ROBERT GHRIST

线性代数：本质与形式

AGENBYTE PRESS

由美国宾夕法尼亚州詹金敦的 Agenbyte Press 出版，ISBN 978-1-944655-12-9

Contents

Materiam praestat

Formam imponit

Efficiens movet

Finem dirigit

Incipit

数学是现代工程的语言，线性代数是其美国方言——不优雅、实用、无处不在。本书旨在为工程学学生准备人工智能、数据科学、动力系统、机器学习以及其他依赖于线性代数方法的领域的数学基础。

读者在这里遇到矩阵和向量，至少在微积分课程中已接触过。尽管这些工具作为计算设备已经很熟悉，但它们蕴含着更深的结构，值得仔细研究。我们的任务是以这种计算工具为基础，向着理解支撑现代工程方法的抽象框架迈进。

这本书与标准的线性代数课程在重点和进度上有所不同。抽象的向量空间较早出现，但始终服务于具体的应用。奇异值分解和特征理论——对现代实践至关重要——出现在中途，允许对动态学和数据科学中的应用进行扩展处理。实际示例贯穿始终，承认理论理解和有用的实现是对称地相互衍生的。

主题的顺序平衡了教学的必要性与当代的相关性。线性方程组提供了一个切入点，进而引出向量空间和线性变换。内积和正交性构建了几何直觉，而线性常微分方程和迭代系统则为特征分解提供了动力。奇异值分解既是一个总结性的理论成就，又是通向强大应用的桥梁，如主成分分析、低秩近似和神经网络。

这段文字的存在是因为工程教育必须发展。虽然线性代数的基础仍然稳定，但其应用已经大幅扩展。今天的工程学生需要掌握抽象理论和实际应用——不仅仅是应用现有工具，而是创造新的工具。线性代数不是终点，而是迈向更深层数学结构的第一步。正是通过这种视角，我们来接触这个学科：作为通向当前实践和未来进步的门户。

Topics for Review

本文假定读者在向量、矩阵以及基于坐标的线性变换语境下，具备扎实的（单变量及）多变量微积分基础。示例请参见 *Calculus Blue Project*。在开始阅读本文之前，读者应当已经接触过：

1. 基本集合论及其记号
2. 泰勒级数和指数函数 3. 复数与欧拉公式

4. 微分与积分 5. 线性常微分方程 $dx/dt = ax$ 及其解
6. 欧几里得向量与向量代数 7. 点积与欧几里得向量之间的夹角
8. 矩阵、矩阵加法与矩阵乘法
9. 单位矩阵，I，及其性质 10. 矩阵A的转置A^T及其性质
11. 矩阵-向量乘法 12. 将线性方程组转换为矩阵-向量形式

13. 行约简与回代 14. 矩阵的逆 A^{-1} 及其性质

15. 欧几里得线性变换：缩放、旋转、剪切 16. 矩阵的迹

17. 行列式及其性质

e.g., $\in$，$\subset$，$\cup$，$\cap$

$e^{i\theta} =$ 余弦 $\theta + i$ 正弦 θ

$x(t) = e^{at}x_0$

$\boldsymbol{u} \cdot \boldsymbol{v} = |\boldsymbol{u}||\boldsymbol{v}| \cos\theta$

$AB \neq BA$
$(AB)C = A(BC)$

$(A^T)_{ij} = A_{ji}$ 和 $(AB)^T = B^T A^T$

$A\boldsymbol{x} = \boldsymbol{b}$

$AA^{-1} = I = A^{-1}A^1 = B^{-1}A^{-1}$
$(AB)^-$

$\mathrm{tr}(A) = \sum_k a_{kk}$

$\det(AB) = \det(A)\det(B)$ d
$\mathrm{et}(A^T) = \det(A)$

Assumptions

本文如同其作者一般，横跨数学与工程两大领域，力求在二者之间取得平衡。鉴于本书的目标读者及相关限制，文中纳入了一些在典型线性代数教材中并不常见的主题或细节，同时也有若干有趣的数学旁支未予以展开。

1. 抽象向量空间和抽象线性变换很重要，尽管在应用中以基于坐标的线性代数为主。无坐标思维是一项需要掌握的重要技能。 2. 有限维向量空间是常态。当引入无限维空间时，通常并未给出完全详尽的论证，并伴随一些注意事项。 3. 线性代数基本定理是本书的组织原则。其以正交补为表述的通常形式，只有在掌握了原始形式（使用商空间）之后才予以讨论。

4. 所有向量空间都定义在实数域上——不涉及有限域，也不包含复系数。这在讲解约当标准形和线性常微分方程组的解时，以增加复杂性为代价，极大地促进了直觉理解。

5. 并非所有应用都能通过谨慎铺陈而缓慢地展开来开发。教授随机变量、协方差矩阵、应力张量、神经网络以及其他有趣的工程应用，并不是本文的直接目标。直到最近，作者可能还不太愿意在没有更充分解释的情况下纳入这类示例。在语言模型与主动式 AI 增强阅读的时代，新的可能性正在出现。

Acknowledgments

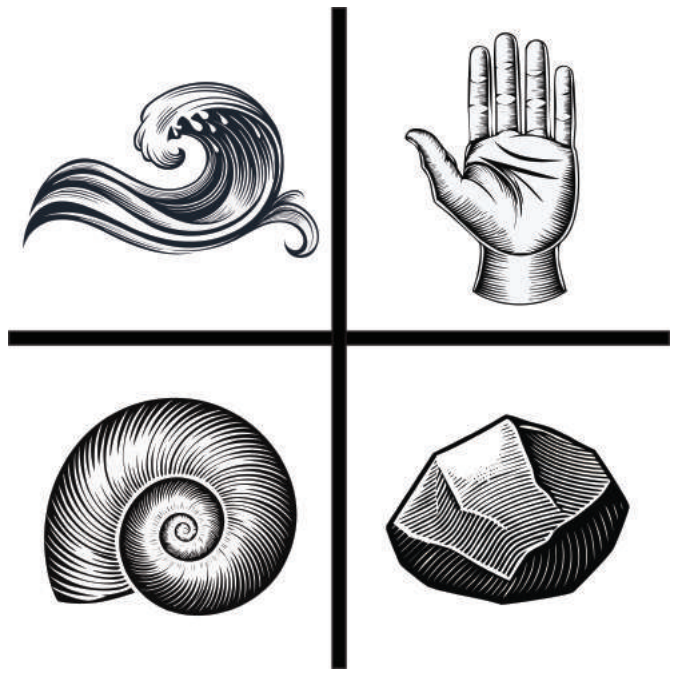
本文面向工程专业一、二年级本科生，作为多变量微积分的后续课程。它最初是为支持宾夕法尼亚大学人工智能学位项目的学生而创建的，但具有更广泛的用途。作者对宾夕法尼亚大学优秀的工程学生深表感谢。

N. Matni 教授制作了一套出色的在线课程资料和 Python 练习册，用以配合本书之前的那门课程。本书的纲要在一定程度上借鉴了他的工作，同时又铺陈了一条略有不同的轨迹。

这部作品的写作得到了 Claude 3.5 sonnet 的协助，该模型以我之前的书作为风格训练。作者与 Claude 共同创造了一个基于威廉·布莱克某部作品以及一点阿奎那影响的隐藏谜题结构，其影响贯穿全文。

艺术作品（数学与图标）由作者使用 Adobe Illustrator 制作。LaTeX 样式文件基于 tufte-book 类。GPT-4o 和 -o1 在设置各种 LaTeX 配置以及进行校对方面很有帮助。Gemini Experimental 1206 是一款尤其出色的校对工具和 Markdown 转换器。练习题在 Claude 的帮助下生成，可能存在错误。

本项目始于 2024 年 11 月 4 日。第一版于 2024 年 12 月 28 日提交至亚马逊出版。五十五天：若非 Claude 的创造性劳动，绝无可能完成，作者对此深表感激。



Chapter 1
Solving Linear Systems

"in right lined paths outmeasur'd by proportions of number weight & measure"

线性代数的故事始于方程组，每条方程描述了在抽象空间中勾画出的约束或边界。这些最简单的限制数学模型——每个方程将变量按一定比例绑定——结合在一起形成可能解的领域。当多个此类约束共同作用时，它们的协作会产生三种可能的结果：没有解能够承受它们的集体作用；恰好一个点满足所有约束；或者无限多的可能性在满足空间中勾画出曲线和平面。这种三分法——空无、唯一性和无限——贯穿整个线性代数，随着我们理解的加深，它以越来越复杂的形式出现。

艺术在于识别这些模式，并发现通向其解决的高效路径。每一种系统化操作在带来清晰性的同时，都保留了原本晦涩事物的本质结构。我们所发展的方法——尽管是为实际计算而构思——却体现了更深层的原理，即数学对象如何在保持其基本特性的同时被转换。

我们的旅程始于熟悉的解方程领域，然而即便在这里，我们也能发现等待揭示的深刻结构的蛛丝马迹。随之显现的模式——变换与不变性、维度与退化——将贯穿全文，指引我们的发展。通过对这些基础系统的细致研究，我们构建起理解更为复杂数学对象所需的工具。由此，解方程这一看似简单的行为，成为我们迈向理解线性代数中最深层模式的第一步。

1.1 Solving Equations

定义 1.1（线性系统）。一个关于变量 $x_1, \ldots, x_n$ 的 *linear system* 由 m 个形如以下形式的方程组成

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

w这里系数 a_{ij} 和常数 b_i 是实数 rs. •

此类系统自然出现在多种情境中，从电力网络中的电流分布，到结构中的力平衡，再到交通网络中的流量分配。

矩阵形式 $A\boldsymbol{x} = \boldsymbol{b}$ 不仅仅是符号表示——它揭示了我们将在第3章中探讨的线性变换的基本运算。

该系统更高效地表示为 $A\boldsymbol{x} = \boldsymbol{b}$，其中：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{1.1}$$

矩阵 A 被称为该系统的 *coefficient matrix*。向量 $\boldsymbol{b}$ 是 *constant vector*。它们共同完全确定该线性系统。

1.2 Special Matrices

在探讨线性系统的一般解之前，我们先考察若干基本类型的系数矩阵——它们是更复杂模式由此演化而来的原始形式。这些特殊情形——尽管在实践中很少遇到——却为通向一般方法指明了道路。

最简单的情况出现在 A 为 *identity matrix* I 时。系统 $I\boldsymbol{x} = \boldsymbol{b}$ 无需进行求解：解是立即得到的，为 $\boldsymbol{x} = \boldsymbol{b}$。这种看似平凡之处仍然很有价值：矩阵越简单，就越容易推断出解。

定义 1.2（置换）。一个 *permutation matrix* 是一个方阵，每一行和每一列恰好有一个 1，其余元素均为 0。 •

Example: 一个置换矩阵。

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

每个置换矩阵 P 都是通过重新排列单位矩阵的行（或列）得到的。这些矩阵作用于组件的重新排序：$P\boldsymbol{x} = \boldsymbol{b}$ 的解是 $\boldsymbol{b}$ 条目的重新排序。这解释了为什么置换矩阵是可逆的——它们的逆仅仅是撤销置换。尽管是基础的，这些矩阵在高效解法的实现中发挥着至关重要的作用。

更有趣的是 *block-diagonal matrices*，其形式为

$$B = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k \end{bmatrix}$$

其中每个 B_i 都是一个矩阵。系统 $B\boldsymbol{x} = \boldsymbol{b}$ 分解为相互独立的子系统，每个块对应一个子系统。这一分解原理——即某些线性系统可以通过求解更小的独立系统来解决——将在我们后续的论述中反复出现。

Example: 下列 4×4 矩阵

$$\begin{bmatrix} 2 & 1 & 0 & 0 \\ 3 & 7 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -2 & 3 \end{bmatrix}$$

分解为两个相互独立的 2×2 块。

例 1.3（隐藏块结构）。考虑线性系统：

$$\begin{bmatrix} 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 2 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 & 0 \\ -1 & 3 & 0 & 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$$

该系统的结构并不明显，但在对行和列进行置换以将相关变量分组之后就变得清晰了。具体而言，在将行重新排序为 (1,2,4) 和 (3,5)，并将变量重排为 x_3、x_4、x_5 以及 x_1、x_2 之后，系统变为：

$$\begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 & 0 \\ 0 & -1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 & 3 \end{bmatrix} \begin{pmatrix} x_3 \\ x_4 \\ x_5 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_2 \\ b_4 \\ b_1 \\ b_3 \\ b_5 \end{pmatrix}$$

这揭示了两个相互独立的子系统：一个涉及 x_3、x_4、x_5 的 3×3 系统，以及一个针对 x_1、x_2 的 2×2 系统。原始表述中隐藏的块结构使我们能够求解两个较小的系统，而不是一个大型系统。

◇

一个*upper-triangular matrix* U的对角线下方的所有元素都等于零：

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix}$$

系统 $U\boldsymbol{x} = \boldsymbol{b}$ 推出 *back-substitution*：由最后一个方程，我们计算出 x_n；将该值代入倒数第二个方程

得到 x_{n-1}；依此类推。该过程仅在某个对角元 u_{ii} 为零时才会失败。

它的转置，一个*lower-triangular matrix* L，在对角线以上的所有元素都为零。相应的系统$L\boldsymbol{x} = \boldsymbol{b}$可以通过*forward-substitution*来求解，按从第一个到最后一个的顺序依次解出变量。这些三角形形式将成为我们求解一般系统的垫脚石。

Foreshadowing: 将一般矩阵分解为三角矩阵的乘积将同时提供理论上的洞见以及求解线性方程组的实用方法。

这些特殊情况提示了一种策略：通过对方程的系统性操作，将一个一般系统转换为其中一种简单形式。

1.3 *Recalling Row Reduction*

通过系统地消除变量来求解线性方程组的方法有着悠久的历史。现代方法在此基础上发展，通过将系数矩阵 A 和常数向量 $\boldsymbol{b}$ 表达为一个单一的对象——*augmented matrix*，写作 $[\,A \mid \boldsymbol{b}\,]$。这个增广矩阵将系统的系数和常数组合成一个数组。

线性系统的解法通过一系列操作进行，每一步将增广矩阵转换为另一个表示等效系统的矩阵（具有相同的解）。

定义 1.4（初等行变换）。矩阵上的 *elementary row operation* 是三种类型之一：

R1：交换任意两行 R2：将任意一行乘以非零标量 R3：将一行的倍数加到另一行

每个都保持相应线性系统的解集。 •

Caveat: 尽管这三种操作看起来很简单，但它们的顺序非常重要。选择不当的操作顺序可能会导致不便和/或数值不稳定。

这些操作虽然简单，却十分强大。第一种，R1，允许对方程进行战略性定位。第二种，R2，实现系数的归一化。第三种，R3，是消元的原子单位——借助它可以系统地从方程中消除变量。

这些操作的目的是将增广矩阵转换为适当简单的形式。

Example: 行阶梯形。

$$\begin{bmatrix} \bullet & * & * & * & * \\ 0 & \bullet & * & * & * \\ 0 & 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

定义 1.5（行阶梯形）。若一个矩阵处于 *row echelon form*，则：

1. 所有零行（如果有的话）出现在底部 2. 每个非零行中的第一个非零元素（*pivot*）出现在其上方行的所有主元的右侧 3. 主元下方列中的所有元素为零

一个行最简形式的矩阵，如果所有枢轴上方的条目也为零，则称其处于 *reduced row echelon form*。 •

达到行梯形形式的过程揭示了线性系统的结构。对应于主元列的变量是 *bound* – 由系统中的其他变量决定。其余的变量是 *free* – 它们可以任意选择，束缚变量会相应调整以保持系统的约束。

Foreshadowing: 有界变量和自由变量之间的区别预示着我们在研究向量空间时将遇到的更深层次结构：维度与约束之间的关系。

如果继续进行行变换，直到超出行最简形，将主元上方的所有元素也设为零，结果是 *reduced row echelon form*。这种形式是线性系统独有的——尽管许多不同的行变换顺序可能都能达到这种形式。通往简化行最简形的路径可能不同；但最终结果是唯一的。

通过这种化简，解空间的维数得以揭示：它等于系统中自由变量的数量。行化简的代数过程与解集的几何解释之间的这种联系体现了线性代数的一个核心主题：计算、代数和几何视角的相互作用。

1.4 *Inverse & Invertibility*

对于方阵 A，系统 $A\boldsymbol{x} = \boldsymbol{b}$ 作为 *determined* 问题的模型具有特殊意义——即方程数量与未知数数量相同的问题。这类系统的可解性取决于一个基本性质：

定义 1.6（非奇异性）。如果下列任一等价条件成立，则方阵 A 是 *nonsingular*：

1. 存在一个矩阵 A^{-1}，使得 $AA^{-1} = A^{-1}A = I$ 2. 系统 $A\boldsymbol{x} = \boldsymbol{b}$ 对每个 $\boldsymbol{b}$ 都有唯一解 3. 系统 $A\boldsymbol{x} = 0$ 只有平凡解 $\boldsymbol{x} = 0$ 4. 行列式非零：$\det A \neq 0$

一个不是非奇异的矩阵称为*singular*。 •

当 A 是非奇异的时，其逆矩阵 A^{-1} 为系统 $A\boldsymbol{x} = \boldsymbol{b}$ 提供了一个直接的解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$。尽管行列式为非奇异性提供了理论上的判别方法，但实际计算需要不同的工具。

Recall:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Foreshadowing: 奇异矩阵作为“压缩空间”的几何解释，在研究特征值（第7章）和奇异值（第10章）时变得深刻。

行简化提供了一种系统的方法来寻找A^{-1}或证明它不存在。形成增广矩阵$[\, A \mid I \,]$并执行行变换。如果A是非奇异的，这将产生$[\, I \mid A^{-1} \,]$—将A变换为I的相同操作将把I变换为A^{-1}。

一个奇异矩阵在行简化过程中通过一行零揭示自己。这种矩阵不可逆地压缩空间，将不同的向量映射到相同的图像。该压缩在系统中表现为

$A\boldsymbol{x} = \boldsymbol{b}$ 要么是不一致（无解），要么是不定性（有无穷多解）。

各种非奇异性刻画的等价性揭示了代数、几何和计算视角之间的深刻联系：

- 代数：$\det(A) \neq 0$
- 解析：$A\boldsymbol{x} = 0$ 只有平凡解
- 计算: $[\, A \mid I \,]$ 简化为 $[\, I \mid A^{-1} \,]$

这些联系预示着线性代数实体的代数、几何与计算属性之间的相互作用。

1.5 *Composition & Elimination*

行化简不仅仅是一系列操作：它是线性变换的复合。每一种初等行变换都可以通过左乘一个适当的 *elementary matrix* 来实现——该矩阵是通过在单位矩阵上执行同样的操作得到的。

例如，要交换矩阵 A 的第 i 行和第 j 行，需要在左侧乘以通过在 I 上执行 R1 得到的矩阵 E。要将第 i 行乘以一个非零常数 c，使用的初等矩阵 E 是通过将 I 的第 i 行按 c 倍缩放形成的：即对 I 应用 R2。要将 c 倍的第 j 行加到第 i 行上，初等矩阵 E 来源于在 I 上执行该 R3 操作。

这些初等矩阵的显著特征是它们的*invertibility*。每一种行操作都可以被还原：

- 行互换是其自身的逆运算。
- 将一行按 c 进行缩放，其逆操作是将同一行按 $1/c$ 进行缩放。
- 将 c 倍的第 j 行加到第 i 行的逆操作是改用 $-c$ 进行同样的操作。

Gaussian elimination 的过程——将矩阵系统地化为行阶梯形——因此可以表示为这三类初等矩阵的复合：

$$E_k \cdots E_2 E_1 A = R$$

其中 R 是行阶梯形，而每个 E_i 都是初等矩阵。$E_k \cdots E_2 E_1$ 的乘积表示这些行操作的累积效果。当 A 可逆时，这一序列将继续直到 $R = I$，从而得到

$$A^{-1} = E_k \cdots E_2 E_1$$

这种将行化简视为可逆线性变换的复合的视角，揭示了线性代数的算法核心。

Examples: 行初等变换矩阵及其逆矩阵：

R1:
$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

R2:
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

R3:
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

尽管高斯消元最初被构想为一种计算方法，但它体现了一个更深层的原理：将复杂的变换分解为一系列简单且可逆的步骤。

1.6 LU Decomposition

我们对初等矩阵和高斯消元的阐述暗示了矩阵分解中更深层的结构。产生上三角矩阵的一系列行操作可以重新组织，从而揭示原矩阵的一种自然分解。

定义 1.7（LU 分解）。方阵 A 的一种 *LU decomposition* 将其表示为乘积 $A = LU$，其中 L 是下三角矩阵（对角线为 1），而 U 是上三角矩阵。

•

矩阵U正是通过不进行行交换的高斯消元所得到的结果；矩阵L记录了消元过程中所使用的乘子。

例 1.8。对于一个 3×3 矩阵，*LU* 分解的形式为：

$$A = \begin{bmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

其中 ℓ_{ij} 是消元乘子。 ◇

这种因式分解是在消元过程中自然产生的。
当我们使用乘子 m，利用第 j 行来消去 (i, j) 元素时，这个相同的乘子会出现在 L 的 (i, j) 位置。上三角矩阵 U 记录了这些消元的结果。因此，我们不是存储一系列初等矩阵，而是将它们的累积效果存储在 L 中。

Caveat: 一个 *LU* 分解的存在假设我们可以在不进行行交换的情况下完成消元。当需要进行行交换时，更一般的 *PLU de-* ~~*composition*~~ 会引入一个置换矩阵 P。

LU 分解的实用性在于其在求解方程组时的高效性。一旦计算完成，因子 L 和 U 使我们能够通过逐次代入求解 $A\boldsymbol{x} = \boldsymbol{b}$：

1. 首先用前向代入法求解 $L\boldsymbol{y} = \boldsymbol{b}$ 2. 然后用回代法求解 $U\boldsymbol{x} = \boldsymbol{y}$

Example: 在电路分析中，人们常常在相同的网络拓扑（A）下，但使用不同的电压或电流源（$\boldsymbol{b}$），反复求解 $A\boldsymbol{x} = \boldsymbol{b}$。LU 分解非常适合这种场景。

当需要求解具有相同系数矩阵但不同右端项的多个线性系统时，计算优势便清晰可见。因式分解只需计算一次，其代价约为对一个 $n \times n$ 矩阵进行 $\frac{2}{3}n^3$ 次运算。随后每一次求解仅需进行前代和回代，共约 $O(n^2)$ 次运算——与重复执行完整的消去过程相比，可显著节省计算量。

这种效率推动了 *LU* 分解在数值线性代数中的广泛应用。从电路分析到结构力学再到流体动力学，线性方程组很少孤立出现。在多个右端项之间复用同一个矩阵分解的能力在实际应用中极其宝贵。

Foreshadowing: LU 分解只是我们将遇到的几种矩阵分解之一。每种分解揭示了矩阵结构的不同方面，并满足不同的计算需求。

LU 分解体现了计算数学中一个反复出现的主题：用增加的存储（显式因子 L 和 U）来换取更少的计算时间。随着我们探索更为复杂的矩阵分解，这一主题将反复出现，而每一种分解都在存储、计算与洞见之间提供其自身的平衡。

1.7 *Pivots & Permutations*

如前所述的高斯消元过程假定我们可以使用任何非零元素作为主元。在实际计算中，这在数值上并不明智。请考虑下面这个系统的消元过程，其系数取自一个数据集：

$$\begin{bmatrix} 0.003 & 7.149 \\ 2.483 & 3.092 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

使用 0.003 作为主元需要除以一个很小的数——这会将其他项中的任何舍入误差放大 1000 倍。先进行行交换可以得到更稳定的消去过程。

这表明需要对我们的消元策略作出修改：在对每一列进行消元之前，我们首先通过置换行来选择一个合适的主元。这样的行交换由置换矩阵来表示。回顾§1.3，置换矩阵 P 是通过重新排列单位矩阵的行得到的；与 P 相乘会实现矩阵行的相应重排。当我们将这种主元选择策略并入消元过程时，我们得到：

定义 1.9（PLU 分解）。矩阵 A 的一个 *PLU decomposition* 将其表示为乘积 $A = P^{-1}LU$，其中：

1. P 是一个置换矩阵 2. L 是对角线为 1 的下三角矩阵 3. U 是上三角矩阵

对于任意非奇异矩阵 A 都存在这样的分解，并且它刻画了带部分主元选取的高斯消去法的步骤。

•

在实际操作中，我们首先对A进行置换，得到PA；然后将其分解为 $PA = LU$。由于P是可逆的，我们可以写成$A = P^{-1}LU$。该系统

$A\boldsymbol{x} = \boldsymbol{b}$ 因此变成

$$P^{-1}LU\boldsymbol{x} = \boldsymbol{b} \quad \Longrightarrow \quad LU\boldsymbol{x} = P\boldsymbol{b}$$

我们通过以下方式解决：

1. 计算 $P\boldsymbol{b}$ (对 $\boldsymbol{b}$ 应用与消元) 中相同的行交换 2. 通过前向代入求解 $L\boldsymbol{y} = P\boldsymbol{b}$ 3. 通过回代求解 $U\boldsymbol{x} = \boldsymbol{y}$

Caveat: 尽管我们将分解写作 $PA = LU$，但实际上我们将 P 存储为置换向量或一系列行交换，而不是显式矩阵。

这种*LU*分解的改进——通过置换引入主元—— exemplifies 了计算数学中的一个更广泛的原则：理论算法往往需要修改以确保在实际中数值稳定性。艺术在于保持基本结构，同时适应实际约束。

BONUS! 这个战略排列是 *preconditioning* 的一个例子。

1.8 Practicalities of Linear Systems

理论通过实践显现出来，因为从计算中涌现出基本模式。尽管到目前为止我们的发展主要强调线性系统的代数结构——它们的解空间、消元方法和矩阵分解——但工程学要求更多。我们不仅要确定解是否存在，还要确定我们是否能够可靠地计算出解。这种抽象数学与实际计算之间的桥梁要求我们理解操作的几何意义及其对有限精度算术的数值敏感性。

行简化到行阶梯形（回忆定义 1.5）不仅揭示了解决方案，还揭示了基本结构。以下这个不完全严格的定义将在第三章中变得至关重要。

定义 1.10（矩阵的秩）。矩阵的 *rank* 是其行最简阶梯形中主元的个数。

•

Example: 一个秩为 2 的 3×3 矩阵将 $\mathbb{R}^3$ 映射到一个平面上，坍缩了一个维度。这种几何图像有助于解释为什么这样的矩阵不可能是非奇异的。

这是衡量矩阵在变换空间方面有效性的一个基本度量。对于一个 $m \times n$ 矩阵 A，其秩满足

$$\text{rank}(A) \leq \min\{m, n\}$$

等号成立意味着*A*具有*full rank*。当*A*是方阵时，满秩等价于非奇异性。

例 1.11（行阶梯形计算）。考虑矩阵

$$A = \begin{bmatrix} 1 & 2 & 0 & 3 & 1 & 2 & 4 \\ 2 & 4 & 0 & 6 & 2 & 5 & 1 \\ 3 & 6 & 0 & 9 & 3 & 7 & 5 \\ 1 & 2 & 0 & 3 & 1 & 1 & 8 \\ 4 & 8 & 0 & 12 & 4 & 9 & 2 \end{bmatrix}$$

Mirabile dictu: (1,1) 位置的元素是一个完美的主元。消去第一列会带来显著的简化；随后消去第六列并稍作重排，得到最终的行阶梯形。

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & -14 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & 2 & 0 & 3 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

出现了几个有趣的特性：

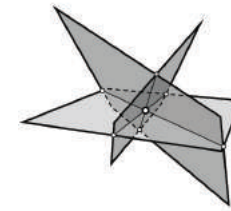
1. 第一行操作揭示了第2到第5行几乎是线性相关的，仅在最后两个条目上有所不同。 2. 该矩阵的秩为3，证明了在阶梯形矩阵中有三个非零行。 3. 第三列全为零，因此无需在此进行消元。 4. 只有在消元之后，前五列之间的依赖关系才变得清晰。

◇

示例 1.12（表面平整度测量）。考虑对加工金属表面的质量控制检查，其中坐标测量机采样五个点以验证平整度。测量值（单位：微米）产生以下坐标：（1.23, 3.41, 502.1），（4.56, −2.17, 498.4），（−2.89, 1.76, 501.3），（0.12, −4.33, 499.7）和（3.45, 2.91, 500.8）。为了评估平整度偏差，我们寻找一个最佳拟合平面 $z = ax + by + c$ 来逼近这些点。每个测量值 (x_i, y_i, z_i) 生成一个方程式在我们的系统中：

$$\begin{bmatrix} 1.23 & 3.41 & 1.0 \\ 4.56 & -2.17 & 1.0 \\ -2.89 & 1.76 & 1.0 \\ 0.12 & -4.33 & 1.0 \\ 3.45 & 2.91 & 1.0 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 502.1 \\ 498.4 \\ 501.3 \\ 499.7 \\ 500.8 \end{pmatrix}$$

这个5×3系统的方程数多于未知数——这是计量学中常见的情况，其中冗余的测量有助于减少单个测量误差的影响。z 坐标集中在 500 微米附近（标称表面高度），偏差表明既有系统性倾斜，也有随机测量噪声。尽管矩阵具有满秩 3，但测量中的小变化可能导致计算系数 a、b 和 c 发生出人意料的大变化。这种对测量扰动的敏感性，对于理解我们计算解的可靠性至关重要，因此我们需要检查不同矩阵在数值稳定性上的差异。$\diamond$



几何视角阐明了这些代数概念。线性系统中的每个方程表示一个位于$\mathbb{R}^n$中的$(n-1)$-维超平面。解集是这些超平面的交集。唯一解对应于n个超平面在一个点上相交；平行的不同超平面没有解；重合或相交于一条直线的超平面则有无限多个解。

Caveat: 尽管这种几何视角有助于我们在二维或三维空间中的直觉，但要小心在更高维度中过度依赖几何思维，因为我们的直觉往往在这些维度中会失效。

这些概念的实际意义在于它们能够预测线性系统的行为，先于尝试求解系统。秩决定了解是否存在解；非奇异性告诉我们该解是否唯一。这种结构性理解指导我们选择求解方法，并帮助我们解释结果。

例 1.13。并非所有矩阵在计算上的可处理性方面都是一样的。考虑求解系统 $A\boldsymbol{x} = \boldsymbol{b}$，其中

$$A = \begin{bmatrix} 1 & 0.999 \\ 0 & 0.001 \end{bmatrix}$$

尽管该矩阵是非奇异的，但$\boldsymbol{b}$的微小变化可能会导致解$\boldsymbol{x}$发生较大变化。这种对扰动的敏感性——无论是来自测量误差、计算中的舍入误差，还是小数位截断——从根本上限制了我们可靠地求解线性系统的能力。

这种灵敏度具有几何意义：A 将单位圆映射为一个极为偏心的椭圆，在一个方向上拉伸的空间是另一个方向的千倍。*condition number* 的 A，记作 cond(A)，通过其最大与最小拉伸因子的比率精确地衡量这种偏心性：

Nota bene: 形式化定义需要第10章中的概念，但这种几何直觉——即某些矩阵对空间的扭曲比其他矩阵更为极端——即便在现在也依然对我们大有裨益。

$$\textsc{cond}(A) = \frac{\text{maximum stretching}}{\text{minimum stretching}}$$

对于上面的矩阵，cond(A) ≈ 2000，表示在求解该系统时，某些方向上的误差可能会被放大2000倍。

实际意义立刻显现：当 cond(A) 很大时，我们称之为 *A ill-conditioned*，并以适当的怀疑态度对待计算得到的解。当 cond(A) 适中（例如小于 100）时，我们对数值结果更有信心。这个单一的数值为哪些线性系统可以可靠求解、哪些需要更加谨慎的处理提供了关键指导。 ◇

关于条件性与精度之间更深层次的关系，将在第 6 章研究最小二乘问题时显现出来，并在第 10 章中借助奇异值分解揭示其几何本质。眼下，这对数值敏感性的初步一瞥已构成关键警示：在线性代数的工坊里，并非所有工具都同样可靠。有些矩阵如同平衡良好的仪器，能够将我们的数学意图转化为可信的结果；而另一些虽然在理论上无懈可击，却在实践中表现出危险的敏感性。

Network Flows: From Graphs to Linear Systems

世界依赖网络运转。供应链将商品从工厂运送到商店；管道在城市之间输送石油和天然气；计算机网络在互联网中转发数据包。尽管这些系统看起来很复杂，其根本行为可归结为求解线性方程组——也正是我们在本章中研究过的那些方程。

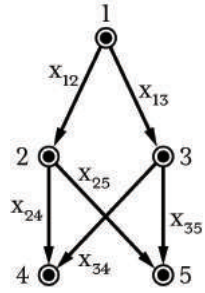
一个（有向）*network* (或*graph*)由通过有向的*edges*连接的*vertices* (或*nodes*)构成。将顶点视为位置，将边视为它们之间的路径。

为了采用一种更为形式化的方法，人们指定一个（有限的）顶点集合V。边由顶点的有序对组成：$E \subset V \times V$，其中顺序蕴含方向性。通常要求一条边中的两个顶点彼此不同。

Think: 在社交网络中，顶点是人，边是两个人之间的社会关系（“朋友”或“关注”）。

考虑一个区域配送网络，包含五个地点，满足以下条件：

- 工厂（节点1）生产200单位
- 两个负责库存分拨的区域仓库（节点2、3）
- 两个零售中心（节点4、5）各需要100个单位



如图所示，运输路线构成一个有向图，其中流量变量 x_{ij} 表示从节点 i 运送到节点 j 的货物数量。工厂向仓库供货，而仓库再向零售中心供货。

流量守恒要求流入等于流出（除源点和汇点外）：

- 在工厂：总流出量等于产量
- 在每个仓库：入库量等于总出库量
- 在每个零售中心：到货量等于需求量

这将生成我们的线性方程组：

$$\begin{aligned} x_{12} + x_{13} &= 200 \quad \text{(factory output)} \\ x_{12} - x_{24} - x_{25} &= 0 \quad \text{(warehouse 1 balance)} \\ x_{13} - x_{34} - x_{35} &= 0 \quad \text{(warehouse 2 balance)} \\ x_{24} + x_{34} &= 100 \quad \text{(retail 4 demand)} \\ x_{25} + x_{35} &= 100 \quad \text{(retail 5 demand)} \end{aligned}$$

将其写为 $A\boldsymbol{x} = \boldsymbol{b}$：

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_{12} \\ x_{13} \\ x_{24} \\ x_{25} \\ x_{34} \\ x_{35} \end{pmatrix} = \begin{pmatrix} 200 \\ 0 \\ 0 \\ 100 \\ 100 \end{pmatrix}$$

该系统具有 LU 分解 $A = LU$，其中：

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad U = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

这种分解的实际价值在供给或需求模式发生变化时便显现出来——这在真实的配送网络中每天都会发生。考虑三种情景：

1. 零售中心4需要150个单位，而中心5只需要50个单位 2. 工厂将产量提高到250个单位 3. 一座仓库暂时关闭，需要重新调整物流流向

对于前两种情况，只有右端项 $\boldsymbol{b}$ 发生变化。一旦计算并存储了 L 和 U，我们就可以通过前向代入和回代来求解每一个新的情形：

$$L\boldsymbol{y} = \boldsymbol{b}_{\text{new}} \quad \text{then} \quad U\boldsymbol{x} = \boldsymbol{y}$$

与计算新的 LU 分解所需的 $O(n^3)$ 成本相比，这只需要 $O(n^2)$ 次运算。第三种情况——网络的结构性变化——需要重新计算分解，这与直觉一致：重大的网络重构需要重新分析，而流量的常规变化可以更高效地处理。

Foreshadowing: 在第6章中，我们将看到优化原理如何帮助在多个可行解中进行选择，选取使成本最小化或效率最大化的流。

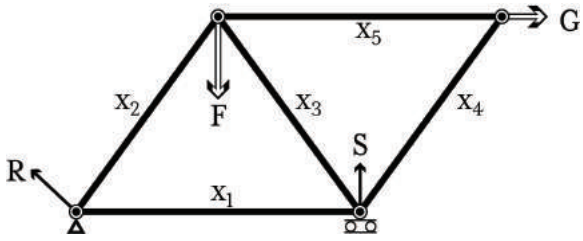
通过网络，第1章中的抽象方程获得了具体的意义——它们成为理解和管理通过相互连接的系统中商品、车辆或信息流动的工具。最初只是对数字和变量的操作，如今演变为解决现实世界分配问题的一个框架。

Structural Analysis: Forces in Trusses

建筑屹立、桥梁跨越，皆源于对力的精妙平衡。最简单的结构单元——在节点处连接成桁架的梁——既具实用价值，又富有数学之美。尽管工程师对这类结构的分析已有数百年历史，其基本行为归结为精确求解本章所建立的线性方程组。

*truss*由通过完全铰接的节点连接的刚性杆件组成，其构件通过纯拉力或纯压力来承受荷载。每个节点（或结点）必须保持平衡，水平方向和竖直方向的力相互平衡。这些平衡条件形成我们的方程组，而材料强度的物理约束使得求解方法的稳定性至关重要。

考虑这个同时承受竖向和水平荷载的五杆桁架：



每个杆件力x_i表示沿其长度方向的拉力（正）或压力（负）。如图所示，外载荷——F向下、G向右——必须分别由支座R和S处的反力来平衡。

各节点的力平衡在水平方向（x）和竖直方向（y）上分别得到方程。对于支座，我们在铰支座（节点1）处引入反力R_x和R_y，在滚支座（节点2）处引入反力R_2。假设该桁架宽4个单位、高2个单位，并将该系统写成$A\boldsymbol{x} = \boldsymbol{b}$，其中所有反力首先分组：

$$\begin{bmatrix} 1 & 0 & 0 & 1 & -0.894 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.447 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -0.894 & -0.894 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.447 & 0.447 & 0 \\ 0 & 0 & 0 & 0 & 0.894 & 0.894 & 0 & 1 \\ 0 & 0 & 0 & 0 & -0.447 & -0.447 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.894 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.447 & 0 \end{bmatrix} \begin{pmatrix} R_x \\ R_y \\ S \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ F \\ G \\ 0 \end{pmatrix}$$

这个系统的LU分解特别有价值。L因子捕捉了通过结构的负载传递，而U揭示了平衡关系的顺序。稀疏模式反映了桁架的物理连接性，使存储和计算变得高效。

这种因式分解使得在负载变化时能够快速重新分析——这是一个至关重要的能力，因为环境力是不断变化的。考虑结构工程师必须分析的三种情境：

1. 风荷载在两个上部节点施加水平力

2. 雪积累不对称地增加垂直载荷
3. 支持结算稍微修改几何系数

前两种情况仅修改右侧 $\boldsymbol{b}$，通过存储的因子实现高效求解。第三种情况——涉及几何变化——需要重新计算系数和因式分解。然而，即便在这里，稀疏性模式保持不变，允许优化后的重因式分解。

Example: 仅1%的成员角度变化就能引起10%的内力变化，强调了本章中开发的稳定数值方法的重要性。

矩阵条件性在结构分析中尤为关键。几乎平行的构件会产生近似线性相关的方程；接近直角的构件则会生成尺度差异巨大的系数。为 PLU 分解引入的主元选取策略正是直接应对这些挑战，即使面对几何形态复杂的桁架，也能确保分析的可靠性。

通过结构分析，第1章的抽象方程式获得了具体的物理意义——它们成为确保建筑物稳固和桥梁安全跨越的工具。最初作为数字操作的内容，最终成为理解力量如何在建筑环境中流动的框架，这一框架通过对矩阵结构和数值稳定性的细致研究变得具有实际意义。

□ ——————————————————————————— □

Exercises: Chapter 1

1. 使用高斯消元法求解下列线性方程组：

$$\begin{aligned} 2x + y - z &= 1 \\ 4x - y + 2z &= 2 \\ -2x + 5y - z &= 3 \end{aligned} \quad : \quad \begin{aligned} 2x + y - z &= 1 \\ 4x - y + 2z &= 2 \\ -2x + 5y - z &= 3 \end{aligned}$$

2. 验证每个矩阵是否

$$A = \begin{bmatrix} -1 & 4 & 2 \\ 0 & 3 & 0 \\ 0 & 2 & -1 \end{bmatrix} \quad : \quad B = \begin{bmatrix} 1 & 3 & 0 & 0 \\ -1 & -4 & 0 & 0 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

通过计算其行列式判断是否可逆。如果它是可逆的，求其逆矩阵。3. 将下列矩阵分解为 LU 形式（不进行主元选取）：

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 7 & -1 \\ -2 & -3 & 6 \end{bmatrix}$$

通过从 L 和 U 重构 A 来验证你的结果。4. 考虑求解 $A\boldsymbol{x} = \boldsymbol{b}$，其中

$$A = \begin{bmatrix} 0.001 & 1 \\ 1 & 1 \end{bmatrix} \quad : \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

在不进行选主元以及应用行置换之后分别求解该系统。通过计算来比较这两种方法的数值稳定性

中间项的大小。这说明了关于选主元的一般原则是什么?

5. 考虑一个由电阻连接的三节点电路。其电导矩阵为

$$G = \begin{bmatrix} 3 & -1 & -2 \\ -1 & 4 & -3 \\ -2 & -3 & 5 \end{bmatrix}$$

通过求解 $G\boldsymbol{v} = \boldsymbol{i}$，找到产生电流 $\boldsymbol{i} = (1、0、-1)^T$ 的节点电压 $\boldsymbol{v}$。

Note: 在电力网络中，G 是对称的，并且由于基尔霍夫定律，其各行和为零。

6. 一个输入-输出经济模型有三个部门，输入矩阵为

$$A = \begin{bmatrix} 0.3 & 0.2 & 0.1 \\ 0.4 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.4 \end{bmatrix}$$

其中 a_{ij} 表示生产一个单位的部门 j 产出所需的部门 i 的产出数量。给定需求 $\boldsymbol{d} = (100, 150, 200)^T$，通过求解 $(I - A)\boldsymbol{x} = \boldsymbol{d}$，找出满足该需求所需的生产水平 $\boldsymbol{x}$。

Note: 在经济模型中，A 通常具有非负的条目，且各列之和小于 1。

7. 一个化学反应器包含三种物种 A、B 和 C，它们按照一级动力学相互转化。速率矩阵为

$$K = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix}$$

如果初始浓度为 $\boldsymbol{c}_0 = (1, 0, 0)^T$，通过求解 $K\boldsymbol{c} = 0$ 并满足质量守恒 $\sum_i c_i = 1$，求稳态浓度。

Note: 由于质量守恒，速率矩阵的行和为零。

8. 设 A 和 B 是 $n \times n$ 矩阵，且假设 $AB = I$。证明 A 和 B 是可逆的，并且 $B = A^{-1}$。

9. 证明任何 n 乘 n 的置换矩阵 P 是单位矩阵的根，具体是 $P^n = I$。对于比 n 更小的幂次，这种情况是否会发生?

10. 令 N 表示一个 k 行 k 列的矩阵，除了在上对角线上有 $+1$ 外，其余全为零：即，$N_{i,j} = 1$ 对于 $j = i + 1$，其他位置为 0。证明 N^p 对于 $p < k$ 非零，且对于 $p \geq k$ 为零。

这种矩阵称为*nilpotent*，因为在足够多次幂之后它会消失（变为零）。

11. 考虑矩阵

$$A = \begin{bmatrix} 1 & \alpha \\ 0 & 1 - \alpha \end{bmatrix}$$

其中 α 是一个参数。对于哪些 α 的取值，A 是条件良好的（条件数小于 10）？对于哪些取值它是病态的（条件数大于 100）？

12. 设

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

是一个旋转矩阵。它的条件数是多少？从几何角度解释为什么这个结果是合理的。

13. 在结构分析问题中，将力 $\boldsymbol{f}$ 与位移 $\boldsymbol{u}$ 联系起来的刚度矩阵具有分块形式

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}$$

如果K_{11}是可逆的，展示如何使用块消元法高效地求解$K\boldsymbol{u} = \boldsymbol{f}$。什么条件确保该方法在数值上是稳定的？14. 高斯消元中的*growth factor*衡量了在过程中条目可能增长的程度。对于矩阵A，它被定义为$\rho(A) = \max_{i,j,k} |a_{ij}^{(k)}| / \max_{i,j} |a_{ij}|$，其中$a_{ij}^{(k)}$表示经过$k$步消元后的$(i, j)$条目。对于矩阵

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1+\epsilon & 1 \\ 1 & 1 & 1+\epsilon \end{bmatrix}$$

证明在不进行主元选取的情况下，当 $\epsilon >$ 很小（$\to 0$）时，增长因子近似为 $1/\epsilon$。找到一个置换 P，使得 PA 的增长因子近似为 1，并解释为什么这表明主元选取对于数值稳定性的重要性。

15. 考虑一个矩阵 A，其对角线元素非零，但超对角线元素要大得多：$|a_{i,i+1}| \gg |a_{ii}|$，对于 $i = 1, \ldots, n-1$。解释为什么在不进行主元选取的情况下计算 LU 分解很可能是不稳定的，循环置换将第一行移到最底部如何可能有所帮助，以及增长因子这一概念如何解释这种现象。

16. 设 A 是一个 $n \times n$ 的矩阵，其元素的大小不超过 1。证明存在一个置换矩阵 P，使得 PA 的 LU 分解中所有主元的大小至少为 $1/n!$。这个界限是否是最优的？

Nota bene: 这表明某些旋转策略总是能够确保枢轴不会变得太小，尽管找到最佳排列通常是不可解的。

17. 考虑一个质量-弹簧系统，其中有两个质量 m_1 和 m_2 通过弹簧连接，弹簧常数分别为 k_1、k_2 和 k_3：

$$\text{wall} \xrightarrow{k_1} m_1 \xrightarrow{k_2} m_2 \xrightarrow{k_3} \text{wall}$$

证明寻找平衡位置需要解一个系统 $A\boldsymbol{x} = \boldsymbol{b}$，其中 A 是对称的并且是三对角矩阵。什么物理原理解释了这种对称性？

18. 记住，在网络流问题中，*incidence matrix* A 的条目为 $a_{ij} = 1$ 如果边 j 进入节点 i，$a_{ij} = -1$ 如果边 j 离开节点 i，否则 $a_{ij} =$ 为 0。证明对于任何具有 n 个节点的连通图，$\text{rank}(A) = n - 1$。这与流的守恒有什么关系？

19. 一个实矩阵 A 被称为 *totally positive*，如果它的所有子式（方阵子矩阵的行列式）都是正的。证明如果 A 是完全正的，那么它的 LU 分解在无需主元素的情况下存在。这对数值稳定性意味着什么？

BONUS! 完全正定矩阵在逼近理论和统计学中自然出现，它们的特殊性质证明是无价的。

20. 矩阵 A 的 *Cayley transform* 定义为 $C = (I + A)(I - A)^{-1}$，当 $(I - A)$ 可逆时。证明如果 A 是块对角矩阵，则 C 也是块对角矩阵，其块是 A 对角块的凯利变换。这对计算可能有什么优势？

21. 考虑求解方程组 $A\boldsymbol{x} = \boldsymbol{b}$，其中 A 是 *strictly diagonally dominant*：$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$，对所有 i 成立。证明 A 是非奇异的，并且在高斯消去中不需要进行行交换。

22. 一个网页排名算法根据链接矩阵 L，为 n 个页面分配重要性得分 $\boldsymbol{x}$，其中 $L_{ij} =$ 为 1 当页面 j 链接到页面 i，否则为 0。

在将 L 的各列归一化使其列和为 1 之后，通过求解来更新得分

$$\boldsymbol{x} = \alpha L\boldsymbol{x} + (1-\alpha)\mathbf{1}/n$$

其中 $\alpha = 0.85$ 和 1 是全为1的向量。证明该系统始终有唯一解。为什么这对于网页搜索很重要?

Chapter 2
Abstract Vector Spaces

“in fear & pale dismay He saw the indefinite space beneath”

从具体到抽象的跃迁标志着本文的第一个重大挑战。在将向量作为有序数列——无论是力、速度还是数据——进行了大量实践之后，我们现在退一步，提出一个更深层的问题：什么 *is* 向量？哪些本质特征使某物具有向量性？

这种抽象并非只是学术上的练习。现代工程中出现的向量往往超越了简单的坐标列表。一个向量可能表示时变信号、高维数据集、图像、多项式或概率分布。我们对这些向量进行的运算——加法、缩放、点积——呼应了欧几里得几何中熟悉的操作，但却从几何中抽离出来，成为纯粹的形式。

考虑一组构成声波的音频样本，或构成数字图像的像素强度。我们可以将两个这样的对象相加（混合声音或融合图像），并对它们进行缩放（改变音量或亮度）。这些操作满足与 $\mathbb{R}^n$ 中的向量加法和标量乘法相同的代数规则。然而，这些对象远非空间中的几何箭头。它们在更一般的意义上是 *vectors*——是 *vector space* 的元素。

Hmmmm... 向量是向量空间中的一个元素，而向量空间是由向量组成的空间。*There must be more than this...*

这种抽象的力量在于它能够在一个共同的框架下统一彼此迥异的情境。无论是在处理微分方程的解、物理学中的量子态，还是机器学习中的特征嵌入，向量空间的基本性质都指导着我们的分析与计算。通过以抽象形式理解这些性质，我们获得了可在现代工程广阔领域中通用的工具。

我们的[illegible]是要小心地构建这个抽象，始终保持 ays a

与具体内容的联系。我们从向量空间的公理化定义开始，并通过熟悉的例子来阐明其基本特征。这一基础将支持我们随后对变换、内积以及促成现代计算方法的更深层结构的研究。

2.1 *Vector Space Axioms*

定义 2.1（向量空间）。一个 *vector space* 由两个要素组成：对象的一个集合 V（称为 *vectors*），以及一个 *scalars* (的域；就我们的目的而言，始终是实数 $\mathbb{R}$)。它们通过两个基本运算联系在一起：

1. 向量加法：一种将任意两个向量 $\boldsymbol{u}$、$\boldsymbol{v} \in V$ 组合以获得一个新向量 $\boldsymbol{u}+\boldsymbol{v} \in V$ 的规则 2. 标量乘法：一种用实数 $c \in \mathbb{R}$ 对任意向量 $\boldsymbol{v} \in V$ 进行缩放以获得一个新向量 $c\boldsymbol{v} \in V$ 的规则

并非所有具有加法和缩放的对象集合都符合向量空间的要求。公理确保向量以保持“向量性质”的方式进行组合和缩放。

这些运算必须满足某些规则——即*vector space axioms*。对于所有向量$\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w} \in V$以及所有标量a、$b \in \mathbb{R}$：

Vector Addition Axioms:

1. 交换律：$\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$ 2. 结合律：$(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$ 3. 零向量：存在一个向量 $0 \in V$，使得对所有 $\boldsymbol{v} \in V$，$\boldsymbol{v}+0=\boldsymbol{v}$ 4. 加法逆元：对每个 $\boldsymbol{v} \in V$，存在一个向量 $-\boldsymbol{v} \in V$，使得 $\boldsymbol{v}+(-\boldsymbol{v})=0$

Scalar Multiplication Axioms:

1. 对向量加法的分配律：$a(\boldsymbol{u}+\boldsymbol{v})=a\boldsymbol{u}+a\boldsymbol{v}$ 2. 对标量加法的分配律：$(a+b)\boldsymbol{v}=a\boldsymbol{v}+b\boldsymbol{v}$ 3. 与标量的结合律：$a(b\boldsymbol{v})=(ab)\boldsymbol{v}$ 4. 单位性：$1\boldsymbol{v}=\boldsymbol{v}$

•

这些例子说明了在保持基本代数结构的同时，我们可以将对“向量”的概念扩展到多远。有关一些初步细节，请参见下一节。

这些公理可能显得吹毛求疵——它们当然适用于 $\mathbb{R}^n$ 中熟悉的向量。当我们考虑更为奇异的空间时，它们的重要性才会显现出来，这些空间中的对象并不 *look* 像向量，例如：

- 在逐点加法和缩放下的连续$\mathbb{R}$值函数
- 在加法和数乘下的多项式
- 在逐项加法与缩放下的泰勒或傅里叶级数
- 线性齐次常微分方程或递推关系的解

这些公理的力量不在于各自的陈述，而在于它们的整体蕴含：任何满足这些规则的对象都会继承向量的基本性质。这意味着为某一个向量空间发展出的技术，往往可以无缝地迁移到其他向量空间。用于求解 $\mathbb{R}^n$ 中线性方程组的方法，经过极少的修改，可能就能用于求解线性微分方程组，或在信号处理滤波器中找到最优系数。

向量空间不仅仅是向量的集合——它们是在加法和数乘运算下具有恰当结构的向量集合。

公理还告诉我们什么是*is not*向量空间。正实数在最小（加法）和最大（乘法）运算下不满足大多数公理（有满足的吗？）。在普通加法和乘法下，整数不满足，因为标量乘法不总是得到一个整数。这些反例有助于我们更清楚地理解使向量空间正常运作的因素。

最后，公理化方法使我们能够证明对 *all* 向量空间都成立的结果，从而省去了逐个例子验证的麻烦。例如，下面的结论在欧几里得空间中当然*seems*显然，但是否在所有可能的世界中都成立就不那么清楚了。

引理 2.2。*In a vector space V, the zero vector is unique.*

Proof. 假设 z 和 z' 是满足零性条件的 V 中的向量。那么：

$$z = z + z' = z',$$

每个等式都源自以下事实：z 和 z' 在加到任何向量时都不起作用。因此，它们是相同的向量。 □

2.2 *A Gallery of Vector Spaces*

抽象的定义通过例子而变得生动。下面各个例子展示了向量空间公理如何在不同情境中体现，从熟悉的到较为奇特的。尽管我们不会对每个例子逐一显式验证这些公理（这是一项枯燥但直接的练习），我们将明确关键要素：向量本身、加法与数乘运算，以及零向量。

尽管我们生活在一个看似三维的世界中，机械系统的构型空间通常具有更高的维度。具有多个旋转关节的机械臂在维度大于三的状态空间中演化。

例 2.3（欧几里得空间）。由有序的 n 元实数组成的空间 $\mathbb{R}^n$ 是我们的原型。在这里，向量是实数的有序列表，并在熟悉的按分量加法和标量乘法作用下进行运算。零向量是所有分量都为零的元组。这是经典物理和工程学所使用的空间，其中 $n =$ 取 2 或 3 时对应于物理空间。◇

例 2.4（矩阵）。所有 $m{\times}n$ 矩阵的集合 $\mathbb{R}^{m\times n}$ 在按元素相加和标量乘法下构成一个向量空间。零矩阵 Z 是 $\mathbb{R}^{m\times n}$ 的“零元”。矩阵空间在工程中无处不在，从计算机图形学的变换矩阵到神经网络的权重矩阵。这里的运算呼应了 $\mathbb{R}^n$ 的运算，尽管对象本身更加结构化。

2×2 矩阵的空间是四维的，尽管从其表象上并不立即显现。这个主题——维度可以隐藏在显而易见之处——将会反复出现。

◇

例 2.5（多项式）。对于每个非负整数 n，我们有次数至多为 n 的多项式空间 $\mathcal{P}_n$。一个典型元素的形式为 $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$。多项式的加法和标量乘法以自然的方式作用于其系数。所有系数均为零的零多项式充当零向量。这些空间用作对更复杂函数的近似，并广泛出现于信号处理和控制理论中。 ◇

Foreshadowing: 这是我们关于无限维向量空间的第一个例子。从有限维到无限维的跃迁是深刻的，并且蕴含着将塑造我们对收敛与逼近理解的诸多惊喜。

例 2.6（函数空间）。考虑在区间 $[a, b]$ 上取值为实数的连续函数空间 $C([a, b])$，其加法和标量乘法按点方式定义。零函数 $z(x) = 0$ 作为零向量，因为 $f + z = f$ 对所有 f 成立。该空间包含所有在受限定义域) 上的多项式 $\mathcal{P}_n$ (，并作为工程中信号空间的一个模型。通常用 $C(D)$ 来表示在定义域 D 上的标量场向量空间 $f\colon D \to \mathbb{R}$。对于具有一定可微性的函数（这在微积分中既有用又熟悉），下面这些用于标量场的记号是标准的：

- $C(D)$: 连续
- $C1(D)$: 连续可微
- $C^\infty(D)$: 无限可微 或 *smooth*
- $C^\omega(D)$: 实解析的

Recall: Real-analytic 这意味着所有泰勒级数在所有点处都收敛并且等于该函数。

◇

例 2.7（线性常微分方程）。线性齐次微分方程的解构成一个向量空间。这里的向量是满足方程 $p(D)x = 0$ 的函数 $x(t)$，其中 $p(D)$ 是微分算子 D 的一个多项式。加法和数乘运算在这些解 $x(t)$ 上逐点作用，而常数函数 $x = 0$ 是零向量。

Nota bene: 类似的空间源自线性递推关系和差分方程。

例如，一个二阶线性 ODE，其形式为

$$a\frac{d^2x}{dt^2} + b\frac{dx}{dt} + cx = 0$$

有 solu tions $x(t)$ 在线性组合下封闭的 an d 因此

形成一个向量空间。这个常微分方程可以使用微分算子 $D = d/dt$ 写为 $(aD^2 + bD + cI)x = 0$. ◇

示例 2.8（数字信号）。数字信号——无论是音频、图像，还是一般数据流——形成各自的向量空间。音频信号可以表示为样本序列或连续时间的函数。加法对应于信号的混合；标量乘法调整幅度。静音起到零向量的作用。对于图像，向量是像素强度的二维数组。这些空间支持信号处理和机器学习中基础的线性运算。◇

示例 2.9（数列与级数）。考虑一个关于变量 x 的形式幂级数集合，这是单变量微积分中常见的内容。忽略收敛性，我们可以将这些幂级数视为向量。给定两个这样的级数，我们可以按项加法（按幂次）；重缩放发生在系数的层面上。零级数（$c_k =$ 对所有 k 为零）是零向量。

回顾一下，幂级数的形式为

$$f = \sum_{k=0}^{\infty} c_k x^k.$$

*convergent*幂级数会形成一个向量空间吗？绝对收敛与条件收敛有区别吗？

在序列的集合上同样存在一个向量空间结构。考虑用 $a = (a_k)$ 代替 $k \in \mathbb{N}$。可以按项相加这些序列，而对序列进行缩放意味着对每一项进行缩放。零序列扮演零向量的角色。有趣的是，这个向量空间“感觉”上与幂级数的向量空间是同一个，尽管它们看起来不同。

Foreshadowing: 在向量空间（以及其他数学领域）中，天生存在的相同性或等价性的概念称为 *isomorphism*。序列和级数的向量空间是 *isomorphic*。

◇

这些例子虽然在性质上各不相同，但都共享向量空间公理所刻画的基本特征。它们的多样性彰显了抽象的力量：为某一个向量空间发展出的技术，往往可以无缝地迁移到其他向量空间中。随着我们继续研究子空间、线性无关性和基，这些例子将作为参照点，在具体情境中为抽象概念提供支撑。

2.3 *Subspaces*

定义 2.10（子空间）。向量空间 V 的一个 *subspace* 是其自身在从 V 继承的运算下形成的向量空间。我们使用符号 $W < V$ 来表示子空间。

关键的洞察是，子空间必须对使向量空间有效的运算封闭。你无法通过加法或缩放逃离一个子空间。

•

这个看似抽象的概念具有直接的实际意义。二次多项式空间 $\mathcal{P}_2$ 包含其中的更简单空间 $\mathcal{P}_1$，即仿射函数空间——一个自然的子空间。矩阵空间 $\mathbb{R}^{n\times n}$ 包含一个对角矩阵的子空间。在数字信号处理领域，满足 $f(t) = f(-t)$ 的偶对称信号集合形成一个子

所有信号的空间。这些并非只是随机的子集，而是保持其母空间基本向量运算的结构化部分。

判断一个子集 $W \subseteq V$ 是否是子空间的检验可归结为检查三个简单的性质：

1. 零向量在 W 中 2. W 在加法下封闭：如果 $\boldsymbol{u}$，$\boldsymbol{v} \in W$，则 $\boldsymbol{u} + \boldsymbol{v} \in W$
3. W 在标量乘法下封闭：如果 $\boldsymbol{v} \in W$ 和 $c \in \mathbb{R}$，则 $c\boldsymbol{v} \in W$

Foreshadowing: 这三个属性不是独立的。由于对于任何向量 $\boldsymbol{v}$，都有 $0\boldsymbol{v} = 0$，第一个实际上是冗余的。这种冗余的提示在我们描述子空间时，预示了即将到来的更深层次的结构结果。

例 2.11（坐标子空间）。在 $\mathbb{R}^n$ 中，坐标平面（以及它们的高维对应物）提供了子空间的自然例子。例如，在 $\mathbb{R}^3$ 中，xy-平面是子空间 $\{(x, y, 0) : x, y \in \mathbb{R}\}$。更一般地，任何经过原点的平面或直线都构成一个子空间。子空间必须包含 0 的要求迫使它们经过原点——一个平移后的平面，无论离原点多近，都不是子空间。

◇

例 2.12 （零空间）。线性齐次系统 $A\boldsymbol{x} = 0$ 的解构成一个子空间，称为 *null space* 的 A。这不仅仅是一个方便的事实，而是线性关系的结果：如果 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 满足该方程，则

Foreshadow: 在第3章介绍线性变换时，我们将使用更一般的术语 *kernel* 来代替零空间。

$$A(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) = c_1A\boldsymbol{x}_1 + c_2A\boldsymbol{x}_2 = \boldsymbol{0}$$

对于任意标量 c_1、c_2。该子空间捕捉了该系统解的本质结构。注意，对于 $\boldsymbol{b} \neq 0$，$A\boldsymbol{x} = \boldsymbol{b}$ 的解确实 *not* 构成一个子空间。◇

例 2.13（列空间）。给定一个矩阵 A，其列的所有可能的线性组合所构成的集合形成 $\mathbb{R}^m$ (的一个子空间 col(A)，其中 m 是行数)。这个 *column space* 表示线性变换 $A\boldsymbol{x} = \boldsymbol{b}$ 的所有可能输出。类似地，还存在一个 *row space*，row(A)——行的组合——它形成 $\mathbb{R}^n$ (的一个子空间，其中 n 是列数)。

Foreshadowing: 对一组向量的所有可能组合很快将被我们称为*span*。

◇

例 2.14（矩阵子空间）。考虑在矩阵加法下由 $n \times n$ 矩阵组成的空间 $\mathbb{R}^{n\times n}$。以下是子空间：

- 上三角矩阵
- 对角矩阵
- 对称矩阵
- 迹为零的矩阵

对每种情况验证子空间性质，为熟练掌握这些公理提供了极好的练习。

◇

交运算保持子空间性质：如果 W_1 和 W_2 是 V 的子空间，那么 $W_1 \cap W_2$ 也是一个子空间。这使我们能够通过寻找已知子空间的共同元素来构造新的子空间。对于子空间的任意交也是如此——这一事实在研究线性约束系统时变得重要。

Caveat: 子空间的并集很少是一个子空间。考虑 $\mathbb{R}^2$ 中通过原点的两条直线——它们的并集在加法下不封闭。

两个子空间W_1和W_2的和，定义为

$$W_1 + W_2 = \{\boldsymbol{w}_1 + \boldsymbol{w}_2 : \boldsymbol{w}_1 \in W_1,\ \boldsymbol{w}_2 \in W_2\}$$

始终是一个子空间。当这些子空间仅有零向量作为公共部分时，也就是说，当 $W_1 \cap W_2 = \{0\}$ 时，我们称之为一个 *direct sum*，记作 $W_1 \oplus W_2$。直和具有一个关键性质：$W_1 \oplus W_2$ 中的每个向量都可以唯一地表示为一个和 $\boldsymbol{w}_1 + \boldsymbol{w}_2$，其中 $\boldsymbol{w}_1 \in W_1$ 和 $\boldsymbol{w}_2 \in W_2$。对于普通的和 $W_1 + W_2$，当子空间发生重叠时，这样的表示不一定是唯一的。

可以将直和看作是把指向“独立方向”的子空间组合起来——就像把完全不同的实轴和虚轴组合起来构建复平面 $\mathbb{C}$。

子空间的概念贯穿线性代数的各个方面，从求解线性方程组这一实际问题，到即将讨论的特征空间和奇异值分解等理论机制。理解向量空间如何分解为更简单的子空间，是实现计算效率和获得理论洞见的关键。

2.4 *Span & Linear Independence*

最简单的子空间源于最基本的向量运算：数乘。向量空间 V 中的一个非零向量 $\boldsymbol{v}$ 生成一条过原点的直线——所有标量倍数的集合 $\{c\boldsymbol{v} : c \in \mathbb{R}\}$。这是 V 的一个一维子空间。当我们允许加法以及数乘时，一个有限的向量集合会生成一个更大的子空间。

定义 2.15（张成）。向量空间 V 中向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ 的 *span* 是它们所有线性组合的集合：

$$\operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \{c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k : c_i \in \mathbb{R}\}$$

向量集合*spans* V，如果V中的每个向量都可以写成该集合中向量的线性组合。

一组向量的张成是包含它们的最小子空间。它包含所有可以通过向量空间中允许的运算由给定向量构造出来的向量。

•

例 2.16（$\mathbb{R}^2$ 中的张成）。在平面中，两个指向不同方向的非零向量 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$ 张成整个 $\mathbb{R}^2$。平面中的任意一点都可以通过适当的线性组合得到。如果这些向量指向相同（或相反）的方向，它们的张成仅仅是一条过原点的直线。

◇

例 2.17（张成多项式）。多项式 1、x 和 x^2 张成空间 $\mathcal{P}_2$——任何二次多项式 $ax^2 + bx + c$ 都是这些基本构件的线性组合。相同的多项式并不能张成 $\mathcal{P}_3$，因为任何线性组合都无法产生三次项。◇

这引出了一个根本性的问题：向量在何种情况下才真正彼此独立?

定义 2.18（线性无关）。向量空间 V 中的一组向量 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ 是 *linearly independent*，如果方程

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$$

仅有平凡解 $c_1 = c_2 = \cdots = c_k = 0$。否则，这些向量为 *linearly dependent*。

线性相关意味着冗余——在不减少张成空间的情况下，可以移除一个或多个向量。线性无关意味着每个向量都贡献了真正新的东西。

如果不存在这样的关系——如果通过线性组合得到 0 的唯一方式是使所有系数都等于零——那么这些向量是 *linearly independent*。

例 2.19 ($\mathbb{R}^n$ 中的依赖性)。$\mathbb{R}^2$ 中的三个向量总是线性相关的。这从直观上看是清楚的：平面可以由两个向量生成，因此第三个向量必定是冗余的。更一般地，如果向量的数量超过了维度，它们必定是线性相关的。◇

例 2.20（多项式的线性无关性）。多项式 1、x 和 x^2 在 $\mathcal{P}_2$ 中是线性无关的。若对所有 x 都有 $a + bx + cx^2 = 0$，则每个系数 a、b、c 都必须为零。然而，将多项式 $x^2 + 1$ 加入该集合会产生线性相关性，因为它可以表示为 1 和 x^2 的线性组合。◇

Foreshadowing: 生成与线性无关之间的相互作用引出了*basis*这一概念——一个张成该空间的线性无关向量集。这个基本概念将组织我们对向量空间的理解。

张成与线性无关这两个概念是相互补充的。张成告诉我们可以构造出哪些向量；线性无关则告诉我们是否在不产生冗余的情况下高效地进行构造。二者共同为理解向量空间及其子空间的结构奠定了基础。

原则上，检验线性无关性是直观的：必须确定一个齐次方程组是否只有平凡解。在实践中，这意味着考察某些标量的集合是否必然全部为零。下面的示例说明了这一过程。

例子 2.21（测试线性独立性）。考虑欧几里得向量 $(1, 2)^T$ 和 $(2, 4)^T$ 在 $\mathbb{R}^2$ 中。为了测试线性独立性，我们检查

$$c_1\begin{pmatrix}1\\2\end{pmatrix} + c_2\begin{pmatrix}2\\4\end{pmatrix} = \mathbf{0} \quad \Rightarrow \quad \begin{aligned} c_1 + 2c_2 &= 0 \\ 2c_1 + 4c_2 &= 0 \end{aligned}$$

第二个方程是第一个的两倍，从而对任意c_2得到$c_1 = -2c_2$。因此，这些向量线性相关 – v_2是v_1的两倍。 ◇

当检验独立性时，沿着零向量进行。关键问题始终是：哪些系数会产生零向量，它们是否必然全部为零?

张成与线性无关的概念为描述向量空间的本质结构提供了所需的语言。它们同时告诉我们能够构建什么（通过张成），以及构建它所需要的条件（通过线性无关）。这种双重视角——我们能做什么与为此所需的条件——将引导我们对该理论的发展。

2.5 *Towards Dimension*

Foreshadowing: 在第11章中，主成分分析发现在数据中存在低维结构；在第12章中，我们通过低维因子来逼近高维矩阵；在第13章中，神经网络将极大维度压缩成关键特征。艺术不在于计算维度，而在于理解何时以及如何在不丢失关键信息的情况下进行降维。

维度这一概念贯穿于我们的物理世界和数学世界。我们谈论三维空间、二维曲面、一维直线。工程师们经常在更高维度中工作：一个具有六个关节的机械臂在六维构型空间中描绘路径；一个拥有数十亿权重的神经网络在一个超出普通想象的空间中运行。

我们的任务是从这些例子中提炼出一个维数的定义，以捕捉其本质特征：需要多少个独立参数才能唯一地确定一个向量？在 $\mathbb{R}^n$ 中，这一点很清楚——我们恰好需要 n 个坐标。对于其他向量空间，我们必须借助生成集和线性无关性来获得指引。

向量空间的一个生成集可能并不高效，包含冗余的向量。一种自然的维数度量会计算生成该空间所需的 *minimal* 个向量的数量。幸运的是，这个最小数量是良好定义的。

引理 2.22 （最小生成集）。*Any two minimal spanning sets of a vector space V have the same size.*

Proof. 设 $S = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m\}$ 和 $T = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n\}$ 是 V 的极小生成集。由于 S 张成 V，每个 $\boldsymbol{w}_j$ 都可以表示为 S 中向量的线性组合：

$$\boldsymbol{w}_j = \sum_{i=1}^{m} c_{ij}\boldsymbol{v}_i$$

对于一些标量 c_{ij}。

我们声明 $m \geq n$。如果不是，那么 $m < n$，我们可以将 n 向量 $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ 写成 m 向量 $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$ 的线性组合。这将意味着 $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ 是线性相关的。

为此，考虑齐次系统

$$\sum_{j=1}^{n} x_j \boldsymbol{w}_j = \mathbf{0}$$

将 $\boldsymbol{w}_j$ 的表达式代入：

$$\sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} c_{ij} \boldsymbol{v}_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} x_j c_{ij} \right) \boldsymbol{v}_i = \mathbf{0}$$

这是一个包含m个方程、n个未知数的齐次方程组。当$m < n$时，这样的方程组必然有非平凡解，这意味着$\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$线性相关。这与$T$作为生成集的极小性相矛盾。

一个对称的论证，将每个 $\boldsymbol{v}_i$ 用 $\boldsymbol{w}_j$ 表示，表明 $n \geq m$。因此 $m = n$。

□

这一显著的结果——即向量空间的所有极小生成集都具有相同的大小——使我们能够无歧义地定义维数。

定义 2.23（维数）。向量空间 V 的 *dimension*，记为 $\dim V$，是 V 的任意一个极小生成集的大小。若不存在有限生成集，则称 V 是 *infinite-dimensional*。

•

这个定义的力量在于它对特定坐标系或几何直观的抽象。它同样适用于多项式、矩阵、信号或微分方程解的空间。让我们考察一些富有启发性的例子。

例 2.24（多项式维数）。考虑次数至多为 n 的多项式空间 $\mathcal{P}_n$。任意这样的多项式具有如下形式

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

一个最小生成集由单项式 $\{1、x、x^2、\ldots、x^n\}$ 组成。因此，维数 $\mathcal{P}_n = n +$ 为 1，而不是 n——我们必须计入常数项！这种微妙的区别提醒我们，维数计算的是参数的数量，而不是最高次数。

◇

例 2.25（矩阵空间）。由实数 m-by-n 矩阵构成的空间 $\mathbb{R}^{m \times n}$ 的维数为 mn。尽管我们将这些数排列成一个矩形数组，维数计算的是元素的总个数。因此 $\mathbb{R}^{2\times 2}$ 的维数是 4，这解释了为什么一般的 2×2 矩阵需要四个参数才能完全确定。

◇

例 2.26（函数空间）。闭区间上的连续函数空间 $C([a, b])$ 没有有限的生成集。考虑函数集合 $\{1, x, x^2, \ldots\}$——没有任何有限子集能够张成该空间，这可由超越函数 e^x 予以说明。这样的空间由于缺乏有限生成集，被称为无限维。

◇

有限维度与无限维度之间的区别是深刻的。有限维度空间可以通过一组有限的参数进行完整描述；而无限维度空间则无法如此简化。这一二分法塑造了我们解决问题的方法：有限维度空间适合使用计算方法，而无限维度空间通常需要通过有限维度子空间进行逼近。

我们对维数的探讨完成了向量空间理论的基础。我们已经从加法与数乘等具体运算，推进到向量空间这一抽象概念，继而到子空间与生成集，最终到维数——衡量空间复杂性的内在尺度。这种从熟悉的 $\mathbb{R}^n$ 领域出发的抽象，为我们研究线性变换做好了准备；在其中，维数将在理解空间如何相互映射方面发挥关键作用。

在接下来的章节中，我们将回到显式的坐标和基，用计算工具来丰富我们的视角。然而，这里所发展的无坐标观点——尤其是维数的基础性——仍将对我们的理解至关重要。抽象结构与具体计算之间的相互作用，正是线性代数在现代工程中力量的核心。

Engineering Signals as Vector Spaces

工程学建立在对信号的测量、分析与控制之上——信号是随时间变化的量，用以编码关于物理系统的信息。在一个时间区间内所有可能信号的集合构成了一个自然的向量空间，尽管它与$\mathbb{R}^n$中熟悉的坐标几何相去甚远。通过向量空间来理解信号，既能阐明其数学结构，也能指导其实用操作。

考虑在固定时间区间 $[0, T]$ 上定义的所有连续信号 $f\colon [0, T] \to \mathbb{R}$ 的集合 $S[0, T]$。两个信号通过逐点组合进行相加，而标量乘法则对信号的幅度进行缩放:

$$(f + g)(t) = f(t) + g(t) \quad : \quad (cf)(t) = c \cdot f(t)$$

Think: 这里的零向量是指在所有时刻恒为零的信号——即任何信号的缺失。它作为加法单位元的角色，与其作为寂静或黑暗的物理含义相呼应。

这些运算并非通过坐标操作来满足我们的向量空间公理，而是通过信号组合的根本性质来实现。

这个无限维空间包含一个自然的有限维子空间序列。对于每个正整数 n，考虑 V_n，由 1 以及前 n 对周期信号张成的空间:

$$\left\{1, \cos\left(\frac{2\pi t}{T}\right), \sin\left(\frac{2\pi t}{T}\right), \cos\left(\frac{4\pi t}{T}\right), \sin\left(\frac{4\pi t}{T}\right), \dots\right\}$$

这些子空间形成一个*filtration*——一个序列$V_0 < V_1 < V_2 < \cdots < S[0, T]$，每个

包含比上一个更复杂的周期性模式。V_n中的信号最多结合n个不同的周期成分。

Historical Note: 傅里叶的激进洞见认为，周期函数以三角函数为基构成一个向量空间，这一观点同时改变了数学与工程学。向量空间的抽象结构阐明了热流与振动中的具体问题。

线性无关性对信号具有特殊意义。考虑在 $[0, T]$ 上振荡一次的信号 $\sin(2\pi t/T)$ 和 $\cos(2\pi t/T)$。没有任何线性组合

$$c_1 \sin(2\pi t/T) + c_2 \cos(2\pi t/T) = 0$$

除平凡的 $c_1 = c_2 = 0$ 外的所有，证明了它们的独立性。在不同频率振荡的信号之间也保持类似的独立性，这使得我们的子空间序列可以在不冗余的情况下增长。

Nota bene: 尽管 $S[0, T]$ 本身是无限维的，但任何实际信号都可以由某个有限维的 V_n 中的元素任意精确地逼近。这一原理支撑了信号处理的很大一部分。

不同的工程背景揭示了其他自然子空间。在 $t = 0$ 处消失的信号构成了一个子空间，模拟从静止状态开始的系统。关于 $T/2$ 对称的信号构成了另一个子空间，反映了时间对称性。每个这样的子空间都捕捉了数学结构和物理意义。

信号处理本身成为信号空间之间变换的研究。滤波器将输入信号映射到输出信号，同时保持向量空间结构。输入的线性组合映射到输出的相同线性组合，反映了数学的优雅性和工程的必要性。向量空间的抽象属性指导着实际信号处理系统的设计和分析。

这种观点——将信号视为抽象空间中的向量，而非仅仅是时间的函数——揭示了协调隐匿的结构。尽管我们可以通过采样值或三角函数系数进行计算，信号的基本特性超越了任何特定的表示形式。向量空间框架不仅提供了形式化的数学工具，更为我们提供了对信号本质及其操作的深刻洞察。

Linear Differential Equations

微积分与线性代数的结合在线性微分方程的理论中展现得淋漓尽致。考虑一个描述某种物理量 $x(t)$ 的方程，其中多个导数以线性方式出现：

$$\frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \cdots + a_1\frac{dx}{dt} + a_0 x = 0$$

尽管看似远离本章研究的向量空间，但当我们通过正确的视角来审视时，便会出现一种显著的结构。

让我们采用简洁的符号 $D = d/dt$ 来表示微分算符。我们的方程变为

$$(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)x = 0$$

或者更简单地 $p(D)x = 0$，其中 $p \in \mathcal{P}_n$ 是一个 n 次的多项式。这个算符符号将微分方程转化为代数对象——这是深层次规律的初步暗示。

一个显著的事实，其证明将在后续章节中给出，是该方程的解构成一个维度恰好为 n 的向量空间。

存在 n 个构成一组基的特殊解，所有其他解都可以通过线性组合由它们产生。这一抽象事实对求解此类方程具有深远的实际意义。

当p完全因式分解时，结构最为清晰：

$$p(D) = (D - \lambda_1)(D - \lambda_2) \cdots (D - \lambda_n)$$

其中 λ_1、...、λ_n 是不同的数字，其含义将在第7章中阐明。目前，请注意最简单的情况 $n = 1$ 导出以下方程：

$$(D - \lambda)x = 0 \quad \Rightarrow \quad \frac{dx}{dt} = \lambda x$$

Historical Note: 多项式根与指数解之间的联系最早由欧拉观察到，尽管完整的向量空间结构直到后来才显现出来。

其解 $x(t) = ce^{\lambda t}$ 你肯定还记得，来自微积分。

当所有 λ_i 都互不相同时，函数 $\{e^{\lambda_1 t}$、$e^{\lambda_2 t}$、……、$e^{\lambda_n t}\}$ 构成解空间的一组基。任何解都具有如下形式：

$$x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t} + \cdots + c_n e^{\lambda_n t}$$

其中系数 c_i 由初始条件确定。验证这些指数函数在 λ_i 彼此不同时是线性无关的，为本章的概念提供了一个极好的练习。

例 2.27（质量-弹簧系统）。无阻尼质量-弹簧系统的经典二阶方程，

$$m\frac{d^2x}{dt^2} + kx = 0$$

呈现为 $p(D)x = 0$，且有 $p(D) = mD^2 + k$。将 $\omega = \sqrt{k/m}$ 写作，可因式分解为：

$$p(D) = m(D + i\omega)(D - i\omega)$$

Foreshadowing: 第7章将揭示数字λ_i的更深层含义，并提供系统化的方法，即使在$p(D)$不能如此容易因式分解的情况下也能找到基解。

从而得到基解 $e^{i\omega t}$ 和 $e^{-i\omega t}$。它们的线性组合通过欧拉公式产生了熟悉的正弦运动 $x(t) = A\cos(\omega t) + B\sin(\omega t)$。$\diamond$

这个例子阐明了一个深刻的原理：抽象的数学结构常常在看似无关的情境中显现出来。本章引入的向量空间并非仅仅是形式化的构造，而是描述物理系统的自然语言。这里仅略作暗示的微分方程与线性代数之间的相互作用，将在第7章和第8章中得到大幅深化。

□ ——————————————————————————— □

Exercises: Chapter 2

1. 考虑 $\mathbb{R}^3$ 中的以下向量：

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 4 \\ -1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix}$$

(a) 通过寻找具体的标量 c_1, c_2, c_3（不全为零），使得 $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = 0$，证明它们是线性相关的 (b) 求 $\text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ 的维数 (c) 找到一个张成同一空间的线性无关子集

2. 考虑由 2×2 矩阵组成的 $\mathbb{R}^{2\times 2}$ 空间。证明下面这组矩阵是线性相关的:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

找出它们之间的线性关系。

3. 考虑 $\mathbb{R}^{2\times 2}$ 中的以下矩阵:

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$$

(a) 证明 A_3 位于 $\text{span}\{A_1, A_2\}$ 通过找到特定的标量 c_1，c_2 使得 $A_3 = c_1A_1 + c_2A_2$ (b) 找到矩阵 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$，它位于 $\text{span}\{A_1, A_2\}$ 并满足 $a + d = 1$ 和 $b + c = 3$

4. 考虑 $\mathbb{R}^4$ 中的以下向量:

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ -1 \end{pmatrix}, \quad \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \quad \boldsymbol{v}_4 = \begin{pmatrix} 8 \\ 7 \\ -1 \\ -3 \end{pmatrix}$$

(a) 将 $\boldsymbol{v}_4$ 表示为 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$ (的线性组合 b) 证明 $\{\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3\}$ 线性无关 （c）$\{\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3\}$ 是否构成 $\mathbb{R}^4$ 的一个张成集？如果不是，找出 $\mathbb{R}^4$ 中一个不在它们张成中的向量

5. 在次数至多为 2 的多项式空间 $\mathcal{P}_2$ 中： (a) 将 $p(x) = 2x^2 - x + 3$ 表示为 $q_1(x) = 1 + x^2$、$q_2(x) = x - x^2$ 和 $q_3(x) = 1 - x$ (b) 的线性组合。 (b) 判断 $\{q_1(x)$、$q_2(x)$、$q_3(x)\}$ 是否张成 $\mathcal{P}_2$ (c)。 (c) 如果它们不能张成 $\mathcal{P}_2$，找出 $\mathcal{P}_2$ 中一个不在它们张成空间内的多项式。

6. 考虑次数至多为 2 的多项式向量空间 $\mathcal{P}_2$。证明多项式 1、$1 + x$ 和 $1 + x + x^2$ 张成 $\mathcal{P}_2$。它们是否线性无关?

