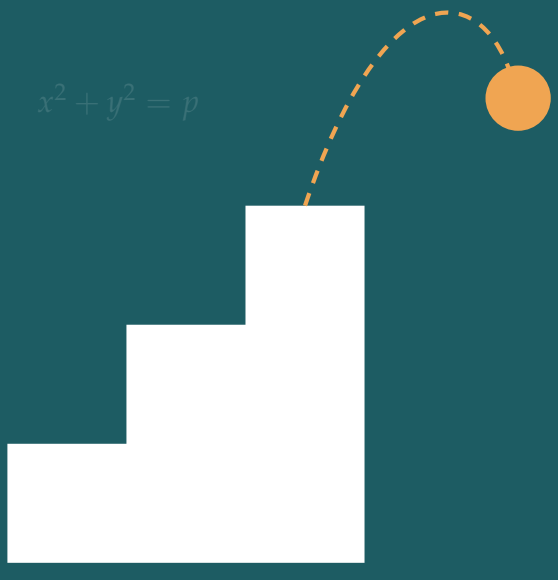
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1991 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

1991 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1991 年详细解答

1991 年 STEP 试卷（STEP I, II 及 III）全部 48 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 三角恒等式 纯数学 3

第 2 题 变率 纯数学 6

第 3 题 复数等比级数 纯数学 10

第 4 题 几何中的最优化 纯数学 14

第 5 题 正则向量组 纯数学 18

第 6 题 驻点与广义积分 纯数学 22

第 7 题 可分离变量微分方程 纯数学 26

第 8 题 代换积分法 纯数学 30

第 9 题 不等式与调和级数 纯数学 34

第 10 题 悬链的平衡 力学 37

第 11 题 滑轮与冲量 力学 41

第 12 题 涉及摩擦力的平衡 力学 45

第 13 题 动量守恒 力学 49

第 14 题 角形的计数 概率与统计 53

第 15 题 态近似 概率与统计 57

第 16 题 条件概率 概率与统计 61

STEP II

第 1 题 二次方程与判别式 纯数学 65

第 2 题 双曲线的渐近线 纯数学 69

第 3 题 对称联立方程 纯数学 73

第 4 题 余弦之和 纯数学 76

第 5 题 积分归约公式 纯数学 80

第 6 题 积分的界限 纯数学 84

第 7 题 函数方程 纯数学 . . . 88
第 8 题 非线性一阶常微分方程 纯数学 . . . 92
第 9 题 矩阵群 纯数学 . . . 96
第 10 题 三维空间中的直线与平面 纯数学 . . . 99
第 11 题 抛体与发射点的距离 力学 . . . 103
第 12 题 受迫弹性振动 力学 . . . 106
第 13 题 刚体转动 力学 . . . 110
第 14 题 相对速度与追踪 力学 . . . 114
第 15 题 期望值的线性性质 概率与统计 . . . 118
第 16 题 二项分布的正态近似 概率与统计 . . . 122

STEP III

第 1 题 级数与展开式 纯数学 . . . 126
第 2 题 几何中的复数 纯数学 . . . 129
第 3 题 分段函数与反函数 纯数学 . . . 133
第 4 题 三维空间中的直线与平面 纯数学 . . . 137
第 5 题 极坐标曲线与常微分方程 纯数学 . . . 141
第 6 题 双曲线性变换 纯数学 . . . 145
第 7 题 利用对称性积分 纯数学 . . . 150
第 8 题 归约公式与求和 纯数学 . . . 154
第 9 题 椭圆的弧长 纯数学 . . . 158
第 10 题 牛顿恒等式 纯数学 . . . 162
第 11 题 刚体平衡 力学 . . . 165
第 12 题 直线上的运动 力学 . . . 169
第 13 题 抛体与天花板的碰撞 力学 . . . 173
第 14 题 悬挂及弹簧的圆周运动 力学 . . . 177
第 15 题 样本中配对的计数 概率与统计 . . . 181
第 16 题 联合分布与边际分布 概率与统计 . . . 185

详细解答

STEP I 1991，第 1 题

纯数学 · 三角恒等式

> **STEP I 1991，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/91-s1-q1

关键思路

角度所满足的条件让我们可以消去其中一个角：由于 $\psi = \frac{1}{2}\pi - (\theta + \phi)$，我们有 $\sin\psi = \cos(\theta + \phi)$，于是该恒等式（identity）便成为一个关于两个自由角的命题。其后，乘积 $\sin\theta\sin\phi$ 与 $\cos(\theta + \phi)$ 结合起来得出 $\cos(\theta - \phi)$，而乘积 $\cos(\theta + \phi)\cos(\theta - \phi)$ 的形式十分简洁。至于第二部分，所选的特殊角令 ψ 等于 θ 的一半，有关 $\frac{1}{10}\pi$ 的资料正是由此而来。

解答

恒等式

由于 $\theta + \phi + \psi = \frac{1}{2}\pi$，

$$\sin\psi = \sin(\tfrac{1}{2}\pi - (\theta + \phi)) = \cos(\theta + \phi).$$

因此恒等式的左方是

$$L = \sin^2\theta + \sin^2\phi + \cos(\theta + \phi)[\cos(\theta + \phi) + 2\sin\theta\sin\phi].$$

在括号之内，$\cos(\theta + \phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$，所以

$$\cos(\theta + \phi) + 2\sin\theta\sin\phi = \cos\theta\cos\phi + \sin\theta\sin\phi = \cos(\theta - \phi).$$

现在我们需要乘积 $\cos(\theta+\phi)\cos(\theta-\phi)$。把复角公式（compound-angle formulae）乘开，

$$\begin{aligned}\cos(\theta+\phi)\cos(\theta-\phi) &= (\cos\theta\cos\phi-\sin\theta\sin\phi)(\cos\theta\cos\phi+\sin\theta\sin\phi)\\ &= \cos^2\theta\cos^2\phi-\sin^2\theta\sin^2\phi\\ &= \cos^2\theta\,(1-\sin^2\phi)-(1-\cos^2\theta)\sin^2\phi\\ &= \cos^2\theta-\sin^2\phi.\end{aligned}$$

因此

$$L=\sin^2\theta+\sin^2\phi+\cos^2\theta-\sin^2\phi=\sin^2\theta+\cos^2\theta=1,$$

这就是所需的恒等式。

$\sin\frac{1}{10}\pi$ 所满足的三次方程

取 $\theta=\phi=\frac{1}{5}\pi$。那么 $\psi=\frac{1}{2}\pi-\frac{2}{5}\pi=\frac{1}{10}\pi$。设

$$x=\sin\tfrac{1}{10}\pi,\qquad s=\sin\tfrac{1}{5}\pi.$$

由恒等式得

$$2s^2+x^2+2s^2x=1,\qquad \text{即}\qquad 2s^2(1+x)=1-x^2. \tag{$*$}$$

> **想一想** 为甚么不直接利用 $\sin\psi=\cos 2\theta=1-2s^2$ 来消去 s?

这个关系式对每一个满足 $\psi=\frac{1}{2}\pi-2\theta$ 的 θ 都成立，所以把 $s^2=\frac{1}{2}(1-x)$ 代入 $(*)$ 只会得出 $(1-x)(1+x)=1-x^2$：这是对的，却毫无内容。我们必须用上 $\frac{1}{5}\pi$ 的某种特殊性质。这个特殊之处就是 $\frac{1}{5}\pi=2\cdot\frac{1}{10}\pi$，所以 $\theta=2\psi$，而二倍角公式（double-angle formula）把 s 与 x 联系起来：

$$s=\sin 2\psi=2\sin\psi\cos\psi,\qquad \text{所以}\qquad s^2=4x^2(1-x^2).$$

把这个结果代入 $(*)$，

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

由于 $0<\frac{1}{10}\pi<\frac{1}{2}\pi$，我们有 $0<x<1$，所以 $1-x^2\neq 0$，可以把它约去：

$$8x^2(1+x)=1,\qquad \text{即}\qquad 8x^3+8x^2-1=0.$$

所以 $\sin\frac{1}{10}\pi$ 满足题目给出的三次方程（cubic），证明完毕。

验算

恒等式可以用 $\theta = \frac{1}{2}\pi$、$\phi = \psi = 0$ 来检验：左方是 $1+0+0+0 = 1$。取 $\theta = \phi = \psi = \frac{1}{6}\pi$ 时，左方是 $3\cdot\frac{1}{4} + 2\cdot\frac{1}{8} = 1$。

至于三次方程，$\sin\frac{1}{10}\pi = \sin 18^\circ \approx 0.309017$，而 $8x^3 + 8x^2 - 1$ 在该处的值在舍入误差范围内等于 0。事实上，这个三次式可以因式分解为 $(2x+1)(4x^2+2x-1)$，其根为 $-\frac{1}{2}$ 和 $\frac{1}{4}(-1 \pm \sqrt{5})$，而 $\frac{1}{4}(\sqrt{5}-1) \approx 0.309017$ 正是大家熟悉的 $\sin 18^\circ$ 的准确值。

常见错误

- 只用 $\sin\psi = \cos 2\theta$ 来消去 s，结果得到恒等式 $0 = 0$；见上面的方框。
- 除以 $1 - x^2$ 却没有说明它为何不等于零。这里 $x = \sin\frac{1}{10}\pi$ 严格介乎 0 与 1 之间。

STEP 之外

三次方程的另外两个根并非偶然：$-\frac{1}{2} = \sin\left(-\frac{1}{6}\pi\right)$，而 $\frac{1}{4}(-1-\sqrt{5}) = -\sin\frac{3}{10}\pi$。这类方程记录了 sin 和 cos 在 π 的有理数倍处的取值之间的代数关系，而要理解这些关系，最好的途径是单位根（roots of unity）：$\cos\frac{1}{5}\pi$ 与五次单位根，即 $z^5 = 1$ 的解，有密切的关系。这就是为甚么以直尺和圆规作正五边形，归根究柢取决于 $\sqrt{5}$。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是从几何角度认识这些想法的绝佳读物。

详细解答

STEP I 1991，第 2 题

纯数学 · 变率

> **STEP I 1991，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/91-s1-q2

关键思路

把融化规律改用半径来表达。对球体而言，$\frac{\mathrm{d}V}{\mathrm{d}t} = 4\pi r^2 \frac{\mathrm{d}r}{\mathrm{d}t}$，而表面积同样是 $4\pi r^2$，所以“体积减少的速率与表面积成正比”其实只是说半径以恒定速率缩短。两个雪球的比例常数相同，因此在任何一段时间内，两个半径缩短的长度都相同。于是所有量都变成单一变量的函数（function），这个变量就是每个半径缩短了的长度 x。

解答

两个半径以相同的恒定速率缩短

设雪球的半径为 r，体积为 $V = \frac{4}{3}\pi r^3$，表面积为 $S = 4\pi r^2$。融化规律是 $\frac{\mathrm{d}V}{\mathrm{d}t} = -kS$，其中 $k > 0$ 为常数，两个雪球的常数相同。由链式法则（chain rule），

$$4\pi r^2 \frac{\mathrm{d}r}{\mathrm{d}t} = -k \cdot 4\pi r^2 \quad \Longrightarrow \quad \frac{\mathrm{d}r}{\mathrm{d}t} = -k,$$

这里两边除以 $4\pi r^2$，只要雪球仍然存在（$r > 0$），这一步便成立。所以经过时间 t 后，两个半径都减少了相同的长度 $x = kt$：头部的半径为 $2R - x$，身体的半径为 $3R - x$，适用于 $0 \le x \le 2R$（头部在 $x = 2R$ 时消失）。

高度减半时的体积

由于头部放在身体之上，Frosty 的高度是两个直径之和：

$$h = 2(2R - x) + 2(3R - x) = 10R - 4x.$$

最初 $h = 10R$，所以高度减半即 $10R - 4x = 5R$，也就是 $x = \frac{5}{4}R$。这个值小于 $2R$，所以头部仍然存在，此时两个半径分别为 $\frac{3}{4}R$ 和 $\frac{7}{4}R$。因子 $\frac{4}{3}\pi$ 在比值中约去，所以

$$\frac{V}{V_0} = \frac{\left(\frac{3}{4}\right)^3 + \left(\frac{7}{4}\right)^3}{2^3 + 3^3} = \frac{(27 + 343)/64}{35} = \frac{370}{2240} = \frac{37}{224}.$$

所以当 Frosty 的高度减半时，他的体积只剩下原来的 $\frac{37}{224} \approx 0.165$：体积随长度的三次方而变化。

$\mathrm{d}V/\mathrm{d}S$ 的最大值

记 $a = 2R - x$ 及 $b = 3R - x$。当 $0 \le x < 2R$ 时，

$$V = \tfrac{4}{3}\pi(a^3 + b^3), \qquad S = 4\pi(a^2 + b^2),$$

又因为 $\frac{\mathrm{d}a}{\mathrm{d}x} = \frac{\mathrm{d}b}{\mathrm{d}x} = -1$，

$$\frac{\mathrm{d}V}{\mathrm{d}x} = -4\pi(a^2 + b^2), \qquad \frac{\mathrm{d}S}{\mathrm{d}x} = -8\pi(a + b).$$

想一想 V 和 S 都是时间（或 x）的函数。那么 $\mathrm{d}V/\mathrm{d}S$ 是甚么意思?

当 x 增加时，S 严格递减（因为 $\frac{\mathrm{d}S}{\mathrm{d}x} < 0$），所以 V 可以看成 S 的函数。由链式法则 $\frac{\mathrm{d}V}{\mathrm{d}S} = \frac{\mathrm{d}V}{\mathrm{d}x} \Big/ \frac{\mathrm{d}S}{\mathrm{d}x}$：

$$\frac{\mathrm{d}V}{\mathrm{d}S} = \frac{a^2 + b^2}{2(a + b)}.$$

现在利用 $b - a = R$，这个关系始终不变。设 $s = a + b = 5R - 2x$。那么 $a^2 + b^2 = \frac{1}{2}\big((a + b)^2 + (b - a)^2\big) = \frac{1}{2}(s^2 + R^2)$，所以

$$\frac{\mathrm{d}V}{\mathrm{d}S} = \frac{s^2 + R^2}{4s} = \frac{s}{4} + \frac{R^2}{4s} = f(s), \text{ 记为此.}$$

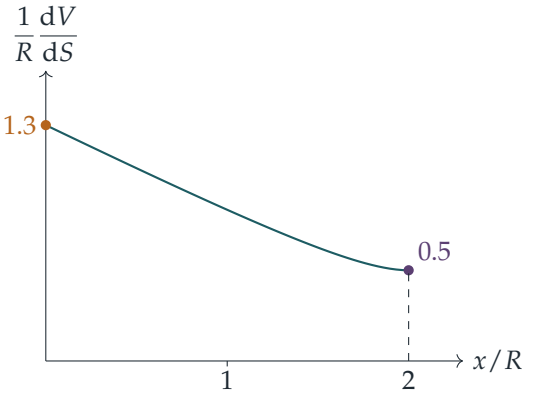
当 x 由 0 增至 $2R$ 时，s 由 $5R$ 减至 R。而

$$f'(s) = \frac{1}{4} - \frac{R^2}{4s^2} = \frac{s^2 - R^2}{4s^2} > 0 \quad \text{当} s > R,$$

所以 f 在 $[R, 5R]$ 上递增，其最大值在 $s = 5R$ 处取得，也就是在最开始的时候（$x = 0$）：

$$\left(\frac{\mathrm{d}V}{\mathrm{d}S}\right)_{\max} = \frac{5R}{4} + \frac{R}{20} = \frac{13R}{10}.$$

（直接计算：$\frac{4R^2 + 9R^2}{2 \cdot 5R} = \frac{13}{10}R$。）其后这个值持续下降，到头部刚好消失时降至 $f(R) = \frac{1}{2}R$。



验算

- 对每个雪球按 $\frac{dV}{dt} = -kS$ 逐步模拟（取 $R = k = 1$），在 $t \approx 1.25$ 时高度减半，体积比为 0.16518，与 $\frac{37}{224} = 0.165178\ldots$ 吻合。
- 模拟所得的 V 和 S 的有限差分（finite differences）自始至终与 $\frac{a^2+b^2}{2(a+b)}$ 相符，而且从不超过 $1.3R$。
- 单位：dV/dS 是体积除以面积，所以是长度，$\frac{13}{10}R$ 正是如此。

常见错误

- 把高度当成两个半径之和，而不是两个直径之和。（在这题中碰巧不会改变 x 的值，但做法是错的。）
- 计算了 V/S 而不是 dV/dS；对单一球体而言，$V/S = r/3$，但 $dV/dS = r/2$。
- 一味寻找驻点（stationary point）：f 的转向点在 $s = R$，即区间的端点，所以最大值在区间的另一端。

STEP 之外

这道题的核心在于：“速率与表面积成正比”这条规律，会化为半径以恒定速率变化，这是几何如何令微分方程（differential equation）变得轻而易举的第一个例子。后半部分是参数式导数（parametric derivative），即沿曲线 $(S(t), V(t))$ 的 dV/dS，以及在闭区间（closed interval）上求最大值时须有的谨慎：最大值可能位于端点，而不是驻点；Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）解释了为何这样的最大值必定存在。至于以微分方程建立模型，可参考 Boyce 与 DiPrima 的 *Elementary Differential*

Equations（初等微分方程）。

详细解答

STEP I 1991，第 3 题

纯数学 · 复数等比级数

> **STEP I 1991，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/91-s1-q3

关键思路

在阿根图（Argand diagram）上，乘以 $re^{i\phi}$ 会把一个向量放大 r 倍，并按逆时针方向转过角 ϕ。路径上每一条线段都是前一条线段按相同的比例缩放、按相同的角度旋转而成，所以每个线段向量都等于前一个乘以一个固定的复数（complex number）。于是这些线段构成一个等比数列（geometric progression），把它们加起来便得出每个顶点的位置。

解答

相邻线段之比

由 P_n 到 P_{n+1} 的向量由 $z_{n+1}-z_n$ 表示，而由 P_{n-1} 到 P_n 的向量则由 z_n-z_{n-1} 表示。两个复数相除时，模（modulus）相除，辐角（argument）相减：

$$\left|\frac{z_{n+1}-z_n}{z_n-z_{n-1}}\right| = \frac{P_nP_{n+1}}{P_{n-1}P_n} = d, \qquad \arg\frac{z_{n+1}-z_n}{z_n-z_{n-1}} = \arg(z_{n+1}-z_n) - \arg(z_n-z_{n-1}) = \theta,$$

因为第二条线段的长度是第一条的 d 倍，而由一条有向线段的方向按逆时针转到下一条的方向，所转的角是 θ。（辐角只确定到相差 2π 的倍数，但这并不影响该复数本身。）因此，对所有 $n \geq 2$，

$$\frac{z_{n+1}-z_n}{z_n-z_{n-1}} = d\,\mathrm{e}^{\mathrm{i}\theta} = d(\cos\theta + \mathrm{i}\sin\theta).$$

z_{n+1} 的公式

设 $w = d\mathrm{e}^{\mathrm{i}\theta}$。第一条线段是 $z_2 - z_1 = 1$，而其后每一条线段都是前一条的 w 倍，所以由归纳法（induction）可得

$$z_{k+1} - z_k = w^{k-1} \qquad (k \geq 1).$$

把 $k = 1, 2, \ldots, n$ 的各式相加，左边逐项相消（telescoping），只剩下 $z_{n+1} - z_1 = z_{n+1}$：

$$z_{n+1} = 1 + w + w^2 + \cdots + w^{n-1} = \frac{w^n - 1}{w - 1} = \frac{d^n \mathrm{e}^{\mathrm{i}n\theta} - 1}{d\mathrm{e}^{\mathrm{i}\theta} - 1}.$$

求等比级数（geometric series）之和时需要 $w \neq 1$，而 $d \neq 1$ 这个条件正是用在这里，因为 $|w| = d$。

$\theta = \frac{1}{3}\pi$、$d = 2$ 的情况

这时 $w = 2\mathrm{e}^{\mathrm{i}\pi/3} = 1 + \mathrm{i}\sqrt{3}$，所以分母非常简单：$w - 1 = \mathrm{i}\sqrt{3}$。由于 $\frac{1}{\mathrm{i}} = -\mathrm{i}$，

$$z_{n+1} = \frac{w^n - 1}{\mathrm{i}\sqrt{3}} = -\frac{\mathrm{i}}{\sqrt{3}}\left(2^n \cos\tfrac{n\pi}{3} - 1 + \mathrm{i}\, 2^n \sin\tfrac{n\pi}{3}\right),$$

从而

$$\operatorname{Re} z_{n+1} = \frac{2^n \sin\frac{n\pi}{3}}{\sqrt{3}}, \qquad \operatorname{Im} z_{n+1} = \frac{1 - 2^n \cos\frac{n\pi}{3}}{\sqrt{3}}.$$

想一想 $\operatorname{Im} z_{n+1}$ 的正负号怎样告诉我们路径穿过实轴（real axis）？

这条路径是连贯的：它是由线段首尾相接而成的一条连续折线。如果某个顶点严格位于实轴下方，而其后某个顶点严格位于实轴上方（或者反过来），那么两者之间的那段路径必定穿过实轴。所以我们要一次又一次地找出位于实轴两侧的顶点。

$\cos\frac{n\pi}{3}$ 的值以 6 为周期重复。取任意整数 $k \geq 1$。

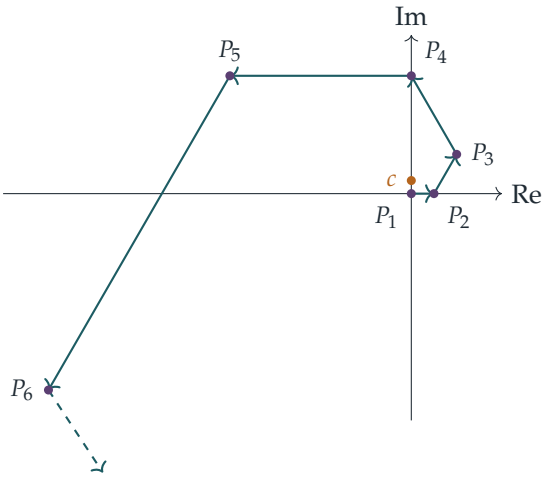
- 当 $n = 6k$ 时，$\cos\frac{n\pi}{3} = \cos 2k\pi = 1$，所以 $\operatorname{Im} z_{6k+1} = \dfrac{1 - 2^{6k}}{\sqrt{3}} < 0$，因为 $2^{6k} > 1$。
- 当 $n = 6k + 3$ 时，$\cos\frac{n\pi}{3} = \cos(2k\pi + \pi) = -1$，所以 $\operatorname{Im} z_{6k+4} = \dfrac{1 + 2^{6k+3}}{\sqrt{3}} > 0$。

因此，对所有 $k \geq 1$，P_{6k+1} 在实轴下方，而 P_{6k+4} 在实轴上方。所以对每个 $k = 1, 2, 3, \ldots$，路径由 P_{6k+1} 到 P_{6k+4} 的那一段都会由下而上穿过实轴。这些是路径上互不相同的部分，每个 k 对应一段，所以路径穿过实轴无限多次。（按同样的论证，路径在 P_{6k+4} 与 P_{6k+7} 之间也会由上而下穿过实轴。）证明至此完成。

事实上，顶点是每三个一组地绕圈的：$\operatorname{Im} z_1, \operatorname{Im} z_2, \ldots$ 的正负号规律是 $0, 0, +, +, +, -, -, -, +, +, +, \ldots$。这背后有一个巧妙的几何原因。把上式写成

$$z_{n+1} = \frac{\mathrm{i}}{\sqrt{3}} - \frac{\mathrm{i}}{\sqrt{3}}\, w^n,$$

可见 P_{n+1} 与定点 $c = \mathrm{i}/\sqrt{3}$ 的距离是 $2^n/\sqrt{3}$，而每走一步，它便绕 c 转过 60°。各顶点位于一条环绕 c 的螺线上，这条螺线每走一步便放大一倍，而 c 刚好在实轴上方不远处，所以螺线会不断扫过实轴。



图中显示 $P_1, \ldots, P_6$；下一个顶点是 $P_7 = -21\sqrt{3}\,\mathrm{i}$，远在下方。路径第一次穿过实轴是在 P_5P_6 上，其后一直留在实轴下方，直至在 P_8P_9 上由下而上穿过，即由 $P_8 = 64 - 21\sqrt{3}\,\mathrm{i}$ 到 $P_9 = 128 + 43\sqrt{3}\,\mathrm{i}$。

验算

- 直接由几何关系计算：$z_3 = 1 + w = 2 + \sqrt{3}\,\mathrm{i}$，而公式给出 $\operatorname{Im} z_3 = (1 - 4\cos\frac{2\pi}{3})/\sqrt{3} = 3/\sqrt{3} = \sqrt{3}$。同样地，再加上 $w^2 = -2 + 2\sqrt{3}\,\mathrm{i}$ 和 $w^3 = -8$，得 $z_5 = -8 + 3\sqrt{3}\,\mathrm{i}$，而在公式中取 $n = 4$，得 $\operatorname{Re} z_5 = 16\sin\frac{4\pi}{3}/\sqrt{3} = -8$ 和 $\operatorname{Im} z_5 = (1+8)/\sqrt{3} = 3\sqrt{3}$。
- 若 $d = 1$，各顶点会落在一个正多边形上，路径会闭合；等比级数的求和公式恰好在这个情况下失效，这正是应有的结果。
- 我们按长度和转角以数值方法建构了这条路径（d、θ 和起点均随机选取），并确认了比值 $d\mathrm{e}^{\mathrm{i}\theta}$、$z_{n+1}$ 的闭合式（closed form），以及在 $\theta = \frac{1}{3}\pi$、$d = 2$ 时上述的正负号规律：在 $z_1, \ldots, z_{41}$ 之中，$\operatorname{Im} z$ 共变号 13 次。

常见错误

- 把比值写成 $\frac{1}{d}\mathrm{e}^{-\mathrm{i}\theta}$：分子是*较后*的那条线段，所以它才是长 d 倍、并转过 $+\theta$ 的那一条。

- 求和时项数差一：由 P_1 到 P_{n+1} 共有 n 条线段，所以级数加到 w^{n-1} 为止。
- 单凭一个位于轴上的顶点便声称路径穿过实轴：你需要的是*严格*位于实轴两侧的顶点（这里 z_1 和 z_2 都在实轴上，所以要由 $k \geq 1$ 开始）。

STEP 之外

规则 $z_{n+1} - z_n = w(z_n - z_{n-1})$ 是一个线性递归关系（linear recurrence），而把每个顶点带到下一个顶点的映射 $z \mapsto c + w(z - c)$，是以其不动点（fixed point）c 为中心的*旋转相似变换*（spiral similarity）：即旋转与放大的结合。这类映射是莫比乌斯变换（Möbius transformation）最简单的例子，把它们反复迭代，便会产生贝壳上常见的等角螺线（又称对数螺线，equiangular (logarithmic) spiral）。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）以这种几何眼光看待复数乘法，阐述得十分精彩。

详细解答

STEP I 1991，第 4 题

纯数学 · 几何中的最优化

> **STEP I 1991，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/91-s1-q4

关键思路

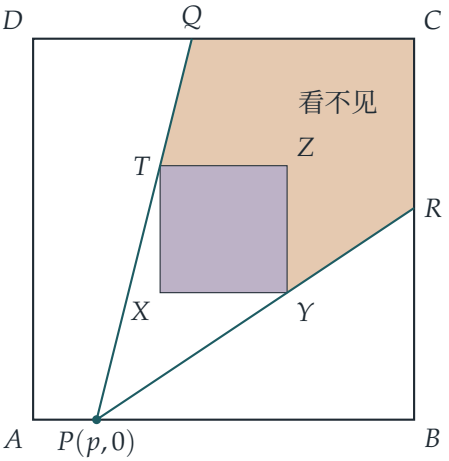
守卫能看见庭院中的每一点，只有被塔楼挡住的点除外，所以我们按照提示，求出从他所在位置 $P=(p,0)$ 望去时塔楼“阴影”的面积。阴影的边界是由 P 出发、刚好擦过塔楼的两条射线。求出这两条射线与围墙的交点，把阴影当作多边形计算其面积，问题便化为在 $[0,4]$ 上求一个单变量函数的最大值。记得检查端点。

解答

建立模型

庭院是大正方形减去塔楼，面积为 $144-16=128$。设守卫位于 $P=(p,0)$，其中 $0\le p\le 4$，并设 $H(p)$ 为他看不见的面积。可见面积为 $128-H(p)$，所以 H 最大时可见面积最小，反之亦然。

由于 $p\le 4$，P 位于直线 $y=4$ 之下，并在直线 $x=4$ 的左方（或在该直线上）。因此从 P 望去，塔楼露出它的底面 XY 和左面 XT，而它的轮廓由顶角 $T(4,8)$ 和 $Y(8,4)$ 决定。顶角 $Z(8,8)$ 位于角 TPY 之内，是看不见的。看不见的区域位于路径 $T\to Z\to Y$ 之外，夹在射线 PT 与 PY 之间。



擦边射线与围墙的交点

由 P 出发穿过 $T(4,8)$ 的射线，每横移 $4-p$ 便上升 8，所以它到达 $y=12$ 时

$$x = p + \tfrac{12}{8}(4-p) = 6 - \tfrac{p}{2}, \qquad \text{由此得} Q = \left(6 - \tfrac{p}{2},\ 12\right).$$

由于 $4 \le 6 - \frac{p}{2} \le 6$，这一点确实在上方的围墙 DC 上。

由 P 出发穿过 $Y(8,4)$ 的射线，每横移 $8-p$ 便上升 4，所以它到达 $x=12$ 时

$$y = \frac{4(12-p)}{8-p}, \qquad \text{由此得} R = \left(12,\ \frac{4(12-p)}{8-p}\right).$$

当 p 由 0 增至 4 时，这个值由 6 增至 8，所以 R 在右方的围墙 BC 上。

看不见的面积

看不见的区域是多边形 $TQCRYZ$。求其面积最简便的方法是

$$H = [PRCQ] - [PYZT],$$

其中 $[\cdots]$ 表示面积：即两条射线之间一直伸延至围墙的整个楔形，减去楔形中位于塔楼远侧两面之前的部分（塔楼前方看得见的地面，连同塔楼本身）。

对 $P(p,0)$、$R(12,q)$、$C(12,12)$、$Q(r,12)$ 使用鞋带公式(shoelace formula)，其中 $q = \frac{4(12-p)}{8-p}$ 及 $r = 6 - \frac{p}{2}$：

$$\begin{aligned}[PRCQ] &= \tfrac{1}{2}\left(pq + (144-12q) + (144-12r) - 12p\right) \\ &= \tfrac{1}{2}\left(288 - 12p - 12r - (12-p)q\right) = 108 - 3p - \frac{2(12-p)^2}{8-p},\end{aligned}$$

这里用了 $12r = 72 - 6p$。同样地，对 $P(p,0)$、$Y(8,4)$、$Z(8,8)$、$T(4,8)$：

$$[\,PYZT\,] = \tfrac{1}{2}\big(4p + 32 + 32 - 8p\big) = 32 - 2p.$$

因此

$$H(p) = 76 - p - \frac{2(12-p)^2}{8-p}.$$

想一想 有没有一个代换（substitution）能令这个函数容易求最大、最小值?

分母不好处理，这提示我们设 $u = 8 - p$，于是 u 的取值范围是 $[4,8]$，而 $12 - p = u + 4$。这样

$$H = 76 - (8-u) - \frac{2(u+4)^2}{u} = 68 + u - \frac{2u^2 + 16u + 32}{u} = 52 - \left(u + \frac{32}{u}\right).$$

第 (i) 部分：最小可见面积

我们需要 H 尽可能大，即 $u + \frac{32}{u}$ 尽可能小。根据算术平均—几何平均不等式（AM–GM inequality），对 $u > 0$，

$$u + \frac{32}{u} \;\geq\; 2\sqrt{32} = 8\sqrt{2},$$

等号恰好在 $u = \frac{32}{u}$ 时成立，即 $u = 4\sqrt{2}$。这个值在 $[4,8]$ 之内（因为 $4\sqrt{2} \approx 5.66$），所以是容许的。于是 $p = 8 - u$，而当守卫位于以下位置时，可见面积最小：

$$\boxed{p = 8 - 4\sqrt{2} \;\approx 2.34 \text{ 个单位，距离点} A}$$

在这一点，他看不见的面积是 $52 - 8\sqrt{2}$，所以他看见 $76 + 8\sqrt{2} \approx 87.3$ 平方单位。

第 (ii) 部分：最大可见面积

现在我们需要 H 尽可能小，即 $f(u) = u + \frac{32}{u}$ 在 $[4,8]$ 上尽可能大。由于 $f'(u) = 1 - \frac{32}{u^2}$，$f$ 在 $[4, 4\sqrt{2}]$ 上递减，在 $[4\sqrt{2}, 8]$ 上递增，所以它的最大值在端点取得。我们有

$$f(4) = 4 + 8 = 12, \qquad f(8) = 8 + 4 = 12,$$

两者相等。所以 H 在 $u = 8$ 和 $u = 4$ 两处，即 $p = 0$ 和 $p = 4$ 两处，都取得其最小值 $52 - 12 = 40$。可见面积最大，为 $128 - 40 = 88$ 平方单位，此时守卫站在

$$\boxed{\text{位于} A\ (p = 0) \text{ 或} E\ (p = 4)}$$

而 AE 上再没有其他点能做到。

验算

- 当 $p = 0$ 时，我们可以直接手算阴影：两条射线与围墙相交于 $(6, 12)$ 和 $(12, 6)$，而看不见的六边形面积为 40，与 $H(0)$ 吻合。
- 当 $p = 4$ 时，穿过 T 的射线沿直线 $x = 4$ 笔直向上，而 $H(4) = 40$ 再次成立。两个端点的值相等是个好兆头：公式没有偏向一边。
- 我们在 600×600 的网格上计数看不见的点，逐一检验由 P 出发的视线是否穿过塔楼。当 $p = 8 - 4\sqrt{2}$ 时，结果是 40.687，而 $52 - 8\sqrt{2} = 40.686$。当 $p = 1$ 时，结果是 40.429，而 $H(1) = \frac{283}{7} \approx 40.429$。

常见错误

- 用 Z 而不是 T 或 Y 作为擦边点。从左下方望去，塔楼的轮廓由 T 和 Y 这两个顶角决定，而不是由远处的那个顶角决定。
- 做第 (ii) 部分时只求出驻点（stationary point）便停下来。驻点给出的是最小可见面积，所以最大值必定来自端点，而这里有两个。

STEP 之外

这是可见性（visibility）问题的一个小例子，这类问题属于计算几何（computational geometry）的研究范围（“美术馆问题”（art gallery problem）问的是：要看守一个多边形房间的每一点，需要多少名守卫）。当中求最大、最小值的一步，展示了分析学中的一个基本事实：闭区间上的连续函数必定取得其最大值和最小值，取得之处不是驻点便是端点。这就是极值定理（extreme value theorem），其严谨证明可见于 Abbott, *Understanding Analysis*。

详细解答

STEP I 1991，第 5 题

纯数学 • 正则向量组

> **STEP I 1991，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/91-s1-q5

关键思路

条件 (ii) 说这些向量（vector）之和为 $\mathbf{0}$，而要把一条向量等式化为一个数，最省力的方法就是与其中一个向量取点积（dot product），条件 (i) 正好告诉我们怎样计算这些点积。此后，整道题目其实是在计数数目：若一个向量与某几个固定向量的点积全部已被指定，它可处的位置就寥寥无几，所以集合内只容得下少数几个向量。

解答

恒等式 $(n-1)\beta = -\alpha^2$

由 (ii)，$\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_n = \mathbf{0}$。将等式两边与 $\mathbf{a}_1$ 取点积。由 (i)，$\mathbf{a}_1 \cdot \mathbf{a}_1$ 一项等于 α^2，而其余 $n-1$ 项 $\mathbf{a}_1 \cdot \mathbf{a}_j$（$j \neq 1$）每项均为 β，所以

$$\alpha^2 + (n-1)\beta = 0, \qquad \text{即} \qquad (n-1)\beta = -\alpha^2,$$

这正是所要证明的。特别地，$\beta = -\alpha^2/(n-1) < 0$。

二维情况

现在 $\alpha^2 = \mathbf{a}_1 \cdot \mathbf{a}_1 = 1$，所以每个 $\mathbf{a}_i$ 都是单位向量（unit vector），并且

$$\beta = -\frac{1}{n-1}.$$

对 $i \geq 2$，设 $\mathbf{a}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$。条件 $\mathbf{a}_1 \cdot \mathbf{a}_i = \beta$ 即是 $x_i = \beta$，于是 $|\mathbf{a}_i| = 1$ 迫使 $y_i = \pm\sqrt{1-\beta^2}$。因此，除 $\mathbf{a}_1$ 以外的每个向量都必定是以下两者之一：

$$\begin{pmatrix} \beta \\ \sqrt{1-\beta^2} \end{pmatrix}, \qquad \begin{pmatrix} \beta \\ -\sqrt{1-\beta^2} \end{pmatrix}.$$

想一想 知道了候选向量之后，怎样限制 n 的大小?

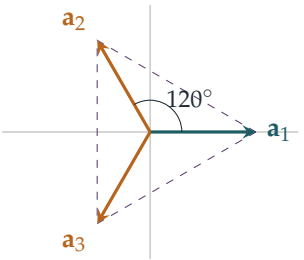
集合中的向量互不相同，所以 $\mathbf{a}_2, \ldots, \mathbf{a}_n$ 是 $n-1$ 个不同的候选向量。候选向量最多只有两个，所以 $n-1 \leq 2$，即 $n \leq 3$。由于 $n \geq 2$，故 $n = 2$ 或 $n = 3$。

$n = 2$。此时 $\beta = -1$，两个候选向量重合，$\mathbf{a}_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$。这确实可行：$\mathbf{a}_1 \cdot \mathbf{a}_2 = -1 = \beta$，而且 $\mathbf{a}_1 + \mathbf{a}_2 = \mathbf{0}$。

$n = 3$。此时 $\beta = -\frac{1}{2}$，$\sqrt{1-\beta^2} = \frac{\sqrt{3}}{2}$，而 $\mathbf{a}_2, \mathbf{a}_3$ 必须是那两个不同的候选向量：

$$\begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix}$$

（次序不拘）。验算：$\mathbf{a}_2 \cdot \mathbf{a}_3 = \frac{1}{4} - \frac{3}{4} = -\frac{1}{2} = \beta$，而三个向量之和为 $\mathbf{0}$。它们指向一个以原点为中心的等边三角形（equilateral triangle）的顶点。



（确定 β 的另一个方法：当 $n \geq 3$ 时，我们有 $-1 < \beta < 0$，所以两个候选向量确实不同，而 $\mathbf{a}_2, \mathbf{a}_3$ 必须分别是这两个向量。于是 $\mathbf{a}_2 \cdot \mathbf{a}_3 = \beta^2 - (1-\beta^2) = 2\beta^2 - 1$，令它等于 β 便得 $(2\beta+1)(\beta-1) = 0$。所以 $\beta = -\frac{1}{2}$，这同样表示 $n = 3$。）

三维情况

同样有 $\alpha = 1$ 及 $\beta = -1/(n-1)$。$n = 2$ 的情况与之前一样：$\beta = -1$，而满足 $\mathbf{a}_1 \cdot \mathbf{a}_2 = -1$ 的单位向量只有 $\mathbf{a}_2 = -\mathbf{a}_1 = (-1, 0, 0)^{\mathrm{T}}$，它位于 x-y 平面上。所以 $\{\mathbf{a}_1, -\mathbf{a}_1\}$ 是其中一个答案。

现在假设 $n \geq 3$，于是 $-\frac{1}{2} \leq \beta < 0$。由于 $\mathbf{a}_2$ 位于 x-y 平面上，沿用二维情况的论证可得

$$\mathbf{a}_2 = \begin{pmatrix} \beta \\ \varepsilon\sqrt{1-\beta^2} \\ 0 \end{pmatrix}, \qquad \varepsilon = \pm 1,$$

而且因为 $\beta \neq \pm 1$，y 分量 $y_2 = \varepsilon\sqrt{1-\beta^2}$ 不等于零。

想一想 对于 $i \geq 3$，这些条件对 $\mathbf{a}_i$ 有甚么要求?

对 $i \geq 3$，设 $\mathbf{a}_i = (x_i, y_i, z_i)^{\mathrm{T}}$。则

$$\begin{aligned} \mathbf{a}_1 \cdot \mathbf{a}_i = \beta &\implies x_i = \beta, \\ \mathbf{a}_2 \cdot \mathbf{a}_i = \beta &\implies \beta^2 + y_2 y_i = \beta \implies y_i = \frac{\beta(1-\beta)}{y_2}, \\ |\mathbf{a}_i| = 1 &\implies z_i^2 = 1 - \beta^2 - y_i^2. \end{aligned}$$

所以对每个 $i \geq 3$，x_i 和 y_i 都相同，而 z_i 只能是两个平方根的其中之一。由于各向量互不相同，$\mathbf{a}_3, \ldots, \mathbf{a}_n$ 当中最多只有两个向量，所以 $n - 2 \leq 2$，即 $n \leq 4$。

$n = 3$。由 (ii)，$\mathbf{a}_3 = -\mathbf{a}_1 - \mathbf{a}_2$，它位于 x-y 平面上。所以这个集合是平面 $z = 0$ 上的一个正则集（regular set），根据二维情况的结果，它就是

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \\ 0 \end{pmatrix},$$

其中 $\mathbf{a}_2$ 可以是后两者的任何一个。（这与上面的公式吻合：$\beta = -\frac{1}{2}$ 给出 $y_3 = -y_2$ 及 $z_3^2 = 1 - \frac{1}{4} - \frac{3}{4} = 0$。）

$n = 4$。此时 $\beta = -\frac{1}{3}$，所以 $\sqrt{1-\beta^2} = \frac{2\sqrt{2}}{3}$，而 $y_2 = \varepsilon\frac{2\sqrt{2}}{3}$。于是

$$y_i = \frac{(-\frac{1}{3})(\frac{4}{3})}{\varepsilon\frac{2\sqrt{2}}{3}} = -\varepsilon\frac{2}{3\sqrt{2}} = -\varepsilon\frac{\sqrt{2}}{3}, \qquad z_i^2 = 1 - \frac{1}{9} - \frac{2}{9} = \frac{2}{3}.$$

由于 $\mathbf{a}_3 \neq \mathbf{a}_4$，它们必须分别取两个相反的值 $z = \pm\sqrt{2/3} = \pm\frac{\sqrt{6}}{3}$。由此得出

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} -\frac{1}{3} \\ \varepsilon\frac{2\sqrt{2}}{3} \\ 0 \end{pmatrix}, \quad \mathbf{a}_3, \mathbf{a}_4 = \begin{pmatrix} -\frac{1}{3} \\ -\varepsilon\frac{\sqrt{2}}{3} \\ \pm\frac{\sqrt{6}}{3} \end{pmatrix},$$

对应 $\varepsilon = \pm 1$ 的每个选择各有一个集合。我们至今只用了涉及 $\mathbf{a}_1$ 和 $\mathbf{a}_2$ 的条件，所以仍须检查其余条件。首先，$\mathbf{a}_3 \cdot \mathbf{a}_4 = \frac{1}{9} + \frac{2}{9} - \frac{6}{9} = -\frac{1}{3} = \beta$。其次是向量之和：$x$ 分量给

出 $1-3\cdot\frac{1}{3}=0$，y 分量给出 $\varepsilon\left(\frac{2\sqrt{2}}{3}-\frac{\sqrt{2}}{3}-\frac{\sqrt{2}}{3}\right)=0$，而 z 分量互相抵销。所以两个集合都是正则集。从几何上看，这四个向量指向一个以原点为中心的正四面体（regular tetrahedron）的顶点。

总结。所有正则集为：$\{\mathbf{a}_1,-\mathbf{a}_1\}$（$n=2$）；$x$-$y$ 平面上的等边三角形（$n=3$）；以及上述两个正四面体（$n=4$，$\varepsilon=\pm1$，两者关于 x-z 平面互为镜像）。不存在 $n\geq 5$ 的正则集。

验算

在每个集合中，向量之和应为零，而且所有两两之间的点积都应相等。对于正四面体，任意两个向量之间的夹角为 $\cos^{-1}(-\frac{1}{3})\approx 109.47°$，即大家熟悉的甲烷键角，而恒等式给出 $(4-1)(-\frac{1}{3})=-1=-\alpha^2$。我们以数值方法检查了全部五个集合（格拉姆矩阵和形心（centroid）准确至 10^{-12}），并在平面上进行穷举搜索，在 n 不超过 11 的范围内，只在 $n=2$ 和 $n=3$ 时找到正则集。

常见错误

- 忘记向量必须*互不相同*：正是这个条件，才把“最多两个候选向量”变成 n 的上界。
- 在 $n=4$ 的情况中，利用 $\mathbf{a}_3,\mathbf{a}_4$ 与 $\mathbf{a}_1,\mathbf{a}_2$ 的点积求出它们之后，没有检查 $\mathbf{a}_3\cdot\mathbf{a}_4=\beta$ 和形心条件。
- 遗漏 $n=2$ 的情况，或遗漏三维情况中正负号 ε 的选择。

STEP 之外

这些集合就是正单纯形（regular simplex）的顶点：在 $(n-1)$ 维空间中的 n 个点，两两之间的距离全部相等。由点积组成的矩阵（matrix）$G=(\mathbf{a}_i\cdot\mathbf{a}_j)$（即*格拉姆矩阵*（Gram matrix））等于 $(\alpha^2-\beta)I+\beta J$，其中 J 是所有元素皆为一的矩阵。它的秩（rank）是 $n-1$，而格拉姆矩阵的秩不能超过空间的维数，由此立即得出在 d 维空间中 $n\leq d+1$。Strang 的 *Introduction to Linear Algebra*（线性代数入门教材）涵盖了秩、点积和格拉姆矩阵。

详细解答

STEP I 1991，第 6 题

纯数学 · 驻点与广义积分

STEP I 1991，第 6 题

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

step.maths.org/questions/91-s1-q6

本题在 *Beyond STEP* 第 3.1 节找出错处中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/spot-the-flaw.html

关键思路

甚么都不要轻信：每一项计算都要亲自重做一遍，并在每个“因此”或“所以”之处问一问，这一步是否真的成立。在 (i) 中，导数（derivative）算错了；而且即使驻点（stationary point）求对了，其后关于极大值、极小值和全局界的推理仍然有毛病。在 (ii) 中，原函数（antiderivative）是错的；更严重的是，被积函数（integrand）在区间内部趋向无穷，所以微积分基本定理（fundamental theorem of calculus）根本不能使用。

解答

第 (i) 部分

第 1 步：驻点。由商法则（quotient rule），

$$g'(x) = \frac{6x^2(x^4+4) - 4x^3(2x^3+3)}{(x^4+4)^2} = \frac{-2x^6 - 12x^3 + 24x^2}{(x^4+4)^2} = \frac{-2x^2\,(x^4+6x-12)}{(x^4+4)^2}.$$

所以 $g'(x) = 0$ 当且仅当 $x = 0$ 或 $h(x) = 0$，其中 $h(x) = x^4 + 6x - 12$。现在 $h(1) = -5$ 而 $h(-1) = -17$，所以 $x = \pm 1$ 并不是驻点：事实上 $g'(1) = \frac{2}{5}$ 而 $g'(-1) = \frac{34}{25}$。原来的论断在第一步就已经不成立。

h 有多少个根？由于 $h'(x) = 4x^3 + 6$ 只在 $x_0 = -(3/2)^{1/3}$ 处为零，h 先递减后递增，所以它最多穿过零两次。另外 $h(0) = -12 < 0$，而且当 $x \to \pm\infty$ 时 $h(x) \to +\infty$，所以

它最少穿过零两次。因此 h 恰好有两个实根 $\beta < 0 < \alpha$。由于 $h(1) < 0 < h(2) = 16$ 及 $h(-3) = 51 > 0 > h(-2) = -8$,

$$1 < \alpha < 2, \qquad -3 < \beta < -2,$$

而数值上 $\alpha \approx 1.386$, $\beta \approx -2.247$。

第 2 步：驻点处的函数值。$g(0) = \frac{3}{4}$ 是对的，$g(1) = 1$ 也是对的，但 $g(-1) = \frac{1}{5}$，而不是 1（何况 ± 1 这两点根本都不是驻点）。在 h 的根 t 处，我们有 $t^4 = 12 - 6t$，所以 $t^3 = (12 - 6t)/t$，并且

$$g(t) = \frac{2(12-6t)/t + 3}{16 - 6t} = \frac{24 - 9t}{t(16-6t)} = \frac{3(8-3t)}{2t(8-3t)} = \frac{3}{2t},$$

其中约去了 $8 - 3t \neq 0$（因为 $t = \frac{8}{3}$ 不是 h 的根）。所以驻点处的函数值是 $g(0) = \frac{3}{4}$、$g(\alpha) = \frac{3}{2\alpha} \approx 1.083$ 和 $g(\beta) = \frac{3}{2\beta} \approx -0.668$。

? **想一想** 如果一个驻点处的函数值比另一个小，较小的那个就一定是极小值吗?

