是 w^2, w^3, ... , w^8，环绕单位圆前进。w 的次方在极形式最佳，因为我们只处理角度 $2\pi/8$, $4\pi/8$, ..., $16\pi/8 = 2\pi$，这 8 个角度是 45°, 90°, 135°, ..., 360°。

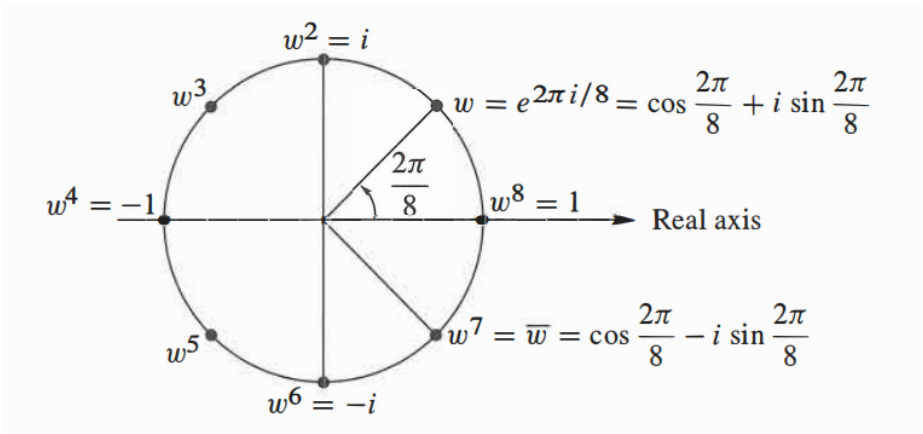


图 9.4：$z^8 = 1$ 的 8 个解是 $1, w, w^2, \ldots, w^7$，其中 $w = (1+i)/\sqrt{2}$。

请尊重版权与译者的劳动成果，侵权必究!

1 的 4 次方根也在图上，他们是 $i, -1, -i, 1$，角度现在是 $2\pi/4$ 或是 90°，第一个根 $w_4 = e^{i2\pi/4}$ 就是 i。甚至 1 的平方根也在，$w_2 = e^{i2\pi/2} = -1$，不要忽视了平方根 1 与–1。FFT 背后的概念就是从一个 **8×8** 傅里叶矩阵(包含 w_8 的次方)往下走到 **4×4** 的矩阵(包含 $w_4 = i$ 的次方)，相同的概念从 4 到 2。利用 F_8 往下到 F_4 以及往上到 F_{16}(还可以往上)的关联性，FFT 使得左乘 F_{1024} 非常快速。

我们描述傅里叶矩阵，先从 $n = 4$ 开始，它的行包含 1 的次方与 w 的次方与 w^2 的次方与 w^3 的次方，这些是 1 的 4 次方根，他们有一个特殊的顺序。

$$\begin{matrix}\text{傅里叶矩阵}\\ n=4\\ w=i\end{matrix}\qquad F=\begin{bmatrix}1&1&1&1\\1&w&w^2&w^3\\1&w^2&w^4&w^6\\1&w^3&w^6&w^9\end{bmatrix}=\begin{bmatrix}1&1&1&1\\1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\end{bmatrix}$$

矩阵是对称($F = F^{\mathrm{T}}$)，它不是埃尔米特，它的主对角线不是实数，但是 $\boldsymbol{F/2}$ 是一个**酉矩阵**，表示 $\left(\frac{1}{2}F^{\mathrm{H}}\right)\left(\frac{1}{2}F\right) = I$：

F 的列得到 $F^{\mathrm{H}}F = 4I$，它的逆矩阵是 $\frac{1}{4}F^{\mathrm{H}}$，也就是 $F^{-1} = \frac{1}{4}\bar{F}$。

逆矩阵从 $w = i$ 变成 $\bar{w} = -i$，让我们从 F 到 $\bar{F}$。当快速傅里叶转换给你一个快速的方法左乘 F，$\bar{F}$与 F^{-1} 也会做同样的事情。

每个列的长度是$\sqrt{n}$，所以酉矩阵是 $Q = F/\sqrt{n}$ 以及 $Q^{-1} = \bar{F}/\sqrt{n}$。我们避免使用 $\sqrt{n}$，只使用 F 与 $\boldsymbol{F^{-1} = \bar{F}/n}$。重点是 F 乘 c_0, c_1, c_2, c_3：

$$\begin{matrix}\text{4 个点的}\\ \text{傅里叶级数}\end{matrix}\qquad \begin{bmatrix}y_0\\y_1\\y_2\\y_3\end{bmatrix}=F\boldsymbol{c}=\begin{bmatrix}1&1&1&1\\1&w&w^2&w^3\\1&w^2&w^4&w^6\\1&w^3&w^6&w^9\end{bmatrix}\begin{bmatrix}c_0\\c_1\\c_2\\c_3\end{bmatrix}\qquad (1)$$

输入是 4 个复数系数 c_0, c_1, c_2, c_3，输出是 4 个函数值 y_0, y_1, y_2, y_3。第一个输出 $y_0 = c_0 + c_1 + c_2 + c_3$ 是傅里叶级数 $\sum c_k e^{ikx}$ 在 $x = 0$ 的值，第二个输出是级数 $\sum c_k e^{ikx}$ 在 $x = 2\pi/4$ 的值：

$$y_1 = c_0 + c_1 e^{i2\pi/4} + c_2 e^{i4\pi/4} + c_3 e^{i6\pi/4} = c_0 + c_1 w + c_2 w^2 + c_3 w^3$$

第三与第四个输出 y_2 与 y_3 是 $\sum c_k e^{ikx}$ 在 $x = 4\pi/4$ 与 $x = 6\pi/4$ 的值，这些是有限傅里叶级数！他们包含 $n = 4$ 项以及在 4 个点计算函数值。这些点 $0, 2\pi/4, 4\pi/4, 6\pi/4$ 都是等距离。

下一个点是 $x = 8\pi/4$ 就是 2π，级数回到 y_0，因为$e^{2\pi i}$与$e^0 = 1$ 相同，每件事情都是周期 4 的循环。这个世界 2 + 2 是 0，因为$(w^2)\,(w^2) = w^0 = 1$。我们遵循方便性**令 j 与 k 从 0 到 $n-1$**(不是 1 到 n)，F 的“第零行”与“第零列”都是 1。

$n\times n$ 矩阵包含 $w=e^{2\pi i/n}$ 的次方：

$$F_n\boldsymbol{c}=\begin{bmatrix}1&1&1&\bullet&1\\1&w&w^2&\bullet&w^{n-1}\\1&w^2&w^4&\bullet&w^{2(n-1)}\\\bullet&\bullet&\bullet&\bullet&\bullet\\1&w^{n-1}&w^{2(n-1)}&\bullet&w^{(n-1)^2}\end{bmatrix}\begin{bmatrix}c_0\\c_1\\c_2\\\bullet\\c_{n-1}\end{bmatrix}=\begin{bmatrix}y_0\\y_1\\y_2\\\bullet\\y_{n-1}\end{bmatrix}=\boldsymbol{y} \tag{2}$$

F_n 是对称但不是埃尔米特，它的列是正交，且 $F_n\overline{F}_n=nI$，则 F_n^{-1} 是 $\overline{F}_n/n$，逆矩阵包含 $\overline{w}_n=e^{-2\pi i/n}$，检视 F 的模式：

在(行 j, 列 k)的单元是 w^{jk}，行零与列零包含 $w^0=1$。

当我们把 F_n 乘 $\boldsymbol{c}$，我们取 n 个点的级数的总和。当我们把 F_n^{-1} 乘 $\boldsymbol{y}$，我们从函数值 $\boldsymbol{y}$ 求出系数 $\boldsymbol{c}$。在 MATLAB 这个指令是 c = fft(y)。矩阵 F 从“频率空间”传送到“物理空间”。 【ω是 omega】

重要注解。许多作者偏好处理 $\omega=e^{-2\pi i/N}$，这是我们的 w 的共轭复数(他们通常使用希腊字母 omega，我会保持两种选择是分开。) 这种选择之下，他们的 DFT 矩阵包含ω而不是 w，它是 $\overline{F}$，我们的 F 的共轭，$\overline{F}$ 从物理空间到频率空间。

$\overline{F}$ 完全合理的选择！MATLAB 使用$\omega=e^{-2\pi i/N}$，DFT 矩阵 fft(eye(N))包含这个数$\omega=\overline{w}$ 的次方。**傅里叶矩阵 $\boldsymbol{F}$ 使用 $\boldsymbol{w}$'s 从 $\boldsymbol{c}$ 重建 $\boldsymbol{y}$，矩阵 $\overline{\boldsymbol{F}}$ 使用 $\boldsymbol{w}$'s 计算傅里叶系数，如同 fft(y)**。

同样重要。 当$f(x)$的周期是 2π。我们把 x 换成$e^{i\theta}$，函数定义成环绕单位圆 (其中 $z=e^{i\theta}$)，离散傅里叶转换与插值法(interpolation)相同，找出多项式 $p(z)=c_0+c_1z+\ldots+c_{n-1}z^{n-1}$ 适配于 n 个值 $f_0,\ldots,f_{n-1}$：

插值法 求出 $c_0,\ldots,c_n$ 使得在 n 个点 $z=1,\ldots,w^{n-1}$ 时，$p(z)=f$。

在这 n 个特殊点的内插值对应的范德蒙矩阵就是傅里叶矩阵。

快速傅里叶转换的一步

我们希望 F 乘 $\boldsymbol{c}$ 越快越好。正常情况下一个矩阵乘一个向量需要 n^2 次个别乘法--矩阵有 n^2 个单元，你可能认为没办法再快了。(如果矩阵有零单元，乘法可以跳过，但是傅里叶矩阵没有零！) 利用单元的特殊模式w^{jk}，F 可以用一种方法分解得到很多零，这就是 **FFT**。

关键概念是连接 $\boldsymbol{F_n}$ 与半幅傅里叶矩阵 $\boldsymbol{F_{n/2}}$，假设 n 是 2 的次方(比如 $n=2^{10}=1024$)，我们会连接 $\boldsymbol{F_{1024}}$ **与两个复制品** $\boldsymbol{F_{512}}$。

请尊重版权与译者的劳动成果，侵权必究!

当 $n=4$，关键在于 F_4 与两个复制品 F_2：

$$\boldsymbol{F_4}=\begin{bmatrix}1&1&1&1\\1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\end{bmatrix}\qquad 与\qquad \begin{bmatrix}\boldsymbol{F}_2&\\&\boldsymbol{F}_2\end{bmatrix}=\begin{bmatrix}1&1&&\\1&i^2&&\\&&1&1\\&&1&i^2\end{bmatrix}$$

左侧是 F_4，没有零。右侧是一个一半是零的矩阵，工作量少了一半。但是等等，这两个矩阵不一样。我们需要两个稀疏又简单的矩阵来完成 FFT 分解：

$$\textbf{FFT 的分解}\quad \boldsymbol{F_4}=\begin{bmatrix}1&&1&\\&1&&i\\1&&-1&\\&1&&-i\end{bmatrix}\begin{bmatrix}1&1&&\\1&i^2&&\\&&1&1\\&&1&i^2\end{bmatrix}\begin{bmatrix}1&&&\\&&1&\\&1&&\\&&&1\end{bmatrix}\tag{3}$$

最后是一个排列矩阵，它把偶数 c's(c_0 与 c_2)放在奇数 c's(c_1 与 c_3)前面。中间的矩阵执行针对偶数 c's 与奇数 c's 的半幅(half-size)转换 F_2，左边的矩阵组合两个半幅输出——以一种产生正确全幅(full-size)输出 $\boldsymbol{y}=F\boldsymbol{c}$ 的方式。

相同的概念可以应用在 $n=1024$ 与 $m=n/2=512$，数字 w 是 $e^{i2\pi/1024}$，它是在单位圆上角度 $\theta=2\pi/1024$。傅里叶矩阵 F_{1024} 充满 w 的次方，FFT 的第一阶段是伟大的分解，是 Cooley 与 Tukey 所发现(1805 年高斯已经预示)：

$$F_{1024}=\begin{bmatrix}I_{512}&D_{512}\\I_{512}&-D_{512}\end{bmatrix}\begin{bmatrix}F_{512}&\\&F_{512}\end{bmatrix}\begin{bmatrix}偶-奇\\排列\end{bmatrix}\tag{4}$$

I_{512} 是单位矩阵，D_{512} 是对角单元为$(1, w, ..., w^{511})$的对角矩阵，两个复制品 F_{512} 是我们预期的，不要忘了他们使用 1 的 512 次方根(就是 w^2 而已!!) 排列矩阵把输入向量 $\boldsymbol{c}$ 的偶数与奇数分开得到 $\boldsymbol{c}'=(c_0, c_2, ..., c_{1022})$与 $\boldsymbol{c}''=(c_1, c_3, ..., c_{1023})$。

下列是与 F_{1024} 的分解有相同意义的代数公式：

(FFT 的一步) 令 $\boldsymbol{m=n/2}$，$\boldsymbol{y}=F_n\boldsymbol{c}$ 的前 m 个与后 m 个分量组合成半幅转换 $\boldsymbol{y}'=F_m\boldsymbol{c}'$与$\boldsymbol{y}''=F_m\boldsymbol{c}''$。方程式(4)展示$n$到$m=n/2$的步骤，如同$I\boldsymbol{y}'+D\boldsymbol{y}''$与$I\boldsymbol{y}'-D\boldsymbol{y}''$：

$$\begin{aligned}y_j&=y_j'+(w_n)^j y_j'',\qquad j=0,\ ...,m-1\\y_{j+m}&=y_j'-(w_n)^j y_j'',\qquad j=0,\ ...,m-1\end{aligned}\tag{5}$$

把 $\boldsymbol{c}$ 分成 $\boldsymbol{c}'$与 $\boldsymbol{c}''$，利用 F_m 转换成 $\boldsymbol{y}'$与 $\boldsymbol{y}''$，然后(5)重建 $\boldsymbol{y}$。

这些公式来自把 $c_0, ..., c_{n-1}$ 分开成为偶 c_{2k} 与奇 c_{2k+1}：w 是 w_n。

$$\boldsymbol{y}=F\boldsymbol{c}\quad y_j=\sum_0^{n-1}w^{jk}c_k=\sum_0^{m-1}w^{2jk}c_{2k}+\sum_0^{m-1}w^{j(2k+1)}c_{2k+1}\qquad 其中\ m=\frac{1}{2}n\tag{6}$$

偶数 c's 进入 $\boldsymbol{c}' = (c_0, c_2, \ldots)$与奇数 c's 进入 $\boldsymbol{c}'' = (c_1, c_3, \ldots)$，得到了转换 $F_m\boldsymbol{c}'$与 $F_m\boldsymbol{c}''$。关键是 $w_n^{\,2} = w_m$，得到 $w_n^{\,2jk} = w_m^{\,jk}$。

改写(6)
$$y_j = \sum (w_m)^{jk} c_k' + (w_n)^j \sum (w_m)^{jk} c_k'' = y_j' + (w_n)^j y_j'' \qquad (7)$$

当 $j \geq m$，(5)中的负号来自从 $(w_n)^j$ 提出 $(w_n)^m = -1$。

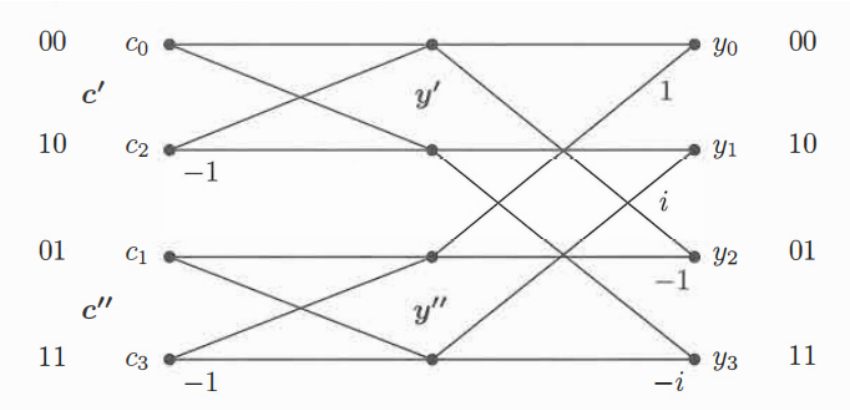
MATLAB 很容易分开偶数 c's 与奇数并且用 w_n^j 去乘，我们使用 conj(F)或等效 MATLAB 的逆转换 ifft，因为 fft 是基于 $\omega = \overline{w} = e^{-2\pi i/n}$。问题 16 显示 F 与 conj(F)是利用行的排列做连接。

MATLAB 的 FFT 步骤从 n 到 $n/2$

```
y' = ifft(c(0:2:n-2))*n/2;
y'' = ifft(c(1:2:n-1))*n/2;
d = w.^(0:n/2-1)';
y = [y' + d.*y''; y' - d.*y''];
```

流向图显示 c'与 c''穿越半幅 F_2，这些步骤依据形状称为“蝴蝶”，则输出 y'与 y''组合(来自 D 的 1 与 i，以及来自–D 的–1 与–i)得到 $\boldsymbol{y} = F_4\boldsymbol{c}$。

从 F_n 到两个 F_m's 的简化几乎减去了一半的工作——在分解中看到的零，这个简化很好但是不伟大，FFT 整体的概念更加有能量。它节省比一半还要多的时间。



使用递归的完整 FFT

如果你已经阅读到这儿，你可能猜到接下来是什么。我们把 F_n 简化到 $F_{n/2}$，然后到 $F_{n/4}$。每个 F_{512} 到 F_{256}，然后 256 到 128，这是一种递归(recursion)

递归是很多快速演算法的基础，这里是步骤 2 的 4 个复制品 F_{256} 与 D(w_{512} 的 256 次方)。偶数的偶数 $c_0, c_4, c_8, \ldots$先出现：

$$\begin{bmatrix} F_{512} & & & \\ & F_{512} & & \\ & & & \end{bmatrix} = \begin{bmatrix} I & D & & \\ I & -D & & \\ & & I & D \\ & & I & -D \end{bmatrix} \begin{bmatrix} F & & & \\ & F & & \\ & & F & \\ & & & F \end{bmatrix} \begin{matrix} \text{挑选} & 0,4,8,\ldots \\ \text{挑选} & 2,6,10,\ldots \\ \text{挑选} & 1,5,9,\ldots \\ \text{挑选} & 3,7,11,\ldots \end{matrix}$$

请尊重版权与译者的劳动成果，侵权必究!

我们会计算个别的乘法次数来观察节省了多少。在 **FFT** 发明之前，计数通常是 $n^2 = (1024)^2$，大概是一百万个乘法。我没有说这是一个很长的时间，当我们有越来越多的转换要做时成本变大很多--这是典型的事情，FFT 的节省也是变多：

大小 $n = 2^l$ 的最终次数从 n^2 降低为 $nl/2$

数字 1024 是 2^{10}，所以 $l = 10$，原始的次数是$(1024)^2$ 降为(5)(1024)，节省因子是 200，一百万变成五千，这是为什么 FFT 有革命性的信号处理。

此处是 $nl/2$ 背后的推理，有 l 层，从 $n = 2^l$ 降到 $n = 1$，每层有 $n/2$ 个来自对角线 D's 的乘法，重组来自较低层的半幅输出，这个得到最终次数 $nl/2$，就是 $\frac{1}{2}n\log_2 n$。

最后一个关于这个惊人演算法的注解，c's 进入 FFT 的顺序有一个惊人的规则，在所有的偶-奇排列之后。写出 0 到 $n-1$ 的二进制(例如 $n = 4$ 的 00, 01, 10, 11)，把每个位数反序：00, 10, 01, 11，得到**比特-反序 0, 2, 1, 3**，其中偶在奇之前(见问题 17。) 完整图形显示 c's 在**比特反序**，$l = \log_2 n$ 步的递归，最终输出 $y_0, \ldots, y_{n-1}$ 就是 F_n 乘 $\boldsymbol{c}$。

本章的结束是非常基础的概念，矩阵乘向量。

问题集 9.3

1 计算方程式(3)的 3 个矩阵乘积并与 F 做比较，哪 6 个单元你需要知道 $i^2 = -1$？

2 逆反方程式(3)的 3 个因子求出 F^{-1} 的快速分解。

3 F 是对称，转置方程式(3)求出新的快速傅里叶转换。

4 F_6 分解的所有单元与 $w_6 = 1$ 的 6 次方方根有关：

$$F_6 = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_3 & \\ & F_3 \end{bmatrix} [P]$$

写出这些矩阵，其中 D 有 $1, w_6, w_6^2$ 以及 F_3 有 $w_3 = w_6^2$。计算乘积！

5 若 $\boldsymbol{v} = (1, 0, 0, 0)$与 $\boldsymbol{w} = (1, 1, 1, 1)$，证明 $F\boldsymbol{v} = \boldsymbol{w}$ 与 $F\boldsymbol{w} = 4\boldsymbol{v}$，因此 $F^{-1}\boldsymbol{w} = \boldsymbol{v}$ 与 $F^{-1}\boldsymbol{v} =$ _____。

6 4×4 傅里叶矩阵 F^2 与 F^4 为何？

7 把向量 $\boldsymbol{v} = (1, 0, 1, 0)$通过 FFT 的 3 个步骤求出 $\boldsymbol{y} = F\boldsymbol{c}$，对 $\boldsymbol{c} = (0, 1, 0, 1)$重做一次。

8 $\boldsymbol{c} = (1, 0, 1, 0, 1, 0, 1, 0)$，利用 FFT 的 3 个步骤计算 $\boldsymbol{y} = F_8\boldsymbol{c}$。对 $\boldsymbol{c} = (0, 1, 0, 1, 0, 1, 0, 1)$重做一次。

网址：linear.neocities.org 邮箱：876152660@qq.com

9 若 $w = e^{2\pi i/64}$，则 w^2 与 $\sqrt{w}$ 在 1 的_____次方根与_____次方根之中。

10 (a) 在单位圆画出所有的 1 的 6 次方根，证明他们的总和是零。

(b) 1 的 3 个立方根为何？他们的总和也是零？

11 傅里叶矩阵 F 的列是循环排列 P(见段落 8.3)的特征向量，计算 PF 求出特征值$\lambda_1, \lambda_2, \lambda_3, \lambda_4$：

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_4 \end{bmatrix}$$

这是 $PF = F\Lambda$或 $P = F\Lambda F^{-1}$，特征向量矩阵(通常是 X)是 F。

12 方程式 $\det(P - \lambda I) = 0$ 是$\lambda^4 = 1$，再次显示特征值是$\lambda =$ _____。哪个排列 P 有特征值= 1 的立方根？

13 (a) C 的两个特征向量是(1, 1, 1, 1)与$(1, i, i^2, i^3)$，求出特征值 e。

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 \\ c_3 & c_0 & c_1 & c_2 \\ c_2 & c_3 & c_0 & c_1 \\ c_1 & c_2 & c_3 & c_0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = e_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{与} \quad C \begin{bmatrix} 1 \\ i \\ i^2 \\ i^3 \end{bmatrix} = e_2 \begin{bmatrix} 1 \\ i \\ i^2 \\ i^3 \end{bmatrix}$$

(b) $P = F\Lambda F^{-1}$ 立即得到 $P^2 = F\Lambda^2 F^{-1}$ 与 $P^3 = F\Lambda^3 F^{-1}$，则 $\boldsymbol{C} = cI + c_1P + c_2P^2 + c_3P^3$ $= F(c_0I + c_1\Lambda + c_2\Lambda^2 + c_3\Lambda^3)\,F^{-1} = \boldsymbol{FEF^{-1}}$。括号中的矩阵 E 是对角矩阵，它包含 C 的_____。

14 从 $E = 2I - \Lambda - \Lambda^3$ 求出“周期”–1, 2, –1 矩阵的特征值，其中 P 的特征值在Λ 中。角落的–1's 使得矩阵变成周期的：

$$C = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \quad \text{有} \quad c_0 = 2, c_1 = -1, c_2 = 0, c_3 = -1.$$

15 **卷积(convolution) = C 的快速乘法：**计算 C 乘一个向量 $\boldsymbol{x}$，我们可以计算 $F(E(F^{-1}\boldsymbol{x}))$，这个直接方法使用 n^2 次个别乘法。已知 E 与 F，第二个方法只使用 $n\log_2 n + n$ 次乘法。有多少来自 E，多少来自 F，多少来自 F^{-1}？

16 **注意。**为什么 $\overline{F}$ 的行 i 与 F 的 $N - i$ 行相同？(数字从 0 到 $N - 1$。)

17 什么是数字 0, 1, ..., 7 的**比特-反序**？将他们写成二进制(基底 2)的 000, 001, ..., 111 然后反转每个比特的顺序，这 8 个数字现在是_____。

请尊重版权与译者的劳动成果，侵权必究!

第十章
应用

10.1 图形与网络

最近这几年我常常看到一个模型，我发现它很基本很有用，所以我经常把它放在前面。这个模型包含由**边**连接的**节点**，称为**图形**(graph)。

图形通常是显示类型的函数 $f(x)$，点-边类型的图形产生矩阵。本段落主要讨论图形的“**关联矩阵(incidence matrix，也有翻译成入射矩阵)**”——这个告诉我们 n 个节点如何用 m 个边连接。通常 $m > n$，边比节点多。

对于 $m \times n$ 矩阵，有两个基础子空间在 $\mathbf{R}^n$，有两个在 $\mathbf{R}^m$，他们是 A 与 A^{T} 的行空间与零空间。他们的维度是 $r, n-r$ 与 $r, m-r$，来自线性代数最重要的定理，定理的第二部分是行空间与列空间的正交性。我们的目的是展示图形的例子来阐述线性代数的基础定理。

当我建立一个**图形**与它的**关联矩阵**，子空间的维度很容易观察，但是我们想要子空间本身——正交性有很大帮助。把子空间与图形的来源结合起来是非常重要的，利用特殊化的关联矩阵，**线性代数的法则**变成**基尔霍夫法则**。不要被“电流”与“电压”分心，这些矩形矩阵是最佳的。

关联矩阵的每个单元是 0 或 1 或–1，在消元法过程仍然会维持。所有的枢轴与乘数是±1，因此 $A = LU$ 的两个因子也只包含 0, 1, –1，零空间矩阵也是如此！全部 4 个子空间的基底向量有意想不到的简单分量。这些矩阵不是教科书捏造出来的，他们来自一个在纯数学与应用数学都是绝对重要的模型。

关联矩阵

图 10.1 显示一个 $m = 6$ 个边与 $n = 4$ 个节点的图形，这个 6×4 矩阵 A 告诉我们哪些点被哪些边连接。第一行的–1, 1, 0, 0 显示第一个边从节点 1 到节点 2(节点 1 的–1 表示箭头往外，节点 2 的 1 表示箭头往内。

A 的行数字是边的数字，列数字 1, 2, 3, 4 是节点的数字。

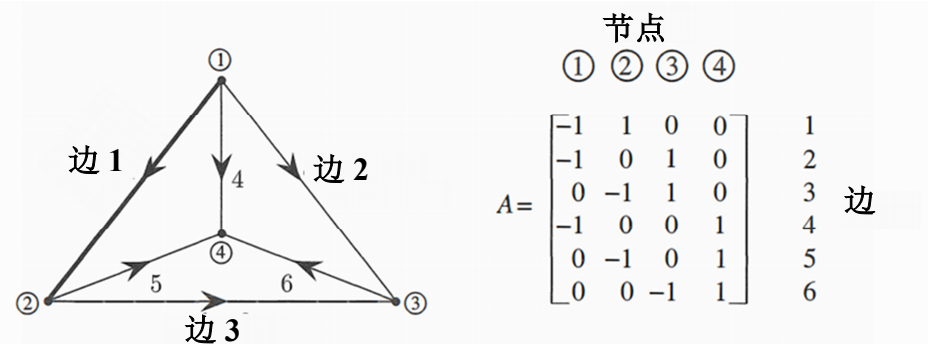


图 10.1：$m = 6$ 个边与 $n = 4$ 个节点的完整图形：6×4 的关联矩阵 A。

观察图形你就可以写出矩阵。第二个图形有相同的 4 个点但是只有 3 个边，它的关联矩阵 B 是 3×4。

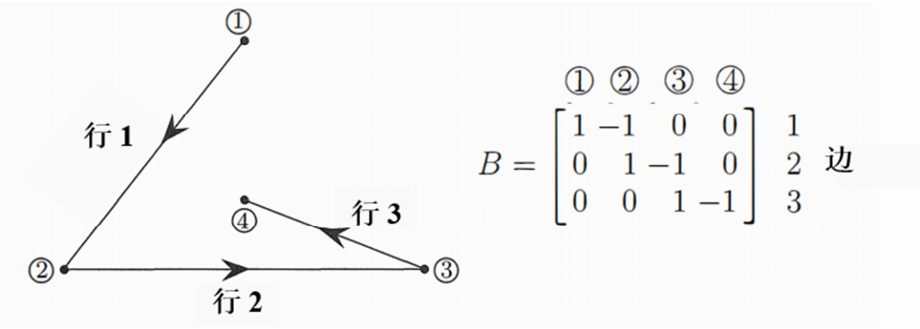


图 10.1*：3 个边与 4 个节与没有回圈的树，则 B 有独立的行。

第一个图形是完整--每个节点配对都有一个边相连，第二个图形是一个**树**(tree)--图形没有**闭回圈**(closed loop)，这是两个极端。边的最大个数是 $n(n-1)/2 = 6$ 以及最小保持连接的个数是 $n-1 = 3$。

消元法把每个图形简化成树，回圈产生 A 的相关行以及在阶梯形式 U 与 R 的零行。检视第一个图形中由边 1, 2, 3 构成的大回圈，会产生 U 的一个零行：

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

这些是典型的步骤，当边 1 与 2 共享节点 1 时，消元法得到没有节点 1 的“捷径(shortcut)边”。如果图形已经有这个捷径边制造出回圈，则消元法得到一整行的零。当灰尘清除我们看到树。

浮现一个概念： **当边形成回圈，行是相关**。无关的行来自树，这是行空间的关键。我们假设图形是连接的，箭头可以两边前进。在一个边，顺着箭头的流动是“正”，反方向的流动是负。这个流动可能是电流或信号或力量--甚至是油或是煤气或是水。

请尊重版权与译者的劳动成果，侵权必究!

当 x_1, x_2, x_3, x_4 是节点的电压，Ax 给出电压差：

$$Ax=\begin{bmatrix}-1&1&0&0\\-1&0&1&0\\0&-1&1&0\\-1&0&0&1\\0&-1&0&1\\0&0&-1&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}x_2-x_1\\x_3-x_1\\x_3-x_2\\x_4-x_1\\x_4-x_2\\x_4-x_3\end{bmatrix} \tag{1}$$

让我再说一次，关联矩阵 A 是差分矩阵，输入向量 $\boldsymbol{x}$ 给出电压，输出向量 $A\boldsymbol{x}$ 给出电压差(沿着边 1 到边 6)。如果电压相等，差是零，这个就是 A 的零空间。

1　零空间包含 $A\boldsymbol{x}=\boldsymbol{0}$ 的解，全部 6 个电压差都是零，表示：全部 4 个电压都相等。每个零空间中的 $\boldsymbol{x}$ 是一个**常数向量**：$\boldsymbol{x}=(c, c, c, c)$。$A$ 的零空间是$\mathbf{R}^n$的一条直线--它的维度是 $n-r=1$。

第二个关联矩阵 B 有相同的零空间，它包含**(1, 1, 1, 1)**：

一维零空间：
对应相同的树

$$B\boldsymbol{x}=\begin{bmatrix}-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

我们可以把所有的电压增加或降低相同的量 c 而不会改变之间的差，电压存在一个“**任意常数**”。把这个跟相同的函数叙述相比，我们可以对函数增加或降低 C 而不会改变导数。

微积分在**不定积分**加一个“$+C$”，图形原理把(c, c, c, c)加到向量 $\boldsymbol{x}$，线性代数把零空间中的任意向量 $\boldsymbol{x}_n$ 加到 $A\boldsymbol{x}=\boldsymbol{b}$ 的一个特定解。

当**定积分**从一个已知点开始，微积分的“$C+$”消失。同理当我们设定 $x_4=0$，零空间消失，未知数 x_4 被移除，而且矩阵 A 与 B 的第四列也消失(这些列用来乘 x_4)。电工程师会说节点 4 被**“接地”**了。

2　行空间包含全部 6 个行的组合，它的维度当然不是 6。方程式 $r+(n-r)=n$ 必须是 $3+1=4$。我们从消元过程看到秩 $r=3$，3 个边之后我们开始形成回圈！新的行不是无关。

我们如何分辨 $\boldsymbol{v}=(v_1, v_2, v_3, v_4)$是否在行空间中？慢的方法是组合行，快的方法是利用正交性：

$\boldsymbol{v}$ 在行空间中　若且唯若　它与零空间中的(1, 1 , 1, 1)垂直。

向量 $\boldsymbol{v}=(0, 1, 2, 3)$测试失败--它的分量总和是 6。向量 $\boldsymbol{v}=(-6, 1, 2, 3)$在行空间中：$-6+1+2+3=0$，这个向量等于 **6(行 1)+5(行 3)+3(行 6)**。

每个 A 的行的总和是零，对于行空间中的每个向量必须是成立的。

3 **列空间**包含全部 4 个列的组合，我们期待有 3 个无关列，因为它有 3 个无关行。A 的前 3 列无关(任意 3 列都是)，但是加上第四列得到零向量，再次说明(1, 1, 1, 1)在零空间。**我们如何分辨一个特定向量 $\boldsymbol{b}$ 是否在关联矩阵的列空间中？**

第一个答案：尝试求解 $A\boldsymbol{x}=\boldsymbol{b}$，这样会错失所有的内在。像以前一样，正交性给出较好的答案。我们现在来到基尔霍夫电路原理的两个著名法则--电压法则与电流法则(**KVL** 与 **KCL**)。他们是线性代数“法则”的自然表述，很高兴在这里看到**左零空间**的关键角色。

第二个答案： $A\boldsymbol{x}$ 是电压差 $x_i - x_j$ 的向量，如果我们沿着图形的一个封闭回圈把所有的差相加，他们抵消之后留下零。环绕边 1, 3, –2(箭头与边 2 的箭头反向)形成的大三角形，这些差会抵消：

差的总和是 0 $\qquad (x_2-x_1)+(x_3-x_2)-(x_3-x_1)=0$

基尔霍夫电压法则：$A\boldsymbol{x}=\boldsymbol{b}$ 的分量，环绕每个回圈的总和是零。

环绕大三角形： $\qquad b_1+b_3-b_2=0$

测试每个回圈，电压法则决定 $\boldsymbol{b}$ 是否在列空间。只有在 $\boldsymbol{b}$ 的分量与 A 的行一样满足所有相同的相关性时，$A\boldsymbol{x}=\boldsymbol{b}$ 才有解，消元法得到 $0=0$，而且 $A\boldsymbol{x}=\boldsymbol{b}$ 是一致。

4 **左零空间**包含 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 的解，它的维度是 $m-r=6-3$：

电流法则
$$A^{\mathrm{T}}\boldsymbol{y}=\begin{bmatrix}-1&-1&0&-1&0&0\\1&0&-1&0&-1&0\\0&1&1&0&0&-1\\0&0&0&1&1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\\y_4\\y_5\\y_6\end{bmatrix}=\begin{bmatrix}0\\0\\0\\0\end{bmatrix} \tag{2}$$

真实的方程式个数是 $r=3$ 不是 $n=4$，理由：4 个方程式的总和得到 $0=0$，第 4 个方程式自动来自前 3 个方程式。

这些方程式代表什么？第 1 个方程式说 $-y_1-y_2-y_4=0$，**进入节点 1 的净流量是零**。第 4 个方程式说 $y_4+y_5+y_6=0$，**进入节点 4 的流量减去离开的流量是零**。方程式 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 非常著名而且很基本：

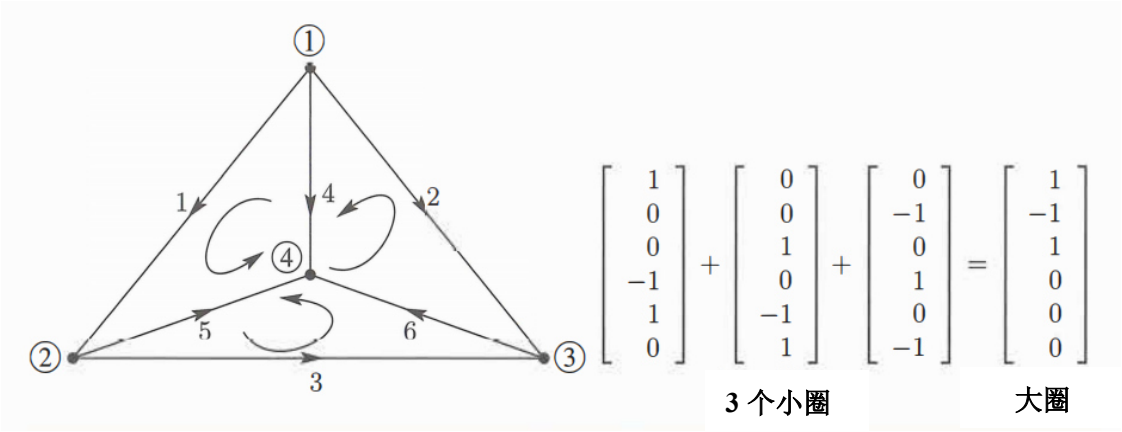
基尔霍夫电流法则：$A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$　在每个节点，流进等于流出。

这个法则值得排在应用数学的第一位，它表述了“守恒”与“连续”与“平衡”，没有流失也没有获得。当电流或是力量平衡时，要求解的方程式是 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$。注意这个美丽的事实，平衡方程式中的矩阵是关联矩阵 A 的转置。

请尊重版权与译者的劳动成果，侵权必究!

$A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 的真正解是什么？电流本身必须平衡，最简单的方法是**环绕一个回圈流动**。如果一个单位的电流环绕大三角形前进(前向边 1 与 3，反向边 2)，这 6 个电流是 $\boldsymbol{y}=(1,-1,1,0,0,0)$，这个满足 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$。每个回圈的电流是**电流法则**的解，在每个节点的流进等于流出。一个较小的回圈是前向边 1，前向边 5，反向边 4，则 $\boldsymbol{y}=(\mathbf{1},0,0,\mathbf{-1},\mathbf{1},0)$也在左零空间。

我们预期 3 个无关 $\boldsymbol{y}$'s: $m-r=6-3=3$，图形中的 3 个小圈是无关。大三角形看起来是第 4 个 $\boldsymbol{y}$，但是这个流量是较小圈流量的总和，环绕 3 个小圈的流量是左零空间的一组基底。



关联矩阵 A 来自有 n 个节点与 m 个边的连接图形，行空间与列空间有维度 $r=n-1$，A 与 A^{T} 的零空间有维度 1 与 $m-n+1$：

$\boldsymbol{N}(A)$	常数向量$(c,c,\ldots,c)$构成 A 的零空间：维度 $=1$。
$\boldsymbol{C}(A^{\mathrm{T}})$	任意树的边给出 A 的 r 个无关行：维度 $r=n-1$。
$\boldsymbol{C}(A)$	**电压法则**：环绕所有回圈的 $A\boldsymbol{x}$ 的分量总和是零：维度 $=n-1$。
$\boldsymbol{N}(A^{\mathrm{T}})$	**电流法则**：$A^{\mathrm{T}}\boldsymbol{y}=$**(流进) – (流出) = 0，利用回圈电流求解**。

对于平面的每个图形，线性代数得到**欧拉公式**：拓扑学第一定理！

(节点个数) – (边个数) + (小圈个数) = 1

这是 $(\boldsymbol{n})-(\boldsymbol{m})+(\boldsymbol{m}-\boldsymbol{n}+\mathbf{1})=\mathbf{1}$。我们例子中的图形是 4 – 6 + 3 = 1。

一个单一的三角形有(3 个节点) – (3 个边) + (1 个回圈)，在一个 9 边 10 节点无回圈的图形，欧拉计数是 10 – 9 + 0。所有的平面图形都会得到答案 1.

下个图形显示一个具有电流源的网络，基尔霍夫电流法则从 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 变成 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{f}$，目的在平衡来自外界的源 $\boldsymbol{f}$。每个节点的流进仍然等于流出，这 6 个边有电导(conductance) $c_1,\ldots,c_6$，以及电流源流进节点 1。电流源从节点 4 流出来保持整体的平衡(入=出)。问题是：**求出 6 个边上的电流 $y_1,\ldots,y_6$。**

网络中的流量现在引导我们从关联矩阵 A 到**拉普拉斯**矩阵 $A^{\mathrm{T}}A$。

电压与电流与 $A^{\mathrm{T}}Ax = f$

我们从节点电压 $\boldsymbol{x} = (x_1, \ldots, x_n)$开始，目前为止我们有 $A\boldsymbol{x}$ 求出沿着边的差 $x_i - x_j$，我们也有电流法则 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$ 来求出边的电流 $\boldsymbol{y} = (y_1, \ldots, y_m)$。如果网络中所有的电阻是 1，欧姆法则会适配 $\boldsymbol{y} = A\boldsymbol{x}$，则 $A^{\mathrm{T}}\boldsymbol{y} = A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{0}$。我们还未到达但是已经接近。如果没有电流源，$A\boldsymbol{x} = \boldsymbol{0}$ 的解会是无流量：$\boldsymbol{x} = \boldsymbol{0}$ 且 $\boldsymbol{y} = \boldsymbol{0}$。我可以看到 3 种方法产生 $\boldsymbol{x} \neq \boldsymbol{0}$ 与 $\boldsymbol{y} \neq \boldsymbol{0}$。

1　指定固定电压 x_i 到一个或多个节点。

2　在一个或多个边加入电池(电压源)。

3　进入一个或多个节点加入电流源，如图 10.2。

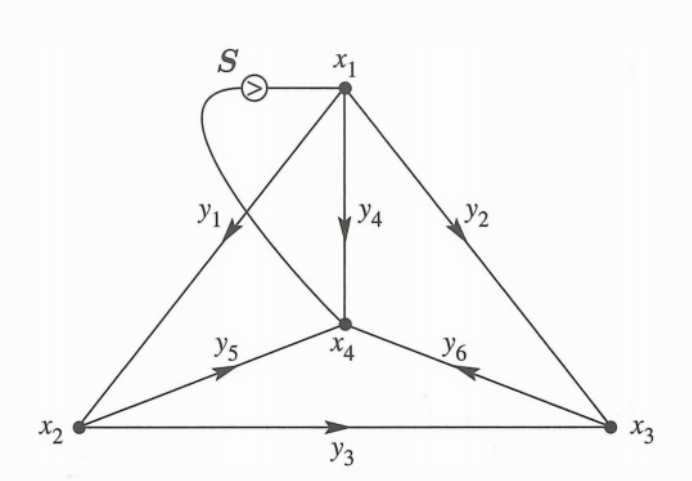


图 10.2：从节点 4 到 1 有一个电流源 S 的网络电流 y_1 到 y_6。

范例　图 10.2 引入一个从节点 4 到 1 的电流源 S，这个电流最终会通过网络回流到节点 4。某个电流 y_4 直接在边 4 前进，其他电流会走比较远的路程，比如从 1 到 2 到 4，或是从 1 到 3 到 4。利用对称，我预期从节点 2 到 3 没有电流($y_3 = 0$)，求解网络方程式会确认这个推论。**这些方程式的矩阵是 $A^{\mathrm{T}}A$，就是图形拉普拉斯矩阵：**

$$\begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}}_{A^{\mathrm{T}}A}$$

请尊重版权与译者的劳动成果，侵权必究！

拉普拉斯矩阵不可逆！因为(1, 1, 1, 1)在 A 与 $A^{\mathrm{T}}A$ 的零空间中，我们无法求解全部 4 个电势(potential)。有一个节点必须**接地**，令 $x_4 = 0$ 移除第 4 行与第 4 列，剩下一个 3×3 可逆矩阵。现在我们针对未知电势 x_1, x_2, x_3 求解 $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{f}$，其中电流源 S 进入节点 1：

$$\text{电压 } A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{f} \qquad \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} S \\ 0 \\ 0 \end{bmatrix} \quad \text{得到} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} S/2 \\ S/4 \\ S/4 \end{bmatrix}$$

$$\text{电压 } \boldsymbol{y} = -A\boldsymbol{x} \qquad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = - \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} S/2 \\ S/4 \\ S/4 \\ 0 \end{bmatrix} = \begin{bmatrix} S/4 \\ S/4 \\ 0 \\ S/2 \\ S/4 \\ S/4 \end{bmatrix}$$

半数的电流直接在边 4 前进，这是 $y_4 = S/2$。从节点 2 到 3 没有电流，对称性指出 $y_3 = 0$ 并且现在得到证明。

承认错误。我记得电流从高电压流向低电压，这个产生 $\boldsymbol{y} = -A\boldsymbol{x}$ 中的负号。当边的电阻不再全部是 1 时，欧姆法则的正确形式应该是 $R\boldsymbol{y} = -A\boldsymbol{x}$。电导比电阻简洁：$C = R^{-1} =$ 对角矩阵。**我们现在提出欧姆法则是 $\boldsymbol{y} = -CA\boldsymbol{x}$。**

网络与 $A^{\mathrm{T}}CA$

在实际的网络中，沿着边的电流是两个数的乘积，一个是边的两个端点电势 $\boldsymbol{x}$ 之间的差，电压是 $A\boldsymbol{x}$ 推动电流。另一个数是 c 是“**电导**”——量测流量通过的难易。

在物理与工程中，c 由材料决定。对电流来说，金属的 c 很高，塑胶的 c 很低。对超级导体来说，c 接近无限大。在经济学，c 量测一个边的容量或成本。

整理一下，从关联矩阵 A 可以得知图形，告诉我们节点-边的连接。网络更深入，会指定每个边一个电导 c，这些数字 $c_1, \ldots, c_m$ 进入**电导矩阵** C——对角矩阵。

对于电阻网络，电导是 1/(电阻)，除了基尔霍夫法则对应整体的电流系统之外，我们与欧姆法则对应每个电流。欧姆法则联结边 1 的电流 y_1 与电压差 $x_2 - x_1$：

欧姆法则：沿着边的电流 = 电导乘电压差

对于所有 m 个电流的欧姆法则是 $\boldsymbol{y} = -CA\boldsymbol{x}$，向量 $A\boldsymbol{x}$ 给出电势差，然后左乘电导 C。结合欧姆法则与基尔霍夫电流法则 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$，我们得到 $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{0}$。

这个几乎是网络流动的中心方程式，唯一有问题的是右侧的零！网络需要外接的能量——一个电压源或是电流源——才能发生一些事情。

符号的注解 我把电路理论的 $\boldsymbol{Ax}$ 改成 $-\boldsymbol{Ax}$，从高电势流向低电势，当 $x_1 - x_2$ 是正数时，有(正)电流从节点 1 到节点 2——$\boldsymbol{Ax}$ 的建立方式是 $x_2 - x_1$。在物理学与电工程的负号，在力学工程与经济学中是正号。$\boldsymbol{Ax}$ 对比 $-\boldsymbol{Ax}$ 一直是头痛的问题，但是不是无法避免的。

应用数学的注解 每个新的应用有它自己的欧姆法则形式，对弹簧来说它是胡克法则，应力 $\boldsymbol{y}$ 是(弹力 C)乘(拉伸 $\boldsymbol{Ax}$)。对于热传导来说，$\boldsymbol{Ax}$ 是温度梯度。对油流量来说它是一个压力梯度。对统计学的最小平方递归(第十二章)C^{-1} 是协方差矩阵。

我的教科书：应用数学与计算科学与工程(Wellesley-cambridge Press)实际建立在 $A^{\mathrm{T}}CA$，这是矩阵以及微分方程式平衡的关键。应用数学比表面看起来还要有组织！新的问题中我学习了关注 $A^{\mathrm{T}}CA$。

问题集 10.1

问题 1-7 与 8-14 有关这些图形的关联矩阵。



1 写出三角形图形的 3×3 关联矩阵 A，第一行有 -1 在列 1 以及 $+1$ 在列 2。什么向量 (x_1, x_2, x_3) 在它的零空间中？你如何知道 $(1, 0, 0)$ 不在它的行空间中？

2 写出三角形图形的 A^{T}，求出零空间中的向量 $\boldsymbol{y}$。$\boldsymbol{y}$ 的分量是边上的电流——环绕这个三角形的电流为何？

3 从第 3 个方程式消去 x_1 与 x_2 得到阶梯矩阵 U，什么树对应到 U 的两个非零行？

$$-x_1 + x_2 = b_1$$
$$-x_1 + x_3 = b_2$$
$$-x_2 + x_3 = b_3$$

请尊重版权与译者的劳动成果，侵权必究！

4 选择一个向量(b_1, b_2, b_3)使得 $A\boldsymbol{x} = \boldsymbol{b}$ 有解，然后另一个向量 $\boldsymbol{b}$ 得到无解。这些 $\boldsymbol{b}$'s 与$(1, -1, 1)$的关系为何？

5 选择一个向量(y_1, y_2, y_3)使得 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{f}$ 有解，然后另一个向量 $\boldsymbol{f}$ 得到无解。这些 $\boldsymbol{f}$'s 与$(1, 1, 1)$的关系为何？方程式 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{f}$ 是基尔霍夫_____法则。

6 计算矩阵乘积得到 $A^{\mathrm{T}}A$。选择一个向量 $\boldsymbol{f}$ 使得 $A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{f}$ 有解，求出 $\boldsymbol{x}$。把这些电势 $\boldsymbol{x}$ 与电流 $\boldsymbol{y} = -A\boldsymbol{x}$ 以及电流源 $\boldsymbol{f}$ 放进三角图形中，电导是 1 因为 $C = I$。

7 电导 $c_1 = 1$ 与 $c_2 = c_3 = 2$，计算乘积 $A^{\mathrm{T}}CA$。的 $\boldsymbol{f} = (1, 0 -1)$，求出 $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f}$ 的解。把电势 $\boldsymbol{x}$ 与电流 $\boldsymbol{y} = -CA\boldsymbol{x}$ 放进三角图形中，此时电流源 $\boldsymbol{f}$ 从节点 3 出，进入节点 1。

8 方形图形有两个回圈，写出对应的 5×4 关联矩阵 A。求出 $A\boldsymbol{x} = \boldsymbol{0}$ 的一个解以及 $A^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$ 的两个解。

9 找出 b's 的两个要求使得 5 个差 $x_2 - x_1, x_3 - x_1, x_3 - x_2, x_4 - x_2, x_4 - x_3$ 等于 b_1, b_2, b_3, b_4, b_5。你已经发现在图形中环绕两个_____的基尔霍夫_____法则。

10 简化 A 得到阶梯形式 U，三个非零行得到哪个图形的关联矩阵？在方形图形中你发现一个树--找出其他 7 个树。

11 计算矩阵乘积得到 $A^{\mathrm{T}}A$，从图形猜测它的单元为何:

(a) $A^{\mathrm{T}}A$ 的对角线说明有多少_____进入美国节点。

(b) 非对角线的–1 或 0 说明哪些节点配对是_____。

12 为什么每个叙述对于 $A^{\mathrm{T}}A$ 都是正确的？答案是针对 $A^{\mathrm{T}}A$ 不是 A。

(a) 它的零空间包含$(1, 1, 1, 1)$，它的秩是 $n - 1$。

(b) 它是正半定不是正定。

(c) 它的 4 个特征值是实数且他们的符号是_____。

13 电导 $c_1 = c_2 = 2$ 与 $c_3 = c_4 = c_5 = 3$，计算矩阵乘积 $A^{\mathrm{T}}CA$。求出 $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f} = (1, 0, 0, -1)$的解。把在节点与边的电势 $\boldsymbol{x}$ 与电流 $\boldsymbol{y} = -CA\boldsymbol{x}$ 放进方形图形中。

14 矩阵 $A^{\mathrm{T}}CA$ 不可逆，什么向量 $\boldsymbol{x}$ 在它的零空间中？为什么 $A^{\mathrm{T}}CA\boldsymbol{x} = \boldsymbol{f}$ 有一个解若且唯若 $f_1 + f_2 + f_3 + f_4 = 0$?