7. 设 V 为由 3×3 矩阵 A 组成的向量空间，满足 $A^T = A$ (对称矩阵)。 (a) 写出 V (的一个生成集) 证明你的生成集是线性无关的 (c) 确定 dim V ( ) 求……的坐标

$$\begin{bmatrix} 2\,1\,0\,1 & -1\,2\,0\,2\,3 \end{bmatrix}$$

相对于你的生成集

8. 设 V 为 2×2 矩阵的向量空间。证明具有如下形式的矩阵的集合

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}$$

是 V 的一个子空间。它的维数是多少?

9. 判断下述集合是否是定义在 [0, 1] 上的连续函数向量空间的一个子空间:
所有满足 $f(0) = 2f(1)$ 的连续函数 f 的集合。证明你的答案。

10. 证明：对于固定常数 c，所有满足 $p(c) = 0$ 的多项式 $p(x)$ 的集合构成一个向量空间。

11. 对于矩阵

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & -1 \\ -1 & -2 & 3 \end{bmatrix}$$

(a) 求行空间 row(A) 的一组基 (b) 求列的一组基
水疗e col(A) （c）证明 dim(row(A)) = dim(col(A)) 在 thi s 情况（d）
在$\mathbb{R}^3$中找一个不在A的列空间中的向量

12. 考虑矩阵

$$A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 2 & 1 \\ 0 & 3 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -1 & -1 \\ -2 & 4 & 2 \\ 0 & 3 & 1 \end{bmatrix}$$

通过找到将它们的行联系起来的显式线性组合，证明 row(A) = row(B)。这对 c ol(A) 与 col(B) 之间的关系说明了什么?

13. 证明：对于向量空间 V 中的任意向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$，$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ 的张成是包含所有向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ 的 V 的最小子空间。（提示：证明它是一个子空间，包含这些向量，并且包含于任何包含这些向量的其他子空间中。）

14. 设 V 为一个向量空间，$U < V$ 为一个子空间。证明：如果 $\boldsymbol{v} \notin U$，那么 $\{\boldsymbol{v}\} \cup U$ 线性相关当且仅当 $\boldsymbol{v} \in U$。

15. 证明：在任意向量空间 V 中，若 $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v}_k\}$ 张成 V，且 $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m\}$ 在 V 中线性无关，则 $m \leq k$。这对不同的张成集说明了什么?

16. 对于向量空间 V 中的向量 $\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3$，证明或反驳：如果 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 线性无关，而 $\{\boldsymbol{v}_1$、$\boldsymbol{v}_2$、$\boldsymbol{v}_3\}$ 线性相关，则 $\boldsymbol{v}_3$ 必须属于 span$\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$。

17. 设 $U = \{(x、x、0) : x \in \mathbb{R}\}$ 以及 $W = \{(0、y、z) : y、z \in \mathbb{R}\}$ 是 $\mathbb{R}^3$ 的子空间。通过证明以下两点来证明 $\mathbb{R}^3 = U \oplus W$：（i）$\mathbb{R}^3$ 中的每个向量都可以写成来自 U 和 W 的向量之和；（ii）同时属于 U 和 W 的唯一向量是 0。

18. 在次数不超过 2 的多项式空间 $\mathcal{P}_2$ 中，令 U 为偶多项式的子空间（其中 $p(-x) = p(x)$），令 W 为奇多项式的子空间（其中 $p(-x) = -p(x)$）。证明 $\mathcal{P}_2 = U \oplus W$。 19. 令 $V = \mathbb{R}^{2\times 2}$ 为 2×2 矩阵的空间。令 U 为上三角矩阵的子空间，W 为严格下三角矩阵的子空间。通过在 U 和 W 中找到一个非零矩阵，证明 $V \neq U \oplus W$。 20. 令 $V = \mathcal{P}_3$，并定义 $U = \{p \in \mathcal{P}_3 : p(0) = 0\}$，以及 $W = \{p \in \mathcal{P}_3 : p(x) = c$，其中 $c\}$ 为某个常数。证明 $V = U \oplus W$。 21. 令 U 和 W 为 $\mathbb{R}^3$ 的子空间。证明如果 $\dim U + \dim W > 3$，则 U 和 W 不能构成直和（也就是说，它们必须具有非零交集）。 22. 令 U 和 W 为向量空间 V 的子空间，使得 $V = U \oplus W$。如果 $\boldsymbol{v} \in V$，证明带有 $\boldsymbol{u} \in U$ 和 $\boldsymbol{w} \in W$ 的表达式 $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$ 必须是唯一的。 23. 令 U_1、U_2、U_3 为向量空间 V 的子空间。证明如果 $V = U_1 \oplus U_2 \oplus U_3$，则任意向量 $\boldsymbol{v} \in V$ 都可以唯一地表示为 $\boldsymbol{v} = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3$，其中 $\boldsymbol{u}_i \in U_i$。

24. 设 U 和 W 是向量空间 V 的有限维子空间。证明:

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$$

Chapter 3
Linear Transformations

"he became what he beheld; he became what he was doing; he was himself transform'd"

数学的本质不在于对象，而在于它们之间的变换。我们迄今为止研究的向量和空间，只有在受到作用时才会“活”起来——被旋转、缩放、投影或以其他方式变换。这样的变换是我们名词的动词，是赋予我们数学宇宙生命的运算。线性变换是那些保持向量空间的基本运算：加法和缩放的变换。这个看似简单的要求——即我们的变换必须尊重向量空间结构——却导致了一个异常丰富的理论，并具有深远的实际意义。

向量空间单独来看是静态的集合。当我们考虑从输入信号到输出响应，从配置到力，从高维数据到低维表示的映射时，线性代数的力量才得以显现。这样的映射，当是线性时，具有一种美妙的结构，既能揭示其理论属性，又能使其实际计算成为可能。

我们的旅程从熟悉的矩阵变换开始，随后逐步升华到更抽象的高度。尽管微分和积分操作远离几何学，它们与矩阵的相关操作在结构上有着深刻的相似性。这种抽象揭示了与任何线性变换相关的四个基本空间——核空间和像空间，分别捕捉消失和得到的部分，余像和余核则衡量效率和缺陷。这些看似不同的空间，通过线性代数的基本定理联系在一起，该定理将线性变换的代数、几何和维度方面统一成一个连贯的整体。

3.1 *Euclidean Transformations*

我们的故事始于熟悉的领域——矩阵在欧几里得空间中作用于向量。从多变量微积分中，我们回想起矩阵 A 乘法如何变换 $\mathbb{R}^n$ 中的向量，将输入向量 $\boldsymbol{x}$ 转换为输出 $A\boldsymbol{x}$。尽管我们机械地执行了这些运算，按列与行计算乘积，但这些变换蕴含着丰富的几何内容，值得在抽象之前细细品味。

最简单的这类变换是对空间进行均匀缩放。矩阵 cI 将每个坐标乘以标量 c，从而以原点为中心对空间进行扩张或收缩。更有趣的是在不同方向上以不同方式缩放的矩阵：

$$\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ y/2 \end{pmatrix}$$

这种变换沿一个轴拉伸空间，同时沿另一个轴压缩——就像一面幽默屋镜子的扭曲被精确地呈现于坐标中。

平面内的旋转由如下形式的矩阵产生

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

绕原点将向量逆时针旋转角度 θ。此类矩阵能够保持长度和角度，这是其结构的结果——见第5章。

更微妙的是 *shear transformations*，例如

$$\begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}$$

Question: What happens with the transpose of this horizontal shear?

它们会按其高度成比例地偏移每一条水平线。这些在倾斜垂直线的同时保持面积——就像一副纸牌被小心地在桌面上滑动。

这些基本变换——缩放、旋转和剪切——结合起来可以生成平面中的所有线性变换。任何 2×2 矩阵都可以理解为这些基本几何操作的组合。这种分解预示着即将出现的更深层结构，当我们学会将矩阵分解为更简单的组成部分时。

这些变换除了通过矩阵乘法实现外，还具有哪些共同特征？首先，它们保持原点不变——零向量保持固定。其次，它们尊重向量加法：和的像等于像的和。第三，它们自然地相互作用。

在标量乘法下：将输入向量加倍，其像也随之加倍。这些性质——在矩阵情形下看似显而易见——将构成我们抽象理论的支架。

Foreshadowing: 我们在矩阵变换中观察到的性质——向量运算的保持——将在抽象设定中定义线性性。

还要考虑这些变换会破坏什么。旋转保持距离不变，但会改变坐标。剪切保持面积不变，却会扭曲角度。缩放既改变距离也改变面积，但保持通过原点的直线不变。对几何特征的这种选择性保持暗示着更深层的不变量——在某些变换类别下保持不变的量或性质。

矩阵变换 $A\boldsymbol{x}$ 将关于空间变换的几何直觉转化为对坐标的代数操作。当我们把这些思想提升到抽象向量空间时，几何与代数之间的这种相互作用仍将存在。尽管我们可能失去直接可视化变换的能力，但核心思想——向量运算的保持、不变量的研究、分解为更简单的部分——将指引我们的发展。

3.2 *Definitions & Implications*

我们对欧几里得变换的经验表明，刻画线性本质的关键特征是：对加法和数乘的保持。许多重要的变换具有这些代数性质，却缺乏明显的几何解释。这促使我们从几何中抽象出来，研究任意向量空间之间的线性变换。

定义 3.1（线性变换）。设 V 和 W 为向量空间。*linear transformation* $T: V \to W$ 是一个满足两个性质的函数：

1. 可加性：$T(\boldsymbol{v_1} + \boldsymbol{v_2}) = T(\boldsymbol{v_1}) + T(\boldsymbol{v_2})$ 对所有 $\boldsymbol{v_1}$、$\boldsymbol{v_2} \in V$ 2
. 齐次性：$T(c\boldsymbol{v}) = cT(\boldsymbol{v})$ 对所有 $c \in \mathbb{R}$ 和 $\boldsymbol{v} \in V$

•

这两个性质结合起来，确保线性变换保持线性组合。这一看似简单的要求却具有深远的影响。

引理 3.2。*A linear transformation* T: $V \to W$ *satisfies:*

1. $T(0) = 0$
2. $T(-\boldsymbol{v}) = -T(\boldsymbol{v})$ *for all* $\boldsymbol{v} \in V$
3. T *preserves linear combinations*

$$T\left(\sum_{i=1}^{n} c_i \boldsymbol{v_i}\right) = \sum_{i=1}^{n} c_i T(\boldsymbol{v_i})$$

线性组合的保持性对子空间有一个重要的推论：线性变换将子空间映射为子空间。

Foreshadowing: 线性组合的保持将是理解线性变换如何与基和坐标系相互作用的关键。

引理 3.3。*If* $T\colon V \to W$ *is linear and* $U < V$*, then* $T(U) < W$*.*

Examples of Linear Transformations

线性变换最简单的例子是上一节中的那些——任意矩阵 A 都通过线性变换 $T_A(\boldsymbol{x}) = A\boldsymbol{x}$ 作用于欧几里得向量。这是如此自然，以至于几乎不值得把它与矩阵本身区分开来。然而，并非所有线性变换在坐标中都如此显式。请考虑以下一些几何性较弱的例子。

例 3.4（微分）。考虑微积分中的微分算子 $D = d/dx$。它满足线性性，即对于可微函数 f 和 g，有 $D(f+g) = Df + Dg$；并且当 c 为一个标量时，有 $D(cf) = c\,Df$。因此，它定义了一个从 $C^\infty(\mathbb{R})$ 到其自身的线性变换。如果偏好有限维向量空间，可以将其限制在多项式上，在这种情况下 $D\colon \mathcal{P}_n \to \mathcal{P}_{n-1}$：

注意到常数多项式的子空间被映射为零。$\mathcal{P}_n$ 的其他子空间会发生什么?

$$D\left(\sum_{i=0}^{n} c_i x^i\right) = \left(\sum_{j=1}^{n} j c_j x^{j-1}\right).$$

在这里我们看到没有几何的线性——和的导数是导数的和，常数可以从导数中提取出来。 ◇

例 3.5（积分）。定积分算子 $\mathrm{I} : C([a,b]) \to \mathbb{R}$ 定义为

$$\mathrm{I}(f) = \int_a^b f(x)dx$$

是线性的，就像求导一样。当我们把注意力限制在多项式以及多项式的子空间上时，会发生什么? ◇

线性变换与线性无关性之间的关系直指其结构的核心。

定义 3.6（单射与满射）。线性变换 $T : V \to W$ 是：

1. *injective* (或 *one-to-one*)，如果不同的输入产生不同的输出：$T(\boldsymbol{v_1}) = T(\boldsymbol{v_2})$ 蕴含 $\boldsymbol{v_1} = \boldsymbol{v_2}$ 2. *surjective* (或 *onto*)，如果 W 中的每个向量都是 V 中某个向量的像：对于每个 $\boldsymbol{w} \in W$，存在 $\boldsymbol{v} \in V$ 使得 $T(\boldsymbol{v}) = \boldsymbol{w}$

•

引理 3.7。 *For a linear transformation $T : V \to W$:*

1. *T is injective if and only if it preserves linear independence*
2. *T is surjective if and only if $T(V) = W$*
3. *T is invertible if and only if it is both injective and surjective*

Caveat: 线性变换可能以两种不同的方式不可逆：要么将不同的向量映射到同一像（非单射），要么在目标空间中遗漏某些向量（非满射）。

这些抽象的力量在于其应用范围之广。无论是变换几何向量、多项式，还是函数，其行为都受同样的原则支配。

3.3 *Isomorphisms*

两个向量空间在本质上何时是相同的？在 $\mathbb{R}^2$ 中的几何向量看起来与线性多项式 $ax + b$ 大不相同，然而二者都允许相同的运算并满足相同的规则。这样的观察促使我们去考察，从线性代数的角度来看，向量空间不可区分究竟意味着什么。

线性变换是两个向量空间之间的定向映射。给定 $T : V \to W$，我们称 V 为 *domain*，并称 W 为 *codomain*。为了作为空间之间的完美字典，T 必须同时具备定义 3.6 中引入的两种性质：它必须既是单射又是满射。这样的变换称为同构：

Nota bene: codomain 的使用可能并不熟悉，因为它源自范畴论。表示单射和满射映射的术语 *monomorphism* 和 *epimorphism* 也是如此，不过我们将不使用这些特定术语。

定义 3.8（同构）。如果在向量空间 V 与 W 之间存在一个 *isomorphism*——一个既是单射又是满射的线性变换，则称它们是 *isomorphic*，记作 $V \cong W$。

•

例 3.9（坐标向量）。变换 $T : \mathbb{R}^2 \to \mathcal{P}_1$ 通过将向量映射为线性多项式通过

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto a + bx$$

它是一个同构。它在几何向量与线性多项式之间提供了一个完美的对应关系，保留了所有向量空间运算。向量的加法对应于多项式的加法；向量的数乘意味着多项式的数乘。

◇

例 3.10（矩阵表示）。2×2 矩阵的空间 $\mathbb{R}^{2\times 2}$ 通过该变换与 $\mathbb{R}^4$ 同构

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

尽管我们通常把矩阵写成方阵的形式，但它们在本质上与向量并无不同——只是以不同的方式对同样的信息进行封装。

◇

Nota bene: 矩阵乘法对向量空间结构而言是不可见的。

并非所有线性变换都能实现这种完美的对应关系。由下式给出的投影 $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$

$$\Pi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

是满射但不是单射——它到达平面中的每一个点，但会将仅在其 z-坐标上不同的所有点折叠为同一点。相反，由下式给出的嵌入 $\iota : \mathbb{R}^2 \to \mathbb{R}^3$

$$\iota \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

是单射但不是满射——它保留了平面中关于向量的所有信息，但遗漏了大部分$\mathbb{R}^3$。

Question: 任何两个具有相同维度的向量空间实际上是同构的吗?

这些例子揭示了一个深刻的真理：不同维度的向量空间无法同构。上面的投影表明，较大的空间无法注入到较小的空间中而不发生塌缩；而嵌入则表明，较小的空间无法满射到较大的空间中而不丢失向量。这个观察——尽管尚未证明——暗示了线性代数中维度的基本性质。

同构的语言不仅仅提供分类，它还提供了一种关于向量空间中哪些特征真正重要的视角。当空间是同构时，我们可以自由地在它们之间转换问题，选择最方便的表示方式。只要我们构建了它们之间的正确词典，$\mathbb{R}^n$的几何直觉就可以应用于多项式、矩阵或信号空间。

例 3.11（多项式导数）。由下式给出的微分算符 $D : \mathcal{P}_2 \to \mathcal{P}_1$

$$D(ax^2 + bx + c) = 2ax + b$$

是满射但不是单射。每个线性多项式都是导数（满射），但常数在微分下消失（非单射）。

◇

3.4 *Image & Kernel*

我们在第二章中遇到的子空间来源于单一向量空间内的运算。线性变换生成它们自己的特征子空间——无论是在定义域还是值域。这些子空间捕捉了变换作用的基本特征，衡量了变换的有效性和缺陷。

在向量空间之间进行线性变换 $T : V \to W$ 的固定。第一个感兴趣的子空间位于余域中。

定义 3.12（像）。T 的 *image*，记作 $\operatorname{im} T$，是与所有可能输出相关的目标空间的子空间：

$$\operatorname{im} T = \{T(\boldsymbol{v}) : \boldsymbol{v} \in V\} < W \tag{3.1}$$

•

这确实是 W 的一个子空间：零向量显然在像中（因为 $T(0) = 0$），如果 $T(\boldsymbol{v_1})$ 和 $T(\boldsymbol{v_2})$ 是像中的任意向量，那么它们的和 $T(\boldsymbol{v_1}) + T(\boldsymbol{v_2}) = T(\boldsymbol{v_1} + \boldsymbol{v_2})$ 也在像中，任何标量倍数也是如此。像度量了变换的“范围”—— W 的多少可以作为输出达到。

定义 3.13（核）。*kernel* (或 *nullspace*) 的 T，记作 $\ker T$，是域的子空间，包含所有在 T 下消失的向量：

$$\ker T = \{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \boldsymbol{0}\} < V \tag{3.2}$$

•

这个同样构成V的一个子空间是显而易见的：零向量显然在核中；如果$\boldsymbol{v_1}$和$\boldsymbol{v_2}$在核中，那么$T(\boldsymbol{v_1} + \boldsymbol{v_2}) = T(\boldsymbol{v_1}) + T(\boldsymbol{v_2}) = 0$，标量倍数的情况类似。核捕捉到的是变换无法“看见”的东西——在其作用下消失的向量。

这些抽象定义凝结了我们早期在线性系统方面的工作。考虑第1章中的矩阵方程 $A\boldsymbol{x} = \boldsymbol{b}$。这定义了一个线性变换 $T_A : \mathbb{R}^n \to \mathbb{R}^m$ 通过 $T_A(\boldsymbol{x}) = A\boldsymbol{x}$ 。关于该系统的标准问题现在具有几何意义：

1. 解是否存在？当且仅当 $\boldsymbol{b} \in \operatorname{im} T_A$ 时存在。 2. 解是否唯一？当且仅当 $\ker T_A = \{0\}$ 时唯一。 3. 如果存在多个解，它们之间如何关联？它们的差异由 $\ker T_A$ 的元素决定。

例 3.14（微积分算子）。微分算子 $D : C^1([a, b]) \to C([a, b])$ 的核由所有常数函数组成

$[a, b]$——这些正是在微分下消失的函数。它的像由所有作为导数出现的连续函数组成，$C([a, b])$ (的一个真子空间，并非每个连续函数都是一个导-数)。

定积分算子 $\mathrm{I} : C([a, b]) \to \mathbb{R}$ 定义为 $\mathrm{I}(f) = \int_a^b f(x)dx$，其（非常大的）核由所有在 $[a, b]$ 上积分为零的函数组成。它的像是整个 $\mathbb{R}$ ——任何实数都可以表示为某个连续函数的积分。

◇

Foreshadowing: 核与像的维数之间的关系将被证明是理解线性变换的基础。本章末尾的基本定理刻画了消失之物与所得之物之间的这种平衡。

这些子空间提供了分析线性变换结构的第一工具。一个变换当且仅当其核只包含零向量时是单射；当且仅当其像是整个陪域时是满射。这些子空间之间的相互作用——它们的维度如何平衡，它们如何分解涉及的空间——引出了接下来的更深层次理论。

3.5 *Rank & Nullity*

向量空间的维数刻画了其规模和复杂性。对于线性变换，我们也寻求类似的规模与复杂性的度量——不是针对单个空间，而是针对该变换在空间之间的作用方式。这些度量自然地来自像空间与核空间的维数。

定义 3.15（秩与零度）。线性变换 $T : V \to W$ 的 *rank* 是其像的维数：

$$\operatorname{rank} T = \dim(\operatorname{im} T) \tag{3.3}$$

T的*nullity*是其核的维数：

$$\operatorname{null} T = \dim(\ker T) \tag{3.4}$$

•

Think: 秩度量变换的有效维数——就像光穿过晶体，一部分被透射，一部分被吸收，它统计的是那些得以存续的独立方向。零度则通过统计丢失的维数来补充这一度量，那些方向如同被完全吸收的光，在变换中彻底消失。

这将定义1.10中关于矩阵秩的朴素概念加以抽象。当T由矩阵A表示时，抽象秩与具体秩一致。

例 3.16（矩阵的秩与零度）。对于一个秩为 2 的 3×4 矩阵 A，变换 $T_A : \mathbb{R}^4 \to \mathbb{R}^3$ 具有：

1. 秩 $T_A = 2$，意味着 $\operatorname{im} T_A$ 是 $\mathbb{R}^3$ 2 中的一个平面 2. 零空间 $T_A = 2$，因为解 $Ax = 0$ 得到一个二维解空间 3. 维数$(\ker T_A)$ + 维数$(\operatorname{im} T_A)$ = 维数$(\mathbb{R}^4) = 4$

这一最后的观察暗示了秩与零度之间更深层的关系。

◇

例 3.17（微积分算子）。微分算子 $D : \mathcal{P}_n \to \mathcal{P}_{n-1}$ 具有：

1. 秩 $D = n$，因为 $\mathcal{P}_{n-1}$ 中的每个多项式都是一个导数2. 零度 $D = 1$，因为只有常数函数在求导下为零3. $\dim(\ker D) + \dim(\text{im } D) = \dim(\mathcal{P}_n) = n + 1$

我们再次看到各个维度保持平衡。 ◇

例 3.18（积分）。考虑定积分算子 $\mathrm{I}\colon C[0,1] \to \mathbb{R}$，由 $\mathrm{I}(f) = \int_0^1 f(x)dx$ 定义。尽管 $C[0,1]$ 是无限维的：

1. rank I = 为 1，因为其像是整个 $\mathbb{R}$ 2. null I 是无限维的，包含所有积分为零的函数 3. 在无限维情形下，维数平衡被打破

Caveat: 在无限维情形下，秩与零度之间的关系变得更加微妙。这里的例子旨在帮助在有限维情况下建立直觉。

◇

这些例子表明，秩、零度以及定义域和值域的维数之间存在着深刻的联系。

3.6 *Quotients*

线性变换不仅通过它们所保持的东西揭示结构，也通过它们所压缩的东西揭示结构。不同向量映射到相同输出的方式暗示了一种自然的组织——将以相同方式变换的向量分组。这一洞见引出了线性代数中最为深刻的构造之一：商空间。

定义 3.19（商空间）。设 V 为一个向量空间，$U < V$ 为其子空间。*quotient space* V/U 是一个向量空间，其元素是 V 中向量的等价类，其中当且仅当向量 v_1 与 $v_2 \in V$ 的差属于 U 时，它们等价：

$$v_1 \sim v_2 \iff v_1 - v_2 \in U$$

$v \in V$ 的等价类，记作 $[v]$，由所有与 v 等价的向量组成：

$$[v] = \{w \in V : w - v \in U\} = v + U$$

V/U 上的向量运算通过代表元来定义：$[v_1] + [v_2] = [v_1 + v_2]$ 以及 $c[v] = [cv]$，对于标量 c。 •

子空间的性质确保 $\sim$ 定义了一个恰当的等价关系：自反性（$v \sim v$）、对称性（$v_1 \sim v_2$ 蕴含 $v_2 \sim v_1$）以及传递性（$v_1 \sim v_2$ 和 $v_2 \sim v_3$ 蕴含 $v_1 \sim v_3$）。

商空间的物理直觉在电网络中自然地显现出来。考虑一个具有 n 个节点的电路，在其中我们测量电压

节点对之间的差异。尽管每个节点都有其自身的电势，但在物理上有意义的测量始终是差值——给每个节点都加上一个常数电压并不会改变任何测量结果。这一观察揭示了商在物理学中的基础性作用。

令 $V = \mathbb{R}^n$ 为节点电压赋值的向量空间。将每个电压配置映射到常数平移下其等价类的投影 Π:

$$\Pi : V \to Q \quad : \quad \boldsymbol{v} \mapsto [\boldsymbol{v}]$$

其核恰好由恒定电压平移组成（c，c，……，c）^{T}。商空间 $V/\ker(\Pi)$ 随后刻画了物理上有意义的电压状态，剥离了它们对参考电势的人为依赖。

例 3.20（核商）。设 $T : V \to W$ 为一个线性变换。相差一个 $\ker T$ 中元素的两个向量被映射到同一个输出:

$$T(\boldsymbol{v_1}) = T(\boldsymbol{v_2}) \iff \boldsymbol{v_1} - \boldsymbol{v_2} \in \ker T$$

商空间 $V/\ker T$ 自然地表示 T 的“有效”输入空间——它刻画了 T 无法区分的输入。 ◇

Foreshadowing: 当我们引入内积时，每个等价类都会有一个唯一的代表元，它与被取商的子空间正交。现在，我们先处理这些等价类本身。

商空间从 V 继承了向量空间结构。加法和标量乘法在等价类上定义:

$$[\boldsymbol{v_1}] + [\boldsymbol{v_2}] = [\boldsymbol{v_1} + \boldsymbol{v_2}] \quad \text{and} \quad c[\boldsymbol{v}] = [c\boldsymbol{v}]$$

这些运算是良好定义的——与我们从等价类中选择哪些代表元无关。

例 3.21（几何商）。首先考虑将 $\mathbb{R}^3$ 按通过原点的一条直线 L 取商。两个点 $\boldsymbol{p}$、$\boldsymbol{q} \in \mathbb{R}^3$ 恰当且仅当它们的差 $\boldsymbol{p} - \boldsymbol{q}$ 属于 L 时属于同一个等价类——也就是说，当它们相差某个与 L 平行的向量时。因而，每个等价类都形成一张与 L 平行的平面，因为沿着 L 平移会使我们保持在同一类中。商空间 $\mathbb{R}^3/L$ 可以被形象地看作所有这类平行平面的集合，并且自然地由一个与 L 垂直的二维平面进行参数化。尽管是抽象的，这个商的维数恰好为 2，因为一旦固定其方向，指定一张平面只需要两个坐标。

现在改为考虑用过原点的平面 P 对 $\mathbb{R}^3$ 取商。等价类是与 P 垂直的直线——相对于 P，每个点都与所有在其正上方或正下方的点等价。商空间 $\mathbb{R}^3/P$ 自然成为一维的，由……参数化

沿着 P 的法向量方向的有符号距离。这个距离为抽象商提供了一个具体的实现：当且仅当两个点在 P 的法向量上的投影相同，或者等价地，当它们沿着彼此平行的垂线与 P 的距离相同，它们才是等价的。◇

商空间的维数同时反映了原空间的维数以及被取商的子空间的维数：

$$\dim(V/U) = \dim V - \dim U$$

这种维度关系对于理解线性变换如何分解空间至关重要。

例子 3.22（积分商）。考虑不定积分算子（或 *antidifferentiation*）D^{-1} 作用于区间上的连续函数 $C([a,b])$。一个原函数总是存在（多亏了 FTIC），但只有在加上常数的情况下才是良好定义的。

You did not forget the $+C$ did you?

要使其成为向量空间之间的线性变换，需要在陪域中使用商空间。令 $U < C([a.b])$ 表示该区间上常值函数的子空间。则商空间 $C([a,b])/U$ 由函数之差为常数的等价类组成。不定积分现在是一个线性变换

$$D^{-1} : C([a,b]) \to C([a,b])/U.$$

◇

例 3.23（平移不变性）。考虑 $\mathbb{R}^d$ 中包含 n 个点的数据集，其以矩阵 $X \in \mathbb{R}^{d\times n}$ 的列表示。在许多应用中，如聚类或模式识别，我们关注的是点之间的相对位置，而非它们在空间中的绝对位置。

令 $V = \mathbb{R}^{d\times n}$ 为所有可能数据集的向量空间。均匀平移的子空间 $U < V$ 由所有列相同的矩阵组成：

$$U = \{\begin{bmatrix} v & v & \cdots & v \end{bmatrix} : v \in \mathbb{R}^d\} = \{v \cdot \mathbf{1}^T : v \in \mathbb{R}^d\}$$

其中 $1 \in \mathbb{R}^n$ 是全为一的向量。U 中的每个矩阵表示通过某个向量 $v \in \mathbb{R}^d$ 对零矩阵的均匀平移。

如果两个数据集 X、$Y \in V$ 的差 $X - Y$ 属于 U，则它们在平移意义下等价——也就是说，存在某个向量 $v \in \mathbb{R}^d$，使得将 v 平移到 X 的每一列上即可得到 Y 中对应的列：

$$X \sim Y \iff X - Y \in U \iff X - Y = v \cdot \mathbf{1}^T \text{ for some } v \in \mathbb{R}^d$$

商空间 V/U 则表示数据集模翻译——它捕捉点配置的内在形状，同时忽略它们在空间中的绝对位置。维度计数揭示了结构：

这给出了V/U与均值为零的数据集子空间之间的显式同构。

- $\dim(V) = dn$ ($\mathbb{R}^d$) 中 n 点的坐标
- $\dim(U) = d$ (平移向量的坐标)
- $\dim(V/U) = dn - d$ (数据集至翻译)

◇

商空间提供了一种正式的方法，用于识别在某些操作下表现相似的向量。当我们通过变换的核对一个域进行商化时，我们得到一个空间，它忠实地表示变换的作用，同时去除冗余。随着我们开发更复杂的工具来分析线性变换，这种观点将证明是无价的。

3.7 *Coimage & Cokernel*

线性变换的图像和核只讲述了一半的故事。正如商空间通过识别行为相似的向量来揭示结构，我们可以通过考察定义域和余域中的商来阐明线性变换的作用。这导致了两个额外的空间，完善了我们对结构的理解。

定义 3.24（余像）。给定一个线性变换 $T : V \to W$，T 的 *coimage* 是商空间

$$\operatorname{coim} T = V / \ker T \tag{3.5}$$

•

Example: 对于一个秩为2的矩阵 $A : \mathbb{R}^4 \to \mathbb{R}^3$，其共影像是二维的，表示影响输出的两个独立输入方向。

共像表示T的“有效”输入空间——它识别T无法区分的输入。投影$\Pi : V \to$ 共像T将每个向量映射到其等价类（模$\ker T$）。

共象有一个自然的解释：它衡量有多少独立的输入方向实际上影响输出。当两个向量在 V 中精确地通过 $\ker T$ 的元素不同，它们会产生相同的输出——因此 $\operatorname{coim} T$ 对 T 所看到的真正不同的输入进行参数化。

我们的第四个基本空间是最隐秘、最晦涩的：

定义 3.25（余核）。给定一个线性变换 $T : V \to W$，*cokernel* 的 T 是商空间

$$\operatorname{coker} T = W / \operatorname{im} T \tag{3.6}$$

•

余核衡量 T 未能达到 W 全体的程度。当 T 是满射时，coker T 是平凡的；否则，它刻画了陪域中的“不可见”空间。

例 3.26（微分算子）。对于导数算子 $D : \mathcal{P}_n \to \mathcal{P}_{n-1}$：1. 余像是（$n$）维的，因为只有常数在 D 下消失 2. 余核是平凡的，因为 $\mathcal{P}_{n-1}$ 中的每个多项式都是一个导数 对于积分 $\mathrm{I} : C[0, 1] \to \mathbb{R}$：1. 余像刻画了相差超过其平均值的函数 2. 余核是平凡的，因为每个实数都是一个积分

◇

Foreshadowing: 这些空间的维度并非相互独立——它们满足一种优美的关系，下一节将揭示这一点。

这四个基本空间——核、像、余核和余像——构成了线性变换的一整套结构不变量。它们看似复杂的关系，将在下文的基本定理中清晰呈现。每一个都衡量了 T 改变空间方式的不同侧面：

- ker T 刻画了消失的部分
- im T 展示了什么是可以实现的
- coim T 揭示独立的输入
- coker T 衡量缺失的输出

舞台已经搭建完毕，可以更深入地理解这些空间如何彼此契合——一种统一性的阐释，不仅解释这些空间是什么，还解释它们为何必须存在以及它们如何相互关联。

3.8 The Fundamental Theorem

数学通过统一性获得清晰。我们所遇到的各种结构——像与核、商、余核与余像——不仅彼此相关，而且在根本上相互交织。它们织就的整体是线性代数中最为优美的成果之一，揭示了线性变换的代数性质与维数性质之间的深刻联系。

考虑有限维向量空间之间的一个线性变换 $T : V \to W$。通过我们的探索，我们发现了四个基本子空间：

1. 像 im T，包含所有可能的输出 2. 核 ker T，包含所有被映射为零的输入 3. 余像 coim T，刻画独立的输入方向 4. 余核 coker T，度量不满射的程度

Example: 对于到 xy-平面的投影 $T : \mathbb{R}^3 \to \mathbb{R}^2$，其核是 z-轴，像是 $\mathbb{R}^2$，余像是定义域中的 xy-平面，而余核是平凡的。

这些空间看似彼此独立，却由一个深刻的定理紧密联系在一起，它不仅解释了它们如何关联，也解释了它们为什么必须如此关联。这就是线性代数基本定理：

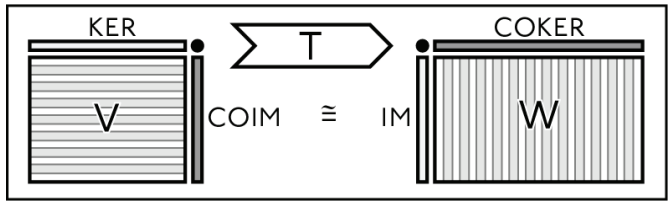
定理 3.27（线性代数基本定理）。 *For any linear transformation $T : V \to W$ between finite-dimensional vector spaces, the following relationships hold and are equivalent:*

1. *The domain and codomain decompose as direct sums:*

$$V \cong \ker T \oplus coim\, T \quad and \quad W \cong im\, T \oplus coker\, T$$

2. *The coimage and image are naturally isomorphic: $coim\, T \cong im\, T$*

"Four mighty ones there are in every Man; a Perfect Unity"



当转换到维度时，基本定理有时被称为 *Rank-Nullity Theorem*：

推论 3.28（秩-零度定理）。 *For a linear transformation between finite-dimensional vector spaces, the dimensions balance in complementary pairs:*

$$\dim V = \dim(\ker T) + \dim(coim\, T) \quad and \quad \dim W = \dim(im\, T) + \dim(coker\, T)$$

Furthermore, the rank connects domain and codomain:

$$\dim(coim\, T) = rank(T) = \dim(im\, T)$$

Otherwise said:

$$\dim V = null(T) + rank(T)$$

例3.29（矩阵秩）。当 T 用矩阵 A 表示时，这些关系解释了为什么：

1. null(A) 与 rank(A) 之和等于 A 的列数 2. row(A)) 的 *row rank* (= 维度和 col(A)) 的 *column rank* (= 维度都等于 rank(A)

Foreshadowing: 当我们在第5章引入内积时，这些代数关系将获得额外的几何意义。

这些来自矩阵代数的熟知事实，是由基本定理所保证的更深层结构关系的体现。 ◇

该定理具有直接的实际意义。在求解线性系统 $T\boldsymbol{x} = \boldsymbol{b}$ 时，我们现在理解到：

1. 当且仅当 $\boldsymbol{b} \in \text{im}(T)$ 时，解存在 2. 当解存在时，其它解与之相差 ker(T) 的元素 3. 任意解都可以唯一地分解为来自 coim(T) 和 ker(T) 的部分 4. 解存在的障碍在于 coker(T)

例 3.30（求解线性系统）。考虑求解 $A\boldsymbol{x} = \boldsymbol{b}$，其中 A 是秩为 2 的 3×4。基本定理告诉我们：

1. 零空间的维数为2 2. 像空间的维数为2 3. 只有当 $\boldsymbol{b}$ 位于一个二维子空间中时才存在解 4. 当解存在时，它们构成一个二维仿射空间

Foreshadowing: 当我们引入内积时，这些分解将为在不存在精确解的情况下寻找最优近似解奠定基础。

◇

基本定理不仅仅是一组关系——它是一种透镜，通过它我们来审视线性变换。无论是在分析电网络、处理信号，还是将模型拟合到数据上，这些结构性关系都引导着我们的理解并指导我们的计算。它所保证的分解在后续章节中与几何结构相结合时，将展现出更为强大的力量。

Graph Topology & Network Structure

当一座城市的电力网络发生故障时，工程师必须迅速识别哪些社区仍然保持连通，以及备用通路位于何处。类似的问题在各种网络中都会出现：信号能否到达电路中的所有神经元？信息能否在社交网络中可靠地流动？计算机网络是否包含可绕过故障的冗余路径？这些关于连通性与韧性的实际关切，共同体现出一种深刻的数学结构，而这种结构通过将线性代数谨慎地应用于网络拓扑而显现出来。

考虑一个有限有向图 $G = (V、E)$，其顶点集为 $V = \{v_1, \ldots, v_n\}$，边集为 $E = \{e_1, \ldots, e_m\}$。每一条边 e 都具有指定的方向，具有起始顶点 e^- 和终止顶点 e^+。从这一离散结构中，我们构造两个基本的向量空间：

- $C_0(G)$: 以顶点 V 为基的向量空间
- $C_1(G)$：以定向边 E 为基元素的向量空间

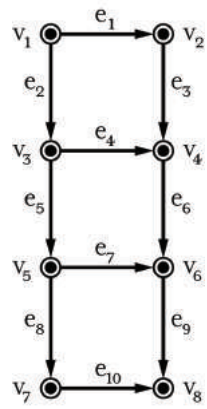
这些空间分别具有维数n和m。尽管它们的元素可以被解释为对顶点或边赋值的数（如电压或电流），但将它们视为抽象向量空间可以澄清它们的基本

结构。

这些空间之间的关系通过一个称为 *boundary operator* $\partial : C_1(G) \to C_0(G)$ 的自然变换显现出来。在基元素上，该算子作用为：

$$\partial(e) = e^+ - e^-$$

线性地扩展到整个 $C_1(G)$。尽管其定义依赖于边的取向，但其基本性质——通过核与余核来刻画——被证明与这些选择无关。



例 3.31（梯形网络）。考虑一个具有八个顶点和十条边的“梯形”网络，其排列与标注如右图所示。边界算子具有如下的矩阵表示：

$$\partial = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

这个网络包含许多环，但只有三个可以被选择为 *independent*。一个简单的代表性环跟随顶部的方形 $e_2 + e_4 - e_3 - e_1$，其中的负号表示穿越与其方向相反的边。这个，连同另外两个显而易见的方形，构成了 $\ker \partial$ 的基础。$\diamond$

这样的 *cycles* — $\ker(\partial)$ 的元素 — 表示通过网络的封闭路径，在每个顶点处“流入”与“流出”相等。并非 $\ker(\partial)$ 的每个元素都对应一个简单的循环；有些代表循环的组合。这个核的维度，记作 β_1，计算了 *independent* 个循环的数量 — 即那些不能表示为更小循环组合的循环。具有更大 β_1 的电网提供更多的备份路径；具有独立循环的神经电路能够维持更复杂的递归模式。

Nota bene: 符号 β 代表 *Betti number*，这是代数拓扑学中的一个基本研究对象。

∂的余核通过其商空间$C_0(G)/\mathrm{im}(\partial)$揭示了互补结构。该空间有效地识别了可以通过网络路径相互到达的顶点。其维度β_0计数了网络的连通分量。具有$\beta_0 > 1$的电网具有需要立即关注的断开区域；具有多个分量的神经网络表示独立的处理模块。

这些数字满足一个显著的关系：

$$\beta_1 - \beta_0 = m - n \tag{3.7}$$

Question: 如果你细分每条边，在每条边的中间添加一个新顶点，并将边分成两部分，会发生什么？

也就是说，独立环的数量减去连通分量的数量等于边数与顶点数之间的过剩。这个方程 —— 同时是 ∂ 的秩-零化定理和图的组合不变量 —— 表达了环与分量之间的基本平衡。添加边通常会创造环（β_1 增加），同时合并分量（β_0 减少）。

例 3.32（点云数据）。现代数据分析通常以从某个基础形状或流形中采样的点开始。给定在 $\mathbb{R}^d$ 中的点 $\{x_1, ..., x_N\}$，我们可以通过连接彼此距离不超过 ε 的点来构建一个图。通过 β_0 和 β_1 测量得到的网络拓扑揭示了基础数据的基本特征：

- β_0 计数数据中的聚类
- β_1 检测孔和循环
- 这些数字对ε的依赖性表征了尺度

这构成了现代*Topological Data Analysis*的基础，其中这些特征在不同尺度上的持续性表明了真实的结构，而非噪声。$\diamond$

该框架的实际价值在于它将关于网络结构的定性问题简化为系统的线性代数。从最初对核和像的抽象研究，到最终作为理解网络的实际工具，这一过程逐渐展开。该基本定理揭示了这些空间之间的深层关系，为分析复杂网络提供了数学基础。

□ ——————————————————————— □

Exercises: Chapter 3

1. 设 $T : \mathbb{R}^2 \to \mathbb{R}^2$ 为由 $T(x, y) = (2x + y, x - y)$ 定义的线性变换。求 T 的核和像。T 是否是单射？T 是否是满射？求 T 的秩和零度。2. 考虑由 $D : \mathcal{P}_2 \to \mathcal{P}_1$ 定义的微分算子 $D(ax^2 + bx + c) = 2ax + b$。计算 $\ker D$ 和 $\operatorname{im} D$ 的维数。验证该变换的秩-零度定理。3. 对于固定向量 $\boldsymbol{v} \in \mathbb{R}^n$，定义 $T_{\boldsymbol{v}} : \mathbb{R}^n \to \mathbb{R}$ 由 $T_{\boldsymbol{v}}(\boldsymbol{x}) = \boldsymbol{v} \cdot \boldsymbol{x}$ 给出。证明这是一个线性变换，并求其核和像。4. 设 V 为 2×2 矩阵的向量空间。定义 $T : V \to V$ 由 $T(A) = A^T$ 给出。证明 T 是线性的。求 T 的核。证明 $T \circ T = I$ (，其中 I 是恒等变换)。T 是否是同构？请证明你的答案。5. 考虑质量算子 $M : C([0, L]) \to \mathbb{R}$ 由 $M(f) = \int_0^L f(x)\, dx$ 定义，其中 f 被解释为线性密度。证明 M 是线性变换。M 是否是满射？计算质心的运算是否为线性变换？6. 设 $T : \mathbb{R}^3 \to \mathbb{R}^2$ 为一个线性变换，且 $\ker T$ 由 $(1, 0, -1)^T$ 张成。$\operatorname{im} T$ 的维数是多少？T 可以是满射吗？求 $\operatorname{coim} T$ 的维数。7. 设 $V = \mathbb{R}^2$ 并定义向量 $\boldsymbol{u}, \boldsymbol{v} \in V$ 为等价的，如果它们仅相差 $\boldsymbol{a} = (1, 1)^T$ 的倍数。证明这是一种等价关系。几何上描述一个等价类是什么样的。证明这个等价关系对应于商空间 V/U，其中 $U = \operatorname{span}(\boldsymbol{a})$。8. 在 $\mathbb{R}^3$ 中，我们说两个向量是等价的，如果它们的第一个坐标相等。证明这定义了一种等价关系。给出商空间的具体描述（提示：它的维数是多少？）。描述生成这个等价关系的子空间。9. 对于线性变换 $S, T : V \to W$，证明 $\operatorname{rank}(S + T) \leq \operatorname{rank}(S) +$。

rank(T)。何时取等号？10. 设 $T : V \to W$ 为线性映射。证明：如果 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ 在 V 中线性无关，且对所有 i 都有 $T(\boldsymbol{v}_i) \neq 0$，则 $\{T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_k)\}$ 在 W 中线性无关。11. 设 $T : V \to W$ 为线性变换。证明：任何包含 $\ker T$ 的 V 的子空间都会被映射到 W 的一个子空间，其维数至多为 $\operatorname{rank} T$。12. 设 V 为有限维，且 $T : V \to V$ 为线性映射。证明：T 可逆当且仅当 $T(\ker T) = \{0\}$。13. 对于线性变换 $S, T : V \to W$，证明 $\ker(S + T)$ 包含 $\ker S \cap \ker T$。给出一个该包含为严格包含的例子。14. 设 $C([0, 1])$ 为区间 $[0, 1]$ 上连续函数的向量空间。我们称两个函数 $f, g \in C([0, 1])$ 等价，如果 $f(0) = g(0)$。证明这是一个等价关系。找出生成该等价关系的子空间。用一种简单的方式从几何上描述商空间。15. 设 $\boldsymbol{v} \in \mathbb{R}^n$ 非零。若 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ 满足 $\boldsymbol{v} \cdot \boldsymbol{x} = \boldsymbol{v} \cdot \boldsymbol{y}$，则称这两个向量等价。证明这是一个等价关系。证明由该等价关系定义的商空间同构于 $\mathbb{R}$。16. 设 V 为向量空间，$U < V$ 为其子空间。对 $\boldsymbol{v} \in V$，证明 V/U 中的等价类 $[\boldsymbol{v}]$ 等于集合 $\boldsymbol{v} + U = \{\boldsymbol{v} + \boldsymbol{u} : \boldsymbol{u} \in U\}$。利用这一点解释为什么 V/U 中的向量加法是良定义的。17. 设 $T : V \to W$ 为线性变换。通过显式构造一个同构并验证其良定义性，证明商空间 $V/\ker T$ 与 $\operatorname{im} T$ 同构。18. 对于线性变换 $T : V \to W$，证明 $\dim(\ker T + \operatorname{im} T) = \dim V$ 当且仅当 $T \circ T = T$。19. 若 $S, T : V \to W$ 为线性变换且 $\ker S = \ker T$，则存在一个同构 $\varphi : \operatorname{im} S \to \operatorname{im} T$。20. 设 V 为有限维向量空间，$U < V$ 为其子空间。证明由 $\pi(\boldsymbol{v}) = [\boldsymbol{v}]$ 定义的商映射 $\pi : V \to V/U$ 是线性变换。它的核是什么？21. 设 V 为 2×2 矩阵的向量空间。若两个矩阵 $A, B \in V$ 的迹相等，则定义它们等价。证明这是一个等价关系，并将该商空间与一个熟悉的向量空间对应起来。22. 考虑微分算子 $D : \mathcal{P}_2 \to \mathcal{P}_1$，其定义为 $D(ax^2 + bx + c) = 2ax + b$。求 $\ker D$ 和 $\operatorname{im} D$。验证该变换的秩–零度定理。求一个线性变换 $S : \mathcal{P}_1 \to \mathcal{P}_2$ 的显式公式，使得 $D \circ S = I_{\mathcal{P}_1}$。

Chapter 4
Bases & Coordinates

"fixing them firm on their base, the bellows began to blowe"

从抽象到具体的过渡与其逆过程同样重要。在抽象向量空间和变换的领域中停留之后，我们现在希望通过坐标和计算使这些概念变得精确且可度量。本章在我们已构建的理论结构与实际工作所需的工具之间架起桥梁。

这些概念源自几何直觉，十分熟悉：我们常常通过相对于所选坐标轴的坐标来描述空间中的点。这些坐标把一个抽象的点转化为一组可以操作的具体数字。然而，这个看似简单的想法——即我们可以系统地为抽象向量赋予数字——却蕴含着令人惊讶的深度。坐标系的选择表面上似乎是任意的，但它却能极大地影响我们解决问题的难易程度以及对结构的理解。

内在性质与其坐标表示之间的这种张力构成了线性代数的核心。向量空间独立于我们选择的任何具体度量方式而存在，然而不做出这些选择我们便无法进行计算。线性变换在几何上起作用，但只有在为其定义域和值域选择基之后，我们才将其编码为矩阵。这些选择——基以及它们所诱导的坐标——在抽象理解与具体计算之间搭建起桥梁。

我们的任务是谨慎地构建这座桥梁，确保它能够跨越鸿沟，承载理论洞见与实践效用。我们从“基”的概念开始——一组既能张成一个空间，又能高效完成这一任务的向量。这些基提供了坐标系，使我们能够将抽象向量转化为具体的数值列表。不同基之间的相互作用引导我们得到坐标变换公式，揭示了如何

几何对象在不同的视角下呈现。

将抽象结构系统地编码为数值形式，使线性代数的算法核心成为可能。为这种计算能力付出的代价，是坐标依赖型计算的激增。我们的挑战是在将坐标作为实用工具加以运用的同时，始终把握与坐标无关的真理。这种抽象理解与具体计算之间的相互作用，将贯穿并引导本书的全部展开。

4.1 *Bases & Spanning Sets*

回顾第2章，一个向量空间的张成集可能包含冗余向量，而线性无关集则可能无法覆盖空间中的所有向量。基的概念综合了这些思想，提供了一组能够高效张成的向量——既没有冗余，也不存在缺漏。

Example: 即使对于像 $\mathbb{R}^2$ 这样简单的对象，也存在无穷多种基的选择。标准基 $\{\hat{\imath}, \hat{\jmath}\}$ 只是众多选择中一个方便的选择。

定义 4.1（基）。向量空间 V 的一个 *basis* 是一组向量 $\mathcal{B}$，它张成 V 且线性无关。

•

这个定义的简洁性掩盖了其力量。基提供了一个最小生成集——所谓最小，是指从基中移除任意一个向量，所得集合将不再张成V。等价地，它提供了一个极大线性无关集——所谓极大，是指加入任何一个向量都会产生线性相关。

BONUS! 单项式基揭示了次数结构；拉格朗日基通过构造在某一个插值点取值为 1、在其他插值点取值为 0 的多项式来简化插值；牛顿基通过其嵌套结构促进递归计算。

例 4.2（多项式基）。二次多项式空间 $\mathcal{P}_2$ 有若干种自然的基的选择，每一种都提供不同的优势：

1. 单项式基 $\{1,\ x,\ x^2\}$
2. 对于插值点 $\{-1, 0, 1$ 的拉格朗日基函数 }Understood. Please provide the source text you would like

$$\left\{ \frac{x(x+1)}{2}, -x^2+1, \frac{x(x-1)}{2} \right\}$$

3. 以 d 和 e $(d \neq e)$ 为节点的牛顿基：

$$\{1, (x-d), (x-d)(x-e)\}$$

为说明起见，考虑多项式 $p(x) = x^2 + x + 1$。在单项式基下，它已经是标准形式：

$$p(x) = 1 \cdot 1 + 1 \cdot x + 1 \cdot x^2$$

在拉格朗日基于点 $\{-1, 0, 1\}$ 处，写出展开式：

$$p(x) = 2 \cdot \frac{x(x+1)}{2} + 1 \cdot (-x^2 + 1) + 2 \cdot \frac{x(x-1)}{2} = x^2 + x + 1$$

这演示了不同的基数如何以对不同计算目的有利的方式表示相同的多项式。 ◇

Foreshadowing: 不同的基揭示了空间结构的不同方面。在第7章中，我们将发现能够阐明线性微分方程作用的基。

任何向量空间都存在一组基，这一点由下面的结果所保证；其证明阐明了张成与线性无关之间的关系：

定理 4.3（基扩张）。 *Let V be a finite-dimensional vector space and $S \subseteq V$ be a linearly independent set. Then S can be extended to a basis of V by adding finitely many vectors. Moreover, if* $\dim V = n$ *and* $|S| = k \leq n$*, then exactly $n - k$ vectors need to be added.*

该定理可借助佐恩引理推广到无限维空间，不过扩展过程变得非构造性的。

Proof. 设 $\{v_1, \ldots, v_n\}$ 为 V 的任意一组基。我们可以通过从该基中逐个添加向量来扩充 S，直到得到一组基；当获得一个生成集时停止。由于 V 中的任何线性无关集的大小至多为 n，该过程在至多 $n - k$ 步后必然终止。

Caveat: 将一个集合扩展为一组基，或从一个生成集中抽取一组基的过程并非唯一——不同的选择会得到不同的基。

□

这一构造性证明揭示了一个基本原理：我们可以通过扩展（不断添加向量直到张成空间）或通过约化（不断移除向量直到线性无关）来构造基。由于有限维性，这两种过程都会终止——而这一关键假设在无限维空间中并不成立。

例 4.4（矩阵基）。2×2 矩阵的空间 $\mathbb{R}^{2\times 2}$ 具有标准基

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \; E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \; E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

这种基使坐标结构清晰，但掩盖了其他性质。例如，该基

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

更好地揭示了其分解为对称部分和反对称部分。

◇

由于引理 2.22，基的一个关键性质是它们的大小都相同：

推论 4.5. *Any two bases of a vector space have the same number of vectors.*

这揭示了维数是空间的一个内在性质，与基的选择无关。

例 4.6（维数计数）。以下维数自然出现：

1. 维数($\mathbb{R}^n$) = n 2. 维数($\mathcal{P}_n$
) = $n+1$ 3. 维数($\mathbb{R}^{m\times n}$)
= mn 4. 维数(sym_n)
= $\frac{1}{2}n(n+1)$ –
每个都统计指定该空间中一个元素所需的最少参数数量。

回顾一下，$\mathcal{P}_n$ 是次数为 $\leq n$ 的多项式空间，而 sym$_n$ 表示对称的 n-by-n 矩阵的空间。

◇

基底为我们提供了刻画向量空间的第一种系统性方法。基底的选择——总带有一定的任意性——以牺牲空间的内在性质来换取具体的可计算性。随着我们发展线性代数的工具体系，这种无坐标性质与依赖坐标的计算之间的张力将反复出现。

4.2 *Coordinates & Components*

基的存在不仅为向量空间提供了一个生成集——它还使得能够将抽象向量系统地转换为具体的数值列表。这个看似平凡的转换，是计算线性代数的基石。它在几何直觉与算法化操作之间架起了桥梁。

关键的观察是，空间中的任意向量都可以唯一地表示为基向量的线性组合。给定向量空间 V 的一组基 $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$，每个向量 $\boldsymbol{v} \in V$ 都有唯一的表示：

$$\boldsymbol{v} = c_1\boldsymbol{b}_1 + c_2\boldsymbol{b}_2 + \cdots + c_n\boldsymbol{b}_n$$

标量 $c_1, \cdots\cdots, c_n$ 称为 $\boldsymbol{v}$ 相对于该基的 *coordinates*。这些坐标的有序列表，写成一个列向量

Caveat: 记号 $[\boldsymbol{v}]_{\mathcal{B}}$ 强调坐标依赖于基的选择。向量在不同的基下具有不同的坐标，尽管向量本身保持不变。

$$[\boldsymbol{v}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

是 th相对于基 $\mathcal{B} = \{\boldsymbol{b}_1$ 的 $\boldsymbol{v}$ 的 e *coordinate vector*, $\cdots\cdots, \boldsymbol{b}_n\}$。

例 4.7（多项式坐标）。考虑 $\mathcal{P}_2$ 中的多项式 $p(x) = 6 + 2x - 3x^2$。在单项式基 $\mathcal{M} = \{1, x, x^2\}$ 下，其坐标向量是

$$[p]_{\mathcal{M}} = \begin{pmatrix} 6 \\ 2 \\ -3 \end{pmatrix}$$

在拉格朗日基 $\mathcal{L} = \{1, x-1, (x-1)(x+1)\}$ 中，同一个多项式具有不同的坐标：

$$[p]_{\mathcal{L}} = \begin{pmatrix} 6 \\ 2 \\ -3/2 \end{pmatrix}$$

多项式保持不变；只有其描述发生变化。 ◇

从向量到坐标的转换保持向量空间运算。如果 $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 的坐标向量分别为 $[\boldsymbol{v}]_{\mathcal{B}}$ 和 $[\boldsymbol{w}]_{\mathcal{B}}$，并且给定标量 a，则：

1. $[\boldsymbol{v} + \boldsymbol{w}]_{\mathcal{B}} = [\boldsymbol{v}]_{\mathcal{B}} + [\boldsymbol{w}]_{\mathcal{B}}$
2. $[a\boldsymbol{v}]_{\mathcal{B}} = a[\boldsymbol{v}]_{\mathcal{B}}$

这种结构的保持意味着坐标向量本身构成一个与原空间同构的向量空间。

Foreshadowing: 在转到坐标表示时向量空间运算得以保持，这解释了为什么矩阵乘法表示了线性变换的复合。

例 4.8（矩阵坐标）。考虑空间 $\mathbb{R}^{2\times 2}$，其标准基由基本矩阵 $\mathcal{E} = \{E_{11}$、E_{12}、E_{21}、$E_{22}\}$ 构成。该矩阵

$$A = \begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix}$$

具有坐标向量

$$[A]_{\mathcal{E}} = \begin{pmatrix} 2 \\ -1 \\ 3 \\ 4 \end{pmatrix}$$

相对于该基。用另一组基表示的同一矩阵

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

需要使用不同的坐标来表达同一变换。 ◇

坐标表示的唯一性——由基向量的线性无关性所保证——使我们能够通过坐标来检验相等性。两个向量相等当且仅当它们的坐标向量

这种将几何或代数性质简化为数值测试的过程展示了坐标如何使计算成为可能。关键在于选择能够使所需计算变得简单的坐标。

相对于任何基底是相等的。这将抽象的向量相等性简化为数值比较。

因此，我们建立了抽象向量与具体数字列表之间的字典。这个字典严重依赖于我们选择的基——一个我们可以根据计算需求自由选择和改变的选择。不同基的选择之间的相互作用，以及向量从不同坐标视角下的表现，自然引出了我们的下一个话题：基变换。

4.3 *Change of Basis*

正确的坐标系统可以将复杂的问题转化为简单的问题。一个振荡的弹簧-质量系统，在笛卡尔坐标系中由耦合方程描述，当以其自然模态观察时，会简化为独立的运动。一个机器人手臂的运动，在工作空间坐标中复杂难以指定，可能在关节角度中沿着简单的路径运动。一个粒子的加速度，在直角坐标系中复杂，但在极坐标系中可能会显著简化。这些视角的变换——这些基变换——不仅仅是数学上的便利，而是理解和控制物理系统的基本工具。

考虑一个向量空间 V，它有两个不同的基 $\mathcal{B}$ 和 $\mathcal{B}'$。一个向量 $\boldsymbol{v} \in V$ 独立于我们如何描述它的方式存在，就像爬山无论我们用米还是英尺来衡量其高度，难度都是一样的。然而，要使用一个向量——进行计算、变换或理解它与其他向量的关系——我们必须选择坐标系。相同的向量在不同的基中有不同的坐标表示：

$$\boldsymbol{v} = \sum_{i=1}^{n} c_i \boldsymbol{b}_i = \sum_{i=1}^{n} c'_i \boldsymbol{b}'_i$$

其中 $[\boldsymbol{v}]_{\mathcal{B}} = (c_1, \ldots, c_n)^T$ 和 $[\boldsymbol{v}]_{\mathcal{B}'} = (c'_1, \ldots, c'_n)^T$ 分别是其在基底 $\mathcal{B}$ 和 $\mathcal{B}'$ 中的坐标向量。

例 4.9 （电场分量）。来自点电荷的电场 $\vec{E}$ 可以在不同的坐标系中测量。在原点附近的电荷 q，我们可以在笛卡尔坐标系中表示 $\vec{E}$：

$$\vec{E} = E_x \hat{\imath} + E_y \hat{\jmath} + E_z \hat{k}$$

或在球坐标系中：

$$\vec{E} = E_\rho \hat{\boldsymbol{e}}_\rho + E_\theta \hat{\boldsymbol{e}}_\theta + E_\phi \hat{\boldsymbol{e}}_\phi$$

这些描述在任意点（x，y，z）与球面坐标（ρ，θ，ϕ）之间的变换由以下公式给出：

$$\begin{pmatrix} E_\rho \\ E_\theta \\ E_\phi \end{pmatrix} = \begin{bmatrix} \sin\phi\cos\theta & \sin\phi\sin\theta & \cos\phi \\ -\sin\theta & \cos\theta & 0 \\ \cos\phi\cos\theta & \cos\phi\sin\theta & -\sin\phi \end{bmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$$

这种正交变换表示一种基变换，对于分析径向对称场尤其有用，其中球面分量通常揭示了在笛卡尔坐标系。变换矩阵遵循数学家的惯例，其中 θ 表示从 x 轴到 xy 平面上的方位角（0 到 2π），ϕ 表示从 z 轴到极角（0 到 π）。 ◇

理解基变换的关键在于通过使用矩阵将新基向量表示为旧基向量的线性组合。

定义 4.10（基变换矩阵）。设 $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ 和 $\mathcal{B}' = \{\boldsymbol{b}'_1, \ldots, \boldsymbol{b}'_n\}$ 是向量空间 V 的基。由 $\mathcal{B}$ 到 $\mathcal{B}'$ 的 *change of basis matrix* 是矩阵 $P = [p_{ij}]$，其条目由唯一表示决定：

$$P = [p_{ij}] \quad : \quad \boldsymbol{b}'_j = \sum_{i=1}^{n} p_{ij}\boldsymbol{b}_i \tag{4.1}$$

j列的P包含了$\boldsymbol{b}'_j$的$\mathcal{B}$坐标，编码了如何用旧基向量表示每一个新的基向量。 •

示例 4.11（信号处理）。在音频处理过程中，声音信号自然地以时域开始——在离散时间点上测量的振幅。为了分析和滤波，我们通常使用离散傅里叶变换（DFT）将其转换到频域。这实际上是基底的变化，其中我们的新基底向量是复指数函数：

$$\boldsymbol{b}'_k = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ e^{-2\pi ik/n} \\ e^{-4\pi ik/n} \\ \vdots \\ e^{-2\pi ik(n-1)/n} \end{pmatrix} \qquad (\quad \diamond$$

Foreshadowing: 在这种情况下，基变换矩阵 P 是 *unitary* 正交矩阵的复数类比），反映了时间域和频率域之间能量的守恒。

给定基变换矩阵 P，我们可以系统地转换坐标系：

$$[\boldsymbol{v}]_{\mathcal{B}} = P[\boldsymbol{v}]_{\mathcal{B}'}$$

这个公式揭示了P从$\mathcal{B}'$坐标系转换到$\mathcal{B}$坐标系——它“撤销”了基变换。相反，为了根据$\mathcal{B}$坐标找到$\mathcal{B}'$坐标，我们求解：

$$[\boldsymbol{v}]_{\mathcal{B}'} = P^{-1}[\boldsymbol{v}]_{\mathcal{B}}$$

例4.12（主应力）。在分析薄平面材料的力学时，*stress tensor* 记录某一点的应力和应变，相对于所选坐标轴是一个对称的 2×2 矩阵 $\sigma \in \mathrm{sym}_2$：

$$\sigma = \begin{bmatrix} \sigma_{xx} & \tau_{xy} \\ \tau_{xy} & \sigma_{yy} \end{bmatrix}$$

总是存在一种方便的基——主应力方向——在该基下应力张量为对角形式：

$$\sigma' = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

通过特征分解（第7章）获得这一基底对于预测材料失效至关重要。这里的换基矩阵 P 由沿主应力方向的单位向量组成。$\diamond$

换基矩阵 P 满足若干性质，反映了坐标变换的本质：

1. P 是可逆的（否则在变换中一些向量会丢失信息） 2. 从 $\mathcal{B}$ 到 $\mathcal{B}'$ 的变化由 P^{-1} (视角反转) 给出 3. 对于第三个基 $\mathcal{B}''$，这些矩阵可以自然地复合（反映传递性）

这些性质确保我们的坐标变换是可逆且一致的——改变视角既不会创造也不会销毁关于所描述向量的信息。这种结构的保持正是线性代数在工程中强大力量的核心：它使我们能够在最符合当前需求的任何坐标系中工作，并且确信可以将结果转换回任何其他视角。

当我们考虑线性变换时，基变换的真正意义才显现出来，因为线性变换的矩阵表示在很大程度上取决于我们对坐标的选择。这种——基、变换及其矩阵表示之间的——关系，引出了相似这一基本概念，见下文的 Definition4.16 。

4.4 Matrix Representations

几何变换独立于我们如何度量它而存在：顺时针旋转90度，无论我们用笛卡尔坐标还是极坐标来描述，都是同一个旋转。然而，为了对变换进行计算——将它们组合、将它们作用于向量、分析其效果——我们必须通过矩阵在坐标中表达它们。抽象变换与其各种矩阵之间的关系

表征揭示了基于坐标的计算的能力与局限性。

设 $T : V \to W$ 为向量空间之间的一个线性变换，其选定的基为 V 的 $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$，以及 W 的 $\mathcal{B}' = \{\boldsymbol{b}'_1, \ldots, \boldsymbol{b}'_m\}$。要构造 T 的矩阵表示，只需记录它对基向量的作用：

$$T(\boldsymbol{b}_j) = \sum_{i=1}^{m} a_{ij} \boldsymbol{b}'_i$$

系数 a_{ij} 构成一个称为 T 相对于基 $\mathcal{B}$ 和 $\mathcal{B}'$ 的 *matrix repre-sentation* 的 $m \times n$ 矩阵 $[T]^{\mathcal{B}}_{\mathcal{B}'}$。元素 a_{ij} 给出 $T(\boldsymbol{b}_j)$ 在基 $\mathcal{B}'$ 中的第 i 个坐标。

Think: 在 $[T]^{\mathcal{B}}_{\mathcal{B}'}$ 中，底部的基 $\mathcal{B}'$ 是我们度量输出（陪域）的地方，而顶部的基 $\mathcal{B}$ 是我们度量输入（定义域）的地方。该矩阵将 $\mathcal{B}$-坐标转换为 $\mathcal{B}'$-坐标，其阅读方向从右到左，就像函数复合一样。

例 4.13（不同基下的旋转）。考虑在 $\mathbb{R}^2$ 中按逆时针方向旋转 $\pi/2$。在标准基 $\mathcal{B} = \{\hat{\imath}, \hat{\jmath}\}$ 下，这个变换具有熟悉的矩阵表示：

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

设$\mathcal{B}'$为由向量$\boldsymbol{v}_1 = (1, 1)^T$和$\boldsymbol{v}_2 = (-1, 2)^T$组成的基。从$\mathcal{B}'$到$\mathcal{B}$的基变换矩阵为：

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$$

在这个新的基下，同样的旋转变换表现为：

$$[T]^{\mathcal{B}'}_{\mathcal{B}'} = P^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} P = \frac{1}{3} \begin{bmatrix} -1 & -5 \\ 2 & 1 \end{bmatrix}$$

尽管这些矩阵看起来相当不同，它们表示的是完全相同的几何变换：将向量逆时针旋转 $\pi/2$。 ◇

矩阵 $[T]^{\mathcal{B}}_{\mathcal{B}'}$ 通过标准的矩阵乘法将输入坐标转换为输出坐标：

$$[T(\boldsymbol{v})]_{\mathcal{B}'} = [T]^{\mathcal{B}}_{\mathcal{B}'} [\boldsymbol{v}]_{\mathcal{B}}$$

该公式概括了线性变换如何与坐标相互作用：先将输入表示为 $\mathcal{B}$ 坐标，然后通过矩阵表示进行相乘，以获得输出的 $\mathcal{B}'$ 坐标。

例4.14（投影到一条直线）。考虑在 $\mathbb{R}^2$ 中到 x-轴的正交投影。在标准坐标下，它的矩阵表示为：

$$[P]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

如果我们将坐标系旋转角度 θ，得到一个新的基 $\mathcal{B}' = \{(\cos\theta, \sin\theta)^T, (-\sin\theta, \cos\theta)^T\}$，同样的投影看起来更为复杂：

$$[P]_{\mathcal{B}'}^{\mathcal{B}'} = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}$$

几何作用保持不变——我们只是通过不同的坐标视角来观察它。 ◇

Example: 当 $\theta = \pi/4$ 时，$\mathcal{B}'$ 的基向量是 $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T$ 和 $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T$。在这些坐标下，投影矩阵变为

$$[P]_{\mathcal{B}'}^{\mathcal{B}'} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

当基发生变化时，矩阵表示会系统地变换。如果 P 是定义域中从 $\mathcal{B}'$ 到 $\mathcal{B}$ 的换基矩阵，而 Q 是余域中从 $\mathcal{D}'$ 到 $\mathcal{D}$ 的换基矩阵，那么：

$$[T]_{\mathcal{D}}^{\mathcal{B}} = Q[T]_{\mathcal{D}'}^{\mathcal{B}'}P^{-1}$$

这个关系揭示了相同变换的不同矩阵表示如何通过相似变换相互关联——这是我们下一节的主题。

例4.15（缩放与反射）。考虑线性变换 $T : \mathbb{R}^2 \to \mathbb{R}^2$，它在 x 方向上按 2 倍缩放，并关于 x 轴反射。在标准坐标下：

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

在旋转了 $\pi/4$ 的坐标系中，这种变换似乎将缩放与反射混合在一起：

$$[T]_{\mathcal{B}'}^{\mathcal{B}'} = \begin{bmatrix} 1/2 & 3/2 \\ 3/2 & 1/2 \end{bmatrix}$$

表观的复杂性并非源自变换本身，而是源于我们所选择的测量系统。

◇

矩阵表示的实践价值在于其可计算性——它们将抽象的变换化约为具体的数值数组。然而，其对坐标的依赖性提醒我们，任何单一的矩阵都无法讲述全部。一个变换在某一基下可能显得简单，而在另一基下则可能变得复杂，正如同一条二次曲线在不同坐标系中呈现出不同的形态。

我们在后续章节中的任务是找到能够揭示线性变换基本特征的基。某些基将对特定变换进行对角化（第7章）；其他基将尊重几何结构（第5章）或优化逼近（第10章）。每种基提供了一个不同的视角，用以观察和理解线性变换。

4.5 *Coordinate-Free Thought*

我们对坐标和矩阵表示的开发呈现了一个根本性的悖论。我们首先将线性变换作为向量空间之间的抽象映射进行研究，独立于任何特定的测量系统来理解其性质。然而，要对这些变换进行计算——将它们应用于向量，进行组合，分析它们的效果——我们必须选择坐标并使用矩阵。变换的内在性质与其依赖坐标的表示之间的这种紧张关系是线性代数的核心所在。

Foreshadowing: 同构向量空间与相似变换之间的关系暗示了数学中更深层的范畴结构。

考虑从某些高维测量过程中提取的数据——可能是成千上万个细胞中的基因表达水平，或神经网络各层之间的激活模式。基础的生物学或计算结构独立于我们选择如何测量它。不同的实验协议或网络架构可能会产生相同基本模式的不同表示。这引出了一个更深层次的问题：何时两个表面上看似不同的表示实际上是等价的?