第 3 步：驻点的性质。不一定。比较驻点处的函数值，完全不能告诉我们驻点的性质；我们必须考察 g' 在驻点两侧的正负号（或者考察 g''）。由因式分解后的形式可见，当 $x \neq 0$ 时，$g'(x)$ 的正负号与 $-h(x)$ 的正负号相同，因为 $x^2 > 0$ 而 $(x^4+4)^2 > 0$。由于在 (β, α) 上 $h < 0$，而在 $[\beta, \alpha]$ 之外 $h > 0$:

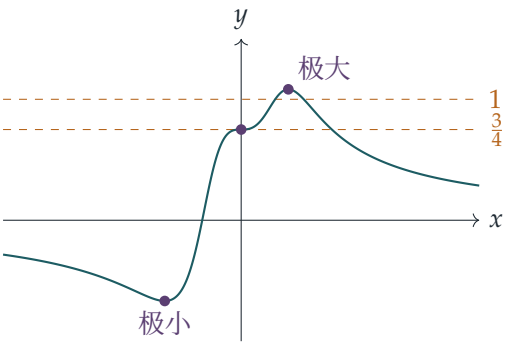
- g 在 $x < \beta$ 时递减，在 (β, α) 上递增，在 $x > \alpha$ 时递减；
- $x = \beta$ 是极小点，而 $x = \alpha$ 是极大点；
- 在 $x = 0$ 处，两侧都有 $g' > 0$，所以 $x = 0$ 是驻拐点（stationary point of inflection），而不是极小点。

第 4 步：全局界。即使局部极大值和极小值都求对了，要得出“对所有 x”都成立的界，还需要更多论据：函数可以在别处超过某个局部极大值，例如当 $x \to \pm\infty$ 时。这里当 $x \to \pm\infty$ 时 $g(x) \to 0$（分母的次数较高），而且由于 $x^4 + 4 > 0$，g 处处连续。把这些与第 3 步的单调性（monotonicity）结合起来，g 由接近 0 的值下降至 $g(\beta)$，再上升至 $g(\alpha)$，然后回落并趋向 0。因此正确的结论是

$$\frac{3}{2\beta} \leq g(x) \leq \frac{3}{2\alpha} \quad \text{对所有} x, \qquad \text{即约为} -0.668 \leq g(x) \leq 1.083,$$

而且两个界都可以取到。

结论。$\frac{3}{4} \leq g(x) \leq 1$ 这个结论在两端都不成立：$g(-1) = \frac{1}{5} < \frac{3}{4}$（其实只要 $2x^3 + 3 < 0$ 就有 $g(x) < 0$），而 $g(\alpha) > 1$ 是因为 $\alpha < \frac{3}{2}$（验证：$h(\frac{3}{2}) = \frac{81}{16} - 3 > 0$，所以 $\alpha < \frac{3}{2}$，从而 $\frac{3}{2\alpha} > 1$）。



第 (ii) 部分

第 1 步：不定积分。题中给出的答案是错的：对 $-3(1-x)^{-4}$ 求导得到 $-12(1-x)^{-5}$，而不是 $(1-x)^{-3}$。幂次应该*增加* 1，由 -3 变为 -2。由于

$$\frac{\mathrm{d}}{\mathrm{d}x}(1-x)^{-2} = -2(1-x)^{-3}\cdot(-1) = 2(1-x)^{-3},$$

正确的结果是

$$\int (1-x)^{-3}\,\mathrm{d}x = \tfrac{1}{2}(1-x)^{-2} + c, \qquad x \neq 1.$$

> **想一想** 用正确的原函数，在 -1 与 3 之间同样得到 $\frac{1}{2}\cdot\frac{1}{4} - \frac{1}{2}\cdot\frac{1}{4} = 0$。那么答案终究还是 0 吗？

第 2 步：定积分。不是。只有当 $F' = f$ 在整个区间上成立，而且 f 在该区间上连续时，计算 $[F(x)]_{-1}^{3}$ 才有根据。这里 $(1-x)^{-3}$ 在 $x = 1$ 处没有定义，而这一点位于 $[-1,3]$ 之内，并且函数在该点附近无界，所以这是一个瑕积分（improper integral），必须定义为两段各自的极限：

$$\int_{-1}^{1-\varepsilon} (1-x)^{-3}\,\mathrm{d}x = \tfrac{1}{2}\left(\varepsilon^{-2} - \tfrac{1}{4}\right) \to +\infty \quad (\varepsilon \to 0^{+}),$$
$$\int_{1+\delta}^{3} (1-x)^{-3}\,\mathrm{d}x = \tfrac{1}{2}\left(\tfrac{1}{4} - \delta^{-2}\right) \to -\infty \quad (\delta \to 0^{+}).$$

两段都发散（diverge），所以这个积分不存在；“$\infty - \infty$”没有值。从几何上看这也合理：被积函数在 $x < 1$ 时为正，在 $x > 1$ 时为负，而两边各自围出的面积都是无穷大。

结论。两步都是错的，而且照原样来说结论并不成立：$\int_{-1}^{3}(1-x)^{-3}\,\mathrm{d}x$ 没有定义。这个“0”只有在挖去一个*对称*的缺口，即 $\varepsilon = \delta$ 时，才算有几分道理：这时两段恰好互相抵消，极限是 0（即柯西主值，Cauchy principal value）。但假如取 $\delta = 2\varepsilon$，两段之和便是 $\frac{1}{2}(\varepsilon^{-2} - \frac{1}{4}\varepsilon^{-2}) \to \infty$，所以答案取决于我们怎样趋近 $x = 1$，这正是这个积分没有定义的原因。

验算

在随机选取的点上，g 的数值导数与因式分解后的公式吻合；$g'(1) = 0.4 \neq 0$。用二分法（bisection）求得 $\alpha \approx 1.3856$，$\beta \approx -2.2467$，并且 $g(\alpha) = 1.0825 = 3/(2\alpha)$，$g(\beta) = -0.6676 = 3/(2\beta)$；在 $[-100, 100]$ 上对 g 取样，可确认这两个值就是全局最大值和最小值。至于 (ii)，在 $[-1, 0.99]$ 上用中点求积法（midpoint quadrature）得到约 4999.87，与 $\frac{1}{2}(10^4 - \frac{1}{4})$ 吻合，而对应的另一段在 $[1.01, 3]$ 上得到 -4999.87。

常见错误

- 靠比较驻点处的函数值来判断驻点的性质。
- 声称得到全局界时，忘记考虑 $x \to \pm\infty$ 时的表现。
- 跨过被积函数趋向无穷的点使用 $[F(x)]_a^b$；务必先检查被积函数在 $[a, b]$ 上是否连续。

STEP 之外

第 (ii) 部分讲的是瑕积分：若积分区间包含奇点（singularity），积分便要定义为极限，而且只有当两个单侧极限各自存在时，积分才收敛（converge）。对称极限，即柯西主值，在复分析（complex analysis）和物理学中确实是很有用的概念，但它是另一回事。第 (i) 部分讲的是局部极值与全局极值之间的差距，数学分析藉极值定理（extreme value theorem）把这一点说得精确。Abbott 的 *Understanding Analysis* 对这两个课题都有仔细的论述。

详细解答

STEP I 1991，第 7 题

纯数学 · 可分离变量微分方程

> **STEP I 1991，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/91-s1-q7

关键思路

两个变化率都是 $(1+V/V_0)^2$ 的常数倍，所以最自然的变量是 $u = 1 + V/V_0$，它满足 $\frac{\mathrm{d}u}{\mathrm{d}t} = \pm c\,u^2$。分离变量（separating the variables）后可见，倒数 $w = 1/u$ 只是随时间线性变化：在增长年以速率 a 下降，在衰退年以速率 b 上升。于是整个问题归结为追踪一条由直线段组成的图像，并回答两个问题：w 会否到达 0（金额变成无限大），又会否到达 1（国库空了）？

解答

建立模型

以年为单位量度 t，并写 $V = kV_0$，于是 $u = 1 + k$，且 $\mathrm{d}V/\mathrm{d}t = V_0\,\mathrm{d}u/\mathrm{d}t$。在增长年，

$$V_0\frac{\mathrm{d}u}{\mathrm{d}t} = aV_0u^2 \quad\Longrightarrow\quad \frac{1}{u^2}\frac{\mathrm{d}u}{\mathrm{d}t} = a \quad\Longrightarrow\quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{u}\right) = -a,$$

同理，在衰退年（只要 $V > 0$），$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{u}\right) = +b$。因此，设

$$w = \frac{1}{1+V/V_0},$$

则 w 在增长期间以速率 a 线性下降，在衰退期间以速率 b 线性上升。注意 $V \geq 0$ 对应 $0 < w \leq 1$：国库空了当且仅当 $w = 1$，而 $V \to \infty$ 当且仅当 $w \to 0^+$。

不等式的意义

开始时 $w = w_0 = \frac{1}{1+k_0}$，而在第一年内 $w = w_0 - at$。于是：

- $k_0 \geq 0$ 只是说国库的存款不可能是负数。
- 金额在整个第一年内保持有限，当且仅当 w 直到 $t = 1$ 都保持正值，即 $w_0 - a > 0$，亦即
$$\frac{1}{1+k_0} > a \iff 1 + k_0 < \frac{1}{a} \iff k_0 < a^{-1} - 1.$$
假如这个条件不成立，V 便会在时间 $t = w_0/a \leq 1$，即第一年之内，变成无限大；对任何描述真实国库的模型来说，这都是荒谬的。

两个不等式合起来还迫使 $a^{-1} - 1 > 0$，即 $a < 1$；我们会在最后用到这一点。

撑过第二年

第一年结束后 $w_1 = \frac{1}{1+k_1}$（事实上 $w_1 = w_0 - a$）。在衰退年内，只要 $w < 1$，就有 $w = w_1 + bt$。国库在年底仍未空，当且仅当 w 到 $t = 1$ 时仍未到达 1：
$$\frac{1}{1+k_1} + b < 1 \iff b < \frac{k_1}{1+k_1}.$$
特别地，我们需要 $b < 1$，这时上式可整理成
$$k_1 > \frac{b}{1-b}.$$
反之，若 $b \geq \frac{k_1}{1+k_1}$，存款会在时间 $t = (1 - w_1)/b \leq 1$ 耗尽，并在该年余下的时间一直维持为零。

第三年的无限增长

> ? **想一想** 假如国库在第二年空了，会怎样？它还有可能在第三年爆破吗？

若国库在第二年年底是空的，第三年便由 $w = 1$ 开始，且 $w = 1 - at$；由于 $a < 1$，它在整年内都保持正值。所以空了的国库不可能在第三年爆破，我们只需考虑存款撑得过第二年的情况。

在这情况下 $w_2 = w_1 + b = \frac{1}{1+k_0} - a + b$，而在第三年 $w = w_2 - at$。金额在第三年内变成无限大，当且仅当 w 在 $t = 1$ 或之前到达 0，即 $w_2 \leq a$：
$$\frac{1}{1+k_0} - a + b \leq a \iff b \leq 2a - \frac{1}{1+k_0}.$$
我们必须检查这与“但不会更早”是否一致。

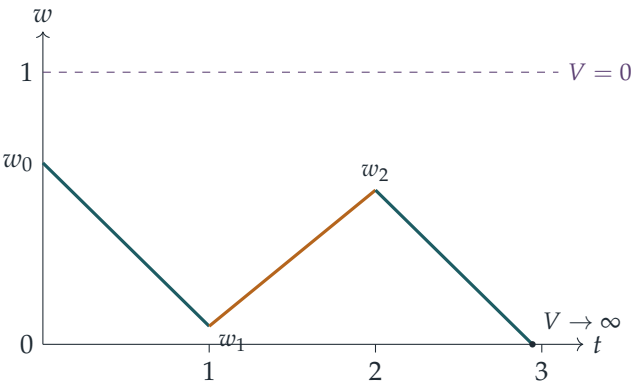
- $k_0 < a^{-1} - 1$ 保证了第一年不会爆破；第二年金额正在减少，也不可能爆破。
- 存款撑得过第二年：若 $w_2 \leq a$，则因 $a < 1$ 而有 $w_2 < 1$，所以存活条件 $w_2 < 1$ 自动成立。

因此，模型预测第三年出现无限增长而不会更早，当且仅当

$$b \le 2a - \frac{1}{1+k_0}.$$

若不等式严格成立，V 会在第三年开始后的时间 $t = w_2/a < 1$ 变成无限大；在等号成立的情况下，$V \to \infty$ 恰好发生在第三年结束之时。（由于 $b > 0$，这只有在 $2a > \frac{1}{1+k_0}$，即 $k_0 > \frac{1}{2a} - 1$ 时才可能发生。）

下图显示 $a = 0.6$、$b = 0.5$、$k_0 = 0.5$ 时 w 对 t 的图像，这是条件成立的一个例子。



验算

在上面的例子中，$k_1 = 14$，存活条件 $14 > \frac{0.5}{0.5} = 1$ 成立；然后 $k_2 = \frac{13}{17} \approx 0.765$，而第三年的爆破发生在 $t = w_2/a \approx 0.944$，与 $2a - \frac{1}{1+k_0} - b = \frac{1}{30} > 0$ 一致。我们亦以数值方法（四阶龙格－库塔法，Runge–Kutta）对这些微分方程（differential equations）作积分，随机选取了 300 组 a、b 和 k_0：k_1 和 k_2 的闭合式与数值结果相差约 10^{-11}，而存活准则和爆破准则在每一个情况下都正确预测了数值结果。

常见错误

- 忘记国库可能在第二年空了。在这题中，由于 $a < 1$，这并无影响，但必须说明。
- 把爆破条件弄反：爆破对应的是 $1/(1+k)$ 到达 0，而不是 k 到达 0。
- 把 $\frac{1}{1+k_1} + b < 1$ 整理成关于 k_1 的条件时，忽略了 $b < 1$ 这个要求。

STEP 之外

方程 $\dot{u} = au^2$ 是有限时间爆破（finite-time blow-up）的标准例子：与指数增长不同，它的解可以在有限时间内变成无限大，所以微分方程的解不一定对所有时间都存在。在大学第一门常微分方程（ODE）课程中，解的局部存在性（local existence）与整体存在性（global existence）之分，正是在这里出现。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）在处理非线性一阶方程时讨论了这一点。

详细解答

STEP I 1991，第 8 题

纯数学 · 代换积分法

> **STEP I 1991，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/91-s1-q8

关键思路

代换（substitution）$y = \pi - x$ 把 $[0, \pi]$ 映射到自身，并且令 $\sin x$ 保持不变，却把碍事的因子 x 变成 $\pi - y$。把原积分与它反射后的版本相加，x 便完全消失。在第 (ii) 部分，被积函数（integrand）由一个函数及其导数构成，因此只要选对一个代换，便可一步到位。

解答

第 (i) 部分

设 $I = \int_0^{\pi} x\,\mathrm{f}(\sin x)\,\mathrm{d}x$。我们需要一个能把 $[0, \pi]$ 两端对调、同时令 $\sin x$ 保持不变的代换，故取 $k = \pi$：令 $x = \pi - y$，于是 $\mathrm{d}x = -\mathrm{d}y$，$\sin x = \sin(\pi - y) = \sin y$，而 $x = 0, \pi$ 分别对应 $y = \pi, 0$。因此

$$\begin{aligned} I &= \int_{\pi}^{0} (\pi - y)\,\mathrm{f}(\sin y)\,(-\mathrm{d}y) = \int_0^{\pi} (\pi - y)\,\mathrm{f}(\sin y)\,\mathrm{d}y \\ &= \pi \int_0^{\pi} \mathrm{f}(\sin y)\,\mathrm{d}y - I, \end{aligned}$$

这是因为 y 只是哑变量（dummy variable），所以最后一个积分就是 I 本身。由此可得

$$I = \frac{\pi}{2} \int_0^{\pi} \mathrm{f}(\sin x)\,\mathrm{d}x.$$

想一想 怎样由 $\frac{\pi}{2}\int_0^{\pi}$ 得出 $\pi \int_0^{\pi/2}$？

由于 $\sin(\pi - x) = \sin x$，$\mathrm{f}(\sin x)$ 在 $[0, \pi]$ 上的图像关于 $x = \frac{1}{2}\pi$ 对称。要严格说明这一

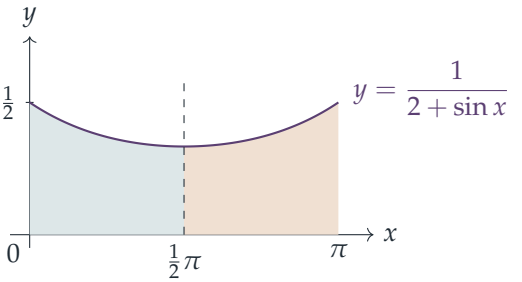
点，可把积分在 $\frac{1}{2}\pi$ 处分成两段，并对后半段使用同一个代换 $x = \pi - y$:

$$\int_{\pi/2}^{\pi} \mathrm{f}(\sin x)\,\mathrm{d}x = \int_{\pi/2}^{0} \mathrm{f}(\sin y)\,(-\mathrm{d}y) = \int_{0}^{\pi/2} \mathrm{f}(\sin y)\,\mathrm{d}y.$$

所以 $\int_0^{\pi} \mathrm{f}(\sin x)\,\mathrm{d}x = 2\int_0^{\pi/2} \mathrm{f}(\sin x)\,\mathrm{d}x$，因此

$$\int_0^{\pi} x\,\mathrm{f}(\sin x)\,\mathrm{d}x = \pi \int_0^{\frac{1}{2}\pi} \mathrm{f}(\sin x)\,\mathrm{d}x,$$

这正是所要证明的结果。



现取 $\mathrm{f}(u) = \dfrac{1}{2+u}$，它在 $u \in [-1,1]$ 上连续（continuous）。于是

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

余下的积分可用 $t = \tan\frac{1}{2}x$ 处理，此时

$$\sin x = \frac{2t}{1+t^2}, \qquad \mathrm{d}x = \frac{2\,\mathrm{d}t}{1+t^2},$$

而 $x = 0, \frac{1}{2}\pi$ 分别对应 $t = 0, 1$。于是

$$\begin{aligned}\int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{2+\sin x} &= \int_0^1 \frac{1}{2+\frac{2t}{1+t^2}} \cdot \frac{2\,\mathrm{d}t}{1+t^2} = \int_0^1 \frac{2\,\mathrm{d}t}{2(1+t^2)+2t} = \int_0^1 \frac{\mathrm{d}t}{t^2+t+1} \\ &= \int_0^1 \frac{\mathrm{d}t}{\left(t+\frac{1}{2}\right)^2 + \frac{3}{4}} = \frac{2}{\sqrt{3}}\left[\tan^{-1}\frac{2t+1}{\sqrt{3}}\right]_0^1 \\ &= \frac{2}{\sqrt{3}}\left(\tan^{-1}\sqrt{3} - \tan^{-1}\frac{1}{\sqrt{3}}\right) = \frac{2}{\sqrt{3}}\left(\frac{\pi}{3} - \frac{\pi}{6}\right) = \frac{\pi}{3\sqrt{3}}.\end{aligned}$$

因此

$$\int_0^{\pi} \frac{x}{2+\sin x}\,\mathrm{d}x = \frac{\pi^2}{3\sqrt{3}} = \frac{\sqrt{3}\,\pi^2}{9}.$$

第 (ii) 部分

想一想 $(\sin^{-1} t)^2$ 的导数是甚么？

根据链式法则（chain rule），

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

这恰好是与余弦相乘的那个因子（的两倍）。所以作代换 $u = (\sin^{-1} t)^2$，其中 $\mathrm{d}u = \dfrac{2\sin^{-1} t}{\sqrt{1-t^2}}\,\mathrm{d}t$。当 $t = 0$ 时，$u = 0$；当 $t = 1$ 时，$\sin^{-1} 1 = \frac{1}{2}\pi$，故 $u = \frac{1}{4}\pi^2$。因此

$$\int_0^1 \frac{(\sin^{-1} t)\cos\left[(\sin^{-1} t)^2\right]}{\sqrt{1-t^2}}\,\mathrm{d}t = \frac{1}{2}\int_0^{\pi^2/4} \cos u\,\mathrm{d}u = \frac{1}{2}\sin\frac{\pi^2}{4}.$$

被积函数在 $t = 1$ 处趋向无穷，所以严格来说这是一个瑕积分（improper integral）。代换把这个问题处理得干净利落：以 $\theta = \sin^{-1} t$ 表示，该积分就是 $\int_0^{\pi/2} \theta\cos\theta^2\,\mathrm{d}\theta$，其被积函数连续，所以积分收敛（converges），而上面求得的值是正确的。（分两步来做，先作 $\theta = \sin^{-1} t$，再作 $u = \theta^2$，也是完全可行的另一种做法。）

答案 $\frac{1}{2}\sin\frac{1}{4}\pi^2$ 是精确值，无法再化简：$\frac{1}{4}\pi^2$ 并不是一个“漂亮”的角，想必这就是考官提醒考生不要动用计算机的原因。

验算

- 以辛普森法则（Simpson's rule）计算，得 $\int_0^{\pi} x/(2+\sin x)\,\mathrm{d}x \approx 1.899406$，而 $\pi^2/(3\sqrt{3}) \approx 1.899406$。
- 快速检查一下是否合理：$\frac{1}{3} \le \frac{1}{2+\sin x} \le \frac{1}{2}$，所以该积分介乎 $\frac{1}{3}\cdot\frac{1}{2}\pi^2 \approx 1.64$ 与 $\frac{1}{2}\cdot\frac{1}{2}\pi^2 \approx 2.47$ 之间。
- 第 (i) 部分的恒等式可用 $\mathrm{f}(u) = 1$ 来检验：两边都等于 $\frac{1}{2}\pi^2$。
- 至于第 (ii) 部分，$\frac{1}{2}\sin(\frac{1}{4}\pi^2) \approx 0.312133$，与 $\int_0^{\pi/2}\theta\cos\theta^2\,\mathrm{d}\theta$ 的数值计算结果吻合。

常见错误

- 做到 $I = \frac{\pi}{2}\int_0^{\pi}\mathrm{f}(\sin x)\,\mathrm{d}x$ 便停下来，没有说明为何可以把积分范围减半。
- 作代换 $t = \tan\frac{1}{2}x$ 时忘记更改积分上下限（上限变成 1，而不是 ∞）。
- 在第 (ii) 部分漏掉因子 $\frac{1}{2}$：$\mathrm{d}u$ 带有一个 2。

STEP 之外

第 (i) 部分的技巧，是利用区间的对称性来计算 $\int x\,g(x)\,\mathrm{d}x$，全程无须求出原函数（antiderivative）。这是一个重要主题的小小例子：要知道一个积分的值，借助对称性往往比直接计算省力得多。第 (ii) 部分则触及瑕积分的理论：变量代换可以把无界的被积函数变成连续函数。Abbott 的 *Understanding Analysis* 严格地建立了变量代换定理（change-of-variables theorem），而 Siklos 的 *Advanced Problems in Mathematics* 则收录了更多这类积分。

详细解答

STEP I 1991，第 9 题

纯数学 · 不等式与调和级数

> **STEP I 1991，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/91-s1-q9

关键思路

在第 (i) 部分，每一条不等式（inequality）都把 x 夹在两个 k 次方根之间，窍门是找出互相矛盾的一个下界（lower bound）和一个上界（upper bound）。要在不用计算机的情况下比较立方根与五次方根，可把两边同时取十五次方。在第 (ii) 部分，我们用（项数）×（最小的项）作为一个长和式的下界。然后把这种形状、互不相交的区块串连起来，使调和级数（harmonic series）要多大有多大。

解答

第 (i) 部分

由于 $x > 1$，x 的每一个幂都是正数，而把两个正数之间的不等式两边同时取正数次方，不等号的方向不变。以下会一直用到这一点。

为了引出矛盾，假设 $n \geq 5$。那么第三条和第五条不等式必然成立：

$$x^3 > 3 \qquad 及 \qquad x^5 < 6.$$

把第一式取五次方，第二式取三次方：

$$x^{15} > 3^5 = 243 \qquad 及 \qquad x^{15} < 6^3 = 216.$$

由此得 $243 < x^{15} < 216$，这是不可能的。所以 $n \geq 5$ 不会发生，因此 $n \leq 4$。

? **想一想** 你怎样知道要选第三条和第五条不等式?

每一条不等式都说明 $k^{1/k} < x < (k+1)^{1/k}$。下界 $k^{1/k}$ 依次为 1、1.41、1.44、1.41、1.38、...，在 $k = 3$ 时最大。上界 $(k+1)^{1/k}$ 则递减，而 $6^{1/5} \approx 1.43$ 是第一个跌到 $3^{1/3} \approx 1.44$

以下的。粗略心算一下便能找到这一对，然后十五次方就能严格地证明它。（$n=4$ 这个值确实可以出现。例如 $x=1.45$ 给出 $x^2=2.1025$、$x^3\approx 3.05$、$x^4\approx 4.42$。）

第 (ii) 部分

设 $n\geq 2$。和式 $\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{n^2}$ 共有 n^2-n 项，每一项都不小于最后一项 $\frac{1}{n^2}$。由于 $n\geq 2$，项数 $n^2-n\geq 2$，所以第一项 $\frac{1}{n+1}$ 严格大于 $\frac{1}{n^2}$。因此

$$\frac{1}{n+1}+\cdots+\frac{1}{n^2}>(n^2-n)\cdot\frac{1}{n^2}=1-\frac{1}{n}.$$

两边同时加上 $\frac{1}{n}$，

$$\frac{1}{n}+\frac{1}{n+1}+\cdots+\frac{1}{n^2}>1,$$

这正是所要证明的。

使调和和超过 10 每一个区块 $\frac{1}{m}+\cdots+\frac{1}{m^2}$（其中 $m\geq 2$）的贡献都大于 1。要同时使用多个区块，它们必须互不重叠，所以我们让每个新区块紧接在上一个区块结束之后开始。定义

$$a_1=2,\qquad a_{k+1}=a_k^2+1,$$

于是 $a_1=2$, $a_2=5$, $a_3=26$, $a_4=677,\ldots$。第 k 个区块由 $\frac{1}{a_k}$ 到 $\frac{1}{a_k^2}$，而下一个区块由 $\frac{1}{a_k^2+1}=\frac{1}{a_{k+1}}$ 开始。所以 $k=1,\ldots,9$ 的各个区块互不相交而且首尾相接，合起来刚好涵盖 $\frac{1}{2},\frac{1}{3},\ldots,\frac{1}{a_9^2}$。因此

$$\sum_{n=1}^{a_9^2}\frac{1}{n}=1+\sum_{k=1}^{9}\left(\frac{1}{a_k}+\cdots+\frac{1}{a_k^2}\right)>1+9=10.$$

想一想 a_9^2 是一个庞大无比的数。我们能否改为给出一个明确的 N？

可以，只要给 a_k 定出上界。我们断言 $a_k<2^{2^k}$。当 $k=1$ 时，由于 $2<4$，断言成立。若 $a_k<2^{2^k}$，则 $a_k\leq 2^{2^k}-1$，所以

$$a_{k+1}=a_k^2+1\leq\left(2^{2^k}-1\right)^2+1=2^{2^{k+1}}-2^{2^k+1}+2<2^{2^{k+1}}.$$

于是由归纳法（induction）可知 $a_9<2^{512}$ 及 $a_9^2<2^{1024}$。再加上更多正数项只会令和变得更大，所以

$$N=2^{1024}\quad\text{给出}\quad\sum_{n=1}^{N}\frac{1}{n}>10.$$

任何更大的 N 同样可行。

另一做法：把各项分成区块 $\frac{1}{2^j+1}+\cdots+\frac{1}{2^{j+1}}$，每个区块有 2^j 项，而且每一项都不小于 $\frac{1}{2^{j+1}}$，由此可见当 $j\geq 1$ 时每个区块都大于 $\frac{1}{2}$，当 $j=0$ 时则等于 $\frac{1}{2}$。所以 $\sum_{n=1}^{2^{18}}\frac{1}{n}>1+\frac{18}{2}=10$，而 $N=2^{18}=262\,144$ 是一个小得多的答案。

验算

- 第 (i) 部分：从数值上看，当 $n=4$ 时，各条不等式剩下 $x\in(3^{1/3},5^{1/4})\approx(1.442,1.495)$。加入第五条不等式后必须有 $x<6^{1/5}\approx 1.431$，于是甚么也不剩。
- 第 (ii) 部分：用准确分数计算，$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=\frac{13}{12}>1$，而且对由 2 到 39 的每一个 n，区块的和都大于 1。
- 调和和在 $N=12\,367$ 时首次超过 10，所以任何正确的答案都必须不小于这个数。2^{18}（此时和约为 13.05）和 2^{1024} 都绰绰有余。

常见错误

- 把不等式取次方时，没有指出两边都是正数（这里由 $x>1$ 保证）。
- 使用区块 $[2,4]$、$[4,16]$、$[16,256]$、...，这些区块共用端点，以致有些项被重复计算了两次。
- 在第 (ii) 部分写了 $\geq$ 而不是 $>$，却没有检查至少有一项严格大于 $\frac{1}{n^2}$。

STEP 之外

第 (ii) 部分是柯西凝聚（Cauchy condensation）这个想法的一个版本。它证明了调和级数发散（diverge），只是发散得慢得惊人：它的部分和（partial sum）的增长有如 $\ln N+\gamma$，其中 $\gamma\approx 0.5772$ 是欧拉－马斯刻若尼常数（Euler–Mascheroni constant）。把和式与积分作比较，便可得出这个增长率，也令最小的 N 容易估算（$e^{10-\gamma}\approx 12\,367$）。Abbott 的 *Understanding Analysis*（一本实分析入门书）对凝聚法和级数的收敛（convergence）有仔细的论述。

详细解答

STEP I 1991，第 10 题

力学 · 悬链的平衡

STEP I 1991，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/91-s1-q10

本题在 *Beyond STEP* 第 11.14 节是抛物线，不是悬链线中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/bridge-parabola.html

关键思路

链上每一个接点都在三个力的作用下处于平衡（equilibrium）：相邻两节链的张力（tension），以及挂在该接点上的重量 w。沿水平方向分解力（resolving），可知每一节链的张力都有相同的水平分量。再沿竖直方向分解，可知每经过一个接点，竖直分量恰好减少 w；由于水平间距固定，各节链的斜率便构成一个等差数列（arithmetic progression）。斜率成等差数列，正是抛物线（parabola）在离散情况下的标记。

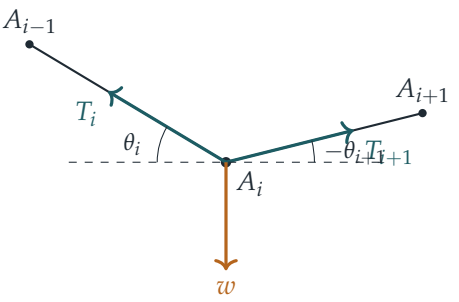
解答

平衡方程

把链视为轻、柔软而且不可伸长（light, flexible and inextensible）的模型，因此每一节链 $A_{i-1}A_i$ 的张力都沿着该节链的方向。设 θ_i 为 $A_{i-1}A_i$ 与水平线所成的角，沿 x 增加的方向向下量度，所以

$$\tan\theta_i = \frac{y_{i-1} - y_i}{h}.$$

（在桥的右半边，θ_i 是负数；以下的论证全都不受其正负号影响。）接点 A_i 受三个力作用：沿 A_iA_{i-1} 方向的 T_i、沿 A_iA_{i+1} 方向的 T_{i+1}，以及吊杆向下的拉力 w。



在 A_i 沿水平和竖直方向分解力，对 $1 \le i \le n+1$（其实对所有 $1 \le i \le 2n+1$ 都成立）：

$$T_i \cos\theta_i = T_{i+1}\cos\theta_{i+1},$$
$$T_i \sin\theta_i - T_{i+1}\sin\theta_{i+1} = w.$$

这就是题目要求的方程。

各个比

第一条方程说明 $T_i\cos\theta_i$ 对每一节链都相同；把这个共同的水平分量记作 H。它是正数，因为链只能拉、不能推。把第二条方程除以 H，并利用 $\tan\theta_i = (y_{i-1}-y_i)/h$，得

$$\frac{H}{h}(y_{i-1}-y_i) - \frac{H}{h}(y_i - y_{i+1}) = w.$$

因此，若以 $V_i = H(y_{i-1}-y_i)/h$ 表示第 i 节链张力的竖直分量，则

$$V_i - V_{i+1} = w \quad (1 \le i \le 2n+1), \qquad \text{所以} \qquad V_i = V_1 - (i-1)w \quad (1 \le i \le 2n+2).$$

想一想 现在只剩下一个未知数 V_1。桥的哪一项条件可以把它确定下来?

两个支点的高度相同。把整条链上各节的下降量加起来，

$$\begin{aligned} 0 = y_0 - y_{2n+2} &= \sum_{i=1}^{2n+2}(y_{i-1}-y_i) = \frac{h}{H}\sum_{i=1}^{2n+2} V_i \\ &= \frac{h}{H}\Big[(2n+2)V_1 - w\big(0+1+\cdots+(2n+1)\big)\Big]. \end{aligned}$$

方括号内的式子是 $(2n+2)V_1 - w(2n+1)(n+1)$，所以 $V_1 = \frac{1}{2}(2n+1)w$。（整体的竖直平衡也指向同一结果：两个支点平均分担总负载 $(2n+1)w$。）因此

$$V_i = \tfrac{1}{2}(2n+3-2i)\,w, \qquad \text{即} \qquad y_{i-1}-y_i = \frac{(2n+3-2i)\,wh}{2H}.$$

取 $i = n+1, n, \ldots, 1$，得到的下降量依次是 $wh/(2H)$ 的 $1, 3, \ldots, 2n+1$ 倍：

$$\frac{h}{y_n - y_{n+1}} = \frac{3h}{y_{n-1}-y_n} = \cdots = \frac{(2n+1)h}{y_0 - y_1} = \frac{2H}{w},$$

这就是所要求的一连串等式。（试卷上最后一个分子印作 $(2n+1)y$；它必定是 $(2n+1)h$，这既符合前面各项的规律，也只有这样，该比才会像其他各个比一样是无量纲的。）

抛物线

设 $k = wh/(2H)$。从 A_0 起把下降量加起来，

$$y_i = y_0 - k\sum_{j=1}^{i}(2n+3-2j) = y_0 - k\big[(2n+3)i - i(i+1)\big] = y_0 - k\,i\,(2n+2-i).$$

记 $m = n+1$，则 $i(2m-i) = m^2 - (i-m)^2$，所以

$$y_i = (y_0 - km^2) + k(i-m)^2 = y_{n+1} + k(i-n-1)^2,$$

这是因为代入 $i = m$ 可知 $y_0 - km^2 = y_{n+1}$。现在 $x_i = x_0 + ih$，所以 $(i-n-1)h = x_i - x_{n+1}$，于是每一点 A_i（$0 \le i \le 2n+2$）都满足

$$y = y_{n+1} + \frac{w}{2Hh}\,(x - x_{n+1})^2.$$

这是同一条抛物线，其竖直的对称轴通过最低点 A_{n+1}，所以 $A_0, A_1, \ldots, A_{2n+2}$ 各点全都在这条抛物线上。两个端点也包括在内：$i = 0$ 和 $i = 2n+2$ 都给出 y_0，正如所料。

验算

- 当 $n = 0$ 时，只有正中间的一根吊杆：公式给出 $V_1 = \frac{1}{2}w$，即链的每一半各承担一半负载，这显然是应有的结果。
- 这个结果与连续的情况吻合：一条水平张力为 H 的缆索，挂上每单位水平长度为 w/h 的均匀负载，会有 $y'' = (w/h)/H$，而我们的抛物线正有 $y'' = w/(Hh)$。
- 我们做了一次数值检验：随机选取 200 座桥（$1 \le n \le 8$，H、w、h 随机选取），以精确分数把各接点的方程当作线性方程组直接求解，并确认了各个比、公式 $y_i = y_0 - k\,i(2n+2-i)$、抛物线，以及每个接点上完整的力平衡。例如当 $n = 2$、$w = 2$、$H = 5$、$h = 1$ 时，下行一段的下降量依次为 1.0, 0.6, 0.2，而每个比都等于 $2H/w = 5$。

常见错误

- 不加说明便以对称为由断言 $y_n = y_{n+2}$。这是对的，但它是由两个支点等高推出来的；上面的求和论证是把它证明出来，而不是假设它成立。
- 忘记张力的水平分量是常数，而尝试逐一求出每个 T_i。本题从头到尾都不需要 T_i 本身（它们等于 $\sqrt{H^2 + V_i^2}$）。

- 在链上升的那一半弄错 θ_i 的正负号。全程使用 $\tan\theta_i = (y_{i-1} - y_i)/h$ 便可避免这个问题。

STEP 之外

只受自身重量而悬垂的链会形成悬链线（catenary）$y = c\cosh(x/c)$；但若链所承受的负载是沿水平方向均匀分布的，例如桥面，则会形成抛物线。本题正是这个事实的离散版本。二阶差分方程 $y_{i-1} - 2y_i + y_{i+1} =$ 常数 是 $y'' =$ 常数 的有限差分（finite difference）形式，而微分方程（differential equations）的数值方法也建基于同一想法。变分法（calculus of variations）处理的正是悬链线及与之相关的曲线。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对绳和缆索的静力学有严谨细致的处理，很值得一读。

详细解答

STEP I 1991，第 11 题

力学 · 滑轮与冲量

STEP I 1991，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/91-s1-q11

关键思路

每个循环分为两个阶段。重锤撞上桩时会立即停下，绳子随之松弛，于是平衡锤向上飞起，再在重力作用下自由落回原处。接着绳子猛然拉紧，两个质量分享动量，然后以第一部分求得的恒定加速度一起运动，直到重锤再次落在桩上。恒定加速度令每个阶段在时间上都对称。因此，一个循环中的每个速率，都是上一个循环中对应速率的固定倍数，等比数列（geometric progression）便随即出现。

解答

恒定加速度

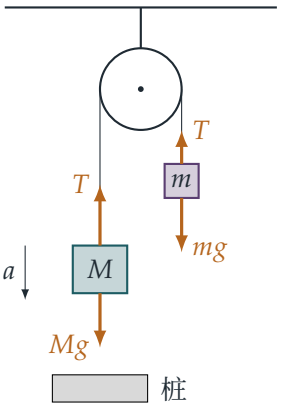
我们把重锤和平衡锤视为质点（particle），并假设绳子轻而不可伸长（light and inextensible），绕过一个光滑的轻滑轮（pulley）。绳子拉紧时，两个质点以相同的速率 v 运动，重锤向下，平衡锤向上。设它们由静止开始移动了距离 x 之后，速率为 v。重锤失去了势能（potential energy）Mgx，平衡锤则增加了 mgx。滑轮光滑，绳子轻而不可伸长，所以没有能量损失，于是

$$\tfrac{1}{2}(M+m)v^2 = (M-m)gx.$$

对时间求导，并利用 $\dot{x} = v$，

$$(M+m)v\dot{v} = (M-m)gv \quad \Longrightarrow \quad \dot{v} = g\,\frac{M-m}{M+m}$$

其中 $v \neq 0$。这是一个常数，记为 a。用牛顿第二定律（Newton's second law）也可得到同样的结果：设 T 为张力（tension），则 $Mg - T = Ma$ 及 $T - mg = ma$，两式相加得 $(M-m)g = (M+m)a$。



从一次碰撞到下一次碰撞

在碰撞前的一刻，重锤以速率 v 向下运动，平衡锤则以相同的速率 v 向上运动。我们分三个阶段追踪这个运动。

撞击。碰撞是完全非弹性（completely inelastic）的，而桩固定不动，所以重锤停下。平衡锤仍以速率 v 向上运动，所以绳子松弛。这时张力帮不了平衡锤，而重锤只受重力作用，被压在桩上，所以重锤留在原处。

自由飞行。绳子松弛时，平衡锤是一个沿铅垂方向运动的抛体（projectile）。它上升后再回到绳子重新拉紧的位置，也就是撞击时它所在的位置。在恒定加速度 g 之下，运动是对称的，所以它回到该处时以速率 v 向下运动，所需时间为

$$\frac{2v}{g}.$$

猛拉。绳子猛然拉紧时，平衡锤以 v 向下运动，重锤则静止。此后两个质量必须以共同的速率 u 运动，平衡锤向下，重锤向上。根据题目容许的假设，平衡锤失去的动量（momentum）等于重锤得到的动量：

$$m(v-u) = Mu \quad \Longrightarrow \quad u = \frac{mv}{M+m}.$$

想一想 为甚么这里不能直接对整个系统使用动量守恒（conservation of momentum）？

两个质量沿相反方向运动，而滑轮又施加了冲量（impulse），所以并不存在一个总动量守恒的方向。真正成立的是：由于绳子是轻的、滑轮是光滑的，冲击张力（impulsive tension）对两个质量的大小同为 J。因此平衡锤沿绳子失去动量 J，重锤沿绳子得到动量 J。这正是题目容许我们假设的那句话。

再次一起运动。现在绳子拉紧，重锤以速率 u 向上运动。根据第一部分，加速度的大小

恒为 a，方向是令较重的质量 M 向下。所以重锤减速至静止，然后落回桩上。恒定加速度令这段运动同样对称：重锤以速率 u 到达桩，所需时间为

$$\frac{2u}{a}.$$

所以在下一次碰撞前的一刻，平衡锤的速率为

$$u = \frac{mv}{M+m},$$

正如所求。

各段时间构成等比数列

记 $k = \dfrac{m}{M+m}$，则 $0 < k < 1$。设 v_0 为第一次撞击前一刻的速率。根据上一部分，第 $(n+1)$ 次撞击前一刻的速率为 $v_n = k^n v_0$。把两个阶段相加，从第 $(n+1)$ 次撞击到下一次撞击的时间为

$$t_n = \frac{2v_n}{g} + \frac{2kv_n}{a} = \left(\frac{2}{g} + \frac{2k}{a}\right) k^n v_0.$$

这是一个常数乘以 k^n，所以相继两次碰撞之间的时间构成等比数列，公比（common ratio）为 $k = m/(M+m)$。

总时间

设 T_0 为从释放到第一次撞击的时间。重锤由静止开始以加速度 a 运动，所以 $v_0 = aT_0$。由于 $0 < k < 1$，这个等比级数（geometric series）收敛（converges），总时间为

$$T_0 + \sum_{n=0}^{\infty} t_n = T_0 + \left(\frac{2}{g} + \frac{2k}{a}\right) \frac{v_0}{1-k}.$$

现在 $1 - k = \dfrac{M}{M+m}$，而 $\dfrac{k}{a} = \dfrac{m}{M+m} \cdot \dfrac{M+m}{g(M-m)} = \dfrac{m}{g(M-m)}$。因此

$$\frac{2}{g} + \frac{2k}{a} = \frac{2}{g}\left(1 + \frac{m}{M-m}\right) = \frac{2M}{g(M-m)},$$

$$\left(\frac{2}{g} + \frac{2k}{a}\right) \frac{v_0}{1-k} = \frac{2M}{g(M-m)} \cdot \frac{M+m}{M} v_0 = \frac{2v_0}{a} = 2T_0.$$

所以碰撞有无限多次，但总共只需有限的时间。重锤最终静止下来所需的时间为

$$T_0 + 2T_0 = 3T_0,$$

即从开始到第一次撞击所需时间的三倍。

验算

- 取极限 $m \to 0$。这时 $k \to 0$ 而 $a \to g$，所以重锤在第一次撞击后不再被提起。余下的时间只有平衡锤的飞行时间 $2v_0/g = 2gT_0/g = 2T_0$，总时间为 $3T_0$，与一般公式所预测的一致。
- 我们以随机的 M、m 和下落高度，用数值方法逐步计算这个运动（例如 $M = 2.403$，$m = 1.571$，故 $k = 0.3953$）。速率之比和时间间隔之比都是 0.3953，而总时间（包括等比级数的尾部）则为 $2.99998\,T_0$。

常见错误

- 忘记了自由飞行的阶段。撞击之后平衡锤继续向上运动，循环中包括它作为抛体所花的时间 $2v/g$。
- 在绳子松弛的阶段用了 a 而不是 g。绳子一旦松弛，平衡锤便只受重力作用。
- 在猛拉的一刻写“总动量守恒”，却把两个速度当作同一方向。滑轮改变了绳子的方向，所以双方分享的是*沿绳子方向*的动量。

STEP 之外

像这部打桩机或一个弹跳的球那样,在有限时间内发生无限多个事件,称为芝诺现象(Zeno behaviour)。这类现象出现在混合系统（hybrid systems）的理论中，当中平滑的运动会被突然的跳跃打断。绳子的猛拉是一个简单的冲击约束（impulsive constraint）。大学一年级的拉格朗日力学（Lagrangian mechanics）会有系统地处理这类约束。想进一步了解冲量、碰撞和受约束运动，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）。

详细解答

STEP I 1991，第 12 题

力学 · 涉及摩擦力的平衡

> **STEP I 1991，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/91-s1-q12

关键思路

木板 AB 是轻的，而且只在两端受力，两端都是光滑铰链（smooth hinge）。所以它施于 CD 的力必定沿 AB 方向。这样问题就简化为单一刚体（rigid body）CD，受三个未知的力作用。对 C 取力矩（taking moments）可求出 AB 中的推力（thrust），再分解力（resolving）便得到 C 处的摩擦力（friction）和法向反作用力（normal reaction）。几何关系（两块木板在 B 相接）把 x 与 α、β 联系起来，正因如此，l 和 x 才会在答案中消去。

解答

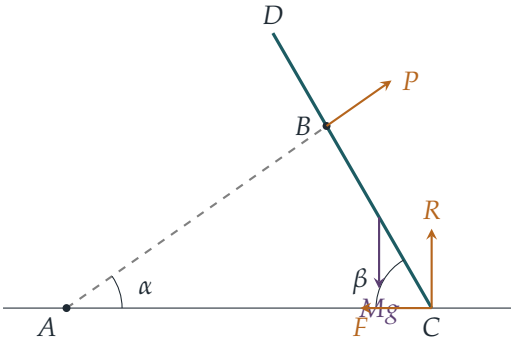
模型

我们把木板视为轻的刚性杆，把油漆工视为质量为 M 的质点（particle），并把 A 和 B 处的铰链视为光滑。地面在 C 处是粗糙的，摩擦系数（coefficient of friction）为 μ。

? **想一想** 为甚么两块木板之间在 B 处的力必定沿 AB 方向?

