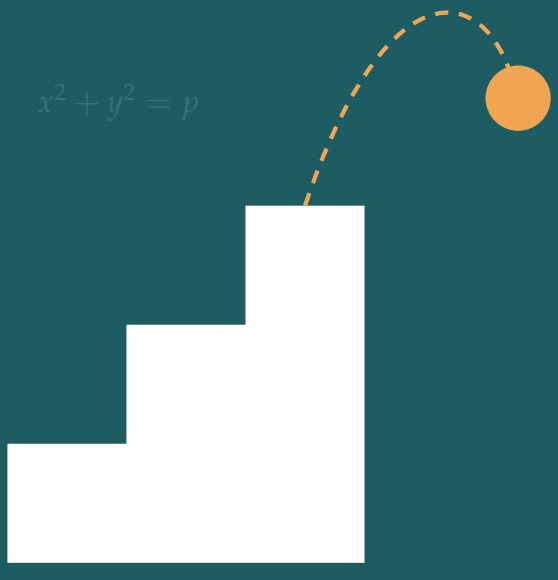
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2019 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

2019 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2019 年详细解答

2019 年 STEP 试卷（STEP I, II 及 III）全部 35 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 三角函数的最优化 纯数学 . . . 3

第 2 题 参数曲线的切线 纯数学 . . . 7

第 3 题 积分法 纯数学 . . . 11

第 4 题 四次方程的因式分解 纯数学 . . . 14

第 5 题 三维向量 纯数学 . . . 18

第 6 题 不等式与界限 纯数学 . . . 23

第 7 题 无理性与同余 纯数学 . . . 27

第 8 题 代换积分法 纯数学 . . . 31

第 9 题 倾倒与摩擦力 力学 . . . 34

第 10 题 抛体运动 力学 . . . 38

第 11 题 几何分布 概率与统计 . . . 41

STEP II

第 1 题 三次曲线的切线 纯数学 . . . 45

第 2 题 反函数的积分 纯数学 . . . 49

第 3 题 多项式根的界限 纯数学 . . . 52

第 4 题 三角函数裂项相消乘积 纯数学 . . . 56

第 5 题 迭代的不动点 纯数学 . . . 59

第 6 题 一阶微分方程 纯数学 . . . 63

第 7 题 向量与纯量积 纯数学 . . . 68

第 8 题 矩阵的积性函数 纯数学 . . . 71

第 9 题 抛体与原点的距离 力学 . . . 74

第 10 题 刚体平衡 力学 . . . 78

第 11 题 角形的计数 概率与统计 . . . 82

第 12 题 连续分布与分位数 概率与统计 86

STEP III

第 1 题 耦合微分方程 纯数学 90
第 2 题 函数方程 纯数学 94
第 3 题 矩阵的不变直线 纯数学 98
第 4 题 根与系数 纯数学 102
第 5 题 代换积分法 纯数学 106
第 6 题 阿根图上的圆 纯数学 110
第 7 题 曲线描绘 纯数学 114
第 8 题 向量与三维几何 纯数学 120
第 9 题 移动半球上的质点 力学 124
第 10 题 斜碰撞 力学 128
第 11 题 稀松稀疏化 概率与统计 132
第 12 题 随机子集 概率与统计 136

详细解答

STEP I 2019，第 1 题

纯数学 · 三角函数的最优化

> **STEP I 2019，第 1 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

一切都取决于直线与两轴的交点。求周界（perimeter）时，定点 $P=(1,k)$ 把斜边（hypotenuse）XY 分成两段，每段都是一个小直角三角形（right-angled triangle）的斜边，而且该小三角形有一条边的长度已知，所以根本不必展开任何平方根。求最小值时，面积是两项之和，而这两项的积是常数（正是算术－几何平均不等式（AM–GM inequality）的典型情况）；至于周界的导数（derivative），只要利用 $\cos^2\theta=(1-\sin\theta)(1+\sin\theta)$，便会漂亮地化简。

解答

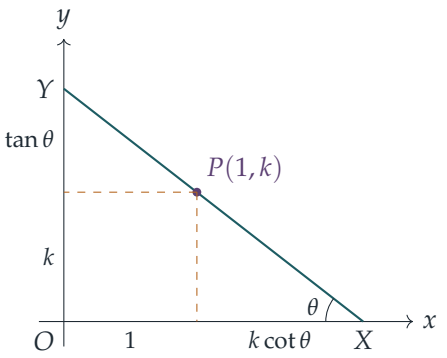
经过 $(1,k)$ 而斜率（gradient）为 $-\tan\theta$ 的直线是

$$y-k=-\tan\theta\,(x-1).$$

代入 $y=0$ 得 $x-1=k\cot\theta$；代入 $x=0$ 得 $y=k+\tan\theta$，所以

$$X=(1+k\cot\theta,\,0),\qquad Y=(0,\,k+\tan\theta).$$

由于 $k>0$ 且 $0<\theta<\frac{1}{2}\pi$，两个截距（intercept）都是正数，因此 OXY 确实是第一象限内的一个直角三角形，而 $P=(1,k)$ 位于其斜边上。



第 (i) 部分

三角形在 O 处是直角，所以

$$\begin{aligned} A &= \tfrac{1}{2}(1 + k\cot\theta)(k + \tan\theta) = \tfrac{1}{2}\big(k + \tan\theta + k^2\cot\theta + k\big) \\ &= k + \tfrac{1}{2}\big(\tan\theta + k^2\cot\theta\big), \end{aligned}$$

这里用了 $\cot\theta\tan\theta = 1$。

求最小值时，$\tan\theta$ 与 $k^2\cot\theta$ 均为正数，而且两者之积是常数 k^2。由算术－几何平均不等式，

$$\tan\theta + k^2\cot\theta \geq 2\sqrt{\tan\theta\cdot k^2\cot\theta} = 2k,$$

等号成立当且仅当 $\tan\theta = k^2\cot\theta$，即 $\tan\theta = k$（由于 $k > 0$，范围内必有某个 θ 达到此值）。因此 $A \geq k + k = 2k$，而

$$A_{\min} = 2k.$$

（用微分亦得同样结果：$\frac{\mathrm{d}A}{\mathrm{d}\theta} = \frac{1}{2}(\sec^2\theta - k^2\operatorname{cosec}^2\theta) = 0$ 当 $\tan^2\theta = k^2$。）取最小值时 $X = (2, 0)$、$Y = (0, 2k)$，所以 P 是 XY 的中点。

第 (ii) 部分

两条直角边分别是 $OX = 1 + k\cot\theta$ 和 $OY = k + \tan\theta$。至于斜边，按图在 P 处把它分开。

- 由 P 到 Y：水平距离是 1，而直线与水平方向成 θ 角，所以 $PY = 1/\cos\theta = \sec\theta$。
- 由 P 到 X：铅垂距离是 k，所以 $PX = k/\sin\theta = k\operatorname{cosec}\theta$。

把四段长度相加，

$$L = 1 + \tan\theta + \sec\theta + k\big(1 + \cot\theta + \operatorname{cosec}\theta\big),$$

正如所求。

想一想 导数中含有 $\sec^2\theta+\sec\theta\tan\theta$。在令它等于零之前，有没有办法把它大大化简?

有的。逐项微分，

$$\frac{\mathrm{d}L}{\mathrm{d}\theta}=\sec^2\theta+\sec\theta\tan\theta-k\big(\operatorname{cosec}^2\theta+\operatorname{cosec}\theta\cot\theta\big)=\frac{1+\sin\theta}{\cos^2\theta}-k\,\frac{1+\cos\theta}{\sin^2\theta}.$$

由于 $\cos^2\theta=(1-\sin\theta)(1+\sin\theta)$、$\sin^2\theta=(1-\cos\theta)(1+\cos\theta)$，而 $1+\sin\theta$ 和 $1+\cos\theta$ 都不为零，所以可以约去:

$$\frac{\mathrm{d}L}{\mathrm{d}\theta}=\frac{1}{1-\sin\theta}-\frac{k}{1-\cos\theta}.$$

它等于零当且仅当 $1-\cos\theta=k(1-\sin\theta)$。所以最小值出现在 $\theta=\alpha$，其中

$$\frac{1-\cos\alpha}{1-\sin\alpha}=k,$$

这正是要证明的。（事实上这是唯一的驻点（stationary point），而且是最小值：当 θ 在 $(0,\frac{1}{2}\pi)$ 上递增时，$\frac{\mathrm{d}L}{\mathrm{d}\theta}$ 的第一项递增，第二项递减，所以 $\frac{\mathrm{d}L}{\mathrm{d}\theta}$ 是递增的，并且只会由负转正一次。）

想一想 怎样才能把最小值中的 k 消去?

把 L 的各项分组并通分:

$$1+\tan\alpha+\sec\alpha=\frac{1+\sin\alpha+\cos\alpha}{\cos\alpha},\qquad 1+\cot\alpha+\operatorname{cosec}\alpha=\frac{1+\sin\alpha+\cos\alpha}{\sin\alpha}.$$

记 $s=\sin\alpha$、$c=\cos\alpha$，且 $k=\dfrac{1-c}{1-s}$，则第二部分变成

$$k\,\frac{1+s+c}{s}=\frac{(1+s+c)(1-c)}{s(1-s)}=\frac{1+s-sc-c^2}{s(1-s)}=\frac{s^2+s-sc}{s(1-s)}=\frac{1+s-c}{1-s},$$

这里用了 $1-c^2=s^2$。因此

$$\begin{aligned}L_{\min}=\frac{1+s+c}{c}+\frac{1+s-c}{1-s}&=\frac{(1+s+c)(1-s)+c(1+s-c)}{c(1-s)}\\&=\frac{1-s^2+c-sc+c+sc-c^2}{c(1-s)}\\&=\frac{2c}{c(1-s)},\end{aligned}$$

因为 $1-s^2-c^2=0$。所以

$$L_{\min}=\frac{2}{1-\sin\alpha}.$$

验算

取 $k=1$。此时 $\tan\theta=1$ 给出 $X=(2,0)$、$Y=(0,2)$ 及 $A=2=2k$，与预测相符。周界的条件 $(1-\cos\alpha)/(1-\sin\alpha)=1$ 给出 $\alpha=\frac{1}{4}\pi$，而 $2/(1-\frac{1}{\sqrt{2}})=4+2\sqrt{2}\approx 6.828$，恰好等于顶点为 $(0,0)$、$(2,0)$、$(0,2)$ 的三角形的周界 $2+2+2\sqrt{2}$。对于其他一系列 k 值，用电脑对 θ 进行搜索，亦证实了两个最小值以及周界最小值出现的位置。

常见错误

- 求斜边时去展开 $\sqrt{(1+k\cot\theta)^2+(k+\tan\theta)^2}$，而不是在 P 处把它分开。
- 直接令未化简的导数等于零，结果迷失在算式中；应先约去 $1+\sin\theta$ 和 $1+\cos\theta$。
- 最终答案中仍保留 k：题目要求只用 α 表示最小值。

STEP 之外

这个简洁的答案背后有一个几何解释。经过 P 而使周界最小的直线，正是在 P 处与一个同时和两轴相切的圆（OXY 的一个旁切圆（excircle））相切的那条直线；此时周界等于由 O 到该圆的切线长的两倍。这个圆经过 P，半径为 r，且 $r=1/(1-\sin\alpha)$。带有“直线经过 P”这类约束条件的最优化（optimisation）问题，正是大学一年级微积分课程中拉格朗日乘数法（Lagrange multipliers）的起点。想看更多同类问题，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

STEP I 2019，第 2 题

纯数学 · 参数曲线的切线

STEP I 2019，第 2 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

这个参数化（parametrisation）的设计，使参数 t 处的斜率（gradient）恰好就是 t，因此每条切线（tangent line）的方程都非常简洁。要找两条切线的交点，只需把立方差（difference of cubes）因式分解（factorise）；而垂直（perpendicular）条件 $pq = -1$ 让我们可以把所有式子改写成以和 $p + q$ 表示。至于两条曲线的公共点，把其中一个参数化代入另一条曲线，会得到一个带有重根（repeated root）的多项式（polynomial），而这个重根正透露了某种几何信息。

解答

切线

设 $x = 3t^2$ 及 $y = 2t^3$，则

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t} = \frac{6t^2}{6t} = t \qquad (t \neq 0).$$

所以在 $(3p^2, 2p^3)$ 处的切线斜率为 p，方程为

$$y - 2p^3 = p\,(x - 3p^2) \quad \Longrightarrow \quad y = px - 3p^3 + 2p^3 = px - p^3,$$

与所求相符。（当 $p = 0$ 时，该点是原点，是一个尖点（cusp）；在该处曲线为 $y = \pm 2(x/3)^{3/2}$，它与 x 轴 $y = 0$ 相切，所以公式仍然成立。）

两条切线的交点

当 $p \neq q$ 时，切线 $y = px - p^3$ 与 $y = qx - q^3$ 的交点满足

$$(p-q)\,x = p^3 - q^3 = (p-q)(p^2+pq+q^2).$$

由于 $p \neq q$，可以两边除以 $p-q$，得 $x = p^2+pq+q^2$，于是

$$y = px - p^3 = p^3 + p^2q + pq^2 - p^3 = pq(p+q).$$

所以两条切线相交于

$$\left(p^2+pq+q^2,\ pq(p+q)\right).$$

两条切线的斜率分别为 p 和 q，所以它们互相垂直当且仅当 $pq = -1$。设 $u = p+q$，则 $p^2+pq+q^2 = (p+q)^2 - pq = u^2+1$，且 $pq(p+q) = -u$。因此两条互相垂直的切线相交于 $(u^2+1,\ -u)$，与所求相符。

想一想 $\widetilde{C}$ 上的每一点，是否真的都来自一对互相垂直的切线?

是的：对任意实数 u，我们需要 $p+q = u$ 及 $pq = -1$，所以 p 和 q 是 $z^2 - uz - 1 = 0$ 的根。其判别式（discriminant）u^2+4 恒为正，所以总有两个相异实根。因此 $\widetilde{C}$ 恰好就是所有可以向 C 作两条互相垂直切线的点所组成的集合。

两条曲线的公共点

消去参数后，$\widetilde{C}$ 是抛物线（parabola）$x = y^2+1$。C 上的点 $(3t^2, 2t^3)$ 位于其上当且仅当

$$3t^2 = 4t^6 + 1 \quad\Longleftrightarrow\quad 4s^3 - 3s + 1 = 0, \quad \text{其中} s = t^2.$$

由于 $s = -1$ 是一个根，可以因式分解：

$$4s^3 - 3s + 1 = (s+1)(4s^2 - 4s + 1) = (s+1)(2s-1)^2.$$

因为 $s = t^2 \geq 0$，唯一可能是 $s = \frac{1}{2}$，所以 $t = \pm\frac{1}{\sqrt{2}}$。于是 $x = 3t^2 = \frac{3}{2}$，$y = 2t^3 = \pm\frac{1}{\sqrt{2}}$。公共点为

$$\left(\tfrac{3}{2},\ \tfrac{1}{\sqrt{2}}\right) \quad \text{及} \quad \left(\tfrac{3}{2},\ -\tfrac{1}{\sqrt{2}}\right).$$

想一想 重复因式 $(2s-1)^2$ 告诉了我们甚么?

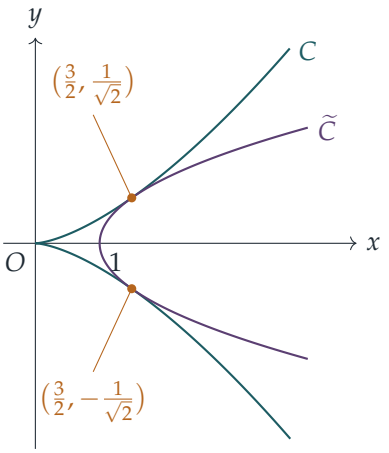
重根通常意味着相切（tangency），这里正是如此。在 $t = \frac{1}{\sqrt{2}}$ 处，C 的斜率为 $t = \frac{1}{\sqrt{2}}$。在 $\widetilde{C}$ 上，由 $x = y^2+1$ 得 $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{2y} = \frac{1}{2/\sqrt{2}} = \frac{1}{\sqrt{2}}$，数值相同。由对称性，下方那一点亦然。所以两条曲线在这两点都只是相切，而不是相交穿过。

草图

从 C 消去 t，得 $y^2 = \frac{4}{27}x^3$，其中 $x \geq 0$。这是一条半立方抛物线（semicubical parabola），关于 x 轴对称，在原点有一个指向左方的尖点。$\widetilde{C}$ 是抛物线 $y^2 = x - 1$，顶点（vertex）为 $(1, 0)$，开口向右。要判断哪条曲线在外，可比较 $x \geq 1$ 时两者 y^2 的值：

$$f(x) = \tfrac{4}{27}x^3 - (x-1), \qquad f'(x) = \tfrac{4}{9}x^2 - 1,$$

所以 f 在 $x \geq 1$ 上的最小值出现在 $x = \frac{3}{2}$，此时 $f(\frac{3}{2}) = \frac{1}{2} - \frac{1}{2} = 0$。因此 $f(x) \geq 0$，且等号只在 $x = \frac{3}{2}$ 时成立。所以 $\widetilde{C}$ 位于 C 的内侧（较接近 x 轴），并在两个公共点与 C 相切。



验算

- 取 $p = 1$，$q = -1$（于是 $pq = -1$，$u = 0$）。切线为 $y = x - 1$ 及 $y = -x + 1$，两者互相垂直，相交于 $(1, 0)$，即 $\widetilde{C}$ 的顶点，与 $(u^2 + 1, -u)$ 的预测一致。
- 取 $p = 2$, $q = -\frac{1}{2}$, 于是 $u = \frac{3}{2}$。切线 $y = 2x - 8$ 与 $y = -\frac{1}{2}x + \frac{1}{8}$ 的交点满足 $\frac{5}{2}x = \frac{65}{8}$，所以 $x = \frac{13}{4} = u^2 + 1$，$y = -\frac{3}{2} = -u$。
- 点 $(1.5, 0.7071)$ 同时满足 $y^2 = \frac{4}{27}x^3$（因为 $\frac{4}{27} \cdot \frac{27}{8} = \frac{1}{2}$）及 $y^2 = x - 1$。电脑验算亦确认 $3t^2 - 4t^6 - 1$ 从不大于零，所以没有其他公共点。

常见错误

- 忘记 $s = t^2$ 不可能是负数，而尝试使用根 $s = -1$。
- 把抛物线画成在公共点穿过 C，而不是在该处与 C 相切。

STEP 之外

曲线 $\widetilde{C}$ 是 C 的正交切线轨迹（orthoptic）：即所有可以向 C 作两条互相垂直切线的点的轨迹（locus）；对抛物线而言，它就是准线（directrix）。以一族切线来描述曲线，会引出包络线（envelope）以及射影几何（projective geometry）中的对偶（duality）概念。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题解》）收录了更多这类题目。

详细解答

STEP I 2019，第 3 题

纯数学 · 积分法

STEP I 2019，第 3 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

把分子和分母同乘以 $1-\sin x$，就能把 $1+\sin x$ 变成 $1-\sin^2 x=\cos^2 x$，而分母是 $\cos x$ 的幂的式子，都可以拆成几个标准部分：$\sec^2 x$、$\sec x\tan x$、$\tan^2 x\sec^2 x$ 和 $\sin x/\cos^4 x$，每一个都可以一眼看出其积分（integral）。另外两个积分其实是同一招的变奏：第二个积分会变成含 $1+\cos x$ 的积分，再用代换（substitution）$x=\frac{1}{2}\pi-u$ 便化为第一类；第三个积分则把因子 $1-\sin x$ 平方。

解答

第一个积分

由于 $(1+\sin x)(1-\sin x)=1-\sin^2 x=\cos^2 x$，

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

我们知道 $\frac{\mathrm{d}}{\mathrm{d}x}\tan x=\sec^2 x$ 及 $\frac{\mathrm{d}}{\mathrm{d}x}\sec x=\sec x\tan x$，所以

$$\int\frac{1}{1+\sin x}\,\mathrm{d}x=\tan x-\sec x+C. \qquad (*)$$

利用 $\tan\frac{1}{4}\pi=1$、$\sec\frac{1}{4}\pi=\sqrt{2}$、$\tan 0=0$ 及 $\sec 0=1$：

$$\int_0^{\frac{1}{4}\pi}\frac{1}{1+\sin x}\,\mathrm{d}x=\left(1-\sqrt{2}\right)-\left(0-1\right)=2-\sqrt{2}.$$

（乘以 $1-\sin x$ 是可以的，因为在整个区间内 $\sin x<1$，所以我们从不会乘以零。）

第二个积分

把分子和分母同乘以 $\cos x$（在 $[\frac{1}{4}\pi, \frac{1}{3}\pi]$ 上不为零），

$$\frac{1}{1+\sec x} = \frac{\cos x}{\cos x + 1} = 1 - \frac{1}{1+\cos x}.$$

想一想 怎样才能把 $(*)$ 用在 $\dfrac{1}{1+\cos x}$ 上?

由于 $\cos x = \sin(\frac{1}{2}\pi - x)$，代入 $u = \frac{1}{2}\pi - x$，于是 $\mathrm{d}x = -\mathrm{d}u$；当 $x = \frac{1}{4}\pi$ 时，$u = \frac{1}{4}\pi$；当 $x = \frac{1}{3}\pi$ 时，$u = \frac{1}{6}\pi$。于是

$$\begin{aligned}\int_{\frac{1}{4}\pi}^{\frac{1}{3}\pi} \frac{1}{1+\cos x}\,\mathrm{d}x = \int_{\frac{1}{6}\pi}^{\frac{1}{4}\pi} \frac{1}{1+\sin u}\,\mathrm{d}u &= \Big[\tan u - \sec u\Big]_{\frac{1}{6}\pi}^{\frac{1}{4}\pi} \\ &= (1-\sqrt{2}) - \Big(\frac{1}{\sqrt{3}} - \frac{2}{\sqrt{3}}\Big) = 1 - \sqrt{2} + \frac{1}{\sqrt{3}},\end{aligned}$$

这里用了 $\tan\frac{1}{6}\pi = 1/\sqrt{3}$ 及 $\sec\frac{1}{6}\pi = 2/\sqrt{3}$。区间长度为 $\frac{1}{3}\pi - \frac{1}{4}\pi = \frac{1}{12}\pi$，所以

$$\int_{\frac{1}{4}\pi}^{\frac{1}{3}\pi} \frac{1}{1+\sec x}\,\mathrm{d}x = \frac{\pi}{12} - \Big(1 - \sqrt{2} + \frac{1}{\sqrt{3}}\Big) = \frac{\pi}{12} + \sqrt{2} - 1 - \frac{\sqrt{3}}{3}.$$

（同样地，也可以把 $\frac{1}{1+\cos x}$ 乘以 $\frac{1-\cos x}{1-\cos x}$，得到 $\operatorname{cosec}^2 x - \cot x \operatorname{cosec} x$，其原函数（antiderivative）为 $\operatorname{cosec} x - \cot x = \tan\frac{1}{2}x$；由于 $\tan\frac{1}{8}\pi = \sqrt{2}-1$，这会得出相同的答案。）

第三个积分

现在把分子和分母同乘以 $(1-\sin x)^2$，使分母变成 $\cos^4 x$：

$$\frac{1}{(1+\sin x)^2} = \frac{(1-\sin x)^2}{\cos^4 x} = \frac{1+\sin^2 x}{\cos^4 x} - \frac{2\sin x}{\cos^4 x}.$$

对于第一部分，在分子中写 $1 = \sin^2 x + \cos^2 x$：

$$\frac{1+\sin^2 x}{\cos^4 x} = \frac{\cos^2 x + 2\sin^2 x}{\cos^4 x} = \sec^2 x + 2\tan^2 x \sec^2 x.$$

想一想 $\tan^2 x \sec^2 x$ 和 $\sin x / \cos^4 x$ 的积分分别是甚么?

两者都是 $f(x)^n f'(x)$ 的形式。取 $f = \tan x$，有 $\frac{\mathrm{d}}{\mathrm{d}x}\tan^3 x = 3\tan^2 x \sec^2 x$；取 $f = \cos x$，有 $\frac{\mathrm{d}}{\mathrm{d}x}(\cos x)^{-3} = 3\sin x\,(\cos x)^{-4}$。因此

$$\int \frac{1}{(1+\sin x)^2}\,\mathrm{d}x = \tan x + \tfrac{2}{3}\tan^3 x - \tfrac{2}{3}\sec^3 x + C.$$

当 $x=\frac{1}{4}\pi$ 时：$\tan=1$ 而 $\sec^3=(\sqrt{2})^3=2\sqrt{2}$，得 $1+\frac{2}{3}-\frac{4}{3}\sqrt{2}$。当 $x=0$ 时：$0+0-\frac{2}{3}$。所以

$$\int_0^{\frac{1}{4}\pi}\frac{1}{(1+\sin x)^2}\,\mathrm{d}x=\frac{5}{3}-\frac{4\sqrt{2}}{3}+\frac{2}{3}=\frac{7-4\sqrt{2}}{3}.$$

验算

- 大小：在 $[0,\frac{1}{4}\pi]$ 上，第一个被积函数（integrand）由 1 递减至 $2-\sqrt{2}\approx0.59$，所以积分值介乎 0.46 与 0.79 之间；$2-\sqrt{2}\approx0.586$ 符合。同样，第三个被积函数是第一个的平方，所以其积分应该较小：$(7-4\sqrt{2})/3\approx0.448$。
- 把你的原函数求导：$\frac{\mathrm{d}}{\mathrm{d}x}(\tan x-\sec x)=\frac{1-\sin x}{\cos^2 x}=\frac{1}{1+\sin x}$。
- 辛普森法则（Simpson's rule）给出三个积分的值为 0.585786、0.098663 及 0.447715，与 $2-\sqrt{2}$、$\frac{\pi}{12}+\sqrt{2}-1-\frac{\sqrt{3}}{3}$ 及 $\frac{1}{3}(7-4\sqrt{2})$ 准确至十位小数。

常见错误

- $\frac{\mathrm{d}}{\mathrm{d}x}\sec x$ 或 $\frac{\mathrm{d}}{\mathrm{d}x}\sec^3 x$ 的正负号出错：两者都带正号，因为 $\frac{\mathrm{d}}{\mathrm{d}x}(\cos x)^{-n}=n\sin x(\cos x)^{-n-1}$。
- 在第一和第三个积分中忘记下限：$\sec 0=1$，不是 0。
- 代换 $u=\frac{1}{2}\pi-x$ 后错误地对调上下限：$\mathrm{d}x=-\mathrm{d}u$ 中的负号正是用来把上下限调回递增次序的。

STEP 之外

乘以“共轭”（conjugate）这个技巧，其实是一个系统方法的特例：代换 $t=\tan\frac{1}{2}x$ 能把任何 $\sin x$ 与 $\cos x$ 的有理函数（rational function）变成 t 的有理函数，然后便可用部分分式（partial fractions）求积分。（我们得到 $\int\frac{\mathrm{d}x}{1+\cos x}$ 的答案是 $\tan\frac{1}{2}x$，正好透露了这一点。）大学一年级的分析（analysis）课程还会探讨这类积分在甚么情况下才存在，例如在 $x=-\frac{1}{2}\pi$ 附近，$1+\sin x$ 等于零。关于这两个概念，可参阅 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）；如想做更多同类风格的积分，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）。

详细解答

STEP I 2019，第 4 题

纯数学 · 四次方程的因式分解

STEP I 2019，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

没有 x^3 项的四次多项式（quartic），其各根（roots）之和为零。因此，若把四个根分成两对，两对各自的和互为相反数，这正是两个二次因式（quadratic factors）的 x 系数分别为 s 和 $-s$ 的原因。接着比较系数（comparing coefficients）得出三条方程，消去 p 和 q 后便剩下一个关于 s^2 的三次方程（cubic）。第 (i) 部分是为了处理最后一步的根式（surds）而设。

解答

第 (i) 部分

把 $m+n\sqrt{2}$ 平方得 $m^2+2n^2+2mn\sqrt{2}$，所以我们需要 $m^2+2n^2=3$ 及 $mn=1$。取 $m=n=1$ 即可：

$$(1+\sqrt{2})^2 = 1+2\sqrt{2}+2 = 3+2\sqrt{2},$$

而由于 $1+\sqrt{2}>0$，它就是正平方根。因此 $m=1$，$n=1$：

$$\sqrt{3+2\sqrt{2}} = 1+\sqrt{2}.$$

同理，$(\sqrt{2}-1)^2 = 3-2\sqrt{2}$，所以 $\sqrt{3-2\sqrt{2}} = \sqrt{2}-1$。两个结果我们都会用到。

第 (ii) 部分

为何存在这样的因式分解。设四个实根为 a、b、c、d，则 $\mathrm{f}(x)=(x-a)(x-b)(x-c)(x-d)$。f 中 x^3 的系数是 $-(a+b+c+d)$，而它等于零，所以 $c+d=-(a+b)$。把根两两配对，

$$\mathrm{f}(x) = \big(x^2-(a+b)x+ab\big)\big(x^2-(c+d)x+cd\big) = \big(x^2+sx+p\big)\big(x^2-sx+q\big),$$

其中 $s=-(a+b)$、$p=ab$、$q=cd$，由于各根皆为实数，它们全都是实数。

三条方程。展开，

$$(x^2+sx+p)(x^2-sx+q)=x^4+(p+q-s^2)x^2+s(q-p)x+pq.$$

与 $x^4-10x^2+12x-2$ 比较：

$$\begin{aligned} p+q-s^2&=-10,\\ s(q-p)&=12,\\ pq&=-2. \end{aligned}$$

关于 s 的方程。第二条方程显示 $s\neq 0$，所以可以除以 s：$q-p=12/s$。第一条方程给出 $p+q=s^2-10$。现在利用恒等式 $(p+q)^2-(q-p)^2=4pq=-8$：

$$(s^2-10)^2-\frac{144}{s^2}=-8.$$

两边乘以 s^2（它不为零）得

$$s^2(s^2-10)^2+8s^2-144=0,$$

此即所求。

s^2 的值。设 $u=s^2$。展开，

$$u(u-10)^2+8u-144=u^3-20u^2+108u-144.$$

试代入 144 的较小因数，$u=2$ 给出 $8-80+216-144=0$，所以 $u-2$ 是一个因式，并且

$$u^3-20u^2+108u-144=(u-2)(u^2-18u+72)=(u-2)(u-6)(u-12).$$

所以 s^2 的三个可能值为 $s^2=2$、6 及 12。

> ? **想一想** 我们只需要一种因式分解，为甚么 s^2 却有三个值?

把四个根分成两对共有三种方法，而每一种配对都给出各自的一对二次因式，且 $s^2=(a+b)^2$。用任何一种都可以解方程；取最小的值只是令根式最整洁。

解 $\mathrm{f}(x)=0$。取 $s^2=2$ 及 $s=\sqrt{2}$（选 $s=-\sqrt{2}$ 只会把两个因式对调）。于是

$$p+q=2-10=-8,\qquad q-p=\frac{12}{\sqrt{2}}=6\sqrt{2},$$

所以 $p=-4-3\sqrt{2}$，$q=-4+3\sqrt{2}$。（验证：$pq=16-18=-2$。）因此

$$\mathrm{f}(x)=\left(x^2+\sqrt{2}\,x-4-3\sqrt{2}\right)\left(x^2-\sqrt{2}\,x-4+3\sqrt{2}\right).$$

第一个因式的判别式（discriminant）为

$$2+4(4+3\sqrt{2})=18+12\sqrt{2}=6(3+2\sqrt{2})=6(1+\sqrt{2})^2,$$

这里用了第 (i) 部分的结果，所以其平方根为 $\sqrt{6}(1+\sqrt{2})=\sqrt{6}+2\sqrt{3}$。其根为

$$x=\frac{-\sqrt{2}\pm(\sqrt{6}+2\sqrt{3})}{2}.$$

第二个因式的判别式为

$$2-4(-4+3\sqrt{2})=18-12\sqrt{2}=6(3-2\sqrt{2})=6(\sqrt{2}-1)^2,$$

其平方根为 $\sqrt{6}(\sqrt{2}-1)=2\sqrt{3}-\sqrt{6}$（为正数，因为 $2\sqrt{3}=\sqrt{12}>\sqrt{6}$）。其根为

$$x=\frac{\sqrt{2}\pm(2\sqrt{3}-\sqrt{6})}{2}.$$

所以 $\mathrm{f}(x)=0$ 的四个解为

$$x=\tfrac{1}{2}(-\sqrt{2}+\sqrt{6}+2\sqrt{3}),\qquad x=\tfrac{1}{2}(-\sqrt{2}-\sqrt{6}-2\sqrt{3}),$$
$$x=\tfrac{1}{2}(\sqrt{2}+2\sqrt{3}-\sqrt{6}),\qquad x=\tfrac{1}{2}(\sqrt{2}-2\sqrt{3}+\sqrt{6}).$$

验算

- 各根的数值约为 2.2497、-3.6639、1.2144 及 0.1998。它们之和为 0，正如没有 x^3 项所要求的；它们之积约为 -2，即常数项。
- 代入 $x\approx 0.1998$：$x^4-10x^2+12x-2\approx 0.0016-0.3992+2.3976-2\approx 0$。
- f 的变号情况吻合：$\mathrm{f}(0)=-2$，$\mathrm{f}(1)=1$，$\mathrm{f}(2)=-2$，$\mathrm{f}(3)=25$，$\mathrm{f}(-4)=46$，所以 $(-4,0)$、$(0,1)$、$(1,2)$ 及 $(2,3)$ 各有一个根。

常见错误

- 没有解释为何 x 系数是 s 和 $-s$：这源于 x^3 的系数为零，即各根之和为零。
- 除以 s 之前忘记指出 $s\neq 0$。
- 把根写成 $\sqrt{18\pm 12\sqrt{2}}$ 便停下：题目要求完全化简，而这正是第 (i) 部分的用途。

STEP 之外

这其实是经过包装的费拉里（Ferrari）四次方程解法：关于 s^2 的三次方程就是预解三次式（resolvent cubic），它的三个根对应把四个根两两配对的三种方法。为何四次方程可以

化为三次方程, 而一般的五次方程（quintic）却根本无法用根式求解（solved by radicals），正是伽罗瓦理论（Galois theory）的主题；这套理论把方程的可解性与其根的对称性联系起来。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是大学一年级迈向这个理论的好途径。

详细解答

STEP I 2019，第 5 题

纯数学 · 三维向量

STEP I 2019，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

两个向量（vector）相等，当且仅当它们的分量（component）全部相同，因此题中每个几何条件都可化为关于 p、q、r 和 s 的几条简单方程。四点共面（coplanar）等价于纯量三重积（scalar triple product）（即一个 3×3 行列式（determinant））等于零。把条件用坐标写出后，每一部分都只是简短的代数运算。

解答

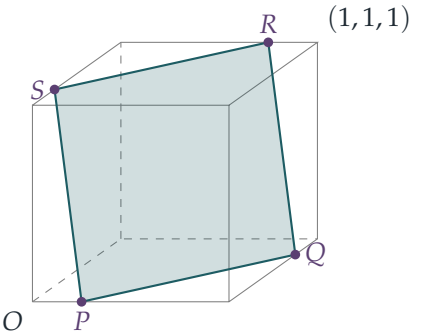
第 (i) 部分

若 $\overrightarrow{PQ}=\overrightarrow{SR}$,则对边 PQ 与 SR 平行且等长,所以 $PQRS$ 是**平行四边形**(parallelogram)。(此时由 $\overrightarrow{PS}=\overrightarrow{PQ}+\overrightarrow{QR}-\overrightarrow{SR}$ 亦得 $\overrightarrow{PS}=\overrightarrow{QR}$。)

若同时有 $|\overrightarrow{PQ}|=|\overrightarrow{PS}|$，则平行四边形的两条邻边相等，因而四条边全部相等：$PQRS$ 是**菱形**（rhombus）。

第 (ii) 部分：共面条件

各点位于单位立方体（unit cube）的棱上，所以 $p,q,r,s\in[0,1]$。这里 P 和 Q 位于面 $z=0$ 上，而 R 和 S 位于面 $z=1$ 上。



四点共面，当且仅当 $\overrightarrow{PQ}$、$\overrightarrow{PS}$ 和 $\overrightarrow{PR}$ 线性相关（linearly dependent），亦即它们的纯量三重积为零。由

$$\overrightarrow{PQ} = \begin{pmatrix} 1-p \\ q \\ 0 \end{pmatrix}, \qquad \overrightarrow{PS} = \begin{pmatrix} -p \\ s \\ 1 \end{pmatrix}, \qquad \overrightarrow{PR} = \begin{pmatrix} r-p \\ 1 \\ 1 \end{pmatrix},$$

沿第一行展开行列式，得

$$\begin{vmatrix} 1-p & q & 0 \\ -p & s & 1 \\ r-p & 1 & 1 \end{vmatrix} = (1-p)(s-1) - q\big(-p-(r-p)\big) + 0$$
$$= (1-p)(s-1) + qr.$$

令其等于零，即得 $rq = (1-s)(1-p)$，正是所求。

如果你想避开行列式，可把经过 P、Q、R 的平面写成 $\overrightarrow{OP} + \lambda\overrightarrow{PQ} + \mu\overrightarrow{PR}$，其上的点为

$$\big(p + \lambda(1-p) + \mu(r-p),\ \lambda q + \mu,\ \mu\big).$$

要 $S = (0, s, 1)$ 在此平面上，由 z 分量需 $\mu = 1$，再由 y 分量得 $\lambda q = s - 1$，由 x 分量得 $r + \lambda(1-p) = 0$。把最后一条方程乘以 q 并代入，得 $rq + (s-1)(1-p) = 0$，即同一条件。（这里用到 P、Q、R 不共线（collinear），这是成立的，因为 R 在顶面，而 P、Q 在底面。）

第 (ii)(a) 部分

我们有

$$\overrightarrow{PQ} = \begin{pmatrix} 1-p \\ q \\ 0 \end{pmatrix}, \qquad \overrightarrow{SR} = \begin{pmatrix} r \\ 1-s \\ 0 \end{pmatrix}.$$

两者的 z 分量已经相同，所以

$$\overrightarrow{PQ} = \overrightarrow{SR} \iff r = 1-p \text{ 且 } q = 1-s \iff p+r = 1 \text{ 且 } q+s = 1.$$

形心（centroid）为

$$\tfrac{1}{4}(\overrightarrow{OP}+\overrightarrow{OQ}+\overrightarrow{OR}+\overrightarrow{OS}) = \tfrac{1}{4}\begin{pmatrix} p+1+r \\ q+1+s \\ 2 \end{pmatrix},$$

而立方体的中心是 $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$。第三个分量总是相同，所以形心位于中心当且仅当 $p+1+r=2$ 且 $q+1+s=2$，这正是同样的两个条件 $p+r=1$、$q+s=1$。因此两个命题等价。

（留意当 $r=1-p$ 且 $s=1-q$ 时，共面条件 $rq=(1-s)(1-p)$ 变成 $(1-p)q=q(1-p)$，自动成立：没有任何位置被排除。）

第 (ii)(b) 部分

由 (a)，$\overrightarrow{PQ}=\overrightarrow{SR}$ 给出 $r=1-p$ 及 $s=1-q$。对于第二个条件，比较长度的平方，并利用 $\overrightarrow{PS}=(-p,\,s,\,1)$ 及 $s=1-q$：

$$\begin{aligned} |\overrightarrow{PQ}|^2 = |\overrightarrow{PS}|^2 &\iff (1-p)^2+q^2 = p^2+(1-q)^2+1 \\ &\iff 1-2p = 2-2q \\ &\iff q = p+\tfrac{1}{2}. \end{aligned}$$

所以

$$q = p+\tfrac{1}{2}, \qquad r = 1-p, \qquad s = \tfrac{1}{2}-p.$$

由于 $q \le 1$ 及 $s \ge 0$，只有当 $0 \le p \le \frac{1}{2}$ 时各点才位于棱上。

> ? **想一想** 哪两个向量给出角 PQR?

