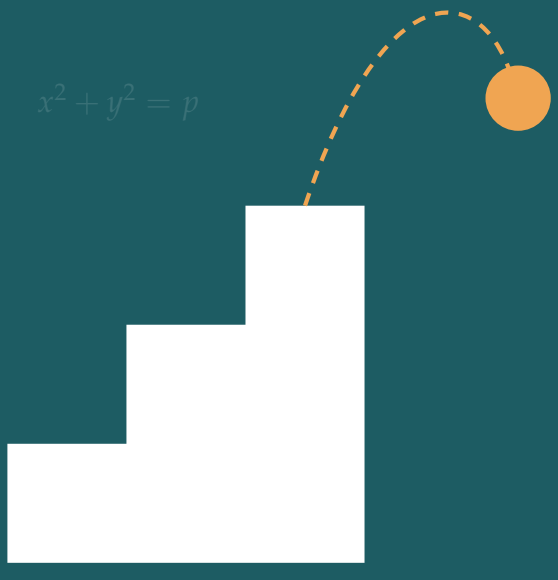
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1987 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

1987 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1987 年详细解答

1987 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 曲线描绘 纯数学 . . . 3

第 2 题 圆填充的面积 纯数学 . . . 7

第 3 题 可分离变量微分方程 纯数学 . . . 11

第 4 题 对数与等比级数 纯数学 . . . 14

第 5 题 代换积分法 纯数学 . . . 16

第 6 题 面积与反函数 纯数学 . . . 20

第 7 题 三角级数 纯数学 . . . 24

第 8 题 代换积分法 纯数学 . . . 27

第 9 题 重心坐标 纯数学 . . . 30

第 10 题 弹性绳 力学 . . . 34

第 11 题 圆周运动与能量 力学 . . . 38

第 12 题 抛体与恢复系数 力学 . . . 41

第 13 题 受力 力学 . . . 44

第 14 题 概率的等比级数 概率与统计 . . . 47

第 15 题 连续随机变量 概率与统计 . . . 51

第 16 题 贝叶斯定理与正态近似 概率与统计 . . . 54

STEP II

第 1 题 代数恒等式 纯数学 . . . 57

第 2 题 三角恒等式 纯数学 . . . 60

第 3 题 共轭根式 纯数学 . . . 63

第 4 题 复数与几何 纯数学 . . . 66

第 5 题 拉格朗日反演级数 纯数学 . . . 70

第 6 题 利用对称性求定积分 纯数学 . . . 74

第 7 题 梯形法则与阶乘 纯数学 78
第 8 题 三维空间中的向量轨迹 纯数学 82
第 9 题 正交矩阵 纯数学 86
第 10 题 群公理 纯数学 89
第 11 题 摩擦力与力矩 力学 92
第 12 题 平衡与能量 力学 95
第 13 题 墙壁的斜碰撞 力学 99
第 14 题 圆锥上的弹性带 力学 104
第 15 题 贝叶斯概率 概率与统计 108
第 16 题 贝叶斯定理 概率与统计 112

STEP III

第 1 题 阶乘的整除性 纯数学 115
第 2 题 三角函数归约公式 纯数学 118
第 3 题 圆角与梅钦公式 纯数学 121
第 4 题 旋转体的体积 纯数学 125
第 5 题 模算术密码 纯数学 128
第 6 题 耦合微分方程 纯数学 131
第 7 题 反正切的级数 纯数学 135
第 8 题 平行四边形面积不等式 纯数学 138
第 9 题 群的直积 纯数学 142
第 10 题 伯努利多项式 纯数学 146
第 11 题 相对速度的最优化 力学 149
第 12 题 变质量转动 力学 153
第 13 题 涉及摩擦力的转动杆 力学 157
第 14 题 张力坍缩 力学 160
第 15 题 期望成本的最优化 概率与统计 164
第 16 题 顺序统计量与泊松分布 概率与统计 167

详细解答

STEP I 1987，第 1 题

纯数学 · 曲线描绘

> **STEP I 1987，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/87-s1-q1

关键思路

用乘积法则（product rule）求导后，会剩下一个永不为零的因子 e^{ax}，乘以 $\cos bx$ 与 $\sin bx$ 的某个组合。因此驻点（stationary points）全部来自同一条方程 $\tan bx =$ 常数，而这条方程的解是等距分布的，相邻两解相距 π/b。从一个驻点移到下一个，指数部分乘以一个固定的因子，而余弦部分则变号，这正正就是一个等比数列（geometric progression）。

解答

驻点

由乘积法则，

$$f'(x) = ae^{ax}\cos bx - be^{ax}\sin bx = e^{ax}(a\cos bx - b\sin bx).$$

由于对所有 x 都有 $e^{ax} > 0$，所以 $f'(x) = 0$ 当且仅当 $a\cos bx = b\sin bx$。

> ? **想一想** 这里可以放心除以 $\cos bx$ 吗?

可以。若 $\cos bx = 0$，则 $\sin bx = \pm 1$，方程便变成 $0 = \pm b$，但因为 $b > 0$，这是不成立的。所以在任何驻点上都有 $\cos bx \neq 0$，两边相除得

$$\tan bx = \frac{a}{b}.$$

设 $\alpha = \arctan(a/b)$。因为 $a/b > 0$，所以 $0 < \alpha < \frac{\pi}{2}$。由于 tan 的周期（period）是 π，方程的解为 $bx = \alpha + n\pi$，因此驻点位于

$$x_n = \frac{\alpha + n\pi}{b}, \qquad n \in \mathbb{Z}, \qquad \text{其中 } \tan\alpha = \frac{a}{b}.$$

考虑两条直角边分别为 a 和 b 的直角三角形，可得 $\cos\alpha = \dfrac{b}{\sqrt{a^2+b^2}}$。由于 $\cos(\theta+n\pi) = (-1)^n\cos\theta$，f 在该处的值为

$$\begin{aligned}\mathrm{f}(x_n) &= \mathrm{e}^{a(\alpha+n\pi)/b}\cos(\alpha+n\pi)\\ &= (-1)^n\frac{b}{\sqrt{a^2+b^2}}\,\mathrm{e}^{a\alpha/b}\,\mathrm{e}^{na\pi/b}.\end{aligned}$$

极大还是极小？ 再求导一次，

$$\mathrm{f}''(x) = \mathrm{e}^{ax}((a^2-b^2)\cos bx - 2ab\sin bx).$$

在驻点上 $b\sin bx = a\cos bx$，所以 $2ab\sin bx = 2a^2\cos bx$，于是

$$\mathrm{f}''(x_n) = \mathrm{e}^{ax_n}\cos bx_n\,(a^2-b^2-2a^2) = -(a^2+b^2)\,\mathrm{f}(x_n).$$

因此当 $\mathrm{f}(x_n) > 0$（n 为偶数）时，x_n 是极大点（maximum）；当 $\mathrm{f}(x_n) < 0$（n 为奇数）时，x_n 是极小点（minimum）。驻点交替出现：极大、极小、极大、...

等比数列

由于 $0 < \alpha < \pi$，$x_n > 0$ 当且仅当 $n \geq 0$。所以 $x > 0$ 的驻点是 $x_0 < x_1 < x_2 < \cdots$，而对每个 $n \geq 0$，

$$\frac{\mathrm{f}(x_{n+1})}{\mathrm{f}(x_n)} = \frac{\mathrm{e}^{a(\alpha+(n+1)\pi)/b}\cos(\alpha+(n+1)\pi)}{\mathrm{e}^{a(\alpha+n\pi)/b}\cos(\alpha+n\pi)} = \mathrm{e}^{a\pi/b}\cdot(-1) = -\mathrm{e}^{a\pi/b},$$

这里用到 $\cos(\theta+\pi) = -\cos\theta$ 及 $\cos(\alpha+n\pi) \neq 0$。这个比值与 n 无关，所以 $\mathrm{f}(x_0), \mathrm{f}(x_1), \ldots$ 组成一个等比数列，首项为 $\dfrac{b\,\mathrm{e}^{a\alpha/b}}{\sqrt{a^2+b^2}}$，公比（common ratio）为 $-\mathrm{e}^{a\pi/b}$，得证。

（同样的计算对负数 n 亦成立，所以事实上所有驻值组成一个向两边无限延伸的等比数列。限制 $x > 0$ 只是给它一个自然的首项。）

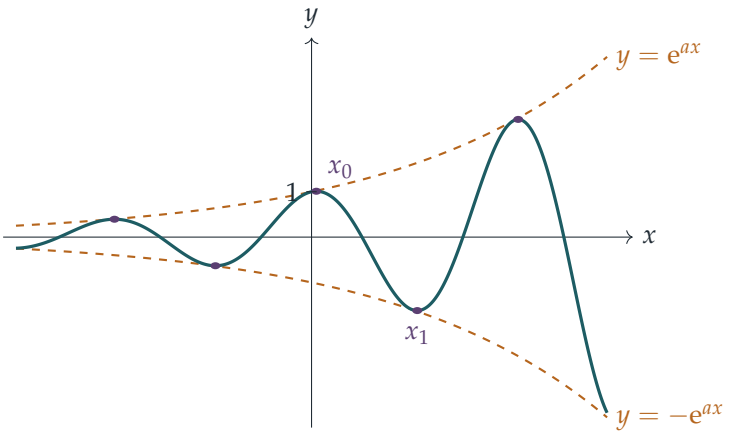
草图

所需的事实：

- 由于 $-1 \leq \cos bx \leq 1$，图像夹在曲线 $y = \mathrm{e}^{ax}$ 与 $y = -\mathrm{e}^{ax}$ 之间。在 $\cos bx = 1$（$bx = 2m\pi$）处，图像与 $y = \mathrm{e}^{ax}$ 相接；在 $\cos bx = -1$（$bx = (2m+1)\pi$）处，则与 $y = -\mathrm{e}^{ax}$ 相接。在这些点上 $\sin bx = 0$，所以 $\mathrm{f}'(x) = \pm a\mathrm{e}^{ax}$，与包络线（envelope）的斜率相同：曲线在该处与包络线相切，而不是转向。
- 零点位于 $bx = \frac{\pi}{2} + m\pi$，相邻两个相距 π/b，而且 $\mathrm{f}(0) = 1$。

- 每个转向点（turning point）$bx = \alpha + n\pi$ 都位于对应的相接点 $bx = n\pi$ 稍后的位置，因为 $0 < \alpha < \frac{\pi}{2}$：不断上升的包络线把每个峰顶稍稍拉向右方。
- 当 $x \to \infty$ 时，振荡幅度无界地增大；当 $x \to -\infty$ 时，$e^{ax} \to 0$，所以 $f(x) \to 0$，但仍然不断振荡。

草图所示为 $a = 0.3$、$b = 2$ 的情况，包络线 $y = \pm e^{ax}$ 以虚线表示。



验算

- 取 $a = b = 1$：此时 $\alpha = \frac{\pi}{4}$，$x_0 = \frac{\pi}{4}$，$f(x_0) = e^{\pi/4}/\sqrt{2} \approx 1.551$，而 $f(x_1) = e^{5\pi/4}\cos\frac{5\pi}{4} \approx -35.89$。比值约为 $\approx -23.14 = -e^{\pi}$。
- 当 $a \to 0$ 时，函数变成 $\cos bx$：$\alpha \to 0$，转向点移到 $bx = n\pi$，比值趋向 -1，得出大家熟悉的 $1, -1, 1, \ldots$
- 我们以数值方法对 500 组随机的 (a, b) 作验证，确认了 $f'(x_n) = 0$、$f(x_n)$ 与上述公式相符、$f''(x_n) = -(a^2 + b^2)f(x_n)$、相邻比值等于 $-e^{a\pi/b}$，以及 f' 没有其他零点。

常见错误

- 把转向点放在曲线与包络线相接之处，即 $bx = n\pi$。这些点的斜率是 $\pm ae^{ax} \neq 0$；真正的转向点稍稍偏右。
- 只取 $bx = \alpha$，忘记了 $\tan bx = a/b$ 的通解（general solution）$bx = \alpha + n\pi$。
- 漏掉比值中的负号：余弦在相邻两个驻点之间变号，所以驻值交替为极大、极小。

STEP 之外

函数 $e^{ax}\cos bx$ 是 $e^{(a+ib)x}$ 的实部(real part),而这类函数正正就是那些辅助方程(auxiliary equation）有复数根的线性微分方程（linear differential equation）$y''+py'+qy=0$ 的解。当 $a<0$ 时，得到的是阻尼振动（damped oscillation），例如有摩擦力的摆；当 $a>0$ 时（如本题），则是一个振荡不断增大的不稳定系统。相邻峰值之比是常数，而这个比值的对数就是工程师所说的对数衰减率（logarithmic decrement）。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）在二阶线性方程一章中详细发展了这个题材；Needham 的 *Visual Complex Analysis*（《直观复分析》）则说明了为甚么乘以 $e^{(a+ib)x}$ 会描出一条螺线。

详细解答

STEP I 1987，第 2 题

纯数学 · 圆填充的面积

> **STEP I 1987，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/87-s1-q2

关键思路

把各个小圆的圆心连起来。这些圆心全都位于半径为 R 的圆上，即恰好在两个大圆的正中间；它们构成一个正 n 边形（regular polygon），而这个多边形的各边都穿过相邻小圆的切点。这个多边形正是划分 Y（在多边形内）和 X（在多边形外）两部分的分界线。因此，每一块面积都可以由三种简单图形得出：多边形、圆盘，以及由多边形顶点截出的小圆扇形（sector）。

解答

n、R 与 r 的关系

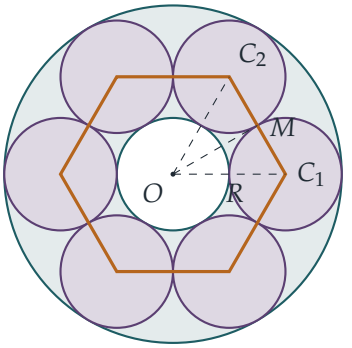
每个小圆都与两个大圆相切，所以它的圆心与公共圆心 O 的距离是 $R - r + r = R$：全部 n 个圆心都在半径为 R 的圆上。由对称性可知，这些圆心等距分布，所以相邻的圆心 C_1、C_2 在 O 处所张的角是 $2\pi/n$。

两个相邻的圆互相相切，所以 $C_1C_2 = 2r$。在等腰三角形（isosceles triangle）OC_1C_2 中，$OC_1 = OC_2 = R$；从 O 向 C_1C_2 的中点 M 作垂线，可得

$$r = C_1M = R\sin\frac{\pi}{n}.$$

这就是所求的关系。（我们需要 $n \geq 3$；若 $n = 2$，内圆的半径会是 $R - r = 0$。）同一个三角形亦给出

$$OM = R\cos\frac{\pi}{n} = \sqrt{R^2 - r^2}.$$



上图（以 $n = 6$ 绘画）用琥珀色显示由圆心构成的多边形。它的各边都穿过相邻圆的切点，例如 M。

面积 Y

这个多边形由 n 个与 OC_1C_2 全等（congruent）的三角形组成，每个的面积是 $\frac{1}{2} \cdot 2r \cdot \sqrt{R^2 - r^2}$。所以

$$\text{多边形的面积} = nr\sqrt{R^2 - r^2}.$$

多边形内有甚么？它的内切圆半径（inradius）是 $OM = \sqrt{R^2 - r^2}$，而 $R^2 - r^2 \geq (R - r)^2$，因为这个不等式（inequality）移项整理后就是 $2r(R - r) \geq 0$。所以半径为 $R - r$ 的内圆盘位于多边形之内。每个顶点处都有一个小圆的扇形。正 n 边形的内角（interior angle）是 $\pi - \frac{2\pi}{n}$，所以每个这样的扇形面积是

$$\tfrac{1}{2}r^2\left(\pi - \frac{2\pi}{n}\right) = \pi r^2\left(\frac{1}{2} - \frac{1}{n}\right).$$

这 n 个扇形和内圆盘彼此不重叠，因为小圆之间以及小圆与内圆之间都只是相切。多边形余下的部分，恰好就是紧贴内圆的 n 个曲边三角形，而它们合起来就是 Y：

$$Y = nr\sqrt{R^2 - r^2} - \pi(R - r)^2 - n\pi r^2\left(\frac{1}{2} - \frac{1}{n}\right),$$

正如所求。

? **想一想** 为甚么边 C_1C_2 正是把 X 与 Y 分开的恰当位置？

边 C_1C_2 由圆心连到圆心，并穿过切点 M，所以它永远不会穿过圆与圆之间的空隙。所有紧贴内圆的空隙都在它靠 O 的一侧，而所有紧贴外圆的空隙都在另一侧。因此沿着多边形切开，便能干净地把两类空隙分开；而且它沿两条半径切开每个小圆，这就是我们得到扇形的原因。

面积 X 与 Z

在多边形之外而外圆之内的区域，我们要除去每个小圆余下的部分。余下的部分是一个优角扇形（reflex sector），角度为 $2\pi-(\pi-\frac{2\pi}{n})=\pi+\frac{2\pi}{n}$，面积为 $\pi r^2\left(\frac{1}{2}+\frac{1}{n}\right)$。因此

$$X=\pi(R+r)^2-nr\sqrt{R^2-r^2}-n\pi r^2\left(\frac{1}{2}+\frac{1}{n}\right).$$

n 个半径为 r 的圆盘给出

$$Z=n\pi r^2.$$

当 $n\to\infty$ 时的极限

两式相加，多边形的项互相抵消，而扇形的项合并为 $n\pi r^2$：

$$\begin{aligned}X+Y&=\pi(R+r)^2-\pi(R-r)^2-n\pi r^2\\&=4\pi Rr-n\pi r^2.\end{aligned}$$

这其实就是圆环（annulus）的面积减去各小圆的面积，理应如此。因此

$$\frac{X+Y}{Z}=\frac{4\pi Rr-n\pi r^2}{n\pi r^2}=\frac{4R}{nr}-1=\frac{4}{n\sin(\pi/n)}-1,$$

这里用了 $r=R\sin(\pi/n)$。当 $n\to\infty$ 时，$\pi/n\to 0$ 且 $\dfrac{\sin(\pi/n)}{\pi/n}\to 1$，所以 $n\sin(\pi/n)\to\pi$，从而

$$\frac{X+Y}{Z}\to\frac{4}{\pi}-1\approx 0.273.$$

验算

- 取 $n=3$、$R=1$，得 $r=\frac{\sqrt{3}}{2}$、$Y\approx 0.0646$ 及 $X\approx 3.7497$。在精细的极坐标网格上进行数值积分，把每一点分类为在某个圆内、在多边形内或在多边形外，结果在 $n=3,5,8,12$ 时均与公式吻合至四位有效数字。
- 比值 $\frac{4}{n\sin(\pi/n)}-1$ 在 $n=10$ 时为 0.2944，在 $n=100$ 时为 0.27345，在 $n=1000$ 时为 0.273240。它正逐渐逼近 $4/\pi-1=0.273240\ldots$。
- 直观地（heuristically）看，当 n 很大时，这些圆就像宽度为 $2r$ 的长条中一排直径为 $2r$ 的圆盘。每个 $2r\times 2r$ 的正方形中，圆盘占 πr^2，空隙占 $4r^2-\pi r^2$，所以比值是 $(4-\pi)/\pi$。

常见错误

- 计算扇形时误用多边形的外角（exterior angle）$2\pi/n$，而不是内角 $\pi-2\pi/n$。

- 计算 Y 时忘记除去内圆盘，或没有验证内圆盘确实位于多边形之内。

STEP 之外

最后的极限（limit）其实就是改头换面的 $\lim_{x\to 0}\frac{\sin x}{x}=1$。正是同一事实，让阿基米德（Archimedes）能够用内接和外切多边形的周长把 π 夹在中间。这道题亦是圆填充密度问题（circle-packing density problem）的一个简化版本，这里问的是：这些圆能填满一个狭窄圆环的多少比例？若想严格处理这类极限，可以阅读 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP I 1987，第 3 题

纯数学 · 可分离变量微分方程

STEP I 1987，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/87-s1-q3

关键思路

这条方程以 v 表示时看似棘手，但左边其实很容易辨认：$x^3v' + x^2v$ 正是 x^2 乘以 xv 的导数（derivative）。因此，题目建议的代换（substitution）把左边变成 x^2y'；而右边改用 y 表示后，变量可以干净利落地分离（separate）。之后只需运用部分分式（partial fractions），再小心选取正负号。

解答

变量代换

设 $y = xv$，由乘积法则（product rule）得 $y' = v + xv'$，因此

$$x^3\frac{\mathrm{d}v}{\mathrm{d}x} + x^2v = x^2\left(x\frac{\mathrm{d}v}{\mathrm{d}x} + v\right) = x^2\frac{\mathrm{d}y}{\mathrm{d}x}.$$

在右边，$x^2v^2 = y^2$ 而 $v = y/x$，所以右边为

$$\frac{1+y^2}{(1+x^2)\,y/x} = \frac{x(1+y^2)}{(1+x^2)\,y}.$$

方程变为

$$x^2\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x(1+y^2)}{(1+x^2)\,y}, \qquad \text{即} \qquad \frac{y}{1+y^2}\,\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{x(1+x^2)},$$

这是把两边除以 x^3（我们在 $x = 1$ 附近考虑，故 $x > 0$），再乘以 $y/(1+y^2)$ 所得。变量已经分离。

积分

左边积分得 $\frac{1}{2}\ln(1+y^2)$。至于右边，用部分分式可得

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

只要把右边通分便可轻易验证。因此，当 $x>0$ 时，

$$\tfrac{1}{2}\ln(1+y^2) = \ln x - \tfrac{1}{2}\ln(1+x^2) + C, \qquad \text{因此} \qquad 1+y^2 = \frac{Ax^2}{1+x^2}$$

其中 $A = \mathrm{e}^{2C}$ 为正常数。

当 $x=1$ 时 $v=1$，故 $y=1$，于是 $2 = A/2$，即 $A=4$。因此

$$y^2 = \frac{4x^2}{1+x^2} - 1 = \frac{3x^2-1}{1+x^2}, \qquad v^2 = \frac{y^2}{x^2} = \frac{3x^2-1}{x^2(1+x^2)}.$$

想一想 应该取哪一个平方根？解在哪个范围内成立？

原方程的分母含 v，所以解永远不能经过 $v=0$；由于解是连续的，它会保持在 $x=1$ 处的符号，即正号。因此取正平方根：

$$\boxed{v(x) = \frac{\sqrt{3x^2-1}}{x\sqrt{1+x^2}}}$$

此解在 $x > 1/\sqrt{3}$ 时成立。当 x 减小至 $1/\sqrt{3}$ 时，v 降至 0，方程便不再成立，所以解不能延伸到该点以下。

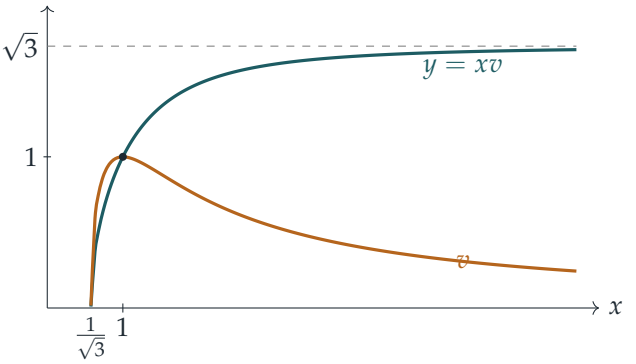
x 很大时

分子分母同除以 x^2，

$$v(x) = \frac{1}{x}\cdot\frac{\sqrt{3-1/x^2}}{\sqrt{1+1/x^2}} \approx \frac{\sqrt{3}}{x},$$

所以当 $x\to\infty$ 时 $v(x)\to 0$。等价地说，$y = xv \to \sqrt{3}$：y 的图像有水平渐近线（horizontal asymptote）$y=\sqrt{3}$，而 $v = y/x$ 则像 $\sqrt{3}/x$ 那样衰减。

事实上，起点正是最高点：设 $u = x^2$，则 $(3u-1)/(u^2+u)$ 的导数的分子为 $-(3u+1)(u-1)$，所以 v 上升至在 $x=1$ 处的最大值 1，然后递减并趋向 0。



验算

- 当 $x = 1$：$v = \sqrt{2}/(1 \cdot \sqrt{2}) = 1$，符合要求。
- 代回验证：由 $1 + y^2 = 4x^2/(1 + x^2)$ 作隐函数微分（implicit differentiation）得 $2yy' = 8x/(1 + x^2)^2$，而 $(1 + y^2)/(x(1 + x^2)) = 4x/(1 + x^2)^2$，所以 $yy'/(1 + y^2) = 1/(x(1 + x^2))$，与预期相符。
- 数值上，以龙格—库塔法（Runge–Kutta method）从 $(1, 1)$ 起逐步求解原方程，得 $v(10) \approx 0.172058$，与公式吻合；而在 $x = 10^4$ 时，公式给出 $xv \approx 1.73205 \approx \sqrt{3}$。

常见错误

- 把极限答成 $\sqrt{3}$：那是 y 的极限，但题目问的是 v，它趋向 0。
- 错误地处理常数的指数：$\ln(1 + y^2) = \ln \frac{x^2}{1+x^2} + 2C$ 应得出 $1 + y^2 = Ax^2/(1 + x^2)$，即一个相乘的常数，而不是 $x^2/(1 + x^2) + 2C$。
- 开平方时写上 $\pm$ 却没有作出取舍：条件 $v(1) = 1$ 及不容许 $v = 0$ 已确定要取正号。

STEP 之外

这里的代换之所以奏效，是因为左边在整体相乘后成为一个恰当导数（exact derivative），这与积分因子（integrating factor）和齐次方程（homogeneous equation）背后的想法如出一辙。处理正负号和定义域时所需的谨慎，让我们初步窥见存在性与唯一性理论（existence and uniqueness theory）：经过某一给定点的解会一直存在，直至遇到令方程失去意义的地方，在这里就是 $v = 0$。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）讨论了可分离方程（separable equation），以及解能延伸多远的问题。

详细解答

STEP I 1987，第 4 题

纯数学 · 对数与等比级数

> **STEP I 1987，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/87-s1-q4

关键思路

以不便处理的数为底的对数（logarithm），最好先化成自然对数（natural logarithm），因为这样底数就会以一个简单的因子出现。这里每个底数都是 2 的幂，而指数 2^n 会降下来成为乘数。把每一项都用 $\ln 2$ 表示之后，这个级数（series）其实就是一个乔装了的几何级数（geometric series）。

解答

改写通项

固定 $n \geq 0$，并设 $c = \log_{2^{2^n}} \mathrm{e}$。根据提示中的定义，这表示

$$\left(2^{2^n}\right)^c = \mathrm{e}.$$

两边取自然对数（两边均为正数），得

$$c \cdot 2^n \ln 2 = \ln \mathrm{e} = 1, \qquad \text{所以} \qquad \log_{2^{2^n}} \mathrm{e} = \frac{1}{2^n \ln 2}.$$

作为对下标的验证，$n = 0, 1, 2$ 分别给出底数 2、4、16，与首三项吻合，其值为 $\frac{1}{\ln 2}$、$\frac{1}{2\ln 2}$、$\frac{1}{4\ln 2}$。

> ? **想一想** 2^n 是从哪里来的?

来自法则 $\ln(a^k) = k \ln a$，取 $a = 2$ 及 $k = 2^n$：底数 2^{2^n} 的自然对数是 $2^n \ln 2$。正是指数的“塔”令每个底数都是前一个的*平方*，而底数平方会令对数减半。

求级数之和

因此级数为

$$\sum_{n=0}^{\infty}(-1)^n \log_{2^{2^n}} \mathrm{e} = \frac{1}{\ln 2}\sum_{n=0}^{\infty}\left(-\frac{1}{2}\right)^n.$$

这是首项为 1、公比为 $r=-\frac{1}{2}$ 的几何级数。由于 $|r|<1$，它收敛（converges），其和为 $\frac{1}{1-r}=\frac{1}{1+\frac{1}{2}}=\frac{2}{3}$。因此

$$\sum_{n=0}^{\infty}(-1)^n \log_{2^{2^n}} \mathrm{e} = \frac{2}{3\ln 2} = \frac{1}{\frac{3}{2}\ln 2} = \frac{1}{\ln\left(2^{3/2}\right)} = \frac{1}{\ln(2\sqrt{2})},$$

这里再次用了 $\ln(a^k)=k\ln a$，以及 $2^{3/2}=2\cdot 2^{1/2}=2\sqrt{2}$。这就是所求的和。

验算

数值上，$\frac{1}{\ln(2\sqrt{2})}\approx 0.96180$。级数的部分和（partial sums）为 1.4427、0.7213、1.0820、0.9017、0.9919、0.9468,…，在此值两侧来回摆动，而差距每次减半，正如交错几何级数应有的表现。取六十项时，结果与 0.961797 在机器精度内一致。

常见错误

- 把 $\log_{2^{2^n}}$ 误读为 $\log_{2^{2n}}$：这样会得出底数 2、16、256,… 及错误的公比 $-\frac{1}{4}$。应该用首几项来核对下标。
- 写成 $\log_a \mathrm{e}=\ln a$，而不是 $\frac{1}{\ln a}$。换底公式（change-of-base rule）给出 $\log_a b=\frac{\ln b}{\ln a}$，而这里 $\ln b=1$。

STEP 之外

一个无穷和究竟是否“有值”，正是分析学（analysis）入门课程的核心问题。几何级数是最基本的参照标准：许多收敛判别法，例如比值判别法（ratio test），都是通过把级数与几何级数比较来运作的。Abbott 的 *Understanding Analysis*（《理解分析》）细致而易读地建立了这套理论。

详细解答

STEP I 1987，第 5 题

纯数学 · 代换积分法

STEP I 1987，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/87-s1-q5

关键思路

这个代换（substitution）的设计，是要令根号下的两个因子都变成完全平方（perfect square）：$x-\alpha$ 变成 $\sin^2\theta$ 的某个倍数，而 $\beta-x$ 则变成 $\cos^2\theta$ 的同一倍数。于是平方根恰好与 $\mathrm{d}x/\mathrm{d}\theta$ 互相抵消，第一个积分的被积函数（integrand）便化为常数。第二个积分多保留了一个因子 $1/x$，只要把分子分母同除以 $\cos^2\theta$，它就变成一个标准的 $\tan\theta$ 积分。

解答

第一个积分，$\alpha<\beta$ 的情况

设 $x=\alpha\cos^2\theta+\beta\sin^2\theta$。利用 $\cos^2\theta+\sin^2\theta=1$，

$$x-\alpha=\alpha\cos^2\theta+\beta\sin^2\theta-\alpha(\cos^2\theta+\sin^2\theta)=(\beta-\alpha)\sin^2\theta,$$
$$\beta-x=\beta(\cos^2\theta+\sin^2\theta)-\alpha\cos^2\theta-\beta\sin^2\theta=(\beta-\alpha)\cos^2\theta,$$

以及

$$\frac{\mathrm{d}x}{\mathrm{d}\theta}=-2\alpha\cos\theta\sin\theta+2\beta\sin\theta\cos\theta=2(\beta-\alpha)\sin\theta\cos\theta.$$

至于积分上下限：$\theta=0$ 时 $x=\alpha$，$\theta=\frac{\pi}{2}$ 时 $x=\beta$。由于在 $(0,\frac{\pi}{2})$ 上 $\mathrm{d}x/\mathrm{d}\theta>0$，当 θ 走遍 $[0,\frac{\pi}{2}]$ 时，x 递增地走遍 $[\alpha,\beta]$ 恰好一次。

在这个范围内 $\sin\theta\geq0$ 及 $\cos\theta\geq0$，而且 $\beta-\alpha>0$，所以

$$\sqrt{(x-\alpha)(\beta-x)}=\sqrt{(\beta-\alpha)^2\sin^2\theta\cos^2\theta}=(\beta-\alpha)\sin\theta\cos\theta.$$

因此

$$\int_\alpha^\beta \frac{\mathrm{d}x}{\sqrt{(x-\alpha)(\beta-x)}} = \int_0^{\pi/2} \frac{2(\beta-\alpha)\sin\theta\cos\theta}{(\beta-\alpha)\sin\theta\cos\theta}\,\mathrm{d}\theta = \int_0^{\pi/2} 2\,\mathrm{d}\theta = \pi,$$

即为所求。（被积函数在两端都趋于无穷大，所以严格来说这是一个反常积分（improper integral）。代换显示它是收敛（converges）的，因为变换后的被积函数是有界的常数 2。）

想一想 我们究竟在哪里用到 $\alpha < \beta$?

只在开平方根的时候：$\sqrt{(\beta-\alpha)^2}$ 等于 $\beta-\alpha$，正正是因为 $\beta-\alpha>0$。一般而言它等于 $|\beta-\alpha|$，而这一点正是下一部分的关键。

第一个积分，$\alpha > \beta$ 的情况

现在积分由较大的数 α 往下积到较小的数 β。在两者之间的区间上，$x-\alpha<0$ 且 $\beta-x<0$，所以乘积仍然是正数，被积函数依然有意义。

代换可以逐字照搬，唯一不同的是现在 $\sqrt{(\beta-\alpha)^2} = \alpha-\beta$，所以

$$\frac{2(\beta-\alpha)\sin\theta\cos\theta}{(\alpha-\beta)\sin\theta\cos\theta} = -2,$$

而积分为

$$\int_0^{\pi/2} (-2)\,\mathrm{d}\theta = -\pi.$$

还有一个更快的看法。把 α 与 β 互换时，式子 $(x-\alpha)(\beta-x)$ 保持不变，所以被积函数与第一种情况是同一个函数，只是两个字母的角色对调了。把上下限倒转会改变正负号，所以

$$\int_\alpha^\beta = -\int_\beta^\alpha = -\pi,$$

这里用了第一个结果，以 β 作为较小的端点。

第二个积分，$0 < \alpha < \beta$ 的情况

用完全相同的代换，同样地抵消后，只剩下因子 $1/x$：

$$\int_\alpha^\beta \frac{\mathrm{d}x}{x\sqrt{(x-\alpha)(\beta-x)}} = \int_0^{\pi/2} \frac{2\,\mathrm{d}\theta}{\alpha\cos^2\theta+\beta\sin^2\theta}.$$

由于 $\alpha,\beta>0$，分母在整个 $[0,\frac{\pi}{2}]$ 上都是正数，所以这是一个正常积分。要计算它，把分子分母同除以 $\cos^2\theta$（对 $0\le\theta<\frac{\pi}{2}$ 成立）：

$$\int_0^{\pi/2} \frac{2\sec^2\theta\,\mathrm{d}\theta}{\alpha+\beta\tan^2\theta}.$$

再设 $t = \tan\theta$，则 $\mathrm{d}t = \sec^2\theta\,\mathrm{d}\theta$，而当 θ 由 0 增至 $\frac{\pi}{2}$ 时，t 由 0 走到 ∞：

$$\int_0^\infty \frac{2\,\mathrm{d}t}{\alpha+\beta t^2} = \frac{2}{\beta}\int_0^\infty \frac{\mathrm{d}t}{t^2+\alpha/\beta} = \frac{2}{\beta}\sqrt{\frac{\beta}{\alpha}}\left[\arctan\left(t\sqrt{\frac{\beta}{\alpha}}\right)\right]_0^\infty$$
$$= \frac{2}{\sqrt{\alpha\beta}}\cdot\frac{\pi}{2} = \frac{\pi}{\sqrt{\alpha\beta}},$$

即为所求。我们两次用到 $\alpha > 0$：一是确保 $1/x$ 在 $[\alpha,\beta]$ 内没有奇点（singularity），二是确保 $\sqrt{\beta/\alpha}$ 是实数。

> ? **想一想** 上限变成了 ∞。这会造成问题吗？

不会。关于 θ 的积分完全是普通的积分；无穷上限之所以出现，只是因为当 $\theta \to \frac{\pi}{2}$ 时 $\tan\theta \to \infty$。关于 t 的积分是收敛的，因为 arctan 趋于 $\frac{\pi}{2}$；而它等于关于 θ 的积分，因为两者都是在 $[0,\Theta]$ 上的积分当 $\Theta \to \frac{\pi}{2}$ 时的极限。

验算

- 缩放：把 x、α、β 换成 kx、$k\alpha$、$k\beta$，第一个积分不变，第二个积分则除以 k。两个答案 π 和 $\pi/\sqrt{\alpha\beta}$ 都符合这个性质。
- 令 $\alpha \to \beta$。这时在整个（极短的）区间上 $1/x \approx 1/\beta$，所以第二个积分应约为 π/β；而 $\pi/\sqrt{\alpha\beta} \to \pi/\beta$，果然如此。
- 数值上，取 $\alpha = 2$、$\beta = 7$，数值积分（quadrature）给出第二个积分为 0.83963，与 $\pi/\sqrt{14} = 0.83963$ 吻合。随机选取 $\alpha < \beta$，第一个积分全部得出 π；把端点互换则得出 $-\pi$。

常见错误

- 在 $\alpha > \beta$ 时仍然答 π。被积函数是正的，但上下限倒过来了，所以答案是 $-\pi$。
- 写下 $\sqrt{(\beta-\alpha)^2\sin^2\theta\cos^2\theta} = (\beta-\alpha)\sin\theta\cos\theta$，却没有检查三个因子的正负号。

STEP 之外

这个代换有一个漂亮的几何意义。由于 $x = \frac{\alpha+\beta}{2} - \frac{\beta-\alpha}{2}\cos 2\theta$，它把以 $[\alpha,\beta]$ 为直径的半圆参数化（parametrises），而 $\sqrt{(x-\alpha)(\beta-x)}$ 正是该半圆在 x 处的高度。第二个积分让你初尝以围道积分（contour integration）计算积分的滋味：在复分析（complex analysis）中，两个结果都可以用留数定理（residue theorem）在短短几行内得出。Needham 的

Visual Complex Analysis（《直观复分析》）对此有精彩的阐述，而反常积分的严谨处理则见于 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP I 1987，第 6 题

纯数学 · 面积与反函数

STEP I 1987，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/87-s1-q6

关键思路

$\mathrm{F}(s)$ 是曲线下方、介乎曲线与 x 轴之间的面积；$\mathrm{G}(t)$ 则是曲线左方、介乎曲线与 y 轴之间的面积。把这两个区域画在同一幅图上，它们合起来覆盖以 $(0,0)$ 和 (s,t) 为顶点的长方形，而且除非顶点 (s,t) 刚好落在曲线上，否则还会有多出来的部分。在第二部分，我们选取 s 使这种情况出现，于是较难的积分 $\int \sin^{-1} y \,\mathrm{d}y$ 便可以从容易的积分 $\int \sin x \,\mathrm{d}x$ 直接得出。

解答

不等式 $(*)$

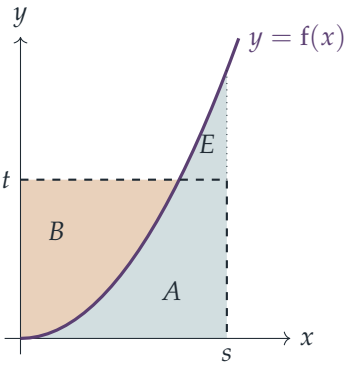
首先看看 f 的一个性质。它是连续（continuous）的，而且每个值最多只取一次，所以它是严格单调（strictly monotonic）的（根据介值定理（intermediate value theorem），一个先上升后下降的连续函数必定会把某个值取两次）。由于 $\mathrm{f}(0) = 0$ 且 $\mathrm{f}(x) \geq 0$，它必定是严格递增（strictly increasing）的。因此 g 亦是递增的，并且对于 $0 \leq x \leq a$、$0 \leq y \leq \mathrm{f}(a)$，

$$y \leq \mathrm{f}(x) \iff \mathrm{g}(y) \leq x.$$

定义两个区域：

$$A = \{(x,y) : 0 \leq x \leq s,\ 0 \leq y \leq \mathrm{f}(x)\}, \qquad \text{面积}\mathrm{F}(s),$$
$$B = \{(x,y) : 0 \leq y \leq t,\ 0 \leq x \leq \mathrm{g}(y)\}, \qquad \text{面积}\mathrm{G}(t),$$

并设 R 为长方形 $0 \leq x \leq s$、$0 \leq y \leq t$，其面积为 st。B 的面积是 $\mathrm{G}(t)$，因为我们是把长度为 $\mathrm{g}(y)$ 的水平窄条积分起来（“对 y 求面积”）。



R 中每一点都位于 A 或 B 之内：若 $(x,y)\in R$，则要么 $y\le \mathrm{f}(x)$，该点在 A 内；要么 $y>\mathrm{f}(x)$，即 $x<\mathrm{g}(y)$，该点在 B 内。两者的重叠部分 $A\cap B$ 只是曲线本身的一段弧，面积为零。因此

$$\mathrm{F}(s)+\mathrm{G}(t)=\text{面积}(A)+\text{面积}(B)\ge\text{面积}(A\cup B)\ge\text{面积}(R)=st,$$

这就是 $(*)$。

想一想 究竟多算了哪些不在长方形内的部分？

多出来的是 $A\cup B$ 中伸出 R 之外的部分 E。若 $t<\mathrm{f}(s)$（如图所示），B 留在 R 之内，但 A 在区域 $\mathrm{g}(t)<x\le s$、$t<y\le \mathrm{f}(x)$ 中伸出了 $y=t$ 之上。由于对每个 $x>\mathrm{g}(t)$ 都有 $\mathrm{f}(x)>t$，而且 f 连续，这个区域的面积为正，所以不等式严格成立。若 $t>\mathrm{f}(s)$，情况相同，只是角色互换：B 伸出了 $x=s$ 的右方。若 $t=\mathrm{f}(s)$，顶点 (s,t) 位于曲线上，$A\cup B$ 恰好就是 R，没有多出来的部分。

所以 $(*)$ **中等号成立当且仅当** $t=\mathrm{f}(s)$，即 (s,t) 位于曲线上的时候。

$\sin^{-1}y$ 的积分

在 $[0,\frac12\pi]$ 上取 $\mathrm{f}(x)=\sin x$。它连续、递增、通过原点，并且留在第一象限，所以上面所有结果都适用，其中 $a=\frac12\pi$、$\mathrm{f}(a)=1$ 及 $\mathrm{g}(y)=\sin^{-1}y$。

给定 $t\in[0,1]$，选取 $s=\sin^{-1}t$，它位于 $[0,\frac12\pi]$ 之内。于是 $t=\sin s=\mathrm{f}(s)$，所以 $(*)$ 中等号成立：

$$\int_0^t \sin^{-1}y\,\mathrm{d}y=\mathrm{G}(t)=st-\mathrm{F}(s).$$

接着计算容易的积分：

$$\mathrm{F}(s)=\int_0^s \sin x\,\mathrm{d}x=1-\cos s.$$

代入 $s=\sin^{-1}t$，

$$\int_0^t \sin^{-1}y\,\mathrm{d}y=t\sin^{-1}t-\bigl(1-\cos(\sin^{-1}t)\bigr),$$

正是所求。

以微分验证

由于 $\sin^{-1} t \in [0, \frac{1}{2}\pi]$，它的余弦非负，所以 $\cos(\sin^{-1} t) = \sqrt{1-t^2}$。因此右方为 $t\sin^{-1} t - 1 + \sqrt{1-t^2}$。对于 $0 \le t < 1$，利用 $\dfrac{\mathrm{d}}{\mathrm{d}t}\sin^{-1} t = \dfrac{1}{\sqrt{1-t^2}}$ 及乘积法则（product rule），

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(t\sin^{-1} t - 1 + \sqrt{1-t^2}\right) &= \sin^{-1} t + \frac{t}{\sqrt{1-t^2}} - \frac{t}{\sqrt{1-t^2}} \\ &= \sin^{-1} t,\end{aligned}$$

根据微积分基本定理（fundamental theorem of calculus），这正是左方的导数。两方在 $t = 0$ 时都等于 0，所以它们在 $[0, 1)$ 上相等，再由连续性，在 $t = 1$ 时亦相等。

验算

- 当 $t = 1$：公式给出 $\frac{1}{2}\pi - 1 \approx 0.5708$。从几何角度看，$y = \sin x$ 左方直至高度 1 的面积，等于长方形 $\frac{1}{2}\pi \times 1$ 减去拱形下方的面积 1。对 $\sin^{-1} y$ 在 $[0, 1]$ 上作数值积分，得 0.57080。
- 当 $t = \frac{1}{2}$：公式给出 $\frac{\pi}{12} - 1 + \frac{\sqrt{3}}{2} \approx 0.12782$，与辛普森法则（Simpson's rule）的结果吻合。
- 对于 $(*)$，试取 $\mathrm{f}(x) = x^2$、$s = 1$、$t = 4$：$\mathrm{F} + \mathrm{G} = \frac{1}{3} + \frac{16}{3} = \frac{17}{3} > 4$；而取 $t = 1 = \mathrm{f}(1)$ 时，则得 $\frac{1}{3} + \frac{2}{3} = 1 = st$。

常见错误

- 写成 $\cos(\sin^{-1} t) = \pm\sqrt{1-t^2}$ 而没有选定符号：$\sin^{-1}$ 的值域决定了它取正值。
- 只说“这些区域覆盖了长方形”，却没有解释为何当 $t \ne \mathrm{f}(s)$ 时不等式严格成立；等号的问题正需要这一点。
- 忘记 f 必须是递增的：整幅图都依赖这一点。

STEP 之外

不等式 $(*)$ 就是杨氏不等式（Young's inequality）。取 $\mathrm{f}(x) = x^{p-1}$，它便成为 $st \le \frac{s^p}{p} + \frac{t^q}{q}$，其中 $\frac{1}{p} + \frac{1}{q} = 1$；这是证明赫尔德不等式（Hölder's inequality）的关键步骤，而赫尔德不等式是分析学中最常用的工具之一。F 与 G 的这种配对，让我们初窥勒让德变换（Legendre

transform），它把拉格朗日力学（Lagrangian mechanics）与哈密顿力学（Hamiltonian mechanics）联系起来。如想了解这一切背后的分析学，可以阅读 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）。

详细解答

STEP I 1987，第 7 题

纯数学 · 三角级数

STEP I 1987，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/87-s1-q7

关键思路

两个级数（series）都是正弦之和，而且角度以相等的步长递增。把 $\sin\big((2k+1)\theta\big)$ 这类项乘以 $2\sin\theta$ 或 $2\cos\theta$，再利用积化和差公式（product-to-sum formula），便可把它化为相邻两项的*差*或*和*，于是整个级数便会裂项相消（telescopes）。唯一要决定的是用哪个因子：$2\sin\theta$ 适用于普通的和，$2\cos\theta$ 则适用于交错（alternating）的和。

解答

准备工作

设 $\theta = \dfrac{2\pi}{23}$。各角为 $2\pi/23,\ 6\pi/23,\ 10\pi/23,\ldots,42\pi/23$，每次增加 $4\pi/23 = 2\theta$，所以它们是

$$(2k+1)\theta \qquad \text{其中} k = 0, 1, \ldots, 10,$$

因为 $21\theta = 42\pi/23$。这样共有 11 项，而在交错级数中，第 k 项的符号是 $(-1)^k$（最后一项 $k=10$ 的符号是 $+$，正合题意）。因此我们要求

$$S_1 = \sum_{k=0}^{10} \sin\big((2k+1)\theta\big), \qquad S_2 = \sum_{k=0}^{10} (-1)^k \sin\big((2k+1)\theta\big).$$

我们会用到 $22\theta = \dfrac{44\pi}{23} = 2\pi - \dfrac{2\pi}{23}$，因此

$$\cos 22\theta = \cos\theta, \qquad \sin 22\theta = -\sin\theta.$$

第一个级数

由积化和差公式 $2\sin A\sin B=\cos(A-B)-\cos(A+B)$，

$$2\sin\theta\,\sin\big((2k+1)\theta\big)=\cos(2k\theta)-\cos\big((2k+2)\theta\big).$$

对 $k=0,\dots,10$ 求和，每个 $1\le k\le 10$ 的 $\cos(2k\theta)$ 都以正负号各出现一次而互相抵消，剩下

$$2\sin\theta\,S_1=\cos 0-\cos 22\theta=1-\cos\theta.$$

由于 $0<\theta<\pi$，$\sin\theta\neq 0$，所以可以相除：

$$S_1=\frac{1-\cos\theta}{2\sin\theta}=\frac{2\sin^2(\theta/2)}{4\sin(\theta/2)\cos(\theta/2)}=\tfrac{1}{2}\tan\frac{\theta}{2},$$

这里用了 $1-\cos\theta=2\sin^2(\theta/2)$ 和 $\sin\theta=2\sin(\theta/2)\cos(\theta/2)$，并约去 $\sin(\theta/2)\neq 0$。由于 $\theta/2=\pi/23$，

$$\boxed{\sin\tfrac{2\pi}{23}+\sin\tfrac{6\pi}{23}+\cdots+\sin\tfrac{42\pi}{23}=\tfrac{1}{2}\tan\tfrac{\pi}{23}.}$$

想一想 为甚么处理交错级数时要乘以 $2\cos\theta$，而不是再乘以 $2\sin\theta$?