木板 AB 没有重量，作用在它上面的力只来自 A 和 B 处的光滑铰链。对 A 取力矩，B 处的力对 A 的力矩为零，所以它沿 BA 方向。于是 A 处的力必定与它大小相等、方向相反。根据牛顿第三定律（Newton's third law），AB 在 B 处以一个沿 AB 方向的力 P 推 CD。我们取 P 的方向为由 A 指向 B（即推力），而下面的计算会确认 $P > 0$。

因此作用在 CD 上的力有：B 处的推力 P，与水平线成仰角 α，方向背离 A；油漆工的重量 Mg，作用点沿木板距 C 为 x；以及 C 处向上的法向反作用力 R 和指向 A 的摩擦力 F。摩擦力必须指向 A，因为推力把 C 推离 A。



几何关系

点 B 是 AB 的顶端，所以它的高度是 $2l\sin\alpha$。它沿 CD 与 C 的距离又是 $2x$，所以它的高度也是 $2x\sin\beta$。因此

$$x\sin\beta = l\sin\alpha, \qquad \text{即} \qquad x = \frac{l\sin\alpha}{\sin\beta}.$$

水平距离 AC 是 AB 和 BC 的水平投影之和：

$$AC = 2l\cos\alpha + 2x\cos\beta.$$

对 CD 取力矩和分解力

对 C 取力矩。P 的作用线（line of action）就是直线 AB，它与 C 的垂直距离是 $AC\sin\alpha$。重量的作用线与 C 的水平距离是 $x\cos\beta$。两个力矩的转向相反，所以

$$P\cdot AC\sin\alpha = Mgx\cos\beta \quad\Longrightarrow\quad P = \frac{Mgx\cos\beta}{(2l\cos\alpha + 2x\cos\beta)\sin\alpha} > 0.$$

沿水平和铅垂方向分解力：

$$F = P\cos\alpha, \qquad R = Mg - P\sin\alpha.$$

代入 P，

$$R = Mg\left(1 - \frac{x\cos\beta}{2l\cos\alpha + 2x\cos\beta}\right) = Mg\,\frac{2l\cos\alpha + x\cos\beta}{2l\cos\alpha + 2x\cos\beta},$$
$$F = Mg\,\frac{x\cos\beta\cos\alpha}{(2l\cos\alpha + 2x\cos\beta)\sin\alpha}.$$

特别地，$R > 0$，所以 C 端一直与地面保持接触。两式相除，

$$\frac{F}{R} = \frac{x\cos\alpha\cos\beta}{\sin\alpha\,(2l\cos\alpha + x\cos\beta)}.$$

现在运用几何关系 $x = l\sin\alpha/\sin\beta$。于是 $x\cos\beta = l\sin\alpha\cot\beta$，而

$$\frac{F}{R} = \frac{l\sin\alpha\cot\beta\cos\alpha}{\sin\alpha\,(2l\cos\alpha + l\sin\alpha\cot\beta)} = \frac{\cos\alpha\cot\beta}{2\cos\alpha + \sin\alpha\cot\beta} = \frac{\cot\alpha\cot\beta}{2\cot\alpha + \cot\beta},$$

其中最后一步把分子和分母同除以 $\sin\alpha$。

木板 CD 只有在摩擦力足以提供这个力时才处于平衡（equilibrium），即 $F \le \mu R$。由于 $R > 0$，这就是

$$\mu \ge \frac{\cot\alpha\cot\beta}{2\cot\alpha + \cot\beta},$$

正是所求。（木板 AB 在 A 处以铰链固定，所以无论它需要多大的力，都能保持不动。）

最后部分

若 B 与 D 重合，则 $2x = 2l$，所以 $x = l$。由几何关系得 $\sin\beta = \sin\alpha$，而两个角都是锐角，所以 $\beta = \alpha$。两块木板与地面构成一个等腰三角形，其中 $AC = 4l\cos\alpha$。摩擦条件变成

$$\mu \ge \frac{\cot^2\alpha}{3\cot\alpha} = \tfrac{1}{3}\cot\alpha, \qquad \text{即} \qquad \tan\alpha \ge \frac{1}{3\mu}.$$

所以当且仅当 $\alpha \ge \alpha_0$ 时才可能平衡，其中 $\tan\alpha_0 = 1/(3\mu)$。

想一想 油漆工在哪里？当 α 变化时，距离朝哪个方向改变？

油漆工站在 CD 的中点，与 C 的距离为 l。以 A 为原点，C 位于 $(4l\cos\alpha, 0)$，而油漆工位于

$$(4l\cos\alpha - l\cos\alpha,\ l\sin\alpha) = (3l\cos\alpha,\ l\sin\alpha).$$

所以油漆工与 A 的距离的平方是

$$d^2 = 9l^2\cos^2\alpha + l^2\sin^2\alpha = l^2\left(1 + 8\cos^2\alpha\right).$$

在 $(0, \frac{\pi}{2})$ 上，这个量随 α 增大而减小。所以最大距离出现在容许的最小角度 $\alpha = \alpha_0$。这时 $\tan\alpha_0 = 1/(3\mu)$，所以

$$\cos^2\alpha_0 = \frac{1}{1 + \tan^2\alpha_0} = \frac{9\mu^2}{1 + 9\mu^2},$$

而

$$d_{\max}^2 = l^2\left(1 + \frac{72\mu^2}{1 + 9\mu^2}\right) = l^2\,\frac{1 + 81\mu^2}{1 + 9\mu^2}.$$

因此油漆工与 A 的最大距离是

$$l\sqrt{\frac{1 + 81\mu^2}{1 + 9\mu^2}},$$

正是所求。这个值是可以达到的，这时摩擦力刚好处于极限状态（limiting friction）。

验算

- 最终答案的极限情况：当 $\mu \to 0$ 时，距离趋向 l，对应 $\alpha \to \frac{\pi}{2}$，两块木板都竖直。当 $\mu \to \infty$ 时，距离趋向 $3l$，对应两块木板平放在地上，油漆工位于 $(3l, 0)$。两者都合理。
- 若 $\beta \to \frac{\pi}{2}$，即 CD 竖直，则 $\cot\beta \to 0$，所需的 μ 趋向 0。竖直的木板不需要摩擦力，与此吻合。
- 数值验算：取 $\alpha = 60^\circ$ 和 $\beta = 70^\circ$，直接解 CD 的三条平衡方程，得 $F/R = 0.13837$，与公式相符。取 $\mu = 0.5$，对 α 逐一扫描，得最大距离为 $2.55704\,l$，而公式给出 $l\sqrt{21.25/3.25} = 2.55704\,l$。

常见错误

- 假设 B 处的力是水平或铅垂的。由于 AB 是轻的，而且两端都以光滑铰链连接，这个力必定沿 AB 方向。
- 在 R 中漏掉推力的铅垂分量。这个分量会减小法向反作用力，令所需的摩擦力更难提供。
- 在最后部分，对所有 α 求最大值，而不是只对可能平衡的角度求最大值。

STEP 之外

两端以光滑销钉连接的轻杆称为*二力构件*（two-force member）。辨认出这类构件，是工程静力学中分析销接框架（桁架，truss）的第一步，其中每一根这样的构件只承受纯拉力或纯推力。最后部分是一个小型的约束优化（constrained optimisation）问题：目标函数是单调的，所以最优值落在可行集（feasible set）的边界上。这个主题在线性规划（linear programming）和优化理论中反复出现。想进一步了解刚体的平衡，可参阅 Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP I 1991，第 13 题

力学 · 动量守恒

STEP I 1991，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/91-s1-q13

关键思路

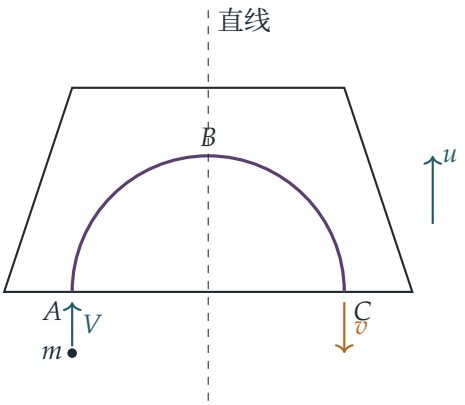
桌面光滑，凹槽也光滑，所以没有任何外力沿直线方向推动系统，也没有任何力作功。由此得到两条守恒定律：沿直线方向的动量（momentum）守恒，以及动能（kinetic energy）守恒。凹槽在出口处的切线与直线平行，所以在出口处两个速度都沿着直线，这时两条定律的用法就与一维弹性碰撞（elastic collision）完全一样。至于固定的薄片（lamina），平均力就是总冲量（impulse）除以所需时间。

关于题目用字，有一点要说明：题目说在 A 和 B 的切线与直线平行，但从图中可见，与直线平行的其实是两端 A 和 C 的切线（在 B 处凹槽与直线成直角）。我们采用 A 和 C，这显然才是题目的原意。

解答

可移动的薄片

把质点（particle）看成点质量，把薄片看成只能沿直线平移的刚体（rigid body），并假设所有接触面都是光滑的。取该直线为轴，以质点在 A 进入凹槽时的运动方向为正方向。



设质点离开 C 时沿直线的速度为 v（以 V 的方向为正），并设薄片此时沿直线的速度为 u。

想一想 为甚么质点在 C 的速度必定沿着直线?

相对于薄片，质点沿凹槽运动，所以它在 C 的相对速度（relative velocity）沿着该处的切线，而这条切线与直线平行。薄片本身只沿直线运动。两者相加，质点在 C 的实际速度同样沿着直线。因此在出口处，每个物体只需要用一个数来描述。

在运动过程中，系统（质点加薄片）所受的外力有两个：一是桌面的法向反作用力（normal reaction），方向垂直于桌面；二是令薄片保持在直线上的力，方向与直线垂直（直线是光滑的）。凹槽对质点的力与质点对凹槽的力都是内力，成对互相抵销。所以沿直线方向的动量守恒：

$$mV = mv + Mu. \tag{1}$$

所有接触面都是光滑的，而且桌面是水平的，所以没有作功，动能守恒。（凹槽的法向反作用力确实分别对质点和薄片作功，但这两个功互相抵销，因为反作用力垂直于相对速度。）所以

$$\tfrac{1}{2}mV^2 = \tfrac{1}{2}mv^2 + \tfrac{1}{2}Mu^2. \tag{2}$$

把 (1) 和 (2) 改写成

$$m(V-v) = Mu, \qquad m(V-v)(V+v) = Mu^2.$$

其中一个解是 $v = V$、$u = 0$：这是在 A 处、一切尚未发生时的情况。在 C 处，质点相对于薄片是向后运动的，所以 $v \neq V$ 而且 $u \neq 0$，于是可以把第二条方程除以第一条，得到

$$V + v = u.$$

（这表示质点相对于薄片的速率在两端都是 V，与弹性碰撞的情况一样。）把 $u = V + v$ 代入 (1)：

$$mV = mv + MV + Mv \quad \Longrightarrow \quad v = -\frac{M-m}{M+m}V, \qquad u = V + v = \frac{2m}{M+m}V.$$

由于 $m < M$，v 是负数：质点从 C 出来时，是沿着来时的方向往回走。所以当质点离开凹槽时，

$$\text{质点的速率} = \frac{M-m}{M+m}V, \qquad \text{薄片的速率} = \frac{2m}{M+m}V,$$

其中薄片沿质点原来的方向运动，质点则沿相反方向运动。

质点确实能够走完整条凹槽：假如它在某一点相对于薄片静止，它的速度就会与薄片相同，这时动能会是 $\frac{1}{2}(mV)^2/(M+m) < \frac{1}{2}mV^2$，与能量守恒矛盾。所以它的相对速率永远不会降至零。

固定不动的薄片

现在薄片不能移动。凹槽对质点的反作用力垂直于质点的速度，所以不作功，质点以恒定速率 V 走过半圆。半圆的长度是 πr，所以质点在凹槽内的时间是

$$T = \frac{\pi r}{V}.$$

质点的动量由向前的 mV 变为向后的 mV，所以薄片给予质点的总冲量是 $2mV$，方向沿直线向后。质点垂直于直线方向的动量在两端都是零，所以冲量的垂直分量是零。

根据牛顿第三定律（Newton's third law），质点给予薄片的冲量是 $2mV$，方向沿直线向前。薄片保持静止，所以栓钉（peg）必须提供相反的冲量，而薄片推向栓钉的总冲量就是 $2mV$，方向与质点原来的运动方向相同。（根据上一段，来自直线的任何横向推力，平均值都是零。）除以 T，栓钉所受的平均力是

$$\frac{2mV}{\pi r/V} = \frac{2mV^2}{\pi r},$$

方向与质点进入凹槽时的方向相同。（题目所说的“average force per unit time”（每单位时间的平均力）正是这个意思：每单位时间的冲量。）

我们也可以直接看出这个结果。质点需要一个指向圆心、大小为 mV^2/r 的力，所以它以 mV^2/r 的力向外推薄片。当连接质点的半径与 CA 所成的角为 θ 时（其中 θ 以均匀的速率由 0 增至 π），这个推力沿直线方向的分量是 $(mV^2/r)\sin\theta$，而 $\sin\theta$ 在 $[0,\pi]$ 上的平均值是 $2/\pi$。

验算

- 若 $M \to \infty$，薄片几乎不动：$u \to 0$，而质点的速率趋向 V，与凹槽固定时的情况一样。
- 若 $m = M$，质点出来时静止，而薄片取得它的全部速率 V，就像两个质量相等的物体之间的弹性碰撞。
- 取 $M = 3$、$m = 1$、$V = 2$：$v = -1$ 而 $u = 1$。动量：$1\cdot(-1) + 3\cdot 1 = 2 = mV$。能量：$\frac{1}{2} + \frac{3}{2} = 2 = \frac{1}{2}\cdot 1\cdot 2^2$。对运动方程作数值模拟（分别用半圆形凹槽和半椭圆形凹

槽），所得数值准确至小数点后六位都相同；而对固定薄片的 $(mV^2/r)\sin\theta$ 作数值积分，得到平均力为 $2mV^2/(\pi r)$，例如当 $m = 0.7$、$r = 1.3$、$V = 2.1$ 时是 1.5117。

常见错误

- 保留 $v = V$、$u = 0$ 这个根：它满足两条方程，但描述的是进入凹槽的一刻，而不是离开的一刻。
- 用了最大的力 mV^2/r，而不是沿直线方向的平均力；横向分量会互相抵销，而沿直线方向的分量只在 B 处才达到 mV^2/r。

STEP 之外

第一部分其实是一次隐藏的碰撞：不论凹槽是甚么形状，重要的只有守恒量和出口方向，这正是不必知道作用力、单凭守恒定律处理散射（scattering）问题背后的想法。在拉格朗日力学（Lagrangian mechanics）中，沿直线的坐标并不出现在拉格朗日量（Lagrangian）之中，所以它的共轭动量（conjugate momentum）自动守恒；这就是对称性与守恒定律之间的联系（诺特定理，Noether's theorem）的起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）正是以这种精神处理动量、冲量和碰撞。

详细解答

STEP I 1991，第 14 题

概率与统计 · 三角形的计数

STEP I 1991，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/91-s1-q14

关键思路

三个长度 $a > b > c$ 能构成非退化三角形（non-degenerate triangle），当且仅当最长的一条严格短于其余两条之和，即 $b + c > a$；另外两个三角不等式（triangle inequality）自动成立。因此固定 a 之后，我们对每个 b 数出夹在 $a - b$ 与 b 之间的 c 值有多少个。这些数目每次递增 2，奇数（或偶数）之和正是由此而来。之后对所有可能的 a 求和，只需用到 $\sum k^2$ 的标准公式。

解答

三角形何时存在？

设 $a > b > c > 0$。非退化三角形须满足的三个三角不等式是

$$b + c > a, \qquad a + c > b, \qquad a + b > c.$$

后两个自动成立，因为 a 本身已经大于 b，也大于 c。所以三根棒能构成非退化三角形，当且仅当 $b + c > a$。（若 $b + c = a$，这个“三角形”是扁平的，面积为零，不计算在内。）

最长的棒长度为偶数

固定 a。我们需要整数 b, c 满足

$$a - b < c < b, \qquad b < a.$$

对于给定的 b，满足 $a - b + 1 \le c \le b - 1$ 的整数 c 共有

$$(b - 1) - (a - b + 1) + 1 = 2b - a - 1,$$

个，前提是这个数为正；否则一个也没有。（留意这里 $c \geq a-b+1 \geq 2$，所以 c 自动是一个真正的棒长。）

现在设 a 为偶数。那么 $2b-a-1$ 是奇数，而它至少为 1 的充要条件是 $b \geq \frac{a}{2}+1$。因此 b 由 $\frac{a}{2}+1$ 取到 $a-1$，得出

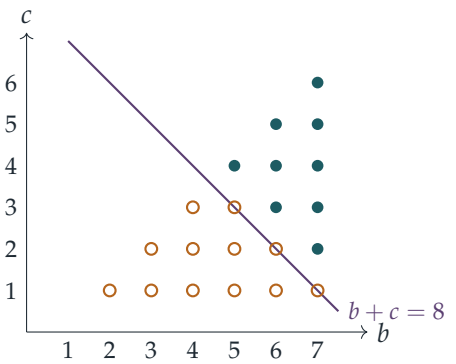
$$\begin{aligned} b = \tfrac{a}{2}+1: &\quad 2b-a-1 = 1, \\ b = \tfrac{a}{2}+2: &\quad 2b-a-1 = 3, \\ &\vdots \\ b = a-1: &\quad 2b-a-1 = a-3. \end{aligned}$$

b 每增加 1，数目便增加 2，所以以这根棒为最长边的三角形数目是

$$1+3+5+\cdots+(a-3),$$

正如题目所求。（当 $a=2$ 时这个和是空的，这是对的：没有三角形的最长边是 2。）首 k 个奇数之和是 k^2，而这里共有 $\frac{a-2}{2}$ 项，所以这个数目是 $\left(\frac{a-2}{2}\right)^2$。

下图显示 $a=8$ 的情况：每一点代表一对 (b,c)，其中 $8>b>c$，而实心点是位于直线 $b+c=8$ 上方的那些点。各列分别有 1、3 和 5 个实心点。



最长的棒长度为奇数

现在设 a 为奇数。那么 $2b-a-1$ 是偶数，而它为正的充要条件是 $b \geq \frac{a+3}{2}$。当 b 由 $\frac{a+3}{2}$ 取到 $a-1$ 时，数目依次为 $2,4,\ldots,a-3$，所以三角形的数目是

$$2+4+6+\cdots+(a-3) = 2\left(1+2+\cdots+\tfrac{a-3}{2}\right) = \frac{a-3}{2}\cdot\frac{a-1}{2} = \frac{(a-1)(a-3)}{4}.$$

（同样，$a=3$ 时正确地得出 0。）

三角形的总数

最长的棒可以是 $3,4,\ldots,2N+1$ 中任何一个（更小的值反正只会得出零）。

偶数 $a=2m$，$m=1,\dots,N$：数目是 $(m-1)^2$，所以这部分合共贡献

$$\sum_{m=1}^{N}(m-1)^2=\sum_{k=1}^{N-1}k^2=\frac{(N-1)N(2N-1)}{6}.$$

奇数 $a=2m+1$，$m=1,\dots,N$：数目是 $\frac{(2m)(2m-2)}{4}=m(m-1)$，所以这部分合共贡献

$$\begin{aligned}\sum_{m=1}^{N}(m^2-m)&=\frac{N(N+1)(2N+1)}{6}-\frac{N(N+1)}{2}\\&=\frac{N(N+1)(2N-2)}{6}=\frac{(N-1)N(N+1)}{3}.\end{aligned}$$

两者相加，能构成非退化三角形的选法数目是

$$T_N=\frac{N(N-1)}{6}\big[(2N-1)+2(N+1)\big]=\frac{N(N-1)(4N+1)}{6}.$$

? **想一想** 为甚么要按 a 的奇偶性（parity）把和式分开，而不是对单一公式求和?

对于给定的 a，当 a 为偶数时数目是 $\frac{(a-2)^2}{4}$，但当 a 为奇数时却是 $\frac{(a-1)(a-3)}{4}=\frac{(a-2)^2-1}{4}$。这是两个不同的 a 的多项式（polynomial），所以写成 $a=2m$ 或 $a=2m+1$ 之后，两者各自变成 m 的简单多项式，可以用标准公式求和。

验证 $N=3$ 的情况

公式给出 $T_3=\frac{3\cdot2\cdot13}{6}=13$。以长度为 $1,\dots,7$ 的棒直接列举，把三角形写成 (a,b,c)：

- $a=4$（一个）：$(4,3,2)$；
- $a=5$（两个）：$(5,4,2),(5,4,3)$；
- $a=6$（四个）：$(6,4,3),(6,5,2),(6,5,3),(6,5,4)$；
- $a=7$（六个）：$(7,5,3),(7,5,4),(7,6,2),(7,6,3),$ $(7,6,4),(7,6,5)$。

合共 $1+2+4+6=13$ 个三角形，与公式吻合（亦与偶数 a 的 $1=1^2,4=2^2$ 及奇数 a 的 $2=2\cdot1,6=3\cdot2$ 吻合）。

概率

随机抽出三根棒而不放回（without replacement），$\binom{2N+1}{3}$ 种无序选法中每一种的可能性均等。由于

$$\binom{2N+1}{3}=\frac{(2N+1)(2N)(2N-1)}{6},$$

构成非退化三角形的概率（probability）是

$$\frac{T_N}{\binom{2N+1}{3}} = \frac{N(N-1)(4N+1)}{(2N+1)\,2N\,(2N-1)} = \frac{(N-1)(4N+1)}{2(4N^2-1)},$$

其中约去了因子 N，并用到 $(2N+1)(2N-1) = 4N^2 - 1$。这就是所求的结果。

验算

- 当 $N = 3$ 时，概率是 $\frac{2\cdot 13}{2\cdot 35} = \frac{13}{35}$，与 $\binom{7}{3} = 35$ 种选法中有 13 个三角形相符。
- 当 $N \to \infty$ 时，概率趋向 $\frac{4N^2}{8N^2} = \frac{1}{2}$。这与连续的版本一致：在 $[0,1]$ 上取三个独立的均匀分布长度，最长者小于其余两者之和的机会是 $\frac{1}{2}$。
- 对 N 由 2 至 29 的每个值，以穷举法（brute force）数遍所有三元组，结果与 T_N 及概率公式完全吻合。

常见错误

- 用了 $b + c \geq a$：等号成立时得到的是扁平三角形，题目已把它排除，于是每个数目都会偏大。
- 把三个三角不等式逐一分开检查，或者忘记了一旦 a 是最长的，就只有 $b + c > a$ 才起作用。
- 把 b 的范围算错一位；应以一个小例子检验，例如 $a = 6$（三角形为 $(6,4,3)$, $(6,5,2)$, $(6,5,3)$, $(6,5,4)$，即 $1 + 3$）。

STEP 之外

像图中那样数算区域内的格点（lattice point），是计算面积的离散版本；而当 N 增大时，这个数目除以 $\binom{2N+1}{3}$ 会趋近单位立方体内的一个体积比。这种由离散均匀分布（uniform distribution）过渡到连续均匀分布的做法，以及折棒问题（broken-stick problem）这类“几何概率”（geometric probability）题目，都是大学一年级的经典课题。想进一步阅读，Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）是个好去处。

详细解答

STEP I 1991，第 15 题

概率与统计 · 正态近似

> **STEP I 1991，第 15 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/91-s1-q15

关键思路

每个 X_j 的平均值(mean)为 0、方差(variance)为 1，而且各次掷币互相独立(independent)，所以 S_n 的平均值为 0、方差为 n；当 n 很大时，中心极限定理（central limit theorem）容许我们把 S_n 视为近似正态分布（normal distribution）。最后一部分看似相仿，却藏着一个陷阱：每个 Y_k 都包含同一次首掷 X_1，所以各个 Y_k 并不独立，而这一次共用的掷币最终支配了整个总分。

解答

每个 X_j 取值 ± 1，概率各为 $\frac{1}{2}$，所以

$$\mathrm{E}(X_j) = 0, \qquad \mathrm{E}(X_j^2) = 1, \qquad \mathrm{Var}(X_j) = 1.$$

各次掷币互相独立，所以 $\mathrm{E}(S_n) = 0$ 及 $\mathrm{Var}(S_n) = n$。根据中心极限定理，当 n 很大时，

$$S_n \approx \mathrm{N}(0,\, n).$$

（等价地，若 H 为掷得正面的次数，则 $H \sim \mathrm{B}(n, \frac{1}{2})$，且 $S_n = H - (n - H) = 2H - n$。）

第 (i) 部分

当 $n = 10\,000$ 时，S_n 的标准差（standard deviation）为 $\sqrt{10\,000} = 100$，所以

$$\mathrm{P}(S_n > 100) \approx \mathrm{P}\left(Z > \frac{100}{100}\right) = 1 - \Phi(1) \approx 0.16.$$

如果想再谨慎一点，可对 H 使用连续性校正（continuity correction）：$S_n > 100$ 即 $2H - 10\,000 > 100$，亦即 $H \geq 5051$。由于 H 的平均值为 5000、标准差为 50，

$$\mathrm{P}(H \geq 5051) \approx \mathrm{P}\left(Z > \frac{5050.5 - 5000}{50}\right) = 1 - \Phi(1.01) \approx 0.156.$$

无论用哪一种方法，答案都约为 0.16。

第 (ii) 部分

标准化（standardising）后，

$$\mathrm{P}(S_n > 0.01n) \approx \mathrm{P}\left(Z > \frac{0.01n}{\sqrt{n}}\right) = 1 - \Phi\left(0.01\sqrt{n}\right).$$

这个值随 n 增大而减小，而且它小于 0.01 当且仅当 $\Phi(0.01\sqrt{n}) > 0.99$。查表得 $\Phi(2.326) = 0.99$，所以我们需要

$$0.01\sqrt{n} > 2.326 \quad \Longleftrightarrow \quad \sqrt{n} > 232.6 \quad \Longleftrightarrow \quad n > 54\,100 \text{（约数）}.$$

所以符合条件的最小 n 大约是 54 000。（若改用 2.33，则得出约 54 300；在这个近似程度下，任何接近 5.4×10^4 的答案都算正确。）

新的计分方法

对于 $2 \leq k \leq n$，首掷与第 k 掷共有四种情况：

首掷	第 k 掷	得分 Y_k	$X_1 + X_k$
H	H	2	$1 + 1 = 2$
T	T	-2	$-1 - 1 = -2$
H	T	0	$1 - 1 = 0$
T	H	0	$-1 + 1 = 0$

最后两栏在每种情况下都相同，所以 $Y_k = X_1 + X_k$，正如题目所求。

平均值与方差。 由线性性质（linearity），$\mathrm{E}(Y_k) = \mathrm{E}(X_1) + \mathrm{E}(X_k) = 0$。由于 $k \geq 2$，X_1 与 X_k 来自不同的两次掷币，因此互相独立，于是

$$\mathrm{Var}(Y_k) = \mathrm{Var}(X_1) + \mathrm{Var}(X_k) = 2.$$

（直接计算：Y_k 取值 2、-2、0 的概率分别为 $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$，所以 $\mathrm{E}(Y_k^2) = 2$。）

极限。 设 $m = n - 1$ 及 $T = Y_2 + Y_3 + \cdots + Y_n$。那么

$$T = \sum_{k=2}^{n} (X_1 + X_k) = mX_1 + S', \qquad \text{其中} S' = X_2 + X_3 + \cdots + X_n.$$

想一想 每个 Y_k 的平均值为 0、方差为 2，就像一个经过缩放的 X_j。为甚么我们不能直接照搬第 (ii) 部分的论证?

因为第 (ii) 部分的论证需要各项互相独立，而各个 Y_k 并不独立：它们全都共用 X_1。事实上，对于 $j \neq k$，$\mathrm{Cov}(Y_j, Y_k) = \mathrm{Var}(X_1) = 1$，而 $\mathrm{Var}(T) = m^2 + m$，并不是 $2m$。所以我们把 X_1 的两个取值分开考虑。留意 S' 是 m 个取值为 ± 1 的项之和，所以 $-m \leq S' \leq m$。

- 若首掷是反面，则 $X_1 = -1$ 且 $T = -m + S' \leq 0$。所以 $T > 0.01m$ 不可能发生。
- 若首掷是正面，则 $X_1 = 1$ 且 $T = m + S'$，所以 $T > 0.01m$ 当且仅当 $S' > -0.99m$。沿用第 (ii) 部分的正态近似，取 $S' \approx \mathrm{N}(0, m)$，

$$\mathrm{P}(S' \leq -0.99m) \approx \Phi\left(-0.99\sqrt{m}\right) \to 0 \quad \text{当} m \to \infty.$$

（其实无须使用近似：根据切比雪夫不等式（Chebyshev's inequality），这个概率至多为 $m/(0.99m)^2 \to 0$。）

由于 S' 与 X_1 独立，

$$\mathrm{P}(T > 0.01m) = \tfrac{1}{2} \cdot 0 + \tfrac{1}{2}\,\mathrm{P}(S' > -0.99m) \;\to\; \tfrac{1}{2} \quad \text{当} n \to \infty.$$

所以该命题是**错误**的：这个概率趋向 $\frac{1}{2}$，而不是 0。用文字来说，当 n 很大时，总分为正基本上就等同于首掷是正面，否则总分便不是正数。

验算

- 第 (i) 部分：对于 $H \sim \mathrm{B}(10\,000, \frac{1}{2})$，精确的二项概率 $\mathrm{P}(H \geq 5051)$ 为 0.15625，与经连续性校正的数值 0.156 吻合。
- 第 (ii) 部分：$\mathrm{P}(S_n > 0.01n)$ 的精确二项数值在 $n = 50\,000$ 时为 0.0125，在 $n = 54\,000$ 附近约为 0.0100，在 $n = 55\,000$ 时为 0.0094，所以临界值确实接近 54 000。
- 最后部分：对 $n \leq 14$ 列举全部 2^n 个结果，所得与 $\frac{1}{2}\,\mathrm{P}(S' > -0.99m)$ 完全一致；在 $n = 14$ 时概率已达 0.49994，而以 $n = 100$ 进行的一次模拟则得出 0.497。

常见错误

- 把 $Y_2, \ldots, Y_n$ 当作互相独立，从而得出 $\mathrm{Var}(T) = 2(n-1)$，令概率看似趋向 0。
- 计算得分的方差时，用了 $\mathrm{Var}(S_n) = n/4$（即正面次数的方差）而不是 n。

STEP 之外

第 (ii) 部分是弱大数定律（weak law of large numbers）的一个定量形式：只要 n 够大，S_n/n 就极不可能远离 0。最后一部分则显示，当各项之间有很强的相依关系时，这条定律会如何失效：这里每一项都共用首掷，所以平均值 $T/m = X_1 + S'/m$ 趋向 X_1 这个随机量，而不是趋向平均值 0。相依序列的大数定律（遍历定理（ergodic theorems）、可交换序列（exchangeable sequences））是大学本科概率论中内容丰富的课题；Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）涵盖了弱大数定律和切比雪夫不等式。

详细解答

STEP I 1991，第 16 题

概率与统计 • 条件概率

STEP I 1991，第 16 题

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

step.maths.org/questions/91-s1-q16

关键思路

把一切按观察时刻 $n = 0, 1, 2, \ldots$ 来整理，即每只爬行动物第一次察看马路之后的分钟数。蟾蜍的等候时间服从几何分布（geometric distribution），所以牠必定会过路。青蛙则不同：牠的规则迫使牠最迟在 $n = 3$ 时作出决定，所以牠的故事只有四个阶段，我们可以逐一列出牠在每个阶段安全过路的概率。有了这张清单，与蟾蜍的比较以及条件概率（conditional probability）都只不过是几个求和而已。

解答

全题假设各次观察时马路的状况互相独立（independent），每次安全的概率为 0.1，并且蟾蜍与青蛙的行动互相独立。这是合理的，因为青蛙是“稍后”才过路，时间并不相同。由于过路所需的时间可以忽略，在第 n 次观察时过路的爬行动物会在时间 n 分钟到达。

蟾蜍

蟾蜍只会在安全时过路，所以牠永远不会被辗过。经过第 $0, 1, \ldots, k$ 次观察后牠仍在等候的概率为 0.9^{k+1}，而这个概率趋向 0。所以牠最终会安全过路，概率为

$$\sum_{k=0}^{\infty} 0.9^k(0.1) = \frac{0.1}{1 - 0.9} = 1.$$

以 T 表示牠过路的时间。那么 $P(T = k) = 0.9^k(0.1)$，另外为方便稍后使用，$P(T > k) = 0.9^{k+1}$。

青蛙

在第 n 次观察时，若牠仍在等候：

- 有 0.1 的概率马路安全，牠安全过路；
- 有 $0.9 \cdot \frac{n}{3}$ 的概率马路不安全但牠照样过路，安全到达的概率为 0.2，被辗过的概率为 0.8；
- 有 $0.9\left(1-\frac{n}{3}\right)$ 的概率牠等待下一次观察。

在 $n=0$ 时牠绝不冒险，而在 $n=3$ 时，不安全也照样过路的机会是 1，所以若牠之前仍未出发，在 $n=3$ 时必定出发。特别地，$n>3$ 的情况永远不会出现。

牠在第 n 次观察时仍在等候的概率为

$$w_0=1,\quad w_1=0.9,\quad w_2=0.9\cdot 0.9\cdot\tfrac{2}{3}=0.54,\quad w_3=0.54\cdot 0.9\cdot\tfrac{1}{3}=0.162.$$

在第 n 次观察时，牠安全到达的概率为 $w_n\left(0.1+0.9\cdot\frac{n}{3}\cdot 0.2\right)=w_n(0.1+0.06n)$。以 s_n 表示这个概率，

$$s_0=1\times 0.1=0.1,\qquad s_1=0.9\times 0.16=0.144,$$
$$s_2=0.54\times 0.22=0.1188,\qquad s_3=0.162\times 0.28=0.04536.$$

因此

$$P(\text{青蛙安全过路})=0.1+0.144+0.1188+0.04536=\boxed{0.40816=\tfrac{2551}{6250}}.$$

牠被辗过的概率就是余下的 $1-0.40816=0.59184$。

两者皆安全，而青蛙较快

设 F 为青蛙安全到达的时间。我们需要两者都安全到达，并且 $F<T$。蟾蜍必定安全，而由独立性，

$$P(\text{青蛙安全过路于}n\text{ 且}T>n)=s_n\,P(T>n)=s_n(0.9)^{n+1}.$$

想一想 为甚么是 $T>n$ 而不是 $T\geq n$?

若两者在同一个观察时刻 n 过路，牠们所用的时间相同，青蛙并非严格地较快。所以蟾蜍在时间 n 的那次观察之后必须仍在等候，其概率为 0.9^{n+1}。对青蛙四个可能的时间求和，

$$\begin{aligned}P&=0.1(0.9)+0.144(0.81)+0.1188(0.729)+0.04536(0.6561)\\&=0.09+0.11664+0.0866052+0.029760696\\&=0.323005896\approx\boxed{0.323}.\end{aligned}$$

写成准确的分数，这就是 $\frac{40375737}{125000000}$。

最后一部分

设 A 为青蛙在第一次观察后 2 分钟仍未到达对面这一事件，R 为牠被辗过这一事件。由于过路是瞬间完成的，在 $n=2$ 那次观察时安全过路的青蛙，在 2 分钟时已经到达。所以 A 是“在 $n=0$、1 或 2 时安全过路”的补事件（complement）：

$$P(A) = 1 - (s_0 + s_1 + s_2) = 1 - 0.3628 = 0.6372.$$

被辗过的青蛙在任何时间都永远不会到达，所以 $R \subseteq A$，并且 $P(R \cap A) = P(R) = 0.59184$。因此

$$P(R \mid A) = \frac{0.59184}{0.6372} = \frac{59184}{63720} = \boxed{\frac{274}{295} \approx 0.929}.$$

这包括了已经被辗过的青蛙（在 $n=1$ 或 $n=2$ 时），以及仍在等候、其后在 $n=3$ 时被辗过的青蛙，正如“在某个阶段”所要求的那样。

（有些读者把这个条件理解为：经过 $n=0$ 和 $n=1$ 两次观察后牠仍未到达，也就是在 $n=2$ 那次察看之前。这种理解得出 $0.59184/(1-0.244) = \frac{137}{175} \approx 0.783$。我们较倾向采用上面的理解，因为恰好在 2 分钟时到达，也算是在牠开始“之后 2 分钟”到达。）

验算

- 青蛙的各种结果必须涵盖所有情况：在 $n=1,2,3$ 时被辗过的概率分别为 $0.9(0.24)$、$0.54(0.48)$ 和 $0.162(0.72)$，即 $0.216+0.2592+0.11664=0.59184$，恰好等于 $1-0.40816$。
- 青蛙安全过路的概率远低于蟾蜍那必然的 1，对没有耐性的一方来说理应如此。
- 以分数作准确枚举，再以蒙地卡罗模拟（Monte Carlo simulation）模拟 10^6 对爬行动物，得出 0.4087、0.3234 和 0.9286，与 0.40816、0.32301 和 0.92881 吻合。

常见错误

- 让青蛙在 $n=0$ 时冒险：按规则，那时 $n/3=0$。
- 忘记青蛙的过程在 $n=3$ 时终止，于是像处理蟾蜍那样为牠求一个几何级数（geometric series）的和。
- 在最后一部分，分母用了 $P(\text{未能安全过路于} n \le 2)$，分子却只放进截至 $n=2$ 为止被辗过的情况。题目问的是在任何阶段被辗过，包括 $n=3$。

STEP 之外

青蛙的旅程是一条小型的马可夫链（*Markov chain*），带有吸收态（absorbing states）（“安全”和“被辗过”），而蟾蜍的等候时间服从几何分布，即无记忆（memoryless）的指数分布（exponential distribution）的离散版本。马可夫链的吸收概率（absorption probabilities）和首达时间（hitting times）是大学第一门概率课程的核心课题。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率入门教科书）对两者都有详细阐述。

详细解答

STEP II 1991，第 1 题

纯数学 · 二次方程与判别式

> **STEP II 1991，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/91-s2-q1

关键思路

一个二次式是某个完全平方（perfect square）的常数倍，当且仅当它有重根（repeated root），亦即它的判别式（discriminant）等于零。其后每一部分，都是把这一个条件应用于系数取决于 λ 的二次式。要找出令判别式等于零的 λ，便会得到一个关于 λ 的二次方程；至于分解式，则是解两条以 f 和 g 为未知量的线性方程而得。

解答

形式 $(*)$ 的条件

展开得

$$a(x+k)^2 = ax^2 + 2akx + ak^2.$$

若 $\mathrm{h}(x)$ 具有这个形式，则比较系数（comparing coefficients）可得 $b = 2ak$ 及 $c = ak^2$，所以

$$b^2 = 4a^2k^2 = 4a \cdot ak^2 = 4ac.$$

反过来说，若 $b^2 = 4ac$，取 $k = \dfrac{b}{2a}$（由于 $a \neq 0$，这是可行的）。于是 $2ak = b$ 且 $ak^2 = \dfrac{b^2}{4a} = c$，所以 $\mathrm{h}(x) = a(x+k)^2$。因此所求条件为

$$b^2 = 4ac,$$

即判别式等于零。

λ 的两个值

我们有

$$\mathrm{f}(x)+\lambda \mathrm{g}(x)=(3+\lambda)x^2+4x-2\lambda.$$

若 $\lambda=-3$，上式变成 $4x+6$。这并不属于形式 $(*)$，因为该形式要求 x^2 的系数不为零。所以 $3+\lambda\neq 0$，而根据上面的条件，我们需要

$$16=4(3+\lambda)(-2\lambda) \iff \lambda^2+3\lambda+2=0 \iff (\lambda+1)(\lambda+2)=0.$$

所以 $\lambda=-1$ 或 $\lambda=-2$，两者皆不等于 -3。具体写出来，就是

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

分解式

想一想 关于 $\mathrm{f}-\mathrm{g}$ 和 $\mathrm{f}-2\mathrm{g}$ 的两项事实，怎样能让我们分别求出 f 和 g?

把上面两条恒等式看成以 $\mathrm{f}(x)$ 和 $\mathrm{g}(x)$ 为未知量的联立线性方程（simultaneous linear equations）。从第一式减去第二式，得

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

然后 $\mathrm{f}(x)=(\mathrm{f}(x)-2\mathrm{g}(x))+2\mathrm{g}(x)=(x+2)^2+4(x+1)^2-2(x+2)^2$，即

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

展开验算：$4(x^2+2x+1)-(x^2+4x+4)=3x^2+4x$，以及 $2(x^2+2x+1)-(x^2+4x+4)=x^2-2$。所以其中一个选择是

$$A=4,\quad B=-1,\quad C=2,\quad D=-1,\quad m=1,\quad n=2.$$

把 m 和 n 的角色对调，便得到等价的答案 $m=2$、$n=1$、$A=-1$、$B=4$、$C=-1$、$D=2$。

最后部分

现在

$$\mathrm{f}(x)+\lambda \mathrm{g}(x)=(3+\lambda)x^2+4x+\lambda\alpha.$$

与之前一样，$\lambda=-3$ 给出线性式 $4x-3\alpha$，它永远不属于形式 $(*)$。当 $\lambda\neq -3$ 时，条件 $b^2=4ac$ 变成

$$16=4(3+\lambda)\lambda\alpha \iff \alpha\lambda^2+3\alpha\lambda-4=0. \qquad (\dagger)$$

我们要求满足 $(\dagger)$ 且 $\lambda\neq -3$ 的 λ 恰好只有一个。

想一想 (†) 会不会有两个根，而其中一个正是不容许的值 $\lambda = -3$?

不会。把 $\lambda = -3$ 代入 (†) 的左方，得 $9\alpha - 9\alpha - 4 = -4 \neq 0$，所以 $\lambda = -3$ 永远不是根。因此，符合要求的 λ 值正好就是 (†) 的根。

- 若 $\alpha = 0$，(†) 变成 $-4 = 0$，没有解。所以 $\alpha \neq 0$。
- 若 $\alpha \neq 0$，(†) 是一个名副其实的关于 λ 的二次方程。它恰好有一个根，当且仅当其判别式等于零：

$$9\alpha^2 + 16\alpha = 0 \iff \alpha(9\alpha + 16) = 0 \iff \alpha = -\tfrac{16}{9},$$

这里用到 $\alpha \neq 0$。

因此

$$\alpha = -\frac{16}{9}.$$

重根为 $\lambda = -\frac{3\alpha}{2\alpha} = -\frac{3}{2}$，而事实上

$$f(x) - \tfrac{3}{2}\left(x^2 - \tfrac{16}{9}\right) = \tfrac{3}{2}x^2 + 4x + \tfrac{8}{3} = \tfrac{3}{2}\left(x + \tfrac{4}{3}\right)^2.$$

验算

代入一个测试值来检查分解式：当 $x = 1$ 时，$4(2)^2 - (3)^2 = 7 = f(1)$，而 $2(2)^2 - 3^2 = -1 = g(1)$。在最后部分，$\frac{3}{2}(x + \frac{4}{3})^2$ 在 $x = 0$ 时的值是 $\frac{3}{2} \cdot \frac{16}{9} = \frac{8}{3}$，与 $\lambda\alpha = (-\frac{3}{2})(-\frac{16}{9})$ 吻合。以电脑测试所有满足 $|p| \leq 60$、$q \leq 12$ 的有理数 $\lambda = p/q$，结果在中间部分恰好找到 $\lambda = -1, -2$，而当 $\alpha = -\frac{16}{9}$ 时只找到 $\lambda = -\frac{3}{2}$。

常见错误

- 忘记 $(*)$ 中 $a \neq 0$ 的要求：当 $\lambda = -3$ 时，该式是线性的，你应该说明为何要排除这个值。
- 在最后部分，只写到“(†) 的判别式等于零”便停下来，没有处理 $\alpha = 0$ 的情况；这时 (†) 根本不是二次方程，而且没有根。

STEP 之外

把 f 和 g 写成两个线性型的平方的组合，就是把一对二次型（quadratic forms）同时对角化（simultaneous diagonalisation）。令 $f + \lambda g$ 退化（degenerate）的 λ 值，就是对应的对称矩阵（symmetric matrices）所满足的方程 $\det(F + \lambda G) = 0$ 的根，这是一个“广义

特征值”（generalised eigenvalue）问题。在最后部分，两个 λ 值重合，这正是这种对角化可能失败的情况。Strang 的 *Introduction to Linear Algebra*（线性代数入门）正是以这种观点处理二次型和特征值。

详细解答

STEP II 1991，第 2 题

纯数学 · 双曲线的渐近线

> **STEP II 1991，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/91-s2-q2

关键思路

二次部分 $3y^2 - 10xy + 3x^2$ 可因式分解（factorise）为 $(3y - x)(y - 3x)$，而这就是整题的关键：只要把因式分解补全，将一次项也包括在内，曲线便成为“两个一次式之积 $= 1$”。沿曲线走到极远处时，其中一个因式必定趋向无限大，另一个因式便必定趋向零；令后者等于零，所得的直线就是一条渐近线（asymptote）。这正是提示所要求的“s 与 t 在极限时的关系”。

解答

切线

把 (†) 对 x 作隐函数微分（implicit differentiation），

$$6y\frac{\mathrm{d}y}{\mathrm{d}x} - 10y - 10x\frac{\mathrm{d}y}{\mathrm{d}x} + 6x + 16\frac{\mathrm{d}y}{\mathrm{d}x} - 16 = 0,$$

所以

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{10y - 6x + 16}{6y - 10x + 16} = \frac{5y - 3x + 8}{3y - 5x + 8}.$$

在 (s, t) 处，斜率（gradient）为 $m = \dfrac{5t - 3s + 8}{3t - 5s + 8}$，而切线（tangent）为 $y = mx + (t - ms)$。把截距（intercept）通分，

$$t - ms = \frac{t(3t - 5s + 8) - s(5t - 3s + 8)}{3t - 5s + 8} = \frac{3t^2 - 10st + 3s^2 + 8t - 8s}{3t - 5s + 8}.$$

由于 (s,t) 在曲线上，$3t^2-10st+3s^2=-16t+16s-15$，所以分子为 $-8t+8s-15=-(8t-8s+15)$。因此切线为

$$y=\left(\frac{5t-3s+8}{3t-5s+8}\right)x-\left(\frac{8t-8s+15}{3t-5s+8}\right),$$

正如所求。（当 $3t-5s+8=0$ 时，切线是铅垂的，此时应把公式理解为去分母后的形式。）

渐近线

尝试把 (†) 写成 $(3y-x+a)(y-3x+b)+k=0$。展开得

$$(3y-x+a)(y-3x+b)=3y^2-10xy+3x^2+(3b+a)y-(b+3a)x+ab,$$

所以我们需要 $3b+a=16$ 及 $b+3a=16$，由此得 $a=b=4$ 及 $ab=16$。因此 (†) 即是

$$(3y-x+4)(y-3x+4)=1.$$

对于点 (s,t)，设

$$u=3t-s+4,\qquad v=t-3s+4,\qquad \text{于是}\qquad uv=1.$$

切线方程中的每一个式子，都是 u、v 与常数的组合。比较 t 和 s 的系数，得

$$\begin{aligned}5t-3s+8&=\tfrac{1}{2}(3u+v),\\3t-5s+8&=\tfrac{1}{2}(u+3v),\\8t-8s+15&=2(u+v)-1=2(u+v)-uv,\end{aligned}$$

其中最后一式用了 $uv=1$。所以切线的斜率和截距为

$$m=\frac{3u+v}{u+3v},\qquad c=\frac{8t-8s+15}{3t-5s+8}=\frac{2(2u+2v-uv)}{u+3v}.$$

想一想 当 $s\to\infty$ 时，u 和 v 会怎样?

消去 t，得 $u-3v=8s-8$，所以 $u-3v\to\infty$。由于 $uv=1>0$，u 和 v 同号。

- 若两者皆为正数，$u-3v\to\infty$ 迫使 $u\to\infty$，于是 $v=1/u\to 0$。
- 若两者皆为负数，$u-3v=3|v|-|u|\to\infty$ 迫使 $|v|\to\infty$，于是 $u=1/v\to 0$。

这就是提示所要的关系：在其中一支上 $t-3s+4\to 0$，在另一支上 $3t-s+4\to 0$。

情况 $v\to 0$，$u\to\infty$。把分子和分母同除以 u，

$$m=\frac{3+v/u}{1+3v/u}\to 3,\qquad c=\frac{2(2+2v/u-v)}{1+3v/u}\to 4,$$

所以切线的极限位置是 $y = 3x - 4$。

情况 $u \to 0$，$v \to \infty$。把分子和分母同除以 v，

$$m = \frac{3u/v + 1}{u/v + 3} \to \tfrac{1}{3}, \qquad c = \frac{2(2u/v + 2 - u)}{u/v + 3} \to \tfrac{4}{3},$$

所以切线的极限位置是 $y = \frac{1}{3}x - \frac{4}{3}$，即 $3y = x - 4$。

因此两条渐近线是 $y = 3x - 4$ 和 $3y = x - 4$。（若改为令 $s \to -\infty$，所得的仍是这两条直线，只是两个情况的角色互换。）留意这两条正是直线 $v = 0$ 和 $u = 0$：从因式分解后的形式一眼便可看出。

夹角与轴

直线 $y = 3x - 4$ 的斜率为 3，所以它与 x 轴的夹角 θ 满足 $\tan\theta = 3$。它与 y 轴的夹角便是 $\frac{\pi}{2} - \theta$，而 $\tan(\frac{\pi}{2} - \theta) = \cot\theta = \frac{1}{3}$。直线 $3y = x - 4$ 的斜率为 $\frac{1}{3}$，所以它与 x 轴的夹角是 $\arctan\frac{1}{3}$，与第一条渐近线和 y 轴的夹角相同。

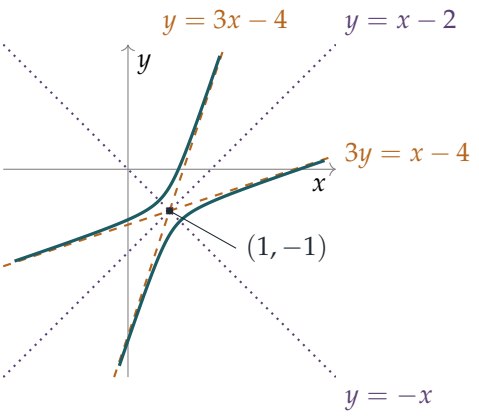
想一想 为甚么由此可知角平分线是甚么?

对直线 $y = x$ 作反射（reflection），会把 x 轴和 y 轴互换；而根据刚才所证，它也会把两条渐近线的方向互换：方向 $(1,3)$ 变成 $(3,1)$。因此两条渐近线之间的夹角关于 $y = x$ 的方向对称，亦即被这个方向平分：两条渐近线的方向与 x 轴分别成 $\arctan 3$ 和 $\arctan\frac{1}{3}$，而这两个角就是 $\frac{\pi}{4} \pm \left(\arctan 3 - \frac{\pi}{4}\right)$。另一条角平分线（bisector）与第一条互相垂直。所以两条角平分线的斜率是 1 和 -1。

双曲线（hyperbola）的轴，就是通过其中心的两条渐近线夹角平分线，而中心是两条渐近线的交点。解 $y = 3x - 4$ 和 $3y = x - 4$：$9x - 12 = x - 4$，所以 $x = 1$，$y = -1$。中心是 $(1,-1)$，而两条轴是

$$y + 1 = x - 1 \iff y = x - 2, \qquad y + 1 = -(x - 1) \iff y = -x.$$

要分辨哪条是哪条：在 $y = -x$ 上有 $u = v = 4 - 4x$，所以由 $uv = 1$ 得 $x = 1 \pm \frac{1}{4}$，顶点（vertices）是 $(\frac{5}{4}, -\frac{5}{4})$ 和 $(\frac{3}{4}, -\frac{3}{4})$。所以 $y = -x$ 是贯轴（transverse axis）。在 $y = x - 2$ 上有 $u = 2x - 2 = -v$，所以 $uv = -(2x - 2)^2 \le 0$，这条直线（即共轭轴（conjugate axis））永不与曲线相交。



验算

- 点 $(\frac{5}{4}, -\frac{5}{4})$ 满足 (†)：$3 \cdot \frac{25}{16} + 10 \cdot \frac{25}{16} + 3 \cdot \frac{25}{16} - 20 - 20 + 15 = 25 - 25 = 0$。
- 以数值检验：取 $u = 10^8$（即 $s \approx 1.25 \times 10^7$），切线公式给出斜率 3.0000000 和截距项 3.99999998；取 $u = -10^{-8}$ 则给出 0.33333333 和 1.33333334。
- 中心 $(1, -1)$ 给出 $u = v = 0$，所以它不在曲线上，这就是切线公式永不出现 $0/0$ 的原因。

常见错误

- 只从 $3k^2 - 10k + 3 = 0$（设 $t \approx ks$）求出斜率 3 和 $\frac{1}{3}$，却忘记截距需要用到下一项：$t = 3s - 4 + o(1)$，而不仅是 $t \approx 3s$。
- 把角平分线取为通过原点，而不是通过中心 $(1, -1)$。

STEP 之外

每一条圆锥曲线（conic）$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0$ 都可写成 $\mathbf{x}^{\mathsf{T}} A \mathbf{x} + \ldots$，其中 $A = \begin{pmatrix} a & h \\ h & b \end{pmatrix}$ 是对称矩阵（symmetric matrix），而曲线的轴沿着 A 的特征向量（eigenvectors）的方向。这里 $A = \begin{pmatrix} 3 & -5 \\ -5 & 3 \end{pmatrix}$ 的特征向量是 $(1, 1)$ 和 $(1, -1)$，对应的特征值（eigenvalues）-2 和 8 正负相反，曲线因而是双曲线。像这样把二次型（quadratic forms）对角化（diagonalising），是大学一年级线性代数（linear algebra）的核心课题；参看 Strang 的 *Introduction to Linear Algebra*（线性代数入门教科书）。

详细解答

STEP II 1991，第 3 题

纯数学 · 对称联立方程

> **STEP II 1991，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/91-s2-q3

关键思路

三个相等的式子给出两条方程，而未知数有三个，所以不能指望求出 x、y、z 本身。我们能做的，是利用当中的循环对称（cyclic symmetry）。把式子两两相减，每个差 $x-y$ 都会变成下一个差 $y-z$ 的倍数；再把三个关系式相乘，几乎所有东西都会约去。至于第二个结果，把共同值记作 k 并消去分母，便得到三个循环的线性关系式，它们以三种不同方式确定 xyz。

解答

证明 $x^2y^2z^2=1$

令首两个式子相等，得

$$x-y=\frac{1}{z}-\frac{1}{y}=\frac{y-z}{yz}.$$

同样地，令第二个与第三个相等，再令第三个与第一个相等，得

$$y-z=\frac{1}{x}-\frac{1}{z}=\frac{z-x}{zx},\qquad z-x=\frac{1}{y}-\frac{1}{x}=\frac{x-y}{xy}.$$

把这三条等式相乘，得

$$(x-y)(y-z)(z-x)=\frac{(y-z)(z-x)(x-y)}{x^2y^2z^2}.$$

由于 x、y 和 z 互不相等（distinct），乘积 $(x-y)(y-z)(z-x)$ 不为零，所以可以把两边同时除以它。这样便剩下 $1=1/(x^2y^2z^2)$，即

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

这正是所要证明的。

> **想一想** 我们究竟在哪里用到了 x、y、z 互不相等且非零这个条件?

