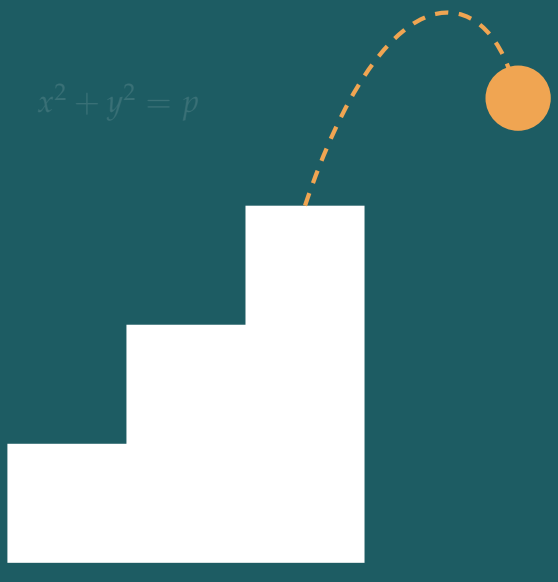
STEP 备试 · 大学数学

Beyond STEP

STEP II 及 III

2020 年详细解答



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP II 及 III

2020 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2020 年详细解答

2020 年 STEP 试卷（STEP II 及 III）全部 24 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 II

第 1 题 代换积分法 纯数学 . . . 3

第 2 题 可分离变量微分方程 纯数学 . . . 7

第 3 题 对数凹数列 纯数学 . . . 11

第 4 题 三角不等式 纯数学 . . . 15

第 5 题 数字和与整除性 纯数学 . . . 19

第 6 题 矩阵的幂与迹 纯数学 . . . 22

第 7 题 复数变换 纯数学 . . . 26

第 8 题 多项式与积分 纯数学 . . . 30

第 9 题 相撞的抛体 力学 . . . 34

第 10 题 能量与平衡 力学 . . . 38

第 11 题 掷硬币中的游程 概率与统计 . . . 42

第 12 题 不公平骰子的不等式 概率与统计 . . . 46

STEP III

第 1 题 归约公式 纯数学 . . . 49

第 2 题 双曲函数曲线描绘 纯数学 . . . 52

第 3 题 几何中的复数 纯数学 . . . 57

第 4 题 反射矩阵与旋转矩阵 纯数学 . . . 61

第 5 题 部分分式 纯数学 . . . 65

第 6 题 极坐标曲线与面积 纯数学 . . . 69

第 7 题 二阶微分方程 纯数学 . . . 74

第 8 题 数列与递降法 纯数学 . . . 77

第 9 题 刚体平衡 力学 . . . 81

第 10 题 简谐运动 力学 . . . 85

第 11 章 随机变量的函数 概率与统计 89
第 12 章 任何分布的独立性 概率与统计 93

详细解答

STEP II 2020，第 1 题

纯数学 · 代换积分法

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

每个被积函数（integrand）都是两个线性因子的幂的乘积，而且指数之和为 -2。对于这类被积函数，合适的代换（substitution）是两个线性因子的比（ratio）（或其平方根）：它能一次过吸收两个括号，剩下的只是一个简单的幂或标准的反正切（arctangent）。第 (i) 部分利用题目给出的代换展示这个过程，第 (ii) 部分要求你仿效这个技巧，而第 (iii) 部分则要求你自行发现它。

解答

第 (i) 部分

设 $x = \dfrac{1}{1-u}$，则有

$$x - 1 = \frac{1-(1-u)}{1-u} = \frac{u}{1-u}, \qquad \frac{\mathrm{d}x}{\mathrm{d}u} = \frac{1}{(1-u)^2}.$$

当 u 取遍 $(0,1)$ 时，x 取遍 $(1,\infty)$，且 $u = 1 - \frac{1}{x} = \frac{x-1}{x}$。由于 $0 < u < 1$，所有根号内的量都是正数，因此

$$x^{3/2}(x-1)^{1/2} = \frac{1}{(1-u)^{3/2}} \cdot \frac{u^{1/2}}{(1-u)^{1/2}} = \frac{u^{1/2}}{(1-u)^2}.$$

这时 $(1-u)^2$ 的因子恰好与 $\mathrm{d}x/\mathrm{d}u$ 互相抵消：

$$\int \frac{1}{x^{3/2}(x-1)^{1/2}}\,\mathrm{d}x = \int \frac{(1-u)^2}{u^{1/2}} \cdot \frac{1}{(1-u)^2}\,\mathrm{d}u = \int u^{-1/2}\,\mathrm{d}u$$
$$= 2u^{1/2} + c = 2\sqrt{\frac{x-1}{x}} + c.$$

第 (ii) 部分

在第 (i) 部分中，$u = \frac{x-1}{x}$ 正是两个线性因子之比。这里我们仿效这个想法，取

$$u = \frac{x-2}{x+1}, \qquad \text{等价地} \qquad x+1 = \frac{3}{1-u}, \quad x-2 = \frac{3u}{1-u},$$

因为 $x-2 = (x+1)-3$。当 $x > 2$ 时，得 $0 < u < 1$，且 $\dfrac{\mathrm{d}x}{\mathrm{d}u} = \dfrac{3}{(1-u)^2}$。于是

$$(x-2)^{3/2}(x+1)^{1/2} = \frac{3^{3/2}u^{3/2}}{(1-u)^{3/2}} \cdot \frac{3^{1/2}}{(1-u)^{1/2}} = \frac{9u^{3/2}}{(1-u)^2},$$