有些人认为等价的表达式

$$AX = XB$$

更令人难忘且引人共鸣。

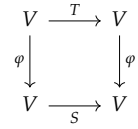
定义 4.16（相似）。若存在一个可逆矩阵 X 使得：则两个矩阵 A 和 B 是 *similar*。

$$B = X^{-1}AX$$

我们写 $A \sim B$ 来表示相似矩阵。 •

这种代数关系精确刻画了何时两个矩阵通过不同坐标系的视角表示同一个线性变换。矩阵 P 编码了将一种视图变换为另一种视图的基变换。更为根本地，我们称两个线性变换 $S, T : V \to V$ 是 *similar*，如果存在一个同构 $\varphi : V \to V$ 使得 $S = \varphi^{-1}T\varphi$。当用坐标表示时，这一抽象概念体现为矩阵相似。

术语*conjugate*在数学中更为常见，但*similar*也完全可以。



当 $S = \varphi^{-1}T\varphi$ 时，该图表可交换，说明相似性可视为由同构 φ 所给出的共轭。

例4.17（几何相似）。考虑平面内一次逆时针旋转90°。在标准坐标中，这表现为:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

如果我们改为使用基 $\{v_1、v_2\}$ 来度量向量，其中:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad \Rightarrow \quad X = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$$

则相同的旋转具有矩阵：

$$B = X^{-1}AX = \begin{bmatrix} 1/3 & -1 \\ 1/3 & 0 \end{bmatrix}$$

尽管这些矩阵看起来非常不同，但它们表示相同的几何操作。

◇

例子 4.18（形状分析）。考虑分析表示为$\mathbb{R}^3$中点云的复杂三维物体的形状。内在形状独立于坐标选择存在——旋转或平移物体不会改变其基本几何形状。相同形状的两种不同坐标表示通过相似变换相关联。数据分析的几何方法基于这一见解，研究在这些变换下保持不变的形状属性。◇

相似矩阵共享一些内在于它们所表示的变换的性质：

1. 它们具有相同的行列式 2. 它们具有相同的秩 3. 它们具有相同的迹（对角线元素之和）

Foreshadowing: 在第七章中，我们将发现相似矩阵也共享特征值——这是它们所表示的变换的另一个内在属性。

这些 *coordinate invariants* 属于变换本身，而非任何特定的矩阵表示。它们形成了一种指纹，能够区分真正不同的变换与仅仅是同一映射的不同坐标视图。

内在属性和坐标依赖属性之间的区别是现代数据分析的指导原则。在研究高维数据时，我们会问：

1. 在合理的变换下，哪些特征是不变的？ 2. 哪些坐标系最能揭示这些特征？ 3. 不同的表示方式如何揭示不同的方面？

Foreshadowing: 在第11章中，我们将看到主成分分析如何发现揭示高维数据中内在低维结构的坐标系统。

这个视角的力量将在接下来的章节中显现出来。我们将发现：

- 尊重几何结构的坐标（第5章）
- 揭示动力学行为的表征（第7章）
- 最适合降维的基（第10章）

每种方法提供了一种不同的视角来观察我们的研究对象，突出了不同的特征，同时保持其本质特征。

例子 4.19（数据可视化）。高维数据集可以通过不同的投影方式，在二维或三维空间中以无数种方式进行可视化。

维度。虽然某些投影可能会掩盖数据的结构，但其他投影则揭示了有意义的模式。艺术在于找到能够展示感兴趣特征的坐标，同时尊重数据中的内在关系。尽管可视化坐标的选择是任意的，但它们揭示的模式——簇、离群值、非线性关系——反映了数据中的真实结构。◇

然而，我们必须时刻牢记变换及其各种表示之间的区别。矩阵是计算工具——强大且必要的工具，但并不是全部。真正的研究对象是变换本身，它们独立于我们选择如何度量它们。坐标无关的概念与基于坐标的计算之间的这种张力贯穿于线性代数的各个方面，从最理论的思想到最实际的应用。

我们的线性代数之旅将不断地在这些视角之间穿梭——在抽象与具体之间，在几何与代数之间，在内在与坐标依赖之间。艺术在于知道何时进行坐标无关的推理，何时利用精心选择的坐标的计算力量。在接下来的章节中，我们将看到这种相互作用揭示从微分方程的结构到神经网络的架构的方方面面。

Robotic Arm Kinematics

机器人操作的数学原理生动展示了不同坐标系统如何揭示同一物理系统的不同方面。机器人臂的运动可以通过多个基底来描述，每个基底揭示了其行为的不同方面。这些坐标选择——以及它们之间的转换——体现了本章中发展出的基本原理。

考虑一个平面机器人臂，具有两个旋转关节，连接两根长度分别为 L_1 和 L_2 的刚性连杆。该机械臂的配置允许两种自然坐标系：

1. 关节空间坐标（θ_1，θ_2），用于测量各个关节的角度 2. 任务空间坐标（x，y），给出末端执行器的位置

这些空间配备了自然基：关节空间的基向量对应于每个关节的微小旋转，而任务空间继承了平面上的标准笛卡尔基。两者之间的变换由以下公式给出：

$$F(\theta_1, \theta_2) = \begin{pmatrix} L_1 \cos\theta_1 + L_2 \cos(\theta_1 + \theta_2) \\ L_1 \sin\theta_1 + L_2 \sin(\theta_1 + \theta_2) \end{pmatrix}$$

尽管该变换 F 是非线性的，但其在任意构型处的导数 $[DF]$ 提供了切空间之间的线性映射——一种将无穷小运动联系起来的基变换矩阵：

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = [DF]_{(\theta_1,\theta_2)} \begin{pmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix}$$

其中：

$$[DF]_{(\theta_1,\theta_2)} = \begin{bmatrix} -L_1 \sin\theta_1 - L_2 \sin(\theta_1+\theta_2) & -L_2 \sin(\theta_1+\theta_2) \\ L_1 \cos\theta_1 + L_2 \cos(\theta_1+\theta_2) & L_2 \cos(\theta_1+\theta_2) \end{bmatrix}$$

Recall: 微积分中的反函数定理保证，当 $[DF]$ 在某一构型处是可逆的（即当 $\det[DF] \neq 0$ 时），F 存在一个局部逆——我们可以唯一地求解为实现期望的末端执行器运动所需的关节角的小变化。当 $[DF]$ 不可逆时，例如机械臂完全伸直时，某些瞬时运动将变得不可能。

该矩阵的各列表示由单位关节速度所产生的任务空间速度——它们构成了一个与构型相关的、用于描述可实现末端执行器运动的基。

更复杂的机械臂揭示了坐标关系的更多方面。设想在保持末端执行器平面运动的情况下，将我们的机械臂扩展为三个关节。此时，关节空间具有基向量 $\{\partial/\partial\theta_1、\partial/\partial\theta_2、\partial/\partial\theta_3\}$，而任务空间仍然是二维的，其基为 $\{\partial/\partial x、\partial/\partial y\}$。导数成为这些空间之间的一个线性变换 $[DF] : \mathbb{R}^3 \to \mathbb{R}^2$，其中：

$$[DF]_{(\theta_1,\theta_2,\theta_3)} = \begin{bmatrix} \dfrac{\partial x}{\partial\theta_1} & \dfrac{\partial x}{\partial\theta_2} & \dfrac{\partial x}{\partial\theta_3} \\ \dfrac{\partial y}{\partial\theta_1} & \dfrac{\partial y}{\partial\theta_2} & \dfrac{\partial y}{\partial\theta_3} \end{bmatrix}$$

该矩阵具有非平凡的核——反映了使末端执行器在瞬时保持不动的关节速度。这种自运动体现了第3章中研究的基本核—像关系，如今在一个具体的机械系统中自然地显现出来。

这些坐标关系的实际意义体现在轨迹规划中。末端执行器的直线运动在任务坐标中虽显优雅，但可能需要复杂精细的关节空间协同。相反，简单的关节轨迹却能在任务空间中描绘出复杂路径。对于三关节机械臂，冗余性进一步丰富了这种关系——同一末端执行器轨迹允许无限多种关节空间实现，对应于穿过 $[DF]$ 核空间的不同路径。

这种表示之间的二重性反映了一个更深层的事实：不存在任何单一的坐标系能够以同等清晰度捕捉复杂系统的所有方面。工程的艺术不仅在于选择合适的坐标，更在于能够根据问题需要在不同表示之间自如切换。关节坐标使动力学和关节约束的问题一目了然，而任务坐标则简化了运动的描述。本章所建立的数学框架将这种艺术从直觉性的技艺转化为系统化的科学，提供了有效处理同一底层现实的多种坐标表示所必需的工具。

Computer Graphics & Coordinate Systems

现代计算机图形学揭示了坐标变换的实际威力。一个虚拟对象——例如飞行模拟器中的航天器——会同时存在于多个坐标系中，而每个坐标系的选择都是为了简化仿真的某些特定方面。理解这些基底如何通过第 4.3 节中的变换相互关联，能够将复杂的几何问题转化为系统化的矩阵计算。

设想我们的虚拟航天器。其几何最初定义在 *body coordinates* 中，在那里，自然基 $\mathcal{B}_b = \{\boldsymbol{b}_1、\boldsymbol{b}_2、\boldsymbol{b}_3\}$ 与航天器的结构对齐：$\boldsymbol{b}_1$ 指向机鼻，$\boldsymbol{b}_2$ 沿右机翼，而 $\boldsymbol{b}_3$ 向下。在这些坐标中，航天器的对称性变得清晰，控制面与坐标平面对齐。航天器上的一点 $\boldsymbol{p}$ 相对于该机体系基具有坐标向量 $[\boldsymbol{p}]_{\mathcal{B}_b}$。

然而，我们的航天器是在一个具有自身坐标系的虚拟世界中运动的。*world basis* $\mathcal{B}_w = \{\boldsymbol{w}_1、\boldsymbol{w}_2、\boldsymbol{w}_3\}$ 通常将 $\boldsymbol{w}_3$ 对齐为竖直方向，而 $\boldsymbol{w}_1$ 和 $\boldsymbol{w}_2$ 则张成地面平面。按照第 4.2 节，从机体坐标基到世界坐标基的坐标变换可由基变换矩阵得到：

$$[\boldsymbol{p}]_{\mathcal{B}_w} = P[\boldsymbol{p}]_{\mathcal{B}_b}$$

该变换矩阵 P 的列表示在世界坐标中表达的机体系基向量，正如示例 7.3 中所构建的那样：

$$P = \begin{bmatrix} | & | & | \\ [\boldsymbol{b}_1]_{\mathcal{B}_w} & [\boldsymbol{b}_2]_{\mathcal{B}_w} & [\boldsymbol{b}_3]_{\mathcal{B}_w} \\ | & | & | \end{bmatrix}$$

每一列展示了一个物体坐标系基向量在世界坐标系中的分解。与第3.1节研究的旋转矩阵一样，这种基变换保持长度和角度——这是刚体至关重要的性质。

虚拟相机引入了另一组基。*camera basis* $\mathcal{B}_c = \{\boldsymbol{c}_1、\boldsymbol{c}_2、\boldsymbol{c}_3\}$ 将虚拟镜头置于原点，其中 $\boldsymbol{c}_3$ 指向视线方向，$\boldsymbol{c}_2$ 与图像的垂直轴对齐。点通过与另一个换基矩阵 Q 的复合变换到这些坐标：

$$[\boldsymbol{p}]_{\mathcal{B}_c} = Q[\boldsymbol{p}]_{\mathcal{B}_w}$$

这些变换的复合——从机体坐标系到世界坐标系再到相机坐标系——体现了第4.3节的核心代数洞见：基变换通过矩阵乘法进行复合。一个点的坐标变换为：

$$[\boldsymbol{p}]_{\mathcal{B}_c} = QP[\boldsymbol{p}]_{\mathcal{B}_b}$$

这个矩阵乘积体现了坐标变换的全部变化，不过为了清晰性和效率，我们通常分别维护各个变换。

该序列中的每一种基都服务于特定的目的：用于物理仿真的物体坐标，用于场景构成的世界坐标，以及用于可见性与渲染的相机坐标。它们之间的变换，尽管表面上看起来复杂，却直接源自我们对坐标的精确认识以及

Example: 当航天器向上俯仰 30° 时，用世界坐标表示的其机体基向量成为换基矩阵 P 的列：

$$\begin{bmatrix} \sqrt{3}/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & \sqrt{3}/2 \end{bmatrix}$$

这些列彼此正交，因为 P 表示刚性旋转。

第4.2节中构建的基。这体现了一个更广泛的原则：当在合适的坐标下审视时，具有挑战性的问题往往会变得可处理。

实际意义远超图形学。在机器人学中，类似的坐标变化通过第4.4节中研究的变换矩阵，将关节角度与末端执行器位置关联。在计算机视觉中，相机坐标系和世界坐标系必须对齐，以实现增强现实。在航天器导航中，机体坐标系与惯性坐标系在导航算法中相互作用。每个应用都基于相同的数学基础：基底的精心构建以及它们之间的变换。

这个坐标系的级联展示了第4.5节中的一个最终关键洞见：基底应选择与我们问题的自然结构相匹配。身体坐标尊重车辆对称性；世界坐标与重力和地形对齐；相机坐标匹配视角几何。没有单一的基底能够优雅地捕捉所有方面——艺术在于为每个子任务选择合适的基底，同时理解它们如何通过本章中发展出的精心矩阵代数相互关联。

□ ——————————————————————————— □

Exercises: Chapter 4

1. 对于 $\mathbb{R}^2$ 的基 $\mathcal{B} = \{v_1, v_2\}$，其中 $v_1 = (1, 2)^T$ 且 $v_2 = (1, -1)^T$，求从 $\mathcal{B}$ 到标准基的换基矩阵 P。2. 考虑 $\mathbb{R}^3$ 中的 $v = (2, 1, -1)^T$ 以及基 $\mathcal{B} = \{(1, 1, 0)^T, (0, 1, 1)^T, (1, 0, 1)^T\}$。求 $[v]_{\mathcal{B}}$。3. 为对称 2×2 矩阵空间 sym_2 找一个基，然后求 $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ 在该基下的坐标向量。4. 求 $p(x) = x^2 - 2x + 1$ 相对于 $\mathcal{P}_2$ 的基 $\{1 + x^2, x - x^2, 1 - x\}$ 的坐标向量。5. 对于 $\mathcal{P}_2$，通过证明其线性无关并张成 $\mathcal{P}_2$，说明 $\{1, 1 + x, 1 + x + x^2\}$ 是一组基。然后用这组基表示 $\{1, x, x^2\}$。6. 给定 $\mathbb{R}^2$ 的一组基 $\mathcal{B} = \{v_1, v_2\}$，证明从 $\mathcal{B}$ 到标准基的换基矩阵的列等于 v_1 和 v_2。7. 设 V 为 2×2 矩阵空间，其基为 $\mathcal{B} = \{E, S, R, D\}$，其中 $E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$，$S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$，$R = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$，$D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$。描述一种方法，用于在该基下求任意 2×2 矩阵的坐标向量。

8. 让 $T : \mathbb{R}^2 \to \mathbb{R}^2$ 逆时针旋转向量 $\pi/4$。求其矩阵在：（a）标准基底，（b）基底 $\{(1, 1)^T, (-1, 1)^T\}$ 中。

9. 设 $T : \mathcal{P}_2 \to \mathcal{P}_2$ 为微分 $T(p) = p'$。求其在标准基 $\{1, x, x^2\}$ 下的矩阵，然后找一个基，使其矩阵表示更简单。

10. 考虑矩阵 $A = \begin{bmatrix} 1 & 2 & 3 \\ & & 4 \end{bmatrix}$ 和 $B = \begin{bmatrix} 5 & -2 & - \\ 3 & 0 & \end{bmatrix}$. 要么找到一个矩阵$P$使得$B = P^{-1}AP$，要么证明不存在这样的$P$。你可以首先检查哪些不变量?

11. 对于 $T:\mathbb{R}^2 \to \mathbb{R}^2$，矩阵 $[T]_{\mathcal{B}_1} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ 在基底 $\mathcal{B}_1$ 中，以及 $[T]_{\mathcal{B}_2} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ 在基底 $\mathcal{B}_2$ 中，求从 $\mathcal{B}_2$ 到 $\mathcal{B}_1$ 的基变换矩阵。 12. 矩阵 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ 和 $B = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$ 是相似的。如何直接证明这一点？尝试找到 P 使得 $B = P^{-1}AP$。这能通过解线性系统来完成吗？ 13. 考虑操作符 $T:\mathcal{P}_2 \to \mathcal{P}_2$ 定义为 $T(p)(x) = p(x+1)$。求其在标准基底 $\{1, x, x^2\}$ 下的矩阵。然后求其在基底 $\{1, x+1, (x+1)^2\}$ 下的矩阵。你观察到了什么模式？解释这个模式为何在几何上会发生。 14. 考虑 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ 作为 $\mathbb{R}^3$ 的基底，其中 $\boldsymbol{v}_1 = (1,1,1)^T, \boldsymbol{v}_2 = (1,1,-2)^T, \boldsymbol{v}_3 = (1,-2,1)^T$。对于 $T:\mathbb{R}^3 \to \mathbb{R}^3$，由叉积 $T(\boldsymbol{x}) = \boldsymbol{x} \times (1,0,0)^T$ 给出，求 T 在该基底下的矩阵。首先求 T 在每个基向量上的作用，然后将这些结果结合起来得到矩阵表示。 15. 对于线性变换 $T:V \to V$：（a）如果 $[T]_{\mathcal{B}}^{\mathcal{B}}$ 对于某个基底 $\mathcal{B}$ 是对角矩阵，几何上这意味着 T 如何作用于向量，（b）证明找到这样一个基底等同于找到那些 T 映射为其自身的标量倍数的向量。 16. 对于相似矩阵 A 和 B：（a）证明 $\det(A) = \det(B)$ 且 $\operatorname{tr}(A) = \operatorname{tr}(B)$；（b）还要证明当 $n >= 1$ 时，A^n 和 B^n 是相似的。*Hint:* 你知道行列式在复合下是如何变化的；对于迹，回顾（或证明）$\operatorname{tr}(XY) = \operatorname{tr}(YX)$。 17. 对于 2×2 矩阵 A，证明如果 A 与对角矩阵相似，则它与任何具有相同对角元素（无论顺序如何）的矩阵相似。给出一个例子，显示两个具有相同对角元素的矩阵不一定是相似的。 18. 令 $\alpha(t) = 1 + t + t^2$ 和 $\beta(t) = 1 - t^2$ 是 $\mathcal{P}_2$ 中的多项式。找到基底 $\mathcal{B}_1$ 和 $\mathcal{B}_2$，使得 $[\alpha]_{\mathcal{B}_1} = (1,0,0)^T$ 和 $[\beta]_{\mathcal{B}_2} = (1,0,0)^T$。这告诉你这些多项式的什么信息？求从 $\mathcal{B}_1$ 到 $\mathcal{B}_2$ 的基变换矩阵。 19. 对于一个基底为 $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ 的向量空间 V，证明任何线性变换 $T:V \to V$ 都由其在基向量上的作用唯一确定。然后证明这意味着相似矩阵必须具有相同的迹，而不使用矩阵乘法。这如何与线性变换的几何性质相关联？

Chapter 5
Inner Products & Orthogonality

"others triangular right angled course maintain; others obtuse, acute Scalene, in simple paths"

我们所居住的空间具有超越简单加法和缩放的结构。就像测量天体弧度的指南针，向量描绘着穿越空间的路径——每个路径都有其自身的大小，每个路径与其他路径形成精确的角度。这些几何概念——长度与垂直性、度量与关系——并非来自随意的约定，而是通过内积仔细定义的向量相互作用的结果。

我们的任务是提取几何测量的精髓——长度、角度和正交性——并将其扩展到超越欧几里得起源的领域。熟悉的向量点积在微积分中表现良好，但它仅仅代表了更深层次结构的一个实例。内积提供了将几何秩序施加于抽象向量空间的机制，使我们能够以尊重向量内在性质的方式度量和比较向量。

这种几何视角改变了我们对线性代数的理解。正交向量，过去通过坐标计算来理解，如今显现为一种基本的组织原则。正交基为计算提供了最优的框架。正交矩阵保持了我们所构建的几何结构。通过内积，前几章中抽象的向量空间获得了形状和实质。

内积的选择塑造了我们对向量空间的看法，突出某些特征的同时掩盖其他特征。不同的内积引发了不同的长度和角度概念，每种概念都适用于特定的应用。有些自然地来源于物理原理，有些则来自统计考虑，还有一些则是出于计算的便利性。

艺术之处在于选择一种内积，在保留基本结构的同时，揭示最值得关注的方面。

我们的发展从具体走向抽象，再回到具体。熟悉的点积在我们沿着公理化的阶梯上升时引导着直觉。尽管我们会通过精心挑选的例子偶尔瞥见无限维空间，我们的关注仍然放在有限维空间上，在那里理论呈现出最纯粹的形态。每一个定义都以数学的必然性建立在前一个之上，然而我们所构建的结构反映了某种根本性的东西：几何直觉与代数法则之间的深刻统一。

5.1 Dot & Inner Products

点积在微积分中首次出现，此后不断扩展。两个向量 $\boldsymbol{u}$、$\boldsymbol{v} \in \mathbb{R}^n$ 通过按坐标逐项相乘并相加来结合：

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i$$

这一运算，尽管通过坐标来定义，却揭示了基本的几何特征：通过 $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ 表示长度，通过 $\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta$ 表示角度，以及当 $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ 时的正交性。如此简单的公式却编码了如此丰富的几何内容，暗示着其中存在更深层次的结构。

考虑哪些性质使得点积在几何上具有意义。首先，它对向量是对称的：$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$。其次，它对每个因子是线性的：$(c\boldsymbol{u} + \boldsymbol{w}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v}) + \boldsymbol{w} \cdot \boldsymbol{v}$。第三，它确保了正长度：$\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$，且仅当 $\boldsymbol{v} = 0$ 时等式成立。这些性质——而非具体的公式——使得几何测量成为可能。

这一见解建议超越 $\mathbb{R}^n$ 进行概括。考虑单位区间上的连续函数空间 $C([0, 1])$。尽管这些向量是曲线而非箭头，我们仍然可能希望测量它们之间的角度或测试它们的正交性。积分公式

$$\langle f, g \rangle = \int_0^1 f(t) g(t)\, dt$$

正好提供了这样的度量。它具有使点积具有几何意义的关键性质：对称性、线性和正定性。当两个函数的乘积积分为零时，它们就被称为“正交”——这是微积分中（以及后来傅里叶分析中）熟悉的概念。

Example: 函数 $\sin(n\pi x)$ 和 $\sin(m\pi x)$ 在该内积下对于不同的整数 n、m 是正交的，这解释了傅里叶正弦级数各项的独立性。

这些例子激发了对捕捉它们共同本质的抽象定义的理解：

定义 5.1（内积）。在一个向量空间 V 上的 *inner product* 是一个函数 $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$，满足对于所有 $\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w} \in V$ 和 $c \in \mathbb{R}$：

1. 对称性：$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$ 2. 线性性：$\langle c\boldsymbol{u} + \boldsymbol{w}, \boldsymbol{v} \rangle = c\langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{w}, \boldsymbol{v} \rangle$ 3. 正定性：$\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$，当且仅当 $\boldsymbol{v} = 0$ 时取等号

配备有内积的向量空间称为一个 *inner product space*。

•

这个看似简约的定义提炼了使几何测量成为可能的本质特征。每一条性质都发挥着至关重要的作用：对称性确保角度具有良好定义；线性性将几何与向量空间结构联系起来；正定性保证长度和距离等概念具有实际意义。

例 5.2（加权内积）。在 $\mathbb{R}^n$ 上，我们不需要对所有坐标给予相同的权重。给定正的权重 a_1，...，a_n，公式

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\boldsymbol{a}} = \sum_{i=1}^{n} a_i u_i v_i$$

定义了一种内积，强调某些组件而非其他组件。这种加权测量在统计学中自然出现，其中权重可能反映测量不确定性，或者在力学中，它们编码质量分布。

◇

例 5.3（矩阵内积）。矩阵空间 $\mathbb{R}^{m \times n}$ 具有若干自然的内积。其中 *Frobenius inner product*，

$$\langle A, B \rangle_F = \operatorname{tr}(A^T B) = \sum_{i,j} a_{ij} b_{ij}$$

将矩阵视为一个长向量，包含其元素。其他内积可能根据矩阵元素的位置或统计显著性对不同的矩阵元素进行加权。

◇

每个内积都会在向量空间上施加其自身的几何结构，从而决定如何度量角度和长度。点积只是众多内积中最基本的一种——是 $\mathbb{R}^n$ 上最初级的内积。正如我们将看到的，内积的选择会深刻影响理论理解和实际计算。

Foreshadowing: 内积的选择塑造了从优化算法到量子测量的一切。我们将看到它的影响在本文中逐渐显现。

内积最基本的结果是它所诱导的长度概念。

定义 5.4（范数）。给定一个内积空间 V，一个向量 $\boldsymbol{v} \in V$ 的 *norm* 定义为：

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

这种诱导范数以与内积结构相一致的方式度量向量的长度。

•

尽管向量空间上存在许多范数，但由内积产生的范数具有特殊的几何性质。内积与其所诱导的范数之间微妙的相互作用，将引导我们在后续各节中对正交性的展开。

5.2 *Angles & Orthogonality*

在物理空间中，向量之间存在夹角。这一看似初等的观察——两个方向可以或多或少地对齐——其意义远远超出几何学。两个函数可以或多或少地相关；两个矩阵可以或多或少地对齐；两个量子态可以或多或少地独立。在每一种情况下，内积都通过一个最初在微积分中初露端倪的深刻公式揭示了这种角度关系。

我们的推导首先需要一个基本不等式——一个确保在任何内积空间中角度都有意义的不等式。

引理 5.5（柯西-施瓦茨不等式）。 *For any vectors $\boldsymbol{u}, \boldsymbol{v}$ in an inner product space,*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|$$

with equality if and only if one vector is a scalar multiple of the other.

Proof. 对于任意实数 t，内积的正定性要求：

$$0 \leq \|\boldsymbol{u} + t\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + 2t\langle \boldsymbol{u}, \boldsymbol{v} \rangle + t^2 \|\boldsymbol{v}\|^2$$

这个关于t的二次式必须对所有t都非负，这只有在其判别式非正时才可能：

$$4\langle \boldsymbol{u}, \boldsymbol{v} \rangle^2 \leq 4\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2$$

等号成立的情形可通过考察该二次方程何时恰有一个根来得到。

□

这个看似技术性的结果具有深远的意义。它保证内积与范数乘积之比的绝对值不可能超过 1——这正是我们通过熟悉的余弦关系来定义角度所需要的。

内积空间中非零向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 之间的 *angle* 是唯一的数 $\theta \in$，位于 $[0, \pi]$，满足：

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} \tag{5.1}$$

当这个角为 $\pi/2$ 时，我们称这些向量是 *orthogonal*，并记为 $\boldsymbol{u} \perp \boldsymbol{v}$。

角在所有内积空间中都存在，这一点令人惊叹。更令人惊叹的是，欧几里得角的熟悉性质如何在抽象环境中得以延续。正交向量的内积为零；小角度对应于几乎平行的向量；当内积为负时会出现钝角。这些性质超越了具体的背景，无论我们是在函数、矩阵，还是量子态之间测量角度。

例5.6（函数角度）。在带有内积$\langle f, g\rangle = \int_a^b f(t)g(t)\,dt$的$C([a,b])$中，当两个函数的乘积积分为零时，它们是正交的。函数之间的角度衡量它们的相关性——即它们在区间内的变化如何对齐。角度较小的函数变化相似；正交的函数彼此独立变化；角度为π的函数变化方向相反。$\diamond$

Example: 函数 $\sin x$ 和 $\cos x$ 在 $[-\pi, \pi]$ 上在该内积下是正交的——这一事实对傅里叶分析至关重要。参见例 5.9。

例 5.7（矩阵对齐）。在 Frobenius 内积下，矩阵 A 与 $B \in \mathbb{R}^{m\times n}$ 所成的角度为：

$$\cos\theta = \frac{\operatorname{tr}(A^T B)}{\|A\|_F \|B\|_F}$$

这衡量了它们的各个元素对齐程度，其中正交矩阵的逐元素相关性为零。对矩阵关系的这种几何解释揭示了隐藏在代数公式中的结构。$\diamond$

正交性在向量分解中显得尤其强大。当 $\boldsymbol{u} \perp \boldsymbol{v}$ 时，勾股定理得以推广：

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

正交向量的范数平方具有这种可加性，使得可以将复杂向量分解为更简单的正交分量——这一原则将指导我们对正交基和投影的研究。

引理 5.8（正交分解）。*Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ be mutually orthogonal nonzero vectors. Then they are linearly independent, and for any scalars $c_1, \ldots, c_k$:*

$$\left\| \sum_{i=1}^{k} c_i \boldsymbol{v}_i \right\|^2 = \sum_{i=1}^{k} c_i^2 \|\boldsymbol{v}_i\|^2$$

Foreshadowing: 当我们构造正交归一基时，这一分解原理将充分发挥其威力，从而实现最优近似和计算方法。

该证明源于内积的分配性质以及正交向量之间交叉项的消失。更为重要的是其含义：正交向量可以独立地组合，它们的

对任何可单独测量的总和的贡献，没有相互干扰。这一独立性原理——即正交分量可以单独分析——贯穿于从信号处理到量子力学的现代应用中。

我们以一个微妙的观察作结：虽然每个内积都会诱导一个范数，但并非每个范数都源自内积。例如，$\mathbb{R}^n$ 上的 ℓ^1 范数和 ℓ^∞ 范数缺乏内积所提供的几何结构。内积范数的特殊之处在于它们如何编码角度——这一能力将在我们发展正交基与变换理论时加以利用。

5.3 *Orthogonal & Orthonormal Bases*

我们迄今遇到的基通常源自于便利或习惯——坐标选择更多是出于习惯而非原则。然而，一些基本质上优于其他基，以尊重我们精心构建的内积结构的方式来衡量向量。这些基源自正交性概念，提供了既适用于理论理解又适用于实际计算的最佳框架。

一组向量 $\{v_1, \dots, v_k\}$ 在内积空间中是 *orthogonal*，如果每个向量都与其他所有向量垂直：

$$\langle v_i, v_j \rangle = 0 \quad \text{for all } i \neq j$$

当这些向量也具有单位长度，使得对所有i，$\|v_i\| = = 1$时，我们称这个集合为*orthonormal*。这样的集合将垂直性的几何优雅与单位向量的计算便利性相结合。

Example: 标准基 $\{\hat{\imath}、\hat{\jmath}、\hat{k}\}$ 对于 $\mathbb{R}^3$ 在点积下是正交归一的——这是一个如此熟悉的事实，以至于我们常常忘记它的重要性。

正交向量的力量在于它们如何分解它们所张成的空间。当 v_1，...，v_k 是正交的时，它们张成的空间中的任何向量都有一个唯一的表示：

$$x = \sum_{i=1}^{k} c_i v_i \quad \text{where} \quad c_i = \frac{\langle x, v_i \rangle}{\|v_i\|^2}$$

系数自然地来源于内积——无需解方程组。当向量是正交归一时，这进一步简化为 $c_i = \langle x、v_i \rangle$，内积本身直接揭示坐标。

示例 5.9（傅里叶级数）。函数 $\{\sin nx，\cos nx\}_{n=1}^{\infty}$ 在 $C[-\pi，\pi]$ 下构成一个正交集合，内积为

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x)\, dx$$

这种正交性——由欧拉发现并被傅里叶加以利用——解释了为什么三角级数能够如此有效地分解周期函数。傅里叶系数作为内积自然地产生，无需分部积分或其他技术手段。

◇

张成一个空间的正交或标准正交集合构成一种格外优雅的基。每个向量都具有可通过内积计算得到的唯一坐标；勾股定理对所有线性组合都成立；几何性质与代数性质完美契合。然而，我们不能仅凭愿望让这样的基出现——必须系统地加以构造。

Gram-Schmidt process 提供了这样的构造。从任意一组基 $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ 出发，我们构造一个张成同一空间的正交基 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$：

$$\begin{aligned} \boldsymbol{v}_1 &= \boldsymbol{b}_1 \\ \boldsymbol{v}_2 &= \boldsymbol{b}_2 - \Pi_{\boldsymbol{v}_1}\boldsymbol{b}_2 \\ \boldsymbol{v}_3 &= \boldsymbol{b}_3 - \Pi_{\boldsymbol{v}_1}\boldsymbol{b}_3 - \Pi_{\boldsymbol{v}_2}\boldsymbol{b}_3 \\ &\vdots \\ \boldsymbol{v}_k &= \boldsymbol{b}_k - \textstyle\sum_{i=1}^{k-1} \Pi_{\boldsymbol{v}_i}\boldsymbol{b}_k \end{aligned}$$

Perspective: 尽管格拉姆-施密特正交化看起来纯属理论，它却是现代推荐系统中关键算法的基础，其中正交的特征向量有助于防止用户表示中的冗余。

其中 $\Pi_{\boldsymbol{v}}\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v}$ 表示正交投影。通过减去其在所有之前向量方向上的分量，确保每个新向量与所有之前的向量正交。

例 5.10（多项式正交化）。考虑带有内积 $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\,dx$ 的空间 $\mathcal{P}_4$。从单项式基 $\{1, x, x^2, x^3, x^4\}$ 出发，Gram–Schmidt 方法（在比例因子意义下）得到 *Legendre polynomials*：

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \tfrac{1}{2}(3x^2 - 1) \\ P_3(x) &= \tfrac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \tfrac{1}{8}(35x^4 - 30x^2 + 3) \end{aligned}$$

值得注意的是，这些多项式不仅在引力理论中发挥着基础性作用，而且在量子力学中也至关重要，在其中它们用于描述角动量态；同时在数值积分中，它们还提供最优的求积点。

这些正交多项式是勒让德在研究引力势时发现的，它们源自在最简单的多项式基上施加正交性这一要求，自然而然地产生。系数日益增加的复杂性反映了这样一个事实：每一个新的多项式都必须与所有先前的多项式保持正交——这一约束导致各项之间的平衡愈发精细复杂。

◇

格拉姆–施密特过程，尽管理论上优雅，但在实践中可能遭受数值不稳定性。每一次投影都会累积计算

可能在最终基中破坏正交性的误差。一种更稳定的方法——*modified Gram-Schmidt*——是按顺序而非同时地应用投影。尽管在数学上等价，这种版本在有限精度算术中能更好地保持正交性。

Foreshadowing: 理论优雅与计算稳定性之间的相互作用预示着矩阵分解与数值线性代数之间更深层的联系。

构造了正交基后，我们可能会问哪些基最适合特定的问题。答案取决于我们希望保持或揭示什么结构：

1. 对于微分方程，特征函数的基底揭示动力学行为 2. 在信号处理中，傅里叶基底揭示频率内容 3. 对于量子系统，能量本征态的基底阐明测量 4. 在数据分析中，主成分基底优化方差的捕获

BONUS! 尽管每个有限维内积空间都存在一组正交规范基，但无限维空间可能无法实现如此完全的正交化。如果你继续学习 *functional analysis*，请记住这一点！

正交基的选择塑造了我们对空间及其向量的看法——这是我们在研究特征值和奇异值时将更深入探讨的主题。

5.4 *Adjoints & Transposes*

矩阵转置这一熟悉的运算蕴含着比乍看之下更深层的结构。当我们为矩阵 A 写下 A^T 时，我们所做的不仅仅是将元素沿对角线反射——而是编码了线性变换与内积之间的一种基本关系。这种关系一旦从其矩阵起源中抽象出来，便为理解变换如何与几何结构相互作用提供了关键。

首先考虑 $\mathbb{R}^n$ 中的矩阵转置及其标准内积。对于任意矩阵 A，其转置 A^T 满足一个关键性质：对所有向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，

$$\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, A^T\boldsymbol{y}\rangle$$

这个看似无害的方程揭示了一个深刻的事实：转置 A^T 不仅仅是一个矩阵运算，而是唯一的线性变换，它保持与 A 的内积关系。

定义 5.11（伴随）。设 V 和 W 为内积空间，且 $T : V \to W$ 为一个线性变换。T 的 *adjoint* 是唯一的线性变换 $T^* : W \to V$，满足：

Nota bene: 伴随算子的存在性与唯一性需要证明——二者都不能从定义中显然得到。存在性需要黎斯表示定理（这种深奥的理论工具我们将避免使用）。

$$\langle T\boldsymbol{v}, \boldsymbol{w}\rangle_W = \langle \boldsymbol{v}, T^*\boldsymbol{w}\rangle_V \tag{5.2}$$

对于所有 $\boldsymbol{v} \in V$ 和 $\boldsymbol{w} \in W$。 •

例 5.12（微分的伴随）。考虑具有内积 $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$ 的微分算子 $D : \mathcal{P}_2 \to \mathcal{P}_1$。其伴随算子

$D^* : \mathcal{P}_1 \to \mathcal{P}_2$ 满足:

$$\int_0^1 (Df)(x)g(x)\,dx = \int_0^1 f(x)(D^*g)(x)\,dx$$

分部积分表明，D^* 同时涉及积分项和边界项——这一关系对微分方程和力学中的变分原理都至关重要。◇

Nota bene: 伴随算子的理论可以扩展到无限维空间，但需要来自泛函分析的额外工具，包括算子的完备性和有界性。

伴随运算在反转线性变换方向的同时保持其代数结构:

引理 5.13. *For linear transformations S and T between finite-dimensional inner product spaces, and for any scalar c:*

1. $(S+T)^* = S^* + T^*$
2. $(cT)^* = cT^*$
3. $(ST)^* = T^*S^*$
4. $(T^*)^* = T$

Example: 对于$\mathbb{R}^2$中的旋转矩阵R，其伴随R^*对应于相反方向的旋转——这解释了为什么$R^TR = I$。

当我们为我们的空间选择正交归一基时，T^* 的矩阵是 T 的矩阵的转置。这解释了我们的记号：抽象的伴随运算将矩阵转置推广到任意内积空间。矩阵转置不过是伴随运算在坐标中的表示。

例 5.14（信号处理）。在数字信号处理中，滤波操作是信号空间之间的线性变换。滤波器的伴随描述了其时间反转的冲激响应——这一关系对于匹配滤波和信号检测至关重要。当滤波器保持信号能量（内积约束）时，其伴随与其逆算子密切相关。

◇

变换与其伴随之间的关系揭示了线性映射的几何性质。一些变换等于其自身的伴随（$T = T^*$）；另一些满足 $T^* = -T$。大多数介于这两个极端之间，它们偏离自伴性的程度刻画了它们对内积结构的扭曲。变换与伴随之间的这种相互作用将引导我们在后续章节中对正交变换的展开。

5.5 *Orthogonal Transformations*

最优雅的变换保留几何结构。　旋转在保持形状的同时改变视角；反射在保持角度的同时颠倒取向。　这样的变换——那些遵循

我们精心构建的内积结构贯穿整个工程学领域，从刚体力学到量子测量再到数据分析。它们的力量在于几何直觉与代数精确性的完美融合。

Note: 术语“正交”在这里意味着保持所有内积关系，而不仅仅是正交关系。一个更准确（尽管不太传统）的名称可能是“内积保持”或“正交归一”。

定义 5.15（正交变换）。在内积空间上的线性变换 $T : V \to V$ 如果保持内积不变，则称其为 *orthogonal*。

$$\langle T\boldsymbol{u}, T\boldsymbol{v}\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle$$

对于所有向量 $\boldsymbol{u}$、$\boldsymbol{v} \in V$。 •

正交变换保持长度和角度：对于所有$\boldsymbol{u}$和$\boldsymbol{v}$,

$$\|T\boldsymbol{v}\| = \|\boldsymbol{v}\| \quad : \quad \cos\theta = \frac{\langle T\boldsymbol{u}, T\boldsymbol{v}\rangle}{\|T\boldsymbol{u}\|\|T\boldsymbol{v}\|} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$$

这些几何约束具有强大的代数后果，将我们这里的工作与前一节中发展起来的伴随理论联系起来：

引理 5.16. *For a linear transformation T on a finite-dimensional inner product space, the following are equivalent:*

1. *T is orthogonal*
2. *$T^*T = TT^* = I$*
3. *$T^* = T^{-1}$*

当我们为我们的空间选择正交标准基时，正交变换具有特别优雅的矩阵表示。如果一个方阵 Q 是 *orthogonal*，则

$$Q^TQ = I = QQ^T \tag{5.3}$$

Foreshadowing: 当我们在第10章学习奇异值分解时，正交矩阵在计算中的强大作用将变得更加清晰；在那里，它们为多种用途提供了最优的坐标变换。

所有 $n \times n$ 正交矩阵的集合记作 $O(n)$。这类矩阵继承了显著的性质，使它们在计算中尤为有价值：

引理 5.17. *An orthogonal matrix Q satisfies:*

1. *Its columns (and rows) form an orthonormal basis*
2. *Its inverse equals its transpose: $Q^{-1} = Q^T$*
3. *It preserves lengths: $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$*
4. *It preserves angles: $(Q\boldsymbol{x})^T(Q\boldsymbol{y}) = \boldsymbol{x}^T\boldsymbol{y}$*
5. *Its determinant is ± 1*

例 5.18（刚体运动）。刚体在三维空间中的取向由一个正交变换来描述。其

在任何正交归一基下的矩阵表示 R 满足 $R^T R = RR^T = I$，且 $\det R = \pm 1$。当 $\det R = 1$ 时，该变换表示纯旋转；当 $\det R = -1$ 时，它包含反射。

这些机械约束——距离和角度的保持——源自这样一个物理原理：刚体在运动过程中保持其形状不变。R 的正交性以代数方式编码了这一基本的几何要求。

这种几何结构解释了为什么不同的物理量在旋转下会以不同方式变换。位置向量按 R 变换，而角动量向量按 $R^{-1} = R^T$ 变换，在尊重刚体运动物理规律的同时保持它们的几何关系。 ◇

例 5.19（信号变换）。离散傅里叶变换（DFT）在适当归一化后，在离散信号空间上提供一种正交变换。其正交性体现为帕塞瓦尔恒等式——时间域与频率域之间信号能量的守恒：

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2$$

其中 $x[n]$ 为时域信号，而 $X[k]$ 为其频域变换。该守恒性恰恰源于 DFT 基向量构成一组正交归一集。 ◇

Relax... 如果这说得通，那就太好了。如果不通，也没关系！ *You are having a dream...*

例 5.20（数据分析）。在高维数据分析中，正交变换为揭示结构提供了最优的坐标变换。主成分分析（PCA）——我们将在第 11 章中详细研究——旨在寻找一种正交的坐标变换，使数据的最大变异轴对齐。正交性确保新的坐标保持不相关，而内积的保持性保证在变换过程中不会丢失任何信息。◇

正交变换的复合仍然是正交的，而正交变换的逆也是正交的。这种在复合与取逆下的封闭性暗示着更深层的代数结构——正交变换在复合运算下构成一个群，尽管我们不再进一步探讨这种抽象。

Caveat: 这里的“群”一词具有精确的数学含义，描述的是一个在某个结合运算下封闭、并且具有单位元和逆元的集合。尽管我们不会展开这一理论，但它解释了正交变换如何组合的许多方面。

正交变换的力量在于其将几何与代数性质融为一体。它们为力学中的刚体运动提供数学框架，在波动方程和量子系统中保持能量，并为数据分析提供最优的坐标变换。它们在工程实践中的无处不在——从机器人学到信号处理再到机器学习——反映了一个更深层的真理：最有用的变换往往是那些保持基本结构的变换。

5.6 The QR Decomposition

我们在第 5.3 节中对正交性的探讨为另一种基本的矩阵分解奠定了基础。Gram-Schmidt 过程通过系统的投影将任意一组基向量转化为正交基——这一过程本身定义了原矩阵的一种自然分解。这种分解称为 QR 分解，它将任意矩阵表示为一个正交矩阵与一个上三角矩阵的乘积。

定义 5.21（QR 分解）。方阵 $A \in \mathbb{R}^{n\times n}$ 的 *QR decomposition* 将其表示为一个乘积

$$A = QR$$

其中 Q 是正交的（$Q^TQ = I$），而 R 是上三角矩阵。当 A 具有满秩时，如果我们要求 R 的对角元素为正，则该分解是唯一的。

•

考虑这种分解如何源自对 A 的列进行正交化。将 $A =$ 写成 $[\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_n]$，Gram-Schmidt 过程产生正交规范向量 $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$，其中每个 $\boldsymbol{q}_k$ 都是通过从 $\boldsymbol{a}_k$ 中减去其在先前向量上的投影而得到的。这些投影的系数连同归一化因子，自然地组装成一个上三角矩阵 R。这一构造揭示了 Q 为何必须是正交的，以及 R 为何呈现三角形形式——A 的每一列仅通过 $\boldsymbol{q}_i$ 表达到其自身的索引为止。

例 5.22（简单 QR 形式）。考虑矩阵

$$A = \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$$

直接计算可得

$$Q = \begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

正交因子 Q 捕捉了 A 中固有的旋转，而三角因子 R 表示缩放。这种几何分解——先进行纯旋转，随后进行缩放——体现了矩阵分解如何揭示其潜在结构。

◇

QR 分解不仅仅提供因式分解——它通过自然的阶段揭示了矩阵的作用。正如第 7 章研究的特征分解以及奇异值分解到

如第10章所述，QR 将一个变换分解为更简单、在几何上有意义的操作：

1. 正交因子 Q 提供了一个新的正交归一基 2. 三角因子 R 描述了在该基下的坐标 3. 它们的乘积 QR 重构了原始变换

引理 5.23（存在性与唯一性）。*Every nonsingular matrix A admits a unique QR decomposition with positive diagonal entries in R. When A is singular, the decomposition exists but loses uniqueness in ways determined by the matrix's rank.*

这种分解为理论分析和实际计算都提供了强有力的工具。方程组通过 QR 因子自然地得到求解：如果 $A\boldsymbol{x} = \boldsymbol{b}$，那么 $R\boldsymbol{x} = Q^T\boldsymbol{b}$ 可化简为回代。正交因子 Q 保持长度和角度，而 R 实现一种简单的三角变换。正如雕塑家的基本技法一样，熟练掌握 QR 分解使得通过对简单操作的精心组合，能够实现日益复杂的应用。

这种分解的力量源于它将几何视角与代数视角融为一体。尽管它是通过代数正交化发现的，但其真正的威力来自这样的几何洞见：任何线性变换都可以自然地通过正交基来分解。几何与代数的这种统一——这一贯穿我们整个论述的发展主题——在 QR 分解中体现得尤为清晰，在这里，抽象的正交性原理转化为切实可行的计算工具。

Text Embeddings & Semantic Search

将文本表示为高维空间中的向量，为内积几何提供了一个引人注目的应用。通过将词语或短语编码为向量（$\mathbb{R}^n$ (通常具有n非常大)的维度），我们可以利用本章中发展出的几何工具来度量语义相似性。内积结构（定义 5.1）将抽象的意义概念转化为具体的计算方法，使现代语言模型能够通过几何运算来处理和理解文本。

关键洞见在于：通过归一化内积来度量的嵌入向量之间的夹角能够刻画语义相似性。对于嵌入向量 $\boldsymbol{v}$、$\boldsymbol{w} \in \mathbb{R}^n$，它们的 *cosine similarity* 定义为：

$$\cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$$

这个mea 当然，范围从 -1 到 1，自然地从 geomet 中显现。 ric 结构-

我们在第5.2节中建立的 ture。具有相似含义的词或短语会产生彼此之间夹角较小的嵌入向量，而不相关的概念在嵌入空间中则几乎正交。

Caveat: 基于嵌入的相似性效果在很大程度上依赖于嵌入向量的质量。现代语言模型通过在大量文本语料库上进行广泛训练，学习这些嵌入，优化几何结构以反映语义关系。

Algorithm 1: Semantic Search via Cosine Similarity

Data: Query embedding $\boldsymbol{q}$, Database of embeddings $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N\}$
Result: Indices of k most similar items
Normalize query: $\boldsymbol{q} \leftarrow \boldsymbol{q}/\|\boldsymbol{q}\|$;
for $i \leftarrow 1$ **to** N **do**
  Normalize database vector: $\boldsymbol{d}_i \leftarrow \boldsymbol{d}_i/\|\boldsymbol{d}_i\|$;
  Compute similarity: $s_i \leftarrow \langle \boldsymbol{q}, \boldsymbol{d}_i \rangle$;
end
Return indices of k largest similarities;

例 5.24（词类比）。训练良好的嵌入空间的一个显著特征是，它们能够通过向量算术来捕捉语义关系。经典的例子

$$\text{king} - \text{man} + \text{woman} \simeq \text{queen}$$

展示了嵌入空间的几何形状如何编码有意义的关系：对于语义上合适的文本嵌入，将 *king* 和 *man* 嵌入的向量差异加到 *woman* 上，产生一个与 *queen* 嵌入在余弦相似度上接近的向量。

◇

示例 5.25（文档检索）。给定一个文档语料库，每个文档通过词嵌入的平均或复杂池化编码为一个向量，语义搜索归结为在余弦相似度下寻找最近邻。与传统的关键词匹配相比，这种几何方法能更好地捕捉主题相似性，因为使用不同但语义相关的术语的文档仍然会产生具有较小角度分离的向量。◇

Foreshadowing: 在第10章中，我们将看到奇异值分解如何通过降维在高维嵌入空间中实现高效的近似最近邻搜索。

这种几何方法的力量超越了简单的相似性搜索。现代语言模型构建了多个嵌入空间层，每个层捕捉语言结构的不同方面。本章中发展出的原理——从内积的抽象性质（定义 5.1）到角度的具体几何（引理 5.16）——为这些复杂的自然语言处理系统提供了数学基础。

基于嵌入的方法在自然语言处理中的成功 exemplifies 一个更广泛的原则：数据中的许多复杂关系可以通过精心设计的几何结构来捕捉。无论是比较文档、分析分子结构，还是处理社交网络，我们开发的内积几何提供了一个强大的框架，用于衡量和利用相似性。

Quantum Measurement & Observable Operators

量子力学或许提供了内积几何最优雅的应用——不仅仅作为一种数学上的便利，而是作为基础

tal 物理定律。本章中发展出的数学结构——内积、正交性以及伴随算子——自然地成为描述量子态及其测量的语言。对于通过物理学接触过量子力学的工程专业学生而言，本节基于有限维线性代数的清晰性，提供了一种全新的视角。

Historical Note: 用于向量（*kets*）的记号 $|\psi\rangle$ 以及用于对偶向量（*bras*）的记号 $\langle\phi|$ 由狄拉克于1939年引入，目的是让内积看起来像夹心括号：$\langle\phi||\psi\rangle$。尽管最初被视为只是巧妙的排版，这种记号在量子计算中被证明异常有效。

关键的洞见在于，量子态不过是内积空间中的向量。量子比特，或称 *qubit*——量子信息的基本单位——存在于一个二维复内积空间 $\mathbb{C}^2$ 中。任何状态都可以写成两个正交归一基态的线性组合，传统上记为 $|0\rangle$ 和 $|1\rangle$（（向量的物理学记号，我们将简要采用）):

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \quad \text{where } |\alpha|^2 + |\beta|^2 = 1$$

归一化条件 $\| \, |\psi\rangle \, \| = 1$ 反映了一个基本的物理原理：概率必须加和为一。

Nota bene: 量子力学传统上使用复向量空间，而本章的关键几何概念——内积、正交性以及伴随——可以自然地从 $\mathbb{R}^n$ 推广到 $\mathbb{C}^n$。主要的变化在于，内积不再是对称的，而是共轭对称的。

当我们测量一个量子态时，本质上是在计算内积。系统处于状态 $|\psi\rangle$ 时测得其处于状态 $|\phi\rangle$ 的概率等于它们内积的模平方：

$$\mathbb{P}(\phi|\psi) = |\,\langle\phi|\,|\psi\rangle\,|^2$$

这种几何解释将量子测量从神秘的现象转变为具体的计算：我们只需将状态向量投影到各种测量方向上，并从得到的内积计算概率。

更一般的测量对应于*observable operators*——与其自身伴随相同的线性变换（在复数情形下在物理学中称为*Hermitian*）。这些算符的本征值给出可能的测量结果，而其本征向量提供相应的测量基。伴随算符与物理测量之间的这种联系，对量子系统中可被观测的内容施加了根本性的约束。

示例 5.26（量子比特测量）。考虑测量处于某一状态的一个量子比特

$$|\psi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

在标准基中。该状态赋予测量0或1的概率相等：

$$\mathbb{P}(0|\psi) = |\langle 0|\,|\psi\rangle|^2 = \left|\left\langle \begin{pmatrix}1\\0\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}\right\rangle\right|^2 = \frac{1}{2}$$

同样地，对于 $\mathbb{P}(1|\psi)$ 也是如此。基态的正交归一性确保这些概率之和为一。$\diamond$

A m 矿石一般的测量对应于可观测的歌剧　　类似 Tor 的

$$A = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$$

该算子显然是自伴的（$A^* = A$），并且具有特征值 ± 1，以及对应的归一化特征向量：

$$\boldsymbol{v}_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{v}_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Caveat: 将量子可观测量限制为自伴算符源于一个物理要求：测量结果必须是实数。自伴算符的本征值始终为实数，正好提供了这一保证。

当我们测量这个可观测量时，我们将得到 $+1$ 或 -1，其概率由将我们的态投影到这些本征向量上来决定。

该框架可以自然地扩展到更高维度的系统。一个由 n 个量子比特组成的寄存器存在于一个 2^n 维的复向量空间中——这既解释了量子计算的强大之处，也解释了其脆弱性。每增加一个量子比特，态空间的维度就会翻倍，从而在操纵量子态时实现巨大的并行性。然而，这种指数级扩展也解释了为什么量子态在经典计算机上如此难以模拟：仅表示 50 个量子比特就需要存储 2^{50} 个复数。

第5.3节的原理在量子纠错中有着直接的应用。通过利用正交子空间将逻辑量子比特编码到更高维的空间中，我们可以通过投影操作检测并纠正某些类型的错误。内积与正交补的数学工具为使量子计算在噪声下保持鲁棒性提供了理论基础。

例 5.27（错误检测）。一种简单的量子错误检测编码将一个逻辑量子比特 $\alpha|0\rangle + \beta|1\rangle$ 编码成一个双量子比特态：

$$\alpha|00\rangle + \beta|11\rangle$$

此编码使用正交状态$|00\rangle$和$|11\rangle$，将$|01\rangle$和$|10\rangle$保留为错误指示符。当发生错误时，通常会将状态移到编码空间的正交补空间——通过投影测量允许进行检测。

◇

量子系统的研究 exemplifies 本章中发展出的几何结构如何在物理法则中自然出现。内积决定测量概率，伴随算子限制可观察算子，正交子空间使得误差修正成为可能。数学和物理的这种统一既为量子工程提供了实用工具，也为几何在物理法则中的作用提供了更深刻的见解。

对于工程专业的学生而言，量子力学因此不再是一堆神秘的公设，而是本文中系统发展起来的线性代数概念的自然应用。在第5.1节中首次以抽象形式出现的内积结构，显现为在最小尺度上物理系统行为的基础。

□ ——————————————————————— □

Exercises: Chapter 5

1. 考虑在 $\mathcal{P}_2$ 上给定的内积 $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$。求 $p(x) = 1 + 2x + x^2$ 的范数。

2. 设 V 为具有内积 $\langle A,\ B\rangle = \text{tr}(A^TB)$ 的 2×2 矩阵空间。求矩阵 $A = \begin{bmatrix}1&0\\0&1\end{bmatrix}$ 与 $B = \begin{bmatrix}1&-1\\1&1\end{bmatrix}$ 之间的角度。

3. 证明 $\langle f, g\rangle = f(0)g(0) + f(1)g(1)$ 定义了 $\mathcal{P}_1$ 上的内积。找到一个与 $p(x) = x$ 正交的向量。

4. 在 $\mathbb{R}^3$ 上，定义 $\langle \boldsymbol{u},\ \boldsymbol{v}\rangle_w = 2u_1v_1 + 3u_2v_2 + u_3v_3$。证明这是一个内积并找到单位长度的向量。

5. 对于矩阵 A 和 B，证明 $\text{tr}(A^TB) = \text{tr}(B^TA)$ 并利用此证明 Frobenius 内积是对称的。

6. 设 $\boldsymbol{u} = (1, 2, 2)^T$ 和 $\boldsymbol{v} = (2, 4, -1)^T$。使用 Gram-Schmidt 过程为 $\text{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ 求一个标准正交基，然后将其扩展为 $\mathbb{R}^3$ 的一个标准正交基。写出得到的正交矩阵 Q 并验证 $Q^TQ = I$。

7. 对于整数 m，$n \geq 0$，证明函数 $\cos(mx)$ 和 $\sin(nx)$ 在 $[-\pi, \pi]$ 上在标准的 L^2 内积下是正交的。

8. 设 $C([0, 2\pi])$ 表示定义在 $[0, 2\pi]$ 上的连续函数，并具有内积 $\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\,dx$。证明 $\{1, \cos x, \sin x\}$ 构成一个正交集，但不是一个正交归一集。求出合适的缩放因子使其成为正交归一集。

9. 考虑带有 Frobenius 内积的 2×2 对称矩阵空间 sym_2。为该空间找到一个正交规范基，并验证其具有正确的维数。

10. 设 A 为矩阵：

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

通过对其进行格拉姆-施密特正交化来求其 QR 分解

列。验证 $Q^TQ = I$ 和 $A = QR$。11. 证明如果 Q 是一个 $\det Q > 0$ 的 2×2 正交矩阵，那么 Q 必须是一个旋转矩阵。如果相反 $\det Q < 0$，你能对它作何判断?

12. 对于平面中的点，定义 $\langle \boldsymbol{p},\ \boldsymbol{q}\rangle_M = p_1q_1 + p_1q_2 + p_2q_1 + 2p_2q_2$。写出矩阵 M，使得该内积等于 $\boldsymbol{p}^TM\boldsymbol{q}$。这是一个有效的内积吗？请证明你的答案。如果有效，在该内积下找出两个正交向量。

13. 在 $\mathbb{R}^2$ 上由 $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 4u_1v_1 + 4u_1v_2 + 4u_2v_1 + 5u_2v_2$ 定义一个内积。求在该内积下 $\boldsymbol{u} = (1, 0)^T$ 与 $\boldsymbol{v} = (0, 1)^T$ 之间的夹角。

14. 考虑一个非零向量 $\boldsymbol{v} \in \mathbb{R}^3$，令 $\Pi = \Pi_{\boldsymbol{v}}$ 表示到 $\text{span}\{\boldsymbol{v}\}$ 上的正交投影。证明 Π 是对称的：$\Pi^T = \Pi$。15. 对于向量 $\boldsymbol{u}$、$\boldsymbol{v} \in \mathbb{R}^n$，证明当且仅当一个向量是另一个向量的标量倍数时，柯西–施瓦茨不等式取等号。16. 证明如果 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是内积空间中的正交向量，则勾股定理成立：$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$。

17. 内积对所有向量 $\boldsymbol{u}$、$\boldsymbol{v}$ 满足 $\|\boldsymbol{u} + \boldsymbol{v}\|^2 + \|\boldsymbol{u} - \boldsymbol{v}\|^2 = 2(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2)$。利用内积的性质证明这一点。

18. 设 A 和 B 为 2 个 $\times$ 2 矩阵。使用 Frobenius 内积 $\langle A, B\rangle = \text{tr}(A^TB)$, 证明如果 A 和 B 作为矩阵空间中的向量是正交的，则 $\text{tr}(AB^T) = 0$。是否存在一种几何解释，用 A 和 B 在 $\mathbb{R}^2$ 中对向量的作用来理解这种正交性?

19. 设 $T : V \to V$ 是一个内积空间上的线性变换。证明：如果对所有 $\boldsymbol{v} \in V$ 都有 $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$，则 $T = 0$。
20. 设 $V = \mathbb{R}^2$ 具有标准内积。对于一个固定的非零向量 $\boldsymbol{a}$，定义 $T : V \to V$ 如下：

$$T\boldsymbol{x} = \boldsymbol{x} - 2\frac{\langle \boldsymbol{x}, \boldsymbol{a}\rangle}{\|\boldsymbol{a}\|^2}\boldsymbol{a}$$

证明 T 保持内积：$\langle T\boldsymbol{x}、T\boldsymbol{y}\rangle = \langle \boldsymbol{x}、\boldsymbol{y}\rangle$ 对所有 $\boldsymbol{x}、\boldsymbol{y} \in V$ 成立。21. 对于内积空间的一个子空间 $W < V$ 和向量 $\boldsymbol{v} \in V$，证明：

$$\|\boldsymbol{v} - \Pi_W\boldsymbol{v}\| \le \|\boldsymbol{v} - \boldsymbol{w}\|$$

对于任意 $\boldsymbol{w} \in W$，当且仅当 $\boldsymbol{w} = \Pi_W\boldsymbol{v}$ 时成立。将此结果几何解释。

22. 平板上点 (x, y) 处的温度由 $T(x, y) = e^{-x^2-y^2}$ 给出。使用内积 $\langle f, g\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)g(x, y)\,dx\,dy$，求 $\|T\|$。
23. 向量 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ 的 *Gram matrix* G 的元素为 $g_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle$。证明 G：是对称的，并且当且仅当这些向量线性相关时其行列式为零。

24. 设 V 为定义在 $[0, 1]$ 上的连续函数空间，其内积为 $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$。求常数 a 和 b，使得 $f(x) = a + bx$ 同时与 1 和 x^2 正交。

25. 对于定义在 $[0, 1]$ 上的连续函数，考虑：

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx + f(0)g(0) + f(1)g(1)$$

证明这定义了一个内积，并求 $f(x) = x(1 - x)$ 的范数。26. 在连续可微函数空间 $C^1([0, 1])$ 上，定义：

$$\langle f, g\rangle = \int_0^1 (f(x)g(x) + f'(x)g'(x))\,dx$$

证明这定义了一个内积，并求 $f(x) = x$ 与 $g(x) = x^2$ 之间的夹角。

27. 设 $\mathcal{P}_3[0, 1]$ 表示次数至多为 3 的多项式，配备内积 $\langle f, g\rangle = \int_0^1 f'(x)g'(x)\,dx$。证明这在满足 $f(0) = 0$ 的多项式子空间上定义了一个有效的内积；然后，在该内积下找到一个同时与 x 和 x^2 正交的多项式。

Chapter 6
Orthogonal Decomposition & Data

"quadrangular the building rose the heavens squared by a line"

几何与代数的结合在正交性概念中达到了第一次顶点。我们构建的内积结构转变了我们对第三章中基本空间和运算的理解。曾经仅仅是代数的内容——核与像、商与补集——通过垂直性获得了几何意义。这一几何视角不仅阐明了理论，而且引导了计算，导致了解决系统和拟合数据的最优方法。

我们的任务是通过正交性的视角重新审视前几章所奠定的基础。一个变换的核不仅在代数上是不可见的，而且在几何上与其共像正交。商空间不仅是代数上的同一化，而且会诱导出正交分解。基本定理本身表达的不仅是维数的核算，更是将定义域和值域几何地分裂为彼此垂直的因子。

这种几何重解释立即带来了实际成果。当线性系统的精确解不存在时，正交投影提供了在自然误差度量下的最佳近似。被噪声污染的数据通过投影到适当的子空间获得干净的表示。内积的抽象机制从理论走向实践，为一系列工程问题提供了最优方法。

前方的道路以全新的视角重新审视熟悉的领域。我们首先通过正交性重新解释四个基本空间。这种几何理解自然引导我们到投影算符——分解的主力工具。有了这些工具，我们回到基本定理，作为正交分解的发源。

空间。最后，我们将这一几何方法应用于最小二乘逼近的实际问题，其中理论引导我们找到过定方程组的最优解。

这种几何与代数的综合——内积和线性变换——为许多现代计算方法奠定了基础。我们在这里发展的原则将指导我们在接下来的章节中研究特征值、奇异值和主成分。

*Foreshadowing*在这里研究的正交投影构成了机器学习中降维的理论基础，其中高维数据被投影到低维子空间，捕捉到关键特征。

6.1 Orthogonal Subspaces & Complements

到目前为止我们所研究的向量空间，通过第 5 章中的内积几何获得了更为丰富的结构。正如单个向量可以彼此垂直一样，整个子空间也可以正交，从而产生自然的几何分解，这些分解既阐明理论，又支持计算。

定义 6.1（正交补）。内积空间 V 的两个子空间 U 和 W 是 *orthogonal*，记作 $U \perp W$，如果一个子空间中的每个向量都与另一个子空间中的每个向量正交：

$$\langle \boldsymbol{u}, \boldsymbol{w} \rangle = 0 \quad \text{for all } \boldsymbol{u} \in U,\ \boldsymbol{w} \in W$$

子空间 $U < V$ 的 *orthogonal complement*，记作 $U^\perp$，是所有与 U 正交的向量所组成的子空间：

$$U^\perp = \{\boldsymbol{v} \in V : \langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0 \text{ for all } \boldsymbol{u} \in U\}$$

•

事实上，$U^\perp$ 确实是一个子空间，这一点可以直接由内积的性质得出：零向量当然与 U 的所有元素正交，而与 U 正交的向量的线性组合仍然与 U 正交。

引理 6.2. *For V a finite-dimensional inner product space and any $U < V$:*

1. $(U^\perp)^\perp = U$
2. 维度 U + 维度 $U^\perp$ = 维度 $V0\}$
3. $U \cap U^\perp = \not{4}$. $V = U \oplus U^\perp$

Proof. 由定义可知 $U \subseteq (U^\perp)^\perp$。对于反向包含，任选 U 的一组正交规范基 $\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k\}$，并将其扩充为 V 的一组正交规范基。该基的补空间张成 $U^\perp$，由此推出维数相等。其余性质由这一显式构造得出。

□

最后一个性质尤为重要：V 中的每个向量都可以唯一地分解为分别位于 U 和 $U^\perp$ 中的正交分量。我们将用 $V = U \boxplus U^\perp$ 来表示将 V 作 *orthogonal direct sum*，分解为彼此互补且正交的子空间 U 和 $U^\perp$。这种 $\boxplus$ 分解将被证明是我们在下一节中发展投影算子的基础。

例6.3（矩阵子空间）。考虑例5.3中的空间$\mathbb{R}^{n\times n}$，其弗罗贝尼乌斯内积为$\langle A, B\rangle = \operatorname{tr}(A^T B)$。对称矩阵的子空间$\operatorname{sym}_n$与反对称矩阵的子空间$\operatorname{skew}_n$互为正交补：

$$\mathbb{R}^{n\times n} = \textsc{sym}_n \boxplus \textsc{skew}_n$$

$$\textsc{sym}_n = \{A : A^T = A\} \quad \text{and} \quad \textsc{skew}_n = \{A : A^T = -A\}$$

这种正交分解揭示，任意矩阵 A 都可以唯一地分解为对称部分和反对称部分：

$$A = \frac{A + A^T}{2} + \frac{A - A^T}{2}$$

这种分解在力学中具有重要应用，其中对称矩阵常常表示应力和应变。 ◇

由正交补所提供的几何视角将改变我们对第3章中基本子空间的理解。此前看似纯粹代数的构造——核与像、商与对偶——将通过正交性获得新的几何意义。这种通过垂直性的重新诠释将引导我们在接下来的章节中发展用于求解方程组和进行数据近似的最优方法。

6.2 *Projections & Quotients*

投影这一概念贯穿于数学与工程领域。阳光投下的影子将三维物体投射到平面上；测量员的地图将地球的曲面投射到平整的纸张上；统计学家将高维数据投射到信息丰富的低维摘要中。这些看似多样的例子共享着一个共同的数学本质：通过系统性的降维，用更简单的对象来近似复杂对象。

我们所构建的内积结构将这一直观概念转化为精确的数学。给定一个子空间 $W < V$，我们试图

通过它们在 W 中的“投影”来近似 V 中的任意向量——这些向量在完全位于 W 中的同时，使其与原始向量的距离最小化。这个几何问题直接引出了正交投影这一概念，它将第 3 章的理论结构与现代数据分析的计算方法统一起来。

定义 6.4（正交投影）。设 $W < V$ 是一个内积空间的子空间。将 *orthogonal projection* 投影到 W 的正交投影是线性变换 $\Pi_W : V \to V$，满足：

1. $\Pi_W \boldsymbol{v} \in W$ 对所有 $\boldsymbol{v} \in V$ (投影到W)2. $\boldsymbol{v} - \Pi_W \boldsymbol{v} \perp W$ 对所有 $\boldsymbol{v} \in V$ (正交投影)

Nota bene: 投影 Π_W 由这些性质唯一确定。尽管其他变换也可能将向量映射到 W，但只有正交投影保持误差的垂直性。

这个看似抽象的定义蕴含着一个强大的优化原理：在内积所诱导的自然距离度量下，$\Pi_W \boldsymbol{v}$ 在 W 内对 $\boldsymbol{v}$ 给出了最佳近似。

引理 6.5（最佳逼近）。*For any $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$:*

$$\|\boldsymbol{v} - \Pi_W \boldsymbol{v}\| \leq \|\boldsymbol{v} - \boldsymbol{w}\|$$

with equality if and only if $\boldsymbol{w} = \Pi_W \boldsymbol{v}$.

Proof. 对于任何 $\boldsymbol{w} \in W$，误差向量 $\boldsymbol{v} - \boldsymbol{w}$ 可以分解为正交分量：

$$\boldsymbol{v} - \boldsymbol{w} = (\boldsymbol{v} - \Pi_W \boldsymbol{v}) + (\Pi_W \boldsymbol{v} - \boldsymbol{w})$$

其中第一项与W正交，第二项位于W中。根据勾股定理：

$$\|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{v} - \Pi_W \boldsymbol{v}\|^2 + \|\Pi_W \boldsymbol{v} - \boldsymbol{w}\|^2$$

正确的项在$\boldsymbol{w} = \Pi_W \boldsymbol{v}$时恰好消失。 □

当 W 具有正交规范基 $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$ 时，投影呈现出特别优雅的形式：

$$\Pi_W \boldsymbol{v} = \sum_{i=1}^{k} \langle \boldsymbol{v}, \boldsymbol{w}_i \rangle \boldsymbol{w}_i$$

每个系数自然地作为与相应基向量的内积出现——无需解任何方程组。该公式揭示了投影作为一种谱分解类型，从 $\boldsymbol{v}$ 中提取出与 W 的基向量对齐的精确分量。

示例 6.6（信号处理）。考虑在 $[-\pi, \pi]$ 上的连续函数空间 $V = C([-\pi, \pi])$，其内积为示例 5.9 中的 L^2。由 $\{1、\cos x、\sin x\}$ 张成的子空间 W 捕捉信号的直流分量和第一谐波分量。投影 Π_W 实现了一个低通滤波器，通过信号的第一傅里叶分量来逼近任意信号。误差 $v - \Pi_W v$ 表示被投影滤除的高频内容。$\diamond$

投影算子具有若干性质，揭示了它们的几何和代数特征：

引理 6.7（投影性质）。*The orthogonal projection Π_W satisfies:*

1. *Idempotence:* $\Pi^2 = \Pi$
2. *Self-adjointness:* $\Pi^* = \Pi$
3. *Complementarity:* $I - \Pi_W = \Pi_{W^\perp}$

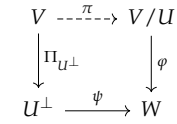
where $W^\perp$ denotes the orthogonal complement of W

.

这些性质反映了投影的几何本质——对其应用两次不会产生额外效果；它尊重内积结构；并且它将空间分解为正交的部分。最后一个性质提供了特别的洞见：投影到 W 和投影到 $W^\perp$ 会将任意向量分解为互补的分量。

Foreshadowing: 这一分解原理将在第10章中发挥其全部威力，其中奇异值分解提供了用于近似数据的最优正交投影序列。

这种分解阐明了我们先前对商空间的研究。当我们用子空间 U 去商 V 时，每个等价类都由相差 U 中元素的向量组成。从正交性的视角来看，我们可以用其在 $U^\perp$ 上的投影来表示每个等价类——即到原点距离最小的唯一成员。商空间 V/U 自然地与 $U^\perp$ 同构，而投影 $\Pi_{U^\perp}$ 提供了一个显式的同构。



Nota bene: 虚线箭头表示自然商映射，而实线箭头表示显式线性变换。

投影与商之间的关系可以通过交换图来可视化。自然的商映射π (虚线)使该图交换：从V到W的两条路径给出等价的结果。

此处 V/U 表示商空间，$U^\perp$ 是正交补空间，而 W 是与二者同构的任意空间（通常取为某个线性变换的像）。该图是可交换的，因为从 V 到 W 的两条路径给出相同的结果——无论是先投影到 $U^\perp$，还是先取商空间 V/U。这种几何视角阐明了为何商空间与正交补为理解线性变换的有效定义域提供了等价的方式。

例6.8（数据中心化）。考虑$\mathbb{R}^d$中的一组向量$\{x_1, \cdots\cdots, x_n\}$。由$1 = (1, \cdots\cdots, 1)^T$张成的子空间$U$表示均匀

翻译。将数据投影到 $U^\perp$ 上，通过减去均值来中心化数据——这是数据分析中的一个基本预处理步骤。商 $\mathbb{R}^d/U$ 捕捉了数据云的内在形状，而不受其绝对位置的影响。$\diamond$

正交投影的力量远远超出了这些基本示例。当线性方程组不存在精确解时，向适当的子空间投影可以得到最优近似。当数据包含噪声时，向信号子空间投影能够实现滤波与压缩。当复杂系统需要简化模型时，向低维空间投影在准确性与复杂性之间取得平衡。这些应用以及更多内容，都源自定义6.4中所蕴含的简单几何原理。

6.3 *The Fundamental Theorem Redux*

线性代数的基本定理在有限维内积空间的视角下，通过正交性的透镜揭示出更深层的结构。最初看似只是一些维数关系的集合，如今以其真实形态呈现出来：关于空间几何分解的陈述。四个基本子空间——核、像、余核与余像——之间的联系不仅体现在维数的计数上，更体现在相互的正交性之中。这种几何化的理解虽受限于有限维情形，却改变了我们对线性变换的认识，为理论洞见与计算上的实用方法提供了基础。

回顾一下，我们用 $A \boxplus B$ 表示一个正交直和：子空间 A 和 B，它们既是互补的（$V = A \oplus B$），又是正交的（$A \perp B$）。

定理 6.9（线性代数基本定理（几何形式））。
Any linear transformation $T\colon V \to W$ between finite-dimensional inner product spaces induces orthogonal decompositions of both domain and codomain:

$$V = \ker T \boxplus (\ker T)^\perp \qquad \text{and} \qquad W = \mathit{im}\, T \boxplus (\mathit{im}\, T)^\perp \tag{6.1}$$

Nota bene: 这里的术语 *naturally isomorphic*意味着这些同构源自几何结构本身。

These decompositions are connected by the following:

1. *The restriction of T to $(\ker T)^\perp$ gives an isomorphism $(\ker T)^\perp \cong \mathit{im}\, T$*
2. *The coimage $V/\ker T$ is naturally isomorphic to $\mathit{im}\, T$*
3. *The cokernel $W / \mathit{im}\, T$ is naturally isomorphic to $(\mathit{im}\, T)^\perp$*
4. *The orthogonal projections $\Pi_{(\ker T)^\perp}$ and $\Pi_{\mathit{im}\, T}$ commute with T*

这些分解通过几何而非代数阐明了四个基本子空间。核表示对T不可见的向量；其正交补刻画了有效的输入。像空间包含所有可能的输出；其正交补衡量了该变换的不足。每一对互补空间都提供了T在其定义域和值域上作用方式的完整视图。

例 6.10（矩阵变换）。对于矩阵 $A \in \mathbb{R}^{m\times n}$，这些分解在计算上就 row($A$) 和 col($A$) 而言具有直接的重要意义，它们分别是行空间和列空间。

1. $\mathbb{R}^n = \ker(A) \boxplus \text{row}(A)^T$ 将输入空间划分 2. $\mathbb{R}^m = \text{col}(A) \boxplus \ker(A^T)$ 将输出空间划分 3. 当精确解不存在时，投影 $\Pi_{\text{row}(A)^T}$ 和 $\Pi_{\text{col}(A)}$ 提供最优的近似解

Foreshadowing: 出现的正交补暗示了第七章中正交矩阵的概念，其中整个变换保持垂直性。 ◇

这些几何分解产生了第3章中的代数关系作为推论：

推论 6.11（秩-零度再论）。 *For a linear transformation $T : V \to W$ between finite-dimensional inner product spaces:*

1. 维度 $V =$ 维度 $\ker T +$ 维度 *im* T *2.* 维度 $W =$ 维度 *im* $T +$ 维度 *coker* T *3.* 维度 *im* $T =$ 维度$(\ker T)^\perp$ 维度 *im* T *and* 维度 *coim* T
4. The rank equals both

该证明源自正交分解，互补子空间的维数之和等于整个空间的维数。此前看似神秘的代数巧合，如今显现为几何分解的自然结果。

这种几何视角指导计算。在求解 $T\boldsymbol{v} = \boldsymbol{w}$ 时：

1. 将 $\boldsymbol{w}$ 投影到 im T 上以检验可解性 2. 若可解，在 $(\ker T)^\perp$ 中找到一个特解 3. 若不可解，将其投影到 im T 上以获得最佳近似

基本定理因此揭示了它不仅仅是代数——它表达了线性变换的基本几何结构。几何与代数的统一提供了理论上的洞察和实际的方法，这一主题将在我们继续讨论最小二乘问题及其应用时得到进一步深化。

6.4 *The Pseudoinverse*

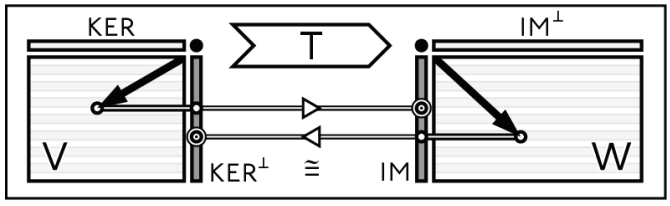
线性代数的基本定理揭示了任何线性变换如何引发四个基本子空间，这些子空间通过定义域和陪域的正交分解相互连接。这个优雅的结构提出了一个自然的问题：我们能否定义一个反向变换，以某种方式撤销原始映射的作用，同时尊重这些几何关系？答案在于*pseudoinverse*——矩阵逆的推广，它即使对于奇异或矩形矩阵，也能提供最优的近似逆。

定义 6.12（伪逆）。对于有限维内积空间之间的线性变换 $T : V \to W$，*pseudoinverse* (或 *Moore-Penrose inverse*) $T^\dagger : W \to V$ 是满足以下所有条件的唯一线性变换:

1. $TT^\dagger T = T$ (第一一致性条件) 2. $T^\dagger TT^\dagger = T^\dagger$ (第二一致性条件) 3. $(TT^\dagger)^* = TT^\dagger$ (第一伴随条件) 4. $(T^\dagger T)^* = T^\dagger T$ (第二伴随条件)

其中，* 表示关于 V 和 W 上内积的伴随运算。

这一抽象定义虽然完备，却可能遮蔽了伪逆深刻的几何意义。通过基本定理的视角，伪逆自然地从对四个基本子空间的正交投影中显现出来:



伪逆 $T^\dagger$ 作为通过基本子空间的 T 的逆，其中 T 先投影到 $(\ker T)^\perp$，再同构地映射到 $\text{im}\, T$；而 $T^\dagger$ 则先投影到 $\text{im}\, T$，再同构地映射到 $(\ker T)^\perp$。

这种对称性揭示了伪逆并非一种权宜的构造，而是一个自然的反向映射，它尊重由我们的内积所施加的正交结构。

对于矩阵，伪逆具有一种特别优雅的形式。当 A 具有满列秩时，其伪逆为:

$$A^\dagger = (A^T A)^{-1} A^T$$

该公式与本章前面研究的正交投影直接相关：$A^T A$ 可逆，恰恰因为 A 具有满列

排名，使得 $\ker(A) = \{0\}$。当 A 具有满秩时，伪逆变为：

$$A^\dagger = A^T(AA^T)^{-1}$$

这些公式说明了伪逆如何将矩阵求逆的概念推广到矩形矩阵，在没有精确逆的情况下，提供最佳的近似逆。

例 6.13 （正交投影）。考虑将正交投影 $\Pi_U : V \to V$ 投影到子空间 $U < V$ 上。它的基本空间是：

- $\ker(\Pi_U) = U^\perp$
- $\operatorname{im}(\Pi_U) = U$

伪逆 $\Pi_U^\dagger$ 等于 Π_U 本身，因为投影已经满足所有四个伪逆条件。实际上，投影是它自身的伪逆，正是因为它已经是幂等的 ($\Pi_U^2 = \Pi_U$) 且自伴的 ($\Pi_U^* = \Pi_U$)。

◇

Ex充足 6.14（满列秩矩阵）。考虑 m 矩阵

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

其具有满列秩。其伪逆由下式给出：

$$A^\dagger = (A^TA)^{-1}A^T = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$$

这个伪逆正确地平衡了A的第一行和第三行中的冗余信息。

◇

定理 6.15（伪逆性质）。*The pseudoinverse $T^\dagger$ satisfies:*

1. *$T^\dagger$ maps $\operatorname{im} T$ isomorphically to $(\ker T)^\perp$*
2. *$T^\dagger$ maps $(\operatorname{im} T)^\perp$ to $\ker T$*
3. *$TT^\dagger T$ is the orthogonal projection onto $(\ker T)^\perp$*
4. *$T^\dagger TT^\dagger$ is the orthogonal projection onto $\operatorname{im} T$*
5. *If T is invertible, then $T^\dagger = T^{-1}$*
6. *$(T^\dagger)^\dagger = T$*
7. *$(T^*)^\dagger = (T^\dagger)^*$*

在一般情况下，当 A 既不是满行秩也不是满列秩时，伪逆的构造变得更加复杂。一种方法是使用正交投影到基本子空间，如图 6.1 所示。首先，我们对图像空间 $\operatorname{im}(A)$ 进行投影。

使用投影算子 $P_{\mathrm{im}(A)}$。然后，我们应用 $\mathrm{im}(A)$ 与 $(\ker A)^\perp$ 之间的同构，随后通过包含映射返回到定义域。

6.5 *Least Squares Approximation*

伪逆将抽象分解转化为实用的计算工具。对于无法存在精确解的系统，它在基本子空间中导航，以提供最佳的近似解。这种最优性——在自然误差度量下找到最佳的近似解——是最小二乘法近似的核心。通过基本定理揭示的几何结构不仅提供了理论理解，还为数据拟合和分析提供了实际方法。

考虑系统 $A\boldsymbol{x} = \boldsymbol{b}$，其中 $A : \mathbb{R}^n \to \mathbb{R}^m$ 与 $m > n$。这类系统通常出现在将模型拟合到数据时：每一行代表一个观测值，每一列是一个待确定的参数。尽管 $\boldsymbol{b}$ 很少位于 $\mathrm{im}\, A$ 中，射影空间 $\mathbb{R}^m$ 的正交分解引导我们找到最优的近似：

$$\mathbb{R}^m = \mathrm{im}\, A \boxplus (\mathrm{im}\, A)^\perp$$

因此，向量$\boldsymbol{b}$可以唯一地分解为$\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2$，其中$\boldsymbol{b}_1 \in \mathrm{im}\, A$，且$\boldsymbol{b}_2 \in (\mathrm{im}\, A)^\perp$。由于$\boldsymbol{b}_1$位于$\mathrm{im}\, A$中，存在$\hat{\boldsymbol{x}}$使得$A\hat{\boldsymbol{x}} = \boldsymbol{b}_1$。该向量$\hat{\boldsymbol{x}}$正是我们所求的最小二乘解，正如下述定理所证实：

定理 6.16（最小二乘解）。*For a full-rank matrix $A \in \mathbb{R}^{m\times n}$ with $m > n$, the system $A\boldsymbol{x} = \boldsymbol{b}$ has unique least squares solution:*

$$\hat{\boldsymbol{x}} = A^\dagger \boldsymbol{b} = (A^T A)^{-1} A^T \boldsymbol{b}$$

This solution minimizes $\|\boldsymbol{b} - A\boldsymbol{x}\|$ over all $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. 在极小值处，误差向量 $\boldsymbol{b} - A\boldsymbol{x}$ 必须与 $\mathrm{im}\, A$ 正交，否则我们可以通过投影来减小其长度。这个正交性条件意味着：

$$\langle \boldsymbol{b} - A\hat{\boldsymbol{x}}, A\boldsymbol{v} \rangle = 0 \quad \text{for all } \boldsymbol{v} \in \mathbb{R}^n$$

因此 $A^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = 0$，从而得到 *normal equations*：

$$A^T A\hat{\boldsymbol{x}} = A^T \boldsymbol{b}$$

当A具有满秩时，$A^T A$为正定，因此可逆，从而得到所述解。该解按……的定义等于$A^\dagger \boldsymbol{b}$

满列秩矩阵的伪逆。伪逆 $A^{\dagger} = (A^T A)^{-1} A^T$ 将我们的抽象分解转化为具体的计算方法，用于寻找最优近似值。

为验证这能将误差最小化，注意对于任意 $\boldsymbol{x}$：

$$\|\boldsymbol{b} - A\boldsymbol{x}\|^2 = \|\boldsymbol{b} - A\hat{\boldsymbol{x}} + A\hat{\boldsymbol{x}} - A\boldsymbol{x}\|^2 = \|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2 + \|A\hat{\boldsymbol{x}} - A\boldsymbol{x}\|^2$$

由于按构造 $(\boldsymbol{b} - A\hat{\boldsymbol{x}}) \perp \operatorname{im} A$。第二项恰在 $\boldsymbol{x} = \hat{\boldsymbol{x}}$ 时消失，从而确认了最优性。 □

Nota bene: 术语“正规方程”体现了几何上的法向性——误差向量与解空间垂直。

正规方程自然地源自将 $\boldsymbol{b}$ 投影到 $\operatorname{im} A$ 上。事实上，矩阵乘积 $A(A^T A)^{-1} A^T$ 正是实现了这一正交投影。通过伪逆，我们将线性代数基本定理的几何结构直接与数据拟合的计算方法联系起来。这种联系把抽象的子空间转化为用于近似的实用工具。

例 6.17（线性回归）。考虑将直线 $y = mx + b$ 拟合到点 $(x_1, y_1), \ldots, (x_n, y_n)$。这将导致一个超定系统：

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

最小二乘解最小化点到直线的竖直距离平方之和——这一准则自然地源自欧几里得内积结构。

对于特定值 $(x_1, y_1) = (1, 2)$，$(x_2, y_2) = (2, 3)$，以及 $(x_3, y_3) = (3, 5)$，我们形成：

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}$$

计算 $A^T A$ 和 $A^T \boldsymbol{b}$：

$$A^T A = \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \quad \text{and} \quad A^T \boldsymbol{b} = \begin{pmatrix} 23 \\ 10 \end{pmatrix}$$

常规方程得出：

$$\begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 23 \\ 10 \end{pmatrix}$$

通过求解，我们得到 $m = 1.5$ 和 $b = 0.5$，从而得到 $y = 1.5x + 0.5$ 作为我们的最佳拟合直线。 ◇

最小二乘法自然地扩展到简单曲线拟合之外。当A的列表示基函数时，我们得到一般线性模型：

$$f(x) = c_1\phi_1(x) + c_2\phi_2(x) + \cdots + c_n\phi_n(x)$$

系数c_i源自最小二乘解，在所选基下提供最优近似。基函数ϕ_i的不同选择会产生不同的近似方案：

- 平滑函数的多项式
- 用于周期数据的三角函数
- 用于局部特征的小波
- 用于分段光滑近似的样条函数

例6.18（多项式拟合）。为将二次多项式$f(x) = ax^2 + bx + c$拟合到数据点(x_i, y_i)，我们构造范德蒙德矩阵：

$$A = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}$$

最小二乘解$\boldsymbol{c} = (a、b、c)^T$使$\sum_{i=1}^{n}(y_i - f(x_i))^2$达到最小。通过伪逆$A^\dagger$，该解在基本子空间中进行导航，从而提供最优近似。

◇

这种几何理解将最小二乘从一种计算方法转变为一种理论原则。当精确解不存在时，对可用解空间进行正交投影，在自然的误差度量下提供了尽可能最优的近似。通过伪逆，基本定理中的抽象子空间——像 $\operatorname{im} A$及其正交补 $(\operatorname{im} A)^\perp$——被直接转化为最优的近似方法。

这种方法的全部威力通过与四个基本子空间的显式联系而显现出来。最小二乘解位于 $(\ker A)^\perp$ 中，使其成为唯一的最小范数解。残差$\boldsymbol{b} - A\hat{\boldsymbol{x}}$ 位于 $(\operatorname{im} A)^\perp$ 中，从而在几何上可解释为模型空间中无法表示的 $\boldsymbol{b}$ 的分量。这种正交分解：

$$\boldsymbol{b} = A\hat{\boldsymbol{x}} + (\boldsymbol{b} - A\hat{\boldsymbol{x}})$$

精确地表达了由基本定理所保证的投影到基本子空间上。伪逆 $A^\dagger$ 进行变换

这种将抽象分解为用于求解最优近似的具体计算方法。

Foreshadowing: 第 10 章中的奇异值分解将为分析和求解最小二乘问题提供一个更为强大的框架，揭示近似解的完整几何结构。

6.6 *Regularized Least Squares*

真实数据中存在噪声。尽管最小二乘逼近的优雅框架在数学上是完备的，但当面对测量误差和不确定性时，可能会表现出脆弱性。设想将一个多项式拟合到含噪声的样本上——增加多项式的阶数可以改善对数据点的拟合，但可能会在数据点之间产生剧烈的振荡。测量拟合度与解的平滑性之间的这种张力暗示了我们几何框架的一种修正，即在保留正交投影本质特征的同时，抑制这种不稳定性。

问题的根源在于我们对最小误差的无约束追求。在存在噪声的测量条件下，最小二乘解可能通过扭曲自身以拟合噪声而非底层信号，从而获得看似很小的残差。我们需要某种方式来偏好更简单、更稳定的解——这种偏好可以通过几何来编码。

例6.19（多项式过拟合）。考虑在[0, 1]上，对带有小随机误差的$f(x) = \cos(2\pi x)$ 的样本，拟合次数不断增加的多项式。随着次数的增加，最小二乘解以越来越高的精度跟踪数据点，但代价是在样本之间出现剧烈振荡。尽管每次拟合都最小化了平方误差，但高次数的解看起来愈发不稳定。

Foreshadowing: 这种在拟合数据与保持简洁性之间的平衡贯穿于数学和工程领域。在后续章节学习数据分析时，我们还会再次遇到它。

◇

岭回归通过对我们的最小二乘框架进行一个简单的修改，提供了一种优雅的解决方案。与仅最小化误差 $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$ 不同，我们添加了一个对较大系数进行惩罚的项：

$$\min_{\boldsymbol{x}} \left(\|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda \|\boldsymbol{x}\|^2 \right)$$

其中 $\lambda > 0$ 控制正则化的强度。这个增广的目标函数保留了我们先前推导的几何特性——它不仅衡量与数据的距离，也衡量解空间中与原点的距离。

从理论角度来看，岭回归用正则化版本（$A^TA + \lambda I$）$^{-1}A^T$替代了伪逆$A^\dagger$，以牺牲精确最优性来换取更好的稳定性和条件性。这种修改后的

伪逆在最小范数解与正则化约束之间取得平衡，在仍然尊重底层正交结构的同时，有效地将系数向零收缩。

几何解释颇具启发性。纯最小二乘将 $\boldsymbol{b}$ 投影到 A 的列空间上。岭回归则改变了这一投影，通过施加额外的正交性约束将解拉向原点。参数 λ 控制这种偏移：取值越大，越偏向较小的系数，但代价是更大的残差。

引理 6.20 （岭解）。*The minimizer of the ridge regression objective satisfies the modified normal equations:*

$$(A^TA + \lambda I)\boldsymbol{x} = A^T\boldsymbol{b}$$

This solution has smaller coefficients than the pure least squares solution but generally larger residual error.

