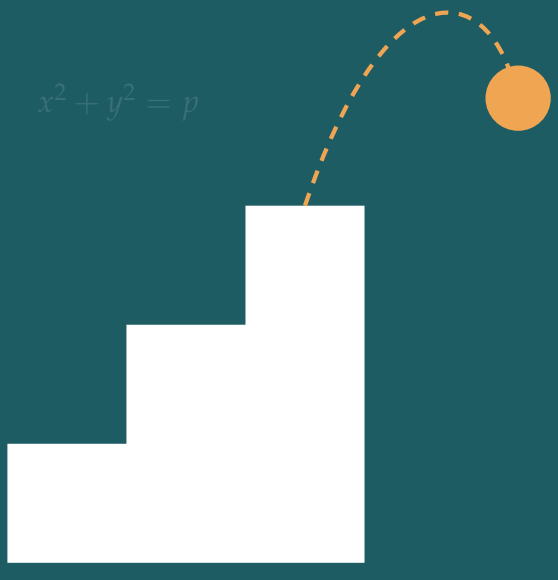
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1988 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

1988 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1988 年详细解答

1988 年 STEP 试卷（STEP I, II 及 III）全部 48 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 曲线描绘 纯数学 . . . 3

第 2 题 驻点 纯数学 . . . 6

第 3 题 几何极值的界限 纯数学 . . . 10

第 4 题 直线交点的计数 纯数学 . . . 13

第 5 题 级数与生成函数 纯数学 . . . 17

第 6 题 复数与几何 纯数学 . . . 20

第 7 题 二次函数导数的界限 纯数学 . . . 24

第 8 题 三维几何 纯数学 . . . 27

第 9 题 代换积分法 纯数学 . . . 30

第 10 题 抛体运动 力学 . . . 33

第 11 题 质心 力学 . . . 36

第 12 题 变力作用下的运动 力学 . . . 39

第 13 题 弹性绳与碰撞 力学 . . . 43

第 14 题 删失正态分布的期望值 概率与统计 . . . 47

第 15 题 二项分布的假设检验 概率与统计 . . . 51

第 16 题 连续分布与条件化 概率与统计 . . . 54

STEP II

第 1 题 部分分式与级数 纯数学 . . . 58

第 2 题 对称代数方程 纯数学 . . . 61

第 3 题 根的对称性 纯数学 . . . 64

第 4 题 阿根图上的轨迹 纯数学 . . . 68

第 5 题 单位根与正切 纯数学 . . . 73

第 6 题 双曲函数不等式 纯数学 . . . 77

第 7 题 代换积分法 纯数学 . . . 81
第 8 题 矩阵与特征向量 纯数学 . . . 85
第 9 题 子群与并集 纯数学 . . . 89
第 10 题 抛物面的向量几何 纯数学 . . . 92
第 11 题 绳缠绕柱子 力学 . . . 96
第 12 题 下落的链条 力学 . . . 100
第 13 题 摩擦力与平衡 力学 . . . 103
第 14 题 转动的刚体 力学 . . . 106
第 15 题 离散均匀分布变量之和 概率与统计 . . . 110
第 16 题 正态分布 概率与统计 . . . 114

STEP III

第 1 题 曲线描绘与界限 纯数学 . . . 117
第 2 题 线性递推关系 纯数学 . . . 121
第 3 题 复数轨迹 纯数学 . . . 124
第 4 题 弧长的最优化 纯数学 . . . 128
第 5 题 递推关系与贷款 纯数学 . . . 131
第 6 题 逐次微分 纯数学 . . . 134
第 7 题 微分方程与正交性 纯数学 . . . 137
第 8 题 抛物线的切线与法线 纯数学 . . . 141
第 9 题 群元素的阶 纯数学 . . . 146
第 10 题 追踪曲线 纯数学 . . . 149
第 11 题 涉及摩擦力的梯子平衡 力学 . . . 153
第 12 题 斜碰撞 力学 . . . 156
第 13 题 有空气阻力的抛体 力学 . . . 160
第 14 题 光滑金属线上的运动 力学 . . . 164
第 15 题 期望值的递推关系 概率与统计 . . . 168
第 16 题 超几何分布 概率与统计 . . . 172

详细解答

STEP I 1988，第 1 题

纯数学 · 曲线描绘

> **STEP I 1988，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/88-s1-q1

关键思路

方程 $n^m = m^n$ 把两个变量分别放在底数和指数的位置，纠缠在一起，处理起来并不方便。取对数（logarithm）后得到 $m\ln n = n\ln m$，再除以 mn 便可把两者完全分开：$\frac{\ln n}{n} = \frac{\ln m}{m}$。因此我们要找两个不同的整数，使 h 在这两处取相同的值，而图像的形状正好告诉我们这样的数对可以出现在哪里。

解答

h 的图像

用商法则（quotient rule）求导，

$$\mathrm{h}'(x) = \frac{x\cdot\frac{1}{x} - \ln x}{x^2} = \frac{1-\ln x}{x^2}.$$

由于 $x^2 > 0$，$\mathrm{h}'(x)$ 的正负号与 $1-\ln x$ 的正负号相同。所以 h 在 $(0,\mathrm{e})$ 上严格递增（strictly increasing），在 (e,∞) 上严格递减（strictly decreasing），并在 $x=\mathrm{e}$ 处取得极大值（maximum），其值为

$$\mathrm{h}(\mathrm{e}) = \frac{1}{\mathrm{e}}.$$

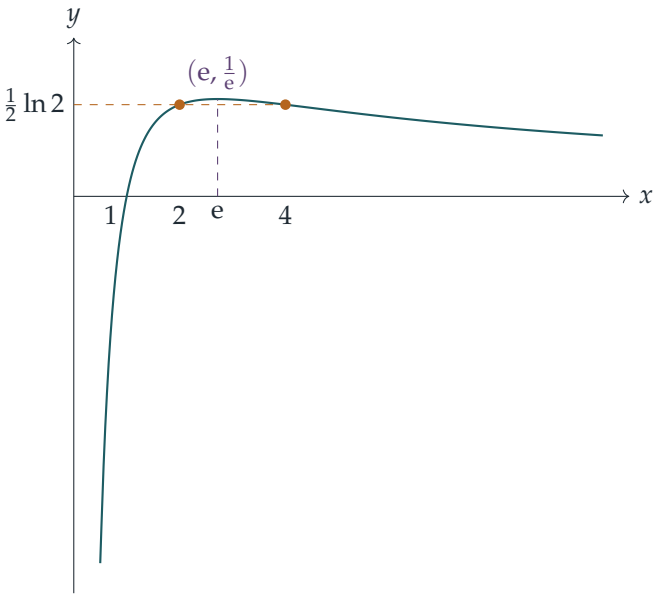
再求导一次，

$$\mathrm{h}''(x) = \frac{-\frac{1}{x}\cdot x^2 - (1-\ln x)\cdot 2x}{x^4} = \frac{2\ln x - 3}{x^3},$$

所以在 $x = \mathrm{e}^{3/2} \approx 4.48$ 处有一个拐点（point of inflection），该处 $\mathrm{h} = \frac{3}{2}\mathrm{e}^{-3/2} \approx 0.335$。曲线在这一点的左方是凹（concave）的，在右方是凸（convex）的。

接着看两端的变化趋势和函数的正负：

- 当 $x \to 0^+$ 时，$\ln x \to -\infty$ 而 $\frac{1}{x} \to +\infty$，所以 $\mathrm{h}(x) \to -\infty$：$y$-轴是一条渐近线（asymptote）。
- $\mathrm{h}(1) = 0$，而 $\mathrm{h}(x)$ 的正负号与 $\ln x$ 相同：在 $(0,1)$ 上为负，当 $x > 1$ 时为正。
- 当 $x \to \infty$ 时，$\frac{\ln x}{x} \to 0$（对数的增长比 x 的任何正数次幂都要慢），而且该处 $\mathrm{h}(x) > 0$，所以曲线从上方趋近 x-轴。



方程 $n^m = m^n$

由于 m 和 n 都是正数，等式两边均为正，可以取对数：

$$n^m = m^n \iff m\ln n = n\ln m \iff \frac{\ln n}{n} = \frac{\ln m}{m} \iff \mathrm{h}(n) = \mathrm{h}(m),$$

其中中间一步除以了 $mn > 0$。所以我们要找满足 $\mathrm{h}(m) = \mathrm{h}(n)$ 的相异正整数。由对称性，不妨设 $m < n$。

想一想 图像上哪些位置可以有两个不同的点处于同一高度?

由于 h 在 $(0,\mathrm{e}]$ 上严格递增，在 $[\mathrm{e},\infty)$ 上严格递减，它在这两个区间上各自都是一对一（one-to-one）的。所以，输出相同的两个相异输入必定分处 e 的两侧：$m < \mathrm{e} < n$。由于 m 是正整数而 $\mathrm{e} \approx 2.718$，故 $m = 1$ 或 $m = 2$。

- 若 $m = 1$，则 $\mathrm{h}(n) = \mathrm{h}(1) = 0$，由此必有 $\ln n = 0$，即 $n = 1 = m$，不合要求。（直接来看：$n^1 = 1^n$ 给出 $n = 1$。）

- 若 $m = 2$，则需要一个整数 $n \geq 3$ 使得 $\mathrm{h}(n) = \frac{1}{2}\ln 2$。而 $n = 4$ 正好符合，因为

$$\mathrm{h}(4) = \frac{\ln 4}{4} = \frac{2\ln 2}{4} = \frac{\ln 2}{2} = \mathrm{h}(2),$$

  也就是 $4^2 = 16 = 2^4$。由于 h 在 $[\mathrm{e}, \infty)$ 上严格递减，它在该处取 $\frac{1}{2}\ln 2$ 这个值至多一次，所以 $n = 4$ 是唯一的可能。

因此，满足 $n^m = m^n$ 的相异正整数对只有

$$(m, n) = (2, 4) \quad \text{及} \quad (m, n) = (4, 2).$$

验算

用电脑以精确的整数运算搜索所有 $1 \leq m, n \leq 300$，只找到 $(2,4)$ 和 $(4,2)$。数值上，$\mathrm{h}(2) = \mathrm{h}(4) = 0.346574\ldots$，仅略低于极大值 $1/\mathrm{e} = 0.367879\ldots$，与图像所示吻合。你也可以看出为甚么 $\mathrm{h}(3)$ 得不出任何解：$\mathrm{h}(3) \approx 0.366$ 大于 $\mathrm{h}(2)$，它在 e 左方的对应点严格介乎 2 与 e 之间，所以不是整数。

常见错误

- 靠试数找到 $(2,4)$ 便就此停笔。题目要求找出*所有*数对，所以用单调性（monotonicity）论证再没有其他数对，才是答案的核心。
- 画草图时忘记渐近性态：曲线在 y-轴附近直坠至 $-\infty$，但当 $x \to \infty$ 时只是缓缓趋向 0，此后不再穿过横轴。

STEP 之外

如果容许实数，则 $(1, \mathrm{e})$ 内的每个 a 都恰好有一个对应的 $b > \mathrm{e}$ 使得 $a^b = b^a$，而这些数对可以简洁地参数化（parametrise）为 $a = (1 + \frac{1}{t})^t$，$b = (1 + \frac{1}{t})^{t+1}$，其中 $t > 0$；当 $t \to \infty$ 时两者都趋向 e，也就是图像顶端两边相会之处。我们用过的极限 $\frac{\ln x}{x} \to 0$ 是比较增长率的第一个例子，大学的第一门分析课程会把它严格化；参见 Abbott, *Understanding Analysis*（一本分析学入门教科书）。

详细解答

STEP I 1988，第 2 题

纯数学 · 驻点

> **STEP I 1988，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/88-s1-q2

关键思路

把 g 看成一个复合函数（composition）：$\mathrm{g} = h \circ \mathrm{f}$，其中 $h(t) = t + 1/t$。由链式法则（chain rule）可知，g' 只会在 $\mathrm{f}' = 0$ 或 $h'(\mathrm{f}) = 0$ 之处为零，而后者即 $\mathrm{f} = \pm 1$。把 $\cos 2x$ 用 $\sin x$ 表示，f 便成为 $\sin x$ 的二次式（quadratic），其大小和形状由此一目了然。

解答

导数

用链式法则对 $\mathrm{g} = \mathrm{f} + \mathrm{f}^{-1}$ 求导，

$$\mathrm{g}'(x) = \mathrm{f}'(x) - \frac{\mathrm{f}'(x)}{\mathrm{f}(x)^2} = \mathrm{f}'(x)\left(1 - \frac{1}{\mathrm{f}(x)^2}\right).$$

再用积法则（product rule）求导一次，并利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(-\mathrm{f}^{-2}\right) = 2\mathrm{f}^{-3}\mathrm{f}'$，

$$\mathrm{g}''(x) = \mathrm{f}''(x)\left(1 - \frac{1}{\mathrm{f}(x)^2}\right) + \frac{2\,\mathrm{f}'(x)^2}{\mathrm{f}(x)^3}.$$

f 的大小

利用 $\cos 2x = 1 - 2\sin^2 x$，并设 $s = \sin x$，

$$\mathrm{f}(x) = 5 + 2s - 2s^2 = \tfrac{11}{2} - 2\left(s - \tfrac{1}{2}\right)^2.$$

当 s 取遍 $[-1,1]$ 时，此式在 $s=\frac{1}{2}$ 处最大（为 $\frac{11}{2}$），而在区间中离 $\frac{1}{2}$ 最远的端点，即 $s=-1$ 处最小，其值为 1。因此

$$1 \le \mathrm{f}(x) \le \tfrac{11}{2} \quad \text{对所有} x,$$

而且只有当 $\sin x=-1$ 时才有 $\mathrm{f}(x)=1$。特别地，f 永不为零，所以 g 处处有定义。

驻点

$\mathrm{g}'(x)=0$ 当且仅当 $\mathrm{f}'(x)=0$ 或 $\mathrm{f}(x)^2=1$；满足此条件的点就是驻点（stationary point）。

f′ 的零点。我们有

$$\mathrm{f}'(x) = -2\sin 2x + 2\cos x = 2\cos x\,(1-2\sin x),$$

它在 $\cos x=0$ 或 $\sin x=\frac{1}{2}$ 时为零；在 $[0,2\pi]$ 内即

$$x = \tfrac{\pi}{6},\ \tfrac{\pi}{2},\ \tfrac{5\pi}{6},\ \tfrac{3\pi}{2}.$$

$\mathrm{f}=\pm 1$ 的点。由于 $\mathrm{f}\ge 1$，这即是 $\mathrm{f}=1$，亦即 $\sin x=-1$，亦即 $x=\frac{3\pi}{2}$，而这一点已在上面的列表之中。

所以 g 在 $[0,2\pi]$ 内恰好有四个驻点。（两个端点不在其中：$\mathrm{g}'(0)=\mathrm{g}'(2\pi)=2\cdot\frac{24}{25}\neq 0$。）

驻点的分类

我们有 $\mathrm{f}''(x) = -4\cos 2x - 2\sin x$。在每一个驻点都有 $\mathrm{f}'=0$，所以 g'' 的第二项消失，得 $\mathrm{g}''=\mathrm{f}''\,(1-\mathrm{f}^{-2})$。

- $x=\frac{\pi}{6}$ 和 $x=\frac{5\pi}{6}$：此时 $\sin x=\frac{1}{2}$，$\cos 2x=\frac{1}{2}$，所以 $\mathrm{f}=\frac{11}{2}$ 且 $\mathrm{f}''=-2-1=-3$。于是 $\mathrm{g}''=-3\big(1-\frac{4}{121}\big)<0$：两点都是**极大点**（maxima），并且

$$\mathrm{g} = \tfrac{11}{2} + \tfrac{2}{11} = \tfrac{125}{22}.$$

- $x=\frac{\pi}{2}$：此时 $\sin x=1$，$\cos 2x=-1$，所以 $\mathrm{f}=5$ 且 $\mathrm{f}''=4-2=2$。于是 $\mathrm{g}''=2\big(1-\frac{1}{25}\big)=\frac{48}{25}>0$：这是**极小点**（minimum），并且 $\mathrm{g}=5+\frac{1}{5}=\frac{26}{5}$。
- $x=\frac{3\pi}{2}$：此时 $\mathrm{f}=1$，所以 $1-\mathrm{f}^{-2}=0$，同时 $\mathrm{f}'=0$。因此 $\mathrm{g}''=0$。

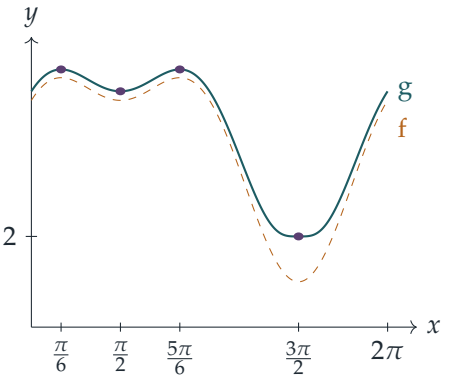
> **想一想** 二阶导数判别法（second-derivative test）在 $x=\frac{3\pi}{2}$ 处失效。还有甚么方法可以判断？

回到 $h(t)=t+1/t$。对于 $t>0$，

$$t+\frac{1}{t}-2=\frac{(t-1)^2}{t}\ge 0,$$

等号只在 $t = 1$ 时成立。由于 $\mathrm{f} > 0$，由此可得对每一个 x 都有 $\mathrm{g}(x) \geq 2$，而等号恰好在 $\mathrm{f}(x) = 1$ 时成立。在 $x = \frac{3\pi}{2}$ 处有 $\mathrm{g} = 2$，所以这是**极小点**（事实上是 g 的全域最小值（global minimum））。

总结。极大点在 $x = \frac{\pi}{6}, \frac{5\pi}{6}$（值为 $\frac{125}{22}$）；极小点在 $x = \frac{\pi}{2}$（值为 $\frac{26}{5}$）和 $x = \frac{3\pi}{2}$（值为 2）。



验算

- 由于 h 在 $t > 1$ 时递增，凡是 $\mathrm{f} > 1$ 之处，g 都应该随 f 一同升降；图像证实 f（虚线）和 g 有相同的转向点（turning point）。
- 数值上，$\mathrm{g}(\frac{\pi}{6}) \approx 5.6818 = \frac{125}{22}$ 而 $\mathrm{g}(\frac{\pi}{2}) = 5.2$；在 $[0, 2\pi]$ 上对 g 作细密扫描，只在 $\frac{\pi}{6}, \frac{\pi}{2}, \frac{5\pi}{6}, \frac{3\pi}{2}$ 找到转向点，最小值为 2.000。
- g'' 的公式与在随机选取的点上以有限差分（finite difference）计算的二阶导数吻合。

常见错误

- 忘记因子 $1 - \mathrm{f}^{-2}$：在本题中它没有带来新的驻点，但这只是因为 $\mathrm{f} = 1$ 恰好发生在 $\mathrm{f}' = 0$ 之处。你仍然必须检查这个因子。
- 由 $\mathrm{g}''(\frac{3\pi}{2}) = 0$ 就断定该点是拐点（point of inflection）。二阶导数为零，单凭这一点并不能说明甚么。

STEP 之外

$x = \frac{3\pi}{2}$ 这个退化点在复合函数中十分典型：如果 f 取极小值之处恰好也是 h 取极小值之处，$h \circ \mathrm{f}$ 在该处就会“更平坦”，二阶导数亦随之为零。大学的数学分析（analysis）以泰勒定理（Taylor's theorem）取代二阶导数判别法，按第一个非零的导数来为这类点分

类，或者（像我们所做的那样）直接用不等式来判断。Abbott 的 *Understanding Analysis* 严谨地建立中值定理（mean value theorem）和泰勒定理，是了解这些内容的好去处。

详细解答

STEP I 1988，第 3 题

纯数学 · 几何极值的界限

> **STEP I 1988，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/88-s1-q3

关键思路

先问一个问题：正方形内的一点若离 O 很远，它可以在哪里？它必定被挤进对角附近的一小块区域。如果 P 和 Q 都离 O 很远，两点便同在这一小块区域内，因此彼此相距很近。$\sqrt{6}-\sqrt{2}$ 这个数的大小恰好令这个论证成立，而它来自一个对称地放在正方形内的等边三角形（equilateral triangle）。

解答

以厘米为单位，取 O 为原点，正方形为 $0 \le x \le 1$、$0 \le y \le 1$。记

$$s = \sqrt{6}-\sqrt{2}, \qquad s^2 = 6+2-2\sqrt{12} = 8-4\sqrt{3} \approx 1.072,$$

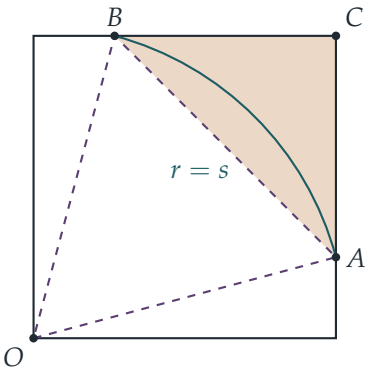
并设 $t = 2-\sqrt{3} \approx 0.268$。定义以下各点：

$$A = (1,t), \qquad B = (t,1), \qquad C = (1,1).$$

于是

$$\begin{aligned} OA^2 = OB^2 &= 1+(2-\sqrt{3})^2 = 8-4\sqrt{3} = s^2, \\ AB^2 &= 2(1-t)^2 = 2(\sqrt{3}-1)^2 = 8-4\sqrt{3} = s^2, \\ AC = BC &= 1-t = \sqrt{3}-1 \approx 0.732 < s. \end{aligned}$$

所以 OAB 是正方形内一个边长为 s 的等边三角形。



蓝绿色的弧是以 O 为圆心、半径为 s 的圆；正方形在这条弧以外的部分，全在涂上阴影的三角形 ABC 之内。

第 1 步：离 O 远的点都在三角形 ABC 内

直线 AB 的方程是 $x+y=1+t$。考虑正方形在这条直线靠 O 一侧的部分，即 $\{x+y\leq 1+t\}$。这是一个五边形，顶点为

$$O=(0,0),\quad (1,0),\quad A,\quad B,\quad (0,1).$$

函数 x^2+y^2 是凸函数（convex function），所以它在这个凸五边形上的最大值必在某个顶点取得。它在各顶点的值依次是 0、1、s^2、s^2、1，其中最大的是 s^2。因此，对正方形内的每一点，

$$x+y\leq 1+t \implies x^2+y^2\leq s^2$$

都成立。反过来说：如果正方形内的一点 (x,y) 与 O 的距离大于 s，则 $x+y>1+t$；又因为 $x\leq 1$ 及 $y\leq 1$，这一点必在三角形 ABC 内。

? **想一想** 为甚么三角形内两点之间的最大距离，必定是它其中一条边的长度?

固定 Q。距离 $|PQ|$ 是 P 的凸函数，所以在三角形上，当 P 是顶点时它取得最大值。现在固定这个顶点，让 Q 变动：根据同样的论证，Q 也可以取为顶点。因此，三角形内两点之间的最大距离就是某两个顶点之间的距离，也就是它最长的边。

第 2 步：导出矛盾

假设 $OP>s$ 而且 $OQ>s$；否则结论已经成立。根据第 1 步，P 和 Q 都在三角形 ABC 内。这个三角形最长的边是 $AB=s$（另外两边的长度是 $\sqrt{3}-1<s$）。所以根据上面方框内的结果，

$$PQ\leq s.$$

由此可见，无论在哪一种情况，OP、PQ、QO 之中至少有一个不超过 $s = (\sqrt{6} - \sqrt{2})$ 厘米，这正是要证明的结论。

这个界已是最佳

取 $P = A$ 及 $Q = B$，便恰好有 $OP = PQ = QO = s$，所以任何更小的数都不行。由于 $\cos 15^\circ = \frac{\sqrt{6}+\sqrt{2}}{4}$，我们有 $s = \sec 15^\circ$：极端情况的三角形就是关于对角线对称放置的等边三角形，其中 OA 和 OB 各与正方形的一条边成 15° 角。

验算

- $s \approx 1.0353$，比正方形的边长稍大，又远小于对角线的长度 $\sqrt{2} \approx 1.414$，这是合理的。
- 数值上，在正方形内随机抽取 10^6 对 P, Q，$\min(OP, PQ, QO)$ 从未超过 s（最大的约为 1.0235）。以爬山法（hill-climbing）搜索这个最小值的最大可能值，结果收敛到 $P \approx (1, 0.26795)$、$Q \approx (0.26795, 1)$，其值为 1.035276，与 A、B 和 s 吻合。

常见错误

- 只论证了当 $OP, OQ > s$ 时 $\angle POQ < 60^\circ$。这是对的，但单凭这一点并不能令 PQ 成为最短的边，所以若不再多下工夫，证明仍未完成。
- 只检查对称的情况。这只说明这个界不能再改进；它并没有证明这个界对所有 P 和 Q 都成立。

STEP 之外

这是一道关于点的装填的小型极值问题（extremal problem）：一个区域内的三点，可以同时彼此相距多远？这类问题会引向装填（packing）与覆盖（covering）问题，也引向集合的*直径*（diameter）这个概念：一个集合的直径与它的凸包（convex hull）的直径相同。正是这个事实令第 2 步变得轻而易举。想做更多同类问题，可参看 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP I 1988，第 4 题

纯数学 · 直线交点的计数

> **STEP I 1988，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/88-s1-q4

关键思路

不要试图一次过想像所有线段。不如改问：两条线段（segment）何时会相交？答案只取决于它们的端点（endpoint）在两条轴上的排列次序，而每一个交叉点因此都对应于在每条轴上各选两点。只要能数出单独一条线段上的交叉点，那些极端的线段以及最后的平均值，只需小心计数便可得出。

解答

把 y 轴上的点按与原点 O 的距离由小至大依次记为 $B_1, \ldots, B_m$，x 轴上的点亦同样记为 $A_1, \ldots, A_n$。以 s_{ij} 表示线段 B_iA_j。

第 (i) 部分

m 个点 B_i 中的每一个都与 n 个点 A_j 中的每一个相连，所以共有

$$mn$$

条线段。

第 (ii) 部分

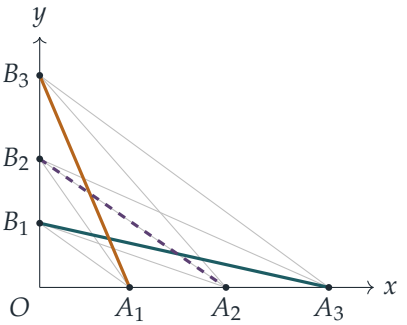
取两条不同的线段 s_{ij} 和 s_{kl}。

- 若 $i = k$ 或 $j = l$，它们有一个公共端点。由同一点出发的两条相异线段只会在该点相遇，而这类相遇点不计算在内。
- 若 $i < k$ 且 $j < l$，则 B_i 和 A_j 都与 O 位于直线 B_kA_l 的同一侧（它们在三角形 OB_kA_l

的两条边 OB_k 和 OA_l 上，而不在 B_k 或 A_l 处）。若一条线段的两个端点都严格位于某直线的同一侧，整条线段便留在该侧，所以 s_{ij} 永不与 s_{kl} 相遇：它“嵌套”在下方。

- 若 $i < k$ 且 $j > l$，则 B_i 与 O 位于直线 B_kA_l 的同一侧，而 A_j 在另一侧。所以 s_{ij} 穿过该直线。同理，A_l 位于直线 B_iA_j 靠 O 的一侧，而 B_k 在另一侧。每条线段都把另一条线段的两个端点分隔开，所以这两条线段恰好交于一点，而该点并非端点。

$i > k$ 的情况相同，只是角色互换。



在图中（$m = n = 3$），蓝绿色的 s_{13} 与琥珀色的 s_{31} 的端点次序相反，所以它们相交。虚线 s_{22} 亦与它们各自相交。

因此，每一对交于内点的线段，都来自选取两点 B_i, B_k（$i < k$）和两点 A_j, A_l（$j < l$），然后取 s_{il} 和 s_{kj}。反过来，每一个这样的选择都恰好给出一对相交的线段。由于没有三条线段通过同一个内点，不同的线段对给出不同的点。因此交点（point of intersection）的数目为

$$\binom{m}{2}\binom{n}{2} = \frac{m(m-1)n(n-1)}{4}.$$

平均值

一条线段上的交叉点。根据上述法则，s_{ij} 与 s_{kl} 相交，当且仅当 $k < i$, $l > j$ 或 $k > i$, $l < j$。所以 s_{ij} 上的交点数目为

$$c_{ij} = (i-1)(n-j) + (m-i)(j-1).$$

当 $m = n$ 时的极端线段。设 $N = n - 1$，$a = i - 1$，$b = j - 1$，则 $0 \le a, b \le N$，且

$$c_{ij} = a(N-b) + (N-a)b = N(a+b) - 2ab = \tfrac{1}{2}\big(N^2 - (N-2a)(N-2b)\big),$$

这可以展开验证。由于 $|N - 2a| \le N$ 及 $|N - 2b| \le N$，乘积 $(N-2a)(N-2b)$ 介乎 $-N^2$ 与 N^2 之间。因此 $0 \le c_{ij} \le N^2$，并且：

- $c_{ij} = N^2$ 仅当 $(N-2a)(N-2b) = -N^2$，即 $\{a,b\} = \{0,N\}$。这就是 s_{1n} 和 s_{n1} 两条线段。
- $c_{ij} = 0$ 仅当 $(N-2a)(N-2b) = N^2$，即 $a = b = 0$ 或 $a = b = N$。这就是 s_{11} 和 s_{nn} 两条线段。

（这里 $N \geq 2$，所以四条线段互不相同。）所以恰好有两条线段上的点数最多，为 $(n-1)^2$，亦恰好有两条线段上的点数最少，为 0。移除它们后，剩下 n^2-4 条线段。

想一想 究竟是对甚么取平均?

我们数的是移除那四条线段之后仍然存在的交点，即两条剩余线段之间的交叉点，每点只数一次。然后除以剩余线段的数目。（如果你改为把每条剩余线段上的交叉点加起来，每个点都会被数两次，而且一些已失去的点仍会被计入。这样会得出 $\frac{1}{2}(n-1)^2$，并非题目所给的答案。见“常见错误”。）

失去的点。线段 s_{11} 和 s_{nn} 上没有交点，所以移除它们不会失去任何点。线段 s_{1n} 和 s_{n1} 上各有 $(n-1)^2$ 个点。它们互相相交（$1 < n$ 且 $n > 1$），所以恰好有一点同时位于两者之上。根据容斥原理（inclusion–exclusion），失去的点数为 $2(n-1)^2 - 1$。

剩余的点。由第 (ii) 部分取 $m = n$，剩余的点数为

$$\begin{aligned}\frac{n^2(n-1)^2}{4} - 2(n-1)^2 + 1 &= \frac{(n-1)^2(n^2-8)+4}{4} \\ &= \frac{n^4 - 2n^3 - 7n^2 + 16n - 4}{4} \\ &= \frac{(n-2)(n^3 - 7n + 2)}{4}.\end{aligned}$$

至于因式分解（factorisation），注意 $n = 2$ 时该四次多项式为 $16 - 16 - 28 + 32 - 4 = 0$，所以 $n-2$ 是一个因式。把 $(n-2)(n^3-7n+2) = n^4 - 2n^3 - 7n^2 + 16n - 4$ 乘开即可确认。

平均值。除以剩余的 $n^2 - 4 = (n-2)(n+2)$ 条线段，并约去 $n - 2 \neq 0$（因为 $n \geq 3$），得平均值为

$$\frac{(n-2)(n^3-7n+2)}{4(n-2)(n+2)} = \frac{n^3-7n+2}{4(n+2)},$$

正如所求。

验算

当 $n = 3$ 时，公式给出 $\frac{8}{20} = \frac{2}{5}$。直接计算：移除 $s_{11}, s_{33}, s_{13}, s_{31}$ 后，剩下 $s_{12}, s_{21}, s_{22}, s_{23}, s_{32}$。其中只有 s_{12} 与 s_{21}，以及 s_{23} 与 s_{32}，端点次序相反，所以 5 条线段上共有 2 个点，得出

$\frac{2}{5}$。我们亦用电脑验算：在两条轴上随机放置点，以精确算术找出每一个交叉点，确认了 mn、$\binom{m}{2}\binom{n}{2}$ 和 c_{ij}。此外亦确认了平均值公式对 3 至 29 的每一个 n 都成立。

常见错误

- 把每条剩余线段上的交叉点加起来。每个点都位于两条线段上，所以这样会重复计算，而且仍然包括了与被移除线段共有的点。结果是 $\frac{1}{2}(n-1)^2$，而非题目所给的公式。
- 忘记 s_{1n} 和 s_{n1} 互相相交，因而减去 $2(n-1)^2$ 而非 $2(n-1)^2-1$。
- 把交叉点数成 $\binom{mn}{2}$ 对线段：大部分线段对不是共用一个端点，就是互相嵌套，根本不会相交。

STEP 之外

本题的核心是一个双射（bijection）：交叉点与在每条轴上各选两点的方法一一对应。这是组合数学（combinatorics）的基本工具。同一幅图亦出现于完全二部图（complete bipartite graph）$K_{m,n}$ 的“交叉数”（crossing number）之中。Zarankiewicz 的画法（正是这一种，只是每条轴的两侧都有点）用了 $\lfloor\frac{m}{2}\rfloor\lfloor\frac{m-1}{2}\rfloor\lfloor\frac{n}{2}\rfloor\lfloor\frac{n-1}{2}\rfloor$ 个交叉点，而这是否总是最少的可能数目，至今仍是未解决的问题（open problem）。想多做以双射计数的题目，可看 Siklos 的 *Advanced Problems in Mathematics*；至于图论（graph theory），可看 Wilson 的 *Introduction to Graph Theory*。

详细解答

STEP I 1988，第 5 题

纯数学 · 级数与生成函数

> **STEP I 1988，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/88-s1-q5

关键思路

把每个因式展开成等比级数（geometric series）（即指数为 -1 的二项式定理（binomial theorem）），然后相乘：c_m 就是一个有限等比级数之和，有简洁的闭合形式（closed form）。只要把 c_m^2 写成 $(a^2)^m$、$(ab)^m$ 和 $(b^2)^m$ 这三个纯幂的组合，最后的级数便会分拆成三个等比级数。余下的工作，就是小心处理分式的代数运算，并留意每个级数在哪里收敛（converge）。

解答

求 c_m

根据指数为 -1 的二项式定理，

$$(1-ax)^{-1}=\sum_{k=0}^{\infty}a^kx^k\quad(|ax|<1),\qquad(1-bx)^{-1}=\sum_{j=0}^{\infty}b^jx^j\quad(|bx|<1).$$

由于 $0<a<b$，条件 $b|x|<1$ 给出 $a|x|<b|x|<1$，所以两个展开式都成立。两个级数都绝对收敛（converge absolutely），因此我们可以把它们相乘，再按 x 的幂次合并各项。x^m 的系数来自满足 $k+j=m$ 的各对指数：

$$c_m=\sum_{k=0}^{m}a^kb^{m-k}=b^m+ab^{m-1}+\cdots+a^{m-1}b+a^m.$$

这是一个等比级数，首项为 b^m，公比为 a/b，共有 $m+1$ 项。由于 $a\neq b$，它的和是

$$c_m=\frac{b^{m+1}-a^{m+1}}{b-a}.$$

（等价地，$(b-a)c_m=b^{m+1}-a^{m+1}$，因为把乘积展开后，各项会逐一相消（telescope）。）

c_m^2 的公式

两边平方，并利用 $(b-a)^2=(a-b)^2$，得

$$c_m^2=\frac{\left(b^{m+1}-a^{m+1}\right)^2}{(b-a)^2}=\frac{a^{2m+2}-2(ab)^{m+1}+b^{2m+2}}{(a-b)^2},$$

这正是所要求的结果。

级数 $\sum c_m^2 x^m$

上一条公式的要点，在于每一项都是一个常数乘以一个 m 次幂：

$$c_m^2x^m=\frac{1}{(a-b)^2}\left(a^2\,(a^2x)^m-2ab\,(abx)^m+b^2\,(b^2x)^m\right).$$

所以 $\sum c_m^2x^m$ 是三个等比级数的组合，公比分别为 a^2x、abx 和 b^2x。

> **想一想** 对哪些 x，这三个等比级数全部收敛?

我们需要 $|a^2x|<1$、$|abx|<1$ 和 $|b^2x|<1$。由于 $0<a^2<ab<b^2$，最后一个条件最强，并且蕴涵另外两个。所以三个级数全部收敛，当且仅当 $|x|<1/b^2$。这时把每个等比级数求和，得

$$\sum_{m=0}^{\infty}c_m^2x^m=\frac{1}{(a-b)^2}\left(\frac{a^2}{1-a^2x}-\frac{2ab}{1-abx}+\frac{b^2}{1-b^2x}\right).$$

把所有项通分，公分母为 $(1-abx)(1-a^2x)(1-b^2x)$。分子是

$$N=a^2(1-abx)(1-b^2x)-2ab(1-a^2x)(1-b^2x)+b^2(1-a^2x)(1-abx).$$

我们按 x 的幂次整理：

$$\begin{aligned}
\text{常数项：}\quad & a^2-2ab+b^2=(a-b)^2,\\
\text{系数，对应}x\text{：}\quad & -a^2(ab+b^2)+2ab(a^2+b^2)-b^2(a^2+ab)\\
& \qquad =a^3b-2a^2b^2+ab^3=ab(a-b)^2,\\
\text{系数，对应}x^2\text{：}\quad & a^2\cdot ab^3-2ab\cdot a^2b^2+b^2\cdot a^3b=0.
\end{aligned}$$

所以 $N=(a-b)^2(1+abx)$。因式 $(a-b)^2$ 可以约去，余下

$$c_0^2+c_1^2x+c_2^2x^2+\cdots=\frac{1+abx}{(1-abx)(1-a^2x)(1-b^2x)}\qquad\text{适用于}-\frac{1}{b^2}<x<\frac{1}{b^2}.$$

这个区间不能再扩大。在 $x=\pm1/b^2$ 处，各项满足

$$|c_m^2x^m|=\frac{\left(b^{m+1}-a^{m+1}\right)^2}{(b-a)^2\,b^{2m}}=\frac{b^2\left(1-(a/b)^{m+1}\right)^2}{(b-a)^2}\to\frac{b^2}{(b-a)^2}\neq 0,$$

所以级数在该处发散（diverge）；当 $|x|>1/b^2$ 时级数同样发散，因为这时各项会无限增大。

验算

取 $a=1, b=2$。这时 $c_m = 2^{m+1}-1$，所以 $c_0, c_1, c_2 = 1, 3, 7$，而直接计算可得 $\frac{1}{(1-x)(1-2x)} = (1+x+x^2+\dots)(1+2x+4x^2+\dots) = 1+3x+7x^2+\cdots$。它们的平方是 $1, 9, 49$。另一方面，$\frac{1}{(1-x)(1-2x)(1-4x)} = 1+7x+35x^2+\cdots$，再乘以 $1+2x$ 便得到 $1+9x+49x^2+\cdots$，与预期相符。数值上，取 $a=0.7$、$b=1.3$ 和 $x=0.3$，把 $c_m^2 x^m$ 加至很多项，得 $4.163878\ldots$，与闭合形式的值吻合。

常见错误

- 把区间写成 $b|x|<1$，即照抄第一部分的条件。新级数的公比是 a^2x、abx、b^2x，所以条件应为 $b^2|x|<1$。
- 让 c_m 停留在求和的形式。闭合形式 $(b^{m+1}-a^{m+1})/(b-a)$ 正是令题目余下部分得以完成的关键。
- 在 N 中 x 的系数出现正负号错误。老老实实地把每个乘积逐项展开，比尝试取巧更快。

STEP 之外

级数 $\sum c_m x^m$ 是数列 c_m 的*生成函数*（*generating function*），而这道题说明：如果一个数列有有理的生成函数，那么由它各项的平方组成的数列也是如此。这是两个有理幂级数的*阿达马乘积*（*Hadamard product*）的一个特例，而这种乘积必定是有理的。这里求得的区间是*收敛半径*（*radius of convergence*）的一个例子：有理函数的幂级数（power series）一直收敛到最近的极点（pole）为止，在这里就是 $x = 1/b^2$。关于生成函数，Wilf 的 *generatingfunctionology* 有十分精彩的阐述；至于幂级数及其收敛性，Abbott 的 *Understanding Analysis* 有严谨细致的讨论。

详细解答

STEP I 1988，第 6 题

纯数学 · 复数与几何

STEP I 1988，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/88-s1-q6

关键思路

在阿根图（Argand diagram）上，向量 $\overrightarrow{P_jP_k}$ 就是复数（complex number）$z_k - z_j$，所以向量相等变成差相等，而垂直（perpendicular）则变成“一个差除以另一个差所得的商是纯虚数（purely imaginary）”。把题目给定的数值代入后，条件 (i) 会以 z 表示出 z_4 和 w，而条件 (ii) 则决定了 z 的虚部（imaginary part）。这时会发现六边形关于其中心对称，凸性（convexity）和最后的角度因而都容易处理。

解答

把题目的陈述翻译成代数

由于 $\overrightarrow{P_jP_k}$ 由 $z_k - z_j$ 表示，陈述 (i) 即是说

$$z_2 - z_1 = z_4 - z_5 \qquad 及 \qquad z_3 - z_2 = z_5 - z_6.$$

至于 (ii)，两个非零向量 u 和 v 互相垂直，当且仅当它们的辐角（argument）相差 $\pm\frac{\pi}{2}$，亦即 u/v 是纯虚数。各点互不相同，所以 $z_4 - z_2 \neq 0$ 及 $z_6 - z_3 \neq 0$，而陈述 (ii) 即是说

$$\frac{z_4 - z_2}{z_6 - z_3} \text{ 是纯虚数}, \qquad 等价于 \qquad \operatorname{Re}\left[(z_4 - z_2)\,\overline{(z_6 - z_3)}\right] = 0.$$

(i) 的推论

把 (i) 的两条等式相加，z_2 和 z_5 两项互相抵消：

$$z_3 - z_1 = z_4 - z_6, \qquad 所以 \qquad z_4 - z_3 = z_6 - z_1,$$

这正是 $\overrightarrow{P_3P_4} = \overrightarrow{P_1P_6}$，即所要证明的结果。

六边形

现在设 $z_1 = 0$、$z_2 = 1$、$z_3 = z$、$z_5 = \mathrm{i}$ 及 $z_6 = w$。由 (i) 的第一条等式得

$$z_4 = z_5 + z_2 - z_1 = 1 + \mathrm{i},$$

而由第二条等式得 $z - 1 = \mathrm{i} - w$，所以

$$w = 1 + \mathrm{i} - z.$$

留意 $z_1 + z_4 = z_2 + z_5 = z_3 + z_6 = 1 + \mathrm{i}$：相对的顶点关于点 $\frac{1}{2}(1 + \mathrm{i})$ 对称，所以这个六边形有对称中心（centre of symmetry）。

至于 (ii)，$z_4 - z_2 = \mathrm{i}$，而 $z_6 - z_3 = w - z = 1 + \mathrm{i} - 2z$。商 $\mathrm{i}/(1 + \mathrm{i} - 2z)$ 是纯虚数，当且仅当 $1 + \mathrm{i} - 2z$ 是实数，亦即 $1 - 2\,\mathrm{Im}(z) = 0$。所以

$$\mathrm{Im}(z) = \tfrac{1}{2}, \qquad \text{于是} \qquad \mathrm{Im}(w) = 1 - \tfrac{1}{2} = \tfrac{1}{2}.$$

设 $z = a + \frac{1}{2}\mathrm{i}$，其中 a 为实数。那么 $w = (1 - a) + \frac{1}{2}\mathrm{i}$，所以 $\mathrm{Re}(w) = 1 - \mathrm{Re}(z)$ 恒成立。六个顶点是

$$(0,0), \quad (1,0), \quad (a, \tfrac{1}{2}), \quad (1,1), \quad (0,1), \quad (1 - a, \tfrac{1}{2}) :$$

即单位正方形在两条垂直边的半高处各加上一个“尖角”。

想一想 怎样才可以不靠图形来检验凸性?