所以

$$\begin{aligned}\int \frac{1}{(x-2)^{3/2}(x+1)^{1/2}}\,\mathrm{d}x &= \int \frac{(1-u)^2}{9u^{3/2}} \cdot \frac{3}{(1-u)^2}\,\mathrm{d}u = \frac{1}{3}\int u^{-3/2}\,\mathrm{d}u \\ &= -\frac{2}{3}u^{-1/2} + c = -\frac{2}{3}\sqrt{\frac{x+1}{x-2}} + c.\end{aligned}$$

求导可以验证这个结果。设 $f = \frac{x+1}{x-2}$，则 $f' = -\frac{3}{(x-2)^2}$，因此

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

第 (iii) 部分

这次题目没有给出代换。由于多了一个因子 $x-1$，被积函数并不是某个比的纯幂，但两个平方根与之前属同一类型。

> **想一想** 前两个答案都是线性因子之比的平方根。如果把这样一个平方根取作新变量，会发生甚么事?

试设

$$t = \sqrt{\frac{x-2}{3x-2}}.$$

当 x 由 2 增加至 ∞ 时，$\frac{x-2}{3x-2} = \frac{1}{3} - \frac{4}{3(3x-2)}$ 由 0 递增并趋向 $\frac{1}{3}$，所以 t 由 0 递增至 $\frac{1}{\sqrt{3}}$，并且始终有 $0 \le t < \frac{1}{\sqrt{3}}$。由 $t^2(3x-2) = x-2$ 解出 x，得

$$x = \frac{2(1-t^2)}{1-3t^2},$$

其中 $1-3t^2 > 0$。由此，

$$x-1 = \frac{1+t^2}{1-3t^2}, \qquad 3x-2 = \frac{4}{1-3t^2},$$

以及

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{2[-2t(1-3t^2)+6t(1-t^2)]}{(1-3t^2)^2} = \frac{8t}{(1-3t^2)^2}.$$

由于 $\sqrt{x-2} = t\sqrt{3x-2}$，两个平方根可以整齐地合并：

$$(x-2)^{1/2}(3x-2)^{1/2} = t\,(3x-2) = \frac{4t}{1-3t^2}.$$

所以整个分母是

$$(x-1)(x-2)^{1/2}(3x-2)^{1/2} = \frac{4t(1+t^2)}{(1-3t^2)^2},$$

而一切再次与 $\mathrm{d}x/\mathrm{d}t$ 互相抵消：

$$\begin{aligned}\int_2^\infty \frac{\mathrm{d}x}{(x-1)(x-2)^{1/2}(3x-2)^{1/2}} &= \int_0^{1/\sqrt{3}} \frac{(1-3t^2)^2}{4t(1+t^2)} \cdot \frac{8t}{(1-3t^2)^2}\,\mathrm{d}t = \int_0^{1/\sqrt{3}} \frac{2}{1+t^2}\,\mathrm{d}t \\ &= 2\Big[\arctan t\Big]_0^{1/\sqrt{3}} = 2\cdot\frac{\pi}{6} = \frac{\pi}{3},\end{aligned}$$

此即所求。

原积分在两端都是反常积分（improper integral），但它是收敛（converge）的。在 $x=2$ 附近，被积函数的表现类似 $\frac{1}{2}(x-2)^{-1/2}$；当 x 很大时，则类似 $\frac{1}{\sqrt{3}}x^{-2}$。代换把它变成一个连续函数在有限区间上的普通积分，因此无需另外取极限。

验算

- 对 (i) 和 (ii) 的答案求导。就 (i) 而言，$\frac{\mathrm{d}}{\mathrm{d}x}2\sqrt{1-\frac{1}{x}} = \frac{1}{x^2}\left(1-\frac{1}{x}\right)^{-1/2} = x^{-3/2}(x-1)^{-1/2}$。
- 正负号：(ii) 的被积函数是正数，所以原函数（antiderivative）必须递增。事实的确如此，因为当 x 增大时，$\sqrt{\frac{x+1}{x-2}}$ 递减并趋向 1。
- 数值验算：用辛普森法则（Simpson's rule）计算 (i) 中的 $\int_{1.5}^{3}$ 和 (ii) 中的 $\int_{2.5}^{4}$，结果与原函数给出的数值一致（0.47829… 和 0.70974…）。把 (iii) 的积分在 $x=3$ 处分拆后直接计算，得 $1.0471975\ldots = \frac{\pi}{3}$。

常见错误

- 在 (iii) 中忘记更改积分上下限。上限是 $t=\frac{1}{\sqrt{3}}$，而不是 ∞；$\frac{\pi}{6}$ 正是由此而来。
- 把 (i) 或 (ii) 的答案写成 u 的式子，而没有按题目要求换回 x。
- 处理 $1-u$ 的半次幂时出错。要小心合并：$(1-u)^{-3/2}(1-u)^{-1/2} = (1-u)^{-2}$。

STEP 之外

由 $\sqrt{(x-a)(x-b)}$ 构成的被积函数，总可以用类似我们的代换化为有理函数。这属于积分 x 与 $\sqrt{\text{二次式}}$ 的有理函数（rational function）的经典理论（欧拉（Euler）代换）。第 (iii) 部分同时也是一个在无限区间上的反常积分，而用极限来为这类积分赋予意义，是大学一年级分析（analysis）的标准课题。Abbott 的 *Understanding Analysis*（《理解分析》）对反常积分有细致的处理，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）则收录了更多这类 STEP 积分题。

