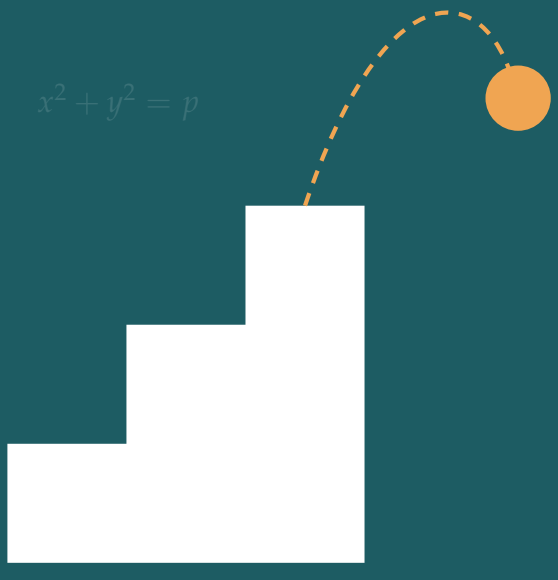
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1990 年详细解答



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP I, II 及 III

1990 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1990 年详细解答

1990 年 STEP 试卷（STEP I, II 及 III）全部 48 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

阅读解答之前，请先认真尝试解题。每个答案均已用数值方法核对，但解答数量如此之多，难免会有疏漏：如果发现有可疑之处，请相信自己的演算，并仔细检查。

由机器从英文翻译而成；尚未经审阅。

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。

目录

STEP I

第 1 题 三角函数的最优化 纯数学 3

第 2 题 单位根筛选法 纯数学 6

第 3 题 法线与平行曲线 纯数学 10

第 4 题 等差数列与差分 纯数学 14

第 5 题 部分分式与曲线描绘 纯数学 17

第 6 题 向量与平行四边形 纯数学 21

第 7 题 一阶线性常微分方程 纯数学 25

第 8 题 三角函数无穷乘积 纯数学 28

第 9 题 切线与面积的界限 纯数学 32

第 10 题 功率与阻力 力学 35

第 11 题 抛体与动量 力学 38

第 12 题 弹簧与能量 力学 42

第 13 题 摩擦力与刚体平衡 力学 46

第 14 题 期望值与最优停止 概率与统计 50

第 15 题 二项分布 概率与统计 53

第 16 题 角分布与假设检验 概率与统计 57

STEP II

第 1 题 多项式不等式 纯数学 61

第 2 题 托勒密定理 纯数学 65

第 3 题 三次曲线的描绘 纯数学 69

第 4 题 直线排布与数学归纳法 纯数学 74

第 5 题 复数与圆 纯数学 78

第 6 题 整数不等式与分数 纯数学 82

第 7 题 微分方程 纯数学 86
第 8 题 分部积分法 纯数学 90
第 9 题 极坐标曲线与弧长 纯数学 93
第 10 题 矩阵多项式方程 纯数学 97
第 11 题 角动量守恒 力学 101
第 12 题 恢复系数与抛体 力学 105
第 13 题 杆的平衡 力学 109
第 14 题 弹性碰撞 力学 112
第 15 题 顺序统计量 概率与统计 116
第 16 题 贝叶斯定理 概率与统计 119

STEP III

第 1 题 倍角与多项式的根 纯数学 122
第 2 题 三维向量 纯数学 126
第 3 题 群公理与共轭 纯数学 130
第 4 题 三角级数 纯数学 133
第 5 题 莱布尼兹定理 纯数学 136
第 6 题 线性变换与椭圆 纯数学 140
第 7 题 隐函数曲线描绘 纯数学 145
第 8 题 恰当微分方程 纯数学 150
第 9 题 双曲函数 纯数学 153
第 10 题 根与渐近近似 纯数学 156
第 11 题 刚体转动 力学 160
第 12 题 移动楔上的质点 力学 164
第 13 题 球内的运动 力学 168
第 14 题 质系 力学 172
第 15 题 几何分布 概率与统计 176
第 16 题 几何概率 概率与统计 180

详细解答

STEP I 1990，第 1 题

纯数学 · 三角函数的最优化

> **STEP I 1990，第 1 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-s1-q1

关键思路

窗帘的每个顶点与圆心 D 的距离都是 r，所以半径 DX 和 DY 把这个四边形分成三个以 D 为顶点的等腰三角形（isosceles triangle）。每个三角形的面积都是 $\frac{1}{2}r^2\sin(\text{顶角})$，而三个顶角之和为 π。这样，一道几何题就变成求几个正弦之和的最大值，而第一部分只需用和化积公式（sum-to-product formula）便可完成，完全不必动用微积分。

解答

窗帘的面积

由于 A、D、C 共线（collinear），在 D 处的各角满足

$$\angle ADX + \angle XDY + \angle YDC = \pi, \qquad \text{所以} \qquad \angle YDC = \pi - \theta - \alpha.$$

这里 $\theta \geq 0$、$\alpha > 0$ 且 $\theta + \alpha \leq \pi$，使得 X 和 Y 在弧上按 A, X, Y, C 的次序排列。

半径 DX 和 DY 把窗帘 $AXYC$ 分成三角形 ADX、XDY 和 YDC，每个三角形都有两条长度为 r 的边相交于 D。对每个三角形使用面积 $= \frac{1}{2}ab\sin C$，得

$$\begin{aligned} \text{面积} &= \tfrac{1}{2}r^2\sin\theta + \tfrac{1}{2}r^2\sin\alpha + \tfrac{1}{2}r^2\sin(\pi - \theta - \alpha) \\ &= \tfrac{1}{2}r^2[\sin\alpha + \sin\theta + \sin(\theta + \alpha)], \end{aligned}$$

因为 $\sin(\pi - u) = \sin u$。

固定 α 时的最佳 θ

当 α 固定时，只有 $\sin\theta + \sin(\theta+\alpha)$ 会变动。根据和化积公式 $\sin P + \sin Q = 2\sin\frac{P+Q}{2}\cos\frac{P-Q}{2}$，

$$\sin\theta + \sin(\theta+\alpha) = 2\cos\tfrac{\alpha}{2}\,\sin\left(\theta+\tfrac{\alpha}{2}\right).$$

> **想一想** 为甚么这样一写，最大值就一目了然?

因子 $2\cos\frac{\alpha}{2}$ 是一个固定的正常数，因为 $0 < \frac{\alpha}{2} < \frac{\pi}{2}$。所以面积最大，恰好就是 $\sin(\theta+\frac{\alpha}{2})$ 最大的时候。当 θ 取遍 $[0, \pi-\alpha]$ 时，角 $\theta+\frac{\alpha}{2}$ 取遍 $[\frac{\alpha}{2}, \pi-\frac{\alpha}{2}]$，而这个区间包含 $\frac{\pi}{2}$。因此正弦可以达到它的最大可能值 1，此时 $\theta+\frac{\alpha}{2} = \frac{\pi}{2}$，即

$$\boxed{\theta = \frac{\pi}{2} - \frac{\alpha}{2} = \frac{\pi-\alpha}{2}.}$$

这时同样有 $\angle YDC = \pi - \theta - \alpha = \frac{\pi-\alpha}{2} = \theta$：最佳的窗帘关于垂直的半径对称，与直觉相符。对这个 α 而言，最大面积是

$$\tfrac{1}{2}r^2\left[\sin\alpha + 2\cos\tfrac{\alpha}{2}\right].$$

让 α 变动

现在我们要在 $0 < \alpha < \pi$ 的范围内求 $f(\alpha) = \sin\alpha + 2\cos\frac{\alpha}{2}$ 的最大值。（对每个 α，最大面积是 $\frac{1}{2}r^2 f(\alpha)$，所以整体的最大面积就是 $\frac{1}{2}r^2$ 乘以 f 的最大值。）求导数（differentiate），得

$$f'(\alpha) = \cos\alpha - \sin\tfrac{\alpha}{2}.$$

设 $s = \sin\frac{\alpha}{2}$，于是 $0 < s < 1$，再利用 $\cos\alpha = 1 - 2\sin^2\frac{\alpha}{2}$：

$$f'(\alpha) = 1 - s - 2s^2 = (1-2s)(1+s).$$

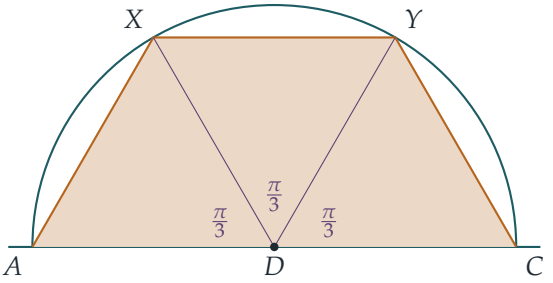
由于 $1+s > 0$，f' 的正负号与 $1-2s$ 相同。因为 s 在 $(0,\pi)$ 上随 α 递增，f' 在 $s < \frac{1}{2}$ 时为正，在 $s > \frac{1}{2}$ 时为负。所以 f 先递增后递减，并在 $\sin\frac{\alpha}{2} = \frac{1}{2}$ 处取得最大值，即 $\alpha = \frac{\pi}{3}$。在该处

$$f\left(\tfrac{\pi}{3}\right) = \sin\tfrac{\pi}{3} + 2\cos\tfrac{\pi}{6} = \tfrac{\sqrt{3}}{2} + \sqrt{3} = \tfrac{3\sqrt{3}}{2},$$

所以窗帘的最大可能面积是

$$\boxed{\tfrac{1}{2}r^2 \cdot \tfrac{3\sqrt{3}}{2} = \frac{3\sqrt{3}}{4}r^2.}$$

此时 $\alpha = \theta = \angle YDC = \frac{\pi}{3}$，所以三个三角形都是等边三角形，而窗帘正是内接（inscribed）于整个圆的正六边形（regular hexagon）的一半。



验算

- 极限情况：当 $\alpha \to 0$ 时，窗帘变成三角形 AXC，其中 X 位于最高点，面积为 $\frac{1}{2} \cdot 2r \cdot r = r^2$，而确实有 $\frac{1}{2} r^2 f(0) = r^2$。当 $\alpha \to \pi$ 时，窗帘塌缩到直径上，而 $f(\pi) = 0$。最大值 $\frac{3\sqrt{3}}{4} r^2 \approx 1.299 r^2$ 比两者都大，同时小于半圆的面积 $\frac{\pi}{2} r^2 \approx 1.571 r^2$，这是必然的。
- 数值验证：我们随机选取 X 和 Y 的位置，把三个三角形的面积公式与鞋带公式（shoelace formula）互相核对；又对若干个 α 的值，用细密的网格搜索找出最佳的 θ；并发现取 $r = 1$ 时，对 θ 和 α 同时搜索所得的最大面积为 $1.299038\ldots = \frac{3\sqrt{3}}{4}$。

常见错误

- 遗漏第三个三角形 YDC，或者把它的角写成 $\pi - \alpha$ 而不是 $\pi - \theta - \alpha$。
- 在第二部分直接取 $\theta = \alpha = \frac{\pi}{3}$ 这条捷径，却没有给出理由。必须用微积分（或对称性的论证）才能证明最佳的窗帘确实是正六边形的一半。

STEP 之外

这个答案是一个经典结果的特殊情况：在内接于同一个圆的所有 n 边形之中，以正多边形的面积最大。有一个简洁的证明，用到 sin 在 $[0, \pi]$ 上的凹性（concavity）和琴生不等式（Jensen's inequality）：若圆心角（central angle）$\phi_1, \ldots, \phi_n$ 之和为 2π，则 $\sum \sin \phi_i \le n \sin \frac{2\pi}{n}$。我们的窗帘就是 $n = 6$ 这个情况的一半。凸性（convexity）和这类不等式是大学一年级分析（analysis）课程中贯穿始终的主题；Abbott 的 *Understanding Analysis* 是正式认识这些概念的好起点，而 Siklos 的 *Advanced Problems in Mathematics* 则收录了更多同类风格的 STEP 题目。

详细解答

STEP I 1990，第 2 题

纯数学 · 单位根筛选法

> **STEP I 1990，第 2 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-s1-q2

本题在 *Beyond STEP* 第 5.2 节单位根筛选法中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/roots-filter.html

关键思路

单位立方根（cube roots of unity）在这里起着筛选器的作用。当我们用二项式定理（binomial theorem）展开 $(1+\omega^r)^n$，再把 $r=1,2,3$ 的结果相加时，每个二项式系数 $\binom{n}{j}$ 都会乘上 $1+\omega^j+\omega^{2j}$；当 3 整除 j 时这个数是 3，否则是 0。因此这个和刚好只保留我们想要的系数。计算每个 $(1+\omega^r)^n$ 的另一个方法，是把 $1+\omega^r$ 写成模—辐角形式（modulus–argument form），再用棣美弗定理（de Moivre's theorem）。

解答

单位立方根

由于 $\omega=\mathrm{e}^{2\pi\mathrm{i}/3}$，我们有 $\omega^3=\mathrm{e}^{2\pi\mathrm{i}}=1$，但 $\omega\neq 1$。于是

$$0=\omega^3-1=(\omega-1)(1+\omega+\omega^2),$$

而因为 $\omega-1\neq 0$，可以把它约去，得到 $1+\omega+\omega^2=0$。

由此可知 $1+\omega^2=-\omega$。乘以 $-1=\mathrm{e}^{-\mathrm{i}\pi}$ 相当于旋转 $-\pi$，所以

$$1+\omega^2=\mathrm{e}^{-\mathrm{i}\pi}\,\mathrm{e}^{2\pi\mathrm{i}/3}=\mathrm{e}^{-\mathrm{i}\pi/3}.$$

因此 $1+\omega^2$ 的模是 1，辐角是 $-\frac{\pi}{3}$。（验算一下：$1+\omega^2=1+\left(-\frac{1}{2}-\frac{\sqrt{3}}{2}\mathrm{i}\right)=\frac{1}{2}-\frac{\sqrt{3}}{2}\mathrm{i}$。）同理，$1+\omega=-\omega^2=\mathrm{e}^{\mathrm{i}\pi/3}$，即它的共轭复数（complex conjugate）。

恒等式

*方法一：二项式定理。*对 $r = 1, 2, 3$,

$$(1+\omega^r)^n = \sum_{j=0}^{n} \binom{n}{j} \omega^{rj}.$$

把三种情况相加,

$$(1+\omega)^n + (1+\omega^2)^n + (1+\omega^3)^n = \sum_{j=0}^{n} \binom{n}{j} (\omega^j + \omega^{2j} + \omega^{3j}).$$

想一想 对一般的整数 j，$\omega^j + \omega^{2j} + \omega^{3j}$ 等于甚么?

由于 $\omega^{3j} = 1$，括号内的式子就是 $1 + \omega^j + \omega^{2j}$；又因为 $\omega^3 = 1$，它只取决于 j 模 3 的值。

- 若 j 是 3 的倍数，则 $\omega^j = 1$，括号内的值是 $1 + 1 + 1 = 3$。
- 若 $j \equiv 1 \pmod 3$，则 $\omega^j = \omega$ 及 $\omega^{2j} = \omega^2$，所以括号内的值是 $1 + \omega + \omega^2 = 0$。
- 若 $j \equiv 2 \pmod 3$，则 $\omega^j = \omega^2$ 及 $\omega^{2j} = \omega^4 = \omega$，所以括号内的值同样是 0。

因此只有 $3 \mid j$ 的项留下来，而

$$(1+\omega)^n + (1+\omega^2)^n + (1+\omega^3)^n = 3\left[\binom{n}{0} + \binom{n}{3} + \cdots + \binom{n}{k}\right],$$

其中 k 是不超过 n 的最大的 3 的倍数。

*方法二：模与辐角。*由第一部分，$1 + \omega = \mathrm{e}^{\mathrm{i}\pi/3}$ 及 $1 + \omega^2 = \mathrm{e}^{-\mathrm{i}\pi/3}$，而 $1 + \omega^3 = 2$。根据棣美弗定理,

$$\begin{aligned}(1+\omega)^n + (1+\omega^2)^n + (1+\omega^3)^n &= \mathrm{e}^{\mathrm{i}n\pi/3} + \mathrm{e}^{-\mathrm{i}n\pi/3} + 2^n \\ &= 2^n + 2\cos\left(\frac{n\pi}{3}\right).\end{aligned}$$

比较两种算法的结果，再除以 3,

$$\binom{n}{0} + \binom{n}{3} + \binom{n}{6} + \cdots + \binom{n}{k} = \frac{1}{3}\left(2^n + 2\cos\left(\frac{n\pi}{3}\right)\right),$$

这就是所要证明的结果。

$n = 25$ 的情况

不超过 25 的最大的 3 的倍数是 24，所以第一个和正正就是恒等式取 $n = 25$ 的情况。现在 $\frac{25\pi}{3} = 8\pi + \frac{\pi}{3}$，所以 $\cos\frac{25\pi}{3} = \cos\frac{\pi}{3} = \frac{1}{2}$，而

$$\binom{25}{0} + \binom{25}{3} + \cdots + \binom{25}{24} = \frac{1}{3}\left(2^{25} + 1\right) = \frac{33554433}{3} = 11184811.$$

第二个和

这里 $n=24$，而我们要的是 $j\equiv 2 \pmod 3$ 的系数 $\binom{24}{j}$。把所有系数 $\binom{24}{j}$ 按 j 除以 3 的余数分组：

$$S_0+S_1+S_2=\sum_{j=0}^{24}\binom{24}{j}=2^{24},$$

其中 S_m 是所有 $j\equiv m \pmod 3$ 的 $\binom{24}{j}$ 之和。

对称性 $\binom{24}{j}=\binom{24}{24-j}$ 把这些项配成对：若 $j\equiv 2$，则 $24-j\equiv -2\equiv 1 \pmod 3$，因为 24 是 3 的倍数。所以 $j\mapsto 24-j$ 把 S_2 的各项与 S_1 的各项一一对应，于是 $S_1=S_2$。

取 $n=24$ 的恒等式可给出 S_0：由于 $\cos 8\pi=1$，

$$S_0=\frac{1}{3}\left(2^{24}+2\right).$$

因此

$$S_2=\frac{1}{2}\left(2^{24}-S_0\right)=\frac{1}{2}\cdot\frac{2\cdot 2^{24}-2}{3}=\frac{2^{24}-1}{3}.$$

把题目给出的数值减半，得 $2^{24}=16777216$，所以

$$\binom{24}{2}+\binom{24}{5}+\cdots+\binom{24}{23}=\frac{16777215}{3}=5592405.$$

验算

- 小数值的情况：当 $n=3$ 时，左边是 $1+1=2$，右边是 $\frac{1}{3}(8+2\cos\pi)=2$；当 $n=4$ 时，左边是 $1+4=5$，右边是 $\frac{1}{3}\left(16+2\cos\frac{4\pi}{3}\right)=\frac{1}{3}(16-1)=5$。
- 整除性：$2^{25}+1$ 和 $2^{24}-1$ 都必须是 3 的倍数；事实确是如此，因为 $2\equiv -1 \pmod 3$。
- 我们用电脑直接把二项式系数相加：恒等式对 199 或以下的每个 n 都成立，第一个和是 11184811，第二个和是 5592405。

常见错误

- 把 $1+\omega^2$ 的辐角写成 $\frac{\pi}{3}$：它位于第四象限，所以辐角主值（principal argument）是 $-\frac{\pi}{3}$（等价地，$\frac{5\pi}{3}$）。
- 忘记了 $r=3$ 的情况，而 2^n 正是由它而来。
- 在第二个和中，试图直接套用公式：公式只给出 3 的倍数的项，所以要处理余数为 2 的项，还需要另一个想法（这里是对称性）。

STEP 之外

这个技巧称为单位根筛选法（roots of unity filter），它是离散傅里叶变换（discrete Fourier transform）的第一个例子：把 $f(\omega^r)$ 在所有 m 次单位根上取平均，就能抽出多项式 f 中属于模 m 某个指定剩余类（residue class）的系数。同样这种在一个对称群上取平均的做法，会在有限群的特征标理论（character theory）以及伯恩赛德引理（Burnside's lemma）等计数论证中再次出现。想进一步了解为何复数令这类计算变得如此自然，可参阅 Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）。

详细解答

STEP I 1990，第 3 题

纯数学 · 法线与平行曲线

> **STEP I 1990，第 3 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-s1-q3

关键思路

只要把向下的单位法向量（unit normal）明确写出来，从而以 x 表示 $\mathbf{s}$ 的坐标，其余一切便水到渠成。第 (i) 部分只需迫使 $\mathbf{s}$ 的 y 坐标等于零。在第 (ii) 部分，题目建议的代换（substitution）令 $\mathbf{s}$ 的坐标变得非常简单，而斜率（gradient）条件则化为一条一阶微分方程（first-order differential equation），这条方程原来恰好是一个全导数（exact derivative）。

解答

准备工作。曲线 $y = \mathrm{f}(x)$ 的切线（tangent）方向为 $(1, \mathrm{f}'(x))$，所以法向量为 $\pm(\mathrm{f}'(x), -1)$。向下的那一个的 y 分量为负，故

$$\mathbf{n} = \frac{1}{N}\begin{pmatrix} \mathrm{f}'(x) \\ -1 \end{pmatrix}, \qquad N = \sqrt{1 + \mathrm{f}'(x)^2},$$

因此

$$\mathbf{s} = \mathbf{r} + \mathrm{p}(x)\mathbf{n} = \begin{pmatrix} x + \dfrac{\mathrm{p}(x)\,\mathrm{f}'(x)}{N} \\ \mathrm{f}(x) - \dfrac{\mathrm{p}(x)}{N} \end{pmatrix}.$$

另外留意，由于 $\mathbf{n}$ 是单位向量，所以 $|\mathbf{r} - \mathbf{s}| = |\mathrm{p}(x)|$。

第 (i) 部分

这里 $\mathrm{f}(x)=x^k$，$\mathrm{f}'(x)=kx^{k-1}$，而 $N=\sqrt{1+k^2x^{2k-2}}$。要使 $\mathbf{s}$ 位于 x 轴上，它的 y 坐标必须为零：

$$x^k-\frac{\mathrm{p}(x)}{N}=0 \quad\Longrightarrow\quad \mathrm{p}(x)=x^k\sqrt{1+k^2x^{2k-2}}.$$

这样选取之后，$\mathbf{s}$ 的 x 坐标为

$$x+x^k\cdot kx^{k-1}=x+kx^{2k-1},$$

它是连续（continuous）而且递增的，当 $x\to 0^+$ 时趋向 0，当 $x\to\infty$ 时趋向无穷大。所以 $\mathbf{s}$ 确实扫过整条正 x 轴，而不只是其中一部分。

现在 $|\mathbf{r}|^2=x^2+x^{2k}$，而 $|\mathbf{r}-\mathbf{s}|^2=\mathrm{p}(x)^2=x^{2k}+k^2x^{4k-2}$（这里 $\mathrm{p}>0$）。所以对每一个 $x>0$ 都有 $|\mathbf{r}|=|\mathbf{r}-\mathbf{s}|$，当且仅当

$$x^2=k^2x^{4k-2}\quad \text{对所有}x>0.$$

取 $x=1$ 便必有 $k^2=1$，所以 $k=1$；反过来，$k=1$ 时两边都等于 x^2。因此 $k=1$。

从几何角度看：在直线 $y=x$ 上，由 (x,x) 出发的向下法线与 x 轴相交于 $(2x,0)$，而以 O、(x,x)、$(2x,0)$ 为顶点的三角形是等腰三角形（isosceles triangle），两条等边的长度均为 $\sqrt{2}\,x$。

第 (ii) 部分

现在 $\mathrm{f}(x)=x^2$，$\mathrm{f}'(x)=2x$，而 $N=\sqrt{1+4x^2}$。写成 $\mathrm{p}=\mathrm{q}N$ 便可消去那些碍事的平方根：

$$\mathbf{s}=\begin{pmatrix}X\\Y\end{pmatrix}=\begin{pmatrix}x+2x\,\mathrm{q}(x)\\x^2-\mathrm{q}(x)\end{pmatrix}.$$

新曲线是以 x 为参数（parameter）给出的，所以它在 $\mathbf{s}$ 的斜率为

$$\frac{\mathrm{d}Y}{\mathrm{d}X}=\frac{\mathrm{d}Y/\mathrm{d}x}{\mathrm{d}X/\mathrm{d}x}=\frac{2x-\mathrm{q}'}{1+2\mathrm{q}+2x\mathrm{q}'}.$$

令它等于 $2x$，即 $y=x^2$ 在 $\mathbf{r}$ 的斜率，然后两边乘以分母：

$$\begin{aligned}&2x-\mathrm{q}'=2x+4x\mathrm{q}+4x^2\mathrm{q}'\\ \Longleftrightarrow\quad &(1+4x^2)\,\mathrm{q}'+4x\,\mathrm{q}=0.\end{aligned}$$

> ? **想一想** 我们能否不除以 q（它可能在某处为零）而解出这条方程?

可以：改为除以 $N=\sqrt{1+4x^2}$，它永不为零。由于 $N'=4x/N$，

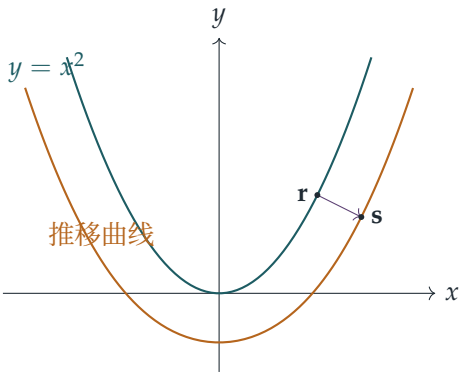
$$\frac{(1+4x^2)\,\mathrm{q}'+4x\,\mathrm{q}}{N}=N\mathrm{q}'+\frac{4x}{N}\,\mathrm{q}=\frac{\mathrm{d}}{\mathrm{d}x}(N\mathrm{q})=\mathrm{p}'(x).$$

所以条件其实就是对所有 x 都有 $\mathrm{p}'(x)=0$，因此

$$\mathrm{p}(x)=c,\qquad \text{为一常数}$$

（等价地，$\mathrm{q}(x)=c/\sqrt{1+4x^2}$）。反过来，每一个常数 c 都可行，因为只要新曲线的斜率有定义，上述每一步都是可逆的。

新曲线是一条*平行曲线*（parallel curve）：抛物线（parabola）上每一点都沿着该点的法线被推移一段固定距离 c。下图所示为 $c=\frac{1}{2}$ 的情况。



有一处细微之处：$\mathrm{d}X/\mathrm{d}x=1+2\mathrm{q}+2x\mathrm{q}'$，代入 $\mathrm{q}=c/N$ 后可化简为 $1+2c/N^3$。当 $c>-\frac{1}{2}$ 时，它恒为正数，所以推移曲线是一条光滑曲线。若 $c\le-\frac{1}{2}$（即向上推移至少 $\frac{1}{2}$，亦即顶点处的曲率半径（radius of curvature）），它会在 $N^3=-2c$ 之处变为零，推移曲线便在该处出现尖点（cusp）。

验算

- 在 (i) 中取 $k=2$ 及 $x=1$：$\mathrm{p}(1)=\sqrt{5}$，而 $\mathbf{s}=(1,1)+\sqrt{5}\cdot\frac{1}{\sqrt{5}}(2,-1)=(3,0)$，一如所求位于 x 轴上；但 $|\mathbf{r}|=\sqrt{2}\neq\sqrt{5}$，所以 $k=2$ 不符合要求。
- 在 (ii) 中，当 $x=0$ 时，常数为 c 的推移曲线通过 $(0,-c)$，而根据对称性，它在该处的斜率为 0，与抛物线顶点的斜率吻合。
- 以数值验证，取 $c=1$ 及 $x=0.8$，用有限差分（finite difference）计算推移曲线的斜率，结果与 $2x=1.6$ 相符；若改用非常数的 $\mathrm{p}=x^2$，得出的斜率则约为 0.41。

常见错误

- 误用向上的法向量 $(-\mathrm{f}',1)/N$：$\mathbf{s}$ 中的正负号会全部相反。

- 把新曲线的斜率写成 $\mathrm{d}Y/\mathrm{d}x$，而不是 $(\mathrm{d}Y/\mathrm{d}x)/(\mathrm{d}X/\mathrm{d}x)$；新曲线的 x 坐标是 X，而不是 x。
- 在 (i) 中，只比较指数（$4k-2=2$）便断定 $k=1$，而没有同时检查系数 $k^2=1$。

STEP 之外

以固定距离推移而成的曲线称为*平行曲线*或*偏移曲线*（offset curve），它们在实际应用中十分重要：铣床（milling machine）刀具的中心，正是沿着所切削形状的平行曲线移动。当偏移距离超过曲率半径时出现的尖点，都位于*渐屈线*（evolute）之上，即曲率中心的轨迹（locus），它是曲线微分几何（differential geometry）中的核心对象。Pressley 的 *Elementary Differential Geometry*（初等微分几何）是一本很好的入门书。

详细解答

STEP I 1990，第 4 题

纯数学 · 等差数列与差分

STEP I 1990，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/90-s1-q4

关键思路

数列的每一项，都等于首项加上各个差的累计总和。等差数列（arithmetic progression）的差是常数，所以求和很容易。至于双重等差数列（double arithmetic progression），各个差本身又构成一个等差数列，于是要多做一层求和，二项式系数（binomial coefficient）$\binom{n}{2}$ 和 $\binom{n}{3}$ 便自然出现。最后一部分则化为关于两个未知差的两条线性方程。

解答

等差数列

设公差（common difference）为 $d = a_2 - a_1$，则 $a_k = a_1 + (k-1)d$。于是

$$\sum_{k=1}^{n} a_k = na_1 + d\sum_{k=1}^{n}(k-1) = na_1 + \frac{n(n-1)}{2}(a_2 - a_1),$$

其中用到 $0 + 1 + \cdots + (n-1) = \frac{1}{2}n(n-1)$。把 a_1 和 a_2 的项分别合并，得

$$a_1 + a_2 + \cdots + a_n = \frac{n}{2}\big[(3-n)a_1 + (n-1)a_2\big].$$

双重等差数列

设各个差为 $\delta_k = b_{k+1} - b_k$。它们构成一个等差数列，首项为 $p = \delta_1 = b_2 - b_1$，公差为

$$e = \delta_2 - \delta_1 = (b_3 - b_2) - (b_2 - b_1) = b_3 - 2b_2 + b_1.$$

因此 $\delta_j = p + (j-1)e$，把首 $k-1$ 个差加起来（即第一部分的公式，把 n 换成 $k-1$），得

$$b_k = b_1 + \sum_{j=1}^{k-1} \delta_j = b_1 + (k-1)p + \frac{(k-1)(k-2)}{2}\, e = b_1 + \binom{k-1}{1} p + \binom{k-1}{2} e.$$

现在对 $k = 1, \ldots, n$ 求和。我们需要两个和。第一个是 $\sum_{k=1}^{n}(k-1) = \binom{n}{2}$，与前面相同。至于第二个，令 $j = k-1$ 由 0 取到 $n-1$，则

$$\begin{aligned}\sum_{j=0}^{n-1} \frac{j(j-1)}{2} &= \frac{1}{2}\left[\frac{(n-1)n(2n-1)}{6} - \frac{(n-1)n}{2}\right] = \frac{n(n-1)}{12}\left[(2n-1)-3\right] \\ &= \frac{n(n-1)(n-2)}{6} = \binom{n}{3}.\end{aligned}$$

所以

$$\sum_{k=1}^{n} b_k = nb_1 + \binom{n}{2}(b_2 - b_1) + \binom{n}{3}(b_3 - 2b_2 + b_1).$$

这已经是一条以 n、b_1、b_2、b_3 表示的公式。如果你想分别看到每个 b_i 的系数，展开后得

$$b_1 + \cdots + b_n = \frac{n(n^2 - 6n + 11)}{6}\, b_1 + \frac{n(n-1)(7-2n)}{6}\, b_2 + \frac{n(n-1)(n-2)}{6}\, b_3.$$

（快速检验：$n = 1, 2, 3$ 分别给出 b_1、$b_1 + b_2$ 和 $b_1 + b_2 + b_3$。）

> **想一想** 有没有更简洁的方法看出 $\binom{n}{3}$ 从何而来?

有：$\binom{k-1}{2}$ 是从 $k-1$ 件东西中选 2 件的方法数目，而曲棍球棒恒等式（hockey-stick identity）$\sum_{k=1}^{n}\binom{k-1}{2} = \binom{n}{3}$（从 n 件中选 3 件时，先固定最大的一件 k）一步便完成求和。每多一层“差的差”，二项式系数便升高一级，正如积分把幂次升高一次。

阶乘数列

现在 $b_1 = 1$，而 p 和 e 如上所定，故 $b_k = 1 + (k-1)p + \frac{1}{2}(k-1)(k-2)e$。所以

$$b_2 = 1 + p, \qquad b_4 = 1 + 3p + 3e, \qquad b_6 = 1 + 5p + 10e.$$

在一个 $c_1 = b_2$、$c_2 = b_4$、$c_3 = b_6$、$c_4 = 220$ 的阶乘数列（factorial progression）中，相邻两项之差依次为 $1!\,d$、$2!\,d$、$3!\,d$，即 d、$2d$、$6d$：

$$\begin{aligned} b_4 - b_2 &= 2p + 3e = d, \\ b_6 - b_4 &= 2p + 7e = 2d, \\ 220 - b_6 &= 219 - 5p - 10e = 6d. \end{aligned}$$

把第一式代入第二式，得 $2p + 7e = 4p + 6e$，所以 $e = 2p$，继而 $d = 8p$。第三条方程变成

$$219 - 5p - 20p = 48p \quad \Longrightarrow \quad 73p = 219 \quad \Longrightarrow \quad p = 3,$$

所以 $e=6$ 及 $d=24$，符合非零的要求。由于每一步都是必然的，这个解是唯一的。

于是 $b_k = 1+3(k-1)+3(k-1)(k-2) = 3k^2-6k+4$，得出数列 $1,4,13,28,49,76,\ldots$，而 $b_2, b_4, b_6, 220 = 4, 28, 76, 220$ 的差确实是 $24, 48, 144 = 1!\cdot 24,\ 2!\cdot 24,\ 3!\cdot 24$。

最后，把 $b_1=1$、$b_2-b_1=3$ 和 $b_3-2b_2+b_1=6$ 代入求和公式：

$$\begin{aligned} 1+b_2+\cdots+b_n &= n+3\cdot\frac{n(n-1)}{2}+6\cdot\frac{n(n-1)(n-2)}{6} \\ &= \frac{n}{2}\left[2+3(n-1)+2(n-1)(n-2)\right] \\ &= \frac{n(2n^2-3n+3)}{2}. \end{aligned}$$

验算

当 $n=1,2,3$ 时，最后的公式给出 1、5 和 18，与 1、$1+4$ 和 $1+4+13$ 吻合。当 $n=6$ 时，公式给出 $3\cdot 57 = 171$，而直接计算亦得 $1+4+13+28+49+76=171$。此外，我们也用电脑以分数作精确运算，对数百个随机数列及所有 $n\le 24$ 验证了两条一般求和公式。

常见错误

- 把阶乘数列的差当作 $d, 2d, 3d$：它们其实是 $1!\,d$、$2!\,d$、$3!\,d$，所以第三个是 $6d$。
- 写成 $b_k = b_1 + kp + \ldots$：b_1 与 b_k 之间只有 $k-1$ 个差。

STEP 之外

这就是有限差分学（calculus of finite differences）：若一个数列的第 m 阶差是常数，它便是 n 的 m 次多项式，而它的部分和则是 $m+1$ 次多项式，最宜用 $\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots$ 这组基来表示（牛顿（Newton）前向差分公式）。把 $\binom{n}{r}$ 求和而得 $\binom{n+1}{r+1}$，正是 $\int x^r\,\mathrm{d}x = \frac{x^{r+1}}{r+1}$ 的离散版本。想多做同类题目，可试试 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）中的问题。

详细解答

STEP I 1990，第 5 题

纯数学 · 部分分式与曲线描绘

STEP I 1990，第 5 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-s1-q5

关键思路

在 (i) 中，二次式可以因式分解，所以用部分分式（partial fractions）便可把积分化为两个对数之差。在 (ii) 中，巨大的幂次看似吓人，但 $1760 = 8 \times 220$，所以这个函数其实就是 $u^8 - u + q$，其中 $u = x^{220} \geq 0$。图像的形状由驻点（stationary points）决定，而根的数目只取决于横轴相对于局部极大值（local maximum）和两个极小值的位置。

解答

第 (i) 部分

由于 $6 \times 15 = 90 = 9 \times 10$ 而 $9 + 10 = 19$，这个二次式可分解为：

$$6x^2 + 19x + 15 = (2x + 3)(3x + 5).$$

设

$$\frac{1}{(2x+3)(3x+5)} = \frac{A}{2x+3} + \frac{B}{3x+5}, \qquad \text{所以} \qquad 1 = A(3x+5) + B(2x+3).$$

代入 $x = -\frac{3}{2}$ 得 $\frac{1}{2}A = 1$，所以 $A = 2$；代入 $x = -\frac{5}{3}$ 得 $-\frac{1}{3}B = 1$，所以 $B = -3$。$2x+3$ 和 $3x+5$ 在 $[1,3]$ 上都是正数，所以无须加上绝对值符号：

$$\int_1^3 \frac{\mathrm{d}x}{6x^2+19x+15} = \int_1^3 \left(\frac{2}{2x+3} - \frac{3}{3x+5}\right)\mathrm{d}x = \Big[\ln(2x+3) - \ln(3x+5)\Big]_1^3$$
$$= \ln\frac{9}{14} - \ln\frac{5}{8} = \ln\frac{72}{70} = \ln\frac{36}{35}.$$

第 (ii) 部分

*对称性与两端的变化。*两个幂次都是偶数，所以 f 是偶函数（even function），其图像对称于 y 轴。当 $|x|$ 很大时，x^{1760} 一项占主导，所以当 $x \to \pm\infty$ 时 $\mathrm{f}(x) \to +\infty$。此外，$\mathrm{f}(0) = q$ 而 $\mathrm{f}(\pm 1) = q$。

*驻点。*求导得

$$\mathrm{f}'(x) = 1760x^{1759} - 220x^{219} = 220\,x^{219}\left(8x^{1540} - 1\right).$$

它在 $x = 0$ 处以及 $x^{1540} = \frac{1}{8}$ 之处为零，后者即 $x = \pm a$，其中

$$a = 8^{-1/1540} \approx 0.9987.$$

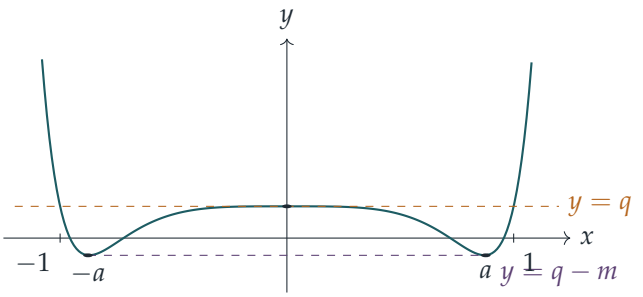
当 $x > 0$ 时，$\mathrm{f}'(x)$ 的正负号与 $8x^{1540} - 1$ 相同：在 $(0, a)$ 上为负，在 $x > a$ 时为正。由于 x^{219} 是奇数次幂，当 $x < 0$ 时正负号刚好相反。所以 f 在 $(-\infty, -a)$ 上递减，在 $(-a, 0)$ 上递增，在 $(0, a)$ 上递减，在 (a, ∞) 上递增：在 $x = 0$ 处有局部极大值 q，在 $x = \pm a$ 处有两个相等的极小值。

> **想一想** 怎样才能求出极小值，而不必跟 $8^{-1/1540}$ 纠缠?

改用 $u = x^{220}$ 来计算。在 $x = \pm a$ 处有 $x^{1540} = u^7 = \frac{1}{8}$，所以 $u = 8^{-1/7}$，而 $u^8 = u \cdot u^7 = \frac{1}{8}u$。因此

$$\mathrm{f}(\pm a) = u^8 - u + q = q - \tfrac{7}{8}u = q - m, \qquad \text{其中} \qquad m = \tfrac{7}{8} \cdot 8^{-1/7} = 7 \cdot 2^{-24/7} \approx 0.650.$$

*草图。*图像呈“W”形：它从高处开始，下降至 $x = -a$ 处的极小值 $q - m$，上升至原点处的局部极大值 q，再下降至 $x = a$ 处的极小值 $q - m$，然后急剧上升。由于幂次极大，真实的图像极为平坦：除非 $|x|$ 非常接近 1，否则 x^{220} 极小，所以在 $(-1, 1)$ 的大部分范围内 $\mathrm{f}(x) \approx q$，只在 $x = \pm 1$（图像在该处回到高度 q）内侧不远处有两个狭窄的凹陷，而在外侧则是近乎垂直的陡壁。下图把凹陷夸大了，好让各项特征清楚可见。



计数根的数目。$\mathrm{f}(x) = 0$ 的根就是图像与 x 轴相交之处，所以我们把 0 与局部极大值 q 和极小值 $q - m$ 作比较。

- $q > m$：极小值 $q - m$ 为正，所以 $\mathrm{f} > 0$ 处处成立：**没有根**。
- $q = m$：图像只在极小值处与横轴相切：**两个根**，$x = \pm a$。
- $0 < q < m$：此时 $q - m < 0 < q$。在四个单调区间 $(-\infty, -a)$、$(-a, 0)$、$(0, a)$、(a, ∞) 的每一个上，f 都是连续（continuous）而且严格单调（strictly monotonic）的，并且由正值变为负值，或由负值变为正值，所以在每个区间上恰有一个根：**四个根**。
- $q = 0$：$\mathrm{f}(x) = x^{220}(x^{1540} - 1)$，它恰好在 $x = 0$ 和 $x = \pm 1$ 处为零：**三个根**。
- $q < 0$：局部极大值 $\mathrm{f}(0) = q$ 为负，所以在整个 $[-a, a]$ 上 $\mathrm{f} < 0$，而在 (a, ∞) 和 $(-\infty, -a)$ 的每一个上，f 都严格单调地由负值变至 $+\infty$：**两个根**。

所以相异根的数目可能是 0、2、3 和 4，其中

$$\begin{aligned} &0 \text{ 个根：} \quad q > m, \\ &2 \text{ 个根：} \quad q < 0 \text{ 或 } q = m, \\ &3 \text{ 个根：} \quad q = 0, \\ &4 \text{ 个根：} \quad 0 < q < m, \end{aligned} \qquad m = \tfrac{7}{8} \cdot 8^{-1/7} = 7 \cdot 2^{-24/7} \approx 0.650.$$

验算

对 (i) 的被积函数使用辛普森法则（Simpson's rule），得 0.0281709，与 $\ln \frac{36}{35} = 0.0281709$ 吻合。至于 (ii)，在 $x = 1$ 附近以细密的数值搜索求 $x^{1760} - x^{220}$ 的极小值，得出在 $x \approx 0.99865$ 处的 -0.650123，与 $-m$ 和 a 一致。取 $q = 1,\ 0.5,\ 0.1,\ 0.001,\ 0,\ -0.01,\ -5$，在细密的格点上计数 f 的变号次数，得出根的数目为 $0, 4, 4, 4, 3, 2, 2$，与上表的预测相符。$q = 0$ 的情况正好用来检验草图是否合理：图像必须通过原点和 $(\pm 1, 0)$。

常见错误

- 遗漏边界情况 $q = 0$（三个根，因为局部极大值正好落在横轴上）和 $q = m$（在极小值处有两个重根）。
- 当 $q < 0$ 时，“凭对称性”声称有两个根，却没有检查 $-a$ 与 a 之间是否别无变化：正是因为局部极大值 q 为负，才排除了该处有额外的根。
- 把草图画成一条弯度平缓的“W”。幂次这么大，图像在 $|x| < 1$ 时几乎平坦地停留在高度 q；即使草图夸大了凹陷，也值得把这一点说明。

STEP 之外

通过确定驻点处的函数值来计数根的数目，让人初尝介值定理（intermediate value theorem）和单调性论证的滋味，而这些内容会在大学第一门分析课程中严格建立起来；代换 $u = x^{220}$ 也显示出 x^n 在 $n \to \infty$ 时的表现：它在 $(-1, 1)$ 上收敛至 0，但并非一致收敛（uniform convergence）。Abbott 的 *Understanding Analysis*（《理解分析》）对连续性和一致收敛的处理既严谨又易读。

详细解答

STEP I 1990，第 6 题

纯数学 · 向量与平行四边形

STEP I 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-s1-q6

关键思路

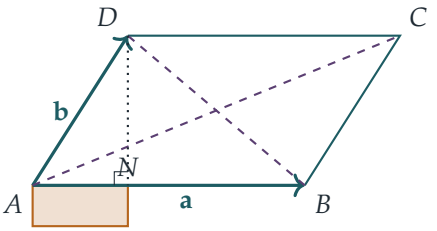
把原点设在 A，并用两条边的向量（vector）$\mathbf{a} = \overrightarrow{AB}$ 和 $\mathbf{b} = \overrightarrow{AD}$ 来描述一切。这样，两条对角线（diagonal）就是 $\mathbf{a}+\mathbf{b}$ 和 $\mathbf{b}-\mathbf{a}$，而每个长度的平方都是一个向量与自身的点积（dot product）。把这些点积展开，整道题就解决了。唯一用到的几何事实是：$|\mathbf{u}\cdot\mathbf{v}|$ 等于 $\mathbf{u}$ 的长度乘以 $\mathbf{v}$ 在 $\mathbf{u}$ 上的投影（projection）的长度。

解答

取 A 为原点，设 $\overrightarrow{AB} = \mathbf{a}$ 及 $\overrightarrow{AD} = \mathbf{b}$。由于 $ABCD$ 是平行四边形（parallelogram），$\overrightarrow{BC} = \overrightarrow{AD} = \mathbf{b}$，所以

$$\overrightarrow{AC} = \mathbf{a}+\mathbf{b}, \qquad \overrightarrow{BD} = \mathbf{b}-\mathbf{a}.$$

对任何向量都有 $|\mathbf{v}|^2 = \mathbf{v}\cdot\mathbf{v}$，而且点积是对称的，并对加法满足分配律。记 $a = |\mathbf{a}|$ 及 $b = |\mathbf{b}|$。



图中，N 是由 D 到直线 AB 的垂足，所以 AN 是 AD 在 AB 上的投影，其长度为 $|\mathbf{a}\cdot\mathbf{b}|/a$。

第 (i) 部分

对边相等，所以左方等于 $2a^2 + 2b^2$。至于对角线，

$$AC^2 = (\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) = a^2 + 2\,\mathbf{a}\cdot\mathbf{b} + b^2,$$
$$BD^2 = (\mathbf{b}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a}) = a^2 - 2\,\mathbf{a}\cdot\mathbf{b} + b^2.$$

两式相加，交叉项互相抵消，得 $AC^2 + BD^2 = 2a^2 + 2b^2 = AB^2 + BC^2 + CD^2 + DA^2$，这正是所要证明的。（这个结果通常称为平行四边形定律（parallelogram law）。）

第 (ii) 部分

这次改为把同样的两式相减，

$$AC^2 - BD^2 = 4\,\mathbf{a}\cdot\mathbf{b}, \qquad \text{所以} \qquad \left|AC^2 - BD^2\right| = 4\,|\mathbf{a}\cdot\mathbf{b}|.$$

