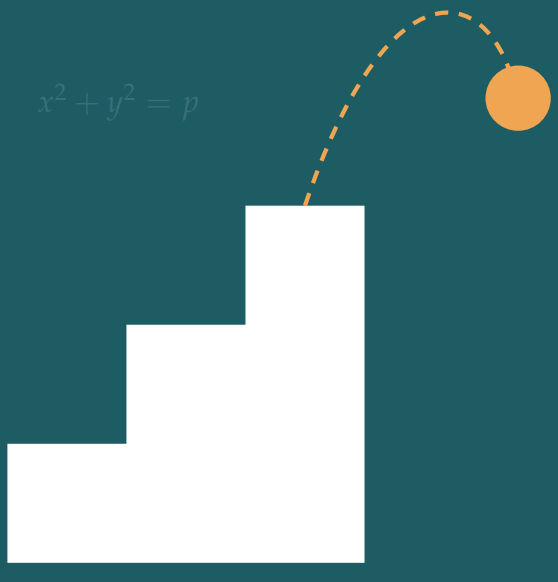
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1994 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

1994 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1994 年详细解答

1994 年 STEP 试卷（STEP I, II 及 III）全部 38 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 2 题 对数微分法 纯数学 . . . 3

第 4 题 反正切积分 纯数学 . . . 7

第 5 题 抛物线上的轨迹 纯数学 . . . 10

第 6 题 复函数的复合 纯数学 . . . 14

第 8 题 代换积分法 纯数学 . . . 17

第 9 题 抛体运动 力学 . . . 20

第 10 题 圆周运动 力学 . . . 24

第 11 题 斜面上的连接质点 力学 . . . 28

第 13 题 指示变量 概率与统计 . . . 32

第 14 题 离散分布的顺序统计量 概率与统计 . . . 35

STEP II

第 1 题 整除性与十进制数字 纯数学 . . . 38

第 2 题 勒让德多项式 纯数学 . . . 41

第 3 题 函数方程 纯数学 . . . 44

第 4 题 椭圆的面积 纯数学 . . . 48

第 5 题 实根的计数 纯数学 . . . 53

第 6 题 三角函数裂项相消级数 纯数学 . . . 57

第 7 题 阿波罗尼奥斯圆 纯数学 . . . 60

第 8 题 二次模型的最优化 纯数学 . . . 64

第 9 题 刚体的平衡 力学 . . . 67

第 10 题 弹性绳与简谐运动 力学 . . . 70

第 11 题 抛体与爆炸 力学 . . . 73

第 12 题 条件概率 概率与统计 . . . 77

第 13 题 赠券收集问题 概率与统计 80
第 14 题 几何概率 概率与统计 84

STEP III

第 1 题 归约公式 纯数学 88
第 2 题 回文四次方程 纯数学 92
第 3 题 球面、平面与扣连 纯数学 96
第 4 题 常微分方程的奇解 纯数学 100
第 5 题 麦克劳林级数 纯数学 104
第 6 题 树根图上的旋转 纯数学 108
第 7 题 群同构 纯数学 112
第 8 题 四平方和恒等式 纯数学 115
第 9 题 等时曲线与碰撞 力学 118
第 10 题 相对速度 力学 122
第 11 题 刚体的平衡 力学 126
第 12 题 随步分析 概率与统计 130
第 13 题 指数分布的寿命 概率与统计 134
第 14 题 几何概率 概率与统计 137

详细解答

STEP I 1994，第 2 题

纯数学 · 对数微分法

STEP I 1994，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/94-s1-q2

关键思路

当变量在底数而指数固定时，可以用幂法则（power rule）；当底数固定而变量在指数时，可以用指数函数的求导法则（exponential rule）。一旦 x 同时出现在这两个位置，两条法则都不能直接套用。解决办法是把所有式子写成 e 的幂，或者等价地，先取对数（logarithm），再作隐函数微分（implicit differentiation）：对数能把难缠的幂化为乘积，而乘积正是我们懂得处理的。

全题假设 $x > 0$（否则 x^x 和 $\ln x$ 都没有意义），而在 (ii) 中底数 a 必须为正数。

解答

我们会反复使用的唯一工具：对 $u > 0$，

$$u^v = \mathrm{e}^{v \ln u},$$

所以若 $y = u^v$，则 $\ln y = v \ln u$，再用链式法则（chain rule）求导，得 $\dfrac{1}{y}\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}}{\mathrm{d}x}(v \ln u)$。

第 (i) 部分

这里指数是常数，所以用的就是普通的幂法则：

$$\frac{\mathrm{d}}{\mathrm{d}x}\, x^a = a\, x^{a-1}.$$

如果你想知道为何它对任意实数 a（而不只是整数）都成立，可以写 $x^a = \mathrm{e}^{a \ln x}$；由链式法则得 $\mathrm{e}^{a \ln x} \cdot \frac{a}{x} = a\, x^a / x = a\, x^{a-1}$。

第 (ii) 部分

现在底数是常数。写成 $a^x = \mathrm{e}^{x\ln a}$，其中 $\ln a$ 只是一个常数，

$$\frac{\mathrm{d}}{\mathrm{d}x}\,a^x = \ln a\;\mathrm{e}^{x\ln a} = a^x \ln a.$$

请留意这与 (i) 有多大分别：答案并不是 $x\,a^{x-1}$。

第 (iii) 部分

设 $y = x^x$。取对数，得 $\ln y = x\ln x$。两边求导，右边用积法则（product rule），

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \ln x + x\cdot\frac{1}{x} = 1 + \ln x,$$

所以

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

想一想 为甚么不能直接用幂法则，得出 $x\cdot x^{x-1} = x^x$?

幂法则假设指数是常数，而指数函数的求导法则假设底数是常数。这里两者都在变。事实上，答案恰好是两个“错误”答案之和：把指数当作固定，得 $x\cdot x^{x-1} = x^x$；把底数当作固定，得 $x^x\ln x$；而真正的导数是 $x^x + x^x\ln x$。这其实是多元链式法则（multivariable chain rule）的变相：当 $y = u^v$ 时，每个变量各自贡献它的变化率。

第 (iv) 部分

设 $y = x^{(x^x)}$，这是一座幂塔，其指数本身就是 x^x。取对数，

$$\ln y = x^x \ln x.$$

用积法则对右边求导，其中 x^x 的导数取自 (iii)：

$$\begin{aligned}\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} &= x^x(1+\ln x)\cdot\ln x + x^x\cdot\frac{1}{x}\\ &= x^x\left(\frac{1}{x} + \ln x + (\ln x)^2\right).\end{aligned}$$

乘以 $y = x^{(x^x)}$，并把 x 的幂合并，

$$\frac{\mathrm{d}}{\mathrm{d}x}\,x^{(x^x)} = x^{(x^x)}\,x^x\left(\frac{1}{x} + \ln x + (\ln x)^2\right) = x^{x^x+x}\left(\frac{1}{x} + \ln x + (\ln x)^2\right).$$

等价地，把 $\frac{1}{x}$ 收进括号内，得 $x^{x^x+x-1}(1 + x\ln x + x(\ln x)^2)$。

第 (v) 部分

想一想 $(x^x)^x$ 不就是 (iv) 中的式子吗?

不是。根据指数定律（law of indices）$(u^m)^n = u^{mn}$，我们有 $(x^x)^x = x^{x\cdot x} = x^{x^2}$，而 (iv) 中的指数是 x^x，不是 x^2。两者碰巧在 $x = 1$ 和 $x = 2$ 时相等（在 $x = 2$ 时都等于 16），但在 $x = 3$ 时，一个是 3^{27}，另一个只是 3^9。

因此设 $y = x^{x^2}$。于是 $\ln y = x^2 \ln x$，由积法则得

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = 2x\ln x + x^2\cdot\frac{1}{x} = x\,(1+2\ln x).$$

因此

$$\frac{\mathrm{d}}{\mathrm{d}x}\,(x^x)^x = x^{x^2}\cdot x\,(1+2\ln x) = x^{x^2+1}\,(1+2\ln x).$$

（你也可以直接对 $(x^x)^x$ 求导：$\ln y = x\ln(x^x) = x^2\ln x$，结果一样。）

验算

- 在 $x = 1$ 时，(iii)、(iv) 和 (v) 的导数全都等于 1：例如在 (iv) 中，$1^{1+1}(1+0+0) = 1$。用中心差分（central difference）$\big(f(1.000001) - f(0.999999)\big)/0.000002$ 计算，结果吻合。
- x^x 在 $x = \mathrm{e}^{-1}$ 处有极小值，该处 $1+\ln x = 0$；而 $(x^x)^x = x^{x^2}$ 在 $x = \mathrm{e}^{-1/2}$ 处有极小值，该处 $1+2\ln x = 0$。两者都可以在图像计算机上看到。
- 我们在 $0.2 \le x \le 2.5$（以及 $0.2 \le a \le 4$）的范围内随机取了 2000 个点，把每条公式与数值导数比较；相对误差始终小于 10^{-9}。

常见错误

- 写成 $\frac{\mathrm{d}}{\mathrm{d}x}x^x = x\cdot x^{x-1}$ 或 $\frac{\mathrm{d}}{\mathrm{d}x}a^x = x\,a^{x-1}$：幂法则要求指数是常数。
- 把 (iv) 和 (v) 当作同一个函数。幂运算不满足结合律（associative）：一般而言 $x^{(x^x)} \neq (x^x)^x$。
- 对 $\ln y$ 求导之后，忘记乘回 y。

STEP 之外

当 a 是无理数时，例如 $x^{\sqrt{2}}$，x^a 究竟是甚么意思？大学里的答案是：先用幂级数（power series）定义 exp，再把 ln 定义为它的反函数，然后对 $u > 0$ 定义 $u^v = \mathrm{e}^{v \ln u}$；这样一来，所有指数定律和上面的导数公式都成为定理。Abbott 的 *Understanding Analysis*，或 Bartle 与 Sherbert 的 *Introduction to Real Analysis*，都对此有严谨的处理。(iii) 中“两个错误答案之和”，正是对偏导数（partial derivative）和多元链式法则的初步一瞥。

详细解答

STEP I 1994，第 4 题

纯数学 · 反正切积分

> **STEP I 1994，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/94-s1-q4

关键思路

第 (i) 部分是一条半角恒等式（half-angle identity），第 (ii) 部分则是标准的“先配方（complete the square），再用 $\tan^{-1}$”积分。最后的积分就是第 (ii) 部分取 $k = \cos\alpha$ 的情况，这时 $\sqrt{1-k^2} = \sin\alpha$，刚好与分子的因子 $\sin\alpha$ 约去。唯一真正要小心的地方在最后：对某个角的正切取 $\tan^{-1}$，只有当该角落在 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 之内时，才会得回原来的角。

解答

第 (i) 部分

利用二倍角公式（double-angle formulae）$1 - \cos\alpha = 2\sin^2\frac{1}{2}\alpha$ 及 $\sin\alpha = 2\sin\frac{1}{2}\alpha\cos\frac{1}{2}\alpha$，

$$\frac{1-\cos\alpha}{\sin\alpha} = \frac{2\sin^2\frac{1}{2}\alpha}{2\sin\frac{1}{2}\alpha\cos\frac{1}{2}\alpha} = \frac{\sin\frac{1}{2}\alpha}{\cos\frac{1}{2}\alpha} = \tan\tfrac{1}{2}\alpha,$$

只要 $\sin\alpha \neq 0$（使等式两边均有定义），上式便成立。这正是所要证明的结果。

第 (ii) 部分

将分母配方：

$$1 - 2kx + x^2 = (x-k)^2 + (1-k^2).$$

由于 $|k| < 1$，常数 $1-k^2$ 为正数，故可写成 $1-k^2 = a^2$，其中 $a = \sqrt{1-k^2} > 0$。（条件 $|k| < 1$ 正是用在这里：否则分母可以因式分解，得到的便会是对数而不是反正切。）作代

换（substitution）$x-k=a\tan\theta$，则 $\mathrm{d}x=a\sec^2\theta\,\mathrm{d}\theta$，

$$\begin{aligned}\int\frac{\mathrm{d}x}{(x-k)^2+a^2}&=\int\frac{a\sec^2\theta}{a^2\tan^2\theta+a^2}\,\mathrm{d}\theta=\int\frac{1}{a}\,\mathrm{d}\theta=\frac{\theta}{a}+C\\&=\frac{1}{\sqrt{1-k^2}}\tan^{-1}\left(\frac{x-k}{\sqrt{1-k^2}}\right)+C,\end{aligned}$$

这就是所求的结果。　（另一做法是对右边求导：由链式法则（chain rule）可得 $\dfrac{1}{a}\cdot\dfrac{1/a}{1+(x-k)^2/a^2}=\dfrac{1}{(x-k)^2+a^2}$。）

最后的积分

取 $k=\cos\alpha$。由于 $0<\alpha<\pi$，我们有 $|k|<1$，所以第 (ii) 部分适用；又因为 $\sin\alpha>0$，可得

$$\sqrt{1-k^2}=\sqrt{\sin^2\alpha}=\sin\alpha$$

（而不是 $-\sin\alpha$，因为在这个区间上 $\sin\alpha>0$）。因此分子的因子 $\sin\alpha$ 与第 (ii) 部分的因子 $1/\sqrt{1-k^2}$ 互相约去：

$$\begin{aligned}\int_0^1\frac{\sin\alpha}{1-2x\cos\alpha+x^2}\,\mathrm{d}x&=\left[\tan^{-1}\left(\frac{x-\cos\alpha}{\sin\alpha}\right)\right]_0^1\\&=\tan^{-1}\left(\frac{1-\cos\alpha}{\sin\alpha}\right)-\tan^{-1}\left(\frac{-\cos\alpha}{\sin\alpha}\right)\\&=\tan^{-1}\left(\tan\tfrac{1}{2}\alpha\right)+\tan^{-1}(\cot\alpha),\end{aligned}$$

其中第一项用了第 (i) 部分，第二项则用了 $\tan^{-1}$ 是奇函数（odd function）这一事实。

> **想一想** $\tan^{-1}(\tan\theta)$ 是否总是等于 θ?

不是：$\tan^{-1}$ 的取值范围是 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$，所以只有当 θ 在这个区间内时，才有 $\tan^{-1}(\tan\theta)=\theta$。我们逐项检查。

- 由于 $0<\alpha<\pi$，我们有 $0<\frac{1}{2}\alpha<\frac{\pi}{2}$，所以 $\tan^{-1}\left(\tan\frac{1}{2}\alpha\right)=\frac{1}{2}\alpha$。
- 写成 $\cot\alpha=\tan\left(\frac{\pi}{2}-\alpha\right)$。由于 $0<\alpha<\pi$，角 $\frac{\pi}{2}-\alpha$ 落在 $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ 之内，所以 $\tan^{-1}(\cot\alpha)=\frac{\pi}{2}-\alpha$。

两项相加，

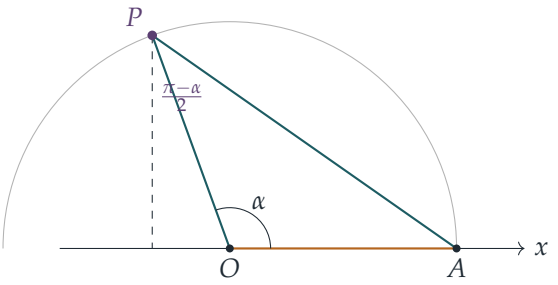
$$\int_0^1\frac{\sin\alpha}{1-2x\cos\alpha+x^2}\,\mathrm{d}x=\frac{\alpha}{2}+\frac{\pi}{2}-\alpha=\frac{\pi-\alpha}{2},$$

这就是所求的结果。

另一做法：一幅图

这个答案背后有一个很漂亮的几何原因。设 $P=(\cos\alpha,\sin\alpha)$ 为单位圆上的一点，O 为原点，$A=(1,0)$。当点 $X=(x,0)$ 沿 x 轴移动时，PX 与过 P 向下的铅垂线之间的夹角 θ 满

足 $\tan\theta = (x-\cos\alpha)/\sin\alpha$，而被积函数（integrand）正好就是 $\mathrm{d}\theta/\mathrm{d}x$。所以这个积分就是 X 由 O 移到 A 时所扫过的角，即角 OPA。三角形 OPA 是等腰三角形，$OP = OA = 1$，在 O 处的顶角为 α，所以两个底角各为 $\frac{1}{2}(\pi-\alpha)$。



验算

- 当 $\alpha = \frac{\pi}{2}$ 时，积分为 $\int_0^1 \frac{\mathrm{d}x}{1+x^2} = \frac{\pi}{4}$，而公式给出 $\frac{1}{2}\left(\pi - \frac{\pi}{2}\right) = \frac{\pi}{4}$。
- 数值上，取 $\alpha = 1$，以辛普森法则（Simpson's rule）计算得 $1.0707963\ldots$，与 $\frac{1}{2}(\pi-1) = 1.0707963\ldots$ 吻合。
- 当 $\alpha \to \pi^-$ 时，对 $[0,1]$ 内的每个 x，被积函数都趋向 0，而 $\frac{1}{2}(\pi-\alpha)$ 亦然。

常见错误

- 写下 $\tan^{-1}(\cot\alpha) = \frac{\pi}{2} - \alpha$，却没有检查 $\frac{\pi}{2} - \alpha$ 是否落在 $\tan^{-1}$ 的值域之内；当 α 在 $(0,\pi)$ 以外时，同样的做法便会出错。
- 把 $\sqrt{1-\cos^2\alpha} = |\sin\alpha|$ 当作 $\sin\alpha$，却没有说明原因；这里的原因是 $0 < \alpha < \pi$。

STEP 之外

当 $\alpha \to 0^+$ 时，答案趋向 $\frac{\pi}{2}$，尽管在 $\alpha = 0$ 时被积函数恒等于零：被积函数在 $x = 1$ 附近变成一个又高又窄的尖峰，而其面积并不缩小。这就是半平面的泊松核（Poisson kernel），而随意交换极限与积分的次序便会得出错误的答案。这种交换在甚么情况下才容许，是大学一年级分析学（analysis）的核心问题；参见 Abbott 的 *Understanding Analysis*。至于“所对的角”这个几何观点，Needham 的 *Visual Complex Analysis* 有十分精彩的阐述。

详细解答

STEP I 1994，第 5 题

纯数学 · 抛物线上的轨迹

> **STEP I 1994，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/94-s1-q5

关键思路

先把几何条件化为代数条件。原点处的角是直角，即位置向量（position vector）$\overrightarrow{OP}$ 与 $\overrightarrow{OQ}$ 互相垂直；对于 $y = x^2$ 上的点，这个条件可简化为 p 与 q 之间一个简洁的关系式。这样一来，R 和 T 两点的坐标都是关于 p 和 q 对称的，所以我们把一切都用 $p+q$ 和 pq 表示，然后消元。

解答

直角的条件

P 和 Q 都不可能是原点，否则该角便没有定义，所以 $p \neq 0$ 且 $q \neq 0$。角 POQ 是直角，当且仅当 $\overrightarrow{OP} = (p, p^2)$ 与 $\overrightarrow{OQ} = (q, q^2)$ 的纯量积（scalar product）为零：

$$pq + p^2q^2 = 0 \quad \Longleftrightarrow \quad pq(1+pq) = 0 \quad \Longleftrightarrow \quad pq = -1,$$

其中最后一步除以 $pq \neq 0$。（等价的说法是：OP 与 OQ 的斜率（gradient）分别为 p 和 q，两者之积为 -1。）

因此我们要找的数对，正是 $q = -1/p$，其中 $p \neq 0$。当 p 取遍所有非零实数时，两数之和

$$s = p + q = p - \frac{1}{p}$$

可取得每一个实数值：任意给定 s，二次方程 $p^2 - sp - 1 = 0$ 的判别式（discriminant）为 $s^2 + 4 > 0$，所以它有实根，而且两根之积为 -1，故两根皆不为零。我们稍后需要这个结论，才能说明下面的轨迹（locus）是完整的曲线，而不只是曲线的一部分。

第 (i) 部分

PQ 的中点（midpoint）是

$$R = \left(\frac{p+q}{2}, \frac{p^2+q^2}{2} \right).$$

设 $R = (x, y)$，则 $p + q = 2x$。利用 $pq = -1$，

$$y = \frac{(p+q)^2 - 2pq}{2} = \frac{4x^2+2}{2} = 2x^2 + 1.$$

由于 $p + q$ 可取得每一个实数值，x 亦然，所以 R 的轨迹是整条抛物线（parabola）

$$y = 2x^2 + 1,$$

其顶点（vertex）为 $(0, 1)$，对称轴为 y 轴。它比 $y = x^2$ 窄，而且完全位于后者的上方，因为 $2x^2 + 1 - x^2 = x^2 + 1 > 0$；这与 R 是凸（convex）曲线上一条弦（chord）的中点这一事实吻合。

第 (ii) 部分

对 $y = x^2$ 求导，可知在 P 的切线（tangent）斜率为 $2p$，所以切线方程为

$$y - p^2 = 2p(x - p), \qquad \text{即} \qquad y = 2px - p^2,$$

同理，在 Q 的切线为 $y = 2qx - q^2$。这两条切线并不平行，因为 $p \neq q$（两者之积为 -1，所以不可能相等）。把两条方程相减，

$$2(p - q)x = p^2 - q^2 \quad \Longrightarrow \quad x = \frac{p+q}{2},$$

其中除以了 $p - q \neq 0$。于是

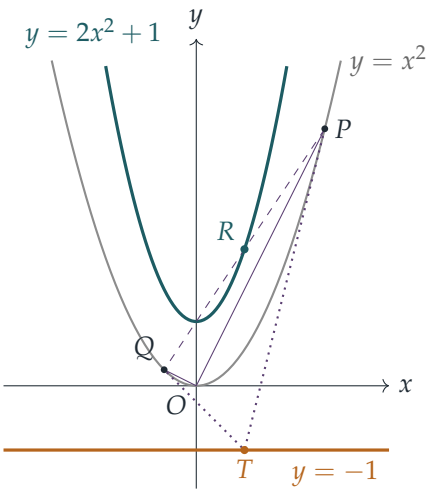
$$y = 2p \cdot \frac{p+q}{2} - p^2 = pq = -1.$$

想一想 T 的 y 坐标是常数。那么轨迹是单独一点、一条线段，还是整条直线?

T 的 x 坐标是 $\frac{1}{2}(p + q)$，而我们已证明它可取得每一个实数值。所以 T 可以位于该直线上的任何位置，T 的轨迹就是整条直线

$$y = -1.$$

还可留意，T 总是在 R 的正下方：两者的 x 坐标都是 $\frac{1}{2}(p + q)$。



图中所示是其中一个位置，$p = 2$，$q = -\frac{1}{2}$：此时 $R = (\frac{3}{4}, \frac{17}{8})$ 在蓝绿色的抛物线上，而 $T = (\frac{3}{4}, -1)$ 在琥珀色的直线上。

验算

取 $p = 1$，$q = -1$：则 $P = (1, 1)$，$Q = (-1, 1)$，而 OP、OQ 就是直线 $y = \pm x$，显然互相垂直。中点是 $(0, 1)$，即 $y = 2x^2 + 1$ 的顶点；两条切线 $y = 2x - 1$ 与 $y = -2x - 1$ 相交于 $(0, -1)$，在 $y = -1$ 上。再取图中的 $p = 2$，$q = -\frac{1}{2}$，则 $2 \cdot (\frac{3}{4})^2 + 1 = \frac{17}{8}$，与预期相符。另外以 10^5 个随机的 p 值作数值检验，结果确认 R 满足 $y = 2x^2 + 1$，误差约在 10^{-10} 以内，而 T 的坐标总是满足 $y = -1$。

常见错误

- 写到 $pq(1 + pq) = 0$ 便停下，没有解释为何可排除 $pq = 0$：这种情况会令 P 或 Q 落在原点。
- 只写出轨迹的方程，却没有检查是否每一个 x 值都能取得，以确定所画的草图应是整条曲线而不是其中一部分。
- 以为 T 位于准线（directrix）上。$y = x^2$ 的准线是 $y = -\frac{1}{4}$；当 PQ 通过焦点（focus）时，两条切线才会在准线上相交，这与 $\angle POQ = 90^\circ$ 是不同的条件。

STEP 之外

这里的弦 PQ 总是通过定点 $(0,1)$：它的方程是 $y=(p+q)x-pq=(p+q)x+1$。通过一个定点的直线族，以及与之对应、由切线相交而成的“极点”（pole）（这里它们位于直线 $y=-1$ 上，即 $(0,1)$ 的极线（polar）），正是射影几何（projective geometry）的起点；射影几何会有系统地研究圆锥曲线（conics）及其极点与极线的关系。想做更多同类题目，可参考 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP I 1994，第 6 题

纯数学 · 复函数的复合

STEP I 1994，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/94-s1-q6

关键思路

要求一个商的实部（real part）和虚部（imaginary part），可把分子和分母同乘以分母的共轭（conjugate）；由于 x 是实数，这个共轭很容易写出来。至于重复施行函数的部分，与其把 (i) 中繁复的实部和虚部再代入公式一次，不如对一般的 z 以*代数方法*把 f 与自身复合（compose），这样简洁得多。一旦把 $\mathrm{f}(\mathrm{f}(z))$ 化成整齐的形式，最后一部分只需一行代入便可完成。

解答

由于 x 是实数，$\mathrm{i}x+1$ 的实部为 1，所以它永不为零，因此 $\mathrm{f}(x)$ 对每个实数 x 都有定义。

第 (i) 部分

因为 x 是实数，$\mathrm{i}x+1$ 的共轭是 $1-\mathrm{i}x$。把分子和分母同乘以它，

$$\begin{aligned}\mathrm{f}(x) = \frac{(\mathrm{i}x-1)(1-\mathrm{i}x)}{(1+\mathrm{i}x)(1-\mathrm{i}x)} &= \frac{\mathrm{i}x+x^2-1+\mathrm{i}x}{1+x^2} \\ &= \frac{x^2-1}{x^2+1} + \mathrm{i}\,\frac{2x}{x^2+1}.\end{aligned}$$

两个分数都是实数，所以它们正是所求的实部和虚部。

第 (ii) 部分

对任何复数 w，$ww^* = (\operatorname{Re} w)^2 + (\operatorname{Im} w)^2$。利用第 (i) 部分，

$$\mathrm{f}(x)\mathrm{f}(x)^* = \frac{(x^2-1)^2+4x^2}{(x^2+1)^2} = \frac{x^4+2x^2+1}{(x^2+1)^2} = 1.$$

所以 $|\mathrm{f}(x)| = 1$：f 把每个实数映射到单位圆（unit circle）上的一点。

第 (iii) 部分

我们不把 (i) 的答案代入 f，而是对一般的 z 进行复合。设 $w = \mathrm{f}(z) = \dfrac{\mathrm{i}z-1}{\mathrm{i}z+1}$，并把 $\mathrm{f}(w) = \dfrac{\mathrm{i}w-1}{\mathrm{i}w+1}$ 的分子和分母同乘以 $\mathrm{i}z+1$：

$$\begin{aligned}\mathrm{f}(\mathrm{f}(z)) &= \frac{\mathrm{i}(\mathrm{i}z-1)-(\mathrm{i}z+1)}{\mathrm{i}(\mathrm{i}z-1)+(\mathrm{i}z+1)} = \frac{-z-\mathrm{i}-\mathrm{i}z-1}{-z-\mathrm{i}+\mathrm{i}z+1}\\ &= \frac{-(1+\mathrm{i})(z+1)}{(1-\mathrm{i})(1-z)} = \frac{1+\mathrm{i}}{1-\mathrm{i}}\cdot\frac{z+1}{z-1} = \mathrm{i}\,\frac{z+1}{z-1},\end{aligned}$$

这是因为 $\dfrac{1+\mathrm{i}}{1-\mathrm{i}} = \dfrac{(1+\mathrm{i})^2}{2} = \mathrm{i}$。

> **想一想** $\mathrm{f}(\mathrm{f}(x))$ 是否对每个实数 x 都有定义?

不是。我们需要 $\mathrm{i}w+1 \neq 0$，即 $\mathrm{f}(x) \neq \mathrm{i}$。由第 (i) 部分，$\mathrm{f}(x) = \mathrm{i}$ 需要 $x^2-1=0$ 及 $2x = x^2+1$，而这只在 $x=1$ 时成立。这与上式分母中的因式 $x-1$ 吻合。所以对实数 $x \neq 1$，

$$\mathrm{f}(\mathrm{f}(x)) = \mathrm{i}\,\frac{x+1}{x-1}, \qquad \text{所以} \qquad \operatorname{Re}\mathrm{f}(\mathrm{f}(x)) = 0, \quad \operatorname{Im}\mathrm{f}(\mathrm{f}(x)) = \frac{x+1}{x-1}.$$

第一次施行 f 把实数线映射到单位圆，第二次则把单位圆映射到虚轴。

第 (iv) 部分

写 $\mathrm{f}(\mathrm{f}(x)) = \mathrm{i}y$，其中 $y = \dfrac{x+1}{x-1}$ 是实数。于是 $\mathrm{i}(\mathrm{i}y) = -y$，所以

$$\mathrm{f}(\mathrm{i}y) = \frac{-y-1}{-y+1} = \frac{y+1}{y-1}.$$

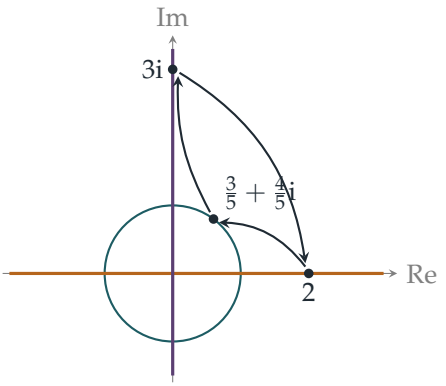
这个式子有定义，因为 $y=1$ 需要 $x+1 = x-1$，而这是不可能的。现在

$$y+1 = \frac{2x}{x-1}, \qquad y-1 = \frac{2}{x-1},$$

所以

$$\mathrm{f}(\mathrm{f}(\mathrm{f}(x))) = x \qquad \text{对每个实数} x \neq 1.$$

施行 f 三次便回到起点：第三次施行把虚轴送回实数线。下图显示 $x=2$ 时的循环，即 $2 \mapsto \frac{3}{5}+\frac{4}{5}\mathrm{i} \mapsto 3\mathrm{i} \mapsto 2$。



验算

- $x=0$：$\mathrm{f}(0)=-1$，它在单位圆上，与 (i) 一致。然后 $\mathrm{f}(-1)=\dfrac{-\mathrm{i}-1}{-\mathrm{i}+1}=-\mathrm{i}$，与 $\mathrm{i}\frac{0+1}{0-1}$ 吻合，而 $\mathrm{f}(-\mathrm{i})=\frac{1-1}{1+1}=0$。
- $x=2$：第 (i) 部分给出 $\frac{3}{5}+\frac{4}{5}\mathrm{i}$，第 (iii) 部分给出 3i，第 (iv) 部分则回到 2。以 10^4 个随机实数 x（以及随机复数 z）作数值检验，结果证实以上每条公式都准确至舍入误差之内。

常见错误

- 把 (i) 的实部和虚部直接代入 f。这样做是可行的，但代数运算繁重，容易出错。以 z 来进行复合则简洁得多。
- 忘记在 (iii) 和 (iv) 中必须排除 $x=1$，因为 $\mathrm{f}(1)=\mathrm{i}$，而 $\mathrm{f}(\mathrm{i})$ 的分母为零。

STEP 之外

形如 $z\mapsto\dfrac{az+b}{cz+d}$ 的函数称为莫比乌斯变换（*Möbius transformation*）。这类映射把圆和直线映射到圆和直线，正如我们在这里所见：实轴到单位圆，再到虚轴，然后返回实轴。把两个这样的变换复合，相当于把 2×2 矩阵 $\begin{pmatrix}a&b\\c&d\end{pmatrix}$ 相乘。事实上，对每个复数 z 都有 $\mathrm{f}(\mathrm{f}(\mathrm{f}(z)))=z$，所以 f 在莫比乌斯变换所组成的群（group）中的阶（order）为 3。此外，在 (i) 中代入 $x=\tan\frac{\theta}{2}$，可得 $\mathrm{f}(x)=-\cos\theta+\mathrm{i}\sin\theta$，这就是大家熟悉的“$t$ 代换”圆参数化（parametrisation）。Needham 的 *Visual Complex Analysis* 中有一章精彩地讲述莫比乌斯变换及其几何。

详细解答

STEP I 1994，第 8 题

纯数学 · 代换积分法

STEP I 1994，第 8 题

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

step.maths.org/questions/94-s1-q8

关键思路

代换(substitution)$\theta = \frac{1}{4}\pi - \phi$ 把区间 $[0, \frac{1}{4}\pi]$ 反射到自身,而复角公式(addition formula)则把 $1 + \tan(\frac{1}{4}\pi - \phi)$ 化为 $2/(1 + \tan\phi)$。于是原来的积分会带着负号在等式右边重现,我们只须把它解出来，而不必计算任何困难的积分。后面两个积分其实是同一个积分的变相：一个经代换 $x = \tan\theta$ 得出，另一个则经半角公式（half-angle formulae）得出。

解答

第一个积分

把这个积分记为 I。设 $\theta = \frac{1}{4}\pi - \phi$，则 $\mathrm{d}\theta = -\mathrm{d}\phi$；当 $\theta = 0$ 时，$\phi = \frac{1}{4}\pi$，而当 $\theta = \frac{1}{4}\pi$ 时，$\phi = 0$。因此

$$I = \int_{\frac{1}{4}\pi}^{0} \ln\bigl(1 + \tan(\tfrac{1}{4}\pi - \phi)\bigr)\,(-\mathrm{d}\phi) = \int_{0}^{\frac{1}{4}\pi} \ln\bigl(1 + \tan(\tfrac{1}{4}\pi - \phi)\bigr)\,\mathrm{d}\phi.$$

利用复角公式及 $\tan\frac{1}{4}\pi = 1$,

$$1 + \tan(\tfrac{1}{4}\pi - \phi) = 1 + \frac{1 - \tan\phi}{1 + \tan\phi} = \frac{2}{1 + \tan\phi}.$$

(在 $[0, \frac{1}{4}\pi]$ 上有 $\tan\phi \geq 0$，所以没有任何式子等于零，而每个对数都是取正数的对数。)故此

$$I = \int_{0}^{\frac{1}{4}\pi} \bigl(\ln 2 - \ln(1 + \tan\phi)\bigr)\,\mathrm{d}\phi = \tfrac{1}{4}\pi \ln 2 - I,$$

因为最后一个积分正是 I，只是哑变量（dummy variable）换了名称。

想一想 我们由始至终没有求出原函数（antiderivative）。为甚么可以这样做?

我们根本不需要原函数。I 只是一个数，而代换给出了这个数必须满足的方程 $I = \frac{1}{4}\pi\ln 2 - I$。解之，得

$$2I = \tfrac{1}{4}\pi\ln 2 \quad\Longrightarrow\quad \int_0^{\frac{1}{4}\pi} \ln(1+\tan\theta)\,\mathrm{d}\theta = \tfrac{1}{8}\pi\ln 2,$$

正是所求。

含 $1+x^2$ 的积分

因子 $\dfrac{\mathrm{d}x}{1+x^2}$ 提示我们设 $x = \tan\theta$，于是 $\mathrm{d}x = \sec^2\theta\,\mathrm{d}\theta = (1+x^2)\,\mathrm{d}\theta$。积分上下限 $x = 0$ 和 $x = 1$ 变成 $\theta = 0$ 和 $\theta = \frac{1}{4}\pi$。因此

$$\int_0^1 \frac{\ln(1+x)}{1+x^2}\,\mathrm{d}x = \int_0^{\frac{1}{4}\pi} \ln(1+\tan\theta)\,\mathrm{d}\theta = \tfrac{1}{8}\pi\ln 2.$$

含 $\sin x$ 和 $\cos x$ 的积分

把它记为 J。最快捷的方法是利用对称性（symmetry）：代换 $x = \frac{1}{2}\pi - u$ 把 sin 和 cos 互换，并把 $[0, \frac{1}{2}\pi]$ 映射到自身，所以

$$J = \int_0^{\frac{1}{2}\pi} \ln\left(\frac{1+\cos u}{1+\sin u}\right)\mathrm{d}u = -J,$$

因此 $J = 0$。（换个说法，$\int_0^{\frac{1}{2}\pi}\ln(1+\sin x)\,\mathrm{d}x$ 与 $\int_0^{\frac{1}{2}\pi}\ln(1+\cos x)\,\mathrm{d}x$ 相等，所以两者之差为零。）

想一想 第一个结果在哪里派上用场?