Proof. 目标函数 $f(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda\|\boldsymbol{x}\|^2$ 在其导数为零的地方最小化。利用内积的性质展开：

$$\begin{aligned} f(\boldsymbol{x}) &= \langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle + \lambda\langle \boldsymbol{x}, \boldsymbol{x}\rangle \\ &= \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle - 2\langle A\boldsymbol{x}, \boldsymbol{b}\rangle + \|\boldsymbol{b}\|^2 + \lambda\|\boldsymbol{x}\|^2 \end{aligned}$$

将关于$\boldsymbol{x}$的导数设为零：

$$2A^TA\boldsymbol{x} - 2A^T\boldsymbol{b} + 2\lambda\boldsymbol{x} = \boldsymbol{0}$$

重新排列得到修改后的正规方程：

$$(A^TA + \lambda I)\boldsymbol{x} = A^T\boldsymbol{b}$$

源文本： 这个正则化系统始终具有唯一解，即使在 A^TA 是奇异的情况下也是如此，因为添加的项 λI 确保了正定性。该解实现了一种有偏伪逆的形式：　　译文：

$$\boldsymbol{x} = (A^TA + \lambda I)^{-1}A^T\boldsymbol{b}$$

它通过以减少参数估计的方差为代价来交换无偏性。参数 λ 控制这种偏差-方差权衡，较大的值会产生较小的系数，但可能导致较大的残差误差。

□

例 6.21（信号平滑）。考虑通过局部多项式拟合对含噪时间序列进行平滑。纯最小二乘会产生过度跟踪噪声的拟合。岭回归通过惩罚较大的系数，得到更为平滑的近似，在抑制高频噪声的同时更好地捕捉潜在趋势。参数 λ 为这种平滑效果提供了直接的控制。

◇

正则化参数 λ 的选择体现了在拟合数据与保持稳定性之间的基本权衡。较小的取值会产生接近纯最小二乘的解；较大的取值则会迫使解趋向于零。不存在通用的选择——合适的平衡取决于噪声水平、问题结构以及最终目的。

Example: 在多项式拟合中，较大的 λ 值会越来越多地抑制高阶项，从而无论形式阶数如何，都能有效限制拟合函数的复杂度。

这种对最小二乘法的修改——通过附加几何约束进行的正则化——体现了计算数学中的一个更为广泛的原则。当理论的优雅遭遇实践的复杂性时，我们往往不是通过放弃既有框架而取得成功，而是通过对其进行审慎的扩展。指导我们发展最小二乘法的几何直觉被证明足够稳健，既能容纳这些实际关切，又能保持其本质特征。

Signal Processing: From Data to Information

从原始测量到有意义信息的转化定义了现代信号处理。每一种传感器——无论是测量温度、压力、加速度，还是电磁波——都会提供数据，而这些数据的自然表示位于对测量噪声取模的商空间中。本章构建的数学框架，尤其是正交投影、伪逆计算及其正则化变体，提供了通过对空间进行系统性分解从噪声中提取信号的工具。

考虑一个以固定时间间隔采样的温度传感器。我们将其测量表示为向量 $\boldsymbol{y} \in \mathbb{R}^n$，其分量将真实温度与随机波动相结合：

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{n}$$

其中 $\boldsymbol{s}$ 表示潜在信号，$\boldsymbol{n}$ 表示噪声。我们的任务是从 $\boldsymbol{y}$ 中恢复 $\boldsymbol{s}$——这一挑战自然引出了第 6.4 节中研究的伪逆运算。

关键的洞见在于，真实的温度变化通常比随机测量噪声发生得更慢。我们可以通过假设 $\boldsymbol{s}$ 位于由低阶多项式张成的子空间附近，从数学上表达这一点。如果我们将基表示为矩阵 A 的列，那么最优恢复变为：

$$\hat{\boldsymbol{s}} = A(A^T A)^{-1} A^T \boldsymbol{y} = A A^{\dagger} \boldsymbol{y}$$

这正是通过伪逆实现的正交投影，在我们的模型子空间中选取最优点。

对于任意时间点窗口 $[t_k - w, t_k + w]$，我们考虑次数至多为 d 的多项式子空间 $\mathcal{P}_d$。基多项式 $\{1, t, t^2, \ldots, t^d\}$ 为分解我们的信号提供了方向，不过为了数值稳定性，我们通常使用正交多项式，如第 5.3 节所述。每一次局部平滑操作都转化为对伪逆的应用，其中岭正则化提供了第 6.6 节所研究的稳定性。

示例 6.22（温度监测）。考虑来自化学反应器的每小时温度测量，表示为 $\boldsymbol{y} \in \mathbb{R}^{24}$ 。原始传感器数据显示出由测量噪声叠加在真实温度趋势上的快速波动。二次多项式空间 $\mathcal{P}_2$ 提供了一个自然的局部近似子空间——其三维结构能够捕捉常数水平、线性趋势和轻微曲率，同时排除高频噪声。

对于我们的设计矩阵 A，其列表示在采样点处评估的基多项式，伪逆 $A^\dagger = (A^TA)^{-1}A^T$ 精确地给出了最优恢复算子。当由于样本间距过近而导致问题变得病态时，正则化伪逆 $(A^TA + \lambda I)^{-1}A^T$ 稳定了我们的解，同时保持近最优性。正则化参数 λ 控制拟合精度和系数稳定性之间的平衡，正如在第6.6节中分析的那样。$\diamond$

商空间视角特别具有启发性。两个仅在高频噪声上有所不同的温度信号在适当的商空间中属于同一等价类。我们开发的伪逆算子提供了一种系统化的方法，从这些等价类中选择规范代表——这些代表最小化了逼近误差和系数幅度。

示例 6.23（心电图处理）。表示为 $\boldsymbol{y} \in \mathbb{R}^n$ 的 ECG 信号包含高频噪声和必须保留的尖锐特征（QRS 复合波）。简单的伪逆投影要么会保留过多噪声，要么会模糊重要的峰值。解决方案来自自适应正则化——通过根据局部信号特性变化正则化参数 λ ，我们可以调整伪逆以尊重局部结构。

具体而言，在过渡急剧的区域，我们减少正则化以保留细节，而在较平坦的区域，我们增加正则化以抑制噪声。这种自适应方法展示了第6.4节中的伪逆框架如何指导实际算法设计，超越简单的最小二乘法。$\diamond$

该框架通过我们对矩阵空间处理中的张量积构造，自然地扩展到多维信号。考虑一个压力传感器阵列，用于监测飞机机翼的结构载荷。它们的读数形成一个矩阵 $Y \in \mathbb{R}^{m\times n}$，但我们期望真实的压力场在机翼表面平滑变化。二维多项式拟合变成了一个伪逆问题，具有结构化设计矩阵，正则化确保了稳定性，正如一维情况中的处理方式。

基本原理依然不变：有意义的信息通常存在于比原始测量更低维的子空间中。通过仔细选择这些子空间，并使用本章中开发的伪逆和正则化投影，我们以数学上有原则的方式将信号与噪声分离。我们所构建的几何直觉——正交性、伪逆和正则化投影——为现代工程中的这一基本任务提供了基础。

Image Processing: Pixels & Polynomials

数字图像为应用本章中开发的正交分解和伪逆方法提供了一个自然的领域。每个灰度图像由一个矩阵 $F \in \mathbb{R}^{m\times n}$ 表示，其条目将有意义的内容与测量噪声结合起来。我们所建立的数学框架——特别是第6.4节的伪逆运算和第6.6节的正则化最小二乘法——提供了通过最优近似进行图像增强的系统工具。

考虑图像平滑的基本问题。关键的见解与我们对伪逆的研究相似：大多数自然图像在局部上可以通过低维子空间的元素进行良好近似。正如第6.4节所展示的，伪逆如何提供对兼容子空间的最优投影，我们可以通过精心应用局部伪逆操作来增强图像。

例6.24（局部表面拟合）。在每个像素位置 (i, j)，强度的一个邻域定义了一个向量，我们使用合适设计矩阵的伪逆对其进行拟合。基多项式

$$\{\boldsymbol{b}_{k\ell}(x, y) = (x - i)^k (y - j)^\ell : k + \ell \leq d\}$$

构成我们设计矩阵 A 的各列。计算 $A^\dagger = (A^T A)^{-1} A^T$ 可为该多项式空间提供最优拟合算子。与第 6.6 节相同，正则化的伪逆 $(A^T A + \lambda I)^{-1} A^T$ 可防止在多项式模型受限的边缘附近出现振荡。

◇

几何解释尤为发人深省。每一次局部拟合都实现了第6.4节中发展的伪逆算子，将图像数据的块映射到精心选择的多项式子空间上。正则化参数 λ 塑造了这一过程，将解拉向更简单的多项式系数，正如第6.6节中分析的岭回归稳定了一般最小二乘问题一样。

例6.25（医学成像）。考虑用于医学诊断的X射线图像，将其表示为包含来自光子计数统计的量子噪声的矩阵 $F \in \mathbb{R}^{m\times n}$。诊断特征的空间在该噪声结构下形成一个自然的商空间。通过适当的正则化以保留临床相关细节的伪逆计算，提供了一种从这些等价类中选择规范代表的系统化方法。第6.4节的框架将参数选择从经验艺术转变为有原则的数学。◇

在第6.4节中介绍的伪逆视角为图像修复（即重建缺失或损坏的像素）提供了特别的洞察。当一些像素缺失时，设计矩阵A仅包含对应于观察到的像素的行。伪逆 $A^\dagger$提供了最优的重建，它既拟合了观察到的数据，又最小化了系数复杂性。由于伪逆最小化了残差和解的范数，它自然地平衡了对周围数据的忠实度和解的简洁性。

彩色图像通过其多个通道引入了额外的结构。一幅 RGB 图像由三个矩阵 F_R、F_G、$F_B \in \mathbb{R}^{m\times n}$ 组成，它们可以从联合分析中获益。本章中提出的伪逆方法可以自然地推广：

- 通过适当的块结构设计矩阵实现通道耦合
- 保持颜色一致性的联合正则化
- 考虑通道间关系的伪逆计算

例 6.26（卫星成像）。卫星影像为多通道数据提供了一个引人注目的例子，其中每个光谱波段都会产生一个矩阵 $F_k \in \mathbb{R}^{m\times n}$。有意义的信息往往位于比原始测量更低维的子空间中。一个结构合理的伪逆计算，落实第 6.4 节中发展的框架，在尊重光谱波段之间耦合关系的同时，将关键特征与传感器噪声分离。

这种多通道正则化体现了对基本伪逆运算的自然扩展，用于处理不同光谱测量之间的结构化冗余。正如伪逆为欠定系统找到最小范数解一样，这些多光谱方法寻找跨通道变化最小的解。

◇

Support Vector Machines: Geometry of Optimal Separation

正交分解、伪逆计算以及最优逼近的原理揭示了工程中的一个根本性挑战：我们如何才能将数据最佳地分离为不同的类别？支持向量机 [SVMs] 通过分离超平面的几何视角来处理这一问题，将分类转化为一个与本章所发展的伪逆算子和优化方法紧密相关的问题。

Think: 类似于第6.4节中的伪逆运算，最优分离超平面通过最小化拟合误差与解复杂性的组合来实现。这些目标之间的权衡与本章贯穿讨论的正则化相呼应。

考虑$\mathbb{R}^d$中的点$\{\boldsymbol{x}_i\}_{i=1}^n$，它们被标记为正类（$y_i = +1$）或负类（$y_i = -1$）。正如我们曾寻求用于近似数据的最优子空间一样，现在我们寻求用于分离各类的最优超平面。这种几何视角将分类从抽象的模式匹配转变为对距离与投影的具体优化。

Example: 在质量控制中，必须将产品的传感器测量结果分类为合格或有缺陷。SVM 的最大间隔原理提供了对测量噪声的鲁棒性，类似于伪逆在传感应用中所提供的稳定性。

一个分离超平面的形式为 $\{\boldsymbol{x} : \boldsymbol{w}^T\boldsymbol{x} + b = 0\}$，其中 $\boldsymbol{w}$ 提供法向方向。这个几何对象通过 $\boldsymbol{w}^T\boldsymbol{x} + b$ 的符号自然地将空间分解为正、负两个半空间。任意点 $\boldsymbol{x}$ 到该超平面的距离通过正交投影得到：

$$d(\boldsymbol{x}) = \frac{|\boldsymbol{w}^T\boldsymbol{x} + b|}{\|\boldsymbol{w}\|}$$

SVM 的几何洞见在于最大化 *margin*——超平面与任一训练点之间的最小距离。在对 $\boldsymbol{w}$ 和 b 进行重新缩放以确保该最小距离等于 $1/\|\boldsymbol{w}\|$ 之后，我们在所有点都被正确分类的约束下寻求最小化 $\|\boldsymbol{w}\|^2$。由此得到的优化反映了在伪逆计算中遇到的权衡：我们在大间隔带来的稳健性（类似于较小的解范数）与拟合训练数据的约束（类似于残差最小化）之间进行平衡。

当数据无法被任何单一超平面分离时，我们引入松弛变量 ξ_i 来衡量误分类的程度。优化问题变为：

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n}\xi_i \quad \text{subject to} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i$$

参数 C 控制着我们在间隔大小与分类错误之间的权衡——这与第 6.4 节中伪逆计算里正则化在拟合与稳定性之间取得平衡的方式完全对应。

考虑这一框架如何指导医学图像分类。每一次扫描都成为高维空间中的一个点，其中坐标表示从图像中提取的强度、纹理和形状。SVM 通过类似于伪逆计算的原理来寻找最优的分离超平面。恰好满足间隔约束的那些点——支持向量——标识出在诊断上最具挑战性的病例。它们的特殊作用自然而然地源于几何结构：这些边界样本完全决定了最优超平面，而其他点即使在各自的半空间内移动，也不会影响解。

Nota bene: 决定间隔的支持向量体现了一种更深层的稀疏性原理，类似于伪逆如何给出最小范数解。两种方法都识别出决定最优行为的关键组成部分。

这种几何视角改变了我们对结构健康监测的 approach。桥梁和建筑物上的振动传感器生成加速度数据流，需要迅速分类为正常或潜在危险。 从这些信号提取的特征成为高维空间中的坐标，SVM 在其中构造其分隔超平面。 边距提供了对传感器噪声和环境变化的关键容忍度 —— 我们需要在测量的不确定性下依然可靠的决策。 接近边距的点识别出需要特别关注的边界结构状态，而松弛变量则允许偶尔的约束违反，而不会影响整体的稳健性。

解答体现了基本定理在 6.4 节中通过伪逆展开的空间分解。一旦找到最优超平面，其法向量 $\boldsymbol{w}$ 会自然地将 $\mathbb{R}^d$ 分解为一条直线（与 $\boldsymbol{w}$ 平行）和一个超平面（与决策边界平行的方向空间）的直和。这种几何分解不仅指导我们的理解，也指导我们的计算：到超平面的距离衡量预测的置信度，而间隔则量化了对输入扰动的鲁棒性。

□ ——————————————————————— □

Exercises: Chapter 6

1. 对于 $\mathbb{R}^3$ 中的向量 $\boldsymbol{v}_1 = (1, 1, 1)^T$ 和 $\boldsymbol{v}_2 = (1, 2, -1)^T$： (a) 求一个与 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ (都正交的向量 $\boldsymbol{w}$ b) 验证 $\{\boldsymbol{v}_1、\boldsymbol{v}_2、\boldsymbol{w}\}$ 构成 $\mathbb{R}^3$ (的一组基 c) 使用正交投影将 (2, 3 , 1)^T 表示为这些基向量的线性组合 2. 设 U 为 $\mathbb{R}^3$ 的一个子空间，由 $\boldsymbol{u}_1 = (1, 1, 0)^T$ 和 $\boldsymbol{u}_2 = (0, 1, 1)^T$ 张成。 (a) 求 $U^\perp$ (的一组基 b) 计算 $\boldsymbol{v} = (2, 1, 3)^T$ 到 U (上的正交投影 c) 验证 $\boldsymbol{v} - \Pi_U\boldsymbol{v}$ 位于 $U^\perp$ 中 3. 在具有内积 $\langle f$ 的向量空间 $\mathcal{P}_2$ 中，$g\rangle = \int_0^1 f(x)g(x)\,dx$： (a) 证明

证明 1 和 $x - \frac{1}{2}$ 是正交多项式。（b）使用 Gram-Schmidt 过程将它们扩展为 $\mathcal{P}_2$ (c) 的一个正交基。求 x^2 在 span$\{1, x - 1$ 上的正交投影。

$\bar{2}\}$

4. 设 $\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$。求：（a）$A$ (b) 的列空间的一组基 A (c) 的零空间的一组基-通过计算适当的内积证明这些空间是正交补

5. 考虑将直线 $y = mx + b$ 拟合到点 (0, 1)、(1, 3) 和 (2, 2)。(a) 建立该最小二乘问题的正规方程 (b) 解出 m 和 b (c) 求残差向量，并验证它与系数矩阵的列空间正交

6. 考虑数据点 (1, 0)、(2, 2)、(3, 1)、(4, 4)。求：(a) 最小二乘直线 $y = mx + b$ (b) 数据向量在系数矩阵列空间上的正交投影 (c) 残差向量，并验证它同时与 1 和 x 正交

7. 设 P 为内积空间 V 的子空间 U 上的正交投影。证明：(a) 对所有 $\boldsymbol{v} \in V$，$P^2 = P$ (b) $P^* = P$ (c) $\|P\boldsymbol{v}\| \leq \|\boldsymbol{v}\|$。

8. 证明如果 U 和 W 是内积空间 V 的正交子空间，则：

$$\|\boldsymbol{u} + \boldsymbol{w}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{w}\|^2$$

对于任意的 $\boldsymbol{u} \in U$ 和 $\boldsymbol{w} \in W$。利用这一点来解释为什么勾股定理对正交投影成立。

9. 设 V 为一个有限维内积空间，$U < V$ 为其一个子空间。证明：(a) $(U^{\perp})^{\perp} = U$ (b) $\dim U + \dim U^{\perp} = \dim V$ (c) V 中的每个向量都可以唯一地表示为来自 U 和 $U^{\perp}$ 的向量之和

10. 对于向量 $\boldsymbol{x}$、$\boldsymbol{y} \in \mathbb{R}^n$ 和 $\alpha > 0$，考虑正则化最小二乘问题：

$$\min_{v \in \mathbb{R}^n} \|\boldsymbol{x} - \boldsymbol{v}\|^2 + \alpha \|\boldsymbol{v} - \boldsymbol{y}\|^2$$

(a) 证明这有唯一解 (b) 用 $\boldsymbol{x}$、$\boldsymbol{y}$ 和 α (给出该解的显式公式 c) 将该解解释为在 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的插值

11. 考虑正则化最小二乘问题：

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \lambda \|\boldsymbol{x}\|^2$$

证明当 $\lambda \to \infty$ 时，解趋于 0；而当 $\lambda \to 0^+$ 时，解趋于原始最小二乘问题的最小范数解。

12. 对于内积空间之间的线性变换 $T : V \to W$，证明 $\ker T$ 与 $\operatorname{im} T^*$ 在 V 中互为正交补。这对 $T\boldsymbol{x} = \boldsymbol{b}$ 与 $T^*T\boldsymbol{x} = T^*\boldsymbol{b}$ 的解之间的关系说明了什么？

13. 设 U 是内积空间 V 的一个子空间。证明 $\boldsymbol{v} \in V$ 具有 mini- m
在其陪集 $\boldsymbol{v} + U$ 中所有向量之间的范数当且仅当 $\boldsymbol{v} \perp$ U.

14. 对于矩阵 A，当 A 具有满列秩时，证明到 col(A) 的正交投影由 $A(A^TA)^{-1}A^T$ 给出。利用这一点解释为什么最小二乘解最小化残差范数。

15. 设 V 为区间 [0, 1] 上的连续函数空间，内积为 $\langle f,\ g\rangle = \int_0^1 f(x)g(x)\,dx$。证明：(a) 函数满足 $f(0) = f(1)$ 的子空间 U 在加法和数乘下是封闭的 (b) 找到一个属于 $U^\perp$ (c) 的非零函数；从几何上描述一个函数对 U 中所有函数都正交意味着什么

16. 考虑由建筑物 HVAC 系统的每小时温度读数 T 和能源使用 E (（以千瓦时 kWh) 为单位）组成的训练数据：

$$(T, E) = \{(68, 42), (72, 45), (75, 48), (71, 44), (69, 43), (74, 47)\}$$

设施管理人员认为，温度读数中的测量误差是能耗读数中的两倍显著。建立并求解一个合适的加权最小二乘问题，以找到最佳的线性关系 $E = aT + b$。

17. 设 A 是一个具有 $m > n$ 的 $m \times n$ 矩阵，并设 $\boldsymbol{b}$ 是一个向量，使得 $A\boldsymbol{x} = \boldsymbol{b}$ 没有解。证明：在所有满足 $A\boldsymbol{x} = \boldsymbol{y}$ 的向量 $\boldsymbol{y}$ 中，向量 $\Pi_{\text{im}\,A}\boldsymbol{b}$ 在欧几里得范数下最接近 $\boldsymbol{b}$。

18. 一家制造商在不同温度 x 下测量材料的拉伸强度 y，获得数据点：

$$(x, y) = \{(20, 45), (30, 42), (40, 38), (50, 35), (60, 30), (70, 28)\}$$

理论表明该关系应为 $y = ae^{bx}$ 的形式。说明如何将其转换为线性最小二乘问题，并求解 a 和 b。

19. 设 P 和 Q 分别是到子空间 U 和 W 的正交投影。证明 PQ 本身是一个正交投影当且仅当 $PQ = QP$。这说明了 U 与 W 之间的关系是什么?

20. 一名工程师测量了受到已知频率 ω 的周期性噪声污染的信号 $s(t)$。在时刻 $t_1, \ldots, t_n$ 的测量值为：

$$y_i = s(t_i) + a\cos(\omega t_i) + b\sin(\omega t_i) + \epsilon_i$$

其中 ϵ_i 表示随机误差。说明如何使用正交投影来估计并去除周期性分量，从而得到对原始信号 $s(t)$ 的更干净的估计。

Chapter 7
Diagonalization & Dynamics

"how is it that all things are chang'd even as in ancient times"

一个持续变化的变换揭示了从静态视角隐藏的模式。一个质量-弹簧系统以特征频率振荡；一个种群按内在速率增长或衰退；一个网络的影响沿着优先通道流动。这些看似不同的现象共享一个共同的数学本质：某些方向在重复变换下保持不变，而沿这些特殊方向的向量则经历纯粹的缩放。这些固定的方向及其相关的缩放因子——*eigenvectors* 和 *eigenvalues*——为理解线性变换随时间变化的方式提供了关键。

考虑一个简单的线性递归模型来描述人口增长：每一代的数量是上一代的一个固定倍数。根据这个倍数是否超过1，人口要么呈指数增长，要么衰退至灭绝。尽管这个例子很基础，但它包含了一个更深刻的真理的种子：线性系统的长期行为通常会简化为沿着特定方向的纯缩放。这些特定方向及其缩放因子不仅决定了人口是繁荣还是灭绝，还决定了机械系统的振动方式、热量的扩散、量子态的演化以及网络如何传递影响。

寻找这样的不变方向将我们引导到 *characteristic polyno- mials* —— 这些方程的根揭示了变换的自然缩放因子。这些 *eigenvalues* 与它们相关联的 *eigenvectors* 提供了对线性变换的新视角。我们寻找与固有缩放对齐的坐标；当这样的坐标存在时，变换呈现出最原始的对角形式。

这个对角化——这种与自然方向对齐的坐标变换——不仅简化了计算。它揭示了变换如何作用的本质，将复杂的运动分解为更简单的

组件。鼓的振动变成纯音的叠加；热流分解成独立的衰减模式。在一个基底下看起来复杂的东西，在对角化的视角下变得简单。

7.1 *The First Order*

最简单的微分方程来自微积分，它作为所有连续时间线性系统的原型。
考虑方程式{v*}

$$\frac{dx}{dt} = \lambda x \tag{7.1}$$

Notation: 使用微分算子 $D = d/dt$ 将在后续证明是有利的。

其中 λ 是一个常数。从微积分中我们知道一般解是指数形式：

$$x(t) = x_0 e^{\lambda t}$$

其中 x_0 是一个常数，可以解释为 *initial condition* $x_0 = x(0)$。这个基本方程包含了更深层次结构的种子，因为 (7.1) 的解在加法和标量乘法下是封闭的。因此，简单解 $e^{\lambda t}$ 作为 (7.1) 的一维解空间的基础。

注意常数 λ 在决定解空间中所有解的定性行为方面的核心作用：

- 当 $\lambda >$ 为 0 时，解呈指数增长
- 当 $\lambda < 0$ 时，解呈指数衰减
- 当 $\lambda =$ 为 0 时，解保持不变

Think: 与 $x_0 = 0$ 相对应的解是特殊的，因为即使其他解增长或收缩，它也保持不变。这样的 *equilibrium* 解是动力系统中研究的核心对象。

这种结构——由特征数 λ 参数化的指数解——将在本章以及接下来的两章中反复出现。更复杂的系统将分解为此类基本解的集合，每一个都以其自身的特征速率增长或衰减。其精髓在于找到这些自然的行为模态。

此处出现的参数 λ 是决定该 ODE 中自然增长或衰减速率的特征值。起初这一点并不明显，但这类 λ 值在最深层次上与矩阵代数和线性变换密切相关。

7.2 *Coupled First-Order Systems*

现实系统很少孤立地演化。捕食者种群依赖其猎物；股价随市场板块波动；神经元在庞大的互联网络中放电。即使是最简单的模型也往往追踪至少两个相互作用的变量。考虑 $x(t)$ 和 $y(t)$，它们的演化取决于——

线性地取决于当前状态的某种组合：

$$\frac{dx}{dt} = ax + by \quad : \quad \frac{dy}{dt} = cx + dy$$

其中 a, b, c, d 为常数。与标量情形不同，这里并不存在显而易见的解。将该线性系统写成矩阵形式会显现出一种熟悉的模式：

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

这种矩阵形式表明有一个值得考察的特殊情况。假设该矩阵是对角的：

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \quad \Leftrightarrow \quad \frac{dx}{dt} = \lambda_1 x \ : \ \frac{dy}{dt} = \lambda_2 y$$

每个都可以通过上一节的方法求解。解将沿着每个坐标轴呈现纯粹的指数增长或衰减：

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \end{pmatrix}$$

其中 c_1 和 c_2 由初始条件决定。

这一观察——即对角系统可以分解为相互独立的标量方程——启示了一种策略。如果我们能以某种方式将原始系统变换为对角形式，其解就会简化为纯指数函数：我们正在寻找能够揭示耦合系统中潜藏的对角结构的坐标。

例 7.1（化学反应）。考虑两种化学物种，其浓度分别为 x_A 和 x_B，它们通过一个以矩阵形式表示的简单反应网络相互作用：

$$\frac{d}{dt}\begin{pmatrix} x_A \\ x_B \end{pmatrix} = \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix}\begin{pmatrix} x_A \\ x_B \end{pmatrix}$$

其中 k_1、$k_2 > 0$ 是反应速率常数。一次耐人寻味的坐标变换揭示了隐藏的简洁性。考虑如下可逆变换

$$\begin{pmatrix} x_A \\ x_B \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ k_2 & -k_1 \end{bmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \quad \Longleftrightarrow \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ k_2 & -k_1 \end{bmatrix}^{-1}\begin{pmatrix} x_A \\ x_B \end{pmatrix}$$

我们可以通过相似性变换来转换微分方程 trans-

形成:

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix}u\\v\end{pmatrix} &= \frac{d}{dt}\left(\begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}^{-1}\begin{pmatrix}x_A\\x_B\end{pmatrix}\right) = \begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}^{-1}\frac{d}{dt}\begin{pmatrix}x_A\\x_B\end{pmatrix}\\ &= \begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}^{-1}\begin{bmatrix}-k_1 & k_2\\ k_1 & -k_2\end{bmatrix}\begin{pmatrix}x_A\\x_B\end{pmatrix}\\ &= \begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}^{-1}\begin{bmatrix}-k_1 & k_2\\ k_1 & -k_2\end{bmatrix}\begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}\begin{pmatrix}u\\v\end{pmatrix} = \begin{bmatrix}0 & 0\\ 0 & -k_1-k_2\end{bmatrix}\begin{pmatrix}u\\v\end{pmatrix}\end{aligned}$$

该对角矩阵在 u 和 $-(k_1+k_2)<$ 上的对角元为 0，在 v 上为 0。因此 u 保持不变，而 v 以指数方式衰减。原始坐标下的解计算为:

$$\begin{pmatrix}x_A\\x_B\end{pmatrix} = \begin{bmatrix}1 & 1\\ k_2 & -k_1\end{bmatrix}\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}c_1 + c_2 e^{-(k_1+k_2)t}\\ c_1k_2 - c_2k_1e^{-(k_1+k_2)t}\end{pmatrix}$$

其中 c_1 和 c_2 取决于初始条件。这种“神奇”的坐标变换预示着更深层的结构——对角线上的零暗示着存在某种守恒原理在起作用（此处为质量守恒）。 ◇

对这种对角表示的探索将指导我们在后续各节中的发展。我们将发现，许多耦合系统可以变换为对角形式，在这种形式下其行为一目了然。而那些抗拒这种对角化的系统则需要更为细致的分析，从而引向第8章的若尔当标准形。贯穿始终，我们的目标仍然是通过寻找能够揭示其本质结构的坐标，来理解线性系统如何演化。

7.3 *Eigenvalues & Eigenvectors*

常微分方程的一般线性系统具有如下形式

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x} \tag{7.2}$$

其中 A 是一个方阵，$\boldsymbol{x}(t)$ 是时间的向量值函数。这类方程在工程领域中无处不在——从机械振动到化学动力学再到种群动力学。它们的解虽然并非立刻显而易见，却是理解线性系统如何演化的关键。

我们在标量方程方面的经验提示了一种策略。若将A替换为一个标量λ，解将呈现为简单的指数形式。

$\boldsymbol{x}(t) = e^{\lambda t}\boldsymbol{x}_0$. 这一观察引出了一个关键问题：是否存在一个 1 维子空间，使得矩阵 A 在其上作用得就像标量乘法一样？这等价于找到一个向量 $\boldsymbol{v} \neq 0$，使得

$$A\boldsymbol{v} = \lambda\boldsymbol{v} \tag{7.3}$$

对于某个标量 λ。这样的特殊子空间 span($\boldsymbol{v}$)，如果存在的话，在 A 下将是 *invariant*，并且对于该子空间中的向量，求解该 ODE 将化简为平凡的一维情形。

Nota bene: 限制 $\boldsymbol{v} \neq 0$ 至关重要。零向量对于任何 λ 都满足 $A0 = \lambda 0$，但并不能告诉我们关于该变换行为的任何信息。

将 A 视为线性变换可以阐明前进的路径。方程（7.3）可以改写为

$$(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$$

其中 I 表示单位矩阵。要使该方程存在非零解 $\boldsymbol{v}$，变换($A - \lambda I$)必须具有非平凡的核。

这恰好发生在（$A - \lambda I$）不可逆时——当它的行列式为零：det($A - \lambda I$) = 0。这是以下基本定义的关键。

定义 7.2（特征值与特征向量）。方阵 A 的 *characteristic polynomial* $p_A(\lambda)$ 是多项式

$$p_A(\lambda) = \det(A - \lambda I) \tag{7.4}$$

一个标量 λ 被称为 *eigenvalue* 的 A，如果它是特征多项式的根：$p_A(\lambda) = 0$。对于每个特征值 λ，任何满足的非零向量 $\boldsymbol{v}$

$$A\boldsymbol{v} = \lambda\boldsymbol{v} \tag{7.5}$$

称为与特征值λ对应的*eigenvector*。 •

Nota bene: 特征值和特征向量是成对的——尽管并不意味着唯一性。

例 7.3。作为一个具体的例子，考虑矩阵

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

其特征多项式为

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3$$

将此设置为零得到特征值 $\lambda_1 = 3$ 和 $\lambda_2 = 1$。对于 $\lambda_1 = 3$，我们解 ($A - 3I$)$\boldsymbol{v} = 0$:

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

同样地，对于 $\lambda_2 = 1$ 我们得到 $\boldsymbol{v}_2 = (1, -1)^T$。每个特征向量揭示了 A 通过简单缩放起作用的一个方向：沿着 $\boldsymbol{v}_1$ 的向量被拉伸 3 倍，而沿着 $\boldsymbol{v}_2$ 的向量保持不变（按 1 进行重缩放）。 ◇

搜索特征值和特征向量因此简化 It seems like your message got cut off. Could you pleas

1. 形成特征多项式 $\det(A - \lambda I)$ 2. 找到它的根（特征值） 3. 对于每个特征值 λ，求解 $(A - \lambda I)\boldsymbol{v} = 0$ 以得到非零 $\boldsymbol{v}$

关于一个系统具有哪种类型以及多少个特征值，以下内容至关重要。

引理 7.4（特征多项式）。*For any matrix $A \in \mathbb{R}^{n\times n}$, its characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$ satisfies:*

1. *The polynomial has degree exactly n*
2. *It has exactly n complex roots (counted with algebraic multiplicity)*
3. *Its coefficients are real, and its complex roots occur in conjugate pairs*

Proof. 对于（1），使用任意的余子式展开来展开 $\det(A - \lambda I)$。项 $(-\lambda)^n$ 是从 $-\lambda I$ 的对角元素的乘积中唯一出现的。行列式展开中的所有其他项包含更少的 λ 因子，因为它们必须至少包含一个来自 A 的元素。因此，$p_A(\lambda)$ 的次数恰好为 n，并且首项系数为 $(-1)^n$。

Example: 矩阵 $J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ 的特征多项式是 $p_J(\lambda) = \lambda^2 + 1$，其根为 $\pm i$，这说明复特征值确实必须以共轭对的形式出现。矩阵为 J 的线性变换做什么?

对于（2），代数学基本定理保证，任何次数为 n 且系数为复数的多项式，恰好有 n 个复根（按重数计）。

对于 (3)，$A - \lambda I$ 的各个元素要么是实常数，要么是关于 λ 的实线性项。此类矩阵的行列式必然得到一个具有实系数的多项式。对于这样的多项式，若 $\alpha + i\beta$ 是一个根，则其复共轭 $\alpha - i\beta$ 也必然是一个根，且具有相同的重数。这是因为实系数多项式的复根总是成共轭对出现：若 $p(a + bi) = 0$，则 $p(a + bi) = p(a - bi) = 0$。

□

特征多项式所蕴含的不仅仅是特征值——其系数揭示了该变换的基本不变量。最引人注目的是特征值与两种基本矩阵量之间的关系：迹和行列式。

引理 7.5（特征值关系）。*For an $n \times n$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$:*

$\lambda^2 -$

1. *The determinant equals their product:* $\det(A) = \prod_{i=1}^{n} \lambda_i$
2. *The trace equals their sum:* $tr(A) = \sum_{i=1}^{n} \lambda_i$

Example: 对于一个 2×2 矩阵 A，特征多项式 $p_A(\lambda) = \mathrm{tr}(A)\lambda + \det(A)$ 使这些关系变得透明。

7.4 *Simple Diagonalization*

当一个线性变换拥有n个互不相同的实特征值时，一种显著的简化就成为可能。特征向量——那些只经历纯粹缩放的特殊方向——为以最简单的形式观察该变换提供了一个自然的基。这种将坐标变换以与特征向量对齐的过程，称为*diagonalization*，揭示了剥离耦合复杂性后的变换本质。

考虑再次查看示例 7.3 中的矩阵：

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

我们找到了特征值 $\lambda_1 = 3$ 和 $\lambda_2 = 1$，对应的特征向量 $\boldsymbol{v}_1 = (1, 1)^T$ 和 $\boldsymbol{v}_2 = (1, -1)^T$。各个特征值方程

$$A\boldsymbol{v}_1 = 3\boldsymbol{v}_1 \quad \text{and} \quad A\boldsymbol{v}_2 = \boldsymbol{v}_2$$

可以通过将特征向量排列为矩阵的列来优雅地组合：

$$A[\boldsymbol{v}_1\ \boldsymbol{v}_2] = [\boldsymbol{v}_1\ \boldsymbol{v}_2] \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

以平方矩阵 $V = [\boldsymbol{v}_1\ \boldsymbol{v}_2]$ 和 $\Lambda = \text{diag}(3, 1)$ 表示，这变得非常简单

$$AV = V\Lambda \tag{7.6}$$

Think: if you have internalized the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$, you may find the matrix form of this equation to be easily memorable. Just be sure to convince yourself as to the ordering of the terms...

当 V 可逆时（当特征值不同时，它必须是可逆的），我们得到对角化 $V^{-1}AV = \Lambda$。这个观察可以推广到任意维度：

定理 7.6（对角化）。 *Let A be an $n \times n$ matrix with n distinct real eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Then:*

1. *The eigenvectors form a basis for $\mathbb{R}^n$*
2. *The matrix $V = [\boldsymbol{v}1 \cdots \boldsymbol{v}_n]$ is invertible*
3. *$V^{-1}AV = \Lambda$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$*

Proof. 特征向量线性无关这一事实源于它们分别对应不同的特征值。确实，假设 $\sum_{i=1}^{n} c_i\boldsymbol{v}_i = 0$。则

$$\mathbf{0} = A\left(\sum_{i=1}^{n} c_i\boldsymbol{v}_i\right) = \sum_{i=1}^{n} c_i\lambda_i\boldsymbol{v}_i$$

从第二个方程中减去 λ_1 倍的第一个方程

Understood. Please provide the source text you would l

$$\sum_{i=2}^{n} c_i(\lambda_i - \lambda_1)\boldsymbol{v}_i = \mathbf{0}$$

由于$\lambda_i \neq \lambda_1$对于$i \geq 2$，我们必须有$c_2 = \cdots = c_n = 0$，因此$c_1 = 0$。因此$\{v_1, \ldots, v_n\}$是线性无关的，使得V是可逆的。最终结论来自我们上面的推导。 □

对角化的威力在于它能简化矩阵幂的计算。当 $A = V\Lambda V^{-1}$ 时，重复相乘就变成了简单的对角缩放：

引理 7.7（矩阵幂）。*If $A = V\Lambda V^{-1}$ is diagonalizable, then for any positive integer k:*

$$A^k = V\Lambda^k V^{-1}$$

Think: how hard is it to compute powers of the diagonal matrix Λ?

Proof. 该结果直接源于矩阵乘法的结合律以及逆矩阵的性质。 □

这种分解揭示了一个变换在反复应用中如何复合：每个特征子空间都会按其特征值反复缩放，而换基矩阵 V 和 V^{-1} 在我们选定的坐标与这些自然的特征子空间之间进行转换。当 $k \to \infty$ 时，其行为变得清晰——它完全由特征值的大小所控制：其影响见第9章。

通过对角化，相似矩阵之间的关系获得了新的意义。当且仅当矩阵是在不同坐标下表示同一线性变换时，它们才是相似的。当这些坐标与一组特征向量基对齐时，矩阵就呈现为对角形式——这是它可能具有的最简单表示。并非每个矩阵都允许这样的对角形式（这是第8章的主题），但一旦可以，我们便同时获得计算上的优势和理论上的洞见。

7.5 *Matrix Exponentials*

回顾第 7.1 节中的一般线性微分方程组：

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x} \tag{7.7}$$

其中 A 是一个方阵。标量情形 $A = \lambda I$ 给出了形如 $e^{\lambda t}\boldsymbol{x}_0$ 的解。这类指数函数确实在一般情况下生成解。下面的定义是关键。

定义 7.8（矩阵指数）。一个方阵 A 的 *matrix exponential* 定义为如下绝对收敛的幂级数

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{A^k}{k!} \tag{7.8}$$

•

这个形式级数继承了标量指数函数的许多性质，其中包括一个至关重要的事实：它满足我们的微分方程：

引理 7.9（矩阵指数解）。*The general solution to the initial value problem*

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}, \quad \boldsymbol{x}(0) = \boldsymbol{x}_0$$

is given by $\boldsymbol{x}(t) = e^{At}\boldsymbol{x}_0$.

Caveat: 尽管该公式类似于标量情形，但直接从级数计算 e^{At} 很少具有实用性。关键在于基于 A 的结构寻找更高效的方法。

最简单的情况出现在 A 为对角矩阵时。对于 $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$，矩阵指数就是 $e^{At} = \operatorname{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})$。每个分量根据其对角元独立演化。

示例 7.10（对角系统）。该系统

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

具有矩阵指数

$$e^{At} = \begin{bmatrix} e^{2t} & 0 \\ 0 & e^{-t} \end{bmatrix}$$

第一个分量呈指数增长，而第二个分量衰减——这种行为完全由特征值决定。

◇

引理 7.11（可对角化矩阵的指数）。*If* $A = V\Lambda V^{-1}$ *is diagonalizable with* $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$*, then*

$$e^{At} = Ve^{\Lambda t}V^{-1}$$

where $e^{\Lambda t} = \operatorname{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})$.

Proof. 由引理 7.7 可知，对于任意正整数 k，$A^k = V\Lambda^k V^{-1}$。因此，在定义 e^{At} 的幂级数中：

$$e^{At} = I + At + \frac{(At)^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{t^k}{k!}A^k = \sum_{k=0}^{\infty} \frac{t^k}{k!}V\Lambda^k V^{-1}$$

矩阵指数级数的绝对收敛性使我们能够将 V 和 V^{-1} 提取出来：

$$e^{At} = V\left(\sum_{k=0}^{\infty} \frac{t^k}{k!}\Lambda^k\right)V^{-1} = Ve^{\Lambda t}V^{-1}$$

其中中间项通过标量指数级数化简为 $\operatorname{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})$.

□

这种分解揭示了特征值如何控制解的长期行为：每个特征方向都会以由其特征值决定的速率经历指数增长或衰减。

例 7.12（人口增长）。考虑一个由两个相互作用的种群组成的简单模型，其中每个物种的增长率既取决于自身的种群数量，也取决于另一物种的种群数量：

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

根据例 7.3，我们知道该矩阵具有特征值 $\lambda_1 = 3$ 和 $\lambda_2 = 1$，其对应的特征向量为 $\boldsymbol{v}_1 = (1, 1)^T$ 和 $\boldsymbol{v}_2 = (1, -1)^T$。因此

$$e^{At} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} e^{3t} & 0 \\ 0 & e^{t} \end{bmatrix}\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

该解揭示了两种基本模态：对称增长，其中两个群体按比例同时增加（e^{3t} 项）；以及非对称增长，其中它们的差异演化得更为缓慢（e^{t} 项）。任何初始条件都会激发这些模态的某种组合，而由于其更大的特征值，对称模态最终将占据主导。

◇

这种特征值与解的定性行为——增长、衰减或振荡——之间的联系，体现了矩阵的代数结构如何决定其流的几何特性。矩阵指数将我们对特征值的静态理解转化为关于系统如何随时间演化的动态图景。

7.6 Higher-Order Equations

一个简单的质量—弹簧系统需要同时跟踪位置和速度；具有电感和电容的电路需要电流和电荷；化学反应网络可能取决于浓度及其变化率。这类系统自然会导向二阶（或更高阶）微分方程。尽管看起来比迄今研究的一阶系统更为复杂，但通过系统地化简为矩阵形式，这些方程同样可以用相同的特征值方法来处理。

正如第2章结尾所预示的，考虑一个n阶线性齐次微分方程：

$$\frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \cdots + a_1\frac{dx}{dt} + a_0 x = 0$$

其中系数 a_k 为常数。使用微分算子 $D = d/dt$，我们可以将其更紧凑地写为

$$p(D)x = (D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0 I)x = 0$$

其中 $p \in \mathcal{P}_n$ 是关于微分算子 D 的一个次数为 n 的多项式，其系数为 a_i，$i = 0 \ldots n-1$（且最高次系数为 1）。这个 n 阶的单个方程可以通过为各阶导数引入新变量而转化为由 n 个一阶方程组成的系统。令

$$x_1 = x, \quad x_2 = \frac{dx}{dt}, \quad x_3 = \frac{d^2x}{dt^2}, \quad \ldots \quad x_n = \frac{d^{n-1}x}{dt^{n-1}}$$

Then 我们的方程以矩阵 f 的形式变成了一个一阶系统　　ORM:

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

Nota bene: 该矩阵具有一种特殊结构——除超对角线上的 1 以及最后一行中的系数 $-a_k$ 之外，其余元素全为零。这样的矩阵称为 *companion* 。

这种变换揭示了原始微分方程与线性代数之间的深刻联系：伴随矩阵的特征值恰好是 *characteristic equation* 的根。

matrices

$$\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 = p(\lambda) = 0$$

通过在多项式算子中将λ替代D获得。当这些特征值是不同的（如我们在本章中所假设的），解可以直接从我们之前关于矩阵指数的工作中得出。

例子 7.13（质量-弹簧系统）。考虑一个质量为 m 的物体，连接到一个弹簧上，弹簧常数为 k，阻尼系数为 c。牛顿第二定律得出

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0$$

其中 $x(t)$ 衡量了从平衡位置的位移。通过除以 m 并设置 $\omega_0^2 = k/m$ (自然频率) 和 $\gamma = c/m$ (阻尼比)，我们得到

$$\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} + \omega_0^2 x = 0$$

这通过设置 $x_1 = x$ 和 $x_2 = \dot{x}$ 转换为一阶形式:

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

对于具体取值 $\omega_0^2 =$ 为 4 且 $\gamma =$ 为 3，伴随矩阵

$$A = \begin{bmatrix} 0 & 1 \\ -4 & -3 \end{bmatrix}$$

具有特征方程 $\lambda^2 + 3\lambda + 4 = 0$，其根为 $\lambda_1 = -1$ 和 $\lambda_2 = -2$。对应的特征向量为

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

该解源自矩阵指数。由于 A 是可对角化的，且

$$V = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad \Lambda = \operatorname{DIAG}(-1, -2) = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$

我们有

$$e^{At} = Ve^{\Lambda t}V^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

对于任意初始条件 $\boldsymbol{x}(0)$，解为 $\boldsymbol{x}(t) = e^{At}\boldsymbol{x}(0)$。位置 $x(t) = x_1(t)$ 按照 $t \to \infty$ 衰减至平衡，其速率由特征值 -1 和 -2 决定。 $\diamond$

这个例子说明了特征值分析如何揭示物理行为。特征值 -1 和 -2 为实数且为负值，表示纯粹的指数衰减且无振荡——这是过度阻尼系统的特征。不同的参数值可能会导致欠阻尼振荡或临界阻尼，这些情况我们将在第8章中探讨。

7.7 *Basis Solutions*

线性齐次微分方程，无论表示为一阶系统还是高阶标量方程，都具有基于基本模态的优雅解结构。这两种视角——系统的向量值解与高阶方程的标量解——为理解线性演化的本质提供了互补的见解。

引理 7.14（解空间）。 *The solutions to both:*

1. *The first-order system* $\dfrac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}$ *in* $\mathbb{R}^n$
2. *The n-th order equation* $(D^n + a_{n-1}D^{n-1} + \cdots + a_1D + a_0)x = 0$

form vector spaces under their natural operations of addition and scalar multiplication.

Proof. 对于该系统，如果 $\boldsymbol{x}_1(t)$ 和 $\boldsymbol{x}_2(t)$ 求解 $D\boldsymbol{x} = A\boldsymbol{x}$，则：

$$\frac{d}{dt}(\boldsymbol{x}_1 + \boldsymbol{x}_2) = A\boldsymbol{x}_1 + A\boldsymbol{x}_2 = A(\boldsymbol{x}_1 + \boldsymbol{x}_2)$$

对于标量方程，如果 $x_1(t)$ 和 $x_2(t)$ 解 $p(D)x = 0$，其中 $p(D)$ 是多项式微分算子，则：

$$p(D)(x_1 + x_2) = p(D)x_1 + p(D)x_2 = 0$$

类似的推理在两种情况下同样适用于数乘。 □

这些解空间通过第7.6节中的伴随矩阵的特征结构紧密相连。下面的定理揭示了特征值如何在两种情形中生成基解：

定理 7.15（基解）。*Let A be diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:*

1. *For the system $D\boldsymbol{x} = A\boldsymbol{x}$ with eigenvectors $\boldsymbol{v}_1, \ldots\ldots, \boldsymbol{v}_n$, a basis for the solution* {v16
$$\boldsymbol{\phi}_i(t) = e^{\lambda_i t}\boldsymbol{v}_i, \quad i = 1, \ldots, n$$
2. *For the scalar equation $P(D)x = 0$, where $P(\lambda)$ is the characteristic polynomial, a basis is:*
$$\phi_i(t) = e^{\lambda_i t}, \quad i = 1, \ldots, n$$

Proof. 对于该系统，直接代入可验证每个 $\boldsymbol{\phi}_i$ 都是一个解：

$$\frac{d}{dt}\boldsymbol{\phi}_i = \lambda_i e^{\lambda_i t}\boldsymbol{v}_i = e^{\lambda_i t}(A\boldsymbol{v}_i) = A\boldsymbol{\phi}_i$$

它们的线性无关性源于特征向量的线性无关性。

对于标量方程，注意到 $p(D)e^{\lambda t} = p(\lambda)e^{\lambda t}$。因此，当 λ 是 p 的一个根时，指数函数 $e^{\lambda t}$ 提供了一个解。为建立这些解的线性无关性，假设我们有一个线性组合为零：

$$\sum_{i=1}^{n} c_i e^{\lambda_i t} = 0 \quad \text{for all } t$$

我们将证明所有系数 c_i 都必须为零。对该方程进行 k 次求导得到：

$$\sum_{i=1}^{n} c_i \lambda_i^k e^{\lambda_i t} = 0 \quad \text{for } k = 0, 1, \ldots, n-1$$

这 n 个方程在系数 c_i 中构成一个线性系统。该系统的矩阵是在互不相同的值 λ_i 上的范德蒙德矩阵，其行列式非零。因此 $c_i = 0$ 对所有 i，从而确立指数解的线性无关性。

□

Nota bene: 一个 *Vandermonde* 矩阵的各个元素具有 $v_{ij} = x_i^{j-1}$ 的形式，其中 x_i 是互不相同的数。其行列式当且仅当 x_i 互不相同时不为零。有关范德蒙德矩阵的性质，见习题。

一般解采取平行的形式：

$$\boldsymbol{x}(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t} \boldsymbol{v}_i \quad \text{and} \quad x(t) = \sum_{i=1}^{n} c_i e^{\lambda_i t}$$

系数 c_i 由初始条件决定——对于系统为 $\boldsymbol{x}(0)$，或者对于标量方程为 $x(0)$，$x'(0)$，$\ldots$，$x^{(n-1)}(0)$。

例 7.16（质量-弹簧系统再论）。来自例 7.13 的质量-弹簧方程：

$$\frac{d^2x}{dt^2} + 3\frac{dx}{dt} + 4x = 0$$

具有特征方程 $\lambda^2 + 3\lambda + 4 = 0$，其根为 $\lambda_1 = -1$ 和 $\lambda_2 = -2$。因此：

1. 作为标量方程，其通解为：

$$x(t) = c_1 e^{-t} + c_2 e^{-2t}$$

2. 作为一个具有特征向量 $\boldsymbol{v}_1 = (1, -1)^T$ 和 $\boldsymbol{v}_2 = (1, -2)^T$ 的系统：

$$\begin{pmatrix} x(t) \\ x'(t) \end{pmatrix} = c_1 e^{-t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + c_2 e^{-2t} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

向量解的第一个分量与标量解匹配，正如它必须的那样。

Nota bene: 一般而言，当方程被正确匹配时，标量解 $\phi_i(t) = e^{\lambda_i t}$ 精确对应于向量解 $\boldsymbol{\phi}_i(t)$ 的第一分量。

◇

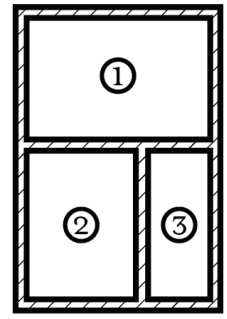
这种并行发展揭示了线性微分方程的本质统一性，无论是作为耦合系统还是高阶标量方程来看。每种视角都有其优点：当我们只关注一个变量时，标量形式通常简化了计算，而系统形式则更好地揭示了几何结构并能推广到耦合方程。

Foreshadowing: 当特征值重合时，两种公式都需要修改。解获得多项式因子，乘以指数函数——这是我们将在第8章探讨的复杂问题。

在标量形式和系统形式之间的选择通常依赖于上下文。物理系统自然耦合多个变量，倾向于使用系统方法。信号处理和控制理论传统上使用标量传递函数，偏好高阶形式。然而，基础的数学结构——由特征值构建的指数解——保持不变，这是我们在应用中将反复利用的统一性。

Multi-Zone Building Temperature Control

多区域建筑的温度动态提供了将特征值分析应用于实际工程问题的优雅方式。考虑一座具有多个房间的建筑，每个房间的温度通过控制系统保持在设定值。



HVAC 输入与与相邻空间的热交换相结合。建筑各处温度随时间的演化揭示了线性系统理论在理解和控制复杂热环境方面的强大作用。

考虑三个相邻的房间，它们共享墙壁，但外部朝向不同。令 $T_i(t)$ 表示在时间 t 时房间 i 的温度。每个房间的温度变化率取决于与相邻房间的热交换（与温度差成正比）、向外界的热损失（与其与环境温度的差成正比），以及 HVAC 输入（受控的供暖或制冷）。

应用牛顿冷却定律和能量守恒可得到一组耦合的微分方程：

$$c_1 \frac{dT_1}{dt} = k_{12}(T_2 - T_1) + k_{13}(T_3 - T_1) - h_1(T_1 - T_a) + u_1$$

$$c_2 \frac{dT_2}{dt} = k_{12}(T_1 - T_2) + k_{23}(T_3 - T_2) - h_2(T_2 - T_a) + u_2$$

$$c_3 \frac{dT_3}{dt} = k_{13}(T_1 - T_3) + k_{23}(T_2 - T_3) - h_3(T_3 - T_a) + u_3$$

其中，c_i 表示房间 i 的热容，k_{ij} 表示房间 i 与 j 之间的热导，h_i 表示与环境温度 T_a 的传热系数，u_i 表示向房间 i 输入的 HVAC 功率。

Nota bene: 这里的建模方法体现了工程学中的一种更普遍的模式：复杂的物理系统通过适当的近似，往往可以化简为耦合的线性微分方程。

为分析该系统，首先考虑相对于环境温度的无输入响应（$u_i = 0$）。令 $x_i = T_i - T_a$表示房间i中的温度偏差。采用现代办公楼一个分区的现实参数：相同的热质量（$c_1 = c_2 = c_3 = 1$）、相邻房间之间的对称耦合（$k_{12} = k_{23} = 1$，$k_{13} = 0$.5），以及不同的外部暴露程度（$h_1 = 2$，$h_2 = 1$，$h_3 = 1.5$），我们得到系统矩阵：

$$A = \begin{bmatrix} -3.5 & 1.0 & 0.5 \\ 1.0 & -2.5 & 1.0 \\ 0.5 & 1.0 & -3.0 \end{bmatrix}$$

A 的特征值决定了建筑的自然热模态。对其进行计算后，揭示出三个不同的实特征值：$\lambda_1 \approx -4.37$、$\lambda_2 \approx -2.83$ 和 $\lambda_3 \approx -1.80$。这些特征值的互异性尤为幸运，因为它使得能够对系统行为进行完整的模态分解，而无需采用第 7.3 节中提出的更为复杂的约当（Jordan）标准形分析。此外，它们的负值保证了渐近稳定性——在任何扰动之后，每个房间的温度最终都会回到环境温度，这与物理直觉一致。

Foreshadowing: 热系统中多个时间尺度的出现预示着更一般的奇异摄动理论，这在许多工程领域中至关重要。

这些特征模态揭示了建筑物的基本热行为。最快的模态，与 λ_1 相关，表示所有房间之间的快速平衡。中等速度的模态显示了房间 2 与其邻近房间之间的温度振荡，而最慢的模态则捕捉到整个系统的逐渐降温。

使用 t 通过第7.5节中开发的对角化方法，我们可以 mpute

通过矩阵指数的完整温度响应：

$$e^{At} = \begin{bmatrix} -0.65 & -0.24 & 0.72 \\ -0.41 & 0.93 & -0.06 \\ -0.64 & -0.28 & -0.69 \end{bmatrix} \begin{bmatrix} e^{-4.37t} & 0 & 0 \\ 0 & e^{-2.83t} & 0 \\ 0 & 0 & e^{-1.80t} \end{bmatrix} \begin{bmatrix} -0.59 & -0.13 & 0.54 \\ -0.25 & 0.87 & 0.12 \\ -0.52 & -0.38 & -0.83 \end{bmatrix}$$

这种分解可以指导实际的 HVAC 控制策略。最快的模态会自然达到平衡，因此控制系统应侧重补偿衰减最慢的模态。温度传感器应放置在能够最好地观测主要特征模态的位置，而建筑设计则可通过调整绝缘层与热质量分布来优化特征值谱，从而改善热响应。

Example: 一个设计良好的建筑可能会在 -2.0 小时$^{-1}$ 处具有其最慢的特征值，这意味着最慢的热模式以约 20 分钟的半衰期衰减。

本文提出的方法可方便地扩展到更大的建筑、更复杂的热网络以及其他扩散型系统。无论是在设计 HVAC 控制系统、优化传感器布置，还是规划建筑改造，特征值分析都能为系统行为提供关键洞见，并指导工程决策。

Heavy Vehicle Suspension Analysis

重度阻尼悬架系统的动力学——例如建筑机械和重型卡车中所采用的系统——为实特征值分析提供了一个富有启发性的应用。与通常表现出振荡行为的乘用车（这一主题将在第8章讨论）不同，重度阻尼系统呈现出纯指数模态，与本章所建立的理论完全一致。

考虑重型车辆单个车轮的“四分之一车辆”模型。令 $x(t)$ 表示车身（簧载质量）的竖向位移，$y(t)$ 表示车轮组件（非簧载质量）相对于其平衡位置的位移。该系统对每个质量分别应用牛顿第二定律：

$$M\ddot{x} = -k(x - y) - c(\dot{x} - \dot{y})$$

$$m\ddot{y} = k(x - y) + c(\dot{x} - \dot{y}) - k_t y$$

其中，M 为四分之一车辆车身质量，m 为车轮组件质量，k 为悬架弹簧刚度，c 为阻尼器系数，k_t 为轮胎弹簧刚度。

为了通过特征理论分析这个四阶系统，我们通过引入状态向量将其转换为一阶形式：

$$\mathbf{z} = \begin{pmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{pmatrix}$$

由此得到矩阵方程$\dot{\mathbf{z}} = A\mathbf{z}$，其中：

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k/M & k/M & -c/M & c/M \\ k/m & -(k + k_t)/m & c/m & -c/m \end{bmatrix}$$

取重型商用车辆的典型取值（$M = 2500$ kg，$m = 200$ kg，$k = 100000$ N/m，$k_t = 800000$ N/m，$c = 15000$ Ns/m），特征值呈现为四个彼此不同的实数值：

$$\lambda_1 = -12.3, \quad \lambda_2 = -8.4, \quad \lambda_3 = -3.2, \quad \lambda_4 = -0.9$$

Think: 特征值之间的宽广分离反映了不同的物理时间尺度：快速的轮胎挠曲、中等的悬架运动，以及缓慢的车身回稳过程。

这些彼此不同的实特征值是过阻尼系统的特征，表明不存在振荡的纯指数衰减模态。通过矩阵指数得到的显式解采用了第7.5节中给出的形式：

$$e^{At} = \sum_{j=1}^{4} e^{\lambda_j t} \boldsymbol{v}_j \boldsymbol{w}_j^T$$

其中 $\boldsymbol{v}_j$ 和 $\boldsymbol{w}_j^T$ 分别为右特征向量和左特征向量。每个模态都以其自身的特征速率衰减，并具有相应的物理解释：

1. λ_1 模式：快速轮胎挠度吸收 2. λ_2 模式：主悬架响应 3. λ_3 模式：车身-悬架耦合运动 4. λ_4 模式：缓慢的最终回稳

这种特征结构通过若干原则指导重型车辆悬架设计：

1. 对于纯阻尼，所有特征值应为实数且为负 2. 时间尺度分离可防止不希望的耦合 3. 模态形状（特征向量）决定有效的传感器布置 4. 最慢的模态（λ_4）决定整体整定时间

Example: 矿用卡车使用多个位移传感器来监测悬架运动，传感器的位置经过选择，以便最佳地观测通过该分析识别出的主导固有模态。

施工设备的设计往往有意追求这种过阻尼行为，以在变化的载荷条件下保持稳定运行。特征向量通过揭示每个衰减模态如何体现在可测量的运动中，从而决定传感器和执行器的最优布置。

扩展到整车动力学引入了额外的自由度——每个角点的垂向运动以及车身的俯仰和侧倾——从而产生更大的矩阵，但原理相似。特征结构揭示了诸如对角相对角之间载荷分担等耦合模态，这些模态在载荷转移条件下对稳定性至关重要。

这一机械应用通过展示特征理论在另一类物理系统中的力量，补充了我们的热分析。热模态源于几何对称性而呈现简单的指数衰减，而悬挂模态则由于机械耦合而表现出分层的衰减速率。二者共同说明了本章所建立的数学框架如何阐明多样的物理现象，并通过线性代数这一统一语言，将四阶微分方程转化为清晰可理解的模态响应。

□ ——————————————————————— □

Exercises: Chapter 7

1. 设 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$。求其特征值和特征向量。利用这些计算 e^{At}，并描述当 $t \to \infty$ 时系统 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 的解的行为。

2. 求 $A = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix}$ 的特征值和特征向量。利用这些结果求解初值问题 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$，其中 $\boldsymbol{x}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$。

3. 计算 $B = \begin{bmatrix} 3 & -2 & 0 & 0 \\ 3 & 0 & -4 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ 的特征值和特征向量。B 是否可对角化?

4. 证明：若 A 是一个三角矩阵（上三角或下三角），则其对角线元素与特征值相同。

This is an extremely useful result.

5. 使用伴随矩阵将方程 $\ddot{x} + 2\dot{x} + 5x = 0$ 转换为一阶系统。说明该矩阵的特征值与原二阶方程的特征方程的根之间的关系。

6. 对于微分方程 $\ddot{x} + 3\dot{x} + 2x = 0$，求所有形如 $e^{\lambda t}$ 的线性无关解。用任意常数 c_i 写出通解。

7. 令 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$，并通过对其级数展开的前四项求和来计算 e^{At}。将其与通过对角化得到的结果进行比较。为什么它们不同?

8. 矩阵 A 的特征多项式为 $p(\lambda) = \lambda^3 - 6\lambda^2 + 11\lambda - 6$。$A$ 是否可对角化?

9. 一个矩阵 A 的特征多项式为 $p(\lambda) = \lambda^4 - 2\lambda^3 - \lambda^2 + 2\lambda$。求它的特征值，并判断 A 是否必然可对角化。你能对 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 的解的长期行为说些什么?

10. 矩阵 A 的迹是其对角元之和。证明迹等于其特征值之和（按重数计算）。

11. 考虑三阶方程 $\dddot{x} + 4\ddot{x} + 5\dot{x} + 2x = 0$。求其解空间的一组基，并表示通解。对于初始条件 $x(0)$、$\dot{x}(0)$ 和 $\ddot{x}(0)$ 的哪些取值，解在 $t \to \infty$ 时保持有界?

12. 证明对于任意可对角化矩阵 A，矩阵 A 和 e^A 具有相同的特征向量。如果 $\boldsymbol{v}$ 是 A 的一个特征向量，其特征值为 λ，那么 e^A 的对应特征值是什么?

13. 证明方阵 A 的行列式等于其特征值的乘积：为简单起见，假设 A 是可对角化的，尽管该结果在一般情况下也成立。

14. 证明如果 A 可对角化且所有特征值均为实数且为负，则 $\lim_{t\to\infty} e^{At} = 0$（零矩阵）。

15. 如果 A 是一个可对角化的 2×2 矩阵，并且 $\operatorname{tr}(A) = 0$，你能对它的特征值说些什么？这对系统 $\dot{\boldsymbol{x}} = A\boldsymbol{x}$ 的解说明了什么?

16. 考虑 $n \times n$ 的方阵 A 和 B。需要什么条件才能 con-

包括看似无害的结果 $e^{(A+B)t} = e^{At}e^{Bt}$。给出一个 2×2 情况的例子，使该公式不成立。

17. 考虑矩阵 $A = \begin{bmatrix} 3 & -1 \\ 2 & 4 \end{bmatrix}$。计算它的特征多项式 $p(\lambda)$，并直接验证 $p(A) = 0$（零矩阵）。这对矩阵与其特征多项式之间的关系有什么启示?

18. *Cayley-Hamilton theorem* 表示每个方阵都满足其特征方程。对于 $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$，通过证明如果 $p(\lambda) = \det(A - \lambda I)$，则 $p(A) = 0$ 来验证这一点。这告诉你关于 $A - 2I$ 的幂有什么信息?

19. 对于任意的 2×2 矩阵 A，证明 $A^2 - \operatorname{tr}(A)A + \det(A)I = 0$。解释这是 Cayley–Hamilton 定理在 2×2 矩阵情况下的一个特例。

20. 假设A是一个3×3矩阵，其特征多项式为$p(\lambda) = \lambda^3 - 7\lambda^2 + 14\lambda - 8$。在不进行任何矩阵运算的情况下，确定$A$的迹和行列式，并给出用$A$的较低次幂表示$A^3$的一个表达式

21. 在$\mathbb{R}^3$上的旋转矩阵R是一个行列式为1的正交矩阵。解释为什么R必须具有1作为特征值，从而绕某个固定轴旋转。这个论证是否可以扩展到$\mathbb{R}^n$中的旋转，适用于所有n？（提示：考虑特征多项式以及奇数维度与偶数维度的区别）。

22. 证明对于任何方阵 A，e^A 的特征值是 $\{e^\lambda\}$，其中 $\{\lambda\}$ 是 A 的特征值。解释为什么这意味着 e^A 总是可逆的。

23. 前面的练习表明，对于 A 一个方阵，e^A 是可逆的。它的逆是什么？猜测逆矩阵比证明它是可逆的更容易吗?

24. 凯莱-哈密顿定理提供了一种计算非奇异矩阵逆的途径。如果 $p(\lambda)$ 是 A 的特征多项式，写出 $p(\lambda) = a_n\lambda^n + \cdots + a_1\lambda + a_0$ 并证明当 A 可逆时：

$$A^{-1} = -\frac{1}{a_0}(a_nA^{n-1} + \cdots + a_2A + a_1I)$$

使用此方法找到A^{-1}对于 $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ 并与其他方法进行比较。

25. 对于具有不同特征值 λ_1 和 λ_2 的 2×2 矩阵 A，使用 Cayley-Hamilton 定理证明：

$$A = \frac{\lambda_2 I - A}{\lambda_2 - \lambda_1}\lambda_1 + \frac{A - \lambda_1 I}{\lambda_2 - \lambda_1}\lambda_2$$

该表达式展示了 A 如何沿其特征空间进行分解，即使不显式计算特征向量也是如此。

26. 对于具有 $x_1 = 1, x_2 = 2$ 和 $x_3 = 3$ 的 3×3 Vandermonde 矩阵，求其特征值并验证它们是实数。然后计算迹并验证其等于 $x_1 + x_2 + x_3$ 加上你应当确定的额外项。这对 x_i 值与 Vandermonde 矩阵的迹之间的普遍关系有何启示?