若乘以 $2\sin\theta$，得到的是差 $\cos(2k\theta)-\cos((2k+2)\theta)$，而交错的符号会令相邻的项相加而非抵消。若乘以 $2\cos\theta$，得到的是和 $\sin((2k+2)\theta)+\sin(2k\theta)$，这时正是交错的符号带来抵消。

第二个级数

由 $2\cos A\sin B=\sin(B+A)+\sin(B-A)$，

$$2\cos\theta\,\sin\big((2k+1)\theta\big)=\sin\big((2k+2)\theta\big)+\sin(2k\theta).$$

乘以 $(-1)^k$ 再求和。对于 $1\le k\le 10$，$\sin(2k\theta)$ 一次来自第 $k-1$ 项，符号为 $(-1)^{k-1}$，另一次来自第 k 项，符号为 $(-1)^k$，所以互相抵消。剩下的是来自 $k=0$ 的 $\sin 0$ 和来自 $k=10$ 的 $(-1)^{10}\sin 22\theta$：

$$2\cos\theta\,S_2=\sin 22\theta=-\sin\theta.$$

由于 $0<\theta<\pi/2$，$\cos\theta\neq 0$，所以

$$S_2=-\frac{\sin\theta}{2\cos\theta}=-\tfrac{1}{2}\tan\theta,$$

即

$$\boxed{\sin\tfrac{2\pi}{23}-\sin\tfrac{6\pi}{23}+\cdots-\sin\tfrac{38\pi}{23}+\sin\tfrac{42\pi}{23}=-\tfrac{1}{2}\tan\tfrac{2\pi}{23}.}$$

如果你喜欢看起来是正的形式，由于 $\tan(\pi-x)=-\tan x$，这也等于 $\frac{1}{2}\tan\frac{21\pi}{23}$。

验算

- 以数值计算，$\frac{1}{2}\tan(\pi/23) \approx 0.068723$，而 $-\frac{1}{2}\tan(2\pi/23) \approx -0.140093$；直接把十一项相加，得到的值在 15 位有效数字（significant figures）内完全相同。
- 第二个答案是负数，这是合理的：部分和（partial sums）在约 ± 0.5 之间摆动，而最后加上的一项 $\sin(42\pi/23) \approx -0.27$ 本身就是负数。
- 同样的方法可得出一般公式 $\sum_{k=0}^{n-1} \sin((2k+1)\theta) = \dfrac{\sin^2 n\theta}{\sin\theta}$ 和 $\sum_{k=0}^{n-1} (-1)^k \sin((2k+1)\theta) = (-1)^{n-1}\dfrac{\sin 2n\theta}{2\cos\theta}$；试取 $n = 1$，两者都化为 $\sin\theta$。

常见错误

- 数错项数：角度每次增加 $4\pi/23$，而不是 $2\pi/23$，所以共有 11 项，而不是 21 项。
- 以为两个答案都是正数。大于 π 的角的正弦是负的，而负号正是来自 $\sin(44\pi/23) = -\sin(2\pi/23)$。

STEP 之外

这些和是几何级数（geometric series）的虚部（imaginary part）：$S_1 = \operatorname{Im}\sum \mathrm{e}^{\mathrm{i}(2k+1)\theta}$，而 S_2 也一样，只是公比为 $-\mathrm{e}^{2\mathrm{i}\theta}$，这为两个答案提供了第二条途径。这类和，例如狄利克雷核（Dirichlet kernel），是傅里叶分析（Fourier analysis）的起点；在傅里叶分析中，函数由正弦和余弦组合而成。Needham 的 *Visual Complex Analysis*（《直观复分析》）借助旋转向量的图像，说明这类和为何有这样的表现。

详细解答

STEP I 1987，第 8 题

纯数学 · 代换积分法

STEP I 1987，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/87-s1-q8

关键思路

代换（substitution）要成为合法的变量转换，前提是当旧变量走过某个区间时，新变量也要平滑地、单向地走过一个区间。这里 $t = 1/x$ 在 $x = 0$ 处趋于无穷，而 $x = 0$ 正好位于 $[-1, 1]$ 的中间，所以“新的上下限”-1 和 1 描述的是一组完全不同的 t 值。只要看一看两个被积函数（integrand）的正负号，就能确认出了问题。

解答

代换为何失效

形式上，设 $x = 1/t$，则 $\mathrm{d}x = -\frac{1}{t^2}\,\mathrm{d}t$，而且

$$\frac{1}{(1+x^2)^2} = \frac{1}{\left(1+\frac{1}{t^2}\right)^2} = \frac{t^4}{(1+t^2)^2}, \qquad \text{所以} \qquad \frac{\mathrm{d}x}{(1+x^2)^2} = \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t,$$

而 $x = \pm 1$ 对应 $t = \pm 1$。这就是“证明”两个积分相等的做法。但这不可能是对的：第一个被积函数处处为正，所以第一个积分是正数；而第二个被积函数处处 ≤ 0，所以第二个积分是负数。

? **想一想** 当 x 由 -1 走到 1 时，$t = 1/x$ 实际上取哪些值?

当 x 由 -1 增加到 0^- 时，t 由 -1 减少到 $-\infty$；在 $x = 0$ 处根本没有对应的 t；而当 x 由 0^+ 走到 1 时，t 由 $+\infty$ 下降到 1。所以 x 的区间 $[-1, 1]$ 对应的是 $t \in (-\infty, -1] \cup [1, \infty)$，即 $[-1, 1]$ 的外面，而不是 $[-1, 1]$ 本身。代换法则要求 $x = \phi(t)$，其中 ϕ 在新上下限之间的整个 t 区间上可导，但 $\phi(t) = 1/t$ 在 $t = 0 \in [-1, 1]$ 处连定义都没有。$(-1, 1)$ 内的 t 值对应 $|x| > 1$，那些值根本在原来的范围以外。

正确地使用这个代换，即在两半上分别进行，例如可得

$$\int_0^1 \frac{\mathrm{d}x}{(1+x^2)^2} = \int_\infty^1 \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t = \int_1^\infty \frac{t^2}{(1+t^2)^2}\,\mathrm{d}t,$$

这是一个正确的命题（下面会加以验证），但它对于 t 在 $[-1,1]$ 上的积分毫无说明。

第一个积分

设 $x = \tan\theta$，则 $\mathrm{d}x = \sec^2\theta\,\mathrm{d}\theta$ 及 $1+x^2 = \sec^2\theta$；这个代换没有问题，因为 tan 在 $[-\frac{\pi}{4}, \frac{\pi}{4}]$ 上平滑而且递增。于是

$$\begin{aligned}\int \frac{\mathrm{d}x}{(1+x^2)^2} &= \int \frac{\sec^2\theta}{\sec^4\theta}\,\mathrm{d}\theta = \int \cos^2\theta\,\mathrm{d}\theta = \int \tfrac{1}{2}(1+\cos 2\theta)\,\mathrm{d}\theta \\ &= \tfrac{1}{2}\theta + \tfrac{1}{4}\sin 2\theta = \tfrac{1}{2}\theta + \tfrac{1}{2}\cdot\frac{\tan\theta}{1+\tan^2\theta} = \tfrac{1}{2}\arctan x + \frac{x}{2(1+x^2)},\end{aligned}$$

这里用了 $\sin 2\theta = 2\sin\theta\cos\theta = 2\tan\theta\cos^2\theta$。因此

$$\int_{-1}^1 \frac{\mathrm{d}x}{(1+x^2)^2} = \left[\tfrac{1}{2}\arctan x + \frac{x}{2(1+x^2)}\right]_{-1}^1 = 2\left(\frac{\pi}{8}+\frac{1}{4}\right) = \boxed{\frac{\pi}{4}+\frac{1}{2}}.$$

第二个积分

把分子写成 $-t^2 = 1-(1+t^2)$：

$$\frac{-t^2}{(1+t^2)^2} = \frac{1}{(1+t^2)^2} - \frac{1}{1+t^2}.$$

由上一部分，第一项在 $[-1,1]$ 上的积分是 $\frac{\pi}{4}+\frac{1}{2}$，而 $\int_{-1}^1 \frac{\mathrm{d}t}{1+t^2} = \left[\arctan t\right]_{-1}^1 = \frac{\pi}{2}$。所以

$$\int_{-1}^1 \frac{-t^2}{(1+t^2)^2}\,\mathrm{d}t = \frac{\pi}{4}+\frac{1}{2}-\frac{\pi}{2} = \boxed{\frac{1}{2}-\frac{\pi}{4}},$$

这是负数（≈ -0.285），正如被积函数的正负号所要求的；而第一个积分约为 1.285。

验算

- 用辛普森法则（Simpson's rule）计算两个积分，得 1.2853981… 和 −0.2853981…，与 $\frac{\pi}{4}+\frac{1}{2}$ 及 $\frac{1}{2}-\frac{\pi}{4}$ 吻合。
- 两个答案之和是 1。这是有原因的：$\frac{1}{(1+x^2)^2} - \frac{x^2}{(1+x^2)^2} = \frac{1-x^2}{(1+x^2)^2}$，它正是 $\frac{x}{1+x^2}$ 的导数（derivative），而 $\left[\frac{x}{1+x^2}\right]_{-1}^1 = 1$。
- 正确的半区间命题也验证无误：$\int_1^\infty \frac{t^2}{(1+t^2)^2}\,\mathrm{d}t = \frac{\pi}{4} - \left(\frac{\pi}{4}-\frac{\pi}{8}-\frac{1}{4}\right) = \frac{\pi}{8}+\frac{1}{4} = \int_0^1 \frac{\mathrm{d}x}{(1+x^2)^2}$。

常见错误

- 只说“$1/t$ 在 $t=0$ 处无定义”，却没有解释 x 的范围 $[-1,1]$ 对应的是 $|t|\geq 1$：真正的重点是两个积分涵盖的是不同的取值集合。
- 把 $\mathrm{d}x=-\frac{1}{t^2}\,\mathrm{d}t$ 的正负号弄错，以致掩盖了那个露出破绽的正负号矛盾。
- 积分 $\cos^2\theta$ 时漏掉 $\frac{1}{4}\sin 2\theta$ 一项。

STEP 之外

代换法则 $\int_a^b f(\phi(t))\phi'(t)\,\mathrm{d}t=\int_{\phi(a)}^{\phi(b)} f(x)\,\mathrm{d}x$ 在大学第一门数学分析（analysis）课程中，会由链式法则（chain rule）和微积分基本定理（Fundamental Theorem of Calculus）证明，而证明过程清楚显示所需的条件：ϕ 必须在整个 $[a,b]$ 上连续可微（continuously differentiable）。这里它在一点上失效，而单单这一点就足以令结果崩溃。Abbott 的 *Understanding Analysis*（《理解分析》）细致地建立了黎曼积分（Riemann integral）及这些定理。

详细解答

STEP I 1987，第 9 题

纯数学 · 重心坐标

> **STEP I 1987，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/87-s1-q9

关键思路

从 A 向 B 移动，就是加上向量（vector）$\overrightarrow{AB} = \mathbf{b} - \mathbf{a}$ 的某个分数倍。做一次，得到的是线段（segment）上的点；做两次，先沿 AB 走，再从该点向 C 走，就能扫遍整个三角形。至于作图部分，窍门是把每种情况改写成我们已经理解的形式：要么是三角形情况的翻转，要么是点 A 加上两条边向量的倍数。

解答

AB 上的点

线段 AB 上的点 P，可以从 A 出发，沿 $\overrightarrow{AB}$ 走全程的 t 倍到达，其中 $0 \le t \le 1$（$t = 0$ 给出 A，$t = 1$ 给出 B）。因此

$$\mathbf{p} = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = (1 - t)\mathbf{a} + t\,\mathbf{b}.$$

设 $\rho = 1 - t$ 及 $\sigma = t$。由于 $0 \le t \le 1$，所以 $\rho \ge 0$ 及 $\sigma \ge 0$，而且 $\rho + \sigma = 1$。这正是所求的形式。反过来，取 $t = \sigma$，每一对这样的 ρ, σ 都给出线段上的一点。

PC 上的点，以及三角形内的点

用完全相同的方法，线段 PC 上的点 Q 的位置向量（position vector）为

$$\mathbf{q} = (1 - s)\mathbf{p} + s\,\mathbf{c}, \qquad 0 \le s \le 1.$$

代入 $\mathbf{p} = \rho\mathbf{a} + \sigma\mathbf{b}$，得

$$\mathbf{q} = (1 - s)\rho\,\mathbf{a} + (1 - s)\sigma\,\mathbf{b} + s\,\mathbf{c}.$$

设 $\lambda = (1-s)\rho$、$\mu = (1-s)\sigma$ 及 $\nu = s$。每一个都是非负（non-negative）数的乘积，所以 $\lambda, \mu, \nu \geq 0$。此外

$$\lambda + \mu + \nu = (1-s)(\rho+\sigma) + s = (1-s) + s = 1.$$

想一想 为甚么三角形内的每一点都位于某条线段 PC 上?

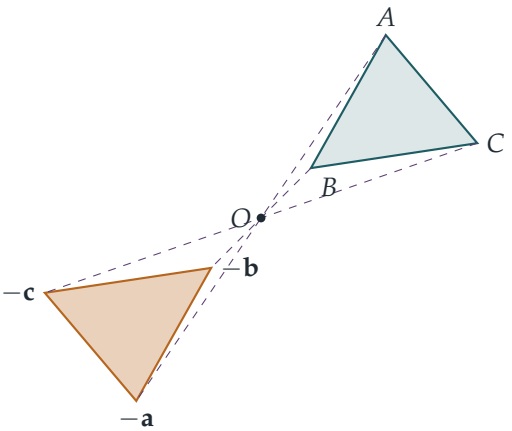
取三角形内（或其边界上）的一点 X。若 $X = C$，则无需证明：取 $s = 1$ 即可。否则，从 C 经过 X 画一条直线，并延伸越过 X。由于 X 位于角 ACB 之内，这条直线保持在 CA 与 CB 之间，因此必定与对边 AB 相交，设交点为 P。由于 X 在三角形内，它位于 C 与 P 之间，所以 X 在线段 PC 上。根据上面的推导，$\mathbf{x} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$，其中 $\lambda, \mu, \nu \geq 0$ 且 $\lambda + \mu + \nu = 1$，命题得证。

第 (i) 部分

设 $\lambda = -\lambda'$、$\mu = -\mu'$ 及 $\nu = -\nu'$。条件变为 $\lambda', \mu', \nu' \geq 0$ 及 $\lambda' + \mu' + \nu' = 1$，而该点为

$$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} = -\left(\lambda'\mathbf{a} + \mu'\mathbf{b} + \nu'\mathbf{c}\right).$$

根据刚才所证，括号内的式子遍历三角形 ABC 的每一点（再加上第一部分的逆命题，每一个这样的组合都在三角形内）。因此所求区域是所有点 $-\mathbf{x}$ 的集合，其中 X 在三角形 ABC 内。换言之，它是以 $-\mathbf{a}$、$-\mathbf{b}$、$-\mathbf{c}$ 为顶点的三角形连同其内部：即三角形 ABC 绕原点 O 旋转半周（half-turn）后的像。



琥珀色阴影三角形就是 (i) 的答案。它的每个顶点都是 ABC 的一个顶点经 O 的点对称像，如虚线所示。

第 (ii) 部分

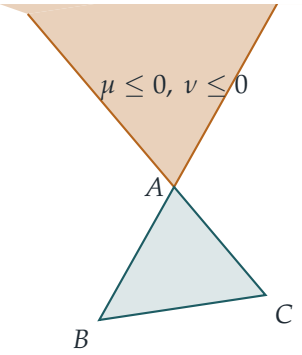
现在 $\lambda = 1 - \mu - \nu$，所以

$$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c} = \mathbf{a} + \mu(\mathbf{b} - \mathbf{a}) + \nu(\mathbf{c} - \mathbf{a}).$$

我们从 A 出发，加上 $\overrightarrow{AB}$ 和 $\overrightarrow{AC}$ 的倍数。设 $\mu = -m$ 及 $\nu = -n$，其中 $m, n \geq 0$，则该点为

$$\mathbf{a} + m(\mathbf{a} - \mathbf{b}) + n(\mathbf{a} - \mathbf{c}), \qquad m \geq 0,\ n \geq 0.$$

向量 $\mathbf{a} - \mathbf{b}$ 指向 BA 越过 A 的延长方向，而 $\mathbf{a} - \mathbf{c}$ 指向 CA 越过 A 的延长方向。两个不平行向量的非负组合恰好填满它们之间的角（由于 $\mathbf{b} - \mathbf{a}$ 与 $\mathbf{c} - \mathbf{a}$ 不平行，平面上每一点都可唯一地写成 $\mathbf{a} + m(\mathbf{a} - \mathbf{b}) + n(\mathbf{a} - \mathbf{c})$，即 m, n 恰有一对）。因此所求区域是以 A 为顶点、以 BA 和 CA 越过 A 的延长线为界的无限楔形区域（wedge），边界射线包括在内。它是角 BAC 的对顶角（vertically opposite angle）。注意题目没有限制 λ，但在这里它自动至少为 1。



琥珀色楔形区域（无限延伸）就是 (ii) 的答案。它的边是边 BA 和 CA 经 A 延长所得。图中青色三角形 ABC 仅供参照。

验算

- 顶点：在 (ii) 中，$(\lambda, \mu, \nu) = (1, 0, 0)$ 给出 A 本身，即楔形的尖端；$(2, -1, 0)$ 给出 $2\mathbf{a} - \mathbf{b}$，即 B 关于 A 的对称点，它位于其中一条边上。在 (i) 中，$(-1, 0, 0)$ 给出 $-\mathbf{a}$，即对称后三角形的一个顶点。
- 重心：在 (i) 中，$\lambda = \mu = \nu = -\frac{1}{3}$ 给出 ABC 重心（centroid）的负值，这正是对称后三角形的重心，与预期相符。
- 数值方面，我们用电脑随机取了 20,000 个三角形，以及满足各组条件的随机系数。结果确认：(i) 的每一点都位于三角形 $(-\mathbf{a}, -\mathbf{b}, -\mathbf{c})$ 内；(ii) 的每一点从 A 出发沿 $\mathbf{a} - \mathbf{b}$ 和 $\mathbf{a} - \mathbf{c}$ 的坐标都是非负的；而 ABC 的每一点都可分解为某条线段 PC 上的点。

常见错误

- 在 (i) 中，在 ABC 附近画出区域。系数之和是 -1 而不是 1，所以该区域是 ABC 绕原点旋转半周的结果，并且取决于 O 的位置。
- 在 (ii) 中，只画出一个三角形。μ 和 ν 没有下界，所以该区域是无界的。
- 把与 BAC 相邻的角涂上阴影（例如 AB 与 CA 延长线之间的角）。那需要 μ, ν 其中之一为正数。

STEP 之外

满足 $\lambda + \mu + \nu = 1$ 的数组 (λ, μ, ν) 称为一点相对于该三角形的重心坐标（barycentric coordinates）。当总和为 1 时，该点与原点的位置无关：把每个位置向量都平移 $\mathbf{d}$，$\lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$ 也会平移 $(\lambda + \mu + \nu)\mathbf{d} = \mathbf{d}$。这就是为甚么 (ii) 的答案是一个真正的几何区域，而 (i) 则取决于 O。三角形本身是 A、B、C 的凸包（convex hull），这个概念是线性规划（linear programming）和凸优化（convex optimisation）的核心。重心坐标的正负号把平面分成七个区域，(ii) 是其中之一。若想从大学一年级的角度认识这些仿射（affine）概念，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

STEP I 1987，第 10 题

力学 • 弹性绳

STEP I 1987，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/87-s1-q10

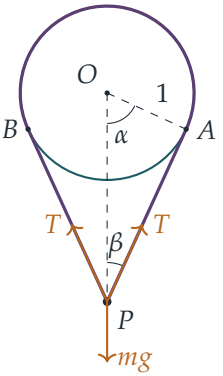
关键思路

质点把橡皮圈的底部从圆柱上拉下来，因此橡皮圈分成两部分：一段仍然紧贴圆柱上方的圆弧，以及两段直线部分；这两段直线沿切线（*tangent*）离开圆柱，并在质点处相交。拉伸后的长度纯粹是几何问题：利用直角三角形，可以求出直线部分的长度，以及它们在轴上所对的角。由于圆柱是光滑（smooth）的，橡皮圈各处的张力（tension）都相同；在质点处分解力，就能把这张力与 mg 联系起来，再用胡克定律（Hooke's law）便可求出 λ。

解答

模型

我们把橡皮圈视为一条自然长度（natural length）为 2π、遵从胡克定律的轻弹性绳，圆柱视为光滑，质点视为点质量。我们在一个竖直横截面上分析：圆柱是一个单位圆，圆心 O 在轴上；质点位于 P，在 O 竖直下方距离 x 处。橡皮圈要离开圆柱，必须有 $x > 1$。



由于图形关于通过 O 的竖直线对称，橡皮圈在两点 A 和 B 离开圆柱，两边各一点。一条拉紧的绳离开光滑曲线时，是沿切线离开的，因此 PA 在 A 点与圆相切，即 $OA \perp AP$。设 $\alpha = \angle AOP$，$\beta = \angle APO$。

拉伸后的长度

三角形 OAP 在 A 处是直角，$OA = 1$，斜边 $OP = x$。因此

$$AP = \sqrt{x^2 - 1}, \qquad \cos\alpha = \frac{1}{x}, \qquad \sin\beta = \frac{1}{x}.$$

除了经底部从 A 到 B 的一段（在 O 处所对的角为 2α）之外，橡皮圈处处与圆柱接触。所以接触部分的圆弧所对的角是 $2\pi - 2\alpha$，在单位圆上其长度就是 $2\pi - 2\alpha$。加上两段直线部分，拉伸后的长度为

$$\boxed{L = 2\sqrt{x^2 - 1} + 2\pi - 2\arccos\frac{1}{x}}$$

其中由于 $x > 1$，$\arccos(1/x)$ 介于 0 与 $\frac{\pi}{2}$ 之间。因此伸长量（extension）为

$$L - 2\pi = 2\sqrt{x^2 - 1} - 2\arccos\frac{1}{x}.$$

? **想一想** 橡皮圈绕着圆柱弯曲。为甚么直线部分的张力与圆弧部分的张力相同?

对于与光滑表面接触的一小段橡皮圈，圆柱施加的力只有垂直于表面的法向力，因此没有任何力可以抵消沿橡皮圈方向的张力变化。所以张力 T 在整条橡皮圈上都是常数，胡克定律可以应用于整条橡皮圈：

$$T = \frac{\lambda\,(L - 2\pi)}{2\pi}.$$

（如果圆柱是粗糙的，张力在圆弧上就可能有变化，这一步便不成立。）

质点的平衡

作用在质点上的力有重量 mg，以及沿 PA 和 PB 方向的张力 T，两者与向上竖直方向的夹角均为 β。由对称性，水平分量互相抵消。竖直方向分解，

$$2T\cos\beta = mg, \qquad \text{其中} \qquad \cos\beta = \frac{\sqrt{x^2-1}}{x},$$

所以

$$T = \frac{mgx}{2\sqrt{x^2-1}}.$$

$x = 2$ 的情况

此时 $\cos\alpha = \frac{1}{2}$，故 $\alpha = \frac{\pi}{3}$，且 $\sqrt{x^2-1} = \sqrt{3}$。拉伸后的长度为

$$L = 2\sqrt{3} + 2\pi - \frac{2\pi}{3} = 2\sqrt{3} + \frac{4\pi}{3}, \qquad \text{所以} \qquad L - 2\pi = 2\sqrt{3} - \frac{2\pi}{3}.$$

张力为 $T = \dfrac{2mg}{2\sqrt{3}} = \dfrac{mg}{\sqrt{3}}$。由胡克定律得

$$\lambda = \frac{2\pi T}{L-2\pi} = \frac{2\pi mg/\sqrt{3}}{2\sqrt{3} - 2\pi/3} = \frac{2\pi mg}{6 - 2\pi/\sqrt{3}} = \frac{\pi mg}{3 - \pi/\sqrt{3}},$$

分子分母同乘以 $\sqrt{3}$，

$$\boxed{\lambda = \frac{\sqrt{3}\,\pi\, mg}{3\sqrt{3} - \pi} \approx 2.65\, mg.}$$

验算

- 当 $x \to 1^+$ 时，$\arccos(1/x) \to 0$ 且 $\sqrt{x^2-1} \to 0$，所以 $L \to 2\pi$：橡皮圈刚好碰到圆柱底部而没有被拉伸，这正合理。同时 $T \to \infty$，这与以下事实吻合：两条接近水平的绳子只有在张力极大时才能承托重物。
- 当 $x = 2$ 时，$L = 2\sqrt{3} + 4\pi/3 \approx 7.653$，比 $2\pi \approx 6.283$ 稍大；而 λ 与 mg 一样具有力的单位。
- 另外，我们取 $\lambda \approx 2.6484\, mg$，以数值方法求总势能（potential energy）$-mgx + \lambda(L - 2\pi)^2/(4\pi)$ 的最小值；最小值出现在 $x = 2$，这同时证实了几何部分和张力部分的结果。

常见错误

- 把直线部分当作从 P 连到水平直径的两个端点。橡皮圈是沿切线离开的，所以它在轴的高度以下离开圆柱，位置与向下竖直方向成角 $\alpha = \arccos(1/x)$。

► 在胡克定律中只用一段直线部分的伸长量。张力是均匀的，所以应该用整条橡皮圈的伸长量 $L - 2\pi$，除以整条的自然长度 2π。

STEP 之外

“在光滑表面上张力处处相同”这一说法，是绞盘方程（capstan equation）在无摩擦情况下的特例：若摩擦系数（coefficient of friction）为 μ，绕过角度 θ 后，张力最多可以改变 $e^{\mu\theta}$ 倍。这就是为甚么一条在柱上绕几圈的绳子能够系住一艘船。上面的能量验算，也让你初步领略力学的变分观点（variational view）：平衡位置就是势能的驻点（stationary point）。若想看大学一年级程度、以这种思路处理力、摩擦和势能的教材，可读 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP I 1987，第 11 题

力学 · 圆周运动与能量

STEP I 1987，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/87-s1-q11

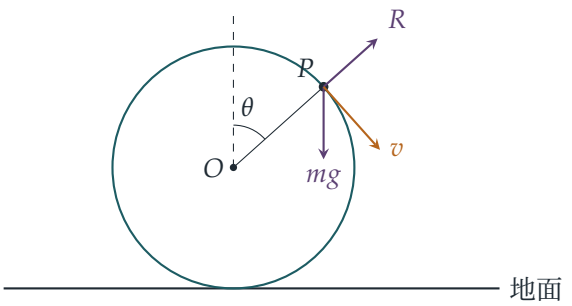
关键思路

质点留在球面上时作圆周运动（circular motion），因此由两条方程支配：能量守恒（conservation of energy）给出其速率，而沿半径方向应用牛顿（Newton）第二定律则给出法向反作用力（normal reaction）。当反作用力降至零的一刻，质点便离开球面。此后质点成为抛射体（projectile），所以其水平速度不再改变；把这一点与由顶部一直到地面的能量守恒结合，便可得出铅直速度，而毋须求出轨迹。

解答

模型

把质点视为一个质点（point mass），置于一个光滑、固定的球面上，球心为 O，位于地面之上高度 r 处（球体搁在地面上）。“轻轻移动”的意思是质点从顶部出发时速率可忽略不计。设 P 为质点所在位置，θ 为 OP 与向上铅垂线之间的角，v 为其速率。作用于质点的力只有重量 mg 及沿 OP 向外的法向反作用力 R。



何处离开球面

质点已下降高度 $r - r\cos\theta$，因此由能量守恒（反作用力与运动方向垂直，故不作功）得

$$\tfrac{1}{2}mv^2 = mgr(1-\cos\theta), \qquad \text{所以} \qquad v^2 = 2gr(1-\cos\theta).$$

质点与球面接触时沿半径为 r 的圆运动，因此其指向 O 的加速度为 v^2/r。沿指向 O 的方向分解：

$$mg\cos\theta - R = \frac{mv^2}{r} = 2mg(1-\cos\theta), \qquad \text{所以} \qquad R = mg(3\cos\theta - 2).$$

在顶部时 $R = mg > 0$，而 R 随 θ 增大而减少。当 $R = 0$ 时质点离开球面，即当

$$\cos\theta = \tfrac{2}{3}, \qquad v^2 = \tfrac{2}{3}gr.$$

> **想一想** 速度与水平线所成的角，是哪一个角？

速度沿圆的切线（tangent）方向，而切线与半径 OP 垂直。半径与铅垂线成角 θ，所以切线与水平线所成的角同样是 θ（把两条线都转一个直角即可）。因此，质点离开球面的一刻，正以水平线之下

$$\cos^{-1}\tfrac{2}{3} \approx 48.2^\circ$$

的角度向下运动。

撞击地面

离开球面的一刻，速度的水平分量为

$$v\cos\theta = \sqrt{\tfrac{2}{3}gr}\cdot\tfrac{2}{3}, \qquad \text{所以} \qquad (v\cos\theta)^2 = \tfrac{8}{27}gr.$$

此后唯一的作用力是重力，所以这个水平分量直至质点到达地面都保持不变。（质点不会再碰到球面：一旦 R 会变成负数，球面向外弯离的速度便快于质点下堕的速度；而质点落地点与通过 O 的铅垂线的水平距离约为 $1.46r$，远离球体。）

现在对整个运动应用能量守恒，由顶部（高度 $2r$）静止开始，一直到地面。质点以速率 V 撞击地面，其中

$$\tfrac{1}{2}mV^2 = mg(2r), \qquad \text{所以} \qquad V^2 = 4gr.$$

设 V_y 为撞击时速度的铅直分量，则 $V^2 = (v\cos\theta)^2 + V_y^2$，所以

$$V_y^2 = 4gr - \tfrac{8}{27}gr = \tfrac{100}{27}gr, \qquad V_y = \frac{10}{3}\sqrt{\frac{gr}{3}}.$$

因此，质点撞击地面时动量（momentum）的铅直分量的大小为

$$mV_y = \frac{10m}{3}\sqrt{\frac{gr}{3}} = \frac{10m\sqrt{3gr}}{9},$$

方向向下。

验算

- 直接用抛射体运动计算：质点在高度 $r + r\cos\theta = \frac{5}{3}r$ 处离开，向下的铅直速率为 $v\sin\theta$，其中 $\sin\theta = \frac{\sqrt{5}}{3}$，所以 $(v\sin\theta)^2 = \frac{10}{27}gr$。再下堕 $\frac{5}{3}r$ 会增加 $2g\cdot\frac{5}{3}r = \frac{90}{27}gr$，再次得到 $V_y^2 = \frac{100}{27}gr$。
- $V_y^2 = \frac{100}{27}gr \approx 3.70gr$ 比 $V^2 = 4gr$ 略小，这是必然的，因为有少量能量留在水平运动之中。
- 以数值方法逐步模拟运动（取 $g = 9.81$、$r = 1.3$、$m = 2$），得出在 48.19° 处离开球面，铅直动量为 13.745，与 $\frac{10m\sqrt{3gr}}{9}$ 吻合；轨迹一直留在球体之外。

常见错误

- 由地面而非由球顶量度下降高度：在球面上，下降的高度是 $r(1-\cos\theta)$。
- 忘记离开球面后水平速度固定不变，误把 $V_y^2 = 4gr$ 写出来，仿佛全部动能都属于铅直方向。
- 误以为速度与水平线成角 $90^\circ - \theta$；那是半径的角度，而非切线的角度。

STEP 之外

“反作用力消失时便离开接触面”这个条件，是约束力（constraint force）的第一个例子：约束力会自行调节，使质点留在某个曲面上，而且只能推、不能拉。在拉格朗日力学（Lagrangian mechanics）中，这类约束从一开始便纳入方程，而反作用力则以拉格朗日乘数（Lagrange multiplier）的形式出现。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）是在更高层次上练习圆周运动和能量方法的绝佳教材。

详细解答

STEP I 1987，第 12 题

力学 · 抛体与恢复系数

STEP I 1987，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/87-s1-q12

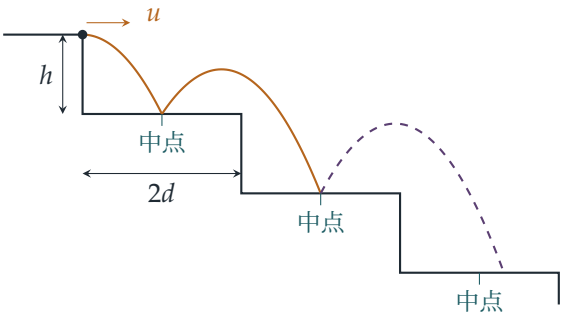
关键思路

在两次反弹之间，粒子是一个抛射体（projectile）；而（台阶光滑时）每次反弹只会把速度的铅直分量反向并按比例缩小，所以水平速率始终不变。这意味著水平距离决定了每段飞行的时间，而飞行时间加上一级台阶的落差又决定了铅直速率。所有量都可以用第一次下落的时间来表示，令代数运算非常简短。

解答

模型

把粒子视为质点（point mass），忽略空气阻力，并假设台阶光滑且水平，因此每次撞击不改变水平速度，而铅直速度则乘以 $-e$。设粒子被踢出时的水平速率为 u，并设 t_1 为第一段飞行的时间。



实线弧是首两段飞行；虚线弧是接下来实际发生的情况（见最后一部分）。

恢复系数

第一段飞行。粒子开始时没有铅直速度，下落 h，并水平移动 d（半个台阶宽度，即到达第一级台阶的中点）。因此

$$h = \tfrac{1}{2}gt_1^2, \qquad d = ut_1.$$

粒子以向下的铅直速率 gt_1 撞击第一级台阶，然后以向上的铅直速率 egt_1 离开，水平速率仍是 u。

第二段飞行。由第一级台阶的中点到第二级台阶的中点，水平距离为 $2d$。由于水平速率仍是 $u = d/t_1$，这段飞行历时

$$T = \frac{2d}{u} = 2t_1.$$

在这段时间内，粒子最终必须下降高度 h。取向上为正，

$$\begin{aligned} -h &= (egt_1)(2t_1) - \tfrac{1}{2}g(2t_1)^2 \\ &= (2e-2)\,gt_1^2 \\ &= (2e-2)(2h), \end{aligned}$$

其中用了 $gt_1^2 = 2h$。两边除以 h 得 $-1 = 4e - 4$，所以

$$e = \tfrac{3}{4}.$$

（在这段飞行中，粒子越过第一级台阶边缘时，位于该台阶上方 $egt_1 \cdot t_1 - \frac{1}{2}gt_1^2 = \frac{1}{2}h$ 的高度，所以它确实能越过边缘。）

能否一直持续下去?

> **想一想** 如果这个规律持续下去，之后每一段飞行必须是怎样的?

每段由一级台阶中点飞到下一级台阶中点的飞行，都以速率 u 水平移动 $2d$，所以每段都历时 $2t_1$，而且每段都必须恰好下降 h。由上面的计算可知，每次反弹后瞬间向上的铅直速率 w 必须满足

$$-h = w(2t_1) - 2gt_1^2 = 2wt_1 - 4h, \qquad \text{所以} \qquad w = \frac{3h}{2t_1} = \tfrac{3}{4}gt_1.$$

因此每次反弹后，粒子都必须以铅直速率 $\frac{3}{4}gt_1$ 离开；由于 $e = \frac{3}{4}$，它到达每一级台阶时的铅直速率都必须恰好是 gt_1。

这在第一级台阶成立，但在第二级台阶并不成立。经过历时 $2t_1$ 的飞行到达第二级台阶时，其铅直速度为

$$\tfrac{3}{4}gt_1 - g(2t_1) = -\tfrac{5}{4}gt_1,$$

所以它以向下速率 $\frac{5}{4}gt_1$（而非 gt_1）撞击第二级台阶，并以向上速率 $\frac{3}{4}\cdot\frac{5}{4}gt_1=\frac{15}{16}gt_1>\frac{3}{4}gt_1$ 反弹。所以这是**不可能的**：粒子不能击中每一级台阶的中点。事实上，它会越过第三级台阶的中点。设第三段飞行历时 st_1，则

$$-h=\tfrac{15}{16}gt_1(st_1)-\tfrac{1}{2}gs^2t_1^2 \quad\Longrightarrow\quad s^2-\tfrac{15}{8}s-1=0 \quad\Longrightarrow\quad s=\frac{15+\sqrt{481}}{16}\approx 2.31,$$

所以它水平移动约 $2.31d$，落在第三级台阶上、中点以外约 $0.31d$ 处（仍在该台阶上，因为远端边缘在 $3d$ 之外）。

验算

- $e=\frac{3}{4}$ 与 g、h 或 d 无关，这是理所当然的：把整道楼梯按比例放大或缩小，并不会改变路径的形状。
- 以 $g=9.8$、$h=d=1$ 及 $e=0.75$ 进行数值模拟，粒子落在 $x=1.000$ 及 $x=3.000$（首两级台阶的中点），然后落在 $x\approx 5.308$，与上面预测的 $3+2.308$ 吻合。
- 若没有反弹（$e=0$），第二段飞行在时间 $2t_1$ 内会下降 $4h$；反弹必须补回其中的 $3h$，即 $2t_1\cdot w=3h$，与所得结果一致。

常见错误

- 把第一段水平距离当成 $2d$ 而不是 d：粒子由边缘出发，落在第一级台阶的中点。
- 在反弹时改变了水平速率。在光滑表面上，冲量（impulse）是铅直的，所以只有铅直分量会改变。
- 因为每段飞行看似相同而得出“可以”的结论：到达时的速率会不断增大，所以反弹也会随之增大。

STEP 之外

弹跳的粒子是*离散动力系统*（discrete dynamical system）的一个简单例子：某次反弹后瞬间的状态，经由一个固定的映射决定下一次反弹后的状态；而问运动能否重复，就是问这个映射的不动点（fixed point）。在这里，不动点（反弹速率为 $\frac{6}{7}gt_1$，只会落低 $\frac{4}{7}h$）与这道楼梯并不相容，所以没有任何周期运动（periodic motion）能符合。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对碰撞与冲量有细致的讨论。

详细解答

STEP I 1987，第 13 题

力学 · 变力

> **STEP I 1987，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/87-s1-q13

关键思路

力的规律在 $x = 0$ 处改变，因此运动分成几个阶段，每个阶段都由一条简单的线性微分方程（differential equation）支配。在正方向的一侧，力是与位移成正比的回复力（restoring force），即简谐运动（simple harmonic motion）；在负方向的一侧，力是与速度成正比的阻力（resistance），令速率呈指数衰减（exponential decay）。一个阶段结束时的位置和速度，就是下一个阶段的初始条件（initial conditions）。真正的问题在于共有多少个阶段：质点究竟会不会回来？

解答

模型

把物体视为 x 轴上只受所给的力作用的质点，并记 $\dot{x} = \mathrm{d}x/\mathrm{d}t$。牛顿第二定律（Newton's second law）$m\ddot{x} = \mathrm{F}$ 给出

$$\ddot{x} = -\mu^2 x \quad (x \geq 0), \qquad \ddot{x} = -\kappa\dot{x} \quad (x < 0).$$

质量互相抵消，因此与答案无关。

第一阶段：$x \geq 0$ 时的简谐运动

由于 $v_0 > 0$，质点立即进入 $x > 0$ 的区域，在那里 $\ddot{x} = -\mu^2 x$。这是角频率（angular frequency）为 μ 的简谐运动，其通解（general solution）为

$$x = A\cos\mu t + B\sin\mu t.$$

条件“$t=0$ 时 $x=0$”给出 $A=0$，而“$t=0$ 时 $\dot{x}=B\mu=v_0$”给出 $B=v_0/\mu$。因此

$$x=\frac{v_0}{\mu}\sin\mu t,\qquad \dot{x}=v_0\cos\mu t.$$

只要 $x\geq 0$，即 $\sin\mu t\geq 0$，这个解便成立，也就是 $0\leq t\leq \pi/\mu$。质点在 $t=\pi/(2\mu)$ 时到达最大位移 v_0/μ，并在

$$t=T=\frac{\pi}{\mu},\qquad \text{此时速度为}\qquad \dot{x}=v_0\cos\pi=-v_0.$$

时返回原点。它回来时的速率与出发时相同，因为在这一侧能量守恒（力 $-m\mu^2x$ 来自位能（potential energy）$\frac{1}{2}m\mu^2x^2$，而位能在原点为零）。

第二阶段：$x<0$ 时的受阻运动

质点以速度 $-v_0$ 进入 $x<0$ 的区域，在那里 $\ddot{x}=-\kappa\dot{x}$。记 $u=\dot{x}$，这就是一阶方程（first-order equation）$\mathrm{d}u/\mathrm{d}t=-\kappa u$，满足“$t=T$ 时 $u=-v_0$”的解为

$$\dot{x}=-v_0\,\mathrm{e}^{-\kappa(t-T)}.$$

利用“$t=T$ 时 $x=0$”积分：

$$x=\int_T^t -v_0\,\mathrm{e}^{-\kappa(s-T)}\,\mathrm{d}s=-\frac{v_0}{\kappa}\left(1-\mathrm{e}^{-\kappa(t-T)}\right).$$

> **想一想** 质点会不会回到 $x\geq 0$，开始第三个阶段?