非零是为了令分数 $1/x$、$1/y$、$1/z$ 有意义。互不相等则让我们可以约去 $(x-y)(y-z)(z-x)$。没有这个条件，结果便不成立：$x=y=z=2$ 满足各方程，但 $x^2y^2z^2=64$。

共同值为 ±1

设 k 为共同值（common value），则

$$x+\frac{1}{y}=k,\qquad y+\frac{1}{z}=k,\qquad z+\frac{1}{x}=k.$$

各式分别乘以 y、z 和 x（全部非零），得

$$xy=ky-1,\qquad yz=kz-1,\qquad zx=kx-1. \tag{$*$}$$

现在把其中第一式乘以 z，并用第二式代换 yz：

$$\begin{aligned} xyz &= k\,yz - z \\ &= k(kz-1)-z \\ &= (k^2-1)z-k. \end{aligned}$$

作代换 $x\to y\to z\to x$ 时，关系式 $(*)$ 保持不变，所以从第二式和第三式出发作同样的计算，可得

$$xyz=(k^2-1)x-k,\qquad xyz=(k^2-1)y-k.$$

把 xyz 的首两个表达式相减，得

$$(k^2-1)(z-x)=0.$$

由于 $z\neq x$，必有 $k^2=1$，所以 $x+\frac{1}{y}$ 的值不是 $+1$ 就是 -1，这正是所要证明的。

额外的收获是：把 $k^2=1$ 代回去，可得 $xyz=-k$。这是得出第一个结果的另一条途径，因为这样便有 $x^2y^2z^2=k^2=1$。

验算

两个值都确实会出现。当 $k=1$ 时，从任意 $x\neq 0,1$ 出发，取 $y=\frac{1}{1-x}$ 及 $z=\frac{1}{1-y}=\frac{x-1}{x}$。于是 $z+\frac{1}{x}=\frac{x-1}{x}+\frac{1}{x}=1$，所以三个式子都等于 1。例如 $x=2$ 给出 $y=-1$，$z=\frac{1}{2}$：确实有 $2-1=-1+2=\frac{1}{2}+\frac{1}{2}=1$，而且 $xyz=-1=-k$。把 (x,y,z) 换成 $(-x,-y,-z)$，便得到 $k=-1$ 的解，例如 $(-2,1,-\frac{1}{2})$。在这里三个数永远互不相等，因为 $x=\frac{1}{1-x}$ 需要 $x^2-x+1=0$，而这条方程没有实根。我们用精确分数检验了约 4000 组这样的三元组，每次都确认 $x^2y^2z^2=1$ 和 $xyz=-k$。

常见错误

- 除以 $(x-y)(y-z)(z-x)$ 时没有说明它为何不为零。题目给出互不相等的条件，正是为此而设。
- 得出 $xyz = 1$ 而不是 $xyz = \pm 1$ 的结论：事实上 $xyz = -k$，所以当共同值为 1 时，$xyz = -1$。

STEP 之外

映射 $t \mapsto \dfrac{1}{k-t}$ 把 x 变成 y，把 y 变成 z，又把 z 变回 x。它是一个莫比乌斯变换（Möbius transformation），对应的矩阵（matrix）是 $\begin{pmatrix} 0 & 1 \\ -1 & k \end{pmatrix}$，而这类映射的复合正好对应于它们的矩阵相乘。这道题其实是在问：这个映射何时有长度为 3 的循环（cycle）？答案 $k^2 = 1$ 正好就是该矩阵的立方为单位矩阵（identity）的倍数之时。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）有精彩的一章，讲述莫比乌斯变换，以及如何按矩阵的迹（trace）把它们分类。

详细解答

STEP II 1991，第 4 题

纯数学 · 余弦之和

STEP II 1991，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/91-s2-q4

本题在 *Beyond STEP* 第 6.2 节高斯的正十七边形中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch06/gauss-17gon.html

关键思路

由于 $5\phi = 2\pi$ 及 $17\theta = 2\pi$，角的大倍数的余弦会折回小倍数的余弦：当 $n\phi = 2\pi$ 时，$\cos(n-k)\phi = \cos k\phi$。因此，当我们用积化和差公式（product formulae）把余弦之和平方或相乘时，所有项都会化回同样的几个余弦。在后半部分，我们把八个余弦 $\cos\theta, \ldots, \cos 8\theta$ 分成两组，每组四个，求出它们的和与积，从而得到一个以 z 为根（root）的二次方程（quadratic）。

解答

五边形：$\phi = 2\pi/5$

我们使用积化和差公式 $2\cos A\cos B = \cos(A+B) + \cos(A-B)$，以及 $\cos^2 A = \frac{1}{2}(1+\cos 2A)$。由于 $5\phi = 2\pi$，

$$\cos 3\phi = \cos(2\pi - 2\phi) = \cos 2\phi, \qquad \cos 4\phi = \cos(2\pi - \phi) = \cos\phi.$$

因此

$$\begin{aligned} y^2 &= \cos^2\phi + 2\cos\phi\cos 2\phi + \cos^2 2\phi \\ &= \tfrac{1}{2}(1+\cos 2\phi) + (\cos 3\phi + \cos\phi) + \tfrac{1}{2}(1+\cos 4\phi) \\ &= \tfrac{1}{2}(1+\cos 2\phi) + (\cos 2\phi + \cos\phi) + \tfrac{1}{2}(1+\cos\phi) \\ &= 1 + \tfrac{3}{2}(\cos\phi + \cos 2\phi) = 1 + \tfrac{3}{2}y. \end{aligned}$$

两边乘以 2，得 $2y^2 = 3y + 2$，正是所求。

此式可因式分解为 $(2y+1)(y-2) = 0$，所以 $y = -\frac{1}{2}$ 或 $y = 2$。但 $2\phi = \frac{4\pi}{5}$ 介乎 $\frac{\pi}{2}$ 与 π 之间，故 $\cos 2\phi < 0$，而 $\cos\phi \le 1$。于是 $y < 1$，这就排除了 $y = 2$，所以

$$y = \cos\tfrac{2\pi}{5} + \cos\tfrac{4\pi}{5} = -\tfrac{1}{2}.$$

$\theta = 2\pi/17$ 时各余弦之和

设 $\omega = \cos\theta + \mathrm{i}\sin\theta$。根据棣美弗定理（de Moivre's theorem），$\omega^k = \cos k\theta + \mathrm{i}\sin k\theta$，并且 $\omega^{17} = \cos 2\pi + \mathrm{i}\sin 2\pi = 1$，而 $\omega \ne 1$。把等比级数（geometric series）求和，

$$\sum_{k=0}^{16} \omega^k = \frac{\omega^{17} - 1}{\omega - 1} = 0.$$

取实部（real parts），得 $\displaystyle\sum_{k=0}^{16} \cos k\theta = 0$，正是所求。

z 的值

记 $c_k = \cos k\theta$。由于 $17\theta = 2\pi$，而且余弦是偶函数（even function），

$$c_{17-k} = c_k \quad 及 \quad c_{-k} = c_k,$$

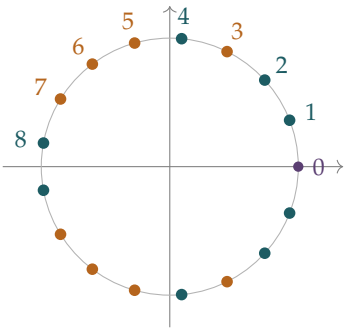
所以每个 c_k 都等于 $c_1, c_2, \ldots, c_8$ 其中之一。特别地，在刚证明的和式中除去 $c_0 = 1$ 后，就得到 $2(c_1 + c_2 + \cdots + c_8) = -1$，所以

$$c_1 + c_2 + \cdots + c_8 = -\tfrac{1}{2}. \qquad (*)$$

z 中的角是 $1, 2, 4, 8$（每个都是前一个的两倍，并按上述方法化简：$16 \to 1$）。其余四个 $3, 5, 6, 7$ 给出与之配对的和

$$w = c_3 + c_5 + c_6 + c_7.$$

下图显示单位圆（unit circle）上的十七个点 ω^k：蓝绿色的点（及其镜像）对应 z 中的余弦，琥珀色的点则对应 w 中的余弦。



和。 由 $(*)$，$z+w=c_1+c_2+\cdots+c_8=-\frac{1}{2}$。

积。 展开 zw 会得到十六个乘积 c_ac_b，其中 $a\in\{1,2,4,8\}$ 而 $b\in\{3,5,6,7\}$，每个乘积都等于 $\frac{1}{2}(c_{a+b}+c_{a-b})$。把每个下标化到 1 至 8 的范围内（利用 $c_{-k}=c_k$ 及 $c_{17-k}=c_k$），所得的各对下标如下：

	$b=3$	$b=5$	$b=6$	$b=7$
$a=1$	4, 2	6, 4	7, 5	8, 6
$a=2$	5, 1	7, 3	8, 4	8, 5
$a=4$	7, 1	8, 1	7, 2	6, 3
$a=8$	6, 5	4, 3	3, 2	2, 1

（例如 $a=8$、$b=7$ 给出 c_{15} 和 c_1，而 $c_{15}=c_2$。）

想一想 我们需要知道出现的是哪些余弦，还是只需知道各出现多少次？

只需知道次数。计数表中各项，下标 $1,2,\ldots,8$ 在 32 个项目中各自恰好出现四次。所以

$$zw=\tfrac{1}{2}\cdot 4\,(c_1+c_2+\cdots+c_8)=2\cdot\left(-\tfrac{1}{2}\right)=-1,$$

这里再次用到 $(*)$。

二次方程。 由于 $z+w=-\frac{1}{2}$ 及 $zw=-1$，z 和 w 这两个数是以下方程的根：

$$t^2+\tfrac{1}{2}t-1=0,\qquad \text{即}\qquad 2t^2+t-2=0,$$

所以 $\{z,w\}=\left\{\dfrac{-1+\sqrt{17}}{4},\ \dfrac{-1-\sqrt{17}}{4}\right\}$。第一个根是正数，第二个是负数。

选取正确的根。 我们证明 $z>0$。由于 $\frac{4}{17}<\frac{1}{4}$，θ 和 2θ 都在 $(0,\frac{\pi}{4})$ 内，所以 c_1 和 c_2 都大于 $\frac{1}{\sqrt{2}}$，而 $c_1+c_2>\sqrt{2}>1$。此外 $4\theta=\frac{8\pi}{17}<\frac{\pi}{2}$，故 $c_4>0$，而且 $c_8\geq -1$。因此

$$z=(c_1+c_2)+c_4+c_8>1+0-1=0,$$

所以 z 必定是正根：

$$z=\frac{-1+\sqrt{17}}{4}=-\frac{1-\sqrt{17}}{4}.$$

验算

以数值计算，$\cos 72^\circ+\cos 144^\circ=0.30902-0.80902=-0.5$。取 $\theta=2\pi/17$，用计算机可得 $z=0.932472+0.739009+0.092268-0.982973=0.780776$，而 $(\sqrt{17}-1)/4=0.780776$；另外 $w=-1.280776=(-1-\sqrt{17})/4$，所以 $z+w=-0.5$ 及 $zw=-1$，与所得结果相符。

常见错误

- 在第一部分忘记排除 $y=2$，或在第二部分忘记排除 w 所对应的根：二次方程给出两个可能的答案，必须另外用正负号的论证来选出其中一个。
- 在 $(*)$ 中把 $k=0$ 也算进去：整个和是 0，但 $c_0=1$，所以 $c_1+\cdots+c_{16}=-1$，除以二才得 $-\frac{1}{2}$。
- 在乘积表中化简下标时出错，例如把 c_9 当作新的一项，而不是 c_8。

STEP 之外

这就是高斯（Gauss）的著名计算，它证明了正 17 边形可以用直尺和圆规作出。按 3 的幂（模 17 的一个生成元（generator））把各余弦分成层层嵌套、大小依次为 8、4、2 和 1 的组，就得到一连串二次方程，所以 $\cos\frac{2\pi}{17}$ 可以只用平方根写出；z 和 w 是这一连串方程的第一步，称为高斯周期（Gaussian periods）。完整的故事属于伽罗瓦理论（Galois theory）：参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）；至于单位根（roots of unity）的算术，可参阅 Davenport 的 *The Higher Arithmetic*（高等算术）。

详细解答

STEP II 1991，第 5 题

纯数学 · 积分归约公式

> **STEP II 1991，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/91-s2-q5

关键思路

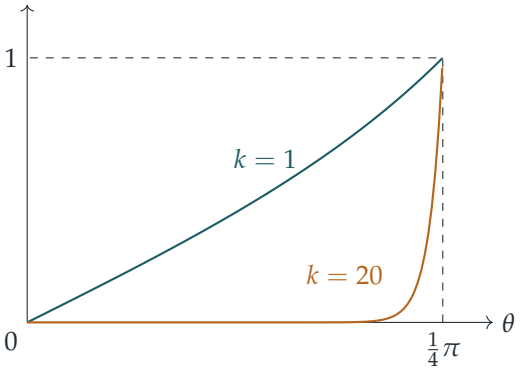
记 $I_k = \int_0^{\pi/4} \tan^k\theta\,\mathrm{d}\theta$。tan 的奇数次幂和偶数次幂都不能直接积分，但把相隔两次的两个幂相加就可以：$\tan^k\theta + \tan^{k+2}\theta = \tan^k\theta\sec^2\theta$，而这正是 $\tan^{k+1}\theta/(k+1)$ 的导数（derivative）。由此得到的归约公式（reduction formula），用归纳法（induction）逐层展开，便得出所求的公式；而 k 很大时的草图告诉你积分会缩小至零，正是这一点把公式变成无穷级数（infinite series）的和。

解答

草图

在 $[0, \frac{1}{4}\pi]$ 上，$0 \le \tan\theta \le 1$，且 $\tan 0 = 0$，$\tan\frac{1}{4}\pi = 1$。当 $k = 1$ 时，图像由 $(0,0)$ 上升至 $(\frac{1}{4}\pi, 1)$，并稍微向上弯曲（起点的斜率为 1，终点的斜率为 $\sec^2\frac{1}{4}\pi = 2$）。

当 $k \gg 1$ 时，$\tan\theta$ 每一个小于 1 的值经过高次乘方后都被压向 0，而两端的值 0 和 1 则保持不变。因此 $\tan^k\theta$ 在区间的大部分范围内都紧贴横轴，到了非常接近 $\theta = \frac{1}{4}\pi$ 之处才急升至 1。事实上，设 $\theta = \frac{1}{4}\pi - u$，其中 u 很小，则 $\tan\theta \approx 1 - 2u$，所以 $\tan^k\theta \approx (1-2u)^k \approx \mathrm{e}^{-2ku}$：这个尖峰的宽度约为 $1/(2k)$，其下的面积约为 $1/(2k)$。



归约公式

对 $k \geq 0$,

$$I_k + I_{k+2} = \int_0^{\pi/4} \tan^k\theta\,(1+\tan^2\theta)\,\mathrm{d}\theta = \int_0^{\pi/4} \tan^k\theta\sec^2\theta\,\mathrm{d}\theta = \left[\frac{\tan^{k+1}\theta}{k+1}\right]_0^{\pi/4} = \frac{1}{k+1}.$$

我们还需要两个初始值:

$$I_0 = \tfrac{1}{4}\pi, \qquad I_1 = \int_0^{\pi/4} \frac{\sin\theta}{\cos\theta}\,\mathrm{d}\theta = \Big[-\ln\cos\theta\Big]_0^{\pi/4} = -\ln\tfrac{1}{\sqrt{2}} = \tfrac{1}{2}\ln 2.$$

奇数次幂

设 $J_n = I_{2n+1}$。在归约公式中取 $k = 2n-1$，得

$$J_n = \frac{1}{2n} - J_{n-1} \qquad (n \geq 1).$$

我们用归纳法证明所求的公式。当 $n=1$ 时: $J_1 = \frac{1}{2} - J_0 = \frac{1}{2} - \frac{1}{2}\ln 2 = (-1)^1\left(\frac{1}{2}\ln 2 - \frac{1}{2}\right)$，这正是 $n=1$ 时的公式。

假设公式对 $n-1 \geq 1$ 成立。则

$$\begin{aligned} J_n &= \frac{1}{2n} - (-1)^{n-1}\left(\tfrac{1}{2}\ln 2 + \sum_{m=1}^{n-1}\frac{(-1)^m}{2m}\right) \\ &= (-1)^n\left(\tfrac{1}{2}\ln 2 + \sum_{m=1}^{n-1}\frac{(-1)^m}{2m}\right) + (-1)^n\frac{(-1)^n}{2n} \\ &= (-1)^n\left(\tfrac{1}{2}\ln 2 + \sum_{m=1}^{n}\frac{(-1)^m}{2m}\right), \end{aligned}$$

其中利用了 $(-1)^n(-1)^n = 1$，把 $\frac{1}{2n}$ 写成括号内 $m=n$ 的一项。根据归纳法，公式对所有正整数 n 成立，证毕。

第一个级数

> **想一想** 我们怎样知道 $J_n \to 0$，而不只是“从草图看来很小”？

我们需要一个确确实实的上界（bound）。在 $[0, \frac{1}{4}\pi]$ 上，$\tan\theta \geq 0$ 且 $\sec^2\theta \geq 1$，所以

$$0 \leq J_n = \int_0^{\pi/4} \tan^{2n+1}\theta \,\mathrm{d}\theta \leq \int_0^{\pi/4} \tan^{2n+1}\theta \sec^2\theta \,\mathrm{d}\theta = \frac{1}{2n+2}.$$

因此当 $n \to \infty$ 时 $J_n \to 0$，这正是 k 很大时的草图所提示的（而且上界 $1/(2n+2)$ 与尖峰面积约为 $1/(2k)$ 相符）。由于 $|J_n|$ 就是公式中括号的绝对值，

$$\left|\tfrac{1}{2}\ln 2 + \sum_{m=1}^{n} \frac{(-1)^m}{2m}\right| = J_n \leq \frac{1}{2n+2} \to 0,$$

所以部分和（partial sum）$\sum_{m=1}^{n} \frac{(-1)^m}{2m}$ 收敛（converge）至 $-\frac{1}{2}\ln 2$。乘以 -1，得

$$\sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{2m} = \tfrac{1}{2}\ln 2.$$

第二个级数

现在对偶数次幂照样做一遍。设 $K_n = I_{2n}$。在归约公式中取 $k = 2n-2$，得

$$K_n = \frac{1}{2n-1} - K_{n-1} \qquad (n \geq 1), \qquad K_0 = \tfrac{1}{4}\pi.$$

完全相同的归纳法（这次由 $K_1 = 1 - \frac{1}{4}\pi$ 开始）给出

$$K_n = \int_0^{\pi/4} \tan^{2n}\theta \,\mathrm{d}\theta = (-1)^n\left(\tfrac{1}{4}\pi + \sum_{m=1}^{n} \frac{(-1)^m}{2m-1}\right):$$

归纳步骤是 $K_n = \frac{1}{2n-1} - (-1)^{n-1}(\cdots) = (-1)^n(\cdots) + (-1)^n\frac{(-1)^n}{2n-1}$，与之前一样。同样的上界给出

$$0 \leq K_n \leq \int_0^{\pi/4} \tan^{2n}\theta \sec^2\theta \,\mathrm{d}\theta = \frac{1}{2n+1} \to 0,$$

所以 $\sum_{m=1}^{n} \frac{(-1)^m}{2m-1} \to -\frac{1}{4}\pi$，即

$$\sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{2m-1} = 1 - \frac{1}{3} + \frac{1}{5} - \cdots = \frac{\pi}{4}.$$

验算

- 取 $n = 1$：$\int_0^{\pi/4} \tan^3\theta \,\mathrm{d}\theta = \frac{1}{2} - \frac{1}{2}\ln 2 \approx 0.15343$，而对该积分使用森逊法则（Simpson's rule）所得的结果，准确至小数点后十位都相同。我们用这个方法验算了 $n = 1, \ldots, 11$ 时 I_{2n+1} 和 I_{2n} 的公式。

- 当 k 很大时，数值计算得 $I_{100} \approx 0.0049995$ 及 $I_{1000} \approx 0.00049999996$，非常接近 $1/(2k)$，与草图的预测一致。
- 两个级数各取一百万项，得 $0.346573\ldots$ 和 $0.785398\ldots$，而 $\frac{1}{2}\ln 2 = 0.346574$，$\frac{1}{4}\pi = 0.785398$。误差大约等于下一项的大小，与上面的上界所提示的一致。

常见错误

- 只说“$\tan^{2n+1}\theta \to 0$，所以积分趋向 0”而不给出上界。被积函数（integrand）在 $\theta = \frac{1}{4}\pi$ 处并不趋向 0，所以必须提出理由，例如利用 $\sec^2\theta \geq 1$ 所得的上界。
- 归纳过程中的正负号错误：把 $(-1)^n$ 留在括号外，并以直接计算核对 $n = 1$ 的情况。

STEP 之外

第一个级数是交错调和级数（alternating harmonic series）$1 - \frac{1}{2} + \frac{1}{3} - \cdots = \ln 2$ 的一半，第二个则是 π 的莱布尼兹（Leibniz）级数。两者都只是条件收敛（conditionally convergent）：重新排列各项可以改变级数的和，这个令人惊讶的事实（黎曼（Riemann）重排定理）在大学第一门分析课程的初段便会遇到。这里的论证，即写出一个精确的误差项再把它夹逼至零，与带余项的泰勒（Taylor）定理是同一个想法。Abbott 的 *Understanding Analysis* 对两者都有仔细的论述。

详细解答

STEP II 1991，第 6 题

纯数学 · 积分的界限

STEP II 1991，第 6 题

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

step.maths.org/questions/91-s2-q6

关键思路

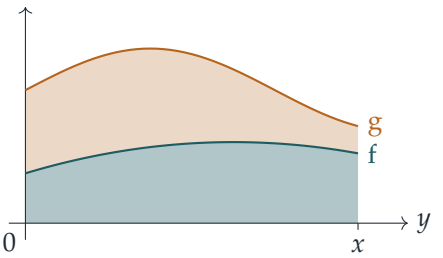
由 0 到 x 的积分（integral）就是面积，而一个较小的非负函数所围出的面积也较小。只要反复运用这一个事实，就能把已知的不等式（inequality）变成新的不等式。把 $0 \le \cos y \le 1$ 积分，得到 $\sin x$ 的一个界（bound）；再把它积分，得到 $\cos x$ 的一个界；再积分一次，便得到 $\sin x$ 更精确的界。之后每个积分的界，都是把被积函数（integrand）夹在两个可以准确积分的函数之间而得出的。

解答

比较积分

由于在 $[0, x]$ 上 $\mathrm{f}(y) \ge 0$，f 的图像位于 y 轴之上或刚好在轴上，所以 $\int_0^x \mathrm{f}(y)\,\mathrm{d}y$ 是图像与轴之间的面积，必为非负。由于 $\mathrm{f}(y) \le \mathrm{g}(y)$，f 之下的区域包含在 g 之下的区域之内，所以它的面积不会更大：

$$0 \le \int_0^x \mathrm{f}(y)\,\mathrm{d}y \le \int_0^x \mathrm{g}(y)\,\mathrm{d}y.$$



（以分析方式说明：$\mathrm{g} - \mathrm{f} \ge 0$，所以 $\int_0^x (\mathrm{g} - \mathrm{f})\,\mathrm{d}y \ge 0$，而积分是线性的。）

$\sin x$ 与 $\cos x$ 的界

取 $0 \le x \le \frac{1}{2}\pi$。对于 $0 \le y \le x$，我们有 $0 \le \cos y \le 1$，所以在上述结果中取 $\mathrm{f} = \cos$、$\mathrm{g} = 1$，得

$$0 \le \int_0^x \cos y \,\mathrm{d}y \le \int_0^x 1 \,\mathrm{d}y, \qquad \text{即} \qquad 0 \le \sin x \le x.$$

这对 $[0, \frac{1}{2}\pi]$ 内的每一个 x 都成立，所以特别地，当 $0 \le y \le x$ 时有 $0 \le \sin y \le y$。再次应用该结果，这次取 $\mathrm{f} = \sin$、$\mathrm{g}(y) = y$：

$$\int_0^x \sin y \,\mathrm{d}y \le \int_0^x y \,\mathrm{d}y \quad \Longrightarrow \quad 1 - \cos x \le \tfrac{1}{2}x^2 \quad \Longrightarrow \quad \cos x \ge 1 - \tfrac{1}{2}x^2.$$

第一个积分

在 $[0, \frac{1}{10}]$ 上（完全位于 $[0, \frac{1}{2}\pi]$ 之内），我们现在有 $1 - \frac{1}{2}x^2 \le \cos x \le 1$，所以

$$3 - \tfrac{1}{2}x^2 \le 2 + \cos x \le 3.$$

> **想一想** 我们可以直接平方再取倒数吗?

可以，因为这里每一个量都是正数：$3 - \frac{1}{2}x^2 \ge 3 - \frac{1}{200} > 0$。对正数而言，平方会保持大小次序，取倒数（reciprocal）则会把次序倒转。然后乘以 $x \ge 0$，得

$$\frac{x}{9} \le \frac{x}{(2+\cos x)^2} \le \frac{x}{(3 - \frac{1}{2}x^2)^2} \qquad (0 \le x \le \tfrac{1}{10}).$$

三个函数全都非负，所以我们可以由 0 积分到 $\frac{1}{10}$。左边给出

$$\int_0^{1/10} \frac{x}{9} \,\mathrm{d}x = \frac{1}{18} \cdot \frac{1}{100} = \frac{1}{1800}.$$

至于右边，分子 x 除正负号外正是 $3 - \frac{1}{2}x^2$ 的导数（derivative），而 $\frac{\mathrm{d}}{\mathrm{d}x}\left(3 - \frac{1}{2}x^2\right)^{-1} = x\left(3 - \frac{1}{2}x^2\right)^{-2}$。所以

$$\int_0^{1/10} \frac{x}{(3 - \frac{1}{2}x^2)^2} \,\mathrm{d}x = \left[\frac{1}{3 - \frac{1}{2}x^2}\right]_0^{1/10} = \frac{200}{599} - \frac{1}{3} = \frac{600 - 599}{1797} = \frac{1}{1797}.$$

因此

$$\frac{1}{1800} \le \int_0^{1/10} \frac{x}{(2+\cos x)^2} \,\mathrm{d}x \le \frac{1}{1797}.$$

$\sin x$ 更精确的界

对于 $0 \le y \le x \le \frac{1}{2}\pi$，我们有 $\cos y \ge 1 - \frac{1}{2}y^2$，即 $0 \le 1 - \cos y \le \frac{1}{2}y^2$（左边的不等式成立是因为 $\cos y \le 1$）。由 0 积分到 x：

$$\int_0^x (1 - \cos y) \,\mathrm{d}y \le \int_0^x \tfrac{1}{2}y^2 \,\mathrm{d}y \quad \Longrightarrow \quad x - \sin x \le \tfrac{1}{6}x^3 \quad \Longrightarrow \quad \sin x \ge x - \tfrac{1}{6}x^3.$$

第二个积分

在 $[0, \frac{1}{10}]$ 上，我们有 $x - \frac{1}{6}x^3 \le \sin x \le x$，所以

$$1 - \tfrac{1}{6}x^3 \le 1 - x + \sin x \le 1.$$

> ? **想一想** 为甚么应该考虑这个组合?

分母中的 $-x$ 正是用来抵消 $\sin x$ 的首项，余下 1 加上一个 x^3 阶的量。下界 $1 - \frac{1}{6}x^3 \ge 1 - \frac{1}{6000}$ 是正数，所以和之前一样，我们先平方，再取倒数（次序倒转），然后乘以 $x^2 \ge 0$：

$$x^2 \le \frac{x^2}{(1 - x + \sin x)^2} \le \frac{x^2}{(1 - \frac{1}{6}x^3)^2}.$$

积分后，左边给出 $\int_0^{1/10} x^2\,\mathrm{d}x = \frac{1}{3}\cdot\frac{1}{1000} = \frac{1}{3000}$。至于右边，$\frac{\mathrm{d}}{\mathrm{d}x}\left(1 - \frac{1}{6}x^3\right)^{-1} = \frac{1}{2}x^2\left(1 - \frac{1}{6}x^3\right)^{-2}$，所以

$$\int_0^{1/10} \frac{x^2}{(1 - \frac{1}{6}x^3)^2}\,\mathrm{d}x = 2\left[\frac{1}{1 - \frac{1}{6}x^3}\right]_0^{1/10} = 2\left(\frac{6000}{5999} - 1\right) = \frac{2}{5999}.$$

因此

$$\frac{1}{3000} \le \int_0^{1/10} \frac{x^2}{(1 - x + \sin x)^2}\,\mathrm{d}x \le \frac{2}{5999}.$$

验算

- 用辛普森法则（Simpson's rule）得 $\int_0^{1/10} x(2 + \cos x)^{-2}\,\mathrm{d}x \approx 0.00055648$，介乎 $1/1800 \approx 0.00055556$ 与 $1/1797 \approx 0.00055648$ 之间（上界极为贴近：两者相差仅约 5×10^{-10}）。
- 同样地，$\int_0^{1/10} x^2(1 - x + \sin x)^{-2}\,\mathrm{d}x \approx 0.000333389$，介乎 $1/3000 \approx 0.000333333$ 与 $2/5999 \approx 0.000333389$ 之间。
- 在 $[0, \frac{1}{2}\pi]$ 内随机抽取的 10^5 个点上，三个逐点不等式全部成立，并在 $x = 0$ 时取等号。

常见错误

- 没有检查正负号就取倒数：只有当 a 与 b 同号时，$a \le b$ 才能推出 $1/a \ge 1/b$。
- 忘记了比较积分的结果需要 f ≥ 0 才能得出左边的不等式；这里是由 $\sin y \ge 0$ 和 $1 - \cos y \ge 0$ 提供这个条件。
- 用 $\cos x \le 1$ 这个界来求积分的上界：$\cos x \le 1$ 令分母小于 9，得出的是下界 $1/1800$，而不是上界。

STEP 之外

反复把 $0 \leq \cos y \leq 1$ 积分，可以得出一整串正负交错的泰勒（Taylor）界 $x - \frac{1}{6}x^3 \leq \sin x \leq x - \frac{1}{6}x^3 + \frac{1}{120}x^5 \leq \cdots$，这是证明 sin 和 cos 的泰勒级数（Taylor series）收敛（converge）于它们本身的其中一种方法。积分的单调性（monotonicity）以及被积函数的比较，都是实分析（real analysis）的基本工具，用来估计余项、误差和极限的界。Abbott 的 *Understanding Analysis*（一本实分析入门书）仔细地建立了积分理论和带余项的泰勒定理（Taylor's theorem）。

详细解答

STEP II 1991，第 7 题

纯数学 · 函数方程

> **STEP II 1991，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/91-s2-q7

关键思路

函数方程（functional equation）对所有 x 和 y 都成立，因此我们可以固定其中一个变量，对另一个变量求导。这样便把它化为一条微分方程（differential equation），把 g' 在两个不同点的值联系起来。然后巧妙地选取其中一个变量的值，就能确定 g' 在各处的值，最后积分便大功告成。全篇我们都跟题目一样，假设 g 和 f 可导（differentiable）。

解答

导数关系式

固定 $y>0$，把 $\mathrm{g}(x)+\mathrm{g}(y)=\mathrm{g}(z)$ 对 x 求导。由于 $\mathrm{g}(y)$ 此时是常数，由链式法则（chain rule）得

$$\mathrm{g}'(x)=\mathrm{g}'(z)\,\frac{\partial z}{\partial x}.$$

由商法则（quotient rule），

$$\frac{\partial z}{\partial x}=\frac{y(x+y+1)-xy}{(x+y+1)^2}=\frac{y^2+y}{(x+y+1)^2},$$

这就是第一个形式。至于第二个形式，留意 $z+1$ 可以漂亮地因式分解：

$$z+1=\frac{xy+x+y+1}{x+y+1}=\frac{(x+1)(y+1)}{x+y+1},\qquad\text{所以}\qquad z(z+1)=\frac{xy(x+1)(y+1)}{(x+y+1)^2}.$$

除以 $x(x+1)$ 后恰好得到 $\dfrac{y(y+1)}{(x+y+1)^2}$，所以

$$\mathrm{g}'(x)=\frac{y^2+y}{(x+y+1)^2}\,\mathrm{g}'(z)=\frac{z(z+1)}{x(x+1)}\,\mathrm{g}'(z),$$

正如所求。值得把它写成以下的对称形式：

$$x(x+1)\,\mathrm{g}'(x) = z(z+1)\,\mathrm{g}'(z). \qquad (\dagger)$$

推导 $2\mathrm{g}'(1) = (u^2+u)\mathrm{g}'(u)$

> **想一想** 我们想得到一个对 $(0,1)$ 内每一个 u 都成立的结论。应该固定哪一个变量？固定在甚么值？

在 $(\dagger)$ 中代入 $x=1$；左边变成 $2\mathrm{g}'(1)$，而 $z = \dfrac{y}{y+2}$。当 y 取遍 $(0,\infty)$ 时，$\dfrac{y}{y+2} = 1 - \dfrac{2}{y+2}$ 由 0 连续地递增并趋向 1，所以它取遍所有满足 $0<u<1$ 的值 u（具体来说，$y = \frac{2u}{1-u}$）。因此

$$2\mathrm{g}'(1) = (u^2+u)\,\mathrm{g}'(u) \qquad \text{对所有} 0<u<1.$$

积分

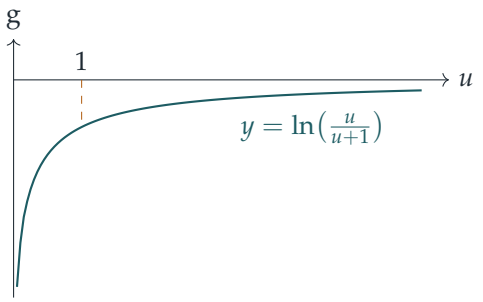
设 $A = 2\mathrm{g}'(1)$，这是一个常数。那么对 $0<u<1$，利用部分分式（partial fractions），

$$\mathrm{g}'(u) = \frac{A}{u(u+1)} = A\left(\frac{1}{u} - \frac{1}{u+1}\right),$$

再积分（各项皆为正数，所以不需要绝对值符号），

$$\mathrm{g}(u) = A\big(\ln u - \ln(u+1)\big) + B = A\ln\left(\frac{u}{u+1}\right) + B.$$

其实这条公式对每一个 $u>0$ 都成立，而不只是 $u<1$。对固定的 x，当 y 由 0 增至 ∞ 时，$z = \dfrac{xy}{x+y+1}$ 由 0 递增并趋向 x，所以我们可以选取 y 使得 $z<1$。这样，由 $(\dagger)$ 和上面的结果可得 $x(x+1)\mathrm{g}'(x) = z(z+1)\mathrm{g}'(z) = A$。因此同样的积分在整个 $(0,\infty)$ 上都适用，而且只有一个常数 B，因为 $(0,\infty)$ 是一个区间。



（图中所示的曲线是 $A=1$、$B=0$ 的情况：它取负值、递增，并且当 $u\to\infty$ 时趋向 0。）

验证 $(*)$ 并求 B

代入 $\mathrm{g}(u) = A\ln\frac{u}{u+1} + B$。$(*)$ 的左边是

$$\mathrm{g}(x) + \mathrm{g}(y) = A\ln\frac{xy}{(x+1)(y+1)} + 2B.$$

至于右边，利用先前求得的 $z+1$ 的因式分解，可得

$$\frac{z}{z+1} = \frac{xy}{x+y+1}\cdot\frac{x+y+1}{(x+1)(y+1)} = \frac{xy}{(x+1)(y+1)},$$

所以

$$\mathrm{g}(z) = A\ln\frac{xy}{(x+1)(y+1)} + B.$$

两边对所有 $x, y > 0$ 都相等，当且仅当 $2B = B$，即 $B = 0$。所以 $\mathrm{g}(u) = A\ln\dfrac{u}{u+1}$ 满足 $(*)$。

能否确定 A?

不能。上面的验证对 A 的每一个值都成立，包括 $A = 0$（零函数）。定义 $A = 2\mathrm{g}'(1)$ 也帮不上忙：对我们的解而言，$\mathrm{g}'(1) = A\left(1-\frac{1}{2}\right) = \frac{A}{2}$，所以 $A = 2\mathrm{g}'(1)$ 自动成立，并没有提供任何新信息。方程 $(*)$ 对 g 而言是线性的，所以一个解的任何倍数仍然是解，而 A 正是这个比例因子。

函数 f

同样的方法也适用。固定 $y > 0$，把 $\mathrm{f}(x) + \mathrm{f}(y) = \mathrm{f}(xy)$ 对 x 求导：

$$\mathrm{f}'(x) = y\,\mathrm{f}'(xy).$$

想一想 在这里，与代入 $x = 1$ 相对应的做法是甚么?

代入 $x = 1$，得 $\mathrm{f}'(1) = y\,\mathrm{f}'(y)$ 对每一个 $y > 0$ 都成立，所以设 $C = \mathrm{f}'(1)$，便有

$$\mathrm{f}'(y) = \frac{C}{y} \qquad\Longrightarrow\qquad \mathrm{f}(y) = C\ln y + D \quad (y > 0).$$

代回原式，$\mathrm{f}(x) + \mathrm{f}(y) = C\ln(xy) + 2D$，而 $\mathrm{f}(xy) = C\ln(xy) + D$，所以 $D = 0$。（等价地，在方程中取 $x = y = 1$ 得 $2\mathrm{f}(1) = \mathrm{f}(1)$，所以 $\mathrm{f}(1) = 0$。）因此 $\mathrm{f}(x) = C\ln x$，正如所求，而 C 同样是任意的。

验算

▶ 取 $x = 1$、$y = 2$：则 $z = \frac{1}{2}$，而两个导数因子分别是 $\frac{6}{16}$ 和 $\frac{\frac{1}{2}\cdot\frac{3}{2}}{2} = \frac{3}{8}$，两者相符。

- 取 $A=1$：$\mathrm{g}(1)+\mathrm{g}(2)=\ln\frac{1}{2}+\ln\frac{2}{3}=\ln\frac{1}{3}$，而 $\mathrm{g}(\frac{1}{2})=\ln\frac{1/2}{3/2}=\ln\frac{1}{3}$。✓
- 我们用电脑以有限差分（finite differences）在 1000 个随机点上核实了导数恒等式，并对 A 的若干个值核实了 $\mathrm{g}(x)+\mathrm{g}(y)-\mathrm{g}(z)$ 恰好等于 B（所以只有当 $B=0$ 时才等于零）。

常见错误

- 对 x 求导时忘记 $\mathrm{g}(y)$ 是常数，或者漏掉链式法则的因子 $\partial z/\partial x$。
- 代入了 $x=1$，却没有检查 z 是否真的取遍整个 $(0,1)$；正是这一步，把单独一条等式变成对每一个 u 都成立的结论。
- “因为积分常数无关重要”而假定 B 必定是 0：其实是方程 $(*)$ 迫使 $B=0$，而这一点需要证明。

STEP 之外

第二条方程是柯西（Cauchy）对数型函数方程：代入 $x=\mathrm{e}^s$ 便把它化为 $h(s+t)=h(s)+h(t)$。这里我们假设了可导性；在大学第一门分析课程中，你会看到只要在单独一点连续（continuity），就已经足以迫使 $h(s)=cs$，而如果完全没有任何正则性条件，就会出现怪异的非线性解（利用哈默尔基（Hamel basis）构造而成）。第一条方程其实是同一回事的变相：$w=\ln\frac{u}{u+1}$ 把 $(*)$ 变换成加性方程。Abbott 的 *Understanding Analysis*（一本实分析入门书）是阅读这类连续性论证的好去处。

详细解答

STEP II 1991，第 8 题

纯数学 · 非线性一阶常微分方程

> **STEP II 1991，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/91-s2-q8

关键思路

两条微分方程（differential equation）都是关于 $\frac{\mathrm{d}y}{\mathrm{d}x}$ 的二次方程（quadratic），所以我们可以先用代数方法解出导数（derivative），然后才积分。每条二次方程有两个根，于是得到两条候选的一阶方程，而 $x = 0$ 处的条件决定了哪一个分支（branch）可取。双曲恒等式（hyperbolic identities），尤其是 $\cosh^2 x - \sinh^2 x = 1$ 和半角公式（half-angle formulae），可令这些根保持简洁。

解答

二次方程

由二次公式（quadratic formula），

$$u = \frac{-2\sinh x \pm \sqrt{4\sinh^2 x + 4}}{2} = -\sinh x \pm \sqrt{\sinh^2 x + 1} = -\sinh x \pm \cosh x,$$

当中用到 $\cosh^2 x - \sinh^2 x = 1$ 和 $\cosh x > 0$。由于 $\cosh x - \sinh x = \mathrm{e}^{-x}$ 而 $\cosh x + \sinh x = \mathrm{e}^{x}$，两个根是

$$u = \mathrm{e}^{-x} \qquad \text{或} \qquad u = -\mathrm{e}^{x}.$$

（验算：两根之积是 -1，两根之和是 $\mathrm{e}^{-x} - \mathrm{e}^{x} = -2\sinh x$，与应有的结果相符。）

第一条微分方程

取 $u = \frac{\mathrm{d}y}{\mathrm{d}x}$，这正是上面的二次方程，所以对每一个 x，要么 $y' = \mathrm{e}^{-x}$，要么 $y' = -\mathrm{e}^{x}$。

想一想 y' 会不会在某些 x 值上取其中一个根，而在其他地方取另一个根?

不会。第一个根恒为正，第二个根恒为负，所以两者永不相遇。导数具有介值性质（intermediate value property）（达布定理（Darboux's theorem）；如果你假设 y' 连续，则用普通的介值定理（intermediate value theorem）即可），所以 y' 不可能由正值跳到负值而不经过 0，但两个根都不会取这个值。因此 y' 自始至终都停留在同一个根上。

在 $x=0$ 处, 两个根是 1 和 -1, 而我们需要 $y'>0$, 所以 $\frac{\mathrm{d}y}{\mathrm{d}x}=\mathrm{e}^{-x}$。积分得 $y=-\mathrm{e}^{-x}+c$, 而由 $x=0$ 时 $y=0$ 可得 $c=1$:

$$\boxed{y=1-\mathrm{e}^{-x}}$$

第二条微分方程

设 $p=\frac{\mathrm{d}y}{\mathrm{d}x}$，则 $p^2\sinh x+2p-\sinh x=0$。

当 $x=0$ 时，$\sinh 0=0$，方程简化为 $2p=0$，所以必定有 $y'(0)=0$。

当 $x\neq 0$ 时，$\sinh x\neq 0$，由二次公式得

$$p=\frac{-2\pm\sqrt{4+4\sinh^2 x}}{2\sinh x}=\frac{-1\pm\cosh x}{\sinh x}.$$

利用 $\cosh x=1+2\sinh^2\frac{x}{2}=2\cosh^2\frac{x}{2}-1$ 和 $\sinh x=2\sinh\frac{x}{2}\cosh\frac{x}{2}$，两个根可化简为

$$\frac{\cosh x-1}{\sinh x}=\frac{2\sinh^2\frac{x}{2}}{2\sinh\frac{x}{2}\cosh\frac{x}{2}}=\tanh\frac{x}{2},$$

$$\frac{-1-\cosh x}{\sinh x}=-\frac{2\cosh^2\frac{x}{2}}{2\sinh\frac{x}{2}\cosh\frac{x}{2}}=-\coth\frac{x}{2}.$$

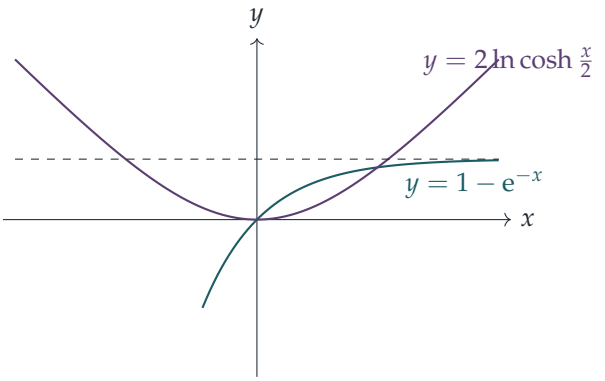
选择分支。第二个根 $-\coth\frac{x}{2}$ 当 $x\to 0$ 时无界，把它积分得 $y=-2\ln\left|\sinh\frac{x}{2}\right|+c$，而无论从哪一侧取 $x\to 0$，这个函数都趋向 $+\infty$。所以在原点的任何一侧，这个分支都不能给出在 $x=0$ 处连续且满足 $y(0)=0$ 的解。因此在 $x=0$ 附近（每一侧）我们都必须取第一个根，而由于 $\tanh 0=0$，这与 $y'(0)=0$ 吻合。两个根永不重合（一个与 x 同号，另一个异号，而且当 $x\neq 0$ 时两者都不为零），所以根据与之前相同的介值论证，对所有 x，解都停留在这个分支上:

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\tanh\frac{x}{2}\quad\text{对所有}x.$$

由于 $\frac{\mathrm{d}}{\mathrm{d}x}\cosh\frac{x}{2}=\frac{1}{2}\sinh\frac{x}{2}$，我们有 $\tanh\frac{x}{2}=2\cdot\frac{\frac{1}{2}\sinh\frac{x}{2}}{\cosh\frac{x}{2}}$，这属于 $2f'/f$ 的形式，所以积分得 $y=2\ln\cosh\frac{x}{2}+c$（因为 $\cosh>0$，无须加上绝对值符号）。由条件 $y(0)=0$ 得 $c=0$，所以

$$\boxed{y=2\ln\cosh\tfrac{x}{2}=\ln\left(\frac{1+\cosh x}{2}\right)}$$

其中第二个形式由 $2\cosh^2\frac{x}{2} = 1 + \cosh x$ 得出。



验算

- 对于 $y = 1 - \mathrm{e}^{-x}$：$y'^2 + 2y'\sinh x = \mathrm{e}^{-2x} + \mathrm{e}^{-x}(\mathrm{e}^x - \mathrm{e}^{-x}) = 1$，符合要求，而且 $y'(0) = 1 > 0$。
- 对于 $y = 2\ln\cosh\frac{x}{2}$，设 $t = \tanh\frac{x}{2}$，则 $\sinh x = \frac{2t}{1-t^2}$；于是 $\sinh x\,(t^2-1) + 2t = -2t + 2t = 0$。这个解是偶函数，这是理所当然的，因为方程在 $x \mapsto -x$、$y \mapsto y$ 之下保持不变。
- 数值上，在 $x = 1$ 处两个解分别是 $1 - \mathrm{e}^{-1} \approx 0.6321$ 和 $2\ln\cosh\frac{1}{2} \approx 0.2402$；在许多随机选取的点代入有限差分导数，两条方程都成立，误差在 10^{-6} 以内。

常见错误

- 把第二条方程除以 $\sinh x$，却没有留意它在 $x = 0$ 处为零，而那里正是初始条件（initial condition）所在之处；原方程迫使 $y'(0) = 0$。
- 在第二部分保留两个分支：$-\coth\frac{x}{2}$ 这个分支不可能满足 $y(0) = 0$，因为它的积分在原点处趋向无穷。

STEP 之外

对 y' 并非线性的方程，例如 $F(x, y, y') = 0$，可以有多个解通过同一点，或者出现没有任何单一分支能够描述的“奇异”解（singular solution），所以存在性与唯一性定理（existence and uniqueness theorems，即 Picard–Lindelöf 定理）都是针对已经解成 $y' = f(x, y)$ 形式的方程而叙述的。在这里两个根互相分隔，所以每个分支都表现良好；

当判别式（discriminant）可以为零时，各分支可以黏合起来，唯一性便不再成立。Boyce 与 DiPrima 的 *Elementary Differential Equations*（基础微分方程教科书）讨论了存在性、唯一性以及这些非线性一阶方程。

详细解答

STEP II 1991，第 9 题

纯数学 · 矩阵群

> **STEP II 1991，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/91-s2-q9

关键思路

一切都建基于一个事实：模 5 的整数构成一个域（field）。每个非零剩余（residue）都有乘法逆元（multiplicative inverse），而非零剩余之积永不为零。正因如此，对角线上的元素才会规规矩矩，封闭性和逆元几乎是不费工夫就得到的。至于最后一部分，可解 $M^2 = I$，把阶（order）为 1 或 2 的元素明确地找出来，然后检验其中两个相乘之后是否仍留在该集合之内。

解答

以下所有运算均以模 5 进行，并在方便时把 4 写成 -1。我们把一般的元素写成

$$M(a,b,c) = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}, \qquad a,c \in \{1,2,3,4\},\ b \in \{0,1,2,3,4\}.$$

G 是一个群

封闭性（closure）。把两个元素相乘，得

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}\begin{pmatrix} d & e \\ 0 & f \end{pmatrix} = \begin{pmatrix} ad & ae+bf \\ 0 & cf \end{pmatrix}.$$

这是一个上三角矩阵（upper triangular matrix），各元素均为模 5 的数。它的对角线元素 ad 和 cf 都不是零：由于 5 是质数，若 $5 \mid ad$，则 $5 \mid a$ 或 $5 \mid d$，但两者皆不成立。所以这个积属于 G。

结合律（associativity）是已知的，因为矩阵乘法符合结合律。

单位元（identity）。$I = M(1,0,1)$ 属于 G，而且对每个 M 都有 $IM = MI = M$。

逆元（inverses）。每个模 5 的非零剩余都有逆元（$1^{-1} = 1,\ 2^{-1} = 3,\ 3^{-1} = 2,\ 4^{-1} = 4$）。所以，给定 $M(a,b,c) \in G$，我们可以构造

$$N = \begin{pmatrix} a^{-1} & -a^{-1}bc^{-1} \\ 0 & c^{-1} \end{pmatrix},$$

由于 $a^{-1}, c^{-1} \neq 0$，它属于 G。利用上面的乘积公式，

$$MN = \begin{pmatrix} aa^{-1} & -aa^{-1}bc^{-1} + bc^{-1} \\ 0 & cc^{-1} \end{pmatrix} = I, \qquad NM = \begin{pmatrix} a^{-1}a & a^{-1}b - a^{-1}bc^{-1}c \\ 0 & c^{-1}c \end{pmatrix} = I.$$

所以每个元素在 G 中都有逆元，因此 G 是一个群（group）。

G 的阶

a 有 4 个选择，b 有 5 个，c 有 4 个，而不同的选择给出不同的矩阵。所以

$$|G| = 4 \times 5 \times 4 = 80.$$

G 是否可交换?

