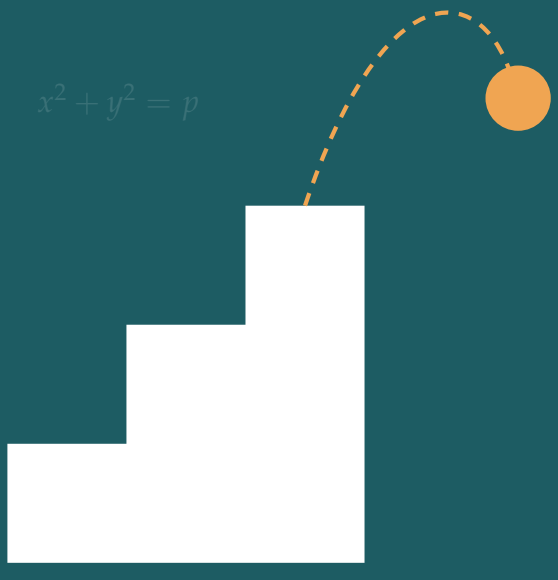
STEP 备试 · 大学数学

Beyond STEP

第 7 章

线性代数与几何



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

第 7 章

线性代数与几何

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

第 7 章：线性代数与几何

本册是 *Beyond STEP* 全书 13 章中的第 7 章。全书、各章的单行本、每一节的可列印工作纸，以及每一道 STEP 试题的详细解答，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

2026 年 11 月首次于网上发表。

本 PDF 内凡属作者本人创作的内容，包括正文、解答、插图、练习及答案，均按照 Creative Commons CC0 1.0 Universal 授权条款贡献予公有领域：你可以为任何目的复制、改编及分享，毋须征求许可。详见 `creativecommons.org/publicdomain/zero/1.0`。

本书讨论的 STEP 试题，版权属 Cambridge University Press & Assessment 所有。本书并无转载这些试题：每一题均附连结，通往官方来源的题目，即 STEP Support Programme 的试题资料库或 OCR 的历届试卷。

如有错漏，欢迎到 `github.com/keithmcnulty/step_book_mcnulty` 提交 issue 指正。

章

7

线性代数与几何

矩阵不是一张数字表。它是空间的一种变换，而且自有其生命。

本章内容

7.1	对矩阵不成立的法则	STEP II 1993
7.2	只伸缩，不转向	STEP III 1990
7.3	甚么才算行列式	STEP II 2019
7.4	回到起点的矩阵	STEP II 2020
7.5	三维空间中的旋转	STEP III 2022
7.6	双曲旋转与相对论	STEP III 1991
7.7	永恒不变的量	STEP III 2024
7.8	四个平方数	STEP III 1994
7.9	只有五种柏拉图立体	STEP III 1999
A	练习答案	第 63 页

空间的变换

我们思考时不依赖基底，写作时也不依赖基底，但到了紧要关头，我们就关上办公室的门，拼命地用矩阵计算。

— Irving Kaplansky，谈他与 Paul Halmos 处理线性代数的方式

在中学，矩阵（matrix）通常以解联立方程的工具出现，或者是变换点的一套步骤：把矩阵乘以一个列向量（column vector），然后读出新的坐标。两者都有用，但它们掩盖了矩阵之所以成为数学中最重要概念之一的原因。矩阵是一个*线性变换*（linear transformation）：一种把空间拉伸、旋转、反射和错切的方式，而且矩阵可以相加、相乘、取幂和求导，就像数一样。只不过它们又不像数。矩阵的乘法取决于次序，两个矩阵可以相乘得零，而且它们有隐藏的不变量（invariant）——迹（trace）、行列式（determinant）、特征值（eigenvalue）——无论矩阵写成甚么样子，这些量都能刻划它的本质。

线性代数为何重要

线性代数（linear algebra）大概是应用最广的数学分支。每一个电脑图像引擎，由电子游戏到电影动画，都是一部做矩阵乘法的机器：角色的每一次旋转、每一个镜头角度、每一个透视投影，都是一个矩阵。互联网上的搜索引擎和推荐系统建基于特征向量（eigenvector）。工程师利用特征值求出桥梁和建筑物的振动频率；统计学家利用它们找出数据中的主要规律；量子力学完全以线性代数的语言写成。而现代的机器学习，核心就是数量极其庞大的矩阵乘法。

线性代数无处不在，是有原因的。线性问题是我们能够彻底解决的问题，而几乎每一个非线性问题，只要放大到足够近，看起来都是线性的。这正是微积分的想法，也意味着几乎所有事物的局部性态，都是用线性代数这种语言来描述的。

你在大学会遇到甚么

线性代数与分析（analysis）并列为大学一年级的两大支柱。你会遇到*向量空间*（vector space，即可以做加法和纯量乘法的集合，平面和三维空间只是最初的例子）以及它们之间的*线性映射*（linear map），并且学到矩阵其实就是用坐标写出来的线性映射。你会学习行列式、特征值和特征向量，以及*谱定理*（spectral theorem），它指出对称矩阵（symmetric matrix）可以藉一个旋转来对角化（diagonalise）。往后的课程处处用到这些内容：在微分方程（第 8 章）中，在几何中（旋转和反射在那里构成群），在描述连续对称的*李群*（Lie

group）理论中，以及在泛函分析（functional analysis）中，也就是无限多维的线性代数。

贯穿本章的问题

三大问题

1. **矩阵在哪些方面像数，又在哪些方面不同？**哪些算术法则仍然成立，哪些不再成立？
2. **矩阵究竟做了甚么？**我们怎样才能看出一个变换的几何，又有哪些数能够不依赖坐标而刻划它？
3. **甚么保持不变？**哪些量在变换之下，或在运动之下保持不变，它们又告诉我们甚么？

本章由矩阵代数的法则开始（Section 7.1），然后转到变换的几何：特征向量、行列式、有限阶矩阵、三维空间中的旋转，以及狭义相对论中的双曲旋转（hyperbolic rotation）。接着是随时间演化的矩阵的守恒量（conserved quantity）、由行列式得出的四平方恒等式（four-square identity），最后以一个计数论证作结，即欧拉（Euler）公式，它把柏拉图立体（Platonic solids）分类。

本章的概念

§	主题	概念
7.1	不成立的法则	非交换性（non-commutativity）；零因子（zero divisor）
7.2	特征向量	变换不会转动的方向
7.3	行列式	由可乘性刻划；有向面积
7.4	有限阶矩阵	迹、行列式与凯莱–哈密顿（Cayley–Hamilton）定理
7.5	三维旋转	由叉积（cross product）得出罗德里格斯（Rodrigues）公式
7.6	洛伦兹推进（Lorentz boosts）	双曲角可以相加
7.7	守恒量	交换子（commutator）的迹为零
7.8	四平方	行列式的可乘性
7.9	柏拉图立体	算两次（double counting）与欧拉公式

如何阅读本章

对本章的每一个矩阵，都把它所做的事画出来。取单位正方形，或单位圆，或几个向量，然后画出它们的像的草图。花两分钟画一幅草图，往往在你动手计算之前就让你看到答案（一个特征向量、一个反例、一条不变直线），事后也有助你检查计算。

7.1 对矩阵不成立的法则

题目

STEP II 1993，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/93-s2-q6`

引子

矩阵（matrix）可以相加，也可以相乘，所以我们很容易把它们当作数来看待。这道题正是要考验这种想法：它列出四条对数成立的“法则”，问哪些对矩阵仍然成立。结果大部分都不成立，而它们不成立的方式很有启发性。矩阵乘法不满足交换律（not commutative）；有些非零矩阵的乘积是零；而一个矩阵的二次方程可以有无穷多个解。

这些“失效”并不是缺陷，反而正是矩阵有用的原因：矩阵描述的是空间的变换（transformation），而以不同的次序进行两个变换，确实会得到不同的结果。1925 年，海森堡（Werner Heisenberg）发现量子力学中的物理量的表现像矩阵而不像数，而 $\mathbf{AB} = \mathbf{BA}$ 不成立，正是不确定性原理（uncertainty principle）的数学形式。

动笔之前

想一想 对实数而言，为甚么 $ab = 0$ 可推出 $a = 0$ 或 $b = 0$？这个论证用到了数的哪一项性质？这项性质对矩阵成立吗？

若 $ab = 0$ 而 $a \neq 0$，两边除以 a 便得 $b = 0$。这个论证需要每个非零的数都有逆（inverse）。非零矩阵却不一定有逆（行列式（determinant）为零的矩阵就没有），所以这个论证行不通。同时也要记住几何上的图像：矩阵是平面的一个变换，而乘积 $\mathbf{AB}$ 的意思是“先做 $\mathbf{B}$，再做 $\mathbf{A}$”。

逐步解题

阶段 1　能否有 $\mathbf{AB}=0$ 但 $\mathbf{BA}\neq 0$?

想一想　从几何角度去想。你希望 $\mathbf{B}$ 把所有东西送到一条直线上，然后 $\mathbf{A}$ 把这条直线压成零。但次序调转时，$\mathbf{A}$ 不应把所有东西都压到会被 $\mathbf{B}$ 消去的地方。

设 $\mathbf{A}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$，它是到 x 轴上的投影（projection），把 y 轴消去；又设 $\mathbf{B}=\begin{pmatrix}0&0\\1&0\end{pmatrix}$，它把 (x,y) 送到 $(0,x)$，即送到 y 轴上。于是

$$\mathbf{AB}=\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix},\qquad \mathbf{BA}=\begin{pmatrix}0&0\\1&0\end{pmatrix}\neq\mathbf{0}.$$

这个断言是**错**的（图 7.1.1）。 ■

阶段 2　平方差

小心展开，保持每个乘积的次序：

$$(\mathbf{A}-\mathbf{B})(\mathbf{A}+\mathbf{B})=\mathbf{A}^2+\mathbf{AB}-\mathbf{BA}-\mathbf{B}^2.$$

这个式子等于 $\mathbf{A}^2-\mathbf{B}^2$，当且仅当 $\mathbf{AB}=\mathbf{BA}$。取第 1 阶段的矩阵，$\mathbf{AB}-\mathbf{BA}=-\mathbf{BA}\neq\mathbf{0}$，所以这个断言是**错**的。 ■

阶段 3　$\mathbf{AX}=0$ 的非零解

想一想　若 $\det\mathbf{A}\neq 0$，你可以对 $\mathbf{AX}=\mathbf{0}$ 做甚么？若 $\det\mathbf{A}=0$，你能否找到一个非零的列向量（column vector）$\mathbf{v}$，使得 $\mathbf{Av}=\mathbf{0}$?

若 $\det\mathbf{A}\neq 0$，则 $\mathbf{A}^{-1}$ 存在，而 $\mathbf{X}=\mathbf{A}^{-1}\mathbf{AX}=\mathbf{0}$：没有非零解。

若 $\det\mathbf{A}=0$，写 $\mathbf{A}=\begin{pmatrix}a&b\\c&d\end{pmatrix}$，其中 $ad=bc$。向量 $\begin{pmatrix}b\\-a\end{pmatrix}$ 和 $\begin{pmatrix}d\\-c\end{pmatrix}$ 满足 $\mathbf{A}\begin{pmatrix}b\\-a\end{pmatrix}=\begin{pmatrix}ab-ba\\cb-da\end{pmatrix}=\mathbf{0}$，另一个向量亦同理。由于 $\mathbf{A}\neq\mathbf{0}$，两者之中至少有一个非零；把它记作 $\mathbf{v}$。那么，两列都是 $\mathbf{v}$ 的矩阵 $\mathbf{X}$ 是非零的，并且 $\mathbf{AX}=\mathbf{0}$。所以这个断言是**对**的。 ■

阶段 4　有多少个平方根?

停一停，试一试 找出非零的 $\mathbf{A}$ 和 $\mathbf{B}$，使得 $\mathbf{X}^2+\mathbf{AX}+\mathbf{B}=\mathbf{0}$ 有无穷多个解。（提示：把左边变成完全平方 $(\mathbf{X}-\mathbf{I})^2$，再想想平方为零的矩阵。）

取 $\mathbf{A}=-2\mathbf{I}$ 和 $\mathbf{B}=\mathbf{I}$，两者都非零。方程变成 $(\mathbf{X}-\mathbf{I})^2=\mathbf{0}$。任何满足 $\mathbf{N}^2=\mathbf{0}$ 的矩阵 $\mathbf{N}$ 都给出一个解 $\mathbf{X}=\mathbf{I}+\mathbf{N}$，而这样的 $\mathbf{N}$ 有无穷多个：

$$\mathbf{N}=\begin{pmatrix}0 & t\\ 0 & 0\end{pmatrix},\qquad \mathbf{N}^2=\mathbf{0}\ \text{对任意实数}t.$$

所以对每一个 t，$\mathbf{X}=\begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix}$ 都是解，这个断言是**错**的。 ■

回顾

- **次序很重要。**每次展开都必须保持乘积的次序；$\mathbf{AB}$ 和 $\mathbf{BA}$ 通常并不相同。
- **零因子（zero divisors）。**非零矩阵相乘可以得零，这就破坏了“二次方程最多有两个根”背后的论证。
- **从变换的角度去想。**只要想像矩阵对平面做了甚么，每个反例都很容易找到。

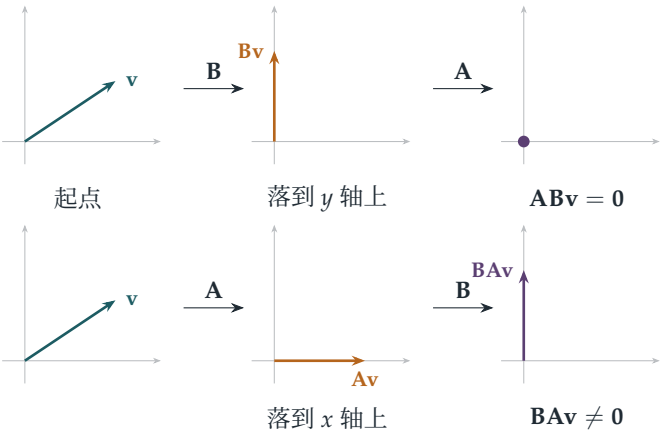


图 7.1.1 为甚么 $\mathbf{AB}=\mathbf{0}$ 但 $\mathbf{BA}\neq\mathbf{0}$。上：$\mathbf{B}$ 把每个向量送到 y 轴上，然后 $\mathbf{A}$（到 x 轴上的投影）把它压成零。下：次序调转时，$\mathbf{A}$ 保留 x 分量，$\mathbf{B}$ 再把它移到 y 轴上，它在那里得以保存下来。

背后的数学

矩阵作为一种代数

凯莱（Arthur Cayley）在 1858 年引入矩阵代数，把矩阵看成可以相加和相乘的单一对象，而不仅仅是系数表。他立刻注意到乘法不满足交换律，并证明了*凯莱—哈密顿定理*（*Cayley–Hamilton theorem*）：每个方阵都满足它自己的特征方程（characteristic equation）。对于迹（trace）为 t、行列式为 d 的 2×2 矩阵，这就是说 $\mathbf{M}^2 - t\mathbf{M} + d\mathbf{I} = \mathbf{0}$，本章会反复用到这个事实（Section 7.4）。2×2 实矩阵构成一个*环*（*ring*）：可以加、减、乘，但乘法不满足交换律，而且存在*零因子*，即乘积为零的非零元素。

无穷多个平方根

第 4 阶段的结论还可以变得更加惊人。方程 $\mathbf{X}^2 = \mathbf{I}$ 有明显的解 $\pm\mathbf{I}$，但平面上对任何一条通过原点的直线的反射（reflection）$\mathbf{X} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ 也都是解，因为反射两次等于甚么也没做。这就是 $\mathbf{I}$ 的一整个圆那么多的平方根。系数为数的 n 次多项式最多有 n 个根，是因为数没有零因子；矩阵却有，于是这个根的数目上限便不再成立。

物理学中的非交换性

在经典力学中，位置 x 和动量 p 都是数，并且 $xp = px$。在量子力学中，它们变成算子（operator，即无穷矩阵），满足 $xp - px = \mathrm{i}\hbar$，其中 $\hbar$ 是普朗克（Planck）常数除以 2π。单凭这一条非交换关系，就可推出海森堡的不确定性原理：位置和动量不可能同时被准确得知。海森堡在 1925 年发现量子力学的法则时，并不知道矩阵是甚么；是玻恩（Max Born）认出海森堡所发明的那条奇怪的乘法法则就是矩阵乘法。

思维技巧

在新的情境中检验熟悉的法则

遇到一种新的对象时，要检查哪些熟悉的法则仍然成立，并弄清楚每条法则究竟依赖哪一项性质。

法则“$ab = 0$ 可推出 $a = 0$ 或 $b = 0$”依赖除法；“$(a-b)(a+b) = a^2 - b^2$”依赖交换律。知道每条法则建基于甚么，你就知道甚么时候仍然可以使用它。

