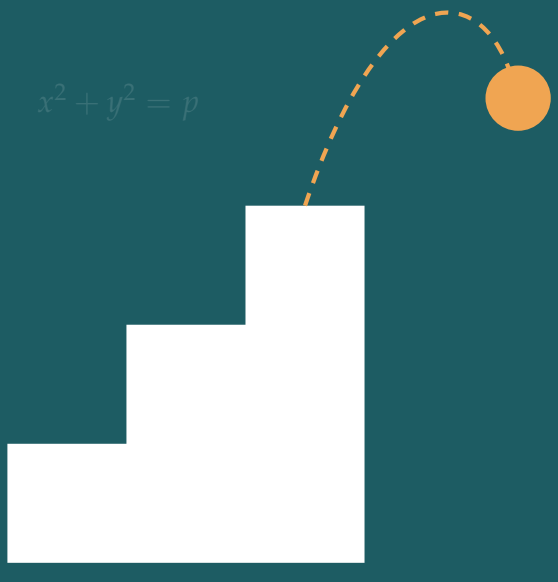
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1989 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

1989 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1989 年详细解答

1989 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 组合区域的面积 纯数学 . . . 3

第 2 题 黎曼和与极限 纯数学 . . . 6

第 3 题 向量与纯量积 纯数学 . . . 10

第 4 题 拉姆齐理论 纯数学 . . . 14

第 5 题 二项式系数恒等式 纯数学 . . . 17

第 6 题 法线与微分方程 纯数学 . . . 20

第 7 题 曲线描绘与最优化 纯数学 . . . 23

第 8 题 倍角公式 纯数学 . . . 27

第 9 题 图像的变换 纯数学 . . . 31

第 10 题 速度相关的阻力 力学 . . . 36

第 11 题 抛体射程的最优化 力学 . . . 39

第 12 题 弹性绳与能量 力学 . . . 42

第 13 题 涉及摩擦力的梯子平衡 力学 . . . 46

第 14 题 条件分布 概率与统计 . . . 49

第 15 题 分布的比较 概率与统计 . . . 53

第 16 题 零和博弈 概率与统计 . . . 56

STEP II

第 1 题 三次方程的三角解法 纯数学 . . . 60

第 2 题 幂级数的系数 纯数学 . . . 64

第 3 题 复双曲函数映射 纯数学 . . . 67

第 4 题 有理函数的描绘 纯数学 . . . 71

第 5 题 极坐标曲线与反射 纯数学 . . . 76

第 6 题 反函数的积分 纯数学 . . . 80

第 7 题 二阶微分方程 纯数学 . . . 84
第 8 题 变换群 纯数学 . . . 87
第 9 题 矩阵指数与方程组 纯数学 . . . 91
第 10 题 直线与平面 纯数学 . . . 94
第 11 题 碰撞质点与冲量 力学 . . . 97
第 12 题 刚体平衡 力学 . . . 101
第 13 题 转动惯量 力学 . . . 105
第 14 题 圆周运动 力学 . . . 110
第 15 题 几何概率 概率与统计 . . . 113
第 16 题 二项分布 概率与统计 . . . 117

STEP III

第 1 题 旋转曲面 纯数学 . . . 121
第 2 题 三维空间中的平面 纯数学 . . . 125
第 3 题 矩阵群 纯数学 . . . 128
第 4 题 曲线描绘与根 纯数学 . . . 132
第 5 题 切比雪夫多项式 纯数学 . . . 137
第 6 题 圆锥曲线的极坐标方程 纯数学 . . . 141
第 7 题 错切矩阵 纯数学 . . . 145
第 8 题 耦合微分方程 纯数学 . . . 149
第 9 题 幂级数求和 纯数学 . . . 152
第 10 题 黎曼和 纯数学 . . . 155
第 11 题 球体的斜碰撞 力学 . . . 159
第 12 题 旋转参考系 力学 . . . 163
第 13 题 简谐运动 力学 . . . 167
第 14 题 摩擦轮与冲量 力学 . . . 171
第 15 题 中位数的分布 概率与统计 . . . 174
第 16 题 概率与假设检验 概率与统计 . . . 178

详细解答

STEP I 1989，第 1 题

纯数学 · 组合区域的面积

> **STEP I 1989，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/89-s1-q1

关键思路

这个区域由半圆（semicircle）组成。先取一个大的半圆盘（half-disc），再减去和加上较小的半圆盘，就可以得到它。所以我们只需要知道每个半径，尤其是下方半圆 CDE 的半径，而这可以由直线 AG 上的各段长度求得。把半圆换成抛物线（parabola），方法完全不变，因为抛物线拱（parabolic arch）有一条简单的面积公式，可以取代 $\frac{1}{2}\pi a^2$ 的角色。

解答

半圆

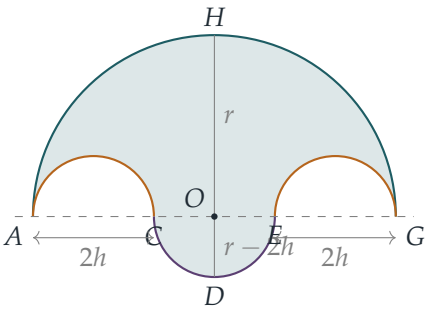
从 A 起沿直线量度。由于 ABC 的半径是 h，所以 $AC = 2h$。同理 $EG = 2h$，而 $AG = 2r$，因为 AHG 的半径是 r。于是

$$CE = 2r - 4h,$$

由于 $2h < r$，这个值是正数。所以半圆 CDE 的半径是 $r - 2h$。它的圆心是 CE 的中点，而根据对称性，这也是 AG 的中点，所以它与大半圆有共同的圆心 O。

从图中可见，ABC 和 EFG 与 H 位于 AG 的同一侧，并且在大半圆之内，而 CDE 则位于另一侧。因此所围成的区域是

$$(\text{半圆盘}AHG) \;-\; (\text{半圆盘}ABC\text{ 和}EFG) \;+\; (\text{半圆盘}CDE).$$



半径为 a 的半圆盘面积是 $\frac{1}{2}\pi a^2$，所以所围成的面积是

$$\begin{aligned}\tfrac{1}{2}\pi r^2 - 2\cdot\tfrac{1}{2}\pi h^2 + \tfrac{1}{2}\pi(r-2h)^2 &= \tfrac{1}{2}\pi(r^2 - 2h^2 + r^2 - 4rh + 4h^2)\\ &= \tfrac{1}{2}\pi(2r^2 - 4rh + 2h^2)\\ &= \pi(r-h)^2.\end{aligned}$$

？ 想一想 HD 有多长?

点 H 是大半圆的最高点，所以它位于通过 O 而垂直于 AG 的直线上，与 O 相距 r。点 D 是 CDE 的最低点，所以它也在同一条垂直线上，位于直线的另一侧，与 O 相距 $r-2h$。所以

$$HD = r + (r-2h) = 2(r-h),$$

而以 HD 为直径的圆，半径是 $r-h$，面积是 $\pi(r-h)^2$。这正是我们求得的面积，符合题目要求。

抛物线

每个半径为 a 的半圆，都换成一个抛物线拱：它建在同一条长度为 $2a$ 的弦（chord）上，顶点（vertex）在弦的中点上方，高度为 a。把原点设在弦的中点。抛物线关于通过顶点的轴对称，所以它的形式是 $y = a - kx^2$。它通过 $(\pm a, 0)$，所以 $k = 1/a$。拱与弦之间的面积是

$$\int_{-a}^{a}\left(a - \frac{x^2}{a}\right)\mathrm{d}x = 2a^2 - \frac{2a^3}{3a} = \frac{4a^2}{3}.$$

（这就是阿基米德（Archimedes）的结果：外接矩形面积的三分之二，即 $\frac{2}{3}\cdot 2a\cdot a$。）

每个拱仍然位于 AG 的同一侧，与原来的半圆一样。我们还应该检查每个小拱是否仍然在大拱之下。两个拱都从 A 出发，斜率（gradient）同为 2，所以在 A 附近这一点并不明显。在距离 A 为 x 之处，其中 $0 \le x \le 2h$，大拱的高度是 $x(2r-x)/r$，而拱 ABC 的高度是 $x(2h-x)/h$。两者之差是

$$\frac{x(2r-x)}{r} - \frac{x(2h-x)}{h} = x^2\left(\frac{1}{h} - \frac{1}{r}\right) \ge 0,$$

所以 ABC 确实在 AHG 之内，而根据对称性，EFG 也是如此。因此同样的加减分解依然适用，只需以 $\frac{4}{3}a^2$ 代替 $\frac{1}{2}\pi a^2$：

$$\begin{aligned}\tfrac{4}{3}r^2 - 2\cdot\tfrac{4}{3}h^2 + \tfrac{4}{3}(r-2h)^2 &= \tfrac{4}{3}(2r^2 - 4rh + 2h^2)\\ &= \tfrac{8}{3}(r-h)^2.\end{aligned}$$

所以换成抛物线后，所围成的面积是 $\frac{8}{3}(r-h)^2$。

验算

- 两个答案都是一个常数乘以 $(r-h)^2$。这是必然的，因为两个计算的结构完全相同：只有单个拱的系数改变了，由 $\frac{\pi}{2}$ 变成 $\frac{4}{3}$。比值 $\frac{8/3}{\pi} \approx 0.849$ 小于 1，这是合理的，因为弦和高度都相同的抛物线拱位于半圆之内。
- 极限情况 $h \to 0$：小拱消失，而 CDE 变成一个半径为 r 的完整下半圆，合起来是一个面积为 πr^2 的完整圆盘。公式 $\pi(r-h)^2$ 与此吻合。
- 数值上，取 $r = 3$ 和 $h = 1$（即试卷图中的比例），面积分别是 $4\pi \approx 12.566$ 和 $\frac{32}{3} \approx 10.667$。对描出的边界以鞋带公式（shoelace formula）计算，取八组随机的 (r,h)，结果与两条公式都吻合，误差在离散化误差范围之内。

常见错误

- 把 CDE 的半径当作 $r-h$ 或 $r-4h$。把 AG 上的各段长度画出来：$CE = 2r - 4h$，所以它的半径是 $r-2h$。
- 把上方的小半圆加上去，而不是减去。它们是从大半圆盘中切走的部分，所以其面积要减掉。

STEP 之外

第一部分的图形是*鞋匠刀形*（arbelos，即“鞋匠的刀”），是阿基米德研究过的一种形状。它的经典形式是一个大半圆，内有两个较小的半圆，而它的面积同样等于以某条精心选取的弦为直径的圆的面积。阿基米德还发现了抛物线弓形的 $\frac{2}{3}$ 法则，所用的极限论证已经很接近积分。在大学里，从他的“穷竭法”（method of exhaustion）走到黎曼积分（Riemann integral）的那一步，是第一门分析课程的核心内容。想了解这一步，Abbott 的 *Understanding Analysis*（分析学入门教材）是很好的选择。

详细解答

STEP I 1989，第 2 题

纯数学 · 黎曼和与极限

> **STEP I 1989，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/89-s1-q2

关键思路

这个古怪的式子，其实只是一些长方形面积之和的伪装。如果直接写出 n 个长方形的总面积，会得到一个含有 $k\ln k$ 的和；那些阶乘（factorial）的对数，不过是把同一个和重新分组而已。只要把两者对上，极限（limit）就是那个积分，而第一部分正好让我们把它算出来。

解答

不定积分

用分部积分法（integration by parts），取 $u=\ln x$ 和 $\frac{\mathrm{d}v}{\mathrm{d}x}=x$，于是 $\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{1}{x}$ 而 $v=\frac{1}{2}x^2$，

$$\begin{aligned}\int x\ln x\,\mathrm{d}x &= \tfrac{1}{2}x^2\ln x-\int \tfrac{1}{2}x^2\cdot\frac{1}{x}\,\mathrm{d}x\\ &= \tfrac{1}{2}x^2\ln x-\tfrac{1}{4}x^2+C.\end{aligned}$$

积分的值

由于 $\ln(1/x)=-\ln x$，被积函数（integrand）是 $-x\ln x$。被积函数在 $x=0$ 处没有定义，所以我们先在 $[\varepsilon,1]$ 上积分，再令 $\varepsilon\to 0^+$：

$$\int_{\varepsilon}^{1} x\ln(1/x)\,\mathrm{d}x=\left[\tfrac{1}{4}x^2-\tfrac{1}{2}x^2\ln x\right]_{\varepsilon}^{1}=\tfrac{1}{4}-\tfrac{1}{4}\varepsilon^2+\tfrac{1}{2}\varepsilon\,(\varepsilon\ln\varepsilon).$$

当 $\varepsilon\to 0$ 时，$\varepsilon^2\to 0$ 且 $\varepsilon\ln\varepsilon\to 0$（即题目给出的事实），所以

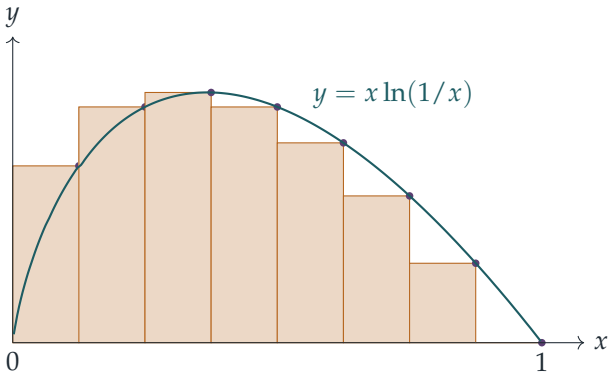
$$\int_0^1 x\ln(1/x)\,\mathrm{d}x=\frac{1}{4}.$$

同一个事实也说明，当 $x \to 0^+$ 时 $f(x) = x\ln(1/x)$ 趋向 0，所以只要定义 $f(0) = 0$，f 就在整个 $[0,1]$ 上连续（continuous）。曲线由 0 上升，在 $x = \frac{1}{\mathrm{e}}$ 处（该处 $f'(x) = -\ln x - 1 = 0$）达到极大值 $\frac{1}{\mathrm{e}}$，然后在 $x = 1$ 处回落到 0。

长方形

把 $[0,1]$ 分成 n 条阔度为 $\frac{1}{n}$ 的长条。第 k 条是 $\left[\frac{k-1}{n}, \frac{k}{n}\right]$，其上的长方形的右上角落在曲线上，所以它的高是

$$f\left(\frac{k}{n}\right) = \frac{k}{n}\ln\frac{n}{k} = \frac{k}{n}(\ln n - \ln k).$$



因此，这些长方形的总面积是

$$\begin{aligned} S_n = \sum_{k=1}^{n} \frac{1}{n}\cdot\frac{k}{n}(\ln n - \ln k) &= \frac{\ln n}{n^2}\sum_{k=1}^{n} k - \frac{1}{n^2}\sum_{k=1}^{n} k\ln k \\ &= \frac{\ln n}{n^2}\cdot\frac{n(n+1)}{2} - \frac{1}{n^2}\sum_{k=1}^{n} k\ln k \\ &= \frac{1}{2}\left(1+\frac{1}{n}\right)\ln n - \frac{1}{n^2}\sum_{k=1}^{n} k\ln k. \end{aligned}$$

第一项已经与我们想要的式子吻合。

？ 想一想 那些阶乘是从哪里来的？

对于 $0 \le m \le n-1$，

$$\ln\frac{n!}{m!} = \ln\big((m+1)(m+2)\cdots n\big) = \sum_{k=m+1}^{n} \ln k.$$

现在把 $m = 0, 1, \ldots, n-1$ 的这些式子加起来，并数一数每个 $\ln k$ 出现了多少次。$\ln k$ 这一项恰好在 $m+1 \le k$ 时出现，即 $m = 0, 1, \ldots, k-1$ 的时候：一共 k 次。因此

$$\ln\frac{n!}{0!} + \ln\frac{n!}{1!} + \cdots + \ln\frac{n!}{(n-1)!} = \sum_{k=1}^{n} k\ln k.$$

（例如取 $n=3$：$\ln 6+\ln 6+\ln 3=\ln 1+2\ln 2+3\ln 3$，两边都等于 $\ln 108$。）所以题目中的式子正正就是 S_n，即 n 个长方形的面积。

极限

当 $n\to\infty$ 时，长方形与曲线贴合得越来越紧密，而由于 f 在 $[0,1]$ 上连续，长方形的总面积趋向曲线下的面积。因此

$$\frac{1}{2}\left(1+\frac{1}{n}\right)\ln n-\frac{1}{n^2}\left[\ln\frac{n!}{0!}+\ln\frac{n!}{1!}+\cdots+\ln\frac{n!}{(n-1)!}\right]=S_n\to\int_0^1 x\ln(1/x)\,\mathrm{d}x=\frac{1}{4},$$

这正是所要证明的结果。

如果你想要比“长方形贴合得越来越紧密”更扎实的说法，这里有一个上界。在每一条长条上，长方形的面积与曲线下的面积相差至多为 $\frac{1}{n}$ 乘以 f 在该长条上的变化幅度。把所有长条上的变化幅度加起来：f 上升了 $\frac{1}{\mathrm{e}}$，又回落了 $\frac{1}{\mathrm{e}}$，所以总和至多为 $\frac{2}{\mathrm{e}}$。因此 $\left|S_n-\frac{1}{4}\right|\le\frac{2}{\mathrm{e}n}\to 0$。

验算

- 对第一部分的答案求导：$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}x^2\ln x-\frac{1}{4}x^2\right)=x\ln x+\frac{1}{2}x-\frac{1}{2}x=x\ln x$。
- 取 $n=2$ 时，式子的值是 $\frac{3}{4}\ln 2-\frac{1}{4}(\ln 2+\ln 2)=\frac{1}{4}\ln 2\approx 0.173$，这正是两个长方形的面积 $\frac{1}{2}\cdot\frac{1}{2}\ln 2+\frac{1}{2}\cdot 0$。
- 数值上，式子的值在 $n=10$ 时约为 0.24559，在 $n=100$ 时约为 0.249937，在 $n=1000$ 时约为 0.2499992，逐步逼近 $\frac{1}{4}$。

常见错误

- 弄丢正负号：被积函数是 $x\ln(1/x)=-x\ln x$，所以第一部分求得的原函数（antiderivative）必须取负，面积是 $+\frac{1}{4}$，而不是 $-\frac{1}{4}$。
- 把 $\frac{1}{2}n(n+1)$ 误写成 $\sum k=\frac{1}{2}n^2$：这样括号内的 $\frac{1}{n}$ 就会不见了。
- 忘记这个积分在 0 处是瑕积分（improper integral）：处理下限时需要用到 $x\ln x\to 0$ 这个事实，而题目已经给出。

STEP 之外

这个长方形面积之和是一个黎曼和（Riemann sum）；对于闭区间上的每一个连续函数，黎曼和都收敛到积分，这是大学分析课程中最早遇到的重要定理之一。我们所用的上界，

考虑的是 f 升降的总幅度，这正是有界变差（bounded variation）函数的想法。反过来看，这道题也提供了关于 $\sum k \ln k$ 的信息；这个量与超阶乘（hyperfactorial）$1^1 2^2 \cdots n^n$ 密切相关，而后者的增长速度可以用类似 $n!$ 的斯特灵公式（Stirling's formula）的思路更精确地求出。Abbott 的 *Understanding Analysis* 仔细地建立了黎曼积分的理论。

详细解答

STEP I 1989，第 3 题

纯数学 · 向量与纯量积

> **STEP I 1989，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/89-s1-q3

关键思路

以 O 为原点（origin），把一切都用两个向量（vector）$\mathbf{a}$ 和 $\mathbf{b}$ 表示。这样，“垂直（perpendicular）”就变成“纯量积（scalar product）为零”，而由于 $\mathbf{a}\cdot\mathbf{a}=\mathbf{b}\cdot\mathbf{b}=1$，这个纯量积只取决于 x、OY 和唯一的未知量 $\mathbf{a}\cdot\mathbf{b}$。利用 $OX+OY=1$，便只剩下一条联系 x 与 $\mathbf{a}\cdot\mathbf{b}$ 的方程。至于上下界，则取决于 x 可以在多大的范围内变动。

解答

条件

取 O 为原点，并记 $k=\mathbf{a}\cdot\mathbf{b}$。由于 C 位于由 A 到 B 的三分之一处，而 D 位于三分之二处，

$$\overrightarrow{OC}=\mathbf{a}+\tfrac{1}{3}(\mathbf{b}-\mathbf{a})=\tfrac{1}{3}(2\mathbf{a}+\mathbf{b}),\qquad \overrightarrow{OD}=\tfrac{1}{3}(\mathbf{a}+2\mathbf{b}).$$

由于 X 在 OA 上，且 $OX=x$ 及 $|\mathbf{a}|=1$，故有 $\overrightarrow{OX}=x\mathbf{a}$。同理 $\overrightarrow{OY}=y\mathbf{b}$，其中 $y=OY=1-x$。所以

$$3\overrightarrow{CY}=(3y-1)\mathbf{b}-2\mathbf{a},\qquad 3\overrightarrow{DX}=(3x-1)\mathbf{a}-2\mathbf{b}.$$

两条直线垂直，当且仅当纯量积为零。展开并利用 $\mathbf{a}\cdot\mathbf{a}=\mathbf{b}\cdot\mathbf{b}=1$，得

$$\begin{aligned}9\,\overrightarrow{CY}\cdot\overrightarrow{DX}&=(3x-1)(3y-1)k-2(3y-1)-2(3x-1)+4k\\&=k\big[(3x-1)(3y-1)+4\big]-2(3x+3y-2).\end{aligned}$$

现在利用 $x+y=1$。于是 $3x+3y-2=1$，并且

$$(3x-1)(3y-1)=9xy-3(x+y)+1=9xy-2.$$

所以垂直的条件变成 $k(9xy+2)=2$。代入 $y=1-x$，即

$$\boxed{\mathbf{a}\cdot\mathbf{b}=\frac{2}{9x(1-x)+2}}$$

（分母至少为 2，所以除以它并无问题）。

上下界

X 位于线段 OA 上，而该线段的长度为 1，所以 $0\le x\le 1$。（这时 $y=1-x$ 也在 $[0,1]$ 之内，所以 Y 自然在 OB 上。）在这个区间上

$$x(1-x)=\tfrac{1}{4}-\left(x-\tfrac{1}{2}\right)^2,$$

所以 $0\le x(1-x)\le\frac{1}{4}$，其中极大值在 $x=\frac{1}{2}$ 取得，极小值在 $x=0$ 或 1 取得。因此

$$2\le 9x(1-x)+2\le\tfrac{9}{4}+2=\tfrac{17}{4}.$$

取倒数会把不等式（inequality）的方向倒转（各项皆为正数），再乘以 2 便得

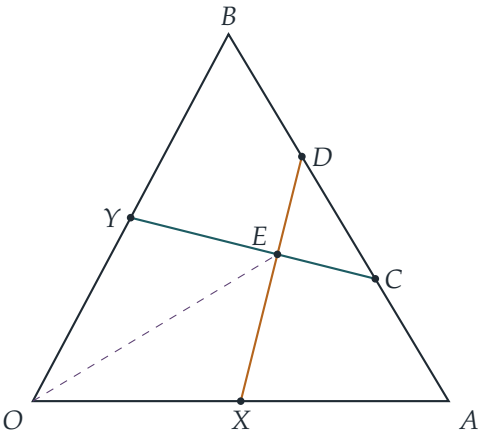
$$\frac{8}{17}\le\mathbf{a}\cdot\mathbf{b}\le 1,$$

正如所求。（取值 1 需要 $\mathbf{a}=\mathbf{b}$，这并不构成真正的三角形，所以实际上 $\mathbf{a}\cdot\mathbf{b}<1$；题目所述的不等式仍然成立。）

最大的角与 OE

? **想一想** 为甚么最大的角对应的是 $\mathbf{a}\cdot\mathbf{b}$ 的最小值?

由于 $|\mathbf{a}|=|\mathbf{b}|=1$，$\mathbf{a}\cdot\mathbf{b}=\cos A\hat{O}B$，而余弦在 $[0,\pi]$ 上是递减的。所以当 $\mathbf{a}\cdot\mathbf{b}=\frac{8}{17}$ 时，即当 $x=\frac{1}{2}$（因而 $y=\frac{1}{2}$）时，这个角最大。这时 X 和 Y 分别是 OA 和 OB 的中点，而这个角为 $\arccos\frac{8}{17}\approx 61.9°$。



现在，把 $\mathbf{a}$ 和 $\mathbf{b}$ 对调，整个图形是对称的：这样对调会交换 $A \leftrightarrow B$、$C \leftrightarrow D$ 和 $X \leftrightarrow Y$，因而交换直线 CY 和 DX。所以它们的交点 E 在对调之下保持不变，亦即它位于对称轴（line of symmetry）上：对某个 t 有 $\overrightarrow{OE} = t(\mathbf{a}+\mathbf{b})$。

DX 上的一般点为

$$\overrightarrow{OX} + s\,\overrightarrow{XD} = \tfrac{1}{2}\mathbf{a} + s\left(\tfrac{1}{3}\mathbf{a} + \tfrac{2}{3}\mathbf{b} - \tfrac{1}{2}\mathbf{a}\right) = \left(\tfrac{1}{2} - \tfrac{s}{6}\right)\mathbf{a} + \tfrac{2s}{3}\mathbf{b}.$$

由于 $\mathbf{a}$ 和 $\mathbf{b}$ 不平行，我们可以比较系数：当 $\frac{1}{2} - \frac{s}{6} = \frac{2s}{3}$ 时，这是 $\mathbf{a}+\mathbf{b}$ 的倍数，由此得 $s = \frac{3}{5}$ 及 $t = \frac{2}{5}$。所以 $\overrightarrow{OE} = \frac{2}{5}(\mathbf{a}+\mathbf{b})$。最后

$$|\mathbf{a}+\mathbf{b}|^2 = \mathbf{a}\cdot\mathbf{a} + 2\,\mathbf{a}\cdot\mathbf{b} + \mathbf{b}\cdot\mathbf{b} = 2 + \tfrac{16}{17} = \tfrac{50}{17},$$

所以

$$OE = \frac{2}{5}\sqrt{\frac{50}{17}} = \frac{2}{5}\cdot\frac{5\sqrt{2}}{\sqrt{17}} = \frac{2\sqrt{2}}{\sqrt{17}} = \boxed{\frac{2\sqrt{34}}{17}} \approx 0.686.$$

验算

- 用坐标表示，取 $\mathbf{a} = (1,0)$ 及 $\mathbf{b} = (\frac{8}{17}, \frac{15}{17})$。于是 $X = (\frac{1}{2}, 0)$，$D = (\frac{11}{17}, \frac{10}{17})$，所以 $\overrightarrow{XD} = (\frac{5}{34}, \frac{10}{17})$。另外 $Y = (\frac{4}{17}, \frac{15}{34})$ 及 $C = (\frac{14}{17}, \frac{5}{17})$，所以 $\overrightarrow{CY} = (-\frac{10}{17}, \frac{5}{34})$。它们的纯量积为 $-\frac{50}{578} + \frac{50}{578} = 0$，正如所料。
- $E = \frac{2}{5}(\frac{25}{17}, \frac{15}{17}) = (\frac{10}{17}, \frac{6}{17})$，而 $|OE| = \frac{\sqrt{136}}{17} = \frac{2\sqrt{34}}{17} \approx 0.6860$。这与直接以数值方法求两条直线交点所得的结果吻合。
- $s = \frac{3}{5}$ 这个值在 $[0,1]$ 之内，所以 E 确实位于 X 与 D 之间，在三角形之内。

常见错误

- 忘记了 $\mathbf{a}\cdot\mathbf{a} = 1$（以致在条件中留下 $|\mathbf{a}|^2$ 的项），或者把两个三等分点（trisection points）的位置调转了。
- 以为最大的角对应最大的 $\mathbf{a}\cdot\mathbf{b}$：角愈大，余弦愈小。
- 写成 $\overrightarrow{OX} = x$ 而不是 $x\mathbf{a}$。点 X 是以向量表示的；x 只是它的长度。

STEP 之外

这里的方法，是把每一点写成两个基向量（basis vectors）的组合，并把几何问题化为纯量积；这正是学习内积（inner product）和格拉姆矩阵（Gram matrix）的起点。一组向量的构形，在不计旋转的意义下，完全由它的纯量积表决定，而在这里该表只有 1、1 和

$\mathbf{a} \cdot \mathbf{b}$。这就是整道题目可以归结为一个数的原因。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对这些想法有详细的阐述。

详细解答

STEP I 1989，第 4 题

纯数学 · 拉姆齐理论

> **STEP I 1989，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/89-s1-q4

本题在 *Beyond STEP* 第 9.2 节小型版的拉姆齐定理中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/ramsey.html

关键思路

先把几何放在一旁：题目对各点所设的条件，只是保证任意三点都构成一个真正的三角形，所以这纯粹是一个为连接六点的 15 条边着色的问题。按照提示，先看其中一点。它连出五条边，而颜色只有两种，所以根据鸽巢原理（pigeonhole principle），这些边之中至少有三条同色。这三条边另一端的三点之间的边，要么与最初那一点合成一个单色三角形，要么自己构成一个单色三角形。

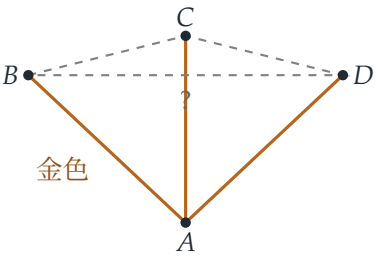
解答

六点

由于任意三点都不共线（collinear），任意三点都是一个真正的三角形的顶点，而它的三条边都在我们所画的线段之中。（任意四点不共面（coplanar）这个条件，在论证中用不着。）

考虑由 A 连出的五条线段 AB, AC, AD, AE, AF。每条线段不是金色就是银色。假如金色的最多两条，银色的也最多两条，线段总数便最多只有四条，这与事实不符。所以其中至少有三条同色。把 $B, \ldots, F$ 各点重新标记，并在有需要时把两种颜色的名称对调，我们可以假设

$$AB,\ AC,\ AD \text{ 全是金色。}$$



现在看看连接另一端三点的三条线段 BC、CD 和 DB。

> **想一想** 无论 BC、CD 和 DB 是甚么颜色，为甚么以下两种情况之中必有一种会给出单色三角形?

只有两种可能。

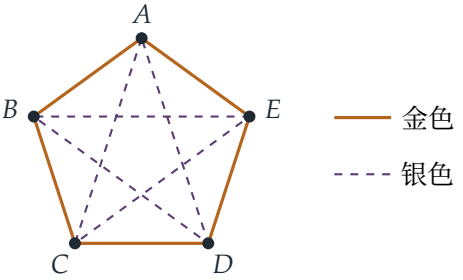
- BC, CD, DB 之中至少有一条是金色。设它是 BC（其余情况只是把字母换一换，道理相同）。那么 AB、AC 和 BC 全是金色，所以三角形 ABC 的边都是同一种颜色。
- BC, CD, DB 之中没有一条是金色。那么三条都是银色，三角形 BCD 的边都是同一种颜色。

无论哪一种情况，都有一个各边同色的三角形，这正是我们要证明的。

五点

若只有五点，每一点只连出四条边，可以分成两条一种颜色、两条另一种颜色，于是上述论证不再成立。要说明结论确实不成立，我们需要找出一种没有单色三角形的着色方法。

把 A, B, C, D, E 五点依次放在一个五边形的顶点上。把五条边 AB, BC, CD, DE, EA 涂成金色，五条对角线 AC, BD, CE, DA, EB（即五角星）涂成银色。



任何三角形都用上五点之中的三点。在五边形的任意三个顶点之中，至少有两个并不相邻（若三个顶点两两相邻，5-圈之内便要有一个三角形），所以每个三角形都有一条银色的边。此外，任意三个顶点之中也至少有两个是相邻的：假如三个顶点两两不相邻，那么

沿五边形走一圈，每两个相继的选定顶点之间都至少隔着一个未被选中的顶点，这就需要 $3+3=6$ 点。所以每个三角形也有一条金色的边。因此没有任何三角形是单色的。

> **想一想** 正五边形是平的，所以它有四个顶点共面。这不是违反了关于一般位置的规定吗？

假如各点真的在同一平面上，那确实违反了规定；但着色只取决于哪两点相连，而与各点的位置无关。把各顶点稍稍提离平面，而且各自升高不同的幅度（例如把第 k 个顶点升到高度 $0.3k^2$，其中 $k=0,1,2,3,4$）。这样便没有三点共线，也没有四点共面，而从上方看下去，图形仍然是这里所画的五边形和五角星，着色相同，也没有单色三角形。

验算

六点的边只有 $2^{15}=32768$ 种着色方法，所以可以用电脑逐一检验：每一种都有单色三角形，而且事实上总有至少两个。至于五点，把五边形着色中全部 $\binom{5}{3}=10$ 个三角形逐一检查，可以确认每个都有两种颜色。再用行列式（determinant）检验，可以确认上面升高了的五边形没有四点共面。

常见错误

- 只说“由对称性”而没有交代假设了甚么：要明确说明，你是从 A 连出的边之中选出占多数颜色的三条，并把那种颜色称为金色。
- 只证明了 A 有三条同色的边便停下来；第二步，即考察 BC、CD、DB，才是这个证明的核心。
- 处理五点时，只画一个草图而没有检查每一个三角形；像“由 A 连出的星形涂金色，其余全涂银色”这样的着色是行不通的。

STEP 之外

这是拉姆齐理论（Ramsey theory）的第一个例子：拉姆齐数（Ramsey number）$R(3,3)$ 是使得完全图（complete graph）K_n 的边的每一种双色着色都含有单色三角形的最小的 n，而本题正好证明了 $R(3,3)=6$。拉姆齐定理（Ramsey's theorem）指出，无论要求的单色完全子图有多大，这样的数都存在，但它们极难计算：就连 $R(5,5)$ 也仍然未知。Wilson 的 *Introduction to Graph Theory*（图论入门）是一本很出色的图论（graph theory）入门书，当中也讲到拉姆齐定理。

详细解答

STEP I 1989，第 5 题

纯数学 · 二项式系数恒等式

STEP I 1989，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/89-s1-q5

关键思路

二项展开式（binomial expansion）是关于 x 的恒等式（identity），所以无论代入甚么数都成立。取 $x=1$ 会把所有系数（coefficient）加起来，取 $x=-1$ 则以正负相间的符号把它们相加；把两者结合，就能把偶数项与奇数项的系数分开。至于 (ii) 中的加权和，$\binom{n}{r}$ 前面的因子 r 正好是对 x^r 求导时产生的，所以我们先对展开式求导（differentiate），然后再代入。

解答

对正整数 n，

$$(1+x)^n=\sum_{r=0}^{n}\binom{n}{r}x^r=\binom{n}{0}+\binom{n}{1}x+\binom{n}{2}x^2+\cdots+\binom{n}{n}x^n,$$

其中 $\binom{n}{r}=\dfrac{n!}{r!\,(n-r)!}$。这对所有实数 x 都成立。

第 (i) 部分

设 $E=\binom{n}{0}+\binom{n}{2}+\cdots+\binom{n}{n}$ 为偶数项系数之和，$O=\binom{n}{1}+\binom{n}{3}+\cdots+\binom{n}{n-1}$ 为奇数项系数之和。由于 n 是偶数，最后一项 $\binom{n}{n}$ 属于 E，而 $\binom{n}{n-1}$ 属于 O，所以 E 与 O 合起来恰好包含每个系数一次。

在展开式中代入 $x=1$，得

$$E+O=2^n.$$

代入 $x=-1$，每个偶数次幂 $(-1)^r$ 都是 $+1$，每个奇数次幂都是 -1，而由于 $n\geq 1$，左

方是 $0^n = 0$。所以

$$E - O = 0.$$

因此 $E = O$，于是 $2E = 2^n$，故

$$E = O = 2^{n-1},$$

这就是所求的结果。

第 (ii) 部分

把展开式两边对 x 求导。常数项消失，而 x^r 变成 rx^{r-1}：

$$n(1+x)^{n-1} = \binom{n}{1} + 2\binom{n}{2}x + 3\binom{n}{3}x^2 + \cdots + n\binom{n}{n}x^{n-1}.$$

现在代入 $x = 1$：

$$n2^{n-1} = \binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n},$$

正如所求。这对所有正整数 n 都成立，不论奇偶。

想一想 有没有不用微积分（calculus）也能看出第 (ii) 部分的方法?

有。由于对 $1 \le r \le n$ 有 $r\binom{n}{r} = r \cdot \dfrac{n!}{r!\,(n-r)!} = n \cdot \dfrac{(n-1)!}{(r-1)!\,(n-r)!} = n\binom{n-1}{r-1}$，在 $(1+x)^{n-1}$ 的展开式中代入 $x = 1$，便知这个和等于 $n\sum_{r=1}^{n}\binom{n-1}{r-1} = n \cdot 2^{n-1}$。用文字来说：从 n 个人中选出一个人数不限的委员会，再从委员会中选出一位主席，这等同于先选主席（n 种方法），再从其余 $n-1$ 个人中任选一个子集（subset）（2^{n-1} 种方法）。

最后的和

把这个和拆成两部分：

$$\sum_{r=0}^{n}\left(r + (-1)^r\right)\binom{n}{r} = \sum_{r=0}^{n} r\binom{n}{r} + \sum_{r=0}^{n}(-1)^r\binom{n}{r}.$$

第一个和中 $r = 0$ 的一项是零，所以第一个和正是 (ii) 中的和，即 $n2^{n-1}$。第二个和是 $(1+x)^n$ 的展开式在 $x = -1$ 时的值，即对所有正整数 n 都有 $(1-1)^n = 0$。（这就是 (i) 中 $E - O = 0$ 的那一步；它只需要 $n \ge 1$，不需要 n 是偶数。）因此

$$\boxed{\sum_{r=0}^{n}\left(r + (-1)^r\right)\binom{n}{r} = n2^{n-1}.}$$

验算

- 当 $n = 2$：系数为 $1, 2, 1$，所以 $E = 1 + 1 = 2 = O$，而 $2^1 = 2$。(ii) 中的和是 $2 + 2 \cdot 1 = 4 = 2 \cdot 2^1$。
- 当 $n = 3$，最后的和是 $(0+1)\cdot 1 + (1-1)\cdot 3 + (2+1)\cdot 3 + (3-1)\cdot 1 = 1 + 0 + 9 + 2 = 12 = 3 \cdot 2^2$。
- 我们对 1 至 29 的每个 n 都精确地验证了全部三个恒等式；例如 $n = 6$ 给出 $192 = 6 \cdot 2^5$。

常见错误

- 忘记了 $\sum (-1)^r \binom{n}{r} = 0$ 对奇数 n 同样成立，因而以为最后的答案取决于 n 的奇偶性（parity）。它只在 $n = 0$ 时才不成立，而这里 n 是正数。
- 在 (ii) 中先代入 $x = 1$ 才求导。代入必须放在最后：一旦 x 成了一个数，就再没有甚么可以求导了。

STEP 之外

把多项式（polynomial）看成储存系数的仓库，然后通过代入、求导或相乘来提取信息，这就是*生成函数*（generating function）的想法，它是组合数学（combinatorics）和概率论中最强大的工具之一。若代入的不是 ± 1 而是复数单位根（roots of unity），便能以同样的方式抽出每隔三项或每隔四项的系数；而对概率生成函数（probability generating function）求导，就得到分布的平均值（mean），正如在这里它给出了 $\sum r\binom{n}{r}$ 一样。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）展示了生成函数的实际应用。

详细解答

STEP I 1989，第 6 题

纯数学 · 法线与微分方程

STEP I 1989，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/89-s1-q6

关键思路

在 P 点的法线（normal）斜率（gradient）为 $-1/\mathrm{f}'(x)$，所以要求 Q，只须把 $X=0$ 代入法线方程。于是 PQ^2 就是水平距离的平方加上垂直距离的平方，而题目给出的长度便化成一条只含 $\mathrm{f}'(x)$ 的方程。唯一要留神的是最后开平方的一步：正负号的取舍由极小点（minimum point）决定。

解答

Q 的 y 坐标

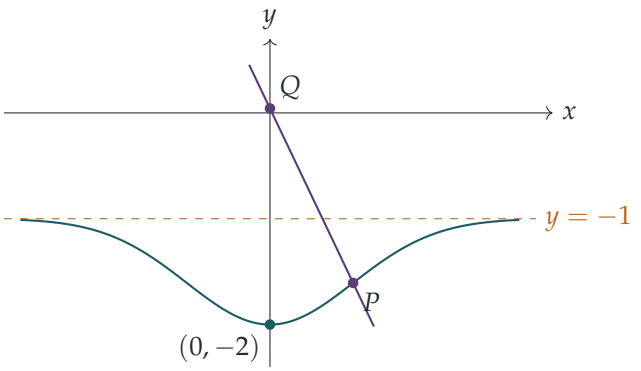
在 $P=(x,\mathrm{f}(x))$ 处，切线（tangent）的斜率为 $\mathrm{f}'(x)$，所以（只要 $\mathrm{f}'(x)\neq 0$）法线的斜率为 $-1/\mathrm{f}'(x)$。以 (X,Y) 表示法线上动点的坐标，法线方程为

$$Y-\mathrm{f}(x)=-\frac{1}{\mathrm{f}'(x)}\,(X-x).$$

代入 $X=0$，法线与 y 轴相交于

$$Y=\mathrm{f}(x)+\frac{x}{\mathrm{f}'(x)},\qquad \text{所以}\qquad Q=\left(0,\ \mathrm{f}(x)+\frac{x}{\mathrm{f}'(x)}\right).$$

（若 $\mathrm{f}'(x)=0$，法线是铅垂线，除非 $x=0$，否则它永不与 y 轴相交。）



微分方程（differential equation）

P 与 Q 的水平距离为 x，垂直距离为 $x/\mathrm{f}'(x)$，所以

$$PQ^2 = x^2 + \frac{x^2}{\big(\mathrm{f}'(x)\big)^2}.$$

令上式等于 $\mathrm{e}^{x^2} + x^2$，两边的 x^2 项互相抵消，得到

$$\frac{x^2}{\big(\mathrm{f}'(x)\big)^2} = \mathrm{e}^{x^2}, \qquad \text{即} \qquad \boxed{\big(\mathrm{f}'(x)\big)^2 = x^2\,\mathrm{e}^{-x^2}.}$$

这等价于 $\mathrm{f}'(x) = \pm x\,\mathrm{e}^{-x^2/2}$。

极小点位于 $(0,-2)$ 的曲线

> ? **想一想** 应该取哪一个正负号？正负号又会否在曲线中途改变？

由于 f′ 连续（continuous）（曲线有极小点，所以我们假设它可微（differentiable）），而 $\mathrm{f}'(x)^2 = x^2\mathrm{e}^{-x^2}$ 只在 $x = 0$ 时为零，因此 f′ 在 $x > 0$ 上或在 $x < 0$ 上都不可能变号：要变号就必须经过零。要使 $(0,-2)$ 成为极小点，f 必须在紧靠 0 的左方递减，在紧靠其右方递增，所以当 $x < 0$ 时 $\mathrm{f}'(x) < 0$，当 $x > 0$ 时 $\mathrm{f}'(x) > 0$。这正好与 x 的正负号相同，因此在两边都有

$$\mathrm{f}'(x) = x\,\mathrm{e}^{-x^2/2}.$$

（若全程取负号，则在 $x = 0$ 处得到极大点；若两边取不同的正负号，则 f′ 的正负号保持不变，即得到拐点（point of inflection）。）

进行积分，由于 $\frac{\mathrm{d}}{\mathrm{d}x}\big(\mathrm{e}^{-x^2/2}\big) = -x\,\mathrm{e}^{-x^2/2}$，

$$\mathrm{f}(x) = -\mathrm{e}^{-x^2/2} + c.$$

由条件 $f(0) = -2$ 得 $-1 + c = -2$，所以 $c = -1$，曲线为

$$\boxed{y = -1 - e^{-x^2/2}.}$$

它在 $x = 0$ 处取得极小值 -2，并在两边向渐近线（asymptote）$y = -1$ 上升，一如略图所示。

验算

- 在 $x = 1$ 处：$f(1) = -1 - e^{-1/2} \approx -1.6065$，而 $f'(1) = e^{-1/2} \approx 0.6065$，所以 $Q \approx (0,\, 0.0422)$，且 $PQ \approx \sqrt{1 + 1.6487^2} \approx 1.9283$，等于 $\sqrt{e+1}$，符合要求。
- 电脑在 2000 个随机点上，以法线与 y 轴的交点重新计算 Q，所得的 PQ 与 $\sqrt{e^{x^2} + x^2}$ 相符，（相对）误差在 10^{-15} 以内。

常见错误

- 误用切线的斜率 $f'(x)$，而不是法线的斜率 $-1/f'(x)$，结果得出 Q 位于 $f(x) - xf'(x)$。
- 开平方时直接写 $f'(x) = xe^{-x^2/2}$，却没有说明为何舍去负号。

STEP 之外

在这一题，一个对曲线上每一点都成立的几何条件，化成了一条一阶微分方程，但方程中的 f' 以平方形式出现，所以要靠额外的资料选定分支，才能确定唯一的解。这类没有就 y' 解出的方程，正是常微分方程（ODE）的唯一性定理（uniqueness theorem）开始需要小心处理的地方，并由此引申到包络线（envelope）和奇解（singular solution）。Boyce 与 DiPrima 的 *Elementary Differential Equations* 涵盖一阶方程及其存在性与唯一性理论。

详细解答

STEP I 1989，第 7 题

纯数学 · 曲线描绘与最优化

STEP I 1989，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/89-s1-q7

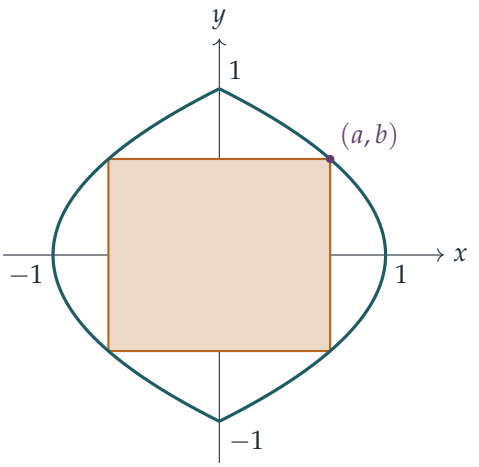
关键思路

这条曲线对两条坐标轴都对称，所以最大的长方形应以原点为中心，并有一个顶点位于第一象限的曲线上。这样一来，原本要在两个变量中搜索的问题，就变成求一个单变量函数的最大值。对于一般的曲线，只要把 $|y|$ 解出来，写成 $1-|x|$ 的幂，同样的想法依然适用。

解答

草图

由于 $y^2 \geq 0$，必须有 $1-|x| \geq 0$，故 $-1 \leq x \leq 1$。当 $x \geq 0$ 时，曲线是 $x = 1-y^2$，即一条开口向左、顶点（vertex）为 $(1,0)$ 的抛物线（parabola）；当 $x \leq 0$ 时，曲线是它的镜像 $x = y^2 - 1$。两段弧在 $(0,\pm 1)$ 相接，并在该处形成尖角（斜率 $\mathrm{d}y/\mathrm{d}x = \mp\frac{1}{2}$ 与 $\pm\frac{1}{2}$ 并不相同），又在 $(\pm 1, 0)$ 垂直地穿过 x 轴。这条曲线是一个闭合的环，围住区域 $|x| + y^2 \leq 1$，形状像一个圆角的菱形。



最大的长方形

设长方形 $[x_1, x_2] \times [y_1, y_2]$ 位于区域 $|x| + y^2 \le 1$ 之内。令 $A = \max(|x_1|, |x_2|)$ 及 $B = \max(|y_1|, |y_2|)$。各顶点为 (x_i, y_j)，其中 i, j 可独立选取，所以必有一个顶点同时满足 $|x| = A$ 和 $|y| = B$。该顶点位于区域之内，故 $A + B^2 \le 1$。于是长方形 $[-A, A] \times [-B, B]$ 也位于区域之内，因为当 $|x| \le A$、$|y| \le B$ 时，

$$|x| + y^2 \le A + B^2 \le 1.$$

它的阔度 $2A \ge x_2 - x_1$，高度 $2B \ge y_2 - y_1$，所以它的面积不会比原来的小。因此我们可以只考虑以原点为中心、顶点为 $(\pm a, \pm b)$ 的长方形；再把顶点向外推，直至碰到曲线为止，便有 $a = 1 - b^2$，其中 $0 \le b \le 1$。

? **想一想** 为甚么只需考虑以原点为中心的长方形?