J 与 I 之间的关系值得一看，这大概也是题目把它放在这里的原因。设 $t = \frac{1}{2}x$，由半角公式得

$$1+\cos x = 2\cos^2 t, \qquad 1+\sin x = (\cos t + \sin t)^2,$$

所以

$$\frac{1+\sin x}{1+\cos x} = \frac{(\cos t+\sin t)^2}{2\cos^2 t} = \tfrac{1}{2}(1+\tan t)^2.$$

当 x 取遍 $[0, \frac{1}{2}\pi]$ 时，t 取遍 $[0, \frac{1}{4}\pi]$，且 $\mathrm{d}x = 2\,\mathrm{d}t$。因此

$$J = \int_0^{\frac{1}{4}\pi} \big(2\ln(1+\tan t) - \ln 2\big)\,2\,\mathrm{d}t = 4I - \tfrac{1}{2}\pi\ln 2 = \tfrac{1}{2}\pi\ln 2 - \tfrac{1}{2}\pi\ln 2 = 0,$$

与对称性的论证吻合。所以

$$\int_0^{\frac{1}{2}\pi} \ln\left(\frac{1+\sin x}{1+\cos x}\right)\mathrm{d}x = 0.$$

验算

用辛普森法则（Simpson's rule）计算，tan 的积分和含 $1+x^2$ 的积分都得出 0.2721982613，与 $\frac{1}{8}\pi\ln 2 = 0.2721982613\ldots$ 相符；第三个积分在机器精度之内等于 0。也可以用手粗略验算：被积函数（integrand）$\ln(1+\tan\theta)$ 由 0 上升至 $\ln 2$，所以 I 应约为 $\frac{1}{4}\pi\ln 2$ 的一半。这恰好就是答案。

常见错误

- 忘记了 $\mathrm{d}\theta = -\mathrm{d}\phi$，同时也忘记了上下限要对调。这两次变号会互相抵消；若只保留其中一次，便会得出 $I = -\frac{1}{4}\pi\ln 2 - I$（以及一个正积分竟得负数答案）。
- 写成 $\tan(\frac{1}{4}\pi - \phi) = 1 - \tan\phi$，漏掉了复角公式中的分母。
- 在第三个积分上花了很长时间，却没有留意到 $x \mapsto \frac{1}{2}\pi - x$ 这个对称性。

STEP 之外

以反射作代换，再把结果与原式相加，从而得到 $\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(a+b-x)\,\mathrm{d}x$，这个技巧在某些书中称为 King 性质（King property）。它同样能给出经典结果 $\int_0^{\frac{1}{2}\pi}\ln\sin x\,\mathrm{d}x = -\frac{1}{2}\pi\ln 2$。在大学里，这类结果属于含参数定积分（definite integrals with parameters）和围道积分（contour integration）的理论，当中一整族的积分都可以不求原函数而算出来。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）展示了复变方法如何得出这类结果，而 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）则有更多运用这类对称技巧的 STEP 式练习。

详细解答

STEP I 1994，第 9 题

力学 • 抛体运动

STEP I 1994，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/94-s1-q9

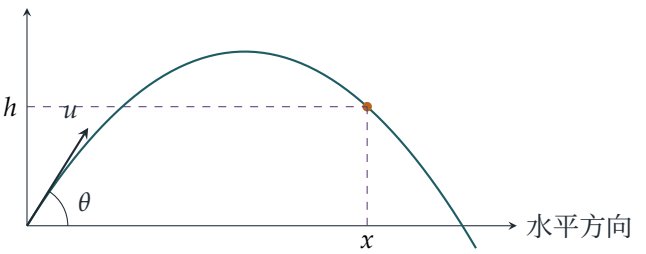
关键思路

发射角 θ 是未知的量，所以要把它消去：水平方向和垂直方向的方程分别含有 $u\cos\theta$ 或 $u\sin\theta$，把两式平方后相加，就可以完全消去 θ。这样得出的方程是 t^2 的二次方程（quadratic），而二次方程只有在判别式（discriminant）非负时才有实根（real roots）。这个条件正好就是射程的上界。

解答

关于 t 的方程

把炮弹看成只受重力作用的质点（particle），从原点以仰角 θ 发射。经过时间 t 后，它的水平距离为 x，高度为 h，所以它与大炮的距离是 $\sqrt{x^2+h^2}$，与题目所述相符。



利用向下的恒定加速度（constant acceleration）g，以及水平方向不受力，

$$x = ut\cos\theta,$$
$$h = ut\sin\theta - \tfrac{1}{2}gt^2.$$

把含 θ 的项分离出来：

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

两式平方后相加，并利用 $\cos^2\theta + \sin^2\theta = 1$，

$$\begin{aligned} u^2t^2 &= x^2 + \left(h + \tfrac{1}{2}gt^2\right)^2 \\ &= x^2 + h^2 + ght^2 + \tfrac{1}{4}g^2t^4, \end{aligned}$$

整理后得

$$\tfrac{1}{4}g^2t^4 - (u^2 - gh)t^2 + h^2 + x^2 = 0,$$

这正是所求。

射程的上界

现在固定 u 和 h，考虑在高度 h 处可以达到哪些水平距离 x。如果炮弹在某时刻 t 位于 (x, h)，那么 $T = t^2$ 就是以下方程的一个实根：

$$\tfrac{1}{4}g^2T^2 - (u^2 - gh)T + (h^2 + x^2) = 0.$$

> **想一想** 这条方程涉及 x、h、u 和 t，但我们想要的是一个只关于 x 的结论。二次方程若有实根，必须满足甚么条件？

有实根就需要判别式非负：

$$\begin{aligned} (u^2 - gh)^2 - 4\cdot\tfrac{1}{4}g^2(h^2 + x^2) &\geq 0 \\ u^4 - 2u^2gh + g^2h^2 - g^2h^2 - g^2x^2 &\geq 0 \\ g^2x^2 &\leq u^2(u^2 - 2gh). \end{aligned}$$

由于 $u^2 > 2gh$，右边为正数，两边取非负平方根得

$$x \leq \frac{u\sqrt{u^2 - 2gh}}{g}.$$

所以在高度 h 处的水平射程永远不会超过这个值，这正是我们要证明的。（如果 $u^2 < 2gh$，则不等式对任何 x 都不成立：高度 h 超过了炮弹所能达到的最大高度 $u^2/2g$。）

上界可以达到

等号成立表示判别式为零，所以方程有重根（double root）

$$t^2 = T = \frac{u^2 - gh}{2\cdot\frac{1}{4}g^2} = \frac{2(u^2 - gh)}{g^2}.$$

这个值是正数（若 $h \geq 0$，因为 $u^2 > 2gh \geq gh$；若 $h < 0$，则显然成立），所以它确实给出一个时刻 $t > 0$。我们仍须检查，是否有某个角 θ 能在 $x = u\sqrt{u^2 - 2gh}/g$ 的情况下给出这个时刻。由两条分量方程，我们需要

$$\sin\theta = \frac{h + \frac{1}{2}gt^2}{ut}, \qquad \cos\theta = \frac{x}{ut}.$$

代入上面 t^2 的值，$h + \frac{1}{2}gt^2 = h + (u^2 - gh)/g = u^2/g$，所以

$$\sin\theta = \frac{u}{gt}, \qquad \cos\theta = \frac{\sqrt{u^2 - 2gh}}{gt}.$$

想一想 我们分别为 $\sin\theta$ 和 $\cos\theta$ 选定了数值。怎样知道有同一个角同时取得这两个值?

两个值都是正数，而且根据 t^2 的值，

$$\sin^2\theta + \cos^2\theta = \frac{u^2 + u^2 - 2gh}{g^2t^2} = \frac{2(u^2 - gh)}{g^2t^2} = 1$$

。所以在 0 与 $\frac{1}{2}\pi$ 之间恰好有一个角 θ 符合要求，即

$$\tan\theta = \frac{u}{\sqrt{u^2 - 2gh}}.$$

以这个角度发射，炮弹在时刻 t 位于高度 h，水平距离恰好是 $u\sqrt{u^2 - 2gh}/g$，所以当 $u^2 > 2gh$ 时，上界总是可以达到的。

留意当 $h > 0$ 时 $\tan\theta > 1$: 要在高于大炮的某个指定高度上射得最远, 发射角要比 45° 更陡。此外，该时刻的垂直速度为 $u\sin\theta - gt = (u^2 - g^2t^2)/(gt) = (2gh - u^2)/(gt) < 0$, 所以炮弹是在下降途中到达高度 h 处的最远点。

验算

- 当 $h = 0$ 时，上界为 u^2/g，而 $\tan\theta = 1$：这就是大家熟悉的平地最大射程，发射角为 45°。
- 当 $u = 20\,\mathrm{m\,s^{-1}}$、$g = 9.8\,\mathrm{m\,s^{-2}}$ 而 $h = 10\,\mathrm{m}$ 时，上界约为 29.15 m，此时 $\theta \approx 54.5°$。
- 我们以数值方法确认了该四次方程（quartic）在随机选取的轨迹上均成立，而且扫描数千个发射角，结果从未超过上界，但可以接近至相差不足 10^{-4}。上面求得的角则恰好达到上界。

常见错误

- 尝试把 x 对 θ 求导来找出最佳的 θ。这样做是可以的，但过程繁复。判别式已经替你完成了求最优值的工作。
- 忘记最后一步：判别式为零只给出 t 的一个值，但你仍须验证 $\sin^2\theta + \cos^2\theta = 1$，以证明确实有一个实数角能给出这个值。

STEP 之外

炮弹以速率 u 发射时所能到达的点构成一个区域，其边界是所有轨迹的 包络线（envelope），即“安全抛物线”（parabola of safety）$y = \frac{u^2}{2g} - \frac{gx^2}{2u^2}$。把本题的上界重新整理，可以看出 (x, h) 可以到达，当且仅当它位于这条抛物线上或其下方。曲线族的包络线同样源自“重根”这个想法，并且会在变分法（calculus of variations）和光学（焦散线，caustics）中再次出现。想进一步了解抛体（projectiles）及其背后的力学，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）。

详细解答

STEP I 1994，第 10 题

力学 · 圆周运动

> **STEP I 1994，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/94-s1-q10

关键思路

这是一个用可伸长的绳系着的圆锥摆（conical pendulum）。质点没有垂直方向的加速度，所以张力（tension）的垂直分量与重量平衡，而水平分量则提供向心力（centripetal force）。新的地方在于圆的半径取决于伸长量（extension），而伸长量又取决于张力，于是虎克定律（Hooke's law）把一切连系起来。至于不等式，则来自这个问题：$\cos\theta$ 取哪些值才真正描述一个圆?

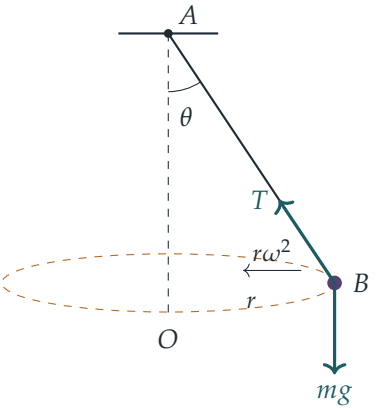
解答

求 $\cos\theta$

把 B 视为质点（particle），把 AB 视为符合虎克定律的轻弹性绳（light elastic string）。设伸长量为 x，张力为 T，则

$$T = \frac{\lambda x}{l}, \qquad AB = l + x.$$

质点绕着 A 正下方的点 O，在半径为 $r = (l + x)\sin\theta$ 的水平圆上运动。作用于质点的力只有它的重量和沿 BA 方向的张力。



沿垂直方向分解（没有垂直加速度）：

$$T\cos\theta = mg. \tag{1}$$

沿水平方向分解，指向圆心 O（加速度为 $r\omega^2$）：

$$T\sin\theta = m\omega^2(l+x)\sin\theta.$$

要是真正的圆，必须 $\sin\theta \neq 0$，所以可以把它约去：

$$T = m\omega^2(l+x). \tag{2}$$

现在由虎克定律得 $\lambda x/l = m\omega^2(l+x)$，即

$$x\left(\lambda - m\omega^2 l\right) = m\omega^2 l^2.$$

由于 $x > 0$（由 (1) 得 $T > 0$, 所以绳处于拉伸状态）而且 $m\omega^2 l^2 > 0$, 可知 $\lambda - m\omega^2 l > 0$, 并且

$$x = \frac{m\omega^2 l^2}{\lambda - m\omega^2 l}, \qquad l + x = \frac{l\lambda}{\lambda - m\omega^2 l}.$$

于是由 (2) 得

$$T = \frac{m\omega^2 l\lambda}{\lambda - m\omega^2 l},$$

再由 (1) 得

$$\boxed{\cos\theta = \frac{mg}{T} = \frac{g\left(\lambda - m\omega^2 l\right)}{l\lambda\,\omega^2}}$$

运动何时可能发生

> ? **想一想** 对 $\cos\theta$ 的每一个限制，在物理上各代表甚么?

要质点在 A 下方的水平圆上运动，而且绳保持拉紧，我们需要 $0 < \theta < \frac{1}{2}\pi$，即 $0 < \cos\theta < 1$。反过来说，如果 ω 代入公式后得出 $0 < \cos\theta < 1$，那么把上述步骤逆推回去（其中 $\lambda - m\omega^2 l > 0$），便会得到满足 (1) 和 (2) 的正伸长量和正张力，所以这个运动确实存在。

上界。由于其余因子均为正数，$\cos\theta > 0$ 成立当且仅当 $\lambda - m\omega^2 l > 0$。这就是

$$\omega^2 < \frac{\lambda}{ml}.$$

物理意义：如果 ω 太大，绳无论伸长多少（只要是有限的）都无法提供所需的向心力，因为多出来的长度所需要的额外的力，增长得比绳所能提供的还要快。

下界。$\cos\theta < 1$ 即是

$$g\lambda - gm\omega^2 l < l\lambda\omega^2 \iff g\lambda < \omega^2 l(\lambda + mg) \iff \omega^2 > \frac{g\lambda}{l(\lambda + mg)},$$

其中我们先除以 $l\lambda\omega^2 > 0$，最后除以 $l(\lambda + mg) > 0$，所以不等号的方向保持不变。物理意义：如果 ω 太小，公式便会要求 $\cos\theta \geq 1$；唯一的可能就是质点静止地悬挂在轴上（$\theta = 0$），而这并不是圆周运动（circular motion）。

把两者合起来，这个运动只有在以下条件下才可能发生：

$$\frac{g\lambda}{l(\lambda + mg)} < \omega^2 < \frac{\lambda}{ml},$$

这正是所要证明的结果。

验算

- 在下界处，$\cos\theta = 1$：这时绳是垂直的，张力为 mg，长度为 $l(\lambda + mg)/\lambda$，恰好就是质点悬挂时的平衡长度。留意 $\omega^2 = g/(\text{该长度})$，即熟悉的圆锥摆临界条件 $\omega^2 > g/L$。
- 当 $\lambda \to \infty$（不可伸长的绳）时，结果变成 $\cos\theta = g/(l\omega^2)$，即标准的圆锥摆，而范围则变成 $\omega^2 > g/l$。
- 数值上，取 $g = 9.8$ 并随机选取 m、l、λ 和 ω，直接用二分法（bisection）解虎克定律和水平方向的方程，所得的 $\cos\theta$ 与公式相同，而且两条分解所得的方程都成立；在 20 000 个随机个案中，运动存在当且仅当不等式成立。

常见错误

- 把半径取为 $l\sin\theta$ 而不是 $(l+x)\sin\theta$：绳是被拉长了的。
- 忘记说明为何可以除以 $\sin\theta$，因而看不出为甚么要排除 $\theta = 0$（从而排除 $\cos\theta = 1$）。
- 只证明了 $\cos\theta < 1$，却忘记 $\cos\theta > 0$（等价于伸长量为正）才给出上界。

STEP 之外

这个上界是失控（runaway）或不稳定性（instability）的一个简单例子：当 $\omega^2 \to \lambda/(ml)$ 时，所需的伸长量会趋向无穷大。同样的效应也出现在旋转的弹性杆以及自旋物体的稳定性之中；而下界则是一个分歧（bifurcation），悬挂的平衡状态在那里让位予圆周运动。想进一步了解旋转参考系中的力学，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）。

详细解答

STEP I 1994，第 11 题

力学 · 斜面上的连接质点

STEP I 1994，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：

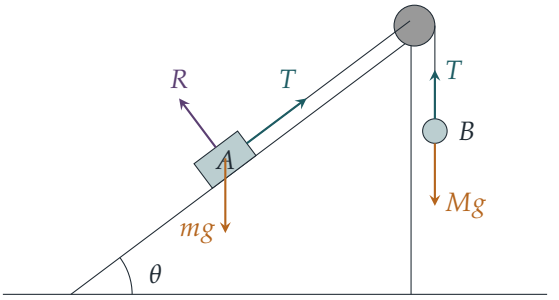
step.maths.org/questions/94-s1-q11

关键思路

先从几何入手。绳子的长度恰好等于轨道的长度，所以当货车位于底部时，小球正好悬在滑轮处。因此小球下落的距离就是斜面的高度 $d\sin\theta$，而货车亦沿斜面向上移动相同的距离。绳子拉紧期间，整个系统以同一个恒定加速度运动。小球一着地，绳子便会松弛，货车只是沿着光滑斜面向上滑行。所以第 (ii) 部分要问的是：货车余下的速率能否把它带到滑轮处。

解答

我们把货车和小球视为质点（particle），轨道视为光滑（smooth），绳子视为轻而不可伸长（light and inextensible），并跨过一个轻而无摩擦的滑轮（pulley）。因此两边的张力（tension）T 相同，而且在绳子拉紧期间，货车和小球的速率相同。



第 (i) 部分

设 a 为小球向下的加速度，这也是货车沿斜面向上的加速度。对小球（沿铅垂方向）和对货车（沿斜面方向）分别应用牛顿第二定律（Newton's second law），得

$$Mg - T = Ma,$$
$$T - mg\sin\theta = ma.$$

法向反作用力（normal reaction）R 与斜面垂直，并不出现在方程之中。把两式相加便可消去 T：

$$a = \frac{(M - m\sin\theta)\,g}{M + m}.$$

系统由静止开始，所以只有当 $a > 0$ 时小球才会移动。（若 $a \le 0$，则甚么也不会发生：当 $a = 0$ 时系统保持平衡；而绳子不能推物，所以 $a < 0$ 是不可能的。）因此，小球落到地面，当且仅当

$$M > m\sin\theta.$$

当货车位于轨道底部时，整条长度为 d 的绳子都贴着斜面，所以小球起初与滑轮同高，离地面的高度为 $d\sin\theta$。它由静止开始以恒定加速度 a 下落这段距离，所以由 $v^2 = u^2 + 2as$ 得

$$v^2 = 2\cdot\frac{(M - m\sin\theta)g}{M + m}\cdot d\sin\theta = \frac{2g(M - m\sin\theta)\,d\sin\theta}{M + m},$$

这正是所要求的结果。

第 (ii) 部分

小球着地时，货车已沿轨道向上移动了 $d\sin\theta$，并正以速率 v 运动。此时它与滑轮之间还余下距离

$$d - d\sin\theta = d(1 - \sin\theta)$$

。

> ? **想一想** 小球着地之后，绳子会怎样?

小球停了下来，但货车仍然沿斜面向上、朝着滑轮运动。所以绳子变得松弛，$T = 0$。这时沿斜面方向唯一的力就是重力的分量，所以货车以 $g\sin\theta$ 减速。它再走一段距离 s 后便会静止，这段距离由 $0 = v^2 - 2g\sin\theta\, s$ 给出：

$$s = \frac{v^2}{2g\sin\theta} = \frac{(M - m\sin\theta)\,d}{M + m}.$$

（其后货车会滑回斜面下方，滑轮附近不会再有任何事情发生。）

如果货车在到达滑轮之前停下，即 $s < d(1-\sin\theta)$，它便不会与滑轮相撞。由于 $d > 0$ 及 $M + m > 0$，这个条件等价于

$$\begin{aligned} M - m\sin\theta &< (M+m)(1-\sin\theta) \\ &= M + m - M\sin\theta - m\sin\theta, \end{aligned}$$

化简后得 $M\sin\theta < m$。

把这个结果与第 (i) 部分结合起来，小球到达地面而货车又在滑轮之前停下，恰好是在以下情况：

$$m\sin\theta < M < \frac{m}{\sin\theta}.$$

若 $M\sin\theta = m$，货车刚好以零速率到达滑轮。那只是轻轻碰到，不算相撞。

验算

- 货车沿轨道向上所能到达的最远距离是 $d\sin\theta + s = \dfrac{Md(1+\sin\theta)}{M+m}$。这个距离小于 d，恰好是在 $M\sin\theta < m$ 的时候，与上面的条件吻合。当 $m \to 0$ 时，它趋向 $d(1+\sin\theta) > d$：一辆没有重量的货车必定会撞上滑轮，正如你所预料的一样。
- 当 $\theta = 90^\circ$ 时，条件 $M < m$ 与第 (i) 部分的 $M > m\sin\theta = m$ 互相矛盾。这是对的：在铅垂的轨道上，货车从滑轮正下方的底部出发，小球从顶部出发，而货车上升的距离恰好等于小球下落的距离，也就是一直升到滑轮处。
- 以数值计算，取 $m = 2$、$M = 1.5$、$\theta = 30^\circ$、$d = 3$ 及 $g = 9.8$，公式给出 $v^2 = 4.2$，而最远点为 $\frac{27}{14} \approx 1.93 < 3$。以细小的时间步长逐步模拟这个运动，得到相同的数值。随机抽取的例子亦证实，相撞恰好在 $M\sin\theta > m$ 时发生。

常见错误

- 把小球下落的距离当作 d 而不是 $d\sin\theta$。小球从滑轮处出发，而滑轮的高度是 $d\sin\theta$。
- 假设小球着地后绳子仍然拉紧，或者完全忘记了滑行阶段，只把 $d\sin\theta$ 与 d 作比较。
- 绳子松弛之后，计算货车的减速度时用了 g 而不是 $g\sin\theta$。

STEP 之外

这道题把运动分成约束条件不同的几个阶段：先是绳子拉紧，然后是绳子松弛。这是*单侧约束*（*unilateral constraint*）最简单的例子：这种约束只能拉，不能推。这类约束在力学中随处可见，从滑落桌面的链条，到机械人学中的接触问题都有。这里用能量方法可以得到

一个快捷的另解：在滑行阶段，货车的动能（kinetic energy）$\frac{1}{2}mv^2$ 转化为势能（potential energy）$mgs\sin\theta$。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对这两种观点都有细致的阐述。

详细解答

STEP I 1994，第 13 题

概率与统计 · 指示变量

> **STEP I 1994，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/94-s1-q13

关键思路

把一个计数写成若干指示变量（indicator variable，每个取值 1 或 0）之和，可以把难算的期望值（expectation）变成易算的，因为无论各部分是否独立（independent），期望值都是线性的。这里各次抽取肯定不独立，但对称性告诉我们每个 X_i 的期望值都相同。然后我们用第二种方法计算 $\mathrm{E}(X)$，即直接由 X 的分布出发；令两个答案相等，几乎不费工夫就得到一条二项式系数（binomial coefficient）恒等式。

解答

全文设 $1 \le n \le M$，并约定当 $b > a$ 或 $b < 0$ 时 $\binom{a}{b} = 0$。

第 (i) 部分

若第 i 个取出的代币是红色，X_i 为 1，否则为 0。因此在和 $X_1 + X_2 + \cdots + X_n$ 之中，每个取出的红色代币恰好贡献 1，每个非红色代币贡献 0。所以这个和数出的正是取出的 n 个代币中红色代币的数目，即 X。

第 (ii) 部分

由于 X_i 只取 0 和 1 两个值，

$$\mathrm{E}(X_i) = 0 \cdot \mathrm{P}(X_i = 0) + 1 \cdot \mathrm{P}(X_i = 1) = \mathrm{P}(\text{第 } i \text{ 个取出的代币是红色}).$$

? **想一想** 第 i 个代币是红色的机会，难道不取决于之前抽到甚么吗？

就条件概率而言，是的；就无条件概率而言，不是。试想像把代币编号，然后逐一全部抽出：M 个代币的每一种排列次序都同样可能出现。因此，位置 i 上的代币同样可能是 M 个代币中的任何一个，正如位置 1 上的代币一样。抽了 n 次便停下来，并不会改变位置 $i \le n$ 上发生的事。由于 M 个代币中有 m 个是红色，

$$\mathrm{E}(X_i) = \mathrm{P}(X_i = 1) = \frac{m}{M} \qquad \text{对每个} i = 1, \ldots, n.$$

（对于 $i = 2$，你可以用较长的方法验证：$\frac{m}{M} \cdot \frac{m-1}{M-1} + \frac{M-m}{M} \cdot \frac{m}{M-1} = \frac{m(M-1)}{M(M-1)} = \frac{m}{M}$。）

第 (iii) 部分

由第 (i) 部分及期望值的线性性质（无需独立性），

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n) = n \cdot \frac{m}{M} = \frac{mn}{M},$$

正是所求。

第 (iv) 部分

取出的 n 个代币所成的集合，同样可能是 $\binom{M}{n}$ 个大小为 n 的子集中的任何一个。要恰好有 k 个红色代币，我们从 m 个红色代币中选 k 个，再从其余 $M-m$ 个中选 $n-k$ 个，所以

$$\mathrm{P}(X = k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}, \qquad k = 0, 1, \ldots, n.$$

按照关于二项式系数的约定，当 $k > m$（红色代币不足）或 $n - k > M - m$（其他代币不足）时，上式正确地等于 0。这就是*超几何分布*（hypergeometric distribution）。

第 (v) 部分

直接由第 (iv) 部分计算 $\mathrm{E}(X)$；$k = 0$ 的一项为零，所以

$$\mathrm{E}(X) = \sum_{k=1}^{n} k\,\mathrm{P}(X = k) = \frac{1}{\binom{M}{n}} \sum_{k=1}^{n} k \binom{m}{k}\binom{M-m}{n-k}.$$

令此式与第 (iii) 部分的结果相等，

$$\sum_{k=1}^{n} k \binom{m}{k}\binom{M-m}{n-k} = \frac{mn}{M}\binom{M}{n}.$$

最后，

$$\frac{n}{M}\binom{M}{n} = \frac{n}{M} \cdot \frac{M!}{n!\,(M-n)!} = \frac{(M-1)!}{(n-1)!\,(M-n)!} = \binom{M-1}{n-1},$$

所以右边等于 $m\binom{M-1}{n-1}$，这就是所求的恒等式。

验算

- 取 $M=5$，$m=2$，$n=3$。第 (iv) 部分给出 $\mathrm{P}(X=0,1,2)=\frac{1}{10},\frac{6}{10},\frac{3}{10}$，三者之和为 1，而 $\mathrm{E}(X)=\frac{6+6}{10}=\frac{6}{5}=\frac{mn}{M}$。恒等式则为 $1\cdot 2\cdot 3+2\cdot 1\cdot 3=12=2\binom{4}{2}$。
- 若 $n=M$（全部取出），则必有 $X=m$，而确实 $\mathrm{E}(X)=mM/M=m$。
- 我们对所有 $M\le 6$ 的袋子，列出代币的每一种排列次序，验证了 $\mathrm{E}(X_i)$、$\mathrm{E}(X)$ 和 $\mathrm{P}(X=k)$；并对每个 $M<30$ 验证了 (v) 中的恒等式。

常见错误

- 声称 $\mathrm{E}(X)=\sum\mathrm{E}(X_i)$ 需要独立性。其实不需要：期望值的线性性质永远成立。
- 把各次抽取当作独立，在 (iv) 中写出二项分布（binomial distribution）$\mathrm{B}(n,\frac{m}{M})$。这样得出的平均值是对的，但概率却是错的，因为抽取是不放回的。

STEP 之外

指示变量法是概率论中最有用的工具之一：它可以求出随机置换（random permutation）的不动点（fixed point）数目的期望值、掷硬币时连串（run）数目的期望值、随机图（random graph）中孤立顶点数目的期望值，而且还可藉 $\mathrm{Var}(\sum X_i)=\sum\mathrm{Var}(X_i)+\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)$ 推广至方差（variance）。以两种方法计算同一个量来证明组合恒等式，这种手法同样普遍适用。Blitzstein 与 Hwang 所著的 *Introduction to Probability*（概率论入门）全书贯彻使用指示变量，是观摩这两种想法实际运用的绝佳去处。

详细解答

STEP I 1994，第 14 题

概率与统计 · 指数分布的顺序统计量

STEP I 1994，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/94-s1-q14

关键思路

每个 T_i 都服从指数分布（exponential distribution），所以第 (i) 部分只是两个标准的分部积分（integration by parts）。至于最后一篇和第一篇文章，窍门在于把关于最大值或最小值的陈述，转化为关于每一个 T_i 的陈述：最后一篇文章在时刻 u 或之前交到，当且仅当所有文章都在那时或之前交到；第一篇文章在时刻 t 仍未交到，当且仅当所有文章在那时都仍未交到。独立性（independence）随即把“所有”变成乘积，再对分布函数（distribution function）求导，便得到密度函数（density）。

解答

第 (i) 部分

以 X 表示任意一个 T_i，其密度函数为 $f(t) = \lambda \mathrm{e}^{-\lambda t}$（$t \geq 0$）。用分部积分，取 $u = t$ 及 $\mathrm{d}v = \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t$（故 $v = -\mathrm{e}^{-\lambda t}$），

$$\mathrm{E}(X) = \int_0^\infty t\,\lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \Big[-t\mathrm{e}^{-\lambda t}\Big]_0^\infty + \int_0^\infty \mathrm{e}^{-\lambda t}\,\mathrm{d}t = 0 + \frac{1}{\lambda} = \frac{1}{\lambda},$$

因为当 $t \to \infty$ 时 $t\mathrm{e}^{-\lambda t} \to 0$（指数函数的增长胜过幂函数）。同样地，

$$\mathrm{E}(X^2) = \int_0^\infty t^2 \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \Big[-t^2\mathrm{e}^{-\lambda t}\Big]_0^\infty + \int_0^\infty 2t\,\mathrm{e}^{-\lambda t}\,\mathrm{d}t = \frac{2}{\lambda}\,\mathrm{E}(X) = \frac{2}{\lambda^2},$$

其中我们认出 $\int_0^\infty t\,\mathrm{e}^{-\lambda t}\,\mathrm{d}t$ 就是 $\frac{1}{\lambda}\mathrm{E}(X)$。因此

$$\mathrm{E}(T_i) = \frac{1}{\lambda}, \qquad \mathrm{Var}(T_i) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

第 (ii) 部分

首先我们需要单独一个 T_i 的分布函数：对于 $u \geq 0$,

$$\mathrm{P}(T_i \leq u) = \int_0^u \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = 1 - \mathrm{e}^{-\lambda u}.$$

现在 $U = \max(T_1, \ldots, T_n)$，而各个 T_i 中最大的一个不超过 u，当且仅当每一个都不超过 u：

$$\{U \leq u\} = \{T_1 \leq u\} \cap \{T_2 \leq u\} \cap \cdots \cap \{T_n \leq u\}.$$

? **想一想** 为甚么这里可以把各个概率相乘?

事件 $\{T_i \leq u\}$ 涉及不同的 T_i，而按假设它们互相独立，所以交集的概率等于各概率的乘积。因此，对于 $u \geq 0$,

$$\mathrm{P}(U \leq u) = \prod_{i=1}^{n} \mathrm{P}(T_i \leq u) = \left(1 - \mathrm{e}^{-\lambda u}\right)^n,$$

正是题目所求。U 的密度函数是这个分布函数的导数；由链式法则（chain rule），

$$f_U(u) = n\left(1 - \mathrm{e}^{-\lambda u}\right)^{n-1} \cdot \lambda \mathrm{e}^{-\lambda u} = n\lambda\, \mathrm{e}^{-\lambda u}\left(1 - \mathrm{e}^{-\lambda u}\right)^{n-1} \qquad (u \geq 0),$$

而当 $u < 0$ 时 $f_U(u) = 0$。

第 (iii) 部分

现在 $T = \min(T_1, \ldots, T_n)$。这里有用的是补事件（complement）：第一篇文章在时刻 t 之后才交到，当且仅当到那时没有一篇交到，即每一个 $T_i > t$。利用 $\mathrm{P}(T_i > t) = \mathrm{e}^{-\lambda t}$，并再次运用独立性，

$$\mathrm{P}(T > t) = \prod_{i=1}^{n} \mathrm{P}(T_i > t) = \left(\mathrm{e}^{-\lambda t}\right)^n = \mathrm{e}^{-n\lambda t} \qquad (t \geq 0).$$

所以 $\mathrm{P}(T \leq t) = 1 - \mathrm{e}^{-n\lambda t}$，求导得

$$f_T(t) = n\lambda\, \mathrm{e}^{-n\lambda t} \qquad (t \geq 0),$$

而当 $t < 0$ 时 $f_T(t) = 0$。

第 (iv) 部分

第 (iii) 部分的密度函数，正是第 (i) 部分的密度函数把 λ 换成 $n\lambda$：T 服从参数为 $n\lambda$ 的指数分布。所以我们可以由第 (i) 部分直接写出

$$\mathrm{E}(T) = \frac{1}{n\lambda}, \qquad \mathrm{Var}(T) = \frac{1}{n^2\lambda^2}.$$

这合乎情理：有 n 名学生时，第一篇文章交到的时间，应该比任何指定一篇快大约 n 倍。

验算

- 当 $n = 1$ 时只有一篇文章，所以 $T = U = T_1$；两个密度函数的确都简化为 $\lambda e^{-\lambda t}$。
- $P(U \le u)$ 由 $u = 0$ 时的 0 上升至 $u \to \infty$ 时的 1，符合分布函数应有的性质。
- 我们以 $n = 5$、$\lambda = 0.7$ 模拟了 2×10^5 个班别。例如，$U \le 2$ 所占的比例为 0.2425，对比 $(1 - e^{-1.4})^5 \approx 0.2427$；$T$ 的平均值为 0.2860，对比 $1/(5 \times 0.7) \approx 0.2857$。用辛普森法则（Simpson's rule）亦验证了 f_U 的积分为 1。

常见错误

- 处理最小值时直接计算 $P(T \le t)$：“至少有一个 $T_i \le t$”并不是交集，所以不能分解成乘积。应该经由 $P(T > t)$ 来计算。
- 忘记说明乘积之所以成立，是因为独立性。
- 写成 $\operatorname{Var}(T_i) = E(T_i^2) = 2/\lambda^2$，忘记减去 $(E\,T_i)^2$。

STEP 之外

T 和 U 分别是样本中最小和最大的*顺序统计量*（order statistics）。独立指数随机变量的最小值仍服从指数分布，而其率参数是各率参数之和，这是泊松过程（Poisson process）和排队论（queueing theory）的基础。由无记忆性（memoryless property）还可得出一个漂亮的结果：相继两篇文章之间的时间间隔是互相独立的指数随机变量，率参数依次为 $n\lambda, (n-1)\lambda, \ldots, \lambda$，所以 $E(U) = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$。这两个结果在 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门），以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率导论）中都有详细讨论。

详细解答

STEP II 1994，第 1 题

纯数学 · 整除性与十进制数字

> **STEP II 1994，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/94-s2-q1

关键思路

乘积的最后几位数字,只取决于各因数的最后几位数字,所以本题的一切运算都是模(modulo）10 的某个幂的算术。末位为 1、3、7 或 9 的数与 10 没有公因数，因此它在模 10 之下是可逆（invertible）的，事实上在模 10 的任何幂之下也是可逆的。中间部分把 2 和 5 这些因数从 n 中抽出来，再把它们变成零。最后部分则利用模 10^{10} 之下的可逆性，迫使乘积的最后十位数字成为我们想要的任何一组数字。

解答

第 1 部分：末位为 1 的倍数

$n \times m$ 的末位数字，就是（n 的末位数字）×（m 的末位数字）的末位数字，因为其余各位数字只会贡献 10 的倍数。由于 n 的末位既不是 5 也不是偶数，所以它的末位是 1、3、7 或 9，而

$$1 \times 1 = 1, \qquad 3 \times 7 = 21, \qquad 7 \times 3 = 21, \qquad 9 \times 9 = 81.$$

因此分别取 $m = 1$、7、3 或 9 即可。这样 $n \times m$ 的末位便是 1，正如所求。

第 2 部分：任意的 n

写成

$$n = 2^a 5^b k,$$

其中 $a, b \geq 0$，而 k 既不能被 2 整除，也不能被 5 整除。于是 k 的末位既不是 5 也不是偶数，所以由第 1 部分可知，存在 m' 使得 km' 的末位是 1。现在设

$$m = 2^b 5^a m'.$$

那么

$$n \times m = 2^{a+b} 5^{a+b} km' = 10^{a+b} (km').$$

乘以 10^{a+b} 只不过是在 km' 之后添上 $a+b$ 个零。所以，若 $a+b=0$，则 $n \times m$ 的末位是 1；否则它的结尾是 1 后面跟着 $a+b$ 个零。这正是我们所需要的。

例如，$n = 40 = 2^3 \cdot 5$ 对应 $k=1$，所以 $m = 2 \cdot 5^3 = 250$，而 $n \times m = 10\,000$。

第 3 部分：齐集十个数字

写成 $n = 10^r q$，其中 $r \geq 0$，而 q 的末位是 1。在末尾添零不会令任何数字消失，所以只需找出 m，使 $q \times m$ 包含每一个数字；这样 $n \times m = 10^r (qm)$ 也会包含它们。

设 $N = 10^{10}$ 及 $X = 1234567890$，后者包含每一个数字。我们的目标是

$$qm \equiv X \pmod{N},$$

也就是说，qm 的最后十位数字是 1234567890。

> **想一想** 为甚么 q 必定有某个倍数，除以 10^{10} 时余数（remainder）是 X?