不是。取 $A = M(1,1,1)$ 和 $B = M(2,0,1)$。则

$$AB = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}, \qquad BA = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix},$$

所以 $AB \neq BA$，G 不是可交换（commutative）的。

阶为 1 或 2 的元素

元素 M 的阶为 1 或 2，当且仅当 $M^2 = I$。由乘积公式，

$$M(a,b,c)^2 = \begin{pmatrix} a^2 & b(a+c) \\ 0 & c^2 \end{pmatrix},$$

所以我们需要 $a^2 = 1$、$c^2 = 1$ 以及 $b(a+c) = 0$。

$1,2,3,4$ 在模 5 下的平方分别是 $1,4,4,1$，所以 $a^2 = 1$ 必然得出 $a = \pm 1$，同理 $c = \pm 1$。

- 若 $a = c$，则 $a + c = \pm 2 \neq 0$。由于 $\mathbb{Z}_5$ 没有零因子（zero divisor），$b(a+c) = 0$ 必然得出 $b = 0$。这给出 I 和 $-I$。
- 若 $c = -a$，则 $a + c = 0$，所以 b 可取任何值。这给出 10 个矩阵 $M(1,b,-1)$ 和 $M(-1,b,1)$，其中 $b = 0,1,\ldots,4$。

所以，由阶为 1 或 2 的元素组成的集合 H 恰好有 12 个成员。

想一想 由自逆元素组成的集合，是否必定封闭?

不是。这个集合包含 I，而且每个元素都是自己的逆元，所以单位元和逆元这两条公理自动成立。唯一可能不成立的是封闭性，而在非交换群中它通常都不成立：若 $X^2 = Y^2 = I$，则 $(XY)^2 = XYXY$，除非 X 与 Y 可交换，否则这未必能化简为 I。

在这里它确实不成立。取 $X = M(1,1,-1)$ 和 $Y = M(1,0,-1)$，两者都属于 H。则

$$XY = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

这个矩阵有 $a = c = 1$ 但 $b = -1 \neq 0$，所以按上面的分类，它不属于 H。（事实上 $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^k = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix}$，所以它的阶是 5。）因此 H 并不封闭，它不是 G 的子群（subgroup）。

如果你认识拉格朗日定理（Lagrange's theorem），还有一个更快捷的论证：子群的阶必须整除群的阶，而 $12 \nmid 80$。

验算

用电脑把全部 80 个矩阵简单枚举一遍，可确认封闭性、逆元公式，以及上述那对矩阵的 $AB \neq BA$。把每个元素平方，找到的正好是预测中 $M^2 = I$ 的 12 个解，而 $XY = M(1,4,1)$ 的阶是 5。用手算可验证 $M(-1,3,1)^2 = I$：非对角线元素是 $3(-1+1) = 0$。

常见错误

- 忘记证明逆元属于 G。它的对角线元素 a^{-1}、c^{-1} 必须不是零。
- 以为 $M^2 = I$ 必然令 M 是对角矩阵，因而漏掉 $b \neq 0$ 的元素 $M(1,b,-1)$ 和 $M(-1,b,1)$。
- 只为 H 检验单位元和逆元这两条公理。不成立的正是封闭性这条公理。

STEP 之外

G 是有限域（finite field）$\mathbb{F}_5$ 上所有可逆 2×2 上三角矩阵组成的群，是 $\mathrm{GL}_2(\mathbb{F}_5)$ 的一个博雷尔子群（*Borel subgroup*）。满足 $a = c = 1$ 的矩阵构成一个阶为 5 的正规子群（normal subgroup），而 G 是由它与对角矩阵以半直积（semidirect product）的方式构成的。有限域上的矩阵群，是群论入门课程中丰富的例子来源。想了解更多，可参阅 Allenby, *Rings, Fields and Groups*（环、域与群）。

详细解答

STEP II 1991，第 10 题

纯数学 · 三维空间中的直线与平面

STEP II 1991，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/91-s2-q10

关键思路

虽然题库把这题归入另一类别，但它其实是一道关于三维空间中直线与平面的题目。影子的尖端，就是刚好掠过棒顶的那道太阳光线与地面相交之处。因此，我们把这道光线写成一条通过太阳与棒顶的直线，把地面写成一个通过棒底并与棒垂直的平面，然后求两者的交点。只要把所有位置都用沿棒方向的单位向量（unit vector）表示，一切便变得容易。

解答

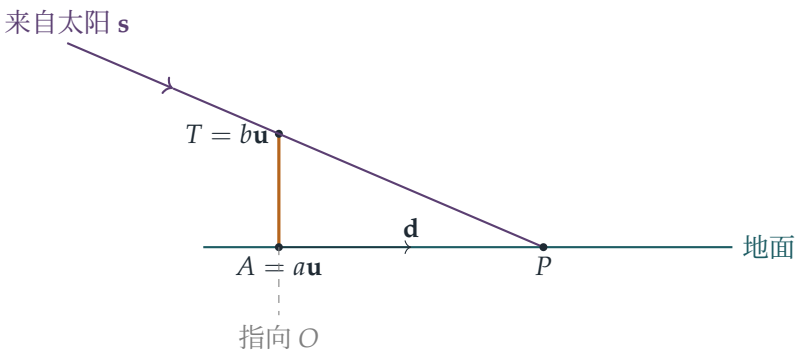
设 O 为地球的中心，并记

$$\mathbf{u} = \begin{pmatrix} 0 \\ \cos\phi \\ \sin\phi \end{pmatrix}, \qquad \mathbf{s} = R\begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}.$$

那么 $\mathbf{u}$ 是沿棒方向的单位向量，棒底为 $A = a\mathbf{u}$，棒顶为 $T = b\mathbf{u}$，而太阳位于 $\mathbf{s}$。我们唯一需要的点积（dot product）是

$$\mathbf{s}\cdot\mathbf{u} = R\sin\theta\cos\phi.$$

为简便起见，记 $c = R\cos\phi\sin\theta$。



通过棒顶的光线

到达影子尖端 P 的光线，刚刚经过棒顶。所以 P 位于通过 $\mathbf{s}$ 与 $b\mathbf{u}$ 的直线上：

$$\mathbf{p} = \mathbf{s} + t\,(b\mathbf{u} - \mathbf{s}) = (1-t)\,\mathbf{s} + tb\,\mathbf{u}$$

其中 t 为某个实数。

地面

地面是水平的，即与棒垂直，并且通过棒底 $a\mathbf{u}$。由于 $\mathbf{u}\cdot\mathbf{u} = 1$，地面就是平面 $\mathbf{r}\cdot\mathbf{u} = a$。

求交点

把直线代入平面的方程，并利用 $\mathbf{u}\cdot\mathbf{u} = 1$ 及 $\mathbf{s}\cdot\mathbf{u} = c$，得

$$(1-t)c + tb = a \qquad \Longrightarrow \qquad t = \frac{c-a}{c-b}.$$

我们可以除以 $c-b$，因为太阳正照着这根棒。在棒顶处，太阳位于地平线之上，当且仅当它位于该处切平面（tangent plane）的另一侧，即 $\mathbf{s}\cdot\mathbf{u} > b$，所以 $c > b$。

> ? **想一想** 这个 P 真的在棒后方的地面上，而不是在太阳背后的某处吗?

由于 $c > b > a$，我们有 $c-a > c-b > 0$，所以 $t > 1$。$t = 0$ 的点是太阳，$t = 1$ 的点是棒顶，因此 $t > 1$ 表示已越过棒顶、背向太阳而行，正如所料。

由 t 的值可得

$$1 - t = \frac{(c-b)-(c-a)}{c-b} = \frac{a-b}{c-b} = -\frac{h}{c-b},$$
$$tb - a = \frac{b(c-a)-a(c-b)}{c-b} = \frac{c(b-a)}{c-b} = \frac{ch}{c-b}.$$

所以由棒底到影子尖端的位移（displacement）为

$$\begin{aligned}\mathbf{d} = \mathbf{p} - a\mathbf{u} &= (1-t)\,\mathbf{s} + (tb - a)\,\mathbf{u} \\ &= \frac{h}{c-b}\left(c\,\mathbf{u} - \mathbf{s}\right) \\ &= \frac{Rh}{c-b}\left[\cos\phi\sin\theta\begin{pmatrix}0\\ \cos\phi\\ \sin\phi\end{pmatrix} - \begin{pmatrix}\cos\theta\\ \sin\theta\\ 0\end{pmatrix}\right] \\ &= \frac{Rh}{c-b}\begin{pmatrix}-\cos\theta\\ -(1-\cos^2\phi)\sin\theta\\ \cos\phi\sin\phi\sin\theta\end{pmatrix}.\end{aligned}$$

由于 $1-\cos^2\phi = \sin^2\phi$ 及 $c - b = R\cos\phi\sin\theta - b$，上式即

$$\mathbf{d} = Rh\,(R\cos\phi\sin\theta - b)^{-1}\begin{pmatrix}-\cos\theta\\ -\sin^2\phi\sin\theta\\ \cos\phi\sin\phi\sin\theta\end{pmatrix},$$

这正是所要求的结果。

中间那一行 $\mathbf{d} = \frac{h}{c-b}(c\,\mathbf{u} - \mathbf{s})$ 值得留意。向量 $c\,\mathbf{u}$ 是 $\mathbf{s}$ 沿棒方向的分量（component），所以 $c\,\mathbf{u} - \mathbf{s}$ 就是太阳位置向量（position vector）的水平部分的负值。影子正正指向与太阳的水平方向相反的一方，与日常经验吻合。

验算

- $\mathbf{d}\cdot\mathbf{u} = \frac{Rh}{c-b}\left(-\sin^2\phi\sin\theta\cos\phi + \cos\phi\sin^2\phi\sin\theta\right) = 0$，所以影子确实贴着地面。
- 取 $\phi = 0$，则棒沿 y 轴方向。这时 $\mathbf{d} = \frac{Rh}{R\sin\theta - b}(-\cos\theta, 0, 0)$。当 $\theta = \frac{\pi}{2}$ 时，太阳在正上方，影子消失。当 R 极大时，影长趋向 $h\cot\theta$。这正是大家熟悉的结果：把地面视为平面、太阳仰角为 θ 时的影长。
- 数值方面，我们随机选取了数千组 R, a, h, θ, ϕ 的值（太阳均在地平线之上），直接以光线与地平面相交的方法求出影子尖端的位置。结果每次都与公式吻合，最大相对误差约为 10^{-13}。

常见错误

- 把地面当作平面 $y = a\cos\phi$，或其他坐标平面。“水平”的意思是与 $\mathbf{u}$ 垂直，即 $\mathbf{r}\cdot\mathbf{u} = a$。
- 一开始就把所有东西按分量展开。把 $\mathbf{u}$ 和 $\mathbf{s}$ 保留为向量，直到最后一行才展开，可以省却大量代数运算。

STEP 之外

在这里，太阳光是通过同一点（太阳）的一族直线，所以影子是棒在地平面上的中心投影（central projection）。当太阳非常遥远时，光线几乎平行，这个投影便变成平行投影（parallel projection）。这正是线性代数（linear algebra）所研究的那类映射，而向量 $c\,\mathbf{u} - \mathbf{s}$ 就是 $\mathbf{s}$ 在棒上的投影减去 $\mathbf{s}$ 本身。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）详细讨论了在直线和平面上的投影。

详细解答

STEP II 1991，第 11 题

力学 · 抛体与发射点的距离

STEP II 1991，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/91-s2-q11

关键思路

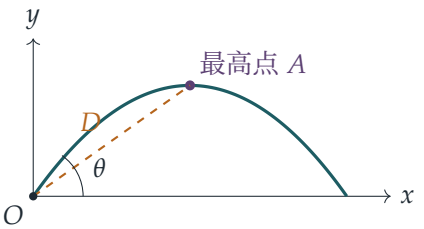
处理砲弹与大砲之间距离的*平方*：这样可以避开平方根，并把两部分都化为多项式（polynomial）问题。在第一部分，最高点距离的平方是 $\sin^2\theta$ 的二次式，所以用配方法（completing the square）（或求导）便可找出最佳角度。在第二部分，砲弹逐渐接近，当且仅当距离的平方正在递减，因此我们要考察它对时间的导数（derivative）在整个飞行过程中的正负。

解答

模型

我们把砲弹视为只受重力作用而自由运动的质点（particle），不计空气阻力，从原点 O（砲手所站之处）以速率 u、仰角（angle of elevation）θ 发射，其中 $0 < \theta < \frac{\pi}{2}$。取 x 为水平方向，y 为铅垂向上。在时间 t，

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



仰角

当 $\dot{y} = u\sin\theta - gt = 0$，即 $t = u\sin\theta/g$ 时，砲弹到达最高点（apex）。此时

$$x_A = \frac{u^2\sin\theta\cos\theta}{g}, \qquad y_A = \frac{u^2\sin^2\theta}{2g}.$$

以 D 表示距离 OA，并设 $s = \sin^2\theta$，于是 $\cos^2\theta = 1 - s$。那么

$$\begin{aligned} D^2 = x_A^2 + y_A^2 &= \frac{u^4}{g^2}\left(\sin^2\theta\cos^2\theta + \tfrac{1}{4}\sin^4\theta\right) \\ &= \frac{u^4}{g^2}\left(s(1-s) + \tfrac{1}{4}s^2\right) = \frac{u^4}{g^2}\left(s - \tfrac{3}{4}s^2\right). \end{aligned}$$

由于损害随距离增加而减少，砲手希望 D 愈大愈好，这等同于令 D^2 愈大愈好。配方得

$$s - \tfrac{3}{4}s^2 = \tfrac{1}{3} - \tfrac{3}{4}\left(s - \tfrac{2}{3}\right)^2,$$

它恰好在 $s = \frac{2}{3}$ 时取得最大值 $\frac{1}{3}$。这个值是可以取到的，因为 $s = \sin^2\theta$ 可取 $(0,1)$ 内的每一个值。所以 $\sin^2\theta = \frac{2}{3}$，$\cos^2\theta = \frac{1}{3}$，从而

$$\boxed{\tan\theta = \sqrt{2}, \qquad \theta = \arctan\sqrt{2} \approx 54.7^\circ}$$

此时最高点与大砲的距离为 $D = \dfrac{u^2}{\sqrt{3}\,g}$。

砲弹在下降期间会否更接近砲手？

设 $r^2 = x^2 + y^2$ 为时间 t 时砲弹与大砲距离的平方。砲弹逐渐接近，当且仅当 r^2 正在递减，所以我们需要知道

$$\tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(r^2) = x\dot{x} + y\dot{y}.$$

的正负。

> ? **想一想** 为甚么不直接把最高点的距离与之后几个点的距离比较一下？

只看几个点，并不能排除距离在它们之间曾经下跌的可能。导数却能一次过告诉我们飞行中每一刻的情况。由 $\dot{x} = u\cos\theta$ 及 $\dot{y} = u\sin\theta - gt$，

$$\begin{aligned} x\dot{x} + y\dot{y} &= u^2 t\cos^2\theta + \left(ut\sin\theta - \tfrac{1}{2}gt^2\right)(u\sin\theta - gt) \\ &= u^2 t - \tfrac{3}{2}ugt^2\sin\theta + \tfrac{1}{2}g^2t^3 \\ &= \tfrac{1}{2}t\left[(gt)^2 - 3u\sin\theta\,(gt) + 2u^2\right], \end{aligned}$$

其中用到 $\cos^2\theta + \sin^2\theta = 1$。当 $t > 0$ 时，其正负与方括号内的式子相同，而该式是 gt 的二次式。配方得

$$(gt)^2 - 3u\sin\theta\,(gt) + 2u^2 = \left(gt - \tfrac{3}{2}u\sin\theta\right)^2 + u^2\left(2 - \tfrac{9}{4}\sin^2\theta\right).$$

在所选的角度下 $\sin^2\theta = \frac{2}{3}$，所以常数项为 $u^2\left(2 - \frac{3}{2}\right) = \frac{1}{2}u^2 > 0$。因此方括号内的式子对每个 t 都严格为正，故 r^2 在整个飞行过程中严格递增（strictly increasing）。

所以答案是**不会**：砲弹在整个下降过程中一直远离砲手，而下降路径上最接近的一点就是最高点本身。这说明砲手只考虑砲弹在最大高度爆炸的捷径是合理的：对这个角度而言，最高点确实是下降路径上最接近的一点，距离为 $u^2/(\sqrt{3}\,g)$。

验算

- 当 $\theta = \frac{\pi}{4}$ 时，最高点位于 $(u^2/2g,\, u^2/4g)$，所以 $D^2 = \frac{5}{16}u^4/g^2$，比我们的最大值 $\frac{1}{3}u^4/g^2$ 略小，正如所料。
- 一般情况的计算显示，只有当 $\frac{9}{4}\sin^2\theta \geq 2$，即 $\sin\theta \geq \frac{2\sqrt{2}}{3}$ 时，砲弹才开始会有接近的情况，对应的角度约为 $70.5°$，远高于 $54.7°$。（极限 $\theta \to \frac{\pi}{2}$ 的情况合乎常理：垂直向上发射的砲弹会笔直地朝大砲落回来。）
- 数值上，取 $u = g = 1$，对 θ 搜索，可得最高点距离的最大值 $0.57735 \approx 1/\sqrt{3}$，出现在 $\theta \approx 54.736°$；在该角度下沿轨迹（trajectory）逐步计算，可见与 O 的距离从不减少；而在 $70.5° \pm 0.6°$ 之间，距离由单调（monotonic）变为不单调。

常见错误

- 把最高点的水平距离或高度最大化，而不是把它与大砲的直线距离最大化。
- 以为砲弹在水平方向不断远离，所以过了最高点后与大砲的距离显然会增加。其实砲弹同时也在下坠，而当发射角度够陡时，它确实会回头接近大砲。

STEP 之外

对 r^2 而非 r 求导，再看 $\mathbf{r}\cdot\dot{\mathbf{r}}$ 的正负，是判断动点正在接近还是远离的标准方法，在向量力学（vector mechanics）以至轨道（orbits）的研究中随处可见。临界角 $\arcsin(2\sqrt{2}/3)$ 是一个经典结果：仰角低于此值时，抛体（projectile）与发射点的距离永远递增。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）阐述了以向量处理运动的方法，令这类论证显得自然而然。

详细解答

STEP II 1991，第 12 题

力学 · 受迫弹性振动

STEP II 1991，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/91-s2-q12

本题在 *Beyond STEP* 第 11.1 节共振中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/resonance.html

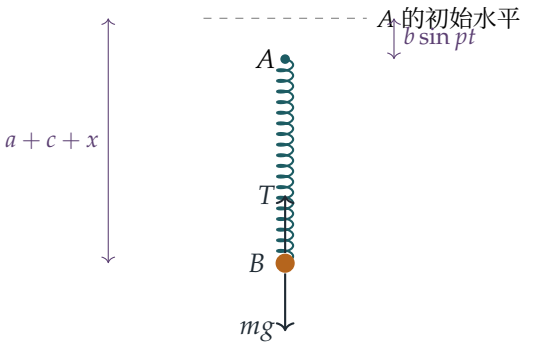
关键思路

张力只取决于伸长量（extension），即质点的位置减去移动端的位置，再减去自然长度（natural length）。把它用 x 和 $b\sin pt$ 表示出来，再利用初始平衡状态把弹性模量（modulus）换成含 c 的式子，运动方程（equation of motion）便自然得出。接下来要处理的是一个受迫简谐振子（forced simple harmonic oscillator）。求解时把特解（particular solution）加到余函数（complementary function）上，并谨记“拉紧”的意思是“伸长量永不为负”。

解答

运动方程

把质点视为质量为 m 的一点，系于一条遵从胡克定律（Hooke's law）、弹性模量为 λ 的轻弹性绳上。所有位置均由 A 的初始位置起向下量度。



在初始平衡状态下，伸长量为 c，而张力与重量平衡：

$$\frac{\lambda c}{a} = mg, \qquad \text{所以} \qquad \frac{\lambda}{ma} = \frac{g}{c} = n^2.$$

在时间 t，质点位于深度 $a + c + x$，而端点 A 位于深度 $b\sin pt$。只要绳子拉紧，其长度为 $a + c + x - b\sin pt$，所以伸长量为

$$e = c + x - b\sin pt,$$

而张力为 $T = \lambda e/a$。对质点沿向下方向分解力，

$$\begin{aligned} m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} &= mg - \frac{\lambda}{a}(c + x - b\sin pt) \\ &= mg - \frac{\lambda c}{a} - \frac{\lambda}{a}(x - b\sin pt) \\ &= -\frac{\lambda}{a}(x - b\sin pt), \end{aligned}$$

其中用了 $\lambda c/a = mg$。两边除以 m，并利用 $\lambda/(ma) = n^2$，得

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -n^2(x - b\sin pt),$$

正是所求。这里需要 $e \geq 0$：假如绳子松弛，张力便会是零，而不是 $\lambda e/a$。

求解 $x(t)$

把方程改写成 $\ddot{x} + n^2x = n^2b\sin pt$。余函数是 $C\cos nt + D\sin nt$。至于特积分（particular integral），试用 $x = K\sin pt$：

$$-p^2K + n^2K = n^2b \quad \Longrightarrow \quad K = \frac{n^2b}{n^2 - p^2},$$

这是成立的，因为 $p < n$ 保证 $n^2 - p^2 \neq 0$。所以

$$x = \frac{n^2b}{n^2 - p^2}\sin pt + C\cos nt + D\sin nt.$$

在 $t = 0$ 时，质点位于平衡位置且处于静止，所以 $x(0) = 0$ 及 $\dot{x}(0) = 0$。由前者得 $C = 0$。由后者得

$$\frac{n^2 bp}{n^2 - p^2} + nD = 0 \quad \Longrightarrow \quad D = -\frac{nbp}{n^2 - p^2}.$$

因此

$$x(t) = \frac{bn}{n^2 - p^2}\,(n \sin pt - p \sin nt),$$

正是所求。

绳子何时保持拉紧

想一想 究竟是甚么必须保持为正：是 x、张力，还是别的东西？

是伸长量 $e = c + x - b \sin pt$。$x(t)$ 的公式只在 $e \geq 0$ 时才描述这个运动，所以绳子自始至终保持拉紧，当且仅当这个不等式对所有 $t \geq 0$ 都成立。代入 x，

$$\begin{aligned} x - b \sin pt &= b \sin pt \left(\frac{n^2}{n^2 - p^2} - 1 \right) - \frac{bnp}{n^2 - p^2} \sin nt \\ &= \frac{bp}{n^2 - p^2}\,(p \sin pt - n \sin nt). \end{aligned}$$

所以充分必要条件（necessary and sufficient condition）是

$$c + \frac{bp}{n^2 - p^2}\,(p \sin pt - n \sin nt) \geq 0 \quad \text{对所有} t \geq 0,$$

又因为 $n^2 - p^2 > 0$，这等价于

$$bp\,(n \sin nt - p \sin pt) \leq c\,(n^2 - p^2) \quad \text{对所有} t \geq 0.$$

（如果 e 只在个别孤立的时刻触及 0，张力在那一瞬间为零，但运动方程仍然成立，绳子亦从未真正松弛。这就是不等式不必严格的原因。）

现在假设 $pb < (n - p)c$。对任何 t，由于 $\sin nt \leq 1$ 和 $-\sin pt \leq 1$，且 $n, p > 0$，

$$n \sin nt - p \sin pt \leq n + p.$$

取 $b > 0$（即振幅），便有 $bp > 0$，所以

$$bp\,(n \sin nt - p \sin pt) \leq bp\,(n + p) < (n - p)c\,(n + p) = c\,(n^2 - p^2).$$

条件成立，而且其实是严格不等式，所以绳子自始至终保持拉紧。（假如容许 b 为负，只要以 $|b|$ 代替 b，同样的论证仍然适用，因为 $|n \sin nt - p \sin pt| \leq n + p$。）

验算

- 取 $t=0$：$x(0)=0$ 及 $\dot{x}(0)=\frac{bn}{n^2-p^2}(np-pn)=0$，与应有的结果相符。
- 当 $p\to 0$ 时，端点 A 几乎不动，而 $x(t)\approx \frac{bn}{n^2}(npt-pnt)\to 0$。一如所料，甚么也没有发生。
- 最小伸长量不可能低于 $c-bp(n+p)/(n^2-p^2)=c-bp/(n-p)$，而此值为正，恰好当 $pb<(n-p)c$。
- 数值上，以随机选取的 m,λ,a,c,b,p 对完整的胡克定律方程作逐步（RK4）积分，结果与 $x(t)$ 的公式相符，误差在 4×10^{-10} 之内。取 $n=1$、$p=1/\sqrt{2}$、$c=1$，而 b 只比 $(n-p)c/p$ 高出 5%，则到 $t=20\,000$ 之前，伸长量确实会变成负数（约为 -0.05）。

常见错误

- 把 x 本身当作伸长量。上端是会移动的，所以伸长量是 $c+x-b\sin pt$。
- 忘记了 $\dot{x}(0)=0$ 会令解中必须出现 $\sin nt$ 项。没有这一项，质点起初的速度便会是 $n^2bp/(n^2-p^2)$。
- 把拉紧的条件写成 $x\geq -c$。只有当 A 固定不动时，这才是对的。

STEP 之外

这是一个无阻尼的受迫振子。它的响应由驱动频率（driving frequency）p 和固有频率（natural frequency）n 混合而成，而因子 $1/(n^2-p^2)$ 在 $p\to n$ 时趋向无穷大，这就是共振（resonance）。当 n/p 是无理数时，(nt,pt) 模 2π 的值会稠密地填满整个范围，所以 $n\sin nt-p\sin pt$ 可以任意接近 $n+p$。于是充分条件 $pb<(n-p)c$ 基本上也是必要条件（等号仍然容许）。关于受迫振动、共振和阻尼（damping），Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）有详尽的论述；微分方程方面，则可参看 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）。

详细解答

STEP II 1991，第 13 题

力学 · 刚体转动

STEP II 1991，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/91-s2-q13

关键思路

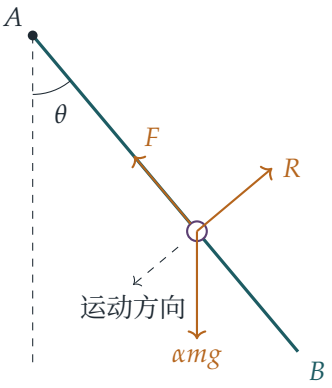
在小环滑动之前，杆和小环如同单一刚体（rigid body）一起绕 A 转动。因此我们可以求出这个组合体的角速率（angular speed，由能量求得）和角加速度（angular acceleration，由力矩（moments）求得），然后用它们算出杆必须对小环施加甚么力，才能令小环继续随杆运动。只有当小环所需的摩擦力（friction）不超过法向反作用力（normal reaction）的 μ 倍时，小环才会留在原位；这个条件首次不成立的角度，就是开始滑动的角度。

解答

模型

我们把小环视为与 A 相距 x 的质点，把支点视为光滑，并把杆视为刚体。设 θ 为杆与向下铅垂线之间的夹角，所以释放时 $\theta = \frac{\pi}{2}$，而杆向下摆动时 θ 递减。以 $\omega = |\dot{\theta}|$ 表示角速率，以 $\beta = |\ddot{\theta}|$ 表示角加速度的大小。

杆对小环施加两个力：一个是垂直于杆的法向反作用力 R，以小环运动的方向为正；另一个是沿杆的摩擦力 F，以指向 A 的方向为正。由于小环是套在杆上的，R 的正负号两种情况都有可能出现。



图中把 R 画成与运动方向相反，因为当 $k^2 > xd$ 时结果正是如此，但在代数运算中我们保留一般的正负号。

刚体的运动

小环固定在杆上时，整体绕 A 的转动惯量（moment of inertia）为 $mk^2 + \alpha m x^2 = m(k^2 + \alpha x^2)$。自释放起，杆的质心（centre of mass）下降了 $d\cos\theta$，小环下降了 $x\cos\theta$，所以由能量守恒（conservation of energy）可得

$$\tfrac{1}{2}m(k^2 + \alpha x^2)\,\omega^2 = mg(d + \alpha x)\cos\theta \quad \Longrightarrow \quad \omega^2 = \frac{2g(d + \alpha x)\cos\theta}{k^2 + \alpha x^2}.$$

对 A 取力矩（支点的力没有力矩），两个重量的总力矩为 $mg(d + \alpha x)\sin\theta$，所以

$$\beta = \frac{g(d + \alpha x)\sin\theta}{k^2 + \alpha x^2}.$$

小环所受的力

小环沿半径为 x 的圆运动。它的加速度有两个分量：沿杆指向 A 的 $x\omega^2$，以及垂直于杆、沿运动方向的 $x\beta$。沿杆向外的方向与向下铅垂线的夹角为 θ，所以重量沿杆向外的分量为 $\alpha mg\cos\theta$，沿运动方向的分量为 $\alpha mg\sin\theta$。

沿杆、指向 A 的方向：$F - \alpha mg\cos\theta = \alpha m x\omega^2$，所以

$$\begin{aligned} F &= \alpha mg\cos\theta\left(1 + \frac{2x(d + \alpha x)}{k^2 + \alpha x^2}\right) \\ &= \frac{\alpha mg\cos\theta\,(k^2 + 3\alpha x^2 + 2xd)}{k^2 + \alpha x^2}. \end{aligned}$$

垂直于杆、沿运动方向：

$$R + \alpha mg\sin\theta = \alpha m x\beta \quad \Longrightarrow \quad R = \alpha mg\sin\theta\left(\frac{x(d + \alpha x)}{k^2 + \alpha x^2} - 1\right) = \frac{\alpha mg\sin\theta\,(xd - k^2)}{k^2 + \alpha x^2}.$$

留意 αx^2 的项在 R 中互相抵消，这就是为甚么只出现 $k^2 - xd$。

想一想 小环倾向朝哪个方向滑动？R 的正负号又代表甚么？

当 $0 < \theta < \frac{\pi}{2}$ 时，我们有 $F > 0$：杆必须把小环拉向 A，所以若摩擦力不足，小环便会向外滑，即滑向 B。若 $k^2 > xd$，则 $R < 0$，即杆正在拖住小环。若只有小环自己，它会比杆上这一点下落得更快，所以杆要顶住它。

何时开始滑动?

只要 $|F| \le \mu|R|$，小环便会固定在杆上。当 $k^2 > xd$ 时，F 和 $|R|$ 都含有同一个正因子 $\alpha m g/(k^2 + \alpha x^2)$，所以条件是

$$\cos\theta\,(k^2 + 3\alpha x^2 + 2xd) \le \mu \sin\theta\,(k^2 - xd), \quad \text{即} \quad \mu\tan\theta \ge \frac{3\alpha x^2 + k^2 + 2xd}{k^2 - xd}.$$

释放时，$\theta = \frac{\pi}{2}$ 而 $F = 0$，所以小环不会立即滑动。当 θ 由 $\frac{\pi}{2}$ 递减至 0 时，$\mu\tan\theta$ 由 $+\infty$ 平稳地递减至 0，而右边是一个固定的正数。所以不等式首次不成立、小环开始向外滑动的时刻，就是在满足下式的唯一角度：

$$\boxed{\mu\tan\theta = \frac{3\alpha x^2 + k^2 + 2xd}{k^2 - xd}}$$

这正是所要求的结果。小环必定在杆到达铅垂位置之前滑动。

第 (i) 部分：$k^2 = xd$

现在在任何角度都有 $R = 0$，所以可提供的最大摩擦力 $\mu|R|$ 为零。但只要 $\theta < \frac{\pi}{2}$，便有 $F > 0$。所以不论 μ 的值是多少，杆一释放，小环便开始沿杆向外滑动。从物理上看，释放时杆上与 A 相距 $x = k^2/d$ 的那一点，开始下落时的加速度恰好是 g（因为在 $\theta = \frac{\pi}{2}$ 时 $x\beta = g$）。所以小环对杆的两侧都没有压力，摩擦力也就无从产生。

第 (ii) 部分：$k^2 < xd$

现在 $R > 0$：杆沿运动方向推动小环，因为杆上这一点在释放时下落的加速度会大于 g。小环同样倾向向外滑动，而条件 $F \le \mu R$ 变成

$$\mu\tan\theta \ge \frac{3\alpha x^2 + k^2 + 2xd}{xd - k^2}.$$

按照与之前相同的论证，小环起初保持不动，并在以下情况开始向外滑动：

$$\mu\tan\theta = \frac{3\alpha x^2 + k^2 + 2xd}{xd - k^2},$$

这就是把题目所给结果中的 $k^2 - xd$ 换成 $|k^2 - xd|$。唯一的分别是法向反作用力现在作用在小环的另一侧。

验算

- 当 $k^2 > xd$ 时，随着 $\mu \to \infty$，滑动角度 $\theta \to 0$：非常粗糙的杆几乎可以把小环一直留住到最低点。当 $k^2 - xd \to 0$ 时，该角度趋向 $\frac{\pi}{2}$，与第 (i) 部分吻合。
- 长度为 $2a$ 的均匀杆有 $d = a$ 及 $k^2 = \frac{4}{3}a^2$，所以临界位置 $x = k^2/d = \frac{4}{3}a$ 就是撞击中心（centre of percussion），即杆释放时以自由落体方式开始运动的那一点。
- 在数值方面，我们对刚体运动进行积分，用有限差分算出小环的加速度，并就随机选取的参数值重新求得 F 和 R，误差在 10^{-5} 之内。例如，当 $k^2 - xd \approx 2.87$ 时，小环首次需要多于 $\mu|R|$ 的摩擦力是在 $\theta \approx 1.3270$，与公式一致。$k^2 < xd$ 的情况亦以同样方式与 (ii) 的公式吻合。

常见错误

- 忘记小环本身对转动惯量和对重量力矩的贡献。在滑动之前，小环是刚体的一部分。
- 使用摩擦定律时，以重量的法向分量代替实际的法向反作用力 R，而这里的法向反作用力还取决于角加速度。
- 在沿杆方向的方程中漏掉向心（centripetal）项 $x\omega^2$，而 $3\alpha x^2$ 中的因子 3 正是由此而来。

STEP 之外

与支点相距 k^2/d 的那一点就是撞击中心。在该处击打，支点不会产生冲击反作用力，这就是板球棒或网球拍有“甜点”的原因。它同时也是与该复摆（compound pendulum）周期相同的单摆的长度。这里所用的方法，是先求出刚体的运动，再问需要甚么内力来维持这种运动，这正是拉格朗日力学（Lagrangian mechanics）中分析约束力（constraint forces）的基础。想进一步了解刚体转动和复摆，可阅读 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）。

详细解答

STEP II 1991，第 14 题

力学 · 相对速度与追踪

STEP II 1991，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：

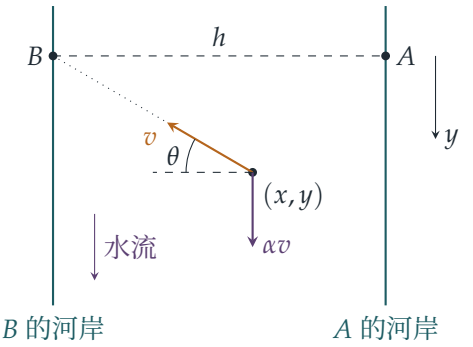
step.maths.org/questions/91-s2-q14

关键思路

船相对于地面的速度，等于船相对于水的速度加上水流的速度。“船头始终指向 B”是一个几何条件，即 $y = x\tan\theta$；对它求导，便得到同一个速度的第二个表达式。把两个表达式对照，就得出一条连系 x 与 θ 的可分离变量（separable）方程。选用代换 $\tan\theta = \sinh(\alpha s)$，是因为它令 $\sec\theta + \tan\theta$ 变成一个指数函数，于是所有式子都化为 s 的简单指数函数。

解答

我们把船当作质点（particle）来建立模型。取 B 为原点，x 由 B 所在的河岸量向 A，y 则沿下游方向量度，因此 A 位于 $(h,0)$，而船由该处出发。



以 θ 表示速度

船头指向 B 所在的河岸，与方向 AB 成 θ 角，并偏向上游（以抵消水流造成的漂移）。因此，以我们的 (x,y) 分量表示，船相对于水的速度（relative velocity）是沿单位向量

$-(\cos\theta, \sin\theta)$ 的 v。再加上水流，即 y 方向上的 αv，便得

$$\begin{pmatrix}\dot{x}\\ \dot{y}\end{pmatrix} = \begin{pmatrix}-v\cos\theta\\ \alpha v - v\sin\theta\end{pmatrix}.$$

指向 B

由位于 (x, y) 的船到 B 的方向是沿 $(-x, -y)$，而这个方向与 AB 成 θ 角，当且仅当 $\tan\theta = y/x$。所以在整个运动过程中，

$$y = x\tan\theta.$$

利用积法则（product rule）和链式法则（chain rule）对 t 求导，

$$\dot{y} = \tan\theta\,\frac{\mathrm{d}x}{\mathrm{d}t} + x\sec^2\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}t},$$

所以速度向量 $(\dot{x}, \dot{y})$ 具有题目要求的形式。

x 与 θ 的微分方程

现在把两个表达式对照。由 x 分量得 $\dot{x} = -v\cos\theta$。把它代入 y 分量，

$$\begin{aligned}\tan\theta\,(-v\cos\theta) + x\sec^2\theta\,\dot{\theta} &= \alpha v - v\sin\theta\\ x\sec^2\theta\,\dot{\theta} &= \alpha v, \qquad (*)\end{aligned}$$

这是因为 $\tan\theta\cos\theta = \sin\theta$，两边的 $-v\sin\theta$ 项互相抵消。特别地，$\dot{\theta} > 0$，所以 θ 由 0 起不断增加，我们可以把 x 看成 θ 的函数。两式相除，

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{\dot{x}}{\dot{\theta}} = \frac{-v\cos\theta\cdot x\sec^2\theta}{\alpha v} = -\frac{x\sec\theta}{\alpha}, \qquad \text{所以} \qquad \alpha\frac{\mathrm{d}x}{\mathrm{d}\theta} = -x\sec\theta,$$

正是所求。

求解 x

这条方程可以分离变量（船到达之前 $x > 0$）：

$$\alpha\int\frac{\mathrm{d}x}{x} = -\int\sec\theta\,\mathrm{d}\theta \quad\Longrightarrow\quad \alpha\ln x = -\ln(\sec\theta + \tan\theta) + C,$$

这里用了标准积分 $\int\sec\theta\,\mathrm{d}\theta = \ln(\sec\theta + \tan\theta)$，它在 $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ 时成立。开始时船在 A，该处 $x = h$ 而 $\theta = 0$，所以 $C = \alpha\ln h$。因此

$$\alpha\ln\frac{x}{h} = -\ln(\sec\theta + \tan\theta), \qquad \text{即} \qquad \left(\frac{x}{h}\right)^{\alpha} = (\sec\theta + \tan\theta)^{-1}.$$

代换 $\tan\theta = \sinh(\alpha s)$

> **想一想** 以 s 表示，$\sec\theta$ 是甚么？

由于 $\sec^2\theta = 1+\tan^2\theta = 1+\sinh^2(\alpha s) = \cosh^2(\alpha s)$，而 $\sec\theta$ 和 $\cosh(\alpha s)$ 都是正数（因为 $0 \le \theta < \frac{\pi}{2}$），我们得到 $\sec\theta = \cosh(\alpha s)$。于是

$$\sec\theta + \tan\theta = \cosh(\alpha s) + \sinh(\alpha s) = \mathrm{e}^{\alpha s},$$

所以 $(x/h)^\alpha = \mathrm{e}^{-\alpha s}$，两边取 α 次方根，得 $x = h\mathrm{e}^{-s}$。

至于变率的方程，把 $\tan\theta = \sinh(\alpha s)$ 对 t 求导：

$$\sec^2\theta\,\dot\theta = \alpha\cosh(\alpha s)\,\dot s.$$

把这个结果和 $x = h\mathrm{e}^{-s}$ 代入 $(*)$，得 $h\mathrm{e}^{-s}\cdot\alpha\cosh(\alpha s)\,\dot s = \alpha v$，再除以 α，

$$h\mathrm{e}^{-s}\cosh(\alpha s)\frac{\mathrm{d}s}{\mathrm{d}t} = v,$$

正是所求。

渡河所需的时间

在 A 处，$\theta = 0$，所以 $s = 0$。由于 $\dot s > 0$，s 不断增加，而当 $x = h\mathrm{e}^{-s} \to 0$，即 $s \to \infty$ 时，船到达 B。（船确实会到达 B：因为 $\alpha < 1$，所以 $y = x\tan\theta = h\mathrm{e}^{-s}\sinh(\alpha s) \to 0$。）分离变量后，渡河时间为

$$\begin{aligned} T &= \frac{h}{v}\int_0^\infty \mathrm{e}^{-s}\cosh(\alpha s)\,\mathrm{d}s = \frac{h}{2v}\int_0^\infty \left(\mathrm{e}^{-(1-\alpha)s} + \mathrm{e}^{-(1+\alpha)s}\right)\mathrm{d}s \\ &= \frac{h}{2v}\left(\frac{1}{1-\alpha} + \frac{1}{1+\alpha}\right) = \frac{h}{2v}\cdot\frac{2}{1-\alpha^2} = \frac{h}{v(1-\alpha^2)}, \end{aligned}$$

其中两个积分都收敛（converge），因为 $1-\alpha > 0$。所以渡河所需的时间是 $hv^{-1}(1-\alpha^2)^{-1}$。

验算

- 没有水流时（$\alpha = 0$），船笔直驶过对岸，$T = h/v$，与预期相符。当 $\alpha \to 1$ 时，所需时间趋向无限大：船若不比水流快，在 B 附近船头正对上游时，便永远无法逆流前进。
- 我们以数值方法逐步计算这个运动（四阶龙格－库塔法（Runge–Kutta），每一步都把船头对准 B）。取 $h = v = 1$ 和 $\alpha = \frac{1}{2}$，船在 $t \approx 1.33333$ 时到达，与 $1/(1-\frac{1}{4}) = \frac{4}{3}$ 吻合。沿着整条路径，$(x/h)^\alpha(\sec\theta+\tan\theta)$ 一直等于 1，误差约为 10^{-13}。

常见错误

- 弄错船头方向的 y 分量的正负号。船会被冲向下游，所以指向 B 就是要部分指向上游；当 $\theta > 0$ 时，这个分量是 $-v\sin\theta$。
- 忘记说明为何 $\sec\theta = +\cosh(\alpha s)$：这需要 θ 一直小于 $\frac{\pi}{2}$，而这是成立的，因为 θ 是由 B 到一个 $x > 0$ 的点的直线所成的角。

STEP 之外

这是一道经典的追逐曲线（pursuit curve）问题：一个物体不断把方向对准目标，而它的路径可由一条关于运动方向的微分方程求得。同样的想法可以描述狗追赶正在走动的主人，或者导弹追踪目标，并且自然地引出曳物线（tractrix）一类的曲线。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教材）很适合用来熟习相对速度，以及在运动参考系（moving frames）中建立运动方程。

详细解答

STEP II 1991，第 15 题

概率与统计 · 期望值的线性性质

STEP II 1991，第 15 题

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

step.maths.org/questions/91-s2-q15

关键思路

把数对 (n_1, n_2) 看成 $m \times m$ 方格表中的一格，全部 m^2 格出现的机会均等。满足 $|n_1 - n_2| = k$ 的格子位于与主对角线平行的两条对角线上，所以很容易数出来。之后的一切都来自*期望值的线性性质*（*linearity of expectation*）：若干个差之和的期望值，等于各个期望值之和，即使相邻的项共用同一个变量、并非独立（independent）亦然。监考员（invigilator）所走的路程也是同一类的和，只是第一段路由一张固定的书桌出发，而不是由随机的一张出发。

解答

$|n_1 - n_2|$ 的分布

由于 n_1 与 n_2 独立，且各自在 $\{1, \ldots, m\}$ 上均匀分布（uniform），m^2 个有序数对 (n_1, n_2) 中每一个的概率都是 $1/m^2$。

固定 k，其中 $1 \le k \le m-1$。满足 $n_1 - n_2 = k$ 的数对是 $(k+1, 1), (k+2, 2), \ldots, (m, m-k)$，共 $m-k$ 对。由对称性，满足 $n_2 - n_1 = k$ 的数对另有 $m-k$ 对，而因为 $k \neq 0$，这两组数对并不重叠。因此

$$\mathrm{P}\big(|n_1 - n_2| = k\big) = \frac{2(m-k)}{m^2},$$

正如所求。（当 $k = 0$ 时有 m 对，所以 $\mathrm{P}(|n_1 - n_2| = 0) = 1/m$。作为验算，$\frac{1}{m} + \sum_{k=1}^{m-1} \frac{2(m-k)}{m^2} = \frac{1}{m} + \frac{m(m-1)}{m^2} = 1$。）

$|n_1 - n_2|$ 的期望值

$k = 0$ 这个值对期望值没有贡献，所以

$$\begin{aligned}\mathrm{E}|n_1 - n_2| &= \sum_{k=1}^{m-1} k \cdot \frac{2(m-k)}{m^2} = \frac{2}{m^2}\left(m\sum_{k=1}^{m-1} k - \sum_{k=1}^{m-1} k^2\right)\\ &= \frac{2}{m^2}\left(\frac{m\cdot m(m-1)}{2} - \frac{(m-1)m(2m-1)}{6}\right)\\ &= \frac{2}{m^2}\cdot\frac{m(m-1)\big(3m-(2m-1)\big)}{6} = \frac{(m-1)(m+1)}{3m} = \frac{m^2-1}{3m},\end{aligned}$$

当中用了标准求和公式 $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$ 和 $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$，并取 $n = m-1$。

$m = 2$ 而有三个整数的情况

当 $m = 2$ 时，共有 $2^3 = 8$ 个机会均等的三元组 (n_1, n_2, n_3)。记 $S = |n_1 - n_2| + |n_2 - n_3|$：

(n_1, n_2, n_3)	111	112	121	122	211	212	221	222
S	0	1	2	1	1	2	1	0

总和是 8，所以 $\mathrm{E}(S) = 8/8 = 1$。题目所述的值是 $2(m^2-1)/(3m) = 2\cdot 3/6 = 1$，两者吻合。

> ? **想一想** $|n_1 - n_2|$ 和 $|n_2 - n_3|$ 两项都涉及 n_2。我们还可以直接把它们的期望值相加吗?