详细解答

STEP II 2020，第 2 题

纯数学 · 可分离变量微分方程

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

这条微分方程（differential equation）可以分离变量（separation of variables）；积分之后，最自然要考察的是乘积 xy，而它原来是 2 的幂。接着利用恒等式（identity）$(x+y)^2-(x-y)^2=4xy$，便可把一切改写成以和 $s=x+y$ 及差 $d=x-y$ 表示。在这两个变量下，每条曲线都变成“d^2 等于某个 s 的函数”，而描绘曲线的关键，就在于找出该函数在哪里非负。

解答

第 (i) 部分

先把右方因式分解：

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{y(kx-1)}{x(1-ky)} \quad\Longrightarrow\quad \frac{1-ky}{y}\,\mathrm{d}y=\frac{kx-1}{x}\,\mathrm{d}x.$$

积分得 $\ln|y|-ky=kx-\ln|x|+c$，即

$$\ln|xy|=k(x+y)+c.$$

在两条曲线上，x 与 y 皆同号，所以 $|xy|=xy$。由于 $k=\ln 2$，取指数得 $xy=A\,2^{x+y}$，其中 A 为正常数。

曲线 C_1。在 $(1,1)$ 处得 $1=4A$，所以 $A=\frac{1}{4}$，且 $4xy=2^{x+y}$。又 $4xy=(x+y)^2-(x-y)^2$，故

$$(x-y)^2=(x+y)^2-2^{x+y},$$

其中 $x,y>0$。这就是所求的结果。

曲线 C_2。在 $(-1,-1)$ 处得 $1 = A \cdot 2^{-2}$，所以 $A = 4$，且 $4xy = 16 \cdot 2^{x+y} = 2^{x+y+4}$。因此

$$(x-y)^2 = (x+y)^2 - 2^{x+y+4},$$

其中 $x, y < 0$。

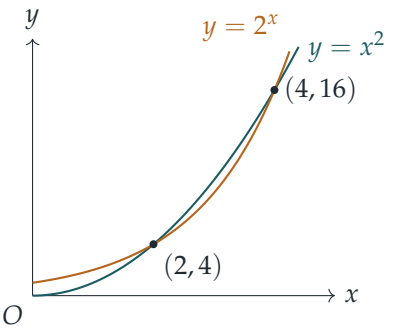
对称性。把 x 与 y 互换，$x+y$ 不变，$(x-y)^2 = (y-x)^2$ 亦不变，所以每条方程都不变。符号条件（$x, y > 0$ 或 $x, y < 0$）同样不变。因此只要 (a,b) 在曲线上，它关于 $y = x$ 的反射点 (b,a) 也在曲线上。故 $y = x$ 是每条曲线的对称轴（line of symmetry）。

第 (ii) 部分

当 $x \geq 0$ 时，曲线 $y = x^2$ 与 $y = 2^x$ 相交于 $(2,4)$ 及 $(4,16)$，而且只有这两个交点。要说明这一点，可比较对数：设 $x > 0$ 时 $g(x) = 2\ln x - x\ln 2$。则 $g'(x) = \frac{2}{x} - \ln 2$，所以 g 在 $x = 2/\ln 2$ 之前递增，之后递减。因此 g 至多有两个零点，而 $g(2) = g(4) = 0$。故

$$x^2 \geq 2^x \iff 2 \leq x \leq 4 \qquad (x > 0),$$

而在 $[0,2)$ 及 $(4,\infty)$ 上，2^x 位于 x^2 之上。



现在取 C_1 上一点，并设 $s = x + y > 0$。C_1 方程的左方是一个平方，所以

$$s^2 - 2^s = (x-y)^2 \geq 0, \qquad \text{即} \qquad s^2 \geq 2^s.$$

由上图可见，这迫使 $2 \leq s \leq 4$。所以 C_1 位于直线 $x + y = 2$ 与 $x + y = 4$ 之间。

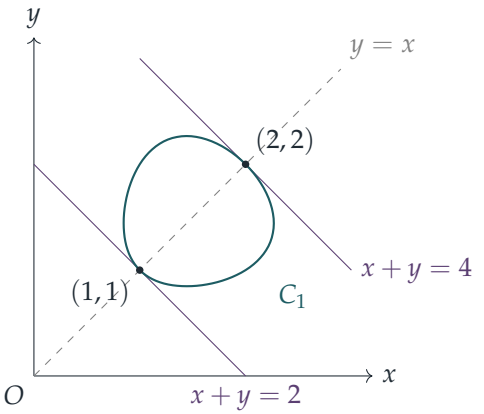
> ? **想一想** $[2,4]$ 中每一个 s 值是否真的都能取到？曲线在两端又是甚么样子？

对每个 $s \in [2,4]$，设 $d = \pm\sqrt{s^2 - 2^s}$，以及 $x = \frac{1}{2}(s+d)$、$y = \frac{1}{2}(s-d)$。由于 $2^s > 0$，有 $|d| < s$，所以 x 与 y 皆为正。因此每个这样的 s 都对应 C_1 上的点，当 $2 < s < 4$ 时更对应两点。所以这条曲线是一个闭环：

- 当 $s = 2$ 及 $s = 4$ 时 $d = 0$，得出 $y = x$ 上的点 $(1,1)$ 及 $(2,2)$。两点都满足 $kxy - y = -(x - kxy)$，所以该处的斜率（gradient）为 -1。闭环在这两点分别与直线 $x + y = 2$ 及 $x + y = 4$ 相切。