顺着次序沿六边形走一圈。六边形是凸的，当且仅当我们在每个顶点都向同一方向转弯。在某个顶点的转向，可用进入该顶点的边 u 与离开该顶点的边 v 的叉积（cross product）$\mathrm{Im}(\bar{u}v) = u_x v_y - u_y v_x$ 的正负号来量度。各边是

$$P_1P_2 = (1,0), \quad P_2P_3 = (a - 1, \tfrac{1}{2}), \quad P_3P_4 = (1 - a, \tfrac{1}{2}),$$
$$P_4P_5 = (-1,0), \quad P_5P_6 = (1 - a, -\tfrac{1}{2}), \quad P_6P_1 = (a - 1, -\tfrac{1}{2}),$$

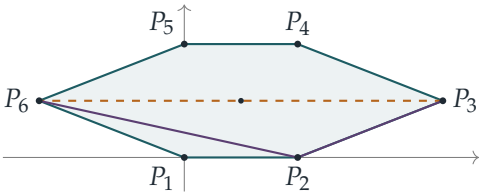
而在 $P_2, P_3, P_4, P_5, P_6, P_1$ 的叉积依次为

$$\tfrac{1}{2}, \qquad a - 1, \qquad \tfrac{1}{2}, \qquad \tfrac{1}{2}, \qquad a - 1, \qquad \tfrac{1}{2}.$$

无论 a 取甚么值，其中四个都是正数，所以六边形是凸的，当且仅当 $a - 1 > 0$。（若 $a = 1$，则 P_3 位于线段 P_2P_4 上，得不到真正的六边形。）从几何上看，P_3 必须位于正方形的边 $x = 1$ 的右方，而根据对称性，P_6 必须位于 $x = 0$ 的左方。所以条件是

$$\boxed{\mathrm{Re}(z) > 1 \quad \text{及} \quad \mathrm{Re}(w) = 1 - \mathrm{Re}(z) \ \ (\text{所以}\mathrm{Re}(w) < 0)}$$

（连同 $\mathrm{Im}(z) = \mathrm{Im}(w) = \frac{1}{2}$，这已由 (ii) 所决定）。



距离 P_3P_6

在 P_2 处，角的两臂是

$$z_3 - z_2 = (a-1) + \tfrac{1}{2}\mathrm{i}, \qquad z_6 - z_2 = -a + \tfrac{1}{2}\mathrm{i}.$$

对于角 $\theta = \angle P_3P_2P_6$（介乎 0 与 π 之间），两臂的点积（dot product）给出 $|P_2P_3|\,|P_2P_6|\cos\theta$，而叉积的大小给出 $|P_2P_3|\,|P_2P_6|\sin\theta$：

$$\text{点积} = -a(a-1) + \tfrac{1}{4} = \tfrac{1}{4}(1 + 4a - 4a^2), \qquad \text{叉积} = (a-1)\cdot\tfrac{1}{2} - \tfrac{1}{2}\cdot(-a) = \tfrac{1}{2}(2a-1).$$

当 $a > 1$ 时叉积为正数，所以它正好就是含 $\sin\theta$ 的那一项，于是

$$\tan\theta = \frac{\frac{1}{2}(2a-1)}{\frac{1}{4}(1+4a-4a^2)} = \frac{2(2a-1)}{1+4a-4a^2}.$$

令它等于 $-\frac{2}{3}$：

$$3(2a-1) = -(1+4a-4a^2) \quad\Longrightarrow\quad 4a^2 - 10a + 2 = 0 \quad\Longrightarrow\quad 2a^2 - 5a + 1 = 0,$$

所以 $a = \frac{1}{4}(5 \pm \sqrt{17})$。取负号得 $a \approx 0.22$，不符合凸性条件 $a > 1$；取正号得 $a \approx 2.28 > 1$。（这时分母 $1 + 4a - 4a^2$ 是负数，所以 θ 是钝角，正切为负数时正应如此。）

最后，P_3 和 P_6 的虚部相同，所以

$$P_3P_6 = |z - w| = |2a - 1| = 2a - 1 = \frac{5+\sqrt{17}}{2} - 1 = \boxed{\frac{3+\sqrt{17}}{2}} \approx 3.56.$$

如果不要求六边形是凸的，用同样的方法（这时分子要改为 $|2a-1|$）会多得一个解 $a = \frac{1}{4}(-1-\sqrt{17})$。这只是镜像 $a \mapsto 1-a$，它把 z 和 w 互换，所得的距离同样是 $\frac{1}{2}(3+\sqrt{17})$。

验算

- 对称性：$z_1 + z_4 = z_2 + z_5 = z_3 + z_6$，所以结果 $\overrightarrow{P_3P_4} = \overrightarrow{P_1P_6}$ 不过是中心对称六边形的第三对对边。
- 取 $a = \frac{1}{4}(5+\sqrt{17}) \approx 2.2808$，直接计算 $\arg\big((z_6 - z_2)/(z_3 - z_2)\big)$，所得的角的正切是 -0.66667，而 $|z - w| \approx 3.5616 = \frac{1}{2}(3+\sqrt{17})$。
- 对由 -5 至 5 的 a 以数值方法检验凸性，确认六边形是凸的当且仅当 $a > 1$。

常见错误

- 弄错向量的方向：$\overrightarrow{P_5P_4}$ 是 $z_4 - z_5$，而不是 $z_5 - z_4$。
- 把垂直写成“斜率之积为 -1”，结果遗漏了铅垂的情况。这里 $\overrightarrow{P_2P_4}$ 是铅垂的，而用商来检验就能干净利落地处理这种情况。
- 保留了根 $a = \frac{1}{4}(5 - \sqrt{17})$，但它并不给出凸六边形（事实上这时 $\tan\theta > 0$）。

STEP 之外

把点看成复数，几何就变成代数：平移是加法，旋转和缩放是乘法。“u/v 是纯虚数”这个检验，正是角度化为辐角的一个例子。以叉积的正负号检验凸性，是计算几何（computational geometry）中标准凸包（convex hull）算法的基础。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）把这种以几何看待复数的观点发挥得淋漓尽致。

详细解答

STEP I 1988，第 7 题

纯数学 · 二次函数导数的界限

> **STEP I 1988，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/88-s1-q7

本题在 *Beyond STEP* 第 3.6 节斜率的界中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/markov-bound.html

关键思路

二次函数（quadratic）由它在三个点的取值完全确定，所以 f' 可以由 $f(1)$、$f(-1)$ 和 $f(0)$ 重新构造出来。一旦把 f' 写成这个形式，假设 $|f| \le 1$ 便控制了三个系数中的每一个，余下的工作交给三角不等式（triangle inequality）就行了。至于最后一部分，我们考察 f 取哪些值时，证明中的每一个不等式都变成等式。

解答

$f'(x)$ 的恒等式

我们有 $f(1) = a+b+c$、$f(-1) = a-b+c$ 和 $f(0) = c$。把右边按 x 的幂次整理：

$$\begin{aligned} &f(1)\left(x+\tfrac{1}{2}\right) + f(-1)\left(x-\tfrac{1}{2}\right) - 2f(0)x \\ &\quad = \left[(a+b+c)+(a-b+c)-2c\right]x + \tfrac{1}{2}\left[(a+b+c)-(a-b+c)\right] \\ &\quad = 2ax + b. \end{aligned}$$

由于 $f'(x) = 2ax+b$，这就是所求的恒等式（identity）。

上界 $|f'(x)| \le 4$

假设只要 $|x| \le 1$，就有 $|f(x)| \le 1$。特别地，$|f(1)|$、$|f(-1)|$ 和 $|f(0)|$ 全都不超过 1。任取满足 $|x| \le 1$ 的 x。由恒等式和三角不等式，

$$\begin{aligned} |f'(x)| &\le |f(1)|\,\left|x+\tfrac{1}{2}\right| + |f(-1)|\,\left|x-\tfrac{1}{2}\right| + 2|f(0)|\,|x| \\ &\le \left|x+\tfrac{1}{2}\right| + \left|x-\tfrac{1}{2}\right| + 2|x|. \end{aligned}$$

> **想一想** 当 $|x| \le 1$ 时，$\left|x+\frac{1}{2}\right| + \left|x-\frac{1}{2}\right|$ 最大可以是多少?

这是 x 到 $\frac{1}{2}$ 和到 $-\frac{1}{2}$ 的距离之和。若 $|x| \le \frac{1}{2}$，则 x 位于这两点之间，距离之和恰好是 1。若 $|x| > \frac{1}{2}$，两点都在 x 的同一侧，距离之和是 $2|x|$。无论哪种情况，当 $|x| \le 1$ 时它都不超过 2。另外 $2|x| \le 2$，所以

$$|f'(x)| \le 2 + 2 = 4 \qquad \text{对于} |x| \le 1,$$

这正是要证明的结果。

（另一个看法：f' 是线性的，所以在 $[-1,1]$ 上它的绝对值（modulus）在端点处最大，而在 $x = \pm 1$ 处的上界是 $|f'(\pm 1)| \le \frac{3}{2} + \frac{1}{2} + 2 = 4$。）

等式成立的一个例子

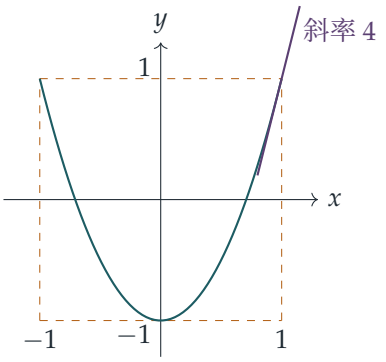
要在 $x = 1$ 处取得等式，上面的每一步在该处都必须恰好取等。在 $x = 1$ 处，系数 $x + \frac{1}{2}$、$x - \frac{1}{2}$ 和 $-2x$ 分别是 $\frac{3}{2}$、$\frac{1}{2}$ 和 -2。因此，要得到 $f'(1) = +4$，我们希望

$$f(1) = 1, \qquad f(-1) = 1, \qquad f(0) = -1.$$

由此得 $c = -1$、$a + b + c = 1$ 和 $a - b + c = 1$，所以 $b = 0$ 而 $a = 2$：

$$a = 2, \quad b = 0, \quad c = -1, \qquad f(x) = 2x^2 - 1.$$

现在检查条件。对于 $|x| \le 1$，我们有 $0 \le x^2 \le 1$，所以 $-1 \le 2x^2 - 1 \le 1$，即 $|f(x)| \le 1$。而 $f'(x) = 4x$，所以 $f'(1) = 4$，且 $x = 1$ 在区间之内。这就是我们需要的例子。



曲线 $y = 2x^2 - 1$（蓝绿色）始终留在正方形 $|x| \le 1$、$|y| \le 1$ 之内，由顶边荡到底边再荡回顶边，因此它在两端达到了可能的最大陡峭程度。

（$\pm(2x^2 - 1)$ 中任何一个都给出 $|\mathrm{f}'(\pm 1)| = 4$。选取 $\mathrm{f}(x) = 2x^2 - 1$ 会得到 $\mathrm{f}'(1) = +4$，而 $-(2x^2 - 1)$ 在 $x = -1$ 处也是如此。）

验算

- 试用 $\mathrm{f}(x) = x^2 + x + 1$：$\mathrm{f}(1) = 3$、$\mathrm{f}(-1) = 1$、$\mathrm{f}(0) = 1$。在 $x = 2$ 处，恒等式给出 $3 \cdot \frac{5}{2} + 1 \cdot \frac{3}{2} - 2 \cdot 2 = 5 = \mathrm{f}'(2)$。
- 对 20,000 个二次函数作随机搜索，每个都经缩放使得 $\max_{|x|\le 1} |\mathrm{f}(x)| = 1$，结果从未发现 $\max_{|x|\le 1} |\mathrm{f}'(x)| = 2|a| + |b|$ 超过 4（最大的约为 3.91），而 $2x^2 - 1$ 则恰好达到 4。

常见错误

- 把 $\left|x + \frac{1}{2}\right|$ 和 $\left|x - \frac{1}{2}\right|$ 分开估计，各以 $\frac{3}{2}$ 为上界。这样只能得到 $|\mathrm{f}'(x)| \le 5$。这两个距离必须合起来一并估计。
- 提出例子后，只在 0 和 ± 1 处检查，而没有在整个区间上验证 $|\mathrm{f}(x)| \le 1$。

STEP 之外

这是马尔可夫不等式（Markov's inequality）在二次情况下的特例：若 n 次多项式（polynomial）p 在 $[-1, 1]$ 上满足 $|p| \le 1$，则在该区间上 $|p'| \le n^2$。等式由切比雪夫多项式（Chebyshev polynomial）$T_n(\cos\theta) = \cos n\theta$ 取得，而 $2x^2 - 1 = T_2(x)$。利用多项式在选定点上的取值来把它写出，就是拉格朗日插值法（Lagrange interpolation），它是数值分析（numerical analysis）和逼近理论（approximation theory）的基本工具。想做更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP I 1988，第 8 题

纯数学 · 三维几何

STEP I 1988，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：

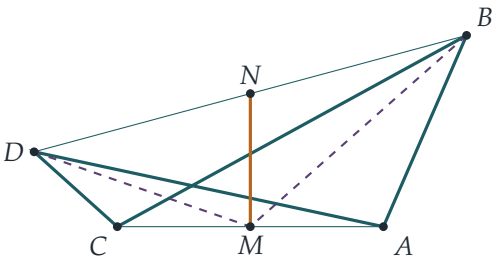
step.maths.org/questions/88-s1-q8

关键思路

每条对角线都把四边形分成两个三角形，而对边相等令这两个三角形全等（congruent）（SSS）。全等三角形的中线（median）相等，所以一条对角线的中点与另一条对角线的两个端点等距。若某点与一条线段的两个端点等距，把它与该线段的中点连起来，所得的直线便垂直于该线段：这其实就是等腰三角形（isosceles triangle）的性质。

解答

设 M 为 AC 的中点，N 为 BD 的中点。



上图是一个空间四边形（skew quadrilateral）的投影；虚线 MB 和 MD 是下文用到的中线，琥珀色的线段是 MN。

首先，$M \neq N$

只有当 M 和 N 是不同的点时，直线 MN 才有意义。假如 $M = N$，对角线 AC 和 BD 便会在一个公共点互相平分。过同一点的两条直线位于同一平面上，所以 A、B、C、D

全都在该平面上，而 $ABCD$ 就会是一个（平面的）平行四边形（parallelogram）。由于 $ABCD$ 是空间四边形，故 $M \neq N$。“空间”这个条件正正用在这里：平行四边形的对边相等，但两个中点却重合。

MN **垂直于** BD

考虑三角形 ABC 和 CDA，它们有公共的对角线 AC。我们有

$$AB = CD, \qquad BC = DA, \qquad CA = AC,$$

所以这两个三角形全等（SSS），其中 $A \leftrightarrow C$、$B \leftrightarrow D$、$C \leftrightarrow A$。在这个对应之下，三角形 ABC 中由 B 到 AC 中点的中线，对应于三角形 CDA 中由 D 到 CA 中点的中线。两条中线的另一端都是同一点 M，所以

$$MB = MD.$$

想一想 为甚么全等三角形的中线必定相等?

中线完全由边长决定。例如，根据余弦定理（cosine rule）（或阿波罗尼奥斯定理（Apollonius's theorem）），三角形 ABC 中由 B 引出的中线 m 满足 $4m^2 = 2AB^2 + 2BC^2 - AC^2$。边长相同，中线也就相同，所以 $MB^2 = \frac{1}{4}(2AB^2 + 2BC^2 - AC^2) = \frac{1}{4}(2CD^2 + 2DA^2 - CA^2) = MD^2$。

现在 M 与 B 和 D 等距，而 N 是 BD 的中点。三角形 MNB 和 MND 有 $MB = MD$、$NB = ND$，以及公共边 MN，所以它们全等（SSS），而 $\angle MNB$ 与 $\angle MND$ 两角相等。又因为 B、N、D 在同一直线上，这两个角互补（supplementary）。所以两者都是直角：$MN \perp BD$。（这两个三角形是真正的三角形，并非退化（degenerate）：假如 M 在直线 BD 上而又有 $MB = MD$，M 便会是中点 N，而这个情况我们已经排除了。）

MN **垂直于** AC

把两条对角线的角色对调，同样的论证依然成立。三角形 ABD 和 CDB 有公共的对角线 BD，而且

$$AB = CD, \qquad AD = CB, \qquad BD = DB,$$

所以它们全等，其中 $A \leftrightarrow C$、$B \leftrightarrow D$、$D \leftrightarrow B$。由 A 和由 C 引出的中线，另一端都是 BD 的中点 N，所以 $NA = NC$。由于 M 是 AC 的中点，而且 $N \neq M$，由上文的等腰三角形论证可得 $MN \perp AC$。

因此，连接两条对角线中点的直线垂直于每一条对角线，这正是所要证明的。

或者用向量

取 M 为原点，于是以位置向量（position vector）表示时有 $C=-A$，而 $\overrightarrow{MN}=\frac{1}{2}(\mathbf{b}+\mathbf{d})$。把题目给出的两个等式平方，

$$AB^2=CD^2 \iff |\mathbf{b}|^2-2\mathbf{a}\cdot\mathbf{b}=|\mathbf{d}|^2+2\mathbf{a}\cdot\mathbf{d},$$
$$BC^2=AD^2 \iff |\mathbf{b}|^2+2\mathbf{a}\cdot\mathbf{b}=|\mathbf{d}|^2-2\mathbf{a}\cdot\mathbf{d}.$$

两式相减得 $4\mathbf{a}\cdot\mathbf{b}=-4\mathbf{a}\cdot\mathbf{d}$，即 $\mathbf{a}\cdot(\mathbf{b}+\mathbf{d})=0$。由于 $\overrightarrow{AC}=-2\mathbf{a}$，这就是说 $\overrightarrow{MN}\perp AC$。改为把原点放在 N，便可用同样的方法得到 $MN\perp BD$。

验算

试取 $A=(1,0,0)$、$B=(1,2,1)$、$C=(-1,0,0)$、$D=(-1,-2,1)$。这时 $AB=CD=\sqrt{5}$，$BC=AD=3$，而且四点不共面（纯量三重积（scalar triple product）$\overrightarrow{AB}\cdot(\overrightarrow{AC}\times\overrightarrow{AD})$ 等于 8）。这里 $M=(0,0,0)$，$N=(0,0,1)$，所以 MN 沿着 z 轴，而 AC 沿着 x 轴，$\overrightarrow{BD}=(-2,-4,0)$ 则没有 z 分量：两者都垂直于 MN。我们还测试了 20 000 个对边相等的随机空间四边形；MN 与任何一条对角线之间夹角的余弦值，从未超过 4×10^{-12}。

常见错误

- 没有解释为甚么 $M\neq N$。缺少了“空间四边形”这个条件，结论便不成立，因为平行四边形根本没有直线 MN。
- 声称“根据对称性”有 $MB=MD$，却没有指明是哪两个全等三角形。
- 只证明了与其中一条对角线垂直。第二条需要用另一对全等三角形。

STEP 之外

在上面的例子中，绕 z 轴（即直线 MN）的半周旋转（half-turn）把 A 与 C 互换，也把 B 与 D 互换。这在一般情况下也成立：每一个对边相等的空间四边形，绕 MN 旋转 180° 后都会与自身重合。（这个旋转固定 M，并把每一个垂直于 MN 的方向反转，所以它把 A 送到 C；同样地，它把 B 送到 D。）反过来说，有了这个半周旋转，对边相等便一目了然。把几何事实看成某种对称（symmetry）——即空间的一个等距变换（isometry）——的结果，是大学一年级几何和群论（group theory）的核心思想；Allenby 的 *Rings, Fields and Groups*（环、域与群）阐述了群论方面的内容，而 Siklos 的 *Advanced Problems in Mathematics*（高等数学难题）则有更多 STEP 风格的几何题。

详细解答

STEP I 1988，第 9 题

纯数学 · 代换积分法

STEP I 1988，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/88-s1-q9

关键思路

只要选对着手的部分，两个积分都可以用标准技巧解决。在 (i) 中，$\ln x$ 与 x 的幂相乘，适合使用分部积分法（integration by parts）：对对数求导，对数便会消失。在 (ii) 中，棘手的是平方根，所以我们把它设为新变量。$\cos x\,\mathrm{d}x$ 正好是代换（substitution）所需要的，而余下的是一个有理函数（rational function），可以用部分分式（partial fractions）处理。

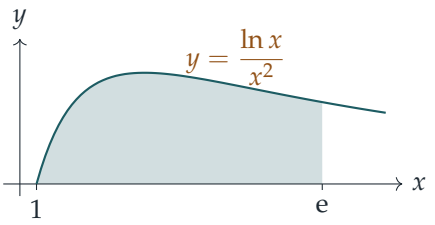
解答

第 (i) 部分

取 $u = \ln x$ 及 $\dfrac{\mathrm{d}v}{\mathrm{d}x} = \dfrac{1}{x^2}$ 作分部积分，于是 $\dfrac{\mathrm{d}u}{\mathrm{d}x} = \dfrac{1}{x}$ 及 $v = -\dfrac{1}{x}$：

$$\begin{aligned}\int_1^{\mathrm{e}} \frac{\ln x}{x^2}\,\mathrm{d}x &= \left[-\frac{\ln x}{x}\right]_1^{\mathrm{e}} + \int_1^{\mathrm{e}} \frac{1}{x^2}\,\mathrm{d}x \\ &= \left(-\frac{1}{\mathrm{e}} - 0\right) + \left[-\frac{1}{x}\right]_1^{\mathrm{e}} \\ &= -\frac{1}{\mathrm{e}} + 1 - \frac{1}{\mathrm{e}} = 1 - \frac{2}{\mathrm{e}}.\end{aligned}$$

这里用到 $\ln \mathrm{e} = 1$ 及 $\ln 1 = 0$。答案是正数，这是必然的：被积函数（integrand）在 $(1, \mathrm{e}]$ 上为正。



阴影部分的面积为 $1 - 2/\mathrm{e} \approx 0.264$。

第 (ii) 部分

被积函数在 $\sin x \neq 0$ 且 $1 + \sin x > 0$ 时有定义。设

$$u = \sqrt{1+\sin x}, \qquad \text{于是} u^2 = 1 + \sin x, \quad 2u\,\mathrm{d}u = \cos x\,\mathrm{d}x, \quad \sin x = u^2 - 1.$$

这里 $u > 0$，而且因为 $\sin x \neq 0$，所以 $u \neq 1$。于是

$$\int \frac{\cos x}{\sin x\sqrt{1+\sin x}}\,\mathrm{d}x = \int \frac{2u}{(u^2-1)\,u}\,\mathrm{d}u = \int \frac{2}{u^2-1}\,\mathrm{d}u.$$

? **想一想** $\cos x$ 到哪里去了？我们需要把它写成 u 的式子吗？

不需要。整个 $\cos x\,\mathrm{d}x$ 都由 $2u\,\mathrm{d}u$ 取代，这正是这个代换如此干净利落的原因。假如分子是别的东西，我们便要用到 $\cos x = \pm\sqrt{1-\sin^2 x}$，还得应付随之而来的种种正负号问题。

由部分分式得 $\dfrac{2}{u^2-1} = \dfrac{1}{u-1} - \dfrac{1}{u+1}$，所以

$$\int \frac{2}{u^2-1}\,\mathrm{d}u = \ln|u-1| - \ln|u+1| + C = \ln\left|\frac{u-1}{u+1}\right| + C.$$

换回 x，

$$\int \frac{\cos x}{\sin x\sqrt{1+\sin x}}\,\mathrm{d}x = \ln\left|\frac{\sqrt{1+\sin x}-1}{\sqrt{1+\sin x}+1}\right| + C.$$

绝对值（modulus）符号不可省略：当 $-1 < \sin x < 0$ 时，我们有 $0 < u < 1$，而 $\dfrac{u-1}{u+1}$ 是负数。（严格来说，由于定义域被 $\sin x$ 的零点分割成多个区间，在被积函数有定义的每个区间上，常数可以各不相同。）

把分数的分子和分母同乘以 $u-1$，可得一个等价的形式。由于 $(u+1)(u-1) = u^2 - 1 = \sin x$，答案亦可写成

$$2\ln\left|\sqrt{1+\sin x} - 1\right| - \ln|\sin x| + C.$$

验算

- 用辛普森法则(Simpson's rule)得 $\int_1^{\mathrm{e}} x^{-2}\ln x\,\mathrm{d}x \approx 0.264241$,与 $1-2/\mathrm{e} = 0.264241\ldots$ 吻合。
- 对于 (ii)，把答案求导：由链式法则（chain rule），$\dfrac{\mathrm{d}}{\mathrm{d}x}\ln\left|\dfrac{u-1}{u+1}\right| = \dfrac{2}{u^2-1}\cdot\dfrac{\cos x}{2u}$，这正是被积函数。数值上，用辛普森法则算得被积函数的 $\int_{0.3}^{1.2}$ 为 0.926006，与原函数（antiderivative）在两端的差一致。在 $\sin x < 0$ 的区间上，结果同样吻合，例如在 $[-1.2, -0.2]$ 上，积分为 -2.361027。

常见错误

- 在 (i) 中，反过来选取 $u = 1/x^2$ 及 $\mathrm{d}v/\mathrm{d}x = \ln x$。这样做原则上可行，但繁复得多；应该对对数求导。
- 在 (ii) 中，漏掉绝对值符号，以致答案在 $\sin x < 0$ 时没有定义。
- 漏掉 $2u\,\mathrm{d}u$ 中的因子 2，令答案减半。

STEP 之外

第 (ii) 部分是一个一般原理的例子：若被积函数是 x 与 $\sqrt{ax+b}$ 的有理函数，经代换 $u = \sqrt{ax+b}$ 后便会变成 u 的有理函数，而有理函数总可以用部分分式来积分。这里 $\sin x$ 扮演了 x 的角色。在大学一年级的分析（analysis）课程中，你会看到故事的另一面：那些为代换法和分部积分法提供依据的定理，以及积分常数（constant of integration）为何可以在定义域的各个分离区间上有所不同。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）对这两方面都有细致的论述。

详细解答

STEP I 1988，第 10 题

力学 · 抛体运动

> **STEP I 1988，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/88-s1-q10

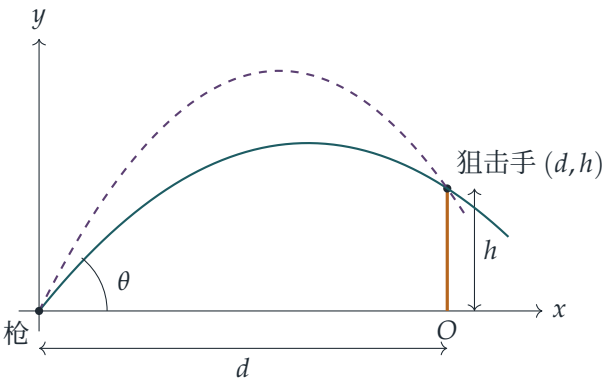
关键思路

把问题反过来想。设枪与树脚的水平距离为 d，然后问：是否存在*任何*一个仰角，使速率为 v 的子弹刚好经过树顶？以 $\tan\theta$ 表示的轨迹方程，是 $\tan\theta$ 的二次方程，所以答案为“是”的充要条件，就是该二次方程有实根。由判别式（discriminant）得出的条件，原来是一个只涉及 d 的不等式。

解答

模型

我们把子弹视为只受重力作用的质点（particle）（不计空气阻力），由地面以速率 v、与水平成 θ 的仰角（angle of elevation）射出，其中 $0 < \theta \leq \frac{\pi}{2}$。由于枪的位置未知，设它与树脚 O 的距离为 $d \geq 0$，并以枪为原点建立坐标轴，x 轴水平指向树，y 轴铅垂向上。于是树顶就是点 (d, h)。



击中狙击手

经过时间 t 后，子弹位于 $x = vt\cos\theta, y = vt\sin\theta - \frac{1}{2}gt^2$。若 $\theta < \frac{\pi}{2}$，我们以 $t = x/(v\cos\theta)$ 消去时间，并利用 $\sec^2\theta = 1 + \tan^2\theta$，得到轨迹（trajectory）

$$y = x\tan\theta - \frac{gx^2}{2v^2}(1 + \tan^2\theta).$$

子弹击中狙击手，当且仅当 (d,h) 位于这条曲线上。先假设 $d > 0$。记 $T = \tan\theta$ 及 $k = \dfrac{gd^2}{2v^2} > 0$，条件就是

$$kT^2 - dT + (h+k) = 0.$$

每个实数 T 都恰好是唯一一个 $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ 的 $\tan\theta$。不过，根会不会是负数呢？两根之积为 $(h+k)/k > 0$，两根之和为 $d/k > 0$，所以只要有实根，两根都必为正数，对应真正的仰角 $0 < \theta < \frac{\pi}{2}$。因此，存在可行的仰角，当且仅当这个二次方程有实根，亦即其判别式非负：

$$\begin{aligned} d^2 - 4k(h+k) \geq 0 &\iff d^2 \geq \frac{2gd^2}{v^2}\left(h + \frac{gd^2}{2v^2}\right) \\ &\iff v^4 \geq 2ghv^2 + g^2d^2 \qquad \text{（除以 } d^2 > 0\text{，再乘以 } v^4\text{）} \\ &\iff d^2 \leq \frac{v^2}{g^2}(v^2 - 2gh). \end{aligned}$$

想一想 假如枪就在树脚，即 $d = 0$，以致上面的轨迹方程派不上用场，那又如何？

这时子弹必定是铅垂向上射出的（$\theta = \frac{\pi}{2}$），而它能到达高度 h，当且仅当 $v^2 \geq 2gh$；这可由 $v^2 = u^2 - 2gs$ 及末速率不小于零得出。这正是 $d = 0$ 时的同一条件，所以该不等式涵盖了所有情况。

由于 $d \geq 0$，取非负平方根便得

$$d \leq \frac{v}{g}\sqrt{v^2 - 2gh}.$$

所以枪位于水平地面上的某一点，而该点与树脚的距离至多为 $(v/g)\sqrt{v^2-2gh}$：它位于以树脚为圆心、以此为半径的圆之内（极端情况下则在圆上），正如所求。平方根是实数，因为我们知道这一枪确实发生了：取 $d = 0$ 时，不等式迫使 $v^2 \geq 2gh$，这不过是说子弹有足够的能量升到高度 h。

这个条件也是充分的。从圆内任何一点出发，该二次方程都有两个正实根，所以有两个可能的仰角（一个较平、一个较陡的射击，如图所示）；而在圆上，两者合而为一。

验算

- 当 $h = 0$ 时，半径为 v^2/g，正是大家熟悉的水平地面上的最大射程。当 $v^2 \to 2gh$ 时，半径缩小至 0：枪只有在树脚笔直向上射击，才仅仅能够击中狙击手。
- 单位：$(v/g)\sqrt{v^2}$ 的单位是 $(\mathrm{m\,s^{-1}})^2/(\mathrm{m\,s^{-2}}) = \mathrm{m}$。
- 数值上，取 $g = 9.8$、$v = 20$ 及 $h = 8$，半径约为 31.8 m。从 $d = 31.8\,\mathrm{m}$ 出发，以唯一容许的仰角 $\tan\theta = v^2/(gd)$ 逐步推算运动，子弹到达树的位置时高度为 8.000 m。再对所有仰角逐一扫描，可确认圆外百分之几的点无法击中目标，而圆内的点则可以。

常见错误

- 只用最大射程 v^2/g 或只用最大高度。狙击手位于高度 h，所以两个坐标必须同时考虑。
- 忘记检查 $\tan\theta$ 的根是否对应水平以上的仰角，或忽略了 $d = 0$ 的铅垂射击。

STEP 之外

对于固定的 v，由枪可以射到的点填满了抛物线 $y = \dfrac{v^2}{2g} - \dfrac{gx^2}{2v^2}$ 以下的区域；这条抛物线是整族轨迹的*包络线*（envelope），又称“安全抛物线”（parabola of safety）。我们的判别式条件，正正就是 $h \leq$ 这条包络线在 $x = d$ 处的高度。曲线族的包络线，在变分法（calculus of variations）和一阶微分方程的理论中还会再次出现。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教材），以同样严谨的方式处理抛体问题，是很好的进阶读物。

详细解答

STEP I 1988，第 11 题

力学 • 质心

STEP I 1988，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/88-s1-q11

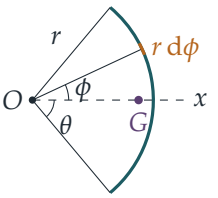
关键思路

对于圆弧，由对称性可知质心（centre of mass）位于平分圆弧的半径上，只需对角度作一次积分，便可求出质心与圆心的距离。至于框架，最值得注意的是它以半圆上的每一点着地都能平衡。地面的反作用力（reaction）总是沿半圆的半径作用，而同时位于所有这些半径上的点只有圆心。因此整个框架的质心必定是半圆的圆心，再对该点取力矩（moments），便得到一条关于 h 的方程。

解答

均匀圆弧

设圆弧的圆心为 O，每单位长度的密度为 ρ。从平分圆弧的半径起量度角 ϕ，则圆弧为 $-\theta \leq \phi \leq \theta$。取 x 轴沿该平分线。由于圆弧关于此轴对称，质心必在此轴上，故只需求 $\bar{x}$。



介乎 ϕ 与 $\phi + \mathrm{d}\phi$ 之间的线元长度为 $r\,\mathrm{d}\phi$，故质量为 $\rho r\,\mathrm{d}\phi$，而它的 x 坐标为 $r\cos\phi$。总

质量为 $\rho r \cdot 2\theta$。因此

$$\bar{x} = \frac{\displaystyle\int_{-\theta}^{\theta} r\cos\phi \; \rho r \, \mathrm{d}\phi}{2\rho r\theta} = \frac{r\big[\sin\phi\big]_{-\theta}^{\theta}}{2\theta} = \frac{r\sin\theta}{\theta}.$$

所以质心位于平分圆弧的半径上，与圆心的距离为

$$\boxed{\bar{x} = \frac{r\sin\theta}{\theta}}$$

对于半圆弧，$\theta = \frac{\pi}{2}$，此距离为 $\dfrac{2r}{\pi}$，这正是下文要用到的情况。

框架：质心必须在哪里？

我们把框架视为刚体（rigid body），由每单位长度密度同为 ρ 的均匀幼杆组成，并假设地面是水平的。设 O 为半圆的圆心，亦即长方形底边的中点。

> **想一想** 为甚么“以半圆上任何一点着地都处于平衡（equilibrium）”就能完全确定质心的位置？

假设框架以半圆上的点 P 着地而静止。地面在 P 点与半圆相切，所以通过 P 的铅垂线垂直于该切线：它就是半径所在的直线 OP。作用于框架的力只有两个，即作用于质心 G 的重量和作用于 P 的反作用力。要达至平衡，两力必须沿同一直线作用，所以 G 位于通过 P 的铅垂线上，亦即直线 OP 上。这对半圆上的每一点 P 都成立。两个不同的点 P 给出两条通过 O 的不同直线，而它们只相交于 O。因此 $G = O$。

反过来说，若 $G = O$，则对每一点 P，重量都沿 OP 作用，刚好穿过接触点，所以框架无论以哪一点接触地面都处于平衡。因此题目的条件恰好等同于“G 位于 O”。

对底边取力矩

对底边所在的直线取力矩，高度由 O 起向上量度（故半圆在该直线之下）。各部分如下：

部分	质量	其质心的高度	力矩
顶杆	$2r\rho$	$2h$	$4rh\rho$
两条侧杆	$2 \times 2h\rho$	h	$4h^2\rho$
底杆	$2r\rho$	0	0
半圆	$\pi r\rho$	$-\dfrac{2r}{\pi}$	$-2r^2\rho$

半圆的质心由圆弧公式取 $\theta = \frac{\pi}{2}$ 得出，而且它在 O 的下*方*。由于 G 的高度为 0，令总力矩等于零，再除以 2ρ：

$$2h^2 + 2rh - r^2 = 0.$$

这是一条关于 h 的二次方程，其根为

$$h = \frac{-2r \pm \sqrt{4r^2 + 8r^2}}{4} = \frac{(-1 \pm \sqrt{3})\, r}{2}.$$

由于 $h > 0$，我们取正根：

$$\boxed{h = \frac{(\sqrt{3} - 1)\, r}{2} \approx 0.366\, r}$$

留意底杆对它本身所在直线的力矩为零。无论框架是否包括底杆，答案都一样，所以对图形的理解稍有不同亦无妨。

验算

- 当 $\theta \to 0$ 时，$\frac{r \sin\theta}{\theta} \to r$：极短的圆弧，其质心就在圆弧本身之上。对于整个圆，$\theta = \pi$，我们得到 0，即圆心，正如对称性所要求。
- 取 $h = \frac{(\sqrt{3}-1)r}{2}$：$2h^2 = (2 - \sqrt{3})r^2$ 而 $2rh = (\sqrt{3} - 1)r^2$，两者相加为 r^2，符合要求。
- 我们把每条杆分成 20 000 小段，以数值方法计算质心。取 $r = 1$ 及 $h \approx 0.36603$，质心与 O 的距离小于 10^{-9}。把 h 增加 10%，质心便上移约 0.03，框架就不能再平衡了。

常见错误

- 误用 $\frac{4r}{3\pi}$，那是半圆形*薄片*（lamina）的结果；半圆形金*属线*应使用 $\frac{2r}{\pi}$。
- 以角度而非长度作加权：线元的质量是 $\rho r\, \mathrm{d}\phi$，因子 r 之所以能约去，只因为它同样出现在总质量之中。
- 没有给出理由便断言 G 位于 O。论证的关键在于反作用力总是沿半径作用。

STEP 之外

圆弧的积分是把质心表示为沿曲线的*线积分*（line integral）$\int \mathbf{r}\, \rho\, \mathrm{d}s$ 的最简单例子。大学一年级的课程会进一步利用二重积分和三重积分处理曲面和立体，而帕普斯（Pappus）定理则把这些质心与旋转体的面积和体积联系起来。这个平衡论证也是摇摆稳定性背后的想法：物体搁在曲面上时，若其质心低于曲率中心（centre of curvature），它便是稳定的。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）正是以这种精神处理质心和刚体平衡。

详细解答

STEP I 1988，第 12 题

力学 • 变力作用下的运动

> **STEP I 1988，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/88-s1-q12

关键思路

对每一种方法，求出溜冰者朝岸边最远能前进多少；只要 d 大于其中至少一种方法的这个距离，她就安全，反之则不然。转弯是典型的圆周运动（circular motion）：一个大小不变、与速度垂直的力会令她沿固定半径的圆运动，而这个半径正是她仍会向前推进的距离。煞停则是变力（variable force）问题：力取决于速率，但我们要求的是距离，这正提示我们把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$。

解答

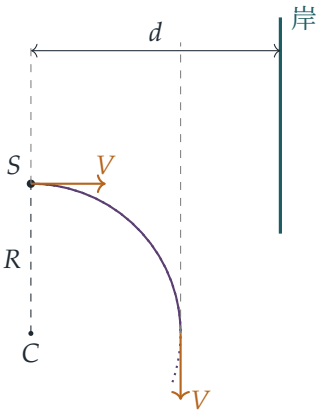
模型

我们把溜冰者视为质量为 M 的质点（particle），置于光滑（smooth）的水平冰面上。她在水平方向所受的唯一的力就是她自己施加的那个力，因此重量与法向反作用力（normal reaction）互相抵消，往后不必再考虑。我们由她的起点朝岸边量度距离 x，所以岸边位于 $x = d$。

方法一：沿圆弧转弯

与速度垂直的力不做功，所以她的速率保持为 V。一个大小恒为 F 的力，若始终与速率恒为 V 的速度成直角，便提供了沿半径为 R 的圆运动所需的向心力（centripetal force），其中

$$F = \frac{MV^2}{R}, \qquad \text{所以} \qquad R = \frac{MV^2}{F}.$$



她由 S 出发，正对着岸边前进，所以圆心 C 位于通过 S 且与岸边平行的直线上，与 S 相距 R。转过角度 θ 之后，她朝岸边前进的距离为 $x = R\sin\theta$，当 $\theta = \frac{\pi}{2}$ 时达到最大值 R；这时她正与岸边平行地运动，其后便会弯离岸边。所以她能避开岸边的条件是

$$d > \frac{MV^2}{F}.$$

（若恰好 $d = MV^2/F$，她会刚好擦过岸边，沿着岸边运动。）

方法二：煞停

现在力沿着她的运动直线正向后作用，所以她沿直线朝岸边运动，并逐渐减速。由于我们要求的是距离而不是时间，把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$，由牛顿（Newton）第二定律得

$$Mv\frac{\mathrm{d}v}{\mathrm{d}x} = -\frac{F(V^2+v^2)}{2V^2}.$$

分离变量（separating the variables），并由 $x = 0$（该处 $v = V$）积分至她停下的位置 $x = X$：

$$\int_V^0 \frac{v}{V^2+v^2}\,\mathrm{d}v = -\frac{F}{2MV^2}\int_0^X \mathrm{d}x,$$
$$\left[\tfrac{1}{2}\ln(V^2+v^2)\right]_V^0 = -\frac{FX}{2MV^2},$$
$$\tfrac{1}{2}\ln\frac{V^2}{2V^2} = -\tfrac{1}{2}\ln 2 = -\frac{FX}{2MV^2},$$

所以她的停止距离（stopping distance）为

$$X = \frac{MV^2}{F}\ln 2.$$

想一想 她真的会停下来吗? 还是她的速率只会趋近于零, 而她永远在缓缓向前挪动?

煞停的力永远不会低于 $\frac{1}{2}F$, 所以她的速率会在有限时间内降至零。具体来说, 把 $M\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{F}{2V^2}(V^2+v^2)$ 积分可得 $\frac{2MV}{F}\left(\arctan 1 - \arctan\frac{v}{V}\right) = t$, 所以她在 $t = \pi MV/(2F)$ 时停下, 其间走过有限的距离 X。(一旦静止, 她便不再施力。) 因此当 $d > X$ 时, 煞停可以救她; 若 $d = X$, 她到达岸边时速率为零, 只是碰到岸边而没有任何撞击。

比较两种方法

由于 $\ln 2 \approx 0.693 < 1$, 停止距离比转弯半径短: 煞停是较佳的方法, 而且凡是转弯行得通的情况, 煞停也行得通。因此, 她能够避免撞上岸边, 当且仅当

$$\boxed{d > \frac{MV^2}{F}\ln 2}$$

(而 $d = \frac{MV^2}{F}\ln 2$ 是临界情况, 这时她到达岸边时刚好静止)。

质点假设

真实的溜冰者是有大小的: 她的身躯和手臂会伸到质心 (centre of mass) 前方一段距离, 所以实际上她所需的 d 要比上述数值大, 多出的大约就是这段伸展的长度。更重要的是, 这些力来自她的冰刀对冰面的作用。要转弯, 她必须向弯内倾侧; 要急煞, 她必须把冰刀扭至横向于运动方向 (即冰球式急停, hockey stop), 这会令她的身体转向侧面。这类转动与姿势, 正是质点模型所忽略的。模型还暗中假设了冰面确实能够提供这些力, 尽管冰面被称为光滑; 即使沿冰刀方向没有摩擦力, 冰刀在横向仍能抓紧冰面。

验算

- 单位: MV^2/F 的单位是 $\mathrm{kg\,m^2\,s^{-2}/N} = \mathrm{m}$, 是长度, 正如所应得的。
- 上下界: 煞停的力介乎 $\frac{1}{2}F$ 与 F 之间。恒力 F 会令她在距离 $MV^2/(2F)$ 内停下, 而恒力 $\frac{1}{2}F$ 则为 MV^2/F; 事实上确有 $\frac{1}{2} < \ln 2 < 1$。
- 数值验证: 随机选取 M、V、F (例如 $R = MV^2/F = 7.0747$), 以龙格—库塔 (Runge–Kutta) 积分法逐步计算两种运动, 得出转弯时最远前进距离为 7.0747, 煞停时停止距离为 $4.9038 = R\ln 2$, 停止时间为 $\pi MV/(2F)$。

常见错误

- 把圆心放在运动直线上（或放在岸边），因而得出前进距离为 $2R$ 或 0；其实圆心应在她的旁边，前进距离才是 R。
- 处理煞停时使用 $\frac{\mathrm{d}v}{\mathrm{d}t}$，然后苦于求不出距离；用 $v\,\frac{\mathrm{d}v}{\mathrm{d}x}$ 只需积分一次便可得出。
- 给出两个各自独立的条件，却没有加以比较。题目问的是她能够脱险的 d 值所成的集合，所以我们需要的是两种方法中较佳的一种。

STEP 之外

写出 $\ddot{x} = v\,\frac{\mathrm{d}v}{\mathrm{d}x}$ 是迈向能量方法的第一步：对于只取决于位置的力，它就是功能定理（work–energy theorem），而在大学力学中，它会发展成势（potential）以及描述运动的相平面（phase plane）图。溜冰者在转向与煞停之间的抉择，也是一个简单的最优控制（optimal control）问题：怎样运用有限的力来避开障碍物，才是最好的方法？Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）正好在高一个层次上讨论取决于速度的力和圆周运动。

详细解答

STEP I 1988，第 13 题

力学 · 弹性绳与碰撞

> **STEP I 1988，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/88-s1-q13

关键思路

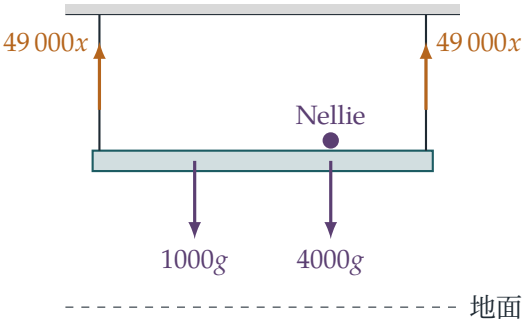
整个过程分为三个阶段，每个阶段各有自己的守恒定律（conservation law）。Nellie 落在木板上是一次突然发生的非弹性碰撞（inelastic collision），所以动量（momentum）守恒，但能量并不守恒。下降至地面的过程只涉及重力和弹性绳，所以能量守恒。Nellie 离开木板之后，能量再次守恒，而我们要问的是：储存的弹性势能是否足以把木板提升至超过绳子回复自然长度（natural length）的位置。

解答

我们把木板看成一个始终保持水平的刚体（rigid body），于是可以把它当作一个质量为 1000 kg、沿铅垂方向运动的质点（particle）。Nellie 是一个质量为 4000 kg 的质点，而绳子是遵守虎克定律（Hooke's law）的轻弹性绳（light elastic string）。两条绳子的伸长量（extension）x 始终相同，所以它们合共施加向上的力

$$2 \cdot \frac{490\,000}{10}\, x = 98\,000\, x \text{ N}.$$

因此这对绳子的作用就如一条劲度（stiffness）为 $K = 98\,000 \text{ N m}^{-1}$ 的弹簧，其弹性势能（elastic energy）为 $\frac{1}{2}Kx^2 = 49\,000\, x^2$ J。



起始位置

Nellie 到来之前，木板悬挂着处于平衡（equilibrium）状态，所以伸长量 e_0 满足

$$98\,000\,e_0 = 1000 \times 9.8 \quad \Longrightarrow \quad e_0 = 0.1\ \text{m}.$$

落在木板上

Nellie 落在木板上之后便留在上面，所以这是一次完全非弹性碰撞（perfectly inelastic collision）。撞击历时极短，而绳子的张力（tension）和各重量都是有限的力，所以它们在撞击期间的冲量（impulse）可以忽略不计。因此动量守恒，而紧接撞击之后的共同速率 v 满足

$$4000 \times 5 = 5000\,v \quad \Longrightarrow \quad v = 4\ \text{m}\,\text{s}^{-1}.$$

想一想 我们可以从 Nellie 到达的一刻起就直接使用能量守恒吗?