在 Q 处的角是 $\overrightarrow{QP}$ 与 $\overrightarrow{QR}$ 之间的夹角（两者都背离 Q）。利用刚求得的值，

$$\overrightarrow{QP} = \begin{pmatrix} p-1 \\ -p-\frac{1}{2} \\ 0 \end{pmatrix}, \qquad \overrightarrow{QR} = \begin{pmatrix} r-1 \\ 1-q \\ 1 \end{pmatrix} = \begin{pmatrix} -p \\ \frac{1}{2}-p \\ 1 \end{pmatrix}.$$

于是

$$\begin{aligned} \overrightarrow{QP}\cdot\overrightarrow{QR} &= -p(p-1)-(p+\tfrac{1}{2})(\tfrac{1}{2}-p) = p-p^2-(\tfrac{1}{4}-p^2) = p-\tfrac{1}{4}, \\ |\overrightarrow{QP}|^2 &= (1-p)^2+(p+\tfrac{1}{2})^2 = 2p^2-p+\tfrac{5}{4}, \\ |\overrightarrow{QR}|^2 &= p^2+(\tfrac{1}{2}-p)^2+1 = 2p^2-p+\tfrac{5}{4}. \end{aligned}$$

两个长度相等，这正是菱形应有的性质，所以

$$\cos PQR = \frac{p-\frac{1}{4}}{2p^2-p+\frac{5}{4}} = \frac{4p-1}{5-4p+8p^2},$$

这是把分子分母同乘以 4 所得。这就是所求的结果。

正方形。 菱形是正方形，当且仅当它的角都是直角，即 $\cos PQR = 0$。分母 $8p^2-4p+5 = 8(p-\frac{1}{4})^2+\frac{9}{2}$ 恒为正，所以这只在 $4p-1=0$ 时发生：

$$p=\tfrac{1}{4}, \qquad q=\tfrac{3}{4}, \qquad r=\tfrac{3}{4}, \qquad s=\tfrac{1}{4}.$$

此时边长的平方为

$$2\cdot\tfrac{1}{16}-\tfrac{1}{4}+\tfrac{5}{4}=\tfrac{9}{8},$$

所以每条边的长度为 $\sqrt{9/8}=\frac{3\sqrt{2}}{4}$。要与 $\frac{21}{20}$ 比较，可比较这两个正数的平方：

$$\left(\tfrac{21}{20}\right)^2 = \tfrac{441}{400} < \tfrac{450}{400} = \tfrac{9}{8}.$$

因此正方形的边长大于 $\frac{21}{20}$。（实际上约为 1.0607。）

验算

- 当 $p=\frac{1}{4}$ 时，四点为 $(\frac{1}{4},0,0)$、$(1,\frac{3}{4},0)$、$(\frac{3}{4},1,1)$、$(0,\frac{1}{4},1)$；它们的平均值是 $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$，与 (a) 的预测一致，而直接计算亦得 $\overrightarrow{QP}\cdot\overrightarrow{QR}=\frac{3}{16}-\frac{3}{16}=0$。
- 当 $p=0$ 时，公式给出 $\cos PQR=-\frac{1}{5}$。直接计算：$\overrightarrow{QP}=(-1,-\frac{1}{2},0)$，$\overrightarrow{QR}=(0,\frac{1}{2},1)$，点积（dot product）为 $-\frac{1}{4}$，长度平方均为 $\frac{5}{4}$，所以 $\cos PQR=-\frac{1}{5}$。✓
- 共面恒等式、(a) 中的等价关系及 $\cos PQR$ 的公式，均已用电脑以精确算术在大量随机数值上验证。

常见错误

- 用 $\overrightarrow{QP}$ 和 $\overrightarrow{RQ}$（一个指向 Q，一个背离 Q）求在 Q 处的角：这会令余弦的符号相反。
- 弄错字母次序：$\overrightarrow{SR}$ 是 $R-S$，不是 $S-R$。
- 单凭计算机数值就断言 $\frac{3\sqrt{2}}{4}>\frac{21}{20}$，而没有作精确比较；由于两者都是正数，把两边平方可以保持论证严谨。

STEP 之外

用行列式判断共面，是一个一般事实在三维的特例：一组向量线性相关，当且仅当由它们组成的矩阵（matrix）的行列式为零。立方体的平面截面也是丰富的例子来源（经过中心并垂直于一条体对角线的平面，截出一个正六边形（regular hexagon））。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）正是从这种几何观点出发，发展行列式、体积与线性相关的理论。

详细解答

STEP I 2019，第 6 题

纯数学 · 不等式与界限

STEP I 2019，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

每条方程直接求解看来都无从入手，但可以把它拆成“某式 = 另一式”的形式，其中左边永远不会小于某个数，而右边永远不会大于同一个数。于是方程只在两边都恰好等于那个数时才成立，而这两个条件各自都很容易求解。每一部分前面的小题，正是为了建立这两个界（bound）。

解答

第 (i) 部分

配方（completing the square），

$$\begin{aligned} 9x^2 - 12x\cos\theta + 4 &= (3x - 2\cos\theta)^2 - 4\cos^2\theta + 4 \\ &= (3x - 2\cos\theta)^2 + 4\sin^2\theta. \end{aligned}$$

平方数永不为负，所以最小值是 $4\sin^2\theta$，当且仅当 $x = \frac{2}{3}\cos\theta$ 时取得。

把这个四次式（quartic）看成 x^2 的二次式（quadratic），

$$12x^2\sin\theta - 9x^4 = -\left(3x^2 - 2\sin\theta\right)^2 + 4\sin^2\theta.$$

所以 $12x^2\sin\theta - 9x^4 \le 4\sin^2\theta$，当且仅当 $x^2 = \frac{2}{3}\sin\theta$ 时取等号。由于 $0 < \theta < \pi$，我们有 $\sin\theta > 0$，所以 x^2 确实可以取这个值，而最大值是 $4\sin^2\theta$。

现在把所给方程中的 x^2 项拆成 $9x^2 - 12x^2\sin\theta$，再移项整理：

$$\underbrace{9x^2 - 12x\cos\theta + 4}_{\ge 4\sin^2\theta} = \underbrace{12x^2\sin\theta - 9x^4}_{\le 4\sin^2\theta}.$$

想一想 一个至少是 $4\sin^2\theta$ 的量，怎样才能等于一个至多是 $4\sin^2\theta$ 的量?

只有两者都等于 $4\sin^2\theta$ 才行。所以方程成立当且仅当同时有

$$x = \tfrac{2}{3}\cos\theta \qquad 和 \qquad x^2 = \tfrac{2}{3}\sin\theta.$$

消去 x：$\frac{4}{9}\cos^2\theta = \frac{2}{3}\sin\theta$，所以 $2\cos^2\theta = 3\sin\theta$，即 $2 - 2\sin^2\theta = 3\sin\theta$。因此

$$2\sin^2\theta + 3\sin\theta - 2 = 0 \quad \Longrightarrow \quad (2\sin\theta - 1)(\sin\theta + 2) = 0.$$

由于 $\sin\theta \neq -2$，得 $\sin\theta = \frac{1}{2}$，所以 $\theta = \frac{\pi}{6}$ 或 $\theta = \frac{5\pi}{6}$。再由 $x = \frac{2}{3}\cos\theta$ 分别得 $x = \frac{1}{\sqrt{3}}$ 或 $x = -\frac{1}{\sqrt{3}}$，而两种情况下都有 $x^2 = \frac{1}{3} = \frac{2}{3}\sin\theta$，符合要求。解为

$$(x,\theta) = \left(\tfrac{1}{\sqrt{3}}, \tfrac{\pi}{6}\right) \quad 和 \quad (x,\theta) = \left(-\tfrac{1}{\sqrt{3}}, \tfrac{5\pi}{6}\right).$$

第 (ii) 部分

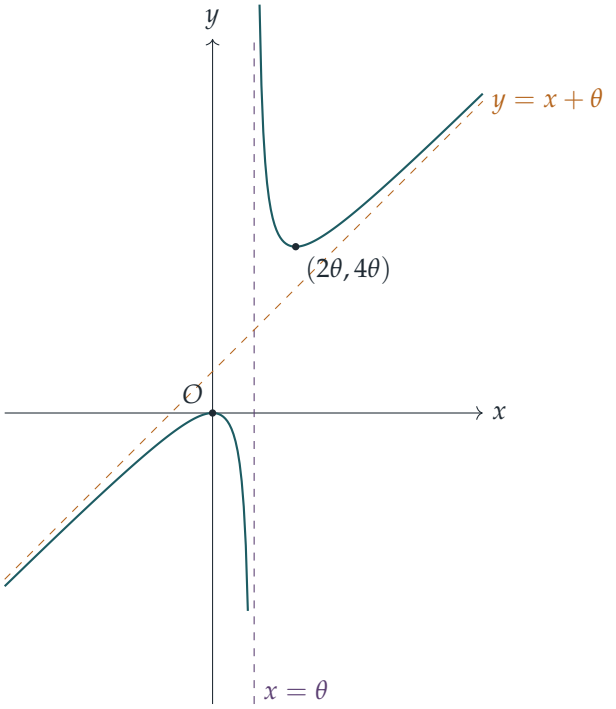
草图。这里 $\theta > 0$。曲线 $y = \dfrac{x^2}{x-\theta}$ 经过原点（而且不在其他地方与坐标轴相交），并有铅垂渐近线（vertical asymptote）$x = \theta$。作除法，

$$\frac{x^2}{x-\theta} = x + \theta + \frac{\theta^2}{x-\theta},$$

所以 $y = x + \theta$ 是一条斜渐近线（oblique asymptote）；当 $x > \theta$ 时曲线在它上方，当 $x < \theta$ 时曲线在它下方。用商法则（quotient rule）求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2x(x-\theta) - x^2}{(x-\theta)^2} = \frac{x(x-2\theta)}{(x-\theta)^2},$$

所以驻点（stationary point）是局部极大点 $(0,0)$ 和局部极小点 $(2\theta, 4\theta)$（斜率在 $x = 0$ 处由 $+$ 变为 $-$，在 $x = 2\theta$ 处由 $-$ 变为 $+$）。



推论。若 $x < \theta$，分子 $x^2 \geq 0$ 而分母为负，所以 $y \leq 0$。若 $x > \theta$，该分支的最低点是 $(2\theta, 4\theta)$，所以 $y \geq 4\theta$。用代数方法说明：当 $x > \theta$ 时，

$$x^2 - 4\theta(x-\theta) = (x-2\theta)^2 \geq 0 \quad \Longrightarrow \quad \frac{x^2}{x-\theta} \geq 4\theta,$$

这里除以了正数 $x - \theta$。所以曲线永远不会取严格介乎 0 与 4θ 之间的值。

三角函数的界。分子满足 $\sin^2\theta\cos^2 x \leq 1 \cdot 1 = 1$，分母满足 $1 + \cos^2\theta\sin^2 x \geq 1$。一个至多为 1 的数除以一个至少为 1（且为正）的数，结果至多为 1：

$$\frac{\sin^2\theta\cos^2 x}{1+\cos^2\theta\sin^2 x} \leq 1.$$

要取等号，必须同时有 $\sin^2\theta\cos^2 x = 1$ 和 $\cos^2\theta\sin^2 x = 0$。另外留意这个分式 ≥ 0，因为它是平方数除以正数。

最后的方程。记 $L = \dfrac{x^2}{4\theta(x-\theta)}$，并以 R 表示右边。把前面的结果除以 $4\theta > 0$，得 $L \leq 0$ 或 $L \geq 1$。但 $0 \leq R \leq 1$。

> ? **想一想** 一个 ≤ 0 或 ≥ 1 的数，能在哪里与 $[0,1]$ 中的数相等?

只能在 0 或 1。所以我们需要 $L = R = 0$ 或 $L = R = 1$。

▶ $L = 0$ 迫使 $x = 0$，但这时 $R = \sin^2\theta > 0$，因为 $0 < \theta < \pi$。无解。

- $L=1$ 表示 $x^2=4\theta(x-\theta)$，即 $(x-2\theta)^2=0$，所以 $x=2\theta$。$R=1$ 需要 $\sin^2\theta\cos^2 x=1$，这迫使 $\sin^2\theta=1$ 和 $\cos^2 x=1$。由于 $0<\theta<\pi$，得 $\theta=\frac{\pi}{2}$，所以 $x=\pi$，而且确实有 $\cos^2\pi=1$ 和 $\cos^2\theta\sin^2 x=0$。

唯一的解是 $x=\pi$，$\theta=\frac{\pi}{2}$。

验算

- 把 $x=\frac{1}{\sqrt{3}},\theta=\frac{\pi}{6}$ 代入四次方程：$9\cdot\frac{1}{9}+(9-6)\cdot\frac{1}{3}-12\cdot\frac{1}{\sqrt{3}}\cdot\frac{\sqrt{3}}{2}+4=1+1-6+4=0$。
- 在最后的方程中取 $x=\pi$，$\theta=\frac{\pi}{2}$：$L=\pi^2/(2\pi\cdot\frac{\pi}{2})=1$，而 $R=1\cdot 1/(1+0)=1$。
- 用电脑在 (x,θ) 的细密网格上搜索，两条方程都没有找到其他近似解，只有在被排除的端点 $\theta\to 0$ 和 $\theta\to\pi$ 附近出现一些虚假的近似解（那里 $\sin\theta\to 0$，两个界都失效）。

常见错误

- 求得 $\sin\theta=\frac{1}{2}$ 就停下来：x 必须满足 $x=\frac{2}{3}\cos\theta$，这对每个 θ 都确定了 x 的正负号，所以 $(\frac{1}{\sqrt{3}},\frac{5\pi}{6})$ 不是解。
- 在最后一部分忘记考虑 $L=R=0$ 的情况；这个情况必须排除，而它之所以能被排除，只是因为 $\sin\theta\neq 0$。
- 声称分式 ≤ 1“因为 $\sin^2$ 和 $\cos^2$ 都不超过 1”，却没有说明为何分母不会破坏这个结论。

STEP 之外

把一条方程夹在一个下界和一个上界之间、而两者只在单一值相遇，这个技巧是分析学（analysis）的常用手法：证明 AM–GM（算术–几何平均）不等式或柯西–施瓦茨（Cauchy–Schwarz）不等式何时取等号，就是这样做的；夹逼定理（squeeze theorem）确定极限，也是靠这个道理。学会问“这个不等式何时取等号?”，往往是找到解法的捷径。想看更多同类思路，可参阅 Abbott 的 *Understanding Analysis*（《理解分析学》），或 Siklos 的习题集 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP I 2019，第 7 题

纯数学 · 无理性与同余

STEP I 2019，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

若以最简形式（lowest terms）写成 $\sqrt{2}+\sqrt{3}=a/b$，平方两次便可消去根式（surd），得到一条关于整数 a 和 b 的多项式（polynomial）方程。接著，我们在模（modulo）某个小质数（prime）下考察这条方程。方程迫使该质数要么同时整除 a 和 b，要么两者都不整除。“两者都整除”与最简形式矛盾，而余数则排除了“两者都不整除”的情况。整个技巧的精髓，在于选出一个能令余数真正产生冲突的质数。

解答

第 (i) 部分

步骤 1。每个整数都可写成 $3k$、$3k+1$ 或 $3k-1$（其中 k 为某整数），因此若 $3\nmid a$，则 $a=3k\pm1$。无论哪种情况，

$$a^2=9k^2\pm6k+1=3(3k^2\pm2k)+1,$$

即比 3 的某个倍数大一。

步骤 3。将 $\dfrac{a}{b}=\sqrt{2}+\sqrt{3}$ 平方，得

$$\frac{a^2}{b^2}=2+2\sqrt{6}+3,\qquad \text{故}\qquad a^2-5b^2=2\sqrt{6}\,b^2.$$

现在根式已单独置于一边，再平方一次便可把它消去：

$$a^4-10a^2b^2+25b^4=24b^4,$$
$$a^4+b^4=10a^2b^2,$$

此即所求。（我们只需要这个方向的蕴涵：步骤 2 中的任何 a, b 都必须满足这条方程。）

步骤 4。假设 $3 \mid a$。将步骤 3 的结果移项，

$$b^4 = 10a^2b^2 - a^4 = a^2\,(10b^2 - a^2),$$

而右方可被 3 整除（甚至可被 9 整除）。若 $3 \nmid b$，则由步骤 1 得 $b^2 = 3m + 1$，于是 $b^4 = (b^2)^2 = 9m^2 + 6m + 1$ 不是 3 的倍数，矛盾。因此 $3 \mid b$。

> **想一想** 步骤 4 处理了 $3 \mid a$ 的情况。那么 3 既不整除 a 也不整除 b 的情况又如何？

步骤 5。步骤 3 中的方程对 a 和 b 是对称的，所以把两者的角色互换后，步骤 4 依然成立：若 $3 \mid b$，则 $3 \mid a$。由于 a 和 b 没有大于一的公因数，它们不会同时是 3 的倍数，故由步骤 4，两者都不是 3 的倍数。步骤 2 至 4 并未涵盖这种情况，而这正是步骤 1 派上用场之处。设 $a^2 = 3r + 1$ 及 $b^2 = 3s + 1$。那么 a^4、b^4 和 a^2b^2 都比 3 的某个倍数大一（两个这类数的乘积仍具有此形式）。在模 3 下考察步骤 3 的方程，

$$a^4 + b^4 \equiv 1 + 1 = 2, \qquad 10a^2b^2 \equiv 1 \cdot 1 \cdot 1 = 1 \pmod 3,$$

因为 $10 \equiv 1$。步骤 3 中方程两边除以 3 的余数不同，这是不可能的。所以这样的 a 和 b 并不存在。此外，$\sqrt{2} + \sqrt{3} > 0$，所以若它是有理数，便可写成 a, b 均为正数且互质（coprime）的形式。因此 $\sqrt{2} + \sqrt{3}$ 是无理数（irrational）。

第 (ii) 部分

假设 $\sqrt{6} + \sqrt{7} = a/b$，其中 a, b 为正整数，且没有大于一的公因数。与之前一样，

$$\frac{a^2}{b^2} = 13 + 2\sqrt{42} \quad \Longrightarrow \quad (a^2 - 13b^2)^2 = 168\,b^4,$$

展开得 $a^4 - 26a^2b^2 + 169b^4 = 168b^4$，即

$$a^4 + b^4 = 26a^2b^2.$$

模 5 下的平方数。若 $5 \nmid a$，则 $a = 5k \pm 1$ 或 $a = 5k \pm 2$，从而

$$a^2 = 5(5k^2 \pm 2k) + 1 \qquad 或 \qquad a^2 = 5(5k^2 \pm 4k) + 4.$$

所以 $a^2 \equiv \pm 1 \pmod 5$，因此 $a^4 \equiv 1 \pmod 5$。

整除性。若 $5 \mid a$，则 $b^4 = a^2(26b^2 - a^2)$ 是 5 的倍数。但若 $5 \nmid b$，则 $b^4 \equiv 1 \pmod 5$，所以必有 $5 \mid b$。由对称性，$5 \mid b$ 蕴涵 $5 \mid a$。由于 a 和 b 互质，两者都不能被 5 整除。

矛盾。现在 $a^4 \equiv b^4 \equiv 1$，而 $a^2b^2 \equiv (\pm 1)(\pm 1) = \pm 1 \pmod 5$。由于 $26 \equiv 1$，

$$a^4 + b^4 \equiv 2, \qquad 26a^2b^2 \equiv \pm 1 \pmod 5.$$

但 $2 \not\equiv \pm 1 \pmod 5$，所以方程不可能成立。因此 $\sqrt{6}+\sqrt{7}$ 是无理数。

想一想 若我们在这里尝试用 3 代替 5，哪一步会失效?

为何 3 行不通。论证的前半部分在模 3 下依然成立：若 $3 \mid a$，则 $3 \mid b^4$，故 $3 \mid b$，所以 a 和 b 都不是 3 的倍数。但此时 $a^4 \equiv b^4 \equiv a^2b^2 \equiv 1 \pmod 3$，而由于 $26 \equiv 2 \pmod 3$，

$$a^4 + b^4 \equiv 2 \qquad 及 \qquad 26a^2b^2 \equiv 2 \pmod 3.$$

两边一致，因此没有矛盾。在模 3 下，每一对非 3 的倍数都满足方程 $a^4 + b^4 = 26a^2b^2$，所以单靠被 3 整除的性质无法排除它。（在第 (i) 部分，系数是 $10 \equiv 1$，而 $2 \neq 1$ 造成了冲突。这里系数是 $26 \equiv 2$，所以冲突消失了。）

验算

- 数值上，$x = \sqrt{2}+\sqrt{3} \approx 3.1463$ 使 $x^4 - 10x^2 + 1 \approx 0$（误差约为 10^{-14}）。同样，$y = \sqrt{6}+\sqrt{7} \approx 5.0952$ 使 $y^4 - 26y^2 + 1 \approx 0$。所以这两条四次方程是正确的。
- 逐一尝试每一对余数：$a^4 + b^4 \equiv 10a^2b^2 \pmod 3$ 的唯一解是 $a \equiv b \equiv 0$，而 $a^4 + b^4 \equiv 26a^2b^2 \pmod 5$ 亦然。但 $a^4 + b^4 \equiv 26a^2b^2 \pmod 3$ 还有 $(1,1)$、$(1,2)$、$(2,1)$ 和 $(2,2)$ 这些解。这正是 3 在第 (ii) 部分行不通的原因。
- 电脑搜索显示，不存在满足任一方程且 $a, b \leq 400$ 的互质正整数 a, b。

常见错误

- 在步骤 4 后便停下。步骤 4 只说明 a 和 b 不能是 3 的倍数，你仍需要步骤 1 去处理两者都不是 3 的倍数的情况。
- 未给理由便断言 $3 \mid b^4 \Rightarrow 3 \mid b$。步骤 1 正好提供了理由：若 $3 \nmid b$，则 b^4 的余数是 1。
- 在第 (ii) 部分，忘记了平方数在模 5 下的余数可以是 4（即 -1），所以 a^2b^2 可以 $\equiv \pm 1$，而不只是 1。

STEP 之外

数 $\sqrt{2}+\sqrt{3}$ 是 $x^4 - 10x^2 + 1$ 的根，而这里的论证实际上是在证明这个多项式没有有理根。有理根定理（rational root theorem）提供了一条更快的途径：任何有理根都必须是 ± 1，而两者都不成立。把多项式在模某个质数下约化，借此证明它没有根或因式，是大学一年级代数中的标准工具（例如用于不可约性（irreducibility）检验）。有趣的是，尽管 $x^4 - 10x^2 + 1$ 在有理数上不可约，它在模每一个质数下却都可以因式分解，这是一个著

名的例子。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）进一步发展了这些概念，而 Davenport 的 *The Higher Arithmetic*（《高等算术》）则是一本关于同余（congruences）的精彩入门书。

详细解答

STEP I 2019，第 8 题

纯数学 · 代换积分法

STEP I 2019，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

题目说明不必求出 f 的表达式，所以每一部分都是要把一个积分改造，直至它的样子完全变成

$$\int_1^X \sqrt{\frac{t-1}{t+1}}\,\mathrm{d}t = \mathrm{f}(X)$$

其中 X 是某个积分上限。两个步骤几乎完成全部工作：一是伸缩（stretch）（$u = kt$），把 $\frac{u-k}{u+k}$ 变成 $\frac{t-1}{t+1}$；二是平移（shift）（$u = t + c$），使分式中的两个数关于零对称。最后一部分还需要多一个代换，把一个商的平方根变成形式合适的函数。

解答

全题都用到这个事实：代换（substitution）会同时改变被积函数（integrand）、微分（differential）以及积分限（limits）；答案正正出现在积分上限之中。

第 (i) 部分

若设 $u = 2t$，被积函数 $\sqrt{\frac{u-2}{u+2}}$ 便会变成 $\sqrt{\frac{t-1}{t+1}}$，因为此时 $\frac{2t-2}{2t+2} = \frac{t-1}{t+1}$。由于 $\mathrm{d}u = 2\,\mathrm{d}t$，而积分限为 $u = 2 \mapsto t = 1$、$u = x \mapsto t = \frac{1}{2}x$，

$$\int_2^x \sqrt{\frac{u-2}{u+2}}\,\mathrm{d}u = \int_1^{x/2} \sqrt{\frac{t-1}{t+1}}\,2\,\mathrm{d}t = 2\,\mathrm{f}\left(\tfrac{1}{2}x\right).$$

条件 $x > 2$ 正好保证 $\frac{1}{2}x > 1$，因此 $\mathrm{f}(\frac{1}{2}x)$ 有定义。这就是所求的结果。

第 (ii) 部分

这里分式中的两个数是 0 和 -4，并不关于零对称。它们的中点是 -2，所以我们平移 2：设 $u = v - 2$。于是 $\frac{u}{u+4} = \frac{v-2}{v+2}$，$\mathrm{d}u = \mathrm{d}v$，而积分限 $u = 0, x$ 变成 $v = 2, x+2$。现在以 $x+2$（它大于 2）代替 x 应用第 (i) 部分，便可完成：

$$\int_0^x \sqrt{\frac{u}{u+4}}\,\mathrm{d}u = \int_2^{x+2} \sqrt{\frac{v-2}{v+2}}\,\mathrm{d}v = 2\,\mathrm{f}\big(\tfrac{1}{2}x+1\big).$$

第 (iii) 部分

分式中的两个数是 5 和 -1，中点是 2。所以设 $u = v + 2$，则

$$\frac{u-5}{u+1} = \frac{v-3}{v+3},$$

而积分限 $u = 5, x$ 变成 $v = 3, x-2$。接着伸缩 3 倍，设 $v = 3t$，使得 $\frac{v-3}{v+3} = \frac{t-1}{t+1}$ 及 $\mathrm{d}v = 3\,\mathrm{d}t$，积分限为 $t = 1, \frac{1}{3}(x-2)$：

$$\begin{aligned}\int_5^x \sqrt{\frac{u-5}{u+1}}\,\mathrm{d}u &= \int_3^{x-2} \sqrt{\frac{v-3}{v+3}}\,\mathrm{d}v = \int_1^{(x-2)/3} \sqrt{\frac{t-1}{t+1}}\,3\,\mathrm{d}t \\ &= 3\,\mathrm{f}\Big(\frac{x-2}{3}\Big).\end{aligned}$$

由于 $x > 5$，自变量 $\frac{x-2}{3}$ 大于 1，这正是应有的情况。（两个代换可合并为单一代换 $u = 3t + 2$。）

第 (iv) 部分

这个被积函数中完全看不到商的平方根，而且次方也不对：分子是 u^2，根号内是 $u^2 + 4$。

> **想一想** 甚么代换可以把 $u^2 + 4$ 变成第 (ii) 部分中的 $u + 4$?

设 $u^2 = w$，则 $u = \sqrt{w}$（在 $[1,2]$ 上 $u > 0$），且 $\mathrm{d}u = \dfrac{\mathrm{d}w}{2\sqrt{w}}$。积分限 $u = 1, 2$ 变成 $w = 1, 4$，而

$$\frac{u^2}{\sqrt{u^2+4}}\,\mathrm{d}u = \frac{w}{\sqrt{w+4}} \cdot \frac{\mathrm{d}w}{2\sqrt{w}} = \frac{1}{2} \cdot \frac{\sqrt{w}}{\sqrt{w+4}}\,\mathrm{d}w = \frac{1}{2}\sqrt{\frac{w}{w+4}}\,\mathrm{d}w,$$

这正是第 (ii) 部分的被积函数。在 0 处把区间分拆，并两次应用第 (ii) 部分（分别取 $x = 4$ 和 $x = 1$），

$$\begin{aligned}\int_1^2 \frac{u^2}{\sqrt{u^2+4}}\,\mathrm{d}u &= \frac{1}{2}\left(\int_0^4 \sqrt{\frac{w}{w+4}}\,\mathrm{d}w - \int_0^1 \sqrt{\frac{w}{w+4}}\,\mathrm{d}w\right) \\ &= \frac{1}{2}\Big(2\,\mathrm{f}(3) - 2\,\mathrm{f}\big(\tfrac{3}{2}\big)\Big) = \mathrm{f}(3) - \mathrm{f}\big(\tfrac{3}{2}\big).\end{aligned}$$

验算

- 每个答案在积分下限处的表现都正确：在 (i) 中取 $x = 2$，得 $2\,\mathrm{f}(1) = 0$；在 (ii) 中取 $x = 0$，得 $2\,\mathrm{f}(1) = 0$；在 (iii) 中取 $x = 5$，得 $3\,\mathrm{f}(1) = 0$。
- 求导是快捷的检验方法。对于 (iii)，$\frac{\mathrm{d}}{\mathrm{d}x}\,3\,\mathrm{f}\left(\frac{x-2}{3}\right) = \mathrm{f}'\left(\frac{x-2}{3}\right) = \sqrt{\frac{x-5}{x+1}}$，与预期相符。
- 以数值计算，(iv) 两边都等于 0.91007；而在 (iii) 中取 $x = 11$，两边都等于 3.19704，两者均由直接数值积分（numerical integration）求得。

常见错误

- 忘记 $\mathrm{d}u = k\,\mathrm{d}t$ 带来的因子，结果得出 $\mathrm{f}(\frac{1}{2}x)$ 而非 $2\,\mathrm{f}(\frac{1}{2}x)$。
- 改了被积函数却没有改积分限，尤其是上限——而答案正正就在上限之中。
- 在 (iv) 中只写出 $\frac{1}{2} \cdot 2\,\mathrm{f}(3)$：第 (ii) 部分给出的是由 0 开始的积分，所以必须减去由 0 到 1 的那一段。

STEP 之外

这里的核心想法是：一整族积分其实只是同一个积分，通过不同的仿射变量代换（affine change of variable）$u = at + b$ 去看而已。辨认出隐藏在表达式中的“标准形式”，是大学分析学以及把微分方程（differential equations）化为标准型（canonical forms）时经常用到的技巧。（事实上，利用双曲函数（hyperbolic functions）可以求出 f 的闭式（closed form），但这道题显示，即使不用它也可以走得很远。）关于代换与积分的严谨理论，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）；如想找更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP I 2019，第 9 题

力学 · 倾倒与摩擦力

STEP I 2019，第 9 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

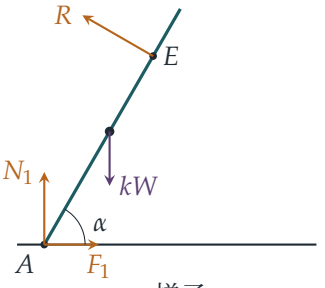
www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

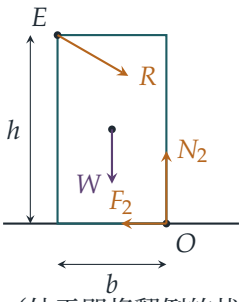
把梯子和箱子看作两个独立的物体，两者之间只靠一个力相连。由于接触处光滑（smooth），这个力垂直于梯子，所以它的作用线由 α 决定，只有大小是未知的。对梯子以其底端取力矩（moments），可求出这个力的大小；对箱子以它将要翻倒（topple）所绕的那条棱取力矩，便得到翻倒的条件。关系式 $h = b\tan\alpha$ 的作用，是让箱子的力矩化简成一个倍角表达式。

解答

把箱子视为均匀的刚性长方体（rigid cuboid），梯子视为轻杆（light rod），油漆工视为质点（particle）。梯子与箱子之间的接触是光滑的；两处与地面的接触都是粗糙（rough）的。设 E 为梯子所接触的那条棱，A 为梯子的底端，O 为箱子底部离梯子最远的那条棱（即箱子翻倒时所绕的棱）。



梯子



箱子（处于即将翻倒的状态）

第 (i) 部分

由于 E 处没有摩擦，箱子对梯子的作用力 R 垂直于梯子。因为 E 位于高度 h，沿梯子量度有 $AE = h/\sin\alpha$。当油漆工位于高度 λh 时，其沿梯子与 A 的距离为 $\lambda h/\sin\alpha$，所以与 A 的水平距离为 $\lambda h\cot\alpha$。

梯子是轻的，所以对 A 的力矩只来自 R 和油漆工的重量（N_1 和 F_1 都作用于 A）。由于 R 垂直于 AE，它的力臂就是 AE 本身：

$$R\cdot\frac{h}{\sin\alpha} = kW\cdot\lambda h\,\frac{\cos\alpha}{\sin\alpha} \quad\Longrightarrow\quad R = k\lambda W\cos\alpha,$$

即为所求。（严格来说，这在攀爬的每一阶段都成立，其中 λh 是油漆工当时的高度。）

第 (ii) 部分

根据牛顿第三定律（Newton's third law），梯子在 E 处以垂直于梯子、指向箱子内部的力 R 推动箱子。它的分量是水平方向的 $R\sin\alpha$（背离梯子，指向 O）及竖直向下的 $R\cos\alpha$。

当箱子即将绕 O 翻倒时，地面的反作用力完全作用于 O，所以 N_2 和 F_2 对 O 没有力矩。对 O 取力矩，以顺时针方向（即箱子翻倒的方向）为正。E 与 O 的水平距离为 b，并位于 O 上方高度 h 处，所以

- 水平分量 $R\sin\alpha$ 作用于高度 h，其力矩为 $+Rh\sin\alpha$；
- 竖直分量 $R\cos\alpha$ 作用于 O 靠梯子一侧、距离为 b 处，其力矩为 $-Rb\cos\alpha$（在那一侧向下压会阻止翻倒）；
- 重量 W 作用于与 O 距离 $\frac{1}{2}b$ 处，其力矩为 $-\frac{1}{2}Wb$。

当总力矩达到零时，箱子开始翻倒：

$$R(h\sin\alpha - b\cos\alpha) = \tfrac{1}{2}Wb.$$

> **想一想** $h = b\tan\alpha$ 在这里有甚么帮助?