- 在 $x = 1/\ln 2 \approx 1.44$ 处斜率为 0，在 $y = 1/\ln 2$ 处斜率为无穷大。在 $x = 1/\ln 2$ 处，闭环上有 $y \approx 0.85$（最低点）及 $y \approx 2.27$（最高点）。
- 它横跨 $y = x$ 的最大宽度出现在 $s^2 - 2^s$ 取最大值之处，即 $s \approx 3.21$，此时 $|d| \approx 1.02$。

这个闭环是一个卵形，关于 $y = x$ 对称，夹在两条直线之间：



第 (iii) 部分

现在 $s = x + y < 0$。设 $t = -s > 0$。与之前一样，我们需要

$$s^2 \geq 2^{s+4} \iff t^2 \geq 2^{4-t} \iff t^2 2^t \geq 16.$$

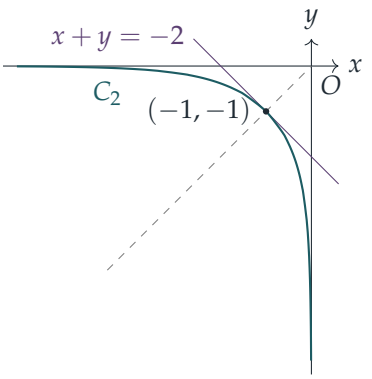
函数 $t^2 2^t$ 在 $t > 0$ 时递增，且在 $t = 2$ 时等于 16。所以条件是 $t \geq 2$，即 $x + y \leq -2$。（等价地，在 $x < 0$ 时描绘 $y = x^2$ 与 $y = 2^{x+4}$，只有当 $x \leq -2$ 时抛物线才在上方。）对每个 $s \leq -2$，用与 (ii) 相同的构造可得两个满足 $x, y < 0$ 的点，唯独 $s = -2$ 只给出单一点 $(-1, -1)$。所以这次 C_2 是无界的：

- 它只在 $(-1, -1)$ 与 $y = x$ 相交，该处斜率同样是 -1，所以它在该点与直线 $x + y = -2$ 相切。
- 当 $x, y < 0$ 时，$kx - 1 < 0$ 且 $1 - ky > 0$，所以

  $$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{y(kx - 1)}{x(1 - ky)} < 0,$$

  因为分子为正而分母为负。所以随着 x 增加，y 持续递减。
- 由 $xy = 4 \cdot 2^{x+y}$：当 $x \to -\infty$ 时，乘积 xy 趋向 0，所以 $y \to 0^-$（事实上 $y \approx 4 \cdot 2^x / x$）。由对称性，当 $x \to 0^-$ 时 $y \to -\infty$。负半轴都是渐近线（asymptote）。

所以 C_2 看起来像第三象限中直角双曲线（rectangular hyperbola）的一支，但它贴近坐标轴的程度紧密得多：



验算

- 对 $4xy = 2^{x+y}$ 作隐函数求导（implicit differentiation）：$4y + 4xy' = k\,2^{x+y}(1+y') = 4kxy(1+y')$。整理后得 $y' = \dfrac{kxy - y}{x - kxy}$，正是原方程。$C_2$ 的情况同样成立。
- $(2,2)$ 在 C_1 上：$4 \cdot 2 \cdot 2 = 16 = 2^4$。在 C_2 上，满足 $x + y = -4$ 的点约为 $(-3.936, -0.064)$，此时 $xy \approx 0.252$，而 $4 \cdot 2^{-4} = 0.25$。上文所引卵形的极端点已用电脑核实。

常见错误

- 积分 $1/x$ 和 $1/y$ 时忘记加上绝对值符号（modulus）。对 C_2 而言这一点很重要，常数应得 $A = 4$，而不是 $\frac{1}{4}$。
- 只证明 $(x+y)^2 \geq 2^{x+y}$ 便停下来。你还必须解释为何这把 $x + y$ 限制在 $[2,4]$ 之内，而这正是 x^2 与 2^x 的比较图派上用场之处。
- 把 C_2 画成穿过坐标轴。在 C_2 上两个坐标都保持为负，而坐标轴是渐近线。

STEP 之外

把坐标转换为 $s = x + y$、$d = x - y$，可将一条对称的方程化为“$d^2 = f(s)$”。曲线恰好存在于 $f(s) \geq 0$ 之处：若该集合是有界区间，曲线便是闭环；若是半直线，曲线便是无界的。同样的推理可以从守恒量（conserved quantity）描绘相图（phase portrait）：$\ln(xy) - k(x+y)$ 沿每条解都是常数，正如 Lotka–Volterra 捕食者—猎物方程（predator–prey equations）中的能量一样；该方程的轨道也正因同样的理由而是闭环。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖了这两个概念。

详细解答

STEP II 2020，第 3 题

纯数学 · 对数凹数列

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

性质 L 表示每个比值 u_{r+1}/u_r 都不大于前一个比值，所以数列一旦停止上升，就不可能再次上升。这正是单峰性（unimodality）。在第 (ii) 部分，递推关系（recurrence）的辅助方程（auxiliary equation）有重根（repeated root）α，而组合 $u_r - \alpha u_{r-1}$ 把它化为一个简单的等比数列（geometric sequence）。其余一切都由这一个代换推出。

解答

第 (i) 部分

假设数列具有性质 L，且 $u_{r-1} \geq u_r$，其中 $2 \leq r \leq n-1$。由于 $u_{r-1} > 0$，我们可以把性质 L 两边除以它：

$$u_{r+1} \leq \frac{u_r^2}{u_{r-1}} = u_r \cdot \frac{u_r}{u_{r-1}} \leq u_r,$$

最后一步用到 $u_r/u_{r-1} \leq 1$ 及 $u_r > 0$。因此 $u_{r-1} \geq u_r \implies u_r \geq u_{r+1}$。