不会。对每个 $t>T$，速度 $-v_0\mathrm{e}^{-\kappa(t-T)}$ 都是负数，所以在整个阶段中 x 不断减少；对所有 $t>T$，它都严格为负，上面的公式永远成立。阻力令质点减速，却永远不会令它反向。当 $t\to\infty$ 时，$\mathrm{e}^{-\kappa(t-T)}\to 0$，因此质点趋近点 $x=-v_0/\kappa$，但永远不会到达。

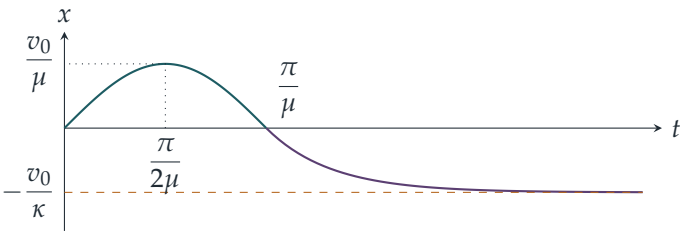
答案

$$x=\begin{cases}\dfrac{v_0}{\mu}\sin\mu t, & 0\leq t\leq\dfrac{\pi}{\mu},\\[2ex] -\dfrac{v_0}{\kappa}\left(1-\mathrm{e}^{-\kappa(t-\pi/\mu)}\right), & t>\dfrac{\pi}{\mu}.\end{cases}$$

两段在 $t=\pi/\mu$ 处都有 $x=0$ 和 $\dot{x}=-v_0$，所以图像在该处是光滑的（斜率连续），尽管 $\ddot{x}$ 从 0 跳到 κv_0。

草图

图像先是轴上方半个正弦曲线的拱形，接着是一条以相同斜率 $-v_0$ 离开原点的曲线，并逐渐趋平，趋近渐近线（asymptote）$x=-v_0/\kappa$。（草图取 $\mu=1$、$\kappa=1$。）



一般而言，峰高 v_0/μ 与渐近线的深度 v_0/κ 互不相关：阻力较弱（κ 较小）时，质点会向左漂移很远。

验算

- 单位：μ 和 κ 的单位都是 s^{-1}，所以 v_0/μ 和 v_0/κ 都是长度，理应如此。
- 极限情况：当 $\kappa \to 0$ 时没有阻力，而确实 $-\frac{v_0}{\kappa}(1-\mathrm{e}^{-\kappa s}) \to -v_0 s$，即以速率 v_0 作匀速运动。
- 数值验证：取 $\mu = 1.7$、$\kappa = 0.9$、$v_0 = 2.2$，用龙格—库塔法（Runge–Kutta method）逐步求解运动方程，在 $0 \le t \le 15$ 范围内与公式的误差在 3×10^{-5} 以内；质点的峰值为 $x = 1.2941 = v_0/\mu$，最终停在 $x = -2.4444 = -v_0/\kappa$ 附近。

常见错误

- 第二阶段的时间由 $t = 0$ 而不是由 $t = \pi/\mu$ 起计：指数必须是 $-\kappa(t - \pi/\mu)$。
- 忘记检查质点会否重新进入 $x \ge 0$；题目要求的是其后的整个运动，所以“没有第三阶段”这个结论需要一句理由支持。

STEP 之外

这里的力在 $x = 0$ 两侧各是一个不同的线性表达式，属于分段线性系统（piecewise-linear system）；把各段的解拼接起来、在切换点匹配位置和速度的技巧，是微分方程入门课程处理这类系统（以及碰撞和摩擦问题）的标准方法。若同时加入两项，得 $\ddot{x} + \kappa\dot{x} + \mu^2 x = 0$，这就是阻尼谐振子（damped harmonic oscillator），它的欠阻尼（underdamped）、临界阻尼（critically damped）和过阻尼（overdamped）三种情况，是任何力学入门课程的核心例子。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对阻尼振子有细致的讨论。

详细解答

STEP I 1987，第 14 题

概率与统计 • 概率的等比级数

STEP I 1987，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/87-s1-q14

关键思路

第一局打完之后，一场仍未分胜负的锦标赛便完全没有自由度可言。每局的胜者要对上该局的裁判（referee）；若锦标赛要继续未分胜负，裁判就必须胜出，因此下一局的胜者是被迫决定的。所以一切只取决于谁赢第 1 局，以及某位选手首次连胜两局的那一刻。之后各概率便化为以周期（period） 3 循环的几何级数（geometric series）。

解答

在整个过程中，每局都由上一局的胜者对上上一局的裁判。上一局的负方在旁担任裁判，因此不可能赢当前这一局。

第 (i) 部分

第 1 局：A 击败 B，由 B 担任第 2 局的裁判。第 2 局是 A 对 C，而 C 胜出，这是可能的。第 3 局是 C 对 B，而 B 胜出，这同样可能。但没有人连胜两局，所以在 ACB 之后锦标赛仍未分胜负，必须继续。因此 ACB **不是**一场完整锦标赛的纪录（它只是一场锦标赛的开头）。

第 (ii) 部分

A 赢了第 1 局后，第 2 局是 A 对 C，由 B 担任裁判。所以 B 不可能赢第 2 局，ABB 是**不可能**的。

第 (iii) 部分

与 (i) 一样，首三局 ACB 是可能的，而第 3 局是 C 对 B，由 A 担任裁判。所以第 4 局是 B 对 A，而 B 可以赢这一局。这时 B 连赢了第 3 和第 4 局，锦标赛结束。因此 $ACBB$ **是**一场完整的锦标赛，由 B 在四局内胜出。

5 局后仍未分胜负

设 X 赢第 1 局，其中 X 是 A 或 B，并设 Y 为另一人。若锦标赛要保持未分胜负，每局都必须由*不是*上一局胜者的选手胜出，而这只能是上一局的裁判。这就迫使纪录为

$$X\,C\,Y\,X\,C\,Y\,X\,\ldots$$

因为每位新的胜者都是上一局的裁判。具体来说，第 2 局是 X 对 C（由 Y 担任裁判），所以 C 必须胜出。第 3 局是 C 对 Y（由 X 担任裁判），所以 Y 必须胜出。第 4 局是 Y 对 X，所以 X 必须胜出，如此类推，周期为 3。

第 1 局谁胜都可以。之后，每多一局保持未分胜负的概率都是 $\frac{1}{2}$，因为各局互相独立（independent），而且每位选手的胜算均等。所以

$$P(n \text{ 局后仍未分胜负}) = \left(\tfrac{1}{2}\right)^{n-1}, \qquad P(5 \text{ 局后仍未分胜负}) = \tfrac{1}{16}.$$

谁胜出，何时胜出

锦标赛在第 n 局（对 $n \geq 2$）结束，当且仅当 $n-1$ 局后仍未分胜负，而第 $n-1$ 局的胜者再次胜出。其概率为 $\left(\frac{1}{2}\right)^{n-2} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^{n-1}$。冠军就是第 $n-1$ 局的胜者，即被迫序列中的第 $(n-1)$ 个字母。记 $k = n-1$，则：

$k \bmod 3$	1	2	0
若 A 赢第 1 局，冠军为	A	C	B
若 B 赢第 1 局，冠军为	B	C	A

所以锦标赛在第 $k+1$ 局结束的概率为 $\left(\frac{1}{2}\right)^{k}$。若 $k \equiv 2 \pmod 3$，这整个概率都归 C。否则它由 A 和 B 平分，各得 $\left(\frac{1}{2}\right)^{k+1}$。

> ? **想一想** 为甚么在 $k \equiv 2$ 时 C 得到全部概率，而在其他时候只得一半?

无论谁赢第 1 局，C 永远是被迫序列中的第二个字母。所以两个分支都在同样的时刻让 C 胜出。A 和 B 在两个分支之间互换角色，所以他们各自只能取得自己占据该位置的那一个分支。

在 5 局或以内胜出

这里 $n = 2,3,4,5$，所以 $k = 1,2,3,4$：

$$\begin{aligned} P(A \text{ 胜出，局数} \le 5) &= \underbrace{\tfrac{1}{4}}_{k=1} + \underbrace{\tfrac{1}{16}}_{k=3} + \underbrace{\tfrac{1}{32}}_{k=4} = \tfrac{11}{32}, \\ P(B \text{ 胜出，局数} \le 5) &= \tfrac{11}{32} \quad \text{（由 } A \text{ 与 } B \text{ 之间的对称性），} \\ P(C \text{ 胜出，局数} \le 5) &= \underbrace{\tfrac{1}{4}}_{k=2}. \end{aligned}$$

验算一下，$\frac{11}{32} + \frac{11}{32} + \frac{8}{32} = \frac{15}{16} = 1 - \frac{1}{16}$，与 5 局后仍未分胜负的概率吻合。

最终胜者

由于 $P(n \text{ 局后仍未分胜负}) = \left(\frac{1}{2}\right)^{n-1} \to 0$，锦标赛以概率 1 结束。所以我们只需对所有 $k \ge 1$ 求和。每个剩余类（residue class）都给出一个公比为 $\frac{1}{8}$ 的几何级数：

$$\begin{aligned} P(C \text{ 胜出}) &= \tfrac{1}{4} + \tfrac{1}{32} + \tfrac{1}{256} + \cdots = \frac{1/4}{1 - 1/8} = \frac{2}{7}, \\ P(A \text{ 胜出}) &= \underbrace{\left(\tfrac{1}{4} + \tfrac{1}{32} + \cdots\right)}_{k \equiv 1} + \underbrace{\left(\tfrac{1}{16} + \tfrac{1}{128} + \cdots\right)}_{k \equiv 0} = \frac{1/4}{7/8} + \frac{1/16}{7/8} = \frac{2}{7} + \frac{1}{14} = \frac{5}{14}, \end{aligned}$$

正如所求。由对称性，$P(B \text{ 胜出}) = \frac{5}{14}$ 亦然，而且确实 $\frac{5}{14} + \frac{5}{14} + \frac{2}{7} = 1$。

所以第一局在旁轮空的 C 处于劣势：$\frac{2}{7} \approx 0.286$，而其余两人各有 $\frac{5}{14} \approx 0.357$。

验算

- 三个在 5 局内胜出的概率之和为 $\frac{15}{16}$，而最终胜出的概率之和为 1。
- 以分数对所有可能的对局序列作精确枚举，确认了 $\frac{1}{16}$、$\frac{11}{32}$、$\frac{11}{32}$、$\frac{1}{4}$。计算至 60 局，得出 A 和 B 的值为 0.357142…，C 的值为 0.285714…。
- 求 C 的另一个快捷方法：设 c 为 C 胜出的概率。第 1 局后，C 立即胜出的概率为 $\frac{1}{4}$。若三局之后局面回到相同的情况（概率为 $\frac{1}{8}$），便重新开始，所以 $c = \frac{1}{4} + \frac{1}{8}c$，得 $c = \frac{2}{7}$。

常见错误

- 说 ACB 是有效的纪录：其中每一局都是可能的，但没有人连胜两局，所以锦标赛还未结束。
- 忘记上一局的负方担任裁判，因而不能赢下一局（这排除了 ABB）。

► 在 $k \equiv 2$ 时只给 C 一半的概率：C 在两个分支中都在这些时刻胜出。

STEP 之外

这个锦标赛是一条马可夫链（Markov chain）：每局之后的状态就只是“上一局谁赢、谁当裁判”，而当链被吸收（absorbed）时冠军便决定了。“验算”一节中的首步分析（first-step argument），即通过以下一步发生的事为条件，为某个概率写出方程，是求吸收概率（absorption probability）的标准方法，而且无须对级数求和。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）在大学一年级程度介绍马可夫链与吸收。

详细解答

STEP I 1987，第 15 题

概率与统计 · 连续随机变量

STEP I 1987，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/87-s1-q15

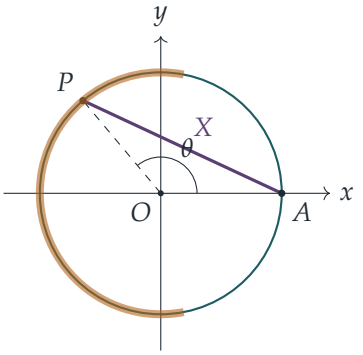
关键思路

“在圆上均匀分布（uniformly distributed）”的意思是 P 在圆心所成的角是均匀分布的，所以自然的做法是设 $P=(\cos\theta,\sin\theta)$，其中 θ 在 $[0,2\pi)$ 上均匀分布。这样，与 $(1,0)$ 的距离就是单位圆的一条弦（chord），而半角公式（half-angle formula）可把它化成单一个正弦。之后，平均值（mean）只需一个积分，方差（variance）几乎完全不用积分，而所求概率则变成一个关于角度区间的问题。

解答

设置

设 θ 为由正 x 轴到 OP 的角，其中 O 为原点。在圆上均匀地选取 P，即是均匀地选取弧长（arc length），所以 θ 在 $[0,2\pi)$ 上均匀分布，概率密度（density）为 $\frac{1}{2\pi}$。记 $A=(1,0)$ 及 $P=(\cos\theta,\sin\theta)$。



于是

$$X^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2 - 2\cos\theta = 4\sin^2\tfrac{\theta}{2},$$

这里用了 $1 - \cos\theta = 2\sin^2\frac{\theta}{2}$。由于 $0 \le \frac{\theta}{2} < \pi$，我们有 $\sin\frac{\theta}{2} \ge 0$，所以取正平方根是稳妥的：

$$X = 2\sin\tfrac{\theta}{2}.$$

平均值

$$\mathrm{E}(X) = \frac{1}{2\pi}\int_0^{2\pi} 2\sin\tfrac{\theta}{2}\,\mathrm{d}\theta = \frac{1}{2\pi}\Big[-4\cos\tfrac{\theta}{2}\Big]_0^{2\pi} = \frac{1}{2\pi}(4+4) = \frac{4}{\pi}.$$

所以 $\mathrm{E}(X) = \dfrac{4}{\pi} \approx 1.273$。

方差

想一想 要求 $\mathrm{E}(X^2)$，我们需要对 $4\sin^2\frac{\theta}{2}$ 积分吗？

不需要：未化成平方的形式更容易处理。$X^2 = 2 - 2\cos\theta$，而 $\cos\theta$ 在整整一圈上的平均值为零，因为 $\int_0^{2\pi}\cos\theta\,\mathrm{d}\theta = 0$。因此 $\mathrm{E}(X^2) = 2$，而

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = 2 - \frac{16}{\pi^2} \approx 0.379.$$

X 大于其平均值的概率

我们需要 $2\sin\frac{\theta}{2} > \frac{4}{\pi}$，即 $\sin\frac{\theta}{2} > \frac{2}{\pi}$。留意 $0 < \frac{2}{\pi} < 1$，所以此不等式有解。设 $\alpha = \arcsin\frac{2}{\pi}$，它位于 $(0, \frac{\pi}{2})$ 内。当 $\frac{\theta}{2}$ 取遍 $[0,\pi)$ 时，正弦值由 0 升至 1 再回落，而它大于 $\sin\alpha$ 当且仅当

$$\alpha < \tfrac{\theta}{2} < \pi - \alpha, \qquad \text{即} \qquad 2\alpha < \theta < 2\pi - 2\alpha.$$

这就是图中加粗的圆弧：圆上位于 A 对面一侧的点。由于 θ 均匀分布，所求概率就是此区间的长度除以 2π：

$$\mathrm{P}\big(X > \mathrm{E}(X)\big) = \frac{2\pi - 4\alpha}{2\pi} = \boxed{1 - \frac{2}{\pi}\arcsin\frac{2}{\pi}} \approx 0.561.$$

验算

- X 的取值范围是 $[0,2]$，而平均值 $4/\pi \approx 1.27$ 位于中点 1 之上。这与图像吻合：在圆上颇大的一部分，弦长都接近其最大值（当 P 经过对面一侧时，X 变化缓慢），所以分布偏向较大的值。

- 基于同样原因，大于平均值的概率应略大于 $\frac{1}{2}$，而 0.561 正是如此。
- 在数值上，以中点法则（midpoint rule）对 2×10^5 个等距角度取平均，得到平均值 1.273240、方差 0.378861 及 $\mathrm{P}=0.56066$；而一个 10^6 点的模拟（simulation）在抽样误差范围内亦与此一致。

常见错误

- 取 $\theta\in(-\pi,\pi]$ 而仍写 $X=2\sin\frac{\theta}{2}$。此时 $\sin\frac{\theta}{2}$ 可以是负数，你需要写 $X=2\left|\sin\frac{\theta}{2}\right|$；使用 $[0,2\pi)$ 就可避免绝对值（modulus）。
- 把 P 的 x 坐标当作在 $[-1,1]$ 上均匀分布。在圆上均匀分布是指在角度（弧长）上均匀，而不是在 x 上均匀。

STEP 之外

这是随机变量变换（transformation of a random variable）的第一个例子：X 的概率密度函数是 $[0,2]$ 上的 $f(x)=\dfrac{2}{\pi\sqrt{4-x^2}}$，属于反正弦（arcsine）类型的分布，而上面每一个答案都可以由它重新推导出来。这道题亦暗示了“随机”一词可以多么难以捉摸：以不同的自然方式随机选取圆的一条弦，会得出不同的答案（贝特朗悖论（Bertrand's paradox））。这两个概念在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中都有进一步发展。

详细解答

STEP I 1987，第 16 题

概率与统计 • 贝叶斯定理与正态近似

> **STEP I 1987，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/87-s1-q16

关键思路

第 (i) 部分其实是乔装了的贝叶斯定理（Bayes' theorem）。两次投票立场相同，是判断该议员属于哪个政党的证据（evidence），因为守成党（Preservative）议员必定如此，而进步党（Progressive）议员只是有时如此。所以我们先更新该议员是守成党议员的概率，然后才预测第三次投票。在第 (ii) 部分，守成党议员的投票是固定的数目，所以所有随机性都来自那 40 位进步党议员。他们投赞成票的人数服从二项分布（binomial distribution），而我们要用正态分布（normal distribution）来近似的，正是这个分布。

解答

第 (i) 部分

设 R 为“被选中的议员是守成党议员”这事件，故 $P(R) = \frac{60}{100} = \frac{3}{5}$，而 $P(R') = \frac{2}{5}$。设 S_2 为“该议员首两次投票立场相同”这事件，S_3 为“第三次投票的立场亦相同”这事件。

守成党议员从不改变主意，所以 $P(S_2 \mid R) = 1$。进步党议员则随机投票，每次投票都独立（independently）地以概率 $\frac{1}{2}$ 投赞成或反对。无论第一次投了甚么，第二次与之相同的概率都是 $\frac{1}{2}$，所以 $P(S_2 \mid R') = \frac{1}{2}$。因此

$$P(S_2) = \tfrac{3}{5} \cdot 1 + \tfrac{2}{5} \cdot \tfrac{1}{2} = \tfrac{4}{5}, \qquad P(R \mid S_2) = \frac{P(R)\,P(S_2 \mid R)}{P(S_2)} = \frac{3/5}{4/5} = \frac{3}{4}.$$

所以，根据我们观察所得，该议员是守成党议员的概率为 $\frac{3}{4}$，是进步党议员的概率为 $\frac{1}{4}$。

? **想一想** 既然该议员是随机选出的，为甚么我们不能直接说 $P(R \mid S_2) = \frac{3}{5}$?

因为信息 S_2 对守成党议员而言比对进步党议员更有可能出现，所以它会改变胜算比

（odds）。观察之前，胜算比是 $3:2$，偏向守成党议员。观察到 S_2 之后，胜算比变成 $3\cdot 1 : 2\cdot \frac{1}{2} = 3:1$。

现在，在已知该议员所属政党的条件下，第三次投票与首两次投票无关，首两次投票的作用只在于透露该议员属于哪个政党。守成党议员必定重复同一立场，进步党议员则以概率 $\frac{1}{2}$ 投出相同立场。所以

$$P(S_3 \mid S_2) = \tfrac{3}{4}\cdot 1 + \tfrac{1}{4}\cdot\tfrac{1}{2} = \frac{7}{8}.$$

（等价地，$P(S_3 \mid S_2) = \dfrac{P(S_2 \cap S_3)}{P(S_2)} = \dfrac{\frac{3}{5}+\frac{2}{5}\cdot\frac{1}{4}}{\frac{4}{5}} = \dfrac{7/10}{4/5} = \dfrac{7}{8}$。）

第 (ii) 部分

考虑就一项守成党政策进行的投票，并假设全部 100 位议员都投票。60 位守成党议员中有比例 α，即 60α 位议员，总是投反对票，其余 $60(1-\alpha)$ 位则总是投赞成票。设 X 为投赞成票的进步党议员人数。每位进步党议员独立地以概率 $\frac{1}{2}$ 投赞成票，所以

$$X \sim \mathrm{B}\left(40, \tfrac{1}{2}\right), \qquad E(X) = 20, \qquad \mathrm{Var}(X) = 40\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = 10.$$

在 100 票中取得简单多数（simple majority），即至少有 51 票赞成：

$$60(1-\alpha) + X \geq 51 \iff X \geq 60\alpha - 9.$$

由于 $n = 40$ 相当大，而 $p = \frac{1}{2}$ 令二项分布对称，我们用 $Y \sim \mathrm{N}(20, 10)$ 来近似 X。这里 $60\alpha - 9$ 是整数（它是议员人数），所以加上连续性校正（continuity correction）后，

$$P(\text{过半数}) = P(X \geq 60\alpha - 9) \approx P(Y > 60\alpha - 9.5) = P\left(Z > \frac{60\alpha - 29.5}{\sqrt{10}}\right),$$

其中 $Z \sim \mathrm{N}(0,1)$。

这个概率随 α 增加而递减，所以恰好当 α 大得足以把它压到 0.8 以下时，便会举行选举。查表得 $P(Z > -0.8416) = 0.8$，所以我们需要

$$\frac{60\alpha - 29.5}{\sqrt{10}} > -0.8416 \iff 60\alpha > 29.5 - 0.8416\sqrt{10} = 26.84$$

$$\iff \alpha > 0.447.$$

由于 60α 必须是整数（议员人数），最小的可能值是 $60\alpha = 27$。所以导致选举的 α 最小值为

$$\alpha = \frac{27}{60} = 0.45.$$

验证这个答案：有 26 位倒戈议员时，取得过半数的近似概率为 $P(Z > -1.107) \approx 0.866$；有 27 位倒戈议员时，则为 $P(Z > -0.791) \approx 0.785 < 0.8$。

验算

- 在 (i) 中，$\frac{7}{8}$ 介乎 $\frac{1}{2}$（已知是进步党议员）与 1（已知是守成党议员）之间，理应如此。以 400,000 位随机议员进行模拟，得出 0.875。
- 在 (ii) 中，精确的二项概率为 $P(X \geq 17) = 0.8659$（26 位倒戈议员）及 $P(X \geq 18) = 0.7852$（27 位倒戈议员）。这些数值与正态近似值在三位小数内一致，并确认分界点位于 26 与 27 位倒戈议员之间，即 $\alpha = 0.45$。

常见错误

- 在 (i) 中答 $\frac{3}{5} \cdot 1 + \frac{2}{5} \cdot \frac{1}{2} = \frac{4}{5}$。这忽略了“该议员已经两次投出相同立场”这项信息。
- 把简单多数当作“至少 50 票”。50 对 50 打和并不算过半数，所以需要 51 票。
- 漏了连续性校正。没有它的话，临界值会变成 $\alpha > 0.456$，从而得出 28 位倒戈议员而不是 27 位；根据精确的二项计算，这是错的。

STEP 之外

第 (i) 部分是贝叶斯更新（Bayesian updating）的一个小例子：数据改变我们对某个隐藏“类型”的信念，然后预测就是对更新后的信念取平均（即后验预测分布（posterior predictive distribution））。第 (ii) 部分的正态近似是中央极限定理（central limit theorem）的一个例子，而这里近乎完美的吻合并非偶然，因为当 $p = \frac{1}{2}$ 时二项分布是对称的。这两个概念在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中都有仔细的阐述。

详细解答

STEP II 1987，第 1 题

纯数学 · 代数恒等式

STEP II 1987，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/87-s2-q1

关键思路

不要急于代入消去其中一个字母，而应该把右边展开（expand）。平方项 a^2、b^2、c^2 在两边都出现，可以互相抵消，剩下的是 x 乘以一个线性表达式（linear expression），而这个表达式恰好是已知条件的某个倍数。一旦看清数字 7 和 11 从何而来，一般结果便水到渠成。

解答

先一次过做一般情况的展开会比较方便。对任意数 p、q、r，

$$(px-a)^2+(qx-b)^2+(rx-c)^2=(p^2+q^2+r^2)x^2-2x(pa+qb+rc)+a^2+b^2+c^2,$$

所以

$$(px-a)^2+(qx-b)^2+(rx-c)^2-(a^2+b^2+c^2)=x\big[(p^2+q^2+r^2)x-2(pa+qb+rc)\big]. \qquad (*)$$

第 (i) 部分

这里 $(p,q,r)=(1,2,3)$，所以 $p^2+q^2+r^2=14$，而 $(*)$ 变成

$$(x-a)^2+(2x-b)^2+(3x-c)^2-(a^2+b^2+c^2)=x\big[14x-2(a+2b+3c)\big].$$

若 $a+2b+3c=7x$，方括号内便是 $14x-14x=0$。因此两边相等，得证。

第 (ii) 部分

现在 $(p,q,r)=(2,3,3)$，所以 $p^2+q^2+r^2=4+9+9=22$，而

$$(2x-a)^2+(3x-b)^2+(3x-c)^2-(a^2+b^2+c^2)=x\big[22x-2(2a+3b+3c)\big].$$

若 $2a+3b+3c=11x$，方括号内便是 $22x-22x=0$，这就证明了 (ii)。

> **想一想** 为甚么 (i) 中是 7，(ii) 中是 11?

它们都是 $p^2+q^2+r^2$ 的一半：$7=\frac{1}{2}(1+4+9)$，$11=\frac{1}{2}(4+9+9)$。这就告诉我们一般的命题应该是甚么。

一般结果

恒等式（identity）$(*)$ 直接证明了以下结果。

对任意实数 p,q,r,a,b,c,x，若

$$2(pa+qb+rc)=(p^2+q^2+r^2)\,x,$$

则

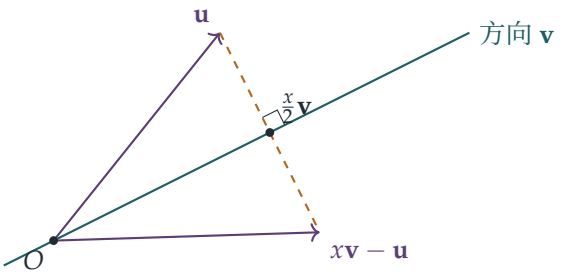
$$a^2+b^2+c^2=(px-a)^2+(qx-b)^2+(rx-c)^2.$$

第 (i) 部分是 $(p,q,r)=(1,2,3)$ 的情况，第 (ii) 部分是 $(p,q,r)=(2,3,3)$ 的情况。

同样的展开对任意多个变量都适用。若 $2\sum_{i=1}^{n}p_ia_i=\left(\sum_{i=1}^{n}p_i^2\right)x$，则 $\sum a_i^2=\sum(p_ix-a_i)^2$。由 $(*)$ 我们还得到一个逆命题（converse）：若 $x\neq 0$ 且恒等式成立，则条件必定成立。

几何上的意义

记 $\mathbf{u}=(a,b,c)$ 及 $\mathbf{v}=(p,q,r)$。条件可写成 $\mathbf{u}\cdot\mathbf{v}=\frac{x}{2}|\mathbf{v}|^2$，意思是 $\frac{x}{2}\mathbf{v}$ 是从 $\mathbf{u}$ 到经过 O、方向为 $\mathbf{v}$ 的直线的垂足（foot of the perpendicular）。于是向量 $x\mathbf{v}-\mathbf{u}=2\cdot\frac{x}{2}\mathbf{v}-\mathbf{u}$ 就是 $\mathbf{u}$ 关于该直线的反射（reflection）。反射保持长度不变，所以 $|x\mathbf{v}-\mathbf{u}|=|\mathbf{u}|$，而这正是那个恒等式。



验算

以 $x=1$、$a=b=0$、$c=\frac{7}{3}$ 试验 (i)。左边是 $\frac{49}{9}$，右边是 $1+4+\left(\frac{2}{3}\right)^2=5+\frac{4}{9}=\frac{49}{9}$。我们还用精确的有理数运算测试了两部分及一般结果，每项各测试 2000 个随机例子（一般情况最多六个变量），并确认了当 $x\neq 0$ 而条件不成立时，恒等式必定不成立。

常见错误

- 一开始就代入 $c=\frac{1}{3}(7x-a-2b)$。这样做行得通，但代数运算很繁琐，而且掩盖了问题的结构。
- 把一般结果写成“$pa+qb+rc=kx$”，却没有交代 k 与 p,q,r 的关系。恒等式要求 $k=\frac{1}{2}(p^2+q^2+r^2)$。

STEP 之外

这个几何图像正是线性代数（linear algebra）的起点。映射 $\mathbf{u}\mapsto 2\frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}-\mathbf{u}$ 是一个反射，属于正交线性映射（orthogonal linear map），而正交映射恰好就是保持长度不变的线性映射。像这样的投影（projection）和反射，是最小二乘法（least squares）以及数值计算中所用的 Householder 变换（Householder transformations）的基础。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）以同样的几何风格发展这些概念。

详细解答

STEP II 1987，第 2 题

纯数学 · 三角恒等式

> **STEP II 1987，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/87-s2-q2

关键思路

这个式子只涉及 $\sin^2 A$、$\sin^2 B$ 和 $\sin^2 C$，而且它是 $\sin^2 C$ 的二次式（quadratic）。关于 $A \pm B \pm C$ 的条件表示 C 与 $A+B$ 或 $A-B$ 有关，因此自然会猜想这个二次式的两个根（roots）就是 $\sin^2(A+B)$ 和 $\sin^2(A-B)$。验证两根之和与两根之积便能证实这个猜想，结果随之而出。

解答

设 $x = \sin^2 A$、$y = \sin^2 B$、$z = \sin^2 C$，并把左边记作 E。于是

$$E = x^2 + y^2 + z^2 - 2yz - 2zx - 2xy + 4xyz.$$

按 z 的幂次整理，

$$\begin{aligned} E &= z^2 - 2z(x + y - 2xy) + (x^2 - 2xy + y^2) \\ &= z^2 - 2(x + y - 2xy)\,z + (x - y)^2. \end{aligned}$$

二次式的根

我们断言

$$E = \big(z - \sin^2(A+B)\big)\big(z - \sin^2(A-B)\big).$$

由于两边都是 z 的首一（monic）二次式，只需验证 $\sin^2(A+B)$ 与 $\sin^2(A-B)$ 的和与积正确即可。

和。展开 $\sin(A\pm B)=\sin A\cos B\pm\cos A\sin B$ 再平方，相加时交叉项互相抵消：

$$\begin{aligned}\sin^2(A+B)+\sin^2(A-B)&=2\sin^2A\cos^2B+2\cos^2A\sin^2B\\&=2x(1-y)+2(1-x)y\\&=2(x+y-2xy).\end{aligned}$$

积。利用平方差（difference of two squares），

$$\sin(A+B)\sin(A-B)=\sin^2A\cos^2B-\cos^2A\sin^2B=x(1-y)-(1-x)y=x-y,$$

所以 $\sin^2(A+B)\sin^2(A-B)=(x-y)^2$。

两者都与 E 的系数吻合，断言得证。

> **想一想** 为甚么只需知道 $\sin^2C$ 等于 $\sin^2(A+B)$ 或 $\sin^2(A-B)$ 就足够?

因为这样 z 就是两个根之一，于是 E 的其中一个因式为零。余下要证明的，就是角的条件恰好给出这一点。

运用条件

假设对某个整数 k 及某些符号 $\varepsilon_1,\varepsilon_2\in\{1,-1\}$，有 $A+\varepsilon_1B+\varepsilon_2C=k\pi$。那么

$$C=\varepsilon_2\big(k\pi-(A+\varepsilon_1B)\big).$$

函数 $\sin^2$ 是偶函数（even function），且周期（period）为 π，因此符号 ε_2 和 $k\pi$ 都不会改变它的值：

$$\sin^2C=\sin^2(A+\varepsilon_1B).$$

所以 $z=\sin^2(A+B)$（当 $\varepsilon_1=1$）或 $z=\sin^2(A-B)$（当 $\varepsilon_1=-1$）。无论哪种情况，E 都有一个因式为零，因此

$$E=0,$$

即为所求。■

更简洁的形式（及逆命题）

每个因式都是两个正弦平方之差，而 $\sin^2P-\sin^2Q=\sin(P+Q)\sin(P-Q)$（即上面的积公式取 $A=P$、$B=Q$）。所以

$$E=\sin(A+B+C)\,\sin(C-A-B)\,\sin(C+A-B)\,\sin(C-A+B).$$

这里的四个角在不计正负号下，正好是 $A+B+C$、$A+B-C$、$A-B+C$ 和 $A-B-C$。这令结果一目了然：$E=0$ 当且仅当四个角 $A\pm B\pm C$ 之一是 π 的倍数。这也给出了逆命题（converse）。

验算

- 取 $A=\frac{\pi}{6}$、$B=\frac{\pi}{3}$、$C=\frac{\pi}{2}$，于是 $A+B-C=0$。那么 $x=\frac{1}{4}$、$y=\frac{3}{4}$、$z=1$，且 $E=\frac{1}{16}+\frac{9}{16}+1-\frac{3}{2}-\frac{1}{2}-\frac{3}{8}+\frac{3}{4}=0$。
- 取 $A=0.3$、$B=0.7$、$C=1.1$（没有特殊关系），E 和四个正弦之积的值都是 0.05538，可见在条件不成立时，这个式子确实不为零。
- 对 10^5 组随机三元组的测试证实了这个因式分解（factorisation），并证实每当强制 $A\pm B\pm C$ 之一为 π 的倍数时，$E=0$ 的误差在 10^{-14} 以内。

常见错误

- 只处理 $A+B+C=\pi$（三角形）的情况。题目容许四种符号组合中的任何一种，以及 π 的任何倍数；$\sin^2$ 的偶性和以 π 为周期的性质可一并处理所有情况。
- 用公式解二次方程，结果被 $\sqrt{xy(1-x)(1-y)}$ 缠住。验证所猜测的根的和与积，就能完全避开开方和正负号的烦恼。

STEP 之外

式子 $x^2+y^2+z^2-2yz-2zx-2xy$ 在数学中随处可见：把 x,y,z 换成 a^2,b^2,c^2，它就是边长为 a、b、c 的三角形面积平方的 -16 倍（海伦公式（Heron's formula）），它也是一个类似判别式（discriminant）的量，决定三条长度能否构成三角形。把对称多项式（symmetric polynomial）看成其中一个变量的二次式，并猜出它的根，是一种在代数及不变量理论（invariant theory）中反复出现的技巧。想做更多同类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

STEP II 1987，第 3 题

纯数学 · 共轭根式

STEP II 1987，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/87-s2-q3

关键思路

设 $\alpha = a + \sqrt{a^2 - b}$ 及 $\beta = a - \sqrt{a^2 - b}$。两者互为共轭（conjugate）：它们的和 $2a$ 与积 b 都是整数，因此 $\alpha^n + \beta^n$ 必定是整数。若 b 所满足的条件迫使较小的共轭数 β 严格介乎 0 与 1 之间，那么 β^n 正是我们所需的小数部分（fractional part），其余结论便随之而来。

解答

在整个解答中，设

$$\alpha = a + \sqrt{a^2 - b}, \qquad \beta = a - \sqrt{a^2 - b}, \qquad \text{所以} \qquad \alpha + \beta = 2a, \quad \alpha\beta = a^2 - (a^2 - b) = b.$$

β 的位置。 首先，这个平方根是实数。由于 $b < 2a - 1$，

$$a^2 - b > a^2 - 2a + 1 = (a - 1)^2 \geq 0.$$

对这条严格不等式（inequality）两边取正平方根，得 $\sqrt{a^2 - b} > a - 1$。另外 $b > 0$，所以 $a^2 - b < a^2$，从而 $\sqrt{a^2 - b} < a$。综合两者，

$$a - 1 < \sqrt{a^2 - b} < a \quad \Longrightarrow \quad 0 < \beta < 1.$$

（这个条件是最佳的：若 $b = 2a - 1$，则 $\sqrt{a^2 - b} = a - 1$，而 $\beta = 1$ 恰好成立。）

第 (i) 部分

由于 $0 < \beta < 1$，对每个正整数 n 都有 $0 < \beta^n < 1$。现在 M 是整数，且 $\beta^n = M + s$，其中 $0 \leq s < 1$，所以 M 是 β^n 的整数部分（integer part），即 0。因此

$$M = 0 \qquad \text{及} \qquad s = \beta^n.$$

第 (ii) 部分

设 $S_n = \alpha^n + \beta^n$。我们断言 S_n 是整数。设 $d = \sqrt{a^2 - b}$，用二项式定理（binomial theorem）展开两个幂，

$$S_n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\left(d^k + (-d)^k\right) = 2 \sum_{k \text{ 偶数}} \binom{n}{k} a^{n-k}(a^2 - b)^{k/2},$$

因为 k 为奇数的项互相抵消。余下的每一项都是整数，所以 $S_n \in \mathbb{Z}$。（等价地，α, β 是 $t^2 - 2at + b = 0$ 的根，所以 $S_n = 2aS_{n-1} - bS_{n-2}$，其中 $S_0 = 2$ 及 $S_1 = 2a$。）

因此

$$\alpha^n = S_n - \beta^n, \qquad S_n \in \mathbb{Z}, \quad 0 < \beta^n < 1.$$

? **想一想** 为甚么这就迫使 $N = S_n$ 及 $r = \beta^n$?

把 $\alpha^n = N - r$ 表示成 N 为整数且 $0 \le r < 1$ 的形式，这种表示是唯一的：N 必须是不小于 α^n 的最小整数。若 $N' - r' = N - r$，且 $r, r' \in [0, 1)$，则 $N - N' = r - r'$ 是一个严格介乎 -1 与 1 之间的整数，所以它等于 0。我们的表达式 $S_n - \beta^n$ 正是这种形式，所以

$$N = S_n = \alpha^n + \beta^n, \qquad r = \beta^n = s.$$

第 (iii) 部分

利用 $r = \beta^n$、$N = \alpha^n + \beta^n$ 及 $(\alpha\beta)^n = b^n$，

$$\begin{aligned} r^2 - Nr + b^n &= \beta^{2n} - (\alpha^n + \beta^n)\beta^n + \alpha^n\beta^n \\ &= \beta^{2n} - \alpha^n\beta^n - \beta^{2n} + \alpha^n\beta^n \\ &= 0, \end{aligned}$$

得证。换句话说，r 是 $t^2 - Nt + b^n = 0$ 较小的根，而这方程的两根是 α^n 和 β^n。

最后部分

取 $a = 8$ 及 $a^2 - b = 63$，所以 $b = 1$，且 $\sqrt{63} = 3\sqrt{7}$。条件 $b < 2a - 1 = 15$ 成立，所以以上所有结果都适用于 $\alpha = 8 + 3\sqrt{7}$:

$$\left(8 + 3\sqrt{7}\right)^n = N - r, \qquad N \in \mathbb{Z},$$

而 α^n 比整数 N 小 r。在第 (iii) 部分中取 $b = 1$，得 $r(N - r) = 1$，即

$$r = \frac{1}{N - r} = \frac{1}{\left(8 + 3\sqrt{7}\right)^n} = \left(8 - 3\sqrt{7}\right)^n.$$

特别地，$r \neq 0$，所以这个幂本身永远不是整数，而且 $r \to 0$。现在 $3\sqrt{7} = \sqrt{63}$ 略小于 $\sqrt{64} = 8$，所以

$$8 + 3\sqrt{7} \approx 16 = 2^4, \qquad \text{因此} \qquad r = \left(8 + 3\sqrt{7}\right)^{-n} \approx 16^{-n} = 2^{-4n}.$$

所以当 n 很大时，$\left(8 + 3\sqrt{7}\right)^n$ 与某个整数相差大约 2^{-4n}，并且略小于该整数。

这个“大约”有多准确？由于 $8 + 3\sqrt{7} \approx 15.937$，我们有 $r = 2^{-cn}$，其中 $c = \log_2(8 + 3\sqrt{7}) \approx 3.994$。这个指数与 $4n$ 的差距在 0.2% 以内，估算正是在这个意义下成立。比值 $r/2^{-4n} = \left(16/(8 + 3\sqrt{7})\right)^n$ 缓慢上升，由 $n = 1$ 时约 1.004 升至 $n = 10$ 时的 1.04。

验算

- $n = 1$：$8 + 3\sqrt{7} \approx 15.93725$，即 $16 - 0.06275$，而 $8 - 3\sqrt{7} \approx 0.06275$，与预测相符。此外，$r^2 - 16r + 1 = 0$ 较小的根是 $8 - \sqrt{63}$。
- $n = 2$：$(8 + 3\sqrt{7})^2 = 127 + 48\sqrt{7} \approx 253.99606$，所以 $N = 254 = S_2 = 2 \cdot 8 \cdot 16 - 2$，而 $r \approx 0.003937$，对比 $2^{-8} \approx 0.003906$。
- 以电脑对数以千计满足 $b < 2a - 1$ 的随机数对 (a, b) 及 $n \leq 7$ 进行验算，确认了 $0 < \beta^n < 1$、$N - \alpha^n = \beta^n$ 及 $r^2 - Nr + b^n = 0$。对 $a = 8$、$b = 1$、$n \leq 10$ 进行的高精度验算确认了 $N - \alpha^n = (8 - 3\sqrt{7})^n$；当 $n = 10$ 时，此值为 9.46×10^{-13}，对比 $2^{-40} \approx 9.09 \times 10^{-13}$。

常见错误

- 只证明了 $\beta < 1$，却忘记 $\beta > 0$（这需要 $b > 0$）。两者都是 $M = 0$ 所必需的。
- 断言 $N = \alpha^n + \beta^n$，却没有说明为甚么 $N - r$ 中的整数部分与小数部分是唯一的。
- 以为 $r/2^{-4n} \to 1$。事实并非如此：近似只是在指数上成立，因为 $8 + 3\sqrt{7}$ 接近 16，但不等于 16。

STEP 之外

数 $\alpha^n + \beta^n$ 是迹（trace）的一个例子：α 是一个代数整数（algebraic integer），而它的共轭数位于单位区间之内，所以它的幂以指数速度趋近整数。这类数称为 Pisot 数（Pisot numbers）；黄金比例（golden ratio）是另一个例子，这就是为甚么 φ^n 接近卢卡斯数（Lucas number）L_n。这里的情况亦与佩尔方程（Pell's equation）密切相关：$8^2 - 7 \cdot 3^2 = 1$，所以 $(8, 3)$ 是 $x^2 - 7y^2 = 1$ 的最小正整数解。Davenport 的 *The Higher Arithmetic*（《高等算术》）对佩尔方程及连分数（continued fractions）有精彩的论述。

详细解答

STEP II 1987，第 4 题

纯数学 · 复数与几何

> **STEP II 1987，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/87-s2-q4

关键思路

乘以 $e^{i\theta}$ 就是一个旋转（rotation）。把三角形的一条边绕某个顶点旋转 $60°$，得到的正是等边三角形（equilateral triangle）。这样，几何条件就化为关于 α, β, γ 的线性方程（linear equation）。若容许两种定向（orientation），条件就变成两个线性表达式中有一个为零；两者之积正是题目中的对称（symmetric）二次式。由于它是对称的，所以可以用三次方程（cubic）的系数（coefficients）来表示。

解答

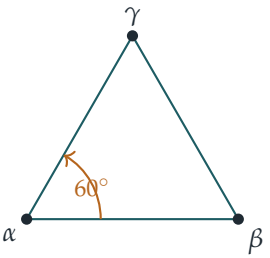
全文中设 $\omega = e^{2\pi i/3}$，于是 $\omega^3 = 1$、$\bar{\omega} = \omega^2$ 及 $1 + \omega + \omega^2 = 0$。

乘以 $e^{i\theta}$

若 $z = re^{i\phi}$，则 $ze^{i\theta} = re^{i(\phi+\theta)}$：模（modulus）不变，辐角（argument）增加 θ。因此，点 $ze^{i\theta}$ 就是点 z 绕原点逆时针（anticlockwise）旋转角 θ 后的像。

更一般地，$\gamma - \alpha = e^{i\theta}(\beta - \alpha)$ 表示由 α 到 γ 的向量，等于由 α 到 β 的向量逆时针旋转 θ 所得；换言之，γ 是 β 绕 α 旋转 θ 后的像。

逆时针方向的等边三角形



相异的点 α, β, γ 按逆时针次序构成等边三角形，当且仅当 γ 是 β 绕 α 逆时针旋转 $\frac{\pi}{3}$ 所得，即

$$\gamma - \alpha = e^{i\pi/3}(\beta - \alpha).$$

（必要性（necessity）：在这样的三角形中，$|\alpha\gamma| = |\alpha\beta|$，而且在 α 处的角为 $\frac{\pi}{3}$，由 $\alpha\beta$ 逆时针量度至 $\alpha\gamma$。充分性（sufficiency）：该方程给出 $|\alpha\gamma| = |\alpha\beta|$，两者夹角为 $\frac{\pi}{3}$，所以三角形是顶角为 $60°$ 的等腰（isosceles）三角形，因而是等边三角形，而且定向是逆时针的。）

把条件写成对称的形式会更简洁。由于 $e^{i\pi/3} = -\omega^2$（因为 $-\omega^2 = e^{i\pi}e^{4\pi i/3} = e^{i\pi/3}$，这里用了 $e^{7\pi i/3} = e^{i\pi/3}$），条件变为

$$\begin{aligned} & \gamma - \alpha = -\omega^2\beta + \omega^2\alpha \\ \iff \quad & \gamma + \omega^2\beta - (1+\omega^2)\alpha = 0 \\ \iff \quad & \gamma + \omega^2\beta + \omega\alpha = 0, \end{aligned}$$

其中用了 $1 + \omega^2 = -\omega$。两边乘以 ω^2（它不为零，所以这一步可逆），得到等价条件

$$\alpha + \omega\beta + \omega^2\gamma = 0.$$

简单验证一下：$1, \omega, \omega^2$ 是一个等边三角形按逆时针次序的顶点，而 $1 + \omega \cdot \omega + \omega^2 \cdot \omega^2 = 1 + \omega^2 + \omega = 0$。

任一定向

任何三个构成等边三角形的相异点，若不是按 α, β, γ 的次序为逆时针，就是按 α, γ, β 的次序为逆时针（循环重排不会改变定向）。由上一部分，对相异的 α, β, γ，

$$\begin{aligned} \text{等边} & \iff \alpha + \omega\beta + \omega^2\gamma = 0 \text{ 或 } \alpha + \omega\gamma + \omega^2\beta = 0 \\ & \iff (\alpha + \omega\beta + \omega^2\gamma)(\alpha + \omega^2\beta + \omega\gamma) = 0. \end{aligned}$$

? **想一想** 怎样看出这个乘积正是我们想要的表达式?