代入后得

$$h\sin\alpha - b\cos\alpha = b\,\frac{\sin^2\alpha - \cos^2\alpha}{\cos\alpha} = -\frac{b\cos 2\alpha}{\cos\alpha}.$$

利用第 (i) 部分的 $R = k\lambda W\cos\alpha$，条件变为

$$-k\lambda W\cos\alpha\cdot\frac{b\cos 2\alpha}{\cos\alpha} = \tfrac{1}{2}Wb \quad\Longrightarrow\quad 2k\lambda\cos 2\alpha + 1 = 0,$$

即为所求。注意这要求 $\cos 2\alpha < 0$，即 $\alpha > 45°$，这与 $h > b$ 一致。

第 (iii) 部分

对箱子分解力（地面对它的反作用力有指向梯子的分量 F_2 和向上的分量 N_2）：

$$F_2 = R\sin\alpha, \qquad N_2 = W + R\cos\alpha.$$

箱子不滑动，所以 $F_2 \le \mu N_2$，即 $\mu \ge \dfrac{R\sin\alpha}{W + R\cos\alpha}$。

> **想一想** 在攀爬过程中的哪一刻，这个条件最难满足？

比值 $\dfrac{R\sin\alpha}{W + R\cos\alpha} = \dfrac{\sin\alpha}{W/R + \cos\alpha}$ 随 R 递增，而 R 随油漆工往上爬而增大。所以对摩擦力要求最大的时刻就是即将翻倒的一刻，我们需要不等式在该处成立。由 (ii)，$k\lambda = -\dfrac{1}{2\cos 2\alpha}$，所以

$$R = -\frac{W\cos\alpha}{2\cos 2\alpha}.$$

于是

$$R\sin\alpha = -\frac{W\sin\alpha\cos\alpha}{2\cos 2\alpha} = -\frac{W\sin 2\alpha}{4\cos 2\alpha},$$

$$W + R\cos\alpha = W\left(1 - \frac{\cos^2\alpha}{2\cos 2\alpha}\right) = W\,\frac{4\cos 2\alpha - (1+\cos 2\alpha)}{4\cos 2\alpha} = W\,\frac{3\cos 2\alpha - 1}{4\cos 2\alpha},$$

这里用了 $2\cos^2\alpha = 1 + \cos 2\alpha$。两式相除，得

$$\mu \ge \frac{-\sin 2\alpha}{3\cos 2\alpha - 1} = \frac{\sin 2\alpha}{1 - 3\cos 2\alpha},$$

即为所求。由于 $45^\circ < \alpha < 90^\circ$ 给出 $\sin 2\alpha > 0$ 及 $\cos 2\alpha < 0$，分子和分母都是正数。

验算

- 取 $\alpha = 60^\circ$，则 $\cos 2\alpha = -\frac{1}{2}$。第 (ii) 部分给出 $k\lambda = 1$，所以 $R = \frac{1}{2}W$。于是 $F_2 = \frac{\sqrt{3}}{4}W$ 及 $N_2 = \frac{5}{4}W$，得 $\mu \ge \frac{\sqrt{3}}{5} \approx 0.346$，与 $\sin 120^\circ / (1 - 3\cos 120^\circ) = \frac{\sqrt{3}}{2} / \frac{5}{2}$ 吻合。
- 当 $\alpha \to 45^\circ$（箱子接近正方体，高度只是略大于宽度），$\cos 2\alpha \to 0$，而 (ii) 要求 $k\lambda \to \infty$：E 处的推力几乎沿着直线 EO，所以几乎没有转动效果。这与几何吻合：由于 $h = b\tan\alpha$，当 α 接近 45° 时，通过 E 的 R 的作用线差不多经过 O。
- 力矩方程、翻倒高度和摩擦力的界限，亦已用电脑就 α、b、k 和 W 的多组随机取值验证。

常见错误

- 把 E 处的力当作水平方向（就像光滑墙壁的情况）。这里的接触是与一条棱接触，所以力垂直于梯子。

- 忘记 R 的竖直分量会加到箱子的法向反作用力（normal reaction）上，因此 $N_2 = W + R\cos\alpha$，而不是 W。
- 对箱子底部错误的那条棱取力矩：箱子是绕离梯子较远的那条棱翻倒的。

STEP 之外

这是双物体静力学（statics）：把每个物体分开考虑，在接触处运用牛顿第三定律，并选取能消去未知量的力矩中心。物体会先滑动还是先翻倒、随着负载增加哪个约束会最先失效，这些问题贯穿工程力学以及接触刚体的研究。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）正是以这种精神处理刚体静力学与摩擦。

详细解答

STEP I 2019，第 10 题

力学 • 抛体运动

STEP I 2019，第 10 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

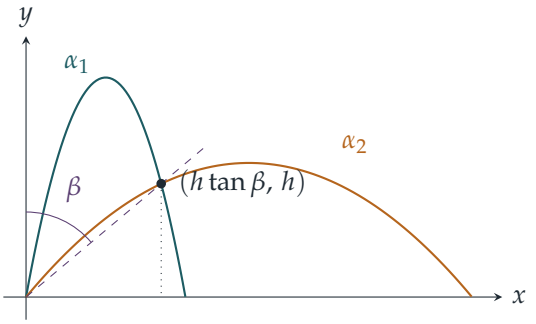
www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

先写出以 x 表示 y 的轨迹方程（trajectory equation），代入那个定点，然后留意 $1/\sin^2\alpha = 1+\cot^2\alpha$ 可把方程化为关于 $c=\cot\alpha$ 的二次方程（quadratic）。之后，(i) 的一切都只是二次方程的代数：根之和与根之积（sum and product of roots）配合 cot 的和角公式（addition formula）就能得出 $\alpha_1+\alpha_2$，而“两个相异的根”则是一个判别式（discriminant）条件。第 (ii) 部分只需用一行能量论证（或 $v^2=u^2+2as$）便可完成。

解答

把粒子视为在重力作用下自由运动的质点（point mass），不计空气阻力。角 α 从向上的铅垂线量起，因此初速度的水平分量为 $u\sin\alpha$，铅垂分量为 $u\cos\alpha$。



上图显示经过该点的两条可能轨迹（以 $\beta=50°$、$k=8$ 绘制）。该点位于经过 O、与铅垂线成 β 角的直线上。

第 (i) 部分

在时间 t，

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

由于 $0 < \beta < 90°$，该点的 $x > 0$，所以 $\sin\alpha \neq 0$，于是可以消去 $t = x/(u\sin\alpha)$：

$$y = x\cot\alpha - \frac{gx^2}{2u^2\sin^2\alpha} = cx - \frac{gx^2}{2u^2}(1+c^2),$$

这里用了 $\csc^2\alpha = 1 + \cot^2\alpha$。现在代入 $x = h\tan\beta$、$y = h$，并除以 $h > 0$：

$$1 = c\tan\beta - \frac{gh}{2u^2}\tan^2\beta\,(1+c^2) = c\tan\beta - \frac{\tan^2\beta}{k}(1+c^2).$$

两边乘以 $k\cot^2\beta$，得 $k\cot^2\beta = ck\cot\beta - 1 - c^2$，即

$$c^2 - ck\cot\beta + 1 + k\cot^2\beta = 0, \qquad (*)$$

此即所求。

(a) 两个值 $c_1 = \cot\alpha_1$ 和 $c_2 = \cot\alpha_2$ 就是关于 c 的二次方程 $(*)$ 的两个根。（它们相异，因为 cot 在 $(0,\pi)$ 上是一一对应（one-to-one）的。）因此，由根之和与根之积，

$$\cot\alpha_1 + \cot\alpha_2 = k\cot\beta, \qquad \cot\alpha_1\cot\alpha_2 = 1 + k\cot^2\beta.$$

第一式就是第一个结果。至于第二个结果，利用 cot 的和角公式：

$$\cot(\alpha_1+\alpha_2) = \frac{\cot\alpha_1\cot\alpha_2 - 1}{\cot\alpha_1 + \cot\alpha_2} = \frac{k\cot^2\beta}{k\cot\beta} = \cot\beta.$$

? **想一想** $\cot(\alpha_1+\alpha_2) = \cot\beta$ 真的能推出 $\alpha_1 + \alpha_2 = \beta$ 吗?

单靠这一点并不能：cot 的周期（period）是 180°，所以我们还需要确定 $\alpha_1+\alpha_2$ 的范围。这里根之和 $k\cot\beta$ 与根之积 $1 + k\cot^2\beta$ 都是正数（因为 $k > 0$ 且 $\cot\beta > 0$），所以两个根 c_1, c_2 都是正数。因此两个角都是锐角，$0 < \alpha_1 + \alpha_2 < 180°$，而 cot 在这个区间上是一一对应的。由于 β 也在这个区间内，所以 $\alpha_1 + \alpha_2 = \beta$。

(b) 要有两个相异的实数值 c，$(*)$ 的判别式必须为正：

$$k^2\cot^2\beta - 4(1 + k\cot^2\beta) > 0.$$

两边乘以 $\tan^2\beta > 0$，

$$k^2 - 4k - 4\tan^2\beta > 0.$$

左边关于 k 的二次式的根为 $k = 2 \pm \sqrt{4 + 4\tan^2\beta} = 2 \pm 2\sec\beta$（利用 $1 + \tan^2\beta = \sec^2\beta$ 及 $\sec\beta > 0$），所以不等式成立当且仅当 $k < 2 - 2\sec\beta$ 或 $k > 2 + 2\sec\beta$。但 $\sec\beta > 1$，所以 $2 - 2\sec\beta < 0 < k$，第一种情况不可能出现。因此

$$k > 2(1 + \sec\beta).$$

第 (ii) 部分

粒子必须升到高度 $h > 0$，所以其初始铅垂速度 $u\cos\alpha$ 为正。在铅垂方向上，运动的加速度恒为 $-g$，而在最高点时铅垂速度为零，所以由 $v^2 = u^2 + 2as$，最大高度为

$$H = \frac{u^2\cos^2\alpha}{2g}.$$

粒子经过高度为 h 的一点，所以 $H \geq h$：

$$\frac{u^2\cos^2\alpha}{2g} \geq h \quad\Longrightarrow\quad \frac{2u^2}{gh}\cos^2\alpha \geq 4 \quad\Longrightarrow\quad k \geq 4\sec^2\alpha,$$

其中先乘以 $4/(gh) > 0$，再除以 $\cos^2\alpha > 0$。等号成立当且仅当点 $(h\tan\beta, h)$ 正是轨迹的最高点。

验算

- 取 $\beta = 50°$、$k = 8$。此时 $2(1+\sec 50°) \approx 5.11 < 8$，所以存在两个角；解 $(*)$ 得 $\alpha_1 \approx 10.29°$ 及 $\alpha_2 \approx 39.71°$，两者之和为 $50°$。以数值方法逐步推进运动，两条轨迹在 $x = h\tan\beta$ 处都达到高度 $1.0000\,h$。
- 对这两个角，$k\cos^2\alpha \approx 7.74$ 及 4.73，都不小于 4，符合 (ii) 的要求。
- 以电脑对数千组随机的 β、h、k 进行检验，证实 $(*)$ 有两个相异根当且仅当 $k > 2(1+\sec\beta)$，并且每次都有 $\alpha_1 + \alpha_2 = \beta$。
- 单位：$u^2/(gh)$ 是无量纲（dimensionless）的，正如 k 应有的那样。

常见错误

- 出于习惯，从水平线量度 α：本题中水平分量是 $u\sin\alpha$，而不是 $u\cos\alpha$。
- 未证明两个角都是锐角，便由 $\cot(\alpha_1+\alpha_2) = \cot\beta$ 得出 $\alpha_1 + \alpha_2 = \beta$。
- 在 (b) 中保留了负数的根 $k < 2 - 2\sec\beta$。

STEP 之外

结果 $\alpha_1 + \alpha_2 = \beta$ 说明两个发射方向关于铅垂线与目标连线之间的角平分线（bisector）对称，这是关于“安全抛物线”（parabola of safety）的一个经典事实：安全抛物线是所有速率为 u 的轨迹的包络线（envelope），对应于两根重合的临界情况 $k = 2(1+\sec\beta)$。曲线族的包络线是大学一年级一个连系微积分与几何的有趣课题。如想阅读更多同类风格的力学，可参考 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP I 2019，第 11 题

概率与统计 · 几何分布

STEP I 2019，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP I 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 12.9 节用不公平的硬币作公平的决定中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/von-neumann.html

关键思路

两人程序的每一轮都是一次独立（independent）试验，要么作出决定，要么不作出决定，因此作出决定的轮次服从几何分布（geometric distribution）。题目中“无论 p 取何值”这个界限，归结为在 $p+q=1$ 的条件下 pq 最大可以是多少，而算术–几何平均不等式（AM–GM inequality）正好回答了这个问题。在第 (ii) 部分，第二轮不再是重新开始，所以我们按第一轮的结果分情况讨论，再用同一个关于 pq 的界限完成证明。

解答

第 (i) 部分

在任何一轮中，两枚硬币是独立的。当两枚都是正面（概率 p^2）或两枚都是反面（概率 q^2）时便作出决定，所以

$$\mathrm{P}(\text{在某一轮作出决定}) = p^2 + q^2, \qquad \mathrm{P}(\text{未作出决定}) = pq + qp = 2pq.$$

验证一下：$p^2 + q^2 + 2pq = (p+q)^2 = 1$。

由于每一轮都重新掷硬币，各轮之间是独立的。在第 n 轮作出决定，即是先有 $n-1$ 轮未作出决定，接着有一轮作出决定，所以

$$\mathrm{P}(\text{作出决定的轮次为}n) = (2qp)^{n-1}(q^2 + p^2),$$

这就是所求的结果。

至于第二个结果，用补事件（complement）最为简便。在第 n 轮或之前仍未作出决定，即是连续 n 轮都未作出决定，其概率为 $(2pq)^n$。因此

$$\mathrm{P}(\text{作出决定的轮次不迟于}n) = 1 - (2pq)^n.$$

（利用 $p^2+q^2 = 1-2pq$，把几何级数（geometric series）$\sum_{r=1}^{n}(p^2+q^2)(2pq)^{r-1}$ 求和，也会得到相同结果。）

现在我们需要一个对所有 p 都成立的 pq 上界（upper bound）。由于 $p, q \geq 0$，算术–几何平均不等式给出

$$\sqrt{pq} \leq \frac{p+q}{2} = \frac{1}{2}, \qquad \text{所以} \qquad pq \leq \frac{1}{4},$$

等号成立当且仅当 $p = q = \frac{1}{2}$。（等价地，$pq = p(1-p) = \frac{1}{4} - (p-\frac{1}{2})^2$。）因此 $0 \leq 2pq \leq \frac{1}{2}$，所以 $(2pq)^n \leq \frac{1}{2^n}$，从而

$$\mathrm{P}(\text{作出决定的轮次不迟于}n) = 1 - (2pq)^n \geq 1 - \frac{1}{2^n}$$

对 p 的每一个值都成立，正如所求。公平硬币是最差的情况。

第 (ii) 部分

设 D_1 为在第 1 轮作出决定的事件，D_2 为在第 2 轮作出决定的事件。由于这两个事件互斥（mutually exclusive），我们需要求的是 $\mathrm{P}(D_1) + \mathrm{P}(D_2)$。

第 1 轮。三枚全是正面或全是反面：$\mathrm{P}(D_1) = p^3 + q^3$。

第 2 轮。若第 1 轮未作出决定，则硬币显示的是两正一反，或两反一正。

- **两正一反**（概率 $3p^2q$，因为三人中任何一人都可以是掷出反面的那位）。掷出正面的两人再掷一次，第三人则把反面的硬币翻转成正面。此时不可能全是反面，而全是正面当且仅当两次新的投掷都是正面，概率为 p^2。这种情况贡献 $3p^2q \cdot p^2 = 3p^4q$。
- **两反一正**（概率 $3pq^2$）。同理，第三枚硬币此时显示反面，我们需要两次新的投掷都是反面，概率为 q^2。这种情况贡献 $3pq^2 \cdot q^2 = 3pq^4$。

第 2 轮的投掷与第 1 轮独立，所以可以把概率相乘。于是

$$\mathrm{P}(D_2) = 3p^4q + 3pq^4 = 3pq\,(p^3+q^3).$$

> **想一想** 有没有简洁的方法整理 $p^3+q^3+3pq(p^3+q^3)$?

先抽出公因式 p^3+q^3，再利用立方和公式及 $p+q=1$：

$$p^3+q^3 = (p+q)(p^2-pq+q^2) = (p+q)^2 - 3pq = 1-3pq.$$

因此

$$\begin{aligned}\mathrm{P}(\text{作出决定的轮次不迟于 } 2) &= (p^3+q^3)(1+3pq)\\ &= (1-3pq)(1+3pq)\\ &= 1-9p^2q^2.\end{aligned}$$

由第 (i) 部分，$0 \le pq \le \frac{1}{4}$，所以 $p^2q^2 \le \frac{1}{16}$，从而

$$\mathrm{P}(\text{作出决定的轮次不迟于 } 2) \ge 1-\frac{9}{16} = \frac{7}{16}$$

无论 p 取何值都成立，且当 $p=\frac{1}{2}$ 时等号成立。证明完毕。

验算

- 当 $p=1$（两面都是正面的硬币）时，立即作出决定。在第 (i) 部分，$1-(2pq)^n = 1$；在第 (ii) 部分，$1-9p^2q^2=1$，与预期相符。
- 在第 (ii) 部分取 $p=\frac{1}{2}$：第 1 轮作出决定的概率为 $\frac{1}{4}$。否则（概率 $\frac{3}{4}$）第 2 轮作出决定的概率为 $\frac{1}{4}$，得 $\frac{1}{4}+\frac{3}{16}=\frac{7}{16}$，所以这个界限是可以达到的。
- 取 $p=0.3$，公式给出 $1-9(0.21)^2 = 0.6031$。用电脑模拟三人程序 200 000 次，得到 0.602。

常见错误

- 在第 (ii) 部分第 1 轮中，忘记乘上表示谁是与众不同那一位的因子 3。
- 把第 (ii) 部分的第 2 轮当作全新的三枚硬币一轮来处理。被翻转的硬币已经固定，所以“全是正面”和“全是反面”只有其中一个可能发生。
- 只在 $p=\frac{1}{2}$ 时证明这个界限。你需要对每一个 p 都有 $pq \le \frac{1}{4}$，而算术–几何平均不等式这类不等式正好给你这个结果。

STEP 之外

第 (i) 部分是一个经典技巧，通常归功于冯·诺伊曼（von Neumann），用来从一枚有偏差的硬币得出公平的决定。在两人程序中，只有当 $p=\frac{1}{2}$ 时，“酒店”和“茶室”的可能性才相等；但如果由一人掷两次，并以 HT 和 TH 作为结果，那么对每一个 p 结果都是公平的。作出决定前所需的轮数是一个几何随机变量（geometric random variable），而关于这类程序多快停止的问题，引出了停时（stopping times）和马可夫链（Markov chains）

的研究。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）对两者都有介绍。

详细解答

STEP II 2019，第 1 题

纯数学 · 三次曲线的切线

STEP II 2019，第 1 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

先写出 $x = a$ 处的切线 (tangent) 通过 $(p, 0)$ 的条件，再代入 $f(x) = (x - p)g(x)$：几乎所有项都互相抵消，只剩下 $g'(a)$ 的某个倍数。因此，对于三次函数 (cubic) 而言，从根 p 出发的切线与曲线相切之处，正是另一个二次因式 (quadratic factor) 取驻值 (stationary) 之处，即另外两个根的正中间。第 (ii) 部分则归结为比较两个斜率 (gradient)，以及把“切线与三次曲线交点”的方程因式分解 (factorise)；这个方程在接触点 (point of contact) 处必定有一个重根 (double root)。

解答

开首的结果

$x = a$ 处的切线为 $y = f(a) + f'(a)(x - a)$。由乘积法则 (product rule)，$f'(x) = g(x) + (x - p)g'(x)$。切线通过 $(p, 0)$ 当且仅当

$$\begin{aligned} 0 &= f(a) + f'(a)(p - a) \\ &= (a - p)g(a) + \big[g(a) + (a - p)g'(a)\big](p - a) \\ &= (a - p)g(a) - (a - p)g(a) - (a - p)^2 g'(a) \\ &= -(a - p)^2 g'(a). \end{aligned}$$

由于 $a \neq p$，因子 $(a - p)^2$ 不为零，所以上式成立当且仅当 $g'(a) = 0$。这就一次过证明了结果的两个方向。

第 (i) 部分

此处 $f(x)=(x-p)g(x)$，其中 $g(x)=A(x-q)(x-r)=A\big(x^2-(q+r)x+qr\big)$，所以 $g'(x)=A\big(2x-(q+r)\big)$。由开首的结果，并因为 $A\neq 0$，

$$g'(a)=0 \iff 2a=q+r.$$

（注意 $a=\frac{1}{2}(q+r)>q>p$，所以 $a\neq p$，符合要求。）

至于斜率，由于 $g'(a)=0$，有 $f'(a)=g(a)+(a-p)g'(a)=g(a)$。利用 $a-q=\frac{1}{2}(r-q)$ 及 $a-r=-\frac{1}{2}(r-q)$，

$$f'(a)=A\cdot\tfrac{1}{2}(r-q)\cdot\big(-\tfrac{1}{2}(r-q)\big)=-\frac{A(r-q)^2}{4}.$$

第 (ii) 部分

现在改为抽出根 r：$f(x)=(x-r)h(x)$，其中 $h(x)=A(x-p)(x-q)$。开首的结果以 r 代替 p 后同样适用（推导中完全没有用到选了哪一个根），所以 $x=c$ 处的切线通过 $(r,0)$ 当且仅当 $h'(c)=0$，即 $c=\frac{1}{2}(p+q)$。与第 (i) 部分完全一样，其斜率为

$$f'(c)=h(c)=A\cdot\tfrac{1}{2}(q-p)\cdot\big(-\tfrac{1}{2}(q-p)\big)=-\frac{A(q-p)^2}{4}.$$

因此，由于 $A\neq 0$，

$$\text{两切线平行} \iff (r-q)^2=(q-p)^2 \iff r-q=q-p \iff q=\tfrac{1}{2}(p+r),$$

其中中间一步用到 $r-q>0$ 及 $q-p>0$，所以可以取正平方根。

> **想一想** 不解三次方程，怎样找出切线与三次曲线的另一个交点？

我们其实已经知道一个根，而且是重复两次的：接触点。在 $x=q$ 处，斜率为 $f'(q)=A(q-p)(q-r)$（乘积法则中另外两项因含有 $x-q$ 而消失），且 $f(q)=0$，所以该处的切线为 $y=A(q-p)(q-r)(x-q)$。它与曲线相交于

$$A(x-p)(x-q)(x-r)=A(q-p)(q-r)(x-q).$$

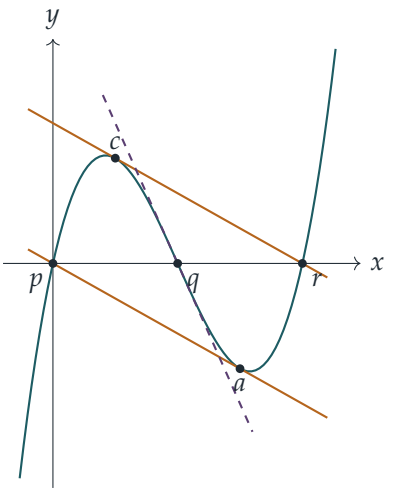
两边除以 $A\neq 0$ 并抽出因子 $x-q$，

$$\begin{aligned}0&=(x-q)\big[(x-p)(x-r)-(q-p)(q-r)\big]\\&=(x-q)\big[x^2-q^2-(p+r)(x-q)\big]\\&=(x-q)^2\,(x+q-p-r).\end{aligned}$$

所以切线与曲线相交于 $x=q$（一个重根，正如接触点所应有的）及 $x=p+r-q$，除此以外再无其他交点。切线不会再与曲线相交，当且仅当这第三个根与 q 重合：

$$p+r-q=q \iff q=\tfrac{1}{2}(p+r).$$

这与之前的条件相同，所以第 (i) 及第 (ii) 部分的两条切线互相平行，当且仅当 $x = q$ 处的切线不再与曲线相交。



草图显示 $q = \frac{1}{2}(p + r)$ 的情况：两条切线（琥珀色）互相平行，而 q 处的切线（虚线）在该点穿过曲线，却不在其他任何地方与曲线相交，因为此时 q 正是拐点（point of inflection）。

验算

- 取 $p = 0$、$q = 1$、$r = 2$、$A = 3$，则 $f(x) = 3x^3 - 9x^2 + 6x$。于是 $f'(x) = 9x^2 - 18x + 6$，得 $f'(\frac{3}{2}) = -\frac{3}{4}$ 及 $f'(\frac{1}{2}) = -\frac{3}{4}$，两者都等于 $-A(r-q)^2/4 = -A(q-p)^2/4 = -\frac{3}{4}$，所以两切线平行；这与预测相符，因为 q 是 p 与 r 的中点（midpoint）。
- $x = \frac{3}{2}$ 处的切线为 $y = -\frac{9}{8} - \frac{3}{4}(x - \frac{3}{2})$，在 $x = 0$ 时其值为 0：它确实通过 $(p, 0)$。
- 以上所有恒等式亦已用电脑以精确算术，就大约二千组随机选取的 $p < q < r$ 及 A 验证过。

常见错误

- 忘记正是 $a \neq p$ 才容许我们约去 $(a-p)^2$；没有这个条件，“仅当”方向便不成立，因为 p 处本身的切线必定通过 $(p, 0)$。
- 写出 $(r-q)^2 = (q-p)^2 \Rightarrow r - q = q - p$，却没有说明为何排除负根。
- 尝试从头解三次方程求交点，而不利用接触点处已知的重根。

STEP 之外

$x = q$ 处的切线之所以特别，是因为此时 q 是拐点，亦即三次曲线的对称中心（centre of symmetry）：每条三次曲线绕其拐点作半周旋转（half-turn）后都保持不变，这正是两条切线变得平行的原因。这里的切线与弦（chord）构作，是迈向三次曲线几何的第一步；在那里，直线与三次曲线的第三个交点被用来定义曲线上点的加法（椭圆曲线（elliptic curves））。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）有更多同类风格的题目。

详细解答

STEP II 2019，第 2 题

纯数学 · 反函数的积分

STEP II 2019，第 2 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

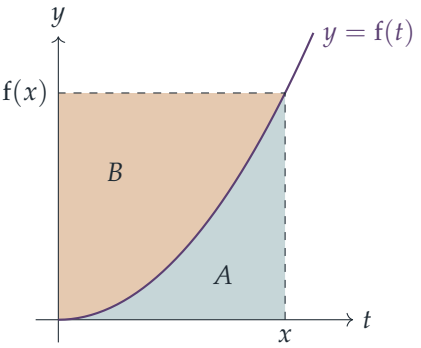
关键思路

$y = \mathrm{f}(t)$ 的图像把以 $(0,0)$ 和 $(x, \mathrm{f}(x))$ 为顶点的长方形分成两块。其中一块是曲线下方的面积，另一块是曲线与 y 轴之间的面积，而后者正是反函数（inverse function）的积分。因此两个积分之和等于长方形的面积。第 (i) 及 (ii) 部分的函数都只是隐式地定义，我们无法直接把它们积分，但它们的反函数却是简单的三次函数。这条恒等式让我们把一个困难的积分换成一个容易的积分。

解答

开首的结果

由于当 $t > 0$ 时 $\mathrm{f}'(t) > 0$，而且 $\mathrm{f}(0) = 0$，所以 f 在 $[0, x]$ 上严格递增（strictly increasing），因此当 $t > 0$ 时 $\mathrm{f}(t) > 0$，而 f 有一个由 $[0, \mathrm{f}(x)]$ 到 $[0, x]$ 的反函数 f^{-1}。同一条曲线既可以用 $y = \mathrm{f}(t)$ 描述，也可以用 $t = \mathrm{f}^{-1}(y)$ 描述。



区域 A 是在 $0 \le t \le x$ 范围内曲线与 t 轴之间的部分，其面积为

$$A = \int_0^x \mathrm{f}(t)\,\mathrm{d}t.$$

区域 B 是在 $0 \le y \le \mathrm{f}(x)$ 范围内曲线与 y 轴之间的部分，由一条条水平窄条组成。位于高度 y 的窄条由 $t = 0$ 延伸至 $t = \mathrm{f}^{-1}(y)$，因此

$$B = \int_0^{\mathrm{f}(x)} \mathrm{f}^{-1}(y)\,\mathrm{d}y.$$

由于曲线递增，并由 $(0,0)$ 走到 $(x,\mathrm{f}(x))$，A 与 B 恰好拼合在一起，互不重叠，填满长方形 $0 \le t \le x$，$0 \le y \le \mathrm{f}(x)$。该长方形的面积是 $x\,\mathrm{f}(x)$，因此

$$\int_0^x \mathrm{f}(t)\,\mathrm{d}t + \int_0^{\mathrm{f}(x)} \mathrm{f}^{-1}(y)\,\mathrm{d}y = x\,\mathrm{f}(x),$$

即为所求。

第 (i) 部分

设 $\phi(y) = y^3 + y$。那么 $\phi'(y) = 3y^2 + 1 > 0$，所以 ϕ 严格递增，而且当 $y \to \pm\infty$ 时 $\phi(y) \to \pm\infty$。因此对每个实数 t，恰好有一个实数 y 使 $\phi(y) = t$，这个 y 就是 $\mathrm{g}(t)$。换句话说，$\mathrm{g} = \phi^{-1}$。

$\mathrm{g}(0) = 0$。代入 $t = 0$，得 $\mathrm{g}(0)\big(\mathrm{g}(0)^2 + 1\big) = 0$。由于 $\mathrm{g}(0)$ 是实数，$\mathrm{g}(0)^2 + 1 \ge 1$，故此 $\mathrm{g}(0) = 0$。

$\mathrm{g}'(t) > 0$。把 $\mathrm{g}(t)^3 + \mathrm{g}(t) = t$ 对 t 作隐函数求导（implicit differentiation），

$$\big(3\,\mathrm{g}(t)^2 + 1\big)\,\mathrm{g}'(t) = 1 \quad\Longrightarrow\quad \mathrm{g}'(t) = \frac{1}{3\,\mathrm{g}(t)^2 + 1} > 0,$$

因为分母至少是 1。

> **想一想** 我们写不出 $\mathrm{g}(t)$ 的表达式，那怎样才能把它积分?

我们根本不需要写出来。g 满足开首结果的条件，而它的反函数是 $\phi(y) = y^3 + y$，很容易积分。以 $x = 2$ 应用该结果时需要知道 $\mathrm{g}(2)$，即 $y^3 + y = 2$ 的实根。这个根是 $y = 1$，而且由于 ϕ 严格递增，它是唯一的实根。因此

$$\int_0^2 \mathrm{g}(t)\,\mathrm{d}t = 2\,\mathrm{g}(2) - \int_0^1 (y^3 + y)\,\mathrm{d}y = 2 - \left(\tfrac{1}{4} + \tfrac{1}{2}\right) = \frac{5}{4}.$$

第 (ii) 部分

比较两个定义，$\mathrm{h}(t)$ 是 $y^3 + y = t + 2$ 的实数解，所以 $\mathrm{h}(t) = \mathrm{g}(t+2)$。代入 $u = t + 2$，

$$\int_0^8 \mathrm{h}(t)\,\mathrm{d}t = \int_2^{10} \mathrm{g}(u)\,\mathrm{d}u = \int_0^{10} \mathrm{g}(u)\,\mathrm{d}u - \int_0^2 \mathrm{g}(u)\,\mathrm{d}u.$$

对于第一个积分，我们以 $x = 10$ 应用开首的结果。由于 $2^3 + 2 = 10$，我们有 $\mathrm{g}(10) = 2$，因此

$$\int_0^{10} \mathrm{g}(u)\,\mathrm{d}u = 10 \times 2 - \int_0^2 (y^3 + y)\,\mathrm{d}y = 20 - (4 + 2) = 14.$$

第二个积分则用第 (i) 部分的结果，

$$\int_0^8 \mathrm{h}(t)\,\mathrm{d}t = 14 - \frac{5}{4} = \frac{51}{4}.$$

另一种方法。我们也可以直接对 h 应用这幅图。这次 $\mathrm{h}(0) = 1$，而不是 0。曲线由 $(0,1)$ 走到 $(8,2)$，而 $\mathrm{h}^{-1}(y) = y^3 + y - 2$。曲线下方的面积等于 8×2 的长方形，减去 0×1 的长方形（这里是空的），再减去在 $1 \le y \le 2$ 范围内曲线左方的面积：

$$16 - \int_1^2 (y^3 + y - 2)\,\mathrm{d}y = 16 - \left(\tfrac{15}{4} + \tfrac{3}{2} - 2\right) = 16 - \tfrac{13}{4} = \tfrac{51}{4}.$$

验算

- 在 $[0,2]$ 上 $0 \le \mathrm{g}(t) \le 1$，所以 (i) 的积分必定介乎 0 与 2 之间。此外，g 在该区间上是凹函数（concave）（它是一个递增凸函数（convex function）的反函数），所以积分至少等于三角形面积 $\frac{1}{2} \cdot 2 \cdot 1 = 1$。答案 $\frac{5}{4}$ 符合这些要求。
- 在 (ii) 中，在 $[0,8]$ 上 $1 \le \mathrm{h}(t) \le 2$，所以积分介乎 8 与 16 之间，而 12.75 亦符合。
- 在多个点上以数值方法解这些三次方程，再用辛普森法则（Simpson's rule）积分，得到 1.25000 和 12.75000，与准确答案吻合。

常见错误

- 反函数积分的上限用错：上限是 $\mathrm{g}(2) = 1$，而不是 2。
- 把开首的结果应用于 h 时，误以为 $\mathrm{h}(0) = 0$。由于 $\mathrm{h}(0) = 1$，图中多了一条窄条，恒等式需要作出调整（或者改用 $\mathrm{h}(t) = \mathrm{g}(t+2)$）。
- 忘记证明 g 的定义是良好的，即 $y^3 + y = t$ 恰好有一个实根。

STEP 之外

开首的恒等式是*杨氏不等式*（Young's inequality）取等号的情况：对于满足 $\mathrm{f}(0) = 0$ 的递增函数 f 及任何 $a, b \ge 0$，都有 $ab \le \int_0^a \mathrm{f} + \int_0^b \mathrm{f}^{-1}$。取 $\mathrm{f}(t) = t^{p-1}$ 便得到 $ab \le a^p/p + b^q/q$，这是分析学中赫尔德不等式（Hölder's inequality）的出发点。g 可导且 $\mathrm{g}' = 1/\phi'(\mathrm{g})$ 这个事实，就是*反函数定理*（inverse function theorem）。Abbott 的 *Understanding Analysis*（《理解分析》）对此有仔细的证明。

详细解答

STEP II 2019，第 3 题

纯数学 · 多项式根的界限

STEP II 2019，第 3 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

三角不等式（triangle inequality）让我们把多项式（polynomial）的每一项换成它的大小，从而估计整个多项式的上界；而当 $|x| < 1$ 时，这些大小构成一个等比级数（geometric series）。这就给出细小的根（root）的一个下界（lower bound）。至于较大的根，窍门是考虑 $1/\omega$：把 f 的系数（coefficient）倒转排列，得到的多项式形式完全相同，于是关于细小根的结果可以再派上用场一次。在第 (ii) 部分，这些界把寻找整数根的范围缩小至只剩四个候选值。

解答

三角不等式

两边均为非负，因此只需比较它们的平方。由于 $x_1x_2 \le |x_1x_2| = |x_1||x_2|$，

$$|x_1 + x_2|^2 = x_1^2 + 2x_1x_2 + x_2^2 \le |x_1|^2 + 2|x_1||x_2| + |x_2|^2 = \big(|x_1| + |x_2|\big)^2,$$

取非负平方根即得 $|x_1 + x_2| \le |x_1| + |x_2|$。

对于 n 个数，我们对 n 用归纳法（induction）。$n = 1$ 的情况是等式，而 $n = 2$ 已经证明。若结果对 $n - 1$ 个数成立，则对 $x_1 + \cdots + x_{n-1}$ 和 x_n 应用两个数的情况，得

$$\begin{aligned} |x_1 + \cdots + x_{n-1} + x_n| &\le |x_1 + \cdots + x_{n-1}| + |x_n| \\ &\le |x_1| + \cdots + |x_{n-1}| + |x_n|, \end{aligned}$$

这正是 n 个数的结果。

第 (i)(a) 部分

我们有 $f(x)-1=a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}+x^n$。先用三角不等式，再用 $|a_i|\le A$，最后对末项用 $1\le A$，得

$$\begin{aligned}|f(x)-1| &\le |a_1||x|+\cdots+|a_{n-1}||x|^{n-1}+|x|^n\\ &\le A\big(|x|+|x|^2+\cdots+|x|^n\big).\end{aligned}$$

因为 $0\le|x|<1$，而无穷级数的所有项均为非负，所以有限的等比和不超过无穷的等比和：

$$|x|+|x|^2+\cdots+|x|^n\le\sum_{k=1}^{\infty}|x|^k=\frac{|x|}{1-|x|}.$$

因此 $|f(x)-1|\le\dfrac{A|x|}{1-|x|}$，得证。

第 (i)(b) 部分

由于 $f(\omega)=0$，我们有 $|f(\omega)-1|=1$，故由 (a) 部分得

$$1\le\frac{A|\omega|}{1-|\omega|}.$$

因为 $1-|\omega|>0$，两边同乘时不等号方向不变：$1-|\omega|\le A|\omega|$，即 $1\le(1+A)|\omega|$，所以

$$|\omega|\ge\frac{1}{1+A}.$$

上界则是显然的：$|\omega|<1\le1+A$。因此 $(*)$ 成立。

第 (i)(c) 部分

现在 $|\omega|\ge1$。下界很容易：由于 $A\ge1>0$，有 $|\omega|\ge1>\frac{1}{1+A}$。真正需要证明的是上界。

若 $|\omega|=1$，则立即有 $|\omega|\le1+A$，因此设 $|\omega|>1$。令 $g(y)=y^n f(1/y)$，即

$$g(y)=1+a_{n-1}y+a_{n-2}y^2+\cdots+a_1y^{n-1}+y^n.$$

> ? **想一想** 为甚么 g 可以使用 (a) 和 (b) 部分的结果?