现在考虑任何具有性质 L 的数列。若 $u_1 \leq u_2 \leq \cdots \leq u_n$，则它是单峰的，且 $k = n$。否则，存在第一个指标 k 使得 $u_k > u_{k+1}$（故 $k \leq n-1$）。由 k 的选取，

$$u_1 \leq u_2 \leq \cdots \leq u_k.$$

由于 $u_k \geq u_{k+1}$，刚才证明的蕴涵（取 $r = k+1$）给出 $u_{k+1} \geq u_{k+2}$；再取 $r = k+2$ 应用一次，得 $u_{k+2} \geq u_{k+3}$，如此以归纳法（induction）一直推至 $u_{n-1} \geq u_n$。因此 $u_k \geq u_{k+1} \geq \cdots \geq u_n$，数列是单峰的。

第 (ii) 部分

对 $2 \le r \le n$，记 $v_r = u_r - \alpha u_{r-1}$。对 $3 \le r \le n$，递推关系给出

$$v_r = 2\alpha u_{r-1} - \alpha^2 u_{r-2} - \alpha u_{r-1} = \alpha\left(u_{r-1} - \alpha u_{r-2}\right) = \alpha v_{r-1}.$$

所以 (v_r) 是公比为 α 的等比数列，且 $v_r = \alpha^{r-2} v_2$，即

$$u_r - \alpha u_{r-1} = \alpha^{r-2}(u_2 - \alpha u_1) \qquad (2 \le r \le n).$$

（当 $r = 2$ 时此式显然成立，而 r 每增加一，两边都乘以 α。）

对 $2 \le r \le n-1$，我们有 $3 \le r+1 \le n$，因此可以代入 $u_{r+1} = 2\alpha u_r - \alpha^2 u_{r-1}$：

$$u_r^2 - u_{r-1}u_{r+1} = u_r^2 - 2\alpha u_r u_{r-1} + \alpha^2 u_{r-1}^2 = (u_r - \alpha u_{r-1})^2.$$

正数且单峰。假设 $u_2 > \alpha u_1 > 0$。由于 $\alpha > 0$，这给出 $u_1 > 0$，而 $u_2 > 0$ 亦成立。至于之后的项，第一个恒等式给出

$$u_r = \alpha u_{r-1} + \alpha^{r-2}(u_2 - \alpha u_1) > \alpha u_{r-1},$$

因为 $\alpha^{r-2} > 0$ 及 $u_2 - \alpha u_1 > 0$。所以若 $u_{r-1} > 0$，则 $u_r > 0$；由归纳法，每一项都是正数。现在第二个恒等式给出：对 $2 \le r \le n-1$，$u_{r-1}u_{r+1} = u_r^2 - (u_r - \alpha u_{r-1})^2 \le u_r^2$。这就是正项数列的性质 L，所以由第 (i) 部分，数列是单峰的。

$u_1 = 1$、$u_2 = 2$ 的情况。设 $w_r = (2-r)\alpha^{r-1} + 2(r-1)\alpha^{r-2}$。则 $w_1 = 1 = u_1$ 及 $w_2 = 2 = u_2$。在这些初始值下，第一个恒等式写成

$$u_r = \alpha u_{r-1} + (2-\alpha)\alpha^{r-2}.$$

假设对某个满足 $3 \le r \le n$ 的 r，有 $u_{r-1} = w_{r-1}$。那么

$$\begin{aligned} u_r &= \alpha\left[(3-r)\alpha^{r-2} + 2(r-2)\alpha^{r-3}\right] + 2\alpha^{r-2} - \alpha^{r-1} \\ &= (2-r)\alpha^{r-1} + 2(r-1)\alpha^{r-2} = w_r, \end{aligned}$$

所以由归纳法，对 $1 \le r \le n$，$u_r = (2-r)\alpha^{r-1} + 2(r-1)\alpha^{r-2}$。（我们同样可以用原来的递推关系进行两步归纳；但一阶的形式较快捷。）

$\alpha = 1 - \frac{1}{N}$ 时的最大值。此时 $\frac{1}{2} \le \alpha < 1$，所以 $u_2 = 2 > \alpha = \alpha u_1 > 0$，由上面的推导，数列是正数且单峰的。要找出峰值，可考察相邻项之差。由 $u_{r+1} = \alpha u_r + (2-\alpha)\alpha^{r-1}$ 及 $\alpha - 1 = -\frac{1}{N}$，

$$u_{r+1} - u_r = (2-\alpha)\alpha^{r-1} - \frac{u_r}{N} = \frac{\alpha^{r-2}}{N}\Big[N\alpha(2-\alpha) - (2-r)\alpha - 2(r-1)\Big].$$

由于 $N\alpha = N-1$ 及 $2-\alpha = \frac{N+1}{N}$，所以 $N\alpha(2-\alpha) = N - \frac{1}{N}$，而 $(2-r)\alpha = 2 - r - \frac{2-r}{N}$。方括号化简为