考虑 N 个数 $q \times 0$, $q \times 1$, …, $q \times (N-1)$，以及它们模 N 的余数。假设其中两个余数相等，譬如 $qs \equiv qt \pmod{N}$，其中 $0 \leq t < s \leq N-1$。那么 10^{10} 整除 $q(s-t)$。但 q 的末位是 1，所以它是奇数，也不是 5 的倍数，因而与 $10^{10} = 2^{10}5^{10}$ 没有公共的质因数。由此可知 $2^{10}5^{10}$ 整除 $s-t$，但这是不可能的，因为 $0 < s-t < N$。所以这 N 个余数各不相同。可能出现的余数只有 N 个，即 $0, 1, \ldots, N-1$，所以每一个都会出现。特别地，存在某个满足 $0 \leq m \leq N-1$ 的 m，使得 $qm \equiv X \pmod{N}$。

这个 m 不是 0，因为 $X \neq 0$。所以 qm 是一个正的倍数，且 $qm \geq X \geq 10^9$。因此它的最后十位数字正好是 $1, 2, 3, 4, 5, 6, 7, 8, 9, 0$，而 X 开首的 1 也确实存在。于是 $q \times m$ 包含 $0, 1, \ldots, 9$ 所有数字，$n \times m$ 亦然。

举例来说，当 $n = 21$ 时，我们可以取 $m = 7\,201\,646\,090$，这时 $21 \times m = 151\,234\,567\,890$。

> **想一想** 关于 n 的假设在哪里用过？为甚么需要它?

我们用它只是为了知道 q 没有因数 2 或 5。没有这一点，余数 $qm \bmod 10^{10}$ 就不能取遍所有的值。例如，2 的每个倍数都是偶数，所以结尾不可能是 …1。正是这个假设让我们可以指定乘积的结尾。（第 2 部分证明了任意的 n 都有一个这种形式的倍数，所以与第 3 部分结合起来，每一个 n 都有一个包含全部十个数字的倍数。）

验算

- 用手验算第 1 部分：$13 \times 7 = 91$ 及 $29 \times 9 = 261$。
- 电脑验算证实，第 1 部分的方法对每一个末位既不是 5 也不是偶数的 $n \le 20\,000$ 都成立。它亦证实第 2 部分的构造对每一个 $n \le 20\,000$，以及 20 000 个不超过 10^{15} 的随机 n 都成立。它还为超过 10 000 个符合所需形式的 n 值计算了第 3 部分的乘数，当中包括末尾有好几个零的数，而在每一个情况下 $n \times m$ 都包含全部十个数字。例子：$21 \times 7\,201\,646\,090 = 151\,234\,567\,890$ 及 $11 \times 4\,657\,687\,990 = 51\,234\,567\,890$。

常见错误

- 在第 3 部分，忘记检查 1234567890 这组数字是否真的完整出现。开首的 1 需要 $qm \ge 10^9$，而这是成立的，因为 $qm \equiv X$ 且 $qm > 0$。
- 在第 2 部分，乘以 10^{a+b} 而不是 $2^b 5^a$。这样做虽然添了零，但 2 和 5 这些原本已在 n 之中的因数却原封不动。

STEP 之外

第 3 部分的鸽巢原理（pigeonhole principle）论证，其实证明了一个一般的事实：若 $\gcd(q, N) = 1$，则 q 在模 N 之下有乘法逆元（multiplicative inverse）。可逆的剩余（residue）构成一个群（group），记作 $(\mathbb{Z}/N\mathbb{Z})^\times$，这是初等数论和 RSA 密码学的起点。另一个技巧可以证明，最后部分对任何 n 都成立：只要 $10^k \ge n$，就有 n 的某个倍数介乎 $X \cdot 10^k$ 与 $X \cdot 10^k + 10^k$ 之间，而这样的倍数是以 1234567890 开首的。Davenport 的 *The Higher Arithmetic* 是一本介绍同余（congruence）与剩余的优秀入门书。

详细解答

STEP II 1994，第 2 题

纯数学 · 勒让德多项式

STEP II 1994，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/94-s2-q2

关键思路

每次求导，因式 $(x-a)^m$ 都会保留下来，只是次数减一，所以 m 重的重根（repeated root）同时也是首 $m-1$ 阶导数（derivative）的根。多项式 $(x^2-1)^n = (x-1)^n(x+1)^n$ 在 $[-1,1]$ 的两个端点都有 n 重根，因此对 $x^m \mathrm{P}_n(x)$ 作分部积分（integration by parts）时，每一个边界项都会消失。每作一次分部积分，就有一阶导数从 $(x^2-1)^n$ 转移到 x^m 上，而 x^m 会比 $(x^2-1)^n$ 更早把导数用尽。

解答

重根与导数

由积法则（product rule），

$$\begin{aligned}\mathrm{P}'(x) &= m(x-a)^{m-1}\mathrm{Q}(x) + (x-a)^m\mathrm{Q}'(x)\\ &= (x-a)^{m-1}\big[m\mathrm{Q}(x) + (x-a)\mathrm{Q}'(x)\big] = (x-a)^{m-1}\mathrm{R}(x),\end{aligned}$$

其中 $\mathrm{R}(x) = m\mathrm{Q}(x) + (x-a)\mathrm{Q}'(x)$ 是多项式，因为多项式的和与积仍是多项式。这正是所要求的形式。

这个结果说明：若多项式含有因式 $(x-a)^k$（$k \geq 1$），则它的导数含有因式 $(x-a)^{k-1}$。由 $k = m$ 开始反复应用这个结果，可得

$$\mathrm{P}^{(r)}(x) = (x-a)^{m-r}\mathrm{R}_r(x) \qquad (0 \leq r \leq m)$$

其中 R_r 为某些多项式。每一步都是容许的，因为我们只在 $r \leq m-1$ 时才对 $\mathrm{P}^{(r)}$ 求导以得出 $\mathrm{P}^{(r+1)}$，而此时指数 $m-r$ 至少为 1。若 $1 \leq r \leq m-1$，则 $m-r \geq 1$，所以 $(x-a)$ 仍是 $\mathrm{P}^{(r)}(x)$ 的因式，因此 $\mathrm{P}^{(r)}(a) = 0$。（这对 $r = 0$ 同样成立，但对 $r = m$ 一般不成立。）

P_n 是 n 次多项式

设 $F(x) = (x^2-1)^n$。由二项式定理（binomial theorem），$F(x)$ 等于 x^{2n} 加上次数不超过 $2n-2$ 的各项。把 x^k 求导 n 次，结果或是 0（若 $k < n$），或是 x^{k-n} 的倍数，所以 $P_n = F^{(n)}$ 是次数不超过 n 的多项式。它的 x^n 项只来自 x^{2n}：

$$\frac{d^n}{dx^n} x^{2n} = 2n(2n-1)\cdots(n+1)\,x^n = \frac{(2n)!}{n!}\,x^n.$$

这个系数不等于零，所以 P_n 的次数恰好是 n。

等于零的积分

由于 $F(x) = (x-1)^n(x+1)^n$，在第一部分取 $a = 1$、$m = n$ 及 $Q(x) = (x+1)^n$，可知当 $1 \le r \le n-1$ 时 $F^{(r)}(1) = 0$。取 $a = -1$ 作同样的论证，可得 $F^{(r)}(-1) = 0$。另外，直接可见 $F(\pm 1) = 0$。所以

$$F^{(r)}(\pm 1) = 0 \qquad \text{对于} 0 \le r \le n-1. \tag{$*$}$$

现在取 $j \ge 1$ 及 $1 \le k \le n$。作分部积分，对 x^j 求导、对 $F^{(k)}$ 积分，可得

$$\begin{aligned}\int_{-1}^{1} x^j F^{(k)}(x)\,dx &= \Big[x^j F^{(k-1)}(x)\Big]_{-1}^{1} - j\int_{-1}^{1} x^{j-1}F^{(k-1)}(x)\,dx \\ &= -j\int_{-1}^{1} x^{j-1}F^{(k-1)}(x)\,dx,\end{aligned}$$

因为 $k-1 \le n-1$，所以由 $(*)$ 可知边界项等于零。

> **想一想** 这个步骤可以用多少次，才会出问题?

每用一次，x 的幂次减一，导数的阶数也减一。我们由 x^m 和 $F^{(n)}$ 开始，所以可以应用 m 次。这个步骤需要 $k \ge 1$，而我们用到的最小的 k 是 $n-m+1 \ge 1$。经过 m 步之后，

$$\int_{-1}^{1} x^m P_n(x)\,dx = (-1)^m m! \int_{-1}^{1} F^{(n-m)}(x)\,dx.$$

（当 $m = 0$ 时，这就是原来的积分。）

若 $0 \le m \le n-1$，则 $n-m \ge 1$，所以最后的被积函数（integrand）本身是一个导数，于是

$$\int_{-1}^{1} x^m P_n(x)\,dx = (-1)^m m!\Big[F^{(n-m-1)}(x)\Big]_{-1}^{1} = 0,$$

这是由 $(*)$ 得出的，因为 $0 \le n-m-1 \le n-1$。这就是所要求的结果。

$m = n$ 的情况

取 $m = n$ 时，同样的化简过程仍然有效（最后一步用到 $k = 1$），并给出

$$\int_{-1}^{1} x^n \mathrm{P}_n(x)\,\mathrm{d}x = (-1)^n n! \int_{-1}^{1} (x^2-1)^n\,\mathrm{d}x = n! \int_{-1}^{1} (1-x^2)^n\,\mathrm{d}x,$$

因为 $(x^2-1)^n = (-1)^n(1-x^2)^n$。被积函数是偶函数（even function），而代换 $x = \sin t$、$\mathrm{d}x = \cos t\,\mathrm{d}t$ 把 $[0,1]$ 变为 $[0, \frac{\pi}{2}]$，且 $1 - x^2 = \cos^2 t$。所以，利用提示中的公式，

$$\int_{-1}^{1} (1-x^2)^n\,\mathrm{d}x = 2\int_{0}^{\pi/2} \cos^{2n+1} t\,\mathrm{d}t = \frac{2^{2n+1}(n!)^2}{(2n+1)!}.$$

因此

$$\boxed{\int_{-1}^{1} x^n \mathrm{P}_n(x)\,\mathrm{d}x = \frac{2^{2n+1}(n!)^3}{(2n+1)!}.}$$

验算

当 $n = 1$ 时，$\mathrm{P}_1(x) = 2x$，所以 $\int_{-1}^{1} x \cdot 2x\,\mathrm{d}x = \frac{4}{3}$，而公式给出 $\frac{8\cdot 1}{6} = \frac{4}{3}$。当 $n = 2$ 时，$\mathrm{P}_2(x) = 12x^2 - 4$。于是 $\int_{-1}^{1}(12x^2-4)\,\mathrm{d}x = 8 - 8 = 0$，且 $\int_{-1}^{1} x(12x^2-4)\,\mathrm{d}x = 0$（被积函数是奇函数），与预期相符。另外 $\int_{-1}^{1} x^2(12x^2-4)\,\mathrm{d}x = \frac{24}{5} - \frac{8}{3} = \frac{32}{15}$，与 $\frac{2^5\cdot 8}{120} = \frac{32}{15}$ 吻合。我们已用电脑对 $n \le 10$ 的所有情况作了精确验证。

常见错误

- 忘记边界项除了需要 $1 \le r \le n-1$ 时 $\mathrm{F}^{(r)}(\pm 1) = 0$ 之外，还需要 $r = 0$ 时也成立。$m = n - 1$ 的最后一步正是用到 $r = 0$。
- 弄错正负号：$(x^2-1)^n = (-1)^n(1-x^2)^n$ 抵销了分部积分所产生的 $(-1)^n$，所以答案是正数。这是合理的，因为 $x^n\mathrm{P}_n(x)$ 的首项系数是正数。

STEP 之外

除了相差一个常数倍之外，P_n 就是 n 次勒让德多项式（Legendre polynomial），而它的这个公式称为罗德里格斯公式（Rodrigues' formula）。第一个结果说明，就内积（inner product）$\int_{-1}^{1} fg\,\mathrm{d}x$ 而言，P_n 与每一个次数较低的多项式都正交（orthogonal）。勒让德级数（Legendre series）和高斯求积法（Gaussian quadrature）背后所依靠的正是这个性质。在球坐标（spherical coordinates）下求解拉普拉斯方程（Laplace's equation）时，勒让德多项式也会出现。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）通过勒让德微分方程（Legendre's differential equation）引入这些多项式。

详细解答

STEP II 1994，第 3 题

纯数学 · 函数方程

STEP II 1994，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/94-s2-q3

关键思路

关于 f，我们只知道一条对*所有* x 和 y 都成立的恒等式（identity），以及一个函数值 $\mathrm{f}(0)=1$，除此之外一无所知。因此唯一的工具就是代入（substitution）：选取适当的 x 和 y，使大部分自变量变成 0 或 a，因为函数在这些点的值我们已经知道（或即将知道）。这条恒等式应该令你联想到 $\cos(x-y)=\cos x\cos y+\sin x\sin y$，其中 $\mathrm{f}(a-x)$ 扮演 $\pm\sin x$ 的角色；这个类比既告诉你可以预期甚么结果，也告诉你例子会是甚么。

解答

以下一律把题目给出的恒等式写成

$$\mathrm{f}(x-y)=\mathrm{f}(x)\,\mathrm{f}(y)-\mathrm{f}(a-x)\,\mathrm{f}(a+y)\qquad \text{对所有}x,y.\qquad (*)$$

下面每一步都只是代入 $(*)$，所以我们没有假设连续性（continuity）、可微性（differentiability）或任何其他性质。

证明 $\mathrm{f}(a)=0$

在 $(*)$ 中取 $x=y=0$：

$$\mathrm{f}(0)=\mathrm{f}(0)^2-\mathrm{f}(a)\,\mathrm{f}(a),\qquad \text{所以}\qquad 1=1-\mathrm{f}(a)^2.$$

因此 $\mathrm{f}(a)^2=0$，而 $\mathrm{f}(a)$ 是实数，故 $\mathrm{f}(a)=0$。

第 (i) 部分

在 $(*)$ 中取 $x = 0$ 和 $y = t$：

$$\mathrm{f}(-t) = \mathrm{f}(0)\,\mathrm{f}(t) - \mathrm{f}(a)\,\mathrm{f}(a+t) = 1 \cdot \mathrm{f}(t) - 0 = \mathrm{f}(t),$$

这里用到 $\mathrm{f}(0) = 1$ 和 $\mathrm{f}(a) = 0$。所以 f 是偶函数（even function）。

第 (ii) 部分

在 $(*)$ 中取 $x = y = a$：

$$\mathrm{f}(0) = \mathrm{f}(a)^2 - \mathrm{f}(0)\,\mathrm{f}(2a), \qquad \text{所以} \qquad 1 = 0 - \mathrm{f}(2a),$$

由此得 $\mathrm{f}(2a) = -1$。

第 (iii) 部分

> **想一想** 哪一个代入可以令右边第一个乘积消失，同时让 t 保持自由？

取 $x = a$ 可以消去 $\mathrm{f}(x)\mathrm{f}(y)$ 这一项（因为 $\mathrm{f}(a) = 0$），并把 $\mathrm{f}(a-x)$ 变成 $\mathrm{f}(0) = 1$，而 y 仍然可以任意选取。当 $x = a$ 时，$(*)$ 变成

$$\mathrm{f}(a-y) = 0 \cdot \mathrm{f}(y) - \mathrm{f}(0)\,\mathrm{f}(a+y) = -\mathrm{f}(a+y) \qquad \text{对所有} y.$$

所以 f “关于 a 是奇的”。现在取 $y = a - t$：这时 $a - y = t$ 而 $a + y = 2a - t$，所以

$$\mathrm{f}(t) = -\mathrm{f}(2a - t),$$

即对所有 t 都有 $\mathrm{f}(2a - t) = -\mathrm{f}(t)$，这正是所要证明的结果。（取 $t = 0$ 便重新得到第 (ii) 部分的结果，这是一个有用的一致性检查。）

第 (iv) 部分

把 (iii) 用两次，中间用一次 (i)。首先，在 (iii) 中把 t 换成 $-t$，得

$$\mathrm{f}(2a + t) = -\mathrm{f}(-t) = -\mathrm{f}(t), \tag{1}$$

这里用到 (i)。所以把自变量平移 $2a$ 会令函数值变号。再平移 $2a$ 一次，即在 (1) 中把 t 换成 $2a + t$，

$$\mathrm{f}(4a + t) = -\mathrm{f}(2a + t) = -\big(-\mathrm{f}(t)\big) = \mathrm{f}(t).$$

所以 f 是周期函数（periodic function），周期为 $4a$（当 $a \neq 0$ 时），正如余弦函数的周期是 $4 \cdot \frac{\pi}{2} = 2\pi$ 一样。

例子

当 $a = \pi/2$ 时。取 $\mathrm{f}(x) = \cos x$。这时 $\mathrm{f}(0) = 1$，而且由于 $\cos(\frac{\pi}{2} - x) = \sin x$ 和 $\cos(\frac{\pi}{2} + y) = -\sin y$，

$$\mathrm{f}(x)\mathrm{f}(y) - \mathrm{f}(\tfrac{\pi}{2} - x)\mathrm{f}(\tfrac{\pi}{2} + y) = \cos x \cos y + \sin x \sin y = \cos(x - y) = \mathrm{f}(x - y).$$

当 $a = -2$ 时。试用 $\mathrm{f}(x) = \cos kx$。由上面的结果可知，我们需要 $\mathrm{f}(a) = 0$，即 $\cos(2k) = 0$，并且周期为 $4|a| = 8$；所以取 $k = \frac{\pi}{4}$，即 $\mathrm{f}(x) = \cos\frac{\pi x}{4}$。验证如下：$\mathrm{f}(0) = 1$，并且设 $u = \frac{\pi x}{4}$、$v = \frac{\pi y}{4}$，则

$$\mathrm{f}(-2-x) = \cos\big(-\tfrac{\pi}{2} - u\big) = \cos\big(\tfrac{\pi}{2} + u\big) = -\sin u, \qquad \mathrm{f}(-2+y) = \cos\big(v - \tfrac{\pi}{2}\big) = \sin v,$$

所以

$$\begin{aligned}\mathrm{f}(x)\mathrm{f}(y) - \mathrm{f}(-2-x)\mathrm{f}(-2+y) &= \cos u \cos v + \sin u \sin v \\ &= \cos(u - v) = \cos\tfrac{\pi(x-y)}{4} = \mathrm{f}(x - y).\end{aligned}$$

这两个函数显然都不是常数函数。（更一般地说，只要 $\cos ka = 0$，$\cos kx$ 对该 a 值便适用，所以当 $a = -2$ 时 $\cos\frac{3\pi x}{4}$ 同样可以。）

验算

在第 (iii) 部分取 $t = 0$，得 $\mathrm{f}(2a) = -\mathrm{f}(0) = -1$，与 (ii) 吻合；在 (iii) 取 $t = a$，得 $\mathrm{f}(a) = -\mathrm{f}(a)$，与 $\mathrm{f}(a) = 0$ 吻合。再以数值检验：对 $\mathrm{f}(x) = \cos\frac{\pi x}{4}$ 和 $a = -2$，取 $x = 1$、$y = 3$，左边是 $\mathrm{f}(-2) = 0$，右边是 $\mathrm{f}(1)\mathrm{f}(3) - \mathrm{f}(-3)\mathrm{f}(1) = \cos\frac{\pi}{4}\cos\frac{3\pi}{4} - \cos\frac{3\pi}{4}\cos\frac{\pi}{4} = 0$。用电脑对两个例子各以 20 000 组随机数对 (x, y) 测试，结果恒等式在 10^{-13} 的误差范围内成立。

常见错误

- 弄错平移半个周期时的正负号：由 (iii) 和 (i) 得 $\mathrm{f}(t + 2a) = -\mathrm{f}(t)$，所以周期是 $4a$，而不是 $2a$。
- 在证明 (i)–(iv) 时假设 f 可微（或假设它就是余弦函数）。题目不容许这样做；只可以使用代入。
- 对于 $a = -2$，只写出 $\cos\frac{\pi x}{4}$ 而没有验证它满足恒等式：符合 $\mathrm{f}(a) = 0$ 和周期的要求只是很好的指引，并不是证明。

STEP 之外

由函数所满足的恒等式来确定该函数，正是*函数方程*（functional equations）这门学问的课题。在本题中，单凭一条恒等式，便足以推出偶函数性质、“半周期”变号以及周期性。在分析学（analysis）课程中会遇到一些经典结果：柯西（Cauchy）方程 $\mathrm{f}(x+y)=\mathrm{f}(x)+\mathrm{f}(y)$ 在加上连续性的条件后，必然得出 $\mathrm{f}(x)=cx$；若没有这类条件，便会出现性质极为古怪的解。同样地，余弦的减法公式在连续函数之中刻划了 $\cos kx$。至于连续性如何在这类论证中发挥作用，可参阅 Abbott 的 *Understanding Analysis*（一本讲解严谨的分析学入门书）。

详细解答

STEP II 1994，第 4 题

纯数学 · 椭圆的面积

STEP II 1994，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/94-s2-q4

关键思路

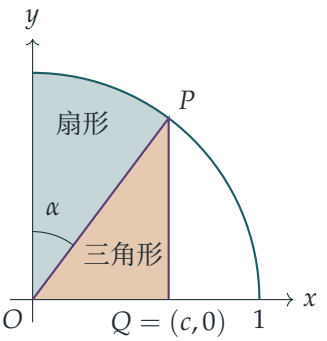
第一个积分是四分之一圆（quarter circle）下方的面积，而这块面积可以干净俐落地分成一个直角三角形和一个扇形（sector），所以毋须作任何代换（substitution）。椭圆（ellipse）是把圆沿两个方向以不同倍数拉伸而成的，而拉伸会把面积乘以一个常数，于是第二个结果立即得出。至于最后一部分，对称性（symmetry）几乎包办了所有工作：两个椭圆互为对直线 $y = x$ 的反射（reflection），所以我们只需求交集（intersection）的八分之一，而这一小块正是第二个结果减去一个三角形。

解答

$(1 - x^2)^{1/2}$ 的积分

所讨论的区域由四分之一圆 $y = (1 - x^2)^{1/2}$、两条坐标轴和直线 $x = c$ 围成。（这条曲线须理解为边界：区域是满足 $0 \le y \le (1 - x^2)^{1/2}$ 和 $0 \le x \le c$ 的点所组成的集合。）它的面积正是我们要求的积分。

设 $P = \left(c, (1 - c^2)^{1/2}\right)$ 为圆上位于 $x = c$ 上方的点，$Q = (c, 0)$，并设 $\alpha = \sin^{-1} c$，于是 $0 < \alpha \le \frac{\pi}{2}$。连接 O 与 P，便把区域分成两块。



- 三角形 OQP 在 Q 处是直角，底为 c，高为 $(1-c^2)^{1/2}$，所以它的面积是 $\frac{1}{2}c(1-c^2)^{1/2}$。
- 余下的部分是单位圆在 OP 与 y 轴之间的扇形。OP 与 y 轴的夹角是 α，因为 P 与 y 轴的水平距离是 $c = 1\cdot\sin\alpha$。半径为 1、角为 α 的扇形，面积是 $\frac{1}{2}\alpha = \frac{1}{2}\sin^{-1}c$。

把两块相加，

$$\int_0^c (1-x^2)^{\frac{1}{2}}\,\mathrm{d}x = \tfrac{1}{2}\left[c(1-c^2)^{\frac{1}{2}} + \sin^{-1}c\right], \qquad 0 < c \le 1,$$

正是所求。（当 $c = 1$ 时，三角形退化成一条线段，我们得到四分之一圆的面积 $\frac{\pi}{4}$。）

椭圆下方的面积

同样地，区域位于曲线上半部分 $y = b\left(1-x^2/a^2\right)^{1/2}$ 之下，其中 $0 \le x \le c$，所以它的面积是

$$\int_0^c b\left(1-\frac{x^2}{a^2}\right)^{\frac{1}{2}}\mathrm{d}x.$$

代入 $x = au$，则 $\mathrm{d}x = a\,\mathrm{d}u$，积分上下限变成 0 和 c/a（由于 $0 < c \le a$，故 $0 < c/a \le 1$），上式变成

$$ab\int_0^{c/a}(1-u^2)^{\frac{1}{2}}\,\mathrm{d}u = \frac{ab}{2}\left[\frac{c}{a}\left(1-\frac{c^2}{a^2}\right)^{\frac{1}{2}} + \sin^{-1}\left(\frac{c}{a}\right)\right],$$

这里用了第一个结果，并把 c 换成 c/a。这就是所求的面积。

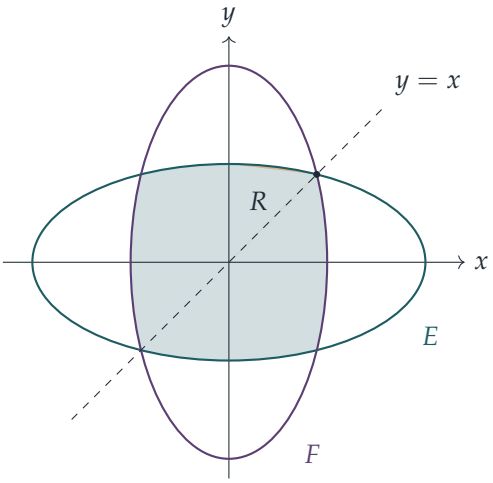
从几何角度看：椭圆是把单位圆沿水平方向拉伸 a 倍、沿垂直方向拉伸 b 倍而成的，而这个拉伸把每一块面积都乘以 ab。

两个椭圆的交集

E 和 F 在 $x \mapsto -x$ 和 $y \mapsto -y$ 之下都保持不变，所以 $E \cap F$ 也是如此，它的面积等于它在象限（quadrant）$x, y \ge 0$ 内那部分面积的四倍。此外，把 x 和 y 对调会把 E 变成 F、把 F 变成 E，所以 $E \cap F$ 对直线 $y = x$ 对称。因此，第一象限内的那四分之一由互为镜像

的两半组成，而

$$\text{面积}(E\cap F)=8\times\text{区域面积}R,\qquad R=\{(x,y)\in E\cap F:\ 0\le x\le y\}.$$



想一想 在 R 之中，两个椭圆里究竟是哪一个构成边界？

对任意一点，

$$\left(\frac{x^2}{b^2}+\frac{y^2}{a^2}\right)-\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)=(x^2-y^2)\left(\frac{1}{b^2}-\frac{1}{a^2}\right).$$

当 $0\le x\le y$ 时，$x^2-y^2\le 0$；又因为 $0<b\le a$，所以 $\frac{1}{b^2}-\frac{1}{a^2}\ge 0$。因此这个差 ≤ 0：在 $y=x$ 的那一侧，E 内的任何一点都自动在 F 内。于是

$$R=\{(x,y)\in E:\ 0\le x\le y\},$$

只有 E 起作用。（这里用到了 $b\le a$；若 $b>a$，两个椭圆的角色便会互换。）

E 的边界在第一象限内与 $y=x$ 相交之处满足 $x^2\left(\frac{1}{a^2}+\frac{1}{b^2}\right)=1$，即交于 $x=y=x_0$，其中

$$x_0=\frac{ab}{\sqrt{a^2+b^2}}.$$

当 $0\le x\le x_0$ 时，E 的上半部分位于 $y=x$ 之上或恰在其上（它从高度 $b>0$ 出发，只在 x_0 处与该直线相交），而当 $x>x_0$ 时则位于该直线之下。所以 R 是 $0\le x\le x_0$ 时介乎 $y=x$ 与曲线 $y=b(1-x^2/a^2)^{1/2}$ 之间的区域：

$$\text{面积}(R)=\underbrace{\int_0^{x_0} b\left(1-\frac{x^2}{a^2}\right)^{\frac12}\mathrm{d}x}_{\text{上一个结果，取}c=x_0}-\underbrace{\frac12 x_0^2}_{\text{三角形，位于此线之下：}y=x}.$$

取 $c=x_0$（并且符合所需的条件 $0<x_0<a$），

$$\frac{c}{a}=\frac{b}{\sqrt{a^2+b^2}},\qquad \left(1-\frac{c^2}{a^2}\right)^{\frac12}=\frac{a}{\sqrt{a^2+b^2}},\qquad \frac{c}{a}\left(1-\frac{c^2}{a^2}\right)^{\frac12}=\frac{ab}{a^2+b^2}.$$

所以

$$\begin{aligned}\text{面积}(E\cap F) &= 8\,\text{面积}(R)\\ &= 4ab\left[\frac{ab}{a^2+b^2}+\sin^{-1}\Big(\frac{b}{\sqrt{a^2+b^2}}\Big)\right]-\frac{4a^2b^2}{a^2+b^2}\\ &= 4ab\sin^{-1}\Big(\frac{b}{\sqrt{a^2+b^2}}\Big),\end{aligned}$$

因为两个代数项恰好互相抵消。这就是所求的面积。

> ? **想一想** 为甚么代数项会互相抵消?

$\frac{ab}{2}\cdot\frac{c}{a}(1-c^2/a^2)^{1/2}=\frac{1}{2}x_0y_0$ 这一项，是圆的图中那个三角形经拉伸后得到的三角形 O、$(x_0,0)$、(x_0,y_0)。这里 $y_0=x_0$，所以它恰好就是我们减去的、位于 $y=x$ 之下的那个三角形。剩下的是一个经拉伸的扇形，这就是答案只含一个反正弦（inverse sine）的原因。

验算

- 若 $a=b$，两个椭圆是同一个圆，公式给出 $4a^2\sin^{-1}\frac{1}{\sqrt{2}}=4a^2\cdot\frac{\pi}{4}=\pi a^2$，与应有的结果相符。
- 当 a 固定而 $b\to 0$ 时，面积的表现近似于 $4ab\cdot\frac{b}{a}=4b^2$：交集几乎就是边长为 $2b$ 的正方形，即 E 那条窄长的水平带与 F 那条窄长的垂直带交叉之处。
- 当 $a=2$、$b=1$ 时，公式给出 $8\sin^{-1}\frac{1}{\sqrt{5}}\approx 3.70918$。对两条边界中较低的一条使用辛普森法则（Simpson's rule），得 3.70918；以 2×10^5 个点作蒙地卡罗（Monte Carlo）估算，得 3.707。头两个结果亦已就多组随机的 a、b、c 以数值求积（quadrature）核对过。

常见错误

- 弄错扇形的角：扇形位于 OP 与 y 轴之间，角是 $\sin^{-1}c$，而不是 $\cos^{-1}c$。
- 没有检查是哪一个椭圆围住该区域。在 $0\le x\le y$ 这一半之中，边界是 E，全因 $b\le a$；若选了另一个椭圆，答案中反正弦里面的 b 便会变成 a。

STEP 之外

线性映射（linear map）$(x,y)\mapsto(ax,by)$ 把每一块面积都乘以 ab；一般而言，线性映射会把面积乘以其行列式（determinant）的绝对值。这正是重积分（multiple integrals）变量代换公式（change-of-variables formula）中雅可比（Jacobian）因子的二维情形，你

会在大学一年级的微积分课程中遇到。Strang 的 *Introduction to Linear Algebra*（线性代数入门）把行列式解释为面积（以及体积）的缩放因子。

详细解答

STEP II 1994，第 5 题

纯数学 · 实根的计数

STEP II 1994，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/94-s2-q5

关键思路

要数 $f(x) = c$ 有多少个根（root），即是问水平线 $y = c$ 与 f 的图像相交多少次，所以关键在于 f 的形状：它在哪里取最小值，在哪里递增、在哪里递减。本题的每个式子都由两个关于某个中心点对称的括号组成。把变量平移，使中心点变成原点，则奇次项互相抵消（四次方的情况）或偶次项互相抵消（三次方的情况），图像的形状便一目了然。

解答

第 (i) 部分

用二项式定理（binomial theorem）展开。在 $(x+1)^4$ 和 $(x-1)^4$ 中，x 的奇次项符号相反，所以相加时互相抵消：

$$f(x) = (x-1)^4 + (x+1)^4 = 2\left(x^4 + 6x^2 + 1\right) = 2x^4 + 12x^2 + 2.$$

因此 f 是偶函数（even function），而且 $f(x) = 2 + 2x^4 + 12x^2 \geq 2$，等号只在 $x = 0$ 时成立。对于 $x \geq 0$，当 $x > 0$ 时 $f'(x) = 8x^3 + 24x > 0$，所以 f 在 $[0, \infty)$ 上严格递增（strictly increasing），而且当 $x \to \infty$ 时 $f(x) \to \infty$。由于 f 是连续（continuous）的，它在 $[0, \infty)$ 上取 $[2, \infty)$ 中的每个值恰好一次；根据对称性，它在 $(-\infty, 0]$ 上亦然。

- 若 $c > 2$：恰有一个根 $x = a > 0$，于是 $x = -a$ 是第二个根；这两个根互不相同，此外再无其他根。所以**恰好有两个**实根。
- 若 $c = 2$：由 $f(x) = 2$ 必得 $2x^4 + 12x^2 = 0$，所以 $x = 0$ 是**唯**一的根。
- 若 $c < 2$：对所有 x 都有 $f(x) \geq 2 > c$，所以**没有**根。

这正是题目所要求的结论。

（另一做法：$2x^4+12x^2+2-c=0$ 是关于 $u=x^2$ 的二次方程，其根为 $u=-3\pm\sqrt{8+c/2}$；只有取 $+$ 号的根才可能非负，而它是正数、零还是负数，分别对应 $c>2$、$c=2$ 和 $c<2$。）

第 (ii) 部分

括号 $x-3$ 和 $x-1$ 以 $x=2$ 为中心，所以设 $y=x-2$。于是

$$(x-3)^4+(x-1)^4=(y-1)^4+(y+1)^4,$$

这正是第 (i) 部分的式子，只是变量换成了 y。由于 $x\mapsto y=x-2$ 是实数之间的一一对应（one-to-one correspondence），这个方程的实根数目与第 (i) 部分相同：

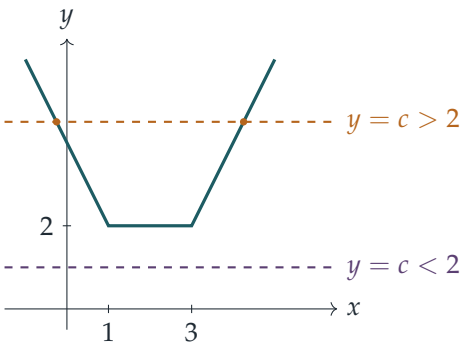
$$\text{有两个根，若}c>2,\qquad \text{有一个根}(x=2)\text{，若}c=2,\qquad \text{没有根，若}c<2.$$

第 (iii) 部分

设 $g(x)=|x-3|+|x-1|$，并在 $x=1$ 和 $x=3$ 所分出的三个区间上分别去掉绝对值（modulus）符号：

$$g(x)=\begin{cases}(3-x)+(1-x)=4-2x, & x<1,\\ (3-x)+(x-1)=2, & 1\le x\le 3,\\ (x-3)+(x-1)=2x-4, & x>3.\end{cases}$$

所以图像是一个平底的 V 形：在 1 与 3 之间恒等于 2，在左边严格递减，在右边严格递增，而且在 $[1,3]$ 之外 $g>2$。



- 若 $c<2$：处处都有 $g(x)\ge 2>c$，所以**没有**根。
- 若 $c=2$：凡满足 $1\le x\le 3$ 的 x 都是根，此范围以外则没有根，所以有**无穷多个**根。

- 若 $c > 2$：在 $[1,3]$ 内没有根；在 $x < 1$ 上，由 $4 - 2x = c$ 得 $x = 2 - \frac{c}{2}$，由于 $c > 2$，它确实小于 1；在 $x > 3$ 上，由 $2x - 4 = c$ 得 $x = 2 + \frac{c}{2} > 3$。所以**恰好有两个**根，即 $x = 2 \pm \frac{c}{2}$。

> **想一想** 为甚么 $c = 2$ 时的答案与第 (ii) 部分相差这么远?