它的形式与 f 完全相同：常数项（constant term）为 1，首项系数（leading coefficient）为 1，其余系数是倒序排列的 $a_{n-1},\dots,a_1$，每个的大小仍然不超过 A。因此 (b) 部分可以用同一个 A 应用于 g。

由于 $\omega \neq 0$，$g(1/\omega) = \omega^{-n} f(\omega) = 0$，所以 $1/\omega$ 是 g 的一个实根，且 $|1/\omega| < 1$。对 g 应用 (b) 部分，得

$$\frac{1}{|\omega|} \geq \frac{1}{1+A}, \qquad \text{所以} \qquad |\omega| \leq 1 + A.$$

因此，无论大小如何，f 的每个实根都满足 $(*)$。（注意 0 永远不是根，因为 $f(0) = 1$。）

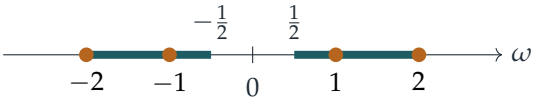
第 (ii) 部分

除以 135 并按升幂排列各项，方程即为 $f(x) = 0$，其中

$$f(x) = 1 - \tfrac{126}{135}x - \tfrac{91}{135}x^2 - \tfrac{100}{135}x^3 - x^4 + x^5.$$

这具有第 (i) 部分的形式，其中 $n = 5$：常数项和首项系数都是 1，其余系数的大小为 $\frac{126}{135}, \frac{91}{135}, \frac{100}{135}, 1$，全部不超过 $A = 1$。所以每个实根都满足

$$\tfrac{1}{2} \leq |\omega| \leq 2.$$



容许范围（涂色部分）内的整数只有 ± 1 和 ± 2。代入原来的五次多项式（quintic）$p(x) = 135x^5 - 135x^4 - 100x^3 - 91x^2 - 126x + 135$：

$$\begin{aligned} p(1) &= 135 - 135 - 100 - 91 - 126 + 135 = -182, \\ p(-1) &= -135 - 135 + 100 - 91 + 126 + 135 = 0, \\ p(2) &= 4320 - 2160 - 800 - 364 - 252 + 135 = 879, \\ p(-2) &= -4320 - 2160 + 800 - 364 + 252 + 135 = -5657. \end{aligned}$$

所以唯一的整数根是 $x = -1$。事实上

$$p(x) = (x+1)\big(135x^4 - 270x^3 + 170x^2 - 261x + 135\big),$$

而四次因式（quartic factor）在 $x = -1$ 时的值为 971，所以 -1 是单根（simple root）。

验算

- 展开 $(x+1)(135x^4 - 270x^3 + 170x^2 - 261x + 135)$，应得回原来的五次多项式。
- 这个五次多项式还有另外两个实根，约为 0.595 和 1.725；两者都在 $[\frac{1}{2}, 2]$ 之内，与第 (i) 部分的预测一致。
- 试取 $f(x) = 1 - 2x + x^2 = (1-x)^2$，其中 $A = 2$：根 1 落在 $[\frac{1}{3}, 3]$ 之内。这些界亦已用电脑在数百个随机多项式上验证过。

常见错误

- 在 (a) 部分忘记 x^n 项：它的系数是 1，而不是某个 a_n；正是 $A \geq 1$ 才让它可以并入等比级数。
- 在 (c) 部分，当 $|\omega| \geq 1$ 时试图直接在 ω 处再用 (a) 部分：此时等比级数不再收敛（converge）。应改为考虑 $1/\omega$。
- 在第 (ii) 部分把 135 的所有因数（divisors）逐一当作候选根测试：这样做是正确的，但第 (i) 部分的界只留下四个需要测试。

STEP 之外

以系数估计多项式的根的范围是一个经典课题；这里的结果是柯西（Cauchy）界（Cauchy's bound）的一个版本，而且同一论证对复根（complex roots）一字不改地成立，只需把绝对值（absolute value）换成模（modulus）。倒转变换 $y^n f(1/y)$ 把细小的根与较大的根互换，是代数和复分析（complex analysis）中的标准工具。若想了解这类界以及更多关于复根几何的内容，可以阅读 Needham 的 *Visual Complex Analysis*（《直观复分析》）。

详细解答

STEP II 2019，第 4 题

纯数学 · 三角函数裂项相消乘积

STEP II 2019，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

若一串余弦的乘积中，角度每次加倍（或减半），只要乘以最小角的正弦，乘积便会伸缩相消（telescopes）：二倍角公式（double-angle formula）$2\sin\theta\cos\theta = \sin 2\theta$ 会把因子逐一“吞掉”。乘积一旦化为闭合形式（closed form），取对数（logarithm）便把它变成和，再求导便把 $\ln\cos$ 变成 $\tan$：正切之和正是由此而来。无穷的版本则可令 $n \to \infty$，并利用两个标准极限得出。

解答

第 (i) 部分

把该乘积记为 P。由于 $\sin\frac{\pi}{9} \neq 0$，我们可以乘以它，并把恒等式使用三次：

$$\begin{aligned} 8\sin\tfrac{\pi}{9}\,P &= 4\big(2\sin\tfrac{\pi}{9}\cos\tfrac{\pi}{9}\big)\cos\tfrac{2\pi}{9}\cos\tfrac{4\pi}{9} = 2\big(2\sin\tfrac{2\pi}{9}\cos\tfrac{2\pi}{9}\big)\cos\tfrac{4\pi}{9} \\ &= 2\sin\tfrac{4\pi}{9}\cos\tfrac{4\pi}{9} = \sin\tfrac{8\pi}{9} = \sin\big(\pi - \tfrac{\pi}{9}\big) = \sin\tfrac{\pi}{9}. \end{aligned}$$

两边除以 $8\sin\frac{\pi}{9}$，得

$$\cos\tfrac{\pi}{9}\cos\tfrac{2\pi}{9}\cos\tfrac{4\pi}{9} = \frac{1}{8}.$$

第 (ii) 部分

这里的角 $x, \frac{x}{2}, \ldots, \frac{x}{2^n}$ 逐次减半，所以最小的是 $x/2^n$。由于 $0 < x/2^n < \frac{1}{2}\pi$，它的正弦不为零。乘以 $\sin(x/2^n)$ 后，因子 $\cos(x/2^n)$ 与它配对，得出 $\frac{1}{2}\sin(x/2^{n-1})$；这又与

$\cos(x/2^{n-1})$ 配对，如此类推，直至 $\cos x$。这 $n+1$ 步每一步都贡献一个因子 $\frac{1}{2}$：

$$\sin\left(\frac{x}{2^n}\right)\prod_{k=0}^{n}\cos\left(\frac{x}{2^k}\right)=\frac{1}{2^{n+1}}\sin 2x,\qquad \text{所以}\qquad \prod_{k=0}^{n}\cos\left(\frac{x}{2^k}\right)=\frac{\sin 2x}{2^{n+1}\sin(x/2^n)}.$$

想一想 求导怎样能把余弦的乘积变成正切的和?

取对数。每个因子都是正数，因为对每个 k 都有 $0<x/2^k<\frac{1}{2}\pi$，而且 $\sin 2x>0$。因此

$$\sum_{k=0}^{n}\ln\cos\left(\frac{x}{2^k}\right)=\ln\sin 2x-(n+1)\ln 2-\ln\sin\left(\frac{x}{2^n}\right).$$

对 x 求导，并由连锁法则（chain rule）得 $\frac{\mathrm{d}}{\mathrm{d}x}\ln\cos(x/2^k)=-\frac{1}{2^k}\tan(x/2^k)$：

$$-\sum_{k=0}^{n}\frac{1}{2^k}\tan\left(\frac{x}{2^k}\right)=2\cot 2x-\frac{1}{2^n}\cot\left(\frac{x}{2^n}\right).$$

乘以 -1，正好得出所求结果：

$$\sum_{k=0}^{n}\frac{1}{2^k}\tan\left(\frac{x}{2^k}\right)=\frac{1}{2^n}\cot\left(\frac{x}{2^n}\right)-2\cot 2x.$$

第 (iii) 部分

把乘积公式除以 $\cos x\neq 0$，并利用 $\sin 2x=2\sin x\cos x$，得

$$\prod_{k=1}^{n}\cos\left(\frac{x}{2^k}\right)=\frac{2\sin x\cos x}{2^{n+1}\sin(x/2^n)\cos x}=\frac{\sin x}{2^n\sin(x/2^n)}.$$

想一想 当 $n\to\infty$ 时，$2^n\sin(x/2^n)$ 会怎样?

设 $\theta=x/2^n$，于是当 $n\to\infty$ 时 $\theta\to 0$。那么

$$2^n\sin\left(\frac{x}{2^n}\right)=x\cdot\frac{\sin\theta}{\theta}\ \longrightarrow\ x\cdot 1=x,$$

因此

$$\prod_{k=1}^{\infty}\cos\left(\frac{x}{2^k}\right)=\lim_{n\to\infty}\frac{\sin x}{2^n\sin(x/2^n)}=\frac{\sin x}{x}.$$

至于级数 (series)，代入 $j=k+2$：那么 $\frac{1}{2^{j-2}}=\frac{1}{2^k}$ 且 $\frac{\pi}{2^j}=\frac{\pi/4}{2^k}$，所以这个和就是第 (ii) 部分的级数取 $x=\frac{\pi}{4}$ 的情形，而此值位于 $(0,\frac{1}{2}\pi)$ 内。取这个 x 时，$2\cot 2x=2\cot\frac{\pi}{2}=0$，所以直至 $k=n$ 的部分和（partial sum）为

$$\frac{1}{2^n}\cot\left(\frac{\pi}{4\cdot 2^n}\right)=\frac{1}{x}\cdot\frac{\theta}{\tan\theta}\qquad\left(\theta=\frac{x}{2^n},\ x=\frac{\pi}{4}\right).$$

当 $n\to\infty$ 时，$\theta\to 0$ 且 $\theta/\tan\theta\to 1$，所以部分和趋于 $1/x$：

$$\sum_{j=2}^{\infty}\frac{1}{2^{j-2}}\tan\left(\frac{\pi}{2^j}\right)=\frac{4}{\pi}.$$

验算

- 第 (i) 部分：$\cos 20^\circ \cos 40^\circ \cos 80^\circ \approx 0.9397 \times 0.7660 \times 0.1736 \approx 0.1250$。
- 第 (ii) 部分取 $n = 0$：乘积就是 $\cos x$，而公式给出 $\sin 2x/(2\sin x) = \cos x$。取 $n = 0$ 时，和的恒等式变成 $\tan x = \cot x - 2\cot 2x$，这正是 cot 的二倍角公式移项后的形式。
- 第 (iii) 部分：级数的首几项为 $\tan \frac{\pi}{4} = 1$、$\frac{1}{2}\tan \frac{\pi}{8} \approx 0.2071$、$\frac{1}{4}\tan \frac{\pi}{16} \approx 0.0497$、…；经电脑验证，总和为 $1.27324\ldots = 4/\pi$。
- 在 $\prod \cos(x/2^k) = \sin x/x$ 中令 $x \to 0$，得 $1 = 1$，理应如此。

常见错误

- 数错 $\frac{1}{2}$ 因子的个数：由 $k = 0$ 到 n 的乘积有 $n + 1$ 个因子，所以分母是 2^{n+1}，而不是 2^n。
- 取对数时没有指出每个因子都是正数；这正是范围 $0 < x < \frac{1}{2}\pi$ 重要的原因。
- 忘记 $2\cot 2x$ 在 $x = \frac{\pi}{4}$ 时为零，或者错用 $x = \frac{\pi}{2}$（这样 $j = 1$ 的项便是 $\tan \frac{\pi}{2}$，而它没有定义）。

STEP 之外

在 $\prod \cos(x/2^k) = \sin x/x$ 中设 $x = \frac{\pi}{2}$，并把每个余弦写成嵌套平方根，便得到韦达（Viète）公式 $\frac{2}{\pi} = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2+\sqrt{2}}}{2} \cdots$，这是数学史上第一个无穷乘积（infinite product）。像我们对有限乘积那样，对无穷乘积的对数逐项求导，是需要理据支持的；分析（analysis）的入门课程会讲清楚何时可以这样做。欧拉（Euler）乘积 $\sin x = x\prod_{n\geq 1}\left(1 - \frac{x^2}{n^2\pi^2}\right)$ 也是从同样的思路发展而来。如想了解严谨的理论背景，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP II 2019，第 5 题

纯数学 • 迭代的不动点

STEP II 2019，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

常数数列（constant sequence）的首项是一个*不动点*（fixed point），即 $\mathrm{f}(x) = x$ 的解。周期（period）为 2 的数列，其首项是 $\mathrm{f}(\mathrm{f}(x)) = x$ 的一个解，而且不是不动点。每个不动点同时也是 $\mathrm{f}(\mathrm{f}(x)) = x$ 的解，所以 $\mathrm{f}(x) - x$ 必定是四次多项式（quartic）$\mathrm{f}(\mathrm{f}(x)) - x$ 的因式（factor）。只要求出另一个因式（它是一个二次式（quadratic）），一切便归结为判别式（discriminant）的问题。

解答

两部分通用的因式分解

在两部分中，$\mathrm{f}(x) = x^2 - px + c$，其中 c 为常数（在 (i) 中 $c = p$，在 (ii) 中 $c = q$）。对任意 x 和 y，

$$\mathrm{f}(y) - \mathrm{f}(x) = (y^2 - x^2) - p(y - x) = (y - x)(y + x - p).$$

令 $y = \mathrm{f}(x)$，再在两边加上 $\mathrm{f}(x) - x$，得

$$\begin{aligned}\mathrm{f}(\mathrm{f}(x)) - x &= \big[\mathrm{f}(\mathrm{f}(x)) - \mathrm{f}(x)\big] + \big[\mathrm{f}(x) - x\big] \\ &= \big(\mathrm{f}(x) - x\big)\big(\mathrm{f}(x) + x - p + 1\big).\end{aligned}$$

我们会用到以下两个推论：

- 数列的周期为 2，当且仅当 a 是 $Q(x) = \mathrm{f}(x) + x - p + 1$ 的根，而*不是* $\mathrm{f}(x) - x$ 的根。事实上，若 $\mathrm{f}(\mathrm{f}(a)) = a$，则 $u_2 = u_0$；由于每一项都由前一项决定，以归纳法（induction）可知对所有 n 都有 $u_{n+2} = u_n$。此时数列不是常数数列，当且仅当 $u_1 \neq u_0$，即 $\mathrm{f}(a) \neq a$。

- 若 x 是不动点，则 $Q(x) = 2x - p + 1$。因此不动点要同时是 Q 的根，只可能是 $x = \frac{1}{2}(p-1)$。

第 (i) 部分

这里 $\mathrm{f}(x) = x^2 - px + p$，所以

$$\mathrm{f}(a) - a = a^2 - (p+1)a + p = (a-1)(a-p).$$

数列是常数数列，当且仅当 $u_1 = u_0$（此时每一项都等于 a），所以所求的值为

$$a = 1 \quad 及 \quad a = p.$$

至于周期 2，另一个因式是

$$Q(x) = x^2 - px + p + x - p + 1 = x^2 + (1-p)x + 1,$$

其判别式为

$$(1-p)^2 - 4 = (p-1-2)(p-1+2) = (p-3)(p+1).$$

- 若 $-1 < p < 3$，判别式为负，Q 没有实根，因此没有实数 a 能给出周期 2。
- 若 $p = 3$ 或 $p = -1$，Q 只有一个根 $x = \frac{1}{2}(p-1)$，分别为 1 或 -1。当 $p = 3$ 时，不动点是 1 和 3；当 $p = -1$ 时，不动点是 1 和 -1。两种情况下 Q 的根都是不动点，所以由此出发的数列是常数数列，而不是周期 2。
- 若 $p > 3$ 或 $p < -1$，Q 有两个相异实根。两者都不是不动点：根据上面第二点，若是不动点则必须等于 $\frac{1}{2}(p-1)$，但

  $$Q\left(\tfrac{1}{2}(p-1)\right) = \tfrac{1}{4}(p-1)^2 - \tfrac{1}{2}(p-1)^2 + 1 = 1 - \tfrac{1}{4}(p-1)^2 \neq 0$$

  因为 $(p-1)^2 > 4$。所以取 a 为任一个根，都得到周期为 2 的数列。

因此，存在能给出周期 2 的 a 值，当且仅当 $p > 3$ 或 $p < -1$。

第 (ii) 部分

现在 $\mathrm{f}(x) = x^2 - px + q$，所以

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

常数数列需要某个 a 使 $\mathrm{f}(a) = a$，即这个二次式的一个实根。

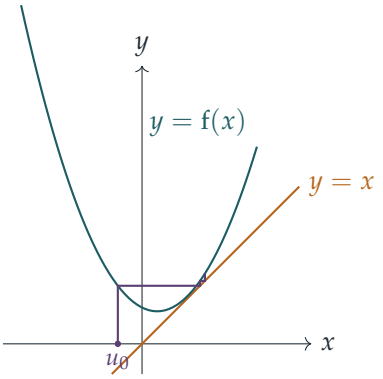
若对所有 x 都有 $\mathrm{f}(x) > x$，则对每个 a 都有 $\mathrm{f}(a) \neq a$，所以没有数列是常数数列。

反过来，假设没有 a 给出常数数列，即 $g(x) = x^2 - (p+1)x + q$ 没有实根。那么 g 是连续（continuous）的而且永不为零，所以它不会变号；又因为当 $x \to \infty$ 时 $g(x) \to +\infty$，它处处为正。（等价地说：判别式为负，而首项系数（leading coefficient）为正。）也就是说，对所有 x 都有 $\mathrm{f}(x) > x$，这便完成了“当且仅当”的证明。

推论。假设不存在常数数列。那么对所有 x 都有 $\mathrm{f}(x) > x$，所以对任何 a，

$$u_{n+1} = \mathrm{f}(u_n) > u_n \quad \text{对所有} n.$$

数列严格递增（strictly increasing），所以 $u_2 > u_1 > u_0$。特别地，$u_2 \neq u_0$，数列不可能有周期 2。



上图展示了这个情况：抛物线（parabola）完全位于 $y = x$ 之上，所以迭代（iterate）形成的“阶梯”只能不断向上攀升。

最后一部分。逆命题（converse）是错误的。取 $p = q = 0$，即 $\mathrm{f}(x) = x^2$。当 $a = 0$ 和 $a = 1$ 时，数列是常数数列。至于周期 2，我们需要 $\mathrm{f}(\mathrm{f}(a)) = a^4 = a$，即 $a(a-1)(a^2+a+1) = 0$。因式 $a^2 + a + 1$ 的判别式为负，所以唯一的实数解是 $a = 0$ 和 $a = 1$，两者都是不动点。因此，虽然有些 a 值给出常数数列，却没有任何 a 值给出周期 2。

想一想 $\mathrm{f}(x) = x^2$ 只是碰巧成立吗?

不是。一般而言，不动点的二次方程的判别式为 $D = (p+1)^2 - 4q$，而 $Q(x) = x^2 + (1-p)x + q - p + 1$ 的判别式是

$$(1-p)^2 - 4(q-p+1) = (p+1)^2 - 4q - 4 = D - 4.$$

与 (i) 一样，Q 的根只有在 Q 的判别式为零时才会是不动点。所以当 $D \geq 0$ 时存在常数数列，当 $D > 4$ 时存在周期 2 的数列。任何满足 $0 \leq D \leq 4$ 的 p、q 都给出一个反例（counterexample）。

验算

- 在 (i) 中取 $p = 4$，$Q(x) = x^2 - 3x + 1$ 的根为 $\frac{1}{2}(3 \pm \sqrt{5}) \approx 2.618,\ 0.382$。从 2.618 出发，$\mathrm{f}(2.618) = 4 + (2.618 - 4)(2.618) \approx 0.382$，而 $\mathrm{f}(0.382) \approx 2.618$：周期为 2，与预测一致。
- 当 $p = 3$ 时，Q 的根是 $x = 1$，而 $\mathrm{f}(1) = 3 + (1 - 3) = 1$：这是不动点，证实边界值须被排除。
- 展开 $(x^2 - (p+1)x + p)(x^2 + (1-p)x + 1)$ 得 $\mathrm{f}(\mathrm{f}(x)) - x$；以电脑在数千组随机的 (p, x) 上检验，连同 (ii) 中“周期 2 当且仅当 $D > 4$”这个条件，均没有发现例外。

常见错误

- 把 $p = 3$ 和 $p = -1$ 也包括在内：此时 $\mathrm{f}(\mathrm{f}(a)) = a$ 有重根，但它是不动点，所以数列是常数数列。
- 忘记在寻找周期 2 的点之前，必须先从 $\mathrm{f}(\mathrm{f}(x)) - x$ 中除去因式 $\mathrm{f}(x) - x$。
- 在最后一部分只凭不等式的方向来论证；要回答“是否成立...?”这类问题，一个明确的反例才是决定性的。

STEP 之外

这是踏入*离散动力系统*（discrete dynamical systems）的第一步。每一个二次映射，经过线性变量代换后，都等价于逻辑斯谛映射（logistic map）$x \mapsto rx(1-x)$。随着参数增大，不动点让位给 2-循环，然后是 4-循环，如此类推，形成一连串倍周期分岔（period-doubling cascade），最终进入混沌（chaos）。你所找到的临界值，即 Q 的判别式为零之处，正是这些分岔（bifurcation）中的第一个。Strogatz 的 *Nonlinear Dynamics and Chaos*（《非线性动力学与混沌》）是一本非常易读的入门书。

详细解答

STEP II 2019，第 6 题

纯数学 · 一阶微分方程

STEP II 2019，第 6 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

两条方程的右方都只通过一个组合依赖 x 和 y：第 (i) 部分是 $y+x$，第 (ii) 部分是 $y-x$，因为那个二次式其实就是 $(y-x)^2+4(y-x)+3$。单凭这一个观察，就能找出直线解，把驻点（stationary point）限制在某些直线上，并且（在第 (i) 部分）把方程化成可以分离变量（separation of variables）的形式。之后，题目中关于解曲线（solution curve）永不相交的提示就能完成余下的工作：一条从两条直线解之间出发的曲线，会永远被困在两者之间。

解答

第 (i) 部分

直线解。若 $y=mx+c$，则 $\frac{\mathrm{d}y}{\mathrm{d}x}=m$，所以我们需要

$$m=mx+c+x+1=(m+1)x+(c+1)\quad \text{对所有}x.$$

比较 x 的系数得 $m=-1$，再比较常数项得 $c+1=m=-1$，所以 $c=-2$。这个解是 $y=-x-2$。

驻点在哪里。在 $y=y_3(x)$ 的驻点上，我们有 $0=\frac{\mathrm{d}y}{\mathrm{d}x}=y+x+1$，所以该点位于直线 $y=-x-1$ 上。

现在假设 $k<-2$。曲线 $y=y_3(x)$ 和直线 $y=-x-2$ 都是解，而在 $x=0$ 时曲线位于直线下方（$k<-2$）。两者永不相遇，所以对每个 x，曲线都留在直线下方：$y_3(x)<-x-2<-x-1$。因此 $y_3(x)+x+1<-1<0$ 处处成立。斜率永不为零（事实上总是小于 -1），所以没有驻点。

每个驻点都是极小点。把方程求导，

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\mathrm{d}y}{\mathrm{d}x} + 1.$$

在驻点上一阶导数（derivative）为 0，所以二阶导数是 $1 > 0$，该点是局部极小点（local minimum）。

求解。设 $Y = y + x$，则 $\frac{\mathrm{d}Y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}x} + 1 = Y + 2$。在直线 $Y = -2$ 上，这就是我们已有的解；否则便分离变量：

$$\int \frac{\mathrm{d}Y}{Y+2} = \int \mathrm{d}x \quad \Longrightarrow \quad \ln|Y+2| = x + \text{常数} \quad \Longrightarrow \quad Y + 2 = Ae^x,$$

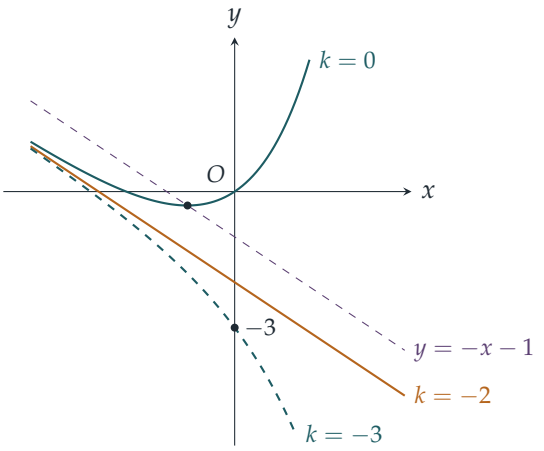
其中 A 是常数（而 $A = 0$ 对应那条直线）。所以 $y = Ae^x - x - 2$，由 $y(0) = k$ 得 $A = k + 2$：

$$y_3(x) = (k+2)e^x - x - 2.$$

这与我们之前的发现一致：$y_3'(x) = (k+2)e^x - 1$，当 $k \le -2$ 时永不为零；当 $k > -2$ 时，它在 $e^x = 1/(k+2)$ 处为零，且只有一次。

三条曲线。

- $k = 0$：$y = 2e^x - x - 2$，经过原点。极小点在 $e^x = \frac{1}{2}$ 处，即 $(-\ln 2,\ \ln 2 - 1) \approx (-0.69, -0.31)$，位于直线 $y = -x - 1$ 上。当 $x \to -\infty$ 时，$2e^x \to 0$，曲线从上方趋近 $y = -x - 2$；当 $x \to \infty$ 时，指数项占主导。
- $k = -2$：就是直线 $y = -x - 2$ 本身。
- $k = -3$：$y = -e^x - x - 2$，经过 $(0, -3)$，一直递减。当 $x \to -\infty$ 时，它从下方趋近 $y = -x - 2$；当 $x \to \infty$ 时，它像 $-e^x$ 一样急降。



第 (ii) 部分

改写右方。把各项归组，

$$x^2+y^2-2xy-4x+4y+3=(y-x)^2+4(y-x)+3=(y-x+1)(y-x+3).$$

不妨设 $u=y-x$，于是 $\frac{\mathrm{d}y}{\mathrm{d}x}=(u+1)(u+3)$。

直线解。若 $y=mx+c$，则 $u=(m-1)x+c$，我们需要

$$m=\big((m-1)x+c+1\big)\big((m-1)x+c+3\big)\quad\text{对所有}x.$$

若 $m\neq 1$，右方是一个真正的 x 的二次式，不可能等于常数，所以 $m=1$。于是 $1=(c+1)(c+3)$，即 $c^2+4c+2=0$，得 $c=-2\pm\sqrt{2}$。两个解是

$$y=x-2+\sqrt{2}\qquad\text{及}\qquad y=x-2-\sqrt{2}.$$

y_4 的驻点。斜率恰好在 $u=-1$ 或 $u=-3$ 时为零，所以任何驻点都位于

$$y=x-1\qquad\text{或}\qquad y=x-3.$$

在这两条直线之间，我们有 $-3<y-x<-1$，所以 $u+1<0<u+3$，斜率 $(u+1)(u+3)$ 是负的。（在两线之外，斜率是正的。）

曲线落在哪里。在 $x=0$ 时，$u=-2$，介乎 $-2-\sqrt{2}\approx-3.41$ 与 $-2+\sqrt{2}\approx-0.59$ 之间。所以在 $x=0$ 时，曲线位于两条直线解之间；又因为解曲线永不相交，

$$x-2-\sqrt{2}<y_4(x)<x-2+\sqrt{2}\quad\text{对所有}x.$$

由两条驻点直线所夹的带状区域（$-3<u<-1$）位于这条带状区域之内。

> ? **想一想** 曲线被困在两条平行线之间，但它在带状区域内会走向哪里？

追踪 $u=y_4-x$。它满足

$$\frac{\mathrm{d}u}{\mathrm{d}x}=(u+1)(u+3)-1=u^2+4u+2=\big(u+2-\sqrt{2}\big)\big(u+2+\sqrt{2}\big),$$

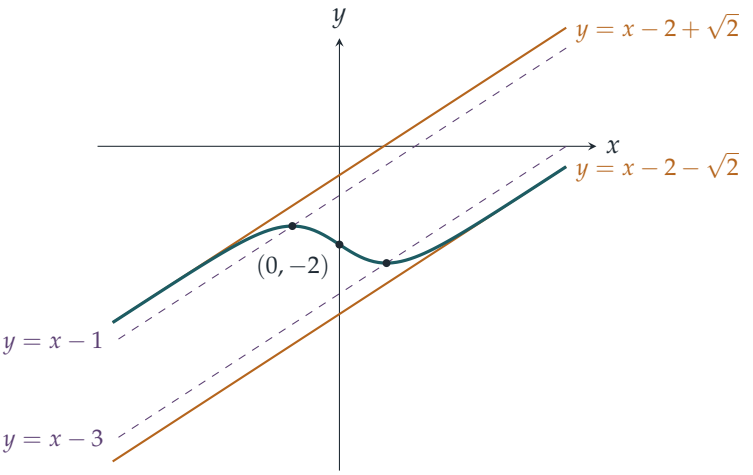
只要 $-2-\sqrt{2}<u<-2+\sqrt{2}$，这就是负数，而这一直成立。所以 u 严格递减且有界，因此当 $x\to\pm\infty$ 时有极限。在这样的极限处，$\frac{\mathrm{d}u}{\mathrm{d}x}$ 必须趋于 0（否则 u 会无界地改变），所以极限就是那两个根：当 $x\to-\infty$ 时 $u\to-2+\sqrt{2}$，当 $x\to+\infty$ 时 $u\to-2-\sqrt{2}$。

所以，由左至右看：

- 在最左方，曲线从下方紧贴 $y=x-2+\sqrt{2}$，斜率接近 1；
- 当 u 下降穿过 -1 时，曲线与 $y=x-1$ 相交于唯一一点，这是一个局部极大点（斜率由正转负）；

- 它以斜率 $(-1)(1) = -1$ 向下穿过 $(0, -2)$；
- 当 u 下降穿过 -3 时，曲线与 $y = x - 3$ 相交于唯一一点，这是一个*局部极小点*；
- 之后它再次上升，从上方紧贴 $y = x - 2 - \sqrt{2}$，斜率趋于 1。

由于 u 严格递减，$u = -1$ 和 $u = -3$ 各自恰好只经过一次，所以恰好有两个驻点。还有一个漂亮的特点：把 (x, y) 换成 $(-x, -4 - y)$，方程和起点 $(0, -2)$ 都保持不变，所以曲线关于 $(0, -2)$ 有半周旋转对称（half-turn symmetry），而极大点和极小点互为关于该点的镜像。



验算

- 在第 (i) 部分代回原式：$y = (k+2)e^x - x - 2$ 给出 $y' = (k+2)e^x - 1$，而 $y + x + 1 = (k+2)e^x - 1$。设 $k = -2$ 便得回那条直线。
- 当 $k = 0$ 时，极小点 $(-\ln 2, \ln 2 - 1) \approx (-0.693, -0.307)$ 确实位于 $y = -x - 1$ 上。
- 在第 (ii) 部分，从 $(0, -2)$ 出发的逐步数值解（经电脑核实）一直留在两条直线之间，在 $y = x - 1$ 上约 $(-0.623, -1.623)$ 处达到极大，在 $y = x - 3$ 上约 $(0.623, -2.377)$ 处达到极小，到了 $x = \pm 6$ 时与渐近线（asymptote）的距离约在 10^{-7} 之内。这两点互为关于 $(0, -2)$ 的像，正如对称性所预测。

常见错误

- 忘记运用不相交的性质：“$k < -2$ 意味着没有驻点”需要曲线对*所有* x 都留在 $y = -x - 2$ 下方，而不只是在 $x = 0$ 时。

- 画 y_4 时让它穿过其中一条直线解，或把极大点和极小点的位置弄反。$(u+1)(u+3)$ 在两条虚线之间的正负号可以判定哪个是哪个。
- 在第 (ii) 部分求 m 和 c 时，未先排除（或确认）$m=1$ 便除以 $m-1$。

STEP 之外

题目开头的提示，就是皮卡－林德洛夫定理（Picard–Lindelöf theorem）中关于唯一性（uniqueness）的那一半：对于光滑的右方，每一点恰好有一个解经过，所以解曲线不能相交。第 (ii) 部分其实是一条乔装的自治（autonomous）方程 $u'=u^2+4u+2$；我们不用任何公式，只凭平衡点（equilibrium）以及 u' 在平衡点之间的正负号，就找出曲线的走向，这正是微分方程定性理论（qualitative theory of differential equations）的起点。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）既涵盖唯一性，也涵盖这种相线（phase line）图像。

详细解答

STEP II 2019，第 7 题

纯数学 · 向量与纯量积

> **STEP II 2019，第 7 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

对于单位向量(unit vector),长度和角度都可以由标量积(scalar product)得出:$|\mathbf{u}-\mathbf{v}|^2 = 2-2\,\mathbf{u}\cdot\mathbf{v}$。因此诀窍是把“总和为零”这个条件拆开，一些向量放在一边，其余放在另一边，然后计算两边长度的平方。这样便把向量方程转化为标量积之间的等式，而这些等式就能告诉我们图形的形状。

解答

以下我们会一直用到这个事实：对于单位向量 $\mathbf{u}$、$\mathbf{v}$,

$$|\mathbf{u}\pm\mathbf{v}|^2 = \mathbf{u}\cdot\mathbf{u}\pm 2\,\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{v} = 2\pm 2\,\mathbf{u}\cdot\mathbf{v}.$$

第 (i) 部分

由于 $\mathbf{a}+\mathbf{b}=-\mathbf{c}$，两边取长度的平方，得

$$2+2\,\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}+\mathbf{b}|^2 = |\mathbf{c}|^2 = 1,$$

所以 $\mathbf{a}\cdot\mathbf{b} = -\frac{1}{2}$，得证。

这个条件对 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 是对称的，所以用同样的方法（写成 $\mathbf{b}+\mathbf{c}=-\mathbf{a}$ 及 $\mathbf{c}+\mathbf{a}=-\mathbf{b}$），可得 $\mathbf{b}\cdot\mathbf{c} = \mathbf{c}\cdot\mathbf{a} = -\frac{1}{2}$。因此

$$AB^2 = |\mathbf{a}-\mathbf{b}|^2 = 2-2\left(-\tfrac{1}{2}\right) = 3,$$

同理 $BC^2 = CA^2 = 3$。三条边的长度都是 $\sqrt{3}$,所以三**角形** ABC **是等边三角形(equilateral triangle)**。(每对位置向量(position vector)之间的夹角都是 $120°$,而由于 $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}) = \mathbf{0}$, O 既是三角形的外心（circumcentre），也是它的形心（centroid）。)

第 (ii) 部分

把条件写成 $\mathbf{a}_1+\mathbf{a}_2=-(\mathbf{a}_3+\mathbf{a}_4)$，两边取长度的平方：

$$2+2\,\mathbf{a}_1\cdot\mathbf{a}_2=2+2\,\mathbf{a}_3\cdot\mathbf{a}_4 \quad\Longrightarrow\quad \mathbf{a}_1\cdot\mathbf{a}_2=\mathbf{a}_3\cdot\mathbf{a}_4.$$

以另外两种可能的方式把向量分组，同样可得

$$\mathbf{a}_1\cdot\mathbf{a}_3=\mathbf{a}_2\cdot\mathbf{a}_4, \qquad \mathbf{a}_1\cdot\mathbf{a}_4=\mathbf{a}_2\cdot\mathbf{a}_3.$$

由于 $A_iA_j^2=2-2\,\mathbf{a}_i\cdot\mathbf{a}_j$，这表示对棱相等：$A_1A_2=A_3A_4$、$A_1A_3=A_2A_4$ 及 $A_1A_4=A_2A_3$。

(a) 四点共面

? **想一想** 四点所在的平面是否一定经过原点?