把它展开就可以了。平方项给出 $\alpha^2+\omega^3\beta^2+\omega^3\gamma^2=\alpha^2+\beta^2+\gamma^2$。交叉项为

$$\begin{aligned}\alpha\beta(\omega^2+\omega)+\alpha\gamma(\omega+\omega^2)+\beta\gamma(\omega^2+\omega^4) &= \alpha\beta(\omega+\omega^2)+\gamma\alpha(\omega+\omega^2)+\beta\gamma(\omega^2+\omega)\\ &= -\alpha\beta-\gamma\alpha-\beta\gamma,\end{aligned}$$

其中用了 $\omega^4=\omega$ 及 $\omega+\omega^2=-1$。因此

$$(\alpha+\omega\beta+\omega^2\gamma)(\alpha+\omega^2\beta+\omega\gamma)=\alpha^2+\beta^2+\gamma^2-\beta\gamma-\gamma\alpha-\alpha\beta,$$

所以相异的点 α,β,γ（按某个次序）构成等边三角形，当且仅当

$$\alpha^2+\beta^2+\gamma^2-\beta\gamma-\gamma\alpha-\alpha\beta=0,$$

得证。

值得留意的是，若点不一定相异，情况会怎样：例如若 $\alpha=\beta$，表达式变为 $(\alpha-\gamma)^2$，只有当 $\gamma=\alpha$ 时才为零。所以表达式为零，恰好是在三点构成等边三角形或三点重合的时候。下文会用到这一点。

三次方程

设 $z^3+az^2+bz+c=0$ 的根为 α,β,γ。则

$$\alpha+\beta+\gamma=-a,\qquad \beta\gamma+\gamma\alpha+\alpha\beta=b,\qquad \alpha\beta\gamma=-c.$$

该表达式是对称的，所以可以用这些量来表示：

$$\begin{aligned}\alpha^2+\beta^2+\gamma^2-(\beta\gamma+\gamma\alpha+\alpha\beta) &= (\alpha+\beta+\gamma)^2-3(\beta\gamma+\gamma\alpha+\alpha\beta)\\ &= a^2-3b.\end{aligned}$$

根据上面的说明，$a^2=3b$ 恰好在各根构成等边三角形或全部相等时成立。三个根都等于 t，即 $z^3+az^2+bz+c=(z-t)^3$，所以 $a=-3t$、$b=3t^2$、$c=-t^3$；也就是 $b=\frac{a^2}{3}$ 及 $c=\frac{a^3}{27}$。因此，各根位于一个（真正的）等边三角形的顶点，当且仅当

$$\boxed{a^2=3b\quad\text{且}\quad 27c\neq a^3}$$

若愿意把三点重合视为退化的等边三角形，条件就只是 $a^2=3b$。

有一个漂亮的方法可以直接看出这一点。代入 $z=w-\frac{a}{3}$ 可消去 z^2 项；当 $a^2=3b$ 时，一次项亦会消失，方程变为 $w^3=\frac{a^3}{27}-c$。它的根是一个非零数的三个立方根，构成以 $w=0$ 为中心的等边三角形，即以 $z=-\frac{a}{3}$ 为中心，而这正是各根的形心（centroid）。

验算

- $z^3-1=0$（$a=b=0$, $c=-1$）的根为 $1,\omega,\omega^2$：构成等边三角形，而确实 $a^2=0=3b$ 且 $27c=-27\neq 0$。

- $z^3 - z = 0$ 的根为 $0, \pm 1$，三点共线（collinear）；此时 $a^2 - 3b = 3 \neq 0$。
- 以数值方法，对 1000 个随机三角形 α、β、$\gamma = \alpha + e^{i\pi/3}(\beta - \alpha)$，我们验证了三角形为逆时针方向的等边三角形，$\alpha + \omega\beta + \omega^2\gamma$ 及该二次表达式（在每种次序下）都为零，而且对随机复数，乘积恒等式及 $a^2 - 3b$ 公式在 10^{-8} 的误差内成立。

常见错误

- 弄错正负号或定向：$e^{i\pi/3}$ 是 $-\omega^2$，而不是 ω；绕一个顶点乘以 ω 旋转（即旋转 120°）并不会得到等边三角形。
- 忘记三重根的情况：单凭 $a^2 = 3b$ 亦容许 $z^3 + az^2 + bz + c = (z + \frac{a}{3})^3$，此时三个“顶点”重合。

STEP 之外

条件 $\alpha + \omega\beta + \omega^2\gamma = 0$ 是三元组 (α, β, γ) 的离散傅里叶变换（discrete Fourier transform）的一个分量；而消去 z^2 项的技巧，是以根式解（solving by radicals）三次方程的第一步，由此可通往伽罗瓦理论（Galois theory）。从几何角度看，这个结果表明：各根构成等边三角形，恰好是导数（derivative）$3z^2 + 2az + b$ 在形心处有二重根（double root）的时候；这是关于导数之根的马登定理（Marden's theorem）的一个特例。Needham 的 *Visual Complex Analysis*（《可视化复分析》）是进一步探索复数乘法几何意义的绝佳读物。

详细解答

STEP II 1987，第 5 题

纯数学 · 拉格朗日反演级数

> **STEP II 1987，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/87-s2-q5

关键思路

拉格朗日（Lagrange）恒等式（Lagrange's identity）就像一部机器：把函数 $\phi(y) = y - \mathrm{f}(y)$ 放进去，取其 n 次方，求导 $n-1$ 次，再把结果加起来。所有功夫都在于把问题设置好，使 ϕ 尽量简单，最好是 y 的单一幂次或单一指数函数，因为这样 ϕ^n 的 $(n-1)$ 阶导数（derivative）就能一下子写出来。在每一部分中，我们都选取适当的 f，使所要解的方程写成 $\mathrm{f}(x) = y$ 的形式，其中 y 取一个方便的值。

解答

全文设 $\phi(y) = y - \mathrm{f}(y)$，于是恒等式可写成

$$\mathrm{f}^{-1}(y) = y + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} \left[\phi(y)^n\right].$$

注意 $n = 1$ 那一项涉及的是零阶导数，也就是 $\phi(y)$ 本身。

我们会反复用到一个事实：对整数 $m \geq n-1$，

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} y^m = \frac{m!}{(m-n+1)!} y^{m-n+1},$$

因为每求导一次，就把当时的幂次乘到前面，并把幂次减一。

第 (i) 部分

取 $\mathrm{f}(x) = \alpha x$，则 $\phi(y) = (1-\alpha)y$，所以 $\phi(y)^n = (1-\alpha)^n y^n$。在上式中取 $m = n$，求导 $n-1$ 次得

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}} \left[(1-\alpha)^n y^n\right] = (1-\alpha)^n \, n! \, y.$$

$n!$ 与 $\frac{1}{n!}$ 互相抵消，所以恒等式的右方为

$$y+\sum_{n=1}^{\infty}(1-\alpha)^n y = y\sum_{n=0}^{\infty}(1-\alpha)^n.$$

这是一个公比为 $1-\alpha$ 的等比级数（geometric series）。由于 $0<\alpha<2$，有 $-1<1-\alpha<1$，因此级数收敛（converges），其和为

$$y\cdot\frac{1}{1-(1-\alpha)} = \frac{y}{\alpha}.$$

另一方面，由 $y=\alpha x$ 直接可得 $x=y/\alpha$，所以 $\mathrm{f}^{-1}(y)=y/\alpha$。两者一致，于是恒等式在这情况下得到验证。条件 $0<\alpha<2$ 正正就是级数收敛的条件。

第 (ii) 部分

方程为 $\mathrm{f}(x)=\frac{1}{2}$，其中

$$\mathrm{f}(x)=x-\tfrac{1}{4}x^3, \qquad \text{所以} \qquad \phi(y)=y-\mathrm{f}(y)=\tfrac{1}{4}y^3.$$

于是 $\phi(y)^n=4^{-n}y^{3n}$，在导数公式中取 $m=3n$（它不小于 $n-1$），得

$$\frac{1}{n!}\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}}\left[4^{-n}y^{3n}\right] = \frac{(3n)!}{n!\,(2n+1)!\,4^n}\,y^{2n+1}.$$

因此拉格朗日恒等式给出

$$\mathrm{f}^{-1}(y)=y+\sum_{n=1}^{\infty}\frac{(3n)!}{n!\,(2n+1)!\,4^n}\,y^{2n+1}.$$

开头的 y 恰好就是通项在 $n=0$ 时的值，因为 $\frac{0!}{0!\,1!\,4^0}\,y^1=y$。所以可以把它并入总和，由 $n=0$ 开始：

$$\mathrm{f}^{-1}(y)=\sum_{n=0}^{\infty}\frac{(3n)!}{n!\,(2n+1)!\,4^n}\,y^{2n+1}.$$

现在代入 $y=\frac{1}{2}$。由于 $4^n\cdot 2^{2n+1}=2^{2n}\cdot 2^{2n+1}=2^{4n+1}$，得

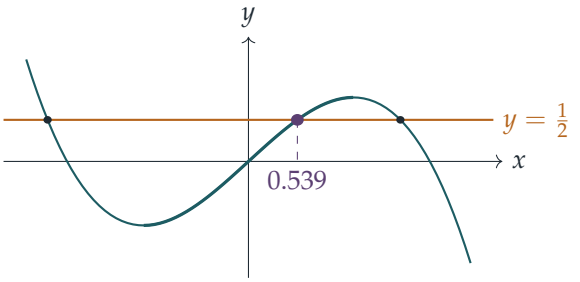
$$x=\mathrm{f}^{-1}\left(\tfrac{1}{2}\right)=\sum_{n=0}^{\infty}\frac{(3n)!}{n!\,(2n+1)!\,2^{4n+1}},$$

又因 $\mathrm{f}(x)=\frac{1}{2}$，这个 x 是 $\frac{1}{2}=x-\frac{1}{4}x^3$ 的一个根，得证。

? **想一想** 这个三次方程有三个实根。级数给出的是哪一个?

函数 f 在整条实数线上并非一一对应（one-to-one），所以“f^{-1}”指的是级数所描述的那一支。这个级数是 y 的幂级数（power series），以 $y+\cdots$ 开始，所以它给出 $\mathrm{f}^{-1}(0)=0$，而且当 y 很小时 $\mathrm{f}^{-1}(y)\approx y$。它是 f 限制在通过原点的递增部分上的反函数，在那里

$\mathrm{f}'(x) = 1 - \frac{3}{4}x^2 > 0$，即 $|x| < 2/\sqrt{3}$。首几项为 $\frac{1}{2} + \frac{1}{32} + \frac{3}{512} + \cdots \approx 0.537$，而总和准确至四位小数为 0.5392：这是最小的正根，已在下图标出。另外两个根约为 1.675 和 -2.214，位于曲线的递减部分。



曲线上较粗的一段是通过原点的递增分支；级数给出的就是这分支上的根，以梅红色标示。

第 (iii) 部分

这里的窍门是把整条方程化成 $\mathrm{f}(x) = y$ 的形式，并使 ϕ 简单。把 $x = \mathrm{e}^{\lambda x}$ 改写为

$$x - \mathrm{e}^{\lambda x} = 0,$$

我们取

$$\mathrm{f}(x) = x - \mathrm{e}^{\lambda x}, \qquad \phi(y) = y - \mathrm{f}(y) = \mathrm{e}^{\lambda y},$$

而所求的是 $x = \mathrm{f}^{-1}(0)$。此时 $\phi(y)^n = \mathrm{e}^{n\lambda y}$，每求导一次就乘以 $n\lambda$，所以

$$\frac{\mathrm{d}^{n-1}}{\mathrm{d}y^{n-1}}\,\mathrm{e}^{n\lambda y} = (n\lambda)^{n-1}\mathrm{e}^{n\lambda y}.$$

（当 $n = 1$ 时，这给出 $\mathrm{e}^{\lambda y}$，正确无误。）拉格朗日恒等式给出

$$\mathrm{f}^{-1}(y) = y + \sum_{n=1}^{\infty} \frac{(n\lambda)^{n-1}}{n!}\,\mathrm{e}^{n\lambda y},$$

代入 $y = 0$，

$$x = \sum_{n=1}^{\infty} \frac{n^{n-1}}{n!}\,\lambda^{n-1} = 1 + \lambda + \frac{3}{2}\lambda^2 + \frac{8}{3}\lambda^3 + \frac{125}{24}\lambda^4 + \cdots.$$

等价地，把指标平移，可得 $x = \sum_{n=0}^{\infty} \frac{(n+1)^{n-1}}{n!}\lambda^n$。

这个级数在 λ 取甚么值时收敛？相邻两项之比为

$$\frac{(n+1)^n\lambda^n/(n+1)!}{n^{n-1}\lambda^{n-1}/n!} = \lambda\left(1 + \frac{1}{n}\right)^{n-1} \to \mathrm{e}\lambda,$$

所以由比值检验法（ratio test），级数在 $|\lambda| < 1/\mathrm{e}$ 时收敛。这是一个自然的界线：$y = x$ 与 $y = \mathrm{e}^{\lambda x}$ 相切，恰好发生在 $\lambda = 1/\mathrm{e}$ 时（切点在 $x = \mathrm{e}$），而当 $\lambda > 1/\mathrm{e}$ 时根本没有实根。当 $0 < \lambda < 1/\mathrm{e}$ 时有两个正根，级数给出较小的那一个，即当 $\lambda \to 0$ 时趋向 1 的那个根。

验算

- 在 (ii) 中，首四项为 $\frac{1}{2}$、$\frac{1}{32}$、$\frac{3}{512}$ 和 $\frac{3}{2048}$。以数值方法加总 200 项得 0.539189，而对 $x^3 - 4x + 2 = 0$ 使用牛顿法（Newton's method）得根 0.5391888…，两者在机器精度内一致。
- 在 (iii) 中，代入 $\lambda = 0$：级数给出 $x = 1$，而确实 $x = \mathrm{e}^0 = 1$。取 $\lambda = 0.2$ 加总级数得 $x \approx 1.29586$，而 $\mathrm{e}^{0.2 \times 1.29586} = 1.29586$。取 $\lambda = -0.3$ 则得 0.78918，同样满足方程。当 $\lambda = 0.2$ 时，另一个根约为 12.71，与级数的值相距甚远。
- 在 (i) 中，试取 $\alpha = 1$：此时 $\phi = 0$，总和中每一项都是零，于是 $\mathrm{f}^{-1}(y) = y$，理应如此。

常见错误

- 忘记 $n = 1$ 那一项不用求导（即零阶导数），因而漏掉或弄错这一项。
- 在 (ii) 中，没有察觉单独的 y 正是通项公式在 $n = 0$ 时的项，因而未能凑成由 $n = 0$ 开始的总和。
- 在 (iii) 中，尝试取 $\mathrm{f}(x) = x\mathrm{e}^{-\lambda x}$ 之类的函数。这样做并无不可，但 ϕ 便成了两项之差，导数会变得一团糟。选取 $\phi = \mathrm{e}^{\lambda y}$ 则令导数变得轻而易举。

STEP 之外

这就是拉格朗日反演定理（Lagrange inversion theorem），通常用复分析（complex analysis）（柯西（Cauchy）留数定理）或形式幂级数（formal power series）来证明。(iii) 中的系数 $(n+1)^{n-1}/n!$ 属于 Lambert W 函数，而 n^{n-1} 是 n 个顶点上的有根标号树（rooted labelled trees）的数目。同样地，(ii) 中的数 $\frac{(3n)!}{n!\,(2n+1)!}$ 计算的是三叉树（ternary trees）的数目，与卡特兰数（Catalan numbers）是近亲。因此，拉格朗日反演是枚举组合学（enumerative combinatorics）的核心工具。Needham 的 *Visual Complex Analysis*（《直观复分析》）提供复分析方面的背景，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）则有更多关于级数的 STEP 练习。

详细解答

STEP II 1987，第 6 题

纯数学 · 利用对称性求定积分

> **STEP II 1987，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/87-s2-q6

关键思路

区间 $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ 是对称的，所以反射 $\theta \mapsto -\theta$ 是自然的第一步：它把 $\sin\theta$ 变成 $-\sin\theta$，第二个形式便不费吹灰之力得到。把两个形式取平均，可消去奇部分（odd part），剩下分母只含 $\sin^2\theta$，再稍作代数处理，就能拆成一个常数加上一个标准的 $\tan\theta$ 积分（integral）。最后一部分的重点在于一个看似无害的步骤：$\sqrt{\cos^2 2\alpha}$ 是 $|\cos 2\alpha|$，而不是 $\cos 2\alpha$。

解答

全文记 $s = \sin 2\alpha$ 及 $c = \cos 2\alpha$，因此 $s^2 + c^2 = 1$。当 $0 < \alpha < \frac{1}{4}\pi$ 时，有 $0 < s < 1$ 及 $0 < c < 1$。由于 $s < 1$，分母 $1 \mp s\sin\theta$ 在整个区间上都是正数，所以所有积分都是常义积分（proper integral）。

I 的第二个形式

代入 $\theta = -\phi$，于是 $\mathrm{d}\theta = -\mathrm{d}\phi$，而上下限互换：

$$I = \int_{\frac{1}{2}\pi}^{-\frac{1}{2}\pi} \frac{\cos^2(-\phi)}{1 - \sin(-\phi)\,s}\,(-\mathrm{d}\phi) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\cos^2\phi}{1 + s\sin\phi}\,\mathrm{d}\phi,$$

这里用了 $\cos(-\phi) = \cos\phi$ 及 $\sin(-\phi) = -\sin\phi$。把 ϕ 改名为 θ，即得所求形式。

含 $\cot^2 2\alpha$ 的表达式

把 I 的两个表达式相加，

$$2I = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos^2\theta \left(\frac{1}{1 - s\sin\theta} + \frac{1}{1 + s\sin\theta}\right)\mathrm{d}\theta = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{2\cos^2\theta}{1 - s^2\sin^2\theta}\,\mathrm{d}\theta,$$

所以

$$I = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\cos^2\theta}{1 - s^2\sin^2\theta}\,\mathrm{d}\theta.$$

现在利用 $1 = \cos^2\theta + \sin^2\theta$ 及 $1 - s^2 = c^2$ 改写分母：

$$1 - s^2\sin^2\theta = \cos^2\theta + c^2\sin^2\theta.$$

? **想一想** 怎样把 $\dfrac{\cos^2\theta}{\cos^2\theta + c^2\sin^2\theta}$ 拆成一个常数加上一个我们能积分的部分?

目标是把分子写成分母的倍数加上一个余项，而这个余项除以 $\cos^2\theta$ 之后不再依赖 θ。由于

$$(\cos^2\theta + c^2\sin^2\theta) - c^2(\cos^2\theta + \sin^2\theta) = (1 - c^2)\cos^2\theta = s^2\cos^2\theta,$$

两边除以 $s^2(\cos^2\theta + c^2\sin^2\theta)$（因为 $s \neq 0$，这是容许的），得

$$\frac{\cos^2\theta}{\cos^2\theta + c^2\sin^2\theta} = \frac{1}{s^2} - \frac{c^2}{s^2}\cdot\frac{1}{\cos^2\theta + c^2\sin^2\theta} = \frac{1}{s^2} - \frac{c^2}{s^2}\cdot\frac{\sec^2\theta}{1 + c^2\tan^2\theta},$$

其中最后一步把分子分母同除以 $\cos^2\theta$。在长度为 π 的区间上积分，并利用 $c^2/s^2 = \cot^2 2\alpha$，得

$$I = \frac{\pi}{\sin^2 2\alpha} - \cot^2 2\alpha \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\sec^2\theta}{1 + \cos^2 2\alpha\tan^2\theta}\,\mathrm{d}\theta,$$

此即所求。

求 I 的值

把余下的积分记为 J。代入 $t = \tan\theta$，于是 $\mathrm{d}t = \sec^2\theta\,\mathrm{d}\theta$；当 θ 由 $-\frac{1}{2}\pi$ 走到 $\frac{1}{2}\pi$ 时，t 走遍整条实数线。于是，当 $c \neq 0$ 时，

$$J = \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{1 + c^2t^2} = \left[\frac{1}{|c|}\arctan\big(|c|\,t\big)\right]_{-\infty}^{\infty} = \frac{\pi}{|c|}.$$

（$|c|$ 正是由代换 $u = |c|t$ 产生的：它使上下限仍按正确次序由 $-\infty$ 到 ∞。）

当 $0 < \alpha < \frac{1}{4}\pi$ 时，$c > 0$，所以 $J = \pi/c$，而

$$I = \frac{\pi}{s^2} - \frac{c^2}{s^2}\cdot\frac{\pi}{c} = \frac{\pi(1-c)}{1-c^2} = \frac{\pi}{1+c} = \frac{\pi}{1+\cos 2\alpha} = \frac{\pi}{2\cos^2\alpha} = \tfrac{1}{2}\pi\sec^2\alpha,$$

这里约去了 $1 - c \neq 0$，并用了 $1 + \cos 2\alpha = 2\cos^2\alpha$。这就是所求结果。

$\frac{1}{4}\pi < \alpha < \frac{1}{2}\pi$ 的情况

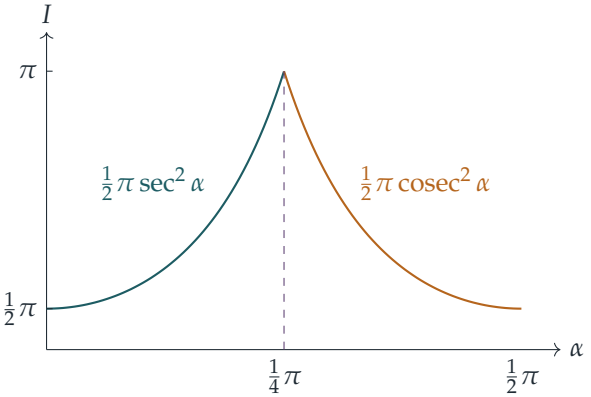
想一想 当 α 越过 $\frac{1}{4}\pi$ 时，论证中有没有任何地方会改变?

几乎没有：我们仍然有 $0 < s < 1$，所以直至以 J 表示 I 的公式为止，每一步都照样成立。唯一的分别是现在 $c = \cos 2\alpha < 0$，所以 $|c| = -c$，而 $J = -\pi/c$。因此

$$I = \frac{\pi}{s^2} + \frac{c^2}{s^2}\cdot\frac{\pi}{c} = \frac{\pi(1+c)}{1-c^2} = \frac{\pi}{1-c} = \frac{\pi}{2\sin^2\alpha} = \tfrac{1}{2}\pi\operatorname{cosec}^2\alpha,$$

这里用了 $1 - \cos 2\alpha = 2\sin^2\alpha$。若盲目照搬第一个答案，便会得出错误的值。

I 对 α 的图像显示两条公式在 $\alpha = \frac{1}{4}\pi$ 处相接，那里两者都等于 π：



验算

- 当 $\alpha = \frac{1}{4}\pi$ 时，$s = 1$，被积函数（integrand）为 $\cos^2\theta/(1-\sin\theta) = 1+\sin\theta$，其积分为 π。两条公式在该处都给出 π，理应如此。
- 当 $\alpha = \frac{1}{6}\pi$ 时：$s = \frac{\sqrt{3}}{2}, c = \frac{1}{2}$，所以 $I = \pi/(1+c) = \frac{2}{3}\pi$，而 $\frac{1}{2}\pi\sec^2\frac{1}{6}\pi = \frac{1}{2}\pi\cdot\frac{4}{3} = \frac{2}{3}\pi$。
- 在 $\frac{1}{4}\pi$ 两侧随机取十个 α 值，用辛普森法则（Simpson's rule）计算原积分，与公式的结果相差少于 10^{-6}；例如 $\alpha \approx 0.6308$ 时，两种方法都得出 $I \approx 2.408759$。
- 当 $\alpha \to 0$ 时，被积函数趋向 $\cos^2\theta$，其积分为 $\frac{1}{2}\pi$，与 $\frac{1}{2}\pi\sec^2 0$ 吻合。

常见错误

- 写成 $\int_{-\infty}^{\infty}\frac{\mathrm{d}t}{1+c^2t^2} = \frac{\pi}{c}$ 而漏了绝对值（modulus）。当 $c < 0$ 时，这是负数，而被积函数为正，这是不可能的。

- 遗漏了 $\mathrm{d}\theta = -\mathrm{d}\phi$ 中的负号，或忘记了它会因上下限互换而抵消。
- 担心 $t = \tan\theta$ 在 $\theta = \pm\frac{1}{2}\pi$ 处无定义：关于 t 的积分是反常积分（improper integral），但它收敛（converges），而这已足够。

STEP 之外

答案中整齐的对称 $\alpha \leftrightarrow \frac{1}{2}\pi - \alpha$ 并非巧合：I 只通过 $\sin 2\alpha$ 依赖 α，而它在这个互换下不变；但像 $\frac{1}{2}\pi\sec^2\alpha$ 这样的公式却会改变。默默地平方再开方的代数运算，正好会掩盖这类问题。$\sin\theta$ 和 $\cos\theta$ 的有理函数在一个完整周期上的积分（例如我们取平均后的被积函数，其周期为 π），在大学里通常用沿单位圆的围道积分（contour integration）来计算，而留数定理（residue theorem）使这类计算几乎变得机械化。如想看一个优美而富几何直观的入门介绍，可参阅 Needham 的 *Visual Complex Analysis*（《直观复分析》）。

详细解答

STEP II 1987，第 7 题

纯数学 · 梯形法则与阶乘

> **STEP II 1987，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/87-s2-q7

关键思路

以单位步长对 $\ln x$ 求梯形和（trapezium sum），结果几乎正好是 $\ln n!$，因为所加的高度正是 $\ln 1, \ln 2, \ldots, \ln n$。由于 $\ln x$ 是凹函数（concave），每一条弦都在曲线下方，所以梯形法则（trapezium rule）总是低估积分；单凭这一个事实，就能决定每条不等式的方向。至于最后一部分，技巧是把细网格上的误差重新缩放，使之变成另一个区间上的单位步长误差，从而可以直接与 $\mathrm{g}(kn)$ 背后的误差比较。

解答

说明法则的依据

把 $[x_0, x_N]$ 分成 N 条狭条 $[x_r, x_{r+1}]$，每条宽度为 h。在每条狭条上，以连接 $(x_r, \mathrm{f}(x_r))$ 与 $(x_{r+1}, \mathrm{f}(x_{r+1}))$ 的直线(弦(chord))代替曲线。弦下方的区域是一个梯形,面积为 $\frac{1}{2}h\{\mathrm{f}(x_r) + \mathrm{f}(x_{r+1})\}$。对 $r = 0, 1, \ldots, N-1$ 求和，每个内部纵坐标（ordinate）$\mathrm{f}(x_1), \ldots, \mathrm{f}(x_{N-1})$ 都在两个相邻梯形中出现，而两端的纵坐标各出现一次，于是得到题中的公式。当 h 很小时，弦与曲线相当接近，所以近似效果良好。

单位步长与 $n! \approx \mathrm{g}(n)$

取 $u = \ln x$、$\mathrm{d}v = \mathrm{d}x$ 作分部积分（integration by parts），

$$I_n = \int_1^n \ln x \, \mathrm{d}x = \Big[x \ln x\Big]_1^n - \int_1^n 1 \, \mathrm{d}x = n \ln n - n + 1.$$

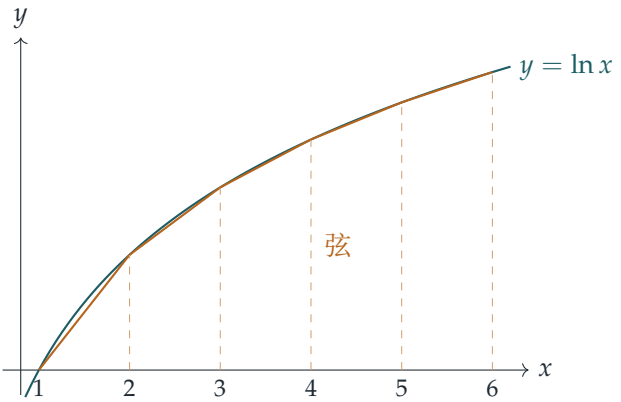
当 $h=1$ 时，纵坐标位于 $x=1,2,\ldots,n$，所以梯形和为

$$\begin{aligned} T_n &= \tfrac{1}{2}\{\ln 1 + 2\ln 2 + \cdots + 2\ln(n-1) + \ln n\} \\ &= \ln 2 + \ln 3 + \cdots + \ln n - \tfrac{1}{2}\ln n = \ln n! - \tfrac{1}{2}\ln n, \end{aligned}$$

这里用了 $\ln 1 = 0$。令 $T_n \approx I_n$，

$$\ln n! \approx \left(n+\tfrac{1}{2}\right)\ln n + 1 - n, \qquad \text{所以} \qquad n! \approx n^{n+\frac{1}{2}}\mathrm{e}^{1-n} = \mathrm{g}(n).$$

为何 $n! < \mathrm{g}(n)$



由于 $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\ln x = -\dfrac{1}{x^2} < 0$，曲线 $y=\ln x$ 是严格凹的：在每条狭条 $[m, m+1]$ 上，除端点外，弦都严格位于曲线下方（见草图）。因此每个梯形的面积都严格小于该狭条上曲线下方区域的面积，把 $n-1 \geq 1$ 条狭条加起来，

$$T_n < I_n \quad \Longrightarrow \quad \ln n! - \tfrac{1}{2}\ln n < n\ln n - n + 1 \quad \Longrightarrow \quad \ln n! < \ln \mathrm{g}(n).$$

由于 exp 是递增函数，所以 $n! < \mathrm{g}(n)$。

宽度为 $1/k$ 的狭条

现在 $h = 1/k$、$N = k(n-1)$，而对 $r = 0,1,\ldots,k(n-1)$，有 $x_r = 1 + \frac{r}{k} = \frac{k+r}{k}$。最后一个纵坐标是 $x_N = n$。把梯形和写成“h 乘以所有纵坐标之和，再减去两端纵坐标各一半”，并利用 $\ln x_0 = 0$，

$$T_n^{(k)} = \frac{1}{k}\left\{\sum_{r=0}^{k(n-1)} \ln\frac{k+r}{k} - \tfrac{1}{2}\ln n\right\}.$$

分子 $k+r$ 依次取 $k, k+1, \ldots, kn$，而分母中共有 $k(n-1)+1$ 个因子 k，所以

$$\sum_{r=0}^{k(n-1)} \ln\frac{k+r}{k} = \ln\frac{(kn)!}{(k-1)!} - \big(k(n-1)+1\big)\ln k = \ln\frac{(kn)!}{k!} - k(n-1)\ln k,$$

因为 $(k-1)!\,k = k!$。令 $kT_n^{(k)} \approx kI_n = kn\ln n - kn + k$ 并整理，

$$\begin{aligned}\ln(kn)! &\approx \ln k! + k(n-1)\ln k + \tfrac{1}{2}\ln n + kn\ln n + k(1-n)\\ &= \ln k! + \left(kn+\tfrac{1}{2}\right)\ln n + k(1-n)(1-\ln k),\end{aligned}$$

即

$$(kn)! \approx k!\, n^{kn+\frac{1}{2}} \left(\frac{\mathrm{e}}{k}\right)^{k(1-n)} =: A,$$

正如所求。

哪一个更接近?

对整数 $a < b$，记误差（error）$\displaystyle\int_a^b \ln t\,\mathrm{d}t -$（单位步长梯形和，区间为$[a,b]$）为 $E(a,b)$。根据上面的凹性论证，$E(a,b) > 0$，而且相邻区间上的误差可以相加：$E(1,b) = E(1,a) + E(a,b)$。

推导 A 的过程中，除了以 I_n 代替 $T_n^{(k)}$ 之外，其余都是精确的，所以

$$\ln A - \ln(kn)! = k\big(I_n - T_n^{(k)}\big).$$

同理，把单位步长的推导中的 n 换成 kn，得 $\ln \mathrm{g}(kn) - \ln(kn)! = E(1,kn)$。

> ? **想一想** 宽度为 $1/k$ 的网格上的误差，怎样才能与宽度为 1 的网格上的误差作比较?

代入 $x = t/k$。纵坐标 $x_r = (k+r)/k$ 变成 $t = k, k+1, \ldots, kn$，而 $\mathrm{d}x = \mathrm{d}t/k$，所以

$$k\big(I_n - T_n^{(k)}\big) = \int_k^{kn} \ln\frac{t}{k}\,\mathrm{d}t \;-\; \big(\text{单位步长梯形和，函数为}\ln\tfrac{t}{k}\text{，区间为}[k,kn]\big).$$

由于 $\ln\frac{t}{k} = \ln t - \ln k$，而梯形法则对常数 $\ln k$ 是精确的，常数项互相抵消，所以上式正是 $E(k,kn)$。因此

$$\ln \mathrm{g}(kn) - \ln(kn)! = E(1,kn) = E(1,k) + E(k,kn) = E(1,k) + \big(\ln A - \ln(kn)!\big).$$

A 和 $\mathrm{g}(kn)$ 都大于 $(kn)!$，而当 $k \geq 2$ 时 $E(1,k) > 0$，所以

$$(kn)! < A < \mathrm{g}(kn) \qquad (k \geq 2).$$

所以无论以差还是以比来衡量，新的近似值 A 都比 $\mathrm{g}(kn)$ 更接近 $(kn)!$。当 $k = 1$ 时两者相同，因为 $A = 1!\,n^{n+\frac{1}{2}}\mathrm{e}^{1-n} = \mathrm{g}(n)$。换句话说，$A$ 对首 k 个因子使用了精确值 $k!$，只对其余部分作近似，因此避开了曲线起始处的误差 $E(1,k)$，而那正是 ln 弯曲得最厉害的地方。

验算

- 在最终公式中取 $k=1$，得回 $\mathrm{g}(n)$，理应如此。
- 取 $n=5$：$5!=120$，而 $\mathrm{g}(5)=5^{5.5}\mathrm{e}^{-4}\approx 128.0$，所以确实 $n!<\mathrm{g}(n)$。
- 取 $k=3$、$n=5$：$15!\approx 1.3077\times 10^{12}$，$A\approx 1.3369\times 10^{12}$，$\mathrm{g}(15)\approx 1.4102\times 10^{12}$，证实 $(kn)!<A<\mathrm{g}(kn)$。
- 比值 $\mathrm{g}(kn)/A=\exp E(1,k)$ 只取决于 k，与 n 无关；当 $k=3$ 时它等于 $\mathrm{g}(3)/3!\approx 1.0548$，与上面的数值吻合。

常见错误

- 数错宽度为 $1/k$ 的和中因子 k 的个数：纵坐标共有 $k(n-1)+1$ 个，而多出来的那个 $\ln k$ 正是把 $(k-1)!$ 变成 $k!$ 的关键。
- 用一幅弦位于曲线上方的草图去论证“$n!<\mathrm{g}(n)$”。对凹函数而言，弦位于曲线下方，所以梯形法则会低估积分。
- 忘记 $\ln A$ 的误差是细网格误差的 k 倍，而不是细网格误差本身。

STEP 之外

这里的 $\mathrm{g}(n)$ 是斯特灵公式（Stirling's formula）$n!\sim\sqrt{2\pi n}\,n^n\mathrm{e}^{-n}$ 的粗略形式。由于误差 $E(1,n)$ 随 n 递增且有界（每条狭条的误差最多是某常数乘以 $1/m^2$），它们收敛（converge），这就解释了为何 $n!/\mathrm{g}(n)$ 趋向常数 $\sqrt{2\pi}/\mathrm{e}\approx 0.922$。欧拉－麦克劳林公式（Euler–Maclaurin formula）把这一点精确化，并给出完整的渐近级数（asymptotic series）。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）讲述了黎曼积分（Riemann integral）以及支撑上述内容的数值积分法则的误差估计。

详细解答

STEP II 1987，第 8 题

纯数学 · 三维空间中的向量轨迹

STEP II 1987，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/87-s2-q8

关键思路

只要把每条方程改写成与位置向量（position vector）为 $\mathbf{a}$ 和 $\mathbf{b}$ 的点 A、B 之间的距离，就能认出它们各自代表甚么。至于等价（equivalence）的部分，一切都可以只用两个数表达：$P = |\mathbf{r}-\mathbf{a}|^2$ 及 $Q = |\mathbf{r}-\mathbf{b}|^2$。这样，(i) 和 (ii) 就变成两条关于 P 和 Q 的简单线性方程，(iii) 和 (iv) 亦然。

解答

以下各处，A、B 和 R 分别是位置向量为 $\mathbf{a}$、$\mathbf{b}$ 和 $\mathbf{r}$ 的点，而 M 是 AB 的中点（midpoint），其位置向量为 $\mathbf{m} = \frac{1}{2}(\mathbf{a}+\mathbf{b})$。我们假设 $\mathbf{a} \neq \mathbf{b}$（$\mathbf{a} = \mathbf{b}$ 的情况留到最后处理），并记 $d = |\mathbf{a}-\mathbf{b}| = AB$。

第 (i) 部分

展开，

$$\begin{aligned}|\mathbf{r}-\mathbf{a}|^2 - |\mathbf{r}-\mathbf{b}|^2 &= \left(|\mathbf{r}|^2 - 2\mathbf{a}\cdot\mathbf{r} + |\mathbf{a}|^2\right) - \left(|\mathbf{r}|^2 - 2\mathbf{b}\cdot\mathbf{r} + |\mathbf{b}|^2\right)\\ &= -2\Big[(\mathbf{a}-\mathbf{b})\cdot\mathbf{r} - \tfrac{1}{2}\left(|\mathbf{a}|^2 - |\mathbf{b}|^2\right)\Big].\end{aligned}$$

所以 (i) 成立当且仅当 $|\mathbf{r}-\mathbf{a}| = |\mathbf{r}-\mathbf{b}|$，即 $RA = RB$。其轨迹（locus）是 AB **的垂直平分面（perpendicular bisector plane）**：经过 M 而法向量（normal）为 $\mathbf{a}-\mathbf{b}$ 的平面。（我们也可以直接看出它是一个平面，因为 (i) 的形式是 $\mathbf{n}\cdot\mathbf{r} =$ 常数，其中 $\mathbf{n} = \mathbf{a}-\mathbf{b} \neq \mathbf{0}$；而 $\mathbf{r} = \mathbf{m}$ 满足此方程，因为 $(\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) = |\mathbf{a}|^2 - |\mathbf{b}|^2$。）

第 (ii) 部分

方程 (ii) 表示 $\overrightarrow{RA}$ 与 $\overrightarrow{RB}$ 互相垂直，或者 R 就是 A 或 B。要辨认这个轨迹，可以配方（complete the square）：

$$\begin{aligned}|\mathbf{r}-\mathbf{m}|^2 &= |\mathbf{r}|^2 - (\mathbf{a}+\mathbf{b})\cdot\mathbf{r} + \tfrac{1}{4}|\mathbf{a}+\mathbf{b}|^2 \\ &= (\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) - \mathbf{a}\cdot\mathbf{b} + \tfrac{1}{4}|\mathbf{a}+\mathbf{b}|^2 \\ &= (\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) + \tfrac{1}{4}|\mathbf{a}-\mathbf{b}|^2.\end{aligned}$$

所以 (ii) 等价于 $|\mathbf{r}-\mathbf{m}| = \frac{1}{2}d$。其轨迹是**以 AB 为直径（diameter）的球面（sphere）**：球心 M，半径 $\frac{1}{2}AB$。这正是“半圆上的圆周角是直角”的三维版本。

第 (iii) 及 (iv) 部分

方程 (iii) 表示 $RA = d/\sqrt{2}$。其轨迹是**以 A 为球心、$AB/\sqrt{2}$ 为半径的球面**。同样地，(iv) 的轨迹是**以 B 为球心、$AB/\sqrt{2}$ 为半径的球面**。

等价性

设 $P = |\mathbf{r}-\mathbf{a}|^2$ 及 $Q = |\mathbf{r}-\mathbf{b}|^2$。第 (i) 部分已证明

$$\text{(i)} \iff P = Q.$$

至于 (ii)，写成 $\mathbf{a}-\mathbf{b} = (\mathbf{a}-\mathbf{r}) - (\mathbf{b}-\mathbf{r})$ 再取平方：

$$d^2 = |\mathbf{a}-\mathbf{r}|^2 + |\mathbf{b}-\mathbf{r}|^2 - 2(\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}) = P + Q - 2(\mathbf{a}-\mathbf{r})\cdot(\mathbf{b}-\mathbf{r}).$$

因此

$$\text{(ii)} \iff P + Q = d^2.$$

方程 (iii) 和 (iv) 分别表示 $P = \frac{1}{2}d^2$ 及 $Q = \frac{1}{2}d^2$。现在

$$\left(P = Q \text{ 且 } P + Q = d^2\right) \iff P = Q = \tfrac{1}{2}d^2.$$

正方向成立，因为代入 $Q = P$ 得 $2P = d^2$。反方向方面，若 $P = Q = \frac{1}{2}d^2$，则显然 $P = Q$ 且 $P + Q = d^2$。以上每一步都是“当且仅当”，所以 (i) 与 (ii) 同时成立，等价于 (iii) 与 (iv) 同时成立，得证。

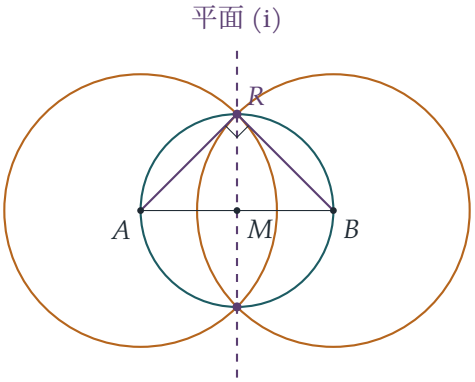
> ? **想一想** 一个平面与一个球面相交，怎会跟两个球面相交得出同一个集合?