可以。期望值的线性性质 $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ 完全不需要独立性。每一对 (n_i, n_{i+1}) 单独来看，都是两个独立而均匀分布的整数，所以每一项的期望值都是 $(m^2-1)/(3m)$。一共有 $r-1$ 项，所以

$$\mathrm{E}\big(|n_1 - n_2| + |n_2 - n_3| + \cdots + |n_{r-1} - n_r|\big) = \frac{(r-1)(m^2-1)}{3m}.$$

$r = 3$ 的情况，就是我们刚才就 $m = 2$ 验证过的结果。

监考员

把书桌沿直线依次编号为 $1, 2, \ldots, m$，于是书桌 i 与书桌 j 相距 $d|i-j|$。设 c_j 为第 j 次召唤所来自的书桌的编号，其中 $j = 1, \ldots, N+1$。我们把召唤者设置为互相独立，且各自在 $\{1, \ldots, m\}$ 上均匀分布。这就是我们对“m 名考生中任何一人召唤的机会均等”的理解。由于考生从不同时召唤，也不会在另一次召唤进行期间召唤，监考员会逐一处理各次召唤，而当第 $j+1$ 次召唤出现时，监考员正站在书桌 c_j 旁。（若同一名考生连续召唤两次，该段路的长度为 0，而公式已经涵盖这种情况。）

因此所走的总距离是

$$D = d\,|c_1 - 1| + d\sum_{j=1}^{N} |c_{j+1} - c_j|.$$

*第一段路。*监考员由书桌 1 出发，而 c_1 是均匀分布的，所以

$$\mathrm{E}|c_1 - 1| = \frac{1}{m}\sum_{i=1}^{m}(i-1) = \frac{m-1}{2}.$$

留意这个值比 $(m^2-1)/(3m)$ 大。由这一列书桌的尽头出发，令第一段路比平均的长。

*其余 N 段路。*这里 $c_1, \ldots, c_{N+1}$ 担当 $n_1, \ldots, n_r$ 的角色，其中 $r = N+1$，所以由上面的结果可得

$$\mathrm{E}\sum_{j=1}^{N} |c_{j+1} - c_j| = \frac{N(m^2-1)}{3m}.$$

*合并。*再用一次线性性质，

$$\begin{aligned}\mathrm{E}(D) &= d\left(\frac{m-1}{2} + \frac{N(m-1)(m+1)}{3m}\right)\\ &= \frac{d(m-1)}{6m}\big(3m + 2N(m+1)\big) = \frac{d(m-1)}{6m}\big[2N(m+1) + 3m\big],\end{aligned}$$

这正是所求的结果。

验算

- 当 $m = 1$ 时只有一张书桌，所以监考员从不走动。公式给出 0，理应如此。
- 当 $N = 0$ 时只有第一次召唤，公式化简为 $d(m-1)\cdot 3m/(6m) = \frac{1}{2}d(m-1)$，即第一段路的期望值。
- 把召唤者的每一个可能序列逐一列举，可确认公式对所有 $m \le 4$ 及 $N \le 3$ 都成立。以 $m = 10$、$N = 20$、$d = 1$ 进行 2×10^5 次模拟，所得平均值为 70.50，而公式给出的是 $\frac{9}{60}(440 + 30) = 70.5$。

常见错误

- 把第一段路当作与其余各段一样处理。监考员由书桌 1 出发，而不是由随机的一张书桌出发，所以该段路的期望值是 $\frac{1}{2}(m-1)$，而不是 $(m^2-1)/(3m)$。
- 只数 $n_1 - n_2 = k$ 的情况，忘记了因 $n_2 - n_1 = k$ 而来的因子 2；或者错误地把 $k = 0$ 的情况也乘以二。
- 以为和式中各项必须互相独立，它们的期望值才可以相加。

STEP 之外

监考员的位置是一条马可夫链（*Markov chain*）：他们下一步往哪里去，只取决于他们现在身在何处。这里的链非常简单，每一步都把过去完全“忘记”。在较复杂的模型中（例如召唤较可能来自附近的书桌），我们便需要这条链的平稳分布（stationary distribution），才能求出长远而言每次召唤平均所走的路程。第一个结果的连续版本是：在 $[0,1]$ 上两个独立而均匀分布的点，平均相距 $\frac{1}{3}$，这正是 $(m^2-1)/(3m)$ 除以 m 后的极限。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）涵盖了期望值的线性性质和马可夫链。

详细解答

STEP II 1991，第 16 题

概率与统计 · 二项分布的正态近似

> **STEP II 1991，第 16 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/91-s2-q16

关键思路

追踪水位*差额*（deficit）：即水位低于坝顶多少。排水令差额稳定上升，而每场骤雨则令它恰好下降 V；溢流就是差额变为负数。在两天假期的问题中，只有一种降雨情况有可能引致溢流，而且届时只有一个均匀分布的时刻起作用。到了暑假，排水已经关掉，于是一切归结为数算骤雨的次数，而这个次数服从二项分布（binomial distribution）。

解答

第 (i) 部分：两天假期

以日为单位，由假期开始的午夜起计算时间，并设 $D(t)$ 为时刻 t 的差额，故 $D(0) = V$。在两场骤雨之间，D 以速率 u 上升，而每场骤雨令它下降 V。只要 D 一旦变为负数，水坝便会溢流。（我们按题目的原意假设，水库存水充足，排水在这两天内可以一直持续。）

设 R_1、R_2 分别为第 1 天和第 2 天下雨的事件，并设下雨的时刻（以一天的分数表示）为 T_1 和 T_2，两者各自在 $[0,1]$ 上服从均匀分布（uniform distribution）。各天互相独立（independent），这一点已隐含在“任何一天下雨的概率为 p”这句话之中。

- **没有下雨：**差额只会增加；不会溢流。
- **只有一天下雨**，时刻为 s：骤雨刚过后 $D = V + us - V = us \geq 0$，所以不会溢流（概率为零的事件 $s = 0$ 除外）。
- **两天都下雨：**第一场骤雨过后 $D = uT_1 \geq 0$，这是安全的。第二场骤雨刚过后，即在时刻 $1 + T_2$，

$$D = V + u(1 + T_2) - 2V = u(1 + T_2) - V.$$

留意 T_1 已经消失：无论第一场骤雨何时落下，排水的速率都一样。

所以，溢流恰好在两天都下雨而且 $u(1+T_2) < V$ 时发生，即 $T_2 < \frac{V}{u} - 1$。若 $\frac{1}{2}V \le u \le V$，则 $\frac{V}{u} - 1$ 这个数位于 $[0,1]$ 之内，而由于 T_2 服从均匀分布，

$$\mathrm{P}(\text{溢流}) = p^2\left(\frac{V}{u} - 1\right).$$

令上式等于 Q：

$$\frac{V}{u} = 1 + \frac{Q}{p^2} \quad \Longrightarrow \quad u = \frac{p^2 V}{p^2 + Q}.$$

条件 $Q < p^2$ 正是令这个结果前后一致的原因：由 $0 < Q/p^2 < 1$ 可得 $\frac{1}{2}V < u < V$，与假设相符。（没有这个条件便无解。溢流的概率永远不会超过 p^2，因为必须两天都下雨。）

第 (i) 部分：暑假

现在 $u = 0$，所以差额永不增加。它由 kV 开始，每场骤雨令它下降 V，因此若 N 为 18 天之中下雨的日数，则

$$\text{溢流} \iff NV > kV \iff N > k.$$

各天互相独立，每天下雨的概率为 $\frac{1}{3}$，所以 $N \sim \mathrm{B}(18, \frac{1}{3})$，其平均值为 6，方差（variance）为 $18 \cdot \frac{1}{3} \cdot \frac{2}{3} = 4$。

> **想一想** 这里哪一种近似才算“合适”？为甚么需要连续性修正（continuity correction）？

由于 $np = 6$ 而 $n(1-p) = 12$，正态近似（normal approximation）$N \approx \mathrm{N}(6,4)$ 是合理的。（泊松近似（Poisson approximation）则不合适，因为 p 并不小。）N 只取整数值，所以对整数 k 而言，事件 $N > k$ 就是事件 $N \ge k+1$，我们以 $X > k + \frac{1}{2}$ 来近似它，其中 $X \sim \mathrm{N}(6,4)$。我们需要

$$\mathrm{P}\left(Z > \frac{k + \frac{1}{2} - 6}{2}\right) < \frac{1}{10} \iff \frac{k - 5.5}{2} > 1.2816 \iff k > 8.06.$$

具体来说，$k = 8$ 给出 $\mathrm{P}(Z > 1.25) \approx 0.106$，太大了；而 $k = 9$ 给出 $\mathrm{P}(Z > 1.75) \approx 0.040$，符合要求。由于溢流与否只取决于 $N > k$ 是否成立，介乎 8 与 9 之间的非整数 k 并不比 $k = 9$ 好。所以工程师应把水库排水至

$$k = 9,$$

即低于坝顶 $9V$。

第 (ii) 部分

现在共有 36 个半天，互相独立，每个半天下雨的概率为 $\frac{1}{6}$。排水量仍然是零，所以溢流同样表示 $N > k$，只是现在 $N \sim \mathrm{B}(36, \frac{1}{6})$。平均值仍然是 6，但方差是 $36 \cdot \frac{1}{6} \cdot \frac{5}{6} = 5$，比 4 大。

平均值相同时，方差较大即表示上尾较重：对每个高于平均值的 k，$\mathrm{P}(N > k)$ 都比之前大。所以所需的 k 较高。以 $\mathrm{N}(6, 5)$ 连同连续性修正作近似，我们需要

$$\frac{k - 5.5}{\sqrt{5}} > 1.2816 \iff k > 8.37,$$

而之前则是 $k > 8.06$。（这里正态近似给出 $\mathrm{P}(N \geq 9) \approx 0.132$，而第 (i) 部分则是 0.106。）所以门槛上升了。若把 k 向上舍入至整数，两种情况仍然都得出 $k = 9$。不过在 $k = 9$ 时，尾部概率现在约为 0.06 而不是 0.04，所以安全余裕较小。

验算

- 以蒙地卡罗模拟（Monte Carlo simulation）模拟两天假期（400,000 次），取 $p = 0.5$、$Q = 0.1$ 及 $u = \frac{0.25}{0.35} V \approx 0.714V$，得出的溢流频率为 0.0996。$p$ 和 Q 取其他数值时，结果同样吻合。
- 极限情况：当 $Q \to p^2$ 时，$u \to \frac{1}{2}V$，这是两天下雨仍可引致溢流的最慢排水速率。当 $Q \to 0$ 时，$u \to V$，在这个速率下，第二场骤雨永远不会造成溢流。
- 精确的二项分布尾部概率印证了正态近似。对 $\mathrm{B}(18, \frac{1}{3})$，$\mathrm{P}(N \geq 9) = 0.1076$ 及 $\mathrm{P}(N \geq 10) = 0.0433$。对 $\mathrm{B}(36, \frac{1}{6})$，$\mathrm{P}(N \geq 9) = 0.1332$ 及 $\mathrm{P}(N \geq 10) = 0.0652$。

常见错误

- 以为在两天假期的问题中，第一场雨的时刻 T_1 有影响。其实它会互相抵消，因为排水速率是恒定的。
- 忘记连续性修正。没有它，不等式便变成 $k > 8.56$，掩盖了 N 是整数这个事实。
- 因为平均值相等，便说 (ii) 的答案“一样”。其实控制尾部的是分布的离散程度。

STEP 之外

雨水在随机的时刻到来，而水则以稳定的速率排走，这是应用概率中的一个经典模型：储存理论（storage theory），由 Moran 为研究水坝而发展出来。它与排队（queues）和保险破产（insurance ruin）问题关系密切。在第 (ii) 部分，若把一天分割成愈来愈多、愈

来愈小的时段，同时保持平均值不变，则 B$(n, 6/n)$ 趋向平均值为 6、方差为 6 的泊松分布（Poisson distribution），而雨水的到来则成为一个泊松过程（Poisson process）。想进一步了解，可参阅 Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）。

详细解答

STEP III 1991，第 1 题

纯数学 · 级数与展开式

STEP III 1991，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/91-s3-q1

关键思路

三个部分都值得先想一想才动手计算。在 (i) 中，部分分式（partial fractions）把和式拆成可以裂项相消（telescope）的几部分。在 (ii) 中，三次式 $1+x+x^2+x^3$ 是一个等比级数（geometric series），因此可以写成两个非常简单的式子之商，而对数亦随之分拆成两个标准级数。在 (iii) 中，先把指数写成级数；由于它由 x^2 开始，指数级数（exponential series）中只有头几项能够到达 x^4。

解答

第 (i) 部分

我们要找出常数，使得

$$\frac{6}{r(r+1)(r+3)} = \frac{A}{r} + \frac{B}{r+1} + \frac{C}{r+3}.$$

利用掩盖法（cover-up rule），$A = \dfrac{6}{1\cdot 3} = 2$，$B = \dfrac{6}{(-1)(2)} = -3$ 及 $C = \dfrac{6}{(-3)(-2)} = 1$。（快速检查一下：$A+B+C=0$，这是必然的，因为当 r 很大时，左方的表现有如 $6/r^3$，并没有 $1/r$ 项。）

系数 $2, -3, 1$ 之和为零，这提示我们把它们拆成 $2-2$ 和 $-1+1$：

$$\frac{6}{r(r+1)(r+3)} = 2\left(\frac{1}{r} - \frac{1}{r+1}\right) - \left(\frac{1}{r+1} - \frac{1}{r+3}\right).$$

现在每个括号都可以裂项相消。第一个给出

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

在第二个之中，两项相隔两步，因此两端各有两项留下：

$$\sum_{r=1}^{n}\left(\frac{1}{r+1}-\frac{1}{r+3}\right)=\frac{1}{2}+\frac{1}{3}-\frac{1}{n+2}-\frac{1}{n+3}.$$

把它们合起来，

$$\begin{aligned}\sum_{r=1}^{n}\frac{6}{r(r+1)(r+3)}&=2-\frac{2}{n+1}-\frac{5}{6}+\frac{1}{n+2}+\frac{1}{n+3}\\&=\frac{7}{6}-\frac{2}{n+1}+\frac{1}{n+2}+\frac{1}{n+3}.\end{aligned}$$

第 (ii) 部分

? **想一想** $1+x+x^2+x^3$ 有没有更简洁的写法?

它是公比为 x 的等比级数，所以当 $x\neq 1$ 时

$$1+x+x^2+x^3=\frac{1-x^4}{1-x}.$$

当 $|x|<1$ 时，$1-x^4$ 和 $1-x$ 均为正数，所以我们可以把对数拆开：

$$\ln(1+x+x^2+x^3)=\ln(1-x^4)-\ln(1-x).$$

利用 $\ln(1-y)=-\sum_{k\geq 1}y^k/k$，此式对 $|y|<1$ 成立（而这里亦有 $|x^4|<1$），

$$-\ln(1-x)=\sum_{k=1}^{\infty}\frac{x^k}{k},\qquad \ln(1-x^4)=-\sum_{m=1}^{\infty}\frac{x^{4m}}{m}.$$

第二个级数只涉及 k 为 4 的倍数的幂 x^k。对于这样的 $k=4m$，合并后的系数为

$$\frac{1}{4m}-\frac{1}{m}=-\frac{3}{4m}=-\frac{3}{k}.$$

因此

$$\ln(1+x+x^2+x^3)=x+\frac{x^2}{2}+\frac{x^3}{3}-\frac{3x^4}{4}+\frac{x^5}{5}+\cdots,$$

而通项为

$$\frac{x^k}{k}\ \text{若}\ 4\nmid k,\qquad -\frac{3x^k}{k}\ \text{若}\ 4\mid k.$$

每个系数都不是零，所以上面列出的五项就是首五个非零项。

第 (iii) 部分

当 $-1<x\leq 1$ 时，我们有 $\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots$，所以

$$u=x\ln(1+x)=x^2-\frac{x^3}{2}+\frac{x^4}{3}-\cdots.$$

想一想 $e^u = 1 + u + \frac{u^2}{2} + \frac{u^3}{6} + \cdots$ 之中，我们需要多少项?

由于 u 由 x^2 开始，u^k 这一项便由 x^{2k} 开始。所以 u^3 及其后各项对 x^4 或以下的项毫无贡献，而从 u^2 我们只需要它的首项 x^4。因此

$$\begin{aligned} e^{x\ln(1+x)} &= 1 + \left(x^2 - \frac{x^3}{2} + \frac{x^4}{3}\right) + \frac{1}{2}x^4 + \cdots \\ &= 1 + x^2 - \frac{x^3}{2} + \frac{5x^4}{6} + \cdots . \end{aligned}$$

（当然，$e^{x\ln(1+x)}$ 其实就是 $(1+x)^x$。）

验算

- 在 (i) 中，$n = 1$ 给出 $\frac{7}{6} - 1 + \frac{1}{3} + \frac{1}{4} = \frac{3}{4}$，与单独一项 $\frac{6}{1\cdot 2\cdot 4}$ 吻合。令 $n \to \infty$，便得无穷项之和 $\frac{7}{6}$。
- 在 (ii) 中，直接展开至二阶：取 $\ln(1+y) \approx y - \frac{y^2}{2}$，其中 $y = x + x^2 + \cdots$，得出 $x + x^2 - \frac{x^2}{2} = x + \frac{x^2}{2}$，与我们的首两项一致。数值上，在 $x = 0.5$ 时，把级数加至很多项，结果与 $\ln 1.875 \approx 0.6286$ 吻合。
- 在 (iii) 中，在 $x = 0.01$ 时，展开式与 $(1.01)^{0.01}$ 相差约 7.4×10^{-11}，其大小属 x^5 的量级，正如所料（下一个系数原来是 $-\frac{3}{4}$）。

常见错误

- 在 (i) 中，忘记了相隔两步的裂项和会在两端各留下两项，而不是一项。
- 在 (ii) 中，把 x^4 的系数写成 $\frac{1}{4} - 1$，但其后却把通项写成 $(-1)^{k+1}x^k/k$ 或类似的式子；其实规律取决于 k 除以 4 的余数。
- 在 (iii) 中，漏掉 $\frac{1}{2}u^2$ 这一项：它贡献了 $\frac{1}{2}x^4$。

STEP 之外

第 (iii) 部分是幂级数（power series）复合的一个小例子，而追踪每一项中最低次幂这个记帐技巧，正是大学里处理渐近展开式（asymptotic expansions）的方法。幂级数何时收敛，以及为何在收敛半径（radius of convergence）之内可以随意把它们相加、相乘和复合，这些问题都是分析学（analysis）入门课程的内容；Abbott 的 *Understanding Analysis*（分析学入门教材）对此有清晰的论述。

详细解答

STEP III 1991，第 2 题

纯数学 · 几何中的复数

STEP III 1991，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/91-s3-q2

关键思路

一个复数（complex number）可以同时记录长度之比和一个角。商 $(z_1 - z_2)/(z_1 - z_3)$ 记下了三角形在 P_1 处两条边的长度之比（即它的模（modulus）），以及由一条边转到另一条边所转过的角（即它的辐角（argument））；而这两项资料，正是相似三角形（similar triangles）的 SAS 判别法所需要的。只要把条件写成行列式（determinant），各部分便归结为把一个行列式因式分解（factorise），或者巧妙地选取一个三角形 $Q_1Q_2Q_3$ 来作比较。

解答

先处理一个印刷错误

按题目所印，左边的分母是 $z_2 - z_3$。这不可能是对的：取 $w_i = z_i$，使两个三角形完全相同。这时右边是 $(z_1 - z_2)/(z_1 - z_3)$，由于 $z_2 \neq z_1$，它与左边并不相等。题目原意的条件，亦即与行列式吻合的条件，是

$$\frac{z_1 - z_2}{z_1 - z_3} = \frac{w_1 - w_2}{w_1 - w_3}. \qquad (*)$$

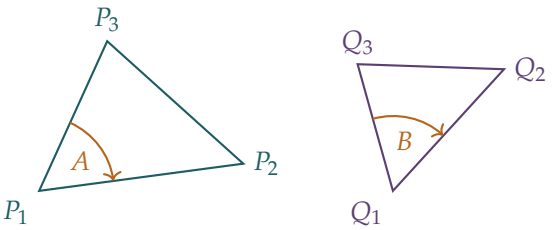
我们要证明的就是这个条件。

相似的条件

设 A 为三角形 $P_1P_2P_3$ 在 P_1 处的角，故 $0 < A < \pi$。以极式（polar form）表示，

$$\frac{z_1 - z_2}{z_1 - z_3} = \frac{z_2 - z_1}{z_3 - z_1} = \frac{P_1P_2}{P_1P_3}\,\mathrm{e}^{\mathrm{i}\theta},$$

其中 θ 是由 $\overrightarrow{P_1P_3}$ 的方向转到 $\overrightarrow{P_1P_2}$ 的方向所转过的角。因此，若 $P_1 \to P_2 \to P_3$ 沿逆时针方向排列，则 $\theta = -A$（由 P_1P_3 出发，要顺时针转才到达 P_1P_2）；若沿顺时针方向排列，则 $\theta = +A$。简而言之，这个商的辐角（取值于 $(-\pi, \pi)$）的大小等于 P_1 处的角，而其正负号则告诉我们三角形的绕向（sense）。对各个 w 而言情况相同，其中 B 是 Q_1 处的角。



同向相似 $\Rightarrow (*)$。若两个三角形相似，且 $P_i \leftrightarrow Q_i$ 互相对应，则 P_1 和 Q_1 处的两边之比相同，即 $P_1P_2/P_1P_3 = Q_1Q_2/Q_1Q_3$，而该处的角亦相等，即 $A = B$。由于两个三角形的绕向相同，两个辐角 $\mp A$ 和 $\mp B$ 的正负号相同，所以它们相等。模相等，辐角也相等，便得出 $(*)$。

$(*) \Rightarrow$ 同向相似。若 $(*)$ 成立，比较两边的模可得 $P_1P_2/P_1P_3 = Q_1Q_2/Q_1Q_3$。比较两边的辐角可得 $A = B$，而且正负号相同，所以两个三角形的绕向相同。两边成比例而夹角相等，正是 SAS 条件，故两个三角形相似，且 $P_i \leftrightarrow Q_i$ 互相对应。（若 $(*)$ 是实数，则两组三点各自共线（collinear），这两个退化的“三角形”仍然相似。）

还可以从另一个角度看同一件事：$(*)$ 恰好表示 $w_i = az_i + b$，其中常数 $a = (w_1 - w_2)/(z_1 - z_2) \neq 0$，$b = w_1 - az_1$。映射 $z \mapsto az + b$ 是倍数为 $|a|$ 的放大（enlargement），再加上转角为 $\arg a$ 的旋转（rotation）和一个平移（translation），而这样的映射保持形状和绕向不变。

行列式形式

去分母后（由于各点互不相同，两个分母均不为零），$(*)$ 等价于

$$(z_1 - z_2)(w_1 - w_3) - (z_1 - z_3)(w_1 - w_2) = 0.$$

展开后，z_1w_1 项互相抵消，余下

$$z_1(w_2 - w_3) - z_2(w_1 - w_3) + z_3(w_1 - w_2) = 0,$$

这正是该行列式沿第一行的展开式。所以 $(*)$ 等同于 $\det = 0$。

处理以下各部分时，用另一个方法计算这个行列式会较为方便。从第二列和第三列分别

减去第一列，行列式的值不变；然后沿最底一行展开，这一行现在是 $(1,0,0)$：

$$\det\begin{pmatrix} z_1 & z_2 & z_3 \\ w_1 & w_2 & w_3 \\ 1 & 1 & 1 \end{pmatrix} = \det\begin{pmatrix} z_2-z_1 & z_3-z_1 \\ w_2-w_1 & w_3-w_1 \end{pmatrix}. \tag{†}$$

第 (i) 部分

取 $w_i = z_i^2$，(†) 变成

$$\begin{aligned}(z_2-z_1)(z_3^2-z_1^2)-(z_3-z_1)(z_2^2-z_1^2) &= (z_2-z_1)(z_3-z_1)\big[(z_3+z_1)-(z_2+z_1)\big] \\ &= (z_2-z_1)(z_3-z_1)(z_3-z_2).\end{aligned}$$

由于各 z_i 互不相同，每个因子均不为零。所以行列式永不为零，两个三角形永不相似（按上述意义）。

第 (ii) 部分

取 $w_i = z_i^3$，并利用 $z_j^3 - z_1^3 = (z_j - z_1)(z_j^2 + z_j z_1 + z_1^2)$：

$$\begin{aligned}(z_2-z_1)(z_3^3-z_1^3)-(z_3-z_1)(z_2^3-z_1^3) &= (z_2-z_1)(z_3-z_1)\big[(z_3^2+z_3z_1)-(z_2^2+z_2z_1)\big] \\ &= (z_2-z_1)(z_3-z_1)(z_3-z_2)(z_1+z_2+z_3),\end{aligned}$$

这是因为 $z_3^2 - z_2^2 + z_1(z_3 - z_2) = (z_3 - z_2)(z_3 + z_2 + z_1)$。首三个因子均不为零，所以行列式为零当且仅当 $z_1 + z_2 + z_3 = 0$。换言之，两个三角形相似，当且仅当形心（centroid）$\frac{1}{3}(z_1 + z_2 + z_3)$ 位于原点。

第 (iii) 部分

窍门是把三角形与重新标记顶点后的自己作比较。取 Q_1, Q_2, Q_3 为 P_2, P_3, P_1，即

$$w_1 = z_2, \qquad w_2 = z_3, \qquad w_3 = z_1.$$

> **想一想** “等边”为甚么会与这样的重新标记有关?

若在 $P_1 \leftrightarrow P_2$、$P_2 \leftrightarrow P_3$、$P_3 \leftrightarrow P_1$ 的对应下相似，则各组对应边必有相同的比 $k > 0$：

$$\frac{P_2P_3}{P_1P_2} = \frac{P_3P_1}{P_2P_3} = \frac{P_1P_2}{P_3P_1} = k.$$

把三者相乘得 $k^3 = 1$，所以 $k = 1$，三边全部相等。反之，若三角形是等边（equilateral）三角形，则绕其中心旋转 $120°$（取适当的方向）会把 $P_1 \to P_2 \to P_3 \to P_1$ 依次对应过去。旋转保持形状和绕向不变，所以该三角形与重新标记后的自己同向相似。因此

$$P_1P_2P_3 \text{ 为等边三角形} \iff \text{它同向相似于} P_2P_3P_1.$$

根据第一部分，右边成立当且仅当

$$0 = \det\begin{pmatrix} z_1 & z_2 & z_3 \\ z_2 & z_3 & z_1 \\ 1 & 1 & 1 \end{pmatrix} = z_1(z_3 - z_1) - z_2(z_2 - z_1) + z_3(z_2 - z_3),$$

即

$$z_1z_3 - z_1^2 - z_2^2 + z_1z_2 + z_2z_3 - z_3^2 = 0 \iff z_2z_3 + z_3z_1 + z_1z_2 = z_1^2 + z_2^2 + z_3^2,$$

这正是所要证明的结果。

验算

- 第 (ii) 部分：$z_1, z_2, z_3 = 1, \omega, \omega^2$（其中 $\omega = \mathrm{e}^{2\pi\mathrm{i}/3}$）的形心为 0，但它们的立方全都等于 1，并非互不相同。改用 $z = 1,\ \mathrm{i},\ -1-\mathrm{i}$。形心为 0，而立方分别是 $1,\ -\mathrm{i},\ 2-2\mathrm{i}$。按 P_1P_2, P_2P_3, P_3P_1 的次序，两个三角形的边长都是 $\sqrt{2}, \sqrt{5}, \sqrt{5}$（在这个例子中它们甚至全等（congruent）），而且两者都沿逆时针方向排列。
- 第 (iii) 部分：对于 $z = 0, 1, \mathrm{e}^{\mathrm{i}\pi/3}$，两边都等于 $\mathrm{e}^{\mathrm{i}\pi/3}$，因为 $1 + \mathrm{e}^{2\mathrm{i}\pi/3} = \mathrm{e}^{\mathrm{i}\pi/3}$。
- 以电脑对 10^4 个随机三角形作测试，结果证实 $w_i = az_i + b$ 总是满足 $(*)$，而经反射的三角形 $w_i = a\bar{z}_i + b$ 则从不满足。测试亦证实了 $(\dagger)$ 的两个因式分解，以及等边三角形的判别准则。至于题目所印的条件（分母为 $z_2 - z_3$），在所有真正相似的情况中无一成立。

常见错误

- 忽略三角形的绕向。单看模，只能得出“容许反射的相似”。互为镜像的三角形，其辐角的正负号相反，所以并不满足 $(*)$。
- 在第 (iii) 部分忘记证明逆命题（converse）：“与重新标记后的自己相似”这个论证必须双向都成立，而 $k^3 = 1 \Rightarrow k = 1$ 这一步用到了 k 是正实数。

STEP 之外

映射 $z \mapsto az + b$ 就是平面上保持定向（orientation-preserving）的相似变换。它们是最简单的莫比乌斯变换（Möbius transformations），而四个点的交比（cross-ratio）则是 $(*)$ 的推广。等边条件亦可因式分解为 $(z_1 + \omega z_2 + \omega^2 z_3)(z_1 + \omega^2 z_2 + \omega z_3) = 0$，这表示 z_1, z_2, z_3 沿其中一个绕向围成一个等边三角形。Needham 的 *Visual Complex Analysis* 把这种以几何眼光看待复数的思考方式发挥得十分精彩。

详细解答

STEP III 1991，第 3 题

纯数学 · 分段函数与反函数

STEP III 1991，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/91-s3-q3

关键思路

绝对值（modulus）不过是一个函数，其内部的式子变号时，它的表达式就跟著改变。这里三个绝对值只在 $x=-1$、0 和 1 处变号，因此在它们把 $x<2$ 分割出来的四个区间上，函数各自都是一个普通的二次函数（quadratic）。一旦求得这四个二次函数，每一个问题（转向点、拐点、连续性、光滑性、反函数）都只是关于简单抛物线及其如何相接的问题。

解答

去掉绝对值

把绝对值内的式子因式分解：$x^2-x=x(x-1)$、$x^2-1=(x-1)(x+1)$ 和 $x^2+x=x(x+1)$。它们的正负号如下：

区间	x^2-x	x^2-1	x^2+x	$\mathrm{f}(x)$
$x\le -1$	$+$	$+$	$+$	x^2-4x-1
$-1\le x\le 0$	$+$	$-$	$-$	$3x^2+1$
$0\le x\le 1$	$-$	$-$	$+$	$1-5x^2$
$1\le x<2$	$+$	$+$	$+$	x^2-4x-1

例如，在 $-1\le x\le 0$ 上，

$$\mathrm{f}(x)=2(x^2-x)+(1-x^2)+2(x^2+x)=3x^2+1,$$

而在 $0\le x\le 1$ 上，$\mathrm{f}(x)=-2(x^2-x)+(1-x^2)-2(x^2+x)=1-5x^2$。在外侧的两个区间上，三个绝对值内的式子全为正，而 $2(x^2-x)+(x^2-1)-2(x^2+x)=x^2-4x-1$。在端点处，其中一个绝对值内的式子为零，所以两边的公式在该处都正确。

连续性与可导性

f 由 $|\cdot|$ 和多项式组合而成，这些全都是连续的，所以 f 在 $x<2$ 上**处处连续（continuous）**。作为验算，相邻的公式在接驳处吻合：$f(-1)=4$、$f(0)=1$、$f(1)=-4$。

至于可导性（differentiability），可比较接驳处的单侧导数（one-sided derivative）。四段的导数分别为 $2x-4$、$6x$、$-10x$ 和 $2x-4$。

- 在 $x=-1$：左侧为 $2(-1)-4=-6$，右侧为 $6(-1)=-6$。两者相同，所以 $f'(-1)=-6$。
- 在 $x=0$：两侧都得 0，所以 $f'(0)=0$。
- 在 $x=1$：左侧为 -10，右侧为 $2-4=-2$。两者不同，所以图像在 $(1,-4)$ 处有一个**尖角（corner）**。

因此 f **并非处处可导**：它恰好只在 $x=1$ 处不可导。

转向点与拐点

在每一段上都有 $f'<0$，只有 $x=0$ 这一点例外：

- $x<-1$ 和 $1<x<2$：由于 $x<2$，$f'(x)=2x-4<0$；
- $-1<x<0$：$f'(x)=6x<0$；
- $0<x<1$：$f'(x)=-10x<0$。

由于 f 连续，而且在每一段上都严格递减，所以它**在整个 $x<2$ 上严格递减（strictly decreasing）**。特别地，它**没有极大点或极小点**，连局部的也没有。

> **想一想** $f'(0)=0$，那么 $(0,1)$ 不就是转向点（turning point）吗？

不是。左边一段 $3x^2+1$ 在 0 处取得极小值，右边一段 $1-5x^2$ 则在该处取得极大值，但在各自公式适用的区间上，两者都是递减的。所以 f 经过 $x=0$ 时仍然在下降，只是切线是水平的。

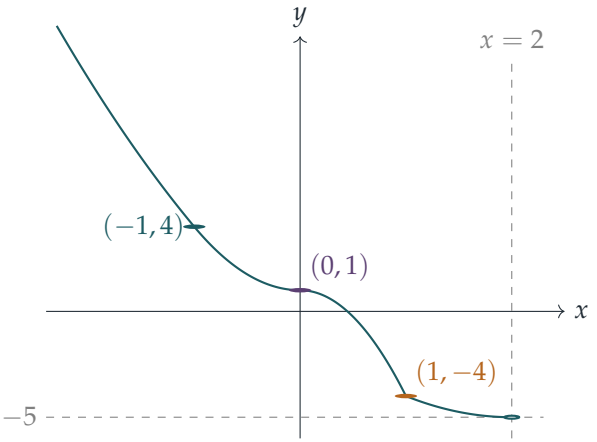
现在看凹凸性（concavity）。各段的二阶导数为

$$f''=2\ (x<-1),\qquad 6\ (-1<x<0),\qquad -10\ (0<x<1),\qquad 2\ (1<x<2).$$

- 在 $x=0$ 处，图像由下凸（convex）变为上凸（concave），而且在该处有切线，所以 $(0,1)$ 是一个**驻拐点（stationary point of inflection）**。
- 在 $x=-1$ 处，f'' 由 2 跳到 6，但仍然为正，所以没有拐点。
- 在 $x=1$ 处，曲线由上凸变回下凸，但该处是尖角，没有切线。按照通常的定义（切线穿过曲线），$(1,-4)$ 不是拐点。作答时值得明确指出这一点。

草图

当 $x \to -\infty$ 时，$\mathrm{f}(x) \to +\infty$。当 $x \to 2^-$ 时，$\mathrm{f}(x) \to 4 - 8 - 1 = -5$，这个值并不会取到。



曲线一直下降。它在 $(0,1)$ 处变平，形成一个水平拐点，并在 $(1,-4)$ 处有一个尖角，斜率在该处由 -10 跳到 -2。

反函数

由于 f 在 $x < 2$ 上连续且严格递减，而且在左端 $\mathrm{f} \to \infty$，在右端 $\mathrm{f} \to -5$，所以它是由 $(-\infty, 2)$ 到 $(-5, \infty)$ 的双射（bijection）。因此 f^{-1} 定义在 $x > -5$ 上。我们利用 $y = \mathrm{f}(x)$，在每一段所覆盖的值域上分别求反函数（inverse）。

- $x \le -1$，值域 $y \ge 4$：$y = x^2 - 4x - 1 = (x-2)^2 - 5$，所以 $(x-2)^2 = y + 5$。这里 $x - 2 < 0$，所以取负根：$x = 2 - \sqrt{y+5}$。
- $-1 \le x \le 0$，值域 $1 \le y \le 4$：$x^2 = (y-1)/3$，而 $x \le 0$，所以 $x = -\sqrt{(y-1)/3}$。
- $0 \le x \le 1$，值域 $-4 \le y \le 1$：$x^2 = (1-y)/5$，而 $x \ge 0$，所以 $x = \sqrt{(1-y)/5}$。
- $1 \le x < 2$，值域 $-5 < y \le -4$：同样有 $(x-2)^2 = y + 5$ 且 $x - 2 < 0$，所以 $x = 2 - \sqrt{y+5}$。

把变量改写为 x：

$$\mathrm{f}^{-1}(x) = \begin{cases} 2 - \sqrt{x+5}, & -5 < x \le -4, \\ \sqrt{\frac{1-x}{5}}, & -4 \le x \le 1, \\ -\sqrt{\frac{x-1}{3}}, & 1 \le x \le 4, \\ 2 - \sqrt{x+5}, & x \ge 4. \end{cases}$$

各公式在共用的端点处吻合。例如，在 $x = 4$ 处两者都得 -1，在 $x = -4$ 处两者都得 1。

验算

- 抽查数值：$\mathrm{f}(\tfrac{1}{2}) = 2 \cdot \tfrac{1}{4} + \tfrac{3}{4} - 2 \cdot \tfrac{3}{4} = -\tfrac{1}{4}$，而 $1 - 5 \cdot \tfrac{1}{4} = -\tfrac{1}{4}$。此外 $\mathrm{f}^{-1}(-\tfrac{1}{4}) = \sqrt{1.25/5} = \tfrac{1}{2}$。
- 直接计算，$\mathrm{f}(-3) = 2 \cdot 12 + 8 - 2 \cdot 6 = 20$。公式给出 $9 + 12 - 1 = 20$，而 $\mathrm{f}^{-1}(20) = 2 - \sqrt{25} = -3$。
- 数值上，对 10^5 个随机选取的 $x < 2$ 都有 $\mathrm{f}^{-1}(\mathrm{f}(x)) = x$；函数在一个细密网格的每一点上都在递减；而 $x = 1$ 处的单侧斜率算出来是 -10 和 -2。

常见错误

- 因为 $\mathrm{f}'(0) = 0$ 就把 $(0, 1)$ 称为极大点或极小点。如果某条抛物线本身的转向点位于该公式适用区间的边缘，那个转向点就无关重要。
- 求反函数时选错平方根。应由 x（或 $x - 2$）在每个区间上的正负号来决定。
- 以为两个公式的接驳处必定是不可导点。在 $x = -1$ 和 $x = 0$ 处，两边的斜率是吻合的。

STEP 之外

分段定义的函数（piecewise-defined function）可以连续而不可导，也可以可导而二阶导数不连续（这里 f'' 在 $x = -1$ 和 $x = 0$ 处跳跃）。反函数定理（inverse function theorem）指出，严格单调（strictly monotonic）的连续函数有连续的反函数，而且在 $\mathrm{f}' \neq 0$ 的地方反函数可导。留意 f^{-1} 在 $x = 1$ 处有一条垂直切线，恰好对应 $\mathrm{f}'(0) = 0$ 之处。Abbott 的 *Understanding Analysis*（数学分析入门读本）对这些概念有细致的阐述。

详细解答

STEP III 1991，第 4 题

纯数学 · 三维空间中的直线与平面

STEP III 1991，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/91-s3-q4

关键思路

同一位观察者 O_1 的两条不同视线（line of sight）都经过 O_1，所以它们张成一个平面（plane），而 P 的路径与这两条视线都相交。一条直线若与经过 O_1 的两条直线相交，而本身又不经过 O_1，就必定位于这两条直线所在的平面上。对 O_2 也是同样道理，所以路径就是两个平面的交线。这样一来，一个关于四条看似异面（skew）的直线的问题，就化为一个叉积（cross product）和一组联立方程（simultaneous equations）。

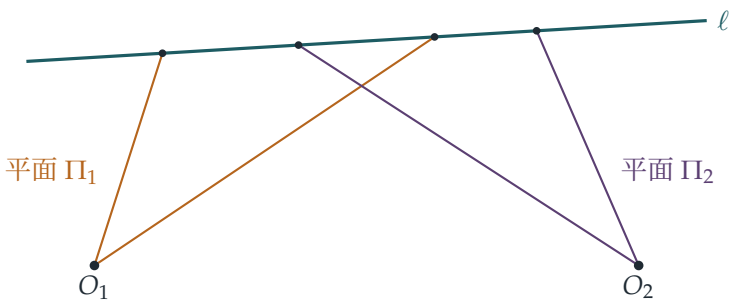
解答

设 P 沿直线 ℓ 运动。四条视线的方向为

$$\mathbf{d}_1 = (2,0,3), \qquad \mathbf{d}_1' = (0,3,4) \qquad （经过O_1），$$

$$\mathbf{d}_2 = (-3,1,3), \qquad \mathbf{d}_2' = (-4,2,5) \qquad （经过O_2）.$$

每条视线都连接一位观察者与 P 的某个位置，所以 ℓ 与这四条视线全部相交。



（仅为示意图：两条琥珀色视线与 ℓ 位于同一平面 Π_1 上，两条紫红色视线与 ℓ 则位于另一平面 Π_2 上。）

为何 ℓ 位于经过每位观察者的一个平面上

首先，ℓ 不可能经过 O_1。假如它经过，那么每当 $P \neq O_1$ 时，从 O_1 出发的视线就是 ℓ 本身，于是从 O_1 出发不可能有两条不同的视线。同理，ℓ 也不经过 O_2。

从 O_1 出发的两条视线是经过 O_1 的不同直线，所以它们只在 O_1 相交，并张成一个平面 Π_1。直线 ℓ 与它们交于两点，而这两点互不相同，因为两条视线唯一的公共点是 O_1，而 ℓ 并不经过 O_1。这两点都在 Π_1 上，所以整条 ℓ 都在 Π_1 上。同样地，ℓ 位于平面 Π_2 上，这个平面经过 O_2，并包含 $\mathbf{d}_2$ 和 $\mathbf{d}_2'$。

> ? **想一想** 难道 ℓ 不可以只是与每条视线各交于一点，而不位于任何特定的平面上吗?

若只有一条视线，的确可以。但这里有两个交点落在经过 O_1 的直线上，而两个相异的点就确定了 ℓ。这两点都在 Π_1 上，所以 ℓ 别无选择：它就是经过这两点的直线，而这条直线位于 Π_1 上。

平面 Π_1

Π_1 的一个法向量（normal）是

$$\mathbf{d}_1 \times \mathbf{d}_1' = (2,0,3) \times (0,3,4) = (0-9,\ 0-8,\ 6-0) = (-9,-8,6),$$

而 Π_1 经过原点，所以

$$\Pi_1 : \quad 9x + 8y - 6z = 0.$$

平面 Π_2

Π_2 的一个法向量是

$$\mathbf{d}_2 \times \mathbf{d}_2' = (-3,1,3) \times (-4,2,5) = (5-6,\ -12+15,\ -6+4) = (-1,3,-2),$$

而 Π_2 经过 $O_2(8a,0,0)$，所以 $-(x-8a) + 3y - 2z = 0$，即

$$\Pi_2 : \quad x - 3y + 2z = 8a.$$

（快速验算：$(-3,1,3)\cdot(-1,3,-2) = 3+3-6 = 0$ 及 $(-4,2,5)\cdot(-1,3,-2) = 4+6-10 = 0$。）

我们假设 $a \neq 0$，否则两位观察者会重合。于是 O_2 不在 Π_1 上（因为 $9\cdot 8a \neq 0$），而 O_1 也不在 Π_2 上，这与上文所说的一致。

路径

法向量 $(-9,-8,6)$ 与 $(-1,3,-2)$ 并不平行，所以 Π_1 与 Π_2 相交于唯一一条直线，而这条直线必定是 ℓ。它的方向与两个法向量都垂直：

$$(-9,-8,6)\times(-1,3,-2)=(16-18,\ -6-18,\ -27-8)=(-2,-24,-35),$$

所以我们可以取方向为 $(2,24,35)$。要找直线上的一点，可设 $y=0$。两条方程变成 $9x=6z$ 和 $x+2z=8a$，所以 $x=\frac{2}{3}z$ 及 $\frac{8}{3}z=8a$。由此得 $z=3a$ 和 $x=2a$。所以 $(2a,0,3a)$ 在 ℓ 上。它也在从 O_1 出发的第一条视线上，正如所料。

因此 P 的路径为

$$\boxed{\frac{x-2a}{2}=\frac{y}{24}=\frac{z-3a}{35}}$$

或等价地，由 $9x+8y-6z=0$ 和 $x-3y+2z=8a$ 这一对方程给出。

最后，ℓ 确实与全部四条视线相交，而且交点都不是观察者所在的点。把 ℓ 写成 $(2a+2\lambda,\ 24\lambda,\ 3a+35\lambda)$，它与四条视线的交点依次为

$$(2a,0,3a),\qquad (0,-24a,-32a),\qquad \left(\tfrac{80a}{37},\tfrac{72a}{37},\tfrac{216a}{37}\right),\qquad \left(\tfrac{56a}{25},\tfrac{72a}{25},\tfrac{36a}{5}\right)$$

。

? **想一想** 当 $a>0$ 时，第二个点的 y 和 z 都是负数。这有问题吗?

题目给出的视线是整条直线，而 $(0,-24a,-32a)$ 满足 $x=0,\ y/3=z/4$。我们只用到这些直线，并没有用到观察者面向哪一方，所以答案成立。如果你把“视线”理解为从观察者出发的射线，这次观测就落在 O_1 的背后。那只是数据的一个古怪之处，并非方法有错。

验算

- 把点 $(2a,0,3a)$ 和方向 $(2,24,35)$ 代入两个平面方程：$9(2)+8(24)-6(35)=18+192-210=0$ 及 $2-72+70=0$，所以这个方向位于两个平面上。
- 取 $a=1$，点 $\left(\frac{80}{37},\frac{72}{37},\frac{216}{37}\right)$ 等于 $O_2+\frac{72}{37}(-3,1,3)$，而且 $\frac{80/37-2}{2}=\frac{3}{37}=\frac{72/37}{24}$。所以它既在视线上，也在路径上。
- 我们对 a 的若干个值，以精确的有理数运算验证了全部四个交点。

常见错误

- 试图求出各对视线的交点：不同观察者在不同时刻的视线，根本不一定相交。

- 忘记 Π_2 经过的是 O_2 而不是原点，结果得出 $x - 3y + 2z = 0$ 而不是 $= 8a$。
- 计算叉积时正负号出错。务必把法向量与两个方向向量分别再作点积验算。

STEP 之外

空间中与四条给定直线都相交的直线有多少条？若四条直线处于一般位置（general position），答案是两条（在复数范围内计算；在实数范围内则可以是零条）。这是枚举几何（enumerative geometry）的一个经典结果，而本题的四条直线比较特殊，因为它们两两一组地经过两位观察者。同一个“视线平面”的想法，也是三角测量（triangulation）和立体视觉中对极几何（epipolar geometry）的基础：两部相机借此还原三维空间中各点的位置。至于直线、平面与交集的线性代数，可阅读 Strang 的 *Introduction to Linear Algebra*（线性代数入门教材）。

详细解答

STEP III 1991，第 5 题

纯数学 · 极坐标曲线与常微分方程

STEP III 1991，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/91-s3-q5

关键思路

先把极坐标（polar coordinates）放在一旁：这条方程其实只是一条以 θ 为变量的二阶常系数线性常微分方程（linear second-order ODE with constant coefficients）。按惯常方法求解（余函数加特别积分），用 $\theta = \frac{\pi}{2}$ 处的条件定出常数，然后再回到几何。由 $r = f(\theta)$ 给出的曲线，若 r 起点和终点皆为 0，而中间一直保持正值，便是一个闭合环圈（closed loop），其面积为 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$。

解答

解微分方程

辅助方程（auxiliary equation）$m^2 + 4 = 0$ 的根为 $m = \pm 2\mathrm{i}$，所以余函数（complementary function）是

$$r = A\cos 2\theta + B\sin 2\theta.$$

至于特别积分（particular integral），$\sin 3\theta$ 并不在余函数之内，所以我们试用 $r = k\sin 3\theta$。于是 $r'' = -9k\sin 3\theta$，而

$$-9k\sin 3\theta + 4k\sin 3\theta = 5\sin 3\theta \quad\Longrightarrow\quad k = -1.$$

因此通解（general solution）为

$$r = A\cos 2\theta + B\sin 2\theta - \sin 3\theta, \qquad \frac{\mathrm{d}r}{\mathrm{d}\theta} = -2A\sin 2\theta + 2B\cos 2\theta - 3\cos 3\theta.$$

在 $\theta = \frac{\pi}{2}$ 处，有 $\cos 2\theta = -1$、$\sin 2\theta = 0$、$\sin 3\theta = -1$ 及 $\cos 3\theta = 0$。两个条件变成

$$r = -A + 1 = 1 \quad\Longrightarrow\quad A = 0, \qquad \frac{\mathrm{d}r}{\mathrm{d}\theta} = -2B = -2 \quad\Longrightarrow\quad B = 1.$$

故此

$$r = \sin 2\theta - \sin 3\theta.$$

曲线是一个闭合环圈

要看出 r 在何处为零、其正负号如何，可利用恒等式 $\sin P - \sin Q = 2\cos\frac{P+Q}{2}\sin\frac{P-Q}{2}$ 作因式分解：

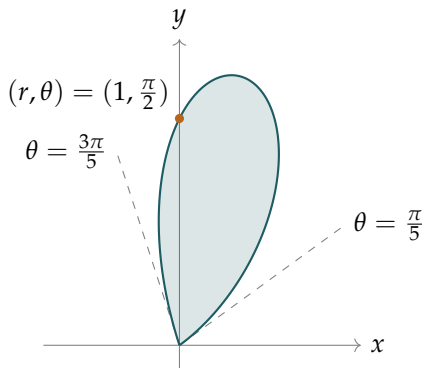
$$r = 2\cos\frac{5\theta}{2}\sin\left(-\frac{\theta}{2}\right) = -2\sin\frac{\theta}{2}\cos\frac{5\theta}{2}.$$

想一想 为甚么因式分解后的形式能解决“闭合环圈”这个问题?