这正是上面的论证：把每条边换成它到坐标轴的两个距离中较大的一个，长方形仍然留在曲线之内（由对称性可知），而且绝不会缩小。

面积为

$$S(b) = 4ab = 4b(1 - b^2) = 4b - 4b^3, \qquad S'(b) = 4 - 12b^2.$$

$S'(b) = 0$ 于 $b = 1/\sqrt{3}$，在此之前 $S' > 0$，在此之后 $S' < 0$，而且 $S(0) = S(1) = 0$，所以这就是最大值。此时 $a = 1 - \frac{1}{3} = \frac{2}{3}$，而

$$S_{\max} = 4 \cdot \frac{2}{3} \cdot \frac{1}{\sqrt{3}} = \frac{8}{3\sqrt{3}} = \frac{8}{\sqrt{27}},$$

正如所求。

一般的曲线

现在考虑 $y^{2m} = (1-|x|)^n$。左边为非负数，而由于 n 是奇数，$(1-|x|)^n$ 与 $1-|x|$ 同号。所以同样有 $|x| \le 1$，取（正的）$2m$ 次方根，得

$$|y| = (1-|x|)^k, \qquad k = \frac{n}{2m} > 0.$$

（假如 n 是偶数，则任何 x 都可以取，曲线会伸延至无穷远，因而不存在最大的长方形：这就是 n 必须是奇数的原因。）这条曲线同样是一个经过 $(\pm 1, 0)$ 和 $(0, \pm 1)$ 的闭合环，对两条坐标轴都对称，围住区域 $|y| \le (1-|x|)^k$。由于 $(1-|x|)^k$ 随 $|x|$ 增加而递减，与之前相同的论证说明，我们可以取顶点为 $(\pm a, \pm b)$ 的长方形，其中 $b = (1-a)^k$，$0 \le a \le 1$。它的面积为

$$S(a) = 4a(1-a)^k.$$

求导（differentiate），得

$$S'(a) = 4(1-a)^{k-1}\big[(1-a) - ka\big] = 4(1-a)^{k-1}\big[1-(1+k)a\big],$$

当 $0 < a < \frac{1}{1+k}$ 时为正，当 $\frac{1}{1+k} < a < 1$ 时为负。由于 $S(0) = S(1) = 0$，最大值出现于 $a = \frac{1}{1+k}$，此时 $1-a = \frac{k}{1+k}$：

$$S_{\max} = \frac{4}{1+k}\left(\frac{k}{1+k}\right)^k = \frac{4k^k}{(1+k)^{k+1}}.$$

代入 $k = \frac{n}{2m}$ 及 $1+k = \frac{n+2m}{2m}$，$2m$ 的各个幂合并为 $(2m)^{-k} \cdot (2m)^{k+1} = 2m$，所以

$$S_{\max} = \frac{8m\, n^{n/(2m)}}{(n+2m)^{(n+2m)/(2m)}} = 8m\left(\frac{n^n}{(n+2m)^{n+2m}}\right)^{1/(2m)}.$$

验算

- 取 $m = n = 1$，得 $8\left(\frac{1}{27}\right)^{1/2} = 8/\sqrt{27} \approx 1.5396$，与第一部分的结果一致。
- 对 $(m,n) = (1,3), (2,1), (2,3), (3,5)$ 在 $a \in [0,1]$ 上作精细的数值搜索，结果与公式吻合；例如 $(m,n) = (2,1)$ 给出 $16/5^{5/4} \approx 2.1400$。随机选取的偏离中心的长方形，面积从未超过这个值。
- 当 n 固定而 $m \to \infty$ 时，$k \to 0$，曲线逐渐填满正方形 $|x| \le 1$、$|y| \le 1$；公式的值趋向 4，正是该正方形的面积。

常见错误

- 漏掉因子 4：顶点 (a,b) 所对应的长方形，阔度是 $2a$，高度是 $2b$。
- 忽略 n 为奇数的重要性：若 n 是偶数，曲线便不是有界的。

STEP 之外

求 $a(1-a)^k$ 的最大值，其实是一道乔装了的加权算术—几何平均不等式（weighted AM–GM）问题：把 $a \cdot \left(\frac{1-a}{k}\right)^k$ 写成 $1+k$ 个和为常数的因子之积，无须微积分便可求得最大值。曲线 $|x|^p + |y|^q = 1$ 与 ℓ^p 范数（norm）的单位球（unit ball）是近亲，后者会在大学一年级的分析和线性代数中出现。想进一步了解这类不等式，可参阅 Siklos 的 *Advanced Problems in Mathematics*；至于最优化（optimisation）背后的分析，则可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP I 1989，第 8 题

纯数学 · 倍角公式

> **STEP I 1989，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/89-s1-q8

关键思路

棣莫弗定理（De Moivre's theorem）把 $\cos n\theta$ 化为 $(\cos\theta + \mathrm{i}\sin\theta)^n$ 的实部（real part），再把 $\sin^2\theta$ 换成 $1-\cos^2\theta$，便得到一个以 $\cos\theta$ 为变量的多项式（polynomial）。至于最后一部分，那个六次多项式（sextic）看来像是刚求得的两个多项式的某种组合，于是我们着手找出这个组合：原来它正是 $\cos 6\theta - \cos 4\theta$ 的一半，而解 $\cos 6\theta = \cos 4\theta$ 十分容易。

解答

以下一律记 $c = \cos\theta$ 及 $s = \sin\theta$，故 $s^2 = 1 - c^2$。

第 (i) 部分

由棣莫弗定理，$\cos 4\theta + \mathrm{i}\sin 4\theta = (c + \mathrm{i}s)^4$。以二项式定理（binomial theorem）展开，并只保留实数项（即 $\mathrm{i}s$ 的幂次为偶数的各项），

$$\begin{aligned}\cos 4\theta &= c^4 - \tbinom{4}{2}c^2s^2 + s^4 \\ &= c^4 - 6c^2(1-c^2) + (1-c^2)^2 \\ &= c^4 - 6c^2 + 6c^4 + 1 - 2c^2 + c^4 \\ &= 8c^4 - 8c^2 + 1,\end{aligned}$$

这正是所求的恒等式（identity）。

第 (ii) 部分

同样地，取 $(c+\mathrm{i}s)^6$ 的实部，得

$$\cos 6\theta = c^6 - 15c^4s^2 + 15c^2s^4 - s^6,$$

其中用到 $\binom{6}{2} = \binom{6}{4} = 15$ 以及 $\mathrm{i}^2 = -1$、$\mathrm{i}^4 = 1$、$\mathrm{i}^6 = -1$。设 $u = c^2$，则 $s^2 = 1-u$：

$$\begin{aligned}\cos 6\theta &= u^3 - 15u^2(1-u) + 15u(1-u)^2 - (1-u)^3 \\ &= u^3 - 15u^2 + 15u^3 + 15u - 30u^2 + 15u^3 - 1 + 3u - 3u^2 + u^3 \\ &= 32u^3 - 48u^2 + 18u - 1 \\ &= 32c^6 - 48c^4 + 18c^2 - 1,\end{aligned}$$

证毕。

方程

把左方记为 $p(x) = 16x^6 - 28x^4 + 13x^2 - 1$。

> ? **想一想** 这个多项式并不是那两个恒等式中的任何一个，它们又怎能派上用场?

比较系数（coefficient）。首项系数 16 是 32 的一半，常数项 -1 是 -2 的一半，这提示我们把 (ii) 减去 (i)：

$$\cos 6\theta - \cos 4\theta = 32c^6 - 56c^4 + 26c^2 - 2 = 2\,p(c).$$

因此，若 $x = \cos\theta$，则

$$p(x) = 0 \iff \cos 6\theta = \cos 4\theta.$$

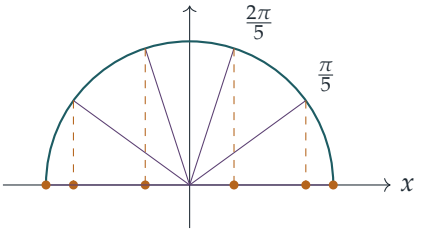
而 $\cos A = \cos B$ 当且仅当存在整数 k 使得 $A = \pm B + 2k\pi$。在这里即是

$$6\theta = 4\theta + 2k\pi \;\Rightarrow\; \theta = k\pi, \qquad \text{或} \qquad 6\theta = -4\theta + 2k\pi \;\Rightarrow\; \theta = \frac{k\pi}{5}.$$

第一组解已包含在第二组之内，所以解为 $\theta = k\pi/5$，而对应的 $x = \cos\theta$ 的值是

$$\cos 0 = 1,\quad \cos\tfrac{\pi}{5},\quad \cos\tfrac{2\pi}{5},\quad \cos\tfrac{3\pi}{5} = -\cos\tfrac{2\pi}{5},\quad \cos\tfrac{4\pi}{5} = -\cos\tfrac{\pi}{5},\quad \cos\pi = -1.$$

k 取其他值时只会重复以上各数。这是六个互不相同的数（余弦在 $[0,\pi]$ 上严格递减），而 p 的次数（degree）为 6，所以它们就是 p 的全部根（root），不论实根还是复根。特别要留意，代换 $x = \cos\theta$ 只能涵盖 $-1 \le x \le 1$，但我们并没有遗漏这个区间以外的根，因为已经再没有其他根可找了。



这些还不算是可接受的答案，因为题目不容许使用数值近似，我们应该给出这些余弦的准确值。由于 $x=1$ 是根，$x=-1$ 也是根（p 是偶函数），所以 x^2-1 是一个因式（factor）。把它除去（或比较系数），

$$p(x)=(x^2-1)(16x^4-12x^2+1).$$

这个四次因式是 x^2 的二次式：

$$x^2=\frac{12\pm\sqrt{144-64}}{32}=\frac{12\pm4\sqrt{5}}{32}=\frac{3\pm\sqrt{5}}{8}.$$

注意

$$\left(\frac{1+\sqrt{5}}{4}\right)^2=\frac{6+2\sqrt{5}}{16}=\frac{3+\sqrt{5}}{8},\qquad\left(\frac{\sqrt{5}-1}{4}\right)^2=\frac{6-2\sqrt{5}}{16}=\frac{3-\sqrt{5}}{8}.$$

x^2 的两个值都是正数，所以各给出两个实数 x。因此方程的实根就是以下六个数：

$$x=\pm1,\qquad x=\pm\frac{1+\sqrt{5}}{4},\qquad x=\pm\frac{\sqrt{5}-1}{4}.$$

额外的收获是：按大小与上面的列表对照（$\cos\frac{\pi}{5}>\cos\frac{2\pi}{5}>0$），便得到两个著名的值 $\cos\frac{\pi}{5}=\frac{1+\sqrt{5}}{4}$ 和 $\cos\frac{2\pi}{5}=\frac{\sqrt{5}-1}{4}$。

验算

- 当 $\theta=0$ 时，两个恒等式都给出 $1=1$；当 $\theta=\frac{\pi}{2}$ 时，(i) 给出 $1=1$，(ii) 给出 $-1=-1$。
- p 的各系数之和为 $16-28+13-1=0$，印证了 $x=1$ 是根。
- 以数值计算，$\frac{1+\sqrt{5}}{4}\approx0.80902=\cos36^\circ$ 及 $\frac{\sqrt{5}-1}{4}\approx0.30902=\cos72^\circ$，而 p 在这两点的值都是零（在舍入误差范围内）。

常见错误

- 遗漏了由 $6\theta=-4\theta+2k\pi$ 得出的解，结果只找到 $x=\pm1$。
- 把答案写成 $\cos\frac{\pi}{5}$ 之类，或写成小数：题目要求的是准确的根式（surd）形式。
- 没有解释为何不存在 $|x|>1$ 的根：数一数次数便可解决。

STEP 之外

把 $\cos n\theta$ 用 $\cos\theta$ 表示的多项式，就是切比雪夫多项式（Chebyshev polynomial）T_n，所以 (i) 和 (ii) 说的是 $T_4(x) = 8x^4 - 8x^2 + 1$ 和 $T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$。切比雪夫多项式在数值分析（numerical analysis）中占有核心地位，它们的根正是多项式插值（polynomial interpolation）的最佳取点。$\cos\frac{\pi}{5}$ 中出现 $\sqrt{5}$，反映了正五边形的几何与黄金比（golden ratio）。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是从几何角度认识棣莫弗定理和单位根（roots of unity）的绝佳读物。

详细解答

STEP I 1989，第 9 题

纯数学 · 图像的变换

> **STEP I 1989，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/89-s1-q9

关键思路

只要画出基本的三次曲线 (cubic)，其余每一个图像都只是同一条曲线经坐标变换后的样子。代换(substitution) $X = ax + b, Y = cy + d$ 不过是在每个方向上先作伸缩(stretch)(或反射（reflection)），再作平移（translation），而且它会把转向点（turning point）变到转向点。因此，全部工夫就在于找出原来三次曲线的两个转向点，并追踪它们去了哪里。

解答

基本图像

写成 $y = \frac{1}{8}(x^3 - 12x) = \frac{1}{8}x(x^2 - 12)$。于是

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{3x^2 - 12}{8} = \frac{3(x-2)(x+2)}{8}, \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{6x}{8} = \frac{3x}{4}.$$

斜率（gradient）在 $x = \pm 2$ 处为零。

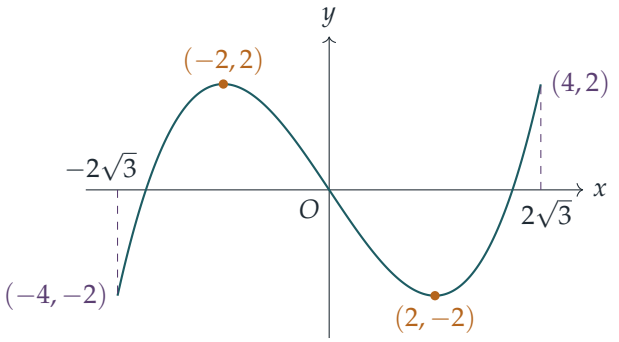
- 在 $x = -2$: $y = \frac{1}{8}(-8 + 24) = 2$, 而 $y'' = -\frac{3}{2} < 0$, 所以 $(-2, 2)$ 是局部极大点 (local maximum)。
- 在 $x = 2$: $y = \frac{1}{8}(8 - 24) = -2$, 而 $y'' = \frac{3}{2} > 0$, 所以 $(2, -2)$ 是局部极小点（local minimum)。

其他有用的特征：

- 曲线与坐标轴相交于 $x(x^2 - 12) = 0$ 之处，即 $x = 0$ 及 $x = \pm 2\sqrt{3} \approx \pm 3.46$。
- 这个函数是奇函数（odd function），所以图像绕原点有 2 阶旋转对称（rotational

symmetry），而原点亦是拐点（point of inflection）；该处的斜率为 $-\frac{3}{2}$。

- 在两端，$y(4) = \frac{1}{8}(64 - 48) = 2$ 及 $y(-4) = -2$。所以曲线恰好在 $x = 4$ 处回升到极大点的高度。



为何转向点会变到转向点

想一想 当我们变换坐标时，怎样知道新的转向点就是旧转向点的像？

若 $X = ax + b$ 及 $Y = cy + d$，其中 $a, c \neq 0$，则由链式法则（chain rule）可得

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \frac{\mathrm{d}Y/\mathrm{d}x}{\mathrm{d}X/\mathrm{d}x} = \frac{c}{a}\frac{\mathrm{d}y}{\mathrm{d}x},$$

它恰好在 $\mathrm{d}y/\mathrm{d}x$ 为零之处为零。所以新的转向点就是 $(-2, 2)$ 和 $(2, -2)$ 的像。（若 $c/a < 0$，图形会被反射，极大点与极小点的角色互换。）从几何上看：

- $X = \frac{1}{2}x$ 把图像沿水平方向压缩，压缩因子为 2；
- $Y = \frac{1}{2}y$ 把图像沿垂直方向压缩，压缩因子为 2；
- 再把 1 加到 X 或 Y 上，则把图像向右或向上平移 1 个单位。

四个变换后的图像

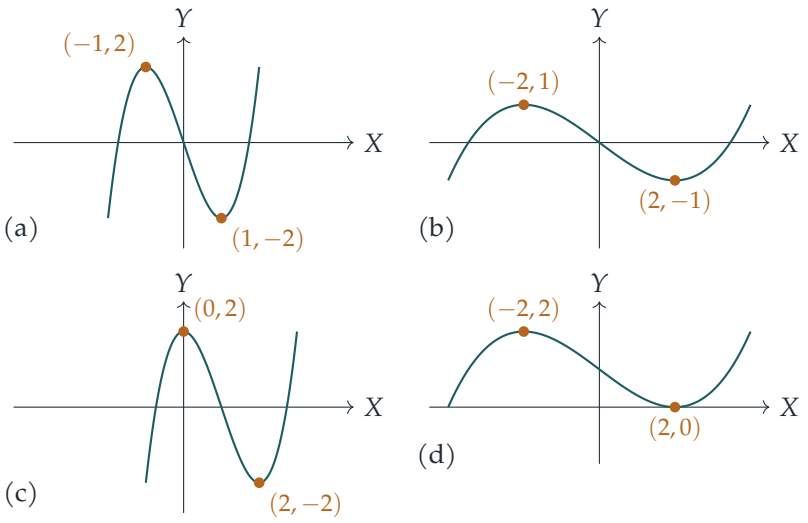
把每个映射分别应用于 $(-2, 2)$、$(2, -2)$ 以及范围 $-4 \leq x \leq 4$：

(a) $X = \frac{1}{2}x$, $Y = y$：极大点 $(-1, 2)$，极小点 $(1, -2)$，范围为 $-2 \leq X \leq 2$。曲线是 $Y = X^3 - 3X$。

(b) $X = x$, $Y = \frac{1}{2}y$：极大点 $(-2, 1)$，极小点 $(2, -1)$，范围为 $-4 \leq X \leq 4$。曲线是 $16Y = X^3 - 12X$。

(c) $X = \frac{1}{2}x + 1$, $Y = y$：极大点 $(0, 2)$，极小点 $(2, -2)$，范围为 $-1 \leq X \leq 3$。这是把图像 (a) 向右移 1 个单位，拐点在 $(1, 0)$。

(d) $X = x, Y = \frac{1}{2}y + 1$：极大点 $(-2,2)$，极小点 $(2,0)$，范围为 $-4 \le X \le 4$。这是把图像 (b) 向上移 1 个单位，拐点在 $(0,1)$；它在极小点处与 X 轴相切。



最后部分

我们需要 $a, c \neq 0$（否则图像会塌缩成一条直线），这样新图像的转向点就是

$$(-2a + b,\ 2c + d) \quad \text{及} \quad (2a + b,\ -2c + d).$$

这两点必须是 $(0,0)$ 和 $(1,1)$，次序不拘，由此得到两组线性方程组。

情况 1：$(-2,2) \mapsto (0,0)$ 及 $(2,-2) \mapsto (1,1)$。于是 $-2a + b = 0$ 及 $2a + b = 1$，所以 $b = \frac{1}{2}, a = \frac{1}{4}$；又 $2c + d = 0, -2c + d = 1$，所以 $d = \frac{1}{2}, c = -\frac{1}{4}$。

情况 2：$(-2,2) \mapsto (1,1)$ 及 $(2,-2) \mapsto (0,0)$。同理可得 $a = -\frac{1}{4}, b = \frac{1}{2}, c = \frac{1}{4}, d = \frac{1}{2}$。

所以其中一个答案是

$$a = \tfrac{1}{4}, \quad b = \tfrac{1}{2}, \quad c = -\tfrac{1}{4}, \quad d = \tfrac{1}{2},$$

另一个答案是 $a = -\frac{1}{4}, b = \frac{1}{2}, c = \frac{1}{4}, d = \frac{1}{2}$。

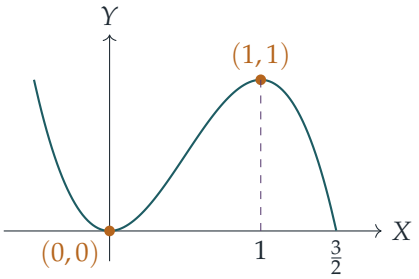
新曲线的方程值得求出来。在情况 1 中，$x = 4X - 2$ 及 $y = 2 - 4Y$，所以

$$\begin{aligned} 8(2 - 4Y) &= (4X - 2)^3 - 12(4X - 2) \\ &= 64X^3 - 96X^2 + 48X - 8 - 48X + 24 \\ &= 64X^3 - 96X^2 + 16, \end{aligned}$$

化简后得

$$Y = 3X^2 - 2X^3 = X^2(3 - 2X).$$

这条曲线有 $Y' = 6X(1-X)$，所以它的转向点确实是位于 $(0,0)$ 的极小点和位于 $(1,1)$ 的极大点。情况 2 给出完全相同的曲线：由于这个三次函数是奇函数，把 (x,y) 换成 $(-x,-y)$ 并不会改变原来的图像，而这正是两种情况之间唯一的分别。



当 $-4 \le x \le 4$ 时，这条曲线由 $(-\frac{1}{2}, 1)$ 延伸至 $(\frac{3}{2}, 0)$。

验算

- 把 $X = 0$ 和 $X = 1$ 代入 $Y = 3X^2 - 2X^3$，分别得 0 和 1，而 $Y' = 6X(1-X)$ 在这两处都等于零。
- 每个变换后的图像都保留了原图像的一个特征：曲线在范围的一端回到某个转向点的高度。例如在 (d) 中，$Y = 2$ 于 $X = 4$，而 $Y = 0$ 于 $X = -4$。
- 数值上，在 $x = 1$ 处我们有 $y = -\frac{11}{8}$；在情况 1 中这给出 $X = 0.75, Y = 0.84375$，而 $3(0.75)^2 - 2(0.75)^3 = 0.84375$。

常见错误

- 弄错伸缩的方向：$X = \frac{1}{2}x$ 令图像变窄，而不是变阔，因为位于 $x = 2$ 的点会移到 $X = 1$。
- 忘记把定义域（domain）也一并变换：图像 (c) 的范围是 $-1 \le X \le 3$，而不是 $-4 \le X \le 4$。
- 最后部分只给出一个解，而没有留意到转向点还有另一种对应次序（给出任何一个已经足够，但第一种情况中的 $c < 0$ 很容易被忽略）。

STEP 之外

映射 $X = ax + b, Y = cy + d$ 是仿射变换（affine transformation），而最后部分说明，任何有两个转向点的三次曲线，都可以用其中一个这样的变换移到标准曲线 $Y = 3X^2 - 2X^3$

上；这就是电脑图像学中所用的“smoothstep”函数。像这样按某个变换群（group of transformations）来为曲线分类，是贯穿几何与线性代数（linear algebra）的一个主题。至于这个故事中与矩阵有关的一面，可参阅 Strang 的 *Introduction to Linear Algebra*（线性代数入门）。

详细解答

STEP I 1989，第 10 题

力学 · 与速度相关的阻力

> **STEP I 1989，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/89-s1-q10

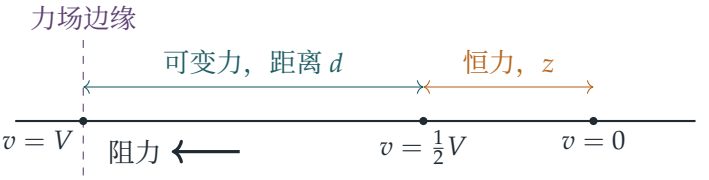
关键思路

阻力取决于速率 v，而非时间，而我们要求的是距离。这正是提示我们把加速度（acceleration）写成 $v\dfrac{\mathrm{d}v}{\mathrm{d}x}$ 的信号：这样一来，牛顿第二定律（Newton's second law）便化为一条直接联系 v 与 x 的可分离变量方程（separable equation）。第二部分是匀减速问题，所以唯一要做的，就是求出转变一刻的力，再让它在 $\pm 10\%$ 的范围内变动。

解答

模型

我们把太空船视为质量为 M、沿直线运动的质点（particle）。设 x 为进入力场后所行的距离，v 为速率，且在 $x = 0$ 时 $v = V$。唯一的力是阻力（resistance），其方向向后，因此速率一直递减。



速率减半所需的距离

取运动方向为正，并把加速度写成 $v\dfrac{\mathrm{d}v}{\mathrm{d}x}$，由牛顿第二定律得

$$M v \frac{\mathrm{d}v}{\mathrm{d}x} = -\frac{M\omega\,(v^2 + V^2)}{v} \qquad \Longrightarrow \qquad \frac{v^2}{v^2 + V^2}\,\mathrm{d}v = -\omega\,\mathrm{d}x.$$

想一想 怎样求 $\dfrac{v^2}{v^2+V^2}$ 的积分?

把分子写成 $(v^2+V^2)-V^2$，于是

$$\frac{v^2}{v^2+V^2}=1-\frac{V^2}{v^2+V^2},\qquad \int\frac{v^2}{v^2+V^2}\,\mathrm{d}v=v-V\arctan\frac{v}{V}+C,$$

这里用了标准积分（standard integral）$\int\frac{\mathrm{d}v}{v^2+V^2}=\frac{1}{V}\arctan\frac{v}{V}$。

设 d 为速率降至 $\frac{1}{2}V$ 时所行的距离。由 $(x,v)=(0,V)$ 积分至 $(d,\frac{1}{2}V)$，

$$\begin{aligned}\omega d&=\int_{V/2}^{V}\frac{v^2}{v^2+V^2}\,\mathrm{d}v=\Big[v-V\arctan\frac{v}{V}\Big]_{V/2}^{V}\\&=\frac{V}{2}-V\Big(\arctan 1-\arctan\tfrac{1}{2}\Big)=V\Big(\frac{1}{2}-\frac{\pi}{4}+\arctan\frac{1}{2}\Big).\end{aligned}$$

因此所求的距离为

$$d=\frac{V}{\omega}\Big(\frac{1}{2}-\frac{\pi}{4}+\arctan\frac{1}{2}\Big)=\frac{V}{\omega}\Big(\frac{1}{2}-\arctan\frac{1}{3}\Big)\approx 0.178\,\frac{V}{\omega}.$$

第二个形式用了 $\arctan 1-\arctan\frac{1}{2}=\arctan\dfrac{1-\frac{1}{2}}{1+\frac{1}{2}}=\arctan\frac{1}{3}$，这来自正切的差角公式（tangent subtraction formula）（两个角都在 $(0,\frac{\pi}{2})$ 之内，所以没有歧义）。两个形式都是完整的答案。

恒力

在转变前的一刻，$v=\frac{1}{2}V$，所以阻力的大小为

$$\frac{M\omega\left(\frac{1}{4}V^2+V^2\right)}{\frac{1}{2}V}=\frac{5}{2}M\omega V.$$

新的恒力 F 与此值相差不超过 10%，所以

$$\frac{9}{10}\cdot\frac{5}{2}M\omega V\le F\le\frac{11}{10}\cdot\frac{5}{2}M\omega V,\qquad \text{即}\qquad \frac{9}{4}M\omega V\le F\le\frac{11}{4}M\omega V.$$

在恒力作用下，减速度（deceleration）$a=F/M$ 是常数，所以由 $v^2=u^2-2as$，取 $u=\frac{1}{2}V$ 及末速率为 0，得

$$z=\frac{(V/2)^2}{2a}=\frac{MV^2}{8F}.$$

这个量随 F 增大而减小，所以最大的力对应最小的 z，反之亦然：

$$\frac{MV^2}{8\cdot\frac{11}{4}M\omega V}\le z\le\frac{MV^2}{8\cdot\frac{9}{4}M\omega V},\qquad \text{亦即}\qquad \boxed{\frac{V}{22\,\omega}\le z\le\frac{V}{18\,\omega}}$$

（若把“相差不超过 10%”理解为严格的范围，则不等式为严格不等式）。假如力完全不变，则 $z=V/(20\omega)$，位于这个范围的中间。

验算

- 量纲（dimensions）：ω 的单位是 s^{-1}（因为 $M\omega v$ 必须是力），所以 V/ω 是长度，正如两个答案所要求的。
- 系数 $\frac{1}{2} - \arctan\frac{1}{3} \approx 0.1782$ 是正数，且小于 $\frac{1}{2}$。这是应该的：太空船的速率介乎 $\frac{1}{2}V$ 与 V 之间，而减速度至少为 $2\omega V$，所以这段行程最多历时 $1/(4\omega)$，所行距离小于 $V/(4\omega)$。
- 取 $V = \omega = 1$，以数值方法逐步计算 $\dot{v} = -\omega(v^2 + V^2)/v$，得出 $v = \frac{1}{2}$ 时的距离为 0.17825；再逐步计算恒力阶段，当力为原来的 110%、100% 及 90% 时，分别得 $z = 0.0455$、0.0500、0.0556，与 $\frac{1}{22}, \frac{1}{20}, \frac{1}{18}$ 吻合。

常见错误

- 用了 $\dfrac{\mathrm{d}v}{\mathrm{d}t}$ 而不是 $v\dfrac{\mathrm{d}v}{\mathrm{d}x}$：这样求得的是时间而不是距离，其后还要多做一次积分。
- 把上下限配反了：力愈大，停下来所需的距离愈短，所以 1.1 倍的力对应的是下限 $V/(22\omega)$。

STEP 之外

写出 $a = v\,\mathrm{d}v/\mathrm{d}x$，是初次窥见一个更宏大的想法：对于只取决于位置或速率的一维力，我们可以用另一个变量取代时间，从而降低方程的阶（order）；而当力只取决于位置时，这就引出能量守恒（energy conservation）以及描述运动的相平面（phase plane）图像。第二部分是一道关于敏感度（sensitivity）的小练习：数据中的不确定性（这里是力的 $\pm10\%$）怎样传递到答案之中（z 大约相差 $\mp10\%$）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）正是以这种精神处理与速度有关的力。

详细解答

STEP I 1989，第 11 题

力学 · 抛体射程的最优化

STEP I 1989，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/89-s1-q11

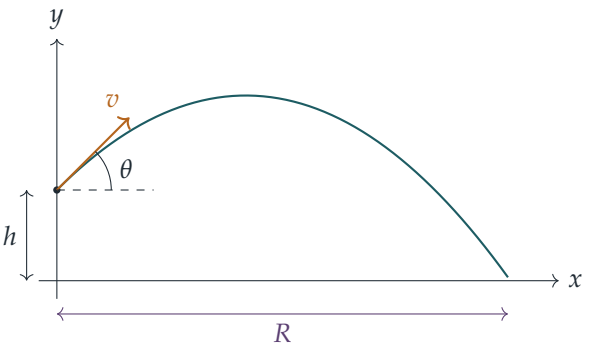
关键思路

飞行时间来自一条关于 t 的二次方程（quadratic equation）：铅球的高度回到零时便落地。取正根，再乘以水平速率，便得到 R；余下的工作只是把平方根整理成题目要求的形式。至于第二部分，“θ 稍为减小，R 会否增大?”其实是问 $\mathrm{d}R/\mathrm{d}\theta$ 在 $\theta = 45^\circ$ 时的正负号，而对称关系 $\sin 45^\circ = \cos 45^\circ$ 令这个导数（derivative）化简得十分俐落。

解答

模型

我们把铅球视为只受重力作用的质点（particle），从点 $(0, h)$ 以速率 v、与水平成仰角 θ 掷出，其中 $0 < \theta < \frac{\pi}{2}$ 及 $h > 0$。取 x 轴为水平方向并指向投掷的方向，y 轴由地面起竖直向上。



射程

经过时间 t 后，

$$x = vt\cos\theta, \qquad y = h + vt\sin\theta - \tfrac{1}{2}gt^2.$$

铅球在 $y = 0$ 时着地，即

$$\tfrac{1}{2}gt^2 - (v\sin\theta)\,t - h = 0 \quad\Longrightarrow\quad t = \frac{v\sin\theta \pm \sqrt{v^2\sin^2\theta + 2gh}}{g}.$$

由于 $h > 0$，两根之积 $-2h/g$ 为负数，所以恰好有一个根是正数；取负号得到的是负根（即掷出之前的时刻），因此我们取正号。于是

$$\begin{aligned} R = v\cos\theta\, t &= \frac{v\cos\theta}{g}\left(v\sin\theta + \sqrt{v^2\sin^2\theta + 2gh}\right) \\ &= \frac{v^2\sin\theta\cos\theta}{g}\left(1 + \sqrt{1 + \frac{2gh}{v^2\sin^2\theta}}\right) \qquad (\text{提出 } v\sin\theta > 0) \\ &= \frac{v^2}{2g}\sin 2\theta\left(1 + \sqrt{1 + \frac{2hg}{v^2\sin^2\theta}}\right), \end{aligned}$$

其中用到 $2\sin\theta\cos\theta = \sin 2\theta$。这就是所要求的结果。

> **想一想** 为甚么可以把 $v\sin\theta$ 以 $v^2\sin^2\theta$ 的形式放进平方根之内?

因为 $\sqrt{a^2 b} = a\sqrt{b}$ 只在 $a \geq 0$ 时成立。这里 $0 < \theta < \frac{\pi}{2}$，所以 $v\sin\theta > 0$，这一步是成立的。（若是向下掷出，这一步便不成立，题目所给的公式也会出错。）

45° 是最佳角度吗?

求导时，用未化简的形式较为方便。写成

$$R(\theta) = \frac{v}{g}\cos\theta\,(v\sin\theta + S), \qquad S = \sqrt{v^2\sin^2\theta + 2gh},$$

于是 $\dfrac{\mathrm{d}S}{\mathrm{d}\theta} = \dfrac{v^2\sin\theta\cos\theta}{S}$。由积法则（product rule），

$$\frac{\mathrm{d}R}{\mathrm{d}\theta} = \frac{v}{g}\left[-\sin\theta\,(v\sin\theta + S) + \cos\theta\left(v\cos\theta + \frac{v^2\sin\theta\cos\theta}{S}\right)\right].$$

在 $\theta = 45^\circ$ 时，设 $\sin\theta = \cos\theta = c = \frac{1}{\sqrt{2}}$。$-vc^2$ 与 $+vc^2$ 两项互相抵销，余下

$$\begin{aligned} \left.\frac{\mathrm{d}R}{\mathrm{d}\theta}\right|_{\theta=45^\circ} &= \frac{vc}{g}\left(-S + \frac{v^2c^2}{S}\right) = \frac{vc}{gS}\left(v^2c^2 - S^2\right) \\ &= \frac{vc}{gS}\left(\tfrac{1}{2}v^2 - \tfrac{1}{2}v^2 - 2gh\right) = -\frac{2vhc}{S} = -\frac{\sqrt{2}\,vh}{\sqrt{v^2 + 4gh}}, \end{aligned}$$

其中最后一步用到 $S=\sqrt{\frac{1}{2}v^2+2gh}=\frac{1}{\sqrt{2}}\sqrt{v^2+4gh}$。

由于 $v>0$ 及 $h>0$，这个导数严格为负。所以在 45° 附近，R 是 θ 的严格递减函数（strictly decreasing function），而当减幅 $\delta\theta>0$ 足够小时，便有 $R(45^\circ-\delta\theta)\approx R(45^\circ)+\frac{\sqrt{2}\,vh}{\sqrt{v^2+4gh}}\,\delta\theta>R(45^\circ)$。**答案是会的：$\theta$ 稍为减小，R 便会增大。**

从物理角度看：铅球从高处掷出，留在空中的时间比在平地上掷出时稍长，而这段额外的时间用来作水平移动更为划算，所以掷得平一点会有好处。只有当 $h=0$ 时，导数才等于零，45° 才是最佳角度。

验算

- 当 $h=0$ 时，公式给出 $R=\frac{v^2}{g}\sin 2\theta$，即熟悉的平地射程，而在 45° 的导数为 0，正如所料。
- 单位：v^2/g 是长度，而 $2hg/v^2$ 是无量纲（dimensionless）的量。
- 数值上，取 $g=9.8$、$v=13\,\mathrm{m\,s^{-1}}$ 及 $h=2\,\mathrm{m}$：$R(45^\circ)\approx 19.05\,\mathrm{m}$，但 $R(40^\circ)\approx 19.10\,\mathrm{m}$。把运动按时间逐步推算，所得结果与 R 的公式吻合；而对许多随机选取的 v 和 h，在 45° 以有限差分（finite difference）计算的导数亦与 $-\sqrt{2}\,vh/\sqrt{v^2+4gh}$ 相符。

常见错误

- 解 t 的二次方程时取错了根，或忘记说明为何舍去负根。
- 推论说“$\sin 2\theta$ 在 45° 时最大，所以 R 也是”。括号内的式子同样取决于 θ（它随 θ 增大而减小），所以这个论证并不完整。

STEP 之外

一般地令 $\mathrm{d}R/\mathrm{d}\theta=0$，可证明从高度 h 掷出的最佳角度满足 $\tan\theta=v/\sqrt{v^2+2gh}$，这个角度略小于 45°（以上述数值计算约为 42°）；现实中的铅球运动员出手角度还要更低，因为角度较平时他们可以推得更用力。像这样对一族轨迹求最优，便会引出这族曲线的包络线（envelope），即“安全抛物线”（parabola of safety），它划出了铅球所能到达的每一点。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教材）对抛体运动（projectile motion）及其各种变化有细致的论述。

详细解答

STEP I 1989，第 12 题

力学 · 弹性绳与能量

> **STEP I 1989，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/89-s1-q12

关键思路

一切都建基于正四面体（regular tetrahedron）的几何：它的质心（centre of mass）与顶点、与面、以及与通过顶点的轴之间的距离各是多少。一旦求得这三个长度，第一部分只不过是虎克定律（Hooke's law）的应用。至于第二部分，对称性确保运动是纯粹的铅垂平移。四面体开始和结束时都处于静止，所以我们根本不需要运动方程（equation of motion）：只要考虑重力势能与弹性势能（elastic potential energy）之间的能量平衡，便可解决问题。

解答

四面体的几何

我们把各杆视为均匀刚杆，把各弦线视为符合虎克定律的轻弹性弦线。每根杆的质量可视为集中在其中点。每个顶点恰好属于六根杆中的三根，所以六个中点的平均等于四个顶点的平均。因此质心 G 就是四面体的形心（centroid）。

面 BCD 是边长为 $2a$ 的等边三角形，所以它的中心 N 与 B、C、D 每一点的距离都是

$$\rho = \frac{2a}{\sqrt{3}}$$

。四面体的高是 $AN = \sqrt{(2a)^2 - \rho^2} = \sqrt{\frac{8}{3}}\,a = \frac{2\sqrt{6}}{3}a$。形心位于 AN 上，在由该面往上四分之一的位置。所以

$$GN = r = \frac{\sqrt{6}}{6}\,a, \qquad GA = GB = GC = GD = R = \frac{\sqrt{6}}{2}\,a.$$

（验算：$R^2 = \rho^2 + r^2 = \frac{4}{3}a^2 + \frac{1}{6}a^2 = \frac{3}{2}a^2$。）

张力

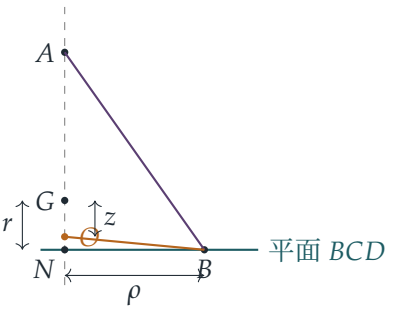
当 X 位于 G 时，每条弦线的长度为 $R = \frac{\sqrt{6}}{2}a$，伸长量（extension）为 $R - a = \frac{\sqrt{6}-2}{2}a > 0$，所以四条弦线全部拉紧。根据虎克定律，每条弦线的张力（tension）为

$$T = \frac{\lambda(R-a)}{a} = \frac{\sqrt{6}-2}{2}\lambda \approx 0.225\,\lambda.$$

XA 断裂之后

断裂之前，四个相等的张力沿着指向 G 的四个方向拉动各顶点。由于 G 是形心，由 G 指向各顶点的单位向量之和为零，所以各弦线不产生净力，桌面承受全部重量 Mg。

当 XA 断裂时，余下三条弦线的合力大小为 T，方向铅垂向上，即等于 XA 原有拉力的负值。在水平方向上，由于绕铅垂线 AN 有三重对称性，三个拉力互相抵销。每个拉力都位于通过该线的铅垂平面内，所以它们不产生扭转力偶（couple），而它们的倾侧力偶亦因对称而互相抵销。因此四面体的运动是纯粹的铅垂平移，AN 始终留在通过 O 的铅垂线上。它之所以离开桌面，是因为 $T > Mg$，这一点我们会在最后确认。



设 z 为 G 高于 O 的高度。那么 B 与通过 O 的铅垂线的水平距离为 ρ，相对于 O 的高度为 $z - r$，所以三条弦线每条的长度都是

$$L(z) = \sqrt{\rho^2 + (z-r)^2}.$$

开始时 $z = 0$，而 $L = \sqrt{\rho^2 + r^2} = R$。在最高点，$BCD$ 与 O 同高，所以 $z = r$，而 $L = \rho = \frac{2}{\sqrt{3}}a$。当 $0 \le z \le r$ 时，L 一直介乎 ρ 与 R 之间，两者都大于 a，所以弦线自始至终保持拉紧。

想一想 怎样才可以不必解一条棘手的非线性运动方程，而求出最大高度?

桌面光滑，弦线为轻弦线，所以机械能（mechanical energy）守恒。由于四面体立即离开桌面，法向反作用力（normal reaction）不作功。四面体在开始时和在最高点时都处于静止，所以动能不必考虑。三条弦线所失去的弹性能量 $3 \cdot \frac{\lambda(L-a)^2}{2a}$，必定等于所获得的

重力势能 Mgr：

$$Mg\,r = \frac{3\lambda}{2a}\Big[(R-a)^2-(\rho-a)^2\Big].$$

把方括号内的式子因式分解，并利用 $R^2-\rho^2 = \frac{3}{2}a^2-\frac{4}{3}a^2 = \frac{1}{6}a^2$，

$$\begin{aligned}(R-a)^2-(\rho-a)^2 &= R^2-\rho^2-2a(R-\rho)\\ &= a^2\Big(\tfrac{1}{6}-\sqrt{6}+\tfrac{4}{\sqrt{3}}\Big).\end{aligned}$$

由于 $\dfrac{3}{2ar} = \dfrac{3}{2a}\cdot\dfrac{6}{\sqrt{6}\,a} = \dfrac{3\sqrt{6}}{2a^2}$，由此可得

$$\begin{aligned}\frac{Mg}{\lambda} &= \frac{3\sqrt{6}}{2}\Big(\tfrac{1}{6}-\sqrt{6}+\tfrac{4}{\sqrt{3}}\Big) = \frac{\sqrt{6}}{4}-9+6\sqrt{2}\\ &\approx 0.6124-9+8.4853 = 0.0977,\end{aligned}$$

准确至 2 位有效数字即为 0.098，正如题目所求。

最后，$Mg \approx 0.098\lambda < 0.225\lambda \approx T$，所以刚断裂后的向上净力 $T-Mg$ 为正数，四面体确实会升离桌面，与题目所述一致。

验算

- 取 $a=1$，我们以坐标表示四个顶点，确认全部六条棱的长度都是 2，四个顶点与形心的距离都是 $\frac{\sqrt{6}}{2}\approx 1.2247$，并重新得出 $T \approx 0.2247\lambda$。
- 取 $Mg/\lambda = 6\sqrt{2}+\frac{\sqrt{6}}{4}-9 \approx 0.09765$，我们由 $z=0$ 处的静止状态出发，以数值方法逐步计算运动方程 $M\ddot{z} = 3T(L)\,\frac{r-z}{L} - Mg$。速度首次回到零时 $z \approx 0.40825$，与 $r=\frac{\sqrt{6}}{6}$ 吻合。

常见错误

- 把 R 当作形心到面的距离。形心位于由顶点往下 $\frac{3}{4}$ 的位置，所以 $R = 3r$。
- 忘记弦线在最高点时仍然处于拉伸状态。每条弦线在该处的能量是 $\frac{\lambda(\rho-a)^2}{2a}$，而不是零。
- 计算伸长量 $L-a$ 时，把 L 取作由 O 到面的中心 N 的距离，而不是由 O 到顶点的距离。

STEP 之外

这里所用的形心技巧（六根杆的中点的平均等于四个顶点的平均）是以下一般事实的一个小例子：一个系统的质心，是其各部分质心按质量加权的平均。把运动视为纯粹平移，靠的是对称性论证；这类论证在刚体动力学（rigid-body dynamics）中会变得有系统，因为物体的运动可分解为质心的平移和绕质心的转动。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对这两个想法都有细致的阐述。

详细解答

STEP I 1989，第 13 题

力学 · 涉及摩擦力的梯子平衡

STEP I 1989，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：

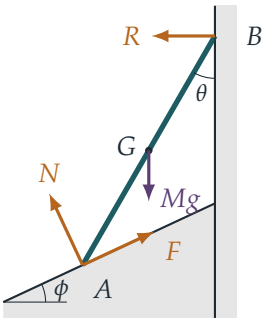
step.maths.org/questions/89-s1-q13

关键思路

这是一道三个力再加摩擦力的平衡（equilibrium）问题，窍门在于选对方向。以梯脚为支点取力矩（moments），可同时消去 N 和 F，立刻得出墙的反作用力。接着，我们不沿水平和铅垂方向分解，而是*沿垂直于斜面和平行于斜面的方向*分解。这样得出的两条方程各自只含未知量 N 和 F 的其中一个，题目给出的公式便直接得出。

解答

我们把梯子模型化为一根长度为 $2a$ 的均匀杆（uniform rod）AB，其质心（centre of mass）G 位于中点。梯脚 A 置于粗糙斜面上，梯顶 B 靠在光滑的墙上，因此墙的反作用力 R 沿水平方向，指向远离墙的一方。在 A 点有垂直于斜面的法向反作用力（normal reaction）N 和沿斜面的摩擦力（friction）F。梯子若要滑动，梯脚会沿斜面向下滑，远离墙壁，所以我们取 F 的方向为沿斜面向上、指向墙壁。（稍后求得的 F 的公式为正数，印证了这个选择。）



墙的反作用力

以 A 为支点取力矩，可消去 N 和 F。梯子与铅垂线的夹角为 θ，所以 B 比 A 高 $2a\cos\theta$，而 G 与 A 的水平距离为 $a\sin\theta$。作用于 B 的水平力 R 的力矩为 $R\cdot 2a\cos\theta$，重量的力矩为 $Mg\cdot a\sin\theta$，转向相反，因此

$$2aR\cos\theta = Mga\sin\theta \qquad \Longrightarrow \qquad R = \tfrac{1}{2}Mg\tan\theta.$$

沿垂直于斜面和平行于斜面的方向分解

想一想 应该沿哪些方向分解，才能令每条方程只含一个未知量?