对于不同的实数 x_1，...，x_n，范德蒙矩阵 V 定义为

$v_{ij} = (x_i)^{j-1}$:

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

首先证明当 n = 为 2 时，$\det(V) = x_2 - x_1$。然后对 n = 为 3 的情况证明 $\det(V) = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$。基于这一模式，用 x_i 的差来表述一般公式 $\det(V)$，并给出一个猜想。证明该公式。当 $x_i = i$ 时，这对特征值说明了什么？

Chapter 8
Eigenvalue Complexities

"With songs of sweetest cadence to the turning spindle & reel"

现实超越了实指数。上一章中那套优雅的理论——以实特征值产生纯粹的增长与衰减——必须直面一个更深层的真理：世界以节律脉动。心脏细胞以同步的波动振荡；动物种群以可预测的周期激增又崩塌；金融市场在繁荣与萧条之间呼吸。甚至我们的思想和社会运动也遵循涨落的模式，观念随着代际更迭而消退又复兴。这种无处不在的周期性似乎挑战了我们精心构建的特征值与指数解的框架。

然而，数学并非破裂，而是弯曲以适应。复特征值的引入化解了我们在描述上的危机，将振荡揭示为与增长或衰减同样自然的现象。当特征值重合时，更为丰富的模式随之涌现——多项式项以乘子形式与我们的指数项相结合，其方式映射出耦合系统中同步行为逐步出现的过程。起初看似对理论的复杂化，最终显现为不可或缺的结构，正是我们试图理解的现象本身所要求的。

这些复杂性——虚特征值、重根及其微妙的相互作用——将我们从简单的对角化引向约当标准形这一尽可能理想的框架。该框架统一了我们对线性演化的处理，同时阐明了几乎相同的特征值如何以微妙的方式产生截然不同的行为。数学本身仿佛在简洁与繁复的两极之间脉动，回响着它所描述的节律。

8.1 Complex Eigenvalues & Oscillation

简谐振子——一个没有阻尼的弹簧上的质量——是我们首次遇到超越真实特征值的行为。运动方程

$$\frac{d^2x}{dt^2} + \omega^2 x = 0$$

具有涉及正弦和余弦的著名解法：

$$x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t)$$

然而，这些解似乎与我们第七章中的理论相冲突，在该理论中所有解都是实指数的组合。通过$x_1 = x$和$x_2 = \dot{x}$转换为一阶形式，得到

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & 0 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

特征多项式 $\det(A - \lambda I) = \lambda^2 + \omega^2$ 具有根 $\lambda = \pm i\omega$ —— 我们第一次遇到复特征值。

例 8.1（复本征向量）。考虑角度 $\pi/3$ 的旋转矩阵：

$$R = \begin{bmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$$

特征方程是 $\lambda^2 - \lambda + 1 = 0$，得到特征值 $\lambda = \frac{1}{2} \pm i\frac{\sqrt{3}}{2} = e^{\pm i\pi/3}$。对于 $\lambda = \frac{1}{2} + i\frac{\sqrt{3}}{2}$，我们解 $(R - \lambda I)\boldsymbol{v} = 0$：

$$-\frac{\sqrt{3}}{2}\begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{0}$$

yielding 特征向量 $\boldsymbol{v} = (1, i)^T$。尽管该特征向量是复数，矩阵 R 将实向量映射到实向量。看似的悖论可以通过注意到 $\boldsymbol{v}$ 及其复共轭 $\boldsymbol{v} = (1, -i)^T$ 在复数域上张成 $\mathbb{R}^2$ 来解决，而它们的实部和虚部在实数域上张成 $\mathbb{R}^2$。

◇

这些复杂的特征值不仅没有使我们的理论失效，反而阐明了它。欧拉公式 $e^{i\theta} = \cos\theta + i\sin\theta$ 显示出我们的三角解只是伪装成复杂指数的形式：

$$c_1 \cos(\omega t) + c_2 \sin(\omega t) = \Re\left(z_1 e^{i\omega t} + z_2 e^{-i\omega t}\right)$$

对于适当的复常数z_1和z_2。振荡运动源自指数为纯虚数的指数解。

通过对基本的 2×2 矩阵的研究，这一见解得以推广

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{8.1}$$

Nota bene: 矩阵 $J = \sqrt{-I}$ 是复数运算中 i 的矩阵类比。

直接计算表明 $J^2 = -I$，这暗示了与复数算术之间的深刻联系。事实上，矩阵指数 e^{Jt} 可以直接由其级数定义计算得到：

$$\begin{aligned} e^{Jt} &= I + tJ + \tfrac{t^2}{2!}J^2 + \tfrac{t^3}{3!}J^3 + \tfrac{t^4}{4!}J^4 + \cdots \\ &= \left(I - \tfrac{t^2}{2!}I + \tfrac{t^4}{4!}I - \cdots\right) + \left(tJ - \tfrac{t^3}{3!}J + \tfrac{t^5}{5!}J - \cdots\right) \\ &= (\cos t)I + (\sin t)J \end{aligned} \tag{8.2}$$

这个欧拉公式的矩阵值类比解释了 J 的几何作用——平面中的纯旋转。

这种矩阵在应用中经常出现——实部 α 控制增长或衰减，而虚部 β 决定振荡频率。

更一般地，考虑一个具有复共轭特征值 $\alpha \pm i\beta$ 的矩阵 A。这个矩阵——和 J 一样——不能在实数域上对角化，但我们可以将其化简为最简单的形式。

引理 8.2（复正规形）。 *Any 2×2 matrix A with complex conjugate eigenvalues $\alpha \pm i\beta$ is similar to $\alpha I + \beta J$. That is, there exists an invertible matrix P such that*

$$P^{-1}AP = \alpha I + \beta J \tag{8.3}$$

Proof. 设 $\boldsymbol{v} = \boldsymbol{u} + i\boldsymbol{w}$ 是特征值 $\alpha + i\beta$ 的一个特征向量。 则

$$A\boldsymbol{v} = (\alpha + i\beta)\boldsymbol{v} \implies A(\boldsymbol{u} + i\boldsymbol{w}) = (\alpha + i\beta)(\boldsymbol{u} + i\boldsymbol{w})$$

令实部和虚部相等：

$$A\boldsymbol{u} = \alpha\boldsymbol{u} - \beta\boldsymbol{w} \quad \text{and} \quad A\boldsymbol{w} = \beta\boldsymbol{u} + \alpha\boldsymbol{w}$$

矩阵 $P = [\boldsymbol{u}\ \ \boldsymbol{w}]$ 是可逆的（因为 $\boldsymbol{v}$ 不可能是纯实数或纯虚数），并且

$$AP = [\alpha\boldsymbol{u} - \beta\boldsymbol{w}\ \ \beta\boldsymbol{u} + \alpha\boldsymbol{w}] = P(\alpha I + \beta J)$$

从而得到期望的相似度。 □

这个 *normal form* 是具有复特征值的 2×2 矩阵最简单的表示形式。基于相似矩阵在指数运算下的行为方式，我们对 A 进行指数化的策略是先进行坐标变换，对 $\alpha I + \beta J$ 求指数，然后再变换回原坐标。欧拉定理再次显现。

引理 8.3（欧拉定理再述）。

$$e^{(\alpha I+\beta J)t} = e^{\alpha t}(\cos(\beta t)I + \sin(\beta t)J) \tag{8.4}$$

Proof. 由于I和J对易，我们有：

$$e^{(\alpha I+\beta J)t} = e^{\alpha tI+\beta tJ} = e^{\alpha tI}e^{\beta tJ}$$

在式（8.2）中计算了 e^{Jt} 之后，我们得出结论：

$$e^{(\alpha I+\beta J)t} = e^{\alpha t}I(\cos(\beta t)I + \sin(\beta t)J) = e^{\alpha t}(\cos(\beta t)I + \sin(\beta t)J)$$

如所声称。 □

这种分解揭示了复特征值如何产生螺旋运动——由旋转调制的指数增长或衰减。

8.2 *Repeated Eigenvalues*

并非所有变换都能保持其特征子空间彼此分离。当特征值重合时，相应的特征向量可能以微妙的方式合并或增殖。这种特征子空间的碰撞——这种完美分离的失效——导致的行为比纯粹的缩放更为丰富，却仍然比复杂的振荡更为简单。理解这种行为需要仔细区分特征值出现的次数，以及它所控制的独立方向的数量。

考虑这些矩阵

$$A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

两者具有特征多项式 $(\lambda - 2)^2$，从而得到特征值 $\lambda = 2$，其 *algebraic multiplicity* 为 2。然而，这些矩阵的行为却大不相同：A_1 在所有方向上仅进行纯缩放，而 A_2 将缩放与剪切相结合。这一区别源于 *geometric multiplicity*——特征空间 $\ker(A - 2I)$ 的维数。

定义 8.4（特征值的重数）。特征值的 *algebraic multiplicity* 是它作为特征多项式的一个根的重数。*geometric multiplicity* 是其特征空间 $\ker(A - \lambda I)$ 的维数。

- 几何重数不超过代数重数。

对于上面的A_1，两种重数都等于2——每个非零向量都是特征值为2的特征向量。对于A_2，代数重数为2，但几何重数只有1，因为只有$(1, 0)^T$的倍数是特征向量。这种特征向量的缺失表明存在更深层的结构，单纯的对角化无法刻画。

当几何重数和代数重数不同时，特征向量本身无法张成空间。我们需要 *generalized eigenvectors* — 向量 $\boldsymbol{w}$ 满足 $(A - \lambda I)^k \boldsymbol{w} = 0$，对于某些 $k > 1$。这些向量生成包含多项式项与指数函数相乘的解，如我们上面的例子所示。

引理 8.5（广义特征空间）。*For eigenvalue λ, the sequence of subspaces*

$$\ker(A - \lambda I) < \ker(A - \lambda I)^2 < \ker(A - \lambda I)^3 < \cdots$$

stabilizes at dimension equal to the algebraic multiplicity of λ.

BONUS! 这种广义特征空间的嵌套——或任何嵌套子空间的序列——称为 *filtration*。滤波在通过分级低维实体（如*cf.*泰勒多项式、傅里叶级数等）来逼近大型复杂空间时非常重要。

例 8.6（广义特征向量）。对于矩阵

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

特征值 $\lambda = 2$ 的代数重数为 3，但几何重数为 1。广义特征向量链

$$\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{w}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{w}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

满足

$$(A - 2I)\boldsymbol{w}_1 = \boldsymbol{0}, \quad (A - 2I)\boldsymbol{w}_2 = \boldsymbol{w}_1, \quad (A - 2I)\boldsymbol{w}_3 = \boldsymbol{w}_2$$

◇

这种更为丰富的结构——由长度递增的链所构成的广义特征向量——为理解具有重特征值的变换提供了关键。尽管这类变换无法对角化，但通过仔细分析广义特征空间如何相互作用，其行为仍然是可以理解的。随着我们在后续章节中发展线性演化的完整理论，这一理解将证明至关重要。

例 8.7（指数化一个简单的块）。考虑矩阵

$$R = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

这个矩阵可以写成 $R = 2I + N$，其中 I 是单位矩阵，并且

$$N = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

是 *nilpotent* — 意味着 N 的某个幂等于零。实际上，$N^3 =$ 为 0，但 $N^2 \neq$ 不为 0。为了求解 e^{Rt}，我们使用一个指数规则：

$$e^{Rt} = e^{(2I+N)t} = e^{2It}e^{Nt} = e^{2t}e^{Nt}$$

由于 I 与任何矩阵都可交换，因此这是有效的。计算 e^{Nt} 的难度并不大，得益于指数级数和幂零性：

$$\begin{aligned} e^{Nt} &= I + Nt + \tfrac{N^2t^2}{2!} + \tfrac{N^3t^3}{3!} + \cdots \\ &= I + Nt + \tfrac{N^2t^2}{2!} \end{aligned}$$

系列终止，因为所有更高次幂的 N 都消失了。计算 N^2 并进行求值得到：

$$e^{Rt} = e^{2t}e^{Nt} = e^{2t}\begin{bmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{2t} & te^{2t} & \frac{t^2}{2}e^{2t} \\ 0 & e^{2t} & te^{2t} \\ 0 & 0 & e^{2t} \end{bmatrix}$$

该解揭示了一种引人注目的结构：指数增长 e^{2t} 由 t 中的多项式项所调制。这种模式——多项式-指数乘积——刻画了具有重特征值的系统的解。幂零部分 N 沿着超对角线向上形成了影响的级联，上方的每一层都会继承其下方的动力学。◇

8.3 The Jordan Canonical Form

追求最简形式将我们带出对角化。当特征值重合或变为复数时，纯粹的缩放让位于更丰富的结构——一种由*Jordan canonical form*捕捉到的影响级联。这一终极分解揭示了线性变换的真实结构，展示了如何从缩放、旋转和广义特征向量的相互作用中出现一般演化。

我们的旅程通过重复特征值和多项式-指数解指向一个更深的统一性。回想一下，具有重复特征值的矩阵可能缺乏一组完整的独立特征向量，需要广义特征向量来生成空间。之前的例子展示了

这种结构在解中如何表现——纯指数通过沿着广义特征向量链的影响级联获得多项式系数。这些模式，表面上看似特殊情况，实际上揭示了所有线性变换必须采取的通用形式。

定理 8.8（若尔当标准形）。*Every square matrix A is similar to a block diagonal matrix J, called its* 若尔当标准形*:*

$$A = VJV^{-1} = V \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_m \end{bmatrix} V^{-1}$$

where each 乔丹块 *J_i has one of two forms:*

1. *For a real eigenvalue λ of algebraic multiplicity k, the Jordan block is the k-by-k matrix:*

$$J_i = \lambda I + N = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

这种块结构与实数情形的结构相对应，其中超对角的 I 矩阵扮演由全 1 构成的幂零矩阵的角色，连接着由 2×2 个复块 C 构成的一条链。

2. *For complex conjugate eigenvalues $\alpha \pm i\beta$ of algebraic muliplicity k, the Jordan block is the $2k$-by-$2k$ block matrix*

$$J_i = \begin{bmatrix} C & I & 0 & \cdots & 0 \\ 0 & C & I & \cdots & 0 \\ 0 & 0 & C & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I \\ 0 & 0 & \cdots & 0 & C \end{bmatrix} \quad : \quad C = \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} \quad : \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where each entry is a 2-by-2 block.

使用复值的若当标准形会更简单；只需要一种类型的块。鉴于我们研究常微分方程解的动机，我们选择保持在实数范围内。

The form J is unique up to the ordering of blocks.

每个 Jordan 块对应一个特征值（实数或复数）及其相关的广义特征向量。块的大小等于该特征值的最长广义特征向量链的长度。当所有特征值均为实数且互不相同（或为实数且具有独立特征向量）时，每个块为 1×1，且我们恢复第 7 章的对角形式。

每个块内的结构揭示了变换的作用。对于实特征值，矩阵 N 在上对角线为 1、其余为 0，是 *nilpotent* —— 其某个幂等于零。这个幂零性解释了我们在解中看到出现的有限多项式项。对于复特征值，2×2 块编码了第 8.1 节中所示的旋转。

例 8.9（约当结构）。该矩阵

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

已经处于约当标准形。它有两个约当块：一个对应特征值 $\lambda = 2$ 的 2×2 块，以及另一个对应 $\lambda = 3$。广义特征向量链的长度分别为 2 和 2。解将包含来自每个块的幂零部分的 te^{2t} 和 te^{3t} 等项。$\diamond$

约当分解为理解矩阵的幂和指数提供了关键。由于 J 是分块对角的，其指数也是分块对角的：

$$e^{At} = Ve^{Jt}V^{-1} = V \begin{bmatrix} e^{J_1 t} & 0 & \cdots & 0 \\ 0 & e^{J_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{J_k t} \end{bmatrix} V^{-1}$$

对于每个实约当块 $J_i = \lambda I + N$，我们有：

$$e^{J_i t} = e^{\lambda t} e^{Nt} = e^{\lambda t} \left(I + Nt + \frac{N^2 t^2}{2!} + \cdots + \frac{N^{m-1} t^{m-1}}{(m-1)!} \right)$$

其中 m 是块的大小。该级数终止，因为 $N^m = 0$。对于复数块，使用先前研究的旋转矩阵会得到类似的公式。

例 8.10（完全若尔当分解）。考虑矩阵

$$A = \begin{bmatrix} 7 & -4 & 4 \\ 4 & -1 & 4 \\ 0 & 0 & 3 \end{bmatrix}$$

其特征多项式（$\lambda - 3$）($\lambda - 3$）($\lambda - 3$）揭示了特征值 $\lambda = 3$，其代数重数为 3。计算广义特征向量

表明几何重数为 1，从而得到若尔当标准形：

$$A = V \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} V^{-1}$$

其中 V 包含广义特征向量。单个若尔当块表明所有广义特征向量构成一条链。解将涉及项 e^{3t}、te^{3t} 和 $t^2e^{3t}/2$。

Nota bene: 计算影响坐标变换V的广义特征向量并不总是直截了当。详见例 8.11。

◇

若尔当标准形揭示了：每一个线性变换在合适的坐标下，都通过以下组合起作用：

1. 纯缩放（对角线项） 2. 旋转（复共轭块）
3. 级联影响（超对角线项）

这种分解提供了线性演化的完整图景，将我们此前对特征值、重根以及复数行为的研究统一到一个连贯的结构之中。

8.4 Finding the Jordan Form

前面各节揭示的约当形之优美结构，如今面临一个实际挑战：给定一个矩阵 A，究竟如何实际计算相似变换 V，将 A 化为约当形？不同于特征值可以从特征多项式中清晰地得到，或特征向量可以通过零空间计算求出，约当结构要求我们仔细考察广义特征向量如何串联成一组基。我们的任务是把理论上的理解转化为实际的计算。

该过程从特征值及其重数开始。对于每个特征值 λ：

1. 代数重数 m 来自特征多项式 2. 几何重数 $k = \dim \ker(A - \lambda I)$ 计算独立特征向量的数量 3. 差值 $m - k$ 揭示了我们需要多少广义特征向量

然而，仅凭这些数字只能讲述故事的一部分。真正的工作在于发现广义特征向量如何形成链条，从而构建我们的变换矩阵 V。

回顾第 8.2 节，我们知道每个特征值 λ 都会生成一系列嵌套的子空间：

$$\ker(A - \lambda I) < \ker(A - \lambda I)^2 < \ker(A - \lambda I)^3 < \cdots$$

当该滤过达到等于 λ 的代数重数的维数时，它会稳定下来。挑战在于从这种嵌套结构中提取广义特征向量链。每一条链都以特征向量 $\boldsymbol{v}_1$ 为起点，并通过求解如下形式的方程逐步构建：

$$(A - \lambda I)\boldsymbol{v}_{i+1} = \boldsymbol{v}_i$$

当存在多条链时，我们必须仔细追踪它们之间的关系，才能正确构建 V。

例 8.11（若尔当变换）。考虑如下 5×5 矩阵：

$$A = \begin{bmatrix} 3 & 2 & 1 & 0 & 2 \\ 0 & 3 & -1 & 0 & 3 \\ 0 & 0 & 3 & 0 & -4 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

特征多项式 $(\lambda - 3)^3(\lambda - 2)^2$ 显示特征值 $\lambda_1 = 3$，代数重数为 3，$\lambda_2 = 2$，代数重数为 2。让我们系统地构造乔丹分解。

对于 $\lambda_1 = 3$，显然 $\ker(A - 3I) = \text{span}(1, 0, 0, 0, 0)^T$，表明几何重数为 1。还需要另外两个向量来完成该链。求解

$$(A - 3I)\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{v}_2 = \begin{pmatrix} a \\ 1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

(其中 a 是一个自由参数)，然后

$$(A - 3I)\boldsymbol{v}_3 = \begin{pmatrix} a \\ 1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{v}_3 = \begin{pmatrix} b \\ a/2 + 1/4 \\ -1/2 \\ 0 \\ 0 \end{pmatrix}$$

(其中 b 是另一个自由参数，) 提供了我们的第一条链 $\boldsymbol{v}_3 \mapsto \boldsymbol{v}_2 \mapsto \boldsymbol{v}_1 \mapsto 0$。

对于 $\lambda_2 = 2$，求解 $(A - 2I)\boldsymbol{v} = 0$ 可得 $\lambda = 2$ 具有几何和

代数重数为2，具有独立的特征向量

$$v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad : \quad v_5 = \begin{pmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

变换矩阵 V 必须按链长度递减的顺序，将这些链作为列组合起来：

$$V = [v_3 \; v_2 \; v_1 \; v_4 \; v_5]$$

检查$a = 0$ 和 $b = 1/12$ 在此示例中足以生成约旦标准形：

Ouch! 术语“*one*”*check*在这里承担了主要作用……

$$J = V^{-1}AV = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

J的结构反映了特征值的代数重数以及在我们系统分解中发现的广义特征向量链的长度。

◇

从这一计算中可以得出若干实用原则：

1. 按照代数重数递减的顺序系统地处理特征值 2. 对于每个特征值，首先找到所有独立的特征向量 3. 从尚未在其他链中使用的特征向量开始构建链 4. 通过按长度递减的顺序排列链来组装 V 5. 通过直接计算 AV 和 VJ 验证结果

两个常见的陷阱值得特别注意。首先，在求解（$A - \lambda I$）$v_{i+1} = v_i$ 的广义特征向量时，解总是存在的，但包含自由参数。不同的选择可以产生不同的但等价的Jordan基。其次，构造V时需要特别注意列的顺序——链必须按顺序排列，以产生J中的正确块结构。

因此，Jordan 标准形的实际计算将本章中发展起来的理论结构统一在一起。广义特征空间的滤过引导我们寻找链，而精心的基的构造则将这些链转化为相似变换 V。

尽管这一过程比单纯的特征值计算需要更为精细的处理，但对这些原理的系统性运用能够得到线性变换最深刻的分解。

8.5 Computing Eigenvalues

在专注于约旦标准型及其计算时，还有一个初步步骤需要考虑：我们如何在实际中计算特征值？特征多项式并不简单——即使是中等大小的矩阵也会产生难以直接求解的方程。大矩阵的特征值需要更复杂的方法来提取。有一个基于不同矩阵分解的显著算法，它通过基本操作的精心组合，将我们的理论理解转化为实际计算。

首先考虑为什么天真方法会失败。一个100x100的矩阵产生一个100次多项式，远远超出了标准求根方法的范围。甚至计算系数本身也证明是危险的——展开$\det(A - \lambda I)$需要进行天文数字级的运算，且伴随灾难性的误差增长。然而，现代软件能够在几秒钟内计算千乘千矩阵的特征值。这个看似矛盾的问题通过迭代而非直接求解得到解决。

算法的优雅来自相似变换——这是我们在特征值研究中反复遇到的主题。如果矩阵 A 和 B 是相似的，它们共享特征值，同时可能提供不同的计算优势。QR 算法通过精心迭代利用这一原理：

定义 8.12 （QR 算法）。给定矩阵 $A_0 = A$，*QR algorithm* 通过以下过程生成一个序列：

1. 因子 $A_k = Q_k R_k$，其中 Q_k 是正交的，R_k 是上三角矩阵 2. 形式 $A_{k+1} = R_k Q_k$ (反向乘以)

每次迭代保持相似性：$A_{k+1} = Q_k^T A_k Q_k$。 •

尽管描述简单，但此过程蕴含着显著的特性。序列 $\{A_k\}$ 收敛到上三角形式，特征值出现在对角线上。共轭复数对自然地以 2x2 块的形式出现，而重复的特征值保持其重数。该算法有效地对所有特征空间执行同时幂迭代，同时通过正交变换保证了至关重要的数值稳定性。

例 8.13（简单迭代）。考虑矩阵

$$A_0 = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

经过五次迭代，我们发现：

$$A_5 \approx \begin{bmatrix} 3.414 & 0.892 & 0.247 \\ 0.000 & 2.000 & 0.731 \\ 0.000 & 0.000 & 0.586 \end{bmatrix}$$

对角线元素收敛到特征值，约为 3.414、2.000 和 0.586——它们是特征多项式 $\lambda^3 - 6\lambda^2 + 11\lambda - 6$ 的根，而我们从未显式地计算过。$\diamond$

对于具有复特征值的矩阵，对角线上自然会出现 2×2 块：

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad \text{representing} \quad \lambda = a \pm ib$$

这种结构使得计算可以完全在实数算术中进行，同时仍然能够捕捉复杂行为——这是我们在第 8.1 节研究旋转时首次遇到的一个原则。

现代实现通过若干改进增强了这一基本迭代：

1. 位移通过将迭代聚焦在可能的特征值附近来加速收敛 2. 隐式更新在降低运算次数的同时保持结构 3. 消去技术将已收敛的特征值隔离，以实现高效处理

Example: 移位 QR 算法先减去特征值的一个估计 μ，进行迭代，然后再加回 μ——当 μ 选择得当时，可显著改善收敛性。

该算法的力量源于它对矩阵结构的尊重。不同于特征多项式的计算——后者为了纯代数而舍弃几何性——QR 迭代保持了关键的关系：

- 正交变换精确保留特征值
- 上三角形式能够直接揭示特征值的近似值
- 相似矩阵会保留若当结构，尽管它可能被掩盖

尽管在数值过程中我们失去了对约当结构的显式访问，但特征值本身却以惊人的精度显现出来。这种在理论理解与实际计算之间的平衡，体现了看似抽象的概念如何引导高效算法的发展。那些在理论上阐明特征值结构的思想，为在实践中求得它们提供了工具。

8.6 Back to Basis

我们对约当标准形的阐述阐明了高阶线性微分方程解的结构。广义特征向量和复块的抽象工具可以直接转化为具体的解的模式——多项式项乘以指数函数，以及由复共轭对产生的三角函数。这样的理解完善了我们在第7章开始建立的基解图景，揭示了这些模式为何以及如何必然出现。

考虑一个 n 阶的一般线性齐次微分方程，其特征多项式 $p \in \mathcal{P}_n$ 的次数为 n：

$$p(D)x = (D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0)x = 0$$

当 $p(\lambda) = 0$ 的根是重根时，伴随矩阵落入约当块。基解直接从第 8.3 节计算的矩阵指数 e^{Jt} 的第一行中得到。对于具有特征值 λ、大小为 k 的约当块，我们得到：

$$e^{Jt} = e^{\lambda t}e^{Nt} = e^{\lambda t}\left(I + Nt + \frac{N^2t^2}{2!} + \cdots + \frac{N^{k-1}t^{k-1}}{(k-1)!}\right)$$

该矩阵的第一行给出了我们的基本解：

定理 8.14（重根的基解）。 *Let λ be a root of the characteristic equation $p(\lambda) = 0$ with algebraic multiplicity k. Then:*

1. *For real λ, the functions*

$$\phi_j(t) = t^j e^{\lambda t}, \quad j = 0, 1, \ldots, k-1$$

   form a basis for the solution space corresponding to λ.

2. *For complex $\lambda = \alpha \pm i\beta$, the functions*

$$\phi_j(t) = t^j e^{\alpha t}\cos(\beta t), \quad \psi_j(t) = t^j e^{\alpha t}\sin(\beta t), \quad j = 0, 1, \ldots, k-1$$

   form a real basis for the solution space corresponding to the conjugate pair.

这些解直接源自第 8.3 节中计算的指数化约当块的第一行。

例8.15（重实根）。该方程

$$\frac{d^3x}{dt^3} - 3\frac{d^2x}{dt^2} + 3\frac{dx}{dt} - x = 0$$

具有特征多项式 $(\lambda - 1)^3 = 0$。基解为

$$\phi_0(t) = e^t, \quad \phi_1(t) = te^t, \quad \phi_2(t) = \frac{1}{2!}t^2e^t$$

与其伴随矩阵的约当形中所见的模式完全一致。这些源自我们先前关于一个 3×3 约当块的例子中计算得到的矩阵指数的第一行。通解是它们的线性组合：

$$x(t) = c_0\phi_0(t) + c_1\phi_1(t) + c_2\phi_2(t)$$

◇

例8.16（阻尼振子）。该方程

$$\frac{d^2x}{dt^2} + 2\gamma\frac{dx}{dt} + \omega^2 x = 0 \tag{8.5}$$

在 $0 < \gamma < \omega$ (欠阻尼情况) 下具有特征方程

$$\lambda^2 + 2\gamma\lambda + \omega^2 = 0$$

其根为 $\lambda = -\gamma \pm i\sqrt{\omega^2 - \gamma^2}$。解

$$x(t) = e^{-\gamma t}\left(c_1\cos(\sqrt{\omega^2-\gamma^2}\,t) + c_2\sin(\sqrt{\omega^2-\gamma^2}\,t)\right)$$

呈现阻尼振荡——以速率 γ 的指数衰减调制频率为 $\sqrt{\omega^2 - \gamma^2}$ 的振荡。

◇

例 8.17（临界阻尼）。在方程（8.5）中，如果我们将阻尼设为精确的取值 $\gamma = \omega$，当阻尼系数恰好与固有频率相平衡时，这就表示临界阻尼。其特征多项式

Example: 在机械系统中，临界阻尼表示在不发生振荡的情况下最快地回到平衡——所有特征值在负实轴上的同一点重合，从而使衰减达到最大。

$$\lambda^2 + 2\omega\lambda + \omega^2 = (\lambda + \omega)^2$$

具有重根 $\lambda = -\omega$。基解现在为

$$\phi_1(t) = e^{-\omega t} \quad \text{and} \quad \phi_2(t) = te^{-\omega t}$$

欠阻尼振荡与纯衰减之间的过渡在这一临界情形中显现——当根合并时，多项式项 t 取代了三角函数。◇

例 8.18（重复的复根）。考虑四阶方程

$$\frac{d^4x}{dt^4} + 2\frac{d^2x}{dt^2} + x = 0$$

其特征多项式 $\lambda^4 + 2\lambda^2 + 1 = (\lambda^2 + 1)^2$ 具有重数为 2 的根 $\lambda = i$。由复解 $t^j e^{it}$，我们得到四个实基解：

$$\phi_1(t) = \cos(t), \quad \phi_2(t) = \sin(t), \quad \phi_3(t) = t\cos(t), \quad \phi_4(t) = t\sin(t)$$

这些直接来自于指数化复数Jordan块的第一行，其中多项式系数调节纯正弦波运动，以创造更复杂的振荡模式。

◇

这些例子展示了乔丹结构与标量解之间的完美对应。每个乔丹块的幂零部分表现为t的幂，而复块生成了对建模振荡至关重要的三角函数。最初看似抽象的矩阵理论揭示了其作为描述物理运动的自然语言。

完整的解通常通过结合每个乔丹块的贡献来得出：

- 真实特征值导致指数衰减或增长
- 复数对产生振荡项
- 重复的根添加多项式因子

这种分解——将运动拆分为其基本模式——为理解线性系统提供了理论洞察和实际方法。

Power Grid Stability

电力电网的稳定性提供了一个实践中约旦标准型形式的引人注目的示范。现代电力网络由数百台发电机通过输电线路连接，每台机器都在同步运动的精妙舞蹈中发挥作用。当这场舞蹈出现失误时，停电可能会像瀑布一样蔓延至整个大陆。《第8章》的数学揭示了通过分析网络的约旦结构，精确地说明了这种不稳定性是如何发展的。

首先考虑一个简单的情况：两个发电机通过输电线连接。每个发电机相对于电网的名义频率具有一个角度 θ_i。从平衡的微小偏差遵循线性化的动态：

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{pmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{pmatrix} + \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \begin{pmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{pmatrix} + \begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \mathbf{0}$$

Historical Note: 2003年的东北大停电影响了5000万人，其起因是过载的输电线路下垂并触及树木。由此导致的同步性丧失生动地表明，局部不稳定性如何通过电网结构传播。

m_i 表示发电机惯性，d_i 阻尼，k 传输线路耦合。通过引入相位角和频率作为状态变量，转换为一阶形式得到一个 4×4 系统矩阵 A。

当生成元具有相同参数时，关键特征便会显现：$m_1 = m_2 = m$ 和 $d_1 = d_2 = d$。随后 A 出现代数重数为二的特征值：

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k/m & k/m & -d/m & 0 \\ k/m & -k/m & 0 & -d/m \end{bmatrix}$$

直接计算表明这些特征值成对出现：

$$\lambda_{1,2} = -\frac{d}{2m} \pm i\sqrt{\frac{2k}{m} - \frac{d^2}{4m^2}}$$

每个的代数重数为 2，但几何重数为 1。其物理意义很清楚：相同的发电机可以同相或反相振荡，而复特征值表明存在振荡运动。

根据第 8.3 节，该矩阵具有若尔当标准形：

$$A = PJP^{-1}$$

其中 J 取实约当标准形，包含两个 2×2 块：

$$J = \begin{bmatrix} C & I_2 \\ 0 & C \end{bmatrix}, \quad C = \begin{bmatrix} -d/(2m) & -\omega \\ \omega & -d/(2m) \end{bmatrix}$$

其中 $\omega = \sqrt{2k/m - d^2/(4m^2)}$ 给出振荡频率。这种实数形式，如第 8.3 节所述，揭示了通过 C 表现出的阻尼振荡以及通过 I_2 表现出的模态之间的耦合。

Example: 现代电网控制系统通过结合基于模型的分析和实时测量，持续评估稳定性边际。当这些与系统特征值相关的评估表明稳定性边际减少或振荡风险增加时，操作员会收到警报以调整系统条件。

非平凡的乔丹结构的出现表明微妙的不稳定性：尽管系统看起来是对称的，但小的扰动可能会激发逐渐增长的振荡，因为生成器失去同步。电网操作员必须仔细监控这些特征值汇聚的近共振条件。

真实的电力网络涉及数百台发电机，导致稳定性矩阵具有极其庞大的维度。计算其约当结构需要第 8.5 节中所发展的精细数值方法。QR 算法被证明至关重要，因为当约当块开始形成时，直接的特征值计算恰恰会变得不稳定——而这正是运行人员必须监控的危险工况区间。

现代网格分析采用了这些方法的复杂变种。与其计算完整的Jordan分解（这在数值上是精细的），操作符通过使用第12.3节中的稳健迭代方案追踪特征值在复平面上的迁移。当特征值接近合并，指示出初步的Jordan块形成时，网络可以重新配置以保持稳定性。

关键的见解是，电力网络中的乔丹结构不仅仅是数学上的好奇——它还表示需要立即关注的物理条件。每个乔丹块中的单位超对角线表示一个不稳定性在发电机之间传播的路径。理解这一结构对于防止级联故障至关重要。

电网稳定性因此例证了乔丹标准形的数学如何塑造关键基础设施。第8.2节中看似抽象的复特征值与几何重数之间的微妙相互作用，在大规模发电机的同步运动中得到了物理表现。现代电网的可靠性基本上依赖于对这一结构的精确数值计算。

Nota bene: 与机械共振不同，电网不稳定通过乔丹块形式可以在没有外部迫使的情况下发生——这是一个令人警觉的提醒，表明仅仅数学结构本身就可能引发物理灾难。

深刻的教训在于，看似完全相同的组件——例如参数匹配的发生器——会由于其相同的特征值而产生微妙的危险。约当标准形精确地揭示了这种对称性如何

使不稳定性得以传播。现在，每个主要的电网控制中心都采用本章中开发的这些数学工具，以维持现代社会所依赖的可靠电力供应。

Chemical Process Network Analysis

在大型化工过程网络中识别弱耦合子系统，展示了约当标准形分析的又一项强大应用。现代化工厂常将数十个反应器、换热器和分离单元组合成复杂网络，其中物料与能量流动在各过程之间形成微妙的耦合。理解这些耦合——尤其是哪些过程可以被视为近似独立——对安全性和高效控制都至关重要。

考虑一个具有六个耦合反应釜的化工装置。在每个反应釜 i 中，关键反应物的浓度 $c_i(t)$ 和温度 $T_i(t)$ 都按照耦合的微分方程演化。反应釜之间的物料流动造成了浓度耦合，而共享的冷却系统引入了热耦合。令 $\boldsymbol{x} = (c_1, T_1, \dots, c_6, T_6)^T$ 表示完整状态。在质量作用动力学和牛顿冷却的标准假设下，系统具有如下形式：

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}$$

对于一个在近稳态运行的典型石油化工过程，我们可能会发现如下的耦合矩阵：

$$A = \begin{bmatrix} A_{11} & A_{12} & \varepsilon_{13} \\ A_{21} & A_{22} & 0 \\ \varepsilon_{31} & 0 & A_{33} \end{bmatrix}$$

其中，每个 A_{ii} 是一个描述两个耦合容器的 4×4 块，A_{ij} 表示相邻单元之间更强的耦合，而 ε_{ij} 表示通过共享公用工程的弱耦合。以环氧乙烷生产装置的现实取值为例：

$$A_{11} = \begin{bmatrix} -2.1 & 0.8 & 0.5 & 0.1 \\ 0.6 & -1.7 & 0.1 & 0.4 \\ 0.5 & 0.1 & -1.9 & 0.7 \\ 0.1 & 0.4 & 0.5 & -1.6 \end{bmatrix}$$

Example: 在环氧乙烷生产中，反应器温度通常在200–300°C之间，这使得通过共享冷却系统进行的热耦合成为一项重要的安全考量。

对于 A_{22} 和 A_{33} 具有相似的结构，而耦合项在强耦合和弱耦合情况下分别具有 0.1 或 0.01 量级的条目。

通过第8.5节的方法计算特征值揭示了三个不同的簇，每个都由耦合强度 ε 参数化：

$$\lambda_1 = -2.2 + O(\varepsilon), \quad \lambda_2 = -2.2 + O(\varepsilon)$$

$$\lambda_3, \lambda_4, \lambda_5 = -1.8 + O(\varepsilon)$$

$$\lambda_6, \dots, \lambda_{12} = -1.5 + O(\varepsilon)$$

这些聚集的特征值表明过程单元之间几乎独立。Jordan分解确认了这一结构：

$$J = \begin{bmatrix} J_1 & \varepsilon E_{12} & \varepsilon E_{13} \\ \varepsilon E_{21} & J_2 & \varepsilon E_{23} \\ \varepsilon E_{31} & \varepsilon E_{32} & J_3 \end{bmatrix}$$

其中，每个J_i对应一簇特征值，而每个E_{ij}都是其元素为一阶量的矩阵，因此非对角块包含ε阶的项。

这 de　组成对过程控制和有着深远的影响　　安全：

1. 每条工艺列可以准独立地控制 2. 扰动基本上局限在各个块内 3. 关键失效模式与约旦块结构一致 4. 传感器布置应遵循子系统边界

与每个约当块相关的广义特征向量准确揭示了扰动在各单元之间的传播方式。例如，1号容器中的一次温度偏移会对其配对的2号容器产生强烈影响，但仅通过$O(\varepsilon)$项对其他工艺列产生较弱的影响。这一认识对于安全系统设计至关重要。

Historical Note: 1974年弗利克斯伯勒灾难等重大化学事故凸显了理解工艺扰动如何在弱耦合系统中传播的重要性。

现代化学工厂设计明确考虑了这种模块化结构。关键工艺被分组以最小化与其他单元的耦合，同时安全系统围绕已识别的子系统边界进行设计。乔丹标准型为量化这些设计决策提供了数学框架。

第8.5节中的计算方法对大型系统至关重要，因为在最需要识别弱耦合之时，直接的特征值计算恰恰会变得数值不稳定。即便直接方法失效，QR 迭代仍能可靠地揭示成簇的特征值。

这一化学工程应用通过展示约旦结构如何阐明弱耦合系统，补充了我们的颤振分析。航空器由于物理对称性而呈现精确重复的特征值，而化学过程则因弱耦合而表现出近似重复。二者共同表明，本章所建立的数学框架能够指导复杂工程系统的分析与设计。

□ ——————————————————————————— □

Exercises: Chapter 8

1. 矩阵

$$A = \begin{bmatrix} 5 & -12 \\ 4 & 5 \end{bmatrix}$$

具有复特征值。求解它们，并用三角函数表示e^{At}。描述该矩阵在t增加时对$\mathbb{R}^2$中向量的几何作用。

2. 求其特征值和广义特征向量

$$A = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

然后使用它们来计算 e^{At}。通过检查它同时满足 $(d/dt)e^{At} = Ae^{At}$ 和 $e^{A0} = I$ 来验证你的答案。

3. 考虑一个 5×5 矩阵

$$A = \begin{bmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

求其约当标准形并计算 e^{At}。哪些解随 $t \to \infty$ 增长?

4. 对于系统

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$$

找出在 $t \to \infty$ 时仍保持有界的所有解。证明它们构成一个子空间，并求其维数。

5. 考虑矩阵

$$A = \begin{bmatrix} -1 & -2 & 1 & 0 & 0 \\ 2 & -1 & 0 & 1 & 0 \\ 0 & 0 & 4 & 1 & 0 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

求它的若尔当标准形，并求一个矩阵 V 使得 $V^{-1}AV$ 给出该标准形。e^{At} 的形式是什么?

6. 对于四阶方程

$$\frac{d^4x}{dt^4} + 4\frac{d^2x}{dt^2} + 4x = 0$$

求所有线性无关的解。（提示：特征多项式可分解为 $(\lambda^2 + 2)^2$）。

7. 考虑矩阵

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

对于每个矩阵：列出所有特征值及其代数重数和几何重数，并识别 Jordan 块及其大小；然后，计算 e^{At} 和 e^{Bt}。

8. 设 A 是一个 4×4 矩阵，其特征多项式为 $p(\lambda) = (\lambda - 5)^4$。$A$ 有多少种不同的约当标准形？画出示意图展示所有可能性。

9. 对于一个具有成对共轭复特征值 $1 \pm 2i$ 和 $-3 \pm i$ 的实 4×4 矩阵 A，可能的约当标准形有哪些？画出展示所有可能性的示意图。

10. 设

$$A = \begin{bmatrix} 1 & -4 & 2 & 0 \\ 1 & 1 & 0 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

找出与其特征值相一致的所有可能的若尔当标准形。哪一个是正确的？说明如何在不计算 V 的情况下确定这一点。

11. 考虑矩阵

$$A = \begin{bmatrix} 3 & -4 & 1 & 0 & 0 \\ 4 & 3 & 0 & 1 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix}$$

求其约当标准形。最小多项式是什么? A 的 ial ？ 12. 证明对于任意大小为 $n \times n$ 的幂零矩阵 N，该级数

$$I + tN + \frac{t^2}{2!}N^2 + \frac{t^3}{3!}N^3 + \cdots + \frac{t^{n-1}}{(n-1)!}N^{n-1}$$

给出 e^{Nt} 正是如此。用这个来找到e^{Nt} 对于具有超对角线为 1 的 4×4 幂零矩阵 N。 13. 设 A 为一个实 $n \times n$ 矩阵。证明：如果 λ 是 A 的一个特征值，其代数重数为 m，则方程 $(A - \lambda I)^m \boldsymbol{x} = 0$ 的解所构成的空间恰好是一个维数为 m 的子空间。 14. 对于一个具有重特征值 λ 的 3×3 矩阵 A，证明恰好存在三种可能的若尔当标准形。画出图示，展示每种情形下的广义特征向量链。 15. 一个 4×4 矩阵 A 的特征多项式为 $(\lambda + 1)^2(\lambda - 3)(\lambda + 2)$。如果 $\ker(A + I)$ 是一维的，那么 A 的若尔当标准形必须是什么？请证明你的结论。 16. 对于一个 $n \times n$ 矩阵 A，其特征值为 λ_1 , . . . , λ_n (，并按重数) 计数，证明 $\operatorname{tr}(e^{At}) = \sum_{k=1}^{n} e^{\lambda_k t}$。利用这一结果说明：如果所有特征值的实部都为负，则当 $t \to \infty$ 时，$\operatorname{tr}(e^{At}) \to 0$。 17. 设 A 为一个实 2×2 矩阵，具有复特征值 $a \pm bi$，其中 $b \neq 0$。证明 A 不能表示为两个具有实特征值的实矩阵的乘积。（提示：考虑矩阵乘法下特征值的变化。）

18. 设 A 为一个实 4×4 矩阵，其特征值为 $3 \pm 4i$ 和 $-2 \pm i$。必须是什么形式

$$\lim_{t \to \infty} \frac{e^{At}}{e^{3t}}$$

取吗？请为您的答案提供理由。

19. 证明对于任何 $n \times n$ 矩阵 A，其 Jordan 分解中的幂零部分 N 满足 $N^n = 0$。给出一个 5×5 的幂零矩阵 N，其中 $N^4 \neq 0$ 但 $N^5 = 0$。

20. 使用Cayley-Hamilton定理证明，如果A是幂零矩阵（即某个幂等于零），则它的特征多项式必须是$p(\lambda) = \lambda^n$，其中n是矩阵的大小。这告诉你关于幂零矩阵的特征值有什么信息?

See the Chapter 7 exercises for the Cayley-Hamilton Theorem.

21. 设 A 为一个 $n \times n$ 矩阵。证明 A (的最小多项式——即湮灭 A) 的最低次数的首一多项式——必须整除特征多项式。这与约当标准形有何关系?

Chapter 9
Linear Iterative Systems

“the sea of Time & Space beat round the Rock in mighty waves”

权力在线性变换的迭代中孕育权力。每次将矩阵应用于向量时，都会将权重向主导方向移动，通过优选路径进行传递，直到某种自然平衡出现。这一基本过程——反复变换向主导模式的收敛——贯穿了理论和计算。尽管比前几章中连续流动和乔丹结构更简单，但这些迭代系统在矩阵结构和渐近行为之间微妙的相互作用中蕴含着独特的美。

核心思想在于矩阵幂与特征值之间的关系。在最简单的情况下，迭代线性变换会导致一个特征值主导结果，将所有向量拉向其特征方向。这个过程将矩阵提炼为其本质轴。求解最大特征值的幂法自然地从迭代过程中产生，将一个理论洞察转化为一个计算工具。

然而，并非所有矩阵都能如此轻易地被分析。特殊结构——正性约束、概率守恒、对称性——都使我们的研究更加复杂且富有启发性。Perron-Frobenius 理论揭示了正性如何强制主导方向的唯一性。随机矩阵在推动系统趋向平衡状态的同时保持总体测度。图矩阵通过其谱编码离散拓扑。每一类矩阵都带来了其独特的谱特征，从而塑造了长期行为。

我们的发展从抽象到具体，从一般的迭代到具有特殊属性的矩阵的结构化类别，这些特殊属性既照亮了理论又推动了应用。我们发现的模式自然地从有限维度扩展到无限维度，从离散步骤到连续

连续极限。贯穿始终，我们坚持这样一种观点：迭代既提供理论洞见，也赋予实践力量——每一次变换的应用都使我们更接近理解其根本本质。

9.1 *At First Iteration*

斐波那契数列为线性迭代内在的奥秘提供了最初的一瞥。每一个数都是其前两个数之和，由此生成了这一数列

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

比这些数字本身更引人注目的是一个隐藏的模式：相邻项之比趋近于一个固定值 $\varphi = (1+\sqrt{5})/2 \approx 1.618$。这种比值的收敛暗示了线性递推关系中更深层的结构。

是的，这就是所谓的 *golden ratio*。不，它并非因为在艺术与自然中无处不在而出现在这里。它出现在这里是因为它是一个主特征值：继续阅读。

斐波那契规则 $x_{n+2} = x_{n+1} + x_n$ 表示一个二阶递推。正如我们在第七章中将二阶微分方程转化为一阶系统一样，我们也可以将这个标量序列重新表示为一个向量迭代。设定

$$\boldsymbol{v}(n) = \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix}$$

将递归转换为

$$\boldsymbol{v}(n+1) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \boldsymbol{v}(n)$$

这个一阶向量系统通过矩阵乘法刻画了相同的演化过程。它的解是明确的，尽管并不一定能够明确计算：

$$\boldsymbol{v}(n) = A^n \boldsymbol{v}(0) = A^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

更一般地，一个随时间演化的离散时间线性系统具有如下形式及其解

$$\boldsymbol{x}(n+1) = A\boldsymbol{x}(n) \quad \Rightarrow \quad \boldsymbol{x}(n) = A^n \boldsymbol{x}(0) \tag{9.1}$$

其中 A 编码了进化规则。正如微分方程 $d\boldsymbol{x}/dt = A\boldsymbol{x}$ 通过无穷小变化生成连续流动，递推关系 (9.1) 通过逐步迭代创建离散轨迹。

这个离散框架在经济系统中有自然的应用，其中规律性的时间周期（季度、年份）带来了固有的粒度。

考虑一个被分为 n 个部门的经济，每个部门生产的商品部分被其他部门消费，部分作为最终产出。*input-output matrix* $A = [a_{ij}]$ 编码了生产需求：入口 a_{ij} 表示生产 j 部门产出一单位所需的 i 部门产出的数量。该系统的演变遵循方程 (9.1)，其中 $\boldsymbol{x}(n)$ 表示在第 n 期的部门产出。

Historical Note: 这种经济模型被称为*Leontief input-output model*，以W. Leontief命名，他在20世纪30年代通过对美国经济的详细研究开发了这一模型。他关于生产中结构性相互依存的研究为他赢得了1973年诺贝尔经济学奖。尽管最初是利用手工艰苦收集的数据开发的，如今这类模型通过自动化的数据收集与计算为全球的经济规划提供依据。

示例 9.1（六部门经济）。考虑一个有六个主要部门的经济：

1. 农业（食品生产） 2. 能源（发电）
3. 制造业（工业品） 4. 交通运输（物流） 5. 服务业（商业/消费） 6. 科技（IT/通信）

输入-输出矩阵 A 捕捉了它们之间的相互依赖关系：

$$A = \begin{bmatrix} 0.15 & 0.08 & 0.10 & 0.05 & 0.20 & 0.05 \\ 0.25 & 0.30 & 0.35 & 0.40 & 0.20 & 0.30 \\ 0.10 & 0.15 & 0.20 & 0.25 & 0.10 & 0.20 \\ 0.15 & 0.12 & 0.15 & 0.15 & 0.10 & 0.08 \\ 0.20 & 0.25 & 0.15 & 0.15 & 0.25 & 0.30 \\ 0.15 & 0.15 & 0.20 & 0.15 & 0.25 & 0.20 \end{bmatrix}$$

每一列表示某一部门单位产出的投入需求。例如，生产一单位制造业产出（第3列）需要0.10单位的农业投入、0.35单位的能源投入，等等。由于不包括劳动和其他外部因素，各列的合计约为0.95。

在迭代过程中出现了一个显著的模式。无论初始条件如何，各部门的相对大小最终趋于固定比例，约为$(0.22, 0.62, 0.34, 0.25, 0.48, 0.39)^T$，而整体经济每个周期大约增长1.09（或9%）。这意味着经过足够的时间后，能源部门（2）将稳定在农业部门（1）的近2.8倍大小，服务部门（5）维持约为技术部门（6）1.2倍的大小，依此类推——尽管初始条件差异极大且存在复杂的相互关系。$\diamond$

Foreshadowing: 从列昂惕夫模型中涌现的均衡增长率预示了正矩阵的佩龙–弗罗贝尼乌斯理论，其中的特殊结构确保了占优正特征值的唯一性。

这种张力——表观上按部门的相互作用的复杂性与渐近行为的简洁性之间——体现了线性系统中的一个更广泛原则。通过多个变量的复杂耦合，往往会简化为由主导模态驱动的更简单模式。我们的任务在

本章旨在理解这一约化，揭示特征结构如何塑造线性迭代的长期行为。

9.2 *Dominance & Convergence*

斐波那契比率的神秘收敛暗示着矩阵幂中更深层的结构。正如比值序列 x_{n+1}/x_n 逼近黄金比例 φ，一般的矩阵迭代也常常呈现出类似的收敛——其行为由一个主导的特征值所支配，推动长期演化。将复杂的迭代化约为简单的尺度变化，反映了一条基本原理：反复的线性变换会将矩阵提炼为其本质特征。

理解此类收敛性的关键在于引理 7.7 中矩阵幂的表示。对于可对角化矩阵 $A = V\Lambda V^{-1}$，其幂具有如下形式

$$A^n = V\Lambda^n V^{-1}$$

其中 Λ^n 只是将每个特征值提升到 n 次幂。当这些特征值在大小上存在差异时，较大的特征值比较小的增长得更快，最终在迭代中占据主导地位。这一观察引出了关键的定义：

定义 9.2（谱半径与支配性）。矩阵 A 的 *spectral radius* ρ_A 是其特征值的最大模：

$$\rho_A = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$$

若$|\lambda_*| = \rho_A$且不存在其他特征值具有该大小，则特征值λ_*是*dominant*。其对应的特征向量$\boldsymbol{v}_*$称为*dominanteigenvector*。

Nota bene: 特征值主导性的概念同样在常微分方程（ODE）和连续时间动力学中占据主导地位；然而，对于连续时间的主导性而言，重要的并不是谱半径——起决定作用的是特征值的实部。为什么？微分算子 D 是移位算子 E 的自然对数。

•

引理 9.3（支配收敛）。 *Let A be diagonalizable with dominant eigenvalue λ_* and corresponding eigenvector $\boldsymbol{v}_*$. Then for any initial vector $\boldsymbol{x}_0$ with nonzero component C along $\boldsymbol{v}_*$,*

$$A^n \boldsymbol{x}_0 \simeq C\lambda_*^n \boldsymbol{v}_* \text{ as } n \to \infty$$

Proof. 对特征值 $\{\lambda_i\}$ 进行编号，使得 $\lambda_* = \lambda_1$。将 $\boldsymbol{x}_0$ 写成 A 的特征基 $\{\boldsymbol{v}_i\}$ 下的表示，并应用 A^n，得到

$$A^n \boldsymbol{x}_0 = A^n \sum_{k=1}^{n} c_k \boldsymbol{v}_k = \sum_{k=1}^{n} c_k \lambda_k^n \boldsymbol{v}_k \tag{9.2}$$

$$= c_1 \lambda_*^n \left(\boldsymbol{v}_* + \sum_{k=2}^{n} \frac{c_k}{c_1} \left(\frac{\lambda_k}{\lambda_*} \right)^n \boldsymbol{v}_k \right) \simeq c_1 \lambda_*^n \boldsymbol{v}_* \tag{9.3}$$

由于对所有$k > 1$，$|\lambda_k/\lambda_1| < 1$，这些比率随着$n \to \infty$呈指数衰减至零。常数$c_1 = C$。 □

这一收敛原理是用于计算主特征值的 *power method* 的基础。以随机向量 $\boldsymbol{x}_0$ 为起点，对 A 进行反复相乘并在每次之后进行归一化，会收敛到主特征方向。收敛速度由第二大特征值幅值与主特征值幅值之比所决定。

例9.4（斐波那契再论）。对于前一节中的斐波那契矩阵，

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

特征值为 $\varphi = (1 + \sqrt{5})/2$ 和 $\psi = (1 - \sqrt{5})/2$，其中 $|\varphi| > 1 > |\psi|$。随着 $n \to \infty$，主导特征值 φ^n 的幂以指数级速度增长得比 $|\psi|^n$ 更快，这解释了为什么相邻的斐波那契数会趋近于比值 φ。

◇

例9.5（经济收敛）。回到例9.1，六部门经济的投入产出矩阵 A 的占优特征值 $\lambda_* = \rho_A \approx$ 为 1.09，并具有相应的占优特征向量

$$\boldsymbol{v}_* \approx (0.22, 0.62, 0.34, 0.25, 0.48, 0.39)^T$$

归一化为单位长度。这同时解释了9%的增长率以及经验上观察到的固定比例——经济结构迫使系统无论初始条件如何都收敛到这些比率。◇

当 A 不可对角化时，Jordan 块会使情况更复杂，但并不会从根本上改变这一图景。Jordan 块的幂

$$J_k = \lambda I + N$$

涉及二项式项：

$$J_k^n = \lambda^n \left(I + \frac{n}{\lambda} N + \frac{n(n-1)}{2\lambda^2} N^2 + \cdots \right)$$

n中的多项式因子会产生修正，但无法克服最大$|\lambda|$的指数级主导性。因此，谱半径仍然控制渐近行为，尽管收敛过程可能由于约当结构而呈现出细微的振荡。

这种将矩阵迭代简化为主导模态的做法体现了一个更广泛的原理：复杂的线性系统往往会显著地简化

在重复的作用下。无论是建模经济增长、人口动态还是网络影响，长期行为通常反映出一个或两个特征方向。关键在于识别何时发生这种简化，并利用由此产生的简化效果。

9.3 *Positivity & Perron-Frobenius Theory*

诸如质量、能量和人口等某些自然量以下界为零；概率与测度同样顽固地拒绝踏入负值领域。当线性变换作用于这些内在为正的量时，其矩阵会继承一种特殊结构，这种结构深刻地塑造了它们的谱性质。正是这种结构——现实世界测量的内在正性——引出了关于特征值与渐近行为的引人注目的结论。

考虑一个矩阵 $A = [a_{ij}]$，其所有元素均为正数：$a_{ij} > 0$ 对于所有 i，j。此类矩阵自然出现在建模耦合增长过程时，其中每个组件都对所有其他组件产生正向影响。Perron-Frobenius 理论揭示了此类矩阵具有独特且简单的谱属性：

定理 9.6（Perron-Frobenius）。*Let A be a square matrix with all entries strictly positive. Then A has a dominant eigenvalue $\lambda_* = \rho_A > 0$, with corresponding dominant eigenvector $\boldsymbol{v}_* > 0$ having all positive components. Moreover, any nonnegative eigenvector of A must lie in the dominant eigenspace.*

Historical Note: O. Perron 首次在 1907 年证明了这些关于正矩阵的结果。G. Frobenius 于 1912 年将其扩展到更一般的非负不可约矩阵，涵盖了经济学和概率理论中的更广泛应用。

这个非凡定理的证明阐明了正性如何塑造谱结构。考虑对任何正向量 $\boldsymbol{x}_0$ 进行 A 的迭代。归一化向量序列 $\boldsymbol{y}_n = A^n\boldsymbol{x}_0/\|A^n\boldsymbol{x}_0\|$ 保持正性，并且一个关键的不等式出现：

$$\min_i \frac{(A\boldsymbol{y})_i}{(\boldsymbol{y})_i} \le \rho_A \le \max_i \frac{(A\boldsymbol{y})_i}{(\boldsymbol{y})_i}$$

对于任何正向量 $\boldsymbol{y}$ 的约束。这些由正性强制的界限限制了谱半径。一个微妙的论证表明，当 $n \to \infty$ 时，这些界限收敛，从而得到主特征值及其正特征向量。保持正性通过振荡或衰减变得不可能，因此迫使所有其他特征值具有严格较小的大小。

有一个更优雅、犀利的证明，使用了代数拓扑，作者对此特别偏爱。

示例 9.7（研究引文网络）。考虑计算机科学中六个主要研究领域之间的影响网络，其中矩阵 B 的条目 b_{ij} 表示相对引文流：在给定年份内，领域 j 中所有引用 *re- ceived by* 论文的引文中，有多少比例来自 *from*。

在领域 i 的论文中，领域级别的平均揭示了引用模式以及一个Perron-Frobenius主导特征向量，如下所示：

$$B = \begin{bmatrix} 0.60 & 0.25 & 0.15 & 0.10 & 0.05 & 0.05 \\ 0.20 & 0.50 & 0.15 & 0.10 & 0.05 & 0.05 \\ 0.10 & 0.10 & 0.50 & 0.15 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.10 & 0.50 & 0.10 & 0.05 \\ 0.03 & 0.05 & 0.05 & 0.10 & 0.65 & 0.10 \\ 0.02 & 0.05 & 0.05 & 0.05 & 0.10 & 0.70 \end{bmatrix} \quad : \quad \boldsymbol{v}_* \approx \begin{pmatrix} 0.220 \\ 0.250 \\ 0.180 \\ 0.148 \\ 0.122 \\ 0.081 \end{pmatrix}$$

Nota bene: 这些数字是人为构造的，但使用矩阵来跟踪引用的原理是合理的，并且在许多其他情境中确实非常有用。

这些对应于：(1) 机器学习，(2) 人工智能，(3) 计算机视觉，(4) 机器人学，(5) 系统与网络，以及 (6) 理论。例如，在考察 AI 论文所收到的全部引用（第 2 列）时，25%（$b_{12} = 0.25$）来自 ML 论文，50%（$b_{22} = 0.50$）来自其他 AI 论文，其余较小比例来自其他领域。Perron–Frobenius 特征向量 $\boldsymbol{v}_*$（对应于 $\lambda_* = 1$）的特征向量）为领域影响力的自然排序或引用的平稳分布提供了依据，其中 AI 和 ML 占主导地位，但仍与应用领域保持显著耦合。主特征值为 $\lambda_* = 1$；次大特征值的幅度为 $|\lambda_2| \approx 0.647$。这揭示了收敛到平稳分布 $\boldsymbol{v}_*$ 的速率；其他特征模态的影响在每个引用周期中都会按 ≈ 0.647 的因子衰减。收敛到最终分布的几个百分点以内发生得相对较快，大致在 N 个周期内，其中 0.647^N 较小（例如，$N \approx 5$ 个周期即可衰减到 10%）。$\diamond$

这些结果的力量超越了严格正的矩阵。若矩阵 A 的所有元素都为 ≥ 0，则称其为 *nonnegative*；若不存在任何坐标置换能够将其变换为块上三角形式，则称其为 *irreducible*。这些性质刻画了底层相互作用网络的连通性：不可约性意味着每个组成部分最终都会影响到其他所有组成部分，无论是直接影响还是通过中介间接影响。

定理 9.8（弗罗贝尼乌斯扩张）。 *The conclusions of the Perron-Frobenius theorem hold for nonnegative irreducible matrices, with the spectral radius* $\rho_A > 0$ *still positive but possibly having complex conjugate eigenvalues of magnitude* ρ_A.

考试 ple 9.9（催化循环）。考虑一个化学反应 n etwork

底物 S_1、S_2、S_3、S_4 在转移矩阵作用下循环转换

$$\begin{array}{ccc} S_1 & \xrightarrow{k_1} & S_2 \\ {\scriptstyle k_4}\uparrow\uparrow & & \downarrow\downarrow{\scriptstyle k_2} \\ S_4 & \xleftarrow[k_3]{} & S_3 \end{array} \quad : \quad A = \begin{bmatrix} 0 & 0 & 0 & k_4 \\ k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & k_3 & 0 \end{bmatrix}$$

每个反应都由酶催化。A的特征多项式为$\lambda^4 - k_1k_2k_3k_4 = 0$，特征值均匀分布在复平面半径为$\rho_A = (k_1k_2k_3k_4)^{1/4}$的圆上。反应的循环性质体现在这两个复特征值对中，而不可约性确保了底物之间存在一个正的稳态分布。

生物化学家称这些网络为 *futile cycles*，当它们似乎除了能量消耗之外没有其他用途时，尽管它们常常在细胞代谢中发挥着关键的调节作用。

◇

正性与谱性质之间的联系体现了一个更为普遍的原则：对矩阵元素的约束往往会迫使其特征结构受到约束。理解这些关系——即矩阵结构如何塑造谱行为——既能提供理论洞见，也能为分析线性系统提供实用工具。Perron–Frobenius 理论或许是这一原则最为优雅的例证，其中简单的正性便能导出关于渐近行为的深刻结论。

9.4 *Stochastic Matrices & Markov Chains*

许多现实世界的过程涉及状态之间的转变，其中概率控制变化。这些例子包括天气模式在晴天和雨天之间的变化、动物种群在领土之间的迁移以及遗传特征在不同世代之间的传递。这类过程，其中未来的状态仅依赖于当前状态而与过去历史无关，自然引出了马尔可夫链的理论——一个通过线性代数统一离散随机过程的框架。

定义 9.10（马尔可夫链）。*Markov chain* 是一列随机变量 $\{X_n\}_{n\geq 0}$，其取值来自状态集合 S，满足 *Markov property*：

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i)$$

从状态i在一步内转移到状态j的概率p_{ij}定义了该链的 *transition matrix* $P = [p_{ij}]$中的一个条目。

Relax... 如果你还没有学习条件概率，可以不用太担心，直接继续。等有机会的时候，了解一下条件概率。现在，先将 P 作为概率矩阵使用。

•

马尔可夫性质——即未来依赖于现在而非过去——将时间演化转化为矩阵迭代。为了理解这一联系，我们必须首先澄清我们的向量代表什么：

定义 9.11（概率分布）。向量 $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ 是一个 *probability distribution* 如果：

1. 非负性：$x_i \geq 0$ 对于所有 i2. 总概率：$\sum_{i=1}^{n} x_i = 1$ 条目 x_i 表示处于状态 i 的概率

.

•

马尔可夫链的转移矩阵必须保持这些概率约束，从而引出我们研究的核心对象：

定义 9.12（随机矩阵）。一个方阵 $P = [p_{ij}]$ 若满足以下条件，则称其为 *stochastic*：

1. 非负性：$p_{ij} \geq 0$ 对于所有 i，j 2. 列随机性：$\sum_{i=1}^{n} p_{ij} = 1$ 对于所有 j

条目 p_{ij} 表示从状态 j 在一步内转移到状态 i 的概率。

Caveat: 许多作者使用行随机矩阵，其中各行是概率分布。如果使用术语 *stochastic*，务必仔细核对指的是哪一种类型。

•

马尔可夫链中概率的演化遵循我们熟悉的迭代模式：

$$\boldsymbol{x}(k+1) = P\boldsymbol{x}(k)$$

其中，$\boldsymbol{x}(k)$ 表示在步骤 k 时在各状态上的概率分布。对 P 的随机性约束确保，如果 $\boldsymbol{x}(k)$ 是一个概率分布，那么 $\boldsymbol{x}(k+1)$ 也将是概率分布。

例 9.13（天气模式）。考虑一个简单的模型，描述每日天气在三种状态之间的转移：晴天（S）、多云（C）和雨天（R）。历史数据表明其转移概率为：

$$P = \begin{bmatrix} 0.7 & 0.3 & 0.2 \\ 0.2 & 0.4 & 0.3 \\ 0.1 & 0.3 & 0.5 \end{bmatrix}$$

按列向下读取：从晴天（第一列）开始，天气以 0.7 的概率转变为晴朗，以 0.2 的概率转变为多云，以 0.1 的概率转变为下雨；从多云或下雨状态的转移同理。

给定初始分布 $\boldsymbol{x}(0) = (1, 0, 0)^T$，表示今天晴天的确定性，明天的分布变为：

$$\boldsymbol{x}(1) = P\boldsymbol{x}(0) = (0.7, 0.2, 0.1)^T$$

展示概率如何在各个状态之间分散。两天后：

$$\boldsymbol{x}(2) = P^2\boldsymbol{x}(0) = (0.55, 0.27, 0.18)^T$$

su暗示着向某种均衡分布收敛 离子。 ◇

随机矩阵的谱性质决定了它们的长期行为。一个马尔可夫链（及其转移矩阵 P）如果可以从任一状态到达任一其他状态（不一定在一步之内），则称其为 *irreducible*。如果对任一状态而言，所有可能返回路径长度的最大公约数为 1，则称其为 *aperiodic*。不可约且非周期的马尔可夫链称为 *ergodic*。每一个列随机矩阵都有一个特别重要的特征值：

引理 9.14（随机谱半径）。*For any stochastic matrix P:*

1. *$\lambda = 1$ is an eigenvalue of P.*
2. *All eigenvalues satisfy $|\lambda| \leq 1$.*
3. *The spectral radius $\rho_P = 1$.*
4. *If P is irreducible, the eigenvalue 1 is simple (algebraic multiplicity one).*
5. *If P is ergodic (irreducible and aperiodic), then 1 is the unique eigenvalue of P with magnitude 1.*
6. *If P is irreducible but periodic with period $h > 1$, there are h distinct eigenvalues on the unit circle.*

Proof. 首先，注意到 $\mathbf{1} = (1, 1, \ldots, 1)^T$ 是 P^T 的一个特征向量，其特征值为 1，因为 P^T (的每一行以及 P) 的每一列之和为 1：

$$P^T \mathbf{1} = \mathbf{1}$$

由于P和P^T的特征值相同，这证明1也是P的一个特征值。

对于谱半径的界，考虑 P 的任意特征值 λ，其对应的特征向量为 $\boldsymbol{v}$。令 i 为 $|v_i|$ 取得最大值的索引。则：

$$\lambda v_i = \sum_{j=1}^{n} p_{ij} v_j$$

取绝对值：

$$|\lambda||v_i| = \left|\sum_{j=1}^{n} p_{ij} v_j\right| \leq \sum_{j=1}^{n} p_{ij}|v_j| \leq |v_i| \sum_{j=1}^{n} p_{ij} = |v_i|$$

由于$|v_i| > 0$，我们有$|\lambda| \leq 1$，从而确立$\rho_P = 1$。第4点中陈述的特征值1的性质是将佩龙–弗罗贝尼乌斯理论适用于随机矩阵所得的标准结果。

□

满足 $P\boldsymbol{\pi} = \boldsymbol{\pi}$ 的分布 $\boldsymbol{\pi}$ 称为 *stationary*——它在转移规则下保持不变。对于不可约链，这样的分布是唯一的。对于遍历链，从任何初始分布出发的迭代都会收敛到这一平稳状态：

定理 9.15（马尔可夫收敛）。 *Let P be an ergodic stochastic matrix (i.e., irreducible and aperiodic). Then:*

1. *There exists a unique probability vector $\boldsymbol{\pi}$ such that $P\boldsymbol{\pi} = \boldsymbol{\pi}$. This $\boldsymbol{\pi}$ is called the* 平稳分布.
2. *For any initial probability vector $\boldsymbol{x}(0)$:*

$$\lim_{k\to\infty} P^k \boldsymbol{x}(0) = \boldsymbol{\pi}$$

3. *The rate of convergence is governed by the magnitude of the second-largest eigenvalue (i.e., the largest $|\lambda|$ such that $|\lambda| < 1$).*

如果 P 是不可约的但周期性的，$P^k\boldsymbol{x}(0)$ 不会收敛到一个单一的向量，而可能表现出周期性极限，或者在Cesaro均值下收敛到 $\boldsymbol{\pi}$。

例 9.16（天气平衡）。回到我们的天气模型，矩阵 P 是不可约的（所有元素都是正数）且非周期性的（例如，$p_{11} > 0$）。求解 $P\boldsymbol{\pi} = \boldsymbol{\pi}$，在 $\sum \pi_i = 1$ 的约束下，得到稳态分布 $\boldsymbol{\pi} \approx (0.455, 0.295, 0.250)^T$。因此，从长远来看，无论初始条件如何，我们预计约 45.5% 的晴天，29.5% 的多云天，和 25% 的雨天。P 的特征值为 $1, \approx 0.422, \approx 0.178$。第二大特征值的绝对值 ≈ 0.422 表明该平衡状态将迅速收敛。

◇

定理 9.15 保证的收敛性解释了许多自然现象。使用随机矩阵的种群遗传学模型预测大种群中的等位基因频率。品牌忠诚度的经济模型预测市场份额的演变。网页排名通过用户浏览模式的随机游走模型得出。在每种情况下，概率守恒的数学原理通过特征结构塑造了长期行为。

例 9.17（种群遗传学）。考虑一个在单一遗传位点上具有两个等位基因 A 和 a 的种群。父代中等位基因 A 的频率 p 决定了在随机交配下，通过过渡矩阵得到的后代基因型概率 (AA，Aa，aa)：

$$P = \begin{bmatrix} p^2 & p^2 & p^2 \\ 2p(1-p) & 2p(1-p) & 2p(1-p) \\ (1-p)^2 & (1-p)^2 & (1-p)^2 \end{bmatrix}$$

其中 Aa 和 aA 代表相同的杂合基因型。该矩阵具有相同的列——每个后代的基因型仅依赖于总体等位基因频率，而不依赖于父母的具体基因型。唯一的平稳分布给出了哈迪-温伯格平衡比例（p^2，$2p(1-p)$，$(1-p)^2$），由于 P 的秩为1结构，这一平衡在仅一代内就得以实现。

◇

随机矩阵中的特殊结构揭示了额外的特性。若一个矩阵 P 的列和行之和都为 1，则称其为 *doubly stochastic* —— 概率在前向和后向时间中都得到保持。此类矩阵自然出现在某些物理系统中，其中的跃迁会守恒某种潜在量。它们的平稳分布必须是均匀的：对于所有 i，$\pi_i =$ 为 $1/n$。

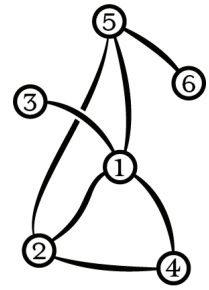
示例 9.18（图上的随机游走）。考虑一个粒子在一个无向图上随机移动，该图有 n 个顶点，每一步它以相等的概率移动到任何相邻的顶点。转移概率形成一个随机矩阵 $P = [p_{ij}]$，其中

$$p_{ij} = \begin{cases} 1/\deg(j) & \text{if vertices } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

在这里，$\deg(j)$ 表示顶点 j 的度数（邻居数）。这个选择使得 P 成为列随机矩阵，尽管不一定是对称的。然而，由于图的无向性质，关系 $\deg(j)p_{ij} = \deg(i)p_{ji}$ 成立。

例如，考虑如图所示的一个包含六个顶点的图。其转移矩阵为：

$$P = \begin{bmatrix} 0 & 1/3 & 1 & 1/2 & 1/3 & 0 \\ 1/4 & 0 & 0 & 1/2 & 1/3 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/3 & 0 & 0 & 0 & 0 \\ 1/4 & 1/3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1/3 & 0 \end{bmatrix}$$



如果该图是连通的且不是二分图（这对这种随机游走意味着非周期性），则该马尔可夫链是遍历的。平稳分布的各个分量与顶点度数成正比：$\pi_1 = 4/14$，$\pi_2 = 3/14$，$\pi_3 = 1/14$，$\pi_4 = 2/14$，$\pi_5 = 3/14$，$\pi_6 = 1/14$。这种非均匀分布反映了随机游走在高阶（高连接度）顶点上停留时间更长的事实，这一原理是许多网络中心性度量的基础。特别地，注意到度数为 4 的顶点 1 获得了平稳概率中最大的份额。$\diamond$

该理论自然地扩展到无限状态空间，尽管在技术上需要谨慎处理。无限图上的随机游走引出了关于常返性与瞬逝性的丰富理论，而尺度极限则导向布朗运动和随机微分方程。贯穿始终，概率守恒与矩阵结构之间的相互作用引导我们理解随机性如何通过线性变换演化为规律性。

9.5 *Symmetric Matrices & Spectra*

数学常常通过结构外在形式与内在本质之间的关系显现出来。正如基因组编码有机体的形状与功能一样，矩阵的某些形态学特征——其数值阵列中的外在模式——指示着隐藏的谱性质，而这些性质决定了这种根本性的行为。简单的结构性条件——对称性（$A^T = A$）——会产生具有深远意义的谱后果。

例如，考虑这些矩阵

$$\begin{bmatrix} 4 & 1 & 2 \\ 1 & 3 & -1 \\ 2 & -1 & 5 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 3 & -2 & 0 \\ -2 & 5 & -1 \\ 0 & -1 & 4 \end{bmatrix}$$

尽管它们的条目差异明显，但它们都具有关于对角线的对称性这一关键特征。这种可见的结构——类似于蝴蝶翅膀的双侧对称——揭示了更深层的模式：这两个矩阵共享的谱特性并非源自具体的数字，而是源自对称模式本身。

定理 9.19（谱定理）。*Let A be a real symmetric matrix. Then:*

1. *All eigenvalues of A are real*
2. *Eigenvectors corresponding to distinct eigenvalues are orthogonal*
3. *A has an orthonormal basis of eigenvectors*

Thus A can be orthogonally diagonalized: $A = Q\Lambda Q^T$ where Q is orthogonal and Λ is diagonal with real entries.

这一显著的结果——对称性强制了特征值的现实性和特征向量的正交性——将抽象矩阵转化为具体的几何对象。每个对称矩阵通过沿着其特征向量所确定的垂直轴伸缩空间来作用。特征值衡量这种形变的尺度，提供了一种坐标无关的变换作用描述。

与对称A相关的二次型揭示了该结构的另一面：

$$q(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}$$

这个标量函数衡量了一种广义的“能量”在方向 $\boldsymbol{x}$ 上。尽管比物理能量更抽象，但这一度量有助于我们理解对称矩阵如何塑造它们作用的空间。以下示例阐明了在不同应用中结构与谱之间的联系。

示例 9.20（相关结构）。给定 n 个 d 变量的测量值，它们的相关矩阵 $[R] = [\text{corr}_{ij}]$ 记录了变量对之间的标准化关系：

$$\text{CORR}_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2}}$$

这个矩阵自然是对称的 ($\text{corr}_{ij} = \text{corr}_{ji}$)，对角线上的元素为 1 ($\text{corr}_{ii} = 1$)。它的谱结构揭示了数据中的基本模式：

- 特征值衡量相关性模式的强度
- 特征向量用于识别相关变量的群组
- 较小的特征值表明存在冗余或依赖关系

例如，在金融数据中，特征向量通常将市场部门分开，而特征值则衡量部门整体与公司特定的变化。在基因表达数据中，特征向量可能识别共调控基因，而特征值则量化调控模式的强度。◇

例9.21（刚体力学）。复杂三维物体的质量可以通过其惯性矩阵（质量加权的）协方差矩阵来表征：

tensor

在力学中，这通常被称为*inertia*。不对称的有质量物体具有三个彼此正交的“自然”转动轴这一直观事实，是谱定理的结果。

$$\mathcal{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix}$$

其中各元素度量物体质量或点分布的二阶矩。$\mathcal{I}$ 的特征结构提供了一种与坐标无关的描述：

- 特征向量给出形状的主轴
- 特征值衡量沿这些轴的空间范围
- 特征值的比值量化了偏离球对称性的程度

这种谱指纹使得形状匹配与分类成为可能，而无需对对象进行显式对齐。一个饮用玻璃杯具有一个较大的特征值（沿其长度方向）以及两个较小且相等的特征值（沿其圆形横截面方向）；一本书具有三个彼此不同的特征值，反映了其矩形比例。

◇

Think: 仅给定点与点之间的成对距离，我们能否重建它们的相对位置？这一逆问题类似于根据原子之间的测量来推断分子的形状，或根据个体之间的相似性来推断社会网络的结构。

示例 9.22（距离矩阵）。考虑一组只有彼此距离已知的 n 抽象点。平方距离形成一个对称矩阵 $D = [d_{ij}]$，其中 d_{ij} 代表平方

点 i 与 j 之间的距离。尽管我们无法直接将这些点可视化，它们的几何结构却隐藏在 D 之中。

居中矩阵 $H = I - \frac{1}{n}[1]$（其中 $[1]$ 表示全 1 矩阵）以及派生矩阵 $B = -\frac{1}{2}HDH$ 起着至关重要的作用。尽管 D 本身可能不是正定的，B 却是对称半正定的，并编码了点云的几何本质：

- 正特征值揭示嵌入维度
- 特征向量重构相对位置
- 谱衰减表明内在维数

这种从距离中提取坐标的过程，被称为 *classical multidimensional scaling*，展示了特征结构如何揭示抽象数据中隐藏的几何关系。

◇

与对称矩阵相关的二次型的极值性质为其结构提供了另一种视角：

引理 9.23（极值）。*For symmetric A, the quadratic form $q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ achieves:*

1. Maximum value $\lambda_{\text{最大}}$ in direction of largest eigenvalue 2. Minimum value $\lambda_{\text{最小}}$ in direction of smallest eigenvalue 3. Intermediate values between $\lambda_{\text{最小}}$ and $\lambda_{\text{最大}}$ elsewhere on the unit sphere $\|\boldsymbol{x}\| = 1$.