15 一个 7 节点 7 边的连接图形有多少回圈？

16 对于 4 节点 6 边 3 回圈的图形加上一个新节点，如果你把它连接到一个旧节点，欧拉公式变成(　　) – (　　) + (　　) = 1。如果你把它连接到两个旧节点，欧拉变成(　　) – (　　) + (　　) = 1。

网址：linear.neocities.org　邮箱：876152660@qq.com

17 假设 A 是来自一个连接图形(但是未知)的 12×9 关联矩阵，

(a) A 有多少无关列？

(b) 对于 $\boldsymbol{f}$ 什么样的条件使得 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{f}$ 有解？

(c) $A^{\mathrm{T}}A$ 的对角线单元给出进入每个节点的边的个数，这些对角线单元的总和为何？

18 为什么一个完整的 6 节点图形有 15 个边？一个连接 6 个点的树有几个边？

注解 **化学计量矩阵(stoichiometric)**是化学中相当重要的“广义”关联矩阵，它的单元显示每个化学片段(每个列)如何进入每个反应(每个行)。

请尊重版权与译者的劳动成果，侵权必究!

10.2 工程中的矩阵

这个段落展示工程问题如何产生对称矩阵 K(通常 K 是正定)，对称与正定的“线性代数理由”是它的形式 $K = A^{\mathrm{T}}A$ 与 $K = A^{\mathrm{T}}CA$。“物理理由”是$\frac{1}{2}\boldsymbol{u}^{\mathrm{T}}K\mathbf{u}$ 的叙述代表能量--能量永远不是负数。矩阵 C，通常是对角线，包含正的物理常数，例如电导或是刚度或是扩散率。

我们的最佳范例来自机械与土木与航空工程，K 是**刚度矩阵**，$K^{-1}\boldsymbol{f}$ 是对应外界力量 $\boldsymbol{f}$ 的结构反应。段落 10.1 转至电工程--矩阵来自网络与电路，我会继续跟上与化学工程有关的练习题！经济学与管理学与工程设计会在本章的后面来到(关键是最佳化(optimization))。

工程有两个方式引导至线性代数，直接与间接：

直接方式　物理问题只有有限个数的片段，联结他们的位置或速度的法则是线性(移动不会太大或太快)，利用矩阵方程式来表述法则。

间接方式　物理系统是“连续”，不是个别的质量，质量密度与力量与速度是 x 或 x, y 或 x, y, z 的函数。利用微分方程式来表述法则。**要找到精确的解，我们利用有限的差方程式或无限的元素方程式来近似**。

两种方式都产生矩阵方程式与线性代数，我真正相信没有矩阵你是无法处理现代的过程。

此处我们表述平衡方程式 $K\boldsymbol{u} = f$，加入运动，$M d^2\boldsymbol{u}/dt^2 + K\boldsymbol{u} = \boldsymbol{f}$ 变成动态，然后我们利用来自 $K\boldsymbol{x} = \lambda M\boldsymbol{x}$ 的特征值，或是有限差分(finite difference)。

微分方程式到矩阵方程式

微分方程式是连续的，我们的基础范例是 $-d^2u/dt^2 = f(x)$。矩阵方程式是离散的，我们的基础范例是 $K_0\boldsymbol{u} = \boldsymbol{f}$。利用第二导数到第二差的步骤，你在一个非常短的空间会看到一个大图。从两个端点 $x = 0$ 与 $x = 1$ 的固定边界条件(fixed boundary condition)开始：

固定-固定边界值问题　$-\dfrac{d^2u}{dt^2} = 1$，其中 $u(0) = 0$ 与 $u(1) = 0$　(1)

这个微分方程式是线性，一个特定解是 $u_p = -\frac{1}{2}x^2$(则 $d^2u/dx^2 = -1$)，我们可以加上任意“零空间中”的函数。我们不是求解 $A\boldsymbol{x} = \boldsymbol{0}$ 的一个向量 $\boldsymbol{x}$，而是求解 $-d^2u/dx^2 = 0$ 的一个函数 $u_n(x)$。(重点：右侧是零。)

零空间的解是 $u_n(x) = C + Dx$ (对应一个二阶微分方程式的 2 维零空间)，完整解是 $u_p + u_n$：

完整解 $-\frac{d^2u}{dx^2}=1$ $u(x)=-\frac{1}{2}x^2+C+Dx$ (2)

现在从两个边界条件求 C 与 D：令 $x=0$ 然后 $x=1$，在 $x=0$，$u(0)=0$ 强制 $C=0$，在 $x=1$，$u(1)=0$ 强制 $-\frac{1}{2}+D=0$，$\boldsymbol{D=\frac{1}{2}}$：

$$u(x)=-\tfrac{1}{2}x^2+\tfrac{1}{2}x=\boldsymbol{\tfrac{1}{2}(x-x^2)}$$ 求解了固定-固定边界值问题 (3)

差分取代微分

要得到矩阵而不是导数，我们有 3 个基础选择--前向或反向或中心**差分**，从第一导数与第一差分开始：【对应微分，取名“差分”】

$$\frac{du}{dx}\approx\frac{u(x+\Delta x)-u(x)}{\Delta x}\text{或是}\frac{u(x)-u(x-\Delta x)}{\Delta x}\text{或是}\frac{u(x+\Delta x)-u(x-\Delta x)}{2\Delta x}$$

在 $x=0$ 与 $x=1$ 之间，我们把区间分成 $n+1$ 个相等片段，这些片段的宽度是 $\Delta x=1/(n+1)$，在这 n 个断点 $\Delta x, 2\Delta x, \ldots$ 的 u 值会是我们矩阵方程式 $K\boldsymbol{u}=\boldsymbol{f}$ 中的未知数 u_1 到 u_n：

需要计算的解：$\boldsymbol{u}=(u_1,u_2,\ldots,u_n)\approx(u(\Delta x),u(2\Delta x),\ldots,u(n\Delta x))$。

零值 $u_0=u_{n+1}=0$ 来自边界值 $u(0)=u(1)=0$。

把**前向差分与反向差分**取代 $-\frac{d}{dx}\left(\frac{du}{dx}\right)=1$ 中的导数：

$$\frac{1}{(\Delta x)^2}\begin{bmatrix}1&-1&0&0\\0&1&-1&0\\0&0&1&-1\end{bmatrix}\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&\boxed{1}\\0&0&-1\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}=\begin{bmatrix}1\\1\\1\end{bmatrix}\tag{4}$$

【原文**粗体 1 部分**写成–1 是错误】

这是我们的矩阵方程式，其中 $n=3$ 与 $\Delta x=1/4$。这两个第一差分是彼此的转置！方程式是 $A^{\mathrm{T}}A\boldsymbol{u}=(\Delta x)^2\boldsymbol{f}$。当我们计算 $A^{\mathrm{T}}A$，我们得到正定第二差分矩阵 K_0：

$\boldsymbol{K_0u=(\Delta x)^2f}$ $\begin{bmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}=\frac{1}{16}\begin{bmatrix}1\\1\\1\end{bmatrix}$ 得到 $\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}=\frac{1}{32}\begin{bmatrix}3\\4\\3\end{bmatrix}$ (5)

请尊重版权与译者的劳动成果，侵权必究！

这个例子有一个很奇妙的事实，就是这些数字 u_1, u_2, u_3 恰好就是正确值！他们与真实的解 $u(x) = \frac{1}{2}(x - x^2)$在这 3 个网格点 $x = \frac{1}{4}, \frac{2}{4}, \frac{3}{4}$的值一致。图 10.3 显示真实解(连续曲线)与近似解 u_1, u_2, u_3(恰好位于曲线上)，这个曲线是抛物线。

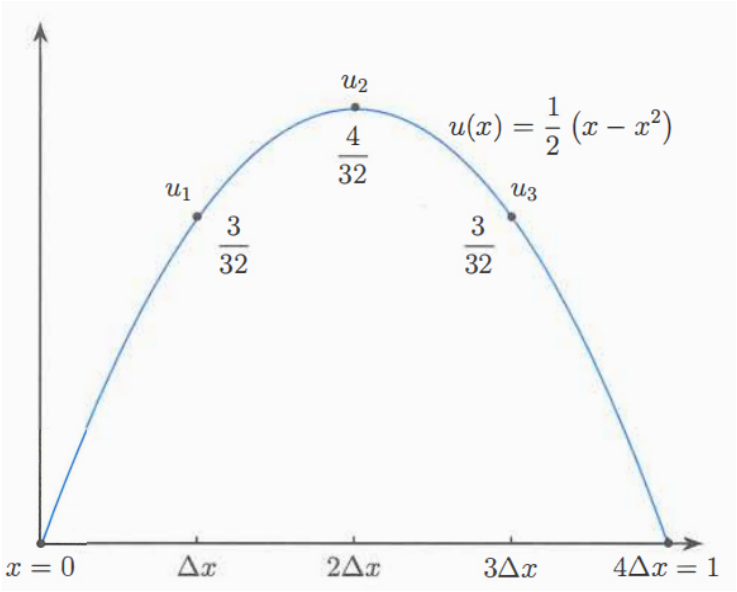


图 10.3：$-\frac{d^2u}{dx^2} = 1$的解与固定-固定边界 $K_0\boldsymbol{u} = (\Delta x)^2\boldsymbol{f}$的解。

如何解释恰好落在 $u(x)$图形上的完美答案？在这矩阵方程式中，$K_0 = A^{\mathrm{T}}A$ 是一个“第二差分矩阵”，它给出了$-\frac{d^2u}{dx^2}$的中心近似。我引入负号是因为第一导数是反对称，第二导数本身就是负数：

$$\frac{d}{dx}\text{的“转置”是}-\frac{d}{dx}\text{，则}(-\frac{d}{dx})(\frac{d}{dx})\text{是正定}$$

你在矩阵 A 与 A^{T} 可以看到，(A =前向差分)的转置是(A^{T} = –反向差分)。我不希望选择一个中心 $u(x + \Delta x) - u(x - \Delta x)$，中心是第一差分的最佳，但是第二差分 $A^{\mathrm{T}}A$ 会从 $u(x + 2\Delta x)$拉伸成为 $u(x - 2\Delta x)$：不好。

现在我们可以说明这个完美答案，确切在真实曲线 $u(x) = ½(x - x^2)$上。第二差分–1, 2, –1 对于直线 $y = x$ 与抛物线是真实的正确！

$$\boldsymbol{y = x} \qquad -\frac{d^2y}{dx^2} = 0 \qquad -(x + \Delta x) + 2x - (x - \Delta x) = 0(\Delta x)^2$$

$$\boldsymbol{y = x^2} \qquad -\frac{d^2y}{dx^2} = -2 \qquad -(x + \Delta x)^2 + 2x^2 - (x - \Delta x)^2 = -2(\Delta x)^2$$

这个奇迹可以延续到 $y = x^3$，正确的$-d^2y/dx^2 = -6$ 是从第二差分获得。但是对于 $y = x^4$ 我们回到了地球，第二差分不再确切适配$-y'' = -12x^2$。这些近似 u_1, u_2, u_3 不会落在 $u(x)$的图形上。

固定端与自由端与变数系数 $c(x)$

要看到两个新的可能性，我会改变方程式只有一个边界条件：

$$-\frac{d}{dx}\left((1+x)\frac{du}{dx}\right)=f(x),\qquad \text{其中 } u(0)=0 \text{ 与 } \frac{du}{dx}(1)=0 \tag{6}$$

端点 $x=1$ 现在是**自由**，在这里没有支撑。“悬吊杆只在顶端固定”，自由端 $x=1$ 不存在力。这个转成 $du/dx=0$ 而不是固定条件 $x=1$ 时，$u=0$。

另一个变化是系数 $c(x)=1+x$，棒子的刚度随着 $x=0$ 到 $x=1$ 变化，可能它的宽度改变，或是材料改变。系数 $1+x$ 带来差分方程式的新矩阵。

由于 u_4 不再固定在 0，它变成新的未知数。反向差分 A 是 4×4，而且左乘 $c(x)=1+x$ 变成对角矩阵 C--在网格点左乘 $1+\Delta x, \ldots, 1+4\Delta x$。下列是 A^{T}, C 与 A：

$$A^{\mathrm{T}}CA=\begin{bmatrix}1&-1&0&0\\0&1&-1&0\\0&0&1&-1\\0&0&0&1\end{bmatrix}\begin{bmatrix}1.25&&&\\&1.5&&\\&&1.75&\\&&&2.0\end{bmatrix}\begin{bmatrix}1&0&0&0\\-1&1&0&0\\0&-1&1&0\\0&0&-1&1\end{bmatrix} \tag{7}$$

矩阵 $K=A^{\mathrm{T}}CA$ 会是对称且正定！对称是因为$(A^{\mathrm{T}}CA)^{\mathrm{T}}=A^{\mathrm{T}}CA^{\mathrm{TT}}=A^{\mathrm{T}}CA$。正定是因为它通过能量测试：$A$ 有无关的列，所以当 $\boldsymbol{x}\neq\boldsymbol{0}$ 就有 $A\boldsymbol{x}\neq\boldsymbol{0}$。

当每个 $\boldsymbol{x}\neq\boldsymbol{0}$ 时就有 $A\boldsymbol{x}\neq\boldsymbol{0}$，**能量** $=\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}CA\boldsymbol{x}=(A\boldsymbol{x})^{\mathrm{T}}C(A\boldsymbol{x})>0$，

当你针对固定-自由组合去计算矩阵乘积 $A^{\mathrm{T}}A$ 与 $A^{\mathrm{T}}CA$，观察 $A^{\mathrm{T}}A$ 的最后角落的 1 是如何取代 2。第 4 个方程式有 u_4-u_3，一个第一(不是第二)差分来自自由边界条件 $du/dx=0$。

注意在 $A^{\mathrm{T}}CA$ 中，c_1, c_2, c_3, c_4 是如何来自方程式(7)的 $c(x)=1+x$，先前这些 c's 就是简单的 1, 1, 1, 1，此处是**固定-自由**矩阵：

$$A^{\mathrm{T}}A=\begin{bmatrix}2&-1&0&0\\-1&2&-1&0\\&-1&2&-1\\&&-1&\mathbf{1}\end{bmatrix}\qquad A^{\mathrm{T}}CA=\begin{bmatrix}c_1+c_2&-c_2&0&0\\-c_2&c_2+c_3&-c_3&0\\0&-c_3&c_3+c_4&-c_4\\0&0&-c_4&\boldsymbol{c_4}\end{bmatrix} \tag{8}$$

请尊重版权与译者的劳动成果，侵权必究!

自由-自由边界条件

假设棒子的两端都是自由，现在 $x=0$ 与 $x=1$ 都有 $du/dx=0$，两端都没有把棒子固定！物理上来说它是不稳定的--没有外力之下可以移动。数学上来说所有的常数函数，比如 $u=1$，都满足这些自由条件。**代数上来说，我们的矩阵 $A^{\mathrm{T}}A$ 与 $A^{\mathrm{T}}CA$ 是不可逆：**

自由-自由范例
未知数 u_0, u_1, u_2
$\Delta x = 0.5$

$$A^{\mathrm{T}}A=\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \qquad A^{\mathrm{T}}CA=\begin{bmatrix} c_0 & -c_0 & \\ -c_0 & c_0+c_1 & -c_1 \\ & -c_1 & c_1 \end{bmatrix}$$

向量(1, 1, 1)在两个零空间中，这个适配了连续问题中的 $u(x)=1$。自由-自由 $A^{\mathrm{T}}A\boldsymbol{u}=\boldsymbol{f}$ 与 $A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f}$ 通常是无解。

在说明更多物理范例之前，我可以写下 6 个这样的矩阵吗？三对角 K_0 在本书出现很多次，现在我们来看它的应用。这些矩阵都是对称，前 4 个正交是正定：

$$K_0=A_0^{\mathrm{T}}A_0=\begin{bmatrix} \mathbf{2} & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{2} \end{bmatrix} \qquad A_0^{\mathrm{T}}C_0A_0=\begin{bmatrix} c_1+c_2 & -c_2 & \\ -c_2 & c_2+c_3 & -c_3 \\ & -c_3 & c_3+c_4 \end{bmatrix}$$

固定-固定 **包含弹簧常数**

$$K_1=A_1^{\mathrm{T}}A_1=\begin{bmatrix} \mathbf{2} & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{1} \end{bmatrix} \qquad A_1^{\mathrm{T}}C_1A_1=\begin{bmatrix} c_1+c_2 & -c_2 & \\ -c_2 & c_2+c_3 & -c_3 \\ & -c_3 & c_3 \end{bmatrix}$$

固定-自由 **包含弹簧常数**

$$K_{\text{奇异}}=\begin{bmatrix} \mathbf{1} & -1 & \\ -1 & 2 & -1 \\ & -1 & \mathbf{1} \end{bmatrix} \qquad K_{\text{圆圈}}=\begin{bmatrix} 2 & -1 & \mathbf{-1} \\ -1 & 2 & -1 \\ \mathbf{-1} & -1 & 2 \end{bmatrix}$$

自由-自由 **周期性 $\boldsymbol{u(0)=u(1)}$**

为了简化，矩阵 $K_0, K_1, K_{\text{奇异}}$ 与 $K_{\text{圆圈}}$ 都有 $C=I$，表示所有的“弹簧常数”是 $c_i=1$。我们引入 $A_0^{\mathrm{T}}C_0A_0$ 与 $A_1^{\mathrm{T}}C_1A_1$ 来展示弹簧常数如何进入矩阵(没有改变它的正定性质)。我们下一个目标是在其他工程问题中观察相同的刚度矩阵。

弹簧线与质量

图 10.4 展示一直线的弹簧连接 3 个质量 m_1, m_2, m_3，固定-固定案例有 4 个弹簧，其中顶端与底部固定，这个得到 K_0 与 $A_0^{\mathrm{T}}C_0A_0$。固定-自由案例只有 3 个弹簧；最下方的质量自由悬挂，这个会得到 K_1 与 $A_1^{\mathrm{T}}C_1A_1$。**自由-自由**问题得到 $\boldsymbol{K}_{\text{奇异}}$。

我们想要质量移动 $\boldsymbol{u}$ 与弹簧张力 $\boldsymbol{y}$ 的方程式：

$$\boldsymbol{u} = (u_1, u_2, u_3) = \textbf{质量的移动}\ (\text{往下为正})$$

$$\boldsymbol{y} = (y_1, y_2, y_3, y_4)\ \text{或}\ (y_1, y_2, y_3) = \textbf{弹簧的张力}$$

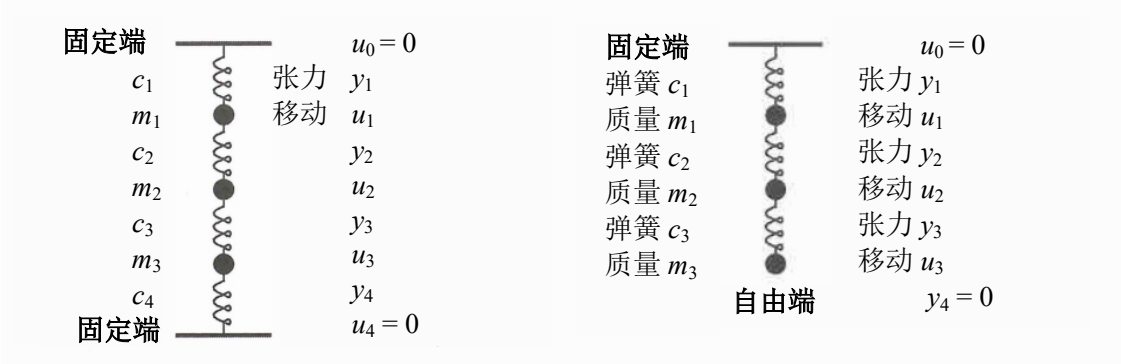


图 10.4：弹簧线与质量：**固定-固定**与**固定-自由**端

当一个质量往下移动，它的位移是正数($u_j > 0$)。对于弹簧来说，张力是正，压缩是负($y_i < 0$)。在张力时，弹簧拉伸，所以把质量往内拉，每个弹簧由自己的胡克法则 $y = ce$ 控制：(拉伸力 $\boldsymbol{y}$) = (弹簧常数 $\boldsymbol{c}$) 乘 (拉伸距离 $\boldsymbol{e}$)。

我们的工作是把这些单一弹簧方程式 $y = ce$ 变成整个系统的向量方程式 $K\boldsymbol{u} = \boldsymbol{f}$，力向量 $\boldsymbol{f}$ 来自引力，引力常数 g 会乘每个质量得到往下的力 $\boldsymbol{f} = (m_1g, m_2g, m_3g)$。

真正的问题在于找出刚度矩阵(**固定-固定**与**固定-自由**)，用 3 个步骤建立最佳的 K，不是一个步骤。我们不把移动 u_j 直接与力 f_i 直接产生关联，我们连接下表的每个向量会更好：

$$\begin{aligned}
\boldsymbol{u} &= \text{每个质量的}\textbf{移动} &&= (u_1, \ldots, u_n)\\
\boldsymbol{e} &= \text{每个弹簧的}\textbf{伸长} &&= (e_1, \ldots, e_m)\\
\boldsymbol{y} &= \text{每个质量的}\textbf{内力} &&= (y_1, \ldots, y_m)\\
\boldsymbol{f} &= \text{每个质量的}\textbf{外力} &&= (f_1, \ldots, f_n)
\end{aligned}$$

应用数学伟大的框架把 $\boldsymbol{u}$ 连接到 $\boldsymbol{e}$ 到 $\boldsymbol{y}$ 到 $\boldsymbol{f}$，则 $A^{\mathrm{T}}CA\boldsymbol{u} = \boldsymbol{f}$：

$$\begin{array}{ccc}
\boxed{\boldsymbol{u}} & & \boxed{\boldsymbol{f}} \\
A\downarrow & & \uparrow A^{\mathrm{T}} \\
\boxed{\boldsymbol{e}} & \xrightarrow{C} & \boxed{\boldsymbol{y}}
\end{array}
\qquad
\begin{array}{ll}
\boldsymbol{e} = A\boldsymbol{u} & A\ \text{是}\ m\times n \\
\boldsymbol{y} = C\boldsymbol{e} & C\ \text{是}\ m\times m \\
\boldsymbol{f} = A^{\mathrm{T}}\boldsymbol{y} & A^{\mathrm{T}}\ \text{是}\ n\times m
\end{array}$$

请尊重版权与译者的劳动成果，侵权必究！

我们会写下这两个例子的矩阵 A 与 C 与 A^{T}，首先是两端固定然后是下端自由。原谅这些矩阵的简洁性，这是他们的重要形式，特别是 A 与 A^{T} 在一起的时候。

伸长度 $\boldsymbol{e}$ 是拉伸距离——弹簧被延伸多少，原始的拉伸是零——系统落在桌面上。当它变成铅锤面竖直，引力开始作用。这些质量往下移动 u_1, u_2, u_3 的距离，每个弹簧被拉伸或压缩 $e_i = u_i - u_{i-1}$，这是端点之间的位移差：

每个弹簧的拉伸

$$
\begin{array}{lll}
\text{第一弹簧} & \boldsymbol{e}_1 = \boldsymbol{u}_1 & (\text{顶端固定所以 } u_0 = 0) \\
\text{第二弹簧} & \boldsymbol{e}_2 = \boldsymbol{u}_2 - \boldsymbol{u}_1 & \\
\text{第三弹簧} & \boldsymbol{e}_3 = \boldsymbol{u}_3 - \boldsymbol{u}_2 & \\
\text{第四弹簧} & \boldsymbol{e}_4 = \quad -\boldsymbol{u}_3 & (\text{底部固定所以 } u_4 = 0)
\end{array}
$$

如果两端移动相同的距离，弹簧就没有拉伸：$u_j = u_{j-1}$ 以及 $e_j = 0$。这 4 个方程式的矩阵是 4×3 的差分矩阵 A 以及 $\boldsymbol{e} = A\boldsymbol{u}$：

$$
\text{拉伸矩阵 (伸长度)} \quad \boldsymbol{e} = A\boldsymbol{u} \text{ 是 } \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{9}
$$

下一个方程式 $\boldsymbol{y} = C\boldsymbol{e}$ 连接弹簧伸长度 $\boldsymbol{e}$ 与弹簧张力 $\boldsymbol{y}$，这是对于每个个别弹簧的胡克法则 $y_i = c_i e_i$，它是与弹簧材料相关的“本质法则”。软的弹簧有小的 $\boldsymbol{c}$，所以一般的力也会产生很大的伸长度 e。在弹簧没有被过度拉伸以及材料塑化之前，胡克线性法则的精确度非常接近实际弹簧。

由于美国弹簧有自己的法则，$\boldsymbol{y} = C\boldsymbol{e}$ 中的矩阵是对角矩阵 C：

$$
\text{胡克法则 } \boldsymbol{y} = C\boldsymbol{e} \quad \begin{array}{l} y_1 = c_1 e_1 \\ y_2 = c_2 e_2 \\ y_3 = c_3 e_3 \\ y_4 = c_4 e_4 \end{array} \text{ 是 } \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} c_1 & & & \\ & c_2 & & \\ & & c_3 & \\ & & & c_4 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \tag{10}
$$

结合 $\boldsymbol{e} = A\boldsymbol{u}$ 与 $\boldsymbol{y} = C\boldsymbol{e}$，弹簧力(张力)是 $\boldsymbol{y} = CA\boldsymbol{u}$。

最终来到平衡方程式，应用数学的最基础法则。来自弹簧的内力与质量上的外力取得平衡，每个质量被上方的弹簧力 y_j **拉或推**，在下方它感觉到弹簧力 y_{j+1} 加上引力 f_j，于是 $y_j = y_{j+1} + f_j$ 或 $f_j = y_j - y_{j+1}$：

$$
\text{力平衡 } \boldsymbol{f} = A^{\mathrm{T}}\boldsymbol{y} \quad \begin{array}{l} f_1 = y_1 - y_2 \\ f_2 = y_2 - y_3 \\ f_3 = y_3 - y_4 \end{array} \text{ 是 } \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \tag{11}
$$

这个矩阵是 $\boldsymbol{A^{\mathrm{T}}}$！力平衡的方程式是 $\boldsymbol{f} = A^{\mathrm{T}}\boldsymbol{y}$，自然转置 $\boldsymbol{e}$ 到 $\boldsymbol{u}$ 矩阵的行与列，得到 $\boldsymbol{f}$ 到 $\boldsymbol{y}$ 的矩阵。这是框架之美，A^{T} 追随 A 出现，这 3 个方程式结合在一起就是 $K\boldsymbol{u} = \boldsymbol{f}$。

$$\left\{\begin{array}{l} \boldsymbol{e}=A\boldsymbol{u} \\ \boldsymbol{y}=C\boldsymbol{e} \\ \boldsymbol{f}=A^{\mathrm{T}}\boldsymbol{y} \end{array}\right\} \quad \begin{array}{l} \text{结合成 } A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f} \text{ 或是 } K\boldsymbol{u}=\boldsymbol{f} \\ K=A^{\mathrm{T}}CA \text{ 是刚度矩阵(力学)} \\ K=A^{\mathrm{T}}CA \text{ 是电导矩阵(网络)} \end{array}$$

有限元素程序花费很大的功夫从成千上万的小片段来组合 $K=A^{\mathrm{T}}CA$，我们发现这 4 个弹簧的 K(**固定-固定**)是 A^{T} 乘 CA：

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} c_1 & 0 & 0 \\ -c_2 & c_2 & 0 \\ 0 & -c_3 & c_3 \\ 0 & 0 & -c_4 \end{bmatrix}=\begin{bmatrix} c_1+c_2 & -c_2 & 0 \\ -c_2 & c_2+c_3 & -c_3 \\ 0 & -c_3 & c_3+c_4 \end{bmatrix}$$

如果所有的弹簧都一样，$c_1=c_2=c_3=c_4=1$，则 $C=I$，刚度矩阵简化成 $A^{\mathrm{T}}A$，它变成特殊的–1, 2, –1 矩阵 K_0。

注意来自工程的 $A^{\mathrm{T}}A$ 与来自线性代数的 LU，来自 4 个弹簧的矩阵 A 是 4×3，来自消元法的三角矩阵是方形。刚度矩阵 K 是由 $A^{\mathrm{T}}A$ 组合而成，再拆散成为 LU。应用数学的一步，其他都是计算数学。每个 K 来自矩形矩阵然后分解成方形矩阵。

我可以列出 $K=A^{\mathrm{T}}CA$ 的一些性质吗？你已经知道大部分：

1. K 是**三对角**，因为质量 3 没有连接到 1。
2. K 是**对称**，因为 C 是对称且 A^{T} 跟随 A。
3. K 是**正定**，因为 $c_i>0$ 且 A 有无关列。
4. K^{-1} 是**满(full)矩阵**(不是稀疏)，所有的**单元都是正数**。

性质 4 得到一个关于 $\boldsymbol{u}=K^{-1}\boldsymbol{f}$ 的重要事实：若所有的力往下作用($f_i>0$)，则所有的移动往下($u_i>0$)。注意“正”与“正定”不相同，K^{-1} 是正(K 不是)，但是两者都是正定。

范例 1　假设所有的 $c_i=c$ 与 $m_j=m$，求出移动 $\boldsymbol{u}$ 与张力 $\boldsymbol{y}$。

所有的弹簧都一样而且质量都一样，但是所有的移动与伸长度与张力不会相同。K^{-1} 包含 $1/c$ 因为 $A^{\mathrm{T}}CA$ 包含 c：

$$\textbf{移动}\qquad \boldsymbol{u}=K^{-1}\boldsymbol{f}=\frac{1}{4c}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} mg \\ mg \\ mg \end{bmatrix}=\frac{mg}{c}\begin{bmatrix} 3/2 \\ 2 \\ 3/2 \end{bmatrix}$$

中间质量的位移 u_2 比 u_1 与 u_3 大，单位是正确的：力 mg 除以每单位长度 c 的力是长度 u，则

$$\textbf{伸长度}\qquad \boldsymbol{e}=A\boldsymbol{u}=\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\frac{mg}{c}\begin{bmatrix} 3/2 \\ 2 \\ 3/2 \end{bmatrix}=\frac{mg}{c}\begin{bmatrix} 3/2 \\ 1/2 \\ -1/2 \\ -3/2 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究!

警告：正常情况下你不可以写成 $K^{-1}=A^{-1}C^{-1}(A^{\mathrm{T}})^{-1}$

这三个矩阵是 $A^{\mathrm{T}}CA$ 混合在一起，他们不能轻易的分开。一般来说，$A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{f}$ 有很多解，4 个方程式的 $A\boldsymbol{u}=\boldsymbol{e}$ 只有 3 个未知数通常是无解，但是 $A^{\mathrm{T}}CA$ 给出在框架内的所有 3 个方程式正确的解。只有当 $m=n$，所有的矩阵是方形，我们才可以从 $\boldsymbol{y}=(A^{\mathrm{T}})^{-1}\boldsymbol{f}$ 到 $\boldsymbol{e}=C^{-1}\boldsymbol{y}$ 到 $\boldsymbol{u}=A^{-1}\boldsymbol{e}$。我们现在会看到。

固定端与自由端

移除第 4 个弹簧，所有的矩阵变成 3×3，模式没有改变！矩阵 A 失去了它的第 4 行以及(当然)A^{T} 失去了它的第 4 列，新的刚度矩阵 K_1 变成方形矩阵的乘积：

$$A_1^{\mathrm{T}}C_1A_1=\begin{bmatrix}1&-1&0\\0&1&-1\\0&0&1\end{bmatrix}\begin{bmatrix}c_1&&\\&c_2&\\&&c_3\end{bmatrix}\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}$$

A^{T} 失去的列与 A 失去的行乘以丢失的 c_4，所以最快找出新 $A^{\mathrm{T}}CA$ 的方法是令旧的矩阵的 $c_4=0$：

$$\textbf{固定-自由}\qquad A_1^{\mathrm{T}}C_1A_1=\begin{bmatrix}c_1+c_2&-c_2&0\\-c_2&c_2+c_3&-c_3\\0&-c_3&c_3\end{bmatrix}\tag{12}$$

范例 2　若 $c_1=c_2=c_3=1$ 与 $C=I$，这是–1, 2, –1 三对角矩阵 K_1。K_1 的最后单元是 1 而不是 2，这是因为底部的弹簧是自由。假设所有的 $m_j=m$：

$$\textbf{固定-自由}\qquad \boldsymbol{u}=K^{-1}\boldsymbol{f}=\frac{1}{c}\begin{bmatrix}1&1&1\\1&2&2\\1&2&3\end{bmatrix}\begin{bmatrix}mg\\mg\\mg\end{bmatrix}=\frac{mg}{c}\begin{bmatrix}3\\5\\6\end{bmatrix}$$

这些移动比固定-固定的情况要来的大。u_1 的数字 3 是因为全部 3 个质量都把第 1 个弹簧往下拉，下一个质量移动了 3 加上来自下方质量额外的 2，第 3 个质量下掉更多(3 + 2 + 1 = 6)。弹簧伸长度 $\boldsymbol{e}=A\boldsymbol{u}$ 显示这些数字 3, 2, 1：

【原文 free-free 是错误，应该是 fixed-fixed】

$$\boldsymbol{e}=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}\frac{mg}{c}\begin{bmatrix}3\\5\\6\end{bmatrix}=\frac{mg}{c}\begin{bmatrix}3\\2\\1\end{bmatrix}$$

两个自由端：K 是奇异

两个自由端表示大麻烦，整个弹簧线都可以移动，A 是 2×3：

$$\textbf{自由-自由}\ \ \boldsymbol{e}=A\boldsymbol{u} \qquad \boldsymbol{e}=\begin{bmatrix}e_1\\e_2\end{bmatrix}=\begin{bmatrix}u_2-u_1\\u_3-u_2\end{bmatrix}=\begin{bmatrix}-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix} \tag{13}$$

现在 $A\boldsymbol{u}=\boldsymbol{0}$ 有一个非零解，**没有弹簧拉伸质量也可以移动**。整条线可以被 $\boldsymbol{u}=(1,1,1)$平移而且得到 $\boldsymbol{e}=(0,0)$：

$$A\boldsymbol{u}=\begin{bmatrix}-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}=\textbf{没有拉伸} \tag{14}$$

$A\boldsymbol{u}=\boldsymbol{0}$ 当然产生 $A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{0}$，则 $A^{\mathrm{T}}CA$ 只有正半定，因为没有 c_1 与 c_4。枢轴会是 c_2 与 c_3 而且没有第 3 个枢轴，秩是 2：

$$\begin{bmatrix}-1&0\\1&-1\\0&1\end{bmatrix}\begin{bmatrix}c_2\\c_3\end{bmatrix}\begin{bmatrix}-1&1&0\\0&-1&1\end{bmatrix}=\begin{bmatrix}c_2&-c_2&0\\-c_2&c_2+c_3&-c_3\\0&-c_3&c_3\end{bmatrix} \tag{15}$$

两个特征值是正数但是 $\boldsymbol{x}=(1,1,1)$是对应$\lambda=0$ 的一个特征向量，我们只能对特殊向量 $\boldsymbol{f}$ 才可以求解 $A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f}$，这些力必须相加成 $f_1+f_2+f_3=0$，或是整条弹簧线(两端自由)会像火箭一样发射。

弹簧圈

第 3 个弹簧会完成从质量 3 回到质量 1 的圆圈，这个不会使得 K 可逆--刚度矩阵 $K_{\text{圆圈}}$矩阵仍然是奇异：

$$A_{\text{圆圈}}{}^{\mathrm{T}}A_{\text{圆圈}}=\begin{bmatrix}1&-1&0\\0&1&-1\\-1&0&1\end{bmatrix}\begin{bmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{bmatrix}=\begin{bmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{bmatrix} \tag{16}$$

枢轴只有 2 与 3/2，特征值是 3 与 3 与 0，行列式是零。当所有的质量一起移动时，零空间仍然包含 $\boldsymbol{x}=(1,1,1)$。移动向量(1, 1, 1)在 $A_{\text{圆圈}}$的零空间中而且 $K_{\text{圆圈}}=A^{\mathrm{T}}CA$。

让我整理一下这个段落？我希望这个例子会帮助你连接微积分与线性代数，利用差分方程式取代微分方程式。如果你的步阶Δx 足够小，你会得到一个完整又满足的解。

$$\textbf{方程式是}\ -\frac{d}{dx}\left(c(x)\frac{du}{dx}\right)=f(x)\text{，}\textbf{其中}\ u(0)=0\ \textbf{与}\left[u(1)\text{或}\frac{du}{dx}(1)\right]=0$$

把棒子分成长度Δx 的 N 个片段，利用 $A\boldsymbol{u}$ 取代 du/dx 以及 $A^{\mathrm{T}}\boldsymbol{y}$ 取代$-dy/dx$，现在 A 与 A^{T} 包含 $1/\Delta x$，端点条件是 $u_0=0$ 与$[u_N=0$ 或 $y_N=0]$。

请尊重版权与译者的劳动成果，侵权必究!

3 个步骤–d/dx 与 $c(x)$与 d/dx 对应到 A^{T} 与 C 与 A：

$$\boldsymbol{f}=A^{\mathrm{T}}\boldsymbol{y} \text{ 与 } \boldsymbol{y}=C\boldsymbol{e} \text{ 与 } \boldsymbol{e}=A\boldsymbol{u} \text{ 得到 } A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f}$$

这是计算科学与工程的基础范例。

1. 利用微分方程式得到问题的模型
2. 把微分方程式离散化得到差分方程式
3. 了解并且求解差分方程式(以及边界条件！)
4. 说明得到的解答；视觉化；如果需要，重新设计。

除了试验与推导之外，数值仿真变成科学的第三分支。波音 777 的电脑设计要比风洞便宜太多了。

应用数学与计算科学与工程(卫斯理-康桥出版社)的两篇文章开发了深入的整体主题——观看课程内容 math.mit.edu/18085，附有视频演讲(这些演讲也在 ocw.mit.edu 以及 Youtube)。我希望这本书帮助你了解计算背后的框架。

问题集 10.2

1 证明 $\det A_0^{\mathrm{T}}C_0A_0 = c_1c_2c_3 + c_1c_3c_4 + c_1c_2c_4 + c_2c_3c_4$。也求出固定-自由范例中的 $\det A_1^{\mathrm{T}}C_1A_1$。

2 求出固定-自由范例中的 $A_1^{\mathrm{T}}C_1A_1$ 的逆矩阵，利用右乘 $A_1^{-1}C_1^{-1}(A_1^{\mathrm{T}})^{-1}$。

3 在自由-自由案例中方程式(15)的 $A^{\mathrm{T}}CA$ 是奇异，把 3 个方程式 $A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f}$ 相加，证明我们需要 $f_1+f_2+f_3=0$。当力 $\boldsymbol{f}=(-1,0,1)$自我平衡时，求 $A^{\mathrm{T}}CA\boldsymbol{u}=\boldsymbol{f}$ 的一个解，再求出所有的解。

4 自由-自由微分方程式的两个端点条件是 $du/dx = 0$:

$$-\frac{d}{dx}\left(c(x)\frac{du}{dx}\right)=f(x)，\text{其中在两个端点有 } \frac{du}{dx}=0$$

两侧积分证明力 $\boldsymbol{f}(x)$必须自我平衡，$\int f(x)\,dx=0$，或者不存在解。完整解是一个特定解加上任意常数，对应 $\boldsymbol{u}=(1,1,1)$的常数在 $A^{\mathrm{T}}CA$ 的零空间中。

5 在固定-自由问题中，矩阵 A 是方形可逆，我们分开求解 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{f}$ 与 $A\boldsymbol{u}=\boldsymbol{e}$。对于微分方程式做同样的事情：

$$\text{求解 } -\frac{d}{dx}=f(x)，\text{其中 } y(1)=0。\text{若 } f(x)=1，\text{画出图形。}$$

6 方程式(6)的 3×3 矩阵 $K_1 = A_1^{\mathrm{T}} C_1 A_1$ 分成 3 个“元素矩阵” $c_1E_1 + c_2E_2 + c_3E_3$。对应每个 c 写下这些片段。展示他们如何来自 $A_1^{\mathrm{T}} C_1 A_1$ 的列乘行，这是有限元素刚度矩阵如何实际组成的过程。

7 针对 5 个弹簧与 4 个质量以及两端固定，矩阵 A 与 C 与 K 为何？当 $C = I$ 求解 $K\boldsymbol{u}$ = ones(4)。

8 比较问题 7 的解 $\boldsymbol{u} = (u_1, u_2, u_3, u_4)$与连续问题$-u'' = mg$ 的解，其中 $u(0) = 0$ 与 $u'(1) = 0$。抛物线 $u(x)$应该在 $x = 1/5, 2/5, 3/5, 4/5$ 对应到 $\boldsymbol{u}$——会存在一个$(\Delta x)^2$的因子作为补偿吗？

9 求解固定-自由问题$-u'' = mg$ 的解，其中 $u(0) = 0$ 与 $u'(1) = 0$。当向量是范例 2 的 $\boldsymbol{u} = (3mg, 5mg, 6mg)$时，在 $x = 1/3, 2/3, 3/$比较 $u(x)$的值。

10 假设 $c_1 = c_2 = c_3 = c_4 = 1$，$m_1 = 2$ 与 $m_2 = m_3 = 1$，针对固定-固定弹簧线求解 $A^{\mathrm{T}}CA\boldsymbol{u} = (2, 1, 1)$。哪个质量移动最多(最大的 u)？

11 (MATLAB) 针对 100 个由 $c = 1$ 的弹簧连接的质量，求出位移 $u(1), ..., u(100)$。每个力是 $f(i) = 0.01$，印出固定-固定与固定-自由时，$\boldsymbol{u}$ 的图形。注意 diag(ones(n, 1), d)是有 n 个 1 沿着对角 d 的矩阵。这个打印指令会画出向量 $\boldsymbol{u}$ 的图形：

```
plot(u, '+'); xlabel('mass number'); ylabel('mass number'); print
```

12 (MATLAB) 化学工程有来自流体速度的第一导数 du/dx，以及来自扩散的 d^2u/dx^2。利用前向差分取代 du/dx，然后是中心差分，再然后是反向差分，其中$\Delta x = 1/8$。画出 3 个数值解的图形，其中

$$-\frac{d^2u}{dx^2} + 10\frac{du}{dx} = 1\text{，其中 } u(0) = u(1) = 0$$

这个**对流-扩散方程式**无所不在，它转换成针对数学财务中的期权价格 Black-Scholes 方程式。

问题 12 发展成在 MIT 中我的计算科学与工程 18.085 课程的第一个 MATLAB 作业。视频在 ocw.mit.edu。

请尊重版权与译者的劳动成果，侵权必究!

10.3 马可夫矩阵，人口与经济

本段落有关**正矩阵**：每个 $a_{ij} > 0$，很快说明关键事实：**最大的特征值是正数实数，它的特征向量也是如此**。在经济学与生态学与人口动态学与随机游走(random walk)中，这些事实还有一段长路要走：

马可夫 $\lambda_{max} = 1$　　人口 $\lambda_{max} > 1$　　消耗 $\lambda_{max} < 1$

λ_{max} 控制 A 的次方，我们先观察$\lambda_{max} = 1$。

马可夫矩阵

一再的使用矩阵 A 乘正数向量 $\boldsymbol{u}_0$：

马可夫矩阵 $\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$　　$\boldsymbol{u}_1 = A\boldsymbol{u}_0$　　$\boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$

在 k 步之后我们得到 $A^k\boldsymbol{u}_0$。向量 $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \ldots$会趋近一个“稳定状态”$\boldsymbol{u}_\infty = (.6, .4)$。最后的结果与开始的向量 $\boldsymbol{u}_0$ 无关，对于每个 $\boldsymbol{u}_0 = (a, 1 - a)$都会收敛至相同的 $\boldsymbol{u}_\infty = (.6, .4)$。问题是为什么？

稳定状态方程式 $A\boldsymbol{u}_\infty = \boldsymbol{u}_\infty$使得 $\boldsymbol{u}_\infty$成为**对应特征值 1 的特征向量**：

稳定状态 $\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .6 \\ .4 \end{bmatrix} = \begin{bmatrix} .6 \\ .4 \end{bmatrix} = \boldsymbol{u}_\infty$

左乘 A 不会改变 $\boldsymbol{u}_\infty$，但是这个不能解释为什么那么多向量 $\boldsymbol{u}_0$ 会得到 $\boldsymbol{u}_\infty$。我们的例子可能有一个稳定状态，但是它不必然是吸引的：

不是马可夫 $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$　　有非吸引的稳定状态　　$B \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

在这个情况下，初始向量 $\boldsymbol{u}_0 = (0, 1)$会得到 $\boldsymbol{u}_1 = (0, 2)$与 $\boldsymbol{u}_2 = (0, 4)$，第二个分量一直加倍。用特征值的语言来说，$B$ 有$\lambda = 1$ 也有$\lambda = 2$--这个产生不稳定。沿着不稳定特征向量的 $\boldsymbol{u}$ 的分量是乘λ，而且$|\lambda| > 1$ 代表爆炸。

本段落有关 A 的两个特殊性质，这些性质可以确保一个稳定的(stable)稳定状态，也定义了一个**正数马可夫矩阵**，以及上方的 A 是一个特定范例：

马可夫矩阵	1. A 的每个单元是正数：$a_{ij} > 0$。 2. A 的每个列的总和是 1。

B 的列 2 的总和是 2 不是 1。当 A 是一个马可夫矩阵，马上得到两个事实：

因为 **1**：A 乘以 $\boldsymbol{u}_0 \geq 0$ 得到一个非负 $\boldsymbol{u}_1 = A\boldsymbol{u}_0 \geq 0$。

因为 **2**：如果 $\boldsymbol{u}_0$ 的分量总和是 1，则 $\boldsymbol{u}_1 = A\boldsymbol{u}_0$ 的分量总和也是 1。

理由：当[1 ... 1] $\boldsymbol{u}_0 = 1$，则 $\boldsymbol{u}_0$ 的分量总和是 1，借由性质 2，对于 A 的每一列都是成立的，利用矩阵乘法[1 ... 1] A = [1 ... 1]：

$$A\boldsymbol{u}_0 \text{ 的分量总和是 } 1 \qquad [1 \ldots 1]\, A\boldsymbol{u}_0 = [1 \ldots 1]\boldsymbol{u}_0 = 1$$

相同的事实可以应用到 $\boldsymbol{u}_2 = A\boldsymbol{u}_1$ 与 $\boldsymbol{u}_3 = A\boldsymbol{u}_2$。**每个向量 $A^k\boldsymbol{u}_0$ 都是非负而且分量的总和是 1**，这些是“**概率向量**”。极限 $\boldsymbol{u}_\infty$也是一个概率向量--但是我们必须证明存在一个极限。我们会证明正数马可夫矩阵的$\lambda_{\max} = 1$。

范例 1　在丹佛租车的比例开始是 1/50 = .02，在丹佛之外是.98。每个月 80%的丹佛车停留在丹佛(20%离开)，同时有 5%的外部车进入(95%停留在外)，这表示用 A 乘分数(fraction) $\boldsymbol{u}_0$ = (.02, .98)：

第一个月　$A = \begin{bmatrix} \mathbf{.80} & \mathbf{.05} \\ \mathbf{.20} & \mathbf{.95} \end{bmatrix}$ 得到 $\boldsymbol{u}_1 = A\boldsymbol{u}_0 = A\begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .065 \\ .935 \end{bmatrix}$

注意.065 + .935 = 1，代表所有的汽车。每个步骤都是用 A 乘：

下个月　$\boldsymbol{u}_2 = A\boldsymbol{u}_1$ = (.09875, .9125)，这是 $A^2\boldsymbol{u}_0$。

所有的向量都是正数，这是因为 A 是正数，每个向量 $\boldsymbol{u}_k$ 的分量总和都是 1。第一分量从 0.02 成长，车子朝着丹佛移动。长期之下会发生什么？

这个段落与矩阵的次方有关，了解 A^k 是对角化的第一个也是最重要的应用，A^k 可以很复杂，对角矩阵Λ^k 很简单，特征向量矩阵 X 连接他们：A^k 等于 $X\Lambda^kX^{-1}$。马可夫矩阵的新应用使用特征值(在Λ中)与特征向量(在 X 中)，我们证明 **$\boldsymbol{u}_\infty$是 A 的一个对应$\lambda = 1$ 的特征向量**。

由于 A 的每个列的分量总和是 1，没有失去没有获得，我们是在移动车辆或人口，没有车辆或人口突然出现(或消失)。这些比例总和是 1 而且矩阵 A 让他们持续，问题是在 k 个时间周期后，他们是如何分布--引导我们到 A^k。

解　$A^k\boldsymbol{u}_0$ 给定丹佛在 k 步之后的进与出比例，我们对角化 A 去了解 A^k，特征值是 $\lambda = 1$ 与$\lambda = .75$(迹是 1.75)。

$A\boldsymbol{x} = \lambda\boldsymbol{x}$　$A\begin{bmatrix} .2 \\ .8 \end{bmatrix} = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix}$　与　$A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = .75\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

起始向量 $\boldsymbol{u}_0$ 组合 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$，这个情况下的系数是 1 与.18：

特征向量的组合　$\boldsymbol{u}_0 = \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

现在用 A 乘得到 $\boldsymbol{u}_1$，特征向量是乘以$\lambda_1 = 1$ 与$\lambda_2 = .75$：

每个 $\boldsymbol{x}$ 乘以λ　$\boldsymbol{u}_1 = 1\begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)(.18)\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

请尊重版权与译者的劳动成果，侵权必究!