延伸练习

1 **甚么时候可交换？** 求所有与 $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ 可交换的 2×2 矩阵。

2 **零的平方根。** 求所有满足 $\mathbf{N}^2 = \mathbf{0}$ 的 2×2 实矩阵 $\mathbf{N}$。（提示：利用凯莱–哈密顿定理证明 $\operatorname{tr}\mathbf{N} = \det\mathbf{N} = 0$。）

3 **反射。** 证明每个反射矩阵 $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ 的平方都是 $\mathbf{I}$，并求出它所对应的反射轴。

4 **凯莱–哈密顿定理。** 以直接计算证明每个 2×2 矩阵都满足 $\mathbf{M}^2 - (\operatorname{tr}\mathbf{M})\mathbf{M} + (\det\mathbf{M})\mathbf{I} = \mathbf{0}$。

延伸阅读

- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge）。最受欢迎的线性代数入门教科书，着重理解。
- **Sheldon Axler, *Linear Algebra Done Right***（Springer；可在网上免费阅读）。取向较为抽象，以线性映射而非矩阵为中心。
- **Jim Baggott, *The Quantum Story***（Oxford University Press）。量子力学的历史，包括海森堡的矩阵力学。

7.2 只伸缩，不转向

题目

STEP III 1990，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/90-s3-q6`

引子

矩阵变换（matrix transformation）通常会把每个向量转到新的方向。不过，许多矩阵都有几个特别的方向，它完全不会转动这些方向，只会沿着它们拉伸或压缩。这些方向上的向量就是矩阵的*特征向量*（eigenvector，源自德文 *eigen*，意思是“自身的”或“特有的”），而拉伸的倍数就是它的*特征值*（eigenvalue）。当特征向量的数目足够时（像本题这类对称矩阵就必定足够），这个变换便一目了然：在合适的坐标下，它只不过是沿两条轴的拉伸。

本题要找出某个矩阵的特征向量，再用它们来理解这个矩阵对一个圆有甚么作用。特征向量是整个应用数学中最有用的概念之一。它们描述桥梁的固有振动、旋转物体的主轴、原子的能级、一团数据中最重要的方向，以及 Google 最初的 PageRank 算法怎样为网页排名。

动笔之前

? **想一想** T 使一个向量“方向不变，只乘以一个纯量”，这是甚么意思？试把它写成一条等式。

意思是存在某个数 λ 使得 $T\mathbf{v} = \lambda\mathbf{v}$：$\mathbf{v}$ 是特征值为 λ 的特征向量。留意题目中的两个向量 $(1, 2)$ 和 $(2, -1)$ 互相垂直。这并非巧合：这个矩阵是*对称*（symmetric）的，而对称矩阵总有互相垂直的特征向量。

逐步解题

阶段 1　特征向量

$$T\begin{pmatrix}1\\2\end{pmatrix}=\frac{2}{5}\begin{pmatrix}9-4\\-2+12\end{pmatrix}=\frac{2}{5}\begin{pmatrix}5\\10\end{pmatrix}=2\begin{pmatrix}1\\2\end{pmatrix},\qquad T\begin{pmatrix}2\\-1\end{pmatrix}=\frac{2}{5}\begin{pmatrix}20\\-10\end{pmatrix}=4\begin{pmatrix}2\\-1\end{pmatrix}.$$

所以 $(1,2)$ 被拉伸至 2 倍，而 $(2,-1)$ 被拉伸至 4 倍。 ■

阶段 2　圆的像

> **想一想**　要求出 E 的方程，可先用 X 和 Y 表示 x 和 y（即求矩阵的逆），再代入 $x^2+y^2=1$。

矩阵 $\frac{2}{5}\begin{pmatrix}9&-2\\-2&6\end{pmatrix}$ 的行列式（determinant）是 $\frac{4}{25}(54-4)=8$，它的逆矩阵（inverse）是 $\frac{1}{20}\begin{pmatrix}6&2\\2&9\end{pmatrix}$。所以 $x=\frac{6X+2Y}{20}$ 及 $y=\frac{2X+9Y}{20}$，而 $x^2+y^2=1$ 变成

$$(6X+2Y)^2+(2X+9Y)^2=400\quad\Longrightarrow\quad 40X^2+60XY+85Y^2=400,$$

两边除以 5，便得 $8X^2+12XY+17Y^2=80$。 ■

> **想一想**　E 是甚么形状？它的面积是多少？试利用特征向量作答。

单位圆在每个方向上的半径都是 1。在 $(2,-1)$ 的方向上，它被拉伸至 4 倍；在与之垂直的 $(1,2)$ 方向上，则被拉伸至 2 倍。所以 E 是一个椭圆（ellipse），半轴长为 4 和 2，分别沿着两个特征向量的方向（图 7.2.1），其面积为 $\pi\times4\times2=8\pi$。（换个说法：任何线性映射（linear map）都会把面积乘以 $|\det T|=8$，而圆的面积是 π。）

阶段 3　X 的最大值

> **想一想**　在 C 上，$X=\frac{2}{5}(9x-2y)$。当 $x^2+y^2=1$ 时，$9x-2y$ 最大可以是多少？

$9x-2y$ 是 $(9,-2)$ 与单位向量 (x,y) 的点积（dot product），所以它的最大值是 $|(9,-2)|=\sqrt{85}$，在 (x,y) 指向 $(9,-2)$ 的方向时取得。因此 X 的最大值是 $\frac{2}{5}\sqrt{85}=2\sqrt{\frac{85}{25}}=2\sqrt{\frac{17}{5}}$。 ■

阶段 4　切线

停一停，试一试 C 上的点 $\frac{1}{5}(3,4)$ 的切线（tangent）是 $3x+4y=5$。线性映射会把切线变成切线。试利用这一点求出 E 的切线。

把 $x=\frac{6X+2Y}{20}$、$y=\frac{2X+9Y}{20}$ 代入 $3x+4y=5$：

$$\frac{3(6X+2Y)+4(2X+9Y)}{20}=5 \quad\Longrightarrow\quad 26X+42Y=100,$$

即 $13X+21Y=50$。 ■

（验算：该点本身的像是 $T(\frac{3}{5},\frac{4}{5})=(\frac{38}{25},\frac{36}{25})$，而 $13\cdot\frac{38}{25}+21\cdot\frac{36}{25}=\frac{494+756}{25}=50$。）

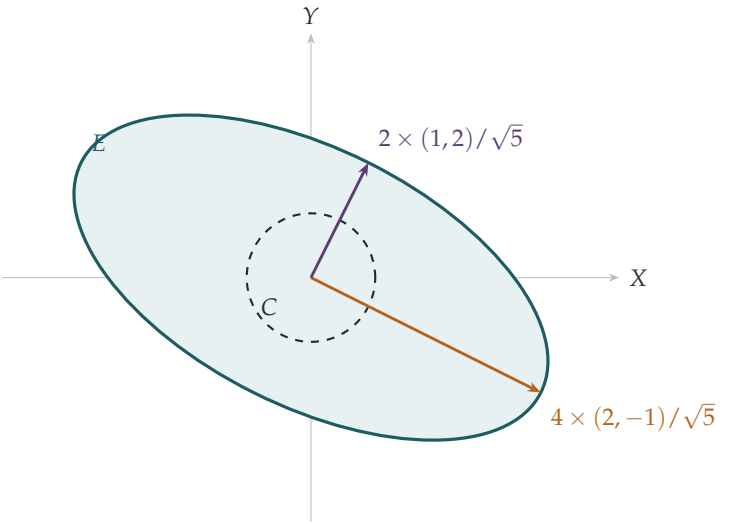


图 7.2.1 单位圆 C（虚线）及它在 T 之下的像 E。沿着特征向量 $(2,-1)$，圆被拉伸至 4 倍；沿着与之垂直的特征向量 $(1,2)$，则被拉伸至 2 倍。所以 E 是一个椭圆，半轴长为 4 和 2，分别沿着这两个方向，面积为 8π。

回顾

- **特征向量是天然的坐标轴。**在特征向量所给出的坐标下，T 只是分别拉伸至 4 倍和 2 倍，它对圆的作用显而易见。
- **行列式量度面积。**E 的面积是 C 的面积的 $|\det T|$ 倍。
- **线性映射保持相切关系。**把圆的切线加以变换，几乎不用计算便得到椭圆的切线。

背后的数学

谱定理

每个实对称矩阵都有一整组互相垂直的特征向量，而且特征值都是实数。这就是*谱定理*（spectral theorem），是大学第一门线性代数（linear algebra）课程的核心结果之一。它告诉我们，每个对称的线性映射，在合适的垂直坐标下，都只是沿各条轴的拉伸。换个说法，每个像 $8X^2 + 12XY + 17Y^2$ 这样的二次型（quadratic form），都可以经旋转化成 $\lambda_1 u^2 + \lambda_2 v^2$ 的形式，所以每条有心圆锥曲线（central conic）都是轴互相垂直的椭圆或双曲线（*主轴定理*，principal axis theorem）。这个想法可追溯至 1820 年代的柯西（Cauchy），他用它来求二次曲面的主轴，以及固体内应力的主轴。

这一切之所以成立，全靠对称性。不对称的矩阵就未必这样合作：旋转 90° 会转动每一个方向，所以它完全没有实特征向量；而错切（shear）$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ 只有一个特征向量方向，所以它不是沿任何一对轴的拉伸。

无处不在的特征向量

以下是特征向量出现的其中几个地方。

- **振动。**由质点和弹簧组成的系统会以*简正模态*（normal mode）振动，即各部分以同一频率同步运动的模式。这些模态是某个矩阵的特征向量，而频率则来自特征值（Section 11.2）。
- **数据。**一团数据点分布得最散的方向，就是其协方差矩阵（covariance matrix）的特征向量。找出这些方向的方法称为*主成分分析*（principal component analysis），是统计学和机器学习的基本工具之一（Section 13.7）。
- **量子力学。**量度一个物理量可能得到的结果，是某个矩阵（或算子）的特征值，而原子的能级也是特征值。
- **网上搜索。**Google 的 PageRank 是这个搜索引擎背后的原始构想：有一个描述网页之间连结的巨型矩阵，PageRank 就按它那个特征值为 1 的特征向量的各个分量来为网页排名。

思维技巧

选择配合变换的坐标

要理解一个线性变换，就找出它不会转动的方向。在这些坐标下，它会变得简单。

这与 Section 5.6 中选费马点（Fermat point）作原点，或 Section 2.8 中以变量代换来简化映射，是同一个想法：选对了坐标，结构便一目了然。

延伸练习

1 **求特征值。**证明 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 的特征值是 $\lambda^2 - (a+d)\lambda + (ad-bc) = 0$ 的根，并验证题目中的矩阵给出 $\lambda = 2$ 和 4。

2 **互相垂直的特征向量。**证明若 $\mathbf{M}$ 是对称矩阵，且 $\mathbf{Mu} = \lambda\mathbf{u}$、$\mathbf{Mv} = \mu\mathbf{v}$，其中 $\lambda \neq \mu$，则 $\mathbf{u}\cdot\mathbf{v} = 0$。（提示：用两种方法计算 $\mathbf{v}\cdot\mathbf{Mu}$。）

3 **旋转。**证明除非 θ 是 π 的倍数，否则旋转矩阵 $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ 没有实特征向量。它的复特征值是甚么？

4 **幂。**利用特征向量，求 $T^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}$。（提示：把 $(1,0)$ 写成 $(1,2)$ 和 $(2,-1)$ 的组合。）

延伸阅读

- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge）。第 6 章是特征值及其应用的出色入门。
- **Kurt Bryan & Tanya Leise, "The $25,000,000,000 eigenvector: the linear algebra behind Google"**，*SIAM Review* 48 (2006)。浅白易读，可免费取阅。
- **3Blue1Brown, *Essence of Linear Algebra***（影片系列，可在网上免费观看）。以精美的动画呈现的，正是本节的几何概念。

7.3 甚么才算行列式

题目

STEP II 2019，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

`www.ocr.org.uk/Images/696590-2019-past-papers.zip`

引子

2×2 矩阵（matrix）的行列式（determinant）是 $ad - bc$。这条公式从何而来？它为甚么如此重要？一个答案是：它是矩阵把面积放大或缩小的倍数。另一个答案，也就是本题要探讨的，是行列式基本上是矩阵上*唯一*能把矩阵乘法变成普通乘法的函数：$\det(\mathbf{MN}) = \det \mathbf{M} \det \mathbf{N}$。

题目给你一个未知函数 f，它具有这一个性质（另加两个用来排除平凡情况的宽松条件），然后要求你一步一步推导出它必然有怎样的表现。与 Section 6.5 的导子一样，这是一道单凭性质来确定一个对象的练习。到最后，f 被迫在两行对调时变号、在奇异矩阵（singular matrix）上取值为零、在错切（shear）上等于 1：这正是行列式的表现。

动笔之前

? **想一想** 题目中每个特殊矩阵各有甚么作用？对一般的 $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$，计算 **JM**、**MJ**、**KM** 和 **PM**。

JM 把 **M** 的两行对调；**MJ** 把它的两列对调。**KM** 把第二行乘以 k。**PM** 把第二行加到第一行上。这些就是高斯消元法（Gaussian elimination）中的*初等行运算*（elementary row operations）。由于 f 把乘积变成乘积，只要知道 f 在这些特殊矩阵上的值，就知道 f 对每种行运算有何反应。

逐步解题

阶段 1　单位矩阵

$\mathrm{f}(\mathbf{M}) = \mathrm{f}(\mathbf{MI}) = \mathrm{f}(\mathbf{M})\,\mathrm{f}(\mathbf{I})$。由于 $\mathrm{f}(\mathbf{M}) \neq 0$，两边相除即得 $\mathrm{f}(\mathbf{I}) = 1$。 ■

阶段 2　对调

$\mathbf{J}^2 = \mathbf{I}$，所以 $\mathrm{f}(\mathbf{J})^2 = \mathrm{f}(\mathbf{I}) = 1$，由此得 $\mathrm{f}(\mathbf{J}) = \pm 1$。由于 $\mathrm{f}(\mathbf{J}) \neq 1$，故 $\mathrm{f}(\mathbf{J}) = -1$。

现在 $\begin{pmatrix} c & d \\ a & b \end{pmatrix} = \mathbf{JM}$，所以它的值是 $\mathrm{f}(\mathbf{J})\mathrm{f}(\mathbf{M}) = -\mathrm{f}(\mathbf{M})$。而 $\begin{pmatrix} d & c \\ b & a \end{pmatrix} = \mathbf{JMJ}$（先对调两行，再对调两列），所以它的值是 $\mathrm{f}(\mathbf{J})^2\mathrm{f}(\mathbf{M}) = \mathrm{f}(\mathbf{M})$。 ■

阶段 3　线性相关的两行

> **想一想**　先考虑一个两行相等的矩阵 $\mathbf{B}$。$\mathbf{JB}$ 是甚么？这对 $\mathrm{f}(\mathbf{B})$ 说明了甚么？

若 $\mathbf{B}$ 的两行都是 (a, b)，把它们对调不会带来任何改变：$\mathbf{JB} = \mathbf{B}$。所以 $\mathrm{f}(\mathbf{B}) = \mathrm{f}(\mathbf{J})\mathrm{f}(\mathbf{B}) = -\mathrm{f}(\mathbf{B})$，这就迫使 $\mathrm{f}(\mathbf{B}) = 0$。

现在若 $\mathbf{A}$ 的两行是 (a, b) 和 $k(a, b)$，则 $\mathbf{A} = \mathbf{KB}$，其中 $\mathbf{B}$ 的两行都等于 (a, b)。所以 $\mathrm{f}(\mathbf{A}) = \mathrm{f}(\mathbf{K})\mathrm{f}(\mathbf{B}) = 0$。 ■（这包括 $k = 0$ 的情况，此时第二行为零。）

阶段 4　错切

> **停一停，试一试**　计算 $\mathbf{P}^2$ 和 $\mathbf{K}^{-1}\mathbf{PK}$。选取 k 使两者相等。由此可得 $\mathrm{f}(\mathbf{P})$ 是甚么？

$\mathbf{P}^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$，而当 $k \neq 0$ 时，

$$\mathbf{K}^{-1}\mathbf{PK} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{k} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}.$$

取 $k = 2$，则 $\mathbf{K}^{-1}\mathbf{PK} = \mathbf{P}^2$。对两边取 f，并利用 $\mathrm{f}(\mathbf{K}^{-1})\mathrm{f}(\mathbf{K}) = \mathrm{f}(\mathbf{I}) = 1$，得

$$\mathrm{f}(\mathbf{P}) = \mathrm{f}(\mathbf{P})^2,$$

所以 $\mathrm{f}(\mathbf{P}) = 0$ 或 1。但 $\mathbf{P}$ 是可逆的 (invertible)，而 $\mathrm{f}(\mathbf{P})\mathrm{f}(\mathbf{P}^{-1}) = \mathrm{f}(\mathbf{I}) = 1$，所以 $\mathrm{f}(\mathbf{P}) \neq 0$。因此 $\mathrm{f}(\mathbf{P}) = 1$。 ■