沿垂直于斜面的方向和沿斜面的方向分解（resolve）。F 在垂直于斜面的方向没有分量，N 在沿斜面的方向没有分量。我们需要求出另外两个力的分量。

- 重量 Mg 有压向斜面的分量 $Mg\cos\phi$ 和沿斜面向下的分量 $Mg\sin\phi$。
- 斜面向着墙壁上升，所以它向上的法线朝远离墙的一方倾斜，而水平力 R（同样指向远离墙的一方）沿该法线的分量为 $R\sin\phi$，即指向斜面之外。沿斜面方向，R 指向远离墙的一方，亦即下坡方向，分量为 $R\cos\phi$。

垂直于斜面：

$$N + R\sin\phi = Mg\cos\phi,$$
$$\text{所以}\quad N = Mg\cos\phi - \tfrac{1}{2}Mg\tan\theta\sin\phi = Mg\left(\cos\phi - \frac{\tan\theta\sin\phi}{2}\right).$$

沿斜面（以沿斜面向上为正）：

$$F = Mg\sin\phi + R\cos\phi \quad \Longrightarrow \quad F = Mg\left(\sin\phi + \frac{\tan\theta\cos\phi}{2}\right).$$

这就是题目要求的两个结果。

最大的角

梯脚不滑动的条件是：接触得以维持，即 $N>0$；以及摩擦力不超过其上限，即 $F\le\mu N$，其中 $\mu = 2$。（这里 $F>0$，所以这是唯一的摩擦条件。）当 $\phi = 45^\circ$ 时，$\sin\phi = \cos\phi = \frac{1}{\sqrt{2}}$，记 $t = \tan\theta$，则

$$F = \frac{Mg}{\sqrt{2}}\left(1+\frac{t}{2}\right), \qquad N = \frac{Mg}{\sqrt{2}}\left(1-\frac{t}{2}\right).$$

所以 $N>0$ 要求 $t<2$，而 $F\le 2N$ 则变成

$$1+\frac{t}{2} \le 2-t \quad\Longleftrightarrow\quad \frac{3t}{2}\le 1 \quad\Longleftrightarrow\quad \tan\theta \le \frac{2}{3}.$$

这比 $t < 2$ 更严格，所以它才是真正起限制作用的条件。由于 $\tan\theta$ 在 $0 < \theta < 90^\circ$ 上递增，梯子能够保持平衡的最大角为

$$\theta = \arctan\tfrac{2}{3} \approx 33.7^\circ,$$

而在这个角度下，摩擦力达到极限值（limiting friction）。

验算

- 当 $\phi = 0$（水平地面）时，公式给出 $N = Mg$ 和 $F = \frac{1}{2}Mg\tan\theta = R$，正是梯子置于水平地面时为人熟悉的结果。
- 当 $\theta = 0$（梯子铅垂）时，墙不起任何作用，而 $N = Mg\cos\phi$，$F = Mg\sin\phi$，与质点静止于斜面上的情况完全一样。
- 数值上，对随机选取的 θ 和 ϕ 直接解水平和铅垂方向的力方程，所得结果与两条公式吻合，误差在 10^{-13} 以内；而当 $\phi = 45^\circ$ 时，比值 F/N 在 $\theta = 33.690^\circ$ 处刚好越过 2。

常见错误

- 弄错 $R\sin\phi$ 一项的正负号。由于斜面向着墙壁上升，R 有一部分把梯脚拉离斜面，令 N 减小。
- 沿水平和铅垂方向分解，然后在解所得的联立方程时出错。这个做法行得通，但沿斜面和垂直于斜面的方向分解要快得多。

STEP 之外

留意斜面越陡，梯子可以倾斜的程度就越小，因为墙的反作用力一方面令 N 减小，另一方面又令 F 增大。这类问题是静力学（statics）的起点：我们选择合适的坐标轴和取矩点，把各个未知量分离开来。同样的想法也是分析桁架（trusses）和框架结构（frameworks）的基础，更广泛而言也是分析稳定性的基础。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对刚体平衡和摩擦有严谨细致的论述。

详细解答

STEP I 1989，第 14 题

概率与统计 · 条件分布

STEP I 1989，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/89-s1-q14

关键思路

一切都建基于一组概率：种子落在每一条阔一里格的地带内的机会。种子当且仅当落在园林或农地时才会长成树，所以树的数目服从二项分布（binomial distribution）。树生长之处的距离 Y，其实就是在“落在陆地上”这个条件下的 L。它的密度函数（density）就是把 f 只保留陆地地带的部分，再按比例放大，使总面积等于 1。最后一部分是假设检验（hypothesis test）：对一个二项计数而言，农地上有 23 棵树是否多得出奇？

解答

各地带的概率

L 的累积分布函数（cumulative distribution function）在 $0 \le x \le 4$ 时为

$$\mathrm{F}(x) = \int_0^x \left(\tfrac{1}{2} - \tfrac{1}{8}t\right)\mathrm{d}t = \frac{x}{2} - \frac{x^2}{16},$$

所以 $\mathrm{F}(1) = \frac{7}{16}$、$\mathrm{F}(2) = \frac{3}{4}$、$\mathrm{F}(3) = \frac{15}{16}$ 及 $\mathrm{F}(4) = 1$。因此

$$\mathrm{P}(\text{园林}) = \mathrm{P}(0 \le L \le 1) = \tfrac{7}{16}, \qquad \mathrm{P}(\text{湖}) = \tfrac{3}{4} - \tfrac{7}{16} = \tfrac{5}{16},$$
$$\mathrm{P}(\text{农地}) = \mathrm{P}(2 \le L \le 3) = \tfrac{15}{16} - \tfrac{3}{4} = \tfrac{3}{16}, \qquad \mathrm{P}(\text{海}) = \tfrac{1}{16}.$$

（由于 L 是连续的，各边界的概率为零，所以边界属于哪一条地带并无影响。）一颗种子长成树的概率为

$$p = \tfrac{7}{16} + \tfrac{3}{16} = \tfrac{5}{8}.$$

设 T 为一年内长出的树的数目。由于 96 颗种子各自独立地被风吹走，每颗长成树的概率为 $\frac{5}{8}$，故 $T \sim \mathrm{B}\left(96, \frac{5}{8}\right)$，而每年新长出的发烧树数目的平均值为

$$\mathrm{E}(T) = 96 \times \tfrac{5}{8} = 60.$$

第 (i) 部分

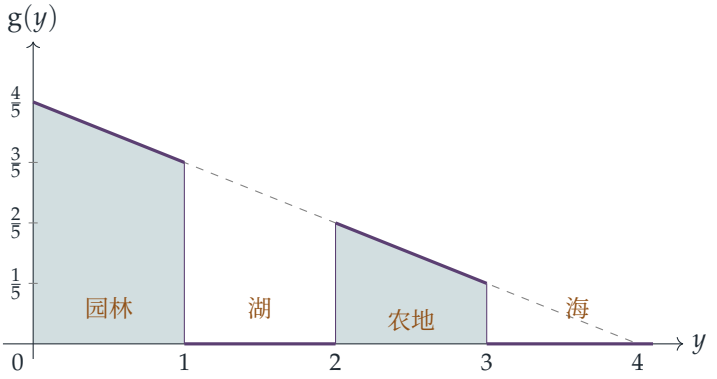
Y 的分布，就是已知种子落在陆地上时 L 的分布。以 $A = [0,1] \cup [2,3]$ 表示陆地。对于 $y \in A$，

$$\mathrm{P}(y \le Y \le y + \delta y) = \frac{\mathrm{P}(y \le L \le y + \delta y)}{\mathrm{P}(L \in A)} \approx \frac{\mathrm{f}(y)\,\delta y}{5/8},$$

而 Y 永远不会取 A 以外的值。所以 Y 的密度函数为

$$\mathrm{g}(y) = \begin{cases} \frac{8}{5}\left(\frac{1}{2} - \frac{1}{8}y\right) = \dfrac{4-y}{5} & \text{当} 0 \le y \le 1 \text{ 或} 2 \le y \le 3, \\ 0 & \text{其他情况。} \end{cases}$$

验算一下，它的面积是 $\frac{8}{5}\left(\frac{7}{16} + \frac{3}{16}\right) = 1$。图像是同一条下降直线上的两段：在园林上由 $\frac{4}{5}$ 降至 $\frac{3}{5}$，在湖上为零，然后在农地上由 $\frac{2}{5}$ 降至 $\frac{1}{5}$。



虚线是 $\frac{8}{5}\mathrm{f}$，画出来以作比较；Y 的密度函数只在陆地上才与它重合。

至于平均值，$\int y(4-y)\,\mathrm{d}y = 2y^2 - \frac{1}{3}y^3$，所以

$$\begin{aligned} \mathrm{E}(Y) &= \frac{1}{5}\left[2y^2 - \tfrac{1}{3}y^3\right]_0^1 + \frac{1}{5}\left[2y^2 - \tfrac{1}{3}y^3\right]_2^3 \\ &= \frac{1}{5}\left(\frac{5}{3}\right) + \frac{1}{5}\left(9 - \frac{16}{3}\right) = \frac{1}{5}\cdot\frac{5}{3} + \frac{1}{5}\cdot\frac{11}{3} = \frac{16}{15}. \end{aligned}$$

所以新树平均生长在城堡以南 $\frac{16}{15}$ 里格处。这个位置刚过了园林，落在湖中，但这并无不妥：平均值不一定是变量可以取到的值。

第 (ii) 部分

设 X 为一年内在农地长出的树的数目。若有依从传统，则放出的种子有 96 颗，每颗独立地以概率 $\frac{3}{16}$ 落在农地，所以

$$X \sim \mathrm{B}\left(96, \tfrac{3}{16}\right), \qquad \mathrm{E}(X) = 18, \qquad \mathrm{Var}(X) = 96 \cdot \tfrac{3}{16} \cdot \tfrac{13}{16} = \tfrac{117}{8} \approx 14.6,$$

由此得标准差（standard deviation）约为 3.82。

想一想 我们究竟在检验甚么？朝哪个方向检验？

零假设（null hypothesis）是传统得到依从：种子有 96 颗，所以 $X \sim \mathrm{B}\left(96, \frac{3}{16}\right)$。农地上的树比平常多，最自然的怀疑是放出的种子多于 96 颗，所以我们采用单尾检验（one-tailed test），问在零假设下出现 23 棵或以上的可能性有多大。

由于 $np = 18$ 和 $n(1-p) = 78$ 都够大，我们采用正态近似（normal approximation）$X \approx \mathrm{N}(18, 14.625)$，并作连续性修正（continuity correction）：

$$\mathrm{P}(X \geq 23) \approx \mathrm{P}\left(Z \geq \frac{22.5 - 18}{3.824}\right) = \mathrm{P}(Z \geq 1.18) \approx 0.12.$$

（二项分布的精确值是 0.121。）所以在正常情况下，农地上有 23 棵或以上的树，大约每八年便有一年出现。这远远未达 5% 水平（甚至 10% 水平）的显著性。这个观察结果与传统得到依从的说法相当吻合，单凭这项证据，国王的怀疑并无充分理由。

作为比较，即使在 5% 水平下，数目也要达到至少 25（精确值 $\mathrm{P}(X \geq 25) \approx 0.049$），国王才算有合理的根据。假如他同时掌握园林的树木数目，便可以检验树的总数 $T \sim \mathrm{B}\left(96, \frac{5}{8}\right)$，这对于判断放出了多少颗种子，会是好得多的证据。

验算

- 四个地带的概率 $\frac{7}{16}, \frac{5}{16}, \frac{3}{16}, \frac{1}{16}$ 相加得 1，而 Y 的密度函数的总面积为 1。
- 用中点法则作数值积分得 $\mathrm{E}(Y) = 1.06667$，与 $\frac{16}{15}$ 吻合。模拟 400 000 颗种子（取 $L = 4 - 4\sqrt{1-U}$，其中 U 在 $[0,1]$ 上均匀分布），结果有 62.5% 落在陆地上，而这些种子的平均落点距离为 1.069。
- 把二项概率逐项精确相加，得 $\mathrm{P}(X \geq 23) = 0.1213$，与正态估计值 0.120 相近。

常见错误

- 直接把 f 当作 Y 的密度函数。它在陆地上的面积只有 $\frac{5}{8}$，所以必须除以 $\frac{5}{8}$。
- 计算了 $\mathrm{E}(L;\, L \in A) = \int_A y\,\mathrm{f}(y)\,\mathrm{d}y = \frac{2}{3}$，却忘记除以 $\mathrm{P}(L \in A)$。

- 检验 $\mathrm{P}(X = 23)$ 而非尾部概率 $\mathrm{P}(X \geq 23)$。X 取任何一个个别数值的可能性本来就不大。

STEP 之外

第 (i) 部分是*条件分布*（conditional distribution）的一个例子：把随机变量限制在某个事件上，再重新归一化。密度函数形式的贝叶斯定理（Bayes' theorem），以及统计学中随处可见的截尾分布（truncated distribution），背后都是同一种运算。第 (ii) 部分让你初尝假设检验和 p 值，大学一年级的课程会用奈曼－皮尔逊（Neyman–Pearson）框架把这些概念严格化。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门教材）讲解条件分布时，附有大量像本题这样带故事情境的例题。

详细解答

STEP I 1989，第 15 题

概率与统计 · 分布的比较

STEP I 1989，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/89-s1-q15

关键思路

以 $12\,\mathrm{km\,h^{-1}}$ 的速率行驶，每公里刚好需时 5 分钟，所以每条路线的行程时间都是一个固定的分钟数，再加上某个随机变量（random variable）的简单倍数：这个变量分别服从均匀分布（uniform distribution）、泊松分布（Poisson distribution）或二项分布（binomial distribution）。于是，每一个关于时间的问题，都变成关于这一个变量的问题。中间的一问则是贝叶斯定理（Bayes′ theorem）的应用，其中三条路线被选中的可能性相等。

解答

行程时间

设 T_A、T_B、T_D 为三条路线的行程时间，以分钟计。由于 $12\,\mathrm{km\,h^{-1}}$ 即每 5 分钟行 1 km：

- Angle Avenue：$T_A = 25 + A$，其中 $A \sim \mathrm{U}(0, 10)$。
- Bend Boulevard：$T_B = 20 + 3B$，其中 $B \sim \mathrm{Po}(4)$。
- Detour Drive：设 X 为绕道的次数。假设五处道路工程互相独立（independent），每处导致绕道的概率均为 $\frac{4}{5}$，则 $X \sim \mathrm{B}(5, \frac{4}{5})$，路程为 $2 + X$ km，而 $T_D = 10 + 5X$。

至少 27 分钟

Angle Avenue。$T_A \geq 27 \iff A \geq 2$，所以

$$\mathrm{P}(T_A \geq 27) = \frac{10 - 2}{10} = \frac{4}{5}.$$

Bend Boulevard。$T_B \geq 27 \iff 3B \geq 7 \iff B \geq 3$，因为 B 是整数。所以

$$\mathrm{P}(T_B \geq 27) = 1 - \mathrm{e}^{-4}\left(1 + 4 + \frac{4^2}{2!}\right) = 1 - 13\mathrm{e}^{-4} \approx 0.7619.$$

Detour Drive。$T_D \geq 27 \iff 5X \geq 17 \iff X \geq 4$。所以

$$\mathrm{P}(T_D \geq 27) = 5\left(\frac{4}{5}\right)^4\frac{1}{5} + \left(\frac{4}{5}\right)^5 = \left(\frac{4}{5}\right)^4\left(1 + \frac{4}{5}\right) = \frac{256}{625}\cdot\frac{9}{5} = \frac{2304}{3125} = 0.73728.$$

已知行程较长，走的是哪条路线?

设 R 为所选的路线，其中 $\mathrm{P}(R = A) = \mathrm{P}(R = B) = \mathrm{P}(R = D) = \frac{1}{3}$，并设 L 为行程时间至少 27 分钟这一事件。根据贝叶斯定理，各个 $\frac{1}{3}$ 因子互相约去：

$$\begin{aligned}\mathrm{P}(R = B \mid L) &= \frac{\mathrm{P}(L \mid B)}{\mathrm{P}(L \mid A) + \mathrm{P}(L \mid B) + \mathrm{P}(L \mid D)} = \frac{1 - 13\mathrm{e}^{-4}}{\frac{4}{5} + 1 - 13\mathrm{e}^{-4} + \frac{2304}{3125}} \\ &= \frac{1 - 13\mathrm{e}^{-4}}{\frac{7929}{3125} - 13\mathrm{e}^{-4}} \approx \frac{0.7619}{2.2992} \approx 0.331.\end{aligned}$$

这个数值只比 $\frac{1}{3}$ 略高，十分合理：三条路线令行程达 27 分钟或以上的可能性几乎相同。

第 (i) 部分

利用 $\mathrm{E}(A) = 5$、$\mathrm{E}(B) = 4$ 和 $\mathrm{E}(X) = 5 \times \frac{4}{5} = 4$，以及期望值的线性性质（linearity of expectation），

$$\mathrm{E}(T_A) = 25 + 5 = 30, \qquad \mathrm{E}(T_B) = 20 + 3 \times 4 = 32, \qquad \mathrm{E}(T_D) = 10 + 5 \times 4 = 30.$$

? **想一想** 有两条路线打成平手。这是计错了吗?

不是。Angle Avenue 和 Detour Drive 的平均时间都刚好是 30 分钟，而 Bend Boulevard 的平均时间是 32 分钟。所以要令平均时间尽量短，我应该选用 Angle Avenue 或 Detour Drive 其中任何一条：两者同样好，而且都胜过 Bend Boulevard。

第 (ii) 部分

现在比较各条路线的 $\mathrm{P}(T < 32)$。

- $T_A < 32 \iff A < 7$，所以 $\mathrm{P}(T_A < 32) = \frac{7}{10} = 0.7$。
- $T_B < 32 \iff 3B < 12 \iff B \leq 3$，所以

$$\mathrm{P}(T_B < 32) = \mathrm{e}^{-4}\left(1 + 4 + 8 + \frac{32}{3}\right) = \frac{71}{3}\mathrm{e}^{-4} \approx 0.4335.$$

- $T_D < 32 \iff 5X < 22 \iff X \le 4$，所以

$$\mathrm{P}(T_D < 32) = 1 - \left(\frac{4}{5}\right)^5 = 1 - \frac{1024}{3125} = \frac{2101}{3125} \approx 0.6723.$$

由于 $0.7 > 0.6723 > 0.4335$，我应该每次都选用 **Angle Avenue**，这样有 70% 的日子可以在 32 分钟内到达。

验算

- 模拟 400 000 次行程，得出 $\mathrm{P}(T \ge 27) \approx 0.801,\ 0.761,\ 0.740$，平均时间 $\approx 30.0,\ 32.0,\ 30.0$ 分钟，$\mathrm{P}(T < 32) \approx 0.700,\ 0.435,\ 0.670$，以及 $\mathrm{P}(R = B \mid L) \approx 0.331$。全部与精确值吻合。
- 检查边界情况是否合理：在 Detour Drive 上，绕道 3 次需时 25 分钟，绕道 4 次则需时 30 分钟，所以“至少 27 分钟”确实等同于“至少绕道 4 次”。

常见错误

- 处理整数条件时不够严谨：$3B \ge 7$ 表示 $B \ge 3$，而不是 $B \ge 2$；$3B < 12$ 表示 $B \le 3$，而不是 $B \le 4$。
- 忘记 Detour Drive 本身有 2 km 的基本路程，或者忘记其他路线上每公里五分钟的骑车时间。
- 在 (i) 中声称只有一条路线胜出。其实两个平均值完全相等。

STEP 之外

第 (i) 和第 (ii) 部分说明，“最好”取决于我们要优化的是甚么：平均值只是概括一个分布的其中一个数字，两条平均值相同的路线，其离散程度和尾部概率可以相差很远。在决策理论（decision theory）中，这就是如何选择损失函数（loss function）的问题。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）涵盖了本题用到的各种分布和贝叶斯定理。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则收录了许多这一类的题目。

详细解答

STEP I 1989，第 16 题

概率与统计 · 零和博弈

STEP I 1989，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/89-s1-q16

本题在 *Beyond STEP* 第 12.14 节混合策略中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/mixed-strategies.html

关键思路

一切都源自一个小小的表：九对猜测各自对应的 B 的赢款。只要把这个表写下来，每个期望值（expected value）都不过是表中各项的加权和。至于最后一部分，把总和按 A 的概率 a_j 分组，“无论 A 怎样做”便化为对三个系数的条件，而要令三者皆为非负，只需解一个小型的线性不等式组。

解答

设 $w(i,j)$ 为在 A 猜 i、B 猜 j 的一局中 B 的赢款（以英镑计）。按规则逐一读出：

- 若 $|i-j|=2$，说 1 的一方付给说 3 的一方两英镑，故 $w(1,3)=2$ 及 $w(3,1)=-2$；
- 若 $|i-j|=1$，说较大数字的一方付出一英镑，故 $w(2,1)=w(3,2)=1$ 及 $w(1,2)=w(2,3)=-1$；
- $w(i,i)=0$。

	B 说 1	B 说 2	B 说 3
A 说 1	0	-1	2
A 说 2	1	0	-1
A 说 3	-2	1	0

留意 $w(j,i) = -w(i,j)$：规则对两名玩家一视同仁，所以把两人的猜测对调，付款的一方也随之对调。

第 (i) 部分

若 A 总是说 N，B 应沿着表的第 N 行看，并总是选取最大的一项。

- $N = 1$：B 总是说 3，每局赢 £2。
- $N = 2$：B 总是说 1，每局赢 £1。
- $N = 3$：B 总是说 2，每局赢 £1。

在每种情况下，这个选择都是唯一的最佳选择，其他任何回应不是一无所获，就是输钱。

第 (ii) 部分

在任何一局中，A 的猜测与 B 的猜测互相独立（independent）（A 自行随机选取，而 B 则暗中决定），所以 A 说 i 而 B 说 j 的概率是 $\frac{1}{3}b_j$。因此 B 的期望赢款是

$$\begin{aligned}\mathrm{E}(W) &= \sum_{j=1}^{3} b_j \cdot \tfrac{1}{3}\big(w(1,j) + w(2,j) + w(3,j)\big) \\ &= \tfrac{1}{3}\big(b_1(0+1-2) + b_2(-1+0+1) + b_3(2-1+0)\big) \\ &= \tfrac{1}{3}\,(b_3 - b_1).\end{aligned}$$

由于 $b_3 \le 1$ 及 $b_1 \ge 0$,此值至多为 $\frac{1}{3}$,而等号恰好在 $b_3 = 1$ 时成立(这时必有 $b_1 = b_2 = 0$)。所以 B 应该总是猜 3，这样每局的期望赢款是 £$\frac{1}{3}$。面对均匀随机选取的 A，猜 3 会赢 £2、输 £1 或打和，概率各为 $\frac{1}{3}$。

第 (iii) 部分

设 A 以概率 a_i 猜 i，B 以概率 b_j 猜 j，两者独立，则

$$\mathrm{E}(W) = \sum_{i=1}^{3}\sum_{j=1}^{3} a_i b_j\, w(i,j).$$

按每个 a_i 合并各项（每次取表中的一行）：

$$\mathrm{E}(W) = \underbrace{(2b_3 - b_2)}_{X}\, a_1 + \underbrace{(b_1 - b_3)}_{Y}\, a_2 + \underbrace{(b_2 - 2b_1)}_{Z}\, a_3.$$

我们希望 $X \ge 0$、$Y \ge 0$ 及 $Z \ge 0$，即

$$b_2 \le 2b_3, \qquad b_3 \le b_1, \qquad 2b_1 \le b_2.$$

把这些不等式串连起来，得 $b_2 \le 2b_3 \le 2b_1 \le b_2$，所以全部都必须是等式：$b_3 = b_1$ 及 $b_2 = 2b_1$。连同 $b_1 + b_2 + b_3 = 1$，便得

$$b_1 = \tfrac{1}{4}, \qquad b_2 = \tfrac{1}{2}, \qquad b_3 = \tfrac{1}{4},$$

这时 $X = Y = Z = 0$，当然都是非负的。所以这样的选择确实存在（而且事实上是唯一的一个）。

想一想 B 能否做得更好，令 X、Y 和 Z 全部严格为正?

不能。对任何 b_j，

$$b_1X + b_2Y + b_3Z = b_1(2b_3 - b_2) + b_2(b_1 - b_3) + b_3(b_2 - 2b_1) = 0,$$

这不过是说，B 与另一个自己对赛时不赚不蚀（这个表是反对称（antisymmetric）的）。以总和为 1 的非负权重对 X, Y, Z 取加权和，若三者皆为正数，结果不可能是零。这就是为甚么上面的不等式只能以等号成立。

现在来作出推论。取 $b = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$，则对任何 a_1, a_2, a_3 的选择，都有

$$\mathrm{E}(W) = 0 \cdot a_1 + 0 \cdot a_2 + 0 \cdot a_3 = 0$$

；更一般地，只要 $X, Y, Z \ge 0$，便有 $\mathrm{E}(W) = Xa_1 + Ya_2 + Za_3 \ge 0$，因为每个 $a_i \ge 0$。B 在不同局的赢款是独立同分布（independent and identically distributed）的（两名玩家的猜测在各局之间互相独立），每局的平均值至少为 0。根据大数定律（law of large numbers），经过许多局之后，B 每局的平均赢款会趋向这个平均值，所以长远而言，无论 A 怎样做，B 都不会输钱。 □

验算

- 第 (ii) 部分是第 (iii) 部分在 $a_1 = a_2 = a_3 = \frac{1}{3}$ 时的特殊情况：$\frac{1}{3}(X + Y + Z) = \frac{1}{3}(b_3 - b_1)$，与之前所得相符。
- 第 (i) 部分是 a 把全部权重放在 N 上的情况：当 $N = 1$ 时，$\mathrm{E}(W) = X = 2b_3 - b_2$，在 $b_3 = 1$ 时取得最大值，即 2。
- 模拟 2×10^5 局，B 采用 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$，A 采用 $(0.6, 0.1, 0.3)$，所得的平均赢款约为每局 -0.004，与 0 吻合。

常见错误

- 把正负号的规定弄反：相差 1 时，付款的是猜较大数字的一方，但相差 2 时，付款的却是猜较小数字的一方。应先把表列出，再检查它是否反对称。

- 从第 (ii) 部分推断 B 应该混合不同的猜测。面对一个固定的混合策略（mixed strategy），最佳回应总是一个纯策略（pure strategy）（这里是总是猜 3）；只有当 B 必须防范 A 的每一个策略时，混合才有好处，正如第 (iii) 部分那样。

STEP 之外

这是一个二人零和博弈（two-player zero-sum game），而第 (iii) 部分是冯·诺伊曼（von Neumann）极小极大定理（minimax theorem）的一个小型例子：B 有一个混合策略，无论 A 怎样做，都能保证他至少取得博弈的值；又因为这个博弈是对称的，所以该值为 0。寻找这样的策略是一个线性规划（linear programming）问题，它把概率与最优化及经济学连系起来。概率方面，可参阅 Grimmett and Welsh, *Probability: An Introduction*（概率入门教材）；至于博弈本身，Binmore 的 *Playing for Real*（博弈论入门读物）是很好的入门书。

详细解答

STEP II 1989，第 1 题

纯数学 · 三次方程的三角解法

STEP II 1989，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/89-s2-q1

关键思路

这条恒等式（identity）说明，当 $z = \cos\theta$ 时，$4z^3 - 3z$ 其实就是乔装了的 $\cos 3\theta$。因此，只要能把三次方程（cubic）化成 $4z^3 - 3z = k$ 的形式，而且 $|k| \le 1$，解这条方程就等同于解 $\cos 3\theta = k$，而后者有三个容易求得的解。两个代换（substitution）各司其职：平移 $y = x + a$ 用来消去平方项，缩放 $z = by$ 则把余下两个系数（coefficient）之比调整为 $4 : 3$。

解答

恒等式

利用复角公式（addition formulae）及 $\sin^2\theta = 1 - \cos^2\theta$，

$$\begin{aligned}\cos 3\theta &= \cos 2\theta\cos\theta - \sin 2\theta\sin\theta \\ &= (2\cos^2\theta - 1)\cos\theta - 2\sin^2\theta\cos\theta \\ &= 2\cos^3\theta - \cos\theta - 2(1 - \cos^2\theta)\cos\theta \\ &= 4\cos^3\theta - 3\cos\theta,\end{aligned}$$

正是所求。

化简三次方程

设 $y = x + a$，则 $x = y - a$，而 $24(y-a)^3 - 72(y-a)^2 + \cdots$ 中 y^2 的系数是 $-72a - 72$。要消去这一项，须取 $a = -1$，即 $x = y + 1$。于是

$$\begin{aligned} &24(y+1)^3 - 72(y+1)^2 + 66(y+1) - 19 \\ &\quad = (24y^3 + 72y^2 + 72y + 24) - (72y^2 + 144y + 72) \\ &\qquad + (66y + 66) - 19 \\ &\qquad\qquad = 24y^3 - 6y - 1, \end{aligned}$$

所以 $(*)$ 变成

$$24y^3 - 6y = 1.$$

现在代入 $y = z/b$：

$$\frac{24}{b^3}z^3 - \frac{6}{b}z = 1.$$

> **想一想** 应该怎样选取 b，才能令左方成为 $4z^3 - 3z$ 的倍数？

我们需要 z^3 与 z 的系数之比为 $4 : -3$，所以

$$\frac{24/b^3}{6/b} = \frac{4}{3} \quad\Longrightarrow\quad \frac{4}{b^2} = \frac{4}{3} \quad\Longrightarrow\quad b^2 = 3.$$

取 $b = \sqrt{3}$。这时方程是 $\dfrac{8}{\sqrt{3}}z^3 - \dfrac{6}{\sqrt{3}}z = 1$，两边乘以 $\frac{\sqrt{3}}{2}$，得

$$4z^3 - 3z = \frac{\sqrt{3}}{2}.$$

所以 $a = -1$、$b = \sqrt{3}$ 及 $k = \frac{\sqrt{3}}{2}$，其中 $z = \sqrt{3}\,(x-1)$。（若改取 $b = -\sqrt{3}$，则 $k = -\frac{\sqrt{3}}{2}$，最后得到的根（root）完全相同。）

求根

由于 $|k| \le 1$，我们可以写 $z = \cos\theta$，而根据恒等式，方程变成

$$\cos 3\theta = \frac{\sqrt{3}}{2} = \cos 30^\circ.$$

所以 $3\theta = 30^\circ,\ 390^\circ,\ 750^\circ$，从而 $\theta = 10^\circ,\ 130^\circ,\ 250^\circ$。这三个角给出三个不同的 $\cos\theta$ 值，而三次方程最多只有三个根，所以这就是全部的根：

$$z = \cos 10^\circ,\quad \cos 130^\circ,\quad \cos 250^\circ.$$

把代换还原，$x = 1 + \dfrac{z}{\sqrt{3}}$，所以

$$x_1 = 1 + \frac{\cos 10^\circ}{\sqrt{3}} = 1.5686\ldots \approx 1.57,$$
$$x_2 = 1 + \frac{\cos 130^\circ}{\sqrt{3}} = 0.6289\ldots \approx 0.629,$$
$$x_3 = 1 + \frac{\cos 250^\circ}{\sqrt{3}} = 0.8025\ldots \approx 0.803.$$

并非每条三次方程都适用

对于一般的首一（monic）三次方程 $x^3 + \alpha x^2 + \beta x + \gamma = 0$，平移的选择别无他法：代入 $x = y - a$ 后，y^2 的系数是 $\alpha - 3a$，所以 $a = \frac{\alpha}{3}$，由此得到 $y^3 + py + q = 0$，其中 $p = \beta - \frac{\alpha^2}{3}$。代入 $y = z/b$ 并乘以 $4b^3$，得

$$4z^3 + 4pb^2 z + 4qb^3 = 0,$$

要与 $4z^3 - 3z$ 吻合，须有 $4pb^2 = -3$，即 $b^2 = -\dfrac{3}{4p}$。只有当 $p < 0$ 时，这才有实数解。

一个反例（counterexample）是

$$x^3 + 3x - 4 = 0.$$

这里 $\alpha = 0$，所以 $a = 0$ 而 $p = 3 > 0$；我们需要 $b^2 = -\frac{1}{4}$，但没有实数 b 能满足这个条件。所以这条方程根本无法化成 $4z^3 - 3z = k$ 的形式。（方程本身确实有根：它可分解为 $(x-1)(x^2 + x + 4) = 0$，而当中二次式的判别式（discriminant）为 $-15 < 0$，所以 $x = 1$ 是它唯一的实根。失效的是这个方法，而不是方程。）

即使 $p < 0$，这个方法也可能在下一步失效。对于 $x^3 - 3x - 4 = 0$，代换 $x = 2z$ 给出 $4z^3 - 3z = 2$；但 $|\cos 3\theta| \le 1$，所以没有实数 θ 能使 $\cos 3\theta = 2$。

验算

- 三个根之和应为 $\frac{72}{24} = 3$，三个根之积应为 $\frac{19}{24}$。事实上，各 cos 项之和为零（$\cos 10^\circ + \cos 130^\circ + \cos 250^\circ = 0$），而数值上 $1.5686 \times 0.6289 \times 0.8025 = 0.7917 \approx \frac{19}{24}$。
- 把 $x = 1.5686$ 代入 $24x^3 - 72x^2 + 66x - 19$，在计算机的准确度之内结果为零；其余两个根亦然。

常见错误

- 求得 $\theta = 10^\circ$ 便停下来：另外两个根来自 $3\theta = 390^\circ$ 和 750°。

- 弄错 a 的正负号：$y = x + a$ 而 $x = y + 1$，即 $a = -1$，不是 $+1$。
- 举反例时，选了一条 $p < 0$ 而且 $|k| \le 1$ 的三次方程，但这个方法其实完全应付得来。在断言方法失效之前，先检查 $b^2 > 0$ 和 $|k| \le 1$ 这两个条件。

STEP 之外

这个方法恰好在三次方程有三个实根时适用，这正是所谓“不可约情形”（casus irreducibilis）：尽管每个答案都是实数，卡尔达诺（Cardano）公式却迫使你绕道复数的立方根。当 $|k| > 1$ 时，把 cos 换成 cosh，同样的想法依然可行，因为 $\cosh 3u = 4\cosh^3 u - 3\cosh u$；对于 $x^3 - 3x - 4 = 0$，这给出实根 $x = 2\cosh\left(\frac{1}{3}\operatorname{arcosh} 2\right) \approx 2.196$。至于哪些多项式方程可以用根式求解（solved by radicals）这个更深入的问题，则通往伽罗瓦理论（Galois theory）；Allenby 的 *Rings, Fields and Groups* 是很好的入门读物，而 Needham 的 *Visual Complex Analysis* 则展示了立方根的几何。

详细解答

STEP II 1989，第 2 题

纯数学 · 幂级数的系数

> **STEP II 1989，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/89-s2-q2

关键思路

一条把函数在 x 处的值与同一函数在 $2x$ 处的值联系起来的三角恒等式（trigonometric identity），会化为幂级数（power series）系数之间的关系。把 x 换成 $2x$，x^n 的系数就乘以 2^n，所以只要把两边都写成幂级数，便可以比较系数（compare coefficients），直接读出所求的关系。有用的窍门是先把整条式乘以 x，使每一项都是在 0 处确实具有幂级数的函数。

解答

以下一律假设 x 很小而且不等于零，因此下面出现的每个级数都收敛（converge）（所涉及的函数在例如 $0 < |x| < \frac{\pi}{4}$ 的范围内性质都良好）。两个收敛的幂级数若在 0 附近的一个区间上相等，它们的系数就相等，所以我们可以比较 x^n 的系数。

a_{n-1} 与 b_n 之间的关系

首先验证 $(*)$：

$$\cot x - \tan x = \frac{\cos^2 x - \sin^2 x}{\sin x \cos x} = \frac{\cos 2x}{\frac{1}{2}\sin 2x} = 2\cot 2x.$$

乘以 x：

$$x\cot x - x\tan x = (2x)\cot(2x).$$

现在把每一项写成级数。

▶ $x\cot x = 1 + \sum_{n\geq 1} b_n x^n$。

- $x\tan x = \sum_{m\geq 0} a_m x^{m+1} = \sum_{n\geq 1} a_{n-1}x^n$，其中取 $n = m+1$。
- 由于只要 x 很小，$2x$ 也很小，我们可以在 $x\cot x$ 的级数中把 x 换成 $2x$：

$$(2x)\cot(2x) = 1 + \sum_{n\geq 1} b_n (2x)^n = 1 + \sum_{n\geq 1} 2^n b_n x^n.$$

因此

$$1 + \sum_{n\geq 1} (b_n - a_{n-1})x^n = 1 + \sum_{n\geq 1} 2^n b_n x^n.$$

常数项相等，而对 $n \geq 1$ 比较 x^n 的系数，得 $b_n - a_{n-1} = 2^n b_n$，即

$$a_{n-1} = (1-2^n)\, b_n \qquad (n \geq 1),$$

这正是所要证明的结果。

想一想 为甚么要先乘以 x 才展开，而不直接处理 $\cot x$ 本身?

$\cot x$ 在 0 处趋向无穷大，所以在该处没有幂级数；题目只给了我们 $x\cot x$ 的级数。把整条恒等式乘以 x 之后，每一项都成为“具有幂级数的函数”的形式，而右边的余切项变成 $(2x)\cot(2x)$，也就是同一个函数在 $2x$ 处的值。这正是我们需要的形式。

系数 c_n

与之配对的恒等式是把差换成和：

$$\cot x + \tan x = \frac{\cos^2 x + \sin^2 x}{\sin x \cos x} = \frac{1}{\frac{1}{2}\sin 2x} = 2\operatorname{cosec} 2x.$$

再次乘以 x，

$$x\cot x + x\tan x = (2x)\operatorname{cosec}(2x).$$

利用与先前相同的展开式，以及把 x 换成 $2x$ 后的 $x\operatorname{cosec} x$ 的级数，

$$1 + \sum_{n\geq 1} (b_n + a_{n-1})x^n = 1 + \sum_{n\geq 1} 2^n c_n x^n,$$

所以对 $n \geq 1$ 有

$$2^n c_n = b_n + a_{n-1}.$$

现在用第一个结果消去 b_n。由于 $n \geq 1$，$1 - 2^n \neq 0$，所以 $b_n = \dfrac{a_{n-1}}{1-2^n}$，于是

$$2^n c_n = a_{n-1}\left(\frac{1}{1-2^n} + 1\right) = a_{n-1}\frac{2-2^n}{1-2^n} = a_{n-1}\frac{2\,(2^{n-1}-1)}{2^n - 1}.$$

除以 2^n，并利用 $2/2^n = 1/2^{n-1}$，

$$c_n = \frac{2^{n-1}-1}{2^n-1}\cdot\frac{1}{2^{n-1}}\,a_{n-1} \qquad (n \geq 1),$$

这就是所求的结果。

验算

由 $\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots$ 可得 $a_1 = 1$, $a_3 = \frac{1}{3}$。第一个结果预测 $b_2 = a_1/(1-4) = -\frac{1}{3}$ 和 $b_4 = a_3/(1-16) = -\frac{1}{45}$，与 $x\cot x = 1 - \frac{1}{3}x^2 - \frac{1}{45}x^4 - \cdots$ 吻合。第二个结果预测 $c_2 = \frac{1}{3}\cdot\frac{1}{2}\cdot 1 = \frac{1}{6}$ 和 $c_4 = \frac{7}{15}\cdot\frac{1}{8}\cdot\frac{1}{3} = \frac{7}{360}$，与 $x\operatorname{cosec} x = 1 + \frac{1}{6}x^2 + \frac{7}{360}x^4 + \cdots$ 吻合。另外留意 $n = 1$ 的情况：$a_0 = 0$ 而 $b_1 = c_1 = 0$，与两条公式都一致。我们把 sin 和 cos 的级数相除，以分数作精确计算，对所有 $n \le 23$ 验证了这两个关系，并用 $\tan 0.3 \approx 0.309336$ 和 $0.3\operatorname{cosec} 0.3 \approx 1.015159$ 核对了这些级数。

常见错误

- 忘记下标的平移：$x\tan x$ 中 x^n 的系数是 a_{n-1}，而不是 a_n。
- 在 $x\cot x$ 中以 $2x$ 代替 x 时，却写成 $2x\cot x$ 或 $x\cot 2x$；恒等式中的那一项必须恰好是 $(2x)\cot(2x)$。
- 除以 $1 - 2^n \neq 0$ 之前，没有指出这一点。

STEP 之外

这里的系数，除了相差一些简单的因子之外，就是伯努利数（Bernoulli numbers）：$x\cot x = \sum_{k\ge 0}(-1)^k \frac{2^{2k}B_{2k}}{(2k)!}x^{2k}$，而题目中的关系则以伯努利数表示出正切和余割的级数。两个幂级数若在 0 附近相等，便可以令它们的系数相等，这个事实（幂级数的恒等定理（identity theorem））会在大学第一个分析课程中证明。想进一步了解，可参阅 Abbott 的 *Understanding Analysis*（分析学入门教材）中关于幂级数的部分，以及 Needham 的 *Visual Complex Analysis*（以几何直观讲解复分析的著作），这些展开式在书中会从复函数自然地产生。

详细解答

STEP II 1989，第 3 题

纯数学 · 复双曲函数映射

STEP II 1989，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/89-s2-q3

关键思路

右边是 $2\sinh(x+\mathrm{i}y)$，所以整道题目其实是在研究映射（map）$u+\mathrm{i}v=\sinh(x+\mathrm{i}y)$。把和的 sinh 展开，便可分成实部（real part）和虚部（imaginary part），从而明确写出 u 和 v。固定 x 或 y 之后，剩下的是一条参数曲线（parametric curve），再利用恒等式 $\cos^2+\sin^2=1$ 和 $\cosh^2-\sinh^2=1$ 便可消去参数。

解答

求 u 和 v

两边除以 2，得 $u+\mathrm{i}v=\sinh(x+\mathrm{i}y)$。利用 $\sinh(\mathrm{i}y)=\mathrm{i}\sin y$ 和 $\cosh(\mathrm{i}y)=\cos y$，

$$\begin{aligned}\sinh(x+\mathrm{i}y)&=\sinh x\cosh(\mathrm{i}y)+\cosh x\sinh(\mathrm{i}y)\\&=\sinh x\cos y+\mathrm{i}\cosh x\sin y.\end{aligned}$$

（如果你喜欢，也可以直接展开 $\frac{1}{2}\left(\mathrm{e}^{x}(\cos y+\mathrm{i}\sin y)-\mathrm{e}^{-x}(\cos y-\mathrm{i}\sin y)\right)$；结果完全一样。）由于 x、y、u、v 全是实数，比较实部和虚部可得

$$u=\sinh x\cos y,\qquad v=\cosh x\sin y.$$

直线 $x=a$

固定 $x=a$ 而让 y 任意变化，则 $u=\sinh a\cos y$，$v=\cosh a\sin y$。由于 $a>0$，$\sinh a$ 和 $\cosh a$ 都是正数，所以可以相除：

$$\left(\frac{u}{\sinh a}\right)^2+\left(\frac{v}{\cosh a}\right)^2=\cos^2 y+\sin^2 y=1.$$

反过来，这个椭圆（ellipse）上的每一点都可以写成 $(\sinh a\cos y, \cosh a\sin y)$ 的形式（对某个 y），因为这正是椭圆的标准参数表示（parametrisation）。所以直线 $x = a$ 对应整个椭圆（y 每走过一个长度为 2π 的区间，椭圆便被描画一次），其半轴（semi-axes）沿 u 轴方向为 $\sinh a$，沿 v 轴方向为 $\cosh a$。

直线 $y = b$

现在固定 $y = b$ 而让 x 任意变化：$u = \sinh x\cos b$, $v = \cosh x\sin b$。题目给出 $0 < \sin b < 1$，所以 $\sin b \neq 0$，而且 $\cos^2 b = 1 - \sin^2 b > 0$，所以 $\cos b \neq 0$ 亦成立。相除后得

$$\left(\frac{v}{\sin b}\right)^2 - \left(\frac{u}{\cos b}\right)^2 = \cosh^2 x - \sinh^2 x = 1.$$

这是一条双曲线（hyperbola），其轴沿 v 轴，顶点（vertices）位于 $(0, \pm\sin b)$。

> **想一想** 为甚么只有一支？这一支又是否全部都被覆盖？

由于 $\cosh x > 0$ 且 $\sin b > 0$，必有 $v > 0$：只能到达上面的一支。而且这一支上的每一点都能到达：任取上支上的一点 (u, v)，选 $x = \sinh^{-1}(u/\cos b)$，使得 $\sinh x\cos b = u$；这时 $\cosh x\sin b$ 就是双曲线方程对于这个 u 的正根，亦即 v。所以直线 $y = b$ 恰好对应其中一支，即 $v > 0$ 的那一支。

交点与直角

点 $x = a$、$y = b$ 同时在两条直线上，所以它的像（image）同时在两条曲线上：

$$(u, v) = (\sinh a\cos b,\ \cosh a\sin b).$$

要找出两条曲线在该点的方向，可把每个参数表示对各自的参数求导（differentiate）。沿椭圆（$x = a$，参数为 y），在 $y = b$ 处：

$$\mathbf{t}_1 = \left(\frac{\mathrm{d}u}{\mathrm{d}y}, \frac{\mathrm{d}v}{\mathrm{d}y}\right) = (-\sinh a\sin b,\ \cosh a\cos b).$$

沿双曲线（$y = b$，参数为 x），在 $x = a$ 处：

$$\mathbf{t}_2 = \left(\frac{\mathrm{d}u}{\mathrm{d}x}, \frac{\mathrm{d}v}{\mathrm{d}x}\right) = (\cosh a\cos b,\ \sinh a\sin b).$$

两个向量都不是零向量，因为它们各有一个分量是 $\cosh a\cos b \neq 0$，所以两者都是真正的切向量（tangent vectors）。它们的纯量积（scalar product）是

$$\mathbf{t}_1 \cdot \mathbf{t}_2 = -\sinh a\sin b\cosh a\cos b + \cosh a\cos b\sinh a\sin b = 0,$$

所以椭圆与双曲线的这一支在该点正交，正如题目所求。

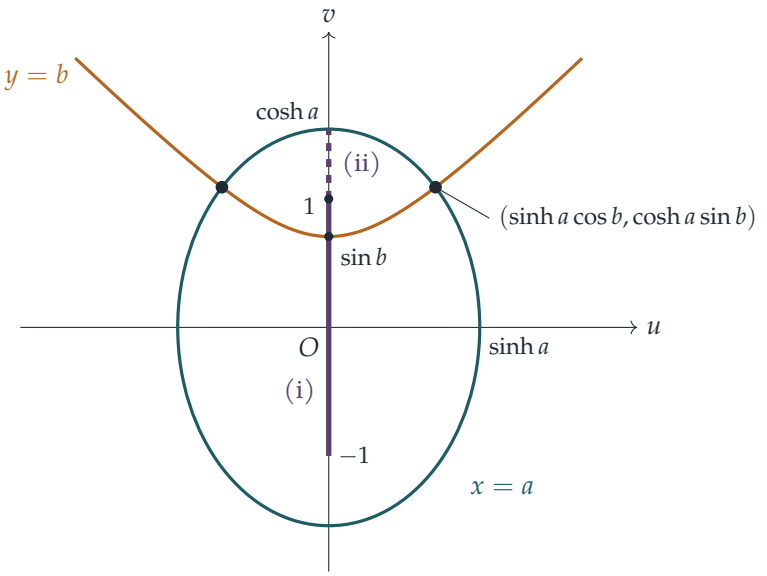
（由于两条曲线都具有 $u \mapsto -u$ 的对称性，它们在 $(-\sinh a\cos b, \cosh a\sin b)$ 亦正交；这一点是 $x = -a$、$y = b$ 的像，也同样是 $x = a$、$y = \pi - b$ 的像。）

略图

(i) $x = 0$ 给出 $u = 0$，$v = \sin y$：即 v 轴上由 $(0, -1)$ 至 $(0, 1)$ 的线段。留意它是椭圆在 $a \to 0$ 时的极限，因为 $\sinh a \to 0$ 而 $\cosh a \to 1$。

(ii) $y = \frac{1}{2}\pi$，$0 \le x \le a$ 给出 $u = 0$，$v = \cosh x$，其值由 1 递增至 $\cosh a$：即 v 轴上由 $(0, 1)$ 向上至椭圆顶端 $(0, \cosh a)$ 的线段。

双曲线这一支的顶点在 $(0, \sin b)$，位于线段 (i) 上。略图中取 $a = 1$，$b = \frac{1}{4}\pi$。



线段 (i) 以实线表示，线段 (ii) 以虚线表示；两者合起来沿 v 轴由 -1 一直向上伸延至椭圆顶端。假如改为取 $y = \frac{1}{2}\pi$ 而让 x 取遍所有实数，便会得到 v 轴上的射线 $v \ge 1$，而且重复覆盖两次；这就是题目排除了的退化情况 $\sin b = 1$。

验算

- 两条圆锥曲线（conics）的焦点（foci）都在 $(0, \pm 1)$：椭圆方面是因为 $\cosh^2 a - \sinh^2 a = 1$，双曲线方面是因为 $\sin^2 b + \cos^2 b = 1$。共焦（confocal）的椭圆和双曲线必定正交，与我们的结果吻合。
- 数值上，取 $a = 1, b = \frac{1}{4}\pi$：$\sinh(1 + \frac{1}{4}\pi\mathrm{i}) \approx 0.8310 + 1.0911\mathrm{i}$，与 $(\sinh 1 \cos \frac{1}{4}\pi,\ \cosh 1 \sin \frac{1}{4}\pi)$ 相符。我们亦在许多随机选取的 (a, b) 上验证了两条曲线的方程，以及纯量积为零。

常见错误

- 把 $\sinh(\mathrm{i}y) = \sin y$ 写成等式，而不是 $\mathrm{i}\sin y$，结果把 u 和 v 的角色对调了。
- 只说“一条双曲线”，却没有解释为甚么只出现 $v > 0$ 的那一支。
- 除以 $\cos b \neq 0$ 之前忘记检查这个条件；条件 $\sin b < 1$ 正是为此而设。

STEP 之外

映射 $z \mapsto \sinh z$ 是保角（conformal）的：凡是导数不为零的地方，它都保持角度不变，所以由直线 $x =$ 常数 和 $y =$ 常数 组成的垂直网格，必定变成由曲线组成的垂直网格，在这里就是共焦的椭圆和双曲线。这就是为甚么在平板或狭缝周围求解拉普拉斯方程（Laplace's equation）时，会出现这类“椭圆坐标”（elliptic coordinates）。Needham 的 *Visual Complex Analysis*（《可视化复分析》）把保角映射讲解得十分精彩，书中有许多正是这一类的图。

详细解答

STEP II 1989，第 4 题

纯数学 · 有理函数的描绘

> **STEP II 1989，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/89-s2-q4

关键思路

由线性因式相乘或相除而成的式子，最容易用取对数求导（logarithmic differentiation）的方法处理，因为对数会把它变成一个和。至于值域（range），不妨把问题倒过来问：对哪些 z，$\mathrm{f}(x) = z$ 有实数解？这条方程是 x 的二次方程（quadratic），所以答案来自判别式（discriminant）。之后的草图，就由零点（zeros）、极点（poles）、$|x|$ 很大时的表现，以及是否有某些 z 值取不到来决定。

关于印出来的题目，要说明一点。条件 $a + d \neq c + d$ 只是说 $a \neq c$，这是我们早已知道的。实际用到的条件是 $a + b \neq c + d$，以下我们一直假设这个条件成立。

解答

导数

在 $\mathrm{f}(x) \neq 0$ 的地方，即 $x \neq a, b$（以及 $x \neq c, d$）时，我们可以对绝对值取对数（logarithms of moduli）：

$$\ln|\mathrm{f}(x)| = \ln|x-a| + \ln|x-b| - \ln|x-c| - \ln|x-d|.$$

求导，并利用 $\frac{\mathrm{d}}{\mathrm{d}x}\ln|x-k| = \frac{1}{x-k}$，得

$$\frac{\mathrm{f}'(x)}{\mathrm{f}(x)} = \frac{1}{x-a} + \frac{1}{x-b} - \frac{1}{x-c} - \frac{1}{x-d}.$$

现在乘以 x。当 $x \neq 0$ 时，每一项变成

$$\frac{x}{x-k} = \frac{1}{1-k/x} = \left(1 - \frac{k}{x}\right)^{-1},$$

于是得到所求的恒等式

$$\frac{x\,\mathrm{f}'(x)}{\mathrm{f}(x)} = \left(1-\frac{a}{x}\right)^{-1} + \left(1-\frac{b}{x}\right)^{-1} - \left(1-\frac{c}{x}\right)^{-1} - \left(1-\frac{d}{x}\right)^{-1}.$$