每个月，另一个$\lambda = .75$乘向量$\boldsymbol{x}_2$，特征向量$\boldsymbol{x}_1$保持不变：

k 步之后　　$$\boldsymbol{u}_k = A^k\boldsymbol{u}_0 = 1^k\begin{bmatrix}.2\\.8\end{bmatrix} + (.75)^k(.18)\begin{bmatrix}-1\\1\end{bmatrix}$$

这个方程式显示发生什么了，**对应$\lambda = 1$的特征向量$\boldsymbol{x}_1$是稳定状态**，另一个特征向量$\boldsymbol{x}_2$会消失，这是因为$|\lambda| < 1$。我们走的步越多，我们越趋近$\boldsymbol{u}_\infty = (.2, .8)$。在极限时，2/10 的车辆在丹佛并且 8/10 在外部。这是马可夫链条(chain)的模式，纵使从$\boldsymbol{u}_0 = (0, 1)$开始：

若A是一个正数马可夫矩阵(单元$a_{ij} > 0$，每个列的总和是 1)，则$\lambda_1 = 1$比其他固有值都要大，特征向量$\boldsymbol{x}_1$是**稳定状态**：

$$\boldsymbol{u}_k = \boldsymbol{x}_1 + c_2(\lambda_2)^k\boldsymbol{x}_2 + \ldots + c_n(\lambda_n)^k\boldsymbol{x}_n \quad \text{永远会趋近} \quad \boldsymbol{u}_\infty = \boldsymbol{x}_1$$

第一点是了解$\lambda = 1$是A的一个特征值。理由：$A - I$的每个列的分量总和是 1 −1 = 0，$A - I$的行的总和是个零行，这些行线性相关，所以$A - I$是奇异，它的行列式是零而且$\lambda = 1$是一个特征值。

第二点是没有特征值可以有$|\lambda| > 1$，有这样的特征值，次方A^k会长大。但是A^k也是一个马可夫矩阵！A^k仍然有总和是 1 的正数单元--所以没有空间变大。

很多关注在于另一个特征值有$|\lambda| = 1$ 的可能性。

范例 2　$A = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$没有稳定状态，因为$\lambda_2 = -1$。

这个矩阵把丹佛内部的车辆全部送到外部，反之亦然。次方A^k在A与I之间交替，第二个特征向量$\boldsymbol{x}_2 = (-1, 1)$在每一步都乘以$\lambda_2 = -1$--但是没有变小：没有稳定状态。

假设A或A的任意次方的单元都是正数--不允许零，在“规则的”或“原始的”情况下，$\lambda = 1$比其他特征值要严格的大，次方A^k趋近秩 1 矩阵，这个矩阵的每个列有稳定状态。

范例 3　(“**每个人都移动**”) 从 3 个群体开始，在每一步，群体 1 的半数到群体 2，一半到群体 3；其他群体也分成两半移动。起始人口是p_1, p_2, p_3，一步之后是：

新人口　　$$\boldsymbol{u}_1 = A\boldsymbol{u}_0 = \begin{bmatrix}0 & \frac{1}{2} & \frac{1}{2}\\\frac{1}{2} & 0 & \frac{1}{2}\\\frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}\begin{bmatrix}p_1\\p_2\\p_3\end{bmatrix} = \begin{bmatrix}\frac{1}{2}p_2 + \frac{1}{2}p_3\\\frac{1}{2}p_1 + \frac{1}{2}p_3\\\frac{1}{2}p_1 + \frac{1}{2}p_2\end{bmatrix}$$

A是马可夫矩阵，没有人出生或死亡。A包含零的单元，这是范例 2 的麻烦。但是在两步之后，零从A^2消失：

两步矩阵　　$$\boldsymbol{u}_2 = A^2\boldsymbol{u}_0 = \begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\\frac{1}{4} & \frac{1}{2} & \frac{1}{4}\\\frac{1}{4} & \frac{1}{4} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}p_1\\p_2\\p_3\end{bmatrix}$$

矩阵 A 的特征值是$\lambda_1 = 1$(因为 A 是马可夫)与$\lambda_2 = \lambda_3 = -1/2$。$\lambda = 1$ 时，特征向量 $\boldsymbol{x}_1 = (1/3, 1/3, 1/3)$会是**稳定状态**。当 3 个相等的人口分成两半移动，人口会再次相等。从 $\boldsymbol{u}_0 = (8, 16, 32)$开始，马可夫链条趋近它的稳定状态：

$$\boldsymbol{u}_0 = \begin{bmatrix} 8 \\ 16 \\ 32 \end{bmatrix} \qquad \boldsymbol{u}_1 = \begin{bmatrix} 24 \\ 10 \\ 12 \end{bmatrix} \qquad \boldsymbol{u}_2 = \begin{bmatrix} 16 \\ 18 \\ 22 \end{bmatrix} \qquad \boldsymbol{u}_3 = \begin{bmatrix} 20 \\ 19 \\ 17 \end{bmatrix}$$

到了 $\boldsymbol{u}_4$ 会把一个人分成两半，这个可不行。在每一步的总人口是 8 +16 +32 = 56，稳定状态是 56 乘(1/3, 1/3, 1/3)。你可以看到 3 个人口趋近，但是永远无法抵达，他们的极限是 56/3。

挑战问题 6.7.16 建立一个来自网站之间链接的马可夫矩阵，稳定状态 $\boldsymbol{u}$ 会给出谷歌等级。谷歌利用追随链接的随机路径(随机浏览)找出 $\boldsymbol{u}_\infty$，特征向量来自每个网站拜访的比例--快速计算稳定状态的方法。

第二特征值$|\lambda_2|$的大小**控制收敛到稳定状态的速度**。

佩龙-弗罗贝尼乌斯定理

有一个矩阵定理支配这个主题，当 $a_{ij} \geq 0$ 时应用**佩龙定理(简写)**，没有要求每个列的总和是 1。我们证明最简洁的形式，当所有的 $a_{ij} > 0$：任意正数矩阵 A (不需要是正定！)。

针对 $A > 0$ 的佩龙定理 **在 $A\boldsymbol{x} = \lambda_{max}\boldsymbol{x}$ 中的数字都是严格正数**

证明 关键概念是检视所有的数字 t，对于某些非负向量 $\boldsymbol{x}$ (除了 $\boldsymbol{x} = \boldsymbol{0}$ 之外)使得 $A\boldsymbol{x} \geq t\boldsymbol{x}$。我们在 $A\boldsymbol{x} \geq t\boldsymbol{x}$ 中允许不等式，这样才能有很多小的正数候选者 t。对于大的数字 t_{max} (已经获得)，我们会证明等式成立：$A\boldsymbol{x} = t_{max}\boldsymbol{x}$。

否则，如果 $A\boldsymbol{x} \geq t_{max}\boldsymbol{x}$ 不是等式，左乘 A。因为 A 是正数会得到严格的不等式 $A^2\boldsymbol{x} > t_{max}A\boldsymbol{x}$，因此正数向量 $\boldsymbol{y} = A\boldsymbol{x}$ 满足 $A\boldsymbol{y} > t_{max}\boldsymbol{y}$，且 t_{max} 增加。这个矛盾强迫等式 $A\boldsymbol{x} = t_{max}\boldsymbol{x}$，而且我们得到一个特征值，它的特征向量 $\boldsymbol{x}$ 是正数，这是因为在等式的左侧，$A\boldsymbol{x}$ 一定是正数。

为了了解没有特征值会比 t_{max} 大，假设 $A\boldsymbol{z} = \lambda\boldsymbol{z}$，由于$\lambda$与 $\boldsymbol{z}$ 可能与负数或复数有关，所以我们取绝对值：使用“三角不等式”有 $|\lambda|\,|\boldsymbol{z}| = |A\boldsymbol{z}| \leq A|\boldsymbol{z}|$，这个$|\boldsymbol{z}|$是一个非负向量，所以这个$|\lambda|$是一个可能的候选者 t，因此$|\lambda|$不可能超过 t_{max} —— 这个必须是λ_{max}。

请尊重版权与译者的劳动成果，侵权必究！

人口成长

把人口分成 3 个年龄群体：年龄< 20，年龄 20 到 39，年龄 40 到 59。在 T 年时，这些群体的数量是 n_1, n_2, n_3，20 年后的人口数量会因为 3 个理由改变：出生，死亡与变老。

1. **生殖** $n_1^{新} = F_1n_1 + F_2n_2 + F_3n_3$ 给出新世代
2. **存活** $n_2^{新} = P_1n_1$ 以及 $n_3^{新} = P_2n_2$ 给出老世代

生育率是 F_1, F_2, F_3(F_2 最大)，莱斯利(Leslie)矩阵 A 可能看起来像这样：

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}^{新} = \begin{bmatrix} F_1 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \begin{bmatrix} .04 & \mathbf{1.1} & .01 \\ .98 & 0 & 0 \\ 0 & \mathbf{.92} & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ \boldsymbol{n}_2 \\ n_3 \end{bmatrix}$$

这是人口数量最简单的投影，在每一步矩阵 A 都相同。在实际模型中，A 会随着时间改变(来自环境或是内部因素)。教授可能想要引入第 4 个群体，年龄 ≥ 60，但是我们不允许这样做。

矩阵 $A \geq 0$ 但不是 $A > 0$，因为 $A^3 > 0$，佩龙定理仍然可以应用。最大的特征值是 $\lambda_{\max} \approx 1.06$。你可以看到世代的移动，从中世代 $n_2 = 1$ 开始：

$$\mathbf{eig}(A) = \begin{matrix} \mathbf{1.06} \\ -1.01 \\ -0.01 \end{matrix} \qquad A^2 = \begin{bmatrix} 1.08 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{1.08} & .01 \\ 0.90 & 0 & 0 \end{bmatrix} \qquad A^3 = \begin{bmatrix} 0.10 & \mathbf{1.19} & .01 \\ 0.06 & \mathbf{0.05} & .00 \\ 0.04 & \mathbf{0.99} & .01 \end{bmatrix}$$

一个快速的开始可以来自 $\boldsymbol{u}_0 = (0, 1, 0)$，中群体会生殖 1.1 然后存活.92。最新与最老的世代在 $\boldsymbol{u}_1 = (1.1, 0, 0.92) = A$ 的列 2，然后 $\boldsymbol{u}_2 = A\boldsymbol{u}_1 = A^2\boldsymbol{u}_0$ 是 A^2 的第二列。早期的数字(暂时状态)与 $\boldsymbol{u}_0$ 有很大关系，但是**渐近(asymptotic)成长率 $\boldsymbol{\lambda}_{\max}$ 对每一个初始值都相同**。它的特征向量 $\boldsymbol{x} = (.63, .58, .51)$显示这 3 个群体一起稳定成长。

Caswell 的书：**矩阵人口模型**强调敏感度分析，这个模型永远都不精确。如果矩阵中的 F's 与 P's 变化 10%，$\lambda_{\max}$ 会小于 1 (表示绝种)吗？问题 19 会显示一个矩阵变化 ΔA 产生特征值变化 $\Delta\lambda = \boldsymbol{y}^{\mathrm{T}}(\Delta A)\boldsymbol{x}$，此处 $\boldsymbol{x}$ 与 $\boldsymbol{y}^{\mathrm{T}}$ 是 A 的右与左特征向量，其中 $A\boldsymbol{x} = d\boldsymbol{x}$ 与 $A^{\mathrm{T}}\boldsymbol{y} = \lambda\boldsymbol{y}$。

经济学的线性代数：消耗矩阵

一篇有关经济学的线性代数的长文章在这里会格格不入，关于一个矩阵的短注解看起来比较合理。**消费矩阵**说明多少的输入产生一个单位的输出，这个在描述经济的生产面。

消耗矩阵 我们有 n 个制造业例如化学品，食物与石油。要生产一个单位的化学品，需要 0.2 单位的化学品与 0.3 单位的食物与 0.4 单位的石油。这些数字进入消耗矩阵 A 的第一行：

$$\begin{bmatrix} \text{化学品输出} \\ \text{食物输出} \\ \text{石油输出} \end{bmatrix} = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix} \begin{bmatrix} \text{化学品输入} \\ \text{食物输入} \\ \text{石油输入} \end{bmatrix}$$

行 2 显示生产食物的输入––大量的化学品与食物以及少量石油，A 的行 3 显示精炼一单位石油需要消耗的输入。美国在 1958 年实际的消耗矩阵包含 83 个制造业，这个模型在 1990's 年更大更精确。我们选择的消耗矩阵有一个便利的特征向量。

现在问题来了：经济能够满足 y_1, y_2, y_3 对应的化学品，食物与石油需求吗？要做到这样，输入 p_1, p_2, p_3 有必要提高一些––因为部分的 $\boldsymbol{p}$ 在生产 y 时消耗掉了。输入 p 以及消耗 Ap，得到输出 $\boldsymbol{p} - A\boldsymbol{p}$，这项净生产满足需求 $\boldsymbol{y}$：

问题 求出一个向量 $\boldsymbol{p}$ 使得 $\boldsymbol{p} - A\boldsymbol{p} = \boldsymbol{y}$ 或是 $\boldsymbol{p} = (I-A)^{-1}\boldsymbol{y}$

很明显的线性代数的问题是 $I-A$ 是否可逆？但是还有更多问题，需求输出向量 $\boldsymbol{y}$ 是非负，A 也是，在 $\boldsymbol{p} = (I-A)^{-1}\boldsymbol{y}$ 的生产水平也必须是非负。实际的问题是：

什么时候 $(I-A)^{-1}$ 是非负矩阵？

这是针对生产经济 $(I-A)^{-1}$ 的一个测试，一个能否满足需求的测试。若 A 比 I 小非常多，则 $A\boldsymbol{p}$ 比 $\boldsymbol{p}$ 小非常多，存在很多输出。若 A 太大，则生产消耗太多而且需求 $\boldsymbol{y}$ 无法满足。

“小”或是“大”决定在 A 的最大特征值 λ_1(是个正数)：

若 $\lambda_1 > 1$ 则 $(I-A)^{-1}$ 有负数单元

若 $\lambda_1 = 1$ 则 $(I-A)^{-1}$ 不存在

若 $\lambda_1 < 1$ 则 $(I-A)^{-1}$ 是我们想要的非负矩阵

重点在于最后一项，这项推理使用了一个 $(I-A)^{-1}$ 的优质公式，我们现在给出这个公式。数学里面最重要的无限级数是**几何级数** $1 + x + x^2 + \ldots$ ，这个级数的总和是 $1/(1-x)$，条件是 x 落在–1 与 1 之间。当 $x = 1$，级数是 $1 + 1 + 1 + \ldots = \infty$。当 $|x| > 1$，x^n 项不会趋近零，这个级数没有机会收敛。

关于 $(I-A)^{-1}$ 的优质公式就是**矩阵的几何级数**：

几何级数 $(I-A)^{-1} = I + A + A^2 + A^3 + \ldots$

请尊重版权与译者的劳动成果，侵权必究!

如果你用 A 乘 $S = I + A + A^2 + \ldots$，除了 I 以外你会得到相同的级数，因此 $S - AS = I$，如果级数收敛的话，级数的总和会等于 $S = (I - A)^{-1}$。如果 **A 所有的特征值都有 $|\lambda| < 1$，这个级数收敛**。

在我们的案例 $A \geq 0$，级数的每一项都是非负，它的总和是 $(I-A)^{-1} \geq 0$。

范例 4 $A = \begin{bmatrix} .2 & .3 & .4 \\ .4 & .4 & .1 \\ .5 & .1 & .3 \end{bmatrix}$ 有 $\lambda_{\max} = .9$ 与 $(I-A)^{-1} = \frac{1}{93}\begin{bmatrix} 41 & 25 & 27 \\ 33 & 36 & 24 \\ 34 & 23 & 36 \end{bmatrix}$。

这个经济是生产型，A 比 I 小，因为 $\lambda_{\max}$ 是 0.9，要满足需求 $\boldsymbol{y}$，从 $\boldsymbol{p} = (I - A)^{-1}\boldsymbol{y}$ 开始，则 $A\boldsymbol{p}$ 是生产中的消耗，剩下 $\boldsymbol{p} - A\boldsymbol{p}$，这就是 $(I-A)\boldsymbol{p} = \boldsymbol{y}$，满足需求。

范例 5 $A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$ 有 $\lambda_{\max} = 2$ 与 $(I-A)^{-1} = -\frac{1}{3}\begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}$。

这个消耗矩阵 A 太大，需求无法满足，因为生产消耗的比它制造的还要多。级数 $I + A + A^2 + \ldots$ 无法收敛到 $(I-A)^{-1}$，这是因为 $\lambda_{\max} > 1$。当 $(I-A)^{-1}$ 实际是负数时，级数是成长的。

相同的方式，$1 + 2 + 4 + \ldots$ 实际上不是 $1/(1-2) = -1$，但也不是完全错误！

问题集 10.3

问题 1-12 有关马可夫矩阵与它们的特征值及次方。

1　求出下列马可夫矩阵的特征值(总和是迹)：

$$A = \begin{bmatrix} .90 & .15 \\ .10 & .85 \end{bmatrix}$$

对应 $\lambda_1 = 1$ 的稳定状态特征向量为何？

2　利用其他的特征向量，对角化问题 1 的马可夫矩阵成为 $A = X\Lambda X^{-1}$：

$$A = \begin{bmatrix} & \end{bmatrix}\begin{bmatrix} 1 & \\ & .75 \end{bmatrix}\begin{bmatrix} & \end{bmatrix}$$

当 $\Lambda^k = \begin{bmatrix} 1 & \\ & .75^k \end{bmatrix}$，$A^k = X\Lambda^k X^{-1}$ 的极限为何？会趋近 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$？

3　求出下列马可夫矩阵的特征值与稳定状态特征向量：

$$A = \begin{bmatrix} 1 & .2 \\ 0 & .8 \end{bmatrix} \quad A = \begin{bmatrix} .2 & 1 \\ .8 & 0 \end{bmatrix} \quad A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

4　对于每个 4×4 马可夫矩阵，A^{T} 的哪个特征向量对应(已知)特征值 $\lambda = 1$？

5 每年会有 2%的年轻人变老以及 3%的年轻人死亡(没有出生)，求出下列的稳定状态：

$$\begin{bmatrix} \text{年轻人} \\ \text{老年人} \\ \text{死亡} \end{bmatrix}_{k+1} = \begin{bmatrix} .98 & .00 & 0 \\ .02 & .97 & 0 \\ .00 & .03 & 1 \end{bmatrix} \begin{bmatrix} \text{年轻人} \\ \text{老年人} \\ \text{死亡} \end{bmatrix}_{k}$$

6 一个马可夫矩阵，$\boldsymbol{x}$ 的分量总和等于 $A\boldsymbol{x}$ 分量的总和。若 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 且$\lambda \neq 1$，证明非稳定特征向量 $\boldsymbol{x}$ 的分量总和是 0。

7 求出矩阵 A 的特征值与稳定状态特征向量。说明为什么 A^k 趋近 A^∞：

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \qquad A^\infty = \begin{bmatrix} .6 & .6 \\ .4 & .4 \end{bmatrix}$$

挑战问题：哪个马可夫矩阵产生稳定状态(.6, .4)?

8 排列矩阵的稳定状态特征向量是(1/4, 1/4, 1/4, 1/4)，当 $\boldsymbol{u}_0 = (0, 0, 0, 1)$，不会趋近到上述向量。$\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4$ 为何？P 的 4 个特征值为何？这些是$\lambda^4 = 1$ 的解？

排列矩阵=马可夫矩阵 $\quad P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

9 证明一个马可夫矩阵的平方仍然是马可夫矩阵。

10 若 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$是一个马可夫矩阵，它的特征值是 1 与____，稳定状态特征向量 $\boldsymbol{x}_1 =$ _____。

11 填空 A 变成一个马可夫矩阵并求出稳定状态特征向量。当 A 是一个对称马可夫矩阵，为什么 $\boldsymbol{x}_1 = (1, ..., 1)$是它的稳定状态？

$$A = \begin{bmatrix} .7 & .1 & .2 \\ .1 & .6 & .3 \\ - & - & - \end{bmatrix}$$

12 一个马可夫微分方程式是 $du/dt = (A - I)u$，而不是 $du/dt = Au$。$A - I$ 的对角线是负数，其他都是正数，每个列的总和是 0 不是 1。

针对 $B = A - I = \begin{bmatrix} -.2 & .3 \\ .2 & -.3 \end{bmatrix}$，求出$\lambda_1$ 与λ_2。为什么 $A - I$ 有$\lambda_1 = 0$？当 $e^{\lambda_1 t}$ 与 $e^{\lambda_2 t}$ 乘上 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 且 $t \to \infty$时，稳定状态为何？

请尊重版权与译者的劳动成果，侵权必究!

问题 13-15 有关经济学的线性代数。

13 范例 4 消耗矩阵的每个行的总和是 0.9，为什么这样会使得特征值是 0.9？特征向量为何？

14 使用 $I-A$ 乘 $I+A+A^2+A^3+\ldots$ 得到 I，级数的总和是$(I-A)^{-1}$。$A=\begin{bmatrix}0 & 1/2\\ 1 & 0\end{bmatrix}$，求 A^2 与 A^3 并且利用这个模式求出级数的总和。

15 下列矩阵中，哪些可以使得 $I+A+A^2+A^3+\ldots$ 得到一个非负矩阵$(I-A)^{-1}$？则经济可以满足需求：

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} \qquad A=\begin{bmatrix}0 & 4\\ .2 & 0\end{bmatrix} \qquad A=\begin{bmatrix}.5 & 1\\ .5 & 0\end{bmatrix}$$

如果需求是 $\boldsymbol{y}=(2,6)$，向量 $\boldsymbol{p}=(I-A)^{-1}\boldsymbol{y}$ 为何？

16 (再次马可夫) 下列矩阵有零行列式，特征值为何？

$$A=\begin{bmatrix}.4 & .2 & .3\\ .2 & .4 & .3\\ .4 & .4 & .4\end{bmatrix}$$

从 $\boldsymbol{u}_0=(1,0,0)$与 $\boldsymbol{u}_0=(100,0,0)$开始，求出 $A^k\boldsymbol{u}_0$ 的极限

17 若 A 是一个马可夫矩阵，为什么 $I+A+A^2+\ldots$ 的总和不是$(I-A)^{-1}$？

18 针对莱斯利矩阵证明 $\det(A-\lambda I)=0$ 得到 $F_1\lambda^2+F_2P_1\lambda+F_3P_1P_2=\lambda^3$。右侧的 λ^3 在$\lambda\to\infty$时会变大，如果 $F_1+F_2P_1+F_3P_1P_2>1$，左侧在$\lambda=1$ 时会变大。在 $\lambda_{\max}>1$ 的情况下两侧会相等：成长。

19 **特征向量的灵敏度：**一个矩阵变化ΔA 产生特征值变化$\Delta\Lambda$，这些变化$\Delta\lambda_1$，...，$\Delta\lambda_n$ 落在$(X\Delta AX^{-1})$的对角线。**挑战：**从 $AX=X\Lambda$开始，特征向量与特征值的变化是ΔX 与$\Delta\Lambda$:

$(A+\Delta A)\,(X+\Delta X)=(X+\Delta X)\,(\Lambda+\Delta\Lambda)$ 变成 $A(\Delta X)+(\Delta A)X=X(\Delta\Lambda)+(\Delta X)\,\Lambda$

忽略小的项$(\Delta A)(\Delta X)$与$(\Delta X)(\Delta\Lambda)$，用 X^{-1} 乘最后一个方程式，从里面的项，$X^{-1}(\Delta A)X$ 的对角线部分给出$\Delta\Lambda$就是我们想要的。为什么外面的项 $X^{-1}A\ (\Delta X)$ 与 $X^{-1}\Delta X\,\Lambda$在对角线会抵消？

说明 $X^{-1}A=\Lambda X^{-1}$，然后 $\mathrm{diag}(\Lambda X^{-1}\Delta X)=\mathrm{diag}(X^{-1}\Delta X\,\Lambda)$。

20 假设 $B>A>0$ 表示每个 $b_{ij}>a_{ij}>0$，佩龙的讨论如何展示$\lambda_{\max}(B)>\lambda_{\max}(A)$?

10.4 线性规划

线性规划(linear programming)是线性代数加上两个概念：**不等式与极小化**。开始点仍然是矩阵方程式 $A\boldsymbol{x} = \boldsymbol{b}$，但是只有非负的解才可以接受，我们需要 $\boldsymbol{x} \geq 0$(表示没有 $\boldsymbol{x}$ 的分量可以是负数)。矩阵是 $n > m$，未知数比方程式多，如果 $A\boldsymbol{x} = \boldsymbol{b}$ 有 $\boldsymbol{x} \geq 0$ 的解，可能会有很多解。线性规划筛选 $\boldsymbol{x}^* \geq 0$ 而且极小化下列成本的解：

成本是 $c_1x_1 + \ldots + c_nx_n$，致胜向量 $\boldsymbol{x}^*$ 是 $A\boldsymbol{x} = \boldsymbol{b}$ 的非负解而且有最低的成本。

于是线性规划的问题从一个矩阵 A 与两个向量 $\boldsymbol{b}$ 与 $\boldsymbol{c}$ 开始：

i) A 有 $n > m$：例如 $A = [1 \quad 1 \quad 2]$ (一个方程式，三个未知数)

ii) $\boldsymbol{b}$ 有 m 个分量对应 m 个方程式 $A\boldsymbol{x} = \boldsymbol{b}$：例如 $\boldsymbol{b} = [4]$

iii) 成本向量 $\boldsymbol{c}$ 有 n 个分量：例如 $\boldsymbol{c} = [5 \quad 3 \quad 8]$

问题就是：在满足 $A\boldsymbol{x} = \boldsymbol{b}$ 且 $\boldsymbol{x} \geq 0$ 的条件下，极小化 $\boldsymbol{c} \cdot \boldsymbol{x}$：

极小化 $5x_1 + 3x_2 + 8x_3$ **满足条件** $x_1 + x_2 + 2x_3 = 4$ 且 $x_1, x_2, x_3 \geq 0$

我们直达主题，没有解释来自何处。线性规划实际上是数学到管理学的最重要应用，快速演算法以及极速代码的发展非常有竞争力。你会看到求出 $\boldsymbol{x}^*$ 比求解 $A\boldsymbol{x} = \boldsymbol{b}$ 要困难很多，这是因为有额外的要求：$\boldsymbol{x}^* \geq 0$ 以及极小化成本 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^*$。在求解范例之后，我们会解释背景以及著名的“单纯形法(simplex method)”与“内点法(interior point method)”。

首先检视“限制”：$A\boldsymbol{x} = \boldsymbol{b}$ 且 $\boldsymbol{x} \geq 0$，方程式 $x_1 + x_2 + 2x_3 = 4$ 给出三维空间的一个平面，非负性质 $x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$ 把平面切成一个三角形，解 $\boldsymbol{x}^*$ 必须落在图 10.5 的三角形 PQR 里面。【原文 8.6 是错误，应该是 10.5】

在三角形里面，$\boldsymbol{x}$ 的分量都是正数。在 PQR 的边缘，一个分量是零。在角点 $\boldsymbol{P}$ 与 $\boldsymbol{Q}$ 与 $\boldsymbol{R}$，两个分量是零。**最佳解 $\boldsymbol{x}^*$ 会是这些角点其中之一**！我们现在说明为什么。

三角形包含所有满足 $A\boldsymbol{x} = \boldsymbol{b}$ 且 $\boldsymbol{x} \geq 0$ 的向量 $\boldsymbol{x}$，这些 $\boldsymbol{x}$ 称为“**可行(feasible)点**”，而且这个三角形是“**可行集合**”。这些点允许成为极小化 $\boldsymbol{c} \cdot \boldsymbol{x}$ 的候选者，这最后一个步骤：

求出在三角形 PQR 中的 $\boldsymbol{x}^*$ 极小化成本 $5x_1 + 3x_2 + 8x_3$

落在平面 $5x_1 + 3x_2 + 8x_3 = 0$ 的向量有零成本，这个平面与三角形没有交会，在满足 $\boldsymbol{x}$ 的要求下，我们无法得到到零成本。所以增加成本 C 直到平面 $5x_1 + 3x_2 + 8x_3 = C$ 与三角形交会。当 C 增加，我们有平行的平面往三角形移动。

请尊重版权与译者的劳动成果，侵权必究！

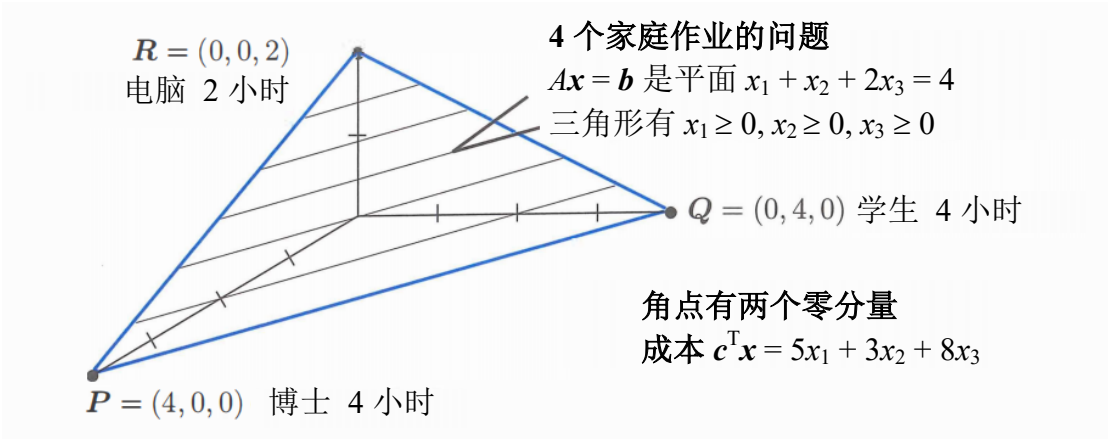


图 10.5：三角形包含所有非负解：$A\boldsymbol{x} = \boldsymbol{b}$ 且 $\boldsymbol{x} \geq 0$。最低成本解 $\boldsymbol{x}^*$是可行集合中的角点 $\boldsymbol{P}$，$\boldsymbol{Q}$，$\boldsymbol{R}$ 的其中之一。

第一个碰触三角形的平面 $5x_1 + 3x_2 + 8x_3 = C$ 有极小成本 C，**碰触的点就是解** $\boldsymbol{x}^*$。碰触点必须是交点 $\boldsymbol{P}$ 或 $\boldsymbol{Q}$ 或 $\boldsymbol{R}$ 的其中之一，一个移动的平面不可能在碰触三角形的一个角点之前碰触到三角形的内部！所以检查在每个角点的成本 $5x_1 + 3x_2 + 8x_3$：

$\boldsymbol{P}$ = (4, 0, 0) 成本 20　　$\boldsymbol{Q}$ = (0, 4, 0) 成本 12　　$\boldsymbol{R}$ = (0, 0, 2) 成本 16

胜利者是 $\boldsymbol{Q}$，所以 $\boldsymbol{x}^* = (0, 4, 0)$是这个线性规划问题的解。

如果成本向量 $\boldsymbol{c}$ 改变，平行平面会倾斜，如果是小的改变，$\boldsymbol{Q}$ 仍然是赢家。对于成本 $\boldsymbol{c} \cdot \boldsymbol{x} = 5x_1 + 4x_2 + 7x_3$，最佳 $\boldsymbol{x}^*$移到 $\boldsymbol{R} = (0, 0, 2)$，极小成本是 $7 \cdot 2 = 14$。

注解 1　某些线性规划**极大化收益**而不是极小化成本，数学差不多一样。平行平面从比较大的 C 开始，而不是小的值。当 C 变小，他们朝着原点移动(而不是离开)，第一个碰触点仍然是角点。

注解 2　需求 $A\boldsymbol{x} = \boldsymbol{b}$ 且 $\boldsymbol{x} \geq 0$ 可能无法满足，方程式 $x_1 + x_2 + x_3 = -1$ 不可能有 $\boldsymbol{x} \geq 0$ 的解，可行集合是空集合。

注解 3　也可能发生可行集合是无界(unbounded)的情况，如果需求是 $x_1 + x_2 - 2x_3 = 4$，很大的正数向量(100, 100, 98)现在是一个候选者，更大的向量(1000, 1000, 998)也是。平面 $A\boldsymbol{x} = \boldsymbol{b}$ 不再切成三角形，两个角点 $\boldsymbol{P}$ 与 $\boldsymbol{Q}$ 仍然是 $\boldsymbol{x}^*$的候选者，但是 $\boldsymbol{R}$ 移到无限远处。

注解 4　一个无界的可行集合，极小成本可能是$-\infty$ (负数无限大)。假设成本是$-x_1 - x_2 + x_3$，则向量(100, 100, 98)使得成本是 $C = -102$，向量(1000, 1000, 998)使得成本是 $C = -1002$。我们现在是收到钱去引入 x_1 与 x_2，而不是付出成本，实际应用上这是不可能的事情。但是这是理论上的可能，A, $\boldsymbol{b}$ 与 $\boldsymbol{c}$ 可以产生非预期的三角形与成本。

原始与对偶问题

第一个问题适配范例中的 A, $\boldsymbol{b}$ 与 $\boldsymbol{c}$，未知数 x_1, x_2, x_3 代表一位博士的工作时数与一位学生的工作时数与一台机器的工作时数，每小时的成本是\$5，\$3 与\$8。(工资这么低我感到抱歉。) 小时的数字不可能是负数：$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0$。博士与学生每小时完成一个家庭作业问题，机器每小时解答两个问题。理论上他们可以共同完成作业，总共有 4 个问题待解答：$x_1 + x_2 + 2x_3 = 4$。

问题是在极小成本 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ 下完成 4 个问题

如果全部三者都在工作，需要一小时完成：$x_1 = x_2 = x_3 = 1$，成本是 $5 + 3 + 8 = 16$。但是理所当然的博士应该被学生的工作捣乱(学生工作很快但是成本低--问题变成实际)，当学生工作两小时机器工作一小时，成本是 $6 + 8 = 14$，而且解决全部 4 个问题。我们在三角形的 $\boldsymbol{QR}$ 边，这是因为博士没有在工作：$x_1 = 0$。但是最佳点是全部由学生解决(在 $\boldsymbol{Q}$)或是全部由机器解决(在 $\boldsymbol{R}$)，这个例子中学生在 4 个小时解决全部 4 个问题只要\$12--极小成本。

只有 $A\boldsymbol{x} = \boldsymbol{b}$ 的一个方程式，角点(0, 4, 0)只有一个非零分量。当 $A\boldsymbol{x} = \boldsymbol{b}$ **有 $\boldsymbol{m}$ 个方程式，角点就有 $\boldsymbol{m}$ 个非零分量**，我们对着 m 个变数求解 $A\boldsymbol{x} = \boldsymbol{b}$，其中 $n - m$ 个自由变数设成零。但是不像第三章，**我们不知道选择哪 $\boldsymbol{m}$ 个变数**。

可能角点的个数是在 n 中选取 m 个分量的方式的个数，这个数字“n 选 m”与博弈及概率极度相关。$n = 20$ 个未知数与 $m = 8$ 个方程式(还是小数字)，“可行集合”可以有 10!/8!/12!个角点，这个数字是(20) (19) ... (13) = 5,079,110,400。

为了极小成本检查 3 个角点还行，检查 50 亿个角点是不可行的方法。下面介绍的**单纯形法**快速多了。

对偶问题(dual problem) 在线性规划中问题都是成对出现，极小值问题与绝对值问题--原始与它的“对偶(dual)”。原始问题是由一个矩阵 A 与两个向量 $\boldsymbol{b}$ 与 $\boldsymbol{c}$ 组成，对偶问题会转置 A 以及交换 $\boldsymbol{b}$ 与 $\boldsymbol{c}$：**极大化 $\boldsymbol{b} \cdot \boldsymbol{y}$**，下列是我们例子的对偶：

一个作弊提供者靠卖答案解答作业问题，一个问题收费 y 元，或一次性 $4y$ 元(注意 $b = 4$ 已经进入成本。) 作弊者必须比博士或是学生伙食机器便宜：$y \leq 5$ 与 $y \leq 3$ 与 $2y \leq 8$。(注意 $\boldsymbol{c} = (5, 3, 8)$如何进入不等式限制)，作弊者极大化收入 $4y$：

对偶问题 极大化 $\boldsymbol{b} \cdot \boldsymbol{y}$ 受限于 $A^{\mathrm{T}}\boldsymbol{y} \leq \boldsymbol{c}$

当 $y = 3$ 出现极大值，收入是 $4y = 12$。对偶问题的极大值(\$12)等于原始问题的极小值(\$12)。极大值 = 极小值就是对偶。

【dual 翻译成“对偶”比“双重”来的好，因为它有**对比**的意思！】

请尊重版权与译者的劳动成果，侵权必究!

如果其中之一的问题有一个最佳向量(x^*或y^*)，则另一个问题也有一样的情形。

极小化成本 $c \cdot x^*$ 等于 极大化收入 $b \cdot y^*$

本书从行图与列图开始，第一个“**对偶定理**”就是有关秩：无关行的个数等于无关列的个数。那个定理，好像这个，对于小矩阵很容易，极小成本 = 极大输入在教科书线性代数与应用中有证明。其中一行可以建立简单定理的一半：**作弊者的输入 b^Ty 不可能超过诚者的成本：**

$$\text{若 } Ax = b,\ x \ge 0,\ A^Ty \le c \text{ 则 } b^Ty = (Ax)^Ty = x^T(A^Ty) \le x^Tc \tag{1}$$

完整的对偶性定理说明当 b^Ty 达到极大值以及 x^Tc 达到极小值时，他们是相等的：$b \cdot y^* = c \cdot x^*$。检视(1)的最后一个步骤，有 $\le$ 符号：

$x \ge 0$ 与 $s = c - A^Ty \ge 0$ 的点积得到 $x^Ts \ge 0$，这是 $x^TA^Ty \le x^Tc$。

等式需要 $x^Ts = 0$，所以对于每个 j，最佳解有 $x_j^* = 0$ 或 $s_j^* = 0$

单纯形法

消元法是线性方程式的工作机器，单纯形法是线性不等式的工作机器。我们不能像消元法一样给出那么多空间，但是概念会更清晰。**单纯形法以较低成本从一个角点到相邻的一个角点**，最终(实际上是很快)到达极小成本的角点。

一个**角点**是一个向量 $x \ge 0$ 满足 m 个方程式 $Ax = b$，其中最多有 m 个正数分量，其余 $n - m$ 个分量是零。(这些是自由变数，反向代入法给出这 m 个基础变数，所有的变数必须是非负或是 x 是一个错误的角点。) 对于相邻角点来说，x 的一个零分量变成正数而且一个正数分量变成零。

单纯形法必须决定哪个分量借由变正“进入”，以及哪个分量借由边零“离开”，交换的原则是缩小成本，这是单纯形法的一个步骤，朝向 x^*前进。

以下是整体的计划，检视目前角点的每个零分量，如果它从 0 变成 1，其他的非零数必须调整才能维持 $Ax = b$。利用反向代入法找出新的 x 然后计算在整体成本 $c \cdot x$ 的变化，这项变化是新分量的“降低成本” r。**进入变数**是给出最负 r 的分量，这是对于一个新变数的单一单位的最多降低成本。

范例 1 假设目前的角点是 $P = (4, 0, 0)$，这是博士负责全部的工作(成本\$20)，如果学生工作一小时，$x = (3, 1, 0)$的成本降低至\$18，减少的成本是 $r = -2$。

如果机器工作一小时，则 $\boldsymbol{x} = (2, 0, 1)$的成本也是\$18，降低的成本也是 $r = -2$。这个情况下单纯形法可以选择学生或机器作为**进入变数**。

尽管这是个小例子，第一步也可能不会马上得到最佳 $\boldsymbol{x}^*$，在这个方法知道要引入多少这样的变数之前，它就要选择进入变数。当进入变数从 0 变成 1，我们计算 r，但是一个单位可能太多或太少，这个方法选择选择**离开变数**(博士)，它移动到图中的 $\boldsymbol{Q}$ 或 $\boldsymbol{R}$。

引入进入变数越多，成本越低。当其中一个正数分量(调整可以维持 $A\boldsymbol{x} = \boldsymbol{b}$)碰到零，程序必须停止。**离开变数**是第一个碰到零的正数 x_i，当这个发生时，会发现一个相邻的角点，然后再出发(从新的角点)发现下个进入或离开变数。

当所有的降低成本都是正数，目前的角点就是最佳 $\boldsymbol{x}^*$。在没有增加 $\boldsymbol{c} \cdot \boldsymbol{x}$ 的情况下，没有零分量可以变成正数，没有新的变数应该进入，问题得到解答(我们也可以证明 $\boldsymbol{y}^*$也找到了)。

注解　通常 $\boldsymbol{x}$ 在 αn 步之内会抵达，其中 α 不是大数字。

范例 2　极小化成本 $\boldsymbol{c} \cdot \boldsymbol{x} = 3x_1 + x_2 + 9x_3 + x_4$，限制是 $\boldsymbol{x} \geq 0$ 与两个方程式 $A\boldsymbol{x} = \boldsymbol{b}$：

$$x_1 + 2x_3 + x_4 = 4 \qquad m = 2 \text{ 个方程式}$$

$$x_2 + x_3 - x_4 = 2 \qquad n = 4 \text{ 个未知数}$$

初始角点是 $\boldsymbol{x} = (4, 2, 0, 0)$的成本是 $\boldsymbol{c} \cdot \boldsymbol{x} = 14$，它有两个非零分量与 $n - m = 2$ 个零分量，零是 x_3 与 x_4。问题是 x_3 或 x_4 应该进入(变成非零)。尝试他们的一个单位：

若 $x_3 = 1$ 与 $x_4 = 0$，则 $\boldsymbol{x} = (2, 1, 1, 0)$的成本是 16

若 $x_4 = 1$ 与 $x_3 = 0$，则 $\boldsymbol{x} = (3, 3, 0, 1)$的成本是 13

与成本 14 做比较，x_3 的降低成本是 $r = 2$，正数但是无用。x_4 的降低成本是 $r = -1$，负数且有用，**进入变数**是 x_4。

x_4 可以进入多少？一个单位的 x_4 使得 x_1 从 4 掉成 3，4 个单位会使得 x_1 从 4 掉成 0(此时 x_2 一路增加到 6)，**离开变数**是 x_1，新的角点是 $\boldsymbol{x} = (0, 6, 0, 4)$，成本只有 $\boldsymbol{c} \cdot \boldsymbol{x} = 10$，这是最佳 $\boldsymbol{x}^*$，但是要确认的话还要尝试另一个来自(0, 6, 0, 4)的单纯形步骤。假设 x_1 或 x_3 尝试进入：

从角点(0, 6, 0, 4)开始　若 $x_1 = 1$ 与 $x_3 = 0$，则 $\boldsymbol{x} = (1, 5, 0, 3)$的成本是 11

若 $x_3 = 1$ 与 $x_1 = 0$，则 $\boldsymbol{x} = (0, 3, 1, 2)$的成本是 14

请尊重版权与译者的劳动成果，侵权必究!

这些成本比 10 高，两者的 r 都是正数--不需要付出而离开，目前的角点(0, 6, 0, 4)是最佳解 $\boldsymbol{x}^*$。

这些计算可以流水线处理，每个单纯形步骤求解 3 个具有相同矩阵 B(这是 $m\times m$ 矩阵维持 A 的 m 个基本列)的线性系统，当一个列进入而一个老的列离开存在一个快速方法更新 B^{-1}，那就是大多数代码组织的单纯形法。

在教科书“计算科学与工程”包含一个有注解的短代码(代码也在 math.mit.edu/cse)。最佳 $\boldsymbol{y}^*$求解 m 个方程式 $A^{\mathrm{T}}\boldsymbol{y}^* = \boldsymbol{c}$，在 m 个 x^*中不是零的分量，然后我们有最佳 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s} = 0$，这是对偶：不是 $\boldsymbol{x}_j^* = \boldsymbol{0}$ 就是 $\boldsymbol{s}^* = \boldsymbol{c} - A^{\mathrm{T}}\boldsymbol{y}^*$中的“松散”有 $\boldsymbol{s}_j^* = \boldsymbol{0}$。

当 $\boldsymbol{x}^* = (0, 4, 0)$是最佳角点 $\boldsymbol{Q}$，作弊者的价格设定为 $y^* = 3$。

内点法

单纯形法沿着可行集合的边移动，最终抵达最佳角点 $\boldsymbol{x}^*$。**内点法是在可行集合的内部移动**(其中 $\boldsymbol{x} > 0$)，这些方法希望更快的到达 $\boldsymbol{x}^*$，他们做得很好。

一个留在内部的方法是在边界设立一个障碍，加入额外的成本，当任何的 x_j 碰到零时对数会爆炸。最佳向量有 $\boldsymbol{x} > 0$，数字θ是个小参数，我们朝向零移动。

障碍问题 极小化 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \theta(\log x_1 + \ldots + \log x_n)$，其中有 $A\boldsymbol{x} = \boldsymbol{b}$ (2)

这个成本是非线性(代数线性规划在不等式已经是非线性)，限制 $x_j > 0$ 已经不需要，这是因为 $\log x_j$ 在 $x_j = 0$ 会变成无限大。

这个障碍给出每个θ 的近似问题，m 个限制 $A\boldsymbol{x} = \boldsymbol{b}$ 有拉格朗日乘数 $y_1, \ldots, y_m$，这是处理限制的好方法。

来自拉格朗日的 $\boldsymbol{y}$ $L(x, y, \theta) = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \theta(\sum \log x_i) - \boldsymbol{y}^{\mathrm{T}}(A\boldsymbol{x} - \boldsymbol{b})$ (3)

$\partial L/\partial y = 0$ 带回 $A\boldsymbol{x} = \boldsymbol{b}$，导数$\partial L/\partial y = x_j$ 很有趣！

障碍问题的最佳 $\dfrac{\partial L}{\partial x_j} = c_j - \dfrac{\theta}{x_j} - \left(A^{\mathrm{T}}y\right)_j = 0$ 就是 $\boldsymbol{x_j s_j = \theta}$ (4)

真正的问题有 $x_j s_j = 0$，障碍问题有 $x_j s_j = \theta$ ，解 $\boldsymbol{x}^*(\theta)$落在到达 $\boldsymbol{x}^*(0)$的中央路径，这些 n 个最佳方程式 $x_j s_j = \theta$ 是非线性，我们利用牛顿法的迭代方法求解。

目前的 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}$ 会满足 $A\boldsymbol{x} = \boldsymbol{b}$，$\boldsymbol{x} \geq 0$ 以及 $A^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c}$，但不是 $x_j s_j = \theta$ 。牛顿方法取一个步阶$\Delta\boldsymbol{x}, \Delta\boldsymbol{y}, \Delta\boldsymbol{s}$，忽略$(\boldsymbol{x} + \Delta\boldsymbol{x})\ (\boldsymbol{s} + \Delta\boldsymbol{s}) = \boldsymbol{\theta}$ 的第二阶项$\Delta\boldsymbol{x}\Delta\boldsymbol{s}$，$x, y, s$ 的修正来自线性方程式：

$$\textbf{牛顿步骤}\qquad \begin{aligned} A\Delta x &= 0 \\ A^{\mathrm{T}}\Delta y + \Delta s &= 0 \\ s_j\Delta x_j + x_j\Delta s_j &= \theta - x_j s_j \end{aligned} \tag{5}$$

牛顿迭代法对于每个θ 有二次收敛，然后θ 趋近 0。对偶间隔 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{s}$ 通常在 20 到 60 步之后会小于 10^{-8}，在我的教科书“计算科学与工程”有详细的牛顿步骤，这个例子有 4 个作业。我没有要求学生把作业全部写完，但是 $\boldsymbol{x}^*$如何出现要熟悉。

这个内点法通常作为商业软件，针对很多的线性与非线性最佳化问题。

问题集 10.4

1 在 xy 平面画出下列的区域：$x+2y=6$ 与 $x\ge 0$ 与 $y\ge 0$。可行集合的哪个点可以极小化成本 $c=x+3y$？哪个点得到极大值成本。这些点都是角点。

2 在 xy 平面画出下列的区域：$x+2y\le 6$，$2x+y\le 6$，$x\ge 0$，$y\ge 0$，它有 4 个角点。哪个点极小化成本 $c=2x-y$？

3 下列区域：$x_1+2x_2-x_3=4$，x_1, x_2, x_3 全部≥ 0 的角点为何？证明成本 x_1+2x_3 在这个可行集合可以是非常负的数。这是一个无界成本的例子：没有极小值。

4 从 $\boldsymbol{x}=(0, 0, 2)$开始，其中机器解答全部 4 个问题花了\$16，移动到 $\boldsymbol{x}=(0, 1, _)$ 求出由学生工作所降低的成本 r(每小时的节省)。求出对应博士的 $\boldsymbol{x}=(1, 0, __)$ 的 r，其中博士工作一小时。

5 从博士角点(4, 0, 0)开始范例 1，其中 $\boldsymbol{c}$ 变成(5, 3, 7)，证明对应机器的 r 比学生的好，纵使学生的总成本较低。单纯形法有两个步骤，第一个对机器，然后对学生求 $\boldsymbol{x}^*$。

6 选择不同的成本向量 $\boldsymbol{c}$，使得博士得到工作。改写对偶问题(作弊者有绝对值收入。)

7 一个 6 问题的作业，博士工作最快得到新的限制 $2x_1+x_2+x_3=6$，则 $\boldsymbol{x}=(2, 2, 0)$显示博士工作两小时与学生工作两小时。$\boldsymbol{c}=(5, 3, 8)$时，这样的 $\boldsymbol{x}$ 能够极小化成本 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$ 吗？

8 这两个问题也是对偶。证明弱对偶，永远有 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\le \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$：

原始问题 极小化 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$，其中 $A\boldsymbol{x}\ge \boldsymbol{b}$ 与 $\boldsymbol{x}\ge \boldsymbol{0}$。

对偶问题 极大化 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$，其中 $A^{\mathrm{T}}\boldsymbol{y}\le \boldsymbol{c}$ 与 $\boldsymbol{y}\ge \boldsymbol{0}$。

请尊重版权与译者的劳动成果，侵权必究！

10.5 傅里叶级数：函数的线性代数

这个段落从**有限维度**到**无限维度**，我想要说明在无限维度空间的线性代数。第一步：回头看。本书从向量与点积与线性组合开始，我们把这些基础概念转换到无限的情况--后面都追随这个。

有无限多分量的向量表示什么？有两种不同的答案，都很棒：

1. 向量是无限长：$\boldsymbol{v} = (v_1, v_2, v_3, ...)$，它可能是$(1, ½, ¼, ...)$。
2. 向量是一个函数$f(x)$，它可能是 $\boldsymbol{v} = \sin x$。

两种方式我们都会使用，傅里叶级数的概念会联结他们。

在向量之后是点积，两个无限向量$(v_1, v_2, v_3, ...)$与$(w_1, w_2, w_3, ...)$的自然点积是一个无限级数：

$$\textbf{点积} \qquad \boldsymbol{v} \cdot \boldsymbol{w} = v_1w_1 + v_2w_2 \tag{1}$$

这个带来新的问题，这是 $\mathbf{R}^n$ 中向量从来没有遇到过的，这个无限级数的总和是一个有限数字吗？级数收敛吗？这是有限与无限第一个也是最大的差别。

当 $\boldsymbol{v} = \boldsymbol{w} = (1, 1, 1, ...)$，总和当然不收敛，这个情况下 $\boldsymbol{v} \cdot \boldsymbol{w} = 1 + 1 + 1 + ...$ 是无限大。由于 $\boldsymbol{v}$ 等于 $\boldsymbol{w}$，我们实际是在计算 $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2$，长度的平方。向量$(1, 1, 1, ...)$有无限的长度，我们不想要这个向量。因为我们在制定规则，我们没必要包含它，我们只允许具备有限长度的向量：

定义 向量 $\boldsymbol{v} = (v_1, v_2, ...)$且函数$f(x)$在我们无限维度的“**希尔伯特空间**”若且唯若$\| \boldsymbol{v} \|$与$\|f\|$是有限：

$$\| \boldsymbol{v} \|^2 = \boldsymbol{v} \cdot \boldsymbol{v} = v_1^2 + v_2^2 + v_3^2 + ... \quad \textbf{总和必须是一个有限数字}$$

$$\|\boldsymbol{f}\|^2 = (\boldsymbol{f}, \boldsymbol{f}) = \int_0^{2\pi} |f(x)|^2 \, dx \qquad \textbf{必须是一个有限积分}$$

范例 1 向量 $\boldsymbol{v} = (1, ½, ¼, ...)$包含在希尔伯特空间，因为它的长度是$2/\sqrt{3}$。我们有一个几何级数的总和是 4/3，$\boldsymbol{v}$ 的长度是平方根：

$$\textbf{长度平方} \qquad \boldsymbol{v} \cdot \boldsymbol{v} = 1 + 1/4 + 1/16 + ... = 1/(1 - 1/4) = 4/3$$

问题 若 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 有有限长度，他们的内积会有多大？

答案 $\boldsymbol{v} \cdot \boldsymbol{w} = v_1w_1 + v_2w_2 + ...$ 的总和也是一个有限数字，我们可以安全的计算内积，苏瓦兹不等式仍然成立：

$$\textbf{苏瓦兹不等式} \qquad |\boldsymbol{v} \cdot \boldsymbol{w}| \le \| \boldsymbol{v} \| \, \| \boldsymbol{w} \| \tag{2}$$

$\boldsymbol{v} \cdot \boldsymbol{w}$与$\| \boldsymbol{v} \| \, \| \boldsymbol{w} \|$的比值仍然是$\theta$($\boldsymbol{v}$与$\boldsymbol{w}$之间的角度)的余弦。纵使在无限维度空间，$|\cos\theta|$也不会超过 1。

现在转移到函数，这些也是“向量”，函数 $f(x), g(x), h(x), \ldots$ 的空间，定义在 $0 \le x \le 2\pi$，必须比 $\mathbf{R}^n$ 稍微大一点。**$f(x)$与 $g(x)$的点积是什么？$f(x)$的长度是什么？**

关键点在连续的情况：利用积分取代总和。不用 v_j 乘 w_j 的总和，点积是 $f(x)$ 乘 $g(x)$的积分。把括号内的“点”换成逗号，点积的词换成内积：

定义 $f(x)$与 $g(x)$的**内积**以及 $f(x)$的**长度平方**是

$$(f,\ g)=\int_0^{2\pi} f(x)g(x)\,dx \qquad \|f\|^2=\int_0^{2\pi}(f(x))^2\,dx \tag{3}$$

函数的定义区间$[0, 2\pi]$可以改成不同的区间，例如$[0, 1]$或$(-\infty, \infty)$，我们选择 2π 是因为我们的第一个例子是 $\cos x$ 与 $\sin x$。

范例 2 $f(x) = \sin x$ 的长度来自于本身的内积：

$$(f,\ f)=\int_0^{2\pi}(\sin x)^2\,dx=\pi, \quad \sin x \text{ 的长度是}\sqrt{\pi}\ 。$$

这是微积分中的一个标准积分——本身线性代数的一部分。把 $\sin^2 x$ 写成½ – ½ $\cos 2x$，我们看出它在平均值½往上往下，平均值乘以区间长度 2π 得到答案π。

更重要：**$\sin x$ 与 $\cos x$ 在函数空间是正交：**$(f, g) = 0$

内积是零 $$\int_0^{2\pi}\sin x\cos x\,dx=\int_0^{2\pi}\frac{1}{2}\sin 2x\,dx=\left[-\frac{1}{4}\cos 2x\right]_0^{2\pi}=0 \tag{4}$$

这个零没有意外，对于科学非常重要，正交性在两个函数 $\sin x$ 与 $\cos x$ 之上，正弦与余弦的无限列表，这个列表包含 $\cos 0x$(这是 1), $\sin x$, $\cos x$, $\sin 2x$, $\cos 2x$, $\sin 3x$, $\cos 3x$, ...