因为在给定的范围内，两个因子的正负号都一目了然。当 $\frac{\pi}{5} \le \theta \le \frac{3\pi}{5}$ 时：

- $\frac{\theta}{2}$ 位于 $\left[\frac{\pi}{10}, \frac{3\pi}{10}\right]$ 之内，所以 $\sin\frac{\theta}{2} > 0$；
- $\frac{5\theta}{2}$ 位于 $\left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$ 之内，所以 $\cos\frac{5\theta}{2} \le 0$，等号只在两个端点成立。

所以在 $\theta = \frac{\pi}{5}$ 及 $\theta = \frac{3\pi}{5}$ 处 $r = 0$，而对严格介乎两者之间的每一个 θ 都有 $r > 0$。因此曲线沿射线 $\theta = \frac{\pi}{5}$ 离开原点，以正的 r 绕行一周，再沿射线 $\theta = \frac{3\pi}{5}$ 返回原点。其间每一条射线 $\theta =$ 常数 都与曲线恰好相交一次（就是对应该 $r > 0$ 的那一点），所以曲线不会自相交。它是单一个闭合环圈，围出区域 $0 \le r \le \sin 2\theta - \sin 3\theta$，$\frac{\pi}{5} \le \theta \le \frac{3\pi}{5}$。



面积

所围的面积为

$$\mathcal{A} = \frac{1}{2}\int_{\pi/5}^{3\pi/5} r^2\,\mathrm{d}\theta = \frac{1}{2}\int_{\pi/5}^{3\pi/5}\left(\sin^2 2\theta - 2\sin 2\theta\sin 3\theta + \sin^2 3\theta\right)\mathrm{d}\theta.$$

利用 $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$ 及 $2\sin P\sin Q = \cos(P-Q) - \cos(P+Q)$，

$$r^2 = 1 - \tfrac{1}{2}\cos 4\theta - \tfrac{1}{2}\cos 6\theta - \cos\theta + \cos 5\theta.$$

记 $D = \sin\frac{2\pi}{5} - \sin\frac{\pi}{5}$。我们在 $\left[\frac{\pi}{5}, \frac{3\pi}{5}\right]$ 上逐项积分，并利用 $\sin(x+2\pi) = \sin x$ 及 $\sin(\pi - x) = \sin x$ 把每个正弦值化为 $\pm\sin\frac{\pi}{5}$ 或 $\pm\sin\frac{2\pi}{5}$：

$$\begin{aligned}
\int 1\,\mathrm{d}\theta &= \tfrac{2\pi}{5},\\
\int -\tfrac{1}{2}\cos 4\theta\,\mathrm{d}\theta &= -\tfrac{1}{8}\left[\sin\tfrac{12\pi}{5} - \sin\tfrac{4\pi}{5}\right] = -\tfrac{1}{8}\left[\sin\tfrac{2\pi}{5} - \sin\tfrac{\pi}{5}\right] = -\tfrac{1}{8}D,\\
\int -\tfrac{1}{2}\cos 6\theta\,\mathrm{d}\theta &= -\tfrac{1}{12}\left[\sin\tfrac{18\pi}{5} - \sin\tfrac{6\pi}{5}\right] = -\tfrac{1}{12}\left[-\sin\tfrac{2\pi}{5} + \sin\tfrac{\pi}{5}\right] = \tfrac{1}{12}D,\\
\int -\cos\theta\,\mathrm{d}\theta &= -\left[\sin\tfrac{3\pi}{5} - \sin\tfrac{\pi}{5}\right] = -D,\\
\int \cos 5\theta\,\mathrm{d}\theta &= \tfrac{1}{5}\left[\sin 3\pi - \sin\pi\right] = 0.
\end{aligned}$$

（第三行中，$\sin\frac{18\pi}{5} = \sin\left(-\frac{2\pi}{5}\right) = -\sin\frac{2\pi}{5}$ 而 $\sin\frac{6\pi}{5} = -\sin\frac{\pi}{5}$。）相加得

$$\int_{\pi/5}^{3\pi/5} r^2\,\mathrm{d}\theta = \frac{2\pi}{5} + \left(-\frac{1}{8} + \frac{1}{12} - 1\right)D = \frac{2\pi}{5} - \frac{25}{24}D,$$

再除以二，

$$\mathcal{A} = \frac{\pi}{5} - \frac{25}{48}D = \frac{\pi}{5} + \frac{25}{48}\left[\sin\frac{\pi}{5} - \sin\frac{2\pi}{5}\right],$$

这正是所求的结果。

验算

- 在 $\theta = \frac{\pi}{2}$ 处：$r = \sin\pi - \sin\frac{3\pi}{2} = 1$ 而 $r' = 2\cos\pi - 3\cos\frac{3\pi}{2} = -2$，符合要求。
- 以数值计算，$\mathcal{A} \approx 0.6283 - \frac{25}{48}(0.3633) \approx 0.4391$。对 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 使用辛普森法则（Simpson's rule），以及对描出的环圈使用鞋带公式（shoelace formula），两者都得出 0.43911。由 $\theta = \frac{\pi}{2}$ 出发以数值方法逐步求解微分方程，可重现 $r = \sin 2\theta - \sin 3\theta$，并确认 r 在两端都回到 0。
- 答案合理：环圈最远达到 $r \approx 1.22$，所张的角为 $\frac{2\pi}{5}$。半径为 1.22、圆心角为该角的扇形，面积约为 0.93，而环圈占去其中略少于一半。

常见错误

- 弄错特别积分的正负号：$(-9+4)k = 5$ 给出 $k = -1$，而不是 $+1$。
- 忘记证明在两个端点之间 $r > 0$。单凭 r 有两个零点并不足以构成环圈：r 有可能在中间变号，那样 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 就不是所围的面积了。
- 化简 $\sin\frac{12\pi}{5}$ 或 $\sin\frac{18\pi}{5}$ 时粗心大意：两者各为 $\pm\sin\frac{2\pi}{5}$，而正负号至关重要。

STEP 之外

这里的方法之所以行得通，是因为方程是线性的：每一个解都是一个特解加上两个独立解 $\cos 2\theta$ 和 $\sin 2\theta$ 的组合，而两个初始条件（initial conditions）恰好选出唯一的一个解。这就是线性常微分方程的存在性与唯一性定理（existence and uniqueness theorem），大学一年级的课程会证明它，并且经常用到。一个相关而又引人入胜的事实是：对于在平方反比力（inverse-square force）作用下的轨道，$u = 1/r$ 满足 $u'' + u =$ 常数，牛顿（Newton）定律正是这样推出圆锥曲线（conic section）轨道的。Boyce 与 DiPrima 的 *Elementary Differential Equations*（常微分方程入门教材）对线性理论有详尽的论述。

详细解答

STEP III 1991，第 6 题

纯数学 · 双曲线性变换

STEP III 1991，第 6 题

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

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

本题在 *Beyond STEP* 第 7.6 节双曲旋转与相对论中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch07/hyperbolic-rotations.html

关键思路

这个矩阵（matrix）的表现与旋转矩阵（rotation matrix）相似，只是以双曲函数（hyperbolic functions）取代了圆函数，而恒等式 $\cosh^2 u - \sinh^2 u = 1$ 以及 cosh 和 sinh 的加法公式（addition formulae）就足以完成全部工作。把斜率（gradient）为 $\tanh v$ 的方向写成向量 $(\cosh v, \sinh v)$，T 对斜率的作用便一目了然：它只是把 u 加到“双曲角”v 上。至于夹角的条件，我们把两条直线夹角的正切用 v_1 和 v_2 表示，再看看甚么必须保持不变。

解答

全篇记 $c = \cosh u$ 及 $s = \sinh u$，于是 $c^2 - s^2 = 1$，而且因为 $u > 0$，所以 $s > 0$。这个变换（transformation）是

$$x' = cx + sy, \qquad y' = sx + cy.$$

双曲线

对任意一点，

$$\begin{aligned} x'^2 - y'^2 &= (cx + sy)^2 - (sx + cy)^2 \\ &= (c^2 - s^2)x^2 + 2csxy - 2csxy - (c^2 - s^2)y^2 \\ &= x^2 - y^2. \end{aligned}$$

所以 T 保持 x^2-y^2 不变，而 $x^2-y^2=1$ 上的每一点都被送到 $x^2-y^2=1$ 上的一点。整条曲线都被完全覆盖，而不仅是被映射到自身之内：把 u 换成 $-u$ 所得的矩阵是 T 的逆矩阵（inverse）（乘开并利用 $c^2-s^2=1$ 即可），而它同样保持 x^2-y^2 不变。所以双曲线（hyperbola）上的每一点都是双曲线上某一点的像（image）。这条曲线变换为自身。

（事实上每一支都变换为自身：若 $x\geq 1$，则 $|y|<x$，所以 $x'=cx+sy>(c-s)x=\mathrm{e}^{-u}x>0$。）

通过原点的不变直线

由于 T 是线性（linear）的，它把原点送到原点，并把通过原点、方向为 $\mathbf{d}$ 的直线送到通过原点、方向为 $T\mathbf{d}$ 的直线。所以我们需要 $T\mathbf{d}$ 平行于 $\mathbf{d}$，即 $\mathbf{d}$ 是特征向量（eigenvector）。试取 $\mathbf{d}=(1,m)$：

$$T\begin{pmatrix}1\\m\end{pmatrix}=\begin{pmatrix}c+sm\\s+cm\end{pmatrix},$$

它平行于 $(1,m)$ 当且仅当 $s+cm=m(c+sm)$，即 $s(1-m^2)=0$。由于 $s\neq 0$，故 $m=\pm 1$。（铅垂方向 $(0,1)$ 被送到 (s,c)，后者并非铅垂，所以不会得出新的直线。）这两条直线是

$$\boxed{y=x\quad\text{及}\quad y=-x}$$

而事实上 $T(1,1)=\mathrm{e}^{u}(1,1)$ 及 $T(1,-1)=\mathrm{e}^{-u}(1,-1)$。它们正是双曲线的渐近线（asymptotes），这与双曲线映射到自身的结果吻合。

斜率：$v'=v+u$

取一条通过点 $\mathbf{p}$、方向为 $\mathbf{d}$ 的直线，其上的点为 $\mathbf{p}+t\mathbf{d}$。由线性可知它的像是 $T\mathbf{p}+tT\mathbf{d}$：一条通过 $T\mathbf{p}$、方向为 $T\mathbf{d}$ 的直线。所以直线的位置无关重要，只有方向才有影响。

斜率为 $\tanh v$ 的直线，方向为 $(1,\tanh v)$，或者乘以 $\cosh v>0$ 后，方向为 $(\cosh v,\sinh v)$。于是由加法公式，

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

因此像直线的斜率为 $\tanh(v+u)$。由于 $\tanh$ 是一对一（one-to-one）的，$\tanh v'=\tanh(v+u)$ 给出

$$v'=v+u,$$

这正是所要证明的。（只有斜率严格介乎 -1 与 1 之间的直线才可以这样表示，这就是题目采用 $\tanh v$ 的原因。）

相等的夹角

若两条直线的斜率为 m_1 和 m_2，且 $1+m_1m_2\neq 0$，则它们之间的锐角 θ 满足

$$\tan\theta=\left|\frac{m_1-m_2}{1+m_1m_2}\right|.$$

这里 $|m_1|,|m_2|<1$，所以 $1+m_1m_2>0$：两条直线永不互相垂直，$\theta\in[0,\frac{\pi}{2})$，而且 θ 由 $\tan\theta$ 唯一决定。

> **想一想** 这条公式看来很像 $\tanh(v_1-v_2)$，但分母的正负号不对。那么它实际上化简成甚么?

代入 $m_i=\tanh v_i$，并把分子和分母同乘以 $\cosh v_1\cosh v_2>0$：

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

变换后的直线有 $v_i'=v_i+u$。它们的差不变，但它们的和增加了 $2u$，所以它们之间的夹角 θ' 满足

$$\tan\theta'=\frac{|\sinh(v_1-v_2)|}{\cosh(v_1+v_2+2u)}.$$

要 ℓ_1 和 ℓ_2 相交，必须有 $v_1\neq v_2$，所以分子不是零。因此

$$\theta=\theta'\iff\cosh(v_1+v_2)=\cosh(v_1+v_2+2u).$$

由于 $\cosh a=\cosh b$ 当且仅当 $a=\pm b$，这表示要么 $v_1+v_2=v_1+v_2+2u$，但因为 $u>0$，这不可能；要么 $v_1+v_2=-(v_1+v_2+2u)$。所以充分必要条件（necessary and sufficient condition）是

$$\boxed{v_1+v_2=-u}$$

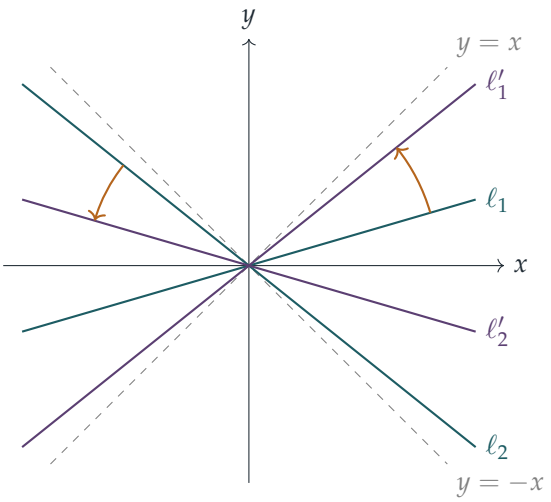
或等价地 $v_1'+v_2'=u$。

两条直线相交于原点时的图像

当 $v_1+v_2=-u$ 时，我们有

$$v_1'=v_1+u=-v_2,\qquad v_2'=v_2+u=-v_1.$$

所以 ℓ_1' 的斜率为 $-\tanh v_2$：它是 ℓ_2 对 x 轴的反射（reflection）。同样，ℓ_2' 是 ℓ_1 的反射。若两条直线通过原点，变换后的一对直线就只是原来那一对对 x 轴的镜像，这就是它们之间的夹角相同的原因。图中采用 $u=0.8$、$v_1=0.3$、$v_2=-1.1$：ℓ_1,ℓ_2 的斜率约为 0.29 和 -0.80，而 ℓ_1',ℓ_2' 的斜率约为 0.80 和 -0.29。每条直线都转向 $y=x$，即被 T 拉伸的那条不变直线。



蓝绿色的直线是 ℓ_1, ℓ_2，紫红色的直线是它们的像；琥珀色的箭头显示 $\ell_1 \to \ell_1'$ 和 $\ell_2 \to \ell_2'$。紫红色的一对是蓝绿色的一对对 x 轴的反射。

验算

- 取 $u = 0.8$、$v_1 = 0.3$、$v_2 = -1.1$，ℓ_1 与 ℓ_2 之间的夹角以及它们的像之间的夹角，数值上都是 $54.92°$。把 v_2 稍作改动，两个夹角便不再相等。
- 当 $u \to 0$ 时，T 变成恒等变换（identity），条件变成 $v_1 + v_2 = 0$：两条直线对 x 轴对称，而恒等映射显然会保持这种对称。
- $T(1,1) = \mathrm{e}^u(1,1)$ 很容易用手算验证，因为 $\cosh u + \sinh u = \mathrm{e}^u$。

常见错误

- 只证明了双曲线映射到自身之内；要用逆变换才能说明没有遗漏任何一点。
- 把夹角公式写成 $\tanh(v_1 - v_2)$；$1 + m_1 m_2$ 中的 $+$ 号使分母变成 $\cosh(v_1 + v_2)$，而在 T 之下改变的正是这个和。
- 忘记了 $\cosh a = \cosh b$ 的根 $v_1 + v_2 = v_1 + v_2 + 2u$，又没有说明它为何不可能。

STEP 之外

这个矩阵是狭义相对论（special relativity）中的洛伦兹变换（Lorentz boost），其中 x 和 y 分别扮演时间和空间的角色（采用光速为 1 的单位）。$x^2 - y^2$ 这个量是不变间隔（invariant

interval），$\tanh v$ 是速度，而 v 是它的快度（rapidity）。$v' = v + u$ 这个结果说明快度可以相加，这其实就是相对论速度加法公式的另一种面貌。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）就是以这种方式建立狭义相对论的。至于以特征向量的观点看不变直线，可参阅 Strang 的 *Introduction to Linear Algebra*（线性代数入门）。

详细解答

STEP III 1991，第 7 题

纯数学 · 利用对称性积分

> **STEP III 1991，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/91-s3-q7

关键思路

$\ln(\sin x)$ 的积分无法通过原函数（antiderivative）求出，但对称性（symmetry）足以解决一切。代换（substitution）$x \mapsto \frac{1}{2}\pi - x$ 会把 sin 与 cos 互换，而 $x \mapsto \pi - x$ 则说明 $\sin x$ 在 $[0, \pi]$ 上的取值，正是它在 $[0, \frac{1}{2}\pi]$ 上的取值重复两次。再写出 $\sin x \cos x = \frac{1}{2}\sin 2x$，便得到一条未知积分必须满足的方程。第 (iii) 部分是一次分部积分（integration by parts），把 $x \cot x$ 化为 $\ln(\sin x)$，而第 (ii) 部分则用来处理边界项（boundary term）。

解答

全题记

$$I = \int_0^{\frac{1}{2}\pi} \ln(\sin x)\,\mathrm{d}x.$$

第 (i) 部分

正弦与余弦的积分相等。代入 $x = \frac{1}{2}\pi - t$，则 $\mathrm{d}x = -\mathrm{d}t$ 且 $\sin x = \cos t$。积分上下限 $0, \frac{1}{2}\pi$ 变成 $\frac{1}{2}\pi, 0$，而负号又把它们调换回来：

$$I = \int_{\frac{1}{2}\pi}^{0} \ln(\cos t)\,(-\mathrm{d}t) = \int_0^{\frac{1}{2}\pi} \ln(\cos t)\,\mathrm{d}t.$$

与 $\sin 2x$ 的关系。把这两个相等的积分相加，并利用 $\ln a + \ln b = \ln(ab)$（由于在开区间

上 $\sin x, \cos x > 0$，此式成立）：

$$\begin{aligned} 2I &= \int_0^{\frac{1}{2}\pi} \ln(\sin x \cos x)\,\mathrm{d}x = \int_0^{\frac{1}{2}\pi} \ln\left(\tfrac{1}{2}\sin 2x\right)\mathrm{d}x \\ &= \int_0^{\frac{1}{2}\pi} \ln(\sin 2x)\,\mathrm{d}x - \tfrac{1}{2}\pi \ln 2. \end{aligned}$$

两边除以二，得 $I = \frac{1}{2}\int_0^{\frac{1}{2}\pi} \ln(\sin 2x)\,\mathrm{d}x - \frac{1}{4}\pi \ln 2$，第一串等式至此证毕。

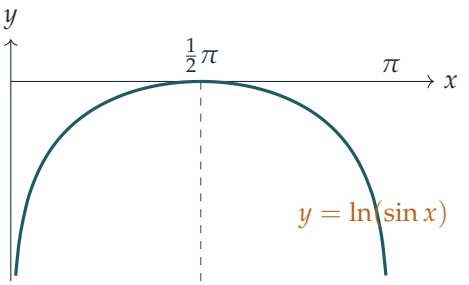
第二串等式。设 $u = 2x$，则 $\mathrm{d}x = \frac{1}{2}\,\mathrm{d}u$，而积分上下限变成 0 和 π：

$$\int_0^{\frac{1}{2}\pi} \ln(\sin 2x)\,\mathrm{d}x = \tfrac{1}{2}\int_0^{\pi} \ln(\sin u)\,\mathrm{d}u.$$

现在把 $[0, \pi]$ 在 $\frac{1}{2}\pi$ 处分成两段。在后半段代入 $u = \pi - v$。由于 $\sin(\pi - v) = \sin v$，而积分上下限 $\frac{1}{2}\pi, \pi$ 变成 $\frac{1}{2}\pi, 0$，

$$\int_{\frac{1}{2}\pi}^{\pi} \ln(\sin u)\,\mathrm{d}u = \int_0^{\frac{1}{2}\pi} \ln(\sin v)\,\mathrm{d}v = I.$$

下图显示了这种对称性：曲线 $y = \ln(\sin x)$ 关于 $x = \frac{1}{2}\pi$ 对称。



所以 $\int_0^{\pi} \ln(\sin u)\,\mathrm{d}u = 2I$，因此

$$\int_0^{\frac{1}{2}\pi} \ln(\sin 2x)\,\mathrm{d}x = \tfrac{1}{2}\int_0^{\pi} \ln(\sin x)\,\mathrm{d}x = \tfrac{1}{2}(2I) = I,$$

正如所求。

计算 I。把第二个结果代入第一个结果，得

$$I = \tfrac{1}{2}I - \tfrac{1}{4}\pi \ln 2 \quad \Longrightarrow \quad I = \int_0^{\frac{1}{2}\pi} \ln(\sin x)\,\mathrm{d}x = -\tfrac{1}{2}\pi \ln 2.$$

这一步需要 I 是一个有限的数，这正是题目容许我们假设积分收敛（convergence）的原因。若没有这个假设，“$I = \frac{1}{2}I + c$”甚么也证明不了。

第 (ii) 部分

设 $x \geq 1$，并取 $u = \sqrt{x}$。则 $u \geq 1$，所以题目给出的不等式适用：

$$\ln \sqrt{x} < \sqrt{x}, \qquad \text{即} \qquad \tfrac{1}{2} \ln x < \sqrt{x}.$$

第一个极限（limit）。当 $x \geq 1$ 时，我们有 $\ln x \geq 0$，而把 $\ln x < 2\sqrt{x}$ 除以 $x > 0$ 可得

$$0 \leq \frac{\ln x}{x} < \frac{2}{\sqrt{x}}.$$

当 $x \to \infty$ 时，右边趋向 0，所以根据夹逼原理（squeeze principle，又称 sandwich principle），$\dfrac{\ln x}{x} \to 0$。

第二个极限。当 $0 < x < 1$ 时，设 $x = 1/t$，则 $t > 1$，且当 $x \to 0^+$ 时 $t \to \infty$。于是根据第一个极限，

$$x \ln x = \frac{1}{t} \ln \frac{1}{t} = -\frac{\ln t}{t} \to 0$$

。所以当 $x \to 0$（只取正值）时，$x \ln x \to 0$。

第 (iii) 部分

首先留意被积函数（integrand）在 0 处并无问题：$x \cot x = \dfrac{x}{\sin x} \cos x \to 1$。由于在 $(0, \frac{1}{2}\pi)$ 上 $\cot x$ 是 $\ln(\sin x)$ 的导数，我们取 $u = x$ 和 $\mathrm{d}v = \cot x \, \mathrm{d}x$，$v = \ln(\sin x)$ 作分部积分。对数在 0 处趋向无穷，所以我们先在 $[\varepsilon, \frac{1}{2}\pi]$ 上计算，然后令 $\varepsilon \to 0^+$：

$$\int_\varepsilon^{\frac{1}{2}\pi} x \cot x \, \mathrm{d}x = \Big[x \ln(\sin x) \Big]_\varepsilon^{\frac{1}{2}\pi} - \int_\varepsilon^{\frac{1}{2}\pi} \ln(\sin x) \, \mathrm{d}x.$$

在上限处，$\ln(\sin \frac{1}{2}\pi) = \ln 1 = 0$。所以边界项只剩下 $-\varepsilon \ln(\sin \varepsilon)$。

> **想一想** 为甚么 $\varepsilon \ln(\sin \varepsilon)$ 趋向 0？第 (ii) 部分讨论的是 $x \ln x$，而不是 $x \ln(\sin x)$。

记 $s = \sin \varepsilon$，当 $\varepsilon \to 0^+$ 时，它取正值而趋向 0。于是

$$\varepsilon \ln(\sin \varepsilon) = \frac{\varepsilon}{\sin \varepsilon} \cdot s \ln s.$$

第一个因子趋向 1（即标准极限 $\sin \varepsilon / \varepsilon \to 1$），而根据第 (ii) 部分，$s \ln s \to 0$。所以边界项趋向 0。

令 $\varepsilon \to 0^+$，并利用第 (i) 部分：

$$\int_0^{\frac{1}{2}\pi} x \cot x \, \mathrm{d}x = 0 - I = \tfrac{1}{2}\pi \ln 2.$$

验算

- 正负号：在该区间上 $\ln(\sin x) \le 0$，所以 I 必为负数，而 $-\frac{1}{2}\pi \ln 2 \approx -1.0888$ 的确是负数。另一方面，在 $(0, \frac{1}{2}\pi)$ 上 $x \cot x > 0$，而我们的答案是正数。
- 大小：$x \cot x$ 在长度为 $\frac{1}{2}\pi \approx 1.571$ 的区间上由 1 下降至 0，所以积分值约为 1.09 是合理的。
- 数值上，以中点求积法（midpoint quadrature）计算可得 $\int_0^{\frac{1}{2}\pi} x \cot x \, \mathrm{d}x \approx 1.0887930$，与 $\frac{1}{2}\pi \ln 2 = 1.0887930\ldots$ 吻合，并且 $\int_0^{\frac{1}{2}\pi} \ln(\sin x) \, \mathrm{d}x \approx -1.08879$。

常见错误

- 转换为 $u = 2x$ 时，漏掉由 $\mathrm{d}x = \frac{1}{2}\,\mathrm{d}u$ 而来的因子 $\frac{1}{2}$，结果得出 $\int_0^{\frac{1}{2}\pi} \ln(\sin 2x) \, \mathrm{d}x = 2I$，之后便无从入手。
- 在第 (iii) 部分，直接写 $[x \ln(\sin x)]_0^{\frac{1}{2}\pi} = 0$，却没有论证在 0 处的极限。第 (ii) 部分正是为了求这个极限而设。

STEP 之外

这里的两个积分都是*瑕积分*（improper integral）：它们的被积函数，或所用的原函数，在某个端点处无界。这类积分是以极限来定义的，正如我们利用 ε 所做的那样。大学第一门分析（analysis）课程会把这一点严格化，并给出判别这类极限何时存在的方法。题目中“可假设”那一句的背后，正是这些判别法。同样的对称技巧还可以求出许多其他积分，例如 $\int_0^{\pi/2} \ln(\tan x) \, \mathrm{d}x = 0$。Abbott 的 *Understanding Analysis* 讲述了支撑这一切的极限与积分理论。

详细解答

STEP III 1991，第 8 题

纯数学 · 归约公式与求和

STEP III 1991，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/91-s3-q8

关键思路

第 (i) 部分是一条归约公式（reduction formula），所以分部积分法（integration by parts）自然是第一步。余下的被积函数（integrand）含有 $\sin x \sin kx$，而 $\cos(k-1)x$ 的复角公式（compound-angle formula）可把它变回同一族的积分。第 (ii) 部分是 $(1+\mathrm{e}^{2\mathrm{i}\theta})^m$ 的二项展开式（binomial expansion）的实部（real part）。第 (iii) 部分则是把 (ii) 积分，再以裂项相消（telescoping）的方法把 (i) 的递推关系（recurrence）逐层展开而得出。

解答

第 (i) 部分

取 $k \geq 1$，并以 $u = \cos^k x$ 和 $\mathrm{d}v = \cos kx\,\mathrm{d}x$ 作分部积分，于是 $v = \frac{1}{k}\sin kx$ 及 $\mathrm{d}u = -k\cos^{k-1}x \sin x\,\mathrm{d}x$：

$$\begin{aligned} I_k &= \left[\tfrac{1}{k}\cos^k x \sin kx\right]_0^\theta + \int_0^\theta \cos^{k-1} x\, \sin x\, \sin kx\,\mathrm{d}x \\ &= \tfrac{1}{k}\cos^k\theta \sin k\theta + \int_0^\theta \cos^{k-1} x\, \sin x\, \sin kx\,\mathrm{d}x, \end{aligned}$$

因为方括号内的式子在 $x = 0$ 时等于零。

? **想一想** 怎样把 $\sin x \sin kx$ 改写成 x 的倍数的余弦?

把 $(k-1)x = kx - x$ 代入复角公式，得

$$\cos(k-1)x = \cos kx \cos x + \sin kx \sin x,$$

所以

$$\sin x \sin kx = \cos(k-1)x - \cos x \cos kx.$$

因此

$$\begin{aligned}\int_0^\theta \cos^{k-1} x \, \sin x \, \sin kx \, \mathrm{d}x &= \int_0^\theta \cos^{k-1} x \cos(k-1)x \, \mathrm{d}x \\ &\quad - \int_0^\theta \cos^k x \cos kx \, \mathrm{d}x \\ &= I_{k-1} - I_k.\end{aligned}$$

代回原式并乘以 k，

$$kI_k = \cos^k \theta \sin k\theta + kI_{k-1} - kI_k,$$

即

$$2kI_k = kI_{k-1} + \cos^k \theta \sin k\theta,$$

这正是所要证明的结果。

第 (ii) 部分

设 $z = \mathrm{e}^{2\mathrm{i}\theta}$。根据二项式定理（binomial theorem），

$$\begin{aligned}(1+z)^m &= \sum_{r=0}^{m} \binom{m}{r} z^r \\ &= \sum_{r=0}^{m} \binom{m}{r} \big(\cos 2r\theta + \mathrm{i} \sin 2r\theta\big).\end{aligned}$$

现在把“半角”抽出来：

$$1 + \mathrm{e}^{2\mathrm{i}\theta} = \mathrm{e}^{\mathrm{i}\theta}\big(\mathrm{e}^{-\mathrm{i}\theta} + \mathrm{e}^{\mathrm{i}\theta}\big) = 2\cos\theta \, \mathrm{e}^{\mathrm{i}\theta},$$

所以

$$(1+z)^m = 2^m \cos^m \theta \, \mathrm{e}^{\mathrm{i}m\theta} = 2^m \cos^m \theta \, (\cos m\theta + \mathrm{i} \sin m\theta).$$

由于 $2^m \cos^m \theta$ 是实数，比较两个表达式的实部，得

$$\sum_{r=0}^{m} \binom{m}{r} \cos 2r\theta = 2^m \cos^m \theta \cos m\theta,$$

这就是所求的恒等式（$r = 0$ 和 $r = m$ 两项分别是 1 和 $\cos 2m\theta$）。

第 (iii) 部分

把第一个表达式记作 S。(ii) 的右边是 I_m 的被积函数的 2^m 倍，所以我们把 (ii) 由 0 积分至 θ。$r = 0$ 一项积分后得 θ，而当 $r \geq 1$ 时，$\int_0^\theta \cos 2rx \, \mathrm{d}x = \frac{\sin 2r\theta}{2r}$。因此

$$\theta + \sum_{r=1}^{m} \binom{m}{r} \frac{\sin 2r\theta}{2r} = 2^m I_m, \qquad \text{所以} \qquad S = 2^m I_m - \theta.$$

现在利用 (i) 计算 $2^m I_m$。把递推关系除以 $2k$ 再乘以 2^k，对 $k \geq 1$ 得

$$2^k I_k = 2^{k-1} I_{k-1} + \frac{2^{k-1}}{k} \cos^k \theta \sin k\theta.$$

想一想 为甚么要乘以 2^k?

若写成 $I_k = \frac{1}{2} I_{k-1} + \ldots$，递推关系每一步都带有因子 $\frac{1}{2}$。改用 $J_k = 2^k I_k$ 便可把它消去：这样 $J_k - J_{k-1}$ 就是一个简单的显式项，而求和时会裂项相消。把 $k = 1, 2, \ldots, m$ 的各条等式相加，中间各项互相抵消：

$$2^m I_m = 2^0 I_0 + \sum_{k=1}^{m} \frac{2^{k-1}}{k} \cos^k \theta \sin k\theta.$$

最后，$I_0 = \int_0^\theta 1 \, \mathrm{d}x = \theta$，所以

$$\begin{aligned} S = 2^m I_m - \theta &= \sum_{k=1}^{m} \frac{1}{k} 2^{k-1} \cos^k \theta \sin k\theta \\ &= \cos\theta \sin\theta + \cos^2\theta \sin 2\theta + \cdots + \frac{1}{m} 2^{m-1} \cos^m \theta \sin m\theta, \end{aligned}$$

这就是所求的结果。

验算

- $m = 1$：所要证明的是 $\frac{1}{2} \sin 2\theta = \cos\theta \sin\theta$，这正是二倍角公式（double-angle formula）。
- $m = 2$：左边是 $\sin 2\theta + \frac{1}{4} \sin 4\theta$，而右边是 $\cos\theta \sin\theta + \cos^2\theta \sin 2\theta = \frac{1}{2} \sin 2\theta + \frac{1}{2} \sin 2\theta (1 + \cos 2\theta)$，两者相同，因为 $\sin 2\theta \cos 2\theta = \frac{1}{2} \sin 4\theta$。
- 数值上，取 $m = 5$ 和 $\theta = 0.7$，两个表达式都等于 1.51963（准确至小数点后 5 位）。对随机选取的 k 和 θ，归约公式亦与用辛普森法则（Simpson's rule）算出的 I_k 值吻合。

常见错误

- 分部积分时弄错正负号：$\frac{\mathrm{d}}{\mathrm{d}x} \cos^k x = -k \cos^{k-1} x \sin x$，而这个负号会与 $\int v \, \mathrm{d}u$ 前面的负号互相抵消。
- 忘记 (ii) 中 $r = 0$ 的一项。它积分后得 θ，并与裂项求和最底端的 $I_0 = \theta$ 互相抵消。

STEP 之外

把 $\cos^m\theta\cos m\theta$ 写成 $\cos 2r\theta$ 的组合，其实就是计算一个傅里叶级数（Fourier series）。这里的级数只有有限项，而逐项积分正是大学一年级数学方法课程中处理傅里叶级数的做法。(i) 中那类归约公式，也是处理 $\int\cos^n x\,\mathrm{d}x$ 这种形式的积分以及沃利斯乘积（Wallis product）的标准工具。Needham 的 *Visual Complex Analysis* 为 (ii) 中用到的技巧 $1+\mathrm{e}^{2\mathrm{i}\theta}=2\cos\theta\,\mathrm{e}^{\mathrm{i}\theta}$ 提供了一幅几何图像。

详细解答

STEP III 1991，第 9 题

纯数学 · 椭圆的弧长

> **STEP III 1991，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/91-s3-q9

关键思路

E_1 是以中心为原点的椭圆（ellipse）方程，而 E_2 是同一条曲线的极坐标（polar）形式，极点位于一个焦点（focus）。把 E_2 化为直角坐标（Cartesian）形式再配方（completing the square），便可看出它的中心和轴在哪里；比较两者的半轴（semi-axes）即可求得 k 和 b。至于积分，周界（perimeter）只是一个数，但每一种参数化（parametrisation）都会给出外表不同的弧长（arc length）积分。

解答

坐标轴、k 和 b

以 (x,y) 表示 E_2 的坐标，并设 $r=\sqrt{x^2+y^2}$，$\theta=\theta_2$。于是 E_2 即 $r=\dfrac{k}{1+e\cos\theta}$，亦即 $r+er\cos\theta=k$，所以

$$r=k-ex.$$

两边平方（这一步没有损失任何资料，因为在曲线上 $r=k/(1+e\cos\theta)>0$），

$$x^2+y^2=k^2-2kex+e^2x^2 \quad\Longrightarrow\quad (1-e^2)x^2+2kex+y^2=k^2.$$

对 x 配方：

$$(1-e^2)\left(x+\frac{ke}{1-e^2}\right)^2+y^2=k^2+\frac{k^2e^2}{1-e^2}=\frac{k^2}{1-e^2},$$

所以

$$\frac{\left(x+\frac{ke}{1-e^2}\right)^2}{k^2/(1-e^2)^2}+\frac{y^2}{k^2/(1-e^2)}=1.$$

这是一个椭圆，在 E_2 坐标轴下中心为 $\left(-\frac{ke}{1-e^2},0\right)$，对称轴与 E_2 的坐标轴平行，半轴为

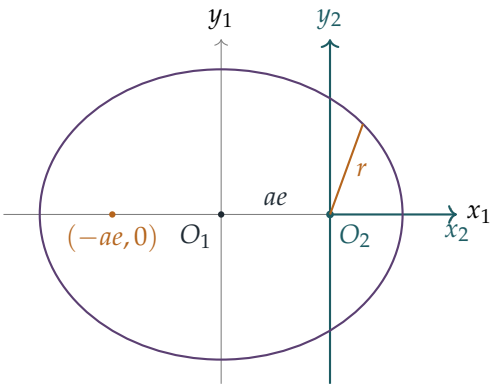
$$A=\frac{k}{1-e^2}\text{（沿 }x\text{ 方向）}，\qquad B=\frac{k}{\sqrt{1-e^2}}\text{（沿 }y\text{ 方向）}.$$

由于 $0<1-e^2<1$，我们有 $A>B$，所以长轴（major axis）位于 E_2 的 x 轴上。对于 E_1，$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 且 $a>b$，所以它的长轴位于 E_1 的 x 轴上。两条曲线是同一条曲线，所以半长轴相等，半短轴也相等：

$$a=\frac{k}{1-e^2},\quad \text{即}\quad k=a(1-e^2),\qquad b^2=\frac{k^2}{1-e^2}=\frac{a^2(1-e^2)^2}{1-e^2}=a^2(1-e^2),$$

这正是所需的两个结果。

中心（即 E_1 的原点）位于 E_2 的 x 轴上 $x=-\frac{ke}{1-e^2}=-ae$ 之处。所以 E_2 的坐标轴就是把 E_1 的坐标轴沿长轴平移距离 ae 而得：E_2 的原点是 E_1 中的点 $(ae,0)$，而由于 $a^2-b^2=a^2e^2$，这一点正是焦点。两条 x 轴位于同一条直线（长轴）上，两条 y 轴互相平行；以 E_1 的坐标表示，$x_1=x_2+ae$，$y_1=y_2$。椭圆对它的两条轴都对称，所以我们同样可以把 E_2 的原点取在另一个焦点 $(-ae,0)$，并把 x 轴反向。无论哪一种取法，$\theta_2=0$ 所对应的点都是最接近该焦点的顶点（vertex）。



两个积分

我们用两种方法计算椭圆上半部分（$y\geq 0$）的长度。

利用 E_1。当 θ_1 由 0 变到 π 时，动点刚好走遍上半部分一次。由 $\dot{x}=-a\sin\theta_1,\dot{y}=b\cos\theta_1$，

$$\dot{x}^2+\dot{y}^2=a^2\sin^2\theta_1+b^2\cos^2\theta_1=a^2-(a^2-b^2)\cos^2\theta_1=a^2\left(1-e^2\cos^2\theta_1\right),$$

其中用到 $a^2-b^2=a^2e^2$。所以半周界为

$$\tfrac{1}{2}P=a\int_0^{\pi}\sqrt{1-e^2\cos^2\theta}\,\mathrm{d}\theta.$$

利用 E_2。这里由于 $r>0$，$y_2=r\sin\theta_2\ge 0$ 当且仅当 $0\le\theta_2\le\pi$。所以 $\theta_2\in[0,\pi]$ 同样刚好描出上半部分一次。在极坐标中，$\left(\frac{\mathrm{d}s}{\mathrm{d}\theta}\right)^2=r^2+\left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2$。由 $r=k(1+e\cos\theta)^{-1}$，

$$\frac{\mathrm{d}r}{\mathrm{d}\theta}=\frac{ke\sin\theta}{(1+e\cos\theta)^2},$$

所以

$$r^2+\left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2=\frac{k^2\left[(1+e\cos\theta)^2+e^2\sin^2\theta\right]}{(1+e\cos\theta)^4}=\frac{k^2\left(1+e^2+2e\cos\theta\right)}{(1+e\cos\theta)^4}.$$

因此，利用 $k=a(1-e^2)$，

$$\tfrac{1}{2}P=a(1-e^2)\int_0^{\pi}\frac{\sqrt{1+e^2+2e\cos\theta}}{(1+e\cos\theta)^2}\,\mathrm{d}\theta.$$

令 $\frac{1}{2}P$ 的两个表达式相等，再除以 $a>0$，便得到所需的恒等式（identity）。

想一想 为甚么两个参数都必须取遍 $[0,\pi]$，而不是其中一个取另一个范围?

两个弧长必须是同一段曲线的长度。θ_1 和 θ_2 是不同的角（一个是从中心量度的离心角（eccentric angle），另一个是从焦点量度的极角），但两者在 $[0,\pi]$ 上都是由右顶点 $(a,0)$ 到左顶点 $(-a,0)$ 扫过上半部分。所以两个积分量度的是同一段弧。

近似值

由于 $0\le e^2\cos^2\theta<1$，二项级数（binomial series）适用：

$$\sqrt{1-u}=1-\tfrac{1}{2}u-\tfrac{1}{8}u^2+O(u^3),\qquad u=e^2\cos^2\theta.$$

$O(u^3)$ 项是 $O(e^6)$，我们把它略去。利用

$$\int_0^{\pi}\cos^2\theta\,\mathrm{d}\theta=\frac{\pi}{2},\qquad\int_0^{\pi}\cos^4\theta\,\mathrm{d}\theta=\frac{3\pi}{8}$$

（第二个由 $\cos^4\theta=\frac{3}{8}+\frac{1}{2}\cos 2\theta+\frac{1}{8}\cos 4\theta$ 得出），我们得到

$$\begin{aligned}\int_0^{\pi}\sqrt{1-e^2\cos^2\theta}\,\mathrm{d}\theta&\approx\pi-\frac{e^2}{2}\cdot\frac{\pi}{2}-\frac{e^4}{8}\cdot\frac{3\pi}{8}=\pi\left(1-\frac{e^2}{4}-\frac{3e^4}{64}\right)\\&=\tfrac{1}{64}\pi\left(64-16e^2-3e^4\right).\end{aligned}$$

根据刚证明的恒等式，第二个积分的值相同。至此全题完成。

验算

- 当 $e=0$ 时，两个积分都等于 π，而 E_2 变成圆 $r=k$，其中 $k=a$：正确。

- 取 $a = 2$，$b = 1$（所以 $e^2 = \frac{3}{4}$），用一个 200 000 边形计算，得周界为 9.6884482，与 $2a$ 乘以第一个积分的值吻合。
- 数值上，当 $e = 0.2$ 时两个积分都是 3.1099371，而近似式给出 3.1099411。两者之差约为 $-0.063e^6$，与级数的下一项 $-\frac{5\pi}{256}e^6 \approx -0.061e^6$ 相符。

常见错误

- 把 E_2 的原点放在中心。分母为 $1 + e\cos\theta$ 的极坐标形式，极点必定在焦点。
- 在极坐标弧长公式中漏掉 $(\mathrm{d}r/\mathrm{d}\theta)^2$ 一项，或者一边在 $[0, 2\pi]$ 上积分，另一边却在 $[0, \pi]$ 上积分。

STEP 之外

积分 $\int_0^{\pi/2} \sqrt{1 - e^2\sin^2\theta}\,\mathrm{d}\theta$ 是第二类完全椭圆积分（complete elliptic integral of the second kind）$E(e)$，所以椭圆的周界是 $4aE(e)$。它不能用初等函数（elementary functions）表示，这正是本题这类级数（以及拉马努金（Ramanujan）那些准确得惊人的近似公式）之所以重要的原因。由椭圆积分再走下去，便是椭圆函数（elliptic functions）和椭圆曲线（elliptic curves），后者是现代数论的核心课题。想做更多这种风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP III 1991，第 10 题

纯数学 · 牛顿恒等式

> **STEP III 1991，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/91-s3-q10

关键思路

这个多项式只有三个非零系数，所以对称函数（symmetric function）Σ_m 几乎全部为零，而题目给出的恒等式（牛顿恒等式，Newton's identity）就简化成一个只有两项的递推关系（recurrence）。次数超过 n 之后，这个恒等式不再适用，但有一个更简单的做法：每个根都满足方程本身，所以把方程乘以 α^k 再对所有根求和，便可把 S_{n+k} 与较早的幂和（power sum）联系起来。

解答

直到 S_{n-1} 的幂和

把 $x^n - qx^{n-1} + r$ 与 $\prod_i (x - \alpha_i) = x^n - \Sigma_1 x^{n-1} + \Sigma_2 x^{n-2} - \cdots + (-1)^n \Sigma_n$ 比较，可得

$$\Sigma_1 = q, \qquad \Sigma_2 = \Sigma_3 = \cdots = \Sigma_{n-1} = 0, \qquad (-1)^n \Sigma_n = r.$$

在题目给出的恒等式中取 $1 \le p \le n-1$。由于 $p \le n-1$，所有含有 $\Sigma_2, \ldots, \Sigma_p$ 的项都等于零。（当 $p = 1$ 时，恒等式就是 $S_1 - \Sigma_1 = 0$。）剩下的是

$$S_p - q\, S_{p-1} = 0 \quad (2 \le p \le n-1), \qquad S_1 = q.$$

所以 $S_1 = q$，而每个幂和都是前一个的 q 倍。对 p 用归纳法（induction），得

$$S_p = q^p \qquad \text{对于} 1 \le p \le n-1,$$

这正是所要证明的结果。另外，由于每个 $\alpha_i^0 = 1$，顺带记下 $S_0 = n$ 会很方便。

S_n、S_{n+1} 和 S_{n+2}

当 $p = n$ 时，恒等式仍然适用，而这次最后一项会出现：

$$S_n - q\,S_{n-1} + (-1)^n n\,\Sigma_n = 0.$$

由于 $(-1)^n\Sigma_n = r$，由此得

$$S_n = q \cdot q^{n-1} - nr = q^n - nr.$$

想一想 题目给出的恒等式到 $p = n$ 就停了。怎样才能再走下去?

我们利用方程本身。每个根都满足 $\alpha_i^n = q\alpha_i^{n-1} - r$。乘以 α_i^k（其中 $k \geq 0$）再对 i 求和，得

$$S_{n+k} = q\,S_{n+k-1} - r\,S_k \qquad (k \geq 0). \tag{$*$}$$

取 $k = 0$，这再次给出 $S_n = q^n - nr$，是一个令人安心的交叉验证。现在

$$\begin{aligned} S_{n+1} &= q\,S_n - r\,S_1 = q^{n+1} - nqr - qr = q^{n+1} - (n+1)qr, \\ S_{n+2} &= q\,S_{n+1} - r\,S_2 = q^{n+2} - (n+1)q^2r - q^2r = q^{n+2} - (n+2)q^2r. \end{aligned}$$

这里用到了 $S_1 = q$ 和 $S_2 = q^2$；它们成立，是因为 $n \geq 5$ 保证了 $2 \leq n-1$。

一般公式

从 $S_n = q^n - nr$、$S_{n+1} = q^{n+1} - (n+1)qr$、$S_{n+2} = q^{n+2} - (n+2)q^2r$ 这个规律，可以猜想

$$S_{n+m} = q^{n+m} - (n+m)\,q^m r \qquad \text{对于} 0 \leq m \leq n-1.$$

我们对 m 用归纳法来证明。$m = 0$ 的情况就是 $S_n = q^n - nr$，上面已经证明。

假设公式对某个 $m-1$ 成立，其中 $1 \leq m \leq n-1$。在 $(*)$ 中取 $k = m$，得

$$\begin{aligned} S_{n+m} &= q\,S_{n+m-1} - r\,S_m \\ &= q\big(q^{n+m-1} - (n+m-1)q^{m-1}r\big) - r\,q^m \\ &= q^{n+m} - (n+m)\,q^m r, \end{aligned}$$

其中用到了 $S_m = q^m$，这是因为 $1 \leq m \leq n-1$ 才成立。这就是 m 的情况下的公式。根据归纳法，公式对所有 $0 \leq m \leq n-1$ 成立，即对 $m < n$ 成立。

想一想 条件 $m < n$ 究竟在哪里用到?

只在 $S_m = q^m$ 这一步。当 $m = n$ 时，我们需要的是 S_n，而它等于 $q^n - nr$，不是 q^n，于是多出来的 $-nr$ 会带进计算之中，规律就此打破。事实上，由 $(*)$ 可得 $S_{2n} = q^{2n} - 2nq^nr + nr^2$。

验算

- 当 $r = 0$ 时，方程是 $x^{n-1}(x - q) = 0$，根为 q 和 0（重复 $n - 1$ 次），所以每个 $S_k = q^k$。我们所有的公式都化为这个结果。
- 当 $q = 0$ 时，各根就是 $-r$ 的 n 次方根。于是对于 $1 \le p \le n - 1$ 有 $S_p = 0$，而 $S_n = n \cdot (-r) = -nr$，与 $q^n - nr$ 吻合。
- 数值上，我们对随机选取的 q, r 和 $5 \le n \le 9$ 求出所有根，再把它们的幂加起来。$S_p = q^p$ 以及 S_{n+m} 的公式（$0 \le m \le n - 1$）都吻合至约 10^{-12} 的误差，而在 $m = n$ 时公式明显不成立。

常见错误

- 忘记恒等式最后一项的因子 p，结果得出 $S_n = q^n - r$ 而不是 $q^n - nr$。
- 弄错 Σ_n 的正负号：常数项是 $(-1)^n \Sigma_n$，所以不论 n 是奇是偶，都有 $(-1)^n n \Sigma_n = nr$。
- 在 $p > n$ 时套用题目给出的恒等式，而照题目所写的形式，它在这个范围并不成立。

STEP 之外

这些就是牛顿恒等式，而 $(*)$ 表明各幂和满足一个线性递推关系（linear recurrence），其特征多项式（characteristic polynomial）正是原来的方程。用矩阵（matrix）的语言来说，S_k 是 A^k 的迹（trace），其中 A 是该多项式的友矩阵（companion matrix）。各根的任何对称多项式（symmetric polynomial）都可以用 Σ_m 表示，这个事实就是对称多项式基本定理（fundamental theorem of symmetric polynomials），由此可以通往伽罗瓦理论（Galois theory）。Allenby 的 *Rings, Fields and Groups*（环、域与群）是了解这套代数如何展开的好去处。

详细解答

STEP III 1991，第 11 题

力学 · 刚体平衡

STEP III 1991，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：

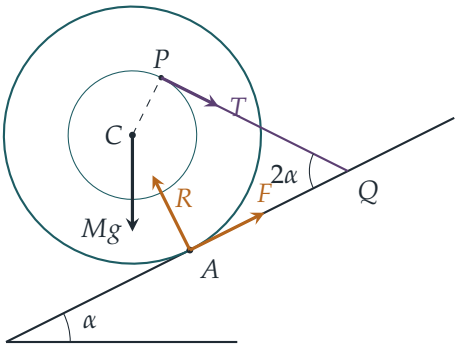
step.maths.org/questions/91-s3-q11

关键思路

圆柱体是一个刚体（rigid body），在四个力的作用下处于平衡（equilibrium）：重力、法向反作用力（normal reaction）、接触点处的摩擦力（friction），以及绳子的张力（tension）。其中重力和法向反作用力都通过圆柱体的中心，所以对中心取力矩（moments），便可把摩擦力与张力直接联系起来。再沿斜面方向和垂直于斜面方向分解力，就得到摩擦力和法向反作用力，余下的工作交给不等式 $F \leq \mu R$ 即可。在第二部分，圆柱体并不“知道”是甚么东西拉住绳子，所以张力不变，我们只需考虑质点。

解答

我们把圆柱体看作均匀刚体，其质心（centre of mass）就是它的中心 C；把绳子看作轻而不可伸长（light and inextensible）；把第二个物体看作质点（particle）。设 A 为圆柱体与斜面的接触点，并设绳子的张力为 T。由于绳子绕在凹槽上，自由的一段 PQ 在 P 点与凹槽相切，所以它的作用线与 C 的垂直距离为 a。由图可见，Q 在斜面上的位置比 A 更高，而绳子把圆柱体由 P 拉向 Q：既沿斜面向上，也压向斜面，与斜面成 2α 角。



虚线所示的半径 CP 长度为 a，并与绳子垂直。我们以 F 表示作用于圆柱体的摩擦力，以沿斜面向上为正。

圆柱体：证明 $\mu \geq \frac{1}{3}\tan\alpha$

对 C 取力矩。重力作用于 C，法向反作用力沿 AC 作用，所以两者对 C 都没有力矩。摩擦力的作用点与 C 相距 $2a$，张力的作用线与之垂直距离为 a，两者使圆柱体朝相反方向转动，所以

$$F \cdot 2a = T \cdot a, \qquad \text{即} \qquad T = 2F.$$

想一想 为甚么摩擦力和张力会使圆柱体朝相反方向转动?