$|x|$ 很大时

当 $|x|$ 远大于 $|k|$ 时，由二项式展开（binomial expansion）得 $(1-k/x)^{-1} = 1+\frac{k}{x}+O\left(\frac{1}{x^2}\right)$。四个首项 1 互相抵消（$1+1-1-1=0$），余下

$$\frac{x\,\mathrm{f}'(x)}{\mathrm{f}(x)} = \frac{a+b-c-d}{x} + O\left(\frac{1}{x^2}\right), \qquad \text{所以} \qquad \mathrm{f}'(x) = \frac{\mathrm{f}(x)}{x^2}\left[(a+b-c-d)+O\left(\tfrac{1}{x}\right)\right].$$

由于 $a+b-c-d \neq 0$，只要 $|x|$ 足够大，方括号内的式子便与 $a+b-c-d$ 同号。另外，当 $|x| \to \infty$ 时 $\mathrm{f}(x) \to 1$，所以在那里 $\mathrm{f}(x) > 0$，而且 $x^2 > 0$。因此，当 $|x|$ 很大时，在实数线的两端，$\mathrm{f}'(x)$ 都与 $a+b-c-d$ 同号。

曲线位于渐近线（asymptote）$y=1$ 的哪一侧，也值得记下来。直接计算，

$$\mathrm{f}(x) - 1 = \frac{(c+d-a-b)x + ab - cd}{(x-c)(x-d)} \approx \frac{c+d-a-b}{x} \quad \text{对于很大的} |x|,$$

而由分子可见，$\mathrm{f}(x) = 1$ 恰好只在一点成立，即 $x_0 = \dfrac{cd-ab}{c+d-a-b}$。

值域的条件：一个方法

把 $\mathrm{f}(x) = z$ 改写成 $(x-a)(x-b) = z(x-c)(x-d)$，这就是二次方程

$$(1-z)x^2 - \big(a+b-(c+d)z\big)x + (ab-cdz) = 0.$$

- 这条方程的根不可能是 $x=c$ 或 $x=d$，因为左边的 $(x-a)(x-b)$ 在那里不等于零。所以二次方程的实根恰好就是 $\mathrm{f}(x)=z$ 的实数解。
- 当 $z \neq 1$ 时，方程有实根，当且仅当判别式 $\Delta(z) = \big(a+b-(c+d)z\big)^2 - 4(1-z)(ab-cdz)$ 满足 ≥ 0。把它展开，乘以 $(c-d)^2 > 0$，再按 z 的幂次整理。设 $P=(a-c)(b-d)$ 和 $Q=(a-d)(b-c)$，有用的恒等式是 $P-Q=(a-b)(c-d)$，所以常数项 $(a-b)^2(c-d)^2$ 等于 $(P+Q)^2-4PQ$。于是题目给出的不等式正好就是 $(c-d)^2\Delta(z) \geq 0$。
- 当 $z=1$ 时，方程是线性的，x 的系数为 $c+d-a-b \neq 0$，所以有一个根；而不等式亦成立，因为两边相差 $(c-d)^2(a+b-c-d)^2 > 0$。

草图

记 $G(z)=\left[(c-d)^2z+P+Q\right]^2-4PQ$。当 z 变化时，平方项取遍 $[0,\infty)$ 内的每一个值，所以 G 的最小值是 $-4PQ$。因此：

$$\text{函数f 的值域是整个}\mathbb{R} \iff PQ\le 0.$$

若 $PQ>0$，则使 $G(z)<0$ 的 z 值构成一个开区间，这些就是取不到的值。

> **想一想** 没有实际的数值，我们怎样知道 PQ 的正负？

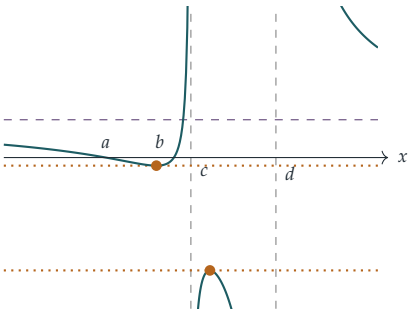
重新组合：$PQ=(a-c)(a-d)\cdot(b-c)(b-d)$。由于 $a<c<d$，第一对是两个负数的积，所以是正数。于是 PQ 与 $(b-c)(b-d)$ 同号，而后者为负，恰好就是 b 位于两个极点 c 与 d 之间的时候。一切都取决于 b 相对于 $c<d$ 的位置，以及 $a+b-c-d$ 的正负：

1. $a<b<c<d$：这时 $a+b<c+d$ 自动成立。
2. $a<c<b<d$：同样有 $a+b<c+d$，因为 $a<c$ 而且 $b<d$。
3. $a<c<d<b$，并且 $a+b<c+d$。
4. $a<c<d<b$，并且 $a+b>c+d$。

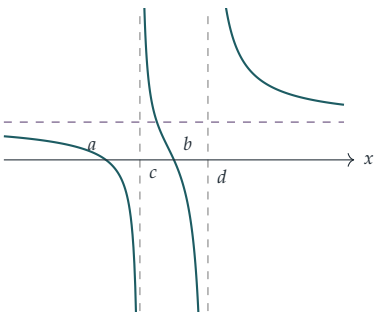
所以只有在情况 2，值域才是整个 $\mathbb{R}$。在情况 1、3 和 4，都有一个区间的值取不到。

还有一件工具，对找出转向点（turning points）很有帮助。若 f 在 x_1 有驻点（stationary point），而 $\mathrm{f}(x_1)=z_1$，则 x_1 是 $(x-a)(x-b)-z_1(x-c)(x-d)=0$ 的重根（repeated root）。（当 $z_1=1$ 时这是线性方程，所以不可能发生。）有重根即是 $G(z_1)=0$。所以驻点处的函数值恰好就是取不到的区间的两个端点。在情况 2，$G>0$ 恒成立，所以完全没有驻点。

f 在每个极点两侧附近的正负，可以靠数一数负因式的个数得出。在每幅图中，垂直虚线是极点 $x=c$、$x=d$，水平虚线是渐近线 $y=1$，琥珀色的点线标出取不到的区间的两个端点，也就是驻点（琥珀色圆点）处的函数值。各曲线是按特定数值（列于每幅图下方）绘画的，但在同一情况下，无论怎样选取数值，形状都是一样的。



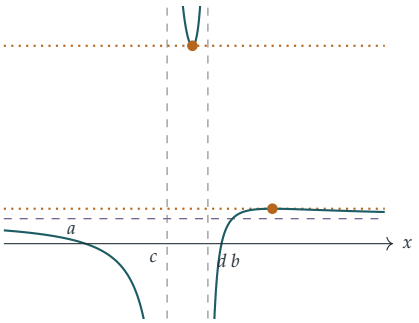
情况 1：$a,b,c,d=-2,0,\frac{1}{2},3$



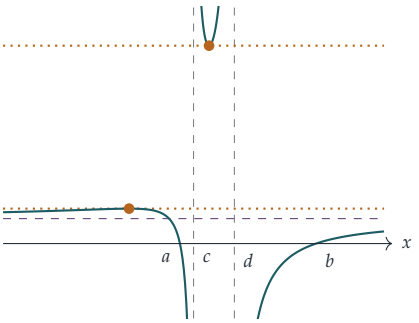
情况 2：$a,b,c,d=-1,1,0,2$

情况 1（$a < b < c < d$）。这里在左边远处 $\mathrm{f} < 1$，在右边远处 $\mathrm{f} > 1$。在 $x < c$ 上，曲线由略低于 1 的位置进来，在 a 处降至 0，在 a 与 b 之间下探至一个负的极小值，然后上升穿过 b，在 c 处趋向 $+\infty$。在两个极点之间，分母的两个因式异号，而分子为正，所以在两端都有 $\mathrm{f} \to -\infty$，中间有一个负的局部极大值（local maximum）。在 d 的右边，曲线由 $+\infty$ 下降至 1。(c, d) 上的局部极大值低于 (a, b) 上的局部极小值（local minimum），严格介乎两者之间的值都取不到。值域不是整个 $\mathbb{R}$。

情况 2（$a < c < b < d$）。这里没有驻点。曲线由略低于 1 的位置下降，穿过 a，在 c 处趋向 $-\infty$；然后由 $+\infty$ 穿过 b，在 d 处趋向 $-\infty$；再由 $+\infty$ 下降至 1。单是中间的一支已经取遍所有实数值。值域是整个 $\mathbb{R}$。



情况 3：$a, b, c, d = -3, 2, 0, \frac{3}{2}$



情况 4：$a, b, c, d = -\frac{1}{2}, \frac{9}{2}, 0, \frac{3}{2}$

情况 3（$a < c < d < b$，$a + b < c + d$）。在 $x < c$ 上，曲线由略低于 1 的位置下降，穿过 a，趋向 $-\infty$。在两个极点之间，分子为负，所以在两端都有 $\mathrm{f} \to +\infty$，中间有一个局部极小值 m。在 d 的右边，曲线由 $-\infty$ 上升穿过 b，在 x_0 处穿过 $y = 1$，到达略高于 1 的局部极大值 M，然后从上方回落至 1。(M, m) 内的值都取不到，其中 $1 < M < m$。值域不是整个 $\mathbb{R}$。

情况 4（$a < c < d < b$，$a + b > c + d$）。这是把情况 3 的外侧表现对调。在左边远处，f 略高于 1 并且递增。它到达局部极大值 $M > 1$，在 $x_0 < a$ 处穿过 $y = 1$，再下降穿过 a，在 c 处趋向 $-\infty$。中间的一支与之前一样呈 U 形，极小值为 m。在 d 的右边，曲线由 $-\infty$ 上升穿过 b，从下方趋向 1。(M, m) 同样取不到。值域不是整个 $\mathbb{R}$。

验算

- 恒等式：随机选取 a, b, c, d, x，用有限差分（finite difference）估算 $x\mathrm{f}'(x)/\mathrm{f}(x)$，结果与右边相符，误差约为 10^{-9}。

- 情况 1，取 $a,b,c,d=-2,0,\frac{1}{2},3$：这里 $P=\frac{15}{2}$，$Q=\frac{5}{2}$，$(c-d)^2=\frac{25}{4}$，而 $G(z)=0$ 给出 $z=-\frac{8}{5}\pm\frac{4}{5}\sqrt{3}$，约为 -0.2144 和 -2.9856。这正好就是解 $\mathrm{f}'(x)=0$ 所得的驻点处的函数值（在 $x\approx-0.515$ 和 $x\approx1.060$）。在 4×10^5 个点上对 f 取样，没有一个值严格落在两者之间。
- 情况 3 和 4（图中所用的数值）：取不到的区间是 $(1.4007, 7.9327)$，同样与驻点处的函数值吻合。在情况 2，$\mathrm{f}'(x)=0$ 没有实数解。

常见错误

- 忘记了 $\mathrm{f}(x)=z$ 只在 $z\neq1$ 时才是二次方程：$z=1$ 这个值需要另外交代一句。
- 以为曲线在两端都从同一侧趋近 $y=1$。事实并非如此：$\mathrm{f}-1\approx(c+d-a-b)/x$ 的正负随 x 而改变。

STEP 之外

把“f 的值域是甚么?”变成“对哪些 z，$\mathrm{f}(x)=z$ 有解?”，就是原像（preimage）的想法；而观察到驻点处的函数值正是两个解合而为一的地方，则是临界值（critical values）理论的起点。像这样的 2 次有理函数（rational function）是一个二对一的映射，它在复数上的表现要简洁得多：在复数上，每个值（按重数（multiplicity）计算）都恰好取到两次。Needham 的 *Visual Complex Analysis* 把这种几何观点发挥得十分精彩。

详细解答

STEP II 1989，第 5 题

纯数学 · 极坐标曲线与反射

STEP II 1989，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/89-s2-q5

关键思路

第 (i) 部分只不过是把 $x = r\cos\theta, y = r\sin\theta$ 代入 $\mathrm{d}y/\mathrm{d}x = (\mathrm{d}y/\mathrm{d}\theta)/(\mathrm{d}x/\mathrm{d}\theta)$。至于镜面，以 O 为中心的极坐标（polar coordinates）是最自然的选择，因为反射光线总是沿着半径回到极点（pole）。这样，反射定律（law of reflection）就能单凭 θ 告诉我们每一点上法线（normal）的方向，而第 (i) 部分则把这一点化为关于 $r(\theta)$ 的一阶微分方程（differential equation），而且这个方程可以分离变量。

解答

第 (i) 部分

极坐标为 (r,θ) 的点，其笛卡儿坐标（Cartesian coordinates）为 $x = r\cos\theta, y = r\sin\theta$。沿着曲线 $r = r(\theta)$，两者都是 θ 的函数，由积法则（product rule）得

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\cos\theta - r\sin\theta, \qquad \frac{\mathrm{d}y}{\mathrm{d}\theta} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\sin\theta + r\cos\theta.$$

由链式法则（chain rule，用于以参数表示的曲线），凡 $\mathrm{d}x/\mathrm{d}\theta \neq 0$ 之处，

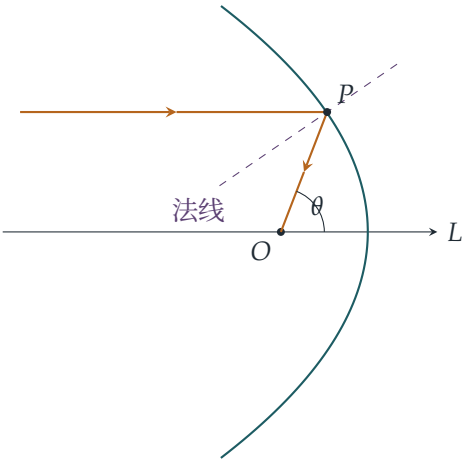
$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}\theta}{\mathrm{d}x/\mathrm{d}\theta} = \frac{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\sin\theta + r\cos\theta}{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\cos\theta - r\sin\theta} = \frac{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\tan\theta + r}{\dfrac{\mathrm{d}r}{\mathrm{d}\theta} - r\tan\theta},$$

其中最后一步把分子和分母同时除以 $\cos\theta$，只要 $\cos\theta \neq 0$ 便可以这样做。这就是所求的公式。

第 (ii) 部分

化为平面曲线问题。任取一个包含 L 的平面 Π，并在 Π 内取镜面上的一点 P。经过 P 且平行于 L 的光线位于 Π 内，O 亦然；因此反射光线 PO 也位于 Π 内。P 处的法线平分 Π 内两条直线的夹角，所以它也位于 Π 内。设镜面与 Π 相交所得的曲线为 C；它在 P 处的切线（tangent）既在 Π 内，又在镜面的切平面（tangent plane）内，所以与法线垂直。于是，在平面 Π 内，C 在 P 处的法线平分入射光线与反射光线的夹角：问题化为一个二维问题。

坐标。在 Π 内采用极坐标，以 O 为极点，始线（initial line）沿着 L，并选定方向使入射光线沿 $\theta = 0$ 的方向前进（如图所示，光线来自 L 上镜面开口以外的那一侧）。设 $P = (r, \theta)$ 在 C 上。



法线的方向。从 P 出发，入射光线沿方向 $(-1, 0)$ 往回走，而反射光线沿方向 $(-\cos\theta, -\sin\theta)$ 射向 O。两者都是单位向量（unit vector），所以它们的和指向两者夹角的平分线（bisector）方向，亦即法线方向：

$$(-1-\cos\theta, -\sin\theta) = -2\cos\tfrac{\theta}{2}\left(\cos\tfrac{\theta}{2}, \sin\tfrac{\theta}{2}\right),$$

这里用了 $1+\cos\theta = 2\cos^2\frac{\theta}{2}$ 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$。所以 P 处的法线与 L 成 $\frac{\theta}{2}$ 角，切线与法线垂直，而 C 在 P 处的斜率（gradient）为

$$m = -\cot\tfrac{\theta}{2}.$$

微分方程。记 $r' = \mathrm{d}r/\mathrm{d}\theta$，$t = \tan\theta$ 及 $\tau = \tan\frac{\theta}{2}$，于是 $t = \dfrac{2\tau}{1-\tau^2}$ 而 $m = -\dfrac{1}{\tau}$。由第 (i) 部分，

$$\frac{r't + r}{r' - rt} = m \quad\Longrightarrow\quad r'(t-m) = -r(1+mt) \quad\Longrightarrow\quad \frac{r'}{r} = -\frac{1+mt}{t-m}.$$

现在

$$1+mt = 1-\frac{2}{1-\tau^2} = -\frac{1+\tau^2}{1-\tau^2}, \qquad t-m = \frac{2\tau}{1-\tau^2}+\frac{1}{\tau} = \frac{1+\tau^2}{\tau(1-\tau^2)},$$

因此

$$\frac{1}{r}\frac{\mathrm{d}r}{\mathrm{d}\theta} = \tan\tfrac{\theta}{2}.$$

想一想 第 (i) 部分在 $\cos\theta = 0$ 之处失效。这会否留下漏洞?

不会。方程 $r'/r = \tan\frac{\theta}{2}$ 只涉及 θ 在 $-\pi < \theta < \pi$ 上的连续函数。除了孤立的值 $\theta = \pm\frac{\pi}{2}$（以及切线垂直的任何孤立点）可能例外，它在每一点都成立，所以由 r' 的连续性，它在这些点上同样成立。（另一个做法是：半径与切线之间的夹角 ψ 满足 $\tan\psi = r/r'$，而由几何关系可直接得出 $\psi = \frac{\pi}{2} - \frac{\theta}{2}$，并无任何例外点。）

求解。两边积分，并注意当 $-\pi < \theta < \pi$ 时 $\cos\frac{\theta}{2} > 0$,

$$\ln r = \int \tan\tfrac{\theta}{2}\,\mathrm{d}\theta = -2\ln\cos\tfrac{\theta}{2} + \text{常数}, \qquad \text{所以} \qquad r = \frac{k}{\cos^2\frac{\theta}{2}} = \frac{2k}{1+\cos\theta}$$

其中 $k > 0$ 为常数（因为 $r > 0$）。令 $\ell = 2k$。于是 $r + r\cos\theta = \ell$，即 $r = \ell - x$。两边平方，并利用 $r^2 = x^2 + y^2$,

$$x^2 + y^2 = (\ell - x)^2 \quad \Longleftrightarrow \quad y^2 = \ell^2 - 2\ell x.$$

这是一条以 L 为轴、以 O 为焦点（focus）的抛物线（parabola）（它就是 $Y^2 = 4aX$，其中 $X = \frac{\ell}{2} - x, a = \frac{\ell}{2}$，所以焦点位于 $x = \frac{\ell}{2} - a = 0$）。由于 Π 是任意一个包含 L 的平面，镜面与每个这样的平面都相交于一条抛物线，正如所求。（事实上，只有当镜面是单一个旋转抛物面（paraboloid of revolution）时，常数 ℓ 才会在每个平面内都相同；本题并不需要这一点。）

验算

- 当 $\theta = 0$ 时得 $r = \frac{\ell}{2}$：顶点位于 O 与准线（directrix）$x = \ell$ 的正中间，这正是以 O 为焦点的抛物线应有的性质。
- 当 $\ell = 2$ 而 $\theta = \frac{\pi}{2}$ 时，该点为 $P = (0,2)$，$y^2 = 4 - 4x$ 在该处的斜率为 $-1 = -\cot\frac{\pi}{4}$，而把方向 $(1,0)$ 对这条切线作反射便得 $(0,-1)$，笔直向下射到 O。
- 数值验证：第 (i) 部分的斜率公式在 2000 条随机曲线上与有限差分（finite differences）的结果吻合（准确至 10^{-8}）；以龙格—库塔法（Runge–Kutta）求解 $r' = r\tan\frac{\theta}{2}$，由 $r(0) = 1.35$ 出发，得 $r(2.5) = 13.5776380$，与 $2.7/(1+\cos 2.5)$ 相符；又在 1000 个随机点上，把平行于 L 的光线经 $y^2 = \ell^2 - 2\ell x$ 反射后，光线都通过 O。

常见错误

- 取错了角平分线：沿两条光线背离 P 方向的两个单位向量，其和给出法线；其差则给出切线。
- 忘记解释为何问题是平面问题：经过 L 的平面包含入射光线和 O，因而也包含反射光线和法线。
- 写成 $\ln r = -2\ln\cos\frac{\theta}{2}$ 而漏掉常数，这样便失去了整族抛物线。

STEP 之外

方程 $r(1+\cos\theta) = \ell$ 是圆锥曲线（conic）的极坐标形式，其焦点位于极点，离心率（eccentricity）为 1；把 1 换成 e 便得到椭圆和双曲线，而在极坐标下求解牛顿（Newton）方程时，行星轨道正是以这种形式出现的。这个反射论证也正是卫星碟形天线和汽车车头灯采用抛物面的原因。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）就是以这种方法推导出圆锥曲线轨道的。

详细解答

STEP II 1989，第 6 题

纯数学 · 反函数的积分

> **STEP II 1989，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/89-s2-q6

关键思路

右边有一个 g 对 y 的积分，所以最自然的做法是把左边的积分改写成以 $y = \mathrm{f}(x)$ 表示，即作代换（substitution）$x = \mathrm{g}(y)$。新的被积函数（integrand）含有 $y\,\mathrm{g}'(y)$，一看便知要用分部积分法（integration by parts），而边界项正好给出那两个长方形。第二条公式的做法完全一样，只是把 x 换成 $x^2/2$，把面积换成绕 y 轴的旋转体体积（volume of revolution）。

解答

第一条公式

由于在 $[a,b]$ 上 $\mathrm{f}'(x) > 0$，f 在该区间上严格递增（strictly increasing），所以它是由 $[a,b]$ 到 $[\alpha,\beta]$ 的双射（bijection），反函数（inverse function）$\mathrm{g}\colon [\alpha,\beta] \to [a,b]$ 存在。此外 g 可微（differentiable），且

$$\mathrm{g}'(y) = \frac{1}{\mathrm{f}'(\mathrm{g}(y))},$$

这个式子有意义，因为 f′ 永不为零。

代入 $x = \mathrm{g}(y)$，于是 $\mathrm{f}(x) = y$，$\mathrm{d}x = \mathrm{g}'(y)\,\mathrm{d}y$。当 $x = a$ 时，$y = \mathrm{f}(a) = \alpha$；当 $x = b$ 时，$y = \beta$。因此

$$\int_a^b \mathrm{f}(x)\,\mathrm{d}x = \int_\alpha^\beta y\,\mathrm{g}'(y)\,\mathrm{d}y.$$

现在用分部积分法，取 $u = y$ 和 $\mathrm{d}v = \mathrm{g}'(y)\,\mathrm{d}y$，所以 $v = \mathrm{g}(y)$：

$$\begin{aligned}\int_\alpha^\beta y\,\mathrm{g}'(y)\,\mathrm{d}y &= \Big[y\,\mathrm{g}(y)\Big]_\alpha^\beta - \int_\alpha^\beta \mathrm{g}(y)\,\mathrm{d}y \\ &= \beta\,\mathrm{g}(\beta) - \alpha\,\mathrm{g}(\alpha) - \int_\alpha^\beta \mathrm{g}(y)\,\mathrm{d}y \\ &= b\beta - a\alpha - \int_\alpha^\beta \mathrm{g}(y)\,\mathrm{d}y,\end{aligned}$$

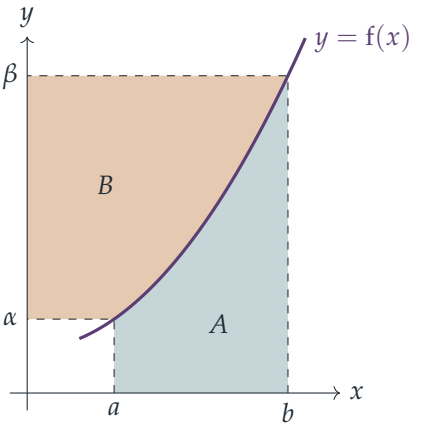
因为 $\mathrm{g}(\beta) = b$ 及 $\mathrm{g}(\alpha) = a$。这就是所求的公式。

几何意义

取 $a > 0$ 及 $\alpha > 0$。由于 f 递增，$0 < a < b$ 且 $0 < \alpha < \beta$，所以曲线 $y = \mathrm{f}(x)$ 由 (a, α) 到 (b, β) 的一段弧位于第一象限（first quadrant）。把公式写成

$$\underbrace{\int_a^b \mathrm{f}(x)\,\mathrm{d}x}_{\text{面积}A} + \underbrace{\int_\alpha^\beta \mathrm{g}(y)\,\mathrm{d}y}_{\text{面积}B} = b\beta - a\alpha.$$

这里 A 是 $a \le x \le b$ 时曲线与 x 轴之间的面积，B 是 $\alpha \le y \le \beta$ 时曲线与 y 轴之间的面积（把图形横过来看，曲线就是 $x = \mathrm{g}(y)$）。两者合起来，恰好填满长方形 $[0, b] \times [0, \beta]$ 除去较小的长方形 $[0, a] \times [0, \alpha]$ 后余下的部分，而这个 L 形区域的面积是 $b\beta - a\alpha$。



第二条公式

用同一个代换 $x = \mathrm{g}(y)$，$\mathrm{d}x = \mathrm{g}'(y)\,\mathrm{d}y$：

$$2\pi \int_a^b x\,\mathrm{f}(x)\,\mathrm{d}x = 2\pi \int_\alpha^\beta y\,\mathrm{g}(y)\,\mathrm{g}'(y)\,\mathrm{d}y.$$

想一想 这一次应该对哪一个因子积分?

留意根据链式法则（chain rule），$\mathrm{g}(y)\,\mathrm{g}'(y)$ 是 $\frac{1}{2}[\mathrm{g}(y)]^2$ 的导数。所以取 $u = y$ 和 $v = \frac{1}{2}[\mathrm{g}(y)]^2$ 作分部积分：

$$\begin{aligned} 2\pi \int_\alpha^\beta y\,\mathrm{g}(y)\,\mathrm{g}'(y)\,\mathrm{d}y &= 2\pi \left(\left[\tfrac{1}{2} y\,[\mathrm{g}(y)]^2\right]_\alpha^\beta - \int_\alpha^\beta \tfrac{1}{2}[\mathrm{g}(y)]^2\,\mathrm{d}y \right) \\ &= \pi(\beta b^2 - \alpha a^2) - \pi \int_\alpha^\beta [\mathrm{g}(y)]^2\,\mathrm{d}y, \end{aligned}$$

这里同样用到 $\mathrm{g}(\beta) = b$ 及 $\mathrm{g}(\alpha) = a$。这就是第二条公式。

第二条公式的几何意义

同样取 $a > 0$、$\alpha > 0$，把上图绕 y 轴旋转一整周。

- 区域 A 扫出一个立体，用圆柱壳法（method of cylindrical shells）计算，其体积为 $2\pi \int_a^b x\,\mathrm{f}(x)\,\mathrm{d}x$：与旋转轴相距 x、高 $\mathrm{f}(x)$、阔 $\mathrm{d}x$ 的窄条，旋转后成为一个体积约为 $2\pi x\,\mathrm{f}(x)\,\mathrm{d}x$ 的圆柱壳。
- 区域 B 扫出一个立体，用圆盘法（discs）计算，其体积为 $\pi \int_\alpha^\beta [\mathrm{g}(y)]^2\,\mathrm{d}y$：位于高度 y 的窄条，旋转后成为一个半径为 $\mathrm{g}(y)$ 的圆盘。
- 两个立体合起来，恰好填满一个半径为 b、高为 β 的圆柱体，再从其底部挖去一个半径为 a、高为 α 的同轴圆柱体，其体积为 $\pi b^2\beta - \pi a^2\alpha$。

所以这条公式是说：圆柱壳体积加圆盘体积，等于两个圆柱体的体积之差。条件 $a > 0$ 和 $\alpha > 0$ 确保两个区域都在旋转轴的同一侧，并且在 x 轴上方，这样两个积分才真正是体积，而两个立体也不会重叠。

验算

试取 $[1,2]$ 上的 $\mathrm{f}(x) = x^2$，于是 $\alpha = 1$，$\beta = 4$，$\mathrm{g}(y) = \sqrt{y}$。第一条公式左边是 $\frac{7}{3}$，右边是 $8 - 1 - \frac{2}{3}(8-1) = \frac{7}{3}$。第二条公式左边是 $2\pi \cdot \frac{15}{4} = \frac{15\pi}{2}$，右边是 $\pi(16-1) - \pi \cdot \frac{15}{2} = \frac{15\pi}{2}$。我们亦用辛普森法则（Simpson's rule）对两条公式作了数值验算，其中 g 以二分法（bisection）求得；所验算的几个函数包括 $[\frac{1}{2}, 2]$ 上的 $x^3 + x$，这时第一条公式两边都等于 5.859375。

常见错误

- 忘记更改积分上下限：代入 $x = \mathrm{g}(y)$ 之后，上下限变成 α 和 β，而不是 a 和 b。

- 没有说明 g 为何存在和可微。条件 $\mathrm{f}' > 0$ 正是为此而设。
- 把区域 B 说成位于曲线之下。其实它位于曲线与 y 轴之间。

STEP 之外

f 与 f^{-1} 的图像互为关于 $y = x$ 的反射，而第一条公式正是这个事实的积分版本。它是杨氏不等式（Young's inequality）背后的想法：若 f 递增、$\mathrm{f}(0) = 0$ 且 $a, b \geq 0$，则 $ab \leq \int_0^a \mathrm{f} + \int_0^b \mathrm{f}^{-1}$；凸分析（convex analysis）和哈密顿力学（Hamiltonian mechanics）中所用的勒让德变换（Legendre transform），背后也是同一个想法。arcsin 和 ln 等反函数的积分，同样可以靠它求出，无需任何巧妙的代换。想看代换与反函数的严谨处理，可参阅 Abbott 的 *Understanding Analysis*（一本实分析入门书）。

详细解答

STEP II 1989，第 7 题

纯数学 · 二阶微分方程

> **STEP II 1989，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/89-s2-q7

关键思路

$x\,y''$ 与 $-b\,y'$ 这两项放在一起看似棘手，但只要更换自变量（independent variable），设 $t = x^{\alpha}$，y'' 便会产生一个一阶导数项，而这一项可以恰好与 $-b\,y'$ 抵消。适当选取 α 做到这一点之后，余下的就是关于 t 的简谐运动（simple harmonic motion）。此后，当 $x \to 0$ 时的表现，就取决于 $\mathrm{d}y/\mathrm{d}x$ 中的因子 x^{b} 如何变化，而这又取决于 b 的正负。

解答

更换变量

设 $t = x^{\alpha}$，其中 α 待定，并把 y 视为 t 的函数。记 $\dot{y} = \mathrm{d}y/\mathrm{d}t$，由链式法则（chain rule）得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \alpha x^{\alpha-1}\dot{y},$$
$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \alpha(\alpha-1)x^{\alpha-2}\dot{y} + \alpha^2 x^{2\alpha-2}\ddot{y},$$

其中第二行用了积法则（product rule）以及 $\frac{\mathrm{d}}{\mathrm{d}x}\dot{y} = \alpha x^{\alpha-1}\ddot{y}$。代入后得

$$x\frac{\mathrm{d}^2y}{\mathrm{d}x^2} - b\frac{\mathrm{d}y}{\mathrm{d}x} = \alpha^2 x^{2\alpha-1}\ddot{y} + \alpha(\alpha-1-b)\,x^{\alpha-1}\dot{y}.$$

若取 $\alpha = b+1$，$\dot{y}$ 项便会消失；由于 $b > -1$，这个值不等于零。此时 $2\alpha - 1 = 2b+1$，于是方程变成

$$(b+1)^2 x^{2b+1}\ddot{y} + x^{2b+1}y = 0 \quad\Longrightarrow\quad \ddot{y} + \frac{y}{(b+1)^2} = 0,$$

这里两边除以了 $x^{2b+1} \neq 0$（因为 $x > 0$）。这是关于 t 的简谐运动，角频率（angular frequency）为 $1/(b+1)$，所以

$$\boxed{y = A\cos\left(\frac{x^{b+1}}{b+1}\right) + B\sin\left(\frac{x^{b+1}}{b+1}\right)}$$

其中 A、B 为任意常数。二阶方程配上两个任意常数：这就是通解（general solution）。

当 $x \to 0$ 时的表现

记 $u = x^{b+1}/(b+1)$。由于 $b+1 > 0$，当 $x \to 0^+$ 时 $u \to 0^+$，所以

$$y \to A\cos 0 + B\sin 0 = A.$$

因此 $y \to 1$ 迫使 $A = 1$。利用 $\mathrm{d}u/\mathrm{d}x = x^b$ 求导，得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = x^b\left(-\sin u + B\cos u\right) \qquad (A = 1).$$

情况 $b > 0$。此时 $x^b \to 0$，而括号内的式子趋向 B，是一个有限值。所以无论 B 取甚么值，都有 $\mathrm{d}y/\mathrm{d}x \to 0$。以下这族函数

$$y = \cos\left(\frac{x^{b+1}}{b+1}\right) + B\sin\left(\frac{x^{b+1}}{b+1}\right), \qquad B \in \mathbb{R},$$

中的每一个都满足两个条件。解是存在的，但并不唯一，正如题目所要求证明的。

> ? **想一想** 对于一般的二阶方程，给定 y 的一个值和 y' 的一个值，就能确定解。为甚么在这里却行不通？

唯一性定理（uniqueness theorem）要求方程可写成 $y'' = f(x, y, y')$ 的形式，而且 f 的性质要够好。两边除以 x 得 $y'' = (b/x)\,y' - x^{2b}y$，而系数 b/x 在 $x = 0$ 处趋向无穷。$x = 0$ 这一点是奇点（singular），所以通常的保证在该处并不适用。

情况 $b = 0$。此时 $x^b = 1$，而 $\mathrm{d}y/\mathrm{d}x \to B$。条件迫使 $B = 0$，得出唯一解 $y = \cos x$。（这其实就是 $y'' + y = 0$。）

情况 $-1 < b < 0$。此时 $x^b \to \infty$。若 $B \neq 0$，则 $Bx^b\cos u \to \pm\infty$，而另一项很小：由于 $\sin u \leq u$，

$$\left|x^b \sin u\right| \leq x^b \cdot \frac{x^{b+1}}{b+1} = \frac{x^{2b+1}}{b+1}.$$

更准确地说，$\sin u/u \to 1$，所以 $x^b \sin u \sim x^{2b+1}/(b+1)$。因此必须有 $B = 0$，这时

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -x^b \sin u \sim -\frac{x^{2b+1}}{b+1}.$$

这个式子趋向 0，当且仅当 $2b+1>0$，即 $b>-\frac{1}{2}$。当 $b=-\frac{1}{2}$ 时它趋向 -2，而当 $b<-\frac{1}{2}$ 时它无界。所以当 $-\frac{1}{2}<b<0$ 时恰好有一个解 $y=\cos\left(x^{b+1}/(b+1)\right)$，而当 $-1<b\le-\frac{1}{2}$ 时则没有解。

结论

这些条件能确定唯一解，恰好是在

$$-\tfrac{1}{2}<b\le 0,$$

的时候，而该解为 $y=\cos\left(x^{b+1}/(b+1)\right)$。



验算

- $b=0$ 时方程为 $y''+y=0$，其解 $A\cos x+B\sin x$ 与公式吻合。
- 数值上，通解在舍入误差范围内满足方程（在随机选取的 x, b, A, B 处用有限差分检验）。取 $b=-\frac{1}{4}, B=0$，我们在 $x=10^{-9}$ 处得到 $y'\approx -4.2\times10^{-5}$，趋向 0。取 $b=-\frac{1}{2}$ 时，该值保持在 -2；而取 $b=1, B=5$ 时，在 $x=10^{-9}$ 处的值为 5×10^{-9}。

常见错误

- 做到“$b\le0$ 时解唯一”就停下来：当 $-1<b\le-\frac{1}{2}$ 时，即使是余弦解也有 $y'\not\to 0$，所以根本没有解。
- 把 $\alpha x^{\alpha-1}\dot{y}$ 对 x 求导时忘记用积法则。多出来的那一项正是用来抵消 $-b\,y'$ 的。

STEP 之外

$x=0$ 这一点是这个方程的*正则奇点*（regular singular point）。在这类点附近，弗罗贝尼乌斯法（Frobenius method）会寻找形如 $x^r\sum c_n x^n$ 的解，而这里的指标方程（indicial equation）的根为 $r=0$ 和 $r=b+1$。在奇点处哪些条件能确定一个解，取决于这些指数，正如我们刚才所见。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）涵盖了级数解和正则奇点。

详细解答

STEP II 1989，第 8 题

纯数学 · 变换群

> **STEP II 1989，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/89-s2-q8

关键思路

一切都建基于一条恒等式（identity）：$\Omega^3 = -1$，所以 Ω 满足第一行的二次方程，我们因而可以把 Ω 的任何次幂化为 Ω 的线性式。一旦知道 R 和 T 怎样调动这四个点，余下的计数工夫便只与这四个点有关，而与公式无关：这个群（group）必须能把 z_0 送到四个点中的每一个，而它本身已包含三个保持 z_0 不动的不同映射。至于最后一部分，不妨问问 R 和 T 属于哪一类置换（permutation）。

解答

恒等式

由于 $\Omega = e^{i\pi/3}$，我们有 $\Omega^3 = e^{i\pi} = -1$，所以

$$0 = \Omega^3 + 1 = (\Omega + 1)(\Omega^2 - \Omega + 1).$$

因为 $\Omega \neq -1$，第二个因式必为零：$\Omega^2 - \Omega + 1 = 0$，这正是所求。

我们会以下列形式运用这条式：

$$\Omega^2 = \Omega - 1, \qquad \Omega^3 = -1, \qquad \Omega^4 = -\Omega.$$

另外留意 $2\Omega - 1 = 2\cos\frac{\pi}{3} - 1 + 2i\sin\frac{\pi}{3} = i\sqrt{3} \neq 0$。

R 和 T 置换这四个点

先看 R：乘以 $\Omega^2 = e^{2i\pi/3}$ 即是绕 0 旋转 $120°$：

$$R(0) = 0, \qquad R(1) = \Omega^2, \qquad R(\Omega^2) = \Omega^4 = -\Omega, \qquad R(-\Omega) = -\Omega^3 = 1.$$

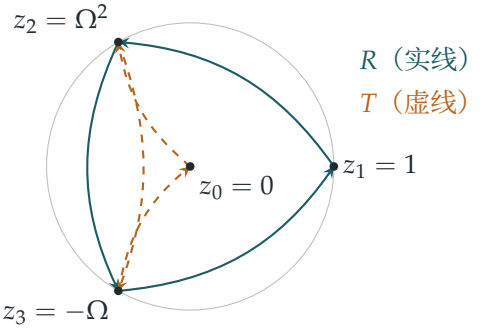
所以 R 保持 z_0 不动，并把 $z_1 \to z_2 \to z_3 \to z_1$。

再看 T：把各点逐一代入，并利用 $\Omega^2 = \Omega - 1$ 化简：

$$\begin{aligned} T(0) &= \frac{\Omega^2}{1} = z_2, \\ T(1) &= \frac{\Omega + \Omega^2}{2\Omega^2 + 1} = \frac{2\Omega - 1}{2\Omega - 1} = 1 = z_1, \\ T(\Omega^2) &= \frac{\Omega^3 + \Omega^2}{2\Omega^4 + 1} = \frac{\Omega - 2}{1 - 2\Omega} = -\Omega = z_3, \\ T(-\Omega) &= \frac{-\Omega^2 + \Omega^2}{-2\Omega^3 + 1} = \frac{0}{3} = 0 = z_0. \end{aligned}$$

在第三行中，$\Omega^3 + \Omega^2 = -1 + (\Omega - 1) = \Omega - 2$，而最后一步成立，是因为 $-\Omega(1 - 2\Omega) = -\Omega + 2\Omega^2 = -\Omega + 2\Omega - 2 = \Omega - 2$。所有分母均不为零（它们分别是 1、$\mathrm{i}\sqrt{3}$、$-\mathrm{i}\sqrt{3}$ 和 3）。

所以 T 保持 z_1 不动，并把 $z_0 \to z_2 \to z_3 \to z_0$。$R$ 和 T 都置换这四个点；两者各是一个 3-循环（cycle）。



这个群的阶至少为 12

设 G 为包含 R 和 T 的最小变换群：即由 R、T 及其逆变换经复合（composition）所能得到的一切变换。G 的每个元素都把集合 $\{z_0, z_1, z_2, z_3\}$ 映射到它自身之上，因为 R 和 T 如此（它们的逆变换亦然，因为那是这个有限集合上的双射（bijection）的逆）。

*三个保持 z_0 不动的映射。*集合 $H = \{I, R, R^2\}$ 包含于 G 之内，其中每个元素都保持 z_0 不动，而且三者是不同的变换（分别是旋转 0°、120°、240°）。

*把 z_0 送到每一个点的映射。*利用上面求得的循环，

$$g_0 = I, \qquad g_1 = R^2T, \qquad g_2 = T, \qquad g_3 = RT$$

满足 $g_k(z_0) = z_k$：例如 $RT(z_0) = R(z_2) = z_3$ 以及 $R^2T(z_0) = R^2(z_2) = R(z_3) = z_1$。

现在考虑十二个乘积 $g_k h$，其中 $k \in \{0,1,2,3\}$ 而 $h \in H$。由于 $h(z_0) = z_0$，我们得到 $g_k h(z_0) = z_k$。

> **想一想** 这十二个乘积当中，会否有两个其实是同一个变换?

不会。若 $g_k h = g_l h'$，把两边同时作用于 z_0：我们得到 $z_k = z_l$，所以 $k = l$。然后在左边复合上 g_k^{-1}，便得 $h = h'$。因此这十二个乘积是 G 中互不相同的元素，于是 $|G| \geq 12$。（用群论的语言来说：z_0 的轨道（orbit）大小为 4，而它的稳定子（stabiliser）大小至少为 3，所以根据轨道－稳定子定理（orbit–stabiliser theorem），$|G| \geq 4 \times 3$。）

最后一部分

有的：这四个点的 24 个置换当中，恰好有一半永远无法产生。

> **想一想** 作为置换，R 和 T 有甚么共通之处?