是的。四点的形心是 $\frac{1}{4}(\mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3+\mathbf{a}_4)=\mathbf{0}$，而平面上各点的形心必定在该平面上。所以该平面包含 O；又由于每个 A_i 与 O 的距离都是 1，四点都在该平面上以 O 为圆心、半径为 1 的圆上。以下我们都在这个平面上讨论。

设 $\mathbf{w}=\mathbf{a}_1+\mathbf{a}_2$，于是 $\mathbf{a}_3+\mathbf{a}_4=-\mathbf{w}$。

若 $\mathbf{w}=\mathbf{0}$，则 $\mathbf{a}_2=-\mathbf{a}_1$ 及 $\mathbf{a}_4=-\mathbf{a}_3$。

若 $\mathbf{w}\neq\mathbf{0}$，把 $\mathbf{a}_3$ 与 $\mathbf{a}_3+\mathbf{a}_4=-\mathbf{w}$ 取点积，并利用 (ii) 的结果：

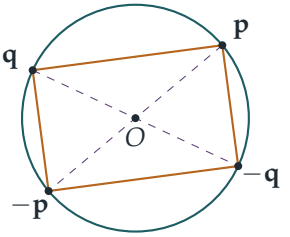
$$\mathbf{a}_3\cdot\mathbf{w}=-(1+\mathbf{a}_3\cdot\mathbf{a}_4)=-(1+\mathbf{a}_1\cdot\mathbf{a}_2)=-\tfrac{1}{2}|\mathbf{w}|^2.$$

所以 $\mathbf{a}_3$ 是一个沿 $\mathbf{w}$ 方向分量固定的单位向量：它的端点落在一条垂直于 $\mathbf{w}$ 的固定直线上，而一条直线与一个圆最多相交于两点。现在 $-\mathbf{a}_1$ 和 $-\mathbf{a}_2$ 正是两个这样的向量，因为

$$(-\mathbf{a}_1)\cdot\mathbf{w}=-(1+\mathbf{a}_1\cdot\mathbf{a}_2)=-\tfrac{1}{2}|\mathbf{w}|^2,$$

$-\mathbf{a}_2$ 亦同理；又因 $A_1\neq A_2$，两者并不相同。因此 $\mathbf{a}_3=-\mathbf{a}_1$ 或 $\mathbf{a}_3=-\mathbf{a}_2$，而 $\mathbf{a}_4=-\mathbf{w}-\mathbf{a}_3$ 就是另一个。

在所有情况下，四个位置向量都是 $\mathbf{p},-\mathbf{p},\mathbf{q},-\mathbf{q}$，其中 $\mathbf{q}\neq\pm\mathbf{p}$ 为某两个单位向量。所以四点是圆上两条直径（diameter）的端点。它们组成的四边形的对角线（diagonal）互相平分（于 O），所以它是平行四边形（parallelogram）；而且对角线相等（长度都是 2），所以它是**长方形（rectangle）**。当且仅当两条直径互相垂直时，它是正方形。



（严格来说，顶点应按圆周上的次序排列。若标记刚好是 $A_1 = \mathbf{p}$、$A_2 = \mathbf{q}$、$A_3 = -\mathbf{q}$、$A_4 = -\mathbf{p}$，按此次序连接各点，画出的会是这个长方形的两条边和两条对角线。）

(b) 正四面体 现在全部六条棱 A_iA_j 都相等，所以全部六个标量积 $\mathbf{a}_i \cdot \mathbf{a}_j$（$i \neq j$）取同一个值，设为 k。把 $\mathbf{a}_1$ 与零和取点积：

$$0 = \mathbf{a}_1 \cdot (\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4) = 1 + 3k,$$

所以 $k = -\frac{1}{3}$。于是每条棱都有

$$A_1A_2^2 = 2 - 2k = 2 + \tfrac{2}{3} = \tfrac{8}{3}, \qquad \text{所以边长为} \qquad \sqrt{\tfrac{8}{3}} = \frac{2\sqrt{6}}{3}.$$

验算

- 以 $\frac{1}{\sqrt{3}}(1,1,1)$、$\frac{1}{\sqrt{3}}(1,-1,-1)$、$\frac{1}{\sqrt{3}}(-1,1,-1)$、$\frac{1}{\sqrt{3}}(-1,-1,1)$ 为顶点的四面体是正四面体（regular tetrahedron），其位置向量都是单位向量且总和为 $\mathbf{0}$，棱长为 $\frac{1}{\sqrt{3}}|(0,2,2)| = \frac{2\sqrt{2}}{\sqrt{3}} = \frac{2\sqrt{6}}{3} \approx 1.633$。
- 与 (i) 比较：三个夹角为 120° 的单位向量给出边长 $\sqrt{3} \approx 1.732$，稍长一点。这是合理的，因为正四面体的顶点在球面上“更挤迫”。
- 用电脑对数千组随机的、总和为零的四个单位向量进行验算，证实了 (ii) 中三条标量积等式全部成立，而且在共面的情况下，各点总是以相反的成对形式 $\pm\mathbf{p}$、$\pm\mathbf{q}$ 出现。

常见错误

- 在 (i) 中，得出 $\mathbf{a} \cdot \mathbf{b} = -\frac{1}{2}$ 便停下来：必须求出全部三个标量积（或全部三条边长），才可断定三角形是等边三角形。
- 在 (a) 中，只凭对边相等便断定是“平行四边形”，而没有检查顶点的排列方式；或忘记证明该平面经过 O。

STEP 之外

一组总和为零、且两两标量积都相等的单位向量，正是*正单形*（*regular simplex*）的顶点：在 n 维空间中共有 $n+1$ 个这样的向量，且 $\mathbf{a}_i \cdot \mathbf{a}_j = -\frac{1}{n}$，这就给出了上面的 $-\frac{1}{2}$ 和 $-\frac{1}{3}$。这类构型见于化学（甲烷的键角 109.5° 就是 $\cos^{-1}(-\frac{1}{3})$）以及编码理论（coding theory）。由所有标量积 $\mathbf{a}_i \cdot \mathbf{a}_j$ 组成的矩阵称为 *Gram 矩阵*（*Gram matrix*），它在不计旋转的意义下决定了一组向量的构型，而这正是整道题背后的想法。想深入了解，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

STEP II 2019，第 8 题

纯数学 · 矩阵的积性函数

> **STEP II 2019，第 8 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 7.3 节甚么才算行列式中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch07/determinant-axioms.html

关键思路

一切都源于单一法则 $\mathrm{f}(\mathbf{MN}) = \mathrm{f}(\mathbf{M})\mathrm{f}(\mathbf{N})$，只要把它套用到精心挑选的乘积上即可。每次的窍门都是找出同一个矩阵的两种不同写法，或者找出一个等于它自身与另一矩阵之乘积的矩阵，然后比较 f 的值。行列式（determinant）正是这样的一个函数，所以它能帮助我们预料答案应该是甚么；但我们绝不能假设 f 就是行列式。

解答

第 (i) 部分

由于 $\mathbf{MI} = \mathbf{M}$，

$$\mathrm{f}(\mathbf{M}) = \mathrm{f}(\mathbf{MI}) = \mathrm{f}(\mathbf{M})\,\mathrm{f}(\mathbf{I}).$$

题目给出 $\mathrm{f}(\mathbf{M}) \neq 0$，因此可以两边同除以它，得 $\mathrm{f}(\mathbf{I}) = 1$。

（若没有这个假设，便无法得出此结论：恒等于零的函数同样是可乘的（multiplicative）。）

第 (ii) 部分

直接相乘可得 $\mathbf{J}^2 = \mathbf{I}$，因此

$$\mathrm{f}(\mathbf{J})^2 = \mathrm{f}(\mathbf{J}^2) = \mathrm{f}(\mathbf{I}) = 1.$$

故此 $\mathrm{f}(\mathbf{J}) = \pm 1$；又因为 $\mathrm{f}(\mathbf{J}) \neq 1$，所以 $\mathrm{f}(\mathbf{J}) = -1$。

> **想一想** 乘以 **J** 对一个矩阵有甚么作用?

左乘 **J** 会交换两行（rows）；右乘则会交换两列（columns）：

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}, \qquad \begin{pmatrix} c & d \\ a & b \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} d & c \\ b & a \end{pmatrix}.$$

对两式分别取 f，并利用 $\mathrm{f}(\mathbf{J}) = -1$，得

$$\mathrm{f}\begin{pmatrix} c & d \\ a & b \end{pmatrix} = -\mathrm{f}\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathrm{f}\begin{pmatrix} d & c \\ b & a \end{pmatrix} = -\mathrm{f}\begin{pmatrix} c & d \\ a & b \end{pmatrix}.$$

第一式给出 $\mathrm{f}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = -\mathrm{f}\begin{pmatrix} c & d \\ a & b \end{pmatrix}$；把第二式乘以 -1，便知道这也等于 $\mathrm{f}\begin{pmatrix} d & c \\ b & a \end{pmatrix}$，正如所求。

第 (iii) 部分

设 **A** 的第一行为 $(a\ \ b)$，第二行为它的 λ 倍。设 **B** 为两行相同的矩阵，并在 **K** 中取 $k = \lambda$。左乘 **K** 会把第二行乘以 k，因此

$$\mathbf{A} = \begin{pmatrix} a & b \\ \lambda a & \lambda b \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} a & b \\ a & b \end{pmatrix} = \mathbf{KB}.$$

交换 **B** 的两行后矩阵保持不变，所以由第 (ii) 部分（取 $c = a$、$d = b$）得

$$\mathrm{f}(\mathbf{B}) = -\mathrm{f}(\mathbf{B}),$$

这迫使 $\mathrm{f}(\mathbf{B}) = 0$。因此 $\mathrm{f}(\mathbf{A}) = \mathrm{f}(\mathbf{K})\,\mathrm{f}(\mathbf{B}) = 0$。这对每个实数 λ 都成立，包括 $\lambda = 0$。

第 (iv) 部分

首先计算这三个矩阵。当 $k \neq 0$ 时，$\mathbf{K}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/k \end{pmatrix}$，而且

$$\mathbf{P}^2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{P}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{K}^{-1}\mathbf{P}\mathbf{K} = \begin{pmatrix} 1 & 0 \\ 0 & 1/k \end{pmatrix}\begin{pmatrix} 1 & k \\ 0 & k \end{pmatrix} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}.$$

接着，以 **K** 作共轭（conjugation）不会改变 f 的值。事实上 $\mathrm{f}(\mathbf{K}^{-1})\mathrm{f}(\mathbf{K}) = \mathrm{f}(\mathbf{K}^{-1}\mathbf{K}) = \mathrm{f}(\mathbf{I}) = 1$，所以

$$\mathrm{f}(\mathbf{K}^{-1}\mathbf{P}\mathbf{K}) = \mathrm{f}(\mathbf{K}^{-1})\,\mathrm{f}(\mathbf{P})\,\mathrm{f}(\mathbf{K}) = \mathrm{f}(\mathbf{P}).$$

? **想一想** k 取哪些值时，$\mathbf{K}^{-1}\mathbf{PK}$ 会变成 $\mathbf{P}^2$ 和 $\mathbf{P}^{-1}$?

取 $k=2$ 得 $\mathbf{K}^{-1}\mathbf{PK}=\mathbf{P}^2$，所以

$$\mathrm{f}(\mathbf{P})=\mathrm{f}(\mathbf{P}^2)=\mathrm{f}(\mathbf{P})^2,$$

因此 $\mathrm{f}(\mathbf{P})=0$ 或 1。

取 $k=-1$ 得 $\mathbf{K}^{-1}\mathbf{PK}=\mathbf{P}^{-1}$，所以 $\mathrm{f}(\mathbf{P}^{-1})=\mathrm{f}(\mathbf{P})$。但 $\mathrm{f}(\mathbf{P})\,\mathrm{f}(\mathbf{P}^{-1})=\mathrm{f}(\mathbf{I})=1$，所以

$$\mathrm{f}(\mathbf{P})^2=1,$$

特别地 $\mathrm{f}(\mathbf{P})\neq 0$。综合两个结果，

$$\mathrm{f}(\mathbf{P})=1.$$

验算

行列式满足所有假设：它是可乘的，$\det\mathbf{I}=1$ 而且 $\det\mathbf{J}=-1$。它亦与我们的所有结论吻合：交换两行或两列会改变其正负号，两行成比例的矩阵的行列式为 0，以及 $\det\mathbf{P}=1$。一个较不明显的例子是 $\mathrm{f}(\mathbf{M})=\operatorname{sgn}(\det\mathbf{M})\,|\det\mathbf{M}|^{0.7}$。我们用电脑对数以千计的随机矩阵作检验，证实这两个函数都满足可乘法则、行与列的交换恒等式、两行成比例时 $\mathrm{f}(\mathbf{A})=0$，以及 $\mathrm{f}(\mathbf{P})=1$。

常见错误

- 假设 $\mathrm{f}=\det$。题目要求你单凭可乘性推导出这些性质；行列式只是一个参考。
- 未加解释便使用 $\mathrm{f}(\mathbf{K}^{-1})=1/\mathrm{f}(\mathbf{K})$。这需要第 (i) 部分的 $\mathrm{f}(\mathbf{I})=1$，以及 $k\neq 0$，使 $\mathbf{K}$ 可逆。
- 停在 $\mathrm{f}(\mathbf{P})^2=\mathrm{f}(\mathbf{P})$，忘记排除 $\mathrm{f}(\mathbf{P})=0$ 的情况。

STEP 之外

像 f 这样保持乘法的映射，是从 2×2 可逆矩阵所成的群（group）到非零实数的一个同态（homomorphism）。第 (iv) 部分是一个经典论证的开端：像 $\mathbf{P}$ 这样的错切矩阵（shear matrix）都是交换子（commutator）的乘积，所以每个这样的同态都把它们映射到 1，最终 f 只取决于行列式。你亦可以证明：行列式是唯一一个对每一行都是线性的、交错的（alternating），并且在 $\mathbf{I}$ 处等于 1 的函数。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）正是从这类性质出发建构行列式，而 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则涵盖同态的内容。

详细解答

STEP II 2019，第 9 题

力学 · 抛体与原点的距离

STEP II 2019，第 9 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 11.10 节会朝你飞回来的球中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/coming-back.html

关键思路

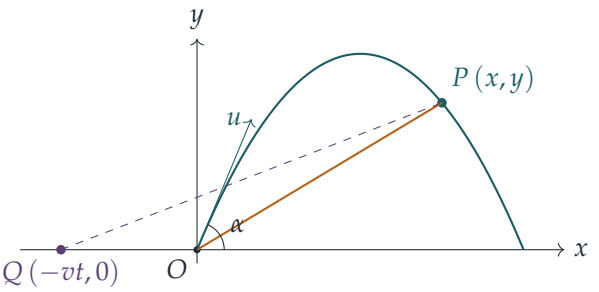
距离本身不易求导，但距离的平方却很容易：OP 递增，当且仅当 $\frac{\mathrm{d}}{\mathrm{d}t}(OP^2) > 0$。对 $x^2 + y^2$ 求导（differentiate），会得到 t 乘以一个关于 t 的二次式，于是整道题就变成“这个二次式何时为正?”，而这可由它的判别式（discriminant）回答。第 (ii) 部分是同一个计算，只是在水平距离上加入 Q 的运动。

解答

把 P 和 Q 视为质点（particle），忽略空气阻力，并以 O 为原点，x 轴取水平方向（沿 P 的运动方向），y 轴取铅直向上。记 $s = \sin\alpha$ 及 $c = \cos\alpha$，由于 $0 < \alpha < \frac{1}{2}\pi$，两者皆为正。在时间 t，

$$x = uct, \qquad y = ust - \tfrac{1}{2}gt^2,$$

而当 $y = 0$ 再次成立时 P 落地，时间为 $T = \dfrac{2us}{g}$。



第 (i) 部分

设 $r = OP$，则 $r^2 = x^2 + y^2$；利用 $\dot{x} = uc$ 及 $\dot{y} = us - gt$，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\right) = 2(x\dot{x} + y\dot{y}) &= 2\left[u^2c^2t + \left(ust - \tfrac{1}{2}gt^2\right)(us - gt)\right] \\ &= 2t\left[u^2c^2 + u^2s^2 - \tfrac{3}{2}usgt + \tfrac{1}{2}g^2t^2\right] \\ &= t\left(g^2t^2 - 3usg\,t + 2u^2\right),\end{aligned}$$

这里用了 $c^2 + s^2 = 1$。由于飞行期间 $r > 0$，$\frac{\mathrm{d}}{\mathrm{d}t}(r^2) = 2r\dot{r}$ 与 $\dot{r}$ 同号；又由于 $t > 0$，其符号与下式相同：

$$f(t) = g^2t^2 - 3usg\,t + 2u^2.$$

这个二次式的首项系数为正，判别式为

$$9u^2s^2g^2 - 8u^2g^2 = u^2g^2\left(9s^2 - 8\right).$$

若 $s < \frac{2\sqrt{2}}{3}$，则 $9s^2 < 8$（因为 $s > 0$），故判别式为负，对所有 t 皆有 $f(t) > 0$。因此在整个飞行过程中 $\dot{r} > 0$：OP 递增。

若 $s > \frac{2\sqrt{2}}{3}$，则判别式为正，f 有两个相异的根

$$t_{1,2} = \frac{u\left(3s \mp \sqrt{9s^2 - 8}\right)}{2g},$$

且当 $t_1 < t < t_2$ 时 $f(t) < 0$。两根皆为正（它们的和 $3us/g$ 与积 $2u^2/g^2$ 皆为正）。我们必须检查这段区间确实出现在 P 仍在空中的时候。

想一想 这个二次式在某处变为负值，但我们怎样知道这发生在 P 落地之前？

计算 f 在落地时刻的值：

$$f(T) = g^2 \cdot \frac{4u^2s^2}{g^2} - 3usg \cdot \frac{2us}{g} + 2u^2 = 2u^2(1 - s^2) = 2u^2c^2 > 0.$$

所以 T 不在两根之间。但 f 的顶点位于 $t = \frac{3us}{2g} < \frac{2us}{g} = T$，所以 T 在顶点右方；既然它不在 $[t_1, t_2]$ 之内，就必然有 $T > t_2$。因此 $0 < t_1 < t_2 < T$，而当 $t_1 < t < t_2$ 时 $\dot{r} < 0$：OP 在 P 落地前的某段时间递减，得证。

第 (ii) 部分

Q 在光滑地面上以恒定速率 v 朝远离 P 那一侧的方向滑行，所以在时间 t 它位于 $(-vt, 0)$。P 与 Q 的水平距离为 $x + vt = wt$，其中

$$w = uc + v > 0,$$

而铅直距离为 y。所以 $QP^2 = w^2t^2 + y^2$，与之前完全一样，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(QP^2\right) &= 2\left[w^2t + \left(ust - \tfrac{1}{2}gt^2\right)(us - gt)\right] \\ &= t\left(g^2t^2 - 3usg\,t + 2w^2 + 2u^2s^2\right).\end{aligned}$$

飞行期间其符号与 $h(t) = g^2t^2 - 3usg\,t + 2(w^2 + u^2s^2)$ 相同，而 h 的判别式为

$$9u^2s^2g^2 - 8g^2\left(w^2 + u^2s^2\right) = g^2\left(u^2s^2 - 8w^2\right).$$

这个判别式为负，当且仅当 $8w^2 > u^2s^2$，即（由于 w 和 us 皆为正）$2\sqrt{2}\,w > us$。代入 $w = uc + v$，

$$2\sqrt{2}\,(uc + v) > us \iff 2\sqrt{2}\,v > \left(\sin\alpha - 2\sqrt{2}\cos\alpha\right)u.$$

所以在题设条件下，对所有 t 皆有 $h(t) > 0$，因此当 $t > 0$ 时 $\frac{\mathrm{d}}{\mathrm{d}t}(QP^2) > 0$，$QP$ 在 P 的整个飞行过程中递增，得证。

与第 (i) 部分相同的论证（现在 $h(T) = 2w^2 > 0$）显示，当不等式严格反转时，QP 确实会在 P 落地前递减一段时间，所以除了等号成立的边界情况外，这个条件其实也是必要条件。

验算

- 在第 (ii) 部分中令 $v = 0$，得 $2\sqrt{2}\cos\alpha > \sin\alpha$，即 $\tan\alpha < 2\sqrt{2}$，即 $\sin\alpha < \frac{2\sqrt{2}}{3}$：正是第 (i) 部分的结果，理应如此，因为此时 Q 一直停在 O。
- 临界角为 $\arcsin\frac{2\sqrt{2}}{3} \approx 70.5°$，与 u 和 g 无关，这正是量纲分析（dimensional analysis）所要求的（唯一可用的无量纲量是 α）。
- 取 $u = 10\,\mathrm{m\,s^{-1}}$、$g = 9.8\,\mathrm{m\,s^{-2}}$ 及 $\sin\alpha = 0.97$，两根为 $t_1 \approx 1.14\,\mathrm{s}$ 和 $t_2 \approx 1.83\,\mathrm{s}$，都在落地时刻 $T \approx 1.98\,\mathrm{s}$ 之前。用电脑逐步模拟整个飞行过程，证实 OP 在这两个时刻之间减少；而且在大量随机的 u、α、v 之下，距离 QP 递增恰好与第 (ii) 部分的不等式成立相吻合。

常见错误

- 在第 (i) 部分只停留于“判别式为正”：你必须证明 OP 的下降发生在 P 落地之前，而不是之后。
- 把 Q 放在 $(vt, 0)$：Q 是朝远离 P 那一侧移动，所以它与 P 的水平距离是 $uct + vt$。
- 漏掉因子 t，或者约去它时没有指出 $t > 0$。

STEP 之外

研究 $\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{r}|^2 = 2\,\mathbf{r}\cdot\dot{\mathbf{r}}$ 的符号，是判断物体正在接近还是远离的标准方法；它在轨道力学（orbital mechanics）和两物体的相对运动（relative motion）中一再出现，而第 (ii) 部分其实就是关于 P 相对于 Q 的位置的问题。一个抛射体若以比约 70.5° 更陡的角度抛出，在飞行的一段时间内会朝抛掷者回来——这个巧妙的事实是力学教科书的热门例子。想了解更多，可参阅 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP II 2019，第 10 题

力学 · 刚体平衡

STEP II 2019，第 10 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

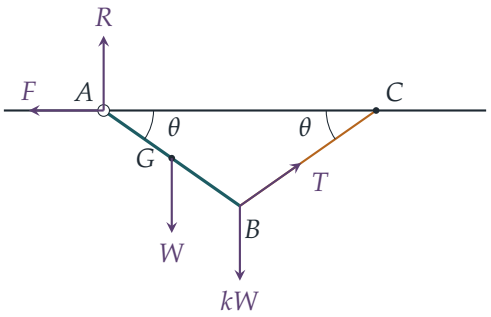
关键思路

由于 $AB = BC = 2a$，而且 A、C 都在轨道上，三角形 ABC 是等腰三角形（isosceles triangle），所以绳子与轨道所成的角也是 θ。掌握了这个几何关系后，这就是一道标准的刚体（rigid body）平衡（equilibrium）问题：对 A 取力矩（moments），可以求出张力（tension）而不牵涉环上的未知力；然后再分解（resolving）各力，便得到法向反作用力（normal reaction）和摩擦力（friction）。最后一部分比较 k 的两个临界值：当 k 增大时，哪一个会先被达到？

解答

模型与受力。 杆是均匀的刚体，所以其重量 W 作用于中点 G，且 $AG = a$。环是轻的，所以轨道对环的力会直接传到杆的 A 端：一个铅垂向上的法向反作用力 R，以及一个沿轨道方向的摩擦力 F。绳子是轻而不可伸长的，张力 T 在 B 点沿 BC 方向作用。额外负载 kW 在 B 点铅垂向下作用。

在三角形 ABC 中，AB 和 BC 两边都等于 $2a$，所以 A 和 C 处的底角相等：绳子与轨道所成的角是 θ。在水平方向上，B 与 A 相距 $2a\cos\theta$，C 与 A 相距 $4a\cos\theta$。



我们设 F 沿轨道指向远离 C 的方向；稍后会看到它必然如此，因为绳子把杆拉向 C。

第 (i) 部分

对 A 取力矩，这样便消去了 R 和 F。重量和负载的力臂分别是 $a\cos\theta$ 和 $2a\cos\theta$。至于张力，把它在 B 点分解为水平分量 $T\cos\theta$（指向 C）和铅垂向上的分量 $T\sin\theta$。铅垂分量的力臂是 $2a\cos\theta$，水平分量的力臂是 $2a\sin\theta$（即 B 在轨道下方的深度），而两者都使杆向同一方向转动，即图中的逆时针方向。所以

$$T\sin\theta\cdot 2a\cos\theta + T\cos\theta\cdot 2a\sin\theta = W\cdot a\cos\theta + kW\cdot 2a\cos\theta,$$
$$4aT\sin\theta\cos\theta = aW(1+2k)\cos\theta.$$

由于当 $0<\theta<90^\circ$ 时 $\cos\theta\neq 0$，我们可以除以 $a\cos\theta$：

$$T=\frac{(2k+1)W}{4\sin\theta}.$$

只要环没有滑动，不论环的情况如何，这个结果都成立。若 $T>W$，绳子便会断裂，即 $(2k+1)W>4W\sin\theta$（两边乘以 $4\sin\theta>0$），亦即

$$2k+1>4\sin\theta,$$

此即所求。

第 (ii) 部分

对整条杆分解各力。铅垂方向：

$$R+T\sin\theta = W+kW \quad\Longrightarrow\quad R=(1+k)W-\frac{(2k+1)W}{4}=\frac{(2k+3)W}{4}.$$

水平方向上，只有 F 和张力的水平分量，所以

$$F=T\cos\theta=\frac{(2k+1)W\cos\theta}{4\sin\theta}.$$

这是正数，证实了摩擦力的方向是远离 C。另外留意 $R>0$，所以环确实是从上方压在轨道上。

想一想 粗糙轨道上的环，究竟在甚么时候会滑动?

摩擦力最多只能提供 μR。若平衡需要 $F > \mu R$，摩擦力便无法令环保持静止，环就会滑动。所以环滑动的条件是

$$\frac{(2k+1)W\cos\theta}{4\sin\theta} > \mu\,\frac{(2k+3)W}{4}.$$

两边乘以 $4\sin\theta/(W\cos\theta) > 0$，得

$$2k+1 > (2k+3)\mu\tan\theta,$$

此即所求。

第 (iii) 部分

记 $t = \tan\theta$ 及 $s = \sin\theta$。每个条件都表示 k 已超过某个临界值。

绳子。由第 (i) 部分，当 $2k+1 > 4s$ 时绳子断裂，即当

$$k > k_{\text{断裂}} = \frac{4s-1}{2}.$$

环。把第 (ii) 部分的条件重新整理:

$$2k+1 > (2k+3)\mu t \iff 2k(1-\mu t) > 3\mu t - 1.$$

这里用到 $\mu t < 1$: 除以 $2(1-\mu t) > 0$ 不会改变不等式（inequality）的方向，所以当

$$k > k_{\text{滑动}} = \frac{3\mu t - 1}{2(1-\mu t)}.$$

时环便会滑动。（若没有 $\mu t < 1$ 这个条件，环永远不会滑动。若 $\mu t \geq 1$，则对所有 k 都有 $(2k+3)\mu t \geq 2k+3 > 2k+1$。这正是题目给出此条件的原因。）

当 k 从零开始增大时，较低的临界值会先被越过。所以环在绳子断裂之前滑动，当且仅当 $k_{\text{滑动}} < k_{\text{断裂}}$。先乘以 $1-\mu t > 0$，

$$\begin{aligned}
\frac{3\mu t-1}{1-\mu t} < 4s-1 &\iff 3\mu t - 1 < (4s-1)(1-\mu t)\\
&\iff 3\mu t - 1 < 4s - 1 - 4s\mu t + \mu t\\
&\iff 2\mu t + 4s\mu t < 4s\\
&\iff \mu t(1+2s) < 2s\\
&\iff \mu < \frac{2\sin\theta}{\tan\theta\,(1+2\sin\theta)} = \frac{2\cos\theta}{1+2\sin\theta},
\end{aligned}$$

最后一步除以了 $t(1+2s) > 0$。这就是所求的条件。

杆起初处于平衡状态，所以在 $k = 0$ 时两个临界值都尚未被越过。由于每个条件都是“k 大于某个值”，两个临界值都只会被越过一次。所以比较两个临界值，确实能决定哪件事先发生。

验算

- 改为对 B 取力矩，会得到相同的答案。A 处的力对 B 的力矩是 $R \cdot 2a\cos\theta - F \cdot 2a\sin\theta$，而重量的力矩是 $Wa\cos\theta$。代入 R 和 F 的值，得 $\frac{1}{2}(2k+3)Wa\cos\theta - \frac{1}{2}(2k+1)Wa\cos\theta = Wa\cos\theta$，恰好与重量的力矩平衡。
- 取 $\theta = 30°$。那么 $k_{\text{断裂}} = \frac{1}{2}$，而临界摩擦系数（coefficient of friction）是 $2\cos 30°/(1+2\sin 30°) = \frac{\sqrt{3}}{2} \approx 0.866$，所以 $\mu t = 0.866 \times 0.577 = \frac{1}{2}$。在这个 μ 下，$k_{\text{滑动}} = \frac{3/2-1}{2\cdot\frac{1}{2}} = \frac{1}{2}$ 也成立，所以两个临界值重合，正如在边界上应有的情况。
- 我们用电脑验证了，对许多随机的 θ 和 k，所有力和力矩都平衡。我们亦对随机的 θ 和 μ 把 k 逐小步增大：环先滑动，当且仅当 $\mu < 2\cos\theta/(1+2\sin\theta)$。

常见错误

- 在对 A 取力矩时忘记了张力的水平分量。它的力臂是 $2a\sin\theta$，并令张力的力矩加倍，$4\sin\theta$ 中的 4 正是由此而来。
- 取 $R = (1+k)W$。绳子承担了部分重量，所以法向反作用力小于总负载。
- 除以 $1 - \mu\tan\theta$ 时没有说明它是正数。题目给出的条件 $\mu\tan\theta < 1$ 正是为此而设。

STEP 之外

这是一个静定（statically determinate）问题：有三个未知量（T、R、F），而平面上的刚体有三条平衡方程。第 (iii) 部分展示了这类结构典型的失效方式：比较每个约束分别在多大负载下失效。工程师也会问同样的问题：结构中哪一个构件会最先失效？Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）仔细讲解刚体静力学和扭矩（torque），并进而讨论转动动力学（rotational dynamics）。

详细解答

STEP II 2019，第 11 题

概率与统计 · 三角形的计数

STEP II 2019，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

三条长度能构成面积非零的三角形（triangle），当且仅当较短的两条长度之和大于最长的一条。因此，整个问题归结为在最大值固定的情况下数算数对的个数；最简洁的数法，是固定中间的数 n_2，再问 n_1 可以取多少个值。第 (i) 部分负责计数；第 (ii) 和第 (iii) 部分只需除以相应的等可能（equally likely）选择总数。

解答

第 (i) 部分

记 $m = n_3$。固定 $n_2 = k$，其中 $k \leq m-1$。那么 n_1 必须满足

$$m - k < n_1 < k,$$

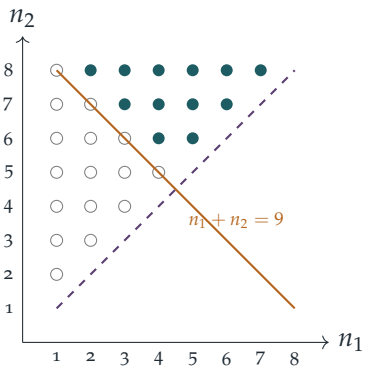
所以 n_1 取遍 $m-k+1, \ldots, k-1$，当 $2k-m-1$ 为正数时，共有 $2k-m-1$ 个选择。（由于 $m-k \geq 1$，条件 $n_1 > 0$ 自动成立。）

当 $m = 9$ 时，个数为 $2k-10$，在 $k = 6,7,8$ 时为正数：

$$2+4+6 = 12.$$

当 $m = 10$ 时，个数为 $2k-11$，在 $k = 6,7,8,9$ 时为正数：

$$1+3+5+7 = 16.$$



上图显示 $n_3 = 9$ 的情况：实心点位于直线 $n_1 + n_2 = 9$ 上方，同时位于对角线 $n_1 = n_2$ 上方，共有 $2 + 4 + 6 = 12$ 个，每个 n_2 对应三角形区域中的一列。

奇数情况， $n_3 = 2n + 1$。当 $n_2 = k$ 时，个数为 $2k - 2n - 2 = 2(k - n - 1)$，在 $k = n + 2, \ldots, 2n$ 时为正数。设 $j = k - n - 1$，它由 1 取到 $n - 1$，

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

偶数情况， $n_3 = 2n$。当 $n_2 = k$ 时，个数为 $2k - 2n - 1$，在 $k = n + 1, \ldots, 2n - 1$ 时为正数。设 $j = k - n$，它由 1 取到 $n - 1$，

$$\sum_{j=1}^{n-1} (2j-1) = (n-1)^2,$$

即首 $n - 1$ 个奇数之和。两条公式都与 $12 = 4 \cdot 3$ 及 $16 = 4^2$ 吻合。

第 (ii) 部分

长度为 N 的棒是三条中最长的，所以这三条棒构成面积非零的三角形，当且仅当所选的两条长度 $a < b$ 满足 $a + b > N$。（其余两条三角不等式（triangle inequality），例如 $a + N > b$，由于 $N > b$ 而自动成立。）因此，有利的选择正是第 (i) 部分中 $n_3 = N$ 时所数算的数对。

两条棒是从余下的 $N - 1$ 条中选出的，所以共有 $\binom{N-1}{2}$ 个等可能的选择。

当 $N = 2n + 1$ 时：共有 $\binom{2n}{2} = n(2n-1)$ 个选择，所以

$$\mathrm{P}(\text{三角形}) = \frac{n(n-1)}{n(2n-1)} = \frac{n-1}{2n-1},$$

此即所求。

当 $N = 2n$ 时：共有 $\binom{2n-1}{2} = (2n-1)(n-1)$ 个选择。要使它非零，我们需要 $n \geq 2$（当 $N = 2$ 时只剩一条棒，无法选出两条）。于是

$$\mathrm{P}(\text{三角形}) = \frac{(n-1)^2}{(2n-1)(n-1)} = \frac{n-1}{2n-1}.$$

颇出人意料的是，尽管 N 小了一，答案却与奇数情况的公式完全相同。

第 (iii) 部分

任意三条不同的棒都有一条最长的，设其长度为 m，其中 $3 \leq m \leq 2M+1$。这个选择能构成三角形，当且仅当另外两条长度构成第 (i) 部分中 $n_3 = m$ 时所数算的其中一个数对。每个有利选择都只在其最长的棒之下被数算一次，所以有利选择的个数为

$$T = \sum_{m=2}^{2M+1} c(m),$$

其中 $c(m)$ 是第 (i) 部分得出的个数。（由于 $c(2) = 0$，把 $m = 2$ 包括在内并无影响。）

> **想一想** $c(m)$ 对奇数和偶数 m 有不同的公式，那该怎样把它们加起来呢?

把它们配对。对每个 $n = 1, \ldots, M$，把偶数值 $m = 2n$ 和奇数值 $m = 2n+1$ 放在一起；这样恰好把 $m = 2, 3, \ldots, 2M+1$ 各涵盖一次。每对的总贡献为

$$(n-1)^2 + n(n-1) = (n-1)(2n-1) = 2n^2 - 3n + 1.$$

利用题目给出的 $\sum k^2$ 公式及 $\sum k = \frac{1}{2}M(M+1)$，

$$\begin{aligned} T &= \sum_{n=1}^{M} \left(2n^2 - 3n + 1\right) \\ &= \tfrac{1}{3}M(M+1)(2M+1) - \tfrac{3}{2}M(M+1) + M \\ &= \tfrac{1}{6}M\left[2(2M^2+3M+1) - 9(M+1) + 6\right] \\ &= \tfrac{1}{6}M\left(4M^2 - 3M - 1\right) \\ &= \tfrac{1}{6}M(4M+1)(M-1). \end{aligned}$$

选出三条棒的等可能选择总数为

$$\binom{2M+1}{3} = \frac{(2M+1)(2M)(2M-1)}{6}.$$

相除，并约去 $M/6$，

$$\mathrm{P}(\text{三角形}) = \frac{M(4M+1)(M-1)}{2M(2M+1)(2M-1)} = \frac{(4M+1)(M-1)}{2(2M+1)(2M-1)},$$

此即所求。

验算

- $M = 1$：棒的长度为 $1, 2, 3$，而 $1 + 2 = 3$ 只得出退化三角形（degenerate triangle），所以概率应为 0，公式亦给出 0。
- $M = 2$：棒的长度为 1 至 5。能构成的三角形有 $\{2, 3, 4\}$、$\{2, 4, 5\}$ 和 $\{3, 4, 5\}$，所以概率为 $3/10$；公式给出 $\frac{9 \cdot 1}{2 \cdot 5 \cdot 3} = \frac{3}{10}$。
- 当 $M \to \infty$ 时，概率趋向 $\frac{4}{8} = \frac{1}{2}$。这与连续版本吻合：若三条长度独立地在 $[0, 1]$ 上均匀分布，较短两条之和大于最长一条的概率为 $\frac{1}{2}$。
- 第 (i) 部分的所有个数，以及第 (ii) 和第 (iii) 部分的所有概率，都已用电脑直接列举每个数对或三元组加以核实，范围为 n 至 39、M 至 19（例如 $M = 3$ 时得 $13/35$）。

常见错误

- 容许 $n_1 + n_2 = n_3$：长度为 $2, 3, 5$ 的三条棒只会平躺成一直线，面积为零，所以不等式必须是严格的。
- 在第 (ii) 部分用了 $\binom{N}{2}$ 而非 $\binom{N-1}{2}$：长度为 N 的棒已经被选走了。
- 在第 (iii) 部分只对 m 的奇数值（或只对偶数值）求和：最长的棒可以是由 3 至 $2M + 1$ 的任何长度。

STEP 之外

按最大元素把计数分拆，然后求和，是组合数学（combinatorics）中的标准技巧；同一思路对最大值的取值积分，便得出连续情况的答案 $\frac{1}{2}$。像上图那样数算某区域内的格点（lattice points），再与该区域的面积比较，正是连接离散概率与连续概率的桥梁。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）收录了许多这类问题，包括经典的“断棍”三角形问题。

详细解答

STEP II 2019，第 12 题

概率与统计 · 连续分布与分位数

STEP II 2019，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

这里的累积分布函数（cumulative distribution function）在 $[0,1]$ 上就是 x^n，所以每个分位数（quantile）都可以准确写出：分布中有比例 p 落在其下方的值是 $p^{1/n}$。于是，比较平均值与某个分位数就变成比较幂，而把两边同时取 n 次方后，问题便化为关于 $\left(1+\frac{1}{n}\right)^n$ 的命题，这正是二项式展开（binomial expansion）最擅长估计的量。

解答

第 (i) 部分

对概率密度函数（probability density function）积分，

$$\mu = \mathrm{E}(X) = \int_0^1 x \cdot nx^{n-1}\,\mathrm{d}x = \int_0^1 nx^n\,\mathrm{d}x = \frac{n}{n+1},$$
$$\mathrm{E}(X^2) = \int_0^1 nx^{n+1}\,\mathrm{d}x = \frac{n}{n+2}.$$

因此

$$\sigma^2 = \mathrm{E}(X^2) - \mu^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n\left[(n+1)^2 - n(n+2)\right]}{(n+1)^2(n+2)}$$
$$= \frac{n\left[n^2+2n+1-n^2-2n\right]}{(n+1)^2(n+2)} = \frac{n}{(n+1)^2(n+2)},$$

得证。

第 (ii) 部分

当 $0 \le x \le 1$ 时，分布函数为

$$\mathrm{F}(x) = \mathrm{P}(X \le x) = \int_0^x nt^{n-1}\,\mathrm{d}t = x^n.$$

取 $n = 2$, 则 $\mathrm{F}(x) = x^2$。下四分位数 (lower quartile) q_1 满足 $q_1^2 = \frac{1}{4}$，上四分位数（upper quartile）q_3 满足 $q_3^2 = \frac{3}{4}$；两者都在 $[0,1]$ 内，所以取正根：

$$q_1 = \tfrac{1}{2}, \qquad q_3 = \tfrac{\sqrt{3}}{2}, \qquad q_3 - q_1 = \frac{\sqrt{3}-1}{2}.$$

由第 (i) 部分，$\sigma^2 = \dfrac{2}{9\cdot 4} = \dfrac{1}{18}$，所以 $\sigma = \dfrac{1}{3\sqrt{2}} = \dfrac{\sqrt{2}}{6}$，而 $2\sigma = \dfrac{\sqrt{2}}{3}$。

我们要证明 $\dfrac{\sqrt{3}-1}{2} < \dfrac{\sqrt{2}}{3}$。两边乘以 6，这等价于

$$3\sqrt{3} - 3 < 2\sqrt{2} \iff 3\sqrt{3} < 3 + 2\sqrt{2}.$$

最后这个不等式两边都是正数，所以它等价于两边平方后的不等式：

$$27 < 9 + 12\sqrt{2} + 8 = 17 + 12\sqrt{2} \iff 10 < 12\sqrt{2},$$

这是成立的，因为 $\sqrt{2} > 1$ 给出 $12\sqrt{2} > 12 > 10$。每一步都是等价的，所以沿着这条推理链倒推回去，可知四分位数间距（interquartile range）小于 2σ。

> ? **想一想** 为甚么在这里两边平方是安全的，但一般情况下却不是?