两组方程描述的都是同一个圆。平面 (i) 经过球面 (ii) 的球心 M，所以它把该球面截出一个大圆（great circle）：圆心 M，半径 $\frac{1}{2}AB$，位于 AB 的垂直平分面上。球面 (iii) 和 (iv) 的半径同为 $AB/\sqrt{2}$，两者之和大于 AB，所以它们确实相交。由对称性，它们相交于垂直平分面上一个以 M 为圆心的圆。在三角形 AMR 中运用毕氏定理（Pythagoras' theorem），其半径为 $\sqrt{\frac{1}{2}AB^2 - \frac{1}{4}AB^2} = \frac{1}{2}AB$，所以是同一个圆。

几何意义

点 R 位于这个圆上，当且仅当三角形 ARB 是等腰直角三角形，且直角在 R。(i) 与 (ii) 这一组是通过三角形的性质来描述：$RA = RB$（等腰）及 $\angle ARB = 90°$（直角）。(iii) 与 (iv) 这一组则通过边长来描述：在斜边（hypotenuse）为 AB 的等腰直角三角形中，两条直角边的长度都是 $AB/\sqrt{2}$。上面的代数运算，其实就是毕氏定理及其逆定理（converse）的向量形式。恒等式 $d^2 = P + Q - 2(\mathbf{a} - \mathbf{r}) \cdot (\mathbf{b} - \mathbf{r})$ 就是三角形 ARB 中的余弦定理（cosine rule）。

下图展示任何一个包含 A 和 B 的平面所截出的截面。平分面显示为虚线，球面 (ii) 显示为青色圆，球面 (iii) 和 (iv) 则显示为琥珀色圆。四者相交于两点，就是解集的圆穿过纸面之处。



退化情况

若 $\mathbf{a} = \mathbf{b}$，则 (i) 变成 $0 = 0$，对所有 $\mathbf{r}$ 都成立，而 (ii)、(iii) 和 (iv) 都表示 $\mathbf{r} = \mathbf{a}$。等价性仍然成立，但每个轨迹都退化（degenerate）为单一点 A（对 (i) 而言则是整个空间）。

验算

取 $\mathbf{a} = (1, 0, 0)$ 及 $\mathbf{b} = (-1, 0, 0)$。那么 (i) 是平面 $x = 0$，(ii) 是 $x^2 + y^2 + z^2 = 1$。方程 (iii) 和 (iv) 是 $(x \mp 1)^2 + y^2 + z^2 = 2$，两式相减又得到 $x = 0$。四者相交于单位圆 $y^2 + z^2 = 1, x = 0$。例如 $R = (0, 1, 0)$ 给出 $RA = RB = \sqrt{2}$ 及 $\overrightarrow{RA} \cdot \overrightarrow{RB} = -1 + 1 = 0$。我们亦以有理数运算，对 5000 组随机选取的 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{r}$ 精确验证了上述所有恒等式，并确认预测的圆上的随机点都满足全部四条方程。

常见错误

- 把 (ii) 说成“一个圆”。在三维空间中它是球面，而 (i) 的轨迹是平面，不是直线。
- 只证明了一个方向。题目要求证明等价，所以要确保每一步都可逆，或者把两个方向都证明。
- 只说两组方程“都得出一个圆”，却没有核实那是同一个圆：同一平面、同一圆心、同一半径。

STEP 之外

这些轨迹每一个都是 **r** 的某个二次函数的等值集（level set），而把两条球面方程相减得出一个平面的技巧，正是两个球面的根面（radical plane）这个概念。更广泛地说，用不同的方程组描述同一个几何对象，是代数几何（algebraic geometry）的起点。在这里，空间中的一个圆由两条方程截出，而有许多不同的方程组合都能做到。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）是巩固向量代数的好去处，特别是点积（dot product）、长度与角度之间的关系。

详细解答

STEP II 1987，第 9 题

纯数学 · 正交矩阵

> **STEP II 1987，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/87-s2-q9

关键思路

把 $\mathbf{B}$ 转置（transpose）之后，条件 $\mathbf{B}^{\mathrm{T}}\mathbf{B}=\mathbf{I}$ 的最左和最右两端各有一个逆矩阵（inverse matrix）。由于这些逆矩阵本身可逆（invertible），我们可以左乘 $\mathbf{I}-\mathbf{A}^{\mathrm{T}}$、右乘 $\mathbf{I}-\mathbf{A}$，而不会损失任何信息。余下的是一条关于 $\mathbf{A}$ 和 $\mathbf{A}^{\mathrm{T}}$ 的多项式恒等式，几乎立刻就能化简。要小心的是每一步都必须可以逆推，因为我们要证明的是“若且仅若”（if and only if）。

解答

我们会用到以下标准结果：

- $(\mathbf{XY})^{\mathrm{T}}=\mathbf{Y}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}$，而且转置是线性（linear）的，所以 $(\mathbf{I}\pm\mathbf{A})^{\mathrm{T}}=\mathbf{I}\pm\mathbf{A}^{\mathrm{T}}$（因为 $\mathbf{I}^{\mathrm{T}}=\mathbf{I}$）；
- 若 $\mathbf{X}$ 是非奇异（non-singular）的，则 $\mathbf{X}^{\mathrm{T}}$ 也是，并且 $(\mathbf{X}^{-1})^{\mathrm{T}}=(\mathbf{X}^{\mathrm{T}})^{-1}$。

第 1 步：$\mathbf{B}$ 的转置。 由于 $\mathbf{I}-\mathbf{A}$ 是非奇异的，它的转置 $\mathbf{I}-\mathbf{A}^{\mathrm{T}}$ 也是非奇异的。把乘积的次序倒转，

$$\mathbf{B}^{\mathrm{T}}=\left((\mathbf{I}-\mathbf{A})^{-1}\right)^{\mathrm{T}}(\mathbf{I}+\mathbf{A})^{\mathrm{T}}=(\mathbf{I}-\mathbf{A}^{\mathrm{T}})^{-1}(\mathbf{I}+\mathbf{A}^{\mathrm{T}}).$$

因此

$$\mathbf{B}^{\mathrm{T}}\mathbf{B}=(\mathbf{I}-\mathbf{A}^{\mathrm{T}})^{-1}\,(\mathbf{I}+\mathbf{A}^{\mathrm{T}})(\mathbf{I}+\mathbf{A})\,(\mathbf{I}-\mathbf{A})^{-1}.$$

第 2 步：以可逆推的方式消去逆矩阵。 设 $\mathbf{P}=\mathbf{I}-\mathbf{A}^{\mathrm{T}}$、$\mathbf{Q}=\mathbf{I}-\mathbf{A}$ 及 $\mathbf{M}=(\mathbf{I}+\mathbf{A}^{\mathrm{T}})(\mathbf{I}+\mathbf{A})$，于是 $\mathbf{B}^{\mathrm{T}}\mathbf{B}=\mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1}$。我们断言

$$\mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1}=\mathbf{I}\quad\Longleftrightarrow\quad\mathbf{M}=\mathbf{P}\mathbf{Q}.$$

（⇒）左乘 $\mathbf{P}$、右乘 $\mathbf{Q}$：$\mathbf{M} = \mathbf{PIQ} = \mathbf{PQ}$。
（⇐）左乘 $\mathbf{P}^{-1}$、右乘 $\mathbf{Q}^{-1}$：$\mathbf{P}^{-1}\mathbf{MQ}^{-1} = \mathbf{P}^{-1}\mathbf{PQQ}^{-1} = \mathbf{I}$。

> ? **想一想** 为甚么不能只是“两边同乘”就了事?

把一条等式乘以一个矩阵，得到的必定仍是正确的等式，但这一步一般不能还原：由 $\mathbf{XY} = \mathbf{XZ}$ 不能推断 $\mathbf{Y} = \mathbf{Z}$，除非 $\mathbf{X}$ 可逆。这里反方向之所以成立，全因 $\mathbf{P}$ 和 $\mathbf{Q}$ 都有逆矩阵，而这正是用上“$\mathbf{I} - \mathbf{A}$ 是非奇异的”（因而 $\mathbf{I} - \mathbf{A}^{\mathrm{T}}$ 也是）这个假设的地方。还要留意，$\mathbf{P}$ 必须保持在左边，$\mathbf{Q}$ 必须保持在右边，因为矩阵乘法不符合交换律（commutative）。

第 3 步：展开。 把 $\mathbf{M} = \mathbf{PQ}$ 的两边展开，并小心保持各因子的次序，

$$(\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A}) = \mathbf{I} + \mathbf{A} + \mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{A},$$
$$(\mathbf{I} - \mathbf{A}^{\mathrm{T}})(\mathbf{I} - \mathbf{A}) = \mathbf{I} - \mathbf{A} - \mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{A}.$$

$\mathbf{I}$ 和 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ 这两项在两边都出现，所以相减后得

$$\mathbf{M} = \mathbf{PQ} \iff 2(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) = \mathbf{O} \iff \mathbf{A} + \mathbf{A}^{\mathrm{T}} = \mathbf{O}.$$

以上每一步都可以逆推：两边同时加上或减去同一个矩阵，以及乘以非零纯量（scalar）2 或 $\frac{1}{2}$。

结论。 把第 1 至第 3 步串连起来，每一环都是等价关系，

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I} \iff (\mathbf{I} + \mathbf{A}^{\mathrm{T}})(\mathbf{I} + \mathbf{A}) = (\mathbf{I} - \mathbf{A}^{\mathrm{T}})(\mathbf{I} - \mathbf{A}) \iff \mathbf{A} + \mathbf{A}^{\mathrm{T}} = \mathbf{O},$$

这正是我们要证明的结果。用文字表达：$\mathbf{B}$ 是正交矩阵（orthogonal matrix），当且仅当 $\mathbf{A}$ 是斜对称矩阵（skew-symmetric matrix）。

验算

- 取 $\mathbf{A} = \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}$，它是斜对称的。这时 $\det(\mathbf{I} - \mathbf{A}) = 1 + t^2 \neq 0$，而
$$\mathbf{B} = \frac{1}{1+t^2}\begin{pmatrix} 1 - t^2 & 2t \\ -2t & 1 - t^2 \end{pmatrix},$$
这是一个旋转矩阵（rotation matrix）（代入 $t = \tan\frac{\theta}{2}$），所以一如所料 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$。
- 1×1 的情况：$\mathbf{A} = (a)$，其中 $a \neq 1$，得 $\mathbf{B} = \left(\frac{1+a}{1-a}\right)$，而 $\mathbf{B}^2 = 1$ 迫使 $1 + a = \pm(1 - a)$，即 $a = 0$（另一个符号不可能成立）。这与 $2a = 0$ 吻合。
- 我们以精确的有理数运算测试了约 2000 个大小由 1 至 5 的随机矩阵：每一个斜对称的 $\mathbf{A}$ 都准确地给出 $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$，其余的则全部不成立。

常见错误

- 没有倒转次序就写 $\mathbf{B}^{\mathrm{T}} = (\mathbf{I}+\mathbf{A}^{\mathrm{T}})(\mathbf{I}-\mathbf{A}^{\mathrm{T}})^{-1}$。（这条式子碰巧是对的，因为 $\mathbf{I}+\mathbf{A}^{\mathrm{T}}$ 与 $(\mathbf{I}-\mathbf{A}^{\mathrm{T}})^{-1}$ 可交换，但这一点需要另行说明理由。）
- 展开 $(\mathbf{I}+\mathbf{A}^{\mathrm{T}})(\mathbf{I}+\mathbf{A})$ 时写成 $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ 而不是 $\mathbf{A}^{\mathrm{T}}\mathbf{A}$。在这一题里两种写法都会抵销，但前提是两边的展开方式必须一致。
- 只证明了一个方向。题目要求的是“若且仅若”，所以要说明每一步为何可以逆推。

STEP 之外

映射 $\mathbf{A} \mapsto (\mathbf{I}+\mathbf{A})(\mathbf{I}-\mathbf{A})^{-1}$ 称为凯莱变换（Cayley transform）。它把构成一个向量空间（vector space）的斜对称矩阵，变成构成一个群（group）的正交矩阵，从而给出旋转的有理参数化（rational parametrisation）。（可以证明，当 $\mathbf{A}$ 是实斜对称矩阵时，$\mathbf{I}-\mathbf{A}$ 必定可逆，而且所得到的正交矩阵恰好是那些没有特征值（eigenvalue）-1 的正交矩阵。）这让我们初次窥见李群（Lie group）与它的李代数（Lie algebra）之间的关系。想进一步了解正交矩阵和斜对称矩阵，Strang 的 *Introduction to Linear Algebra*（线性代数入门）是很好的读物。

详细解答

STEP II 1987，第 10 题

纯数学 · 群公理

> **STEP II 1987，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/87-s2-q10

关键思路

一切都发生在下标（subscripts）之中，所以把每个运算转化为下标 $1,2,\ldots,N$ 上的普通算术：第一个运算是“取最大值”，第二个运算是“取两者之差的绝对值，再加一”。然后逐一检验四条公理（axioms）：封闭性（closure）、结合律（associativity）、单位元（identity）、逆元（inverses）。要说明某条公理不成立，只需一个具体的反例（counterexample）；要说明某条公理成立，则需要一般性的论证。

解答

全文中，以 $j\circ k$ 表示 $a_j\circ a_k$ 的下标，$*$ 亦同理。两个元素相等，当且仅当它们的下标相等。

运算 $\circ$

这里 $a_j\circ a_k=a_{\max(j,k)}$。

封闭性成立。若 $1\le j,k\le N$，则 $\max(j,k)$ 是 j 和 k 其中之一，所以它属于 $\{1,\ldots,N\}$，因此 $a_j\circ a_k\in S$。

结合律成立。对任意 j,k,l，

$$(a_j\circ a_k)\circ a_l=a_{\max(\max(j,k),l)}=a_{\max(j,k,l)}=a_{\max(j,\max(k,l))}=a_j\circ(a_k\circ a_l),$$

因为两种加括号的方式都得出三个下标中最大的一个。

单位元存在：它就是 a_1。每个下标都至少是 1，所以 $\max(1,k)=\max(k,1)=k$，因此对每个 k 都有 $a_1\circ a_k=a_k\circ a_1=a_k$。

逆元不成立。a_j 的逆元应是某个 a_k，使得 $a_j\circ a_k=a_1$，即 $\max(j,k)=1$。但 $\max(j,k)\ge j$，

所以当 $j \geq 2$ 时这是不可能的。例如，对每个 k，$a_2 \circ a_k$ 都是 a_2 或排在其后的元素，永远不会是 a_1。只有 a_1 有逆元（就是它自己）。

所以 S 在 $\circ$ 下满足封闭性、结合律和单位元，但不满足逆元公理，因此不是群（group）。

想一想 会不会有其他元素也是单位元，以致逆元的论证需要重做?

不会：单位元是唯一的。若 e 和 e' 都是单位元，则 $e = e \circ e' = e'$。所以“$a_j \circ a_k$ 等于单位元”中的单位元只可能是 a_1，上面的论证是完整的。

运算 $*$

这里 $a_j * a_k = a_{|j-k|+1}$。当 $N = 4$ 时，下标的运算表为

$*$	1	2	3	4
1	1	2	3	4
2	2	1	2	3
3	3	2	1	2
4	4	3	2	1

表中可见的规律对每个 N 都成立。

封闭性成立。若 $1 \leq j, k \leq N$，则 $0 \leq |j-k| \leq N-1$，所以 $1 \leq |j-k|+1 \leq N$，因此 $a_j * a_k \in S$。

单位元存在：它就是 a_1。由于 $k \geq 1$, $|1-k| = k-1$, 所以 $a_1 * a_k = a_k$; 同样地 $a_k * a_1 = a_k$（因为 $|j-k| = |k-j|$，这个运算是可交换的（commutative））。

逆元成立。$a_j * a_j = a_{0+1} = a_1$，即单位元，所以每个元素都是自己的逆元。（这就是表中对角线上的一列 1。）

结合律不成立。取 a_2, a_2, a_3（这需要 $N \geq 3$）：

$$(a_2 * a_2) * a_3 = a_1 * a_3 = a_3,$$
$$a_2 * (a_2 * a_3) = a_2 * a_{|2-3|+1} = a_2 * a_2 = a_1.$$

两者不同，所以 $*$ 不满足结合律。

所以 S 在 $*$ 下满足封闭性、单位元和逆元，但不满足结合律，因此同样不是群。

想一想 哪里用到了 $N > 2$?

只在结合律的反例中用到，因为它用了 a_3。当 $N = 2$ 时，运算表就是 2 阶群的运算表（a_1 为单位元，$a_2 * a_2 = a_1$），而 $*$ 是满足结合律的；至于 $\circ$，逆元的论证只需要 a_2，所以 $\circ$ 在 $N = 2$ 时同样不成群。

验算

用穷举法（brute force）检查小的情况：对 2 至 9 的每个 N，我们列出所有三元组来检验结合律，并直接搜索单位元和逆元。对于 $\circ$：封闭、满足结合律、单位元为 a_1，且只有 a_1 可逆。对于 $*$：封闭、单位元为 a_1、每个元素都是自己的逆元，而结合律对每个 $N \geq 3$ 都不成立（包括在 (a_2, a_2, a_3) 处），但在 $N = 2$ 时成立。你可以用一个 3×3 的运算表，一分钟内亲手完成 $N = 3$ 的情况。

常见错误

- 以为“取最大值”不像乘法，便断言 $\circ$ 没有单位元。应检查最小的元素：a_1 正是单位元。
- 尝试一般地证明 $*$ 满足结合律，或者只检验含有 a_1 的三元组（这些三元组必然满足结合律，因为 a_1 是单位元）。真正需要的只是一个具体的反例。
- 忘记检验封闭性。封闭性很容易被想当然，但这里它需要用到上界 $|j - k| \leq N - 1$。

STEP 之外

一个带有满足结合律的运算和单位元的集合，例如 S 在 $\circ$ 下，称为么半群（monoid）；全序集（totally ordered set）上的“取最大值”是一个标准例子，它同时也是一个半格（semilattice）。一个满足封闭性、单位元和逆元但不满足结合律的集合，例如 S 在 $*$ 下，说明结合律是一条真正独立的公理：它不能从其余三条推导出来。若想了解群的各条公理如何互相配合，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 1987，第 11 题

力学 · 摩擦力与力矩

STEP II 1987，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/87-s2-q11

关键思路

支点的作用力是未知的，我们也不需要求它，所以我们对支点 O 取力矩（moment），因为该力对 O 没有力矩。杆是处于三维空间中的，但圆环所在的平面给出两个自然的方向：垂直于平面的方向（法向反作用力（normal reaction）沿此方向作用）和平面内的方向（摩擦力（friction）在此方向作用）。把重力也按同样方式分解，一条力矩方程给出法向反作用力，另一条给出摩擦力。两者之比就是 μ。

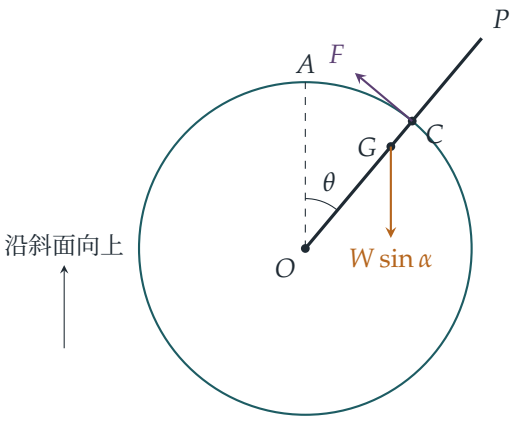
解答

模型。 杆 OP 是一条重量为 W 的均匀杆，其质心（centre of mass）G 位于中点，所以 $OG = \frac{1}{2}b$。杆以光滑支点（pivot）固定于圆环的圆心 O。杆横跨圆环，搁在圆周上的一点 C，所以 $OC = a$。由于 O 和 C 都在圆环所在的平面上，杆也位于该平面内。设 A 为圆环的最高点，于是 OA 沿最大斜率线（line of greatest slope）直指向上，并设 $\angle AOC = \theta$。

作用力。 作用于杆上的力有：

- 作用于 G 的重力 W；
- 光滑支点 O 处的作用力，此力未知，但对 O 没有力矩；
- C 处的接触力。杆与圆环的金属线在 C 相交。杆沿半径方向，金属线沿切线方向，两者都在平面内，所以它们的公共法线垂直于平面。因此法向反作用力 R 垂直于圆环所在的平面。摩擦力 F 位于平面内。杆唯一可能的运动是绕 O 转动，这会令 C 沿圆环的切线移动。所以摩擦力沿该切线作用，与杆垂直。

把重力分解为垂直于平面（指向平面内）的分量 $W\cos\alpha$，以及在平面内沿最大斜率线向下的分量 $W\sin\alpha$。下图是垂直俯视该平面所见的情况，A 在最上方。R 在 C 处指出纸面。



对 O 取力矩。 由于支点光滑，其余各力对经过 O 的每一条轴的总力矩都是零。其中两条轴能给出有用的方程。

*在平面内、经过 O 并垂直于杆的轴。*只有垂直于平面的力对这条轴有力矩，即距离为 $\frac{1}{2}b$ 的 $W\cos\alpha$ 和距离为 a 的 R。平面内的力 F 和 $W\sin\alpha$ 要么与这条轴相交，要么与它平行，所以对它没有力矩。因此

$$R\,a = W\cos\alpha\cdot\tfrac{1}{2}b, \qquad \text{所以} \qquad R = \frac{Wb\cos\alpha}{2a}.$$

*垂直于平面、经过 O 的轴。*现在只有平面内的力才起作用。杆与沿斜面向上的方向成 θ 角，所以 $W\sin\alpha$ 垂直于杆的分量是 $W\sin\alpha\sin\theta$，作用于距离 $\frac{1}{2}b$ 处。摩擦力本身已垂直于杆，作用于距离 a 处。因此

$$F\,a = W\sin\alpha\sin\theta\cdot\tfrac{1}{2}b, \qquad \text{所以} \qquad F = \frac{Wb\sin\alpha\sin\theta}{2a}.$$

重力在平面内的分量倾向令杆摆离 A，向圆环底部摆下去，而摩擦力则阻止这个运动。

想一想 摩擦力会不会也有沿杆方向的分量?

沿杆方向的力经过 O,所以对 O 没有力矩。它既不能促进也不能阻碍转动,而支点可以提供沿杆方向的任何力。唯一可能的运动是绕 O 转动, 所以对极限平衡（limiting equilibrium）而言，有关的摩擦力是沿切线方向的分量，也就是 F。我们把摩擦力视为完全沿切线作用。

极限平衡。 在极限平衡时 $F = \mu R$:

$$\frac{Wb\sin\alpha\sin\theta}{2a} = \mu\,\frac{Wb\cos\alpha}{2a} \qquad \Longrightarrow \qquad \boxed{\mu = \tan\alpha\,\sin\theta}$$

（我们取 $0 \le \theta \le \pi$）。只要 $\tan\alpha \sin\theta \le \mu$，杆便能在角 θ 处保持平衡。留意 a、b 和 W 全都消去了。

验算

- 若 $\theta = 0$ 或 $\theta = \pi$，杆沿最大斜率线放置。此时重力在平面内的分量正好沿杆方向拉，不需要摩擦力：$\mu = 0$。当 $\alpha = 0$ 时亦然，因为圆环水平时重力在平面内没有分量。
- 当 $\theta = \frac{1}{2}\pi$，即杆处于水平时，所需的摩擦力最大。此时 $\mu = \tan\alpha$，与倾角为 α 的斜面上的木块情况吻合。
- 我们做了一次数值验算：对 α, θ, a, b, W 随机取 2000 组值，把各力写成三维向量（vector），解对 O 的向量力矩方程求 R 和 F，发现 F/R 与 $\tan\alpha \sin\theta$ 的差距在 4×10^{-15} 以内。例如 $\alpha = 30°$、$\theta = 60°$ 给出 $\mu = \frac{1}{2}$。

常见错误

- 以为 C 处的法向反作用力沿半径方向，就像杆靠在铅垂平面中的圆环内侧那样。这里杆是横跨在金属线上的，所以法线垂直于圆环所在的平面。
- 用分解力的方法而不取力矩。这样会引入支点作用力的三个未知分量，而题目并不需要它们。

STEP 之外

这里我们对经过 O 的两条互相垂直的轴取力矩。这是单一向量方程 $\sum \mathbf{r} \times \mathbf{F} = \mathbf{0}$ 的捷径，该方程表示对 O 的总转矩（torque）为零。把转矩视为向量，以及刚体（rigid body）在三维空间中的平衡条件，都是大学一年级力学的标准内容，并会进一步引向刚体绕固定点的转动。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）正是以这种方式，把转矩和角动量（angular momentum）作为向量来发展。

详细解答

STEP II 1987，第 12 题

力学 · 平衡与能量

STEP II 1987，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/87-s2-q12

关键思路

“在每个位置都处于平衡（equilibrium）”其实是关于势能（potential energy）的陈述：如果珠子（bead）可以停在线上的任何位置，那么珠子沿线滑动时，系统的势能就不会改变，所以势能在线上必须是常数。用珠子的高度及它与环的距离写出这个能量，便立即得到曲线的方程。同一事实也令第二部分变得简洁：势能既然是常数，能量守恒（conservation of energy）便告诉我们珠子与质点的总动能（kinetic energy）是常数，余下的只是一些把两个速率联系起来的几何。

解答

我们把珠子和悬挂的质量视为质点，绳视为轻而不可伸长（inextensible），环和线视为光滑。把环放在原点 O，y 垂直向上量度。设珠子位于极坐标（polar coordinates）(r,θ)，其中 θ 由水平方向逆时针量度，因此珠子相对于 O 的高度是 $y = r\sin\theta$。设 L 为绳的长度。

曲线的方程

由 O 到珠子的一段绳长 r，所以悬挂的质点位于环下方距离 $L - r$ 处。取 O 为零势能水平，系统的势能是

$$V = mgr\sin\theta - mg(L - r) = mg\,(r + r\sin\theta) - mgL.$$

线是光滑的，所以它对珠子的反作用力垂直于线，能使系统沿线移动的力只来自 V。系统在某位置处于平衡，当且仅当 V 在该处对沿线的移动取驻值。在每个位置都平衡，意味着沿整条线都有 $\mathrm{d}V/\mathrm{d}s = 0$（其中 s 为弧长（arc length）），所以 V 在线上是常数：

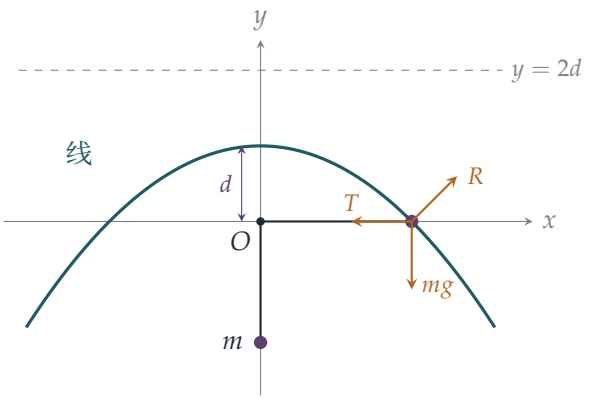
$$r(1 + \sin\theta) = c \quad \text{对某常数} c.$$

用分解力的方法也可以看到这一点。平衡时，悬挂的质点令整条绳的张力（tension）为 mg（环是光滑的）。在珠子上，指向 O 的张力 mg 的切向分量是 $-mg\,\frac{\mathrm{d}r}{\mathrm{d}s}$，而重力的切向分量是 $-mg\,\frac{\mathrm{d}y}{\mathrm{d}s}$。两者必须抵消，所以 $\frac{\mathrm{d}}{\mathrm{d}s}(r+y)=0$，这正是同一条件。

求 c：由于 $1+\sin\theta\le 2$，我们有 $r=c/(1+\sin\theta)\ge c/2$，等号在 $\theta=\frac{\pi}{2}$ 时成立（即线上位于环正上方的点）。所以环到线的最短距离是 $c/2$，由此得 $c=2d$，而曲线是

$$\boxed{r(1+\sin\theta)=2d} \qquad \text{即} \qquad r=\frac{2d}{1+\sin\theta}.$$

（如果改为由向下的铅垂线量度 θ，同一曲线写成 $r(1-\cos\theta)=2d$。）由于 $r=2d-y$，线上每一点与 O 的距离都等于它与水平线 $y=2d$ 的距离。所以这条线是一条抛物线（parabola），焦点（focus）在环上，准线（directrix）在环上方高度 $2d$ 处；以直角坐标表示为 $x^2=4d(d-y)$，顶点为 $(0,d)$。



图中珠子与环处于同一水平；张力 $T=mg$、重力 mg 和法向反作用力（normal reaction）R 互相平衡，因为抛物线在该处的斜率恰好是 -1。

珠子的速率

现在珠子以速率 u 沿线移动。绳保持拉紧，长度不变，所以悬挂的质点仍位于环下方距离 $L-r$ 处。

想一想 悬挂的质点移动得多快?

它在环下方的深度是 $L-r$，所以它的速度是垂直的，速率为 $|\dot r|$。注意 $\dot r$ 只是珠子速度的径向部分，而不是它的整个速率 u。

一旦开始运动，张力便不再是 mg，但它对系统不做净功：绳不可伸长而环光滑，所以张力对珠子所做的功与它对悬挂质点所做的功互相抵消。线的反作用力垂直于珠子的运动，不做功。所以能量守恒：

$$\tfrac{1}{2}mu^2+\tfrac{1}{2}m\dot r^2+V=\text{常数}.$$

但由第一部分，在线上 $V = mg\cdot 2d - mgL$ 是常数。因此

$$u^2 + \dot r^2 = \text{常数}. \qquad (*)$$

接着把 $\dot r$ 与 u 联系起来。对 $r = 2d/(1+\sin\theta)$ 求导，

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = -\frac{2d\cos\theta}{(1+\sin\theta)^2} = -\frac{r\cos\theta}{1+\sin\theta}, \qquad \text{所以} \qquad \dot r^2 = r^2\dot\theta^2\,\frac{\cos^2\theta}{(1+\sin\theta)^2} = r^2\dot\theta^2\,\frac{1-\sin\theta}{1+\sin\theta},$$

这里用了 $\cos^2\theta = (1-\sin\theta)(1+\sin\theta)$。由曲线方程，$1+\sin\theta = 2d/r$ 及 $1-\sin\theta = 2-2d/r = 2(r-d)/r$，所以

$$\frac{1-\sin\theta}{1+\sin\theta} = \frac{r-d}{d}, \qquad \text{得} \qquad \dot r^2 = \frac{r-d}{d}\,r^2\dot\theta^2.$$

在极坐标中，速率满足 $u^2 = \dot r^2 + r^2\dot\theta^2$，所以

$$u^2 = r^2\dot\theta^2\left(\frac{r-d}{d}+1\right) = \frac{r}{d}\,r^2\dot\theta^2, \qquad \text{因此} \qquad \dot r^2 = \frac{r-d}{r}\,u^2.$$

想一想 当 $r = d$ 时，$\dot r$ 是多少?

当 $r = d$ 时，珠子位于顶点，$\theta = \frac{\pi}{2}$，此处 r 取最小值，且 $\mathrm{d}r/\mathrm{d}\theta = 0$。所以该处 $\dot r = 0$（公式 $\dot r^2 = \frac{r-d}{r}u^2$ 也这样说），悬挂的质点瞬间静止。因此 $(*)$ 中的常数是 v^2：

$$u^2 + \frac{r-d}{r}\,u^2 = v^2 \quad\Longrightarrow\quad u^2\,\frac{2r-d}{r} = v^2 \quad\Longrightarrow\quad u = \left(\frac{r}{2r-d}\right)^{\frac12} v,$$

由于 u 和 v 是速率，取正根。这就是所求的结果。

验算

- 当 $r = d$ 时，公式给出 $u = v$，理应如此。
- 当 $r \to \infty$ 时，$u \to v/\sqrt{2}$。在很远处，抛物线的切线几乎正好沿着半径方向，所以 $\dot r \approx u$，而动能 $\frac12 mv^2$ 由珠子与质点平均分享。
- 在环的水平上的平衡：点 $(2d, 0)$ 在曲线上，该处张力 mg（水平）与重力 mg（垂直）合成一个与水平成 45° 的力，垂直于抛物线在该处斜率为 -1 的切线。
- 数值上，取 $d = 1.3$ 及 $v = 2$，我们由顶点出发，沿抛物线对珠子与质点系统的运动方程进行数值积分。在 $r = 2.8905$ 处，珠子的速率是 1.606309，与 $\sqrt{r/(2r-d)}\,v$ 相差约 10^{-8}。我们亦在线上 1000 个随机点验证了静止珠子所受的切向力为零。

常见错误

- 忘记悬挂质点的动能，结果对所有 r 都得出 $u = v$。势能是常数，但动能是由两者分享的。
- 把质点的速率当作 u 而不是 $|\dot{r}|$。只有珠子运动的径向部分才会把绳拉过环。
- 由另一条线量度 θ，却照搬 $r(1+\sin\theta) = 2d$。任何正确的极坐标形式都可以，但要说明 θ 是由哪条线量度的。

STEP 之外

系统在势能取驻值之处处于平衡，这个想法就是*虚功原理*（principle of virtual work），它引向拉格朗日力学（Lagrangian mechanics）。在拉格朗日力学中，动能以一个广义坐标（generalised coordinate）（这里是 θ）表示，运动方程由此直接得出，完全不必求张力。由于 V 在这条线上是常数，珠子和质点的运动就好像完全没有受力一样，而我们的关系式不过是动能 $\frac{1}{2}m r^2(2r-d)\dot{\theta}^2/d$ 守恒。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）以这种精神讲解能量方法和受约束的运动。

详细解答

STEP II 1987，第 13 题

力学 · 与墙壁的斜碰撞

> **STEP II 1987，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/87-s2-q13

关键思路

光滑（smooth）的台边只能改变垂直于它的速度分量（component）。这个分量会反向并乘以 e，而沿台边方向的分量则保持不变。两条相邻的台边分别作用于两个不同的分量，所以每个分量各被反向并乘以 e 一次。至于闭合路径，我们用坐标追踪圆盘由一条台边到另一条台边的运动，并选取唯一可自由调节的量，即初始斜率（slope），使圆盘回到出发点。

解答

模型。圆盘视为一个质点（particle），在设有光滑台边的光滑水平台面上滑动，因此两次碰撞之间没有任何力作用于它，它以等速沿直线运动。碰撞时，台边施加的冲量（impulse）垂直于台边（因为台边光滑），所以速度的平行分量不变。根据牛顿恢复定律（Newton's law of restitution），垂直分量反向并乘以 e。我们设 $0 < e \leq 1$。

沿台面的两条边设坐标轴，使台面为 $0 \leq x \leq L$、$0 \leq y \leq B$。

第 (i) 部分：两条相邻的台边

设圆盘的速度为 (u, v)。相邻的台边是一条形如 $x =$ 常数 的台边和一条形如 $y =$ 常数 的台边，碰撞次序不限。撞击 x 台边使 $(u, v) \mapsto (-eu, v)$，撞击 y 台边使 $(u, v) \mapsto (u, -ev)$。两者各撞一次后，无论次序如何，速度均为

$$(-eu,\ -ev) = -e\,(u, v).$$

这是原速度的纯量倍数，所以最后的路径与原路径平行，正如所求。此时圆盘沿相反方向运动，速率减为原来的 e 倍。

第 (ii) 部分：从台边出发的闭合路径

设圆盘静止于台边 $y=0$ 上的点 $P=(a,0)$，且 $a:(L-a)=z:1$，于是

$$a=\frac{zL}{1+z},\qquad L-a=\frac{L}{1+z},\qquad 0<a<L.$$

从上方俯视台面，x 向右、y 向上，由台边 $y=0$ 开始顺时针方向，依次遇到 $x=0$、$y=B$，然后是 $x=L$。若比例是从另一端量度，则把 z 换成 $1/z$，论证不变。以速度 $(-u,v)$ 击出圆盘，其中 $u,v>0$，并记斜率为 $k=v/u$。四段路程上的速度依次为

$$(-u,v)\xrightarrow{x=0}(eu,v)\xrightarrow{y=B}(eu,-ev)\xrightarrow{x=L}(-e^2u,-ev)\xrightarrow{y=0}\cdots$$

- 第 1 段，由 $(a,0)$ 至 $x=0$：水平走过距离 a，上升 ak，所以撞击左台边于 $W_1=(0,\,ak)$。
- 第 2 段，斜率为 $v/(eu)=k/e$：上升 $B-ak$，因此横向移动 $e(B-ak)/k$。它在 $W_2=\big(e(B/k-a),\,B\big)$ 碰到顶边。
- 第 3 段，斜率为 $-k$：横向移动 $L-x_{W_2}$，下降 $k(L-x_{W_2})$，所以在高度 $y_{W_3}=B-kL+e(B-ak)$ 处碰到右台边。
- 第 4 段，斜率为 k/e：下降 y_{W_3}，向回移动 $e\,y_{W_3}/k$，到达 $y=0$ 时位于

$$x_{\text{end}}=L-\frac{e\,y_{W_3}}{k}=(1+e)L+e^2a-\frac{e(1+e)B}{k}.$$

我们要求 $x_{\text{end}}=a$。由此得 $e(1+e)B/k=(1+e)L-(1-e^2)a=(1+e)\big(L-(1-e)a\big)$，所以

$$k=\frac{eB}{L-(1-e)a}=\frac{B}{L}\cdot\frac{e(1+z)}{1+ez}.$$

由于 $L-(1-e)a>L-a>0$，对每个 $z>0$，这都是正数。

想一想 我们完成了吗？我们假设了圆盘依左、上、右、下的次序碰到台边。

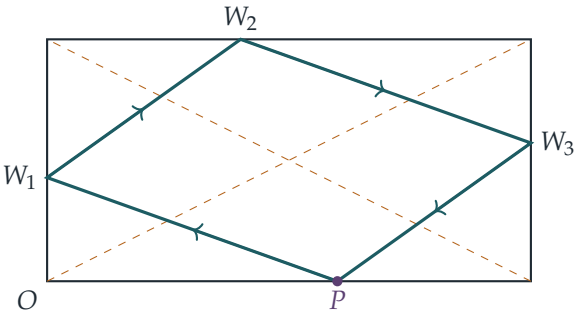
还未完成。我们仍须验证每个碰撞点确实落在我们假设的那条边上。记 $D=L-(1-e)a=L-a+ea$，则 $k=eB/D$。于是

$$y_{W_1}=ak=\frac{eaB}{D},$$
$$x_{W_2}=e\Big(\frac{D}{e}-a\Big)=L-a,$$
$$y_{W_3}=B(1+e)-k(L+ea)=\frac{B\big[(1+e)(L-a+ea)-eL-e^2a\big]}{D}=\frac{B(L-a)}{D}.$$

由于 $0<a<L$ 及 $e>0$，我们有 $0<ea<D$ 及 $0<L-a<D$。所以 $0<y_{W_1}<B$、$0<x_{W_2}<L$ 及 $0<y_{W_3}<B$。每次碰撞都落在我们假设的那条边（不含端点）上，不会

撞到角位。因此对每个 z，以上述斜率 k 击出的圆盘会依次经过左、上、右台边，然后回到 P。

由第 (i) 部分，第 1 段与第 3 段平行，第 2 段与第 4 段亦平行，所以这条闭合路径是内接于台面的平行四边形（parallelogram）$PW_1W_2W_3$。



虚线是对角线（diagonal）。图中路径取 $L = 2B$、$e = \frac{1}{2}$ 及 $z = 1.5$。

第 (iii) 部分：从任意内点出发的两种击法

*每一种这样的击法都是第 (ii) 部分的路径之一。*设圆盘从内点（interior point）Q 以速度 $\mathbf{w}$ 出发，依顺时针次序在四条台边上的 V_1, V_2, V_3, V_4 反弹，然后回到 Q。应用第 (i) 部分两次：四次反弹后速度为 $(-e)^2\mathbf{w} = e^2\mathbf{w}$，与原速度方向相同。因此最后一段 V_4Q 与第一段 QV_1 位于同一直线上，且方向相同。所以 $V_4 \to Q \to V_1$ 是同一段直线路程，圆盘的路径是一条从台边上的 V_4 出发、顺时针在其余三条台边反弹后回到 V_4 的闭合路径。第 (ii) 部分的推导显示，这种路径的斜率由其起点唯一决定。每条这样的路径在每条边上各有一个顶点，所以它们全都是第 (ii) 部分中的平行四边形 Π_a（$0 < a < L$），而 Q 位于 Π_a 的某条边上。反之，若 Q 位于某个 Π_a 的一条边上，从 Q 沿该边按行进方向击出即可。因此我们要数出通过 Q 的（平行四边形，边）组合的数目。

*当 z 改变时，各边如何扫过台面。*当 z 由 0 增至 ∞，a 由 0 增至 L。由上面的公式，y_{W_1} 由 0 增至 B，$x_{W_2} = L - a$ 由 L 减至 0，而 $y_{W_3} = B/\big(1 + ea/(L-a)\big)$ 由 B 减至 0。所以 Π_a 起初是通过 O 的对角线（当 $a \to 0$ 时，各顶点聚集到角位 O 和 (L, B)），最后变成另一条对角线。现在考虑其中一条边，例如 PW_1。它在角位 O 处截出一个三角形，两条直角边分别是沿底边的 a 和沿左边向上的 y_{W_1}。两条直角边都随 a 严格递增，所以这些三角形严格地一个套一个，其斜边永不相交。它们由无到有连续地增长，直至成为台面在由 $(L, 0)$ 至 $(0, B)$ 的对角线以下的一半。准确地说，若 $Q = (q_1, q_2)$ 位于这一半之中，则函数

$$f(a) = \frac{q_1}{a} + \frac{q_2}{y_{W_1}(a)}$$

是连续（continuous）且严格递减（strictly decreasing）的。当 $a \to 0$ 时它趋于 ∞，当

$a \to L$ 时它趋于 $q_1/L + q_2/B < 1$，所以恰好有一个 a 使 $f(a) = 1$（即 Q 位于 PW_1 上）。若 Q 不在这一半之中，则对每个 a 都有 $f(a) > 1$，PW_1 永不通过 Q。同样的论证适用于每条边：

- PW_1 扫过由 $(L, 0)$ 至 $(0, B)$ 的对角线以下的半个台面一次；
- W_2W_3 扫过该对角线以上的半个台面一次；
- W_1W_2 扫过由 O 至 (L, B) 的对角线以上的半个台面一次；
- W_3P 扫过该对角线以下的半个台面一次。

一个不在台边上、也不在对角线上的点 Q，恰好位于每条对角线的某一侧。所以恰好有两条边通过它，而且分属两个不同的平行四边形。这就是那两种击法：每一种都使圆盘沿该边出发，顺时针绕经四条台边后回到 Q。（若 Q 位于对角线上，其中一种击法会退化为射向角位的一击。）

验算

- 当 $e = 1$ 时，对每个 z 斜率都是 $k = B/L$，所以所有边都与对角线平行。这正是对完全弹性台边使用常见的镜像展开（unfolding）方法所得的结果。
- 取 $L = 2$、$B = 1$、$e = \frac{1}{2}$、$z = 1$：则 $a = 1$、$k = \frac{1}{3}$，碰撞点为 $(0, \frac{1}{3})$、$(1, 1)$ 及 $(2, \frac{2}{3})$。最后一段有 $\mathrm{d}x/\mathrm{d}y = e/k = \frac{3}{2}$，所以下降 $\frac{2}{3}$ 时向回移动 1，到达 $(1, 0) = P$。
- 在一个逐次计算圆盘由一边到另一边的数值模拟中，三个部分的结论都成立：随机选取的台面、e 值和 z 值都给出预测的碰撞点，而随机选取的内点都恰好位于两条闭合路径上。

常见错误

- 使用镜像（展开）技巧：当 $e < 1$ 时，反射角不等于入射角，所以这技巧失效。
- 由 $x_{\text{end}} = a$ 求得 k 后，没有验证每次碰撞是否落在预期的边上。
- 在第 (iii) 部分中，假设从 Q 出发的击法必须沿某一特定的边开始。正是第 (i) 部分的论证把第一段与最后一段连成同一段直线路程。

STEP 之外

这类路径称为撞球（billiards）问题。当 $e = 1$ 时，它本身就是一个完整的学科，以反射和展开来研究，并通向动力系统（dynamical systems）。这里的台边会损失能量，而“一

圈前的速度 $\mapsto$ 一圈后的速度”这个映射纯粹是乘以 e^2 的缩放。这就是为何对每个 e 都存在闭合轨道，尽管圆盘每绕一圈都会减速。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对冲量和恢复有仔细的讲解。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）收录了更多同类型的题目。

详细解答

STEP II 1987，第 14 题

力学 • 圆锥上的弹性带

> **STEP II 1987，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/87-s2-q14

关键思路

由于橡皮圈在圆锥上始终保持为一个水平圆，它的整个状态只需一个数便可描述：它沿母线（generator）量度与顶点（vertex）的距离 s。因此我们只需追踪一小段元素。张力（tension）产生的向内拉力有一个沿圆锥面的分量，推动该元素沿母线移动；另有一个垂直于圆锥面的分量，被光滑的圆锥抵消。把 s 换成长度 x，便得到简谐运动（simple harmonic motion），而橡皮圈在四分之一个周期（period）后变得松弛（slack）。

解答

模型。 我们把橡皮圈视为一条质量为 m、沿圆周均匀分布的细弹性弦。圆锥是光滑的，所以它对每一小段的反作用力都垂直于圆锥面；我们亦忽略重力。由对称性，每一小段的运动方式都相同，所以橡皮圈始终是一个以轴为圆心的圆。设它沿母线量度与顶点的距离为 $s(t)$。由于每条母线与轴成角 α，它的半径便是 $s\sin\alpha$，因此

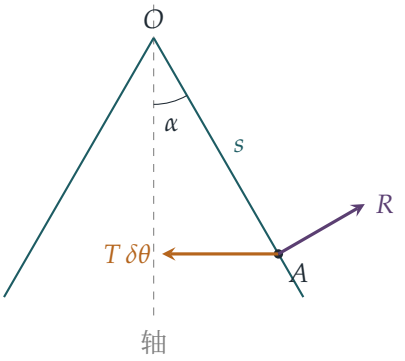
$$x = 2\pi s \sin\alpha.$$

当橡皮圈被拉伸时（$x > l$），胡克定律（Hooke's law）给出张力 $T = \dfrac{\lambda(x-l)}{l}$。

作用于一小段的力。 取一小段在轴上所对的角为 $\delta\theta$。它占整条橡皮圈的 $\delta\theta/2\pi$，所以其质量为 $\dfrac{m\,\delta\theta}{2\pi}$。作用于它的力有两个：

- 题目给出、由张力产生的力 $T\,\delta\theta$，垂直于轴并指向轴；
- 圆锥的法向反作用力（normal reaction）R，垂直于经过该小段的母线。

下图显示包含轴和该小段的平面。



该小段被限制在圆锥面上，所以只能沿母线 OA 移动。我们沿这个方向及其垂直方向分解力。力 $T\,\delta\theta$ 垂直于轴，而母线与轴成角 α，所以 $T\,\delta\theta$ 与方向 $\overrightarrow{AO}$ 之间的夹角是 $90^\circ - \alpha$。因此：

- 沿 $\overrightarrow{AO}$（指向顶点）的方向，$T\,\delta\theta$ 的分量是 $T\,\delta\theta\cos(90^\circ-\alpha) = T\,\delta\theta\sin\alpha$，而 R 在此方向没有分量；
- 在垂直于圆锥面的方向，由于该方向没有运动，$R = T\,\delta\theta\cos\alpha$。

由于 $R \geq 0$，橡皮圈确实一直紧贴圆锥。

> ? **想一想** 为甚么我们不能直接说向内的力 $T\,\delta\theta$ 令半径缩小?