两者各是一个 3-循环，而 3-循环是偶置换（even permutation）：例如 $(z_1\, z_2\, z_3) = (z_1\, z_3)(z_1\, z_2)$，是两个对换（transposition）的乘积。它们的逆也是 3-循环。偶置换的复合仍是偶置换，所以由 R 和 T 产生的每一个点置换都是偶置换。四件物件只有 12 个偶置换，所以其余 12 个，即奇置换，永远不会产生。例如，把 z_0 和 z_1 对调而保持 z_2 和 z_3 不动，是单独一个对换，因而是奇置换，不可能得到。

事实上，结合上一部分，这就完全确定了 G。R 和 T 都是莫比乌斯变换（Möbius map）（T 有 $ad - bc = \Omega - 2\Omega^4 = 3\Omega \neq 0$），而保持三个点不动的莫比乌斯变换必为恒等变换，所以 G 的每个元素都由它怎样置换这四个点所决定。因此 $|G| \leq 12$，再配合上一部分，$|G| = 12$：G 在这些点上的作用，正是全体偶置换所成的群 A_4。

验算

- 以数值计算，$|\Omega^2 - \Omega + 1| \approx 2 \times 10^{-16}$，而在四个点上计算 R 和 T 的值，所得的正是上述两个 3-循环。
- $R^3 = I$ 显然成立；而 $T^3 = I$ 亦成立，正如 3-循环所应有的性质（已在一个随机点上检验，准确至 10^{-16}）。
- 用电脑生成 R 和 T 的所有复合，恰好得到 12 个不同的映射，全部诱导出偶置换，其中 z_0 被移到全部四个点，并有 3 个映射保持它不动。

常见错误

- 论证说“G 含有阶（order）为 3 的元素，所以 $|G| \geq 12$”。单凭阶为 3 的元素，只能得出 3 整除 $|G|$；因子 4 是来自在四个点上的可迁性（transitivity）。
- 没有化简 Ω 的幂，以致 $T(\Omega^2) = -\Omega$ 看似不成立。应先把一切化为 $a + b\Omega$ 的形式。

STEP 之外

$|G| =$ (轨道大小) $\times$ (稳定子大小) 这个计数结果就是轨道－稳定子定理，而全体偶置换构成交错群（alternating group）A_4，它也是正四面体的旋转群。这里出现这个现象，背后有个很漂亮的原因：莫比乌斯变换保持四个点的交比（cross-ratio）不变，而这四个点的交比是 $\mathrm{e}^{-\mathrm{i}\pi/3}$，正是那个特殊的“等非调和”（equianharmonic）值，对这个值而言，恰好只有偶置换保持交比不变。关于群和群作用（group action），可参阅 Allenby 的 *Rings, Fields and Groups*；关于莫比乌斯变换和黎曼球面（Riemann sphere），可参阅 Needham 的 *Visual Complex Analysis*。

详细解答

STEP II 1989，第 9 题

纯数学 · 矩阵指数与方程组

STEP II 1989，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/89-s2-q9

关键思路

一切都系于一个小小的计算：$\mathbf{A}^2$ 原来就是单位矩阵（identity matrix）。于是 $\mathbf{A}$ 的每个偶次幂都是 $\mathbf{I}$，每个奇次幂都是 $\mathbf{A}$，所以该级数可拆成指数级数（exponential series）的偶数项部分和奇数项部分，也就是 $\cosh t$ 和 $\sinh t$。至于微分方程（differential equation），我们仿照纯量（scalar）情况下的积分因子（integrating factor）法，由矩阵 $\mathbf{F}$ 担当 $\mathrm{e}^{\int a\,\mathrm{d}t}$ 的角色。

解答

矩阵 F

乘开可得

$$\mathbf{A}^2=\begin{pmatrix}-3&-1\\8&3\end{pmatrix}\begin{pmatrix}-3&-1\\8&3\end{pmatrix}=\begin{pmatrix}9-8&3-3\\-24+24&-8+9\end{pmatrix}=\begin{pmatrix}1&0\\0&1\end{pmatrix}=\mathbf{I}.$$

因此对所有 $k\geq 0$，$\mathbf{A}^{2k}=(\mathbf{A}^2)^k=\mathbf{I}$ 及 $\mathbf{A}^{2k+1}=\mathbf{A}^{2k}\mathbf{A}=\mathbf{A}$。写成 $\mathbf{I}=\frac{t^0}{0!}\mathbf{A}^0$，并把级数按 n 为偶数和奇数拆开（矩阵级数的每个元素都是绝对收敛（absolutely convergent）纯量级数的重排，所以这样做是容许的），

$$\begin{aligned}\mathbf{F}&=\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}\mathbf{A}^{2k}+\sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}\mathbf{A}^{2k+1}\\&=\mathbf{I}\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}+\mathbf{A}\sum_{k=0}^{\infty}\frac{t^{2k+1}}{(2k+1)!}=\mathbf{I}\cosh t+\mathbf{A}\sinh t,\end{aligned}$$

当中用到麦克劳林级数（Maclaurin series）$\cosh t=\sum t^{2k}/(2k)!$ 和 $\sinh t=\sum t^{2k+1}/(2k+1)!$。这正是所求的形式。

逆矩阵

设 $\mathbf{G} = \mathbf{I}\cosh t - \mathbf{A}\sinh t$。由于 $\mathbf{I}$ 与 $\mathbf{A}$ 可交换（commute），我们可以像处理普通数字一样乘开：

$$\mathbf{FG} = \mathbf{I}\cosh^2 t - \mathbf{A}^2\sinh^2 t + (\mathbf{A} - \mathbf{A})\cosh t \sinh t = \mathbf{I}(\cosh^2 t - \sinh^2 t) = \mathbf{I},$$

当中用到 $\mathbf{A}^2 = \mathbf{I}$ 和 $\cosh^2 t - \sinh^2 t = 1$。同样的计算给出 $\mathbf{GF} = \mathbf{I}$，所以 $\mathbf{F}^{-1} = \mathbf{I}\cosh t - \mathbf{A}\sinh t$。

导数

逐个元素求导（矩阵 $\mathbf{I}$ 和 $\mathbf{A}$ 都是常数矩阵），

$$\frac{\mathrm{d}\mathbf{F}}{\mathrm{d}t} = \mathbf{I}\sinh t + \mathbf{A}\cosh t, \qquad \text{而} \qquad \mathbf{FA} = \mathbf{A}\cosh t + \mathbf{A}^2\sinh t = \mathbf{A}\cosh t + \mathbf{I}\sinh t.$$

两者相同，所以 $\dfrac{\mathrm{d}\mathbf{F}}{\mathrm{d}t} = \mathbf{FA}$。另外留意 $\mathbf{FA} = \mathbf{AF}$，因为 $\mathbf{F}$ 是由 $\mathbf{I}$ 和 $\mathbf{A}$ 组合而成的。

微分方程

在纯量方程 $r' + ar = 0$ 中，我们乘以 e^{at}，它的导数是 $\mathrm{e}^{at}a$，于是左方恰好变成某个式子的导数。矩阵 $\mathbf{F}$ 正有这个性质，即 $\mathbf{F}' = \mathbf{FA}$，所以我们在方程的*左边*乘上 $\mathbf{F}$：

$$\mathbf{F}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} + \mathbf{FAr} = \mathbf{0}.$$

根据积法则（product rule）（只要保持各因子的次序，它对矩阵乘积同样成立），

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{Fr}) = \frac{\mathrm{d}\mathbf{F}}{\mathrm{d}t}\mathbf{r} + \mathbf{F}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{FAr} + \mathbf{F}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t},$$

这恰好就是左方。所以方程说的是

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{Fr}) = \mathbf{0},$$

而 $\mathbf{Fr}$ 是一个常向量 $\mathbf{c}$。

> ? **想一想** 为甚么要在左边乘而不是在右边乘？又怎样求出 $\mathbf{c}$？

乘积 $\mathbf{rF}$ 根本没有意义（$\mathbf{r}$ 是一个 2×1 的列向量），而且积法则得出的是 $\mathbf{FAr}$，只有当 $\mathbf{F}$ 在左边时，它才与 $\mathbf{F}(\mathbf{Ar})$ 这一项吻合。至于常向量，代入 $t = 0$：这时 $\cosh 0 = 1$ 而 $\sinh 0 = 0$，所以 $\mathbf{F} = \mathbf{I}$，并且

$$\mathbf{c} = \mathbf{I}\,\mathbf{r}(0) = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

现在在 $\mathbf{Fr} = \mathbf{c}$ 的左边乘上前面已求得的 $\mathbf{F}^{-1}$：

$$\mathbf{r} = \mathbf{F}^{-1}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \cosh t\begin{pmatrix}\alpha\\ \beta\end{pmatrix} - \sinh t\begin{pmatrix}-3 & -1\\ 8 & 3\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \cosh t\begin{pmatrix}\alpha\\ \beta\end{pmatrix} - \sinh t\begin{pmatrix}-3\alpha-\beta\\ 8\alpha+3\beta\end{pmatrix}.$$

读出各分量，

$$\begin{aligned} x(t) &= \alpha\cosh t + (3\alpha+\beta)\sinh t,\\ y(t) &= \beta\cosh t - (8\alpha+3\beta)\sinh t, \end{aligned}$$

正如所求。

验算

- 在 $t = 0$ 时，公式给出 $x = \alpha$、$y = \beta$，理应如此。
- 直接代入：$x' = \alpha\sinh t + (3\alpha+\beta)\cosh t$，而 $-(-3x-y) = 3x+y = (3\alpha+\beta)\cosh t + \alpha\sinh t$（用到 $9\alpha+3\beta-8\alpha-3\beta = \alpha$）。所以 $x' = -$(第一行，取自$\mathbf{Ar}$)，y 的情况亦类似。
- 数值上，把级数加到第 60 项，结果与 $\mathbf{I}\cosh t + \mathbf{A}\sinh t$ 吻合；而取 $\alpha = 1$、$\beta = 0$ 时，以龙格－库塔（Runge–Kutta）法解这个方程组，得出 $x(2) \approx 14.6428$，与 $\cosh 2 + 3\sinh 2$ 相符。

常见错误

- 在右边乘上 $\mathbf{F}$，又或者把 $\mathbf{AF}$ 和 $\mathbf{FA}$ 随意互换，却没有说明为何在这里可以这样做（它们可交换，是因为 $\mathbf{F}$ 是 $\mathbf{I}$ 和 $\mathbf{A}$ 的组合）。
- 拆开级数时漏掉 $\mathbf{I}$ 这一项（即 $n = 0$ 的一项），结果失去了 $\cosh t$ 的常数项 1。
- 把 $-\mathbf{A}$ 乘到 $(\alpha,\beta)^{\mathrm{T}}$ 上时弄错正负号。

STEP 之外

矩阵 $\mathbf{F}$ 就是矩阵指数（matrix exponential）$\mathrm{e}^{t\mathbf{A}}$，而这里的方法说明 $\mathbf{r}' = \mathbf{Mr}$ 的解总是 $\mathbf{r} = \mathrm{e}^{t\mathbf{M}}\mathbf{r}(0)$；本题中 $\mathbf{M} = -\mathbf{A}$，所以答案才会涉及 $\mathbf{F}^{-1} = \mathrm{e}^{-t\mathbf{A}}$。一般而言，计算 $\mathrm{e}^{t\mathbf{M}}$ 的方法是把 $\mathbf{M}$ 对角化（diagonalising）：我们的 $\mathbf{A}$ 的特征值（eigenvalues）是 ±1（这正是 $\mathbf{A}^2 = \mathbf{I}$ 的含意），$\mathrm{e}^{\pm t}$ 以至 cosh 和 sinh 就是由此而来。Strang 的 *Introduction to Linear Algebra*（线性代数入门）有一章把矩阵指数和微分方程组讲得很好，而 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）则详细讨论线性方程组。

详细解答

STEP II 1989，第 10 题

纯数学 · 直线与平面

> **STEP II 1989，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/89-s2-q10

关键思路

空间中的两条直线不交于唯一一点，有三种情况：两者是同一条直线；两者平行；或者两者是*异面直线*（skew lines），分处不同平面而互不相遇。因此“恰好交于一点”需要两个条件：两个方向不平行，而且两条直线位于同一平面上。向量积（cross product）可以同时处理这两个条件。至于那两个平面，其交线的方向就是两个法向量（normal）的向量积，而最后一部分其实只是第一部分换了个样子再出现一次。

解答

第一部分：两条直线何时恰好交于一点

两条直线相交之处满足 $\mathbf{a}+\lambda\mathbf{u}=\mathbf{b}+\mu\mathbf{v}$，即

$$\lambda\mathbf{u}-\mu\mathbf{v}=\mathbf{b}-\mathbf{a}.$$

要它们是真正的直线，需要 $\mathbf{u}\neq\mathbf{0}$ 和 $\mathbf{v}\neq\mathbf{0}$。于是：

- 若 $\mathbf{u}$ 与 $\mathbf{v}$ 平行（$\mathbf{u}\times\mathbf{v}=\mathbf{0}$），两条直线要么平行而相异（没有公共点），要么完全重合（有无限多个公共点）。无论哪一种，都不是恰好一点。
- 若 $\mathbf{u}\times\mathbf{v}\neq\mathbf{0}$，则 $\mathbf{u}$ 与 $\mathbf{v}$ 张成（span）一个平面，而方程有解当且仅当 $\mathbf{b}-\mathbf{a}$ 位于该平面上，即与 $\mathbf{u}\times\mathbf{v}$ 垂直。解若存在则必唯一，因为 $\mathbf{u}$ 与 $\mathbf{v}$ 线性无关（independent）。

所以条件是

$$\mathbf{u}\neq\mathbf{0},\quad \mathbf{v}\neq\mathbf{0},\quad \mathbf{u}\times\mathbf{v}\neq\mathbf{0},\quad (\mathbf{b}-\mathbf{a})\cdot(\mathbf{u}\times\mathbf{v})=0.$$

（头两个条件已由第三个推出，但仍值得写明。）若最后一个条件不成立，两条直线便是异

面直线。

第二部分：两个平面

两个法向量是 $\mathbf{n}_1 = (1,1,b)$ 和 $\mathbf{n}_2 = (b,b,1)$，而

$$\mathbf{n}_1 \times \mathbf{n}_2 = \left(1-b^2,\ b^2-1,\ 0\right) = (1-b^2)(1,-1,0).$$

两个平面不交于一条直线，当且仅当它们的法向量平行；在这里即 $1-b^2=0$，所以 $b=1$ 或 $b=-1$。

- $b=1$：两条方程都变成 $x+y+z=3$。
- $b=-1$：两条方程是 $x+y-z=1$ 和 $-x-y+z=-1$，同样是同一个平面。

所以在这两种情况下，两个平面都是重合，而不是平行而分开。

现在取 $b \neq \pm 1$。交线沿 $\mathbf{n}_1 \times \mathbf{n}_2$ 的方向，所以可以取 $\mathbf{u} = (1,-1,0)$。要找出线上的点，留意 x 和 y 只以 $s = x+y$ 的形式出现：

$$s + bz = b + 2, \qquad bs + z = 2b + 1.$$

从第二式减去第一式的 b 倍，得 $(1-b^2)z = 1-b^2$，所以 $z=1$（由于 $b^2 \neq 1$，可以相除），继而 $s=2$。因此交线是 $x+y=2$，$z=1$，与 b 无关。

我们需要在这条直线上找一点 $\mathbf{a}$ 使得 $\mathbf{a}\cdot\mathbf{u} = 0$，即 $x-y=0$。连同 $x+y=2$，得 $x=y=1$，所以

$$\mathbf{r} = \begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda\begin{pmatrix}1\\-1\\0\end{pmatrix}, \qquad b \neq \pm 1.$$

点 $\mathbf{a}$ 就是由原点到这条直线的垂足。

最后部分

坐标同时满足两条方程 $(*)$ 的点，当 $b \neq \pm 1$ 时组成上述直线，当 $b = \pm 1$ 时则组成整个平面。我们把这两种情况分开处理。

> ? **想一想** 当 $b = \pm 1$ 时，同时满足两条方程的点组成甚么集合?

是整个平面，而不是空集，因为我们已发现两个平面重合。在这些情况下，答案因此完全不同。

情况 $b \neq \pm 1$。我们需要题目所给的直线与直线 $\mathbf{r} = (1,1,1) + \lambda(1,-1,0)$ 恰好交于一点。应用第一部分，取 $\mathbf{a} = (1,1,1)$，$\mathbf{u} = (1,-1,0)$，$\mathbf{b} = (0,0,c)$，$\mathbf{v} = (1,d,0)$。于是

$$\mathbf{u}\times\mathbf{v} = (0,\,0,\,d+1), \qquad (\mathbf{b}-\mathbf{a})\cdot(\mathbf{u}\times\mathbf{v}) = (-1,-1,c-1)\cdot(0,0,d+1) = (c-1)(d+1).$$

我们需要 $\mathbf{u}\times\mathbf{v}\neq\mathbf{0}$，所以 $d\neq -1$，于是 $(c-1)(d+1)=0$ 迫使 $c=1$。所以条件是 $c=1$ 和 $d\neq -1$。（直接验证：题目所给的直线上 $z=c$，所以只有当 $c=1$ 时它才与 $z=1$ 相遇，而此时 $x+y=\mu(1+d)=2$ 在 $d\neq -1$ 时有唯一解 $\mu=2/(1+d)$。）

情况 $b=\pm 1$。现在我们需要题目所给的直线与一个平面恰好交于一点。点 $(\mu,d\mu,c)$ 位于 $x+y+z=3$ 上（对应 $b=1$）当 $\mu(1+d)=3-c$，位于 $x+y-z=1$ 上（对应 $b=-1$）当 $\mu(1+d)=1+c$。每条方程对 μ 恰有一个解，当且仅当 $1+d\neq 0$；若 $d=-1$，直线与平面平行，方程或无解，或有无限多个解。这里 c 可取任何值。

答案。恰有一个这样的点，当且仅当

$$d\neq -1 \quad 及 \quad (c=1 \text{ 或 } b=\pm 1),$$

可以简洁地写成 $d\neq -1$ 及 $(b^2-1)(c-1)=0$。

验算

- 把 $(1+\lambda,1-\lambda,1)$ 代入 $(*)$：$2+b=b+2$ 及 $2b+1=2b+1$，对所有 b 和 λ 都成立。
- 取 $b=2$，$c=1$，$d=3$，交点对应 $\mu=\frac{1}{2}$，即 $(\frac{1}{2},\frac{3}{2},1)$：$\frac{1}{2}+\frac{3}{2}+2=4=b+2$ 及 $1+3+1=5=2b+1$。
- 用电脑以精确算术测试了 $\{-3,\ldots,3,\frac{1}{2},-\frac{5}{3}\}$ 中所有的 b,c,d（共 729 个情况），直接求解 μ，结果解的数目为一，恰好就是上述条件成立之时。

常见错误

- 只因两条直线不平行，就说它们交于一点。在三维空间中，它们可以是异面直线。
- 以为当 $b=\pm 1$ 时两个平面平行而分开，从而断定完全没有这样的点。检查一下右边的常数：两个平面是重合的。
- 除以 $1-b^2$ 而没有说明它为何不等于零。

STEP 之外

这其实是一个小型的线性方程组（system of linear equations），只是换了个样子：三个未知数的两条方程，当系数矩阵的秩（rank）为 2 时，解组成一条直线；当秩为 1 时，解组成一个平面（或者无解）。加入直线的参数式后，它就变成一个关于 μ 的超定方程组（overdetermined system），而“恰有一个解”就是一个关于秩与相容性（consistency）的命题。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）正是由这里出现的行图像（平面相交）和列图像（向量的组合）出发，建立起这一切。

详细解答

STEP II 1989，第 11 题

力学 · 连接质点与冲量

STEP II 1989，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/89-s2-q11

关键思路

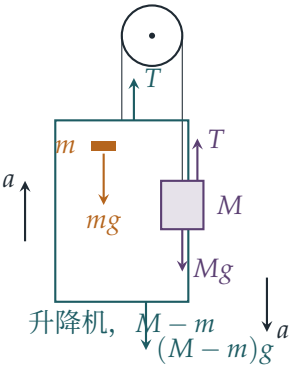
升降机与平衡锤一开始互相平衡，所以升降机的质量 M 必定已包括瓷砖在内。瓷砖一旦脱落，升降机便比平衡锤轻，于是向上加速，而瓷砖则自由下坠，地板因而升上来迎向瓷砖。处理碰撞时，把升降机、缆索和平衡锤视为沿缆索运动的单一系统。这个系统的“沿缆索的动量”只会因重力而改变，而碰撞产生的冲量（impulse）则属于系统的内力。

解答

建立模型

我们把升降机、平衡锤和瓷砖都当作质点（particle），并假设缆索轻而不可伸长（light and inextensible），绕过一个光滑而轻的滑轮（pulley）。导轨是光滑的，所以不会施加任何垂直方向的力。瓷砖掉下之前，系统处于静止状态，两边的质量分别为 M 和 M。因此升降机的质量 M 已包括瓷砖。（否则升降机一边的重量是 $(M+m)g$，另一边只有 Mg，系统不可能静止。）

瓷砖脱落之后，升降机的质量为 $M-m$，平衡锤的质量为 M，而瓷砖是一个自由质点。



到达地板所需的时间

设升降机以加速度 a 向上加速，则平衡锤以加速度 a 向下加速。设 T 为缆索的张力（tension）。由牛顿第二定律（Newton's second law）可得

$$\begin{aligned} &\text{升降机（向上）：} && T-(M-m)g=(M-m)a,\\ &\text{平衡锤（向下）：} && Mg-T=Ma. \end{aligned}$$

把两式相加便可消去 T：

$$a=\frac{mg}{2M-m},\qquad T=M(g-a)=\frac{2M(M-m)g}{2M-m}>0,$$

所以缆索一直保持拉紧。

> ? **想一想** 为甚么不能直接说瓷砖在重力作用下坠落了距离 h？

因为地板也在移动。瓷砖以 g 向下加速，地板以 a 向上加速，两者都由静止开始。因此两者之间的距离以恒定的相对加速度（relative acceleration）收窄：

$$g+a=g+\frac{mg}{2M-m}=\frac{2Mg}{2M-m}.$$

这段距离最初是 h，所以瓷砖在 $\frac{1}{2}(g+a)t^2=h$ 时着地，由此可得

$$t^2=\frac{2h}{g+a}=\frac{(2M-m)h}{Mg},\qquad t=\sqrt{\frac{(2M-m)h}{Mg}}.$$

这并不是题目所印的式子，题目中出现的是 $2(M-m)$ 而不是 $2M-m$。两者只在 $m=0$ 时才一致，而我们的结果会在下文以数值方法逐步模拟运动来验证。题目所印的式子，需要升降机以 $mg/(M-m)$ 加速才会成立。如果忘记了平衡锤也要被加速，并把张力当作维持在 Mg，就会得到这个结果。我们相信题目所印的公式是原题的误植或疏忽。题目其余部分完全不依赖这个式子。

碰撞

在撞击前的一刻，升降机向上运动，平衡锤向下运动，两者的速率都是 $u = at$。瓷砖则以速率 $v = gt$ 向下运动。于是

$$(2M - m)u = (2M - m) \cdot \frac{mg}{2M - m}\, t = mgt = mv.$$

现在考虑碰撞本身。取“正方向”为升降机向上／平衡锤向下，即缆索带动两者一起运动的方向。撞击历时极短，所以有限的力（各物体的重量）所产生的冲量可以忽略。仅有的冲量是瓷砖与地板之间的冲量（它是“升降机连瓷砖”这个整体的内力），以及缆索中的冲击张力（impulsive tension）J。设 w 为碰撞后的共同速率（升降机连瓷砖，质量为 M，向上；平衡锤向下）。碰撞是完全非弹性的（perfectly inelastic），所以瓷砖最终随地板一起运动。于是

$$\begin{aligned} \text{升降机连瓷砖：} &\quad Mw - \big[(M - m)u - mv\big] = J, \\ \text{平衡锤：} &\quad Mw - Mu = -J. \end{aligned}$$

把两式相加，

$$2Mw = (M - m)u + Mu - mv = (2M - m)u - mv = 0,$$

所以 $w = 0$：升降机停了下来。（另外 $J = Mu > 0$，所以缆索确实是被猛然拉紧而不是变松，这与我们的假设一致。）撞击之后，两边的质量再次是 M 和 M，互相平衡，所以一切保持静止。

> ? **想一想** 为甚么在撞击前的一刻，总的“沿缆索的动量”是零?

撞击之前，作用在这三个物体上而又沿运动方向的外力只有重力，按上面所定的正方向计算，其净效果为 $Mg - (M - m)g - mg = 0$。（滑轮的反作用力不作功，而缆索属于系统内部。）所以这个动量自始至终保持在初始值 0。碰撞同样不能改变它。

能量损失

系统由静止开始，最后也归于静止，所以损失的机械能（mechanical energy）等于开始与结束之间重力势能（gravitational potential energy）的减少量。在坠落期间，设升降机上升了 s_1（因此平衡锤下降了 s_1），并设瓷砖下降了 s_2。瓷砖走完了原来相距的 h，所以 $s_1 + s_2 = h$。势能的变化为

$$(M - m)g\, s_1 - Mg\, s_1 - mg\, s_2 = -mg(s_1 + s_2) = -mgh.$$

碰撞之后再没有任何东西移动，所以损失的能量是 mgh，正如题目所求。撞击前一刻所有的动能（kinetic energy）都在碰撞中消失。留意这个论证完全没有用到 t 的值。

验算

- 若 $m \to 0$，则 $a \to 0$ 且 $t \to \sqrt{2h/g}$，即自由落体落在静止的地板上。两条公式都通过这个测试，所以它无法分辨两者。
- 取 $M = 5$、$m = 1$、$h = 1$ 及 $g = 9.8$，以数值方法逐步模拟运动，得 $t \approx 0.4286\,\text{s}$，与 $\sqrt{9/49} \approx 0.4286$ 吻合。题目所印的公式则给出 $\sqrt{8/49} \approx 0.404\,\text{s}$。
- 在同一个模拟中，撞击前一刻的动能等于 mgh，准确至五位有效数字，而碰撞后的速率为零，误差只在舍入范围之内。

常见错误

- 把升降机的质量当作 M 另加瓷砖：这样系统一开始就不可能静止。
- 忘记了平衡锤的惯性（inertia），以致算出升降机的加速度为 $mg/(M-m)$。
- 在碰撞中只对升降机和瓷砖使用线动量守恒（conservation of linear momentum），忽略了缆索中的冲击张力。

STEP 之外

这里所用的“沿缆索的动量”是一种广义动量（generalised momentum）：对于只有一个自由度（degree of freedom）的系统，其坐标就是已绕过滑轮的缆索长度。拉格朗日力学（Lagrangian mechanics）把这套做法系统化。凡是力在其上没有净效果的坐标，都对应一个守恒的广义动量，而冲击约束（impulsive constraint）也可用同样的方法处理。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）仔细讨论了以绳和滑轮连接的系统，以及碰撞中的冲量和能量。

详细解答

STEP II 1989，第 12 题

力学 · 刚体平衡

> **STEP II 1989，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/89-s2-q12

关键思路

作用在薄片（lamina）上的力只有三个：它的重量和两根钉子的反作用力（reaction）。光滑的钉子只会沿垂直于所接触的边的方向施力，而在长方形中，这就表示每个反作用力都平行于另一条边。一个物体若在三个互不平行的力作用下保持平衡（equilibrium），这三个力的作用线（line of action）必须交于一点。因此，两个反作用力的交点必定位于通过中心的铅垂线上。两个结果都可由这一事实，加上两根钉子等高这个条件推导出来。

解答

建立模型

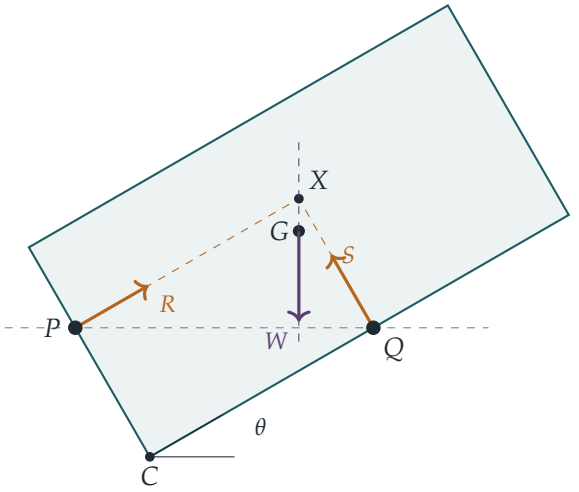
我们把薄片看成一个重量为 W、中心为 G 的均匀刚性长方形，并把两根钉子看成位于同一高度的光滑点 P 和 Q。设 C 为最低的顶点。通过 C 而长度为 $2a$ 的边搁在 P 上，其中 $CP = p$；通过 C 而长度为 $2b$ 的边搁在 Q 上，其中 $CQ = q$。$2b$ 边与水平线成锐角 θ，所以（与它垂直的）$2a$ 边与铅垂线成角 θ。

最方便的做法是沿两条边设立坐标：由 C 出发，沿 $2b$ 边量度 u，沿 $2a$ 边量度 v。于是

$$C = (0,0), \qquad P = (0,p), \qquad Q = (q,0), \qquad G = (b,a).$$

点 (u,v) 相对于 C 的水平位移和铅垂位移为

$$x = u\cos\theta - v\sin\theta, \qquad y = u\sin\theta + v\cos\theta.$$



两个反作用力交于通过 G 的铅垂线上

P 处的反作用力 R 垂直于 $2a$ 边，所以它平行于 $2b$ 边：它的作用线是 $v = p$。同理，Q 处的反作用力 S 沿 $u = q$ 作用。这两条直线交于 $X = (q, p)$，即长方形 $CPXQ$ 的第四个顶点。

对 X 取力矩（moment），两个反作用力的力矩均为零，所以重量的力矩也必须为零：通过 G 的铅垂线通过 X。换句话说，X 和 G 的水平坐标相同：

$$q\cos\theta - p\sin\theta = b\cos\theta - a\sin\theta, \qquad \text{即} \qquad (q-b)\cos\theta = (p-a)\sin\theta. \qquad (1)$$

（分解各力可得 $R = W\sin\theta$ 和 $S = W\cos\theta$，两者皆为正数，所以钉子确实能够提供这些力。）

第一个结果

由于 $0 < \theta < \frac{\pi}{2}$，$\cos\theta$ 和 $\sin\theta$ 都是正数，所以由 (1) 可知 $q - b$ 与 $p - a$ 同号。若 $p < a$，则 (1) 的右边为负数，所以 $q - b < 0$，即 $q < b$。这正是所求的结果：若该顶点与 $2a$ 边上的钉子相距少于 a，则它与另一根钉子相距少于 b。

> ? **想一想** 为甚么问题还未完结？d 在哪里出现？

方程 (1) 只用到力矩条件。它还未涉及 d，而且 p 和 q 并非互相独立：两根钉子等高，并且相距 d。现在我们把这两个事实用上。

与 d 的关系

两根钉子位于同一高度，所以 P 和 Q 的 y 坐标相等：

$$p\cos\theta = q\sin\theta, \qquad \text{所以} \qquad p = q\tan\theta. \tag{2}$$

它们的水平距离为 d：

$$d = x_Q - x_P = q\cos\theta + p\sin\theta = q\cos\theta + \frac{q\sin^2\theta}{\cos\theta} = \frac{q}{\cos\theta}. \tag{3}$$

（等价地说，三角形 CPQ 以 C 为直角，斜边（hypotenuse）$PQ = d$ 是水平的，所以 $q = d\cos\theta$ 及 $p = d\sin\theta$。）

现在把 $q = d\cos\theta$ 和 $p = d\sin\theta$ 代入 (1) 的第一个形式的左边：

$$\begin{aligned} b\cos\theta - a\sin\theta &= q\cos\theta - p\sin\theta \\ &= d\cos^2\theta - d\sin^2\theta \\ &= d\cos 2\theta, \end{aligned}$$

这就是所求的结果。

验算

- 正方形的情况 $a = b$、$\theta = \frac{\pi}{4}$：两边都等于零。根据对称性，对角线 CG 是铅垂的，所以结果吻合；而且这对任何 d 都成立，正方形对称地放置时理应如此。
- 取 $a = 1$、$b = 2$、$\theta = \frac{\pi}{6}$，平衡时需要 $q = d\cos\theta$ 及 $d = (b\cos\theta - a\sin\theta)/\cos 2\theta \approx 2.4641$。由此得 $p \approx 1.2321 > a$ 及 $q \approx 2.1340 > b$，与 (1) 吻合。以 $R = W\sin\theta$ 和 $S = W\cos\theta$ 对 G 取力矩，结果为零。
- 数值上，超过 10^5 个随机的平衡位置（随机选取 a、b、θ，而钉子落在边上）全部满足该恒等式，误差在 10^{-14} 以内，而且从不违反“$p < a \Rightarrow q < b$”。

常见错误

- 把钉子的反作用力当作铅垂。光滑接触的施力方向垂直于*接触面*（这里是薄片的边），而不是垂直于地面。
- 混淆了哪一条边与水平线成角 θ：$2a$ 边是与*铅垂线*成角 θ。

STEP 之外

三力原理是刚体（rigid body）静力学一般理论的特殊情况：一个力系处于平衡，当且仅当它的合力以及它对任意一点的合力矩同时为零。到了大学，这会发展成力螺旋（wrench）和虚功（virtual work）的语言。若用虚功的方法，可以在薄片与两根钉子保持接触的约束下，把 G 的高度减至最小，从而求出这个平衡位置。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对刚体静力学和力矩（torque）有严谨的论述。

详细解答

STEP II 1989，第 13 题

力学 · 转动惯量

> **STEP II 1989，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/89-s2-q13

本题在 *Beyond STEP* 第 11.7 节杆中的辛普森法则中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/simpsons-rod.html

关键思路

一切都建基于转动惯量（moment of inertia）的两个事实。平行轴定理（parallel axis theorem）指出，对任何一条轴的转动惯量，都由质量、质心（centre of mass）以及对通过质心的平行轴的转动惯量所决定。此外，转动惯量具有可加性（additive），所以我们可以把一个物体切成若干小块，再把每一块换成等效的东西。薄片（lamina）就是一叠幼杆，所以只要知道怎样替换一根杆，就能替换整块薄片。

解答

对每一条轴均等效

设 ℓ 为任意一条轴，并设 ℓ_O 为通过 O 的平行轴，与 ℓ 相距 d。物体与质点（particle）系统的质量同为 m，质心同为 O，所以可以对两者分别应用平行轴定理：

$$I_\ell^{\text{物体}} = I_{\ell_O}^{\text{物体}} + md^2, \qquad I_\ell^{\text{质点}} = I_{\ell_O}^{\text{质点}} + md^2.$$

由于 ℓ_O 通过 O，根据假设 $I_{\ell_O}^{\text{物体}} = I_{\ell_O}^{\text{质点}}$。因此两式的右边相等，于是对每一条轴 ℓ 都有 $I_\ell^{\text{物体}} = I_\ell^{\text{质点}}$，这正是所要证明的。

均匀杆

把均匀杆（uniform rod）沿 x 轴放置，由 $x=-a$ 伸展至 $x=a$，于是它的质心 O 就是原点。在 $x=-a$、0、a 处分别放置质量为 p、q、r 的质点。要令总质量为 m 且质心位于 O，我们需要

$$p+q+r=m, \qquad -pa+ra=0,$$

所以 $r=p$ 及 $q=m-2p$。

现在取任意一条通过 O、与杆成角 θ 的轴。杆上位于 x 的一点与这条轴的距离为 $|x|\sin\theta$。因此，对任何一条通过 O 的轴，转动惯量都是 $\sin^2\theta$ 乘以 $\sum(\text{质量})\,x^2$，我们只需令这一个和相等便可。对于杆，每个微元的质量为 $\frac{m}{2a}\,\mathrm{d}x$，所以

$$\int_{-a}^{a}\frac{m}{2a}\,x^2\,\mathrm{d}x=\frac{m}{2a}\cdot\frac{2a^3}{3}=\tfrac{1}{3}ma^2,$$

而对于质点，这个和是 $pa^2+q\cdot 0+pa^2=2pa^2$。当 $p=\frac{1}{6}m$ 时，两者对每个 θ 都相等。所以这根杆动力等效（dynamically equivalent）于

$$\tfrac{1}{6}m \text{ 于每一端}, \qquad \tfrac{2}{3}m \text{ 于中点}.$$

根据第一部分，两者对任何一条轴的转动惯量都相同，而不限于通过 O 的轴。

长方形薄片

设长方形的中心为 O，边长为 $2a$（平行于 x 轴）及 $2b$，于是它每单位面积的质量为 $M/(4ab)$。把它切成平行于 x 轴的幼条。位于高度 y、阔度为 δy 的幼条是一根长度为 $2a$、质量为 $\delta M=\frac{M}{2b}\,\delta y$ 的均匀杆。

> **想一想** 幼条的中心并不在 O。为甚么我们仍然可以把每条幼条换成它的三个质点？

第一部分的用处正在于此。每条幼条与它的三个质点对每一条轴的转动惯量都相同，而不只是对通过幼条本身中心的轴。转动惯量可以相加，所以对任何一条轴，薄片的转动惯量都与全部替换质点所组成的系统相同。这个系统的总质量同样是 M，质心同样是 O，因为每条幼条的替换质点与该幼条有相同的质量和质心。

位于高度 y 的幼条换成：在 $(\pm a,y)$ 两点各放 $\frac{1}{6}\delta M=\frac{M}{12b}\,\delta y$，以及在 $(0,y)$ 放 $\frac{2}{3}\delta M=\frac{M}{3b}\,\delta y$。当 y 由 $-b$ 变到 b 时，这些质点填满三根长度为 $2b$ 的均匀杆：

- 沿两边 $x=\pm a$ 各有一根，每根的质量为 $\frac{M}{12b}\cdot 2b=\frac{1}{6}M$；
- 沿中线 $x=0$ 有一根，质量为 $\frac{M}{3b}\cdot 2b=\frac{2}{3}M$。

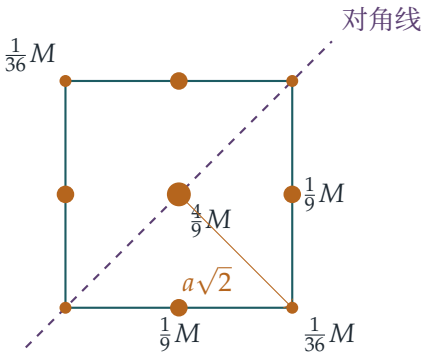
现在以同样的可加性论证，对这三根杆逐一应用均匀杆的结果：

- 每根边杆（$\frac{1}{6}M$）变成：在两端（即长方形的角）各有 $\frac{1}{6}\cdot\frac{1}{6}M=\frac{1}{36}M$，以及在中点（即长度为 $2b$ 的边的中点）有 $\frac{2}{3}\cdot\frac{1}{6}M=\frac{1}{9}M$；
- 中央的杆（$\frac{2}{3}M$）变成：在两端（即长度为 $2a$ 的边的中点）各有 $\frac{1}{6}\cdot\frac{2}{3}M=\frac{1}{9}M$，以及在中心有 $\frac{2}{3}\cdot\frac{2}{3}M=\frac{4}{9}M$。

所以每个角有 $\frac{1}{36}M$，每条边的中点有 $\frac{1}{9}M$，中心有 $\frac{4}{9}M$。总和为 $\frac{4}{36}M+\frac{4}{9}M+\frac{4}{9}M=M$，而由对称性可知质心为 O。这就是所求的等效系统。

正方形对其对角线

对于边长为 $2a$ 的正方形，把四个角放在 $(\pm a,\pm a)$，并取对角线 $y=x$。点 (x_0,y_0) 与这条直线的距离为 $|x_0-y_0|/\sqrt{2}$。



- 角 (a,a)、$(-a,-a)$ 和中心都在对角线上，没有贡献。
- 角 $(a,-a)$、$(-a,a)$ 与对角线的距离为 $2a/\sqrt{2}=a\sqrt{2}$。
- 每个中点 $(\pm a,0)$、$(0,\pm a)$ 与对角线的距离为 $a/\sqrt{2}$。

因此

$$I_{\text{对角线}}=2\cdot\tfrac{1}{36}M\cdot 2a^2+4\cdot\tfrac{1}{9}M\cdot\tfrac{1}{2}a^2=\tfrac{1}{9}Ma^2+\tfrac{2}{9}Ma^2=\tfrac{1}{3}Ma^2.$$

非均匀杆

设杆的长度为 $2a$，沿 x 轴由 $x=0$ 伸展至 $x=2a$，线密度（line density）为 kx。那么

$$m=\int_0^{2a}kx\,\mathrm{d}x=2ka^2,\qquad\text{所以}\qquad k=\frac{m}{2a^2}.$$

它的质心位于

$$\bar{x}=\frac{1}{m}\int_0^{2a}kx^2\,\mathrm{d}x=\frac{k}{m}\cdot\frac{8a^3}{3}=\frac{4a}{3},$$

并且

$$\int_0^{2a} kx^3\,\mathrm{d}x = \frac{k\,(2a)^4}{4} = 4ka^4 = 2ma^2.$$

与均匀杆的情况一样，所有点都在同一条直线上，所以对一条通过质心、与杆成角 θ 的轴，转动惯量为 $\sin^2\theta\sum(\text{质量})(x-\bar{x})^2$。若质量与质心都相同，则 $\sum(\text{质量})(x-\bar{x})^2 = \sum(\text{质量})x^2 - m\bar{x}^2$，所以只需令 $\sum(\text{质量})\,x^2$ 相等便已足够，而这正是对通过端点 $x=0$ 的垂直轴的转动惯量。

设质点为：$x=0$ 处的 p_0、$x=a$ 处的 p_1 及 $x=2a$ 处的 p_2。令质量、一阶矩（first moment）和二阶矩（second moment）分别相等，得出方程

$$\begin{aligned} p_0 + p_1 + p_2 &= m, \\ p_1 a + 2p_2 a &= \tfrac{4}{3}ma, \\ p_1 a^2 + 4p_2 a^2 &= 2ma^2. \end{aligned}$$

把第三式（除以 a^2）减去第二式（除以 a），得 $2p_2 = \frac{2}{3}m$，所以 $p_2 = \frac{1}{3}m$。于是 $p_1 = \frac{4}{3}m - \frac{2}{3}m = \frac{2}{3}m$，最后 $p_0 = m - \frac{2}{3}m - \frac{1}{3}m = 0$。

位于较轻一端的质点质量为零，所以只需要两个质点：中点的 $\frac{2}{3}m$ 和较重一端的 $\frac{1}{3}m$。

验算

- 边长为 $2a$ 的正方形薄片，对一条通过中心、位于薄片平面内且平行于一边的轴，其转动惯量的标准结果是 $\frac{1}{3}Ma^2$。根据垂直轴定理（perpendicular axis theorem）和对称性，每一条通过中心而位于平面内的轴都给出相同的值，所以对角线也必然如此。结果吻合。
- 对于非均匀杆，两个质点给出质量 m、一阶矩 $\frac{2}{3}m\cdot a + \frac{1}{3}m\cdot 2a = \frac{4}{3}ma$（所以质心位于 $\frac{4}{3}a$）以及二阶矩 $\frac{2}{3}ma^2 + \frac{4}{3}ma^2 = 2ma^2$，全部符合要求。
- 在数值上，我们用细密的点质量网格，把每个物体与它的质点系统对空间中 100–200 条随机的轴（不通过中心）作比较。均匀杆和非均匀杆的误差小于 10^{-7}，长方形的误差为 4×10^{-5}（网格误差），而正方形对其对角线在 $M=a=1$ 时给出 0.33332。

常见错误

- 把每条幼条换成它的三个质点，却没有说明为甚么可以这样做。幼条的中心并不在 O，所以你需要第一部分关于“任何轴”的结果。
- 在最后一部分，只施加质量和质心两个条件。两条方程不足以确定三个质量：正是转动惯量的条件迫使 $p_0 = 0$。

STEP 之外

对于在空间中运动的刚体（rigid body），在动力学上起决定作用的是它的质量、质心和惯性张量（inertia tensor）。惯性张量是一个对称的 3×3 矩阵，可给出对每一条通过质心的轴的转动惯量。两个物体若这三者都相同，在相同的力作用下运动便完全一样，这就是“动力等效”这个名称贴切的原因。用几个点质量去配合一个分布的头几阶矩，也是数值积分（numerical integration）背后的想法。均匀杆的权重 $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$ 正好就是辛普森法则（Simpson's rule），而若除了权重之外连位置也一并选取，便会引出高斯求积法（Gaussian quadrature）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics* 详细讲述了惯性张量和刚体运动。

详细解答

STEP II 1989，第 14 题

力学 · 圆周运动

STEP II 1989，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：

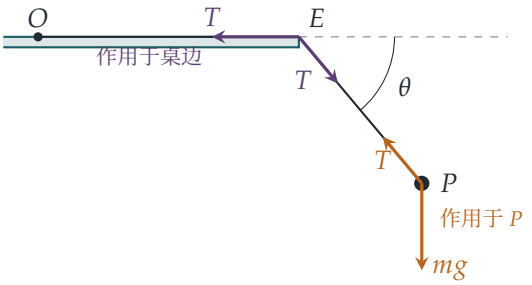
step.maths.org/questions/89-s2-q14

关键思路

质点一经释放，绳子留在桌面上的那一段便保持不动：它在固定端与桌边之间被拉紧。因此桌边相当于一个固定的支点，质点绕着它沿半径为 a 的圆运动，情况与由水平位置释放的单摆（simple pendulum）完全相同。由能量可求得速率，由径向运动方程（radial equation of motion）可求得张力（tension）。桌边是光滑的，所以绳子在桌边两侧的张力相同，而作用于桌边的力，就是沿两段绳子拉扯的两个相等张力的向量和。

解答

模型。 绳子是轻而不可伸长的（light and inextensible），桌面光滑而且很薄，质点是一个点质量（point mass）。设 O 为绳子的固定端，E 为绳子跨过桌边的那一点，于是 $OE = l - a$，而悬垂部分 EP 的长度为 a。设 θ 为悬垂部分从水平方向（即它起初所指、背离桌子的方向）向下转过的角。



由于 O 固定，而绳子一直保持拉紧（下文会验证 $T > 0$），OE 这一段留在桌面上静止不动。因此 P 沿以 E 为圆心、半径为 a 的圆运动。

速率。 桌面光滑，而张力始终与 P 的运动方向垂直，所以两者都不作功，机械能守恒（conservation of mechanical energy）。当绳子转过角 θ 时，P 下降了高度 $a\sin\theta$，故由静止开始，

$$\tfrac{1}{2}mv^2 = mga\sin\theta \qquad \Longrightarrow \qquad v^2 = 2ga\sin\theta.$$

张力。 沿 PE 向圆心 E 的方向分解：张力沿 PE 作用，而重量沿 PE 的分量为 $mg\sin\theta$，方向背离 E（当 $0<\theta<\pi$ 时，P 位于 E 的水平以下）。所以

$$T - mg\sin\theta = \frac{mv^2}{a} = 2mg\sin\theta, \qquad \text{由此得} \qquad \boxed{T = 3mg\sin\theta.}$$

当 $0<\theta<\pi$ 时这个值为正，所以绳子自始至终保持拉紧。在 $\theta=\pi$ 时有 $v=0$：质点刚好到达桌子底面、与 E 同高的位置，然后往回摆。因此运动发生在 $0\le\theta\le\pi$ 的范围内。

作用于桌边的力。 桌边光滑，所以绳子在桌边两侧的张力都是 T。绳子以两个大小为 T 的拉力压在桌边上：一个沿 EO，水平指向固定端；另一个沿 EP。取 x 轴为水平背离桌子的方向，y 轴为铅垂向上。沿 EO 和 EP 的单位向量（unit vector）分别是 $(-1,0)$ 和 $(\cos\theta,-\sin\theta)$，所以作用于桌边的力为

$$\mathbf{F} = T\,(\cos\theta - 1,\ -\sin\theta).$$

它的大小为

$$|\mathbf{F}| = T\sqrt{(1-\cos\theta)^2+\sin^2\theta} = T\sqrt{2-2\cos\theta} = 2T\sin\tfrac{\theta}{2}$$

这里用了 $1-\cos\theta = 2\sin^2\frac{\theta}{2}$，以及当 $0\le\theta\le\pi$ 时 $\sin\frac{\theta}{2}\ge 0$。（从几何上看，两个夹角为 $\pi-\theta$ 的相等的力 T，合力为 $2T\cos\frac{\pi-\theta}{2}$，结果相同。）代入 T 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$，

$$|\mathbf{F}| = 6mg\sin\theta\sin\tfrac{\theta}{2} = 12mg\,s^2\sqrt{1-s^2}, \qquad \text{其中} s=\sin\tfrac{\theta}{2}\in[0,1].$$

想一想 怎样才能干净俐落地求 $s^2\sqrt{1-s^2}$ 的最大值?

由于它是非负的，只需求它的平方的最大值，即 $h(u)=u^2(1-u)$，其中 $u=s^2\in[0,1]$。于是 $h'(u)=2u-3u^2=u(2-3u)$，当 $0<u<\frac{2}{3}$ 时为正，当 $\frac{2}{3}<u\le 1$ 时为负。所以 h 在 $[0,1]$ 上的最大值出现在 $u=\frac{2}{3}$，此时 $h=\frac{4}{9}\cdot\frac{1}{3}=\frac{4}{27}$。因此 $s^2\sqrt{1-s^2}$ 的最大值为 $\sqrt{4/27}=\dfrac{2}{3\sqrt{3}}$，而

$$|\mathbf{F}|_{\max} = 12mg\cdot\frac{2}{3\sqrt{3}} = \frac{8mg}{\sqrt{3}},$$

正如所求。它出现在 $\sin^2\frac{\theta}{2}=\frac{2}{3}$ 时，即 $\cos\theta = 1-2\sin^2\frac{\theta}{2} = -\frac{1}{3}$，所以 $\theta\approx 109.5°$：这时质点已经越过桌边的正下方，正往桌子底下摆回去。

验算

- 在 $\theta = \frac{\pi}{2}$（绳子铅垂）时，速率为 $\sqrt{2ga}$，而 $T = 3mg$，这正是由水平位置释放的单摆的熟悉结果。这时作用于桌边的力为 $T\sqrt{2} = 3\sqrt{2}\,mg \approx 4.24mg$，略小于 $8mg/\sqrt{3} \approx 4.62mg$，与最大值出现在铅垂位置之后这一点吻合。
- 在 $\theta = 0$ 和 $\theta = \pi$ 时，质点在与桌边同高的位置瞬时静止，所以 $T = 0$，桌边不受任何力，与公式所示一致。
- 我们对运动方程 $\ddot{\theta} = (g/a)\cos\theta$ 作了数值积分：由径向方程求得的 T 与 $3mg\sin\theta$ 相符，误差在 10^{-12} 以内；而作用于桌边的力的最大值出现在 $\theta \approx 1.9106$（即 $\arccos(-\frac{1}{3})$），其值为 $4.6188\,mg$，与 $8/\sqrt{3}$ 相符。

常见错误

- 把作用于桌边的力当作只有 T，或只取悬垂绳子拉力的铅垂部分。桌面上的那段绳子也会水平地拉扯桌边，两个拉力必须以向量方式相加。
- 求最大值时只考虑到 $\theta \le \frac{\pi}{2}$ 为止。质点会继续摆到桌子底下，而最大值正是出现在这个较后的阶段。

STEP 之外

这里的张力是一种*约束力*（constraint force）：它不作功，其大小恰好足以令质点保持在圆周上。在拉格朗日力学（Lagrangian mechanics）中，这类力会从运动方程中完全消失，事后才再求回来（就像我们利用径向方程所做的那样），或者藉拉格朗日乘数（Lagrange multipliers）求得。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）在这个程度上对圆周运动、约束和能量方法有细致的论述。

详细解答

STEP II 1989，第 15 题

概率与统计 · 几何概率

STEP II 1989，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/89-s2-q15

关键思路

“在周界上随机”选取的一点，是按长度均匀分布（uniform distribution）的，所以一切都由总长度 $\pi+2$ 出发。三角形总有一个顶点在圆心 O，所以它的面积是 $\frac{1}{2}ab\sin C$，其中 a 和 b 是两点与 O 的距离，C 是在 O 处的夹角。这样，每个条件情况都化为一个简单的期望值（expectation），再用全期望定律（law of total expectation）把它们合并起来。

解答

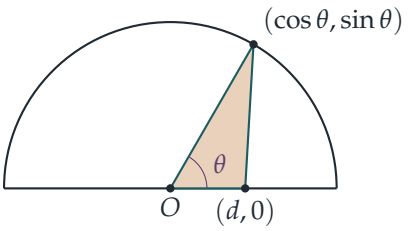
把 O 放在原点，直径沿 x 轴由 $(-1,0)$ 到 $(1,0)$，圆弧位于 $y\geq 0$。周界的长度是 $2+\pi$：直径占 2，圆弧占 π。由于每一点都沿周界均匀分布，

$$\mathrm{P}(\text{一点落在直径上})=\frac{2}{\pi+2},\qquad \mathrm{P}(\text{一点落在圆弧上})=\frac{\pi}{\pi+2}.$$

已知该点落在圆弧上时，它是 $(\cos\theta,\sin\theta)$，其中 θ 在 $[0,\pi]$ 上均匀分布。已知该点落在直径上时，它是 $(d,0)$，其中 d 在 $[-1,1]$ 上均匀分布。

两点互相独立（independent），而“恰好一点在直径上”可以是第一点在直径上，也可以是第二点在直径上，所以

$$\mathrm{P}(\text{恰好一点在直径上})=2\cdot\frac{2}{\pi+2}\cdot\frac{\pi}{\pi+2}=\frac{4\pi}{(\pi+2)^2}.$$



第 (i) 部分

设其中一点是直径上的 $(d,0)$，另一点是圆弧上的 $(\cos\theta,\sin\theta)$。由 O 出发的两条边长度分别是 $|d|$ 和 1，它们之间的夹角是 θ（若 $d<0$，夹角是 $\pi-\theta$，其正弦值相同）。所以

$$A=\tfrac{1}{2}\,|d|\sin\theta.$$

在这个情况下，d 和 θ 互相独立，所以乘积的期望值等于期望值的乘积。由于 d 在 $[-1,1]$ 上均匀分布，$|d|$ 在 $[0,1]$ 上均匀分布，故 $\mathrm{E}|d|=\frac{1}{2}$。另外，

$$\mathrm{E}(\sin\theta)=\frac{1}{\pi}\int_0^{\pi}\sin\theta\,\mathrm{d}\theta=\frac{2}{\pi}.$$

因此

$$\mathrm{E}(A\mid\text{恰好一点在直径上})=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{\pi}=\frac{1}{2\pi},$$

正是所求。

第 (ii) 部分

现在两点都在圆弧上，角度分别是 θ_1 和 θ_2，两者互相独立，并在 $[0,\pi]$ 上均匀分布。两点与 O 的距离都是 1，而它们在 O 处的夹角是 $X=|\theta_1-\theta_2|$，位于 $[0,\pi]$ 之内。所以

$$A=\tfrac{1}{2}\sin X.$$

? **想一想** 为甚么我们可以使用题目给出的 X 的密度函数?