现在来解释 $|\mathbf{a}\cdot\mathbf{b}|$ 的意义。若 θ 是 $\mathbf{a}$ 与 $\mathbf{b}$ 之间的夹角，则 $|\mathbf{a}\cdot\mathbf{b}| = a\cdot b|\cos\theta|$，而 $b|\cos\theta|$ 正好是 AD 在 AB 上的投影的长度 AN。所以 $|\mathbf{a}\cdot\mathbf{b}|$ 就是以 AB 和 AD 在 AB 上的投影为边的长方形（图中的阴影部分）的面积。

> **想一想** 题目声称这个结果对任何一条边和它任何一条邻边都成立。为甚么我们不需要分别做八次计算？

平行四边形的每条边都是向量 $\pm\mathbf{a}$ 或 $\pm\mathbf{b}$，而与它相邻的两条边，除了正负号之外，就是另一个向量。例如，边 BC 是 $\mathbf{b}$，它的邻边 AB 和 CD 分别是 $\mathbf{a}$ 和 $-\mathbf{a}$。改变正负号并不会改变向量的长度，也不会改变投影的长度，因为投影的长度是 $|\mathbf{u}\cdot\mathbf{v}|/|\mathbf{u}|$。所以无论选哪一条边和哪一条邻边，所得的长方形面积都同样是 $|\mathbf{a}\cdot\mathbf{b}|$：它或是 $a\times$（向量$\mathbf{b}$ 投影于$\mathbf{a}$），或是 $b\times$（向量$\mathbf{a}$ 投影于$\mathbf{b}$），而两者都等于 $|\mathbf{a}\cdot\mathbf{b}|$。因此，在所有情况下，

$$\left|AC^2 - BD^2\right| = 4\times\big(\text{该长方形的面积}\big). \qquad \blacksquare$$

最后部分

第 (ii) 部分来自两条边的点积。现在把边和对角线的角色对调，考虑两条对角线的点积：

$$\overrightarrow{AC}\cdot\overrightarrow{BD} = (\mathbf{a}+\mathbf{b})\cdot(\mathbf{b}-\mathbf{a}) = b^2 - a^2,$$

因为交叉项 $\mathbf{a}\cdot\mathbf{b}$ 和 $-\mathbf{b}\cdot\mathbf{a}$ 互相抵消。因此

$$\left|AB^2 - AD^2\right| = |\overrightarrow{AC}\cdot\overrightarrow{BD}|.$$

仿照第 (ii) 部分的论证，便得到以下结果。

结果。$\left|AB^2 - AD^2\right|$ 等于以下长方形的面积：它的两边分别是任何一条对角线，以及另一条对角线在它上面的投影。

证明。若 φ 是两条对角线之间的夹角，则 $\left|\overrightarrow{AC}\cdot\overrightarrow{BD}\right| = AC\cdot BD\,|\cos\varphi|$。这里 $BD\,|\cos\varphi|$ 是 BD 在 AC 上的投影的长度，而 $AC\,|\cos\varphi|$ 是 AC 在 BD 上的投影的长度。所以无论从哪一条对角线出发，面积都相同，而且由上面的式子可知它等于 $\left|AB^2 - AD^2\right|$。■

想让它看起来更像 (ii)，可设 O 为两条对角线互相平分的那一点。这样 $AC = 2OA$，而 BD 在 AC 上的投影是 OB 在 OA 上的投影的两倍。所以 $\left|AB^2 - AD^2\right|$ 也等于以 OA 和 OB 在 OA 上的投影为边的长方形面积的 4 倍：半条对角线现在担当了边的角色。

由此还可得出两个漂亮的推论（corollary）。两条对角线相等，当且仅当 $\mathbf{a}\cdot\mathbf{b} = 0$，亦即当且仅当该平行四边形是长方形。两条对角线互相垂直，当且仅当 $a = b$，亦即当且仅当它是菱形（rhombus）。

验算

取 $A = (0,0)$、$B = (4,0)$、$D = (1,2)$，所以 $C = (5,2)$。这样 $AC^2 = 29$ 及 $BD^2 = 13$，而 $2(16+5) = 42 = 29+13$，验证了 (i)。至于 (ii)，$AC^2 - BD^2 = 16$，而 AD 在 AB 上的投影长度为 1，所以长方形的面积是 $4\times 1 = 4$，而 $16 = 4\times 4$。至于最后部分，$AB^2 - AD^2 = 16 - 5 = 11$，而 $\overrightarrow{AC}\cdot\overrightarrow{BD} = (5,2)\cdot(-3,2) = -11$。我们还用 10^5 个随机平行四边形，就每一种边和邻边的选法测试了全部三个结果；最大的偏差约为 10^{-13}，只是舍入误差而已。

常见错误

- 写成 $\overrightarrow{BD} = \mathbf{a} - \mathbf{b}$ 或 $\mathbf{a} + \mathbf{b}$。正负号对 BD^2 没有影响，但把方向写对可以避免之后出现混乱。
- 忘记取绝对值：当 A 处的角是钝角时，$\mathbf{a}\cdot\mathbf{b}$ 是负数，但面积不可能是负数，所以题目写的是 $\left|AC^2 - BD^2\right|$。
- 在最后部分用了整条对角线，却写上因子 4。用整条对角线时因子是 1；只有用半条对角线时才会出现因子 4。

STEP 之外

恒等式 $|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2 = 4\,\mathbf{a}\cdot\mathbf{b}$ 就是极化恒等式（polarisation identity）：它表明单凭长度就可以还原出点积。它与第 (i) 部分的平行四边形定律一起，是关于内积空间（inner product space）的关键事实。一个范数（norm）来自某个内积，当且仅当它满足平行四边形定律，这是约当（Jordan）与冯·诺伊曼（von Neumann）的一条定理。Strang 的 *Introduction to Linear Algebra*（线性代数入门）是个好去处，可以看到点积、投影和正

交性（orthogonality）如何在任意维数中发展出来。

详细解答

STEP I 1990，第 7 题

纯数学 · 一阶线性常微分方程

> **STEP I 1990，第 7 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-s1-q7

关键思路

设 $y = uv$ 之后，我们有两个未知函数，但其实只需要一个，所以大可对其中一个多加一个条件。最自然的选择，是让 u 消去含 P 的那一项，余下的微分方程（differential equation）便可直接积分。这其实是乔装了的积分因子法（integrating-factor method）：$1/u$ 正正就是积分因子。

解答

代换

设 $y = uv$，由乘积法则（product rule）得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = u\dfrac{\mathrm{d}v}{\mathrm{d}x} + v\dfrac{\mathrm{d}u}{\mathrm{d}x}$，于是方程变成

$$u\frac{\mathrm{d}v}{\mathrm{d}x} + v\left(\frac{\mathrm{d}u}{\mathrm{d}x} + Pu\right) = Q.$$

如果我们选取 u 使它满足一阶方程（first order equation）

$$\frac{\mathrm{d}u}{\mathrm{d}x} + Pu = 0,$$

括号内的式子便等于零，余下

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{Q}{u},$$

这正是题目所要求的、只以 x、Q 和 u 表示 $\frac{\mathrm{d}v}{\mathrm{d}x}$ 的方程。（要作除法，必须有 $u \neq 0$；$u' + Pu = 0$ 的非零解是 $u = A\mathrm{e}^{-\int P\,\mathrm{d}x}$，其中 $A \neq 0$，这些解永不为零。）

题中的方程

这里 $P=-\dfrac{2}{x+1}$，$Q=(x+1)^{5/2}$。留意 Q 只在 $x\geq -1$ 时才有定义（作为实数），而 P 在 $x=-1$ 处没有定义，所以我们在 $x>-1$ 上求解，此时 $x+1>0$。

> **想一想** 我们需要 u 的通解（general solution）吗？还是任何一个解都可以？

任何一个非零解都可以：最终答案中的任意常数，会在积分求 v 时出现。于是我们用分离变量法（separating variables）解

$$\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{2u}{x+1}$$

得 $\ln|u|=2\ln(x+1)+$ 常数，并取

$$u=(x+1)^2.$$

（验算：$u'=2(x+1)=\frac{2u}{x+1}$。）于是

$$\frac{\mathrm{d}v}{\mathrm{d}x}=\frac{(x+1)^{5/2}}{(x+1)^2}=(x+1)^{1/2},\qquad \text{所以}\qquad v=\tfrac{2}{3}(x+1)^{3/2}+C.$$

因此通解是

$$y=uv=\tfrac{2}{3}(x+1)^{7/2}+C(x+1)^2.$$

代入 $y(1)=0$：

$$0=\tfrac{2}{3}\cdot 2^{7/2}+4C\quad\Longrightarrow\quad C=-\tfrac{1}{6}\cdot 2^{7/2}=-\tfrac{1}{6}\cdot 8\sqrt{2}=-\frac{4\sqrt{2}}{3}.$$

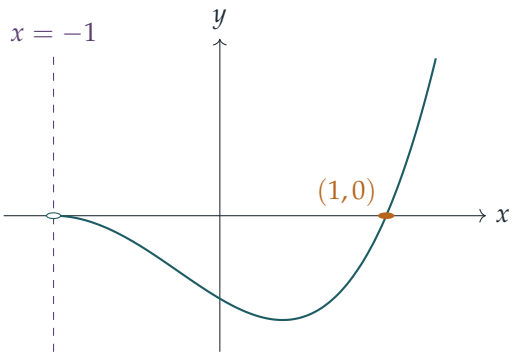
所以

$$\boxed{y=\tfrac{2}{3}(x+1)^{7/2}-\tfrac{4\sqrt{2}}{3}(x+1)^2=\tfrac{2}{3}(x+1)^2\left[(x+1)^{3/2}-2\sqrt{2}\right]}$$

解在何处成立

> **想一想** y 的公式在 $x=-1$ 处是有意义的。为甚么不把 $x=-1$ 包括在内？

这个解在 $x>-1$ 时成立。当 $x<-1$ 时，右边的 $(x+1)^{5/2}$ 不是实数，所以微分方程本身在那里毫无意义（我们的 $(x+1)^{7/2}$ 也一样）。在 $x=-1$ 处，$\frac{2y}{x+1}$ 这一项没有定义，所以方程在那里同样不成立，尽管按我们的公式，当 $x\to -1^+$ 时 $y\to 0$。在整个区间 $x>-1$ 上，由于 $x+1>0$，上面每一步都站得住脚，而且解经过题目给定的点 $x=1$，这一点正在此区间之内。



空心圆圈标示被排除的端点 $x=-1$。

验算

- 代回原方程：$y'=\frac{7}{3}(x+1)^{5/2}-\frac{8\sqrt{2}}{3}(x+1)$，而 $\frac{2y}{x+1}=\frac{4}{3}(x+1)^{5/2}-\frac{8\sqrt{2}}{3}(x+1)$；两者之差是 $(x+1)^{5/2}$，符合要求。
- $y(1)=\frac{2}{3}\cdot 4\left(2\sqrt{2}-2\sqrt{2}\right)=0$。
- 特殊值：$y(0)=\frac{2-4\sqrt{2}}{3}\approx -1.219$，$y(3)=\frac{256-64\sqrt{2}}{3}\approx 55.16$。由 $y(1)=0$ 出发，对方程作数值积分（龙格—库塔法，Runge–Kutta），所得结果与这两个值吻合至九位有效数字。

常见错误

- 弄错 P 的正负号：这里 $P=-\frac{2}{x+1}$，所以 $u=(x+1)^2$，而不是 $(x+1)^{-2}$。
- 积分求 v 时漏掉常数，结果没有余地去满足 $y(1)=0$。
- 把范围写成 $x\geq -1$：方程在 $x=-1$ 处没有定义。

STEP 之外

选取 u 来消去麻烦的一项，正是*参数变易法*（variation of parameters）背后的想法：先解齐次方程（homogeneous equation），然后让那个“常数”变动起来，以吸收外加项（forcing）。同样的技巧也适用于二阶线性方程；至于解在何处成立这个问题，则发展成存在与唯一性定理（existence and uniqueness theorem），它保证在 P 和 Q 皆连续的任何区间上都有唯一的解。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）对两者都有论述。

详细解答

STEP I 1990，第 8 题

纯数学 · 三角函数无穷乘积

STEP I 1990，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/90-s1-q8

关键思路

二倍角公式（double-angle formula）$\sin 2\theta = 2\sin\theta\cos\theta$ 可把每个余弦化成两个正弦之比：$\cos\theta = \dfrac{\sin 2\theta}{2\sin\theta}$。把这些比值相乘，几乎所有项都会互相抵消，情况就如裂项相消的和（telescoping sum）一样。一旦把有限乘积化成一个简洁的分数，令 $n \to \infty$ 时便只需用到标准极限（limit）：当 $x \to 0$ 时 $\dfrac{\sin x}{x} \to 1$。

解答

关于 α 的条件

以下每一步都要除以某个 $m \geq 0$ 所对应的 $\sin(\alpha/2^m)$，所以必须确保这些值永不为零。若 $\sin(\alpha/2^m) = 0$，则对某个整数 j 有 $\alpha/2^m = j\pi$，于是 $\alpha = (2^m j)\pi$ 是 π 的整数倍，而这种情况已被排除。因此对所有 $m \geq 0$ 都有 $\sin(\alpha/2^m) \neq 0$。

两个余弦的乘积

两次运用 $\sin 2\theta = 2\sin\theta\cos\theta$，先取 $\theta = \alpha/2$，再取 $\theta = \alpha/4$，得

$$\begin{aligned}\sin\alpha &= 2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}\\ &= 2\left(2\sin\frac{\alpha}{4}\cos\frac{\alpha}{4}\right)\cos\frac{\alpha}{2} = 4\sin\frac{\alpha}{4}\cos\frac{\alpha}{2}\cos\frac{\alpha}{4}.\end{aligned}$$

由于 $\sin(\alpha/4) \neq 0$，我们可以除以 $4\sin(\alpha/4)$，得出

$$\cos\frac{\alpha}{2}\cos\frac{\alpha}{4} = \frac{\sin\alpha}{4\sin(\alpha/4)},$$

这正是所需的结果。

n 个余弦的乘积

记 $P_n = \cos\frac{\alpha}{2}\cos\frac{\alpha}{4}\cdots\cos\frac{\alpha}{2^n}$。我们对 n 用归纳法（induction）证明

$$P_n = \frac{\sin\alpha}{2^n\sin(\alpha/2^n)}.$$

当 $n=1$ 时，上式即 $\cos\frac{\alpha}{2} = \dfrac{\sin\alpha}{2\sin(\alpha/2)}$，也就是把二倍角公式除以 $2\sin(\alpha/2)\neq 0$ 所得的结果。（$n=2$ 的情况就是刚才证明的结果。）

现假设公式对某个 $n \geq 1$ 成立。在二倍角公式中取 $\theta = \alpha/2^{n+1}$，得 $\sin\frac{\alpha}{2^n} = 2\sin\frac{\alpha}{2^{n+1}}\cos\frac{\alpha}{2^{n+1}}$。因此

$$\begin{aligned} P_{n+1} = P_n\cos\frac{\alpha}{2^{n+1}} &= \frac{\sin\alpha}{2^n\sin(\alpha/2^n)}\cos\frac{\alpha}{2^{n+1}} \\ &= \frac{\sin\alpha\,\cos(\alpha/2^{n+1})}{2^n\cdot 2\sin(\alpha/2^{n+1})\cos(\alpha/2^{n+1})} = \frac{\sin\alpha}{2^{n+1}\sin(\alpha/2^{n+1})}. \end{aligned}$$

约去 $\cos(\alpha/2^{n+1})$ 是合理的：它是 $\sin(\alpha/2^n)$ 的一个因子，而后者不为零，所以它本身也不为零。这正是 $n+1$ 时的公式，故由归纳法可知，公式对所有正整数 n 都成立。

令 $n\to\infty$

无穷乘积（infinite product）的意思，是部分积（partial products）P_n 的极限。设 $x_n = \alpha/2^n$，则当 $n\to\infty$ 时 $x_n\to 0$，并且

$$2^n\sin\frac{\alpha}{2^n} = \alpha\cdot\frac{\sin x_n}{x_n} \longrightarrow \alpha\cdot 1 = \alpha.$$

> **想一想** 为甚么可以放心地把 $\lim P_n = \dfrac{\sin\alpha}{\alpha}$ 上下倒转?

因为这个极限不是零：对我们所考虑的 α，有 $\alpha\neq 0$ 及 $\sin\alpha\neq 0$。所以 $P_n\to\dfrac{\sin\alpha}{\alpha}\neq 0$，无穷乘积存在，而以它除 $\sin\alpha$ 便得

$$\frac{\sin\alpha}{\cos\frac{\alpha}{2}\cos\frac{\alpha}{4}\cos\frac{\alpha}{8}\cdots} = \lim_{n\to\infty}\frac{\sin\alpha}{P_n} = \lim_{n\to\infty}2^n\sin\frac{\alpha}{2^n} = \alpha,$$

这就是所需的公式。

$\pi/2$ 的公式

取 $\alpha = \pi/2$，它不是 π 的整数倍，并留意 $\sin\alpha = 1$。于是

$$\frac{\pi}{2} = \frac{1}{\cos\frac{\pi}{4}\cos\frac{\pi}{8}\cos\frac{\pi}{16}\cdots}.$$

余下要证明的，是这些余弦正是那些嵌套平方根（nested square roots）。对 $k \geq 1$，记 $c_k = \cos(\pi/2^{k+1})$。那么 $c_1 = \cos\frac{\pi}{4} = \sqrt{\frac{1}{2}}$。在半角公式（half-angle formula）$\cos^2\theta = \frac{1}{2} + \frac{1}{2}\cos 2\theta$ 中取 $\theta = \pi/2^{k+2}$，得

$$c_{k+1}^2 = \tfrac{1}{2} + \tfrac{1}{2}c_k, \qquad \text{所以} \qquad c_{k+1} = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}c_k},$$

这里取正根，是因为 $0 < \pi/2^{k+2} < \pi/2$，所以 $c_{k+1} > 0$。因此

$$c_1 = \sqrt{\tfrac{1}{2}}, \quad c_2 = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}}, \quad c_3 = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}}}, \quad \ldots,$$

而每个因子都是把前一个因子代入 $t \mapsto \sqrt{\frac{1}{2} + \frac{1}{2}t}$ 所得。这些正好就是所需乘积中的各个因子，所以

$$\frac{\pi}{2} = \frac{1}{\sqrt{\frac{1}{2}}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}}\cdots}.$$

验算

- 取 $\alpha = \pi/2$ 及 $n = 1$，有限乘积的公式给出 $\cos\frac{\pi}{4} = \dfrac{1}{2\sin(\pi/4)} = \dfrac{1}{\sqrt{2}}$，结果正确。
- 嵌套根式（nested radicals）各部分积的倒数依次为 1.4142、1.5307、1.5607、1.5683、1.5702,…；乘到第 60 个因子后得到 1.5707963267949，在机器精度之内与 $\pi/2$ 吻合。
- 对随机选取的 α 以及不超过 14 的 n，有限乘积的恒等式在数值上同样成立；而对 $\alpha = 2.5$ 和 $\alpha = -4$ 等数值，$\sin\alpha / P_{60}$ 也能重现 α。

常见错误

- 除以 $\sin(\alpha/2^n)$ 而没有说明它为何不为零。这正是设下条件 $\alpha \neq k\pi$ 的原因。
- 写出 $1/\lim P_n = \lim(1/P_n)$ 而没有指出 $\lim P_n \neq 0$。
- 在半角公式一步中，忘记说明为何取正平方根。

STEP 之外

最后的公式就是 1593 年的韦达（Viète）乘积，是历史上第一个写下来的 π 的精确无穷表达式。到了大学，你会在分析学（analysis）中正式学习无穷乘积：若一个由非零因子组成的乘积，其部分积趋向一个非零的极限，便称该乘积收敛（converge），而这正是我

们刚才检查过的一点。欧拉（Euler）的乘积 $\sin\alpha = \alpha\prod_{k\geq 1}\left(1-\frac{\alpha^2}{k^2\pi^2}\right)$ 则是与它一脉相承而更为深刻的结果。Abbott 的 *Understanding Analysis* 是一本平易近人的书，可以从中看到这些极限如何得到严谨的处理。

详细解答

STEP I 1990，第 9 题

纯数学 · 切线与面积的界限

> **STEP I 1990，第 9 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-s1-q9

关键思路

曲线 $y = 1/x$ 向上弯曲（即它是凸（convex）的），所以每一条切线（tangent）都位于曲线之下。因此，两条切线下方的两个梯形（trapezium）都包含在曲线下方由 $x = 1$ 至 $x = b$ 的区域之内，而该区域的面积为 $\ln b$。比较这两个面积，便得到对数的下界（lower bound）。同样道理，一个位于曲线之上的长方形则给出上界（upper bound）。

解答

点 C

对于 $y = 1/x$，我们有 $\frac{\mathrm{d}y}{\mathrm{d}x} = -1/x^2$。所以在 $(t, 1/t)$ 的切线为

$$y - \frac{1}{t} = -\frac{1}{t^2}(x - t), \qquad \text{即} \qquad y = \frac{2}{t} - \frac{x}{t^2}.$$

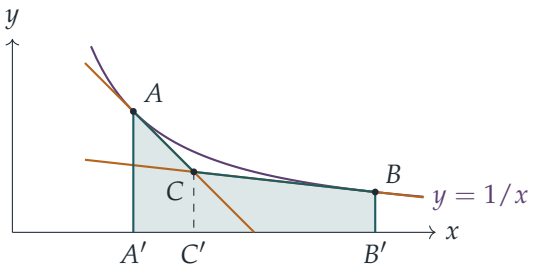
取 $t = 1$，在 A 的切线为 $y = 2 - x$。取 $t = b$，在 B 的切线为 $y = \frac{2}{b} - \frac{x}{b^2}$。在两线相交之处，

$$2 - x = \frac{2}{b} - \frac{x}{b^2} \quad \Longrightarrow \quad x\left(1 - \frac{1}{b^2}\right) = 2\left(1 - \frac{1}{b}\right).$$

由于 $b > 1$，我们可以除以 $1 - \frac{1}{b} \neq 0$，余下 $x\left(1 + \frac{1}{b}\right) = 2$。因此

$$C = \left(\frac{2b}{b+1}, \frac{2}{b+1}\right),$$

其中 y 坐标由 $y = 2 - x$ 求得。注意 $1 < \frac{2b}{b+1} < b$：左边的不等式（inequality）即 $b + 1 < 2b$，右边的即 $2 < b + 1$。两者皆因 $b > 1$ 而成立，所以 C 在水平方向上位于 A 与 B 之间。



此图按 $b=3$ 绘画。阴影区域由该两个四边形（quadrilateral）组成。

面积之和

每个四边形都有两条铅垂的边，所以它是梯形，其面积等于阔度乘以两条铅垂边的平均值。

对于 $ACC'A'$，阔度为 $\frac{2b}{b+1}-1=\frac{b-1}{b+1}$，而两条铅垂边分别为 1 和 $\frac{2}{b+1}$：

$$\text{面积}(ACC'A')=\frac{1}{2}\cdot\frac{b-1}{b+1}\left(1+\frac{2}{b+1}\right)=\frac{(b-1)(b+3)}{2(b+1)^2}.$$

对于 $CBB'C'$，阔度为 $b-\frac{2b}{b+1}=\frac{b(b-1)}{b+1}$，而两条铅垂边分别为 $\frac{2}{b+1}$ 和 $\frac{1}{b}$：

$$\text{面积}(CBB'C')=\frac{1}{2}\cdot\frac{b(b-1)}{b+1}\left(\frac{2}{b+1}+\frac{1}{b}\right)=\frac{(b-1)(3b+1)}{2(b+1)^2}.$$

两式相加，分子合并为 $(b-1)(4b+4)=4(b-1)(b+1)$，所以

$$\text{面积}(ACC'A')+\text{面积}(CBB'C')=\frac{2(b-1)}{b+1}.$$

不等式

> ? **想一想** 为甚么两个梯形必定包含在曲线之下?

对于 $x>0$，曲线与在 $(t,1/t)$ 的切线之间的差距为

$$\frac{1}{x}-\left(\frac{2}{t}-\frac{x}{t^2}\right)=\frac{t^2-2tx+x^2}{t^2x}=\frac{(x-t)^2}{t^2x}\geq 0.$$

所以对所有 $x>0$，每条切线都位于曲线上或曲线之下。阴影区域的上边界，在 $[1,\frac{2b}{b+1}]$ 上是在 A 的切线，在 $[\frac{2b}{b+1},b]$ 上是在 B 的切线。两段都位于 $y=1/x$ 之下，而该区域位于 x 轴之上。所以阴影区域包含在曲线下方由 $x=1$ 至 $x=b$ 的区域之内，而

$$\frac{2(b-1)}{b+1}\leq\int_1^b\frac{\mathrm{d}x}{x}=\ln b.$$

至于上界，当 $x \geq 1$ 时 $\frac{1}{x} \leq 1$，所以曲线下方的区域位于以 $A'B'$ 为底、高为 1 的长方形之内。因此

$$\ln b = \int_1^b \frac{\mathrm{d}x}{x} \leq \int_1^b 1\,\mathrm{d}x = b - 1.$$

现设 $z > 0$ 并取 $b = 1 + z$，于是 $b > 1$，符合要求。这时 $b - 1 = z$ 及 $b + 1 = 2 + z$，两个不等式合起来就是

$$\frac{2z}{2+z} \leq \ln(1+z) \leq z,$$

这正是我们要证明的结果。（事实上，当 $z > 0$ 时两个不等式都是严格的，因为曲线与作为边界的直线只在孤立的点相交。）

验算

- 当 $b = 1$ 时，C 会是 $(1,1) = A$，而面积之和会是 0，与预期相符。
- 当 $b = 2$ 时：$C = (\frac{4}{3}, \frac{2}{3})$，面积之和为 $\frac{2}{3} \approx 0.667$，仅略小于 $\ln 2 \approx 0.693$，而后者又小于 1。
- 对于细小的 z，$\frac{2z}{2+z} = z - \frac{z^2}{2} + \cdots$ 与 $\ln(1+z) = z - \frac{z^2}{2} + \cdots$ 直至 z^2 项都一致，所以下界在 0 附近非常精确。

我们以精确运算对若干个 b 值验算了坐标和两个梯形的面积，并对 10^5 个介乎 10^{-6} 与 10^6 之间的随机 z 值验算了该不等式。

常见错误

- 忘记检查 C 是否位于 $x = 1$ 与 $x = b$ 之间。缺少这一步，那些“四边形”未必是你所想像的形状。
- 声称“从图中可见”切线位于曲线之下。用 $(x - t)^2$ 作出的那一行计算，才令它成为一个证明。

STEP 之外

用简单的图形从上方和下方与函数的图像作比较，正是黎曼和（Riemann sum）背后的想法，也是大学第一门分析（analysis）课程中严格定义积分的方法。凸函数（convex function）的切线位于其图像之下，这个事实是凸函数理论的起点。由此可推出延森不等式（Jensen's inequality）以及许多经典不等式，包括算术平均—几何平均不等式（AM–GM）。Abbott 的 *Understanding Analysis*（一本分析入门教科书）正是按这种精神建立积分理论的。

详细解答

STEP I 1990，第 10 题

力学 · 功率与阻力

STEP I 1990，第 10 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-s1-q10

关键思路

“稳定速率”（steady speed）即是没有加速度，所以跑手的驱动力（driving force）恰好与各阻力平衡，而功率（power）等于驱动力乘以速率。唯一要留神的地方，是空气阻力（air resistance）取决于相对于空气的速率，而这个速率在去程和回程并不相同。至于题目的后半部分，应该考虑总能量（energy）而不是功率：每一程所消耗的能量等于阻力乘以距离，于是固定的能量预算就化为两个速率之间的一个线性条件。

解答

稳定速率与完成赛事的时间

把跑手看成沿直线运动的质点（particle）。设 v（单位为 $\mathrm{m\,s^{-1}}$）为跑手在某一程的对地速率，D（单位为 N）为跑手产生的向前驱动力。阻力分别为 50 N 和 $2w$ N，两者分别阻碍相对于地面和相对于空气的运动。



沿跑道方向分解各力，由于以稳定步速前进时加速度为零，所以

$$D = 50 + 2w.$$

跑手作功的速率等于驱动力乘以其作用点的速率，所以 $300 = Dv$。

去程。风以 $3\,\mathrm{m\,s^{-1}}$ 的速率顺着跑手的运动方向吹。若 $v > 3$，跑手相对于空气以 $w = v - 3$ 向前运动，空气阻力向后作用。（下文会验证 $v > 3$。）于是 $D = 50 + 2(v-3) = 44 + 2v$，

而

$$v_1(44+2v_1)=300 \iff v_1^2+22v_1-150=0 \iff v_1=-11+\sqrt{271}\approx 5.462,$$

这里取正根。这个值确实大于 3，所以假设成立。（另一个看法：当 $v=3$ 时没有空气阻力，所需功率只是 $3\times 50=150<300\,\mathrm{W}$，所以跑手必定跑得比风快。）

回程。这时跑手逆风而跑，所以 $w=v+3$，而 $D=50+2(v+3)=56+2v$：

$$v_2(56+2v_2)=300 \iff v_2^2+28v_2-150=0 \iff v_2=-14+\sqrt{346}\approx 4.601.$$

时间。忽略起跑和折返时的加速过程，

$$T_0=\frac{5000}{v_1}+\frac{5000}{v_2}\approx 915.4+1086.7=2002.1,$$

所以完成赛事需时 2002 s（准确至最接近的秒），即大约 33 分 22 秒。

另一种策略

现在假设跑手去程以 x_1 W、回程以 x_2 W 的功率作功，稳定速率分别为 v_1 和 v_2。力的平衡关系不变，所以只要 $v_1>3$，

$$x_1=v_1(44+2v_1), \qquad x_2=v_2(56+2v_2).$$

去程所需时间为 $5000/v_1$，所以去程消耗的能量为

$$x_1\cdot\frac{5000}{v_1}=5000(44+2v_1),$$

这正好是阻力乘以距离，理应如此。同样地，回程消耗 $5000(56+2v_2)$。因此总能量为

$$E=5000\big(100+2(v_1+v_2)\big).$$

想一想 固定的总能量在哪里派上用场?

跑手消耗的总能量必须与原来的赛事相同，所以 E 是一个固定的数。由于 E 只通过 v_1+v_2 取决于两个速率，这个和就被确定了：

$$v_1+v_2=\frac{E}{10\,000}-50,$$

无论 x_1 和 x_2 怎样选取都是如此。这就证明了所求结果。它的值就是原来赛事中的值，

$$S=v_1+v_2=\sqrt{271}+\sqrt{346}-25\approx 10.063.$$

最佳策略。现在完成赛事的时间为

$$T=5000\left(\frac{1}{v_1}+\frac{1}{v_2}\right)=\frac{5000\,(v_1+v_2)}{v_1v_2}=\frac{5000\,S}{v_1v_2}.$$

由于 S 固定，当 $v_1 v_2$ 最大时这个时间最小。根据算术—几何平均不等式（AM–GM inequality），$v_1 v_2 \le \left(\frac{S}{2}\right)^2$，等号恰好在 $v_1 = v_2 = \frac{S}{2}$ 时成立。所以最佳方案是来回都以相同的速率跑，即 $\frac{S}{2} \approx 5.032\,\mathrm{m\,s^{-1}}$。这仍然比风快，所以去程中 x_1 的公式依然有效。最短时间为

$$T_{\min} = \frac{5000\,S}{S^2/4} = \frac{20\,000}{S} \approx 1987.4\ \mathrm{s}.$$

跑手节省了 $T_0 - T_{\min} \approx 2002.1 - 1987.4 = 14.7\,\mathrm{s}$，即大约 15 秒，正如题目所求。所用功率约为去程 $x_1 \approx 272\,\mathrm{W}$、回程 $x_2 \approx 332\,\mathrm{W}$：跑手在顺风时放松一点，逆风时则加把劲。

验算

- 能量是否一致：$300 \times 2002.105 \approx 600\,632\,\mathrm{J}$，而 $5000(100 + 2 \times 10.0632) \approx 600\,632\,\mathrm{J}$ 亦然。
- 若没有风，两程都会得出 $v^2 + 25v - 150 = 0$，所以 $v = 5$，两种方法就会完全一致。正是有风，才有可能节省时间。
- 我们以数值方法验证过：随机选取 1000 个 x_1 的值（x_2 由能量预算确定），都有 $v_1 + v_2 = 10.0632$；没有任何选择能快过 1987.4 s；节省的时间为 14.66 s。

常见错误

- 在空气阻力一项中用了对地速率，而不是相对于空气的速率；又或者在其中一程把风的正负号弄反了。
- 以为来回用相同的*功率*才算合理：其实是来回用相同的*速率*才令时间最短。
- 试图单凭功率方程证明 $v_1 + v_2$ 为常数。你需要用到总能量固定这个条件。

STEP 之外

最后一部分是一个小型的约束优化（constrained optimisation）问题：在一个线性约束条件下，求 $1/v_1 + 1/v_2$ 的最小值。这里用 AM–GM 不等式就能解决，但一般的工具是大学一年级课程会学到的拉格朗日乘数法（method of Lagrange multipliers），而同样的想法也支配着单车和赛艇运动中的最佳配速。若想看对功、功率和阻力的严谨论述，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP I 1990，第 11 题

力学 · 抛体与动量

> **STEP I 1990，第 11 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-s1-q11

关键思路

整个过程分为三个阶段：先是普通的抛体（projectile）运动，直至飞行的最高点；然后是一次瞬时碰撞（接住炮弹）；最后是一次瞬时“爆炸”（推开炮弹）。在两个瞬时阶段中，重力来不及起作用，所以炮弹与超人组成的系统动量守恒（conservation of momentum），而且在每个方向上各自守恒。此后炮弹再次成为抛体，从已知的高度以已知的向下速率出发。

解答

我们把炮弹和超人都视为质点（particle），忽略空气阻力，并假设炮弹从水平地面发射，最后亦落在水平地面上。取 x 为沿运动方向的水平方向，y 为垂直向上。

直至飞行的最高点

炮弹初速度的分量为

$$u_x = v\cos\tfrac{\pi}{3} = \tfrac{1}{2}v, \qquad u_y = v\sin\tfrac{\pi}{3} = \tfrac{\sqrt{3}}{2}v.$$

在最高点，垂直速度为零，所以所需时间和到达的高度分别是

$$t_1 = \frac{u_y}{g} = \frac{\sqrt{3}\,v}{2g}, \qquad H = \frac{u_y^2}{2g} = \frac{3v^2}{8g}.$$

此刻炮弹正以速率 $\frac{1}{2}v$ 水平运动。

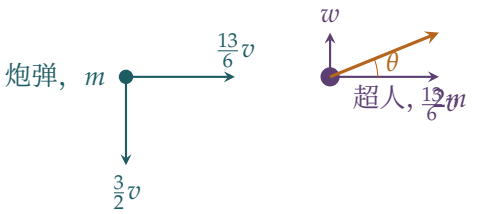
接住炮弹

超人以 $3v$ 的速率水平飞行，从后方接住炮弹，然后两者一起运动。在（瞬时的）接住过程中水平动量守恒，所以若 V 是两者的共同速率，则

$$2m(3v) + m\left(\tfrac{1}{2}v\right) = 3mV \quad \Longrightarrow \quad V = \frac{13v}{6}.$$

此刻两者都没有垂直速度，所以合起来的垂直动量为零。

推开炮弹



超人现在把炮弹垂直向下推。这一推是内部冲量（impulse），所以两者的总动量在两个方向上都保持不变。

水平方向：炮弹保持水平速率 $\frac{13}{6}v$。总水平动量仍然是 $3m \cdot \frac{13}{6}v$，所以超人的水平速率同样保持为 $\frac{13}{6}v$。（这并不意外：冲量是垂直方向的。）

垂直方向：总垂直动量原本为零，所以若超人以速率 w 向上运动，则

$$2m\,w - m\left(\tfrac{3}{2}v\right) = 0 \quad \Longrightarrow \quad w = \tfrac{3}{4}v.$$

> ? **想一想** 重力自始至终都在作用，为甚么我们仍然可以使用动量守恒?

接住和推开都是瞬时发生的。重力是有限的力，所以在长度为零的时间区间内，它的冲量为零；真正起作用的只有超人与炮弹之间巨大而短暂的内力，而这些内力成对地互相抵消。

炮弹余下的飞行

炮弹现在从高度 $H = \frac{3v^2}{8g}$ 出发，垂直向下的速率为 $\frac{3}{2}v$。取 s 以向下为正，炮弹经过时间 t_2 后到达地面，其中

$$\tfrac{3}{2}v\,t_2 + \tfrac{1}{2}gt_2^2 = \frac{3v^2}{8g}.$$

两边乘以 $8g$，得到一个关于 gt_2 的二次方程：

$$4(gt_2)^2 + 12v(gt_2) - 3v^2 = 0 \quad \Longrightarrow \quad gt_2 = \frac{-12v + \sqrt{144v^2 + 48v^2}}{8} = \frac{-12 + 8\sqrt{3}}{8}\,v,$$

由于 $t_2 > 0$，这里取正根。所以

$$t_2 = \left(\sqrt{3} - \tfrac{3}{2}\right)\frac{v}{g}.$$

（这是正数，因为 $\sqrt{3} \approx 1.732 > 1.5$。）总飞行时间为

$$\begin{aligned} T = t_1 + t_2 &= \frac{\sqrt{3}}{2}\frac{v}{g} + \left(\sqrt{3} - \frac{3}{2}\right)\frac{v}{g} \\ &= \frac{3(\sqrt{3}-1)}{2}\frac{v}{g} \;\approx\; 1.10\,\frac{v}{g}. \end{aligned}$$

超人的运动方向

紧接推开炮弹之后，超人的速度分量为水平方向 $\frac{13}{6}v$ 和向上 $\frac{3}{4}v$。因此他的轨迹起初与水平成仰角 θ，其中

$$\tan\theta = \frac{\frac{3}{4}v}{\frac{13}{6}v} = \frac{9}{26}, \qquad \theta = \arctan\frac{9}{26} \approx 19.09^\circ,$$

所以该角准确至最接近的度数是 19°。

验算

- 若没有超人，炮弹会在时间 $2t_1 = \sqrt{3}\,v/g \approx 1.73\,v/g$ 落地。把炮弹向下推理应缩短飞行时间，而 $T \approx 1.10\,v/g$ 的确较小。
- 质量比：炮弹的质量是超人的一半，所以它得到的垂直速率应是超人的两倍，方向相反。事实正是如此：$\frac{3}{2}v = 2 \times \frac{3}{4}v$。
- 数值验证：取 $v = 13.7$ 和 $g = 9.81$，以 10^{-6} 为步长逐步计算炮弹从高度 H 下落的过程，得 $t_2 \approx 0.32407\,\mathrm{s}$，与 $(\sqrt{3} - \frac{3}{2})v/g$ 吻合，而 $T \approx 1.5335\,\mathrm{s} = 1.0981\,v/g$。

常见错误

- 忘记了接住炮弹这一步：推开之后没有使用共同速率 $\frac{13}{6}v$，而是用了炮弹本身的水平速率 $\frac{1}{2}v$，或超人的 $3v$。
- 忽略质量，以为超人向上的速率是 $\frac{3}{2}v$。
- 取了二次方程的负根，或把 t_2 加到 $2t_1$ 而不是 t_1 上。

STEP 之外

接住炮弹是一次完全非弹性碰撞（perfectly inelastic collision），而推开炮弹则是它的逆过程，即一次爆炸：同一条守恒定律在两个方向上都适用，只是动能（kinetic energy）在前者中损失，在后者中得到补充。在大学一年级的课程中，这会发展成冲力（impulsive force）和系统质心（centre of mass）的一般理论：无论系统各部分之间互相推撞得多么猛烈，质心的运动只取决于外力。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）正是本着这种精神处理动量、冲量和质心。

详细解答

STEP I 1990，第 12 题

力学 · 弹簧与能量

STEP I 1990，第 12 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：

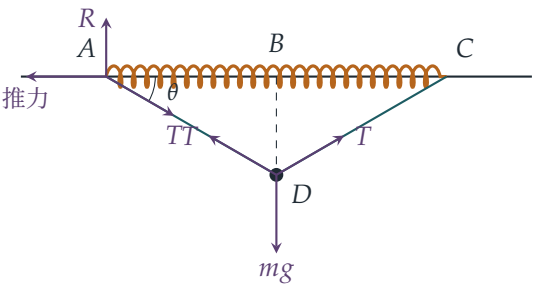
step.maths.org/questions/90-s1-q12

关键思路

这里除了质点之外，所有物件都是轻的，所以两端 A 和 C 所受的力必须完全平衡：绳子对每一端的水平拉力，等于被压缩的弹簧的推力。这就给出平衡条件。至于运动，能量是最自然的工具：唯一的动能是质点的动能，所以质点静止的时刻，正正就是势能（potential energy）回到初始值的时刻。

解答

模型。弹簧和绳子都是轻的，绳子不可伸长（inextensible），而杆是光滑的，所以杆只能沿垂直方向推弹簧。由于整个系统关于通过 B 的垂直线对称，质点 D 一直位于 B 的正下方，且 $AD = DC = l$。设 T 为每半段绳子的张力（tension），θ 为每半段绳子与杆所成的角。



平衡

当 $\theta = \pi/6$ 时，AC 的一半长度是 $l\cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}l$，所以

$$AC = \sqrt{3}\,l < 2l,$$

弹簧被压缩了 $(2-\sqrt{3})l$。根据虎克定律（Hooke's law），它对 A（以及对 C）施加向外的推力（thrust）

$$\frac{\lambda(2-\sqrt{3})l}{2l} = \lambda\Big(1-\frac{\sqrt{3}}{2}\Big).$$

在 D 沿垂直方向分解力：$2T\sin\frac{\pi}{6} = mg$，所以 $T = mg$。

在 A 沿水平方向分解力：

> **想一想** A 处明明甚么也没有，为甚么可以在那里列出力的平衡?

弹簧的末端是一个轻的点：质量为零的物体若受到任何净力，加速度便会是无限大，所以它所受的力必须在每一刻都平衡（而不仅是在平衡状态下）。杆是光滑的，所以它的反作用力（reaction）R 是垂直的，水平方向就只有绳子和弹簧在起作用：

$$T\cos\frac{\pi}{6} = \lambda\Big(1-\frac{\sqrt{3}}{2}\Big) \quad\Longrightarrow\quad mg\,\frac{\sqrt{3}}{2} = \lambda\,\frac{2-\sqrt{3}}{2},$$

所以

$$mg = \lambda\,\frac{2-\sqrt{3}}{\sqrt{3}} = \lambda\Big(\frac{2}{\sqrt{3}}-1\Big),$$

这正是所要证明的结果。

势能

设 $AC = 2xl$，则绳子拉紧时 $0 < x \le 1$。AC 的一半是 xl，而 $AD = l$，所以由毕氏定理（Pythagoras' theorem），D 在杆下方的深度为

$$BD = l\sqrt{1-x^2}$$

。弹簧的伸长量是 $2xl - 2l$，所以它的弹性势能（elastic energy）是 $\frac{\lambda(2l-2xl)^2}{2\cdot 2l} = \lambda l(1-x)^2$。取杆为重力势能（gravitational potential energy）的零位，

$$V(x) = \lambda l(1-x)^2 - mgl\sqrt{1-x^2}.$$

作为一致性的检查，

$$V'(x) = -2\lambda l(1-x) + \frac{mglx}{\sqrt{1-x^2}},$$

而在 $x = \frac{\sqrt{3}}{2}$ 处（这时 $\cos\theta = x$，所以 $\theta = \frac{\pi}{6}$），上式为零的充要条件正是 $mg\sqrt{3} = \lambda(2-\sqrt{3})$，也就是我们在上面求得的条件。

从 B 落下

起初质点位于 B，所以绳子沿着杆平放，$AC = 2l$ 而 $x = 1$。于是 $V(1) = 0$，系统由静止开始运动。

杆是光滑的，绳子不可伸长，所以没有能量损失；又因为弹簧和绳子都是轻的，唯一的动能（kinetic energy）就是质点的 $\frac{1}{2}mv^2$。因此在整个运动过程中

$$\tfrac{1}{2}mv^2 + V(x) = V(1) = 0.$$

> ? **想一想** 为甚么“静止下来”的意思是 $V(x) = 0$，而不是 $V'(x) = 0$?