只有当 a 和 b 都非负时，平方才能保持不等式 $a < b$（例如 $-3 < 2$，但 $9 > 4$）。正因如此，我们先把不等式整理成 $3\sqrt{3} < 3 + 2\sqrt{2}$，使两边明显都是正数，而不是直接把仍含有减法的 $3\sqrt{3} - 3 < 2\sqrt{2}$ 平方。

第 (iii) 部分

由二项式定理,

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

而第 $(k+1)$ 项是

$$\binom{n}{k} x^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\,x^k.$$

中位数与下四分位数。 由于 $\mathrm{F}(x) = x^n$，中位数（median）m 满足 $m^n = \frac{1}{2}$，下四分位数 q 满足 $q^n = \frac{1}{4}$，所以

$$m = \left(\tfrac{1}{2}\right)^{1/n}, \qquad q = \left(\tfrac{1}{4}\right)^{1/n}.$$

μ、m、q 全都是正数，而 $t \mapsto t^n$ 在 $t > 0$ 上严格递增（strictly increasing），所以我们可以通过比较它们的 n 次方来比较它们。另外

$$\mu^n = \left(\frac{n}{n+1}\right)^n = \frac{1}{\left(1+\frac{1}{n}\right)^n}.$$

记 $A = \left(1+\frac{1}{n}\right)^n$。那么

$$\mu < m \iff \mu^n < \tfrac{1}{2} \iff A > 2, \qquad \mu > q \iff \mu^n > \tfrac{1}{4} \iff A < 4.$$

μ 小于中位数。 在展开式中代入 $x = \frac{1}{n}$。每一项都非负，所以 A 至少等于首三项之和：

$$A \geq 1 + n \cdot \frac{1}{n} + \frac{n(n-1)}{2} \cdot \frac{1}{n^2} = 2 + \frac{n-1}{2n} > 2,$$

其中最后一个不等式是严格的，因为 $n > 1$。（当 $n \geq 2$ 时，展开式中确实有第三项。）因此 $\mu < m$。

μ 大于下四分位数。 取 $x = \frac{1}{n}$，第 $(k+1)$ 项是

$$\binom{n}{k}\frac{1}{n^k} = \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

因为那 k 个分数 $\frac{n-j}{n}$（$0 \leq j \leq k-1$）每个都在 0 与 1 之间。（当 $k = 0$ 时，该项是 $1 = \frac{1}{0!}$。）对 $k = 0, 1, \ldots, n$ 求和，

$$A \leq 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} < 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots < 4,$$

中间的不等式成立，是因为无穷级数（infinite series）还包含更多正项；最后一个不等式则是题目容许我们假设的结果。因此 $A < 4$，所以 $\mu > q$。

综合两者，对每个整数 $n > 1$，下四分位数都小于 μ，而 μ 又小于中位数。

> ? **想一想** 为甚么比较 n 次方就能确定 μ 与 m 的大小关系？

对正数 a 和 b，$a < b$ 若且唯若 $a^n < b^n$，因为 t^n 在 $t > 0$ 上严格递增。我们的三个数全是正数，所以取 n 次方不会损失任何信息，而且这样做能消去 m 和 q 中麻烦的 n 次方根。

验算

- 当 $n = 2$：$\mu = \frac{2}{3}$，而 $\sigma^2 = \frac{2}{9 \cdot 4} = \frac{1}{18}$，与 $\mathrm{E}(X^2) - \mu^2 = \frac{1}{2} - \frac{4}{9}$ 吻合。
- 数值上，当 $n = 2$ 时，四分位数间距是 0.366，而 $2\sigma = 0.471$。

- 当 $n=2$ 时，三个值分别是 $q=0.5$、$\mu\approx 0.667$、$m\approx 0.707$；当 $n=10$ 时，它们是 $q\approx 0.871$、$\mu\approx 0.909$、$m\approx 0.933$。对 2 至 2000 的每个 n，顺序 $q<\mu<m$ 都已用电脑验证。
- 当 $n\to\infty$ 时，$A=\left(1+\frac{1}{n}\right)^n\to \mathrm{e}\approx 2.718$，稳稳地落在 2 与 4 之间。

常见错误

- 没有先确认两边都是正数，就直接把 $3\sqrt{3}-3<2\sqrt{2}$ 平方。
- 取倒数时弄错不等号方向：$\mu^n<\frac{1}{2}$ 对应的是 $A>2$，而不是 $A<2$。
- 声称第 $(k+1)$ 项等于 $\frac{1}{k!}$，而不是至多为 $\frac{1}{k!}$。

STEP 之外

$\left(1+\frac{1}{n}\right)^n$ 随 n 递增并趋向极限 e；而你逐项与 $\sum 1/k!$ 比较所证出的界，正是证明 $\mathrm{e}=\sum_{k\geq 0}1/k!$ 的第一步，这是大学第一门数学分析（analysis）课程中的标准结果。这个分布本身是 Beta$(n,1)$ 分布（Beta distribution）；它也是 n 个在 $[0,1]$ 上独立均匀分布的随机变量之最大值的分布，因为 $\mathrm{P}(\max\leq x)=x^n$。想深入了解这两条线索，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》），以及 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2019，第 1 题

纯数学 · 耦合微分方程

> **STEP III 2019，第 1 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

当 $k = 1$ 时，方程组的右方正好是把 $x + \mathrm{i}y$ 乘以一个复数常数所得的结果，因此这对方程可以合并为关于 $z = x + \mathrm{i}y$ 的单一方程 $\dot{z} = cz$。这样一来，运动就变成旋转配合指数衰减（exponential decay）：一条螺线（spiral）。当 $k = 0$ 时，耦合只朝一个方向：x 不依赖 y，所以我们先解出 x，再把它代入 y 的方程。

解答

第 (i) 部分

当 $k = 1$ 时，我们有 $\dot{x} = -x - y$ 及 $\dot{y} = x - y$。设 $z = x + \mathrm{i}y$，则

$$\begin{aligned}\dot{z} = \dot{x} + \mathrm{i}\dot{y} &= (-x - y) + \mathrm{i}(x - y) \\ &= (-1 + \mathrm{i})(x + \mathrm{i}y) = (-1 + \mathrm{i})z,\end{aligned}$$

这可以通过展开 $(-1 + \mathrm{i})(x + \mathrm{i}y) = -x - \mathrm{i}y + \mathrm{i}x - y$ 来确认。由 $z(0) = 1$，解为

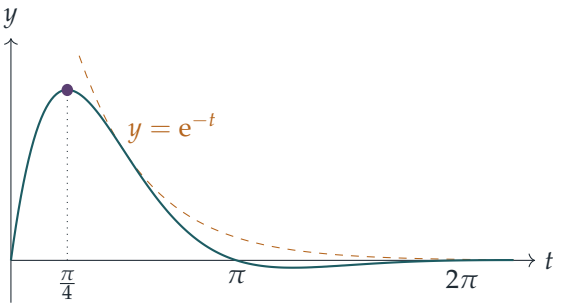
$$z = \mathrm{e}^{(-1+\mathrm{i})t} = \mathrm{e}^{-t}(\cos t + \mathrm{i}\sin t),$$

因此

$$x = \mathrm{e}^{-t}\cos t, \qquad y = \mathrm{e}^{-t}\sin t.$$

（如果你想避免使用复数（complex numbers）：把第一条方程微分，再消去 y，得到 $\ddot{x} + 2\dot{x} + 2x = 0$，其辅助方程（auxiliary equation）的根为 $-1 \pm \mathrm{i}$，然后利用 $x(0) = 1$、$\dot{x}(0) = -1$。）

y 对 t 的图像。这是一条位于包络线（envelope）$\pm e^{-t}$ 之内的正弦波，所以是一个阻尼振荡（damped oscillation）。它在 $t = n\pi$ 处为零，而其转向点（turning point）位于 $\dot{y} = x - y = e^{-t}(\cos t - \sin t) = 0$ 之处，即 $t = \frac{\pi}{4} + n\pi$。每个相继的峰值都缩小 $e^{-\pi} \approx 0.04$ 倍，因此在第一个隆起之后，若按真实比例绘画，振荡几乎看不见。



路径。以极式表示，$z = e^{-t}e^{it}$ 的模（modulus）为 $r = e^{-t}$，辐角（argument）为 $\theta = t$。因此路径是螺线

$$r = e^{-\theta}, \qquad \theta \geq 0,$$

从 $(1,0)$ 出发，以逆时针方向卷向原点，只有在 $t \to \infty$ 时才到达原点。起始时 $\dot{x} = -1$ 及 $\dot{y} = 1$，所以它离开 $(1,0)$ 时向左上方移动。每转一圈，与原点的距离便缩小 $e^{-2\pi} \approx 0.002$ 倍，因此按真实比例绘画时只看得见第一圈。

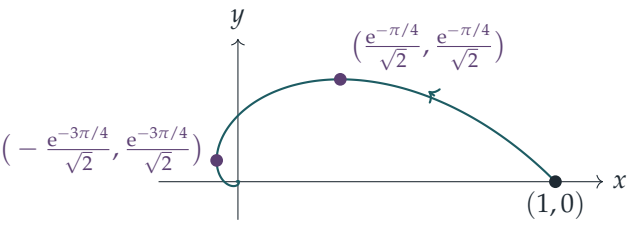
y 的最大值。y 的转向点位于 $t = \frac{\pi}{4} + n\pi$。在 $t = \frac{\pi}{4}$ 时，$y = e^{-\pi/4}/\sqrt{2} > 0$。对所有 $t \geq \frac{\pi}{2}$，我们有 $y \leq e^{-t} \leq e^{-\pi/2}$，而 $e^{-\pi/2} < e^{-\pi/4}/\sqrt{2}$，因为这等价于 $\sqrt{2} < e^{\pi/4} \approx 2.19$。在 $[0, \frac{\pi}{2}]$ 上唯一的转向点是 $t = \frac{\pi}{4}$。所以 y 的最大值出现在

$$\left(\frac{e^{-\pi/4}}{\sqrt{2}}, \frac{e^{-\pi/4}}{\sqrt{2}}\right) \approx (0.322,\ 0.322).$$

x 的最小值。现在 $\dot{x} = -x - y = -e^{-t}(\cos t + \sin t)$，它在 $\tan t = -1$ 时为零，即 $t = \frac{3\pi}{4} + n\pi$。要使 $x < 0$，需要 $\cos t < 0$，这就选出 $t = \frac{3\pi}{4} + 2n\pi$，得 $x = -e^{-t}/\sqrt{2}$；由于 e^{-t} 是递减的，所以在这些 t 中取最小者时 x 最负，即 $t = \frac{3\pi}{4}$。（最小值必定出现在其中一个转向点，因为 $x(0) = 1 > 0$，而且当 $t \to \infty$ 时 $x \to 0$。）所以 x 的最小值出现在

$$\left(-\frac{e^{-3\pi/4}}{\sqrt{2}}, \frac{e^{-3\pi/4}}{\sqrt{2}}\right) \approx (-0.067,\ 0.067).$$

两点都位于直线 $y = \pm x$ 上，这正如预期：在螺线上，y 在路径朝 x 方向前进之处取最大值，而 x 在路径朝 y 方向前进之处取最小值。



第 (ii) 部分

当 $k=0$ 时，第一条方程单独就是 $\dot{x}=-x$，所以由 $x(0)=1$，

$$x=e^{-t}.$$

于是 $\dot{y}+y=e^{-t}$。乘以积分因子 (integrating factor) e^{t}，得 $\frac{d}{dt}\left(ye^{t}\right)=1$，所以 $ye^{t}=t+c$，再由 $y(0)=0$ 得 $c=0$：

$$y=te^{-t}.$$

想一想 怎样把它化为路径的方程?

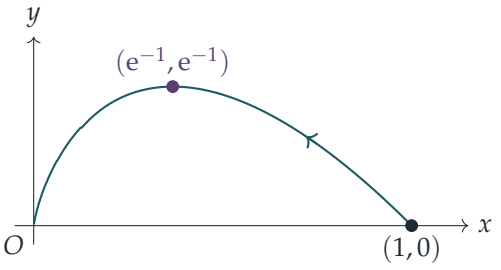
消去 t：由于 $x=e^{-t}$，我们有 $t=-\ln x$，因此

$$y=-x\ln x, \qquad 0<x\leq 1.$$

当 t 由 0 增至 ∞ 时，x 由 1 递减趋向 0，所以质点从 $(1,0)$ 出发，沿这条曲线向左移动，当 $t\to\infty$ 时趋近原点。

绘图要点：

- $\frac{dy}{dx}=-\ln x-1$，在 $x=e^{-1}$ 时为零。所以最高点是 (e^{-1},e^{-1})，在 $t=1$ 时到达。
- 在 $(1,0)$ 处斜率为 -1，与 $t=0$ 时的 $\dot{x}=-1$、$\dot{y}=1$ 吻合。
- 当 $x\to 0^{+}$ 时，$x\ln x\to 0$，所以 $y\to 0$；但斜率 $-\ln x-1\to+\infty$，所以曲线进入原点时与 y 轴相切。
- 对 $0<x<1$ 有 $y>0$，所以整条路径都在 x 轴之上。



留意与第 (i) 部分的对比：当 $k = 0$ 时不再有任何旋转，所以质点从不穿越坐标轴，只是沿曲线弯入原点。

验算

- 代回原方程：对于 $x = \mathrm{e}^{-t}\cos t$，$\dot{x} = -\mathrm{e}^{-t}\cos t - \mathrm{e}^{-t}\sin t = -x - y$，符合要求；$\dot{y}$ 的验算方法相同。两组解亦都满足 $x(0) = 1$、$y(0) = 0$。
- 以细小时间步长对方程作数值推进，可重现两组解，并确认在第 (i) 部分中 y 的最大值约为 0.3224，出现在 $t \approx 0.785$；x 的最小值约为 -0.0670，出现在 $t \approx 2.356$；而在第 (ii) 部分中 y 在 $t = 1$ 时达到峰值 $1/\mathrm{e} \approx 0.3679$。
- 对 (ii) 的合理性检查：$y = t\mathrm{e}^{-t}$ 是熟悉的“重根”形式，这是合理的，因为该方程组的矩阵有重特征值（repeated eigenvalue）-1。

常见错误

- 把 x 的最小值取在 $t = -\frac{\pi}{4}$ 或 $\frac{7\pi}{4}$：那里同样有 $\dot{x} = 0$，但那里 $x > 0$（是局部极大值）。要检查 $\cos t$ 的正负号。
- 忘记螺线只有在 $t \to \infty$ 时才到达原点，或把它画成顺时针卷入：$\theta = t$ 递增，所以它是逆时针卷入的。
- 在 (ii) 中漏掉原点处的铅垂切线：$y \to 0$ 但 $\mathrm{d}y/\mathrm{d}x \to \infty$。

STEP 之外

任何线性方程组（linear system）$\dot{\mathbf{x}} = A\mathbf{x}$ 的行为都由 A 的特征值决定。这里 $A = \begin{pmatrix} -1 & -k \\ 1 & -1 \end{pmatrix}$ 的特征值是 $-1 \pm \sqrt{-k}$：当 $k = 1$ 时它们是一对复数 $-1 \pm \mathrm{i}$，给出稳定螺旋点（stable spiral）；当 $k = 0$ 时有重特征值 -1，给出退化结点（degenerate node）；而当 $k < 0$ 时则会得到两个实特征值。以这种方法对不动点（fixed point）分类（即相平面分析（phase-plane analysis））是大学一年级的核心课题。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）对此有详细讲解，并附有大量图像。

详细解答

STEP III 2019，第 2 题

纯数学 · 函数方程

> **STEP III 2019，第 2 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 3.8 节由加法公式决定的函数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/addition-laws.html

关键思路

函数方程（functional equation）有一个特点：只要知道函数在某一点附近的表现，就能确定整个函数。代入 $x = y = 0$ 可以确定函数在 0 的值；然后在一般的 x 处写出差商（difference quotient），再利用函数方程把它化成在 0 处的差商，而我们知道这个差商的极限（limit）是 k。这样便得到一条微分方程（differential equation），再配合先前求得的初始值（initial value）把它解出来。

解答

第 (i) 部分

在 0 的值。在函数方程中代入 $x = y = 0$：$f(0) = f(0)^2$，所以 $f(0) = 0$ 或 $f(0) = 1$。

假设 $f(0) = 0$。代入 $y = 0$，对每个 x 都有 $f(x) = f(x)f(0) = 0$，所以 f 是零函数，而 $f'(0) = 0$。这与 $f'(0) = k \neq 0$ 矛盾。因此 $f(0) = 1$。

导数（derivative）。由 $(*)$ 及取 $y = h$ 的函数方程，

$$f'(x) = \lim_{h\to 0} \frac{f(x)f(h) - f(x)}{h} = f(x) \lim_{h\to 0} \frac{f(h) - 1}{h}.$$

由于 $f(0) = 1$，最后的极限就是 $\lim_{h\to 0} \frac{f(0+h) - f(0)}{h}$，根据 $(*)$ 在 $x = 0$ 的情况，它等于

$\mathrm{f}'(0) = k$。所以

$$\mathrm{f}'(x) = k\,\mathrm{f}(x),$$

即为所求。（我们可以把 $\mathrm{f}(x)$ 提到极限之外，因为它与 h 无关。）

求解。与其除以 f（这需要 $\mathrm{f} \neq 0$），不如考虑 $u(x) = \mathrm{e}^{-kx}\mathrm{f}(x)$。于是

$$u'(x) = \mathrm{e}^{-kx}\big(\mathrm{f}'(x) - k\,\mathrm{f}(x)\big) = 0,$$

所以 u 是常数，等于 $u(0) = \mathrm{f}(0) = 1$。因此

$$\mathrm{f}(x) = \mathrm{e}^{kx}.$$

这确实满足所有条件：$\mathrm{e}^{k(x+y)} = \mathrm{e}^{kx}\mathrm{e}^{ky}$，而且它在 0 的导数是 k。

第 (ii) 部分

在 0 的值。代入 $x = y = 0$:

$$\mathrm{g}(0) = \frac{2\mathrm{g}(0)}{1+\mathrm{g}(0)^2} \quad\Longrightarrow\quad \mathrm{g}(0)\big(\mathrm{g}(0)^2 - 1\big) = 0,$$

所以 $\mathrm{g}(0) \in \{0, 1, -1\}$。由于对所有 x 都有 $|\mathrm{g}(x)| < 1$，必然有 $\mathrm{g}(0) = 0$。另外留意，$|\mathrm{g}| < 1$ 保证了 $1 + \mathrm{g}(x)\mathrm{g}(y) > 0$，所以函数方程永远不会出现除以零的情况。

差商。利用取 $y = h$ 的函数方程，并把各项通分，

$$\mathrm{g}(x+h) - \mathrm{g}(x) = \frac{\mathrm{g}(x) + \mathrm{g}(h) - \mathrm{g}(x) - \mathrm{g}(x)^2\mathrm{g}(h)}{1 + \mathrm{g}(x)\mathrm{g}(h)} = \frac{\mathrm{g}(h)\big(1 - \mathrm{g}(x)^2\big)}{1 + \mathrm{g}(x)\mathrm{g}(h)}.$$

除以 h,

$$\frac{\mathrm{g}(x+h) - \mathrm{g}(x)}{h} = \big(1 - \mathrm{g}(x)^2\big) \cdot \frac{\mathrm{g}(h)}{h} \cdot \frac{1}{1 + \mathrm{g}(x)\mathrm{g}(h)}.$$

? **想一想** 当 $h \to 0$ 时，右边每个因子会怎样变化?

第一个因子不涉及 h。由于 $\mathrm{g}(0) = 0$, 中间的因子就是 $\dfrac{\mathrm{g}(0+h) - \mathrm{g}(0)}{h}$, 它趋向 $\mathrm{g}'(0) = k$。至于最后的因子,我们需要 $\mathrm{g}(h) \to \mathrm{g}(0) = 0$:这是成立的,因为 g 在 0 可微(differentiable),因此在 0 连续（continuous)。所以最后的因子趋向 1，而

$$\mathrm{g}'(x) = k\big(1 - \mathrm{g}(x)^2\big).$$

求解。因为 $|\mathrm{g}(x)| < 1$，因子 $1 - \mathrm{g}(x)^2$ 永不为零，所以我们可以分离变量（separation of variables)：

$$\int \frac{\mathrm{d}g}{1 - g^2} = \int k\,\mathrm{d}x.$$

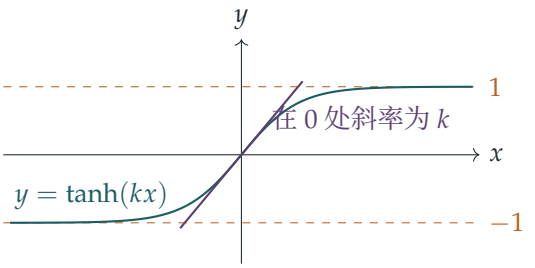
利用部分分式（partial fractions），$\dfrac{1}{1-\mathrm{g}^2} = \dfrac{1}{2}\left(\dfrac{1}{1+\mathrm{g}} + \dfrac{1}{1-\mathrm{g}}\right)$，而 $1 \pm \mathrm{g}$ 两者皆为正数，所以

$$\frac{1}{2}\ln\frac{1+\mathrm{g}(x)}{1-\mathrm{g}(x)} = kx + c.$$

在 $x = 0$ 时左边等于 $\frac{1}{2}\ln 1 = 0$，所以 $c = 0$。因此 $\dfrac{1+\mathrm{g}}{1-\mathrm{g}} = \mathrm{e}^{2kx}$，解出 g，

$$\mathrm{g}(x) = \frac{\mathrm{e}^{2kx}-1}{\mathrm{e}^{2kx}+1} = \tanh(kx).$$

这确实可行：$|\tanh(kx)| < 1$；加法公式（addition formula）$\tanh(a+b) = \dfrac{\tanh a + \tanh b}{1 + \tanh a \tanh b}$ 正正就是这条函数方程；而 $\tanh(kx)$ 在 0 的导数是 k。



验算

- 在 (i) 中，取 $\mathrm{f}(x) = \mathrm{e}^{kx}$，$k = 1$，$x = 0.3$，$y = 0.5$：$\mathrm{e}^{0.3}\mathrm{e}^{0.5} = 1.3499 \times 1.6487 = 2.2255 = \mathrm{e}^{0.8}$。
- 在 (ii) 中，取 $k = 1$，$x = 0.3$，$y = 0.5$：$\tanh 0.3 = 0.29131$，$\tanh 0.5 = 0.46212$，而 $(0.29131 + 0.46212)/(1 + 0.29131 \times 0.46212) = 0.66404 = \tanh 0.8$。
- 两个答案亦已用电脑对大量随机的 k、x、y 作出验证，包括导数关系 $\mathrm{f}' = k\mathrm{f}$ 及 $\mathrm{g}' = k(1-\mathrm{g}^2)$。
- 两部分是互相关连的：$\tanh(kx) = \dfrac{\mathrm{e}^{2kx}-1}{\mathrm{e}^{2kx}+1}$ 是由第 (i) 部分的指数函数构成，只是以 $2k$ 代替 k。

常见错误

- 没有给出理由便舍弃 $\mathrm{f}(0) = 0$ 或 $\mathrm{g}(0) = \pm 1$。条件 $k \neq 0$ 及 $|\mathrm{g}| < 1$ 正是为了排除这些情况而设的。
- 在分母中令 $\mathrm{g}(h) \to 0$ 却没有说明原因：这需要 g 在 0 连续，而连续性来自可微性。

► 除以 f 或 $1 - g^2$ 之前没有检查它是否非零。

STEP 之外

这些都是柯西型函数方程（*Cauchy-type functional equations*）的例子。大学第一个分析（analysis）课程会证明：在远比可微性弱的假设下，例如只在某一点连续，指数方程 $f(x+y) = f(x)f(y)$ 便迫使 $f(x) = e^{kx}$（或 $f \equiv 0$）；而如果没有任何正则性（regularity）条件，则会出现极其怪异的解。tanh 方程正是狭义相对论（special relativity）中的速度相加定律，在那里速度的 $\tanh^{-1}$（称为快度（rapidity））只需直接相加。想了解背后的分析，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2019，第 3 题

纯数学 · 矩阵的不变直线

> **STEP III 2019，第 3 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

点 (x,y) 是不变点（invariant point），当且仅当 $(\mathbf{A}-\mathbf{I})$ 把它映到零，由此得到两条线性方程（linear equations）。第 (i) 和第 (ii) 部分的所有内容，都源于追问：这两条方程何时描述同一条直线？其条件是 $(a-1)(d-1)=bc$，亦即 $\det(\mathbf{A}-\mathbf{I})=0$。在第 (iii) 部分，我们把直线上的一般点代入变换，然后比较系数。

解答

第 (i) 部分

点 (x,y) 是不变点，当 $ax+by=x$ 且 $cx+dy=y$，即

$$(a-1)x=-by, \tag{1}$$

$$(d-1)y=-cx. \tag{2}$$

把 (1) 和 (2) 相乘，得 $(a-1)(d-1)xy=bc\,xy$，所以对 L_1 上每一点 (x,y) 都有

$$\big((a-1)(d-1)-bc\big)xy=0$$

，此即所求。

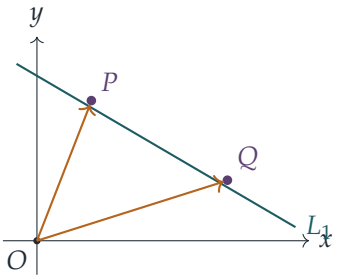
至于第二个结果，我们需要 L_1 上一个满足 $xy\neq 0$ 的点，这样才可以除以 xy。

> ? **想一想** 如果 L_1 上每一点都有 $x=0$ 或 $y=0$，那又如何？

集合 $xy=0$ 正是两条坐标轴。一条位于两轴之并集内的直线必定是其中一条坐标轴，所以除非 L_1 是坐标轴，否则它必含一个 $xy\neq 0$ 的点。于是除以 xy，便得 $(a-1)(d-1)=bc$。

剩下的是坐标轴的情况。若 L_1 是 y 轴，则每个 $(0,y)$ 都是不变点，而 (1) 和 (2) 给出对所有 y 皆有 $by=0$ 及 $(d-1)y=0$。所以 $b=0$ 且 $d=1$，而 $(a-1)(d-1)=bc$ 两边都是 0。若 L_1 是 x 轴，同理得 $c=0$ 且 $a=1$，两边同样是 0。因此在所有情况下都有 $(a-1)(d-1)=bc$。

现在假设 L_1 不经过原点。在 L_1 上取两个不同的点 P 和 Q，其位置向量 (position vectors) 分别为 $\mathbf{p}$ 和 $\mathbf{q}$。



由于 O、P、Q 不共线，$\mathbf{p}$ 与 $\mathbf{q}$ 不平行，所以每个向量都可写成 $\mathbf{v}=\lambda\mathbf{p}+\mu\mathbf{q}$。由线性性 (linearity)，

$$\mathbf{Av}=\lambda\mathbf{Ap}+\mu\mathbf{Aq}=\lambda\mathbf{p}+\mu\mathbf{q}=\mathbf{v},$$

所以平面上每一点都是不变点，且 $\mathbf{A}=\mathbf{I}$。(此时 $a=d=1$ 及 $b=c=0$，与 $(a-1)(d-1)=bc$ 吻合。)

第 (ii) 部分

假设 $(a-1)(d-1)=bc$。我们需要找一条直线，使 (1) 和 (2) 在其上都成立。

情况 $b\neq 0$。取直线 $(a-1)x+by=0$；由于 $b\neq 0$，这确实是一条通过原点的直线。直线上每一点都满足 (1)。在这条直线上 $y=-\frac{(a-1)x}{b}$，所以

$$cx+(d-1)y=cx-\frac{(a-1)(d-1)}{b}\,x=cx-\frac{bc}{b}\,x=0,$$

因此 (2) 亦成立。所以 $(a-1)x+by=0$ 是一条不变点直线 (line of invariant points)。

情况 $b=0$。此时条件变为 $(a-1)(d-1)=0$，所以 $a=1$ 或 $d=1$。

- 若 $a\neq 1$，则 $d=1$。方程 (1) 变成 $(a-1)x=0$，所以 $x=0$；而 (2) 变成 $cx=0$，在 $x=0$ 时成立。不变点直线就是 y 轴 $x=0$。
- 若 $a=1$，则 (1) 对每一点都成立，我们只需 (2)：$cx+(d-1)y=0$。
  - 若 c 与 $d-1$ 不同时为零，这是一条通过原点的直线，而且是一条不变点直线。
  - 若 $c=0$ 且 $d=1$，则 $\mathbf{A}=\mathbf{I}$，每一点都是不变点，每一条直线都是不变点直线。

所以只要 $(a-1)(d-1)=bc$，就存在一条不变点直线。

第 (iii) 部分

由于 L_2 不经过原点，可把 L_2 写成 $y = mx + k$，其中 $k \neq 0$。L_2 上的一般点是 $(x,\, mx + k)$，其像为

$$(X, Y) = \big((a + bm)x + bk,\ (c + dm)x + dk\big).$$

这条直线是不变直线（invariant line），当且仅当对每个实数 x 都有 $Y = mX + k$：

$$(c + dm)x + dk = m(a + bm)x + mbk + k.$$

两边都是 x 的一次式，且对所有 x 都相等，所以系数必须相同：

$$\text{常数项：}\quad dk = mbk + k, \tag{3}$$

$$\text{一次项系数，关于}x\text{：}\quad c + dm = am + bm^2. \tag{4}$$

想一想 我们在哪里用到 L_2 不经过原点这个事实？

我们在 (3) 中用到它：因为 $k \neq 0$，可以除以 k，得 $d - 1 = mb$。把 $d = 1 + mb$ 代入 (4)，

$$c = am + bm^2 - dm = am + bm^2 - m - bm^2 = m(a - 1).$$

因此

$$bc = bm(a - 1) = (d - 1)(a - 1),$$

此即所求结果。（若 $k = 0$，方程 (3) 不提供任何信息，这正是通过原点的直线可以在不满足该条件下仍是不变直线的原因。旋转 π 就是一例：它把每条通过 O 的直线映到自身，但 $(a-1)(d-1) = 4 \neq 0 = bc$。）

验算

- 试取 $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$：此时 $(a-1)(d-1) = -2 = bc$。第 (ii) 部分预测不变直线为 $2x + 2y = 0$，即 $y = -x$，而确实有 $\mathbf{A}(1,-1)^{\mathrm{T}} = (1,-1)^{\mathrm{T}}$。
- 对于第 (iii) 部分，取 $m = 2$、$k = 1$、$a = 3$、$b = 1$。则 $d = 1 + mb = 3$ 及 $c = m(a-1) = 4$。$y = 2x + 1$ 上的点 $(1, 3)$ 映到 $(6, 13)$，而 $13 = 2 \times 6 + 1$。此外 $(a-1)(d-1) = 4 = bc$。
- 我们用电脑验证了数以千计的随机矩阵：所找到的每条满足 $k \neq 0$ 的不变直线 $y = mx + k$ 都有 $d - 1 = mb$、$c = m(a-1)$ 及 $(a-1)(d-1) = bc$；而每条不经过原点的不动点直线都来自 $\mathbf{A} = \mathbf{I}$。

常见错误

- 在第 (i) 部分除以 xy，却没有处理 L_1 是坐标轴的情况。
- 在第 (ii) 部分 $b = 0$ 时漏掉某个子情况，特别是 $\mathbf{A} = \mathbf{I}$，此时每条直线都符合。
- 在第 (iii) 部分混淆了不变直线（点仍留在直线上）与不变点直线（点本身不动），因而错误地设 $X = x$。

STEP 之外

条件 $(a-1)(d-1) = bc$ 就是 $\det(\mathbf{A} - \mathbf{I}) = 0$，表示 1 是 $\mathbf{A}$ 的特征值（eigenvalue）。通过原点的不变点直线，就是该特征值的特征空间（eigenspace）。第 (iii) 部分说明，一条不经过原点的不变直线同样迫使 1 成为特征值。关系式 $c = m(a-1)$ 和 $d - 1 = mb$ 表示 $\mathbf{A} - \mathbf{I}$ 的第二行是第一行的 m 倍，所以 $\mathbf{A} - \mathbf{I}$ 的秩（rank）最多是 1，并把每一点沿平行于 L_2 的方向移动。关于特征值、特征向量（eigenvectors）及线性映射的几何，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

STEP III 2019，第 4 题

纯数学 · 根与系数

STEP III 2019，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

交替出现的正负号正是韦达定理（Vieta's formulas）中的符号：以 $r_1,\dots,r_n$ 为根的首一多项式（monic polynomial）是 $x^n - e_1x^{n-1} + e_2x^{n-2} - \cdots + (-1)^n e_n$，其中 e_k 是从各根中任取 k 个相乘所得的所有乘积之和。因此，“自反”（reflexive）其实只是说对每个 k 都有 $e_k(a_1,\dots,a_n) = a_k$。这样，较小的情况就只是几条联立方程，而各根的平方和 $e_1^2 - 2e_2$ 则给出 (ii) 中的恒等式。一旦证得 $a_n = 0$，就可以抽出因式 x，把多项式逐层剥减。

解答

以下一律以 e_k 表示 $a_1,\dots,a_n$ 的第 k 个初等对称函数（elementary symmetric function）。根据韦达定理，P 的根是 $a_1,\dots,a_n$ 若且唯若

$$e_k(a_1,\dots,a_n) = a_k \qquad \text{对} k = 1,2,\dots,n.$$

第 (i) 部分

一次。 $\mathrm{P}(x) = x - a_1$ 的根是 a_1，所以所有这类多项式都是自反的：对任意实数 a，$\mathrm{P}(x) = x - a$。

二次。 我们需要 $a_1 + a_2 = a_1$ 及 $a_1a_2 = a_2$。由第一条得 $a_2 = 0$，此时第二条自动成立。所以对任意实数 a，$\mathrm{P}(x) = x^2 - ax = x(x-a)$；它的根 a 和 0 确实分别是 a_1 和 a_2。

三次。我们需要

$$a_1 + a_2 + a_3 = a_1,$$
$$a_1a_2 + a_1a_3 + a_2a_3 = a_2,$$
$$a_1a_2a_3 = a_3.$$

由第一条得 $a_3 = -a_2$。于是第二条变成 $a_1(a_2 + a_3) + a_2a_3 = a_2$，即 $-a_2^2 = a_2$，所以 $a_2 = 0$ 或 $a_2 = -1$。

- 若 $a_2 = 0$，则 $a_3 = 0$，而第三条方程变成 $0 = 0$。所以 a_1 可以任意取值：$\mathrm{P}(x) = x^3 - ax^2 = x^2(x - a)$。
- 若 $a_2 = -1$，则 $a_3 = 1$，而第三条方程给出 $-a_1 = 1$，所以 $a_1 = -1$。由此得

  $$\mathrm{P}(x) = x^3 + x^2 - x - 1 = (x + 1)^2(x - 1),$$

  它的根 $-1, -1, 1$ 确实是 a_1, a_2, a_3。

所以次数不超过 3 的自反多项式是

$$x - a, \qquad x^2 - ax, \qquad x^3 - ax^2 \quad (a \in \mathbb{R}), \qquad \text{以及} \qquad x^3 + x^2 - x - 1.$$

第 (ii) 部分

各根的平方和是

$$a_1^2 + a_2^2 + \cdots + a_n^2 = e_1^2 - 2e_2,$$

因为把 e_1 平方后，每个 a_i^2 出现一次，而每个乘积 a_ia_j（$i < j$）出现两次。对自反多项式而言，$e_1 = a_1$ 且 $e_2 = a_2$，所以

$$a_1^2 + a_2^2 + \cdots + a_n^2 = a_1^2 - 2a_2, \qquad \text{即} \qquad 2a_2 = -a_2^2 - a_3^2 - \cdots - a_n^2,$$

这就是所求的结果。（这对每个 $n \geq 2$ 都成立。）

现在假设各 a_k 都是整数，$a_n \neq 0$，并用反证法（proof by contradiction）假设 $n > 3$。把 a_2^2 移到另一边并配方（completing the square），得

$$(a_2 + 1)^2 = 1 - \left(a_3^2 + a_4^2 + \cdots + a_n^2\right).$$

想一想 整数条件如何把这条方程“夹紧”?