这一优化原理解释了为什么许多迭代过程会收敛到特征向量。例如，用于计算主特征值的幂法便是自然而然的结果：反复与 A 相乘会放大沿着最大特征值方向的分量。

像 Lanczos 迭代这样更复杂的算法利用对称性与正交性之间的联系，高效地计算多个特征值。

示例 9.24（投资组合优化）。考虑为 n 个资产选择投资权重 $\boldsymbol{w} = (w_1, \ldots, w_n)^T$，其中 w_i 代表投资于资产 i 的财富比例。这些权重必须满足两个基本约束：它们的总和必须为一（完全投资可用资金），并且通常必须是非负的（不允许卖空）：

$$\sum_{i=1}^{n} w_i = 1 \quad \text{and} \quad w_i \geq 0$$

该投资组合的风险取决于资产收益如何共同变动，这由它们的 *covariance matrix* $[C] = [C_{ij}]$ 所刻画。每个条目 C_{ij} 衡量资产 i 和 j 的收益如何共同变化——正值表示它们往往同涨同跌，而负值则表明走势相反。

这个协方差矩阵自然是对称的 ($C_{ij} = C_{ji}$)，将我们的金融问题与对称矩阵的理论联系起来。总投资组合风险，通过其方差来衡量，呈现出二次表达式 $\boldsymbol{w}^T[C]\boldsymbol{w}$ 的形式。在最简单的表述中，我们的优化问题变为：

$$\begin{array}{ll} \text{minimize} & \boldsymbol{w}^T[C]\boldsymbol{w} \\ \text{subject to} & \boldsymbol{w}^T\mathbf{1} = 1 \end{array}$$

其中 1 表示全为 1 的向量。尽管其外观简单，但当通过拉格朗日方法分析时，这一约束最小化展现出惊人的深度：

Recall: 为最小化 $F(\boldsymbol{x})$，在约束 $G(\boldsymbol{x}) = 0$ 下，拉格朗日函数 $\mathcal{L}(\boldsymbol{x}, \lambda) = F(\boldsymbol{x}) - \lambda G(\boldsymbol{x})$ 导出方程 $\nabla_{\boldsymbol{x}}\mathcal{L} = 0$ 和 $G(\boldsymbol{x}) = 0$。这里 $F(\boldsymbol{w}) = \boldsymbol{w}^T[C]\boldsymbol{w}$，以及 $G(\boldsymbol{w}) = \boldsymbol{w}^T 1 - 1$。

$$\mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{w}^T[C]\boldsymbol{w} - \lambda(\boldsymbol{w}^T\mathbf{1} - 1)$$

令 $\nabla_{\boldsymbol{w}}\mathcal{L} =$ 为 0，则得到:

$$2[C]\boldsymbol{w} - \lambda\mathbf{1} = \mathbf{0}$$

与约束方程一起，这个系统确定了最优权重：

$$\begin{array}{rcl} \boldsymbol{w} & = & \frac{1}{2}[C]^{-1}\mathbf{1}\lambda \\ 1 & = & \boldsymbol{w}^T\mathbf{1} = \frac{1}{2}\lambda\mathbf{1}^T[C]^{-1}\mathbf{1} \end{array}$$

该解揭示了一种基本关系：

Think: 约束 $\boldsymbol{w}^T 1 = 1$ 定义了 $\mathbb{R}^n$ 中的一个超平面。解出现在二次型的水平集与该超平面相切的地方——这是一个几何图像，解释了为什么约束优化常常产生如此优雅的解。

$$\boldsymbol{w} = \frac{[C]^{-1}\mathbf{1}}{\mathbf{1}^T[C]^{-1}\mathbf{1}}$$

这种转变——从在投资约束下最小化风险到解决线性系统——体现了一个更深层次的模式。许多涉及对称矩阵的约束优化问题通过拉格朗日乘数法转化为这样的系统。[C]的特征结构决定了最优投资组合权重和可实现的最小风险。在实际操作中，这种分析自然地扩展到通过马科维茨均值-方差框架来包括预期收益，其中风险最小化通过仔细利用协方差结构来平衡收益最大化。

◇

这种结构对称性与谱特性之间的亲密关系 exemplifies 数学中的一个更深层次的模式：形态学约束通常编码谱的本质。就像一个生物体的形态反映了它的基因组一样，一个矩阵的可见模式编码了决定其基本行为的隐藏属性。这个主题将在我们探讨网络科学中的结构化矩阵并在第10章发展奇异值分解时进一步深化。

9.6 *Networked Behavior & Consensus*

影响在网络中的流动塑造了从观点形成到经济行为再到人工智能的一切。当网络中的代理基于其邻居的数值更新自身状态时，简单的局部规则也能涌现出复杂的全局模式。理解这种集体行为需要将第 9.3–9.5 节的谱理论与交互网络的具体拓扑结构相结合。

定义 9.25（图拉普拉斯）。对于一个具有 n 个顶点的无向图，*graph Laplacian* $L = [L_{ij}]$ 是一个 $n \times n$ 矩阵，其条目为：

$$L_{ij} = \begin{cases} d_i & \text{if } i = j \\ -1 & \text{if vertices } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

其中 d_i 表示顶点 i 的 *degree*——与其相连的边的数量。等价地，若 $D = \mathrm{diag}(d_1, \ldots, d_n)$ 是 *degree matrix*，且 $A = [a_{ij}]$ 是 *adjacency matrix*，其中相连的顶点取 $a_{ij} =$ 为 1，否则为 0，则 $L = D - A$。

•

这个矩阵虽然定义简单，但通过其谱携带着关于网络拓扑的深刻信息。它对向量的作用度量了量如何沿着边扩散，使其成为理解集体动力学的基础。

考虑一个由 n 个智能体组成的网络，每个智能体持有一个实值 $x_i(t)$，该值通过局部平均在离散时间中演化：

$$x_i(t+1) = \sum_{j \in N(i)} w_{ij} x_j(t)$$

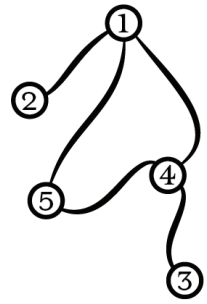
其中 $N(i)$ 表示代理 i 的邻居，权重 w_{ij} 表示交互强度。将状态向量记为 $\boldsymbol{x}(t)$，这一局部更新规则可写为矩阵迭代：

$$\boldsymbol{x}(t+1) = W\boldsymbol{x}(t)$$

权重矩阵 $W = [w_{ij}]$ 通常采用 $W = D^{-1}A$ (等权重的邻居) 或 $W = I - \epsilon L$ (基于分歧的梯度下降) 的形式，这与我们在第9.4节对随机矩阵的研究相联系。

例9.26（意见动力学）。考虑一个政治话语网络　其中在社交媒体上有五个大致的人群群体，其连接关系如图所示（这些不是个体，而是相互作用的

具有汇总意见的群体）。第1组和第4组处于最具影响力的位置，各自与另外三个群体相连。每个群体都会基于其邻居观点的加权平均来更新对一项政治提案的意见：



$$W = \begin{bmatrix} 1/4 & 1/2 & 0 & 1/4 & 1/3 \\ 1/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/4 & 0 \\ 1/4 & 0 & 0 & 1/4 & 1/3 \\ 1/4 & 0 & 1/2 & 1/4 & 1/3 \end{bmatrix}$$

W 的列随机性确保每一次新的意见分布都保持为一个规范的概率分布。从两极化的初始意见 $\boldsymbol{x}(0) = (1, -1, -1, 1, -1)^T$ 出发，迭代表明系统稳定地收敛于共识，其中连接度更高的群体通过其多条影响路径对最终取值施加更大的影响。$\diamond$

当网络是连通的（如第9.3节所述，矩阵 W 是不可约的），并且权重满足某些平衡条件时，会出现一种显著的现象：所有智能体都会收敛到一致。这种收敛性直接源自第9.3节的佩龙–弗罗贝尼乌斯理论，但从网络的视角可以揭示拓扑结构如何塑造达成一致的路径。

定理 9.27（网络共识）。 *Let W be an $n \times n$ column-stochastic matrix that is also* 遍历的 *(i.e., irreducible and aperiodic). Let $\boldsymbol{\pi}$ be the unique stationary probability distribution for W (the eigenvector corresponding to eigenvalue $\lambda_1 = 1$, normalized such that $\sum_i \pi_i = 1$). Then for any initial state vector $\boldsymbol{x}(0) \in \mathbb{R}^n$:*

$$\lim_{t\to\infty} W^t \boldsymbol{x}(0) = \left(\sum_{i=1}^{n} x_i(0)\right) \boldsymbol{\pi}$$

项 $\sum x_i(0)$ 是初始状态的总和。如果 $\boldsymbol{x}(0)$ 本身是一个概率分布，则该和为 1，系统收敛到 $\boldsymbol{\pi}$。如果 W 仅是不可约且周期性的，$W^t\boldsymbol{x}(0)$ 通常不会收敛到单一向量。

Moreover, the rate of convergence to this consensus state (or its scaled version) is determined by $|\lambda_2|$, where λ_2 is an eigenvalue of W with the second-largest magnitude (so $|\lambda_2| < 1$). The convergence is geometric, with the error typically decreasing by a factor of approximately $|\lambda_2|$ at each step.

该定理通过谱性质将网络结构与集体行为联系起来。$\boldsymbol{\pi}$ 的各个分量表示最终一致状态中每个节点的相对影响，而 $|\lambda_2|$ 衡量一致性出现的速度。具有良好混合性质的高度连通网络具有较大的谱隙并且收敛迅速，正如我们在第 9.4 节对随机矩阵所见。

示例 9.28（资产价格形成）。考虑一个市场网络，在这个网络中，交易者根据他们交易伙伴的价格调整对资产的估值。权重矩阵 W 反映了各方之间的信任关系和交易量。产生的共识价格代表了市场效率的一种形式，连接良好的交易者（具有较高 π_i 值）对价格发现有更大的影响。谱隙 $1 - |\lambda_2|$ 决定了市场达到这一均衡的速度——这是理解市场稳定性和对冲击反应的关键。
◇

例 9.29（机器人群集）。在平面内运动的一群机器人可以通过对速度进行局部平均来实现协调运动。每个机器人 i 具有位置 $\boldsymbol{p}_i$ 和速度 $\boldsymbol{v}_i$，其更新规则如下：

$$\boldsymbol{v}_i(t+1) = \boldsymbol{v}_i(t) + \epsilon \sum_{j \in N_i(t)} (\boldsymbol{v}_j(t) - \boldsymbol{v}_i(t))$$

其中 $N_i(t)$ 包含位于机器人 i 通信范围内的机器人。这可以写成具有时变拉普拉斯矩阵的迭代形式：

$$\boldsymbol{v}(t+1) = (I - \epsilon L(t))\boldsymbol{v}(t)$$

$L(t)$ 的谱同时决定了群集行为的稳定性以及向一致运动收敛的速率。该示例展示了为静态网络开发的工具如何自然地扩展到动态图拓扑。

◇

除了简单的平均之外，图拉普拉斯算子通过其与网络结构的联系，使得更为复杂的分析成为可能。对于一个连通图：

1. 零特征值是单的，其特征向量为 1 2. 最小的非零特征值衡量连通性 3. 相应的特征向量揭示自然的网络簇

这些性质源自第9.5节所研究的对称性，为工程化集体行为提供了分析工具和设计原则。

例9.30（供应链网络）。回到第9.1节的投入产出模型，拉普拉斯谱揭示了对中断的脆弱性。与费德勒向量（最小非零特征值对应的特征向量）分量相对应的行业在压力下往往最先发生分裂，从而识别出经济网络中的自然断裂带。该特征值的大小量化了网络对这类分裂的稳健性。

◇

这些原理在技术与自然界中的现代应用比比皆是：

- 社交网络塑造舆论形成和信息传播
- 金融网络传播冲击并决定系统性风险
- 机器人群体通过局部互动进行协调
- 供应链在效率与韧性之间取得平衡

在每种情况下，网络结构与谱特性之间的相互作用决定了系统层面的行为。

我们开发的工具——从随机矩阵到对称谱——在网络动力学的研究中融合在一起。这一综合揭示了局部规则和全局拓扑如何结合产生集体智能，并为接下来的章节指向更深层的数学和实际应用。本章的最终应用将探讨一个特别优雅的应用：从这些原理发展而来的PageRank算法，用于组织早期的网络。

Spectral Graph Theory & Vibrations

音乐器具的数学来源于弦和膜的振动。被拨动的弦会形成驻波，其频率决定了音高，而鼓膜则以复杂的模式振动，创造出其特有的音色。尽管这些物理系统看起来与本章中发展起来的矩阵理论相距甚远，但它们的数学本质体现在 *graph Laplacians* 的特征值中——这一联系通过谱理论的深层原理，将离散网络分析转化为连续的谐波。

首先考虑一根两端固定的弦，将其离散化为由相同弹簧连接的 n 个点。点 i 相对于平衡位置的位移 x_i 会感受到来自其相邻点的回复力，其大小与位移差成正比：

$$m\frac{d^2x_i}{dt^2} = \kappa(x_{i+1} - x_i) - \kappa(x_i - x_{i-1})$$

其中 m 是每个点的质量，κ 为弹簧常数。用矩阵形式表示：

$$m\frac{d^2\boldsymbol{x}}{dt^2} = -\kappa L\boldsymbol{x}$$

其中 L 是 *graph Laplacian*：

$$L = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Note: 图拉普拉斯算子是对称的半正定矩阵，其特征值直接决定振动频率。

L 的特征向量表示驻波模式——一种振动模态，其中所有点以相同的频率但不同的振幅作正弦运动。相应的特征值 λ_i 通过 $\omega_i = \sqrt{k\lambda_i/m}$ 决定这些频率。随着 n 的增加，这些离散模态收敛到连续解：

$$v_k(x) = \sin\left(\frac{k\pi x}{L}\right), \quad \omega_k = \frac{k\pi}{L}\sqrt{\frac{T}{\rho}}$$

Think: 离散特征模态向连续特征模态的收敛揭示了图论如何自然地扩展到连续系统。

其中，L 为弦长，T 为张力，ρ 为线密度。

这个图谱与物理振动之间的联系扩展到更高的维度。对于由三角网格近似的鼓面，关注网格的顶点和边，并将图拉普拉斯算子定义为顶点集V上的方阵：

$$L_{ij} = \begin{cases} \deg(i) & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

其中 $\deg(i)$ 计算与顶点 i 相邻的顶点数量，而 $i \sim j$ 表示相邻顶点。特征值再次决定振动频率，而特征向量描述模态形状——在随时间振荡时保持其形态的位移模式。

Example: 圆形鼓膜的振型形成了非凡的 *Bessel functions*，作为网格特征向量的极限而出现。

一个显著的结果将这些谱与纯粹的图论联系起来：拉普拉斯算子的零特征值的重数等于连通分量的数量。这揭示了一种优雅的对偶性——正如低特征值描述缓慢的振动一样，它们也刻画了基本的拓扑性质。第二小的特征值衡量图的连通性程度，数值越大表示连通性越强，正如在我们对随机矩阵的研究中，主导特征值决定了收敛速率。

Nota bene: 特征值与图连通性之间的联系，与本章前文中主导特征值控制网络收敛性的方式相呼应。

谱方法在图分析中的应用自然地扩展到划分问题中，即我们希望将顶点划分为若干组，使组内连接密集而组间连接稀疏。与第二小特征值对应的特征向量为这一离散问题提供了一个连续松弛：根据各顶点在该特征向量中对应分量的符号来进行划分。这种谱聚类方法往往能够揭示网络中的自然社区结构。

对于加权图，其中边具有不同的强度，定义度矩阵 $D = \mathrm{diag}(d_1, \ldots, d_n)$，其中 $d_i = \sum_j w_{ij}$ 表示与顶点 i 相连的边权重之和。归一化拉普拉斯矩阵 $\mathcal{L} = D^{-1/2}LD^{-1/2}$ 在保持对称性的同时考虑了这些不同的连接强度。其特征值位于 $[0, 2]$ 区间内，提供了一种不依赖于绝对边权的图结构的归一化度量。

离散图与连续振动之间的深刻联系揭示了数学物理中的一种深刻统一性。特征值和特征向量同时编码了物理振荡和抽象图的性质，而拉普拉斯算子则提供了结构与动力学之间的桥梁。这种合成——通过谱理论将离散与连续结合——展示了基本数学原理如何阐明看似不相关的现象。

PageRank: The Flow of Web Authority

万维网或许是人类历史上构建的最大网络——数十亿个页面通过超链接相互连接，在数字空间中引导注意力与信息的流动。与第9.1节研究的离散时间系统类似，这一庞大网络呈现出内在的信息流动模式，这些模式可以通过严谨的数学分析来理解。对这一空间进行组织的挑战，促成了随机矩阵与迭代收敛最为优雅的应用之一：谷歌的 PageRank 算法。

考虑一名在网页之间沿着链接浏览的网络冲浪者，将其行为建模为第 9.4 节中提出的那类马尔可夫链。在每一步中，他们要么以概率 α 跟随一条随机选择的外出链接，要么以概率 $1-\alpha$ 跳转到网络中的任意一个随机页面。这一过程生成一个 *transition matrix* $P = [p_{ij}]$：

$$p_{ij} = \alpha \frac{a_{ij}}{d_j} + \frac{1-\alpha}{n}$$

其中，$a_{ij} =$ 在页面 j 链接到页面 i (时为 1，否则为 0)，d_j 是页面 j (的外链数量；如果 $d_j =$ 为 0，该项通常通过假设等概率跳转到所有页面来处理)，并且 n 是页面总数。项 $(1-\alpha)/n$ 表示随机传送的概率。

Historical Note: 加入随机传送（通常 $\alpha \approx 0.85 < 1$ ）确保当 $n > 0$ 时，P 是一个 *primitive* 矩阵：其所有元素都严格为正。一个本原随机矩阵必然是 *ergodic* (不可约且非周期的)，从而保证第 9.4 节中研究的收敛性质。

这种构造使 P 成为一个列随机矩阵。由于 $0 < \alpha < 1$ 且 $n > 0$，所有元素 p_{ij} 都严格为正。严格为正的随机矩阵是遍历的。应用第 9.4 节中发展出的理论（具体而言是定理 9.15），我们知道该矩阵 P 必然具有唯一的平稳概率分布 $\boldsymbol{\pi}$，满足 $\boldsymbol{\pi} = P\boldsymbol{\pi}$。该 *PageRank vector* $\boldsymbol{\pi}$ 通过其长期访问概率来衡量每个页面的重要性。

正如第9.6节中研究的共识问题一样，$\boldsymbol{\pi}$ 的计算自然采用迭代方法（幂方法）。从任意初始概率分布 $\boldsymbol{x}_0$ (（通常为均匀分布）) 出发，通过对 P 的反复乘法会收敛到 $\boldsymbol{\pi}$:

$$\boldsymbol{x}_{k+1} = P\boldsymbol{x}_k \quad \text{and} \quad \lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{\pi}$$

P 的性质确保每一次迭代都保持为概率分布，并且会收敛到唯一的极限 $\boldsymbol{\pi}$。

例 9.31（简单网络）。考虑一个由四个页面组成的微小网络，其链接结构由邻接矩阵 A 给出。在阻尼因子 $\alpha =$ 为 0.85 的情况下，迭代收敛到：

$$\boldsymbol{\pi} \approx \begin{pmatrix} 0.17 \\ 0.31 \\ 0.35 \\ 0.17 \end{pmatrix}$$

该分布同时反映了局部链接结构和全局网络位置。与第 9.6 节中的网络中心性度量类似，具有更多入路径的页面往往会获得更高的得分。$\diamond$

收敛速率与所有遍历的马尔可夫链一样，由 P 的第二大特征值的大小决定，记为 $|\lambda_2(P)|$。该

P 的特征值与归一化邻接矩阵部分的特征值相关，其关系为 $\lambda_k(P) = \alpha\lambda_k(M_{norm}) + (1-\alpha)/n$，适用于对应于 $\lambda_k(M_{norm})$ 的特征向量在其与全 1 向量正交时，且 $\lambda_1(P) = 1$。因此，$|\lambda_2(P)| \approx \alpha|\lambda_2(M_{norm})|$。对于典型的 Web 图以及 $\alpha \approx 0.85$，$|\lambda_2(P)|$ 通常接近于 α。

例 9.32（收敛行为）。对于我们的四页面示例，跟踪连续迭代揭示了几何收敛：

$$\|\boldsymbol{x}_k - \boldsymbol{\pi}\| \approx |\lambda_2(P)|^k \|\boldsymbol{x}_0 - \boldsymbol{\pi}\|$$

其中 $|\lambda_2(P)| < 1$。对于 $\alpha = 0.85$，该值通常接近 0.85，从而实现相当快的收敛。这与我们的马尔可夫链分析所预测的行为一致。$\diamond$

网络图为第9.6节所研究的网络结构提供了一个完美示例，在那里拓扑结构塑造着动态行为。PageRank 的成功源于它将本章的两个关键原则结合在一起：遍历随机矩阵的收敛性，以及影响在网络中的流动。这种综合——将马尔可夫链理论与网络动力学相融合——通过纯数学改变了网页搜索。

该框架可自然地扩展到其他重要性沿网络边传播的情境，只要构造能够确保唯一收敛的遍历转移矩阵：

- 学术论文引文网络排名
- 衡量用户影响力的社交网络
- 揭示系统重要性的经济网络

PageRank 例证了本章所发展的数学如何塑造现代世界。我们所研究的随机矩阵、网络动力学和迭代方法通过严谨的理论相结合，将混乱的网络加以组织。这种将抽象数学转化为实用计算的过程，展示了严谨的数学分析在当代工程中的深远效用。

□ ——————————————————————————— □

Exercises: Chapter 9

1. 设 $\begin{bmatrix} 0&1&0&1&0&1&0&1&0 \end{bmatrix}$。计算 A 的前五次幂，并描述你观察到的任何模式。这对对应图上的随机游走说明了什么？ 2. 对于转移矩阵 $\begin{bmatrix} 0.5&0.3&0.2&0.4&0.4&0.3&0.1&0.3&0.5 \end{bmatrix}$，计算 P^2 和 P^3。P 的幂似乎在发生什么变化？请使用随机矩阵理论解释这种行为。 3. 考虑随机矩阵 $P = \begin{bmatrix} 0.8&0.2&0.3&0.7 \end{bmatrix}$。求其平稳分布。是否

这个分布是否唯一？请使用佩龙–弗罗贝尼乌斯理论解释原因。

4. 设 G 为一个无向图，其顶点为 $\{1, 2, 3, 4\}$，以及边

$$E = \{(1,2), (2,3), (3,4), (4,1), (2,4)\}$$

写出它的邻接矩阵 A 和度矩阵 D；然后计算图拉普拉斯矩阵 $L = D - A$ 并求其特征值。这些结果告诉你关于该图连通性的什么信息？

5. 一名赌徒以 $2 开始，玩一个游戏：以 $1/3 的概率赢得 $1，以 $2/3 的概率输掉 $1。当他们要么破产（$0），要么达到目标 $3 时就停止游戏。将其建模为一个具有状态 {0,1,2,3} 的马尔可夫链。求转移矩阵，并计算最终达到目标状态的概率。

6. 一个儿童玩具有三个按钮：红色、蓝色和绿色。按下任意一个按钮时，都会播放一段旋律，并按照转移矩阵随机转移到点亮另一个按钮（也可能仍然是同一个按钮）。

$$P = \begin{bmatrix} 0.2 & 0.4 & 0.3 \\ 0.5 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.3 \end{bmatrix}$$

如果首先按下红色按钮，那么在恰好三次按压之后绿色按钮亮起的概率是多少？在经过大量按压之后，每个按钮亮起的时间占比各是多少？

7. 考虑一个具有三个部门的投入—产出经济模型，其中 input 矩阵是 $A = \begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.2 \\ 0.4 & 0.3 & 0.1 & 0.2 \\ 0.4 & \end{bmatrix}$ 找到平衡生产水平。Perron-Frobenius 定理告诉你关于这个经济体稳定性的什么信息？

8. 设 A 为一个简单无向图的邻接矩阵。证明 A^k 的第 (i, j) 个元素计数了从顶点 i 到顶点 j 的长度为 k 的行走的个数。

9. 若矩阵 P 的行和与列和均为 1，则称其为双随机矩阵。证明：如果 P 是双随机的，则向量 $1/n$ （（其中 n 是维数 ））必然是一个平稳分布。该分布是否一定唯一？

10. 对于一个连通的无向图，随机游走拉普拉斯矩阵定义为 $L_{rw} = I - D^{-1}A$，其中 D 是度矩阵，A 是邻接矩阵。证明 L_{rw} 总是具有特征值 0，其对应的特征向量为 1。

11. 设 P 为一个不可约随机矩阵。证明如果 P 有一个特征值 λ，且 $|\lambda| =$ 为 1，则 λ 必须是单位根（即，存在某个正整数 k 使得 $\lambda^k =$ 为 1）。

12. 考虑一个由 4 个智能体组成的网络，其中每个智能体根据其邻居意见的加权平均值来更新自己的意见。若该网络是一个正方形（4-环），写出其更新矩阵，并确定意见收敛到一致所需的速度。

13. 设 G 为一张图，L 为其拉普拉斯矩阵。证明特征值 0 的重数等于 G 中连通分量的个数。

14. 对于一个不可约的随机矩阵 P，证明对于任何与 1 垂直的向量 $\boldsymbol{x}$，有 $\|P^n\boldsymbol{x}\| \leq \|\boldsymbol{x}\|$，其中 $\|\cdot\|$ 是标准欧几里得范数。 15. 设 P 是表示马尔可夫链转移概率的随机矩阵。定义 *mean first passage time* m_{ij} 为从状态 i 开始到达状态 j 的期望步数。证明这些时间满足方程 $m_{ij} = 1 + \sum_{k\neq j} p_{ki}m_{kj}$ 对于 $i \neq j$。 16. 考虑一个马尔可夫链，其转移矩阵为 P，假设状态 j 是 *absorb- ing* (，意味着 $p_{jj} = 1$ 且 $p_{ij} = 0$ 对于 $i \neq j$)。证明如果链从状态 $i \neq j$ 开始，最终吸收在状态 j 的概率等于 (i, j) 条目，来自于 $(I - Q)^{-1}R$，其中 Q 是 P，去掉了行和列 j，而 R 是 P 的列 j，去掉了条目 j。
17. 设 P 为不可约马尔可夫链的转移矩阵，$\boldsymbol{\pi}$ 为其稳态分布。*time-reversed* 链的转移矩阵为$\tilde{P}$，其中

$$\tilde{p}_{ij} = \frac{\pi_i}{\pi_j} p_{ji}$$

证明$\tilde{P}$ 是随机的，并且与 P 具有相同的平稳分布。这告诉你链的可逆性有什么信息?

18. 概率分布 $\boldsymbol{\pi}$ 的熵定义为 $H(\boldsymbol{\pi}) = -\sum_i \pi_i \ln(\pi_i)$。对于一个随机矩阵 P，证明如果 $\boldsymbol{\pi}$ 是一个平稳分布，那么 $\boldsymbol{\pi}$ 在所有满足 $P\boldsymbol{x} = \boldsymbol{x}$ 的分布 $\boldsymbol{x}$ 中最大化熵。

19. 考虑一个立方体的图（8个顶点，12条边）。写下它的邻接矩阵并计算它的谱。谱告诉你关于在这个图上进行随机游走的什么信息?

20. 无向图中一个顶点 *clustering coefficient* 的聚类系数定义为该顶点的 i 的邻居对中，彼此相连的比例。给定图的邻接矩阵，写出计算一个顶点聚类系数的公式，公式中使用矩阵运算。应用此公式计算图中每个顶点的聚类系数，邻接矩阵为:

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

21. 网络中节点的 *eigenvector centrality* $\boldsymbol{x}$ 定义为 $\lambda\boldsymbol{x} = A\boldsymbol{x}$ 的解，其中 A 是邻接矩阵，λ 是 A 的谱半径。给定一个由邻接矩阵表示的网络:

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

使用幂迭代法来近似特征向量中心性得分（每次迭代后进行归一化）。这些得分如何与每个节点的度数相关? 解释为什么一些节点的中心性高于其度数所暗示的水平。

Chapter 10
Singular Value Decomposition

"build we the Mundane Shell around the Rock of Albion"

每个变换都蕴藏着隐藏的对称性，存在于其表面复杂性之下。就像埋藏在看似无形岩石中的晶体结构，这些模式只有通过仔细的挖掘和分析才能显现出来。前几章中发展出的特征分解通过迭代作用揭示了方阵的结构。然而，尽管这个工具非常强大，它仅仅揭示了线性变换底层模式的一部分。一个更为基础的分解方法正等待被发掘。

特征理论对方阵的限制并非偶然——特征值和特征向量自然地来源于迭代，而迭代需要一种变换将空间映射到自身。然而，线性变换的几何本质——它如何拉伸和旋转空间——超出了这种自同构的范畴。每一个线性变换，无论是方阵还是矩形矩阵，都有一个标准分解，揭示其内在的几何特征。这个分解不仅揭示了偏好的方向，还揭示了输入空间和输出空间之间的基本关系，这些关系仅仅依靠特征理论是无法看出的。

我们的任务是深入挖掘线性变换的表面结构，发现其中更深层次的模式。我们从矩阵如何将球体变换为椭球体的几何直觉开始，揭示自然的输入和输出方向。在这些基础上，产生了奇异值分解，它为理解线性变换的普遍性提供了理论见解和实用工具。

这种分解的强大之处在于它融合了代数和几何的视角。在坐标系中看似抽象的因式分解，几何上则表现为一系列最优的简单变换。这一最优性原则——即奇异值分解

提供在各种精确意义下的最佳近似——将我们的理论理解转化为数据压缩、信号处理和降维的实用方法。我们在这里开发的工具将对从图像处理到机器学习的现代应用至关重要。

10.1 Spheres, Ellipsoids, & Singular Values

线性变换 $A \in \mathbb{R}^{m\times n}$ 的几何意义可以通过其在定义域 $\mathbb{R}^n$ 的单位球面上的作用生动地揭示出来。单位球面由 $\|\boldsymbol{x}\|^2 = 1$ 或 $\boldsymbol{x}^T\boldsymbol{x} = 1$ 定义，在 A 的作用下被变形为值域 $\mathbb{R}^m$ 中的一个椭球体。理解这种变形是关键。

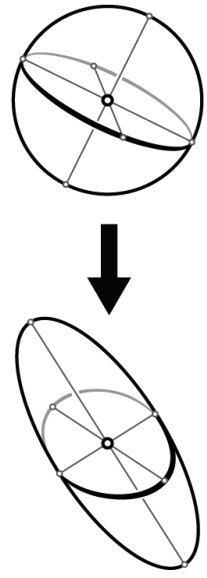
该输出椭球的形状和取向由对称正半定矩阵 A^TA 决定。变换后向量 $A\boldsymbol{x}$ 的平方长度由下式给出：

$$\|A\boldsymbol{x}\|^2 = (A\boldsymbol{x})^T(A\boldsymbol{x}) = \boldsymbol{x}^T(A^TA)\boldsymbol{x}.$$

根据谱定理（定理 9.19），A^TA (作为一个 $n \times n$ 对称矩阵)，具有 n 实的、非负的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$。令 $V = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_n]$ 为一个正交矩阵，其列向量是 A^TA 的相应正交归一特征向量。这些特征向量 $\boldsymbol{v}_i$ 表示输入空间 $\mathbb{R}^n$ 中的主方向。当 $\boldsymbol{x}$ 是这些特征向量之一，例如 $\boldsymbol{x} = \boldsymbol{v}_i$ ，则

$$\|A\boldsymbol{v}_i\|^2 = \boldsymbol{v}_i^T(A^TA)\boldsymbol{v}_i = \boldsymbol{v}_i^T(\lambda_i\boldsymbol{v}_i) = \lambda_i\|\boldsymbol{v}_i\|^2 = \lambda_i.$$

因此，矩阵A将主输入方向$\boldsymbol{v}_i$拉伸了$\sqrt{\lambda_i}$倍。这些拉伸因子是变换A的基础。—



特征值 λ_i 特指 A^TA 的特征值。

定义 10.1（奇异值）。设 $A \in \mathbb{R}^{m\times n}$。$n \times n$ 矩阵 A^TA 是对称且半正定的，因此其特征值 λ_i 为实且非负。A 的 *singular values*，记为 σ_i，是这些特征值的平方根：$\sigma_i = \sqrt{\lambda_i}$。它们通常按降序排列：—

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

非零奇异值的个数是 $r = \text{rank}(A^TA) = \text{rank}(A)$. 对应的 A^TA 的正交归一特征向量 $\boldsymbol{v}_i$ 被称为 A 的 *right singular vectors*.

•

在 A 下，单位球面的像是 $\mathbb{R}^m$ 中的一个椭球（如果 A 秩不足，则可能是退化的）。这个椭球的半轴

与某些向量 $\boldsymbol{u}_i \in \mathbb{R}^m$ 对齐，并且其长度等于非零奇异值 σ_i。$\mathbb{R}^n$ 中的方向 $\boldsymbol{v}_i$ 由 A 映射到这些半轴向量：$A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$。向量 $\boldsymbol{u}_i$ 将是 *left singular vectors* 。

例10.2（图像变换）。考虑这个 2×2 矩阵：

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \quad \Rightarrow \quad A^T A = \begin{bmatrix} 10 & 5 \\ 5 & 5 \end{bmatrix}$$

$A^T A$ 的特征多项式是 $\lambda^2 - 15\lambda + 25 = 0$。其特征值为 $\lambda_1 = (15 + \sqrt{125})/2 \approx 13.09$ 和 $\lambda_2 = (15 - \sqrt{125})/2 \approx 1.91$。奇异值为 $\sigma_1 = \sqrt{\lambda_1} \approx 3.618$ 和 $\sigma_2 = \sqrt{\lambda_2} \approx 1.382$。$\mathbb{R}^2$ 中的单位圆经过 A 变换成一条椭圆，其半轴长度为 σ_1 和 σ_2。这些半轴在 $\mathbb{R}^2$ (the domain) 中的方向由 $A^T A$ 的特征向量给出。$\diamond$

这一几何图像——将主要输入方向（$A^T A$ 的特征向量）映射到按比例缩放的主要输出方向——构成了奇异值分解的直观基础。

10.2 Constructing the SVD

一个线性变换 A 将正交归一的主输入方向映射为相互正交的主输出方向，并按奇异值进行缩放，这一几何直觉直接导向其最基本的分解。我们的目标是找到正交矩阵 U 和 V 以及一个矩形对角矩阵 Σ，使得 $A = U\Sigma V^T$。

Step 1: Finding V and the Singular Values σ_i.

如第 10.1 节所述，矩阵 $A^T A$ 是一个 $n \times n$ 对称正半定矩阵。由谱定理，存在一个 $n \times n$ 正交矩阵 $V = [\boldsymbol{v}_1 \dots \boldsymbol{v}_n]$，其列向量是 $A^T A$ 的正交归一特征向量，以及一个 $n \times n$ 对角矩阵 $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$，其对应的特征值为非负，使得 $A^T A = V \Lambda V^T$。A 的奇异值为 $\sigma_i = \sqrt{\lambda_i}$，按 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ 排序。令 r 为 A 的秩，这也等于非零奇异值的个数。列向量 $\boldsymbol{v}_1, \dots, \boldsymbol{v}_r$ 构成 $(\ker A)^\perp = \operatorname{im}(A^T)$ 的一个正交归一基，而 $\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n$ 构成 $\ker A = \ker(A^T A)$ 的一个正交归一基。

Step 2: Defining the Left Singular Vectors U.

For 每个 $i = 1, \ldots, r$ (其中 $\sigma_i > 0$)，定义向量 $\boldsymbol{u}_i \in \mathbb{R}^m$ 由

$$\boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i.$$

这些 r 向量是正交归一的。为此，考虑它们的内积：

$$\begin{aligned}\boldsymbol{u}_i^T \boldsymbol{u}_j &= \left(\frac{1}{\sigma_i} A\boldsymbol{v}_i\right)^T \left(\frac{1}{\sigma_j} A\boldsymbol{v}_j\right) = \frac{1}{\sigma_i \sigma_j} \boldsymbol{v}_i^T (A^T A)\boldsymbol{v}_j \\ &= \frac{1}{\sigma_i \sigma_j} \boldsymbol{v}_i^T (\lambda_j \boldsymbol{v}_j) \quad (\text{since } \boldsymbol{v}_j \text{ is an eigenvector of } A^T A) \\ &= \frac{\sigma_j^2}{\sigma_i \sigma_j} (\boldsymbol{v}_i^T \boldsymbol{v}_j).\end{aligned}$$

由于 $\{\boldsymbol{v}_k\}$ 是一个正交归一集，$\boldsymbol{v}_i^T \boldsymbol{v}_j = \delta_{ij}$。因此，$\boldsymbol{u}_i^T \boldsymbol{u}_j = \frac{\sigma_j}{\sigma_i} \delta_{ij}$。对于 $i = j$，$\boldsymbol{u}_i^T \boldsymbol{u}_i =$ 为 1。对于 $i \neq j$，$\boldsymbol{u}_i^T \boldsymbol{u}_j =$ 为 0。因此，$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ 是在 $\mathbb{R}^m$ 中的 r 个向量所构成的一个正交归一集。这些向量构成 A 的列空间的一个正交归一基，即 $\text{im}(A)$。

如果 $r < m$，则集合 $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ 不能张成 $\mathbb{R}^m$ 的全部。我们可以通过选择另外 $m - r$ 个正交归一向量 $\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m\}$ 将其扩展为 $\mathbb{R}^m$ 的一组完整的正交归一基，这些向量构成 $(\text{im } A)^\perp = \ker(A^T)$ 的一组基。令 $U = [\boldsymbol{u}_1 \ldots \boldsymbol{u}_r \ldots \boldsymbol{u}_m]$ 为 $m \times m$ 的正交矩阵，其列为这些左奇异向量。

Step 3: Defining the Matrix Σ.

设 Σ 为 $m \times n$ 矩阵，其在 $i = 1, \ldots, p = \min(m, n)$ 时的元素 $\Sigma_{ii} = \sigma_i$，其余所有元素均为零。若 $r < p$，则 $\sigma_{r+1}, \ldots, \sigma_p$ 为零。

在如此构造了 U、Σ 和 V 之后，我们对于 $i = 1, \ldots, r$ 有 $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$，并且对于 $i = r + 1, \ldots, n$ (有 $A\boldsymbol{v}_i = 0$，因为这些 $\boldsymbol{v}_i$ 属于 $\ker A$)。这组向量方程可以写成矩阵形式 $AV = U\Sigma'$，其中 Σ' 是一个 $m \times n$ 矩阵，其前 r 个对角元素为 $\sigma_1, \ldots, \sigma_r$，其余元素均为零。这个 Σ' 正是我们的 Σ。由于 V 是正交的，$V^{-1} = V^T$，因此 $AV = U\Sigma$ 推出 $A = U\Sigma V^T$。

Nota bene: U 的各列同样是 AA^T 的正交归一特征向量，而 AA^T 的非零特征值为 $\sigma_1^2, \cdots\cdots, \sigma_r^2$，与 $A^T A$ 相同。也可以通过对 AA^T 进行对角化来求得 U 以及 σ_i^2。

这种构造导出了中心定理：

定理 10.3（奇异值分解）。 *Every matrix $A \in \mathbb{R}^{m \times n}$ admits a decomposition*

$$A = U\Sigma V^T \tag{10.1}$$

where:

1. $U \in \mathbb{R}^{m\times m}$ *is an orthogonal matrix whose columns are the left singular vectors of* A.
2. $V \in \mathbb{R}^{n\times n}$ *is an orthogonal matrix whose columns are the right singular vectors of* A.
3. $\Sigma \in \mathbb{R}^{m\times n}$ *is a rectangular diagonal matrix, where the diagonal entries* $\Sigma_{ii} = \sigma_i$ *are the singular values of* A*, ordered* $\sigma_1 \geq \sigma 2 \geq \cdots \geq \sigma_p \geq 0$ *with* $p =$ 最小值$\{m,\ n\}$.

The singular values σ_i *are uniquely determined.*

该 SVD 揭示了 A 的基本作用：它将第 i 个右奇异向量 $\boldsymbol{v}_i$ 映射为 σ_i 倍的第 i 个左奇异向量 $\boldsymbol{u}_i$：

$$A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \quad \text{for } i = 1, \ldots, \min(m, n).$$

如果 $\sigma_i = 0$，则 $A\boldsymbol{v}_i = 0$。变换 A 本质上通过以下方式起作用：
1. 旋转/反射输入空间，使基向量 $\boldsymbol{e}_i$ 与 $\boldsymbol{v}_i$ (的 V^T) 作用对齐。 2. 沿新的坐标轴按 σ_i 对这些已对齐的向量进行缩放（Σ 的作用）。 3. 将结果旋转/反射到输出空间，使缩放后的坐标轴与 $\boldsymbol{u}_i$ (的 U) 作用对齐。

SVD 可以说是最重要的矩阵分解，它为理解矩阵的结构、几何、秩以及数值性质提供了深刻的洞见。其应用在后续各节和章节中将予以探讨，范围十分广泛。

10.3 Interpreting the SVD

奇异值分解 $A = U\Sigma V^T$ 所提供的远不止是对矩阵 $A \in \mathbb{R}^{m\times n}$ 的一种代数分解；它揭示了线性变换 A 所表示的内在几何与基本结构。在第 10.2 节中构建了 U、Σ 和 V 之后，我们现在考察它们的含义，以及它们如何与秩、四个基本子空间等核心概念相联系，并理解 A 在其定义域上的作用。

位于 Σ 对角线上的奇异值 σ_i，是该变换沿特定正交方向的拉伸因子或“增益”。最大的奇异值 σ_1，恰好是 A 的 *spectral norm* (或 2-范数)，表示 A 能将任意单位向量拉伸的最大倍数：

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \sigma_1.$$

有序序列 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$（其中 $r = \text{rank}(A)$）表明该变换中不同模式的相对重要性。

这些奇异值的快速衰减，正如我们将在第11章和第12章中看到的那样，表明矩阵A (以及它可能表示的数据)可以很好地由一个较低秩的矩阵来近似。

这是降维和数据压缩技术的基石，如主成分分析 {v*}。

V = 的列 $[\boldsymbol{v}_1 \dots \boldsymbol{v}_n]$ 是 *right singular vectors*。每个 $\boldsymbol{v}_i$ 代表一个 *principal input direction*。当 A 代表数据（例如，行作为观察值，列作为特征）时，这些 $\boldsymbol{v}_i$ 对应于特征空间中的主要方向或固有模式。例如，如果 A 是一个文档-术语矩阵，$\boldsymbol{v}_i$ 可能与潜在主题对齐。

向量 $\boldsymbol{v}_i$ 是定义域中的方向，通过 A 映射到由单位球体变换形成的椭球体的半轴。

$U = [\boldsymbol{u}_1 \dots \boldsymbol{u}_m]$ 的各列是 *left singular vectors*，构成输出空间 $\mathbb{R}^m$ 的一组正交归一基。A 对其右奇异向量的基本作用是将它们映射为其左奇异向量的按比例缩放版本：

$$A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \quad \text{for } i = 1, \dots, \min(m, n).$$

如果 $\sigma_i = 0$，则 $A\boldsymbol{v}_i = 0$。集合 $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}$ (对应于非零 σ_i)，构成 A 的像（或列空间）$\text{im}(A)$ 的正交规范基。这些 $\boldsymbol{u}_i$ 是 *principal output directions* 。在数据上下文中，当 A 的行是观测值时，$U\Sigma$ (的列，特别是 $U_r\Sigma_r$，其中 U_r 和 Σ_r 包含前 r 个分量) 可以解释为转换后数据在主输出方向基中的坐标。

SVD 提供了一个引人注目的几何细化，并为围绕线性代数基本定理（定理 6.9）的概念提供了明确的构造。回想一下，任何线性变换 A : $\mathbb{R}^n \to \mathbb{R}^m$ (使用矩阵表示 T) 都与四个基本子空间相关。SVD 不仅确认了它们的存在及维度关系，还为每个子空间提供了正交规范基。

- *column space* 或 *image* 的 A，$\text{im}(A) \subset \mathbb{R}^m$：左奇异向量 $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}$ 对应于非零奇异值 $\sigma_1, \dots, \sigma_r$ 形成 $\text{im}(A)$ 的正交标准基。因此，$\dim(\text{im } A) = r$。
- A 的 *null space* 或 *kernel*，$\ker(A) \subset \mathbb{R}^n$：对应于零奇异值（如果 $r < n$ ）的右奇异向量 $\{\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n\}$ 构成 $\ker(A)$ 的一组正交规范基。因此，$\dim(\ker A) = n - r$。
- *row space* 的 A，$\text{im}(A^T) \subset \mathbb{R}^n$：自 $A^T = V\Sigma^T U^T$ 起，右奇异向量 $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r\}$ (是 A^T 的左奇异向量，对应于非零 σ_i)，构成了 $\text{im}(A^T)$ 的正交归一基。这个空间也是 $(\ker A)^\perp$。因此，$\dim(\text{im } A^T) = r$。

- *left null space* 的 A，$\ker(A^T) \subset \mathbb{R}^m$：与零奇异值相关的位置所对应的左奇异向量 $\{\boldsymbol{u}_{r+1}, \dots, \boldsymbol{u}_m\}$（若 $r < m$）构成 $\ker(A^T)$ 的一组正交归一基。该空间也等于 $(\text{im } A)^\perp$。因此，$\dim(\ker A^T) = m - r$。

A、r的秩可以立即通过非零奇异值的数量来识别。秩-零度定理，dim(ker A) + dim(im A^T) = n (或$(n-r)+r=n$)，以及其在A^T中的对应定理，因而通过SVD变得显式。

Think: SVD 优雅地将 $\mathbb{R}^n = \text{im}(A^T) \boxplus \ker(A)$ 和 $\mathbb{R}^m = \text{im}(A) \boxplus \ker(A^T)$ 分解，其中 A 在 $\text{im}(A^T)$ 与 $\text{im}(A)$ 之间充当（按 σ_i 缩放的）同构。

SVD 将 A 表示为 r 个秩一矩阵之和，通常称为 SVD 展开：

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T.$$

每一项 $\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$ 都是一个秩一矩阵，表示一个外积。该展开将 A 表示为这些基本秩一“层”的线性组合，并按其对应奇异值 σ_i 的大小排序。这种形式对低秩近似（第12章）至关重要，其中通过仅保留前 k 项对该和进行截断，可得到 $A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}^T$。

$_i$，最佳的 rank-k 近似于 A。

最后，SVD 提供了定义任意矩阵 A 的 *pseudoinverse* $A^\dagger$ 的最通用且稳定的方法，将第 6 章中的概念加以扩展。若 $A = U\Sigma V^T$，其伪逆由下式给出：

$$A^\dagger = V\Sigma^\dagger U^T.$$

这里，$\Sigma^\dagger$ 是一个 $n \times m$ 的矩形对角矩阵。如果 $\Sigma_{ii} = \sigma_i > 0$，则 $(\Sigma^\dagger)_{ii} = 1/\sigma_i$。如果 $\Sigma_{ii} = 0$，则 $(\Sigma^\dagger)_{ii} = 0$。$\Sigma^\dagger$ 的所有非对角元素均为零。

伪逆 $A^\dagger$ 在可能的情况下有效地“反转”A 的作用：

- 它通过“撤销”σ_i的缩放，将向量从im(A)映射回im(A^T)：如果 $\boldsymbol{y} = \sigma_i \boldsymbol{u}_i \in \text{im}(A)$，则$A^\dagger \boldsymbol{y} = \boldsymbol{v}_i$。
- 它将来自 $(\text{im } A)^\perp = \ker(A^T)$ 的向量映射到 $\mathbb{R}^n$ 中的零向量。

如第6章所述，$A^\dagger \boldsymbol{b}$ 给出了 $A\boldsymbol{x} = \boldsymbol{b}$ 的最小范数最小二乘解。$A^\dagger$ 的 SVD 构造使得这一结论对任意 A 都成立。此外，$AA^\dagger = U_r U_r^T$ (，其中 $U_r = [\boldsymbol{u}_1 \ldots \boldsymbol{u}_r]$) 是到 im($A$) 的正交投影；以及 $A^\dagger A = V_r V_r^T$ (，其中 $V_r = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_r]$) 是到 im($A^T$) 的正交投影。

因此，SVD 为理解线性变换 A 提供了一把“万能钥匙”：它为其定义域和值域提供最优的正交归一基（V 和 U），给出沿这些主轴的缩放因子（Σ）、其秩、四个基本子空间的显式基，以及其广义逆的稳健定义（$A^\dagger$）。

从本质上讲，SVD 揭示了任何线性变换，无论其初始矩阵表示多么复杂，都可以被理解为三个基本几何操作的序列：一次旋转/反射（V^T）、沿坐标轴的缩放（Σ），以及另一次旋转/反射（U）。这一深刻的洞见是许多现代线性代数应用的核心。

10.4 Invariance & Natural Structure

奇异值的出现不仅仅是计算上的产物，而是该变换的基本不变量——当通过不同的正交规范基来观察时仍保持不变的量。如果 Q_1 和 Q_2 是正交矩阵，那么变换 $B = Q_1 A Q_2^T$ 与 A 表示的是同一个底层映射，只是通过不同的坐标来观察。然而，它的奇异值与 A 的奇异值完全一致，度量的是与我们选择的测量参考系无关的内在伸缩因子。

从这种几何视角自然地出现了两种矩阵范数：

定义 10.4（矩阵范数）。对于矩阵 $A \in \mathbb{R}^{m\times}$ [n]Understood. Please provide the source text you would like

1. 该*spectral norm* (或*2-norm*)测量最大拉伸 ng:

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \sigma_1$$

2. *Frobenius norm* 测量总能量：

$$\|A\|_F = \left(\sum_{i,j} a_{ij}^2\right)^{1/2} = \left(\sum_{i=1}^{p} \sigma_i^2\right)^{1/2}$$

Compare: Frobenius 范数源自第 5 章例 5.3 中的内积。它提供了一种自然的近似误差度量，有效地统计了所有矩阵元素上的平方差总量。

•

这些范数通过定义域和陪域中的基本度量 A^TA 和 AA^T 与 SVD 结构深度关联：

$$A^TA = V\Sigma^T\Sigma V^T = \sum_{i=1}^{p} \sigma_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^T \quad \text{and} \quad AA^T = U\Sigma\Sigma^T U^T = \sum_{i=1}^{p} \sigma_i^2 \boldsymbol{u}_i \boldsymbol{u}_i^T$$

这些矩阵之间非零特征值的相等性如今自然地源自奇异值结构。

更为引人注目的是，奇异值如何控制变换的复合。尽管特征值在矩阵乘法下可能呈爆炸式增长，奇异值却满足精妙的不等式：

引理 10.5（奇异值交错）。*For matrices A and B of compatible size with singular values in descending order:*

$$\sigma_i(AB) \le \sigma_i(A)\sigma_1(B)$$

Proof Idea. 关键洞见来自于*minimax principle*：第 k 个奇异值等于在所有秩为（$k-1$）的近似中的最小谱范数。对于任意 k 维子空间 S：

$$\sigma_k(AB) = \min_{\text{rank}(X)<k} \|AB - X\|_2 \le \|A\|_2 \sigma_k(B) = \sigma_1(A)\sigma_k(B)$$

通过同时利用 A 和 B 的 SVD，更为细致的论证确立了完整的不等式。

□

这种在复合下对奇异值的控制在特征值方面并没有直接的对应。即使两个对称正定矩阵的乘积也可能具有复特征值，而它们的奇异值仍然是实数且表现良好。这种在复合下的稳定性有助于解释为何在分析迭代过程和误差传播时，奇异值往往比特征值更为有用。

奇异值也提供了衡量矩阵秩的最自然的度量：

$$\operatorname{rank}(A) = \#\{\sigma_i > 0\}$$

这个等式阐明了秩的几何意义，即对独立伸展方向的计数。更微妙的是，较小的奇异值表明一些方向是*nearly*依赖的——这是我们将在第12章中探讨的、用于数值计算秩的关键洞见。

这些关系——通过范数、复合和秩——反映了奇异值如何刻画线性变换内在特性的不同侧面。它们的整体力量在于将代数、几何和计算的视角统一到一个连贯的框架中，用以理解线性映射。

矩阵的奇异值不仅仅提供最优近似——它们还能精确量化矩阵在变换过程中对空间的扭曲程度。回顾第 1 章中将条件数作为数值敏感性度量的概念。现在，SVD 框架使我们能够严格地定义这一概念：

定义 10.6（条件数）。对于一个非奇异矩阵 A，其奇异值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$，*condition number* 是其比值

$$\operatorname{cond}(A) = \frac{\sigma_1}{\sigma_n}$$

按奇异值从大到小排列。对于奇异矩阵，我们设 $\operatorname{cond}(A) = \infty$。

•

这一定义阐明了为什么条件数用于度量敏感性。在求解 $A\boldsymbol{x} = \boldsymbol{b}$ 时，SVD 表明 A 会将某些方向拉伸 σ_1，同时将其他方向压缩 σ_n。因此，比值 σ_1/σ_n 给出了在计算解时相对误差可能被放大的程度的上界。更准确地说，右端项中的小扰动 $\delta\boldsymbol{b}$ 可能导致解的变化，其幅度最高可达 $\operatorname{cond}(A)$ $\|\delta\boldsymbol{b}\|$。

条件数在正交变换下保持不变：如果 Q_1 和 Q_2 是正交的，那么 $\operatorname{cond}(Q_1 A Q_2) = \operatorname{cond}(A)$。这种不变性反映了条件性衡量的是内在敏感性

Example: 第1章例1.13中的矩阵，

$$A = \begin{bmatrix} 1 & 0.999 \\ 0 & 0.001 \end{bmatrix}$$

具有奇异值 $\sigma_1 \approx 2$ 和 $\sigma_2 \approx 0.001$，这解释了其约为 2000 的条件数。

而不是特定坐标选择的产物。事实上，cond(A) 可以在不依赖 SVD 的情况下刻画为

$$\text{COND}(A) = \|A\|\|A^{-1}\|$$

其中 $\|\cdot\|$ 表示在正交变换下不变的任意矩阵范数。

这种关于条件性的几何解释——将其视为衡量 A 如何把单位球变换得多么偏心——解释了它在数值计算中的根本重要性。具有较大条件数的矩阵会将某些方向映射到几乎为零，使得无论我们采用何种数值方法，都难以从输出中准确恢复输入。

奇异值也阐明了第3章中研究的基本子空间。一个变换的核恰好对应于具有零奇异值的右奇异向量：

$$\ker(A) = \text{span}\{\boldsymbol{v}_i : \sigma_i = 0\}$$

而其像空间由对应非零奇异值的左奇异向量张成：

$$\text{im}(A) = \text{span}\{\boldsymbol{u}_i : \sigma_i > 0\}$$

Think: SVD 为四个基本子空间——核、像、余核和余像——各自提供了一组正交归一基。这一对基本定理的几何精化不仅揭示了维数，还给出了自然的坐标系。

这种理解将诸如秩和零度（零空间的维数）等抽象概念转化为具体的几何度量。为零的奇异值表示在该变换下塌陷的一个方向；这类零值的数量就是零度。非零奇异值度量了每个保留下来的方向被拉伸或压缩的程度；它们的数量给出了秩。

例 10.7（矩阵补全）。考虑一个包含缺失条目的矩阵，例如来自不完整调查的数据：

$$A = \begin{bmatrix} 1 & ? & 2 \\ 2 & 1 & ? \\ ? & 2 & 3 \end{bmatrix}$$

如果我们怀疑真实矩阵是低秩的（意味着各个条目之间存在许多依赖关系），SVD 提出了一种自然的补全策略：填充缺失条目，使非零奇异值的数量最小化。这一几何原理——真实数据往往位于接近低维子空间的位置——对现代数据科学具有深远影响。◇

虽然奇异值在按降序排列时是唯一确定的，但奇异向量表现出受限的非唯一性，这反映了基本的对称性：

- 对应于不同非零奇异值的向量仅在符号上确定

- 向量共享单一值时，可以在它们的子空间内旋转。
- 对应于零奇异值的向量只需要形成核的正交标准基。

示例 10.8（图像压缩）。作为 $m \times n$ 矩阵存储的灰度图像通常具有满秩——每个奇异值都是非零的。然而，大多数奇异值可能非常小，表示在图像视觉内容中贡献较小的方向。将这些小的奇异值设置为零，实际上减少了秩，同时保持了基本特征。这个低秩近似的过程将在第 12 章详细研究，示范了奇异值如何通过几何直觉指导实际计算。◇

这种几何性的理解将我们对线性变换的看法，从仅仅是计算对象，转变为具有内在特性的结构化实体。奇异值分解（SVD）不仅揭示了如何对矩阵进行分解，更揭示了如何解读线性映射自身所编织的最深层模式。这些关于伸缩、秩以及基本子空间的模式，将在接下来的章节中引导我们同时推进理论理解与实用算法的构建。

10.5 Inner Products & the SVD

奇异值分解自然地推广到任意内积空间之间的线性变换。给定有限维内积空间之间的 $T : V \to W$，第 5 章中定义的伴随算子 $T^* : W \to V$ 使我们能够形成 $T^*T : V \to V$ 和 $TT^* : W \to W$。这些自伴算子扮演着 A^TA 和 AA^T 的角色，其谱性质决定了 SVD 结构。

Definition: 对于向量 $\boldsymbol{u} \in W$ 和 $\boldsymbol{v} \in V$，张量积 $\boldsymbol{u} \otimes \boldsymbol{v}$ 表示将 $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{v} \rangle \boldsymbol{u}$ 发送的秩-一算子。这将矩阵外积 $\boldsymbol{u}\boldsymbol{v}^T$ 推广到任意内积空间。

更精确地说，令 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ 是 T^*T 的正交归一特征向量，对应特征值 $\{\sigma_1^2, \ldots, \sigma_n^2\}$。对于每一个非零奇异值 σ_i，向量 $\boldsymbol{u}_i = \frac{1}{\sigma_i} T\boldsymbol{v}_i$ 是良好定义的，这些向量在 W 中形成一个正交归一集合。该变换随后可分解为：

$$T = \sum_{i=1}^{r} \sigma_i (\boldsymbol{u}_i \otimes \boldsymbol{v}_i)$$

当 $r =$ 排名(T) 时，以正交标准基表示，这种抽象分解恰好得到之前开发的矩阵分解 $A = U\Sigma V^T$。

这种无坐标的视角揭示了SVD不仅仅是一个矩阵分解，而是线性变换的一个基本性质。

在有限维内积空间之间。有限维性的假设至关重要——在无限维情形下，情况会变得更加微妙。尽管希尔伯特空间之间的紧算子允许一种类似的谱分解，其具有可数个趋于零的奇异值，但一般的有界算子可能完全不存在这样的分解。有限维与无限维之间的这一界限，标志着线性变换结构中的一次深刻转变。

BONUS! 无限维空间中紧算子的谱理论引出了与积分方程、量子力学和泛函分析的深刻联系。SVD 在其中作为积分核的 *Schmidt decomposition*出现。

然而，对于有限维空间之间的变换，内积结构已经足够——没有它，我们无法构造伴随算子或度量正交性，通过奇异向量在输入与输出空间之间建立的那种优美联系也将消失。内积恰恰提供了使 SVD 自然涌现所需的几何结构，而不依赖于任何坐标选择。

例10.9（有限维函数空间）。考虑由次数不超过 n 的多项式构成的空间 $\mathcal{P}_n$，并在区间 $[0, 1]$ 上配备 L^2 内积：$\langle f,\ g\rangle = \int_0^1 f(t)g(t)\,dt$。微分算子 $D : \mathcal{P}_n \to \mathcal{P}_{n-1}$ 是线性的，而其伴随算子 $D^* : \mathcal{P}_{n-1} \to \mathcal{P}_n$ 同时涉及积分项和边界项。尽管我们处理的是函数，这些多项式空间的有限维性保证了 SVD 的存在性与唯一性。

◇

例10.10（积分算子）。考虑由下式定义的积分算子 $T : C([0, 1]) \to C([0, 1])$

$$(Tf)(x) = \int_0^x f(t)\,dt$$

当为 $C[0, 1]$ 赋予标准内积 $\langle f,\ g\rangle = \int_0^1 f(t)g(t)\,dt$ 时，该算子是有界的。其伴随算子可通过分部积分得到：

$$(T^*f)(x) = \int_x^1 f(t)\,dt$$

T他构造了算子 T^*T，随后得到了一个特别好的 f ORM：

$$(T^*Tf)(x) = \int_0^1 \min(x, t)f(t)\,dt$$

这是一个具有对称核的积分算子的例子。虽然我们无法显式写出它的奇异值和奇异向量（它们涉及某些微分方程的解），但我们知道它们必然在 $C([0, 1])$ 中构成一组完备的正交归一集。这一无限的奇异值序列趋于零，从而保证该算子是紧的——这与有限维情形形成关键区别，在有限维情形中，除非恰好为零，奇异值都会与零保持有界距离。

这个例子暗示了积分算子更深层的理论，其中 SVD 不再表现为有限求和，而是作为无穷级数出现。几何直觉依然成立——我们仍在将变换分解为正交分量——但维度的无穷性引入了有限维中不存在的收敛性与完备性等微妙问题。 ◇

Latent Semantic Structure in Text

隐藏在人类话语土壤之中的，是生长为意义的文本之根。奇异值分解揭示了这些潜在结构，将我们关于 *words are known by the company they keep* 的直觉感受转化为更为精确的数学洞见。通过对词语共现模式的细致分析，我们能够揭示赋予语言非凡表达力的深层关系。

考虑一个通过词频表示的文档集合。每个文档成为高维空间中的一个向量，其中每个维度对应一个可能的词。完整的语料库形成一个词项-文档矩阵 A，其中条目 a_{ij} 表示词 i 在文档 j 中的（加权）出现次数。这个矩阵虽然稀疏且维度很高，但包含着 SVD 可以揭示的丰富结构。

奇异值分解 $A = U\Sigma V^T$ 揭示了基本的语义模式：

- 左奇异向量（U 的列）揭示了倾向于共同出现的词簇
- 右奇异向量（V 的列）用于识别文档主题
- 奇异值衡量这些语义关联的强度

更值得注意的是，尽管这种分解完全基于词共现模式运行，它往往能够捕捉到真实的语义关系。具有相似含义的词往往出现在相似的语境中，从而在词—文档矩阵中形成平行的行。SVD 会检测到这些平行性，并在其奇异向量中将相关术语分组。

例 10.11（科学摘要分析）。考虑分析一组物理学摘要。前几个左奇异向量通常会揭示清晰的语义分组：

1. 实验术语：“测量”“观察”“数据”“实验” 2. 理论术语：“模型”“理论”“预测”“框架” 3. 量子术语：“状态”“叠加”“纠缠”“量子比特”
这些分组自然地从共现模式中涌现，而算法并未被提供任何显式的语义知识。

◇

奇异值本身就讲述了一个有趣的故事。它们通常呈现幂律衰减：少数值很大，随后是大量较小的值。这表明大多数语义内容存在于一个低维子空间中——这一现象使得高效的文档索引和语义搜索成为可能。

若干实用性的改进（如词频–逆文档频率加权）以及更为复杂的现代方法共同增强了这些 SVD 基础。诸如 *Word2Vec* 之类的词嵌入为词语创建了稠密的向量表示，能够捕捉细微的语义关系。然而，这些方法仍然体现了一个根本性的洞见：意义源自关联模式——而 SVD 正是独特地擅长揭示这些模式。

Historical Note: 潜在语义分析（LSA）诞生于20世纪80年代末，使用奇异值分解（SVD）作为其基础性的数学工具。将矩阵分解应用于语言的这一做法，改变了理论语言学以及实用的信息检索系统。

这一应用例证了一个更深层的真理：线性代数能够照亮看似无结构的数据中的结构。正如奇异值分解（SVD）揭示了几何变换中拉伸的优选方向，它也在高维的人类语言空间中揭示了自然的语义轴。因此，本章所发展的数学不仅提供了计算工具，还为意义如何从模式中涌现提供了真正的洞见。

Sensor Networks & The Geometry of Measurement

传感器是工业体的神经——数据中心中的温度探头，智能手机中的加速度计，工业工厂中的压力表。每个设备都测量现实的外部方面，但这些测量存在来自制造差异、环境条件和故障的系统误差。奇异值分解提供了一个优雅的框架，通过测量的基础几何结构来整合和平滑这些误差。

考虑一个包含 n 个传感器的数组，这些传感器在 m 个不同的时间或条件下测量相同的物理量。在理想的世界中，这些测量值仅因已知的物理变化而有所不同。现实证明情况更为复杂——每个传感器都有其自己的增益、偏移和噪声特性。一个 *measurement matrix* $M \in \mathbb{R}^{m \times n}$ 包含了这些被污染的观测值：

$$M_{ij} = g_i(s_j + \eta_{ij}) + b_i$$

其中 s_j 是在时间 j 的真实信号，g_i 和 b_i 分别是传感器 i 的 *gain* 和 *bias*，η_{ij} 表示噪声。

此测量矩阵的SVD揭示了显著的结构。对每个传感器的读数进行中心化（减去其均值）后，我们得到：

$$M = U\Sigma V^T = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T$$

主导奇异向量 $\boldsymbol{v}_1$ 往往捕捉到真实的潜在信号，而后续的奇异向量则揭示系统性的误差模式：

- $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}^T{}_1$ 近似于物理变化
- $\sigma_2 \boldsymbol{u}_2 \boldsymbol{v}^T{}_2$ 通常显示增益变化效应
- 高阶项可以捕捉更细微的误差模式

示例 10.12（温度传感器阵列）。考虑一个由 100 个温度传感器监控的服务器机房，每分钟采样一次。在一小时的运行时间内（60 个样本），我们得到一个 60×100 的测量矩阵。SVD 通常揭示：

1. 第一奇异值 $\sim$ 比第二个大 $10\times$ 倍，反映了真实的温度变化 2. 第二奇异向量与传感器位置相关，显示空间偏差 3. 第三及以后捕捉各种漂移和噪声模式　这种分解能够同时实现数据清洗和传感器故障检测。

◇

奇异值本身提供了重要的诊断信息。定义测量矩阵的*effective rank*为：

$$r_{\text{eff}} = \left(\sum_{i=1}^{r} \sigma_i^2\right)^2 \Big/ \sum_{i=1}^{r} \sigma_i^4$$

该量始终介于 1 与 $r = \text{rank}(M)$ 之间，用于衡量数据中存在多少个独立模式。取值接近 1 表明测量较为干净，主要由物理信号主导；取值越大则表明显著的系统性误差。

Nota bene: 这种有效秩公式在随机矩阵理论和量子力学中以参与比的形式出现，用于衡量一个系统中有多少个分量在其中发挥显著作用。

更复杂的分析使用完整的SVD结构来校准传感器阵列。如果$\boldsymbol{v}_1$近似于真实信号方向，我们可以通过将每个传感器的响应与此参考进行比较，来估计每个传感器的增益：

$$\hat{g}_i = \frac{\langle \boldsymbol{m}_i, \boldsymbol{v}_1 \rangle}{\|\boldsymbol{v}_1\|^2}$$

其中 $\boldsymbol{m}_i$ 是 i 列的 M (居中)。

Nota bene: 更复杂的标定模型可以通过对奇异向量的细致分析加以解决。

这种传感器校准方法揭示了物理测量中一个更深层的事实：尽管原始数据往往看起来复杂且受到污染，但其底层信号通常存在于一个低维子空间中。系统性误差并非产生纯粹的噪声，而是生成具有特征性的几何模式，SVD 能够自然地将其检测并分离出来。现代传感器网络通过用于时变校准的滑动窗口分析以及在大规模阵列上的分布式计算扩展了这些原理，但核心洞见依然不变：测量误差看似复杂，然而在合适的坐标系下往往呈现出异常简单的结构。

Tensors & Multi-Linear Algebra

数据很少服从二维表示。尽管矩阵为分析成对关系提供了强大的工具，现实往往需要更高维的结构。一组随时间变化的 RGB 图像；跨不同平台和情境的用户与产品的交互；深度神经网络在多层中对多样输入作出响应的激活——这类数据自然地组织成称为 *ten- sors* 的多维数组。正如我们发展过程中从向量到矩阵的过渡一样，从矩阵跃迁到张量通过对熟悉原理的谨慎扩展揭示了新的模式。

首先考虑形式结构。一个k阶的*tensor*为每一种k个索引的选择赋予一个数：

$$\mathcal{T} = [t_{i_1 i_2 \cdots i_k}], \quad 1 \le i_j \le n_j$$

正如矩阵通过增加第二个索引来推广向量，张量将这种索引扩展到任意维度。这种代数定义与多变量微积分中遇到的微分形式自然地联系起来——k-形式恰好是一个交替的 k-张量，通过多线性映射来度量定向的 k 维体积。

Nota bene: 微积分中研究的微分形式表示一种特殊的张量，它们在对其输入进行奇置换时会变号。它们的反对称性使其非常适合用于积分，正如对称张量在深度学习中显得自然而然一样。

线性代数中熟悉的运算在这一情境下自然地扩展。给定一个三阶张量 $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$，我们可以通过 *fibers* (固定除一个索引外的所有索引) 或 *slices* (固定除两个索引外的所有索引) 来提取矩阵：

$$\mathcal{T}_{:jk} = [t_{ijk}]_{i=1}^{n_1}, \quad \mathcal{T}_{i::} = [t_{ijk}]_{j,k=1}^{n_2,n_3}$$

这些部分为张量的结构提供了窗口，正如行向量和列向量阐明了矩阵一样。

Example: 在图像嵌入的张量中，纤维可能表示跨相似图像的特征轨迹，而切片则捕捉固定语义层级上的特征关系。

将奇异值分解扩展到张量揭示了深刻的微妙性。尽管矩阵允许分解为正交因子的唯一分解，张量却抗拒如此整洁的因式分解。*CP decomposition* 提供了一种自然的推广：

$$\mathcal{T} = \sum_{r=1}^{R} \sigma_r \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r$$

Nota bene: CP 代表 CANDECOMP / PARAFAC，正如你可能已经猜到的那样。

通过外积将张量表示为秩一分量之和。尽管优雅，这种分解很少能够实现精确的低秩表示。*Tucker decomposition* 提供了更大的灵活性：

$$\mathcal{T} = \mathcal{S} \times_1 U \times_2 V \times_3 W$$

其中 $\mathcal{S}$ 表示一个小的 *core tensor*，而 $\times_k$ 表示沿 k 模态的乘法。这种结构——通过因子矩阵展开的集中核心——呼应了 SVD 如何揭示矩阵中的低维结构。

例 10.13（神经网络分析）。现代深度网络将其学习到的权重组织为张量，其索引涵盖输入通道、输出通道以及空间维度。作用于图像数据的卷积层使用四阶张量 $\mathcal{W} \in \mathbb{R}^{c_{\text{out}} \times c_{\text{in}} \times h \times w}$ 来变换输入特征图。理解信息如何在这些结构中流动需要张量分析：

$$\text{output}[i, x, y] = \sum_{j,p,q} \mathcal{W}[i, j, p, q] \cdot \text{input}[j, x + p, y + q]$$

这些权重的张量分解揭示了学习到的特征层级，同时通过降低参数量实现高效计算。 ◇

除了原始数据的组织之外，张量为现代机器学习系统提供了自然的结构。语言模型通过注意力张量处理序列，这些张量捕捉了词元、位置和特征维度之间的关系。计算机视觉模型通过保持空间关系的张量运算来构建分层表示，同时学习抽象特征。推荐系统通过张量分解来建模用户、物品和上下文之间的复杂交互，从而捕捉潜在模式。

示例 10.14（多模态学习）。考虑一个处理带有文本描述的图像的深度学习系统。每个图像-文本对在各自的空间中生成嵌入，但它们之间的关系形成一个三阶张量 $\mathcal{T} \in \mathbb{R}^{n \times d_1 \times d_2}$，其中：

- n 索引训练示例
- d_1 表示图像嵌入维度
- d_2 表示文本嵌入维度

张量结构刻画了不同模态如何相互作用——其分解揭示了共享的语义空间，从而实现跨模态检索与生成。

◇

张量的数学通过现代应用不断演进。训练算法利用张量结构实现高效的梯度计算。神经网络架构通过张量分解压缩其参数。基础模型借助张量运算同时处理多种模态。每一项发展都揭示了这些基本对象的新方面，将抽象的多线性映射转化为人工智能的实用工具。

□ ——————————————————————————— □

Exercises: Chapter 10

1. 对于以下矩阵，A，通过以下步骤计算奇异值分解：(1) 计算 A^TA 和 AA^T；(2) 求各自的特征值和特征向量；(3) 构造矩阵 U、Σ 和 V。

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix} \quad : \quad A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 2 & -1 \end{bmatrix}$$

2. *Hilbert matrix* H_n 具有条目 $h_{ij} = \frac{1}{i+j-1}$。例如：

$$H_3 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

使用 SVD 计算 $\text{cond}(H_3)$；然后解释当 $n \to \infty$ 时 $\text{cond}(H_n)$ 的渐近行为。

3. 证明对任意矩阵 A：$\text{tr}(A^TA) = \sum_{i=1}^{p} \sigma_i^2$，其中 $p = \min\{m, n\}$。4. 对于方阵 A，证明 $|\det(A)| = \prod_{i=1}^{n} \sigma_i$。然后利用这一点说明 A 当且仅当其所有奇异值均非零时是可逆的。

这将行列式的代数概念与奇异值作为伸缩因子的几何意义联系起来。

5. 对于一个非奇异矩阵 A，令 H 为 AA^T (的正平方根——为什么这是良好定义的？)。证明存在某个正交矩阵 Q 使得 $A = HQ$。 6. 如果 A 是可逆的，你如何利用 SVD 快速计算 A^{-1}？ *Ex- plain.* 7. 设 $A + A^T$ 和 $A - A^T$ 的奇异值至多为 A 的两倍。何时取等号?

8. 证明对于任意非奇异矩阵 A 以及正交矩阵 Q_1、Q_2：

$$\text{COND}(Q_1 A Q_2) = \text{COND}(A)$$

9. 对于尺寸相容的矩阵 A 和 B，证明：

$$\text{COND}(AB) \leq \text{COND}(A)\text{COND}(B)$$

何时成立等式?

10. 设 $\boldsymbol{x} \in \mathbb{R}^n$ 为单位向量。矩阵 A 的 *Rayleigh quotient* 定义为 $R_A(\boldsymbol{x}) = \boldsymbol{x}^T A^T A \boldsymbol{x}$。证明：

(a) $\sigma_n^2 \leq R_A(\boldsymbol{x}) \leq \sigma_1^2$ 对所有单位向量 $\boldsymbol{x}$
(b) 当且仅当 $\boldsymbol{x}$ 是相应的右奇异向量时，任一界中等号成立

11. 对于一个对称正定矩阵 A，证明其奇异值等于其特征值。对于一般矩阵 A，在什么条件下，其奇异值等于其特征值的绝对值?

12. 一个实矩阵 A 如果 $AA^T = A^T A$，则称为 *normal*。正常矩阵的 SVD 有哪些有趣的性质?

13. 设 A 是一个 $m \times n$ 矩阵，具有 $m > n$ 和满秩列。证明 $A^T A$ 的条件数是 A 的条件数的平方。这个结果对求解最小二乘问题有什么实际意义?

14. 考虑通过仅保留 k 个最大奇异值获得的矩阵 A 的秩-k 近似 A_k。证明：

$$\text{COND}(A_k) \leq \text{COND}(A)$$

当且仅当 $k =$ 的秩为 A 时，成立相等。解释为什么这意味着低秩近似通常比原始矩阵具有更好的条件。

15. 设 V 和 W 是具有正交归一基 $\{\boldsymbol{e}_i\}$ 和 $\{\boldsymbol{f}_j\}$ 的有限维内积空间。给定一个线性变换 $T : V \to W$，其 *matrix elements* 分别为 $a_{ij} = \langle T\boldsymbol{e}_i$ 和 $\boldsymbol{f}_j \rangle$。证明：

(a) $\sum_{i,j} |a_{ij}|^2 = \sum_k \sigma_k^2$，其中 σ_k 是 T (b) 的奇异值。该和与正交归一基的选择无关

16.（挑战）设 V 是一个有限维内积空间，且 $T : V \to V$ 是一个线性变换。T 的 *numerical range* 定义为：

$$W(T) = \{\langle T\boldsymbol{v}, \boldsymbol{v} \rangle : \boldsymbol{v} \in V, \|\boldsymbol{v}\| = 1\}$$

证明如果 T 是正规矩阵（即 $TT^* = T^*T$），则 $W(T)$ 是 T 的特征值的凸包。当 T 不是正规矩阵时，这个集合与奇异值有什么关系?

17.（挑战）设 $A = U\Sigma V^T$ 为矩阵 A 的 SVD。通过仅保留 Σ 中前 k 个奇异值并将其余置为零来定义 A_k。证明在所有秩至多为 k 的矩阵 B 中，A_k 使 $\|A - B\|_F$ 取得最小值。

This optimality property of the SVD truncation will be explored further in Chapter 12 on low-rank approximation.