质点静止时 $v = 0$，根据能量守恒（energy conservation），这表示 $V(x) = 0$。条件 $V'(x) = 0$ 描述的是平衡位置，那是质点运动得最快的地方，而不是它停下来的地方。所以 x 满足

$$\lambda(1-x)^2 = mg\sqrt{1-x^2}, \qquad \text{即} \qquad (1-x)^2 = \left(\frac{2}{\sqrt{3}} - 1\right)\sqrt{1-x^2}.$$

根 $x = 1$ 是起始位置。对于 $x < 1$，我们可以把两边除以 $\sqrt{1-x} > 0$ 再平方，得到一个较简洁的三次方程（cubic）：

$$(1-x)^3 = k^2(1+x), \qquad k = \frac{2}{\sqrt{3}} - 1 \approx 0.15470, \quad k^2 \approx 0.023932.$$

这里平方是稳妥的，因为 $(1-x)^{3/2} = k\sqrt{1+x}$ 的两边都是正数。设 $f(x) = (1-x)^3 - k^2(1+x)$。则

$$f(0.65) = 0.042875 - 0.039488 \approx +0.0034,$$
$$f(0.66) = 0.039304 - 0.039727 \approx -0.0004,$$

所以在 0.65 与 0.66 之间有一个根，而且非常接近 0.66。由于 $f'(x) = -3(1-x)^2 - k^2 < 0$，$f$ 是严格递减（strictly decreasing）的，所以这是唯一满足 $x < 1$ 的根。质点下一次静止时 $x \approx 0.66$（更精确地说 $x \approx 0.6589$），这时 D 在杆下方的深度约为 $0.75l$。

验算

- 在 $\theta = \pi/6$ 处，取 $\lambda = l = m = 1$ 及 $g = \frac{2}{\sqrt{3}} - 1$，$A$ 处的水平拉力和弹簧的推力都等于 0.13397，而 V 在 $x = \frac{\sqrt{3}}{2}$ 处的数值导数为零（且 $V'' > 0$，所以这个平衡是稳定的）。
- 用二分法（bisection）求得根 $x = 0.658861$；另外，就 D 的深度 y，由静止于 $y = 0$ 开始，以 RK4 逐步计算运动方程 $m\ddot{y} = mg - 2\lambda(1-x)\,y/(lx)$，结果显示质点首次停下时 $x = 0.658861$，两者吻合。

- 平衡时的值 $x = \frac{\sqrt{3}}{2} \approx 0.866$ 位于 0.66 与 1 之间，这是必然的：质点围绕平衡位置振荡。

常见错误

- 以为弹簧是被拉长的。这里 $AC < 2l$，所以弹簧是被压缩的，它把 A 和 C 向外推开。
- 以 $V'(x) = 0$ 来求静止位置。这样求得的是平衡位置，那里速率最大。
- 忘记了重力势能在杆的下方是负的，结果得出一条在 $(0, 1)$ 内无解的能量方程。

STEP 之外

“无质量的物体所受的净力必为零”这个想法是一个约束（constraint），而不是一条运动定律。这类问题用拉格朗日力学（Lagrangian mechanics）处理最为干净利落：你只需用单一个坐标（这里是 x）写下一个能量函数，约束力根本不会出现。势能 $V(x)$ 还可以通过 V'' 给出围绕平衡位置的小振荡（small oscillations）的周期。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）正是以这种精神讲解能量方法和小振荡。

详细解答

STEP I 1990，第 13 题

力学 · 摩擦力与刚体平衡

STEP I 1990，第 13 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：

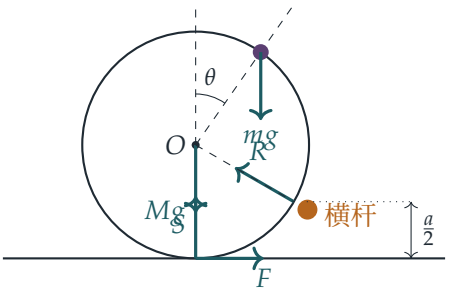
step.maths.org/questions/90-s1-q13

关键思路

先处理质点（particle）：它只是一个置于粗糙斜面上的普通质点，而该斜面就是圆柱体在接触点处的切平面（tangent plane）。然后留意，由于质点处于平衡（equilibrium），它施于圆柱体的总力就只是它的重量，并且通过接触点铅垂向下作用。对圆柱体的中心取力矩（moments），除了这个力和来自平面的摩擦力（friction）之外，其余各力全部消失，这是最快捷的入手方法。

解答

把质点视为质量为 m 的点质量，把圆柱体视为质量为 M 的均匀刚体（rigid body）圆柱，其质心（centre of mass）O 位于轴上。横杆是光滑的，所以它只施加法向力；平面和圆柱体表面则是粗糙的，摩擦系数（coefficient of friction）为 μ。一切都发生在通过 O 的铅垂横截面内，以下我们只考虑这个截面。把横杆放在右边。



图中显示作用于圆柱体的各个力（质点位置的 mg 会在第二部分解释）。箭咀并非按比例绘画。

质点的平衡

在接触点 P，向外的法线与向上的铅垂线成 θ 角，所以该处的切平面是一个与水平成 θ 角的斜面。作用于质点的力有：它的重量 mg、沿向外法线方向的法向反作用力（normal reaction）N，以及沿切线、指向斜面上方的摩擦力 f。沿法线方向及垂直于法线方向分解，

$$N = mg\cos\theta, \qquad f = mg\sin\theta.$$

只要 $f \le \mu N$，摩擦力便足以提供这个力，即

$$\tan\theta \le \mu, \qquad \text{等价于} \qquad \theta \le \arctan\mu.$$

（我们需要 $0 \le \theta < 90^\circ$ 才能使 N 为正，而由于质点位于圆柱体的上半部，这是成立的。）

放手后圆柱体的平衡

> **想一想** 质点以一个法向力和一个摩擦力压在圆柱体上。我们需要分开处理这两个力吗？

不需要。根据牛顿第三定律（Newton's third law），圆柱体在 P 点受到 $-N$ 和 $-f$。由于 N 与 f 合起来抵销质点的重量，它们的和是铅垂向上的 mg，所以质点施于圆柱体的总力是铅垂向下、作用于 P 点的 mg。它的作用线与 O 的水平距离为 $a\sin\theta$，位于横杆的一侧。

作用于圆柱体的其他力有：

- 它的重量 Mg，作用于 O；
- 横杆的反作用力 R，垂直于圆柱体表面，因此沿半径指向 O。横杆的接触点高度为 $a/2$，即在 O 之下 $a/2$，所以这条半径与向下的铅垂线成 60°。因此 R 有一个背离横杆的水平分量 $\frac{\sqrt{3}}{2}R$ 和一个向上的分量 $\frac{1}{2}R$；
- 在最低点，有向上的法向反作用力 S 和水平方向的摩擦力 F（取其方向为指向横杆）。

对 O 取力矩。Mg、R 和 S 的作用线全部通过 O。质点的 mg 的力矩为 $mga\sin\theta$（使圆柱体转向横杆），而 F 作用于 O 之下距离 a 处，转向相反。所以

$$Fa = mga\sin\theta \quad \Longrightarrow \quad F = mg\sin\theta.$$

这是正数，所以摩擦力确实如图所示指向横杆。（若没有横杆，圆柱体会向横杆那一侧滚动；这就是为甚么质点必须在那一侧，横杆才会有用。）

水平方向分解：$\frac{\sqrt{3}}{2}R = F$，所以

$$R = \frac{2mg\sin\theta}{\sqrt{3}} > 0,$$

这与横杆只能推而不能拉相符。

铅垂方向分解：

$$S = (M+m)g - \tfrac{1}{2}R = (M+m)g - \frac{mg\sin\theta}{\sqrt{3}},$$

由于 $\sin\theta < 1 < \sqrt{3}$ 且 $M > 0$，这是正数，所以圆柱体不会离开平面。

只要 $F \le \mu S$，平面便能提供所需的摩擦力：

$$m\sin\theta \le \mu\left[(M+m) - \frac{m\sin\theta}{\sqrt{3}}\right] \iff m\sin\theta\left(\sqrt{3}+\mu\right) \le \sqrt{3}\,\mu\,(M+m).$$

所以，放手之后，只要 $\tan\theta \le \mu$ 与

$$\boxed{\sin\theta \le \frac{\sqrt{3}\,\mu\,(M+m)}{m\,(\sqrt{3}+\mu)}}$$

同时成立，两个物体便能保持平衡。（光滑的横杆处不需要摩擦力，而质点与圆柱体之间的摩擦力已在第一部分处理。）

最后部分

当 $m = 7M$ 及 $\mu = \frac{3}{4}$ 时，质点的条件是 $\tan\theta \le \frac{3}{4}$，即 $\sin\theta \le \frac{3}{5}$。圆柱体的条件是

$$\sin\theta \le \frac{\sqrt{3}\cdot\frac{3}{4}\cdot 8M}{7M\,(\sqrt{3}+\frac{3}{4})} = \frac{24\sqrt{3}}{7(4\sqrt{3}+3)} = \frac{24\sqrt{3}\,(4\sqrt{3}-3)}{7\cdot 39} = \frac{24(4-\sqrt{3})}{91}.$$

哪一个较严格？我们有

$$\frac{24(4-\sqrt{3})}{91} < \frac{3}{5} \iff 120(4-\sqrt{3}) < 273 \iff 207 < 120\sqrt{3} \iff 42849 < 43200,$$

这是成立的。所以圆柱体在平面上滑动才是真正起限制作用的条件（仅仅如此），而能保持平衡的最大 θ 值是

$$\theta_{\max} = \arcsin\frac{24(4-\sqrt{3})}{91} \approx 36.7^\circ,$$

仅略小于 $\arctan\frac{3}{4} \approx 36.9^\circ$。

验算

- 极限情况 $M \to \infty$：圆柱体的条件变得形同虚设，因为非常重的圆柱体总能靠摩擦力固定；只剩下 $\tan\theta \le \mu$。
- 当 $\theta \to 0$ 时，质点位于顶部，它的重量通过 O，而 $F = R = 0$：正如所料，不需要横杆或摩擦力。

- 数值上，随机选取 20 000 组 M, m, a, μ, θ，直接由原始的力向量解出三条平衡方程，结果每次都与方框内的条件吻合；当 $m = 7M$、$\mu = 0.75$ 时，对 θ 作精细扫描，得到 $\theta_{\max} \approx 36.737^\circ$，而 $\sin\theta_{\max} \approx 0.59814$。

常见错误

- 把横杆的力当作水平方向。它垂直于圆柱体表面，所以沿半径方向，在这里是与水平成 30° 向上。
- 忘记横杆的力有向上的分量 $\frac{1}{2}R$，它会减小 S，从而减小平面所能提供的摩擦力。
- 做到 $\theta = \arctan\frac{3}{4}$ 便停下来，没有与圆柱体的条件比较，而后者（稍为）更严格。

STEP 之外

这里用到的技巧，是把质点的接触力换成一个等于其重量的合力（resultant）。由此可以初步窥见一个想法：作用于刚体的任何力系，都可以化简为通过某个选定点的单一力再加上一个力偶（couple）。具有多个摩擦接触点的刚体静力学（statics）是大学一年级的标准课题，而“哪一个接触点先滑动”这个问题，会引申到更微妙的情况：摩擦力并不能单靠方程确定。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对力、力矩和刚体平衡有严谨的论述。

详细解答

STEP I 1990，第 14 题

概率与统计 · 期望值与最优停止

> **STEP I 1990，第 14 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-s1-q14

关键思路

由于每次抽出的球都会放回，各次抽球互相独立（independent），所以每次抽到白球、红球或黑球的概率都固定不变。对 Juggins 而言，我们只需知道连续三次抽球都“存活”的机会，以及一次存活抽球的平均得分。对 Muggins 而言，问题在于取舍：多抽一次，平均可多得少许分数，却要冒失去已累积的全部分数的风险，而这个风险随 N 增大而增大。比较“再抽一次的期望得分”与“就此停手的得分”，便得出一个临界值（threshold）。

解答

在任何一次抽球中，设 W、R、B 分别为抽到白球、红球、黑球的事件。由于球会放回，每次抽球都与其他各次独立，且

$$\mathrm{P}(W)=\frac{5}{10}=\frac{1}{2},\qquad \mathrm{P}(R)=\frac{3}{10},\qquad \mathrm{P}(B)=\frac{2}{10}=\frac{1}{5}.$$

Juggins 的策略

Juggins 最终得零分，当且仅当他首三次抽球中出现黑球。否则三次抽球全都不是黑球，他便停手。因此

$$\mathrm{P}(\text{最终得分为0})=1-\mathrm{P}(\text{三次抽球均非黑球})=1-\left(\frac{4}{5}\right)^3=1-\frac{64}{125}=\frac{61}{125}.$$

至于期望得分（expected score），设 S 为他的最终得分。若首三次抽球全都不是黑球（概率为 $\frac{64}{125}$），则 $S=X_1+X_2+X_3$，其中 X_i 是已知第 i 次抽球不是黑球时该次的得分。已知某次抽球不是黑球，则它是白球的概率为 $\frac{1/2}{4/5}=\frac{5}{8}$，是红球的概率为 $\frac{3/10}{4/5}=\frac{3}{8}$。因此

$$\mathrm{E}(X_i)=1\cdot\frac{5}{8}+2\cdot\frac{3}{8}=\frac{11}{8}.$$

在其余所有情况下 $S=0$。所以

$$\mathrm{E}(S)=\tfrac{64}{125}\cdot 3\cdot\tfrac{11}{8}=\tfrac{264}{125}=2.112.$$

想一想 为甚么这里用 $\frac{5}{8}$ 和 $\frac{3}{8}$，而不是 $\frac{1}{2}$ 和 $\frac{3}{10}$?

我们已经以“存活”为条件 (conditioning), 即已知三次抽球每次都不是黑球, 所以各概率必须重新按比例调整, 使其总和为 1。换一个不用条件概率的说法: $\mathrm{E}(S)$ 是对 $2^3=8$ 个白／红序列求 (概率) $\times$ (得分) 的总和。每次抽球的期望贡献为 $1\cdot\frac{1}{2}+2\cdot\frac{3}{10}=\frac{11}{10}$, 而只有当另外两次抽球也存活时 (概率为 $\frac{16}{25}$), 这份贡献才得以保留。由此同样得出 $3\cdot\frac{11}{10}\cdot\frac{16}{25}=\frac{264}{125}$。

Muggins 的抉择

假设 Muggins 现有 N 分，并再抽一次（然后停手)。他的得分会变成 $N+1$、$N+2$ 或 0，所以

$$\mathrm{E}(\text{再抽一次后的得分})=\tfrac{1}{2}(N+1)+\tfrac{3}{10}(N+2)+\tfrac{1}{5}\cdot 0=\tfrac{4}{5}N+\tfrac{11}{10}.$$

平均而言，再抽一次胜过停手，当且仅当

$$\tfrac{4}{5}N+\tfrac{11}{10}>N\iff\tfrac{11}{10}>\tfrac{1}{5}N\iff N<5.5.$$

抽一次球的平均得益是 $\frac{11}{10}$，平均损失则是 $\frac{1}{5}N$，即抽到黑球的机会乘以所押上的分数。由于得分是整数，这提示了以下策略:

得分为 5 或以下时继续抽球；一旦达到 6 或以上便立即停手。

想一想 只比较“多抽一次”与“停手”，看来有点短视。在 $N\geq 6$ 时先抽一次，并打算稍后再抽，会不会更有利?

不会。得分一旦达到 6 或以上，其后任何一次抽球都只会令得分仍然是 6 或以上（或以 0 分结束游戏)，所以在往后每一个决策点，“只看一次抽球”的比较仍然会说“停手”。从这样的得分出发，每多抽一次都会降低最终得分的期望值，所以由此开始的最佳方案就是立即停手。在 6 分以下，多抽一次已经胜过停手，而可以继续抽下去这个选择只会锦上添花。因此这个临界值法则不仅合理，更是最佳策略。

采用这个策略，设 $V(N)$ 为由 N 分开始的最终得分期望值。则当 $N\geq 6$ 时 $V(N)=N$，而当 $N\leq 5$ 时 $V(N)=\frac{1}{2}V(N+1)+\frac{3}{10}V(N+2)$。由上而下逐一计算:

N	6	5	4	3	2	1	0
$V(N)$	6	5.1	4.35	3.705	3.1575	2.690	2.292

所以 Muggins 平均每局约得 2.29 分，而 Juggins 则是 2.112 分。

验算

- 以电脑精确地列举了所有最多三次抽球的序列，确认 $\mathrm{P}(0) = \frac{61}{125} = 0.488$ 及 $\mathrm{E}(S) = \frac{264}{125}$。以蒙地卡罗（Monte Carlo）方法模拟 2×10^5 局，得出 0.488 和 2.111。
- 在 $N = 5$ 时，抽一次得 $\frac{4}{5} \cdot 5 + 1.1 = 5.1 > 5$。在 $N = 6$ 时则得 $5.9 < 6$。所以转捩点在 5 与 6 之间。
- 对所有策略进行价值迭代（value iteration），得出相同的法则。若改为在 ≥ 5 或 ≥ 7 时停手，平均得分较低，分别为 2.261 和 2.265，而在 6 时停手则为 2.292。

常见错误

- 把 $\mathrm{E}(S)$ 算成 $3 \times \frac{11}{10}$，忘记了其后抽到黑球会把先前的分数一笔勾销。
- 把条件平均值 $\frac{11}{8}$ 再乘以无条件概率，以致把存活因子重复计算了一次。
- 把临界值写成 $N \leq 5.5$，却没有指出 N 是整数，所以法则应是“5 分时抽，6 分时停”。

STEP 之外

Muggins 的问题是一个最优停止（optimal stopping）问题。“当现在停手胜过多走一步时便停手”这个法则称为一步前瞻（one-step look-ahead）法则；只要像本题一样，法则指示停手的那些状态所组成的集合一经进入便无法离开，这个法则就是最优的。同样的想法可引申至马可夫决策过程（Markov decision processes）和动态规划（dynamic programming）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门）涵盖期望值和马可夫链（Markov chain）的背景知识，而 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则有许多类似的首步分析（first-step analysis）。

详细解答

STEP I 1990，第 15 题

概率与统计 • 负二项分布

STEP I 1990，第 15 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-s1-q15

关键思路

第一部分的做法，是把事件拆成两个独立（independent）的部分：最后一次掷币的结果，以及之前所有掷币的结果。至于其余部分，关键在于硬币没有记忆。每当刚掷出一个正面，等待下一个正面的过程便重新开始，所以我们只需知道一个服从几何分布（geometric distribution）的等待时间有多大机会是偶数或奇数。

解答

以下一律记 $q = 1 - p$。各次掷币互相独立，每次掷出正面的概率为 p。

第 k 个正面的分布

设 T_k 为第 k 个正面出现时的掷币次数。事件 $\{T_k = n\}$ 发生，当且仅当

- 第 n 次掷币是正面，并且
- 在第 $1, 2, \ldots, n-1$ 次掷币之中恰好有 $k-1$ 个正面（因而有 $n-k$ 个反面）。

两个条件缺一不可。假如第 n 次掷币是反面，第 k 个正面便会在第 n 次掷币之前或之后出现，而不是在这一次。假如头 $n-1$ 次掷币中的正面数目不同，第 n 次掷币便不会给出第 k 个正面。反过来，两个条件合起来亦已足够。

这两个事件涉及的掷币次数互不重叠，所以它们是独立的。第一个事件的概率是 p。至于第二个事件，那 $k-1$ 个正面在头 $n-1$ 次掷币中的位置共有 $\binom{n-1}{k-1}$ 种选法。每一种排列都是由 $k-1$ 个正面和 $n-k$ 个反面组成的一个特定序列，根据独立性，其概率为 $p^{k-1}q^{n-k}$，

而不同的排列是互斥（mutually exclusive）的。因此

$$\mathrm{P}(T_k = n) = p \cdot \binom{n-1}{k-1} p^{k-1} q^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \qquad n \geq k,$$

这正是所要证明的结果。

由偶数到偶数

设 $G = T_k - T_{k-1}$ 为第 $(k-1)$ 个正面出现之后还需要再掷的次数（其中 $T_0 = 0$）。第 T_{k-1} 次掷币之后的各次掷币，与此前发生的一切互相独立。所以 G 与 T_{k-1} 独立，并且服从几何分布

$$\mathrm{P}(G = g) = q^{g-1} p, \qquad g = 1, 2, \ldots$$

（即第一部分中 $k = 1$ 的情况）。

> ? **想一想** 如果 T_{k-1} 是偶数，T_k 在甚么情况下是偶数?

由于 $T_k = T_{k-1} + G$，如果 T_{k-1} 是偶数，则 T_k 是偶数当且仅当 G 是偶数。因为 G 与 T_{k-1} 独立，以 T_{k-1} 为偶数作为条件（conditioning）并不会改变 G 的分布。因此

$$\mathrm{P}(T_k \text{ 为偶数} \mid T_{k-1} \text{ 为偶数}) = \mathrm{P}(G \text{ 为偶数}).$$

把等比级数（geometric series）求和，

$$\mathrm{P}(G \text{ 为偶数}) = p\left(q + q^3 + q^5 + \cdots\right) = \frac{pq}{1-q^2} = \frac{pq}{(1-q)(1+q)} = \frac{q}{1+q},$$
$$\mathrm{P}(G \text{ 为奇数}) = p\left(1 + q^2 + q^4 + \cdots\right) = \frac{p}{1-q^2} = \frac{1}{1+q}.$$

所以所求的概率是

$$\frac{q}{1+q} = \frac{1-p}{2-p}.$$

答案与 k 无关。

恰好有两个正面落在偶数次掷币

现在一直掷到第三个正面出现为止。设 $G_1 = T_1$、$G_2 = T_2 - T_1$ 和 $G_3 = T_3 - T_2$ 为三段等待时间。根据同样的无记忆性（memoryless）论证，它们互相独立，而且各自服从上述的几何分布。记

$$e = \mathrm{P}(G_i \text{ 为偶数}) = \frac{q}{1+q}, \qquad o = \mathrm{P}(G_i \text{ 为奇数}) = \frac{1}{1+q}.$$

我们从第 0 次掷币开始，这是偶数。间隔为偶数时，当前位置的奇偶性（parity）保持不变；间隔为奇数时，奇偶性便会改变。下表列出 (G_1, G_2, G_3) 的八种可能奇偶组合，以及相应的 (T_1, T_2, T_3) 的奇偶性：

G_1	G_2	G_3	T_1	T_2	T_3	偶数次的正面数目
偶	偶	偶	偶	偶	偶	3
偶	偶	奇	偶	偶	奇	**2**
偶	奇	偶	偶	奇	奇	1
偶	奇	奇	偶	奇	偶	**2**
奇	偶	偶	奇	奇	奇	0
奇	偶	奇	奇	奇	偶	1
奇	奇	偶	奇	偶	偶	**2**
奇	奇	奇	奇	偶	奇	1

其中有三种情况恰好有两个正面落在偶数次掷币，它们的概率分别是 e^2o、eo^2 和 o^2e。由于各段间隔互相独立，所求的概率是

$$\begin{aligned} e^2o + 2eo^2 = eo(e+2o) &= \frac{q}{(1+q)^2}\cdot\frac{q+2}{1+q} \\ &= \frac{q(2+q)}{(1+q)^3} = \frac{(1-p)(3-p)}{(2-p)^3}. \end{aligned}$$

验算

- 验算第一部分：把 $\mathrm{P}(T_k = n)$ 对 $n \geq k$ 求和，结果应该是 1。这就是负二项级数 $\sum_{n\geq k}\binom{n-1}{k-1}q^{n-k} = (1-q)^{-k}$。
- 两个奇偶概率 $\frac{q}{1+q}$ 与 $\frac{1}{1+q}$ 之和为 1。当 p 接近 1 时，第一个正面几乎必定立即出现，即落在奇数次掷币，所以 $\mathrm{P}(G \text{ 为偶数}) \to 0$，与公式给出的结果一致。
- 对于公平硬币（$p = \frac{1}{2}$），最后的答案是 $\frac{(1/2)(5/2)}{(3/2)^3} = \frac{10}{27} \approx 0.370$。一次进行了 2×10^5 轮的模拟得出 0.369。直接对三段等待时间求和，所得结果在 $p = 0.3$、0.5 和 0.8 时均与公式吻合，误差在 10^{-9} 以内。

常见错误

- 在第一部分使用 $\binom{n}{k}$。最后一次掷币必定是正面，所以只有头 $n-1$ 次掷币可以自由安排。
- 在第二部分把负二项分布对偶数的 n 求和。利用无记忆的间隔，便可完全避开这一步。
- 在最后一部分，计算落在偶数位置的正面时，把各个位置当作互相独立。它们并不独立，因为每个位置都是由前一个位置推算出来的。

STEP 之外

第 k 个正面出现的时间服从负二项（negative binomial）分布。把它写成若干个独立几何间隔之和，正是泊松（Poisson）过程的离散版本：泊松过程的到达时间，同样是由独立的指数等待时间构成的。最后一部分记录奇偶性的做法，其实是一个双状态的马可夫链（Markov chain），这是大学一年级概率课程中很早便会遇到的课题。Grimmett and Welsh 的 *Probability: An Introduction*（概率入门教科书）涵盖了以上所有内容。

详细解答

STEP I 1990，第 16 题

概率与统计 · 三角分布与假设检验

STEP I 1990，第 16 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-s1-q16

关键思路

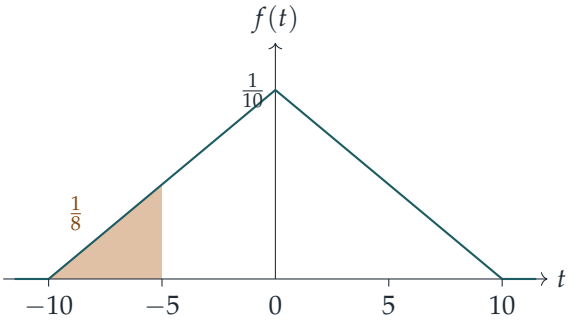
概率密度函数（probability density function）的图像是一个三角形，所以我们需要的每一个概率都是某个小三角形的面积，而累积分布函数（cumulative distribution function）会替我们完成所有工作。第二点要留意的是，只有 8 时的那班巴士才有可能在上午 8 时前后到站：7 时的那班巴士在 7.10 之前已经离开，而 9 时的那班巴士不可能在 8.50 之前到达。所以早上“赶上巴士”的真正意思是“比 8 时的那班巴士更早到站”。最后一部分于是变成一个只有一个参数的二项模型（binomial model）和一个显著性检定（significance test）。

解答

T 的密度函数

密度由 $t=-10$ 处的 0 线性上升至 $t=0$ 处的峰值，然后以相同的速率下降至 $t=10$ 处的 0，所以它在 $[-10,10]$ 上的形式是 $f(t)=k(10-|t|)$，在其他地方则为零。总面积等于一个底为 20、高为 $10k$ 的三角形的面积，所以 $\frac{1}{2}\cdot 20\cdot 10k=1$，由此得 $k=\frac{1}{100}$：

$$f(t)=\begin{cases}\dfrac{10+t}{100}, & -10\le t\le 0,\\ \dfrac{10-t}{100}, & 0\le t\le 10,\\ 0, & \text{其他情况。}\end{cases}$$



为方便稍后使用，先求左半部分的累积分布函数：它是一个底为 $a+10$、高为 $f(a)$ 的三角形的面积，

$$\mathrm{P}(T<a)=\tfrac{1}{2}(a+10)\cdot\frac{a+10}{100}=\frac{(a+10)^2}{200},\qquad -10\le a\le 0,$$

而根据关于 $t=0$ 的对称性，当 $0\le a\le 10$ 时，$\mathrm{P}(T>a)=\dfrac{(10-a)^2}{200}$。图中阴影部分的面积是 $\mathrm{P}(T<-5)=\frac{25}{200}=\frac{1}{8}$。

在上午 8.10 或之前赶上巴士

我们假设我会在车站一直等到有巴士来为止。预定 8 时到站的巴士在上午 8 时之后 T 分钟到达，其中 $-10\le T\le 10$；上一班巴士在 7.10 之前已经离开，而下一班不可能在 8.50 之前到达。所以，我能在 8.10 或之前赶上巴士，当且仅当我比 8 时的那班巴士更早到达车站，而这时 8.10 的期限自然会得到满足。

如果我在上午 8 时之后 a 分钟到达车站，那么当 $T>a$ 时我便能赶上这班车（恰好出现 $T=a$ 的机会为零）。三种情况分别给出

$$\begin{aligned}
7{\cdot}55,\ a=-5:&\quad \mathrm{P}(T>-5)=1-\tfrac{25}{200}=\tfrac{7}{8},\\
7{\cdot}58,\ a=-2:&\quad \mathrm{P}(T>-2)=1-\tfrac{64}{200}=\tfrac{17}{25},\\
8{\cdot}02,\ a=2:&\quad \mathrm{P}(T>2)=\tfrac{64}{200}=\tfrac{8}{25}.
\end{aligned}$$

我在 7·55 到达车站的概率是 0.5，在 7·58 到达的概率是 $0.5\times 0.8=0.4$，在 8.02 到达的概率是 $0.5\times 0.2=0.1$。我睡过头与否和巴士毫无关系，所以根据全概率公式（law of total probability），

$$\begin{aligned}
\mathrm{P}(\text{在 8.10 或之前赶上巴士})&=0.5\cdot\tfrac{7}{8}+0.4\cdot\tfrac{17}{25}+0.1\cdot\tfrac{8}{25}\\
&=0.4375+0.272+0.032=\frac{1483}{2000}=0.7415.
\end{aligned}$$

新时钟是否不准?

首先假设时钟是准确的，那么我每天早上都在 7.55 到达车站。设 T_8 和 T_9 分别为 8 时和 9 时那两班巴士的到站时间，以整点之后的分钟数计算；它们互相独立（independent），各自的密度都是 f。我在上午 9 时之后仍在等车，当且仅当我错过了 8 时的巴士（$T_8 < -5$）而且 9 时的巴士迟到（$T_9 > 0$）。所以在任何一个早上，

$$p = \mathrm{P}(T_8 < -5)\,\mathrm{P}(T_9 > 0) = \tfrac{1}{8} \cdot \tfrac{1}{2} = \tfrac{1}{16}.$$

? **想一想** 检定应该朝哪个方向进行：5 次是“太多”还是“太少”?

时钟调校错误，可以令 p 向任何一个方向偏移。假设时钟快了 d 分钟。那么我到达车站的真实时间是 $7.55 - d$，而时钟显示 9.00 时的真实时间是 $9.00 - d$，所以等到“上午 9 时”之后的概率变成

$$g(d) = \mathrm{P}(T_8 < -5 - d)\,\mathrm{P}(T_9 > -d).$$

时钟快了（$d > 0$）会令我更早到站，所以我很少会错过 8 时的巴士，于是 $g(d) < \frac{1}{16}$；时钟稍为慢了则会产生相反的效果（例如 $g(-2.5) \approx 0.079$）。所以，合理而不偏颇的做法是采用双尾检定（two-sided test）。

设 X 为这 100 个早上之中，按新时钟计算我等到上午 9 时之后的早上数目。不同巴士的到站时间互相独立，而我每天都在同一时间到站，所以各个早上是独立试验，并且

$$H_0：\text{时钟准确，}\quad X \sim \mathrm{B}\big(100, \tfrac{1}{16}\big);\qquad H_1：\text{时钟调校错误，因此} p \neq \tfrac{1}{16}.$$

在 H_0 之下，我们预期有 $\mathrm{E}(X) = 6.25$ 次，而 $\mathrm{Var}(X) = 100 \cdot \frac{1}{16} \cdot \frac{15}{16} \approx 5.86$，所以标准差（standard deviation）约为 2.42。观察所得的 5 只是略低于平均值。准确地计算，

$$\mathrm{P}(X \le 5) = \sum_{k=0}^{5} \binom{100}{k} p^k (1-p)^{100-k} \approx 0.40 \qquad \text{其中} p = \tfrac{1}{16},$$

而采用连续性校正（continuity correction）的正态近似（normal approximation）则给出 $\Phi\big((5.5 - 6.25)/2.42\big) = \Phi(-0.31) \approx 0.38$。这与 0.025（5% 双尾检定的下尾）相去甚远，事实上与任何常用的显著性水平都相去甚远。所以我们不拒绝 H_0：**没有证据显示新时钟调校错误**。

这并不等于有证据显示时钟是准确的。事实上 $g(d) = g(-5 - d)$，因为密度是对称的（把两班巴士的角色对调即可），所以一个慢了 5 分钟的时钟所给出的 $p = \frac{1}{16}$，与准确的时钟完全相同。这类数据无论收集多少个早上，都无法把这两种情况区分开来。

验算

▶ 密度的面积为 1，而三角形公式与对 f 作数值积分的结果吻合，例如 $\mathrm{P}(T < -2) = 0.32$。

- 答案 0.7415 介乎最好的情况 $\frac{7}{8}$（每次都在 7.55 到达）与最坏的情况 0.32（每次都在 8.02 到达）之间，这是理所当然的。模拟 10^6 个早上，所得结果为 0.741。
- 模拟 10^6 对 (T_8, T_9)，所得结果为 0.0622，对应 $\frac{1}{16} = 0.0625$；而准确的二项尾部概率 $\mathrm{P}(X \leq 5)$ 的数值为 0.4002。

常见错误

- 忘记了要等到 9 时之后，必须先错过 8 时的巴士：答案 $p = \frac{1}{2}$（只考虑“9 时的巴士迟到”）忽略了第一个因子。
- 以为“在 8.10 或之前赶上巴士”需要另外检查 $T \leq 10$：8 时的巴士必定在 8.10 或之前到达，所以只有 $T > a$ 才是关键。
- 数据显示次数略为偏少，却针对“太多”进行单尾检定（one-tailed test）；时钟不准可以令 p 向任何一个方向移动。

STEP 之外

三角形密度正是两个独立均匀随机变量（uniform random variables）之和的分布，这是卷积（convolution）的第一个例子，而大学一年级的概率课程正是用卷积来求和的密度。最后一部分让我们浅尝统计学上的可识别性（identifiability）：对称性 $g(d) = g(-5 - d)$ 意味着两个不同的参数值会产生完全相同的数据，所以没有任何检定能够把它们区分开来。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）对卷积和随机变量之和有清晰的讲解。

详细解答

STEP II 1990，第 1 题

纯数学 · 多项式不等式

> **STEP II 1990，第 1 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-s2-q1

关键思路

没有实根（real roots）的四次多项式（quartic），其实就是一个恒为正的四次多项式，所以我们要设法把它拆成若干个我们确知为非负（non-negative）的部分：完全平方（perfect square），或者在某个区间（interval）上正负号一目了然的乘积。没有哪一种拆法能处处适用，但不同的拆法在不同的区间上奏效，合起来便覆盖整条实数线。第一部分把这些部件直接交给我们；到了最后一部分，就要自己动手去找。

解答

两个非负的表达式

把四次式因式分解（factorise），并对二次式配方（completing the square），

$$x^4 - 2x^3 + x^2 = x^2\left(x^2 - 2x + 1\right) = x^2(x-1)^2,$$
$$x^2 - 8x + 17 = (x-4)^2 + 1.$$

第一个是两个平方的乘积，所以对任何实数 x 都 ≥ 0。第二个是一个平方加 1，所以至少是 1，自然也是非负的。

第一个方程

设 $p(x) = x^4 - 2x^3 + x^2 - 8x + 17$。留意 p 并不是上面两个表达式之和（那个和含有 $2x^2$，而不是 x^2），所以不能简单地把两者相加。我们改为在每个区间上用不同的方式拆开 p，从而证明 $p(x) > 0$ 处处成立。

想一想 题目给了两个非负的表达式，但它们加起来并不等于 $p(x)$。我们还可以怎样利用它们?

两者都可以在实数线的某一段上“借用”：在一个区间上，我们保留四次式 $x^2(x-1)^2$，然后直接处理余下的 $17-8x$；在另一个区间上，我们保留二次式 $(x-4)^2+1$，然后直接处理余下的 x^4-2x^3。

情况 $x \le 0$。 这时 $x^4 \ge 0$、$-2x^3 \ge 0$（因为 $x^3 \le 0$）、$x^2 \ge 0$ 以及 $-8x \ge 0$。所以 $p(x)$ 的每一项都是非负的，再加上常数项 17，总和至少是 17。因此 $p(x) > 0$。

情况 $0 < x \le 2$。 写成

$$p(x) = x^2(x-1)^2 + (17-8x).$$

第一部分 ≥ 0，而由于 $x \le 2$，我们有 $17-8x \ge 17-16 = 1$。所以 $p(x) \ge 1 > 0$。

情况 $x > 2$。 写成

$$p(x) = x^3(x-2) + \left(x^2 - 8x + 17\right) = x^3(x-2) + (x-4)^2 + 1.$$

由于 $x > 2$，x^3 和 $x-2$ 都是正数，所以 $x^3(x-2) > 0$；而 $(x-4)^2+1 \ge 1$。所以 $p(x) > 1 > 0$。

这三个区间覆盖了所有实数，而在每个区间上都有 $p(x) > 0$，所以 $p(x) = 0$ 没有实根。

第二个方程

设 $q(x) = x^4 - x^3 + x^2 - 4x + 4$。我们沿用同样的方法，不过这次要自己选择拆法。

想一想 应该在哪里把实数线分段？又应该把哪些项配成一组?

我们希望每一组在各自的区间上都有明显的正负号。$4-4x = 4(1-x)$ 这一组在 $x \le 1$ 时是非负的。当 $x \ge 1$ 时，$x^4 - x^3 = x^3(x-1)$ 这一组是非负的，而余下的 $x^2 - 4x + 4 = (x-2)^2$ 是一个完全平方。所以最自然的分界点是 $x = 1$。

情况 $x \le 1$。 写成

$$q(x) = \left(x^4 - x^3 + x^2\right) + 4(1-x) = x^2\left[\left(x - \tfrac{1}{2}\right)^2 + \tfrac{3}{4}\right] + 4(1-x).$$

第一项 ≥ 0，而且只在 $x = 0$ 时才等于零（方括号内的值至少是 $\frac{3}{4}$）。第二项因 $x \le 1$ 而 ≥ 0，而且只在 $x = 1$ 时才等于零。两项不可能同时为零，所以 $q(x) > 0$。

情况 $x \geq 1$。 写成

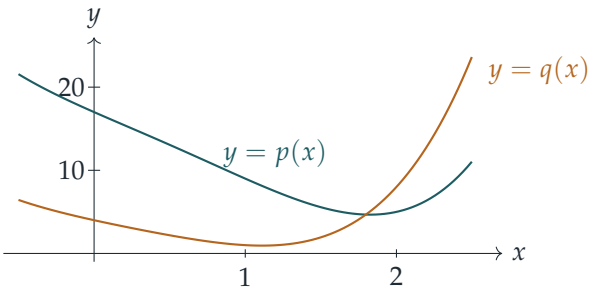
$$q(x) = x^3(x-1) + \left(x^2 - 4x + 4\right) = x^3(x-1) + (x-2)^2.$$

当 $x \geq 1$ 时，x^3 和 $x-1$ 都是非负的，所以第一项 ≥ 0，而且只在 $x = 1$ 时才等于零。第二项是一个平方，只在 $x = 2$ 时才等于零。两者同样不可能同时为零，所以 $q(x) > 0$。

所以对所有实数 x 都有 $q(x) > 0$，而 $q(x) = 0$ 没有实根。

验算

- 把每一种拆法展开，确保它确实还原成 p 或 q。例如 $x^3(x-1) + (x-2)^2 = x^4 - x^3 + x^2 - 4x + 4$。
- 在 $-5 \leq x \leq 8$ 的范围内以 10^{-4} 为步长作数值扫描，得出 p 的最小值约为 4.67（在 $x \approx 1.83$ 附近），而 q 的最小值约为 0.94（在 $x \approx 1.11$ 附近）。两者都明显是正数，正如下图所示。



常见错误

- 把题目给出的两个表达式相加，然后声称其和就是 $p(x)$：其实并不是，因为 x^2 项对不上。
- 由“非负项加非负项”就断定 $q(x) > 0$。这样只能得出 $q(x) \geq 0$；要排除有根的可能，必须证明两项不可能同时为零。
- 以为 $-2x^3$ 或 $-8x$ 永远是负数。当 $x \leq 0$ 时，它们非但无害，反而有帮助。

STEP 之外

把多项式（polynomial）写成平方和（sum of squares）来证明它是正的，是一个很有份量的课题。每一个非负的单变量多项式，都是两个实系数多项式的平方之和，但希尔伯

特（Hilbert）证明了这个结论在多变量的情况下并不成立，由此引出了他的第十七个问题。时至今日，人们在最优化（optimisation）问题中会利用电脑去搜索这类分解。如果想学习更多这一类不等式技巧，Siklos 的 *Advanced Problems in Mathematics* 是下一步的好选择。

详细解答

STEP II 1990，第 2 题

纯数学 · 托勒密定理

STEP II 1990，第 2 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-s2-q2

关键思路

第一部分是一条三角恒等式（trigonometric identity），只要把每个正弦之积都化成余弦之差，它便迎刃而解；然后利用条件 $A+B+C+D=\pi$ 改写 $\cos(A+C)$ 和 $\cos(A-C)$。第二部分是托勒密定理（Ptolemy's theorem）。圆心 O 就是提示：半径为 r 的圆上每条弦（chord）的长度都是 $2r\sin$(所对角的一半，而该角的顶点为O)，而围绕 O 的四个半角之和为 π，这正是第一部分的条件。

解答

恒等式

我们三次使用积化和差公式（product-to-sum formula）

$$\sin X\sin Y=\tfrac{1}{2}\big[\cos(X-Y)-\cos(X+Y)\big]$$

。在左方，

$$\sin(A+B)\sin(A+D)=\tfrac{1}{2}\big[\cos(B-D)-\cos(2A+B+D)\big],$$
$$\sin B\sin D=\tfrac{1}{2}\big[\cos(B-D)-\cos(B+D)\big],$$

所以 $\cos(B-D)$ 两项互相抵消，得

$$\sin(A+B)\sin(A+D)-\sin B\sin D=\tfrac{1}{2}\big[\cos(B+D)-\cos(2A+B+D)\big].$$

在右方，

$$\sin A\sin C=\tfrac{1}{2}\big[\cos(A-C)-\cos(A+C)\big].$$

现在利用 $C = \pi - A - B - D$。由此可得

$$A + C = \pi - (B + D), \qquad \text{所以} \quad \cos(A + C) = -\cos(B + D),$$
$$A - C = (2A + B + D) - \pi, \qquad \text{所以} \quad \cos(A - C) = -\cos(2A + B + D),$$

当中用到 $\cos(\pi - \theta) = -\cos\theta$ 和 $\cos(\theta - \pi) = -\cos\theta$。因此

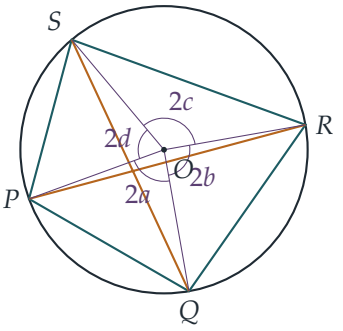
$$\sin A \sin C = \tfrac{1}{2}\big[-\cos(2A + B + D) + \cos(B + D)\big],$$

这与左方相同。恒等式得证。

圆

设圆的半径为 r。沿圆周由 P 到 Q 到 R 到 S 再回到 P，设弧（arc）PQ、QR、RS、SP 在 O 所对的角分别为 $2a$、$2b$、$2c$、$2d$。由于各点按此次序排列，这四段弧互不重叠，并合成整个圆，所以

$$a, b, c, d > 0, \qquad 2a + 2b + 2c + 2d = 2\pi, \qquad \text{即} \quad a + b + c + d = \pi.$$



想一想 若一条弦所对的弧在圆心所对的角为 2θ，这条弦有多长?

若 $0 < 2\theta \le \pi$，由 O 和弦的两个端点组成的三角形是等腰（isosceles）三角形，两腰长 r，顶角为 2θ。由 O 所作的垂线同时平分弦和顶角，所以弦长为 $2r\sin\theta$。若 $\pi < 2\theta < 2\pi$，则三角形内位于 O 的角为 $2\pi - 2\theta$，弦长为 $2r\sin(\pi - \theta) = 2r\sin\theta$。所以在任何情况下

$$\text{弦长} = 2r\sin\theta.$$

把这个结果应用于四条边，以及两条对角线（diagonal）（弧 $P \to Q \to R$ 所对的角为 $2(a + b)$，弧 $Q \to R \to S$ 所对的角为 $2(b + c)$），便得

$$PQ = 2r\sin a, \quad QR = 2r\sin b, \quad RS = 2r\sin c, \quad PS = 2r\sin d,$$

$$PR = 2r\sin(a + b), \qquad QS = 2r\sin(b + c).$$

注意 $a+b$ 或 $b+c$ 可能大于 $\frac{\pi}{2}$；这就是我们需要为大于半圆的弧验证该公式的原因。

除以 $4r^2$ 之后，要证明的结果变成

$$\sin a \sin c + \sin b \sin d = \sin(a+b)\sin(b+c),$$

或等价地，

$$\sin(b+a)\sin(b+c) - \sin a \sin c = \sin b \sin d.$$

这正是第一部分的恒等式，其中 $A=b$、$B=a$、$D=c$ 及 $C=d$；这样代入是容许的，因为 $b+a+d+c=\pi$。再乘回 $4r^2$，得

$$PQ \times RS + QR \times PS = PR \times QS,$$

这就是所要证明的结果。

验算

- 内接于半径为 $\frac{1}{\sqrt{2}}$ 的圆的正方形，边长为 1，对角线长 $\sqrt{2}$：$1\cdot 1 + 1\cdot 1 = 2 = \sqrt{2}\cdot\sqrt{2}$。
- 对任何边长为 p 和 q 的长方形，托勒密定理变成 $p^2+q^2 = (\text{对角线})^2$：毕氏定理（Pythagoras's theorem）是它的特殊情况。对于 3×4 的长方形，$9+16=25=5\times 5$。
- 以 10^5 组随机的 A, B, D 值（其中 $C=\pi-A-B-D$）对恒等式作数值检验，最大误差约为 5×10^{-15}；另外 10^5 个随机的圆内接四边形（cyclic quadrilateral）均满足最后的结果，相对误差在 10^{-15} 以内。若各点的次序被打乱（例如沿圆周依次为 P, R, Q, S），等式便不成立，可见次序确实重要。

常见错误

- 使用以圆上某点为顶点的圆周角（inscribed angle），却没有说明它们对着哪一段弧。按题目的提示改用圆心角，便可避免这个问题。
- 忘记对角线所截的弧可以大于半圆；公式 $2r\sin\theta$ 依然成立，但需要上文那一行的理由来支持。
- 把错误的角配对到 A, B, C, D：A 和 C 必须是只在右方出现的两个角，在这里即 b 和 d。

STEP 之外

托勒密定理有一个影响深远的推广：对平面上任意四点，$PQ\cdot RS + QR\cdot PS \geq PR\cdot QS$，而等号成立当且仅当这些点依次位于同一个圆（或同一条直线）上。有一个巧妙的证明：把各

点视为复数(complex numbers)p, q, r, s,留意代数恒等式 $(p-q)(r-s)+(q-r)(p-s)=(p-r)(q-s)$，再应用三角不等式（triangle inequality）；等号成立的情况，正是复平面的几何派上用场之处。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是欣赏这类论证的绝佳读物，书中也介绍了反演（inversion），而反演同样能简洁地证明这个结果。

详细解答

STEP II 1990，第 3 题

纯数学 · 三次曲线的描绘

STEP II 1990，第 3 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-s2-q3

关键思路

把这个三次式（cubic）写成 $f(x) = x(x^2 - 2bx + c^2)$。两个判别式（discriminant）决定了一切：$f'(x)$ 的判别式告诉我们曲线有没有转向点（turning point），而二次因式的判别式则告诉我们曲线除了原点之外，还会不会在别处与 x 轴相交。题目的七种情况，不过是 b 相对于两个临界值 $\frac{\sqrt{3}}{2}c$ 和 c 的各种位置。三次曲线一旦画好，$y^2 = f(x)$ 几乎是顺手可得：只保留 $f(x) \geq 0$ 的部分，取平方根，再对 x 轴作反射。

解答

各种情况都成立的事实

设 $f(x) = x^3 - 2bx^2 + c^2x = x(x^2 - 2bx + c^2)$。则

$$f'(x) = 3x^2 - 4bx + c^2, \qquad f''(x) = 6x - 4b.$$

- *转向点*。f' 的判别式是 $16b^2 - 12c^2 = 4(4b^2 - 3c^2)$。所以若 $2b > c\sqrt{3}$，f' 有两个相异实零点；若 $2b = c\sqrt{3}$，有一个重零点；若 $2b < c\sqrt{3}$，则没有零点。有零点时，它们的和是 $\frac{4}{3}b \geq 0$，积是 $\frac{1}{3}c^2 \geq 0$，所以两个都不是负数：所有转向点都位于 $x \geq 0$ 的范围内。
- *根*。$f(0) = 0$ 恒成立。二次因式是 $(x - b)^2 + c^2 - b^2$，所以若 $b < c$，它没有实零点；若 $b = c$，有二重零点 $x = b$；若 $b > c$，则有两个零点 $b \pm \sqrt{b^2 - c^2}$。
- *负的 x*。当 $x < 0$ 时，x^3、$-2bx^2$ 和 c^2x 每一项都 ≤ 0，而且 $x^3 < 0$，所以 $f(x) < 0$。曲线在 y 轴左边的部分全部位于 x 轴下方。
- *形状*。当 $x \to \pm\infty$ 时 $f(x) \to \pm\infty$；曲线在原点的斜率（gradient）是 $f'(0) = c^2$；唯

一的拐点（point of inflection）在 $x=\frac{2}{3}b$，f'' 在该处变号。

由于只要 $c>0$ 就有 $c\sqrt{3}<2c$，所以对固定的 c，各种情况是按 b 由小至大的次序排列的：先是没有转向点，然后是驻拐点，再然后是转向点都在横轴上方，接着是曲线在一个根处与横轴相切，最后是三个根。

三次曲线

(i) $2b<c\sqrt{3}$。这时对所有 x 都有 $f'(x)>0$（它的判别式是负数，首项系数是正数），所以 f 严格递增。此外 $b<\frac{\sqrt{3}}{2}c<c$（注意这里 $c>0$），所以原点是唯一的根。曲线以斜率 c^2 上升穿过原点，并在 $x=\frac{2}{3}b$ 处有一个拐点。

(ii) $2b=c\sqrt{3}\neq 0$。这时 $f'(x)=3(x-\frac{2}{3}b)^2\geq 0$, 所以 f 仍然递增, 但在 $x=\frac{2}{3}b=c/\sqrt{3}$ 处有一瞬间是水平的。这也正是 $f''=0$ 的地方，所以它是一个驻拐点（stationary point of inflection）。利用 $c^2=\frac{4}{3}b^2$,

$$f\left(\tfrac{2}{3}b\right)=\tfrac{2}{3}b\left(\tfrac{4}{9}b^2-\tfrac{4}{3}b^2+\tfrac{4}{3}b^2\right)=\tfrac{8}{27}b^3=\tfrac{\sqrt{3}}{9}c^3>0.$$

同样有 $b<c$，所以原点是唯一的根。

(iii) $c\sqrt{3}<2b<2c$。这时 f' 有两个零点 $x_1<x_2$，两个都是正数，于是在 $x_1=\frac{1}{3}\left(2b-\sqrt{4b^2-3c^2}\right)$ 处有极大点（maximum），在 $x_2=\frac{1}{3}\left(2b+\sqrt{4b^2-3c^2}\right)$ 处有极小点（minimum）。

> ? **想一想** 极小点会不会跌到 x 轴下方?