只有这两个力对 C 有力矩，所以它们的力矩必须互相抵消。单靠绳子在 P 点的拉力，圆柱体会转动而沿斜面向上滚。A 点处沿斜面向上的摩擦力则使它朝另一方向转动。假如我们猜错了摩擦力的方向，下面的方程只会给出 $F < 0$，所以这个正负号并不是额外的假设。

沿斜面方向分解（以沿斜面向上为正）：

$$F + T\cos 2\alpha = Mg\sin\alpha \quad \Longrightarrow \quad F(1 + 2\cos 2\alpha) = Mg\sin\alpha.$$

垂直于斜面方向分解：张力有一个压向斜面的分量 $T\sin 2\alpha$，所以

$$R = Mg\cos\alpha + T\sin 2\alpha = Mg\cos\alpha + 2F\sin 2\alpha.$$

代入 $F = Mg\sin\alpha/(1 + 2\cos 2\alpha)$，

$$\begin{aligned} R(1 + 2\cos 2\alpha) &= Mg\big(\cos\alpha + 2\cos 2\alpha\cos\alpha + 2\sin 2\alpha\sin\alpha\big) \\ &= Mg\big(\cos\alpha + 2\cos(2\alpha - \alpha)\big) = 3Mg\cos\alpha. \end{aligned}$$

因此

$$F = \frac{Mg\sin\alpha}{1 + 2\cos 2\alpha}, \qquad R = \frac{3Mg\cos\alpha}{1 + 2\cos 2\alpha}.$$

圆柱体保持与斜面接触，所以 $R > 0$。由于 α 是锐角，$\cos\alpha > 0$，于是 $1 + 2\cos 2\alpha > 0$，这表示 $F > 0$ 亦成立。两式相除，

$$\frac{F}{R} = \frac{\sin\alpha}{3\cos\alpha} = \tfrac{1}{3}\tan\alpha.$$

要保持平衡，摩擦力必须满足 $F \le \mu R$，所以 $\mu \ge F/R = \frac{1}{3}\tan\alpha$，正是所要证明的结果。

质点：摩擦系数的最小值

绳子留在原来的位置，所以圆柱体的位置完全相同，它的三条方程也不变。特别是，张力仍然是

$$T = 2F = \frac{2Mg\sin\alpha}{1 + 2\cos 2\alpha}.$$

当 $\tan\alpha = \frac{1}{2}$ 时，我们有 $\sin\alpha = \frac{1}{\sqrt{5}}$ 和 $\cos\alpha = \frac{2}{\sqrt{5}}$，所以

$$\cos 2\alpha = \frac{1 - \tan^2\alpha}{1 + \tan^2\alpha} = \frac{3}{5}, \qquad \sin 2\alpha = \frac{2\tan\alpha}{1 + \tan^2\alpha} = \frac{4}{5},$$

而 $1 + 2\cos 2\alpha = \frac{11}{5}$。因此

$$T = \frac{2Mg}{\sqrt{5}} \cdot \frac{5}{11} = \frac{2\sqrt{5}}{11} Mg.$$

（圆柱体同时需要 $\mu \ge \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$。）

现在考虑位于 Q 的质点。绳子把它拉向 P，也就是说，拉力有沿斜面向下的分量 $T\cos 2\alpha$，以及离开斜面的分量 $T\sin 2\alpha$。它的重力 $3Mg$ 同样有沿斜面向下的分量，所以作用于质点的摩擦力 F' 必定沿斜面向上。设作用于质点的法向反作用力为 R'。

沿斜面方向：

$$F' = 3Mg\sin\alpha + T\cos 2\alpha = \frac{3Mg}{\sqrt{5}} + \frac{3}{5} \cdot \frac{2\sqrt{5}}{11} Mg = \frac{33\sqrt{5}}{55} Mg + \frac{6\sqrt{5}}{55} Mg = \frac{39\sqrt{5}}{55} Mg.$$

垂直于斜面方向：

$$R' = 3Mg\cos\alpha - T\sin 2\alpha = \frac{6Mg}{\sqrt{5}} - \frac{4}{5} \cdot \frac{2\sqrt{5}}{11} Mg = \frac{66\sqrt{5}}{55} Mg - \frac{8\sqrt{5}}{55} Mg = \frac{58\sqrt{5}}{55} Mg.$$

这是正数，所以质点确实留在斜面上。平衡的条件是 $F' \le \mu' R'$，其中 μ' 是质点与斜面之间的摩擦系数（coefficient of friction），所以

$$\mu' \ge \frac{F'}{R'} = \frac{39}{58}.$$

摩擦系数的最小值是 $\dfrac{39}{58}$。

验算

- 改为对 A 取力矩，得 $T\,a(1+2\cos 2\alpha) = 2a\,Mg\sin\alpha$，张力相同。（由 A 到绳子的垂直距离是 $a+2a\cos 2\alpha$。）
- 当 $\alpha \to 0$ 时，斜面变成水平，张力和摩擦力都消失，而下界 $\frac{1}{3}\tan\alpha \to 0$，与预期相符。
- 数值验算：随机选取 10 000 个小于 60° 的锐角 α，用坐标建立几何图形，解出平衡方程，确认 $F/R = \frac{1}{3}\tan\alpha$。当 $\tan\alpha = \frac{1}{2}$ 时，结果是 $T \approx 0.4066\,Mg$、$F' \approx 1.5856\,Mg$、$R' \approx 2.3580\,Mg$，以及 $F'/R' \approx 0.67241 = 39/58$。

常见错误

- 忘记张力会把圆柱体压向斜面，令 R 增大。漏掉 $T\sin 2\alpha$ 会得出错误的比值。
- 弄错绳子对质点的拉力方向。绳子把质点拉向 P，所以它减少了作用于质点的法向反作用力，并加大了沿斜面向下的力。
- 在第二部分从头重新计算张力。圆柱体的平衡没有改变，所以 T 也没有改变。

STEP 之外

刚体静力学（statics）归结为两条向量方程：合力为零，以及对任何一点的总力矩为零。选好这一点是最主要的技巧；本题选的是中心，因为有两个未知的力通过它。到了大学，这会发展成对力系及其等效“力螺旋”（wrench）的研究。摩擦力问题还会引出平衡如何被破坏的问题：是滑动、滚动，还是翻倒。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对刚体静力学和转动有细致的论述。

详细解答

STEP III 1991，第 12 题

力学 · 曲线上的运动

> **STEP III 1991，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/91-s3-q12

关键思路

质点被约束在椭圆上，所以它的位置完全由单一参数 θ 决定。把位置向量（position vector）对时间求导两次，再把加速度投影到向内的单位法向量（inward unit normal）上：含 $\ddot{\theta}$ 的项会消去，因为它们沿切线方向。至于第二部分，当法向反作用力（normal reaction）降至零时，质点便离开表面；而管是光滑的，所以由能量守恒（energy conservation）可求得每一点的速率。

解答

法向加速度

我们把质点视为在光滑曲线上运动的一点，因此它只受两个力：向下的重量 mg，以及管所施的法向反作用力 R。

位置为 $\mathbf{r} = (a\cos\theta,\ b\sin\theta)$。利用链式法则（chain rule）对时间求导，

$$\dot{x} = -a\sin\theta\,\dot{\theta}, \qquad \dot{y} = b\cos\theta\,\dot{\theta},$$
$$\ddot{x} = -a\sin\theta\,\ddot{\theta} - a\cos\theta\,\dot{\theta}^2, \qquad \ddot{y} = b\cos\theta\,\ddot{\theta} - b\sin\theta\,\dot{\theta}^2.$$

切线方向为 $(-a\sin\theta,\ b\cos\theta)$，所以 $(b\cos\theta,\ a\sin\theta)$ 是一个法线方向（两者的点积（dot product）为 0）。在 $\theta = 0$ 时，它等于 $(b, 0)$，指向椭圆之外，因此向内的单位法向量为

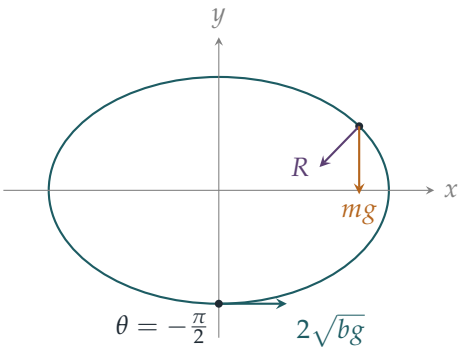
$$\mathbf{n} = -\frac{(b\cos\theta,\ a\sin\theta)}{D}, \qquad D = \sqrt{a^2\sin^2\theta + b^2\cos^2\theta}.$$

于是沿向内法线方向的加速度为

$$\begin{aligned}\ddot{\mathbf{r}}\cdot\mathbf{n} &= -\frac{1}{D}\Big[b\cos\theta\big(-a\sin\theta\,\ddot\theta - a\cos\theta\,\dot\theta^2\big) + a\sin\theta\big(b\cos\theta\,\ddot\theta - b\sin\theta\,\dot\theta^2\big)\Big]\\ &= -\frac{1}{D}\Big[-ab\,\dot\theta^2\big(\cos^2\theta+\sin^2\theta\big)\Big] = \frac{ab\,\dot\theta^2}{\sqrt{a^2\sin^2\theta + b^2\cos^2\theta}},\end{aligned}$$

因为两个含 $\ddot\theta$ 的项互相抵消。这正是所求的结果。

离开表面



最低点对应 $\theta=-\frac{\pi}{2}$，该处 $y=-b$。管是光滑的，所以反作用力不作功，能量守恒。设质点在参数为 θ 的点上速率为 v，则

$$\tfrac12 mv^2 + mgb\sin\theta = \tfrac12 m(4bg) - mgb \quad\Longrightarrow\quad v^2 = 2gb(1-\sin\theta).$$

由第一部分，$v^2 = \dot x^2+\dot y^2 = D^2\dot\theta^2$，所以 $\dot\theta^2 = v^2/D^2$。

现在沿向内法线 $\mathbf{n}$ 的方向分解各力。反作用力 R 沿 $\mathbf{n}$ 作用（质点在内侧，所以管向内推，而且只要仍有接触便有 $R\ge 0$）。重量为 $(0,-mg)$，它沿 $\mathbf{n}$ 的分量是 $(0,-mg)\cdot\mathbf{n} = mga\sin\theta/D$。由牛顿第二定律（Newton's second law）得

$$R + \frac{mga\sin\theta}{D} = \frac{m\,ab\,\dot\theta^2}{D} = \frac{m\,ab\,v^2}{D^3}, \qquad \text{所以} \qquad R = \frac{ma}{D^3}\Big(bv^2 - g\sin\theta\,D^2\Big).$$

? **想一想** 为甚么重量的分量是 $+mga\sin\theta/D$，而不是带负号?

在上半部分（$\sin\theta>0$），向内法线有向下的分量，所以重力在那里有助把质点拉向内侧，情况就如在铅垂圆周的顶部一样。R 和重量的这个分量共同提供向心加速度（centripetal acceleration）。

现在代入 $a=\frac32 b$ 和 $v^2 = 2gb(1-\sin\theta)$，并记 $s=\sin\theta$。那么

$$D^2 = \tfrac94 b^2 s^2 + b^2(1-s^2) = b^2\big(1+\tfrac54 s^2\big),$$

所以

$$bv^2 - g\sin\theta\, D^2 = gb^2\left[2(1-s) - s\left(1+\tfrac{5}{4}s^2\right)\right] = -\tfrac{1}{4}gb^2\left(5s^3+12s-8\right).$$

由于 $ma/D^3 > 0$，R 的正负号与 $-(5s^3+12s-8)$ 的正负号相同。

质点在 R 首次变为零之处离开表面。函数 $p(s) = 5s^3+12s-8$ 严格递增（strictly increasing），因为 $p'(s) = 15s^2+12 > 0$，而且 $p(-1) = -25 < 0$，$p(1) = 9 > 0$。所以 p 在 $(-1,1)$ 内恰有一个根。质点由最低点上升时，s 由 -1 开始递增，在 s 到达该根之前一直有 $R > 0$。此外，对所有 $s < 1$ 都有 $v^2 = 2gb(1-s) > 0$，所以质点到达该处时仍在沿管向上运动。因此 R 首次变为零、质点离开表面的位置满足

$$5\sin^3\theta + 12\sin\theta - 8 = 0,$$

正如所求。数值上，$\sin\theta \approx 0.584$，所以质点在略高于轴的水平之处离开。

验算

- 若 $a = b = r$，法向加速度变成 $r\dot{\theta}^2 = v^2/r$，即熟悉的圆周运动结果。
- 在最低点，$R = m\left(\frac{bv^2}{a^2}+g\right)$，其中 $v^2 = 4bg$，其值为正（该处的曲率半径（radius of curvature）为 a^2/b），与 $p(-1) = -25 < 0$ 吻合。
- 取 $b = 1$、$a = 1.5$ 和 $g = 9.81$，我们由最低点开始，逐步对以 θ 表示的运动方程作数值积分。反作用力首次变为零时 $\sin\theta \approx 0.58379$，与三次方程的根 0.58377 相符（$\theta \approx 35.7°$）。

常见错误

- 把 θ 当作极角（polar angle）。除非 $a = b$，否则点 $(a\cos\theta, b\sin\theta)$ 与 x 轴的夹角并不是 θ，法线也不沿半径方向。
- 以 v^2/b 或 v^2/a 作为法向加速度。在椭圆上，曲率半径逐点不同，而这正是第一部分所计算的。
- 弄错重量法向分量的正负号，结果三次式变成 $5s^3 - 4s + \ldots$，得不到合理的根。

STEP 之外

第一部分是法向加速度公式 v^2/ρ 的特殊情况，其中 ρ 是曲率半径。事实上，这个椭圆的 $\rho = D^3/(ab)$。把加速度分解为切向和法向两部分（弗勒内标架（Frenet frame）），是

大学一年级力学和微分几何（differential geometry）的标准内容。以单一坐标 θ 处理约束，自然引向拉格朗日力学（Lagrangian mechanics）。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）对约束运动，以及极坐标和内禀坐标（intrinsic coordinates），都有很好的讲解。

详细解答

STEP III 1991，第 13 题

力学 · 抛体与天花板的碰撞

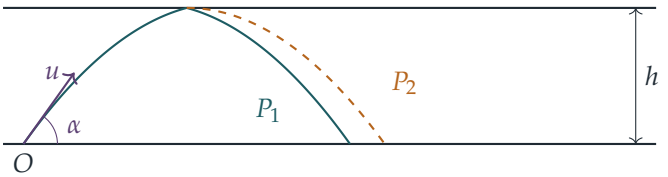
> **STEP III 1991，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/91-s3-q13

关键思路

天花板是光滑的，所以水平速度 $u\cos\alpha$ 始终不变，每个射程都只是 $u\cos\alpha$ 乘以飞行时间。全部工夫都在铅垂方向的运动上，而能量方程（energy equation）可以立即给出质点到达天花板时的铅垂速率。经过弹性碰撞（elastic impact）之后，质点下落的过程与上升的过程完全一样，只是倒过来播放；经过非弹性碰撞（inelastic impact）之后，它就只是从静止落下。

解答

把两个质点都视为在重力作用下自由运动的质点（point mass），不计空气阻力。记 $u=\sqrt{8gh}$、$s=\sin\alpha$ 及 $c=\cos\alpha$。以 O 为原点，取 x 沿水平方向，y 铅垂向上。



到达天花板

初始铅垂速度为 us，而在上升期间

$$\dot{y}^2=u^2s^2-2gy=8gh\,s^2-2gy.$$

质点要到达高度 h，必须在该处有 $\dot{y}^2\geq 0$，即 $8ghs^2\geq 2gh$，所以 $s^2\geq\frac{1}{4}$。由于 α 是锐角，$s>0$，故此条件即 $\sin\alpha\geq\frac{1}{2}$，亦即 $\alpha\geq\frac{1}{6}\pi$。（当 $\alpha=\frac{1}{6}\pi$ 时，质点恰好擦过天花板。）当

它确实到达天花板时，它在该处的向上速率为

$$w = \sqrt{8ghs^2 - 2gh} = \sqrt{2gh}\,\sqrt{4s^2 - 1}.$$

P_1 的射程

在上升途中，铅垂速度以 g 的变率由 us 均匀地减至 w，所以到达天花板所需的时间为

$$t_1 = \frac{us - w}{g}.$$

> **想一想** P_1 从天花板回落到地面需要多少时间?

碰撞是完全弹性的，而天花板是光滑的，所以铅垂速度变为向下的 w，水平速度则不变。于是，由天花板向下的铅垂运动，就是上升时的铅垂运动在时间上倒过来进行（从高度 h 出发，以速率 w 朝向地面，并以 g 加速）。因此它到达地面时的铅垂速率为 us，所需时间同样是 t_1。直接计算亦可：铅垂速率以 g 的变率由 w 增至 $\sqrt{w^2 + 2gh} = us$。

总飞行时间为 $2t_1$，所以射程（range）为

$$\begin{aligned} R_1 &= uc \cdot \frac{2(us - w)}{g} = \frac{2\sqrt{8gh}\,c}{g}\left(2\sqrt{2gh}\,s - \sqrt{2gh}\sqrt{4s^2 - 1}\right) \\ &= \frac{2\sqrt{8gh}\sqrt{2gh}}{g}\,c\left(2s - \sqrt{4s^2 - 1}\right) = 8h\cos\alpha\left(2\sin\alpha - \sqrt{4\sin^2\alpha - 1}\right), \end{aligned}$$

当中用到 $us = \sqrt{8gh}\,s = 2\sqrt{2gh}\,s$ 及 $\sqrt{8gh}\sqrt{2gh} = 4gh$。这就是所求的射程。

P_2 的射程与差距 D

P_2 沿相同的路径上升至天花板，所需时间为 t_1。碰撞是完全非弹性的，所以它的铅垂速度变为零；天花板是光滑的，所以水平速度仍然是 uc。其后它从静止下落高度 h，所需时间为 $\sqrt{2h/g}$。因此

$$R_2 = uc\left(t_1 + \sqrt{\tfrac{2h}{g}}\right), \qquad R_2 - R_1 = uc\left(\sqrt{\tfrac{2h}{g}} - t_1\right).$$

现在

$$t_1 = \frac{\sqrt{2gh}\,(2s - \sqrt{4s^2 - 1})}{g} = \sqrt{\frac{2h}{g}}\,(2s - \sqrt{4s^2 - 1})$$

而 $uc\sqrt{2h/g} = \sqrt{8gh}\sqrt{2h/g}\,c = 4hc$，所以

$$R_2 - R_1 = 4h\cos\alpha\left(1 - 2\sin\alpha + \sqrt{4\sin^2\alpha - 1}\right).$$

当 $\frac{1}{6}\pi < \alpha < \frac{1}{2}\pi$ 时，此式为正：对于 $s > \frac{1}{2}$，我们有 $(2s-1)^2 = 4s^2 - 1 - 2(2s-1) < 4s^2 - 1$，所以 $2s - 1 < \sqrt{4s^2 - 1}$。因此 P_2 总是落在较远处（弹性反弹令 P_1 更快地被抛回地面），而

$$D = 4h\cos\alpha\left(1 - 2\sin\alpha + \sqrt{4\sin^2\alpha - 1}\right).$$

求 D 的极大值

设 $f(\alpha) = D/(4h) = c(1-2s) + cr$，其中 $r = \sqrt{4s^2-1}$，定义于 $\frac{1}{6}\pi \le \alpha < \frac{1}{2}\pi$。留意 $f(\frac{1}{6}\pi) = 0$，当 $\alpha \to \frac{1}{2}\pi$ 时 $f \to 0$（因为 $c \to 0$），而在两者之间 $f > 0$。求导（differentiate），并利用 $\frac{\mathrm{d}r}{\mathrm{d}\alpha} = \frac{4sc}{r}$，得

$$f'(\alpha) = -s(1-2s) - 2c^2 - sr + \frac{4sc^2}{r}.$$

代入 $c^2 = 1-s^2$，再乘以 $r > 0$（当 $\alpha > \frac{1}{6}\pi$）：

$$\begin{aligned} rf'(\alpha) &= (4s^2 - s - 2)\,r - sr^2 + 4s(1-s^2) \\ &= (4s^2 - s - 2)\,r - s(8s^2 - 5). \end{aligned}$$

当 $\alpha = \frac{1}{4}\pi$ 时，$s = \frac{1}{\sqrt{2}}$ 而 $r = 1$，得出 $\left(2 - \frac{1}{\sqrt{2}} - 2\right) - \frac{1}{\sqrt{2}}(4-5) = 0$。所以 $\alpha = \frac{1}{4}\pi$ 是一个驻点（stationary point）。

> **想一想** 我们怎样知道它就是极大值，而不是别的驻点？

在任何驻点，都有 $(4s^2 - s - 2)r = s(8s^2 - 5)$。两边平方并利用 $r^2 = 4s^2 - 1$，展开后因式分解，得

$$(4s^2 - s - 2)^2(4s^2 - 1) - s^2(8s^2 - 5)^2 = -2(2s^2 - 1)(8s^3 - s^2 - 2s - 2),$$

所以每个驻点都满足 $s^2 = \frac{1}{2}$ 或 $q(s) = 8s^3 - s^2 - 2s - 2 = 0$。在 $[\frac{1}{2}, 1]$ 上，$q'(s) = 24s^2 - 2s - 2 > 0$，所以 q 是递增的；由于 $q(0.8) = -0.144 < 0 < 0.0985 = q(0.82)$，它在该区间内唯一的根位于 $(0.8, 0.82)$ 之内。但对于这样的 s，

$$4s^2 - s - 2 < 4(0.82)^2 - 0.8 - 2 < 0, \qquad 8s^2 - 5 > 8(0.8)^2 - 5 > 0,$$

所以左边 $(4s^2 - s - 2)r$ 为负，右边 $s(8s^2 - 5)$ 为正：这个根是平方时产生的，并不是驻点。因此 $\alpha = \frac{1}{4}\pi$ 是 $(\frac{1}{6}\pi, \frac{1}{2}\pi)$ 内唯一的驻点。由于 f 是连续的，在区间内部为正，而在两端为零（或趋向零），它在区间内部必有极大值，而这极大值必定出现在这个驻点。所以当 $\alpha = \frac{1}{4}\pi$ 时 D 最大，且

$$D_{\max} = 4h \cdot \tfrac{1}{\sqrt{2}}\left(1 - \sqrt{2} + 1\right) = 4h(\sqrt{2} - 1).$$

验算

- 当 $\alpha = \frac{1}{6}\pi$ 时，质点刚好触及天花板，所以两个质点的结果应该一致：果然 $D = 0$，而 $R_1 = 8h \cdot \frac{\sqrt{3}}{2} = 4\sqrt{3}\,h$ 正是惯常的射程 $u^2 \sin 2\alpha / g$。
- 取 $h = 1.7\,\mathrm{m}$，在 $\alpha = 1$ 时以数值方法逐步计算两个运动，得 $R_1 \approx 2.420\,\mathrm{m}$ 及 $D \approx 2.464\,\mathrm{m}$，与公式吻合。对 α 作细密搜索，得出 $D/(4h)$ 的极大值位于 $\alpha = 0.78540 \approx \frac{1}{4}\pi$，其值为 $0.41421 \approx \sqrt{2} - 1$。

常见错误

- 由 $h = ust - \frac{1}{2}gt^2$ 解出 t_1 时取了较大的根；那其实是假如没有天花板，质点在下落途中经过高度 h 的时间。
- 写成 $D = R_1 - R_2$，这是负数：非弹性碰撞的质点落在较远处。
- 得出 $f'(\frac{1}{4}\pi) = 0$ 便停下来，没有证明这个驻点是极大值。

STEP 之外

对 P_1 所用的时间反演（time reversal）论证，让我们初次窥见一个深刻的事实：在保守力（conservative force）作用下，牛顿（Newton）定律在时间倒流时形式不变；完全弹性碰撞保持这种对称性，而非弹性碰撞（会损失能量）则破坏了它。介乎 0 与 1 之间的恢复系数（coefficient of restitution），会给出介乎这里求得的两个射程之间的结果。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）正是以这种精神处理碰撞与能量守恒。

详细解答

STEP III 1991，第 14 题

力学 · 涉及弹簧的圆周运动

STEP III 1991，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：

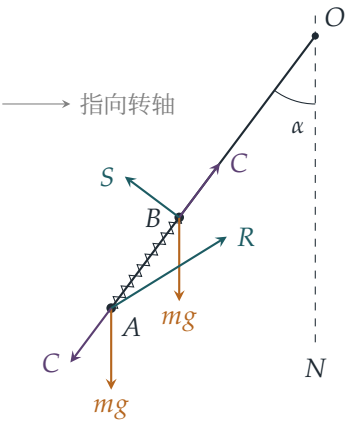
step.maths.org/questions/91-s3-q14

关键思路

所有物体都以相同的角速率（angular speed）绕 ON 作水平圆周运动，所以每个质点都有已知的加速度 $\omega^2 r$，方向指向转轴，此外再无其他加速度。小环套在一根光滑（smooth）的杆上，对它来说最简洁的方程是沿杆方向的那一条，因为未知的法向反作用力（normal reaction）不会在其中出现。接着，弹簧的力把小环与 A 处的质点联系起来，而杆在该处的反作用力的方向就决定了 k。

解答

模型。杆是轻而光滑的，A 和 B 是质量为 m 的质点，弹簧沿着杆位于 B 与 A 之间。设 $\alpha = \sin^{-1}\frac{3}{5}$，于是 $\sin\alpha = \frac{3}{5}$，$\cos\alpha = \frac{4}{5}$，并且 $\omega^2 = \frac{5g}{2a}$。设 $OB = x$，则 $AB = 2a - x$。



这里 C 是弹簧中的推力（thrust）（我们稍后会发现弹簧是被压缩的，所以它把 B 推向 O，并把 A 推离 O），S 是光滑杆对 B 的法向反作用力（垂直于杆，位于铅垂面内，指向有待确定），而 R 是杆对 A 的反作用力。由于 ω 是常数，而且 B 相对于杆静止，所以没有横越铅垂面的加速度，因此所有的力都在该平面内。

长度 OB

B 沿半径为 $x\sin\alpha$ 的水平圆周运动，所以它的加速度是 $\omega^2 x\sin\alpha$，方向水平指向 ON。沿杆按 $O\to A$ 的方向（向下并远离转轴）分解（resolve）。加速度在这个方向上的分量是 $-\omega^2 x\sin\alpha\cdot\sin\alpha$，因为加速度指向转轴，而杆与铅垂线的夹角是 α。

最简单的做法，是把弹簧作用于 B 的力设为张力（tension）T（把 B 拉向 A），让正负号自行在计算中显现。由胡克定律（Hooke's law），

$$T=\frac{kmg\left((2a-x)-a\right)}{a}=\frac{kmg(a-x)}{a}.$$

因为杆是光滑的，反作用力 S 没有沿杆方向的分量。所以，沿 OA 方向，

$$mg\cos\alpha+\frac{kmg(a-x)}{a}=-m\omega^2 x\sin^2\alpha.$$

代入 $\cos\alpha=\frac{4}{5}$、$\sin^2\alpha=\frac{9}{25}$ 和 $\omega^2=\frac{5g}{2a}$，右边就是 $-\frac{9gx}{10a}m$。两边除以 mg 再乘以 $10a$：

$$\begin{aligned}8a+10k(a-x)&=-9x\\(10k+8)a&=(10k-9)x,\end{aligned}$$

所以

$$OB=x=\frac{(10k+8)a}{10k-9},$$

这正是要证明的结果。留意 $x>a$（答案要为正数，必须有 $10k>9$），所以 $AB=2a-x<a$：弹簧是被压缩的，推力为

$$C=-T=\frac{kmg(x-a)}{a}=\frac{17k\,mg}{10k-9},$$

其中用到 $x-a=\dfrac{17a}{10k-9}$。

求 k

想一想 为甚么 A 处的反作用力不像 B 处的那样垂直于杆？

A 处的质点是固定在杆上的，而不是套在杆上，所以杆可以沿任何方向推它或拉它。光滑这个条件只限制了作用于滑动小环的力。因此我们把 R 写成分量：R_h 水平指向转轴，R_v 铅垂向上。

A 沿半径为 $2a\sin\alpha = \frac{6a}{5}$ 的圆周运动，所以它的加速度是 $\omega^2 \cdot \frac{6a}{5} = 3g$，方向指向转轴。推力 C 沿 OA 作用于 A，方向远离 O，所以它的分量是远离转轴的 $C\sin\alpha$ 和向下的 $C\cos\alpha$。因此

$$\text{水平（指向转轴）：}\quad R_h - \tfrac{3}{5}C = 3mg,$$
$$\text{铅垂：}\quad R_v - mg - \tfrac{4}{5}C = 0.$$

设 $C = c\,mg$。反作用力与水平线的夹角是 $\tan^{-1}\frac{13}{21}$，所以

$$\frac{R_v}{R_h} = \frac{1+\frac{4}{5}c}{3+\frac{3}{5}c} = \frac{13}{21} \implies 21 + \tfrac{84}{5}c = 39 + \tfrac{39}{5}c \implies 9c = 18, \quad c = 2.$$

（当 $c > 0$ 时，R_h 和 R_v 都是正数，所以反作用力确实指向上方和内侧，正负号并无歧义。）于是

$$\frac{17k}{10k-9} = 2 \implies 17k = 20k - 18 \implies k = 6.$$

所以 $k = 6$，由此得 $OB = \frac{68a}{51} = \frac{4}{3}a$，稳妥地介乎 a 与 $2a$ 之间，而推力为 $C = 2mg$。（这时作用于 A 的反作用力是 $R_h = \frac{21}{5}mg$，$R_v = \frac{13}{5}mg$。）

作用于小环的反作用力

> **想一想** 应该沿哪个方向分解，才能以最少的工夫求出 S?

垂直于杆的方向：这样弹簧的力就不会出现。在铅垂面内取垂直于 OA 的单位方向，指向上方并远离转轴；它与水平线的夹角是 α，所以它的分量是向外的 $\cos\alpha$ 和向上的 $\sin\alpha$。设 S 为作用于 B 的反作用力在这个方向上的分量。重量的贡献是 $-mg\sin\alpha$，而加速度（$\omega^2 x\sin\alpha$，指向转轴）的贡献是 $-\omega^2 x\sin\alpha\cos\alpha$。所以

$$S - mg\sin\alpha = -m\omega^2 x\sin\alpha\cos\alpha.$$

代入 $x = \frac{4}{3}a$：

$$m\omega^2 x\sin\alpha\cos\alpha = m\cdot\frac{5g}{2a}\cdot\frac{4a}{3}\cdot\frac{12}{25} = \frac{8}{5}mg, \qquad \text{所以} \qquad S = \frac{3}{5}mg - \frac{8}{5}mg = -mg.$$

因此杆与小环之间的反作用力的大小是

$$\boxed{mg},$$

而负号告诉我们，它垂直于杆，方向是向下并指向转轴。

验算

- 对 B 作完整的向量验算（水平分量以指向转轴为正，铅垂分量以向上为正），取 $C = 2mg$ 沿 AO 作用，以及上面求得的 $S = mg$。水平方向：$\frac{3}{5}(2mg) + \frac{4}{5}(mg) = 2mg$，而

$m\omega^2 x\sin\alpha = m\cdot\frac{5g}{2a}\cdot\frac{4a}{3}\cdot\frac{3}{5} = 2mg$。铅垂方向：$\frac{4}{5}(2mg) - \frac{3}{5}mg - mg = 0$。两者都平衡。

- 当 $k\to\infty$ 时，弹簧变成刚性，而 $OB\to a$，即弹簧的自然长度（natural length），与预期相符。
- 数值方面，取 $m = g = a = 1$，我们以 200 个随机的 k 值精确地验证了 OB 的公式，发现夹角只在 $k = 6$ 时才等于 $\tan^{-1}\frac{13}{21}$，并得到 $|S| = 1.0$。由一个小位移开始，逐步计算 B 沿杆的运动，结果它一直留在 $x\approx 1.3333$ 附近。

常见错误

- 假设弹簧被拉长，然后对 $x > a$ 感到困惑。应让张力的正负号由代数运算自行得出。
- 把 A 处的反作用力当作垂直于杆。只有滑动的小环才受这个限制。
- 圆周运动的半径用了 x 而不是 $x\sin\alpha$。

STEP 之外

在随杆一起转动的参考系中，B 受到一个向外的“离心力”（centrifugal force）$m\omega^2 r$，沿杆方向的方程就变成一道普通的平衡问题。把这个观点再推进一步，就会得到有效势（effective potential）、科里奥利力（Coriolis force）（在这里，一旦 B 沿杆移动，它就由杆的反作用力抵销），以及平衡的稳定性：$x = \frac{4}{3}a$ 是否稳定？Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教材）正是以这个程度讨论非惯性参考系（non-inertial frames）和转动系统。

详细解答

STEP III 1991，第 15 题

概率与统计 · 样本中配对的计数

> **STEP III 1991，第 15 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/91-s3-q15

关键思路

每一组 k 张牌出现的机会均等，所以每个概率（probability）都等于符合条件的手牌数目除以 $\binom{2n}{k}$。要数出具有指定对子结构的手牌，可以分阶段构造。首先选定哪些种类以完整的一对出现，然后选定另外哪些种类只出现一张，最后决定每个单张种类的两张牌中，我们手上拿的是哪一张。这个计数一旦完成，题目的每一部分都只是同一条公式的特殊情况。

解答

把 $2n$ 张牌看成全部可以互相区分（同一种类的两张牌不妨编号为 1 和 2）。这样，不放回地抽出 k 张牌时，$\binom{2n}{k}$ 个大小为 k 的子集各有相同的概率。当某一种类的两张牌都被抽中时，我们说手牌中有该种类的一个对子（pair）。

四张牌，没有对子

没有对子的手牌包含四个不同的种类，每个种类各一张。选种类有 $\binom{n}{4}$ 种方法，每个种类选哪一张牌有 2 种方法，所以共有 $2^4\binom{n}{4}$ 手牌。由于

$$\binom{2n}{4} = \frac{2n(2n-1)(2n-2)(2n-3)}{24} = \frac{4n(n-1)(2n-1)(2n-3)}{24},$$

我们得到

$$\mathrm{P}(\text{没有对子}) = \frac{16 \cdot n(n-1)(n-2)(n-3)/24}{4n(n-1)(2n-1)(2n-3)/24} = \frac{4(n-2)(n-3)}{(2n-1)(2n-3)},$$

正如所求。（这里 $n \geq 4$ 保证了四个不同的种类确实存在。）

四张牌，恰好一个对子

先选出以对子出现的种类（n 种方法）。其余两张牌必须来自余下 $n-1$ 个种类中的两个不同种类，选法有 $\binom{n-1}{2}$ 种，而每个种类选哪一张牌又有 2 种选择。手牌的数目是

$$n \cdot \frac{(n-1)(n-2)}{2} \cdot 4 = 2n(n-1)(n-2),$$

所以

$$\mathrm{P}(\text{恰好一个对子}) = \frac{24 \cdot 2n(n-1)(n-2)}{4n(n-1)(2n-1)(2n-3)} = \frac{12(n-2)}{(2n-1)(2n-3)}.$$

k 张牌，恰好 r 个对子

假设手牌恰好有 r 个对子。它们占去 $2r$ 张牌。余下的 $k-2r$ 张牌全部属于不同的种类，而且这些种类都不是那 r 个成对的种类，否则对子就会更多。于是我们数：

- 成对的种类：$\binom{n}{r}$ 种方法；
- 只出现一张的 $k-2r$ 个种类，从其余 $n-r$ 个种类中选出：$\binom{n-r}{k-2r}$ 种方法；
- 每个单张种类的两张牌中，我们拿的是哪一张：2^{k-2r} 种方法。

每一手恰好有 r 个对子的牌，都恰好由一组这样的选择得出。因此

$$\boxed{\mathrm{P}(\text{恰好}r\text{ 个对子}) = \frac{2^{k-2r}\binom{n}{r}\binom{n-r}{k-2r}}{\binom{2n}{k}}}$$

想一想 这个概率会不会等于零？它与四张牌的答案是否吻合？

单张的种类需要满足 $k-2r \le n-r$，即 $r \ge k-n$。若 $k > n$，任意 k 张牌必定包含至少 $k-n$ 个对子，而 $\binom{n-r}{k-2r} = 0$ 自动处理了这个情况。取 $k=4$，$r=0$ 的情况给出 $2^4\binom{n}{4}/\binom{2n}{4}$，$r=1$ 则给出 $2^2\,n\binom{n-1}{2}/\binom{2n}{4}$，正好就是上面的两个计数。

k 为偶数，所有牌都成对

设 $k = 2m$。手牌由 m 个对子组成，当且仅当它是 m 个完整种类的并集（union）。这样的手牌有 $\binom{n}{m}$ 手（这就是一般公式取 $r = m$ 的情况）。所以

$$\mathrm{P} = \frac{\binom{n}{m}}{\binom{2n}{2m}} = \frac{n!}{m!\,(n-m)!} \cdot \frac{(2m)!\,(2n-2m)!}{(2n)!}.$$

要化简这个式子，可把阶乘（factorial）中的偶数因子和奇数因子分开归组：

$$(2j)! = \big(2 \cdot 4 \cdots 2j\big)\big(1 \cdot 3 \cdots (2j-1)\big) = 2^j\, j!\,\big(1 \cdot 3 \cdots (2j-1)\big).$$

记 $O_j = 1\cdot 3\cdots(2j-1)$，并设 $O_0 = 1$。那么 $(2j)!/j! = 2^j O_j$，所以

$$\mathrm{P} = \frac{2^m O_m \cdot 2^{n-m} O_{n-m}}{2^n O_n} = \frac{O_m\, O_{n-m}}{O_n}.$$

把 O_n 除以 O_{n-m}，剩下的是由 $2n-2m+1 = 2n-k+1$ 到 $2n-1$ 的奇数，所以

$$\mathrm{P}(\tfrac{1}{2}k \text{ 个对子}) = \frac{1\times 3\times 5\times\cdots\times(k-1)}{(2n-1)(2n-3)\cdots(2n-k+1)},$$

正如所求。

另外还有一个简洁的直接论证。把整副牌洗匀，以最上面的 k 张牌作为手牌。看看最顶的那张牌。它的拍档出现在其余 $2n-1$ 个位置中任何一个的机会均等，而其中 $k-1$ 个位置在手牌之内，所以拍档在手牌中的概率是 $\frac{k-1}{2n-1}$。如果它确实在手牌中，就把这一对拿走。剩下的是 $2n-2$ 张牌的一个随机排列，而手牌就是其中最上面的 $k-2$ 张。重复这个论证得到 $\frac{k-3}{2n-3}$，如此类推，直到 $\frac{1}{2n-k+1}$。把这些条件概率（conditional probability）相乘，便得到同一个乘积。

验算

- 四张牌时共有三种情况：0、1 和 2 个对子，而 2 个对子的概率是 $\binom{n}{2}/\binom{2n}{4} = \frac{3}{(2n-1)(2n-3)}$。果然

  $$4(n-2)(n-3) + 12(n-2) + 3 = 4n^2 - 8n + 3 = (2n-1)(2n-3),$$

  所以三个概率之和为 1。
- 取 $n=5$ 和四张牌，三个概率分别是 $\frac{8}{21}$、$\frac{4}{7}$ 和 $\frac{1}{21}$。用电脑对 $4\le n\le 8$ 以及由 3 到 $2n-1$ 的每个 k 穷举所有手牌，结果与一般公式及全部成对的公式均相符。
- 取 $k=2$，全部成对的公式给出 $\frac{1}{2n-1}$，即第二张牌与第一张牌配成一对的机会。

常见错误

- 用 $\binom{2n-2r}{k-2r}$ 从余下全部 $2n-2r$ 张牌中选出单张的牌。这样会容许其中两张属于同一种类，从而产生额外的对子。
- 遗漏因子 2^{k-2r}。每个只出现一张的种类，拿到的可以是它两张牌中的任何一张。

STEP 之外

恰好有 r 个对子的手牌数目，是典型的“先选结构，再填内容”式计数，而乘积 $1\cdot 3\cdots(2m-1)$（即双阶乘（double factorial）$(2m-1)!!$）数的是把 $2m$ 件物件分成若干对的方法数

目。双阶乘还会以标准正态分布（standard normal distribution）的偶数阶矩（moment）的身分再次出现。同样的配对问题，即随机样本中有多少个对子得以保存，也见于经典的“失袜”问题和生日类型的问题。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）为这类计数打下良好的基础。

详细解答

STEP III 1991，第 16 题

概率与统计 · 联合分布与边际分布

STEP III 1991，第 16 题

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

step.maths.org/questions/91-s3-q16

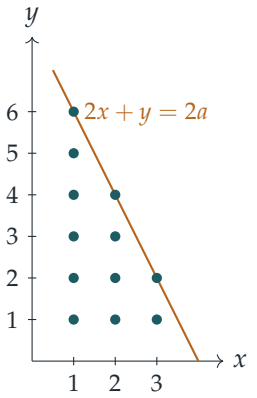
关键思路

这里的一切都是在一个由格点（lattice points）组成的三角形上的有限和。先把区域画出来：对每个 x，我们要知道哪些 y 是容许的；对每个 y，又有哪些 x 是容许的。一旦求和的上下限定对了，每个边际分布（marginal distribution）都是一个等差数列（arithmetic progression）之和。唯一要留神的地方，是对给定的 y，容许的最大 x 是 $\lfloor (2a-y)/2 \rfloor$，而这取决于 y 是偶数还是奇数。

解答

区域

我们需要 $x \geq 1$、$y \geq 1$ 和 $2x + y \leq 2a$。由于 $y \geq 1$，必须有 $2x \leq 2a - 1$，所以 $x \leq a - 1$。对于固定的 x（$1 \leq x \leq a-1$），容许的值是 $y = 1, 2, \ldots, 2a - 2x$。同样地，因为 $x \geq 1$，所以 $y \leq 2a - 2$。



图中所示为 $a = 4$ 的情况：第 x 列有 $2a - 2x$ 个点。

X 的边际分布与 c 的值

固定 x，并记 $n = 2a - 2x$。把等差数列 $2x + 1, \ldots, 2x + n$ 求和，

$$\begin{aligned} \mathrm{P}(X = x) &= c \sum_{y=1}^{n} (2x + y) = c \left(2xn + \frac{n(n+1)}{2} \right) = c \, \frac{n}{2} \, (4x + n + 1) \\ &= c \, (a - x)(2a + 2x + 1), \qquad x = 1, 2, \ldots, a - 1, \end{aligned}$$

而在其他情况下 $\mathrm{P}(X = x) = 0$。

所有概率之和必须等于 1。设 $k = a - x$，它由 1 取到 $a - 1$；于是 $2a + 2x + 1 = 4a + 1 - 2k$，并且

$$\begin{aligned} \sum_{k=1}^{a-1} k(4a + 1 - 2k) &= (4a + 1) \frac{(a-1)a}{2} - 2 \cdot \frac{(a-1)a(2a-1)}{6} \\ &= a(a-1) \left(\frac{4a+1}{2} - \frac{2a-1}{3} \right) = \frac{a(a-1)(8a+5)}{6}. \end{aligned}$$

令 c 乘以此和等于 1，便得

$$c = \frac{6}{a(a-1)(8a+5)},$$

正如所求。

Y 的边际分布

固定 y，其中 $1 \le y \le 2a - 2$。容许的 x 是 $1, 2, \ldots, N$，其中 N 是满足 $2N \le 2a - y$ 的最大整数。对这 N 项求和，

$$\sum_{x=1}^{N} (2x + y) = N(N + 1) + yN = N(N + 1 + y).$$

y **为偶数**。此时 $2a-y$ 是偶数，$N=\frac{1}{2}(2a-y)$，所以 $N+1+y=\frac{1}{2}(2a+2+y)$，并且

$$\mathrm{P}(Y=y)=c\cdot\frac{(2a-y)(2a+2+y)}{4}=\frac{3(2a-y)(2a+2+y)}{2a(a-1)(8a+5)},$$

其中用了 c 的值。这就是所要求的结果。

想一想 当 y 是奇数时，为甚么不能用 $N=\frac{1}{2}(2a-y)$?

当 y 是奇数时，$\frac{1}{2}(2a-y)$ 不是整数。满足 $2x\le 2a-y$ 的最大整数 x 是 $N=\frac{1}{2}(2a-y-1)$：向下取整把那半个舍去了。

y **为奇数**。取 $N=\frac{1}{2}(2a-1-y)$，便得 $N+1+y=\frac{1}{2}(2a+1+y)$，所以

$$\mathrm{P}(Y=y)=c\cdot\frac{(2a-1-y)(2a+1+y)}{4}=\frac{3(2a-1-y)(2a+1+y)}{2a(a-1)(8a+5)}.$$

（这两种情况可以写成 $3\big((2a+1)^2-(y+1)^2\big)$ 和 $3\big(4a^2-(y+1)^2\big)$ 除以同一个分母。）

Y 的期望值

把 $y\,\mathrm{P}(Y=y)$ 求和，就要把偶数和奇数的 y 分开处理。较简洁的做法是回到联合分布（joint distribution），逐列求和。对于固定的 x，同前面一样取 $n=2a-2x$，

$$\sum_{y=1}^{n}y(2x+y)=2x\cdot\frac{n(n+1)}{2}+\frac{n(n+1)(2n+1)}{6}=\frac{n(n+1)}{6}\big(6x+2n+1\big).$$

再次设 $k=a-x$，于是 $n=2k$，$6x+2n+1=6a-2k+1$。每列之和变成

$$\frac{2k(2k+1)(6a+1-2k)}{6}=\frac{k(2k+1)(6a+1-2k)}{3}.$$

展开得 $k(2k+1)(6a+1-2k)=(6a+1)(2k^2+k)-4k^3-2k^2$。利用求和至 $a-1$ 的标准公式，

$$\sum_{k=1}^{a-1}(2k^2+k)=\frac{a(a-1)(2a-1)}{3}+\frac{a(a-1)}{2}=\frac{a(a-1)(4a+1)}{6},$$

$$\sum_{k=1}^{a-1}k^3=\frac{a^2(a-1)^2}{4},\qquad\sum_{k=1}^{a-1}k^2=\frac{a(a-1)(2a-1)}{6}.$$

所以

$$\begin{aligned}&\sum_{k=1}^{a-1}k(2k+1)(6a+1-2k)\\&\quad=a(a-1)\left(\frac{(6a+1)(4a+1)}{6}-a(a-1)-\frac{2a-1}{3}\right)\\&\quad=\frac{a(a-1)}{6}\big(24a^2+10a+1-6a^2+6a-4a+2\big)\\&\quad=\frac{a(a-1)(6a^2+4a+1)}{2}.\end{aligned}$$

因此

$$\mathrm{E}(Y) = \frac{c}{3}\cdot\frac{a(a-1)(6a^2+4a+1)}{2} = \frac{6}{a(a-1)(8a+5)}\cdot\frac{a(a-1)(6a^2+4a+1)}{6}$$
$$= \frac{6a^2+4a+1}{8a+5}.$$

验算

- 当 $a = 2$ 时只有两个点，$(1,1)$ 和 $(1,2)$，权重分别为 $3c$ 和 $4c$，所以 $c = \frac{1}{7}$。这与 $6/(2\cdot 1\cdot 21)$ 一致。于是 $\mathrm{E}(Y) = \frac{3+8}{7} = \frac{11}{7}$，而公式给出 $\frac{33}{21} = \frac{11}{7}$。
- X 和 Y 的边际分布公式（奇偶两种情况）以及 $\mathrm{E}(Y)$，已对 a 由 2 至 29 的每个值用分数作精确枚举核对。例如，$a = 3$ 给出 $\mathrm{E}(Y) = \frac{67}{29}$。
- 当 a 很大时，$\mathrm{E}(Y) \approx \frac{3}{4}a$。权重 $2x + y$ 偏重靠近斜边的点，所以 Y 明显高于在三角形上取相等权重时所得的 $\frac{2}{3}a$。

常见错误

- 让 x 一直取到 a：$x = a$ 这一点会迫使 $y \le 0$。
- 对奇数 y 使用 $N = \frac{1}{2}(2a - y)$，而它并不是整数。
- 试图从分两种情况的边际分布求 $\mathrm{E}(Y)$。这样做是行得通的，但沿联合分布的各列求和，就可以完全避开奇偶之分。

STEP 之外

从联合分布求边际分布，就是把另一个变量求和消去，而难处总在于区域的上下限。对于连续变量，同样的工作变成在二重积分（double integral）中选取积分限，而交换求和次序（正如我们求 $\mathrm{E}(Y)$ 时所做的）则变成交换积分次序。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）涵盖了离散和连续变量的联合分布与边际分布。

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

1991 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。