列表中的每个函数都与列表中的其他函数正交。

傅里叶级数

一个函数 $f(x)$的傅里叶级数是它的正弦与余弦的扩展：

$$f(x) = a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + b_2\sin 2x + \ldots \tag{5}$$

我们有一组正交基底！“函数空间”中的向量是这些正弦与余弦的组合。从 2π 到 4π 的区间，所有的函数重复回到 0 到 2π，他们是周期函数，重复距离是 2π。

请尊重版权与译者的劳动成果，侵权必究！

记住：列表是无限的，傅里叶级数是无限级数。我们避免 $\boldsymbol{v} = (1, 1, 1, ...)$是因为它的长度是无限，现在我们避免像½ + cos x + cos $2x$ + cos $3x$ 这样的函数。(注解：这是π 乘上著名的“delta 函数”$\delta(x)$，它是在单一点上的无限“尖波(spike)”。在 $x = 0$ 它的高度是½ + 1 + 1 + ...是无限大，其他在 $0 < x < 2\pi$ 的所有点，级数的总和以某种平均方式得到零。) $\delta(x)$的积分是零，但是$\int \delta^2(x)dx = \infty$，所以 delta 函数不允许进入希尔伯特空间。 【原文少了 dx】

计算一个典型总和$f(x)$的长度：

$$\begin{aligned}(f,\ f) &= \int_0^{2\pi} (a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + ...)^2\, dx \\ &= \int_0^{2\pi} \left(a_0{}^2 + a_1{}^2\cos^2 x + b_1{}^2\sin^2 x + a_2{}^2\cos^2 2x + ...\right)dx\end{aligned}$$

$$\boldsymbol{\|f\|^2 = 2\pi a_0{}^2 + \pi(a_1{}^2 + b_1{}^2 + a_2{}^2 + ...\ )} \qquad (6)$$

第一行到第二行的步骤使用了正交性，所有像乘积 cos x cos $2x$ 的积分都会得到零。第二行包含剩下的项--每个正弦与余弦的平方的积分。第三行计算了这些积分。(1^2 的积分是 2π，其他的积分得到π。) 如果我们除以他们的长度，我们的函数变成正交单位：

$$\frac{1}{\sqrt{2\pi}},\ \frac{\cos x}{\sqrt{\pi}},\ \frac{\sin x}{\sqrt{\pi}},\ \frac{\cos 2x}{\sqrt{\pi}},\ ....$$ **是我们的函数空间的一组基底**

这些是单位向量，我们可以使用系数 $A_0, A_1, B_1, A_2, ...$组合他们得到一个函数 $F(x)$，则 2π与π's 就从长度公式提出。

函数长度=向量长度 $\|F\|^2 = (F, F) = A_0{}^2 + A_1{}^2 + B_1{}^2 + A_2{}^2 + ...$ (7)

这里有一个重点，$f(x)$与 $F(x)$都适用。**只有当系数向量有有限长度时，**函数才有**有限长度**。傅里叶级数给了我们一个在函数与向量的希尔伯特空间之间完美的适配，数在 L^2，它的傅里叶系数在 l^2。

> 当希尔伯特空间包含 $f(x)$的傅里叶系数的向量 $\boldsymbol{v} = (a_0, a_1, b_1, ...)$，则函数空间包含 $f(x)$，两者都有有限长度。

范例 3 假设 $f(x)$是个“方形波”，在 $0 \le x < \pi$ 时等于 1，在$\pi \le x < 2\pi$ 时掉到–1，然后+1 与–1 一直重复。这个$f(x)$是个类似正弦的奇函数，所有的余弦系数都是零。我们会求出它的傅里叶级数，只包含正弦：

方形波 $$f(x) = \frac{4}{\pi}\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + ...\right] \qquad (8)$$

这个函数的长度是$\sqrt{2\pi}$，这是因为在每一点$(f(x))^2$ 是$(-1)^2$ 或是$(+1)^2$：

$$\|\ f\ \|^2 = \int_0^{2\pi} (f(x))^2\ dx = \int_0^{2\pi} 1\, dx = 2\pi$$

在 $x = 0$ 正弦是零而且傅里叶级数也是零，这是从–1 跳到+1 的一半。当 $x = \pi/2$ 时傅里叶级数也很有趣，在这个点方形波等于 1，(8)中的正弦在+1 与–1 交替：

π 的公式
$$1 = \frac{4}{\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots\right) \tag{9}$$

乘π 得到那个著名数字魔法般的公式 $4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots\right)$。

傅里叶系数

我们如何求出乘上正弦与余弦的系数 a's 与 b's？给定一个函数 $f(x)$，我们想要知道它的傅里叶系数 a_k 与 b_k：

傅里叶级数 $f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \ldots$

这是求出 a_1 的方法，两侧同乘 $\cos x$，然后从 0 到 2π积分。

关键是正交性！除了 $\cos^2 x$ 之外，右侧所有的积分都是零：

系数 a_1
$$\int_0^{2\pi} f(x)\cos x\,dx = \int_0^{2\pi} a_1 \cos^2 x\,dx = \pi a_1 \tag{10}$$

除以π 你得到 a_1。要求出其他的 a_k，用 $\cos kx$ 乘傅里叶级数，从 0 到 2π 积分。利用正交性，所以只留下 $a_k \cos^2 kx$ 的积分，这个积分是πa_k，除以π ：

$$\boldsymbol{a_k} = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx \quad \text{相似的} \quad \boldsymbol{b_k} = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx \tag{11}$$

a_0 是例外，这次我们乘 $\cos 0x = 1$，1 的积分是 2π ：

常数项
$$\boldsymbol{a_0} = \frac{1}{2\pi}\int_0^{2\pi} f(x)\cdot 1\,dx = f(x)\text{的平均值} \tag{12}$$

我使用这些公式去求方程式(8)的方形波的傅里叶系数，$f(x)$ cos kx 的积分是零，$f(x)\sin kx$ 的积分是 $4/k$，此时 k 是奇数。

比较 $\mathbf{R}^n$ 中的线性代数

无限维度的希尔伯特空间很像 n 维空间 $\mathbf{R}^n$，假设非零向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 在 $\mathbf{R}^n$ 中正交，我们想要把向量 $\boldsymbol{b}$(而不是函数 $f(x)$)写成这些 $\boldsymbol{v}$'s 的组合：

有限正交级数
$$\boldsymbol{b} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n \tag{13}$$

使用 $\boldsymbol{v}_1^{\mathrm{T}}$ 同时乘两侧，利用正交性，所以 $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_2 = 0$，只留下 c_1 项：

系数 c_1
$$\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{b} = c_1\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1 + \boldsymbol{0} + \ldots + \boldsymbol{0} \quad \text{因此} \quad c_1 = \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1 \tag{14}$$

分母 $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{v}_1$ 是长度平方，好像方程式(11)的π。分子 $\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{b}$ 是好像 $\int f(x)\cos x\,dx$ 的内积。**当基底向量正交时，很容易求出系数。**

请尊重版权与译者的劳动成果，侵权必究!

我们只是在做一维的投影，找出沿着每个基底向量的分量。

当向量是单位正交时，这个公式会更好，我们在 Q 有单位向量，分母 $\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{v}_k$ 全部是 1，你知道 $c_k = \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{b}$ 的另一个形式：

$$\textbf{c's 的方程式} \qquad c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n = \boldsymbol{b} \quad \text{或是} \quad [\boldsymbol{v}_1 \ldots \boldsymbol{v}_n]\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \boldsymbol{b}$$

$$Q\boldsymbol{c} = \boldsymbol{b} \quad \text{得到} \quad \boldsymbol{c} = Q^{\mathrm{T}}\boldsymbol{b}, \quad \text{一行一行的 } c_k = \boldsymbol{q}_k^{\mathrm{T}}\boldsymbol{b}。$$

傅里叶级数很像一个有无限多正交列的矩阵，这些列是基底函数 1, cos x, sin x, ... 除以他们的长度之后我们有一个“无限正交矩阵”，它的逆矩阵是它的转置 Q^{T}。正交性让我们在积分时把一个级数的项简化成单一项。

问题集 10.5

1 假设 $j \neq k$，针对三角恒等式 $2\cos jx \cos kx = \cos(j + k)x + \cos(j - k)x$ 取积分，证明 $\cos jx$ 与 $\cos kx$ 正交。当 $j = k$ 时结果为何？

2 当积分从 $x = -1$ 到 $x = 1$ 时，证明 $1, x, x^2 - \frac{1}{3}$ 是正交。把 $f(x) = 2x^2$ 写成这些正交函数的组合。

3 求出一个与 $\boldsymbol{v} = (1, ½, ¼, ...)$ 正交的向量 $\boldsymbol{w} = (w_1, w_2, w_3, ...)$。计算它的长度 $\|\boldsymbol{w}\|$。

4 前 3 个勒让德多项式是 $1, x, x^2 - \frac{1}{3}$，选择 c 使得第 4 个多项式 $x^3 - cx$ 与前 3 个多项式正交。积分都是从 -1 到 1。

5 范例 3 的方形波 $f(x)$ 从 1 跳到 -1，证明

$$\int_0^{2\pi} f(x)\cos x\,dx = 0 \qquad \int_0^{2\pi} f(x)\sin x\,dx = 4 \qquad \int_0^{2\pi} f(x)\sin 2x\,dx = 0$$

哪 3 个傅里叶系数来自这些积分？

6 方形波有 $\|f\|^2 = 2\pi$，则(6)给出对于 π^2 什么样的惊人总和？

7 画出方形波的图形。然后用手画出级数中两个正弦项的总和，或由机器画出 2, 3 及 10 项的总和。著名的**吉布斯现象**是超调(overshoot)跳跃的震荡(这个不会因为项数变多而消失)。

8 求出下列希尔伯特空间的向量长度

(a) $\boldsymbol{v} = \left(\frac{1}{\sqrt{1}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{8}}, \ldots\right)$

(b) $\boldsymbol{v} = (1, a, a^2, ...)$

(c) $f(x) = 1 + \sin x$

9 计算傅里叶系数 a_k 与 b_k，其中 $f(x)$定义在 0 到 2π ：

(a) $f(x) = 1, 0 \le x \le \pi$; $f(x) = 0, \pi < x < 2\pi$

(b) $f(x) = x$

10 当 $f(x)$有周期 2π，为什么从$-\pi$ 到π 的积分与 0 到 2π 的积分相等？如果 $f(x)$是一个奇函数，$f(-x) = f(x)$，证明 $\int_{-\pi}^{\pi} f(x)dx = 0$。奇函数只有正弦项，偶函数只有余弦项。

11 使用三角恒等式求出 $f(x)$的傅里叶级数的两项：

(a) $f(x) = \cos^2 x$ (b) $f(x) = \cos(x + \pi/3)$ (c) $f(x) = \sin^3 x$

12 函数 1, cos x, sin x, cos $2x$, sin $2x$, ...是希尔伯特空间的一组基底，把前 5 项函数的导数写成相同的前 5 项函数的组合。这些函数的 5×5“微分矩阵”为何？

13 求出中心在 $x = 0$ 的方形波脉冲 $F(x)$的傅里叶系数 a_k 与 b_k：$F(x) = 1/h, |x| \le h/2$ 以及 $F(x) = 0$，$h/2 < |x| \le \pi$。当 $h \to 0$，这个 $F(x)$趋近一个 delta 函数，求出 a_k 与 k_k 的极限。

计算科学与工程的段落 4.1 说明正弦级数，余弦级数，完全级数与复数级数 $\sum c_k e^{ikx}$，也在 **math.mit.edu/cse**。

本书的段落 9.3 介绍**离散傅里叶转换**，这是**向量的傅里叶级数**，可以利用**快速傅里叶转换**计算。这个快速的演算法迅速来自特殊的复数 $z = e^{i\theta} = \cos\theta + i\sin\theta$，此时角度$\theta = 2\pi k/n$。

请尊重版权与译者的劳动成果，侵权必究!

10.6 电脑图形

电脑图形处理图像，图像会搬迁，比例会调整，三维投影在二维，所有的重要运算由矩阵完成--但是这些矩阵的形状会很惊人。

三维空间的转换由 4×4 矩阵完成，你可能会预期是 3×3，这项改变的理由是 4 种关键运算其中之一无法使用 3×3 的矩阵乘法完成。以下是 4 种运算：

平移 (原点移到另一点 $P_0 = (x_0, y_0, z_0)$)。

调整比例 (所有的方向都是 c 倍，或是不同的因子 c_1, c_2, c_3。)

旋转 (环绕通过原点的轴，或是通过 P_0 的轴。)

投影 (投影在通过原点的平面或是通过 P_0 的平面。)

平移最简单--每个店都加上(x_0, y_0, z_0)，但是这个表示线性！没有 3×3 矩阵可以移到原点，所以我们改变原点的坐标值到(0, 0, 0, 1)，这是为什么矩阵是 4×4 的原因。点(x, y, z)的“齐次(homogeneous)坐标”是$(x, y, z, 1)$，我们现在展示如何工作。

1. 平移 沿着向量 $\boldsymbol{v}_0$ 移动整个 3 维空间，原点移动到(x_0, y_0, z_0)，这个向量 $\boldsymbol{v}_0$ 加到 $\mathbf{R}^3$ 的每个点 $\boldsymbol{v}$。使用齐次坐标，4×4 矩阵 T 把整个空间移动 $\boldsymbol{v}_0$：

平移矩阵 $$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_0 & y_0 & z_0 & 1 \end{bmatrix}$$

重要：电脑图形处理行向量，我们有行乘矩阵而不是矩阵乘列，你可以快速的检查$[0 \quad 0 \quad 0 \quad 1]T = [x_0 \quad y_0 \quad z_0 \quad 1]$。

要把(0, 0, 0)与(x, y, z)移动 $\boldsymbol{v}_0$，改变到齐次坐标(0, 0, 0, 1)与$(x, y, z, 1)$。乘 T，一个行向量乘 T 得到一个行向量。每个 $\boldsymbol{v}$ 移动到 $\boldsymbol{v} + \boldsymbol{v}_0$：

$$[x \quad y \quad z \quad 1]\, T = [x{+}x_0 \quad y{+}y_0 \quad z{+}z_0 \quad 1]$$

这个输出说明任意 $\boldsymbol{v}$ 会移到哪儿。(它到了 $\boldsymbol{v}{+}\boldsymbol{v}_0$)。平移已经由一个矩阵完成，这个在 $\mathbf{R}^3$ 无法做到。

2. 重调比例 要让一个图像适配一整页，我们改变它的宽度与高度，一台复印机会重调一张图 90%的比例。在线性代数，我们乘 0.9 倍的单位矩阵，这个矩阵通常是平面的 2×2 与立体的 3×3。在电脑图形，已经是齐次坐标的情况下，这个矩阵是等比例放大：

重调平面比例 $$S = \begin{bmatrix} .9 & & \\ & .9 & \\ & & 1 \end{bmatrix}$$

重调立体比例 $$S = \begin{bmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

重要：S 不是 cI，我们维持“1”在右下角，则$[x, y, 1]$乘 S 在齐次坐标是正确的。原点仍然停留在正常位置，这是因为$[0 \quad 0 \quad 1]S = [0 \quad 0 \quad 1]$。

然后我们把那个 1 改成 c，结果会很奇怪，点(cx, cy, cz, c)与$(x, y, z, 1)$相同，这个齐次坐标的特殊性质是乘 cI 不会移动点。$\mathbf{R}^3$ 的原点有齐次坐标$(0, 0, 0, 1)$以及对应每个非零 c 的$(0, 0, 0, c)$，这是隐藏在“齐次”这个词背后的概念。

调整比例可以在不同方向有不同的比例，要把一个“全页”图片适配成半页，y 方向比例要调成 1/2。要建立边距，把 x 方向比例设成 3/4。图形矩阵是对角但不是 2×2，要调整平面需要 3×3，要调整空间需要 4×4：

$$\textbf{调整矩阵} \quad S = \begin{bmatrix} \frac{3}{4} & & \\ & \frac{1}{2} & \\ & & 1 \end{bmatrix} \qquad \textbf{与} \qquad S = \begin{bmatrix} c_1 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \\ 0 & 0 & c_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

最后的矩阵调整 x, y, z 方向的比例是 c_1, c_2, c_3，所有矩阵额外列留下额外的 1 在每个向量的末端。

综整：调整矩阵 S 与平移矩阵 T 有相同的大小，他们可以相乘。要先平移再调整，计算乘积 $\boldsymbol{v}TS$，要先调整再平移，计算乘积 $\boldsymbol{v}ST$。他们不一样吗？是的。

$\mathbf{R}^3$ 的点(x, y, z)有在 $\mathbf{P}^3$ 的齐次坐标$(x, y, z, 1)$，这个“投影空间”与 $\mathbf{R}^4$ 不同，它仍然是 3 维。要达成这样的事情，(cx, cy, cz, c)与$(x, y, z, 1)$是相同的点，投影空间 $\mathbf{P}^3$ 的这些点实际上是 $\mathbf{R}^4$ 中通过原点的直线。

电脑图形使用“仿射(affine)”转换，线性加移动。一个仿射转换 T 是利用 4×4 矩阵在 $\mathbf{P}^3$ 执行，其中的矩阵有一个特殊的第 4 列：

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix} = \begin{bmatrix} T(1,0,0) & 0 \\ T(0,1,0) & 0 \\ T(0,0,1) & 0 \\ T(0,0,0) & 1 \end{bmatrix}$$

一般的 3×3 矩阵告诉我们 3 个输出，这个告诉我们 4 个。一般的输出来自输入(1, 0, 0)与(0, 1, 0)与(0, 0, 1)，当这个转换是线性时，3 个输出会显示所有事情，当这个转换是仿射时，这个矩阵额外包含了来自(0, 0, 0)的输出，我们知道移动量。

3. 旋转 在 $\mathbf{R}^2$ 与 $\mathbf{R}^3$ 的旋转由正交矩阵 Q 完成，行列式是+1。(行列式–1 时，我们得到额外通过镜子的镜射。) 当你使用齐次坐标时，就会引入额外的列！

$$\textbf{平面旋转} \quad Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \text{变成} \qquad R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

这个矩阵环绕原点旋转平面，我们该如何环绕点(4, 5)旋转一个平面？这个答案带出了齐次坐标的美丽。把(4, 5)平移到(0, 0)，旋转θ，然后把(0, 0)平移回去(4, 5):

$$\boldsymbol{v}T_RT^{+} = [x \quad y \quad 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -5 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 5 & 1 \end{bmatrix}$$

我不会去计算乘积。这个点一次应用一个矩阵：$\boldsymbol{v}$ 平移到 $\boldsymbol{v}T_$，旋转到 $\boldsymbol{v}T_R$，平移回去 $\boldsymbol{v}T_RT^{+}$。因为每个点[x y 1]是一个行向量，$T_$先作用，旋转的中心点(4, 5)--或是写成(4, 5, 1)--首先移到(0, 0, 1)，旋转没有改变它，然后 T^{+}把它平移回去(4, 5, 1)，这是它全部该做的。点(4, 6, 1)移到(0, 1, 1)，旋转θ，再移回去。

在三维，每个旋转环绕一个轴，轴不移动--它是$\lambda = 1$ 的特征向量直线。假设轴在 z 方向，Q 中的 1 要单独离开 z 轴，R 中额外的 1 要单独离开原点：

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 与 \quad R = \begin{bmatrix} & & & 0 \\ & Q & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

现在假设旋转是环绕单位向量 $\boldsymbol{a} = (a_1, a_2, a_3)$，有了轴 $\boldsymbol{a}$，适配进入 R 的旋转矩阵 Q 有 3 个部分：

$$Q = (\cos\theta)I + (1 - \cos\theta) \begin{bmatrix} a_1^2 & a_1a_2 & a_1a_3 \\ a_1a_2 & a_2^2 & a_2a_3 \\ a_1a_3 & a_2a_3 & a_3^2 \end{bmatrix} - \sin\theta \begin{bmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{bmatrix}. \quad (1)$$

轴没有移动，这是因为 $\boldsymbol{a}Q = \boldsymbol{a}$。当 $\boldsymbol{a}$ = (0, 0, 1)是在 z 方向，这个 Q 变成前面的 Q--对应环绕 z 轴的旋转。

线性转换 Q 永远进入 R 的左上方块，它的下方我们看到零，因为旋转让原点留在正确位置。当这些不是零，这个转换是仿射而且原点会移动。

4. 投影 在线性代数课程中，大部分的平面通过原点。实际生活中，大部分都不会。通过原点的平面是向量空间，其他平面是仿射空间，有时候称为“**平** (flats)”。仿射空间来自平移一个向量空间。

我们想要把三维向量投影在平面，从通过原点的平面开始，平面的单位法向量是 $\boldsymbol{n}$，(我们维持 $\boldsymbol{n}$ 是列向量。) 平面上的向量满足 $\boldsymbol{n}^{\mathrm{T}}\boldsymbol{v} = 0$。一般对于平面的投影矩阵是 $I - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$，要投影向量，用这个矩阵乘。向量 $\boldsymbol{n}$ 投影到零，在平面上的向量 $\boldsymbol{v}$ 投影到自己：

$$(I - nn^{\mathrm{T}})n = n - n(n^{\mathrm{T}}n) = 0 \quad 与 \quad (I - nn^{\mathrm{T}})v = v - n(n^{\mathrm{T}}v) = v.$$

在齐次坐标中投影矩阵变成 4×4(但是原点没有移动)：

$$\textbf{投影在平面 } \boldsymbol{n}^{\mathrm{T}}\boldsymbol{v}=0 \quad P=\begin{bmatrix} & I-\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}} & & 0 \\ & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

现在投影到一个没有通过原点的平面 $\boldsymbol{n}^{\mathrm{T}}(\boldsymbol{v}-\boldsymbol{v}_0)=0$，平面上有一个点是 $\boldsymbol{v}_0$，这是仿射空间(或者**平**)，它很像 $A\boldsymbol{x}=\boldsymbol{b}$ 当右侧不是零的解。一个特定解 $\boldsymbol{v}_0$ 加到零空间--产生一个**平**。

投影到这个**平**有三个步骤，使用 $T_$ 把 $\boldsymbol{v}_0$ 平移到原点，沿着 $\boldsymbol{n}$ 的方向投影，再沿着行向量 $\boldsymbol{v}_0$ 平移回去：

$$\textbf{投影到一个平} \quad T_PT_{+}=\begin{bmatrix} I & 0 \\ -\boldsymbol{v}_0 & 1 \end{bmatrix}\begin{bmatrix} I-\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} I & 0 \\ \boldsymbol{v}_0 & 1 \end{bmatrix}$$

我情不自禁的注意到 $T_$ 与 T_{+}是可逆：平移与平移回去，他们很像第二章的基础矩阵。

这个练习会包含镜射矩阵，也称为“**镜子矩阵**”，这些是电脑图形需要的第 5 种形式。镜射会把每个点移动投影两倍的距离--镜射穿过平面来到另一侧，所以镜子矩阵需要把投影 $I-\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$ 改成 $I-2\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$。

矩阵 P 给出“平行”投影，所有的点移动到与 $\boldsymbol{n}$ 平行，直到抵达平面。电脑图形的另一种选择是“透视”投影，因为它含有透视缩短效果，因此更为流行。有了透视，当物体靠近时看起来更大。投影直线朝着眼睛而来，而不是维持在与 $\boldsymbol{n}$ 平行--眼睛是投影的中心，这是二维照片的深度所在。

电脑图形的基本问题从一个镜头与一个观看位置开始，理想情况下，屏幕上的图像是观看者看到的景象。最简单的图像指定一个比特给每一个小图元素--称为像素(pixel)，它是亮或暗，这个给出没有明暗的黑或白。你可能不同意，事实上我们对于红，绿，蓝三种颜色指定 0 与 2^8 之间的明暗度，这个表示每个像素有 8 * 3 = 24 比特，乘上像素的数目，需要很大的记忆体！

根本上来说，光栅帧缓冲区指挥电子束，它想电视机一样扫描，品质控制在像素的数目以及每个像素的比特数目。在这个领域，标准的教科书是“电脑图形：原理与实务”，作者 Hughes, Van Dam, Mcguire, Skylar, Feiner 与 Akeley(第三版，爱德森-卫斯理，2014)。Ronald Goldman 与 Tony DeRose 的注解是很棒的参考。

请尊重版权与译者的劳动成果，侵权必究！

主要观念的复习

1. 电脑图形需要移动运算 $T(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{v}_0$ 以及线性运算 $T(\boldsymbol{v}) = A\boldsymbol{v}$。
2. $\mathbf{R}^n$ 中的移动可以利用阶数 $n + 1$ 的矩阵执行，使用齐次坐标。
3. 当所有矩阵有 0, 0, 0, 1 在最后一列，[x y z 1]的额外分量 1 会保持。

问题集 10.6

1 $\mathbf{R}^3$ 的典型点是 $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$，坐标向量 $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ 是(1, 0, 0), (0, 1, 0), (0, 0, 1)，点的坐标是(x, y, z)。

这个点在电脑图形中是 $x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ +**原点**，它的齐次坐标是(___, ___, ___)，对于这个相同点的其他坐标是(___, ___, ___)。

2 线性转换 T 由已知的 $T(\boldsymbol{i})$, $T(\boldsymbol{j})$, $T(\boldsymbol{k})$决定，对于仿射转换，我们还需要 T(_____)。输入点$(x, y, z, 1)$转换到 $xT(\boldsymbol{i}) + yT(\boldsymbol{j}) + zT(\boldsymbol{k})$ + _____。

3 4×4 矩阵 T 对应沿着(1, 4, 3)的平移，矩阵 T_1 对应沿着(0, 2, 5)的平移，计算乘积 TT_1，TT_1 是沿着_____的平移。

4 写出调整常数 $\boldsymbol{c}$ 的 4×4 比例矩阵 S，计算 ST 以及 TS，其中 T 是平移(1, 4, 3)。要环绕中心点(1, 4, 3)放大照片，你会选择 $\boldsymbol{v}ST$ 或 $\boldsymbol{v}TS$?

5 什么样的调整矩阵 S(齐次坐标，所以是 3×3)可以把标准页 8.5×11 变成 1×1 平方页？

6 什么样的 4×4 矩阵可以把一个立方体的一个角点移动到原点，然后所有的边长都放大两倍？这个立方体的角点原来在(1, 1, 2)。

7 当方程式(1)的 3 个矩阵乘单位向量 $\boldsymbol{a}$，证明他们给出$(\cos\theta)\boldsymbol{a}$ 与$(1 - \cos\theta)\boldsymbol{a}$ 与 $\mathbf{0}$。相加得到 $\boldsymbol{a}Q = \boldsymbol{a}$ 而且旋转轴没有移动。

8 若 $\boldsymbol{b}$ 与 $\boldsymbol{a}$ 垂直，方程式(1)的 3 个矩阵乘 $\boldsymbol{b}$ 得到$(\cos\theta)\boldsymbol{b}$ 与 $\mathbf{0}$ 与一个垂直 $\boldsymbol{b}$ 的向量，所以 $Q\boldsymbol{b}$ 与 $\boldsymbol{b}$ 有一个夹角 θ，**这是旋转**。

9 投影在平面 $\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3}z = 0$ 的 3×3 投影矩阵 $I - \boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$ 为何？在齐次坐标加上 0, 0, 0, 1 作为 P 中的额外行与额外列。

10 使用相同的4×4矩阵P，计算$T_{-}PT_{+}$求出在平面$\frac{2}{3}x+\frac{2}{3}y+\frac{1}{3}z=0$的投影矩阵。平移$T_{-}$把平面(选择一个)上的一个点移到(0, 0, 0, 1)，逆矩阵T_{+}把它平移回去。

11 把(3, 3, 3)投影在这些平面。使用问题9的P以及问题10的$T_{-}PT_{+}$。

12 如果你把一个方形投影到一个平面，你会得到数目形状？

13 如果你把一个立方体投影到一个平面，这个投影的轮廓为何？令投影平面与立方体的一个对角垂直。

14 穿过平面$\boldsymbol{n}^{\mathrm{T}}\boldsymbol{v}=0$的镜射的3×3镜子矩阵是$M=I-2\boldsymbol{n}\boldsymbol{n}^{\mathrm{T}}$。求出点(3, 3, 3)对于平面$\frac{2}{3}x+\frac{2}{3}y+\frac{1}{3}z=0$的镜射点。

15 求出点(3, 3, 3)对于平面$\frac{2}{3}x+\frac{2}{3}y+\frac{1}{3}z=1$的镜射点。采取3个使用4×4矩阵的步骤：$T_{-}$负责平移使得平面通过原点，把平移点(3, 3, 3, 1) T_{-}对于那个平面做镜射，然后使用 T_{+}平移回去。

16 原点(0, 0, 0, 1)到点(x, y, z, 1)的向量是之间的差$\boldsymbol{v}$ = _____，在齐次坐标，向量的末端是_____，所以我们加_____到一个点，而不是加一个点到一个点。

17 如果你只乘每个点最后的坐标得到(x, y, z, c)，你调整整个空间的比例是____，这是因为点(x, y, z, c)与(　,　　,　　)相同。

请尊重版权与译者的劳动成果，侵权必究！

10.7 密码学的线性代数

1 密码学可以使用字母作为有限场：信息中的字母变成数字 0, 1, 2, ..., $p-1$。

2 数字相加也相乘(模 p)，除以 p 保留余数。

3 希尔密码利用机密矩阵 E(模 p)乘以信息方块。

4 解码时，利用逆矩阵 D(模 p)乘每个方块，不是很保密的密码。

密码学是有关信息的编码与解码，银行一直用这个处理金融信息。令人惊讶的是现代演算法可以处理极端深入的数学，“**椭圆曲线**”扮演密码学的一部分，如同安德鲁-威尔士做的有关费马最后定理耸动的证明。

本段落不会太深入！但是会是我们第一个有限场与有限向量空间的经验，$\mathbf{R}^n$ 的场包含全部的实数，“模运算”的场只包含 p 个整数 0, 1, ..., $p-1$。在 $\mathbf{R}^n$ 存在无限多的向量——现在只会在信息空间(message space)有长度 n 的 p^n 个信息。字母从 A 到 Z 是有限的(如同 $p = 26$)。

本段落的代码很容易破解——他们对于实际的安全性太简单。电脑的能力要求更复杂的密码学，这是因为电脑的强大很容易侦测到小的编码矩阵。但是一个矩阵代码(希尔密码)允许我们看到线性代数的新工作方式。

我们所有编码与解码的计算都会是“**模 p**”，但是线性无关与基底与逆矩阵与行列式的中心观念在改变中存活。我们会做“有限场的线性代数”，以下是模 p 的意义：

$27 \equiv 2$ (模 5)　意思是 $27-2$ **可以整除 5**
$y \equiv x$ (模 p)　意思是 $y-x$ **可以整除 p**

y 除以 5 会有 5 个可能的余数 $x = 0, 1, 2, 3, 4$，所有余数为 0 的数字 5, –5, 10, –10, ... 与零“全等(congruent)”(模 5)。数字 6, –4, 11, –19, ...与 $x = 1$ 全等(模 5)。

我们使用全等这个词代表符号≡，我们称为“**模运算**”，每个整数 y 产生 $x = 0, 1, 2, ..., p-1$ 的数字之一。

如果 p 是一个素数(prime)就是最好的原理，$p = 26$ 个字母从 A 到 Z，我们很不幸的没有从一个素数 p 开始，密码学可以处理这个问题。

模运算

线性代数是基于向量的线性组合，现在我们的向量$(x_1, ..., x_n)$是整数字串，整数限定在 $x = 0, 1, ..., p-1$。当我们以“模 p”处理时，所有的计算都会得到这些整数。这表示每个范围之外的整数 y 除以 p 得到余数 x：

$y = qp + x$	$y \equiv x$(模 p)	y 除以 p 得到余数 x

加法模 3　$10 \equiv 1$(模 3)与 $16 \equiv 1$(模 3)与 $10 + 16 \equiv 1 + 1$(模 3)

我可以 10 + 16，然后 26 除以 3 得到余数 2。

或是我直接余数相加 1 + 1 得到相同的答案 2。

加法模 2　$11 \equiv 1$(模 2)与 $17 \equiv 1$(模 2)与 $11 + 17 \equiv 28 \equiv 0$(模 2)。

余数相加 1 + 1 但不是 2，最终的步骤是 $2 \equiv 0$(模 2)。

加法模 p 是完全合理，乘法模 p 也是一样，此处 $p = 3$：

$10 \equiv 1$(模 3)乘 $16 \equiv 1$(模 3)得到 1 乘 $1 \equiv 1$　　$160 \equiv 1$(模 3)

$5 \equiv 2$(模 3)乘 $8 \equiv 2$(模 3)得到 2 乘 $2 \equiv 1$　　$40 \equiv 1$(模 3)

结论：我们可以安全的执行加法与乘法的模 p，所以我们可以取线性组合，这是线性代数的关键运算。**但是可以除吗？**

在实数场中，倒数是 $1/y$(除了 $y = 0$ 以外的任意数)，这表示：我们找得到另外一个实数 z 使得 $yz = 1$，可逆性是一个场的要求。**可逆性的模 p 永远成立吗？**对于任意 $y = 1, ..., p-1$，我们找得到数字 $z = 1, ..., p-1$ 使得 $yz \equiv 1$ 模 p？

这些例子 $3^{-1} \equiv 4$(模 11)与 $2^{-1} \equiv 6$(模 11) 与 $5^{-1} \equiv 9$(模 11)都是成功的。你可以求解 $7z \equiv 1$(模 11)吗？倒数会是逆反矩阵的关键。

让我们展示当 p 不是一个素数时，逆反模 p 会存在一个问题，例子是 $p = 26$ 分解成 2 乘 13，则 **$y = 2$ 不可能有倒数 z(模 26)**。因为 2 与 26 都是偶数，不可能满足 $2z \equiv 1$(模 26)的要求。

相似地当 p 是 25，5 没有倒数 z，我们不可能求解 $5z \equiv 1$(模 25)，数字 $5z - 1$ 不可能是 5 的倍数，所以它不可能是 25 的倍数。

每个 $y(0 < y < p)$存在倒数若且唯若 p 是素数

倒数需要 $y, 2y, 3y, ..., py$ 在除以 p 时有不同的余数。

如果 my 与 ny 有相同的余数 x，则$(m-n)y$ 可以被 p 整除。

素数 p 可以整除 $m - n$ 或 y，不可能两者都可以。

所以 $y, ..., py$ 有不同的余数：**其中之一的余数必须是 $x = 1$。**

请尊重版权与译者的劳动成果，侵权必究！

恩尼格玛密码机与希尔密码

莱斯特希尔在 1929 年的数学月刊发表了它的密码(编码与解码系统)，这个概念很简单，但是在某些的地方它开启了密码学从语言到数学的变革。那个时代的代码主要是字母混合与信息重排，二次世界大战德国海军使用恩格尼玛是一个巨大的进步--使用的机器很像初始的计算机。英国人设立布莱切利园破解恩格尼玛，他们雇佣拼图解答者与主修语言的人，运气好的他们碰巧得到**艾伦图灵**。

我不知道你有没有看过关于他的电影：模仿者游戏，很多都是虚构的(就好像心灵捕手与美丽心灵)，但是破解恩格尼玛代码的核心概念是正确的，利用在编码以及传播的人性弱点。德国海军司令部公开发送他们的代码顺序--知道这个代码太复杂无法破解(如果没有这些弱点的话)。破解代码需要英国的电子设备去撤销德国的电子设备，也需要天才。

艾伦图灵当然是个天才--英格兰最杰出的数学家，他的一生超级悲惨于 1954 年去世。Andrew Hodges 写的传记非常棒，图灵在波兰被侵略的第二天来到布莱切利园，这是温斯顿丘吉尔的信任使得他有快速又完整的支持。

恩格尼玛机器有齿轮及轮子，希尔密码只需要一个矩阵，这是我们现在要说明的代码，使用线性代数。你会看到解码与矩阵的关系，所有的步骤使用模运算，乘与逆模 p。

我会遵循 Salisbury 州立大学的 Spickler 教授的简洁说明，这些在网站看得到 facultyftp.salisbury.edu/despickler/personal/index.asp。

矩阵的模运算

对于 $A\boldsymbol{x}$(矩阵乘向量)来说，加法与减法与乘法都是我们全部需要的，要乘模 p，我们可以像平常一样计算 A 中的整数乘 $\boldsymbol{x}$ 中的整数--用它的值模 p 后取代 $A\boldsymbol{x}$ 的每个单元。

关键问题：我们何时可以求解 $A\boldsymbol{x} \equiv \boldsymbol{b}$(模 p)？我们仍然有 4 个子空间 $\boldsymbol{C}(A)$, $\boldsymbol{N}(A)$, $\boldsymbol{C}(A^{\mathrm{T}})$, $\boldsymbol{N}(A^{\mathrm{T}})$？他们仍然是配对正交？只要 A 有非零的行列式模 p，是否一定存在逆矩阵模 p？我很开心的说最后 3 个答案是“是的”。(但是逆反问题需要 p 是一个素数)。

我们使用高斯消元法可以求出 A^{-1}(模 p)，如同段落 2.5 一样简化[A　　I]成为[I　A^{-1}]，或者我们可以使用行列式与余因子矩阵 C 的公式 $A^{-1} = (\det A)^{-1}C^{\mathrm{T}}$。我会处理 2×2 整数矩阵 A 的模 3：

$$[A \quad I]=\begin{bmatrix}\mathbf{2} & \mathbf{0} & 1 & 0\\ \mathbf{2} & \mathbf{1} & 0 & 1\end{bmatrix}\rightarrow\begin{bmatrix}2 & 0 & 1 & 0\\ 0 & 1 & 2 & 1\end{bmatrix}\rightarrow \begin{matrix}\text{行 1 乘}\\ 2^{-1}\equiv 2\end{matrix}\rightarrow\begin{bmatrix}1 & 0 & \mathbf{2} & \mathbf{0}\\ 0 & 1 & \mathbf{2} & \mathbf{1}\end{bmatrix}=\begin{bmatrix}I & A^{-1}\end{bmatrix}$$

纯属偶然的 $A^{-1}\equiv A$！计算 A 乘 A 模 3 确实得到单位矩阵：

$$A^2=\begin{bmatrix}2&0\\2&1\end{bmatrix}\begin{bmatrix}2&0\\2&1\end{bmatrix}=\begin{bmatrix}4&0\\6&1\end{bmatrix}\equiv\begin{bmatrix}1&0\\0&1\end{bmatrix}\ (\text{模 }3)$$

A 的行列式是 2，来自段落 5.3 的余因子公式同样得到 $A^{-1}\equiv A$：

$$\begin{bmatrix}2&0\\2&1\end{bmatrix}^{-1}=2^{-1}\begin{bmatrix}1&-0\\-2&2\end{bmatrix}\equiv 2\begin{bmatrix}1&-0\\-2&2\end{bmatrix}\equiv\begin{bmatrix}2&0\\2&1\end{bmatrix}\ (\text{模 }3)$$

定理　A^{-1} 存在模 p 若且唯若 $(\det A)^{-1}$ 存在模 p

要求是：$\det A$ 与 p 没有公因子。

希尔密码的加密

原始的密码是用 A 到 Z 的字母，总数是 $p=26$。希尔选择一个 $n\times n$ 的加密矩阵 E 使得 $\det E$ 无法被 2 或 13 整除，则 $\det E$ 的数字有一个模 26 的倒数，矩阵 E 也有模 26 的逆矩阵。逆矩阵 $E^{-1}\equiv D$(模 26)会是解码信息的解密矩阵。

现在转换信息的每个字母变成 0 到 25 的数字，A = 0 到 Z = 25 的明显选择是可以接受的，这是因为矩阵会使得密码变强。

忽略空间并且把信息分成大小为 n 的方块 $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$，然后用加密矩阵 E 去乘每个信息方块(模 p)，编码信息是 $E\boldsymbol{v}_1, E\boldsymbol{v}_2, \ldots$，然后你知道解码器会怎么做。

Spikler 的范例有 $D=E^{-1}=\begin{bmatrix}2&3&15\\5&8&12\\1&13&4\end{bmatrix}^{-1}\equiv\begin{bmatrix}10&19&16\\4&23&7\\17&5&19\end{bmatrix}\ (\text{模 }26)$

$\det E=583\equiv 11(\text{模 }26)$

当然一个破译器并不知道 E 或 D，方块大小 n 通常也是未知，希尔心中的 n 不会太大，电脑可以快速的发现 E 与 D。

我不太确定希尔密码在使用很大的矩阵以及很大的素数 p 的情况下会太难破解，应用二次编码到已编码信息，使用不同的方块大小 n_2 与大矩阵 E_2 以及大素数 p。

有限场与有限向量空间

在代数中，一个场 **F** 是一个纯量的集合，这些纯量可以加，乘，倒数(除了 0 以外)，熟悉的例子是实数 **R** 以及复数 **C** 以及有理数 **Q**(包含每个整数的比值 p/q)。从一个场你建立向量 $\boldsymbol{v}=(f_1, \ldots, f_n)$，从向量的线性组合你建立向量空间，所以线性代数从一个场 **F** 开始。

我从一本从场开始的书教学 10 年，在去 $\mathbf{R}^n$ 的途中，我丢失了一大堆学生，这是一个信号--如果我们想要课程有用，就不能不适当的强调。

请尊重版权与译者的劳动成果，侵权必究!

我相信正确的方法是先了解 $\mathbf{R}^n$ 以及它的子空间，就像你现在做的一样。然后你可以去检视其他场与向量空间，心中可以自然的存在一个疑问：当场不是 **R** 时，会有什么新奇事情？

这些页的内容是在询问**有限场**(finite field)的问题，可能性变得更加限制但是也有高度乐趣。初始点(不是终点)是有限场 $\mathbf{F}_p$，它只包含数字 0, 1, ..., $p-1$ 而且 p 是一个素数。我首先关注场 $\mathbf{F}_2$，它只有两个数字“**0**”与“**1**”，你可以想象 **0** 与 **1** 是“偶数”与“奇数”，因为偶数与奇数的加法与乘法的规则都是遵守的：**偶 + 奇 = 奇，偶 × 奇 = 偶**。

加法表

	0	1
0	0	1
1	1	0

乘法表

	0	1
0	0	0
1	0	1

这是加法与乘法“模 2”。

从场 $\mathbf{F}_2$ 来看，我们可以建立比如 $\boldsymbol{v} = (0, 0, 1)$与 $\boldsymbol{w} = (1, 0, 1)$的向量，每个向量有 3 个分量，每个分量有 2 个选择：总共有 $2^3 = 8$ 个不同向量在向量空间$(\mathbf{F}_2)^3$ 中。

a) 零维子空间只包含 $\mathbf{0} = (0, 0, 0)$。

b) 一维子空间包含 **0** 与类似 $\boldsymbol{v}$ 的向量，注意 $\boldsymbol{v} + \boldsymbol{v} = \mathbf{0}$!

c) 二维子空间有一组类似 $\boldsymbol{v}$ 与 $\boldsymbol{w}$ 的基底，以及 4 个向量 $\mathbf{0}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v} + \boldsymbol{w}$。

d) 满三维子空间$(\mathbf{F}_2)^3$ 有 8 个向量。

什么是$(\mathbf{F}_2)^3$ 可能的基底？标准基底包含(1, 0, 0)与(0, 1, 0)与(0, 0, 1)，这些向量是线性无关，他们生成$(\mathbf{F}_2)^3$，他们系数 0 与 1 的 8 个组合填满整个$(\mathbf{F}_2)^3$。

什么是乘这些向量的矩阵？这些矩阵会是 1×3 或 2×3 或 3×3，当他们是 3×3，我们会问他们是否可逆，他们的行列式只可能是 0(奇异矩阵)或 1(可逆矩阵)。判断下列矩阵是否可逆的乐趣，我就留给你们了。**你怎么求出他们的逆矩阵**？

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

在 $\mathbf{F}_2$ 有 2^9 个可能的矩阵，我猜大部分都是奇异。

结束对于 $\mathbf{F}_2$ 的讨论，我提出一个有 $2^2 = 4$ 个成员的一个场。它不是来自乘法(模 4)，因为 4 不是素数，2 乘 2 会得到 0(2 没有倒数)：不是一个场。但是我们可以从 $\mathbf{F}_2$ 的 0 与 1 开始，再发明另外两个数 a 与 $1 + a$––假设他们遵守两个规则：**$(a + a = 0)$与$(a \times a = 1 + a)$**，则 a 与 $1 + a$ 是倒数。不明显!