不可以。一头大象落在木板上并留在上面时，大量动能（kinetic energy）会转化为热和声音：这里 $\frac{1}{2}\cdot 4000\cdot 5^2 = 50\,000\,\text{J}$ 变成了 $\frac{1}{2}\cdot 5000\cdot 4^2 = 40\,000\,\text{J}$，即损失了 $10\,000\,\text{J}$。能量只是从紧接落下之后的一刻起才守恒。

下降过程

设 h 为木板最初离地面的高度。木板和 Nellie（总质量 $5000\,\text{kg}$）以速率 $4\,\text{m}\,\text{s}^{-1}$ 出发，在低了距离 h 的地面处静止下来，该处的伸长量为 $0.1+h$。做功的力只有重力和绳子的拉力，所以

$$\underbrace{\tfrac{1}{2}\cdot 5000\cdot 4^2}_{\text{动能损失}} + \underbrace{5000\cdot 9.8\,h}_{\text{重力势能损失}} = \underbrace{49\,000\left[(0.1+h)^2 - 0.1^2\right]}_{\text{弹性势能增加}},$$

$$40\,000 + 49\,000\,h = 49\,000\,(0.2h + h^2).$$

两边除以 49 000 并整理，得

$$h^2 - 0.8\,h - \tfrac{40}{49} = 0, \qquad \text{所以} \qquad h = 0.4 + \sqrt{0.16 + \tfrac{40}{49}} = \frac{14 + 2\sqrt{299}}{35},$$

由于 $h > 0$，这里取正根。因此

$$h \approx 1.39\ \text{m}.$$

在这个阶段中绳子一直保持拉紧，因为木板下降时伸长量只会增加。

Nellie 离开之后

Nellie 轻轻地走下木板，所以只剩下木板（1000 kg）静止在地面处，伸长量为

$$X = 0.1 + h \approx 1.488\ \text{m}.$$

如果木板上升得足够高，使伸长量降至零，绳子便会松弛（slack），这表示木板要上升高度 X。绳子拉紧期间能量守恒，因此只要储存的弹性势能大于势能的增加，木板便会到达该高度：

$$\tfrac{1}{2}KX^2 = 49\,000\,X^2 \approx 108\,500\ \text{J}, \qquad 1000 \cdot 9.8 \cdot X \approx 14\,600\ \text{J}.$$

前者远大于后者，所以木板到达自然长度位置时仍在向上运动，动能约为 93 900 J（速率约为 $13.7\,\text{m}\,\text{s}^{-1}$）。在这一点之上，绳子要比自然长度更短才行，所以它们是松弛的。**会，绳子会变得松弛。**

> ? **想一想** 我们不可以直接用简谐运动吗?

可以,只要绳子仍然拉紧。木板独自绕着其平衡伸长量 0.1 m 作简谐运动(simple harmonic motion，简称 SHM)，而它从平衡位置之下 $h \approx 1.39$ m 处出发，所以振幅（amplitude）是 1.39 m。振动的最高点会在伸长量为 $0.1 - 1.39 < 0$ 之处，而这对拉紧的绳子来说是不可能的。所以绳子必定在木板到达最高点之前松弛。（事实上，只要 $h > 0.1$，SHM 的论证便已给出答案。）

绳子一旦松弛，木板便只在重力作用下自由运动，再上升 $13.7^2/(2 \times 9.8) \approx 9.6$ m。这小于 10 m，所以木板（仅仅!）不会在回落之前碰到绳子的固定点。

验算

- 把 $h = 1.3881$ 代入能量方程：两边都等于 108 017 J（准确至最接近的焦耳）。
- 由 $x = 0.1$、$\dot{x} = 4$ 出发，以数值方法逐步计算 $\ddot{x} = g - Kx/5000$，得出木板下降 1.3881 m 后停下。

- 逐步计算 Nellie 离开后木板的运动，显示伸长量降至零时木板正以 $13.7\,\mathrm{m\,s^{-1}}$ 向上运动，而最高点在地面之上 $11.07\,\mathrm{m}$，与 $X + 9.58$ 吻合。
- 单位：$490\,000\,\mathrm{N} / 10\,\mathrm{m}$ 给出 $\mathrm{N\,m^{-1}}$，所以能量方程中每一项的单位都是焦耳。

常见错误

- 在落下的一刻使用能量守恒，这样得出的共同速率是 $\sqrt{20} \approx 4.47\,\mathrm{m\,s^{-1}}$，而不是 $4\,\mathrm{m\,s^{-1}}$。
- 由自然长度起计算弹性势能，却忘记了初始伸长量 $0.1\,\mathrm{m}$：增加的能量是 $49\,000[(0.1 + h)^2 - 0.1^2]$，而不是 $49\,000\,h^2$。
- 只用了一条绳子的劲度，而不是两条。

STEP 之外

这对绳子是一条非线性弹簧（nonlinear spring）：拉长时它会拉，但压缩时却不能推。这类会在不同运动方程之间切换的系统，称为分段线性（piecewise-linear）振子或“碰撞”振子（“impact” oscillator），即使是简单的例子，其行为也可以出乎意料地复杂。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对能量、碰撞和振动有严谨的论述，而这些正是本题的全部基础。

详细解答

STEP I 1988，第 14 题

概率与统计 · 删失正态分布的期望值

> **STEP I 1988，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/88-s1-q14

关键思路

变量 X_M 其实就是 $\min(X, M)$：把正态变量（normal variable）中所有高于 M 的部分一律压低到 M。它的期望值（expectation）可拆成两部分：一个在 $x < M$ 上的积分，以及一个常数乘以尾部概率（tail probability）；而这个积分很容易计算，因为 $x\phi(x)$ 除了正负号之外，正是 $\phi(x)$ 的导数。游客的故事只是同一个想法换了个包装。一队巴士所接载的乘客人数是 $\min(\text{需求}, \text{载客量})$，经正态近似（normal approximation）之后，这正好就是一个 X_M。

解答

X_M 的期望值

由于在 $\{X < M\}$ 上 $X_M = X$，而在 $\{X \geq M\}$ 上 $X_M = M$，

$$\mathrm{E}(X_M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M\,\mathrm{P}(X \geq M) = \int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x \;+\; M(1-\Phi(M)).$$

> **想一想** 有没有一个函数，其导数是 $x\phi(x)$?

有。由 $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$，按链式法则（chain rule）可得 $\phi'(x) = -x\phi(x)$。因此

$$\int_{-\infty}^{M} x\,\phi(x)\,\mathrm{d}x = \Big[-\phi(x)\Big]_{-\infty}^{M} = -\phi(M) + 0,$$

因为当 $x \to -\infty$ 时 $\phi(x) \to 0$。故此

$$\mathrm{E}(X_M) = -\phi(M) + M(1-\Phi(M)),$$

这正是所求的结果。

作为合理性检查，令 $M \to \infty$ 可得 $\mathrm{E}(X_M) \to 0 = \mathrm{E}(X)$，因为 $M(1-\Phi(M)) \to 0$。令 $M \to -\infty$ 则得 $\mathrm{E}(X_M) \approx M$，与预期相符。

巴士

设 N 为某一次到访中被指往巴士站的游客人数。我们假设 1024 名游客各自独立地被指路，往两边的概率各为 $\frac{1}{2}$。于是

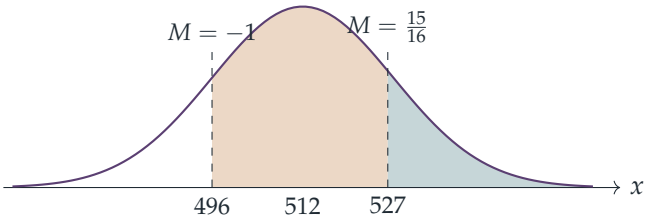
$$N \sim \mathrm{B}(1024, \tfrac{1}{2}), \qquad \mathrm{E}(N) = 512, \qquad \mathrm{Var}(N) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256 = 16^2.$$

由于 $n = 1024$ 这么大，N 可以非常准确地用 $512 + 16X$ 来近似，其中 X 服从标准正态分布（standard normal distribution）。

有 16 辆巴士时，载客量是 $16 \times 31 = 496$。有 17 辆时则是 527。载客量为 C 的车队接载 $\min(N, C)$ 名乘客，所以它每次到访的利润（以 vloska 计）是

$$\begin{aligned}\min(N,C) &\approx \min(512 + 16X,\ C) \\ &= 512 + 16\min\Big(X, \frac{C-512}{16}\Big) = 512 + 16\,X_M, \qquad \text{其中} M = \frac{C-512}{16}.\end{aligned}$$

因此 $C = 496$ 对应 $M = -1$，而 $C = 527$ 对应 $M = \frac{15}{16}$。



当需求落在琥珀色的区域时，第 17 辆巴士接载的是超出 496 的那部分乘客。当需求落在蓝绿色的尾部时，它载满 31 人。

利用 $\phi(1) \approx 0.24197$ 和 $\Phi(1) \approx 0.84134$，并留意 $1 - \Phi(-1) = \Phi(1)$，

$$\mathrm{E}(X_{-1}) = -\phi(1) - \big(1 - \Phi(-1)\big) = -\phi(1) - \Phi(1) \approx -0.24197 - 0.84134 = -1.08332.$$

利用 $\phi(0.9375) \approx 0.25707$ 和 $1 - \Phi(0.9375) \approx 0.17425$，

$$\mathrm{E}(X_{15/16}) \approx -0.25707 + 0.9375 \times 0.17425 \approx -0.25707 + 0.16336 = -0.09371.$$

因此，多一辆巴士令每次到访的期望利润增加

$$16\big(\mathrm{E}(X_{15/16}) - \mathrm{E}(X_{-1})\big) \approx 16 \times 0.98961 \approx 15.83 \text{ vloska}.$$

每年有 50 次到访，所以多一辆巴士的价值大约是

$$50 \times 15.83 \approx 792 \text{ vloska（每年）}.$$

这就是值得考虑的每年贿款上限。付出的金额一旦超过期望的额外利润，合作社平均而言就会得不偿失。

> **想一想** 既然 N 是离散的，我们不是应该使用连续性校正（continuity correction）吗？

这里并不需要。连续性校正的用处，在于近似 $\mathrm{P}(N \le c)$ 这类概率，因为这时被取平均的函数在整数处有跳跃。函数 $k \mapsto \min(k, C)$ 没有跳跃，不作校正的正态近似对它已经极为准确。以二项分布（binomial distribution）作精确计算，结果是 791.68，与上面的数字吻合至小数点后两位。把载客量移动 $\frac{1}{2}$ 反而会令结果变差，得出 775。

验算

- 以中点法则作数值积分（midpoint quadrature）计算 $\int_{-\infty}^{M} x\phi(x)\,\mathrm{d}x + M(1-\Phi(M))$，对多个 M 值所得的结果都与公式吻合至小数点后八位。以 4×10^5 个样本对 $\min(X, M)$ 进行蒙地卡罗模拟（Monte Carlo simulation），结果亦吻合。
- 对 $\mathrm{B}(1024, \frac{1}{2})$ 作精确求和，得出每次到访的额外期望利润为 15.8337，即每年 791.68，与正态近似相符。
- 粗略的合理性检查：只要 $N > 496$，多出的那辆巴士就有收入，这种情况出现的概率约为 0.84，而且它往往满座。因此每次到访的数字比 31 的一半略多，是合理的。

常见错误

- 忘记换算比例：$\min(N, C)$ 是 $512 + 16X_M$，所以 $\mathrm{E}(X_M)$ 之差必须先乘以标准差（standard deviation）16，再乘以 50。
- $M = -1$ 时的正负号错误：$M(1-\Phi(M))$ 一项是 $-(1-\Phi(-1)) = -\Phi(1)$，而不是 $-(1-\Phi(1))$。

STEP 之外

$\mathrm{E}\min(X, M)$ 这个量是一个*删失*（censored）正态变量的期望值。与它密切相关的表达式，例如 $\mathrm{E}(X-M)^+ = \phi(M) - M(1-\Phi(M))$，是存货理论（inventory theory，即“报贩问题”（newsvendor problem））以及布莱克–舒尔斯（Black–Scholes）期权定价的骨

干。恒等式 $\mathrm{E}\min(N, C+1) - \mathrm{E}\min(N, C) = \mathrm{P}(N > C)$ 提供了一个简洁的方法，让人看清每多一个单位载客量的价值。想进一步了解这些概念，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）。

详细解答

STEP I 1988，第 15 题

概率与统计 • 二项分布的假设检验

STEP I 1988，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/88-s1-q15

关键思路

把“超过一半的得分来自犯规”化为犯规次数与入球数之间的一个简单不等式（inequality）。以入球数作为条件（conditioning）之后，余下的工作只是数一数有多少个等可能的犯规次数符合要求。至于第二部分，要问的是：按照 B 分布，在 300 次试验中预期会出现多少次九次犯规的情况？假如甚么都没有改变，观察到的次数又有多出人意料？

解答

概率

就某场比赛中的某一支球队而言，设 G 为该队的入球数，F 为对方向该队犯规的次数。那么 G 与 F 独立（independent），且

$$\mathrm{P}(G=0),\ \ldots,\ \mathrm{P}(G=4) \;=\; \tfrac{1}{9},\ \tfrac{2}{9},\ \tfrac{3}{9},\ \tfrac{2}{9},\ \tfrac{1}{9}, \qquad \mathrm{P}(F=f)=\tfrac{1}{10} \quad (f=0,1,\ldots,9).$$

该队从入球得 $2G$ 分，从犯规得 F 分，合共 $2G+F$ 分。该队超过一半的得分来自犯规，当且仅当

$$F > \tfrac{1}{2}(2G+F) \iff F > 2G.$$

（假如该队一分也没有得到，则 $F=0=2G$，这种情况不应计算在内，而严格不等式正好把它排除了。）

给定 $G=g$，符合要求的 F 值为 $2g+1,\ldots,9$，即十个等可能的值当中的 $9-2g$ 个。因此，利用 F 与 G 的独立性，

$$\mathrm{P}(F>2G \mid G=g)=\frac{9-2g}{10}, \qquad g=0,1,2,3,4,$$

得出 $\frac{9}{10}, \frac{7}{10}, \frac{5}{10}, \frac{3}{10}, \frac{1}{10}$。根据全概率公式（law of total probability），

$$\mathrm{P}(F > 2G) = \frac{1\cdot 9 + 2\cdot 7 + 3\cdot 5 + 2\cdot 3 + 1\cdot 1}{9\times 10} = \frac{45}{90} = \frac{1}{2}.$$

答案如此简洁，是有原因的：入球数的分布关于 2 对称，而 $(9-2g)/10$ 是 g 的线性函数，所以它的平均值等于它在 $g=2$ 的值，即 $\frac{1}{2}$。

犯规次数的分布有没有改变?

一支球队的犯规次数，就是对手被犯规的次数，所以在旧规则下，每支球队在一场比赛中犯规 9 次的概率为 $\frac{1}{10}$。设 X 为 300 个球队场次当中，球队犯规 9 次的场次数目。我们检验

$$H_0:\ p = \tfrac{1}{10}\ \text{（没有改变）} \qquad \text{对} \qquad H_1:\ p < \tfrac{1}{10},$$

其中 p 是一支球队在一场比赛中犯规 9 次的概率。备择假设（alternative hypothesis）是单侧（one-sided）的，因为修改规则的目的，是令犯规次数偏高的情况变得*较不*可能出现。

> **想一想** 这里采用二项分布（binomial）模型是否合理?

我们需要这 300 个球队场次互相独立，而且每个场次的概率 p 都相同。题目已说明同一场比赛之内的独立性，而假设不同场次的比赛之间也互相独立，亦属合理。因此在 H_0 之下，$X \sim \mathrm{B}(300, \frac{1}{10})$，且

$$\mathrm{E}(X) = 30, \qquad \mathrm{Var}(X) = 300\times\tfrac{1}{10}\times\tfrac{9}{10} = 27, \qquad \text{标准差}\sqrt{27}\approx 5.20.$$

我们观察到 $X=12$，比期望值 30 少了 18，相当于约 3.5 个标准差（standard deviation）。采用正态近似（normal approximation）并加上连续性校正（continuity correction），

$$\mathrm{P}(X\le 12) \approx \Phi\left(\frac{12.5-30}{\sqrt{27}}\right) = \Phi(-3.37) \approx 0.0004.$$

精确的二项概率还要更小，约为 0.0001。无论采用哪一个数值，假如甚么都没有改变，这样极端的结果在一千个球季之中也出现不到一次。它远远超出常用的 5% 甚至 1% 显著性水平（significance level），所以我们拒绝 H_0。这*确实*是有力的证据，显示犯规次数的分布已经改变，而且是朝着所期望的方向改变：犯规九次的场次变得较为罕见。

有两点保留值得说明。这个检验只显示 P(9 次犯规) 下降了，并没有说明分布的其余部分如何改变（尽管那是最自然的解释）。此外，统计学也无法证明这个下降是由修改规则*造成*的，因为两个球季之间可能还有其他事情改变了。

验算

▶ 入球数的概率 $\frac{1}{9}(1+2+3+2+1)$ 之和为 1，理应如此。

- 以精确分数逐一列举全部 5×10 个数对 (g, f)，得出 $\mathrm{P}(F > 2G) = \frac{1}{2}$；而模拟 200,000 场比赛的结果为 0.4997。
- 直接把二项概率加起来，对于 $X \sim \mathrm{B}(300, 0.1)$ 得出 $\mathrm{P}(X \leq 12) = 9.2 \times 10^{-5}$。正态近似（$3.8 \times 10^{-4}$）与平均值为 30 的泊松近似（Poisson approximation）（1.7×10^{-4}）都一致显示这个概率极小。

常见错误

- 把 F 与 G 比较，而不是与 $2G$ 比较：一个入球值两分。
- 计算 $F \geq 2G$ 而不是 $F > 2G$。“超过一半”是严格的不等关系，而一分未得的球队，并没有任何得分来自犯规。
- 以 150 场比赛作为样本大小，而不是 300 个球队场次；或者在计算之前忘记写明零假设（null hypothesis）和二项分布模型。

STEP 之外

第二部分是对二项比例所作的假设检验（hypothesis test），而我们求得的尾部概率就是 p 值（p-value）。到了大学，你会学到似然比（likelihood ratio）检验和卡方检验（chi-squared test），它们会把观察所得的整个犯规次数分布与旧分布作比较，而不只是比较其中一个值。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）把概率方面的内容（条件概率、二项分布与正态近似）讲解得十分清楚。

详细解答

STEP I 1988，第 16 题

概率与统计 · 连续分布与条件化

> **STEP I 1988，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/88-s1-q16

关键思路

第一部分是一道归一化（normalisation）与积分的练习：题目只给出密度函数（density）到相差一个常数的程度，所以我们必须要求总概率为 1，从而求出这个常数；而 $t^2/(1+t^2)^2$ 的积分，只要把它拆成 $1/(1+t^2)$ 与 $t/(1+t^2)$ 的导数，便迎刃而解。其后的检验制度只是一棵简短的分情况树形图。只有车龄 4 和 5 时的检验才有影响（车龄 3 之前没有汽车会生锈），所以我们按汽车在第四次检验时是否已经生锈来分情况，而最后一部分就是在同一棵树形图上运用贝叶斯定理（Bayes' theorem）。

解答

开始生锈时间的分布

设汽车在车龄 $3+T$ 时开始生锈，其中 $T \geq 0$ 的密度函数为 $f(t) = k\,t^2/(1+t^2)^2$，k 为某常数。关键的积分来自以下观察：

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

只要把括号内的式子通分，以 $(1+t^2)^2$ 为公分母，便可验证：分子为 $(1+t^2)-(1-t^2) = 2t^2$。因此

$$\int_0^t \frac{s^2}{(1+s^2)^2}\,\mathrm{d}s = \frac{1}{2}\left(\arctan t - \frac{t}{1+t^2}\right).$$

当 $t \to \infty$ 时，$\arctan t \to \frac{\pi}{2}$ 而 $t/(1+t^2) \to 0$，所以在 $[0,\infty)$ 上的积分为 $\frac{\pi}{4}$。要使 f 成为密度函数，我们需要 $k \cdot \frac{\pi}{4} = 1$，故 $k = 4/\pi$。于是

$$F(t) = \mathrm{P}(\text{开始生锈时车龄小于} 3+t) = \mathrm{P}(T < t) = \frac{2}{\pi}\left(\arctan t - \frac{t}{1+t^2}\right), \quad t \geq 0.$$

（留意 $F(t) \to 1$：每辆汽车终究都会生锈。）

我们需要以下两个数值：

$$F(1) = \frac{2}{\pi}\left(\frac{\pi}{4} - \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{\pi} \approx 0.1817,$$
$$F(2) = \frac{2}{\pi}\left(\arctan 2 - \frac{2}{5}\right) \approx 0.4502.$$

第五周年检验不合格

车龄 3 之前没有汽车生锈，所以车龄 1、2 和 3 时的检验必定合格（恰好在第三次检验那一刻开始生锈的概率为零）。汽车要在车龄 5 时不合格，必须在车龄 4 时合格，然后在车龄 5 时不合格，而不合格的前提是车龄 5 时已经生锈。锈蚀永远不会消失，所以这件事有两种互斥（disjoint）的发生方式。

- 车龄 4 时已经生锈（概率为 $F(1)$）。它以概率 $\frac{1}{2}$ 通过第四次检验，然后在仍然生锈的情况下，以概率 $\frac{1}{2}$ 在第五次检验不合格。贡献：$\frac{1}{4}F(1)$。
- 在车龄 4 与 5 之间开始生锈（概率为 $F(2) - F(1)$）。它必定通过第四次检验，并以概率 $\frac{1}{2}$ 在第五次检验不合格。贡献：$\frac{1}{2}\big(F(2) - F(1)\big)$。

这里我们按题目的原意假设：一辆生锈汽车在各次检验的结果彼此独立（independent），亦与开始生锈的时间独立。

> ? **想一想** 为甚么我们不用乘上通过车龄 1、2 和 3 时各次检验的概率?

因为这些检验都在任何汽车有可能生锈之前进行，所以每辆汽车都会通过；该因子为 1。

把两种情况相加，

$$\begin{aligned}\mathrm{P}(\text{检验不合格时的车龄为5}) &= \tfrac{1}{4}F(1) + \tfrac{1}{2}\big(F(2) - F(1)\big) = \tfrac{1}{2}F(2) - \tfrac{1}{4}F(1)\\ &= \frac{1}{\pi}\left(\arctan 2 - \frac{2}{5}\right) - \frac{1}{8} + \frac{1}{4\pi}\\ &= \boxed{\frac{1}{\pi}\left(\arctan 2 - \frac{3}{20}\right) - \frac{1}{8} \approx 0.1797.}\end{aligned}$$

已知在第五次检验后被销毁，求第四次检验时已生锈的概率

一辆在第五次检验后随即被销毁的汽车，正正就是在该次检验不合格（而之前各次全部合格）的汽车，亦即刚才求得的事件。设 A 为汽车在第四次检验时已经生锈的事件，D 为它在第五次检验不合格的事件。上面第一点所述的正是 $A \cap D$，所以

$$\mathrm{P}(A \cap D) = \tfrac{1}{4}F(1) = \frac{1}{8} - \frac{1}{4\pi} = \frac{\pi - 2}{8\pi}.$$

根据条件概率（conditional probability）的定义，

$$\mathrm{P}(A \mid D) = \frac{\mathrm{P}(A \cap D)}{\mathrm{P}(D)} = \frac{\dfrac{\pi - 2}{8\pi}}{\dfrac{1}{\pi}\left(\arctan 2 - \dfrac{3}{20}\right) - \dfrac{1}{8}}.$$

分子和分母同乘以 8π，

$$\boxed{\mathrm{P}(A \mid D) = \frac{\pi - 2}{8\arctan 2 - \frac{6}{5} - \pi} = \frac{5(\pi - 2)}{40\arctan 2 - 6 - 5\pi} \approx 0.253.}$$

所以在五年时被销毁的汽车当中，只有约四分之一是在已经生锈的情况下侥幸通过第四次检验的；大部分都是在第五年内才生锈。

验算

- $F(0) = 0$，而且当 $t \to \infty$ 时 $F(t) \to 1$，符合分布函数（distribution function）应有的性质。对 $\frac{4}{\pi}\int_0^2 t^2/(1+t^2)^2\,\mathrm{d}t$ 使用辛普森法则（Simpson's rule）得 0.45018，与 $F(2)$ 吻合。
- 模拟 400 000 辆汽车（抽样得出开始生锈的时间，再进行每年的检验），得出五年时不合格的概率为 0.1789，条件概率为 0.255，而精确值分别为 0.1797 和 0.2528。
- 合理性检查：$\mathrm{P}(D)$ 必须小于 $\frac{1}{2}F(2) \approx 0.225$（即车龄五年时已生锈然后不合格的概率，当中不考虑较早被淘汰的情况），而事实的确如此。

常见错误

- 忘记归一化：密度函数只是与 $t^2/(1+t^2)^2$ 成正比，漏掉因子 $4/\pi$ 会令其后每个答案都出错。
- 把“车龄 5 时已生锈”当作概率为 $F(2)$ 的单一情况，然后乘以 $\frac{1}{2}$。这样做忽略了那些在车龄 4 时已生锈、并在该次检验被淘汰的汽车。
- 在最后一部分，以“车龄 5 时已生锈”而非“车龄 5 时不合格”作为条件。

STEP 之外

这里的函数 F 是一个分布函数，而比值 $f(t)/(1-F(t))$ 称为风险率（hazard rate），即已知汽车至今仍未生锈的条件下，瞬间生锈的可能性；它是存活分析（survival analysis）与可靠性理论（reliability theory）的核心对象。最后一部分是滤波（filtering）的一个小

例子：从不完美的观测（各次检验）推断隐藏的状态（是否生锈）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）正是以这种精神讲解连续分布与条件概率。

详细解答

STEP II 1988，第 1 题

纯数学 · 部分分式与级数

> **STEP II 1988，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/88-s2-q1

关键思路

把 f 分解成部分分式（partial fractions），整个问题便归结为把两个简单的分式 $\frac{1}{x-1}$ 和 $\frac{1}{x-2}$ 展开成等比级数（geometric series）。唯一真正要决定的，是每个分式应该朝哪个方向展开。等比级数 $\sum r^n$ 只在 $|r|<1$ 时收敛（converge），所以对每个分式，我们都把分母两项之中绝对值较大的一项提出来。当 $|x|$ 先后越过 1 和 2 时，这个选择便会改变，这就是为甚么会有三个不同的展开式。

解答

部分分式

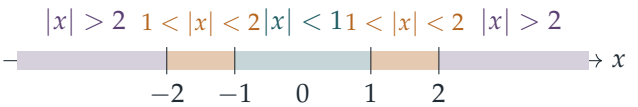
设 $\dfrac{1}{(x-1)(x-2)} = \dfrac{A}{x-1} + \dfrac{B}{x-2}$。由遮盖法（cover-up rule）得 $A = \frac{1}{1-2} = -1$ 及 $B = \frac{1}{2-1} = 1$，所以

$$\mathrm{f}(x) = \frac{1}{x-2} - \frac{1}{x-1}.$$

我们需要的两件基本工具都来自等比级数

$$\frac{1}{1-r} = \sum_{n=0}^{\infty} r^n \qquad (|r|<1),$$

其中 r 取 x、$\frac{x}{2}$、$\frac{1}{x}$ 或 $\frac{2}{x}$ 其中之一，视乎在所讨论的范围内哪一个的模（modulus）小于 1。



范围 $|x|<1$

这里 $|x|<1$，同时 $|\frac{x}{2}|<\frac{1}{2}<1$，所以两个分式都可以展开成 x 的幂：

$$-\frac{1}{x-1}=\frac{1}{1-x}=\sum_{n=0}^{\infty}x^n,$$
$$\frac{1}{x-2}=-\frac{1}{2}\cdot\frac{1}{1-\frac{x}{2}}=-\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^n.$$

两式相加，得

$$\mathrm{f}(x)=\sum_{n=0}^{\infty}x^n-\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^n\qquad(|x|<1),$$

正是所求。

范围 $1<|x|<2$

第二个分式维持不变：我们仍然有 $|\frac{x}{2}|<1$，所以 $\frac{1}{x-2}=-\frac{1}{2}\sum_{n\geq0}(\frac{x}{2})^n$，与之前完全一样。

> **想一想** 为甚么第一个分式不能继续使用 $\sum x^n$?

因为现在 $|x|>1$，各项 x^n 并不趋向零，所以 $\sum x^n$ 发散（diverge）。我们改为从分母提出较大的一项 x，余下的公比是 $\frac{1}{x}$，而 $|\frac{1}{x}|<1$：

$$-\frac{1}{x-1}=-\frac{1}{x}\cdot\frac{1}{1-\frac{1}{x}}=-\frac{1}{x}\sum_{n=0}^{\infty}x^{-n}=-\sum_{n=0}^{\infty}x^{-(n+1)}=-\sum_{n=1}^{\infty}x^{-n},$$

其中最后一步只是把 $n+1$ 重新记作 n。把两部分相加，得

$$\mathrm{f}(x)=-\sum_{n=1}^{\infty}x^{-n}-\frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^n\qquad(1<|x|<2),$$

正是所求。

范围 $|x|>2$

第一个分式沿用上一个范围的展开式，因为 $|\frac{1}{x}|<\frac{1}{2}<1$。但现在 $|\frac{x}{2}|>1$，所以第二个分式也要用同样的方法倒转过来展开：提出 x，余下公比 $\frac{2}{x}$，而 $|\frac{2}{x}|<1$：

$$\frac{1}{x-2}=\frac{1}{x}\cdot\frac{1}{1-\frac{2}{x}}=\frac{1}{x}\sum_{n=0}^{\infty}\left(\frac{2}{x}\right)^n=\sum_{n=0}^{\infty}2^nx^{-(n+1)}=\sum_{n=1}^{\infty}2^{n-1}x^{-n}.$$

因此，当 $|x|>2$ 时，

$$\mathrm{f}(x)=\sum_{n=1}^{\infty}2^{n-1}x^{-n}-\sum_{n=1}^{\infty}x^{-n}=\sum_{n=1}^{\infty}(2^{n-1}-1)x^{-n}.$$

$n=1$ 的系数是 $2^0-1=0$，所以这个级数其实由 x^{-2} 开始：

$$\boxed{\mathrm{f}(x)=\sum_{n=2}^{\infty}(2^{n-1}-1)x^{-n}=\frac{1}{x^2}+\frac{3}{x^3}+\frac{7}{x^4}+\frac{15}{x^5}+\cdots \qquad (|x|>2).}$$

（两个级数各自收敛，所以逐项合并是合理的。）无论写成两个级数的形式，还是合并后的形式，都是正确答案。

验算

- 主导项。当 $|x|$ 很大时，$\mathrm{f}(x)=\frac{1}{x^2-3x+2}\approx\frac{1}{x^2}$，与最后一个展开式的首项吻合。$x^{-1}$ 的系数等于零并非巧合：两个部分分式的 $\frac{1}{x}$ 项大小相等、符号相反。
- 代入一个值。当 $x=3$ 时，$\mathrm{f}(3)=\frac{1}{2\cdot1}=\frac{1}{2}$，而 $\sum_{n\geq2}(2^{n-1}-1)3^{-n}=\frac{1}{2}\sum_{n\geq2}(\frac{2}{3})^n-\sum_{n\geq2}(\frac{1}{3})^n=\frac{1}{2}\cdot\frac{4}{3}-\frac{1}{6}=\frac{1}{2}$。同样地，$\mathrm{f}(\frac{1}{2})=\frac{4}{3}$ 和 $\mathrm{f}(\frac{3}{2})=-4$ 亦与头两个展开式吻合。
- 数值上，在各自范围内随机选取数百个点（正负皆有），三个级数（各取 4000 项求和）与 f 的值都相符，误差约为 10^{-14}。

常见错误

- 在 $\frac{1}{x-2}=-\frac{1}{2}\cdot\frac{1}{1-x/2}$ 中弄错正负号：提出来的因子是 -2，不是 2。
- 乘以 $\frac{1}{x}$ 之后，求和指标差了一：$\frac{1}{x}\sum_{n\geq0}x^{-n}$ 由 x^{-1} 开始，所以重新标记后的和由 $n=1$ 开始。
- 没有检查公比的模是否小于 1，就在每个范围都使用同一个展开式。

STEP 之外

这三个展开式，是复分析（complex analysis）中洛朗级数（Laurent series）在实变量上的投影。对于复数 z，函数 $\frac{1}{(z-1)(z-2)}$ 在 $|z|<1$、$1<|z|<2$ 和 $|z|>2$ 这三个区域内各有不同的洛朗展开式：这三个区域是一个圆盘和两个圆环（annulus），以通过极点（pole）的圆为边界，而 $|x|=1$ 和 $|x|=2$ 在复平面上正好变成这些圆。这类展开式中 z^{-1} 的系数称为留数（residue），原来它们决定了围道积分（contour integral）的值。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）以优美的几何方式讲解了幂级数（power series）和洛朗级数。

详细解答

STEP II 1988，第 2 题

纯数学 · 对称代数方程

STEP II 1988，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/88-s2-q2

关键思路

先去分母，再把两条方程相减。右边的平方差（difference of two squares）会产生因式 $y-z$，正好与左边出现的 $y-z$ 对应，而假设 $y \neq z$ 恰恰让我们可以把它约去。此后，一切都归结为一条关于 x、y 和 z 的对称式（symmetric expression），余下的工作就交给对称性完成。

解答

证明 $x+y+z=a$

由于 $y \neq 0$ 及 $z \neq 0$，我们可以把两边乘开：

$$2ay - 3y^2 = (z-x)^2, \tag{1}$$
$$2az - 3z^2 = (x-y)^2. \tag{2}$$

由 (1) 减去 (2)，并在右边运用平方差，

$$\begin{aligned} 2a(y-z) - 3(y^2-z^2) &= (z-x)^2 - (x-y)^2 \\ &= \big[(z-x)-(x-y)\big]\big[(z-x)+(x-y)\big] \\ &= (y+z-2x)(z-y). \end{aligned}$$

左边是 $(y-z)\big[2a-3(y+z)\big]$，所以

$$(y-z)\big[2a-3(y+z)\big] = -(y-z)(y+z-2x).$$

想一想 究竟在哪一步需要用到 $y \neq z$?

就在这里：因为 $y - z \neq 0$，我们可以把它除去，得到

$$2a - 3y - 3z = 2x - y - z \quad \Longrightarrow \quad 2a = 2x + 2y + 2z,$$

所以 $x + y + z = a$，正如所求。

证明第三条方程

把 $a = x + y + z$ 代回 (1)：

$$y(2x + 2y + 2z - 3y) = (z - x)^2 \implies 2xy - y^2 + 2yz = z^2 - 2zx + x^2,$$

整理后得

$$x^2 + y^2 + z^2 - 2xy - 2yz - 2zx = 0. \tag{3}$$

最引人注目的是，(3) 对 x、y 和 z 是对称的。现在看看我们要证明甚么。由于 $x \neq 0$，方程 $2a - 3x = \dfrac{(y-z)^2}{x}$ 等价于

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

而代入 $a = x + y + z$ 后，这就是

$$x(2y + 2z - x) = y^2 - 2yz + z^2 \iff x^2 + y^2 + z^2 - 2xy - 2yz - 2zx = 0,$$

这正是 (3)。因此

$$2a - 3x = \frac{(y-z)^2}{x}.$$

(事实上，三条方程都只是 (3) 连同 $a = x + y + z$，各自从其中一个变量的角度写出来而已。)

当 $y = z$ 时，最后一条方程可以成立吗?

可以。若 $y = z$，题目给出的两条方程变成同一条方程 $2a - 3y = \dfrac{(y-x)^2}{y}$，而最后一条方程则变成 $2a - 3x = 0$，即 $a = \frac{3}{2}x$。把 $a = \frac{3}{2}x$ 代入第一条方程，

$$3x - 3y = \frac{(x-y)^2}{y} \iff 3y(x - y) = (x - y)^2 \iff (x - y)(x - 4y) = 0.$$

所以，举例说，$x = y = z = 1$ 及 $a = \frac{3}{2}$ 满足全部三条方程：

$$2a - 3y = 0 = \frac{(z-x)^2}{y}, \qquad 2a - 3z = 0 = \frac{(x-y)^2}{z}, \qquad 2a - 3x = 0 = \frac{(y-z)^2}{x}.$$

另一个例子是 $x=4$、$y=z=1$、$a=6$：这时头两条方程都有 $2a-3y=9=\frac{(1-4)^2}{1}$，而 $2a-3x=0=\frac{0^2}{4}$。

所以，最后一条方程并非仅当 $y\neq z$ 时才成立：条件 $y\neq z$ 保证它成立，但当 $y=z$ 时它也可以成立。当 $y=z$ 时失去的是这个保证。例如 $x=1$、$y=z=2$、$a=\frac{13}{4}$ 满足题目给出的两条方程（两边各等于 $\frac{1}{2}$），但 $2a-3x=\frac{7}{2}\neq 0$。还要留意，第一个例子中 $x+y+z=3\neq a$，所以当 $y=z$ 时，连第一个结论也可能不成立。

验算

- 取 $x=9$、$y=4$、$z=1$，它们满足 (3)：$81+16+1-72-8-18=0$。这时 $a=14$，而 $2a-3y=16=\frac{(1-9)^2}{4}$、$2a-3z=25=\frac{(9-4)^2}{1}$ 及 $2a-3x=1=\frac{(4-1)^2}{9}$。三条方程全部成立。
- 一次数值检验在 10,000 组随机有理数三元组上精确地验证了恒等式 $(1)-(2)=(y-z)\big(2a-2(x+y+z)\big)$，并在 (3) 上约 1,900 个满足 $y\neq z$ 的随机点确认了全部三条方程，另加上文三个 $y=z$ 的例子。

常见错误

- 除以 $y-z$ 而不说明为何可以这样做。题目之所以给出这个假设，全是为了这一步。
- 最后一部分只答“不可以”或“可以”而没有举例。只要明确给出一组 $y=z$ 的三元组，问题便解决了。
- 以为 $y=z$ 会令最后一条方程不成立。它只是令保证消失而已。

STEP 之外

方程 (3) 是一个二次型（quadratic form），它的零点集是 (x,y,z) 空间中的一个双锥面（double cone）。若 $x,y,z>0$，设 $x=p^2$、$y=q^2$、$z=r^2$：则 (3) 可因式分解为

$$-(p+q+r)(-p+q+r)(p-q+r)(p+q-r)=0,$$

这是边长为 p,q,r 的三角形面积平方的 16 倍（希罗公式，Heron's formula）。所以 (3) 的意思是 $\sqrt{x},\sqrt{y},\sqrt{z}$ 构成一个退化三角形（degenerate triangle），例如上面验算中的 $\sqrt{9}=\sqrt{4}+\sqrt{1}$。二次型及其分类是大学一年级线性代数（linear algebra）的标准课题；参见 Strang，*Introduction to Linear Algebra*（线性代数入门教科书）。

详细解答

STEP II 1988，第 3 题

纯数学 • 根的对称性

> **STEP II 1988，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/88-s2-q3

关键思路

不要先从系数入手，而应先考虑根的集合。题目的性质是说，根的集合在运算 $k \mapsto 1/k$ 之下保持不变，所以两个根要么互换，要么各自不动。加入第二个运算 $k \mapsto 1-k$ 后，同样的推理可以再用一次。至于三次方程，把这两个运算反复施行，由一个根最多可得出六个值，而三次方程只容得下三个根：因此各根必须落在特殊的点上，使六个值之中有些重合。

解答

倒数性质

设两根为 α 和 β（可能相等，也可能是复数），则 $\alpha+\beta=-b$ 及 $\alpha\beta=c$。首先，0 不可能是根，因为 0^{-1} 不存在；所以 $c\neq 0$。

映射（map）$k \mapsto 1/k$ 必须把每个根送到一个根，所以 $1/\alpha \in \{\alpha, \beta\}$。

- 若 $1/\alpha=\beta$，则 $\alpha\beta=1$，即 $c=1$。反过来，若 $c=1$，则两根皆非零，且 $\beta=1/\alpha$，所以每个根的倒数（reciprocal）就是另一个根，而这性质对任何实数 b 都成立。（这包括 $b=\mp 2$ 时的重根（repeated root）± 1，以及 $|b|<2$ 时位于单位圆（unit circle）上的复数根。）
- 若 $1/\alpha=\alpha$ 但 $1/\alpha\neq\beta$，则 $\alpha=\pm 1$ 且 $\alpha\beta\neq 1$。把这性质应用于 β：它的倒数不是 α（因为 $\alpha\beta\neq 1$），所以 $1/\beta=\beta$，于是 $\beta=\pm 1$ 亦然。由于 $\alpha\beta\neq 1$，两根必定是 1 和 -1，得出 $x^2-1=0$：$b=0$，$c=-1$。这方程确实具有该性质。

所以该性质成立，当且仅当

$$c = 1 \ (b \text{ 为任何实数}), \qquad \text{或} \qquad b = 0,\ c = -1.$$

加入性质 $k \mapsto 1-k$

以 $1-k$ 取代 $1/k$，重复同样的论证：要么 $1-\alpha = \beta$，于是 $\alpha+\beta = 1$；要么两根都不动，即 $1-\alpha = \alpha$ 及 $1-\beta = \beta$，于是 $\alpha = \beta = \frac{1}{2}$，同样有 $\alpha+\beta = 1$。无论哪种情况都有 $b = -1$；反过来，$b = -1$ 使两根成为 α 和 $1-\alpha$，所以该性质成立。

$b = 0$、$c = -1$ 这情况不符合此条件，所以我们需要 $c = 1$ 及 $b = -1$：

$$b = -1, \qquad c = 1,$$

得出 $x^2 - x + 1 = 0$，其根为 $\frac{1}{2} \pm \frac{\sqrt{3}}{2}\mathrm{i}$。每个根的模（modulus）都是 1，实部都是 $\frac{1}{2}$，所以 $1/k$ 和 $1-k$ 都等于共轭复数（complex conjugate）$\bar{k}$，即另一个根。

三次方程

现设 $x^3 + px^2 + qx + r = 0$ 的系数为实数，并同时具有两个性质。一如之前，0 不是根，1 也不是（否则 $1 - 1 = 0$ 便会是根）。

实系数三次方程至少有一个实根，称之为 a。反复施行这两个运算，以下每一个都是根：

$$a, \quad \frac{1}{a}, \quad 1-a, \quad \frac{1}{1-a}, \quad 1-\frac{1}{a} = \frac{a-1}{a}, \quad \frac{1}{(a-1)/a} = \frac{a}{a-1}.$$

（第四个是第三个的倒数，第五个是 1 − 第二个，第六个是第五个的倒数。）这是六个实数，但三次方程最多只有三个相异的根，所以其中必有两个重合。

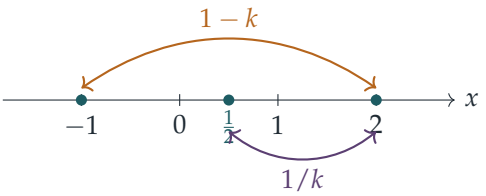
> ? **想一想** “其中两个相等”这句话，怎样确定 a 的值?