回顾

- **特殊元素揭示函数的面貌。**把乘法性质应用于 $\mathbf{I}$、$\mathbf{J}$、$\mathbf{K}$ 和 $\mathbf{P}$，便得出 f 在每个矩阵上的值，以及它对每种行运算的影响。
- **与自身的变换相等的元素。**两行相等的矩阵经行对调后保持不变，而行对调会改变 f 的符号；所以 f 在该处必为零。
- **共轭（conjugation）保持 f 不变。**$\mathrm{f}(\mathbf{K}^{-1}\mathbf{P}\mathbf{K}) = \mathrm{f}(\mathbf{P})$，这个事实在代数中一再用到。

背后的数学

每个乘性函数都是行列式的函数

题目在 $\mathrm{f}(\mathbf{P}) = 1$ 处便停下，但这个论证是可以完成的。每个可逆矩阵都是初等矩阵（elementary matrices，即行对调、行缩放和错切）的乘积，而题目已说明 f 在每一种上的表现。由此可知 $\mathrm{f}(\mathbf{M}) = \phi(\det \mathbf{M})$，其中 ϕ 是某个单一实数变量的函数，并满足 $\phi(xy) = \phi(x)\phi(y)$。行列式本身就是 $\phi(x) = x$ 的情况，但 $\phi(x) = x^3$ 或 $\phi(x) = |x|^{1/2}\operatorname{sign}(x)$ 同样可行。换言之，*行列式是矩阵上唯一真正崭新的乘性函数*。用群论（group theory）的语言来说，可逆矩阵所成的群之中，凡是交换群（commutative group）所能看见的一切，行列式都已尽数捕捉。

行列式作为面积

从几何上看，2×2 矩阵把单位正方形变成一个平行四边形，而 $\det \mathbf{M}$ 就是它的*有向面积*（signed area）：若定向（orientation）保持不变则为正，若定向反转则为负（图 7.3.1）。这样，乘法性质的意义便一目了然：相继进行两个变换，面积的比例因子便会相乘。行对调会反转定向，所以 $\det \mathbf{J} = -1$；错切把正方形推斜成面积相同的平行四边形，所以 $\det \mathbf{P} = 1$；而两行线性相关的矩阵会把正方形压扁，所以它的行列式为零。题目的每一步都有对应的图像。



图 7.3.1 行列式作为有向面积。两个基向量（basis vectors）的像（琥珀色、梅紫色）张成一个平行四边形。错切保持面积不变；对调基向量会反转定向；线性相关的两行（或两列）把正方形压缩到一条直线上。相继进行两个变换，有向面积便会相乘，这正是性质 $\mathrm{f}(\mathbf{M}\mathbf{N}) = \mathrm{f}(\mathbf{M})\mathrm{f}(\mathbf{N})$。

高维的行列式

对于 $n \times n$ 矩阵，行列式是有向体积的比例因子，并由三个性质所刻划：它对每一行都是线性的，两行对调时它会变号，而且它在单位矩阵上等于 1。单凭这三点，你就能推导出熟悉的展开公式，而高斯消元法则提供了一个高效的计算方法。行列式出现在多重积分的换元公式（雅可比行列式，Jacobian）之中，出现在逆矩阵的公式之中，并且如我们在 Section 5.4 所见，也出现在多项式的判别式（discriminant）之中。

思维技巧

> **用特殊输入探测未知函数**
>
> 当一个函数只能通过某个性质来认识时，把这个性质应用于你能想到的最简单的特殊输入，看看它会迫使甚么结果成立。
>
> 在这里，单位矩阵、一个对调、一个缩放和一个错切，已足以确定几乎一切。选取那些等于自身的平方、逆或共轭的输入，尤其有效。

延伸练习

1 **其他例子。** 证明 $\mathrm{f}(\mathbf{M}) = (\det \mathbf{M})^3$ 满足题目的所有条件，并证明 $\mathrm{f}(\mathbf{M}) = |\det \mathbf{M}|$ 满足乘法性质，但不满足 $\mathrm{f}(\mathbf{J}) \neq 1$。

2 **列运算。** 证明对于任何第二列是第一列的倍数的矩阵，f 的值同样为零。

3 **缩放。** 设 $\mathbf{K}_k = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}$。证明 $\mathrm{f}\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix} = \mathrm{f}(\mathbf{K}_k)$，并证明 $\mathrm{f}(\mathbf{K}_k)\mathrm{f}(\mathbf{K}_l) = \mathrm{f}(\mathbf{K}_{kl})$。

4 **归结为行列式。** 较难 证明每个可逆的 2×2 矩阵都可以写成形如 $\mathbf{J}$、$\mathbf{K}_k$ 和 $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ 的矩阵的乘积。由此推导出 $\mathrm{f}(\mathbf{M})$ 只取决于 $\det \mathbf{M}$。

延伸阅读

- **Sheldon Axler, *Linear Algebra Done Right***（Springer；网上免费阅览）。从行列式的性质出发来定义行列式，与本题的精神一脉相承。
- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge）。第 5 章从三条规则出发建立行列式的理论。

- **3Blue1Brown, “The determinant”**（影片，网上免费观看）。以精美的动画呈现行列式的面积诠释。

7.4 回到起点的矩阵

题目

STEP II 2020，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

`www.ocr.org.uk/Images/696589-2020-past-papers.zip`

引子

有些变换（transformation）只要重复足够多次，就会令一切回到原位。旋转 90° 要四步，反射则只要两步。哪些 2×2 矩阵（matrix）具有这种性质，即对某个 n 有 $\mathbf{M}^n = \mathbf{I}$？又怎样判断？

这道题令人意外之处，在于只需两个数就足以判定：迹（trace）和行列式（determinant）。你不必知道那四个元素，只需知道 $a + d$ 和 $ad - bc$。而且答案对现实世界有一个惊人的推论：它解释了为甚么晶体可以有 2 重、3 重、4 重和 6 重对称，却永远没有 5 重对称。这条晶体学“定律”屹立了两个世纪，直至 1982 年的一项发现才被打破，而该发现更赢得了诺贝尔奖。

动笔之前

想一想 每个 2×2 矩阵都满足 $\mathbf{M}^2 - (\operatorname{tr}\mathbf{M})\mathbf{M} + (\det\mathbf{M})\mathbf{I} = \mathbf{0}$（这是凯莱－哈密顿定理（Cayley–Hamilton theorem）；把它乘开来验证一下）。这对处理有关 $\mathbf{M}^2$ 的问题有甚么帮助？

它把 $\mathbf{M}^2$ 用 $\mathbf{M}$ 和 $\mathbf{I}$ 表示出来：$\mathbf{M}^2 = t\mathbf{M} - d\mathbf{I}$，其中 $t = \operatorname{tr}\mathbf{M}$，$d = \det\mathbf{M}$。于是，关于 $\mathbf{M}^2$ 的条件就变成一个把 $\mathbf{M}$、t 和 d 联系起来的条件。这就是整道题的关键。

逐步解题

阶段 1　平方的迹

$$\mathbf{M}^2 = \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{pmatrix},$$ 所以 $\operatorname{tr}(\mathbf{M}^2) = a^2 + 2bc + d^2 = (a+d)^2 - 2(ad-bc)$。 ■

阶段 2　$\pm\mathbf{I}$ 的平方根

> **想一想**　假设 $\mathbf{M}^2 = \mathbf{I}$。把它代入 $\mathbf{M}^2 = t\mathbf{M} - d\mathbf{I}$。若 $t \neq 0$，这说明 $\mathbf{M}$ 是怎样的矩阵?

若 $\mathbf{M}^2 = \mathbf{I}$，则 $\mathbf{I} = t\mathbf{M} - d\mathbf{I}$，所以 $t\mathbf{M} = (1+d)\mathbf{I}$。若 $t \neq 0$，则 $\mathbf{M}$ 是 $\mathbf{I}$ 的倍数，设为 $c\mathbf{I}$，而 $c^2 = 1$，所以 $\mathbf{M} = \pm\mathbf{I}$，但这已被排除。因此 $t = 0$，于是 $(1+d)\mathbf{I} = \mathbf{0}$，所以 $d = -1$。

反过来，若 $t = 0$ 且 $d = -1$，则 $\mathbf{M}^2 = t\mathbf{M} - d\mathbf{I} = \mathbf{I}$，而且 $\mathbf{M} \neq \pm\mathbf{I}$，因为 $\operatorname{tr}(\pm\mathbf{I}) = \pm 2 \neq 0$。 ■

至于 $\mathbf{M}^2 = -\mathbf{I}$：代入后得 $t\mathbf{M} = (d-1)\mathbf{I}$。若 $t \neq 0$，则 $\mathbf{M} = c\mathbf{I}$，其中 $c^2 = -1$，这对实数 c 是不可能的。所以 $t = 0$ 且 $d = 1$。反过来，由 $t = 0$ 和 $d = 1$ 可得 $\mathbf{M}^2 = -\mathbf{I}$。 ■

阶段 3　四次方根

> **停一停，试一试**　设 $\mathbf{N} = \mathbf{M}^2$，则 $\mathbf{M}^4 = \mathbf{I}$ 即是 $\mathbf{N}^2 = \mathbf{I}$。根据第 (ii) 部分，要么 $\mathbf{N} = \pm\mathbf{I}$，要么 $\operatorname{tr}\mathbf{N} = 0$ 且 $\det\mathbf{N} = -1$。为甚么第二种情况不可能发生?

$\det\mathbf{N} = \det(\mathbf{M}^2) = (\det\mathbf{M})^2 \geq 0$，所以 $\det\mathbf{N} \neq -1$。因此 $\mathbf{N} = \mathbf{M}^2 = \pm\mathbf{I}$。逆命题显然成立。 ■

至于 $\mathbf{M}^4 = -\mathbf{I}$：把第 (ii) 部分应用于 $\mathbf{N} = \mathbf{M}^2$，可知它成立当且仅当 $\operatorname{tr}(\mathbf{M}^2) = 0$ 且 $\det(\mathbf{M}^2) = 1$，即 $t^2 - 2d = 0$ 且 $d^2 = 1$。由于 $t^2 = 2d \geq 0$，必须有 $d = 1$，于是 $t^2 = 2$。所以

$$\mathbf{M}^4 = -\mathbf{I} \iff \det\mathbf{M} = 1 \text{ 且 } \operatorname{tr}\mathbf{M} = \pm\sqrt{2}. \qquad \blacksquare$$

阶段 4　一个不是旋转的 I 的八次方根

任何满足 $\mathbf{M}^4 = -\mathbf{I}$ 的矩阵都有 $\mathbf{M}^8 = \mathbf{I}$。旋转 45° 的矩阵的迹是 $\sqrt{2}$，行列式是 1；我们想找另一个迹和行列式都相同、但不是旋转的矩阵。取

$$\mathbf{M} = \begin{pmatrix} \sqrt{2} & -1 \\ 1 & 0 \end{pmatrix}, \qquad \operatorname{tr}\mathbf{M} = \sqrt{2}, \quad \det\mathbf{M} = 1.$$

根据第 3 步，$\mathbf{M}^4 = -\mathbf{I}$，所以 $\mathbf{M}^8 = \mathbf{I}$。但 $\mathbf{M}$ 既不是旋转也不是反射，因为旋转和反射都保持长度不变，而 $\mathbf{M}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\sqrt{2}\\1\end{pmatrix}$ 的长度是 $\sqrt{3} \neq 1$。 ■

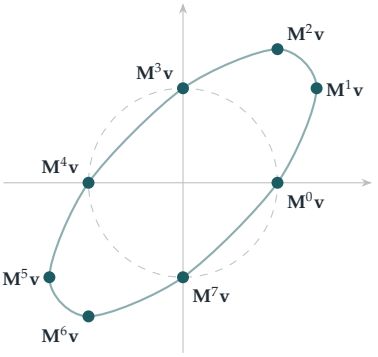


图 7.4.1 $\mathbf{v} = (1,0)$ 在 $\mathbf{M} = \begin{pmatrix}\sqrt{2} & -1\\1 & 0\end{pmatrix}$ 作用下的轨道（orbit）。八步之后它回到起点。这些点并不在同一个圆（虚线）上，所以 $\mathbf{M}$ 不是旋转，但它们确实位于一个椭圆上：$\mathbf{M}$ 是在斜坐标下看到的 45° 旋转。

回顾

- **凯莱－哈密顿定理把幂化为线性组合（linear combination）。** 2×2 矩阵的每一个幂都是 $\mathbf{M}$ 和 $\mathbf{I}$ 的组合，其系数只取决于迹和行列式。
- **平方的行列式是非负的。** 这个简单的事实在第 (iii) 部分排除了一整种情况。
- **迹和行列式是不变量（invariant）。** 两个迹和行列式都相同的矩阵（且都不是 $\mathbf{I}$ 的倍数），不论外表是否相似，取幂时的表现都一样。

背后的数学

相似矩阵

为甚么说第 4 步的 $\mathbf{M}$ 是“伪装了的旋转”？若存在可逆（invertible）矩阵 $\mathbf{S}$ 使得 $\mathbf{M} = \mathbf{SRS}^{-1}$，则称两个矩阵 $\mathbf{M}$ 和 $\mathbf{R}$ 相似（similar）：它们在不同的坐标下表示同一个变换。相似矩阵有相同的迹和行列式，而且（对于不是 $\mathbf{I}$ 的倍数的 2×2 矩阵）逆命题也成立。所以任何迹为 $\sqrt{2}$、行列式为 1 的实矩阵都与 45° 的旋转相似，而在 $\mathbf{S}$ 所给出的斜坐标下，它就只是在旋转（图 7.4.1）。一般而言，满足 $\mathbf{M}^n = \mathbf{I}$ 的 2×2 实矩阵必定与某个旋转或反射相似。

为甚么晶体没有五重对称

晶体是原子有规律、重复的排列。晶体的旋转对称把原子所成的格（lattice）映射到它自身，所以在以这个格为基础的坐标下，它由一个元素全是整数的矩阵给出。因此它的迹是整数。但是旋转角为 θ 的旋转，在任何坐标下的迹都是 $2\cos\theta$，而 $2\cos\theta$ 只有在等于 $-2, -1, 0, 1$ 或 2 时才是整数，也就是对应阶（order）为 2、3、4 和 6（以及 1）的旋转。五重对称需要 $2\cos 72^\circ = \frac{\sqrt{5}-1}{2}$，而这不是整数。这就是*晶体学限制定理*（crystallographic restriction theorem）。1982 年，Dan Shechtman 观察到一种看似具有五重对称的合金，起初还遭人嘲笑；后来才知道那是一种*准晶体*（quasicrystal），有序但没有周期性，就像彭罗斯铺砖（Penrose tilings）一样，因此不在该定理的涵盖范围之内。他于 2011 年获颁诺贝尔化学奖。

思维技巧

> **寻找能刻划本质行为的不变量**
>
> *当题目涉及一个有很多参数的对象时，问问自己：是否只需这些参数的几个组合，即那些转换观点后仍然不变的组合，就已经足够？*
>
> 在这里，四个元素化约为两个不变量，即迹和行列式，因为这个问题（“$\mathbf{M}^n = \mathbf{I}$ 是否成立?”）并不取决于坐标。找到合适的不变量，往往是解决分类问题的关键。

延伸练习

1. **三阶的旋转。**试做 STEP II 2021，第 7 题（载于 OCR 提供下载的 2021 年试卷，`www.ocr.org.uk/Images/696588-2021-past-papers.zip`）。它从一个具有不寻常性质的旋转矩阵出发，最后得出一个结论，适用于所有立方为单位矩阵（identity matrix）的 2×2 实矩阵。

2. **六阶。**求迹和行列式须满足的条件，使得 $\mathbf{M}^6 = \mathbf{I}$，但 $\mathbf{M}^2 \neq \mathbf{I}$ 且 $\mathbf{M}^3 \neq \mathbf{I}$。

3. **晶体学限制。**证明：若 $\mathbf{M}$ 的元素全是整数，$\det \mathbf{M} = 1$，且对某个 n 有 $\mathbf{M}^n = \mathbf{I}$，则 $\operatorname{tr} \mathbf{M} \in \{-2, -1, 0, 1, 2\}$。找出一个阶为 6 的整数矩阵。

4. **以凯莱－哈密顿定理求幂。**利用 $\mathbf{M}^2 = t\mathbf{M} - d\mathbf{I}$，把 $\mathbf{M}^3$ 和 $\mathbf{M}^4$ 表示成 $p\mathbf{M} + q\mathbf{I}$ 的形式，并借此重新推导 $\mathbf{M}^4 = -\mathbf{I}$ 的条件。