对于四次方，由于两项都是严格凸（strictly convex）的，它们的和在中心点有唯一而尖锐的最小值。对于绝对值，当 x 在 1 与 3 之间移动时，其中一个距离增加多少，另一个距离便恰好减少多少，所以它们的和在整个区间上都保持为 2（即 1 与 3 之间的距离）。

第 (iv) 部分

同样设 $y = x - 2$。对于三次方，互相抵消的是 y 的偶次项：

$$(x-3)^3 + (x-1)^3 = (y-1)^3 + (y+1)^3 = 2y^3 + 6y.$$

设 $h(y) = 2y^3 + 6y$。则对任意实数 y 都有

$$h'(y) = 6y^2 + 6 \geq 6 > 0$$

, 所以 h 在整条实数线上严格递增, 因而每个值最多取一次。此外, 当 $y \to \infty$ 时 $h(y) \to \infty$, 当 $y \to -\infty$ 时 $h(y) \to -\infty$，而 h 是连续的，所以根据介值定理（intermediate value theorem），它取每个实数值至少一次。

因此，对**每一个**实数 c，方程都**恰好有一个**实根。

验算

- 在中心点 $x = 2$，两个四次方的式子都等于 $1 + 1 = 2$，与 $c = 2$ 的情况吻合，而且 $(2-3)^3 + (2-1)^3 = 0$。
- 在第 (iv) 部分, $c = 8$ 应给出 $x = 3$: 的确 $0^3 + 2^3 = 8$, 而取 $y = 1$ 时 $h(1) = 2 + 6 = 8$。
- 在第 (iii) 部分取 $c = 4$，根 $x = 0$ 和 $x = 4$ 分别给出 $3 + 1 = 4$ 和 $1 + 3 = 4$。

我们以数值方法确认了各展开式在随机选取的点上成立，并对一系列 c 值（例如在 (i) 和 (ii) 中取 $c = 1.9$、2.5、100，在 (iv) 中取 c 由 -1000 至 5000），在细密的网格上数算每个函数减去 c 后的变号次数：所得次数分别是 0、2、2，以及恒为 1，与上述结论一致。

常见错误

- 在第 (iii) 部分 $c = 2$ 时答“一个根”或“两个根”：整个区间 $1 \leq x \leq 3$ 上的点全都是根。

- 在第 (iv) 部分，只说三次方程“有一个或三个实根”，却没有判断是哪一种情况。$h'(y) > 0$ 处处成立这一事实便足以定夺。
- 忘记检查 $x = 2 - \frac{c}{2}$ 是否真的落在使用公式 $4 - 2x$ 的区域 $x < 1$ 之内。

STEP 之外

这里的论证建基于两个定理，大学第一门分析（analysis）课程会仔细证明它们：其一是介值定理（区间上的连续函数会取遍两个函数值之间的每一个值），其二是导数为正的函数必定严格递增，这可由中值定理（mean value theorem）推出。第 (iii) 部分亦让你初尝凸性（convexity）和 ℓ^1 距离的滋味：到两点的距离之和，在连接这两点的整条线段上都取得最小值。Abbott 的 *Understanding Analysis* 一书严谨地阐述了这些概念。

详细解答

STEP II 1994，第 6 题

纯数学 · 三角函数裂项相消级数

> **STEP II 1994，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/94-s2-q6

关键思路

一切都建基于一条小小的恒等式（identity）：$\tan x = \cot x - 2\cot 2x$。它把和式中的每一项化为形如 $\frac{1}{2^r}\cot\frac{\theta}{2^r}$ 的相邻两项之差，于是和式可以裂项相消（telescope），而归纳步骤之所以行得通，靠的也正是这条恒等式。至于无穷和，只需用到极限（limit）$\frac{\sin\varphi}{\varphi}\to 1$。

解答

一条有用的恒等式

对任何使 $\sin x$、$\cos x$ 和 $\sin 2x$ 皆不为零的 x，

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

所以

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

我们只会在 $x = \theta/2^r$（其中 $r \geq 1$）的情况下使用 $(*)$。由于 $0 < \theta < \pi$，这样的 x 必在 $(0, \frac{\pi}{2})$ 之内，因此 $\sin x$、$\cos x$ 和 $\sin 2x$ 全为正数，我们写下的每一个 tan 和 cot 都有定义。

以归纳法求有限和

以 S_n 表示左方 $\sum_{r=1}^{n}\frac{1}{2^r}\tan\frac{\theta}{2^r}$，并设 $P(n)$ 为命题

$$S_n = \frac{1}{2^n}\cot\frac{\theta}{2^n} - \cot\theta.$$

基础情况 $n=1$。在 $(*)$ 中代入 $x=\theta/2$，再除以二：

$$S_1=\tfrac{1}{2}\tan\tfrac{\theta}{2}=\tfrac{1}{2}\cot\tfrac{\theta}{2}-\cot\theta,$$

这就是 $P(1)$。

归纳步骤。假设 $P(k)$ 对某个 $k\geq 1$ 成立。在 $(*)$ 中代入 $x=\theta/2^{k+1}$，于是 $2x=\theta/2^k$，再除以 2^{k+1}：

$$\frac{1}{2^{k+1}}\tan\frac{\theta}{2^{k+1}}=\frac{1}{2^{k+1}}\cot\frac{\theta}{2^{k+1}}-\frac{1}{2^k}\cot\frac{\theta}{2^k}.$$

把此式与 $P(k)$ 相加，

$$\begin{aligned}S_{k+1}&=S_k+\frac{1}{2^{k+1}}\tan\frac{\theta}{2^{k+1}}\\&=\left(\frac{1}{2^k}\cot\frac{\theta}{2^k}-\cot\theta\right)+\left(\frac{1}{2^{k+1}}\cot\frac{\theta}{2^{k+1}}-\frac{1}{2^k}\cot\frac{\theta}{2^k}\right)\\&=\frac{1}{2^{k+1}}\cot\frac{\theta}{2^{k+1}}-\cot\theta,\end{aligned}$$

这就是 $P(k+1)$。根据数学归纳法（induction），$P(n)$ 对所有 $n\geq 1$ 和所有 $\theta\in(0,\pi)$ 成立，命题得证。

想一想 假如你从未见过恒等式 $(*)$，怎样才能把它找出来？

从归纳法倒推回去。由 $n=k$ 到 $n=k+1$，右方的变化是 $\frac{1}{2^{k+1}}\cot\frac{\theta}{2^{k+1}}-\frac{1}{2^k}\cot\frac{\theta}{2^k}$，左方的变化是 $\frac{1}{2^{k+1}}\tan\frac{\theta}{2^{k+1}}$。乘以 2^{k+1} 并记 $x=\theta/2^{k+1}$，可见两边的变化相等，当且仅当 $\tan x=\cot x-2\cot 2x$。换言之，题目本身已告诉你要证明哪一条恒等式；之后只需用倍角公式（double-angle formulae）验证便可。

无穷和

各项全为正数，而且我们已有部分和（partial sums）的精确公式，所以只需求出 S_n 在 $n\to\infty$ 时的极限。设 $\varphi_n=\theta/2^n$，则 $\varphi_n\to 0^+$ 且 $\frac{1}{2^n}=\frac{\varphi_n}{\theta}$。于是

$$\frac{1}{2^n}\cot\frac{\theta}{2^n}=\frac{\varphi_n}{\theta}\cdot\frac{\cos\varphi_n}{\sin\varphi_n}=\frac{1}{\theta}\cdot\frac{\varphi_n}{\sin\varphi_n}\cdot\cos\varphi_n\longrightarrow\frac{1}{\theta}\cdot 1\cdot 1=\frac{1}{\theta},$$

这里用到了标准极限：当 $\varphi\to 0$ 时，$\frac{\sin\varphi}{\varphi}\to 1$ 及 $\cos\varphi\to 1$。由于 $\cot\theta$ 与 n 无关，

$$\sum_{r=1}^{\infty}\frac{1}{2^r}\tan\frac{\theta}{2^r}=\lim_{n\to\infty}S_n=\frac{1}{\theta}-\cot\theta,$$

这正是所求的结果。特别地，此级数（series）对所有 $\theta\in(0,\pi)$ 均收敛（converge）。

想一想 为甚么可以说“无穷和等于 S_n 的极限”？

这正是无穷级数之和的定义：$\sum_{r=1}^{\infty}a_r$ 的意思就是 $\lim_{n\to\infty}\sum_{r=1}^{n}a_r$，前提是该极限存在。我们已把极限算了出来，从而证明它存在，所以毋须再做其他工夫。

验算

- $\theta = \frac{\pi}{2}$：右方为 $\frac{2}{\pi} - 0 \approx 0.63662$。以数值方法把级数的首 60 项相加，同样得到 0.6366198。
- $\theta = 1$：级数与 $1 - \cot 1$ 均得 0.3579074。
- θ 很小时：$\tan \frac{\theta}{2^r} \approx \frac{\theta}{2^r}$，所以级数约为 $\theta \sum 4^{-r} = \frac{\theta}{3}$。这与展开式 $\frac{1}{\theta} - \cot\theta = \frac{\theta}{3} + O(\theta^3)$ 吻合。
- 当 $\theta \to \pi^-$ 时，$\cot\theta \to -\infty$，而首项 $\frac{1}{2}\tan\frac{\theta}{2} \to +\infty$，所以两边同时趋向无穷大，正如所料。

常见错误

- 把 $(*)$ 中的因子 2 弄错，例如写成 $\tan x = \cot x - \cot 2x$。可用 $x = \frac{\pi}{4}$ 检验：$1 = 1 - 2\cdot 0$。
- 因为 $\frac{1}{2^n} \to 0$ 便声称 $\frac{1}{2^n}\cot\frac{\theta}{2^n} \to 0$。余切的增长速度与 $2^n/\theta$ 相若，所以乘积趋向 $\frac{1}{\theta}$，而不是 0。
- 忘记说明为何各个正切和余切都有定义：正是条件 $0 < \theta < \pi$ 确保每一个角 $\theta/2^r$ 都在 $(0, \frac{\pi}{2})$ 之内。

STEP 之外

函数 $\frac{1}{\theta} - \cot\theta$ 属于一个著名的家族：对部分分式展开（partial-fraction expansion）$\cot\theta = \frac{1}{\theta} + \sum_{k\geq 1}\frac{2\theta}{\theta^2 - k^2\pi^2}$ 求导，可得巴塞尔和（Basel sum）$\sum 1/k^2 = \pi^2/6$，而欧拉（Euler）正是从 $\sin\theta$ 的无穷乘积（infinite product）得出这类结果。对韦达（Viète）乘积 $\frac{\sin\theta}{\theta} = \prod_{r\geq 1}\cos\frac{\theta}{2^r}$ 取对数再求导，恰好得到本题的级数，这是理解此结果的另一途径。如何令这类对无穷和与无穷乘积的运算变得严谨，是大学一年级分析（analysis）课程的课题；可参阅 Abbott, *Understanding Analysis*（一本数学分析入门教科书）。

详细解答

STEP II 1994，第 7 题

纯数学 · 阿波罗尼奥斯圆

> **STEP II 1994，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/94-s2-q7

关键思路

从距离 r 处观看高度为 h 的旗杆顶，仰角（angle of elevation）的正切是 h/r，所以“仰角相等”其实只是说：士兵与两个杆脚的距离保持固定比 $h_1 : h_2$。把这个条件两边平方，得到的方程正好是第一部分的形式，因此轨迹（locus）是一个圆（即阿波罗尼奥斯圆（circle of Apollonius））。至于最后一部分，把旗杆放在一般位置，并把每个圆心写成两个杆脚的加权组合；恒等式便自然得出，而且其系数之和为零，正是这一点保证了三点共线（collinear）。

解答

圆

由于 $a > 0$，我们可以除以 a，然后配方（complete the square）：

$$x^2 + y^2 + \frac{2g}{a}x + \frac{2f}{a}y + \frac{c}{a} = 0 \iff \left(x + \frac{g}{a}\right)^2 + \left(y + \frac{f}{a}\right)^2 = \frac{g^2 + f^2 - ac}{a^2}.$$

因为 $f^2 + g^2 > ac$，右边是正数，所以这是一个圆，其

$$\text{圆心为} \left(-\frac{g}{a}, -\frac{f}{a}\right), \qquad \text{半径为} \frac{\sqrt{g^2 + f^2 - ac}}{a}.$$

（若没有这个条件，这个“圆”便会是单独一点，或者是空集。）

两支旗杆

把高度为 h_1 的旗杆的杆脚放在原点，另一支的杆脚放在 $(d, 0)$。若士兵位于 (x, y)，与两个杆脚的距离分别是 r_1 和 r_2，则仰角 θ_1, θ_2 满足 $\tan\theta_i = h_i / r_i$。两个角都在 $(0, \frac{\pi}{2})$ 之内，

而 tan 在这个区间上是一对一（one-to-one）的，所以

$$\theta_1 = \theta_2 \iff \frac{h_1}{r_1} = \frac{h_2}{r_2} \iff h_2^2 r_1^2 = h_1^2 r_2^2,$$

最后一步可以逆推，因为所有量都是正数。（士兵不可能站在杆脚上：$r_1 = 0$ 会迫使 $r_2 = 0$ 同时成立。）所以路径是

$$h_2^2(x^2+y^2) = h_1^2\big((x-d)^2+y^2\big) \iff (h_2^2-h_1^2)(x^2+y^2) + 2h_1^2 d\,x - h_1^2 d^2 = 0.$$

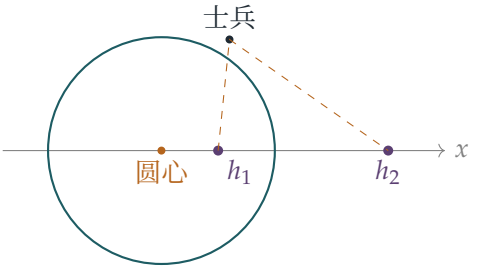
为旗杆编号，使得 $h_2 > h_1$（若两者高度不同，总有一支较高）。这样，上式就是第一部分的方程，其中

$$a = h_2^2 - h_1^2 > 0, \quad g = h_1^2 d, \quad f = 0, \quad c = -h_1^2 d^2,$$

而且

$$g^2 + f^2 - ac = h_1^4 d^2 + (h_2^2 - h_1^2)h_1^2 d^2 = h_1^2 h_2^2 d^2 > 0.$$

所以士兵沿着一个圆步操，圆心为 $\left(-\dfrac{h_1^2 d}{h_2^2 - h_1^2}, 0\right)$，半径为 $\dfrac{h_1 h_2 d}{h_2^2 - h_1^2}$。圆心位于通过两个杆脚的直线上，在较矮旗杆的外侧。这个圆围绕着较矮的旗杆，这很合理：要令矮杆看起来与高杆一样高，就必须站得离它近一些。



（图中取 $h_1 = 1$，$h_2 = 2$，$d = 3$：圆心是 $(-1, 0)$，半径是 2。）

若 $h_1 = h_2$，$x^2 + y^2$ 的项互相抵消，方程变成 $2h_1^2 d\,x = h_1^2 d^2$，即 $x = \frac{1}{2}d$。士兵沿着连接两个杆脚的线段的垂直平分线（perpendicular bisector）行走，那是一条直线而不是圆。这正是上述各圆的极限情况：当 $h_2 \to h_1$ 时，半径趋向无限大。

三支旗杆

现在我们无法把三个杆脚全部放在同一条坐标轴上，所以设高度为 h_i 的旗杆立于 $\mathbf{p}_i = (p_i, q_i)$。以 $r_i^2 = (x-p_i)^2 + (y-q_i)^2$ 重复上面的论证，负责旗杆 i 和 j 的士兵所走的路径是

$$(h_j^2 - h_i^2)(x^2+y^2) - 2(h_j^2 p_i - h_i^2 p_j)x - 2(h_j^2 q_i - h_i^2 q_j)y + \text{常数} = 0.$$

这是一个圆（移动原点不会改变 g^2+f^2-ac 的值，所以它仍然是 $h_i^2h_j^2|\mathbf{p}_i-\mathbf{p}_j|^2>0$；若 $h_j<h_i$，则先把全式乘以 -1）。利用第一部分，其圆心 $\mathbf{c}_{ij}=(x_{ij},y_{ij})$ 满足

$$(h_j^2-h_i^2)\,\mathbf{c}_{ij}=h_j^2\mathbf{p}_i-h_i^2\mathbf{p}_j.$$

把 i 和 j 对调，这条式子不变（两边同时变号），所以 $\mathbf{c}_{31}$ 的下标次序怎样写都没有关系。

> **想一想** 怎样不必解出任何未知量，就把三条这样的公式变成题目给出的恒等式?

恒等式中每一项在 x_{ij} 旁边都已经有因子 $\pm(h_j^2-h_i^2)$，所以我们可以直接代入上面的公式，并使用向量，把 x 坐标和 y 坐标一并处理：

$$\begin{aligned}
h_3^2(h_1^2-h_2^2)\,\mathbf{c}_{12}&=-h_3^2(h_2^2\mathbf{p}_1-h_1^2\mathbf{p}_2)=-h_2^2h_3^2\,\mathbf{p}_1+h_1^2h_3^2\,\mathbf{p}_2,\\
h_1^2(h_2^2-h_3^2)\,\mathbf{c}_{23}&=-h_1^2(h_3^2\mathbf{p}_2-h_2^2\mathbf{p}_3)=-h_1^2h_3^2\,\mathbf{p}_2+h_1^2h_2^2\,\mathbf{p}_3,\\
h_2^2(h_3^2-h_1^2)\,\mathbf{c}_{31}&=\;\;h_2^2(h_3^2\mathbf{p}_1-h_1^2\mathbf{p}_3)=\;\;h_2^2h_3^2\,\mathbf{p}_1-h_1^2h_2^2\,\mathbf{p}_3.
\end{aligned}$$

三式相加，各项两两抵消，所以

$$h_3^2(h_1^2-h_2^2)\,\mathbf{c}_{12}+h_1^2(h_2^2-h_3^2)\,\mathbf{c}_{23}+h_2^2(h_3^2-h_1^2)\,\mathbf{c}_{31}=\mathbf{0}.$$

其 x 分量就是所求的 x_{ij} 之间的关系式，而 y 分量则是 y_{ij} 之间的同一关系式。

> **想一想** 为甚么三点之间有一个线性关系，就能令它们共线?

把三个系数记为 $A=h_3^2(h_1^2-h_2^2)$，$B=h_1^2(h_2^2-h_3^2)$，$C=h_2^2(h_3^2-h_1^2)$。各高度都是正数而且互不相同，所以 A,B,C 全都不是零。关键的额外事实是它们的和为零：

$$A+B+C=h_1^2h_3^2-h_2^2h_3^2+h_1^2h_2^2-h_1^2h_3^2+h_2^2h_3^2-h_1^2h_2^2=0.$$

所以 $C=-A-B$，关系式变成

$$A(\mathbf{c}_{12}-\mathbf{c}_{31})+B(\mathbf{c}_{23}-\mathbf{c}_{31})=\mathbf{0},\qquad\text{所以}\qquad\mathbf{c}_{12}-\mathbf{c}_{31}=-\frac{B}{A}\,(\mathbf{c}_{23}-\mathbf{c}_{31}).$$

由 $\mathbf{c}_{31}$ 指向另外两个圆心的向量互相平行（或者其中一个是零向量，若有两个圆心重合），所以三个圆心位于同一条直线上。假如没有 $A+B+C=0$，一个线性关系根本证明不了甚么：任意三点都满足某个这类的关系。

验算

- 取 $h_1=1$，$h_2=2$，$d=3$，我们的公式给出圆心 $(-1,0)$ 和半径 2。点 $(1,0)$ 与两个杆脚的距离是 1 和 2，而 $(-3,0)$ 的距离是 3 和 6：两者之比都是 $1:2$，正如所料。

- 我们做了一次数值检验：在每个圆上随机取点，确认两个仰角相等。此外，对 2000 组位置随机的三支旗杆，也确认了恒等式成立，以及三个圆心共线。
- 预期共线的另一个理由：若某点同时在其中两个圆上，则 $r_1 : r_2 = h_1 : h_2$ 且 $r_2 : r_3 = h_2 : h_3$，所以 $r_1 : r_3 = h_1 : h_3$，它也在第三个圆上。当这些圆相交时，它们全都通过相同的两点，所以它们的圆心全都在这两点的垂直平分线上。

常见错误

- 直接写 $h_1/r_1 = h_2/r_2$，却没有说明为何仰角相等就得出正切相等（两个角都是锐角）。
- 忘记检查 $g^2 + f^2 > ac$：第一部分只在这个条件下才给出一个圆。
- 单凭恒等式就推断共线，而没有用到 $A + B + C = 0$ 和 $A, B, C \neq 0$。

STEP 之外

与两个定点的距离成固定比 $k \neq 1$ 的点所组成的集合，是一个*阿波罗尼奥斯圆*。取遍 k 的所有值，便得到一个*共轴圆系*（*coaxal system*），即其中任意两个圆都有相同根轴（radical axis）的一族圆。这类圆族正是反演几何（inversive geometry）的起点；在复平面上，它们就是圆 $|z - a| = k|z - b|$，而莫比乌斯变换（Möbius transformation）会把这些圆互相映射。Coxeter 与 Greitzer 的 *Geometry Revisited*（重温几何）涵盖共轴圆和反演；Needham 的 *Visual Complex Analysis*（视觉化的复分析）则展示复数方面的观点。

详细解答

STEP II 1994，第 8 题

纯数学 · 二次模型的最优化

> **STEP II 1994，第 8 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/94-s2-q8

关键思路

追踪一个送货周期内的金钱流向。存货以稳定的速率 r 减少，直至售罄为止，所以 $S(t)$ 只是一条简单的斜线，而仓租就是 b 乘以这条斜线下的面积。于是每个周期的利润是 X 的二次函数（quadratic）。在容许范围 $0 < X \le rT$ 内求它的最大值，便得出所需条件，而假设 $T \ge p/b$ 的作用，正是令抛物线（parabola）的顶点（vertex）落在这个范围之内。

解答

为何 $X \le rT$

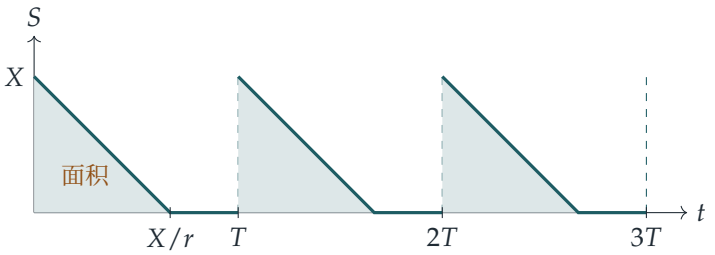
在 T 小时内，她最多只能卖出 rT 件货。若 $X > rT$，下一辆货车到达时，货架上至少仍有 $X - rT$ 件货。此后每个周期都会令这批卖不出的剩货至少再增加 $X - rT$ 件，所以经过 n 个周期后，剩货至少有 $n(X - rT)$ 件。这个数量会无止境地增长，按每件每小时 b 计算的仓租亦然。她要为这些货品付款（还要付仓租），但它们不会带来额外的销售，因为每个周期的销量无论如何都以 rT 为上限。因此每次订货多于 rT 只会增加成本，她应取 $X \le rT$。

存货量

当 $X \le rT$ 时，每个周期开始时恰好有 X 件货。在连续模型（continuous model）中，只要有存货，她便以速率 r 售货，所以

$$S(t) = \begin{cases} X - rt, & 0 \le t \le X/r, \\ 0, & X/r \le t \le T. \end{cases}$$

图像是一条直线，由 X 下降至 $t = X/r$ 时的零，然后保持在零，直到 $t = T$ 下一次送货为止（届时跳回 X）。经过数个周期，便形成锯齿形的图像。



每个周期的利润

在长 T 小时的一个周期内：

- 销售。全部 X 件货都会售出（因为 $X \le rT$），每件售 £$(p+q)$，收入为 £$(p+q)X$。
- 送货。£$(a+qX)$。
- 仓租。在一段短时间 δt 内存放 $S(t)$ 件货的费用为 $bS(t)\,\delta t$，所以总数是

$$b\int_0^T S(t)\,\mathrm{d}t = b\int_0^{X/r}(X-rt)\,\mathrm{d}t = b\cdot\tfrac{1}{2}\cdot X\cdot\frac{X}{r} = \frac{bX^2}{2r},$$

  即阴影三角形的面积乘以 b。
- 营运开支。£cT。

含 qX 的项互相抵销，所以每个周期的利润是

$$P(X) = pX - \frac{bX^2}{2r} - a - cT.$$

可见 p 才是每件货的真正利润：q 只是转嫁给顾客而已。

生意何时可以有利可图?

T 固定时，她可以在 $0 < X \le rT$ 的范围内选择 X。配方（completing the square）后，

$$P(X) = \frac{p^2 r}{2b} - a - cT - \frac{b}{2r}\left(X - \frac{pr}{b}\right)^2.$$

想一想 最佳的 X 真的在容许范围之内吗?

抛物线在 $X = pr/b$ 处达到最高点，而我们需要 $pr/b \le rT$，即 $T \ge p/b$。这正是题目的假设。所以她可以选 $X = pr/b$，此时

$$P = \frac{p^2 r}{2b} - (a + cT).$$

这个值为正，当且仅当 $p^2 r > 2b(a + cT)$，即

$$p^2 > 2\,\frac{(a + cT)b}{r}.$$

因此，若这个不等式（inequality）成立，订货量 $X = pr/b$ 便能令生意有利可图，这正是所要证明的。（事实上 pr/b 是最佳的选择，所以当 $T \geq p/b$ 时，这个条件也是必要的。）

反之，若 $T < p/b$，顶点便落在 rT 之外。这时 P 在整个容许范围内都是递增的，她最多只能取 $X = rT$。这就是为何要得出这个条件，必须有关于 T 的假设。

验算

- 取 $a = 10$、$b = 1$、$r = 10$、$p = 4$、$c = 1$、$T = 5$。则 $T \geq p/b = 4$，且 $p^2 = 16 > 2(15)(1)/10 = 3$。最佳订货量为 $X = 40 \leq rT = 50$，利润为 $160 - 10 - 80 - 5 = 65 = \frac{16 \cdot 10}{2} - 15$。
- 电脑就 30 组随机参数，以数值方法逐步计算一个周期内的 $S(t)$，结果与 $P(X)$ 相差不超过 10^{-7}。另就 2000 个满足 $T \geq p/b$ 的随机个案，对 X 作网格搜索，证实最大值为 $\frac{p^2 r}{2b} - a - cT$。
- 单位：$p^2 r/b$ 的单位是 $(£/件)^2 \times (件/小时)/(£/件/小时) = £$，与 a 和 cT 相同。

常见错误

- 把全部 X 件货在整整 T 小时内的仓租都计算在内（即 bXT），而不是按实际存货量计算的 $b\int S\,\mathrm{d}t$。
- 忘记检查 $X = pr/b$ 是否满足 $X \leq rT$。整道题只有这一步用到条件 $T \geq p/b$。

STEP 之外

这是运筹学（operations research）中经典的经济订货量（economic order quantity）模型。在每次订货的固定成本与每件货每单位时间的存货成本之间取得平衡，会得出一条平方根法则。在本题中，若她还可以选择 T，则要令利润率 P/T 最大，便须在 a 与仓租之间作出取舍。这类约束优化（constrained optimisation）问题（这里是顶点与边界 $X = rT$ 之间的比较）进一步发展下去，便是线性规划（linear programming）、非线性规划（nonlinear programming）以及 Karush–Kuhn–Tucker 条件。若想养成建立和检验这类模型的习惯，Siklos 的 *Advanced Problems in Mathematics* 是很好的参考书。

详细解答

STEP II 1994，第 9 题

力学 · 刚体的平衡

STEP II 1994，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/94-s2-q9

关键思路

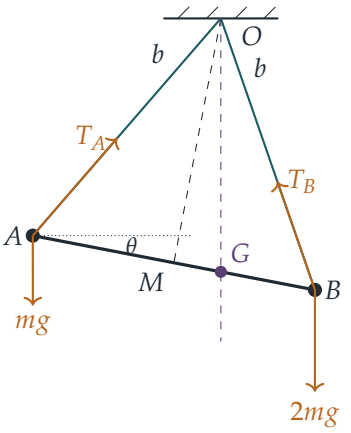
把杆、两条绳和两个质点视为一个刚体（rigid body）。作用于它的外力只有两个重量和 O 点的支承力，因此要达致平衡（equilibrium），作用于质心（centre of mass）G 的合重量，其作用线必须通过 O：即 G 悬于 O 的正下方。这样一来，求角度就变成纯粹的等腰三角形几何问题。至于张力（tension），则对 B 取力矩（moments），这样可消去作用于 B 的两个未知力。

解答

模型。杆和绳都是轻的，绳不可伸长，而 A、B 是质点（particles）。由于 $OA = OB = b$ 且 $AB = 2a$，三角形 OAB 是等腰三角形。设 M 为 AB 的中点；则 $OM \perp AB$，并由毕氏定理（Pythagoras）得

$$OM = h = \sqrt{b^2 - a^2}.$$

较重的质点 B 悬得较低，杆与水平线成角 θ。



角度

考虑整个系统：杆、绳和质点合在一起。作用于它的外力是 A 处的重量 mg、B 处的重量 $2mg$，以及 O 点支承所施的力。对 O 取力矩，支承力没有力矩，所以两个重量对 O 的总力矩必须为零。两个重量互相平行，故可合成为一个通过两质点质心 G 的单一力 $3mg$。它对 O 的力矩为零，当且仅当 G 位于通过 O 的铅垂线上。所以 G 在 O 的正下方。

现在求 G 的位置。它在 AB 上，且 $AG : GB = 2m : m = 2 : 1$，所以 $AG = \frac{2}{3} \cdot 2a = \frac{4a}{3}$，因此

$$MG = AG - AM = \frac{4a}{3} - a = \frac{a}{3},$$

而 G 在 M 靠近 B 的一侧。

> ? **想一想** 为甚么角 MOG 等于 θ?