把这六个映射称为 $f_1, \ldots, f_6$，其中 f_1 为恒等映射（identity）。每一个映射的逆映射（inverse）都在这列表之中：$\frac{1}{k}$、$1-k$ 和 $\frac{k}{k-1}$ 各自是自己的逆映射，而 $\frac{1}{1-k}$ 和 $\frac{k-1}{k}$ 则互为逆映射。还可以验证，把任意两个映射复合（compose），所得的仍是列表中的映射。所以若 $f_i(a) = f_j(a)$ 而 $i \neq j$，施以 f_i^{-1} 便得 $a = g(a)$，其中 $g = f_i^{-1} \circ f_j$ 是六个映射之一，且不是恒等映射。所以 a 是五个非恒等映射其中之一的不动点（fixed point）：

$$\frac{1}{a} = a \iff a = \pm 1, \qquad \frac{1}{1-a} = a \iff a^2 - a + 1 = 0,$$
$$1-a = a \iff a = \tfrac{1}{2}, \qquad \frac{a-1}{a} = a \iff a^2 - a + 1 = 0,$$
$$\frac{a}{a-1} = a \iff a(a-2) = 0.$$

二次式 $a^2 - a + 1$ 没有实根（它就是先前的那个二次式），而 $a \neq 0, 1$。因此 $a \in \{-1, \frac{1}{2}, 2\}$。

无论 a 是其中哪一个，上面六个值恰好取遍 -1、2 和 $\frac{1}{2}$：映射 $1-k$ 把 -1 和 2 互换，并保持 $\frac{1}{2}$ 不动；而 $1/k$ 则把 2 和 $\frac{1}{2}$ 互换，并保持 -1 不动。



所以 -1、2 和 $\frac{1}{2}$ 是该三次方程的三个相异根。首一（monic）三次多项式由它的三个根完全决定，所以它必定是

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

反过来，这个三次方程确实具有两个性质，因为它的根集 $\{-1, \frac{1}{2}, 2\}$ 在 $k \mapsto 1/k$ 和 $k \mapsto 1-k$ 之下都被送回自身。所以

$$x^3 - \tfrac{3}{2}x^2 - \tfrac{3}{2}x + 1 = 0$$

是唯一符合条件的三次方程，正如题目所求。

验算

代入 $x=2$：$8-6-3+1=0$；$x=-1$：$-1-\frac{3}{2}+\frac{3}{2}+1=0$。各根之积为 $-1\cdot 2\cdot\frac{1}{2} = -1 = -r$，各根之和为 $\frac{3}{2} = -p$。留意这三次方程的系数，除正负号外，顺读和倒读都一样（$1, -\frac{3}{2}, -\frac{3}{2}, 1$）：这种回文（palindromic）形状正是 $k \mapsto 1/k$ 所必然导致的。以电脑测试 20,000 组随机的 (b, c)，结果印证了倒数条件；而随机选取远离 $0, \pm 1, \frac{1}{2}, 2$ 的实数 a，每次都得出六个相异的值。

常见错误

- 第一部分只答“$c = 1$”：每个根都是自身倒数的那种情况，还会多出 $b = 0$、$c = -1$。
- 忽略了 $\frac{1}{2} \pm \frac{\sqrt{3}}{2}\mathrm{i}$ 这一对根，它在两个映射之下都是封闭的。正是由实根出发，才能把它排除：实系数三次方程必有一个实根，而该实根的六个值不可能包含这两个数。
- 只证明了题目给出的三次方程符合条件；题目要求的是证明它是唯一的一个。

STEP 之外

六个映射 $k \mapsto k,\ 1/k,\ 1-k,\ \frac{1}{1-k},\ \frac{k-1}{k},\ \frac{k}{k-1}$ 在复合运算下构成一个群（group），它是对称群（symmetric group）S_3 的一个复本；它置换 $0, 1, \infty$ 这三个点，而那六个值就是四个

点的交比（cross-ratio）所有可能的取值。计数的那一步正是轨道－稳定子定理（orbit–stabiliser theorem）：只有在被某个非恒等元素固定的点上，轨道（orbit）的长度才会少于六。Allenby 的 *Rings, Fields and Groups*（环、域与群）阐述群在集合上的作用，而 Needham 的 *Visual Complex Analysis*（复分析的视觉化讲解）则把这些映射展示为平面上的莫比乌斯变换（Möbius transformation）。

详细解答

STEP II 1988，第 4 题

纯数学 · 阿根图上的轨迹

> **STEP II 1988，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/88-s2-q4

关键思路

把复数写成 $a+\mathrm{i}b$，再把“实数”或“虚数”这个条件化为一条关于 a 和 b 的实方程。这样便会在阿根图（Argand diagram）上得到一条熟悉的曲线。接着，把平方看成平面上的一个映射（map），将 x 和 y（或 p 和 q）用 a 和 b 表示，然后消去参数，便可找出该曲线被映射到哪里。

关于题目的印刷文字，要说明一点：原文写的是 w^2-2x，但 x 在该处尚未引入，而最后部分（出现 t^2-2t）显示了题目原意的条件。我们把它理解为w^2-2w 是实数。

解答

第 (i) 部分

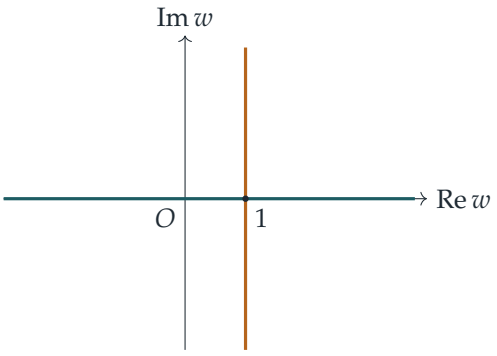
设 $w=a+\mathrm{i}b$，其中 a,b 为实数。则

$$w^2-2w=(a^2-b^2-2a)+\mathrm{i}\,(2ab-2b)=(a^2-b^2-2a)+2\mathrm{i}\,b(a-1).$$

这个数是实数，当且仅当其虚部（imaginary part）为零，即

$$b(a-1)=0,\qquad \text{即}\qquad b=0\ \text{或}\ a=1.$$

因此 w 的轨迹（locus）是实轴（real axis）连同垂直线 $\mathrm{Re}\,w=1$。两条直线相交于 $w=1$。



第 (ii) 部分

我们把两条直线分开来逐一平方。

实轴。若 $b=0$，则 $w=a$ 是实数，而 $w^2=a^2$。所以 $y=0$ 及 $x=a^2\geq 0$，而当 a 取遍所有实数时，a^2 会取遍 $[0,\infty)$ 内的每一个值。轨迹的这一部分是非负的 x 轴：即半直线（half-line）$y=0$，$x\geq 0$。

直线 $a=1$。若 $w=1+\mathrm{i}b$，则

$$w^2=1-b^2+2\mathrm{i}b, \qquad \text{所以} \qquad x=1-b^2, \quad y=2b.$$

利用 $b=\frac{1}{2}y$ 消去参数，得

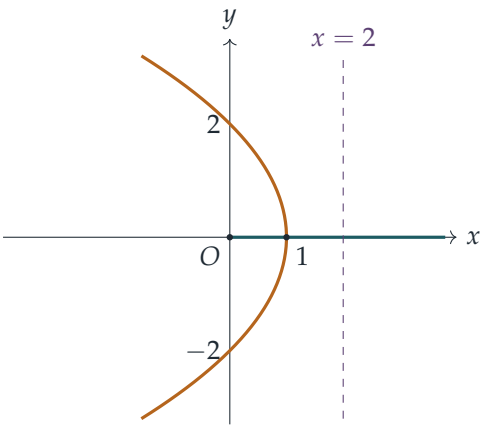
$$x=1-\tfrac{1}{4}y^2, \qquad \text{或} \qquad y^2=4(1-x)=-4(x-1).$$

只要取 $b=\frac{1}{2}y$，这条曲线上的每一点都能取到。它是一条抛物线（parabola），其轴沿 x 轴，顶点（vertex）在 $(1,0)$，开口朝向 x 的负方向。与标准式 $Y^2=4cX$ 比较，这里 $c=-1$，而 $X=x-1$。所以焦点（focus）在 $X=-1$，即原点；准线（directrix）是 $X=1$，即直线 $x=2$。事实上 $x^2+y^2=x^2+4-4x=(2-x)^2$，所以曲线上每一点与 O 的距离都等于它与直线 $x=2$ 的距离。抛物线与 y 轴相交于 $(0,\pm 2)$。

? **想一想** 为甚么这条半直线不会延伸到 x 的负值?

w 平面上的实轴被覆盖了两次（分别由 a 和 $-a$），但实数的平方永远不会是负数。所以整条实轴对折到非负的那一半上。同样地，直线 $\mathrm{Re}\,w=1$ 只覆盖抛物线一次，因为 $b\mapsto 2b$ 是一对一（one-to-one）的。

所以 (x,y) 的轨迹是抛物线 $y^2=4(1-x)$（焦点在原点，准线为 $x=2$），连同半直线 $y=0$，$x\geq 0$。这条半直线由焦点出发，穿过顶点 $(1,0)$；该顶点是 $w=1$ 的像，亦即第 (i) 部分两条直线的交点的像。



最后部分

设 $t = a + \mathrm{i}b$。这里的“虚数”是指纯虚数（purely imaginary），所以我们需要 $t^2 - 2t$ 的实部（real part）为零：

$$a^2 - b^2 - 2a = 0, \qquad \text{即} \qquad (a-1)^2 - b^2 = 1.$$

所以 t 位于一条以 1 为中心的等轴双曲线（rectangular hyperbola）上，其顶点在 $t = 0$ 和 $t = 2$。在这条曲线上，$b^2 = a^2 - 2a = a(a-2) \geq 0$，所以 $a \leq 0$ 或 $a \geq 2$。

现在由 $t^2 = p + \mathrm{i}q$ 得

$$p = a^2 - b^2 = 2a, \qquad q = 2ab,$$

其中我们用了双曲线的方程 $a^2 - b^2 = 2a$ 来化简 p。所以 $a = \frac{1}{2}p$，于是

$$q^2 = 4a^2b^2 = 4a^2(a^2 - 2a) = p^2\left(\tfrac{1}{4}p^2 - p\right), \qquad \text{所以} \qquad 4q^2 = p^3(p-4).$$

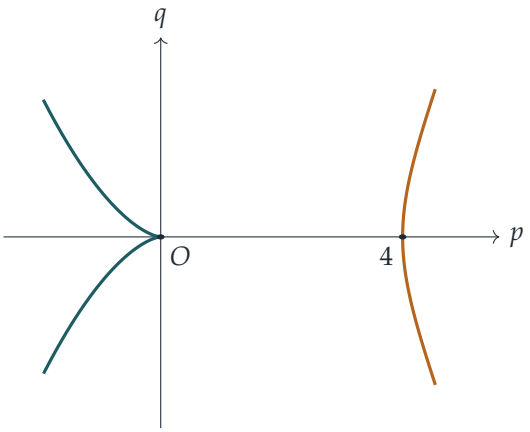
条件 $a \leq 0$ 或 $a \geq 2$ 变成 $p \leq 0$ 或 $p \geq 4$，这正好就是 $p^3(p-4) \geq 0$ 成立的范围。反过来，在这条曲线上任取一点。若 $p \neq 0$，设 $a = \frac{1}{2}p$ 及 $b = q/p$。则 $b^2 = q^2/p^2 = \frac{1}{4}p(p-4) = a^2 - 2a$，所以 t 在双曲线上，而且 $t^2 = (a^2 - b^2) + 2\mathrm{i}ab = p + \mathrm{i}q$。若 $p = 0$，则 $q = 0$，而这一点来自 $t = 0$。所以轨迹就是整条曲线

$$4q^2 = p^3(p-4).$$

它的形状：

- 它对 p 轴对称，因为方程中只出现 q^2。
- 在 $0 < p < 4$ 的范围内没有任何点。曲线分为两段，分别来自双曲线的两支。
- 右边一段由 $(4, 0)$ 开始，该点是 $t = 2$ 的像。在该点附近 $q^2 \approx 16(p-4)$，所以切线（tangent）是垂直的。

- 左边一段通过原点，即 $t = 0$ 的像。在该点附近 $q^2 \approx -p^3$，所以 $|q| \approx |p|^{3/2}$，形成一个指向右方的尖点（cusp），两臂都与 p 轴相切。
- 当 $|p|$ 很大时，$|q| \approx \frac{1}{2}p^2$，所以两段都像抛物线般向外张开。



（图中纵向比例经过压缩。蓝绿色的一段来自双曲线上 $\operatorname{Re} t \le 0$ 的一支，琥珀色的一段来自 $\operatorname{Re} t \ge 2$ 的一支。）

验算

- 第 (i) 部分：$w = 1 + 2\mathrm{i}$ 给出 $w^2 - 2w = (-3 + 4\mathrm{i}) - (2 + 4\mathrm{i}) = -5$，这是实数。第 (ii) 部分：它的平方 $-3 + 4\mathrm{i}$ 满足 $16 = 4(1 - (-3))$。
- 最后部分：$t = -1 + \mathrm{i}\sqrt{3}$ 满足 $1 - 3 + 2 = 0$。它的平方是 $-2 - 2\sqrt{3}\,\mathrm{i}$，而 $4q^2 = 48 = (-8)(-6) = p^3(p - 4)$。
- 我们以数千个随机点作了数值检验，确认每个轨迹都映射到所述的曲线上，而且每条曲线上的每一点都能取到。

常见错误

- 在第 (i) 部分忘记了实轴，以致在第 (ii) 部分漏掉那条半直线。又或者画了整条 x 轴，而不是只画 $x \ge 0$ 的部分。
- 没有利用 $a^2 - b^2 = 2a$ 来得出 $p = 2a$。不用这一步的话，消去参数的过程会繁复得多。
- 在最后的草图中漏掉了 $0 < p < 4$ 的空隙，或原点处的尖点。

STEP 之外

平方映射 $z \mapsto z^2$ 是最简单的保角映射（conformal map）。它把角度加倍，这就是为甚么这里的直线和双曲线会变成抛物线和带尖点的曲线。在 $z = 0$ 处导数为零，角度不再保持不变，尖点正是由此而来。在流体流动和静电学中，有一个标准技巧，就是利用这类映射把难处理的区域变成容易处理的区域。Needham 的 *Visual Complex Analysis*（《视觉复分析》）以图像把这些映射解释得十分精彩。

详细解答

STEP II 1988，第 5 题

纯数学 · 单位根与正切

STEP II 1988，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/88-s2-q5

关键思路

设 $z = \cos\theta + \mathrm{i}\sin\theta$，并用二项式定理（binomial theorem）展开 z^7。其虚部（imaginary part）是 $\sin 7\theta$，把它除以 $\cos^7\theta$ 后，就变成 $t = \tan\theta$ 的多项式。七次单位根（roots of unity）令 $\sin 7\theta$ 等于零，于是我们毫不费力就得到一个 7 次多项式的七个根。之后，正切之积不过是各根之积，可以直接从常数项读出。

解答

六次方程的根

由棣美弗定理（de Moivre's theorem），$(\cos\theta + \mathrm{i}\sin\theta)^7 = \cos 7\theta + \mathrm{i}\sin 7\theta$。设 $c = \cos\theta$，$s = \sin\theta$。用二项式定理展开左方，并只保留含 $\mathrm{i}s$ 奇数次幂的项，得

$$\sin 7\theta = 7c^6 s - 35c^4 s^3 + 21c^2 s^5 - s^7.$$

只要 $\cos\theta \neq 0$，我们就可以除以 c^7，并设 $t = \tan\theta$：

$$\frac{\sin 7\theta}{\cos^7\theta} = 7t - 35t^3 + 21t^5 - t^7 = -t\,(t^6 - 21t^4 + 35t^2 - 7). \qquad (*)$$

$z^7 = 1$ 的根是 $z = \cos\theta + \mathrm{i}\sin\theta$，其中 $\theta = \frac{2k\pi}{7}$，而对这些角有 $\sin 7\theta = \sin 2k\pi = 0$，即 $z^7 = 1$ 的虚部。改用 $\theta = \frac{k\pi}{7}$，$k = 0, 1, \ldots, 6$，会较为简洁，这同样给出 $\sin 7\theta = \sin k\pi = 0$（而且得到的正切值集合相同，因为 tan 的周期是 π）。

对上述每一个 θ：

- $\cos\theta \neq 0$，因为 $\frac{k\pi}{7} = \frac{\pi}{2}$ 需要 $2k = 7$；所以 $(*)$ 成立；

► $\sin 7\theta = 0$，所以由 $(*)$ 可知，$t = \tan\frac{k\pi}{7}$ 这个数是 $p(t) = t^7 - 21t^5 + 35t^3 - 7t$ 的根。

$\tan\frac{k\pi}{7}$ $(k = 0, \ldots, 6)$ 这七个值互不相同：各角 $0, \frac{\pi}{7}, \ldots, \frac{6\pi}{7}$ 是 $[0, \pi)$ 中互不相同的点，没有一个等于 $\frac{\pi}{2}$，而 tan 在 $[0, \pi) \setminus \{\frac{\pi}{2}\}$ 上是一对一（one-to-one）的（在 $[0, \frac{\pi}{2})$ 上为正且递增，在 $(\frac{\pi}{2}, \pi)$ 上为负，同样递增，两段的取值范围分别是 $[0, \infty)$ 和 $(-\infty, 0)$）。

? **想一想** 我们怎样知道已经找到所有的根，而不只是其中一部分？

一个 7 次多项式最多有 7 个根。我们已找到 7 个互不相同的根，所以这就是全部的根。$k = 0$ 给出 $t = 0$，即来自因式 t 的那个根。因此 $t^6 - 21t^4 + 35t^2 - 7 = 0$ 的根恰好是

$$t = \tan\frac{k\pi}{7}, \qquad k = 1, 2, 3, 4, 5, 6,$$

也就是 $\pm\tan\frac{\pi}{7}$、$\pm\tan\frac{2\pi}{7}$、$\pm\tan\frac{3\pi}{7}$，因为 $\tan\frac{(7-k)\pi}{7} = -\tan\frac{k\pi}{7}$。

乘积 $\tan\frac{2\pi}{7}\tan\frac{4\pi}{7}\tan\frac{6\pi}{7}$

对于偶数次的首一多项式（monic polynomial），各根之积等于常数项。所以

$$\left(\tan\tfrac{\pi}{7}\right)\left(-\tan\tfrac{\pi}{7}\right)\left(\tan\tfrac{2\pi}{7}\right)\left(-\tan\tfrac{2\pi}{7}\right)\left(\tan\tfrac{3\pi}{7}\right)\left(-\tan\tfrac{3\pi}{7}\right) = -7,$$

亦即

$$\left(\tan\tfrac{\pi}{7}\tan\tfrac{2\pi}{7}\tan\tfrac{3\pi}{7}\right)^2 = 7.$$

三个角都在 $(0, \frac{\pi}{2})$ 内，所以每个正切值都是正数，而乘积就是正平方根：$\tan\frac{\pi}{7}\tan\frac{2\pi}{7}\tan\frac{3\pi}{7} = \sqrt{7}$。

现在利用 $\tan(\pi - x) = -\tan x$：$\tan\frac{4\pi}{7} = -\tan\frac{3\pi}{7}$ 及 $\tan\frac{6\pi}{7} = -\tan\frac{\pi}{7}$。两个负号互相抵消，所以

$$\tan\frac{2\pi}{7}\tan\frac{4\pi}{7}\tan\frac{6\pi}{7} = \tan\frac{2\pi}{7}\tan\frac{3\pi}{7}\tan\frac{\pi}{7} = \sqrt{7},$$

正如所求。

$n = 9$ 和 $n = 11$ 的情况

同样的论证适用于任何奇数 $n = 2m + 1$。取 $(\cos\theta + \mathrm{i}\sin\theta)^n$ 的虚部，再除以 $\cos^n\theta$，得

$$\frac{\sin n\theta}{\cos^n\theta} = \sum_{r=0}^{m}(-1)^r\binom{n}{2r+1}t^{2r+1} = t\,q_n(t),$$

其中 q_n 的次数是 $2m = n - 1$，首项系数（leading coefficient）是 $(-1)^m$，常数项是 $\binom{n}{1} = n$。具体写出来，

$$q_9(t) = t^8 - 36t^6 + 126t^4 - 84t^2 + 9,$$
$$q_{11}(t) = -t^{10} + 55t^8 - 330t^6 + 462t^4 - 165t^2 + 11.$$

与之前完全一样（n 是奇数，所以没有任何 $\frac{k\pi}{n}$ 等于 $\frac{\pi}{2}$），q_n 的根就是 $n-1$ 个互不相同的数 $\tan\frac{k\pi}{n}$，$k=1,\dots,n-1$，它们可配对成 $\pm\tan\frac{k\pi}{n}$，其中 $k=1,\dots,m$。由于次数是偶数，q_n 各根之积是（常数项）/（首项系数）：

$$(-1)^m\prod_{k=1}^{m}\tan^2\frac{k\pi}{n}=\frac{n}{(-1)^m}\quad\Longrightarrow\quad\prod_{k=1}^{m}\tan\frac{k\pi}{n}=\sqrt{n},$$

这里取正根，因为这些角全都在 $(0,\frac{\pi}{2})$ 内。当 $n=9$ 时，这就是 q_9 的常数项 9 除以首项系数 1；当 $n=11$ 时，则是 $11/(-1)=-11=(-1)^5\prod\tan^2$。无论哪种情况，$\tan\frac{k\pi}{n}$ 对 $k=1,\dots,m$ 的乘积，在 $n=9$ 时是 3，在 $n=11$ 时是 $\sqrt{11}$。

> **想一想** 题目所求的乘积，其角是 $\frac{2\pi}{n},\frac{4\pi}{n},\dots$，而不是 $\frac{\pi}{n},\frac{2\pi}{n},\dots$。我们怎样由一个转到另一个？正负号又会怎样变化？

每个偶数倍 $2j\le m$ 本来就在 $1,\dots,m$ 这列数之中。每个偶数 $2j>m$（即 $2j>\frac{n}{2}$）给出 $\tan\frac{2j\pi}{n}=-\tan\frac{(n-2j)\pi}{n}$，而 $n-2j$ 是奇数，且不超过 m。由 1 到 m 的每个数都恰好这样出现一次，因为偶数来自第一类，奇数来自第二类。所以所求的乘积是 $\pm\sqrt{n}$，而介乎 $\frac{n}{2}$ 与 n 之间的每个偶数各带来一个负号。

- $n=9$：偶数有 $2,4,6,8$；其中大于 4.5 的是 $6,8$，共两个负号。所以
$$\tan\frac{2\pi}{9}\tan\frac{4\pi}{9}\tan\frac{6\pi}{9}\tan\frac{8\pi}{9}=\sqrt{9}=3.$$
- $n=11$：偶数有 $2,4,6,8,10$；其中大于 5.5 的是 $6,8,10$，共三个负号。所以
$$\tan\frac{2\pi}{11}\tan\frac{4\pi}{11}\tan\frac{6\pi}{11}\tan\frac{8\pi}{11}\tan\frac{10\pi}{11}=-\sqrt{11}.$$

（当 $n=7$ 时，大于 3.5 的偶数是 $4,6$，共两个负号，所以该乘积是正数。）

验算

- $n=9$ 有一个简单的局部验算：$\tan\frac{6\pi}{9}=\tan\frac{2\pi}{3}=-\sqrt{3}$，而数值上 $\tan\frac{2\pi}{9}\tan\frac{4\pi}{9}\tan\frac{8\pi}{9}\approx 0.8391\times 5.6713\times(-0.3640)\approx -1.732=-\sqrt{3}$，所以总乘积是 3。
- 用计算机计算，$\tan\frac{2\pi}{7}\tan\frac{4\pi}{7}\tan\frac{6\pi}{7}\approx 2.645751$，而 $\sqrt{7}\approx 2.645751$；$n=11$ 的乘积是 $\approx -3.316625=-\sqrt{11}$。
- 代入 $\tan\frac{\pi}{7}\approx 0.4816$，得 $t^6-21t^4+35t^2-7\approx 0$，与预期相符。

常见错误

- 除以 $\cos^7\theta$ 时没有检查它不等于零，或者没有论证七个正切值互不相同。缺少这几点，就未能证明已经找到所有的根。

- 把 $n = 11$ 的正负号弄错：乘积的平方很容易求得，但正负号要靠数清楚有多少个负的正切值来决定。
- 忘记 $a_d t^d + \cdots + a_0$ 各根之积是 $(-1)^d a_0 / a_d$；这在 $n = 11$ 时尤其重要，因为其首项系数是 -1。

STEP 之外

$\tan \frac{k\pi}{n}$ 这些数都是代数数（algebraic number），而多项式 q_n（当 n 是质数时）就是它们在有理数上的最小多项式（minimal polynomial）。像 $\prod \tan \frac{k\pi}{n} = \sqrt{n}$ 这类恒等式，其实是高斯和（Gauss sums）的影子；高斯和说明 $\sqrt{\pm p}$ 位于由 p 次单位根生成的域（field）之中，这是代数数论（algebraic number theory）的一块基石。Davenport 的 *The Higher Arithmetic* 是进入这一系列想法的平易入门读物，而 Needham 的 *Visual Complex Analysis* 则让人从几何角度感受单位根。

详细解答

STEP II 1988，第 6 题

纯数学 · 双曲函数不等式

STEP II 1988，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/88-s2-q6

关键思路

三个函数在 $x = 0$ 时的值都是零，所以只需证明它们在 $x > 0$ 时递增（increasing）便可。题目中各导数（derivative）是刻意安排的，使每一部分都为下一部分铺路：(ii) 的导数包含 (i) 的函数，而 (iii) 的导数则是 (ii) 经重新缩放后的版本。至于 f，对数微分法（logarithmic differentiation）把 f′ 的正负号化为 (iii) 中函数的正负号，这显然正是设置该部分的用意。

解答

以下我们一直使用这个原理：若 $F(0) = 0$ 且对所有 $x > 0$ 都有 $F'(x) > 0$，则对所有 $x > 0$ 都有 $F(x) > 0$。（根据中值定理（mean value theorem），存在某个 $\xi \in (0, x)$ 使得 $F(x) = F(x) - F(0) = xF'(\xi)$。）

第 (i) 部分

设 $g(x) = x - \tanh x$。则 $g(0) = 0$，且

$$g'(x) = 1 - \operatorname{sech}^2 x = \tanh^2 x,$$

这里用了 $1 - \tanh^2 x = \operatorname{sech}^2 x$。当 $x > 0$ 时，我们有 $\tanh x > 0$，所以 $g'(x) > 0$。因此当 $x > 0$ 时，$x - \tanh x > 0$。

第 (ii) 部分

设 $h(x) = x\sinh x - 2\cosh x + 2$。则 $h(0) = 0 - 2 + 2 = 0$，且

$$h'(x) = \sinh x + x\cosh x - 2\sinh x = x\cosh x - \sinh x = \cosh x\,(x - \tanh x).$$

当 $x > 0$ 时，$\cosh x > 0$，而且由第 (i) 部分知 $x - \tanh x > 0$，所以 $h'(x) > 0$。因此当 $x > 0$ 时，$h(x) > 0$。

第 (iii) 部分

设 $k(x) = 2x\cosh 2x - 3\sinh 2x + 4x$。则 $k(0) = 0$，且

$$\begin{aligned} k'(x) &= 2\cosh 2x + 4x\sinh 2x - 6\cosh 2x + 4 \\ &= 4x\sinh 2x - 4\cosh 2x + 4. \end{aligned}$$

想一想 这个组合我们在哪里见过?

令 $u = 2x$，上式便是 $2u\sinh u - 4\cosh u + 4 = 2\,(u\sinh u - 2\cosh u + 2) = 2h(u)$。所以

$$k'(x) = 2h(2x) > 0 \quad \text{对于} x > 0,$$

这是由第 (ii) 部分得出的，因为 $2x > 0$。因此当 $x > 0$ 时，$k(x) > 0$。

f 没有转向点

当 $x > 0$ 时，$\mathrm{f}(x)$ 中每个因子都是正数，所以 $\mathrm{f}(x) > 0$，我们可以取对数:

$$\ln \mathrm{f}(x) = \ln x + \tfrac{1}{3}\ln\cosh x - \ln\sinh x.$$

求导，得

$$\frac{\mathrm{f}'(x)}{\mathrm{f}(x)} = \frac{1}{x} + \frac{\sinh x}{3\cosh x} - \frac{\cosh x}{\sinh x} = \frac{3\sinh x\cosh x + x\sinh^2 x - 3x\cosh^2 x}{3x\sinh x\cosh x}.$$

现在把分子用倍角表示，利用 $\sinh x\cosh x = \frac{1}{2}\sinh 2x$、$\sinh^2 x = \frac{1}{2}(\cosh 2x - 1)$ 和 $\cosh^2 x = \frac{1}{2}(\cosh 2x + 1)$:

$$\begin{aligned} 3\sinh x\cosh x + x\sinh^2 x - 3x\cosh^2 x &= \tfrac{3}{2}\sinh 2x + \tfrac{x}{2}(\cosh 2x - 1) - \tfrac{3x}{2}(\cosh 2x + 1) \\ &= \tfrac{3}{2}\sinh 2x - x\cosh 2x - 2x \\ &= -\tfrac{1}{2}\big(2x\cosh 2x - 3\sinh 2x + 4x\big). \end{aligned}$$

于是

$$\mathrm{f}'(x) = -\frac{\mathrm{f}(x)\,\big(2x\cosh 2x - 3\sinh 2x + 4x\big)}{6x\sinh x\cosh x}.$$

当 $x > 0$ 时，由第 (iii) 部分知括号内的式子为正，而 $\mathrm{f}(x)$、x、$\sinh x$ 和 $\cosh x$ 全部为正。所以对每个 $x > 0$ 都有 $\mathrm{f}'(x) < 0$：f' 永不为零，故 f 没有转向点（turning point），事实上 f 在 $x > 0$ 上严格递减。

草图

我们需要知道曲线在两端的表现。

当 $x \to 0^+$ 时：$x/\sinh x \to 1$ 且 $(\cosh x)^{1/3} \to 1$，所以 $\mathrm{f}(x) \to 1$。更仔细地说，$(\cosh x)^{1/3} = 1 + \frac{1}{6}x^2 + O(x^4)$ 而 $x/\sinh x = 1 - \frac{1}{6}x^2 + O(x^4)$，所以 x^2 项互相抵消；每个级数再多取一项，便得

$$\mathrm{f}(x) = 1 - \tfrac{1}{45}x^4 + O(x^6).$$

所以曲线离开（不包括在内的）点 $(0,1)$ 时非常平坦，斜率为零。

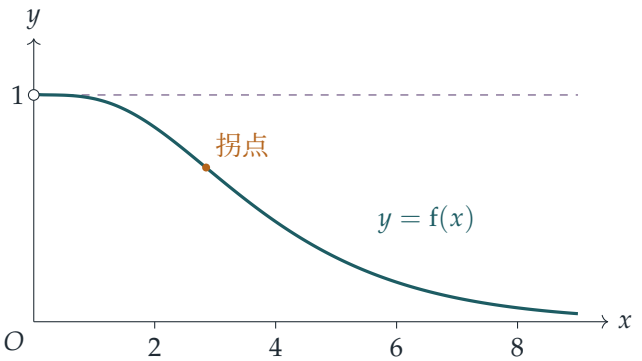
当 $x \to \infty$ 时：$\cosh x \sim \frac{1}{2}e^x$ 且 $\sinh x \sim \frac{1}{2}e^x$，所以

$$\mathrm{f}(x) \sim \frac{x\,(\frac{1}{2}e^x)^{1/3}}{\frac{1}{2}e^x} = 2^{2/3}\,x\,e^{-2x/3} \to 0,$$

因为指数函数的增长胜过 x 的幂。x-轴是一条渐近线（asymptote），而由于 $\mathrm{f} > 0$，曲线始终在它的上方。

> **想一想** 曲线开始时平坦，结束时也平坦，却一直递减。这必然导致甚么？

在 0 附近，曲线向下弯（它是 $1 - x^4/45$）；而当 x 很大时，它是一个衰减的指数函数，向上弯。所以两者之间必定至少有一个拐点（point of inflection）；数值计算显示恰有一个，在 $x \approx 2.85$ 附近，该处 $\mathrm{f} \approx 0.68$。一些函数值：$\mathrm{f}(1) \approx 0.983$，$\mathrm{f}(2) \approx 0.858$，$\mathrm{f}(4) \approx 0.441$，$\mathrm{f}(6) \approx 0.174$。



验算

- 在 $x = 1$ 处：$1 - \tanh 1 \approx 0.238$、$\sinh 1 - 2\cosh 1 + 2 \approx 0.089$ 和 $2\cosh 2 - 3\sinh 2 + 4 \approx 0.643$，全部为正。
- 恒等式 $\mathrm{f}'/\mathrm{f} = -k(x)/(6x\sinh x\cosh x)$ 可在任何一点检验：在 $x = 1$ 处两边都约为 -0.0592，而 f 的数值导数亦与此吻合。
- 数值上，当 x 很小时 $(1 - \mathrm{f}(x))/x^4 \approx 0.0222 \approx \frac{1}{45}$，而当 x 很大时 $\mathrm{f}(x)/(2^{2/3}xe^{-2x/3}) \to 1$。

常见错误

- 只证明导数为正，却忘记检查函数在 $x = 0$ 时为 0；两者缺一不可。
- 试图直接用商法则和积法则对 f 求导。这样做是可行的，但对数导数简洁得多，而且能让人看清它与 (iii) 的联系。
- 把曲线画成穿过 x-轴，或从原点出发。曲线从高度 1 开始（但取不到这个值），并且一直保持为正。

STEP 之外

由 $F(0) = 0$ 和 $F' > 0$ 证明 $F(x) > 0$ 的技巧，正是中值定理的实际应用；而把它连环使用（如本题般依次经过 k'、h'、g'），是分析学中证明函数之间不等式的标准方法。展开式 $\mathrm{f}(x) = 1 - x^4/45 + \cdots$ 让我们初尝运算幂级数（power series）的滋味，大学一年级的分析课程会为此奠定严格的基础。要阅读这两方面的内容，Abbott 的 *Understanding Analysis* 是个好选择。

详细解答

STEP II 1988，第 7 题

纯数学 · 代换积分法

STEP II 1988，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/88-s2-q7

关键思路

这里每一个等式都来自一次代换（substitution），难处在于找出合适的代换。第一步需要一个把 $[1,2]$ 送到 $[0,1]$、并把 $2-2x+x^2$ 变成含有 $1+t^2$ 的式子的代换：莫比乌斯映射（Möbius map）$x=2/(1+t)$ 恰好做到这一点。一旦出现 $1+t^2$，取 $t=\tan\theta$ 便是顺理成章的一步。以 θ 表示的被积函数（integrand）关于 $\theta=\frac{1}{8}\pi$ 对称，而在半个区间上把 $t=\tan\theta$ 还原，便会带出 $\tan\frac{1}{8}\pi=\sqrt{2}-1$。最后的推论是把同样的代换反过来使用。

解答

全篇之中，k 是任意实常数。所有要取幂的底数在所涉及的范围内都是正数，所以非整数次幂不会造成任何问题。

从 $[1,2]$ 到 $[0,1]$

设 $x=\dfrac{2}{1+t}$，于是当 $x=1$ 时 $t=1$，当 $x=2$ 时 $t=0$，并且 $\mathrm{d}x=-\dfrac{2}{(1+t)^2}\,\mathrm{d}t$。那么

$$2-2x+x^2=\frac{2(1+t)^2-4(1+t)+4}{(1+t)^2}=\frac{2(1+t^2)}{(1+t)^2},\qquad x^{k+1}=\frac{2^{k+1}}{(1+t)^{k+1}}.$$

所以被积函数变成

$$\frac{2^k(1+t^2)^k}{(1+t)^{2k}}\cdot\frac{(1+t)^{k+1}}{2^{k+1}}\cdot\frac{2}{(1+t)^2}=\frac{(1+t^2)^k}{(1+t)^{k+1}},$$

而 $\mathrm{d}x$ 中的负号把积分上下限由 $\int_1^0$ 对调为 $\int_0^1$：

$$I=\int_0^1\frac{(1+t^2)^k}{(1+t)^{k+1}}\,\mathrm{d}t.$$

三角形式

现在设 $t = \tan\theta$，其中 θ 由 0 变到 $\frac{1}{4}\pi$。那么 $1 + t^2 = \sec^2\theta$，$\mathrm{d}t = \sec^2\theta\,\mathrm{d}\theta$，并且

$$1 + t = \frac{\cos\theta + \sin\theta}{\cos\theta} = \frac{\sqrt{2}\cos\left(\frac{1}{4}\pi - \theta\right)}{\cos\theta},$$

这里用到 $\cos\theta + \sin\theta = \sqrt{2}\left(\cos\frac{1}{4}\pi\cos\theta + \sin\frac{1}{4}\pi\sin\theta\right)$。被积函数变成

$$\sec^{2k}\theta \cdot \frac{\cos^{k+1}\theta}{\left[\sqrt{2}\cos\left(\frac{1}{4}\pi - \theta\right)\right]^{k+1}} \cdot \sec^2\theta = \frac{\cos^{-(k+1)}\theta}{\left[\sqrt{2}\cos\left(\frac{1}{4}\pi - \theta\right)\right]^{k+1}} = \frac{1}{\left[\sqrt{2}\cos\theta\cos\left(\frac{1}{4}\pi - \theta\right)\right]^{k+1}},$$

因为 $\cos\theta$ 的幂次合计为 $-2k + (k+1) - 2 = -(k+1)$。两个余弦在 $[0, \frac{1}{4}\pi]$ 上都是正数，所以这对任何实数 k 都成立。这就得出 I 的第二个形式。

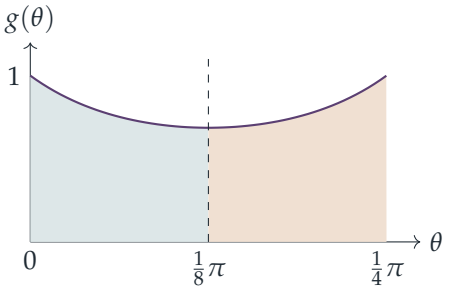
想一想 为甚么在 $[0, \frac{1}{4}\pi]$ 上的积分应该是在 $[0, \frac{1}{8}\pi]$ 上的积分的两倍?

把以 θ 表示的被积函数记作 $g(\theta)$。把 θ 换成 $\frac{1}{4}\pi - \theta$，只不过是把 $\cos\theta$ 和 $\cos(\frac{1}{4}\pi - \theta)$ 这两个因子对调，所以 $g(\frac{1}{4}\pi - \theta) = g(\theta)$：$g$ 的图像关于 $\theta = \frac{1}{8}\pi$ 对称。（你也可以直接看出这一点：$\sqrt{2}\cos\theta\cos(\frac{1}{4}\pi - \theta) = \frac{1}{2}(1 + \sqrt{2}\cos(2\theta - \frac{1}{4}\pi))$，它只取决于 θ 与 $\frac{1}{8}\pi$ 的距离。）严格来说，代入 $\phi = \frac{1}{4}\pi - \theta$，

$$\int_{\pi/8}^{\pi/4} g(\theta)\,\mathrm{d}\theta = \int_0^{\pi/8} g(\tfrac{1}{4}\pi - \phi)\,\mathrm{d}\phi = \int_0^{\pi/8} g(\phi)\,\mathrm{d}\phi,$$

所以 $[0, \frac{1}{4}\pi]$ 的两半贡献相等，而

$$I = 2\int_0^{\pi/8} \frac{\mathrm{d}\theta}{\left[\sqrt{2}\cos\theta\cos\left(\frac{1}{4}\pi - \theta\right)\right]^{k+1}}.$$



图中所示是 $k = 1$ 时的 g：两个着色的半边互为镜像。

由此得出积到 $\sqrt{2}-1$ 的积分

现在在 $[0,\frac{1}{8}\pi]$ 上把代换 $t=\tan\theta$ 反过来做。上面计算的每一步都保持不变；只有上限移到了 $t=\tan\frac{1}{8}\pi$。由 $\tan\frac{1}{4}\pi=\dfrac{2\tau}{1-\tau^2}=1$，其中 $\tau=\tan\frac{1}{8}\pi$，可得 $\tau^2+2\tau-1=0$，所以 $\tau=-1\pm\sqrt{2}$，而正根就是 $\tan\frac{1}{8}\pi=\sqrt{2}-1$。因此

$$I=2\int_0^{\sqrt{2}-1}\frac{(1+t^2)^k}{(1+t)^{k+1}}\,\mathrm{d}t,$$

这正是所求。把这个结果与同一被积函数的 $I=\int_0^1$ 比较，我们发现在 $[0,\sqrt{2}-1]$ 上的积分和在 $[\sqrt{2}-1,1]$ 上的积分*两者都*等于 $\frac{1}{2}I$。这就是我们接下来要用的事实。

推论

左方。写成 $\dfrac{2-2x^2+x^4}{x^2}=\dfrac{2-2u+u^2}{u}$，其中 $u=x^2$。那么 $\mathrm{d}u=2x\,\mathrm{d}x$，所以 $\dfrac{\mathrm{d}x}{x}=\dfrac{\mathrm{d}u}{2u}$，而 u 由 1 变到 2：

$$\int_1^{\sqrt{2}}\left(\frac{2-2x^2+x^4}{x^2}\right)^k\frac{1}{x}\,\mathrm{d}x=\frac{1}{2}\int_1^2\frac{(2-2u+u^2)^k}{u^k}\cdot\frac{1}{u}\,\mathrm{d}u=\frac{1}{2}I.$$

右方。被积函数是 $\dfrac{(2-2x+x^2)^k}{x^{k+1}}$，与 I 的定义中的相同，但积分区间是 $[1,\sqrt{2}]$。再次使用第一个代换 $x=2/(1+t)$。这时 $x=\sqrt{2}$ 对应 $1+t=\sqrt{2}$，即 $t=\sqrt{2}-1$，所以

$$\begin{aligned}\int_1^{\sqrt{2}}\left(\frac{2-2x+x^2}{x}\right)^k\frac{1}{x}\,\mathrm{d}x&=\int_{\sqrt{2}-1}^1\frac{(1+t^2)^k}{(1+t)^{k+1}}\,\mathrm{d}t\\&=I-\int_0^{\sqrt{2}-1}\frac{(1+t^2)^k}{(1+t)^{k+1}}\,\mathrm{d}t=I-\tfrac{1}{2}I=\tfrac{1}{2}I.\end{aligned}$$

两方都等于 $\frac{1}{2}I$，所以两者相等，这就是所要证明的结果。

验算

- $k=0$：$I=\int_1^2\frac{1}{x}\,\mathrm{d}x=\ln 2$，而最后那个恒等式的两方都是 $\int_1^{\sqrt{2}}\frac{1}{x}\,\mathrm{d}x=\frac{1}{2}\ln 2$，正如所料。
- $k=1$：$I=\int_1^2\left(\frac{2}{x^2}-\frac{2}{x}+1\right)\mathrm{d}x=2-2\ln 2\approx 0.61371$。对 I 的全部五个形式，以及最后那个恒等式的两方，使用辛普森法则（Simpson's rule）计算，结果吻合至小数点后十位（每一方都是 ≈ 0.30685）。对于非整数和负数的 k，例如 $k=0.5$、$k=-0.7$ 和 $k=4.3$，结果同样吻合。

常见错误

- 把 $\cos\theta + \sin\theta$ 写成单一个余弦时漏掉因子 $\sqrt{2}$，结果多出一个 $2^{(k+1)/2}$。
- 试图求出 I 的值。对一般的 k，它并没有闭合形式（closed form），也不需要有：每一步都只是变量代换或对称性。
- 在推论中，没有留意到右方只是 I 的一部分，需要用上“由此得出”的那个结果才能完成。

STEP 之外

映射 $x \mapsto 2/(1+t)$ 是一个莫比乌斯变换（Möbius transformation），又称分式线性变换（fractional linear transformation）。这类映射把圆和直线送到圆和直线，对于由二次式构成的被积函数，它们是最自然的变量代换。找出被积函数隐藏的对称性，就像这里的 $\theta \mapsto \frac{1}{4}\pi - \theta$，是计算没有初等原函数（antiderivative）的定积分（definite integral）时的常用技巧。想进一步了解莫比乌斯变换，可阅读 Needham 的 *Visual Complex Analysis*（以图像方式讲解复分析）。想看更多这种风格的积分，可参考 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP II 1988，第 8 题

纯数学 · 矩阵与特征向量

> **STEP II 1988，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/88-s2-q8

关键思路

模型中的一年是一个线性映射（linear map），因此相当于乘以一个 2×2 矩阵（matrix）M。比例永远不变的种群，就是 M 只会把它按比例放大或缩小的种群，也就是一个*特征向量*（*eigenvector*）。一旦求得两个特征向量，任何初始种群都可以分拆成沿这两个方向的两部分，而每一部分每年只是乘以它自己的特征值（eigenvalue）。于是，长远的命运就取决于哪个特征值较大，以及它前面的系数的正负。

解答

保持不变的比例

把种群数量写成列向量（column vector），模型中的一年就是

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = M\begin{pmatrix} A \\ B \end{pmatrix}, \qquad M = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}.$$

比例 $A:B$ 保持不变，当且仅当存在某个数 λ 使得 $(A',B')=\lambda(A,B)$，即 (A,B) 是 M 的特征向量。特征值满足

$$\det(M-\lambda I)=\left(\tfrac{1}{4}-\lambda\right)\left(\tfrac{3}{2}-\lambda\right)+\tfrac{3}{8}=\lambda^2-\tfrac{7}{4}\lambda+\tfrac{3}{4}=0,$$

所以 $4\lambda^2-7\lambda+3=(\lambda-1)(4\lambda-3)=0$，得出 $\lambda=1$ 或 $\lambda=\frac{3}{4}$。

- $\lambda=1$：由第一行得 $\frac{1}{4}A+\frac{3}{4}B=A$，所以 $A=B$。比例为 1，而这样的种群每年都完全不变。
- $\lambda=\frac{3}{4}$：由第一行得 $\frac{1}{4}A+\frac{3}{4}B=\frac{3}{4}A$，所以 $A=\frac{3}{2}B$。比例为 $\frac{3}{2}$；以第二行验算：$\frac{3}{2}B-\frac{1}{2}\cdot\frac{3}{2}B=\frac{3}{4}B$。两个物种每年都缩减至原来的 $\frac{3}{4}$。

不用矩阵也可以得到同样的结果：若 $A = rB$，则新的比例为 $\dfrac{\frac{1}{4}r + \frac{3}{4}}{\frac{3}{2} - \frac{1}{2}r}$，令它等于 r 便得 $2r^2 - 5r + 3 = 0$，所以 $r = 1$ 或 $r = \frac{3}{2}$。无论用哪种方法，比例只有在等于 1 或 $\frac{3}{2}$ 时才能保持不变，正如题目所要求。我们取 $x = 1$ 和 $y = \frac{3}{2}$。

其后各年

设

$$\begin{pmatrix} a \\ b \end{pmatrix} = p \begin{pmatrix} 1 \\ 1 \end{pmatrix} + q \begin{pmatrix} \frac{3}{2} \\ 1 \end{pmatrix}.$$

比较各分量，得 $p + \frac{3}{2}q = a$ 和 $p + q = b$。两式相减，$\frac{1}{2}q = a - b$，所以

$$q = 2(a - b), \qquad p = b - q = 3b - 2a.$$

由于 M 是线性的，一年之后第一部分乘以 1，第二部分乘以 $\frac{3}{4}$；n 年之后（只要模型仍然适用），我们有

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = (3b - 2a) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2(a - b)\left(\frac{3}{4}\right)^n \begin{pmatrix} \frac{3}{2} \\ 1 \end{pmatrix},$$

即

$$a_n = (3b - 2a) + 3(a - b)\left(\tfrac{3}{4}\right)^n,$$
$$b_n = (3b - 2a) + 2(a - b)\left(\tfrac{3}{4}\right)^n.$$

由于 $(\frac{3}{4})^n \to 0$，种群数量趋向 $(3b - 2a,\ 3b - 2a)$，这是平衡点（equilibrium）所组成的直线 $a = b$ 上的一点。

分类之前，先确认一件可以放心的事：土豚永远不会出问题。若 $a_n \geq 0$ 且 $b_n > 0$，则 $a_{n+1} = \frac{1}{4}a_n + \frac{3}{4}b_n > 0$，所以只要水牛仍然存活，土豚的数量也保持为正数。因此唯一需要留意的是 b_n 的正负，而我们只需考虑平面上 $a \geq 0$、$b > 0$ 的区域。