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

因此

$$u_{r+1} - u_r = \frac{\alpha^{r-2}}{N^2}\Big(N^2+1-r(N+1)\Big).$$

想一想 $N^2+1-r(N+1)$ 究竟在哪里变号？它有可能等于零吗？

它只在 $r = \dfrac{N^2+1}{N+1} = N-1+\dfrac{2}{N+1}$ 时为零。由于 $N \geq 2$，我们有 $0 < \frac{2}{N+1} < 1$，所以这个值严格介乎整数 $N-1$ 与 N 之间。由于 $\alpha^{r-2} > 0$：

- 当 $r \leq N-1$ 时，$u_{r+1} > u_r$；
- 当 $r \geq N$ 时，$u_{r+1} < u_r$。

因此 $u_1 < u_2 < \cdots < u_N$ 且 $u_N > u_{N+1} > \cdots > u_n$。由于 $N \leq n$，项 u_N 存在，而 u_r 在 $r = N$ 时取最大值（严格最大，没有并列）。

验算

取 $N = 5$, 则 $\alpha = \frac{4}{5}$。递推关系给出 1, 2, 2.56, 2.816, 2.8672, 2.78528, 2.62144, ..., 在 $r = 5$ 升至峰值，之后下降，与所述相符。差的公式预测 $u_6 - u_5 = \alpha^3(26-30)/25 = -0.08192$，结果吻合。此外，$u_3 = 2\alpha \cdot 2 - \alpha^2 = 2.56$ 与闭合形式 $-\alpha^2 + 4\alpha$ 一致。这些恒等式及峰值位置亦已用电脑以精确算术，就多组 α、u_1、u_2 的值及 N 直至 29 验证过。

常见错误

- 在第 (i) 部分除以 u_{r-1} 时，没有指出它是正数。对于含负项的数列，这个论证并不成立。
- 在第 (ii) 部分引用第 (i) 部分之前，忘记证明各项为正数，而第 (i) 部分需要正性。
- 只证明 $u_{N+1} < u_N$ 及 $u_{N-1} < u_N$。由于单峰性，这已经足够，但你必须说明这一点；较简洁的做法是找出每一个差的正负号。

STEP 之外

性质 L 称为对数凹性（log-concavity）：$\ln u_r$ 是 r 的凹函数（concave function）。对数凹数列在组合数学和概率中随处可见。二项式系数（binomial coefficients）$\binom{n}{r}$ 就是一个例子，而牛顿（Newton）不等式指出，只有实根的多项式，其系数是对数凹的。第 (ii) 部分的递推关系是一个带重根的线性递推关系，这就是为甚么解具有 $(A+Br)\alpha^r$ 的形式，正如微分方程（differential equations）出现重根时会得到 $(A+Bx)\mathrm{e}^{\alpha x}$。关于这种类比在微分方程方面的内容，可参阅 Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）；至于同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

STEP II 2020，第 4 题

纯数学 · 三角不等式

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

一切都建基于三角不等式（triangle inequality）：每条边都严格短于另外两条边之和。事实上，只需检查一条不等式，就是关于最长边的那一条，因为另外两条会自动成立。至于逆命题（converse），可以用圆规把三角形画出来：先画下一条边，再从它的两个端点各画一个圆。

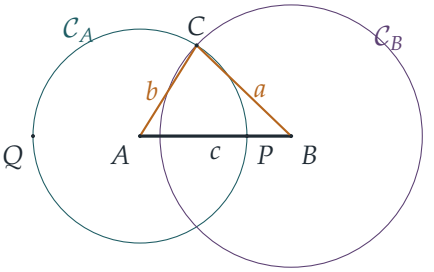
解答

第 (i) 部分

把顶点记为 A、B、C，其中 $BC = a$、$CA = b$、$AB = c$。直线段 AB 是由 A 到 B 的最短路径，所以长度为 $b + a$ 的路径 $A \to C \to B$ 至少是 c。它不可能等于 c：这只会在 C 位于线段 AB 上时发生，而那时根本没有三角形。因此 $c < a + b$。对另外两对顶点用同样的论证，便得到 $a < b + c$ 及 $b < a + c$。

第 (ii) 部分

画一条长度为 c 的线段 AB。以 A 为圆心、b 为半径画圆 $\mathcal{C}_A$，再以 B 为圆心、a 为半径画圆 $\mathcal{C}_B$。若两圆相交于直线 AB 以外的一点 C，则 ABC 是一个三角形，且 $CA = b$、$CB = a$、$AB = c$。



为甚么两圆必定相交？设 P 和 Q 为 $\mathcal{C}_A$ 与直线 AB 的交点，其中 P 与 B 位于 A 的同一侧。那么

$$PB = |c - b|, \qquad QB = c + b.$$

由于 $c - b < a$ 和 $b - c < a$ 正是已知不等式中的两条,所以 $|c - b| < a$;而 $c + b > a$ 就是第三条。因此，当一点沿 $\mathcal{C}_A$ 的上半圆由 P 移动到 Q 时，它与 B 的距离会连续（continuously）地由一个小于 a 的值变成一个大于 a 的值，所以在某处必定等于 a。这一点同时位于两个圆上，而且不在直线 AB 上，因为 $\mathcal{C}_A$ 在该直线上的点只有 P 和 Q，而它们与 B 的距离都不是 a。这就是所求的点 C，所以边长为 a、b、c 的三角形存在。

第 (iii) 部分

第 (i) 和第 (ii) 部分合起来说明：三个正数是某三角形的三边，当且仅当每一个都小于另外两个之和。记三者中最大的为 ℓ。若 ℓ 小于另外两个之和，则另外两条不等式立即成立，因为较小的每个数都不超过 ℓ，而余下的那个数是正数。所以我们只需检验最长的长度。