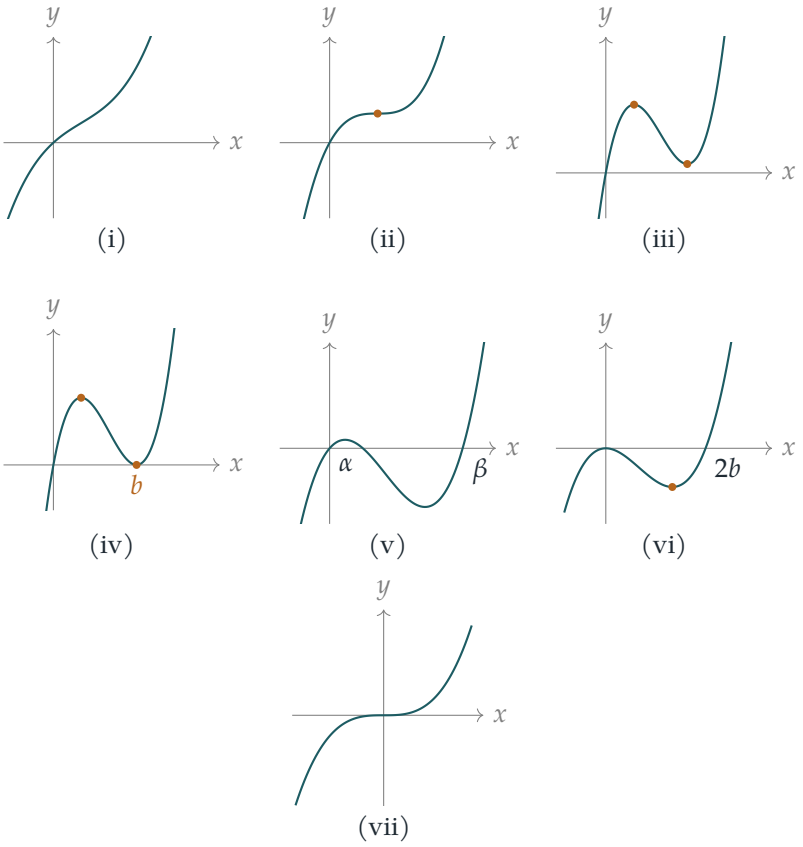
不会。由于 $b<c$，因式 $(x-b)^2+c^2-b^2$ 对每一个 x 都是正数，所以对所有 $x>0$ 都有 $f(x)>0$。因此两个转向点都在横轴上方，曲线仍然只在原点穿过横轴。它先上升，稍微回落，然后一直上升。

(iv) $b=c\neq 0$。这时 $f(x)=x(x-b)^2$。$x=b$ 是二重根，表示曲线在该处与横轴相切；这就是极小点 $(b,0)$。$f'(x)=(x-b)(3x-b)$ 的另一个零点给出极大点 $\left(\frac{1}{3}b,\frac{4}{27}b^3\right)$。

(v) $b>c>0$。这时有三个根 $0<\alpha<\beta$，其中 $\alpha=b-\sqrt{b^2-c^2}$，$\beta=b+\sqrt{b^2-c^2}$（因为 $\sqrt{b^2-c^2}<b$，所以两个都是正数）。在相邻两个根之间曲线必定要转向，所以极大点在 $(0,\alpha)$ 内，该区间上 $f>0$；极小点在 (α,β) 内，该区间上 $f<0$。所以 $f\geq 0$ 恰好在 $[0,\alpha]$ 和 $[\beta,\infty)$ 上成立。

(vi) $c=0,b\neq 0$。这时 $f(x)=x^2(x-2b)$，$f'(x)=x(3x-4b)$。原点是二重根，所以它是极大点，曲线在该处与横轴相切；另外在 $\left(\frac{4}{3}b,-\frac{32}{27}b^3\right)$ 处有极小点。曲线在 $x=2b$ 处穿过横轴。

(vii) $b=c=0$。这就是 $y=x^3$，在原点有一个驻拐点。



这些草图取 $c=1$，在 (i)–(iv) 中 $b=\frac{1}{2}$, $\frac{\sqrt{3}}{2}$, 0.99, 1，在 (v) 中 $(b,c)=(1,0.8)$，在 (vi) 中 $b=0.6$。各小图的纵向比例并不相同，好让 (iii) 和 (iv) 中细小的起伏能够看得见。

曲线 $y^2=f(x)$

以下三条法则可把每条三次曲线变成曲线 $y^2=f(x)$：

- 在 $f(x)<0$ 的地方没有点，尤其是 $x<0$ 的地方没有点；
- 在 $f(x)\geq 0$ 的地方，$y=\pm\sqrt{f(x)}$，所以曲线对 x 轴对称，并且在 f 最大的地方最高；
- 求导可得 $2y\dfrac{\mathrm{d}y}{\mathrm{d}x}=f'(x)$。

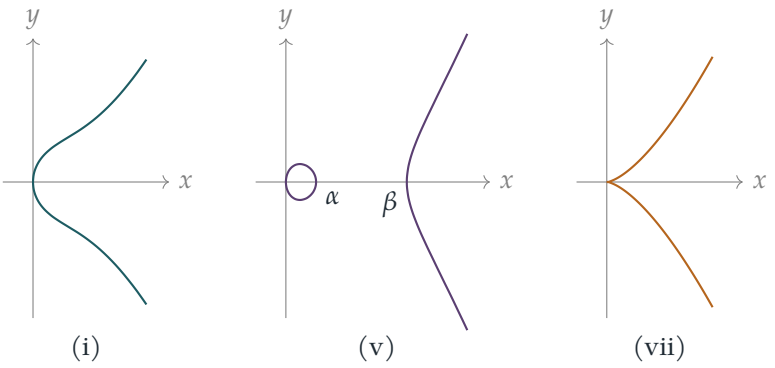
想一想 曲线与 x 轴相交的地方是甚么样子的?

在 f 的单根(simple root)x_0 处,我们有 $f'(x_0)\neq 0$,而当 $y\to 0$ 时 $\dfrac{\mathrm{d}x}{\mathrm{d}y}=\dfrac{2y}{f'(x)}\to 0$。所以该处的切线(tangent)是垂直的,上下两半平滑地接合。在这样的根附近 $y^2\approx f'(x_0)(x-x_0)$，那是一条横放的抛物线。重根（repeated root）的表现则有所不同，情况 (vii) 就是例子。

情况 (i)。$f(x) \geq 0$ 恰好当 $x \geq 0$ 时成立，而 0 是单根（$f'(0) = c^2 > 0$）。所以曲线是单独一条无界的分支。它以垂直切线离开原点，在该处看来像 $y^2 \approx c^2 x$；由于 f 递增，$|y| = \sqrt{f(x)}$ 稳步增加，最终增长得像 $x^{3/2}$。上半部分起初是凹（concave）的，像 $\sqrt{x}$，最后变成凸（convex）的，像 $x^{3/2}$，所以它有一个拐点（由对称性，下半部分也有）。

情况 (v)。$f(x) \geq 0$ 在 $[0, \alpha]$ 和 $[\beta, \infty)$ 上成立，而三个根都是单根。所以曲线分成互不相连的两部分：一个封闭的卵形线，经过 $(0,0)$ 和 $(\alpha, 0)$，在这两点的切线都是垂直的，并且在三次曲线极大点的 x 坐标处最宽（就 y 方向而言）；以及一条无界的分支，从 $(\beta, 0)$ 以垂直切线出发，像情况 (i) 的曲线那样向右张开。在三次曲线取负值的空隙 $\alpha < x < \beta$ 中，甚么都没有。

情况 (vii)。$y^2 = x^3$，所以 $x \geq 0$ 且 $y = \pm x^{3/2}$。这里当 $x \to 0$ 时 $\mathrm{d}y/\mathrm{d}x = \pm\frac{3}{2}x^{1/2} \to 0$：两半分别从上方和下方到达原点，并且都与 x 轴相切，形成一个尖角，称为尖点（cusp）。



(i) (v) (vii)

这些图所用的 b 和 c 的值与相应的三次曲线相同，两轴的比例相等。

验算

- 把 $b = c = 0$ 代入任何一个一般性的结论，结果都应该与 (vii) 吻合；例如 (ii) 中的驻点值 $\frac{8}{27}b^3$ 和 (vi) 中的极小值 $-\frac{32}{27}b^3$ 都收缩到原点。
- 取 $c = 1$ 和 $b = 0.99$（情况 (iii)），转向点大约在 $(0.340, 0.150)$ 和 $(0.980, 0.020)$，正如所说，两个都在横轴上方。取 $(b, c) = (1, 0.8)$，根是 0、0.4、1.6，极小值约为 -0.388，在 $x \approx 1.147$ 处取得，位于 $(0.4, 1.6)$ 之内。
- 就情况 (i)、(iii) 和 (v) 随机选取数千组 (b, c) 作数值测试，结果确认了转向点和根的数目与位置，以及 $f \geq 0$ 成立的集合。

常见错误

- 把情况 (iii) 的极小点画在横轴下方：当 $b < c$ 时，三次曲线对所有 $x > 0$ 都取正值。
- 画 $y^2 = f(x)$ 时，在它于单根处穿过 x 轴的地方画成一个角，或者画成水平切线。该处的切线是垂直的。
- 忘记了 $y^2 = f(x)$ 在三次曲线取负值的地方没有点，所以在情况 (v) 中，卵形线与分支是不相连的。

STEP 之外

若曲线 $y^2 = x^3 + Ax^2 + Bx + C$ 中的三次式有相异的根，这条曲线就是椭圆曲线（elliptic curve），是现代数论的核心对象之一。情况 (v) 有三个实根，给出一个卵形线加一条分支的两部分图形；情况 (i) 只有一个实根，给出单独一部分。当有根重合时，曲线就变成奇异（singular）的：情况 (vii) 有一个尖点，而情况 (iv) 则会给出一条在 $(b, 0)$ 处自相交的曲线。这类曲线上的点竟然可以用一种奇妙的方式“相加”，想初步了解，可参阅 Silverman 与 Tate 的 *Rational Points on Elliptic Curves*（椭圆曲线上的有理点）。

详细解答

STEP II 1990，第 4 题

纯数学 · 直线排布与数学归纳法

> **STEP II 1990，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/90-s2-q4

本题在 *Beyond STEP* 第 9.1 节直线分割平面中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/lazy-caterer.html

关键思路

逐条直线地看这幅图。每加入一条新直线，它与每条旧直线都恰好相交一次，而且交点各不相同，所以它被切成的段数是可以预知的。每一段都有两个作用：它本身是一条新的线段，同时又把一个原有的区域一分为二。数清楚加入一条新直线所带来的变化，便得到递推关系（recurrence），而这些递推关系很容易求和。

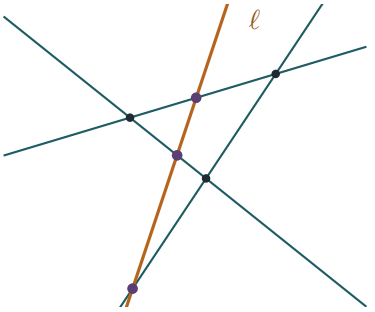
解答

小的情况

在下文中，“线段”（segment）是指某条直线上相邻两个交点之间的一段，包括每条直线两端的两段无限长部分（射线，ray）；只有一条直线时，整条直线算作一条线段。画出最初几种情况，可得：

n	1	2	3	4
线段	1	4	9	16
区域	2	4	7	11

例如，三条直线围成一个三角形：每条直线被切成一条有限的边和两条射线，共得 9 条线段；而区域则有三角形本身、三个与它的边相接的区域，以及三个与它的顶点相接的区域，合共 7 个。线段的数目令人猜想答案是 n^2。



图中有三条直线（蓝绿色）和第四条直线 ℓ（琥珀色）。ℓ 与旧直线相交于三个紫红色的点，这三点把它切成四段，每一段各把七个旧区域的其中一个分成两个，于是得到 11 个区域。

线段的数目

固定 n 条直线中的一条。由于没有两条直线平行（parallel），它与其余 $n-1$ 条直线各相交一次；又由于没有三条直线共点（concurrent），这 $n-1$ 个交点各不相同（假如有两个交点重合，便会有三条直线在该处相交）。所以这条直线上有 $n-1$ 个不同的点，把它切成 n 段：$n-2$ 条有限线段和 2 条射线。对 n 条直线逐一这样做，可得

$$\text{线段的数目} = n \cdot n = n^2.$$

每条线段只被数了一次，因为每条线段都恰好位于一条直线上。

如果你喜欢用归纳法（induction），可设 $S(n)$ 为线段的数目。在 $n-1$ 条处于一般位置（general position）的直线上加入第 n 条直线，新直线被切成 n 段，而它所穿过的 $n-1$ 条旧直线，每条都有一条线段被分成两条，即每条各多出一条线段。所以

$$S(n) = S(n-1) + n + (n-1) = S(n-1) + 2n - 1,$$

而 $S(1) = 1$，若 $S(n-1) = (n-1)^2$，则 $S(n) = (n-1)^2 + 2n - 1 = n^2$。

区域的数目

设 $R(n)$ 为 n 条处于一般位置的直线所分成的区域数目，所以 $R(1) = 2$。从 $n-1$ 条这样的直线出发，加入第 n 条直线 ℓ。如上所述，ℓ 与旧直线相交于 $n-1$ 个不同的点，并被切成 n 段。

取其中一段。它的内部不与任何旧直线相交，所以它位于单一个旧区域之内，并且由边界到边界（或一直伸延至无穷远）横贯该区域。因此它把该区域分成两个。

想一想 这 n 段之中，会否有两段位于同一个旧区域内，以致新增的区域少于 n 个？

不会。每个区域都是若干个半平面（half-plane）的交集（即位于每条旧直线的某一固定侧），所以它是凸（convex）的。一条直线与一个凸集的交集是单一个区间，所以 ℓ 最多只穿过每个旧区域一次，不同的段位于不同的区域。此外，ℓ 没有经过的区域则完全不受影响。

所以加入 ℓ 恰好把 n 个旧区域一分为二，其余的保持不变：

$$R(n) = R(n-1) + n.$$

现在使用归纳法。当 $n = 1$ 时，公式 $\frac{1}{2}(n^2+n+2)$ 给出 2，这是正确的。若 $R(n-1) = \frac{1}{2}\big((n-1)^2+(n-1)+2\big) = \frac{1}{2}(n^2-n+2)$，则

$$R(n) = \tfrac{1}{2}(n^2-n+2) + n = \tfrac{1}{2}(n^2+n+2),$$

所以公式对所有 $n \geq 1$ 成立，证明完毕。（等价地，$R(n) = 1 + (1+2+\cdots+n) = 1 + \frac{1}{2}n(n+1)$，其中开首的 1 是没有任何直线时的整个平面。）

验算

- 当 $n = 1, 2, 3, 4$ 时，公式给出 $2, 4, 7, 11$，与所画的图吻合。
- 欧拉公式（Euler's formula）提供了一个独立的验算方法。把所有无限长的射线在单一个“无穷远点”（point at infinity）处连接起来。这样便有 $V = \binom{n}{2} + 1$ 个顶点、$E = n^2$ 条边和 F 个面，且 $V - E + F = 2$，所以 $F = n^2 - \binom{n}{2} + 1 = \frac{1}{2}(n^2+n+2)$。
- 我们用电脑随机画出最多 8 条系数为整数的直线，通过找出交点的位置来数线段，并在每个交点附近及远处选取样本点，以这些样本点“位于每条直线哪一侧”的不同组合数目作为区域的数目。例如，8 条直线每次都得出 64 条线段和 37 个区域。

常见错误

- 忘记每条直线上的两条无限长射线，结果得出 $n(n-2)$ 而不是 n^2。
- 说新直线增加 $n-1$ 个区域（每个交点一个），而不是 n 个（每段一个）。
- 没有用上题设条件：若没有“无三线共点”这个条件，一条直线上的交点便不一定各不相同，数目亦会减少。

STEP 之外

这是超平面排列（hyperplane arrangement）最简单的情况。把同样的论证提高一个维度，可证明 n 个处于一般位置的平面把空间切成 $\frac{1}{6}(n^3+5n+6)$ 块，因为每个新平面本身都被旧平面切成 $R(n-1)$ 个区域。用欧拉公式所作的验算，让我们初次窥见欧拉示性数（Euler characteristic），它是拓扑学（topology）的一个核心概念。若想读一章讲述欧拉公式及其应用的精彩文字，可参阅 Aigner 与 Ziegler 的 *Proofs from THE BOOK*（《数学天书中的证明》）。

详细解答

STEP II 1990，第 5 题

纯数学 · 复数与圆

STEP II 1990，第 5 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-s2-q5

关键思路

在圆 $|z| = r$ 上，取共轭（conjugation）就等于做除法：$\bar{z} = r^2/z$。这样一来，“这个比是纯虚数（purely imaginary）”（判断两个方向互相垂直的准则）就化成一个简洁的多项式条件 $lm + pq = 0$。此后，题目余下的部分只是一路追踪下去：每条垂直的弦（chord）都以形如 $z \mapsto c/z$ 的公式给出新的点，你只须把这些映射复合（compose）起来，看看结果是甚么。

解答

垂直的判别准则

设 S 的半径为 r，则 $|l| = |m| = |p| = |q| = r$，因此

$$\bar{l} = \frac{r^2}{l}, \qquad \bar{m} = \frac{r^2}{m}, \qquad \bar{p} = \frac{r^2}{p}, \qquad \bar{q} = \frac{r^2}{q}.$$

弦 LM 和 PQ 的方向分别是 $l - m$ 和 $p - q$，由于各点互不相同，两者皆不为零。两个非零复数所代表的方向互相垂直，当且仅当它们的商是纯虚数，所以设

$$w = \frac{l - m}{p - q}.$$

那么 w 是纯虚数当且仅当 $w + \bar{w} = 0$。现在

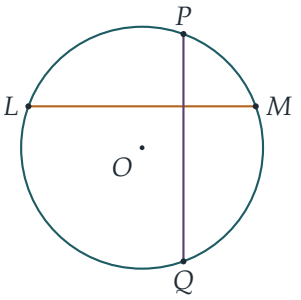
$$\bar{w} = \frac{r^2/l - r^2/m}{r^2/p - r^2/q} = \frac{(m - l)/(lm)}{(q - p)/(pq)} = \frac{l - m}{p - q} \cdot \frac{pq}{lm} = w\,\frac{pq}{lm},$$

所以

$$w + \bar{w} = w\left(1 + \frac{pq}{lm}\right) = \frac{w}{lm}\,(lm + pq).$$

由于 $w \neq 0$ 且 $lm \neq 0$，上式等于零当且仅当 $lm + pq = 0$。所以 $LM \perp PQ$ 当且仅当 $lm + pq = 0$，这正是所要证明的。

（用垂直平分线（perpendicular bisector）的做法，其实是同一个想法换上几何的外衣：弦 LM 的垂直平分线通过 O，而弦的方向是该平分线方向的 i 倍。）



$l = re^{i\cdot 160^\circ}$, $m = re^{i\cdot 20^\circ}$
$p = re^{i\cdot 70^\circ}$, $q = re^{-i\cdot 70^\circ}$
$lm = -r^2$, $pq = r^2$

三个点

以 a_1, a_2, a_3 表示 A_1, A_2, A_3，并以 z 表示 A_1'。根据上述准则，“$A_1'A_2' \perp A_1A_2$”即是 $z\,a_2' = -a_1a_2$，所以

$$a_2' = -\frac{a_1a_2}{z}.$$

这个点确实在 S 上，因为 $|a_2'| = r \cdot r/r = r$；而且它是 S 上唯一符合要求的点。同理

$$a_3' = -\frac{a_2a_3}{a_2'} = \frac{a_3}{a_1}z,$$
$$a_1'' = -\frac{a_3a_1}{a_3'} = -\frac{a_1^2}{z}.$$

因此 $A_1'' = A_1'$ 当且仅当 $z = -a_1^2/z$，即

$$z^2 = -a_1^2 \iff z = \pm i\,a_1.$$

这两个值的模（modulus）都是 r，所以都在 S 上；又因为 $a_1 \neq 0$，它们并不相同。所以 A_1' 恰好有两个位置：就是把 A_1 绕 O 旋转 $\pm 90^\circ$ 所得的两点。

? **想一想** 映射 $z \mapsto -a_1^2/z$ 对这个圆做了甚么？

在 S 上有 $1/z = \bar{z}/r^2$，所以 $-a_1^2/z = -(a_1^2/r^2)\,\bar{z}$，这是先取共轭、再作旋转（rotation）：也就是把圆沿一条直径作反射（reflection）。反射恰好只固定镜线的两个端点，这就是为何有两个解。事实上，每一步 $z \mapsto -a_ka_{k+1}/z$ 都是这样的反射，而这一点解释了下文的一切。

四个点

继续同样的追踪，只是最后一步改用 a_4：

$$a_2' = -\frac{a_1 a_2}{z}, \qquad a_3' = \frac{a_3}{a_1} z,$$
$$a_4' = -\frac{a_3 a_4}{a_3'} = -\frac{a_1 a_4}{z}, \qquad a_1'' = -\frac{a_4 a_1}{a_4'} = z.$$

所以无论 A_1' 取 S 上哪一点，都有 $A_1'' = A_1'$。

n 个点

对 k 用归纳法（induction），这样追踪下去可得

$$a_k' = \frac{a_k}{a_1} z \quad (k \text{ 为奇数}), \qquad a_k' = -\frac{a_1 a_k}{z} \quad (k \text{ 为偶数}),$$

因为若 a_k' 具有其中一种形式，则 $a_{k+1}' = -a_k a_{k+1}/a_k'$ 便具有另一种形式。最后一步是 $a_1'' = -a_n a_1 / a_n'$：

- 若 n 是偶数，$a_1'' = -a_n a_1 \cdot \dfrac{z}{-a_1 a_n} = z$，所以对每一个 A_1'，A_1'' 都与 A_1' 重合；
- 若 n 是奇数，$a_1'' = -a_n a_1 \cdot \dfrac{a_1}{a_n z} = -\dfrac{a_1^2}{z}$，所以与 $n = 3$ 的情况完全一样，A_1'' 与 A_1' 重合的情况恰好只有 A_1' 的两个位置，即 $z = \pm \mathrm{i} a_1$。

用文字来说：偶数次反射合成一个旋转，而在这里它原来就是恒等映射（identity）；奇数次反射则合成一个反射，它在圆上有两个不动点（fixed point）。

验算

- 图中的例子：取 $l = r\mathrm{e}^{\mathrm{i}\cdot 160^\circ}$、$m = r\mathrm{e}^{\mathrm{i}\cdot 20^\circ}$ 时，弦 LM 是水平的；取 $p = r\mathrm{e}^{\pm \mathrm{i}\cdot 70^\circ}$ 时，弦 PQ 是垂直的；而确实有 $lm = r^2 \mathrm{e}^{\mathrm{i}\pi} = -r^2 = -pq$。
- 数值验证：在半径随机的圆上随机取点，并让 n 由 3 取到 9，沿着垂直的弦逐步追踪，每一步所得的点都落回 S 上；当 n 是偶数时，最后准确地回到起点；当 n 是奇数时，沿圆周扫描 200 000 个起始位置，恰好找到两个不动点，位于 $\pm \mathrm{i} a_1$。

常见错误

- 以“w 是实数”来检验垂直。平行的弦给出实数比；垂直的弦给出的是纯虚数比。
- 忘记 $\bar{z} = r^2/z$ 需要 $|z| = r$；正是因为四个点全都在以 O 为圆心的同一个圆上，这个准则才会如此简洁。

► 在 n 个点的情况中，忘记答案取决于 n 的奇偶性（parity）。

STEP 之外

把映射 $z \mapsto c/z$ 限制在圆上，所得的就是反射，而反射的复合正是研究等距变换群（isometry group）的起点：圆的对称构成正交群（orthogonal group）$O(2)$，当中两个反射复合起来就是一个旋转。像 $z \mapsto c/z$ 和 $z \mapsto (az+b)/(cz+d)$ 这样的映射称为莫比乌斯变换（Möbius transformations），是复分析（complex analysis）和几何的核心。Needham 的 *Visual Complex Analysis* 把这两种观点都阐述得十分精彩。

详细解答

STEP II 1990，第 6 题

纯数学 · 整数不等式与分数

> **STEP II 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-s2-q6

关键思路

一切都建基于整数（integer）的一个事实：整数之间的严格不等式（strict inequality）可以收紧 1。若 $a < b$，则 $b \geq a+1$；若 $d < c$，则 $d \leq c-1$。这样一来，(ii) 就变成一行便可完成的估计。其后各部分都是把 (ii) 应用于巧妙选取的整数：先把分数不等式中的分母消去，再把所得的整数不等式代回 (ii)。

解答

以下所有字母均代表正整数（positive integer）。

第 (i) 部分

由于 $b > a \geq 1$ 而 b 是整数，故 $b \geq 2$。两边乘以 $c > 0$ 得 $bc \geq 2c$。

由于 $c > 1$ 而 c 是整数，故 $c \geq 2$，所以 $2c = c + c \geq c + 2$。合起来便得 $bc \geq 2c \geq 2 + c$，正是所求。

第 (ii) 部分

因为两个不等式都是严格的，而各数均为整数，所以 $b \geq a+1$ 及 $d \leq c-1$。在前者用上 $c > 0$，在后者用上 $a > 0$，

$$\begin{aligned} bc - ad &\geq (a+1)c - a(c-1) \\ &= ac + c - ac + a \\ &= a + c. \end{aligned}$$

所以 $bc-ad \geq a+c$，而等号成立当且仅当 $b=a+1$ 及 $d=c-1$。

（这与 (i) 所用的“收紧 1”技巧相同；(i) 只是热身，之后不会再用到。）

第 (iii) 部分

利用 $b,d>0$，把两个不等式乘开：

$$\frac{a}{b}<p \iff a<bp, \qquad p<\frac{c}{d} \iff dp<c.$$

现在应用 (ii)，把 b 换成 bp，把 d 换成 dp（两者都是正整数）。所需的假设 $a<bp$ 和 $dp<c$ 正是我们已有的，所以

$$(bp)c-a(dp) \geq a+c, \qquad \text{即} \qquad (bc-ad)p \geq a+c.$$

第 (iv) 部分

消去分母（全部为正数），得到两个整数不等式：

$$aq<bp \qquad \text{及} \qquad dp<cq.$$

此外 $\frac{a}{b}<\frac{c}{d}$，所以 $ad<bc$，而 $bc-ad$ 是一个正整数；因此我们可以除以它而无须把任何不等号反转。

> **想一想** 应该由哪四个数来担当 (ii) 中 a,b,c,d 的角色?

我们有两个严格不等式，而 (ii) 正好需要两个：“第一个 $<$ 第二个”和“第四个 $<$ 第三个”。配对的自然方法有两种，每种各给出其中一个下界。

p 的下界。取 $aq<bp$ 作为“$a<b$”那个不等式，取 $dp<cq$ 作为“$d<c$”那个。把第 (ii) 部分应用于 $aq,\ bp,\ cq,\ dp$，得

$$(bp)(cq)-(aq)(dp) \geq aq+cq, \qquad \text{所以} \qquad pq(bc-ad) \geq q(a+c).$$

先除以 $q>0$，再除以 $bc-ad>0$，

$$p \geq \frac{a+c}{bc-ad}.$$

q 的下界。现在把角色对调：取 $dp<cq$ 作为“$a<b$”那个不等式，取 $aq<bp$ 作为“$d<c$”那个。把第 (ii) 部分应用于 $dp,\ cq,\ bp,\ aq$，得

$$(cq)(bp)-(dp)(aq) \geq dp+bp, \qquad \text{所以} \qquad pq(bc-ad) \geq p(b+d).$$

除以 $p>0$，再除以 $bc-ad>0$，

$$q \geq \frac{b+d}{bc-ad}.$$

两个下界都已证明。（第 (iii) 部分是第一个下界在 $q=1$ 时的特殊情况。）

介乎 $8/9$ 与 $9/10$ 之间的分数

取 $a=8$、$b=9$、$c=9$、$d=10$。那么

$$bc-ad=81-80=1,$$

所以任何严格介乎 $8/9$ 与 $9/10$ 之间的分数 p/q（其中 p,q 为正整数，不论是否最简分数（lowest terms））都满足

$$p\geq\frac{8+9}{1}=17,\qquad q\geq\frac{9+10}{1}=19.$$

因此，小于 20 的分母必定恰好是 $q=19$。这时我们需要

$$\frac{8}{9}<\frac{p}{19}<\frac{9}{10}\iff\frac{152}{9}<p<\frac{171}{10}\iff 16.88\ldots<p<17.1,$$

所以 $p=17$。而 $17/19$ 的确落在这个区间内，因为 $8\cdot19=152<153=9\cdot17$ 及 $17\cdot10=170<171=9\cdot19$。

因此，介乎 $8/9$ 与 $9/10$ 之间而分母小于 20 的分数只有

$$\frac{17}{19}.$$

由于下界 $q\geq19$ 适用于分数的每一种写法，而不限于最简形式，所以像 $34/38$ 这类未约简的写法，也不可能以更小的分母混进来。

验算

- $17/19$ 是中间分数（mediant）$\frac{8+9}{9+10}$，而 (iv) 中两个下界都由它达到，这是个好兆头，说明这些下界是最佳的。
- 小数：$8/9\approx0.8889$，$17/19\approx0.8947$，$9/10=0.9$。
- 我们对不超过 29 的所有数值检验了 (i) 和 (ii)，以约 60,000 个随机例子检验了 (iii) 和 (iv)，并用穷举法（brute force）列出所有介乎 $8/9$ 与 $9/10$ 之间而 $q\leq19$ 的 p/q：只有 $17/19$ 出现。

常见错误

- 在 (iv) 中除以该数之前，忘记说明 $bc-ad>0$。
- 在最后一部分只用 p 的下界。它容许 q 最大为 19，但单凭它并不能排除更小的分母；真正起作用的是 q 的下界。
- 在 (iv) 中只从 (ii) 得到一个下界：第二个下界需要把两个不等式的角色对调。

STEP 之外

满足 $bc - ad = 1$ 的两个分数 $a/b < c/d$ 称为**法里相邻分数**（Farey neighbours），而本题证明了关于它们的关键事实：介乎两者之间而分母最小的分数，就是它们的中间分数 $\frac{a+c}{b+d}$。不断插入中间分数，便可建构出法里数列（Farey sequences）和 Stern–Brocot 树（Stern–Brocot tree）；它们把每个正有理数（rational）恰好列出一次，并与连分数（continued fractions）以及实数的最佳有理逼近（best rational approximations）密切相关。Davenport 的 *The Higher Arithmetic* 是阅读连分数与逼近理论的绝佳读物。

详细解答

STEP II 1990，第 7 题

纯数学 · 微分方程

> **STEP II 1990，第 7 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-s2-q7

关键思路

以指数函数作试解（trial solution），可把时滞方程（delay equation）化为关于 p 的普通方程，因为时间平移只是把 e^{pt} 乘以一个常数。方程 $(*)$ 同时含有 p 和 e^{-p}，所以无法精确求解；但全式乘以 e^{p} 后，所有含 p 的部分都归到同一边：方程的根就是单一条曲线 $y=(p+1)\mathrm{e}^{p}$ 与水平线 $y=k$ 的交点。于是，数算根的个数就变成描绘一条曲线的问题。

解答

关于 p 的方程

设 $\mathrm{f}(t)=A\mathrm{e}^{pt}$，其中 $A\neq 0$（我们要的是非零解）。则 $\mathrm{f}'(t)=pA\mathrm{e}^{pt}$，而 $\mathrm{f}(t-1)=A\mathrm{e}^{p(t-1)}=\mathrm{e}^{-p}A\mathrm{e}^{pt}$，所以 (†) 变成

$$A\mathrm{e}^{pt}\left(p+1-k\mathrm{e}^{-p}\right)=0 \quad \text{对所有}t.$$

由于 $A\neq 0$ 且 $\mathrm{e}^{pt}>0$，上式对所有 t 成立，当且仅当

$$p+1=k\mathrm{e}^{-p},$$

这正是 $(*)$。因此，这种形式的解存在，恰好是当 p 为 $(*)$ 的实根之时，而此时任何 A 都可以。

有多少个实根?

把 $(*)$ 乘以 $\mathrm{e}^{p}>0$，得到等价的方程

$$g(p)=k, \qquad \text{其中}g(p)=(p+1)\mathrm{e}^{p}.$$

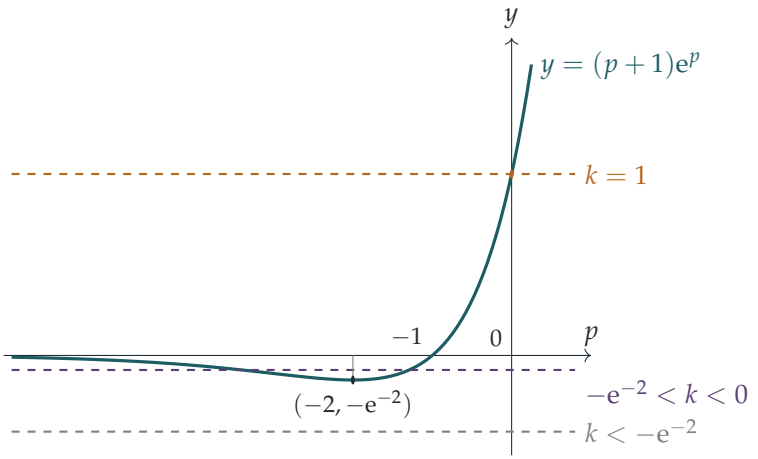
所以我们要数的是曲线 $y = g(p)$ 与水平线 $y = k$ 的交点数目。

> **想一想** $y = (p+1)\mathrm{e}^p$ 的图像是甚么模样？它在哪里取得最小值？两端的走势又如何？

求导得 $g'(p) = \mathrm{e}^p + (p+1)\mathrm{e}^p = (p+2)\mathrm{e}^p$，当 $p < -2$ 时为负，当 $p > -2$ 时为正。所以 g 在 $(-\infty, -2]$ 上递减，在 $[-2, \infty)$ 上递增，并有全域最小值（global minimum）

$$g(-2) = -\mathrm{e}^{-2}.$$

在两端：当 $p \to \infty$ 时 $g(p) \to \infty$；当 $p \to -\infty$ 时 $g(p) \to 0$（指数衰减压倒线性增长）；当 $p < -1$ 时因子 $p+1$ 为负，所以 g 从下方趋近 0。曲线只在 $p = -1$ 处穿过横轴，且 $g(0) = 1$。



从图中读出结论（上面的单调性（monotonicity）令论证严格）：

- $k > 0$：在 $(-\infty, -1]$ 上有 $g \le 0 < k$，而在 $(-1, \infty)$ 上 g 由 0 严格递增至 ∞，所以**恰有一个**根，且它满足 $p > -1$。
- $-\mathrm{e}^{-2} < k < 0$：在 $(-\infty, -2]$ 上，g 由接近 0 的值严格递减至 $-\mathrm{e}^{-2}$，所以它取值 k 一次；在 $[-2, -1]$ 上它由 $-\mathrm{e}^{-2}$ 严格递增至 0，所以它再取值 k 一次；当 $p > -1$ 时，$g > 0 > k$。因此有**两个**根，一个在 $(-\infty, -2)$ 内，一个在 $(-2, -1)$ 内。
- $k = -\mathrm{e}^{-2}$：直线与曲线在最低点相切，得出**唯**一的根 $p = -2$。
- $k < -\mathrm{e}^{-2}$：直线位于整条曲线之下，所以**没有**根。

所以 $(*)$ 的实根数目视乎 k 而为 0、1 或 2，正如题目所求；而形如 $A\mathrm{e}^{pt}$ 的解存在，恰好是当

$$k \ge -\mathrm{e}^{-2}, \quad k \ne 0.$$

（题目已排除 $k = 0$ 这个值；它会给出单一的根 $p = -1$。）

解在甚么时候衰减?

> ? **想一想** 哪些 p 会令 $Ae^{pt} \to 0$? 不解 $(*)$ 的话，怎样判断根相对于 0 的位置?

在 $A \neq 0$ 之下，$Ae^{pt} \to 0$（当 $t \to +\infty$）当且仅当 $p < 0$。我们从来不需要把根明确求出，只需知道它们相对于 0 的位置。

- 若 $-e^{-2} \leq k < 0$，每个根都在 $(-\infty, -1)$ 内（事实上当 $p \geq -1$ 时 $g \geq 0$，所以 $g(p) = k < 0$ 必然导致 $p < -1$）。所有这些解都会衰减。
- 若 $k > 0$，唯一的根位于 g 严格递增的范围内，所以 $p < 0 \iff g(p) < g(0) \iff k < 1$。当 $k = 1$ 时，根是 $p = 0$，给出常数解 $f = A$，它并不趋向零；当 $k > 1$ 时，根为正数，解会增长。

因此，当 $t \to +\infty$ 时解趋向零，恰好是当

$$-e^{-2} \leq k < 1, \quad k \neq 0.$$

验算

- $k = 1$：$p = 0$ 满足 $0 + 1 = 1 \cdot e^{0}$，而 $f = A$ 显然是 $f' + f - f(t-1) = 0$ 的解。这正好是衰减与增长之间的分界，合乎预期。
- $k = -e^{-2} \approx -0.1353$：$p = -2$ 给出 $-1 = -e^{-2}e^{2}$。数值上，$k = -0.13$ 给出两个根 $p \approx -2.311$ 和 $p \approx -1.742$，分别在 -2 的两侧，而 $k = -0.14$ 则没有根。
- 物理意义：没有反馈（feedback）时（$k \to 0$），系统是 $f' = -f$，衰减率为 $p = -1$，这正是 $|k|$ 很小时根所在的位置。强的正反馈（$k > 1$）会盖过阻尼（damping），令解增长。

常见错误

- 把 $y = p + 1$ 与 $y = ke^{-p}$ 画在一起比较，却忽略了当 $k < 0$ 时直线可以与曲线相交两次，也可以完全不相交；改用 $g(p) = (p+1)e^{p}$，相切（tangency）的条件会较容易找到。
- 遗漏端点 $k = -e^{-2}$；它确实给出一个（重）根 $p = -2$，因而有一个衰减的解。
- 把 $k = 1$ 纳入衰减的范围：$p = 0$ 给出的是常数解，而不是衰减的解。

STEP 之外

方程 (†) 是一个时滞微分方程（delay differential equation），而 $(*)$ 是它的特征方程（characteristic equation）。与一般的线性常微分方程（ODE）不同，它有无穷多个复数根 p，而通解是由所有这些根构成的，所以要完全稳定，就需要每个复数根的实部都是负数；这是控制理论（control theory）研究的课题。$(p+1)\mathrm{e}^p = k$ 的实根是朗伯（Lambert）W 函数的值：$p+1 = W(k\mathrm{e})$，而 $-\mathrm{e}^{-1}$（这里即 $k\mathrm{e} = -\mathrm{e}^{-1}$）正好是 W 的分支点（branch point）。关于线性常微分方程和特征方程，可参看 Boyce 与 DiPrima 的 *Elementary Differential Equations*。

详细解答

STEP II 1990，第 8 题

纯数学 · 分部积分法

> **STEP II 1990，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/90-s2-q8

关键思路

微分方程（differential equation）的每一项都是，或者几乎是，一个全导数（exact derivative）。由于 y 和 y' 在 $t=0$ 与 $t=1$ 取相同的值，这些全导数的积分都等于零。每一条关系式的做法，都是把方程乘以一个合适的函数，然后反复使用分部积分法（integration by parts），直至所有边界项（boundary term）都消失。不妨先列出能令边界项消失的事实：$x(0)=0$、$y(0)=y(1)$、$y'(0)=y'(1)$，以及（待我们证明之后）$x(1)=0$。

解答

第一条恒等式

由于 $x'=y$，被积函数（integrand）$xy=xx'$ 正是 $\frac{1}{2}x^2$ 的导数。所以

$$\int_0^1 x(t)\,y(t)\,\mathrm{d}t=\Big[\tfrac{1}{2}x(t)^2\Big]_0^1=\tfrac{1}{2}[x(1)]^2,$$

因为 $x(0)=\int_0^0 y(u)\,\mathrm{d}u=0$。这就是所要求的结果。

证明 $x(1)=0$

把 $(*)$ 的中间项写成全导数：

$$(y^2-1)\,y'=\frac{\mathrm{d}}{\mathrm{d}t}\Big(\tfrac{1}{3}y^3-y\Big).$$

现在把 $(*)$ 由 0 到 1 逐项积分：

$$\Big[y'\Big]_0^1+\Big[\tfrac{1}{3}y^3-y\Big]_0^1+\int_0^1 y\,\mathrm{d}t=0.$$

第一个括号等于 $y'(1)-y'(0)=0$，第二个括号等于零则是因为 $y(1)=y(0)$。根据 x 的定义，余下的积分就是 $x(1)$。因此 $x(1)=0$。

关系式 $\int y^2=\frac{1}{3}\int y^4$

把 $(*)$ 乘以 x，然后在 $[0,1]$ 上积分。这样会得到三个积分，我们逐一处理。

*最后一项。*由第一条恒等式和 $x(1)=0$，可得

$$\int_0^1 x\,y\,\mathrm{d}t=\tfrac{1}{2}[x(1)]^2=0.$$

*y'' 项。*我们使用分部积分法，对 x 求导（于是 $x'=y$）：

$$\begin{aligned}\int_0^1 x\,y''\,\mathrm{d}t&=\Big[x\,y'\Big]_0^1-\int_0^1 y\,y'\,\mathrm{d}t\\&=x(1)y'(1)-x(0)y'(0)-\Big[\tfrac{1}{2}y^2\Big]_0^1=0,\end{aligned}$$

因为 $x(1)=x(0)=0$ 而且 $y(1)=y(0)$。

> **想一想** 我们应该怎样对 $x\,(y^2-1)\,y'$ 进行分部积分？应该对哪一个因子积分？

*中间项。*对因子 $(y^2-1)y'$ 积分（我们已经知道它是 $\frac{1}{3}y^3-y$ 的导数），并对 x 求导：

$$\begin{aligned}\int_0^1 x\,(y^2-1)\,y'\,\mathrm{d}t&=\Big[x\big(\tfrac{1}{3}y^3-y\big)\Big]_0^1-\int_0^1 y\big(\tfrac{1}{3}y^3-y\big)\,\mathrm{d}t\\&=0-\tfrac{1}{3}\int_0^1 y^4\,\mathrm{d}t+\int_0^1 y^2\,\mathrm{d}t.\end{aligned}$$

边界项消失，是因为 x 在两个端点都等于零。

把三部分相加，结果为零：

$$\int_0^1 y^2\,\mathrm{d}t-\tfrac{1}{3}\int_0^1 y^4\,\mathrm{d}t=0,\qquad\text{即}\qquad\int_0^1[y(t)]^2\,\mathrm{d}t=\tfrac{1}{3}\int_0^1[y(t)]^4\,\mathrm{d}t,$$

这正是所要求的结果。

关系式 $\int y'^2=\int y^2$

> **想一想** 我们想得到 y'^2 的积分。乘以哪一个函数才会产生它？

对 $y\,y''$ 进行分部积分会产生 $-\int y'^2$，所以这次我们把 $(*)$ 乘以 y 再积分。逐项计算：

$$\int_0^1 y\,y''\,\mathrm{d}t=\Big[y\,y'\Big]_0^1-\int_0^1 y'^2\,\mathrm{d}t=-\int_0^1 y'^2\,\mathrm{d}t,$$

$$\int_0^1 (y^3-y)\,y'\,\mathrm{d}t=\Big[\tfrac{1}{4}y^4-\tfrac{1}{2}y^2\Big]_0^1=0.$$

边界项 $y(1)y'(1) - y(0)y'(0)$ 等于零，因为 y 和 y' 的值都会重复。第二行等于零，是因为 $y(1) = y(0)$。余下的就是

$$-\int_0^1 y'^2\,\mathrm{d}t + \int_0^1 y^2\,\mathrm{d}t = 0, \qquad \text{所以} \qquad \int_0^1 [y'(t)]^2\,\mathrm{d}t = \int_0^1 [y(t)]^2\,\mathrm{d}t.$$

验算

第一条恒等式对任何 y 都成立。取 $y = \cos 3t + t^2 - 0.7e^t$，以数值积分（quadrature）计算可得 $\int_0^1 xy\,\mathrm{d}t \approx 0.338191$，与 $\frac{1}{2}x(1)^2$ 相等。

$(*)$ 并没有在时间 1 之后重复的非平凡解（non-trivial solution）（见下文）。不过，对于任何在时间 T 之后重复的解，同样的论证可以一字不改地套用到 $[0, T]$ 上。我们以数值方法对 $(*)$ 积分，直至解落在它的周期轨道（periodic orbit）上，其周期为 $T \approx 6.6633$。在一个周期内，我们求得 $\int y \approx 10^{-9}$、$\int y^2 \approx 13.72222$、$\frac{1}{3}\int y^4 \approx 13.72222$ 和 $\int y'^2 \approx 13.72228$，这些数值在步长误差范围内吻合。

常见错误

- 忘记第三部分需要用到 $x(1) = 0$。没有这个条件，$\int xy$ 和各个边界项都不会消失。
- 在中间项中对 x 积分，而不是对 $(y^2 - 1)y'$ 积分。正是利用了 $x' = y$，积分才能化简成 y 的幂。
- 说边界项等于零是“因为 $y(0) = 0$”。题目并没有给出这个条件。边界项消失，是因为在 0 和 1 处的值相等。

STEP 之外

方程 $(*)$ 是范德波尔振荡器（van der Pol oscillator），是自持振荡电路的经典模型。它只有一条周期轨道，而你所证明的关系式，就是沿一个循环的能量平衡。这些关系式还揭示了一个意想不到的事实。由于 $\int_0^1 y = 0$，由傅里叶级数（Fourier series）得出的维廷格（Wirtinger）不等式（又称庞加莱（Poincaré）不等式）给出 $\int_0^1 y'^2 \geq 4\pi^2 \int_0^1 y^2$。结合最后一条关系式，这就迫使 $y \equiv 0$。所以在时间 1 之后重复的解只有平凡解，因为真正的周期约为 6.66。想进一步了解周期轨道和极限环（limit cycle），可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）。

详细解答

STEP II 1990，第 9 题

纯数学 · 极坐标曲线与弧长

STEP II 1990，第 9 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-s2-q9

关键思路

两条曲线都以 θ 表示 r，所以可以逐条射线比较：从原点出发的每条半直线上，抛物线（parabola）恰好有一点，而这一点不是在心脏线（cardioid）之内，就是在其外。这样立刻就知道抛物线在哪里进入和离开心脏线。至于长度，心脏线最好保留极坐标（polar）形式（其弧长（arc length）的被积函数（integrand）用半角恒等式（half-angle identity）便可大幅化简），而抛物线则用直角坐标（Cartesian）形式容易得多。