延伸阅读

- **M. A. Armstrong, *Groups and Symmetry***（Springer）。内容包括晶体学限制和墙纸图案的分类。
- **Marjorie Senechal, *Quasicrystals and Geometry***（Cambridge University Press）。准晶体与非周期有序结构的数学。
- **Gilbert Strang, *Introduction to Linear Algebra***（Wellesley-Cambridge）。相似矩阵与凯莱—哈密顿定理。

7.5 三维空间中的旋转

题目

STEP III 2022，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2022 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696587-2022-past-papers.zip`

引子

在二维空间中，旋转（rotation）一个点很容易：乘以 $e^{i\theta}$ 便可。到了三维空间，事情就困难得多。没有一个复数可以让我们直接相乘，绕不同轴的旋转不可交换（commute），而要写出“绕通过 $(1,2,2)$ 的轴旋转 $37°$”的矩阵，看来更是令人却步。

本题利用叉积（cross product），为任意的轴和任意的角构造出这个矩阵。所得的结果就是*罗德里格旋转公式*（Rodrigues' rotation formula），由罗德里格（Olinde Rodrigues）于 1840 年发表。机械人技术、电脑图像和太空船导航每天都用到它，而它亦揭示了一个优美的事实：与 $\mathbf{n}$ 作叉积，就是“绕 $\mathbf{n}$ 旋转 $90°$”，而每一个绕 $\mathbf{n}$ 的旋转都由它构成，正如 $e^{i\theta} = \cos\theta + i\sin\theta$ 是由乘以 i 构成一样。

动笔之前

想一想 当 $\mathbf{r}$ 垂直于 $\mathbf{n}$ 时，$\mathbf{n}\times\mathbf{r}$ 在几何上是甚么样子？当 $\mathbf{r}$ 平行于 $\mathbf{n}$ 时又如何？

若 $\mathbf{r} \parallel \mathbf{n}$，则 $\mathbf{n}\times\mathbf{r} = \mathbf{0}$。若 $\mathbf{r} \perp \mathbf{n}$，则 $\mathbf{n}\times\mathbf{r}$ 的长度与 $\mathbf{r}$ 相同（因为 $|\mathbf{n}| = 1$，而夹角是 $90°$），与两者皆垂直，而且根据右手定则（right-hand rule），它就是 $\mathbf{r}$ 绕 $\mathbf{n}$ 逆时针旋转 $90°$ 的结果（沿 $\mathbf{n}$ 向下看）。因此 f 消去 $\mathbf{r}$ 沿 $\mathbf{n}$ 方向的分量（component），并把垂直的分量旋转 $90°$。这幅图像足以解释整条题目。

逐步解题

阶段 1　作两次叉积

$\mathrm{f}(\mathbf{r}) = \mathbf{n}\times\mathbf{r} = (bz-cy,\ cx-az,\ ay-bx)$。再与 $\mathbf{n}=(a,b,c)$ 作一次叉积，$\mathrm{f}(\mathrm{f}(\mathbf{r}))$ 的 x 分量是

$$b(ay-bx)-c(cx-az) = -x(b^2+c^2)+aby+acz. \qquad \blacksquare$$

由于 $a^2+b^2+c^2=1$，上式等于 $-x(1-a^2)+aby+acz = a(ax+by+cz)-x$，这正是 $(\mathbf{n}.\mathbf{r})\mathbf{n}-\mathbf{r}$ 的 x 分量。其余分量可由对称性得出，所以 $\mathrm{f}(\mathrm{f}(\mathbf{r})) = (\mathbf{n}.\mathbf{r})\mathbf{n}-\mathbf{r}$。 $\blacksquare$

> **想一想**　设 $\mathbf{r}=\mathbf{r}_{\parallel}+\mathbf{r}_{\perp}$，其中 $\mathbf{r}_{\parallel}=(\mathbf{n}.\mathbf{r})\mathbf{n}$ 是沿 $\mathbf{n}$ 方向的分量。用这两部分表示，$\mathrm{f}(\mathrm{f}(\mathbf{r}))$ 是甚么?

$\mathrm{f}(\mathrm{f}(\mathbf{r})) = \mathbf{r}_{\parallel}-\mathbf{r} = -\mathbf{r}_{\perp}$。施行 f 两次，会消去平行的部分，并把垂直的部分绕 $\mathbf{n}$ 旋转 180°，亦即把它反向（图 7.5.1）。这就是 $\mathrm{i}^2=-1$ 的类比。

阶段 2　旋转

> **想一想**　计算 $\mathrm{g}(\mathbf{n})$。然后对于 $\mathbf{r}\perp\mathbf{n}$ 的情况计算 $\mathrm{g}(\mathbf{r})$，并利用 $\mathrm{f}(\mathrm{f}(\mathbf{r}))=-\mathbf{r}$ 对这样的 $\mathbf{r}$ 成立这一点。

$\mathrm{f}(\mathbf{n}) = \mathbf{n}\times\mathbf{n} = \mathbf{0}$，所以 $\mathrm{g}(\mathbf{n})=\mathbf{n}$：轴保持不动。若 $\mathbf{r}\perp\mathbf{n}$，则 $\mathrm{f}(\mathrm{f}(\mathbf{r}))=-\mathbf{r}$，于是

$$\mathrm{g}(\mathbf{r}) = \mathbf{r}+\sin\theta\,(\mathbf{n}\times\mathbf{r})-(1-\cos\theta)\mathbf{r} = \cos\theta\,\mathbf{r}+\sin\theta\,(\mathbf{n}\times\mathbf{r}).$$

由于 $\mathbf{n}\times\mathbf{r}$ 是 $\mathbf{r}$ 绕 $\mathbf{n}$ 转过 90° 的结果，长度不变，所以上式就是 $\mathbf{r}$ 绕 $\mathbf{n}$ 旋转 θ 的结果，与平面上的 $(\cos\theta+\mathrm{i}\sin\theta)z$ 如出一辙。由于 g 是线性（linear）的，而每个向量都可写成 $\mathbf{n}$ 的倍数与一个垂直于 $\mathbf{n}$ 的向量之和，所以 g 就是**绕通过原点、方向为 $\mathbf{n}$ 的轴旋转角 θ 的旋转**（从 $\mathbf{n}$ 的尖端望去为逆时针方向）。 $\blacksquare$

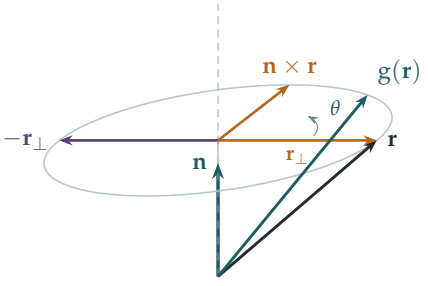


图 7.5.1　罗德里格公式。向量 $\mathbf{r}$ 分解为沿轴 $\mathbf{n}$ 的部分和与轴垂直的部分 $\mathbf{r}_{\perp}$，后者扫出一个圆。与 $\mathbf{n}$ 作叉积，会把 $\mathbf{r}_{\perp}$ 转过 90°（琥珀色）；作两次叉积则得 $-\mathbf{r}_{\perp}$（紫红色）。旋转后的向量 $\mathrm{g}(\mathbf{r})$ 是 $\mathbf{r}_{\parallel}+\cos\theta\,\mathbf{r}_{\perp}+\sin\theta\,(\mathbf{n}\times\mathbf{r})$。

阶段 3 反射

> **停一停，试一试** 利用第 (i) 部分化简 $\mathrm{h}(\mathbf{s}) = -\mathbf{s} - 2\mathrm{f}(\mathrm{f}(\mathbf{s}))$。它对沿 $\mathbf{n}$ 方向的向量有甚么作用？对垂直于 $\mathbf{n}$ 的向量又有甚么作用？

$\mathrm{h}(\mathbf{s}) = -\mathbf{s} - 2\big((\mathbf{n}.\mathbf{s})\mathbf{n} - \mathbf{s}\big) = \mathbf{s} - 2(\mathbf{n}.\mathbf{s})\mathbf{n}$。垂直于 $\mathbf{n}$ 的向量保持不变，而 $\mathrm{h}(\mathbf{n}) = -\mathbf{n}$。所以 h 是**对通过原点且垂直于 n 的平面的反射（reflection）**。 ■

回顾

- **分解为平行和垂直的部分。**一旦把 $\mathbf{r}$ 分解为沿 $\mathbf{n}$ 方向和与之垂直的分量，每一步都变得清晰。
- **线性减省工夫。**要辨认一个线性映射（linear map），只需知道它对几个精心挑选的向量有甚么作用便足够。
- **复数的类比。**在垂直于 $\mathbf{n}$ 的向量上，f 的表现就像乘以 i：$\mathrm{f}^2 = -1$，而 $\mathrm{g} = \cos\theta + \sin\theta\,\mathrm{f}$。

背后的数学

欧拉旋转定理

三维空间中每一个固定原点的旋转，都是绕某条通过原点的轴的旋转。欧拉（Euler）于 1775 年证明了这一点。用矩阵的语言来说，每个 3×3 旋转矩阵都有一个特征值（eigenvalue）为 1 的特征向量（eigenvector），也就是旋转轴。罗德里格公式给出其逆命题：对每一条轴 $\mathbf{n}$ 和每一个角 θ，旋转矩阵是

$$\mathbf{R} = \mathbf{I} + \sin\theta\,\mathbf{K} + (1-\cos\theta)\,\mathbf{K}^2, \qquad \mathbf{K} = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

其中 $\mathbf{K}$ 是 $\mathbf{r} \mapsto \mathbf{n}\times\mathbf{r}$ 的矩阵。由于 $\mathbf{K}^3 = -\mathbf{K}$，指数级数（exponential series）$e^{\theta\mathbf{K}} = \mathbf{I} + \theta\mathbf{K} + \frac{\theta^2}{2!}\mathbf{K}^2 + \cdots$ 恰好化简成这条公式，它就是 $e^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$ 的三维版本。旋转与“无穷小旋转”$\mathbf{K}$ 的指数之间的这种联系，正是李群（Lie groups）理论的起点。

旋转不可交换

把一本书平放，先绕垂直轴旋转 90°，再绕水平轴旋转 90°；然后从头再来，以相反的次序做这两个旋转。书最终的位置并不相同。三维空间中旋转的复合（composition）不符

合交换律，与二维的情况不同。这正是 Section 6.7 的四元数（quaternion）有用的原因之一：绕 $\mathbf{n}$ 旋转 θ 对应于四元数 $\cos\frac{\theta}{2} + \sin\frac{\theta}{2}(a\mathrm{i} + b\mathrm{j} + c\mathrm{k})$，而复合旋转就是把四元数相乘，不可交换性已包含在其中。罗德里格本人在 1840 年的论文中，已找到旋转的复合公式，比哈密顿（Hamilton）发明四元数还早三年。

思维技巧

分解为不变的部分

要理解一个变换，可把空间分成变换处理得简单的几块：在这里，就是轴（保持不动）和与轴垂直的平面（被旋转）。

这与 Section 7.2 中特征向量的想法相同：找出变换所保持的子空间（subspace），然后在每一块上分别研究它。

延伸练习

1 **平面上的旋转。**试做 STEP III 2017, 第 2 题（`step.maths.org/questions/17-s3-q2`）。该题利用复数，证明平面上的两个旋转复合起来仍是一个旋转，只有一种例外。找出这个例外，以及此时复合的结果是甚么。

2 **一个具体的旋转。**求绕方向为 $(1,1,1)$ 的轴旋转 120° 的矩阵。证明它把各坐标轴循环置换。

3 **$\mathbf{K}$ 的幂。**证明 $\mathbf{K}^2\mathbf{r} = (\mathbf{n}.\mathbf{r})\mathbf{n} - \mathbf{r}$ 及 $\mathbf{K}^3 = -\mathbf{K}$，并由指数级数推导出罗德里格公式。

4 **两次反射构成一个旋转。**证明先后对两个通过原点、互成 α 角的平面作反射，结果是绕两平面的交线旋转 2α。

延伸阅读

- **Simon Altmann, *Rotations, Quaternions, and Double Groups***（Dover）。关于三维旋转的权威论著，当中包括罗德里格的贡献。
- **Simon Altmann, "Hamilton, Rodrigues, and the quaternion scandal"**, *Mathematics Magazine* 62 (1989)。一篇简短而生动的历史叙述。
- **John Stillwell, *Naive Lie Theory***（Springer）。为大学本科生而写的旋转、四元数与李群入门。

7.6 双曲旋转与相对论

题目

STEP III 1991，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/91-s3-q6`

引子

旋转矩阵（rotation matrix）$\begin{pmatrix} \cos u & -\sin u \\ \sin u & \cos u \end{pmatrix}$ 把点沿着圆 $x^2 + y^2 = 1$ 移动，而先旋转 u 再旋转 w，就等于旋转 $u + w$。把 cos 和 sin 换成 cosh 和 sinh，再调整一个正负号，便得到本题的矩阵。它改为把点沿着双曲线（hyperbola）$x^2 - y^2 = 1$ 移动，而且它的“角”同样可以相加。这就是*双曲旋转*（hyperbolic rotation）。

这并非只是一件趣闻。若以 x 代表时间、y 代表距离（采用光速为 1 的单位），这个矩阵就是*劳仑兹推进*（Lorentz boost）：它是爱因斯坦（Einstein）狭义相对论（special relativity）中发现的变换，联系着两个以恒定相对速度运动的观察者各自的量度结果。“双曲角”u 称为*快度*（rapidity），而快度可以相加这一事实，正好解释了我们在 Section 3.8 遇过的那条奇怪的速度相加定律。

动笔之前

想一想 $\cosh u$ 和 $\sinh u$ 满足哪一条恒等式，与 $\cos^2 u + \sin^2 u = 1$ 相对应？$\cosh(u + v)$ 和 $\sinh(u + v)$ 的加法公式又是甚么？

$\cosh^2 u - \sinh^2 u = 1$，以及

$$\cosh(u + v) = \cosh u \cosh v + \sinh u \sinh v,$$
$$\sinh(u + v) = \sinh u \cosh v + \cosh u \sinh v.$$

你可以由定义 $\cosh u = \frac{1}{2}(e^u + e^{-u})$ 和 $\sinh u = \frac{1}{2}(e^u - e^{-u})$ 验证这些公式。点

$(\cosh v, \sinh v)$ 位于双曲线 $x^2 - y^2 = 1$ 上，正如 $(\cos v, \sin v)$ 位于圆上一样。

逐步解题

阶段 1　双曲线保持不变

$$\begin{aligned} x'^2 - y'^2 &= (x\cosh u + y\sinh u)^2 - (x\sinh u + y\cosh u)^2 \\ &= (x^2 - y^2)(\cosh^2 u - \sinh^2 u) = x^2 - y^2, \end{aligned}$$

因为交叉项 $2xy\cosh u \sinh u$ 互相抵消。所以 T 保持 $x^2 - y^2$ 不变，特别是把双曲线 $x^2 - y^2 = 1$ 映射到它自身。 ■

> **想一想**　哪些通过原点的直线会被映射到自身？（试找特征向量（eigenvector）。）

$T\begin{pmatrix}1\\1\end{pmatrix} = (\cosh u + \sinh u)\begin{pmatrix}1\\1\end{pmatrix} = e^u\begin{pmatrix}1\\1\end{pmatrix}$，而 $T\begin{pmatrix}1\\-1\end{pmatrix} = e^{-u}\begin{pmatrix}1\\-1\end{pmatrix}$。所以直线 $y = x$ 和 $y = -x$ 各自被映射到自身：一条被拉长 e^u 倍，另一条被缩短至 e^{-u} 倍。这两条直线正是双曲线的渐近线（asymptote）。 ■

阶段 2　双曲角相加

斜率（gradient）为 $\tanh v$ 的直线，其方向向量为 $(\cosh v, \sinh v)$。它的像的方向为

$$T\begin{pmatrix}\cosh v\\ \sinh v\end{pmatrix} = \begin{pmatrix}\cosh u\cosh v + \sinh u\sinh v\\ \sinh u\cosh v + \cosh u\sinh v\end{pmatrix} = \begin{pmatrix}\cosh(u+v)\\ \sinh(u+v)\end{pmatrix},$$

其斜率为 $\tanh(u+v)$。线性映射（linear map）把平行线变成平行线，所以无论直线是否通过原点，这个结论都成立。因此 $v' = v + u$。 ■

阶段 3　夹角何时保持不变？

> **想一想**　斜率分别为 m_1、m_2 的两条直线，其夹角 ψ 满足 $\tan\psi = \left|\frac{m_1 - m_2}{1 + m_1 m_2}\right|$。当 $m_i = \tanh v_i$ 时，化简这个式子。（提示：把所有项通分，以 $\cosh v_1 \cosh v_2$ 为分母。）

利用 $\tanh v_1 - \tanh v_2 = \dfrac{\sinh(v_1 - v_2)}{\cosh v_1 \cosh v_2}$ 和 $1 + \tanh v_1 \tanh v_2 = \dfrac{\cosh(v_1 + v_2)}{\cosh v_1 \cosh v_2}$，得

$$\tan\psi = \left|\frac{\sinh(v_1 - v_2)}{\cosh(v_1 + v_2)}\right|.$$

变换之后，$v_1 - v_2$ 不变，而 $v_1 + v_2$ 则变成 $v_1 + v_2 + 2u$。

> **停一停，试一试** 甚么时候 $\cosh(v_1 + v_2) = \cosh(v_1 + v_2 + 2u)$?