(A) 必然成立。 三个长度都是正数。利用 $c < a + b$，

$$c + 1 < a + b + 1 < (a + 1) + (b + 1),$$

另外两条不等式同理可得。

(B) 有时成立，但并非必然。 对于等边三角形（equilateral triangle），三个长度是 $1, 1, 1$，确实能组成三角形。但取 $a = b = 1$ 及 $c = \frac{1}{10}$，由于 $1 < 1 + \frac{1}{10}$，这是一个真正的三角形。三个长度变成 1、10、$\frac{1}{10}$，而 $10 > 1 + \frac{1}{10}$，所以不可能组成三角形。

(C) 永不成立。 三个长度关于 a、b、c 对称（symmetric），所以可以适当标记，使 $a \geq b \geq c$。那么

$$|c - a| = a - c = (a - b) + (b - c) = |a - b| + |b - c|.$$

最大的长度等于另外两个之和，所以严格不等式不成立。（若 a, b, c 中有两个相等，则其中一个长度为 0，同样不成立。）

(D) 必然成立。 把 a、b、c 中任意两个互换，集合 $\{a^2+bc,\ b^2+ca,\ c^2+ab\}$ 都不变。(例如互换 a 和 b，得到 b^2+ac、a^2+cb、c^2+ba。) 所以只需对每个三角形证明三条不等式中的一条，例如 $a^2+bc<(b^2+ca)+(c^2+ab)$。两者之差为

$$\begin{aligned}(b^2+ca)+(c^2+ab)-(a^2+bc) &= (b^2-bc+c^2)+a(b+c)-a^2\\ &= \left(b-\tfrac{1}{2}c\right)^2+\tfrac{3}{4}c^2+a(b+c-a).\end{aligned}$$

由于 $c>0$，第一个括号是正数；又由于 $a>0$ 且 $a<b+c$，所以 $a(b+c-a)>0$。因此两者之差为正数。另外两条不等式经重新标记即可得出，而三个长度显然都是正数，所以它们必然能组成三角形。

第 (iv) 部分

a、b、c 的角色可以互换，所以可以适当标记，使 $a\geq b\geq c$。$f(a)$、$f(b)$、$f(c)$ 三者都是正数。由于 f 是递增 (increasing) 函数，$f(a)\geq f(b)\geq f(c)$，所以根据第 (iii) 部分开首的说明，我们只需证明

$$f(a)<f(b)+f(c).$$

? **想一想** 关于 $f(x)/x$ 的条件，怎样帮助我们比较 $f(a)$ 与 $f(b)+f(c)$?

它让我们可以用 $f(a)$ 的倍数，从下方为 $f(b)$ 和 $f(c)$ 定界。由于 $f(x)/x$ 随 x 增大而递减，且 $a\geq b$，

$$\frac{f(b)}{b}\geq\frac{f(a)}{a},\qquad\text{所以}\qquad f(b)\geq\frac{b}{a}f(a),$$

而等号只在 $b=a$ 时才可能成立。同理 $f(c)\geq\frac{c}{a}f(a)$。相加，得

$$f(b)+f(c)\ \geq\ \frac{b+c}{a}f(a)\ >\ f(a),$$

其中最后一步用到 $b+c>a$ (由三角形得出) 及 $f(a)>0$。所以三个新长度中最大的一个小于另外两个之和，因此 $f(a)$、$f(b)$、$f(c)$ 是某三角形的三边。

验算

- 对于 (D)，试取 $a=2$、$b=3$、$c=4$：三个长度是 16、17、22，而 $22<16+17$。把恒等式重新标记以配合最长的长度 c^2+ab，可预测差值为 $(a^2-ab+b^2)+c(a+b-c)=(4-6+9)+4\cdot1=11$，而事实上 $16+17-22=11$。
- 对于 (C)，三角形 $3,4,5$ 给出 $1,1,2$：正如预测，是一个退化成直线的“扁平”三角形。
- 用电脑对 200 000 个随机三角形进行检验，证实了 (A)、(D) 和 (C) (必然、必然、永不)，发现 (B) 对其中约 40% 成立，并证实第 (iv) 部分对 $f(x)=\sqrt{x}$、$\ln(1+x)$、$x/(1+x)$、$x^{0.3}$ 及 $1-\mathrm{e}^{-x}$ 成立。

常见错误

- 只测试了一个三角形，便回答 (B) 是“永不”或“必然”：要回答“有时”，需要每种情况各举一个例子。
- 在 (D) 中只检查了一条不等式，却没有解释为何另外两条也成立：能够支持这一点的，是对称性（或关于最长边的说明）。
- 在 (iv) 中，当 a 可能等于 b 时写出 $f(a)/a < f(b)/b$；此时不等式应为 $\leq$，而严格不等号必须来自 $b + c > a$。

STEP 之外

三角不等式是度量（metric）的三条公理（axioms）之一；度量是距离的抽象概念，是大学一年级分析（analysis）和拓扑学（topology）的核心。第 (iv) 部分在那里是一个经典结果：若 d 是度量，而 f 是递增函数，且 $f(0) = 0$、$f(x)/x$ 递减，则 $f \circ d$ 也是度量；这就是为甚么 $d/(1 + d)$ 是构造有界（bounded）度量的标准方法。Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）是接触这些概念的好起点。

详细解答

STEP II 2020，第 5 题

纯数学 · 数字和与整除性

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

10 的每一个幂除以 9 的余数（remainder）都是 1，所以一个数与它的数字和（digit sum）在模（modulo）9 下必定相同。单凭这一点，每条方程都能化成一个很强的整除（divisibility）条件。接著用大小估计收尾：数字和只以约 $9n$ 的速度增长，而数本身则以约 10^{n-1} 的速度增长，因此只剩下寥寥几个候选数需要检验。