Chapter 11
Principal Component Analysis

"to stretch across the heavens & step from star to star"

深层模式隐藏在高维数据中，直接观察无法看到，但却是理解复杂系统的基础。挑战不在于收集数据——现代科学和工程产生了大量的观察数据——而在于从通常跨越数十或数百维度的测量中提取有意义的结构。

Example: 一个单一的人类基因组包含大约20,000个基因，这些基因的表达水平在不同条件下有所变化。理解这种变化需要在一个20,000维的空间中寻找模式。

考虑数字图像，其中每个像素强度代表一个维度，或金融市场，其中成千上万的证券在微妙的关联中波动。即使是简单的物理系统，在完全仪器化后，也会生成维度远超人类直觉理解的测量值。在这样的空间中，重要特征通常沿着少数几个关键方向集中，就像矿藏通过地质过程集中一样。

主成分分析（PCA）提供了揭示这些基本模式的数学工具。通过仔细研究测量值的变化和相关性，PCA揭示了与数据内在结构对齐的自然坐标。这些坐标按其在捕捉变异中的重要性排序，能够实现降维和模式发现。

Foreshadowing: 主成分分析（PCA）与神经网络（第13章）之间的联系非常深刻：两者都旨在将高维数据转化为更有意义的表示。

此分析的基础来源于我们在第10章中关于奇异值的工作。在那里，我们看到任何线性变换都可以分解为沿主轴的正交拉伸。主成分分析（PCA）将这一几何洞察应用于数据矩阵，其中行表示观测值，列表示测量的变量。这些矩阵的奇异向量揭示了理解变异的自然坐标，而奇异值则衡量了每个方向上模式的强度。

Nota bene: PCA 起源于统计分析，追溯到 Pearson (1901)，尽管 SVD 更为现代。

我们的论述从统计学基础出发，经由几何直觉，走向实际应用。尽管我们分析的数据乍看之下可能显得杂乱无章，但细致的数学挖掘往往会揭示其内在的简单性。正如通过谨慎地去除多余材料，雕塑从粗石中显现出来一样，有意义的低维结构也通过有原则的降维从高维数据中浮现。我们的任务是同时发展理解这一过程的理论，以及在实践中实现它的工具。

11.1 *Covariance & Correlation*

方差的故事始于旋转。一个绕其质心旋转的刚体所经历的转动阻力并非由总质量决定，而是由质量在空间中的分布决定。熟悉的标量转动惯量 $I = \int r^2\, dm$ 衡量的是绕单一轴的这种阻力，但完整的描述需要完整的 *inertia tensor*：

$$\mathcal{I} = [\mathcal{I}_{ij}] \quad : \quad \mathcal{I}_{ij} = \int (r^2 \delta_{ij} - x_i x_j)\, dm$$

Definition: 当$i = j$时，克罗内克δ δ_{ij}的值为1，否则为0。

这种机械视角——质量围绕中心的分布及其对不同旋转的阻抗——为我们如今所构建的统计结构提供了出人意料的深刻洞见。

首先考虑一个单一的随机变量 Z。其 *mean* 或 *expectation* $\mu = \mathbb{E}(Z)$ 充当质心，而其 *variance* $\mathbb{V}(Z) = \mathbb{E}((Z-\mu)^2)$ 衡量围绕该中心的离散程度——这与质量分布相对于其质心的标量转动惯量完全类比。*standard deviation* $\sigma = \sqrt{\mathbb{V}}$ 类似于力学中的回转半径，提供了这种离散的特征长度尺度。

Think: 正如固体对旋转的抗性取决于其绕各轴的质量分布一样，数据集的统计结构取决于测量值在不同方向上围绕其均值的分布方式。

在大多数情况下，数据是离散的而非连续的，我们可以将Z表示为向量$\mathbb{Z} = (z_1, \ldots, z_n)^T$。由此，我们得到基本的统计度量：

$$\mathbb{E}(Z) = \frac{1}{n}\sum_{i=1}^{n} z_i \quad \text{and} \quad \mathbb{V}(Z) = \frac{1}{n}\sum_{i=1}^{n} (z_i - \mathbb{E}(Z))^2$$

Nota bene: 在方差前面，人们常常看到的是 $1/(n-1)$，而不是 $1/n$，尤其是在统计学的语境中。这反映了在从样本中估计均值时损失了一个自由度。对于进行数据科学和降维的目的而言，$1/n$ 是更为合适的缩放方式，我们将在全文中采用它。

在数据科学和统计学中，人们通常会*centers*数据，将Z转换为均值为零的$\hat{Z} = Z - \mathbb{E}(Z)$。随后，这引出了对方差的几何解释：将其视为 $\mathbb{V}(Z) = \hat{Z}^T\hat{Z}$，而标准差被解释为维度尺度的长度：

$$\sigma = \sqrt{\mathbb{V}} = \frac{1}{\sqrt{n}}\sqrt{\hat{Z}^T\hat{Z}} = \frac{1}{\sqrt{n}}\left\|\hat{Z}\right\|.$$

这在　诠释是理解几何 o 的关键　　f 数据。

两个随机变量会发生什么？协方差和相关系数是关键的度量。给定随机变量 Y 和 Z，它们的 *covariance*

$$\begin{aligned}\operatorname{cov}(Y,Z) &= \mathbb{E}(\hat{Y}\hat{Z}) = \mathbb{E}((Y-\mathbb{E}(Y))(Z-\mathbb{E}(Z))) \\ &= \frac{1}{n}\hat{Y}\cdot\hat{Z}\end{aligned}$$

衡量它们共同变化的趋势。类似于惯性矩阵中的非对角项，协方差刻画了不同变化方向之间的耦合。正协方差表示 Y 的大值往往与 Z 的大值同时出现，而负协方差则暗示相反——当一个变量高于其均值时，另一个往往低于其均值。

这一事实——它是一个点积（按维度缩放）——应当能抚慰那些熬过传统统计学课程的学生。沉浸在点积所激发的几何想象之中，读者或许已经能猜到接下来会发生什么。

为了确定两个数据向量之间的对齐程度（或不对齐程度），人们将 *correlation* 定义为对协方差进行重新缩放，使其取值范围介于 -1 和 $+1$ 之间，其中相关性为零表示独立性。用公式表示，相关性就成为一个熟悉的朋友：

$$\operatorname{corr}(Y,Z) = \frac{\operatorname{cov}(Y,Z)}{\sigma(Y)\,\sigma(Z)} = \frac{\hat{Y}\cdot\hat{Z}}{\|\hat{Y}\|\|\hat{Z}\|} = \cos\theta(\hat{Y},\hat{Z}).$$

Truth: 相关性不等于因果性；但它是余弦相似度。

它是两个中心化数据点之间的夹角。

11.2 *Matrices & Data*

单个数据向量对应于点云中的一个点。那么我们应当如何着手处理整个数据集呢？读者此时想必已经看到了未来：一组数据点变成一组向量，被汇集成一个数据矩阵。

对于一个（任意）有序的 d 个变量 Z_1、……、Z_d 的集合，将每个变量中心化到$\hat{Z}_i$，并将它们排列成一个具有 d 列的中心化数据矩阵 $\mathcal{X}$。在实践中，$\mathcal{X}$的列对应于各个变量，而行对应于样本或运行。

为了从数据中估计这些量，我们将观测结果整理成一个 *data matrix* $\mathcal{X} \in \mathbb{R}^{n\times d}$，其中：

- 每一行代表一次观测；
- 每一列对应一个变量；
- 条目x_{ij}是来自观测i的第j次测量。

例如 ce，考虑在三个w处的每日温度测量 eather

超过一年的站点：

$$\mathcal{X} = \begin{bmatrix} 72 & 70 & 68 \\ 75 & 74 & 71 \\ 65 & 63 & 62 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

通过减去列均值进行中心化（类似于力学中转换到质心坐标系），协方差矩阵变为：

$$[C] = \frac{1}{n}\mathcal{X}^T\mathcal{X} = \begin{bmatrix} 25.3 & 23.1 & 20.4 \\ 23.1 & 24.7 & 19.8 \\ 20.4 & 19.8 & 22.9 \end{bmatrix},$$

其中我们假设 $\mathcal{X}$ 已经被中心化。

对角线元素显示了各站点的温度方差——站点1的变异性略高于其他站点。较大的正的非对角项表明站点之间存在强相关性，这符合相邻地点经历相似天气模式的预期。然而，这种相关性并非完美，站点对（1,2）的关系强于涉及站点3的组合，表明站点3在地理位置上可能更为遥远。

该协方差矩阵是对称半正定的——这一性质自然地源于其构造方式。其对角元素是各个变量的方差，而非对角项则度量成对关系。

正如惯性矩阵的特征值衡量绕主轴旋转的阻力一样，协方差矩阵的特征结构揭示了数据中基本的变化模式。为了在不受尺度影响的情况下更好地理解这些模式，我们有时会通过 *corre- lation matrix* $[R] = [R_{ij}]$ 进行归一化，其中

$$R_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j} = \frac{\operatorname{cov}(Z_i, Z_j)}{\sqrt{\mathbb{V}(Z_i)\mathbb{V}(Z_j)}}$$

这会将所有条目缩放到区间 $[-1, 1]$，仅度量线性关系的强度和方向。对于我们的温度数据，我们首先从协方差矩阵的对角线条目中提取标准差：

$$\sigma_1 = \sqrt{25.3} \approx 5.03^\circ, \quad \sigma_2 = \sqrt{24.7} \approx 4.97^\circ, \quad \sigma_3 = \sqrt{22.9} \approx 4.79^\circ$$

这些我　确保每个站点的典型变化。完整的　　相关-

转换矩阵变为：

$$[R] = \begin{bmatrix} 1.000 & 0.923 & 0.846 \\ 0.923 & 1.000 & 0.831 \\ 0.846 & 0.831 & 1.000 \end{bmatrix}$$

协方差矩阵和相关矩阵将抽象的统计关系转化为具体的几何对象。它们的特征向量标识了主要的变异轴，而它们的特征值则衡量了沿这些轴的变异强度。这种机械直觉、统计理论和几何结构的融合为我们将要开发的降维方法奠定了基础。即使是我们简单的温度示例也暗示了数据集可能蕴藏着隐藏的简洁性：尽管我们测量了三个变量，强相关性却暗示了潜在的模式，等待通过主成分分析来揭示。

11.3 Principal Components

协方差矩阵刻画了我们的数据在测量空间中沿不同方向的变化方式。然而，这些与原始变量对齐的方向可能会掩盖更为简单的潜在模式。正如第 6 章中的投影算子揭示了最优的近似子空间一样，我们现在寻求与数据内在结构对齐的坐标，而不是由任意测量选择所决定的坐标。

考虑一个中心化的数据矩阵 $\mathcal{X} \in \mathbb{R}^{n \times d}$，其中每一行代表 d 个变量的一个观测值，每一列具有零均值。在第10章研究的奇异值分解提供了我们所需的变换：

定义 11.1（主成分）。给定一个中心化的数据矩阵 $\mathcal{X}$，其 *principal components* 是从 SVD $\mathcal{X} = U\Sigma V^T$ 中得到的右奇异向量 $\boldsymbol{v}_1, \dots, \boldsymbol{v}_d$，按奇异值递减顺序排列。每个成分 $\boldsymbol{v}_k$ 表示在原始变量空间中的一个方向，该方向捕捉了在考虑了先前成分后的最大剩余变异。

•

这些主成分将我们的原始变量转化为新的特征，捕捉数据的变化结构：

定义 11.2 (PC 分数)。给定一个主成分 $\boldsymbol{v}_k$，对应于观察值 $\boldsymbol{x} \in \mathbb{R}^d$ 的 *principal component score* 是其在该方向上的投影 $z_k = \boldsymbol{x}^T \boldsymbol{v}_k$。所有观察值沿着 $\boldsymbol{v}_k$ 的分数形成 k 阶 *score vector* $\boldsymbol{z}_k = \mathcal{X}\boldsymbol{v}_k$。

•

正如第6章的正交投影将向量分解为最优近似的分量一样，这些得分向量将我们的数据分解为重要性递减的正交特征。奇异值与变异之间的关系自然显现：如果 σ_k 是 $\mathcal{X}$ 的第 k 个奇异值，那么 $\lambda_k = \sigma_k^2/(n-1)$ 给出了沿第 k 个主成分方向的得分方差。随着我们沿着各个分量推进，这些方差逐渐减小，反映了每一个后续方向如何捕获数据中剩余的最大变异。

例子 11.3（基因表达数据）。考虑不同细胞类型中成千上万个基因的基因表达测量。我们数据矩阵的每一行代表一个细胞，而列则记录不同基因的表达水平：

$$\mathcal{X} = \begin{bmatrix} \leftarrow & \text{cell 1} & \rightarrow \\ \leftarrow & \text{cell 2} & \rightarrow \\ & \vdots & \\ \leftarrow & \text{cell n} & \rightarrow \end{bmatrix} \begin{matrix} \text{gene 1} \\ \text{gene 2} \\ \vdots \\ \text{gene d} \end{matrix}$$

尽管每个细胞的状态存在于维度为 $d \approx 20{,}000$ 的空间中，生物学约束通常会将变化限制在一个维度更低的流形中。主成分分析通过方向 $\boldsymbol{v}_k$ 揭示了这些约束，这些方向通常对应于基本的调控程序或细胞状态转变。

Foreshadowing: 通过PCA实现的降维预示了神经网络（第13章）如何通过低维特征学习表示高维数据。

将数据投影到前两个主成分上，产生二维可视化：

$$\begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ \vdots & \vdots \\ z_{n1} & z_{n2} \end{bmatrix} = \mathcal{X}[\boldsymbol{v}_1 \; \boldsymbol{v}_2]$$

由此得到的点的散点图（z_{i1}, z_{i2}）往往揭示出相似细胞类型的簇或细胞分化的梯度——这些模式在原始的高维空间中是不可见的。

◇

例 11.4（市场收益）。股票市场指数中股票的日收益提供了另一个富有启发性的例子。$\mathcal{X}$ 的每一行表示一个交易日，而列表示不同的股票。主成分提取出基本的市场因子：

- 第一个成分 $\boldsymbol{v}_1$ 通常反映市场整体的变动
- 后续组件通常与行业特定的差异相一致

- 后续组件可能揭示公司特定的效应

相应的得分衡量了各个因子在不同日期对收益的影响强度，从而提供了对市场行为的自然分解。

◇

Nota bene: 这一系列方向推广了第 4 章中的正交基，如今经过优化以捕捉数据中的变化，而非任意的坐标选择。

前k个成分所捕获的总方差占比，提供了一个衡量它们在多大程度上概括我们数据的指标：

$$r_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i} = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{d} \sigma_i^2}$$

当该比率在 k 较小时趋近于 1 时，我们就找到了一个能够捕获数据大部分变异的低维表示。

在实践中，缩放方式的选择会对 PCA 所发现的模式产生至关重要的影响。由此出现了两种标准方法：

1. *Covariance PCA*：直接使用中心化数据，保留相对尺度 2. *Correlation PCA*：先将每个变量标准化为单位方差

前者强调绝对变化幅度大的方向；后者则不受尺度影响，关注相关性的模式。第 11.5 节将详细探讨这些选择及其影响。

因此，主成分分析提供了一种系统的方法，用与数据中变异对齐的自然轴来替代任意的测量坐标。与第6章的最优投影和第10章的奇异向量类似，这些方向源自基本的几何与代数性质，下一节将对其进行仔细考察。

11.4 Optimality Properties

主成分不仅仅为数据分析提供了便利的坐标——它们自然地源自基本的优化原理。类似于第6章中的正交投影在几何空间中最小化近似误差，主成分在通过低维摘要表示高维数据时同样最小化误差。

首先考虑寻找一个能够最好地捕捉我们数据中变化的单一方向的问题。给定中心化的观测$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, 我们寻求一个单位向量 $\boldsymbol{v}$，使投影的方差最大：

$$\text{maximize} \quad \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i^T \boldsymbol{v})^2 \quad \text{subject to} \quad \|\boldsymbol{v}\| = 1$$

这种优化具有深刻的几何意义：我们寻找数据展现最大离散度的方向。将我们中心化的 $\mathcal{X}$ 记为

数据矩阵，这个目标变为：

$$\frac{1}{n-1}\boldsymbol{v}^T\mathcal{X}^T\mathcal{X}\boldsymbol{v} = \boldsymbol{v}^T[C]\boldsymbol{v}$$

Think: 该优化在两个相互竞争的目标之间取得平衡：在尽可能捕获更多变异的同时，保持各成分之间的正交性。

满足 $\boldsymbol{v}^T\boldsymbol{v} = 1$。

定理 11.5（主成分最优性）。 *The first principal component $\boldsymbol{v}_1$ maximizes $\boldsymbol{v}^T[C]\boldsymbol{v}$ subject to $\|\boldsymbol{v}\| = 1$. Each subsequent component $\boldsymbol{v}_k$ maximizes this same objective subject to orthogonality with all previous components.*

Proof. 根据 Rayleigh-Ritz 原理，在满足 $\|\boldsymbol{v}\| = 1$ 的条件下，$\boldsymbol{v}^T[C]\boldsymbol{v}$ 的最大值等于 $[C]$ 的最大特征值，并由其对应的特征向量达到。后续分量的优化遵循同一原理，即在移除先前分量之后，对剩余变差应用该原理。

□

这种优化视角揭示了 PCA 的根本特性：它提供了一个与数据中变异最优对齐的正交坐标系。每个成分在保持与先前成分正交的同时，提取剩余的最大方差。

例 11.6（图像压缩）。考虑一幅以像素强度矩阵表示的灰度图像。尽管在形式上是高维的，自然图像由于空间相关性，往往将变化集中在少数几个方向上。PCA 揭示了这种低维结构：前几个主成分捕捉到连贯的图像特征，而后续成分通常代表噪声。

Foreshadowing: 第12章将发展支撑此类压缩任务的矩阵近似理论，补充我们当前的统计视角。

例如，仅保留前 k 个分量即可通过以下方式近似原始图像：

$$\hat{\mathcal{X}}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

其中 σ_i 为奇异值，$\boldsymbol{u}_i$、$\boldsymbol{v}_i$ 分别为左、右奇异向量。保留的方差比例等于第 11.3 节中的 r_k，为近似质量提供了一种自然的度量。

◇

除了方差最大化之外，主成分还具有若干等价的最优性性质：

1. 他们最小化 k 维表示的均方重构误差 2. 在高斯假设下，他们最大化数据与其投影之间的互信息

Example: 在金融投资组合分析中，PCA 往往揭示按其对整体市场方差贡献度排序的风险因子——这种分解对风险管理至关重要。

3. 它们为保持成对距离提供了最优的线性降维

重构误差视角尤为启发性。给定观测 $\{\boldsymbol{x}_i\}$，考虑通过以下方式对每个进行近似：

$$\hat{\boldsymbol{x}}_i = \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j$$

其中 z_{ij} 是分数，$\boldsymbol{v}_j$ 是单位向量。主成分最小化：

$$\frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2$$

对于所有k正交归一向量$\{\boldsymbol{v}_j\}$的选择。这种最优性直接与第6章的投影算子相关：PCA提供了基于平方误差的最佳秩-k近似。

示例 11.7 （信号去噪）。考虑一个被噪声污染的时间序列。如果基础信号的结构比噪声更简单，主成分通常通过以下方式实现去噪：

1. 将含噪数据投影到主成分上 2. 仅保留方差高于噪声水平的成分 3. 仅使用这些成分进行重建

PCA 的最优性确保在适当的噪声假设下，该过程能够最小化均方误差。

◇

尽管我们一直关注统计最优性，但通过奇异值分解，这些相同的原理同样可以从纯线性代数中得到。第12章将展开这一互补视角，展示PCA的最优性性质如何在保持计算效率的同时推广到任意矩阵逼近问题。

11.5 Preprocessing & Scaling

真实数据很少以我们理论开发中假定的完美形式到达。考虑通过五个传感器监控一个自动化制造过程：

$$\mathcal{X} = \begin{bmatrix} 82.3 & 1205 & 4.2 & 7.1 & 156 \\ 85.1 & 1198 & 4.1 & 7.2 & 162 \\ 79.8 & 1210 & 4.3 & 6.8 & 145 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad : \quad \text{cols} \Rightarrow \begin{array}{l} \text{Temperature (°C)} \\ \text{Pressure (kPa)} \\ \text{Flow (L/s)} \\ \text{pH} \\ \text{Conductivity } (\mu\text{S/cm}) \end{array}$$

每一行代表一个测量值，但变量在尺度上有显著差异。直接应用协方差主成分分析（PCA）将完全受到压力测量的主导，而 pH 值或流量等潜在重要变化可能会在噪声中消失。然而盲目地对每个变量进行标准化可能会丢失有意义的尺度信息。

例子 11.8（规模效应）。对于上述制造数据，不同预处理选择下的第一个主成分揭示了截然不同的模式：

协方差PCA得到$v_1 \approx (0.002, 0.999, 0.001, 0.000, 0.003)^T$，基本上只捕捉到压力变化。在标准化为单位方差后，相关PCA给出$v_1 \approx (0.51, 0.48, -0.42, 0.38, 0.44)^T$，揭示了所有测量值之间的协调变化。选择方法从根本上决定了我们能发现哪些模式。

◇

超越尺度的限制，真实数据受到测量误差、传感器故障以及极端但真实事件的污染。我们需要系统的方法来识别需要特殊处理的观测值。关键的见解在于，不仅仅根据原始值来衡量距离，而是以一种能够考虑到数据中自然尺度和关系的方式来衡量：

定义 11.9（马哈拉诺比斯距离）。对于来自具有均值$\bar{x}$和协方差 $[C]$ 的集合中的观测值 x，*Mahalanobis distance* 为：

$$d_M(x) = \sqrt{(x - \bar{x})^T [C]^{-1} (x - \bar{x})}$$

Nota bene: 矩阵 $[C]^{-1}$ 在高维中起到平方倒数标准差 $1/\sigma^2$ 的作用，考虑到变量之间的尺度和相关性。

该度量同时考虑了量级和相关性结构，在衡量观测值与典型模式的偏差程度时。逆协方差矩阵 $[C]^{-1}$ 确保了距离能够正确反映每个方向上的自然变异性。

•

对于我们的工程应用，这个距离提供了一种自然的方式来识别明显偏离典型模式的观测值。经验表明，马氏距离大于典型点的两倍的观测值值得仔细调查。

Think: 在一维中，d_M 简化为以标准差为单位测量的与均值的距离：$d_M(x) = |x - \mu|/\sigma$。这与我们对简单测量中“异常”值的直觉相关。

示例 11.10（异常值检测）。回到我们的制造数据，大多数观察值的马哈拉诺比斯距离在 1.5 到 3 个单位之间。然而，有一个测量值：

$$x = (84.2, 1203, 12.8, 7.0, 158)^T$$

产量 $d_M = 4.9$，远超出典型范围。虽然每个单独的测量结果看似合理，但它们的结合暗示了一个过程

需要调查的异常情况。12.8 L/s 的流量在绝对值上并不极端，但与观测到的温度和压力不一致。 ◇

缺失测量构成了最后一个挑战。当缺失相对罕见（< 的条目中占 5%）时，直接省略受影响的观测即可。更大范围的缺失需要谨慎处理，以避免使我们的分析产生偏倚。简单的均值插补——用变量的平均值替换缺失值——往往会扭曲相关结构。更为复杂的迭代方案利用部分 PCA 结果来估计缺失值，但这类方法存在施加人为模式的风险。

预处理的选择从根本上决定了 PCA 能够发现哪些模式。不同的选择服务于不同的目的：

- 当绝对尺度具有意义时，协方差 PCA 会保留这些绝对尺度
- 相关性 PCA 揭示纯粹的关系型模式
- 稳健方法为了在受污染的数据下保持可靠性而牺牲效率

在对数据质量进行细致审查并以清晰的分析目标为指导的前提下，审慎的工程判断必须为这些选择提供依据。

11.6 Statistical Significance

对我们中心化的数据矩阵 $\mathcal{X}$ 的奇异值分解不仅提供了用于降维的最优方向，也给出了这些方向重要性的自然度量。每一个奇异值 σ_i 都量化了其对应方向对数据结构的贡献强度——这与第9章中提出的支配性概念完全类比。关键在于确定在何处截断这一序列，在表示的保真度与模型的简洁性之间取得平衡。

考虑使用逐渐增加的奇异值数量对 $\mathcal{X}$ 的一系列部分近似：

$$\mathcal{X}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

保留变异与总变异的比率为我们提供了衡量近似质量的第一个指标：

$$r_k = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{d} \sigma_i^2}$$

这个量，类似于第9.2节中的谱比，用于衡量 k 个成分在多大程度上捕捉了数据的本质结构。

例 11.11（振动分析）。考虑来自桥梁结构上传感器的加速度测量，得到如下奇异值：

$$\sigma_1 = 12.5, \quad \sigma_2 = 9.5, \quad \sigma_3 = 3.6, \quad \sigma_4 = 1.3, \quad \sigma_5 = 1.1, \ldots$$

在 σ_2 之后的急剧下降表明存在两种占主导地位的振动模态。正如第9章中占主导的特征值支配了渐近行为一样，这些占主导的奇异值识别出对桥梁运动至关重要的方向。前两个分量捕捉比例

$$r_2 = \frac{\sigma_1^2 + \sigma_2^2}{\sum_{i=1}^{d} \sigma_i^2} \approx 0.85$$

的 全变差——对其主导性的定量度量 这。 ◇

Nota bene: 在分析被噪声污染的数据时，奇异值的急剧下降往往标志着从信号到噪声的过渡——这一模式可以通过随机矩阵理论得到理论解释，但即使在简单示例中也清晰可见。

这一奇异值序列与最优逼近理论直接相关。根据第11.4节，每个 $\mathcal{X}_k$ 在平方误差意义下为 $\mathcal{X}$ 提供最佳的秩-k 逼近。因此，截断水平 k 在逼近精度与模型复杂度之间进行权衡——这种权衡通过奇异值谱得以量化。

例 11.12（化学过程数据）。通过多个传感器监测的化学反应器得到的归一化奇异值下降得更为平缓：

$$\sigma_1 = 2.05, \quad \sigma_2 = 1.76, \quad \sigma_3 = 1.52, \quad \sigma_4 = 1.18, \quad \sigma_5 = 0.89, \ldots$$

未出现明显的主导性，反映了过程变量之间复杂的耦合关系。累计比例 $r_4 \approx 0.85$ 表明，保留四个成分即可捕获最显著的变异，同时过滤传感器噪声。 ◇

在实践中，保留多少个成分的选择受益于系统性的验证。通过将数据划分为训练集和测试集，我们可以评估不同截断水平在新观测上的泛化能力。在数据某一部分中占主导地位的成分，应当在其他部分中也保持其主导性——这一原则有助于区分真实结构与抽样伪影。

示例 11.13（材料光谱学）。考虑光谱测量情形，其中每个样本在不同波长上产生数千个强度读数。在样本数量有限的情况下，区分显著成分变得至关重要。交叉验证表明，尽管奇异值在 σ_{20} 之后仍然持续存在，但使用超过 5-6 个成分进行预测往往在留出数据上表现更差——这表明尽管在训练数据中看似占据主导地位，仍然发生了过拟合。◇

样本量从根本上影响我们对所识别成分的信心。在分析 d 个变量的 n 个观测时，超过索引 $\min\{n, d\}$ 的奇异值必定为零——这一事实直接源自 SVD 的矩阵结构。更微妙的是，n/d 的比值会影响非零奇异值的稳定性：相对于变量数量，观测值过少会产生虚假的表观结构。

这些考虑将我们对SVD主导性的理论理解转化为数据分析的实际指导。尽管缺乏纯矩阵代数的确定性，但它们提供了系统的方法来识别真正主导的成分，同时防止过度解释。真正的考验不在于抽象的意义，而在于所保留的成分是否能更好地预测、控制或理解我们研究的系统。

11.7 Beyond Linear PCA

现实很少屈从于纯粹的线性描述。尽管主成分在空间的线性变换中揭示了结构，许多数据集却蕴含着内在的非线性——它们的本质模式像藤蔓一样扭曲和弯曲，通过测量空间生长，超越了它们的线性支持。温度传感器的读数可能随着时间呈正弦波变化；化学反应速率通过非线性动力学耦合；手写数字的图像描绘了复杂的流形，远离任何线性子空间。理解这些数据需要将我们的数学框架扩展到超越目前支撑我们发展的线性领域。

先考虑为什么线性 PCA 可能会失败。当数据位于弯曲的表面或流形附近时，任何线性投影都无法捕捉其真实结构。主成分尽管在线性近似意义下是最优的，却可能完全错失潜在的简洁性。就像弯曲金属丝投下的影子看起来纠缠并自相交一样，线性投影往往会遮蔽而非揭示非线性数据的真实形态。

从这些例子中得出的关键见解是，我们必须调整降维的概念，以尊重数据的局部结构。与其寻求全局线性投影，不如通过切空间在局部上逼近数据，切空间沿着数据的曲线和弯曲。这个观点——即非线性结构在足够小的尺度上通常表现得几乎是线性的——为现代流形学习方法提供了基础。

示例 e 11.14 (单细胞基因组学)。考虑基因表达 n mea-

来自正在经历分化的细胞的测量结果。尽管这些数据是在成千上万维的空间中测得的（每个基因一维），但发育过程通常随着细胞从类干细胞状态走向特化类型而沿着分支路径推进。任何线性投影都无法捕捉这种分支结构——PCA 可能只会呈现一团纠缠的混乱，而其中实际上存在具有生物学意义的进程。只有当我们尊重其弯曲、分支的几何结构时，数据真正的简洁性才会显现出来，从而揭示出与已知生物学相呼应的发育轨迹。◇

超越线性的一条途径是通过 *kernel methods*。我们不再直接在测量空间中工作，而是通过核函数 $k(\boldsymbol{x}、\boldsymbol{y})$ 将数据隐式地映射到一个更高维的特征空间，用以衡量观测之间的相似性。在这一扩展后的空间中，我们执行标准的主成分分析（PCA）——不过此时是作用于核矩阵 $K = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]$，而非原始测量的协方差。

Foreshadowing: 内核方法的隐式特征映射预示了神经网络如何学习表示，将数据转化为更有意义的空间。

该 *kernel PCA* 通过精心选择核函数揭示了非线性结构。径向基核

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\sigma^2}\right)$$

度量局部相似性，使该方法能够跟随数据中的弯曲模式。尽管在大型数据集上计算开销很大，核PCA在数学上提供了线性方法与非线性方法之间的一座优雅桥梁。

BONUS! 除了此处讨论的方法之外，拓扑数据分析（TDA）还提供了用于理解数据形状的工具。虽然这超出了我们当前的范围，TDA 通过诸如 *persistent homology* 等线性代数工具，为数据几何提供了强有力的洞见。

通过考察线性 PCA 如何优化重构误差，可以得到一种不同的思路。与其将自己限制在线性映射上，不如寻找能够在压缩数据的同时保留其本质结构的一般函数。这种视角自然引出了 *au- toencoders*——通过一个低维瓶颈层来重构其输入而训练的神经网络。尽管我们将其详细讨论推迟到第13章，但自编码器代表了对 PCA 降维原理的一个根本性推广。

局部方法提供了另一条超越线性的路径。与其寻找全局结构，诸如 *locally linear embedding* (LLE) 之类的技术通过从其邻居重构每个点来保持几何结构。如果 $\boldsymbol{x}_i$ 表示我们的第 i 个观测，而 $\mathcal{N}(i)$ 表示其最近邻集合，LLE 首先通过求解以下问题来计算权重 w_{ij}:

$$\min_{w_{ij}} \sum_{i=1}^{n} \left\| \boldsymbol{x}_i - \sum_{j \in \mathcal{N}(i)} w_{ij} \boldsymbol{x}_j \right\|^2 \quad \text{subject to} \quad \sum_{j \in \mathcal{N}(i)} w_{ij} = 1$$

这些 w 八次捕捉通过线性近似的局部几何。 tions. A

第二次优化随后找到低维坐标 $\boldsymbol{y}_i$，以保持这些关系：

$$\min_{\boldsymbol{y}_i} \sum_{i=1}^{n} \left\| \boldsymbol{y}_i - \sum_{j \in \mathcal{N}(i)} w_{ij} \boldsymbol{y}_j \right\|^2$$

Think: LLE的两阶段优化反映了流形如何在局部呈线性——首先捕捉局部结构，然后找到尊重这些局部模式的全局坐标。

这个从局部到全局的原则——即复杂结构在适当分解时通常会简化为更简单的部分——在数学和工程中随处可见。正如积分将复杂函数简化为简单部分的总和，或有限元通过简单的补丁近似复杂的区域，流形学习方法通过局部近似构建全球理解。

这些非线性方法的出现标志着我们对降维的思考方式发生了深刻的转变。不再局限于线性投影，我们可以寻求真正捕捉数据在空间中如何自我组织的表示。这种自由既带来了力量，也带来了挑战——我们在表达能力上获得了提升，但牺牲了独特性和计算简便性。实践中，方法的选择受到这些因素平衡的影响：

- 核主成分分析（Kernel PCA）提供了数学优雅性，但扩展性较差
- 局部方法捕捉细节结构，但可能忽略全局模式
- 自编码器提供了灵活性，但需要大量数据和计算

然而，这些挑战不应掩盖一个基本的洞察：即降维，正确理解的含义，是找到能保留本质结构的简化表示。无论是通过核函数、本地近似，还是学习到的变换，目标始终是将复杂数据提炼到其有意义的本质。这一原则——高维数据通常具有等待被发现的低维结构——将引导我们在神经网络及更远领域的发展。

从线性PCA到现代非线性方法的路径回顾了数学思维中的一个更广泛的模式。正如经典线性代数发展到涵盖无限维空间和非线性算子，我们理解数据的工具也必须超越最初产生它们的线性框架。我们探索过的方法代表了朝向这一更广阔愿景迈出的第一步——这一愿景将在我们深入神经网络和人工智能的第13章时继续展开。

Decoding Neural Population Activity

神经元群体的协同放电通过电活动模式对复杂行为进行编码。现代实验技术使得能够同时记录数百个神经元的活动，每个神经元都贡献一个随时间变化的信号，既反映局部计算，也反映全脑状态。尽管单个神经元呈现出复杂且常常含噪的动力学，其群体活动在通过降维的视角进行分析时却展现出异常清晰的结构。

考虑一个典型的运动皮层实验，在该实验中，研究人员记录了 $n = 256$ 个神经元的活动，同时被试进行朝向不同目标的到达运动。每个神经元的放电率随着时间变化，产生一个在毫秒分辨率下采样的瞬时群体活动向量 $\boldsymbol{r}(t) \in \mathbb{R}^n$。在一个包含 100 次持续 500 毫秒的到达运动的实验过程中，我们得到中心化的数据矩阵：

$$R = \begin{bmatrix} \leftarrow & \boldsymbol{r}(0) - \bar{\boldsymbol{r}} & \rightarrow \\ \leftarrow & \boldsymbol{r}(1) - \bar{\boldsymbol{r}} & \rightarrow \\ & \vdots & \\ \leftarrow & \boldsymbol{r}(50000) - \bar{\boldsymbol{r}} & \rightarrow \end{bmatrix} \in \mathbb{R}^{50000 \times 256}$$

其中 $\bar{\boldsymbol{r}}$ 表示时间平均放电率向量。样本协方差矩阵 $[C] = \frac{1}{T-1} R^T R \in \mathbb{R}^{256 \times 256}$ 捕捉了神经元对在放电模式中的共变情况。

一个具体的例子揭示了这份神经数据中显著的结构。在最近的一次实验中，归一化的神经活动矩阵的前十个奇异值为：

$$\sigma_1 = 128.4, \quad \sigma_2 = 84.2, \quad \sigma_3 = 52.1, \quad \sigma_4 = 31.5, \quad \sigma_5 = 18.7$$

$$\sigma_6 = 12.3, \quad \sigma_7 = 8.1, \quad \sigma_8 = 5.4, \quad \sigma_9 = 3.8, \quad \sigma_{10} = 2.6$$

由前 k 个分量解释的方差比例 源文本：ts：翻译文本：

$$r_k = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{256} \sigma_i^2}$$

仅用五个成分就达到了$r_5 = 0.86$。这种显著的降维——在保留86%方差的同时，从256个神经元降至5个主成分——表明神经计算发生在一个远低于原始细胞数量所暗示的维度空间中。

主成分本身作为 $[C]$ 的特征向量计算得出，揭示了运动编码的不同方面。第一个成分 $\boldsymbol{v}_1$ 显示出整个群体中的协调激活：

$$\boldsymbol{v}_1 = \begin{bmatrix} 0.31 \\ 0.28 \\ 0.33 \\ \vdots \\ 0.29 \end{bmatrix} \quad \text{with entries} \quad [\boldsymbol{v}_1]_j \approx \frac{1}{\sqrt{n}} \pm 0.1$$

这种近乎均匀的加权表明一种全局活动模式，或许反映了整体运动的活力。第二个成分呈现出清晰的解剖学组织：

$$\boldsymbol{v}_2 = \begin{bmatrix} 0.42 \\ 0.38 \\ -0.35 \\ \vdots \\ -0.41 \end{bmatrix}$$

对偏好前向伸手的神经元赋予正权重，对偏好后向伸手的神经元赋予负权重
。

为了将这些成分与行为进行验证，我们可以在每个时间 t 将神经数据投影到我们的主成分上：

$$z_i(t) = \boldsymbol{v}_i^T(\boldsymbol{r}(t) - \bar{\boldsymbol{r}})$$

所得评分$z_i(t)$揭示了神经轨迹与运动之间的紧密耦合：

- $z_1(t)$ 与运动速度（$r = 0.82$）高度相关
- $z_2(t)$ 预测到达方向 ($r = 0.76$ 与目标角度)
- $z_3(t)$ 和 $z_4(t)$ 捕捉握力调节

与第 11.4 节中的制造数据类似，神经记录需要仔细的预处理。常见的挑战包括：

1. 神经元之间放电率高度可变（有些细胞的放电率比其他细胞高出100倍） 2. 由于脑状态变化导致的非平稳基线活动 3. 来自共享输入的时间相关噪声 4. 由于神经信号短暂丢失而产生的缺失数据

在协方差和基于相关性的PCA之间的选择尤为关键。对于上述数据，基于相关性的分析通过防止高度活跃的神经元主导，揭示了额外的结构：

$$[\tilde{C}]_{ij} = \frac{[C]_{ij}}{\sqrt{[C]_{ii}[C]_{jj}}}$$

基于相关性的主成分强调神经元之间的协同，而不受其绝对放电率的影响。这通常更能反映潜在的计算过程，因为诸如电极放置等神经元特异性的因素可能会人为地抬高某些放电率。

现代实验通常同时记录来自多个大脑区域的数据。考虑来自初级运动皮层（M1）和背侧前运动皮层（PMd）的数据：

$$R = \begin{bmatrix} R_{M1} \\ R_{PMd} \end{bmatrix} \quad \text{where} \quad R_{M1} \in \mathbb{R}^{50000\times 256} \quad \text{and} \quad R_{PMd} \in \mathbb{R}^{50000\times 128}$$

PCA o 在此，合并的数据揭示了既有区域特定的，也有跨区域的 模式。

一些主成分的载荷主要集中在单一领域:

$$v_1 = \begin{bmatrix} v_{M1} \\ \mathbf{0} \end{bmatrix} \quad \text{or} \quad v_2 = \begin{bmatrix} \mathbf{0} \\ v_{PMd} \end{bmatrix}$$

而其他则反映了协同计算:

$$v_3 = \begin{bmatrix} v_{M1} \\ v_{PMd} \end{bmatrix} \quad \text{with} \quad \|v_{M1}\| \approx \|v_{PMd}\|$$

PCA揭示的低维结构直接影响大脑-
计算机接口 (BCI) 设计。现代 BCI 并不是试图从单个神经元解码预期的运动，而是首先将神经活动投影到数据驱动的主成分上:

$$\hat{x}(t) = W \begin{bmatrix} v_1^T(r(t) - \bar{r}) \\ \vdots \\ v_k^T(r(t) - \bar{r}) \end{bmatrix}$$

其中 W 将 k 主成分得分映射到解码的运动变量$\hat{x}(t)$。这种方法证明了极高的鲁棒性——即使个别神经元丧失，主成分通常仍然保持稳定，因为可以利用剩余神经元群体的协调活动。

统计显著性框架对于验证低维神经结构至关重要。考虑将我们的伸手数据划分为训练集和测试集:

$$R_{train} = R_{1:40000,:} \quad \text{and} \quad R_{test} = R_{40001:50000,:}$$

在R_{train}上计算主成分，并在R_{test}上评估解释方差，有助于区分可靠模式与过拟合。当神经维度在不同数据划分中与行为持续一致地对齐，同时保持较高的解释方差时，我们对其计算相关性更有信心。

除了基础研究之外，降维还推动了复杂的临床应用。使用脑机接口（BCI）的瘫痪患者需要在神经活动存在变异的情况下获得可靠的控制信号。基于 PCA 的解码器通过提取与运动相关的稳定维度，即使单个神经元的调谐特性发生变化，也能保持性能。这种数学理论与临床实践的融合，生动地展示了降维如何在加深我们对神经计算理解的同时，促进实用的神经工程应用。

对神经元群体活动的分析既展示了 PCA 的强大能力，也揭示了其局限性。尽管线性方法揭示了引人注目的结构，但大脑内在的非线性与动力学表明，仍有更为丰富的模式有待发现。就目前而言，PCA 为我们提供了观察神经元群体如何共同编码行为的最清晰窗口——这充分证明了降维在现代神经科学中的深远价值。

Eigenfaces

人类大脑在识别面孔方面的非凡能力长期以来激发了数学方法在图像分析中的应用。在这些方法中，*eigenfaces* 以其优雅的简洁性和对计算机视觉的深远影响而脱颖而出。尽管现代深度学习方法在实际应用中已取而代之，但特征脸或许最清晰地展示了 PCA 如何从高维数据中提取有意义的结构。

考虑一幅大小为 $m \times n$ 像素的灰度图像。尽管我们自然会将其视为一个由强度值组成的矩形数组，但我们可以将其重塑为 $\mathbb{R}^{mn}$ 中的一个单一向量。一幅分辨率仅为 100×100 的人脸图像因此成为 $\mathbb{R}^{10000}$ 中的一个点——这是一个维度高得令人望而却步、无法进行直接分析的空间。然而，人脸尽管复杂，却表现出强烈的统计规律性，PCA 可以将其揭示出来。

给定一组 N 张人脸图像 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$，每一张都被重塑为一个向量，我们首先通过减去平均人脸来对数据进行中心化：

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \quad : \quad \tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \bar{\boldsymbol{x}}$$

经中心化的图像构成数据矩阵 $\mathcal{X} \in \mathbb{R}^{(mn)\times N}$ 的各列。其奇异值分解 $\mathcal{X} = U\Sigma V^T$ 得到的特征脸是 U 的各列，对应于最大的奇异值。

这些特征脸在重新变回图像格式时，呈现出引人注目的结构。前几个捕捉的是大尺度特征——整体脸型、眼睛和嘴巴的位置，以及主要的光照变化。随后的特征脸编码逐渐更精细的细节，较后的分量往往表示噪声或个体化特征。通常只需40–50个分量就足以捕捉人脸图像的基本结构，从而实现两个数量级的降维。

任何人脸图像 $\boldsymbol{x}$ 都可以通过投影到前 k 个特征脸上来近似表示：

$$\boldsymbol{x} \approx \bar{\boldsymbol{x}} + \sum_{i=1}^{k} c_i \boldsymbol{u}_i \quad \text{where} \quad c_i = (\boldsymbol{x} - \bar{\boldsymbol{x}})^T \boldsymbol{u}_i$$

系数 c_i 构成了一个 *face code*——一种低维表示，用于捕捉图像的关键特征。这种编码实现了高效的人脸识别：相似的人脸会产生相似的系数，从而可以在降维后的空间中通过简单的距离度量进行识别。

人脸图像的预处理对于有效分析至关重要。除了基本的居中之外，还需要仔细关注：

1. 面部特征的对齐与缩放 2. 光照归一化 3. 背景去除 4. 面部表情处理

对所得特征脸的质量产生显著影响。人脸空间中的马氏距离提供了一种自然的方法，用于检测与典型人脸显著偏离的图像——这是在实际系统中评估可靠性的一项有价值的测试。

现代人脸识别在很大程度上已经转向能够处理姿态、光照和表情更大变化的深度学习方法。然而，特征脸

该方法的核心洞见——高维人脸图像位于一个低维流形附近——仍然成立。事实上，现代神经网络的中间层往往学习到与特征脸（eigenfaces）惊人相似的表示，这表明这些模式反映的是基本结构，而非仅仅是算法伪影。

BONUS! 特征脸（Eigenface）技术可以自然地扩展到其他类型的图像分析。研究人员已将类似的方法应用于医学图像（用于癌症检测的特征肿瘤）、卫星数据（用于地形分类的特征景观）以及工业检测（用于质量控制的特征缺陷）。

除了其实用影响之外，特征脸方法还展示了线性代数如何将抽象理论转化为工程解决方案。第10章的奇异值分解成为降维的实用工具；第11.5节的预处理原则得到具体应用；第11.6节的统计考量指导实现选择。这种——数学、统计学与工程的——综合，使机器开始接近人类在模式识别和场景理解方面的能力。

□ ——————————————————————————— □

Exercises: Chapter 11

1. 以下测量数据来自两个用于监测一个简单过程的传感器：

$$\begin{bmatrix} 1.0 & 0.8 \\ 0.8 & 2.0 \\ -0.5 & -0.2 \\ -1.2 & -1.5 \end{bmatrix}$$

计算样本均值和协方差矩阵。求主成分，并创建一个散点图，展示原始数据点以及第一主成分的方向。该主成分解释了多少比例的方差?

2. 一个 3×3 相关矩阵的特征值为 2.4、0.5 和 0.1。在未看到原始矩阵的情况下：

（a）求其迹和行列式 （b）计算第一主成分所捕获的方差比例 （c）你可以就原始变量之间的相关性得出什么结论?

3. 给定中心化的数据矩阵 $\mathcal{X} \in \mathbb{R}^{n \times 3}$，其列分别表示温度、压力和浓度测量值：

（a）写出用于计算主成分得分的显式公式（b）说明如何将新的测量值 $\boldsymbol{x}$ 投影到这些成分上（c）解释为何在进行该计算之前进行标准化可能很重要

4. 一个化学反应器通过三个传感器进行监测，分别测量温度（°C）、压力（kPa）和流量（L/s）。这些测量得到协方差矩阵

$$[C] = \begin{bmatrix} 16 & 8 & 0 \\ 8 & 25 & -4 \\ 0 & -4 & 9 \end{bmatrix}$$

求其特征值和特征向量。前两个主成分捕获了总方差的多少比例? 给出这些主成分可能揭示反应堆行为的物理解释。

5. 一个振动分析系统在机器上的三个点记录位移，从而得到协方差矩阵

$$[C] = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 6 \end{bmatrix}$$

对于测量 $\boldsymbol{x} = (0.5, 2.8, -1.2)^T$:

（a）计算其在前两个主成分上的投影

(b) 向量的能量保留了多少比例?

(c) 首先标准化变量会如何改变你的分析?

6. 考虑一系列逐个到达的测量值。描述一种高效的算法来:

(a) 增量更新样本均值 (b) 在不存储所有历史数据的情况下更新协方差矩阵 (c) 在需要时定期更新主成分

为什么这种增量方法在实践中可能具有价值?

7. 在分析半导体制造时，三个质量指标生成相关矩阵

$$[R] = \begin{bmatrix} 1.0 & 0.8 & 0.7 \\ 0.8 & 1.0 & 0.9 \\ 0.7 & 0.9 & 1.0 \end{bmatrix}$$

找到其主要成分并解释它们揭示了关于制造过程的哪些模式。是什么可能解释所有变量之间的强相关性?

8. 来自材料在循环加载下的应变测量时间序列产生特征值:

$$\sigma_1 = 12.3, \quad \sigma_2 = 5.1, \quad \sigma_3 = 0.8, \quad \sigma_4 = 0.3, \quad \sigma_5 = 0.2$$

根据这个衰变模式，您会保留多少个分量？请从数学角度和材料行为的物理意义上说明您的答案。

9. 安装在桥梁结构上的四个加速度计记录垂直位移（毫米）与协方差矩阵

$$[C] = \begin{bmatrix} 10 & -2 & 4 & 1 \\ -2 & 8 & 0 & 3 \\ 4 & 0 & 6 & -1 \\ 1 & 3 & -1 & 5 \end{bmatrix}$$

为了捕获总方差的90%，你会保留多少个主成分？这些主导模态在桥梁振动模式方面可能具有怎样的物理解释?

10. 考虑监测一个化学过程，其中温度范围为 150–200 °C，压力为 2000–2500 kPa，流量为 0.5–2.0 L/s。解释为什么协方差主成分分析（PCA）在这里可能会产生误导，并提出一种预处理策略。你如何通过交叉验证来验证你的选择?

11. 在监测一座工业炉时，来自五个传感器的温度测量值（°C）得到的均值向量为 $\bar{\boldsymbol{x}} = (845, 835, 840, 850, 855)^T$，以及测量值 $\boldsymbol{x} = (850, 823, 841, 868, 859)^T$，其协方差矩阵为

$$[C] = \begin{bmatrix} 100 & 85 & 82 & 75 & 70 \\ 85 & 120 & 90 & 80 & 75 \\ 82 & 90 & 110 & 85 & 80 \\ 75 & 80 & 85 & 90 & 75 \\ 70 & 75 & 80 & 75 & 95 \end{bmatrix}$$

计算 $\boldsymbol{x}$ 相对于均值的马氏距离。如果典型测量的马氏距离小于 3，那么这次读数是否值得进一步调查？

12. 设 $\mathcal{X}$ 为一个中心化的数据矩阵。证明：相关矩阵的主成分（在将每个变量标准化为单位方差之后得到）与协方差矩阵的主成分不同，除非所有原始变量具有相同的方差。在实践中，什么时候你会更倾向于使用基于相关矩阵的 PCA 而不是基于协方差矩阵的 PCA？

13. 对于马氏距离 $d_M^2(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}})^T [C]^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}})$，证明它在保持协方差结构的线性变换下是不变的。说明其等值集是与主成分对齐的椭球。

14. 一个制造质量检测系统对每个零件记录五个测量值。经中心化的数据矩阵的奇异值递减如下：

$$\sigma_k = 10 e^{-0.8k}, \quad k = 1, \dots, 5$$

证明保留前两个成分可以捕获至少 95% 的总方差。奇异值的指数衰减暗示了制造变异的内在维度性。

15. 考虑在执行 PCA 之前将数据划分为训练集和验证集的问题。请解释：(a) 如何仅使用训练数据正确地计算主成分；(b) 如何在验证数据上度量重构误差；(c) 为什么这种交叉验证方法能更好地估计泛化误差。

16. 考虑一个数据集，其中大多数测量值聚集在均值附近，但偶尔会出现异常值。解释为何基于协方差矩阵的PCA可能会受到这些异常值的过度影响。提出一种使用第11.5节中的概念的鲁棒替代方法。你如何验证你的方法更好地捕捉了真实的数据结构？

17. 化学过程中的测量值相关矩阵几乎是块对角的，块内有较强的相关性，而块之间的相关性较弱。解释这种结构揭示了过程的哪些信息？如何利用这些信息设计基于PCA的分层监控策略？

18. 解释数据矩阵的奇异值谱如何揭示潜在过程的内在维数。你将如何区分：(a) 真实的低维结构；(b) 由有限采样导致的随机相关性；以及 (c) 系统性的测量伪影。请结合第 11.6 节中的相关概念来支持你的回答。

Chapter 12
Low Rank Approximation

“do I not stretch the heavens abroad or fold them up like a garment”

一个满秩矩阵蕴含着微妙形式的冗余——被细节的繁杂所掩盖的模式，结构被数据的极度丰富所遮蔽。矩阵的基本特征往往不在于其表面上的复杂性，而在于其潜藏在底层的简洁形式。这个洞察——矩阵可以通过较低秩的版本来进行逼近——改变了我们理论上的理解和实际计算。艺术在于找到这种逼近方式，既保留了基本特征，又去除了多余部分。

第10章中开发的奇异值分解为我们提供了第一次如何通过更简单的版本逼近矩阵的机会。就像雕塑家通过去除多余的材料来揭示形态一样，我们通常通过仅保留矩阵的主导奇异值和向量来捕捉矩阵的本质特征。这种低秩近似的过程——即寻找复杂数据的最佳简单表示——与矩阵分析的抽象理论以及现代计算的实际挑战有着深刻的联系。

考虑一个矩阵 A，其条目表示数字图像中的像素强度。尽管在形式上具有较高的秩（每个条目可能是独立的），但真实图像的结构化特性通常允许从远少于的参数中进行惊人的精确重建。本章中开发的理论基础解释了为什么这种压缩是可能的，以及如何实现最优压缩。这些相同的原理指导着数据分析、信号处理以及新兴的矩阵补全领域中的降维 – 在这些领域中，我们寻求从部分观测中恢复完整矩阵。

前方的道路从抽象的近似理论通过实际的算法到现代应用。我们从有趣的...

低秩近似的根本极限，阐明了什么是可能的，什么是不可能的。这自然引出了能够高效实现这些极限的计算方法，即便对于直接计算不可行的海量数据集亦然。该理论还扩展到处理受污染或不完整的数据，反映了真实世界应用的杂乱现实。贯穿始终，我们坚持这样的观点：有效的近似既需要数学理解，也需要可靠的工程判断。

12.1 Optimal Approximation

第10章中介绍的奇异值分解揭示了任何矩阵都可以表示为若干秩一项之和，每一项都刻画了数据中不同的变化方向。这种分解暗示了一条自然的近似途径：与其保留所有项，不如只保留那些对应于最大奇异值的项。这样的截断提出了关于最优性的根本问题——矩阵能够在多大程度上被其更简单的版本所近似，以及这些近似应采取何种形式?

让我们首先形式化我们所说的秩约束近似：

定义 12.1（秩-k 近似）。对于矩阵 $A \in \mathbb{R}^{m\times n}$ 和整数 $k \leq \text{rank}(A)$, *rank-k approximation* 是任意满足 $\text{rank}(B) = k$ 的矩阵 $B \in \mathbb{R}^{m\times n}$。这种近似的误差通常使用第 5 章中引入的矩阵范数来度量，特别是：

1. The Frobenius norm: $\|A - B\|_F = \sqrt{\sum_{i,j}(a_{ij} - b_{ij})^2}$
2. The spectral norm: $\|A - B\|_2 = \sigma_1(A - B)$

其中 $\sigma_1(A - B)$ 表示差分矩阵的最大奇异值。

•

给定矩阵A及其奇异值分解$A = U\Sigma V^T$，我们可以通过仅保留前k个奇异值来构造一个自然的秩-k近似：

$$A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

这种截断的 SVD 得到一个秩恰好为 k 的矩阵，因为每一项 $\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$ 都是秩一的，并且这些项线性组合。该近似的误差呈现出一种尤为优雅的形式：

Compare: 奇异值的截断与第11章中 PCA 的降维相呼应，但这里是通过矩阵近似而非统计学的视角来审视。

$$\|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$$

其中 $r = \text{rank}(A)$. 每个被截断的奇异值都会恰好以其平方对总近似误差作出贡献。

这种构造的深远重要性源于其最优性性质：

定理 12.2（Eckart-Young-Mirsky）。 *Let $A \in \mathbb{R}^{m\times n}$ have singular value decomposition $A = U\Sigma V^T$. Then for any matrix B of rank at most k:*

$$\|A - A_k\|_F \leq \|A - B\|_F$$

That is, the truncated SVD A_k provides the best possible rank-k approximation to A in the Frobenius norm.

Historical Note: 尽管它因20世纪30年代的相关工作而得名，但矩阵通过 SVD 获得最优低秩近似的核心洞见源自天文学；1901年，卡尔·皮尔逊在其中用它将平面拟合到含噪数据。

Proof. 设 B 的秩至多为 k，并考虑矩阵 $A - B$。根据基本定理（定理 3.27），其零空间的维数至少为 $n - k$。该空间必然与由 $\{\boldsymbol{v}_{k+1}, \ldots\ldots, \boldsymbol{v}_n\}$ 张成的子空间有非平凡的交集，而该子空间的维数也为 $n - k$。令 $\boldsymbol{x}$ 为该交集中的一个单位向量。

那么 $\|A - B\|_F^2 \geq \|(A - B)\boldsymbol{x}\|^2 = \|A\boldsymbol{x}\|^2$，因为 $B\boldsymbol{x} = 0$。但 $A\boldsymbol{x}$ 位于 span$\{\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_r\}$ 中，并且其范数至少为 σ_{k+1}。因此：

$$\|A - B\|_F^2 \geq \sigma_{k+1}^2 = \|A - A_k\|_F^2$$

截断SVD最优性的证明 □

这一卓越的结果不仅适用于 Frobenius 范数，而且推广到任何酉不变的矩阵范数——包括谱范数 $\|A\|_2 = \sigma_1$ 和核范数 $\|A\|_* = \sum_i \sigma_i$，我们将在第 12.4 节中加以探讨。其几何直觉依然成立：A_k 捕捉了数据中最重要的 k 个变化方向，这些方向由奇异值来度量。

例12.3（图像压缩）。考虑一幅以像素强度存储为 1024×1024 矩阵 A 的灰度图像。尽管形式上秩为1024，但大多数奇异值都小到可以忽略。一个秩为50的近似 A_{50} 通常能够捕捉到视觉上重要的特征，同时将存储量从超过一百万个数值减少到仅：

- 50 个奇异值（$\sigma_1, \ldots, \sigma_{50}$）
- 50 个左奇异向量（$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{50}$）
- 50 个右奇异向量 ($\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{50}$)

总计约100,000 个数值——在几乎没有感知损失的情况下，存储量减少了 90%。 ◇

截断 SVD 不仅仅提供最优近似——它揭示了矩阵近似的基本极限。对于任意秩至多为 k 的矩阵 B：

$$\|A - B\|_2 \geq \sigma_{k+1}(A)$$

这个由 A_k 达到的下界，精确地量化了降秩矩阵能够多好地近似 A。这种关系不仅限于谱范数，还扩展到其他重要的矩阵范数：

引理 12.4（近似界）。*For the truncated SVD approximation A_k:*

1. $\|A - A_k\|_2 = \sigma_{k+1}$
2. $\|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$

Moreover, these bounds are optimal: no rank-k matrix can achieve smaller error in any of these norms.

这些最优逼近结果的力量，通过与受约束的逼近问题进行对比而最为清晰地显现。当我们要求额外的性质——精确保持某些条目，或维持原矩阵的非负性——闭式解通常不再存在。然而，这类受约束的问题仍然可以通过数值优化来处理，我们将在第12.3节中探讨相关技术。

最优逼近的深远影响远不止于简单的数据压缩。通过揭示矩阵逼近的基本极限，Eckart–Young–Mirsky 定理为以下方面提供了理论基础：

- 统计学中的主成分分析
- 自然语言处理中的潜在语义分析
- 动力系统中的模型降阶
- 推荐系统中的协同过滤

在每一种情况下，最优低秩近似的存在将看似复杂的问题转化为可管理的计算。艺术之处不在于寻找这些近似——截断的 SVD 会自动给出它们——而在于选择合适的秩 k，以在精度与简洁性之间取得平衡。这一选择由奇异值谱所引导，体现了模型复杂度与数据保真度之间的根本张力，这种张力贯穿于现代数据科学之中。

12.2 *Approximation in Practice*

截断奇异值的优雅最优性，尽管在数学上是完备的，但仍然留下了实现层面的关键问题。在处理真实数据时——无论来自传感器阵列、图像序列，还是网络流量——我们必须在理论最优性与存储、计算和可靠性等实际约束之间取得平衡。这些工程方面的考量通过对数据自然结构方式的细致思考，将抽象的逼近理论转化为可运行的系统。

首先考虑存储方面的影响。通过对矩阵 $A \in \mathbb{R}^{m\times n}$ 进行截断 SVD 得到的秩为 k 的近似需要存储：

1. k 奇异值 $\sigma_1, \ldots, \sigma_k$ 2. k 左奇异向量 $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k \in \mathbb{R}^m$ 3. k 右奇异向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$

总计 $k(m+n+1)$ 个数字。当 $k(m+n+1) < mn$ 时，这会产生压缩。然而，这种粗略的参数计数忽略了一个关键问题：什么秩足以捕捉本质行为？

示例 12.5（建筑传感器网络）。考虑一栋配备了 100 个温度传感器的建筑，这些传感器以每小时一次的频率记录一整年的测量数据。由此得到的数据矩阵 $A \in \mathbb{R}^{8760\times 100}$ 名义上需要 876,000 个数值才能完全描述。每个条目 a_{ij} 给出了在第 i 个小时内传感器 j 的温度。物理现实表明存在很强的冗余性：热量在空间中平滑扩散，同时遵循清晰的日变化和季节性模式。

奇异值谱以定量方式揭示了这种结构：

$$\sigma_1 = 1247.3, \quad \sigma_2 = 423.1, \quad \sigma_3 = 89.4, \quad \sigma_4 = 12.7, \quad \sigma_5 = 3.2, \ldots$$

这种快速衰减——每个奇异值大约是其前驱的四分之一——确认了强烈的低秩结构。通过奇异向量可以揭示其物理意义：$\boldsymbol{u}_1$ 捕捉季节变化，$\boldsymbol{u}_2$ 反映日周期，而 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 显示主要的热区。一个秩为4的近似值使得均方根误差低于0.1°C，同时减少了95%的存储量。$\diamond$

不同的误差度量建议不同的截断策略。第12.1节讨论的Frobenius范数衡量典型的重构误差：

$$\|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}$$

在谱范数 $\|A - A_k\|_2 = \sigma_{k+1}$ 限制最坏情况下的误差时——这在近似值输入到控制系统时尤为关键，因为误差

可能会加重。对于我们的温度监测示例，谱范数界限确保在保留六个奇异值时，重构值与测量数据的差异不超过 0.5°C。

奇异值谱本身为近似质量提供了自然的指标：

1. 陡峭的下降表明存在清晰的截断点，例如我们的温度数据中物理模态能够被清楚地区分 2. 若呈现逐渐衰减且没有明显间隙，则警示可能不存在自然的低秩近似 3. 相近奇异值的成簇分布暗示特征之间存在耦合，需要联合保留

例12.6（图像序列分析）。以每秒30帧、1024×1024分辨率存储的视频流，名义上每秒需要超过3000万个数值。然而，大多数运动在空间和时间上都显得平滑。通过将每一秒视为一个$1024 \times (1024 \cdot 30)$矩阵，低秩近似通常可以在保持视觉质量的同时实现20:1的压缩率。奇异向量通过其不同的时间结构，自然地将持久的背景特征与运动元素分离开来：

- 静态背景元素出现在具有较大奇异值的前几个奇异向量中
- 移动物体在具有中等奇异值的后续向量中显现出来
- 传感器噪声集中在具有小奇异值的向量中

这种分解让人联想到第11章中研究的主成分，揭示了矩阵近似如何自动发现数据中有意义的结构。

◇

当数据持续到达或分布在多个传感器上时，直接计算SVD可能变得不可行。这种情况需要我们在第12.3节中开发的流式算法，基于第9章研究的主导原则。通过迭代，主导方向自然出现的数学洞察将理论最优性转化为实际计算。

物理约束往往会限制可接受的近似方式，而不只是简单的秩约简：

- 重构值的非负性
- 关键传感器读数的精确保存
- 相邻测量之间的有界导数
- 守恒定律或其他物理不变量

定义 12.7（约束低秩近似）。给定矩阵 $A \in \mathbb{R}^{m \times n}$ 和约束集合 $\mathcal{C}$，*constrained rank-k approximation prob- lem* 寻找：

$$\min_{B \in \mathcal{C}} \|A - B\|_F \quad \text{subject to} \quad \operatorname{rank}(B) \leq k$$

常见的约束包括：

1. 非负性: $\mathcal{C} = \{B : b_{ij} \geq 0\}$ 2. 元素级界限: $\mathcal{C} = \{B : l_{ij} \leq b_{ij} \leq u_{ij}\}$ 3. 精确匹配: $\mathcal{C} = \{B : b_{ij} = a_{ij} \text{ 对于 } (i, j) \in \mathcal{I}\}$

其中 $\mathcal{I}$ 表示必须被精确保留的条目的索引对集合。

•

尽管此类受约束的近似很少像第12.1节中的 Eckart–Young–Mirsky 定理那样具有闭式解，截断 SVD 往往能为寻求物理上有效近似的数值优化提供出色的初始化。我们将在第12.3节探讨用于求解此类受约束问题的算法。

低秩近似的实际威力恰恰在于理论结构与自然数据属性相契合之时显现。正如第11章中的主成分在高维数据中揭示了模式一样，截断的奇异值揭示了矩阵测量中的冗余。这种数学优雅与工程现实的融合，将抽象的优化理论转化为用于数据压缩与分析的实用工具。

12.3 *Algorithms at Scale*

第9章中研究的迭代系统揭示了矩阵幂如何自然地放大主导方向并抑制其他方向。正是这一原理——通过迭代而涌现的主导性——为我们提供了高效计算奇异向量的第一条途径。尽管对大型矩阵进行直接的 SVD 计算不可行，但通过精心的迭代，可以高效地提取低秩近似所需的主导奇异向量。

Nota bene: 这里的归一化不仅仅是为了方便——它在保留我们所需的方向信息的同时，防止数值溢出。

定义 12.8（幂迭代）。对于矩阵 $A \in \mathbb{R}^{m \times n}$，*power method* 的迭代为：

$$\boldsymbol{x}_{k+1} = \frac{A\boldsymbol{x}_k}{\|A\boldsymbol{x}_k\|}$$

从随机初始向量$\boldsymbol{x}_0$开始。当A的主奇异值σ_1严格大于σ_2时，这些迭代将收敛到主右奇异向量$\boldsymbol{v}_1$。

•

这种收敛性直接源自奇异向量展开：将 $\boldsymbol{x}_0 = \sum_i c_i \boldsymbol{v}_i$ 写成右奇异向量基下的表示：

$$A^k \boldsymbol{x}_0 = \sum_{i=1}^{r} \sigma_i^k c_i \boldsymbol{u}_i = \sigma_1^k \left(c_1 \boldsymbol{u}_1 + \sum_{i=2}^{r} \left(\frac{\sigma_i}{\sigma_1} \right)^k c_i \boldsymbol{u}_i \right)$$

由于 $\sigma_i/\sigma_1 < 1$ 对于 $i > 1$，重复乘法会放大主导方向并抑制其他方向。收敛速度取决于间隙比 σ_2/σ_1——第一和第二奇异值之间存在较大间隙可确保快速收敛。

Compare: 与第9章的支配性概念类似，幂迭代会放大最强的模式，同时自然地抑制较弱的模式。自然界本身也是如此运作：规模更大的人口增长得更快，更强的信号会淹没更弱的信号。

对于低秩近似，我们不仅需要一个，而是需要多个奇异向量。*block power method* 通过同时对多个向量进行迭代来解决这一问题：

$$X_{k+1} = \operatorname{orth}(A X_k)$$

其中，$\operatorname{orth}(\cdot)$ 通过 QR 分解为列空间生成一组正交归一基。这一过程提取了主导右奇异子空间的近似基，但谨慎的重新正交化对数值稳定性至关重要。

定理12.9（块幂收敛）。 *Let $A \in \mathbb{R}^{m \times n}$ have singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_r > 0$. For block size p, the block power method converges to the span of the first p right singular vectors at rate:*

$$\| \sin \Theta(X_k, V_p) \|_2 \leq \left(\frac{\sigma_{p+1}}{\sigma_p} \right)^k$$

where $\Theta(X_k, V_p)$ denotes the principal angles between the subspace spanned by X_k and that spanned by the first p right singular vectors.

Think: 主角夹角的正弦衡量子空间之间的不对齐——就像在依靠星辰导航时测量我们偏离真北有多远。

Krylov 子空间方法通过在更丰富的空间中工作，实现了显著更快的收敛。它们不只是保留当前迭代，而是维护整个子空间：

$$\mathcal{K}_k(A, \boldsymbol{x}) = \operatorname{span}\{\boldsymbol{x}, A\boldsymbol{x}, A^2\boldsymbol{x}, \ldots, A^{k-1}\boldsymbol{x}\}$$

该 *Lanczos process* 通过一种极其高效的三项递推构造了该空间的一组正交归一基：

$$\beta_{j+1} \boldsymbol{q}_{j+1} = A\boldsymbol{q}_j - \alpha_j \boldsymbol{q}_j - \beta_j \boldsymbol{q}_{j-1} \tag{12.1}$$

Historical Note: 兰佐斯于1950年发现了这一优雅的递推关系，但直到现代计算机使大规模计算成为可能，其力量才得以显现。宛如一段预言性的诗句，它的意义只有随着时间才逐渐浮现。

其中系数 α_j、β_j 在正交化过程中自然产生。该递推在保持简洁性的同时——每一步仅需一次矩阵乘法——并通过对正交性的精心管理来确保数值稳定性。

Lanczos 向量 $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k\}$ 为奇异向量提供了极好的近似，且误差随着 k 的增长而迅速减小。更为显著的是，系数的三对角矩阵：

$$T_k = \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \beta_3 & \\ & \beta_3 & \alpha_3 & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

捕捉了 A 的本质谱——其特征值迅速收敛到 A 的奇异值。将大型稀疏问题化简为小型三对角形式，体现了精心的算法设计如何高效地提取结构。

现代应用常常涉及规模如此之大的矩阵，以至于即便一次矩阵-向量乘法也代价高昂。随机化通过一个出奇简单的想法提供了一条强有力的前进路径：将我们的大矩阵投影到一个小的随机子空间上。考虑草图矩阵：

$$Y = A\Omega \quad \text{where} \quad \Omega \in \mathbb{R}^{n\times(k+p)}$$

具有相互独立的标准正态分布元素。尽管表面上看似鲁莽，这种随机投影却能以惊人的精度保留主导奇异子空间。

定理 12.10（随机投影）。 *Let $A \in \mathbb{R}^{m\times n}$ have singular value decomposition $A = U\Sigma V^T$, and let $\Omega \in \mathbb{R}^{n\times(k+p)}$ have independent standard normal entries. Then with probability at least $1 - 5e^{-p}$:*

$$\|A - Q_k Q_k^T A\|_2 \leq (1 + 11\sqrt{k/p})\sigma_{k+1}$$

where Q_k has orthonormal columns spanning the range of $Y = A\Omega$.

BONUS! 因子 $(1 + 11\sqrt{k/p})$ 揭示了一个深刻的事实：仅使用比我们的目标秩（$p \approx k$）略大的随机投影，我们就能达到近乎最优的精度。

对于观测值按顺序到达的流式数据：

$$A_t = A_{t-1} + \boldsymbol{u}_t \boldsymbol{v}_t^T$$

Think: 流式计算的挑战呼应了第9章的迭代系统——在处理无限的更新序列时，我们如何保持关键的结构?

我们希望在不存储完整矩阵的情况下保持低秩近似。*incremental SVD* 提供了一种优雅的解决方案：给定秩为 k 的近似 $A_{t-1} \approx U_{t-1}\Sigma_{t-1}V_{t-1}^T$，我们可以通过对大小不超过 $(k+1)\times(k+1)$ 的小矩阵进行 SVD，有效地更新它，同时保持固定秩。

定理 12.11（增量误差增长）。 *The error in incremental rank-k approximation grows as:*

1. $O(\sqrt{t})$ *in Frobenius norm*
2. $O(\log t)$ *in spectral norm*

where t is the number of rank-one updates processed.

这种受控的退化使得在保持近似最优性的同时能够处理几乎无限的数据流。从幂方法，经由随机化算法到流式更新的演进，反映了计算数学中的一种更广泛模式：当问题规模超越传统极限时，我们通过谨慎地放宽要求来保持准确性——以接受概率性保证或近似最优性来换取巨大的效率提升。

实践表明，算法选择应遵循以下指南：

1. 使用幂迭代快速估计主导方向 2. 当需要多个奇异向量时采用块方法 3. 在中等规模下选择 Lanczos 以获得最高精度 4. 对于极大规模问题采用随机化方法 5. 对流式数据使用增量更新

每种方法在准确性、效率和实现复杂性之间进行不同的权衡。

12.4 *Incompleteness & Reconstruction*

真实世界的数据常常存在缺失项。推荐系统只掌握用户偏好的极小一部分；传感器网络会遭遇偶发性故障；实验室测量只能捕捉蛋白质之间的某些相互作用。这些情形共享一种共同的数学结构：我们只观测到矩阵中的部分条目，而其他条目仍然未知；然而，由于自然约束或内在模式，底层数据往往具有低秩结构。

定义 12.12（矩阵补全问题）。令 $M \in \mathbb{R}^{m\times n}$ 为一个未知矩阵，并令 $\Omega \subset \{1,\dots,m\} \times \{1,\dots,n\}$ 表示一组被观测到的索引。给定观测 $\{m_{ij} : (i,j) \in \Omega\}$，该 *matrix com- pletion problem* 在其具有低秩这一假设下，旨在恢复 M。形式化地，我们旨在求解：

$$\min_{X} \operatorname{rank}(X) \quad \text{subject to} \quad x_{ij} = m_{ij} \text{ for all } (i,j) \in \Omega$$

Example: 在协同过滤中，我们只观测到可能的用户-物品评分中的极小一部分，但希望通过利用潜在的低秩结构来预测未观测到的偏好。

由于矩阵秩的离散性，这一优化问题被证明具有挑战性。通过第 12.1 节中引入的核范数，可以得到一种强有力的松弛方法：

定义 12.13（核范数）。矩阵 A 的 *nuclear norm*，记作 $\|A\|_*$，等于其奇异值之和：

$$\|A\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(A)$$

Foreshadowing: 核范数在矩阵补全中的作用预示了现代深度学习如何利用精心选择的正则化来引导学习表示中所需的结构。

该规范提供了矩阵秩的凸松弛，因为秩(A)等于非零奇异值的数量。

•

如果 M 的秩为 r，则它可以分解为 $M = UV^T$，其中 $U \in \mathbb{R}^{m\times r}$ 和 $V \in \mathbb{R}^{n\times r}$。部分观测提供了约束：

$$\sum_{k=1}^{r} u_{ik}v_{jk} = m_{ij} \quad \text{for all } (i,j) \in \Omega$$

其中 u_{ik} 和 v_{jk} 分别是 U 和 V 的条目。当这些方程足够约束因素时，恢复变得可能。

定理 12.14（矩阵补全）。 *Let $M \in \mathbb{R}^{m\times n}$ be a rank-r matrix with singular value decomposition $M = U\Sigma V^T$, where $\sigma_r(M) > 0$. Let Ω contain entries sampled uniformly at random. Then with probability at least $1 - cn^{-3}$ (for some constant c), M is the unique solution to:*

$$\min_X \|X\|_* \quad \textit{subject to} \quad x_{ij} = m_{ij} \textit{ for all } (i,j) \in \Omega$$

provided:

1. $|\Omega| \geq C\mu r(m+n)\log^2(m+n)$ *entries are observed*
2. *The singular vectors $\{\boldsymbol{u}_i\}$ and $\{\boldsymbol{v}_i\}$ satisfy the* 不相干性条件*:*

$$\max_{i,j} \left\{ \frac{m}{r}\|\Pi_U \boldsymbol{e}_i\|^2, \frac{n}{r}\|\Pi_V \boldsymbol{e}_j\|^2 \right\} \leq \mu$$

   where Π_U and Π_V denote projection onto the column spaces of U and V respectively

Here C and μ are numerical constants independent of matrix dimensions.

Think: 不相干性确保信息在矩阵中均匀传播，而不是集中在少数几个条目中——就像统计学中要求良好的实验设计一样。

非相干性条件至关重要——它确保信息在矩阵中均匀传播，而不是集中在少数几个元素上。当奇异向量与坐标轴高度对齐时，单个条目可能会对恢复变得至关重要。随机采样确保我们能够捕获关于所有方向的足够信息。

示例 12.15（电影推荐）。考虑一个矩阵，其中行表示用户，列表示电影，条目给出评分（1-5 星）。大多数用户仅对一小部分电影进行评分，但仍然存在模式

显现出来：品味相似的用户对电影给出相似的评分；相似类型的电影获得相关的评分。这些模式在数学上表现为低秩结构。

对于一个包含1000名用户和1000部电影的数据集，传统方法可能需要全部一百万个潜在评分。然而，如果底层偏好仅依赖于40个潜在因子（如类型偏好、制作水准等），矩阵补全理论表明，我们可以从大约80,000个已观测评分中恢复出准确的预测——这种显著的减少使得实用的推荐系统成为可能。◇

尽管理论上优雅，核范数最小化在大规模问题中需要谨慎的算法处理。*proximal gradient method* 通过迭代提供了一种高效的方法：

$$\begin{aligned} Y_k &= X_k - \eta_k \nabla f(X_k) \\ X_{k+1} &= \mathcal{S}_{\lambda\eta_k}(Y_k) \end{aligned}$$

其中 $f(X)$ 衡量对观测条目的保真度，η_k 控制步长，$\mathcal{S}_\tau$ 表示奇异值软阈值化：

$$\mathcal{S}_\tau(Y) = U\,\mathrm{DIAG}(\max\{\sigma_i - \tau, 0\})V^T$$

用于 SVD $Y = U\Sigma V^T$。该算法有效地在低秩矩阵空间中执行梯度下降。

更复杂的模型能够处理现实世界中的复杂性，如噪声和系统性偏差。*robust matrix completion* 问题对我们的优化进行了修改，以允许误差：

$$\min_{X,S} \|X\|_* + \lambda\|S\|_1 \quad \text{subject to} \quad x_{ij} + s_{ij} = m_{ij} \text{ for } (i,j) \in \Omega$$

这里，S 捕获稀疏误差，而 X 保持低秩结构。参数 λ 在这些相互竞争的目标之间进行平衡——取值越大，越倾向于低秩而非误差容忍。

例12.16（传感器网络）。分布在建筑物各处的温度传感器网络应呈现出显著的低秩结构——相邻的传感器记录相似的数值，而日周期和季节性模式会系统性地影响所有传感器。当部分传感器失效时，矩阵补全可以根据仍然存活的测量对其读数进行插值。定理12.14的不相干条件可转化为物理上的洞见：当传感器在空间中均匀分布、避免过于密集的簇或过于稀疏的区域时，恢复效果最佳。◇

矩阵补全的理论保证揭示了一个深刻的真理：结构往往会使最初看似必不可少的东西变得多余。正如第 10 章的奇异值分解揭示了满矩阵中的冗余性一样，补全理论表明，当底层模式存在时，部分观测就已足够。这一原则——结构使得从不完整数据中进行重建成为可能——贯穿于现代数据科学的诸多领域，从压缩感知到神经网络训练。

12.5 Robust Matrix Factorization

数据很少以足够完美的状态到达，以至于可以直接进行近似。尽管我们所发展的低秩方法在模式隐藏于高维空间时展现出强大的能力，真实测量往往遭受更为根本性的破坏。传感器可能会完全失效，而不只是引入噪声；实验流程会引入系统性偏差；恶意行为者可能会故意污染观测。然而，在许多情况下，潜在的低秩结构仍然在这些扭曲之下得以保持，正如晶体的基本对称性即便其表面受损也依然存在。

首先考虑矩阵观测中污染是如何体现的：

1. 单个测量中的粗大误差 2. 影响整行或整列的系统性偏差 3. 针对特定模式的对抗性篡改 4. 来自测量不确定性的随机噪声

这些不同形式的污染需要不同的数学处理，但它们共享一个共同主题：它们表示相对于潜在低秩结构的稀疏偏离。

定义 12.17（鲁棒主成分分析）。*robust principal component analysis* 问题旨在将观测矩阵 M 分解为：

$$M = L + S + N$$

其中：

- L 具有低秩（捕捉真实模式）
- S 是稀疏的（表示粗大误差）
- N 包含少量随机噪声

无噪声版本（$N = 0$）在保留基本特性的同时，便于进行优雅的理论分析。

Compare: 与第11章的主成分类似，鲁棒PCA寻求数据中的基本模式——但现在会显式地对异常污染和噪声进行建模并将其分离。

这一洞见——即腐败往往只影响一小部分条目，而大多数观测仍然可靠——暗示结合

将第12.4节中的核范数与ℓ_1范数结合，以同时促进低秩和稀疏性：

定理 12.18（主成分追踪）。 *Let $M = L_0 + S_0$ where L_0 has rank r and S_0 has at most k nonzero entries with random signs. Under the conditions:*

1. $\text{rank}(L_0) \leq \rho_r$ 分钟$\{m, n\}$
2. $\|S_0\|_0 \leq \rho_s mn$
3. *The singular vectors of L_0 satisfy the incoherence condition of Theorem 12.14*

the solution to:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad \textit{subject to} \quad L + S = M$$

exactly recovers L_0 and S_0 with high probability when $\lambda = 1/\sqrt{\max\{m, n\}}$ and ρ_r, ρ_s are sufficiently small constants.