解答

两条曲线

对抛物线而言，$r + r\cos\theta = 1$ 即 $r = 1 - x$。两边平方，得 $x^2 + y^2 = (1-x)^2$，所以

$$y^2 = 1 - 2x.$$

这是一条焦点（focus）在原点、顶点（vertex）在 $(\frac{1}{2}, 0)$、开口向左的抛物线。原点位于它凹的一侧，即 $y^2 < 1 - 2x$ 的一侧；我们称这一侧为抛物线的内部。

心脏线 $r = 4(1+\cos\theta)$ 的尖点（cusp）在原点（当 $\theta = \pi$ 时），并在 $\theta = 0$ 时到达 $x = 8$。

两曲线的交点

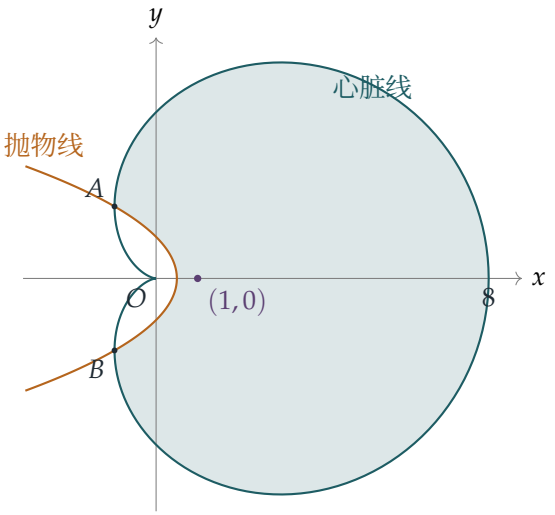
固定一个方向 θ，其中 $-\pi < \theta < \pi$，使得 $1 + \cos\theta > 0$。在这条射线上，抛物线与原点的距离为 $1/(1+\cos\theta)$，心脏线的边界与原点的距离为 $4(1+\cos\theta)$，而心脏线的内部就是射线上位于边界之前的那一段。所以抛物线上的点在心脏线之内，当且仅当

$$\frac{1}{1+\cos\theta} < 4(1+\cos\theta) \iff (1+\cos\theta)^2 > \tfrac{1}{4} \iff \cos\theta > -\tfrac{1}{2} \iff |\theta| < \tfrac{2\pi}{3},$$

其中我们乘以正数 $1+\cos\theta$，并取了正平方根。等号在 $\theta=\pm\frac{2\pi}{3}$ 时成立，此时 $r=2$，得出两个交点

$$A=(-1,\sqrt{3}),\qquad B=(-1,-\sqrt{3}).$$

所以抛物线在心脏线之内的部分是一段连续不断的弧，由 B 经顶点 $(\frac{1}{2},0)$ 到 A，其余部分全在心脏线之外。这段弧连接心脏线边界上的两点，因此把心脏线的内部分成两部分：一部分在抛物线之内（包含原点，以心脏线在尖点附近的小环为边界），另一部分在抛物线之外（以心脏线外侧的大弧为边界）。



（草图中可见一个巧妙的细节：$x=4(1+\cos\theta)\cos\theta=(2\cos\theta+1)^2-1\geq -1$，所以 A 和 B 也是心脏线上最左边的点。）

边界的长度

$r=1,\theta=0$ 所表示的点是 $(1,0)$。它在心脏线之内 $(1<8)$，又在抛物线之外 $(0^2>1-2)$，所以位于阴影部分。这部分的边界由心脏线的弧 $-\frac{2\pi}{3}\leq\theta\leq\frac{2\pi}{3}$ 和抛物线由 B 到 A 的弧组成。

心脏线的弧。在极坐标中，$\mathrm{d}s=\sqrt{r^2+(\mathrm{d}r/\mathrm{d}\theta)^2}\,\mathrm{d}\theta$。这里 $\mathrm{d}r/\mathrm{d}\theta=-4\sin\theta$，所以

$$r^2+\left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2=16\left(1+2\cos\theta+\cos^2\theta+\sin^2\theta\right)=32(1+\cos\theta)=64\cos^2\tfrac{\theta}{2}.$$

想一想 我们可以直接写 $\sqrt{64\cos^2(\theta/2)}=8\cos(\theta/2)$ 吗？

可以，但全因为在整个范围内 $\cos\frac{\theta}{2}\geq 0$：当 $|\theta|\leq\frac{2\pi}{3}$ 时，有 $|\frac{\theta}{2}|\leq\frac{\pi}{3}$。利用图形对 x 轴的对称性，

$$L_{\text{心脏线}}=2\int_0^{2\pi/3}8\cos\tfrac{\theta}{2}\,\mathrm{d}\theta=32\left[\sin\tfrac{\theta}{2}\right]_0^{2\pi/3}=32\cdot\tfrac{\sqrt{3}}{2}=16\sqrt{3}.$$

抛物线的弧。以 y 为参数（parameter）：$x=\frac{1}{2}(1-y^2)$，所以 $\mathrm{d}x/\mathrm{d}y=-y$，而 $\mathrm{d}s=\sqrt{1+y^2}\,\mathrm{d}y$，其中 y 由 $-\sqrt{3}$ 变到 $\sqrt{3}$。由对称性，

$$L_{\text{抛物线}}=2\int_0^{\sqrt{3}}\sqrt{1+y^2}\,\mathrm{d}y.$$

设 $y=\sinh u$，$\mathrm{d}y=\cosh u\,\mathrm{d}u$，被积函数变成 $\cosh^2 u=\frac{1}{2}(1+\cosh 2u)$，所以

$$\int\sqrt{1+y^2}\,\mathrm{d}y=\tfrac{1}{2}u+\tfrac{1}{4}\sinh 2u=\tfrac{1}{2}\left(\operatorname{arsinh}y+y\sqrt{1+y^2}\right)+C.$$

当 $y=\sqrt{3}$ 时，$\sqrt{1+y^2}=2$，而 $\operatorname{arsinh}\sqrt{3}=\ln(\sqrt{3}+2)$，所以

$$L_{\text{抛物线}}=\left[\operatorname{arsinh}y+y\sqrt{1+y^2}\right]_0^{\sqrt{3}}=\ln(2+\sqrt{3})+2\sqrt{3}.$$

总长度。两者相加，

$$L=16\sqrt{3}+2\sqrt{3}+\ln(2+\sqrt{3})=18\sqrt{3}+\ln(2+\sqrt{3}),$$

这正是所要求的结果。

验算

- 整条心脏线的长度是 $\int_{-\pi}^{\pi}8\cos\frac{\theta}{2}\,\mathrm{d}\theta=32$，而我们所用的外侧弧长 $16\sqrt{3}\approx 27.71$ 占了其中大部分，与草图吻合：尖点附近的小环长度为 $32-16\sqrt{3}\approx 4.29$。
- 抛物线的弧长约为 4.78，比由 B 到顶点 $(\frac{1}{2},0)$ 再到 A 的两条直线段的总长度 $2\sqrt{1.5^2+3}\approx 4.58$ 稍长，这是理所当然的。
- 数值上，对两条曲线的极坐标弧长被积函数用中点法则（midpoint rule）作数值积分（quadrature），所得总和为 32.49387，与 $18\sqrt{3}+\ln(2+\sqrt{3})\approx 32.49387$ 一致。

常见错误

- 选错了部分：$(1,0)$ 在抛物线之外（在其顶点 $x=\frac{1}{2}$ 的右方），所以它所在部分的边界用的是心脏线外侧的大弧，而不是尖点附近的小环。
- 用 $8\cos\frac{\theta}{2}$ 一口气把心脏线由 0 积分到 2π，但这个式子在 $\theta>\pi$ 时变成负数；被积函数应是 $8|\cos\frac{\theta}{2}|$。

STEP 之外

两条曲线都是圆锥曲线（conic）或蚶线（limaçon），写成以焦点或尖点为极点（pole）的形式；而 $r=\ell/(1+e\cos\theta)$ 正是圆锥曲线在轨道力学（orbital mechanics）中出现的

形式，太阳位于焦点。弧长公式 $\int \sqrt{r^2 + r'^2}\,\mathrm{d}\theta$ 是参数曲线（parametrised curve）长度 $\int |\mathbf{r}'(t)|\,\mathrm{d}t$ 的特殊情况，后者在大学一年级的向量微积分（vector calculus）中会进一步引出曲率（curvature）和线积分（line integral）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）用圆锥曲线的极坐标形式推导出开普勒（Kepler）轨道。

详细解答

STEP II 1990，第 10 题

纯数学 · 矩阵多项式方程

> **STEP II 1990，第 10 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-s2-q10

关键思路

行列式（determinant）把矩阵（matrix）的乘积化为普通数字的乘积，而实数的乘积为零，必定是其中一个因子为零。至于那个三次式，$\mathbf{M}$ 与 $\mathbf{I}$ 可交换（commute），所以 $\mathbf{M}$ 的矩阵多项式的因式分解，与 x 的多项式完全一样。上三角矩阵（upper triangular matrix）的表现则十分理想：乘积的对角线元素就是各对角线元素的乘积，所以对角线很快便可处理好，只有右上角才需要真正花工夫。

解答

满足 $\mathbf{AB}=\mathbf{0}$ 的矩阵

由于乘积的行列式等于行列式的乘积，

$$\det\mathbf{A}\,\det\mathbf{B}=\det(\mathbf{AB})=\det\mathbf{0}=0.$$

左方是两个实数的乘积，所以其中至少一个为零：$\det\mathbf{A}=0$，或 $\det\mathbf{B}=0$，或两者皆为零。这就是所需的结果。

若 $\det\mathbf{B}\neq 0$，则 $\mathbf{B}^{-1}$ 存在，把 $\mathbf{AB}=\mathbf{0}$ 右乘 $\mathbf{B}^{-1}$，得

$$\mathbf{A}=\mathbf{ABB}^{-1}=\mathbf{0}\,\mathbf{B}^{-1}=\mathbf{0}.$$

所以 $\mathbf{A}$ 必定是零矩阵（zero matrix）。

不过，两个行列式皆为零并不足够。取

$$\mathbf{A}=\mathbf{B}=\begin{pmatrix}1&0\\0&0\end{pmatrix}.$$

则 $\det \mathbf{A} = \det \mathbf{B} = 0$，但 $\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \mathbf{0}$。

三次式的因式分解

首先把多项式 $f(x) = x^3 + 2x^2 - 5x - 6$ 因式分解。试代 6 的细小因数：

$$\begin{aligned} f(-1) &= -1 + 2 + 5 - 6 = 0, \\ f(2) &= 8 + 8 - 10 - 6 = 0, \\ f(-3) &= -27 + 18 + 15 - 6 = 0, \end{aligned}$$

由于首项系数为 1，所以 $f(x) = (x+1)(x-2)(x+3)$。

现在展开 $(\mathbf{M} + p\mathbf{I})(\mathbf{M} + q\mathbf{I})(\mathbf{M} + r\mathbf{I})$。因为 $\mathbf{I}$ 与 $\mathbf{M}$ 可交换（而且 $\mathbf{I}^2 = \mathbf{I}$），展开的结果与数字的情况完全相同：

$$\mathbf{M}^3 + (p+q+r)\mathbf{M}^2 + (pq+qr+rp)\mathbf{M} + pqr\,\mathbf{I}.$$

与上面的多项式比较，我们可取

$$p = 1, \qquad q = -2, \qquad r = 3$$

(次序不拘), 于是对任何方阵 $\mathbf{M}$, 都有 $\mathbf{M}^3 + 2\mathbf{M}^2 - 5\mathbf{M} - 6\mathbf{I} = (\mathbf{M}+\mathbf{I})(\mathbf{M}-2\mathbf{I})(\mathbf{M}+3\mathbf{I})$。

> ? **想一想** 我们现在可否断定三个因子之中必有一个是零矩阵?

不可以。第一部分已经说明，矩阵的乘积可以为零，而没有任何一个因子为零；我们只知道至少有一个因子是奇异（singular）的。所以我们必须动手计算。

上三角矩阵的解

设 $\mathbf{M} = \begin{pmatrix} a & c \\ 0 & b \end{pmatrix}$。每个因子都是上三角矩阵，而且右上角的元素同样是 c：

$$\mathbf{M} + k\mathbf{I} = \begin{pmatrix} a+k & c \\ 0 & b+k \end{pmatrix}.$$

对于三个对角线为 (d_i, e_i)、右上角元素全为 c 的上三角矩阵，乘开后得

$$\begin{aligned} &\begin{pmatrix} d_1 & c \\ 0 & e_1 \end{pmatrix} \begin{pmatrix} d_2 & c \\ 0 & e_2 \end{pmatrix} \begin{pmatrix} d_3 & c \\ 0 & e_3 \end{pmatrix} \\ &= \begin{pmatrix} d_1 d_2 d_3 & c\,(d_1 d_2 + d_1 e_3 + e_2 e_3) \\ 0 & e_1 e_2 e_3 \end{pmatrix}. \end{aligned}$$

取 $d_1, d_2, d_3 = a+1, a-2, a+3$ 及 $e_1, e_2, e_3 = b+1, b-2, b+3$，方程便化为三个条件：

$$\begin{aligned}(a+1)(a-2)(a+3) &= 0,\\ (b+1)(b-2)(b+3) &= 0,\\ c\left[(a+1)(a-2)+(a+1)(b+3)+(b-2)(b+3)\right] &= 0.\end{aligned}$$

头两个条件表示 $a, b \in \{-1, 2, -3\}$。把第三个条件中的括号展开，

$$(a^2-a-2)+(ab+3a+b+3)+(b^2+b-6) = a^2+ab+b^2+2(a+b)-5.$$

> **想一想** 要看出这个括号何时为零，有没有比逐一试验全部九对数值更聪明的方法？

有。若 $a \neq b$，括号等于

$$\frac{a^3-b^3}{a-b}+2\frac{a^2-b^2}{a-b}-5\frac{a-b}{a-b} = \frac{f(a)-f(b)}{a-b},$$

由于 $f(a) = f(b) = 0$，此式为 0。所以当 $a \neq b$ 时，无论 c 是甚么，第三个条件都成立。若 $a = b$，括号等于 $3a^2+4a-5 = f'(a)$。在 $a = -1, 2, -3$ 时，其值分别为 -6、15 和 10（全部不为零，因为 f 有三个相异的根），所以第三个条件迫使 $c = 0$。

因此，解为：

- $\mathbf{M} = \begin{pmatrix} a & c \\ 0 & b \end{pmatrix}$，其中 a 和 b 是 $\{-1, 2, -3\}$ 中两个不同的数，而 c 是任意实数（共六族）；
- 三个纯量矩阵（scalar matrix）$-\mathbf{I}$、$2\mathbf{I}$ 和 $-3\mathbf{I}$（即 $a = b$ 的情况，此时必有 $c = 0$）。

验算

取 $a = -1$、$b = 2$、$c = 5$。则 $\mathbf{M}+\mathbf{I} = \begin{pmatrix} 0 & 5 \\ 0 & 3 \end{pmatrix}$、$\mathbf{M}-2\mathbf{I} = \begin{pmatrix} -3 & 5 \\ 0 & 0 \end{pmatrix}$ 及 $\mathbf{M}+3\mathbf{I} = \begin{pmatrix} 2 & 5 \\ 0 & 5 \end{pmatrix}$。头两个的乘积已经是 $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$，所以无论 c 取何值，整个乘积都是零。另一方面，$\mathbf{M} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ 给出 $\mathbf{M}^3+2\mathbf{M}^2-5\mathbf{M}-6\mathbf{I} = \begin{pmatrix} 0 & 15 \\ 0 & 0 \end{pmatrix} \neq \mathbf{0}$，与 $f'(2) = 15$ 吻合。以电脑搜索大量有理数 a, b, c，所得的解恰好就是上述各解；因式分解亦已用随机的 3×3 整数矩阵验证。

常见错误

- 由 $(\mathbf{M}+\mathbf{I})(\mathbf{M}-2\mathbf{I})(\mathbf{M}+3\mathbf{I})=\mathbf{0}$ 断定 $\mathbf{M}$ 必定是 $-\mathbf{I}$、$2\mathbf{I}$ 或 $-3\mathbf{I}$。第一部分正是为了提醒你不要犯这个错。
- 遗漏 $a=b$ 的情况，或以为在这情况下 c 同样可以任意取值。
- 为第一部分举例时，不慎写出一个 $\mathbf{AB}=\mathbf{0}$ 的例子；要检查你的乘积确实不是零。

STEP 之外

这个答案让我们初次窥见最小多项式（minimal polynomial）：一个矩阵满足某个具相异根的多项式，当且仅当它可对角化（diagonalisable），而且其特征值（eigenvalue）都在这些根之中。我们那些 $a\neq b$ 的解，对任何 c 都可对角化；而 $\begin{pmatrix} a & c \\ 0 & a \end{pmatrix}$ 在 $c\neq 0$ 时则不可对角化，这正是它不成立的原因。右上角出现的均差（divided difference）$\big(f(a)-f(b)\big)/(a-b)$，在 $a=b$ 时变成 $f'(a)$，一般的矩阵函数正是这样定义的。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对特征值和对角化有清晰的讲解。

详细解答

STEP II 1990，第 11 题

力学 · 角动量守恒

STEP II 1990，第 11 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-s2-q11

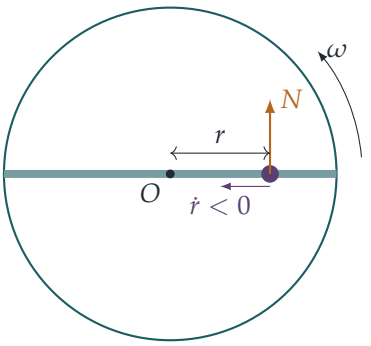
关键思路

圆盘与质点合起来是一个系统：对垂直轴没有外力矩（torque），亦没有外力作功，所以有两个量守恒：对该轴的角动量（angular momentum）和动能（kinetic energy）。由这两条方程可直接求得 ω 和 $\mathrm{d}r/\mathrm{d}t$。把两个变率相除便可消去时间，得出一条描述路径的可分离变量方程（separable equation），即 r 作为 θ 的函数。

解答

模型

把质点视为置于光滑凹槽内的点质量，并假设转轴光滑（轴承无摩擦）。从上方俯视，质点位于凹槽上距离 r 处，沿凹槽方向的速度为 $\dot{r}$，垂直于凹槽方向的速度为 $r\omega$，其中 $\omega = \dot{\theta}$。



质点所受的水平力只有凹槽的反作用力 N，方向垂直于凹槽（凹槽光滑，故沿凹槽方向没有分量）。圆盘则受到 $-N$。

角动量

重力和垂直方向的反作用力均与转轴平行，而光滑的转轴不施加力矩，所以对转轴的总角动量是常数。圆盘的贡献是 $mk^2\omega$，质点的贡献是 $m\,r\cdot r\omega = mr^2\omega$。与 $t=0$ 时比较：

$$mk^2\omega + mr^2\omega = mk^2\Omega + ma^2\Omega \quad\Longrightarrow\quad \omega = \frac{\Omega(k^2+a^2)}{k^2+r^2}.$$

能量

> **想一想** 质点横向移动时，反作用力 N 正好横向推它。这样不会改变动能吗？

它会改变质点的动能，但不会改变总动能。N 对质点作功的功率为 $N\cdot r\omega$，而作用于圆盘上大小相等、方向相反的力，作用点同样在距离 r 处，产生力矩 $-Nr$，故作功的功率为 $-Nr\omega$。两者互相抵消，所以总动能守恒：

$$\tfrac12 mk^2\omega^2 + \tfrac12 m\left(\dot r^2 + r^2\omega^2\right) = \tfrac12 mk^2\Omega^2 + \tfrac12 m\left(V^2 + a^2\Omega^2\right).$$

因此

$$\dot r^2 = V^2 + (k^2+a^2)\Omega^2 - (k^2+r^2)\omega^2.$$

利用已知条件 $V^2 = \Omega^2 a^2(k^2+a^2)/k^2$，

$$V^2 + (k^2+a^2)\Omega^2 = \Omega^2(k^2+a^2)\,\frac{a^2+k^2}{k^2} = \frac{\Omega^2(k^2+a^2)^2}{k^2},$$

而由角动量的结果可得 $(k^2+r^2)\omega^2 = \Omega^2(k^2+a^2)^2/(k^2+r^2)$。于是

$$\dot r^2 = \Omega^2(k^2+a^2)^2\left(\frac{1}{k^2} - \frac{1}{k^2+r^2}\right) = \frac{\Omega^2(k^2+a^2)^2\,r^2}{k^2(k^2+r^2)}.$$

质点向转轴移动时，$\dot r<0$，所以取负平方根（其中 $r>0$）：

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -\frac{\Omega r(k^2+a^2)}{k(k^2+r^2)^{1/2}},$$

这正是所求。

路径

由于 $\omega = \mathrm{d}\theta/\mathrm{d}t > 0$，由链式法则（chain rule）得

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{\mathrm{d}r/\mathrm{d}t}{\mathrm{d}\theta/\mathrm{d}t} = -\frac{\Omega r(k^2+a^2)}{k(k^2+r^2)^{1/2}}\cdot\frac{k^2+r^2}{\Omega(k^2+a^2)} = -\frac{r(k^2+r^2)^{1/2}}{k},$$

即 $k\,\mathrm{d}r/\mathrm{d}\theta = -r(k^2+r^2)^{1/2}$。留意 Ω 和 a 已经消失。

现在设 $u = 1/r$，则 $\mathrm{d}r/\mathrm{d}\theta = -u^{-2}\,\mathrm{d}u/\mathrm{d}\theta$，而 $r(k^2+r^2)^{1/2} = u^{-2}(1+k^2u^2)^{1/2}$（因为 $u > 0$）。方程变成

$$k\frac{\mathrm{d}u}{\mathrm{d}\theta} = \sqrt{1+k^2u^2}.$$

这个平方根提示我们使用双曲代换（hyperbolic substitution）：设 $ku = \sinh\phi$。那么 $k\,\mathrm{d}u/\mathrm{d}\theta = \cosh\phi\,\mathrm{d}\phi/\mathrm{d}\theta$，且 $\sqrt{1+\sinh^2\phi} = \cosh\phi$，所以

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = 1 \quad\Longrightarrow\quad \phi = \theta + \alpha.$$

当 $\theta = 0$ 时 $r = a$，所以 $\sinh\alpha = ku = k/a$。因此 $k/r = \sinh(\theta+\alpha)$，即

$$r\sinh(\theta+\alpha) = k, \qquad \sinh\alpha = \frac{k}{a}.$$

质点永远不会到达转轴

> **想一想** 当 $\theta \to \infty$ 时，$r = k/\sinh(\theta+\alpha)$ 趋向 0。圆盘有可能在有限时间内转过无限大的角度吗?

不可能。由于 $r^2 \geq 0$，$\omega = \Omega(k^2+a^2)/(k^2+r^2) \leq \Omega(k^2+a^2)/k^2$，所以到时间 t 为止，圆盘转过的角度最多为 $\theta \leq \Omega(k^2+a^2)t/k^2$，这是有限的。对任何有限的 θ，$\sinh(\theta+\alpha)$ 都是有限的，所以 $r = k/\sinh(\theta+\alpha) > 0$。此外，对每个 $r > 0$，$\dot{r}$ 都严格为负，所以质点永远不会停下来再折返：这些公式对所有 t 都成立，而 r 不断减小并趋向 0，却永远不会到达。

（另一做法是直接从 $\mathrm{d}r/\mathrm{d}t$ 的方程入手：由 a 走到 r 所需的时间为 $\frac{k}{\Omega(k^2+a^2)}\int_r^a \frac{\sqrt{k^2+x^2}}{x}\,\mathrm{d}x$，而在 $x = 0$ 附近，被积函数的表现有如 k/x，其积分在 $r \to 0$ 时以对数方式发散（diverge）。）

验算

- 在 $t = 0$ 时，$r = a$ 给出 $\omega = \Omega$ 和 $\dot{r} = -\Omega a\sqrt{k^2+a^2}/k = -V$，与初始数据吻合；而 $\theta = 0$ 给出 $r = k/\sinh\alpha = a$。
- 当 θ 很大时，$r \approx 2k\mathrm{e}^{-(\theta+\alpha)}$：质点以指数方式螺旋向内，与永远不会到达的结论一致。
- 数值上，对随机选取的 k, a, Ω 逐步计算运动方程（$\ddot{r} = r\omega^2$，角动量固定），所得的 $\dot{r}$、能量和 $r\sinh(\theta+\alpha) = k$ 均吻合至约 10^{-12}。取 $k = a = \Omega = 1$ 时，到达 $r = 10^{-8}$ 所需的时间约为 9.32，增长方式有如 $\frac{1}{2}\ln(1/r)$。

常见错误

- 以为质点本身的能量守恒：凹槽的反作用力对质点作功，守恒的只是总能量。

- 在质点的动能或角动量中漏掉横向速度 $r\omega$。
- 只凭 θ 有限时 $r > 0$ 便断定质点永远不会到达转轴，却没有检查 θ 在有限时间内是否保持有限。

STEP 之外

关于 V 的特殊条件，令质点恰好落在永远向 $r = 0$ 缓缓靠近的轨迹上：在旋转参考系（rotating frame）中，$r = 0$ 是一个不稳定平衡点（unstable equilibrium），而这条轨迹就是它的*稳定流形*（*stable manifold*）。V 只要稍大一点，质点便会穿过中心；稍小一点，质点便会折返。这类想法，连同耦合物体的守恒定律，可见于 Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*。

详细解答

STEP II 1990，第 12 题

力学 · 恢复系数与抛体

STEP II 1990，第 12 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：

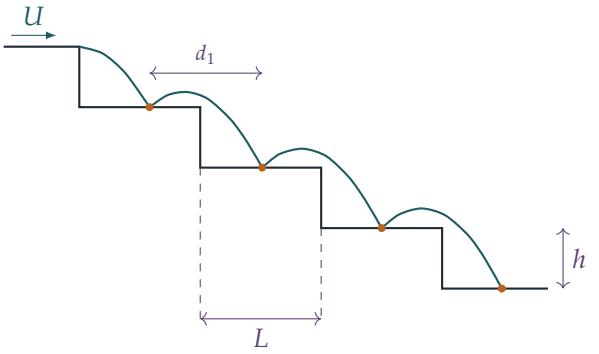
step.maths.org/questions/90-s2-q12

关键思路

楼梯是光滑（smooth）的，所以水平速度永远不变：它始终是 U，而每一段水平距离都只是 U 乘以一段飞行时间。真正要花工夫的是垂直方向的运动：每次反弹令垂直速率乘以 e，而每下降 h 则令其平方增加 $2gh$。由此可得每次反弹时垂直速率的平方所满足的一个简单递推关系（recurrence）。至于最后一部分，想想反弹点必须落在哪里：在一道无穷无尽的楼梯上，每一级恰好落一次。

解答

我们把质点（particle）看成一个点，在两次反弹之间只受重力作用；又把楼梯看成光滑的水平面，每一级比上一级低 h，相邻梯级边缘的水平距离为 L。



相邻两次反弹之间的距离

水平运动。楼梯是光滑的，所以梯级每次施加的冲量（impulse）都纯粹沿垂直方向，而在两次反弹之间唯一的力就是重力。因此水平速度任何时候都是 U，一段历时 T 的飞行所走的水平距离为 UT。

垂直运动。设 w_n 为质点第 n 次反弹时撞上梯级的（向下）速率。质点离开最高一级时没有垂直速度，然后下跌 h，所以

$$w_1^2 = 2gh.$$

在第 n 次反弹时，由于梯级固定不动，根据牛顿恢复定律（Newton's law of restitution），质点以速率 ew_n 向上弹起。之后它必须落在低 h 的下一级上，所以由 $v^2 = u^2 + 2as$（取向下为正，位移为 $+h$）得：

$$w_{n+1}^2 = e^2 w_n^2 + 2gh.$$

由 $w_1^2 = 2gh$ 出发，反复应用这关系，

$$\begin{aligned} w_n^2 &= 2gh\left(1 + e^2 + e^4 + \cdots + e^{2(n-1)}\right) \\ &= 2gh\,\frac{1-e^{2n}}{1-e^2} = 2gh\,\alpha_n^2, \end{aligned}$$

这里用了等比级数（geometric series）求和（因 $e^2 \neq 1$ 而成立）。所以 $w_n = \sqrt{2gh}\,\alpha_n$ 对每个 $n \geq 1$ 都成立；用归纳法（induction）简单验证一下便可使论证严谨，因为 $e^2 \cdot 2gh\,\alpha_n^2 + 2gh = 2gh\big(e^2\alpha_n^2 + 1\big)$，而 $e^2\alpha_n^2 + 1 = \dfrac{e^2 - e^{2n+2} + 1 - e^2}{1-e^2} = \alpha_{n+1}^2$。

> **想一想** 由第 n 次反弹到第 $(n+1)$ 次反弹，飞行时间有多长?

垂直速度以恒定的变率 g（向下）改变。它开始时是向上的 ew_n，结束时是向下的 w_{n+1}，所以总共改变了 $ew_n + w_{n+1}$，飞行时间为

$$T_n = \frac{ew_n + w_{n+1}}{g} = \frac{\sqrt{2gh}}{g}\left(e\alpha_n + \alpha_{n+1}\right) = \left(\frac{2h}{g}\right)^{\frac{1}{2}}\left(e\alpha_n + \alpha_{n+1}\right).$$

（这样便不必解一条关于 T_n 的二次方程：我们早已知道末速率。）乘以水平速率 U，

$$d_n = U\left(\frac{2h}{g}\right)^{\frac{1}{2}}\left(e\alpha_n + \alpha_{n+1}\right),$$

正是所求。

一道很长的楼梯

由于 $0 \leq e < 1$，我们有 $e^{2n} \to 0$，所以当 $n \to \infty$ 时 $\alpha_n \to \dfrac{1}{\sqrt{1-e^2}}$，因此

$$d_n \to D = U\left(\frac{2h}{g}\right)^{\frac{1}{2}}\frac{1+e}{\sqrt{1-e^2}} = U\left(\frac{2h}{g}\right)^{\frac{1}{2}}\left(\frac{1+e}{1-e}\right)^{\frac{1}{2}},$$

其中用了 $1-e^2=(1-e)(1+e)$。

> **想一想** 为甚么这个极限距离必须恰好等于 L?

由最高一级的边缘起沿水平方向量度 x。下面第一级占据 $0\le x\le L$，一般而言，最高一级以下的第 n 级占据 $(n-1)L\le x\le nL$。质点在每一级上都反弹一次，所以它第 n 次反弹的位置（设为 x_n）满足

$$(n-1)L\le x_n\le nL,\qquad \text{所以}\qquad \frac{x_n}{n}\to L,\ \text{当}n\to\infty.$$

另一方面，$x_n=x_1+d_1+d_2+\cdots+d_{n-1}$，而 $d_k\to D$。若 $D>L$，则由某一项起，每个 d_k 都比 L 大至少一个固定的量，于是质点每反弹一次，便相对梯级边缘多走一段固定距离，最终会越过某一级而不落在其上；若 $D<L$，质点则逐步落后，最终会在同一级上落两次。（严格来说，各项趋向 D 时，其平均值 x_n/n 亦趋向 D。）无论是哪种情况，只要楼梯够长，质点都不会每级恰好反弹一次。所以对于很大的 N，我们需要 $D=L$：

$$U\left(\frac{2h}{g}\right)^{\frac12}\left(\frac{1+e}{1-e}\right)^{\frac12}=L\quad\Longrightarrow\quad U=\left(\frac{L^2g}{2h}\right)^{\frac12}\left(\frac{1-e}{1+e}\right)^{\frac12},$$

正是所求。

验算

- 特殊值。$\alpha_1=1$，所以 $d_1=U\sqrt{2h/g}\,(e+\sqrt{1+e^2})$。当 $e=0$ 时，质点每次反弹后垂直方向完全停下，所以每段飞行都只是下跌 h 的自由下落，需时 $\sqrt{2h/g}$：果然 $d_n=U\sqrt{2h/g}$，而最后的条件变成 $U\sqrt{2h/g}=L$，完全符合预期。
- 量纲。$\sqrt{2h/g}$ 是时间，所以 d_n 是长度；$\sqrt{L^2g/h}$ 是速率。
- 数值检验。对随机选取的 U、h 和 e（取 $g=9.8$）以数值方法逐步模拟这运动，所得的每个 d_n 与公式相差都在 10^{-6} 以内。取 $e=\frac12$、$L=0.3$ 和 $h=0.2$，公式给出 $U\approx0.857$，而其后的反弹位置在 2000 级楼梯上都依次落在相继的梯级上。

常见错误

- 把每次反弹都当成由静止下跌 h，或由上一次反弹的高度下跌：质点每次反弹后都会上升，所以它的垂直速率会一级一级地累积起来。
- 写成 $w_{n+1}=ew_n+\sqrt{2gh}$：相加的是速率的平方，而不是速率本身。

STEP 之外

最后一部分其实是一个关于平均值的命题：若 $d_n \to D$，则其累计平均值亦满足 $\frac{1}{n}\sum d_k \to D$（切萨罗定理，Cesàro's theorem），这是实分析（real analysis）入门阶段的一道标准习题。还要留意，关于 U 的条件是*必要*（necessary）条件，而非充分（sufficient）条件：最初几次反弹的距离比 L 短，而若 e 接近 1，累积起来的差额可以超过起步时领先的 x_1，那么任何发射速率都行不通。想看数列极限的严谨处理，可参阅 Abbott 的 *Understanding Analysis*；至于碰撞的力学，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*。

详细解答

STEP II 1990，第 13 题

力学 · 杆的平衡

STEP II 1990，第 13 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-s2-q13

关键思路

由于钉子光滑（smooth）而绳子为轻绳（light），绳上各处的张力（tension）都相同，所以杆受到两个相等的拉力，分别沿 PO 和沿 QO 方向。它们的合力（resultant）平分角 POQ，而要达致平衡（equilibrium），这个合力必须与重量沿同一条铅垂线互相抵销。因此，过 O 的铅垂线就是角 POQ 的角平分线，并且通过 G，余下的工作交给角平分线定理（angle bisector theorem）便可。至于那个不等式，其实只是说三角形 OPQ 是一个真正的三角形。

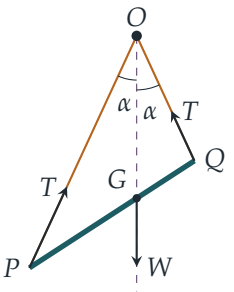
解答

求 OP 和 OQ

把杆模型化为一枝刚性杆，其重心（centre of gravity）为 G，其中 $PG = a + d$ 及 $GQ = a - d$；我们取 $0 \leq d < a$，使得 G 位于杆上，在中点靠 Q 的一侧（或正好在中点）。绳子是不可伸长（inextensible）的轻绳，钉子细小而光滑，所以 OP 和 OQ 两段绳的张力 T 相同，并且

$$OP + OQ = 2b.$$

作用在杆上的力只有三个：作用于 G 的重量 W、在 P 处沿 PO 方向的张力 T，以及在 Q 处沿 QO 方向的张力 T。



两个张力的作用线相交于 O，所以（对 O 取力矩（moments））重量的作用线也必须通过 O：过 O 的铅垂线通过 G。此外，两个张力的大小相等，所以它们的合力指向两个单位向量之和的方向，这两个单位向量一个沿 PO，一个沿 QO，而这正是两者夹角的平分线方向。这个合力必须是铅垂的，才能与 W 平衡。因此铅垂线 OG 平分角 POQ；把每一半记作 α。

想一想 知道 OG 平分角 POQ，能告诉我们关于各长度的甚么资料?

在三角形 OPQ 中应用角平分线定理，由 O 引出的角平分线把 PQ 按邻边之比分割：

$$\frac{OP}{OQ} = \frac{PG}{GQ} = \frac{a+d}{a-d}.$$

（直接的看法：三角形 OPG 和 OQG 有共同的由 O 引出的高，所以它们的面积之比为 $PG : GQ$；而它们的面积又分别是 $\frac{1}{2}\,OP\cdot OG\sin\alpha$ 和 $\frac{1}{2}\,OQ\cdot OG\sin\alpha$。）所以对某个 $k>0$，有 $OP = k(a+d)$ 及 $OQ = k(a-d)$，而由 $OP+OQ = 2b$ 得 $2ka = 2b$，故 $k = b/a$：

$$OP = \frac{b(a+d)}{a}, \qquad OQ = \frac{b(a-d)}{a}.$$

不等式

设 $\theta = \angle POQ = 2\alpha$。由于杆并非铅垂，$O$ 不在直线 PQ 上，所以 OPQ 是一个真正的三角形，且 $0<\theta<\pi$。以 $PQ = 2a$ 代入余弦定理（cosine rule），得

$$\begin{aligned} 4a^2 &= OP^2 + OQ^2 - 2\,OP\cdot OQ\cos\theta \\ &= \frac{b^2}{a^2}\Big[(a+d)^2 + (a-d)^2 - 2(a^2-d^2)\cos\theta\Big] \\ &= \frac{2b^2}{a^2}\Big[(a^2+d^2) - (a^2-d^2)\cos\theta\Big]. \end{aligned}$$

现在，因为 $\theta\neq 0$，所以 $\cos\theta<1$；又因为 $d<a$，所以 $a^2-d^2>0$。于是

$$4a^2 > \frac{2b^2}{a^2}\Big[(a^2+d^2) - (a^2-d^2)\Big] = \frac{4b^2d^2}{a^2},$$

亦即 $b^2d^2 < a^4$。取正平方根（其中 $b > 0$，$d \geq 0$），得 $bd < a^2$，所以

$$d < \frac{a^2}{b},$$

这正是所要证明的结果。

换一个等价的说法，这就是三角不等式（triangle inequality）$OP - OQ < PQ$：这里 $OP - OQ = 2bd/a$，而 $2bd/a < 2a$ 正是同一个条件。（另一个不等式 $\cos\theta > -1$ 给出 $OP + OQ > PQ$，即 $b > a$，这只是说绳子的长度足以绕过钉子而已。）

验算

- 若 $d = 0$（均匀杆），则 $OP = OQ = b$，杆水平悬挂，正如对称性所要求的；不等式 $0 < a^2/b$ 自动成立。
- 当 $d \to a^2/b$ 时，$OP - OQ \to 2a = PQ$，三角形被压扁，杆变成铅垂：超过这个值之后，杆唯一的悬挂方式就是铅垂悬挂。
- 取 $a = 1$，$b = 2$，$d = 0.3$：$OP = 2.6$，$OQ = 1.4$，$\cos\theta \approx 0.648$。把这个三角形摆放成 OG 为铅垂，两个张力的水平分量互相抵销，而对 O 的力矩在机器精度之内平衡。对绳子的所有位置逐一扫描，确认了这是唯一的非铅垂平衡位置，并且一旦 $d \geq a^2/b$ 便不存在这样的位置。

常见错误

- 假设两段绳的张力不同。绳子绕过光滑钉子时，两段的张力相等，而正是这一点令 OG 成为角平分线。
- 把角平分线的比例弄反了：较长的一段绳 OP 连接的是离 G 较远的一端。
- 以为这个平衡是稳定的。事实上它是不稳定的：当 $d = 0$ 时，水平的杆的 G 在 $\sqrt{b^2 - a^2}$ 的距离处位于 O 之下，而铅垂悬挂时则在其下方 b 的距离处，位置更低。题目只问及平衡。

STEP 之外

相对于杆而言，O 的轨迹（locus）是一个以 P 和 Q 为焦点（foci）的椭圆，而平衡位置就是 G 的高度取驻值之处。这就是虚功原理（principle of virtual work），由此可以通往拉格朗日力学（Lagrangian mechanics）。它也解释了为甚么相等的张力会平分夹角：这正是椭圆的反射性质。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）是进一步学习静力学和稳定性的好选择。

详细解答

STEP II 1990，第 14 题

力学 · 弹性碰撞

STEP II 1990，第 14 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-s2-q14

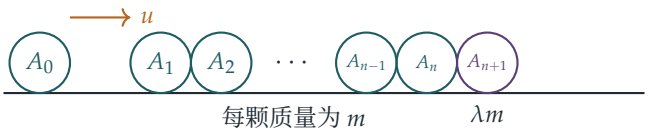
本题在 *Beyond STEP* 第 10.1 节牛顿摆中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch10/newtons-cradle.html

关键思路

我们不知道冲量（impulse）怎样沿着这列弹珠传递，所以无法逐次追踪每一下碰撞。我们能做的，是用三个条件去检验每个提出的最终运动：动量（momentum）守恒、动能（kinetic energy）守恒，以及弹珠永远不会互相穿过。第三个条件很容易被遗忘，但每一部分的结论正是由它决定的。若一颗弹珠的正后方紧贴着一颗静止的弹珠，它就不能向后运动；而一颗弹珠的速度也不能比它正前方的弹珠快。

解答

我们把弹珠看成在一条光滑直线（即槽）上的质点（particle），速度以 A_0 原来的运动方向为正。碰撞之前，总动量是 mu，总动能是 $\frac{1}{2}mu^2$，两者都守恒。



由于弹珠不能重叠，任何可能的最终运动都必须符合以下两点。

- 若 A_k 静止，而 A_{k+1} 在它的正前方并与它接触，则 A_{k+1} 的速度不能是负数。
- 若 A_k 与 A_{k+1} 互相接触，则 A_k 的速度不能大于 A_{k+1} 的速度。

第 (i) 部分

假设只有 A_n 和 A_{n+1} 运动，速度分别为 v 和 w。两式同除以 m，得

$$v + \lambda w = u,$$
$$v^2 + \lambda w^2 = u^2.$$

把 $v = u - \lambda w$ 代入能量方程，得

$$u^2 - 2\lambda uw + \lambda^2 w^2 + \lambda w^2 = u^2 \quad \Longrightarrow \quad \lambda w\big((1+\lambda)w - 2u\big) = 0.$$

根 $w = 0$ 给出 $v = u$。这样 A_n 就会向前撞进静止的 A_{n+1}，这是不可能的。另一个根是

$$w = \frac{2u}{1+\lambda}, \qquad v = u - \lambda w = \frac{(1-\lambda)u}{1+\lambda}.$$

若 $\lambda < 1$： 则 $v > 0$，所以 A_n 离开静止的 A_{n-1}。此外

$$w - v = \frac{(1+\lambda)u}{1+\lambda} = u > 0,$$

所以 A_{n+1} 离开 A_n。两条守恒定律都成立，而且没有弹珠重叠，因此这个运动是可能的，其中 $A_0, \ldots, A_{n-1}$ 全部静止。

若 $\lambda > 1$： 则 $v < 0$，所以 A_n 必须向后运动。但 A_{n-1} 就在它的正后方，与它接触而且静止。因此，唯一同时满足两条守恒定律的运动在物理上并不可能，所以当 $\lambda > 1$ 时，这类运动并不存在。

第 (ii) 部分

假设只有 A_{n-1}、A_n 和 A_{n+1} 运动，速度分别为 a、b 和 c。由于 $n \geq 3$，弹珠 A_{n-2} 存在，它静止不动，并从后面与 A_{n-1} 接触。所以 $a \geq 0$。由于三颗运动中的弹珠不能互相超越，$a \leq b \leq c$。因此 a、b 和 c 全部非负。

> **想一想** 两条守恒定律都给出关于 a、b、c 的方程。怎样可以不解方程而比较它们?

我们把动量方程平方，再减去能量方程。由守恒得 $a + b + \lambda c = u$ 和 $a^2 + b^2 + \lambda c^2 = u^2$，所以

$$(a + b + \lambda c)^2 - (a^2 + b^2 + \lambda c^2) = u^2 - u^2 = 0.$$

展开左方，得

$$\lambda(\lambda - 1)c^2 + 2ab + 2\lambda c(a + b) = 0.$$

当 $\lambda > 1$ 且 $a, b, c \geq 0$ 时,左方每一项都非负,所以每一项都必须是零。特别地,$\lambda(\lambda-1)c^2 = 0$，由此得 $c = 0$。于是 $0 \leq a \leq b \leq c = 0$，三颗弹珠全部静止。由于 $u > 0$，这与动量方程矛盾。因此当 $\lambda > 1$ 时，这样的运动并不存在。

第 (iii) 部分

(ii) 的论证适用于任何一组运动中的弹珠：若所有速度都非负，而 A_{n+1} 运动，则动能会太小。所以当 $\lambda > 1$ 时，必定有某颗弹珠向后运动。除了 A_0 之外，每颗弹珠的正后方都有另一颗弹珠，所以 A_0 自然是最有可能反弹的一颗。

想一想 我们应该试哪两颗弹珠?

试试 A_0 和 A_{n+1}，速度分别为 p 和 q，而 $A_1, \ldots, A_n$ 保持静止。所得的两条方程与 (i) 中的相同，只是以 p, q 代替 v, w，而非平凡解（non-trivial solution）是

$$p = \frac{(1-\lambda)u}{1+\lambda} = -\frac{(\lambda-1)u}{\lambda+1}, \qquad q = \frac{2u}{1+\lambda}.$$

与 (i) 一样，动量和能量都守恒。这个运动在物理上也是可能的：

- $p < 0$，所以 A_0 向后运动，离开 A_1，而 A_0 后面没有任何东西。
- $q > 0$，所以 A_{n+1} 离开静止的 A_n。

所以其中一个可能的运动是：A_0 以速率 $\dfrac{(\lambda-1)u}{\lambda+1}$ 反弹，A_{n+1} 以速率 $\dfrac{2u}{\lambda+1}$ 向前运动，而 $A_1, \ldots, A_n$ 保持静止。

如果我们把这次碰撞看成相邻弹珠之间一连串独立的碰撞，结果也是如此。每对质量相等的弹珠会交换速度，所以速率 u 一直传递到 A_n。当 A_n 撞上 A_{n+1} 时，它以速度 p 反弹，然后这个向后的速度再沿着这列弹珠传回 A_0。

验算

- 当 $\lambda = 1$ 时，所有弹珠完全相同，两个答案都变成 A_{n+1} 以速率 u 向前运动，而其余全部静止，这就是大家熟悉的“牛顿摆”（Newton's cradle）。
- 取 $\lambda = 2$ 和 $u = 3$，第 (iii) 部分给出 $p = -1$ 和 $q = 2$。动量：$-1 + 2 \times 2 = 3$。能量：$1 + 2 \times 4 = 9 = 3^2$。
- 我们就 $n = 3$ 和 $n = 6$ 以数值方法逐步计算了这一连串两两之间的弹性碰撞（elastic collision）。当 $\lambda = 2$ 时，最终速度是 $-\frac{1}{3}u$ 和 $\frac{2}{3}u$；当 $\lambda = 4$ 时，则是 $-0.6u$ 和 $0.4u$，与预测完全相符。当 $\lambda = \frac{1}{2}$ 时，模拟结束时 A_n 和 A_{n+1} 分别以 $\frac{1}{3}u$ 和 $\frac{4}{3}u$ 运动，与 (i) 吻合。

常见错误

- 解了两条守恒方程便停下来。每一部分都取决于所得的解在物理上是否可能，即取决于各速度的正负，以及弹珠会否互相穿过。
- 在 (i) 中保留 $w = 0$、$v = u$ 这个根。这个根只是描述碰撞之前的运动。
- 忘记了在 (ii) 中，正是 $n \geq 3$ 保证了 A_{n-1} 后面有一颗静止的 A_{n-2}。

STEP 之外

对于一列互相接触的弹珠，单凭两条守恒定律并不能确定结果。这道题真正要问的是：在众多满足守恒量的解之中，哪些在物理上是可能的。要在它们之间作出选择，就需要一个更仔细的接触模型，例如一连串两两之间的碰撞，或把弹珠当作刚硬的弹簧。把弹珠当作弹簧来建立模型，就会引向质点链中的波动理论。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）就是以这种精神讲解碰撞和守恒定律的。

详细解答

STEP II 1990，第 15 题

概率与统计 · 顺序统计量

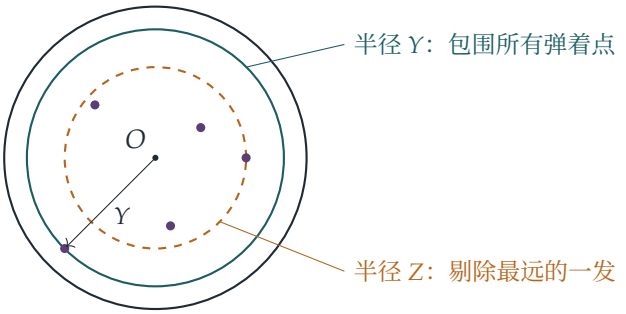
STEP II 1990，第 15 题

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-s2-q15

关键思路

以 O 为圆心、能包围一组弹着点的最小圆，其半径等于最远一发与圆心的距离，所以 Y 是若干个独立（independent）距离的最大值。最大值不超过 y，当且仅当每一发都落在距离 y 之内，于是累积分布函数（cumulative distribution function）便化为一个乘积。剔除最远的一发之后，我们要求的半径是第二大的距离，而它的分布函数可以通过数算有多少发可以落在给定的圆之外而求得。



解答

设 $R_1, \ldots, R_n$ 为 n 发射击的弹着点与 O 的距离。它们互相独立（因为各发是独立射出的），而且每一个的分布函数都是

$$P(R_i \le t) = t^2, \qquad 0 \le t \le 1.$$

Y 的分布

以 O 为圆心、包含所有弹着点的最小圆，半径为 $Y = \max(R_1, \ldots, R_n)$。对于 $0 \le y \le 1$，

$$P(Y \le y) = P(R_1 \le y, \ldots, R_n \le y) = \prod_{i=1}^{n} P(R_i \le y) = (y^2)^n = y^{2n},$$

其中中间一步用到独立性。求导后，Y 的概率密度函数（probability density function）为

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} y^{2n} = 2ny^{2n-1}, \qquad 0 \le y \le 1,$$

正如题目所求。（它在 $[0,1]$ 之外为零，而且积分为 1，理应如此。）

圆的面积为 πY^2，所以

$$E(\pi Y^2) = \pi \int_0^1 y^2 \cdot 2ny^{2n-1}\,\mathrm{d}y = 2n\pi \int_0^1 y^{2n+1}\,\mathrm{d}y = \frac{2n\pi}{2n+2} = \frac{n\pi}{n+1}.$$

剔除最远的一发

我们需要 $n \ge 2$，这样才会有弹着点剩下。剔除最远的一发之后，以 O 为圆心的最小包围圆的半径为 Z，即 $R_1, \ldots, R_n$ 之中第二大的一个。

> ? **想一想** 怎样用“有多少发落在半径为 z 的圆之外”来描述事件 $Z \le z$?