加	0	1	a	$1+a$
0	0	1	a	$1+a$
1	1	0	$1+a$	a
a	a	$1+a$	0	1
$1+a$	$1+a$	a	1	0

乘	0	1	a	$1+a$
0	0	0	0	0
1	0	1	a	$1+a$
a	0	a	$1+a$	1
$1+a$	0	$1+a$	1	a

超过 $p = 2$ 时，我们的场是 $\mathbf{F}_p$，其中 p 是所有的素数，他们使用模 p 的加法与乘法，他们是代码的字母，他们在$(\mathbf{F}_p)^n$ 提供分量给向量 $\boldsymbol{v} = (f_1, ..., f_n)$。这些场 $\mathbf{F}_p$ 是最常使用的有限场。

其他唯一的有限场 p^k 个成员，上述的例子 0, 1 , a, $1 + a$ 有 $2^2 = 4$ 个成员。我们把它留在这里然后安全的回到 $\mathbf{R}$。

问题集 10.7

1 如果你乘 n 个数(偶或奇)，何时会得到奇数的答案？以(模 2)的语言来说：如果你乘 0’与 1’s，何时答案是 1？

2 如果你加 n 个数(偶或奇)，何时会得到奇数的总和？如果你加 0’与 1’s(模 2)，何时答案是 1？

3 (a) 若 $y_1 \equiv x_1$ 与 $y_2 \equiv x_2$，为什么 $y_1 + y_2 \equiv x_1 + x_2$(模 p)？
建议：$y_1 = pq_1 + x_1$ 与 $y_2 = pq_2 + x_2$，现在 $y_1 + y_2$。
(b) 你能确认 $x_1 + x_2$ 小于 p？不！给出一个例子，存在一个较小的 x 使得$(y_1 + y_2) = x$(模 p)？

4 $p = 39$ 不是素数，找出一个 a 使得不存在倒数 z(模 39)。这表示 $az \equiv 1$(模 39)无解。找出一个矩阵 A 不存在逆矩阵 Z(模 39)，这表示 $AZ \equiv I$(模 39)无解。

5 证明 $y \equiv x$(模 p)可以得到$-y \equiv -x$(模 p)。

6 找出一个有无关列的 $\mathbf{R}^2$ 矩阵，但是在(模 5)下，矩阵有相关列。

7 2×2 矩阵由 0’s 与 1’s 组成，在(模 2)下，全部的可逆矩阵为何？

8 在模运算(模 11)下，A 的行空间仍然与零空间正交吗？在(模 11)下，这些子空间的基底仍然是基底吗？

9 (希尔密码) 把 THISWHOLEBOOKISINCODE 分成 3 个字母的方块。把每个字母用一个 1 到 26 的数字(正常顺序)取代，用一个 3×3 矩阵 L(对角线及对角线下方都是 1)乘每个方块，编码后的信息(数字表示)为何？你又如何解码？

10 假设你知道原始信息(平面文字)，假设你也看到编码后信息，你如何开始检测希尔密码的矩阵？对一个很长的信息，你期待成功吗？

请尊重版权与译者的劳动成果，侵权必究!

第十一章
数值线性代数

1 数值线性代数的目的是**速度与精确与稳定**：$n > 10^3$ 或 10^6。

2 矩阵可以是满或稀疏或带状或结构：每个都有特殊演绎法。

3 消元法的精确度由**条件数**$\|A\|\ \|A^{-1}\|$决定。

4 格莱姆-施密特的计算通常使用**豪斯霍德镜射** $H = I - 2uu^{\mathrm{T}}$ 来求 Q。

5 固有值使用 ***QR* 迭代** $A_0 = Q_0R_0 \to R_0Q_0 = A_1 = Q_1R_1 \to \to A_n$。

6 **平移 *QR*** 更好：移到 $A_k - c_kI = R_kQ_k$，移回 $A_{k+1} = R_kQ_k + c_kI$。

7 若 $S^{-1}T$ 所有的固有值有$|\lambda| < 1$，则迭代 $Sx_{k+1} = b - Tx_k$ 可以解答$(S + T)x = b$。

8 迭代法通常使用**预条件子** P，把 $Ax = b$ 变成 $PAx = Pb$，其中 $PA \approx I$。

9 **共轭梯度与 GMRES** 是克雷洛夫方法；参考 Trefethen-Bau(以及其他书本)。

11.1 实际的高斯消元法

数值线性代数是为了快速求解与精确求解的奋斗，我们需要效率但也要避免不稳定。高斯消元法中，主要的自由度(一直可得)是**交换方程式**。本段落说明为了速度何时交换行，以及为了精确何时交换行。

精确的关键是避免不必要的大数字，通常也是要求我们避免小数字！一个好计划是“**部分枢轴**”，在每个新列选择最大可用枢轴，我们会看到为什么这个枢轴策略会进入计算机程序当中。

其他的行交换是要节省消元步骤，实际上大部分的大型矩阵都是**稀疏**——几乎全部的单元都是零。

当方程式的顺序安排可以产生**非零窄带**(nonzero narrow band)时，消元法最快速。带状中的零在消元过程中被填满--这些零被摧毁而且不会节省电脑时间。

段落 11.2 有关不能避免的不稳定，它是内建在问题中，敏感度用**条件数(condition number)**来量测。然后 11.3 描述如何运用迭代求解 $A\boldsymbol{x} = \boldsymbol{b}$，我们不是直接求解，而是电脑求解一个较简单的问题很多次，每个答案 $\boldsymbol{x}_k$ 得到新的猜测 $\boldsymbol{x}_{k+1}$。优质的迭代(**共轭梯度法**是超级棒)，$\boldsymbol{x}_k$ 快速收敛到 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$。

最快速的超级电脑

最新的超级计算记录是 IBM 与 Los Alamos 在 2008 年 5 月发表，**路跑者**测试是第一个达成每秒 10^{15} 浮点数运算：每秒 1 千万亿的计算。这个记录的衡量标准是一个大型浓密系统 $A\boldsymbol{x} = \boldsymbol{b}$：利用线性代数测试电脑速度。

那台机器在 2013 年关闭！TOP500 计划为世界上 500 个最高能电脑系统评定等级，当我在 2015 年 10 月写下这一页时，前四是来自中国的 NUDT，美国的 CRAY 与 IBM，以及日本的富士通，他们全部使用 LINUX 基础的系统，所有的向量处理器都在 top500。

往前看，Summit 是被期待有第一名的实力，每秒 150-300 千万亿。奥巴马总统刚刚命令这项百万兆级系统的发展(1000 千万亿)，直到现在都是遵循每 14 个月加倍的 Moore 定律。

LAPACK 软件是利用“部分枢轴”执行消元法，与本书最大的不同是使用大型子矩阵以及不使用单一数字。图形处理单元(GPU's)都是成功所必需，视频市场使得科学计算相对渺小，导致芯片惊人的加速发展。

在 IBM 的蓝色基因之前，一个关键课题是计算一个千万亿机器需要 4 核心处理器的个数：32,000 个。新的架构用电更少，但是混合设计有它的成本：一个代码需要 3 个个别编译与明确指令来移动所有的资料。请参阅 SIAM 的优秀文章(siam.org, 7 月，2008)，以及网站 www.lanl.gov/roadrunner。

我们对于矩阵计算的想法反映在高度最佳化 **BLAS(基本线性代数子程序)**，他们有层次 1, 2, 3：

层次 1　向量 $a\boldsymbol{u} + \boldsymbol{v}$ 的组合：$O(n)$个工作

层次 2　矩阵-向量乘法 $A\boldsymbol{u} + \boldsymbol{v}$：$O(n^2)$个工作

层次 3　矩阵-矩阵乘法 $AB + C$：$O(n^3)$个工作

层次 1 是消元步骤(从行 i 减去 l_{ij} 乘行 j)，层次 2 可以一次消除一整列，一个高性能求解器在层次 3 的 **BLAS** 已经很足够(AB 有 $2n^3$ 的移动与 $2n^2$ 资料，好的工作比例。)

传送资料与提取储存限制了平行处理的速度，在主记忆体与浮点计算之间的高速快取必须全面使用！顶级速度需要针对消元法的**方块矩阵法**。

大变化现在来临，在芯片层次的平行处理。

请尊重版权与译者的劳动成果，侵权必究!

舍去误差与部分枢轴

直到现在，任意枢轴(当然是非零)都可以接受，实际上小的枢轴是危险的，当不同大小的数字相加会引起大灾难。电脑用有效位数(digit，比如 3 位小数，很小的机器)保存一个固定数，10,000 + 1 的总和舍去变成 10,000，“1”完全丢失，看看问题的解如何变化：

$$\begin{aligned} .0001u+v&=1\\ -u+v&=0 \end{aligned} \qquad \text{从系数矩阵开始} \qquad A=\begin{bmatrix} .0001 & 1\\ -1 & 1 \end{bmatrix}$$

如果我们接受.0001 作为枢轴，消元法会把 10,000 乘行 1 加到行 2，舍去后留下

$$10{,}000v=10{,}000 \quad \text{而不是} \quad 10{,}001v=10{,}000$$

计算得到答案 $v=1$ 与真实 $v=.9999$ 接近。但是之后的方向代入法把错误的 $v=1$ 放进 u 的方程式：

$$.0001u+1=1 \quad \text{而不是} \quad .0001u+.9999=1$$

第一个方程式得到 $u=0$，正确的答案(检视第二个方程式)是 $u=1.000$。因为丢失矩阵中的“1”，我们丢失了解。**从 10,001 到 10,000 的小变化导致答案从 $u=1$ 到 $u=0$**(100%错误！)。

如果我们交换行，纵使这部小电脑也能找正确答案到 3 位小数：

$$\begin{aligned} -u+v&=0\\ .0001u+v&=1 \end{aligned} \quad\rightarrow\quad \begin{aligned} -u+v&=0\\ v&=1 \end{aligned} \quad\rightarrow\quad \begin{aligned} u&=1\\ v&=1 \end{aligned}$$

原始的枢轴是.0001 与 10,000——差劲的比例，交换行之后，确切的枢轴是–1 与 1.0001——良好比例，计算的枢轴–1 与 1 跟确切值非常接近。小的枢轴带来数值的不稳定，补救的方法就是**部分枢轴(partial pivoting)**。这里是我们抵达并且搜寻列 k 作为最佳枢轴的策略：

选择行 k 或下方最大的数字，交换它的行与行 k

完全枢轴的策略也会寻找后面的列来找到最大枢轴，跟行一样，它也会交换列。这个代价是不合理的，所有的主流代码都是使用部分枢轴。用一个比例常数乘行或列是非常值得的，如果上述第一个方程式是 $u+10{,}000v=10{,}000$，我们不调整比例，则 1 看起来很像优质枢轴，使得我们可能会错失重要的行交换。

对于正定矩阵来说，不需要行交换，接受枢轴出现的方式是安全的。小枢轴会发生，但是矩阵不会因为行交换而改善。当它的条件数很高时，问题是在于矩阵而不是在于代码，这个情况下输出是无可避免的对于输入很敏感。

读者现在了解实际上电脑如何求解 $A\boldsymbol{x}=\boldsymbol{b}$——**使用部分枢轴的消元法**，与原理描述比较——**求 A^{-1} 然后计算 $A^{-1}\boldsymbol{b}$**——这很花时间。但是在电脑时间，消元法很快速。我相信消元法也是行空间与零空间代数的最佳途径。

这是有关成本的实际问题，应用消元法需要多少次的个别运算才能求解 $A\boldsymbol{x}=\boldsymbol{b}$？这个决定在我们能够处理问题的大小。

首先观察 A，什么样的改变逐渐进入 U？当我们从行 2 减去一个乘数乘行 1，我们做了 n 次运算。首先是除以枢轴，找出乘数 l，对沿着这一行的其他 $n-1$ 个单元来说，这项运算是一个“乘--减”。方便起见，我们把它当成单一运算来计数。如果你认为用 l 去乘与从现存单元减去是两个运算，就把我们的计数乘 2。

矩阵 A 是 $n\times n$，这项运算计数应用到在第一行下方的其他全部 $n-1$ 行，它需要 n 乘 $n-1$ 次运算，或是 n^2-n，用来产生第一枢轴下方的零。检查：除了第一行的 n 个单元，全部 n^2 个单元都改变了。

当消元法降到 k 个方程式，行变短了，我们只需要 k^2-k 次运算(而不是 n^2-n)去清除枢轴下方的列，对于 $1\le k\le n$ 是成立的。最后一个步骤不需要运算($1^2-1=0$)；前向消元完成。得到 U 的全部计数是 k^2-k 从 k 等于 1 到 n 的总和：

$$\left(1^2+\ldots+n^2\right)-\left(1+\ldots+n\right)=\frac{n\left(n+1\right)\left(2n+1\right)}{6}-\frac{n\left(n+1\right)}{2}=\frac{n^3-n}{3}$$

这些是前 n 个数与其平方总和的已知公式，代入 $n=100$ 得到一百万减一百--然后除以 3 (这个换成工作站的一秒钟)。我们忽略 n 与 n^3 的比较，得到我们的主要结论：

对应前向消元(A 到 U，产生 L)的“乘—减计数”是 $n^3/3$

这个表示 $n^3/3$ 次乘法与减法，加倍的 n，成本会变成 8 倍(因为 n 是立方)。100 个方程式很简单，1000 更贵，10000 个浓密矩阵接近不可能完成。我们需要更快的电脑或是一大堆零或是更新的概念。

在方程式的右侧，步骤走的更快，我们对单一数字运算，不是一整行，**每个右侧需要确切的 n^2 次运算**。往下与往上我们是在求解两个三角形系统，前向的 $L\boldsymbol{c}=\boldsymbol{b}$ 与反向的 $U\boldsymbol{x}=\boldsymbol{c}$。反向代入法中，最后的未知数只需要除以最后一个枢轴，在它上方的方程式需要两个运算--代入 x_n 然后除以它的枢轴。第 k 步需要 k 次“乘--减”运算，全部的方向代入法需要

$$\left(1+1+2\ldots+n\right)=\frac{n\left(n+1\right)}{2}\approx\frac{n^2}{2}\quad\text{次运算}$$

前向的部分很类似。n^2 的总数确切等于计算 $A^{-1}\boldsymbol{b}$ 的乘法次数！这个产生了高斯消元法胜过 $A^{-1}\boldsymbol{b}$ 的两大优势：

1　消元法需要 $n^3/3$ 次乘法与减法，相较之下 A^{-1} 需要 n^3。

2　若 A 是带状，L 与 U 也会是带状：相较之下 A^{-1} 充满非零数。

请尊重版权与译者的劳动成果，侵权必究！

带状矩阵

当 A 有“**优质零**”时，运算计数会改善，一个优质零是停留在 L 与 U 的零。**最佳零是在一行的开始**，他们不需要消元步骤(乘数是零)，所以我们也要找出在 L 的相同优质零，对于三对角矩阵 A(以及图 11.1 的带状矩阵)：

三对角
双对角
乘
双对角

$$\begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \\ & & & 1 \end{bmatrix}.$$

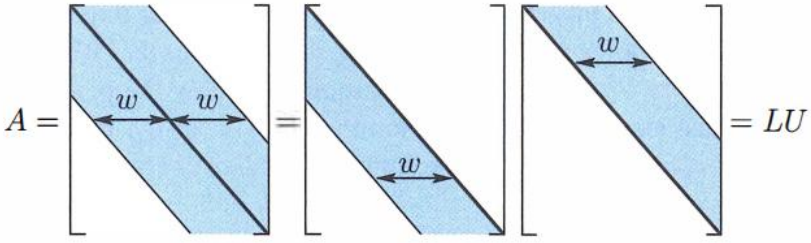


图 11.1：带状矩阵的 $A = LU$，A 的优质零在 L 与 U 还是零

这些零对于运算计数产生完全的改变，对于“**半-带宽**”w：

一个带状矩阵在 $|i-j| > w$ 时，有 $a_{ij} = 0$

于是 $w = 1$ 是对角矩阵，$w = 2$ 是三对角，$w = n$ 是浓密。枢轴行的长度最多是 w，在任意枢轴下方不会有超过 $w-1$ 的非零数。消元法的每个步骤在 $w(w-1)$次运算后完成，仍然维持带状结构。有 n 个列要去除，所以

带状矩阵执行消元法(A 到 L 与 U)需要小于 w^2n 次运算

对于带状矩阵，计数正比于 n 而不是 n^3，它也正比于 w^2。一个满矩阵有 $w = n$ 而且我们回到 n^3。针对确切的计数，记得带宽在右下角掉到小于 w(空间不够)：

$$\text{带状}\ \frac{w(w-1)(3n-2w+1)}{3} \qquad \text{浓密}\ \frac{n(n-1)(n+1)}{3} = \frac{n^3-n}{3}$$

在 $A\boldsymbol{x} = \boldsymbol{b}$ 的右侧，从 $\boldsymbol{b}$ 求 $\boldsymbol{x}$，成本大约是 $2wn$ (与平常的 n^2 对比)。重点：**针对一个带状矩阵的运算次数与 n 成正比**，这是非常的快速。假设我们不计算 A^{-1}，阶数 10000 的三对角矩阵很便宜。反矩阵一个零都没有：

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \quad \text{有} \quad A^{-1} = U^{-1}L^{-1} = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 3 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

我们实际上知道 A^{-1} 还不如知道 L 与 U 来得好，用 A^{-1} 去乘需要完整的 n^2 个步骤，求解 $L\boldsymbol{c} = \boldsymbol{b}$ 与 $U\boldsymbol{x} = \boldsymbol{c}$ 只需要 $2wn$。

带状结构实际上是常见的，当矩阵反映近邻的连接情形：$a_{13}=0$ 与 $a_{14}=0$，因为 1 不是 3 与 4 的近邻。

我们以高斯-乔丹与格莱姆-施密特-豪斯霍尔德的计数作为结束：

A^{-1} 的成本 n^3 次乘法与减法。	QR 的成本 $2n^3/3$ 步。

在 $AA^{-1}=I$ 中，A^{-1} 的第 j 列是 $A\boldsymbol{x}_j=I$ 的第 j 列的解，左侧平常的成本 $n^3/3$(这是一次的成本！L 与 U 没有重复。) 对于 I 的第 j 列的特殊节省是来自它的前 $j-1$ 个零，直到消元法抵达行 j 之前，右侧都不需要工作。前向成本是 $(n-j)^2/2$ 而不是 $n^2/2$。对 j 累加，前向消元的 n 个右侧总成本是 $n^3/6$，如果我们实际想要用 A^{-1}，最终对应 A^{-1} 的总成本是 n^3：

$$\text{对 } A^{-1} \qquad \frac{n^3}{3}(L\text{与}U)+\frac{n^3}{6}(\text{前向})+n\left(\frac{n^2}{2}\right)(\text{反向代入})=n^3 \tag{1}$$

正交化　(A 到 Q)：来自消元法的关键差异是每个乘数由一个点积决定，需要 n 个运算，其中消元法只是除以枢轴，然后有 n 个“乘—减”运算去移除来自列 k 沿着列 $j<k$ 的投影(段落 4.4)，组合成本是 $2n$，其中消元法是 n，因子 2 是正交的成本。当消元法把一个单元变成零时，我们是在把点积变成零。

警告　要评价一个演绎法，只计数运算次数是不够的，在“失败计数”之上是稳定的研究(豪斯霍尔德获胜)以及资料流。

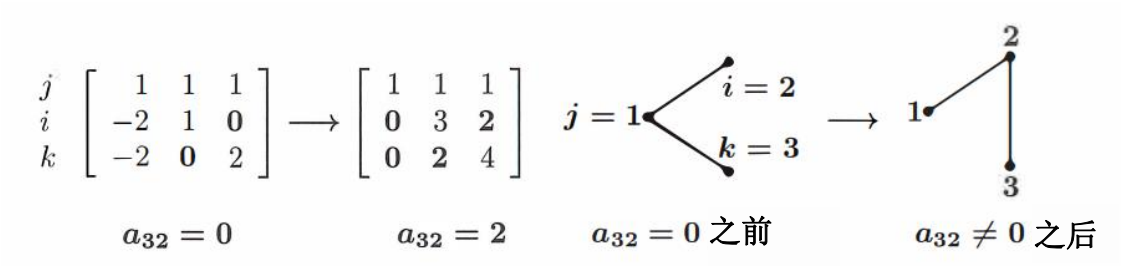
重新排序稀疏矩阵

对于常数宽度 w 的带状矩阵，行的排序是最佳。但是对于大部分实际计算的稀疏矩阵来说，带宽并不是常数而且在带状里面存在很多零，当消元法进行时，这些零会被填满—他们丢失了。我们需要重新计数方程式的个数来降低填满，因此可以加速消元法。

一般来说，我们希望把零移到较为前面的行与列，后面的行与列本身就比较短。“近似极小次数(**a**pproximate **m**inimum **d**egree)”演绎法在稀疏 MATLAB 是贪婪的—在没有计算全部结果的情况下选择要消去的行。我们在接近末端时可以得到近似的满矩阵，但是得到 LU 的总运算次数仍然非常小。要找到 L 与 U 中绝对极小的零是一个 NP-难题，太昂贵了，**amd** 是一个可以接受的折中。

当方形格点的每个点连接到它的 4 个最近邻居时，填满是著名的，不可能把所有的格点编号使得邻居维持在一起！如果我们使用格子的行去编号，来到上述格点会是一个漫长的等待。

请尊重版权与译者的劳动成果，侵权必究!



我们只需要非零数的位置，而不是他们的确切值。思考非零数的图形：若 $a_{ij} \neq 0$ 则节点 i 与 j 相连，观察消元法如何产生非零(新的边)，这是我们要避免的。

指令 nnz(L)计算下三角 L 的非零数的个数，find(L)会列表，spy(L)显示他们。**colamd** 与 **symamd** 是较好的排序(排列 P)产生 AP 与 $P^{\mathrm{T}}AP$ 的填满--选择下方有最少非零数的枢轴。

快速正交化

有 3 种方法得到重要的分解 $A = QR$，格莱姆-施密特求出 Q 中的正交单位向量，R 是上三角，这是因为格莱姆-施密特步骤的顺序。现在我们检视较好的方法(豪斯霍尔德与吉文斯)，使用特殊简单 Q's 的乘积，我们知道这些 Q 是正交。

消元法得到 $A = LU$，正交化得到 $A = QR$，我们不想要三角 L，我们想要正交 Q。L 是来自消元法 E's 的乘积，有 1 在对角线且乘数 l_{ij} 在下方，Q 会是正交矩阵的乘积。

有两个简单的正交矩阵可以取代 E's，**镜射矩阵** $I - 2\boldsymbol{uu}^{\mathrm{T}}$ 是以豪斯霍尔德命名，**平面旋转矩阵**是以吉文斯命名。把 xy 平面旋转 θ 的简单矩阵是 Q_{21}：

在 1-2 平面的吉文斯旋转 $\quad Q_{21} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

如同使用 E_{21} 一样的使用 Q_{21}，产生在位置(2, 1)的零，这个决定在 θ。比尔哈格在“应用数值线性代数”给出例子：

$$Q_{21}A = \begin{bmatrix} .6 & .8 & 0 \\ \mathbf{-.8} & \mathbf{.6} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{90} & -153 & 114 \\ \mathbf{120} & -79 & -223 \\ 200 & -40 & 395 \end{bmatrix} = \begin{bmatrix} 150 & -155 & -110 \\ \mathbf{0} & 75 & -225 \\ 200 & -40 & 395 \end{bmatrix}$$

这个零来自$-.8(90) + .6(120)$，不需要找出 θ，我们只需要 $\cos\theta$ ：

$$\cos\theta = \frac{90}{\sqrt{90^2 + 120^2}} \quad \text{与} \quad \sin\theta = \frac{-120}{\sqrt{90^2 + 120^2}} \tag{2}$$

现在我们攻击(3, 1)单元。旋转会在行 3 与列 1，数字 $\cos\theta$与 $\sin\theta$由 150 与 200 决定，而不是 90 与 120。

$$Q_{31}Q_{21}A=\begin{bmatrix} .6 & 0 & .8 \\ 0 & 1 & 0 \\ -\mathbf{.8} & 0 & \mathbf{.6} \end{bmatrix}\begin{bmatrix} \mathbf{150} & . & . \\ 0 & . & . \\ \mathbf{200} & . & . \end{bmatrix}=\begin{bmatrix} 250 & -125 & 250 \\ 0 & 75 & -225 \\ \mathbf{0} & 100 & 325 \end{bmatrix}$$

再一步到 R，(3, 2)单元要移除。现在 $\cos\theta$与 $\sin\theta$来自 75 与 100，旋转会在行 2 与列 3:

$$Q_{32}Q_{31}Q_{21}A=\begin{bmatrix} 1 & 0 & 0 \\ 0 & .6 & .8 \\ 0 & -\mathbf{.8} & \mathbf{.6} \end{bmatrix}\begin{bmatrix} 250 & -125 & . \\ 0 & \mathbf{75} & . \\ 0 & \mathbf{100} & . \end{bmatrix}=\begin{bmatrix} 250 & -125 & 250 \\ 0 & 125 & 125 \\ 0 & \mathbf{0} & 375 \end{bmatrix}$$

我们得到上三角 R，Q 是什么？把平面旋转 Q_{ij} 移到另一侧求出 $A = QR$——就好像你把消元矩阵 E_{ij} 移到另一侧求得 $A = LU$ 一般：

$$Q_{32}Q_{31}Q_{21}A = R \quad \text{就是} \quad A = (Q_{21}^{-1}Q_{31}^{-1}Q_{32}^{-1})R = QR \tag{3}$$

每个 Q_{ij} 的逆矩阵是 Q_{ij}^{T}(旋转$-\theta$)，每个 E_{ij} 的逆矩阵不是正交矩阵！LU 与 QR 相似，但是 L 与 Q 不相同。

豪斯霍尔德镜射比旋转来的快，因为每次清除对角线下方的一整列，观察 A 的第一列 $\boldsymbol{a}_1$ 如何变成 R 的列 $\boldsymbol{r}_1$：

由 H_1 镜射 $\quad H_1 = I - 2\boldsymbol{u}_1\boldsymbol{u}_1^{\mathrm{T}}$

$$H_1\,\boldsymbol{a}_1 = \begin{bmatrix} \|\boldsymbol{a}_1\| \\ 0 \\ . \\ 0 \end{bmatrix} \quad \text{或} \quad \begin{bmatrix} -\|\boldsymbol{a}_1\| \\ 0 \\ . \\ 0 \end{bmatrix} = \boldsymbol{r}_1 . \tag{4}$$

长度没有改变，$\boldsymbol{u}_1$ 是在 $\boldsymbol{a}_1 - \boldsymbol{r}_1$ 的方向。我们有单位向量 $\boldsymbol{u}_1$ 中的 $n-1$ 个单元得到 $\boldsymbol{r}_1$ 中的 $n-1$ 个零。(旋转有一个角θ 得到一个零。) 当我们抵达列 k，我们在单位向量 $\boldsymbol{u}_k$ 中有 $n-k$ 个可用选择，这个产生 $\boldsymbol{r}_k$ 的 $n-k$ 个零。我们只需要储存 $\boldsymbol{u}$'s 与 $\boldsymbol{r}$'s 就可以知道最终的 Q 与 R：

H_i 的逆矩阵是 H_i $\quad (H_{n-1} \ldots H_1)A = R \quad$ 就是 $\quad A = (H_1 \ldots H_{n-1})R = QR \tag{5}$

这是 LAPACK 对于 19 世纪格莱姆-施密特的改善，Q 确切是正交。

段落 11.3 说明 $A = QR$ 如何使用在其他大型线性代数的计算——**固有值问题**，因子 QR 反向得到 $A_1 = RQ$ 也就是 $Q^{-1}AQ$。由于 A_1 与 A 相似，固有值没有改变，A_1 分解成 Q_1R_1，再反转顺序得到 A_2。惊人的是在 $A_1, A_2, A_3, \ldots$对角线下方的单元会越来越小，我们就可以鉴别固有值，这是针对 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 的“QR 方法”，数值线性代数的一大成功。

请尊重版权与译者的劳动成果，侵权必究!

问题集 11.1

1 分别使用与不使用行交换最大化枢轴，求出两个枢轴：

$$A=\begin{bmatrix}.001 & 0\\ 1 & 1000\end{bmatrix}$$

有行交换去极大化枢轴，为什么 L 没有单元大于 1？求出 3×3 矩阵 A，其中$|a_{ij}| \le 1$ 且$|l_{ij}| \le 1$ 但是第三枢轴 $=4$。

2 利用消元法计算希尔伯特矩阵 A 的确切逆矩阵，然后舍去三位小数之后的数，再次计算 A^{-1}。

$$\text{不良条件矩阵} \qquad A=\text{hilb}(3)=\begin{bmatrix}1 & \frac{1}{2} & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4}\\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5}\end{bmatrix}$$

3 对于相同的 A 计算 $\boldsymbol{b}=A\boldsymbol{x}$，其中 $\boldsymbol{x}=(1,1,1)$与 $\boldsymbol{x}=(0,6,-3.6)$。小的变化$\Delta\boldsymbol{b}$导致大的变化$\Delta\boldsymbol{x}$。

4 求出 8×8 希尔博特矩阵 $a_{ij}=1/(i+j-1)$的固有值(使用电脑)。方程式 $A\boldsymbol{x}=\boldsymbol{b}$ 的$\|\boldsymbol{b}\|=1$，$\|\boldsymbol{x}\|$可以到多大？如果 $\boldsymbol{b}$ 的舍去误差小于 10^{-16}，出现在 $\boldsymbol{x}$ 的误差会有多大？见段落 9.2。

5 针对带状矩阵(带宽 w)进行反向代入，说明求解 $U\boldsymbol{x}=\boldsymbol{c}$ 的乘法次数大约是 wn。

6 如果你知道 L, U, Q, R，求解 $LU\boldsymbol{x}=\boldsymbol{b}$ 或是求解 $QR\boldsymbol{x}=\boldsymbol{b}$ 会比较快？

7 证明逆反一个 $n\times n$ 的上三角矩阵的乘法次数大约是 $n^3/6$。对 I 的列使用反向代入，从 1's 往上。

8 在每一列选择最大可用的枢轴(部分枢轴)，把每个 A 分解成 $PA=LU$：

$$A=\begin{bmatrix}1 & 0\\ 2 & 2\end{bmatrix} \quad \text{与} \quad A=\begin{bmatrix}1 & 0 & 1\\ 2 & 2 & 0\\ 0 & 2 & 0\end{bmatrix}$$

9 把 1's 放在一个 4×4 三对角矩阵的三个中心对角线，求出 6 个零单元的余因子。这些单元是 A^{-1} 的非零数。

10 (C. Van Loan 建议) 求出 LU 分解并且利用消元法求解，其中 $\varepsilon=10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}, 10^{-15}$：

$$\begin{bmatrix}\varepsilon & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}1+\varepsilon\\ 2\end{bmatrix}$$

真实的 $\boldsymbol{x}$ 是(1, 1)，对于每个ε 制作一个误差表。交换两个方程式再求解——误差几乎消失不见。

网址：linear.neocities.org 邮箱：876152660@qq.com

11 (a) 选择 $\sin\theta$与 $\cos\theta$来三角化 A，并求 R：

吉文斯旋转 $$Q_{21}A=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}1 & -1\\ 3 & 5\end{bmatrix}=\begin{bmatrix}* & *\\ 0 & *\end{bmatrix}=R$$

(b) 选择 $\sin\theta$与 $\cos\theta$ 使得 QAQ^{-1} 成为三角形，固有值为何？

12 用一个平面旋转 Q_{ij} 乘 A，会改变 A 的哪些单元？ 当 $Q_{ij}A$ 在右侧乘 Q_{ij}^{-1}，会改变哪些单元？

13 计算 $Q_{ij}A$ 需要多少次乘法与多少次加法？仔细安排整个旋转序列得到 $2n^3/3$ 次乘法与 $2n^3/3$ 加法——与使用镜射的 QR 相同，是 LU 的两倍。

挑战问题

14 (**转动机械手**) 机械手产生任意 3×3 旋转矩阵 A，他们来自环绕 x, y, z 轴的平面旋转，则 $Q_{32}Q_{31}Q_{21}A=R$，其中 A 是正交，所以 R 是 I！三个机械手转动在 $A=Q_{21}^{-1}Q_{31}^{-1}Q_{32}^{-1}$。3 个角是“欧拉角”而且 $\det Q=1$ 避免镜射。从选择 $\cos\theta$ 与 $\sin\theta$ 开始，使得

$$Q_{21}A=\begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}\frac{1}{3}\begin{bmatrix}-1 & 2 & 2\\ 2 & -1 & 2\\ 2 & 2 & -1\end{bmatrix}\text{有零在(2, 1)位置}$$

15 创建 10×10 第二差分矩阵 K = toeplitz([2 **–1** zeros(1, 8)])，利用 $KK=K$(randperm(10), randperm(10))随机排列行与列，利用[L, U] = lu(K)与[LL, UU] = lu(KK)的分解，然后利用 nnz(L)与 nnz(LL)计数非零数。这个情况下 L 是完美三对角，但不是 LL。 【负号与之 1 间不能有空格】

16 这个矩阵 K 的另一种排序是对于网点交替涂上红色与黑色，这个排列 P 把正常的 1 到 10 变成 1, 3, 5, 7, 9, 2, 4, 6, 8, 10：

红黑排序 $PKP^{\mathrm{T}}=\begin{bmatrix}2I & D\\ D^{\mathrm{T}} & 2I\end{bmatrix}$，求出矩阵 D。

很多有趣的试验都是可能的，如果你传送好的点子，他们可以在网站 math.mit/linearalgebra 上持续。我也推荐学习指令 B = sparse(A)，在 find(B)之后，其中 find(B)会列出非零单元以及 lu(B)会使用稀疏格式的 L 与 U 分解 B。只有非零数会被计算，其中 ordinary(dense)MATLAB 也会计算所有的零。

17 Jeff Stuart 创建一个学生活动聪明的展示**不良条件**：

$$\begin{bmatrix}1 & 1.0001\\ 1 & 1.0000\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}3.0001+e\\ 3.0000+E\end{bmatrix}\qquad \begin{matrix}\text{误差是}\\ e\text{与}E\end{matrix}\qquad \begin{matrix}x=2-10000\left(e-E\right)\\ y=1+10000\left(e-E\right)\end{matrix}$$

当这些方程式用来展示几乎平行的长棍，一个小的震动会导致交点(x, y)的大跳跃，误差 e 与 E 放大 10000 倍。

请尊重版权与译者的劳动成果，侵权必究！

11.2 范数与条件数

我们如何量测矩阵的大小？对向量来说，长度是$\|\boldsymbol{x}\|$，对矩阵来说，范数是$\|A\|$。这个词“范数”有时候向量也用在向量，而不用长度，它一直使用在矩阵，而且有很多方式去量测$\|A\|$。我们检视所有的“矩阵范数”要求，然后选择一个。

弗罗贝尼乌斯取所有的$|a_{ij}|^2$然后相加，它的范数$\|A\|_F$是平方根，这个是把A当成一个有n^2个分量的长向量：有时候很有帮助，但不是这里的选择。

我偏好从向量的范数开始，三角不等式说明$\|\boldsymbol{x}+\boldsymbol{y}\|$不会比$\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$大，$2x$或$-2x$的长度加倍成为$2\|\boldsymbol{x}\|$。相同的规则可以应用在矩阵的范数：

$$\|A+B\|\le\|A\|+\|B\| \qquad \text{与} \qquad \|cA\|=|c|\,\|A\| \tag{1}$$

第二个对于矩阵范数的要求是新的，这是矩阵乘法。范数$\|A\|$控制从$\boldsymbol{x}$到$A\boldsymbol{x}$的成长，以及B到AB的成长：

成长因子$\|A\|$
$$\|A\boldsymbol{x}\|\le\|A\|\,\|\boldsymbol{x}\| \qquad \text{与} \qquad \|AB\|\le\|A\|\,\|B\| \tag{2}$$

这个得到定义$\|A\|$的一个自然方式，矩阵的范数：

矩阵的范数是最大的比值 $\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$：
$$\|A\|=\max_{\boldsymbol{x}\neq\boldsymbol{0}}\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \tag{3}$$

$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$永远不大于$\|A\|$(它的绝对值)，意思是$\|A\boldsymbol{x}\|\le\|A\|\,\|\boldsymbol{x}\|$。

范例 1 若A是单位矩阵I，比值是$\|\boldsymbol{x}\|/\|\boldsymbol{x}\|$，因此$\|I\|=1$。若$A$是正交矩阵$Q$，长度再次被维持：$\|Q\boldsymbol{x}\|=\|\boldsymbol{x}\|$，比值仍然得到$\|Q\|=1$。一个正交$Q$很好计算：误差不会成长。

范例 2 角矩阵的范数是它的最大单元(使用绝对值)：

$$A=\begin{bmatrix}2&0\\0&3\end{bmatrix}\text{的范数是}\|A\|=3,\quad \text{固有向量}\ \boldsymbol{x}=\begin{bmatrix}0\\1\end{bmatrix}\text{有}\ A\boldsymbol{x}=3\boldsymbol{x}$$

固有值是3，对这个A(不是全部的A)来说，最大的固有值等于范数。

对于正定对称矩阵来说，范数是 $\|A\|=\lambda_{\max}(A)$。

选择$\boldsymbol{x}$是对应最大固有值的固有向量，则$\|A\boldsymbol{x}\|/\|\boldsymbol{x}\|$等于$\lambda_{\max}$，重点是没有其他的$\boldsymbol{x}$可以使得这个比值更大。矩阵是$A=Q\Lambda Q^{\mathrm{T}}$而且正交矩阵$Q$与$Q^{\mathrm{T}}$维持长度不变，所以极大化比值是$\|\Lambda\boldsymbol{x}\|/\|\boldsymbol{x}\|$，范数是对角矩阵$\Lambda$的最大固有值。

对称矩阵 假设 A 是对称但不是正定，$A = Q\Lambda Q^{\mathrm{T}}$ 仍然成立，则范数是$|\lambda_1|, |\lambda_2|, ..., |\lambda_n|$的最大值。我们取绝对值，这是因为范数只跟长度有关。对固有向量来说，$\|A\boldsymbol{x}\| = \|\lambda\boldsymbol{x}\| = |\lambda|$乘$\|\boldsymbol{x}\|$，这个给出极大比值的 $\boldsymbol{x}$ 是对应极大值$|\lambda|$的固有向量。

非对称矩阵 假设 A 不是对称，固有值不可以量测它的实际大小，范数可能比任意固有值都来的大。一个非常不对称的例子有$\lambda_1 = \lambda_2 = 0$，但是它的范数不是零：

$$\|A\| > \lambda_{\max} \qquad A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{有范数} \quad \|\boldsymbol{A}\| = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|} = 2$$

向量 $\boldsymbol{x} = (0, 1)$给出 $A\boldsymbol{x} = (2, 0)$，比值是 2/1，这是极大比值$\|A\|$，纵使 $\boldsymbol{x}$ 不是固有向量。

是对称矩阵 $A^{\mathrm{T}}A$ 有固有向量 $\boldsymbol{x} = (0, 1)$，而不是非对称的 A 有这个固有向量。范数实际上是由 $A^{\mathrm{T}}A$ 的最大固有值决定：

A(对称或不对称)的范数是$\lambda_{\max}(A^{\mathrm{T}}A)$的平方根：

$$\|\boldsymbol{A}\|^2 = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{Ax}\|^2}{\|\boldsymbol{x}\|^2} = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \lambda_{\max}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right) \tag{3}$$

$\lambda_{\max}(A) = 0$ 的非对称例子有$\lambda_{\max}(A^{\mathrm{T}}A) = 4$：

$$A = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \text{产生} \quad A^{\mathrm{T}}A = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \text{有} \lambda_{\max} = 4，\text{所以范数是} \|A\| = \sqrt{4}\ 。$$

对任意的 $\boldsymbol{A}$ 选择 $\boldsymbol{x}$ 是对应 $A^{\mathrm{T}}A$ 最大固有值$\lambda_{\max}$ 的固有向量，方程式(4)的比值是 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\left(\lambda_{\max}\right)\boldsymbol{x}$ 除以 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，这个是$\lambda_{\max}$。

没有 $\boldsymbol{x}$ 可以给出更大的比值，对称矩阵 $A^{\mathrm{T}}A$ 有固有值$\lambda_1, ..., \lambda_n$ 以及正交单位固有向量 $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$，每个 $\boldsymbol{x}$ 是这些向量的组合。尝试在比值中的这个组合，不要忘记 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = \boldsymbol{0}$：

$$\frac{\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{\left(c_1\boldsymbol{q}_1 + ... + c_n\boldsymbol{q}_n\right)^{\mathrm{T}}\left(c_1\lambda_1\boldsymbol{q}_1 + ... + c_n\lambda_n\boldsymbol{q}_n\right)}{\left(c_1\boldsymbol{q}_1 + ... + c_n\boldsymbol{q}_n\right)^{\mathrm{T}}\left(c_1\boldsymbol{q}_1 + ... + c_n\boldsymbol{q}_n\right)} = \frac{c_1^{\,2}\lambda_1 + ... + c_n^{\,2}\lambda_n}{c_1^{\,2} + ... + c_n^{\,2}}$$

最大比值$\lambda_{\max}$ 是当所有的 c's 是零，除了乘$\lambda_{\max}$ 的那个 c。

注解 1 方程式(4)的比值是对称矩阵 $A^{\mathrm{T}}A$ 的莱利商(Rayleigh quotient)，它的极大值是最大固有值$\lambda_{\max}(A^{\mathrm{T}}A)$，极小值比值是$\lambda_{\min}(A^{\mathrm{T}}A)$。如果你把任意 $\boldsymbol{x}$ 代入莱利商 $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} / \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$，你会确保得到一个介于$\lambda_{\min}(A^{\mathrm{T}}A)$与$\lambda_{\max}(A^{\mathrm{T}}A)$之间的数。

请尊重版权与译者的劳动成果，侵权必究!

注解 2 $\|A\|$等于 A 的最大奇异值σ_{max}，这些奇异值$\sigma_1, ..., \sigma_r$ 是 A^TA 的正数固有值的平方根，当然会有$\sigma_{max} = (\lambda_{max})^{1/2}$。由于 $A = U\Sigma V^T$ 中的 U 与 V 是正交，$\|A\|$ 等于σ_{max}。

A 的条件数

段落 9.1 展示了舍去误差可以很严重，有些系统很敏感，有些不敏感，对于误差的敏感度可以用**条件数**(condition number)来量测。这是本书特意提出误差的第一章，我们希望估测他们改变 $\boldsymbol{x}$ 多少。

原始方程式是 $A\boldsymbol{x} = \boldsymbol{b}$，假设因为舍去与量测误差造成右侧变成 $\boldsymbol{b} + \Delta\boldsymbol{b}$，解会变成 $\boldsymbol{x} + \Delta\boldsymbol{x}$，我们的目的是估测来自方程式的变化$\Delta\boldsymbol{b}$ 造成的变化$\Delta\boldsymbol{x}$。两式相减给出**误差方程式** $A(\Delta\boldsymbol{x}) = \Delta\boldsymbol{b}$：

$$\text{从 } A(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{b} + \Delta\boldsymbol{b} \text{ 减去 } A\boldsymbol{x} = \boldsymbol{b} \quad \text{得到} \quad A(\Delta\boldsymbol{x}) = \Delta\boldsymbol{b} \tag{5}$$

误差$\Delta\boldsymbol{x} = A^{-1}\Delta\boldsymbol{b}$，当 A^{-1} 很大(A 近乎是奇异)误差就很大。当$\Delta\boldsymbol{b}$ 指向错误的方向时——被 A^{-1} 放大最多的方向，误差$\Delta\boldsymbol{x}$ 会特别大。最差的误差是$\|\Delta\boldsymbol{x}\| = \|A^{-1}\|\,\|\Delta\boldsymbol{b}\|$。

误差界限$\|A^{-1}\|$有一个很严重的缺点，我们让 A 乘 1000，A^{-1} 就会除以 1000，矩阵看起来是 1000 倍的好，但是单纯的调整比例并没有改变问题的本质。$\Delta\boldsymbol{x}$ 的确会除以 1000，但是确切的解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 也会除以 1000，**相对误差**$\|\Delta\boldsymbol{x}\|/\|\boldsymbol{x}\|$仍然没有改变，跟 $\boldsymbol{b}$ 的相对误差比较的应该是 $\boldsymbol{x}$ 的相对误差。

比较相对误差现在产生“条件数” $c = \|A\|\,\|A^{-1}\|$，A 乘以 1000 不会改变这个数，因为 A^{-1} 会除以 1000，条件数 c 维持不变，它可以量测 $A\boldsymbol{x} = \boldsymbol{b}$ 的敏感度。

解的误差小于 $c = \|A\|\,\|A^{-1}\|$ 乘 问题的误差：

$$\textbf{条件数 } c \qquad \frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \le c\,\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} \tag{6}$$

如果问题误差是ΔA (A 的误差而不是 $\boldsymbol{b}$ 的误差)，仍然是 c 控制$\Delta\boldsymbol{x}$：

$$\textbf{A 的误差}\Delta A \qquad \frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x} + \Delta\boldsymbol{x}\|} \le c\,\frac{\|\Delta A\|}{\|A\|} \tag{7}$$

证明 原始方程式是 $\boldsymbol{b} = A\boldsymbol{x}$，误差方程式(5)是$\Delta\boldsymbol{x} = A^{-1}\Delta\boldsymbol{b}$，应用矩阵范数的关键性质$\|A\boldsymbol{x}\| \le \|A\|\,\|\boldsymbol{x}\|$：

$$\|\boldsymbol{b}\| \le \|A\|\,\|\boldsymbol{x}\| \qquad \text{且} \qquad \|\Delta\boldsymbol{x}\| \le \|A^{-1}\|\,\|\Delta\boldsymbol{b}\|$$

两式的左侧相乘得到$\|\boldsymbol{b}\|\,\|\Delta\boldsymbol{x}\|$，右侧相乘得到 $c\|\boldsymbol{x}\|\,\|\Delta\boldsymbol{b}\|$，两侧同时除以$\|\boldsymbol{b}\|\,|\boldsymbol{x}\|$，左侧现在是相对误差$\|\Delta\boldsymbol{x}\|/\|\boldsymbol{x}\|$，右侧现在是方程式(6)的上界限。

当误差是在矩阵，相同的条件数 $c = \|A\|\,\|A^{-1}\|$也会出现，在误差方程式中我们有ΔA 而不是$\Delta\boldsymbol{b}$：

$$\text{从 } (A + \Delta A)(\boldsymbol{x} + \Delta\boldsymbol{x}) = \boldsymbol{b} \text{ 减去 } A\boldsymbol{x} = \boldsymbol{b} \qquad \text{得到} \qquad A(\Delta\boldsymbol{x}) = -(\Delta A)(\boldsymbol{x} + \Delta\boldsymbol{x})$$

s 使用 A^{-1} 乘最后一个方程式，然后取范数得到方程式(7)：

$$\|\Delta\boldsymbol{x}\| \le \|A^{-1}\|\,\|\Delta A\|\,\|\boldsymbol{x} + \Delta\boldsymbol{x}\| \quad \text{或是} \quad \frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x} + \Delta\boldsymbol{x}\|} \le \|A\|\,\|A^{-1}\|\,\frac{\|\Delta A\|}{\|A\|}$$

结论 误差有两个进入方式，他们从误差ΔA 或$\Delta\boldsymbol{b}$ 开始——一个错误的矩阵或一个错误的 $\boldsymbol{b}$。问题的错误被放大(很多或是一些)进入解的误差$\Delta\boldsymbol{x}$，这个与自己相关的误差是有界的，界限是条件数 c。

误差$\Delta\boldsymbol{b}$ 与电脑的舍去以及 $\boldsymbol{b}$ 的原始量测有关，误差ΔA 也与消元法的步骤有关。小枢轴倾向于制造 L 与 U 的大误差，$L + \Delta L$ 乘 $U + \Delta U$ 等于 $A + \Delta A$，当ΔA 或条件数很大时，误差$\Delta\boldsymbol{x}$ 将是无法接受。

范例 3 当 A 是对称，$c = \|A\|\,\|A^{-1}\|$来自固有值：

$$A = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \text{有范数 } 6, \quad A^{-1} = \begin{bmatrix} \frac{1}{6} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \text{有范数} \frac{1}{2}。$$

这个 A 是对称正定，它的范数是$\lambda_{\max} = 6$，A^{-1} 的范数是 $1/\lambda_{\min} = \frac{1}{2}$。计算两个范数的乘积得到条件数$\|A\|\,\|A^{-1}\| = \lambda_{\max}/\lambda_{\min}$：

$$\textbf{正定 } \boldsymbol{A} \textbf{ 的条件数} \qquad c = \lambda_{\max}/\lambda_{\min} = 6/2 = 3$$

范例 4 维持相同的 A，固有值是 6 与 2。要使得 $\boldsymbol{x}$ 小，沿着第一固有向量(1, 0)选择 $\boldsymbol{b}$。要使得$\Delta\boldsymbol{x}$ 大，沿着第二固有向量(0, 1)选择$\Delta\boldsymbol{b}$。$\boldsymbol{x} = \boldsymbol{b}/6$，$\Delta\boldsymbol{x} = \Delta\boldsymbol{b}/2$，比值$\|\Delta\boldsymbol{x}\| / \|\boldsymbol{x}\|$确切是 $c = 3$ 乘比值$\|\Delta\boldsymbol{b}\| / \|\,\boldsymbol{b}\,\|$。

这个展示了条件数$\|A\|\,\|A^{-1}\|$允许的最差误差实际上是会发生的，这里有一个经验法则，高斯消元法实验验证：电脑会损失 $\log c$ 位数的舍去误差。

请尊重版权与译者的劳动成果，侵权必究!

问题集 11.2

1 求出下列正定矩阵的范数$\|A\| = \lambda_{max}$与条件数 $c = \lambda_{max}/\lambda_{min}$：

$$\begin{bmatrix} .5 & 0 \\ 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}$$

2 从$\lambda_{max}(A^TA)$的平方根求出范数与条件数与$\lambda_{min}(A^TA)$，A 没有正定，我们找 A^TA！

$$\begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

3 从$\|A\|$与$\|B\|$的定义(3)，说明这两个不等式：

$$\|AB\boldsymbol{x}\| \le \|A\| \, \|B\boldsymbol{x}\| \le \|A\| \, \|B\| \, \|\boldsymbol{x}\|$$

从$\|AB\boldsymbol{x}\|$到$\|\boldsymbol{x}\|$，推导$\|AB\| \le \|A\| \, \|B\|$，这是使用矩阵范数的关键，$A^n$ 的范数不会大于$\|A\|^n$。

4 使用$\|AA^{-1}\| \le \|A\| \, \|A^{-1}\|$证明条件数最小是 1.