假设两条直线不平行（$v_1 \neq v_2$），则两个夹角相等，当且仅当 $\cosh(v_1 + v_2) = \cosh(v_1 + v_2 + 2u)$。由于 cosh 是偶函数，并且在 $[0, \infty)$ 上严格递增，这表示 $v_1 + v_2 = \pm(v_1 + v_2 + 2u)$。取正号得 $u = 0$，而这已被排除，所以

$$v_1 + v_2 = -u. \qquad \blacksquare$$

阶段 4 图像

当 $v_1 + v_2 = -u$ 时，变换后的角为 $v_1' = v_1 + u = -v_2$ 和 $v_2' = v_2 + u = -v_1$。所以 ℓ_1 的像的斜率为 $-\tanh v_2$：它是 ℓ_2 对 x 轴的反射，而 ℓ_2 的像则是 ℓ_1 的反射。对于通过原点的直线，这个变换把两条直线对调并加以反射，这就解释了为何它们之间的夹角不变（图 7.6.1）。

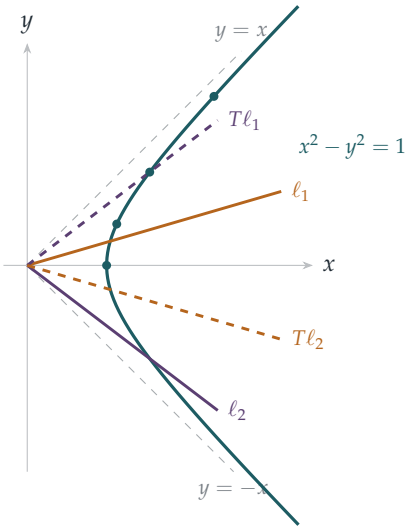


图 7.6.1 把 $u = 0.7$ 的双曲旋转施于通过原点的直线 ℓ_1、ℓ_2，其中 $v_1 = 0.3$，$v_2 = -1.0$，因此 $v_1 + v_2 = -u$。ℓ_1 的像（紫红色虚线）是 ℓ_2 的镜像，而 ℓ_2 的像（琥珀色虚线）是 ℓ_1 的镜像，所以两线之间的夹角保持不变。双曲线上的蓝绿色点按双曲角等距分布（$v = 0, 0.5, 1, 1.5$）；T 把每一点沿曲线再移前 0.7，到达新的位置。

回顾

- **不变的二次型（quadratic form）。** T 保持 $x^2 - y^2$ 不变，正如旋转保持 $x^2 + y^2$ 不变。

- **特征向量就是渐近线。**保持不变的直线是 $y = \pm x$，分别按 $e^{\pm u}$ 的倍数拉长和缩短。
- **双曲角可以相加。**以 $\tanh v$ 把方向参数化之后，T 的作用便化为简单的加法 $v \mapsto v + u$。

背后的数学

狭义相对论

在爱因斯坦的狭义相对论（1905 年）中，以恒定相对速度运动的不同观察者，对时间和距离的量度并不一致，但他们对 $t^2 - x^2$ 这个量（取光速 $c = 1$）却意见一致，这个量称为时空间隔（spacetime interval）。保持它不变的变换就是劳仑兹变换（Lorentz transformation），而在一个空间维度的情况下，它们恰好就是本题的矩阵，其中 x 为时间、y 为距离。一个以速度 $v = \tanh u$ 运动的观察者，所用的坐标与我们的坐标就是由这个矩阵联系起来，其中 $\cosh u = \frac{1}{\sqrt{1-v^2}}$，即时间膨胀（time dilation）中著名的因子 γ。闵可夫斯基（Hermann Minkowski）在 1908 年看出，这使时空成为一种四维几何，其“距离”之中带有一个负号：“从今以后，空间本身和时间本身，都注定要消退成为影子。”

题目的物理意义

- **不变直线** $y = \pm x$ 是光线的路径。它们对每个观察者都相同：光速对所有人都一样，这正是相对论的奠基原理。
- **双曲线** $x^2 - y^2 = 1$ 是与原点相距一个单位“原时”（proper time）的所有事件所成的集合，而原时是由以各种可能的恒定速度运动的观察者所量度的。
- **快度相加。**通过原点而斜率为 $\tanh v$ 的直线，是一个以速度 $\tanh v$ 运动的物体的路径。快度为 u 的推进把它的速度变为 $\tanh(v+u) = \frac{\tanh v + \tanh u}{1 + \tanh v \tanh u}$，这就是 Section 3.8 中爱因斯坦的速度相加公式。速度不能直接相加，快度却可以。

而在 Section 5.8 中，我们见过同一条加法公式，它是庞加莱圆盘（Poincaré disc）中的距离法则：相对论中所有可能的速度所成的集合，以快度来量度，就是一个双曲空间（hyperbolic space）。

思维技巧

找出合适的参数化

选用一个参数来描述你的对象，使变换在这个参数之下变得简单。

若以斜率描述方向，T 的作用是一条繁复的公式；若以 v 描述方向，其中斜率为 $\tanh v$，它就只是加上 u。同样的想法，曾令旋转在使用角度时变得简单，也曾在 Section 2.8 中令映射 $x \mapsto x^2 - 2$ 变得简单。

延伸练习

1 **推进的复合。**证明 $T(u)T(w) = T(u+w)$，其中 $T(u)$ 是题目中的矩阵。$T(u)^{-1}$ 是甚么？

2 **时间膨胀。**一艘太空船以光速的 0.8 倍航行。求它的快度 u 和 $\cosh u$。若地球上过了 10 年，船上过了多少时间？（利用 $t_{\text{船}} = t_{\text{地球}} / \cosh u$。）

3 **双曲面积。**证明由双曲线 $x^2 - y^2 = 1$、x 轴，以及由原点至 $(\cosh v, \sinh v)$ 的线段所围成的面积为 $\frac{v}{2}$，正如圆的相应面积为 $\frac{\theta}{2}$。为甚么 T 必定保持这类面积不变？

4 **挤压。**把坐标转换为 $p = x + y$，$q = x - y$。证明 T 变成 $p \mapsto e^u p$，$q \mapsto e^{-u} q$，而双曲线则变成 $pq = 1$。试在这组坐标下，以几何方式描述这个变换。

延伸阅读

- **Edwin Taylor & John Wheeler, *Spacetime Physics***（W. H. Freeman）。以快度和时空图讲解狭义相对论，适合具备高级程度（A-level）数学基础的学生。
- **Tevian Dray, *The Geometry of Special Relativity***（CRC Press）。把相对论看成双曲三角学，与本节的精神完全一致。
- **Albert Einstein, *Relativity: The Special and the General Theory***（多种版本）。爱因斯坦亲自撰写的通俗读物，至今仍然十分易读。

7.7 永恒不变的量

题目

STEP III 2024，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2024 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/725753-2024-past-papers.zip`

引子

物理学中一些最深刻的事实是*守恒定律*（conservation laws）：无论运动多么复杂，能量、动量和角动量始终保持不变。当一个系统有足够多的守恒量时，它的运动可以精确地解出；守恒量太少时，运动则可能是混沌的。

这道题展示了一种制造守恒量的绝妙方法。若一个矩阵按照 $\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t} = \mathbf{MN} - \mathbf{NM}$ 随时间变化，那么无论它的元素变化得多么剧烈，它的迹（trace）和行列式（determinant），以至它的特征值（eigenvalues），都永远不变。这样的一对 $(\mathbf{M}, \mathbf{N})$ 称为*拉克斯对*（Lax pair），以彼得．拉克斯（Peter Lax）命名。他在 1968 年提出这个想法，用来解释*孤立子*（solitons）的惊人表现：孤立子是一种孤立波，彼此穿过之后丝毫不变。

动笔之前

想一想 表达式 $\mathbf{MN} - \mathbf{NM}$ 称为 $\mathbf{M}$ 与 $\mathbf{N}$ 的*交换子*（commutator）。它在甚么时候等于零？根据第 (i) 部分，它的迹是多少？

它等于零，当且仅当 $\mathbf{M}$ 与 $\mathbf{N}$ 可交换。而根据第 (i) 部分，它的迹是 $\mathrm{tr}(\mathbf{MN}) - \mathrm{tr}(\mathbf{NM}) = 0$，恒常如此。这一个事实，就是题目中每一条守恒定律的起点。

逐步解题

阶段 1　迹

设 $\mathbf{M}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$ 及 $\mathbf{N}=\begin{pmatrix}p & q\\ r & s\end{pmatrix}$，则 $\mathbf{MN}$ 的对角线元素是 $ap+br$ 和 $cq+ds$，所以 $\operatorname{tr}(\mathbf{MN})=ap+br+cq+ds$。把 $\mathbf{M}$ 与 $\mathbf{N}$ 的角色对调，得 $pa+qc+rb+sd$，结果相同。显然 $\operatorname{tr}(\mathbf{M}+\mathbf{N})=\operatorname{tr}\mathbf{M}+\operatorname{tr}\mathbf{N}$。 ■

阶段 2　行列式的导数

? **想一想**　用积法则求 $\det\mathbf{M}=ad-bc$ 的导数。然后利用 $\mathbf{M}^{-1}=\frac{1}{\det\mathbf{M}}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$，计算 $\mathbf{M}^{-1}\frac{\mathrm{d}\mathbf{M}}{\mathrm{d}t}$ 及其迹。

以字母上的一点表示导数，则 $\frac{\mathrm{d}}{\mathrm{d}t}\det\mathbf{M}=\dot a d+a\dot d-\dot b c-b\dot c$。而

$$\mathbf{M}^{-1}\dot{\mathbf{M}}=\frac{1}{\det\mathbf{M}}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}\begin{pmatrix}\dot a & \dot b\\ \dot c & \dot d\end{pmatrix}$$

的对角线元素是 $\frac{d\dot a-b\dot c}{\det\mathbf{M}}$ 和 $\frac{-c\dot b+a\dot d}{\det\mathbf{M}}$。两者之和是 $\frac{1}{\det\mathbf{M}}\frac{\mathrm{d}}{\mathrm{d}t}\det\mathbf{M}$。 ■

阶段 3　三个守恒量

? **想一想**　对 $\dot{\mathbf{M}}=\mathbf{MN}-\mathbf{NM}$ 取迹。至于行列式，则用第 (ii) 部分：$\mathbf{M}^{-1}(\mathbf{MN}-\mathbf{NM})$ 是甚么？

迹：$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\mathbf{M}=\operatorname{tr}\dot{\mathbf{M}}=\operatorname{tr}(\mathbf{MN})-\operatorname{tr}(\mathbf{NM})=0$。

行列式：根据第 (ii) 部分，

$$\frac{1}{\det\mathbf{M}}\frac{\mathrm{d}}{\mathrm{d}t}\det\mathbf{M}=\operatorname{tr}(\mathbf{M}^{-1}(\mathbf{MN}-\mathbf{NM}))=\operatorname{tr}\mathbf{N}-\operatorname{tr}(\mathbf{M}^{-1}\mathbf{NM})=\operatorname{tr}\mathbf{N}-\operatorname{tr}\mathbf{N}=0,$$

当中用到第 (i) 部分的以下形式：$\operatorname{tr}(\mathbf{M}^{-1}(\mathbf{NM}))=\operatorname{tr}((\mathbf{NM})\mathbf{M}^{-1})=\operatorname{tr}\mathbf{N}$。

平方的迹：$\operatorname{tr}(\mathbf{M}^2)=(\operatorname{tr}\mathbf{M})^2-2\det\mathbf{M}$（见 Section 7.4），由于两项都是常数，所以它也是常数。 ■

阶段 4　一个显式解

✎ **停一停，试一试**　设 $\mathbf{N}=t\mathbf{I}+t\mathbf{E}$，其中 $\mathbf{E}=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$，计算 $\mathbf{MN}-\mathbf{NM}$，并写出 a、b、c、d 所满足的微分方程。按以下次序求解：先解 c，然后 a 和 d，最后 b。

$t\mathbf{I}$ 这部分与任何矩阵都可交换，所以 $\mathbf{MN}-\mathbf{NM}=t(\mathbf{ME}-\mathbf{EM})$。现在 $\mathbf{ME}=\begin{pmatrix}0 & a\\ 0 & c\end{pmatrix}$

而 $\mathbf{EM} = \begin{pmatrix} c & d \\ 0 & 0 \end{pmatrix}$，所以

$$\dot{a} = -tc, \qquad \dot{b} = t(a-d), \qquad \dot{c} = 0, \qquad \dot{d} = tc.$$

因此 $c = C$ 是常数；于是 $a = A - \frac{1}{2}Ct^2$ 及 $d = D + \frac{1}{2}Ct^2$；继而 $\dot{b} = t(A - D - Ct^2)$，所以 $b = B + \frac{1}{2}(A-D)t^2 - \frac{1}{4}Ct^4$。故此

$$\mathbf{M} = \begin{pmatrix} A - \frac{1}{2}Ct^2 & B + \frac{1}{2}(A-D)t^2 - \frac{1}{4}Ct^4 \\ C & D + \frac{1}{2}Ct^2 \end{pmatrix}. \qquad \blacksquare$$

验算：迹是 $A + D$，而行列式是 $(A - \frac{1}{2}Ct^2)(D + \frac{1}{2}Ct^2) - C\big(B + \frac{1}{2}(A-D)t^2 - \frac{1}{4}Ct^4\big) = AD - BC$，一如所料是常数（图 7.7.1）。

阶段 5 另一条方程

想一想 若 $\dot{\mathbf{M}} = \mathbf{MN}$，则 $\mathbf{N} = \mathbf{M}^{-1}\dot{\mathbf{M}}$，而根据第 (ii) 部分，$\operatorname{tr}\mathbf{N}$ 就是 $\det\mathbf{M}$ 的对数导数（logarithmic derivative）。$\mathbf{M}$ 的迹能否保持不变，而它的行列式却在改变?