第二大的距离不超过 z，当且仅当最多只有一发与 O 的距离大于 z。每一发各自独立地以概率 z^2 落在距离 z 之内，所以落在半径为 z 的圆之外的发数服从二项分布（binomial distribution），$\mathrm{B}(n, 1-z^2)$。因此，对于 $0 \le z \le 1$，

$$\begin{aligned} P(Z \le z) &= P(\text{没有一发在圆外}) + P(\text{恰好一发在圆外}) \\ &= z^{2n} + n(1-z^2)z^{2(n-1)} \\ &= nz^{2n-2} - (n-1)z^{2n}. \end{aligned}$$

求导便得 Z 的密度函数：

$$f_Z(z) = 2n(n-1)z^{2n-3} - 2n(n-1)z^{2n-1} = 2n(n-1)z^{2n-3}(1-z^2), \qquad 0 \le z \le 1.$$

因此新圆的面积的期望值（expected value）为

$$\begin{aligned} E(\pi Z^2) &= 2n(n-1)\pi \int_0^1 \left(z^{2n-1} - z^{2n+1}\right)\mathrm{d}z \\ &= 2n(n-1)\pi \left(\frac{1}{2n} - \frac{1}{2n+2}\right) \\ &= 2n(n-1)\pi \cdot \frac{2}{2n(2n+2)} = \left(\frac{n-1}{n+1}\right)\pi, \end{aligned}$$

这正是所求的结果。

验算

- 当 $n = 1$ 时，第一个答案是 $\pi/2$：单独一发的 R^2 在 $[0,1]$ 上服从均匀分布（uniform distribution）（因为 $P(R^2 \le u) = u$），所以面积 πR^2 的期望值是 $\pi/2$。
- 当 $n = 2$ 时，第二个答案是 $\pi/3$，即两发之中较近一发所对应的面积期望值；连同第一个答案 $2\pi/3$，两者的平均是 $\pi/2$，即单独一发时的值，这是必然的。
- 以 2×10^5 次试验及 $n = 5$ 进行模拟（simulation），得到的面积期望值为 2.618 和 2.095，而理论值为 $5\pi/6 \approx 2.618$ 和 $4\pi/6 \approx 2.094$。

常见错误

- 写成 $P(Z \le z) = z^{2(n-1)}$，仿佛余下的弹着点是一个全新的、大小为 $n-1$ 的样本；被剔除的一发并不是随机选出的，所以留下来的弹着点会偏向 O。
- 忘记了因子 n，它代表落在圆外的是哪一发。
- 先求 $E(Y)$ 再把它平方：面积的期望值是 $\pi E(Y^2)$，而不是 $\pi\big(E(Y)\big)^2$。

STEP 之外

这些就是顺序统计量（order statistics）。由于 $U_i = R_i^2$ 在 $[0,1]$ 上服从均匀分布，这些面积其实就是 n 个均匀随机变量的顺序统计量，而第 k 小的一个（取自 n 个独立的均匀随机变量）服从贝塔分布（beta distribution），平均值为 $k/(n+1)$；本题的两个答案分别是 $k = n$ 和 $k = n-1$ 的情况。一个简洁的图像是：n 个随机点把 $[0,1]$ 分割成 $n+1$ 段期望长度相等的间隙。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）涵盖了这类变换和顺序统计量。

详细解答

STEP II 1990，第 16 题

概率与统计 • 贝叶斯定理

STEP II 1990，第 16 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-s2-q16

关键思路

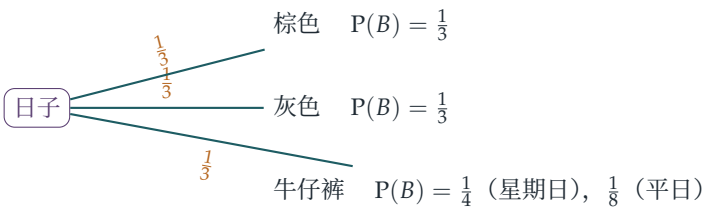
这其实是乔装了的贝叶斯定理（Bayes' theorem）。照片告诉我们当天系的是黑色领带，而我们要求的是已知这一点时，那天是某一天的概率。所以我们先顺着推：对每一类日子，用一幅按裤子分支的树形图（tree diagram），求出系黑色领带的概率。然后再把条件概率（conditional probability）倒转过来。星期一的规则只改动树形图的其中一条分支，所以第二部分几乎不用多花工夫。

解答

设 S 为照片摄于星期日的事件，B 为我系着黑色领带的事件。照片出自七天中任何一天的可能性均等，所以在看照片之前，$\mathrm{P}(S) = \frac{1}{7}$。除了（稍后的）星期一之外，每一天穿每条裤子的概率都是 $\frac{1}{3}$。

各类日子系黑色领带的机会

- **棕色裤子**。领巾已被排除，所以领带是从黑色、波点和军团领带三者中随机选出：$\mathrm{P}(B) = \frac{1}{3}$。
- **灰色裤子**。波点领带已被排除，所以是从黑色领带、军团领带和领巾中选：$\mathrm{P}(B) = \frac{1}{3}$。
- **星期日穿牛仔裤**。一定会系领带，从全部四条中选：$\mathrm{P}(B) = \frac{1}{4}$。
- **平日穿牛仔裤**。有 $\frac{1}{2}$ 的概率穿运动衫而不系领带。否则就是四条领带的其中一条：$\mathrm{P}(B) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$。



沿各分支相加（即按裤子划分的全概率公式（law of total probability）），

$$\mathrm{P}(B \mid \text{星期日}) = \tfrac{1}{3}\left(\tfrac{1}{3}+\tfrac{1}{3}+\tfrac{1}{4}\right) = \tfrac{1}{3}\cdot\tfrac{11}{12} = \tfrac{11}{36} = \tfrac{22}{72},$$
$$\mathrm{P}(B \mid \text{平日}) = \tfrac{1}{3}\left(\tfrac{1}{3}+\tfrac{1}{3}+\tfrac{1}{8}\right) = \tfrac{1}{3}\cdot\tfrac{19}{24} = \tfrac{19}{72}.$$

第一个答案

根据贝叶斯定理，星期日的先验概率（prior）为 $\frac{1}{7}$，六个平日合计为 $\frac{6}{7}$，

$$\mathrm{P}(S \mid B) = \frac{\frac{1}{7}\cdot\frac{22}{72}}{\frac{1}{7}\cdot\frac{22}{72}+\frac{6}{7}\cdot\frac{19}{72}} = \frac{22}{22+6\times 19} = \frac{22}{136} = \frac{11}{68},$$

正如所求。（公因数 $\frac{1}{7\times 72}$ 会约去，这就是为甚么值得把所有分数都通分成以 72 为分母。）

> ? **想一想** 为甚么答案不直接是 $\mathrm{P}(B \mid \text{星期日})$，或者 $\frac{1}{7}$?

$\mathrm{P}(B \mid \text{星期日}) = \frac{11}{36}$ 答的是另一条问题：它是已知日子时系这条领带的机会，而不是已知领带时那天是星期日的机会。而 $\frac{1}{7}$ 则忽略了证据。星期日系黑色领带的可能性稍高一点（$\frac{22}{72}$ 对 $\frac{19}{72}$），因为那天没有不系领带这个选择，所以这张照片把星期日的概率由 $\frac{1}{7} = \frac{11}{77}$ 轻轻推高至 $\frac{11}{68}$。

加入星期一的规则

现在星期一必定穿牛仔裤，而星期一仍然是平日，所以运动衫这个选择依然适用：

$$\mathrm{P}(B \mid \text{星期一}) = 1\cdot\tfrac{1}{2}\cdot\tfrac{1}{4} = \tfrac{1}{8} = \tfrac{9}{72}.$$

星期日和其余五个平日不变。每一天的先验概率仍然是 $\frac{1}{7}$，可以约去，所以

$$\mathrm{P}(S \mid B) = \frac{\frac{22}{72}}{\frac{22}{72}+\frac{9}{72}+5\cdot\frac{19}{72}} = \frac{22}{22+9+95} = \frac{22}{126} = \frac{11}{63}.$$

这比 $\frac{11}{68}$ 稍大一点。这是合理的：星期一系黑色领带的可能性降低了（由 $\frac{19}{72}$ 变成 $\frac{1}{8}$），所以一张系黑色领带的照片会令星期一的可能性下降，而稍为更倾向星期日。

验算

- 每一天各条领带的概率，连同“不系领带”，加起来等于 1。例如，平日不系领带的机会是 $\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{6}$。
- 模拟 4×10^5 个随机日子，得出系黑色领带的日子当中星期日所占的比例约为 0.161（准确值 $\frac{11}{68}\approx0.1618$），加入星期一的规则后则约为 0.174（准确值 $\frac{11}{63}\approx0.1746$）。

常见错误

- 忘记了星期一穿牛仔裤时，仍然有一半机会穿运动衫：星期一是平日。星期一若用 $\frac{1}{4}$ 而不是 $\frac{1}{8}$，会得出 $\frac{22}{135}$。
- 穿棕色或灰色裤子时从全部四条领带中选。这两条裤子各自只剩下三条领带可选。

STEP 之外

把“已知假设时证据的概率”转化为“已知证据时假设的概率”，正是贝叶斯推断（Bayesian inference）的核心。有一个简洁的说法：后验赔率（posterior odds）等于先验赔率乘以似然比（likelihood ratio）。这里星期日对平日的先验赔率是 1 : 6，似然比是 22 : 19，于是得出 22 : 114，即 11 : 57。大学课程会由此发展出贝叶斯统计学（Bayesian statistics），当中整个分布都以同样的方式更新。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门）有一章讲条件概率，写得十分出色，并附有许多像本题这样的例子。

详细解答

STEP III 1990，第 1 题

纯数学 · 倍角与多项式的根

> **STEP III 1990，第 1 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-q1

关键思路

九等于 3×3，所以只要把正切的三倍角公式（triple-angle formula）用两次，就可以得到 $\tan 9\theta$。这比展开 $(\cos\theta+\mathrm{i}\sin\theta)^9$ 省事得多。因式分解后的形式还告诉我们，在 $\theta=\pi/9,\,2\pi/9,\,4\pi/9$ 时是哪一个因式等于零：这些角正好是令 $\tan^2 3\theta=3$ 的角。由此得出一个以 $u=\tan^2\theta$ 为未知数的三次方程（cubic），余下的工作就交给韦达公式（Vieta's formulas，即多项式的根与系数之间的关系）。

解答

$\tan 9\theta$ 的公式

设 $t=\tan\theta$ 及 $T=\tan 3\theta$。三倍角公式可由 $\tan(2\theta+\theta)$ 推出，亦可由棣美弗定理（de Moivre's theorem）取 $n=3$ 得出，它给出

$$T=\frac{3t-t^3}{1-3t^2}=\frac{t(t^2-3)}{3t^2-1},\qquad \tan 9\theta=\frac{3T-T^3}{1-3T^2}=\frac{T(3-T^2)}{1-3T^2}.$$

现在把 $3-T^2$ 和 $1-3T^2$ 都写成以 $(1-3t^2)^2$ 为公分母的形式：

$$\begin{aligned}(1-3t^2)^2(3-T^2)&=3(1-3t^2)^2-t^2(3-t^2)^2\\&=(3-18t^2+27t^4)-(9t^2-6t^4+t^6)\\&=-(t^6-33t^4+27t^2-3),\end{aligned}$$

$$\begin{aligned}(1-3t^2)^2(1-3T^2)&=(1-3t^2)^2-3t^2(3-t^2)^2\\&=(1-6t^2+9t^4)-(27t^2-18t^4+3t^6)\\&=-(3t^6-27t^4+33t^2-1).\end{aligned}$$

两式相除时，因子 $-(1-3t^2)^{-2}$ 互相约去，所以

$$\tan 9\theta = \frac{t(t^2-3)}{3t^2-1} \cdot \frac{t^6-33t^4+27t^2-3}{3t^6-27t^4+33t^2-1},$$

这就是所求的结果。

若改用棣美弗定理，则 $\tan 9\theta = \operatorname{Im}(1+\mathrm{i}t)^9 / \operatorname{Re}(1+\mathrm{i}t)^9$。这样便要再验证

$$9t-84t^3+126t^5-36t^7+t^9 \quad \text{和} \quad 1-36t^2+126t^4-84t^6+9t^8$$

确实可以分解成上面的分子和分母。它们的确可以，但经由 $\tan 3\theta$ 这条路，因式分解是现成的，不费工夫。

三次方程

设 $P(t) = t^6-33t^4+27t^2-3$。上面第一个计算表明

$$3-\tan^2 3\theta = -\frac{P(t)}{(1-3t^2)^2} \qquad \text{只要} 3t^2 \neq 1.$$

现在取 $\theta = \pi/9,\ 2\pi/9,\ 4\pi/9$。这时 $3\theta = \pi/3,\ 2\pi/3,\ 4\pi/3$，而每一个情况都有 $\tan 3\theta = \pm\sqrt{3}$，所以 $\tan^2 3\theta = 3$。每一个情况下 3θ 都不是 $\pi/6$ 的奇数倍，所以 $3t^2 \neq 1$（否则 $\tan 3\theta$ 便没有定义）。因此，对这三个角都有 $P(\tan\theta) = 0$。

在 $(0, \pi/2)$ 之内，这些正好是令 $\tan 9\theta = 0$ 的原因在于因式 P，而不在于 t 或 t^2-3 的角。事实上，$t \neq 0$，而且因为 $\theta \neq \pi/3$，所以 $t^2 \neq 3$。

想一想 我们怎样知道这三个数就是该三次方程的全部根，而且没有重根?

令 $u = t^2$。$\tan^2(\pi/9)$、$\tan^2(2\pi/9)$、$\tan^2(4\pi/9)$ 每一个都是

$$u^3-33u^2+27u-3 = 0.$$

的根。这三个值互不相同，因为 tan 在 $(0, \pi/2)$ 上取正值且严格递增 (strictly increasing)，而 $\pi/9 < 2\pi/9 < 4\pi/9 < \pi/2$。三次方程最多有三个根，所以这三个数正好就是它的全部根。这就是所求的整数系数三次方程。

乘积

根据韦达公式，$u^3-33u^2+27u-3$ 的根之积是 3:

$$\tan^2\left(\tfrac{\pi}{9}\right)\tan^2\left(\tfrac{2\pi}{9}\right)\tan^2\left(\tfrac{4\pi}{9}\right) = 3.$$

三个角都在 $(0, \pi/2)$ 之内，所以每个正切值都是正数，我们取正平方根:

$$\tan\left(\tfrac{\pi}{9}\right)\tan\left(\tfrac{2\pi}{9}\right)\tan\left(\tfrac{4\pi}{9}\right) = \sqrt{3}.$$

六次方之和

设三个根为 u_1, u_2, u_3。由该三次方程可知

$$e_1 = u_1 + u_2 + u_3 = 33, \qquad e_2 = u_1u_2 + u_2u_3 + u_3u_1 = 27, \qquad e_3 = u_1u_2u_3 = 3.$$

我们要求的是 $u_1^3 + u_2^3 + u_3^3$，也就是三个正切值的六次方之和。

> **想一想** 不把根求出来，怎样由 e_1, e_2, e_3 得到立方和?

每个根都满足 $u_i^3 = 33u_i^2 - 27u_i + 3$。把这三条等式相加，得

$$\sum u_i^3 = 33\sum u_i^2 - 27\sum u_i + 9.$$

这里 $\sum u_i = 33$，而

$$\sum u_i^2 = e_1^2 - 2e_2 = 1089 - 54 = 1035.$$

因此

$$\begin{aligned}\sum u_i^3 &= 33 \times 1035 - 27 \times 33 + 9\\ &= 34155 - 891 + 9\\ &= 33273.\end{aligned}$$

也就是说，

$$\tan^6\left(\tfrac{\pi}{9}\right) + \tan^6\left(\tfrac{2\pi}{9}\right) + \tan^6\left(\tfrac{4\pi}{9}\right) = 33273,$$

正如所求。等价地，$\sum u_i^3 = e_1^3 - 3e_1e_2 + 3e_3 = 35937 - 2673 + 9$。

验算

- 以数值计算，$\tan^2(\pi/9) \approx 0.1325$、$\tan^2(2\pi/9) \approx 0.7041$ 及 $\tan^2(4\pi/9) \approx 32.163$。它们的和是 33.000，积是 3.000。
- 三个正切值的积是 $1.7320508\ldots = \sqrt{3}$。
- 上述三个值的立方和是 33273.0。它几乎全部来自 $\tan^6(4\pi/9) \approx 33272.4$。
- 随机取数千个 θ 的值，$\tan 9\theta$ 的公式与计算机的结果吻合至约十二位有效数字。
- 简单情况：取 $t = 1$，即 $\theta = \pi/4$。公式给出 $\frac{1\cdot(-2)(-8)}{2\cdot 8} = 1 = \tan(9\pi/4)$。

常见错误

- 未先检查 $3t^2 \neq 1$（或分母不为零），就断定那个六次因式等于零。

- 把乘积取为 $-\sqrt{3}$。每个角都是锐角，所以每个正切值都是正数。
- 写成 $\sum u_i^3 = e_1^3 - 3e_1e_2$ 而漏掉 $+3e_3$ 一项，结果得出 33264。

STEP 之外

三次多项式 $u^3 - 33u^2 + 27u - 3$ 没有有理根，所以它在 $\mathbb{Q}$ 上不可约（irreducible）。它又有三个实根。伽罗华理论（Galois theory）表明，这类三次多项式的根不能单靠实数根式表示出来（即所谓*不可约情形*（casus irreducibilis））。正因如此，像 $\tan^2(\pi/9)$ 这样的三角表达式才是写出这些根的自然方式。至于根的幂和，一般由牛顿恒等式（Newton's identities）支配。想了解棣美弗定理的几何一面，可读 Needham 的 *Visual Complex Analysis*（以图像方式讲解复分析）。代数方面，可读 Allenby 的 *Rings, Fields and Groups*（环、域与群的入门书）。

详细解答

STEP III 1990，第 2 题

纯数学 · 三维向量

STEP III 1990，第 2 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-q2

关键思路

四点全部位于同一条曲线 $t \mapsto (t^3, t^2, t)$ 上，这条曲线称为三次挠线（twisted cubic）。两条相交的直线必定位于同一平面上，所以第 (i) 部分归结为证明 O、A、B、C 永不共面（coplanar），而这是一个行列式（determinant），它可以像范德蒙行列式（Vandermonde determinant）那样因式分解。在第 (ii) 部分，条件 $ab = 1$ 令纯量积（scalar product）和长度都化为单一个量 $a^2 + 1/a^2$ 的函数。

解答

首先，“相异”告诉我们甚么？参数为 t 的点，其第三个坐标就是 t 本身，所以不同的点有不同的参数。因此 a、b、c 互不相同，而由于 O 是 $t = 0$ 所对应的点，a、b、c 均不为零。

第 (i) 部分

假设直线 OA 与 BC 确实相交。通过同一点的两条直线位于同一平面上（若两线刚好重合，任何包含它们的平面皆可）。于是 O、A、B、C 全部位于同一个通过原点的平面上，而位置向量（position vector）$\mathbf{a} = \overrightarrow{OA}$、$\mathbf{b} = \overrightarrow{OB}$、$\mathbf{c} = \overrightarrow{OC}$ 就会线性相关（linearly dependent）。这种情况恰好在以下条件成立时出现：

$$D = \begin{vmatrix} a^3 & a^2 & a \\ b^3 & b^2 & b \\ c^3 & c^2 & c \end{vmatrix} = 0.$$

从第一行抽出因子 a，第二行抽出 b，第三行抽出 c，

$$D = abc\begin{vmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{vmatrix}.$$

把首两行各自减去第三行，然后沿最后一列展开：

$$D = abc\begin{vmatrix} (a-c)(a+c) & a-c & 0 \\ (b-c)(b+c) & b-c & 0 \\ c^2 & c & 1 \end{vmatrix} = abc\,(a-c)(b-c)\begin{vmatrix} a+c & 1 \\ b+c & 1 \end{vmatrix}$$
$$= abc\,(a-b)(a-c)(b-c).$$

每个因子都不为零：$a, b, c \neq 0$，而且 a、b、c 互不相同。所以 $D \neq 0$，四点不共面，因此直线 OA 与 BC 不可能相交。（两线亦不平行，所以它们是异面直线（skew lines）。）

想一想 为甚么只需证明四点不共面便已足够?

因为相交的直线必定张成一个平面。假如 OA 与 BC 相交于点 P，那么通过 P 并包含这两条直线的平面就会包含 O、A、B 和 C。我们已证明这样的平面并不存在。

第 (ii) 部分

设 $b = 1/a$。那么 $a \neq 0$，并且 $a \neq \pm 1$，否则 $b = a$，A 与 B 便会重合。纯量积为

$$\mathbf{a}\cdot\mathbf{b} = (ab)^3 + (ab)^2 + ab = 3,$$

而长度的平方为

$$|\mathbf{a}|^2 = a^2(a^4 + a^2 + 1), \qquad |\mathbf{b}|^2 = \frac{1}{a^6} + \frac{1}{a^4} + \frac{1}{a^2} = \frac{a^4 + a^2 + 1}{a^6}.$$

两式相乘，得 $|\mathbf{a}|^2|\mathbf{b}|^2 = (a^4 + a^2 + 1)^2/a^4$，所以 $|\mathbf{a}||\mathbf{b}| = (a^4 + a^2 + 1)/a^2$，这是正数。记 $\theta = \angle AOB$，则

$$\cos\theta = \frac{3a^2}{a^4 + a^2 + 1} = \frac{3}{u+1}, \qquad \text{其中} u = a^2 + \frac{1}{a^2}.$$

哪些值可以出现。根据算术几何平均不等式(AM–GM)(或由于 $a^2 + a^{-2} - 2 = (a - a^{-1})^2$)，$u \geq 2$，等号仅在 $a^2 = 1$ 时成立，而这个情况已被排除。所以 $u > 2$，因此

$$0 < \cos\theta = \frac{3}{u+1} < \frac{3}{3} = 1,$$

由此得 $0 < \theta < \frac{1}{2}\pi$。其他角度都不可能出现。

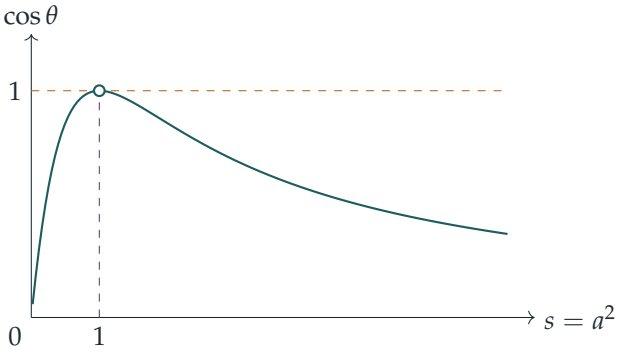
每个这样的值都会出现。任取满足 $0 < \theta < \frac{1}{2}\pi$ 的 θ，于是 $0 < \cos\theta < 1$。我们需要 $u = 3/\cos\theta - 1$，而这个值大于 2。方程 $a^2 + 1/a^2 = u$ 是以 a^2 为未知数的二次方程 $(a^2)^2 - u\,a^2 + 1 = 0$，它有一个根

$$a^2 = \frac{u + \sqrt{u^2 - 4}}{2} > 1,$$

由于 $u > 2$，这个根是实数。取 a 为其正平方根，并取 $b = 1/a$。那么 $a > 1 > b > 0$，所以 $ab = 1$，$A \neq B$，两点都不是 O，并且 $\cos\angle AOB = 3/(u+1) = \cos\theta$。因此 $\angle AOB = \theta$，因为两个角都在 $[0, \pi]$ 之内，而余弦函数在这个区间上是一对一（one-to-one）的。

所以 $\angle AOB$ 可取 $0 < \angle AOB < \frac{1}{2}\pi$ 内的每一个值，而不能取其他值，正如题目所求。

下图显示 $\cos\theta$ 对 $s = a^2$ 的图像。曲线在被排除的值 $s = 1$ 处升至 1，并在两端趋向 0，所以它的值恰好填满 $(0, 1)$。



验算

- 试取 $a = 2$、$b = \frac{1}{2}$：$A = (8, 4, 2)$ 和 $B = (\frac{1}{8}, \frac{1}{4}, \frac{1}{2})$ 给出 $\mathbf{a}\cdot\mathbf{b} = 1 + 1 + 1 = 3$，而公式给出 $\cos\theta = 12/21 = 4/7$，所以 $\theta \approx 55.2°$，确实在区间之内。
- 当 $a \to 1$ 时，点 A 与 B 合而为一，$\theta \to 0$。当 $a \to \infty$ 时，A 几乎沿著 x 轴远去，而 B 几乎沿著 z 轴趋近 O，所以 $\theta \to \frac{1}{2}\pi$。这正是区间的两端。
- D 的因式分解已用数千组随机有理数 a、b、c 作精确验证，角度公式及其值域亦已用数值方法验证。

常见错误

- 忘记说明为何 a、b、c 互不相同且不为零。正是这些事实令 $D \neq 0$。
- 没有排除 $a = \pm 1$，单凭 $\cos\theta < 1$ 就下结论。严格不等式 $u > 2$ 才是排除 $\theta = 0$ 的关键。

- 只证明“没有其他值”。题目还要求你证明区间内每一个值都能取到，这需要明确给出一个 a。

STEP 之外

曲线 (t^3, t^2, t) 就是三次挠线，是代数几何（algebraic geometry）中的标准例子。第 (i) 部分所说的，就是这条曲线上任意四点都不共面，而你计算的行列式是一个范德蒙行列式，它在多项式插值（polynomial interpolation）和线性独立（linear independence）的论证中会再次出现。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对行列式和线性独立有清晰的讲解。

详细解答

STEP III 1990，第 3 题

纯数学 · 群公理与共轭

> **STEP III 1990，第 3 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-q3

关键思路

三个部分都只需用到群公理（group axioms），而每一部分中有用的一步，都是把一个元素改写成以另一个元素表示。阶（order）为 2 的元素是它自己的逆元（inverse），这使 (i) 变成一行计算。在 (ii) 和 (iii) 中，所用的工具是*共轭*（*conjugation*）：$x \mapsto c^{-1}xc$ 保持乘积，所以它把幂映到幂。

全文中我们把 $x * y$ 写作 xy，并以 e 表示单位元（identity）。

解答

第 (i) 部分

元素 x 的阶为 2，当且仅当 $x \neq e$ 且 $x^2 = e$；第二个条件即是说 $x^{-1} = x$。所以

$$a^{-1} = a, \qquad b^{-1} = b, \qquad (ab)^{-1} = ab.$$

对任何群元素，都有 $(ab)^{-1} = b^{-1}a^{-1}$，因为 $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = e$。因此

$$ab = (ab)^{-1} = b^{-1}a^{-1} = ba,$$

所以 a 与 b 可交换（commute）。

第 (ii) 部分

首先留意 dc 是 cd 的一个共轭元：

$$c^{-1}(cd)\,c = (c^{-1}c)\,dc = dc.$$

想一想 为甚么写成 $dc = c^{-1}(cd)c$ 对处理阶有帮助?

因为共轭可以穿过幂。对任何 x 和任何正整数 n,

$$(c^{-1}xc)^n = (c^{-1}xc)(c^{-1}xc)\cdots(c^{-1}xc) = c^{-1}x^n c,$$

因为中间每一对相邻的 $c\,c^{-1}$ 都互相抵消。取 $x = cd$，便得 $(dc)^n = c^{-1}(cd)^n c$。于是

$$(dc)^n = e \iff c^{-1}(cd)^n c = e \iff (cd)^n = cec^{-1} = e,$$

其中第二步是左乘 c 并右乘 c^{-1}（反方向则乘以 c^{-1} 和 c）。所以满足 $(cd)^n = e$ 的正整数 n，正好就是满足 $(dc)^n = e$ 的那些正整数。阶就是这样的 n 中最小的一个，所以 cd 与 dc 的阶相同。若不存在这样的 n，两者的阶都是无限，所以命题在这个情况下同样成立。

第 (iii) 部分

已知 $c^{-1}bc = b^r$。对正整数 s，用 (ii) 中逐项相消（telescoping）的论证，可得

$$c^{-1}b^s c = (c^{-1}bc)^s = (b^r)^s = b^{sr}.$$

当 $s = 0$ 时同样成立（两边都是 e）。当 s 为负数时，设 $s = -t$，其中 $t > 0$；则 $c^{-1}b^s c = \left(c^{-1}b^t c\right)^{-1}$，因为 $(c^{-1}b^t c)(c^{-1}b^{-t}c) = e$。所以

$$c^{-1}b^s c = \left(b^{tr}\right)^{-1} = b^{-tr} = b^{sr}.$$

因此对每个整数 s，都有 $c^{-1}b^s c = b^{sr}$。

至于第二个结果，我们对 $n \geq 1$ 使用归纳法（induction），并同时处理所有整数 s。$n = 1$ 的情况就是刚才所证明的结果。假设 $c^{-n}b^s c^n = b^{sr^n}$ 对所有整数 s 成立。则

$$\begin{aligned} c^{-(n+1)}\, b^s\, c^{n+1} &= c^{-1}\left(c^{-n}b^s c^n\right)c \\ &= c^{-1}\, b^{sr^n}\, c \\ &= b^{(sr^n)r} = b^{sr^{n+1}}, \end{aligned}$$

其中最后一行用了第一个结果，并以整数 sr^n 代替 s。归纳法至此完成，所以 $c^{-n}b^s c^n = b^{sr^n}$ 对所有正整数 n 成立。

想一想 为甚么我们要特地为负数 s 证明第一个结果?

若 r 是负数（例如在二面体群（dihedral group）中 $c^{-1}bc = b^{-1}$），则即使 s 是正数，sr^n 也可能是负数，而归纳步骤正是以 sr^n 代替 s 来应用第一个结果。为所有整数 s 证明第一个结果，归纳法便不论 r 的正负都行得通。

验算

用小的群来测试。在正方形的对称群中，沿两条对角线的反射及它们的乘积（旋转半周）的阶都是 2，而这两个反射确实可交换，与 (i) 的预测一致。在 S_3 中，取 $c = (12)$ 和 $d = (13)$：则 cd 和 dc 就是那两个 3-循环（3-cycle），阶都是 3。至于 (iii)，在二面体群中，反射 c 与旋转 b 满足 $c^{-1}bc = b^{-1}$，所以 $r = -1$，而公式预测 $c^{-n}b^s c^n = b^{(-1)^n s}$，这是对的：以同一个反射共轭两次，等于甚么也没有做。我们亦用电脑验证了全部三个命题，穷举了 S_4 中所有合适的元素（就 (i) 而言还包括 S_5），以及模 13 之下由映射 $x \mapsto ux + v$ 所组成的群。

常见错误

- 写成 $(ab)^{-1} = a^{-1}b^{-1}$。次序是会倒转的；正是这个倒转令 (i) 得以成立。
- 在 (ii) 中，只证明 $(cd)^n = e \Rightarrow (dc)^n = e$。两个方向都要证明（或利用对称性：把 c 和 d 的角色互换），而阶为无限的情况亦应提及。
- 在 (iii) 中，以三点省略号代替归纳法，为第二个结果写出一个所谓的“证明”。

STEP 之外

映射 $x \mapsto c^{-1}xc$ 是这个群的一个*自同构*（*automorphism*）：它是由 G 到自身的双射（bijection），并且保持运算，所以它保持元素的阶，并把幂映到幂。第 (iii) 部分是*半直积*（*semidirect products*）中的关键关系：由 b 和 c 生成并满足 $c^{-1}bc = b^r$ 的群，包括二面体群（$r = -1$）以及模一个质数之下由映射 $x \mapsto ux + v$ 所组成的群。Allenby 的 *Rings, Fields and Groups*（环、域与群）对共轭及这些例子有仔细的阐述。

详细解答

STEP III 1990，第 4 题

纯数学 · 三角级数

STEP III 1990，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/90-s3-q4

关键思路

两个级数（series）都是余弦之和，而且角度按等差递增，所以有两件顺手的工具。对于单纯的和，可乘以 $2\sin\beta$ 并使用积化和差公式（product-to-sum formula）：每一项都拆成两个正弦之差，整个和便会裂项相消（telescope）。对于二项式的和，以及最后的恒等式，可把 $\cos\phi$ 看成 $\mathrm{e}^{\mathrm{i}\phi}$ 的实部（real part）。这样，二项式定理（binomial theorem）或等比级数（geometric series）公式便会代劳，最后取实部或虚部（imaginary part）就得出答案。

解答

第一个级数

把这个和记作 S。根据积化和差公式 $2\sin A\cos B=\sin(B+A)-\sin(B-A)$，

$$2\sin\beta\cos(\alpha+2r\beta)=\sin\big(\alpha+(2r+1)\beta\big)-\sin\big(\alpha+(2r-1)\beta\big).$$

把 $r=0,1,\ldots,n$ 的各式相加，和便裂项相消：对应 r 的第一个正弦，与对应 $r+1$ 的第二个正弦互相抵消。因此只剩下两项：

$$\begin{aligned}2\sin\beta\,S&=\sin\big(\alpha+(2n+1)\beta\big)-\sin(\alpha-\beta)\\&=2\cos(\alpha+n\beta)\sin\big((n+1)\beta\big),\end{aligned}$$

这里用了 $\sin P-\sin Q=2\cos\frac{P+Q}{2}\sin\frac{P-Q}{2}$，其中 $\frac{P+Q}{2}=\alpha+n\beta$ 及 $\frac{P-Q}{2}=(n+1)\beta$。由于 $\sin\beta\neq0$，我们可以除以它，得

$$\boxed{\sum_{r=0}^{n}\cos(\alpha+2r\beta)=\frac{\sin(n+1)\beta\,\cos(\alpha+n\beta)}{\sin\beta}}$$

留意 $\alpha + n\beta$ 正是首尾两个角的平均值，这可用来粗略核对答案的形式是否合理。

二项式级数

把这个和记作 T。由于 $\cos\phi = \operatorname{Re} e^{i\phi}$，而二项式系数（binomial coefficient）都是实数，

$$T = \operatorname{Re} \sum_{r=0}^{n} \binom{n}{r} e^{i\alpha} \left(e^{2i\beta}\right)^r = \operatorname{Re}\left[e^{i\alpha}\left(1 + e^{2i\beta}\right)^n\right]$$

这是根据二项式定理得出的。现在抽出“半角”因子：

$$1 + e^{2i\beta} = e^{i\beta}\left(e^{-i\beta} + e^{i\beta}\right) = 2\cos\beta\, e^{i\beta}.$$

因此 $e^{i\alpha}(1 + e^{2i\beta})^n = 2^n \cos^n\beta\, e^{i(\alpha+n\beta)}$，而由于 $2^n\cos^n\beta$ 是实数，

$$\boxed{\sum_{r=0}^{n} \binom{n}{r} \cos(\alpha + 2r\beta) = 2^n \cos^n\beta\, \cos(\alpha + n\beta)}$$

（这一个其实并不需要 $\sin\beta \neq 0$。）

最后的恒等式

式中出现正割，表示我们需要 $\cos\theta \neq 0$，这是题目隐含的条件。第 r 项是 $\cos r\theta\, \sec^r\theta = \operatorname{Re}\left(e^{i\theta}\sec\theta\right)^r$，因为 $\sec^r\theta$ 是实数。于是设

$$w = e^{i\theta}\sec\theta = \frac{\cos\theta + i\sin\theta}{\cos\theta} = 1 + i\tan\theta.$$

左方就是 $\operatorname{Re}(1 + w + \cdots + w^n)$。

> **想一想** 为甚么在这里可以放心使用等比级数公式？w 又为甚么有帮助？

这条公式需要 $w \neq 1$，而 $w - 1 = i\tan\theta \neq 0$ 正正是因为 $\sin\theta \neq 0$。题目的条件就是用在这里。引入 w 的好处在于 $w - 1$ 是纯虚数（purely imaginary），所以除以它很容易，而 w^{n+1} 则可由指数形式 $w = \sec\theta\, e^{i\theta}$ 轻易算出。设 $t = \tan\theta$，

$$1 + w + \cdots + w^n = \frac{w^{n+1} - 1}{w - 1} = \frac{\sec^{n+1}\theta\, e^{i(n+1)\theta} - 1}{it}.$$

对任意复数 z 及实数 $t \neq 0$，我们有 $\operatorname{Re}\dfrac{z}{it} = \operatorname{Re}\dfrac{-iz}{t} = \dfrac{\operatorname{Im} z}{t}$。分子中的 -1 虚部为零，所以

$$\sum_{r=0}^{n} \cos r\theta\, \sec^r\theta = \frac{\sec^{n+1}\theta\, \sin(n+1)\theta}{\tan\theta} = \sec^{n+1}\theta\, \sin(n+1)\theta \cdot \frac{\cos\theta}{\sin\theta}$$
$$= \frac{\sin(n+1)\theta\, \sec^n\theta}{\sin\theta},$$

这就是所求的结果。

*与第一部分的联系。*把全式乘以 $\cos^n\theta$，恒等式便变成

$$\sum_{r=0}^{n}\cos r\theta\cos^{n-r}\theta=\frac{\sin(n+1)\theta}{\sin\theta}.$$

你也可以对 n 用归纳法（induction）证明这个形式：若 C_n 为左方，则 $C_n=\cos\theta\,C_{n-1}+\cos n\theta$，而 $\sin n\theta\cos\theta+\cos n\theta\sin\theta=\sin(n+1)\theta$ 便完成归纳步骤。

验算

- $n=0$：三条公式都化为各自唯一的首项（$\cos\alpha$、$\cos\alpha$ 和 1）。
- 在第一个和中，$\beta\to 0$ 时 $\frac{\sin(n+1)\beta}{\sin\beta}\to n+1$，所以 $S\to(n+1)\cos\alpha$，这与 $n+1$ 个相等的项之和吻合。在第二个和中，$\beta=0$ 给出 $2^n\cos\alpha$，因为 $\sum\binom{n}{r}=2^n$。
- $\theta=\pi/4$，$n=2$：左方是 $1+\frac{1}{\sqrt{2}}\cdot\sqrt{2}+0=2$，右方是 $\sin\frac{3\pi}{4}\cdot 2/\sin\frac{\pi}{4}=2$。
- 数值上，随机选取 20 000 组 α、β、θ 及 $n\le 15$，三条公式与直接求和的结果都吻合，相对误差在 10^{-9} 以内。

常见错误

- 忘记第一个级数共有 $n+1$ 项（由 $r=0$ 至 n），结果得出 $\sin n\beta$ 而不是 $\sin(n+1)\beta$。
- 把 $1+\mathrm{e}^{2\mathrm{i}\beta}$ 写成 $2\cos\beta$，漏掉因子 $\mathrm{e}^{\mathrm{i}\beta}$，令答案中的角度出现偏移。
- 在最后一部分，把分子和分母分别取实部。你必须先把分母化为实数。

STEP 之外

把 cos 写成指数的实部，三角函数的和便化为等比或二项式的和。这正是傅里叶分析（Fourier analysis）的起点：第一个和本质上就是狄利克雷核（Dirichlet kernel）$\sin\big((n+\frac{1}{2})x\big)/\sin\frac{x}{2}$，它决定了傅里叶级数（Fourier series）如何收敛。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）以优美的几何观点，解释了为甚么乘以 $\mathrm{e}^{\mathrm{i}\theta}$ 就是旋转，以及为甚么这类和最终能收拢成简洁的公式。

详细解答

STEP III 1990，第 5 题

纯数学 · 莱布尼兹定理

> **STEP III 1990，第 5 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-q5

关键思路

帕斯卡法则（Pascal's rule）只不过是含阶乘（factorial）的分数运算，把两项通分即可。莱布尼兹公式（Leibniz's formula）给出乘积的 n 阶导数（derivative），其证明是对 n 用数学归纳法（induction）：把 n 阶的公式再求导一次，每个混合项都会出现两次，而帕斯卡法则正好把每一对合并起来。至于最后一部分，直接把 $y = \sin^{-1} x$ 求导 n 次是行不通的。我们改为找出 y 所满足的一条简单二阶方程，再对它应用莱布尼兹公式。

（试卷上印的是 $y = \sin^{-1} y$。这显然是 $y = \sin^{-1} x$ 的误印，我们采用后者。）

解答

帕斯卡法则

回想 $\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$，其中 $0! = 1$。当 $1 \le r \le n$ 时，$r-1$ 和 $n-r$ 均为非负数，所以下面每个阶乘都有意义。由于 $r! = r\,(r-1)!$ 及 $(n-r+1)! = (n-r+1)\,(n-r)!$，公分母是 $r!\,(n-r+1)!$：

$$\begin{aligned}\binom{n}{r} + \binom{n}{r-1} &= \frac{n!}{r!\,(n-r)!} + \frac{n!}{(r-1)!\,(n-r+1)!} \\ &= \frac{n!\,(n-r+1)}{r!\,(n-r+1)!} + \frac{n!\,r}{r!\,(n-r+1)!} \\ &= \frac{n!\,(n+1)}{r!\,(n+1-r)!} = \frac{(n+1)!}{r!\,(n+1-r)!} = \binom{n+1}{r}.\end{aligned}$$

莱布尼兹公式

我们假设 u 和 v 可按需要求导任意多次。用求和记号表示，要证明的是

$$(uv)^{(n)} = \sum_{r=0}^{n} \binom{n}{r} u^{(n-r)} v^{(r)},$$

其中 $z^{(0)}$ 表示 z 本身。我们对 $n \geq 1$ 用数学归纳法证明。

基础步骤（base case） $n = 1$。右边是 $u^{(1)}v + uv^{(1)}$，由乘法法则（product rule）可知它就是 $(uv)'$。

归纳步骤（inductive step）。假设公式对某个 $n \geq 1$ 成立。把它求导，对每一项使用乘法法则：

$$(uv)^{(n+1)} = \sum_{r=0}^{n} \binom{n}{r} u^{(n+1-r)} v^{(r)} + \sum_{r=0}^{n} \binom{n}{r} u^{(n-r)} v^{(r+1)}.$$

在第二个和式中设 $s = r+1$，它便成为 $\sum_{s=1}^{n+1} \binom{n}{s-1} u^{(n+1-s)} v^{(s)}$。现在两个和式都由形如 $u^{(n+1-r)}v^{(r)}$ 的项组成。把第一个和式的 $r = 0$ 项和第二个和式的 $s = n+1$ 项分拆出来：

$$\begin{aligned}(uv)^{(n+1)} &= u^{(n+1)}v + \sum_{r=1}^{n} \left[\binom{n}{r} + \binom{n}{r-1} \right] u^{(n+1-r)} v^{(r)} + uv^{(n+1)} \\ &= u^{(n+1)}v + \sum_{r=1}^{n} \binom{n+1}{r} u^{(n+1-r)} v^{(r)} + uv^{(n+1)},\end{aligned}$$

这里用了帕斯卡法则，因为 $1 \leq r \leq n$，所以法则适用。由于 $\binom{n+1}{0} = \binom{n+1}{n+1} = 1$，上式即

$$(uv)^{(n+1)} = \sum_{r=0}^{n+1} \binom{n+1}{r} u^{(n+1-r)} v^{(r)},$$

这正是 $n+1$ 时的公式。根据数学归纳法，公式对所有 $n \geq 1$ 成立。

$\sin^{-1} x$ 的导数

设 $y = \sin^{-1} x$，其中 $-1 < x < 1$。则

$$y^{(1)} = \frac{1}{\sqrt{1-x^2}}, \qquad \text{所以} \qquad (1-x^2)\left(y^{(1)}\right)^2 = 1.$$

把此式对 x 求导，

$$(1-x^2) \cdot 2y^{(1)}y^{(2)} - 2x\left(y^{(1)}\right)^2 = 0.$$

由于在 $(-1,1)$ 上 $y^{(1)} > 0$，我们可以除以 $2y^{(1)}$：

$$(1-x^2)\, y^{(2)} - x\, y^{(1)} = 0. \qquad (*)$$

这就是 $n = 0$ 时所要求的结果。

想一想 为甚么先把 $y^{(1)}$ 平方再求导，而不直接把 $(1-x^2)^{-1/2}$ 求导?