左边至少是 0，所以 $a_3^2 + \cdots + a_n^2 \leq 1$。但 a_n 是非零整数，所以 $a_n^2 \geq 1$。因此这个和恰好等于 1：$a_n^2 = 1$，而其余各项都是零，所以 $a_3 = a_4 = \cdots = a_{n-1} = 0$，同时 $a_2 = -1$。由于 $n \geq 4$，数列 $a_3, \ldots, a_{n-1}$ 并非空的，特别地 $a_3 = 0$。现在，最后一条韦达方程说各根之积等于 a_n：

$$a_n = e_n = a_1a_2a_3 \cdots a_n = 0,$$

因为 $a_3 = 0$。这与 $a_n \neq 0$ 矛盾。所以 $n \leq 3$。

第 (iii) 部分

抽出因式 x。假设 P 是次数 $n \geq 2$ 的自反多项式，且 $a_n = 0$。那么 $\mathrm{P}(x) = x\,\mathrm{Q}(x)$，其中

$$\mathrm{Q}(x) = x^{n-1} - a_1 x^{n-2} + \cdots + (-1)^{n-1} a_{n-1}.$$

P 的根是 $a_1, \ldots, a_{n-1}, a_n$，其中 $a_n = 0$，而因式 x 对应其中一个根 0。所以 Q 的根就是余下的 $a_1, \ldots, a_{n-1}$：Q 是自反的，而且系数相同（同样是整数）。反过来，若 Q 是次数 m 的自反多项式，系数为 $a_1, \ldots, a_m$，那么 $x\,\mathrm{Q}(x)$ 的系数是 $a_1, \ldots, a_m, 0$，根是 $a_1, \ldots, a_m, 0$，所以它也是自反的。

逐层剥减。从任何整系数自反多项式 P 出发。只要它的次数至少是 2 而最后一个系数是 0，就抽出一个因式 x。这个过程最终停在 $\mathrm{P}(x) = x^m \mathrm{R}(x)$，其中 $m \geq 0$，而 R 是次数为 k 的整系数自反多项式，且要么 $k = 1$，要么最后一个系数 a_k 非零。

- 若 $a_k \neq 0$，则由第 (ii) 部分得 $k \leq 3$（若 $k \leq 3$ 则无需证明）。根据第 (i) 部分的列表：$k = 1$ 时得 $x - a$，其中 $a \neq 0$；$k = 2$ 时最后一个系数必定是 0，所以没有；$k = 3$ 时唯一满足 $a_3 \neq 0$ 的是 $x^3 + x^2 - x - 1$。
- 若 $k = 1$ 且 $a_1 = 0$，则 $\mathrm{R}(x) = x$。

把 $\mathrm{R}(x) = x$ 并入 $x - a$ 中 $a = 0$ 的情况，再利用上述的逆命题得知每个乘以 x^m 的倍式都是自反的，便知整系数自反多项式恰好是

$$x^m(x - a) \quad (m \geq 0,\ a \in \mathbb{Z}),$$
$$x^m(x^3 + x^2 - x - 1) = x^m(x+1)^2(x-1) \quad (m \geq 0).$$

验算

- $x^2(x+1)^2(x-1) = x^5 + x^4 - x^3 - x^2$ 的 $a_1, \ldots, a_5 = -1, -1, 1, 0, 0$，根是 $-1, -1, 1, 0, 0$：它是自反的。(ii) 中的恒等式变成 $2(-1) = -(1 + 1 + 0 + 0)$，成立。
- 用电脑搜索所有 $n \leq 7$、$a_1, \ldots, a_n$ 取 $\{-3, \ldots, 3\}$ 中整数的情况，除上列者外没有找到其他自反多项式；对 $n = 2, 3$ 在较小的有理数中搜索，亦印证了第 (i) 部分的列表。

常见错误

- 遗漏含自由参数的多项式族：$x^2 - ax$ 和 $x^3 - ax^2$ 对每个实数 a 都是自反的，而不只是对某些特殊值。

- 在 (ii) 中，求得 $a_n = \pm 1$、$a_2 = -1$ 便停下来。矛盾其实来自各根之积：当 $n \geq 4$ 时，$a_3, \ldots, a_{n-1}$ 中有零，所以积为零。
- 在 (iii) 中，只说“$a_n = 0$，所以除以 x”，却没有验证商式仍然是自反的：一个零根对应因式 x，而余下的根要与余下的系数吻合。

STEP 之外

方程组 $e_k(a_1, \ldots, a_n) = a_k$ 是多项式方程组的一个小例子，其解构成一个代数簇（algebraic variety），在这里是若干直线与孤立点的并集。“平方和 $= e_1^2 - 2e_2$”这个事实是牛顿（Newton）恒等式（Newton's identities）中的第一条；这些恒等式把各根的幂和与系数联系起来，是对称函数理论的基础。“若干整数的平方之和不超过 1”这一步，是丢番图（Diophantine）论证中的典型手法。Davenport 的 *The Higher Arithmetic*（《高等算术》）和 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）都是进一步探索这些概念的好起点。

详细解答

STEP III 2019，第 5 题

纯数学 · 代换积分法

STEP III 2019，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

第 (i) 部分中的函数把 $[0,\infty)$ 光滑且递增地映射到 $[0,1)$ 上。因此只要把它乘以一个常数，就能把一个 y 积分（积分区间是 $[0,c)$ 的一部分）转化为一个 x 积分（积分区间是 $[0,\infty)$ 的一部分）。这样做亦会令 c^2-y^2 变成一个经过伪装的完全平方，从而消去那个麻烦的平方根。根号消失后，剩下的是一个有理（rational）被积函数（integrand），可以用题目注释中给出的反正切（arctangent）积分来处理。倒数代换（reciprocal substitution）$y=1/u$ 则把后面的积分联系回第一个积分。

解答

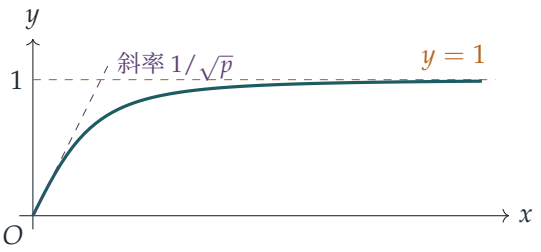
第 (i) 部分

设 $f(x)=x(x^2+p)^{-1/2}$ 且 $p>0$，则 $f(0)=0$，且

$$f'(x)=(x^2+p)^{-1/2}-x^2(x^2+p)^{-3/2}=\frac{p}{(x^2+p)^{3/2}}>0,$$

$$f''(x)=-\frac{3px}{(x^2+p)^{5/2}}<0 \quad \text{对于} x>0.$$

因此曲线以斜率（gradient）$1/\sqrt{p}$ 离开原点，并稳定上升，同时向下弯曲（即曲线是凹（concave）的）。由于当 $x>0$ 时 $f(x)=1/\sqrt{1+p/x^2}$，故当 $x\to\infty$ 时 $f(x)\to 1$，且由下方趋近，所以 $y=1$ 是曲线永远达不到的渐近线（asymptote）。



第 (ii) 部分

按试卷印刷的版本，代换式的分子是 b。这样行不通：它虽然令 b^2-y^2 变得整齐，却留下 $c^2(x^2+p)-b^2x^2$ 的平方根，而这并不是完全平方。真正可行的代换（亦即出题者的原意）分子是 c：

$$y=\frac{cx}{\sqrt{x^2+p}},\qquad p>0.$$

由第 (i) 部分，这是由 $x\in[0,\infty)$ 映射到 $y\in[0,c)$ 上的递增映射，而这正是 $\sqrt{c^2-y^2}$ 有定义的范围。于是

$$c^2-y^2=\frac{c^2(x^2+p)-c^2x^2}{x^2+p}=\frac{c^2p}{x^2+p},$$

$$b^2-y^2=\frac{b^2p+(b^2-c^2)x^2}{x^2+p},$$

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{cp}{(x^2+p)^{3/2}}\quad\text{（由第 (i) 部分）}.$$

由于 $c>0$ 及 $p>0$，有 $\sqrt{c^2-y^2}=c\sqrt{p}/\sqrt{x^2+p}$。代入得

$$I=\int\frac{x^2+p}{b^2p+(b^2-c^2)x^2}\cdot\frac{\sqrt{x^2+p}}{c\sqrt{p}}\cdot\frac{cp}{(x^2+p)^{3/2}}\,\mathrm{d}x=\int\frac{\sqrt{p}}{b^2p+(b^2-c^2)x^2}\,\mathrm{d}x.$$

取 $p=1$，得

$$I=\int\frac{1}{b^2+(b^2-c^2)x^2}\,\mathrm{d}x,$$

这正是所求的结果。所用的代换是 $y=cx/\sqrt{x^2+1}$，其反函数是 $x=y/\sqrt{c^2-y^2}$。

第一个积分。取 $b=\sqrt{3}$ 及 $c=\sqrt{2}$。于是 $x=y/\sqrt{2-y^2}$，因此 $y=1$ 对应 $x=1$，而当 $y\to\sqrt{2}$ 时 $x\to\infty$。原被积函数在 $y=\sqrt{2}$ 处趋于无穷，其增长方式如 $(2-y^2)^{-1/2}$，属于可积（integrable）的类型。代换后，积分变成 $[1,\infty)$ 上一个收敛（convergent）的反常积分（improper integral）：

$$\int_1^{\sqrt{2}}\frac{\mathrm{d}y}{(3-y^2)\sqrt{2-y^2}}=\int_1^{\infty}\frac{\mathrm{d}x}{3+x^2}=\frac{1}{\sqrt{3}}\Big[\tan^{-1}\frac{x}{\sqrt{3}}\Big]_1^{\infty}$$

$$=\frac{1}{\sqrt{3}}\Big(\frac{\pi}{2}-\frac{\pi}{6}\Big)=\frac{\pi}{3\sqrt{3}}.$$

想一想 新的积分中出现的是 $3y^2-1$ 和 $2y^2-1$，而旧的积分中则是 $3-y^2$ 和 $2-y^2$。哪一种代换能把 y^2 与常数的角色互换?

第二个积分。 设 $y=1/u$，其中 $u>0$，于是 $\mathrm{d}y=-\mathrm{d}u/u^2$。那么

$$3y^2-1=\frac{3-u^2}{u^2},\qquad \sqrt{2y^2-1}=\frac{\sqrt{2-u^2}}{u},$$

而积分上下限 $y=\frac{1}{\sqrt2}$ 和 $y=1$ 分别变成 $u=\sqrt2$ 和 $u=1$。所以

$$\int_{1/\sqrt2}^{1}\frac{y\,\mathrm{d}y}{(3y^2-1)\sqrt{2y^2-1}}=\int_{\sqrt2}^{1}\frac{1}{u}\cdot\frac{u^2}{3-u^2}\cdot\frac{u}{\sqrt{2-u^2}}\cdot\left(-\frac{\mathrm{d}u}{u^2}\right)=\int_1^{\sqrt2}\frac{\mathrm{d}u}{(3-u^2)\sqrt{2-u^2}}.$$

这正好就是第一个积分，因此其值同样是 $\dfrac{\pi}{3\sqrt3}$。

第 (iii) 部分

同样的倒数代换 $y=1/u$ 会把被积函数变成 $u/\big((3-u^2)\sqrt{2-u^2}\big)$。多出来的因子 u 提示我们设 $w=\sqrt{2-u^2}$，因为这样便有 $3-u^2=1+w^2$。把两步合并，就得到单一个代换:

$$w=\frac{\sqrt{2y^2-1}}{y}=\sqrt{2-\frac{1}{y^2}},\qquad \text{故}\qquad w^2=2-\frac{1}{y^2},\quad 2w\,\mathrm{d}w=\frac{2}{y^3}\,\mathrm{d}y.$$

当 $y\in[\frac{1}{\sqrt2},1]$ 时，w 由 0 递增至 1。此外，$\mathrm{d}y=wy^3\,\mathrm{d}w$、$\sqrt{2y^2-1}=wy$，以及 $3y^2-1=y^2\big(3-\frac{1}{y^2}\big)=y^2(1+w^2)$。因此

$$\int_{1/\sqrt2}^{1}\frac{\mathrm{d}y}{(3y^2-1)\sqrt{2y^2-1}}=\int_0^1\frac{wy^3\,\mathrm{d}w}{y^2(1+w^2)\cdot wy}=\int_0^1\frac{\mathrm{d}w}{1+w^2}=\frac{\pi}{4}.$$

（$y=1/\sqrt2$ 处的奇点（singularity）同样只属于 $(2y^2-1)^{-1/2}$ 这种可积的类型，而代换把它消除了。）

这其实是经过伪装的第 (ii) 部分方法。设 $x=1/w$，便得 $y=x/\sqrt{2(x^2-1)}$，这正是取 $c^2=\frac12$ 及 $p=-1$ 时的一般代换。这就是第 (i) 部分容许 p 为负数的原因。

验算

- $\pi/(3\sqrt3)\approx0.6046$，而 $\pi/4\approx0.7854$。先用三角代换消除端点的奇点，再对原积分进行数值积分（numerical integration），两个数值均得到证实。
- 第一个积分的粗略大小检查：在 $[1,\sqrt2]$ 上有 $1\le3-y^2\le2$，所以积分值介乎 $\frac12$ 倍与 1 倍的 $\int_1^{\sqrt2}(2-y^2)^{-1/2}\,\mathrm{d}y=\frac{\pi}{4}$ 之间。由此得出的范围是 0.39 至 0.79，而 0.6046 稳稳落在其中。

- 在某一点检验第 (ii) 部分的恒等式：取 $b=\sqrt{3}$、$c=\sqrt{2}$ 及 $x=1$，得 $y=1$。此时被积函数的值是 $\frac{1}{2}$，$\mathrm{d}y/\mathrm{d}x=\sqrt{2}/2^{3/2}=\frac{1}{2}$，两者之积是 $\frac{1}{4}=1/(3+1)$，与预期相符。

常见错误

- 坚持使用代换式中印出来的 b，结果卡住：需要消去的平方根是 $\sqrt{c^2-y^2}$，所以缩放因子必须是 c。
- 转换积分上下限时出错。这里 $y=\sqrt{2}=c$ 对应的是 $x=\infty$，而不是某个有限的 x 值。
- 漏掉 $\mathrm{d}y=-\mathrm{d}u/u^2$ 中的负号，或忘记把上下限调换回来，以致一个正的被积函数得出负的答案。

STEP 之外

这里用到的代换属于有理化代换（rationalising substitution）。它们把含有二次式平方根的被积函数变成有理函数，之后再用部分分式（partial fractions）或 $\tan^{-1}$ 便可完成。在大学一年级的分析（analysis）课程中，你会学到这里所用反常积分背后的理论：含无穷上下限或无界被积函数的积分何时收敛，以及为何代换法对这类积分仍然成立。Bartle 与 Sherbert 合著的 *Introduction to Real Analysis*（《实分析导论》）对这两方面都有仔细的论述。

详细解答

STEP III 2019，第 6 题

纯数学 · 阿根图上的圆

> **STEP III 2019，第 6 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

开首那条式子其实就是乔装了的 $|z-a|^2-r^2$，所以轨迹（locus）是与 a 距离为 r 的点的集合。至于在 $\omega=1/z$ 之下的像，先代入 $z=1/\omega$，消去分母，再把结果整理回与原方程相同的形式。一旦化成这个形式，便可以比较系数（matching coefficients），直接读出圆心和半径。最后比较圆心和半径，就能判断 C' 与 C 何时是同一个圆。

解答

由于 $(z-a)^*=z^*-a^*$，

$$|z-a|^2=(z-a)(z^*-a^*)=zz^*-a^*z-az^*+aa^*,$$

所以题目给出的方程正好说 $|z-a|^2=r^2$，即 $|z-a|=r$（因为 $r>0$）。因此 C 是**圆心为 a、半径为 r 的圆**。

更一般地，若 $s>0$，方程 $ww^*-bw^*-b^*w+bb^*-s^2=0$ 表示圆心为 b、半径为 s 的圆。我们会把这个当作模板使用。

第 (i) 部分

首先，$z=0$ 不在 C 上：否则需要 $|a|=r$，即 $a^*a=r^2$，而这已被排除。所以对 C 上每一点，$\omega=1/z$ 都有定义，而且 $\omega\neq0$。另外 $z=1/\omega$，所以 Q 在 C' 上当且仅当 $\omega\neq0$ 且

$$\left|\frac{1}{\omega}-a\right|=r \iff |1-a\omega|=r|\omega| \iff (1-a\omega)(1-a^*\omega^*)=r^2\omega\omega^*.$$

展开并合并同类项，

$$(a^*a-r^2)\,\omega\omega^*-a\omega-a^*\omega^*+1=0.$$

设 $k = a^*a - r^2 = |a|^2 - r^2$，由题设它是一个非零实数。两边除以 k，

$$\omega\omega^* - \frac{a^*}{k}\omega^* - \frac{a}{k}\omega + \frac{1}{k} = 0.$$

这与模板吻合，其中 $b = a^*/k$（由于 k 是实数，所以 $b^* = a/k$），前提是 $bb^* - s^2 = 1/k$，即

$$s^2 = \frac{|a|^2}{k^2} - \frac{1}{k} = \frac{|a|^2 - k}{k^2} = \frac{r^2}{k^2}.$$

这是正数，所以 C' 确实是一个圆，其

$$\text{圆心为}\frac{a^*}{|a|^2 - r^2}, \qquad \text{半径为}\frac{r}{\big||a|^2 - r^2\big|}.$$

（注意 $\omega = 0$ 不满足这方程，因为它会给出 $1 = 0$。所以这个圆上每一点确实都是某个在 C 上的 z 的 $1/z$。）

? 想一想 两个圆何时相同?

两个圆相同，当且仅当它们有相同的圆心和相同的半径。所以 $C' = C$ 要求

$$\frac{r}{|k|} = r \qquad \text{及} \qquad \frac{a^*}{k} = a.$$

第一个条件给出 $|k| = 1$，即 $k^2 = 1$：

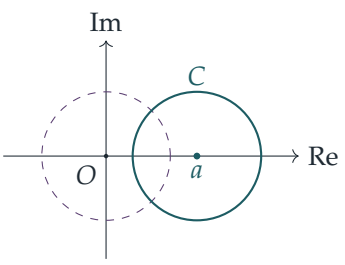
$$\big(|a|^2 - r^2\big)^2 = 1.$$

所以 $k = 1$ 或 $k = -1$，再由第二个条件得：

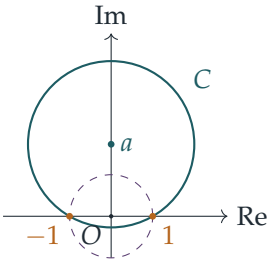
- 若 $k = 1$：$a^* = a$，所以 a 是**实数**；
- 若 $k = -1$：$a^* = -a$，所以 a 是**虚数**（纯虚数（purely imaginary））。

（若 $a = 0$，则两者皆是，此时 $r = 1$。）反过来，上述任一情况，加上半径条件，的确会令圆心和半径一致。

草图。若 a 是实数，则 $|a|^2 = r^2 + 1 > r^2$，所以圆心在实轴上，与 O 的距离为 $\sqrt{r^2+1}$，而原点位于 C 的外面。这个圆与实轴相交于 $a - r$ 和 $a + r$，两者之积为 $a^2 - r^2 = 1$，这是必然的，因为 $1/z$ 会把它们互换。若 a 是虚数，则 $|a|^2 = r^2 - 1$，所以这情况需要 $r \geq 1$。圆心在虚轴上，而原点位于 C 的里面。设 $a = \mathrm{i}c$，与实轴的交点满足 $x^2 + c^2 = r^2$，所以 $x = \pm 1$。在两种情况下，C 都与单位圆（虚线）成直角相交（正交（orthogonal）），因为 $|a|^2 = r^2 \pm 1$（左图中 $a = \sqrt{2}$，$r = 1$；右图中 $a = \mathrm{i}\sqrt{3}$，$r = 2$）。



a 为实数



a 为虚数

若 a 是实数，它当然也可以是负数，得出的是左图关于虚轴的镜像；同样，虚数 a 也可以位于实轴下方。

第 (ii) 部分

现在 $z = 1/\omega^*$，所以 Q 在轨迹上当且仅当

$$\left|\frac{1}{\omega^*} - a\right| = r \iff |1 - a\omega^*| = r|\omega^*| \iff |1 - a^*\omega| = r|\omega|,$$

这里是在第一个模（modulus）里面取共轭（conjugate）（一个数与它的共轭有相同的模）。这正是第 (i) 部分的方程，只是把 a 换成 a^*。所以由第 (i) 部分，轨迹是一个圆，其

$$\text{圆心为}\frac{a}{k}, \qquad \text{半径为}\frac{r}{|k|}, \qquad k = |a|^2 - r^2.$$

要这个圆就是 C，需要 $|k| = 1$ 及 $a/k = a$。若 $a \neq 0$，第二个条件迫使 $k = 1$，而条件就只有这些：条件正好是

$$|a|^2 = r^2 + 1,$$

它对 a 的辐角（argument）没有任何限制。（若 $a = 0$，则与之前一样需要 $r = 1$。）

? **想一想** 你能找出一个既非实数、亦非虚数的具体 a 吗?

取 $a = 1 + \mathrm{i}$ 及 $r = 1$。那么 $|a|^2 = 2 = r^2 + 1$，所以 Q 的轨迹就是 C，但 a 既不是实数，也不是纯虚数。所以答案是**否**：a 并不一定是实数或虚数。

验算

- 取 $a = \sqrt{5}, r = 2$（于是 $|a|^2 - r^2 = 1$），C 上的点 $z = \sqrt{5} \pm 2$ 映到 $1/(\sqrt{5} \pm 2) = \sqrt{5} \mp 2$，它们仍然在 C 上。
- 在第 (ii) 部分取 $a = 1 + \mathrm{i}$，$r = 1$：点 $z = 1$ 在 C 上，而 $1/z^* = 1$ 也在。点 $z = 2 + \mathrm{i}$ 给出 $1/z^* = (2 + \mathrm{i})/5 = 0.4 + 0.2\mathrm{i}$，而 $|(0.4 + 0.2\mathrm{i}) - (1 + \mathrm{i})|^2 = 0.36 + 0.64 = 1$。但 $1/z = 0.4 - 0.2\mathrm{i}$ 不在 C 上，所以这个圆在第 (i) 部分的映射下并不固定。

- 两部分中的像圆亦已用电脑验证：把随机圆上的大量点映射过去，再量度它们与预测圆心的距离。

常见错误

- 把 C' 的圆心写成 $a/(|a|^2 - r^2)$，而不是 $a^*/(|a|^2 - r^2)$。读出圆心需要看 ω^* 的系数，而不是 ω 的系数。
- 忘记半径 $r/\left||a|^2 - r^2\right|$ 中的绝对值。没有它的话，$|a| < r$ 的情况会得出负的半径。
- 在第 (ii) 部分照搬第 (i) 部分而答“是”，没有留意到共轭把条件 $a^* = \pm a$ 消除了。

STEP 之外

映射 $z \mapsto 1/z^*$ 是关于单位圆的反演（*inversion in the unit circle*），而 $z \mapsto 1/z$ 则是先作这个反演，再关于实轴作反射。两者都属于莫比乌斯变换（Möbius transformations）这个家族，这类变换总是把圆和直线映到圆和直线。第 (ii) 部分就是一个经典事实：一个圆在反演下固定，当且仅当它与反演圆成直角相交，即 $|a|^2 = r^2 + 1$（或者它就是单位圆本身）。Needham 的 *Visual Complex Analysis*（《直观复分析》）以大量图像，优美地发展了这些概念。

详细解答

STEP III 2019，第 7 题

纯数学 · 曲线描绘

STEP III 2019，第 7 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

方程只含 x^2 和 y^2，所以它是 x^2（或 y^2）的二次方程（quadratic）。这一下子给了我们两件工具：一是判别式（discriminant），它告诉我们曲线能取到哪些 y 值；二是以 y 表示 x^2 的明确公式。要看曲线在原点附近的形状，只保留最低次的项；要看远处的形状，则只保留最高次的项。在 $a = b$ 这个特殊情况下，方程可以完全因式分解（factorise）。

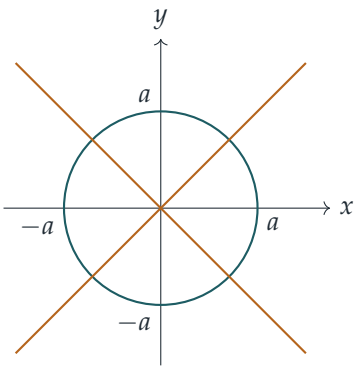
解答

第 (i) 部分

设 $a = b$，把所有项移到同一边，并把四次（quartic）项和二次项分组：

$$\begin{aligned} 0 &= (y^4 - x^4) - a^2(y^2 - x^2) \\ &= (y^2 - x^2)(y^2 + x^2) - a^2(y^2 - x^2) \\ &= (y^2 - x^2)(x^2 + y^2 - a^2). \end{aligned}$$

因此一点在曲线上，当且仅当 $y = \pm x$ 或 $x^2 + y^2 = a^2$：魔鬼曲线（Devil's Curve）就是两条直线 $y = x$ 和 $y = -x$，加上以原点为圆心、半径为 a 的圆。



第 (ii)(a) 部分

现在 $a=2$、$b=\sqrt{5}$，曲线为 $y^4-5y^2=x^4-4x^2$。固定 $y\geq 0$，并记 $X=x^2$。一点 (x,y)（其中 $x\geq 0$）在曲线上，当且仅当 $X\geq 0$ 满足

$$X^2-4X-(y^4-5y^2)=0.$$

要有实根，判别式必须非负：

$$\begin{aligned}16+4(y^4-5y^2)&=4(y^4-5y^2+4)=4(y^2-1)(y^2-4)\\&=4(y+1)(y+2)(y-1)(y-2)\geq 0.\end{aligned}$$

由于 $y\geq 0$，因子 $y+1$ 和 $y+2$ 都是正数，所以需要 $(y-1)(y-2)\geq 0$，即

$$0\leq y\leq 1 \quad 或 \quad y\geq 2.$$

反过来，每个这样的 y 都确实会出现。两个根是

$$X=2\pm\sqrt{D}, \qquad D=(y^2-1)(y^2-4),$$

其中较大的根至少是 2，所以它是合法的 x^2。下面会用到这条 x^2 的公式。

第 (ii)(b) 部分

在原点附近，四次项 y^4 和 x^4 与 y^2 和 x^2 相比微不足道。所以曲线近似于

$$-5y^2\approx -4x^2, \qquad 即 \qquad y\approx \tfrac{2}{\sqrt{5}}\,x \quad (x,y\geq 0).$$

曲线沿着斜率(gradient)为 $2/\sqrt{5}\approx 0.89$ 的直线离开原点，略低于直线 $y=x$。((a) 的公式也吻合：对于根 $X=2-\sqrt{D}$，用二项展开(binomial expansion)可得 $x^2=\frac{5}{4}y^2+O(y^4)$。)

在离原点很远处，最高次的项占主导：$y^4\approx x^4$，所以 $y\approx x$。我们还可以做得更好。把方程写成

$$(y^2-x^2)(y^2+x^2)=5y^2-4x^2 \quad\Longrightarrow\quad y^2-x^2=\frac{5y^2-4x^2}{y^2+x^2}.$$

由于 $y/x \to 1$，右边趋向 $\frac{5-4}{2} = \frac{1}{2}$。所以 $y^2 \approx x^2 + \frac{1}{2}$，从而

$$y \approx x + \frac{1}{4x}.$$

曲线以直线 $y = x$ 为渐近线（asymptote），并从上方趋近它。事实上，除原点外曲线从不与这条直线相交：把 $y = x$ 代入方程，得 $x^2 = 0$。

第 (ii)(c) 部分

对 $y^4 - 5y^2 = x^4 - 4x^2$ 关于 x 进行隐函数微分（implicit differentiation），

$$(4y^3 - 10y)\frac{\mathrm{d}y}{\mathrm{d}x} = 4x^3 - 8x \qquad \Longrightarrow \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2x(x^2-2)}{y(2y^2-5)}.$$

切线（*tangent*）平行于 x 轴：需要分子为零而分母不为零。所以 $x = 0$ 或 $x = \sqrt{2}$。

- $x = 0$：此时 $y^4 = 5y^2$，所以 $y = 0$ 或 $y = \sqrt{5}$。在 $(0, \sqrt{5})$，分母为 $\sqrt{5}(10-5) \neq 0$，所以这一点符合。
- $x = \sqrt{2}$：此时 $y^4 - 5y^2 = 4 - 8$，所以 $y^4 - 5y^2 + 4 = 0$，得 $y = 1$ 或 $y = 2$。这两处的分母分别是 -3 和 6，都不为零。

> **想一想** 原点处分子也是零，为甚么我们不把它算进去?

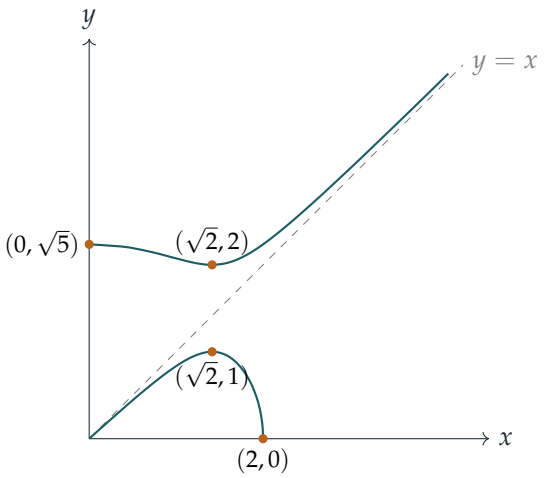
在 $(0,0)$，分母也是零，所以公式给出 $0/0$，甚么也说明不了。(b) 部分已经解决了这个问题：曲线以斜率 $2/\sqrt{5}$（而不是 0）穿过原点。

所以切线在 $(0, \sqrt{5})$、$(\sqrt{2}, 1)$ 和 $(\sqrt{2}, 2)$ 处是水平的。

切线平行于 y 轴：需要分母为零而分子不为零，所以 $y = 0$ 或 $y^2 = \frac{5}{2}$。但 $\sqrt{5/2} \approx 1.58$ 严格介乎 1 与 2 之间，已被 (a) 部分排除。当 $y = 0$ 时得 $x^4 = 4x^2$，所以 $x = 0$（又是原点，如上所述须排除）或 $x = 2$。在 $(2,0)$，分子为 $2 \cdot 2 \cdot 2 = 8 \neq 0$。所以唯一这样的点是 $(2,0)$。

$x \geq 0$、$y \geq 0$ 部分的草图 利用 $x^2 = 2 \pm \sqrt{D}$，曲线分为两段。

- $0 \leq y \leq 1$ **的一个环**。根 $2 - \sqrt{D}$ 由 0（在 $y = 0$）增至 2（在 $y = 1$），而 $2 + \sqrt{D}$ 则由 4 减至 2。所以曲线以斜率 $2/\sqrt{5}$ 离开原点，升至 $(\sqrt{2}, 1)$ 处的极大值，再下降，在 $(2,0)$ 处垂直地与 x 轴相交。
- $y \geq 2$ **的一翼**。当 $2 \leq y \leq \sqrt{5}$ 时，根 $2 - \sqrt{D}$ 由 2 减至 0；超过 $y = \sqrt{5}$ 后它变成负数，不再是合法的 x^2。根 $2 + \sqrt{D}$ 则由 2 开始无界地增大。所以这一支由 $(0, \sqrt{5})$ 开始，切线水平，下降至 $(\sqrt{2}, 2)$ 处的极小值，然后上升，从上方趋近 $y = x$。

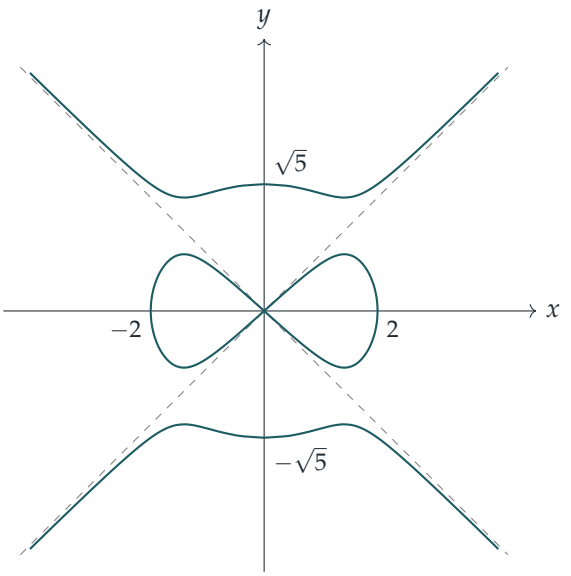


第 (iii) 部分

方程只涉及 x^2 和 y^2，所以曲线关于 y 轴对称（symmetric）（把 x 换成 $-x$），也关于 x 轴对称（把 y 换成 $-y$）。把四分之一的曲线分别对两条轴作反射：

- 环变成介乎 $x=-2$ 与 $x=2$ 之间的“8”字形（figure-of-eight），且 $|y|\leq 1$。两瓣沿直线 $y=\pm\frac{2}{\sqrt{5}}x$ 在原点相交，转向点（turning point）位于 $(\pm\sqrt{2},\pm 1)$。
- 翼变成两支，一支在 $y\geq 2$，另一支在 $y\leq -2$。每支在 $(0,\pm\sqrt{5})$ 处有局部极大值（local maximum）或局部极小值，转向点位于 $(\pm\sqrt{2},\pm 2)$，两支都从外侧趋近渐近线 $y=\pm x$。

曲线没有任何部分位于带状区域 $1<|y|<2$ 内。



验算

- 每个标示的点都满足方程。例如在 $(\sqrt{2}, 2)$，两边都等于 -4；在 $(0, \sqrt{5})$，两边都等于 0。
- 点 $(1, 1.5)$ 位于禁区带内。在该处左边为 $2.25 \times (-2.75) \approx -6.19$，而右边 $x^2(x^2-4)$ 永远不会小于 -4（这是它在 $x^2 = 2$ 处的最小值）。这正是判别式论证的缩影。
- 在翼上取 $y = 100$，公式给出 $x^2 = 2 + \sqrt{(10^4-1)(10^4-4)}$。于是 $4x(y-x) \approx 1.0002$，与 $y - x \approx \frac{1}{4x}$ 吻合。在原点附近，环上的比值 y/x 趋向 $0.8944 \approx 2/\sqrt{5}$。

常见错误

- 把 $(0,0)$ 列为切线水平（或垂直）的点。在该处 $\mathrm{d}y/\mathrm{d}x$ 是 $0/0$，真正的斜率要由 (b) 得出。
- 把环画成沿 $y = x$ 离开原点，或把翼画成与 $y = x$ 相交。环的起始斜率是 $2/\sqrt{5} < 1$，而翼严格保持在 $y = x$ 上方。
- 忘记条件 $x^2 \geq 0$——正是这个条件令内侧的根在 $y = \sqrt{5}$ 处终止。

STEP 之外

魔鬼曲线是一条代数曲线（algebraic curve）：即多项式 $F(x,y)$ 的零点集。原点是一个奇点（singular point），在该处两个偏导数（partial derivative）F_x 和 F_y 都等于零。F 的最低次项给出该处的切线，这正是我们在 (b) 所做的；而最高次项则给出渐近线的方向。到了大学，这些概念会在隐函数定理（implicit function theorem）和代数几何（algebraic geometry）中再次出现。Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多 STEP 风格的曲线描绘题，并附有详尽评注。

详细解答

STEP III 2019，第 8 题

纯数学 · 向量与三维几何

STEP III 2019，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

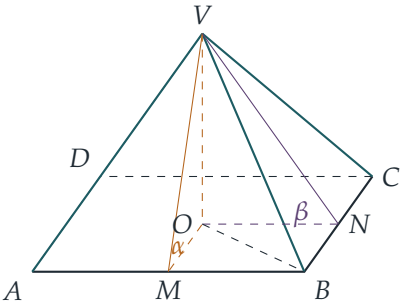
先为棱锥（pyramid）建立坐标。每个斜面都包含一条底边，而底边是水平的，因此该斜面的法向量（normal）必与这条底边垂直。这就迫使法向量落在一个铅垂平面内，而在这个平面上，斜面与底面之间的夹角可以直接读出。之后，两个斜面之间的夹角可由两条法向量的点积（dot product）求得，其余部分就只是细心的三角代数运算，再加上一次算术－几何平均不等式（AM–GM）。

解答

设底面的中心为原点 O，$\mathbf{i}$ 沿 AB 方向，$\mathbf{j}$ 沿 BC 方向，$\mathbf{k}$ 向上。取

$$A=(-a,-b,0),\quad B=(a,-b,0),\quad C=(a,b,0),\quad D=(-a,b,0),\quad V=(0,0,h),$$

其中 $a,b,h>0$。



设 $M=(0,-b,0)$ 为 AB 的中点，$N=(a,0,0)$ 为 BC 的中点。由于 $VA=VB$，直线 VM 位于面 AVB 内并与 AB 垂直；底面上的 OM 亦与 AB 垂直。因此该斜面与底面之间的

夹角就是角 VMO，并且

$$\tan\alpha = \frac{h}{b}, \qquad \text{同样地} \qquad \tan\beta = \frac{h}{a}.$$

第 (i) 部分

面 AVB 包含方向 $\overrightarrow{AB} \parallel \mathbf{i}$ 及 $\overrightarrow{MV} = b\,\mathbf{j} + h\,\mathbf{k}$。其一条法向量为

$$\mathbf{i} \times (b\,\mathbf{j} + h\,\mathbf{k}) = -h\,\mathbf{j} + b\,\mathbf{k}.$$

除以其长度 $\sqrt{h^2 + b^2}$，并利用 $\tan\alpha = h/b$，便得单位法向量（unit normal）

$$\mathbf{n}_1 = -\sin\alpha\,\mathbf{j} + \cos\alpha\,\mathbf{k}.$$

它的 $\mathbf{j}$ 分量为负，所以它指向棱锥外面（面 AVB 位于 $y < 0$ 的一侧）。

用完全相同的方法，面 BVC 包含 $\mathbf{j}$ 及 $\overrightarrow{NV} = -a\,\mathbf{i} + h\,\mathbf{k}$，其向外的单位法向量为

$$\mathbf{n}_2 = \sin\beta\,\mathbf{i} + \cos\beta\,\mathbf{k}.$$

因此

$$\mathbf{n}_1 \cdot \mathbf{n}_2 = \cos\alpha\cos\beta.$$

想一想 两条法向量之间的夹角，究竟是两个面之间的夹角，还是它的补角（supplement）？

考虑一个与公共棱 VB 垂直的截面。在截面上，两个面显示为两条直线，相交成内角 $\pi - \theta$，而每条向外法向量都与所属直线垂直，并指离立体。把其中一个面经内角转到另一个面上，向外法向量亦会转过同样的角度；但在截面上，两条法向量都指向外面，所以它们之间的夹角是 $\pi - (\pi - \theta) = \theta$。（想像一个立方体：两面成 $\frac{\pi}{2}$，法向量亦成 $\frac{\pi}{2}$；把棱锥压扁，两面的夹角趋向 π，而法向量的夹角则趋向 0。）因此

$$\cos\theta = \mathbf{n}_1 \cdot \mathbf{n}_2 = \cos\alpha\cos\beta,$$

此即所求。由于 $0 < \cos\alpha\cos\beta < 1$，$\theta$ 是锐角，这与两个面相交成钝角的情况一致。

第 (ii) 部分

从 V 所作垂线的垂足是 O，所以 BV 与底面之间的夹角是角 VBO，且

$$\cot\phi = \frac{OB}{h} = \frac{\sqrt{a^2 + b^2}}{h}.$$

因此

$$\cot^2\phi = \frac{a^2}{h^2} + \frac{b^2}{h^2} = \cot^2\beta + \cot^2\alpha,$$

这里用了 $\cot\beta = a/h$ 及 $\cot\alpha = b/h$。这就是第一个结果。

至于第二个结果，记 $p = \cos^2\alpha$，$q = \cos^2\beta$，则由第 (i) 部分得 $\cos^2\theta = pq$。于是

$$\cot^2\phi = \frac{p}{1-p} + \frac{q}{1-q} = \frac{p+q-2pq}{(1-p)(1-q)}.$$

由于 $\cos^2\phi = \dfrac{\cot^2\phi}{1+\cot^2\phi}$，我们需要

$$1+\cot^2\phi = \frac{(1-p)(1-q)+p+q-2pq}{(1-p)(1-q)} = \frac{1-pq}{(1-p)(1-q)},$$

两者相除，得

$$\cos^2\phi = \frac{p+q-2pq}{1-pq} = \frac{\cos^2\alpha+\cos^2\beta-2\cos^2\theta}{1-\cos^2\theta}.$$

（由于 $\cos\theta < 1$，分母为正。）

由于 $\cos\alpha$ 和 $\cos\beta$ 均为正，根据算术—几何平均不等式，

$$\cos^2\alpha+\cos^2\beta \;\geq\; 2\cos\alpha\cos\beta = 2\cos\theta,$$

又因 $1-\cos^2\theta > 0$，

$$\cos^2\phi \;\geq\; \frac{2\cos\theta-2\cos^2\theta}{1-\cos^2\theta}.$$

想一想 这个下界怎样告诉我们有关 ϕ 和 θ 的信息?