OG 是铅垂的，而 OM 垂直于杆。把铅垂线旋转到 OM 上的旋转，正是把水平线转到杆上的那个旋转，所以 OG 与 OM 之间的夹角等于杆与水平线之间的夹角，即 θ。

三角形 OMG 在 M 处是直角，所以

$$\tan\theta = \frac{MG}{OM} = \frac{a/3}{\sqrt{b^2 - a^2}} = \frac{a}{3\sqrt{b^2 - a^2}},$$

正如所求。

AO 的张力

再次把杆和质点合起来考虑。作用于这个系统的力是沿 AO 的 T_A、沿 BO 的 T_B，以及两个重量。（杆内的力是内力。）对 B 取力矩，便可消去 T_B 和重量 $2mg$。

由 B 到直线 OA 的垂直距离，等于三角形 OAB 面积的两倍除以 OA。该面积为 $\frac{1}{2} \cdot 2a \cdot h = ah$，所以这距离是 $2ah/b$。A 处重量 mg 的力臂（moment arm）等于 A 与 B 之间的水平

距离，即 $2a\cos\theta$。这两个力矩方向相反，令它们互相平衡，得

$$T_A \cdot \frac{2ah}{b} = mg \cdot 2a\cos\theta \quad \Longrightarrow \quad T_A = \frac{mgb\cos\theta}{h}.$$

由 $\tan\theta = \dfrac{a}{3h}$，且 θ 为锐角，可得 $\cos\theta = \dfrac{3h}{\sqrt{9h^2+a^2}}$，而 $9h^2 + a^2 = 9b^2 - 8a^2$。因此

$$T_A = \frac{mgb}{h} \cdot \frac{3h}{\sqrt{9b^2-8a^2}} = \frac{3mgb}{\sqrt{9b^2-8a^2}}.$$

（以同样的论证对 A 取力矩，可得 $T_B = 2T_A$：每条绳的张力与它所支承的质量成正比。）

验算

- 若 $a \to 0$，两条绳变成铅垂，而 $T_A \to 3mgb/(3b) = mg$，正好是质点 A 所需的张力。
- 对整个系统沿铅垂方向分解力（resolving），结果应为 $3mg$。取 $a=1$、$b=2$、$m=1$、$g=9.8$，得 $T_A \approx 11.11\,\mathrm{N}$ 和 $T_B \approx 22.22\,\mathrm{N}$。把它们沿绳的方向分解，铅垂分量的总和为 $29.4\,\mathrm{N} = 3mg$。
- 以数值方法验算：随机选取 2000 组 (a,b,m)，把杆置于所预测的角度，然后分别对每个质点求解力的平衡。两个质点所得的杆内力一致，而且是推力（thrust），张力则与公式相符，误差约为 10^{-15}。

常见错误

- 把 G 放在距离 M 为 $\frac{2a}{3}$ 之处，或放在错误的一侧。质心较接近较重的质点，与中点相距 $\frac{a}{3}$。
- 把 T_A 对 B 的力臂当作 $2a$ 或 b。所需的是由 B 到直线 OA 的垂直距离。
- 答案仍以 θ 表示，但题目要求以 m、g、a 和 b 表示。

STEP 之外

从一点自由悬挂的物体，静止时质心必在该点的正下方；这个原理其实是一个更深刻事实的最简单情况：稳定平衡使势能（potential energy）取最小值，而这里的势能不过是 $3mg$ 乘以 G 的高度。到了大学，这会发展成虚功原理（virtual work principle）和拉格朗日力学（Lagrangian mechanics），当中求平衡位置的方法是对能量求导，而不是分解力。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对刚体静力学和能量方法有细致的论述。

详细解答

STEP II 1994，第 10 题

力学 · 弹性绳与简谐运动

STEP II 1994，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/94-s2-q10

关键思路

着眼于绳子的伸长量（extension），而不是拖卡的位置。由于货车的速度恒定，伸长量所满足的方程与拖卡加速度的方程完全相同，只差一个正负号，而这个方程正是简谐运动（simple harmonic motion）。因此绳子会在半个振动周期内保持拉紧，然后放开拖卡，而此时拖卡已比货车走得更快。之后便只是一道普通的匀速追及问题。

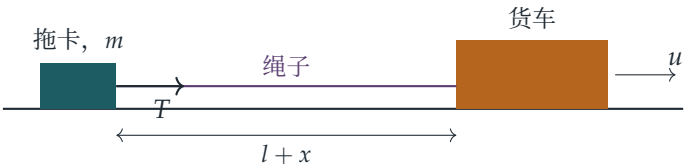
解答

建立模型

把拖卡视为质点（particle），把绳子视为轻绳，拉紧时服从胡克定律（Hooke's law），松弛时不施任何力，并忽略摩擦。由绳子首次拉紧的一刻起计时间 t，距离则沿运动方向量度。在 $t = 0$ 时，货车以速率 u 行驶（并一直保持这个速率），拖卡静止，而货车在拖卡前方距离 l 处。

设拖卡位于 $y(t)$，其中 $y(0) = 0$，则货车位于 $l + ut$。绳子的伸长量为

$$x = (l + ut) - y - l = ut - y.$$



绳子拉紧期间

当 $x > 0$ 时，张力（tension）为 $T = \lambda x/l$，它是拖卡所受的唯一水平力，把拖卡向前拉。所以

$$m\ddot{y} = \frac{\lambda x}{l}.$$

由于 $x = ut - y$ 而 u 为常数，$\ddot{x} = -\ddot{y}$，因此

$$\ddot{x} = -\omega^2 x, \qquad \text{其中} \omega^2 = \frac{\lambda}{ml}.$$

这就是简谐运动。初始条件（initial conditions）为 $x(0) = 0$（绳子刚刚拉紧）和 $\dot{x}(0) = u - \dot{y}(0) = u$，所以

$$x = \frac{u}{\omega}\sin\omega t, \qquad \dot{x} = u\cos\omega t.$$

拖卡的速度为 $\dot{y} = u - \dot{x} = u(1 - \cos\omega t)$。

> **想一想** 按照简谐运动的公式，过了一段时间后 x 会变成负数。那时公式还成立吗？

不成立。方程 $\ddot{x} = -\omega^2 x$ 用到张力 $\lambda x/l$，而这只在 $x > 0$ 时适用；绳子不能推东西。公式在 $0 \le t \le \pi/\omega$ 内成立，此时 $\sin\omega t \ge 0$，而绳子再次松弛的时刻为

$$t_1 = \frac{\pi}{\omega} = \pi\left(\frac{lm}{\lambda}\right)^{1/2}.$$

在那一刻伸长量为零，所以拖卡又一次落后货车距离 l，而它的速度为

$$\dot{y}(t_1) = u(1 - \cos\pi) = 2u.$$

绳子松弛之后

当 $t > t_1$ 时，拖卡以 $2u$ 行驶，货车以 u 行驶，所以两者之间的距离（在 $t = t_1$ 时为 l）以稳定的速率 $2u - u = u$ 缩短。此后这段距离小于 l，所以绳子保持松弛，拖卡仍然不受任何水平力，它的速率确实维持在 $2u$。再过时间 l/u，这段距离便缩短至零。

因此，拖卡与货车相撞的时刻是绳子首次拉紧后的

$$t_1 + \frac{l}{u} = \pi\left(\frac{lm}{\lambda}\right)^{1/2} + \frac{l}{u}$$

正如题目所求。

验算

- 单位：λ 是一个力，所以 lm/λ 的单位是 $\mathrm{m}\cdot\mathrm{kg}/\mathrm{N} = \mathrm{s}^2$，两项都是时间。

- 转换参考系（frame）：绳子的作用就像一条弹簧，令拖卡与一辆质量实际上无限大的货车作弹性碰撞（elastic collision）。在货车的参考系中，拖卡以 u 接近，再以 u 沿相反方向离开，所以它最后相对地面的速率是 $2u$，与上面的结果一致。
- 对拖卡的运动方程作数值积分（numerical integration）（绳子松弛时把力关掉），随机选取 m、l、λ、u 的值，结果与公式吻合；例如选取 m, l, λ, u 使得 $\pi/\omega \approx 2.9944$ 及 $l/u \approx 2.3017$ 时，得出的碰撞时间为 5.2962，而公式的预测值为 5.2961。

常见错误

- 把简谐运动的解用于所有时间，忘记了伸长量回到零时绳子会松弛。
- 直接就拖卡的位置 y 写出运动方程，得到 $m\ddot{y} = \lambda(ut - y)/l$，这是一个受迫方程（forced equation）。它是可以解的，但代换 $x = ut - y$ 能立即把它化成非受迫方程。
- 漏掉最后的 l/u：绳子松弛时，拖卡仍然落后货车整整一个自然长度（natural length）。

STEP 之外

转到货车的参考系后，整个过程就变成拖卡与一条固定在不动墙上的弹簧之间的一次弹性“碰撞”：拖卡以速率 u 进来，以速率 u 离开，而不论速率是多少，接触的时间都恰好是半个周期。这是说明伽利略不变性（Galilean invariance）和能量守恒（energy conservation）如何约束碰撞的最简单例子；这个想法贯穿大学一年级的力学，并引向散射理论（scattering theory）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）正是以这种精神处理振动和碰撞。

详细解答

STEP II 1994，第 11 题

力学 · 抛体与爆炸

STEP II 1994，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/94-s2-q11

关键思路

爆炸发生在一个已知的高度，即垂直飞行的最高点，而那一刻炮弹是静止的。因此两块碎片由同一点出发，速度大小相等、方向相反。若用一个角 θ 来描述方向，每块碎片的落地时间都可由一条二次方程（quadratic equation）求得，而两个落地点之间的距离就成为 θ 的单一函数。之所以出现两种情况，是因为该函数的最大值要么位于容许范围的端点，要么位于范围内部的转向点（turning point）。

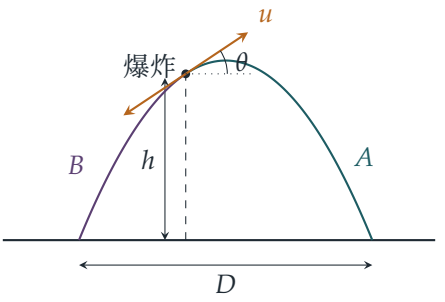
解答

建立模型

我们把炮弹和碎片都视为只受重力作用的质点（particle）。炮弹以初速 v 上升，所以它在以下高度静止：

$$h = \frac{v^2}{2g}, \qquad \text{即} \qquad 2gh = v^2.$$

把刚爆炸后的两块碎片称为 A 和 B。它们的速度大小相等、方向相反，所以两者都在通过爆炸点的同一个垂直平面内。在该平面内取 x 轴为水平方向，y 轴垂直向上，爆炸点位于 $(0, h)$。我们可以适当地命名碎片，使 A 不向下运动。于是 A 的速度为 $(u\cos\theta,\ u\sin\theta)$，$B$ 的速度为 $(-u\cos\theta,\ -u\sin\theta)$，其中 $0 \le \theta \le \frac{\pi}{2}$。记 $s = \sin\theta$ 及 $c = \cos\theta$，故 $c \ge 0$。



落地时间

爆炸后的时间 t，碎片 A 的高度为 $h + ust - \frac{1}{2}gt^2$。它落地时

$$\tfrac{1}{2}gt^2 - us\,t - h = 0.$$

两根之积为 $-2h/g < 0$，所以恰好有一个正根：

$$t_A = \frac{us + \sqrt{u^2s^2 + 2gh}}{g} = \frac{us + \sqrt{u^2s^2 + v^2}}{g}.$$

对于 B，垂直速度为 $-us$，所以我们把 s 换成 $-s$：

$$t_B = \frac{-us + \sqrt{u^2s^2 + v^2}}{g}.$$

这个值是正数，因为 $\sqrt{u^2s^2 + v^2} > us$。

落地点之间的距离

在水平方向上，A 沿 x 轴正方向移动了 $uc\,t_A$，B 沿负方向移动了 $uc\,t_B$，所以它们落在发射点的两侧。因此两个落地点之间的距离为

$$D = uc\,(t_A + t_B) = \frac{2uc\sqrt{u^2s^2 + v^2}}{g}.$$

由于 $D \geq 0$，求 D 的最大值等同于求 D^2 的最大值。设 $x = s^2 \in [0,1]$，则 $c^2 = 1 - x$。于是

$$\frac{g^2D^2}{4u^2} = (1-x)(v^2 + u^2x) = f(x), \qquad \text{记作此式,}$$

并且

$$f(x) = v^2 + (u^2 - v^2)x - u^2x^2.$$

这是 x 的二次函数，首项系数为负，其顶点（vertex）位于

$$x^* = \frac{u^2 - v^2}{2u^2}.$$

想一想 f 的顶点是最大值点，那么为甚么答案不总是 $f(x^*)$?

只有 $x \in [0,1]$ 才是容许的，因为 $x = \sin^2\theta$。留意 $x^* < \frac{1}{2}$ 恒成立，所以顶点永远不会超出上端。顶点是否在容许范围内，只取决于它的正负。

情况 $u \le v$。此时 $x^* \le 0$。在 $[0,1]$ 上，我们位于一条开口向下的抛物线（parabola）的顶点右方，所以 f 在该区间上递减，最大值在 $x = 0$ 处取得，而 $f(0) = v^2$。这对应 $\theta = 0$，即水平方向的爆炸。因此

$$D_{\max} = \frac{2u}{g}\sqrt{v^2} = \frac{2uv}{g}.$$

情况 $v \le u$。此时 $0 \le x^* < \frac{1}{2}$，所以顶点在容许范围内，并给出最大值：

$$\begin{aligned} f(x^*) &= v^2 + \frac{(u^2-v^2)^2}{2u^2} - \frac{(u^2-v^2)^2}{4u^2} = v^2 + \frac{(u^2-v^2)^2}{4u^2} \\ &= \frac{4u^2v^2 + u^4 - 2u^2v^2 + v^4}{4u^2} = \frac{(u^2+v^2)^2}{4u^2}. \end{aligned}$$

所以

$$D_{\max} = \frac{2u}{g}\cdot\frac{u^2+v^2}{2u} = \frac{u^2+v^2}{g}.$$

这在 $\sin^2\theta = \dfrac{u^2-v^2}{2u^2}$ 时发生，所以碎片是与水平方向成一个角度飞出的。

当 $u = v$ 时，两条公式都给出 $2u^2/g$，所以两个答案在两种情况的分界处一致，这是必然的。证明至此完成。

验算

- 单位：uv/g 和 $(u^2+v^2)/g$ 都是（速率）^{2}/（加速度），即长度。
- 取 $g = 9.8$、$u = 30$ 及 $v = 10$，最佳角度为 $\theta \approx 41.8°$，而由 D 的一般公式和 $(u^2+v^2)/g$ 都得出 $D = 1000/9.8 \approx 102.04\,\text{m}$。水平方向的爆炸只给出 $2\cdot 30\cdot 10/9.8 \approx 61.2\,\text{m}$。
- 以数值方法逐步计算两条轨迹，并在细密的网格上扫描 θ，对数百组随机的 (u,v) 都能重现这两个最大值。
- 当 $v = 0$ 时，炮弹在地面爆炸，答案变成 u^2/g。这正是以速率 u、仰角 $45°$ 发射的抛体（projectile）通常的最大射程，理应如此：B 立即撞入地面，而 $x^* = \frac{1}{2}$ 给出 $\theta = 45°$。

常见错误

- 假设最佳的爆炸方向是水平的。这只在 $u \le v$ 时才成立。

- 对两块碎片使用相同的飞行时间。向上抛出的那一块在空中停留得较久。
- 求最大值时忘记了约束条件 $0 \leq \sin^2\theta \leq 1$，因而漏掉了分情况讨论。

STEP 之外

这个简洁的答案有一个漂亮的解释。质心（centre of mass）只是笔直下坠，而相对于质心，两块碎片以相对速率 $2u$ 沿一条直线互相远离。但它们在不同的时间着地，所以问题就变成在约束条件下求最优值，而最大值会随着参数的变化由边界转移到内部。这种转换会在大学一年级最优化（optimisation）课程的拉格朗日乘数（Lagrange multiplier）和 Karush–Kuhn–Tucker 条件中再次出现。关于动量守恒（conservation of momentum）、质心参考系和更深入的抛体问题，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）。

详细解答

STEP II 1994，第 12 题

概率与统计 · 条件概率

STEP II 1994，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/94-s2-q12

关键思路

第一次掷骰一旦定出“目标点数”(point)s,其后每一次掷骰都是同一个独立(independent)试验，只有三种结果：掷出 s（胜）、掷出 7（负），或掷出其他点数（继续）。因此，在给定目标点数的条件下，游戏结束于第几次掷骰服从几何分布（geometric distribution），而整道题目就是按目标点数分情况讨论，再用全概率公式（law of total probability）加起来。最后一部分是同样的计算，只是换了另一个作为条件的事件。

解答

设 S 为两颗公平骰子的点数和。数算 36 个等可能的结果，可得

$$\mathrm{P}(S=s)=\frac{6-|s-7|}{36},\qquad s=2,3,\ldots,12,$$

所以点数和 $2,3,4,5,6,7,8,9,10,11,12$ 的概率为 $\frac{1}{36}\times(1,2,3,4,5,6,5,4,3,2,1)$。

第一次掷骰即获胜

若第一次的点数和是 7 或 11，她立即获胜：

$$\mathrm{P}(\text{获胜掷次为1})=\frac{6}{36}+\frac{2}{36}=\frac{2}{9}.$$

已知掷骰多于一次，在第 n 次掷骰获胜

她掷骰多于一次，当且仅当第一次的点数和是目标点数 $4,5,6,8,9,10$ 其中之一，其概率为

$$\mathrm{P}(M)=\frac{3+4+5+5+4+3}{36}=\frac{2}{3},$$

其中 M 是事件“她掷骰多于一次”。

固定一个目标点数 s，并记 $p_s = \mathrm{P}(S = s)$。其后每一次掷骰都与之前的掷骰独立，而在该次掷骰中，

$$\mathrm{P}(\text{点数和为}s) = p_s, \qquad \mathrm{P}(\text{点数和为}7) = \tfrac{1}{6}, \qquad \mathrm{P}(\text{两者皆非}) = r_s = 1 - p_s - \tfrac{1}{6}.$$

（由于 $s \neq 7$，这两个事件互斥（disjoint）。）她在第 $n \geq 2$ 次掷骰获胜，当且仅当第 $2, \ldots, n-1$ 次掷骰全是“两者皆非”，而第 n 次掷出 s。所以，由独立性，

$$\mathrm{P}(\text{获胜掷次为}n \mid \text{目标点数为}s) = r_s^{n-2}\, p_s.$$

目标点数可以配成概率相等的三对（4 与 10、5 与 9、6 与 8）。记 $p_s = k/36$，其中 $k = 3, 4, 5$；则 $r_s = (30-k)/36$，分别为 $\frac{3}{4}$、$\frac{13}{18}$、$\frac{25}{36}$。由全概率公式，对于 $n \geq 2$，

$$\mathrm{P}(\text{获胜掷次为}n) = \sum_s p_s \cdot r_s^{n-2} p_s = 2\sum_{k=3}^{5} \left(\frac{k}{36}\right)^2 \left(\frac{30-k}{36}\right)^{n-2}.$$

在第 $n \geq 2$ 次掷骰获胜必然意味着 M 发生，所以除以 $\mathrm{P}(M) = \frac{2}{3}$ 就相当于把每个系数乘以 $\frac{3}{2}$：

$$2 \cdot \frac{k^2}{1296} \cdot \frac{3}{2} = \frac{k^2}{432} \quad \text{得出} \quad \frac{9}{432} = \frac{1}{48}, \quad \frac{16}{432} = \frac{1}{27}, \quad \frac{25}{432}.$$

因此

$$\mathrm{P}(\text{获胜掷次为}n \mid M) = \frac{1}{48}\left(\frac{3}{4}\right)^{n-2} + \frac{1}{27}\left(\frac{13}{18}\right)^{n-2} + \frac{25}{432}\left(\frac{25}{36}\right)^{n-2},$$

这正是所要求的结果。

想一想 为甚么不可以只用一个“平均”目标点数和单一个几何分布来计算？

因为继续下去的概率 r_s 取决于目标点数。三个几何分布的混合（mixture）本身并不是几何分布，所以答案中有三个不同的幂，而不是只有一个。

已知掷骰多于 m 次

设 M_m 为事件“她掷骰多于 m 次”，其中 $m > 1$。这个事件发生，当且仅当第一次掷骰定出目标点数 s，而第 $2, \ldots, m$ 次掷骰全是“两者皆非”，所以

$$\mathrm{P}(M_m) = \sum_s p_s\, r_s^{m-1} = \frac{2}{36}\sum_{k=3}^{5} k\left(\frac{30-k}{36}\right)^{m-1}.$$

若 $n \leq m$，在第 n 次掷骰获胜与 M_m 不能同时发生，所以条件概率（conditional probability）为 0。若 $n > m$，在第 n 次掷骰获胜必然意味着 M_m 发生，所以

$$\mathrm{P}(\text{获胜掷次为}n \mid M_m) = \frac{\mathrm{P}(\text{获胜掷次为}n)}{\mathrm{P}(M_m)} = \frac{9\left(\frac{3}{4}\right)^{n-2} + 16\left(\frac{13}{18}\right)^{n-2} + 25\left(\frac{25}{36}\right)^{n-2}}{36\left[3\left(\frac{3}{4}\right)^{m-1} + 4\left(\frac{13}{18}\right)^{m-1} + 5\left(\frac{25}{36}\right)^{m-1}\right]},$$

其中因子 $\frac{2}{1296}$ 和 $\frac{2}{36}$ 已经约去。等价地，把分子和分母同乘以 $\frac{1}{432}$，得

$$\frac{\frac{1}{48}\left(\frac{3}{4}\right)^{n-2}+\frac{1}{27}\left(\frac{13}{18}\right)^{n-2}+\frac{25}{432}\left(\frac{25}{36}\right)^{n-2}}{\frac{1}{4}\left(\frac{3}{4}\right)^{m-1}+\frac{1}{3}\left(\frac{13}{18}\right)^{m-1}+\frac{5}{12}\left(\frac{25}{36}\right)^{m-1}}, \qquad n>m.$$

留意分母正是 $\mathrm{P}(M_m \mid M)$，这是理所当然的：分子是先前以 M 为条件的答案，而 $M_m \subseteq M$。

验算

- 取 $m=1$ 时，分母变成 $\frac{1}{4}+\frac{1}{3}+\frac{5}{12}=1$，还原为上一部分的结果。
- 当 $n=2$ 时，条件概率公式给出 $\frac{1}{48}+\frac{1}{27}+\frac{25}{432}=\frac{50}{432}=\frac{25}{216}\approx 0.1157$；直接计算，$\frac{3}{2}\cdot\frac{2(9+16+25)}{1296}$ 给出相同的值。
- 当 $m=2$、$n=3$ 时，答案是 $\frac{107}{930}\approx 0.1151$。我们用分数把游戏逐步精确推演，核对了所有公式；另外模拟了 4×10^5 局游戏，结果在抽样误差范围内吻合。

常见错误

- 忘记除以 $\mathrm{P}(M)=\frac{2}{3}$，结果得到系数 $\frac{1}{72}$、$\frac{2}{81}$、$\frac{25}{648}$。
- 以单一个公比写成 $\mathrm{P}(M_m)=\left(\text{某数}\right)^{m-1}$：作为条件的事件同样是对各个目标点数的混合。
- 把“两者皆非”的掷骰次数算成 $n-1$ 而不是 $n-2$：第 1 次掷骰用来定出目标点数，第 n 次掷骰则是获胜的一掷。

STEP 之外

这个游戏是一条小型的马可夫链（Markov chain），有“胜”和“负”两个吸收态（absorbing states），而这里的吸收时间是几何分布的混合。以掷过 m 次仍未结束为条件，会令混合的权重移向维持得最久的目标点数（这里是 4 和 10，即最难再次掷出的点数）；这是一个简单的例子，说明无记忆（memoryless）分布的混合本身并不具有无记忆性。Grimmett and Welsh, *Probability: An Introduction*（概率论入门）以这个程度讲解条件概率和马可夫链。

详细解答

STEP II 1994，第 13 题

概率与统计 · 赠券收集问题

STEP II 1994，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/94-s2-q13

本题在 *Beyond STEP* 第 9.11 节集齐一套中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/coupon-collector.html

关键思路

这是经典的“优惠券收集问题”（coupon collector problem），虽然资料库把它归入别的类别，它其实是纯粹的概率（probability）题。把集齐全套所需的漫长等待拆成 n 段较短的等待：先等第一张新卡，再等第二张，如此类推。每段短等待都服从几何分布（geometric distribution），所以平均值很容易求得，再用期望值的线性性质（linearity of expectation）把它们加起来。总和是 n 乘以一个调和和（harmonic sum），而题中的积分不等式正是用来告诉我们这个和有多大。

解答

等待一张新卡

我们假设每包内有 n 张卡的其中一张，每张出现的概率均为 $\frac{1}{n}$，而且各包之间互相独立（independent）。

假设我已经拥有 r 张不同的卡，其中 $0 \leq r \leq n-1$。打开一包得到新卡的概率是

$$p = \frac{n-r}{n} > 0,$$

得到旧卡的概率则是 $q = 1 - p = \frac{r}{n}$。设 X 为我直至第一次得到新卡为止（包括该包在内）所打开的包数。由于各包互相独立，$X = k$ 表示先有 $k-1$ 张旧卡，然后是一张新卡，

所以

$$\mathrm{P}(X=k)=q^{k-1}p, \qquad k=1,2,3,\ldots$$

因此

$$\mathrm{E}(X)=\sum_{k=1}^{\infty} kpq^{k-1}=p\,(1-q)^{-2}=\frac{p}{p^2}=\frac{1}{p},$$

其中用到当 $|q|<1$ 时 $\sum_{k\geq 1} kq^{k-1}=(1-q)^{-2}$（把几何级数（geometric series）求导，或者用二项式定理展开 $(1-q)^{-2}$）。这里 $0\leq q<1$，因为 $r\leq n-1$。所以

$$\mathrm{E}(X)=\frac{n}{n-r},$$

这正是所要证明的结果。

（一个更简洁的做法：第一包要么给出新卡，我们就此停下；要么以概率 q 令我们回到起点，所以 $\mathrm{E}(X)=1+q\,\mathrm{E}(X)$，由此得 $\mathrm{E}(X)=1/(1-q)=1/p$。这个做法假设了 $\mathrm{E}(X)$ 是有限的，而上面的级数已确认了这一点。）

积分不等式

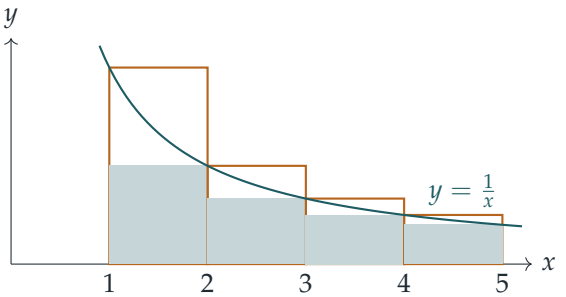
设 $r\geq 1$。当 $r\leq x\leq r+1$ 时，我们有

$$\frac{1}{r+1}\leq\frac{1}{x}\leq\frac{1}{r},$$

因为 $1/x$ 在 $x>0$ 时是递减的。在长度为 1 的区间上积分，这些不等式仍然成立，所以

$$\frac{1}{r+1}\leq\int_r^{r+1}\frac{1}{x}\,\mathrm{d}x\leq\frac{1}{r}.$$

在图中，曲线 $y=1/x$ 在 $x=r$ 与 $x=r+1$ 之间的面积，介乎高度为 $\frac{1}{r+1}$ 的矮长方形与高度为 $\frac{1}{r}$ 的高长方形之间。



（为清楚起见，纵轴的比例已经拉长。）涂上阴影的长方形位于曲线之下；只画出外框的长方形则高于曲线。

现在把 $r = 1, 2, \ldots, n-1$ 的不等式加起来，其中 $n \geq 2$。各个积分会连接成一个：

$$\sum_{r=1}^{n-1} \frac{1}{r+1} \leq \int_1^n \frac{1}{x}\,\mathrm{d}x \leq \sum_{r=1}^{n-1} \frac{1}{r}.$$

中间的一项是 $\ln n - \ln 1 = \ln n$，而在左边的和式中，我们把 $r+1$ 改记为 r，它由 2 取到 n。所以

$$\sum_{r=2}^{n} \frac{1}{r} \leq \ln n \leq \sum_{r=1}^{n-1} \frac{1}{r},$$

这正是所要证明的结果。

集齐全套

设 X_r 为我恰好拥有 r 张不同的卡期间所打开的包数，以给出第 $r+1$ 张卡的那一包作结。集齐全套所需的总包数是

$$T = X_0 + X_1 + \cdots + X_{n-1}.$$

> **想一想** 各个 X_r 取决于之前发生过的事。我们仍然可以直接把它们的期望值相加吗?

可以。期望值的线性性质 $\mathrm{E}(A+B) = \mathrm{E}(A) + \mathrm{E}(B)$ 完全不需要独立性。（事实上，无论之前的经过如何，X_r 都服从 $p = \frac{n-r}{n}$ 的几何分布，所以第一部分的结果适用于每一个 X_r。）因此

$$\mathrm{E}(T) = \sum_{r=0}^{n-1} \frac{n}{n-r} = n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1}\right) = n\sum_{k=1}^{n} \frac{1}{k},$$

其中记 $k = n - r$。

记 $H_n = \sum_{k=1}^{n} \frac{1}{k}$。不等式的两边分别给出

$$H_n = 1 + \sum_{k=2}^{n} \frac{1}{k} \leq 1 + \ln n, \qquad H_n = \sum_{k=1}^{n-1} \frac{1}{k} + \frac{1}{n} \geq \ln n + \frac{1}{n}.$$

所以

$$n\ln n + 1 \leq \mathrm{E}(T) \leq n\ln n + n.$$

上下界之间的差距小于 n，当 n 很大时，这与 $n\ln n$ 相比微不足道（两者之比 $1/\ln n \to 0$）。所以当 n 很大时，我预计大约要吃

$$n\ln n \text{ 包。}$$

例如，100 位数学家就意味着大约 460 包 Cruncho。

验算

- $r = 0$：第一包必定是新卡，而 $n/(n-0) = 1$。$r = n-1$：只欠某一张指定的卡，它出现的概率是 $\frac{1}{n}$，所以我们预计需要 n 包，与 $n/1$ 吻合。
- $n = 2$：$\mathrm{E}(T) = 2(1+\frac{1}{2}) = 3$，而上下界分别给出 $2\ln 2 + 1 \approx 2.39$ 和 $2\ln 2 + 2 \approx 3.39$。
- 以 $n = 20$ 模拟 10^5 次完整的收集过程，平均需要 71.9 包，而 $20H_{20} \approx 71.95$；上下界是 60.9 和 79.9。

常见错误

- 由 $r = 0$ 开始把积分不等式相加：界 $\frac{1}{r}$ 在该处不成立，而且 $\int_0^1 \frac{1}{x}\,\mathrm{d}x$ 发散（diverge）。求和必须由 $r = 1$ 开始。
- 以为各个 X_r 必须互相独立，才可以把期望值相加。线性性质并不需要独立性。
- 答 $\ln n$ 或 n：包数的期望值的增长速度与 $n\ln n$ 相若，比 n 快。

STEP 之外

差 $H_n - \ln n$ 递减至一个极限，即欧拉－马斯刻若尼常数（Euler–Mascheroni constant）$\gamma \approx 0.5772$，所以更精确的答案是 $\mathrm{E}(T) \approx n(\ln n + \gamma)$；当 $n = 100$ 时，这给出大约 519 包，而不是 460 包。像这样把和与积分作比较，就是级数的积分审敛法（integral test），在大学第一门分析课程的初段就会遇到。优惠券收集问题本身亦会在算法分析中再次出现，也会在研究要随机洗牌多少次才能把一副牌洗匀时出现。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率入门教科书）是一个好去处，可以看到几何等待时间与期望值的线性性质如何一并运用。

详细解答

STEP II 1994，第 14 题

概率与统计 · 几何概率

> **STEP II 1994，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/94-s2-q14

关键思路

“在圆盘上均匀分布（uniform）”的意思是概率与面积成正比。因此，落点与圆心的距离不超过 r 的概率，就是较小圆盘的面积除以镖靶的面积，由此立刻得到 R 的分布函数（distribution function）。第 (ii) 部分于是化为三件事：每次投掷的单一概率 p、一个几何分布式的“至少成功一次”计算，以及求一个整数变量函数的最小值。最后一项我们不用微积分，而是考察相邻两项之差。

解答

第 (i) 部分

把每枝飞镖的落点看成在半径为 a 的圆盘上均匀分布的一点，于是飞镖落在镖靶上任何区域的概率，就是该区域的面积除以 πa^2。

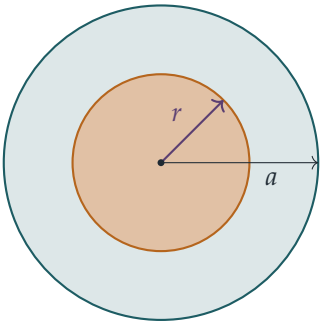
对于 $0 \le r \le a$，事件 $R \le r$ 表示飞镖落在半径为 r 的同心圆盘内，所以

$$\mathrm{P}(R \le r) = \frac{\pi r^2}{\pi a^2} = \frac{r^2}{a^2}.$$

求导可得 R 的概率密度函数（probability density function）为

$$f(r) = \frac{2r}{a^2} \qquad (0 \le r \le a),$$

而在其他情况下 $f(r) = 0$。



注意 R 在 $[0,a]$ 上并非均匀分布：半径越大，面积越多（薄圆环越长），这就是密度随 r 线性增长的原因。

现在

$$\mathrm{E}(R) = \int_0^a r \cdot \frac{2r}{a^2}\,\mathrm{d}r = \frac{2}{a^2}\cdot\frac{a^3}{3} = \frac{2a}{3},$$
$$\mathrm{E}(R^2) = \int_0^a r^2 \cdot \frac{2r}{a^2}\,\mathrm{d}r = \frac{2}{a^2}\cdot\frac{a^4}{4} = \frac{a^2}{2},$$

所以

$$\mathrm{Var}(R) = \mathrm{E}(R^2) - \big(\mathrm{E}(R)\big)^2 = \frac{a^2}{2} - \frac{4a^2}{9} = \frac{a^2}{18},$$

这正是所要证明的方差（variance）。

第 (ii) 部分

（我们把题目中的“with $a/\sqrt{10}$”理解为“within $a/\sqrt{10}$”。）由第 (i) 部分，单次投掷落在距圆心 $a/\sqrt{10}$ 以内的概率为

$$p = \frac{(a/\sqrt{10})^2}{a^2} = \frac{1}{10}.$$

我们假设各次投掷互相独立（independent）（每次都是镖靶上一个全新的均匀随机点）。那么 m 次投掷中没有一次落在 $a/\sqrt{10}$ 以内的概率为 $(0.9)^m$，所以

$$\mathrm{P}(\text{他赢回£12}) = 1 - (0.9)^m.$$

他的损失是所付的 $£(12+m)$ 减去赢回的金额，所以他的期望损失（expected loss，以英镑计）为

$$L(m) = 12 + m - 12\big(1-(0.9)^m\big) = m + 12(0.9)^m.$$

每多投一次，肯定要多付 £1，但赢得 £12 的机会也随之增加；我们需要找出两者在何处取得平衡。

想一想 一个只在整数上有定义的函数，怎样求它的最小值?