BONUS! 定理12.18中的优化问题虽然是凸的，但在大规模问题中需要谨慎的算法处理。*alternating* A*direction method of multipliers* (ADMM)通过迭代优化提供了一种高效的方法。

例12.19（视频监控）。考虑一台固定相机记录的场景，其中大部分变化来自静态背景前的少数运动物体。得到的数据矩阵 M 的行索引为帧，列索引为像素。背景贡献了一个低秩分量 L (，因为它变化缓慢，甚至几乎不变)，而运动物体产生稀疏变化 S。尽管偶尔存在传感器故障或光照变化，鲁棒 PCA 仍能成功分离这些分量：

$$\begin{bmatrix} \text{frame 1} \\ \text{frame 2} \\ \vdots \\ \text{frame } n \end{bmatrix} = \underbrace{\begin{bmatrix} \text{background} \\ \text{background} \\ \vdots \\ \text{background} \end{bmatrix}}_{\text{rank} \approx 1-4} + \underbrace{\begin{bmatrix} \text{moving objects} \\ \text{moving objects} \\ \vdots \\ \text{moving objects} \end{bmatrix}}_{\text{sparse}}$$

◇

更现实的场景往往需要在我们的分解中引入额外的结构。当扰动呈现出模式——例如传感器校准中的系统性偏差、对评分的协同操纵——我们可以通过刻画这种结构的专门范数来建模 S：

$$\min_{L,S} \|L\|_* + \lambda \Omega(S) \quad \text{subject to} \quad L + S = M$$

w在这里，$\Omega(\cdot)$ 编码了我们对腐败 p 的假设　　模式。

示例 12.20（协同过滤）。在推荐系统中，一些用户会故意操纵评分以提升或贬低某些

项目。与其将这些视为独立的干扰，我们可以通过组稀疏性来建模它们：

$$\Omega(S) = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} s_{ij}^2}$$

这行-wise $\ell_{2,1}$ 范数鼓励将整个用户识别为操控性用户，而不是独立地对待每个评分。 ◇

当噪声水平在不同观测之间变化时，加权变体显得很有价值：

$$\min_{L,S} \|L\|_* + \lambda \|W \odot S\|_1 \quad \text{subject to} \quad L + S = M$$

其中 W 包含逐项权重，$\odot$ 表示 Hadamard（逐元素）积。这使我们能够在保持对潜在受损测量的怀疑的同时，更加信任可靠的测量。

定理 12.21（稳定恢复）。*Under the conditions of Theorem 12.18, if* $M = L_0 + S_0 + N$ *where* $\|N\|_F \leq \epsilon$*, then the solution* (L, S) *to the robust PCA problem satisfies:*

$$\|L - L_0\|_F^2 + \|S - S_0\|_F^2 \leq C\epsilon^2$$

for some constant C *depending only on matrix dimensions.*

稳健矩阵分解的理论保证揭示了一个基本原理：即使在观察数据严重受损的情况下，只要损坏本身表现出某种形式的简单性（如稀疏性），结构仍然可以被恢复。这个原理——即当通过适当的数学视角观察时，模式变得更加清晰——将引导我们在第十三章中开发神经网络。

The Netflix Prize

像考古学家从残缺的卷轴中重建古代文本一样，现代推荐系统面临着从稀疏的观察中推断完整意义的挑战。2006-2009年的Netflix奖将这种隐喻式的重建转化为精确的数学，展示了低秩矩阵逼近理论如何使机器在前所未有的规模上预测人类偏好。这个百万美元的挑战将本章中发展出的抽象数学与实用的机器学习相结合，揭示了矩阵分解与协同过滤之间的深刻联系。

数学图景似乎很清晰：一个庞大但稀疏的矩阵 $R \in \mathbb{R}^{m \times n}$，包含大约一亿条评分，来自 $m = 480,000$ 名用户，遍布

$n = 17{,}700$ 部电影。每个条目 r_{ij} 表示一名用户对一部电影的 1–5 星评分，其中超过 98% 的条目缺失——大多数用户只对可用电影中的极少一部分进行评分。形式化的目标很简单：最小化预测的均方根误差（RMSE）：

$$\text{RMSE} = \sqrt{\frac{1}{|\Omega|}\sum_{(i,j)\in\Omega}(r_{ij} - \hat{r}_{ij})^2}$$

Historical Note: 10% 的改进目标是基于 Netflix 的内部分析而选定的，该分析表明这一阈值将为用户体验带来有意义的提升。几乎没有人预料到实现这一目标需要三年。

其中 Ω 表示观测到的评分集合，$\hat{r}_{ij}$ 表示预测评分。

获胜的解决方案揭示了理论与实践之间的深刻统一。其核心在于奇异值分解及其衍生方法——正是第12.4节为矩阵补全而开发的那些工具。然而，现实世界的推荐系统必须应对我们纯净的理论所忽略的复杂性：

1. 时间效应：观看模式随时间变化 2. 用户偏差：有些人给分慷慨，另一些人较为严苛 3. 物品偏差：有些电影的评分始终高于其他电影 4. 隐式反馈：未对电影评分也传达了信息

纯粹的低秩近似虽在理论上优雅，但事实证明并不足够。获胜的算法通过对评分矩阵进行细致的分解来应对这些复杂性。 对于用户 u 在时间 t 对物品 i 的评分，预测评分采用如下形式：

$$\hat{r}_{ui}(t) = \mu + b_u(t) + b_i(t) + \boldsymbol{q}_i^T\left(\boldsymbol{p}_u + |\mathcal{N}(u)|^{-\frac{1}{2}}\sum_{j\in\mathcal{N}(u)}\boldsymbol{y}_j\right)$$

这里：

- μ 表示全局平均评分
- $b_u(t)$ 并且 $b_i(t)$ 捕捉随时间变化的用户和物品偏置
- 向量 $\boldsymbol{q}_i$、$\boldsymbol{p}_u \in \mathbb{R}^f$ 编码 f 潜在特征
- 对 $\mathcal{N}(u)$ 的求和包含了隐式反馈

这种分解体现了一种深刻的洞见：人类偏好虽然复杂，但往往可以归结为更简单模式之间的相互作用。正如第 ?? 节所展示的，任意矩阵可以分解为秩一分量，Netflix Prize 揭示了客户偏好如何从可解释的特征中涌现。这些参数是通过求解正则化优化问题来学习得到的：

$$\min_{b,\boldsymbol{p},\boldsymbol{q},\boldsymbol{y}}\sum_{(u,i)\in\Omega}(r_{ui} - \hat{r}_{ui})^2 + \lambda\left(\|b_u\|^2 + \|b_i\|^2 + \|\boldsymbol{p}_u\|^2 + \|\boldsymbol{q}_i\|^2 + \sum_{j\in\mathcal{N}(u)}\|\boldsymbol{y}_j\|^2\right)$$

与第12.4节中提出的核范数正则化相匹配。

实现同样揭示了关于稳健矩阵分解的重要经验。原始评分出人意料地嘈杂，受以下因素污染：

- 用户对星级评分的解读各不相同
- 评分模式中的时间漂移
- 用户选择对哪些电影进行评分时的选择偏差
- 来自竞争服务的对抗性评分

这些挑战要求使用在第12.5节中开发的强大技术。通过仔细建模和消除系统性偏差，同时对极端评分保持怀疑态度，获胜团队实现了10.06%的RMSE改进——几乎刚刚突破了比赛开始时看似不可能的门槛。他们的成功展示了理论洞察如何促进实践工程：没有矩阵分解的数学基础，任何算法的巧思都无法成功。

Netflix Prize 的遗产远不止于电影推荐。其洞见如今为整个数字经济中的个性化提供动力，从电子商务到音乐流媒体，再到社交媒体内容排序。它们都应用同一核心原则：人类偏好尽管看似混乱，但在适当的数学视角下，往往可以用低秩近似来刻画。

Nota bene: 尽管由于工程复杂性，Netflix 从未实施获胜的算法，但这场竞赛的洞见从根本上改变了行业构建推荐系统的方式。

Network Anomaly & Detection

在互联网脉动的动脉中，流量像血液通过血管一样在端点之间流动，承载着现代通信的命脉。每个数据包从源到目的地的旅程都会在路由器日志和网络监控器中留下痕迹，形成庞大的流量数据矩阵。然而，在这个数字循环系统中，隐藏着正常模式——电子邮件、网页浏览和视频流的规律性节奏——以及可能暗示安全漏洞或即将发生故障的危险异常。本章中发展出的低秩近似理论提供了强大的工具，用于区分这些模式和背景噪声。

考虑一个包含 n 个节点、随时间交换流量的网络。每次测量都会产生一个流量矩阵 $X(t) \in \mathbb{R}^{n \times n}$，其中条目 $x_{ij}(t)$ 表示在时间区间 t 内从节点 i 流向节点 j 的数据量。将这些快照在 m 个时间段内收集起来，得到一个三阶张量 $\mathcal{X} \in \mathbb{R}^{n \times n \times m}$。尽管看起来很复杂，这个流量张量通常可以通过低秩近似获得极其简洁的描述。

Historical Note: 早期的网络异常检测尝试集中在简单的统计离群值测试上。2000年代初期，矩阵分解技术的应用通过揭示复杂模式，彻底改变了这一领域，这些模式是简化方法无法发现的。

关键的见解来自于检查张量的结构。正常的网络流量展示出强烈的模式：

1. 空间相关性：邻近节点往往共享相似的流量特征2. 时间周期性：日周期和周周期在正常使用中占主导3. 内在维度较低：大多数流量沿着可预测的路径流动

这些模式建议将每个时间片 $X(t)$ 分解为三个组成部分：

$$X(t) = L(t) + S(t) + N(t)$$

其中 $L(t)$ 捕捉低秩正常流量模式；$S(t)$ 表示稀疏异常流量；$N(t)$ 包含测量噪声。

这种分解与第12.5节中提出的稳健PCA框架完全一致。通过求解：

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad \text{subject to} \quad L + S = X$$

对于每个时间切片，我们将常规流量模式与潜在异常分开。

核范数 $\|L\|_*$ 鼓励正常流量中的低秩结构，而 ℓ_1 范数 $\|S\|_1$ 促进稀疏异常。

例12.22（DDoS 攻击检测）。在分布式拒绝服务（DDoS）攻击期间，许多被攻陷的机器向单一目标洪泛流量。这在 $S(t)$ 中形成了一种典型模式：在同一列中出现许多小的条目（多个来源）指向同一个目的地。即使单个流量看起来无害，稀疏性约束也能自然地凸显这种协同异常。

考虑一个拥有1000个节点、正遭受攻击的网络。正常流量分量 $L(t)$ 通常具有 10–20 的秩，反映了合法的通信模式。攻击期间，$S(t)$ 通过单一列中的数值升高揭示了攻击，尽管每个单独的源仅贡献了总流量的一小部分。⋄

更为复杂的分析通过张量分解来利用时间结构。通过将时间切片堆叠起来，将交通张量以矩阵形式表示：

$$\mathbf{X} = \begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(m) \end{bmatrix}$$

Nota bene: 时间分辨率的选择对会涌现出哪些模式具有关键性影响。粒度过细会放大噪声；粒度过粗则会错失重要结构。大多数实现同时使用多个时间尺度。

通过其奇异值谱揭示出额外的模式。正常流量通常在前20-30个奇异值之后出现陡降，而异常则会对该谱产生明显的扰动。

鲁棒分解框架应对了若干实际挑战：

- 缺失来自故障监视器的数据
- 数据包计数引起的量化效应
- 时变网络拓扑
- 混合正常与异常流量

这些问题直接关联到第12.4节的矩阵补全理论。正如 Netflix 必须预测未观测到的评分一样，网络分析必须在保持对受损观测具有鲁棒性的同时，推断缺失的流量测量。

现代实现通过流式算法扩展了这些思想，这些算法实时处理流量数据。第12.3节中开发的增量SVD技术使得可以在有限的内存要求下持续监控网络健康。当出现异常时，它们在主奇异向量上的投影通常揭示了它们的性质：

- 端口扫描呈现为稀疏的行
- DDoS攻击创建密集的列
- 蠕虫传播呈现出典型的对角线模式
- 数据外泄表现为持续的点对点流量

然而，这种力量带来了责任。检测恶意流量的相同技术可能会被用来去匿名化用户或追踪通信模式。随着网络在现代生活中变得愈加重要，安全与隐私之间的平衡变得愈加微妙。这些担忧反映了

我们遇到了推荐系统——强大但本身并不具备伦理性的数学工具。

将低秩近似应用于网络流量分析体现了一个更为普遍的原则：复杂的现实世界系统往往蕴含着有待发现的更为简单的内在结构。正如第 10 章中的奇异值分解在抽象向量空间中揭示了优选方向，这里发展出的技术揭示了数字通信中的基本模式。随着我们在第 13 章迈向神经网络，这一洞见——高维数据往往分布在接近低维流形的区域——将变得愈发重要。

□ ——————————————————————————— □

Exercises: Chapter 12

1. 设 $\begin{bmatrix} 4 & 2 & 1 & 2 & 3 & 0 & 1 & 0 & 2 \end{bmatrix}$。通过 SVD 截断求最佳的秩为 2 的近似 A_2。将 Frobenius 范数误差 $\|A - A_2\|_F$ 与埃卡特–杨–米尔斯基定理所保证的理论最小值进行比较。

2. 幂法以 $|\sigma_2/\sigma_1|$ 的速率线性收敛。对于矩阵

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

从 $x_0 = (1, 0)^T$ 开始计算三次迭代。估计收敛速率并与理论进行比较。什么样的初始向量会给出最快的收敛?

3. 图像压缩通常在进行低秩近似之前先划分为若干块。对于矩阵

$$A = \begin{bmatrix} 100 & 98 & 45 & 43 \\ 97 & 95 & 47 & 44 \\ 42 & 45 & 153 & 155 \\ 44 & 46 & 152 & 154 \end{bmatrix}$$

比较以下两种情况下的误差：（a）对完整矩阵进行秩为 2 的近似，与（b）对每个 2×2 块进行秩为 1 的近似。解释哪种方法更好地捕捉了块结构。

4. 一个化学过程生成了被稀疏误差污染的秩为2结构的噪声测量：

$$M = \begin{bmatrix} 2.1 & 4.2 & -15.0 & 8.1 \\ 4.0 & 8.1 & 12.2 & 16.0 \\ 6.2 & 12.0 & 18.1 & 23.9 \\ 8.0 & 16.2 & 24.1 & 32.2 \end{bmatrix}$$

利用数据中的模式，识别可能的离群点，并提出一个干净的秩-2结构。论证你的分解的合理性。

5. 考虑稳健PCA问题 $M = L + S$ 其中

$$M = \begin{bmatrix} 1 & 2 & 10 \\ 2 & 4 & 3 \\ 3 & 6 & 4 \end{bmatrix}$$

解释为什么 (3, 3) 项很可能表示一个异常值。找到一种合理的分解，将其分解为低秩 L 和稀疏 S 分量。

6. 一个秩为 3 的矩阵 $A \in \mathbb{R}^{5\times 4}$ 其前两个奇异值 $\sigma_1 =$ 为 10 和 $\sigma_2 =$ 为 5。无需显式计算 A_1，求解：

(a) 对于任意秩为1的矩阵 B (b)，可达到的最小误差 $\|A - B\|_F$ (b) 最佳秩-2与秩-1近似之间的弗罗贝尼乌斯范数差 $\|A_2 - A_1\|_F$ (c) 未知的奇异值 σ_3

7. 考虑矩阵补全问题，其中我们观测到部分条目

$$M = \begin{bmatrix} 2 & ? & 1 \\ ? & 4 & ? \\ 1 & ? & 2 \end{bmatrix}$$

找出一个与这些观测一致的秩为 1 的矩阵，或者证明不存在。存在多少个秩为 2 的补全？将你的答案与定理 12.14 的采样条件联系起来。

8. 一个传感器网络每小时在5个位置测量温度。三天后，两个传感器同时失效。数据矩阵变为：

$$M = \begin{bmatrix} 72 & 70 & 68 & ? & ? \\ 75 & 74 & 71 & ? & ? \\ 69 & 67 & 65 & ? & ? \\ 73 & 71 & 69 & ? & ? \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

解释每日和空间模式如何创建低秩结构。如果第一个奇异值解释了85

9. 证明对于任意矩阵 A 和索引 k，相继的最佳秩近似之间的差满足

$$\|A_k - A_{k-1}\|_F^2 = \sigma_k^2$$

用此来解释为什么考察比值 σ_k/σ_{k+1} 有助于选择截断秩。

10. 设 A 的 SVD 为 $A = U\Sigma V^T$，并考虑当新列 $\boldsymbol{c}$ 到来时更新一个秩为 k 的近似。证明从头开始计算 $(A|\boldsymbol{c})_k$ 是不必要的——该更新只需要 $O(mk)$ 次运算，其中 m 是行维度。你将如何为流式数据维护一个近似的 SVD?

11. 随机化 SVD 算法计算 $Y = AQ$，其中 Q 的列是正交归一的。考虑如下误差界：

$$\|A - QQ^T A\|_2 \leq (1 + \epsilon)\sigma_{k+1}$$

ϵ 依赖于过采样。解释为什么这激励使用 Q 作为秩-k 近似的基础。你如何在实践中选择过采样参数?

12. 对于大小兼容的矩阵 A 和 B，证明

$$\|\Pi_{\text{im}\, A} B\|_F \leq \|B\|_F$$

其中 $\Pi_{\text{im}\, A}$ 表示正交投影到 A 的像上。何时等式成立? 这与最优低秩近似有什么关系?

13. 设 A 为一个具有块的矩阵:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

证明 $\text{rank}(A) \geq \max\{\text{rank}(A_{11}), \text{rank}(A_{22})\}$。然后构造一个示例，展示这个界限可以是严格的。这如何影响分块低秩近似的策略?

14. 设 A 为一个秩为 r 的矩阵，其最小的非零奇异值为 σ_r。证明如果 $\|E\|_2 < \sigma_r/2$，则 $\text{rank}(A + E) \geq r$。利用此证明为什么从噪声数据中估计秩需要检查奇异值间隙，而不仅仅是计算非零值的个数。

15. 解释为什么当缺失条目形成规则性的模式而不是随机出现时，矩阵补全通常会失败。构造一个简单的 3×3 示例来展示这一现象。这与定理 12.14 中的非相干性条件有何关系?

16. 对于两个秩为 k 的矩阵 A 和 B，证明它们的乘积 AB 的秩至多为 k。然后通过构造达到等号的矩阵来表明该界是紧的。这个结论如何为近似矩阵乘法的算法提供启示?

17. 设 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ 是 $\mathbb{R}^m$ 中的单位向量。Gram 矩阵 G 的元素为 $g_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$。证明 G 的秩至多为 m，并将其奇异值与子空间之间的主角度联系起来。这将如何有助于分析迭代方法的收敛性?

18. 加权低秩逼近问题旨在最小化:

$$\|W \odot (A - X)\|_F \quad \text{subject to} \quad \text{rank}(X) \leq k$$

其中 $\odot$ 表示按元素乘法，且 W 具有非负元素。解释为什么该问题的解不再由截断 SVD 给出。提出一种基于交替最小化的迭代算法来解决该问题。

19. 对于任意矩阵 A，证明核范数 $\|A\|_*$ 等于最优值

$$\min_{UV^T = A} \frac{1}{2}(\|U\|_F^2 + \|V\|_F^2)$$

使用此方法解释为什么核范数最小化在矩阵补全中促进低秩解。

20. 考虑一个 $n \times n$ 矩阵 A，其奇异值呈指数衰减 $\sigma_k = 2^{-k}$。对于误差容限 ϵ:

（a）求用于 Frobenius 范数近似所需的最小秩 （b）比较存储完整形式与低秩形式的内存需求

(c) 分析幂迭代与随机化方法的相对效率。你的结论如何依赖于矩阵维数 n？ 21.（挑战）设 A 为一个秩为 r、条件数为 κ 的矩阵。证明在进行 t 次迭代后，幂迭代中的误差满足

$$\|\sin\Theta(x_t, v_1)\| \le (1/\kappa)^t \|\tan\Theta(x_0, v_1)\|$$

其中 Θ 表示子空间之间的角度。这揭示了关于初始化敏感性的什么？

Chapter 13
Neural Networks & AI

“multitudes without number work incessant: the hewn stone is plac’d in beds of mortar mingled with the ashes of Vala”

线性代数近期被综合进机器智能，标志着在数学与社会层面上的一次深刻转变。通过十二个章节，我们建立了对线性变换的细致理解——它们的基本子空间与商空间，通过奇异值展现的几何，通过特征理论体现的动力学，以及通过低秩结构实现的近似。

推理超越线性，需要能够捕捉自然与人工系统中普遍存在的微妙非线性的工具。神经网络正提供了这样的工具，通过精心的组合与训练，将我们的线性基础转化为人工智能的构建基石。

首先考虑为什么仅有线性性不足以解决问题。一个图像识别系统必须以某种方式将像素强度映射到对象类别；一个语言模型必须将词序列转换为有意义的表示；一个机器人必须把传感器读数转化为恰当的动作。然而，无论多么精心选择，任何纯线性变换都无法捕捉这些本质上非线性的关系。尽管如此，我们所发展的数学仍然至关重要——不是作为完整的解决方案，而是作为神经网络构建其表达能力的基础语法。

这种构造通过简单元素的组合来实现。神经网络的每一层先执行线性变换，随后进行非线性 *activation function*。尽管单个层很简单，这些层组合起来却能逼近极其复杂的映射。第12章研究的低秩近似暗示了这种能力——正如截断的奇异值能够捕捉数据中的主导模式，训练过的

网络通过其分层结构学习提取并变换相关特征。理解这一学习过程需要将我们的线性代数洞见与优化理论、概率论以及信息几何相结合。

现代神经架构已经远远超越了简单的分层网络。卷积网络通过专门化的线性运算来利用空间结构；注意力机制根据上下文动态地重新加权其计算；残差连接创建了捷径，使优化更加容易。然而，在这些精巧设计之下，潜藏着我们在线性代数学习中熟悉的模式——条件数在训练稳定性中的作用、秩在压缩中的重要性，以及高维表示的几何结构。即便是处理数十亿参数的最大规模语言模型，其能力也建立在对本质上线性运算的精心组合之上。

我们的任务是弥合经典数学与现代人工智能之间的鸿沟。我们从神经网络的基本架构入手，展示非线性如何将简单的矩阵乘法转化为通用函数逼近。这自然引出基于梯度的学习算法，其中来自矩阵微积分的洞见指导高效的优化。随后，现代架构作为这些原则的精心改进而出现，每一项创新都扎根于数学理解之中。贯穿始终，我们坚持这样一种视角：神经网络并非对经典方法的断裂，而是其自然演进——通过有原则的组合与训练，将线性的构件转化为学得的智能。

探索神经网络如何实现其卓越能力——以及它们可能面临的根本性限制——推动了当代人工智能研究的很大一部分。尽管我们不可能解决所有谜题，但本文所建立的数学基础为分析与创新提供了不可或缺的工具。当我们探索人工智能的前沿，从大型语言模型到机器人控制时，我们将反复看到线性代数的洞见如何同时照亮实际工程以及关于学习与智能本质的更深层次理论问题。

13.1 Beyond Linear Transformations

现实抵制被简化为纯线性描述。尽管前几章已经构建了一个强大的框架，用于理解线性变换——通过奇异值理解其几何，通过特征理论理解其动力学，通过低秩结构理解其近似。

——自然要求更多。定义智能的模式，无论是自然的还是人工的，都描绘出远超任何线性子空间的曲线与流形。如同光穿过晶体发生弯折，信息必须经过非线性变换，才能显露其更深层的结构。

考虑线性的根本性局限。任何纯线性变换都无法分离线性不可分的点；仅靠矩阵乘法无法刻画 XOR 函数；任何线性运算序列都不能表示哪怕是简单的逻辑判定。然而，大脑却能毫不费力地完成这些壮举，通过神经活动的级联将原始感官数据转化为抽象理解。这揭示了一个深刻的原则：智能并非源自孤立的线性运算，而是源自它们与非线性的精心组合。

Historical Note: XOR 函数仅在其输入中恰好有一个为 1 时输出 1：

x_1	x_2	XOR
0	0	0
0	1	1
1	0	1
1	1	0

明斯基和帕珀特于1969年出版的著作*Perceptrons*证明，单层网络无法学习这一简单函数，从而在事实上使神经网络研究停滞了十多年。解决方案需要引入隐藏层——这一见解要等到20世纪80年代反向传播革命才得以实现。

该组合的数学基础通过非线性操作实现，这些操作逐个分量地转换它们的输入：

定义 13.1（激活函数）。一个 *activation function* $\sigma : \mathbb{R} \to \mathbb{R}$ 是一个按元素作用于向量的非线性函数。常见的选择包括：

1. ReLU（修正线性单元）： $\sigma(x) = \max\{0, x\}$ 2. Sigmoid： $\sigma(x) = 1/(1 + e^{-x})$ 3. 双曲正切函数： $\sigma(x) = \tanh(x)$

当作用于向量 $\boldsymbol{x}$ 时，我们写作 $\sigma(\boldsymbol{x}) = (\sigma(x_1), \ldots, \sigma(x_n))^T$。 •

这些看似简单的函数将线性运算转化为强大的计算构建模块。例如，ReLU 函数实现了一种基本形式的稀疏性——将负值设置为零，同时保留正值。尽管其导数在零处不连续，但这一不连续性对高效学习至关重要。Sigmoid 和双曲正切函数提供了平滑的过渡，连接渐近极限，但现代实践更倾向于使用 ReLU，因为它在计算上更简单，且具有有益的梯度特性。

Historical Note: 早期的神经网络使用 sigmoid 激活函数，类似于生物神经元的放电频率。2011 年左右转向 ReLU 标志着一个关键的进步，通过改进的梯度流使得更深的网络成为可能。

现在考虑线性运算和非线性运算如何组合。给定输入向量 $\boldsymbol{x} \in \mathbb{R}^n$、*weight matrix* $W \in \mathbb{R}^{m \times n}$ 和 *bias vector* $\boldsymbol{b} \in \mathbb{R}^m$，单个神经网络层计算：

$$\boldsymbol{h} = \sigma(W\boldsymbol{x} + \boldsymbol{b}) \tag{13.1}$$

这种由矩阵乘法、向量加法以及逐元素非线性组成的组合构成了神经计算的基本构件。尽管每个操作都很简单，但它们的组合却赋予了非凡的表达能力：

引理 13.2（普适逼近）。*A neural network with a single hidden layer of sufficient width can approximate any continuous function on a compact domain to arbitrary precision. More precisely, given any continuous function* $f : [0, 1]^n \to \mathbb{R}$ *and error* $\epsilon > 0$*, there exist weights* W_1, W_2 *and biases* $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ *such that:*

$$\|f(\boldsymbol{x}) - W_2\sigma(W_1\boldsymbol{x} + \boldsymbol{b}_1) + \boldsymbol{b}_2\|_\infty < \epsilon$$

for all $\boldsymbol{x} \in [0, 1]^n$.

这一理论保证尽管强有力，但几乎无法提供实践指导。现代网络之所以取得卓越性能，并非依赖宽度，而是依赖深度——通过精心组合多层线性与非线性操作的交替。这种分层结构既映射了生物神经网络（不同脑区以层级方式处理信息），也契合一种数学观念：复杂函数往往可以自然地分解为更简单的阶段。

例子 13.3（图像识别）。考虑识别手写数字的任务。每个图像作为像素强度的矩阵到达——这是高维空间中的一个点。人类手写的变化创建了远离任何线性子空间的复杂流形。典型的卷积神经网络通过以下阶段转换这些图像：

1. 线性卷积检测局部模式 2. ReLU 激活引入非线性 3. 池化操作提供不变性 4. 进一步的层级结合特征

Keep going... 有关CNN的更多信息，请参见本章末尾。

没有任何单一的线性变换能够分离这些数字类别，但它们与非线性的组合却能够实现显著的准确性。 ◇

神经网络的力量不在于它们各个操作的复杂性——线性变换和激活函数都保持简单——而在于这些操作如何结合以逼近复杂的函数。就像矿物的晶体结构从简单的原子排列中形成，智能也从基本数学操作的精心组合中涌现。

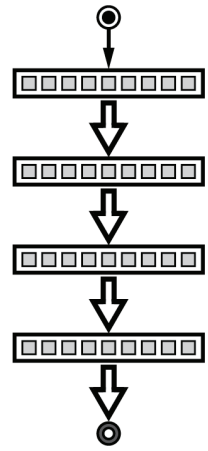
这一视角改变了我们对神经网络在计算什么的理解。我们不再将它们视为黑箱，而是将其看作由易于理解的数学运算所构成、且可以学习的组合。前几章研究的线性变换为信息流动提供了框架，而激活函数则引入了复杂计算所必需的非线性。这种线性与非线性的统一

非线性操作——连续变换和离散决策——为现代人工智能提供了基础。

接下来的章节将探讨这些基本构建块如何组合成复杂的架构，网络如何通过基于梯度的优化进行学习，以及现代创新如注意力机制和残差连接如何扩展这些基本原理。在整个过程中，我们保持这样的视角：神经网络不仅不是经典数学的突破，而是其自然演化——通过有原则的组合和训练将线性代数运算转化为学习到的智能。

13.2 *Network Architecture & Matrix Factorization*

第13.1节讨论的线性和非线性运算的分层组合暗示了神经计算的自然架构。像第10章的奇异值分解一样，神经网络将复杂的变换分解为更简单的组件——不是通过解析推导的奇异向量，而是通过学习的权重矩阵，通过非线性函数分隔开来。这一视角将神经网络从生物学隐喻转变为数学实体，揭示了经典矩阵分析与现代机器学习之间的深刻联系。



考虑一个具有 L 层的神经网络。每一层执行由方程 (13.1) 描述的变换，但现在我们将这些操作链式连接在一起：

$$\begin{aligned} \boldsymbol{h}_1 &= \sigma(W_1\boldsymbol{x} + \boldsymbol{b}_1) \\ \boldsymbol{h}_2 &= \sigma(W_2\boldsymbol{h}_1 + \boldsymbol{b}_2) \\ &\vdots \\ \boldsymbol{y} &= W_L\boldsymbol{h}_{L-1} + \boldsymbol{b}_L \end{aligned} \tag{13.2}$$

其中 $\boldsymbol{x}$ 表示输入，$\boldsymbol{h}_i$ 是 *hidden layers*，$\boldsymbol{y}$ 提供输出。没有非线性函数 σ 时，这将简化为仅仅是矩阵乘法 $W_L \cdots W_1$ —— 正是我们在矩阵分解的上下文中研究过的形式。激活函数将这一线性组合转化为更具表现力的形式，但矩阵乘法的基本结构仍然是至关重要的。

Think: 如果没有激活函数，这种组合将退化为纯粹的矩阵乘法 $W_L \cdots W_1$。非线性带来了超越矩阵分解的表达能力，但也将一个凸优化问题转变为非凸问题。

定义 13.4（前馈神经网络）。*feedforward neural network* 是一个函数 $f\colon \mathbb{R}^n \to \mathbb{R}^m$，由权重矩阵 $\{W_i\}$ 和偏置向量 $\{\boldsymbol{b}_i\}$ 参数化，通过连续应用方程（13.2）来转换其输入。层 i 的 *width* 等于 W_i 的行数，而 *depth* 等于总层数 L。

•

该c 在网络深度和宽度之间的选择呈现出基本的 心理的

与低秩近似中遇到的权衡类似。就像截断不同数量的奇异值在近似精度与复杂度之间取得平衡一样，神经网络的宽度和深度控制其表达能力。宽而浅的网络有效地学习目标函数的大规模基扩展——就像在低秩近似中保留许多奇异值一样。深度网络则组合更简单的变换，可能更有效地捕捉层次结构：

定理 13.5（深度网络表达能力）。*There exist functions that can be computed by deep neural networks with polynomial size but require exponential size when restricted to bounded depth. More precisely, for any n there exists a function $f\colon \{0,1\}^n \to \{0,1\}$ computable by a network of depth $O(n)$ and total size $O(n)$ that requires size $2^{n/\log n}$ when restricted to depth $O(\log n)$.*

Historical Note: 2015 年残差网络（ResNets）的引入使得训练超过 100 层的网络成为可能，打破了此前的深度壁垒。这一洞见源于提出了 *"what if layers learned differences rather than absolute transformations?"*

现代架构通过呼应矩阵条件化思想的创新来增强这一基本结构。*Skip connections* 允许信息直接绕过各层：

$$\boldsymbol{h}_{i+1} = \sigma(W_i \boldsymbol{h}_i + \boldsymbol{b}_i) + \boldsymbol{h}_i$$

与第12章的正则化技术类似，这些残差路径通过为梯度流提供直接通道来提高数值稳定性。它们将学习任务从近似目标函数转变为学习其改进——这通常是一个条件更好的优化问题。类似地，*normalization layers* 对其输入进行标准化：

$$\hat{\boldsymbol{h}}_i = \frac{\boldsymbol{h}_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}}$$

Compare: 就像第 1 章中通过列缩放改善矩阵条件数一样，归一化层通过对中间尺度的精细控制来稳定计算。关键区别在于，这些尺度是从数据中学习得到的，而不是通过解析方式计算出来的。

其中 μ_i、σ_i 估计均值和标准差。该操作通过稳定中间表示的尺度来控制权重矩阵的条件数，这与第 1 章中研究的列缩放改善矩阵条件性的作用类似。

这些架构选择体现了一个深刻的原则：神经网络的有效性不仅源于其表达能力，还源于其结构如何促成稳定的优化。正如精心选择的基变换在经典线性代数中揭示了结构一样，现代网络架构通过对运算的周到组合，为高效学习创造了通路。下一节将探讨这一学习过程如何通过系统地应用矩阵微积分来运作。

13.3 Chains & Backpropagation

神经网络训练的挑战不在于理解要优化什么——显然我们寻求使预测误差最小的参数——而在于计算参数的微小变化如何影响网络输出。在跨越多层、可能多达数十亿个参数的情况下，直接计算导数似乎复杂得令人绝望。然而，多元微积分的链式法则恰恰提供了我们所需的工具，将看似难以处理的计算转化为一种优雅的递归算法。

Historical Note: 反向传播最早出现在 Werbos 于1974年的论文中，随后被忽视。直到20世纪80年代，多位研究者的并行发现才认识到它对神经网络的重要性。其关键洞见——通过递归计算系统地应用链式法则——源自20世纪60年代的反向模式自动微分技术。

首先考虑我们希望计算的对象的数学结构。给定一个具有参数Ψ (编码所有权重和偏置)的网络，我们试图最小化某个损失函数$\mathcal{L}(\Psi)$，它衡量在训练数据上的预测误差。挑战在于计算$[\partial\mathcal{L}/\partial\Psi]$，即损失对参数的导数。尽管$\mathcal{L}$最终是标量值，但它源自许多运算的复杂组合。

损失函数的常见选择包括：

- 均方误差：$\mathcal{L}(\boldsymbol{y},\ \hat{\boldsymbol{y}}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- 用于分类的交叉熵：$\mathcal{L}(\boldsymbol{y},\hat{\boldsymbol{y}}) = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$
- 用于鲁棒性的 L1 损失：$\mathcal{L}(\boldsymbol{y},\ \hat{\boldsymbol{y}}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

Nota bene: loss 和 *cost function* 这两个术语在文献中可互换使用，其中“loss”在现代机器学习实践中更为常见。

其中$\boldsymbol{y}$表示真实值，$\hat{\boldsymbol{y}}$表示网络的预测值。

网络的每一层都会对其输入执行一次简单的变换。对于具有权重$W_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$和偏置$\boldsymbol{b}_\ell \in \mathbb{R}^{n_\ell}$的一层，我们计算：

$$\boldsymbol{h}_\ell = \sigma(W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell)$$

其中σ是逐分量应用的非线性激活函数，而$\boldsymbol{h}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$表示前一层的激活值。关于参数的该操作的导数是直接的：

$$\left[\frac{\partial}{\partial W_\ell}(W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell)\right] = \begin{bmatrix}\boldsymbol{h}_{\ell-1}^T & \mathbf{0}\end{bmatrix} \quad \text{and} \quad \left[\frac{\partial}{\partial \boldsymbol{b}_\ell}(W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell)\right] = I$$

挑战在于通过链式法则将这些简单的部分结合起来。

定义 13.6（误差信号）。位于第ℓ层的 *error signal* $[\delta_\ell]$是损失函数相对于该层预激活输出的导数：

$$[\delta_\ell] = \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{z}_\ell}\right] \in \mathbb{R}^{1\times n_\ell}$$

其中$\boldsymbol{z}_\ell = W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell \in \mathbb{R}^{n_\ell}$表示该层的预激活值。 •

引理 13.7（反向传播规则）。*Let $\mathcal{L}$ be a scalar loss function of network output. For any layer ℓ in a neural network:*

$$[\delta_\ell] = [\delta_{\ell+1}]\, W_{\ell+1}\, \mathrm{DIAG}(\sigma'(\boldsymbol{z}_\ell))$$

where σ' denotes the derivative of the activation function. The parameter gradients are then:

$$\left[\frac{\partial \mathcal{L}}{\partial W_\ell}\right] = [\delta_\ell]^T\, \boldsymbol{h}_{\ell-1}^T \quad \textit{and} \quad \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{b}_\ell}\right] = [\delta_\ell]^T$$

Proof. 通过链式法则：

$$\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{z}_\ell}\right] = \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{z}_{\ell+1}}\right] \left[\frac{\partial \boldsymbol{z}_{\ell+1}}{\partial \boldsymbol{h}_\ell}\right] \left[\frac{\partial \boldsymbol{h}_\ell}{\partial \boldsymbol{z}_\ell}\right]$$

结果来自对每个因子进行计算：$[\partial \boldsymbol{z}_{\ell+1}/\partial \boldsymbol{h}_\ell] = W_{\ell+1}$ 和 $[\partial \boldsymbol{h}_\ell/\partial \boldsymbol{z}_\ell] = \mathrm{diag}(\sigma'(\boldsymbol{z}_\ell))$，并仔细注意每个矩阵乘积的维度。

□

例13.8（简单神经网络）。考虑一个具有两个隐藏层、宽度分别为 n_1 和 n_2、处理输入 $\boldsymbol{x} \in \mathbb{R}^{n_0}$ 的网络：

$$\begin{aligned} \boldsymbol{h}_1 &= \sigma(W_1 \boldsymbol{x} + \boldsymbol{b}_1) \\ \boldsymbol{h}_2 &= \sigma(W_2 \boldsymbol{h}_1 + \boldsymbol{b}_2) \\ \boldsymbol{y} &= W_3 \boldsymbol{h}_2 + \boldsymbol{b}_3 \end{aligned}$$

其中 $W_1 \in \mathbb{R}^{n_1 \times n_0}$、$W_2 \in \mathbb{R}^{n_2 \times n_1}$ 和 $W_3 \in \mathbb{R}^{n_3 \times n_2}$。对于目标为 $\boldsymbol{t}$ 的均方误差损失 $\mathcal{L} = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{t}\|^2$，反向传播计算：

$$\begin{aligned} [\delta_3] &= [\boldsymbol{y} - \boldsymbol{t}]^T \\ [\delta_2] &= [\delta_3]\, W_3\, \mathrm{DIAG}(\sigma'(\boldsymbol{z}_2)) \\ [\delta_1] &= [\delta_2]\, W_2\, \mathrm{DIAG}(\sigma'(\boldsymbol{z}_1)) \end{aligned}$$

这些误差信号随后通过与存储的激活值进行矩阵乘法来产生参数梯度。

◇

Nota bene: 在反向传播过程中对矩阵维度的细致跟踪揭示了该算法为何能为每个参数矩阵产生尺寸正确的梯度。

算法的效率源于对计算的精心组织：

- 前向传播存储激活值 $\boldsymbol{h}_\ell$ 和预激活值 $\boldsymbol{z}_\ell$
- 反向传播从输出到输入计算误差信号 $[\delta_\ell]$
- 参数梯度通过使用存储的值进行简单矩阵运算得出

这种连接前向与反向传递的递归结构呼应了我们在第 3 章首次遇到的互补子空间。只是

作为核和图像分解的线性变换，反向传播的前向和反向传递分解了信息在网络中的流动。误差信号 δ_ℓ 精确地衡量了变化是如何通过这一结构向后传播的。

定理 13.9（反向传播复杂度）。*For a network with Λ layers each of width at most n, backpropagation computes all parameter derivatives in time $O(\Lambda n^2)$ using storage $O(\Lambda n)$.*

这种卓越的效率——深度线性、宽度平方——将神经网络训练从理论上的可能性转变为实践中的现实。像第10章中研究的矩阵分解一样，反向传播通过精心分解计算来实现其强大功能。链式法则提供了数学基础，而深思熟虑的算法设计则将这一洞察转化为高效的实现。

Foreshadowing: 反向传播的效率在我们扩展到具有数十亿参数的深度网络时变得至关重要。下一节将展示如何通过随机采样进一步减少计算成本。

反向传播的基本洞察——影响像信息一样在网络中反向流动——不仅仅提供了计算效率。它揭示了神经网络如何通过系统地测量参数影响来学习，这一过程由精确的数学原理指导。微积分与计算图的融合将基于梯度的学习这一抽象可能性转化为实际现实。

13.4 Stochastic Gradient Descent

规模的数学促使了概率思维。尽管反向传播提供了一种优雅的算法来计算梯度，但将其应用于每一个训练样本对于大型数据集来说是过于昂贵的。就像河流在复杂地形中寻找高效的路径一样，随机逼近通过精心采样将精确但不可处理的计算转化为高效的估计。这个原则——随机性可以加速计算同时保持准确性——贯穿于现代机器学习。

定义 13.10（随机梯度下降）。设 $\{\Psi_t\}_{t\geq 0}$ 表示一个参数向量序列，按照以下方式迭代更新：

$$\Psi_{t+1} = \Psi_t - \eta_t \left[\frac{\widehat{\partial \mathcal{L}}}{\partial \Psi}\right]^T$$

其中 $\eta_t > 0$ 是学习率，而 $[\widehat{\partial \mathcal{L}}/\partial \Psi]$ 表示损失梯度的随机估计。

•

考虑我们损失函数 $\mathcal{L}$ 在 n 个样本数据集上的真实梯度平均值：

$$\left[\frac{\partial \mathcal{L}}{\partial \Psi}\right] = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial \mathcal{L}_i}{\partial \Psi}\right]$$

计算这一结果需要对数据集进行完整的遍历，但总和的结构暗示了一个自然的近似方法。通过均匀随机地采样一个大小为 $b \ll n$ 的小批量 $\mathcal{B}$，我们可以得到一个无偏估计：

$$\left[\frac{\widehat{\partial \mathcal{L}}}{\partial \Psi}\right] = \frac{1}{b}\sum_{i\in\mathcal{B}}\left[\frac{\partial \mathcal{L}_i}{\partial \Psi}\right]$$

这种随机梯度为现代神经网络训练提供了基础。

Historical Note: “梯度下降”一词源于早期对标量值函数的关注，这些函数的导数自然地形成梯度向量。现代用法已涵盖矩阵导数，尽管这一名称仍然沿用。

引理13.11（小批量性质）。*Let σ^2 denote the variance of individual gradient estimates. The mini-batch gradient estimator satisfies:*

1. *Unbiasedness:* $\mathbb{E}\left[\frac{\widehat{\partial \mathcal{L}}}{\partial \Psi}\right] = \left[\frac{\partial \mathcal{L}}{\partial \Psi}\right]$
2. *Variance bound:* $\mathbb{V}\left[\frac{\widehat{\partial \mathcal{L}}}{\partial \Psi}\right] = \mathbb{E}\left\|\left[\frac{\widehat{\partial \mathcal{L}}}{\partial \Psi}\right] - \left[\frac{\partial \mathcal{L}}{\partial \Psi}\right]\right\|_F^2 \leq \frac{\sigma^2 b}{}$

where $\|\cdot\|_F$ denotes the Frobenius norm.

例子 13.12（二分类）。考虑训练一个简单的网络，用于使用逻辑损失分类 $\mathbb{R}^2$ 中的点。给定参数 $\boldsymbol{w} \in \mathbb{R}^2$ 和 $c \in \mathbb{R}$，该网络计算：

$$p(x) = \frac{1}{1 + e^{-(\boldsymbol{w}^T\boldsymbol{x} + c)}}$$

对于由点 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ 组成且 $y_i \in \{0, 1\}$ 的数据集，损失变为：

$$\mathcal{L}(\boldsymbol{w}, c) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log p(\boldsymbol{x}_i) + (1 - y_i)\log(1 - p(\boldsymbol{x}_i))\right]$$

当小批量大小为 $b = 2$ 时，每次迭代：

1. 均匀随机采样两个点 i, j
2. 计算预测结果 $p(\boldsymbol{x}_i), p(\boldsymbol{x}_j)$
3. 使用平均梯度更新参数：

$$\begin{aligned} \boldsymbol{w}_{t+1} &= \boldsymbol{w}_t - \tfrac{1}{2}\eta_t \textstyle\sum_{k\in\{i,j\}}(p(\boldsymbol{x}_k) - y_k)\boldsymbol{x}_k \\ c_{t+1} &= c_t - \tfrac{1}{2}\eta_t \textstyle\sum_{k\in\{i,j\}}(p(\boldsymbol{x}_k) - y_k) \end{aligned}$$

随机更新在保持收敛性的同时，被证明比全批量计算高效得多。

◇

像第12章中研究的低秩近似一样，小批量导数通过牺牲精度来提高计算效率。通过动量方法，矩阵条件数的联系显现出来，动量方法将一阶更新转换为二阶信息的近似：

$$\begin{aligned} \boldsymbol{m}_t &= \beta \boldsymbol{m}_{t-1} + (1-\beta)\left[\frac{\partial \widehat{\mathcal{L}}}{\partial \Psi}\right]^T \\ \Psi_{t+1} &= \Psi_t - \eta_t \boldsymbol{m}_t \end{aligned}$$

其中 $\beta \in [0, 1)$ 控制动量，而 $\boldsymbol{m}_0 =$ 为 0。这种导数的平均化减少了方差，同时加速了收敛，特别是在优化面类似于病态系统特征的拉长谷时。

Nota bene: 像第9章的迭代方法一样，动量在步骤之间积累信息，以加速收敛。参数 $\beta = 0.9$ 在许多应用中证明是有效的。

为了证明SGD收敛性的结果，需要引入一些额外的技术术语。它们对我们的叙述并非必需，但为了精确性，仍然包括在内。

定义 13.13（平滑性与强凸性）。一个可微函数 $f\colon \mathbb{R}^n \to \mathbb{R}$ 是：

1. *L-smooth* 如果其梯度是利普希茨连续的，且参数为 $L > 0$：

$$\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\| \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

Ouch! 是的，如果你真的想学习人工智能，你可能需要更多的数学知识。

2. *μ-strongly convex* 如果对于某些 $\mu > 0$：

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

实际上，L-光滑性提供了梯度变化的上界，而μ-强凸性则确保了所有方向上最小的曲率。

Nota bene: 这些条件与矩阵条件性直接相关：对于二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A \boldsymbol{x}$，比值 L/μ 等于 A 的条件数。

定理 13.14 （SGD 收敛性）。*Let $\mathcal{L}$ be μ-strongly convex and L-smooth. For learning rate schedule $\eta_t = \frac{2}{\mu(t+1)}$ and mini-batch size b, stochastic gradient descent converges in expectation:*

$$\mathbb{E}[\|\Psi_t - \Psi^*\|^2] \leq \frac{4L}{b\mu^2(t+1)} \max\{\|\Psi_0 - \Psi^*\|^2, \sigma^2\} = O\left(\frac{1}{t}\right)$$

where Ψ^* denotes the optimal parameters and σ^2 bounds the variance of individual gradients. For non-convex losses typical in deep learning, convergence to local minima occurs under suitable regularity conditions.

Interpretation: 误差至少以 $1/t$ 的速度下降，其速率由问题的条件数 L/μ 以及梯度噪声水平 σ^2/b 所控制。

现代实践表明，对这些基本原则进行了若干改进：

1. 梯度裁剪以限制更新幅度： 2. 随时间递减的学习率调度：

3. 深度网络的逐层自适应学习率 4. 通过加入随机噪声进行正则化

每项改进都针对优化中的特定挑战，同时保持随机近似的核心思想。

Historical Note: 2010 年代，自适应方法（如 Adam 和 RMSprop）的发展通过使训练对超参数选择更加鲁棒而改变了深度学习，并有效地为每个参数学习独立的学习率。

随机梯度下降的卓越有效性源自深刻的数学原理。与第6章的最小二乘方法类似，后者通过可行的计算来近似最优解，小批量梯度为优化提供了合适的方向。这些估计中固有的噪声往往被证明是有益的，既能帮助网络摆脱糟糕的局部极小值，又能提供隐式正则化。随机近似与经典优化的这种融合，将神经网络训练从理论上的可能性转变为实践中的现实。

13.5 Attention & Transformers

线性代数与智能的融合在 Transformer 架构的注意力机制中达到了最充分的体现。正如神圣的铁匠以圆规与天平丈量缥缈的光芒，注意力引入了一项深刻的创新：其矩阵的元素本身是从数据中动态生成的。每个元素不再只是固定的权重，而是一种被学习到的关系——通过系统性的内积计算得到的相关性度量。这种机制——构思优雅却在应用中威力强大——将本文中所研究的固定权重转化为自适应的变换，使计算能够聚焦于真正需要之处。

注意力背后的直觉自然地源自信息检索系统。设想一个图书馆，每本书既有描述性元数据（标题、主题、关键词），也有实际内容。用户的搜索查询必须以某种方式与这些元数据进行匹配，以确定应当检查并组合哪些书的内容。这个过程——计算查询与众多可能的键之间的相关性分数，然后利用这些分数来组合值——为神经注意力提供了模板。

Think: 查询-键-值范式概括了我们自然组织信息的方式。查询表达我们所寻求的内容，键提供可搜索的描述符，而值则承载要被检索的实际内容。

数学从三种已学习的投影开始，这些投影将输入向量转换为承担不同角色的表示。给定输入向量 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$，我们首先将每个向量投影为一个查询向量，用于表示该位置在寻找什么；一个键向量，用于编码它向他人提供什么；以及一个值向量，包含其实际内容。这些角色分别通过已学习的矩阵 W_Q、W_K 和 W_V 产生：

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

其中 X 将输入向量堆叠为行。像坐标 tra nsfor-

与第4章中研究的变换相比，这些投影将输入特征与注意力计算的自然轴对齐——但现在这些轴本身是通过学习涌现出来的。

定义 13.15（注意力机制）。给定输入矩阵 $X \in \mathbb{R}^{n \times d}$，*attention mechanism* 通过三个阶段计算输出 Y：

$$A = QK^T, \quad \tilde{A}_{ij} = \frac{\exp(a_{ij}/\sqrt{d_k})}{\sum_{j=1}^{n} \exp(a_{ij}/\sqrt{d_k})}, \quad Y = \tilde{A}V \tag{13.3}$$

其中 d_k 表示查询/键空间的维度。指数归一化在确保权重为正且总和为一的同时，增强了对强匹配的关注。

Historical Note: 注意力机制起源于2015年的神经机器翻译，但其真正的威力直到2017年的 Transformer 架构出现才得以显现。

考虑这种机制如何处理自然语言。在诗句“Tyger, tyger, burning bright”中，每个词的表示都会经过注意力计算。查询投影学会捕捉每个词所需要的上下文——“burning”可能会查询它所修饰的对象。键投影学会捕捉每个词所提供的上下文——“tyger”作为主语形成了强烈的键信号。值投影学会编码可以被有效组合的意义——也许捕捉到“bright”修饰的是燃烧的性质。随后，注意力权重决定了这些意义如何在不同位置之间流动。

这一机制的显著有效性源于其将经典计算与自适应计算的融合。尽管它由熟悉的操作——矩阵乘法和归一化——构成，但它们的精心组合创造了一种能够适应输入的变换。核心矩阵乘积 QK^T 通过内积衡量所有位置对之间的相关性，而归一化确保这些分数形成合适的注意力分布。

定理 13.16（注意力性质）。*The attention mechanism satisfies:*

$$\textit{row-stochastic:} \quad \sum_{j=1}^{n} \tilde{a}_{ij} = 1 \quad \textit{and} \quad \tilde{a}_{ij} \geq 0$$

$$\textit{permutation equivariant:} \quad \pi(Y) = \mathit{Attention}(\pi(X))$$

for any permutation π. The computation requires $O(n^2$ 最大$\{d_k, d_v\})$ operations and $O(n^2)$ memory.

Transformer 架构通常并非计算单一的注意力模式，而是并行地采用多个注意力“头”。

每个头学习不同的查询/键/值投影，使网络能够同时捕获多种类型的关系。随后，一个学习得到的输出变换 W_O 将这些并行流进行合并：

$$\text{MultiHead}(X) = W_O \begin{bmatrix} \text{head}_1 \\ \vdots \\ \text{head}_h \end{bmatrix}$$

Think: 类似于残差网络中的并行路径或 SVD 的多个奇异向量，多头注意力使网络能够同时捕获关系的不同方面。

完整的 Transformer 架构将这些注意力模块与若干呼应前几章概念的创新相结合。层归一化像第一章中的列缩放一样稳定计算。残差连接如第九章所示，使梯度得以顺畅传播。位置编码通过三角函数为序列引入顺序信息，而前馈层则通过标准线性变换处理注意力的输出。每个组件都建立在本文贯穿始终所发展的基本原理之上，然而它们的组合却超越了各部分之和。

注意力的计算需求会随着序列长度呈二次增长，因为计算所有键–查询交互需要进行矩阵乘法 $QK^T \in \mathbb{R}^{n \times n}$。现代实现通过稀疏注意力模式、巧妙的矩阵乘法算法以及架构创新来应对这一问题。然而，其基础操作仍然是矩阵乘法——正是自第 1 章以来一直指导我们发展的同一操作。

注意力机制体现了经典数学如何转化为现代人工智能。尽管它是由矩阵和内积构建的，但它引入了一个新的动态层次，其中变换会根据输入进行调整。这种将永恒数学原理与学习适应相融合的方式，推动了自然语言处理、计算机视觉和其他领域的显著进展。注意力机制将线性代数的精确性与智能所需的灵活性结合在一起。

13.6 Representation Learning

虽然前面的章节揭示了网络如何学习去近似函数，更深层的洞见来自于理解它们如何发现使这种近似变得自然的表示。这一过程——学习最优的特征空间，而不仅仅是拟合函数

Foreshadowing: 从手工设计到学习到的表示的演进，与我们从解析的矩阵分解到训练的神经网络的历程相呼应。

– 将前几章的抽象机制与现代人工智能的卓越能力联系起来。就像第10章中从矩阵结构中自然涌现的奇异向量一样，学习到的表征-

呈现捕捉数据中的基本模式，但通过精心构造的变换超越了线性。

首先考虑经典特征工程与学习到的表示之间的差异。手工设计的特征提取器基于领域知识应用固定的变换：用于图像的边缘检测器、用于信号的频率分析、用于文本的句法解析树。尽管源自专家经验，这类特征仍受人类直觉的限制。神经网络通过学习自身的变换超越了这一限制，并通过端到端训练发现最适合其任务的表示空间。

定义 13.17（表示空间）。用于数据 $\mathcal{X}$ 的 *representation space* $\mathcal{H}$ 是一个配备了学习得到的变换 $f\colon \mathcal{X} \to \mathcal{H}$ 的向量空间，该变换将输入映射为特征。在具有 L 层的深度网络中，每一层隐藏层都通过其学习到的权重和非线性定义这样一个空间 $\mathcal{H}_\ell$：

$$\mathcal{H}_\ell = \{\boldsymbol{h}_\ell = \sigma_\ell(W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell) : \boldsymbol{h}_{\ell-1} \in \mathcal{H}_{\ell-1}\}$$

whe 其中 W_ℓ、$\boldsymbol{b}_\ell$ 是学习得到的参数，而 σ_ℓ 提供非线性 earity。 •

这种学习到的结构与*manifold hypothesis*有着深刻的联系——该原理认为，高维数据往往集中在较低维的非线性曲面附近。尽管第11章的PCA方法发现了数据的最优*linear*投影，深度网络通过变换的组合来学习*nonlinear*嵌入。每一层的权重定义了一个映射：

Think: 你见过的每一幅图像都存在于图像空间中的一个高维子域，而该子域的补集具有更高的维度。

$$\boldsymbol{h}_\ell = \sigma(W_\ell \boldsymbol{h}_{\ell-1} + \boldsymbol{b}_\ell)$$

从而逐步将输入转换为更具结构化的表示。非线性函数 σ 使这些映射能够沿着弯曲的流形进行，而跳跃连接和注意力机制则在表示空间中提供了捷径。

引理 13.18（表示分解）。*Let f_ℓ denote the transformation implemented by layer ℓ of a neural network. Each f_ℓ admits decomposition:*

$$f_\ell = \sigma_\ell \circ \tilde{f}_\ell$$

where $\tilde{f}_\ell$ is linear and σ_ℓ provides nonlinearity. The full network learns representations by composing such transformations:

$$f = f_L \circ f_{L-1} \circ \cdots \circ f_1$$

Moreover, each linear component $\tilde{f}_\ell$ induces a decomposition of its domain through the fundamental subspaces:

$$\mathcal{H}_{\ell-1} = \ker(\tilde{f}_\ell) \boxplus (\ker \tilde{f}_\ell)^\perp$$

Nota bene: 第三章的基本子空间在每一层学习到的变换中自然地再次出现，只是它们如今是通过优化而非分析被发现的。

The network learns both the transformations and their associated subspace decompositions through training.

例13.19（视觉表示）。考虑卷积网络如何通过其各层对图像数据进行变换。早期表示捕捉诸如边缘和纹理等局部模式——这些特征让人联想到手工设计的滤波器。更深层学习到越来越抽象的特征：

- 层 1：边缘和方向
- 第2层：纹理和简单形状
- 第3层：对象部件和复杂模式
- 第4层：完整的物体与场景

这种层级结构在训练过程中自然出现，每一层都会学习能够简化下一层任务的表示。这种涌现背后有着精确的数学模式：每一层的权重 W_ℓ 都实现了线性映射，其奇异值分解揭示了所学习的特征层次结构。

$$W_\ell = U_\ell \Sigma_\ell V_\ell^T$$

右奇异向量 V_ℓ 捕捉输入模式，而左奇异向量 U_ℓ 提供变换后的表示基。

Example: 当 $\boldsymbol{v}_i$ 捕捉到面部特征而 $\boldsymbol{u}_i$ 提供其变换后的表示时，可能会涌现出一个人脸检测器——就像奇异向量与数据中的自然模式对齐一样。

◇

通过第12章的近似理论视角，学习到的表示的力量变得清晰。正如低秩矩阵分解能够捕捉数据中的主导模式一样，深度网络学习到的特征使复杂任务在变换后的空间中近似线性。这一原则——良好的表示会将原本复杂的问题线性化——在现代架构中反复出现：

定理 13.20（表示线性化）。*Let $f: \mathcal{X} \to \mathcal{Y}$ be a smooth function between manifolds, and let $\{\boldsymbol{h}_\ell\}_{\ell=1}^L$ be the sequence of representations learned by a deep network trained to approximate f. Then under suitable regularity conditions:*

1. *The representation error $\|\boldsymbol{h}_\ell - \boldsymbol{h}_{\ell-1}\|$ decreases with depth*
2. *The nonlinearity of mappings between successive layers decreases*
3. *The final layers implement approximately linear transformations*

BONUS! 这种渐进式线性化与通用逼近理论相联系：深度网络可以通过学习到的变换逐步将其图形拉直，从而表示任何连续函数。

示例 13.21（词嵌入）。将词表示为向量提供了一个关于学习到的结构的鲜明例子。与通过词性或语义类别等人工设计的特征来编码词语不同，现代语言模型直接从文本中的共现模式中学习嵌入。一个词的表示源自一个学习得到的线性变换 $W \in \mathbb{R}^{d \times v}$，该变换将独热编码映射为稠密向量：

$$\boldsymbol{e}_w = W\boldsymbol{x}_w$$

其中 $\boldsymbol{x}_w \in \mathbb{R}^v$ 表示词 w 的独热编码，$d \ll v$ 是嵌入维度。W 的行构成了一个学习到的基，捕捉了语义关系：

1. 相似的词在嵌入空间中聚集在一起 2. 向量之间的差异编码类比关系 3. 投影到主成分上揭示语义场 嵌入矩阵 W 有效地学习到一个低维流形，平滑地刻画语言结构。

Caveat: 低维约束 $d \ll v$ 至关重要——它迫使网络发现高效的表示，而不是记忆表层模式。

◇

表征学习的研究自然地与第11章中发展起来的统计视角相联系。正如主成分分析通过最大化数据中的可解释方差一样，学习到的表征通过其架构和训练来优化隐含的目标。然而，与 PCA 唯一的最优投影不同，学习到的表征可能会在不同的训练运行或架构选择之间发生变化。这种变异性往往是有益的，因为不同的随机初始化可以发现互补的特征空间。

更正式地说，考察一个深度网络如何通过连续的表征逐步变换其输入空间。每一层都在隐藏空间之间定义一个映射 $f_\ell : \mathcal{H}_{\ell-1} \to \mathcal{H}_\ell$，最终汇聚为某个最终表征 $\mathcal{H}_L$。网络的力量不在于任何单一的变换，而在于这一系列映射如何逐步构建结构。每一层都会对其输入进行塑形与细化，直到复杂模式从简单的基础中涌现。

表示学习的数学揭示了一个深刻的原理：智能并非来自原始计算，而是来自能够将复杂模式简化的学习变换。这一见解——即好的表示使困难的问题变得简单——引导我们向将线性代数最终合成到人工智能中的方向发展。下一节将探讨这些学习到的结构如何最终与第三章首次遇到的基本定理相连接。

引领着我们从经典数学走向现代人工智能的旅程。

13.7 Deep Linear Algebra

智能的数学通过基本模式显现出来。尽管神经网络似乎超越了前几章中发展起来的结构，但其最深层的架构回响并放大了我们在第三章首次遇到的基本定理。在每一层，学习到的表示之间的每一次变换都保留了贯穿我们整个发展过程的本质分解。这种统一——经典分解与学习结构的统一——标志着我们数学旅程的最终转变。

思考神经网络如何塑造其表征空间。每一层实现的不仅是线性变换后接非线性，而是一种学习到的分解，反映了线性代数基本定理中的核–像关系。当具有权重矩阵 W 的一层将 $\mathbb{R}^n$ 映射到 $\mathbb{R}^m$ 时，它诱导出两种互补的分解：

$$\mathbb{R}^n = \ker(W) \boxplus (\ker W)^\perp \quad \text{and} \quad \mathbb{R}^m = \operatorname{im}(W) \boxplus (\operatorname{im} W)^\perp$$

与前几章研究的固定分解不同，这些空间是在训练过程中涌现的——网络会发现哪些方向需要保留，哪些方向需要塌缩。这种通过学习获得的适应性解释了为何神经网络常常能够捕捉到手工设计特征难以发现的模式：它们将其基本分解与数据中的自然结构对齐。

例子 13.22（学习的分解）。考虑通过狭窄的隐藏层压缩数据的自编码器。编码器 $E : \mathbb{R}^n \to \mathbb{R}^k$ 和解码器 $D : \mathbb{R}^k \to \mathbb{R}^n$ ($k \ll n$) 通过学习到的变换实现维度减少，其基本空间获得深刻的意义：

- *kernel* 的 $\ker(E)$ 精确地捕捉了在训练过程中被视为无关的那些特征
- *image* $\operatorname{im}(E)$ 提供了学习到的表示——一个最优的 k 维摘要
- *cokernel* $\operatorname{coker}(E)$ 用于衡量几何重建损失
- *coimage* 的 $\operatorname{coim}(E)$ 揭示了在模去无关变化下的有效特征空间

该网络通过在最小化重建误差的同时遵循由基本定理所保证的正交结构来发现这些空间。每个空间的涌现并非通过数学规定，而是

through learned adaptation to data. ◇

这个原理——神经网络学习任务相关的基本分解版本——在现代架构中随处可见。在卷积网络中，每一层的核捕捉到被认为无关的局部模式，而其图像保留了任务相关的特征。注意力机制实现了学习到的等价关系，通过识别相似的标记，有效地构建了自适应的商空间。残差连接通过这些分解创建了捷径，为信息流提供了直接路径，同时保持了底层结构。

四个基本空间，最初作为抽象概念出现在第3章中，因此在神经网络如何学习和处理信息的方式中找到了它们最深刻的体现：

- *kernel* 从零空间演变到学习到的不变性，捕捉应该视为等效的输入变异。
- *image* 从范围变换为学习的流形，提供有意义地编码信息的表示。
- *cokernel* 测量的不仅仅是代数缺陷，还有学习能力。
- *coimage* 显示为最优商空间，识别与任务相关的特征

通过训练，这些空间自然地与数据中固有的结构对齐。核学习任务相关的不变性；图像约束可能的表示；余核指导架构设计；而余像捕捉有效特征。每一层发现自己版本的这些基本分解，将我们的抽象分析转化为学习到的智能。

我们的线性代数探索以与其开始时相同的方式结束：通过表面上不同的概念的深刻统一。基本定理首次揭示了所有线性变换背后的互补子空间，在神经网络的自适应分解中达到了最完整的表达。从这个角度看，人工智能并非一种革命性的断裂，而是一种革命性的精炼——通过这些页面中开发的结构，将永恒的数学原理转化为学习形式。就像从矩阵结构中出现的奇异值或捕捉渐近行为的特征空间一样，线性代数的基本空间在现代人工智能的学习模式中重新显现。

Convolutional Neural Networks

人类视觉与机器视觉的融合，最深刻地体现在神经网络处理图像的方式中。当人类观看一张照片时，我们感知的并非只是像素，而是由形状、纹理和边缘之间关系编织而成的有意义的模式。这种深层结构——视觉信息的自然几何——指导着现代人工智能的设计。卷积神经网络之所以具备非凡的能力，正是因为它们将这些感知原则直接编码进其架构之中，把标准网络中致密的矩阵运算转化为专门化的变换，从而映射并复现人类与机器视觉解析视觉世界的方式。

首先考虑图像是如何编码其信息的。一个大小为 $h \times w$ 的灰度图像呈现为一个强度矩阵，但将其视为任意数组会丢弃关键的空间关系。当为了神经网络处理而被重塑为向量时，一张普通的 256×256 图像就变成了 $\mathbb{R}^{65536}$ 中的一个点——这是一个维度极高的空间，在其中学习任意变换被证明是极其低效的。然而，我们试图发现的模式——那些将猫与狗或肿瘤与健康组织区分开的特征——依赖于相邻像素之间的关系，而非任意的长程连接。

这一局部性原则通过专门化的权重矩阵转化为数学结构，这些矩阵编码了空间关系。与学习任意变换不同，卷积层将其线性运算限制为遵循视觉数据的网格状拓扑。一个小型滤波器 $K \in \mathbb{R}^{k \times k}$ 定义了一个局部特征检测器，依次检查图像的每个区域——相比全连接层的 $hw \times hw$ 矩阵，参数数量大幅减少。

这种方法的力量源于其将生物学启发与数学结构相融合。正如视觉皮层通过局部感受野来处理输入一样，卷积网络通过精心受约束的线性变换学习分层的特征检测器。早期层通常会发现简单的模式：

- 不同方向的边缘检测器
- 颜色对比边界
- 局部纹理元素

它们在后续层中结合，表示越来越复杂的特征，这些特征均由对更简单模式的学习组合而涌现。

现代架构通过借鉴矩阵条件化思想的创新来增强这一基本结构。批量归一化层在空间位置上对其输入进行标准化：

Historical Note: 尽管受到生物视觉的启发，卷积网络只有在经过严谨的数学分析、揭示了恰当初始化和正则化的重要性之后，才取得了实际上的成功。2012 年 AlexNet 的突破性表现展示了理论洞见如何促成工程实践。

$$\hat{x}_{ijk} = \frac{x_{ijk} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}}$$

其中 k 表示特征通道的索引。类似于第 1 章中改善矩阵条件性的缩放技术，这一操作在训练过程中稳定了梯度流。同样，残差连接创建了捷径，使网络即使在深度很大的情况下也能更容易地进行优化：

$$Y = X + \sigma(K_2 * \sigma(K_1 * X))$$

这种结构让人联想到第3章中研究的分裂方式，在保持由卷积所施加的空间结构的同时，为梯度流动提供了直接路径。

卷积网络的卓越有效性源于生物启发与数学原理的这种融合。与本章研究的低秩近似类似，它们通过精心设计的约束实现效率——并非通过限制表达能力，而是通过将视觉数据的已知属性编码到其矩阵运算中。早期层学习局部化的特征检测器：

- 垂直和水平边缘检测器
- 中心-周围对比滤波器
- 简单的纹理元素

每一个都代表一种通过优化发现的专门线性变换，同时受架构设计约束，以尊重视觉数据的结构。

这些学到的特征往往展现出显著的普适性。在自然图像上训练的网络，无论其最终任务是什么，都会稳定地发现相似的早期层特征，这表明这些模式反映的是视觉数据中的基本结构，而非特定任务的伪影。后续层的专门化程度更加明显，但在针对相关任务训练的网络之间，仍然能看到可辨识的对应关系。

卷积网络的成功对表示学习具有深远影响。它们能够自动发现有效特征，而非依赖手工设计的变换，这表明通过受约束的优化，结构如何从数据中涌现，背后存在更深层的原理。这一主题——精心设计的架构能够学习到自然的表示——贯穿于现代机器学习之中，从语言模型到蛋白质结构预测皆是如此。

然而，挑战仍然存在。正是那些促成高效学习的架构约束，有时反而会成为限制：

- 翻译不变性可能是不可取的
- 长程依赖被忽略
- 空间结构可能与数据不匹配

像 Transformer 这样的现代变体通过学习到的注意力机制解决了这些局限性，但仍然保留了一个核心洞见：架构约束能够实现高效学习。

卷积网络背后的数学原理远不止于计算机视觉。它们的根本洞见——局部结构至关重要，且应在架构层面加以编码——启发了面向众多领域的专用网络：

- 用于分子建模的图卷积
- 气候数据的球面卷积
- 用于时间序列的时序卷积

每一种方法都在尊重特定领域结构的同时对基础数学进行调整，并保持这样一个核心原则：架构约束能够实现高效学习。

贝佛nd 它们的实际影响，卷积网络体现了 h 噢 类-

经典数学通过精心的工程转化为现代机器学习。贯穿全文所研究的线性变换奠定了基础，而经过深思熟虑的架构约束使对自然表示的高效学习成为可能。随着人工智能持续迅猛发展，这一原则——结构促成学习——很可能将愈发显得重要。

Large Language Models & Algebraic Reasoning

线性代数融入智能的终极综合，最深刻地体现在庞大神经网络如何处理语言之中。当人类阅读文本时，我们感知到的并非仅是词语的序列，而是由语法、语境与知识交织而成的复杂意义之网。这种深层结构——语言信息的自然几何——指导着现代语言模型的设计。通过对注意力机制与学习到的表示进行精心组合，这些系统将本文中发展出的数学转化为表面上的理解，达成了在短短数年前看起来仍如奇迹般的能力。

首先考虑语言如何编码其意义。每个词或子词标记都会映射到高维空间中的一个学习得到的向量——$\mathbb{R}^d$中的一个点，其中通常$d \approx$为1024或更大。然而，将这些嵌入视为任意向量会丢弃语言的深层结构。在处理短语“the cat sat on the mat”时，语言模型必须以某种方式捕捉的不仅是单个词的含义，还有它们的语法关系、语义角色以及更广泛的上下文。我们试图建模的模式——区分疑问句与陈述句或检测细微含义的特征——源自这些表示之间复杂的相互作用。

这种关系结构通过本章前面研究的注意力机制转化为数学形式。不同于应用固定的变换，每一层都会在所有成对的词元之间计算动态关系：

$$A_{ij} = \frac{\boldsymbol{q}_i^T \boldsymbol{k}_j}{\sqrt{d}} \quad \text{and} \quad \tilde{A}_{ij} = \frac{\exp(A_{ij})}{\sum_k \exp(A_{ik})}$$

由此得到的注意力权重$\tilde{A}$实现了学习到的等价关系，有效地构建了按照其上下文角色对标记进行分组的商空间。正如第 3 章中首次遇到的基本分解一样，这些动态商空间组织了网络中的信息流动。

这种方法的力量源自其将数学上的优雅与实践中的有效性相融合。正如前面章节中的商空间通过识别等价元素揭示结构一样，注意力机制通过学习哪些标记应当相互作用来发现上下文模式。早期层通常捕捉表层属性：

- 基本句法关系
- 局部词汇关联
- 简单的指称模式

这些在后续层中结合，以表示越来越抽象的关系，每个关系都是从简单模式的学习组合中生成的。

现代语言模型的巨大规模——拥有数千亿个参数——揭示了超出较小系统所能显现的原理。就像巨大的晶体，其微观结构产生了涌现的性质，这些网络发展出似乎超越其基本组成部分的能力：

Historical Note: 术语“注意力”来源于神经机器翻译，在那里这些机制首次展示了它们的强大。它们的真正潜力直到2017年随着变换器架构的出现才显现出来，直接推动了现代语言模型的发展。

- 从文本示例中进行少量学习
- 零-shot转移到新任务
- 明显的推理和推断

然而，它们的基础仍然是贯穿本文所研究的矩阵运算——注意力通过内积计算相关性分数，前馈层实现学习到的变换，层归一化稳定梯度。

这种由规模带来的智能涌现，揭示了关于表征与学习的深刻问题。对每个词元进行处理的向量和矩阵，不仅编码词义，还编码知识片段、推理模式以及隐含技能。正如第10章中从矩阵结构中自然涌现的奇异向量一样，这些学得的表征捕捉了语言中的基本模式——只是如今它们是通过优化而非分析被发现的。

考虑语言模型如何处理提示“完成模式：2, 4, 6, ...”。每个标记的嵌入通过注意力层和前馈计算层进行转换：

$$\boldsymbol{h}_\ell = \text{FFN}_\ell(\text{Attention}_\ell(\boldsymbol{h}_{\ell-1})) + \boldsymbol{h}_{\ell-1}$$

残差连接让人联想到第 3 章中研究的拆分方式，使信息能够直接流动，而注意力机制则提取相关模式。不知为何，正是从这些数学运算中涌现出了识别并延续序列的能力——这一看似简单的能力，实际上需要对语言和数字概念进行高度复杂的处理。

然而，挑战依然存在。正是那些促成这种涌现的架构特征，也同时施加了限制：

- 注意力随序列长度呈二次增长
- 模型可能会生成看似合理但不正确的回答
- 他们能力的基础仍然有些不透明

这些局限性提醒我们，尽管当前的系统取得了显著的成果，但它们只是沿途的一步，而不是终点。

支撑语言模型的数学原理远远超出了文本处理的范畴。它们的根本性洞见——注意力机制使信息得以动态流动，规模化促成涌现，精心的架构设计引导学习——已在各个领域激发了应用：

- 蛋白质结构预测
- 科学发现
- 代码生成与分析

每种方法都将基础数学适应于新的情境，同时保持着结构化计算能够实现复杂行为的核心原则。

除了其实用影响之外，大型语言模型还展示了经典数学如何通过精心的工程化转化为现代人工智能。贯穿本文所研究的线性变换提供了基础，而深思熟虑的架构选择使对自然表征的高效学习成为可能。随着这些系统持续快速推进，这一原则——结构促成智能——很可能将变得愈发重要。

从基础线性代数到语言模型的旅程，呼应了我们文本的更广泛主题：永恒的数学原理如何通过精心的发展转化为实用工具。这些系统从计算这一原始材料中锻造出理解。它们的成功并不意味着传统数学被取代，而是表明它通过我们所创造的结构获得了新的表达。在这一视角下，大型语言模型并非对经典原则的断裂，而是其自然演进——通过有原则的规模化组合，将线性代数运算转化为表观智能。

□ ——————————————————————— □

Exercises: Chapter 13

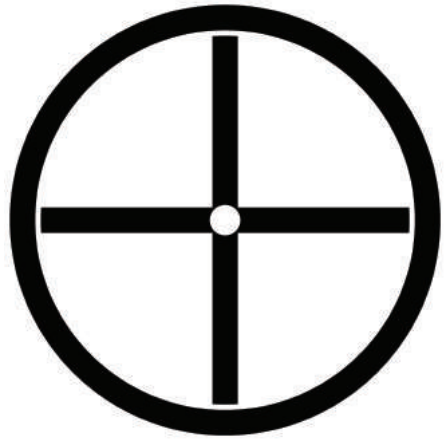
1. 设 $W \in \mathbb{R}^{m\times n}$ 和 $\boldsymbol{b} \in \mathbb{R}^m$ 定义一个神经网络层，计算 $\boldsymbol{h} = \max\{0, W\boldsymbol{x} + \boldsymbol{b}\}$。证明在没有 ReLU 激活 $\max\{0, \cdot\}$ 的情况下，这将化简为第 3 章中的变换。然后当 $n = m =$ 为 2 时，明确描述该变换在线性情况下输入空间的各个区域，并用示意图加以说明。

2. 对于 softmax 函数 $\sigma_i(\boldsymbol{x}) = e^{x_i} / \sum_{j=1}^{n} e^{x_j}$：证明对任意输入 $\boldsymbol{x}$，输出之和为 1；推导导数矩阵的各元素 $\partial\sigma_i/\partial x_j = \sigma_i(\delta_{ij} - \sigma_j)$；并解释为什么对所有输入加上同一个常数不会改变输出。
3. 考虑层归一化 $\hat{\boldsymbol{h}} = \gamma(\boldsymbol{h} - \mu)/\sqrt{\sigma^2 + \epsilon} + \beta$，其中 $\overline{\mu、\ \sigma^2}$ 为激活的均值/方差，γ、β 为学习到的参数。说明输出具有均值 β 和方差 γ^2，并将其与第1章中的列缩放联系起来。这如何有助于控制权重矩阵的条件数?

4. 对于一个将词汇映射到d维向量的词嵌入矩阵$W \in \mathbb{R}^{d\times v}$：证明其行空间的维数至多为$d$；将这一点与第12章的低秩近似联系起来；并解释相似语境对W中对应行意味着什么。

5. 残差连接 $\boldsymbol{h}_{\text{out}} = \boldsymbol{h}_{\text{in}} + F(\boldsymbol{h}_{\text{in}})$ 将输入直接加到变换后的输出上。用 F 的雅可比矩阵表示 $\partial\boldsymbol{h}_{\text{out}}/\partial\boldsymbol{h}_{\text{in}}$，并解释这如何有助于深层网络训练，将其与第 4 章的条件性概念联系起来。 6. 设 $E : \mathbb{R}^n \to \mathbb{R}^k$ 和 $D : \mathbb{R}^k \to \mathbb{R}^n$ 为具有 $k < n$ 的自编码器的编码/解码映射。识别 E 的四个基本子空间，解释 $\ker(E)$ 如何与被丢弃的信息相关，以及 $\operatorname{coker}(E)$ 如何与重构误差相关。将其与第 3 章中发展的理论联系起来。 7. 对于注意力权重 $A_{ij} = \exp(\boldsymbol{q}_i^T \boldsymbol{k}_j/\sqrt{d})/\sum_l \exp(\boldsymbol{q}_i^T \boldsymbol{k}_l/\sqrt{d})$：证明每一行之和为 1；分析当 $d \to 0$ 和 $d \to \infty$ 时的行为；并将其与第 9 章中的随机矩阵联系起来。 8. 多头注意力通过 $\text{MultiHead}(Q, K, V)$ 将 h 个并行计算组合在一起 =

$W_O[\text{head}_1; \ldots; \text{head}_h]$。说明当 $h = 1$ 时这会化简为标准注意力，并解释为什么多头可能更好地捕捉不同类型的关系。将其与 SVD 的奇异向量如何捕捉不同模式联系起来。 9. 考虑一个神经网络，应用 L 层 $\boldsymbol{h}_{\ell+1} = \sigma(W_\ell \boldsymbol{h}_\ell + \boldsymbol{b}_\ell)$。证明在没有激活函数的情况下，这会化简为矩阵乘法；解释非线性如何使其能够逼近弯曲流形；并将其与第 13.6 节中的表征学习联系起来。 10.（挑战）对于一个使用位置编码 $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n\}$ 且序列为 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ 的 Transformer：证明注意力可以通过内积检测相对位置；说明正弦编码如何促进这一点；并将其与第 5 章的正交性概念联系起来。 11.（挑战）设 f 为一个参数数量多于训练点 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$ 的神经网络，通过最小化均方误差进行训练。通过分析相关雅可比矩阵的秩，证明它可以实现零训练误差。将其与第 13.1 节的通用逼近定理联系起来。



& these again surrounded by
Four Wonders of the Almighty Incomprehensible
Pervading all amidst & round about
Fourfold each in the other reflected
They are named Life's in Eternity
Four Starry Universes going forward
From Eternity to Eternity

关于作者

罗伯特·格里斯特（博士，康奈尔大学，应用数学，1995年）是宾夕法尼亚大学数学与电气与系统工程系的安德里亚·米切尔PIK教授。他是应用代数拓扑领域的公认领导者，研究方向包括网络、机器人学、信号处理、数据分析、优化等。他是一位获奖的研究者、教师和数学及其应用的解说者，目前担任宾夕法尼亚大学工程与应用科学学院本科教育副院长。

他是几本书的作者，如：*Elementary Applied Topology* 和 *Calculus Blue Guide*；也是 YouTube 视频系列的创作者，包括 *Calculus BLUE, Calculus GREEN, & Applied Dynamical Systems*。

在他的空闲时间，他以*colimit*的名字发布数学艺术和动画。