5 为什么 I 是唯一具有$\lambda_{max} = \lambda_{min} = 1$ 的对称正定矩阵？其他具有$\|A\| = 1$ 与$\|A^{-1}\| = 1$ 的矩阵必须有 $A^TA = I$。这些是_____矩阵：完美条件。

6 正交矩阵有范数$\|Q\| = 1$。若 $A = QR$ 证明$\|A\| \le \|R\|$同时$\|R\| \le \|A\|$，则$\|A\| = \|Q\| \, \|R\|$。找出一个例子 $A = LU$，其中$\|A\| < \|L\| \, \|U\|$。

7 (a) 哪个著名的不等式给出对于任意 $\boldsymbol{x}$，$\|(A + B)\boldsymbol{x}\| \le \|A\boldsymbol{x}\| + \|B\boldsymbol{x}\|$？

(b) 为什么矩阵范数的定义(3)得到$\|A + B\| \le \|A\| + \|B\|$？

8 证明若λ是 A 的任意固有值，则$|\lambda| \le \|A\|$，从 $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 开始。

9 “光谱(spectral)半径”$\rho(A) = |\lambda_{max}|$是固有值的最大绝对值，展示一个 2×2 矩阵使得$\rho(A + B) \le \rho(A) + \rho(B)$以及$\rho(AB) \le \rho(A) \, \rho(B)$都是错误，光谱半径不能作为一个范数。

10 (a) 说明为什么 A 与 A^{-1} 有相同的条件数。

(b) 说明为什么 A 与 A^T 有相同的范数，基于$\lambda(A^TA)$与$\lambda(AA^T)$。

11 估测不良条件矩阵 $A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$的条件数。

12 为什么 A 的行列式不能很好的作为范数？为什么不能很好的作为条件数？

13 给定下列矩阵计算 $\boldsymbol{b}-A\boldsymbol{y}$ 与 $\boldsymbol{b}-A\boldsymbol{z}$：

$$\boldsymbol{b}=\begin{bmatrix}.217\\.254\end{bmatrix}\quad A=\begin{bmatrix}.780 & .563\\.913 & .659\end{bmatrix}\quad \boldsymbol{y}=\begin{bmatrix}.341\\-.087\end{bmatrix}\quad \boldsymbol{z}=\begin{bmatrix}.999\\-1.0\end{bmatrix}$$

$\boldsymbol{y}$ 与 $\boldsymbol{z}$ 哪个比较接近 $A\boldsymbol{x}=\boldsymbol{b}$ 的解？两个方式回答问题：比较残余量 $\boldsymbol{b}-A\boldsymbol{y}$ 与 $\boldsymbol{b}-A\boldsymbol{z}$，然后比较 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 与真实的 $\boldsymbol{x}=(1,-1)$。两个答案都是正确的，有时候我们希望小的残余量，有时候希望小的$\Delta\boldsymbol{x}$。

14 (a) 计算问题 13 的 A 的行列式，计算 A^{-1}。

(b) 可能的话计算$\|A\|$与$\|A^{-1}\|$，证明 $c>10^6$。

问题 15-19 有关与平常$\|\boldsymbol{x}\|=\sqrt{\boldsymbol{x}\cdot\boldsymbol{x}}$不同的向量范数。

15 $\boldsymbol{x}=(x_1,\ldots,x_n)$的“$l^1$ 范数”以及“l^∞范数”是

$$\|\boldsymbol{x}\|_1=|x_1|+\ldots+|x_n|\quad 与\quad \|\boldsymbol{x}\|_\infty=\max_{1\le i\le n}|x_i|$$

计算 $\mathbf{R}^5$ 的两个向量的 $\|\boldsymbol{x}\|$ 与 $\|\boldsymbol{x}\|_1$ 与 $\|\boldsymbol{x}\|_\infty$：

$$\boldsymbol{x}=(1,1,1,1,1)\qquad \boldsymbol{x}=(.1,.7,.3,.4,.5)$$

16 证明 $\|\boldsymbol{x}\|_\infty\le\|\boldsymbol{x}\|\le\|\boldsymbol{x}\|_1$ ，从苏瓦兹不等式证明这些比值$\|\boldsymbol{x}\|/\|\boldsymbol{x}\|_\infty$与$\|\boldsymbol{x}\|_1/\|\boldsymbol{x}\|$不会大于$\sqrt{n}$。哪个向量$(x_1,\ldots,x_n)$使得比值大于$\sqrt{n}$？

17 所有的向量范数必须满足三角不等式，证明

$$\|\boldsymbol{x}+\boldsymbol{y}\|_\infty\le\|\boldsymbol{x}\|_\infty+\|\boldsymbol{y}\|_\infty\quad 与\quad \|\boldsymbol{x}+\boldsymbol{y}\|_1\le\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_1$$

18 向量范数也必须满足$\|c\boldsymbol{x}\|=|c|\,\|\boldsymbol{x}\|$，除了 $\boldsymbol{x}=\mathbf{0}$ 时范数必须是正数。下列哪些是 $\mathbf{R}^2$ 向量(x_1,x_2)的范数？

$\|\boldsymbol{x}\|_A\le|x_1|+2|x_2|$　　$\|\boldsymbol{x}\|_B=\min(x_1,x_2)$

$\|\boldsymbol{x}\|_C\le\|\boldsymbol{x}\|+\|\boldsymbol{x}\|_\infty$　　$\|\boldsymbol{x}\|_D=\|A\boldsymbol{x}\|$　(答案与 A 有关)

挑战问题

19 证明 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\le\|\boldsymbol{x}\|_1\,\|\boldsymbol{y}\|_\infty$，借由选择分量 $y_i=\pm1$ 使得 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ 尽可能的大。

20 $-1,2,-1$ 差分矩阵 K 的固有值是$\lambda=2-2\cos(j\pi/(n+1))$，当 n 增加时，估测$\lambda_{\min}$与$\lambda_{\max}$与 $c=\mathbf{cond}(K)=\lambda_{\max}/\lambda_{\min}$：$c\approx Cn^2$，其中常数 C 为何？

当 $n=10,100,1000$ 时，测试估测值 $\mathbf{eig}(K)$与 $\mathbf{cond}(K)$。

请尊重版权与译者的劳动成果，侵权必究!

11.3 迭代法与预条件子

直到现在，我们都是直接求解 $A\boldsymbol{x}=\boldsymbol{b}$，我们接受自来的 A，我们使用消元法与行交换处理它。我们现在要检视**迭代(iterative)法，这个方法把比较简单的矩阵 $\boldsymbol{S}$ 取代 $\boldsymbol{A}$**，差分 $T=S-A$ 移到方程式的右侧。利用 S 取代 A，问题变得容易求解。但是要有代价——简单的系统要重复求解。

迭代方法很容易发现，只要把 A(仔细地)分成 $S-T$。

$$\textbf{改写}\quad A\boldsymbol{x}=\boldsymbol{b} \qquad\qquad S\boldsymbol{x}=T\boldsymbol{x}+\boldsymbol{b} \tag{1}$$

新奇的地方在于迭代求解(1)，每个猜测的 $\boldsymbol{x}_k$ 得到下一个 $\boldsymbol{x}_{k+1}$：

$$\textbf{单纯迭代}\qquad S\boldsymbol{x}_{k+1}=T\boldsymbol{x}_k+\boldsymbol{b} \tag{2}$$

从任意 $\boldsymbol{x}_0$ 开始，求解 $S\boldsymbol{x}_1=T\boldsymbol{x}_0+\boldsymbol{b}$，继续 $S\boldsymbol{x}_2=T\boldsymbol{x}_1+\boldsymbol{b}$，100 次迭代很正常——通常要更多。当(或是如果) $\boldsymbol{x}_{k+1}$ 足够接近 $\boldsymbol{x}_k$ 时——或是残余量 $\boldsymbol{r}_k=\boldsymbol{b}-A\boldsymbol{x}_k$ 接近零，停止迭代程序。我们希望消元法更快得到接近的真实解，当 $\boldsymbol{x}_k$ 收敛，他们的极限 $\boldsymbol{x}_\infty$ 确切是方程式(1)的解：$S\boldsymbol{x}_\infty=T\boldsymbol{x}_\infty+\boldsymbol{b}$ 表示 $A\boldsymbol{x}_\infty=\boldsymbol{b}$。

分割 $A=S-T$ 的两个目的是**加速每个步骤以及快速收敛**。每个步骤的速度与 S 有关而且收敛速度与 $S^{-1}T$ 有关：

1 方程式(2)应该很容易求解 $\boldsymbol{x}_{k+1}$，“**预条件子(preconditioner)**” S 应该是 A 的对角或是三角部分。一个快速方法是 $S=L_0U_0$，其中这些因子与确切的 $A=LU$ 相比多了很多零。这是“不完全 LU”。

2 差 $\boldsymbol{x}-\boldsymbol{x}_k$ (这是误差 $\boldsymbol{e}_k$)应该快速到零，从方程式(1)减去方程式(2)消去 $\boldsymbol{b}$，剩下**误差 $\boldsymbol{e}_k$ 的方程式**：

$$\textbf{误差方程式}\qquad S\boldsymbol{e}_{k+1}=T\boldsymbol{e}_k \qquad \textbf{意思是} \qquad \boldsymbol{e}_{k+1}=S^{-1}T\boldsymbol{e}_k \tag{3}$$

在每一步 $S^{-1}T$ 乘上误差，若 $S^{-1}T$ 够小，它的次方很快到零，但是什么是**够小**？

极端的分割是 $S=A$ 与 $T=0$，第一步的迭代就是原始的 $A\boldsymbol{x}=\boldsymbol{b}$，收敛是完美的而且 $S^{-1}T$ 是零。但是这一步的成本是我们想要避免的，S 的选择是在每一步的速度(简单的 S)与快速收敛(S 接近 A)之间的战斗。这里是一些 S 的选择法：

J $S=A$ 的对角部分(这个迭代称为**雅克比**方法)

GS $S=A$ 的包含对角的下三角(**高斯-赛德尔**方法)

ILU $S=$ 近似 L 乘近似 U (**不完全 LU** 方法)

我们第一个问题是纯粹的线性代数：**什么时候 x_k's 收敛到 x**？答案会揭露数字 $|\lambda|_{\max}$ 控制着收敛，在雅克比与高斯-赛德尔的例子中，我们会计算“光谱半径”$|\lambda|_{\max}$，它是**迭代矩阵** $B = S^{-1}T$ 的最大固有值。

光谱半径$\rho(B)$控制收敛

方程式(3)是 $\boldsymbol{e}_{k+1} = S^{-1}T\boldsymbol{e}_k$，每一个迭代步骤都用相同的矩阵 $B = S^{-1}T$ 乘上误差，在 k 步之后误差 $\boldsymbol{e}_k = B^k\boldsymbol{e}_0$，如果 $B = S^{-1}T$ 的次方趋近零，则误差趋近零。看看 B 的固有值——特别是最大固有值——如何控制矩阵次方 B^k 是美丽的。

> 次方 B^k 趋近零若且唯若 B 的每个固有值都有 $|\lambda| < 1$。
>
> **B 的光谱半径控制收敛速率：** $\rho = \max|\lambda(B)|$

收敛的测试是 $|\lambda|_{\max} < 1$，实数固有值必须在 -1 与 1 之间，复数固有值 $\lambda = a + ib$ 必须有 $|\lambda|^2 = a_2 + b^2 < 1$。光谱半径“$\rho$ (读成 ro)”从零到 $B = S^{-1}T$ 的固有值的最大距离，这是 $\rho = |\lambda|_{\max}$。

要了解为什么 $|\lambda|_{\max} < 1$ 是必要条件，假设初始误差 $\boldsymbol{e}_0$ 刚好是 B 的一个固有向量，一步之后误差是 $B\boldsymbol{e}_0 = \lambda\boldsymbol{e}_0$，$k$ 步之后是 $B^k\boldsymbol{e}_0 = \lambda^k\boldsymbol{e}_0$。如果我们从固有向量开始，我们维持在这个固有向量——当 $|\lambda| < 1$ 时，因子 λ^k 才会到零，这个条件适用每一个固有向量。

要了解为什么 $|\lambda|_{\max} < 1$ 是误差趋近零的充分条件，假设 $\boldsymbol{e}_0$ 是固有向量的组合：

$$\boldsymbol{e}_0 = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n \quad \text{得到} \quad \boldsymbol{e}_k = c_1(\lambda_1)^k\boldsymbol{x}_1 + \ldots + c_n(\lambda_n)^k\boldsymbol{x}_n \tag{4}$$

这是固有向量的特点！当我们用 B 去乘，相当于用 λ_i 去乘每个固有向量 $\boldsymbol{x}_i$。如果所有的 $|\lambda_i| < 1$，则方程式(4)确保 $\boldsymbol{e}_k$ 走到零。

范例 1 $B = \begin{bmatrix} .6 & .5 \\ .6 & .5 \end{bmatrix}$ 有 $\lambda_{\max} = 1.1$ $\qquad B' = \begin{bmatrix} .6 & 1.1 \\ 0 & .5 \end{bmatrix}$ 有 $\lambda_{\max} = .6$。

B^2 是 1.1 乘 B，B^3 是 $(1.1)^2$ 乘 B，B 的次方会爆炸。B' 的次方刚好相反，矩阵 $(B')^k$ 在对角线有 $(.6)^k$ 与 $(.5)^k$，非对角线的单元也跟 $\rho^k = (.6)^k$ 有关，这些设定收敛速度。

注解 当固有向量不足，方程式(4)就不正确。当我们丢失固有向量回到乔丹形式而且 B 不能对角化：

$$\textbf{乔丹形式 } \boldsymbol{J} \qquad B = MJM^{-1} \quad \text{与} \quad B = MJ^kM^{-1} \tag{5}$$

段落 8.3 显示 J 与 J^k 由一个重复固有值的“**方块**”组成：

$$J \text{ 的 } 2\times2 \text{ 方块的次方是 } \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{bmatrix}$$

请尊重版权与译者的劳动成果，侵权必究！

如果$|\lambda|<1$则这些次方趋近零，来自双重固有值的额外因子k被衰减因子λ^{k-1}压制，这个适用每一个方块：

是否可对角化：　收敛 $B^k \to 0$ 而且它的速度依赖$\rho = |\lambda|_{\max} < 1$

雅克比对比高斯-赛德尔

我们现在分割 A 来求解一个特定的 2×2 问题，注意数字$|\lambda|_{\max}$。

$$A\boldsymbol{x} = \boldsymbol{b} \qquad \begin{matrix} 2u - v = 4 \\ -u + 2v = -2 \end{matrix} \quad \text{的解是} \quad \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \tag{6}$$

第一个分割是**雅克比方法**，维持 A 的对角线(这是 S)在左侧，把 A 的非对角线部分移到右侧(这是 T)，开始迭代：

雅克比方法　$S\boldsymbol{x}_{k+1} = T\boldsymbol{x}_k + \boldsymbol{b}$　$\begin{matrix} 2u_{k+1} = v_k + 4 \\ 2v_{k+1} = u_k - 2 \end{matrix}$

从 $u_0 = v_0 = 0$ 开始，第一步得到 $u_1 = 2$ 与 $v_1 = -1$，继续下去：

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 2 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 2 \\ -1/4 \end{bmatrix} \quad \begin{bmatrix} 15/8 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 2 \\ -1/16 \end{bmatrix} \quad \text{趋近} \quad \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

这个显示了收敛。在第 1, 3, 5 步，第二分量是–1, –1/4, –1/16，每两步下降 4 倍。误差方程式是 $S\boldsymbol{e}_{k+1} = T\boldsymbol{e}_k$ ：

误差方程式　$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \boldsymbol{e}_{k+1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{e}_k$　或是　$\boldsymbol{e}_{k+1} = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \boldsymbol{e}_k$ (7)

最后的矩阵 $S^{-1}T$ 有固有值½与–½，所以他的光谱半径是$\rho(B) = ½$：

$$B = S^{-1}T = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \quad \text{有} |\lambda_{\max}| = ½ \quad \text{且} \quad \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}^2 = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

在这个特殊例子的两个步骤确实得到误差乘¼。重要的信息是：当 A 的主对角线比非对角线部分来的大，雅克比方法可以很好的工作。对角线部分是 S，剩下的是$-T$，我们希望对角线部分支配。

固有值$\lambda = ½$是不平常的小，迭代 10 次降低误差 $2^{10} = 1024$。更典型更昂贵的是$|\lambda|_{\max} = .99$ 或.999。

高斯-赛德尔方法把 A 的整个下三角部分留给 S：

高斯-赛德尔　$\begin{matrix} 2u_{k+1} & = v_k + 4 \\ -u_{k+1} + 2v_{k+1} & = -2 \end{matrix}$　或是　$\begin{matrix} u_{k+1} = \frac{1}{2}v_k + 2 \\ v_{k+1} = \frac{1}{2}u_{k+1} - 1 \end{matrix}$ (8)

注意变化。来自第一个方程式的新 u_{k+1} 马上用在第二个方程式，如果是雅克比，我们保存旧的 u_k 直到整个步骤完成。现在是高斯-赛德尔，新的值立刻加入而且旧的 u_k 被摧毁，这个减少一半的储存，它也会加速迭代(通常)，而且它的成本不会比雅克比方法来的高。

从另一个初始值 $u_0 = 0$ 与 $v_0 = -1$ 开始测试：

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \begin{bmatrix} 3/2 \\ -1/4 \end{bmatrix} \quad \begin{bmatrix} 15/8 \\ -1/16 \end{bmatrix} \quad \begin{bmatrix} 63/32 \\ -1/64 \end{bmatrix} \quad \text{趋近} \quad \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

在第一分量的误差是 2, 1/2, 1/8, 1/32，在第二分量的误差是–1, –1/4, –1/16, –1/32，我们一个步骤就除以 4，而不是两个步骤，**高斯-赛德尔是雅克比的两倍快**。当 A 是正定三对角，我们有 $\rho_{GS} = (\rho_J)^2$：

$$S = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix} \quad \text{与} \quad T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{而且} \quad S^{-1}T = \begin{bmatrix} \mathbf{0} & \frac{1}{2} \\ 0 & \frac{\mathbf{1}}{\mathbf{4}} \end{bmatrix}$$

高斯-赛德尔的固有值是 0 与¼，比较雅克比的½与–½。

往前推进一下，我们可以描述**连续超松弛(overrelaxation)方法(SOR)**。这个新概念引入一个参数 ω (omega)进入迭代，选择数字 ω 使得 $S^{-1}T$ 的光谱半径尽可能小。

改写 $A\boldsymbol{x} = \boldsymbol{b}$ 成为 $\omega A\boldsymbol{x} = \omega\boldsymbol{b}$，**SOR** 中的矩阵 S 有原始 A 的对角线，但是对角线下方我们使用 ωA，在右侧的 T 是 $S - \omega A$：

$$\textbf{SOR} \qquad \begin{aligned} 2u_{k+1} &= (2-2\omega)u_k + \omega v_k + 4\omega \\ -\omega u_{k+1} + 2v_{k+1} &= \omega u_k + (2-2\omega)v_k - 2\omega \end{aligned} \tag{9}$$

这个看起来更复杂，但是电脑跑起来一如既往的快。**SOR** 很像高斯-赛德尔，有一个可以调整的数字 ω，最佳 ω 得到快速的迭代。

我会把阶数 n 的最有价值测试矩阵记录在案，它是我们的最爱–1, 2, –1 的三对角矩阵 K。对角线是 $2I$，下方与上方是–1's，我们的例子有 $n = 2$，得到 $\cos \pi/3 = ½$，这是上述的雅克比固有值。特别注意这个 $|\lambda|_{max}$ 是高斯-赛德尔的平方：

–1, 2, –1 阶数 n 的矩阵 K 的分割得到这些 B 的固有值：

雅克比(S = 0, 2, 0 矩阵)：　$S^{-1}T$ 有 $|\lambda|_{max} = \cos\dfrac{\pi}{n+1}$

高斯-赛德尔(S = –1, 2, 0 矩阵)：　$S^{-1}T$ 有 $|\lambda|_{max} = \left(\cos\dfrac{\pi}{n+1}\right)^2$

SOR (具有最佳 ω)：　$S^{-1}T$ 有 $|\lambda|_{max} = \left(\cos\dfrac{\pi}{n+1}\right)^2 \Big/ \left(1+\sin\dfrac{\pi}{n+1}\right)^2$

让我说明清楚：你不应该针对–1, 2, –1 矩阵使用任何迭代法！三对角矩阵使用消元法已经非常快速(确切的 LU)。迭代法是针对大型的稀疏矩阵，其中矩阵有远离中心对角线的零，这些零在确切的 L 与 U 制造了更多的非零，这项“**填满**”就是为什么消元法昂贵的原因。

请尊重版权与译者的劳动成果，侵权必究!

我们多提一个分割，“**不完全** LU”的概念是把 L 与 U 中小的非零行恢复到零，这样留下的三角矩阵 L_0 与 U_0 再次是稀疏矩阵。这个分割有 $S = L_0U_0$ 在左侧。每个步骤很快速：

不完全 LU　　$L_0U_0\boldsymbol{x}_{k+1} = (L_0U_0 - A)\boldsymbol{x}_k + \boldsymbol{b}$

在右侧我们执行稀疏的矩阵-向量乘法，不用 L_0 乘 U_0，这些是矩阵。用 U_0 乘 $\boldsymbol{x}_k$，再用 L_0 乘那个向量，在左侧我们执行前向与反向代入法。如果 L_0U_0 接近 A，则 $|\lambda|_{\max}$ 很小，几个步骤过后会得到接近的答案。

多格点与共轭梯度

我不能留下雅克比与高斯-赛德尔是优秀方法的印象，通常误差的“低频率”部分衰减非常慢，需要很多次迭代。这里是两个得到重大改善的重要概念，**多格点(Multigrid)**可以在 $O(n)$ 步内解决阶数 n 的问题，具有一个优秀的**预条件子**，**共轭梯度(conjugate gradient)**变成数值线性代数中最流行也最有效的演绎法之一。

多格点　求解小问题使用粗格点，每次迭代便宜又快速，然后在粗格点值之间内插，得到全幅大小问题的快速开头。多格点可能走 4 阶层的下与回。

共轭梯度　如同 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - A\boldsymbol{x}_k + \boldsymbol{b}$ 的正常迭代，会在每个阶段牵涉到用 A 乘，如果 A 是稀疏，这就不贵：$A\boldsymbol{x}_k$ 是我们想要做的，它多加了一个基底向量至“**克雷洛夫空间**”，这个空间包含我们的近似值。但是 $\boldsymbol{x}_{k+1}$ 不是 $\boldsymbol{x}_0, A\boldsymbol{x}_0, ..., A^k\boldsymbol{x}_0$ 的**最佳组合**，正常的迭代有效但是离最佳还很远。

共轭梯度在每个步骤选择**最佳组合** $\boldsymbol{x}_k$，额外的成本(超出一个 A 的乘法)不多，我们会给出 CG 迭代，强调这个方法是为了对称正定矩阵而创建。当 A 不是对称，一个好选择是 GMRES，当 $A = A^{\mathrm{T}}$ 不是正定时，有 MINRES。环绕制造每个连续 $\boldsymbol{x}_k$ 的最佳选择的概念，已经建立了一个高能迭代方法的世界。

我的教科书“计算科学与工程”详细描述了多格点与 CG，在数值线性代数的书本中，Trefethen-Bau 的书值得流行(其他也很棒)，Golub-Van 是一个提高。

问题集重建了从 $\boldsymbol{x}_{k-1}$ 到 $\boldsymbol{x}_k$ 的每个共轭梯度循环的 5 个步骤，我们计算新的近似 $\boldsymbol{x}_k$，新的残余量 $\boldsymbol{r}_k = \boldsymbol{b} - A\boldsymbol{x}_k$，与新的搜寻方向 $\boldsymbol{d}_k$ 寻找下个 $\boldsymbol{x}_{k+1}$。

我写下这些针对原始矩阵 A 的步骤，但是**预条件子** S 可以使得收敛更快。我们的原始方程式是 $A\boldsymbol{x} = \boldsymbol{b}$，预条件方程式是 $S^{-1}A\boldsymbol{x} = S^{-1}\boldsymbol{b}$，代码的小变化得出预条件共轭梯度法——求解正定系统领先的迭代法。

最大的竞争者是直接消元法，把方程式重新排序取得 A 中的零的最大优势，要超过高斯还是不容易的。

固有值的迭代法

我们从 $A\boldsymbol{x} = \boldsymbol{b}$ 回到 $A\boldsymbol{x} = \lambda\boldsymbol{x}$，迭代法是线性方程式的一种选择，他们是固有值问题的必需品。一个 $n \times n$ 矩阵的固有值是一个 n 次多项式的根，$A - \lambda I$ 的行列式从 $(-\lambda)^n$ 开始，本书不能留下固有值只能这样计算的印象！从 $\det(A - \lambda I) = 0$ 开始处理是一个很差劲的方法---除了 n 很小。

当 $n > 4$ 就没有公式求解 $\det(A - \lambda I) = 0$，更糟糕的是 λ's 可能非常不稳定与敏感。从 A 本身开始处理会更好，逐渐的变成对角或是三角。(则固有值出现在对角线。) LAPACK 函数库有很好的电脑代码---个别的副程序在 www.netlib.org/lapack 是免费的，这些函数库组合了较早的 LINPACK 与 EISPACK，具有很多改善(在层次 3 BLAS 使用矩阵-矩阵运算)，它是一个在高性能电脑上针对线性代数的 Fortran 77 程序的集合。对于你的或我的电脑，我们需要的是高品质矩阵包装。对于具有平行处理的超级电脑，要用到 ScaLAPACK 与方块消元法。

我们会简单讨论针对计算固有值的**次方(power)**法与 QR 法(LAPACK 选择的)，给出详细代码就没有意义了。

1　次方法与逆次方法。 从任意向量 $\boldsymbol{u}_0$ 开始，用 A 乘得到 $\boldsymbol{u}_1$，再次用 A 乘得到 $\boldsymbol{u}_2$。如果 $\boldsymbol{u}_0$ 是固有向量的组合，会有 A 乘每个固有向量等效于 λ_i 乘 $\boldsymbol{x}_i$，k 个步骤之后我们有 $(\lambda_i)^k$：

$$\boldsymbol{u}_k = A^k\boldsymbol{u}_0 = c_1(\lambda_1)^k\boldsymbol{x}_1 + \ldots + c_n(\lambda_n)^k\boldsymbol{x}_n \tag{10}$$

当次方法继续时，**最大的固有值开始支配**，向量 $\boldsymbol{u}_k$ 指向支配固有向量 $\boldsymbol{x}_1$。从马可夫矩阵我们看到这个：

$$A = \begin{bmatrix} .9 & .3 \\ .1 & .7 \end{bmatrix} \text{ 有 } \lambda_{\max} = 1\text{，对应固有向量 } \begin{bmatrix} .75 \\ .25 \end{bmatrix}$$

从 $\boldsymbol{u}_0$ 开始，每个步骤用 A 乘：

$$\boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \boldsymbol{u}_1 = \begin{bmatrix} .9 \\ .1 \end{bmatrix} \qquad \boldsymbol{u}_2 = \begin{bmatrix} .84 \\ .16 \end{bmatrix} \quad \text{趋近} \quad \boldsymbol{u}_\infty = \begin{bmatrix} .75 \\ .25 \end{bmatrix}$$

收敛速度与第二大固有值 λ_2 对最大固有值 λ_1 的比值有关，我们不希望 λ_1 太小，我们希望 λ_2/λ_1 很小。这里 $\lambda_2 = 0.6$ 与 $\lambda_1 = 1$，则次方法就太慢了。

存在一个找出最小固有值的方法吗? ---这是应用上最重要的。是的，使用**逆次方法**：用 A^{-1} 乘 $\boldsymbol{u}_0$ 而不用 A。我们从来都不想计算 A^{-1}，我们实际是求解 $A\boldsymbol{u}_1 = \boldsymbol{u}_0$。借由节省的 LU 因子，下个步骤 $A\boldsymbol{u}_2 = \boldsymbol{u}_1$ 很快速，步骤 k 会有 $A\boldsymbol{u}_k = \boldsymbol{u}_{k-1}$：

请尊重版权与译者的劳动成果，侵权必究!

逆次方法 $$\boldsymbol{u}_k = A^{-k}\boldsymbol{u}_0 = \frac{c_1\boldsymbol{x}_1}{(\lambda_1)^k} + \ldots + \frac{c_n\boldsymbol{x}_n}{(\lambda_n)^k} \tag{10}$$

现在最小的固有值$\lambda_{\min}$在掌握之中，当它很小时，因子 $1/\lambda_{\min}{}^k$ 就很大。为了高速度，我们借由平移矩阵到 $A - \lambda^* I$ 使得$\lambda_{\min}$更小。

这项平移不会改变固有向量，(λ^*可能来自 A 的对角线，如果来自莱利商 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 会更好。) 如果λ^*接近$\lambda_{\min}$，则$(A - \lambda^* I)^{-1}$会有非常大的固有值$(\lambda_{\min} - \lambda^*)^{-1}$。**每个平移逆反次方步骤**使用大的数乘固有向量，使得固有向量很快就会支配。

2　*QR* 法　这是一个数值线性代数的重要成就，六十年前，固有值计算很慢而且不精确，我们甚至不知道求解 $\det(A - \lambda I) = 0$ 是糟糕的方法。雅克比较早曾经建议 A 应该逐渐变成三角形--然后固有值自动出现在对角线，他使用 2×2 旋转产生非对角线的零。(不幸的是较早的零又变成非零，但是雅克比方法使得平行电脑恢复一部分流行。) ***QR* 方法**目前是固有值计算的领导者。

基本的步骤是分解 A 成为 QR，其中 A 的固有值是我们想要的。回忆格莱姆-施密特(段落 4.4)，Q 有正交单位列与 R 是三角形。固有值的关键概念是：**反转 *Q* 与 *R***。新的矩阵(相同λ's)是 $A_1 = RQ$，因为 $A = QR$ 与 $A_1 = Q^{-1}AQ$ 相似，所以 RQ 的固有值没有改变：

$$\boldsymbol{A_1 = RQ}\text{ 有相同的}\boldsymbol{\lambda} \quad QR\boldsymbol{x} = \lambda\boldsymbol{x} \quad \text{得到} \quad RQ(Q^{-1}\boldsymbol{x}) = \lambda(Q^{-1}\boldsymbol{x}) \tag{12}$$

继续这个程序。把新矩阵 A_1 分解成 Q_1R_1，然后反转因子变成 R_1Q_1，这是相似矩阵 A_2 而且再次没有改变固有值。惊人的是这些固有值开始出现在对角线，很快的 A_4 的最后单元得到一个精确的固有值。这个情况下，我们移除最后一行与最后一列得到一个较小的矩阵，然后继续寻找下一个固有值。

两个额外的概念使得这个方法成功，一个是在分解成 QR 之前，把矩阵平移 I 的倍数，然后 RQ 再平移回去得到 A_{k+1}：

把 $A_k - c_k I$ 分解成 Q_kR_k，下一个矩阵是 $A_{k+1} = R_kQ_k + c_k I$

A_{k+1} 与 A_k 有相同的固有值，而且与原始的 $A_0 = A$ 有相同的固有值。一个优质的平移是选择 c 靠近一个(未知的)固有值，这个更精确的固有值出现在 A_{k+1} 对角线--这个告诉我们下一个到 A_{k+2} 的步骤应该选择的较好 c。

第二个概念是在 QR 方法开始之前先获得非对角线的零，一个消元步骤 E 会做这样的事情，或是一个吉文斯旋转，但是不要忘了 E^{-1}(否则λ会改变)：

$$EAE^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ \mathbf{1} & \mathbf{4} & 5 \\ \mathbf{1} & \mathbf{6} & 7 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 3 \\ \mathbf{1} & 9 & 5 \\ \mathbf{0} & \mathbf{4} & 2 \end{bmatrix} \quad \textbf{相同的}\boldsymbol{\lambda}\textbf{'s}$$

我们必须在**次对角线**留下非零数 1 与 4，更多的 E's 可以移除他们，但是 E^{-1} 会再次填满，这是一个**海森堡矩阵**(一个非零的)次对角线。

在 QR 方法过程中，左下角的的零仍然维持是零。每个 QR 分解的运算次数从 $O(n^3)$ 降到 $O(n^2)$。

Golub 与 Van Logan 给出一个在海森堡矩阵的 QR 一步平移的范例，这个平移是 $7I$，把 A 的所有对角线单元提出 7(然后平移回去 A_1)：

$$A=\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 0 & 0.001 & 7\end{bmatrix} \text{ 产生 } A_1=\begin{bmatrix}-.54 & 1.69 & 0.835\\ .31 & 6.53 & -6.656\\ 0 & 0.00002 & 7.012\end{bmatrix}$$

把 $A-7I$ 分解成 QR 产生 $A_1=RQ+7I$。注意这个非常小的数 0.00002，对角单元 7.012 几乎是 A_1 的确切固有值，所以也是 A 的固有值。另一个在 A_1 的平移 $7.012I$ 的 QR 步骤会给出极佳的精确度。

对于大型稀疏矩阵的固有值，我会参考 ARPACK。问题 25-27 描述基底正交化的 Arnoldi 迭代法--当 A 是对称时每个步骤只有 3 项，矩阵变成三对角：计算固有值的奇妙开始。

问题集 11.3

问题 1-12 有关 $Ax=b$ 的迭代法。

1 把 $A\boldsymbol{x}=\boldsymbol{b}$ 变成 $\boldsymbol{x}=(I-A)\boldsymbol{x}+\boldsymbol{b}$，这样分割对应的 S 与 T 为何？什么样的矩阵 $S^{-1}T$ 控制 $\boldsymbol{x}_{k+1}=(I-A)\boldsymbol{x}_k+\boldsymbol{b}$ 的收敛？

2 若λ是 A 的一个固有值，则____是 $B=I-A$ 的固有值。若 A 的实数固有值介于_____与_____之间，B 的实数固有值有绝对值小于 1。

3 说明为什么 $\boldsymbol{x}_{k+1}=(I-A)\boldsymbol{x}_k+\boldsymbol{b}$ 的迭代对于 $A=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}$ 无法收敛？

4 为什么 B^k 的范数不会大于$\|B\|^k$？则$\|B\|<1$ 确保次方 B^k 趋近零(收敛)。由于 $|\lambda|_{\max}$ 小于$\|B\|$，所以前述的事实毫无惊讶。

5 若 A 是奇异，则所有的分割 $A=S-T$ 必定失败。由 $A\boldsymbol{x}=\boldsymbol{0}$ 证明 $S^{-1}T\boldsymbol{x}=\boldsymbol{x}$，所以矩阵 $B=S^{-1}T$ 有$\lambda=1$，因此迭代失败。

6 把 2's 换成 3's，求出雅克比方法对应的 $S^{-1}T$ 的固有值：

$$S\boldsymbol{x}_{k+1}=T\boldsymbol{x}_k+\boldsymbol{b} \text{ 是 } \begin{bmatrix}3 & 0\\ 0 & 3\end{bmatrix}\boldsymbol{x}_{k+1}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\boldsymbol{x}_k+\boldsymbol{b}$$

7 针对问题 6，求出高斯-赛德尔方法对应的 $S^{-1}T$ 的固有值：

$$\begin{bmatrix}3 & 0\\ -1 & 3\end{bmatrix}\boldsymbol{x}_{k+1}=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\boldsymbol{x}_k+\boldsymbol{b}$$

高斯-赛德尔对应的$|\lambda|_{\max}$ 等于雅克比对应的$|\lambda|_{\max}$？

请尊重版权与译者的劳动成果，侵权必究!

8 对于任意 2×2 矩阵$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$，证明对应高斯-赛德尔的$|\lambda|_{\max}$等于$|bc/ad|$，对应雅克比的$|\lambda|_{\max}$等于$|bc/ad|^{1/2}$。我们需要 $ad \neq 0$ 使得矩阵 S 可逆。

9 写出高斯-赛德尔的一个电脑代码(MATLAB 或其他)，你可以从 A 定义 S 与 T，或是直接从单元 a_{ij} 设定迭代回圈。针对阶数 10, 20, 50 的–1, 2, –1 矩阵 A 进行测试，其中 $\boldsymbol{b} = (1, 0, ..., 0)$。

10 在分量 i 的高斯-赛德尔迭代使用较早的 $\boldsymbol{x}^{\text{new}}$ 的部分：

高斯-赛德尔 $$x_i^{\text{new}} = x_i^{\text{old}} + \frac{1}{a_{ii}}\Big(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{\text{new}} - \sum_{j=i}^{n} a_{ij}x_j^{\text{old}}\Big).$$

如果每个 $\boldsymbol{x}_i^{\text{new}} = \boldsymbol{x}_i^{\text{old}}$，然后证明解 $\boldsymbol{x}$ 是正确的？对应雅克比方法时，公式要如何修改？对于 **SOR**，在括号外插入ω。

11 方程式(10)除以λ_1^k，说明为什么$|\lambda_2/\lambda_1|$控制次方法的收敛速度。创建一个矩阵 A 使得这个方法不能收敛。

12 马可夫矩阵 $A = \begin{bmatrix} .9 & .3 \\ .1 & .7 \end{bmatrix}$有$\lambda = 1$ 与 0.6，次方法 $\boldsymbol{u}_k = A^k\boldsymbol{u}_0$ 收敛到$\begin{bmatrix} .75 \\ .25 \end{bmatrix}$，求出 A^{-1} 的固有值。逆次方法 $\boldsymbol{u}_{-k} = A^{-k}\boldsymbol{u}_0$ 收敛到什么数字(在你乘 0.6^k 之后)？

13 对角线–1, 2, –1 的阶数 $n - 1$ 的三对角矩阵有固有值$\lambda_j = 2 - 2\cos(j\pi/n)$，为什么最小的固有值近似于$(j\pi/n)^2$？逆次方法收敛的速度是$\lambda_1/\lambda_2 \approx 1/4$。

14 对于 $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$应用次方法 $\boldsymbol{u}_{k+1} = A\boldsymbol{u}_k$ 三次，其中初始值是 $\boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$，次方法收敛到什么样的固有向量？

15 对于 A = –1, 2, –1 矩阵，应用逆次方法 $\boldsymbol{u}_{k+1} = A^{-1}\boldsymbol{u}_k$ 三次，初始值与前述相同，这些 $\boldsymbol{u}_k$'s 趋近什么样的固有向量？

16 求取固有值的 QR 方法中，当 A 平移使得 $A_{22} = 0$，证明 2, 1 单元从 $A = QR$ 中的 $\sin\theta$ 掉落成 RQ 中的$-\sin^3\theta$。(计算 R 与 RQ。) 这项“立方收敛”使得这个方法成功：

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & 0 \end{bmatrix} = QR = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & ? \\ 0 & ? \end{bmatrix}$$

17 如果 A 是正交矩阵，它的 QR 分解有 Q = ____与 R = ____，因此 QR = ____。当 QR 方法无法使用时，这些是稀有的例子。

18 平移 QR 法把 $A - cI$ 分解成 QR，证明下一个矩阵 $A_1 = RQ + cI$ 等于 $Q^{-1}AQ$，因此 A_1 与 A 有_____的固有值(但是 A_1 接近三角形。)

19 当 $A = A^{\mathrm{T}}$，“Lanczos 方法”求出 a's 与 b's 与正交单位 $\boldsymbol{q}$'s 使得 $A\boldsymbol{q}_j = b_{j-1}\boldsymbol{q}_{j-1} + a_j\boldsymbol{q}_j + b_j\boldsymbol{q}_{j+1}$ (其中 $\boldsymbol{q}_0 = \boldsymbol{0}$)，用 $\boldsymbol{q}_j^{\mathrm{T}}$ 乘前述方程式求出 a_j 的公式。这个方程式显示 $AQ = QT$，其中 T 是一个三对角矩阵。

20 问题 19 的方程式来自下列的回圈，其中 $b_0 = 1$ 与 $\boldsymbol{r}_0$ =任意 $\boldsymbol{q}_1$：

$$\boldsymbol{q}_{j+1} = \boldsymbol{r}_j / b_j\ ;\ \ j = j+1;\ \ a_j = \boldsymbol{q}_j^{\mathrm{T}} A\boldsymbol{q}_j\ ;\ \ \boldsymbol{r}_j = A\boldsymbol{q}_j - b_{j-1}\boldsymbol{q}_{j-1} - a_j\boldsymbol{q}_j\ ;\ \ b_j = \|\boldsymbol{r}_j\|$$

写出一个代码并在–1, 2, –1 矩阵 A 测试，$Q^{\mathrm{T}}Q$ 应该是 I。

21 假设 A 在 QR 方法中是三对角且对称，从 $A_1 = Q^{-1}AQ$ 证明 A_1 对称。写出 $A_1 = RAR^{-1}$，证明 A_1 也是三对角。(如果证明 A_1 的下半部是三对角，依据对称性质，则上半部分也是三对角。) 对称三对角矩阵是 QR 方法的最佳开始方式。

问题 22-25 提出两个基础迭代，每个步骤包含 $A\boldsymbol{q}$ 或 $A\boldsymbol{d}$。

大型矩阵的关键点在于矩阵-向量乘法比矩阵-矩阵乘法快很多。需要的创建从向量 $\boldsymbol{b}$ 开始，重复乘法得到 $A\boldsymbol{b}, A^2\boldsymbol{b}, \ldots$，但是这些向量不是正交。**“Arnoldi 迭代”**针对格莱姆-施密特概念建立的空间，创建一组正交单位基底 $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$：针对前面的 $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{n-1}$，正交化每个新的 $A\boldsymbol{q}_n$，由 $\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{n-1}\boldsymbol{b}$ 生成的“Krylov 空间”具有更好的基底 $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$。

这里的伪代码是数值线性代数中两个最重要的演绎法，Arnoldi 给出优质基底，CG 给出 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 的优质近似。

Arnoldi 迭代法	**针对正定 A 的共轭梯度迭代**	
$\boldsymbol{q}_1 = \boldsymbol{b}/\|\boldsymbol{b}\|$	$\boldsymbol{x}_0 = 0, \boldsymbol{r}_0 = \boldsymbol{b}, \boldsymbol{d}_0 = \boldsymbol{r}_0$	
for $n = 1$ **to** $N-1$	**for** $n = 1$ **to** N	
$\boldsymbol{v} = A\boldsymbol{q}_n$	$\alpha_n = (\boldsymbol{r}_{n-1}^{\mathrm{T}}\boldsymbol{r}_{n-1})/(\boldsymbol{d}_{n-1}^{\mathrm{T}}A\boldsymbol{d}_{n-1})$	步阶长度 $\boldsymbol{x}_{n-1}$ 到 $\boldsymbol{x}_n$
for $j = 1$ **to** n	$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} + \alpha_n\boldsymbol{d}_{n-1}$	近似解
$h_{jn} = \boldsymbol{q}_j^{\mathrm{T}}\boldsymbol{v}$	$\boldsymbol{r}_n = \boldsymbol{r}_{n-1} - \alpha_n A\boldsymbol{d}_{n-1}$	新的残余量 $\boldsymbol{b} - A\boldsymbol{x}_n$
$\boldsymbol{v} = \boldsymbol{v} - h_{jn}\boldsymbol{q}_j$	$\beta_n = (\boldsymbol{r}_n^{\mathrm{T}}\boldsymbol{r}_n)/(\boldsymbol{r}_{n-1}^{\mathrm{T}}\boldsymbol{r}_{n-1})$	改善步阶
$h_{n+1,n} = \|\boldsymbol{v}\|$	$\boldsymbol{d}_n = \boldsymbol{r}_n + \beta_n\boldsymbol{d}_{n-1}$	下个搜寻方向
$\boldsymbol{q}_{n+1} = \boldsymbol{v}/h_{n+1,n}$	% 注意：只有一个矩阵-向量乘法 $A\boldsymbol{q}$ 与 $A\boldsymbol{d}$	

对于共轭梯度，残余量 $\boldsymbol{r}_n$ 是正交，搜寻方向是“A-正交”：所有的 $\boldsymbol{d}_j^{\mathrm{T}}A\boldsymbol{d}_k = 0$。这个迭代法利用极小化误差 $\boldsymbol{e}^{\mathrm{T}}A\boldsymbol{e}$ 求解 $A\boldsymbol{x} = \boldsymbol{b}$，其中极小化是针对 $\boldsymbol{b}, A\boldsymbol{b}, \ldots, A^{n-1}\boldsymbol{b}$ 生成的“Krylov 空间”中的所有向量。这是奇妙的演绎法。

22 对于对角矩阵 A= diag([1　2　3　4])与向量 $\boldsymbol{b} = (1, 1, 1, 1)$，走完一个 Arnoldi 步骤，求出正交单位向量 $\boldsymbol{q}_1$ 与 $\boldsymbol{q}_2$。

请尊重版权与译者的劳动成果，侵权必究!

23 Arnoldi 的方法是找出 Q 使得 $AQ = QH$(一列一列的处理)：

$$AQ = \begin{bmatrix} Aq_1 & \cdots & Aq_N \end{bmatrix} = \begin{bmatrix} q_1 & \cdots & q_N \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \cdot & h_{1N} \\ h_{21} & h_{22} & \cdot & h_{2N} \\ 0 & h_{32} & \cdot & \cdot \\ 0 & 0 & \cdot & h_{NN} \end{bmatrix} = QH$$

H 是一个“海森堡矩阵”，其中有一个非零的次对角线。当 A 是对称，存在一个重要的事实：**海森堡矩阵 $H = Q^{-1}AQ = Q^{\mathrm{T}}AQ$ 是一个对称矩阵，因此它是三对角矩阵**。说明上述这句话。

24 三对角 H(当 A 是对称)给出 **Lanczos 迭代**：

只有三个项 $$\boldsymbol{q}_{j+1} = \left(A\boldsymbol{q}_j - h_{j,\,j}\,\boldsymbol{q}_j - h_{j-1,\,j}\,\boldsymbol{q}_{j-1}\right)/h_{j+1,\,j}$$

从 $H = Q^{-1}AQ$，为什么 H 的固有值与 A 的固有值相同？对于大型矩阵，“Lanczos 方法”借由停在较小的三对角矩阵 H_k 来计算带头固有值，书中的 QR 方法应用在计算 H_k 的固有值。

25 应用共轭梯度(CG)方法求解 $A\boldsymbol{x} = \boldsymbol{b}$ = ones(100, 1)，其中 A 是–1, 2, –1 的第二差分矩阵 A = toeplitz([2 **–1** zeros(1, 98)])。沿着确切解 $\boldsymbol{x}^*$，从 CG 画出 $\boldsymbol{x}_{10}$ 到 $\boldsymbol{x}_{20}$。(它的 100 个分量是 $x_i = (ih - i^2h^2)/2$，其中 $h = 1/101$。“plot(i, x(i))”应该是一个抛物线。) 【原文– 1 之间空格是错误】

26 对于非对称矩阵，光谱半径 $\rho = \max|\lambda_i|$ 不是一个范数。但是在大的 n 时，$\|A^n\|$ 会像 ρ^n 一样的成长或衰减。针对 A = [1 1; 0 1.1]，使用指令 **norm** 比较前述的数字。

$A^n \to 0$ 若且唯若 $\rho < 1$。当 $A = S^{-1}T$，这是收敛的关键。

第十二章
概率与统计的线性代数

12.1 平均值，方差与概率

我们从本章 3 个基本词语开始：平均值，方差与概率，在我开始写任何公式之前那，先做个粗略的意义说明：

平均值是平均数值或是期望数值。

方差σ^2 量测与平均值 m 的距离的平方。

n 个不同事件的**概率**是正数 $p_1, ..., p_n$，他们的总和是 1。

当然，平均值很容易了解，我们从它开始。但是马上我们有两个不同的情况需要你保持清楚，一方面我们可以从完整的试验得到结果(样本值)，另一方面我们可以从未来的试验得到期望的结果(期望值)。让我给出一些例子：

样本值 5 个随机大一新生的年龄 18, 17, 18, 19, 17

样本平均值 (18 + 17 + 18 + 19 + 17)/5 = 17.8

概率 大一新生班级的年龄是 17(20%), 18(50%), 19(30%)

一个随机大一新生有**期望年龄 E[x]**是 (.2)17 + (.5)18 + (.3)19 = 18.1

两个数字 17.8 与 18.1 都是正确的平均。样本平均值从来自一个完整试验的 N 个样本 $x_1, ..., x_N$ 开始，他们的平均值是 N 个观察样本的平均：

$$\textbf{样本平均值} \qquad m = \mu = \frac{1}{N}(x_1 + x_2 + ... + x_N) \tag{1}$$

请尊重版权与译者的劳动成果，侵权必究!

x 的期望值从年龄 $x_1, ..., x_n$ 的概率 $p_1, ..., p_n$ 开始：

$$\textbf{期望值} \qquad m = \mathrm{E}[x] = p_1x_1 + p_2x_2 + ... + p_nx_n \tag{2}$$

这是 $\boldsymbol{p} \cdot \boldsymbol{x}$。注意 $m = \mathrm{E}[x]$告诉我们期望什么，$m = \mu$告诉我们得到什么。

借由取得很多样本(很大的 N)，样本的结果会与概率接近，“**大数法则**”说明当样本大小 N 增加，样本的平均值会**收敛**到它的期望值 $\mathrm{E}[x]$，这个概率是 1。一个公平的铜板有概率 $p_0 = ½$的尾与 $p_1 = ½$的头，则 $\mathrm{E}[x] = (½)0 + (½)1$。在 N 次铜板的投掷出现头的比例是样本平均值，期望会趋近 $\mathrm{E}[x] = ½$。

这个并不是说如果我们看到的尾比头多，下个样本会倾向出现头，机会仍然是 50-50。前 100 或 1000 次投掷确实会影响样本平均值，但是 1000 次投掷不会影响它的极限—因为你要除以 $N \to \infty$。

方差(环绕平均值)

方差σ^2 量测与期望平均值 $\mathrm{E}[x]$的**期望距离**(平方)，**样本方差 S^2** 量测与样本平均值的**实际距离**(平方)，**平方根是标准差σ 或 S**。一次考试之后，我邮寄μ与 S 到班级，我不知道期望平均值与方差，因为我不知道每个分数的概率 p_1 到 p_{100}。(教学 50 年，我仍然不知道期望什么。)

偏差永远是与平均值的偏差—样本或期望，我们在检视环绕平均值 $x = m$“分布”的大小，从 N 个样本开始。

$$\textbf{样本方差} \qquad S^2 = \frac{1}{N-1}\left[(x_1 - m)^2 + ... + (x_N - m)^2 \right] \tag{3}$$

样本年龄 $x = 18, 17, 18, 19, 17$ 有平均值 $m = 17.8$，样本有方差 0.7：

$$S^2 = \frac{1}{4}\left[(.2)^2 + (-.8)^2 + (.2)^2 + (1.2)^2 + (-.8)^2\right] = \frac{1}{4}(2.8) = \mathbf{0.7}$$

当我们计算平方时，负号消失。请注意！统计学家除以 $N - 1 = 4$(不是 $N = 5$)使得 S^2 是σ^2 的一个“无偏差(unbiased)”估测。一个自由度已经用在样本平均值。

一个重要的等式来自分割$(x - m)^2 = x^2 - 2mx + m^2$：

$$(x_i - m)^2 \text{ 的总和} = (x_i^2 \text{ 的总和}) - 2m(x_i \text{ 的总和}) + (m^2 \text{ 的总和})$$

$$= (x_i^2 \text{ 的总和}) - 2m(Nm) + Nm^2$$

$$\boldsymbol{(x_i - m)^2 \text{ 的总和} = (x_i^2 \text{ 的总和}) - Nm^2} \tag{4}$$

这是借由 $x_1^2 + ... + x_N^2$ 的总和求取$(x_1 - m)^2 + ... + (x_N - m)^2$ 的等效方式。

【原文排版错误】

现在从概率 p_i(不会是负数！)而不是从样本开始，我们求出**期望值而不是样本值**。方差σ^2 是统计学的重要数字。

$$\text{方差} \qquad \sigma^2 = \mathrm{E}\,[\,(x-m)^2\,] = p_1(x_1-m)^2 + \ldots + p_n(x_n-m)^2\,] \tag{5}$$

我们把与期望值 $m = \mathrm{E}(x)$的距离平方，我们没有样本，只有期望，我们知道概率但是我们不知道试验结果。

范例 1　求出大一新生年龄的方差σ^2。

解　年龄 $x_i = 17, 18, 19$ 的概率是 $p_i = 0.2, 0.5$ 与 0.3，期望值是 $m = \sum p_i x_i = 18.1$，方差使用这些样本概率：

$$\begin{aligned}\sigma^2 &= (0.2)(\mathbf{17-18.1})^2 + (0.5)(\mathbf{18-18.1})^2 + (0.3)(\mathbf{19-18.1})^2 \\ &= (0.2)(\mathbf{1.21}) + (0.5)(\mathbf{0.01}) + (0.3)(\mathbf{0.81}) = \mathbf{0.49}.\end{aligned}$$

标准方差是平方根$\sigma = 0.7$。

这是量测 17, 18, 19 环绕 E[x]的分布情况，加权是概率.2, .5, .3。

连续概率分布

直到现在我们才允许 n 个可能的结果 $x_1, \ldots, x_n$。对于年龄 17, 18, 19，我们只有 $n = 3$。如果我们用天量测而不是用年，会有 1000 个可能的结果(太多了)，最好是允许介于 17 到 20 之间的每个数——一个可能年龄的**连续**。对应年龄 x_1, x_2, x_3 的概率 p_1, p_2, p_3 必须移到一个**概率分布(probability distribution)** $p(x)$，对应一整个 $17 \le x \le 20$ 的连续范围。

说明概率分布的最佳方法是给你**两个例子**，他们是**均匀分布(uniform)**与**正态分布**(normal)。第一个(均匀)很简单，正态分布最重要。

均匀分布　假设年龄在 17.0 到 20.0 之间均匀分布，两个数字之间的年龄发生的机会相等，当然任何一个确切的年龄完全没有机会。当你要命中确切数字 17.1 或 $17 + \sqrt{2}$ 的概率将是零。你真实能够提供(假设我们的均匀分布)的是**一个随机大一新生的年龄小于 x 的机会 $F(x)$：**

年龄小于 $x = 17$ 的机会是 $F(17) = 0$　　$x \le 17$ 不会发生

年龄小于 $x = 20$ 的机会是 $F(20) = 1$　　$x \le 20$ 一定发生

年龄小于 x 的机会是 $\boldsymbol{F(x) = \frac{1}{3}(x-17)}$　　**F 的值从 0 到 1**

公式 $F(x) = \frac{1}{3}(x-17)$在 $x = 17$ 时 $F = 0$，表示 $x \le 17$ 不会发生。在 $x = 20$ 时给出 $F = 1$，表示 $x \le 20$ 一定发生。针对均匀模型，介于 17 与 20 的**累积(cumulative)分布 $F(x)$**的图形是线性增加。

请尊重版权与译者的劳动成果，侵权必究!