比较相邻的值。对于 $m \geq 1$，

$$L(m+1) - L(m) = 1 + 12(0.9)^m(0.9-1) = 1 - 1.2\,(0.9)^m.$$

由于 $(0.9)^m$ 随 m 增大而减小，这个差随 m 递增，所以当差为负时 L 递减，一旦差变为正便递增。差为负当且仅当 $(0.9)^m > \frac{1}{1.2} = \frac{5}{6}$。现在

$$(0.9)^1 = 0.9 > \tfrac{5}{6}, \qquad (0.9)^2 = 0.81 < \tfrac{5}{6},$$

所以 $L(2) - L(1) < 0$，而对每个 $m \geq 2$ 都有 $L(m+1) - L(m) > 0$。因此

$$L(1) > L(2) < L(3) < L(4) < \cdots,$$

而事实上 $L(1) = 11.80$，$L(2) = 11.72$，$L(3) = 11.748$。(他决意要玩，所以不容许 $m = 0$；无论如何 $L(0) = 12$ 也较大。)

他应选择投 $m = 2$ 次，期望损失为 £11.72。由于这个最小值是严格的，$m = 2$ 是唯一符合要求的值。

验算

- 在单位圆盘内模拟 4×10^5 个均匀分布的点，得出平均值 0.667、平方的平均值 0.500 及方差 $0.0556 \approx \frac{1}{18}$，与 $\frac{2}{3}$、$\frac{1}{2}$ 及 $\frac{1}{18}$ 吻合；约有 10% 的点落在距圆心 $1/\sqrt{10}$ 以内。
- 准确值为 $L(1) = \frac{59}{5}$，$L(2) = \frac{293}{25}$，$L(3) = \frac{2937}{250}$，而在 $1 \leq m \leq 99$ 范围内逐一搜索，亦确认 $m = 2$ 为最小值所在。
- 合理性检查：方差 $a^2/18$ 远小于 $[0, a]$ 上均匀分布变量的方差 $a^2/12$，因为 R 集中在靠近边缘的位置。

常见错误

- 把 R 当作在 $[0, a]$ 上均匀分布，这样会得出方差 $a^2/12$。
- 忘记固定的 £12 只是成本的一部分：损失是 $12 + m$ 减去期望赢得的金额，而不只是 m。
- 令 $m + 12(0.9)^m$ 的导数等于零会得出 $m \approx 2.3$，这并不是答案；你必须比较它两侧的整数值。

STEP 之外

由平面上一个均匀分布的点求 R 的分布，是把联合密度（joint density）变换到极坐标（polar coordinates）的最简单例子，其中 $\mathrm{d}A = r\,\mathrm{d}r\,\mathrm{d}\theta$ 里的雅可比（Jacobian）因子 r，正是令 R 的密度与 r 成正比的原因。联合密度的变量代换，以及在离散选择上求最优，都是大学一年级概率课程的标准课题。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率入门教材）涵盖了这两个课题。

详细解答

STEP III 1994，第 1 题

纯数学 · 归约公式

STEP III 1994，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/94-s3-q1

关键思路

对于第一个积分,把 $\operatorname{sech} t$ 写成 $\dfrac{\cosh t}{\cosh^2 t} = \dfrac{\cosh t}{1+\sinh^2 t}$：分子正是 $\sinh t$ 的导数(derivative)，所以这其实是一个乔装了的 arctan。至于归约公式（reduction formula），最自然的做法是分出一个因子 $\operatorname{sech}^2 t$，即 $\tanh t$ 的导数，然后用分部积分法（integration by parts）。恒等式（identity）$\tanh^2 t = 1 - \operatorname{sech}^2 t$ 随即令 I_n 在右边重新出现，我们再把它合并起来。

解答

$\operatorname{sech} t$ 的积分

把分子和分母同乘以 $\cosh t$，并利用 $\cosh^2 t = 1 + \sinh^2 t$，

$$\int_0^x \operatorname{sech} t \,\mathrm{d}t = \int_0^x \frac{\cosh t}{1+\sinh^2 t}\,\mathrm{d}t.$$

设 $s = \sinh t$，则 $\mathrm{d}s = \cosh t \,\mathrm{d}t$，积分上下限变成 0 和 $\sinh x$，所以

$$\int_0^x \operatorname{sech} t \,\mathrm{d}t = \int_0^{\sinh x} \frac{\mathrm{d}s}{1+s^2} = \arctan(\sinh x).$$

另一个同样好的做法是取 $\operatorname{sech} t = \dfrac{2\mathrm{e}^t}{\mathrm{e}^{2t}+1}$ 并设 $u = \mathrm{e}^t$，所得答案的形式是 $2\arctan(\mathrm{e}^x) - \frac{\pi}{2}$。两者是一致的：它们在 $x = 0$ 时都等于零，而且导数都是 $\operatorname{sech} x$。（这个以 x 为变量的函数称为古德曼函数（Gudermannian）。）

归约公式

取 $n \geq 2$，并写 $\operatorname{sech}^n t = \operatorname{sech}^{n-2} t \cdot \operatorname{sech}^2 t$。用分部积分法，取

$$u = \operatorname{sech}^{n-2} t, \qquad \frac{\mathrm{d}v}{\mathrm{d}t} = \operatorname{sech}^2 t,$$

于是 $v = \tanh t$，而根据链式法则（chain rule）（利用 $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{sech} t = -\operatorname{sech} t \tanh t$），

$$\frac{\mathrm{d}u}{\mathrm{d}t} = (n-2)\operatorname{sech}^{n-3} t \cdot (-\operatorname{sech} t \tanh t) = -(n-2)\operatorname{sech}^{n-2} t \tanh t.$$

于是

$$\begin{aligned} I_n &= \left[\operatorname{sech}^{n-2} t \tanh t\right]_0^x + (n-2)\int_0^x \operatorname{sech}^{n-2} t \tanh^2 t \,\mathrm{d}t \\ &= \operatorname{sech}^{n-2} x \tanh x + (n-2)\int_0^x \operatorname{sech}^{n-2} t \left(1 - \operatorname{sech}^2 t\right) \mathrm{d}t \\ &= \operatorname{sech}^{n-2} x \tanh x + (n-2) I_{n-2} - (n-2) I_n, \end{aligned}$$

其中在 0 处的边界项等于零，因为 $\tanh 0 = 0$。

> **想一想** 我们似乎兜了一个圈：I_n 又在右边出现了。这是个问题吗？

完全不是，这正是我们想要的。右边的 I_n 的系数是 $-(n-2)$，而不是 $+1$，所以把它移到左边便得到 $(n-1)I_n$，而我们可以除以 $n-1 \neq 0$。归约公式就是

$$\boxed{(n-1) I_n = \operatorname{sech}^{n-2} x \tanh x + (n-2)\, I_{n-2} \qquad (n \geq 2).}$$

求 I_5

奇数的一串由 $I_1 = \arctan(\sinh x)$ 开始。代入 $n = 3$：

$$2I_3 = \operatorname{sech} x \tanh x + I_1, \qquad \text{所以} \qquad I_3 = \tfrac{1}{2}\operatorname{sech} x \tanh x + \tfrac{1}{2}\arctan(\sinh x).$$

代入 $n = 5$：

$$4I_5 = \operatorname{sech}^3 x \tanh x + 3I_3,$$

所以

$$I_5 = \tfrac{1}{4}\operatorname{sech}^3 x \tanh x + \tfrac{3}{8}\operatorname{sech} x \tanh x + \tfrac{3}{8}\arctan(\sinh x).$$

求 I_6

偶数的一串由 $I_0 = \int_0^x 1 \,\mathrm{d}t = x$ 开始，或者更直接地由 $I_2 = \int_0^x \operatorname{sech}^2 t \,\mathrm{d}t = \tanh x$ 开始（在公式中代入 $n = 2$ 可以确认这一点：$I_2 = \tanh x + 0$）。代入 $n = 4$：

$$3I_4 = \operatorname{sech}^2 x \tanh x + 2I_2, \qquad \text{所以} \qquad I_4 = \tfrac{1}{3}\operatorname{sech}^2 x \tanh x + \tfrac{2}{3}\tanh x.$$

代入 $n = 6$：

$$5I_6 = \operatorname{sech}^4 x \tanh x + 4I_4,$$

所以

$$I_6 = \tfrac{1}{5}\operatorname{sech}^4 x \tanh x + \tfrac{4}{15}\operatorname{sech}^2 x \tanh x + \tfrac{8}{15}\tanh x.$$

偶数次幂的“其他方法”

偶数次幂有一个更简洁的直接做法。设 $u = \tanh t$，则 $\mathrm{d}u = \operatorname{sech}^2 t \,\mathrm{d}t$，而 $\operatorname{sech}^2 t = 1 - u^2$，所以

$$I_6 = \int_0^x (\operatorname{sech}^2 t)^2 \operatorname{sech}^2 t \,\mathrm{d}t = \int_0^{\tanh x} (1-u^2)^2 \,\mathrm{d}u = \tanh x - \tfrac{2}{3}\tanh^3 x + \tfrac{1}{5}\tanh^5 x.$$

在先前的答案中把 $\operatorname{sech}^2 x$ 换成 $1 - \tanh^2 x$ 再展开，所得的正是这个式子，所以两种形式是一致的。

验算

- 当 $x \to \infty$ 时，$\tanh x \to 1$、$\operatorname{sech} x \to 0$ 且 $\arctan(\sinh x) \to \frac{\pi}{2}$，所以 $I_5 \to \frac{3\pi}{16} \approx 0.58905$ 而 $I_6 \to \frac{8}{15}$。在 $[0, 30]$ 上用森逊法则（Simpson's rule）得出 0.589049 和 0.533333。
- 在 $x = 1$ 处，森逊法则给出 $I_5 \approx 0.561566$ 和 $I_6 \approx 0.518343$，与公式吻合；我们亦在 200 个随机的 x 值上，对 $2 \le n \le 8$ 检验过归约公式。
- I_5 和 I_6 都是 x 的奇函数（odd function），这是必然的，因为 sech 是偶函数（even function）。

常见错误

- 弄错 $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{sech} t$ 的正负号：它是 $-\operatorname{sech} t \tanh t$，而正是这个负号令边界项成为正数。
- 把两个 I_n 合并之后，除以 n 而不是 $n - 1$。
- 把答案留作不定积分（indefinite integral）：积分上下限是 0 和 x，所以没有任意常数，而且每一项在 $x = 0$ 时都应该等于零。

STEP 之外

像这样的归约公式，让你可以用递归（recursion）的方法求出一整族积分的值；同一个想法也是沃利斯（Wallis）公式的基础，该公式处理 $\int_0^{\pi/2} \sin^n t \,\mathrm{d}t$，并由此引出一个关于 π 的著名无穷乘积（infinite product）。这里求得的极限 $\int_0^{\infty} \operatorname{sech}^n t \,\mathrm{d}t$ 是贝塔函数

（Beta function）的值，而贝塔函数是大学一年级分析和概率课程中的核心对象。Siklos 的 *Advanced Problems in Mathematics* 有更多归约公式的练习，而 Abbott 的 *Understanding Analysis* 则为反常积分（improper integral）提供严格的处理。

详细解答

STEP III 1994，第 2 题

纯数学 · 回文四次方程

STEP III 1994，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/94-s3-q2

关键思路

(i) 中的四次方程是回文（palindromic）的：它的系数由前往后读和由后往前读都一样。除以 x^2 之后，x^2 与 x^{-2} 配成一对，x 与 x^{-1} 配成一对，而这两对都可以用 $y = x + x^{-1}$ 表示，于是四次方程（quartic）便简化成 y 的二次方程（quadratic）。(ii) 没有提示，所以我们要找出同一种对称性（symmetry）的变奏：只要容许加入一个缩放因子，系数 1, 1, −10, −4, 16 同样可以配对，由此想到代换（substitution）$y = x - 4/x$。

解答

第 (i) 部分

常数项为 $1 \neq 0$，所以 $x = 0$ 不是根（root），我们可以除以 x^2：

$$x^2 + 10x + 26 + \frac{10}{x} + \frac{1}{x^2} = 0, \qquad \text{即} \qquad \left(x^2 + \frac{1}{x^2}\right) + 10\left(x + \frac{1}{x}\right) + 26 = 0.$$

设 $y = x + x^{-1}$，则 $y^2 = x^2 + 2 + x^{-2}$，所以 $x^2 + x^{-2} = y^2 - 2$。方程变成

$$y^2 - 2 + 10y + 26 = 0 \iff y^2 + 10y + 24 = 0 \iff (y + 4)(y + 6) = 0,$$

所以 $y = -4$ 或 $y = -6$。现在回到 x，把 $x + x^{-1} = y$ 两边乘以 x（因为 $x \neq 0$，这一步同样没有问题）：

$$x + \frac{1}{x} = -4 \iff x^2 + 4x + 1 = 0 \iff x = -2 \pm \sqrt{3},$$

$$x + \frac{1}{x} = -6 \iff x^2 + 6x + 1 = 0 \iff x = -3 \pm 2\sqrt{2}.$$

这是四个互不相同的数，而四次方程最多只有四个根，所以方程的解是

$$x = -2 \pm \sqrt{3}, \qquad x = -3 \pm 2\sqrt{2}.$$

留意这些根是以互为倒数（reciprocal）的形式成对出现的：$(-2+\sqrt{3})(-2-\sqrt{3})=1$ 和 $(-3+2\sqrt{2})(-3-2\sqrt{2})=1$。回文多项式（polynomial）必然如此，因为把 x 换成 $1/x$ 再乘以 x^4，便会得回同一条方程。

第 (ii) 部分

想一想 系数 1, 1, −10, −4, 16 并不对称。有没有类似 (i) 的代换仍然行得通?

比较外侧的两对系数。常数项是首项系数（leading coefficient）的 $16=4^2$ 倍，而 x 的系数是 x^3 的系数的 -4 倍。因此各项可以这样配对：

$$x^4+16=x^2\left(x^2+\frac{16}{x^2}\right),\qquad x^3-4x=x^2\left(x-\frac{4}{x}\right).$$

这就是把 $1/x$ 换成 $-4/x$ 之后的回文规律。由于 $x=0$ 不是根（常数项是 16），除以 x^2 可得

$$\left(x^2+\frac{16}{x^2}\right)+\left(x-\frac{4}{x}\right)-10=0.$$

设 $y=x-4/x$。则 $y^2=x^2-8+16/x^2$，所以 $x^2+16/x^2=y^2+8$，而

$$y^2+8+y-10=0 \iff y^2+y-2=0 \iff (y-1)(y+2)=0,$$

得 $y=1$ 或 $y=-2$。把 $x-4/x=y$ 乘以 x：

$$x-\frac{4}{x}=1 \iff x^2-x-4=0 \iff x=\frac{1\pm\sqrt{17}}{2},$$
$$x-\frac{4}{x}=-2 \iff x^2+2x-4=0 \iff x=-1\pm\sqrt{5}.$$

这同样是四个互不相同的根，所以全部的根就是这些：

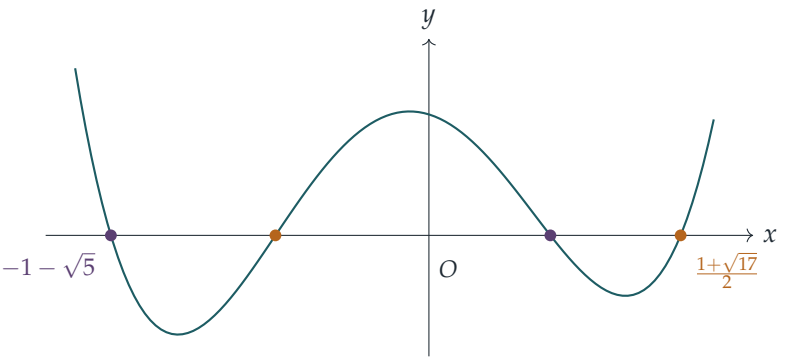
$$x=\frac{1\pm\sqrt{17}}{2},\qquad x=-1\pm\sqrt{5}.$$

这次各对根的积是 -4 而不是 1：对称性现在是 $x\mapsto -4/x$。

作为验算，以上的运算相当于因式分解（factorisation）

$$x^4+x^3-10x^2-4x+16=(x^2-x-4)(x^2+2x-4),$$

你可以展开来确认：x^3 项给出 $2-1=1$，x^2 项给出 $-4-2-4=-10$，而 x 项给出 $4-8=-4$。



(ii) 中四次多项式的图像与横轴相交四次：琥珀色的根来自 $x^2-x-4=0$，紫红色的根来自 $x^2+2x-4=0$。

验算

- 根的积：在 (i) 中，四个根相乘得 $1\cdot 1=1$，即常数项；在 (ii) 中，它们相乘得 $(-4)(-4)=16$，同样是常数项。
- 根的和：在 (i) 中，$(-4)+(-6)=-10$，即 x^3 系数的相反数；在 (ii) 中，$1+(-2)=-1$，情况相同。
- 数值上，$x=-2+\sqrt{3}\approx -0.268$ 和 $x=-3-2\sqrt{2}\approx -5.828$ 都令 (i) 中的四次多项式的值为零，误差在 10^{-9} 之内；而 $x=\frac{1}{2}(1+\sqrt{17})\approx 2.562$ 和 $x=-1-\sqrt{5}\approx -3.236$ 在 (ii) 中亦然。

常见错误

- 忘记交叉项 (cross term)：$x^2+x^{-2}=y^2-2$，而不是 y^2；在 (ii) 中则是 $x^2+16x^{-2}=y^2+8$，这里是加号，因为 $(x-4/x)^2$ 的交叉项是 -8。
- 在 (ii) 中尝试 $y=x+4/x$。这样可得 $x^2+16/x^2=y^2-8$，却无法得出 x^3-4x，因为它的正负号要求用 $x-4/x$。
- 求出 y 的值便停下来。题目要求的是 x，而每个 y 的值对应两个 x 的值。

STEP 之外

这个技巧之所以行得通，是因为这些多项式在某个代换之下保持不变 (invariant)：(i) 中是 $x\mapsto 1/x$，(ii) 中是 $x\mapsto -4/x$。y 这个量在该代换之下不变，而任何在该代换下不

变的多项式表达式，都可以用 y 来表示。这是对伽罗瓦理论（Galois theory）的初步一瞥：在这套理论中，多项式各根之间的对称性决定了对应的方程可以怎样求解。Allenby 的 *Rings, Fields and Groups*（环、域与群）是了解这些想法如何发展下去的好去处。

详细解答

STEP III 1994，第 3 题

纯数学 · 球面、平面与扣连

> **STEP III 1994，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/94-s3-q3

关键思路

每个圆各自位于一个平面上，而这两个平面相交于直线 L。因此 C_1 能够穿过 C_2 所在平面的地方，就只有它在 L 上的两个点，反过来也是一样。于是，两个圆是否互相扣连，归根究底是一个一维的问题：L 上的四个点沿着这条直线是怎样排列的。

解答

平面与球面

设一个球面（sphere）的球心为 Q、半径为 r，并设 d 为 Q 到平面 Π 的距离。设 F 为由 Q 到 Π 的垂足（foot of the perpendicular）。对于 Π 上的一点 X，由毕氏定理（Pythagoras）得 $QX^2 = d^2 + FX^2$，所以 X 在球面上当且仅当 $FX^2 = r^2 - d^2$。因此：

- 若 $d > r$，平面与球面不相交；
- 若 $d = r$，平面与球面相切于唯一的一点 F（它是一个切平面（tangent plane））；
- 若 $d < r$，交集是 Π 上的一个圆，圆心为 F，半径为 $\sqrt{r^2 - d^2}$。当 $d = 0$ 时，这是一个大圆（great circle），半径为 r。

已知资料

配方（completing the square）后，S_2 为

$$x^2 + (y-3)^2 + (z-2)^2 = 3,$$

球心为 $(0,3,2)$，半径为 $\sqrt{3}$。球面 S_1 的球心为 O，半径为 $\sqrt{7}$。

O 到 P_1 的距离为 $\frac{|-1|}{\sqrt{10}} = \frac{1}{\sqrt{10}} < \sqrt{7}$，所以 C_1 确实是一个圆，半径为 $\sqrt{7 - \frac{1}{10}}$。$(0,3,2)$ 到 P_2 的距离为 $\frac{|0-3+1|}{\sqrt{2}} = \sqrt{2} < \sqrt{3}$，所以 C_2 是一个半径为 $\sqrt{3-2} = 1$ 的圆。它的圆心就是垂足 $(0,3,2) + (1,-1,0) = (1,2,2)$；验证一下：$1 - 2 + 1 = 0$。

直线 L

把两个平面方程相减，得 $2x - 2 = 0$，所以 $x = 1$，继而 $y = x + 1 = 2$，而 z 可取任意值。所以

$$L = \{(1,2,t) : t \in \mathbb{R}\},$$

即通过 $(1,2,0)$ 的铅垂线。

在 S_1 上：$1 + 4 + t^2 = 7$，所以 $t = \pm\sqrt{2}$。因此 L 与 S_1 相交于

$$(1,2,\sqrt{2}) \quad \text{和} \quad (1,2,-\sqrt{2}).$$

在 S_2 上：$1 + 1 + (t-2)^2 = 3$，所以 $t - 2 = \pm 1$。因此 L 与 S_2 相交于

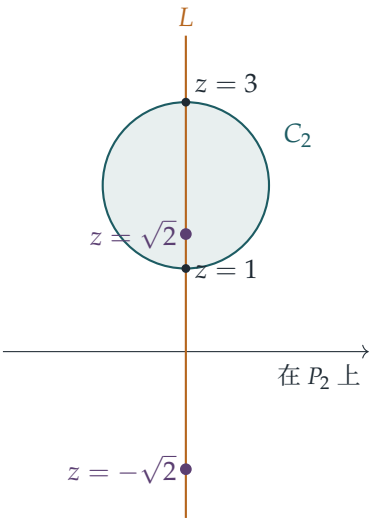
$$(1,2,1) \quad \text{和} \quad (1,2,3).$$

由于 L 位于 P_1 上，L 在 S_1 上的点正好就是 L 在 C_1 上的点；同样地，L 在 S_2 上的点就是 L 在 C_2 上的点。特别地，$(1,2,1)$ 和 $(1,2,3)$ 是 C_2 一条直径的两个端点，这与它的圆心 $(1,2,2)$ 和半径 1 吻合。

两个圆是否互相扣连?

首先，两个圆并不相交。任何公共点都必定位于 $P_1 \cap P_2 = L$ 上，而 C_1 在 L 上的点（$z = \pm\sqrt{2}$）与 C_2 在 L 上的点（$z = 1,3$）并不相同。

设 D_2 为 P_2 上以 C_2 为边界的平面圆盘（disc）。它与 L 的交集是由 $(1,2,1)$ 到 $(1,2,3)$ 的直径，即满足 $1 \le t \le 3$ 的点 $(1,2,t)$。



图中所示是平面 P_2，当中 C_2 及其圆盘以阴影表示。紫色的点是 C_1 穿过 P_2 的位置。

想一想 为甚么 C_1 在这两个点真的是穿过 P_2，而不只是与它相切?

直线 L 把平面 P_1 分成两个半平面（half-plane），而由于 P_1 沿着 L 横越 P_2，这两个半平面分别位于 P_2 的两侧。圆 C_1 与 L 相交于两个不同的点，所以 L 把 C_1 切成两段弧，每个半平面各有一段。因此 C_1 的一段弧完全位于 P_2 的一侧，另一段弧则完全位于另一侧，而 C_1 在 $(1, 2, \sqrt{2})$ 和 $(1, 2, -\sqrt{2})$ 穿过 P_2，此外再无其他穿越点。

现在 $1 < \sqrt{2} < 3$，所以在 $(1, 2, \sqrt{2})$ 的穿越点位于圆盘 D_2 之内。但 $-\sqrt{2} < 1$，所以在 $(1, 2, -\sqrt{2})$ 的穿越点位于圆盘之外。沿 C_1 绕一圈，我们恰好穿过 C_2 所张成的圆盘一次，然后在圆盘之外越过 P_2 返回。换言之，C_1 穿过 C_2 一次，就像链条上相扣的两个环。

沿着 L，这四个点的次序是

$$-\sqrt{2} \ < \ 1 \ < \ \sqrt{2} \ < \ 3,$$

所以来自 C_1 的一对点与来自 C_2 的一对点互相交错（interleave）。正正是这一点令 C_1 穿过 D_2 一次。（假如两对点都没有把对方分隔开，C_1 就会穿过 D_2 两次或完全不穿过，两个圆便不会互相扣连。）因此**圆 C_1 与 C_2 互相扣连**（linked）。

我们无法把它们拉开：任何把它们变成两个相距甚远的圆的连续移动，都必须把 C_1 穿过 C_2 的圆盘的次数由奇数变为偶数。穿越点总是成对地出现和消失，除非其中一个圆穿过另一个圆，而在不剪断的情况下把它们分开，是不容许这样做的。

验算

把各点逐一代入：例如 $(1,2,1)$ 给出 $3-2-1=0$、$1-2+1=0$ 和 $1+4+1-12-4+10=0$。数值方面，我们把两个圆参数化（parametrise），并计算高斯（Gauss）的环绕积分（linking integral）$\frac{1}{4\pi}\oint\oint \frac{(\mathbf{r}_1-\mathbf{r}_2)\cdot(\mathrm{d}\mathbf{r}_1\times\mathrm{d}\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|^3}$，得到 1.002（不扣连的一对圆会给出 0）。我们亦求得两个圆之间的最短距离约为 0.09，所以它们十分接近，但并不相碰。

常见错误

- 只检查两个圆是否相交。它们并不相交，但单凭这一点完全无法说明它们是否扣连。
- 忘记了 C_1 与 P_2 的穿越点必定位于 L 上，反而在别处寻找。
- 弄错 S_2 的球心：$-6y$ 和 $-4z$ 两项给出的球心是 $(0,3,2)$，而不是 $(0,-3,-2)$。

STEP 之外

两条互不相交的闭曲线（closed curve）是否扣连，是由一个整数不变量（invariant）来量度的，称为*环绕数*（linking number）。计算它的方法，可以是数算其中一条曲线穿过另一条曲线所张成的曲面的带符号穿越次数（这正是我们所做的），也可以是使用“验算”一节中高斯的二重积分（double integral）。环绕数是踏入纽结理论（knot theory）和代数拓扑（algebraic topology）的第一步。Colin Adams 的 *The Knot Book*（一本纽结理论入门书）是十分易读的入门读物。

详细解答

STEP III 1994，第 4 题

纯数学 · 常微分方程的奇解

STEP III 1994，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/94-s3-q4

本题在 *Beyond STEP* 第 8.1 节当解不唯一中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/non-uniqueness.html

关键思路

两边开平方：方程即 $\mathrm{d}y/\mathrm{d}x = \pm 2\sqrt{y}$，马上可以分离变量（separation of variables）。更值得留意的是，每一条抛物线（parabola）$y = (x-c)^2$ 都满足这个方程，所以方程的解不过是把 $y = x^2$ 沿 x 轴滑动所得的各个副本。一旦看出这幅图像，其后各部分要问的，就是哪些副本通过哪些点，以及所有副本有甚么共通之处。

解答

通解（general solution）

由于 $(\mathrm{d}y/\mathrm{d}x)^2 \geq 0$，任何解都有 $y \geq 0$。在 $y > 0$ 之处，可以写成

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm 2\sqrt{y} \quad \Longrightarrow \quad \frac{1}{2\sqrt{y}}\frac{\mathrm{d}y}{\mathrm{d}x} = \pm 1 \quad \Longrightarrow \quad \sqrt{y} = \pm(x-c)$$

其中 c 为常数，而无论取哪一个符号，平方后都得到

$$y = (x-c)^2.$$

值得直接验证整条抛物线都是解，包括左右两半和顶点（vertex）：$y' = 2(x-c)$，所以对每一个 x 都有 $(y')^2 = 4(x-c)^2 = 4y$。

? **想一想** 两个符号得出同一条公式。那么“±”到哪里去了？

它藏在每条抛物线的两半之中。在 $y=(x-c)^2$ 上，斜率（gradient）$2(x-c)$ 在顶点左方等于 $-2\sqrt{y}$，在右方等于 $+2\sqrt{y}$，所以单单一条抛物线已经用上两个符号。到我们要令曲线通过某一点时，符号的选择便以 c 的选择这个形式重新出现。

通过 (a,b^2) 的解

我们需要 $(a-c)^2=b^2$，所以 $a-c=\pm b$，即 $c=a-b$ 或 $c=a+b$。两个解是

$$y=(x-a+b)^2 \qquad 及 \qquad y=(x-a-b)^2.$$

两者并不相同，因为 $b\neq 0$ 令 $a-b\neq a+b$。（假如 $b=0$，两者便会合并成一条顶点在 $(a,0)$ 的抛物线；这时该点亦会位于解 $y=0$ 之上，这个解我们在最后会再遇到。）把 b 换成 $-b$ 只会把两个解对调，这是理所当然的，因为点 (a,b^2) 并无改变。

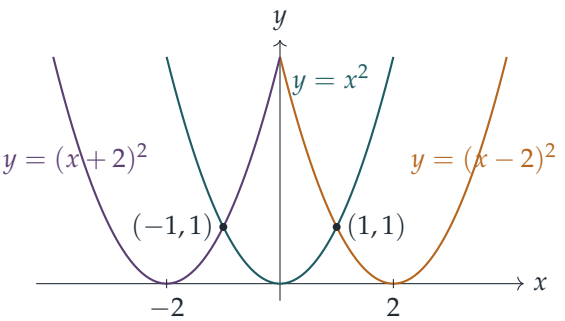
点 $(a_1,1)$ 和 $(a_2,1)$

取 $b=1$：通过 $(a,1)$ 的解是 $y=(x-a+1)^2$ 和 $y=(x-a-1)^2$。抛物线 $y=(x-c)^2$ 通过原点，当且仅当 $c=0$，所以我们需要 $a-1=0$ 或 $a+1=0$。因此

$$a_1=1, \qquad a_2=-1.$$

通过 $(1,1)$ 的解是 $y=x^2$（来自 $c=a_1-1=0$）和 $y=(x-2)^2$。通过 $(-1,1)$ 的解是 $y=(x+2)^2$ 和 $y=x^2$（来自 $c=a_2+1=0$）。

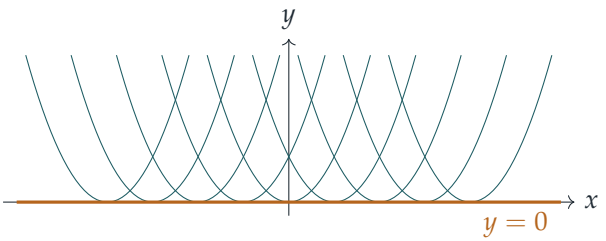
对这两点来说，通过原点的解都是同一条曲线 $y=x^2$，所以它们的图像重合。这并不令人意外：$(1,1)$ 和 $(-1,1)$ 正是 $y=x^2$ 上高度为 1 的两点，分别位于顶点两侧。另外两个解是 $y=(x-2)^2$ 和 $y=(x+2)^2$，即把 $y=x^2$ 分别向右和向左平移 2。



在 $(1,1)$，曲线 $y=x^2$ 的斜率是 $+2$，而 $y=(x-2)^2$ 的斜率是 -2：通过同一点的两个解总是在该点相交，斜率为 $\pm 2b$，每个平方根各占其一。

曲线族及其公切线（common tangent）

改变 a 和 b，只不过是令 $c = a \pm b$ 取遍所有实数，所以这个曲线族就是 $y = x^2$ 的所有水平平移（translate）。



每一条抛物线都在其顶点 $(c, 0)$ 与 x 轴相切，该处的斜率为 0。所以直线 $y = 0$ 是整个曲线族的公切线（它是这族曲线的包络线（*envelope*））。

它之所以重要，是因为 $y = 0$ 本身也是一个解：这时 $y' = 0$，而 $(y')^2 = 0 = 4y$。但无论 c 取甚么值，它都不是其中一条抛物线：它是一个奇异解（*singular solution*），分离变量时因为除以 $\sqrt{y}$ 而遗漏了它。在每一点 $(c, 0)$，解 $y = 0$ 与抛物线 $y = (x - c)^2$ 有相同的值和相同的斜率，所以通过 x 轴上各点的解并不唯一。事实上我们可以把它们拼接起来：对于 $p < q$，

$$y = \begin{cases} (x-p)^2, & x < p, \\ 0, & p \le x \le q, \\ (x-q)^2, & x > q \end{cases}$$

有连续的导数，并且处处满足方程。所以严格来说，通过 (a, b^2) 的“两个解”是指在该点附近与那两条抛物线吻合的两个解；离开该点之后，解一旦到达 x 轴，便可以沿着该轴走一段才离开。

验算

- 代回原式：对于 $y = (x - a \pm b)^2$，在 $x = a$ 时得 $y = b^2$，而 $(y')^2 = 4y$ 恒成立。
- 在 (a, b^2) 的两个斜率是 $2(a - c) = \pm 2b$，它们的平方都是 $4b^2 = 4y$，符合要求。
- 数值上，以中心差分（central difference）计算导数，确认了随机选取的抛物线以及上述取 $p = -1$、$q = 1$ 的拼接曲线都满足 $(y')^2 = 4y$，亦确认了通过 $(1, 1)$ 和 $(-1, 1)$ 的解分别有 $c \in \{0, 2\}$ 和 $c \in \{-2, 0\}$，两者只共有 $c = 0$。

常见错误

- 只写 $y' = 2\sqrt{y}$，结果通过每一点只得到半条抛物线。
- 忘记了 $y = 0$：它是一个真正的解，而且正正是最后一部分所要找的公切线。
- 以为通过 $(a_1, 1)$ 和 $(a_2, 1)$ 而又经过原点的两个解是不同的曲线;其实两者都是 $y = x^2$。

STEP 之外

这个方程是解的唯一性（uniqueness）失效的标准例子。皮卡–林德勒夫定理（Picard–Lindelöf theorem）保证，当 f 对 y 满足利普希茨（Lipschitz）条件时，$y' = f(x, y)$ 通过一点的解恰好只有一个；但 $\pm 2\sqrt{y}$ 在 $y = 0$ 附近并不满足利普希茨条件，而那里正是抛物线与奇异解相遇之处。单参数曲线族的包络线，以及它们为克莱罗（Clairaut）方程一类方程所带来的奇异解，是常微分方程（ODE）入门课程中的经典课题。Boyce and DiPrima, *Elementary Differential Equations*（常微分方程入门教科书），以类似这个的例子讨论了解的存在性和唯一性。

详细解答

STEP III 1994，第 5 题

纯数学 · 麦克劳林级数

> **STEP III 1994，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/94-s3-q5

关键思路

把 $\sin^{-1}x$ 求导一次会得到一个平方根，再求导一次，便得到 f'' 与 f' 之间一个不含平方根的关系式。然后利用莱布尼兹定理（Leibniz's theorem），就可以把这一条方程变成所有高阶导数的递推关系（recurrence）。在 $x=0$ 处，大部分的项都会消失，于是得到麦克劳林系数（Maclaurin coefficients）的一个简单递推关系。至于 g，利用对数定律便可把一切化为我们熟悉的 $\ln(1\pm x)$ 的级数。

解答

微分方程

对于 $-1<x<1$，

$$\mathrm{f}'(x)=\frac{1}{\sqrt{1-x^2}}=(1-x^2)^{-1/2},\qquad \mathrm{f}''(x)=-\tfrac{1}{2}(1-x^2)^{-3/2}\cdot(-2x)=\frac{x}{(1-x^2)^{3/2}}.$$

把第二式乘以 $1-x^2$，得 $(1-x^2)\mathrm{f}''(x)=x(1-x^2)^{-1/2}=x\mathrm{f}'(x)$，所以

$$(1-x^2)\mathrm{f}''(x)-x\mathrm{f}'(x)=0,$$

这正是所需的结果。

递推关系

利用莱布尼兹定理 $(uv)^{(n)}=\sum_r \binom{n}{r}u^{(r)}v^{(n-r)}$，把这个恒等式求导 n 次。在第一个乘积中，取 $u=1-x^2$。它的各阶导数依次为 $-2x$、-2，之后全为 0，所以只有三项留下来：

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[(1-x^2)\mathrm{f}''\right]=(1-x^2)\mathrm{f}^{(n+2)}+n(-2x)\mathrm{f}^{(n+1)}+\binom{n}{2}(-2)\mathrm{f}^{(n)}.$$

在第二个乘积中 $u = x$，只有两项留下来：

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[x\mathrm{f}'\right] = x\mathrm{f}^{(n+1)} + n\mathrm{f}^{(n)}.$$

两式相减，并利用 $2\binom{n}{2} = n(n-1)$，

$$\begin{aligned} 0 &= (1-x^2)\mathrm{f}^{(n+2)} - (2n+1)x\mathrm{f}^{(n+1)} - \big(n(n-1)+n\big)\mathrm{f}^{(n)} \\ &= (1-x^2)\mathrm{f}^{(n+2)} - (2n+1)x\mathrm{f}^{(n+1)} - n^2\mathrm{f}^{(n)}, \end{aligned}$$

这就是所需的结果。它对每个 $n \geq 1$ 都成立，而当 $n = 0$ 时，它就是最初的那条方程。（如果你喜欢，也可以改用归纳法（induction）来证明：把 n 的情形求导一次，再整理各项，便得到 $n+1$ 的情形。这其实是同一个计算，只是逐步进行而已。）

f 的麦克劳林级数

> **想一想** x 取哪个值，会令这个递推关系容易使用？

代入 $x = 0$。中间的一项消失，而 $1 - x^2 = 1$，所以

$$\mathrm{f}^{(n+2)}(0) = n^2\,\mathrm{f}^{(n)}(0) \qquad (n \geq 0).$$

我们由 $\mathrm{f}(0) = \sin^{-1} 0 = 0$ 和 $\mathrm{f}'(0) = 1$ 出发。

偶数阶导数。由于 $\mathrm{f}(0) = 0$，根据递推关系，每个偶数阶导数在 0 处都等于零。这与我们的预期相符，因为 $\sin^{-1}$ 是奇函数（odd function）。

奇数阶导数。取 $n = 1, 3, 5, \ldots$，

$$\mathrm{f}'''(0) = 1^2, \quad \mathrm{f}^{(5)}(0) = 3^2 \cdot 1^2, \quad \mathrm{f}^{(7)}(0) = 5^2 \cdot 3^2 \cdot 1^2,$$

而一般地

$$\mathrm{f}^{(2k+1)}(0) = \big(1 \cdot 3 \cdot 5 \cdots (2k-1)\big)^2 \qquad (k \geq 1).$$

因此

$$\sin^{-1} x = x + \frac{1^2}{3!}x^3 + \frac{1^2 \cdot 3^2}{5!}x^5 + \frac{1^2 \cdot 3^2 \cdot 5^2}{7!}x^7 + \cdots = x + \frac{x^3}{6} + \frac{3x^5}{40} + \frac{5x^7}{112} + \cdots.$$

x^{2k+1} 的系数可以化简得很整齐。写出 $(2k+1)! = \big(1 \cdot 3 \cdots (2k-1)\big)\big(2 \cdot 4 \cdots 2k\big)(2k+1)$，再约去一个奇数乘积：

$$a_{2k+1} = \frac{\big(1 \cdot 3 \cdots (2k-1)\big)^2}{(2k+1)!} = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2 \cdot 4 \cdot 6 \cdots (2k)} \cdot \frac{1}{2k+1}.$$

由比值审敛法（ratio test）可知，这个级数在 $|x| < 1$ 时收敛，因为 $a_{2k+3}/a_{2k+1} \to 1$。至于它在该范围内确实等于 $\sin^{-1} x$，则可以把 $(1-x^2)^{-1/2}$ 的二项级数（binomial series）逐项积分来确认，所得的系数完全相同。

函数 g

根据对数定律，$\mathrm{g}(x) = \frac{1}{2}\big(\ln(1+x) - \ln(1-x)\big)$。对于 $|x| < 1$，

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \qquad \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots.$$

两者相减时，偶数次幂互相抵消，奇数次幂则变成两倍。所以

$$\mathrm{g}(x) = x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1}.$$

（你也许会认出 g 就是 $\tanh^{-1} x$。）

比较系数

> ? **想一想** 两个系数都有因子 $\frac{1}{2k+1}$。那么还剩下甚么要比较?