角 θ 量度的是沿单位半圆的弧长，所以在圆弧上均匀地选取两点，等同于在一条长度为 π 的线段上均匀地选取两点。角 X 正是两点之间的距离，所以 X 在 $[0,\pi]$ 上的密度函数（density）是 $2(\pi-x)/\pi^2$。于是

$$\mathrm{E}(\sin X)=\frac{2}{\pi^2}\int_0^{\pi}(\pi-x)\sin x\,\mathrm{d}x.$$

把积分拆开，$\int_0^{\pi}\pi\sin x\,\mathrm{d}x=2\pi$；再取 $u=x$ 作分部积分（integration by parts），

$$\int_0^{\pi}x\sin x\,\mathrm{d}x=\Big[-x\cos x\Big]_0^{\pi}+\int_0^{\pi}\cos x\,\mathrm{d}x=\pi+0=\pi.$$

所以 $\int_0^\pi (\pi - x)\sin x\,\mathrm{d}x = 2\pi - \pi = \pi$，由此得 $\mathrm{E}(\sin X) = \dfrac{2}{\pi}$，以及

$$\mathrm{E}(A \mid \text{两点都不在直径上}) = \frac{1}{2}\cdot\frac{2}{\pi} = \frac{1}{\pi},$$

正是所求。

A 的期望值

还有第三种情况：若两点都落在直径上，这个“三角形”是退化（degenerate）的（三点都在 x 轴上），所以 $A = 0$。两点都不在直径上的概率是 $\left(\frac{\pi}{\pi+2}\right)^2$。根据全期望定律，

$$\begin{aligned}\mathrm{E}(A) &= \frac{4\pi}{(\pi+2)^2}\cdot\frac{1}{2\pi} + \frac{\pi^2}{(\pi+2)^2}\cdot\frac{1}{\pi} + \frac{4}{(\pi+2)^2}\cdot 0\\ &= \frac{2+\pi}{(\pi+2)^2}\\ &= \frac{1}{2+\pi},\end{aligned}$$

这就是我们要证明的结果。

验算

- 三种情况的概率 $\frac{4}{(\pi+2)^2}$、$\frac{4\pi}{(\pi+2)^2}$ 和 $\frac{\pi^2}{(\pi+2)^2}$ 相加得 $\frac{(2+\pi)^2}{(\pi+2)^2} = 1$。
- $\mathrm{E}(A) \le \frac{1}{2}$，即不超过最大可能的面积（两条半径互相垂直），这是必然的；事实上 $\frac{1}{2+\pi} \approx 0.1945$。
- 以蒙地卡罗模拟（Monte Carlo simulation）抽取 2×10^6 对点，得出 P(恰好一点) ≈ 0.4757（准确值为 0.4754），条件平均值为 0.1591 和 0.3184（准确值为 $\frac{1}{2\pi} \approx 0.1592$ 和 $\frac{1}{\pi} \approx 0.3183$），以及 $\mathrm{E}(A) \approx 0.1944$。

常见错误

- 在 P(恰好一点) 中漏掉因子 2：两点之中任何一点都可以是落在直径上的那一点。
- 把“在直径上”的概率当作 $\frac{1}{2}$：选点是按长度均匀分布，而不是按落在边界的哪一段来平均分配。
- 在第 (ii) 部分使用 $\sin\theta_1\sin\theta_2$ 或类似的式子，而不是两条半径之间夹角的正弦。

STEP 之外

把期望值按各种情况的划分（partition）拆开，即 $\mathrm{E}(A) = \sum_i \mathrm{E}(A \mid B_i)\,\mathrm{P}(B_i)$，就是全期望定律，它是大学一年级概率论中最常用的工具之一；其一般形式 $\mathrm{E}(A) = \mathrm{E}\big(\mathrm{E}(A \mid Y)\big)$ 更是无处不在，由分支过程（branching processes）到马可夫链（Markov chains）都会用到。第 (ii) 部分引用的 X 的密度函数，本身就是一道不错的练习：它来自正方形 $[0, \pi]^2$ 内一条带状区域的面积。这类题目属于几何概率（geometric probability）；Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）以合适的程度讲解条件期望（conditional expectation）和联合密度函数（joint densities）。

详细解答

STEP II 1989，第 16 题

概率与统计 · 二项分布

STEP II 1989，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/89-s2-q16

关键思路

把整批产品分成表现不同的两部分：样本中的 n 件小配件，以及从未被抽查的 N 件。每一部分的次品数目都服从二项分布（binomial distribution），而且两个数目互相独立（independent）。一切取决于样本：如果样本中有两件或以上次品，整批最终会完美无瑕，否则一切维持不变。因此，若一批产品最后有 $k \geq 1$ 件次品，样本中必定只有 0 或 1 件次品，其余的次品则在另外 N 件之中。

解答

记 $q = 1 - p$。设 X 为样本中次品的数目，Y 为该批其余 N 件中次品的数目。每件小配件各自独立地以概率（probability）p 为次品，所以

$$X \sim \mathrm{B}(n, p), \qquad Y \sim \mathrm{B}(N, p),$$

而且 X 与 Y 独立，因为它们所数算的是两组互不相交的小配件中的次品。

设 K 为经过品质管制后该批产品中次品的数目。若 $X \geq 2$，所有次品都会被修正，于是 $K = 0$。若 $X \leq 1$，则甚么都不会修正，于是 $K = X + Y$。所以

$$K = \begin{cases} X + Y, & X \leq 1, \\ 0, & X \geq 2. \end{cases}$$

K 的分布

取 $1 \le k \le N$。由于 $k \ge 1$，事件 $K = k$ 不可能来自 $X \ge 2$ 的情况，所以它恰好在 $X = 0,\ Y = k$ 或 $X = 1,\ Y = k-1$ 时发生。这两个事件互斥，而由独立性可得

$$\begin{aligned}\mathrm{P}(K=k) &= \mathrm{P}(X=0)\,\mathrm{P}(Y=k) + \mathrm{P}(X=1)\,\mathrm{P}(Y=k-1)\\ &= q^n\binom{N}{k}p^k q^{N-k} + npq^{n-1}\binom{N}{k-1}p^{k-1}q^{N-k+1}\\ &= \left[\binom{N}{k} + n\binom{N}{k-1}\right]p^k q^{N+n-k}.\end{aligned}$$

两项中 p 的幂次相同，q 的幂次也相同，所以才能合并。现在把二项式系数（binomial coefficient）通分，公分母为 $k!\,(N-k+1)!$：

$$\binom{N}{k} = \frac{N!\,(N-k+1)}{k!\,(N-k+1)!}, \qquad n\binom{N}{k-1} = \frac{n\,N!\,k}{k!\,(N-k+1)!}.$$

相加后，分子为 $N!\,(N-k+1+nk) = N!\,[N+1+k(n-1)]$，所以

$$\mathrm{P}(K=k) = \frac{[N+1+k(n-1)]\,N!}{(N-k+1)!\,k!}\,p^k(1-p)^{N+n-k},$$

正如所求。

> **想一想** 当 $k = 0$ 时会怎样?

这条公式只在 $k \ge 1$ 时成立。当 $k = 0$ 时，它给出 $\frac{(N+1)N!}{(N+1)!}q^{N+n} = q^{N+n}$，这是完全没有次品的概率。但只要 $X \ge 2$，$K = 0$ 同样会发生，所以实际上

$$\mathrm{P}(K=0) = q^{N+n} + \mathrm{P}(X \ge 2) = q^{N+n} + 1 - q^n - npq^{n-1}.$$

题目中的“$k \le N$”应理解为 $1 \le k \le N$。

$k = N+1$ 的情况

要使 $K = N+1$，需要 $X \le 1$ 且 $X + Y = N+1$。由于 $Y \le N$，这就迫使 $X = 1$ 且 $Y = N$，所以

$$\mathrm{P}(K=N+1) = npq^{n-1}\cdot p^N = np^{N+1}q^{n-1}.$$

把 $k = N+1$ 代入公式，得到系数

$$\frac{[N+1+(N+1)(n-1)]\,N!}{0!\,(N+1)!} = \frac{(N+1)\,n\,N!}{(N+1)!} = n,$$

而 q 的幂为 $q^{N+n-(N+1)} = q^{n-1}$。所以公式给出 $np^{N+1}q^{n-1}$，结果正确。（当 $k \ge N+2$ 时概率为零，因为 $K \le N+1$。）

次品数目的期望值

我们可以对 $k\,\mathrm{P}(K=k)$ 求和，但利用以 X 和 Y 表示 K 的写法会快得多。以 $\mathbf{1}_{\{X\le 1\}}$ 表示事件 $X\le 1$ 的指示变量（indicator），则

$$K=(X+Y)\,\mathbf{1}_{\{X\le 1\}}=X\,\mathbf{1}_{\{X\le 1\}}+Y\,\mathbf{1}_{\{X\le 1\}}.$$

对于第一项，$X\,\mathbf{1}_{\{X\le 1\}}$ 在 $X=1$ 时等于 1，其余情况等于 0，所以它的期望值（expectation）是 $\mathrm{P}(X=1)=npq^{n-1}$。对于第二项，$Y$ 与 X 独立，所以

$$\mathrm{E}\left[Y\,\mathbf{1}_{\{X\le 1\}}\right]=\mathrm{E}[Y]\,\mathrm{P}(X\le 1)=Np\left(q^n+npq^{n-1}\right).$$

相加，并提出因子 pq^{n-1}，

$$\begin{aligned}\mathrm{E}[K]&=npq^{n-1}+Np\,q^{n-1}(q+np)\\&=pq^{n-1}\left[n+N(1-p)+Nnp\right]\\&=\left[N+n+pN(n-1)\right]p(1-p)^{n-1},\end{aligned}$$

正如所求。

验算

- 各概率之和应为 1。取 $n=2$、$N=1$：$\mathrm{P}(K=1)=(2+1)pq^2=3pq^2$，$\mathrm{P}(K=2)=2p^2q$，以及 $\mathrm{P}(K=0)=q^3+p^2$，三者相加得 $q^3+3pq^2+3p^2q+p^3=1$（用到 $p^2=p^2(p+q)$）。
- 若 n 极大，样本中有次品的批次几乎全部都会被修正，期望值趋向 0，这与因子 $(1-p)^{n-1}$ 吻合。
- 我们以精确分数验算了每一条公式，范围是 $2\le n\le 6$、$0\le N\le 6$ 以及若干个 p 值，并且模拟了整个程序：取 $n=4$、$N=6$、$p=0.3$ 时，公式给出 $\mathrm{E}[K]\approx 1.585$，而模拟 200 000 批所得的结果是 1.581。

常见错误

- 把“两件或以上”的规则套用在整批的次品总数上。次品会否被修正，只由样本中的数目 X 决定。
- 忘记 $K=0$ 还有来自已修正批次的额外贡献，所以公式在 $k=0$ 时不成立。
- 以硬算求和的方法计算 $\mathrm{E}[K]$。这样做是可行的，但用指示变量把 K 拆开，再利用独立性，会简短得多。

STEP 之外

把一个随机变量（random variable）写成若干部分各自乘以指示变量之和，再运用期望值的线性性质（linearity of expectation），是大学一年级概率论中最有用的技巧之一。当只需要平均值时，这个方法可以免却求出整个分布的工夫。这道题目也是品质管制中验收抽样（acceptance sampling）的一个小型模型，当中我们要选择 n，在检测成本与漏网的次品之间取得平衡。想进一步了解，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导论），后者全书贯彻使用指示变量。

详细解答

STEP III 1989，第 1 题

纯数学 · 旋转曲面

> **STEP III 1989，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/89-s3-q1

关键思路

球面是半圆的旋转曲面（surface of revolution），所以球带（zone）的面积是一个标准的旋转曲面积分。出人意表的是，被积函数（integrand）原来是常数，这正是答案只取决于两平面之间距离的原因。在第二部分，曲面分成一个圆锥（由 OP 生成）和一个球带（由弧 PR 生成）。这时第一个结果便担起了最吃力的工作，而相切的条件则提供了我们所需的直角三角形。

关于题目印出的方程，要说明一下：若右边是 c^2，圆便会通过 O，而常数 a 则从未出现。题目原意的曲线显然是

$$(x-c)^2+y^2=a^2,$$

即圆心为 $C=(c,0)$、半径为 a 的圆。由于 $c>a$，原点在圆外，所以由 O 出发的切线（tangent）存在。我们就这条曲线来解题。

解答

球带的面积

取一个半径为 r、球心在原点的球面，并选取 x 轴与两个平面垂直，这两个平面是 $x=h_1$ 和 $x=h_2$，其中 $-r\le h_1<h_2\le r$。球面由上半圆 $y=\sqrt{r^2-x^2}$ 绕 x 轴旋转 2π 而成，所以球带的面积是

$$S=\int_{h_1}^{h_2}2\pi y\sqrt{1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2}\,\mathrm{d}x.$$

这里 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -\dfrac{x}{\sqrt{r^2-x^2}}$，所以当 $|x| < r$ 时

$$y\sqrt{1+\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2} = \sqrt{r^2-x^2}\sqrt{\frac{r^2-x^2+x^2}{r^2-x^2}} = \sqrt{r^2-x^2}\cdot\frac{r}{\sqrt{r^2-x^2}} = r.$$

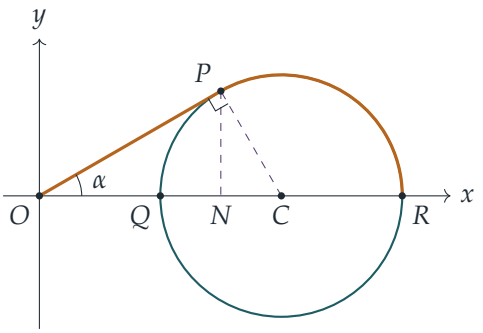
? **想一想** 在 $x = \pm r$ 处，$\mathrm{d}y/\mathrm{d}x$ 是无穷大，这时会怎样?

导数（derivative）在两极处趋向无穷大，但乘积 $y\sqrt{1+(y')^2}$ 在整个 $(-r, r)$ 上都等于 r，所以它是有界的。若其中一个平面与球面相切（$h_1 = -r$ 或 $h_2 = r$），这个积分虽属反常积分（improper integral），却无伤大雅：先积分至 $h_2 - \varepsilon$，再令 $\varepsilon \to 0$ 即可。（另一个做法是以 $x = r\cos\theta$、$y = r\sin\theta$ 作参数化（parametrise），这样便完全避开了这个问题。）因此

$$S = \int_{h_1}^{h_2} 2\pi r\,\mathrm{d}x = 2\pi r\,(h_2 - h_1),$$

也就是 $2\pi \times$ (半径) $\times$ (两平面之间的距离)，正是所求。

曲面



设 $C = (c, 0)$ 为圆心，N 为由 P 至 x 轴的垂足。那么 $Q = (c-a, 0)$，$R = (c+a, 0)$。弧 PR 是指不经过 Q 的那一段（图中以琥珀色标示）：由 P 到 R 的另一段弧会经过 Q 并绕过圆的底部，把它旋转便会扫出整个球面。

切线与过切点的半径互相垂直，所以角 OPC 是直角。在三角形 OPC 中，$OC = c$ 而 $CP = a$，所以由毕氏定理（Pythagoras' theorem）得

$$OP = \sqrt{c^2 - a^2}.$$

设 α 为角 POC。那么 $\sin\alpha = \frac{a}{c}$ 而 $\cos\alpha = \frac{\sqrt{c^2-a^2}}{c}$，由此得

$$PN = OP\sin\alpha = \frac{a\sqrt{c^2-a^2}}{c}, \qquad ON = OP\cos\alpha = \frac{c^2-a^2}{c}.$$

圆锥。把 OP 绕 x 轴旋转，得到一个斜高（slant height）为 $l = OP$、底半径为 PN 的圆锥。它的曲面面积是 $\pi \times$ (底半径) $\times$ (斜高)，即

$$\pi \cdot \frac{a\sqrt{c^2 - a^2}}{c} \cdot \sqrt{c^2 - a^2} = \frac{\pi a(c^2 - a^2)}{c}.$$

球带。把弧 PR 旋转，得到球面（球心 C，半径 a）上夹在两个平面之间的部分：一个是过 N 且垂直于轴的平面，另一个是在 R 处的切平面（tangent plane）。这两个平面相距

$$NR = (c + a) - \frac{c^2 - a^2}{c} = \frac{ac + a^2}{c} = \frac{a(c + a)}{c}$$

，所以由第一个结果，球带的面积是

$$2\pi a \cdot \frac{a(c + a)}{c} = \frac{2\pi a^2(c + a)}{c}.$$

总和。把两部分相加，

$$\begin{aligned} \text{面积} &= \frac{\pi a}{c}\left[(c^2 - a^2) + 2a(c + a)\right] \\ &= \frac{\pi a}{c}\left(c^2 + 2ac + a^2\right) \\ &= \frac{\pi a(c + a)^2}{c}. \end{aligned}$$

验算

- 取 $a = 1$，$c = 2$。那么 $P = \left(\frac{3}{2}, \frac{\sqrt{3}}{2}\right)$，圆锥的面积是 $\frac{3\pi}{2}$，球带的面积是 3π，所以总面积是 $\frac{9\pi}{2} \approx 14.137$。以 20 000 条短线段逼近曲线，再把所得各个圆台（frustum）的面积相加，得 14.13717。
- 当 $a \to c$ 时，P 趋近 O，圆锥随之消失。曲面变成半径为 c 的整个球面，而果然 $\pi c(2c)^2/c = 4\pi c^2$。
- 当 a 固定而 $c \to \infty$ 时，面积的增长与 πac 相若，这正是一个斜高约为 c、底半径约为 a 的细长圆锥的面积。这与我们的预期相符。

常见错误

- 以 OC 或 ON 代替 OP 作为圆锥的斜高。
- 由 C 或 Q 起量度球带的厚度，而不是由 N（即由 P 所作垂线的垂足）起量度。
- 使用公式 $2\pi rh$ 时取 $r = c$ 而非 $r = a$：被截的球面是以 C 为球心、半径为 a 的那一个，而不是任何以 O 为球心的球面。

STEP 之外

球带的结果就是阿基米德（Archimedes）定理。把球面沿水平方向投影到外接圆柱上，面积保持不变，这就是兰伯特（Lambert）等积地图投影背后的想法。它也解释了为甚么在球面上均匀地选取一点时，该点的高度服从均匀分布（uniform distribution）。旋转曲面的面积，是大学一年级向量微积分（vector calculus）课程中曲面积分（surface integral）$\iint |\mathbf{r}_u \times \mathbf{r}_v| \, \mathrm{d}u \, \mathrm{d}v$ 的第一个例子。若想从优美的几何角度认识曲率（curvature）与曲面，可参阅 Needham 的 *Visual Differential Geometry and Forms*（以图像讲解微分几何与微分形式）。

详细解答

STEP III 1989，第 2 题

纯数学 · 三维空间中的平面

> **STEP III 1989，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/89-s3-q2

关键思路

引入坐标：东、北和上。这样每个平面都是一个图像 $z = px + qy + r$，而三个点就能确定一个平面。平面有多陡，只取决于 z 沿水平方向变化得多快，所以所求的角可由 p 和 q 得出。第二部分有一条巧妙的捷径：矿层在地面以下的深度同样是位置的线性函数（linear function），而矿层恰好在深度为零之处露出地面。

解答

建立坐标

取 A 为原点，x 轴指向东，y 轴指向北，z 轴铅垂向上（所有长度以米为单位）。于是

$$A = (0,0,0), \qquad B = (0,100,-5), \qquad C = (60,100,-10),$$

因为 B 比 A 低 5 m，而 C 又比 B 再低 5 m。钻孔（bore-hole）是铅垂的，所以钻孔碰到矿层（seam）的点就在这些点的正下方：

$$A' = (0,0,-95), \qquad B' = (0,100,-50), \qquad C' = (60,100,-86).$$

矿层的倾角

矿层并非铅垂，因为 A', B', C' 位于三个在平面图（plan）上不共线（collinear）的点的上方。所以我们可以把它写成 $z = px + qy + r$。

- 在 A'：$r = -95$。
- 由 A' 到 B'，y 增加 100，z 增加 45，所以 $100q = 45$，即 $q = \frac{9}{20}$。

▶ 由 B' 到 C'，x 增加 60，z 的变化为 -36，所以 $60p = -36$，即 $p = -\frac{3}{5}$。

因此矿层为

$$z = -\tfrac{3}{5}x + \tfrac{9}{20}y - 95,$$

它的一个法向量（normal vector）是 $\mathbf{n} = \left(\frac{3}{5}, -\frac{9}{20}, 1\right)$，或者乘以 20 后得 $(12, -9, 20)$。

两个平面之间的夹角等于它们的法线之间的夹角。水平面的法向量是 $\mathbf{k} = (0,0,1)$，所以矿层与水平面的夹角 θ 满足

$$\cos\theta = \frac{|(12,-9,20)\cdot(0,0,1)|}{\sqrt{12^2+9^2+20^2}} = \frac{20}{\sqrt{144+81+400}} = \frac{20}{\sqrt{625}} = \frac{20}{25} = \frac{4}{5}.$$

故此矿层与水平面的倾角为 $\cos^{-1}\left(\frac{4}{5}\right)$，正如题目所求。

（另一个看法：沿最大斜率线（line of greatest slope），水平方向每走一米，z 的变化为 $\sqrt{p^2+q^2} = \sqrt{\frac{9}{25} + \frac{81}{400}} = \frac{3}{4}$。所以 $\tan\theta = \frac{3}{4}$，是一个 3–4–5 三角形，于是 $\cos\theta = \frac{4}{5}$。）

矿层在何处露出地面

地面是一个通过原点的平面。与之前的做法一样，由 A 向北走到 B，地面每 100 下降 5；由 B 向东走到 C，地面每 60 下降 5，所以

$$z = -\tfrac{1}{12}x - \tfrac{1}{20}y.$$

想一想 我们真的需要联立求解两个平面方程吗?

不尽然。设 $d(x,y)$ 为在水平位置 (x,y) 处矿层在地面以下的深度。它是两个线性函数之差，所以本身也是线性的：

$$d(x,y) = \left(-\tfrac{1}{12}x - \tfrac{1}{20}y\right) - \left(-\tfrac{3}{5}x + \tfrac{9}{20}y - 95\right) = \tfrac{31}{60}x - \tfrac{1}{2}y + 95.$$

作为验算，$d(0,0) = 95$、$d(0,100) = 45$ 及 $d(60,100) = 45 + 31 = 76$，正是三个钻孔的深度。（其实我们可以直接由这些数据写出 d：向北每走 100 m，深度减少 50；向东每走 60 m，深度增加 31。）

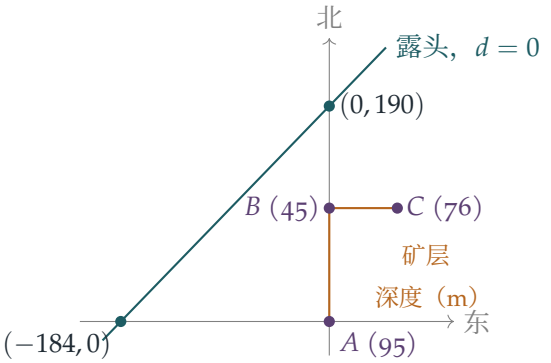
矿层在 $d = 0$ 之处露出地面，在平面图上这是直线

$$\tfrac{31}{60}x - \tfrac{1}{2}y + 95 = 0, \qquad \text{即} \qquad y = \tfrac{31}{30}x + 190.$$

沿这条直线，每向东移动 30 m，就同时向北移动 31 m。所以这条直线与正北方向成角 φ，其中

$$\tan\varphi = \frac{\text{东}}{\text{北}} = \frac{30}{31}, \qquad \varphi = \tan^{-1}\left(\tfrac{30}{31}\right) \approx 44.06^\circ.$$

因此这条直线的方位角（bearing）是 $\tan^{-1}\left(\frac{30}{31}\right) \approx 044^\circ$。若从相反方向来读，则是 224°；一条直线有这两个方位角，用哪一个来描述它都可以。这条直线的走向略偏于东北方向以西，在平面图上它通过 A 正北 190 m 处。



验算

- 矿层的法向量 $(12, -9, 20)$ 的长度恰好是 25，所以 $\cos\theta = \frac{20}{25}$，是一个整齐的 3–4–5 答案，与题目所预告的一致。
- 深度公式 d 重现了全部三个钻孔的深度，所以两个平面都是对的。
- 数值上，$\tan^{-1}(30/31) = 44.0608°$。在直线 $y = \frac{31}{30}x + 190$ 上随机取点，两个平面给出的高度相同，只差舍入误差。

常见错误

- 把每个钻孔由孔口起计的深度直接拿来用，仿佛各孔口处于同一水平。B 和 C 分别比 A 低 5 m 和 10 m，所以在该两处矿层位于 A 的水平以下 50 m 和 86 m，而不是 45 m 和 76 m。
- 给出矿层内一条水平线的方位角（即 $-\frac{3}{5}x + \frac{9}{20}y$ 为常数之处，方位角为 $\tan^{-1}\frac{3}{4} \approx 037°$），而不是露头（outcrop）的方位角。由于地面也是倾斜的，两者并不相同。
- 把 $\tan\varphi =$ 东/北 与北/东混淆。方位角是由北起顺时针方向量度的。

STEP 之外

这是地质学家的“走向与倾角”（strike and dip）问题，所用的工具属于大学一年级的向量几何：以 $\mathbf{n}\cdot\mathbf{r} = c$ 表示平面、通过法线求平面之间的夹角，以及交线。深度这一招是一个普遍原理的例子：两个仿射函数（affine function）之差仍是仿射函数，所以两个不平行平面的交集就是某一个线性函数的零点集（zero set）。Strang 的 *Introduction to Linear Algebra*（线性代数入门）把这幅图像融入线性方程的几何之中。

详细解答

STEP III 1989，第 3 题

纯数学 · 矩阵群

> **STEP III 1989，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/89-s3-q3

关键思路

$\mathbf{M}$ 是转角为 $2\pi/m$ 的旋转（rotation），所以 $\mathbf{M}^m = \mathbf{I}$，而它各次幂之和是一个等比级数（geometric series）：把这个和乘以 $\mathbf{I}-\mathbf{M}$，各项便会逐一相消（telescope）。至于递推关系（recurrence），只要写出 $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$，规律便一目了然。之后的群（group）全靠一条恒等式：把其中一个表达式代入另一个，结果只是把下标相加，所以这些元素的运算方式就像模（modulo）m 的整数一样。

解答

$\mathbf{M}$ 的各次幂之和

$\mathbf{M}$ 是绕原点逆时针旋转 $\theta = 2\pi/m$ 的矩阵。连续施行 k 次，就是旋转 $k\theta$，所以

$$\mathbf{M}^k = \begin{pmatrix} \cos k\theta & -\sin k\theta \\ \sin k\theta & \cos k\theta \end{pmatrix}, \qquad \text{特别地} \qquad \mathbf{M}^m = \begin{pmatrix} \cos 2\pi & -\sin 2\pi \\ \sin 2\pi & \cos 2\pi \end{pmatrix} = \mathbf{I}.$$

（由于 $\mathbf{M}^{k+1} = \mathbf{M}\mathbf{M}^k$，$\mathbf{M}^k$ 的公式可用归纳法（induction）由复角公式（compound-angle formulae）推出。）

设 $\mathbf{S} = \mathbf{I} + \mathbf{M} + \cdots + \mathbf{M}^{m-1}$。这些矩阵全是 $\mathbf{M}$ 的幂，所以它们可交换（commute），而乘积中各项逐一相消：

$$(\mathbf{I}-\mathbf{M})\mathbf{S} = \mathbf{I} - \mathbf{M}^m = \mathbf{I} - \mathbf{I} = \mathbf{O}.$$

? **想一想** 我们可以直接“约去”因子 $\mathbf{I}-\mathbf{M}$ 吗？

只有当 $\mathbf{I}-\mathbf{M}$ 可逆（invertible）时才可以：非零矩阵的乘积可以是 $\mathbf{O}$。这里

$$\det(\mathbf{I}-\mathbf{M}) = (1-\cos\theta)^2 + \sin^2\theta = 2-2\cos\theta,$$

而由于 $m>1$，我们有 $0<\theta\le\pi$，所以 $\cos\theta\neq 1$，行列式（determinant）不等于零。在 $(\mathbf{I}-\mathbf{M})\mathbf{S}=\mathbf{O}$ 的左边乘以 $(\mathbf{I}-\mathbf{M})^{-1}$，便得 $\mathbf{S}=\mathbf{O}$，这正是所求的结果。

$\mathbf{X}_k$ 的公式

算出最初几项，

$$\mathbf{X}_1 = \mathbf{P}\mathbf{X}_0+\mathbf{Q},\quad \mathbf{X}_2 = \mathbf{P}^2\mathbf{X}_0+\mathbf{P}\mathbf{Q}+\mathbf{Q},\quad \mathbf{X}_3 = \mathbf{P}^3\mathbf{X}_0+(\mathbf{P}^2+\mathbf{P}+\mathbf{I})\mathbf{Q},$$

由此可猜想，对 $k\ge 0$，

$$\mathbf{X}_k = \mathbf{P}^k\mathbf{X}_0+\mathbf{S}_k\mathbf{Q},\qquad \text{其中}\quad \mathbf{S}_k = \mathbf{I}+\mathbf{P}+\cdots+\mathbf{P}^{k-1}\quad(\mathbf{S}_0=\mathbf{O}).$$

当 $k=0$ 时，这式子就是 $\mathbf{X}_0=\mathbf{X}_0$，显然成立。假设它对某个 $k\ge 0$ 成立，则

$$\begin{aligned}\mathbf{X}_{k+1} &= \mathbf{P}\mathbf{X}_k+\mathbf{Q} = \mathbf{P}^{k+1}\mathbf{X}_0+(\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^k)\mathbf{Q}+\mathbf{Q}\\ &= \mathbf{P}^{k+1}\mathbf{X}_0+\mathbf{S}_{k+1}\mathbf{Q},\end{aligned}$$

所以它对 $k+1$ 也成立，按归纳法，它对所有 $k\ge 0$ 都成立。要保持因子的次序：$\mathbf{P}$ 自始至终都是从左边乘上去的，而 $\mathbf{P}$ 与 $\mathbf{Q}$ 不一定可交换。

群

把 $\mathbf{X}_i$ 看成表达式 $\mathbf{P}^i\mathbf{X}_0+\mathbf{S}_i\mathbf{Q}$，即一条把任意矩阵 $\mathbf{X}_0$ 变成另一个矩阵的规则。以 $\mathbf{X}_j$ 代替 $\mathbf{X}_0$，得

$$\begin{aligned}\mathbf{X}_i * \mathbf{X}_j &= \mathbf{P}^i\left(\mathbf{P}^j\mathbf{X}_0+\mathbf{S}_j\mathbf{Q}\right)+\mathbf{S}_i\mathbf{Q}\\ &= \mathbf{P}^{i+j}\mathbf{X}_0+\left(\mathbf{P}^i\mathbf{S}_j+\mathbf{S}_i\right)\mathbf{Q}.\end{aligned}$$

现在 $\mathbf{P}^i\mathbf{S}_j = \mathbf{P}^i+\cdots+\mathbf{P}^{i+j-1}$，再加上 $\mathbf{S}_i=\mathbf{I}+\cdots+\mathbf{P}^{i-1}$，刚好就是 $\mathbf{S}_{i+j}$。所以，对任意 $\mathbf{P}$，

$$\mathbf{X}_i * \mathbf{X}_j = \mathbf{X}_{i+j}.$$

（这并不令人意外：先按递推关系走 j 步，再多走 i 步，就是走了 $i+j$ 步。）

现在取 $\mathbf{P}=\mathbf{M}$。由第一部分，$\mathbf{M}^m=\mathbf{I}$ 且 $\mathbf{S}_m=\mathbf{O}$，所以

$$\mathbf{X}_m = \mathbf{X}_0,$$

也就是说，$\mathbf{X}_m$ 是令每个矩阵保持不变的表达式。于是对任意 k，$\mathbf{X}_{k+m}=\mathbf{X}_k*\mathbf{X}_m=\mathbf{X}_k$，所以这列表达式以周期 m 重复出现，而

$$G=\{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\ldots\}=\{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_m\}$$

是一个有限集。我们逐一检验群公理。

- 封闭性（*closure*）。$\mathbf{X}_i * \mathbf{X}_j = \mathbf{X}_{i+j}$，而 $i+j \geq 1$，所以它属于 G。
- 结合律（*associativity*）。$(\mathbf{X}_i * \mathbf{X}_j) * \mathbf{X}_k$ 与 $\mathbf{X}_i * (\mathbf{X}_j * \mathbf{X}_k)$ 都等于 $\mathbf{X}_{i+j+k}$。（更一般地说，代入就是函数的复合，而函数的复合必定满足结合律。）
- 单位元（*identity*）。$\mathbf{X}_m * \mathbf{X}_j = \mathbf{X}_{m+j} = \mathbf{X}_j$ 且 $\mathbf{X}_j * \mathbf{X}_m = \mathbf{X}_j$，所以 $\mathbf{X}_m$ 是单位元。
- 逆元（*inverses*）。对 $1 \leq i < m$，$\mathbf{X}_i * \mathbf{X}_{m-i} = \mathbf{X}_{m-i} * \mathbf{X}_i = \mathbf{X}_m$，而 $\mathbf{X}_m$ 的逆元就是它自己。每个逆元都属于 G。

所以 G 在运算 $*$ 下是一个有限群。

事实上，它刚好有 m 个元素：$\mathbf{X}_1, \ldots, \mathbf{X}_m$ 是互不相同的表达式，因为它们的系数 $\mathbf{M}^1, \ldots, \mathbf{M}^m$ 是转角各不相同的旋转，转角都在 $(0, 2\pi]$ 之内。由于每个元素都是 $\mathbf{X}_1$ 的“幂”，G 就是阶（order）为 m 的循环群（cyclic group）。

验算

- 当 $m = 2$ 时，$\mathbf{M} = -\mathbf{I}$，和为 $\mathbf{I} + (-\mathbf{I}) = \mathbf{O}$。当 $m = 4$ 时，$\mathbf{M}$ 是四分之一圈的旋转，而 $\mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \mathbf{M}^3 = \mathbf{I} + \mathbf{M} - \mathbf{I} - \mathbf{M} = \mathbf{O}$。
- 以数值计算，对由 2 至 39 的每个 m，各次幂之和都是 $\mathbf{O}$（在舍入误差之内）；对随机选取的 $\mathbf{P}$、$\mathbf{Q}$、$\mathbf{X}_0$，$\mathbf{X}_k$ 的公式与递推关系吻合；而当 $\mathbf{P} = \mathbf{M}$，$\mathbf{Q}$ 和 $\mathbf{X}_0$ 随机选取时，按递推关系走 m 步后刚好回到 $\mathbf{X}_0$。

常见错误

- 没有检查 $\mathbf{I} - \mathbf{M}$ 是否可逆便把它约去。
- 写成 $\mathbf{Q}\mathbf{S}_k$ 或 $\mathbf{X}_0\mathbf{P}^k$：矩阵乘法不可交换，而递推关系中 $\mathbf{P}$ 是放在左边的。
- 把 $\mathbf{X}_0$ 当作单位元：按题目所列，$\mathbf{X}_0$ 并不在 G 之内；单位元是 $\mathbf{X}_m$，它之所以等于 $\mathbf{X}_0$，全因 $\mathbf{S}_m = \mathbf{O}$。

STEP 之外

每个 $\mathbf{X}_i$ 都是一个仿射映射（*affine map*）$\mathbf{X} \mapsto \mathbf{A}\mathbf{X} + \mathbf{B}$，而 $\mathbf{A}$ 可逆的仿射映射在复合运算下构成一个群；这里我们找到的是一个阶为 m 的循环子群（cyclic subgroup），它由单独一个映射生成，而这个映射的 m 次幂是单位元。$\mathbf{I} + \mathbf{M} + \cdots + \mathbf{M}^{m-1} = \mathbf{O}$ 这个事实，正是“m 次单位根（roots of unity）之和为零”的矩阵版本；而背后的一般想法，即

一个群以变换的方式作用于一个集合，是大学第一门群论（group theory）课程的核心。Allenby 的 *Rings, Fields and Groups* 对此有详细阐述。

详细解答

STEP III 1989，第 4 题

纯数学 · 曲线描绘与根

STEP III 1989，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/89-s3-q4

关键思路

把分式拆成多项式加一个真分式（proper fraction），斜渐近线（oblique asymptote）便立刻出现，求导数（derivative）也变得容易。结果发现导数处处大于 2，所以三段分支各自由 $-\infty$ 一直上升到 $+\infty$。题目后半部的每一条方程，其实都是“曲线 = 某条较简单的曲线”的变相写法。两者之差在每段分支上的严格单调性（strict monotonicity），便能准确数出交点的数目。

解答

曲线

由于 $2x(x^2-5) = 2x(x^2-4) - 2x$，

$$y = 2x - \frac{2x}{x^2-4}.$$

这个函数是奇函数（odd function），所以曲线关于原点有 2 阶旋转对称（rotational symmetry）。曲线只在 $x(x^2-5)=0$ 之处与坐标轴相交，即 $x=0$ 及 $x=\pm\sqrt{5}$。

渐近线。分母在 $x=\pm 2$ 时为零（此时分子为 $\mp 4 \neq 0$），所以 $x=2$ 和 $x=-2$ 是铅垂渐近线（vertical asymptote）。当 $|x|\to\infty$ 时，$\frac{2x}{x^2-4}\to 0$，所以 $y=2x$ 是斜渐近线。此外，$y-2x = -\frac{2x}{x^2-4}$ 在 x 为很大的正数时是负的，在 x 为很大的负数时是正的。因此曲线在右方从下方趋近 $y=2x$，在左方则从上方趋近。

斜率。对拆开后的式子求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2 - \frac{2(x^2-4) - 2x\cdot 2x}{(x^2-4)^2} = 2 + \frac{2(x^2+4)}{(x^2-4)^2} > 2.$$

所以曲线没有驻点（stationary point），而 y 在 $x<-2$、$-2<x<2$ 和 $x>2$ 这三个区间上各自严格递增。再求导一次，得 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}=-\dfrac{4x(x^2+12)}{(x^2-4)^3}$，所以唯一的拐点（point of inflection）是原点。

原点处的切线。在 $x=0$ 处斜率为 $2+\frac{2\cdot4}{16}=\frac{5}{2}$，所以该处的切线（tangent）是

$$y=\tfrac{5}{2}x.$$

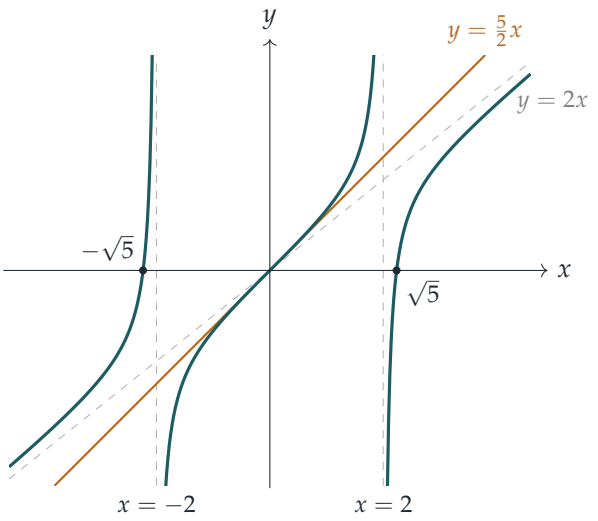
在 $x=\pm2$ 附近的性态。在 $x=2$ 附近，$-\frac{2x}{x^2-4}$ 这一项起主导作用。当 $x\to2^+$ 时它趋向 $-\infty$，当 $x\to2^-$ 时趋向 $+\infty$；由对称性，当 $x\to-2^-$ 时它趋向 $+\infty$，当 $x\to-2^+$ 时趋向 $-\infty$。因此：

- 当 $x<-2$ 时，y 由 $-\infty$ 上升至 $+\infty$，并在 $-\sqrt{5}$ 处穿过横轴；
- 当 $-2<x<2$ 时，y 由 $-\infty$ 上升至 $+\infty$，并经过原点；
- 当 $x>2$ 时，y 由 $-\infty$ 上升至 $+\infty$，并在 $\sqrt{5}$ 处穿过横轴。

还有一个事实稍后会用得着。当 $0<x<2$ 时，

$$y-\tfrac{5}{2}x=\frac{x\left[4(x^2-5)-5(x^2-4)\right]}{2(x^2-4)}=\frac{-x^3}{2(x^2-4)}>0,$$

所以在这一段上，曲线位于原点处切线的上方（由对称性，当 $-2<x<0$ 时则位于切线下方）。



第 (i) 部分

$x = 0$（左边为 0，右边为 8）和 $x = \pm 2$（左边为 -16，右边为 0）都不是根。所以我们可以除以 $2x(x^2-4)$，方程变成

$$\frac{2x(x^2-5)}{x^2-4} = \frac{5x-2}{2x} = \frac{5}{2} - \frac{1}{x}.$$

因此，方程的根就是曲线与双曲线（hyperbola）$H\colon y = \frac{5}{2} - \frac{1}{x}$ 各交点的 x-坐标。我们逐段分支来看。

- $x < -2$：这里 H 的斜率为 $\frac{1}{x^2} < \frac{1}{4}$，而曲线的斜率 > 2。所以两者之差（曲线 $- H$）严格递增。当 $x \to -\infty$ 时它趋向 $-\infty$，当 $x \to -2^-$ 时趋向 $+\infty$，因为 H 一直介乎 $\frac{5}{2}$ 与 3 之间。恰有一个根。
- $-2 < x < 0$：曲线上 $y < 0$（三个因子 x、x^2-5、x^2-4 全是负的），但 H 上 $y > \frac{5}{2}$。没有根。
- $0 < x < 2$：由上面的事实，$y > \frac{5}{2}x$。另外

  $$\tfrac{5}{2}x - \left(\tfrac{5}{2} - \tfrac{1}{x}\right) = \frac{5x^2-5x+2}{2x} > 0,$$

  因为 $5x^2-5x+2$ 的判别式（discriminant）为 $25-40 < 0$。所以曲线严格位于 H 的上方。没有根。
- $x > 2$：与第一种情况一样，两者之差严格递增（斜率 $> 2 - \frac{1}{4}$）。它由 $-\infty$ 走到 $+\infty$，因为 H 一直介乎 2 与 $\frac{5}{2}$ 之间。恰有一个根。

所以方程 (i) 恰有**两**个实根，一个小于 -2，一个大于 2。

> **想一想** (ii) 的方程两边都有平方，怎样才能从曲线上把它读出来?

除以 $(x^2-4)^2$。左边变成 y^2，所以方程说的是 $y^2 = x^2 + 1$。

第 (ii) 部分

同样，$x = \pm 2$ 不是根（左边为 16，右边为 0）。所以方程就是 $y^2 = x^2 + 1$，其中 y 是曲线的纵坐标（ordinate）。方程的根就是曲线与等轴双曲线（rectangular hyperbola）$y^2 - x^2 = 1$ 的两支 $y = \pm\sqrt{x^2+1}$ 的交点。

上支。设 $f(x) = y - \sqrt{x^2+1}$。则

$$f'(x) = \frac{\mathrm{d}y}{\mathrm{d}x} - \frac{x}{\sqrt{x^2+1}} > 2 - 1 > 0,$$

所以 f 在三个区间上各自严格递增。在每个区间上，y 由 $-\infty$ 走到 $+\infty$，而 $\sqrt{x^2+1}$ 保持有限（它的增长只与 $|x|$ 相若，而当 $x \to +\infty$ 时对手是 $2x$；当 $x \to -\infty$ 时，f 的两项都是负的）。所以 f 由 $-\infty$ 走到 $+\infty$，每个区间上恰有一个根：合共三个。

下支。映射 $(x,y) \mapsto (-x,-y)$ 把曲线映到自身，并把下支映到上支。所以曲线与下支也恰有三个交点。

这两组根不会重叠：对同一个 x，曲线只有一个 y，它不可能既正又负。所以方程 (ii) 恰有**六**个实根。（作为验算，它是一条 6 次多项式方程，即 $3x^6 - 33x^4 + 92x^2 - 16 = 0$，所以它的根全是实数。）

? **想一想** (iii) 的变量是 z 而不是 x，而且 z 可能是负数。怎样才能把曲线用上?