> ? **想一想** b_n 会不会有一段时间跌至零以下，然后回升并趋向一个正的极限（limit）？

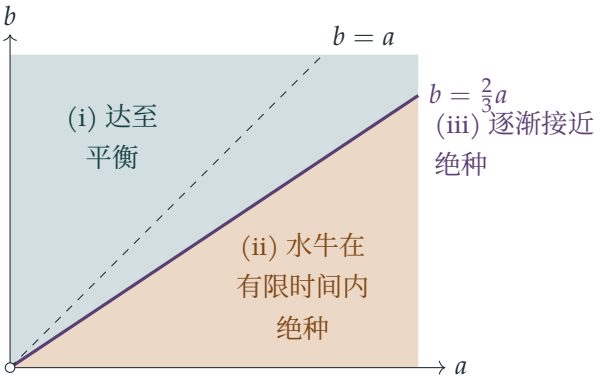
不会。在 b_n 的公式中，只有 $(\frac{3}{4})^n$ 与 n 有关，而它由 1 稳步递减至 0。所以 b_n 由 $b_0 = b$ 单调（monotonic）地移向它的极限 $3b - 2a$，而每个 b_n 都介乎 b 与 $3b - 2a$ 之间。这样分类就干净利落了。

三个区域

(i) 若 $3b - 2a > 0$，即 $b > \frac{2}{3}a$，则 b 和 $3b - 2a$ 都是正数，所以每个 b_n 都是正数，模型永远适用。种群趋向平衡点 $(3b - 2a,\ 3b - 2a)$，而这个平衡点并非零。这是直线 $b = \frac{2}{3}a$ 上方的区域（包括 b 轴，即 $a = 0$）。

(ii) 若 $3b-2a<0$，即 $b<\frac{2}{3}a$，则 $a>\frac{3}{2}b>b$，所以 $a-b>0$，而 b_n 严格递减并趋向负数 $3b-2a$。因此对某个有限的 n 有 $b_n\le 0$：水牛在有限时间内绝种。这是直线 $b=\frac{2}{3}a$ 与 a 轴之间的区域。

(iii) 若 $3b-2a=0$，即 $b=\frac{2}{3}a$ 且 $a>0$，则 $p=0$，种群纯粹是 $(\frac{3}{2},1)$ 的倍数：$b_n=b\,(\frac{3}{4})^n$，它对每个 n 都是正数，但趋向 0。水牛（以及土豚）逐渐接近绝种，却永远不会真正绝种。这只是半直线（half-line）$b=\frac{2}{3}a$，$a>0$。



虚线 $b=a$ 由已经处于平衡状态的种群组成；区域 (i) 内的每个起点都会向它移动，并且沿着与直线 $b=\frac{2}{3}a$ 平行的方向滑动（$(\frac{3}{2},1)$ 方向的分量不断缩小，而 $(1,1)$ 分量则保持不变）。

验算

- 取 $n=0$，公式给出 $a_0=3b-2a+3a-3b=a$ 和 $b_0=3b-2a+2a-2b=b$，正如所应得的结果。
- 由 $(a,b)=(1,3)$ 出发，它位于区域 (i)。按模型计算一年，得 $(\frac{1}{4}+\frac{9}{4},\ \frac{9}{2}-\frac{1}{2})=(\frac{5}{2},4)$，而公式给出 $7-6(\frac{3}{4})=\frac{5}{2}$ 和 $7-4(\frac{3}{4})=4$。极限为 $(7,7)$。
- 我们用电脑以分数作精确运算，由 3000 个随机起点迭代这个模型，确认了闭合形式（closed form）在 40 年内都成立，也确认了 $3b-2a$ 的正负在每一个情况下都能正确预测结果，包括 $b=\frac{2}{3}a$ 上的点，例如 $(3,2)$ 在五年后变成 $(\frac{729}{1024},\frac{243}{512})$，即起点的 $(\frac{3}{4})^5$ 倍。

常见错误

- 把比例上下颠倒：题目问的是土豚对水牛的比例，所以答案是 1 和 $\frac{3}{2}$，而不是 1 和 $\frac{2}{3}$。

- 忘记模型在 $b_n \leq 0$ 时便告终止，于是声称每个种群都趋向 $(3b - 2a, 3b - 2a)$，连“极限”为负数的情况也包括在内。
- 把区域 (iii) 说成一块面积：它只是半直线 $b = \frac{2}{3}a$，即 (i) 与 (ii) 之间的边界。

STEP 之外

这是一个离散线性动力系统（discrete linear dynamical system），而这里的方法（先对角化（diagonalise），再把每个特征值取 n 次方）正是大学一年级线性代数计算 M^n 的做法。特征值 1 给出一条由不动点（fixed point）组成的直线，而特征值 $\frac{3}{4} < 1$ 则给出一个不断衰减的方向；把时间改为连续，并以 $e^{\lambda t}$ 取代 λ^n，同样的图像就支配着线性微分方程组（systems of linear differential equations）。Strang 的 *Introduction to Linear Algebra* 有一章很好地讲解差分方程（difference equation）和矩阵的幂，而 Boyce 与 DiPrima 的 *Elementary Differential Equations* 则处理连续的版本。

详细解答

STEP II 1988，第 9 题

纯数学 · 子群与并集

> **STEP II 1988，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/88-s2-q9

本题在 *Beyond STEP* 第 6.6 节子群的并集中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch06/subgroup-unions.html

关键思路

用反证法（proof by contradiction）。假如两个子群（subgroup）都不是整个群，就取一个只属于 G_2 的元素，再取一个只属于 G_1 的元素，然后问：它们的乘积落在哪里？无论把乘积放进哪一个子群，乘法的封闭性（closure）和逆元（inverse）都会导出矛盾。换成三个子群，乘积就多了一个去处。最自然的尝试，是找一个每个元素都是自身逆元的群。

解答

两个子群

假设 G 的每个元素都属于 G_1 或 G_2，并为了导出矛盾，假设 $G_1 \neq G$ 且 $G_2 \neq G$。

由于 $G_1 \neq G$ 而 $G_1 \subseteq G$，存在元素 $a \in G$ 使得 $a \notin G_1$。每个元素都属于 G_1 或 G_2，所以 $a \in G_2$。

同理，由于 $G_2 \neq G$，存在元素 $b \in G$ 使得 $b \notin G_2$，于是 $b \in G_1$。

因此我们有

$$a \in G_2, \quad a \notin G_1, \qquad\qquad b \in G_1, \quad b \notin G_2.$$

现在考虑乘积 $ab \in G$。按假设，它属于 G_1 或 G_2。

- 若 $ab \in G_1$：由于 $b \in G_1$ 而 G_1 是子群，故 $b^{-1} \in G_1$，于是由封闭性得

$$a = (ab)\,b^{-1} \in G_1,$$

  这与 $a \notin G_1$ 矛盾。

- 若 $ab \in G_2$：由于 $a \in G_2$，故 $a^{-1} \in G_2$，于是

$$b = a^{-1}(ab) \in G_2,$$

  这与 $b \notin G_2$ 矛盾。

两种情况都不可能发生，所以我们的假设不成立：$G_1 = G$ 或 $G_2 = G$。 ■

想一想 我们在哪里用到了 G 是有限群这个条件?

完全没有用到。这个论证只用到子群的性质（封闭性和逆元），对任何群都成立，不论有限还是无限。换句话说：一个群永远不会是两个真子群（proper subgroup）的并集（union）。

三个子群

取 H 为克莱因四元群（Klein four-group）

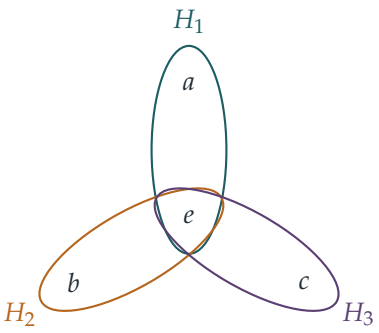
$$H = \{e,\ a,\ b,\ c\}, \qquad a^2 = b^2 = c^2 = e, \quad ab = ba = c,\ bc = cb = a,\ ca = ac = b.$$

具体来说，H 就是 $\{(0,0),(1,0),(0,1),(1,1)\}$ 在每个坐标各自模 2 相加之下所成的群，也可以看成非正方形的长方形的对称群（symmetry group）（恒等变换、两个反射和半周旋转）。

每个非单位元的元素的阶（order）都是 2，所以

$$H_1 = \{e,a\}, \qquad H_2 = \{e,b\}, \qquad H_3 = \{e,c\}$$

每一个都是子群：它包含 e，在乘法下封闭（唯一非平凡的乘积是 $a \cdot a = e$，b 和 c 的情况相同），而且每个元素都是自身的逆元。它们各有 2 个元素，所以没有一个等于 H。但它们合起来包含了 e、a、b 和 c，也就是 H 的全部元素。



所以三个真子群是可以覆盖一个群的。另一个例子是四元数群（quaternion group）$Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$，它是三个循环子群（cyclic subgroup）$\{\pm 1, \pm i\}$、$\{\pm 1, \pm j\}$ 和 $\{\pm 1, \pm k\}$ 的并集，每个子群的阶都是 4。

留意第一部分的论证为何在这里失效：当 $a \notin H_2$ 而 $b \notin H_1$ 时，乘积 $ab = c$ 可以溜进第三个子群 H_3。

验算

- 以穷举法搜索下列各群的所有子群：阶为 1 至 12 的循环群、S_3、S_4、阶为 8 的二面体群（dihedral group）、Q_8、$\mathbb{Z}_2^2$ 和 $\mathbb{Z}_2^3$（共 1278 个由真子群组成的有序对），结果找不到任何一对的并集等于整个群。
- 同一次搜索亦确认 $\{e, a\} \cup \{e, b\} \cup \{e, ab\}$ 就是整个克莱因四元群，而 Q_8 是 $\langle i\rangle$、$\langle j\rangle$、$\langle k\rangle$ 的并集。有趣的是，所测试的循环群，以及 S_3 和 S_4，没有一个是三个真子群的并集。
- 对第一部分作一个简单的合理性检查：在循环群中，生成元（generator）不属于任何真子群，所以循环群无论用多少个真子群都无法覆盖。

常见错误

- 用计数来论证：两个真子群各自最多有 $\frac{1}{2}|G|$ 个元素，但它们都含有 e，所以并集最多有 $|G| - 1$ 个元素。这个计数证明是正确的（根据拉格朗日定理（Lagrange's theorem）），而且简短，但它确实需要有限性和拉格朗日定理；两者都必须写明。
- 取了 $a \notin G_1$ 和 $b \notin G_2$，却忘记使用 $a \in G_2$ 和 $b \in G_1$。缺少这两个事实，就推不出矛盾。
- 给出的三个子群并非全是真子群，或者根本不是子群（例如 $\{e, a, b\}$，由于 $ab = c$，它并不封闭）。

STEP 之外

哪些群是三个真子群的并集？Scorza 的一个定理（1926 年）给出了答案：恰好是那些有商群（quotient）同构（isomorphic）于克莱因四元群的群，这也正是克莱因群成为基本例子的原因。同样的“不能是两个的并集”这一事实也出现在线性代数（linear algebra）中：无限域（infinite field）上的向量空间（vector space）永远不会是有限个真子空间（proper subspace）的并集。若想找一本群论的入门书，可以试试 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 1988，第 10 题

纯数学 · 抛物面的向量几何

> **STEP II 1988，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/88-s2-q10

关键思路

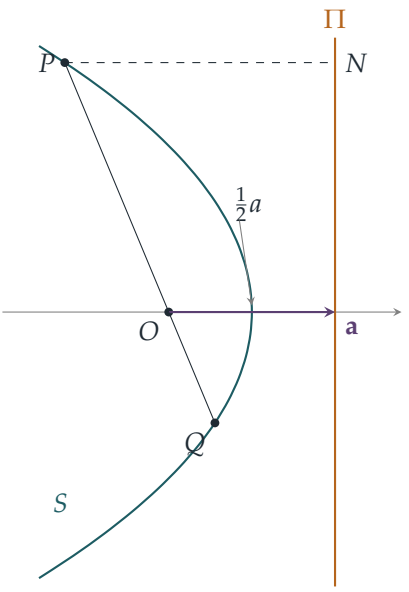
方程中的两个部分各有清晰的几何意义：r 是与 O 的距离，而 $(a^2 - \mathbf{a}\cdot\mathbf{r})/a$ 是与平面 $\mathbf{r}\cdot\mathbf{a} = a^2$ 的有向距离（signed distance）。因此 S 是一个旋转抛物面（paraboloid of revolution），以 O 为焦点（focus），以该平面为准面（directrix）。至于其余部分，可沿着通过 O 的直线来处理：把直线 OP 上的点写成 $\lambda\mathbf{p}$，把通过 O 且同时垂直于 $\mathbf{a}$ 和 $\mathbf{p}$ 的直线上的点写成 $\mu\,\mathbf{a}\times\mathbf{p}$，然后代入 S 的方程。λ 的正负号十分重要，因为 $|\lambda\mathbf{p}| = |\lambda|p$。

解答

轨迹

设 Π 为平面 $\mathbf{r}\cdot\mathbf{a} = a^2$。它的单位法向量（unit normal）是 $\hat{\mathbf{a}} = \mathbf{a}/a$，与 O 的距离为 a，并位于 $\mathbf{a}$ 所指的一侧。由点 $\mathbf{r}$ 到 Π 的垂直距离是

$$d(\mathbf{r}) = |\,a - \hat{\mathbf{a}}\cdot\mathbf{r}\,| = \frac{|a^2 - \mathbf{a}\cdot\mathbf{r}|}{a}.$$



上图显示 S 被一个包含 $\mathbf{a}$ 的平面所截出的截面。点 P 与 O 的距离，等于它与其在 Π 上的垂足 N 的距离。曲面与轴相交于 $\frac{1}{2}\mathbf{a}$，即 O 与 Π 的正中间。

S 上的点是等距的。若 $\mathbf{a}\cdot\mathbf{r}+ar=a^2$，则 $a^2-\mathbf{a}\cdot\mathbf{r}=ar\geq 0$。因此

$$d(\mathbf{r})=\frac{a^2-\mathbf{a}\cdot\mathbf{r}}{a}=\frac{ar}{a}=r=|OR|.$$

等距的点都在 S 上。假设 $r=d(\mathbf{r})$，即 $ar=|a^2-\mathbf{a}\cdot\mathbf{r}|$。若 $a^2-\mathbf{a}\cdot\mathbf{r}\geq 0$，这正是 S 的方程。否则 $\mathbf{a}\cdot\mathbf{r}>a^2$，而该条件表示 $\mathbf{a}\cdot\mathbf{r}=ar+a^2>ar$。这与 $\mathbf{a}\cdot\mathbf{r}\leq ar$（柯西–施瓦茨不等式，Cauchy–Schwarz）矛盾。所以在 Π 另一侧的点都不会等距，而 S 恰好就是与 O 和 Π 等距的所有点所组成的集合。这是一个旋转抛物面，其旋转轴是通过 O 且方向为 $\mathbf{a}$ 的直线。

以下两个推论稍后会用到。对于 S 上的点 P，

$$\mathbf{a}\cdot\mathbf{p}=a^2-ap. \qquad (*)$$

此外 $a^2=\mathbf{a}\cdot\mathbf{p}+ap\leq 2ap$，所以 $p\geq\frac{1}{2}a$，从而 $2p-a\geq 0$。等号只在 $\mathbf{p}$ 是 $\mathbf{a}$ 的正数倍时成立，即 P 是顶点（vertex）$\frac{1}{2}\mathbf{a}$ 的时候。

第二个交点 Q

直线 OP 上的点是 $\lambda\mathbf{p}$，其中 $|\lambda\mathbf{p}|=|\lambda|p$。代入 S 的方程得

$$\lambda\,\mathbf{a}\cdot\mathbf{p}+a|\lambda|p=a^2.$$

想一想 为甚么 Q 不可能与 P 位于 O 的同一侧?

若 $\lambda > 0$，左方是 $\lambda(\mathbf{a}\cdot\mathbf{p} + ap) = \lambda a^2$，所以 $\lambda = 1$，这就是 P 本身。（从几何上看，由焦点出发的射线与抛物面最多相交一次。）因此 Q 对应 $\lambda < 0$，这时 $|\lambda| = -\lambda$。利用 $(*)$，

$$\lambda(\mathbf{a}\cdot\mathbf{p} - ap) = a^2 \quad\Longrightarrow\quad \lambda(a^2 - 2ap) = a^2 \quad\Longrightarrow\quad \lambda = -\frac{a}{2p-a}.$$

只要 $2p > a$，即 P 不是顶点，这个值便如所要求般是负数。（在顶点处，直线 OP 就是轴，它与 S 只相交一次。）因此

$$\boxed{\mathbf{q} = -\frac{a}{2p-a}\,\mathbf{p}} \qquad \text{及} \qquad |OQ| = \frac{ap}{2p-a}.$$

点 T 和 T'

通过 O 且同时垂直于 $\mathbf{a}$ 和 $\mathbf{p}$ 的直线是 $\mathbf{t} = \mu\,\mathbf{a}\times\mathbf{p}$。（我们需要 $\mathbf{a}\times\mathbf{p} \neq \mathbf{0}$，所以 P 不在轴上。）由于 $\mathbf{a}\cdot\mathbf{t} = 0$，$S$ 的方程简化为 $at = a^2$，即 $|\mathbf{t}| = a$。所以 T 和 T' 位于 S 与通过 O 且平行于 Π 的平面相交所得的圆上，并且

$$|\mu|\,|\mathbf{a}\times\mathbf{p}| = a.$$

想一想 在不知道 $\mathbf{a}$ 与 $\mathbf{p}$ 之间夹角的情况下，怎样求出 $|\mathbf{a}\times\mathbf{p}|$?

利用恒等式（identity）$|\mathbf{a}\times\mathbf{p}|^2 = a^2p^2 - (\mathbf{a}\cdot\mathbf{p})^2$，它其实就是 $\sin^2\theta = 1 - \cos^2\theta$ 的另一种写法。然后代入 $(*)$:

$$\begin{aligned}|\mathbf{a}\times\mathbf{p}|^2 &= a^2p^2 - (a^2 - ap)^2\\ &= a^2p^2 - a^4 + 2a^3p - a^2p^2\\ &= a^3(2p-a).\end{aligned}$$

因此

$$|\mu| = \frac{a}{a^{3/2}\sqrt{2p-a}} = \frac{1}{\sqrt{a(2p-a)}} = \frac{1}{\sqrt{2ap-a^2}},$$

而 T 和 T' 的位置向量（position vector）是 $\pm\dfrac{1}{\sqrt{2ap-a^2}}\,\mathbf{a}\times\mathbf{p}$，正如所求。

三角形 PQT 的面积

P、O、Q 三点共线（collinear），而 O 在 P 与 Q 之间。向量 $\mathbf{t}$ 平行于 $\mathbf{a}\times\mathbf{p}$，所以它垂直于 $\mathbf{p}$，从而垂直于直线 PQ。因此 OT 是三角形 PQT 中由 T 引出的高，其长度为 a。底边是

$$|PQ| = p + \frac{ap}{2p-a} = \frac{p(2p-a)+ap}{2p-a} = \frac{2p^2}{2p-a}.$$

因此

$$\text{面积} = \tfrac{1}{2}\cdot|PQ|\cdot|OT| = \tfrac{1}{2}\cdot\frac{2p^2}{2p-a}\cdot a = \frac{ap^2}{2p-a},$$

正如所求。把 T 换成 T'，结果同样成立。

验算

- 取 $\mathbf{a} = (2,0,0)$ 和 $P = (-1.25,3,0)$，即图中的那一点。这时 $\mathbf{a}\cdot\mathbf{p} + ap = -2.5 + 2(3.25) = 4 = a^2$，所以 P 在 S 上。公式给出 $\mathbf{q} = -\frac{2}{4.5}\mathbf{p} = (0.556,-1.333,0)$，而 $\mathbf{a}\cdot\mathbf{q} + aq = 1.111 + 2.889 = 4$，所以 Q 也在 S 上。
- 取 P 为正焦弦（latus rectum）的一个端点，于是 $\mathbf{p}\perp\mathbf{a}$ 且 $p = a$。这时 $\mathbf{q} = -\mathbf{p}$，由对称性可知这是正确的。面积是 $\frac{1}{2}(2a)(a) = a^2$，而公式给出 $a\cdot a^2/a = a^2$。
- 两个焦距满足 $\dfrac{1}{OP} + \dfrac{1}{OQ} = \dfrac{1}{p} + \dfrac{2p-a}{ap} = \dfrac{2}{a}$，这正是抛物线经典的焦弦（focal chord）性质。
- 在数值上，我们随机选取了 20 000 组 $\mathbf{a}$ 和 P，检验了 Q、T、T' 和面积。最大误差约为 10^{-11}。

常见错误

- 对所有 λ 都写成 $|\lambda\mathbf{p}| = \lambda p$。这样唯一的解就是 $\lambda = 1$，便无法找到 Q。
- 在第一部分忘记证明逆命题(converse)。位于 Π 另一侧的点满足 $|a^2 - \mathbf{a}\cdot\mathbf{r}| = \mathbf{a}\cdot\mathbf{r} - a^2$，我们必须排除这些点。
- 用 $\overrightarrow{PQ}$ 与 $\overrightarrow{PT}$ 的向量积（cross product）来计算面积。这样做是可行的，但比起留意到 $OT\perp PQ$ 要冗长得多。

STEP 之外

这就是圆锥曲线（conic）的焦点–准线定义，只是放在三维空间中，并且不依赖坐标。设 e 为离心率（eccentricity），则曲面 $e\,\mathbf{a}\cdot\mathbf{r} + ar = e\,a^2$（与 O 的距离等于与 Π 的距离的 e 倍）给出其他以 O 为一个焦点的旋转二次曲面（quadric of revolution）。关系式 $1/OP + 1/OQ = 2/a$ 是极方程（polar equation）$\ell/r = 1 + e\cos\theta$ 的向量形式，而这个方程在轨道力学（orbital mechanics）中占有核心地位。若想进一步学习向量代数，Strang 的 *Introduction to Linear Algebra*（线性代数入门）是很好的下一步；Siklos 的 *Advanced Problems in Mathematics*（高等数学难题）则收录了更多这类风格的题目。

详细解答

STEP II 1988，第 11 题

力学 · 绳缠绕柱子

> **STEP II 1988，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/88-s2-q11

关键思路

绳子未缠上圆柱的部分，总是沿*切线*（tangent）离开圆柱。因此质点的位置就是“接触点加上一个长度等于未缠绕绳长的切向量”，而这很容易用角 θ 表示出来。求导之后，有两项互相抵消，速度原来与绳子垂直。于是张力（tension）不做功，速率保持在初始值，而速率公式就成为一条关于 θ 的可分离变量微分方程（separable differential equation）。

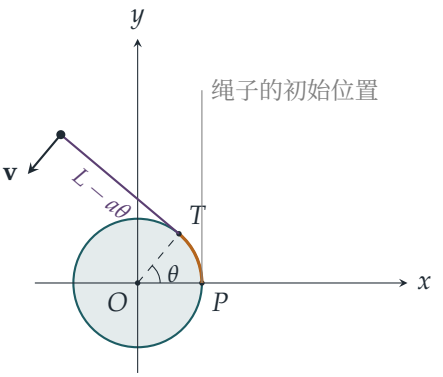
解答

建立模型

我们把质点看成光滑桌面上的一个质点（point mass），系着一条始终拉紧的轻而不可伸长的绳子（light inextensible string）。重力与法向反作用力（normal reaction）都沿铅垂方向，互相抵消，所以唯一的水平力就是张力。

取以 O 为原点的坐标轴，并把 P 放在 $(a, 0)$。绳子起初垂直于 OP，所以它在 P 处与圆柱相切，我们设质点从 (a, L) 出发。（若取镜像的情况，只是把 y 坐标的正负号反转。）当绳子沿逆时针方向缠上圆柱，缠过角 θ 之后，长度 $a\theta$ 的一段贴在由 P 起的圆弧上，而绳子离开圆柱的位置是

$$T = a(\cos\theta,\ \sin\theta).$$



绳子未缠上的部分长 $L-a\theta$，沿 T 处的切线伸展。指向绳子展开方向的单位切向量（unit tangent）是 $(-\sin\theta,\ \cos\theta)$：它垂直于 OT，而且当 $\theta=0$ 时等于 $(0,1)$，与起始位置吻合。

位置与速度

因此质点的位置向量（position vector）是

$$\mathbf{r}=a\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}+(L-a\theta)\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}=\begin{pmatrix}a\cos\theta-(L-a\theta)\sin\theta\\ a\sin\theta+(L-a\theta)\cos\theta\end{pmatrix}.$$

利用链式法则（chain rule）和乘积法则（product rule）对 t 求导，

$$\begin{aligned}\dot{\mathbf{r}}&=a\dot\theta\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}-a\dot\theta\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}+(L-a\theta)\dot\theta\begin{pmatrix}-\cos\theta\\ -\sin\theta\end{pmatrix}\\&=-(L-a\theta)\dot\theta\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}.\end{aligned}$$

首两项互相抵消：接触点沿圆柱移动所带来的效果，恰好被未缠上的绳子缩短所抵销。

由于 $(\cos\theta,\sin\theta)$ 是单位向量，$L-a\theta>0$，而且绳子缠上圆柱期间 $\dot\theta>0$，所以速率是

$$|\dot{\mathbf{r}}|=(L-a\theta)\dot\theta,$$

正是所求。

碰到圆柱所需的时间

想一想 绳子一直拉着质点，为甚么质点的速率仍会保持等于 v?

速度平行于 $(\cos\theta,\sin\theta)$，而绳子（因而张力）则沿 $(-\sin\theta,\cos\theta)$ 方向。两者互相垂直，所以张力不做功。桌面光滑，张力是唯一的水平力。因此动能（kinetic energy）保持不变，速率也随之恒等于 v。张力只会改变速度的方向。

所以，在质点到达圆柱之前的所有时刻 t，

$$(L - a\theta)\,\frac{\mathrm{d}\theta}{\mathrm{d}t} = v, \qquad \theta(0) = 0.$$

分离变量，并由 0 积分至 t，

$$L\theta - \tfrac{1}{2}a\theta^2 = vt.$$

当整条绳子都缠上圆柱时，质点便碰到圆柱，即 $a\theta = L$，所以 $\theta = L/a$。于是

$$vt = \frac{L^2}{a} - \frac{L^2}{2a} = \frac{L^2}{2a}, \qquad \text{所以} \qquad t = \frac{L^2}{2av},$$

这就是所求的时间。

验算

- 当 $\theta = 0$ 时，公式给出 $\mathbf{r} = (a, L)$ 及 $\dot{\mathbf{r}} = (-L\dot{\theta},\, 0)$，与初始的绳子垂直，理应如此。
- 单位：$L^2/(av)$ 的单位是 $\mathrm{m}^2/(\mathrm{m}\cdot\mathrm{m\,s^{-1}}) = \mathrm{s}$。
- 另一个理解答案的方法：质点整条路径的长度是

  $$\int_0^{L/a} (L - a\theta)\,\mathrm{d}\theta = \frac{L^2}{2a},$$

  并以恒定速率 v 走完。
- 我们用数值方法模拟了这个运动：一条非常硬的绳子缠绕在圆柱上，取 $a = 1$、$L = 3$、$v = 2$。模拟所得的时间与 $L\theta - \frac{1}{2}a\theta^2 = vt$ 相符，误差在 10^{-5} 以内；而把 $\dot{\theta} = v/(L - a\theta)$ 积分至 $a\theta = L$，得到 2.25，即 $L^2/(2av)$。

常见错误

- 使用关于 O 的角动量守恒（conservation of angular momentum）。张力并不通过 O（它沿切线作用），所以关于 O 的角动量并不守恒。能量才是合适的工具。
- 把未缠上的绳子写成沿 $(\cos\theta, \sin\theta)$ 方向，即沿径向，而不是沿切线方向。
- 忘记速率公式需要 $\dot{\theta} > 0$。该向量的绝对值是 $(L - a\theta)|\dot{\theta}|$。

STEP 之外

质点所描出的路径是圆的*渐伸线*（involute），即从线轴上拉紧展开的一条线，其末端所描出的曲线。齿轮的齿之所以造成那种形状，正是因为渐伸线；它也是惠更斯（Huygens）

研究摆钟时的核心。它的弧长 $\int (L - a\theta)\,\mathrm{d}\theta$ 是平面曲线微分几何（differential geometry，包括渐屈线（evolute）、渐伸线和曲率（curvature））的一个入门例子。至于力学方面，包括能量与角动量在甚么情况下守恒、甚么情况下不守恒，可阅读 Kleppner and Kolenkow, *An Introduction to Mechanics*（一本大学力学入门教科书）。

详细解答

STEP II 1988，第 12 题

力学 · 下落的链条

STEP II 1988，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：

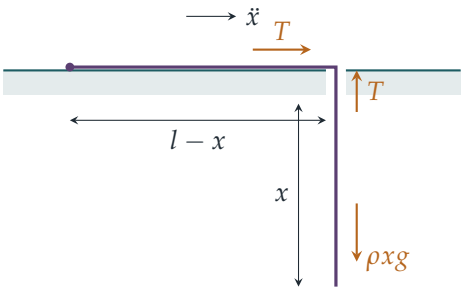
step.maths.org/questions/88-s2-q12

关键思路

由于绳子不可伸长（inextensible），绳上每一小段的速率都相同，所以整条绳子就像一个物体，而推动它的力就是垂在桌面以下那一段的重量。随着愈来愈多绳子穿过小孔，这个重量不断增大，于是得到一条以下垂长度为未知量的简单线性微分方程（linear differential equation）。它的解是双曲余弦（hyperbolic cosine），而所求的时间则由反解这个函数得出。

解答

整个系统完全没有摩擦：桌面光滑（smooth），小孔既小且光滑（所以它只改变绳子的方向，而不会沿绳子方向施加任何力），而绳子每单位长度的质量为 ρ。设在时间 t 时垂在桌面以下的长度为 x，则留在桌面上的长度为 $l-x$。起初 $x=l-d$ 且 $\dot{x}=0$，而当 $x=l$ 时绳端到达小孔顶部。由于桌子的高度大于 l，在此之前下垂部分不会碰到地面。



设 T 为绳子在小孔处的张力（tension）。由于小孔光滑，紧贴小孔上方和下方的张力同为 T。

桌面上的部分。它的质量为 $\rho(l-x)$，以加速度 $\ddot{x}$ 沿水平方向移向小孔（因为绳子不可伸长，它与下垂部分移动得一样快）。由于桌面光滑，作用在它上面的水平力只有张力 T。所以

$$\rho(l-x)\,\ddot{x} = T.$$

下垂的部分。它的质量为 ρx，以加速度 $\ddot{x}$ 向下运动，所受的力是向下的重量 $\rho x g$ 和向上的张力 T：

$$\rho x\,\ddot{x} = \rho x g - T.$$

把两条方程相加，便可消去 T：

$$\rho l\,\ddot{x} = \rho g x, \qquad \text{即} \qquad \ddot{x} = n^2 x, \quad n = \sqrt{\frac{g}{l}}.$$

想一想 小孔两边的绳子长度一直在变。对每一段分别使用 $F = ma$，真的没有问题吗？

没有问题。每条方程都是在某一瞬间，应用于该瞬间位于小孔那一边的那一段绳子；整条绳子以同一速率 $\dot{x}$ 一起运动，所以整条绳子的动量（momentum）变化率是 $\rho l\ddot{x}$，而沿运动方向有分量的外力，就只有下垂部分的重量 $\rho x g$。（桌面的反作用力抵销了水平部分的重量，而小孔只是令绳子转向。）因此合并后的方程 $\rho l\ddot{x} = \rho g x$ 正是整条绳子的牛顿第二定律（Newton's second law）。下文的能量验算亦印证了这一点。

$\ddot{x} = n^2 x$ 的通解（general solution）是 $x = A\cosh nt + B\sinh nt$。由 $t = 0$ 时 $\dot{x} = 0$ 得 $B = 0$，再由 $t = 0$ 时 $x = l - d$ 得 $A = l - d$：

$$x = (l-d)\cosh nt.$$

绳端在 $x = l$ 时到达小孔顶部，即在满足下式的时间 t_1：

$$\cosh nt_1 = \frac{l}{l-d}.$$

由于 $l/(l-d) > 1$，此方程恰有一个正数解。利用 $\operatorname{arcosh} u = \ln\left(u + \sqrt{u^2-1}\right)$，有

$$\sqrt{u^2-1} = \frac{\sqrt{l^2-(l-d)^2}}{l-d} = \frac{\sqrt{2ld-d^2}}{l-d},$$

所以所需时间为

$$\boxed{t_1 = \sqrt{\frac{l}{g}}\,\operatorname{arcosh}\!\left(\frac{l}{l-d}\right) = \sqrt{\frac{l}{g}}\,\ln\left(\frac{l+\sqrt{2ld-d^2}}{l-d}\right).}$$

验算

- **能量**。取桌面为势能（potential energy）的零点。下垂部分的质心（centre of mass）位于深度 $x/2$ 处，所以它的势能是 $-\frac{1}{2}\rho g x^2$。由静止开始，能量守恒（conservation of energy）给出 $\frac{1}{2}\rho l\dot{x}^2 = \frac{1}{2}\rho g\left(x^2 - (l-d)^2\right)$，对 t 求导便再次得到 $l\ddot{x} = gx$。
- d **很小**。如果 d 极小，几乎整条绳子都垂着，它应以接近 g 的加速度下落，所需时间约为 $\sqrt{2d/g}$。事实上，当 d 很小时，由 $\cosh nt_1 = 1 + \frac{d}{l} + \cdots$ 和 $\cosh nt_1 \approx 1 + \frac{1}{2}n^2t_1^2$ 可得 $t_1^2 \approx 2d/(n^2 l) = 2d/g$。
- $d \to l$。如果几乎没有绳子垂下，绳子几乎不会动，于是 $t_1 \to \infty$，正如对数式所显示的那样。
- **数值**。取 $l = 2\,\mathrm{m}$、$d = 1.5\,\mathrm{m}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$，公式给出 $t_1 \approx 0.932\,\mathrm{s}$，与对这个运动逐步进行数值积分所得的结果吻合。

常见错误

- 只对下垂部分写出 $\rho x\ddot{x} = \rho x g$，忘记了它还要拖动桌面上的绳子（换句话说，是用错了质量：被加速的是整条绳子，质量为 ρl）。
- 尝试用 $x = A\cos nt$ 形式的解：方程是 $\ddot{x} = +n^2 x$，所以运动是按指数增长，而不是振动。
- 把结束条件定为 $x = d$ 而不是 $x = l$：绳端到达小孔时，整段长度 l 都垂在下面。

STEP 之外

这是一个质量分布随时间变化的问题：总质量固定不变，但受重力驱动的那一部分随位置而改变，由此得到一条线性方程，其解为指数函数。假如绳子是从无摩擦的桌边滑落，同时有一堆原本静止的绳子不断被扯动起来，动量便会在非弹性（inelastic）的“猛扯”中损失，能量亦不再守恒；分辨这两类情况，是大学力学课程中的经典课题。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对质量流动的系统以及链条和绳索的问题有细致的讨论。

详细解答

STEP II 1988，第 13 题

力学 · 摩擦力与平衡

STEP II 1988，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：

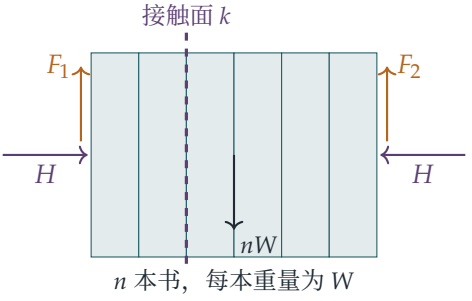
step.maths.org/questions/88-s2-q13

关键思路

托住书本的每一个垂直力都是摩擦力（friction），而任何接触面上的摩擦力都不能超过该处法向反作用力（normal reaction）的 μ 倍。因此我们需要两样东西：每个接触面上的法向反作用力（由水平方向的平衡（equilibrium）很容易求得），以及每个接触面必须提供的摩擦力（考虑适当选取的一组书本在垂直方向的平衡）。然后我们找出哪一个接触面最接近滑动。把相邻的一组书本视为一个整体，可以令代数运算保持简短。

解答

模型。 设有 n 本相同的书，每本重量为 W，书的封面保持垂直。每只手以水平力 $H \leq Q$ 压向书本，并提供向上的摩擦力；把这两个摩擦力记作 F_1（左手）和 F_2（右手）。由于每只手的垂直力正正就是这个摩擦力，所以我们需要 $F_1, F_2 \leq P$。题目指明可以忽略转动效应，因此我们只需考虑每本书或每组书本的力平衡。



法向反作用力。 把书本由左至右编号为 $1,2,\ldots,n$，并设接触面 k 为第 k 本书与第 $k+1$ 本书之间的接触面（$1 \le k \le n-1$）。作用于第 $1,\ldots,k$ 本书这一组书本上的水平力，只有左手的推力 H 和来自第 $k+1$ 本书的法向反作用力。所以该反作用力同样是 H：每一个接触面，不论是手与书之间还是书与书之间，法向反作用力都是 H，因此每个接触面所能提供的摩擦力大小最多为 μH。

双手。 在垂直方向上，对整排书本而言，

$$F_1 + F_2 = nW.$$

F_1, F_2 各自最多为 P（臂力），亦最多为 $\mu H \le \mu Q$（摩擦力）。记 $L = \min(P, \mu Q)$。于是 $nW = F_1 + F_2 \le 2L$，所以

$$n \le \frac{2L}{W} = \frac{2\min(P, \mu Q)}{W}.$$

这是必要条件。若此条件成立，最佳的做法是尽力压紧，即 $H = Q$（H 愈大，只会提高每一个摩擦力的上限），并把负荷平均分配，即 $F_1 = F_2 = \frac{1}{2}nW \le L$。这样双手都在各自的限度之内。

> ? **想一想** 即使双手托得住，书本会不会在中间某处互相滑过?

书本之间的接触面。 我们直接验证这一点。对第 $1,\ldots,k$ 本书这一组书本应用垂直方向的平衡。它的重量 kW 由手提供的 $F_1 = \frac{1}{2}nW$ 以及来自第 $k+1$ 本书的摩擦力 f_k（以向上为正）所承托：

$$f_k = kW - \tfrac{1}{2}nW = \left(k - \tfrac{n}{2}\right) W.$$

对于 $1 \le k \le n-1$，我们有 $\left|k - \frac{n}{2}\right| \le \frac{n}{2} - 1$，等号在 $k = 1$ 和 $k = n-1$ 时成立，即紧贴最外侧两本书的接触面。所以任何接触面所需的最大摩擦力为 $\left(\frac{n}{2} - 1\right)W$，比每只手所提供的 $\frac{1}{2}nW$ 少 W。由于每个接触面的上限与手的一样，都是 μQ，所以只要双手不滑动，就没有任何两本书会互相滑动。（从物理上看：最外侧的每本书由手托住，而它向内传递的摩擦力，就是手的摩擦力减去它本身的重量。）

答案。 这排书本能被提起，当且仅当 $nW \le 2\min(P, \mu Q)$。所以书本数目的最大值为

$$\boxed{n_{\max} = \left\lfloor \frac{2\min(P,\, \mu Q)}{W} \right\rfloor}$$

即同时不超过 $2P/W$ 和 $2\mu Q/W$ 的最大整数。哪一个上限起作用，便说明了是甚么限制着图书馆管理员：当 $P < \mu Q$ 时是提举的臂力，当 $\mu Q < P$ 时则是握力。

验算

- 取 $W = 10\,\mathrm{N}$、$P = 60\,\mathrm{N}$、$Q = 100\,\mathrm{N}$ 及 $\mu = 0.4$。则 $\min(60, 40) = 40$，而 $n_{\max} = 8$。有 8 本书时，每只手提供 $40\,\mathrm{N}$ 的摩擦力，刚好达到其上限，而最吃力的接触面需要 $30\,\mathrm{N} < 40\,\mathrm{N}$。若加上第九本书，每只手便需要提供 $45\,\mathrm{N}$。
- 只有一本书（$n = 1$）时没有书本之间的接触面，我们只需要 $W \le 2\min(P, \mu Q)$，这显然是正确的。
- 我们用电脑对 W, P, Q, μ 的 3000 组随机取值求解了整组不等式（任意 $H \le Q$、任意的 F_1, F_2 分配、每一个接触面），所得的最大可行 n 每次都与公式吻合。

常见错误

- 忘记了手臂的垂直力就是摩擦力，所以 P 和 μQ 这两个限制是作用于同一个力之上。
- 未经验证便假设中间的书本是最薄弱之处。书本之间的力比双手处的力小，而不是大。
- 只给出 $2\min(P, \mu Q)/W$ 而没有取下整（floor）：答案必须是整数本书。

STEP 之外

这里的摩擦力是静不定的（statically indeterminate）：平衡条件只确定了 $F_1 + F_2$，而我们问的是有没有某一种分配方式能令每个接触面都不超出其限度。这类问题其实是小型的线性规划（linear programme），而“检查是否存在某个可容许的平衡状态”这个想法，同样是结构极限分析（limit analysis）的基础。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）是一个好去处，可以看到摩擦力与平衡如何得到严谨的处理。

详细解答

STEP II 1988，第 14 题

力学 • 转动的刚体

STEP II 1988，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/88-s2-q14

关键思路

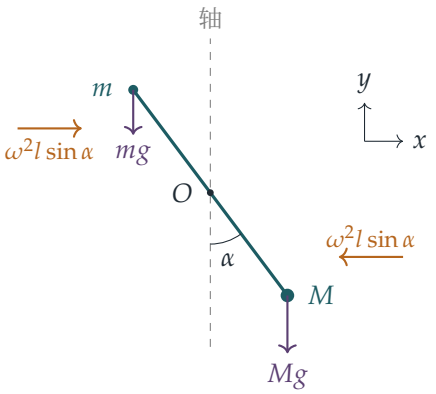
在稳定状态（steady state）下，相对于包含杆的那个旋转铅垂平面，一切都静止不动，所以每个质点（particle）只是以恒定速率沿水平圆周运动。这样每个质点的加速度就完全确定了。接着由牛顿第二定律（Newton's second law）可知杆对每个质点所施的力，再加上关于轻杆（light rod）的两个事实，问题便迎刃而解：作用于轻杆的各力必须平衡，而且（因为杆可绕轴自由摆动）这些力对轴的力矩（moment）也必须平衡。

解答

模型。 把两个质量视为装在轻质刚性杆上的质点，杆在 O 处光滑地枢接，因此轴不会沿其自身方向施加力偶（couple）。我们在包含杆的铅垂平面内讨论，这个平面以角速率（angular speed）ω 绕通过 O 的铅垂线转动。取 x 为此平面内的水平方向，y 为铅垂向上，原点为 O。

设 M 位于 $(l\sin\alpha,\ -l\cos\alpha)$，$m$ 位于 $(-l\sin\alpha,\ l\cos\alpha)$；稍后便会看到，较重的质点必定是在下方的那个。每个质点都以恒定速率，沿半径为 $l\sin\alpha$ 的水平圆周绕铅垂轴运动，所以其加速度大小为 $\omega^2 l\sin\alpha$，方向水平指向轴：

$$\mathbf{a}_M = (-\omega^2 l\sin\alpha,\ 0), \qquad \mathbf{a}_m = (\omega^2 l\sin\alpha,\ 0).$$



（琥珀色箭头表示加速度，紫红色箭头表示重量。）

杆所施的力。 设 $\mathbf{R}_M$ 和 $\mathbf{R}_m$ 分别为杆对 M 和 m 所施的力。这些力不一定沿着杆的方向：刚性杆也可以向侧面施力。由牛顿第二定律得 $\mathbf{R}_M - Mg\,\mathbf{j} = M\mathbf{a}_M$，所以

$$\mathbf{R}_M = \left(-M\omega^2 l\sin\alpha,\ Mg\right), \qquad \mathbf{R}_m = \left(m\omega^2 l\sin\alpha,\ mg\right).$$

根据牛顿第三定律（Newton's third law），两个质点对杆所施的力，在 M 的一端为 $-\mathbf{R}_M$，在 m 的一端为 $-\mathbf{R}_m$。

> ? **想一想** 轴在 O 处对杆施加一个未知的力。在不知道这个力的情况下，我们怎样取力矩?