把它化简，再与 $\cos^2\theta$ 比较。约去正因子 $1-\cos\theta$，

$$\frac{2\cos\theta(1-\cos\theta)}{(1-\cos\theta)(1+\cos\theta)} = \frac{2\cos\theta}{1+\cos\theta}.$$

记 $c = \cos\theta \in (0,1)$，

$$\frac{2c}{1+c} - c^2 = \frac{c\,(2-c-c^2)}{1+c} = \frac{c\,(1-c)(2+c)}{1+c} > 0.$$

所以 $\cos^2\phi \geq \dfrac{2c}{1+c} > \cos^2\theta$。$\phi$ 和 θ 均位于 $(0, \frac{\pi}{2})$ 内，而余弦在此区间内为正且递减，所以由 $\cos\phi > \cos\theta$ 得 $\phi < \theta$。

验算

- 正方形底面，且 h 等于边长的一半（$a = b = h$）：$\alpha = \beta = \frac{\pi}{4}$，所以 $\cos\theta = \frac{1}{2}$，$\theta = 60°$。直接计算，$\cot\phi = \sqrt{2}$，所以 $\cot^2\phi = 2 = 1+1$，且 $\phi \approx 35.26° < 60°$。此外 $\cos^2\phi = \frac{2}{3}$，与 $(\frac{1}{2}+\frac{1}{2}-\frac{1}{2})/(1-\frac{1}{4}) = \frac{2}{3}$ 相符。
- 以电脑对 10^5 个随机棱锥进行验算，通过经 VB 的截面直接量度两面之间的夹角，结果与 $\cos\theta = \cos\alpha\cos\beta$ 及 ϕ 的两条公式均吻合，而且每次都得到 $\phi < \theta$。

常见错误

- 一个面取向内法向量，另一个面却取向外法向量：这样点积的正负号便会出错，得出的是 $\cos(\pi-\theta)$。
- 沿棱 VA 而非沿斜高（slant height）VM 量度 α；两个平面之间的夹角必须在与其交线垂直的方向上量度。
- 只推到 $\cos^2\phi \geq \dfrac{2\cos\theta}{1+\cos\theta}$ 便停下，而没有证明这个值大于 $\cos^2\theta$。

STEP 之外

两个平面之间的夹角等于它们法向量之间的夹角（至多相差一个补角），这正是多变量微积分（multivariable calculus）和微分几何（differential geometry）中用法向量描述曲面的起点。恒等式 $\cos\theta = \cos\alpha\cos\beta$ 其实亦是球面毕氏定理（spherical Pythagoras theorem）的变相形式：三条法向量 $\mathbf{k}$、$\mathbf{n}_1$、$\mathbf{n}_2$ 构成一个球面直角三角形。若想加深对点积和叉积（cross product）的体会，Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）是个不错的起点。

详细解答

STEP III 2019，第 9 题

力学 • 移动半球上的质点

STEP III 2019，第 9 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

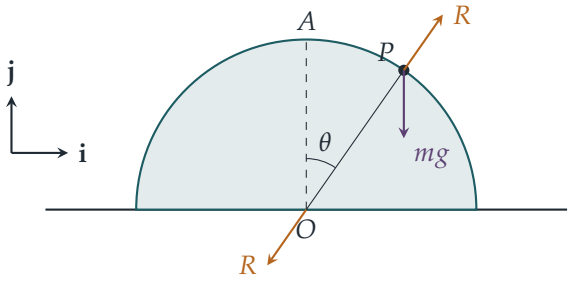
www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

两个接触面都是光滑（smooth）的，所以作用于整个系统（质点加半球）的外力都没有水平分量：水平动量守恒（conservation of horizontal momentum），这把半球的速率与质点的角速率联系起来。由于除重力外没有其他力做功，能量亦守恒（conservation of energy），这两条定律合起来便得出第 (ii) 部分。至于质点何时离开表面，最干净的判断准则是法向反作用力（normal reaction）变为零；题目建议的向量经过特意挑选，使我们取分量时，未知的 $\ddot{\theta}$ 会消失。

解答

模型：P 是一个质点，半球是刚体（rigid body），所有接触面都光滑。两者接触期间，P 与半球之间唯一的力是沿 OP 方向的法向反作用力 R，而桌面则对半球施加铅垂方向的推力。



作用于半球的反作用力（为清晰起见画在 O 处）与作用于 P 的反作用力大小相等、方向相反。

第 (i) 部分

在时间 t，球心 O 已移至 $-s\,\mathbf{i}$，而 P 与它的距离为 a，方向为 $\sin\theta\,\mathbf{i}+\cos\theta\,\mathbf{j}$。因此

$$\mathbf{r}=(a\sin\theta-s)\,\mathbf{i}+a\cos\theta\,\mathbf{j},$$

对 t 求导（对 $\sin\theta$ 和 $\cos\theta$ 使用连锁法则（chain rule)），得

$$\dot{\mathbf{r}}=(a\dot\theta\cos\theta-\dot s)\,\mathbf{i}-a\dot\theta\sin\theta\,\mathbf{j},$$

得证。

作用于系统的外力是两个重量和桌面的反作用力，全部都是铅垂方向的，所以总水平动量一直保持其初始值，即零（我们假设微小的初始位移所带来的速率可以忽略)。半球的速度是 $-\dot s\,\mathbf{i}$，所以

$$m(a\dot\theta\cos\theta-\dot s)-M\dot s=0\quad\Longrightarrow\quad\dot s=\frac{m}{m+M}\,a\dot\theta\cos\theta=(1-k)\,a\dot\theta\cos\theta,$$

因为 $1-k=\dfrac{m}{m+M}$。代入 $\dot{\mathbf{r}}$ 的 $\mathbf{i}$ 分量，

$$a\dot\theta\cos\theta-(1-k)a\dot\theta\cos\theta=k\,a\dot\theta\cos\theta,\qquad\text{所以}\qquad\dot{\mathbf{r}}=a\dot\theta\,(k\cos\theta\,\mathbf{i}-\sin\theta\,\mathbf{j}).$$

第 (ii) 部分

各反作用力所做的净功为零：桌面的反作用力与半球的运动方向垂直；而 P 与半球之间的一对力 $\pm R$ 所做的总功为零，因为 P 一直留在表面上（两者的相对速度沿着表面，与 R 垂直)。所以动能加势能守恒。P 已下降了 $a(1-\cos\theta)$，而起初所有物体（实际上）都是静止的，所以

$$\tfrac12 ma^2\dot\theta^2\left(k^2\cos^2\theta+\sin^2\theta\right)+\tfrac12 M\dot s^2=mga(1-\cos\theta).$$

利用第 (i) 部分，$M\dot s^2=M(1-k)^2a^2\dot\theta^2\cos^2\theta$。于是 $a^2\dot\theta^2\cos^2\theta$ 的总系数为

$$mk^2+M(1-k)^2=\frac{mM^2+Mm^2}{(m+M)^2}=\frac{mM}{m+M}=mk.$$

因此 $\tfrac12 ma^2\dot\theta^2\left(k\cos^2\theta+\sin^2\theta\right)=mga(1-\cos\theta)$，两边除以 $\tfrac12 ma$，

$$a\dot\theta^2\left(k\cos^2\theta+\sin^2\theta\right)=2g(1-\cos\theta).$$

第 (iii) 部分

设 $\mathbf{u}=\sin\theta\,\mathbf{i}+k\cos\theta\,\mathbf{j}$。对第 (i) 部分的最后结果求导，

$$\ddot{\mathbf{r}}=a\ddot\theta\,(k\cos\theta\,\mathbf{i}-\sin\theta\,\mathbf{j})-a\dot\theta^2\,(k\sin\theta\,\mathbf{i}+\cos\theta\,\mathbf{j}).$$

第一个括号与 $\mathbf{u}$ 垂直，因为 $k\cos\theta\sin\theta-\sin\theta\,k\cos\theta=0$。所以

$$\ddot{\mathbf{r}}\cdot\mathbf{u}=-a\dot\theta^2\left(k\sin^2\theta+k\cos^2\theta\right)=-k\,a\dot\theta^2.$$

想一想 为甚么这个向量特别有用?

它消去了我们一无所知的 $\ddot{\theta}$ 项，只剩下第 (ii) 部分已给出的 $\dot{\theta}^2$。现在与 P 的牛顿第二定律（Newton's second law）作比较：

$$m\ddot{\mathbf{r}} = -mg\,\mathbf{j} + R(\sin\theta\,\mathbf{i} + \cos\theta\,\mathbf{j}).$$

取沿 $\mathbf{u}$ 的分量，并与上面的结果比较，

$$-mk\,a\dot{\theta}^2 = -mgk\cos\theta + R\left(\sin^2\theta + k\cos^2\theta\right),$$

所以

$$R\left(\sin^2\theta + k\cos^2\theta\right) = mk\left(g\cos\theta - a\dot{\theta}^2\right).$$

当 R 降至零时，质点便离开半球。由于 $k > 0$，在 $\theta = \alpha$ 时得

$$a\dot{\theta}^2 = g\cos\alpha.$$

三次方程。 设 $c = \cos\alpha$，并在 $\theta = \alpha$ 时使用第 (ii) 部分，配合 $\sin^2\alpha = 1 - c^2$：

$$gc\left(kc^2 + 1 - c^2\right) = 2g(1 - c) \quad\Longrightarrow\quad (1-k)\,c^3 - 3c + 2 = 0.$$

设 $f(x) = (1-k)x^3 - 3x + 2$，其中 $0 < 1 - k < 1$，因为 m 和 M 都是正数。

- 对于 $-1 \le x \le 1$，$f'(x) = 3\left((1-k)x^2 - 1\right) < 0$，所以 f 在该区间严格递减（strictly decreasing），在 $[-1, 1]$ 内最多只有一个根。
- $f\left(\frac{2}{3}\right) = \frac{8}{27}(1-k) - 2 + 2 = \frac{8}{27}(1-k) > 0$，而 $f(1) = -k < 0$。

所以 $[-1, 1]$ 内唯一的根严格位于 $\frac{2}{3}$ 与 1 之间。由于 $\cos\alpha$ 必须位于 $[-1, 1]$ 内，它就是这个根，因此

$$\cos\alpha > \tfrac{2}{3}.$$

（这与 R 的公式一致：代入第 (ii) 部分可见 $g\cos\theta - a\dot{\theta}^2$ 与 $-f(\cos\theta)$ 同号，它在 $\theta = 0$ 时为正，并在这个根处首次变为零，所以 P 确实一直保持接触，直至那一刻。）

验算

- 若 $M \to \infty$，则 $k \to 1$，半球保持不动，三次方程变为 $-3c + 2 = 0$：即质点从固定球面滑落时大家熟悉的 $\cos\alpha = \frac{2}{3}$。半球会移动，令质点较早离开，即在较大的 $\cos\alpha$ 处离开。
- 当两者质量相等时，$1 - k = \frac{1}{2}$，三次方程 $c^3 - 6c + 4 = 0$ 可因式分解为 $(c-2)(c^2 + 2c - 2) = 0$，得 $\cos\alpha = \sqrt{3} - 1 \approx 0.732$，大于 $\frac{2}{3}$。

- 用电脑对完整运动作逐步数值模拟（numerical simulation），测试了多个质量比，过程中证实了第 (i) 和第 (ii) 部分的关系式，并发现当 $m = M$ 时反作用力在 $\cos\theta \approx 0.732$ 处变为零，与三次方程吻合。

常见错误

- 未经论证便使用固定球面的条件 $a\dot{\theta}^2 = g\cos\theta$：这里半球有加速度，所以质点的加速度并不只是绕固定点的圆周运动。这条件在离开的一刻恰好成立，只是因为 $R = 0$ 意味着半球在那一刻没有水平加速度。
- 在能量方程中漏掉半球的动能 $\frac{1}{2}M\dot{s}^2$。
- 只证明了有一个根位于 $\frac{2}{3}$ 与 1 之间，却没有排除 $[-1, 1]$ 内存在另一个根。

STEP 之外

这是一个具两个自由度（degrees of freedom）的系统，拉格朗日力学（Lagrangian mechanics）可以有系统地处理它：用 θ 和 s 写出动能和势能，由于 s 不出现在拉格朗日量（Lagrangian）中，水平动量守恒便自动出现（这是一个“可忽略”坐标（ignorable coordinate），由诺特（Noether）定理与一种对称性联系起来）。像 R 这样的约束力（constraint forces）则在之后才求出，正如我们在这里所做的一样。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）是先建立牛顿力学直觉的好起点。

详细解答

STEP III 2019，第 10 题

力学 • 斜碰撞

STEP III 2019，第 10 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

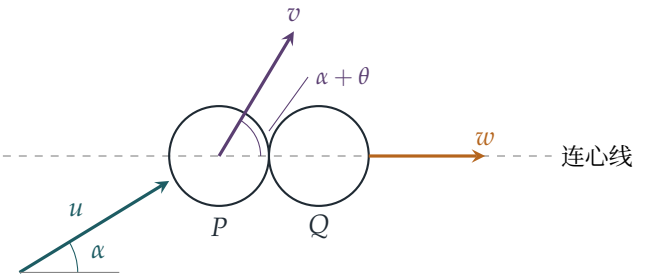
www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

把每个速度分解为沿连心线 (line of centres) 的分量和垂直于连心线的分量。由于两球都是光滑的，冲量 (impulse) 只沿连心线作用，所以 P 的垂直分量保持不变，而 Q 则沿该直线运动。动量守恒 (conservation of momentum) 和牛顿实验定律 (Newton's experimental law) 在连心线方向上的应用，就跟一维碰撞完全一样。余下的只是三角学。

解答

把 P 和 Q 视为放在光滑水平桌面上、质量同为 m 的光滑球体。取碰撞瞬间的连心线为 x 方向，由 P 指向 Q；并设 y 方向在运动平面内与之垂直，指向 P 初速度所在的一侧。



第 (i) 部分

碰撞前，P 沿连心线的分量为 $u\cos\alpha$，垂直于连心线的分量为 $u\sin\alpha$。

光滑球体之间的接触力沿连心线作用。因此 P 在垂直于该直线的方向上不受冲量，P 的垂直分量保持不变。此外，Q 起初静止，而且只受沿连心线的冲量，所以它以速率 w 沿

该直线运动。

> **想一想** 为甚么 P 的新方向与连心线所成的角是 $\alpha+\theta$，而不是 $\alpha-\theta$?

P 所受的冲量沿连心线指向后方。它把 P 的 x 分量由 $u\cos\alpha$ 减小至下面求得的 v_x，而不影响 y 分量。因此速度方向偏离连心线，与该直线所成的角由 α 增大至 $\alpha+\theta$。下面我们会证实当 $e<1$ 时 $v_x>0$，所以 $0<\alpha+\theta<\frac{\pi}{2}$。

碰撞后，P 沿连心线的分量为 $v\cos(\alpha+\theta)$，垂直于连心线的分量为 $v\sin(\alpha+\theta)$。令垂直分量相等，得

$$u\sin\alpha = v\sin(\alpha+\theta),$$

正如所求。

沿连心线方向，由动量守恒（两者质量相等，所以可约去 m）得

$$u\cos\alpha = v\cos(\alpha+\theta)+w. \tag{1}$$

代入第一个结果中的 $u=v\sin(\alpha+\theta)/\sin\alpha$，

$$\begin{aligned} w &= \frac{v\sin(\alpha+\theta)\cos\alpha}{\sin\alpha} - v\cos(\alpha+\theta) \\ &= \frac{v\left[\sin(\alpha+\theta)\cos\alpha-\cos(\alpha+\theta)\sin\alpha\right]}{\sin\alpha} = \frac{v\sin\theta}{\sin\alpha}, \end{aligned}$$

这里用了 $\sin\big((\alpha+\theta)-\alpha\big)$ 的展开式。所以

$$w=\frac{v\sin\theta}{\sin\alpha}.$$

这个结果有一个漂亮的诠释。向量形式的动量守恒给出 $\mathbf{u}=\mathbf{v}+\mathbf{w}$，沿垂直于 $\mathbf{u}$ 的方向分解，便得 $0=v\sin\theta-w\sin\alpha$。

第 (ii) 部分

沿连心线应用牛顿实验定律：分离速率等于接近速率的 e 倍：

$$w-v\cos(\alpha+\theta)=eu\cos\alpha. \tag{2}$$

代入第 (i) 部分的 $w=v\sin\theta/\sin\alpha$ 和 $u=v\sin(\alpha+\theta)/\sin\alpha$：

$$\frac{v\sin\theta}{\sin\alpha}-v\cos(\alpha+\theta)=\frac{ev\sin(\alpha+\theta)\cos\alpha}{\sin\alpha}.$$

两边乘以 $\sin\alpha/v$（因为 $v>0$ 且 $\sin\alpha\neq0$，所以可以这样做），整理后得

$$\sin\theta=\cos(\theta+\alpha)\sin\alpha+e\sin(\theta+\alpha)\cos\alpha,$$

正如所求。

要求 $\tan\theta$，先把两个复角展开：

$$\sin\theta = (\cos\theta\cos\alpha - \sin\theta\sin\alpha)\sin\alpha + e(\sin\theta\cos\alpha + \cos\theta\sin\alpha)\cos\alpha.$$

把含 $\sin\theta$ 的项移到左边：

$$\sin\theta\left(1 + \sin^2\alpha - e\cos^2\alpha\right) = (1+e)\cos\theta\sin\alpha\cos\alpha.$$

两边除以 $\cos\theta\cos^2\alpha$，并记 $t = \tan\alpha$，利用 $1/\cos^2\alpha = 1 + t^2$：

$$\tan\theta\left(1 + t^2 + t^2 - e\right) = (1+e)\,t, \qquad \text{所以} \qquad \tan\theta = \frac{(1+e)\tan\alpha}{1 - e + 2\tan^2\alpha}.$$

由于 $e < 1$，分母为正，所以 $\tan\theta > 0$，而 θ 的确是一个锐角偏转。

$\tan\theta$ 的最大值

当 α 取遍所有锐角时，$t = \tan\alpha$ 取遍 $(0, \infty)$ 中的每一个值。对于 $t > 0$，可以把分子和分母同除以 t：

$$\tan\theta = \frac{1+e}{\dfrac{1-e}{t} + 2t}.$$

> **想一想** 不用微积分，怎样求这个式子的最大值?

尽量令分母最小。$(1-e)/t$ 和 $2t$ 都是正数，所以由算术－几何平均不等式（AM–GM inequality），

$$\frac{1-e}{t} + 2t \geq 2\sqrt{\frac{1-e}{t}\cdot 2t} = 2\sqrt{2(1-e)},$$

等号成立当且仅当 $(1-e)/t = 2t$，即 $\tan\alpha = \sqrt{(1-e)/2}$。这是一个真正的锐角，所以最小值确实能够达到。因此 $\tan\theta$ 的最大值为

$$\tan\theta_{\max} = \frac{1+e}{2\sqrt{2(1-e)}}.$$

（对 t 求导，会得到相同的驻点 $2t^2 = 1 - e$。）

验算

- 接近弹性的极限 $e \to 1$：两球交换沿连心线的分量，所以 P 只剩下 $u\sin\alpha$，并沿垂直于连心线的方向运动。于是 $\alpha + \theta = \frac{\pi}{2}$，而 $\tan\theta = \cot\alpha$。公式给出 $\tan\theta = 2t/(2t^2) = 1/t$，两者一致。最大值趋于无穷大，这是应有的结果，因为当 $\alpha \to 0$ 时 $\theta \to \frac{\pi}{2}$。
- 当 $e = \frac{1}{2}$ 时，最大值为 $\frac{3/2}{2\sqrt{1}} = \frac{3}{4}$，在 $\tan\alpha = \frac{1}{2}$（约 26.6°）时达到。以细小步长扫描 α，证实最大值为 0.75。用电脑模拟大量随机碰撞，亦重现了题目中全部四个结果。
- α 很小（近乎正面碰撞）：$\tan\theta \approx \frac{1+e}{1-e}\,\alpha \to 0$，所以正面碰撞不会造成偏转，与预期相符。

常见错误

- 把牛顿实验定律应用于整个速率，而不是沿连心线的分量。恢复（restitution）只作用于冲量的方向。
- 把 P 的新角度写成 $\alpha - \theta$。冲量减去了部分沿连心线的分量，所以 P 是转离该直线的。
- 忘记正是 $e < 1$ 令 $1 - e + 2\tan^2\alpha$ 为正，也忘记 $\tan\alpha$ 取遍整个 $(0, \infty)$，因此算术－几何平均不等式的等号情况确实能够达到。

STEP 之外

光滑球体的斜碰（oblique collision）是散射（scattering）最简单的例子。在这道题中，同样两条守恒定律分别沿一条直线及垂直于该直线应用，而这个想法可以推广到粒子物理和气体分子运动论（kinetic theory of gases）。对于等质量的弹性碰撞，结果就是著名的“碰撞后两速度互相垂直”。在质心参考系（centre-of-mass frame）中，每次碰撞都只是把相对速度旋转（在本题中还会把它缩短），这是理解上述所有结果的一个更简洁的方法。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）正是从这个观点讨论碰撞的。

详细解答

STEP III 2019，第 11 题

概率与统计 · 泊松稀疏化

STEP III 2019，第 11 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

关键思路

一切都建基于以到店顾客人数或取沙人数作条件（conditioning），然后对一个泊松（Poisson）级数求和。第 (i) 部分是经典的泊松分布“稀疏化”（thinning）：把二项概率与泊松概率相乘之后，对那个看不见的总数求和，结果会收缩成一个指数函数。一旦知道取沙人数服从泊松分布，第 (ii) 及 (iii) 部分便只需要求 $(1-k)^Y$ 的期望值（expectation），而那又是另一个指数级数。

解答

设 N 为一天内的顾客人数，故 $N \sim \mathrm{Po}(\lambda)$；设 Y 为取沙的人数。我们假设顾客之间的决定互相独立（independent），并且与 N 独立，因此在 $N = n$ 的条件下，有 $Y \sim \mathrm{B}(n, p)$。

第 (i) 部分

对 $y = 0, 1, 2, \ldots$，以 N 作条件（只有 $n \geq y$ 的项有贡献）：

$$\begin{aligned} \mathrm{P}(Y = y) &= \sum_{n=y}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{y} p^y (1-p)^{n-y} \\ &= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \sum_{n=y}^{\infty} \frac{\big(\lambda(1-p)\big)^{n-y}}{(n-y)!} \qquad \text{由于} \frac{\lambda^n}{n!} \cdot \frac{n!}{y!\,(n-y)!} = \frac{\lambda^y \lambda^{n-y}}{y!\,(n-y)!} \\ &= \mathrm{e}^{-\lambda} \frac{(p\lambda)^y}{y!} \mathrm{e}^{\lambda(1-p)} = \mathrm{e}^{-p\lambda} \frac{(p\lambda)^y}{y!}, \end{aligned}$$

其中我们令 $m = n - y$，并用了 $\sum_{m \geq 0} x^m / m! = \mathrm{e}^x$。这正是 $\mathrm{Po}(p\lambda)$ 的概率，所以 $Y \sim \mathrm{Po}(p\lambda)$，得证。

第 (ii) 部分

每位取沙的顾客都会留下原有沙量的 $1-k$ 比例。所以经过 Y 位取沙顾客后，剩下的质量是 $S(1-k)^Y$，而被取走的质量是

$$T = S\big(1-(1-k)^Y\big).$$

由第 (i) 部分，

$$\mathrm{E}\big[(1-k)^Y\big] = \sum_{y=0}^{\infty}(1-k)^y\,\mathrm{e}^{-p\lambda}\frac{(p\lambda)^y}{y!} = \mathrm{e}^{-p\lambda}\,\mathrm{e}^{p\lambda(1-k)} = \mathrm{e}^{-kp\lambda}.$$

因此，由期望值的线性性质（linearity of expectation），

$$\mathrm{E}(T) = S\big(1-\mathrm{E}[(1-k)^Y]\big) = \big(1-\mathrm{e}^{-kp\lambda}\big)S,$$

这正是所求的结果。

第 (iii) 部分

想一想 怎样把“取走沙的 k 比例”转化为关于某一粒特定沙粒的概率?

由于沙粒极多，而且没有任何一粒被特别挑出，所以每当有人取走剩余沙的 k 比例时，一粒仍在袋中的沙粒被取走的概率是 k，并且与之前发生的事独立。（等价地说，金沙粒落在某一份沙中的机会，等于该份沙占总质量的比例。）

设 G 为店员取得金沙粒这一事件。在 $Y=y$ 的条件下，金沙粒在全部 y 位顾客手下幸存的概率是 $(1-k)^y$，之后店员取得它的概率是 k。所以

$$\mathrm{P}(G \mid Y=y) = k(1-k)^y,$$

再利用第 (ii) 部分求得的期望值，

$$\mathrm{P}(G) = k\,\mathrm{E}\big[(1-k)^Y\big] = k\,\mathrm{e}^{-kp\lambda}.$$

这与第 (ii) 部分吻合：金沙粒在当天幸存的概率是 $\mathrm{e}^{-kp\lambda}$，恰好就是剩余沙量的期望比例。

*$k=0$ 的情况。*此时 $\mathrm{P}(G)=0$。这是合理的：当 $k=0$ 时，没有人（包括店员）会取走任何沙，所以金沙粒一直留在袋中。

*$k\to 1$ 的极限。*此时 $\mathrm{P}(G)\to \mathrm{e}^{-p\lambda}$，即 $\mathrm{P}(Y=0)$。当 k 接近 1 时，第一位取沙的顾客几乎会把沙全部取走，金沙粒几乎必然也在其中。店员基本上只有在没有任何顾客取沙时才能得到金沙粒，而那时店员会取走（几乎）整袋沙。

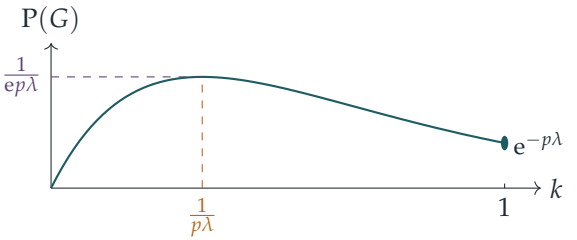
*求最大值。*记 $\mu = p\lambda$，并对 $0\le k<1$ 设 $f(k) = k\,\mathrm{e}^{-\mu k}$。则

$$f'(k) = \mathrm{e}^{-\mu k}(1-\mu k),$$

它在 $k < 1/\mu$ 时为正，在 $k > 1/\mu$ 时为负。由于 $\mu > 1$，点 $k = 1/\mu$ 位于 $(0,1)$ 内，所以 f 在该点之前递增、之后递减，该点给出 $[0,1)$ 上的最大值。因此概率在

$$k = \frac{1}{p\lambda},$$

时取得最大值，而最大概率是 $\dfrac{1}{\mathrm{e}\,p\lambda}$。（若 $p\lambda \le 1$，则 f 在整个 $[0,1)$ 上递增，没有最大值，只有当 $k \to 1$ 时的上确界（supremum）$\mathrm{e}^{-p\lambda}$。这就是题目限定 $p\lambda > 1$ 的原因。）



上图显示 $p\lambda = 3$ 时的 $\mathrm{P}(G) = k\,\mathrm{e}^{-kp\lambda}$。

验算

- 第 (i) 部分通过明显的检验：$p = 1$ 时还原为 $\mathrm{Po}(\lambda)$，而 $p = 0$ 时必然有 $Y = 0$。
- 在第 (ii) 部分中，$k = 0$ 给出取走的沙量为零；而当 $\lambda \to \infty$ 时，取走质量的期望值趋向 S，理应如此。
- 取 $\lambda = 3$、$p = 0.6$、$k = 0.4$，公式给出 $\mathrm{P}(G) = 0.4\,\mathrm{e}^{-0.72} \approx 0.1947$。以电脑模拟 200 000 天，得到 0.1953。
- 取 $p\lambda = 3$，在细密网格上 $k\,\mathrm{e}^{-3k}$ 的最大值出现在 $k \approx 0.3333$，其值为 $0.1226 \approx 1/(3\mathrm{e})$。

常见错误

- 把取走的沙量写成 kSY，仿佛每位顾客都取走*原来*沙量的 k 比例。其实每位顾客取走的是*剩下*沙量的一部分，所以各次取走的量构成一个等比数列（geometric sequence）。
- 把 $\mathrm{E}[(1-k)^Y]$ 算成 $(1-k)^{\mathrm{E}(Y)} = (1-k)^{p\lambda}$。非线性函数的期望值并不等于期望值的函数。
- 忘记检查 $1/(p\lambda)$ 是否真的落在容许范围 $0 \le k < 1$ 内。条件 $p\lambda > 1$ 正是用在这里。

STEP 之外

第 (i) 部分就是泊松分布的稀疏化性质。第 (ii) 及 (iii) 部分中出现的期望值 $\mathrm{E}[s^Y] = \mathrm{e}^{\mu(s-1)}$，是 $\mathrm{Po}(\mu)$ 的概率母函数（probability generating function）在 $s = 1-k$ 处的值。利用母函数，稀疏化几乎一行就能证明：若 $\mathrm{E}[s^N] = \mathrm{e}^{\lambda(s-1)}$，则 $\mathrm{E}[s^Y] = \mathrm{E}[(1-p+ps)^N] = \mathrm{e}^{\lambda p(s-1)}$。在泊松过程（Poisson process）的理论中，这就成为以下命题：对泊松过程的点作独立标记，所得的仍然是泊松过程。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）对母函数及这些概念有清晰的论述。

详细解答

STEP III 2019，第 12 题

概率与统计 · 随机子集

STEP III 2019，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696590-2019-past-papers.zip

本题在 *Beyond STEP* 第 9.12 节随机子集中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/random-subsets.html

关键思路

从 S 中均匀随机（uniformly at random）地选取一个子集（subset），等同于逐一考虑整数 $1, 2, \ldots, n$，并为每个整数掷一枚公平硬币，以决定它是否放入子集。一旦看清这一点，题中每个事件（event）都变成关于每个整数各自的陈述，各整数之间互相独立（independent），而答案就是某个小概率的 n 次方。

解答

数算子集

对 n 个整数中的每一个决定它“在内”或“在外”，便完全确定了 S 的一个子集。每个整数有 2 种选择，各自独立地作出，而不同的选择模式给出不同的子集。因此，连同 $\varnothing$（全部在外）和 S（全部在内）计算在内，子集的数目为

$$2 \times 2 \times \cdots \times 2 = 2^n.$$

想一想 为甚么可以把各整数视为独立的掷硬币？

设 A 在 T 上均匀分布。2^n 种在内／在外模式各对应一个子集，所以每种模式的概率都是 $2^{-n} = (\frac{1}{2})^n$。这正是掷 n 枚独立公平硬币时得到某一特定模式的概率。因此事件“$1 \in A$”、“$2 \in A$”、...、“$n \in A$”互相独立，各自的概率为 $\frac{1}{2}$。由于 $A_1, \ldots, A_m$ 彼此之

间亦是独立选取的，所以全部 mn 个事件“$i \in A_k$”都互相独立，各自的概率为 $\frac{1}{2}$。

第 (i) 部分

2^n 个子集中恰好有一半包含 1（把每个包含 1 的子集与移走 1 后的同一子集配对）。所以

$$\mathrm{P}(1 \in A_1) = \tfrac{1}{2}.$$

第 (ii) 部分

对整数 i，设 E_i 为“i 并非同时属于 A_1 和 A_2”这一事件。那么 $A_1 \cap A_2 = \varnothing$ 当且仅当每个 E_i 都发生。由独立性，

$$\mathrm{P}(i \in A_1 \text{ 且 } i \in A_2) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \qquad \text{所以} \qquad \mathrm{P}(E_i) = \tfrac{3}{4}.$$

事件 E_i 只取决于整数 i 的硬币，所以 $E_1, \ldots, E_n$ 互相独立，并且

$$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \mathrm{P}(E_1) \cdots \mathrm{P}(E_n) = \left(\tfrac{3}{4}\right)^n,$$

此即所求。

同一论证适用于任意数目的集合。整数 i 属于 $A_1 \cap A_2 \cap A_3$ 的概率为 $(\frac{1}{2})^3 = \frac{1}{8}$，所以它不在该交集（intersection）中的概率为 $\frac{7}{8}$，从而

$$\mathrm{P}(A_1 \cap A_2 \cap A_3 = \varnothing) = \left(\tfrac{7}{8}\right)^n.$$

一般而言，i 同时属于全部 m 个集合的概率为 2^{-m}，所以

$$\mathrm{P}(A_1 \cap A_2 \cap \cdots \cap A_m = \varnothing) = \left(1 - \frac{1}{2^m}\right)^n.$$

第 (iii) 部分

现在 $A_1 \subseteq A_2$ 的意思是没有任何整数属于 A_1 而不属于 A_2。对每个整数 i，（是否在 A_1? 是否在 A_2?）共有四种等可能的情况，其中只有“在 A_1、不在 A_2”一种是不容许的。所以每个整数以概率 $\frac{3}{4}$ 符合要求，且互相独立，因此

$$\mathrm{P}(A_1 \subseteq A_2) = \left(\tfrac{3}{4}\right)^n.$$

（这与第 (ii) 部分吻合：$A_1 \subseteq A_2$ 当且仅当 A_1 与 A_2 的补集（complement）不相交（disjoint），而均匀随机子集的补集仍然是均匀随机的。）

至于 $A_1 \subseteq A_2 \subseteq A_3$，考虑某一整数 i 在 (A_1, A_2, A_3) 中的归属模式，以 1 表示“在内”，0 表示“在外”。链（chain）条件表示一旦 i 属于某个 A_k，它就属于其后所有集合，所以模式必须是以下其中之一：

$$000, \quad 001, \quad 011, \quad 111.$$

这是 8 种等可能模式中的 4 种，所以每个整数符合要求的概率为 $\frac{4}{8}=\frac{1}{2}$，从而

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3)=\left(\frac{1}{2}\right)^n.$$

对于由 m 个集合组成的链，容许的模式是一段 0 之后接一段 1：它由 1 开始出现的位置决定，这位置是 $1,2,\ldots,m$ 之一，或者“从不出现”。这样在 2^m 种模式中共有 $m+1$ 种容许模式，所以

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m)=\left(\frac{m+1}{2^m}\right)^n.$$

验算

- 当 $m=2$ 时，一般公式给出 $(1-\frac{1}{4})^n$ 和 $(\frac{3}{4})^n$；当 $m=3$ 时，给出 $(\frac{7}{8})^n$ 和 $(\frac{4}{8})^n$，与特殊情况吻合。当 $m=1$ 时，给出 $(\frac{1}{2})^n$（只有 $A_1=\varnothing$ 符合）和 1（单一集合必然是一条链），正如所料。
- 当 $n=1$ 时，所有情况都可以用手列出。例如在四种选择中，只有 $A_1=A_2=\{1\}$ 的交集非空，得出 $\frac{3}{4}$。
- 用电脑列出所有集合选择的验算，确认了各公式在 n 和 m 较小时均成立。例如当 $n=3$ 时，$\mathrm{P}(A_1\cap A_2=\varnothing)=\frac{27}{64}$，而 $\mathrm{P}(A_1\subseteq A_2\subseteq A_3)=\frac{1}{8}$。

常见错误

- 以“A_1 的每个元素都必须在 A_2 中”为理由，答 $\mathrm{P}(A_1\subseteq A_2)$ 为 $\left(\frac{1}{2}\right)^n$，却没有考虑到 A_1 本身的大小也是随机的。
- 把链的模式数目算成 m 而非 $m+1$，漏掉了整数不属于任何集合的那种模式。

STEP 之外

把随机子集表示为一串独立的掷硬币结果（即它的*指示向量*（indicator vector））是组合学（combinatorics）中的标准工具，而这里对链的计数正是*偏序集*（posets）及布尔格（Boolean lattice）理论的起点。同样的想法只需很少额外功夫，便可得出链的期望（expected）数目，或随机集合覆盖 S 的概率。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（《概率导论》）正是以这种精神发展独立性及指示变量（indicator variables）的理论。

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

2019 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。