当然可以。取 $\mathbf{M} = \begin{pmatrix} 1 & t \\ t & 1 \end{pmatrix}$，其中 $|t| < 1$（此时 $\det\mathbf{M} = 1 - t^2 \neq 0$）。它的迹恒等于 2。定义 $\mathbf{N} = \mathbf{M}^{-1}\dot{\mathbf{M}}$，使得 $\dot{\mathbf{M}} = \mathbf{MN}$：

$$\mathbf{N} = \frac{1}{1-t^2}\begin{pmatrix} 1 & -t \\ -t & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{1-t^2}\begin{pmatrix} -t & 1 \\ 1 & -t \end{pmatrix},$$

这是一个非奇异（non-singular）矩阵，而且当 $t \neq 0$ 时，$\operatorname{tr}\mathbf{N} = \frac{-2t}{1-t^2} \neq 0$。所以答案是**否定的**，$\operatorname{tr}\mathbf{N}$ 不一定等于零。 $\blacksquare$

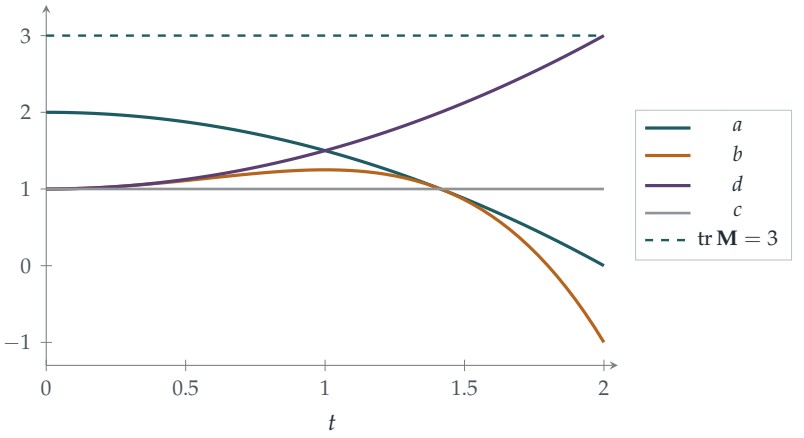


图 7.7.1　第 4 阶段中 $\mathbf{M}(t)$ 的各个元素，其中 $A = 2$、$B = 1$、$C = 1$、$D = 1$。四个元素之中除了 c 以外全都在变，但迹始终是 3，行列式始终是 1。所以特征值 $\frac{3\pm\sqrt{5}}{2}$ 永远不变：这个流是等谱（isospectral）的。

回顾

- **交换子的迹等于零。**由 $\operatorname{tr}(\mathbf{MN}) = \operatorname{tr}(\mathbf{NM})$ 出发，每一条守恒定律都随之而来。
- **对数导数。**第 (ii) 部分把行列式的导数化为一个迹，而迹是线性的，容易处理。
- **在显式解中检查守恒。**我们确认了 $\mathbf{M}(t)$ 的迹和行列式都是常数，借此验算它的公式。

背后的数学

孤立子与拉克斯对

1834 年，苏格兰工程师 John Scott Russell 策马追踪一个沿爱丁堡附近的联合运河（Union Canal）前进的孤立波，追了超过一英里，波形始终没有改变。这类波由科特韦赫－德弗里斯（*Korteweg–de Vries*，简称 *KdV*）方程描述。到了 1960 年代，Norman Zabusky 与 Martin Kruskal 的电脑实验显示，两个这样的“孤立子”可以相撞，然后原封不动地离开，仿佛它们是粒子一样。要解释这个现象，需要无穷多个守恒量。彼得. 拉克斯在 1968 年证明，KdV 方程可以写成 $\dot{L} = LP - PL$，其中 L 和 P 是一对微分算子（differential operators）；于是，与这道题完全一样，L 的特征值是守恒的，而它们正好提供了那无穷多条守恒定律。拥有拉克斯对的系统称为可积（integrable）系统，它们属于少数可以精确求解的非线性系统。

等谱流

由于 $\mathbf{M}$ 的迹和行列式决定了它的特征多项式（characteristic polynomial）$\lambda^2 - (\operatorname{tr}\mathbf{M})\lambda + \det\mathbf{M}$，$\mathbf{M}$ 的特征值不随时间改变。流 $\dot{\mathbf{M}} = [\mathbf{M}, \mathbf{N}]$ 使 $\mathbf{M}$ 在特征值固定的矩阵所组成的集合之内移动。事实上，解总是具有 $\mathbf{M}(t) = \mathbf{S}(t)^{-1}\mathbf{M}(0)\mathbf{S}(t)$ 的形式，其中 $\mathbf{S}(t)$ 是某个可逆矩阵：这个矩阵始终与它的初始值相似（similar）。同一个想法贯穿整个物理学：在量子力学中，可观测量（observable）$\mathbf{A}$ 按 $\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \frac{\mathrm{i}}{\hbar}[\mathbf{H}, \mathbf{A}]$ 演化，其中 $\mathbf{H}$ 是能量，所以凡是与能量可交换的量都是守恒的。

雅可比公式

第 (ii) 部分对任何大小的可逆矩阵都成立，称为雅可比公式（Jacobi's formula）。最稳妥的写法是不用对数，写成 $\frac{\mathrm{d}}{\mathrm{d}t}\det\mathbf{M} = \det\mathbf{M}\,\operatorname{tr}\left(\mathbf{M}^{-1}\dot{\mathbf{M}}\right)$，因为 $\det\mathbf{M}$ 可能是负数；对数形式 $\frac{\mathrm{d}}{\mathrm{d}t}\ln\det\mathbf{M} = \operatorname{tr}\left(\mathbf{M}^{-1}\dot{\mathbf{M}}\right)$ 需要 $\det\mathbf{M} > 0$（一般情况下则用 $\ln|\det\mathbf{M}|$）。由此可得一个推论：若 $\dot{\mathbf{M}} = \mathbf{A}\mathbf{M}$，其中 $\mathbf{A}$ 是常数矩阵，则 $\det\mathbf{M}(t) = e^{t\operatorname{tr}\mathbf{A}}\det\mathbf{M}(0)$。在微分方程的研究中，这个结果告诉你相空间（phase space）中的体积如何增长或缩小；对于经典力学的方程，这个迹等于零，体积保持不变，这个结果称为刘维尔定理（Liouville's theorem）。

思维技巧

> **找出不变的东西**
>
> 当某样东西在演化时，试找出由它构成而又保持不变的量。这些量限制了运动，并往往揭示出它的结构。
>
> 在这里，迹和行列式之所以守恒，是因为变化率是一个交换子。在力学中，守恒量来自对称性（诺特定理，Noether's theorem），而找出守恒量往往是解题的第一步。

延伸练习

1. **验算解。**通过求导，验证第 4 阶段的矩阵 $\mathbf{M}(t)$ 满足 $\dot{\mathbf{M}} = \mathbf{M}\mathbf{N} - \mathbf{N}\mathbf{M}$。

2. **相似性。**证明：若 $\mathbf{M}(t) = \mathbf{S}^{-1}\mathbf{M}_0\mathbf{S}$，其中 $\mathbf{M}_0$ 是常数矩阵，而 $\mathbf{S}(t)$ 满足 $\dot{\mathbf{S}} = \mathbf{S}\mathbf{N}$，则 $\dot{\mathbf{M}} = \mathbf{M}\mathbf{N} - \mathbf{N}\mathbf{M}$。（提示：$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S}^{-1} = -\mathbf{S}^{-1}\dot{\mathbf{S}}\mathbf{S}^{-1}$。）

3 **体积增长。**设 $\mathbf{M}(t)=\begin{pmatrix} e^{2t} & 0 \\ 0 & e^{-t} \end{pmatrix}$。求满足 $\dot{\mathbf{M}}=\mathbf{M}\mathbf{N}$ 的 $\mathbf{N}$，并验证 $\frac{\mathrm{d}}{\mathrm{d}t}\ln\det\mathbf{M}=\operatorname{tr}\mathbf{N}$。

4 **更高次幂。**直接证明：当 $\dot{\mathbf{M}}=\mathbf{M}\mathbf{N}-\mathbf{N}\mathbf{M}$ 时，对每个正整数 k 都有 $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}(\mathbf{M}^k)=0$。

延伸阅读

- **P. G. Drazin & R. S. Johnson, *Solitons: An Introduction***（Cambridge University Press）。大学本科程度的标准入门书，内容包括拉克斯对。
- **Thierry Dauxois & Michel Peyrard, *Physics of Solitons***（Cambridge University Press）。书中讲述了 Scott Russell 那个孤立波的故事。
- **Dwight Neuenschwander, *Emmy Noether's Wonderful Theorem***（Johns Hopkins University Press）。讲解对称性如何带来守恒定律。

7.8 四个平方数

题目

STEP III 1994，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/94-s3-q8`

引子

7 不是两个平方数之和，也不是三个平方数之和，却是四个平方数之和：$7 = 4 + 1 + 1 + 1$。1770 年，拉格朗日（Joseph-Louis Lagrange）证明了每一个正整数都是四个平方数之和（容许使用 0^2）。这条定理早在 1621 年已由 Bachet 提出猜想，费马（Fermat）亦声称自己证得，但一如以往没有留下证明；欧拉（Euler）为此钻研了四十年，始终差一步未能完成。

关键的工具由欧拉于 1748 年发现，正正就在这道题目之中：两个四平方和的积，本身也是一个四平方和。这表示只需对质数（prime number）证明拉格朗日定理，其余情况靠相乘便可得出。题目利用复矩阵（complex matrix）的行列式（determinant），寥寥数行便推导出这条恒等式（identity），而这些矩阵其实就是 Section 6.7 中的四元数（quaternion），只是换了个面貌。

动笔之前

想一想 回想两平方的版本：$(a^2 + b^2)(p^2 + q^2)$ 是两个平方数之和，因为对复数 $z = a + b\mathrm{i}$、$w = p + q\mathrm{i}$ 有 $|z|^2|w|^2 = |zw|^2$。矩阵可以怎样把这个结果推广到四个平方？

想法是找出一种由四个实数构成的“数”，它的“大小的平方”是 $a^2 + b^2 + c^2 + d^2$，而且具有可乘性（multiplicative）。矩阵的行列式正有可乘性：$\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$。因此我们要找一个由 a, b, c, d 构成的矩阵，其行列式为 $a^2 + b^2 + c^2 + d^2$，并且这类矩阵相乘后仍然是同一形式。

逐步解题

阶段 1　行列式是四个平方数之和

想一想　$\begin{pmatrix} z_1 & z_2 \\ -z_2^* & z_1^* \end{pmatrix}$ 的行列式是甚么？若 $z_1 = a + b\mathrm{i}$ 及 $z_2 = c + d\mathrm{i}$，以 a, b, c, d 表示又是甚么？

$\det = z_1 z_1^* + z_2 z_2^* = |z_1|^2 + |z_2|^2 = a^2 + b^2 + c^2 + d^2$。同样地，取 $z_3 = p + q\mathrm{i}$ 及 $z_4 = r + s\mathrm{i}$，第二个矩阵的行列式为 $p^2 + q^2 + r^2 + s^2$。

阶段 2　乘积具有相同形式

相乘得

$$\begin{pmatrix} z_1 & z_2 \\ -z_2^* & z_1^* \end{pmatrix}\begin{pmatrix} z_3 & z_4 \\ -z_4^* & z_3^* \end{pmatrix} = \begin{pmatrix} z_1 z_3 - z_2 z_4^* & z_1 z_4 + z_2 z_3^* \\ -z_2^* z_3 - z_1^* z_4^* & -z_2^* z_4 + z_1^* z_3^* \end{pmatrix}.$$

设 $w_1 = z_1 z_3 - z_2 z_4^*$ 及 $w_2 = z_1 z_4 + z_2 z_3^*$，则下面一行是 $(-w_2^*, w_1^*)$，所以乘积是 $\begin{pmatrix} w_1 & w_2 \\ -w_2^* & w_1^* \end{pmatrix}$，形式相同。它的行列式是 $|w_1|^2 + |w_2|^2$，而由于行列式可以相乘，

$$(a^2 + b^2 + c^2 + d^2)(p^2 + q^2 + r^2 + s^2) = |w_1|^2 + |w_2|^2.$$

阶段 3　读出四个平方

停一停，试一试　把 $w_1 = (a + b\mathrm{i})(p + q\mathrm{i}) - (c + d\mathrm{i})(r - s\mathrm{i})$ 及 $w_2 = (a + b\mathrm{i})(r + s\mathrm{i}) + (c + d\mathrm{i})(p - q\mathrm{i})$ 展开，写成实部（real part）和虚部（imaginary part）。

$$w_1 = (ap - bq - cr - ds) + \mathrm{i}(aq + bp + cs - dr),$$
$$w_2 = (ar - bs + cp + dq) + \mathrm{i}(as + br - cq + dp).$$

所以

$$L_1 = ap - bq - cr - ds, \qquad L_2 = aq + bp + cs - dr,$$
$$L_3 = ar - bs + cp + dq, \qquad L_4 = as + br - cq + dp,$$

每一个对 a, b, c, d 及对 p, q, r, s 都是线性（linear）的，并且 $(a^2 + b^2 + c^2 + d^2)(p^2 + q^2 + r^2 + s^2) = L_1^2 + L_2^2 + L_3^2 + L_4^2$。 ■

例如，$7 = 2^2 + 1^2 + 1^2 + 1^2$ 而 $3 = 1^2 + 1^2 + 1^2 + 0^2$。取 $(a, b, c, d) = (2, 1, 1, 1)$ 及 $(p, q, r, s) = (1, 1, 1, 0)$，由公式得 $(L_1, L_2, L_3, L_4) = (0, 2, 4, 1)$，而果然 $0^2 + 2^2 + 4^2 + 1^2 = 21 = 7 \times 3$。

回顾

- **可乘性是关键。**这条恒等式来自 $\det(\mathbf{AB}) = \det\mathbf{A}\det\mathbf{B}$，正如两平方恒等式来自 $|zw| = |z||w|$ 一样。
- **对乘法封闭（closure）。**$\begin{pmatrix} z_1 & z_2 \\ -z_2^* & z_1^* \end{pmatrix}$ 这个特殊形式在相乘之下保持不变，结果才会是另一个四平方和。
- **细心记账十分重要。**最后一步虽属例行运算，却容易出错；正负号和共轭（conjugate）都必须小心处理。

背后的数学

拉格朗日四平方和定理

拉格朗日的证明大致如下。根据欧拉恒等式，只需证明每个质数 p 都是四个平方数之和。先用一个计数论证，证明存在某个倍数 mp（其中 $m < p$）是四个平方数之和。然后用一个巧妙的“递降法”（descent）证明：若 mp 是四个平方数之和而 $1 < m < p$，则对某个更小的 m'，$m'p$ 也是四个平方数之和；如此重复下去，便会到达 $m = 1$。（可与 Section 1.5 中费马的递降法比较。）1834 年，雅可比（Jacobi）更进一步，求出把 n 写成四个平方数之和的方法的确切数目：n 的所有非 4 的倍数的因数之和的 8 倍。例如，1 有 8 种写法（± 1 可放在四个位置中的任何一个）。

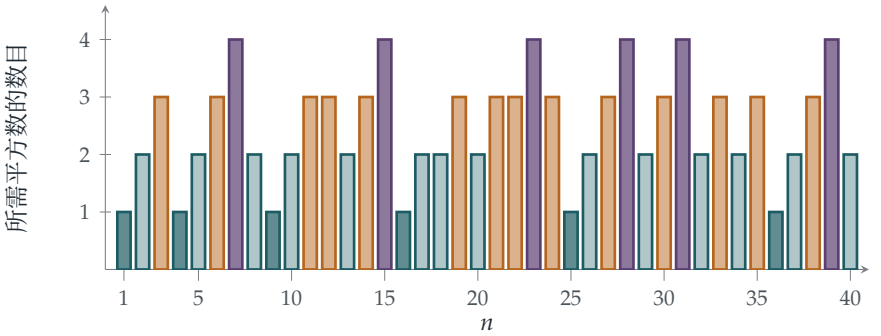


图 7.8.1 由 1 至 40 的每个 n 最少需要多少个平方数才能组成。根据拉格朗日定理，每个数最多只需四个。需要四个的数（紫红色）是 $7, 15, 23, 28, 31, 39, \ldots$：根据勒让德（Legendre）和高斯（Gauss）的一条定理，它们正好是形如 $4^k(8m+7)$ 的数。

二、三、四

不同维数之间的规律相当微妙。两平方和来自复数，而一个数是两个平方数之和，当且仅当它的质因数中形如 $4k+3$ 的都以偶数次幂出现（Section 1.6）。四平方和来自四元数，而每个数都是四平方和。但三平方和却完全没有乘法恒等式（并不存在一个“大小”具可乘性的三维数系；这正是哈密顿（Hamilton）研究三元组时失败的原因，见 Section 6.7），相应的定理也较难：勒让德和高斯证明了，除非 $n = 4^k(8m+7)$，否则 n 是三个平方数之和（图 7.8.1）。四以上，八维的八元数（octonion）给出一条八平方恒等式，由 Degen 和 Graves 发现；而胡尔维兹（Hurwitz）于 1898 年证明，这类恒等式只在 1、2、4 和 8 个平方的情况下存在。

华林问题

拉格朗日定理是华林（Edward Waring）于 1770 年提出的一个问题的第一个情况：对每个 k，是否存在一个数 $g(k)$，使得每个正整数都是最多 $g(k)$ 个 k 次幂之和？希尔伯特（David Hilbert）于 1909 年证明了确实存在。现在我们知道 $g(2)=4$、$g(3)=9$（例如 $23 = 2\cdot 2^3 + 7\cdot 1^3$ 需要九个立方数）以及 $g(4)=19$。

思维技巧

寻找具可乘性的结构

要证明某个性质在乘法之下得以保持，可寻找一个代数结构（复数、四元数、矩阵），使该性质就是其中某个对象的“大小”，而大小具有可乘性。

一旦找到这样的结构，乘积恒等式便唾手可得，而一个关于所有数的问题，也就化为一个关于质数的问题。

延伸练习

1. **一个实例。**利用 $L_1,\dots,L_4$ 的公式，取 $(a,b,c,d)=(p,q,r,s)=(2,1,1,1)$，把 $49=7\times 7$ 写成四个平方数之和，并验算你的答案。它与显而易见的 $49=7^2+0^2+0^2+0^2$ 是否相同？

2. **四平方中的两平方。**证明若 $c=d=r=s=0$，这条恒等式便简化为 Section 1.6 的两平方恒等式。

3 **需要四个平方的数。**考虑平方数模 8 的余数，证明形如 $8m+7$ 的数都不是三个平方数之和。

4 **雅可比公式。**数一数把 2 表示成四个平方数之和的方法有多少种（次序和正负号不同均视为不同），并验证答案是 $8\times(1+2)=24$。

延伸阅读

- **Harold Davenport, *The Higher Arithmetic***（Cambridge University Press）。第 V 章证明了拉格朗日四平方和定理。
- **John Conway & Derek Smith, *On Quaternions and Octonions***（A. K. Peters）。讲述四元数的算术，包括利用四元数整数证明四平方和定理。
- **Aigner & Ziegler, *Proofs from THE BOOK***，“Representing numbers as sums of two squares”（把数表示成两个平方数之和）一章。与本节相配的两平方版本。