当 $z > 0$ 时，我们令 $z = x^2$，其中 $x > 0$。至于 $z \le 0$ 的情况，则另外直接处理。

第 (iii) 部分

左边永不为负，所以任何根都必须满足 $z + 1 \ge 0$，即 $z \ge -1$。另外 $z = 0$ 不是根（左边为 0，右边为 16），$z = 4$ 也不是（左边为 64，右边为 0）。

满足 $z > 0$ 的根。令 $z = x^2$，其中 $x > 0$，使每个正根 z 恰好对应一个 $x > 0$。方程变成 $4x^4(x^2-5)^2 = (x^2-4)^2(x^2+1)$。除以 $x^2(x^2-4)^2$ 得

$$y^2 = 1 + \frac{1}{x^2}, \qquad \text{即} \qquad y = \pm s(x), \quad s(x) = \sqrt{1+x^{-2}}.$$

这里 s 由 $+\infty$（当 $x \to 0^+$）递减至 1（当 $x \to \infty$），而 $s'(x) = -\dfrac{1}{x^3\sqrt{1+x^{-2}}}$，所以 $|s'(x)| < \frac{1}{x^3}$。

- $y = s(x)$：函数 $y - s$ 的导数为 $y' - s' > 2$。在 $(0,2)$ 上，它由 $-\infty$（因为 $y \to 0$ 而 $s \to \infty$）走到 $+\infty$。在 $(2,\infty)$ 上，它同样由 $-\infty$ 走到 $+\infty$。每个区间上各有一个根，所以共有两个根。
- 在 $(0,2)$ 上的 $y = -s(x)$：曲线在那里满足 $y > 0 > -s$。没有根。
- 在 $(2,\infty)$ 上的 $y = -s(x)$：函数 $y + s$ 的导数为 $y' + s' > 2 - \frac{1}{8} > 0$，并由 $-\infty$ 走到 $+\infty$。有一个根。

所以恰有三个正根 z。

满足 $-1 \le z < 0$ 的根。写 $z = -t$，其中 $0 < t \le 1$。方程变成

$$4t^2(t+5)^2 = (t+4)^2(1-t).$$

在 $(0,1]$ 上，左边严格递增，由 0（当 $t \to 0$）增至 144。右边严格递减，由 16 减至 0。两者之差严格递增并且变号，所以恰有一个根。

因此方程 (iii) 恰有**四**个实根：三个正根，另一个在 $(-1,0)$ 内。展开后，它是四次方程（quartic）$4z^4-41z^3+107z^2-8z-16=0$，所以它的四个根全是实数。

验算

- 在 $x=1000$ 处，$y-2x\approx -0.002$，印证了曲线是从下方趋近 $y=2x$。
- 用数值计算，(i) 的根为 $x\approx -2.143$ 和 2.401，一如所料，外侧两段分支上各有一个。
- (ii) 的根为 $x\approx \pm 0.431,\ \pm 2.155,\ \pm 2.484$，对双曲线的每一支而言，曲线的每段分支上各有一个。
- (iii) 的根为 $z\approx -0.332,\ 0.473,\ 4.799,\ 5.310$。后三个给出 $x=\sqrt{z}\approx 0.688,\ 2.191,\ 2.304$，曲线上相应的值为 $y\approx 1.77,\ -1.10,\ 1.09$，与分情况讨论的结果吻合。

常见错误

- 没有先检查 $x=0$ 和 $x=\pm 2$ 不是根，就除以 x 或 x^2-4。
- 在 (ii) 中，只数与 $y=+\sqrt{x^2+1}$ 的交点而忘记了下支，结果得出 3 而不是 6。
- 在 (iii) 中，代入 $z=x^2$ 而忘记了 z 可以是负数，因而漏掉 $(-1,0)$ 内的根。

STEP 之外

这里的计数论证，建基于介值定理（intermediate value theorem，即变号的连续函数必有零点），再加上严格单调性（零点最多只有一个）。把这两个事实严格地建立起来，是大学分析课程最先要做的工作之一。Abbott 的 *Understanding Analysis* 对两者都有细致而易读的论述。笛卡儿（Descartes）符号法则（rule of signs）和施图姆（Sturm）序列是数算多项式实根数目的代数方法，与这里所用的几何方法对照起来，相当有趣。

详细解答

STEP III 1989，第 5 题

纯数学 · 切比雪夫多项式

> **STEP III 1989，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/89-s3-q5

关键思路

设 $x = \cosh\theta$。这样 $e^{\pm\theta}$ 正好就是 $x \pm \sqrt{x^2-1}$，而 $y = \cosh n\theta$ 是它们的 n 次幂的平均值。用二项式定理（binomial theorem）展开这些幂，$\sqrt{x^2-1}$ 的奇次幂会互相抵消，余下的就是 x 的多项式（polynomial）。只要把这个多项式写成整齐的形式，每个系数（coefficient）不过是几个二项式系数（binomial coefficient）之和。

解答

闭合形式

设 $\theta = \cosh^{-1} x$，则 $\theta \geq 0$ 且 $x = \cosh\theta$。于是 $\sinh\theta \geq 0$ 而 $\sinh^2\theta = x^2 - 1$，所以 $\sinh\theta = \sqrt{x^2-1}$。因此

$$e^{\theta} = \cosh\theta + \sinh\theta = x + \sqrt{x^2-1}, \qquad e^{-\theta} = \cosh\theta - \sinh\theta = x - \sqrt{x^2-1}.$$

故此

$$y = \cosh n\theta = \frac{e^{n\theta} + e^{-n\theta}}{2} = \frac{(x+\sqrt{x^2-1})^n + (x-\sqrt{x^2-1})^n}{2},$$

这正是所要证明的结果。

为何 y 是奇多项式

记 $s = \sqrt{x^2-1}$。由二项式定理，

$$(x+s)^n + (x-s)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} s^j \left(1 + (-1)^j\right) = 2 \sum_{j \text{ 为偶数}} \binom{n}{j} x^{n-j} s^j.$$

s 的奇次幂互相抵消，而每个偶次幂 $s^{2m} = (x^2-1)^m$ 都是 x 的多项式。所以

$$y = \sum_{m=0}^{\lfloor n/2 \rfloor} \binom{n}{2m} x^{n-2m}(x^2-1)^m.$$

现在取 $n = 2k+1$，则 m 由 0 取到 k。每一项都是奇次幂 $x^{2k+1-2m}$ 乘以一个 x^2 的多项式，所以只含 x 的奇次幂，其中最高的是 x^{2k+1}。因此 $y = a_0x + a_1x^3 + \cdots + a_kx^{2k+1}$。

要读出各系数，设 $j = k-m$（即 $j = 0,\ldots,k$），并利用 $\binom{2k+1}{2k-2j} = \binom{2k+1}{2j+1}$：

$$y = \sum_{j=0}^{k} \binom{2k+1}{2j+1} x^{2j+1}(x^2-1)^{k-j}.$$

指标为 j 的那一项由 x^{2j+1} 开始。所以 x^{2r+1} 只来自 $j \le r$ 的各项，而在其中第 j 项，我们需要 $(x^2-1)^{k-j}$ 中的 $(x^2)^{r-j}$ 项，即 $\binom{k-j}{r-j}(-1)^{k-r}$。因此

$$a_r = (-1)^{k-r} \sum_{j=0}^{r} \binom{2k+1}{2j+1}\binom{k-j}{r-j}. \qquad (*)$$

a_0

当 $r = 0$ 时，只有 $j = 0$ 有贡献：

$$a_0 = (-1)^k \binom{2k+1}{1} = (-1)^k(2k+1).$$

第 (i) 部分

当 $r = 1$ 时，$j = 0, 1$ 有贡献：

$$\begin{aligned} a_1 &= (-1)^{k-1}\left[(2k+1)\binom{k}{1} + \binom{2k+1}{3}\right] = (-1)^{k-1}(2k+1)\left[k + \frac{2k(2k-1)}{6}\right] \\ &= (-1)^{k-1}(2k+1)\,k\,\frac{3+2k-1}{3} = (-1)^{k-1}\,\frac{2k(k+1)(2k+1)}{3}, \end{aligned}$$

这正是所要证明的结果。

第 (ii) 部分

这里需要 $k \ge 2$，x^5 才会出现。当 $r = 2$ 时，$j = 0, 1, 2$ 有贡献：

$$a_2 = (-1)^k\left[(2k+1)\binom{k}{2} + \binom{2k+1}{3}\binom{k-1}{1} + \binom{2k+1}{5}\right].$$

每一项都有因子 $(2k+1)k(k-1)$；把它抽出，

$$\begin{aligned} a_2 &= (-1)^k(2k+1)k(k-1)\left[\frac{1}{2}+\frac{2k-1}{3}+\frac{(2k-1)(2k-3)}{30}\right] \\ &= (-1)^k(2k+1)k(k-1)\,\frac{15+(20k-10)+(4k^2-8k+3)}{30} \\ &= (-1)^k(2k+1)k(k-1)\,\frac{4(k+1)(k+2)}{30} = (-1)^k\,\frac{2(k-1)k(k+1)(k+2)(2k+1)}{15}. \end{aligned}$$

处理第三项时，我们用了 $\binom{2k+1}{5} = \frac{(2k+1)(2k)(2k-1)(2k-2)(2k-3)}{120} = \frac{(2k+1)k(k-1)(2k-1)(2k-3)}{30}$。

> **想一想** 我们的答案有一个因子 $(k+1)$，而 (ii) 中印出的公式却没有。哪一个才对？

用 $k=2$ 来检验。这时 $y=\cosh(5\theta)$，其中 $x=\cosh\theta$，而 $\cosh 5\theta = 16x^5-20x^3+5x$，所以 $a_2=16$。我们的公式给出 $2\cdot1\cdot2\cdot3\cdot4\cdot5/15=16$。印出的公式则给出 $2\cdot1\cdot2\cdot4\cdot5/15=\frac{16}{3}$，这不可能正确，因为各个 a_r 都是整数。所以试卷有印刷错误：漏了因子 $(k+1)$。正确的结果就是上面所示的那一个。

系数之和

在多项式中代入 $x=1$。由于恒等式对 $x\geq 1$ 成立，而 $x=1$ 在容许范围之内，

$$\sum_{r=0}^{k} a_r = y(1) = \cosh\left(n\cosh^{-1}1\right) = \cosh 0 = 1.$$

验算

- $k=1$: $\cosh 3\theta = 4x^3-3x$。公式给出 $a_0=-3$ 和 $a_1=2\cdot1\cdot2\cdot3/3=4$，而 $-3+4=1$。
- $k=3$：$\cosh 7\theta = 64x^7-112x^5+56x^3-7x$。公式给出 $a_0=-7$、$a_1=56$ 和 $a_2=-2\cdot2\cdot3\cdot4\cdot5\cdot7/15=-112$，而各系数之和为 1。
- 我们用电脑把 $(*)$ 与递推关系（recurrence）$T_{m+1}=2xT_m-T_{m-1}$ 对照，检验了所有 $k\leq 14$ 的情况，并在随机选取的 $x\geq 1$ 处把多项式与 $\cosh(n\cosh^{-1}x)$ 作比较。

常见错误

- 忘记了 x^{2r+1} 的系数来自展开式中的好几项，而不只是一项。
- 正负号出错：a_r 的每一项贡献都带有相同的符号 $(-1)^{k-r}$，所以应先把它抽出。
- 硬要把代数运算凑成 (ii) 中印出的形式。只要检验一个小的情况，例如 $k=2$，便知道那是印错了的。

STEP 之外

多项式 y 就是切比雪夫多项式（Chebyshev polynomial）$T_n(x)$，通常以 $T_n(\cos\phi) = \cos n\phi$ 来定义。我们的双曲函数（hyperbolic function）版本是同一个恒等式在 $x \geq 1$ 时的情形，而两者一致，是因为多项式由它在任何一个区间上的值完全确定。切比雪夫多项式在逼近论（approximation theory）中占有核心地位，因为在所有 n 次首一多项式（monic polynomial）之中，$2^{1-n}T_n$ 在 $[-1, 1]$ 上的最大绝对值是最小的。在任何数值分析（numerical analysis）的入门课程中，它们都会再次出现。如果想多做一些这类代数运算的 STEP 式练习，可以试试 Siklos 的 *Advanced Problems in Mathematics*（STEP 进阶数学题集）。

详细解答

STEP III 1989，第 6 题

纯数学 · 圆锥曲线的极坐标方程

> **STEP III 1989，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/89-s3-q6

关键思路

在极坐标（polar coordinates）中，一条不经过原点的直线的 $1/r$ 等于 $\cos\theta$ 与 $\sin\theta$ 的组合，而把它写成 $A\cos\theta + B\cos(\theta-\gamma)$，就可以与圆锥曲线（conic）本身的 $\cos\theta$ 项完全配对。只要 $e\cos\theta$ 的部分一致，弦（chord）的方程便只需处理单独一个余弦项，其余的就交给关于 $\theta=\gamma$ 的对称性。让弦收缩成一点便得到切线（tangent），而有了切线方程，平分线（bisector）性质几乎是自动得出的。

解答

全文之中，(r,θ) 为极坐标，其中 $x = r\cos\theta$，$y = r\sin\theta$，并且容许 $r<0$（点 (r,θ) 与 $(-r,\theta+\pi)$ 是同一点）。圆锥曲线的焦点（focus）S 位于原点。

直线的极坐标形式

由于 $lx+my=1$ 不可能经过原点，直线上 $r\neq 0$，两边除以 r 得

$$\frac{1}{r} = l\cos\theta + m\sin\theta.$$

把题目建议的形式展开，

$$A\cos\theta + B\cos(\theta-\gamma) = (A + B\cos\gamma)\cos\theta + B\sin\gamma\,\sin\theta.$$

若 $B\sin\gamma = m$ 及 $A + B\cos\gamma = l$，两式便对所有 θ 都相等。由于 $\sin\gamma\neq 0$，可取

$$B = \frac{m}{\sin\gamma}, \qquad A = l - m\cot\gamma,$$

这就是所求的极坐标方程。反过来说，任何方程 $1/r = A\cos\theta + B\cos(\theta-\gamma)$，只要 A、B 不会令两个系数同时为零，都是一条直线 $lx+my=1$，其中 $l = A + B\cos\gamma$，$m = B\sin\gamma$。

留意这个形式在 $(r,\theta)\to(-r,\theta+\pi)$ 之下保持不变，所以无论 r 取哪个符号，它描述的都是整条直线。

弦 PQ

设弦的方程具有 $a/r = A\cos\theta + B\cos(\theta-\gamma)$ 的形式。在 P 和 Q 处有 $\theta = \gamma\mp\delta$ 及 $a/r = 1+e\cos\theta$，所以我们需要

$$A\cos(\gamma\mp\delta) + B\cos(\mp\delta) = 1 + e\cos(\gamma\mp\delta).$$

因为对两点而言都有 $\cos(\mp\delta)=\cos\delta$，取 $A=e$，$B\cos\delta = 1$ 便可同时满足两条方程，而 $\cos\delta\neq 0$ 保证可以解出 B。两个相异的点确定一条直线，所以弦 PQ 为

$$\boxed{\frac{a}{r} = e\cos\theta + \sec\delta\,\cos(\theta-\gamma).}$$

（若 $\cos\delta = 0$，则 P 和 Q 位于 $\theta=\gamma\mp\frac{\pi}{2}$，而 PQ 是一条焦弦（focal chord）；经过原点的直线没有 $1/r=\ldots$ 这种形式的方程，这就是该情况被排除的原因。）

> **想一想** 我们怎样知道这条弦只在 P 和 Q 与圆锥曲线相交?

对每个 θ，圆锥曲线上恰有一点，满足 $a/r = 1+e\cos\theta$。弦经过该点，当且仅当 $1+e\cos\theta = e\cos\theta+\sec\delta\cos(\theta-\gamma)$，即 $\cos(\theta-\gamma)=\cos\delta$，所以 $\theta=\gamma\pm\delta$。$e\cos\theta$ 项互相抵消，这正是选取 $A=e$ 的用意所在。

在 $\theta=\gamma$ 处的切线

设 $\theta=\gamma$ 所对应的点为 G，该点的切线就是当 P 和 Q 向 G 靠拢时，亦即 $\delta\to 0$ 时，弦 PQ 的极限。弦的方程连续地依赖于 δ，而 $\sec\delta\to 1$，所以切线为

$$\boxed{\frac{a}{r} = e\cos\theta + \cos(\theta-\gamma).}$$

作为印证，这条直线与圆锥曲线相交于 $\cos(\theta-\gamma)=1$ 之处，即只在 G 相交，切线理应如此。更具决定性的验证是：切线的直角坐标（cartesian）形式为 $(e+\cos\gamma)x+(\sin\gamma)y=a$，其法向量（normal vector）为 $(e+\cos\gamma,\ \sin\gamma)$。圆锥曲线为 $r=a-ex$；沿曲线求微分得 $(x\,\mathrm{d}x+y\,\mathrm{d}y)/r = -e\,\mathrm{d}x$，即 $(e+x/r)\,\mathrm{d}x+(y/r)\,\mathrm{d}y=0$。在 G 处，$x/r=\cos\gamma$ 及 $y/r=\sin\gamma$，所以曲线在该处的法向量正是同一个向量。

最后一部分

把焦点 S 置于原点，圆锥曲线便是 $a/r=1+e\cos\theta$，并设 L 和 M 分别为 $\theta=\alpha$ 和 $\theta=\beta$ 所对应的点。记

$$\gamma = \tfrac{1}{2}(\alpha+\beta),\qquad \delta = \tfrac{1}{2}(\beta-\alpha),$$

于是 L 和 M 位于 $\theta=\gamma-\delta$ 和 $\gamma+\delta$，与上文的 P 和 Q 完全一样。根据上一个结果，在 L 和 M 处的切线分别为

$$\frac{a}{r}=e\cos\theta+\cos(\theta-\alpha),\qquad \frac{a}{r}=e\cos\theta+\cos(\theta-\beta).$$

在 T 处两式同时成立，所以 $\cos(\theta-\alpha)=\cos(\theta-\beta)$。因此，要么 $\theta-\alpha=\theta-\beta+2k\pi$，这会令 $L=M$；要么

$$\theta-\alpha=-(\theta-\beta)+2k\pi \quad\Longrightarrow\quad \theta=\tfrac{1}{2}(\alpha+\beta)+k\pi=\gamma+k\pi.$$

所以 T 位于经过 S、方向为 $\theta=\gamma$ 的直线上，且

$$r_T=\frac{a}{e\cos\gamma+\cos\delta}$$

（两条切线平行、T 不存在的情况，恰好是 $e\cos\gamma+\cos\delta=0$ 之时）。

射线 SL 和 SM 分别指向 $\alpha=\gamma-\delta$ 和 $\beta=\gamma+\delta$ 的方向（这里取通常的情况，即在 L 和 M 处 $r>0$，椭圆（ellipse）或抛物线（parabola）便是如此）。直线 $\theta=\gamma$ 与两条射线所成的角都是 δ，所以直线 ST 平分角 LSM。（若 $r_T<0$，则 T 位于反方向的射线上，ST 平分的是优角（reflex angle）LSM；角 LST 与角 MST 仍然相等。）

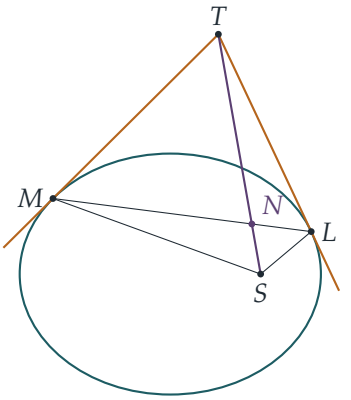
要找出 ST 与 LM 的交点，把该点称为 N。弦 LM 为 $a/r=e\cos\theta+\sec\delta\cos(\theta-\gamma)$，而在 ST 上有 $\theta=\gamma$，所以

$$\frac{a}{SN}=e\cos\gamma+\sec\delta \quad\Longrightarrow\quad SN=\frac{a\cos\delta}{1+e\cos\gamma\cos\delta}.$$

利用 $2\cos\gamma\cos\delta=\cos(\gamma-\delta)+\cos(\gamma+\delta)=\cos\alpha+\cos\beta$，上式即

$$\boxed{N\text{ 位于}\theta=\tfrac{1}{2}(\alpha+\beta),\quad r=\frac{2a\cos\frac{1}{2}(\beta-\alpha)}{2+e\cos\alpha+e\cos\beta}.}$$

当 LM 是焦弦（$\cos\delta=0$）时，这给出 $N=S$，正如所料。



验算

- 试试圆的情况（$e = 0$，半径为 a）：切线变成 $a/r = \cos(\theta - \gamma)$，即 $x\cos\gamma + y\sin\gamma = a$，正是 $x^2 + y^2 = a^2$ 为人熟悉的切线。
- 三角形 LSM 中的角平分线定理（angle bisector theorem）指出 $LN : NM = SL : SM$。取 $a = 1.5$，$e = 0.6$，$\alpha = 40°$，$\beta = 160°$（即图中的情况），得 $SN \approx 0.791$，而该比例经数值验算后成立。
- 我们用电脑在数以千计随机选取的圆锥曲线和点上，测试了弦与切线的方程、ST 的方向、两角相等的性质，以及 SN 的公式。

常见错误

- 解 $\cos(\theta - \alpha) = \cos(\theta - \beta)$ 时只取 $+$ 号。这个分支得不出任何结果（它需要 $\alpha = \beta$），有用的解来自 $-$ 号。
- 没有善用 A 的选择自由：取 $A = e$，才能令 $e\cos\theta$ 项在与圆锥曲线求交点时互相抵消。

STEP 之外

ST 平分 $\angle LSM$ 这个结果，是圆锥曲线一个经典的焦点性质，而直线 LM 则是 T 关于该圆锥曲线的极线（*polar*）。极点与极线的对偶（pole–polar duality）是通往射影几何（projective geometry）的大门；在射影几何中，所有非退化的圆锥曲线都是等价的，相切与关联都变成纯粹的代数关系。Brannan、Esplen 与 Gray 合著的 *Geometry*（《几何学》）是一本很好的大学一年级程度读物，讲述圆锥曲线与射影几何。

详细解答

STEP III 1989，第 7 题

纯数学 · 错切矩阵

> **STEP III 1989，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/89-s3-q7

关键思路

用两个单位向量（unit vector）来处理，一切都最简单：$\mathbf{u}$ 沿直线 $L\colon y = x\tan\alpha$ 的方向，而 $\mathbf{n}$ 与它垂直。这个错切（shear）把每一点沿平行于 $\mathbf{u}$ 的方向移动，移动量与该点到 L 的有向距离（signed distance）成正比，所以 $\mathbf{M} = \mathbf{I} + k\,\mathbf{u}\mathbf{n}^{\mathsf{T}}$。至于第二部分，一个矩阵与 $\mathbf{I} + k\mathbf{A}$ 可交换（commute），当且仅当它与 $\mathbf{A}$ 可交换，这样便可撇除一切枝节。

解答

记 $s = \sin\alpha$ 及 $c = \cos\alpha$。由于 α 是锐角，故 $s > 0$ 且 $c > 0$。设

$$\mathbf{u} = \begin{pmatrix} c \\ s \end{pmatrix}, \qquad \mathbf{n} = \begin{pmatrix} -s \\ c \end{pmatrix},$$

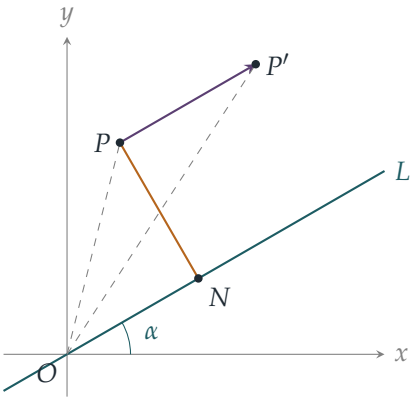
则 $\mathbf{u}$ 是沿 L 方向、与 x 轴成角 α 的单位向量，而 $\mathbf{n}$ 是指向 L 左上方一侧的单位法向量（unit normal）。

求 M

对于 $P = (x, y)$，设

$$d = \mathbf{n} \cdot \overrightarrow{OP} = y\cos\alpha - x\sin\alpha,$$

即由 P 到 L 的有向距离。于是 $PN = |d|$，而 $d > 0$ 当且仅当 P 位于 L 的上方。



取 P 在 L 的上方，故 $d > 0$。由 (i)，$\overrightarrow{PP'}$ 沿 $\mathbf{u}$ 的方向，所以 $\overrightarrow{PP'} = \lambda\mathbf{u}$，其中 $\lambda > 0$。条件 (ii) 与此相符：二维叉积（cross product）为

$$\overrightarrow{OP} \times \overrightarrow{OP'} = \overrightarrow{OP} \times (\overrightarrow{OP} + \lambda\mathbf{u}) = \lambda\,(xs - yc) = -\lambda d < 0,$$

所以由 OP 转到 OP' 是顺时针方向，符合要求。条件 (iii) 给出 $\lambda = k\,PN = kd$。（要这些条件同时成立，必须有 $k > 0$。）因此，对于 L 上方的每一点 P，

$$\overrightarrow{OP'} = \overrightarrow{OP} + kd\,\mathbf{u} = \overrightarrow{OP} + k\,\mathbf{u}\,(\mathbf{n} \cdot \overrightarrow{OP}).$$

? **想一想** 我们只用了 L 上方的点。为甚么这样已足以完全确定 $\mathbf{M}$?

T 是线性（linear）的，而映射 $\mathbf{v} \mapsto \mathbf{v} + k\,\mathbf{u}(\mathbf{n} \cdot \mathbf{v})$ 也是线性的。两者在 L 上方的开半平面（open half-plane）上一致，而这个半平面包含两个线性独立（linearly independent）的向量（例如 $\mathbf{n}$ 和 $\mathbf{n} + \mathbf{u}$）。两个线性映射若在一组基（basis）上一致，便完全相等。所以对于所有向量，

$$\mathbf{M} = \mathbf{I} + k\,\mathbf{u}\mathbf{n}^{\mathsf{T}} = \mathbf{I} + k\begin{pmatrix} c \\ s \end{pmatrix}\begin{pmatrix} -s & c \end{pmatrix} = \mathbf{I} + k\begin{pmatrix} -sc & c^2 \\ -s^2 & sc \end{pmatrix}.$$

亦即

$$\mathbf{M} = \begin{pmatrix} 1 - k\sin\alpha\cos\alpha & k\cos^2\alpha \\ -k\sin^2\alpha & 1 + k\sin\alpha\cos\alpha \end{pmatrix},$$

正如所求。作为合理性检查，由 $\mathbf{n} \cdot \mathbf{u} = 0$ 得 $\mathbf{M}\mathbf{u} = \mathbf{u}$，所以 L 上每一点都固定不动。此外 $\det\mathbf{M} = 1 - k^2s^2c^2 + k^2s^2c^2 = 1$，正是错切应有的行列式（determinant）值。

可交换的条件

设 $\mathbf{A} = \mathbf{u}\mathbf{n}^{\mathsf{T}} = \begin{pmatrix} -sc & c^2 \\ -s^2 & sc \end{pmatrix}$ 及 $\mathbf{X} = \begin{pmatrix} p & q \\ r & t \end{pmatrix}$。由于 $\mathbf{M} = \mathbf{I} + k\mathbf{A}$，而 $\mathbf{I}$ 与任何矩阵都可交换，

$$\mathbf{X}\mathbf{M} - \mathbf{M}\mathbf{X} = k(\mathbf{X}\mathbf{A} - \mathbf{A}\mathbf{X}).$$

因为 $k \neq 0$，$\mathbf{X}$ 与 $\mathbf{M}$ 可交换，当且仅当它与 $\mathbf{A}$ 可交换。乘开得

$$\mathbf{X}\mathbf{A} = \begin{pmatrix} -psc - qs^2 & pc^2 + qsc \\ -rsc - ts^2 & rc^2 + tsc \end{pmatrix}, \qquad \mathbf{A}\mathbf{X} = \begin{pmatrix} -psc + rc^2 & -qsc + tc^2 \\ -ps^2 + rsc & -qs^2 + tsc \end{pmatrix},$$

所以

$$\mathbf{X}\mathbf{A} - \mathbf{A}\mathbf{X} = \begin{pmatrix} -(qs^2 + rc^2) & -c^2(t-p) + 2qsc \\ -s^2(t-p) - 2rsc & qs^2 + rc^2 \end{pmatrix}.$$

这个矩阵为零，当且仅当

$$\begin{aligned} &\text{(a)} \quad qs^2 + rc^2 = 0, \\ &\text{(b)} \quad c^2(t-p) = 2qsc, \\ &\text{(c)} \quad s^2(t-p) = -2rsc. \end{aligned}$$

由于 $c \neq 0$，(b) 即 $t - p = 2q\tan\alpha$。由于 $s \neq 0$，(c) 即 $t - p = -2r\cot\alpha$。

> **想一想** 这里有三条方程，但题目所述的条件只包含两条。(a) 到哪里去了？

它可由另外两条推出。若 $2q\tan\alpha = -2r\cot\alpha$，两边乘以 $\frac{1}{2}sc$ 便得 $qs^2 = -rc^2$，这正是 (a)。所以

$$\mathbf{X}\mathbf{M} = \mathbf{M}\mathbf{X} \iff t - p = 2q\tan\alpha = -2r\cot\alpha.$$

两个方向都成立：以上每一步都是等价关系。这就是所求的充分必要条件（necessary and sufficient condition）。

验算

- 取 $\alpha = \frac{\pi}{4}$、$k = 2$，得 $\mathbf{M} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$。点 $(0,1)$ 位于 $y = x$ 上方，距离为 $\frac{1}{\sqrt{2}}$，它被移到 $(1,2)$。这是沿 45° 方向移动了 $\sqrt{2} = 2 \times \frac{1}{\sqrt{2}}$，而且转向是顺时针的。
- $\mathbf{X} = \mathbf{M}$ 本身满足这个条件：$t - p = 2ksc$，$2q\tan\alpha = 2kc^2 \cdot \frac{s}{c} = 2ksc$，以及 $-2r\cot\alpha = 2ks^2 \cdot \frac{c}{s} = 2ksc$。$\mathbf{I}$ 也满足这个条件，三个式子都等于零。
- 以数值方法检验：随机选取 α、k 和 L 上方的 P，$\mathbf{M}$ 都重现了 (i)–(iii)。满足条件的随机矩阵与 $\mathbf{M}$ 可交换，误差在 10^{-8} 之内；不满足条件的随机矩阵则不可交换。

常见错误

- 对每一点都使用不带正负号的距离 PN。位移是有向距离的 k 倍。L 下方的点向相反方向移动，正因如此 T 才是线性的。
- 就可交换的条件只证明了“若”（或只证明了“仅若”）的方向，或忘记说明 $(1,1)$ 元素并没有带来新的条件。

STEP 之外

这里与 $\mathbf{M}$ 可交换的矩阵，恰好就是形如 $a\mathbf{I}+b\mathbf{A}$ 的矩阵：请验证 $t-p=2q\tan\alpha=-2r\cot\alpha$ 是一个双参数族。这是一个一般事实的特例。对于一个不是 $\mathbf{I}$ 的倍数的 2×2 矩阵，与它可交换的矩阵恰好就是 $a\mathbf{I}+b\mathbf{M}$。对于 $n\times n$ 矩阵，当最小多项式（minimal polynomial）等于特征多项式（characteristic polynomial）时，相应的结论（可交换的矩阵都是 $\mathbf{M}$ 的多项式）成立。Strang 的 *Introduction to Linear Algebra*（线性代数入门教材）阐述了秩一矩阵（rank-one matrix）$\mathbf{u}\mathbf{n}^{\mathsf{T}}$ 以及这背后的思想。

详细解答

STEP III 1989，第 8 题

纯数学 · 耦合微分方程

STEP III 1989，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/89-s3-q8

关键思路

第一条方程把 y 用 x 和 $\mathrm{d}x/\mathrm{d}t$ 表示出来。因此只要把它再微分一次，出现的 $\mathrm{d}y/\mathrm{d}t$ 就可以用第二条方程代换，y 便完全消失。剩下的是一条二阶常系数线性方程（second-order linear equation with constant coefficients）。它的辅助方程（auxiliary equation）有重根（repeated root），而其中一个强迫项（forcing term）正好与这个根吻合。这种共振（resonance）才是本题真正的考验。

解答

消去 y

以字母上方的点表示对 t 的导数。把 $\dot{x} = 4(x - y)$ 微分，得

$$\ddot{x} = 4\dot{x} - 4\dot{y} = 4\dot{x} - 4x + 48\left(\mathrm{e}^{2t} + \mathrm{e}^{-2t}\right),$$

其中 $\dot{y}$ 用了第二条方程代入。因此

$$\boxed{\ddot{x} - 4\dot{x} + 4x = 48\left(\mathrm{e}^{2t} + \mathrm{e}^{-2t}\right),}$$

式中不含 y。

求解 x

余函数（*complementary function*）。辅助方程 $m^2 - 4m + 4 = (m-2)^2 = 0$ 有重根 $m = 2$，所以余函数是

$$x_{\mathrm{c}} = (A + Bt)\,\mathrm{e}^{2t}.$$

特别积分（*particular integral*）。由于方程是线性的，我们可以把两个强迫项逐一处理。

对于 $48\mathrm{e}^{-2t}$，试用 $x = D\mathrm{e}^{-2t}$。代入后得 $D(4+8+4)\mathrm{e}^{-2t} = 48\mathrm{e}^{-2t}$，所以 $D = 3$。

> ? **想一想** 对于 $48\mathrm{e}^{2t}$ 这一项，为甚么 $C\mathrm{e}^{2t}$ 行不通，甚至连 $Ct\,\mathrm{e}^{2t}$ 也行不通？

e^{2t} 和 $t\,\mathrm{e}^{2t}$ 都已经在余函数之中，所以方程左边会把它们变成 0。因为 2 是二重根，我们必须升到 $t^2\mathrm{e}^{2t}$。试用 $x = Ct^2\mathrm{e}^{2t}$。一个简洁的代入方法是写成 $x = u\,\mathrm{e}^{2t}$。这样

$$\dot{x} = (\dot{u} + 2u)\mathrm{e}^{2t}, \qquad \ddot{x} = (\ddot{u} + 4\dot{u} + 4u)\mathrm{e}^{2t},$$

所以

$$\ddot{x} - 4\dot{x} + 4x = \ddot{u}\,\mathrm{e}^{2t}.$$

取 $u = Ct^2$，上式便是 $2C\mathrm{e}^{2t}$，所以 $2C = 48$，即 $C = 24$。（这个代换同时说明了为甚么余函数是 $(A + Bt)\mathrm{e}^{2t}$：我们需要 $\ddot{u} = 0$。）

因此通解（general solution）是

$$x = (A + Bt)\,\mathrm{e}^{2t} + 24t^2\mathrm{e}^{2t} + 3\mathrm{e}^{-2t}.$$

初始条件（*initial conditions*）。题目只给出 $x(0) = 0$ 和 $y(0) = 0$。二阶方程还需要 $\dot{x}(0)$，而原来的第一条方程正好提供了它：$\dot{x}(0) = 4\big(x(0) - y(0)\big) = 0$。

由 $x(0) = 0$ 得 $A + 3 = 0$，所以 $A = -3$。微分得

$$\dot{x} = \big(B + 2A + 2Bt + 48t + 48t^2\big)\mathrm{e}^{2t} - 6\mathrm{e}^{-2t},$$

所以 $\dot{x}(0) = B + 2A - 6 = 0$，得 $B = 12$。因此

$$\boxed{x = \left(24t^2 + 12t - 3\right)\mathrm{e}^{2t} + 3\mathrm{e}^{-2t}.}$$

求 y

不必再解另一条微分方程，因为第一条方程给出 $y = x - \frac{1}{4}\dot{x}$。代入 $A = -3$ 和 $B = 12$，

$$\dot{x} = \left(48t^2 + 72t + 6\right)\mathrm{e}^{2t} - 6\mathrm{e}^{-2t},$$

所以

$$\begin{aligned} y &= \left(24t^2 + 12t - 3\right)\mathrm{e}^{2t} + 3\mathrm{e}^{-2t} - \left(12t^2 + 18t + \tfrac{3}{2}\right)\mathrm{e}^{2t} + \tfrac{3}{2}\mathrm{e}^{-2t} \\ &= \left(12t^2 - 6t - \tfrac{9}{2}\right)\mathrm{e}^{2t} + \tfrac{9}{2}\mathrm{e}^{-2t}. \end{aligned}$$

即

$$\boxed{y = \tfrac{3}{2}\left(8t^2 - 4t - 3\right)\mathrm{e}^{2t} + \tfrac{9}{2}\mathrm{e}^{-2t}.}$$

验算

- 当 $t = 0$ 时：$y(0) = -\frac{9}{2} + \frac{9}{2} = 0$。这个条件从未直接用上，所以这是一项真正的验算。
- 第二条方程我们只用来消去 y，它也应该成立。微分得

$$\dot{y} = \left(24t^2 + 12t - 15\right)\mathrm{e}^{2t} - 9\mathrm{e}^{-2t},$$

而

$$x - 12\left(\mathrm{e}^{2t} + \mathrm{e}^{-2t}\right) = \left(24t^2 + 12t - 15\right)\mathrm{e}^{2t} - 9\mathrm{e}^{-2t}.$$

两者相符。

- 数值方面，用龙格—库塔法（Runge–Kutta method）由 $(0,0)$ 出发逐步推算原来的方程组，得 $x(1) \approx 244.2449$ 和 $y(1) \approx 11.6926$。公式给出 $33\mathrm{e}^2 + 3\mathrm{e}^{-2}$ 和 $\frac{3}{2}\mathrm{e}^2 + \frac{9}{2}\mathrm{e}^{-2}$，数值与上述结果相同。

常见错误

- 对共振项试用 $Ct\,\mathrm{e}^{2t}$。根 2 是重根，所以我们需要 $t^2\mathrm{e}^{2t}$。
- 看不出第二个初始条件从何而来。$\dot{x}(0) = 0$ 这个值隐藏在第一条方程之中，再配合 $x(0) = y(0)$ 便可得出。
- 另外为 y 解一条二阶方程。这样会引入两个新常数，之后还要把它们与 x 配对。利用 $y = x - \frac{1}{4}\dot{x}$ 便可避免这个麻烦。

STEP 之外

消去 y 是解线性方程组（linear system）的其中一种方法。大学的做法是把它写成 $\dot{\mathbf{v}} = M\mathbf{v} + \mathbf{f}(t)$，其中 $M = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix}$。这里 M 有重特征值（eigenvalue）2，但只有一个特征向量（eigenvector），所以不能对角化（diagonalise），而 $t\,\mathrm{e}^{2t}$ 项正是由此而来。若尔当标准形（Jordan normal form）和矩阵指数（matrix exponential）e^{Mt} 可以把这一点说得精确。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）涵盖线性方程组，包括重特征值的情况。

详细解答

STEP III 1989，第 9 题

纯数学 · 幂级数求和

STEP III 1989，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/89-s3-q9

关键思路

每个级数（series）都是一个熟悉的幂级数（power series）$\sum a_r x^r$ 在某个特定的 x 处的取值，所以要做的就是认出它是哪一个。几何级数（geometric series）$\frac{1}{1-x} = 1 + x + x^2 + \cdots$ 是头两个级数的源头。对它求导，会带出一个因子 r；对它积分，则会除以 r。至于第三个，奇数的连乘积除以阶乘，正是指数为负分数的二项展开式（binomial expansion）所产生的形式。

解答

以下我们一直会用到：当 $|x| < 1$ 时，

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots = \sum_{r=1}^{\infty} x^{r-1}.$$

题目说明不必理会收敛（convergence）的问题，所以我们可以放心逐项求导、逐项积分。（我们用到的三个 x 值全都稳稳落在收敛区间（interval of convergence）之内，所以这样做其实是合理的。）

第 (i) 部分

通项是 $r\left(\frac{1}{2}\right)^{r-1}$，所以我们要求的是 $\sum rx^{r-1}$ 在 $x = \frac{1}{2}$ 处的值。把几何级数逐项求导，

$$\sum_{r=1}^{\infty} rx^{r-1} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}.$$

代入 $x=\frac{1}{2}$：

$$1+\frac{2}{2}+\frac{3}{2^2}+\frac{4}{2^3}+\cdots=\frac{1}{\left(\frac{1}{2}\right)^2}=4.$$

第 (ii) 部分

通项是 $\frac{1}{r}\left(\frac{1}{2}\right)^{r-1}$。求导可以抵销除以 r 这一步，所以这次我们改为积分。把几何级数由 0 积分至 x，

$$\sum_{r=1}^{\infty}\frac{x^r}{r}=\int_0^x\frac{\mathrm{d}t}{1-t}=-\ln(1-x).$$

这里的幂次高了一次，所以我们除以 x（其中 $x\neq 0$）：

$$\sum_{r=1}^{\infty}\frac{1}{r}x^{r-1}=-\frac{\ln(1-x)}{x}.$$

代入 $x=\frac{1}{2}$：

$$1+\frac{1}{2}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{2^2}+\cdots=-\frac{\ln\frac{1}{2}}{\frac{1}{2}}=2\ln 2.$$

第 (iii) 部分

设

$$c_k=\frac{1\times 3\times\cdots\times(2k-1)}{k!},\qquad c_0=1.$$

? **想一想** 你以前在哪里见过奇数的连乘积除以 $k!$?

就在 $(1+y)^{-1/2}$ 的二项展开式里。y^k 的系数是

$$\binom{-\frac{1}{2}}{k}=\frac{\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2k-1}{2}\right)}{k!}=\frac{(-1)^k}{2^k}\cdot\frac{1\times 3\times\cdots\times(2k-1)}{k!}=\left(-\frac{1}{2}\right)^k c_k.$$

因此，取 $y=-2x$，正负号和 2 的幂便互相抵销：

$$(1-2x)^{-1/2}=\sum_{k=0}^{\infty}\binom{-\frac{1}{2}}{k}(-2x)^k=\sum_{k=0}^{\infty}c_kx^k=1+x+\frac{1\times 3}{2!}x^2+\frac{1\times 3\times 5}{3!}x^3+\cdots,$$

此式在 $|2x|<1$ 时成立。

现在仔细看看题目给出的级数。它的通项是 $c_k\cdot\frac{1}{3^{k-1}}=3c_k\left(\frac{1}{3}\right)^k$，而且由 $k=2$ 开始（首项是 $\frac{1\times 3}{2!}\times\frac{1}{3}$）。所以它等于上面的级数在 $x=\frac{1}{3}$ 处的值，去掉 $k=0$ 和 $k=1$ 两项之后，

再乘以 3：

$$\begin{aligned}\sum_{k=2}^{\infty} c_k \frac{1}{3^{k-1}} &= 3\left(\sum_{k=0}^{\infty} c_k \left(\tfrac{1}{3}\right)^k - c_0 - c_1 \cdot \tfrac{1}{3}\right)\\ &= 3\left(\left(1-\tfrac{2}{3}\right)^{-1/2} - 1 - \tfrac{1}{3}\right)\\ &= 3\left(\sqrt{3} - \tfrac{4}{3}\right)\\ &= 3\sqrt{3} - 4.\end{aligned}$$

这里 $\left(\frac{1}{3}\right)^{-1/2} = \sqrt{3}$，而 $x = \frac{1}{3}$ 满足 $|2x| < 1$。

验算

- 第 (i) 部分：部分和（partial sums）1, 2, 2.75, 3.25, 3.5625, … 逐步向 4 靠近，而把 200 项加起来，在机器精度之内得到 4。
- 第 (ii) 部分：$2\ln 2 \approx 1.3863$，而部分和 1, 1.25, 1.3333, 1.3646, … 与此吻合。
- 第 (iii) 部分：各项为 0.5, 0.2778, 0.1620, …，把 400 项加起来得到 1.196152…，与 $3\sqrt{3} - 4 \approx 1.1962$ 相符。再做一个简单的合理性检查：答案必须大于首项 0.5，而且是正数；这就排除了忘记去掉 $k = 1$ 一项的情况（那样会得到 $3\sqrt{3} - 3 \approx 2.196$，由于各项大致按几何级数递减，这个值太大了）。

常见错误

- 在 (ii) 中，忘记除以 x：$-\ln(1-x)$ 在 $x = \frac{1}{2}$ 处的值是 $\ln 2$，只有答案的一半。
- 在 (iii) 中，没有留意到级数由 $k = 2$ 开始，或者漏掉了由 $\frac{1}{3^{k-1}} = 3 \cdot \frac{1}{3^k}$ 而来的额外因子 3。

STEP 之外

幂级数在其收敛半径（radius of convergence）之内可以逐项求导和逐项积分，原因是幂级数在严格位于该半径之内的每一个闭区间上都一致收敛（converges uniformly）。这是大学一年级分析学（analysis）的标准结果，也正因如此，“不必考虑收敛问题”这句话在这里是稳妥的。指数并非整数的二项级数（binomial series），正是这一类幂级数。Abbott 的 *Understanding Analysis*（分析学入门教材）对一致收敛和幂级数有清晰的讲解。

详细解答

STEP III 1989，第 10 题

纯数学 · 黎曼和

> **STEP III 1989，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/89-s3-q10

本题在 *Beyond STEP* 第 4.1 节甚么是积分？中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/riemann.html

关键思路

连续整数之积（product of consecutive integers）是 r^5 的“下降幂”（falling power）版本，而且可以用裂项相消（telescoping）求和，与 r^5 的积分如出一辙：两个相邻的六项乘积之差，等于一个五项乘积的六倍。第 (i) 和第 (ii) 部分分别给出一个略大于和一个略小于 $\sum r^5$ 的乘积之和。第 (iii) 部分再把 x^5 的两个黎曼和（Riemann sum）夹在这两个界之间，而两个界趋向同一个极限。

解答

第 (i) 部分

设 $P(r) = r(r+1)(r+2)(r+3)(r+4)(r+5)$。那么 $P(r)$ 与 $P(r-1) = (r-1)r(r+1)(r+2)(r+3)(r+4)$ 有五个公共因子 $r(r+1)\cdots(r+4)$，所以

$$\begin{aligned} P(r) - P(r-1) &= r(r+1)(r+2)(r+3)(r+4)\big[(r+5) - (r-1)\big] \\ &= 6\,r(r+1)(r+2)(r+3)(r+4). \end{aligned}$$

由 $r = 1$ 加至 n，右边裂项相消：

$$\begin{aligned} \sum_{r=1}^{n} r(r+1)(r+2)(r+3)(r+4) &= \tfrac{1}{6}\sum_{r=1}^{n}\big[P(r) - P(r-1)\big] = \tfrac{1}{6}\big[P(n) - P(0)\big] \\ &= \tfrac{1}{6}\,n(n+1)(n+2)(n+3)(n+4)(n+5), \end{aligned}$$

因为 $P(0)=0$。（对 n 使用归纳法（induction）同样可行；归纳步骤用的正是同一个因式分解。）

至于推论：若 $r\geq 1$，则对 $k=1,2,3,4$，每个因子都满足 $r+k>r>0$，所以

$$r^5 < r(r+1)(r+2)(r+3)(r+4) \qquad (r\geq 1).$$

把 $r=1,\ldots,n$ 的这些严格不等式（strict inequality）相加，便得 $\sum_{r=1}^{n} r^5 < \frac{1}{6}n(n+1)(n+2)(n+3)(n+4)(n+5)$，正是所求。

第 (ii) 部分

现在我们需要一个位于 r^5 之下的乘积，所以把各因子向下平移：改用 $(r-4)(r-3)(r-2)(r-1)r$。设 $Q(r)=(r-4)(r-3)(r-2)(r-1)r(r+1)$，同样的因式分解给出

$$\begin{aligned} Q(r)-Q(r-1) &= (r-4)(r-3)(r-2)(r-1)r\big[(r+1)-(r-5)\big] \\ &= 6\,(r-4)(r-3)(r-2)(r-1)r. \end{aligned}$$

由 $r=0$ 加至 $n-1$：

$$\begin{aligned} \sum_{r=0}^{n-1}(r-4)(r-3)(r-2)(r-1)r &= \tfrac{1}{6}\big[Q(n-1)-Q(-1)\big] \\ &= \tfrac{1}{6}\,(n-5)(n-4)(n-3)(n-2)(n-1)n, \end{aligned}$$

因为 $Q(-1)=(-5)(-4)(-3)(-2)(-1)\cdot 0=0$。

> **想一想** 对和式中的每一个 r，$(r-4)(r-3)(r-2)(r-1)r<r^5$ 是否都成立？

不尽然，而这正是条件 $n>1$ 发挥作用的地方。我们逐项比较：

- $r=0$：两边都是 0，所以只有等式成立。
- $r=1,2,3,4$：乘积中有一个因子为零，所以乘积是 $0<r^5$。
- $r\geq 5$：对 $k=1,2,3,4$，每个因子都满足 $0<r-k<r$，所以乘积 $<r^5$。

因此对所有 $r\geq 0$，都有 $(r-4)(r-3)(r-2)(r-1)r\leq r^5$，而且对每个 $r\geq 1$，不等式都是严格的。若 $n>1$，由 $r=0$ 至 $n-1$ 的和包含 $r=1$ 这一项，所以最少有一个不等式是严格的，于是

$$\sum_{r=0}^{n-1} r^5 > \tfrac{1}{6}(n-5)(n-4)(n-3)(n-2)(n-1)n.$$

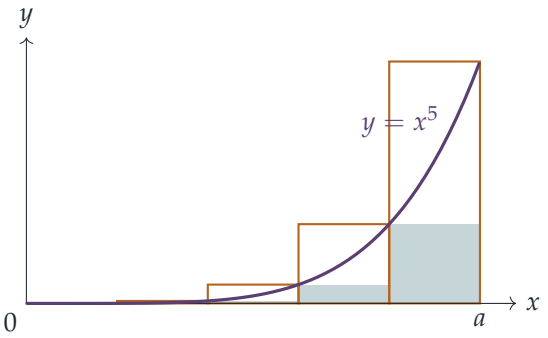
（当 $n=1$ 时，两边都是 0，这就是需要该条件的原因。）

第 (iii) 部分

取 $\mathrm{f}(x) = x^5$，这是一个递增函数（increasing function）。定义中的两个和是

$$L_n = \sum_{r=0}^{n-1} \frac{a}{n}\left(\frac{ra}{n}\right)^5 = \frac{a^6}{n^6}\sum_{r=0}^{n-1} r^5, \qquad U_n = \sum_{r=1}^{n} \frac{a}{n}\left(\frac{ra}{n}\right)^5 = \frac{a^6}{n^6}\sum_{r=1}^{n} r^5.$$

当 $a > 0$ 时，它们分别是曲线下方和上方各长方形的面积之和：



（涂色部分：L_n；框线部分：U_n；图中 $n = 5$。）不过，以下的代数运算并不需要 $a > 0$：对任何 $a \neq 0$ 都有 $a^6 > 0$，而 $a = 0$ 的情况显然成立，因为每个和都是 0。

对 L_n 使用第 (ii) 部分（其中 $n > 1$），对 U_n 使用第 (i) 部分，

$$L_n > \frac{a^6}{6n^6}(n-5)(n-4)(n-3)(n-2)(n-1)n = \frac{a^6}{6}\prod_{k=1}^{5}\left(1-\frac{k}{n}\right),$$

$$U_n < \frac{a^6}{6n^6}n(n+1)(n+2)(n+3)(n+4)(n+5) = \frac{a^6}{6}\prod_{k=1}^{5}\left(1+\frac{k}{n}\right).$$

此外 $U_n - L_n = \dfrac{a^6}{n^6}\left(n^5 - 0^5\right) = \dfrac{a^6}{n} > 0$，所以综合起来

$$\frac{a^6}{6}\prod_{k=1}^{5}\left(1-\frac{k}{n}\right) < L_n < U_n < \frac{a^6}{6}\prod_{k=1}^{5}\left(1+\frac{k}{n}\right).$$

当 $n \to \infty$ 时，每个括号都趋向 1，所以外侧的两个界都趋向 $\frac{1}{6}a^6$。根据夹逼定理（squeeze theorem，又称 sandwich theorem），L_n 和 U_n 都收敛（converge），而且收敛至同一个极限 $\frac{1}{6}a^6$。所以定义中的两个极限都存在而且相等，因此

$$\int_0^a x^5\,\mathrm{d}x = \tfrac{1}{6}a^6.$$

验算

- 第 (i) 部分的小数值情况：当 $n = 1$ 时，$1\cdot2\cdot3\cdot4\cdot5 = 120 = \frac{1}{6}\cdot720$；当 $n = 2$ 时，$120 + 720 = 840 = \frac{1}{6}\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7$。

- 第 (i) 和第 (ii) 部分的恒等式（identity），以及两个不等式，对 200 以内的每个 n 都准确成立（已用电脑核对）。
- 取 $a = 1$ 和 $n = 1000$：$L_n \approx 0.166167$，$U_n \approx 0.167167$，分别位于 $\frac{1}{6} \approx 0.166667$ 的两侧，而且 $U_n - L_n = 0.001 = a^6/n$。

常见错误

- 声称对所有 $r \geq 0$ 都有 $(r-4)\cdots r < r^5$：在 $r = 0$ 时其实是等式，所以严格不等式要靠 $r = 1$ 这一项，亦即需要 $n > 1$。
- 只证明 L_n 和 U_n 有界（bounded），或只证明 $U_n - L_n \to 0$。定义要求两个极限都*存在*；夹逼定理则一并给出存在性和极限值。

STEP 之外

第 (iii) 部分是黎曼（Riemann）积分（或称达布（Darboux）积分）的一次亲手实践：若一个函数的下和（lower sum）与上和（upper sum）可以夹在一起，该函数便是可积（integrable）的；对单调函数（monotonic function）而言，两者之差恒为 $(b-a)\,|f(b) - f(a)|/n$，所以闭区间（closed interval）上的每个单调函数都可积。第 (i) 和第 (ii) 部分的乘积分别是“上升阶乘幂”（rising factorial power）和“下降阶乘幂”（falling factorial power），它们令有限差分演算（calculus of finite differences）看起来与普通的微积分一模一样。Abbott 的 *Understanding Analysis*（一本实分析入门书）正是以这种方式建立积分理论。

详细解答

STEP III 1989，第 11 题

力学 · 球体的斜碰撞

STEP III 1989，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/89-s3-q11

关键思路

首先确定碰撞一刻的几何关系：它决定了连心线（line of centres）的方向，其余一切都由此而来。然后套用光滑（smooth）球体斜碰撞（oblique impact）而其中一个球受约束（constraint）时的标准方法。两球之间的冲量（impulse）沿连心线作用。牛顿恢复定律（Newton's law of restitution）适用于沿该直线的速度分量。至于动量守恒（conservation of momentum），由于绳子或支点（pivot）本身也提供冲量，这里并不成立；我们改用不受约束影响的那一条方程：有绳子时用 A 的水平动量，有杆时用对 O 的角动量（angular momentum）。