直接把 $(1-x^2)^{-1/2}$ 求导，得 $y^{(2)} = x(1-x^2)^{-3/2}$，再乘以 $1-x^2$ 同样得到 $(*)$。两种做法都可行。重点是要得出一条方程，其系数是 x 的多项式（polynomial），而且对 y 的各阶导数而言是线性（linear）的，因为这正是莱布尼兹公式处理得干净俐落的情况：多项式系数求导几次便变成零。

现在把 $(*)$ 求导 n 次，其中 $n \geq 1$，对每个乘积应用莱布尼兹公式。对第一个乘积，取 $u = y^{(2)}$ 和 $v = 1-x^2$。则 $v^{(1)} = -2x$，$v^{(2)} = -2$，而所有更高阶的导数均为零，所以只剩下三项：

$$\left((1-x^2)\,y^{(2)}\right)^{(n)} = (1-x^2)\,y^{(n+2)} + \binom{n}{1}(-2x)\,y^{(n+1)} + \binom{n}{2}(-2)\,y^{(n)}.$$

（当 $n=1$ 时最后一项不存在，这与 $\binom{1}{2} = 0$ 吻合。）对第二个乘积，取 $u = y^{(1)}$ 和 $v = x$，所以只剩下两项：

$$\left(x\,y^{(1)}\right)^{(n)} = x\,y^{(n+1)} + n\,y^{(n)}.$$

两式相减，$(*)$ 的 n 阶导数是

$$(1-x^2)\,y^{(n+2)} - 2nx\,y^{(n+1)} - n(n-1)\,y^{(n)} - x\,y^{(n+1)} - n\,y^{(n)} = 0.$$

合并同类项，$-2nx - x = -(2n+1)x$ 及 $-n(n-1) - n = -n^2$，所以

$$(1-x^2)\,y^{(n+2)} - (2n+1)x\,y^{(n+1)} - n^2 y^{(n)} = 0,$$

这正是所要求的结果。它对所有 $n \geq 0$ 及 $(-1,1)$ 内的所有 x 成立。

验算

- 在莱布尼兹公式中取 $n=1$：得到乘法法则。取 $n=2$：$(uv)'' = u''v + 2u'v' + uv''$，你可以把 $u'v + uv'$ 求导来核实。
- 在 $x=0$ 处，最后的结果给出 $y^{(n+2)}(0) = n^2 y^{(n)}(0)$。配合 $y^{(1)}(0) = 1$，可得 $y^{(3)}(0) = 1$ 和 $y^{(5)}(0) = 9$，所以麦克劳林级数（Maclaurin series）的开首是 $x + \frac{x^3}{6} + \frac{9x^5}{120} = x + \frac{x^3}{6} + \frac{3x^5}{40}$，这正是 $\sin^{-1} x$ 的已知级数。
- 我们作了数值验算：对所有 $n < 60$ 检验帕斯卡法则；用 300 对随机多项式精确检验莱布尼兹公式；并利用 $(1-x^2)^{-1/2}$ 的泰勒系数（Taylor coefficients），在 $(-1,1)$ 内的六个点上对 n 直至 27 检验最后的恒等式。相对残差小于 10^{-15}。

常见错误

- 在归纳步骤中把第二个和式重新编排指标后，弄错二项式系数（binomial coefficient）。应先把 $r = 0$ 和 $r = n + 1$ 这两个首尾项分开写出，然后才应用帕斯卡法则。
- 试图直接把 $\sin^{-1} x$ 求导 n 次，而不先找出方程 $(*)$。
- 把 $1 - x^2$ 求导时，漏掉因子 $\binom{n}{2}$ 或负号。

STEP 之外

先找出一条系数为多项式的线性微分方程（differential equation），再把它求导 n 次，这个技巧是求泰勒系数递推关系（recurrence）的标准方法。它直接通往常微分方程（ODE）的幂级数解法（power-series solutions），勒让德（Legendre）函数和贝塞尔（Bessel）函数正是由此而来。莱布尼兹公式也是二项式定理（binomial theorem）在微积分中的版本：若 $u = \mathrm{e}^{ax}$ 而 $v = \mathrm{e}^{bx}$，它便化为 $(a+b)^n = \sum \binom{n}{r} a^{n-r} b^r$。想进一步了解级数解法，可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）。

详细解答

STEP III 1990，第 6 题

纯数学 · 线性变换与椭圆

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

本题在 *Beyond STEP* 第 7.2 节只伸缩，不转向中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch07/eigenvectors.html

关键思路

那两个特殊的向量是矩阵的特征向量（eigenvector），它们道出了整个几何图像：T 沿一条直线按某个倍数拉伸，又沿与之垂直的另一条直线按另一个倍数拉伸，于是单位圆变成一个椭圆（ellipse），其两轴就落在这两条直线上。代数方面，要求 E 的方程，最简洁的做法是先求 T 的逆，把 x 和 y 用 X 和 Y 表示，再代入 $x^2+y^2=1$。面积按行列式（determinant）的倍数缩放；又因为 T 把直线映射为直线，所以 E 的切线（tangent）就是 C 的切线的像。

解答

设 $M=\frac{2}{5}\begin{pmatrix}9 & -2\\ -2 & 6\end{pmatrix}$，则 T 就是左乘 M。

两个特殊方向

计算可得

$$M\begin{pmatrix}1\\2\end{pmatrix} = \frac{2}{5}\begin{pmatrix}9-4\\-2+12\end{pmatrix} = \frac{2}{5}\begin{pmatrix}5\\10\end{pmatrix} = 2\begin{pmatrix}1\\2\end{pmatrix},$$
$$M\begin{pmatrix}2\\-1\end{pmatrix} = \frac{2}{5}\begin{pmatrix}18+2\\-4-6\end{pmatrix} = \frac{2}{5}\begin{pmatrix}20\\-10\end{pmatrix} = 4\begin{pmatrix}2\\-1\end{pmatrix}.$$

所以 $\begin{pmatrix}1\\2\end{pmatrix}$ 方向不变，并乘以 2；而 $\begin{pmatrix}2\\-1\end{pmatrix}$ 方向不变，并乘以 4，正如题目所求。留意这两个方向互相垂直，因为 $1\cdot 2 + 2\cdot(-1) = 0$。

E 的方程

M 的行列式为

$$\det M = \left(\tfrac{2}{5}\right)^2 (9\cdot 6 - (-2)(-2)) = \tfrac{4}{25}\cdot 50 = 8,$$

其值非零，故 T 可逆（invertible），且

$$M^{-1} = \frac{1}{8}\cdot\frac{2}{5}\begin{pmatrix}6&2\\2&9\end{pmatrix} = \frac{1}{20}\begin{pmatrix}6&2\\2&9\end{pmatrix}.$$

因此，点 (X,Y) 是 (x,y) 的像，当且仅当

$$x = \frac{3X+Y}{10}, \qquad y = \frac{2X+9Y}{20}.$$

点 (X,Y) 在 E 上，当且仅当它的原像在 C 上，即当且仅当 $x^2+y^2=1$。代入得

$$\begin{aligned}\frac{(3X+Y)^2}{100} + \frac{(2X+9Y)^2}{400} = 1 &\iff 4(9X^2+6XY+Y^2) + (4X^2+36XY+81Y^2) = 400\\ &\iff 40X^2+60XY+85Y^2 = 400\\ &\iff 8X^2+12XY+17Y^2 = 80,\end{aligned}$$

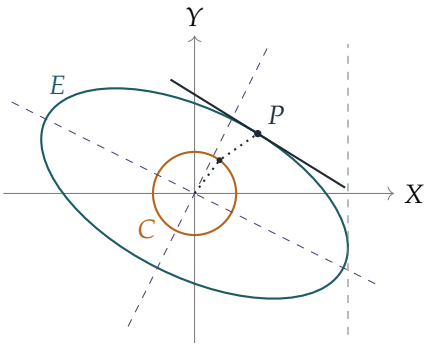
这就是所求的 E 的方程。

面积

线性映射（linear map）会把每一个面积乘以 $|\det M|$。C 所围成的圆盘面积为 π，所以 E 所围成的区域面积为

$$8\pi.$$

这与几何图像吻合：E 是一个椭圆，其半轴长为 2（沿 $\begin{pmatrix}1\\2\end{pmatrix}$ 方向）和 4（沿 $\begin{pmatrix}2\\-1\end{pmatrix}$ 方向），而 $\pi\cdot 2\cdot 4 = 8\pi$。



图中显示 C、椭圆 E 及其两轴（虚线）、与 E 相切的铅垂线 $X = 2\sqrt{17/5}$，以及在 P 的切线，该点是 $\frac{1}{5}(3,4)$ 的像。

X 的最大值

把 E 的方程看成 Y 的二次方程：

$$17Y^2 + 12XY + (8X^2 - 80) = 0.$$

某个 X 值会在 E 上出现，当且仅当此方程有实根 Y，即当且仅当其判别式（discriminant）非负：

$$\begin{aligned} 144X^2 - 4 \cdot 17(8X^2 - 80) \geq 0 &\iff 144X^2 - 544X^2 + 5440 \geq 0 \\ &\iff 400X^2 \leq 5440 \\ &\iff X^2 \leq \tfrac{68}{5}. \end{aligned}$$

所以 E 上每一点都满足 $X \leq \sqrt{68/5} = 2\sqrt{17/5}$，而且等号可以取到（此时判别式为零，得出唯一的一点，其 $Y = -\frac{6X}{17}$）。因此 X 在 E 上的最大值为 $2\sqrt{17/5}$，正如题目所求。

想一想 有没有办法不用判别式，直接从圆看出这个结果？

有。在 E 上有 $X = \frac{2}{5}(9x - 2y)$，其中 $x^2 + y^2 = 1$。设 $(x, y) = (\cos\theta, \sin\theta)$，则表达式 $9\cos\theta - 2\sin\theta$ 的最大值为 $\sqrt{81 + 4} = \sqrt{85}$（即 $R\cos(\theta + \alpha)$ 的形式）。所以最大的 X 是 $\frac{2}{5}\sqrt{85} = 2\sqrt{85/25} = 2\sqrt{17/5}$，与上面的结果一致。

切线

C 上的点 $\frac{1}{5}(3,4)$ 被映射到

$$P = \frac{2}{5}\begin{pmatrix} 9 \cdot \frac{3}{5} - 2 \cdot \frac{4}{5} \\ -2 \cdot \frac{3}{5} + 6 \cdot \frac{4}{5} \end{pmatrix} = \frac{2}{5}\begin{pmatrix} \frac{19}{5} \\ \frac{18}{5} \end{pmatrix} = \begin{pmatrix} \frac{38}{25} \\ \frac{36}{25} \end{pmatrix}.$$

C 在 $\frac{1}{5}(3,4)$ 的切线垂直于该处的半径，所以它的方程是

$$3x + 4y = 5.$$

由于 T 是线性而且可逆的，它把直线映射为直线，并且把一条与 C 恰好交于一点的直线，映射为一条与 E 恰好交于一点的直线。所以 E 在 P 的切线就是这条直线的像。代入 $x = \frac{3X+Y}{10}$ 和 $y = \frac{2X+9Y}{20}$：

$$\begin{aligned}\frac{3(3X+Y)}{10} + \frac{4(2X+9Y)}{20} = 5 &\iff 6(3X+Y) + 4(2X+9Y) = 100\\ &\iff 26X + 42Y = 100,\end{aligned}$$

所以 E 在 P 的切线是

$$13X + 21Y = 50.$$

验算一下：$13 \cdot \frac{38}{25} + 21 \cdot \frac{36}{25} = \frac{494+756}{25} = 50$，所以 P 在这条直线上。另一个做法是对 $8X^2 + 12XY + 17Y^2 = 80$ 作隐函数微分（implicit differentiation），得出在 P 的法线（normal）方向为 $(16X + 12Y,\ 12X + 34Y) = \frac{1}{25}(1040, 1680)$，它与 $(13,21)$ 平行，于是得出同一条直线。

验算

- $(1,0)$ 的像是 $\frac{2}{5}(9,-2)$，而 $\frac{4}{25}\left(8 \cdot 81 - 12 \cdot 18 + 17 \cdot 4\right) = \frac{4}{25} \cdot 500 = 80$，所以它在 E 上。
- 两轴的端点 $\pm 2 \cdot \frac{1}{\sqrt{5}}(1,2)$ 和 $\pm 4 \cdot \frac{1}{\sqrt{5}}(2,-1)$ 都满足 E 的方程；例如 $\frac{4}{5}(8 + 24 + 68) = 80$。
- 数值上，在 C 上取样 200 000 个点，得出最大值 $X \approx 3.687818$，与 $2\sqrt{17/5} \approx 3.687818$ 相符；而用鞋带公式（shoelace formula）计算像多边形的面积，得 $25.13274 \approx 8\pi$。

常见错误

- 代入 $x^2 + y^2 = 1$ 时用了 M 而不是 M^{-1}：要求像曲线的方程，需要把旧坐标用新坐标表示。
- 计算行列式时忘记了因子 $\frac{2}{5}$：它贡献的是 $\left(\frac{2}{5}\right)^2$，而不是 $\frac{2}{5}$。
- 求出了像点 P，却用 C 在 $\frac{1}{5}(3,4)$ 的斜率来写 E 的切线：线性映射保持相切关系，但不保持斜率。

STEP 之外

这里的矩阵是对称（symmetric）的，这正是它的特征向量互相垂直的原因：这是谱定理（spectral theorem）的二维情形，该定理指出每个实对称矩阵都可以用一个旋转来对角化（diagonalise）。它解释了为甚么任何二次型（quadratic form）$ax^2 + 2hxy + by^2$ 所描述的圆锥曲线（conic）都有互相垂直的主轴，而该二次型的矩阵的特征值（eigenvalue）可以让你立刻知道它的形状。Strang 的 *Introduction to Linear Algebra*（线性代数入门）以大量几何图像阐述特征值、对称矩阵和二次型。

详细解答

STEP III 1990，第 7 题

纯数学 · 隐函数曲线描绘

STEP III 1990，第 7 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-q7

关键思路

隐函数微分法（implicit differentiation）做两次，便得到每一点的 $\mathrm{d}y/\mathrm{d}x$ 和 $\mathrm{d}^2y/\mathrm{d}x^2$，而这些正是泰勒级数（Taylor series）的系数。至于不等式，窍门在于留意到这条方程只是 y 的二次方程：点 (x,y) 要存在，这个二次方程必须有实根，所以它的判别式（discriminant）非负。把二次方程直接解出来，便可看清整条曲线：两条分支（branch），在判别式等于零之处相接。

解答

以下一律记 $F(x,y)=xy^2+x^3+a^2y-a^3$。代入可知在 P、Q 和 R 均有 $F=0$；例如在 R 得 $a\cdot a^2+a^3-a^3-a^3=0$。

在 P、Q 和 R 的泰勒级数

把 y 视为 x 的函数，将 $F(x,y)=0$ 对 x 求导，

$$y^2+2xyy'+3x^2+a^2y'=0,\qquad \text{所以}\qquad y'=-\frac{y^2+3x^2}{2xy+a^2}.$$

把第一条方程再求导一次（$2xyy'$ 一项给出 $2yy'+2xy'^2+2xyy''$）：

$$4yy'+2xy'^2+6x+(2xy+a^2)\,y''=0.$$

在每一点，分母 $2xy+a^2$ 都不等于零，所以这些方程确定了 y' 和 y''，继而有 $y=y_0+y'h+\frac{1}{2}y''h^2+\cdots$。

在 $P(0,a)$： $y' = -a^2/a^2 = -1$。于是 $4a(-1) + a^2 y'' = 0$，所以 $y'' = 4/a$，而

$$y = a - h + \frac{2}{a}h^2 + \cdots.$$

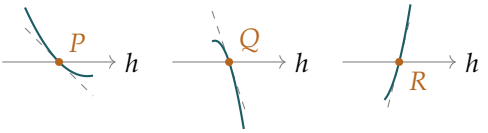
在 $Q(a,0)$： $y' = -3a^2/a^2 = -3$。于是 $0 + 2a(9) + 6a + a^2 y'' = 0$，所以 $y'' = -24/a$，而

$$y = -3h - \frac{12}{a}h^2 + \cdots.$$

在 $R(a,-a)$：这里 $2xy + a^2 = -a^2$，所以 $y' = -4a^2/(-a^2) = 4$。于是 $4(-a)(4) + 2a(16) + 6a - a^2 y'' = 0$，即 $22a = a^2 y''$，所以 $y'' = 22/a$，而

$$y = -a + 4h + \frac{11}{a}h^2 + \cdots.$$

因此在 P 附近，曲线以斜率（gradient）-1 下降，并且向上凹（concave up，即弯向切线上方）；在 Q 附近，曲线以斜率 -3 急降，并且向下凹（concave down）；在 R 附近，曲线以斜率 4 上升，并且向上凹。



（取 $a = 1$ 时的局部形状；虚线为切线。）

不等式

把 C 的方程写成 y 的二次方程：

$$x\,y^2 + a^2\,y + (x^3 - a^3) = 0.$$

若 $x \neq 0$ 而 (x,y) 在 C 上，这个二次方程便有实根 y，所以它的判别式非负：

$$a^4 - 4x(x^3 - a^3) \geq 0 \quad \Longleftrightarrow \quad 4x^4 - 4a^3 x - a^4 \leq 0.$$

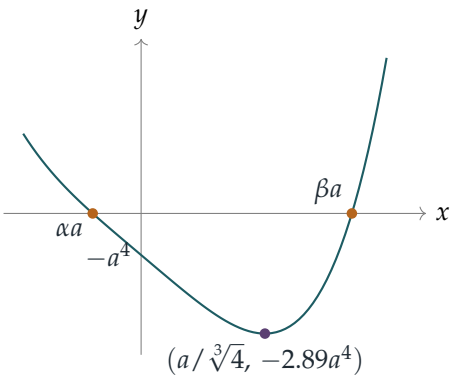
若 $x = 0$，左边等于 $-a^4 < 0$，所以不等式在该处同样成立。结果得证。

$y = 4x^4 - 4a^3 x - a^4$ 的图像

试题原文印的是 $4x^4 - 4a^3 - a^4$，这肯定是误植，应为我们刚刚遇到的四次多项式（quartic）$g(x) = 4x^4 - 4a^3 x - a^4$。我们有 $g(0) = -a^4$ 及 $g'(x) = 16x^3 - 4a^3$，后者只在 $x_0 = a/\sqrt[3]{4} \approx 0.63a$ 处等于零，g 在该处取得极小值。利用 $x_0^3 = a^3/4$，

$$g(x_0) = a^3 x_0 - 4a^3 x_0 - a^4 = -a^4\left(1 + 3 \cdot 4^{-1/3}\right) \approx -2.89a^4.$$

由于当 $x \to \pm\infty$ 时 $g \to +\infty$，而 g 先递减后递增，所以它恰好有两个零点 $x = \alpha a$ 和 $x = \beta a$，其中 $\alpha \approx -0.246$ 和 $\beta \approx 1.072$ 是 $4t^4 - 4t - 1 = 0$ 的根（它们没有简洁的闭合形式，我们也用不着）。此外 $g''(x) = 48x^2 \geq 0$，所以图像是凸（convex）的。



（以 $a = 1$ 绘画。）所以 C 上每一点都满足 $\alpha a \le x \le \beta a$：曲线位于一条垂直的带状区域之内。

曲线 C

当 $x \neq 0$ 时，解二次方程得出两条分支

$$y_\pm = \frac{-a^2 \pm \sqrt{D}}{2x}, \qquad D = a^4 + 4a^3x - 4x^4,$$

它们恰好在该带状区域上取实数值，并在 $D = 0$ 之处相接，即 $x = \alpha a$ 和 $x = \beta a$。在那里 $y = -a^2/(2x)$，而切线是垂直的，因为那正是 y' 的分母 $2xy + a^2$ 等于零的地方。由此得到 $(\beta a,\ -0.47a)$ 和 $(\alpha a,\ 2.03a)$ 两点。

> **想一想** 当 $x \to 0$ 时，两条分支会怎样？

把分子和分母同乘以 $\sqrt{D} + a^2$，并利用 $D - a^4 = 4x(a^3 - x^3)$，得

$$y_+ = \frac{2(a^3 - x^3)}{a^2 + \sqrt{D}},$$

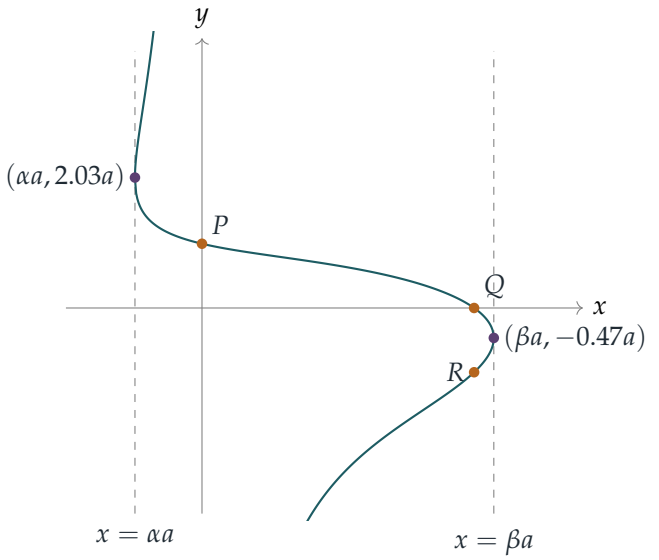
它在 $x = 0$ 处完全没有问题，并且在该处等于 a。所以 y_+ 是一条通过 P 的光滑分支。另一条分支在 x 很小时满足 $y_- \approx -a^2/x$，所以当 $x \to 0^+$ 时 $y_- \to -\infty$，当 $x \to 0^-$ 时 $y_- \to +\infty$。这就是沿 y 轴的渐近线（asymptote）。

再多几项事实，有助确定草图的形状。

- 曲线没有水平切线：$y' = 0$ 需要 $y^2 + 3x^2 = 0$，即 $(0,0)$，而这一点不在 C 上。
- 曲线只在 $x^3 = a^3$ 之处与 $y = 0$ 相交，即 Q；并且只在 P 与 $x = 0$ 相交。
- 在 $x = a$ 处，$D = a^4$，所以 $y_+ = 0$（点 Q）而 $y_- = -a$（点 R）。

综合以上各点，C 是一条连通的曲线。它在 y 轴左边不远处沿 y_- 从 $+\infty$ 下降，到达 $(\alpha a, 2.03a)$ 处的垂直切线。然后它沿 y_+ 折回，经过 P（斜率 -1），再下降经过 Q（斜率

-3），到达 $(\beta a, -0.47a)$ 处的垂直切线，接着沿 y_- 折回，经过 R（斜率 4），在 y 轴右边不远处下降至 $-\infty$。曲线在 P 向上凹，在 Q 向下凹，所以两点之间有一个拐点（point of inflection）。



（以 $a = 1$ 绘画；虚线是上文求得的带状区域的边界。）

验算

- 取 $a = 1.7$，我们在每一点附近的 $x = x_0 + h$ 处解出二次方程的 y。每当 h 减半，各个两项泰勒级数的误差便缩小约 8 倍，这正是我们预期的 h^3 表现。
- 随机取 10^5 个 x 值，二次方程有实根的情况，恰好就是 $4x^4 - 4a^3x - a^4 \leq 0$ 的情况。
- 用二分法（bisection）得 $\alpha \approx -0.2463$ 和 $\beta \approx 1.0724$。在 $x = \pm 0.001$ 处（取 $a = 1.7$），一个根接近 a，另一个根接近 $-a^2/x \approx \mp 2890$。

常见错误

- 漏掉因子 $\frac{1}{2}$：h^2 的系数是 $\frac{1}{2}y''$，不是 y''。
- 为了使用判别式而除以 x，却没有另外检查 $x = 0$ 的情况。
- 没有察觉 y_+ 光滑地通过 P。写成 $(-a^2 + \sqrt{D})/(2x)$ 时，它在 $x = 0$ 处看似奇异，其实不然。

STEP 之外

这一题中，我们利用隐式定义的函数的泰勒展开式，找出曲线在某一点附近的表现。保证在 $\partial F/\partial y \neq 0$ 之处 y 是 x 的光滑函数的定理，就是**隐函数定理**（implicit function theorem），它是多元分析（multivariable analysis）的一个重要结果。带状区域两端的垂直切线，正好出现在 $\partial F/\partial y = 2xy + a^2$ 等于零的点。这条曲线是平面上的三次曲线（cubic），而研究这类曲线会引领我们走向代数几何（algebraic geometry）和椭圆曲线（elliptic curve）。想做更多这种风格的题目，可试试 Siklos 的 *Advanced Problems in Mathematics*（进阶数学问题集）。

详细解答

STEP III 1990，第 8 题

纯数学 · 恰当微分方程

> **STEP III 1990，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/90-s3-q8

关键思路

展开 $\dfrac{\mathrm{d}}{\mathrm{d}x}(py'+qy)$，再与 $Py''+Qy'+Ry$ 比较：这样便完全确定了 p 和 q，而条件 $P''-Q'+R=0$ 正好令最后一个系数吻合。至于题中的微分方程（differential equation），这个条件成立，所以积分一次就能把二阶方程直接化为一阶线性方程，再用积分因子（integrating factor）便可完成。

解答

恰当性的判别准则

对任意函数 p、q 和 y，由乘积法则（product rule）得

$$\frac{\mathrm{d}}{\mathrm{d}x}(py'+qy)=py''+(p'+q)y'+q'y. \qquad (*)$$

仅当。假设 p 和 q（两者均与 y 无关）使 $Py''+Qy'+Ry$ 对每一个 y 都等于 $(*)$。取 $y=1$，得 $R=q'$。取 $y=x$，得 $Q+Rx=(p'+q)+q'x$，所以 $Q=p'+q$。取 $y=\frac{1}{2}x^2$，得 $P+Qx+\frac{1}{2}Rx^2=p+(p'+q)x+\frac{1}{2}q'x^2$，所以 $P=p$。因此

$$P''-Q'+R=p''-(p''+q')+q'=0.$$

若。假设 $P''-Q'+R=0$。取

$$p=P,\qquad q=Q-P'.$$

那么 $p'+q=P'+Q-P'=Q$，而利用该条件可得 $q'=Q'-P''=R$。由 $(*)$，对每一个 y 都有 $\dfrac{\mathrm{d}}{\mathrm{d}x}(py'+qy)=Py''+Qy'+Ry$，这正是所要证明的。

以上的推导同时说明 p 和 q 必定是 P 和 $Q-P'$，下面我们就用这个方法把它们求出来。

微分方程

这里 $P=x-x^4$、$Q=1-7x^3$ 及 $R=-9x^2$，所以

$$P''-Q'+R=-12x^2-(-21x^2)-9x^2=0.$$

判别准则适用，其中 $p=P=x-x^4$ 及 $q=Q-P'=(1-7x^3)-(1-4x^3)=-3x^3$。方程变成

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(x-x^4)y'-3x^3y\right]=(x^3+3x)\mathrm{e}^x.$$

反复运用分部积分法（integration by parts），得 $\int x^3\mathrm{e}^x\,\mathrm{d}x=(x^3-3x^2+6x-6)\mathrm{e}^x$ 及 $\int 3x\mathrm{e}^x\,\mathrm{d}x=(3x-3)\mathrm{e}^x$，所以

$$(x-x^4)y'-3x^3y=(x^3-3x^2+9x-9)\mathrm{e}^x+C.$$

当 $x=2$ 时，$y=2\mathrm{e}^2$ 及 $y'=0$，所以左边是 $-3\cdot 8\cdot 2\mathrm{e}^2=-48\mathrm{e}^2$，右边是 $5\mathrm{e}^2+C$。因此 $C=-53\mathrm{e}^2$。

> **想一想** 左边的 $(x-x^4)y'-3x^3y$ 并不是 $(x-x^4)y$ 的导数。那么第二次积分该怎样做？

抽出因子 x：左边是 $x\left[(1-x^3)y'-3x^2y\right]=x\dfrac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^3)y\right]$。（这就是常用的积分因子：$y'-\frac{3x^2}{1-x^3}y$ 的积分因子是 $\exp\int\frac{-3x^2}{1-x^3}\,\mathrm{d}x=1-x^3$。）由于解位于 $x=2$ 附近，可以除以 x，于是

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(1-x^3)y\right]=(x^2-3x+9)\mathrm{e}^x-\frac{9\mathrm{e}^x}{x}-\frac{53\mathrm{e}^2}{x}.$$

再用分部积分法，得 $\int(x^2-3x+9)\mathrm{e}^x\,\mathrm{d}x=(x^2-5x+14)\mathrm{e}^x$。$\mathrm{e}^x/x$ 这一项没有初等原函数（elementary antiderivative），所以我们把它保留为由起点 $x=2$ 开始的积分：

$$(1-x^3)y=(x^2-5x+14)\mathrm{e}^x-53\mathrm{e}^2\ln\frac{x}{2}-9\int_2^x\frac{\mathrm{e}^t}{t}\,\mathrm{d}t+D.$$

当 $x=2$ 时，左边是 $-7\cdot 2\mathrm{e}^2=-14\mathrm{e}^2$，右边是 $8\mathrm{e}^2+D$，所以 $D=-22\mathrm{e}^2$。因此

$$\boxed{y=\frac{1}{1-x^3}\left[(x^2-5x+14)\mathrm{e}^x-22\mathrm{e}^2-53\mathrm{e}^2\ln\frac{x}{2}-9\int_2^x\frac{\mathrm{e}^t}{t}\,\mathrm{d}t\right]}$$

适用于 $x>1$。y'' 的系数 $x-x^4$ 在 $x=1$ 处为零，所以经过 $x=2$ 的解定义在 $x>1$ 上，而不一定能延伸至 $x=1$ 以外。

验算

- 当 $x = 2$ 时，方括号内的值是 $8e^2 - 22e^2 = -14e^2$，除以 $1 - 8 = -7$ 得 $y = 2e^2$，符合要求。
- 由 $x = 2$ 开始以数值方法逐步求解原方程（四阶龙格－库塔法，Runge–Kutta），得 $y(3) \approx 7.90306$，与公式的结果吻合，其中的积分以辛普森法则（Simpson's rule）计算。两者在 $x = 1.2$、1.5 和 4 处亦一致。
- 展开 $\frac{\mathrm{d}}{\mathrm{d}x}\left[(x - x^4)y' - 3x^3y\right]$，可还原为 $(x - x^4)y'' + (1 - 7x^3)y' - 9x^2y$。

常见错误

- 第一次积分后漏掉常数 C。它在这里并不是零，而且对数项正是由它产生的。
- 硬要把 $\int e^x/x\,\mathrm{d}x$ 写成闭合形式（closed form）。这是做不到的，所以应把它保留为下限是 2 的定积分，这样代入初始条件也方便得多。

STEP 之外

若表达式 $Py'' + Qy' + Ry$ 是某个函数的全导数，便称它为恰当（exact）的。表达式 $P'' - Q' + R$ 就是伴随（adjoint）方程 $(Pz)'' - (Qz)' + Rz = 0$ 的左边在 $z = 1$ 时的值。一般而言，伴随方程的任何一个解 z 都是积分因子，能使 $z(Py'' + Qy' + Ry)$ 成为恰当的。由伴随方程再往前走，便是自伴随（self-adjoint）问题，即施图姆－刘维尔（Sturm–Liouville）问题，以及格林函数（Green's function）。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（一本初等微分方程教科书）在二阶线性方程的习题中，涵盖了恰当方程和伴随方程。

详细解答

STEP III 1990，第 9 题

纯数学 · 双曲函数

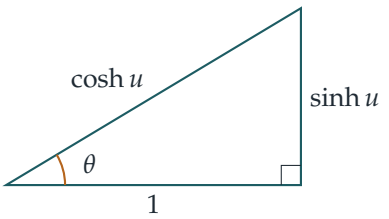
STEP III 1990，第 9 题

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-q9

关键思路

关系式 $\tan\theta = \sinh u$ 等同于 $\sec^2\theta = 1 + \sinh^2 u = \cosh^2 u$，因此两个变量的三角函数（trigonometric functions）与双曲函数（hyperbolic functions）完全对应：$\cos\theta = \operatorname{sech} u$、$\sin\theta = \tanh u$、$\sec\theta = \cosh u$。画一个边长为 1、$\sinh u$ 和 $\cosh u$ 的直角三角形，这些关系便一目了然。有了这本“字典”，每一部分都只需先做一次简短的求导，再把结果翻译回来。



解答

在整个解答中，由于 $0 \le \theta < \frac{1}{2}\pi$，我们有 $\tan\theta \ge 0$，所以 $\sinh u \ge 0$ 且 $u \ge 0$。此外 $\cos\theta > 0$ 且 $\cosh u > 0$。

第 (i) 部分

利用 $1 + \tan^2\theta = \sec^2\theta$ 和 $1 + \sinh^2 u = \cosh^2 u$，

$$\sec^2\theta = 1 + \tan^2\theta = 1 + \sinh^2 u = \cosh^2 u.$$

$\sec\theta$ 和 $\cosh u$ 均为正数，所以 $\sec\theta = \cosh u$，即 $\cos\theta = \operatorname{sech} u = v$，这正是所要证明的结果。

值得顺带记下与之配对的结果：$\sin\theta = \tan\theta\cos\theta = \sinh u \operatorname{sech} u = \tanh u$。

第 (ii) 部分

把 $\tan\theta = \sinh u$ 对 u 求导（利用链式法则（chain rule）作隐函数求导（implicit differentiation)），

$$\sec^2\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}u} = \cosh u \quad\Longrightarrow\quad \frac{\mathrm{d}\theta}{\mathrm{d}u} = \cosh u\cos^2\theta = \cosh u\operatorname{sech}^2 u = \operatorname{sech} u = v,$$

其中 $\cos\theta$ 的表达式来自第 (i) 部分。

第 (iii) 部分

由于 $\frac{\mathrm{d}}{\mathrm{d}u}\cosh u = \sinh u$，

$$\frac{\mathrm{d}v}{\mathrm{d}u} = \frac{\mathrm{d}}{\mathrm{d}u}\big(\cosh u\big)^{-1} = -\frac{\sinh u}{\cosh^2 u} = -\operatorname{sech} u\tanh u.$$

查一查字典，$\operatorname{sech} u\tanh u = \cos\theta\sin\theta = \frac{1}{2}\sin 2\theta$，所以

$$\sin 2\theta = -2\,\frac{\mathrm{d}v}{\mathrm{d}u}.$$

至于二阶导数（second derivative），可运用积法则（product rule），并配合 $\frac{\mathrm{d}}{\mathrm{d}u}\tanh u = \operatorname{sech}^2 u$：

$$\begin{aligned}\frac{\mathrm{d}^2v}{\mathrm{d}u^2} &= -\Big[(-\operatorname{sech} u\tanh u)\tanh u + \operatorname{sech} u\cdot\operatorname{sech}^2 u\Big]\\ &= \operatorname{sech} u\big(\tanh^2 u - \operatorname{sech}^2 u\big).\end{aligned}$$

乘以 $-\cosh u$ 便可消去 $\operatorname{sech} u$：

$$-\cosh u\,\frac{\mathrm{d}^2v}{\mathrm{d}u^2} = \operatorname{sech}^2 u - \tanh^2 u = \cos^2\theta - \sin^2\theta = \cos 2\theta,$$

这就是第二个结果。

第 (iv) 部分

现在把 u 和 v 看成 θ 的函数。由第 (i) 部分，$v = \cos\theta$ 且 $\cosh u = \sec\theta$；而第 (ii) 部分给出 $\frac{\mathrm{d}u}{\mathrm{d}\theta} = 1\big/\frac{\mathrm{d}\theta}{\mathrm{d}u} = 1/v$（因为 $v > 0$，此式成立）。所以

$$\frac{\mathrm{d}u}{\mathrm{d}\theta} = \sec\theta, \qquad \frac{\mathrm{d}^2u}{\mathrm{d}\theta^2} = \sec\theta\tan\theta,$$
$$\frac{\mathrm{d}v}{\mathrm{d}\theta} = -\sin\theta, \qquad \frac{\mathrm{d}^2v}{\mathrm{d}\theta^2} = -\cos\theta.$$

想一想 在埋头苦算之前，先想一想：第 (iv) 部分的恒等式，照试卷上所印的样子看来完全正确吗？

并不正确。按试卷所印，第一项含有 $\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}$，于是左方为

$$\begin{aligned}\sec\theta\cdot\sec\theta\tan\theta-\sin\theta\sec\theta\tan\theta+\sec^2\theta &= \sec^2\theta\tan\theta+(\sec^2\theta-\tan^2\theta)\\ &= \sec^2\theta\tan\theta+1\ \geq 1,\end{aligned}$$

它永远不会等于零。题目原意的恒等式（identity），第一项应为 $\frac{\mathrm{d}^2v}{\mathrm{d}\theta^2}$（把 v 误印成 u）：

$$\frac{\mathrm{d}u}{\mathrm{d}\theta}\frac{\mathrm{d}^2v}{\mathrm{d}\theta^2}+\frac{\mathrm{d}v}{\mathrm{d}\theta}\frac{\mathrm{d}^2u}{\mathrm{d}\theta^2}+\left(\frac{\mathrm{d}u}{\mathrm{d}\theta}\right)^2=0.$$

代入上述四个导数，

$$\sec\theta(-\cos\theta)+(-\sin\theta)\sec\theta\tan\theta+\sec^2\theta=-1-\tan^2\theta+\sec^2\theta=0,$$

因为 $\sec^2\theta=1+\tan^2\theta$。这就证明了修正后的结果。

还有一个更简洁的看法：根据积法则，首两项合起来就是 $\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\mathrm{d}u}{\mathrm{d}\theta}\frac{\mathrm{d}v}{\mathrm{d}\theta}\right)$，而 $\frac{\mathrm{d}u}{\mathrm{d}\theta}\frac{\mathrm{d}v}{\mathrm{d}\theta}=-\sec\theta\sin\theta=-\tan\theta$，其导数为 $-\sec^2\theta=-\left(\frac{\mathrm{d}u}{\mathrm{d}\theta}\right)^2$。

验算

取 $u=\ln 2$，则 $\sinh u=\frac{3}{4}$、$\cosh u=\frac{5}{4}$ 且 $\tan\theta=\frac{3}{4}$：由 3–4–5 三角形可得 $\cos\theta=\frac{4}{5}=\operatorname{sech}u$ 及 $\sin 2\theta=\frac{24}{25}$，而 $-2\,\mathrm{d}v/\mathrm{d}u=2\cdot\frac{4}{5}\cdot\frac{3}{5}=\frac{24}{25}$。数值方面，在 2000 个随机点上以有限差分（finite difference）计算的导数，与 (i)–(iii) 及修正后的 (iv) 吻合，误差约为 10^{-7}，并确认试卷所印的 (iv) 左方等于 $1+\sec^2\theta\tan\theta$。

常见错误

- 写出 $\cos\theta=\pm\operatorname{sech}u$ 却没有说明为何应取正号：决定正负号的是范围 $0\leq\theta<\frac{1}{2}\pi$。
- 忘记 $\frac{\mathrm{d}u}{\mathrm{d}\theta}$ 只不过是 $\frac{\mathrm{d}\theta}{\mathrm{d}u}$ 的倒数，反而从头对 $\operatorname{arsinh}(\tan\theta)$ 求导。

STEP 之外

函数 $\theta(u)=\arctan(\sinh u)$ 称为*古德曼函数*（Gudermannian function），它毋须借助复数（complex numbers）便把圆函数与双曲函数联系起来；其反函数（inverse function）$u=\ln(\sec\theta+\tan\theta)=\int_0^\theta\sec t\,\mathrm{d}t$ 正好是麦卡托（Mercator）地图上的纵坐标。当你写下 $\cosh(\mathrm{i}x)=\cos x$ 时，同样的配对也会出现；Needham 的 *Visual Complex Analysis*（《可视化复分析》）是从几何角度欣赏这些联系的绝佳读物。

详细解答

STEP III 1990，第 10 题

纯数学 · 根与渐近近似

STEP III 1990，第 10 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-q10

关键思路

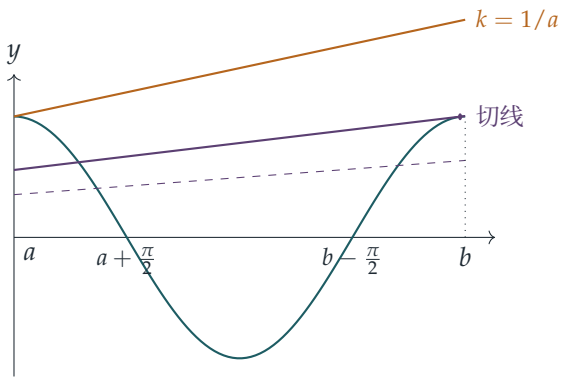
在题目所给的区间内，$\sin x$ 刚好走完一个完整周期，起点和终点都在峰顶，而直线 $y = kx$ 则恒为正，并缓缓上升。根只可能出现在 $\sin x > 0$ 的地方，而这样的地方分为两段：一段是 $\sin x$ 下降的部分（最多相交一次），另一段是 $\sin x$ 上升而且是凹函数（concave）的部分（最多相交两次）。值得留意的临界情况，是直线刚好与第二个拱形相切之时；把切点写成在区间右端点左方一小段距离 δ 处，相切条件便化为题目所给的方程。

解答

以下一律记

$$a = (4n+1)\frac{\pi}{2}, \qquad b = (4n+5)\frac{\pi}{2} = a + 2\pi,$$

于是区间就是 $a < x < b$。由于 $\sin a = \sin b = 1$，$\sin x$ 的图像由峰顶出发，在 $x = a + \frac{\pi}{2}$ 处降至 0，其后一直为负，直至 $x = a + \frac{3\pi}{2} = b - \frac{\pi}{2}$，然后再回升至 b 处的峰顶。



上图（按 $n = 1$ 的比例绘画，原点远在左方图外）以蓝绿色显示 $\sin x$，以琥珀色显示通过 $(a, 1)$ 的直线，以紫红色实线显示切线（tangent），另以虚线显示一条斜率较小、与曲线相交三次的直线。

根的可能数目

由于 $k > 0$ 且 $x > 0$, 故 $kx > 0$, 所以每个根都位于 $\sin x > 0$ 之处：不是在 $I_1 = (a,\, a + \frac{\pi}{2})$ 内，就是在 $I_2 = (b - \frac{\pi}{2},\, b)$ 内。

*在 I_1 上。*这里 $\sin x$ 严格递减（strictly decreasing），而 kx 严格递增（strictly increasing），所以 $kx - \sin x$ 严格递增，最多只有一个根。在右端点处 $kx - \sin x > 0$；在 $x = a$ 处其值为 $ka - 1$。因此，若 $ka < 1$，则 I_1 内恰有一个根；若 $ka \geq 1$，则没有根（此时在整个区间上 $kx - \sin x > 0$）。

*在 I_2 上。*设 $f(x) = \sin x - kx$。这里 $f''(x) = -\sin x < 0$，所以 f 是凹函数，在 I_2 上最多有两个零点。此外，f 在左端点处 $f < 0$，而 $f(b) = 1 - kb$。

- 若 $kb < 1$，则 f 由负变正，而凹函数在上升途中只能变号一次，所以恰有一个根。当 $kb = 1$ 时结论相同：此时 $f(b) = 0$ 且 $f'(b) = \cos b - k = -k < 0$，所以在 b 的左侧附近 $f > 0$，于是 I_2 内有一个根（b 本身这个零点不计在内）。
- 若 $kb > 1$，则 f 在两端均有 $f < 0$，所以根的数目是 0、1 或 2，分别对应 f 在 I_2 上的最大值为负、为零或为正。最大值为 0 即相切的情况：直线与拱形在某点 $t \in I_2$ 相切，设此时的斜率为 k_t。当 k 由很大的值逐渐减小时，直线随之下降，所以当 $k > k_t$ 时没有根，当 $k = k_t$ 时有一个根，而当 $\frac{1}{b} < k < k_t$ 时有两个根。

在切点处 $k_t t = \sin t < 1$ 且 $t > a$，所以 $k_t < 1/a$；又因为相切的情况满足 $kb > 1$，故 $k_t > 1/b$。把两段的结果合起来：

k 的范围	I_1 内的根数	I_2 内的根数	总数
$k \leq 1/b$	1	1	2
$1/b < k < k_t$	1	2	3
$k = k_t$	1	1	2
$k_t < k < 1/a$	1	0	1
$k \geq 1/a$	0	0	0

所以方程在该区间内可以有 0、1、2 或 3 个根，正如题目所求。

恰有一个根

由上表可见，恰有一个根的充要条件是 $k_t < k < \dfrac{1}{a} = \dfrac{2}{(4n+1)\pi}$。余下的工作是描述 k_t。

想一想 怎样把“直线与曲线相切”写成方程?