7.9 只有五种柏拉图立体

题目

STEP III 1999，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/99-s3-q4`

引子

正多边形（regular polygon）有无穷多种：三角形、正方形、五边形，如此类推，任何边数都有。你或会以为正多面体（regular polyhedron）也有无穷多种。其实恰好只有五种：正四面体（tetrahedron）、正方体（cube）、正八面体（octahedron）、正十二面体（dodecahedron）和正二十面体（icosahedron）。柏拉图（Plato）把其中四种与四大元素（火、土、气、水）对应，第五种则对应宇宙；欧几里得（Euclid）的《几何原本》（*Elements*）是有史以来影响最深远的数学著作，全书最后就是证明再没有其他正多面体；开普勒（Kepler）更用它们建构了一个太阳系模型。

本题只用寥寥数行就证明了这个分类，所依据的只是欧拉（Euler）在 1750 年发现的一条非凡公式：对任何凸多面体（convex polyhedron），$V - E + F = 2$。一条关于计数的公式，化成了一个关于形状的定理，而 2 这个数原来正是球面的“指纹”。

动笔之前

想一想 用正方体验证欧拉公式：它有多少个顶点（vertex）、多少条棱（edge）、多少个面（face）？以正方形为底的棱锥体（pyramid）又如何？

正方体有 $V = 8$，$E = 12$，$F = 6$，而 $8 - 12 + 6 = 2$。以正方形为底的棱锥体有 5 个顶点、8 条棱和 5 个面，而 $5 - 8 + 5 = 2$。任何没有洞的多面体，不论多么不规则，这条公式都成立。

想一想　在正多面体中，每个面有 m 条棱，而每条棱是两个面的公共边。由此可得 E 与 F 之间的甚么方程？同样地，E 与 V 之间又有甚么方程？

用两种方法计数棱与面的关联：F 个面各有 m 条棱，而每条棱被数了两次（它所属的两个面各数一次），所以 $mF = 2E$。同样地，每个顶点连着 n 条棱，而每条棱有两个端点，所以 $nV = 2E$。这种“用两种方法计数同一样东西”的技巧称为*双重计数*（double counting）。

逐步解题

阶段 1　F 的公式

由 $mF = 2E$ 和 $nV = 2E$，可得 $E = \frac{mF}{2}$ 和 $V = \frac{mF}{n}$。代入欧拉公式，

$$\frac{mF}{n} - \frac{mF}{2} + F = 2 \quad\Longrightarrow\quad F\,\frac{2m - mn + 2n}{2n} = 2 \quad\Longrightarrow\quad F = \frac{4n}{2m + 2n - mn}.$$

而 $4 - (n-2)(m-2) = 4 - (mn - 2m - 2n + 4) = 2m + 2n - mn = h$。所以 $F = \frac{4n}{h}$。∎

阶段 2　m 和 n 可以取哪些值？

想一想　m（每个面的边数）和 n（每个顶点处的面数）最小可以是多少？又既然 $F > 0$，h 必须是正还是负？

多边形最少要有 3 条边，所以 $m \geq 3$；立体的每个顶点处最少要有 3 个面相交，所以 $n \geq 3$。由于 $F > 0$，我们需要 $h > 0$，即 $(n-2)(m-2) < 4$。两个因子都至少是 1，所以 $(m-2, n-2)$ 只可能是 $(1,1)$、$(1,2)$、$(2,1)$、$(1,3)$ 和 $(3,1)$。

停一停，试一试　就这五种可能逐一计算 h，然后求 F，再求 $E = \frac{mF}{2}$ 和 $V = \frac{mF}{n}$。指出每种情况是哪一个立体。

m	n	h	F	E	V	立体
3	3	3	4	6	4	正四面体
3	4	2	8	12	6	正八面体
4	3	2	6	12	8	正方体
3	5	1	20	30	12	正二十面体
5	3	1	12	30	20	正十二面体

可能的情况只有这些，所以正多面体最多只有五种；而这五种全部存在（欧几里得说明了如何逐一构作）。∎

回顾

- **双重计数。**用两种方法计数棱与面、棱与顶点的关联，得出 $mF = 2E$ 和 $nV = 2E$。
- **由正值得出不等式。**$F > 0$ 这个简单的要求，把无穷多个候选者筛剩五个。
- **存在性要另行处理。**这个论证只说明最多有五种；每一种是否真的存在，还须实际构作出来才能确定。

背后的数学

欧拉公式与拓扑学

欧拉在 1750 年提出 $V - E + F = 2$，但他的证明有漏洞；这条公式其后经历了一段引人入胜的历史，当中有一次又一次的尝试证明、反例（counterexample）和修补。一个“画框”形的多面体，中间穿了一个正方形的洞，有 $V - E + F = 0$；由两个正方体在一个顶点处相连而成的多面体，则有 $V - E + F = 3$。拉卡托斯（Imre Lakatos）把这段历史写成一本名著 *Proofs and Refutations*（《证明与反驳》），探讨数学概念如何在反例的磨砺下变得精确。最终的答案是：$V - E + F$ 只取决于曲面在拓扑学（topology）那种宽松意义下的形状：凡是可以变形成球面的曲面，它等于 2；环面（torus，即冬甩形）是 0；有 g 个洞的曲面则是 $2 - 2g$。这个数称为欧拉示性数（Euler characteristic），是拓扑学最早的概念之一；拓扑学研究的，就是经拉伸和弯曲后仍然不变的性质。

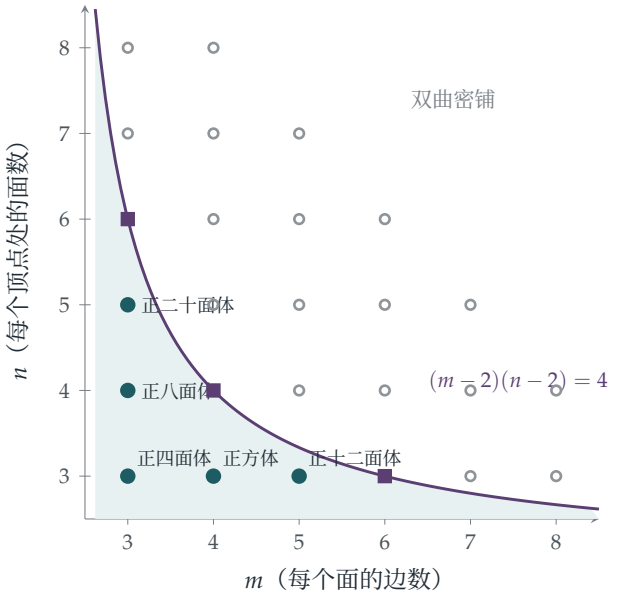


图 7.9.1　满足 $m, n \geq 3$ 的所有可能数对 (m, n)。在曲线 $(m-2)(n-2)=4$ 下方（阴影部分）的是五种柏拉图立体（蓝绿色）。在曲线上的是平面的三种密铺（紫红色方点）：分别用三角形（$m=3$，$n=6$）、正方形（4,4）和六边形（6,3）。曲线上方的每一点（空心圆圈）都对应双曲平面的一种正则密铺。

球面以外

当 $h=0$ 或 $h<0$ 时又会怎样？当 $(m-2)(n-2)=4$ 时，有三组解：(3,6)、(4,4) 和 (6,3)，即六个三角形、四个正方形或三个六边形在一点相接，各角之和恰好是 360°。这就是平面的三种正则密铺（regular tiling），也就是浴室地砖和蜂巢的图案。而当 $(m-2)(n-2)>4$ 时，一个顶点处各角之和大于 360°，这在平面上不可能出现，但在 Section 5.8 的双曲平面（hyperbolic plane）上却完全可行。双曲平面有无穷多种正则密铺，图 7.9.1 中曲线上方的每一点各对应一种；艾雪（Escher）的木刻版画 *Circle Limit*（《圆极限》）系列，画的就是其中几种。所以，柏拉图立体的分类其实是一个三分法（trichotomy）的第一种情况，这个三分法（球面、平面、双曲平面）贯穿整个几何学。

对称群

柏拉图立体的对称构成群（group）：正四面体有 12 个旋转，正方体和正八面体有 24 个（两者互为对偶（dual）：把面与顶点互换，$m \leftrightarrow n$），正十二面体和正二十面体则有 60 个。这个阶（order）为 60 的正二十面体群是一个“单”群（simple group）；正如我们在 Section 6.4 提过，正是它令五次方程（quintic）不能用根式（radicals）求解。克莱因（Felix Klein）写了一整本书 *Lectures on the Icosahedron*（《正二十面体讲义》，1884），把

正二十面体与五次方程的解法联系起来。

思维技巧

双重计数

用两种不同的方法计数同一组东西，然后令两个结果相等。

这里所计数的，是所有（棱，面）配对所成的集合，其中棱在该面上。双重计数是组合数学（combinatorics）中最强大的工具之一，在第 9 章会一再出现。

延伸练习

1 **对偶性。**证明把 m 和 n 互换，会令 V 和 F 互换，而 E 保持不变。哪几对柏拉图立体互为对偶？哪一个是自对偶的？

2 **足球。**传统的足球由 12 个五边形和 20 个六边形组成，每个顶点处有三个面相接。利用双重计数和欧拉公式，验证 $V - E + F = 2$。再证明更一般的结果：任何这样由五边形和六边形组成、每个顶点处有三个面相接的多面体，都恰好有 12 个五边形。

3 **环面。**造一个“画框”形的多面体（一个正方形方块，中间有一个正方形的洞），数出 V、E 和 F。验证 $V - E + F = 0$。

4 **顶点处的角。**正 m 边形的内角（interior angle）是 $180°(1 - \frac{2}{m})$。证明 n 个这样的角能围着凸立体的一个顶点拼合（总和小于 $360°$），当且仅当 $(m-2)(n-2) < 4$，从而得出这个分类的第二个证明。

延伸阅读

- **David Richeson, *Euler's Gem***（Princeton University Press）。讲述 $V - E + F = 2$ 的历史，以及它如何引出拓扑学。非常适合高中第六学级（sixth form）的学生阅读。
- **Imre Lakatos, *Proofs and Refutations***（Cambridge University Press）。一场关于欧拉公式和证明本质的对话。
- **H. S. M. Coxeter, *Regular Polytopes***（Dover）。论述正多面体及其高维同类的经典著作。

我们学到了甚么

本章开首我们说过，矩阵（matrix）是空间的一种变换（transformation），自有其生命；我们还提出了三个问题：矩阵与数有何相似、有何不同，矩阵究竟做了甚么，以及甚么保持不变。以下是本章九道题目所揭示的答案。

1. 矩阵与数有何相似，又有何不同？

矩阵可以相加和相乘，但 Section 7.1 指出了哪些熟悉的法则不再成立。乘法不符合交换律（commutative law），所以 $(\mathbf{A}-\mathbf{B})(\mathbf{A}+\mathbf{B})$ 未必等于 $\mathbf{A}^2-\mathbf{B}^2$；矩阵中存在零因子（zero divisor），所以 $\mathbf{AB}=\mathbf{0}$ 并不保证 $\mathbf{BA}=\mathbf{0}$；而一条二次矩阵方程可以有无限多个根。每一项法则的失效都有几何解释，可以从这些矩阵对平面的作用来理解。至关重要的是，有一项性质保留了下来：行列式（determinant）可乘，即 $\det(\mathbf{AB})=\det\mathbf{A}\det\mathbf{B}$。Section 7.3 说明这项性质基本上刻划了行列式，而 Section 7.8 则用它证明了欧拉（Euler）的四平方恒等式（four-square identity），这正是拉格朗日（Lagrange）定理的关键一步；该定理指出每个正整数都是四个平方数之和。

2. 矩阵究竟做了甚么？

有几节通过找出合适的观点来回答这个问题。

- **特征向量（eigenvector）**（Section 7.2）是矩阵不会转动的方向。在以特征向量为轴的坐标下，对称矩阵（symmetric matrix）只不过是一个伸缩，而圆的像是一个椭圆，其轴和面积都可以直接读出。
- **行列式**（Section 7.3）是带正负号的面积比例因子：错切（shear）保持面积不变，对换会把定向（orientation）反转，奇异矩阵（singular matrix）则把平面压扁。
- **迹（trace）与行列式**（Section 7.4）通过凯莱–哈密顿定理（Cayley–Hamilton theorem），决定一个矩阵会否在若干步之后回到单位矩阵（identity matrix）。满足 $\mathbf{M}^8=\mathbf{I}$ 的矩阵未必是旋转；它是在歪斜坐标下看到的旋转。又由于整数矩阵的迹是整数，晶体不可能有五重对称。
- **三维空间中的旋转**（Section 7.5）可由向量积（cross product）通过罗德里格斯公式（Rodrigues′ formula）构造出来，这是 $e^{i\theta}=\cos\theta+i\sin\theta$ 在三维的对应版本。

- **双曲旋转（hyperbolic rotation）**（Section 7.6）使点沿双曲线移动，并把它们的双曲角相加。它们就是狭义相对论中的洛伦兹变换（Lorentz boost），而它们的不变直线（invariant line）就是光的路径。

3. 甚么保持不变？

不变量（invariant）无处不在。旋转保持 $x^2 + y^2$ 不变，洛伦兹变换保持 $x^2 - y^2$ 不变，而相似矩阵（similar matrices）有相同的迹和行列式。在 Section 7.7 中，一个按 $\dot{\mathbf{M}} = \mathbf{MN} - \mathbf{NM}$ 演化的矩阵，不论其元素如何变化，它的迹和行列式都保持不变，因而它的特征值（eigenvalue）也保持不变：这就是孤立子（soliton）各条守恒定律背后的机制。而在 Section 7.9 中，不变量是一个数，$V - E + F = 2$，它对每个凸多面体（convex polyhedron）都成立，并把正多面体限定为仅有五种。这个数称为欧拉示性数（Euler characteristic），是拓扑学（topology）中最早出现的不变量之一。

4. 线性代数工具箱

方法	你在哪里用过
几何反例	失效的法则（7.1）
以特征向量为自然坐标轴	椭圆（7.2）；洛伦兹变换（7.6）
用特殊矩阵作试探	行列式（7.3）
凯莱–哈密顿定理	有限阶（7.4）
分解为平行和垂直部分	罗德里格斯公式（7.5）
选择合适的参数	快度（rapidity）（7.6）
$\mathrm{tr}(\mathbf{MN}) = \mathrm{tr}(\mathbf{NM})$	守恒量（7.7）
行列式的可乘性	四平方和（7.8）
重复计数（double counting）	柏拉图立体（Platonic solids）（7.9）

下一步

- 在 Section 11.2 中，特征值和特征向量支配着弹簧与质点系统的振动。
- 在 Section 13.7 中，协方差矩阵（covariance matrix）的特征向量会在统计学中出现。
- 微分方程（differential equation）和动力系统（dynamical system）是第 8 章的主题。

如果你想在入读大学之前再走远一点

- **3Blue1Brown, *Essence of Linear Algebra***（线性代数的本质）：一套免费的影片系列，展示矩阵究竟做了甚么。请先看这个。

- **Gilbert Strang, *Introduction to Linear Algebra***（线性代数导论）：经典的入门课程，十分着重理解和应用。
- **Sheldon Axler, *Linear Algebra Done Right***（线性代数应该这样学）：较为抽象的处理手法，以线性映射（linear map）为中心；可在网上免费阅读。
- **David Richeson, *Euler's Gem***（欧拉的瑰宝）：欧拉公式与拓扑学的诞生。

Kaplansky 曾打趣说，他和 Halmos 思考时不用坐标，但关起门来却“拼命地”用矩阵计算。这句话正好道出本章想说明的事：矩阵既具体又可以计算，但每一项计算背后都有几何（一个伸缩、一个旋转、一个投影、一个洛伦兹变换），还有一个不变量，它不取决于你怎样把它写下来。学会在这两种观点之间来回转换，是数学家所能拥有的最宝贵技能之一。

延伸练习答案

以下只是简短的答案和提示，并非完整解答：这里给出结果和关键步骤，让你核对自己的做法。请先认真做过每道练习，然后才看答案。

7.1 对矩阵不成立的法则

1. 设 $\mathbf{X}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$，比较 $\mathbf{XP}$ 与 $\mathbf{PX}$，必有 $c=0$ 及 $a=d$：所求的矩阵（matrix）为 $\begin{pmatrix}a & b\\ 0 & a\end{pmatrix}$。