对轴取力矩。 轴的力作用于 O，所以它对 O 没有力矩；而杆可自由摆动，所以轴沿其自身方向（即通过 O 并垂直于我们这个平面的直线）不提供力偶。轻杆本身没有角动量（angular momentum），所以两端的力对这条直线的力矩必须互相抵消。作用于 (x,y) 的力 (F_x,F_y)，其力矩为 $xF_y - yF_x$（以逆时针方向为正）。对于作用于 $(l\sin\alpha,-l\cos\alpha)$ 的 $-\mathbf{R}_M$：

$$l\sin\alpha\,(-Mg) - (-l\cos\alpha)(M\omega^2 l\sin\alpha) = -Mgl\sin\alpha + M\omega^2 l^2\sin\alpha\cos\alpha,$$

而对于作用于 $(-l\sin\alpha,\,l\cos\alpha)$ 的 $-\mathbf{R}_m$：

$$(-l\sin\alpha)(-mg) - l\cos\alpha\,(-m\omega^2 l\sin\alpha) = mgl\sin\alpha + m\omega^2 l^2\sin\alpha\cos\alpha.$$

令两者之和为零，并除以 $l\sin\alpha \neq 0$（因为 $\alpha > 0$）：

$$(M+m)\,\omega^2 l\cos\alpha = (M-m)\,g \quad\Longrightarrow\quad \omega^2 = \left(\frac{M-m}{M+m}\right)\frac{g}{l\cos\alpha},$$

其中除以了 $\cos\alpha > 0$，这正是所需的结果。

假如 M 反而是在上方的质点，同样的计算（把 M 和 m 对调）会得出 $(M+m)\omega^2 l\cos\alpha = (m-M)g < 0$，这是不可能的。所以较重的质点确实如所假设的那样，位于 O 的下方。

轴所受的反作用力。 杆是轻的，所以作用于它的合力为零：轴对杆所施的力为 $\mathbf{R}_M + \mathbf{R}_m$，因此杆对轴所施的力为

$$\mathbf{F} = -(\mathbf{R}_M + \mathbf{R}_m) = \Big((M-m)\,\omega^2 l\sin\alpha,\ -(M+m)\,g\Big).$$

这个力指向下方，其水平分量则在 M 的一侧指向外。若 ϕ 是它与铅垂向下方向的夹角，则代入 ω^2 的值，可得

$$\begin{aligned}\tan\phi &= \frac{(M-m)\,\omega^2 l\sin\alpha}{(M+m)\,g} = \frac{M-m}{M+m}\cdot\frac{l\sin\alpha}{g}\cdot\frac{M-m}{M+m}\cdot\frac{g}{l\cos\alpha}\\ &= \left(\frac{M-m}{M+m}\right)^2\tan\alpha.\end{aligned}$$

由于 $0 \le \phi < \frac{1}{2}\pi$，$\phi = \tan^{-1}\left[\left(\frac{M-m}{M+m}\right)^2\tan\alpha\right]$，这就是所需的结果。

验算

- 若 $m = 0$，这个系统就是一个长度为 l 的圆锥摆（conical pendulum），而公式给出我们熟悉的 $\omega^2 = g/(l\cos\alpha)$。这时反作用力的角度为 α：杆沿着自身方向拉，就像一条绳一样。
- 若 $M = m$，杆在 O 处平衡，所以 $\omega = 0$，反作用力是铅垂的，大小为 $2Mg$；两条公式都与此相符。
- 数值上，取 $M = 3$、$m = 1$、$l = 0.7$、$\alpha = 0.6$ 和 $g = 9.8$，可得 $\omega^2 \approx 8.48$，反作用力的角度约为 9.71°。把杆的角度所满足的运动方程（equation of motion）由静止于 α 开始逐步数值计算，角度一直保持在 α；而在三维空间中由两个质点的位置计算它们的加速度，亦得出相同的角度。

常见错误

- 假设杆对质点所施的力沿着杆的方向。事实并非如此：刚性杆可以传递侧向的力，这正是我们要取力矩的原因。
- 忘记了 m 的加速度同样指向轴，因此与 M 的加速度位于相反的一侧。反作用力的水平分量会部分抵消，得出的是 $M-m$ 而不是 $M+m$。
- 把较重的质点放在上方。力矩方程显示这种情况不可能发生。

STEP 之外

杆对 O 的角动量并不平行于铅垂的自转轴，所以它会随着系统旋转而改变方向。要改变角动量就需要力偶，而这个力偶由轴提供，其方向垂直于我们刚才所用的那个方向。这就是运用惯性张量（inertia tensor）的刚体动力学（rigid-body dynamics）的起点，也解释了为甚么平衡不良的车轮会在轴承上摇晃。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对此有仔细的论述。

详细解答

STEP II 1988，第 15 题

概率与统计 • 离散均匀分布变量之和

STEP II 1988，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/88-s2-q15

关键思路

每份试卷的分数在 $\{0, 1, \ldots, m\}$ 上均匀分布（uniform distribution），所以每一组可能的分数出现的机会均等，而概率（probability）就是符合条件的分数组合数目除以所有组合的总数。两份试卷的数目只须简单列举便可得到；三份试卷的数目，则可先以第三份试卷的分数为条件，再把两份试卷的数目相加而得（这就是*卷积*（convolution））。之后利用对称性（symmetry）（把每份试卷的分数作 $k \mapsto m - k$ 的变换），最高的那个范围几乎不费工夫便可得出。

动笔之前先提醒一句：试卷上印出的三份试卷公式有一处印刷错误。下文会推导出正确的公式，并指出印刷版本错在哪里。

解答

设 X_1, X_2, X_3 为各份试卷的分数。由于各份试卷独立评分，它们是独立（independent）的，而且各自在 $\{0, 1, \ldots, m\}$ 上均匀分布。因此每个有序对 (X_1, X_2) 的概率是 $(m+1)^{-2}$，每个有序三元组的概率是 $(m+1)^{-3}$。记 $N_2(n)$ 和 $N_3(n)$ 分别为分数取自 $\{0, \ldots, m\}$ 而总和为 n 的有序对和有序三元组的数目。

两份试卷

我们需要 $X_1 = k$ 及 $X_2 = n - k$，而且两者都在 $[0, m]$ 之内，所以

$$\max(0,\, n - m) \le k \le \min(m,\, n).$$

- 若 $0 \le n \le m$：k 由 0 取到 n，所以 $N_2(n) = n + 1$。

▶ 若 $m < n \le 2m$：k 由 $n-m$ 取到 m，所以 $N_2(n) = m - (n-m) + 1 = 2m - n + 1$。

因此

$$P(X_1 + X_2 = n) = \begin{cases} \dfrac{n+1}{(m+1)^2}, & 0 \le n \le m, \\ \dfrac{2m-n+1}{(m+1)^2}, & m < n \le 2m, \end{cases}$$

这就是题目要求的 $m < n \le 2m$ 的结果，连同 $0 \le n \le m$ 的结果。这个分布形如“帐篷”：每一步上升一，直到 $n = m$，然后逐步下降。

三份试卷

以第三份试卷的分数为条件：若 $X_3 = k$，则需要 $X_1 + X_2 = n - k$，所以

$$N_3(n) = \sum_{k=0}^{m} N_2(n-k) = \sum_{j=n-m}^{n} N_2(j),$$

其中 $j = n - k$，而当 $j < 0$ 或 $j > 2m$ 时 $N_2(j) = 0$。

情况 $0 \le n \le m$。 只有 $j = 0, 1, \ldots, n$ 有贡献（负的 j 给出零），而且它们全都落在帐篷上升的一段：

$$N_3(n) = \sum_{j=0}^{n} (j+1) = \frac{(n+1)(n+2)}{2}, \qquad P(X_1 + X_2 + X_3 = n) = \frac{(n+1)(n+2)}{2(m+1)^3}.$$

情况 $m < n \le 2m$。 这时 $1 \le n - m \le m < n \le 2m$，所以 $j = n-m, \ldots, n$ 这个范围横跨帐篷的顶点。我们在顶点处把它分成两段：

$$N_3(n) = \sum_{j=n-m}^{m} (j+1) + \sum_{j=m+1}^{n} (2m - j + 1).$$

第一个和是把由 $n - m + 1$ 到 $m + 1$ 的整数相加：

$$\sum_{j=n-m}^{m} (j+1) = \frac{(m+1)(m+2) - (n-m)(n-m+1)}{2}.$$

在第二个和中，设 $i = 2m - j + 1$，它由 $2m - n + 1$ 取到 m：

$$\sum_{j=m+1}^{n} (2m - j + 1) = \frac{m(m+1) - (2m-n)(2m-n+1)}{2}.$$

把四个乘积展开，

$$\begin{aligned} 2N_3(n) &= (m^2 + 3m + 2) - (n^2 - 2mn + m^2 + n - m) \\ &\quad + (m^2 + m) - (4m^2 - 4mn + n^2 + 2m - n) \\ &= 6mn - 3m^2 - 2n^2 + 3m + 2. \end{aligned}$$

除以 $(m+1)^3$ 个机会均等的三元组，

$$P(X_1+X_2+X_3=n)=\frac{6mn-3m^2-2n^2+3m+2}{2(m+1)^3},\qquad m<n\le 2m.$$

? **想一想** 这并不是题目印出的公式。究竟谁对？

印刷版本以 $-4m^2$ 代替了 $-3m^2$，又以 $(m+1)^2$ 代替了 $(m+1)^3$；两处都是印刷错误。只须看一个例子便可定夺：取 $m=1$ 及 $n=2$，由 $\{0,1\}$ 取出的三个分数，在 8 种机会均等的情形中恰好有 3 种总和为 2，所以概率是 $\frac{3}{8}$。我们的公式给出 $\frac{12-3-8+3+2}{16}=\frac{3}{8}$，而印刷的公式却给出 $\frac{12-4-8+3+2}{8}=\frac{5}{8}$。无论如何，分母必定是 $(m+1)^3$，因为这正是机会均等的结果的总数。再多验算一次：当 $n=m$ 时，我们中间范围的公式给出 $\frac{(m+1)(m+2)}{2(m+1)^3}$，与第一种情况吻合；这是理所当然的，因为两个和在该处正好相接。

情况 $2m<n\le 3m$。 把每个分数 X_i 换成 $m-X_i$。新的分数同样在 $\{0,\dots,m\}$ 上均匀分布，而三个新分数仍然独立，所以

$$\begin{aligned}P(X_1+X_2+X_3=n)&=P\big((m-X_1)+(m-X_2)+(m-X_3)=n\big)\\&=P(X_1+X_2+X_3=3m-n).\end{aligned}$$

当 $2m<n\le 3m$ 时，有 $0\le 3m-n<m$，所以可把第一种情况中的 n 换成 $3m-n$ 来应用：

$$P(X_1+X_2+X_3=n)=\frac{(3m-n+1)(3m-n+2)}{2(m+1)^3},\qquad 2m<n\le 3m.$$

（同样的对称性，以 $2m-n$ 代替 $3m-n$，可把两份试卷在 $n\le m$ 时的结果变成 $n>m$ 时的结果。）

验算

- 小例子：当 $m=1$ 时，三条公式对 $n=0,1,2,3$ 分别给出 $\frac{1}{8},\frac{3}{8},\frac{3}{8},\frac{1}{8}$，它们的总和是 1，并且与掷三次公平硬币的二项分布（binomial distribution）一致。
- 中间范围的公式在 $n\mapsto 3m-n$ 之下对称：在 $6mn-3m^2-2n^2$ 中把 n 换成 $3m-n$，式子保持不变，正如对称性论证所要求的。
- 对由 1 到 8 的每一个 m 作完全列举，五条公式全部与实际数目完全相符；例如 $m=10$ 及 $n=15$ 时，两种方法都给出 $\frac{91}{1331}$。

常见错误

- 忘记 $N_2(j)$ 有两条不同的公式，在中间范围的求和中由头到尾都用 $j+1$。

- 一心想凑出印刷的中间公式，于是“硬砌”代数运算。如果仔细计算的结果与题目给出的答案不符，先用一个极小的例子检验，不要急于怀疑自己。

STEP 之外

独立随机变量（random variable）之和的分布，就是它们各自分布的*卷积*，而我们的和 $\sum_k N_2(n-k)$ 正是如此。概率母函数（probability generating function）可把这个过程变成机械式的运算：数目 $N_3(n)$ 就是 $(1+t+\cdots+t^m)^3$ 中 t^n 的系数。随着试卷的数目愈来愈多，分布的形状会由帐篷变成平滑的隆起，并趋近正态曲线（这就是中心极限定理（central limit theorem））。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）以合适的程度讲解母函数和中心极限定理。

详细解答

STEP II 1988，第 16 题

概率与统计 · 正态分布

STEP II 1988，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/88-s2-q16

关键思路

关于两根的一切，都由单一个随机变量（random variable）B 决定，所以每一个关于 X_1 和 X_2 的问题，都应该转化为“B 落在哪个区间”的问题。判别式（discriminant）告诉我们两根何时为实数，而韦达定理（Vieta's formulas）（两根之和为 $-2B$，两根之积为 1）令 (i) 和 (ii) 的条件很容易转换。之后，每个答案都是标准正态变量（standard normal variable）的条件概率（conditional probability）或条件期望值（conditional expectation），而求期望值只需计算一个简短的积分。

解答

全题设 $B \sim \mathrm{N}(0,1)$，其密度函数（density）为 $\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}$，分布函数（distribution function）为 Φ。设 R 为“两根皆为实数”这个事件。

实根

$X^2 + 2BX + 1$ 的判别式为 $4B^2 - 4$，所以两根为实数当且仅当 $B^2 \geq 1$，即 $|B| \geq 1$。由正态分布（normal distribution）的对称性，

$$\mathrm{P}(R) = \mathrm{P}(|B| \geq 1) = 2\big(1 - \Phi(1)\big) = 2(1 - 0.84134) = 0.31731\ldots \approx 0.317.$$

（是否包括等根的情况 $|B| = 1$ 并无分别，因为它的概率为零。）

当两根为实数时，它们是

$$X_{1,2} = -B \pm \sqrt{B^2 - 1}, \qquad X_1 + X_2 = -2B, \qquad X_1 X_2 = 1.$$

第 (i) 部分

想一想 怎样把“两根都大于 $\frac{1}{5}$”转化为关于 B 的条件?

首先，两根必须都是正数。两根之积为 $1>0$，所以它们同号，而这个符号就是两根之和 $-2B$ 的符号。因此我们需要 $B<0$，连同 R 即表示 $B\le -1$。记 $b=-B\ge 1$；这时两根为 $b\pm\sqrt{b^2-1}$，都是正数。

由于两根之积为 1，较小的根是较大的根的倒数：

$$b-\sqrt{b^2-1}=\frac{1}{b+\sqrt{b^2-1}}.$$

所以较小的根大于 $\frac{1}{5}$ 当且仅当较大的根小于 5，而这时较大的根自然也大于 $\frac{1}{5}$。因此条件就是

$$\sqrt{b^2-1}<5-b.$$

这需要 $b<5$，而当 $1\le b<5$ 时两边皆为非负数，所以可以把两边平方而不改变不等式：

$$b^2-1<25-10b+b^2 \iff 10b<26 \iff b<2.6.$$

（$b\ge 5$ 的值无论如何都不符合，因为 $2.6<5$。）因此，在两根为实数的条件下，两根都大于 $\frac{1}{5}$ 当且仅当 $-2.6<B\le -1$，于是

$$\begin{aligned}\mathrm{P}\big(X_1,X_2>\tfrac{1}{5}\,\big|\,R\big)&=\frac{\mathrm{P}(-2.6<B\le -1)}{\mathrm{P}(|B|\ge 1)}=\frac{\Phi(2.6)-\Phi(1)}{2\big(1-\Phi(1)\big)}\\&=\frac{0.99534-0.84134}{0.31731}=0.4853\ldots\approx 0.485.\end{aligned}$$

第 (ii) 部分

由于 $|X_1+X_2|=2|B|$，我们要求的是 $\mathrm{E}\big(2|B|\,\big|\,|B|\ge 1\big)$。要计算给定某事件下的条件期望值，我们在该事件上积分，再除以该事件的概率：

$$\mathrm{E}\big(|B|\,\big|\,|B|\ge 1\big)=\frac{1}{\mathrm{P}(|B|\ge 1)}\int_{|z|\ge 1}|z|\,\phi(z)\,\mathrm{d}z=\frac{2\int_1^\infty z\,\phi(z)\,\mathrm{d}z}{2\big(1-\Phi(1)\big)},$$

这里用到了 ϕ 的对称性。这个积分很容易计算，因为 $z\,\phi(z)$ 正是 $\phi(z)$ 的导数的负值：

$$\int_1^\infty z\,\phi(z)\,\mathrm{d}z=\Big[-\phi(z)\Big]_1^\infty=\phi(1)=\frac{\mathrm{e}^{-1/2}}{\sqrt{2\pi}}=0.24197\ldots$$

因此

$$\mathrm{E}\big(|X_1+X_2|\,\big|\,R\big)=\frac{2\phi(1)}{1-\Phi(1)}=\frac{2\times 0.24197}{0.15866}=3.050\ldots\approx 3.05.$$

验算

- 试取 $B = -2.6$：方程为 $X^2 - 5.2X + 1 = 0$，两根恰好是 5 和 $\frac{1}{5}$，正是 (i) 的边界情况。
- (ii) 的答案必须大于 2，因为在 R 上 $2|B| \geq 2$。答案确实如此，而且合理：正态分布在 1 以外的尾部很薄，所以 $|B|$ 通常只比 1 大一点（它的条件平均值约为 1.53）。
- 以 2×10^6 个标准正态的 B 值进行模拟，得出 0.3177、0.4855 和 3.049，与 0.317、0.485 和 3.05 吻合。

常见错误

- 忘记取条件：在 (i) 和 (ii) 中，我们必须除以 $\mathrm{P}(R) = 0.317$。无条件概率 $\Phi(2.6) - \Phi(1) = 0.154$ 并不是答案。
- 在 (i) 中把 $B \geq 1$ 也计算在内。这时两根都是负数，所以没有一个大于 $\frac{1}{5}$。
- 把 $\mathrm{E}|X_1 + X_2|$ 当作 $|\mathrm{E}(X_1 + X_2)|$ 来计算，而后者由对称性可知等于 0。

STEP 之外

$\phi(1)/(1 - \Phi(1))$ 这个量是*截尾*（truncated）正态分布的平均值，而比值 $\phi(a)/(1 - \Phi(a))$ 称为逆米尔斯比率（inverse Mills ratio）；它出现于统计学（例如设限数据（censored data）的模型），也出现于可靠性理论（reliability theory），在那里它是正态分布的风险率（hazard rate）。关于随机多项式（random polynomials）的问题，例如系数随机的多项式平均有多少个实根，构成了概率论中一个丰富得出人意表的领域。关于条件期望值和连续分布，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门），或 Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论导引）。

详细解答

STEP III 1988，第 1 题

纯数学 · 曲线描绘与界限

STEP III 1988，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/88-s3-q1

关键思路

绘画草图时，因子 x^2e^{-x} 永不为负，所以 y 的正负完全由 $1+x$ 决定。有了这一点，再加上一次求导，以及曲线在两端和 $x=-1$ 附近的表现，整条曲线的形状便清楚了。至于积分，最干净利落的做法是把 x^2 除以 $1+x$：这样被积函数（integrand）的大部分积分后恰好互相抵消，余下的部分则很容易夹在两个简单的界之间。

解答

图像

设 $f(x)=\dfrac{x^2e^{-x}}{1+x}$，定义于所有 $x\neq -1$。

正负与零点。由于 $x^2\geq 0$ 而 $e^{-x}>0$，当 $x<-1$ 时 $f(x)<0$，当 $x>-1$ 时 $f(x)\geq 0$，而且只有在 $x=0$ 时才有 $f(x)=0$。因子 x^2 表示曲线在原点只是与 x 轴相切，而不是穿过它。

渐近线（asymptote）与两端的表现。

- 当 $x\to -1$ 时，分子趋向 $e>0$，分母趋向 0，所以 $x=-1$ 是一条铅垂渐近线；当 $x\to -1^-$ 时 $f\to -\infty$，当 $x\to -1^+$ 时 $f\to +\infty$。
- 当 $x\to +\infty$ 时，$f(x)\approx xe^{-x}\to 0^+$，因为指数函数的增长胜过幂函数。所以 $y=0$ 是右方的渐近线，曲线从上方趋近它。
- 当 $x\to -\infty$ 时，$\dfrac{x^2}{1+x}\approx x\to -\infty$ 而 $e^{-x}\to +\infty$，所以 $f(x)\to -\infty$，而且下降得非常快。

驻点（stationary point）。根据商法则（quotient rule），

$$f'(x) = \frac{(2x - x^2)\mathrm{e}^{-x}(1+x) - x^2\mathrm{e}^{-x}}{(1+x)^2} = \frac{\mathrm{e}^{-x}\left(2x + x^2 - x^3 - x^2\right)}{(1+x)^2}$$
$$= \frac{x\left(2 - x^2\right)\mathrm{e}^{-x}}{(1+x)^2}.$$

所以在 $x = 0$ 和 $x = \pm\sqrt{2}$ 时 $f'(x) = 0$。因子 e^{-x} 和 $(1+x)^2$ 都是正数，所以 f' 的正负与 $x(2 - x^2)$ 的正负相同：

区间	$x < -\sqrt{2}$	$-\sqrt{2} < x < -1$	$-1 < x < 0$	$0 < x < \sqrt{2}$	$x > \sqrt{2}$
$f'(x)$	$+$	$-$	$-$	$+$	$-$

由此得出：

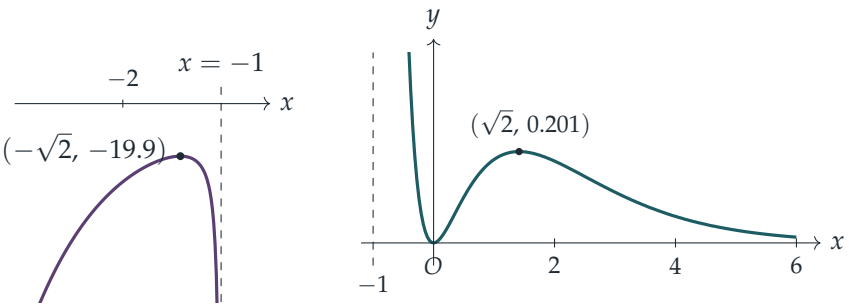
- 在 $x = \sqrt{2}$ 处有一个局部极大点（local maximum），该处 $y = \dfrac{2\mathrm{e}^{-\sqrt{2}}}{1+\sqrt{2}} = 2(\sqrt{2} - 1)\mathrm{e}^{-\sqrt{2}} \approx 0.201$；
- 在原点有一个局部极小点（local minimum）；
- 在左边分支上的 $x = -\sqrt{2}$ 处有一个局部极大点，该处 $y = \dfrac{2\mathrm{e}^{\sqrt{2}}}{1-\sqrt{2}} = -2(\sqrt{2} + 1)\mathrm{e}^{\sqrt{2}} \approx -19.9$。

想一想 一条完全位于横轴下方的分支，怎可能有极大点？

在 $x < -1$ 的范围内，曲线随着 $x \to -\infty$ 从 $-\infty$ 升上来，到了渐近线 $x = -1$ 又再跌回 $-\infty$，所以它必定在中间某处转向。最高点位于 $x = -\sqrt{2}$，而这一点仍然远在横轴之下。这条分支是一个悬在横轴下方的“小丘”，丘顶的高度约为 -19.9。

在右边，曲线从紧贴 $x = -1$ 右侧的 $+\infty$ 下降，到原点时降至 0，然后上升至 $x = \sqrt{2}$ 处那个小小的极大点，其后衰减至 0。曲线在原点附近向上凹，在极大点处向下凹，所以在 $x = \sqrt{2}$ 的两侧各有一个拐点（point of inflection）（数值上约在 $x \approx 0.35$ 和 $x \approx 2.5$ 附近）。绘画草图时，我们不需要知道它们的准确位置。

两条分支在纵向上的尺度相差极大，所以我们把它们分开绘画。在左边，y 一直下降至约 -70。在右边，我们把 y 轴的比例放大，让 0.2 这个极大值清楚可见。



积分

设 $I=\int_0^\infty \frac{x^2\mathrm{e}^{-x}}{1+x}\,\mathrm{d}x$。把 x^2 除以 $1+x$，得

$$\frac{x^2}{1+x}=x-1+\frac{1}{1+x}.$$

现在 $\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x=1$（取 $u=x$ 作分部积分（integration by parts)），而 $\int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x=1$，所以首两项的贡献是 $1-1=0$，于是

$$I=\int_0^\infty \frac{\mathrm{e}^{-x}}{1+x}\,\mathrm{d}x.$$

（三个积分全部收敛（converge），所以把积分拆开是合理的。）当 $x>0$ 时，我们有 $0<\frac{1}{1+x}<1$，所以

$$0<\frac{\mathrm{e}^{-x}}{1+x}<\mathrm{e}^{-x}\qquad(x>0).$$

这些不等式在整个区间上都是严格的，而且各函数均为连续，所以积分后不等式仍然严格成立：

$$0<I<\int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x=1.$$

所以 I 介乎 0 与 1 之间，正是题目所求。

另一种做法。我们可以略去除法这一步。当 $x>0$ 时，$\frac{x^2}{1+x}<\frac{x^2}{x}=x$，所以 $0<f(x)<x\mathrm{e}^{-x}$，从而 $0<I<\int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x=1$。不过那个除法仍然值得认识，因为它清楚显示了 I 究竟从何而来。

验算

- 在一千个随机选取的点上以数值方法对 f 求导，结果与 $f'(x)=x(2-x^2)\mathrm{e}^{-x}/(1+x)^2$ 吻合。

- 在 $[0,60]$ 上使用辛普森法则（Simpson's rule）得出 $I \approx 0.59635$，而 $\int_0^\infty \frac{\mathrm{e}^{-x}}{1+x}\,\mathrm{d}x$ 亦给出相同的数值，稳稳地落在 0 与 1 之间。
- 从草图作粗略估算：在 $x > 0$ 的范围内，曲线是一个高度约为 0.2 的小丘，到了 $x = 6$ 已大致消退。以此为高和底的三角形，面积约为 0.6，与 0.596 相符。

常见错误

- 遗漏了 $x = -\sqrt{2}$ 处的驻点，因为它位于“另一条”分支上；又或者因为该处 y 是负数，便把它说成极小点。
- 说曲线在原点穿过横轴。二重因子 x^2 令曲线只是与横轴相切。
- 单凭 $\mathrm{e}^{-x} \le 1$ 便声称 $I < 1$。这个界只会得出 $\int_0^\infty \frac{x^2}{1+x}\,\mathrm{d}x$，而这个积分是发散（diverge）的。

STEP 之外

I 的值是 $\mathrm{e}\,E_1(1) \approx 0.5963$，其中 $E_1(t) = \int_t^\infty \mathrm{e}^{-u}u^{-1}\,\mathrm{d}u$ 是指数积分（exponential integral）。这个数称为 Gompertz 常数（Gompertz constant）。它不能用初等函数（elementary functions）表示，这正是题目只要求找出上下界的原因。对于无法求出确切值的积分，拿它与能够求值的积分作比较，从而作出估计，是分析学（analysis）的一项核心技巧（即瑕积分（improper integral）的“比较判别法”（comparison test））。Abbott 的 *Understanding Analysis*（一本分析学入门教科书）对此有细致的阐述。

详细解答

STEP III 1988，第 2 题

纯数学 · 线性递推关系

STEP III 1988，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/88-s3-q2

关键思路

常系数线性递归关系（linear recurrence with constant coefficients）的表现就像一条二次方程：幂 x^n 是它的解，当且仅当 x 是辅助方程（auxiliary equation）的根。由于只要知道首两项，递归关系便决定了其后每一项，所以只需配对 u_0 和 u_1，就能得到唯一的解。最后一部分的系数是可变的，但因子 $(n+1)$、$(n+2)$、$(n+3)$ 提示我们把全式除以它们的乘积，然后用一个简单的代换（substitution），便可把它化为第一种的常系数递归关系。

解答

（试卷上把递归关系的首项系数印成 α。这显然是 a 的误植，应与二次式 ax^2+bx+c 一致，我们就按此理解。）由于这条二次方程有两个根，它确实是二次的，所以 $a \neq 0$。

公式满足递归关系

设 $u_n = A\alpha^n + B\beta^n$，其中 A、B 为任意常数。那么

$$\begin{aligned} au_{n+2} + bu_{n+1} + cu_n &= A\alpha^n\left(a\alpha^2 + b\alpha + c\right) + B\beta^n\left(a\beta^2 + b\beta + c\right) \\ &= A\alpha^n \cdot 0 + B\beta^n \cdot 0 = 0, \end{aligned}$$

因为 α 和 β 是 $ax^2+bx+c=0$ 的根。所以递归关系对每个 $n \geq 0$ 都成立。

取题目所给的特定 A 和 B 时，公式亦能重现初始值。当 $n=0$，

$$A + B = \frac{(u_1 - \beta u_0) - (u_1 - \alpha u_0)}{\alpha - \beta} = \frac{(\alpha - \beta)u_0}{\alpha - \beta} = u_0,$$

而当 $n=1$，

$$A\alpha+B\beta=\frac{\alpha u_1-\alpha\beta u_0-\beta u_1+\alpha\beta u_0}{\alpha-\beta}=\frac{(\alpha-\beta)u_1}{\alpha-\beta}=u_1.$$

由于两根相异，所以可以除以 $\alpha-\beta$。

它是唯一的解

想一想 我们究竟是对甚么作归纳？为甚么每次需要两项？

假设 $u_0,u_1,u_2,\ldots$ 是任何一个满足该递归关系的数列（sequence），并设 A、B 如上由它的 u_0 和 u_1 构造而成。记 $w_n=A\alpha^n+B\beta^n$。我们对 n 用归纳法（induction）证明以下命题：

$$P(n):\quad u_n=w_n\text{ 及}u_{n+1}=w_{n+1}.$$

每个新的项都取决于它之前的两项，所以命题必须同时包含连续两项。

基础步骤（*base case*）。$P(0)$ 是说 $u_0=w_0$ 及 $u_1=w_1$，这正是我们在上面验证过的。

归纳步骤（*inductive step*）。假设 $P(n)$ 成立。(u_n) 和 (w_n) 都满足递归关系，而且 $a\neq0$，所以

$$u_{n+2}=-\frac{bu_{n+1}+cu_n}{a}=-\frac{bw_{n+1}+cw_n}{a}=w_{n+2}.$$

连同 $u_{n+1}=w_{n+1}$，这就是 $P(n+1)$。

所以 $P(n)$ 对所有 $n\geq0$ 成立，即对每个 n 都有 $u_n=A\alpha^n+B\beta^n$。递归关系的任何解都具有这个形式，其中 A 和 B 由 u_0 和 u_1 决定。这条公式给出的就是唯一的解。

数列 v_n

每个系数都是 $(n+1)$、$(n+2)$、$(n+3)$ 这三个因子中其中两个的乘积，而各自缺少的因子都不相同。除以 $(n+1)(n+2)(n+3)$（当 $n\geq0$ 时它永不为零），得

$$8\,\frac{v_{n+2}}{n+3}-2\,\frac{v_{n+1}}{n+2}-\frac{v_n}{n+1}=0.$$

每个分数都具有 $v_k/(k+1)$ 的形式。因此设

$$w_n=\frac{v_n}{n+1},$$

递归关系便变成

$$8w_{n+2}-2w_{n+1}-w_n=0,\qquad w_0=0,\quad w_1=\tfrac12.$$

辅助方程是 $8x^2-2x-1=0$，即 $(2x-1)(4x+1)=0$。它的根 $\alpha=\frac12$ 和 $\beta=-\frac14$ 是相异实根，所以第一部分的结果适用，其中 $\alpha-\beta=\frac34$：

$$A=\frac{w_1-\beta w_0}{\alpha-\beta}=\frac{1/2}{3/4}=\frac23,\qquad B=\frac{w_1-\alpha w_0}{\beta-\alpha}=\frac{1/2}{-3/4}=-\frac23.$$

因此 $w_n = \frac{2}{3}\left(\left(\frac{1}{2}\right)^n - \left(-\frac{1}{4}\right)^n\right)$。第一部分亦告诉我们，这是 w_n 唯一的可能，因而也是 v_n 唯一的可能。乘回 $n+1$，得

$$\boxed{v_n = \frac{2(n+1)}{3}\left(\frac{1}{2^n} - \left(-\frac{1}{4}\right)^n\right)} \qquad (n = 0, 1, 2, \ldots).$$

验算

公式给出 $v_0 = \frac{2}{3}(1-1) = 0$ 和 $v_1 = \frac{4}{3}\left(\frac{1}{2} + \frac{1}{4}\right) = 1$，符合要求。以 $n = 0$ 和 $n = 1$ 用人手计算原来的递归关系，得 $16v_2 = 6v_1 + 6v_0$，所以 $v_2 = \frac{3}{8}$；然后 $48v_3 = 16v_2 + 12v_1$，所以 $v_3 = \frac{3}{8}$。公式的结果与此吻合：$v_2 = 2\left(\frac{1}{4} - \frac{1}{16}\right) = \frac{3}{8}$ 及 $v_3 = \frac{8}{3}\left(\frac{1}{8} + \frac{1}{64}\right) = \frac{3}{8}$。我们以有理数运算，对 n 直至 41 精确地核对过公式与递归关系相符，亦以数百组随机的二次方程和初始值核对过第一部分。

常见错误

- 每次只对一项作归纳。单凭 $u_n = w_n$，无法推出任何关于 u_{n+1} 的结论：归纳假设必须涵盖连续两项。
- 忘记说明为何可以除以 a（因为方程有两个根，所以它不为零）和除以 $\alpha - \beta$（两根相异）。
- 直接把第一部分应用于 v_n。它的系数取决于 n，所以第一部分适用的，是经重新缩放的数列 $w_n = v_n/(n+1)$。

STEP 之外

二阶线性递归关系的解构成一个二维向量空间（vector space），而 α^n、β^n 是它的一组基（basis）。这与微分方程（differential equation）$ay'' + by' + cy = 0$ 及其解 $e^{\alpha x}$ 和 $e^{\beta x}$ 的情况完全平行。把递归关系写成 $\mathbf{u}_{n+1} = M\mathbf{u}_n$，其中 M 是一个 2×2 矩阵，便可看出 α 和 β 是 M 的特征值（eigenvalue），而重根则对应于不能对角化（diagonalise）的矩阵。Strang 的 *Introduction to Linear Algebra*（线性代数入门）正是这样通过特征值来处理差分方程（difference equation），而 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）则为微分方程建立了平行的理论。

详细解答

STEP III 1988，第 3 题

纯数学 · 复数轨迹

STEP III 1988，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/88-s3-q3

关键思路

映射 $w = (z+\mathrm{i})/(z-\mathrm{i})$ 可以倒过来用：解出 z 得 $z = \mathrm{i}(w+1)/(w-1)$，于是任何关于 z 的条件都变成关于 w 的条件。至于那个圆，则改用参数表示（parametrisation）来处理，因为这时 $z-\mathrm{i}$ 就只是 $2\mathrm{e}^{\mathrm{i}\theta}$。对于实数和虚数两种情况，可从模（modulus）和辐角（argument）着眼：实数 z 与 i 和 $-\mathrm{i}$ 等距。

解答

参数形式

曲线 $|z-\mathrm{i}| = 2$ 是以 i 为圆心、半径为 2 的圆，所以

$$z = \mathrm{i} + 2\mathrm{e}^{\mathrm{i}\theta}, \qquad -\pi < \theta \le \pi,$$

或者，设 $z = x + \mathrm{i}y$，则 $x = 2\cos\theta$ 及 $y = 1 + 2\sin\theta$。

我们还会用到逆映射（inverse map）。由 $w(z-\mathrm{i}) = z+\mathrm{i}$ 得 $z(w-1) = \mathrm{i}(w+1)$，所以当 $w \neq 1$ 时

$$z = \frac{\mathrm{i}(w+1)}{w-1}.$$

没有任何 z 会给出 $w = 1$，因为那需要 $z+\mathrm{i} = z-\mathrm{i}$。每个 $w \neq 1$ 都来自恰好一个 $z \neq \mathrm{i}$。

第 (i) 部分

代入参数形式，$z+\mathrm{i} = 2\mathrm{i} + 2\mathrm{e}^{\mathrm{i}\theta}$ 及 $z-\mathrm{i} = 2\mathrm{e}^{\mathrm{i}\theta}$，所以

$$w = \frac{2\mathrm{i} + 2\mathrm{e}^{\mathrm{i}\theta}}{2\mathrm{e}^{\mathrm{i}\theta}} = 1 + \mathrm{i}\,\mathrm{e}^{-\mathrm{i}\theta} = 1 + \mathrm{e}^{\mathrm{i}(\pi/2-\theta)}.$$

当 θ 绕行一周时，$\pi/2-\theta$ 也恰好取遍每个角度一次。所以 w 描出整个圆

$$|w-1|=1,$$

其圆心为 1，半径为 1。例如，$z=3\mathrm{i}$ 给出 $w=2$，而 $z=-\mathrm{i}$ 给出 $w=0$。点 $w=1$ 不在这个圆上，这与它永远取不到的事实一致。

第 (ii) 部分

若 $z=x$ 是实数，它与 i 的距离等于它与 $-\mathrm{i}$ 的距离，所以 $|x+\mathrm{i}|=|x-\mathrm{i}|$，而且

$$|w|=\frac{|x+\mathrm{i}|}{|x-\mathrm{i}|}=1.$$

具体地说，把分子和分母同乘以 $x+\mathrm{i}$，

$$w=\frac{(x+\mathrm{i})^2}{x^2+1}=\frac{x^2-1}{x^2+1}+\mathrm{i}\,\frac{2x}{x^2+1}.$$

想一想 单位圆上的每一点都能取到吗?

不，还差一点。任取满足 $|w|=1$ 且 $w\neq 1$ 的 w，并写成 $w=\mathrm{e}^{\mathrm{i}\phi}$。那么

$$z=\mathrm{i}\,\frac{\mathrm{e}^{\mathrm{i}\phi}+1}{\mathrm{e}^{\mathrm{i}\phi}-1}=\mathrm{i}\,\frac{2\cos(\phi/2)}{2\mathrm{i}\sin(\phi/2)}=\cot\tfrac{\phi}{2},$$

这是实数。所以每个这样的点都能取到，但 $w=1$ 永远取不到：它是 $x\to\pm\infty$ 时的极限。所求轨迹（locus）是单位圆 $|w|=1$ 除去 $w=1$ 这一点。例如，$x=0$ 给出 $w=-1$，而 $x=1$ 给出 $w=\mathrm{i}$。

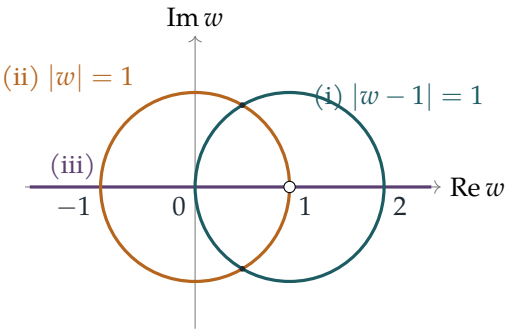
第 (iii) 部分

若 $z=\mathrm{i}y$，其中 y 为实数且 $y\neq 1$，则

$$w=\frac{\mathrm{i}(y+1)}{\mathrm{i}(y-1)}=\frac{y+1}{y-1}=1+\frac{2}{y-1},$$

这是实数。当 y 取遍 1 以外的所有实数时，分数 $2/(y-1)$ 取遍每个非零实数值，所以 w 取遍 1 以外的每个实数值。（反过来，若 $w=u$ 是实数且 $u\neq 1$，则 $z=\mathrm{i}(u+1)/(u-1)$ 是虚数。）所求轨迹是实轴除去 $w=1$ 这一点。这里 $z=0$ 给出 $w=-1$；如果你把“虚数”理解为不包括 $z=0$，那么 $w=-1$ 也要除去。

略图



蓝绿色的圆是 (i) 的轨迹，琥珀色的圆（除去点 1）是 (ii) 的轨迹，紫色的轴（除去点 1）是 (iii) 的轨迹。空心圆点标示 $w = 1$，这一点永远取不到。

验算

- (i) 中的圆与实轴相交于 $z = \pm\sqrt{3}$。这两点应该映射到同时位于蓝绿色圆和琥珀色圆上的点，而事实上 $w(\sqrt{3}) = (\sqrt{3}+\mathrm{i})^2/4 = \frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i} \approx 0.5 + 0.866\mathrm{i}$，它确实既在 $|w| = 1$ 上，也在 $|w-1| = 1$ 上。
- (i) 中的圆与虚轴相交于 $3\mathrm{i}$ 和 $-\mathrm{i}$，它们分别映射到 2 和 0：正好是蓝绿色圆与实轴相交之处。
- 在每条轨迹上各取 10^4 个随机点作数值检验，结果确认了 $|w-1| = 1$、$|w| = 1$ 和 $\operatorname{Im} w = 0$，并显示逆映射能还原出 z。

常见错误

- 忘记 (ii) 和 (iii) 中缺少了 $w = 1$。它对应于“$z = \infty$”。
- 在 (i) 中展开成 x 和 y，结果迷失在代数运算之中。参数表示令 $z - \mathrm{i}$ 变成单一个指数式，余下的只需一行。

STEP 之外

映射 $z \mapsto (z+\mathrm{i})/(z-\mathrm{i})$ 是一个莫比乌斯变换（Möbius transformation）。这类映射总是把圆和直线映射成圆和直线，这正是你在三个部分中所发现的。要理解它们，最好放在黎曼球面（Riemann sphere）上看：在那里加入一个“无穷远点”（point at infinity），便填补了 $w = 1$ 处的缺口。它的近亲凯莱变换（Cayley transform）$z \mapsto (z-\mathrm{i})/(z+\mathrm{i})$ 把上

半平面（upper half-plane）映射到单位圆盘（unit disc）上。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）对这一切有十分优美的几何处理。

详细解答

STEP III 1988，第 4 题

纯数学 · 弧长的最优化

> **STEP III 1988，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/88-s3-q4

关键思路

把道路的总长度写成函数 $L(b)$，然后求导（differentiate）。位于铅垂平面内的两段路必须贴着地面走，所以它们的长度是抛物线的弧长（arc length）；半圆的长度是 πb。把弧长积分对其下限求导，根本不需要做任何积分，而 $L'(b)$ 的正负号就决定了一切。

（题目印出的横截面公式左边是 $x=$；当然应该是 $z=$。）

解答

建立长度的表达式

设 $h(b)=(a^2-b^2)/(2a)$，即山丘在距离对称轴 b 处的高度。道路必须铺在地面上，所以在 (x,z) 平面内由 $(-R,0)$ 到 $(-b,h(b))$ 的路径别无选择：它先沿平地由 $x=-R$ 走到 $x=-a$，再沿抛物线由 $x=-a$ 上行至 $x=-b$。第 (iii) 段是它的镜像。

在山丘上 $z=(a^2-x^2)/(2a)$，所以 $\dfrac{\mathrm{d}z}{\mathrm{d}x}=-\dfrac{x}{a}$，而由 $x=b$ 到 $x=a$ 的弧长是

$$\int_b^a \sqrt{1+\frac{x^2}{a^2}}\,\mathrm{d}x.$$

半圆 (ii) 是水平的，位于高度 $h(b)$，半径为 b（它是山丘上通过 $(\pm b,h(b))$ 的纬圈），所以长度是 πb。因此

$$L(b)=2(R-a)+2\int_b^a \sqrt{1+\frac{x^2}{a^2}}\,\mathrm{d}x+\pi b,\qquad 0\le b\le a.$$

这个式子也涵盖了 $b=0$（没有半圆；道路直接越过山顶）和 $b=a$（道路绕着山脚走）两种情况。

求最小值

根据微积分基本定理（fundamental theorem of calculus），对下限求导，所得的是被积函数（integrand）在该处的值再加上负号：

$$L'(b) = \pi - 2\sqrt{1 + \frac{b^2}{a^2}}.$$

想一想 驻点（stationary point）会在哪里？它在容许的范围内吗？

$L'(b) = 0$ 需要 $1 + b^2/a^2 = \pi^2/4$，即 $b = \frac{a}{2}\sqrt{\pi^2 - 4} \approx 1.21a$。这比 a 大，所以落在容许范围 $0 \le b \le a$ 之外。事实上，当 $0 \le b \le a$ 时，我们有

$$2\sqrt{1 + \frac{b^2}{a^2}} \le 2\sqrt{2} \approx 2.83 < \pi,$$

所以在整个区间上 $L'(b) > 0$。L 严格递增（strictly increasing），其最小值出现在左端点 $b = 0$。

建议： 取 $b = 0$。道路应该从山丘的一边直上，再从另一边直下，不必绕半圆的弯路。

省下多少路程

顺带一提，原函数（antiderivative）是

$$\int \sqrt{1 + \frac{x^2}{a^2}}\,\mathrm{d}x = \frac{x}{2}\sqrt{1 + \frac{x^2}{a^2}} + \frac{a}{2}\sinh^{-1}\frac{x}{a} + C,$$

（代入 $x = a\sinh t$），所以

$$L(0) = 2(R - a) + a\left(\sqrt{2} + \ln(1 + \sqrt{2})\right) \approx 2(R - a) + 2.296a,$$

而绕山脚走则得 $L(a) = 2(R - a) + \pi a \approx 2(R - a) + 3.142a$。越过山顶大约可省下 $0.85a$。

验算

- 答案与 R 无关，这是理所当然的：不论 b 是多少，平地部分的总长度都是 $2(R - a)$。
- 粗略检查一下是否合理：山丘只有 $a/2$ 高、$2a$ 宽，所以越过它不可能比直线距离 $2a$ 多出很多；而绕山脚的半圆长 πa。果然 $2.296a$ 介乎 $2a$ 与 πa 之间。
- 数值上，取 $a = 1$ 和 $R = 3$，在 $[0, 1]$ 的 201 个点上以数值积分（quadrature）计算 $L(b)$，最小值出现在 $b = 0$；而 $L'(a) = \pi - 2\sqrt{2} \approx 0.313$ 仍然是正数。

常见错误

- 令 $L'(b) = 0$ 并写下 $b = \frac{a}{2}\sqrt{\pi^2 - 4}$，却没有留意到它比 a 大。
- 第 (i) 段和第 (iii) 段用了直线（弦）的长度：那样的路会穿山而过。
- 忘记了闭区间（closed interval）上的最小值可以出现在端点。

STEP 之外

在曲面上寻找最短路径，就是测地线（geodesics）的问题，在变分法（calculus of variations）或微分几何（differential geometry）的入门课程中会遇到。在旋转曲面（surface of revolution）上，克莱罗关系式（Clairaut's relation）简洁地刻画了测地线；国王的这一族道路，只是该问题中一个小小的单参数切片。这里用到的两个分析学事实——对积分的上下限求导，以及闭区间上的连续函数（continuous function）必定取得最小值——在 Abbott 的 *Understanding Analysis* 中有严谨的证明。

详细解答

STEP III 1988，第 5 题

纯数学 · 递推关系与贷款

STEP III 1988，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/88-s3-q5

关键思路

先写下债务在一天之内的变化。每天债务先乘以 $1+k$（利息），然后加上（挖掘成本）或减去（收入）一个固定金额。形如 $S_n = (1+k)S_{n-1} +$ 常数 的递推关系（recurrence）逐层展开后，就是一个等比级数（geometric series）。最后一部分则归结为一个比较：隧道通车那一刻，每日收入是否多于债务的利息？