直线 $y = kx$ 与 $y = \sin x$ 在 $x = t$ 处相切，即两个图像在该处相交，而且斜率相等：

$$\sin t = k_t t, \qquad \cos t = k_t.$$

点 t 位于 I_2 内，所以可写 $t = b - \delta$，其中 $0 < \delta < \frac{\pi}{2}$。由于 $\sin b = 1$ 及 $\cos b = 0$，

$$\sin t = \sin(b - \delta) = \cos\delta, \qquad \cos t = \cos(b - \delta) = \sin\delta.$$

所以由斜率条件得 $k_t = \sin\delta$，而相交条件则变成

$$\cos\delta = \left((4n+5)\frac{\pi}{2} - \delta\right)\sin\delta.$$

因此，恰有一个根的条件是

$$\sin\delta < k < \frac{2}{(4n+1)\pi},$$

其中 δ 如上所述，这正是我们要证明的结果。

n 很大时的近似值

把关于 δ 的方程除以 $\sin\delta$（由于 $0 < \delta < \frac{\pi}{2}$，它不等于零）：

$$\cot\delta = b - \delta, \qquad \text{即} \qquad \tan\delta = \frac{1}{b-\delta}.$$

由于 $\delta < \frac{\pi}{2}$，右边小于 $1/(b - \frac{\pi}{2})$，而当 n 很大时这个值很小。所以 δ 很小，$\tan\delta \approx \delta$，且 $b - \delta \approx b$，由此得

$$\delta \approx \frac{1}{b} = \frac{2}{(4n+5)\pi}.$$

要求下一级的近似值，可写 $\delta = \frac{1}{b} + \frac{c}{b^3}$，并在两边保留至 b^{-3} 的项。左边是 $\tan\delta = \delta + \frac{1}{3}\delta^3 + \cdots = \frac{1}{b} + \frac{c + \frac{1}{3}}{b^3} + \cdots$。右边是

$$\frac{1}{b-\delta} = \frac{1}{b}\left(1 - \frac{\delta}{b}\right)^{-1} = \frac{1}{b}\left(1 + \frac{\delta}{b} + \cdots\right) = \frac{1}{b} + \frac{1}{b^3} + \cdots,$$

其中在修正项里用了 $\delta \approx 1/b$。比较 b^{-3} 的项，得 $c + \frac{1}{3} = 1$，所以 $c = \frac{2}{3}$，于是

$$\delta \approx \frac{1}{b} + \frac{2}{3b^3} = \frac{2}{(4n+5)\pi} + \frac{16}{3(4n+5)^3\pi^3}.$$

（等价的做法是：由 $\delta = 1/b$ 出发，把迭代（iteration）$\delta \mapsto \arctan\frac{1}{b-\delta}$ 进行一步，得到 $\arctan\frac{b}{b^2-1}$，它的展开式与上式相同。）由此可知，k 的区间的下端是 $\sin\delta \approx \frac{1}{b} + \frac{1}{2b^3}$。

验算

- 当 $n=1$ 时，以数值方法解 $\cos\delta=(\frac{9\pi}{2}-\delta)\sin\delta$ 得 $\delta=0.070973$，而 $1/b=0.070736$ 及 $1/b+2/(3b^3)=0.070971$：改良后的近似值准确了约 150 倍。当 $n=10$ 时，误差降至约 5×10^{-10}。
- 取 $n=1,2,3,5$，并从表中每一行各取一个 k 值，在细密的格点上数算 $kx-\sin x$ 的变号次数，所得的根数 $2,3,1,0$ 与预测完全吻合；而在 $\sin\delta$ 与 $2/((4n+1)\pi)$ 之间随机选取 k，则每次都得到一个根。

常见错误

- 忘记了相切的情况，或者把 $1/b$、k_t 和 $1/a$ 的大小次序弄错，以致根的数目数错。
- 在第二级近似中只用 $\tan\delta\approx\delta$，结果得到 $\delta\approx b/(b^2-1)\approx1/b+1/b^3$。$\tan\delta$ 中的 $\frac{1}{3}\delta^3$ 项与修正项属同一阶，必须保留。

STEP 之外

当方程含有一个很小或很大的参数时，逐步求出其根越来越精确的近似值，这就是*渐近分析*（asymptotic analysis）：把一个以 $1/b$ 的幂写成的级数代入方程，然后比较系数，做法与这里完全一样。把同样的想法应用于 $\tan x=x$，便可描述该方程的大根，这些根会在热传导和振动问题中出现。如想进一步了解这类展开式背后的分析学，可参阅 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门教材）；Siklos 的 *Advanced Problems in Mathematics* 则有更多 STEP 风格的练习。

详细解答

STEP III 1990，第 11 题

力学 • 刚体转动

STEP III 1990，第 11 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：

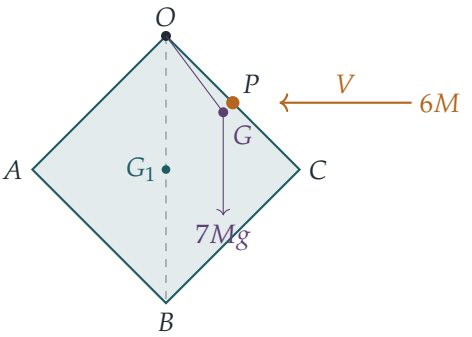
step.maths.org/questions/90-s3-q11

关键思路

碰撞在一瞬间发生，其间位于 O 的转轴会施加一个很大的冲力（impulsive force），所以线动量（linear momentum）并不守恒。不过这个冲量（impulse）作用于 O，因此对 O 没有力矩（moment）：在整个碰撞过程中，对转轴的角动量（angular momentum）守恒。碰撞之后，整个系统是一个绕固定轴转动的刚体（rigid body），只有重力作功，于是能量守恒（conservation of energy）便可告诉我们它能否越过最高点。

解答

我们把薄片（lamina）视为均匀的刚性正方形，把质点视为在碰撞一刻黏附在薄片上的点质量，并假设转轴光滑。取 O 为原点。当 OB 铅垂向下时，正方形的中心 G_1 位于深度 $\sqrt{2}a$ 处（即对角线 $2\sqrt{2}a$ 的一半），而 A 和 C 分别位于 OB 两侧、水平距离为 $\sqrt{2}a$ 之处，深度同样是 $\sqrt{2}a$。因此 OC 的中点 P 与 O 的水平距离为 $a/\sqrt{2}$，深度为 $a/\sqrt{2}$，并且 $OP = a$。



（图中所示的质点从右方飞来；稍后我们会看到，从左方飞来也会得到相同的答案。）

对转轴的转动惯量

对于边长为 $2a$ 的均匀正方形薄片，它对通过中心并垂直于薄片的轴的转动惯量（moment of inertia）是 $\frac{1}{12}M\big((2a)^2+(2a)^2\big)=\frac{2}{3}Ma^2$（根据垂直轴定理（perpendicular axis theorem），把薄片平面内两条轴的转动惯量相加，每条均为 $\frac{1}{3}Ma^2$）。由于 $OG_1=\sqrt{2}\,a$，根据平行轴定理（parallel axis theorem），

$$I_{\text{薄片}}=\tfrac{2}{3}Ma^2+M\big(\sqrt{2}\,a\big)^2=\tfrac{8}{3}Ma^2.$$

质点最后停在 P，与 O 相距 a，所以合成后的物体有

$$I=\tfrac{8}{3}Ma^2+6Ma^2=\tfrac{26}{3}Ma^2.$$

碰撞

想一想 为甚么处理这次碰撞时不能使用线动量守恒?

因为转轴是固定的：要令 O 保持不动，转轴必须在碰撞期间向薄片施加一个冲量，而这个冲量是未知的。不过它的作用线通过 O，所以它对 O 的力矩为零，而重力在一瞬间内不会产生冲量。因此对转轴的角动量守恒。

在碰撞前的一刻，质点对 O 的角动量等于它的动量 $6MV$ 乘以由 O 至其（水平）运动路线的垂直距离，这个距离就是 P 的深度，即 $a/\sqrt{2}$。薄片此时静止。碰撞刚结束时，整个物体以角速率（angular speed）ω 转动，所以

$$\tfrac{26}{3}Ma^2\,\omega=6MV\cdot\frac{a}{\sqrt{2}}\qquad\Longrightarrow\qquad\omega=\frac{9V}{13\sqrt{2}\,a}.$$

于是紧接碰撞之后的动能（kinetic energy）为

$$\tfrac{1}{2}I\omega^2=\tfrac{1}{2}\cdot\tfrac{26}{3}Ma^2\cdot\frac{81V^2}{338a^2}=\frac{27}{26}MV^2.$$

作完整的旋转

合成物体的质心（centre of mass）G（质量为 $7M$）相对于 O 的水平位移和深度分别是

$$\frac{M\cdot 0+6M\cdot\frac{a}{\sqrt{2}}}{7M}=\frac{3\sqrt{2}}{7}a,\qquad\frac{M\cdot\sqrt{2}\,a+6M\cdot\frac{a}{\sqrt{2}}}{7M}=\frac{4\sqrt{2}}{7}a,$$

所以 $OG=\frac{\sqrt{2}}{7}a\sqrt{3^2+4^2}=\frac{5\sqrt{2}}{7}a$。

这个物体是刚体，并且钉在 O 上，所以（与系在绳上的质点不同）它不可能“掉离”所走的圆形轨道；它能作完整的旋转，当且仅当它在势能（potential energy）最高的位置仍

有正的动能，而这个位置就是 G 在 O 正上方的时候。由起初在 O 之下深度 $\frac{4\sqrt{2}}{7}a$ 处，到在 O 之上高度 $\frac{5\sqrt{2}}{7}a$ 处，G 上升了

$$\frac{4\sqrt{2}}{7}a+\frac{5\sqrt{2}}{7}a=\frac{9\sqrt{2}}{7}a,$$

无论物体向哪一个方向转动都是如此，所以质点从哪一边飞来并无影响。碰撞之后只有重力作功（光滑的转轴不作功），所以根据能量守恒，我们需要

$$\frac{27}{26}MV^2>7Mg\cdot\frac{9\sqrt{2}}{7}a=9\sqrt{2}\,Mga,$$

即

$$\boxed{V^2>\frac{26\sqrt{2}}{3}ga.}$$

质量 M 被约去了，这是必然的，因为题目中每一个质量都是 M 的倍数。（若刚好取等号，物体只会缓缓趋近最高位置而永远不会越过它，所以这是严格不等式。）

验算

- 转动惯量、质心（它与向下铅垂线所成角的 tan 值为 $\frac{3}{4}$）以及动能系数 $\frac{27}{26}$ 均已用数值方法核实，其中第一项是在倾斜的正方形上以网格积分求得的。
- 取 $a=1.3$，$g=9.81$，以数值方法逐步求解运动方程 $I\ddot{\theta}=-7Mg\,x_G$：当 V^2 为 $\frac{26\sqrt{2}}{3}ga\approx 15.6$ 的 98% 时，物体会摆回来；而在 102% 时，它会转完一整圈，无论质点从哪一边飞来都是如此。
- 对能量作合理性检查：质点原有的 $3MV^2$ 动能之中，只有 $\frac{27}{26}MV^2$ 在碰撞后保留下来，这正符合完全非弹性碰撞（perfectly inelastic collision）应有的情况。

常见错误

- 在碰撞过程中使用线动量守恒：转轴会施加冲量，所以线动量并不守恒。
- 以距离 $OP=a$ 作为质点动量的力臂，而没有使用到其运动路线的垂直距离 $a/\sqrt{2}$。
- 忘记质点黏上之后质心已经偏离 OB，所以升至最高点的高度并不单单是原来深度的两倍。

STEP 之外

这是刚体绕固定轴转动的力学：由于转轴上有冲击反作用力，在碰撞中守恒的量是角动量，而不是线动量。一个精彩的延伸课题是撞击中心（centre of percussion）：在这一点上

施加打击, 转轴完全不会产生冲击反作用力, 这正是板球棒有“甜蜜点”的原因。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对刚体转动和角冲量有细致的论述。

详细解答

STEP III 1990，第 12 题

力学 • 移动楔上的质点

> **STEP III 1990，第 12 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-q12

关键思路

质点下滑期间有两个量守恒：水平动量（horizontal momentum）（桌面光滑，所以没有水平方向的外力）和能量（所有表面都是光滑的）。这两条定律，加上质点相对于楔块始终沿斜面 AC 运动这一事实，就确定了所有速度。之后的反弹和飞行只是抛体运动（projectile motion），而楔块则以恒定速率滑行。

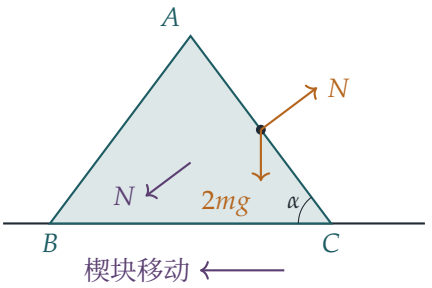
解答

建立模型

把质点视为质量为 $2m$ 的点质量（point mass），并把楔块（wedge）视为质量为 m 的刚体（rigid body）。所有接触面都是光滑的。在通过 A、B、C 的铅垂截面上考虑问题。取 B 为原点，$C = (6a, 0)$，$A = (3a, 4a)$。A 的高度是 $4a$，因为三角形是等腰三角形，$AB = AC = 5a$，而底边的一半是 $3a$。设 α 为角 ACB。则

$$\sin\alpha = \frac{4}{5}, \qquad \cos\alpha = \frac{3}{5}.$$

质点沿 AC 下滑，向 x 轴正方向运动，而楔块则向 x 轴负方向后退。



设 u 为质点相对于楔块的速率，方向沿 AC 向下。设 V 为楔块的速度，其方向是水平的。质点的速度等于楔块的速度加上质点相对于楔块的速度：

$$\mathbf{v}_P = \big(V + u\cos\alpha,\ -u\sin\alpha\big) = \big(V + \tfrac{3}{5}u,\ -\tfrac{4}{5}u\big).$$

下滑期间的动量与能量

外力只有重力和桌面的铅垂反作用力，所以水平动量守恒。它起初为零，因此

$$2m\big(V + \tfrac{3}{5}u\big) + mV = 0 \quad\Longrightarrow\quad V = -\tfrac{2}{5}u.$$

所以质点的速度是 $\big(\frac{1}{5}u,\ -\frac{4}{5}u\big)$，楔块的速度是 $-\frac{2}{5}u$。

> ? **想一想** 为甚么不能单独对质点使用能量守恒?

作用于质点的法向反作用力（normal reaction）垂直于斜面 AC，但斜面本身正在移动。因此质点的实际速度并不沿 AC 方向，而 N 对质点作功。只有整个系统的能量才守恒，因为这一对法向力所作的总功为零。

质点到达 AC 底部时，已下降了高度 $4a$。对整个系统应用能量守恒，得

$$\begin{aligned}\tfrac{1}{2}(2m)\Big(\tfrac{1}{25} + \tfrac{16}{25}\Big)u^2 + \tfrac{1}{2}m\cdot\tfrac{4}{25}u^2 &= 2mg\cdot 4a\\ \tfrac{19}{25}u^2 &= 8ga\\ u^2 &= \frac{200ga}{19}.\end{aligned}$$

质点自始至终都留在斜面上。按加速度分解各力，可知法向反作用力是常数 $N = \frac{10}{19}mg > 0$。

碰撞时各物体的位置

动量始终为零，所以系统的质心（centre of mass）没有水平运动。设质点到达桌面时，质点的水平位移为 x_P，楔块的水平位移为 x_W。则 $2m\,x_P + m\,x_W = 0$。相对于楔块，质点

由 A 走到 C，水平距离为 $3a$，所以 $x_P - x_W = 3a$。因此

$$x_P = a, \qquad x_W = -2a.$$

在该瞬间，质点位于 $x = 3a + a = 4a$，而顶角 C 位于 $6a - 2a = 4a$。两者在同一位置，这是必然的，因为质点刚好到达 C。

反弹与飞行

桌面固定而且光滑，而碰撞是完全弹性（perfectly elastic）的。所以质点的水平速度不变，铅垂速度则反向。它以速度 $\left(\frac{1}{5}u,\ \frac{4}{5}u\right)$ 离开桌面。它返回桌面所需的飞行时间为

$$T = \frac{2 \cdot \frac{4}{5}u}{g} = \frac{8u}{5g}.$$

飞行期间，它向右水平移动了距离 $\frac{1}{5}u \cdot T = \dfrac{8u^2}{25g}$。

质点离开楔块之后，楔块不再受水平力作用，所以它继续以速率 $\frac{2}{5}u$ 向左运动。在时间 T 内，它移动了 $\frac{2}{5}u \cdot T = \dfrac{16u^2}{25g}$。质点向右运动而楔块向左运动，所以质点不可能再与楔块相遇。

D 与 E 之间的距离是这两段距离之和，因为两者都由 $x = 4a$ 出发，并向相反方向运动：

$$DE = \frac{8u^2}{25g} + \frac{16u^2}{25g} = \frac{24u^2}{25g} = \frac{24}{25} \cdot \frac{200a}{19} = \frac{192}{19}\,a,$$

正如所求。

验算

- 碰撞时的总动能：$\frac{19}{25}mu^2 = \frac{19}{25} \cdot \frac{200}{19}mga = 8mga$。这等于所失去的位能。
- 动量：$2m \cdot \frac{1}{5}u + m \cdot \left(-\frac{2}{5}u\right) = 0$。
- 取 $a = g = m = 1$，以数值方法逐步计算运动，得 $DE \approx 10.1053$，而 $\frac{192}{19} = 10.10526\ldots$。

常见错误

- 把质点的速度当作沿 AC 方向。那只是它相对于移动中的楔块的速度。
- 单独对质点写出 $\frac{1}{2}(2m)u^2 = 2mg \cdot 4a$。这忽略了楔块的动能和 N 所作的功。
- 由 C 的原来位置量度 DE，而不是由碰撞点 $x = 4a$ 量度；在那一刻，质点和 C 都位于该处。

STEP 之外

在光滑桌面上滑动的楔块，是带有约束（constraints）的系统的经典例子。拉格朗日力学（Lagrangian mechanics）处理这类系统的方法，是选取适当的坐标（这里是楔块的位置和质点沿斜面下滑的距离），使约束力根本无须出现。这样，水平动量守恒便源于问题在水平平移下的对称性，这是诺特定理（Noether's theorem）的一个例子。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）涵盖了运动参考系，以及多体系统中的动量守恒。

详细解答

STEP III 1990，第 13 题

力学 • 球内的运动

> **STEP III 1990，第 13 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-q13

关键思路

整个过程分为两个阶段：先是在球面上作圆周运动（circular motion），直至法向反作用力（normal reaction）消失为止；然后是作为抛体（projectile）的自由飞行。离开球面的条件决定了离开一刻的速率，并可用 θ 表示；其后只需依照提示去做：把抛体重返球面的条件 $x^2 + y^2 = a^2$ 写成关于 t 的方程。由于离开条件令其中几项互相抵销，这条方程会化简得非常漂亮。

解答

模型

我们把 P 视为质点（particle），并把球面（半径 a）视为光滑（smooth），因此 P 与球面接触时，所受的力只有它的重量 mg 和指向 O 的法向反作用力 R。运动在通过 O 的一个铅垂平面内进行；我们取 O 为原点，x 轴为水平方向，y 轴铅垂向上，并设 P 沿球面的右边向上爬升。

P 若要离开球面，就必定是在高于 O 的水平之处离开：在该水平以下，重力沿 PO 的分量背向 O，所以当 P 仍在圆上运动时，R 在那里不可能为零。因此 $0 < \theta < \frac{\pi}{2}$。

离开球面

设 P 的速率为 v 时，OP 与向上的铅垂线所成的角为 θ，于是 P 位于 $a\cos\theta$ 的高度（相对于 O 而言）。沿指向 O 的方向分解，重量在该方向的分量为 $mg\cos\theta$，所以

$$R + mg\cos\theta = \frac{mv^2}{a}.$$

P 在 $R = 0$ 时离开球面，所以在该刻

$$v^2 = ag\cos\theta.$$

（由能量守恒（conservation of energy）可知，抛射速率必须是 u，其中 $u^2 = v^2 + 2ga(1 + \cos\theta) = ag(2 + 3\cos\theta)$，所以只要适当选取 u，$(0, \frac{\pi}{2})$ 内的每一个 θ 都可以达到。随着 P 向上爬升，v 减小，R 亦减小，所以 R 正是在这一点首次变为零。）

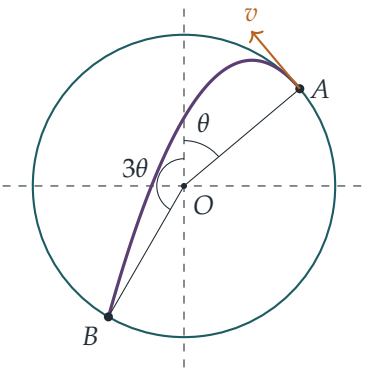
自由飞行

把离开的一刻记为 $t = 0$，此时 P 位于 $(a\sin\theta, a\cos\theta)$，其速度沿切线（tangent）方向，与 OP 垂直，并指向上方、越过球顶：

$$\text{速度} = v(-\cos\theta, \sin\theta).$$

此后只有重力作用，所以

$$x = a\sin\theta - vt\cos\theta, \qquad y = a\cos\theta + vt\sin\theta - \tfrac{1}{2}gt^2.$$



P 在 A 离开球面，并且（如下文所证）在 B 再次撞上球面。

P 位于球面上的条件是 $x^2 + y^2 = a^2$。展开后，关于 t 的常数项和一次项分别是

$$a^2\sin^2\theta + a^2\cos^2\theta = a^2 \qquad \text{及} \qquad -2avt\sin\theta\cos\theta + 2avt\cos\theta\sin\theta = 0,$$

所以

$$\begin{aligned} x^2 + y^2 - a^2 &= v^2t^2\cos^2\theta + v^2t^2\sin^2\theta - gt^2\left(a\cos\theta + vt\sin\theta\right) + \tfrac{1}{4}g^2t^4 \\ &= \left(v^2 - ag\cos\theta\right)t^2 - gv\sin\theta\, t^3 + \tfrac{1}{4}g^2t^4 \\ &= gt^3\left(\tfrac{1}{4}gt - v\sin\theta\right), \end{aligned}$$

最后一步用了 $v^2 = ag\cos\theta$。

想一想 为甚么 $t=0$ 是三重根（triple root），而不只是单根?

在离开的一刻，单靠重力就刚好提供了沿圆运动所需、指向 O 的加速度 v^2/a。因此抛物线起初不但与球面有相同的方向，*而且*有相同的曲率（curvature）：它先紧贴着球面，然后才向球内偏离。这就是为甚么除了 t 项之外，连 t^2 项也抵销了。

当 $0<t<4v\sin\theta/g$ 时，上式为负，所以 P 严格位于球内，而它下一次与球面相遇的时刻是

$$T=\frac{4v\sin\theta}{g}.$$

落点位置

在 $t=T$ 时，利用 $vT\sin\theta=4v^2\sin^2\theta/g=4a\cos\theta\sin^2\theta$ 和 $\frac{1}{2}gT^2=8v^2\sin^2\theta/g=8a\cos\theta\sin^2\theta$，可得

$$y=a\cos\theta+4a\cos\theta\sin^2\theta-8a\cos\theta\sin^2\theta=a\cos\theta\left(1-4\sin^2\theta\right).$$

（这就是 $a\cos3\theta$，因为 $\cos3\theta=4\cos^3\theta-3\cos\theta=\cos\theta(1-4\sin^2\theta)$；同样地 $x=-a\sin3\theta$，所以撞击点 B 位于另一边、与向上的铅垂线成 3θ 角之处，正如图中所示。）

P 撞在低于 O 的水平之处，当且仅当 $y<0$。由于 $0<\theta<\frac{\pi}{2}$，故 $\cos\theta>0$，所以

$$y<0\iff1-4\sin^2\theta<0\iff\sin\theta>\tfrac{1}{2}\iff\theta>\tfrac{\pi}{6},$$

最后一步成立，是因为 sin 在 $(0,\frac{\pi}{2})$ 上递增。当 $\theta=\frac{\pi}{6}$ 恰好成立时，P 撞上球面的位置恰好与 O 同一水平。所以 θ 必须超过的角是

$$\boxed{\theta=\frac{\pi}{6}=30^\circ}$$

而每一个 θ 只要满足 $30^\circ<\theta<90^\circ$，撞击点确实都低于 O 的水平。

验算

- $\theta=60^\circ$：公式给出 $y=a\cos180^\circ=-a$，所以 P 落回最低点，即抛射点。若取 $\theta=45^\circ$，它落在与向上的铅垂线成 135° 之处；两者都有 $y<0$，与预期相符。
- 当 $\theta\to0$ 时，$T\to0$ 且 $B\to A$：一个在球顶仅仅离开球面的质点，会立即落回球面上。
- 以数值模拟（逐步计算球面上的运动直至 $R=0$，然后计算抛体运动直至它与球面相遇）检验 $\theta=10^\circ,25^\circ,29^\circ,30.5^\circ,31^\circ,45^\circ,60^\circ,80^\circ$ 各情况，结果与 $T=4v\sin\theta/g$ 和 $y=a\cos3\theta$ 吻合至小数点后四位；例如 $\theta=31^\circ$ 给出 $y/a=-0.0523$。

常见错误

- 弄错速度的方向:速度与 OP 垂直,并且仍然向上、越过球顶,其分量为 $(-v\cos\theta,\ v\sin\theta)$,而不是 $(v\sin\theta,\ v\cos\theta)$。
- 忘记先代入 $v^2 = ag\cos\theta$ 才去解 $x^2 + y^2 = a^2$,结果得到一条棘手的四次方程(quartic),而不是简洁的因式分解(factorisation)。
- 把 $1 - 4\sin^2\theta < 0$ 两边相除时,没有指出 $\cos\theta > 0$(这是因为 P 在高于 O 的水平之处离开)。

STEP 之外

$t = 0$ 处的三重根,其实是关于**密切**(osculation)的一个陈述:抛物线与圆在离开点有相同的切线和相同的曲率,所以两者有二阶切触。曲率和密切圆(osculating circle)是曲线的微分几何(differential geometry)中的核心概念;在力学中,每当我们把加速度分解为切向和法向两部分时,它们都会再次出现。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*(力学入门教科书)对极坐标(polar coordinates)下的运动,以及这类受约束的运动,都有细致的论述。

详细解答

STEP III 1990，第 14 题

力学 · 力系

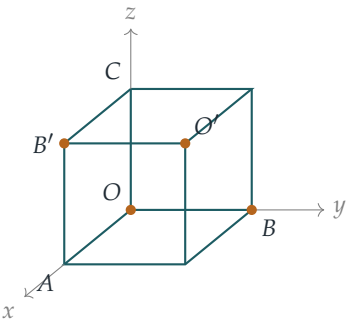
> **STEP III 1990，第 14 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/90-s3-q14

关键思路

作用于刚体(rigid body)上的力系,归根究底只由两个向量决定:合力(resultant)$\mathbf{R} = \sum \mathbf{F}_i$,以及对某一选定点的力矩（moment）$\mathbf{G}_P = \sum(\mathbf{r}_i - \mathbf{p}) \times \mathbf{F}_i$。把基点由 O 移到 P，力矩按一条简单的规则改变：$\mathbf{G}_P = \mathbf{G}_O - \mathbf{p} \times \mathbf{R}$。本题每一部分，都是把这条规则用于不同的 P。

解答

把正方体视为边长为 a 的刚体，于是 $A = (a,0,0)$，$B = (0,a,0)$，$C = (0,0,a)$，而与 B 相对的顶点是 $B' = (a,0,a)$。



求 λ、μ、ν

作用于 O 的力对 O 没有力矩，所以只需考虑其余三个力。利用 $(x,y,z)\times(p,q,r)=(yr-zq,\ zp-xr,\ xq-yp)$：

$$\overrightarrow{OO'}\times\begin{pmatrix}\lambda\\0\\1\end{pmatrix}=a\begin{pmatrix}1\\1\\1\end{pmatrix}\times\begin{pmatrix}\lambda\\0\\1\end{pmatrix}=a\begin{pmatrix}1\\\lambda-1\\-\lambda\end{pmatrix},$$

$$\overrightarrow{OB}\times\begin{pmatrix}-1\\0\\2\end{pmatrix}=a\begin{pmatrix}0\\1\\0\end{pmatrix}\times\begin{pmatrix}-1\\0\\2\end{pmatrix}=a\begin{pmatrix}2\\0\\1\end{pmatrix},$$

$$\overrightarrow{OB'}\times\begin{pmatrix}1\\\mu\\\nu\end{pmatrix}=a\begin{pmatrix}1\\0\\1\end{pmatrix}\times\begin{pmatrix}1\\\mu\\\nu\end{pmatrix}=a\begin{pmatrix}-\mu\\1-\nu\\\mu\end{pmatrix}.$$

相加后，对 O 的总力矩为 $a\,(3-\mu,\ \lambda-\nu,\ 1-\lambda+\mu)$。令每个分量为零，得 $\mu=3$，继而 $\lambda=1+\mu=4$，再得 $\nu=\lambda=4$：

$$\lambda=4,\qquad \mu=3,\qquad \nu=4.$$

以上结果完全不取决于 α,β,γ，而合力为

$$\mathbf{R}=\begin{pmatrix}\alpha+\lambda-1+1\\\beta+\mu\\\gamma+1+2+\nu\end{pmatrix}=\begin{pmatrix}\alpha+4\\\beta+3\\\gamma+7\end{pmatrix},\qquad \mathbf{G}_O=\mathbf{0}.$$

第 (i) 部分

当 $\alpha=\beta=\gamma=0$ 时，$\mathbf{R}=(4,3,7)$。两个力系等效（equivalent），是指它们有相同的合力，并且对某一点（因而对每一点）有相同的力矩。所以作用于 B 的单一力必定就是 $\mathbf{R}$ 本身，而力偶（couple）必须等于原力系对 B 的力矩：

$$\mathbf{G}_B=\mathbf{G}_O-\overrightarrow{OB}\times\mathbf{R}=\mathbf{0}-a\begin{pmatrix}0\\1\\0\end{pmatrix}\times\begin{pmatrix}4\\3\\7\end{pmatrix}=-a\begin{pmatrix}7\\0\\-4\end{pmatrix}=a\begin{pmatrix}-7\\0\\4\end{pmatrix}.$$

? **想一想** 为甚么对 B 的力矩等于 $\mathbf{G}_O-\overrightarrow{OB}\times\mathbf{R}$?

对每一个力，$(\mathbf{r}_i-\mathbf{b})\times\mathbf{F}_i=\mathbf{r}_i\times\mathbf{F}_i-\mathbf{b}\times\mathbf{F}_i$；对 i 求和便得 $\mathbf{G}_O-\mathbf{b}\times\mathbf{R}$。

因此，等效力系是一个作用于 B 的力 $(4,3,7)$，连同一个力矩为 $a(-7,0,4)$ 的力偶。验算一下：作用于 B 的力加上这个力偶，对 O 的力矩等于 $a(7,0,-4)+a(-7,0,4)=\mathbf{0}$，正如所料。

第 (ii) 部分

现在 $\mathbf{R} = (2+4,\ 3+3,\ 2+7) = (6,6,9) = 3(2,2,3)$，而 $\mathbf{G}_O = \mathbf{0}$ 依然成立。在位置向量（position vector）为 $\mathbf{r} = (x,y,z)$ 的点 P，

$$\mathbf{G}_P = \mathbf{G}_O - \mathbf{r}\times\mathbf{R} = -\mathbf{r}\times\mathbf{R}.$$

由于 $\mathbf{R}\neq\mathbf{0}$，上式为零当且仅当 $\mathbf{r}$ 与 $\mathbf{R}$ 平行（或 $\mathbf{r}=\mathbf{0}$）。所以所求轨迹（locus）是通过 O、方向为 $(2,2,3)$ 的直线：

$$\frac{x}{2} = \frac{y}{2} = \frac{z}{3}.$$

从物理上看，整个力系等效于沿这条直线作用的单一力 $(6,6,9)$。

当正方体平移（不带转动）时，每一点的位移都相同，所以各力所作的功（work done）是 $\mathbf{R}\cdot\mathbf{s}$，其中 $\mathbf{s}$ 是位移。若沿该直线、朝 $(2,2,3)$ 的方向移动距离 d，则 $\mathbf{s} = \dfrac{d}{\sqrt{17}}(2,2,3)$，而

$$W = (6,6,9)\cdot\frac{d}{\sqrt{17}}(2,2,3) = \frac{51d}{\sqrt{17}} = 3\sqrt{17}\,d.$$

题目没有为这段距离命名；每单位距离所作的功是 $3\sqrt{17}\approx 12.37$ 个单位。由于 $\mathbf{R}$ 沿着这条直线，这其实就是 $|\mathbf{R}|\,d = \sqrt{36+36+81}\,d = \sqrt{153}\,d$，数值相同。（若沿直线反方向移动，则得 $-3\sqrt{17}\,d$。）

验算

- 取 $\lambda=4$、$\mu=3$、$\nu=4$，三个对 O 的力矩分别是 $a(1,3,-4)$、$a(2,0,1)$ 和 $a(-3,-3,3)$，其和为零。
- 在 (ii) 中，取 $P=(2,2,3)$：$\mathbf{r}\times\mathbf{R} = (2,2,3)\times(6,6,9) = \mathbf{0}$。在 $P=(1,0,0)$，力矩为 $-(0,-9,6)\neq\mathbf{0}$。
- 以电脑用精确分数运算，核实了每一个力矩，检查了 (i) 中的等效力系对 20 个随机点的力矩均与原力系相同，并确认每单位距离所作的功为 $3\sqrt{17}\approx 12.369$。

常见错误

- 把 $\mathbf{r}\times\mathbf{F}$ 的次序弄错：次序颠倒，所有正负号都会反转。
- 把 (i) 中的力偶当作 $+\overrightarrow{OB}\times\mathbf{R}$。把力由 O 移到 B，所需的补偿力偶带相反的符号。
- 逐一计算作用于不同顶点的各个力所作的功再相加。在纯平移中，只有合力才起作用。

STEP 之外

作用于刚体上的任何力系，都可以化简为沿某条直线的单一力，连同一个绕同一条直线的力偶。这称为*力螺旋*（wrench），其螺距（pitch）为 $\mathbf{R}\cdot\mathbf{G}/|\mathbf{R}|^2$。在 (ii) 中螺距为零，所以力螺旋只是一个单一力，而零力矩轨迹就是它的作用线（line of action），即中心轴（central axis）。这些概念通往刚体动力学（rigid-body dynamics），以及机械人学中的旋量理论（screw theory）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）以恰当的程度讲解力矩和刚体。

详细解答

STEP III 1990，第 15 题

概率与统计 • 几何分布

> **STEP III 1990，第 15 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-q15

关键思路

每次投掷出现 A 的概率（probability）是 $\frac{1}{4}$，出现 B 的概率是 $\frac{3}{4}$，而且各次投掷互相独立（independent）。因此事件 $\{M = m,\ N = n\}$ 就是某一串特定的投掷结果，其概率是一个乘积。这个乘积可以拆成只与 m 有关的因子乘以只与 n 有关的因子，而这一点几乎包办了所有工作：边际分布（marginal distribution）是几何分布（geometric distribution），两个变量互相独立，其后每一个概率都是一个等比级数（geometric series）。

解答

设单次投掷中 $p = \mathrm{P}(A) = \frac{3}{12} = \frac{1}{4}$，$q = \mathrm{P}(B) = \frac{9}{12} = \frac{3}{4}$。各次投掷互相独立。

事件 $\{M = m,\ N = n\}$ 发生，当且仅当最初的 $m + n + 1$ 次投掷为

$$\underbrace{A\,A\cdots A}_{m}\ \underbrace{B\,B\cdots B}_{n}\ A,$$

所以根据独立性，对于 $m \geq 0$ 及 $n \geq 1$，

$$\mathrm{P}(M = m,\ N = n) = \left(\tfrac{1}{4}\right)^m \left(\tfrac{3}{4}\right)^n \tfrac{1}{4} = \left(\tfrac{1}{4}\right)^{m+1} \left(\tfrac{3}{4}\right)^n.$$

（这串投掷结果以概率 1 终会出现一个 B，然后再出现一个 A，所以这些概率之和为 1；下面的边际分布可以印证这一点。）

第 (i) 部分

求 M 的边际分布时，对 $n \geq 1$ 求和，并利用 $\sum_{n\geq 1}\left(\frac{3}{4}\right)^n = \dfrac{3/4}{1-3/4} = 3$：

$$\mathrm{P}(M=m) = \left(\tfrac{1}{4}\right)^{m+1}\cdot 3 = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}, \qquad m = 0,1,2,\ldots$$

求 N 的边际分布时，对 $m \geq 0$ 求和，并利用 $\sum_{m\geq 0}\left(\frac{1}{4}\right)^{m+1} = \dfrac{1/4}{1-1/4} = \frac{1}{3}$：

$$\mathrm{P}(N=n) = \tfrac{1}{3}\left(\tfrac{3}{4}\right)^{n} = \tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1}, \qquad n = 1,2,3,\ldots$$

两个结果都可以直接理解：M 是第一个 B 出现之前 A 的个数，而 N 则是一连串 B 开始之后，等待第一个 A 出现所需的次数。两者各自的总和都是 1。

第 (ii) 部分

我们利用 $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$，其中 $|x|<1$（把等比级数求导即得）。于是

$$\mathrm{E}(M) = \tfrac{3}{4}\sum_{m\geq 1} m\left(\tfrac{1}{4}\right)^{m} = \tfrac{3}{4}\cdot\tfrac{1}{4}\cdot\frac{1}{(3/4)^2} = \tfrac{1}{3},$$

$$\mathrm{E}(N) = \tfrac{1}{4}\sum_{n\geq 1} n\left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4}\cdot\frac{1}{(1/4)^2} = 4.$$

独立性

把两个边际分布相乘，

$$\mathrm{P}(M=m)\,\mathrm{P}(N=n) = \tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\cdot\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1} = \left(\tfrac{1}{4}\right)^{m+1}\left(\tfrac{3}{4}\right)^{n} = \mathrm{P}(M=m,\ N=n)$$

对所有 $m \geq 0$、$n \geq 1$ 都成立。所以 M 与 N 确实互相独立。

> ? **想一想** 一连串 A 的长度，难道不会影响一连串 B 的长度吗？为甚么不会？

一连串 A 在第一个 B 出现时结束，而由该次投掷起，整个过程重新开始：余下的投掷与之前发生的一切互相独立。一连串 B 的长度只取决于这些较后的投掷，所以知道 M 的值，并不能告诉我们任何关于 N 的资料。联合概率（joint probability）可以分解为两个因子，正是这个事实的代数形式。

涉及 N 与 M 的概率

对于整数 $k \geq 1$，把几何分布的尾部求和：

$$\mathrm{P}(N>k) = \sum_{n\geq k+1}\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{n-1} = \tfrac{1}{4}\cdot\frac{(3/4)^k}{1-3/4} = \left(\tfrac{3}{4}\right)^{k}.$$

这条公式同时给出 $\mathrm{P}(N>0)=1$，由于 $N\geq 1$，这是正确的。

求 $\mathrm{P}(N>M)$ 时，以 M 的值为条件，并利用独立性：

$$\begin{aligned}\mathrm{P}(N>M) &= \sum_{m\geq 0}\mathrm{P}(M=m)\,\mathrm{P}(N>m) = \sum_{m\geq 0}\tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\left(\tfrac{3}{4}\right)^{m}\\ &= \tfrac{3}{4}\sum_{m\geq 0}\left(\tfrac{3}{16}\right)^{m} = \tfrac{3}{4}\cdot\tfrac{16}{13} = \tfrac{12}{13}.\end{aligned}$$

至于 $\mathrm{P}(N=M)$，由于 $N\geq 1$，两者的共同值必须至少为 1：

$$\begin{aligned}\mathrm{P}(N=M) &= \sum_{m\geq 1}\tfrac{3}{4}\left(\tfrac{1}{4}\right)^{m}\cdot\tfrac{1}{4}\left(\tfrac{3}{4}\right)^{m-1} = \tfrac{3}{64}\sum_{m\geq 1}\left(\tfrac{3}{16}\right)^{m-1}\\ &= \tfrac{3}{64}\cdot\tfrac{16}{13} = \tfrac{3}{52}.\end{aligned}$$

最后，$N>M$、$N=M$、$N<M$ 这三个事件互斥（exclusive）且穷尽（exhaustive）所有可能，所以

$$\mathrm{P}(N<M) = 1-\tfrac{12}{13}-\tfrac{3}{52} = \tfrac{52-48-3}{52} = \tfrac{1}{52},$$

正是所求。

验算

- 两个平均值都符合直觉：A 很少出现，所以开头的一连串 A 很短（平均为 $\frac{1}{3}$），而我们平均要等 $1/p=4$ 次投掷，才有一个 A 出现来结束一连串 B。
- 直接计算，$\mathrm{P}(M>n)=\left(\frac{1}{4}\right)^{n+1}$，所以 $\mathrm{P}(N<M)=\sum_{n\geq 1}\frac{1}{4}\left(\frac{3}{4}\right)^{n-1}\left(\frac{1}{4}\right)^{n+1}=\frac{1}{64}\sum_{n\geq 1}\left(\frac{3}{16}\right)^{n-1}=\frac{1}{64}\cdot\frac{16}{13}=\frac{1}{52}$，与用减法所得的结果一致。$\mathrm{P}(N>k)=\left(\frac{3}{4}\right)^{k}$ 的直接解释是：第一个 B 之后，接着的 k 次投掷必须全部是 B。
- 把联合分布精确求和，得 $\mathrm{P}(N>M)=0.92308$，$\mathrm{P}(N=M)=0.05769$，$\mathrm{P}(N<M)=0.01923$；模拟 200 000 组投掷，所得结果为 0.923、0.058、0.019，样本平均值则为 0.334 和 3.99。

常见错误

- 在 $\mathrm{P}(N=M)$ 的求和中包括了 $m=0$：N 永远不可能是 0。
- 忘记 M 可以是 0（第一次投掷可能就是 B），以致 M 的边际分布移了位，算出错误的平均值。

STEP 之外

这里两段连串的长度互相独立, 是因为一连串独立试验会在任何只由过去决定的时刻“重新开始”，例如第一个 B 出现的时刻。这就是强马可夫性质（strong Markov property），它是更新理论（renewal theory）和马可夫链（Markov chain）分析的基础，而这些都是大学一、二年级的课题。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教科书）以恰到好处的程度讲解几何分布、独立性和马可夫链。

详细解答

STEP III 1990，第 16 题

概率与统计 · 几何概率

STEP III 1990，第 16 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-q16

本题在 *Beyond STEP* 第 12.21 节折断一根棒中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/broken-stick.html

关键思路

三段总长固定的线段，恰好在每一段都小于总长的一半时，才可以组成三角形。三角形一旦存在，它是钝角三角形（obtuse triangle）的充要条件，就是最长边的平方大于另外两边的平方和（余弦定理（cosine rule））。两部分都可化为关于随机变量（random variable）的条件，而所求的概率在 (i) 中变成区间的长度，在 (ii) 中变成区域的面积。

解答

第 (i) 部分

三段的长度分别是 X、$\frac{1}{2}(1-X)$ 和 $\frac{1}{2}(1-X)$，总长为 1。三个正数长度可以组成三角形，当且仅当每一个都小于另外两个之和，即小于周界的一半，在这里就是 $\frac{1}{2}$。只要 $X>0$，两段相等的线段都满足 $\frac{1}{2}(1-X)<\frac{1}{2}$，所以唯一的条件是 $X<\frac{1}{2}$。设 T 为可以组成三角形这个事件，则

$$\mathrm{P}(T)=\mathrm{P}\left(0<X<\tfrac{1}{2}\right)=\tfrac{1}{2}.$$

一个三角形是钝角三角形，当且仅当最长边的平方大于另外两边的平方和。哪一条边最长，则取决于 X。

- 若 $X\le\frac{1}{3}$，则 $X\le\frac{1}{2}(1-X)$，所以两段相等的线段之一是最长边。这时要检验的条件是 $\frac{1}{4}(1-X)^2>\frac{1}{4}(1-X)^2+X^2$，即 $X^2<0$，这是不可能发生的。这个三角形是等腰

三角形（isosceles triangle），底边不比两条等边长，所以它永远不是钝角三角形。

- 若 $\frac{1}{3} < X < \frac{1}{2}$，最长边是 X，而三角形是钝角三角形，当且仅当

$$X^2 > 2 \cdot \tfrac{1}{4}(1-X)^2 \iff X > \frac{1-X}{\sqrt{2}},$$

这里取正平方根，因为两边都是正数。整理后得 $X(\sqrt{2}+1) > 1$，即 $X > \dfrac{1}{\sqrt{2}+1} = \sqrt{2}-1 \approx 0.414$，这个值确实落在 $\left(\frac{1}{3}, \frac{1}{2}\right)$ 之内。

设 O 为三角形是钝角三角形这个事件，则 $O \cap T$ 就是事件 $\sqrt{2}-1 < X < \frac{1}{2}$。由于 X 服从均匀分布（uniform distribution），

$$\mathrm{P}(O \mid T) = \frac{\mathrm{P}(\sqrt{2}-1 < X < \frac{1}{2})}{\mathrm{P}(T)} = \frac{\frac{1}{2} - (\sqrt{2}-1)}{\frac{1}{2}} = 1 - 2\sqrt{2} + 2 = 3 - 2\sqrt{2},$$

正如所求。

第 (ii) 部分

这个构作总是可行的。 以 $(1,0)$、$(1,1)$、$(0,1)$ 为顶点的三角形 R，就是满足 $x \le 1$、$y \le 1$ 和 $x + y \ge 1$ 的点所组成的集合。题目给出的三条边 x、y、$2-x-y$ 总长为 2，它们可以组成一个三角形（可能是退化（degenerate）的），当且仅当三者皆非负，而且没有一个超过周界的一半，即 1：

- $0 \le x \le 1$ 和 $0 \le y \le 1$ 在 R 上成立；
- $2 - x - y \le 1$ 其实就是 $x + y \ge 1$，而 $2 - x - y \ge 0$ 可由 $x \le 1$、$y \le 1$ 推出。

所以 R 上的每一点都给出一个三角形。只有在 R 的边上，三角形才是退化的，而这些边的概率为零。

$\angle ABC$ 何时是钝角？ 位于 B 的角夹在 BA 与 BC 之间，所以它的对边是 $CA = y$。根据余弦定理，$\angle ABC$ 是钝角，当且仅当 $CA^2 > AB^2 + BC^2$：

$$y^2 > x^2 + (2-x-y)^2.$$

展开 $(2-x-y)^2 = (2-x)^2 - 2(2-x)y + y^2$，$y^2$ 项互相抵消，条件变成

$$2(2-x)y > x^2 + (2-x)^2 = 2x^2 - 4x + 4.$$

由于 $x \le 1$，我们有 $2 - x > 0$，所以可以把两边除以 $2(2-x)$ 而不必改变不等号的方向：

$$y > \frac{x^2 - 2x + 2}{2 - x},$$

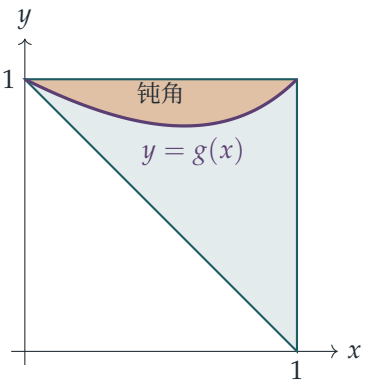
正如所求。把右边记作 $g(x)$。

想一想 曲线 $y=g(x)$ 真的一直留在三角形 R 之内吗?

我们需要它在 $0\le x\le 1$ 时位于直线 $y=1-x$ 与 $y=1$ 之间，这样钝角区域就正好是 R 在曲线上方的部分。事实上，

$$1-g(x)=\frac{(2-x)-(x^2-2x+2)}{2-x}=\frac{x(1-x)}{2-x}\ge 0,\qquad g(x)-(1-x)=\frac{x}{2-x}\ge 0,$$

而且 $g(0)=g(1)=1$。所以曲线由 $(0,1)$ 走到 $(1,1)$，中途降到 $y=1$ 之下，但从不低于斜边 $x+y=1$。



概率。 三角形 R 的面积是 $\frac12$，所以联合密度（joint density）在 R 上等于 2，于是

$$\mathrm{P}(\angle ABC\text{ 为钝角})=2\times\text{面积}\{(x,y):g(x)<y\le 1,\ 0\le x\le 1\}=2\int_0^1\big(1-g(x)\big)\,\mathrm{d}x.$$

要对 g 积分，可代入 $u=2-x$，于是 $x=2-u$，而

$$x^2-2x+2=(2-u)^2-2(2-u)+2=u^2-2u+2,\qquad g=\frac{u^2-2u+2}{u}=u-2+\frac{2}{u}.$$

当 x 由 0 变到 1 时，u 由 2 变到 1，而 $\mathrm{d}x=-\mathrm{d}u$，所以

$$\begin{aligned}\int_0^1 g(x)\,\mathrm{d}x&=\int_1^2\left(u-2+\frac{2}{u}\right)\mathrm{d}u=\left[\tfrac12u^2-2u+2\ln u\right]_1^2\\&=(2-4+2\ln 2)-\left(\tfrac12-2\right)=2\ln 2-\tfrac12.\end{aligned}$$

因此

$$\mathrm{P}(\angle ABC\text{ 为钝角})=2\left(1-2\ln 2+\tfrac12\right)=3-4\ln 2,$$

正如所求。

验算

- 数值上，$3-2\sqrt{2}\approx 0.1716$，$3-4\ln 2\approx 0.2274$。以 10^6 次试验进行蒙地卡罗模拟（Monte Carlo simulation），结果约为 0.1719 和 0.2278。
- 两个答案都在 0 与 1 之间。在 (ii) 中，曲线在两端都碰到 $y=1$：在 $x=0$ 时边 BC 消失，在 $x=1$ 时边 AB 消失，所以两者都是退化情况，正如所料。
- 以辛普森法则（Simpson's rule）计算 $\int_0^1 g(x)\,\mathrm{d}x$，得 0.886294，与 $2\ln 2-\frac{1}{2}$ 吻合。

常见错误

- 在 (i) 中忘记取条件：出现钝角三角形的无条件概率是 $\frac{3}{2}-\sqrt{2}$，只有所求答案的一半。
- 在 (ii) 中，余弦定理用错了边。位于 B 的角的对边是 $CA=y$，而不是 AB。
- 忘记 R 上的密度是 2 而不是 1，结果答案只有应有的一半。

STEP 之外

把概率化为长度或面积，就是*几何概率*（geometric probability），而随机折断一根棒正是其中一个经典问题：在两个独立的均匀随机点把棒折断，可以组成三角形的机会是 $\frac{1}{4}$。第 (ii) 部分让你初次接触连续联合分布（continuous joint distribution），当中的概率是联合密度在某个区域上的二重积分（double integral）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）有讲解联合密度，并收录了好几道这类题目。

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

1990 年详细解答. 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。