2. 由凯莱—哈密顿定理（Cayley–Hamilton theorem）得 $\mathbf{N}^2=t\mathbf{N}-d\mathbf{I}$，所以 $t\mathbf{N}=d\mathbf{I}$。若 $t\neq 0$，则 $\mathbf{N}$ 是 $\mathbf{I}$ 的倍数，而 $\mathbf{N}^2=\mathbf{0}$ 迫使 $\mathbf{N}=\mathbf{0}$，与 $t\neq 0$ 矛盾。所以 $t=0$，从而 $d=0$。答案是：$\begin{pmatrix}a & b\\ c & -a\end{pmatrix}$，其中 $a^2+bc=0$。

3. 平方后的对角线元素为 $\cos^2\theta+\sin^2\theta=1$，非对角线元素为 0。保持不变的向量（vector）是 $(\cos\frac{\theta}{2},\sin\frac{\theta}{2})$，所以它是对通过原点、倾角为 $\frac{\theta}{2}$ 的直线的反射（reflection）。

4. 设 $\mathbf{M}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$，则 $\mathbf{M}^2=\begin{pmatrix}a^2+bc & b(a+d)\\ c(a+d) & d^2+bc\end{pmatrix}$；减去 $(a+d)\mathbf{M}$ 再加上 $(ad-bc)\mathbf{I}$，便得 $\mathbf{0}$。

7.2 只伸缩，不转向

1. $\mathbf{Mu}=\lambda\mathbf{u}$ 且 $\mathbf{u}\neq\mathbf{0}$，必须有 $\det(\mathbf{M}-\lambda\mathbf{I})=0$，这就是该二次方程。这里迹（trace）为 6，行列式（determinant）为 $\frac{4}{25}(54-4)=8$，所以 $\lambda^2-6\lambda+8=0$：$\lambda=2$（对应 $(1,2)$）和 4（对应 $(2,-1)$）。

2. $\mathbf{v}\cdot\mathbf{Mu}=\lambda\,\mathbf{v}\cdot\mathbf{u}$，又因为 $\mathbf{M}$ 是对称矩阵（symmetric matrix），$\mathbf{v}\cdot\mathbf{Mu}=\mathbf{Mv}\cdot\mathbf{u}=\mu\,\mathbf{v}\cdot\mathbf{u}$。所以 $(\lambda-\mu)\,\mathbf{u}\cdot\mathbf{v}=0$。

3. 特征方程（characteristic equation）$\lambda^2-2\lambda\cos\theta+1=0$ 的判别式（discriminant）为 $-4\sin^2\theta$，除非 θ 是 π 的倍数，否则必为负数。特征值（eigenvalue）是 $\mathrm{e}^{\pm\mathrm{i}\theta}$。

4. $(1,0)=\frac{1}{5}(1,2)+\frac{2}{5}(2,-1)$，所以 $T^n(1,0)=\frac{2^n}{5}(1,2)+\frac{2\cdot 4^n}{5}(2,-1)=\frac{1}{5}\big(2^n+4^{n+1},\,2^{n+1}-2\cdot 4^n\big)$。当 $n=1$ 时，结果是 $(\frac{18}{5},-\frac{4}{5})$，正如所料。

7.3 甚么才算行列式

1. $(\det \mathbf{MN})^3 = (\det \mathbf{M})^3 (\det \mathbf{N})^3$，而 $(\det \mathbf{J})^3 = -1 \neq 1$。函数 $|\det \mathbf{M}|$ 具有可乘性（multiplicative），但 $|\det \mathbf{J}| = 1$。
2. $\begin{pmatrix} a & \lambda a \\ c & \lambda c \end{pmatrix} = \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \begin{pmatrix} 1 & \lambda \\ 0 & 0 \end{pmatrix}$，而第二个因子的第二行是第一行的 0 倍，所以它的 f 值为 0。
3. $\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{J}\mathbf{K}_k\mathbf{J}$，所以它的值是 $\mathrm{f}(\mathbf{J})^2\mathrm{f}(\mathbf{K}_k) = \mathrm{f}(\mathbf{K}_k)$。另外 $\mathbf{K}_k\mathbf{K}_l = \mathbf{K}_{kl}$。
4. 行简化（row reduction）可把任何可逆矩阵（invertible matrix）写成初等矩阵（elementary matrix）的乘积：行对换（$\mathbf{J}$）、行倍乘（$\mathbf{K}_k$ 或 $\mathbf{J}\mathbf{K}_k\mathbf{J}$），以及错切（shear）（$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ 或它经 $\mathbf{J}$ 共轭（conjugate）后所得的矩阵）。错切的 $\mathrm{f} = 1$（做法如第 (iv) 部分，利用 $\mathbf{K}^{-1}\mathbf{P}\mathbf{K}$）。此外，$\mathbf{J}$ 是若干错切与 $\mathbf{K}_{-1}$ 的乘积，所以 $\mathrm{f}(\mathbf{K}_{-1}) = \mathrm{f}(\mathbf{J}) = -1$。因此 $\mathrm{f}(\mathbf{M}) = \mathrm{g}(\det \mathbf{M})$，其中 $\mathrm{g}(k) = \mathrm{f}(\mathbf{K}_k)$；若 $\mathbf{M}$ 是奇异矩阵（singular matrix），则由第 (iii) 部分得 $\mathrm{f}(\mathbf{M}) = 0$。

7.4 回到起点的矩阵

1. 由 $|1 + \mathrm{e}^{\mathrm{i}\phi}|^2 = 2 + 2\cos\phi = 1$ 得 $\phi = \pm\frac{2\pi}{3}$，所以 $\mathbf{R}^3 = \mathbf{I}$。至于 $\mathbf{S}$：它的特征值是 1 的立方根。假如两个特征值都是 1，则由于 $x^3 - 1$ 的根各不相同，$\mathbf{S}$ 可对角化（diagonalisable），因而等于 $\mathbf{I}$。实矩阵的非实数特征值必以共轭对的形式出现，所以特征值是 $\mathrm{e}^{\pm 2\pi\mathrm{i}/3}$：$\det \mathbf{S} = 1$ 及 $\operatorname{tr} \mathbf{S} = -1$。
2. 特征值必定是六次本原单位根（primitive sixth roots of unity）$\mathrm{e}^{\pm\mathrm{i}\pi/3}$（若选取其他六次单位根并与其共轭配对，都会得出 $\mathbf{M}^2 = \mathbf{I}$ 或 $\mathbf{M}^3 = \mathbf{I}$）。所以 $\operatorname{tr} \mathbf{M} = 1$ 及 $\det \mathbf{M} = 1$。
3. 特征值是乘积为 1 的单位根，所以它们是 $\mathrm{e}^{\pm\mathrm{i}\theta}$（或同为 ± 1），而迹 $2\cos\theta$ 是 $[-2, 2]$ 内的整数。阶（order）为 6 的例子：$\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$。
4. $\mathbf{M}^3 = (t^2 - d)\mathbf{M} - td\mathbf{I}$ 及 $\mathbf{M}^4 = (t^3 - 2td)\mathbf{M} - d(t^2 - d)\mathbf{I}$。要使 $\mathbf{M}^4 = -\mathbf{I}$（$\mathbf{M}$ 不可能是 $\mathbf{I}$ 的倍数），须有 $t(t^2 - 2d) = 0$ 及 $d(t^2 - d) = 1$。$t = 0$ 给出 $-d^2 = 1$，不可能成立；所以 $t^2 = 2d$，$d^2 = 1$，于是 $d = 1$，$t^2 = 2$。

7.5 三维空间中的旋转

1. $z \mapsto a + \mathrm{e}^{\mathrm{i}\theta}(z - a)$。复合变换是 $z \mapsto \mathrm{e}^{\mathrm{i}(\theta+\phi)}z + k$，其中 $k = \mathrm{e}^{\mathrm{i}\phi}a(1 - \mathrm{e}^{\mathrm{i}\theta}) + b(1 - \mathrm{e}^{\mathrm{i}\phi})$；若 $\mathrm{e}^{\mathrm{i}(\theta+\phi)} \neq 1$，它有不动点（fixed point）$c = \frac{k}{1 - \mathrm{e}^{\mathrm{i}(\theta+\phi)}}$。若 $\theta + \phi = 2\pi$，它就是位

移为 $(b-a)(1-\mathrm{e}^{-\mathrm{i}\theta})$ 的平移（translation）。

2. $\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}$：它保持 $(1,1,1)$ 不变，迹为 $0=1+2\cos 120^\circ$，并把 $\mathbf{i}\to\mathbf{j}\to\mathbf{k}\to\mathbf{i}$。

3. $\mathbf{K}^2\mathbf{r}=\mathbf{n}\times(\mathbf{n}\times\mathbf{r})=(\mathbf{n}.\mathbf{r})\mathbf{n}-\mathbf{r}$，于是 $\mathbf{K}^3\mathbf{r}=\mathbf{n}\times\big((\mathbf{n}.\mathbf{r})\mathbf{n}-\mathbf{r}\big)=-\mathbf{K}\mathbf{r}$。在级数（series）中，奇数次幂给出 $(\theta-\frac{\theta^3}{3!}+\cdots)\mathbf{K}=\sin\theta\,\mathbf{K}$，偶数次幂（由 $\mathbf{K}^2$ 起）给出 $(\frac{\theta^2}{2!}-\frac{\theta^4}{4!}+\cdots)\mathbf{K}^2=(1-\cos\theta)\mathbf{K}^2$。

4. 复合变换保持交线上的每一点不变，而且行列式为 $(-1)^2=1$。在每个垂直于该交线的平面上，它是对两条夹角为 α 的直线先后作反射的乘积，即转角为 2α 的旋转（rotation）。

7.6 双曲旋转与相对论

1. $T(u)T(w)$ 的元素是 $\cosh u\cosh w+\sinh u\sinh w=\cosh(u+w)$ 和 $\sinh u\cosh w+\cosh u\sinh w=\sinh(u+w)$。所以 $T(u)^{-1}=T(-u)$。

2. 由 $\tanh u=0.8$ 得 $u=\frac12\ln\frac{1.8}{0.2}=\ln 3\approx 1.099$ 及 $\cosh u=\frac12(3+\frac13)=\frac53$。在飞船上，过去的时间是 $10\times\frac35=6$ 年。

3. 面积是 $\frac12\oint(x\,\mathrm{d}y-y\,\mathrm{d}x)$。两条直边没有贡献；沿着双曲线（hyperbola），取 $x=\cosh s$，$y=\sinh s$，被积函数（integrand）是 $(\cosh^2 s-\sinh^2 s)\,\mathrm{d}s=\mathrm{d}s$，得出 $\frac{v}{2}$。T 的行列式为 $\cosh^2 u-\sinh^2 u=1$，所以它保持面积不变；它把双曲线映射到自身，并把参数为 v 的点移到参数为 $v+u$ 的点。

4. $x'+y'=(\cosh u+\sinh u)(x+y)=\mathrm{e}^u p$ 及 $x'-y'=\mathrm{e}^{-u}q$；另外 $x^2-y^2=pq$。在这组坐标下，T 是一个挤压映射（squeeze）：沿渐近线（asymptote）$y=x$ 拉伸 e^u 倍，沿 $y=-x$ 压缩至 e^{-u} 倍，使各点沿着每条双曲线 $pq=$ 常数滑动。

7.7 永恒不变的量

1. $\dot a=-Ct=-tc$，$\dot d=Ct=tc$，$\dot c=0$，以及 $\dot b=(A-D)t-Ct^3=t(a-d)$，与 $\mathbf{M}\dot{\mathbf{N}}-\dot{\mathbf{N}}\mathbf{M}=t\begin{pmatrix}-c&a-d\\0&c\end{pmatrix}$ 吻合。

2. $\dot{\mathbf{M}}=-\mathbf{S}^{-1}\dot{\mathbf{S}}\mathbf{S}^{-1}\mathbf{M}_0\mathbf{S}+\mathbf{S}^{-1}\mathbf{M}_0\dot{\mathbf{S}}=-\mathbf{N}\mathbf{M}+\mathbf{M}\mathbf{N}$，当中用到 $\dot{\mathbf{S}}=\mathbf{S}\mathbf{N}$。

3. $\mathbf{N}=\mathbf{M}^{-1}\dot{\mathbf{M}}=\begin{pmatrix}2&0\\0&-1\end{pmatrix}$，而 $\ln\det\mathbf{M}=2t-t=t$，其导数（derivative）为 $1=\operatorname{tr}\mathbf{N}$。

4. 利用迹的循环性质（cyclic property），$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}(\mathbf{M}^k)=\sum_j\operatorname{tr}(\mathbf{M}^j\dot{\mathbf{M}}\mathbf{M}^{k-1-j})=$

$k\,\mathrm{tr}(\mathbf{M}^{k-1}\dot{\mathbf{M}})$；再用一次循环性质，$\mathrm{tr}(\mathbf{M}^{k-1}(\mathbf{MN}-\mathbf{NM})) = \mathrm{tr}(\mathbf{M}^k\mathbf{N}) - \mathrm{tr}(\mathbf{M}^{k-1}\mathbf{NM}) = 0$。

7.8 四个平方数

1. $(L_1, L_2, L_3, L_4) = (1,4,4,4)$，而 $1+16+16+16 = 49$。这个表示法与 $7^2+0^2+0^2+0^2$ 不同。

2. 取 $c = d = r = s = 0$：$L_1 = ap - bq$，$L_2 = aq + bp$，$L_3 = L_4 = 0$，即 $(a^2+b^2)(p^2+q^2) = (ap-bq)^2 + (aq+bp)^2$。

3. 平方数模（modulo）8 只可能是 0、1 或 4。三个这样的数之和模 8 可得 $0,1,2,3,4,5,6$，但永远不会是 7。

4. $2 = (\pm1)^2 + (\pm1)^2 + 0^2 + 0^2$：选出两个非零的位置（6 种方法）及它们的正负号（4 种方法），共 24 种。由于 2 不能被 4 整除，雅可比（Jacobi）公式给出 $8\sigma(2) = 8\times 3 = 24$。

7.9 只有五种柏拉图立体

1. 设 $h = 2m + 2n - mn$，则 $V = \frac{4m}{h}$，$E = \frac{2mn}{h}$ 及 $F = \frac{4n}{h}$；把 m 和 n 对调，V 和 F 也随之对调。正方体与正八面体（octahedron）互为对偶（dual），正十二面体（dodecahedron）与正二十面体（icosahedron）亦然；正四面体（tetrahedron）则是自对偶（self-dual）的。

2. $F = 32$，$E = \frac{1}{2}(12\cdot 5 + 20\cdot 6) = 90$ 及 $V = \frac{1}{3}(180) = 60$，而 $60 - 90 + 32 = 2$。一般而言，若有 P 个五边形和 H 个六边形，则 $E = \frac{1}{2}(5P+6H)$ 及 $V = \frac{1}{3}(5P+6H)$，所以 $V - E + F = -\frac{1}{6}(5P+6H) + P + H = \frac{P}{6}$。这个值等于 2，所以 $P = 12$。

3. 把顶面和底面（两者都有洞）各分成四个梯形（trapezium）。于是 $V = 16$，$F = 4+4+4+4 = 16$ 及 $E = 12+12+8 = 32$，所以 $V - E + F = 0$。

4. $180^\circ n(1 - \frac{2}{m}) < 360^\circ$ 即 $n(m-2) < 2m$，也就是 $mn - 2m - 2n < 0$，或 $(m-2)(n-2) < 4$。当 $m, n \geq 3$ 时，解为 $(3,3)$、$(3,4)$、$(4,3)$、$(3,5)$ 和 $(5,3)$。

Beyond STEP

STEP 是大多数中学生一生中应考过最艰深的数学考试，也是最出色的考试之一：题目简短、陌生而深刻，当中最精彩的，其实是乔装了的大学数学小品。

本书精选最有趣的 STEP 试题，由 1987 年首次开考之前的样本试卷（specimen papers），到 2025 年的试卷，并以每一道题目作为一扇门。每道题目都会仔细解答，沿途提出优秀学生会问的问题，然后带你穿过这扇门：这个想法从何而来，到了大学会演变成甚么，以及接下来该读些甚么。

本书为正在准备 STEP、有天分又有志投身数学的学生而写。他们想要的不止是模拟试卷：他们想开始像数学系本科生那样思考，并一探前方有甚么正等着自己。

- 共 13 章，涵盖纯数学（pure mathematics）、力学（mechanics）、概率（probability）与统计（statistics）
- 136 道 STEP 历届试题，每题均附引导式解答
- 每道题目背后的大学数学
- 附答案的延伸练习，以及延伸阅读

全书共 13 章，本册为第 7 章。 全书各章及所有详细解答均可免费阅读、下载及分享，网址为

beyondstep.keithmcnulty.org

文字内容已贡献予公有领域（CC0 1.0）。STEP 试题 © Cambridge University Press & Assessment。