解答

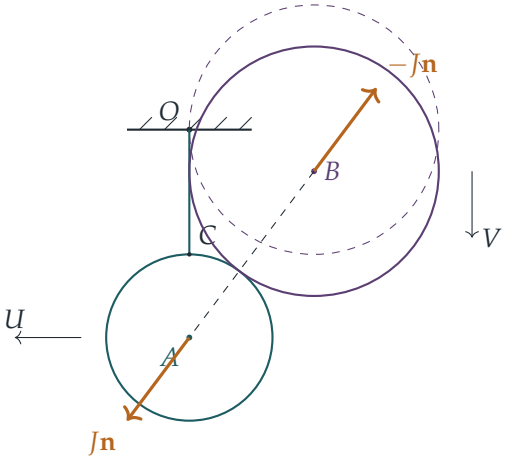
碰撞时的几何关系

我们把两个球都视为光滑而均匀（uniform），并把绳子视为轻而不可伸长（light and inextensible）。碰撞前绳子垂直悬挂。由于 C 位于直线 OA 上，我们有 $OA = 3a + 2a = 5a$，而 A 在 O 的正下方。

球 B 起初的位置使 $OB = 3a$ 为水平，然后垂直下落。设它下落的距离为 h。碰撞时两球相切，所以 $AB = 2a + 3a = 5a$。A 与 B 的水平距离是 $3a$，所以两者的垂直距离是 $\sqrt{25a^2 - 9a^2} = 4a$。因此 B 在 A 之上 $4a$ 处，亦即在 O 的水平线之下 a 处，故 $h = a$。

所以连心线构成一个 3–4–5 三角形：由 B 到 A，它沿水平方向（朝向 OA）走 $3a$，再向下走 $4a$。设 $\mathbf{n}$ 为由 B 指向 A 的单位向量（unit vector），并取 x 轴为水平，方向由 OA 指向 B。那么

$$\mathbf{n} = \left(-\tfrac{3}{5}, -\tfrac{4}{5}\right).$$



虚线的球标示 B 的起始位置，实线的球则标示它在碰撞时的位置。

有绳子时的碰撞

设 J 为两球之间冲量的大小，所以 A 受到 $J\mathbf{n}$，而 B 受到 $-J\mathbf{n}$。两球都是光滑的，所以 $J\mathbf{n}$ 沿连心线作用，并通过 A。绳子的冲量沿 OC 作用，同样通过 A。所以没有任何东西令球 A 自转，A 的运动就如同一个质点（particle），在以 O 为圆心、半径为 $5a$ 的圆上运动。紧接碰撞之后绳子仍然垂直，所以 A 沿水平方向运动，速率为 U，方向为 x 的负方向。

*球 A，水平方向。*绳子不提供水平冲量，所以

$$3mU = \tfrac{3}{5}J, \qquad \text{即} \qquad J = 5mU.$$

（在垂直方向，绳子提供一个向上的冲量 $\frac{4}{5}J$，抵消了 $J\mathbf{n}$ 向下的部分。这个量是正数，所以绳子确实保持拉紧。）

*球 B。*碰撞前 $\mathbf{v}_B = (0, -V)$，它沿 $\mathbf{n}$ 的分量是 $\frac{4}{5}V$。冲量 $-J\mathbf{n}$ 只改变这个分量，并使它减少 $J/(25m)$：

$$\mathbf{v}'_B \cdot \mathbf{n} = \tfrac{4}{5}V - \frac{J}{25m} = \tfrac{4}{5}V - \tfrac{1}{5}U.$$

？ 想一想 为甚么我们不能直接沿连心线运用动量守恒？

绳子对 A 的冲量是垂直的，而它沿 $\mathbf{n}$ 有一个分量。所以沿 $\mathbf{n}$ 的总动量会改变。唯一没有外来冲量作用的方向是水平方向，而那正是我们对 A 所用的方程。

*恢复定律。*沿 $\mathbf{n}$ 方向，靠近速率（speed of approach）是 $\frac{4}{5}V - 0$，因为 A 起初静止。分离速率（speed of separation）是 $\mathbf{v}'_A \cdot \mathbf{n} - \mathbf{v}'_B \cdot \mathbf{n}$，其中 $\mathbf{v}'_A \cdot \mathbf{n} = (-U)(-\frac{3}{5}) = \frac{3}{5}U$。所以

$$\tfrac{3}{5}U - \left(\tfrac{4}{5}V - \tfrac{1}{5}U\right) = \tfrac{3}{4} \cdot \tfrac{4}{5}V \quad \Longrightarrow \quad \tfrac{4}{5}U = \tfrac{7}{5}V \quad \Longrightarrow \quad U = \tfrac{7}{4}V.$$

有杆时的碰撞

现在球与杆组成单一个刚体（rigid body），绕 O 转动。支点可以提供任何方向的冲量，所以水平动量不再守恒。但支点的冲量作用于 O，所以我们对 O 取矩（take moments）。根据平行轴定理（parallel axis theorem），它对 O 的转动惯量（moment of inertia）是

$$I = \tfrac{2}{5}(3m)(2a)^2 + 3m(5a)^2 = \tfrac{24}{5}ma^2 + 75ma^2 = \tfrac{399}{5}ma^2.$$

冲量 $J\mathbf{n}$ 的作用线通过 A，而它位于 O 之下 $5a$ 的深度。只有它的水平分量 $\frac{3}{5}J$ 对 O 有矩，而该矩为 $5a \cdot \frac{3}{5}J = 3aJ$。若刚体在紧接碰撞之后的角速率（angular speed）是 ω，则 A 以速率 $U' = 5a\omega$ 沿水平方向运动，并且

$$I\omega = 3aJ \quad\Longrightarrow\quad \tfrac{399}{5}ma^2 \cdot \frac{U'}{5a} = 3aJ \quad\Longrightarrow\quad J = \tfrac{133}{25}\,mU'.$$

关于 B 的方程和恢复定律的方程与之前完全相同，因为它们只涉及沿 $\mathbf{n}$ 的分量。所以

$$\tfrac{3}{5}U' - \left(\tfrac{4}{5}V - \frac{J}{25m}\right) = \tfrac{3}{5}V \quad\Longrightarrow\quad \left(\tfrac{3}{5} + \tfrac{133}{625}\right)U' = \tfrac{7}{5}V \quad\Longrightarrow\quad \tfrac{508}{625}U' = \tfrac{7}{5}V,$$

所以 $U' = \frac{875}{508}V$。把这个结果与 $U = \frac{7}{4}V$ 比较，

$$\frac{U'}{U} = \frac{875}{508} \cdot \frac{4}{7} = \frac{500}{508} = \frac{125}{127},$$

所以 A 在紧接碰撞之后的速率是 $125U/127$，正如题目所求。

答案比之前稍小，这是合理的。有绳子时，A 抵抗水平运动的表现就像一个质量为 $3m$ 的质点。有杆时，球还必须开始转动，而它在 A 处的等效水平质量变成 $I/(5a)^2 = \frac{399}{125}m$，比 $3m$ 略大。

验算

- 能量：取 $m = a = V = 1$，碰撞前的动能（kinetic energy）是 12.5。碰撞后，有绳子时是 11.625，有杆时约为 11.58。两者都较小，在 $e < 1$ 时必然如此。
- 弹性极限情况：取 $e = 1$ 并用绳子，同样的步骤给出 $\frac{4}{5}U = \frac{8}{5}V$，所以 $U = 2V$，而动能恰好守恒。这确认了正负号无误。
- 用电脑以精确分数求解两次碰撞，取 $a = m = V = 1$。结果是 $U = 7/4$ 和 $U' = 875/508$，得出比值 $125/127$。电脑亦核对了 B 的切向速度（tangential velocity）不变，以及 B 在接触前下落的距离为 a。

常见错误

- 沿连心线运用总动量守恒。绳子或支点会提供外来冲量，而该冲量在这个方向上有分量。

- 忘记杆上的球会转动。它对 O 的转动惯量除了 $3m(5a)^2$ 之外，还需要 $\frac{2}{5}(3m)(2a)^2$ 这一项。
- 把连心线当作成 45° 角，或当作垂直。相切的条件决定了它构成一个 3–4–5 三角形。

STEP 之外

受约束的碰撞让我们初次接触*刚体的冲击运动*（impulsive motion of rigid bodies）。一般的法则是：固定支点上的冲量，不会出现在对该支点的角动量方程之中。由此可引伸出*撞击中心*（centre of percussion）的概念：敲击一个装有支点的物体时，若击中这一点，支点便完全感受不到冲量，这就是板球棒有“甜蜜点”的原因。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）正是以这个程度讲解角动量和刚体冲量。

详细解答

STEP III 1989，第 12 题

力学 · 旋转参考系

> **STEP III 1989，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/89-s3-q12

关键思路

平面是光滑（smooth）的，而杆是轻（light）杆并可绕枢轴自由转动，所以作用于 P 的水平力只能沿杆的方向。因此，P 的加速度在垂直于 AP 方向上的分量必定为零。只要求出这个分量，令它等于零，便得到关于 ϕ 的单摆方程（pendulum equation），再用惯常的“乘以 $\dot\phi$ 再积分”技巧，便可求出 $\dot\phi$。

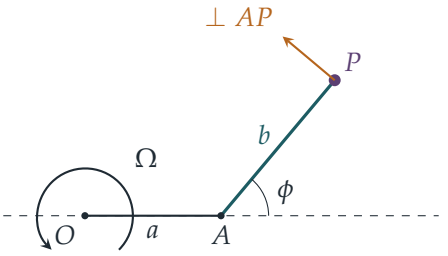
解答

模型与设置

把 P 看成光滑平面上的一个质点（particle），把杆看成轻而刚硬的杆，可绕 A 自由转动。我们在一个固定的惯性参考系（inertial frame）中考虑问题。用复数表示平面上的位置，以 O 为原点，实轴取为 $t=0$ 时 OA 的方向，并取 $\Omega>0$ 为逆时针方向。由于 A 固定在平面上，它沿半径为 a 的圆运动；而 AP 与 OA 的延长线成角 ϕ，该延长线本身已转过 Ωt。因此，记 $\theta=\Omega t+\phi$，便有

$$z_A = a\,\mathrm{e}^{\mathrm{i}\Omega t}, \qquad z_P = a\,\mathrm{e}^{\mathrm{i}\Omega t} + b\,\mathrm{e}^{\mathrm{i}\theta}.$$

这里 $\mathrm{e}^{\mathrm{i}\theta}$ 是沿 AP 方向的单位向量，而 $\mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta}$ 是垂直于 AP、指向 ϕ 增加方向的单位向量。



加速度的垂直分量

求导，其中 $\dot\theta = \Omega + \dot\phi$ 及 $\ddot\theta = \ddot\phi$（因为 Ω 是常数），得

$$\dot z_P = \mathrm{i}a\Omega\,\mathrm{e}^{\mathrm{i}\Omega t} + \mathrm{i}b(\Omega+\dot\phi)\,\mathrm{e}^{\mathrm{i}\theta},$$
$$\ddot z_P = -a\Omega^2\,\mathrm{e}^{\mathrm{i}\Omega t} + b\big(\mathrm{i}\ddot\phi - (\Omega+\dot\phi)^2\big)\mathrm{e}^{\mathrm{i}\theta}.$$

要把加速度分解为沿 AP 和垂直于它的分量，可写 $\mathrm{e}^{\mathrm{i}\Omega t} = \mathrm{e}^{\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}\phi} = \mathrm{e}^{\mathrm{i}\theta}(\cos\phi - \mathrm{i}\sin\phi)$。于是

$$\ddot z_P = \big[-a\Omega^2\cos\phi - b(\Omega+\dot\phi)^2\big]\mathrm{e}^{\mathrm{i}\theta} + \big[\,b\ddot\phi + a\Omega^2\sin\phi\,\big]\,\mathrm{i}\,\mathrm{e}^{\mathrm{i}\theta}.$$

所以，P 的加速度在垂直于 AP 方向（指向 ϕ 增加的一方）的分量是

$$\boxed{\,b\,\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2} + a\Omega^2\sin\phi\,}$$

（第一项是我们熟悉的 $b\ddot\phi$，来自绕 A 的圆周运动。第二项来自 A 本身指向 O 的向心加速度（centripetal acceleration）$a\Omega^2$，再分解到垂直于杆的方向。）

求 $\mathrm{d}\phi/\mathrm{d}t$

? **想一想** 为甚么加速度的这个分量必定为零?

作用于 P 的力有：它的重量和法向反作用力（normal reaction）（两者都是铅垂的，并且互相抵消）；没有摩擦力（平面是光滑的）；以及杆所施的力。一根可绕 A 自由转动的轻杆，只能沿自身的长度方向推或拉。所以在垂直于 AP 的方向上没有水平力，由牛顿第二定律（Newton's second law）得

$$b\ddot\phi + a\Omega^2\sin\phi = 0.$$

这就是单摆方程。乘以 $\dot\phi$ 再积分，得

$$\tfrac{1}{2}b\dot\phi^2 = a\Omega^2\cos\phi + C.$$

初始条件。开始时 O, A, P 在同一直线上。由于 $b > a$，P 不可能位于 O 与 A 之间，我们取 P 在 A 之外的一方，所以在 $t = 0$ 时 $\phi = 0$。这时两个指数函数的值都是 1，而

$$\dot{z}_P(0) = \mathrm{i}\big[(a+b)\Omega + b\dot{\phi}(0)\big],$$

它垂直于 OP（即实轴）。留意 $(a+b)\Omega$ 正是平面上位于 P 正下方那一点的速率，而 $b\dot{\phi}(0)$ 则是 P 相对于平面的速率。题目印出的速率是 $(a+2\sqrt{ab})\Omega$，但这看来是 $(a+b+2\sqrt{ab})\Omega = (\sqrt{a}+\sqrt{b})^2\Omega$ 的误印，因为后者才能得出题目所述的结果（见文末的备注）。采用这个值，P 相对于平面的速率是 $2\sqrt{ab}\,\Omega$，所以

$$\dot{\phi}(0) = 2\Omega\sqrt{a/b}.$$

把 $\phi = 0$ 代入能量方程，得 $\frac{1}{2}b \cdot 4a\Omega^2/b = a\Omega^2 + C$，所以 $C = a\Omega^2$，而

$$\tfrac{1}{2}b\dot{\phi}^2 = a\Omega^2(1+\cos\phi) = 2a\Omega^2\cos^2\tfrac{\phi}{2}, \qquad \text{所以} \qquad \dot{\phi}^2 = \frac{4a\Omega^2}{b}\cos^2\tfrac{\phi}{2}.$$

开始时 $\dot{\phi} > 0$，而 $\dot{\phi}$ 只有经过零才能改变正负号，这需要 $\cos\frac{\phi}{2} = 0$，即 $\phi = \pm\pi$。所以只要 $-\pi < \phi < \pi$，我们便取正根：

$$\boxed{\frac{\mathrm{d}\phi}{\mathrm{d}t} = 2\Omega\sqrt{\frac{a}{b}}\cos\frac{\phi}{2}}$$

P 永不越过 OA

> **想一想** ϕ 难道不可以慢慢爬升到 π，然后越过它吗？

只有当 ϕ 是 π 的倍数时，P 才会在直线 OA 上（当 $\phi = \pi$ 时，它位于 O 之外、距离为 $b-a$ 的地方）。由于当 $0 \le \phi < \pi$ 时 $\dot{\phi} > 0$，ϕ 由 0 开始递增，而回到这条直线上的唯一途径就是到达 $\phi = \pi$。我们解出这个方程，从而证明这种情况永不发生。分离变量（separating the variables），并设 $k = \sqrt{a/b}$，得

$$\int_0^{\phi} \sec\tfrac{u}{2}\,\mathrm{d}u = 2k\Omega t \quad\Longrightarrow\quad \ln\Big(\sec\tfrac{\phi}{2} + \tan\tfrac{\phi}{2}\Big) = k\Omega t.$$

设 $s = \sin\frac{\phi}{2}$，则 $\sec\frac{\phi}{2} + \tan\frac{\phi}{2} = \frac{1+s}{\sqrt{1-s^2}} = \sqrt{\frac{1+s}{1-s}}$，所以左边等于 $\frac{1}{2}\ln\frac{1+s}{1-s} = \operatorname{artanh} s$。因此

$$\sin\frac{\phi}{2} = \tanh\Big(\Omega t\sqrt{\tfrac{a}{b}}\Big).$$

对于每个 $t > 0$，右边都严格介乎 0 与 1 之间，所以对所有 $t > 0$ 都有 $0 < \phi < \pi$。于是 P 始终严格地留在直线 OA 的同一侧。当 $t \to \infty$ 时，它趋近这条直线（趋近 O 之外的那一点），但永远不会到达，所以 P 永不越过 OA。 ■

（不解方程的做法：$\phi = \pi, \dot{\phi} = 0$ 是单摆方程的一个平衡点（equilibrium），而在它附近，$\psi = \pi - \phi$ 满足 $\dot{\psi} \approx -k\Omega\psi$，所以 ψ 以指数形式衰减，永不为零。）

关于印出的速率的备注。如果我们按字面把速率取为 $(a+2\sqrt{ab})\Omega$，则 $\dot\phi(0)=(2k-1)\Omega$，而同样的积分给出 $\dot\phi^2=\Omega^2\left[(2k-1)^2-4k^2\sin^2\frac{\phi}{2}\right]$。这并不是一个简洁的式子。这时 P 会绕 $\phi=0$ 来回摆动（每摆一次都越过 OA 的延长线），如果 $b<16a$；又或者转足一整圈，越过 OA，如果 $b>16a$。所以印出的数值不可能是出题者的原意。

验算

- 令 $a=0$，A 便落在 O 上，这时 $\dot\phi\equiv 0$：P 只是随着平面以速率 $b\Omega$ 一起转动。这是对的，因为初始速率变成了 $b\Omega$。
- 单位：$\Omega\sqrt{a/b}$ 是一个变率，所以它的单位是 s^{-1}，正如 $\dot\phi$ 应有的单位。
- 我们以数值方法（RK4）逐步求解 $b\ddot\phi=-a\Omega^2\sin\phi$，取 $a=1, b=3, \Omega=1.3$。在 $t=2$ 时，得出 $\phi\approx 2.2644$ 和 $\dot\phi\approx 0.6375$，与 $2\arcsin\tanh(\Omega t\sqrt{a/b})$ 和 $2\Omega\sqrt{a/b}\cos\frac{\phi}{2}$ 相符，准确至 10 位有效数字。在 $t=10$ 时，$\phi\approx 3.1394<\pi$。

常见错误

- 忘记了参考系本身对 P 的初始速度所贡献的 $(a+b)\Omega$，把题目给出的速率当成相对于平面的速率。
- 把 P 的加速度只写成绕 A 作圆周运动的加速度。枢轴 A 本身也在加速，$a\Omega^2\sin\phi$ 这一项正是由此而来。
- 把 $a\Omega^2\sin\phi$ 的正负号弄错。如果符号错了，$\phi=0$ 便会是不稳定的，这不符合物理实际，因为“离心”拉力是背向 O 的。

STEP 之外

从旋转的平面上看，P 是一个单摆，悬挂在大小为 $a\Omega^2$、方向沿 OA 延长线的均匀“重力”之中。科里奥利力（Coriolis force）沿杆的方向作用，所以它不做功，可以不予考虑。题目的初始条件令运动恰好落在 分界线（*separatrix*）上：这是来回摆动与越过顶点之间的临界情况，此时单摆需要无限长的时间才到达倒立位置。旋转参考系（rotating frames）、虚拟力（fictitious forces）以及这类相平面（phase plane）图像，可参阅 Kleppner and Kolenkow, *An Introduction to Mechanics*（一本大学力学入门教科书）。

详细解答

STEP III 1989，第 13 题

力学 · 简谐运动

> **STEP III 1989，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/89-s3-q13

关键思路

每条弦线都由 O（在桌面之下）出发，向上穿过所属的小孔，再沿桌面连到 P。桌下的一段长度恒为 a，正好等于自然长度（natural length），所以每条弦线的伸长量（extension）就是 P 与该小孔之间的距离。这样每个张力（tension）都与一个向量成正比；用位置向量（position vector）来表示，五个力只需一行便可加起来，因为五个小孔对称排列，它们的位置向量之和为零。

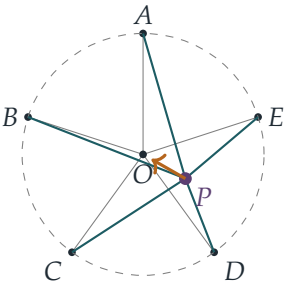
解答

建立模型

把 P 视为光滑桌面上的质点（particle），弦线视为轻弹性弦线（light elastic string），小孔视为光滑，因此每条弦线上各处的张力相同。取桌面上的 O 为原点，设 $\mathbf{r}$ 为 P 的位置向量，并设 $\mathbf{a}_1, \ldots, \mathbf{a}_5$ 为 $A, \ldots, E$ 的位置向量，每个的长度均为 a。

穿过第 i 个小孔的弦线由两段组成：一段是由 O（在桌面之下，位于中心的正下方）到小孔的垂直部分，长度为 a；另一段在桌面上，长度为 $|\mathbf{a}_i - \mathbf{r}|$。弦线的总长度是 $a + |\mathbf{a}_i - \mathbf{r}|$，所以伸长量是 $|\mathbf{a}_i - \mathbf{r}|$。这个量永不为负，因此无论 P 在哪里，每条弦线都是拉紧的（或当 P 位于小孔时刚好松弛）：不会有弦线松掉而令方程改变。根据虎克定律（Hooke's law），张力为 $\frac{\lambda}{a}|\mathbf{a}_i - \mathbf{r}|$，而在桌面上它把 P 拉向小孔，方向为单位向量 $\frac{\mathbf{a}_i - \mathbf{r}}{|\mathbf{a}_i - \mathbf{r}|}$。所以第 i 条弦线作用于 P 的力是

$$\frac{\lambda}{a}|\mathbf{a}_i - \mathbf{r}| \cdot \frac{\mathbf{a}_i - \mathbf{r}}{|\mathbf{a}_i - \mathbf{r}|} = \frac{\lambda}{a}(\mathbf{a}_i - \mathbf{r}).$$



蓝绿色的线是从上方俯视时桌面上的五条弦线；琥珀色箭头表示作用于 P 的合力（resultant force），它正正指向 O。

第一部分

作用于 P 的垂直方向的力（重量与法向反作用力）互相平衡。把五个水平方向的力相加，

$$\mathbf{F} = \frac{\lambda}{a}\sum_{i=1}^{5}(\mathbf{a}_i - \mathbf{r}) = \frac{\lambda}{a}\Big(\sum_{i=1}^{5}\mathbf{a}_i\Big) - \frac{5\lambda}{a}\mathbf{r}.$$

? **想一想** 为甚么 $\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4 + \mathbf{a}_5 = \mathbf{0}$?

把整个图形绕 O 旋转 72°，五个小孔只是互相置换，所以总和 $\mathbf{s} = \sum \mathbf{a}_i$ 保持不变。但旋转 72° 后仍然不变的向量只有零向量，所以 $\mathbf{s} = \mathbf{0}$。（换句话说，五次单位根（fifth roots of unity）之和为零。）

因此由牛顿第二定律（Newton's second law）可得

$$m\ddot{\mathbf{r}} = -\frac{5\lambda}{a}\mathbf{r}, \qquad \text{即} \qquad \ddot{\mathbf{r}} = -\omega^2\mathbf{r} \quad \text{其中} \quad \omega^2 = \frac{5\lambda}{am}.$$

P 由 F 从静止开始运动，所以 $\mathbf{r}(0) = \overrightarrow{OF}$ 及 $\dot{\mathbf{r}}(0) = \mathbf{0}$。解为 $\mathbf{r} = \cos(\omega t)\,\overrightarrow{OF}$，所以 P 沿直线 FO 来回运动（去到 F 关于 O 的反射点再折返），而它相对于 O 的位移满足简谐运动（simple harmonic motion）的方程。周期（period）为

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{am}{5\lambda}},$$

这正是所需的结果。（注意这对任何起点 F 都成立，并不限于小位移：力的规律是完全线性的。）

最后部分

现在弦线 PAO 的弹性模量（modulus）为 $k\lambda$，所以它作用于 P 的力是 $\frac{k\lambda}{a}(\mathbf{a}_1-\mathbf{r})$。利用 $\mathbf{a}_2+\mathbf{a}_3+\mathbf{a}_4+\mathbf{a}_5=-\mathbf{a}_1$，

$$\begin{aligned}\mathbf{F}&=\frac{k\lambda}{a}(\mathbf{a}_1-\mathbf{r})+\frac{\lambda}{a}\sum_{i=2}^{5}(\mathbf{a}_i-\mathbf{r})\\&=\frac{k\lambda}{a}(\mathbf{a}_1-\mathbf{r})+\frac{\lambda}{a}(-\mathbf{a}_1-4\mathbf{r})\\&=\frac{\lambda}{a}\big[(k-1)\mathbf{a}_1-(k+4)\mathbf{r}\big]=-\frac{(k+4)\lambda}{a}\,(\mathbf{r}-\mathbf{r}_0)\,,\qquad \mathbf{r}_0=\frac{k-1}{k+4}\,\mathbf{a}_1.\end{aligned}$$

所以平衡位置（equilibrium position）是 OA 上的点 $\mathbf{r}_0$（该处 $\mathbf{F}=\mathbf{0}$），并且

$$\ddot{\mathbf{r}}=-\frac{(k+4)\lambda}{am}(\mathbf{r}-\mathbf{r}_0).$$

若把 P 由 $\mathbf{r}_0$ 沿 OA 移开，然后从静止释放，则 $\mathbf{r}-\mathbf{r}_0$ 起初与 $\mathbf{a}_1$ 平行，加速度又恒与 $\mathbf{r}-\mathbf{r}_0$ 平行，而初速度为零。所以与之前完全一样，$\mathbf{r}-\mathbf{r}_0=\cos(\omega' t)(\mathbf{r}(0)-\mathbf{r}_0)$：$P$ 沿 OA 作直线运动，以 $\mathbf{r}_0$ 为中心作简谐运动，角频率（angular frequency）为 ω'，其中

$$\omega'^2=\frac{(k+4)\lambda}{am}.$$

周期为 $T/2$，当且仅当 $\omega'=2\omega$，即 $\omega'^2=4\omega^2$：

$$\frac{(k+4)\lambda}{am}=\frac{20\lambda}{am}\quad\Longrightarrow\quad k+4=20\quad\Longrightarrow\quad k=16.$$

当 $k=16$ 时，平衡位置是 $\mathbf{r}_0=\frac{15}{20}\mathbf{a}_1$，即由 O 到 A 的四分之三处。（要令模型成立，振幅（amplitude）必须小于 $\frac{1}{4}a$，使 P 不会到达 A 处的小孔。）

验算

- 代入 $k=1$ 便回到第一部分：$\mathbf{r}_0=\mathbf{0}$ 及 $\omega'^2=5\lambda/(am)$。
- 单位：λ 是一个力，所以 $\lambda/(am)$ 的单位是 $\mathrm{N/(m\,kg)}=\mathrm{s^{-2}}$，而 T 是时间。
- 我们在 1000 个随机位置把五个张力向量相加（取 $a=1.3$、$m=0.7$、$\lambda=2.1$），结果恰好是 $-5\lambda\mathbf{r}/a$；而当 $k=16$ 时恰好是 $-20\lambda(\mathbf{r}-\mathbf{r}_0)/a$。由 $(0.4,-0.3)$ 出发，以数值方法逐步计算运动，P 在时间 $T\approx 1.85$ 后回到起点，并一直留在通过 O 的直线上；当 $k=16$ 时由 $\mathbf{r}_0+0.1a\,\hat{\mathbf{a}}_1$ 出发，则在 $T/2$ 后回到起点。

常见错误

- 把伸长量当作 $|\mathbf{a}_i-\mathbf{r}|-a$ 或 $|\mathbf{r}|$：自然长度 a 刚好全部用于桌下的那一段。

- 用角度和三角学来分解五个张力。这样做是可以的，但用向量的话，各项一眼便可看出互相抵消。
- 假设新的振动仍以 O 为中心：较硬的弦线会把平衡位置拉向 A。

STEP 之外

二维空间中的力 $\mathbf{F} = -c\,\mathbf{r}$ 给出各向同性谐振子（isotropic harmonic oscillator）：一般的运动（初速度带有横向分量时）是以平衡点为中心的椭圆，而且每条轨道的周期都相同。它与平方反比定律（inverse-square law）是仅有的两种令每条有界轨道都闭合的有心力（central force），这就是伯特兰定理（Bertrand's theorem）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）正是以这种向量方式处理振子和有心力。

详细解答

STEP III 1989，第 14 题

力学 · 滑轮与冲量

> **STEP III 1989，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/89-s3-q14

关键思路

第 (i) 部分：把圆盘切成许多薄圆环，每个圆环上的质量全都与圆心相距同一距离，因此质量和转动惯量（moment of inertia）都只是对 r 的单一积分。第 (ii) 部分：整个系统只有一个自由度（degree of freedom），即 D 的深度，所以一条运动方程（equation of motion）便可求出加速度。当 B 突然被截停时，绳的约束（constraint）立刻告诉我们 D 此后可以运动得多快，而冲量（impulse）正是它的动量（momentum）的改变。

解答

第 (i) 部分

（第 (i) 部分所指的轴是通过圆心 O 的轴；试卷上的“C”是印刷错误。）设与 O 相距 r 处的密度为 $\rho = kr$，其中 k 为常数。半径 r 与 $r + \mathrm{d}r$ 之间的薄圆环面积为 $2\pi r\,\mathrm{d}r$，故质量为 $2\pi kr^2\,\mathrm{d}r$，而且环上每一点与轴的距离都是 r。因此

$$m = \int_0^a 2\pi kr^2\,\mathrm{d}r = \frac{2\pi ka^3}{3}, \qquad I = \int_0^a r^2 \cdot 2\pi kr^2\,\mathrm{d}r = \frac{2\pi ka^5}{5}.$$

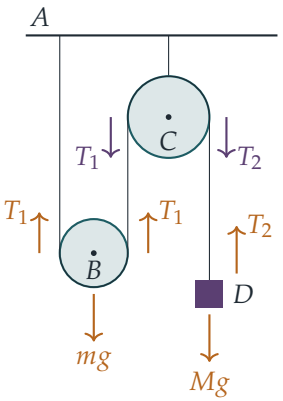
两式相除，得 $I/m = \frac{3}{5}a^2$，所以

$$I = \tfrac{3}{5}ma^2,$$

这正是所要证明的结果。（它比均匀圆盘的 $\frac{1}{2}ma^2$ 大，这是合理的：较多的质量集中在边缘附近。）

第 (ii) 部分：B 被截停前的运动

模型：B 是质量为 m 的光滑滑轮（pulley），C 是 (i) 中那种圆盘，质量为 m、半径为 a，所以它绕其固定轴的转动惯量是 $\frac{3}{5}ma^2$。绳是轻而不可伸长的，并且在 C 上不打滑。由于 B 是光滑的，AB 和 BC 两段绳的张力（tension）T_1 相同。把 CD 段绳的张力记为 T_2；由于 C 是粗糙的，T_2 不一定等于 T_1。



设 D 以加速度 f 向下运动。支承 B 的两段绳各自缩短的长度，是 CD 伸长量的一半，所以 B 以加速度 $\frac{1}{2}f$ 上升；又因为绳在 C 上不打滑，C 的角加速度（angular acceleration）是 f/a。运动方程为：

$$\begin{aligned} D\text{（向下）：}&\quad Mg - T_2 = Mf,\\ B\text{（向上）：}&\quad 2T_1 - mg = \tfrac{1}{2}mf,\\ C\text{（对其轴取矩）：}&\quad (T_2 - T_1)a = \tfrac{3}{5}ma^2 \cdot \frac{f}{a}. \end{aligned}$$

把 $T_2 = Mg - Mf$ 和 $T_1 = \frac{1}{2}mg + \frac{1}{4}mf$ 代入最后一条方程，

$$Mg - Mf - \tfrac{1}{2}mg - \tfrac{1}{4}mf = \tfrac{3}{5}mf \quad\Longrightarrow\quad f\left(M + \tfrac{17}{20}m\right) = g\left(M - \tfrac{1}{2}m\right),$$

所以

$$f = \frac{10(2M - m)g}{20M + 17m}.$$

这是常数，而且正因为 $m < 2M$ 才是正数：题目给出这个条件，原因就在这里。它说明 D 确实会由静止开始下降。因此 t 秒后，D 向下运动的速率为

$$v = ft = \frac{10(2M - m)gt}{20M + 17m},$$

B 以 $\frac{1}{2}v$ 上升，而 C 以角速率 v/a 转动。

（用能量验算可得相同的 f：当 D 下降了 x 时，动能（kinetic energy）为 $\frac{1}{2}\left(M + \frac{1}{4}m + \frac{3}{5}m\right)\dot{x}^2$，而失去的势能（potential energy）为 $(M - \frac{1}{2}m)gx$。）

第 (ii) 部分：冲量

> ? **想一想** B 被截停之后，D 还能运动吗?

只要绳保持拉紧，D 的速率总是 B 的速率的两倍，因为 A 是固定的，而 AB 和 BC 的长度每次都改变相同的量。一旦 B 静止，这两段的长度都固定了，于是再没有绳可以绕过 C，D 也必须停下来，C 亦然。绳确实保持拉紧：D 正在远离 C，所以它把 CD 拉紧；而仍在转动的 C 则从 BC 一侧把绳往上扯，所以那一段也被拉紧。（事实上，算出来的冲击张力（impulsive tension）在 CD 中是 Mv，在 BC 中是 $(M+\frac{3}{5}m)v$，两者皆为正数，印证了这一点。）

所以 D 的速度由向下的 v 降至零，而（绳）作用于 D 的冲量等于它的动量改变：

$$\text{冲量施于} D = Mv = \frac{10M(2M-m)gt}{20M+17m} \quad \text{，方向铅垂向上。}$$

验算

- 若 $m \to 0$（轻滑轮），则 $f \to g$，冲量为 Mgt：D 只是自由下落，一如所料。
- 若 $m = 2M$，冲量为零：系统处于平衡状态，根本不会开始运动。
- 取 $M = 3$、$m = 2.5$、$g = 9.8$、$t = 1.3$，以数值方法逐步模拟运动，得 $Mv \approx 13.0507$，与公式吻合，而所得的动能亦与失去的势能相符。用数值积分算出的转动惯量比值是 0.6000。

常见错误

- 把 C 两边的张力当作相同。滑轮 C 有质量而且粗糙，正是两边张力之差令它转动。
- 忘记 B 的速率只有 D 的一半；这使有效质量中出现的是 $\frac{1}{4}m$，而不是 m。
- 只给出冲量的大小而不说明方向：它向上作用于 D。

STEP 之外

用单一坐标把动能写成 $\frac{1}{2}M_{\text{有效}}\dot{x}^2$，正是拉格朗日力学（Lagrangian mechanics）的起点；在那里，像滑轮几何这样的约束从一开始就已内置，张力根本不会出现。像截停 B 这类突然施加的约束，也可用同样的方法，以冲量方程处理。Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）以这种风格讲述刚体转动（rigid-body rotation）和冲量。

详细解答

STEP III 1989，第 15 题

概率与统计 • 中位数的分布

> **STEP III 1989，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/89-s3-q15

关键思路

$2n+1$ 个数值的中位数（median）落在 z 附近，当且仅当其中一个数值落在 z 处的一个极小区间内，另有 n 个数值在它之下、n 个在它之上。第 (i) 及第 (ii) 部分给出了“全部在下”和“全部在上”这两部分的概率，再数一数有多少种方法选定哪些数值担当哪个角色，便得到密度。之后，利用密度的积分等于 1 这个事实，不费吹灰之力就能求出该积分；把同一技巧用于以 $n+1$ 代替 n 的情况，便得到方差。

解答

在整个解答中，X 在 $[-c,c]$ 上的密度为 $\frac{1}{2c}$，所以对于 $-c \le k \le c$，

$$\mathrm{P}(X > k) = \frac{c-k}{2c}, \qquad \mathrm{P}(X < k) = \frac{c+k}{2c}.$$

第 (i) 部分

这 n 个数值互相独立（independent），所以把概率相乘：

$$\mathrm{P}(\text{全部 } n \text{ 个数值均大于} k) = \left(\frac{c-k}{2c}\right)^n.$$

第 (ii) 部分

同理，

$$\mathrm{P}(\text{全部 } n \text{ 个数值均小于} k) = \left(\frac{c+k}{2c}\right)^n.$$

中位数的密度

设 $X_1, \ldots, X_{2n+1}$ 为样本，Z 为其中位数，即第 $(n+1)$ 小的数值。由于 X 是连续的，出现相同数值的概率为 0，所以我们不予理会。每个 X_i 都在 $[-c, c]$ 内，所以 Z 也在其中。

取 $(-c, c)$ 内的 z 和一个小的 $\delta z > 0$。略去 $(\delta z)^2$ 阶的项后，事件 $z < Z \leq z + \delta z$ 发生，当且仅当

- 其中一个数值落在 $(z, z+\delta z]$ 内，
- 其余数值中有 n 个小于 z，并且
- 余下的 n 个数值大于 $z + \delta z$。

（有两个或以上数值落在这个小区间内的情况，其概率为 $(\delta z)^2$ 阶，在取极限时无关重要。）

想一想 把这 $2n+1$ 个数值分配到这三个角色，共有多少种方法?

落在小区间内的数值有 $2n+1$ 种选择，然后从其余数值中选出 n 个放在 z 之下，有 $\binom{2n}{n}$ 种方法。于是共有

$$(2n+1)\binom{2n}{n} = \frac{(2n+1)!}{(n!)^2}$$

种分配方法。这些分配互斥（mutually exclusive），而且每一种的概率相同。对于一种固定的分配，由独立性以及第 (i) 和第 (ii) 部分（取 $k = z$），其概率为

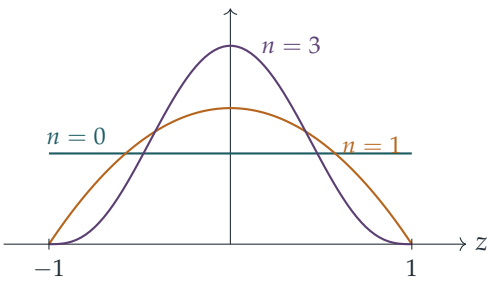
$$\frac{\delta z}{2c}\left(\frac{c+z}{2c}\right)^n \left(\frac{c-z-\delta z}{2c}\right)^n.$$

对所有分配求和，除以 δz，再令 $\delta z \to 0$，便得 Z 的密度为

$$\begin{aligned} g(z) &= \frac{(2n+1)!}{(n!)^2} \cdot \frac{1}{2c} \cdot \frac{(c+z)^n (c-z)^n}{(2c)^{2n}} \\ &= \frac{(2n+1)!}{(n!)^2 (2c)^{2n+1}} (c^2 - z^2)^n, \qquad -c \leq z \leq c, \end{aligned}$$

其中用到 $(c+z)(c-z) = c^2 - z^2$。这就是所要求的结果。

下图显示 $c = 1$ 时的密度。当 $n = 0$ 时，“中位数”只是单独一个数值，服从均匀分布（uniform distribution）。随着 n 增大，中位数愈来愈集中在 0 附近。



积分

概率密度（probability density）在其取值范围上的积分等于 1，所以

$$\frac{(2n+1)!}{(n!)^2(2c)^{2n+1}}\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z=1,$$

因此

$$\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z=\frac{(n!)^2(2c)^{2n+1}}{(2n+1)!}.$$

这对每个整数 $n\geq 0$ 都成立，稍后我们就会用到这一点。

平均值与方差

密度 g 是 z 的偶函数（even function），所以 $zg(z)$ 是奇函数，它在对称区间 $[-c,c]$ 上的积分为零：

$$\mathrm{E}(Z)=0.$$

因此 $\mathrm{var}(Z)=\mathrm{E}(Z^2)$。记 $K_n=\dfrac{(2n+1)!}{(n!)^2(2c)^{2n+1}}$ 及 $I_n=\displaystyle\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z=\frac{1}{K_n}$。

想一想 怎样才能不展开而求出 $z^2(c^2-z^2)^n$ 的积分？

写成 $z^2=c^2-(c^2-z^2)$。于是

$$\begin{aligned}\mathrm{E}(Z^2)&=K_n\int_{-c}^{c}\bigl(c^2-(c^2-z^2)\bigr)(c^2-z^2)^n\,\mathrm{d}z\\&=K_n\bigl(c^2I_n-I_{n+1}\bigr)=c^2-\frac{K_n}{K_{n+1}},\end{aligned}$$

其中用到 $K_nI_n=1$ 和 $I_{n+1}=1/K_{n+1}$，后者就是以 $n+1$ 代替 n 的积分结果。现在

$$\frac{K_n}{K_{n+1}}=\frac{(2n+1)!}{(2n+3)!}\cdot\frac{((n+1)!)^2}{(n!)^2}\cdot\frac{(2c)^{2n+3}}{(2c)^{2n+1}}=\frac{(n+1)^2\cdot 4c^2}{(2n+2)(2n+3)}=\frac{2(n+1)c^2}{2n+3}.$$

因此

$$\mathrm{var}(Z)=c^2\left(1-\frac{2n+2}{2n+3}\right)=\frac{c^2}{2n+3}.$$

验算

- 当 $n=0$ 时，中位数就是 X 的一个数值。密度变成 $\frac{1}{2c}$，积分为 $2c$，方差为 $\frac{c^2}{3}$，正是 $[-c,c]$ 上均匀分布的方差。
- 当 $n=1$ 及 $c=1$ 时，积分公式给出 $\int_{-1}^{1}(1-z^2)\,\mathrm{d}z=\frac{1\cdot 8}{6}=\frac{4}{3}$，结果正确。
- 以 10^5 个大小为 5 的样本（$n=2$）并取 $c=1.7$ 进行模拟，得到中位数的方差为 0.415，而 $c^2/7\approx 0.413$。

常见错误

- 忘记了选择中间数值所带来的因子 $2n+1$，只用了 $\binom{2n}{n}$。
- 漏掉中间数值本身的密度所带来的因子 $\frac{1}{2c}$，结果得到的幂是 $(2c)^{2n}$ 而不是 $(2c)^{2n+1}$。
- 用二项式展开 $(c^2-z^2)^n$ 来求 $\mathrm{E}(Z^2)$。这样做是可行的，但所得的和很难化简。

STEP 之外

中位数是样本的*顺序统计量*（order statistics）之一。同样的计数论证可以给出任何连续分布的 m 个数值中第 r 小者的密度，即 $\frac{m!}{(r-1)!\,(m-r)!}F^{r-1}(1-F)^{m-r}f$。对于均匀分布，这些都是*贝塔分布*（beta distributions），而你所推导出的积分正是贝塔函数 $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ 的一个特例。方差 $c^2/(2n+3)$ 以 $1/n$ 的速度缩小，这是迈向样本中位数的渐近正态性（asymptotic normality）的第一步。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）对顺序统计量和贝塔分布有很好的讲解。

详细解答

STEP III 1989，第 16 题

概率与统计 · 概率与假设检验

> **STEP III 1989，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/89-s3-q16

关键思路

三个“是／否”特征把人口分成八格，而那八项陈述不过是关于各格比例的线性方程。先填绿眼的各格（那里的条件最具体），再填淡褐眼的各格。之后的统计部分全靠一个数，即一个鲁里坦尼亚人是窄头的概率，并问在这个数值之下，观察所得是否令人意外。

解答

试卷的印刷本有三处笔误。我们把 (iv) 中的“arrow-headed”读作 *narrow-headed*（窄头），把 (vi) 中的“fair-headed”读作 *fair-haired*（金发），把 (viii) 中的“hazel-haired”读作 *hazel-eyed*（淡褐眼）。这是唯一合理的读法，因为三个特征分别是头发（金或深色）、眼睛（绿或淡褐）和头形（窄或阔）。

窄头者所占的比例

随机选出一个鲁里坦尼亚人。以 F、D 表示金发和深色头发，以 G、H 表示绿眼和淡褐眼，以 N、B 表示窄头和阔头。已知 $\mathrm{P}(F) = \frac{1}{4}$ 及 $\mathrm{P}(G) = \frac{1}{5}$。

绿眼的各格。由 (v)，$\mathrm{P}(N \mid G) = \mathrm{P}(B \mid G) = \frac{1}{2}$，所以

$$\mathrm{P}(G \cap N) = \mathrm{P}(G \cap B) = \frac{1}{10}.$$

由 (vi)，绿眼的阔头者平均分为金发和深色头发两半：

$$\mathrm{P}(F \cap G \cap B) = \mathrm{P}(D \cap G \cap B) = \frac{1}{20}.$$

由 (iv)，$\mathrm{P}(F \cap G \cap N) = 0$，所以 $\mathrm{P}(D \cap G \cap N) = \frac{1}{10}$。

淡褐眼的各格。设

$$a=\mathrm{P}(F\cap H\cap B),\quad b=\mathrm{P}(F\cap H\cap N),\quad c=\mathrm{P}(D\cap H\cap B),\quad d=\mathrm{P}(D\cap H\cap N).$$

- 金发：绿眼的金发者占 $0+\frac{1}{20}$，所以 $a+b=\frac{1}{4}-\frac{1}{20}=\frac{1}{5}$。
- 淡褐眼：$a+b+c+d=\frac{4}{5}$，所以 $c+d=\frac{3}{5}$。
- 陈述 (vii)：$\mathrm{P}(B\cap D)=\frac{1}{20}+c=\frac{1}{2}$，所以 $c=\frac{9}{20}$，因此 $d=\frac{3}{20}$。
- 陈述 (viii)：$\mathrm{P}(F\cap B\mid H)=\mathrm{P}(D\cap N\mid H)$；两边乘以 $\mathrm{P}(H)$ 得 $a=d=\frac{3}{20}$，于是 $b=\frac{1}{20}$。

全部八个比例，以二十分之一为单位，列出如下：

	$G\cap N$	$G\cap B$	$H\cap N$	$H\cap B$
F	0	1	1	3
D	2	1	3	9

它们全部非负，而且加起来是 20，所以这个模型并无矛盾。把窄头的两栏相加，

$$\mathrm{P}(N)=\frac{0+2+1+3}{20}=\frac{3}{10}.$$

所以按照这个模型，有 30% 的鲁里坦尼亚人是窄头的。

六个阔头的相识

假如我这六个相识是一个随机样本（random sample），而且彼此独立（independent），那么六人全是阔头的机会是

$$\left(\tfrac{7}{10}\right)^6=0.1176\ldots\approx 0.12.$$

想一想 一件概率为 12% 的事件，算得上是否定模型的有力证据吗?

算不上。一件大约每 8 次就有 1 次发生的事并不罕见；在 5% 这类惯用的显著性水平（significance level）下，它都不显著。这些数据还有另一个弱点：一个人的相识并不是随机样本。他们可能来自同一家庭、同一地区或同一社交圈子，所以未必彼此独立，也未必能代表整个人口。这项观察顶多只提供极轻微的怀疑理由。

一个 200 人的随机样本

设 X 为一个 200 人的随机样本中窄头者的数目。人口很大，所以各次抽取实际上是独立的，而在这个模型之下，$X\sim\mathrm{B}(200,\frac{3}{10})$，且

$$\mathrm{E}(X)=60,\qquad \mathrm{Var}(X)=200\times 0.3\times 0.7=42,\qquad \sqrt{42}\approx 6.481.$$

由于 np 和 $n(1-p)$ 都很大，我们采用正态近似（normal approximation）$X \approx \mathrm{N}(60, 42)$，并加上连续性校正（continuity correction）：

$$\mathrm{P}(X \leq 50) \approx \Phi\left(\frac{50.5-60}{\sqrt{42}}\right) = \Phi(-1.466) \approx 0.071.$$

所以在这个模型之下，窄头者不多于 50 人的情况大约有 7% 的机会出现。

想一想 我们现在应该否定这个模型吗？

在 5% 的水平下不应该：以 $p = 0.3$ 对 $p < 0.3$ 作单尾检验（one-tailed test），所得的 p 值约为 0.07。不过，这个结果确实令人对模型有些怀疑。样本比例 $\frac{50}{200} = 0.25$ 比预测值低了约 1.5 个标准差（standard deviation）。它的方向也与我那些相识的情况一致，他们当中的窄头者同样比预期少。在相信模型所给的 30% 这个数字之前，多收集一些数据才是明智的做法。（假如检验是双尾的，p 值便约为 0.14，证据就更弱了。）

验算

- 八个条件每一个都可以直接从表中读出，例如 (vii)：$1 + 9 = 10$ 个二十分之一 $= \frac{1}{2}$。
- 用电脑核对，可确认各格的比例是唯一解，$\mathrm{P}(N) = \frac{3}{10}$，以及 $(0.7)^6 = 0.117649$。精确的二项概率 $\mathrm{P}(X \leq 50) = 0.0695$ 与正态近似值 0.0713 相当接近。

常见错误

- 把 (v) 或 (vi) 读成关于整个人口的陈述，而不是条件（conditional）陈述：在绿眼的一组之内“同样可能”（as likely as），意思是把 $\frac{1}{5}$ 平分成两半。
- 忘记连续性校正。用 50 而不用 50.5 会得到 $\Phi(-1.543) \approx 0.061$。
- 把相识当作随机样本而不加任何说明。

STEP 之外

最后一部分是一个假设检验（hypothesis test），而那两项证据引出一个很自然的问题：几项薄弱的证据应该怎样合并起来？大学的统计学以似然（likelihood）来回答这个问题（对贝叶斯学派（Bayesian）而言，还有后验概率（posterior probability）），它们衡量 p 的每一个值对全部数据合起来解释得有多好。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）提供所需的概率背景，而 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）则对贝叶斯观点有一段浅白易读的介绍。

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

1989 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。