该小段不能直接向轴移动，因为圆锥挡住了去路。只有 $T\,\delta\theta$ 沿圆锥面的部分会产生加速度，其余部分被反作用力抵消。因子 $\sin\alpha$ 正是由此而来；这与光滑斜面上的珠子沿斜面只受 $mg\sin\theta$ 的道理相同。

运动方程。 对该小段沿母线应用牛顿第二定律（Newton's second law），并取 s 以远离顶点为正方向，得

$$\frac{m\,\delta\theta}{2\pi}\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -T\,\delta\theta\sin\alpha \qquad \Longrightarrow \qquad \frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{2\pi T\sin\alpha}{m}.$$

$\delta\theta$ 被消去,所以每一小段都服从同一条方程,这证实了橡皮圈保持圆形。由于 $x = 2\pi s\sin\alpha$ 而 α 是常数，所以

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = 2\pi\sin\alpha\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{4\pi^2 T\sin^2\alpha}{m} = -\frac{4\pi^2\lambda}{ml}(x-l)\sin^2\alpha,$$

其中用了 $T = \lambda(x-l)/l$。此式在橡皮圈拉紧时成立，即对 $0 < t < t_0$ 成立，整理后便得所求方程。 □

求 t_0。设 $y = x - l$ 及 $\omega^2 = \dfrac{4\pi^2\lambda\sin^2\alpha}{ml}$，则 $\ddot{y} = -\omega^2 y$，即绕 $y = 0$ 的简谐运动。设 $x_0 > l$ 为初始长度（橡皮圈是被推上去的，所以开始时处于拉伸状态）。橡皮圈由静止释放，所以 $\dot{y}(0) = 0$，故

$$x - l = (x_0 - l)\cos\omega t.$$

当 x 首次降至 l，即 $\cos\omega t = 0$ 首次成立时，橡皮圈变得松弛，所以 $\omega t_0 = \frac{\pi}{2}$。由于 $\omega = 2\pi\sin\alpha\sqrt{\lambda/(ml)}$，

$$t_0 = \frac{\pi}{2\omega} = \frac{1}{4\sin\alpha}\sqrt{\frac{ml}{\lambda}}.$$

留意 t_0 与橡皮圈被拉伸了多少无关，这是简谐运动的典型特征：无论振幅（amplitude）多大，四分之一周期都是一样的。

验算

- 单位：λ 是一个力，所以 ml/λ 的单位是 $\mathrm{kg\,m/N} = \mathrm{s}^2$，而 t_0 是一个时间。
- 极限情况：当 $\alpha \to \frac{\pi}{2}$ 时，圆锥变平成为一个平面，橡皮圈直接向中心收缩，$t_0 = \frac{1}{4}\sqrt{ml/\lambda}$。当 $\alpha \to 0$ 时，圆锥变成圆柱，张力没有沿圆柱面的分量，而 $t_0 \to \infty$，正合预期。
- 数值验证：取 $\alpha = 30°$、$m = 1$、$l = 1$ 及 $\lambda = 4$，公式给出 $t_0 = \frac{1}{4}$。对随机选取的参数，以数值方法逐步求解该小段的运动方程（并另外在三维空间中直接计算一条精细多边形橡皮圈所受的净张力），均能准确到多个小数位地重现因子 $\sin\alpha$ 及 t_0。

常见错误

- 忘记圆锥的反作用力，对半径 r 写出 $\frac{m\,\delta\theta}{2\pi}\ddot{r} = -T\,\delta\theta$。这样会完全漏掉因子 $\sin^2\alpha$。
- 把该小段的质量写成 m，而不是 $\frac{m\,\delta\theta}{2\pi}$。
- 以整个周期或半个周期作为 t_0：橡皮圈由最大拉伸开始，所以经过四分之一周期便回到自然长度（natural length）。

STEP 之外

题目给出的力 $T\,\delta\theta$ 是一个普遍事实的小角度版本：一条张力为 T、曲率半径（radius of curvature）为 ρ 的弦，每单位长度向侧面施加大小为 T/ρ 的力。把同一想法推广到曲面，便能解释为何肥皂泡内的压强大于外面的压强（拉普拉斯定律（Laplace's law））。受限于曲面上的运动，最简洁的处理方法是拉格朗日力学（Lagrangian mechanics），在其中反

作用力根本无需出现。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）收录了许多关于受张力的弦及受约束运动的题目。

详细解答

STEP II 1987，第 15 题

概率与统计 · 几何概率

> **STEP II 1987，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/87-s2-q15

关键思路

先不要想位置，改为考虑时间。每辆车都会在一段时间区间内占用交叉口，而碰撞发生的充要条件，正是这两段时间区间有重叠。把这个重叠条件写成两条不等式（inequality），(u_1, u_2) 平面上的碰撞区域就变成夹在两条过原点直线之间的楔形区域。由于 u_1, u_2 在 $(0, V)$ 上均匀分布（uniform），点 (u_1, u_2) 便在一个正方形上均匀分布，所以所求概率（probability）就是正方形中落在楔形区域内的面积比例，而题目的三种情况，对应的正是两条直线从正方形哪一边穿出。

解答

碰撞条件

我们把火车和货车看作以恒定速率 $u_1, u_2 > 0$ 沿两条相交于一点的直路行驶（交叉口的宽度忽略不计，或已计入车长之内）。碰撞发生当且仅当在同一时刻，两辆车各有某部分正处于交叉口。

火车车头在时刻 d_1/u_1 到达交叉口，车尾在时刻 $(d_1 + l_1)/u_1$ 驶离，因此火车占用交叉口的时间为

$$\frac{d_1}{u_1} < t < \frac{d_1 + l_1}{u_1}.$$

同样，货车占用交叉口的时间为 $d_2/u_2 < t < (d_2 + l_2)/u_2$。两个开区间（open interval）$(a_1, b_1)$ 与 (a_2, b_2) 有重叠，当且仅当每一个都在另一个结束之前开始：$a_1 < b_2$ 且 $a_2 < b_1$。所以碰撞发生当且仅当

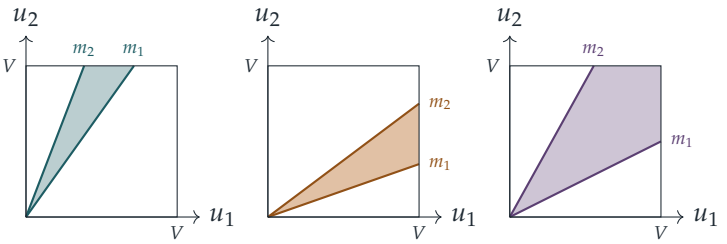
$$\frac{d_1}{u_1} < \frac{d_2 + l_2}{u_2} \quad \text{且} \quad \frac{d_2}{u_2} < \frac{d_1 + l_1}{u_1}.$$

两边乘以正数 u_1u_2 再整理，即得

$$\boxed{\frac{d_2}{d_1+l_1}\,u_1 \;<\; u_2 \;<\; \frac{d_2+l_2}{d_1}\,u_1.}$$

（若某条不等式取等号，两车只是车头碰车尾地刚好接触；这种情况的概率为零，所以算不算进去都没有影响。）

记 $m_1=\dfrac{d_2}{d_1+l_1}$ 及 $m_2=\dfrac{d_2+l_2}{d_1}$。由于 $d_2<d_2+l_2$ 且 $d_1+l_1>d_1$，有 $m_1<m_2$，因此碰撞区域就是第一象限中夹在直线 $y=m_1x$ 与 $y=m_2x$ 之间的楔形区域。下图以阴影显示它在正方形 $0<x,y<V$ 内的部分，这正是下一步所需要的。



(A) $d_1+l_1\le d_2$　(B) $d_2+l_2\le d_1$　(C) 其他情况

火车车尾比货车车头更接近交叉口时的概率

由于 u_1 与 u_2 互相独立（independent），且各自在 $(0,V)$ 上均匀分布，数对 (u_1,u_2) 在正方形 $(0,V)^2$ 上的联合密度（joint density）为常数 $1/V^2$。所以碰撞的概率为

$$\frac{\text{楔形区域在正方形内的面积}}{V^2}.$$

想一想 这两条直线分别从正方形的哪一边穿出?

若 $m\ge1$，直线 $y=mx$ 从正方形的上边穿出（于 $(V/m,\,V)$）；若 $m\le1$，则从右边穿出（于 $(V,\,mV)$）。在这情况下 $d_1+l_1<d_2$，所以 $m_1=d_2/(d_1+l_1)>1$，从而 $m_2>m_1>1$：两条直线都从上边穿出，如图 (A) 所示。

此时楔形区域在正方形内的部分，是顶点为 $(0,0)$、$(V/m_2,V)$ 和 $(V/m_1,V)$ 的三角形。它在上边的底长为 $V/m_1-V/m_2$，高为 V，因此

$$\begin{aligned}\mathrm{P}(\text{碰撞}) &= \frac{1}{V^2}\cdot\frac{1}{2}V\left(\frac{V}{m_1}-\frac{V}{m_2}\right)=\frac{1}{2}\left(\frac{d_1+l_1}{d_2}-\frac{d_1}{d_2+l_2}\right)\\ &=\frac{(d_1+l_1)(d_2+l_2)-d_1d_2}{2d_2(d_2+l_2)}=\frac{l_1l_2+l_2d_1+l_1d_2}{2d_2(l_2+d_2)},\end{aligned}$$

因为 d_1d_2 项互相抵消。这就是所求的结果。

其余两种情况

货车车尾比火车车头更接近交叉口， $d_2 + l_2 < d_1$。此时 $m_2 = (d_2 + l_2)/d_1 < 1$，所以 $m_1 < m_2 < 1$，两条直线都从右边穿出，如图 (B) 所示。楔形区域是顶点为 $(0,0)$、$(V, m_1 V)$、$(V, m_2 V)$ 的三角形：在边 $x = V$ 上的底长为 $(m_2 - m_1)V$，高为 V。所以

$$\mathrm{P}(\text{碰撞}) = \frac{1}{2}(m_2 - m_1) = \frac{1}{2}\left(\frac{d_2 + l_2}{d_1} - \frac{d_2}{d_1 + l_1}\right) = \frac{l_1 l_2 + l_1 d_2 + l_2 d_1}{2 d_1 (l_1 + d_1)}.$$

这正是把第一个答案中火车与货车的角色对调（$l_1 \leftrightarrow l_2$，$d_1 \leftrightarrow d_2$）所得的结果；这是理所当然的，因为对调两辆车就相当于把图形沿 $y = x$ 作反射。

两辆车都不完全在对方前面， $d_2 \le d_1 + l_1$ **且** $d_1 \le d_2 + l_2$。此时 $m_1 \le 1 \le m_2$，所以直线 $y = m_1 x$ 从右边穿出，而 $y = m_2 x$ 从上边穿出，如图 (C) 所示。最简单的做法是从正方形中减去两个不会碰撞的三角形：在 $y = m_1 x$ 下方的那个面积为 $\frac{1}{2} V \cdot m_1 V$，而在 $y = m_2 x$ 左方（即它与 u_2 轴之间）的那个面积为 $\frac{1}{2} V \cdot V / m_2$。因此

$$\mathrm{P}(\text{碰撞}) = 1 - \frac{m_1}{2} - \frac{1}{2 m_2} = 1 - \frac{d_2}{2(d_1 + l_1)} - \frac{d_1}{2(d_2 + l_2)}.$$

在各情况的交界处，答案互相吻合：当 $d_2 = d_1 + l_1$（即 $m_1 = 1$）时，第一条和第三条公式都给出 $\frac{1}{2} - \frac{d_1}{2(d_2 + l_2)}$；在 $d_1 = d_2 + l_2$ 处亦然。留意所有概率都与 V 无关：把两个速率按同一比例放大或缩小，并不会改变两车是否相撞。

验算

- 取 $l_1 = 1$、$l_2 = 2$、$d_1 = 1$、$d_2 = 5$（情况 A），公式给出 $\frac{9}{70} \approx 0.1286$；以 4×10^5 组随机速率作模拟，得 0.1283。在情况 B 和 C 中的模拟（例如 $l_1 = l_2 = 1$、$d_1 = 1$、$d_2 = 1.5$，得 0.425）亦吻合至 0.002 以内。
- 在情况 A 中令 $l_1, l_2 \to 0$，概率为 0：两个点只有在 $u_2/u_1 = d_2/d_1$ 恰好成立时才会在交叉口相遇。
- 在情况 C 中取 $d_1 = d_2$ 而两车都很长，则 $m_1 \to 0$ 且 $m_2 \to \infty$，所以概率趋于 1，这是合理的。

常见错误

- 只写一条不等式，例如“火车在货车驶离交叉口之前到达”；两个区间有重叠需要两个条件同时成立。
- 在情况 C 中套用三角形面积公式，但那里的区域其实是四边形；应先确定每条直线从正方形哪一边穿出。

STEP 之外

这里的方法是把关于独立均匀随机变量的概率转化为面积，称为几何概率（geometric probability）；它是把 $\mathrm{P}((X,Y)\in A)$ 计算为联合密度的二重积分（double integral）的最简单例子。同一想法也是一些经典问题的基础，例如布丰（Buffon）投针问题，以及两人在随机时刻到达的“相遇问题”。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（《概率导论》）正是以这种方式发展联合密度及其应用。

详细解答

STEP II 1987，第 16 题

概率与统计 • 贝叶斯定理

STEP II 1987，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/87-s2-q16

关键思路

这是贝叶斯定理（Bayes' theorem）最朴素的形式。两个“是”的报告可以有两种来源：说法属实而两位朋友都说真话，或者说法不实而两人都说谎。我们按每条途径的先验概率（prior probability）和出现这对报告的概率为它加权，然后问总数中有多大比例来自“属实”这条途径。第二部分的做法相同。唯一的新要点是，朋友的回答只涉及 A，所以我们必须按 A 是否在雪柜里，把三种可能情况分类。

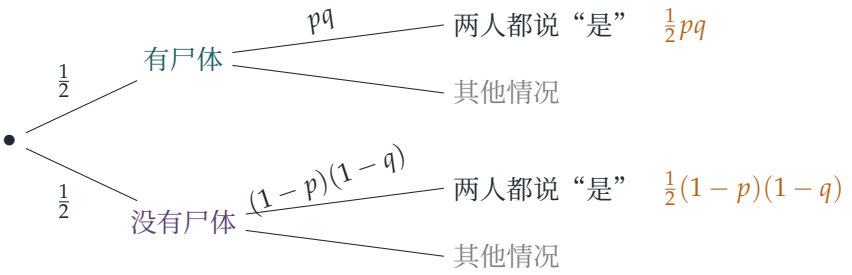
解答

第一个估计

设 F 为雪柜里有尸体这一事件，Y 为 P 和 Q 都说有尸体这一事件。今天早上 $\Pr(F) = \Pr(F') = \frac{1}{2}$。

我们把每位朋友看作回答一条是非题：说真话就是给出正确答案，说谎就是给出错误答案。朋友是否诚实与雪柜里有甚么无关，而且 P 和 Q 的行为互相独立（independent）。所以：

- 若有尸体，两人都说“是”当且仅当两人都说真话，所以 $\Pr(Y \mid F) = pq$；
- 若没有尸体，两人都说“是”当且仅当两人都说谎，所以 $\Pr(Y \mid F') = (1-p)(1-q)$。



根据条件概率（conditional probability）的定义（贝叶斯定理），

$$\begin{aligned}\Pr(F \mid Y) &= \frac{\Pr(F)\Pr(Y \mid F)}{\Pr(F)\Pr(Y \mid F) + \Pr(F')\Pr(Y \mid F')} = \frac{\frac{1}{2}pq}{\frac{1}{2}pq + \frac{1}{2}(1-p)(1-q)} \\ &= \frac{pq}{pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 2pq},\end{aligned}$$

这就是所求的估计。由于先验概率是平分的，因子 $\frac{1}{2}$ 互相抵消。

> **想一想** 为甚么答案不就是 pq，即两人都说真话的概率?

因为我们已经知道一些信息：他们的说法一致。恰好一人说谎的结果现在已不可能出现，因为那会产生一个“是”和一个“否”。我们必须在剩下的情况上重新按比例调整。在两条剩下的途径中，“两人都说真话”的权重是 pq，“两人都说谎”的权重是 $(1-p)(1-q)$。除以两者之和，正是这个重新调整。（先验概率 $\frac{1}{2}$ 也有作用。若先验概率不平分，它们就不会抵消。）

第二个估计

现在肯定有尸体。三个等可能的情况，各自的先验概率为 $\frac{1}{3}$，分别是

$$\text{只有 } A, \qquad \text{只有 } B, \qquad A \text{ 和 } B \text{ 都在.}$$

设 Y_A 为两位朋友都说 A 在雪柜里这一事件。按照与之前相同的模型，他们的回答只取决于 A 是否在雪柜里：

- 在“只有 A”和“两者都在”的情况下，A 在雪柜里，所以每种情况都有 $\Pr(Y_A) = pq$；
- 在“只有 B”的情况下，A 不在，所以 $\Pr(Y_A) = (1-p)(1-q)$。

因此

$$\begin{aligned}\Pr(\text{两者都在} \mid Y_A) &= \frac{\frac{1}{3}pq}{\frac{1}{3}pq + \frac{1}{3}(1-p)(1-q) + \frac{1}{3}pq} \\ &= \frac{pq}{2pq + 1 - p - q + pq} = \frac{pq}{1 - p - q + 3pq}.\end{aligned}$$

留意这些报告并没有直接告诉我们任何关于 B 的信息。它们唯一的作用，是改变 A 在场的概率。根据第一部分（以先验概率 $\frac{2}{3}$ 代替 $\frac{1}{2}$），这个概率变成 $2pq/(1-p-q+3pq)$。已知 A 在场时，“只有 A”和“两者都在”仍然等可能，所以把它减半。这样得出相同的答案。

验算

- 若 $p=q=\frac{1}{2}$，朋友的话毫无用处，答案回到先验概率：$\frac{1}{4}/\frac{1}{2}=\frac{1}{2}$ 及 $\frac{1}{4}/\frac{3}{4}=\frac{1}{3}$。
- 若 $p=q=1$，朋友绝不出错，答案变成 1（有尸体）和 $\frac{1}{2}$（A 在，所以 B 是否也在是一半一半）。
- 当 $p=q=\frac{3}{4}$ 时，答案是 $\frac{9}{10}$ 和 $\frac{9}{19}$。对 200 组随机的 (p,q)，逐一列举所有状态与说真话情况组合的精确计算均与答案吻合。以 $p=0.7$、$q=0.6$ 进行 4×10^5 次试验的模拟，第二部分得出 0.4375，与 $\frac{0.42}{0.96}=0.4375$ 相符。

常见错误

- 答 pq，忘记了“两人都说谎”同样会产生两个“是”。
- 在第二部分，分子只放“两者都在”的情况，却在分母中漏掉“只有 A”的情况。这两种情况都与 A 在场相符。

STEP 之外

这就是贝叶斯更新（Bayesian updating）。先验信念乘以证据的*似然*（likelihood），然后归一化。一个简洁的写法是胜算（odds）形式：后验胜算等于先验胜算乘以似然比（likelihood ratio）$\frac{pq}{(1-p)(1-q)}$。这样写就清楚显示，互相独立的证人只需把各自的比率相乘。同样的计算是医学检测、垃圾邮件过滤器以及法庭上综合证据的基础。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）有一章精彩地讲解条件概率，收录了许多正是这类风格的趣题。

详细解答

STEP III 1987，第 1 题

纯数学 · 阶乘的整除性

> **STEP III 1987，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/87-s3-q1

关键思路

阶乘（factorial）$(n-1)! = 1 \times 2 \times \cdots \times (n-1)$ 包含了每一个小于 n 的正整数。如果 n 可以写成两个小于 n 的不同数的乘积，这两个数都会在阶乘中出现，于是 n 整除这个阶乘。所以问题的核心其实是：哪些 n 不能这样拆开？质数（prime）显然不能，而质数的平方则要仔细看看。亲手试算最初十来个数值，答案便呼之欲出。

解答

先做实验

值得先列一个简单的表：

n	1	2	3	4	5	6	7	8
$(n-1)!$	1	1	2	6	24	120	720	5040
$n \mid (n-1)!$？	是	否	否	否	否	是	否	是

不成立的情况是 $2, 3, 4, 5, 7$：即目前为止的质数，再加上 4。我们现在证明这个规律会一直延续下去。

断言。 n 不整除 $(n-1)!$，当且仅当 n 是质数或 $n = 4$。

$n = 1$

这时 $(n-1)! = 0! = 1$，而 1 整除 1。所以 $n = 1$ 不属于这个集合。

$n = p$ 为质数

$(p-1)!$ 的每个因子 $1, 2, \ldots, p-1$ 都是小于 p 的正数，所以没有一个能被 p 整除。由于 p 是质数，假如 p 整除这个乘积，它就必须整除其中一个因子（欧几里得引理，Euclid's lemma）。因此 $p \nmid (p-1)!$，即每个质数都属于这个集合。

n 为合成数

设 p 为 n 的最小质因数，并写成 $n = pm$。因为 n 是合成数（composite），所以 $m > 1$，而且 m 的每个质因数都不小于 p，故 $m \geq p$。以下分两种情况。

情况 1： $m > p$。 这时 $1 < p < m < n$，所以 p 和 m 是数列 $1, 2, \ldots, n-1$ 中两个不同的数。因此它们的乘积 $pm = n$ 整除 $(n-1)!$。

情况 2： $m = p$，**即** $n = p^2$。

> **想一想** 为甚么情况 1 的论证在这里行不通？可以用甚么来代替？

因子 p 在数列 $1, \ldots, n-1$ 中只出现一次，所以不能把它用两次。我们改用 p 和 $2p$。若 $p \geq 3$，则

$$2p < p \cdot p = n,$$

所以 p 和 $2p$ 是数列 $1, 2, \ldots, n-1$ 中两个不同的数。因此它们的乘积 $2p^2 = 2n$ 整除 $(n-1)!$，于是 n 也整除它。

若 $p = 2$，这个论证便失效，因为 $2p = 4 = n$ 不在数列之中。事实上，$n = 4$ 时得 $3! = 6$，而 $4 \nmid 6$。所以 4 属于这个集合。

结论

除 4 以外，每个合成数 n 都整除 $(n-1)!$，而 $n = 1$ 亦然。因此所求的集合是

$$\{\, n : n \text{ 是质数} \,\} \cup \{4\} = \{2, 3, 4, 5, 7, 11, 13, 17, \ldots\}.$$

验算

上面的表在 $n \leq 8$ 时与断言吻合。再试一个质数的平方：$n = 9$ 时得 $8! = 40320 = 9 \times 4480$，正如情况 2 所预测（用的是因子 3 和 6）。用电脑对每个 $n \leq 2000$ 直接计算 $(n-1)! \bmod n$，结果找不到任何违反断言的例外。

常见错误

- 论证“写成 $n = ab$，其中 $a, b < n$，所以 $ab \mid (n-1)!$”时，没有坚持要求 $a \neq b$。这个论证会错误地“证明” $4 \mid 3!$。
- 忘记了 $n = 1$。由于 $0! = 1$，它不属于这个集合。
- 说质数 p 不整除 $(p-1)!$ 是“因为 p 不是其中一个因子”。真正的原因是 p 是质数，而且不整除任何一个因子；合成数即使不是乘积中的因子，也可以整除该乘积。

STEP 之外

对于质数，有一个更精确的结果，称为威尔逊定理（Wilson's theorem）：$(p-1)! \equiv -1 \pmod{p}$。它与本题合起来，便得出一个（非常慢的）质性测试（primality test）：$n > 1$ 是质数，当且仅当 $(n-1)! \equiv -1 \pmod{n}$。威尔逊定理的证明方法，是把每个数与它模 p 的逆元（inverse）配对，这让人初尝整数模 p 的群（group）结构。Davenport 的 *The Higher Arithmetic* 对此有十分精彩的论述。

详细解答

STEP III 1987，第 2 题

纯数学 · 三角函数归约公式

> **STEP III 1987，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/87-s3-q2

关键思路

递推公式（reduction formula）的来源，是先对“边界项” $\cos^m x\cos nx$ 求导，再用复角公式（compound-angle formula）把 $\sin x\cos nx$ 化为含有 $\sin(n-1)x$ 的式子。有了这条公式之后，两部分的关键都在于计算该边界项在积分上下限的值：在 0 和 π 处，它会互相抵消还是保留下来，取决于奇偶性（parity）；而在 $\frac{\pi}{2}$ 处，因子 $\cos^m x$ 令它变为零。

（题目印出的公式写作 $-\cos^m\cos nx$；应为 $-\cos^m x\cos nx$。）

解答

递推公式

用积法则（product rule）对 $\cos^m x\cos nx$ 求导：

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\cos^m x\cos nx\right) = -m\cos^{m-1}x\,\sin x\cos nx \;-\; n\cos^m x\sin nx.$$

第一项还不是我们想要的形式。由复角公式

$$\sin(n-1)x = \sin nx\cos x - \cos nx\sin x \quad \text{可得} \quad \sin x\cos nx = \cos x\sin nx - \sin(n-1)x.$$

代入后，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\cos^m x\cos nx\right) &= -m\cos^m x\sin nx + m\cos^{m-1}x\sin(n-1)x - n\cos^m x\sin nx\\ &= -(m+n)\cos^m x\sin nx + m\cos^{m-1}x\sin(n-1)x.\end{aligned}$$

对两边积分（这里需要 $m\geq 1$ 及 $n\geq 1$，使得 $m-1, n-1$ 均为非负），

$$\cos^m x\cos nx = -(m+n)I_{m,n} + mI_{m-1,n-1},$$

即

$$(m+n)I_{m,n} = -\cos^m x \cos nx + mI_{m-1,n-1},$$

这正是所需的结果（惯常的任意常数已并入不定积分（indefinite integral）之中）。

第 (i) 部分

设 $J_{m,n} = \int_0^\pi \cos^m x \sin nx \,\mathrm{d}x$。把公式应用于 0 与 π 之间，并利用 $\cos^m \pi \cos n\pi = (-1)^m(-1)^n$，对 $m, n \geq 1$ 有：

$$(m+n)J_{m,n} = 1 - (-1)^{m+n} + mJ_{m-1,n-1}.$$

若 m 与 n 的奇偶性相同，则 $m+n$ 为偶数，边界项消失，于是

$$J_{m,n} = \frac{m}{m+n} J_{m-1,n-1}.$$

此外，$m-1$ 与 $n-1$ 的奇偶性仍然相同，所以我们可以继续递推下去。

> **想一想** 这一连串的递推会在哪里停下来？为甚么在那里的值是零？

每一步都把两个指标各减 1，所以经过 $\min(m,n)$ 步之后，其中一个指标为 0，而另一个指标因奇偶性相同，必为偶数。因此只需证明对每个偶数 $k \geq 0$，都有 $J_{k,0} = 0$ 及 $J_{0,k} = 0$。

- $J_{k,0} = \int_0^\pi \cos^k x \sin 0 \,\mathrm{d}x = 0$，因为被积函数（integrand）恒等于零。
- 对偶数 $k \geq 2$，$J_{0,k} = \int_0^\pi \sin kx \,\mathrm{d}x = \left[-\frac{1}{k}\cos kx\right]_0^\pi = \frac{1}{k}\left(1-(-1)^k\right) = 0$。

因此，对 $\min(m,n)$ 使用归纳法（induction），可知只要 m, n 同为偶数或同为奇数，便有 $J_{m,n} = 0$。

或用其他方法：代换 $x = \pi - t$ 把 $\cos^m x$ 变为 $(-1)^m \cos^m t$，把 $\sin nx$ 变为 $-(-1)^n \sin nt$，所以 $J_{m,n} = (-1)^{m+n+1} J_{m,n}$，由此可见当 $m+n$ 为偶数时必有 $J_{m,n} = 0$。被积函数关于 $x = \frac{\pi}{2}$ 反对称（antisymmetric）。

第 (ii) 部分

为了引入余弦，写 $\sin^2 x = 1 - \cos^2 x$：

$$\int_0^{\pi/2} \sin^2 x \sin 3x \,\mathrm{d}x = K_{0,3} - K_{2,3}, \qquad K_{m,n} = \int_0^{\pi/2} \cos^m x \sin nx \,\mathrm{d}x.$$

对 $m \geq 1$，边界项为 $-\left[\cos^m x \cos nx\right]_0^{\pi/2} = 0 - (-1) = 1$，因为 $\cos\frac{\pi}{2} = 0$。所以

$$(m+n)K_{m,n} = 1 + mK_{m-1,n-1} \qquad (m, n \geq 1).$$

由 $K_{0,1} = \int_0^{\pi/2} \sin x \,\mathrm{d}x = 1$ 出发：

$$3K_{1,2} = 1 + K_{0,1} = 2, \qquad K_{1,2} = \tfrac{2}{3},$$
$$5K_{2,3} = 1 + 2K_{1,2} = \tfrac{7}{3}, \qquad K_{2,3} = \tfrac{7}{15}.$$

另外 $K_{0,3} = \left[-\frac{1}{3}\cos 3x\right]_0^{\pi/2} = \frac{1}{3}$，因为 $\cos\frac{3\pi}{2} = 0$。因此

$$\int_0^{\pi/2} \sin^2 x \sin 3x \,\mathrm{d}x = \frac{1}{3} - \frac{7}{15} = -\frac{2}{15}.$$

验算

第 (ii) 部分可以直接用积化和差公式（product-to-sum formulae）计算：$\sin^2 x \sin 3x = \frac{1}{2}\sin 3x - \frac{1}{4}(\sin 5x + \sin x)$，它在 $[0, \frac{\pi}{2}]$ 上的积分为 $\frac{1}{6} - \frac{1}{4}\left(\frac{1}{5} + 1\right) = -\frac{2}{15}$。答案为负数是合理的：在 $(\frac{\pi}{3}, \frac{\pi}{2})$ 上 $\sin 3x < 0$，而这正是 $\sin^2 x$ 最大的地方。用辛普森法则（Simpson's rule）算得 -0.133333，并且对许多 m, n 以数值方法验证了递推公式和第 (i) 部分。

常见错误

- 在 (ii) 中尝试使用代换 $x = \frac{\pi}{2} - t$：它把 $\sin^2 x$ 变为 $\cos^2 t$，却把 $\sin 3x$ 变为 $-\cos 3t$，所以该积分并不属于 $I_{m,n}$ 的形式。
- 忘记递推公式只适用于 $m, n \geq 1$，所以 (i) 中的一连串递推需要另外处理有一个指标等于 0 的基本情况（base case）。

STEP 之外

像这样的递推公式，只是一个更大的故事的开端：三角函数乘积在一整个周期上的积分，正是傅里叶级数（Fourier series）背后的正交关系（orthogonality relations）；在那里，像 (i) 中那样的对称性论证，让人一眼就能判断哪些系数会消失。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（初等微分方程）对傅里叶级数及其在求解微分方程（differential equation）上的应用，有一段浅白易读的介绍。

详细解答

STEP III 1987，第 3 题

纯数学 · 辐角与梅钦公式

STEP III 1987，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/87-s3-q3

本题在 *Beyond STEP* 第 5.1 节梅钦公式中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/machin.html

关键思路

两部分都建基于复数（complex number）的极式（polar form）。在 (i) 中，点 $(|x|, |y|)$ 与原点的距离为 $|z|$，所以 $|x|\cos\alpha + |y|\sin\alpha$ 等于 $|z|$ 乘以一个余弦值。在 (ii) 中，复数相乘时，辐角（argument）相加。因此，若由 $5-\mathrm{i}$ 和 $1+\mathrm{i}$ 这类数组成的乘积结果非常接近实数，它的辐角便会给出一条关于反正切（inverse tangent）的恒等式。唯一真正需要小心的，是辐角中 2π 的不确定性。

解答

第 (i) 部分

设 $r = |z| = \sqrt{x^2+y^2}$。若 $r = 0$，则 $x = y = 0$，各项均为 0，故以下假设 $r > 0$。

*右边的不等式。*两边均为非负数，所以只需比较两边的平方：

$$(|x| + |y|)^2 = x^2 + y^2 + 2|x||y| \geq x^2 + y^2 = |z|^2,$$

这是因为 $|x||y| \geq 0$。取非负平方根，得 $|z| \leq |x| + |y|$，等号成立当且仅当 $x = 0$ 或 $y = 0$。

*左边的不等式。*点 $(|x|, |y|)$ 位于闭第一象限内，与原点的距离为 r，所以存在角 $\varphi \in [0, \frac{\pi}{2}]$ 使得

$$|x| = r\cos\varphi, \qquad |y| = r\sin\varphi.$$

于是，由复角公式（compound-angle formula），

$$|x|\cos\alpha + |y|\sin\alpha = r(\cos\varphi\cos\alpha + \sin\varphi\sin\alpha) = r\cos(\alpha-\varphi) \le r = |z|,$$

因为对任意实数 α 都有 $\cos(\alpha-\varphi) \le 1$。这就证明了两条不等式。左边的等号在 $\alpha = \varphi + 2k\pi$ 时成立，即 $(\cos\alpha, \sin\alpha)$ 与 $(|x|, |y|)$ 方向相同之时。

（左边不等式的另一证法：由柯西–施瓦兹不等式（Cauchy–Schwarz inequality），$|x|\cos\alpha + |y|\sin\alpha \le \sqrt{x^2+y^2}\,\sqrt{\cos^2\alpha+\sin^2\alpha} = |z|$。）

第 (ii) 部分：第一条公式

逐步展开：

$$\begin{aligned}(5-\mathrm{i})^2 &= 25 - 10\mathrm{i} - 1 = 24 - 10\mathrm{i},\\ (5-\mathrm{i})^4 &= (24-10\mathrm{i})^2 = 576 - 480\mathrm{i} - 100 = 476 - 480\mathrm{i},\\ (5-\mathrm{i})^4(1+\mathrm{i}) &= 476 + 476\mathrm{i} - 480\mathrm{i} + 480 = 956 - 4\mathrm{i} = 4(239-\mathrm{i}).\end{aligned}$$

现在比较辐角。设 $\theta = \tan^{-1}\frac{1}{5}$。由于 $5-\mathrm{i}$ 位于第四象限，它的辐角为 $-\theta$，而 $1+\mathrm{i}$ 的辐角为 $\frac{\pi}{4}$。同理，$4(239-\mathrm{i})$ 的辐角为 $-\tan^{-1}\frac{1}{239}$。复数相乘时辐角相加，但可能相差 2π 的整数倍，所以

$$\frac{\pi}{4} - 4\theta = -\tan^{-1}\frac{1}{239} + 2k\pi \quad \text{对于某个整数} k.$$

想一想 我们怎样知道 $k = 0$?

我们估计两边的范围。由于当 $0 < t < \frac{\pi}{2}$ 时 $\tan t > t$，我们有 $0 < \theta < \frac{1}{5}$，所以 $\frac{\pi}{4} - 4\theta$ 介乎 $\frac{\pi}{4} - \frac{4}{5} > -1$ 与 $\frac{\pi}{4} < 1$ 之间。另外，$-\tan^{-1}\frac{1}{239}$ 介乎 -1 与 0 之间。$(-1, 1)$ 内的两个数相差小于 $2 < 2\pi$，所以 $2k\pi = 0$，即 $k = 0$。整理后得

$$\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239},$$

正是所要证明的结果。

第 (ii) 部分：第二条公式

按同样的道理，我们需要一个乘积，其辐角为

$$\frac{\pi}{4} - 3\tan^{-1}\frac{1}{4} - \tan^{-1}\frac{1}{20} - \tan^{-1}\frac{1}{1985},$$

即 $(4-\mathrm{i})^3(20-\mathrm{i})(1985-\mathrm{i})(1+\mathrm{i})$，并希望它是一个正实数。把因子 $1985-\mathrm{i}$ 留到最后：

$$\begin{aligned}(4-\mathrm{i})^2 &= 16-8\mathrm{i}-1 = 15-8\mathrm{i},\\ (4-\mathrm{i})^3 &= (15-8\mathrm{i})(4-\mathrm{i}) = 60-15\mathrm{i}-32\mathrm{i}-8 = 52-47\mathrm{i},\\ (4-\mathrm{i})^3(20-\mathrm{i}) &= 1040-52\mathrm{i}-940\mathrm{i}-47 = 993-992\mathrm{i},\\ (4-\mathrm{i})^3(20-\mathrm{i})(1+\mathrm{i}) &= 993+993\mathrm{i}-992\mathrm{i}+992 = 1985+\mathrm{i}.\end{aligned}$$

所以到目前为止的乘积是 $1985+\mathrm{i}$，而最后一个因子 $1985-\mathrm{i}$ 正是它的共轭（conjugate）：整个乘积为 1985^2+1，是正实数。换一个等价的做法，不乘最后一个因子，直接比较辐角：

$$\frac{\pi}{4}-3\varphi-\psi = \tan^{-1}\frac{1}{1985}+2k\pi, \qquad \varphi=\tan^{-1}\frac{1}{4},\ \psi=\tan^{-1}\frac{1}{20}.$$

同样有 $k=0$：由于 $0<\varphi<\frac14$ 及 $0<\psi<\frac{1}{20}$，左边介乎 $\frac{\pi}{4}-\frac45>-1$ 与 $\frac{\pi}{4}<1$ 之间，而右边减去 $2k\pi$ 后位于 $(0,1)$ 内。整理后得

$$\frac{\pi}{4} = 3\tan^{-1}\frac14+\tan^{-1}\frac{1}{20}+\tan^{-1}\frac{1}{1985}.$$

验算

- 模（modulus）很适合用来检查展开的结果。$|5-\mathrm{i}|^8\,|1+\mathrm{i}|^2 = 26^4\cdot 2 = 913\,952$，而 $956^2+4^2 = 913\,952$ 亦然。同样地，$17^3\cdot 401\cdot 2 = 3\,940\,226 = 1985^2+1$。
- 数值上，$4\tan^{-1}\frac15\approx 0.789582$，减去 $\tan^{-1}\frac{1}{239}\approx 0.004184$ 得 $0.785398\approx\frac{\pi}{4}$。另外，$3\tan^{-1}\frac14+\tan^{-1}\frac{1}{20}\approx 0.784894$，加上 $\tan^{-1}\frac{1}{1985}\approx 0.000504$ 同样得到 $\frac{\pi}{4}$。
- 在 (i) 中，试取 $z=3-4\mathrm{i}$：不等式变成 $3\cos\alpha+4\sin\alpha\le 5\le 7$，而 $3\cos\alpha+4\sin\alpha = 5\cos(\alpha-\varphi)$ 在 $\tan\alpha=\frac43$ 时取得 5。

常见错误

- 把 $\arg(zw)=\arg z+\arg w$ 当作精确的等式，却从不检查 2π 的倍数。正是对各个反正切值的范围估计，才令最后的等式成立。
- 以 $5+\mathrm{i}$ 代替 $5-\mathrm{i}$：这样 $(5+\mathrm{i})^4(1+\mathrm{i}) = -4+956\mathrm{i}$，结果接近纯虚数而非接近实数。正负号必须选得令乘积接近实数。
- 在 (i) 中，忘记 α 是任意的，所以 $\cos\alpha$ 或 $\sin\alpha$ 可能是负数。余弦形式可一次过处理所有 α。

STEP 之外

第一条公式是梅钦公式（Machin's formula，1706 年）。级数 $\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots$ 在 x 很小时收敛得很快，前人把它与梅钦公式结合，以人手计算出 π 的许多位小数。寻找这类公式，其实是关于高斯整数（Gaussian integers）$a + b\mathrm{i}$ 的因式分解问题，当中 $1 + \mathrm{i}$、$5 \pm \mathrm{i}$ 和 $4 \pm \mathrm{i}$ 的表现有如质数。这是代数数论（algebraic number theory）的初步体验。Davenport 的 *The Higher Arithmetic* 讨论了二平方和与高斯整数，而 Needham 的 *Visual Complex Analysis* 则展示了复数相乘的几何意义。

详细解答

STEP III 1987，第 4 题

纯数学 · 旋转体的体积

STEP III 1987，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：

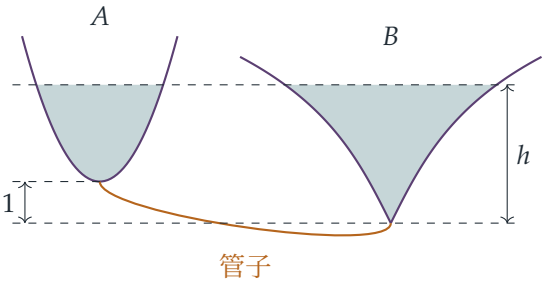
step.maths.org/questions/87-s3-q4

关键思路

水只是由一个漏斗流到另一个漏斗，所以水的体积守恒（conserved）。每个漏斗都是绕 y 轴旋转而成的旋转体（solid of revolution），因此水深为 d 时，水的体积是 $\int_0^d \pi x^2 \,\mathrm{d}y$，其中 x^2 要以 y 表示。求得两个体积作为深度的函数之后，唯一还要小心的，就是两个不同的基准高度：A 和 B 中的水深相差一个单位。

解答

我们假设管子的体积可以忽略不计，并把题目中第二个“closed”理解为“opened”（否则水根本无法流动）。由每个漏斗的底部起向上量度 y。



以深度表示的体积函数

对于 A，曲线为 $y = x^2$，所以在高度 y 处的水平截面是一个圆盘，面积为 $\pi x^2 = \pi y$。A 中水深为 d 时，水的体积是

$$V_A(d) = \int_0^d \pi y \,\mathrm{d}y = \frac{\pi d^2}{2}.$$

对于 B，曲线 $y = 2\sinh^{-1} x$（其中 $x > 0$）的反函数（inverse）是 $x = \sinh \frac{y}{2}$。利用 $\sinh^2 \frac{y}{2} = \frac{1}{2}(\cosh y - 1)$，$B$ 中水深为 d 时，水的体积是

$$V_B(d) = \int_0^d \pi \sinh^2 \frac{y}{2} \,\mathrm{d}y = \frac{\pi}{2} \int_0^d (\cosh y - 1) \,\mathrm{d}y = \frac{\pi}{2} (\sinh d - d).$$

水量守恒

开始时，A 盛有深度为 4 的水，而 B 是空的，所以总体积是

$$V_A(4) = \frac{16\pi}{2} = 8\pi.$$

最后，两个水面都在 B 的底部之上高度 h 处。因此 B 中的水深等于 h。A 的底部比 B 的底部高一个单位，所以 A 中的水深是 $h - 1$。

> ? **想一想** 为甚么两个水面最终会处于同一高度?

两个漏斗和管子构成一个连通而静止的水体。水中的压强只取决于高度，而两个自由水面都处于大气压强之下，所以它们位于同一高度。题目已经告诉我们这一点，但这也解释了为甚么两个水深之差恰好等于两个底部的高度差。

水既没有增加，也没有减少，所以

$$\frac{\pi(h-1)^2}{2} + \frac{\pi}{2} (\sinh h - h) = 8\pi.$$

乘以 $2/\pi$ 并展开，得

$$\begin{aligned} (h-1)^2 + \sinh h - h &= 16 \\ h^2 - 2h + 1 + \sinh h - h &= 16 \\ h^2 - 3h + \sinh h &= 15, \end{aligned}$$

这就是所要求的方程。

中间一行的左方只在 $h \geq 1$ 时才是合理的模型，因为这时 A 才真正盛有水。在这个范围内，它的导数 $2(h-1) + \cosh h - 1$ 是正数，所以方程恰好有一个符合题意的根。数值上 $h \approx 3.33$，所以 A 中的水深由 4 降至约 2.33。

验算

- 以二分法（bisection）求得 $h \approx 3.3272$。于是 $V_A(h-1) + V_B(h) \approx 25.1327$，与 $8\pi \approx 25.1327$ 吻合。
- 对 $\int_0^d \pi \sinh^2(y/2)\,\mathrm{d}y$ 使用辛普森法则（Simpson's rule），所得结果在多个 d 值下都与 $\frac{\pi}{2}(\sinh d - d)$ 一致（例如 $d = 1.7$ 时，两种方法都得出 1.48540）。
- 合理性：$h - 1 < 4$，所以 A 中的水位下降了；水从中流出，水位理应下降。

常见错误

- 在两个漏斗中都用深度 h，忘记了 A 的底部高出一个单位。这样会得出 $h^2 - h + \sinh h = 16$。
- 对 $\pi y^2\,\mathrm{d}x$ 积分（即绕 x 轴旋转），而不是对 $\pi x^2\,\mathrm{d}y$ 积分。
- 把 $y = 2\sinh^{-1} x$ 的反函数写成 $x = 2\sinh y$ 或 $x = \frac{1}{2}\sinh y$。正确的反函数是 $x = \sinh\frac{y}{2}$。

STEP 之外

最后的方程把多项式和双曲函数（hyperbolic function）混在一起，所以没有闭合形式（closed-form）的解，必须用数值方法求解；而上面的单调性（monotonicity）论证，正好保证了二分法或牛顿–拉弗森法（Newton–Raphson）之类的方法能找到那个唯一具物理意义的根。大学一年级的数值分析（numerical analysis）会利用介值定理（intermediate value theorem）和收敛速度（convergence rate）把这一点严格化。至于背后的分析学，可参阅 Abbott 的 *Understanding Analysis*（一本分析学入门教科书）。

详细解答

STEP III 1987，第 5 题

纯数学 · 模算术密码

STEP III 1987，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/87-s3-q5

关键思路

这套密码其实是两个独立的移位密码（shift cipher）：偶数位置用一个（密钥为 n_0），奇数位置用另一个（密钥为 n_1），两者都再加上额外的移位 j。先把已知的移位 j 除去。余下的，就是信息的两半各自移了一个未知常数；而模（modulo）50 的移位是双射（bijection），所以相等的数仍然相等，不同的数仍然不同。那三个 43 正是泄露 n_0 的指纹。之后，取一个模 50 的和，便可确定 n_1。

解答

只有 n_0 和 n_1 模 50 的值才有关系，所以我们取 $0 \le n_0, n_1 \le 49$。每个 a_k 都介乎 0 与 49 之间，所以只要知道 a_k 模 50 的剩余（residue），它的值便确定了。

除去已知的移位

对于偶数位置，设 $c_j = b_{2j} - j$，其中 $j = 1, \ldots, 6$。对于奇数位置，设 $d_j = b_{2j+1} - j$，其中 $j = 0, \ldots, 5$。于是

$$a_{2j} \equiv c_j - n_0, \qquad a_{2j+1} \equiv d_j - n_1 \pmod{50}.$$

由加密后的信息可得：

j	1	2	3	4	5	6
b_{2j}	27	36	35	40	37	48
c_j	26	34	32	36	32	42

j	0	1	2	3	4	5
b_{2j+1}	35	2	15	8	40	24
d_j	35	1	13	5	36	19

求 n_0

在模 50 之下减去一个固定的 n，是一对一（one-to-one）的运算。所以，奇偶性（parity）相同的两个位置上的 a 相等，当且仅当它们对应的 c（或 d）相等。

> ? **想一想** 那三个 43 可以放在哪里?