对于 $n \geq 1$，x^{2n+1} 的系数在 g 中是 $\dfrac{1}{2n+1}$，而在 f 中是

$$\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdots \frac{2n-1}{2n} \cdot \frac{1}{2n+1}$$

。这 n 个因子 $\frac{2r-1}{2r}$ 每一个都严格介乎 0 与 1 之间，而且因为 $n \geq 1$，这样的因子至少有一个。所以它们的乘积严格小于 1，从而

$$\frac{1 \cdot 3 \cdots (2n-1)}{2 \cdot 4 \cdots (2n)} \cdot \frac{1}{2n+1} < \frac{1}{2n+1}.$$

因此 g 中的系数较大，这正是所需的结果。当 $n = 0$ 时，两个系数都等于 1，这就是题目把这个情形排除在外的原因。

验算

- 在 $x = -\frac{1}{2}$ 处，把级数求和至 x^{59} 一项，得 -0.5235987756，与 $\sin^{-1}(-\frac{1}{2}) = -\pi/6$ 吻合。
- 系数 $1, \frac{1}{6}, \frac{3}{40}, \frac{5}{112}$ 与直接由 $\mathrm{f}^{(n+2)}(0) = n^2 \mathrm{f}^{(n)}(0)$ 生成的系数一致。$\frac{1}{6} < \frac{1}{3}$、$\frac{3}{40} < \frac{1}{5}$、$\frac{5}{112} < \frac{1}{7}$ 这几个比较亦如所述成立。
- 由于 f 的每个系数都不超过 g 中对应的系数，对于 $0 < x < 1$ 应有 $\sin^{-1} x < \tanh^{-1} x$。在 $x = 0.5$ 处：$0.5236 < 0.5493$。

常见错误

- 使用莱布尼兹定理时漏掉 $\binom{n}{2}$ 的一项，以致最后的系数变成 n 而不是 n^2。
- 写成 $\mathrm{f}^{(2k+1)}(0) = 1 \cdot 3 \cdots (2k-1)$ 而漏掉平方。递推关系的每一步都是乘以 n^2，而不是 n。
- 计算麦克劳林系数时，只除以 $(2k)!$ 或 $(2k+1)$，而不是除以 $(2k+1)!$。

STEP 之外

比值 $\frac{1\cdot3\cdots(2n-1)}{2\cdot4\cdots(2n)}$ 等于 $\binom{2n}{n}/4^n$，而当 n 很大时，它的表现近似于 $1/\sqrt{\pi n}$。这是沃利斯乘积（Wallis's product）和斯特灵公式（Stirling's formula）的推论。$\sin^{-1} x$ 的级数在 $x = 1$ 处同样收敛，其和为 $\pi/2$，但要证明这一点，需要用到阿贝尔定理（Abel's theorem），该定理处理幂级数（power series）在收敛区间（interval of convergence）端点的情况。当一个函数的各阶导数满足某条微分方程时，便可推出其系数的递推关系；这就是常微分方程（ODE）的级数解法，Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）对此有所讲解。至于泰勒级数（Taylor series）在甚么情况下真正等于原来的函数，严谨的证明可参阅 Abbott 的 *Understanding Analysis*（数学分析入门教材）。

详细解答

STEP III 1994，第 6 题

纯数学 · 阿根图上的旋转

> **STEP III 1994，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/94-s3-q6

关键思路

以点 p 为中心、转角为 θ 的旋转（rotation）就是映射 $z \mapsto p + \mathrm{e}^{\mathrm{i}\theta}(z-p)$：先把 p 移到原点，旋转，再移回去。把若干个这样的映射复合（compose）起来，结果必定形如 $z \mapsto \lambda z + \mu$，而“$P_{\text{最后}} = P_1$ 对每一个 P_1 都成立”就表示这个映射是恒等映射（identity），所以 $\lambda = 1$ 且 $\mu = 0$。余下的工作，就是从几何角度解读这两条等式。

解答

求 z_5

正方向的四分之一圈旋转相当于乘以 i，所以把 P_1 绕 A 旋转便得

$$z_2 = a + \mathrm{i}(z_1 - a) = (1-\mathrm{i})a + \mathrm{i}z_1.$$

同理，$z_3 = (1-\mathrm{i})b + \mathrm{i}z_2$，$z_4 = (1-\mathrm{i})c + \mathrm{i}z_3$，以及 $z_5 = (1-\mathrm{i})d + \mathrm{i}z_4$。逐步代回去，

$$\begin{aligned} z_5 &= (1-\mathrm{i})\big(d + \mathrm{i}c + \mathrm{i}^2 b + \mathrm{i}^3 a\big) + \mathrm{i}^4 z_1 \\ &= (1-\mathrm{i})\big(d - b + \mathrm{i}(c-a)\big) + z_1. \end{aligned}$$

P_5 何时回到 P_1？

四个四分之一圈合起来是一整圈，所以 z_1 的系数恰好是 1，而

$$z_5 - z_1 = (1-\mathrm{i})\big(d - b + \mathrm{i}(c-a)\big),$$

这完全与 z_1 无关。复合映射是一个纯粹的平移（translation）。由于 $1-\mathrm{i}\neq 0$,

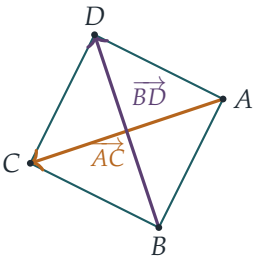
$$\begin{aligned} z_5 = z_1 \text{ 对所有} z_1 &\iff d-b+\mathrm{i}(c-a)=0 \\ &\iff \mathrm{i}(c-a)=b-d \\ &\iff a-c=\mathrm{i}(b-d), \end{aligned}$$

其中最后一步是把两边同乘以 $-\mathrm{i}$（而 $-\mathrm{i}\cdot\mathrm{i}=1$）。每一步都可逆，所以这就是所要求的“若且仅若”（if and only if）。（如果条件不成立，平移量便不是零，于是无论怎样选取 P_1，都有 $P_5\neq P_1$。）

> **想一想** $a-c=\mathrm{i}(b-d)$ 说明了图形 $ABCD$ 有甚么特点?

乘以 -1：条件变成 $c-a=\mathrm{i}(d-b)$。这里 $c-a$ 代表向量 $\overrightarrow{AC}$，而 $d-b$ 代表向量 $\overrightarrow{BD}$。乘以 i 会把向量旋转 $\pi/2$ 而不改变其长度。所以：

$\overrightarrow{AC}$ 是 $\overrightarrow{BD}$ 沿正方向转过一个直角所得；也就是说，四边形 $ABCD$ 的对角线（diagonal）AC 与 BD 长度相等且互相垂直，而 AC 是由 BD 逆时针转四分之一圈而得。



在图中的例子里，$\overrightarrow{BD}=-1+3\mathrm{i}$，而 $\overrightarrow{AC}=-3-\mathrm{i}=\mathrm{i}(-1+3\mathrm{i})$。定向（orientation）是有关系的：顶点按顺时针次序标记的正方形满足这个条件，但按逆时针次序标记的则不满足（那时 $\overrightarrow{AC}=-\mathrm{i}\,\overrightarrow{BD}$）。

三次转角为 θ 的旋转

设 $w=\mathrm{e}^{\mathrm{i}\theta}$。与之前完全一样，每一步都是 $z\mapsto(1-w)p+wz$，所以

$$\begin{aligned} z_4 &= (1-w)c+w\big((1-w)b+w\big((1-w)a+wz_1\big)\big) \\ &= (1-w)\big(c+wb+w^2a\big)+w^3z_1. \end{aligned}$$

这就是 $\lambda z_1+\mu$，其中 $\lambda=w^3$，$\mu=(1-w)(c+wb+w^2a)$。它对所有 z_1 都等于 z_1，若且仅若 $\lambda=1$ 且 $\mu=0$（取 $z_1=0$ 得 $\mu=0$，再取 $z_1=1$ 得 $\lambda=1$）。

转角。$w^3=1$ 表示 3θ 是 2π 的倍数。我们把 $\theta\neq 0$ 理解为“不是整数圈”，所以 $w\neq 1$；因此 w 是一个非实数的三次单位根（cube root of unity），而

$$\theta=\pm\tfrac{2\pi}{3}\pmod{2\pi}.$$

各点。由于 $w \neq 1, \mu = 0$ 变成 $c + wb + w^2 a = 0$。乘以 w 并利用 $w^3 = 1$，得 $a + w^2 b + wc = 0$；又因为 $1 + w + w^2 = 0$（因 $w^3 = 1$，$w \neq 1$），我们可以改写 $a = -(w + w^2)a$，所以

$$w(c-a) + w^2(b-a) = 0 \quad \Longrightarrow \quad c - a = -w\,(b-a) = \mathrm{e}^{\mathrm{i}(\theta+\pi)}(b-a) = \mathrm{e}^{\mathrm{i}(\theta-\pi)}(b-a).$$

> **想一想** 当 $\theta = \pm 2\pi/3$ 时，$\mathrm{e}^{\mathrm{i}(\theta-\pi)}$ 是甚么？

当 $\theta = 2\pi/3$ 时，它是 $\mathrm{e}^{-\mathrm{i}\pi/3}$；当 $\theta = -2\pi/3$ 时，它是 $\mathrm{e}^{\mathrm{i}\pi/3}$（因为 $\mathrm{e}^{-5\pi\mathrm{i}/3} = \mathrm{e}^{\mathrm{i}\pi/3}$）。所以 $\overrightarrow{AC}$ 是 $\overrightarrow{AB}$ 转过 $\mp\pi/3$ 所得，长度不变：ABC 是**等边**（equilateral）三角形，当 $\theta = 2\pi/3$ 时 A, B, C 按**顺时针**次序排列，当 $\theta = -2\pi/3$ 时则按**逆时针**次序排列。反过来，以上每一步都可逆，所以这些条件也能推出 $\mu = 0$。

由于 A, B, C 位于单位圆上，这个圆就是三角形的外接圆（circumcircle），而等边三角形的外心（circumcentre）就是它的中心，所以三个顶点平均分布在 $|z| = 1$ 上。两种情况可以简洁地合并成一句：

$$\theta = \pm\tfrac{2\pi}{3} \pmod{2\pi}, \qquad b = a\,\mathrm{e}^{-\mathrm{i}\theta}, \quad c = b\,\mathrm{e}^{-\mathrm{i}\theta}.$$

用文字来说，三次旋转各转三分之一圈，而 A, B, C 平均分布在单位圆上，其排列方向与旋转方向相反。

验算

- 用电脑随机选取 A, B, C, D，再对随机选取的点施行四次四分之一圈旋转，结果与 z_5 的公式吻合，误差在 10^{-9} 之内；取 $c = a - \mathrm{i}(b-d)$ 时，每个 P_1 都回到原位，而把 c 稍微移动 0.1 之后便不再成立。
- 在单位圆上随机选取 a，并取 $\theta = \pm 2\pi/3$，则 $b = a\mathrm{e}^{-\mathrm{i}\theta}$、$c = b\mathrm{e}^{-\mathrm{i}\theta}$ 对随机的 z_1 都给出 $z_4 = z_1$；把标记次序倒转或改用其他 θ 则不成立。
- 手算：把 $z_1 = 0$ 依次绕顺时针标记的单位正方形 $a = 0$、$b = \mathrm{i}$、$c = 1 + \mathrm{i}$、$d = 1$ 旋转。你会得到 $0 \to 0 \to 1 + \mathrm{i} \to 1 + \mathrm{i} \to 0$。回到原位了。

常见错误

- 把绕 A 的旋转写成 $z \mapsto \mathrm{i}z$，忘记了要先把 A 移到原点。
- 忽略定向：对角线之间的旋转方向，以及 $\theta = \pm 2\pi/3$ 两种情况下三角形各自的排列方向，都是条件的一部分。

STEP 之外

满足 $|\lambda| = 1$ 的映射 $z \mapsto \lambda z + \mu$ 就是平面上保持定向的等距变换（isometry），它们构成一个群（group）。转角可以相加，所以若干个旋转的复合是一个转角为各角总和的旋转；若总和是整数圈，则是一个平移：这就是 $z_5 - z_1$ 结果是常数的原因。Needham 的 *Visual Complex Analysis* 把复数的这种几何观点阐述得十分精彩。

详细解答

STEP III 1994，第 7 题

纯数学 · 群同构

STEP III 1994，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/94-s3-q7

关键思路

两个群（group）要同构（isomorphic），就必须共有同构映射（isomorphism）所保持的每一项结构性质。其中最有用的两项，是各元素的阶（order）的清单，以及该群是否可交换（commutative，即阿贝尔群，abelian）。只要列出各元素的阶，已足以把 S_3 和 Z_6 区分开来。要证明两个群确实同构，我们就写出一个明确的双射（bijection），并验证它与运算相容。最后一部分归结为交换性：凡是由复数组成的群都是阿贝尔群，而 S_3 不是。

解答

S_3 和 Z_6 的元素

把三件物件标记为 $1,2,3$，并采用循环记法（cycle notation），由右至左复合，即 $\sigma\tau$ 表示“先做 τ，再做 σ”。六个置换（permutation）列于下表。

S_3 的元素	e	(12)	(13)	(23)	(123)	(132)
阶	1	2	2	2	3	3

对换（transposition）把两件物件互换，做两次便一切还原，所以它的阶是 2。像 (123) 这样的 3-循环把 $1\to 2\to 3\to 1$，要施行三次才能使每件物件回到原位，所以它的阶是 3。

在 $Z_6=\{0,1,2,3,4,5\}$ 中，单位元（identity）是 0，而 a 的阶是满足 $na\equiv 0 \pmod 6$ 的最小 $n\geq 1$，即 $6/\gcd(a,6)$。

Z_6 的元素	0	1	2	3	4	5
阶	1	6	3	2	3	6

例如，$4+4=8\equiv 2$ 而 $4+4+4=12\equiv 0$，所以 4 的阶是 3。

S_3 和 Z_6 同构吗?

不同构。同构映射 θ 保持元素的阶：若 g 的阶是 n，则 $\theta(g)^k=\theta(g^k)$，而由于 θ 是单射（injective），这个元素是单位元当且仅当 g^k 是单位元。因此，同构的群中，每一种阶的元素数目都相同。但 Z_6 有两个阶为 6 的元素，S_3 却一个也没有，所以两者并不同构。（两者在阶为 2 的元素上也有分别：S_3 有三个，Z_6 只有一个。）

还有第二个理由，我们在最后会再用到。Z_6 是阿贝尔群，因为加法可交换。S_3 则不是：

$$(12)(13)=(132),\qquad (13)(12)=(123).$$

就第一个乘积而言，(13) 把 $1\to 3$，然后 (12) 保持 3 不变，所以 $1\to 3$。同理 $3\to 2$ 及 $2\to 1$。同构映射把可交换的一对元素映到可交换的一对元素，所以非阿贝尔群不可能与阿贝尔群同构。

C_6 与 Z_6 同构

定义 $\phi: Z_6\to C_6$ 为 $\phi(k)=\alpha^k$，其中 $k=0,1,\dots,5$。

- 双射。当 $k=0,\dots,5$ 时，各数 $\alpha^k=\mathrm{e}^{\mathrm{i}k\pi/3}$ 的辐角（argument）为 $0,\frac{\pi}{3},\dots,\frac{5\pi}{3}$。这些是 $[0,2\pi)$ 内互不相同的角，所以这六个值是单位圆上互不相同的点。因此 ϕ 是单射，从而是两个大小同为 6 的集合之间的双射。
- 运算。设 $a,b\in Z_6$，并写 $a+b=6q+r$，其中 $r=(a+b)\bmod 6$。由于 $\alpha^6=\mathrm{e}^{2\pi\mathrm{i}}=1$，

$$\phi(a)\,\phi(b)=\alpha^a\alpha^b=\alpha^{a+b}=(\alpha^6)^q\alpha^r=\alpha^r=\phi\big((a+b)\bmod 6\big).$$

所以 ϕ 把模 6 加法变成复数乘法，它是一个同构映射。因此 $C_6\cong Z_6$。作为核对，α 和 α^5（即 1 和 5 的像）的阶是 6，α^2 和 α^4 的阶是 3，而 $\alpha^3=-1$ 的阶是 2，与 Z_6 的表吻合。

C 有与 S_3 同构的子群吗?

> **想一想** 所有由复数组成的群都共有甚么性质?

它们全都是阿贝尔群。$(\mathbb{C},+)$ 的子群（subgroup）H 是阿贝尔群，因为对所有复数 z,w 都有 $z+w=w+z$。非零复数在乘法下所成的群的子群 H 是阿贝尔群，因为 $zw=wz$。子群沿用大群的同一运算，所以继承了交换性。

现在假设某个这样的 H 有同构映射 $\theta: S_3\to H$。把它应用于上面那对不可交换的元素。由于 H 是阿贝尔群，

$$\theta\big((12)(13)\big)=\theta(12)\,\theta(13)=\theta(13)\,\theta(12)=\theta\big((13)(12)\big),$$

这里以乘法形式书写（若是加法的情况，把乘积换成和即可）。因为 θ 是单射，由此会得出 $(12)(13) = (13)(12)$，而这是错的。所以复数的子群，不论是加法的还是乘法的，都**没有**一个与 S_3 同构。

验算

- 在每个群中，元素的阶都整除 6（拉格朗日定理，Lagrange's theorem）：S_3 中是 $1, 2, 3$，Z_6 中是 $1, 2, 3, 6$。
- 我们用电脑重新计算了 S_3、Z_6 和 C_6 中每个元素的阶，确认了 $(12)(13) \neq (13)(12)$，并对全部 36 对 (a, b) 以数值方法检验了 $\alpha^{(a+b) \bmod 6} = \alpha^a \alpha^b$，误差在 10^{-12} 以内。

常见错误

- 说“两者都有六个元素，所以同构”。大小相同是必要条件，但远远不是充分条件。
- 说 S_3 中阶为 3 的元素是 (12) 和 (13)。对换的阶是 2；只有 3-循环的阶才是 3。
- 在最后一部分，只排除了有限的乘法子群（单位根，roots of unity）。交换性的论证一次过涵盖所有子群，不论有限还是无限、加法还是乘法。

STEP 之外

这道题讲的是不变量（invariant）：同构的群必须共有的性质，例如元素的阶和交换性。事实上，每个阶为 6 的群都与 Z_6 或 S_3 其中之一同构。更一般地，非零复数在乘法下所成的群的任何有限子群都是循环群（cyclic group），所以它是一个由单位根组成的群。这两个结果，你在第一门群论课程的初段就会遇到。Allenby 的 *Rings, Fields and Groups*（环、域与群）写得平易近人，很适合用来阅读这些内容。

详细解答

STEP III 1994，第 8 题

纯数学 · 四平方和恒等式

STEP III 1994，第 8 题

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

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

本题在 *Beyond STEP* 第 7.8 节四个平方数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch07/four-squares.html

关键思路

形如 $\begin{pmatrix} z & w \\ -w^* & z^* \end{pmatrix}$ 的矩阵 (matrix), 其行列式 (determinant) 为 $zz^* + ww^* = |z|^2 + |w|^2$, 是四个实数平方之和。若把 a, b, c, d 装进一个这样的矩阵，再把 p, q, r, s 装进另一个，则由 $\det(AB) = \det A \det B$ 便得到左边。要做的工夫是证明乘积 AB 具有同样的特殊形式，因为这样它的行列式也是四个平方之和，而这些平方正是 L_i^2。

解答

准备工作

对复数（complex number）z 和 w，记

$$M(z,w) = \begin{pmatrix} z & w \\ -w^* & z^* \end{pmatrix}, \qquad \det M(z,w) = zz^* - w(-w^*) = |z|^2 + |w|^2.$$

取

$$z_1 = a + b\mathrm{i}, \quad z_2 = c + d\mathrm{i}, \quad z_3 = p + q\mathrm{i}, \quad z_4 = r + s\mathrm{i}.$$

则

$$\det M(z_1, z_2) = a^2 + b^2 + c^2 + d^2, \qquad \det M(z_3, z_4) = p^2 + q^2 + r^2 + s^2.$$

乘积具有相同的形式

乘开得

$$M(z_1,z_2)\,M(z_3,z_4)=\begin{pmatrix} z_1z_3-z_2z_4^* & z_1z_4+z_2z_3^* \\ -z_2^*z_3-z_1^*z_4^* & -z_2^*z_4+z_1^*z_3^* \end{pmatrix}.$$

设 $w_1=z_1z_3-z_2z_4^*$ 及 $w_2=z_1z_4+z_2z_3^*$，即第一行的两个元素。由于乘积的共轭（conjugate）等于共轭的乘积，而且 $(z^*)^*=z$，

$$w_1^*=z_1^*z_3^*-z_2^*z_4,\qquad -w_2^*=-z_1^*z_4^*-z_2^*z_3,$$

而这两个正好就是右下角和左下角的元素。所以

$$M(z_1,z_2)\,M(z_3,z_4)=M(w_1,w_2).$$

想一想 乘积具有这个特殊形状，为甚么这样重要?

对一般的 2×2 复矩阵而言，行列式根本不是平方和。正因为乘积仍然是 $M(w_1,w_2)$ 的形式，它的行列式才是 $|w_1|^2+|w_2|^2$，从而可拆成四个实数平方。

取行列式

利用 $\det(AB)=\det A\,\det B$，

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

余下只须求出 w_1 和 w_2 的实部（real part）和虚部（imaginary part）：

$$\begin{aligned} w_1&=(a+b\mathrm{i})(p+q\mathrm{i})-(c+d\mathrm{i})(r-s\mathrm{i})\\ &=\big[(ap-bq)+(aq+bp)\mathrm{i}\big]-\big[(cr+ds)+(dr-cs)\mathrm{i}\big]\\ &=(ap-bq-cr-ds)+(aq+bp+cs-dr)\mathrm{i},\end{aligned}$$

$$\begin{aligned} w_2&=(a+b\mathrm{i})(r+s\mathrm{i})+(c+d\mathrm{i})(p-q\mathrm{i})\\ &=\big[(ar-bs)+(as+br)\mathrm{i}\big]+\big[(cp+dq)+(dp-cq)\mathrm{i}\big]\\ &=(ar-bs+cp+dq)+(as+br-cq+dp)\mathrm{i}.\end{aligned}$$

因此可取

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

于是 $|w_1|^2+|w_2|^2=L_1^2+L_2^2+L_3^2+L_4^2$，得出

$$(a^2+b^2+c^2+d^2)(p^2+q^2+r^2+s^2)=L_1^2+L_2^2+L_3^2+L_4^2.$$

每个 L_i 都是若干项之和，每项的形式为（a,b,c,d 中的一个字母）$\times$（p,q,r,s 中的一个字母），所以它对两组变量分别都是线性（linear）的，符合题目要求。（答案并不唯一：改变任何 L_i 的正负号，或以不同方式选取 z_j，都会得出其他合用的组合。）

验算

- 取 $(a,b,c,d)=(1,2,3,4)$ 及 $(p,q,r,s)=(5,6,7,8)$。左边为 $30\times 174=5220$。由公式得 $(L_1,L_2,L_3,L_4)=(-60,12,30,24)$，而 $3600+144+900+576=5220$。
- 设 $c=d=r=s=0$。则 $L_3=L_4=0$，恒等式（identity）简化为 $(a^2+b^2)(p^2+q^2)=(ap-bq)^2+(aq+bp)^2$，即由 $|z_1z_3|=|z_1||z_3|$ 得出的二平方和恒等式。
- 电脑验算已就八个变量的 20,000 组随机有理数取值，确认恒等式精确成立，并以数值方式确认了 $M(z_1,z_2)M(z_3,z_4)=M(w_1,w_2)$。

常见错误

- 相乘时忘记第二行的共轭，以致破坏了乘积的特殊形式。
- $(c+d\mathrm{i})(r-s\mathrm{i})$ 中的正负号出错：$-ds\,\mathrm{i}^2$ 一项会变成 $+ds$。

STEP 之外

这就是欧拉（Euler）四平方和恒等式。矩阵 $M(z,w)$ 是哈密顿（Hamilton）四元数（quaternion）的一个具体模型：它们在乘法下封闭（closed），而行列式就是“长度”的平方，具有可乘性（multiplicative）。这个恒等式是拉格朗日（Lagrange）定理（每个正整数都是四个平方之和）证明中的关键一步，因为它把问题归结为质数（prime）的情况。Davenport 的 *The Higher Arithmetic*（一本数论入门经典）涵盖平方和的课题，而 Needham 的 *Visual Complex Analysis*（以几何观点讲解复分析）则为二平方和的情况提供几何直观。

详细解答

STEP III 1994，第 9 题

力学 · 等时曲线与碰撞

STEP III 1994，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/94-s3-q9

本题在 *Beyond STEP* 第 11.4 节等时曲线中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/tautochrone.html

关键思路

沿曲面的能量守恒（energy conservation）给出一个首次积分（first integral），而碗的形状 $y = s^2/(4k)$ 把它化为以弧长（arc-length）s 表示的简谐运动（simple harmonic motion）能量方程。简谐运动的周期与振幅无关，所以每个由静止释放的质点，无论从哪里出发，都在相同的时间后到达 V。正因如此，最后一部分的两个质点必定在 V 相遇，而且一次又一次地相遇。

解答

常数

我们把质点看作光滑曲面上的一个质点模型（point mass）。作用在它身上的力有两个：重量和法向反作用力（normal reaction）。质点由静止出发，所以它的速度一开始并没有垂直于“通过起点和轴的铅垂平面”的水平分量。两个力都没有垂直于该平面的分量（轴对称曲面的反作用力位于通过轴的平面之内），所以质点一直留在这个铅垂截面内。它沿曲线 $s = 2\sqrt{ky}$ 运动，速率为 $|\mathrm{d}s/\mathrm{d}t|$。

曲面是光滑的，所以法向反作用力与速度垂直，不作功。只有重力作功，因此机械能守恒：

$$\tfrac{1}{2}m\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 + mgy = \text{常数}.$$

乘以 $2/m$ 可见 $\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 + 2gy$ 是常数，正如所求。质点在高度 h 由静止释放，故该常数为 $2gh$。

到达 V 所需的时间

在曲线上，$y = s^2/(4k)$。代入后得

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 + \frac{g}{2k}s^2 = 2gh.$$

对 t 求导，再除以 $2\,\mathrm{d}s/\mathrm{d}t$（质点在运动时此步成立），得

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t^2} = -\frac{g}{2k}s,$$

这是关于 s 的简谐运动，其中 $\omega^2 = \frac{g}{2k}$，中心在 $s = 0$，即点 V。

想一想 为甚么由释放到抵达 V 的时间是四分之一个周期?

质点由静止从端点位置 $s_0 = 2\sqrt{kh}$ 出发，所以 $s = s_0 \cos \omega t$。当 $\omega t = \frac{\pi}{2}$ 时它首次到达 $s = 0$，这是周期 $2\pi/\omega$ 的四分之一。所以到达 V 所需的时间是

$$T = \frac{\pi}{2\omega} = \frac{\pi}{2}\sqrt{\frac{2k}{g}} = \pi\sqrt{\frac{k}{2g}},$$

正如所求。这个结果与 h 无关：这个碗是一条等时曲线（tautochrone）。

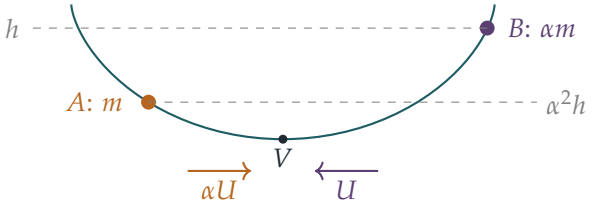
为方便稍后使用，质点在 V 的速率是 ωs_0；或者直接在能量方程中取 $y = 0$，得 $\sqrt{2gh}$。

两个质点

把两个质点看作在 V 正面碰撞的质点。该处的切线是水平的，所以碰撞沿一条水平线发生。把质量为 m 的质点称为质点 A（在高度 $\alpha^2 h$ 释放），把质量为 αm 的质点称为质点 B（在另一侧的高度 h 释放）。

它们在 V 相遇。根据上一部分，每个质点不论起始高度如何，都在时间 T 后到达 V。它们同时释放，所以同时到达 V，并在时间 T 于该处碰撞。

碰撞前一刻的速率。记 $U = \sqrt{2gh}$。由能量方程，A 到达时的速率是 $\sqrt{2g\alpha^2 h} = \alpha U$，$B$ 到达时的速率是 U，两者运动方向相反。取 A 的运动方向为正。



碰撞前的动量是

$$m(\alpha U) + \alpha m(-U) = 0.$$

这就是题目如此选取质量和高度的原因。

第一次碰撞之后。设碰撞后的速度为 v_A 和 v_B。由动量守恒（conservation of momentum）和牛顿实验定律（Newton's experimental law）得

$$mv_A + \alpha m\, v_B = 0, \qquad v_B - v_A = \alpha\big(\alpha U - (-U)\big) = \alpha(1+\alpha)U.$$

由第一条方程得 $v_B = -v_A/\alpha$。代入第二条方程得 $-v_A(1+\alpha)/\alpha = \alpha(1+\alpha)U$，所以

$$v_A = -\alpha^2 U, \qquad v_B = \alpha U.$$

两个速度都反向，并乘以 α。（总动量为零时，相对速度和每个质点各自的速度全都乘以 $-\alpha$。）所以每个质点都沿碗中自己那一侧往回上升。由能量守恒，A 升至高度 $\alpha^4 h$，B 升至高度 $\alpha^2 h$。

规律重复出现。两个质点在同一时刻，即碰撞后时间 T，各自在新的最高点静止下来。然后各自再用时间 T 返回 V。所以它们在时间 $3T$ 再次于 V 碰撞，此时 A 来自高度 $\alpha^4 h$，B 来自高度 $\alpha^2 h$。这正是把 h 换成 $\alpha^2 h$ 之后的初始状况。总动量再次为零，两个速度也再次反向并乘以 α。

由归纳法（induction），第 n 次碰撞在时间 $(2n-1)T$ 于 V 发生，而在这次碰撞之后

$$A \text{ 升至高度} \alpha^{2n+2} h, \qquad B \text{ 升至高度} \alpha^{2n} h,$$

各自留在碗中它出发的那一侧。两个质点在各自的一侧振荡，并在等间隔的时刻 $T, 3T, 5T, \ldots$ 于 V 碰撞。每次碰撞时速率都按因子 α 减小，所以高度按因子 α^2 减小，动能损失的比例为 $1-\alpha^2$。碰撞永不停止，因为相邻两次碰撞的间隔总是 $2T$。振幅按等比方式递减，当 $t \to \infty$ 时两个质点都趋向静止于 V，但不会在有限时间内真正静止在那里。

验算

- 量纲（dimensions）：k 是长度（因为 $s^2 = 4ky$），所以 $\sqrt{k/g}$ 是时间。
- 数值方面,我们对随机选取的 k 和 h 以数值求积法(quadrature)计算了 $\int_0^{s_0} \mathrm{d}s/\sqrt{2g(h-y)}$。取 $k=1$、$g=9.81$ 时，结果是 $0.7092\ldots = \pi\sqrt{1/19.62}$，与 h 无关。
- 我们以数值方法逐步计算两个质点的运动（取 $\alpha = 0.6$ 和 $\alpha = 0.3$），并在每次相遇时应用碰撞定律。碰撞发生在 $t/T = 1, 3, 5, 7$，每次都在 V，而且总动量为零。取 $\alpha = 0.6$ 时，第一次碰撞后的高度是 $0.1296h = \alpha^4 h$ 和 $0.36h = \alpha^2 h$。
- 极限情况 $\alpha \to 1$：质量相等、高度相等，而且是完全弹性碰撞（perfectly elastic），所以每个质点都永远反弹回到起始高度，与预期相符。

常见错误

- 以为 V 距离起点半个周期：质点从振荡的端点出发，所以到 V 只需四分之一个周期。
- 在牛顿定律中忘记方向反转：要使用带正负号的速度，而不只是速率。
- 声称质点在有限次碰撞后静止下来。碰撞之间的时间间隔是固定的，并不会缩短（这与球在地板上弹跳不同），所以有无限多次碰撞，分布在无限长的时间之内。

STEP 之外

内禀方程（intrinsic equation）$s^2 = 4ky$ 描述的是一条摆线（cycloid）。这道题重新发现了惠更斯（Huygens）的等时曲线：在倒置的摆线上，滑到最低点所需的时间与起点无关。惠更斯利用这一点设计出周期与摆幅无关的摆钟。同一条曲线也是最速降线（brachistochrone），即下降最快的曲线，它可用变分法（calculus of variations）求得，而变分法是大学二年级力学的标准课题。想进一步了解振荡和碰撞，可参阅 Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP III 1994，第 10 题

力学 · 相对速度

STEP III 1994，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：

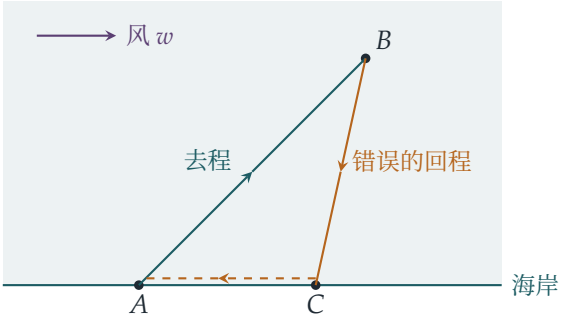
step.maths.org/questions/94-s3-q10

关键思路

一切都是向量加法（vector addition）：飞机相对地面的速度，等于它相对空气的速度加上风的速度。把海岸放在 x 轴上，以 Alphaport 为原点，并把去程的空速（air velocity）写成分量（components）。把航向（heading）反过来，空速会反向，但风不会反向，因此回程时，运动的向北部分和向东部分表现截然不同，而这正是第一部分的全部要点。至于第二部分，两个正确的地速（ground speed）来自同一个二次方程（quadratic），于是韦达定理（Vieta's formulas）就能代劳。

解答

我们把飞机看成一个质点（particle），它相对空气以大小为 v 的恒定速度运动，而空气则以恒定速度 w 向正东移动（来自西面的风是吹向东面的）。取 Alphaport A 为原点，x 轴向东，y 轴向北，于是海岸就是直线 $y = 0$，而岛屿位于 $y > 0$ 的一侧。Betaville B 位于 $\left(\frac{d}{\sqrt{2}}, \frac{d}{\sqrt{2}}\right)$。速率以 $\mathrm{km\,h^{-1}}$ 为单位，时间以小时为单位。（试卷上把这两个地方称为 Alphabet/Alphaport 和 Betaville/Betaport；它们指的是同样的地方。）



去程

设去程的空速为 (a, b)，则 $a^2 + b^2 = v^2$。相对地面的速度是 $(a + w,\ b)$，它必须指向东北，所以两个分量相等而且为正：

$$a + w = b > 0.$$

她以 y 方向的速率 b 飞过 y 方向的距离 $\frac{d}{\sqrt{2}}$，而去程所需时间为 t，所以

$$bt = \frac{d}{\sqrt{2}} = (a + w)t. \tag{1}$$

错误的回程

采用相反的航向时，她的空速是 $(-a, -b)$，但风仍然加上 $(w, 0)$，所以她相对地面的速度是

$$(w - a,\ -b).$$

? **想一想** 她会在哪里到达海岸？在 Alphaport 的哪一边？

向北方向的运动不受风影响：她的 y 坐标由 $y = \frac{d}{\sqrt{2}}$ 起以速率 b 递减，所以由 (1) 可知，她恰好在时间 t 之后到达 $y = 0$，与去程所需的时间相同。在这段时间内，她的 x 坐标改变了 $(w - a)t$，所以她越过海岸的位置是

$$x = \frac{d}{\sqrt{2}} + (w - a)t = (a + w)t + (w - a)t = 2wt,$$

这里再次用到 (1)。因此她降落在 C，位于 Alphaport 以东 $2wt$ 处：去程时，风把她推向东面，而她选定的航向正是为了抵销这个影响；但回程时，航向所作的修正反了过来，风的推力却没有反过来，所以偏移加倍。由于 $2wt > 0$，沿海岸转向西面的确是正确的方向。她以空速 v 逆风向正西飞行，地速为 $v - w > 0$（因为 $v > w$）。沿海岸的一段需时 $\dfrac{2wt}{v - w}$，所以整个回程需时

$$t + \frac{2wt}{v - w} = \frac{(v - w) + 2w}{v - w}\, t = \left(\frac{v + w}{v - w}\right) t \text{ 小时},$$