解答

挖掘隧道期间的债务

利率为 $100k\%$，即 £S 的债务每天的利息是 £kS。设 S_n 为第 n 天结束时的债务（以英镑计），其中 $S_0 = 0$。在第 $n \le T$ 天结束时，公司要借入 c 支付挖掘费用，另借入 kS_{n-1} 支付利息，所以

$$S_n = S_{n-1} + kS_{n-1} + c = (1+k)S_{n-1} + c.$$

由 $S_0 = 0$ 开始，反复运用这个关系，

$$\begin{aligned} S_n &= c + (1+k)S_{n-1} \\ &= c + c(1+k) + (1+k)^2 S_{n-2} \\ &= \cdots = c[1 + (1+k) + (1+k)^2 + \cdots + (1+k)^{n-1}]. \end{aligned}$$

这是一个共有 n 项、公比（common ratio）为 $1+k \neq 1$（因为 $k > 0$）的等比级数，所以

$$S_n = \frac{c[(1+k)^n - 1]}{k}, \qquad n \le T,$$

正是所求的结果。（为严谨起见，可用归纳法（induction）验证：公式给出 $S_0 = 0$；若公式对 S_{n-1} 成立，则 $(1+k)S_{n-1} + c = \frac{c}{k}[(1+k)^n - (1+k) + k]$，这正是 S_n 的公式。）

隧道通车后的债务

由第 $T+1$ 天起，不再需要挖掘。每天债务增加利息 kS，而公司还款 d，所以只要债务仍然是正数，就有

$$S_{T+j} = (1+k)S_{T+j-1} - d.$$

这与之前的递推关系相同，只是 c 换成了 $-d$，而且现在是由 S_T 而不是由 0 开始。展开 m 次，

$$\begin{aligned} S_{T+m} &= (1+k)^m S_T - d[1 + (1+k) + \cdots + (1+k)^{m-1}] \\ &= (1+k)^m S_T - \frac{d[(1+k)^m - 1]}{k}. \end{aligned}$$

代入第一部分求得的 S_T，

$$S_{T+m} = \frac{c[(1+k)^T - 1](1+k)^m - d[(1+k)^m - 1]}{k}.$$

我们需要 $S_{T+m} > 0$，因为这个递推关系只在债务尚未还清时适用。到了还清债务的那一天，公司只需偿还所欠的余额，而不是整笔 d。

债务何时还清

把含 $(1+k)^m$ 的项合并：

$$S_{T+m} = \frac{d}{k} - \frac{(1+k)^m}{k}\Big(d - c[(1+k)^T - 1]\Big).$$

想一想 从金钱的角度看，条件 $d/c > (1+k)^T - 1$ 究竟是甚么意思？

把它乘以 c：条件变成 $d > c[(1+k)^T - 1] = kS_T$。换句话说，第一天的收入多于隧道通车时债务的利息。因此债务在第 $T+1$ 天下降，利息亦随之减少，于是债务下降得愈来愈快。以下的代数运算把这个想法严格地表达出来。

假设 $d > c[(1+k)^T - 1]$，并记 $\delta = d - c[(1+k)^T - 1] > 0$。那么

$$S_{T+m} = \frac{d - \delta(1+k)^m}{k}.$$

由于 $1+k > 1$，当 $m \to \infty$ 时 $(1+k)^m \to \infty$，所以 $\delta(1+k)^m$ 终会超过 d。这条公式只在债务为正数时成立，因此不可能永远成立：必定有第一天，余下的债务加上当天的利息不超过 d，公司就在那一天把债务还清。事实上，债务在第 $T+m$ 天还清，其中 m 是满足下式的最小值：

$$(1+k)^m \geq \frac{d}{\delta}, \qquad \text{即} \qquad m \geq \frac{\ln(d/\delta)}{\ln(1+k)}.$$

所以，若 $d/c > (1+k)^T - 1$，公司最终能还清债务，正是所求的结果。

这个条件也是必要的。若 $d \le kS_T$，则 $\delta \le 0$；又因为 $S_T = (d-\delta)/k$，

$$S_{T+m} - S_T = \frac{-\delta[(1+k)^m - 1]}{k} \ge 0$$

对所有 m 都成立。债务永远不会低于 S_T，所以永远无法还清。

验算

- 代入 $n=1$：$S_1 = c$。这是对的，因为第 1 天并无欠款，所以无须支付利息。
- 在 S_{T+m} 的公式中代入 $m=0$：得出 S_T，与预期相符。
- 若刚好有 $d = kS_T$，则 $\delta = 0$，而且对所有 m 都有 $S_{T+m} = d/k = S_T$：收入仅仅足够支付利息，债务始终不变。这是临界情况。
- 取 $c=100$、$k=0.01$、$T=30$ 逐日模拟，得出 $S_T \approx 3478.49$，与公式吻合。门槛是 $d > 34.78$。取 $d=50$ 时，债务在 $m=120$ 天后还清，与 $\ln(50/15.22)\,/\,\ln 1.01 \approx 119.6$ 相符。

常见错误

- 在第 1 天就计算利息，或者把级数数成 $n+1$ 项。由于 $S_0 = 0$，第一天的成本刚好是 c。
- 忘记第 T 天之后级数是由 S_T 而不是由 0 开始：$(1+k)^m S_T$ 这一项很容易漏掉。
- 说债务能还清是因为“d 大于 c”。真正要比较的，是 d 与通车时的利息 kS_T。

STEP 之外

递推关系 $S_n = aS_{n-1} + b$ 是最简单的*一阶线性差分方程*（linear first-order difference equation）。它的解是不动点（fixed point）$b/(1-a)$ 加上 a^n 的某个倍数，这正是我们那条公式的形式：d/k 加上 $(1+k)^m$ 的某个倍数。这个倍数是正还是负，决定了债务是不断增长还是得以还清；同样的想法会在不动点的稳定性（stability）和线性常微分方程（linear ODE）中再次出现。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）把差分方程与微分方程并列讨论，而 Siklos 的 *Advanced Problems in Mathematics*（STEP 进阶题集）则有更多关于递推关系的 STEP 练习。

详细解答

STEP III 1988，第 6 题

纯数学 · 逐次微分

> **STEP III 1988，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/88-s3-q6

关键思路

直接把一个乘积求导 1988 次是行不通的，但正弦与余弦的乘积可以改写成一个和。这样每一项 $\sin kx$ 的导数（derivative）便以四步为一个循环，每求导一次就多出一个因子 k。至于零点（zero），$\sin 3x$ 前面那个庞大的因子 3^{1988} 意味着导数只可能在 $\sin 3x$ 几乎为零的地方等于零，所以我们只需看 $\sin 3x$ 第一个零点稍后的位置，然后作线性化（linearise）。

解答

第 1988 阶导数

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

$$\mathrm{f}(x) = \sin 2x\cos x = \tfrac{1}{2}\left(\sin 3x + \sin x\right).$$

把 $\sin kx$ 求导四次，得

$$\sin kx \;\to\; k\cos kx \;\to\; -k^2\sin kx \;\to\; -k^3\cos kx \;\to\; k^4\sin kx,$$

所以每求导四次，效果只是把 $\sin kx$ 乘以 k^4。（等价地说，$\sin kx$ 的第 n 阶导数是 $k^n\sin\left(kx+\frac{1}{2}n\pi\right)$。）由于 $1988 = 4\times 497$，

$$\mathrm{f}^{(1988)}(x) = \tfrac{1}{2}\left(3^{1988}\sin 3x + \sin x\right).$$

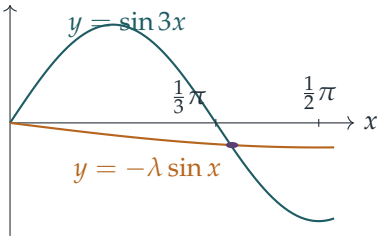
最小的正零点

我们要找出满足下式的最小 $x>0$：

$$3^{1988}\sin 3x + \sin x = 0. \qquad (*)$$

在 $(0, \frac{1}{3}\pi]$ 内没有零点。当 $0 < x < \frac{1}{3}\pi$ 时，$0 < 3x < \pi$，所以 $\sin 3x$ 和 $\sin x$ 都是正数，$(*)$ 的左边为正。在 $x = \frac{1}{3}\pi$ 处，左边等于 $0 + \frac{\sqrt{3}}{2} > 0$。

紧接 $\frac{1}{3}\pi$ 之后有一个零点。在 $x = \frac{1}{2}\pi$ 处，左边是 $3^{1988}\sin\frac{3}{2}\pi + 1 = 1 - 3^{1988} < 0$。左边是连续（continuous）的，所以由介值定理（intermediate value theorem），在 $(\frac{1}{3}\pi, \frac{1}{2}\pi)$ 内必有零点，而最小的正零点就是其中的第一个。把它写成 $x = \frac{1}{3}\pi + \epsilon$，其中 $0 < \epsilon < \frac{1}{6}\pi$。



草图把 $(*)$ 表示为 $\sin 3x = -\lambda \sin x$，其中 $\lambda = 3^{-1988}$（图中画得远比实际大）。两条曲线首次相交之处，就在 $\sin 3x$ 于 $\frac{1}{3}\pi$ 穿过零之后不远；对于真正的、极其微小的 λ，交点几乎被推到 $\frac{1}{3}\pi$ 上。

由于 $\sin(\pi + 3\epsilon) = -\sin 3\epsilon$，方程 $(*)$ 变成

$$3^{1988}\sin 3\epsilon = \sin\left(\tfrac{1}{3}\pi + \epsilon\right). \qquad (**)$$

想一想 我们怎样知道 ϵ 小得足以使用 $\sin 3\epsilon \approx 3\epsilon$?

$(**)$ 的右边最多是 1，所以 $\sin 3\epsilon \le 3^{-1988}$。由于 $0 < 3\epsilon < \frac{1}{2}\pi$，而在 $[0, \frac{1}{2}\pi]$ 上有 $\sin t \ge \frac{2t}{\pi}$（正弦曲线在该区间内位于其弦（chord）的上方），我们得到 $3\epsilon \le \frac{\pi}{2}3^{-1988}$。所以 ϵ 小得难以想像，这就说明了以下的近似是合理的。

ϵ 既然这样小，便有 $\sin 3\epsilon = 3\epsilon\,(1 + O(\epsilon^2))$ 和 $\sin(\frac{1}{3}\pi + \epsilon) = \frac{\sqrt{3}}{2} + O(\epsilon)$。代入 $(**)$，

$$3^{1989}\epsilon \approx \frac{\sqrt{3}}{2}, \qquad \text{所以} \qquad \epsilon \approx \frac{3^{-1989}\sqrt{3}}{2} = \frac{1}{3}\cdot\frac{3^{-1988}\sqrt{3}}{2},$$

其相对误差（relative error）的阶为 3^{-1988}。

关于试卷上印出的数值。 试卷要求证明 $\epsilon \approx \frac{1}{2}3^{-1988}\sqrt{3}$，这是此处所得数值的三倍。因子 3 来自 $\sin 3\epsilon \approx 3\epsilon$，而下面的数值验算也确认答案是 3^{-1989}，而不是 3^{-1988}。所以题目看来有一处笔误：它原意要用的方法正是上面的方法，只是 3 的幂次差了一。如果你得到的是 3^{-1989}，请相信自己的计算。

验算

- 同样的论证适用于任何 4 的倍数 n：$3^n \sin 3x + \sin x$ 的第一个正零点位于 $\frac{1}{3}\pi + \epsilon_n$，其中 $\epsilon_n \approx \frac{1}{2}3^{-(n+1)}\sqrt{3}$。当 $n = 4$ 时，数值求解得 $\epsilon_4 = 3.5713 \times 10^{-3}$，而预测值是 $\frac{1}{2}\sqrt{3}/3^5 = 3.5640 \times 10^{-3}$（比值为 1.0021）。当 $n = 8$ 时，比值是 1.000025；到了 $n = 40$ 及以上，比值在机器精度内等于 1。它与 $\frac{1}{2}\sqrt{3}/3^n$ 的比值则趋向 $\frac{1}{3}$。
- 导数公式可以用最初几个情况来检验：$\mathrm{f}^{(4)}(x) = \frac{1}{2}(81\sin 3x + \sin x)$，你可以把 $\frac{1}{2}(\sin 3x + \sin x)$ 求导四次来确认。

常见错误

- 弄错正负号或相位：只有在求导次数是四的倍数之后，$\sin kx$ 才会回到 $+k^n \sin kx$。
- 忘记排除 $(0, \frac{1}{3}\pi)$ 内的零点；正是那里的正负号论证，才使 $\frac{1}{3}\pi + \epsilon$ 成为最小的正零点。
- 写成 $\sin(\pi + 3\epsilon) \approx \epsilon$ 而不是 -3ϵ，漏掉了正负号或因子 3。

STEP 之外

第二部分是*摄动理论*（perturbation theory）的一个简单例子：一个很大的参数迫使解非常贴近某个已知的根，而在该根附近作线性化便得出首项修正，误差的大小则由泰勒级数（Taylor series）的下一项控制。利用 $O(\cdot)$ 记号和误差界，把“近似等于”说得精确，是大学第一门分析课程的核心主题；参见 Abbott 的 *Understanding Analysis*。

详细解答

STEP III 1988，第 7 题

纯数学 · 微分方程与正交性

STEP III 1988，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/88-s3-q7

本题在 *Beyond STEP* 第 8.6 节量子振子中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/quantum-oscillator.html

关键思路

第一部分纯粹是代入（substitution）：把试探函数（trial function）求导两次，再比较 x 各次幂的系数，便可确定 λ。第二部分是经典的“朗斯基技巧”（Wronskian trick）：组合 $y_m y_n' - y_n y_m'$ 的导数中，麻烦的 $\omega^2 x^2$ 项会互相抵消，只剩下 $y_m y_n$ 的一个倍数。之后把这个导数在整条实数线上积分就很容易，因为在 $\pm\infty$ 处的条件令边界项（boundary terms）全部消失。

解答

全篇以撇号表示 $\mathrm{d}/\mathrm{d}x$。把方程写成以下形式会比较方便：

$$y_n'' = \left(\omega^2 x^2 - (2n+1)\omega\right) y_n. \tag{$*$}$$

$n = 0$ 的情况

取 $y_0 = \mathrm{e}^{-\lambda x^2}$，则

$$y_0' = -2\lambda x\,\mathrm{e}^{-\lambda x^2}, \qquad y_0'' = (4\lambda^2 x^2 - 2\lambda)\,\mathrm{e}^{-\lambda x^2}.$$

代入 $n = 0$ 时的 $(*)$，并约去非零因子 $\mathrm{e}^{-\lambda x^2}$，我们需要

$$4\lambda^2 x^2 - 2\lambda = \omega^2 x^2 - \omega \quad \text{对所有} x.$$

比较常数项得 $\lambda = \frac{1}{2}\omega$，这时 x^2 项亦相符，因为 $4\lambda^2 = \omega^2$。（$4\lambda^2 = \omega^2$ 的另一个根 $\lambda = -\frac{1}{2}\omega$ 不满足常数项的要求。）

$n = 1$ 的情况

取 $y_1 = x\,\mathrm{e}^{-\lambda x^2}$，则

$$\begin{aligned} y_1' &= (1 - 2\lambda x^2)\,\mathrm{e}^{-\lambda x^2}, \\ y_1'' &= \big(-4\lambda x - 2\lambda x(1 - 2\lambda x^2)\big)\mathrm{e}^{-\lambda x^2} = (4\lambda^2 x^3 - 6\lambda x)\,\mathrm{e}^{-\lambda x^2}. \end{aligned}$$

$n = 1$ 时的方程 $(*)$ 要求 $y_1'' = (\omega^2 x^2 - 3\omega)\,x\,\mathrm{e}^{-\lambda x^2}$，即

$$4\lambda^2 x^3 - 6\lambda x = \omega^2 x^3 - 3\omega x \quad \text{对所有} x,$$

而这同样恰好在 $\lambda = \frac{1}{2}\omega$ 时成立。因此同一个常数对两个函数都适用：

$$\lambda = \tfrac{1}{2}\omega, \qquad y_0 = \mathrm{e}^{-\omega x^2/2}, \qquad y_1 = x\,\mathrm{e}^{-\omega x^2/2}.$$

无穷远处的条件

由于 $\lambda = \frac{1}{2}\omega > 0$，凡是形如（多项式）$\times \mathrm{e}^{-\lambda x^2}$ 的函数，当 $x \to \pm\infty$ 时都趋向 0，因为 $\mathrm{e}^{-\lambda x^2}$ 衰减的速度比 x 的任何次幂增长的速度都要快。函数 y_0、y_0'、y_1 和 y_1' 全部属于这种形式（对应的多项式分别是 1、$-2\lambda x$、x 和 $1 - 2\lambda x^2$），所以这四个函数在两端都趋向 0，符合要求。

恒等式

根据积法则（product rule），$y_m' y_n'$ 项互相抵消：

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(y_m y_n' - y_n y_m'\big) = y_m' y_n' + y_m y_n'' - y_n' y_m' - y_n y_m'' = y_m y_n'' - y_n y_m''.$$

现在对 y_n'' 和 y_m'' 都使用 $(*)$：

$$\begin{aligned} y_m y_n'' - y_n y_m'' &= \big(\omega^2 x^2 - (2n+1)\omega\big) y_m y_n - \big(\omega^2 x^2 - (2m+1)\omega\big) y_m y_n \\ &= \big((2m+1) - (2n+1)\big)\omega\, y_m y_n \\ &= 2(m-n)\,\omega\, y_m y_n, \end{aligned}$$

这就是所要求的结果。

正交性

> ? **想一想** 在尝试证明这个无穷积分等于零之前，我们怎样知道它确实存在？

我们先在有限区间上处理，再让两个端点分别趋向无穷；这样存在性便自然而然地得到。记 $W = y_m y_n' - y_n y_m'$。对任意 $S, R > 0$，把恒等式由 $-S$ 积分至 R，得

$$2(m-n)\,\omega \int_{-S}^{R} y_m y_n \,\mathrm{d}x = W(R) - W(-S).$$

当 $R \to \infty$ 时，根据题目所给的条件，y_m, y_n, y_m', y_n' 各自趋向 0，所以 $W(R) \to 0$。同理，当 $S \to \infty$ 时 $W(-S) \to 0$。因此右边趋向 0，故积分 $\int_{-\infty}^{\infty} y_m y_n \,\mathrm{d}x$ 收敛（converges），并且

$$2(m-n)\,\omega \int_{-\infty}^{\infty} y_m y_n \,\mathrm{d}x = 0.$$

若 $m \neq n$，则 $2(m-n)\omega \neq 0$，因为 $\omega > 0$。两边除以它，得

$$\int_{-\infty}^{\infty} y_m(x)\, y_n(x)\,\mathrm{d}x = 0,$$

正如所求。请留意，这里只用到方程本身和两端的条件：我们从来不需要 y_m 或 y_n 的具体公式。

验算

- 当 $m = 0$、$n = 1$ 时，结果可以直接看出：$y_0 y_1 = x\,\mathrm{e}^{-\omega x^2}$ 是奇函数（odd function），所以它在整条实数线上的积分是 0。
- 一个没那么显然的检验：$y_2 = (2\omega x^2 - 1)\mathrm{e}^{-\omega x^2/2}$ 满足 $n = 2$ 时的方程（不妨自己试试）。利用 $\int_{-\infty}^{\infty} \mathrm{e}^{-\omega x^2}\,\mathrm{d}x = \sqrt{\pi/\omega}$ 和 $\int_{-\infty}^{\infty} x^2 \mathrm{e}^{-\omega x^2}\,\mathrm{d}x = \frac{1}{2\omega}\sqrt{\pi/\omega}$，可得 $\int y_0 y_2\,\mathrm{d}x = 2\omega \cdot \frac{1}{2\omega}\sqrt{\pi/\omega} - \sqrt{\pi/\omega} = 0$。
- 数值方面，取 $\lambda = \frac{1}{2}\omega$ 时，方程在随机点上的残差（residual）对 $n = 0, 1, 2$ 都小于 10^{-6}；以有限差分（finite differences）检验，导数恒等式准确至 10^{-4}；而在 $[-30, 30]$ 上使用辛普森法则（Simpson's rule），对 $\omega = 0.5, 1, 2.5$，$\int y_0 y_1$、$\int y_0 y_2$、$\int y_1 y_2$ 全部小于 10^{-15}。

常见错误

- 只比较 x^2 的系数而得出 $\lambda = \pm\frac{1}{2}\omega$：常数项排除了负号（无穷远处的衰减条件亦然）。
- 忘记说明边界项为何消失：这正是同时用到 y_n 和 y_n' 两个条件的地方。
- 除以 $m - n$ 时，没有指出它不等于零，亦没有指出 $\omega > 0$。

STEP 之外

这就是量子谐振子（quantum harmonic oscillator）：y_n 是它的能量态，$(2n+1)\omega$ 是它的能级（相差一个比例常数），而这些函数是埃尔米特多项式（Hermite polynomials）乘以一个高斯函数（Gaussian）。这个正交性（orthogonality）论证是施图姆－刘维尔理论（Sturm–Liouville theory）的一个特例；该理论证明，自伴（self-adjoint）微分算子对应于不同特征值（eigenvalue）的特征函数（eigenfunction）必定正交，正如对称矩阵的特征向量（eigenvector）一样。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（微分方程入门教材）涵盖边值问题和施图姆－刘维尔理论。

详细解答

STEP III 1988，第 8 题

纯数学 · 抛物线的切线与法线

> **STEP III 1988，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/88-s3-q8

关键思路

把每一条直线都用参数（parameter）t 表示，求出各对直线的交点，再把这些顶点代入行列式（determinant）。这些行列式看来很繁复，但它们对 t_1, t_2, t_3 是反对称（antisymmetric）的，所以每一个都必定是 $(t_1 - t_2)(t_2 - t_3)(t_3 - t_1)$ 的倍数。只要用简单的行运算和列运算（row and column operations），就可以把这个因子提出来，余下的就是两个面积之间的比较。

（印刷的题目要求的是 A_3；这是 A_2 的误印，即法线三角形的面积。）

解答

切线与法线

沿着 $x = at^2, y = 2at$ 把 $y^2 = 4ax$ 对 t 求导，得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{2a}{2at} = \dfrac{1}{t}$。在 $(at^2, 2at)$ 的切线（tangent）是 $y - 2at = \frac{1}{t}(x - at^2)$，即

$$ty = x + at^2.$$

这个形式在 $t = 0$ 时也成立（在顶点的切线是 $x = 0$）。法线（normal）的斜率（gradient）是 $-t$：

$$y + tx = 2at + at^3.$$

我们全程假设 $a \neq 0$。

切线三角形

把在 t_i 和 t_j 的切线方程相减，得 $(t_i - t_j)y = a(t_i^2 - t_j^2)$。由于 $t_i \neq t_j$，我们可以相除，所以 $y = a(t_i + t_j)$，于是 $x = t_i y - a t_i^2 = a t_i t_j$。因此

$$\text{在} t_i, t_j \text{ 的两条切线相交于} \big(a t_i t_j,\ a(t_i + t_j)\big).$$

各切线的斜率 $1/t_i$ 互不相同（或者其中一条是铅垂的），所以没有两条是平行的，三角形确实存在。记 $s = t_1 + t_2 + t_3$。与 P_1 相对的顶点是 $(a t_2 t_3,\ a(s - t_1))$，其余两个顶点类似。从首两列各提出因子 a，然后把第 3 列的 $-s$ 倍加到第 2 列，再把第 2 列变号，

$$\det \begin{pmatrix} a t_2 t_3 & a(s - t_1) & 1 \\ a t_1 t_3 & a(s - t_2) & 1 \\ a t_1 t_2 & a(s - t_3) & 1 \end{pmatrix} = -a^2 \det \begin{pmatrix} t_2 t_3 & t_1 & 1 \\ t_1 t_3 & t_2 & 1 \\ t_1 t_2 & t_3 & 1 \end{pmatrix} = a^2 D,$$

当中我们对调了首两列，并设

$$D = \det \begin{pmatrix} t_1 & t_2 t_3 & 1 \\ t_2 & t_1 t_3 & 1 \\ t_3 & t_1 t_2 & 1 \end{pmatrix}.$$

要计算 D，把第 1 行换成第 1 行 − 第 2 行，把第 2 行换成第 2 行 − 第 3 行。第 1 行变成 $(t_1 - t_2,\ t_3(t_2 - t_1),\ 0) = (t_1 - t_2)(1,\ -t_3,\ 0)$，而第 2 行变成 $(t_2 - t_3)(1,\ -t_1,\ 0)$。沿最后一列展开，

$$D = (t_1 - t_2)(t_2 - t_3) \det \begin{pmatrix} 1 & -t_3 \\ 1 & -t_1 \end{pmatrix} = (t_1 - t_2)(t_2 - t_3)(t_3 - t_1).$$

因此

$$A_1 = \tfrac{1}{2} a^2 \left| (t_1 - t_2)(t_2 - t_3)(t_3 - t_1) \right|,$$

由于各 t_i 互不相同，这个面积不等于零。

法线三角形

把在 t_i 和 t_j 的法线方程相减，得 $(t_i - t_j)x = 2a(t_i - t_j) + a(t_i^3 - t_j^3)$，所以除以 $t_i - t_j \neq 0$ 后，

$$x = 2a + a(t_i^2 + t_i t_j + t_j^2), \qquad y = 2a t_i + a t_i^3 - t_i x = -a t_i t_j (t_i + t_j).$$

各法线的斜率 $-t_i$ 互不相同，所以同样地每一对法线都相交。对于与 P_1 相对的顶点，利用 $t_2 + t_3 = s - t_1$，并记 $e_1 = t_2 t_3$（以及 $e_2 = t_1 t_3, e_3 = t_1 t_2$）。那么

$$t_2^2 + t_2 t_3 + t_3^2 = (t_2 + t_3)^2 - t_2 t_3 = (s - t_1)^2 - e_1, \qquad -t_2 t_3 (t_2 + t_3) = -s e_1 + p,$$

其中 $p = t_1t_2t_3$。所以这个顶点是 $a\big(2 + (s - t_1)^2 - e_1,\ -se_1 + p\big)$，其余两个顶点类似。现在提出 a^2，并进行不改变行列式的值的列运算：从第 1 列减去第 3 列的 2 倍，从第 2 列减去第 3 列的 p 倍，然后从第 2 列提出因子 $-s$：

$$\begin{aligned}\frac{1}{a^2}\det(\text{法线三角形的顶点}) &= \det\Big((s-t_i)^2 - e_i \quad -se_i \quad 1\Big)_{i=1,2,3}\\ &= -s\det\Big((s-t_i)^2 - e_i \quad e_i \quad 1\Big)_{i=1,2,3},\end{aligned}$$

当中每个式子只显示有代表性的第 i 行。接着把第 2 列加到第 1 列，这样就消去了 $-e_i$，再展开 $(s - t_i)^2 = s^2 - 2st_i + t_i^2$。当中的 s^2 是第 3 列的倍数，可以消去，而行列式对第 1 列是线性的：

$$\det\Big(t_i^2 - 2st_i \quad e_i \quad 1\Big) = \det\Big(t_i^2 \quad e_i \quad 1\Big) - 2s\det\Big(t_i \quad e_i \quad 1\Big) = E - 2sD,$$

这里我们如此记这两个行列式。第二个行列式正好就是 D。

想一想 我们怎样计算 E，即各行为 $(t_i^2,\ e_i,\ 1)$ 的行列式?

使用与计算 D 时相同的行运算。第 1 行 − 第 2 行是 $(t_1^2 - t_2^2,\ t_3(t_2 - t_1),\ 0) = (t_1 - t_2)(t_1 + t_2,\ -t_3,\ 0)$，而第 2 行 − 第 3 行是 $(t_2 - t_3)(t_2 + t_3,\ -t_1,\ 0)$。所以

$$\begin{aligned}E &= (t_1 - t_2)(t_2 - t_3)\big[-t_1(t_1 + t_2) + t_3(t_2 + t_3)\big]\\ &= (t_1 - t_2)(t_2 - t_3)\big[(t_3^2 - t_1^2) + t_2(t_3 - t_1)\big]\\ &= (t_1 - t_2)(t_2 - t_3)(t_3 - t_1)(t_1 + t_2 + t_3) = sD.\end{aligned}$$

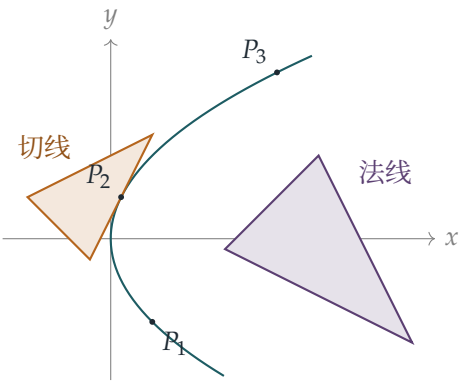
把以上各步合起来，

$$\det(\text{法线三角形的顶点}) = a^2 \cdot (-s)\,(sD - 2sD) = a^2s^2D.$$

取绝对值的一半，并与 $A_1 = \frac{1}{2}a^2|D|$ 比较，

$$A_2 = (t_1 + t_2 + t_3)^2A_1,$$

这就是所求的结果。



图中取 $a = 1$ 及 $t_1, t_2, t_3 = -1, \frac{1}{2}, 2$，所以 $s = \frac{3}{2}$。切线三角形（琥珀色）的面积是 $\frac{27}{8}$，而法线三角形（紫红色）的面积是 $\frac{243}{32} = \left(\frac{3}{2}\right)^2 \cdot \frac{27}{8}$。

法线何时共点？

没有两条法线是平行的，所以它们两两相交于三点 Q_{12}, Q_{13}, Q_{23}。如果三条法线共点（concurrent），这三点就会重合，三角形的面积为零。反过来，假设面积为零但三点并非全部相同，例如 $Q_{12} \neq Q_{13}$。这两点都在 P_1 的法线上，所以通过它们的直线就是该法线。由于三点共线（collinear），Q_{23} 也在 P_1 的法线上。于是 Q_{23} 同时在三条法线上，所以 $Q_{23} = Q_{12}$（在 P_1 和 P_2 的两条法线的唯一公共点），同理 $Q_{23} = Q_{13}$，这就产生矛盾。所以

$$\text{三条法线共点} \iff A_2 = 0 \iff (t_1 + t_2 + t_3)^2 A_1 = 0 \iff t_1 + t_2 + t_3 = 0,$$

这里用到 $A_1 \neq 0$，因为各 t_i 互不相同。所求的条件是 $t_1 + t_2 + t_3 = 0$。

验算

- 取 $t_1, t_2, t_3 = 0, 1, 2$ 及 $a = 1$，切线三角形的顶点是 $(2,3)$、$(0,2)$、$(0,1)$，面积为 1，而 $\frac{1}{2}|D| = \frac{1}{2}|(-1)(-1)(2)| = 1$。
- 在 $t = 1, 2, -3$（总和为零，$a = 1$）的法线全部通过 $(9, -6)$。例如，在 $t = 1$ 的法线是 $y + x = 3$。
- 我们用电脑以分数作精确计算，对大约 2,800 组随机选取的 a 和互不相同的 t_i 求出各顶点，每一次都确认了 $A_2 = (t_1 + t_2 + t_3)^2 A_1$。

常见错误

- 忘记说明三条直线确实构成一个三角形。我们需要的是两两不平行，而这里可由斜率互不相同得出。
- 未加理由就从 $A_2 = 0$ 断定法线共点。面积为零只说明三个交点共线，上面那段简短的论证才把它提升为“相同”。
- 把 3×3 行列式完全展开。这样做是可行的，但几乎肯定会出现正负号错误。应先用行运算提出因子 $(t_i - t_j)$。

STEP 之外

因子 $(t_1 - t_2)(t_2 - t_3)(t_3 - t_1)$ 是一个范德蒙行列式（Vandermonde determinant），而 $E = sD$ 中的余因子 $t_1 + t_2 + t_3$ 是一个对称多项式（symmetric polynomial）。任何反对称多项式都是范德蒙乘积乘以一个对称多项式，这个事实在行列式和对称函数的理论中会再次出现。法线三角形与抛物线的渐屈线（evolute）有关，渐屈线是抛物线的曲率中心（centres of curvature）所描出的曲线，也就是各法线的包络线（envelope）。Strang 的 *Introduction to Linear Algebra*（线性代数入门）对行列式及其行运算有很好的讲解。

详细解答

STEP III 1988，第 9 题

纯数学 · 群元素的阶

STEP III 1988，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/88-s3-q9

关键思路

第 (i) 部分其实是带余除法（division algorithm）的变相应用：把 N 写成 n 的倍数加上余数，而 n 的最小性迫使余数为零。第 (ii) 部分之所以成立，是因为以 g 作共轭（conjugation）会保持乘积，所以在一串共轭元素的乘积中，夹在中间的每一对 $g^{-1}g$ 都会互相抵消。在第 (iii) 部分，我们反复以 g 作共轭；由于 $g^5 = e$，做五次之后 h 便回到自身，从而确定 h 的阶（order）。

解答

第 (i) 部分

根据带余除法，可写 $N = qn + r$，其中 q 为整数，且 $0 \le r < n$。（这对任何整数 N 都成立，包括零和负整数。）于是，由于 $g^n = e$，

$$e = g^N = (g^n)^q g^r = e^q g^r = g^r.$$

假如 r 是正数，它便是一个小于 n 而又满足 $g^r = e$ 的正整数，这与 $n = o(g)$ 是符合此条件的最小正整数这个定义矛盾。因此 $r = 0$，即 $N = qn$，故 n 整除 N。

第 (ii) 部分

$m = 0$：两边都等于 e。

$m \ge 1$：我们对 m 使用归纳法（induction）。$m = 1$ 的情况显然成立。若 $gh^m g^{-1} =$

$(ghg^{-1})^m$，则

$$(ghg^{-1})^{m+1} = (ghg^{-1})^m\,(ghg^{-1}) = (gh^mg^{-1})(ghg^{-1})$$
$$= gh^m(g^{-1}g)hg^{-1} = gh^{m+1}g^{-1},$$

归纳法至此完成。（把乘积 $(ghg^{-1})(ghg^{-1})\cdots(ghg^{-1})$ 全部写出来，便会逐项相消：中间的每一个 $g^{-1}g$ 都抵消掉。）

$m \le -1$：首先，ghg^{-1} 的逆元（inverse）是 $gh^{-1}g^{-1}$，因为

$$(ghg^{-1})(gh^{-1}g^{-1}) = gh(g^{-1}g)h^{-1}g^{-1} = ghh^{-1}g^{-1} = e.$$

现在写 $m = -k$，其中 $k \ge 1$。以元素 h^{-1} 代替 h，套用正数情况的结果，得

$$gh^mg^{-1} = g(h^{-1})^kg^{-1} = (gh^{-1}g^{-1})^k = \left((ghg^{-1})^{-1}\right)^k = (ghg^{-1})^{-k} = (ghg^{-1})^m.$$

所以这个恒等式对每一个整数 m 都成立。

第 (iii) 部分

先用 $ghg^{-1} = h^2$，再用第 (ii) 部分并取 $m = 2$，

$$g^2hg^{-2} = g\,(ghg^{-1})\,g^{-1} = gh^2g^{-1} = (ghg^{-1})^2 = (h^2)^2 = h^4,$$

正如所求。

同样的步骤在每一个阶段都适用。我们断言：对所有 $k \ge 0$，$g^khg^{-k} = h^{2^k}$。$k = 0$ 时这显然成立；若它对 k 成立，则由第 (ii) 部分并取 $m = 2^k$，

$$g^{k+1}hg^{-(k+1)} = g\,(g^khg^{-k})\,g^{-1} = gh^{2^k}g^{-1} = (ghg^{-1})^{2^k} = (h^2)^{2^k} = h^{2^{k+1}}.$$

取 $k = 5$，并利用 $g^5 = e$（因此 $g^{-5} = e$ 亦成立），得

$$h = g^5hg^{-5} = h^{32}, \qquad \text{所以} \qquad h^{31} = e$$

（两边乘以 h^{-1} 即得）。

> **想一想** 我们知道 $h^{31} = e$。为甚么由此可知 $o(h)$ 恰好是 31，而不仅仅是不超过 31?

由第 (i) 部分，$o(h)$ 整除 31。由于 31 是质数（prime），$o(h)$ 只能是 1 或 31。但 $o(h) = 1$ 即表示 $h = e$，而题目已说明并非如此。因此

$$o(h) = 31.$$

这样的群（group）确实存在，所以题目的条件并无矛盾。考虑 $\{0, 1, \ldots, 30\}$ 上的两个置换（permutation）：$h(x) = x + 1$ 和 $g(x) = 2x$，运算皆以 31 为模。那么 $g^5(x) = 32x = x$，而且 $ghg^{-1}(x) = 2\left(\frac{x}{2} + 1\right) = x + 2 = h^2(x)$，同时 h 的阶确实是 31。

验算

- 出现质数 31 是合理的：以 g 作共轭，对 h 的各次幂的作用就是“把指数乘以 2”，而加倍五次后必须回到起点，所以 $2^5 \equiv 1$（模 $o(h)$）。换句话说，$o(h)$ 整除 $2^5 - 1 = 31$。
- 我们以置换 $g(x) = 2x$ 和 $h(x) = x + 1$（模 31）作电脑验算，确认了 $g^5 = e$、$ghg^{-1} = h^2$、$g^k h g^{-k} = h^{2^k}$（$k = 0, \ldots, 5$）以及 $o(h) = 31$。此外，亦在数百个随机置换上，就 $-12 \le m \le 12$ 检验了 (i) 和 (ii)。

常见错误

- 在 (ii) 中只证明 $m \ge 1$ 的情况。题目说的是*任何整数*，所以 $m = 0$ 和负数 m 都需要各自的（简短）论证。
- 直接由 $h^{31} = e$ 断定 $o(h) = 31$。这只能得出 $o(h) \mid 31$；还需要用到 31 是质数以及 $h \neq e$ 这个条件。

STEP 之外

映射 $x \mapsto gxg^{-1}$ 是 G 的一个*自同构*（automorphism），即保持结构的双射（bijection），而第 (ii) 部分正是这个事实的一个推论。第 (iii) 部分是*半直积*（semidirect product）的一个小例子：由 g 和 h 生成的群的阶是 $5 \times 31 = 155$，而它之所以存在，全赖 2 在模 31 之下的阶是 5。这种构造在小阶群的分类中占有核心地位。若想进一步了解阶、共轭和拉格朗日定理（Lagrange's theorem），Allenby 的 *Rings, Fields and Groups*（环、域与群）是很好的读物。

详细解答

STEP III 1988，第 10 题

纯数学 · 追踪曲线

> **STEP III 1988，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/88-s3-q10

本题在 *Beyond STEP* 第 8.5 节格力犬追逐格力犬中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/pursuit.html

关键思路

不要尝试分别追踪四只移动中的狗。整个设置绕中心旋转四分之一圈后维持不变，所以四只狗始终位于一个以 O 为中心的正方形的四个顶点上：这个正方形会缩小、会旋转，但始终是正方形。这就确定了 A 的速度与直线 AO 之间的夹角，而速度与半径之间的夹角固定，正正就是产生等角螺线（equiangular spiral）$r = \lambda \mathrm{e}^{-k\theta}$ 的条件。

解答

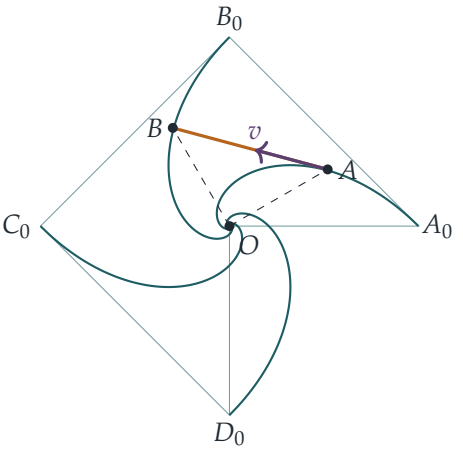
四只狗始终位于正方形上

取正方形的中心为原点 O，并按逆时针方向为四只狗编号，使得 B 位于 A 逆时针旋转四分之一圈之处，C 位于 B 逆时针旋转四分之一圈之处，如此类推。

把整幅图绕 O 旋转 $\frac{\pi}{2}$。A, B, C, D 的起始位置分别变成 B, C, D, A 的起始位置，而“以速率 v 直奔下一只狗”这条规则对每只狗都是一样的。所以旋转后的运动，同样是由相同的起始布局出发、遵从相同规则的运动。由给定起点出发的运动是唯一的（只要四只狗尚未相遇，速度就是位置的光滑函数），所以旋转后的运动就是实际的运动。换言之，在每一瞬间，B 的位置就是 A 的位置绕 O 旋转 $\frac{\pi}{2}$ 所得的位置。C 和 D 亦然。所以 $ABCD$ 始终是一个以 O 为中心的正方形。

速度与半径之间的夹角

设 A 在极坐标（polar coordinates）中位于 (r,θ)。那么 B 位于 $(r,\theta+\frac{\pi}{2})$，所以 $OA=OB=r$，而 $\angle AOB=\frac{\pi}{2}$。三角形 OAB 是等腰直角三角形，所以 $\angle OAB=\frac{\pi}{4}$。



图中显示四条路径由起始顶点 A_0,B_0,C_0,D_0 向内盘旋，并标出某一瞬间 A 和 B 两只狗的位置。A 的速度（紫红色）沿 AB 方向。

A 的速度大小为 v，方向沿 AB，与向内的半径 AO 成 $\frac{\pi}{4}$ 角。把它分解为径向分量（radial component，以向外为正）和横向分量（transverse component，沿 θ 增加的方向）：

$$\dot r=-v\cos\frac{\pi}{4},\qquad r\dot\theta=v\sin\frac{\pi}{4}.$$

横向分量为正，是因为 B 位于 A 的逆时针方向，所以 A 被拉着逆时针转动。

> ? **想一想** 为甚么可以把这两条方程相除?