六个 d_j 各不相同，所以奇数位置上是六个不同的数，其中最多只有一个 43。因此最少有两个 43 在偶数位置。在 c_j 当中，唯一重复的是 32，它恰好出现两次（$j = 3, 5$），不多也不少。所以偶数位置上恰好有两个相等的数，而它们必定就是那两个 43，因为 43 是信息中唯一出现多于一次的数。因此

$$32 - n_0 \equiv 43 \pmod{50} \quad \Longrightarrow \quad n_0 \equiv -11 \equiv 39 \pmod{50}.$$

现在 $a_{2j} \equiv c_j - 39 \equiv c_j + 11$，由此得

$$(a_2, a_4, a_6, a_8, a_{10}, a_{12}) = (37, 45, 43, 47, 43, 3),$$

这是因为 $42 + 11 = 53 \equiv 3$。这六个数全部都在信息之中，这是必然的。

求 n_1

从信息中取走这六个数，余下的就是奇数位置上的数

$$S = \{1, 7, 23, 24, 39, 43\},$$

所以 $\{d_j - n_1\}$ 在模 50 之下必须等于 S。

> ? **想一想** 怎样才可以不必把五十个值逐一尝试，就求出 n_1?

把两边分别加起来。如果两个集合相同，它们的和在模 50 之下也相同：

$$\sum d_j - 6n_1 \equiv \sum_{s \in S} s \pmod{50}, \qquad \text{即} \qquad 109 - 6n_1 \equiv 137.$$

所以 $6n_1 \equiv -28 \equiv 22 \pmod{50}$。全式除以 2（它同时整除模数），得 $3n_1 \equiv 11 \pmod{25}$。由于 $3 \times 17 = 51 \equiv 1$，我们得到 $n_1 \equiv 11 \times 17 = 187 \equiv 12 \pmod{25}$。所以 $n_1 = 12$ 或 $n_1 = 37$。

和的条件只是必要条件（necessary condition），并非充分条件（sufficient condition），所以两个值都要检验。若 $n_1 = 37$，则 $d_0 - 37 = -2 \equiv 48$，而这个数不在信息之中，所以 $n_1 = 37$ 不成立。若 $n_1 = 12$，则

$$(a_1, a_3, a_5, a_7, a_9, a_{11}) \equiv (23, -11, 1, -7, 24, 7) \equiv (23, 39, 1, 43, 24, 7),$$

这正好就是 S。所以 $n_1 = 12$。

信息

把两半交错排列，原来的信息是

$$23,\ 37,\ 39,\ 45,\ 1,\ 43,\ 43,\ 47,\ 24,\ 43,\ 7,\ 3,$$

其中 $n_0 \equiv 39$ 及 $n_1 \equiv 12 \pmod{50}$。以上论证显示这是唯一的答案：n_0 的值别无选择，而 n_1 的两个候选值中只有一个可行。

验算

把答案重新加密。例如 $b_{12} \equiv a_{12} + 39 + 6 = 48$ 及 $b_7 \equiv a_7 + 12 + 3 = 58 \equiv 8$，两者都正确。用电脑把全部 2500 对 (n_0, n_1) 逐一尝试，会发现恰好只有一对，即 $(39, 12)$，能解密出题目所给的那组数，而且得出的正是上面的信息。

常见错误

- 弄错下标。奇数位置由 b_1 开始，对应 $j = 0$；但偶数位置由 b_2 开始，对应 $j = 1$。
- 把 $6n_1 \equiv 22 \pmod{50}$ 除以 6，仿佛 6 是可逆（invertible）的。其实不然，因为 $\gcd(6, 50) = 2$，而这个同余式在模 50 之下有两个解。
- 找到一个可行的 n_1 值，却没有说明其他值都不可行。题目要求你解释所用的方法，所以唯一性需要论证。

STEP 之外

这是凯撒（Caesar）密码，亦即移位密码，而这里的破解方法是频率分析（frequency analysis）的一个小例子：移位会保留字母重复出现的规律。在大学的入门课程中，由 $6n \equiv 22$ 得出两个解的那一步，属于线性同余式（linear congruence）$ax \equiv b \pmod{m}$ 的理论：这类同余式有解，当且仅当 $\gcd(a, m)$ 整除 b。仿射密码（affine cipher）$x \mapsto ax + b$，以及其后的 RSA，都是建基于同样的想法。Davenport 的 *The Higher Arithmetic* 对同余的讲解十分精彩。

详细解答

STEP III 1987，第 6 题

纯数学 · 耦合微分方程

STEP III 1987，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/87-s3-q6

关键思路

处理一对耦合的一阶方程（coupled first-order equations），最好的做法是把它们化为只含一个未知函数的二阶方程。利用第一条方程，可把 y 用 x 和 $\mathrm{d}x/\mathrm{d}t$ 表示，再代入第二条方程，便得到一条关于 x 的受迫振子（forced oscillator）方程。答案的整体性质，就取决于这个振子的固有频率（natural frequency）与外力项 $2\cos t$ 的频率相比如何。

解答

以字母上方加点表示对 t 的导数。

消去 y

由第一条方程，

$$y = \tfrac{1}{5}\left(\dot{x} + 2x\right), \qquad \text{所以} \qquad \dot{y} = \tfrac{1}{5}\left(\ddot{x} + 2\dot{x}\right).$$

把两者代入第二条方程，再乘以 5，

$$\ddot{x} + 2\dot{x} + 5ax - 2\left(\dot{x} + 2x\right) = 10\cos t,$$
$$\ddot{x} + (5a - 4)\,x = 10\cos t.$$

含 $\dot{x}$ 的项互相抵消，所以没有阻尼（damping）：这是一个纯粹的受迫振子，固有频率为 $\sqrt{5a-4}$（当 $5a > 4$ 时）。

转换初始条件

题目给出 $x(0)=0$ 和 $\dot{y}(0)=0$，但关于 x 的方程需要的是 $x(0)$ 和 $\dot{x}(0)$。把 $t=0$ 代入原来的第二条方程：

$$\dot{y}(0)+a\,x(0)-2y(0)=2 \quad\Longrightarrow\quad 0+0-2y(0)=2 \quad\Longrightarrow\quad y(0)=-1.$$

然后由第一条方程在 $t=0$ 的情况，得 $\dot{x}(0)=5y(0)-2x(0)=-5$。留意这两步都没有用到 a 的值，所以

$$x(0)=0, \qquad \dot{x}(0)=-5 \qquad \text{对所有}a.$$

> **想一想** 条件是关于 $\mathrm{d}y/\mathrm{d}t$ 的，而不是关于 y 的。我们该怎样运用它？

微分方程本身已把 x、y 及它们的导数在任何一个时刻的值联系起来。在 $t=0$ 时，第二条方程把关于 $\dot{y}(0)$ 的资料转化为关于 $y(0)$ 的资料，第一条方程再把它转化为 $\dot{x}(0)$，而这正是关于 x 的二阶方程所需要的。

$a=1$ 的情况

这时方程为

$$\ddot{x}+x=10\cos t.$$

余函数（complementary function）是 $A\cos t+B\sin t$。外力项 $\cos t$ 本身就是齐次方程（homogeneous equation）的解（这就是共振（resonance）），所以试用 $C\cos t$ 是行不通的，我们改试 $x_p=t\,(C\cos t+D\sin t)$。求导两次，

$$\ddot{x}_p+x_p=-2C\sin t+2D\cos t,$$

所以 $C=0$ 及 $D=5$，得 $x_p=5t\sin t$。因此

$$x=A\cos t+B\sin t+5t\sin t.$$

由条件 $x(0)=0$ 得 $A=0$。于是 $\dot{x}=B\cos t+5\sin t+5t\cos t$，所以 $\dot{x}(0)=B=-5$。因此

$$\boxed{x=5(t-1)\sin t}$$

至于 y，利用 $y=\frac{1}{5}(\dot{x}+2x)$ 以及 $\dot{x}=5\sin t+5(t-1)\cos t$：

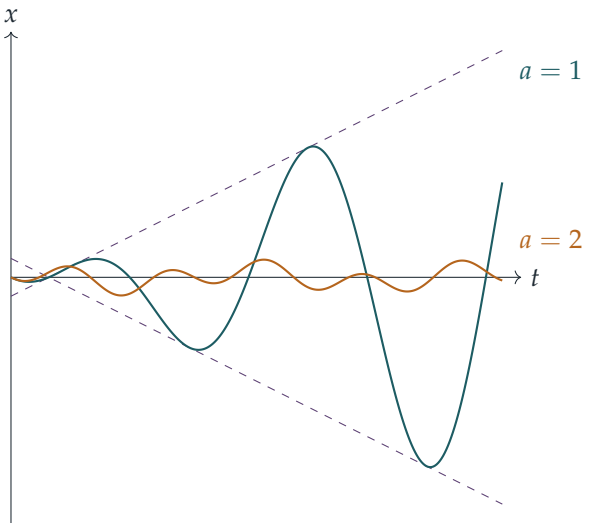
$$y=\sin t+(t-1)\cos t+2(t-1)\sin t, \qquad \text{即} \qquad \boxed{y=(t-1)\cos t+(2t-1)\sin t}$$

简单验算一下：$y(0)=-1$，而由 $\dot{y}=\cos t-(t-1)\sin t+2\sin t+(2t-1)\cos t$ 得 $\dot{y}(0)=1-1=0$，符合要求。

$a > 1$ 的情况

这时 $5a - 4 > 1$，所以可写 $\omega = \sqrt{5a-4}$，其中 $\omega > 1$。方程 $\ddot{x} + \omega^2 x = 10\cos t$ 不再处于共振状态，因为外力频率 1 与固有频率 ω 不同。所以：

- 余函数是 $A\cos\omega t + B\sin\omega t$，振动周期为 $2\pi/\omega$（比 2π 短），而不再是 2π。
- 特积分（particular integral）只是 $\cos t$ 的倍数，没有因子 t，所以不会出现 $t\sin t$ 这类项。
- 因此 x（以及 $y = \frac{1}{5}(\dot{x} + 2x)$）是频率为 1 和 ω 的正弦型函数之和。两者对所有 t 都保持有界（bounded），而不像 $a = 1$ 时那样，振幅随 t 线性增长。只有当 ω 是有理数时，运动才是周期性的；一般而言，它是两个振动的叠加（superposition），两者互相产生“拍”（beat）的现象。



验算

- 把 $x = 5(t-1)\sin t$ 和 $y = (t-1)\cos t + (2t-1)\sin t$ 代回原来的两条方程，两者均恒等成立。逐步数值求解（由 $x = 0$、$y = -1$ 出发的四阶龙格－库塔法（Runge–Kutta））所得结果，在 $0 \le t \le 10$ 上与公式相差不超过 10^{-10}；例如 $x(10) = 45\sin 10 \approx -24.481$。
- 当 $a > 1$ 时，完整的解原来是 $x = \frac{2}{a-1}(\cos t - \cos\omega t) - \frac{5}{\omega}\sin\omega t$。当 $a \to 1$ 时，有 $\omega - 1 \approx \frac{5}{2}(a-1)$ 及 $\cos t - \cos\omega t \approx (\omega - 1)\, t\sin t$，由此重新得到 $5t\sin t - 5\sin t$。共振情况的答案正是非共振情况的极限（limit）。

常见错误

- 直接用 $y(0) = 0$，而没有从 $\dot{y}(0)$ 的条件推出 $y(0) = -1$。
- 当 $a = 1$ 时试用 $C\cos t + D\sin t$ 作为特积分：它是余函数的一部分，所以只会得出 $0 = 10\cos t$。
- 另外引入新的任意常数来独立求解 y：一旦知道 x，y 就由 $y = \frac{1}{5}(\dot{x} + 2x)$ 完全确定。

STEP 之外

这类线性方程组其实就是 $\dot{\mathbf{v}} = M\mathbf{v} + \mathbf{f}(t)$，其中 $\mathbf{v} = (x, y)$，而 M 是常数矩阵；M 的特征值（eigenvalue），在这里是 $\pm\sqrt{4 - 5a}$，决定了自由运动是增长、衰减还是振动。当外力频率与特征值 $\pm\mathrm{i}$ 吻合时，便会出现共振，而这正好就是 $a = 1$。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（微分方程入门教材）既讲解如何用矩阵处理方程组，也讨论受迫振动以及共振和拍的现象。

详细解答

STEP III 1987，第 7 题

纯数学 · 反正切的级数

STEP III 1987，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/87-s3-q7

关键思路

$\tan^{-1} t$ 的级数（series）来自对 $1/(1+x^2)$ 求积分，而 $1/(1+x^2)$ 正是公比为 $-x^2$ 的等比级数（geometric series）之和。把等比级数在有限项之后截断，会留下一个精确的余项（remainder），再对这个余项积分，便得到误差项（error term）。在 $[0,1]$ 上，因子 $1/(1+x^2)$ 介乎 $\frac{1}{2}$ 与 1 之间，于是误差便被夹在两个容易计算的积分之间。

解答

恒等式

对任意实数 x，首项为 1、公比为 $-x^2$ 的有限等比级数给出

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

此式成立，因为公比 $-x^2$ 永远不等于 1。移项整理，得

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - \cdots + (-1)^n x^{2n} + (-1)^{n+1}\frac{x^{2n+2}}{1+x^2}.$$

这是一个精确的恒等式（identity），并非近似。把两边由 0 积分至 t。由于 $\int_0^t \frac{\mathrm{d}x}{1+x^2} = \tan^{-1} t$ 及 $\int_0^t x^{2r}\,\mathrm{d}x = \frac{t^{2r+1}}{2r+1}$，

$$\tan^{-1} t = t - \frac{t^3}{3} + \frac{t^5}{5} - \cdots + \frac{(-1)^n t^{2n+1}}{2n+1} + (-1)^{n+1}\int_0^t \frac{x^{2n+2}}{1+x^2}\,\mathrm{d}x,$$

这正是所求。此式对所有实数 t 均成立。

上下界

记 $S_n(t) = \sum_{r=0}^{n} \frac{(-1)^r t^{2r+1}}{2r+1}$。由上述恒等式，

$$\left|\tan^{-1} t - S_n(t)\right| = \int_0^t \frac{x^{2n+2}}{1+x^2}\,\mathrm{d}x,$$

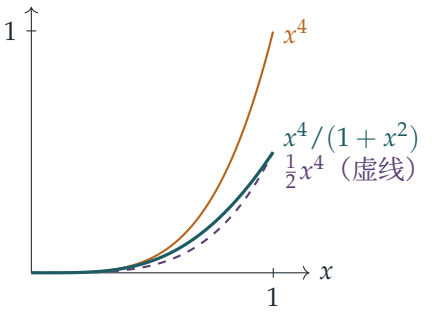
右边的绝对值符号可以略去，因为当 $t \geq 0$ 时，被积函数（integrand）非负，而且积分是由小到大进行的。

> **想一想** 一个无法以闭合形式（closed form）求值的积分，我们可以怎样估计？

用常数来界定被积函数中难处理的部分。当 $0 \leq x \leq 1$ 时，我们有 $1 \leq 1+x^2 \leq 2$，所以

$$\tfrac{1}{2}x^{2n+2} \leq \frac{x^{2n+2}}{1+x^2} \leq x^{2n+2}.$$

下图显示 $n = 1$ 的情况：被积函数（蓝绿色）被夹在 x^4 与 $\frac{1}{2}x^4$ 之间。



把这些不等式由 0 积分至 t（由于 $t \geq 0$，不等号方向保持不变），得

$$\frac{t^{2n+3}}{2(2n+3)} \leq \left|\tan^{-1} t - S_n(t)\right| \leq \frac{t^{2n+3}}{2n+3} \qquad (0 \leq t \leq 1).$$

这就是所要证明的结果。

π 的级数

取 $t = 1$，此时 $\tan^{-1} 1 = \frac{\pi}{4}$ 而 $S_n(1) = \sum_{r=0}^{n} \frac{(-1)^r}{2r+1}$。由上界（upper bound）可得

$$\left|\pi - 4\sum_{r=0}^{n} \frac{(-1)^r}{2r+1}\right| \leq \frac{4}{2n+3} \to 0 \quad \text{当} n \to \infty,$$

所以 $4\sum_{r=0}^{n} \frac{(-1)^r}{2r+1} \to \pi$。

想一想 两个界之中，哪一个告诉我们这个近似很差？

是下界（lower bound）：它说明误差不可能很小。把 $t=1$ 时左边的不等式乘以 4，

$$\left|\pi - 4\sum_{r=0}^{n}\frac{(-1)^r}{2r+1}\right| \geq \frac{4}{2(2n+3)} = \frac{2}{2n+3}.$$

若 $n \leq 98$，则 $2n+3 \leq 199 < 200$，所以 $\frac{2}{2n+3} > \frac{2}{200} = 10^{-2}$。因此对所有 $n \leq 98$，误差至少为 10^{-2}（事实上严格大于此数）。由于误差小于 $4/(2n+3)$，一百项只换来大约两个小数位的准确度，所以用这个级数来计算 π 是非常缓慢的。

验算

- $n=0$，$t=1$：上下界给出 $\frac{1}{6} \leq \left|\frac{\pi}{4}-1\right| \leq \frac{1}{3}$。事实上 $1-\frac{\pi}{4} \approx 0.215$。
- 以数值计算，当 $n=98$ 时误差为 $|\pi - 4S_{98}(1)| \approx 0.01010$，仅仅高于 10^{-2}。当 $n=99$ 时误差约为 0.0099998，仅仅低于此值。所以 98 正好是分界点。我们亦以数值积分（quadrature）对随机的 t 和 n 验证了恒等式，并对数千个随机的 $t \in [0,1]$ 验证了双边的界。

常见错误

- 把 $1/(1+x^2)$ 展开成无穷级数，然后逐项积分。这样做需要 $|x|<1$，还要说明为何可以交换求和与积分的次序。采用带有精确余项的有限形式，则两者皆可避免。
- 用上界来论证误差很大。只有下界才能做到这一点。

STEP 之外

这其实是带积分形式余项的泰勒定理（Taylor's theorem），只是针对一个函数亲手推导出来。$t=1$ 时的级数就是莱布尼兹（Leibniz）级数，又称马德哈瓦（Madhava）–格雷果里（Gregory）级数。它的误差大约为 $1/(2n)$，这是各项递减缓慢的交错级数（alternating series）的典型表现。较快的公式，例如梅钦（Machin）公式 $\frac{\pi}{4} = 4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}$，是把同一个级数用于较小的 t，此时 t^{2n+3} 会迅速缩小。若想了解余项与收敛（convergence）的严格处理，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP III 1987，第 8 题

纯数学 · 平行四边形面积不等式

STEP III 1987，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/87-s3-q8

关键思路

平行四边形（parallelogram）的对角线（diagonal）互相平分（bisect），因此把它分成四个三角形，而这些三角形的面积可以用 $\frac{1}{2}ab\sin C$ 立即写出。这样便得到一条以对角线及其夹角表示面积的公式，题目开首的两个结果都可由此直接推出。集合 A 是两条带状区域相交之处，所以是一个平行四边形。只要求出它的面积，第二个结果便告诉我们它较长的对角线必定够长，而这条对角线的两个端点正是我们要找的点。

解答

以对角线表示面积

设两条对角线的长度为 p 和 q，相交所成的角为 θ，其中 $0 < \theta < \pi$。平行四边形的对角线互相平分，所以把它分成四个三角形，每个三角形都有两条边 $\frac{1}{2}p$ 和 $\frac{1}{2}q$ 在中心相交。其中两个三角形在该处的角为 θ，另外两个则为 $\pi - \theta$。由于 $\sin(\pi - \theta) = \sin\theta$，每个三角形的面积都是 $\frac{1}{2} \cdot \frac{p}{2} \cdot \frac{q}{2}\sin\theta$，所以

$$\text{面积} = 4 \cdot \tfrac{1}{8}pq\sin\theta = \tfrac{1}{2}pq\sin\theta.$$

若 p 和 q 固定，则当 $\sin\theta$ 最大时面积最大，即 $\theta = \frac{\pi}{2}$ 之时。因此面积取最大值，当且仅当两条对角线互相垂直，而最大面积为 $\frac{1}{2}pq$。

以较长对角线表示的上界

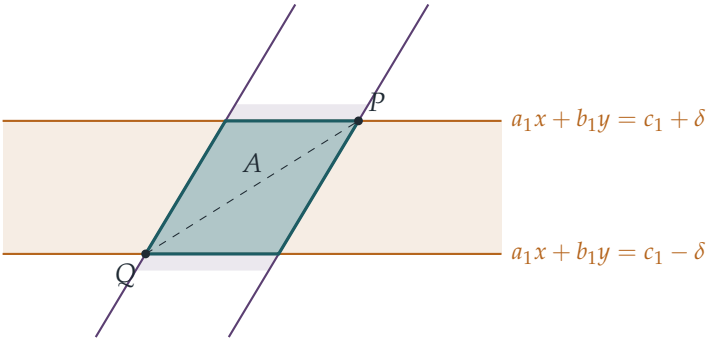
设 $p \geq q$，即 p 是较长的对角线（或两条等长对角线的其中一条）。先用 $\sin\theta \leq 1$，再用 $q \leq p$，得

$$\text{面积} = \tfrac{1}{2}pq\sin\theta \leq \tfrac{1}{2}pq \leq \tfrac{1}{2}p^2.$$

所以面积最多是较长对角线平方的一半。等号成立需要 $\sin\theta = 1$ 及 $q = p$，故只有正方形才能取等号。

集合 A 是平行四边形

假设 $\delta > 0$。（若 $\delta = 0$，则上界为 0，把 A 中任意一点取两次即可。）不等式（inequality）$|a_1x + b_1y - c_1| \leq \delta$ 的意思是 $c_1 - \delta \leq a_1x + b_1y \leq c_1 + \delta$。因此它描述的是两条平行线 $a_1x + b_1y = c_1 \pm \delta$ 之间的闭带状区域（closed strip）。同理，第二条不等式给出第二条带状区域。条件 $a_1b_2 \neq a_2b_1$ 表示法向量（normal vector）$\mathbf{n}_1 = (a_1, b_1)$ 与 $\mathbf{n}_2 = (a_2, b_2)$ 不平行。所以两条带状区域相交，而且 $\mathbf{n}_1$ 和 $\mathbf{n}_2$ 都不是零向量。它们的交集（intersection）A 是一个平行四边形，由两对平行线围成。



图中显示第一条带状区域（琥珀色）和第二条带状区域的一部分（紫红色）。它们的交集 A（蓝绿色）是一个平行四边形，虚线是它较长的对角线。

A 的面积

平行线 $a_1x + b_1y = c_1 \pm \delta$ 之间的距离为

$$h_1 = \frac{2\delta}{\sqrt{a_1^2 + b_1^2}},$$

这里用了点到直线的距离公式。同理，第二条带状区域的阔度为 $h_2 = 2\delta / \sqrt{a_2^2 + b_2^2}$。设 φ 为两组直线之间的夹角，其中 $0 < \varphi < \pi$。在 A 的各边之中，位于第二条带状区域边界上的两条边以角 φ 横越第一条带状区域，所以每条的长度为 $h_1 / \sin\varphi$。这两条边之间

的垂直距离为 h_2，所以

$$\text{面积} A = \frac{h_1 h_2}{\sin\varphi}.$$

想一想 怎样由各系数求出 $\sin\varphi$?

两条直线之间的夹角等于它们法线之间的夹角，或该角的补角（supplement），而两者的正弦值相同。对于向量 $\mathbf{n}_1 = (a_1, b_1)$ 和 $\mathbf{n}_2 = (a_2, b_2)$，它们所张成的平行四边形面积为 $|\mathbf{n}_1||\mathbf{n}_2|\sin\varphi$。这个面积亦等于 $|a_1 b_2 - a_2 b_1|$，即 2×2 行列式（determinant）的绝对值。因此

$$\sin\varphi = \frac{|a_1 b_2 - a_2 b_1|}{\sqrt{a_1^2 + b_1^2}\,\sqrt{a_2^2 + b_2^2}}.$$

代入后，平方根互相抵消：

$$\text{面积} A = \frac{2\delta}{\sqrt{a_1^2 + b_1^2}} \cdot \frac{2\delta}{\sqrt{a_2^2 + b_2^2}} \cdot \frac{\sqrt{a_1^2 + b_1^2}\,\sqrt{a_2^2 + b_2^2}}{|a_1 b_2 - a_2 b_1|} = \frac{4\delta^2}{|a_1 b_2 - a_2 b_1|}.$$

所求的两点

设 PQ 为 A 较长的对角线，并记 $P = (x_1, y_1)$ 及 $Q = (x_2, y_2)$。这两点是 A 的顶点，而带状区域包含其边界线，所以它们都在 A 之内。由第一部分的上界，

$$\frac{4\delta^2}{|a_1 b_2 - a_2 b_1|} = \text{面积} A \le \tfrac{1}{2} PQ^2,$$

所以

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 = PQ^2 \ge \frac{8\delta^2}{|a_1 b_2 - a_2 b_1|},$$

这正是所要证明的。

验算

- 取 $a_1 = b_2 = 1$ 及 $a_2 = b_1 = 0$，则 A 是正方形 $|x - c_1| \le \delta$，$|y - c_2| \le \delta$。它的对角线平方为 $(2\delta)^2 + (2\delta)^2 = 8\delta^2$，刚好等于上界。这是合理的，因为正方形正是第一部分中等号成立的情况。
- 缩放：若把 a_1、b_1 和 c_1 都加倍，第一条带状区域的阔度减半，A 的面积减半，而 $|a_1 b_2 - a_2 b_1|$ 则加倍。公式的变化与此吻合。
- 我们用电脑测试了 10^5 个随机例子。我们解边界方程求出 A 的顶点，再用鞋带公式（shoelace formula）计算面积，每次都与 $4\delta^2/|a_1 b_2 - a_2 b_1|$ 相符。（较长对角线）^{2} 与上界之比的最小值为 1.0007，从未低于 1。

常见错误

- 把带状区域的阔度当作 2δ。其实 $|a_1x + b_1y - c_1| \leq \delta$ 的阔度是 $2\delta / \sqrt{a_1^2 + b_1^2}$。
- 忘记说明为何所选的点在 A 之内。原因是它们是顶点，而 A 包含其边界。
- 把 A 的面积写成 h_1h_2，这只在两条带状区域互相垂直时才正确。

STEP 之外

这个面积计算是一个基本事实的特例：矩阵为 M 的线性映射（linear map）会把所有面积乘以 $|\det M|$。这里的映射 $(x, y) \mapsto (a_1x + b_1y,\ a_2x + b_2y)$ 把 A 映到一个面积为 $4\delta^2$ 的正方形，所以 A 的面积为 $4\delta^2 / |\det M|$。这就是在二重积分（double integral）中作变量代换时遇到的雅可比（Jacobian）因子。Strang 的 *Introduction to Linear Algebra*（线性代数入门）正是以这种方式，把行列式看作面积和体积来阐述。

详细解答

STEP III 1987，第 9 题

纯数学 · 群的直积

> **STEP III 1987，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/87-s3-q9

关键思路

$G \times H$ 中的一切运算都是逐个坐标进行的，所以每一条群公理（group axiom）以及 $G \times H$ 的每一项性质，都归结为 G 和 H 各自具备同样的性质。至于判断真假的部分，答“真”需要给出一般的证明，而答“假”则只需找出一个确凿的障碍。同构（isomorphic）的群，凡是单靠乘法就能描述的性质都完全一致，例如群的大小和各元素的阶（order）。所以我们就比较这些性质。

解答

以下一律以 e_G 和 e_H 表示 G 和 H 的单位元（identity），并以 g^{-1} 和 h^{-1} 表示在 G 和 H 中的逆元（inverse）。

$G \times H$ 是一个群

我们逐一检查四条公理。

封闭性（*closure*）。若 $g_1, g_2 \in G$ 且 $h_1, h_2 \in H$，则 $g_1 * g_2 \in G$ 且 $h_1 \circ h_2 \in H$，因为 G 和 H 都是封闭的。所以 $(g_1, h_1)(g_2, h_2) = (g_1 * g_2,\ h_1 \circ h_2) \in G \times H$。

结合律（*associativity*）。对任意三个元素，

$$\begin{aligned}\big((g_1, h_1)(g_2, h_2)\big)(g_3, h_3) &= \big((g_1 * g_2) * g_3,\ (h_1 \circ h_2) \circ h_3\big) \\ &= \big(g_1 * (g_2 * g_3),\ h_1 \circ (h_2 \circ h_3)\big) \\ &= (g_1, h_1)\big((g_2, h_2)(g_3, h_3)\big),\end{aligned}$$

其中中间一步用到 G 和 H 中的结合律。

单位元。序偶 (e_G, e_H) 满足

$$(e_G, e_H)(g,h) = (e_G * g,\ e_H \circ h) = (g,h),$$

同理 $(g,h)(e_G,e_H) = (g,h)$。所以它是单位元。

逆元。给定 (g,h)，序偶 (g^{-1},h^{-1}) 属于 $G \times H$，并且

$$(g,h)(g^{-1},h^{-1}) = (g * g^{-1},\ h \circ h^{-1}) = (e_G, e_H),$$

而次序倒过来相乘，所得的积同样是 (e_G,e_H)。所以每个元素都有逆元。

四条公理全部成立，所以 $G \times H$ 在这个乘法之下是一个群。

第 (i) 部分：真

若 G 和 H 是阿贝尔群（*abelian group*），则 $G \times H$ 也是。对任意两个元素，

$$(g_1,h_1)(g_2,h_2) = (g_1 * g_2,\ h_1 \circ h_2) = (g_2 * g_1,\ h_2 \circ h_1) = (g_2,h_2)(g_1,h_1),$$

这里用到 G 和 H 中的交换律（commutativity）。

若 $G \times H$ 是阿贝尔群，则 G 和 H 也是。任取 $g_1, g_2 \in G$。由于 $G \times H$ 是阿贝尔群，

$$(g_1 * g_2,\ e_H) = (g_1,e_H)(g_2,e_H) = (g_2,e_H)(g_1,e_H) = (g_2 * g_1,\ e_H).$$

两个序偶相等，其第一个坐标必须相等，所以 $g_1 * g_2 = g_2 * g_1$，即 G 是阿贝尔群。以同样方法使用序偶 (e_G,h_1) 和 (e_G,h_2)，可证明 H 是阿贝尔群。

> ? **想一想** 为甚么在第二个方向的证明中需要用到单位元 e_H?

我们需要某个第二坐标，才能由 g_1 和 g_2 造出 $G \times H$ 的元素。任何固定的 $h \in H$ 都可以，因为比较时只有第一个坐标起作用。群必定含有单位元，所以 e_H 是一个稳妥的选择，我们确知它存在。

第 (ii) 部分：真

设 $\widetilde{G} = \{(g,e_H) : g \in G\}$，并定义 $\phi\colon G \to \widetilde{G}$ 为 $\phi(g) = (g,e_H)$。

$\widetilde{G}$ 是 $G \times H$ 的子群（subgroup）。它包含单位元 (e_G,e_H)。它对乘积封闭，因为 $(g_1,e_H)(g_2,e_H) = (g_1 * g_2, e_H)$。它对取逆元封闭，因为 $(g,e_H)^{-1} = (g^{-1},e_H)$。

ϕ 是一个同构（isomorphism）。按定义，它映满 $\widetilde{G}$（即满射，onto）。它是一对一的（one-to-one），因为 $(g_1,e_H) = (g_2,e_H)$ 必然推出 $g_1 = g_2$。它保持运算，因为

$$\phi(g_1 * g_2) = (g_1 * g_2,\ e_H) = (g_1,e_H)(g_2,e_H) = \phi(g_1)\phi(g_2).$$

所以 $G \times H$ 包含子群 $\widetilde{G}$，而它与 G 同构。

第 (iii) 部分：假

把两个群都写成加法形式：$\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$，运算为模 n 加法。在 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 中，每个元素 $x = (a, b)$ 都满足

$$x + x = (2a, 2b) = (0, 0),$$

所以每个元素的阶都是 1 或 2。相反，在 $\mathbb{Z}_4$ 中，$1 + 1 = 2 \neq 0$，所以 1 的阶是 4。

假设存在同构 $\phi\colon \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2$。那么

$$\phi(2) = \phi(1 + 1) = \phi(1) + \phi(1) = (0, 0) = \phi(0),$$

因为由上面的计算可知 $\phi(1) + \phi(1) = (0, 0)$。最后一个等号成立，是因为同构把单位元映到单位元。于是 $\phi(2) = \phi(0)$ 而 $2 \neq 0$，所以 ϕ 不是一对一的，这就产生矛盾。因此 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 与 $\mathbb{Z}_4$ 不同构。（简而言之，$\mathbb{Z}_4$ 是循环群（cyclic group），而 $\mathbb{Z}_2 \times \mathbb{Z}_2$ 不是。）

第 (iv) 部分：假

同构是双射（bijection），所以同构的群有相同数目的元素。现在 $|S_n| = n!$，因为 n 件物件共有 $n!$ 种排列；而 $|G \times H| = |G|\,|H|$，因为每个序偶都是由自由选取的 g 和 h 组成。所以

$$|S_2 \times S_3| = 2! \times 3! = 12, \qquad |S_6| = 6! = 720.$$

两者大小不同，所以 $S_2 \times S_3$ 与 S_6 不同构。

验算

- 对一些细小的例子（例如 $\mathbb{Z}_2 \times \mathbb{Z}_3$、$\mathbb{Z}_2 \times S_3$、$S_3 \times S_3$ 和 $S_2 \times S_3$）逐一穷举检查，可见这个积满足全部四条公理，单位元为 (e_G, e_H)。当且仅当两个因子都是阿贝尔群时，它才是阿贝尔群，而 $g \mapsto (g, e_H)$ 是一个单射（injective）同态（homomorphism）。
- $\mathbb{Z}_2 \times \mathbb{Z}_2$ 中各元素的阶是 $1, 2, 2, 2$，而 $\mathbb{Z}_4$ 中则是 $1, 2, 4, 4$，所以这两个群确实不同。
- 列出 $S_2 \times S_3$ 的元素，共有 12 个，而 S_6 则有 720 个。

常见错误

- 在 (i) 中只证明了“当且仅当”的其中一个方向。逆命题（converse）才是需要技巧的部分，要把 e_H 放在第二个坐标。
- 在 (ii) 中说“G 是 $G \times H$ 的子群”。严格来说，G 连 $G \times H$ 的子集也不是。真正的子群是 $G \times \{e_H\}$，它与 G 同构。

- 在 (iii) 中指出两个群的阶都是 4 便就此打住。大小相等是同构的必要条件，但并非充分条件，所以我们需要找出一个能区分这两个群的性质。

STEP 之外

这就是*直积*（direct product）的构造，是由旧群造出新群的最简单方法。大学第一门群论课的一个重要结果是：每个有限阿贝尔群都是循环群的直积。与此相关的一个事实是：$\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ 当且仅当 $\gcd(m,n) = 1$，这正是中国剩余定理（Chinese remainder theorem）在群论中的面貌。第 (iii) 部分就是这个结论不成立的最小例子。想了解更多，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）。

详细解答

STEP III 1987，第 10 题

纯数学 · 伯努利多项式

> **STEP III 1987，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/87-s3-q10

关键思路

定义规则只通过导数（derivative）来确定 P_n，因此最自然的做法，是对差式 $P_n(x+1)-P_n(x)-nx^{n-1}$ 的导数运用归纳法（induction）：归纳假设令它变为零，于是只剩下一个常数，而积分条件就把这个常数确定下来。之后，所求的和是一个裂项和（telescoping sum）；至于立方和，我们只需要求出 P_4 除常数项以外的部分，反正常数项会互相抵消。

解答

差分恒等式

对 $n \geq 1$，设

$$Q_n(x) = P_n(x+1) - P_n(x) - nx^{n-1}.$$

我们用归纳法证明：对所有 x，$Q_n(x) = 0$。

基础情况 $n = 1$。由于 $P_1' = 1 \cdot P_0 = 1$，故对某个常数 c 有 $P_1(x) = x + c$。于是 $P_1(x+1) - P_1(x) = 1 = 1 \cdot x^0$，所以 $Q_1 \equiv 0$。（事实上，由 $\int_0^1 (x+c)\,\mathrm{d}x = 0$ 可得 $c = -\frac{1}{2}$，但我们用不着它。）

归纳步骤。假设 $n \geq 2$ 且 $Q_{n-1} \equiv 0$。求导，并利用 $P_n' = nP_{n-1}$，得

$$\begin{aligned} Q_n'(x) &= nP_{n-1}(x+1) - nP_{n-1}(x) - n(n-1)x^{n-2} \\ &= n\left[P_{n-1}(x+1) - P_{n-1}(x) - (n-1)x^{n-2}\right] \\ &= nQ_{n-1}(x) = 0. \end{aligned}$$

所以 Q_n 是常数，并且等于它在 $x = 0$ 的值。因为 $n \geq 2$，nx^{n-1} 这一项在 $x = 0$ 时为零，

所以

$$Q_n(0) = P_n(1) - P_n(0).$$

> **想一想** 我们怎样知道 $P_n(1) = P_n(0)$? 积分条件说的是 P_n 的积分，而不是它在个别点的取值。

把积分条件用在低一级的多项式上。根据微积分基本定理（fundamental theorem of calculus），

$$P_n(1) - P_n(0) = \int_0^1 P_n'(x)\,\mathrm{d}x = n\int_0^1 P_{n-1}(x)\,\mathrm{d}x = 0,$$

这是因为 $n-1 \geq 1$，所以条件 $\int_0^1 P_{n-1}\,\mathrm{d}x = 0$ 适用。（这正是 $n=1$ 的情况必须另外处理的原因：$\int_0^1 P_0\,\mathrm{d}x = 1$。）因此 $Q_n \equiv 0$，归纳法至此完成：

$$P_n(x+1) - P_n(x) = nx^{n-1} \qquad \text{对所有} n \geq 1.$$

幂和

在恒等式中代入 $x = m$，并对 $m = 0, 1, \ldots, k$ 求和。右边逐项相消：

$$\begin{aligned} n\sum_{m=0}^{k} m^{n-1} &= \sum_{m=0}^{k}\big[P_n(m+1) - P_n(m)\big] \\ &= \big[P_n(1) - P_n(0)\big] + \big[P_n(2) - P_n(1)\big] + \cdots + \big[P_n(k+1) - P_n(k)\big] \\ &= P_n(k+1) - P_n(0), \end{aligned}$$

这正是所求。（当 $n=1$ 时，$m=0$ 的一项是 0^0，必须理解为 1；这样两边都等于 $k+1$。）

立方和

取 $n=4$，所以我们需要 P_4。由 $P_0 = 1$ 开始逐级向上，每次积分，并选取常数使得在 $[0,1]$ 上的积分为零：

$$\begin{aligned} P_1(x) &= x - \tfrac{1}{2}, \\ P_2(x) &= x^2 - x + \tfrac{1}{6} && \text{因为 } \textstyle\int_0^1 (x^2 - x)\,\mathrm{d}x = -\tfrac{1}{6}, \\ P_3(x) &= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x && \text{因为 } \textstyle\int_0^1 (x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x)\,\mathrm{d}x = \tfrac{1}{4} - \tfrac{1}{2} + \tfrac{1}{4} = 0, \\ P_4(x) &= x^4 - 2x^3 + x^2 + c \end{aligned}$$

其中 c 为某个常数。我们不需要求 c：它在 $P_4(k+1) - P_4(0)$ 中会抵消。记 $y = k+1$，则

$$P_4(y) - P_4(0) = y^4 - 2y^3 + y^2 = y^2(y-1)^2 = (k+1)^2k^2.$$

所以在求和公式中取 $n=4$，得

$$\sum_{m=0}^{k} m^3 = \frac{k^2(k+1)^2}{4} = \left(\frac{k(k+1)}{2}\right)^2.$$

当 $k=1000$ 时，$\frac{1}{2}k(k+1)=\frac{1}{2}\cdot 1000\cdot 1001=500500$，所以

$$\sum_{m=0}^{1000} m^3=(500500)^2,$$

这就是所要求的结果。

验算

- 较小的 k：$k=2$ 给出 $0+1+8=9=\left(\frac{2\cdot 3}{2}\right)^2$，而 $k=3$ 给出 $36=6^2$。
- 同样的方法用于 $n=2$，得 $2\sum m=P_2(k+1)-P_2(0)=(k+1)^2-(k+1)=k(k+1)$，即大家熟悉的三角形数（triangular number）公式。留意立方和正好是它的平方，这是一个值得发现的巧妙之处。
- 用电脑以精确分数构造 $P_1,\dots,P_{12}$，可在随机选取的有理点上验证差分恒等式，并对 $k\le 39$ 验证求和公式。直接把立方相加，得 $250\,500\,250\,000=500500^2$。（我们略过的那个常数，原来是 $c=-\frac{1}{30}$。）

常见错误

- 试图把 Q_n 在 $[0,1]$ 上积分，借此确定 Q_n 这个常数。这样会引入 $\int_1^2 P_n$，而我们对它一无所知。在 $x=0$ 取值，再运用关于 P_{n-1} 的条件，要简洁得多。
- 忘记了 $P_n(1)=P_n(0)$ 这一步需要 $n\ge 2$，所以基础情况必须是 $n=1$，并要亲手验证。
- 花时间去求 P_4 的常数项，而它其实会抵消。

STEP 之外

伯努利多项式（Bernoulli polynomials）有生成函数（generating function）$\dfrac{te^{xt}}{e^t-1}=\sum_{n\ge 0}P_n(x)\dfrac{t^n}{n!}$，而它们在 0 的值就是伯努利数（Bernoulli numbers）$1,-\frac{1}{2},\frac{1}{6},0,-\frac{1}{30},\dots$。它们是欧拉—麦克劳林公式（Euler–Maclaurin formula）的核心；这条公式把一个和与一个积分及其误差项作比较。它们也出现在欧拉（Euler）对 $\sum 1/n^{2k}$ 的求值之中。想进一步了解这些想法，可参阅 Graham、Knuth 与 Patashnik 合著的 *Concrete Mathematics*（《具体数学》），书中对幂和与伯努利数有深入的论述。

详细解答

STEP III 1987，第 11 题

力学 · 相对速度的最优化

STEP III 1987，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：

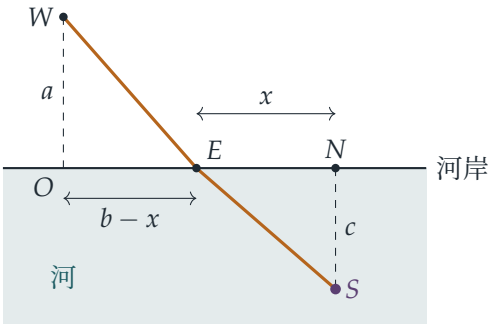
step.maths.org/questions/87-s3-q11

关键思路

把总时间写成单一变量 x 的函数，然后令其导数（derivative）等于零。在静水中，这其实就是光学中的折射（refraction）问题，答案是斯涅耳定律（Snell's law）。有水流时，唯一要多做的工夫是计算游泳时间：水流速度与她的游泳速度大小相同，所以两者之和是一个菱形（rhombus）的对角线，游泳时间因而成为 x 的一个出奇地简单的函数。

解答

把这名女子视为质点（particle），她在陆地上以恒定速率 u 运动，在水中则以相对于水的恒定速率 v 运动，两段路程都沿直线进行。取河岸为 X 轴，以下游为正方向，田野为 $Y > 0$，并设女子位于 $W = (0, a)$。于是树桩位于 $S = (b, -c)$，河岸上离树桩最近的点是 $N = (b, 0)$，而她在 $E = (b - x, 0)$ 下水，此点在 N 的上游，距离为 x。



她跑步的距离是 $WE = \sqrt{a^2 + (b-x)^2}$，游泳的距离是 $ES = \sqrt{x^2 + c^2}$。

静水

总时间为

$$T(x) = \frac{\sqrt{a^2 + (b-x)^2}}{u} + \frac{\sqrt{x^2 + c^2}}{v},$$

所以

$$T'(x) = -\frac{b-x}{u\sqrt{a^2 + (b-x)^2}} + \frac{x}{v\sqrt{x^2 + c^2}}.$$

函数 $s \mapsto s/\sqrt{k^2 + s^2}$ 是递增的，所以 T' 的两项都随 x 增加而增加。因此 T' 严格递增，T 恰好有一个驻点（stationary point），而这个驻点就是它的最小值点。由于 $T'(0) < 0 < T'(b)$，这一点位于 $0 < x < b$ 之内。所求的方程是

$$\boxed{\frac{b-x}{u\sqrt{a^2 + (b-x)^2}} = \frac{x}{v\sqrt{x^2 + c^2}}}$$

或等价地 $v(b-x)\sqrt{x^2 + c^2} = ux\sqrt{a^2 + (b-x)^2}$。若以 WE 和 ES 与河岸法线（normal）所成的角 α 和 β 表示，这就是 $\dfrac{\sin\alpha}{u} = \dfrac{\sin\beta}{v}$，亦即折射的斯涅耳定律。