让我画出 $F(x)$与它的导数 $p(x)$ = “**概率密度(probability density)函数**”的图形。

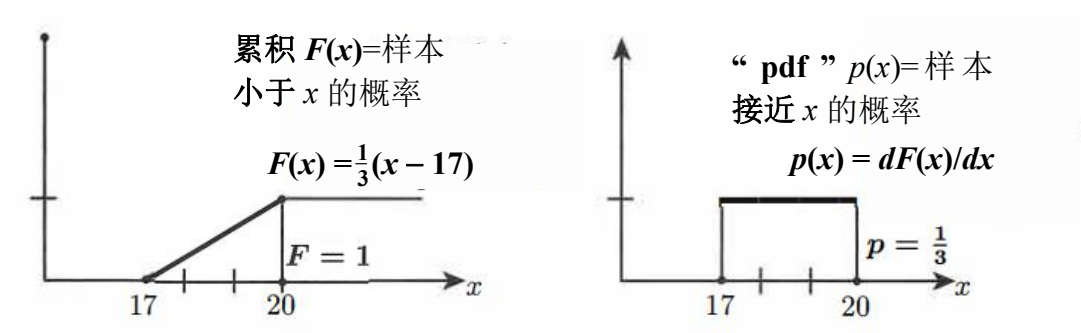


图 12.1：$F(x)$是累积分布而且它的导数 $p(x) = dF(x)/dx$ 是**概率密度函数(pdf)**。对于均匀分布，$p(x)$在 17 到 20 之间是常数，$p(x)$图形下的总面积是总概率 $F = 1$。

你可以说 $p(x)dx$ 是一个样本**落在 x 与 $x + dx$ 之间的概率**，这是“无限小的真实”：$p(x)dx = F(x + dx) - F(x)$，这里是整个事实：

$$\boldsymbol{F = p \text{ 的积分}} \qquad \boldsymbol{a \le x \le b \text{ 的概率}} = \int_a^b p(x)\,dx = F(b) - F(a) \tag{6}$$

$F(b)$是 $x \le b$ 的概率，我减去 $F(a)$维持 $x \ge a$，留下了 $a \le x \le b$。

$p(x)$的平均值与方差

一个概率分布的平均值 m 与方差 σ^2 为何？前面我们累加 $p_i x_i$ 得到平均值(期望值)，连续分布的情况下，我们**积分 $xp(x)$**：

$$\text{平均值} \qquad \boldsymbol{m = E[x]} = \int xp(x)\,dx = \int_{x=17}^{20} (x)\left(\frac{1}{3}\right) dx = 18.5$$

对于均匀分布来说，平均值 m 在 17 与 20 的中间，在中间点 $m = 18.5$ 以下的随机值 x 的概率是 $\boldsymbol{F(m) = 1/2}$。

在 MATLAB 中，x = rand(1)会在 0 与 1 之间均匀选择一个随机数字，期望平均值是 $m = 1/2$。从 0 到 x 的区间有概率 $F(x) = x$，在平均值 m 以下的区间永远有概率 $F(m) = 1/2$。

方差是与平均值距离的**平方的平均值**，如果有 N 个结果，σ^2 是 $p_i(x_i - m)^2$ 的总和。对于连续的随机变数 x，总和变成积分：

$$\text{方差} \qquad \boldsymbol{\sigma^2 = E[\,(x - m)^2\,]} = \int p(x)(x - m)^2\,dx \tag{7}$$

当年龄在 $17 \le x \le 20$ 之间均匀分布，积分可以平移到 $0 \le x \le 3$：

$$\sigma^2 = \int_{17}^{20} \frac{1}{3}(x-18.5)^2\,dx = \int_0^3 \frac{1}{3}(x-1.5)^2\,dx = \frac{1}{9}(x-1.5)^3\Big|_{x=0}^{x=3} = \frac{2}{9}(1.5)^3 = \frac{3}{4}$$

这是典型的例子，这里是对于均匀 $p(x)$的完整轮廓，0 到 a。

对于 $0 \le x \le a$ 的均匀分布，密度 $p(x) = 1/a$　累积分布 $F(x) = x/a$

平均值 $m = a/2$ 中间值　　**方差** $\sigma^2 = \int_0^a \frac{1}{a}\left(x - \frac{a}{2}\right)^2 dx = \frac{a^2}{12}$　　(8)

平均值是 a 的倍数，方差是 a^2 的倍数。$a = 3$ 时，$\sigma^2 = 9/12 = 3/4$。对于一个介于 0 到 1(平均值是 1/2)的随机数，方差是$\sigma^2 = 1/12$。

正态分布：钟形曲线

正态分布也称为高斯分布，它是所有概率密度函数 $p(x)$中**最重要**的函数，它的超级重要性的理由来自重复一个实验以及实验的平均结果。实验有他们自己的分布(例如头与尾)，平均值会趋近正态分布。

中央极限定理(central limit theorem) (非正式)　任意概率分布在 $N \to \infty$时，它的 N 个样本的平均值会趋近正态分布。

从“标准正态分布”开始，它是对称于 $x = 0$，所以它的平均值是 $m = 0$，它被选成一个标准方差$\sigma^2 = 1$，称为 **N**(0, 1)。

标准正态分布　　$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$　　(9)

$p(x)$的图形是图 12.2 的**钟形曲线**，标准事实是：

总概率= 1　　$\int_{-\infty}^{\infty} p(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$

平均值 E[x] = 0　　$m = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-x^2/2} dx = 0$

方差 E[x^2] = 1　　$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-0)^2\, e^{-x^2/2} dx = 1$

请尊重版权与译者的劳动成果，侵权必究!

零平均值很简单，因为对一个奇函数积分，把 x 换成$-x$ 显示“积分 $=-$积分”，所以积分必须是 $m=0$。

另外两个积分应用问题 12 的概念得到 1，图 12.2 显示正态分布 $\mathbf{N}(0,\sigma)$的 $p(x)$的图形，以及累积分布 $F(x)=p(x)$的积分的图形。从 $p(x)$的对称性，你看到平均值 $=$ 零，从 $F(x)$你看到针对民意调查非常重要的实际近似：

随机样本落在$-\sigma$与σ之间的概率是 $\boldsymbol{F(\sigma)-F(-\sigma)\approx 2/3}$。

$$\text{这是因为 } \int_{-\sigma}^{\sigma} p(x)\,dx=\int_{-\infty}^{\sigma} p(x)\,dx-\int_{-\infty}^{-\sigma} p(x)\,dx=F(\sigma)-F(-\sigma)$$

近似的，随机 x 落在-2σ与 2σ(离开平均值的距离小于两个标准差)之间的概率是 $F(2\sigma)-F(-2\sigma)\approx 0.95$。如果你有一个实验结果距离平均值比 2σ还远，它几乎可以确定不是偶然发生：机会 $=0.05$。药物测试要寻找更严格的确认，比如概率 0.001。寻找希格斯玻色子(Higgs boson)时使用远离偶然的 5σ 标准差的超严格测试。

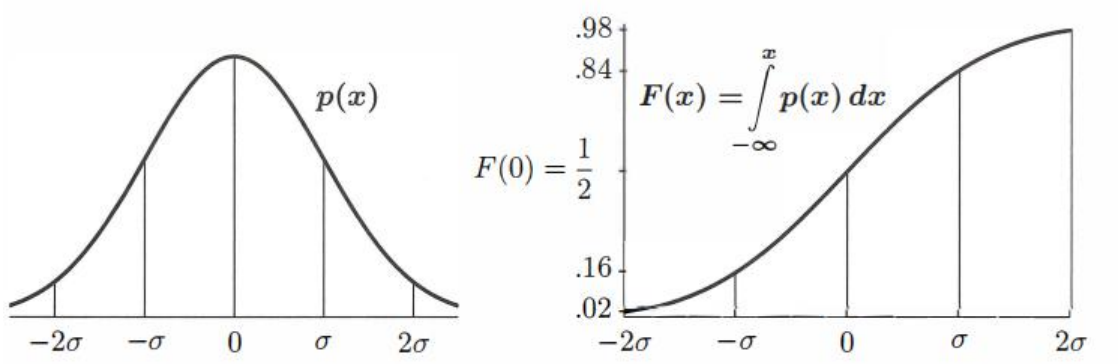


图 12.2：标准正态分布 $p(x)$有平均值 $m=0$ 与$\sigma=1$。

有任意平均值 m 与标准差σ 的正态分布，是来自平移与拉伸标准的 $\mathbf{N}(0,1)$。x 平移到 $x-m$，拉伸 $x-m$ 成为$(x-m)/\sigma$。

高斯密度 $\boldsymbol{p(x)}$
正态分布 $\mathbf{N}(\boldsymbol{m},\ \boldsymbol{\sigma})$

$$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-m)^2/2\sigma^2} \tag{10}$$

$p(x)$的积分是 $F(x)$——随机样本小于 x 的概率，微分 $p(x)dx=F(x+dx)-F(x)$是一个随机样本落在 x 与 $x+dx$ 之间的概率。没有简单的公式可以积分$e^{-x^2/2}$，所以累积分布 $F(x)$需要仔细计算与列表。

范例 2　**假设 x 是 1 或–1，概率是相等的 $p_1 = p_{-1} = 1/2$。**

平均值是 $m = ½(1) + ½(-1) = 0$，方差是 $\sigma^2 = ½(1)^2 + ½(-1)^2 = 1$。关键问题是平均值 $A_N = (x_1 + ... + x_N)/N$，无关的 x_i 是±1，而且我们除以他们的总和 N，A_N 的期望平均值仍然是零。**大数法则**说明这个样本平均值趋近零的概率是 1，A_N 趋近零会有多快？**它的方差 σ_N^2 为何？**

$$\text{依据线性}\quad \sigma_N^2 = \frac{\sigma^2}{N^2} + \frac{\sigma^2}{N^2} + ... + \frac{\sigma^2}{N^2} = N\frac{\sigma^2}{N^2} = \frac{1}{N},\ \text{这是因为}\sigma^2 = 1 \qquad (11)$$

范例 3　**改变输出，从 1 或–1 改成 $x = 1$ 或 $x = 0$，维持 $p_1 = p_0 = 1/2$。**

新的平均值是 $m = 1/2$，这是 0 与 1 的中间值。方差移到 $\sigma^2 = 1/4$：

$$\boldsymbol{m} = ½(1) + ½(0) = \mathbf{1/2} \quad \text{与} \quad \boldsymbol{\sigma^2} = ½(1 - ½)^2 + ½(0 - ½)^2 = \mathbf{1/4}$$

平均 A_N 现在有平均值 1/2 与方差 $\dfrac{1}{4N^2} + ... + \dfrac{1}{4N^2} = \dfrac{1}{4N} = \boldsymbol{\sigma_N^2}$ (12)

这个 σ_N 是范例 2 的 σ_N 的一半，这个必须是正确的，因为新的值域是 0 到 1，这是 –1 到 1 的一半。范例 2-3 展示一个线性法则。

新 0 – 1 变数 $x_{新}$ 是 $½x_{旧} + ½$，所以平均值 m 增加到 1/2 以及方差变成 $(1/2)^2$ 倍。一个平移改变 m 与调整比例改变 σ^2。

线性　$\boldsymbol{x_{新} = ax_{旧} + b}$　有　$\boldsymbol{m_{新} = am_{旧} + b}$　与　$\boldsymbol{\sigma^2_{新} = a^2\sigma^2_{旧}}$ (13)

这里是来自 3 个测试的结果：N 次试验的随机 0 或 1 的平均。

[从 N=**100** 有 **48** 个 1's] [从 N=**10000** 有 **5035** 个 1's] [从 N=**40000** 有 **19967** 个 1's]

标准化 $\boldsymbol{X} = (x - m)/\sigma = (A_N - ½)/2\sqrt{N}$ 是 **[–.40] [.70] [–.33]**

中央极限定理说明很多次投掷铜板会趋近一个正态分布，让我们开始观察这是如何发生：**二项式趋近正态**。

每次投掷出现头的概率是 1/2。$N = 3$ 次投掷出现三次头的概率是 $(1/2)^3 = 1/8$。出现两次头一次尾的概率是 3/8，这是来自 3 个序列 HHT 与 HTH 与 THH。数字 1/8 与 3/8 是 $\left(\frac{1}{2} + \frac{1}{2}\right)^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \mathbf{1}$ 的片段。3 次投掷出现头的平均数是 1.5。

$$\textbf{平均值}\quad m = (3\ \text{次头})\frac{1}{8} + (2\ \text{次头})\frac{3}{8} + (1\ \text{次头})\frac{3}{8} + 0 = \frac{3}{8} + \frac{6}{8} + \frac{3}{8} = \mathbf{1.5}\ \ \text{次头}$$

请尊重版权与译者的劳动成果，侵权必究！

投掷 N 次，范例 3(或常识)得到平均值 $m = \sum x_i p_i = \boldsymbol{N/2}$ 次头。

方差σ^2是基于离开平均值$N/2$的距离平方，$N = 3$时方差是$\sigma^2 = 3/4$(就是$N/4$)。要求得σ^2，我们累加$(x_i - m)^2 p_i$，其中 $m = 1.5$：

$$\sigma^2 = (3-1.5)^2\frac{1}{8} + (2-1.5)^2\frac{3}{8} + (1-1.5)^2\frac{3}{8} + (0-1.5)^2\frac{1}{8} = \frac{9+3+3+9}{32} = \frac{3}{4}$$

对于任意的 N，方差$\sigma_N^2 = N/4$，则$\sigma_N = \sqrt{N}/2$。

图 12.3 显示在 $N = 4$ 次投掷中，出现人头次数 0, 1, 2, 3 的概率是如何接近一个钟形高斯。这个高斯的中心是平均值 $N/2 = 2$，要变成标准高斯(平均值 0 与方差 1)，我们平移然后拉伸图形。如果 x 是 N 次投掷出现人头的次数--N 个 0-1 结果的平均值--则 x 平移到它的平均值 $m = N/2$ 然后利用$\sigma = \sqrt{N}/2$ 调整比例，产生标准的 X：

平移与调整比例 $\quad X = \dfrac{x-m}{\sigma} = \dfrac{x - N/2}{\sqrt{N}/2} \quad (N = 4 \text{ 有 } X = x - 2)$

减去 m 是“中心化”或“消除趋势”， X 的平均值是零

除以σ 是“正态化”或“标准化”，X 的方差是 1

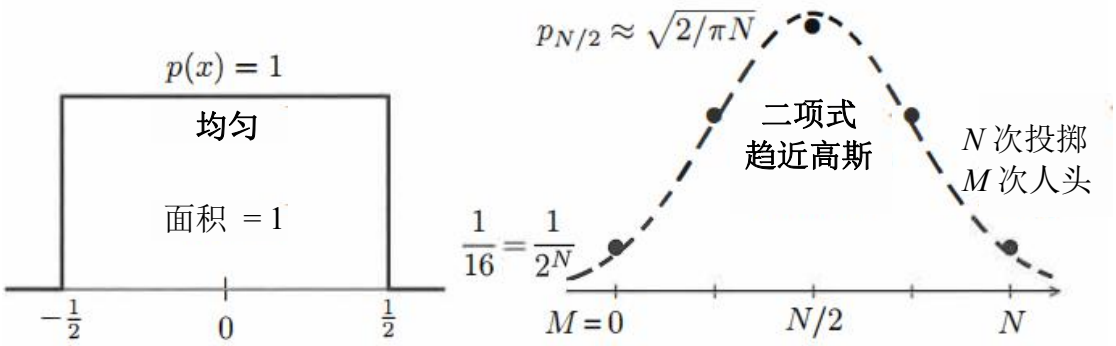


图 12.3：4 次投掷出现人头次数的概率 $p = (1, 4, 6, 4, 1)/16$，这些 p_i 趋近一个高斯分布，其中$\sigma^2 = N/4$，中心点在 $m = N/2$。对于 X，中央极限定理给出收敛到正态分布 $\mathbf{N}(0, 1)$。

看到中央极限定理在中心点 $X = 0$ 给出正确答案是开心的，在这个点的因子$e^{-X^2/2}$等于 1。我们知道 N 次投掷铜板的方差是$\sigma^2 = N/4$，钟形曲线的中心有高度$1/\sqrt{2\pi\sigma^2} = \sqrt{2/N\pi}$。

投掷铜板分布 p_0 到 p_n(二项式分布)的中心点高度为何？当 $N = 4$，人头次数 0, 1, 2, 3, 4 的概率来自$\left(\frac{1}{2} + \frac{1}{2}\right)^4$

中心概率 6/16 $\quad \left(\dfrac{1}{2} + \dfrac{1}{2}\right)^4 = \dfrac{1}{16} + \dfrac{4}{16} + \dfrac{6}{16} + \dfrac{4}{16} + \dfrac{1}{16} = 1$

问题 8 的二项式定理告诉我们对于任意偶数 N 的中心概率：

中心概率(N/2 次人头，N/2 次尾)是 $\dfrac{1}{2^N}\dfrac{N!}{(N/2)!\,(N/2)!}$

$N=4$，这些阶层得到 4!/2!2!= 24/4 = 6。针对大的 N，Stirling 的公式 $\sqrt{2\pi N}(N/e)^N$ 是一个接近 $N!$ 的近似值，使用 Stirling 公式取代 $N!$与两次的$(N/2)!$ ：

【原文少了两个'!'】

投掷铜板中心概率的极限 $$p_{N/2}\approx\frac{1}{2^N}\frac{\sqrt{2\pi N}(N/e)^N}{\pi N(N/2e)^N}=\frac{\sqrt{2}}{\sqrt{\pi N}}=\frac{1}{\sqrt{2\pi}\,\sigma} \tag{14}$$

最后一个步骤我们针对铜板投掷问题使用了方差$\sigma^2=N/4$，结果$\dfrac{1}{\sqrt{2\pi}\,\sigma}$适配高斯的中心值(上述)。中央极限定理成立：当 $N\to\infty$，“二项式分布”趋近正态分布。

蒙特卡罗估测法

科学计算必须处理资料中的误差，金融计算要处理不确定的数字与不必然的预测，所有的应用数学必须移到**接受输入的不确定与估测输出的方差**。

如何估测方差？通常概率分布 $p(x)$是未知，我们能做的就是尝试不同的输入 $\boldsymbol{b}$ 与计算输出 $\boldsymbol{x}$ 然后取平均。这是**蒙特卡罗法**的最简单形式(以位于 Reviera 的赌场命名，在那儿我看过一次由于赌注是否及时落地引起的争斗。) 蒙特卡罗法利用样本平均$(x_1+...+x_n)/N$ 取得期望值 $\mathbf{E}[x]$的近似值。

请了解每个 x_k 都是很花成本去计算的，我们不只是投掷铜板，每个样本来自一个资料 b_k 的集合。蒙特卡罗法随机选择资料 b_k，它计算输出 x_k，然后取这些 x's 的平均值。$\mathbf{E}[x]$要有不错的精确度通常需要很多样本 b 以及巨大的计算成本。利用$(x_1+...+x_n)/N$ 取得 $\mathbf{E}[x]$近似值的误差的阶数通常是$1/\sqrt{N}$，当 N 增加会有缓慢的改善。

这个$1/\sqrt{N}$ 估测出现在方程式(11)的投掷铜板中，对 N 个无关样本 x_k 取平均，方差σ^2 降为方差σ^2/N。

“类似-蒙特卡罗”有时候可以把方差降低到σ^2/N^2：一个巨大的不同！非常仔细的选择输入 b_k--不只是随机。这个 QMC 方法在期刊 Acta Numerica 2013 审阅，更新的概念“多阶层蒙特卡罗”在 Acta Numerica 2015 由 Michale Giles 发表，这里是如何工作的原理：

2 阶层蒙特卡罗 $$\mathbf{E}[x]\approx\frac{1}{N}\sum_1^N y(b_k)+\frac{1}{N^*}\sum_1^{N^*}[x(b_k)-y(b_k)]\,.$$

请尊重版权与译者的劳动成果，侵权必究!

这个概念是 $x-y$ 比原始的 x 有一个较小的方差σ^*，因此在相同的 E[x]的精确度下 N^*会比 N 小。我们做 N 次便宜的仿真求出 y's，每次的成本是 C，我们只做 N^*次与 x's 有关的昂贵仿真，每次的成本是 C^*，全部的计算成本是 $NC + N^*C^*$。

微积分针对固定总成本取整体方差的极小值，最佳比值 N^*/N 是$\sqrt{C/C^*}\ \sigma^*/\sigma$。三阶层蒙特卡罗会仿真 x, y 与 z ：

$$\mathrm{E}[x] \approx \frac{1}{N}\sum_{1}^{N} z(b_k) + \frac{1}{N^*}\sum_{1}^{N^*}[y(b_k) - z(b_k)] + \frac{1}{N^{**}}\sum_{1}^{N^{**}}[x(b_k) - y(b_k)].$$

Giles 最佳化 $N, N^*, N^{**}, \ldots$ 维持 E[x] ≤ 固定的 E_0，而且有提供 MATLAB 代码。

复习：平均值与方差的三个公式

对于 m 与σ^2 的公式所有概率与统计的起点，有三种不同的方式：**样本值** X_i，期望值(离散 p_i)与期望值的值域(连续 $p(x)$)，这里是平均值与方差：。

	m	S^2 或 σ^2
样本 X_1 到 X_N	$m = \dfrac{X_1 + \cdots + X_N}{N}$	$S^2 = \dfrac{(X_1 - m)^2 + \cdots + (X_N - m)^2}{N-1}$
概率 p_i 的 n 个可能输出	$m = \sum_1^n p_i x_i$	$\sigma^2 = \sum_1^n p_i (x_i - m)^2$
概率密度的输出值域	$m = \int x\, p(x)dx$	$\sigma^2 = \int (x-m)^2 p(x)dx$

一个自然的问题：为什么第一行没有概率 p？这些公式如何相通？答案：我们期望一个样本的函数 p_i 是 $X = x_i$，如果真实成立，$X = x_i$ 重复 $p_i N$ 次，则行 1 与行 2 得到相同的 m。

当我们处理样本时，我们不知道 p_i，我们只能记录每个输出 X 来临的次数，我们得到的是经验值而不是期望平均值。

问题集 12.1

1 加 7 到每个输出 x，平均值与方差发生什么？新的样本平均值，期望平均值与新的方差为何？

2 我们知道：1/3 的整数可以被 3 整除，1/7 的整数可以被 7 整除。可以被 3 或被 7 或同时整除的比例是多少？

3 假设你用相同的概率 1/1000 从 1 到 1000 取得样本，样本的最后位数是 0, ..., 9 的概率 p_0 到 p_9 为何？最后一位数的期望平均值为何？方差σ^2 为何？

4 再次从 1 到 1000 取样，但是这次观察样本平方的末位数，这个平方的末端可能是 $x = 0, 1, 4, 5, 6$ 或 9，这些概率 $p_0, p_1, p_4, p_5, p_6, p_9$ 为何？数字 x 的(期望)平均值 m 与方差σ^2 为何？

5 (有一些困难) 再次使用相同的概率从 1 到 1000 取样，令 x 是第一位数(如果数字是 15，$x = 1$)。这些 $x = 1, ..., 9$ 的概率 p_1 到 p_9(总和是 1)为何？x 的平均值与方差为何？

6 假设问题 5 你有 $N = 4$ 个样本 157, 312, 696, 602，这些数字的平方的第一位数 x_1 到 x_4 为何？样本平均值μ为何？样本方差 S^2 为何？记得除以 $N - 1 = 3$，而不是 $N = 4$。

7 方程式(4)给出 S^2 的第二等效形式(使用样本的方差)：

$$S^2 = \frac{1}{N-1}\left((x_i - m)^2\text{的总和}\right) = \frac{1}{N-1}\left(\left(x_i^{\,2}\text{ 的总和}\right) - Nm^2\right)$$

验证期望方差σ^2的适配等式(使用 $m = \sum p_i x_i$：

$$\sigma^2 = \ p_i(x_i - m)^2\text{的总和} = \left(p_i x_i^{\,2}\text{ 的总和}\right) - m^2$$

8 来自人口的全部 24 个样本产生相同的年龄 $x = 20$，则样本平均值μ与样本方差 S^2 为何？如果 $x = 20$ 或 21，分别有 12 次，答案为何？

9 如同 541 页的电脑实验：找出一百万次随机 0-1 样本的平均值 $A_{1000000}$！求出 $X = \left(A_N - \frac{1}{2}\right)\Big/2\sqrt{N}$。

10 在 N 次投掷铜板得到 i 次人头的概率 p_i 是二项式数字 $b_i = \binom{N}{i}$ 除以 2^N，b_i 的总和是$(1 + 1)^N = 2^N$，所以概率 p_i 的总和是 1。

$$p_0 + ... + p_N = \left(\frac{1}{2} + \frac{1}{2}\right)^N = \frac{1}{2^N}(b_0 + ... + b_N) \quad \text{其中} \quad b_i = \frac{N!}{i!\,(N-i)!}$$

$$N = 4 \text{ 得到 } b_0 = \frac{24}{24},\ b_1 = \frac{24}{(1)(6)},\ b_2 = \frac{24}{(2)(2)},\ p_i = \frac{1}{16}(1, 4, 6, 4, 1)$$

注意 $b_i = b_{N-i}$。问题：确认平均值 $m = 0p_0 + ... + Np_N = N/2$。

11 任意函数$f(x)$的期望值是 $\mathrm{E}[f] = \sum p_i f(x_i)$ 或 $\int p(x) f(x)\,dx$ (离散概率或连续概率)。假设平均值是 $\mathrm{E}[x] = m$ 以及方差是 $\mathrm{E}[(x - m)^2] = \sigma^2$，求 $\mathrm{E}[x^2]$。

12 证明标准正态分布 $p(x)$的总概率是如同要求的 $\int p(x)\,dx = 1$，一个著名的把戏是 $\int p(x)\,dx$ 乘 $\int p(y)\,dy$，然后计算对于全部 x 与 y(−∞到∞)的积分。这个把戏是把 $dx\,dy$ 换成 $r\,dr\,d\theta$(极坐标的 $x^2 + y^2 = r^2$)，说明下列的每个步骤：

$$2\pi \int_{-\infty}^{\infty} p(x)\,dx \int_{-\infty}^{\infty} p(y)\,dy = \iint_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy = \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} e^{-r^2/2}\,r\,dr\,d\theta = \mathbf{2\pi}.$$

请尊重版权与译者的劳动成果，侵权必究!

12.2 协方差矩阵与联合概率

当我们一次进行 M 个不同试验时，线性代数进入。我们可以量测年龄与身高与重量(N 个人，每人 $M = 3$ 个量测值)，每个试验有自己的平均值，所以我们有一个向量 $\boldsymbol{m} = (m_1, m_2, m_3)$包含这 M 个平均值。这些可以是年龄与身高与重量样本平均值，或是 m_1, m_2, m_3 可以是基于已知概率的年龄与身高与重量的期望平均值。

当我们关注方差时就会引入矩阵，每个试验有一个样本方差 S_i^2 或一个基于离开平均值距离平方的期望$\sigma_i^2 = \mathrm{E}[(x_i - m_i)^2]$，这 m 个数字$\sigma_1^2, \ldots, \sigma_M^2$ 会落在矩阵的对角线。目前为止我们还没有联结这 M 个平行试验的关系，他们量测 M 个不同的随机变数，但是试验没有必要是无关！

如果我们量测小孩子的年龄与身高与重量($\boldsymbol{a}, \boldsymbol{h}, \boldsymbol{w}$)，这些结果会有强烈的相关性，大一点的小孩通常比较高比较重。假设这些平均值 m_a, m_h, m_w 已知，则$\sigma_a^2, \sigma_h^2, \sigma_w^2$ 是年龄与身高与重量的个别方差，**新的数字是好像σ_{ah} 的协方差(covariance)，这是年龄乘身高。**

$$\text{协方差} = \sigma_{ah} = \mathrm{E}\,[\,(\text{年龄} - \text{平均年龄})\,(\text{身高} - \text{平均身高})\,] \tag{1}$$

这个定义需要仔细观察，要计算σ_{ah}，知道每个年龄的概率与每个身高的概率是不够的，我们必须知道**每个配对(年龄与身高)的联合(joint)概率**，这是因为年龄与身高相关。

p_{ah} = 一个随机小孩有年龄= a 与身高= h 的概率：同时拥有。

p_{ij} = 试验 1 得到 x_i 与试验 2 得到 y_j 的概率。

假设试验 1(年龄)的平均值是 m_1，试验 2(身高)的平均值是 m_2，(1)中试验 1 与 2 之间的协方差关注年龄 x_i 与身高 y_j 的**所有配对**：

$$\text{协方差} \qquad \sigma_{12} = \sum_{\text{全部}\,i,\ j}\sum p_{ij}\,(x_i - m_1)\big(y_j - m_2\big) \tag{2}$$

要捕捉“联合概率 p_{ij} 的概念”我们从两个小例子开始。

范例 1 分开投掷两个铜板，1 代表人头，0 代表尾，结果可以是(1, 1)或(1, 0)或(0, 1)或(0, 0)，这 4 个结果的概率都是 $p_{11} = p_{10} = p_{01} = p_{00} = \mathbf{1/4}$。无关试验有

$$(i,\ j)\text{的概率} = (i\ \text{的概率})\ (j\ \text{的概率})$$

范例 2 把铜板粘在一起，面朝同一个反向。可能性只有(1, 1)与(0, 0)，这些有概率 1/2 与 1/2。概率 p_{10} 与 p_{01} 是零，(1, 0)或(0, 1)不会发生，这是因为铜板粘在一起：同时人头或同时尾。

试验 1 与 2 的概率矩阵 $\quad P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad P = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

让我在 P 停留久一点，来展示它在良好的矩阵表示法中。矩阵显示每个配对(x_i, y_j)的概率--从(x_1, y_1) = (头，头)与(x_1, y_2) = (头，尾)开始，注意行的总和 p_i 与列的总和 P_j 以及全部总和=1.。

概率矩阵 $P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$ $\begin{matrix} p_{11} + p_{12} = \boldsymbol{p}_1 \\ p_{21} + p_{22} = \boldsymbol{p}_2 \end{matrix}$ $\begin{pmatrix} \text{第一个} \\ \text{铜板} \end{pmatrix}$

(第二个铜板)列的总和 $\boldsymbol{P_1}$ $\boldsymbol{P_2}$ **4 个单元总和是 1**

这些数字 p_1, p_2 与 P_1, P_2 称为矩阵 P 的**边际**(marginal)：

$p_1 = p_{11} + p_{12} =$ 来自铜板 1 的人头机会(铜板 2 可以是头或尾)

$P_1 = p_{11} + p_{21} =$ 来自铜板 2 的人头机会(铜板 1 可以是头或尾)

范例 1 展示无关变数，每个概率 $p_{ij} = p_i$ 乘 p_j (本例中 1/2 乘 1/2 得到 $p_{ij} = 1/4$)，这个情况下**协方差σ_{12}会是零**，来自第一个铜板的头或尾无法给出第二个铜板的信息。

无关试验的零协方差σ_{12} $V = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ **=对角协方差矩阵**

无关试验有$\sigma_{12} = 0$，这是因为方程式(2)的每个 $p_{ij} = (p_i)$乘(p_j)。

$$\sigma_{12} = \sum_i \sum_j (p_i)(p_j)(x_i - m_1)(y_j - m_2) = \left[\sum_i (p_i)(x_i - m_1)\right]\left[\sum_j (p_j)(y_j - m_2)\right] = [0][0].$$

粘在一起的铜板展示完美的关联，一个出现头表示另一个也是，协方差σ_{12} 从 0 变成$\sigma_1\sigma_2 = 1/4$--这是σ_{12}的最大可能值：

$$\textbf{平均值} = \frac{1}{2} \qquad \boldsymbol{\sigma_{12}} = \frac{1}{2}\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right) + 0 + 0 + \frac{1}{2}\left(0 - \frac{1}{2}\right)\left(0 - \frac{1}{2}\right) = \frac{1}{4}$$

来自铜板 1 的头或尾得到来自铜板 2 的头或尾的完整信息：

粘合铜板得到最大可能的协方差
奇异协方差矩阵：行列式 = 0

$$V_{\text{粘合}} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

永远有$\boldsymbol{\sigma_1^2\sigma_2^2 \geq \sigma_{12}^2}$，于是$\sigma_{12}$介于$-\sigma_1\sigma_2$与$\sigma_1\sigma_2$之间，协方差矩阵 V 是**正定**(或是粘合铜板的奇异情况下，V 是**正半定**。) 这是关于 M 个试验的 $M \times M$ 协方差矩阵的重要事实。

注意来自 N 次试验的**样本协方差矩阵 $\boldsymbol{S}$** 当然是正半定，每个新样本 X = (年龄，身高，重量)贡献出**样本平均值 $\boldsymbol{\overline{X}}$** 与 S，每个项$(X_i - \overline{X})(X_i - \overline{X})^{\mathrm{T}}$是正半定，然后我们相加得到 S：

$$\overline{X} = \frac{X_1 + \ldots + X_N}{N} \qquad \boldsymbol{S} = \frac{(X_1 - \overline{X})(X_1 - \overline{X})^{\mathrm{T}} + \cdots + (X_N - \overline{X})(X_N - \overline{X})^{\mathrm{T}}}{N - 1} \quad (3)$$

请尊重版权与译者的劳动成果，侵权必究!

回到试验 1 与 2(两个铜板)之间的期望方差σ_{12}：

$$\sigma_{12} = [(\text{输出 1}–\text{平均 1})\text{乘}(\text{输出 2}–\text{平均 2})]\text{的期望值}$$
$$= \sum_{\text{全部 } i,\,j}\sum p_{ij}(x_i - m_1)(y_j - m_2) \tag{4}$$

$p_{ij} \geq 0$ 是看到试验 1 出现 x_i 与试验 2 出现 y_j 的概率。某些输出配对必须出现，因此 N^2 个概率 p_{ij} 的总和是 1。

总概率(所有的配对)是 1
$$\sum_{\text{全部 } i,\,j}\sum p_{ij} = 1 \tag{5}$$

这里我们需要另一个事实。在试验 1 固定一个特定输出 x_i，允许所有的输出 y_j 出现在试验 2，把出现$(x_i, y_1), (x_i, y_2), ..., (x_i, y_n)$的概率相加：

P 的行总和 p_i
$$\sum_{j=1}^{n} p_{ij} = \text{试验 1 出现 } x_i \text{ 的概率 } \boldsymbol{p_i} \tag{6}$$

某些 y_j 必须出现在试验 2！不论两个铜板是完全分开或是粘合一起，我们仍然得到铜板 1 是人头的概率 $p_H = p_{HH} + p_{HT}$ 是 1/2：

(分开) $p_{HH} + p_{HT} = 1/4 + 1/4 = \mathbf{1/2}$　　(粘合) $p_{HH} + p_{HT} = 1/2 + 0 = \mathbf{1/2}$

基础的推理允许我们写出一个矩阵公式，这个公式在试验 1 与试验 2 的个别方差 σ_1^2 与 σ_2^2 之外，引入伴随的协方差σ_{12}。把每个配对(i, j)的矩阵 V_{ij} 相加，我们得到整体的协方差矩阵 V：

协方差矩阵
$\boldsymbol{V = \sum\sum V_{ij}}$
$$\boldsymbol{V} = \sum_{\text{全部 } i,\,j}\sum p_{ij}\begin{bmatrix} (x_i - m_1)^2 & (x_i - m_1)(y_j - m_2) \\ (x_i - m_1)(y_j - m_2) & (y_j - m_2)^2 \end{bmatrix} \tag{7}$$

在非对角线，这是对应协方差σ_{12} 的方程式(2)。在对角线，我们得到正常的方差σ_1^2 与σ_2^2，我会仔细的展示我们如何使用方程式(6)得到 $V_{11} = \sigma_1^2$。允许所有的 j 得到试验 1 出现 x_i 的概率 p_i：

$$\boldsymbol{V}_{11} = \sum_{\text{全部 } i,\,j}\sum p_{ij}(x_i - m_1)^2 = \sum_{\text{全部 } i}(x_i \text{ 的概率})(x_i - m_1)^2 = \boldsymbol{\sigma}_1^{\,2} \tag{8}$$

请观察两次，这是利用一个公式(7)产生整个协方差矩阵的关键。这个公式的美妙之处在于它结合 2×2 矩阵 V_{ij}，并且(7)中对应每个输出 i, j 配对的矩阵 V_{ij} 是**正半定**：

V_{ij} 有对角线单元 $p_{ij}(x_i - m_1)^2 \geq 0$ 与 $p_{ij}(y_j - m_2)^2 \geq 0$ 以及 $\det(V_{ij}) = 0$

这个矩阵 V_{ij} 有秩 1，方程式(7)计算 p_{ij} 乘列 U 乘行 U^{T}:

$$\begin{bmatrix} (x_i-m_1)^2 & (x_i-m_1)(y_j-m_2) \\ (x_i-m_1)(y_j-m_2) & (y_j-m_2)^2 \end{bmatrix} = \begin{bmatrix} x_i-m_1 \\ y_j-m_2 \end{bmatrix} \quad [x_i-m_1 \quad y_j-m_2] \tag{9}$$

每个矩阵 UU^{T} 是正半定，所以整个矩阵 V(利用加权 $p_{ij} \geq 0$ 结合这些矩阵 UU^{T})**最少是正半定**--V 可能是**正定**。

除非试验是相关，否则协方差矩阵 V 是正定。

现在我们从两个变数 x 与 y 来到 M 个变数，比如年龄-身高-重量，来自每个试验的输出是一个有 M 个分量的向量 $\boldsymbol{X}$, (每个小孩有一个 3 个分量的年龄-身高-重量的向量。) 协方差矩阵 V 现在是 $M \times M$，V 是来自输出向量 $\boldsymbol{X}$ 与它们的平均值 $\overline{\boldsymbol{X}} = \mathrm{E}[\boldsymbol{X}]$：

协方差矩阵 $$\boldsymbol{V} = \mathbf{E}\left[(\boldsymbol{X} - \overline{\boldsymbol{X}})\,(\boldsymbol{X} - \overline{\boldsymbol{X}})^{\mathbf{T}} \right] \tag{10}$$

记得 $\boldsymbol{X}\boldsymbol{X}^{\mathbf{T}}$ 与 $\overline{\boldsymbol{X}}\,\overline{\boldsymbol{X}}^{\mathbf{T}}$ = (列) (行)是 $M \times M$ 矩阵。

$M = 1$(一个变数)，你看到 $\overline{\boldsymbol{X}}$ 是平均值 m 与 V 是 σ^2(段落 12.1)。$M = 2$(两个铜板)，你看到 $\overline{\boldsymbol{X}}$ 是平均值(m_1, m_2)与 V 适配方程式(10)，期望值 E 永远是**输出乘概率**的总合。对年龄-身高-重量来说，输出可以是 $\boldsymbol{X}$ = (5 岁，31 吋，48 磅)并且它的概率是 $\boldsymbol{p}_{5,31,48}$。

现在是一个新概念，取任意的线性组合 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X} = c_1X_1 + \ldots + c_MX_M$，当 $\boldsymbol{c} = (6, 2, 5)$，这个就是 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}$ = 6(年龄) + 2(身高) + 5(重量)。基于线性我们知道它的期望值是 $\mathrm{E}[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}] = \boldsymbol{c}^{\mathrm{T}}\,\mathrm{E}[\boldsymbol{X}] = \boldsymbol{c}^{\mathrm{T}}\,\overline{\boldsymbol{X}}$。

$\mathrm{E}[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}] = \boldsymbol{c}^{\mathrm{T}}\,\mathrm{E}[\boldsymbol{X}]$ = 6(年龄期望值) + 2(身高期望值) + 5(重量期望值)，不止如此，我们也知道 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}$ 的方差 σ^2 ：

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X} \text{ 的方差 } = \mathrm{E}\left[(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{c}^{\mathrm{T}}\,\overline{\boldsymbol{X}})\,(\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{c}^{\mathrm{T}}\,\overline{\boldsymbol{X}})^{\mathrm{T}}\right]$$

$$= \boldsymbol{c}^{\mathrm{T}}\,\mathrm{E}\left[(\boldsymbol{X} - \overline{\boldsymbol{X}})\,(\boldsymbol{X} - \overline{\boldsymbol{X}})^{\mathrm{T}}\right]\boldsymbol{c} = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c} \tag{11}$$

现在是关键点：$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}$ 的方差不可能是负数，所以 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c} \geq 0$，依据能量测试 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c} \geq 0$，所以协方差矩阵 V 是正半定。

协方差矩阵 V 开启概率与线性代数的联结：V 等于 $Q\Lambda Q^{\mathrm{T}}$，其中固有值 $\lambda_i \geq 0$ 而且有正交单位向量 $\boldsymbol{q}_1$ 到 $\boldsymbol{q}_M$。

对角化协方差矩阵的意义是求出 M 个无关试验作为原始 M 个试验的组合。

请尊重版权与译者的劳动成果，侵权必究!

告白 我不是很喜欢基于 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{c} \geq 0$ 的证明，期望值符号 E 葬送了**联合概率**的关键概念。请允许我直接证明 V 是正半定(至少是针对年龄-身高-重量的例子)，这个证明就是 $\boldsymbol{V}$ **是每个(年龄-身高-重量)组合的联合概率** $\boldsymbol{p_{ahw}}$，**乘上正半定矩阵** $\boldsymbol{UU^{\mathrm{T}}}$，这里的 U 是 $\boldsymbol{X}-\overline{\boldsymbol{X}}$：

$$\boldsymbol{V}=\sum_{\text{all } a,h,w} p_{ahw}\,\boldsymbol{U}\,\boldsymbol{U}^{\mathrm{T}} \qquad \boldsymbol{U}=\begin{bmatrix}\text{年龄}\\ \text{身高}\\ \text{重量}\end{bmatrix}-\begin{bmatrix}\text{平均年龄}\\ \text{平均身高}\\ \text{平均重量}\end{bmatrix}. \tag{12}$$

这个确切如同方程式(7)中 2×2 铜板投掷矩阵 V，现在 $M=3$。

期望值符号 E 的值也允许 *pdf*'s(例如对应连续随机变数 x 与 y 与 z 的概率密度函数 $p(x,y,z)$。) 如果我们允许年龄与身高与重量是所有的数字，而不是年龄 $i=0, 1, 2, 3$ ....，则我们需要 $p(x,y,z)$而不是 p_{ijk}，本书本段落的总和都要改成积分，但是我们仍然有 $V=\mathrm{E}[\boldsymbol{UU}^{\mathrm{T}}]$：

协方差矩阵 $$V=\iiint p(x,y,z)\,\boldsymbol{UU}^{\mathrm{T}}\,dx\,dy\,dz$$ **其中** $$\boldsymbol{U}=\begin{bmatrix}x-\overline{x}\\ y-\overline{y}\\ z-\overline{z}\end{bmatrix}. \tag{13}$$

永远有 $\iiint p=1$。范例 1-2 强调了 p 如何给出对角 V 或奇异 V：

无关变数 x, y, z $\quad p(x,y,z)=\boldsymbol{p_1(x)\,p_2(y)\,p_3(z)}$

相关变数 x, y, z $\quad p(x,y,z)=\boldsymbol{0}$，除了 $cx+dy+ez=0$ 之外

$z=x+y$ 的平均值与方差

从样本平均值开始，我们有 x 的 N 个样本，他们的平均值(average)是 m_x。我们也有 y 的 N 个样本，他们的平均值是 m_y。$z=x+y$ 的**样本平均值很明显是** $\boldsymbol{m_z=m_x+m_y}$：

总和的平均 = 平均的总和 $$\frac{1}{N}\sum_{1}^{N}(x_i+y_i)=\frac{1}{N}\sum_{1}^{N}x_i+\frac{1}{N}\sum_{1}^{N}y_i. \tag{14}$$

很开心看到这么简单的事情。

乘σ^2(段落 12.1)。$M=2$(两个铜板)，$z=x+y$ 的期望平均值看起来不简单，但是**它必须如同 E[*z*] = E[*x*] + E[*y*]的形式出现**，有一个方法可以看出来。

他们的配对(x_i, y_j)的联合概率是 p_{ij}，它的值与试验是否无关产生关联，这是我们不知道的。

但是总和 $z = x + y$ 的平均值，与 x 及 y 是否无关或是相关并没有关联，期望值仍然是相加：

$$\mathbf{E}[x+y] = \sum_i \sum_j p_{ij}\left(x_i + y_j\right) = \sum_i \sum_j p_{ij} x_i + \sum_i \sum_j p_{ij} y_j \tag{15}$$

所有的总和从 1 到 N，我们可以采用任意顺序相加。在右侧的第一项，沿着概率矩阵 P 的行 i 累加 p_{ij} 得到 p_i，这个双重累加得到 $\mathrm{E}[x]$：

$$\sum_i \sum_j p_{ij} x_i = \sum_i \left(p_{i1} + \ldots + p_{iN}\right) x_i = \sum_i p_i x_i = \mathrm{E}[x]$$

对于最后一项，沿着矩阵的列 j 累加 p_{ij} 得到 y_j 的概率 P_j。这些配对(x_1, y_j)与(x_2, y_j)与...与(x_N, y_j)是全部的 y_j：

$$\sum_i \sum_j p_{ij} y_j = \sum_j \left(p_{1j} + \ldots + p_{Nj}\right) y_j = \sum_j p_j y_j = \mathrm{E}[y]$$

现在方程式(15)说明了 **E[*z*] = E[*x*] + E[*y*]**。

$z = x + y$ 的方差又是如何？联合概率 p_{ij} 与协方差σ_{xy}会参与其中，让我把 $x + y$ 的方差分成 3 个简单的片段：

$$\begin{aligned}\boldsymbol{\sigma_z^2} &= \sum\sum p_{ij}(x_i + y_j - m_x - m_y)^2 \\ &= \sum\sum p_{ij}(x_i - m_x)^2 + \sum\sum p_{ij}(y_j - m_y)^2 + 2\sum\sum p_{ij}(x_i - m_x)(y_j - m_y)\end{aligned}$$

第一个片段是$\boldsymbol{\sigma_x^2}$，第二个片段是$\boldsymbol{\sigma_y^2}$，第三个片段是 $\boldsymbol{2\sigma_{xy}}$。

$$\boldsymbol{z = x + y \text{ 的方差} \quad 是 \quad \sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}} \tag{16}$$

Z=AX 的协方差矩阵

有一个好方法可以看出 $z = x + y$ 的σ_z^2，想象(x, y)是一个向量 $\boldsymbol{X}$，再想象 1×2 矩阵 $A = [1 \quad 1]$乘向量 $\boldsymbol{X}$，则 $\boldsymbol{AX}$ 是 $z = x + y$。方程式(16)的方差σ_z^2 写成矩阵表示法得到：

$$\sigma_z^2 = [1 \quad 1]\begin{bmatrix}\sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2\end{bmatrix}\begin{bmatrix}1 \\ 1\end{bmatrix} \quad \text{就是} \quad \boldsymbol{\sigma_z^2 = AVA^{\mathrm{T}}} \tag{17}$$

你可以看到(17)的$\sigma_z^2 = AVA^{\mathrm{T}}$ 与(16)的$\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$ 是一致。

现在是重点，向量 $\boldsymbol{X}$ 有来自 M 个(不只是 2)试验的 M 个分量，这些试验会有一个 $M \times M$ 的协方差矩阵 $\boldsymbol{V_X}$。矩阵 A 可能是 $K \times M$，则 $\boldsymbol{AX}$ 是一个具有 M 个输出的 K 个组合的向量(不是 2 个输出的 1 一个组合 $x + y$)。

向量 $\boldsymbol{Z} = \boldsymbol{AX}$ 有长度 K，它的协方差矩阵是 $K \times K$ 的 $\boldsymbol{V_Z}$，则协方差矩阵的伟大规则——方程式(17)只是一个 1×2 的例子——是这个美丽公式：$\boldsymbol{AX}$ 的协方差矩阵是 A($\boldsymbol{X}$ 的协方差矩阵)A^{T}：

$$\boldsymbol{Z = AX \text{ 的协方差矩阵是} \quad V_Z = AV_XA^{\mathrm{T}}} \tag{18}$$

对我来说，这个简洁公式展示了矩阵乘法的美丽，我不会去证明这个公式，只是赞赏它。它在应用中经常使用——段落 12.3 会介绍。

请尊重版权与译者的劳动成果，侵权必究!