由于四只狗尚未相遇时 $r\dot\theta=v/\sqrt{2}>0$，θ 随时间严格递增。所以我们可以用 θ 代替 t 作为路径上的参数，再由链式法则（chain rule）得

$$\frac{\mathrm{d}r}{\mathrm{d}\theta}=\frac{\dot r}{\dot\theta}=\frac{-v/\sqrt{2}}{v/(\sqrt{2}r)}=-r.$$

求出路径

由于 $r>0$，我们可以分离变量（separate variables）：$\int\frac{\mathrm{d}r}{r}=-\int\mathrm{d}\theta$，所以 $\ln r=-\theta+c$，即

$$r=\lambda\mathrm{e}^{-\theta}$$

其中 $\lambda=\mathrm{e}^c$ 是常数。这就是所求的结果。如果我们选取 OA_0 为始线，使得开始时 $\theta=0$，那么 λ 就是 A 与 O 的初始距离，对于边长为 a 的正方形，即 $a/\sqrt{2}$。若选取其他始线，λ

只会相差一个常数因子。（如果我们按顺时针方向为四只狗编号，则以顺时针方向量度 θ 便会得到相同的方程。）

推广至 n 只狗

对于位于正 n 边形各顶点上的 n 只狗（按逆时针方向编号，每只追逐下一只），把旋转角改为 $\frac{2\pi}{n}$，同样的论证依然成立。这些狗始终构成一个以 O 为中心的正 n 边形。所以若 A 位于 (r,θ)，下一只狗 B 就位于 $(r,\theta+\frac{2\pi}{n})$。

三角形 OAB 是等腰三角形，顶角为 $\frac{2\pi}{n}$，所以每个底角为

$$\angle OAB = \frac{1}{2}\left(\pi - \frac{2\pi}{n}\right) = \frac{\pi}{2} - \frac{\pi}{n}.$$

如前分解 A 的速度，

$$\dot r = -v\cos\left(\frac{\pi}{2}-\frac{\pi}{n}\right) = -v\sin\frac{\pi}{n}, \qquad r\dot\theta = v\sin\left(\frac{\pi}{2}-\frac{\pi}{n}\right) = v\cos\frac{\pi}{n}.$$

由于 $n \geq 3$，$\cos\frac{\pi}{n} > 0$，所以 θ 同样稳步递增，并且

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = -r\tan\frac{\pi}{n} \qquad \Longrightarrow \qquad r = \lambda\,\mathrm{e}^{-\theta\tan(\pi/n)}.$$

当 $n=4$ 时，$\tan\frac{\pi}{4}=1$，我们重新得到 $r=\lambda\mathrm{e}^{-\theta}$。每只狗都沿着一条等角螺线（又称对数螺线，logarithmic spiral）前进，以固定的角 $\frac{\pi}{2}-\frac{\pi}{n}$ 与每条半径相交。

一个有趣的副产品：$\dot r = -v\sin\frac{\pi}{n}$ 是常数，所以这些狗在时间 $r_0/(v\sin\frac{\pi}{n})$ 之后于 O 相遇，其中 r_0 是与 O 的初始距离。在这段时间内，每只狗跑过的距离是有限的 $r_0/\sin\frac{\pi}{n}$，尽管 $\theta\to\infty$，牠绕着 O 转了无限多圈。（对于正方形，这个距离是 a，即边长。）

验算

- 当 $n=4$ 时，结果就是一般公式在 $\tan\frac{\pi}{4}=1$ 时的特殊情况。
- 当 $n\to\infty$ 时，$\tan\frac{\pi}{n}\to 0$，路径趋向一个圆：每只狗所追逐的邻居，在圆上几乎与牠并排。
- 我们以数值方法模拟了这场追逐（用 RK4 方法，取 $n=3,4,5,6,8$ 只狗，由半径 1 出发）。算出的 r 与 $\mathrm{e}^{-\theta\tan(\pi/n)}$ 吻合，相对误差为 10^{-11}。当 $n=4$ 时，到达 $r=0.02$ 所需的时间和路径长度均算得 $1.386 = 0.98\sqrt{2}$，与 $(1-0.02)/\sin\frac{\pi}{4}$ 所预测的一致。

常见错误

- 未先论证这些狗始终位于正多边形上，就直接写下 $\dot r$ 和 $r\dot\theta$。正是靠对称性论证，我们才知道 A 处的角。

- 用了多边形的内角（正方形为 $\frac{\pi}{2}$），而不是 AB 与半径 AO 之间的夹角（$\frac{\pi}{4}$）。
- 忘记了横向速度分量是 $r\dot{\theta}$，而不是 $\dot{\theta}$。

STEP 之外

这是一道经典的追逐问题（pursuit problem）。对称性那一步，是微分方程（differential equations）中一个普遍原理的例子：如果方程和初始数据在某个对称变换下都保持不变，那么解的唯一性就迫使解本身也保持不变。与每条半径以固定角相交的曲线，就是等角螺线。以复数形式表示，它们是 $z = \mathrm{e}^{(-k+\mathrm{i})\theta}$，所以整个运动是平面上一个连续的“旋转相似变换”（spiral similarity）。Needham 的 *Visual Complex Analysis*（以几何观点讲解复分析）阐述了这种几何观点，而 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）则涵盖了解的存在性与唯一性。

详细解答

STEP III 1988，第 11 题

力学 · 涉及摩擦力的梯子平衡

STEP III 1988，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/88-s3-q11

关键思路

先弄清几何关系：地面、斜坡和梯子构成一个三角形，它的各个角你其实已经知道，由此便可判断光滑斜坡对梯顶的推力朝哪个方向。之后就是处理梯子问题的标准步骤：分解（resolve）两次，对梯脚取力矩（moments），再把所需的摩擦力（friction）与 μ 乘以法向反作用力（normal reaction）作比较。由于工人的位置只出现在力矩方程之中，一条通用公式便可同时回答两部分。

解答

准备工作

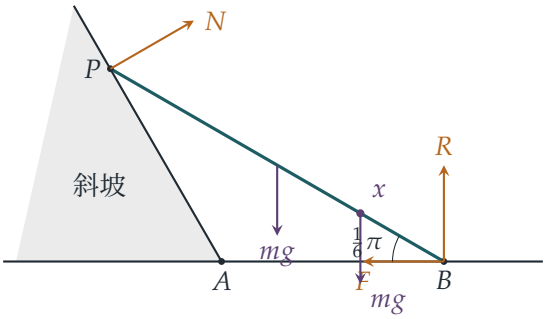
把梯子视为均匀杆（uniform rod），把工人视为质点（particle）。设 A 为斜坡的底端，B 为梯脚，P 为梯顶，梯顶位于斜坡上。

几何关系。在三角形 ABP 中，B 处的角为 $\frac{1}{6}\pi$（即梯子的倾角）。斜坡与铅垂线成 $\frac{1}{6}\pi$，所以它在 A 的另一侧与水平线成 $\frac{1}{3}\pi$。因此三角形内 A 处的角为 $\pi - \frac{1}{3}\pi = \frac{2}{3}\pi$，余下 $\frac{1}{6}\pi$ 便是 P 处的角。所以这是一个等腰三角形（isosceles triangle），$AP = AB$，而由正弦定理（sine rule）可得

$$AB = \frac{l \sin \frac{1}{6}\pi}{\sin \frac{2}{3}\pi} = \frac{l}{\sqrt{3}},$$

与题目给出的距离吻合。题目的数据互相一致，P 的确在斜坡上。

斜坡反作用力的方向。斜坡是光滑（smooth）的，所以它对梯子的反作用力 N 垂直于斜坡。斜坡与水平线成 $\frac{1}{3}\pi$，故其法线与水平线成 $\frac{1}{6}\pi$，指向斜上方并离开斜坡，朝着 B 的一方。因此 N 的水平分量为 $\frac{\sqrt{3}}{2}N$，把梯子推离斜坡；铅垂分量为 $\frac{1}{2}N$，方向向上。



方程。设工人沿梯子向上距离梯脚 x。在 B 处，地面施加向上的法向反作用力 R，以及沿水平方向指向斜坡的摩擦力 F（除此之外唯一的水平力是 N 的分量，它把梯脚推离斜坡，所以摩擦力必须与之相反）。这里的距离是由 B 起计。

沿水平和铅垂方向分解：

$$F = \tfrac{\sqrt{3}}{2}N, \qquad\qquad R + \tfrac{1}{2}N = 2mg.$$

对 B 取力矩，这样 R 和 F 便不会出现。由 B 到梯子上距离梯脚 s 的一点，水平距离为 $s\cos\frac{1}{6}\pi = \frac{\sqrt{3}}{2}s$。$N$ 的作用线与水平线成 $\frac{1}{6}\pi$，而 BP 则在水平线的另一侧与之成 $\frac{1}{6}\pi$，所以 N 与梯子的夹角为 $\frac{1}{3}\pi$。它对 B 的力矩是 $Nl\sin\frac{1}{3}\pi = \frac{\sqrt{3}}{2}Nl$。因此

$$\tfrac{\sqrt{3}}{2}Nl = mg\cdot\tfrac{\sqrt{3}}{2}\cdot\tfrac{l}{2} + mg\cdot\tfrac{\sqrt{3}}{2}x \quad\Longrightarrow\quad N = mg\left(\frac{1}{2}+\frac{x}{l}\right).$$

想一想 为甚么 N 的力矩会这样简单？

由 B 到 N 的作用线的垂直距离是 $l\sin\frac{1}{3}\pi$，因为 N 与梯子相交的角度是 $\frac{1}{3}\pi$。又或者把 N 分解成分量：水平部分 $\frac{\sqrt{3}}{2}N$ 作用于高度 $\frac{1}{2}l$ 处，铅垂部分 $\frac{1}{2}N$ 作用于水平距离 $\frac{\sqrt{3}}{2}l$ 处。两者的力矩都是 $\frac{\sqrt{3}}{4}Nl$，加起来就是 $\frac{\sqrt{3}}{2}Nl$。

于是当 $R = 2mg - \frac{1}{2}N = mg\left(\frac{7}{4} - \frac{x}{2l}\right) > 0$ 对 $0 \le x \le l$ 成立，梯脚便一直留在地面上。只要 $F \le \mu R$，梯子就不会滑动，即

$$\mu \ge \frac{F}{R} = \frac{\sqrt{3}\,N}{4mg - N} = \frac{\sqrt{3}\,(l+2x)}{7l-2x}.$$

这个比值随 x 增加而增加：分子增加，而恒为正数的分母减少。所以工人愈往上爬，所需的摩擦力便愈大。

第 (i) 部分

工人一到达梯子的中点，梯子便滑动。所以极限摩擦（limiting friction）恰好在 $x = \frac{1}{2}l$ 时出现，而由于 F/R 随 x 递增，在此之前不会出现。这时 $N = mg$、$F = \frac{\sqrt{3}}{2}mg$ 及 $R = \frac{3}{2}mg$，

所以

$$\mu = \frac{F}{R} = \frac{\sqrt{3}/2}{3/2} = \frac{1}{\sqrt{3}}.$$

第 (ii) 部分

若 $F \le \mu R$ 对每一个 $x \in [0, l]$ 都成立，工人便能到达梯顶。由于 F/R 随 x 递增，最坏的情况是 $x = l$。这时 $N = \frac{3}{2}mg$、$F = \frac{3\sqrt{3}}{4}mg$ 及 $R = \frac{5}{4}mg$，所以摩擦系数（coefficient of friction）的最小值为

$$\mu = \frac{3\sqrt{3}/4}{5/4} = \frac{3\sqrt{3}}{5} \approx 1.04.$$

验算

- 梯上无人时（$x = 0$），公式给出 $\mu \ge \sqrt{3}/7 \approx 0.247$。这比两个答案都小，正合预期。
- 把 $l = 1$、$m = 1$、$g = 9.8$ 代入三条平衡方程作数值计算，并把各个力写成向量，可重现 $N/mg = 1$ 和 $\frac{3}{2}$，以及比值 $F/R = 0.57735\ldots$ 和 $1.03923\ldots$。计算亦确认，在梯子上的 101 个位置，F/R 都是递增的。

常见错误

- 把斜坡的反作用力当作水平方向，就像处理铅垂墙壁那样。这里它向水平线上方倾斜了 $\frac{1}{6}\pi$，而它向上的分量会令 R 减小。
- 沿铅垂方向分解时忘记了工人的重量，结果写成 $R + \frac{1}{2}N = mg$ 而不是 $2mg$。

STEP 之外

第 (ii) 部分其实是对各种配置所作的一个小型最优化（optimisation）：你需要找出负载在所有位置之中 F/R 的最坏情况。工程师正是这样找出结构的临界负载情况。要想像条件 $F \le \mu R$，一个有用的方法是**摩擦锥**（cone of friction）：地面的总反作用力必须落在与法线成 $\arctan\mu$ 的角度之内，而当所需的反作用力离开这个锥的一刻，梯子便会滑动。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对刚体静力学和摩擦有仔细的论述。

详细解答

STEP III 1988，第 12 题

力学 · 斜碰撞

STEP III 1988，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/88-s3-q12

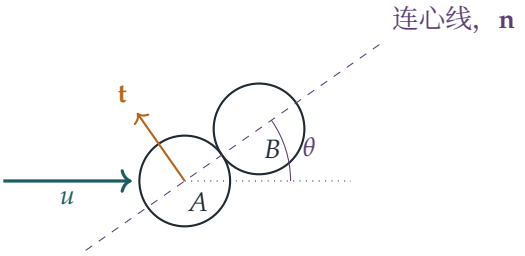
关键思路

对光滑球体而言，碰撞瞬间的冲量（impulse）只沿连心线（line of centres）作用。因此我们把每个速度分解为沿连心线的分量和垂直于连心线的分量：垂直分量完全不受影响，而沿连心线方向则是一个普通的一维碰撞。于是一切都取决于一个未知量，即连心线与第一个球原来速度之间的夹角。

解答

建立模型

把两球视为质量同为 m 的光滑球体，在光滑桌面上滑动，因此碰撞期间唯一的水平力就是沿连心线的接触冲量。设第一个球 A 以速率 u 运动，并设碰撞时连心线与 A 的速度成角 θ，其中 $0 \le \theta < \frac{1}{2}\pi$（若 $\theta = \frac{1}{2}\pi$，两球只是擦过，甚么也不会发生）。取单位向量（unit vector）$\mathbf{n}$ 沿连心线（由 A 指向 B），$\mathbf{t}$ 与之垂直。



碰撞前，A 沿 $\mathbf{n}$ 的分量为 $u\cos\theta$，沿 $\mathbf{t}$ 的分量为 $u\sin\theta$；B 静止。

沿 $\mathbf{t}$ 方向：两球在这个方向上都没有受到冲量，所以 A 保持 $u\sin\theta$，B 保持 0。

沿 $\mathbf{n}$ 方向：设碰撞后的分量为 v_A 和 v_B。由动量守恒（conservation of momentum）（质量互相约去）和牛顿恢复定律（Newton's law of restitution）得

$$v_A + v_B = u\cos\theta, \qquad v_B - v_A = e\,u\cos\theta,$$

所以

$$v_A = \tfrac{1}{2}(1-e)\,u\cos\theta, \qquad v_B = \tfrac{1}{2}(1+e)\,u\cos\theta.$$

因此碰撞后

$$\mathbf{v}_A = \tfrac{1}{2}(1-e)u\cos\theta\,\mathbf{n} + u\sin\theta\,\mathbf{t}, \qquad \mathbf{v}_B = \tfrac{1}{2}(1+e)u\cos\theta\,\mathbf{n}.$$

两个速度之间的夹角

B 沿连心线运动，而两个速度之间的夹角为锐角，当且仅当它们的纯量积（scalar product）为正。这里

$$\mathbf{v}_A \cdot \mathbf{v}_B = \tfrac{1}{4}(1-e)(1+e)\,u^2\cos^2\theta > 0,$$

因为 $e < 1$、$u \neq 0$，而且当 $0 \le \theta < \frac{1}{2}\pi$ 时 $\cos\theta > 0$。故此两个速度之间的夹角小于 $\frac{1}{2}\pi$，这正是所要证明的。（对于正面碰撞，即 $\theta = 0$，两球沿同一直线运动，夹角为 0。）

留意 $e < 1$ 在哪里用到：若 $e = 1$，则 $v_A = 0$，两球恰好成直角分开，这就是大家熟悉的桌球结果。

最大偏转角

设 ϕ 为 A 的偏转角（angle of deflection），即 $\mathbf{v}_A$ 与原来方向 $\cos\theta\,\mathbf{n} + \sin\theta\,\mathbf{t}$ 之间的夹角。记 $k = \frac{1}{2}(1-e)$，则 $0 < k \le \frac{1}{2}$，且 $\mathbf{v}_A = u(k\cos\theta\,\mathbf{n} + \sin\theta\,\mathbf{t})$。对方向向量 $(\cos\theta, \sin\theta)$ 和 $(k\cos\theta, \sin\theta)$ 使用纯量积和二维叉积（cross product），得

$$|\mathbf{v}_A|\cos\phi = u\,(k\cos^2\theta + \sin^2\theta),$$
$$|\mathbf{v}_A|\sin\phi = u\,(\cos\theta\sin\theta - k\cos\theta\sin\theta) = u\,(1-k)\sin\theta\cos\theta.$$

第一式为正，所以 ϕ 是锐角；两式相除（再把分子和分母同除以 $\cos^2\theta$），并记 $t = \tan\theta \ge 0$，得

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

想一想 怎样才能不经繁琐的求导而求出这个式子的最大值?

由于 ϕ 是锐角，求 ϕ 的最大值等同于求 $\tan\phi$ 的最大值。当 $t > 0$ 时，

$$\tan\phi = \frac{1-k}{k/t+t} \quad \text{而} \quad \frac{k}{t} + t \ge 2\sqrt{k}$$

这是由算术平均—几何平均不等式（AM–GM inequality）得出的，等号成立当且仅当 $t = \sqrt{k}$。（当 $t = 0$ 时，$\phi = 0$。）所以最大值在 $\tan\theta = \sqrt{k}$ 时取得，且

$$\tan\phi_{\max} = \frac{1-k}{2\sqrt{k}}.$$

要转换成正弦，留意 $(1-k)^2 + (2\sqrt{k})^2 = 1 - 2k + k^2 + 4k = (1+k)^2$，所以在两条直角边为 $1-k$ 和 $2\sqrt{k}$ 的直角三角形中，斜边为 $1+k$，于是

$$\sin\phi_{\max} = \frac{1-k}{1+k} = \frac{\frac{1}{2}(1+e)}{\frac{1}{2}(3-e)} = \frac{1+e}{3-e},$$

当中用到 $1-k = \frac{1}{2}(1+e)$ 和 $1+k = \frac{1}{2}(3-e)$。由于 $\phi_{\max}$ 是锐角，

$$\phi_{\max} = \sin^{-1}\left(\frac{1+e}{3-e}\right),$$

这就是所求的结果。它出现于连心线与原来运动方向成角 $\tan^{-1}\sqrt{\frac{1}{2}(1-e)}$ 之时。

验算

- $e \to 1$：公式给出 $\sin^{-1}1 = \frac{1}{2}\pi$，与完全弹性（perfectly elastic）的情况一致：此时 A 失去沿连心线的全部速度，所以它总是沿垂直于连心线的方向运动，而一次接近正面的撞击会使它以接近直角的方向偏离原来的路径。
- $e = 0$：$\sin\phi_{\max} = \frac{1}{3}$，所以 $\phi_{\max} \approx 0.3398$ rad（19.5°）。
- $e = \frac{1}{2}$：$\sin\phi_{\max} = \frac{3}{5}$，$\phi_{\max} \approx 0.6435$ rad，出现于 $\theta = \tan^{-1}\frac{1}{2} \approx 0.4636$。以数值方法对 200 000 个 θ 值逐一计算偏转角，结果与这两个数字吻合；而 200 000 次随机碰撞中，两个速度之间的夹角从未达到或超过 $\frac{1}{2}\pi$。

常见错误

- 把恢复定律套用于整个速率，而不是沿连心线的分量。
- 忘记 A 的速度的垂直分量保持不变；正是这一点才令 A 发生偏转。
- 对反正切函数求导来直接求 ϕ 的最大值；改用 $\tan\phi$（因为 ϕ 是锐角，这样做是合理的）会简洁得多。

STEP 之外

这里的分解方法，其实就是牛顿第二定律（Newton's second law）的冲量形式：冲量只会在其作用的方向上改变动量。在弹性碰撞的情况下，直角这个结果是一个巧妙的参考

系（frame of reference）论证的特例：在质心参考系（centre-of-mass frame）中，弹性碰撞保持每个球的速率不变，只改变其方向，而变换回实验室参考系（lab frame）后，便一次过得出所有可能的结果。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学导论）正是这样在质心参考系中处理碰撞的。

详细解答

STEP III 1988，第 13 题

力学 · 有空气阻力的抛体

STEP III 1988，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/88-s3-q13

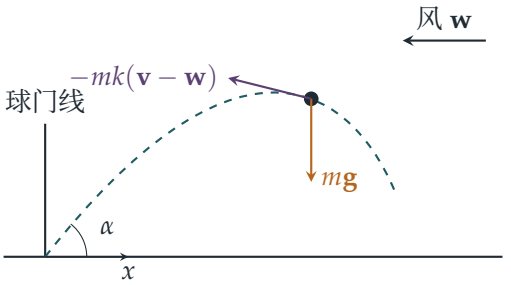
关键思路

空气阻力与速度成正比时，运动方程（equation of motion）是线性的，因此每个分量都可以用积分因子（integrating factor）求解。要看清最后一部分，最简洁的方法是问：终端速度（terminal velocity）是甚么？速度会由初始值按指数规律趋向终端速度，所以若这两个向量平行，速度便永远不会离开通过原点的同一条直线。于是皮球沿着同一条直线路径飞出去，再飞回来。

解答

运动方程

我们把皮球看成质量为 m 的质点（particle）。作用于它的力有两个：重量 $m\mathbf{g}$，以及空气阻力。皮球相对于空气的速度是 $\mathbf{v}-\mathbf{w}$，而阻力与这个相对速度方向相反，所以阻力是 $-mk(\mathbf{v}-\mathbf{w})$。



由牛顿第二定律（Newton's second law）得

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\mathbf{g} - mk(\mathbf{v} - \mathbf{w}).$$

两边除以 m 并移项，

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + k\mathbf{v} = \mathbf{g} + k\mathbf{w},$$

这正是所要证明的结果。

时间 t 时的位置

取开球点为原点，x 轴为水平方向并指向开球的方向，y 轴垂直向上。记 $V = |\mathbf{v}_0|$、$W = |\mathbf{w}|$ 及 $g = |\mathbf{g}|$。皮球是迎着风踢出的，所以

$$\mathbf{w} = (-W, 0), \qquad \mathbf{g} = (0, -g), \qquad \mathbf{v}_0 = (V\cos\alpha, V\sin\alpha).$$

设 $\mathbf{v} = (u, v)$，运动方程的两个分量便是

$$\frac{\mathrm{d}u}{\mathrm{d}t} + ku = -kW, \qquad \frac{\mathrm{d}v}{\mathrm{d}t} + kv = -g.$$

把第一式乘以积分因子 e^{kt}，得 $\frac{\mathrm{d}}{\mathrm{d}t}\left(u\,\mathrm{e}^{kt}\right) = -kW\mathrm{e}^{kt}$，所以 $u\,\mathrm{e}^{kt} = -W\mathrm{e}^{kt} + C$。由于 $t = 0$ 时 $u = V\cos\alpha$，故 $C = V\cos\alpha + W$，于是

$$u = -W + (V\cos\alpha + W)\,\mathrm{e}^{-kt}.$$

用完全相同的方法，

$$v = -\frac{g}{k} + \left(V\sin\alpha + \frac{g}{k}\right)\mathrm{e}^{-kt}.$$

再由 0 至 t 积分一次，并利用 $t = 0$ 时 $x = y = 0$ 以及 $\int_0^t \mathrm{e}^{-ks}\,\mathrm{d}s = \frac{1}{k}\left(1 - \mathrm{e}^{-kt}\right)$，得

$$x = -Wt + \frac{V\cos\alpha + W}{k}\left(1 - \mathrm{e}^{-kt}\right),$$

$$y = -\frac{g}{k}\,t + \frac{1}{k}\left(V\sin\alpha + \frac{g}{k}\right)\left(1 - \mathrm{e}^{-kt}\right).$$

写成向量形式，设 $\mathbf{u}_\infty = \mathbf{w} + \mathbf{g}/k$（即终端速度，亦即运动方程的常数解），上式便是

$$\mathbf{r} = \mathbf{u}_\infty t + \frac{1}{k}\left(\mathbf{v}_0 - \mathbf{u}_\infty\right)\left(1 - \mathrm{e}^{-kt}\right).$$

乌龙球

想一想 $\tan\alpha = g/(kW)$ 这个条件，说明了终端速度的甚么性质？

终端速度是 $\mathbf{u}_\infty = (-W, -g/k) = -W\left(1, g/(kW)\right)$。条件给出 $g/(kW) = \tan\alpha$，所以 $\mathbf{u}_\infty = -W(1, \tan\alpha)$，它与 $\mathbf{v}_0$ 平行（方向相反）。由向量公式可见，$\mathbf{r}$ 此时始终是两个沿 $(\cos\alpha, \sin\alpha)$ 方向的向量的组合，所以皮球沿直线 $y = x\tan\alpha$ 运动。

若要直接从分量看出这一点，可利用 $g/k = W\tan\alpha$：

$$V\sin\alpha + \frac{g}{k} = V\sin\alpha + W\tan\alpha = \tan\alpha\,(V\cos\alpha + W),$$

所以

$$y = \tan\alpha\left[-Wt + \frac{V\cos\alpha + W}{k}\left(1 - \mathrm{e}^{-kt}\right)\right] = x\tan\alpha \quad \text{对所有} t \geq 0.$$

因此，皮球位于地面（即 $y = 0$）当且仅当 $x = 0$（因为 $\tan\alpha \neq 0$）。余下要检查的，是这情况确实在某个 $t > 0$ 发生。设 $f(t) = x(t)$。则 $f(0) = 0$ 而 $f'(0) = V\cos\alpha > 0$，所以刚踢出之后 $x > 0$。此外

$$f''(t) = -k(V\cos\alpha + W)\,\mathrm{e}^{-kt} < 0,$$

所以 f 是严格凹函数（strictly concave），而且由于有 $-Wt$ 一项，当 $t \to \infty$ 时 $f(t) \to -\infty$。因此 f 恰好有一个正零点 T，并且在该处 $f'(T) < 0$（一个由 0 升起的凹函数，在跌回 0 时必定正在递减）。

在时间 T，皮球回到原点，即球门线上开球的那一点，而且水平速度为负，也就是说正朝着球门移动。它沿着飞出去时的那条直线飞回来，并在地面高度越过球门线，远低于横楣。既然守门员不在位，这便是一个乌龙球。

验算

- 当 $k \to 0$ 时，$\frac{1}{k}(1 - \mathrm{e}^{-kt}) \to t$ 而 $\frac{1}{k}\left(t - \frac{1}{k}(1 - \mathrm{e}^{-kt})\right) \to \frac{1}{2}t^2$，所以 $x \to Vt\cos\alpha$ 而 $y \to Vt\sin\alpha - \frac{1}{2}gt^2$：这正是普通的抛体运动（projectile motion）。
- 当 $t \to \infty$ 时，速度趋向 $(-W, -g/k)$，即运动方程的常数解：皮球最终随风飘移，同时以终端速率下坠。
- 以龙格—库塔法（Runge–Kutta）对运动方程作数值积分，在随机选取多组 V, α, W, k 的情况下，结果与 x 和 y 的公式相符，误差约为 10^{-13}。当 $\tan\alpha = g/(kW)$ 时，计算所得的路径一直留在 $y = x\tan\alpha$ 上，并在 $x = 0$ 处着地，而且 $u < 0$；若取例如 $V = 20$、$\alpha = 0.6$、$W = 5$、$k = 0.5$，皮球则落在前场 17.1 m 处。

常见错误

- 把阻力取为 $-mk\mathbf{v}$（相对于地面），而不是 $-mk(\mathbf{v} - \mathbf{w})$。
- 弄错 $\mathbf{w}$ 的正负号：皮球是*迎着*风踢出的，所以 $\mathbf{w}$ 指向球门的方向。
- 得出 $y = x\tan\alpha$ 便停下来，没有检查皮球是否真的回到 $x = 0$，以及回到该处时是否正朝着球门移动。

STEP 之外

方程 $\dot{\mathbf{v}} + k\mathbf{v} = \mathbf{F}$（其中 $\mathbf{F}$ 为常数）是线性系统松弛至平衡状态（equilibrium）的最简单例子：偏差 $\mathbf{v} - \mathbf{F}/k$ 按 e^{-kt} 的方式衰减。线性阻力只对细小而缓慢的物体才符合实际；足球实际所受的阻力大致与其速率的平方成正比，这令方程变成非线性，并使水平与垂直方向的运动互相耦合。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对这两种阻力都有讨论。

详细解答

STEP III 1988，第 14 题

力学 · 光滑金属线上的运动

STEP III 1988，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：

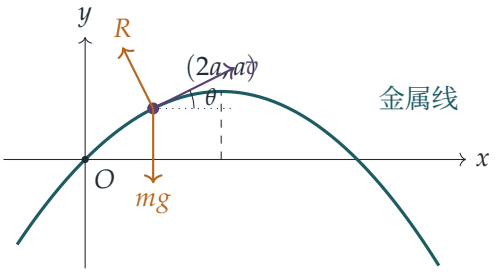
step.maths.org/questions/88-s3-q14

关键思路

如果金属线对小珠完全不施力，小珠就只是一个抛体（projectile），仅在重力作用下运动。所以第 (i) 部分问的是：一个沿金属线方向抛出的自由抛体，能否刚好沿着金属线的抛物线（parabola）运动？把两条抛物线对照起来即可。至于第 (ii) 部分，金属线是光滑的，所以能量守恒（conservation of energy）。真正要做的功夫，是利用 $\tan\theta = \mathrm{d}y/\mathrm{d}x$，把速率和高度用斜角 θ 表示出来。

解答

我们把小珠视为一个质量为 m 的质点（particle），套在一条固定的光滑金属线上。作用在它身上的力只有两个：垂直向下的重量 mg，以及金属线的法向反作用力（normal reaction）R，后者与金属线垂直。



以下会反复用到关于金属线的两个事实。在一般的点上，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 1 - \frac{x}{2a},$$

所以在原点处斜率为 1，小珠出发时的方向与 x-轴成 45° 仰角。由于金属线是一条开口向下的抛物线，它的最高点位于 $x = 2a, y = a$。

第 (i) 部分

假设在整个运动过程中 $R = 0$。那么重力就是唯一的力，小珠的运动便与一个以初速率 V、仰角 45° 抛出的抛体完全相同。它的水平速度保持不变，恒为 $V/\sqrt{2}$，所以 $x = Vt/\sqrt{2}$。同时

$$y = \frac{V}{\sqrt{2}}\,t - \tfrac{1}{2}gt^2.$$

以 $t = \sqrt{2}\,x/V$ 消去时间，得到轨迹（trajectory）

$$y = x - \frac{gx^2}{V^2}.$$

只有当这条曲线在小珠所到达的每一个 x 处都与金属线 $y = x - x^2/(4a)$ 重合，它才是小珠的路径。这就需要

$$\frac{g}{V^2} = \frac{1}{4a}, \qquad \text{即} \qquad V = 2\sqrt{ag}.$$

反过来，取 $V = 2\sqrt{ag}$。一个从 O 以这个速率、仰角 45° 抛出的自由质点，会沿着曲线 $y = x - x^2/(4a)$ 运动，而这正是金属线本身。因此它沿着金属线运动，每一刻都有 $R = 0$，而这个运动在反作用力为零的情况下满足小珠的运动方程（equation of motion）。由于 $x = Vt/\sqrt{2}$ 稳定地增加，小珠会越过位于 $(2a, a)$ 的顶点，并沿另一边一直向下运动，永不停止。在整个过程中，它对金属线不施加任何压力，符合题目要求，而所求的速率是 $V = 2\sqrt{ag}$。

> **想一想** 套在金属线上的小珠，怎可能在完全没有反作用力的情况下运动？金属线不是要把它留在线上吗？

只有当单靠重力会令小珠偏离曲线时，金属线才需要施力。换了任何其他速率，情况正是如此。V 较大时，抛体的抛物线较平坦，所以金属线必须把小珠向下拉。V 较小时，抛物线弯曲得较厉害，所以金属线必须把小珠托住。而刚好在 $V = 2\sqrt{ag}$ 时，重力已经提供了曲线所需的加速度，金属线便无事可做。

第 (ii) 部分

现在 V 是任意的。速度沿着金属线的切线（tangent）方向，所以

$$\tan\theta = \frac{\mathrm{d}y}{\mathrm{d}x} = 1 - \frac{x}{2a}, \qquad \text{所以} \qquad x = 2a(1 - \tan\theta).$$

（如果小珠向后滑回，它的速度便沿切线指向相反方向，角度改变 π。这时 $\tan\theta$ 和 $\cos^6\theta$ 都不变，所以下文不受任何影响。）

*以 θ 表示速率。*把 $x = 2a(1 - \tan\theta)$ 对 t 求导，得

$$\dot{x} = -2a\sec^2\theta\,\dot{\theta}.$$

由于 $\dot{y} = \tan\theta\,\dot{x}$，速率 v 满足

$$v^2 = \dot{x}^2 + \dot{y}^2 = \dot{x}^2(1 + \tan^2\theta) = \dot{x}^2\sec^2\theta = \frac{4a^2\dot{\theta}^2}{\cos^6\theta}.$$

*以 θ 表示高度。*记 $T = \tan\theta$，于是 $x = 2a(1 - T)$，

$$\begin{aligned} y = x - \frac{x^2}{4a} &= 2a(1 - T) - a(1 - T)^2 \\ &= a(1 - T)\big(2 - (1 - T)\big) \\ &= a(1 - T)(1 + T) = a(1 - \tan^2\theta). \end{aligned}$$

*能量。*金属线是光滑的，而 R 与速度垂直，所以 R 不作功，机械能（mechanical energy）守恒。小珠从 O 以速率 V 出发，所以

$$\tfrac{1}{2}mv^2 + mgy = \tfrac{1}{2}mV^2, \qquad \text{即} \qquad v^2 + 2gy = V^2.$$

代入 v^2 和 y 的表达式，得

$$\frac{4a^2\dot{\theta}^2}{\cos^6\theta} + 2ga(1 - \tan^2\theta) = V^2,$$

这就是所要证明的结果。

验算

- 在 $t = 0$ 时，我们有 $\theta = 45^\circ$，所以 $\tan^2\theta = 1$，第二项消失。这时第一项就是 $v^2 = V^2$，与预期相符。
- 取第 (i) 部分的 $V^2 = 4ag$，则 $\dot{x} = V/\sqrt{2}$ 为常数，所以 $\dot{\theta} = -\dot{x}\cos^2\theta/(2a)$。第一项变成 $2ag\sec^2\theta = 2ag(1 + \tan^2\theta)$。加上 $2ga(1 - \tan^2\theta)$ 便得 $4ag = V^2$。两部分的结果互相吻合。
- 数值方面，我们用龙格—库塔法（Runge–Kutta method）沿金属线逐步计算小珠的运动方程。取 $a = 1$、$g = 9.8$ 和 $V = 2\sqrt{9.8}$，算出的法向反作用力（以每单位质量计）一直到 $x \approx 13$ 都保持在 10^{-15} 以下。把 V 增加 30% 后，它在起点处约为 4.8。对于随机选取的 V 值，包括一些令小珠向后滑回的情况，第 (ii) 部分的恒等式都成立，相对误差约为 10^{-13}。

常见错误

- 只证明了在抛出的一刻 $R = 0$。题目要求的是一个*自始至终*都没有反作用力的运动，这正是我们要把整条轨迹与金属线对照的原因。
- 对 $\tan\theta$ 求导时，忘记了链式法则（chain rule）所带来的因子 $\sec^2\theta$。这样第一项中便会出现 $\cos^2\theta$，而不是 $\cos^6\theta$。

STEP 之外

第 (ii) 部分是这个运动的一个首次积分（first integral）：能量守恒把一条二阶方程化为只含单一坐标（这里是 θ）的一阶方程。在大学一年级的课程中，这会发展成处理约束运动（constrained motion）的拉格朗日（Lagrangian）方法：你只需选取一个能描述约束的坐标，完全不必计算反作用力。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对约束运动和能量方法有细致的处理，很值得一读。

详细解答

STEP III 1988，第 15 题

概率与统计 · 期望值的递推关系

STEP III 1988，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/88-s3-q15

关键思路

对当天的第一本书取条件（conditioning）。这本书一旦上了架，当天余下的过程就与书架较短的全新一天完全一样，所以长度为 n 的书架的期望（expected）本数，与长度为 $n-1$ 和 $n-2$ 时的期望本数有关。由此得到一条线性差分方程（difference equation），解法与二阶递推关系（recurrence）一样：齐次（homogeneous）部分用辅助二次方程（auxiliary equation）处理，再为右边的常数 1 找一个特解（particular solution）。

解答

以下一律设 N_n 为书架长度为 n 的一天内放上书架的书本数目，因此 $M(n) = \mathrm{E}(N_n)$。各本书是厚是薄互相独立（independent），每本为厚书的概率是 p。题目说明书本数量总是多于书架所能容纳的，所以每天最终必定有某本书放不下。

第 (i) 部分

当 $n = 0$ 时，书架完全没有空位。第一本书不论厚薄都放不下，所以它和其后每一本书都要放到手推车上。因此 $N_0 = 0$ 恒成立，而 $M(0) = 0$。

第 (ii) 部分

当 $n = 1$ 时，考虑第一本书。

- 若它是薄书（概率为 q），它放得下，并刚好放满书架。下一本书至少需要一个单位，所以放不下，自此所有书都放到手推车上。故 $N_1 = 1$。

- 若它是厚书（概率为 p），它需要两个单位，放不下，故 $N_1 = 0$。

因此 $M(1) = q \cdot 1 + p \cdot 0 = q$。

第 (iii) 部分

设 $n \geq 2$, 同样对第一本书取条件。不论它是哪一种书都放得下, 因为它最多只需要 $2 \leq n$ 个单位。

- 若它是薄书（概率为 q），它上架后余下 $n-1$ 个单位。
- 若它是厚书（概率为 p），它上架后余下 $n-2$ 个单位。

> **想一想** 为甚么在第一种情况下，其后上架书本数目的期望值恰好是 $M(n-1)$?

第一本书之后，其余的书依次到来，它们是厚是薄与第一本书独立，概率亦相同，而且按同一规则摆放：一直上架，直到有一本放不下为止。第一本书唯一改变了的是剩余空间，现在是 $n-1$ 个单位。所以当天余下的过程，就是书架长度为 $n-1$ 的完整一天的翻版，其后上架书本数目的期望值是 $M(n-1)$。同理，若第一本是厚书，这个期望值就是 $M(n-2)$。（根据第 (i) 部分，当 $n-2=0$ 时这也成立。）

把第一本书也计算在内，由全期望定律（law of total expectation）得

$$\begin{aligned} M(n) &= q\big(1 + M(n-1)\big) + p\big(1 + M(n-2)\big) \\ &= (p+q) + qM(n-1) + pM(n-2), \end{aligned}$$

而由于 $p+q=1$,

$$M(n) - qM(n-1) - pM(n-2) = 1 \qquad (n \geq 2),$$

这正是所要证明的。

闭合公式

先考虑齐次方程 $M(n) - qM(n-1) - pM(n-2) = 0$。代入 $M(n) = \lambda^n$ 试解，得辅助方程

$$\lambda^2 - q\lambda - p = 0.$$

由于 $q = 1-p$，$\lambda = 1$ 是一个根（$1-q-p=0$），而两根之积为 $-p$，所以

$$\lambda^2 - q\lambda - p = (\lambda - 1)(\lambda + p).$$

因此对任意 A 和 B，$A(-p)^n$ 和常数 B 都满足齐次方程。

因为 1 是一个根，常数无法产生右边的 1，所以我们改试 Cn。代入得

$$\begin{aligned} Cn - qC(n-1) - pC(n-2) &= C\big[(1-q-p)n + q + 2p\big] \\ &= C(q+2p) = C(1+p). \end{aligned}$$

当 $C = \dfrac{1}{1+p}$ 时，上式等于 1；由于 $1+p > 0$，这个取值没有问题。于是由线性性质（linearity），不论 A 和 B 取何值，$M(n) = A(-p)^n + B + Cn$ 对所有 $n \geq 2$ 都满足 (iii) 中的方程。

现在代入初始值。由第 (i) 部分得

$$M(0) = A + B = 0, \qquad \text{所以} B = -A,$$

而由第 (ii) 部分得

$$M(1) = -Ap + B + C = -A(1+p) + \frac{1}{1+p} = 1 - p.$$

因此

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

所以

$$A = \frac{p^2}{(1+p)^2}, \qquad B = -\frac{p^2}{(1+p)^2}, \qquad C = \frac{1}{1+p}.$$

即

$$M(n) = \frac{n}{1+p} - \frac{p^2}{(1+p)^2}\big(1 - (-p)^n\big).$$

由于 (i)、(ii) 和 (iii) 确定了 $M(0)$、$M(1)$，继而逐一确定其后每一个值，所以这不仅是一个可能的解，更是唯一的解。

验算

- $p = 0$（全是薄书）：$M(n) = n$，理应如此。
- $p = 1$（全是厚书）：$M(n) = \frac{n}{2} - \frac{1}{4}\big(1 - (-1)^n\big)$，当 n 为偶数时是 $\frac{n}{2}$，当 n 为奇数时是 $\frac{n-1}{2}$，即 $\lfloor n/2 \rfloor$。正确。
- 取 $p = q = \frac{1}{2}$，直接由递推关系计算：$M(2) = 1 + \frac{1}{4} = \frac{5}{4}$，而公式给出 $\frac{2}{3} - \frac{1}{9} \cdot \frac{3}{4} = \frac{5}{4}$。
- 当 n 很大时，每本书平均占用 $1+p$ 个单位，所以我们预期大约有 $n/(1+p)$ 本书；公式与此吻合，另加一个有界的修正项，来自书架末端浪费掉的空间。

我们已把这个闭合公式与上架过程的精确枚举结果核对（$n \leq 12$，并取多个 p 值），亦以模拟作了核对；例如当 $p = \frac{1}{2}$ 及 $n = 5$ 时，两者都给出 $M(5) = \frac{103}{32} \approx 3.219$。

常见错误

- 以为厚书放不下之后，薄书仍然可以上架。题目说明其后所有书都放到手推车上，这正是 $M(1) = q$ 而不是更大的数值的原因。
- 以常数作为特解来试：因为 $\lambda = 1$ 是辅助方程的解，常数会被齐次部分吸收，所以需要用 Cn。

STEP 之外

对第一步取条件，并留意到余下的部分是同一问题的较小翻版，这种方法称为*首步分析*（first-step analysis）。它是在马可夫链（Markov chain）中求期望首达时间（hitting time）和吸收概率（absorption probability）的标准方法。我们的答案亦说明了*更新定理*（renewal theorem）：若把大小随机的物件装进长度 n 之内，能装下的数目的增长速度与 n 除以平均大小相若。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）既涵盖随机漫步（random walk）的差分方程，亦有马可夫链的入门介绍。

详细解答

STEP III 1988，第 16 题

概率与统计 · 超几何分布

> **STEP III 1988，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/88-s3-q16

关键思路

不放回抽取（drawing without replacement）令各次抽取互相依赖，逐次分析会十分繁琐。有两个想法可以绕过这个困难。第一，数算不计次序的选法，从而得到红球数目的分布。第二，利用对称性（symmetry）：假想把整个瓮以随机次序全部抽空，则每个球出现在任何指定位置的机会均等。把 X 写成指示变量（indicator）X_i 之和，便可求得期望值（mean）和方差（variance），完全不需要知道分布。我们只需要各对变量之间的协方差（covariance）。

解答

恰好有 k 个红球的概率

把各个球视为可区分的。抽出的 n 个球组成一个大小为 n 的集合，而每个这样的集合出现的机会均等，所以共有 $\binom{M}{n}$ 个等可能的结果。要组成一个恰好含 k 个红球的集合，须从 m 个红球中选 k 个，并从 $M-m$ 个绿球中选 $n-k$ 个。因此

$$\mathrm{P}(\text{恰好}k\text{ 个红球}) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}},$$

其中沿用惯常的约定：若 $b<0$ 或 $b>a$，则 $\binom{a}{b}=0$。（这已涵盖不可能发生的情况，例如 $n-k>M-m$。）这就是超几何（hypergeometric）分布。

第 (i) 部分

假想一直抽下去，直至瓮中再没有球为止。M 个（可区分的）球被抽出的次序是一个均匀随机的排列：全部 $M!$ 种次序出现的机会均等。抽了 n 次便停下来并不会改变首 n 个位置，所以我们可以用完整的排列来处理。

固定一个位置 i。某个指定的球位于位置 i 的排列共有 $(M-1)!$ 个，这个数目对每个球都相同。所以位置 i 上的球是 M 个球中任何一个的机会均等，而

$$\mathrm{P}(X_i = 1) = \frac{m\,(M-1)!}{M!} = \frac{m}{M},$$

正是所求。

> **想一想** 第二个球的颜色，不是明明取决于第一个球是甚么颜色吗?

在已知第一个球的情况下，确实如此。但我们要求的是无条件（unconditional）概率。当 $i = 2$ 时，由全概率定律（law of total probability）可得

$$\mathrm{P}(X_2 = 1) = \frac{m}{M}\cdot\frac{m-1}{M-1} + \frac{M-m}{M}\cdot\frac{m}{M-1} = \frac{m(M-1)}{M(M-1)} = \frac{m}{M}.$$

对称性论证则一次过为每个 i 完成了同样的计算。

第 (ii) 部分

固定两个位置 $i \neq j$（所以 $M \geq 2$）。数算这两个位置上都是红球的排列。位置 i 上的球有 m 个选择，然后位置 j 有 $m-1$ 个选择，其余的球则有 $(M-2)!$ 种排法。因此

$$\mathrm{P}(X_i = 1 \text{ 及 } X_j = 1) = \frac{m(m-1)(M-2)!}{M!} = \frac{m(m-1)}{M(M-1)},$$

正是所求。换句话说，位置 (i, j) 上的有序球对，是 $M(M-1)$ 个由相异的球组成的有序对中任何一个的机会均等。

X 的期望值和方差

X 数算首 n 个抽出的球当中红球的数目。根据期望值的线性性质（linearity of expectation）——这并不需要独立性（independence）——再加上 $\mathrm{E}(X_i) = \mathrm{P}(X_i = 1)$，

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{nm}{M}.$$

至于方差，可用 $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$。把和的平方展开，并利用 $X_i^2 = X_i$（因为 X_i 只取 0 或 1），

$$X^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j = X + \sum_{i\neq j} X_i X_j,$$

其中第二个和遍历满足 $i \neq j$ 的 $n(n-1)$ 个有序对 (i,j)。而 $X_iX_j = 1$ 当且仅当两个球都是红球，所以由第 (ii) 部分得 $\mathrm{E}(X_iX_j) = \dfrac{m(m-1)}{M(M-1)}$。因此，当 $M \geq 2$ 时，

$$\begin{aligned}\mathrm{Var}(X) &= \frac{nm}{M} + \frac{n(n-1)m(m-1)}{M(M-1)} - \frac{n^2m^2}{M^2} \\ &= \frac{nm}{M^2(M-1)}\Big[M(M-1) + (n-1)(m-1)M - nm(M-1)\Big].\end{aligned}$$

方括号内的式子展开后为

$$\begin{aligned}&M^2 - M + nmM - nM - mM + M - nmM + nm \\ &\quad = M^2 - nM - mM + nm = (M-m)(M-n).\end{aligned}$$

因此

$$\mathrm{Var}(X) = \frac{nm(M-m)(M-n)}{M^2(M-1)}.$$

（若 $M = 1$，则 $n \leq 1$，根本没有球对，而 X 是常数，所以 $\mathrm{Var}(X) = 0$。）

设 $p = m/M$ 会令结果更具启发性。这时

$$\mathrm{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p) \cdot \frac{M-n}{M-1}.$$

这正是二项分布（binomial distribution）的方差乘以一个修正因子，而当 $n \geq 2$ 时，这个因子小于 1。

验算

- $n = M$：我们把所有球都抽出，所以必定有 $X = m$。公式给出期望值 m 和方差 0，后者源于因子 $M - n$。
- $n = 1$：X 是单独一个伯努利（Bernoulli）变量，其 $p = m/M$，而方差化简为 $p(1-p)$。
- $M = 6$、$m = 2$、$n = 3$：公式给出 $\mathrm{Var}(X) = \frac{3\cdot2\cdot4\cdot3}{36\cdot5} = \frac{2}{5}$。把全部 6! 种次序逐一列举，结果完全吻合；$M \leq 6$ 的每一个情况亦然。

常见错误

- 把各个 X_i 当作互相独立，于是得出二项分布的方差 $np(1-p)$。其实它们是负相关（negatively correlated）的：早段抽到红球，会令其后抽到红球的机会降低。
- 忘记了当球对是有序的时候，二重和 $\sum_{i\neq j}$ 共有 $n(n-1)$ 项，而不是 $\binom{n}{2}$ 项。

STEP 之外

把一个计数写成指示变量之和，是概率论中最强大的技巧之一。即使分布本身根本无从写出，这个方法仍能给出期望值和方差。因子 $\frac{M-n}{M-1}$ 就是抽样调查中所用的有限总体修正（finite population correction）。它说明了为何一个访问 1000 人的民意调查，对一个国家所能提供的信息，与对一个无限总体所能提供的几乎一样多。Blitzstein 与 Hwang 所著的 *Introduction to Probability*（概率论入门教材）全书贯彻使用指示变量，并详细讨论超几何分布。

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

1988 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。