流动的河水

现在河水以速率 v 向下游流动，而她相对于水的游泳速率也是 v。要沿直线 ES 前进，她必须选择一个游向（heading），与下游方向成向下的角 θ，使她相对于地面的速度沿 ES 的方向。

想一想 她沿 ES 实际移动得有多快？为甚么她必须在树桩的上游下水？

她相对于地面的速度是水流速度加上她的游泳速度：

$$v\begin{pmatrix}1\\0\end{pmatrix} + v\begin{pmatrix}\cos\theta\\-\sin\theta\end{pmatrix} = 2v\cos\tfrac{\theta}{2}\begin{pmatrix}\cos\frac{\theta}{2}\\-\sin\frac{\theta}{2}\end{pmatrix},$$

这里用了 $1 + \cos\theta = 2\cos^2\frac{\theta}{2}$ 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$。从几何上看，两个速度的长度相等，所以它们构成一个菱形，而两者之和平分它们之间的夹角。因此她相对于地面的路径与下游方向成向下的角 $\phi = \theta/2$。因为 $0 \le \theta \le \pi$，我们有 $0 \le \phi \le \frac{\pi}{2}$，所以她只能向下游漂移：树桩必须在 E 的下游，即 $x > 0$。要使路径沿着 ES，我们需要

$$\tan\phi = \frac{c}{x}, \qquad \text{所以} \qquad \cos\phi = \frac{x}{\sqrt{x^2 + c^2}},$$

而她沿 ES 的速率是 $2v\cos\phi = \dfrac{2vx}{\sqrt{x^2 + c^2}}$。游泳时间等于距离除以速率：

$$\frac{\sqrt{x^2 + c^2}}{2vx/\sqrt{x^2 + c^2}} = \frac{x^2 + c^2}{2vx}.$$

所以总时间为

$$T(x) = \frac{\sqrt{a^2+(b-x)^2}}{u} + \frac{x^2+c^2}{2vx} = \frac{\sqrt{a^2+(b-x)^2}}{u} + \frac{1}{2v}\left(x + \frac{c^2}{x}\right), \qquad x > 0,$$

而

$$T'(x) = -\frac{b-x}{u\sqrt{a^2+(b-x)^2}} + \frac{1}{2v}\left(1 - \frac{c^2}{x^2}\right).$$

当 $x > 0$ 时，两项同样随 x 严格递增。此外，当 $x \to 0^+$ 时 $T'(x) \to -\infty$，而当 x 很大时 $T'(x) > 0$，所以恰好有一个驻点，并且它给出最短时间。令 $T'(x) = 0$，再乘以 $2vx^2u\sqrt{a^2+(b-x)^2}$（此式为正），得

$$2vx^2(b-x) = u(x^2-c^2)\left[a^2+(b-x)^2\right]^{\frac{1}{2}},$$

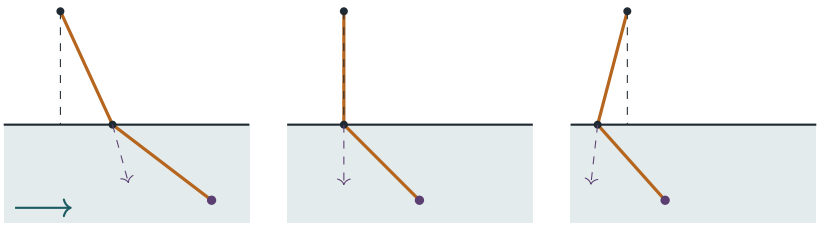
这正是所要求的结果。

三幅草图

不必解这个方程，也能从中看出 x 的位置。因子 $[a^2+(b-x)^2]^{1/2}$ 是正数，所以 $b-x$ 和 $x-c$ 必须同号（或同时为零）。

- $b > c$：这时 $c < x < b$。假如 $x \le c$，便会有 $b-x > 0 \ge x-c$；假如 $x \ge b$，便会有 $b-x \le 0 < x-c$；两者都不可能。她在 $b-x > 0$ 处下水，位于 O 的下游，而该点是河岸上离她最近的点。由于 $x > c$，她的游泳路径与河岸所成的角 $\phi < 45°$，而她的游向 $\theta = 2\phi < 90°$ 偏向下游。
- $b = c$：这时按同样的论证可得 $x = c = b$。她直接跑到 O，然后正对着对岸横渡（$\theta = 90°$），水流则带着她沿一条与河岸成 $45°$ 的路径前进。
- $b < c$：这时 $b < x < c$，所以 $b-x < 0$。她要向上游跑去下水，沿一条比 $45°$ 更陡的路径游泳，而她的游向 $\theta > 90°$ 部分指向上游。

以下草图按比例绘画，取 $a = 3$、$c = 2$、$u = 2$、$v = 1$，以及 $b = 4, 2, 1$。数值上，$x \approx 2.62$、$x = 2$ 和 $x \approx 1.79$。水流向右；E 处的虚线箭头表示她相对于水的游向。



$b > c$　　$b = c$　　$b < c$

验算

- 若 $b = c$，则 $x = c$ 使题目所给方程的两边都等于零，与上面所得一致。
- 正对着对岸横渡（$\theta = 90°$）：她每横渡 v，便向下游漂移 v，所以路径成 45°。公式给出速率 $2v\cos 45° = \sqrt{2}\,v$，路径长度为 $\sqrt{2}\,c$，所以时间是 c/v，这显然是应有的结果。
- 根据算术平均—几何平均不等式（AM–GM），游泳时间 $(x^2 + c^2)/(2vx)$ 在 $x = c$ 时最小。这是合理的：假如跑步不花任何时间，她会在 N 上游距离 c 处下水，然后正对着对岸横渡。
- 数值上，对 a, b, c, u, v 的 2000 组随机取值直接求 T 的最小值，所得的最小值点每次都满足两个方程，而且在第二种情况下都位于 b 与 c 之间。逐步模拟她在水流中游泳的过程，亦验证了游泳时间。例如，$a = 3$、$b = 4$、$c = 2$、$u = 2$、$v = 1$ 给出 $x \approx 2.621$。

常见错误

- 把水流速率直接加到她的速率上，仿佛她是对准树桩直游过去。她必须偏着游，使合速度（resultant velocity）指向树桩。
- 忘记了有水流时只有 $x > 0$ 才可能：由于两个速率相等，她根本无法逆流而上。
- 过早把关于 x 的方程两边平方。这样会失去正负号的信息，而正是这些信息把 x 定在 b 与 c 之间。

STEP 之外

第一个答案是斯涅耳定律，它源自费马原理（Fermat's principle）：光沿所需时间最短的路径行进。选择整条路径以使时间最短，正是变分法（calculus of variations）所研究的课题。与此密切相关的是策梅洛（Zermelo）导航问题：水流因地而异，如何驾驶一艘船穿越其中，才能最快到达目的地。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对相对速度和运动参考系有细致的讨论。

详细解答

STEP III 1987，第 12 题

力学 · 变质量转动

STEP III 1987，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/87-s3-q12

关键思路

这是一道变质量（variable mass）问题，所以不能盲目套用 $F = ma$（或它在转动中的对应公式）。稳妥的做法是回到基本原理：比较时刻 t 与时刻 $t + \delta t$ 绕枢轴（pivot）的角动量（angular momentum），其间有少量已燃烧的燃料离开了系统。枢轴是光滑的，而重力沿铅垂方向，所以没有任何力对通过枢轴的铅垂轴产生力矩（moment），因此该角动量守恒。

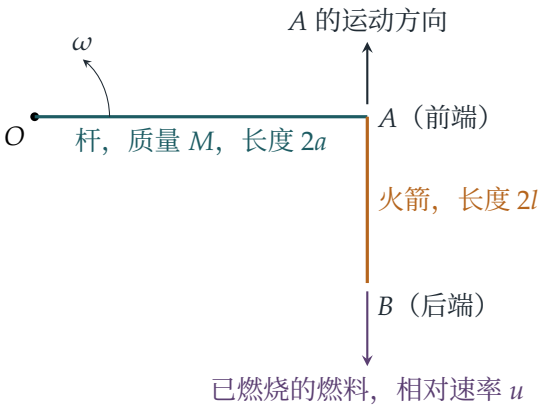
解答

建立模型

把杆和火箭都看成均匀刚体杆，在水平面上绕通过枢轴 O 的铅垂轴转动。设杆为 OA，其中 $OA = 2a$。取单位向量 $\mathbf{e}_r$ 沿 OA 方向，$\mathbf{e}_\theta$ 与它垂直并指向转动的方向，又设 $\mathbf{k}$ 为铅垂向上的单位向量，于是 $\mathbf{e}_r \times \mathbf{e}_\theta = \mathbf{k}$。

火箭的前端在 A，并指向前方，即 A 的运动方向。因此火箭沿 $\mathbf{e}_\theta$ 方向放置，由 A 向后延伸至其后端 B，其中

$$\overrightarrow{OB} = 2a\,\mathbf{e}_r - 2l\,\mathbf{e}_\theta.$$



转动惯量

杆绕其一端的转动惯量（moment of inertia）为 $\frac{1}{3}M(2a)^2 = \frac{4}{3}Ma^2$。

火箭的中心位于 $2a\,\mathbf{e}_r - l\,\mathbf{e}_\theta$，与 O 的距离为 $\sqrt{4a^2+l^2}$。它绕自身中心的转动惯量为 $\frac{1}{3}ml^2$，所以由平行轴定理（parallel axis theorem），它的贡献是 $\frac{1}{3}ml^2 + m(4a^2+l^2)$。总和为

$$I = \tfrac{4}{3}Ma^2 + m\left(4a^2 + \tfrac{4}{3}l^2\right).$$

I 如何变化

由于 $m = 2\alpha l$ 及 $\dfrac{\mathrm{d}m}{\mathrm{d}t} = -\alpha\beta$，我们有 $\dfrac{\mathrm{d}l}{\mathrm{d}t} = -\frac{1}{2}\beta$。用积法则（product rule）对 I 求导，

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\mathrm{d}m}{\mathrm{d}t}\left(4a^2 + \tfrac{4}{3}l^2\right) + m\cdot\tfrac{8}{3}l\,\frac{\mathrm{d}l}{\mathrm{d}t}\\ &= -\alpha\beta\left(4a^2 + \tfrac{4}{3}l^2\right) + 2\alpha l\cdot\tfrac{8}{3}l\cdot\left(-\tfrac{1}{2}\beta\right) = -\alpha\beta\,(4a^2+4l^2).\end{aligned}$$

留意 $4a^2 + 4l^2 = OB^2$：转动惯量下降的速率，恰好等于从后端 B 移走质量时所得的速率。这是合理的，因为一枚均匀的火箭若要保持均匀，而前端又固定在 A，就只能从后端缩短。

喷出燃料的角动量

在由 t 到 $t+\delta t$ 的短时间内，有质量为 $\delta\mu = \alpha\beta\,\delta t$ 的已燃烧燃料从 B 离开。燃烧之前，它随火箭一起运动，速度为 $\mathbf{v}_B = \omega\mathbf{k}\times\overrightarrow{OB}$。燃烧之后，它的速度是 $\mathbf{v}_B - u\,\mathbf{e}_\theta$，因为它相对于火箭是向正后方运动的。这时它绕 O 的角动量为

$$\begin{aligned}\delta\mu\,\overrightarrow{OB}\times(\mathbf{v}_B - u\,\mathbf{e}_\theta) &= \delta\mu\left[\omega\,OB^2\,\mathbf{k} - (2a\,\mathbf{e}_r - 2l\,\mathbf{e}_\theta)\times u\,\mathbf{e}_\theta\right]\\ &= \delta\mu\left[\omega(4a^2+4l^2) - 2au\right]\mathbf{k},\end{aligned}$$

其中用到 $\mathbf{e}_\theta \times \mathbf{e}_\theta = \mathbf{0}$，以及当 $\mathbf{r}$ 垂直于 $\mathbf{k}$ 时 $\mathbf{r} \times (\omega\mathbf{k} \times \mathbf{r}) = \omega r^2 \mathbf{k}$。

> **想一想** 为甚么在排气的那一项中，$\overrightarrow{OB}$ 只有 $2a\,\mathbf{e}_r$ 这部分起作用?

排气的相对速度沿 $\mathbf{e}_\theta$ 方向，与火箭在同一直线上。所以 $\overrightarrow{OB}$ 沿火箭方向的分量不产生力矩。无论燃料在火箭上哪一点离开，由 O 到排气（以及推力）作用线的垂直距离都只是 $OA = 2a$。

角动量守恒

枢轴是光滑的，所以它的反作用力对通过 O 的铅垂轴没有力矩；重力沿铅垂方向，所以也没有力矩。因此整个系统（烟花连同在 δt 内燃烧的燃料）绕该轴的角动量在 t 与 $t+\delta t$ 两个时刻相同：

$$I(t+\delta t)\,\omega(t+\delta t) + \delta\mu\left[\omega(4a^2+4l^2) - 2au\right] = I(t)\,\omega(t),$$

误差只有 $(\delta t)^2$ 阶的项。两边除以 δt，并令 $\delta t \to 0$，其中 $\delta\mu/\delta t = \alpha\beta$，得

$$\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) + \alpha\beta\,\omega(4a^2+4l^2) - 2\alpha\beta au = 0.$$

展开 $\dfrac{\mathrm{d}}{\mathrm{d}t}(I\omega) = I\dfrac{\mathrm{d}\omega}{\mathrm{d}t} + \omega\dfrac{\mathrm{d}I}{\mathrm{d}t}$，并代入 $\dfrac{\mathrm{d}I}{\mathrm{d}t} = -\alpha\beta(4a^2+4l^2)$，两个含 ω 的项恰好互相抵消，余下

$$I\frac{\mathrm{d}\omega}{\mathrm{d}t} = 2\alpha\beta au.$$

这正是“转动惯量 × 角加速度 = 推力的力矩”：推力（thrust）为 $\alpha\beta u$，沿火箭方向作用，与 O 的垂直距离为 $2a$。最后，

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{2\alpha\beta au}{\frac{4}{3}Ma^2 + 2\alpha l\left(4a^2 + \frac{4}{3}l^2\right)} = \frac{6\alpha\beta au}{4Ma^2 + 8\alpha l(3a^2+l^2)}$$
$$= \frac{3\alpha\beta au}{2\left[Ma^2 + 2\alpha l(3a^2+l^2)\right]},$$

正如所求。只要还有燃料剩下，即 $l > 0$，这个结果就成立。

验算

- 没有火箭时（$l \to 0$），结果变成 $\frac{4}{3}Ma^2\,\dot{\omega} = 2a \cdot \alpha\beta u$：推力乘以力臂，与预期相符。
- 量纲：$\alpha\beta u$ 是一个力（α 是每单位长度的质量，β 是一个速率），所以分子是力 × 长度，而分母是转动惯量。
- 数值上，取 $M = 1.3$、$a = 0.7$、$\alpha = 0.9$、$\beta = 0.4$、$u = 2.5$、$l(0) = 1.1$ 及 $\omega(0) = 0.3$，我们把运动分成多步来模拟：从后端逐一喷出 2×10^5 小块，每次都令角动量守恒。在 $t = 4.95$ 时，这样得出 $\omega \approx 2.39817$，而对公式积分则得出 2.39816。

常见错误

- 只对烟花本身写出 $\frac{\mathrm{d}}{\mathrm{d}t}(I\omega) =$ 力矩。烟花并不是封闭系统，除非你同时计及燃料带走的角动量，否则这样会得出错误的答案。
- 把排气的绝对速度当作向后的 u，而不是相对于运动中的火箭的 u。
- 忘记 l 会改变，以致忽略了火箭的转动惯量是通过 m 和 l 两者而随时间变化的。

STEP 之外

这里的抵消是一个普遍的事实。若物体失去质量时，离开的质量带有当地的刚体速度再加上一个相对排气速度，则运动方程就是瞬时物体的普通运动方程，另加一个由相对速度产生的“推力”。这就是火箭方程（rocket equation）的转动版本。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）仔细讨论了变质量系统和角动量，程度正好与本题相若。

详细解答

STEP III 1987，第 13 题

力学 · 涉及摩擦力的转动杆

STEP III 1987，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/87-s3-q13

关键思路

只要质点仍紧贴在杆上不滑动，杆与质点便作为一个刚体（rigid body）一同绕铰链转动，因此以整体的转动惯量（moment of inertia）应用能量守恒（conservation of energy），即可立刻求得 $\dot{\theta}^2$。要找出何时开始滑动，我们再单独考虑质点：它绕铰链作圆周运动，所以在径向和横向应用牛顿第二定律（Newton's second law），便能准确知道杆须提供多大的摩擦力（friction）和法向反作用力（normal reaction）。当所需的摩擦力达到法向反作用力的 $\frac{1}{2}$ 时，滑动便开始。

解答

运动方程

把杆视为绕光滑铰链 O 转动的均匀刚性杆，并把质点视为与 O 相距 a 的质点。在质点滑动之前，它一直与 O 保持距离 a，并随杆一同转动。

系统对 O 的转动惯量为

$$I = \tfrac{4}{3}(3m)a^2 + ma^2 = 4ma^2 + ma^2 = 5ma^2,$$

这里用了长度为 $2l$ 的均匀杆对其一端的转动惯量 $\frac{4}{3}Ml^2$。

铰链是光滑的，而质点与杆之间的摩擦力和法向反作用力都是系统的内力，所以它们所作的净功为零。当杆转过角度 θ 时，杆的中心和质点（两者与 O 的距离均为 a）各下降了 $a\sin\theta$。由于系统从静止开始运动，能量守恒给出

$$\tfrac{1}{2}(5ma^2)\dot{\theta}^2 = 3mg\,a\sin\theta + mg\,a\sin\theta = 4mga\sin\theta,$$

即

$$5a\dot{\theta}^2 = 8g\sin\theta,$$

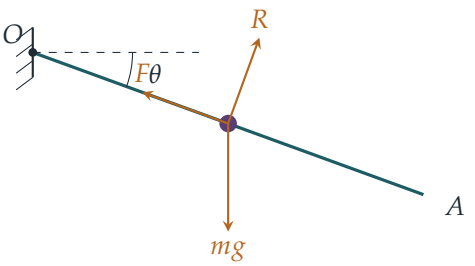
这正是所要证明的结果。对 t 求导得 $10a\dot{\theta}\ddot{\theta} = 8g\cos\theta\,\dot{\theta}$，所以（当 $\dot{\theta} \neq 0$ 时，而运动一旦开始这便成立）

$$a\ddot{\theta} = \tfrac{4}{5}g\cos\theta.$$

（对 O 取力矩也可得到同样的结果：$5ma^2\ddot{\theta} = 4mga\cos\theta$。）

质点何时滑动

现在单独考虑质点。作用在它身上的力有：重量 mg、垂直于杆的法向反作用力 R，以及沿杆方向的摩擦力 F。由于质点在以 O 为圆心、半径为 a 的圆上运动，它的加速度有一个沿杆指向 O 的分量 $a\dot{\theta}^2$，以及一个垂直于杆、沿 θ 增加方向的分量 $a\ddot{\theta}$。



我们取 F 沿杆指向 O，稍后会看到这正是质点所需的方向。沿杆方向分解，以指向 O 为正：重量有一个背离 O 的分量 $mg\sin\theta$，所以

$$F - mg\sin\theta = ma\dot{\theta}^2 = \tfrac{8}{5}mg\sin\theta \quad\Longrightarrow\quad F = \tfrac{13}{5}mg\sin\theta.$$

垂直于杆方向分解，以 θ 增加的方向（即由杆向下）为正：重量的分量为 $mg\cos\theta$，所以

$$mg\cos\theta - R = ma\ddot{\theta} = \tfrac{4}{5}mg\cos\theta \quad\Longrightarrow\quad R = \tfrac{1}{5}mg\cos\theta.$$

想一想 重力看来是把质点沿斜面向下拉，为甚么质点反而需要指向铰链的摩擦力?

起作用的不只是重力：质点要继续绕 O 作圆周运动，就需要一个指向 O 的净力 $ma\dot{\theta}^2$。重力沿杆的分量却指向相反方向，即背离 O，所以摩擦力既要提供向心力（centripetal force），又要抵销重力的分量。这就是为甚么 F 算出来是指向 O 为正，并且随 $\sin\theta$ 增长；而 R 很小，是因为杆正在质点下方“坠离”（在 $\theta = 0$ 时，质点有向下的加速度 $\frac{4}{5}g$，所以杆只承托它重量的 $\frac{1}{5}$）。

当 $0 < \theta < \frac{\pi}{2}$ 时，F 和 R 均为正，所以质点一直与杆保持接触。只要 $F \le \frac{1}{2}R$，质点便相对于杆保持静止，即

$$\tfrac{13}{5}mg\sin\theta \le \tfrac{1}{2}\cdot\tfrac{1}{5}mg\cos\theta \iff \tan\theta \le \tfrac{1}{26}.$$

在 $\theta = 0$ 时，我们有 $F = 0 < \frac{1}{2}R$，而 $F/R = 13\tan\theta$ 随 θ 递增，所以质点最初开始滑动（向外，背离铰链）是在

$$\tan\theta = \tfrac{1}{26},$$

之时，这正是所要证明的结果。

验算

- 在 $\theta = 0$ 时，力矩方程给出角加速度（angular acceleration）$\frac{4g}{5a}$；所以质点的加速度为 $\frac{4}{5}g$，小于 g，这与 $R = \frac{1}{5}mg > 0$ 一致。
- 以数值方法逐步计算刚体运动（取 $g = 9.81$，$a = 0.7$），所得结果与 $5a\dot\theta^2 = 8g\sin\theta$ 吻合，误差在 10^{-12} 以内；而直接由质点的计算加速度求得的摩擦力和反作用力，亦与 $\frac{13}{5}mg\sin\theta$ 和 $\frac{1}{5}mg\cos\theta$ 相符。比值 F/R 最初达到 $\frac{1}{2}$ 时，$\theta \approx 0.03845$ 弧度，即 $\arctan\frac{1}{26} \approx 2.20^\circ$。

常见错误

- 忘记向心项，把质点当作在斜面上处于平衡来分解力；这样不但得出错误的角度，连摩擦力的方向也会弄错。
- 正确地使用了 $\frac{1}{3}(3m)(2a)^2$，却忘记把质点的 ma^2 加进转动惯量。
- 像在固定斜面上那样取 $R = mg\cos\theta$：杆正在加速，所以反作用力要小得多。

STEP 之外

这道题让我们初尝在转动参考系（rotating frame）中处理问题的滋味：从杆上观察，质点受到一个向外的“离心力”（centrifugal force）$ma\dot\theta^2$ 和一个“欧拉力”（Euler force）$-ma\ddot\theta$，而这些虚拟力（fictitious forces）正是我们求得的那些项。大学本科的力学课程会有系统地发展这套理论，并会讨论质点开始滑动后出现的科里奥利力（Coriolis force）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对刚体转动和非惯性参考系（non-inertial frames）有十分清晰的讲解。

详细解答

STEP III 1987，第 14 题

力学 • 引力坍缩

STEP III 1987，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/87-s3-q14

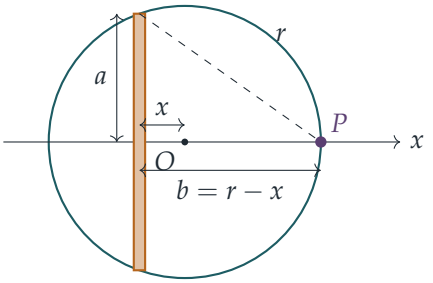
关键思路

实心球体可视为一叠薄圆盘，因此整个球体所施的力，就是把题目给出的圆盘公式沿着穿过质点（particle）的轴线积分。那个平方根看似麻烦，但只要留意到，当质点位于球面上时，根号内的式子可以因式分解，问题便迎刃而解。之后，关于云团的部分就是牛顿第二定律（Newton's second law）配合链式法则（chain rule）$\frac{\mathrm{d}v}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}r}$，而代换（substitution）$r = R\cos^2\alpha$ 的用意，正是要整理第二次积分时出现的平方根。

解答

球面上的力

把球心置于原点，质点 P 置于 x 轴上 $x = r$ 之处。把球体切成与轴线垂直的圆盘。位于 x 处的薄片（其中 $-r \le x \le r$）厚度为 δx，半径为 a，其中 $a^2 = r^2 - x^2$，而 P 位于该薄片的轴线上，与它相距 $b = r - x \ge 0$。



关键的化简在于平方根：

$$a^2+b^2=(r^2-x^2)+(r-x)^2=2r^2-2rx=2r(r-x),$$

所以

$$\frac{b}{(a^2+b^2)^{1/2}}=\frac{r-x}{\sqrt{2r(r-x)}}=\sqrt{\frac{r-x}{2r}}.$$

每个圆盘都沿轴线把 P 拉向球心，因此各个力的方向相同，直接相加即可。令 $\delta x\to 0$，合力为

$$\begin{aligned}F&=2\pi mk\rho\int_{-r}^{r}\left(1-\sqrt{\frac{r-x}{2r}}\right)\mathrm{d}x\\&=2\pi mk\rho\left(2r-\frac{1}{\sqrt{2r}}\int_0^{2r}u^{1/2}\,\mathrm{d}u\right)\qquad(u=r-x)\\&=2\pi mk\rho\left(2r-\frac{1}{\sqrt{2r}}\cdot\tfrac{2}{3}(2r)^{3/2}\right)=2\pi mk\rho\left(2r-\tfrac{4}{3}r\right)=\tfrac{4}{3}\pi mk\rho\, r.\end{aligned}$$

球体的质量为 $M=\frac{4}{3}\pi r^3\rho$，所以 $\frac{4}{3}\pi\rho=M/r^3$，于是

$$F=\frac{kmM}{r^2},$$

方向指向球心，正如题目所求。

云团的运动方程

把云团中的质点看成一个质量为 M 的连续均匀球体，并考虑球面上一个质量为 m、与球心相距 r 的质点。根据上面的结果，它所受的唯一的力是指向球心的 kmM/r^2。取 r 为坐标（以向外为正），由牛顿第二定律得

$$m\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=-\frac{kmM}{r^2}.$$

现在 $\frac{\mathrm{d}r}{\mathrm{d}t}=-v$，所以 $\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=-\frac{\mathrm{d}v}{\mathrm{d}t}$，而根据链式法则，

$$\frac{\mathrm{d}v}{\mathrm{d}t}=\frac{\mathrm{d}v}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}t}=-v\frac{\mathrm{d}v}{\mathrm{d}r}.$$

因此 $\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=v\frac{\mathrm{d}v}{\mathrm{d}r}$，约去 m 后得

$$v\frac{\mathrm{d}v}{\mathrm{d}r}=-\frac{kM}{r^2}.$$

想一想 圆盘的结果是针对固定的球体而得出的。为甚么可以把它用于正在坍缩的云团?

任何一刻的力，只取决于该刻质量的分布，所以我们需要云团始终保持为均匀球体，而且球面以内的质量始终是 M。事实正是如此：一个从半径 λR 处由静止开始运动的质点（其中 $0<\lambda\le 1$），其内侧的质量为 $\lambda^3 M$，而只要 $r(t)$ 满足 $\ddot r=-kM/r^2$，$\lambda r(t)$ 便满足 $\ddot s=-k\lambda^3 M/s^2$。所以每个质点与球心的距离，始终是球面半径的同一个比例：云团按均匀的比例因子收缩，保持均匀，而且没有质点会越过球面上的质点。（质点外侧的质量只会对内部的质点有影响，而球面之外并无任何质点。）

由静止开始的坍缩

分离变量（separate the variables），并由 $r=R$（该处 $v=0$）起积分：

$$\tfrac12 v^2=\int_R^r -\frac{kM}{s^2}\,\mathrm{d}s=kM\left(\frac1r-\frac1R\right)=\frac{kM(R-r)}{rR}.$$

加速度始终向内，而云团由静止开始，所以 r 一直递减，全程 $v\ge 0$。

现在代入 $r=R\cos^2\alpha$，当 r 由 R 向 0 递减时，α 由 $t=0$ 时的 0 向 $\frac\pi2$ 递增。这时 $R-r=R\sin^2\alpha$，所以

$$v^2=\frac{2kM\,R\sin^2\alpha}{R^2\cos^2\alpha}=\frac{2kM}{R}\tan^2\alpha,\qquad v=\left(\frac{2kM}{R}\right)^{1/2}\tan\alpha,$$

这里取正根，因为 $v\ge 0$，而且当 $0\le\alpha<\frac\pi2$ 时 $\tan\alpha\ge 0$。

> ? **想一想** 怎样由 v 作为 α 的函数，得出 t 作为 α 的函数?

利用 $v=-\frac{\mathrm{d}r}{\mathrm{d}t}$ 和 $\frac{\mathrm{d}r}{\mathrm{d}\alpha}=-2R\cos\alpha\sin\alpha$：

$$\frac{\mathrm{d}t}{\mathrm{d}\alpha}=-\frac1v\frac{\mathrm{d}r}{\mathrm{d}\alpha}=\frac{2R\sin\alpha\cos\alpha}{(2kM/R)^{1/2}\tan\alpha}=\left(\frac{R^3}{2kM}\right)^{1/2}2\cos^2\alpha=\left(\frac{R^3}{2kM}\right)^{1/2}(1+\cos 2\alpha).$$

积分，并利用 $\alpha=0$ 时 $t=0$，得

$$t=\left(\frac{R^3}{2kM}\right)^{1/2}\left(\alpha+\tfrac12\sin 2\alpha\right),$$

这就是 $r=R\cos^2\alpha$ 的时刻，正如题目所求。

验算

- 时间很短时：当 α 很小，$t\approx 2\alpha(R^3/2kM)^{1/2}$ 而 $R-r\approx R\alpha^2$，所以 $R-r\approx\frac12(kM/R^2)\,t^2$：即由静止开始、加速度为常数 kM/R^2 的运动，与起初阶段的预期完全吻合。
- 令 $\alpha\to\frac\pi2$，得出整个坍缩所需的时间为 $\frac\pi2(R^3/2kM)^{1/2}$；即使 $v\to\infty$，这个时间仍是有限的。

- 数值上，把 200 000 个薄圆盘的力相加，所得结果与 kmM/r^2 的相对误差约为 2×10^{-9}。取 $k=1.7$、$M=2.3$、$R=1.9$，对 $\ddot{r}=-kM/r^2$ 逐步作数值计算，直至 $\alpha=0.8$ 所对应的时刻，得 $r=0.922260$，与 $R\cos^2 0.8=0.922260$ 吻合。

常见错误

- 由球心而非由质点量度 b：应为 $b=r-x$，而不是 x。
- 忘记 v 是向内的速率，这会令 $\frac{\mathrm{d}^2r}{\mathrm{d}t^2}=v\frac{\mathrm{d}v}{\mathrm{d}r}$ 中的正负号相反。
- 为 v 取了负平方根，结果算出的时间是负数。

STEP 之外

第一部分是牛顿（Newton）球壳定理（shell theorem）的特殊情况：在球对称物体之外，引力的作用就如全部质量都集中在球心一样。大学课程会用引力的高斯定律（Gauss's law）$\nabla\cdot\mathbf{g}=-4\pi G\rho$ 简洁地证明它。你求得的坍缩时间，就是天体物理学家所说的自由落体时间（free-fall time），用来估算气体云坍缩形成恒星有多快；而以 α 作参数的表示法，也正是描述径向开普勒（Kepler）轨道的那一个。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）正是以这种方式处理有心力（central force）和球壳定理。

详细解答

STEP III 1987，第 15 题

概率与统计 · 期望成本的最优化

> **STEP III 1987，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/87-s3-q15

关键思路

与其直接求 E_n 的最小值，不如看看多做一次测试会带来甚么改变。比较“做完 n 次便停”与“做完 $n+1$ 次便停”这两个策略时，只有至今已得到 n 次阳性结果的病人才有分别。对这些病人来说，多做一次测试要花 £10，而若病人患的是 Steppe 综合症而测试结果呈阴性，便可省下 £100 的治疗费。因此一切取决于 $E_{n+1} - E_n$ 的正负号，而这个正负号最多只会改变一次。

解答

设 F 为病人患上 Fenland 热病的事件（event），故 $\mathrm{P}(F) = p$，并设 $S = F'$ 为患上 Steppe 综合症。我们假设 Steppe 病人的各次测试结果互相独立（independent），每次呈阳性的概率为 q，其中 $0 < q < 1$。不论谁接受治疗，治疗费都是 £100。

第 (i) 部分

求 E_n。若病人患的是 F，每次测试都呈阳性。因此 n 次测试全部都要做，然后再接受治疗，费用为 $10n + 100$。

若病人患的是 S，第 k 次测试（其中 $1 \le k \le n$）只会在首 $k-1$ 次测试全部呈阳性时才进行，其概率为 q^{k-1}。因此测试次数的期望值（expected value）为

$$\sum_{k=1}^{n} q^{k-1} = \frac{1-q^n}{1-q},$$

而病人只会在 n 次测试全部呈阳性时才接受治疗，其概率为 q^n。根据期望值的线性性质

(linearity of expectation),

$$E_n = p\,(10n + 100) + (1 - p)\left(\frac{10(1 - q^n)}{1 - q} + 100q^n\right).$$

由此亦得 $E_0 = 100$，这正合预期，因为此时所有人都接受治疗。

差值。 F 的部分增加了 10。在 S 的部分，$\dfrac{1 - q^{n+1}}{1 - q} - \dfrac{1 - q^n}{1 - q} = q^n$，所以

$$\begin{aligned} E_{n+1} - E_n &= 10p + (1 - p)\left(10q^n + 100q^{n+1} - 100q^n\right) \\ &= 10p + 10(1 - p)q^n(1 + 10q - 10) \\ &= 10p - 10(1 - p)q^n(9 - 10q), \end{aligned}$$

这正是所要证明的结果。

解读这个结果：有 p 的概率，第 $n + 1$ 次测试白白用在热病病人身上。有 $(1 - p)q^n$ 的概率，这次测试用在 Steppe 病人身上，要花 £10，但它有 $1 - q$ 的概率可省下 £100。期望净节省为 $100(1 - q) - 10 = 10(9 - 10q)$。

代入数值。 记 $D_n = E_{n+1} - E_n$。当 $p = 10^{-4}$ 及 $q = 10^{-2}$ 时，$9 - 10q = 8.9$ 而 $1 - p = 0.9999$，所以

$$D_n = 10^{-3} - 89 \times 0.9999 \times 10^{-2n}.$$

> **想一想** 为甚么只需找出 D_n 在何处变号便已足够?

这里 $9 - 10q > 0$，而 q^n 随 n 增大而严格递减（strictly decreasing），所以 D_n 随 n 严格递增。因此当 $D_n < 0$ 时数列（sequence）E_n 下降，一旦 $D_n > 0$ 便上升，而最小值出现在第一个使 $D_n > 0$ 的 n。现在

$$\begin{aligned} D_2 &= 10^{-3} - 0.0088991 < 0, \\ D_3 &= 10^{-3} - 0.000088991 = 0.000911\ldots > 0. \end{aligned}$$

所以 $E_0 > E_1 > E_2 > E_3$ 且 $E_3 < E_4 < E_5 < \cdots$，医生应选 $n = 3$。（此时期望费用约为 £10.11，而立即治疗则要 £100。）

第 (ii) 部分

取 $q_0 = \frac{9}{10}$，并假设 $q > \frac{9}{10}$。那么 $9 - 10q < 0$，所以

$$D_n = 10p + 10(1 - p)q^n(10q - 9).$$

对任意 $p \in [0, 1]$，这两项都是非负的。它们不可能同时为零：若 $p > 0$，第一项为正；若 $p = 0$，第二项为 $10q^n(10q - 9) > 0$。因此对任意 $n \geq 0$ 都有 $D_n > 0$，即

$$E_0 < E_1 < E_2 < \cdots.$$

不论 p 取何值，医生都应选 $n = 0$，立即治疗。

这个 q_0 的值不能再调低。若 $q < \frac{9}{10}$，则当 p 足够小时，$D_0 = 10p - 10(1-p)(9-10q)$ 为负，所以至少测试一次会比 $n = 0$ 更好。用文字来说，当 $q > \frac{9}{10}$ 时，Steppe 病人的测试呈阴性的机会不足十分之一，所以即使未把热病病人计算在内，一次 £10 的测试平均省下的治疗费也不足 £10。

验算

- 直接计算 $n = 1$：$E_1 = 10 + 100\big(p + (1-p)q\big)$。取 $p = 0$ 得 $E_1 - E_0 = 10 - 100(1-q)$，与公式给出的 $-10q^0(9-10q)$ 吻合。
- 当 $p = 10^{-4}$ 及 $q = 10^{-2}$ 时，公式给出 $E_0 = 100$、$E_1 \approx 11.010$、$E_2 \approx 10.1210$、$E_3 \approx 10.11309$ 及 $E_4 \approx 10.11400$。最小值出现在 $n = 3$，与上面所得一致。
- 以分数作精确枚举，已就数百组随机的 (p, q) 及 $n \le 8$ 验证了 E_n 的公式和差值恒等式；另外以 $p = 0.3$、$q = 0.5$、$n = 2$ 进行模拟，得出平均费用为 63.98，而精确值为 64。

常见错误

- 向 Steppe 病人收取全部 n 次测试的费用。整个过程在第一次出现阴性结果时便停止，所以第 k 次测试只有 q^{k-1} 的概率会进行。
- 找出第一个使 $D_n > 0$ 的 n，却没有说明 D_n 为何是单调的（monotonic）。少了这一步，数列 E_n 原则上有可能在之后再次下降。
- 在第 (ii) 部分忘记了 $p = 0$ 的情况。正因为这个情况，条件才是 $q > q_0$ 而不是 $q \ge q_0$：当 $q = \frac{9}{10}$ 且 $p = 0$ 时，每个 E_n 都等于 100。

STEP 之外

这是*最优停止*（optimal stopping）和决策论（decision theory）的一个小例子：一连串带有成本的行动、一个随机的状态，以及一条决定何时停止的法则。比较相邻的策略（即“边际分析”（marginal analysis）），并证明差值只变号一次，是处理这类问题的标准方法。这些想法会进一步通往贝叶斯决策论（Bayesian decision theory）和动态规划（dynamic programming）。Grimmett and Welsh, *Probability: An Introduction*（概率论入门教材）涵盖了支撑这些理论的条件期望（conditional expectation）工具。

详细解答

STEP III 1987，第 16 题

概率与统计 · 顺序统计量与泊松分布

STEP III 1987，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/87-s3-q16

关键思路

在第 (i) 部分，事件“$A \le k$ 且 $B \ge 1-k$”直接处理起来颇为棘手，但它的补集（complement）却是两个简单事件的并集：“所有 X_i 都大于 k”和“所有 X_i 都小于 $1-k$”。余下的工作交给容斥原理（inclusion–exclusion）便可。在第 (ii) 部分，得奖人数服从二项分布（binomial distribution），其中 n 极大而 p 极小，这正是泊松分布（Poisson distribution）大派用场的地方。因此我们先求出恰有一位或两位得奖者的泊松概率的最大值，然后再确保近似所带来的误差不会把答案推过 60%。

解答

第 (i) 部分

设 E 为事件 $A > k$，F 为事件 $B < 1-k$。那么 $A > k$ 表示每个 X_i 都大于 k，而 $B < 1-k$ 表示每个 X_i 都小于 $1-k$。由独立性（independence），

$$\mathrm{P}(E) = (1-k)^n, \qquad \mathrm{P}(F) = (1-k)^n, \qquad \mathrm{P}(E \cap F) = (1-2k)^n,$$

其中最后一式成立，是因为 $E \cap F$ 表示每个 X_i 都落在 $(k, 1-k)$ 之内，而这个区间的长度为 $1-2k > 0$。

我们要求的事件是 $E \cup F$ 的补集，所以由容斥原理得

$$p_n = 1 - \mathrm{P}(E \cup F) = 1 - 2(1-k)^n + (1-2k)^n.$$

由于 $0 < 1-2k < 1-k < 1$，两个幂都趋向 0，因此

$$p_n \to 1 \quad \text{当} n \to \infty.$$

评注。无论把 k 取得多小，只要观测值够多，我们便几乎可以肯定：有某个观测值与 0 相距不超过 k，又有另一个观测值与 1 相距不超过 k。换言之，样本最小值收敛于 0，样本最大值收敛于 1（均指依概率收敛），所以样本全距（sample range）$B-A$ 趋向 1，即整个区间的长度。这正是为甚么（举例说）A 和 B 是未知均匀分布（uniform distribution）两个端点的自然估计量（estimator）。

第 (ii) 部分

设 $N = 4\,000\,000$，并设 X 为得奖人数。各读者是否得奖互相独立，每人得奖的概率为 p，所以 $X \sim \mathrm{B}(N,p)$。记 $q = 1-p$。你能保住工作，当且仅当 $X=1$ 或 $X=2$，其概率为

$$K(p) = Npq^{N-1} + \binom{N}{2} p^2 q^{N-2}.$$

从泊松分布的角度看。由于 N 极大而 p 很小，X 近似服从平均值为 $\lambda = Np$ 的泊松分布，所以

$$K(p) \approx h(\lambda), \qquad \text{其中} h(\lambda) = \mathrm{e}^{-\lambda}\Big(\lambda + \frac{\lambda^2}{2}\Big).$$

求导得

$$h'(\lambda) = \mathrm{e}^{-\lambda}\Big(1+\lambda-\lambda-\frac{\lambda^2}{2}\Big) = \mathrm{e}^{-\lambda}\Big(1-\frac{\lambda^2}{2}\Big),$$

它在 $0<\lambda<\sqrt{2}$ 时为正，在 $\lambda>\sqrt{2}$ 时为负。所以 h 在 $\lambda=\sqrt{2}$ 处取得最大值：

$$h(\sqrt{2}) = \mathrm{e}^{-\sqrt{2}}(\sqrt{2}+1) \approx 0.2431 \times 2.4142 \approx 0.5869.$$

你的最佳选择是 $p = \sqrt{2}/N \approx 3.5\times 10^{-7}$，但即使如此，你保住工作的机会也只有约 58.7%。这就是题目所要求的推理。

想一想 泊松分布只是一个近似。真正的二项概率会不会悄悄越过 0.6?

不会。我们可以直接为 $K(p)$ 找出上界。由于 $q \le 1$ 及 $\binom{N}{2} \le \frac{1}{2}N^2$，

$$K(p) = q^{N-2}\Big(Npq + \binom{N}{2}p^2\Big) \le q^{N-2}\Big(\lambda + \frac{\lambda^2}{2}\Big).$$

对所有 p 都有 $1-p \le \mathrm{e}^{-p}$（直线 $1-p$ 是凸曲线 e^{-p} 在 $p=0$ 处的切线），所以 $q^{N-2} \le \mathrm{e}^{-(N-2)p} = \mathrm{e}^{-\mu}$，其中 $\mu = (N-2)p$。设 $c = N/(N-2)$，则 $\lambda = c\mu$，且 $c>1$。于是

$$K(p) \le \mathrm{e}^{-\mu}\Big(c\mu + \frac{c^2\mu^2}{2}\Big) \le c^2\,\mathrm{e}^{-\mu}\Big(\mu+\frac{\mu^2}{2}\Big) = c^2 h(\mu) \le c^2 h(\sqrt{2}).$$

这里 $c^2 = \big(1-\frac{2}{N}\big)^{-2} < 1.000002$，所以无论 p 取甚么值，

$$K(p) < 1.000002 \times 0.58694 < 0.587 < 0.6.$$

因此，无论你怎样选择 p，保住工作的机会都少于 60%，这正是所要证明的。

验算

- 在第 (i) 部分，取 $n = 1$ 得 $p_1 = 1 - 2(1-k) + (1-2k) = 0$。这是对的：当 $k < \frac{1}{2}$ 时，单独一个观测值不可能既 $\le k$ 又 $\ge 1-k$。取 $n = 2$ 则得 $p_2 = 2k^2$，即两点分别落在两端的窄带内（次序不拘）的概率。
- 以 $n = 10$、$k = 0.3$ 进行模拟，结果为 0.944，而公式给出 0.9436。
- 在 $\lambda = Np$ 的细密网格上计算精确的二项概率 $K(p)$，得出最大值为 0.58694，在 $\lambda \approx 1.414$ 处取得，与泊松值在小数点后五位吻合。

常见错误

- 写成 $p_n = \mathrm{P}(A \le k)\,\mathrm{P}(B \ge 1-k)$。同一个样本的最小值和最大值并不独立。
- 直接对 p 求二项概率的最大值。这样做是可行的，但代数运算相当繁琐；改用 $\lambda = Np$ 来处理，问题便化为一行就能完成的微积分题。
- 做到泊松近似便停笔，没有留意到 0.587 与 0.6 相当接近，值得补上一些论证。

STEP 之外

第 (i) 部分让我们初尝次序统计量（order statistics）和依概率收敛（convergence in probability）的滋味：样本的最小值和最大值会收敛于支集（support）的两端。第 (ii) 部分是泊松的“稀有事件定律”（law of rare events），而我们所用的上界，正是以下一般事实的一个简单例子：当 p 很小时，二项分布与泊松分布在全变差（total variation）的意义下非常接近。这两个想法在 Grimmett 与 Welsh 的 *Probability: An Introduction*，以及 Blitzstein 与 Hwang 的 *Introduction to Probability* 中都有详细阐述。

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

1987 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。