相关因子ρ

相关因子ρ_{xy}与协方差σ_{xy}有很大的关联，他们都是量测无关性或相关性。从调整比例或是“标准化”随机变数 x 与 y 开始，新 $\boldsymbol{X}=\boldsymbol{x}/\sigma_x$ 与 $\boldsymbol{Y}=\boldsymbol{y}/\sigma_y$ 有$\sigma_x^2=\sigma_y^2=\mathbf{1}$。这就好像把一个向量 $\boldsymbol{v}$ 除以它的长度得到长度 1 的单位向量 $\boldsymbol{v}/\|\boldsymbol{v}\|$一样。

x 与 y 的相关因子是 X 与 Y 的协方差，如果原始 x 与 y 的协方差是σ_{xy}，调整比例至 X 与 Y 后会除以σ_x 与σ_y：

$$\textbf{相关因子}\quad \rho_{xy}=\frac{\sigma_{xy}}{\sigma_x\sigma_y}=\frac{x}{\sigma_x}\ \text{与}\ \frac{y}{\sigma_y}\ \textbf{的协方差}\qquad \textbf{永远有}\quad -1\le\rho_{xy}\le 1$$

零协方差得到零相关，无关随机变数产生$\rho_{xy}=0$。

我们知道永远有$\sigma_{xy}^2\le\sigma_x^2\sigma_y^2$(协方差矩阵 V 至少是正半定)，则$\rho_{xy}^2\le 1$。接近$\rho=+1$ 的相关因子在相同的方向有强烈的相关性：通常结果都是一样。负的相关因子表示当 x 大于平均值时，y 倾向小于平均值：结果在相反的方向。

范例 3　当假设 y 刚好是$-x$，投掷铜板的输出是 $x=0$ 或 1，相同的投掷有输出 $y=0$ 或-1。一个公正铜板的平均值 m_x 是 1/2 与 m_y 是$-1/2$。协方差是$\sigma_{xy}=-\sigma_x\sigma_y$，这个协方差除以$\sigma_x\sigma_y$ 得到$\rho_{xy}=-1$。这个情况下相关矩阵(correlation matrix)R 有零的行列式(奇异且半定)：　【原文写成 correlation 是错误】

$$\textbf{相关矩阵}\ R=\begin{bmatrix}1 & \rho_{xy}\\ \rho_{xy} & 1\end{bmatrix}\qquad \text{当}\ \boldsymbol{y=-x},\ \boldsymbol{R}=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$$

R 永远有 1's 在对角线，这是因为我们正态化成为$\sigma_X=\sigma_Y=1$。R 对应 x 与 y 的相关矩阵，也是对应 $X=x/\sigma_x$ 与 $Y=y/\sigma_y$ 的协方差矩阵。

数字ρ_{xy}也称为皮尔森(Pearson)系数。

范例 4　假设随机变数 x, y, z 是无关，矩阵 R 为何？

答案　R 是一个单位矩阵，全部 3 个相关因子$\rho_{xx}, \rho_{yy}, \rho_{zz}$基于定义都是 1。原文无关，所以 3 个交叉相关因子$\rho_{xy}, \rho_{xz}, \rho_{yz}$都是零。

当我们重调每个行与列的比例，相关矩阵 R 来自协方差矩阵 V。把每个行 i 与列 i 除以第 i 个标准方差σ_i。

(a)　$R=DVD$，其中对角矩阵 $D=\text{diag}(1/\sigma_1, ..., 1/\sigma_M)$。

(b)　若协方差 V 是正定，相关矩阵 $R=DVD$ 也是正定。

已解范例

12.2A 假设 x 与 y 是无关的随机变数，其中平均值是 0 与方差是 1，对应 $\boldsymbol{X}=(x, y)$的协方差矩阵 $\boldsymbol{V_X}$ 是 2×2 单位矩阵。具有 3 分量的向量 $\boldsymbol{Z}=(x, y, ax+by)$的平均值 $\boldsymbol{m_Z}$ 与协方差矩阵 $\boldsymbol{V_Z}$ 为何？

解 $\boldsymbol{Z}$ **经由** $\boldsymbol{A}$ **与** $\boldsymbol{X}$ **产生关联** $\boldsymbol{Z}=\begin{bmatrix} x \\ y \\ ax+by \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\boldsymbol{AX}$

向量 $\boldsymbol{m_X}$ 包含 $\boldsymbol{X}$ 的 M 个分量的平均值，向量 $\boldsymbol{m_Z}$ 包含 $\boldsymbol{Z}=\boldsymbol{AX}$ 的 K 个分量的平均值，连接 $\boldsymbol{X}$ 与 $\boldsymbol{Z}$ 的两个平均值的矩阵必须是线性：$\boldsymbol{m_Z}=\boldsymbol{Am_X}$。$ax+by$ 的平均值是 am_x+bm_y。

$\boldsymbol{Z}$ 的协方差矩阵是 $\boldsymbol{V_Z}=AA^{\mathrm{T}}$，当 $\boldsymbol{V_X}$ 是 2×2 单位矩阵时：

$$\boldsymbol{V_Z}=(\boldsymbol{Z}=(\boldsymbol{x, y, ax+by})\text{的协方差矩阵})=\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{bmatrix}\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix}=\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ a & b & a^2+b^2 \end{bmatrix}$$

说明：x 与 y 是无关所以 $\sigma_{xy}=0$，则 x 与 $ax+by$ 的协方差是 a，y 与 $ax+by$ 的协方差是 b，这些刚好来自 $ax+by$ 的两个无关部分。最终，方程式(18)给出 $ax+by$ 的**方差**：

$$\text{使用 } \boldsymbol{V_Z}=\boldsymbol{AV_XA^{\mathrm{T}}} \qquad \sigma^2_{ax+by}=\sigma^2_{ax}+\sigma^2_{by}+2\sigma_{ax,by}=a^2+b^2+0$$

3×3 矩阵 $\boldsymbol{V_Z}$ 是奇异，它的行列式是 $a^2+b^2-a^2-b^2=0$。第 3 个分量 $z=ax+by$ 完全与 x 及 y 相关，$\boldsymbol{V_Z}$ 的秩只有 2。

GPS 范例 来自 GPS 卫星的信号包含它的出发时间，接收器的时钟给出到达时间，接收器把传播时间乘上光速，就知道与卫星的距离。来自 4 个或以上的卫星定位接收器的位置(使用最小平方！)

一个问题：光的速度在电离层(ionosphere)会改变，但是对于附近的接收器来说，所需的修正几乎都一样。如果一个接收器在已知位置，我们可以从那个位置取差分，**差分 GPS** 降低误差的方差：

$$\text{差分矩阵 } \boldsymbol{A}=[1 \quad -1] \qquad \text{协方差矩阵 } \boldsymbol{V_Z}=\boldsymbol{AV_XA^{\mathrm{T}}}$$

$$\boldsymbol{V_Z}=[1 \quad -1]\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}=\sigma_1^2-2\sigma_{12}+\sigma_2^2$$

光速中的误差消失了，达成“厘米定位”的精确度。(关键概念在 320 页，全球定位系统演绎法，Borre, Strang) GPS 的世界就是时间，空间与惊人的精确度。

请尊重版权与译者的劳动成果，侵权必究!

问题集 12.2

1 (a) 当投掷铜板的概率是 p 与 $1-p$，计算方差σ^2。(尾 $=0$，头 $=1$)。

(b) N 个无关铜板的投掷(0 或 1)的总和是 N 次试验后出现人头的次数，规则总和的方差规则(16-17-18)给出$\sigma^2=$ _____。

2 试验 3 的结果 x_1 到 x_n 之间的协方差σ_{kl}为何？试验 5 的结果 y_1 到 y_n 之间的协方差为何？你的公式看起来像方程式(2)的σ_{12}，则协方差矩阵 V 的(3, 5)与(5, 3)单元是$\sigma_{35}=\sigma_{53}$。

3 $M=3$ 个试验，协方差矩阵 V 是 3×3，存在一个输出是 x_i 与 y_j 与 z_k 的概率 p_{ijk}，写出一个如同方程式(7)的矩阵 V 的公式。【原文多了 variance】

4 $M=3$ 个无关试验，平均值分别是 m_1, m_2, m_3，方差分别是$\sigma_1^2, \sigma_2^2, \sigma_3^2$，求出协方差矩阵 V?

问题 5-9 有关已知 $X=x_i$ 得到 $Y=y_j$ 的条件概率(conditional probability)。表示法 **Prob($Y=y_j \mid X=x_i$) =** 给定 $X=x_i$，结果是 y_j 的概率。

范例 1 铜板 1 与铜板 2 粘合，则 Prob(Y =人头，当 X =人头)是 1。

范例 2 无关的铜板投掷：X 的信息与 Y 无关，知道 X 没有用处。

则 Prob(Y =人头 $\mid X$ =人头)与 Prob(Y =人头)相等。

5 说明条件概率的**加法规则**：

$\text{Prob}(Y=y_j)=$ 针对所有输出 x_i 的 $\text{Prob}(Y=y_j \mid X=x_i)$的累加。

6 $n\times n$ 矩阵 P 包含**联合概率** $p_{ij}=\text{Prob}(X=x_i$ 且 $Y=y_j)$。说明为什么

条件概率 $\text{Prob}(Y=y_j \mid X=x_i)$ 等于 $\dfrac{p_{ij}}{p_{i1}+\ldots+p_{in}}=\dfrac{p_{ij}}{p_i}$。

7 下列的联合概率矩阵有 $\text{Prob}(x_1, y_2)=0.3$，求出 $\text{Prob}(y_2 \mid x_1)$与 $\text{Prob}(x_1)$。

$$P=\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}=\begin{bmatrix} .1 & .3 \\ .2 & .4 \end{bmatrix}$$

单元 p_{ij} 的总和是 1

某些 i 与 j 必须发生

8 说明条件概率的**乘积规则**：

$p_{ij}=\text{Prob}(X=x_i$ 且 $Y=y_j)$等于 $\text{Prob}(Y=y_j \mid X=x_i)$ 乘 $\text{Prob}(X=x_i)$。

9 从问题 8 的乘积规则，推导 p_{ij} 的贝叶斯(Bayes)定理：

$$\text{Prob}\,(Y=y_j \text{ 且 } X=x_i)=\frac{\text{Prob}\,(X=x_i \mid Y=y_j)\,\text{Prob}\,(Y=y_j)}{\text{Prob}\,(X=x_i)}$$

“贝叶斯学派”使用先前的信息，“频率学派”只使用取样信息。

12.3 多元高斯与加权最小平方

正态概率密度 $p(x)$ (高斯)只跟两个数有关：

平均值 m 与方差σ^2　　$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-m)^2/2\sigma^2} \tag{1}$$

$p(x)$的图形是个中心在 $x=m$ 的钟形曲线，连续变数 x 可以在$-\infty$与∞之间的任意位置。如果概率接近 2/3，随机 x 会落在 $m-\sigma$ 与 $m+\sigma$ 之间(离开平均值 m 的距离小于一个标准差)。

$$\int_{-\infty}^{\infty}p(x)\,dx=1 \quad 与 \quad \int_{m-\sigma}^{m+\sigma}p(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}e^{-X^2/2}dX\approx\frac{2}{3} \tag{2}$$

这个积分有一个从 x 到 $X=(x-m)/\sigma$ 的变数变换，把指数简化成$-X^2/2$ 以及把积分的极限简化成–1 到 1。就连来自 p 的σ也在积分外面消失，这是因为 $dX=dx/\sigma$。每个高斯变成**标准高斯** $p(X)$，其中平均值 $m=0$ 与方差$\sigma^2=1$，就称为 $p(x)$：

标准正态分布 $N(0,1)$　有　$$p(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \tag{3}$$

对 $p(x)$从$-\infty$到 x 积分得到累积分布 $F(x)$：随机样本小于 x 的概率，这个项概率在 $x=0$(平均值)时 $F=1/2$。

二维高斯

现在我们有 $M=2$ 个高斯随机变数 x 与 y，分别有平均值 m_1 与 m_2，以及方差σ_1^2与σ_2^2。如果他们是无关，他们的概率密度 $p(x,y)$就是 $\boldsymbol{p_1(x)}$**乘** $\boldsymbol{p_2(y)}$。当变数是无关时，概率直接相乘：

无关 x 与 y　　$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-(x-m_1)^2/2\sigma_1^2}e^{-(y-m_2)^2/2\sigma_2^2} \tag{4}$$

x 与 y 的协方差会是$\sigma_{12}=0$，协方差矩阵 V 是对角矩阵，方差σ_1^2 与σ_2^2 永远在 V 的主对角线。$p(x,y)$的指数就是 x-指数与 y-指数的总和，很好的注意到这两个指数可以结合成为$-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{m})$，其中 V^{-1} 在中间：

$$-\frac{(x-m_1)^2}{2\sigma_1^{\,2}}-\frac{(y-m_2)^2}{2\sigma_2^{\,2}}=-\frac{1}{2}\begin{bmatrix}x-m_1 & y-m_2\end{bmatrix}\begin{bmatrix}\sigma_1^{\,2} & 0\\ 0 & \sigma_2^{\,2}\end{bmatrix}^{-1}\begin{bmatrix}x-m_1\\ y-m_2\end{bmatrix} \tag{5}$$

请尊重版权与译者的劳动成果，侵权必究!

我们准备好抛弃无关性质，当 V 不再是对角矩阵时，具有 V^{-1} 的指数(5)仍然是正确。**现在的高斯与向量 $\boldsymbol{m}$ 及矩阵 $\boldsymbol{V}$ 有关。**

当 $M = 2$，第一个变数 x 会给出第二个变数 y 的部分信息(反之亦然)，也许 y 的某个部分由 x 决定，其他部分是真实无关。正是 $M\times M$ 协方差矩阵 V 负责 M 个变数 $\boldsymbol{x} = x_1, ..., x_M$ 之间的相关性，它的逆矩阵 V^{-1} 进入 $p(x)$：

多元高斯概率分布

$$p(x) = \frac{1}{\left(\sqrt{2\pi}\right)^M \sqrt{\det V}} e^{-(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}} V^{-1}(\boldsymbol{x}-\boldsymbol{m})/2} \tag{6}$$

向量 $\boldsymbol{x} = (x_1, ..., x_M)$与 $\boldsymbol{m} = (m_1, ..., m_M)$包含随机变数与他们的平均值，引进 2π 的平方根的 M 次方与 V 的行列式是为了使得总概率等于 1。让我使用线性代数验证，我使用对称矩阵 $V = Q\Lambda Q^{\mathrm{T}}$ 的固有值λ与正交单位固有向量 $\boldsymbol{q}$，所以 $\boldsymbol{V^{-1} = Q\Lambda^{-1}Q^{\mathrm{T}}}$：

$$\boldsymbol{X = x - m} \qquad \boldsymbol{(x-m)^{\mathrm{T}}V^{-1}(x-m) = X^{\mathrm{T}}Q\Lambda^{-1}Q^{\mathrm{T}}X = Y^{\mathrm{T}}\Lambda^{-1}Y}$$

很棒！组合 $\boldsymbol{Y = Q^{\mathrm{T}}X = Q^{\mathrm{T}}(x-m)}$是统计无关(statistical independent)，他们的协方差矩阵Λ是对角矩阵。

利用 V 的固有向量矩阵 Q 进行对角化的步骤与把随机变数“解除关联”是相同的，对于新的变数 $X_1, ..., X_M$ 的协方差是零。这是线性代数协助微积分计算多维积分的重点。 【原文 m 是错误】

当我们执行中心化把变数 $\boldsymbol{x}$ 减去 m 得到 $\boldsymbol{X}$，然后旋转变数得到 $\boldsymbol{Y = Q^{\mathrm{T}}X}$，此时 $p(x)$的积分并没有改变，矩阵Λ是对角矩阵！所以我们把积分分割成 M 个分别已知的一维积分：

$$\int \cdots \int e^{-\boldsymbol{Y^{\mathrm{T}}\Lambda^{-1}Y}/2}\, \boldsymbol{dY} = \left(\int_{-\infty}^{\infty} e^{-y_1^2/2\lambda_1}\, dy_1\right) \cdots \left(\int_{-\infty}^{\infty} e^{-y_M^2/2\lambda_M}\, dy_M\right)$$

$$= \left(\sqrt{2\pi\lambda_1}\right) \cdots \left(\sqrt{2\pi\lambda_M}\right) = \left(\sqrt{2\pi}\right)^M \sqrt{\det V}. \tag{7}$$

V 的行列式(也是Λ的行列式)是固有值$(\lambda_1), ..., (\lambda_M)$的乘积，则(7)给出应该除以的正确数字使得方程式(6)的 $p(x_1, ..., x_M)$有预期的积分=1。

$p(x)$的平均值与方差也是 M 维的积分。利用 V 的固有向量进行对角化的相同概念，以及引进 $\boldsymbol{Y = Q^{\mathrm{T}}X}$ 可以求出这些积分：

平均值的向量 $\boldsymbol{m}$

$$\int \cdots \int \boldsymbol{x}\, p(\boldsymbol{x})\, \boldsymbol{dx} = (m_1, m_2, \ldots) = \boldsymbol{m} \tag{8}$$

协方差矩阵 V

$$\int \cdots \int (\boldsymbol{x}-\boldsymbol{m})\, p(\boldsymbol{x}) (\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\, \boldsymbol{dx} = \boldsymbol{V}. \tag{9}$$

结论：对应概率密度函数 $p(x)$的方程式(6)拥有完美想要的全部性质。

加权最小平方

在第四章，最小平方从无解的系统 $A\boldsymbol{x} = \boldsymbol{b}$ 开始，我们选择 $\hat{\boldsymbol{x}}$ 来极小化误差$\|\boldsymbol{b} - A\boldsymbol{x}\|^2$，这个引导我们得到最小平方方程式 $A^{\mathrm{T}}A\,\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$。最佳 $A\,\hat{\boldsymbol{x}}$ 是 $\boldsymbol{b}$ 在 A 的列空间的投影，但是这个距离平方 $E = \|\boldsymbol{b} - A\boldsymbol{x}\|^2$ 是我们想要极小化的正确误差量测吗？

如果 $\boldsymbol{b}$ 中的量测误差是无关的随机变数，其中平均值 $m = 0$ 与方差$\sigma^2 = 1$ 以及是一个正态分布，此时高斯会说**好的**：使用最小平方。如果误差不是无关或是他们的方差不相等，高斯会说**不行**：使用**加权**最小平方，本段落会展示误差的优质量测是 $\boldsymbol{E} = (\boldsymbol{b} - A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}^{-1}(\boldsymbol{b} - A\boldsymbol{x})$，针对最佳 $\hat{\boldsymbol{x}}$ 的方程式使用协方差矩阵 V：

$$\textbf{加权最小平方} \qquad A^{\mathrm{T}}V^{-1}A\,\hat{\boldsymbol{x}} = A^{\mathrm{T}}V^{-1}\boldsymbol{b} \tag{10}$$

最重要的范例是 $\boldsymbol{b}$ 有 m 个无关的误差，这些误差有方差$\sigma_1^2, \ldots, \sigma_m^2$。依据无关性质，$V$ 是对角矩阵，优质加权 $1/\sigma_1^2, \ldots, 1/\sigma_m^2$ 来自 V^{-1}，我们的目标是加权 $\boldsymbol{b}$ 的误差使得**方差 $= 1$**：

$$\textbf{加权 } \boldsymbol{b} \textbf{ 的无关误差的最小平方} \qquad \textbf{极小化} \quad E = \sum_{i=1}^{m} \frac{(\boldsymbol{b} - A\boldsymbol{x})_i^2}{\sigma_i^2} \tag{11}$$

借由加权这些误差，我们是在“白色化”杂讯。“**白色杂讯**”是一个无关误差的快速描述，其中的误差是基于零平均值与$\sigma^2 = 1$ 的标准高斯 $\mathbf{N}(0, 1)$。

让我写下求出最佳$\hat{\boldsymbol{x}}$过程中抵达方程式(10)与(11)的步骤：

从 $A\boldsymbol{x} = \boldsymbol{b}$ 开始 (m 个方程式，n 个未知数，$m > n$，无解)。

每个右侧 b_i 有零平均值与方差σ_i^2，b_i 是无关。

第 i 个方程式除以σ_i，每个 b_i/σ_i 有方差=1。

上述的除法把 $A\boldsymbol{x} = \boldsymbol{b}$ 变成 $V^{-1/2}A\boldsymbol{x} = V^{-1/2}\boldsymbol{b}$，其中 $V^{-1/2} = \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_m)$。

这些加权方程式的正常最小平方有 $A \to V^{-1/2}A$ 与 $\boldsymbol{b} \to V^{-1/2}\boldsymbol{b}$。

$$(V^{-1/2}A)^{\mathrm{T}}(V^{-1/2}A)\,\hat{\boldsymbol{x}} = (V^{-1/2}A)^{\mathrm{T}}V^{-1/2}\boldsymbol{b} \qquad \textbf{是} \qquad A^{\mathrm{T}}V^{-1}A\,\hat{\boldsymbol{x}} = A^{\mathrm{T}}V^{-1}\boldsymbol{b} \tag{12}$$

因为 V^{-1} 的 $1/\sigma^2$，更多可靠的方程式(较小的σ)得到较重的加权，这是加权最小平方的重点。

这些对角加权(未耦合方程式)是最频繁也是最简单的，他们应用到 b_i 的无关误差。当这些量测不是无关，V 不再是对角--但是(12)仍然是正确的加权方程式。

实际上，找出所有的协方差是很困难的工作，对角化 V 比较简单。

请尊重版权与译者的劳动成果，侵权必究!

估测 $\hat{x}$ 的方差

还有一个重点：通常重要的方程式不是对应量测值 $\boldsymbol{b}$ 的一个特定集合的最佳 $\hat{\boldsymbol{x}}$，这只是一个样本！真实的目的是要知道整个试验的可靠性，这个由**估测 $\hat{\boldsymbol{x}}$ 的方差**来量测(作为可靠度)。首先，$\boldsymbol{b}$ 的零平均值给出 $\hat{\boldsymbol{x}}$ 的零平均值，然后联结输入 $\boldsymbol{b}$ 的方差 $\boldsymbol{V}$ 与输出 $\hat{\boldsymbol{x}}$ 的方差 $\boldsymbol{W}$ 的公式变得如此美丽：

对应 $\hat{x}$ 的方差-协方差矩阵 W　　$$\mathbf{E}[(\hat{\boldsymbol{x}}-\boldsymbol{x})(\hat{\boldsymbol{x}}-\boldsymbol{x})^{\mathrm{T}}]=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1} \tag{13}$$

最小可能的方差来自最佳可能的加权，就是 V^{-1}。

这个关键公式是段落 12.2 的完美应用，**如果 $\boldsymbol{b}$ 有协方差矩阵 $\boldsymbol{V}$，则 $\hat{\boldsymbol{x}}=\boldsymbol{L}\boldsymbol{b}$ 有协方差矩阵 $\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}^{\mathrm{T}}$**。前面的方程式(12)告诉我们 L 是 $(A^{\mathrm{T}}V^{-1}A)^{-1}\,A^{\mathrm{T}}V^{-1}$，现在代入 LVL^{T} 可以看到方程式(13)出现了：

$$\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}^{\mathrm{T}}=(A^{\mathrm{T}}V^{-1}A)^{-1}\,A^{\mathrm{T}}V^{-1}\quad V\quad V^{-1}A(A^{\mathrm{T}}V^{-1}A)^{-1}=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}$$

这是输出的协方差矩阵 W，我们的最佳估测 $\hat{\boldsymbol{x}}$。范例的时间到了。

范例 1　假设医生量测你的心率 x 三次($m=3$，$n=1$)：

$$\begin{matrix}x=b_1\\x=b_2\\x=b_3\end{matrix}\quad 是\quad A\boldsymbol{x}=\boldsymbol{b}\quad 其中\quad A=\begin{bmatrix}1\\1\\1\end{bmatrix}\quad 与\quad V=\begin{bmatrix}\sigma_1^{\,2}&0&0\\0&\sigma_2^{\,2}&0\\0&0&\sigma_3^{\,2}\end{bmatrix}$$

方差是 $\sigma_1^2=1/9$ 与 $\sigma_2^2=1/4$ 与 $\sigma_3^2=1$。当量测值 b_3 比 b_2 与 b_1 不可靠，你越来越紧张。这 3 个量测值都包含某些信息，所以他们都进入最佳(估测) $\hat{\boldsymbol{x}}$：

$$V^{-1/2}A\boldsymbol{x}=V^{-1/2}\boldsymbol{b}\quad 是\quad \begin{matrix}3x=3b_1\\2x=2b_2\\1x=1b_3\end{matrix}\quad 得到\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{A}\,\hat{\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{b}$$

$$\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}9&&\\&4&\\&&1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}\hat{\boldsymbol{x}}=\begin{bmatrix}1&1&1\end{bmatrix}\begin{bmatrix}9&&\\&4&\\&&1\end{bmatrix}\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

$$\hat{\boldsymbol{x}}=\frac{9b_1+4b_2+b_3}{14}\quad \textbf{是 } \boldsymbol{b_1, b_2, b_3} \textbf{ 的加权平均}$$

方差σ_1最小，所以最多的加权放在b_1。$\hat{\boldsymbol{x}}$的方差有美丽的公式

$W = (A^{\mathrm{T}}V^{-1}A)^{-1} = 1/14$：

$$\hat{\boldsymbol{x}}\text{ 的方差 } \left(\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & & \\ & 4 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} = \frac{1}{14} \text{ 小于 } \frac{1}{9}$$

高斯的蓝色定理(网站有证明)说明我们的$\hat{\boldsymbol{x}} = L\boldsymbol{b}$是$A\boldsymbol{x} = \boldsymbol{b}$的解的最佳线性无偏见估测，任何其他无偏见选择$\boldsymbol{x}^* = L^*\boldsymbol{b}$的方差都会比$\hat{\boldsymbol{x}}$的方差来的大。所有的无偏见选择都有$L^*A = I$，所以一个确切的$A\boldsymbol{x} = \boldsymbol{b}$会产生正确的答案$\boldsymbol{x} = L^*\boldsymbol{b} = L^*A\boldsymbol{x}$。

注解： 我必须加上刚开始不要极小化误差平方的理由。第一个理由：这个$\hat{\boldsymbol{x}}$通常有很多小的分量，小数的平方非常小，当我们极小化时他们出现了。理解稀疏矩阵比较容易—只有一些非零数。统计学家通常偏好极小化**非平方误差：$|(\boldsymbol{b} - A\boldsymbol{x})_i|$的总和**，这个误差量测是$L^1$不是$L^2$。因为绝对值，$\hat{\boldsymbol{x}}$的方程式变成非线性(实际是片段线性。)

快速的新演绎法非常快速的计算稀疏$\hat{\boldsymbol{x}}$，未来可能属于L^1。

卡尔曼滤波器

“卡尔曼滤波器”是动态最小平方的最伟大演绎法，**动态**这个词表示持续有新的量测值$\boldsymbol{b}_k$进来，所以最佳估测$\hat{\boldsymbol{x}}$持续改变(基于所有的$\boldsymbol{b}_0, ..., \boldsymbol{b}_k$)。不止如此，矩阵$A$也在改变，所以$\hat{\boldsymbol{x}}_2$会是我们的最新解$\boldsymbol{x}_k$的最佳最小平方估测，其中$\boldsymbol{x}_k$是**直到时间 2 的历史最新解，这项历史包含观察方程式与更新方程式(状态方程式)**：

$$A_0\boldsymbol{x}_0 = \boldsymbol{b}_0 \quad \boldsymbol{x}_1 = F_0\boldsymbol{x}_0 \quad A_1\boldsymbol{x}_1 = \boldsymbol{b}_1 \quad \boldsymbol{x}_2 = F_1\boldsymbol{x}_1 \quad A_2\boldsymbol{x}_2 = \boldsymbol{b}_2 \tag{14}$$

卡尔曼的概念是一次引进一个方程式，每个方程式存在误差。基于每个新方程式，我们对目前的$\boldsymbol{x}_k$更新成最佳估测$\hat{\boldsymbol{x}}_k$。但是历史不能忘记！新的估测$\hat{\boldsymbol{x}}_k$使用了过去所有的观察$\boldsymbol{b}_0$到$\boldsymbol{b}_{k-1}$以及所有的状态方程式$\boldsymbol{x}_{\text{new}} = F_{\text{old}}\boldsymbol{x}_{\text{old}}$，这是一个大型与成长的问题。

更重要的一点，每个最小平方方程式使用对应每个$\boldsymbol{b}_k$的误差的协方差矩阵V进行加权，甚至存在一个对应更新方程式$\boldsymbol{x}_{k+1} = F_k\boldsymbol{x}_k$的误差的协方差矩阵$C_k$，则最佳$\hat{\boldsymbol{x}}_2$会与$\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2$与$V_0, V_1, V_2$与$C_1, C_2$有关。写出$\hat{\boldsymbol{x}}_k$的优质方法是写成前面的$\hat{\boldsymbol{x}}_{k-1}$的更新。

让我聚焦在一个没有矩阵F_k与协方差C_k的简化问题，我们每一步都在估测相同的真实$\boldsymbol{x}$。我们如何从$\hat{\boldsymbol{x}}_0$得到$\hat{\boldsymbol{x}}_1$?

旧的 $A_0\boldsymbol{x}_0 = \boldsymbol{b}_0$得到加权方程式 $A_0^{\mathrm{T}}V_0^{-1}A_0\hat{\boldsymbol{x}}_0 = A_0^{\mathrm{T}}V_0^{-1}\boldsymbol{b}_0$ (15)

新的 $\begin{bmatrix} A_0 \\ A_1 \end{bmatrix}\hat{\boldsymbol{x}}_1 = \begin{bmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b}_1 \end{bmatrix}$得到下列$\hat{\boldsymbol{x}}_1$的加权方程式：

请尊重版权与译者的劳动成果，侵权必究!

$$[A_0^{\mathrm{T}} \quad A_1^{\mathrm{T}}]\begin{bmatrix} V_0^{-1} & \\ & V_1^{-1} \end{bmatrix}\begin{bmatrix} A_0 \\ A_1 \end{bmatrix}\hat{\boldsymbol{x}}_1 = [A_0^{\mathrm{T}} \quad A_1^{\mathrm{T}}]\begin{bmatrix} V_0^{-1} & \\ & V_1^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b}_1 \end{bmatrix} \tag{16}$$

是的，我们求解了新问题也忘记了旧问题。但是旧的解 $\hat{\boldsymbol{x}}_0$ 需要处理以便在 $\hat{\boldsymbol{x}}_1$ 中使用，我们寻找的就是 $\hat{\boldsymbol{x}}_0$ 的**更新**：

卡尔曼更新从 $\hat{\boldsymbol{x}}_0$ 到 $\hat{\boldsymbol{x}}_1$　　$\hat{\boldsymbol{x}}_1 = \hat{\boldsymbol{x}}_0 + K_1(\boldsymbol{b}_1 - A_1\hat{\boldsymbol{x}}_0)$　　(17)

这项更新修正是旧状态是 $\hat{\boldsymbol{x}}_0$ 与新量测 $\boldsymbol{b}_1$ 之间的不适配 $\boldsymbol{b}_1 - A_1\hat{\boldsymbol{x}}_0$——用**卡尔曼增益矩阵 $\boldsymbol{K_1}$** 去乘。K_1 的公式来自比较(15)的解 $\hat{\boldsymbol{x}}_0$ 与(16)的解 $\hat{\boldsymbol{x}}_1$，我们基于新资料 $\boldsymbol{b}_1$ 把 $\hat{\boldsymbol{x}}_0$ 更新到 $\hat{\boldsymbol{x}}_1$，**我们也把协方差矩阵 $\boldsymbol{W_0}$ 更新到 $\boldsymbol{W_1}$**。记得来自方程式(13)有 $W_0 = (A_0^{\mathrm{T}}V_0^{-1}A_0)^{-1}$。更新它的逆矩阵到 W_1^{-1}：

$\hat{\boldsymbol{x}}_1$ 误差的协方差 $\boldsymbol{W}_1$　　$\boldsymbol{W}_1^{-1} = \boldsymbol{W}_0^{-1} + \boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{V}_1^{-1}\boldsymbol{A}_1$　　(18)

卡尔曼增益矩阵 $\boldsymbol{K}_1$　　$\boldsymbol{K}_1 = \boldsymbol{W}_1\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{V}_1^{-1}$　　(19)

这是卡尔曼滤波器的中心，注意 W_k 的重要性，这些矩阵量测整个程序的可靠度，其中向量 $\hat{\boldsymbol{x}}_k$ 基于特定的量测值 $\boldsymbol{b}_0$ 到 $\boldsymbol{b}_k$ 来估测目前的状态。

当状态 $\hat{\boldsymbol{x}}_k$ 也在变化(基于矩阵 F_k)，整章与整本书都在说明动态卡尔曼滤波器。在使用新资料 $\boldsymbol{b}$ 修正之后，利用 F 可以预测 $\hat{\boldsymbol{x}}_k$，最好停止这里。

本页是有关**迭代最小平方**：加入新资料 $\boldsymbol{b}_k$ 与更新 $\hat{\boldsymbol{x}}$ 及 W_k：基于所有资料的最佳当前估测，以及它的协方差矩阵。

问题集 12.3

1　相同变数 x 的两个量测得到 $\boldsymbol{x} = b_1$ 与 $\boldsymbol{x} = b_2$，假设平均值是零与方差是 σ_1^2 与 σ_2^2，误差是无关：V 是对角且对角单元是 σ_1^2 与 σ_2^2。写出如同 $A\boldsymbol{x} = \boldsymbol{b}$ 的两个方程式 (A 是 2×1)。如同课本范例 1，基于 b_1 与 b_2，求出最佳估测 $\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{x}} = \frac{b_1/\sigma_1^{\,2} + b_2/\sigma_2^{\,2}}{1/\sigma_1^{\,2} + 1/\sigma_2^{\,2}} \qquad \mathrm{E}\left[\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^{\mathrm{T}}\right] = \left(\frac{1}{\sigma_1^{\,2}} + \frac{1}{\sigma_2^{\,2}}\right)^{-1}$$

2　(a)　问题 1 中，假设第二个量测值 b_2 变成**超级精确**而且它的方差 $\sigma_2 \to 0$，当 σ_2 抵达零时，最佳估测 $\hat{\boldsymbol{x}}$ 为何？

(b)　相反的情况是 $\sigma_2 \to \infty$ 而且没有 b_2 的信息，现在基于 b_1 与 b_2 的最佳估测 $\hat{\boldsymbol{x}}$ 为何？

3 若 x 与 y 无关，概率分别是 $p_1(x)$与 $p_2(y)$，则 $p(x, y) = p_1(x)p_2(y)$。借由分离双重积分成为单一积分(–∞到∞)，证明

$$\iint p(x,y)\,dx\,dy = \mathbf{1} \qquad 与 \qquad \iint (x+y)p(x,y)\,dx\,dy = \boldsymbol{m}_1 + \boldsymbol{m}_2$$

4 继续问题 3，对于无关的 x 与 y 证明 $p(x, y) = p_1(x)p_2(y)$有

$$\iint (x-m_1)^2\,p(x,y)\,dx\,dy = {\sigma_1}^2 \qquad \iint (x-m_1)(y-m_2)p(x,y)\,dx\,dy = 0$$

所以 2×2 协方差矩阵 V 是对角而且对角单元是_____。

5 证明 2×2 协方差矩阵 V 的逆矩阵是

$$V^{-1} = \begin{bmatrix} {\sigma_1}^2 & \sigma_{12} \\ \sigma_{12} & {\sigma_2}^2 \end{bmatrix}^{-1} = \frac{1}{1-\rho}\begin{bmatrix} 1/{\sigma_1}^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/{\sigma_2}^2 \end{bmatrix} \quad 相关因子是 \rho = \sigma_{12}/\sigma_1\sigma_2$$

这个产生双变数高斯的指数 $-(\boldsymbol{x}-m)^{\mathrm{T}}\,V^{-1}\,(\boldsymbol{x}-m)$。

6 假设 $\hat{\boldsymbol{x}}_k$ 是 $b_1, ..., b_k$ 的平均值，新量测值 b_{k+1} 抵达，我们想要新的平均值 $\hat{\boldsymbol{x}}_{k+1}$。卡尔曼更新方程式(17)是

$$\textbf{新平均} \qquad \hat{\boldsymbol{x}}_{k+1} = \hat{\boldsymbol{x}}_k + \frac{1}{k+1}(b_{k+1} - \hat{\boldsymbol{x}}_k)$$

验证 $\hat{\boldsymbol{x}}_{k+1}$ 是 $b_1, ..., b_{k+1}$ 的正确平均值。

7 也验证更新方程式(18)，其中平均值 $\hat{\boldsymbol{x}}$ 的方差是 $W_{k+1} = \sigma^2/(k+1)$，我们假设 $W_k = \sigma^2/(k)$而且 b_{k+1} 有方差 $V = \sigma^2$。

8 **(稳定模型)** 问题 6-7 是**静态**最小平方，所有的样本平均值是 $\hat{\boldsymbol{x}}_k$ 是相同 x 的估测值。要使得卡尔曼滤波器变成**动态**，也包含一个状态方程式 $\hat{\boldsymbol{x}}_{k+1} = F\boldsymbol{x}_k$，其中本身的误差方差是 s^2。动态最小平方问题在 k 最佳时允许 x **移动**：

$$\begin{bmatrix} 1 & & \\ -F & 1 & \\ & 1 & \end{bmatrix}\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ 0 \\ b_1 \end{bmatrix} \quad 其中方差是 \begin{bmatrix} \sigma^2 \\ s^2 \\ \sigma^2 \end{bmatrix}$$

当 $F = 1$，这 3 个方程式的两侧同时除以σ, s, σ，利用最小平方求出 $\hat{\boldsymbol{x}}_0$ 与 $\hat{\boldsymbol{x}}_1$，使得新进的 b_1 有更多加权。卡尔曼滤波器在全球定位演绎法中发表(Borre 与 Strang，卫斯理-康桥出版社)。

请尊重版权与译者的劳动成果，侵权必究!

来自 A 的改变导致的 A^{-1} 的改变

最后一页把本书的开头(逆矩阵与秩 1 矩阵)与本书的结尾(动态最小平方与滤波器)做个联结，从基本公式开始：

$$M = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \text{ 的逆矩阵是 } M^{-1} = \boldsymbol{I} + \frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{\boldsymbol{1} - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}$$

快速证明是 $MM^{-1} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + (\boldsymbol{1} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\dfrac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{\boldsymbol{1} - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} = \boldsymbol{I}$。

如果 $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} = 1$ 则 M 不可逆($M\boldsymbol{u} = \boldsymbol{0}$)。这里 $\boldsymbol{v}^{\mathrm{T}} = \boldsymbol{u}^{\mathrm{T}} = [1 \quad 1 \quad 1]$：

范例 $M = I - \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ 的逆矩阵是 $M^{-1} = I + \dfrac{1}{\boldsymbol{1} - \boldsymbol{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

但是我们不是一直从单位矩阵开始，很多应用需要逆反 $\boldsymbol{M} = \boldsymbol{A} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$。在我们求解 $A\boldsymbol{x} = \boldsymbol{b}$ 之后，我们期望一个秩 1 的改变得到 $M\boldsymbol{y} = \boldsymbol{b}$。前述的除以 $1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$ 会变成除以 $c = 1 - \boldsymbol{v}^{\mathrm{T}}A^{-1}\boldsymbol{u} = 1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{z}$：

步骤 1　求解 $A\boldsymbol{z} = \boldsymbol{u}$ 然后计算 $1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{z}$。

步骤 2　若 $c \neq 0$ 则 $M^{-1}\boldsymbol{b}$ 是 $\boldsymbol{y} = \boldsymbol{x} + \dfrac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{c}\boldsymbol{z}$

假设 A 很容易处理，A 可能已经应用消元法变成 LU，则这个 Sherman-Woodbury-Morrison 公式是求解 $M\boldsymbol{y} = \boldsymbol{b}$ 是最快速方法。这里有三个问题来结束本书!

9　当 $A = I$ 与 $\boldsymbol{u}^{\mathrm{T}} = \boldsymbol{v}^{\mathrm{T}} = [1 \quad 2 \quad 3]$与 $\boldsymbol{b}^{\mathrm{T}} = [2 \quad 1 \quad 4]$，使用步骤 1-2 求 $\boldsymbol{y}$。

10　“更新公式”的步骤 2 宣称 $M\boldsymbol{y} = (A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})\left(\boldsymbol{x} + \dfrac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{c}\boldsymbol{z}\right) = \boldsymbol{b}$，简化成为

$\dfrac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}}{c}\left(1 - c - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{z}\right) = \boldsymbol{0}$。由于 $c = 1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{z}$，前式成立。

11　当 A 有一个新行 $\boldsymbol{v}^{\mathrm{T}}$，最小平方方程式的 $A^{\mathrm{T}}A$ 变成 M：

$M = [A^{\mathrm{T}} \quad \boldsymbol{v}] \begin{bmatrix} A \\ \boldsymbol{v}^{\mathrm{T}} \end{bmatrix} = A^{\mathrm{T}}A + \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} = A^{\mathrm{T}}A$ 中的秩 1 变化。

为什么这个乘法是正确？更新的 $\hat{\boldsymbol{x}}_{\text{new}}$ 来自步骤 1 与 2。这里有 M^{-1} 的 4 个公式作为参考，前两个是当变化是 $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ 时，在上方已经给出。公式 3 与 4 超越秩 1，允许矩阵 U, V, W。

1	$M = I - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$	$M^{-1} = I + \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}/(1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})$ **(秩 1 变化)**
2	$M = A - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$	$M^{-1} = A^{-1} + A^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}A^{-1}/(1 - \boldsymbol{v}^{\mathrm{T}}A^{-1}\boldsymbol{u})$
3	$M = I - UV$	$M^{-1} = I_n + U(\boldsymbol{I_m} - \boldsymbol{VU})^{-1}V$
4	$M = A - \boldsymbol{UW^{-1}V}$	$M^{-1} = A^{-1} + A^{-1}U(\boldsymbol{W} - \boldsymbol{VA^{-1}U})^{-1}VA^{-1}$

公式 **4** 是工程中的“矩阵逆反辅助定理”，直到现在才看到！卡尔曼滤波器在求解方块三对角系统时，每个步骤都用到公式 **4**。

矩阵分解

1. $\boldsymbol{A=LU}$ = (1's 在对角线的下三角 L) (枢轴在对角线的上三角 U)

   要求：高斯消元法从方形 A 简化到 U 没有行交换。

2. $\boldsymbol{A=LDU}$ = (1's 在对角线的下三角 L) (枢轴矩阵 D 是对角) (1's 在对角线的上三角 U)

   要求：没有行交换，提出 D 中的枢轴留在 U 的对角线。若 A 对称则 U 是 L^{T} 以及 $\boldsymbol{A=LDL^{\mathrm{T}}}$。

3. $\boldsymbol{PA=LU}$ (排列矩阵 P 是为了避免枢轴位置的零)

   要求：A 可逆，则 P, L, U 可逆， P 事先做完 A 的所有行交换，允许正常的 LU。交替：$A_1 = L_1P_1U_1$。

4. $\boldsymbol{EA=R}$ ($m\times m$ 可逆 E) (任意 $m\times n$ 矩阵 A) = rref(A)。

   要求：无！简化行阶梯形式 R 有 r 个枢轴行与 r 个枢轴列，包含单位矩阵。E 的最后 $m-r$ 个行是 A 的左零空间的一组基底；他们乘 A 得到 R 中的 $m-r$ 个零行。E^{-1} 的前 r 个列是 A 的列空间的一组基底。

5. $\boldsymbol{S=CC^{\mathrm{T}}}$ = (下三角) (上三角)，其中 $\sqrt{D}$ 在两个对角线。

   要求：S 是对称正定(D 全部的 n 个枢轴都是正数)，**乔列斯基** C = chol(S)有 $C^{\mathrm{T}} = L\sqrt{D}$，所以 $S = C^{\mathrm{T}}C = LDL^{\mathrm{T}}$。

6. $\boldsymbol{A=QR}$ = (Q 有正交单位列) (上三角 R)

   要求：A 有无关列，这些列在 Q 利用格莱姆-施密特或豪斯霍德程序正交单位化。若 A 是方形则 $Q^{-1} = Q^{\mathrm{T}}$。

7. $\boldsymbol{A=X\Lambda X^{-1}}$ = (X 有固有向量) (Λ有固有值) (X^{-1} 有左固有向量)

   要求：A 必须有 n 个无关固有向量。

8. $\boldsymbol{S=Q\Lambda Q^{\mathrm{T}}}$= (正交矩阵 Q) (实数固有值矩阵Λ) (Q^{T} 是 Q^{-1})

   要求：S 是实数与对称：$S^{\mathrm{T}} = S$，这是光谱定理。

请尊重版权与译者的劳动成果，侵权必究!

9. $\boldsymbol{A = B\,J\,B^{-1}}$ = (B 有广义固有向量) (J 有乔丹方块) (B^{-1})

要求：A 是任意方形矩阵，这个乔丹形式有对应 A 的每个固有值的方块，每个方块只有一个固有值。

10. $\boldsymbol{A = U\,\Sigma\,V^{T}}$ = (正交 U 是 $m\times m$) ($m\times n$ 奇异值矩阵，σ_1, ..., σ_r 在对角线) (正交 V 是 $n\times n$)

要求：无。这个奇异值分解(SVD)有 AA^{T} 的固有向量在 U 以及 $A^{\mathrm{T}}A$ 的固有向量在 V；$\sigma_i = \sqrt{\lambda_i\left(A^{\mathrm{T}}A\right)} = \sqrt{\lambda_i\left(AA^{\mathrm{T}}\right)}$。这些奇异值是$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$，使用列-行乘法：

$$\boldsymbol{A = U\,\Sigma\,V^{T}} = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + ... + \sigma_r\boldsymbol{u}_r\boldsymbol{v}_r^{\mathrm{T}}$$

若 S 是对称正定，则 $U = V = Q$ 与$\Sigma = \Lambda$以及 $S = Q\Lambda Q^{\mathrm{T}}$。

11. $\boldsymbol{A^{+} = U\,\Sigma^{+}\,V^{T}}$ = (正交 $n\times n$) ($n\times m$ Σ的伪逆，$1/\sigma_1$, ..., $1/\sigma_r$ 在对角线) (正交 $m\times m$)

要求：无。伪逆 A^{+}有 $A^{+}A =$ 对 A 的行空间的投影，以及 AA^{+} =对 A 的列空间的投影。若 A 可逆则 $A^{+} = A^{-1}$。$A\boldsymbol{x} = \boldsymbol{b}$ 的最短最小平方解是 $\boldsymbol{x}^{+} = A^{+}\boldsymbol{b}$，这是 $A^{\mathrm{T}}A\boldsymbol{x}^{+} = A^{\mathrm{T}}\boldsymbol{b}$ 的解。

12. $\boldsymbol{A = QS}$ = (正交矩阵 Q) (对称正定矩阵 S)

要求：A 可逆。这个极分解有 $S^2 = A^{\mathrm{T}}A$，若 A 是奇异则因子 S 是半定，逆反极分解 $A = KQ$ 有 $K^2 = AA^{\mathrm{T}}$，两者都有来自 SVD 的 $Q = UV^{\mathrm{T}}$。

13. $\boldsymbol{A = U\,\Lambda\,U^{-1}}$ = (幺正 U) (固有值矩阵Λ) (U^{-1} 就是$U^{\mathrm{H}} = \overline{U}^{\mathrm{T}}$)

要求：A 是正态：$A^{\mathrm{H}}A = AA^{\mathrm{H}}$，它的正交单位(可能是复数)固有向量是 U 的列。除非 $S = S^{\mathrm{H}}$，λ是复数：厄米矩阵。

14. $\boldsymbol{A = Q\,T\,Q^{-1}}$ = (幺正 Q) (三角 T，λ's 在对角线) ($Q^{-1} = Q^{\mathrm{H}}$)

要求： 任意方形 A 的舒尔三角化，存在矩阵 Q 有正交单位列使得 $Q^{-1}AQ$ 变成三角形：段落 6.4。

15. $\boldsymbol{F_n} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_{n/2} & \\ & F_{n/2} \end{bmatrix} \begin{bmatrix} \text{偶 - 奇} \\ \text{排列} \end{bmatrix}$ = 迭代 **FFT** 的一步

要求：F_n = 傅里叶矩阵有单元 w^{jk} 其中 $w^n = 1$：$F_n\overline{F}_n = nI$。D 有 1, w, ..., $w^{n/2-1}$ 在对角线。$n = 2^l$ 时，快速傅里叶转换会计算 $F_n\boldsymbol{x}$，只需要来自 D's 的 l 阶段的 $nl/2 = (n\log_2 n)/2$ 次乘法。

线性代数的六个伟大定理

维度定理	对应一个向量空间的所有基底都有相同的向量个数。
计数定理	列空间的维度 + 零空间的维度 = 列的个数。
秩定理	列空间的维度 = 行空间的维度，这是秩。
基础定理	A 的行空间与零空间是 $\mathbf{R}^n$ 中的正交互补。
SVD	存在正交单位基底($\boldsymbol{v}$'s 与 $\boldsymbol{u}$'s 对应行与列空间)使得 $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$。
光谱定理	若 $A^{\mathrm{T}} = A$，存在正交单位 $\boldsymbol{q}$'s 使得 $A\boldsymbol{q}_i = \lambda_i \boldsymbol{q}_i$ 与 $A = Q\Lambda Q^{\mathrm{T}}$。

概括线性代数（ 矩阵 A 是 $n \times n$ ）

非奇异	奇异
A 是可逆	A 不可逆
列是无关	列是相关
行是无关	行是相关
行列式不是零	行列式是零
$A\boldsymbol{x} = \boldsymbol{0}$ 有一个解 $\boldsymbol{x} = \boldsymbol{0}$	$A\boldsymbol{x} = \boldsymbol{0}$ 有无限多解
$A\boldsymbol{x} = \boldsymbol{b}$ 有一个解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$	$A\boldsymbol{x} = \boldsymbol{b}$ 无解或无限多解
A 有 n 个(非零)枢轴	A 有 $r < n$ 个枢轴
A 有满秩 $r = n$	A 的秩 $r < n$
简化行阶梯形式是 $R = I$	R 最少有一个零行
列空间是全部的 $\mathbf{R}^n$	列空间的维度 $r < n$
行空间是全部的 $\mathbf{R}^n$	行空间的维度 $r < n$
所有的固有值都是非零	零是 A 的一个固有值
$A^{\mathrm{T}}A$ 是对称正定	$A^{\mathrm{T}}A$ 只有半定
A 有 n 个(正数)奇异值	A 有 $r < n$ 个奇异值

请尊重版权与译者的劳动成果，侵权必究!