正如所求。

正确的回程

把沿直线 AB 的相对地面速度写成 $\pm(p, p)$，其中 $\sqrt{2}\, p$ 是地速。空速等于相对地面的速度减去风速，而且大小为 v。

去程时，相对地面的速度为 (p_1, p_1)，空速为 $(p_1 - w,\ p_1)$：

$$(p_1 - w)^2 + p_1^2 = v^2 \iff 2p_1^2 - 2wp_1 + (w^2 - v^2) = 0.$$

回程时，相对地面的速度为 $(-p_2,-p_2)$，其中 $p_2>0$，空速为 $(-p_2-w,\,-p_2)$：

$$(p_2+w)^2+p_2^2=v^2 \iff 2(-p_2)^2-2w(-p_2)+(w^2-v^2)=0.$$

所以 p_1 和 $-p_2$ 都是同一个二次方程 $2p^2-2wp+(w^2-v^2)=0$ 的根。它的两根之积为 $\frac{w^2-v^2}{2}<0$，所以一个根为正，一个根为负；p_1 是正的那个，$-p_2$ 是负的那个，因此它们就是两个相异的根。根据韦达定理，

$$p_1+(-p_2)=w,\qquad p_1(-p_2)=\frac{w^2-v^2}{2},$$

即 $p_1-p_2=w$ 及 $p_1p_2=\frac{1}{2}(v^2-w^2)$。

来回的距离都是 d，所以去程所需时间为 $t=\dfrac{d}{\sqrt{2}\,p_1}$，而正确回程所需时间为 $\dfrac{d}{\sqrt{2}\,p_2}$。因此

$$\begin{aligned}\frac{d}{\sqrt{2}\,p_2}&=t+\frac{d}{\sqrt{2}}\left(\frac{1}{p_2}-\frac{1}{p_1}\right)=t+\frac{d}{\sqrt{2}}\cdot\frac{p_1-p_2}{p_1p_2}\\&=t+\frac{d}{\sqrt{2}}\cdot\frac{2w}{v^2-w^2}=t+\frac{\sqrt{2}\,wd}{v^2-w^2}\ \text{小时},\end{aligned}$$

这就是所求的结果。

验算

- 若 $w=0$，两个回程时间都化为 t，在无风的情况下理应如此。
- 取 $v=300$、$w=60$、$d=400$：$t\approx 1.1785\,\mathrm{h}$，错误的回程（以数值方法逐步模拟运动而求得）需时 $1.7678\,\mathrm{h}=\frac{360}{240}t$，而正确的回程需时 $1.5713\,\mathrm{h}$，与 $t+\frac{\sqrt{2}\cdot 60\cdot 400}{86400}$ 吻合。对许多组 $v>w$ 和 d 作随机试算，结果在舍入误差范围内一致。
- 错误的回程总是比正确的回程慢：这就是不等式 $\frac{2wt}{v-w}>\frac{\sqrt{2}\,wd}{v^2-w^2}$，即 $\sqrt{2}(v+w)t>d$；它之所以成立，是因为去程的地速 $\frac{d}{t}=\sqrt{2}p_1=\sqrt{2}(p_2+w)$ 小于 $\sqrt{2}(v+w)$，原因是 $p_2<v$。

常见错误

- 把相对地面的速度反向，而不是把空速反向：飞行员掉头时，风并不会跟着掉头。
- 以为风是吹向西面的。“来自西面”的意思是风把飞机带向东面。
- 用求根公式逐一解两个二次方程，再把繁复根式（surds）的倒数相减；只要看出 p_1 和 $-p_2$ 是同一个二次方程的根，就可以完全避开这些计算。

STEP 之外

这是相对速度（relative velocity）最简单的形式：飞机在地面参考系（frame）中的速度，等于它在空气参考系中的速度加上空气的速度，也就是一次伽利略（Galilean）参考系变换。同一套记帐方法，只要做得仔细，就是导航、大学一年级力学中的渡河问题，以及（当它在高速下失效时）狭义相对论（special relativity）的动机的基础。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对运动参考系有详尽的论述。

详细解答

STEP III 1994，第 11 题

力学 • 刚体的平衡

STEP III 1994，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/94-s3-q11

关键思路

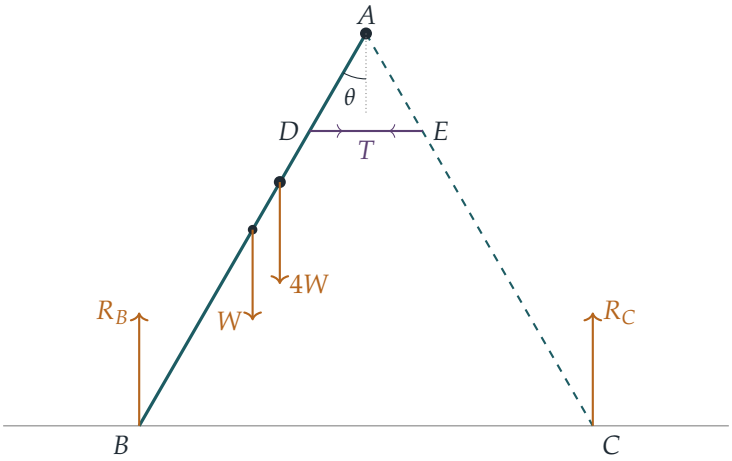
由于地面光滑（smooth），地面施加的力只能沿铅垂方向，所以把整座梯子视为一个整体，便可立即求出 C 处的反作用力（reaction）。接下来，巧妙之处全在轻质的梯脚 AC：作用于它的力只有三个（地面的反作用力、绳的张力和铰链的力），因此对铰链 A 取矩（taking moments），便可消去未知的铰链力，并把绳的张力（tension）与 C 处的反作用力联系起来。虎克定律（Hooke's law）把张力与梯子的角度连在一起，由此可把 β 表示成该角度的函数，然后求其最大值。

解答

建立模型

我们把 AB 视为均匀（uniform）杆，把 AC 视为轻质（light）杆，两者长度均为 $4a$，在 A 处以光滑铰链（hinge）相连，并把绳视为轻质弹性绳。那人则视为质点（particle）。由对称性（$AD = AE$ 及 $AB = AC$），绳 DE 是水平的。设每条梯脚与铅垂线的夹角为 θ，其中 $0 < \theta < \frac{\pi}{2}$。

由于地面光滑，B 和 C 处的反作用力 R_B 和 R_C 都沿铅垂方向。AB 的重量 W 作用于其中点，与 B 的水平距离为 $2a\sin\theta$；那人的重量 $4W$ 的作用点与 B 的水平距离为 $\beta a\sin\theta$。张力 T 把 D 和 E 互相拉近。



C 处的反作用力

对整座折梯以 B 为支点取矩（铰链力和张力都是内力，所以不会出现），其中 $BC = 8a\sin\theta$：

$$R_C \cdot 8a\sin\theta = W \cdot 2a\sin\theta + 4W \cdot \beta a\sin\theta \quad\Longrightarrow\quad R_C = \tfrac{1}{4}W(1+2\beta).$$

沿铅垂方向分解力，得 $R_B = 5W - R_C = \frac{1}{4}W(19-2\beta)$；由于 $\beta \le 4$（梯级在梯子上），此值为正。所以梯脚永远不会离开地面。

绳的张力

绳的长度为 $DE = 2a\sin\theta$，自然长度（natural length）为 $\frac{a}{4}$。假如绳是松弛的，则 $T = 0$，而下文对 AC 以 A 为支点取矩便会迫使 $R_C = 0$，这是不可能的。所以绳处于拉伸状态，$\sin\theta > \frac{1}{8}$，而由虎克定律得

$$T = \frac{W\left(2a\sin\theta - \frac{a}{4}\right)}{a/4} = W(8\sin\theta - 1).$$

轻质梯脚 AC 的平衡

> ? **想一想** 为甚么要单独考虑 AC，而不是 AB?

梯脚 AC 是轻质的，所以作用于它的力只有三个：C 处的 R_C、E 处的张力 T，以及 A 处的铰链力。对 A 取矩便可消去未知的铰链力。张力沿水平方向，作用点在 A 之下、铅垂深度为 $a\cos\theta$ 之处；反作用力 R_C 沿铅垂方向，作用点与 A 的水平距离为 $4a\sin\theta$。所以

$$R_C \cdot 4a\sin\theta = T \cdot a\cos\theta.$$

代入 R_C 和 T：

$$W(1+2\beta)\sin\theta = W(8\sin\theta-1)\cos\theta,$$
$$\text{所以}\quad 1+2\beta = 8\cos\theta-\cot\theta,$$
$$\text{而}\quad \beta = \tfrac{1}{2}\left(8\cos\theta-\cot\theta-1\right).$$

对于那人的每个位置 β，平衡（equilibrium）可以成立，当且仅当这条方程有解 θ（且 $\sin\theta > \frac{1}{8}$）。所以 β 的最大可能值，由 $f(\theta) = 8\cos\theta - \cot\theta$ 的最大值决定。

求最大值

求导数，得

$$f'(\theta) = -8\sin\theta + \operatorname{cosec}^2\theta = \frac{1-8\sin^3\theta}{\sin^2\theta}.$$

由于 $\sin\theta$ 在 $(0, \frac{\pi}{2})$ 上递增，当 $\sin\theta < \frac{1}{2}$ 时上式为正，当 $\sin\theta > \frac{1}{2}$ 时上式为负。所以 f 在 $\sin\theta = \frac{1}{2}$（即 $\theta = \frac{\pi}{6}$）时取得最大值，此时

$$f\left(\tfrac{\pi}{6}\right) = 8\cdot\tfrac{\sqrt{3}}{2} - \sqrt{3} = 3\sqrt{3}, \qquad \beta_{\max} = \tfrac{1}{2}\left(3\sqrt{3}-1\right) \approx 2.10.$$

此值小于 4，所以梯级确实在梯子上；而且 $\sin\theta = \frac{1}{2} > \frac{1}{8}$，所以绳是拉紧的，与我们的假设相符。

梯级的高度

梯级位于由 B 起沿 BA 方向距离 βa 之处，而 BA 与铅垂线的夹角为 θ，所以它的高度为

$$h = \beta a\cos\theta = \tfrac{1}{2}\left(3\sqrt{3}-1\right)a\cdot\tfrac{\sqrt{3}}{2} = \frac{(9-\sqrt{3})\,a}{4} \approx 1.82a.$$

验算

- 当 $\theta = \frac{\pi}{6}$ 并取 $W = a = 1$ 时：$T = 8\cdot\frac{1}{2} - 1 = 3$，$R_C = \frac{1}{4}(1+3\sqrt{3}-1) = \frac{3\sqrt{3}}{4} \approx 1.299$，而确实有 $R_C\cdot 4\sin\theta = 2.598 = T\cos\theta$。
- 以数值方法验算：分别对两条梯脚解出力和力矩的方程，并让 θ 扫遍 $(0, \frac{\pi}{2})$，结果发现 β 的最大值为 2.09808，出现于 $\theta \approx 30°$，对应的高度为 $1.81699a$。
- 极限情况也合理：当 $\sin\theta \to \frac{1}{8}$ 时，绳变得松弛，β 便须为负数；当 $\theta \to \frac{\pi}{2}$ 时，绳的力臂 $a\cos\theta$ 趋于零，所以 β 同样必须下降。最大值出现在两者之间，正如所料。

常见错误

- 以为 A 处的铰链力是水平的。一般而言并非如此，因为两条梯脚所受的负荷不同；对 A 取矩，便完全毋须用到这个力。
- 把整段长度 $DE = 2a\sin\theta$ 当作伸长量（extension）。伸长量应为 $2a\sin\theta - \frac{a}{4}$。
- 答案写成 $\beta_{\max}$（沿梯子量度的距离），而不是题目所问的高度；求高度还须多乘一个因子 $a\cos\theta$。

STEP 之外

这是一道静力学（statics）题目，其中的位形（configuration，即 θ）并非已知，而是由弹性绳决定，所以各个平衡位置构成一个单参数族，而问题便变成在这个族上求最优化（optimisation）。在大学里，这正是虚功（virtual work）和势能（potential energy）方法背后的想法：平衡位置就是势能的驻点（stationary point），而追问哪些平衡是稳定的，便会引出分歧（bifurcation）理论和突变理论（catastrophe theory）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）是深入学习力学的好起点。

详细解答

STEP III 1994，第 12 题

概率与统计 · 首步分析

> **STEP III 1994，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/94-s3-q12

关键思路

把各局两局两局地成对来看。领先 2 局的情况只可能在打完偶数局之后出现，而每打完一对局，两名球员各发球一次，于是发球权又回到原来的一方。这样，整场比赛就变成一个简单的重复试验：每一对局不是由 A 全胜，就是由 B 全胜，又或者各胜一局，而各胜一局的话，双方便回到平手的局面。标准网球其实是同一回事，只是头两对局要特别处理。

解答

全篇记

$$a = p_A p_B, \qquad b = q_A q_B, \qquad s = p_A q_B + q_A p_B.$$

在连续的一对局中（两名球员各发球一局），a 是 A 两局全胜的概率（probability），b 是 B 两局全胜的概率，s 是两人各胜一局的概率（或是 A 保住发球局而 B 也保住发球局，或是双方都破发）。我们按题目的原意，假设不同局的结果互相独立（independent）。结果只有这三种，所以

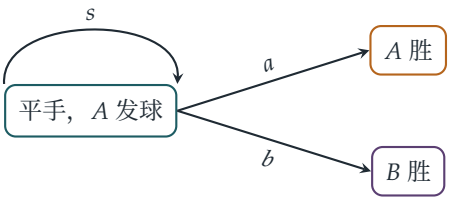
$$a + b + s = (p_A + q_A)(p_B + q_B) = 1.$$

简化网球

每打一局，比分之差就改变 ± 1，所以打完偶数局后分差是偶数，打完奇数局后分差是奇数。因此，恰好领先 2 局的情况，最早只可能在第 $2, 4, 6, \ldots$ 局之后出现；而如果打完偶数局后比赛仍未结束，双方必定是平手。

于是把比赛分成一对一对（第 1–2 局、第 3–4 局、…）。由平手、轮到 A 发球的局面开始，

下一对局有概率 s 令比赛以 A 胜出告终，有概率 b 令比赛以 B 胜出告终，另有概率 s 令我们回到平手、又轮到 A 发球的局面。这时的情况与开始时完全一样，所以



$$P_{\text{short}} = a + sP_{\text{short}} \quad \Longrightarrow \quad P_{\text{short}} = \frac{a}{1-s} = \frac{a}{a+b} = \frac{p_A p_B}{p_A p_B + q_A q_B},$$

其中用到 $1 - s = a + b$。（除以 $1 - s$ 并无问题，除非 $s = 1$；那是每一对局都各胜一局、比赛永远不会结束的退化情况。）

如果由 B 先发球，每一对局就是先打一局 B 的发球局，再打一局 A 的发球局。A 两局全胜、B 两局全胜、两人各胜一局的概率分别是 $p_B p_A$、$q_B q_A$ 和 $q_B p_A + p_B q_A$：仍然是同样的 a、b、s，因为一对局之中两局的先后次序并无影响。同样的论证给出同样的答案，所以 P_{short} 与谁先发球无关。

4 局内分出胜负

未打完 4 局之前，比赛不可能分出胜负。打完 4 局后（即两对局，由 A 先发球），分差是偶数，而比赛分出胜负，当且仅当比分是 4–0、3–1、1–3 或 0–4。用两对局来表示：

- A 以 4–0 胜出：A 两对局全胜，概率为 $a^2 = p_A^2 p_B^2$；
- 0–4：B 两对局全胜，概率为 $b^2 = q_A^2 q_B^2$；
- A 以 3–1 胜出：A 全胜其中一对局，另一对局各胜一局，两种次序都可以，概率为 $2as$；
- 1–3：同理，概率为 $2bs$。

相加得

$$\Pr(4 \text{ 局内分出胜负}) = p_A^2 p_B^2 + q_A^2 q_B^2 + 2(p_A p_B + q_A q_B)(p_A q_B + q_A p_B),$$

正是所求。

概率相等时的标准网球

如果比赛在 4 局内未分胜负，比分就是 2–2。这种情况的概率是 $s^2 + 2ab$（两对局都各胜一局，或两名球员各全胜一对局，两种次序都可以）。由 2–2 开始，“领先 2 局”的规则正好就是简化网球由平手开始的规则，而第 5 局由 A 发球；所以由这里开始，A 胜出的

概率是 P_{short}。因此，一般而言，

$$P_{\text{stan}} = a^2 + 2as + (s^2 + 2ab)P_{\text{short}}.$$

现在取 $p_A = p_B = p$ 和 $q_A = q_B = q$，于是 $a = p^2$，$b = q^2$，$s = 2pq$，而 $P_{\text{short}} = \dfrac{p^2}{p^2+q^2}$。那么

$$P_{\text{stan}} = p^4 + 4p^3q + \frac{6p^4q^2}{p^2+q^2}.$$

想一想 有没有办法比较 P_{stan} 与 P_{short}，而无须把所有式子按 p 和 q 的幂展开？

有：把简化网球同样在头四局上展开。把 $P_{\text{short}} = a + sP_{\text{short}}$ 连用两次，

$$P_{\text{short}} = a + as + s^2P_{\text{short}}.$$

用文字来说：A 全胜第一对局；或者第一对局各胜一局而 A 全胜第二对局；或者两对局都各胜一局，一切重新开始。从 P_{stan} 的公式减去此式，s^2P_{short} 两项互相抵消：

$$\begin{aligned} P_{\text{stan}} - P_{\text{short}} &= a^2 + as - a + 2abP_{\text{short}} \\ &= a(a+s-1) + 2abP_{\text{short}} \\ &= -ab + \frac{2a^2b}{a+b} \qquad (\text{因为 } a+s-1 = -b) \\ &= \frac{ab(a-b)}{a+b}. \end{aligned}$$

这个结果对任何 p_A、p_B 都成立。代入 $a = p^2$ 和 $b = q^2$，

$$P_{\text{stan}} - P_{\text{short}} = \frac{p^2q^2(p^2-q^2)}{p^2+q^2} = \frac{p^2q^2(p-q)(p+q)}{p^2+q^2} = \frac{p^2q^2(p-q)}{p^2+q^2},$$

因为 $p+q=1$。这就是所求的结果。

个中原因从运算过程中已可看出：两种赛制只在 A 全胜第一对局而 B 全胜第二对局（或次序相反）时才有分别。简化网球把胜利判给全胜第一对局的一方；标准网球则由 2–2 继续打下去。两种次序的概率都是 ab，所以两者之差是 $ab(2P_{\text{short}} - 1)$，对较强的球员有利。

验算

- 若 $p = q = \frac{1}{2}$，两者之差为 0，按对称性理应如此：两个概率都是 $\frac{1}{2}$。
- 取 $p = 0.6$：$P_{\text{short}} = 0.36/0.52 \approx 0.6923$，$P_{\text{stan}} = 0.1296 + 0.3456 + 0.124416/0.52 \approx 0.7145$，两者之差 0.02215 与 $0.0576 \times 0.2/0.52$ 吻合。

- 4 局内分出胜负的概率 $a^2+b^2+2(a+b)s$，与打成 2–2 的概率 s^2+2ab，相加得 $(a+b+s)^2=1$，理应如此。
- 按实际规则用电脑模拟，取 $p_A=0.7$ 和 $p_B=0.45$，无论谁先发球，A 在简化网球中胜出的次数都约占 65.6%，与 $0.315/0.48=0.65625$ 吻合。

常见错误

- 忘记 3–1 可以由两种次序得出（先各胜一局再全胜，或先全胜再各胜一局），因而漏掉因数 2。
- 以为每一局的胜出概率都相同：其实它取决于谁发球，这正是成对分组如此有用的原因。
- 写出一个无穷等比级数（infinite geometric series），却没有留意到它其实就是方程 $P=a+sP$。

STEP 之外

这种逐对分析的论证，是带有吸收态（absorbing states）的马可夫链（Markov chain）的入门例子，而方程 $P=a+sP$ 就是首步分析（first-step analysis），是求吸收概率（absorption probabilities）的标准方法。同样的想法可以求出赢得一整盘网球或一局壁球的概率，并进一步引向赌徒破产问题（gambler's ruin）和随机漫步（random walks）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）正好以这个程度讲解随机漫步和马可夫链。

详细解答

STEP III 1994，第 13 题

概率与统计 · 指数分布的寿命

> **STEP III 1994，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/94-s3-q13

关键思路

一切都源自指数分布（exponential distribution）寿命的生存函数（survival function）$\mathrm{P}(T > t) = \mathrm{e}^{-\lambda t}$。以“已存活至 t_0”为条件，只不过是把两个这样的指数相除，著名的无记忆性（memoryless property）正是由此而来。直至两条绳都失效为止的时间，是两个独立（independent）寿命的最大值，而最大值不超过 t 恰好等价于每个寿命都不超过它。因此我们先求出它的分布函数（distribution function），然后求导。至于最后一部分，“两条绳在同一场表演中失效”是指两条绳在同一个区间 $((n-1)h,\, nh]$ 内失效，再对 n 求和。这个和是一个等比级数（geometric series）。

解答

已在使用中的绳

设 T 为该绳的总寿命，则对 $t \geq 0$ 有

$$\mathrm{P}(T > t) = \int_t^{\infty} \lambda \mathrm{e}^{-\lambda u}\, \mathrm{d}u = \mathrm{e}^{-\lambda t}.$$

已知该绳已经撑过了 t_0 小时，即 $T > t_0$。设 $S = T - t_0$ 为它还能再用的时间。对 $s \geq 0$，事件 $\{T > t_0 + s\}$ 包含于 $\{T > t_0\}$ 之内，所以

$$\mathrm{P}(S > s \mid T > t_0) = \frac{\mathrm{P}(T > t_0 + s)}{\mathrm{P}(T > t_0)} = \frac{\mathrm{e}^{-\lambda(t_0+s)}}{\mathrm{e}^{-\lambda t_0}} = \mathrm{e}^{-\lambda s}.$$

因此，剩余时间 S 的概率密度（density）在 $s \geq 0$ 时为 $\lambda \mathrm{e}^{-\lambda s}$（在 $s < 0$ 时为 0）。无论 t_0 是多少，它都服从参数同样为 λ 的指数分布。

诠释。用过的绳与新绳完全一样好。绳并不“记得”自己过去用了多久，无论绳龄多大，

它在接下来一小段时间 δt 内失效的机会都约为 $\lambda\,\delta t$（即失效率（failure rate）为常数）。这种模型适用于由突发的随机意外所造成的失效。但对真实的绳来说并不切合实际：绳会磨损、起毛，所以旧绳应该更容易在短期内失效。采用失效率随绳龄递增的模型会更好。表演者坚持使用新绳，在现实中是明智的，尽管在这个模型之下并无分别。

直至两条新绳都失效的时间

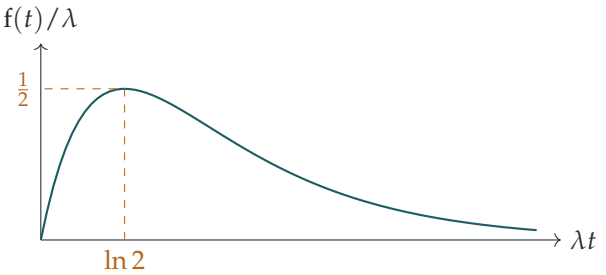
设 T_1, T_2 为两条新绳的寿命。两者互相独立，各自的分布函数为 $\mathrm{P}(T_i \le t) = 1 - \mathrm{e}^{-\lambda t}$ $(t \ge 0)$。两条绳到时刻 t 都已失效，恰好等价于较迟的那次失效已经发生，所以我们要求的时间是 $M = \max(T_1, T_2)$。现在 $M \le t$ 当且仅当 $T_1 \le t$ 并且 $T_2 \le t$，所以由独立性得

$$\mathrm{P}(M \le t) = \mathrm{P}(T_1 \le t)\,\mathrm{P}(T_2 \le t) = \left(1 - \mathrm{e}^{-\lambda t}\right)^2 \qquad (t \ge 0),$$

而对 $t < 0$ 则有 $\mathrm{P}(M \le t) = 0$。利用链式法则（chain rule）求导，

$$\mathrm{f}(t) = 2\left(1 - \mathrm{e}^{-\lambda t}\right) \cdot \lambda \mathrm{e}^{-\lambda t} = 2\lambda \mathrm{e}^{-\lambda t}\left(1 - \mathrm{e}^{-\lambda t}\right) \quad (t \ge 0),$$

而对 $t < 0$ 则有 $\mathrm{f}(t) = 0$，正如题目所求。



与单一指数寿命不同，M 在 $t = 0$ 处的密度为 0，因为两次失效不可能一开始就同时发生。密度在 $\mathrm{e}^{-\lambda t} = \frac{1}{2}$ 处达到最高点。

两条绳都在第 n 场表演中失效

第 n 场表演对应表演时间的区间 $\big((n-1)h,\, nh\big]$。记 $q = \mathrm{e}^{-\lambda h}$。对其中一条绳，

$$\mathrm{P}\big((n-1)h < T_i \le nh\big) = \mathrm{e}^{-\lambda(n-1)h} - \mathrm{e}^{-\lambda nh} = q^{n-1}(1-q).$$

想一想 我们应否把 $\mathrm{f}(t)$ 在第 n 场表演的时段上积分?

不应该。$\int_{(n-1)h}^{nh} \mathrm{f}(t)\,\mathrm{d}t$ 是第二次失效发生在第 n 场表演中的概率。第一条绳可能早在之前的某场表演中已经失效。我们需要的是每一条绳都在第 n 场表演中失效，而两条绳互相独立，所以

$$p_n = \left(q^{n-1}(1-q)\right)^2 = \mathrm{e}^{-2\lambda(n-1)h}\left(1 - \mathrm{e}^{-\lambda h}\right)^2.$$

两条绳在同一场表演中失效

对 $n = 1, 2, 3, \ldots$，各个“两条绳都在第 n 场表演中失效”的事件互斥（mutually exclusive），所以我们把它们的概率相加。这是一个公比为 $q^2 < 1$ 的等比级数（因为 $\lambda h > 0$）：

$$\sum_{n=1}^{\infty} p_n = (1-q)^2 \sum_{n=1}^{\infty} q^{2(n-1)} = \frac{(1-q)^2}{1-q^2} = \frac{1-q}{1+q}.$$

现在把分子和分母同乘以 $\mathrm{e}^{\lambda h/2}$：

$$\frac{1-\mathrm{e}^{-\lambda h}}{1+\mathrm{e}^{-\lambda h}} = \frac{\mathrm{e}^{\lambda h/2} - \mathrm{e}^{-\lambda h/2}}{\mathrm{e}^{\lambda h/2} + \mathrm{e}^{-\lambda h/2}} = \tanh\left(\frac{\lambda h}{2}\right),$$

这就是所要求的结果。

验算

- $\int_0^{\infty} \mathrm{f}(t)\,\mathrm{d}t = \left[(1-\mathrm{e}^{-\lambda t})^2\right]_0^{\infty} = 1$，所以 f 确实是一个概率密度。
- 极限情况：当 $h \to \infty$（一场永不完结的表演）时，答案 $\tanh(\lambda h/2) \to 1$。当 $h \to 0$ 时，它趋向 0，因为两个连续型寿命几乎不可能落在同一个极短的区间内。当 h 很小时，它约为 $\lambda h/2$。
- 数值上，取 $\lambda = 0.7$ 和 $h = 0.8$，级数与 $\tanh(0.28)$ 都给出 0.27291。模拟 2×10^5 对绳所得的结果为 0.2746，而 $p_1 \approx 0.1839$ 亦与模拟所得的 0.1848 吻合。

常见错误

- 用了*最小值*而非最大值。最小值是直至*第一条*绳失效的时间，它服从参数为 2λ 的指数分布。
- 把 f 在第 n 场表演的时段上积分来求 p_n。这些概率加起来是 1，而不是 $\tanh(\lambda h/2)$。

STEP 之外

无记忆性刻划了指数分布：它是唯一具有这个性质的连续型分布。在可靠性理论（reliability theory）中，“失效率”（风险函数，hazard function）$\mathrm{f}(t)/\mathrm{P}(T > t)$ 只有在指数分布下才是常数。磨损则以递增的风险函数来建模，例如采用韦布尔分布（Weibull distribution）。独立变量的最大值和最小值是*次序统计量*（order statistics）的两个极端情况，你会在大学的第一门概率课程中遇到它们。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门教材）涵盖了以上所有内容。

详细解答

STEP III 1994，第 14 题

概率与统计 · 几何概率

> **STEP III 1994，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/94-s3-q14

关键思路

圆周角（inscribed angle）等于所对的弧在圆心所张的角的一半。因此，三角形的三个角正好是 π 乘以 P、Q、R 把圆分成的三段弧，而弧长以占圆周的比例来量度。问题于是变成：一个圆怎样被随机分成三段？而这其实是一个关于三角形内面积的问题。

解答

建立模型

以占圆周的比例来量度圆上的位置，这样圆的“长度”就是 1。由旋转对称性（rotational symmetry），P 的位置无关重要，所以把 P 放在 0。于是 Q 和 R 的位置是独立的，并在 $[0,1)$ 上均匀分布（uniformly distributed）。

设 $U < V$ 为 Q 和 R 的位置由小至大的排列。数对 (U,V) 在区域 $0 < u < v < 1$ 上均匀分布，而这个区域的面积是 $\frac{1}{2}$，所以它的联合密度（joint density）在该区域上是 2。（两点重合的概率是 0，下文所有边界情况亦然。）三段弧是

$$X = U, \qquad Y = V - U, \qquad Z = 1 - V, \qquad X + Y + Z = 1.$$

变量代换 $(u,v) \mapsto (x,y) = (u, v-u)$ 的雅可比行列式（Jacobian）是 1，所以 (X,Y) 在三角形

$$T = \{(x,y) : x > 0,\ y > 0,\ x + y < 1\},$$

上的密度是 2，而这个三角形的面积是 $\frac{1}{2}$。因此，下文每一个概率都是 T 内某块面积的两倍。

想一想 为甚么三角形的每个角都是 π 乘以一段弧?

三角形在 R 的角是一个圆周角，它所对的是不包含 R 的那段弧 PQ。根据圆周角定理（inscribed angle theorem），它等于该弧在圆心所张的角的一半。一段占圆周比例为 x 的弧，在圆心所张的角是 $2\pi x$，所以在 R 的角是 πx。每个顶点的情况都一样，所以 PQR 的三个角是 πX、πY 和 πZ（它们的和是 π，正如所料）。因此

$$\text{某个角大于}k\pi \iff \text{对应的弧大于}k.$$

设 A_X、A_Y 和 A_Z 分别为事件 $X > k$、$Y > k$ 和 $Z > k$。我们要求 $\mathrm{P}(A_X \cup A_Y \cup A_Z)$。

一段弧大于 k

对于 $0 \le k \le 1$:

- A_X 是 T 中 $x > k$ 的部分：一个两条直角边均为 $1-k$ 的三角形，面积是 $\frac{1}{2}(1-k)^2$。
- A_Y 是 $y > k$ 的部分：同样的计算给出面积 $\frac{1}{2}(1-k)^2$。
- A_Z 是 $x + y < 1 - k$ 的部分：位于角落、两条直角边均为 $1-k$ 的三角形，面积是 $\frac{1}{2}(1-k)^2$。

乘以二，得 $\mathrm{P}(A_X) = \mathrm{P}(A_Y) = \mathrm{P}(A_Z) = (1-k)^2$。

$\frac{1}{2} \le k \le 1$ 的情况

最多只有一段弧可以大于 k：假如有两段，它们的总和就会大于 $2k \ge 1$，但三段弧加起来才是 1。所以事件 A_X、A_Y、A_Z 互斥（mutually exclusive），而

$$\mathrm{P}(\text{某个角大于}k\pi) = 3(1-k)^2,$$

正是所求。

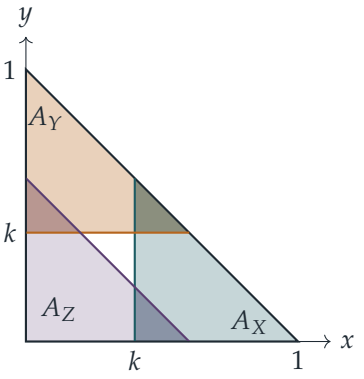
$\frac{1}{3} \le k \le \frac{1}{2}$ 的情况

现在可以有两段弧同时大于 k，所以我们使用容斥原理（inclusion–exclusion）。

两段弧。 $A_X \cap A_Y$ 是 T 中 $x > k$、$y > k$、$x + y < 1$ 的部分。代入 $x = k + s$、$y = k + t$，条件变成 $s, t > 0$、$s + t < 1 - 2k$，这是一个两条直角边均为 $1 - 2k \ge 0$ 的三角形，面积是 $\frac{1}{2}(1-2k)^2$。所以 $\mathrm{P}(A_X \cap A_Y) = (1-2k)^2$。另外两对的结果相同：例如 $A_X \cap A_Z$ 是 $x > k$、$y > 0$、$x + y < 1 - k$，同样是一个两条直角边均为 $1 - 2k$ 的三角形。

三段弧。 三段弧全部大于 k 就需要 $X + Y + Z > 3k \ge 1$，这是不可能的。所以三个事件的交集的概率是 0。

下图显示 $k = 0.4$ 时的 T：三个位于角落的三角形是 A_X、A_Y、A_Z，颜色较深的重叠部分是两两的交集。



由容斥原理，

$$\begin{aligned}\mathrm{P}(A_X \cup A_Y \cup A_Z) &= 3(1-k)^2 - 3(1-2k)^2 + 0 \\ &= 3\big[(1-k)-(1-2k)\big]\big[(1-k)+(1-2k)\big] \\ &= 3k(2-3k).\end{aligned}$$

所以当 $\frac{1}{3} \le k \le \frac{1}{2}$ 时，至少有一个角大于 $k\pi$ 的概率是

$$3k(2-3k) = 6k - 9k^2.$$

验算

- 在 $k = \frac{1}{2}$ 时，两条公式都给出 $\frac{3}{4}$，所以它们连续地衔接起来。这就是那个经典结果：随机的圆内接三角形是钝角三角形的概率为 $\frac{3}{4}$。
- 在 $k = \frac{1}{3}$ 时，我们得到 $3 \cdot \frac{1}{3} \cdot 1 = 1$。这是对的：最大的角永远不小于 $\frac{\pi}{3}$，因为三个角加起来是 π。
- 在 $k = 1$ 时，我们得到 0，因为没有角可以大于 π。
- 模拟 4×10^5 个随机三角形，直接由坐标计算各个角，在 $k = 0.4$ 时得到 0.9602（公式给出 0.96），在 $k = 0.6$ 时得到 0.4783（公式给出 0.48）。

常见错误

- 用了整段弧 PQ，而不是不包含第三个顶点的那段弧。两段弧 PQ 之中，只有一段是 R 所对的弧。

- 把 X、Y、Z 当作独立的均匀随机变量。它们受 $X+Y+Z=1$ 所约束，而老老实实处理这个约束的方法，就是面积图。
- 在 $k=\frac{1}{2}$ 以下仍然套用 $3(1-k)^2$。当 $k<\frac{1}{2}$ 时，它把有两个大角的三角形重复计算了，而在 $k=\frac{1}{3}$ 时它会给出 $\frac{4}{3}>1$。

STEP 之外

弧 (X,Y,Z) 在单纯形（simplex）$x+y+z=1$ 上均匀分布。这就是狄利克雷分布（Dirichlet distribution），它描述随机点把一个区间或一个圆切开后留下的“间距”。有 n 个点时，同样的容斥原理给出 $\mathrm{P}(\text{最大间距}>k)=\sum_j(-1)^{j+1}\binom{n}{j}(1-jk)_+^{n-1}$，这条公式可追溯至 Whitworth。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）对几何概率和联合密度有很好的论述。

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

1994 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。