解答

第 (i) 部分

设 $x = \sum_{r=0}^{n-1} a_r 10^r$ 及 $\mathrm{d}(x) = \sum_{r=0}^{n-1} a_r$，则

$$x - \mathrm{d}(x) = \sum_{r=0}^{n-1} a_r \left(10^r - 1\right).$$

每一项都有 $a_r \geq 0$ 及 $10^r - 1 \geq 0$，所以总和非负。此外，$10^r - 1 = 99\ldots9$（r 个 9）$= 9 \times 11\ldots1$，可被 9 整除（当 $r = 0$ 时它等于 0）。所以每一项都可被 9 整除，因此 $x - \mathrm{d}(x)$ 亦可被 9 整除。 ■

用同余（congruence）的语言来说：对每个正整数 x，都有 $x \equiv \mathrm{d}(x) \pmod 9$。下文会反复用到这个结果。

第 (ii) 部分

等价关系。由于 $\mathrm{d}(x) \equiv x \pmod 9$，

$$x - 44\,\mathrm{d}(x) \equiv x - 44x = -43x \equiv 2x \pmod 9,$$

因为 $-43 = -45 + 2$。所以 $x - 44\,\mathrm{d}(x)$ 是 9 的倍数，当且仅当 $9 \mid 2x$。由于 $\gcd(2,9) = 1$，$9 \mid 2x$ 当且仅当 $9 \mid x$。这样便一次过证明了两个方向。 ■

上下界。假设 $x = 44\,\mathrm{d}(x)$，且 x 有 n 位数字。每个数字最多是 9，所以 $\mathrm{d}(x) \le 9n$，而 $x = 44\,\mathrm{d}(x) \le 396n$。另外，一个 n 位数的首位数字 $a_{n-1} \ge 1$，所以 $x \ge 10^{n-1}$。合起来得

$$10^{n-1} \le x \le 396n.$$

当 $n = 5$ 时，这需要 $10\,000 \le 1980$，并不成立。此外，若对某个 $n \ge 5$ 有 $10^{n-1} > 396n$，则 $10^n = 10 \cdot 10^{n-1} > 3960n \ge 396(n+1)$，所以由归纳法（induction）可知，不等式 $10^{n-1} \le 396n$ 对每个 $n \ge 5$ 都不成立。因此 $n \le 4$。 ■

> ? **想一想** 怎样才能把搜索范围由 9999 以内的所有数，缩减到只剩几个?

利用刚刚证明的等价关系。若 $x = 44\,\mathrm{d}(x)$，则 $x - 44\,\mathrm{d}(x) = 0$，这是 9 的倍数，所以 $9 \mid x$。于是 $9 \mid 44\,\mathrm{d}(x)$，而由于 $\gcd(44,9) = 1$，这迫使 $9 \mid \mathrm{d}(x)$。因此 $x = 44\,\mathrm{d}(x)$ 是 $44 \times 9 = 396$ 的倍数。

求解。由于 $n \le 4$，我们有 $x \le 396 \times 4 = 1584$，所以 x 是 396, 792, 1188, 1584 其中之一。它们的数字和全都是 18，所以在每种情况下 $44\,\mathrm{d}(x) = 792$。唯一符合的是

$$x = 792, \qquad \mathrm{d}(792) = 18, \quad 44 \times 18 = 792.$$

每一个解都在这四个候选数之中，而只有 792 符合，所以再没有其他值。 ■

第 (iii) 部分

记 $y = \mathrm{d}(x)$ 及 $z = \mathrm{d}(y)$，于是方程为 $x = 107z$，且 $z = x/107 > 0$。

整除性。由第 (i) 部分，$x \equiv y \equiv z \pmod 9$。所以

$$0 = x - 107z \equiv z - 107z = -106z \pmod 9.$$

由于 $106 = 99 + 7$，所以 $9 \mid 7z$；又因 $\gcd(7,9) = 1$，得 $9 \mid z$。

大小。第 (i) 部分亦给出：对任何正整数 m，$\mathrm{d}(m) \le m$。所以若 x 有 n 位数字，则 $z = \mathrm{d}(y) \le y \le 9n$，以及

$$10^{n-1} \le x = 107z \le 963n.$$

当 $n = 5$ 时，这需要 $10\,000 \le 4815$，并不成立。用与第 (ii) 部分相同的归纳法（现在是 $10^n > 9630n \ge 963(n+1)$）可证明它对所有 $n \ge 5$ 都不成立，所以 $n \le 4$。

> ? **想一想** 上界 $z \le 36$ 仍然留下好几个 9 的倍数。可以做得更好吗?

可以，方法是再取一次数字和。由于 $n \le 4$，$y = \mathrm{d}(x) \le 36$。在 $1, 2, \ldots, 36$ 之中，最大的数字和是 $\mathrm{d}(29) = 11$，因为小于 37 的两位数十位数字最多是 3，而十位数字为 3 时

个位数字最多是 6（数字和最多是 9），至于十位数字为 2 时则可以有 $2+9=11$。所以 $1 \le z \le 11$，而 z 是 9 的倍数。因此 $z=9$，以及

$$x = 107 \times 9 = 963.$$

验证：$\mathrm{d}(963) = 18$，$\mathrm{d}(18) = 9$，而 $107 \times 9 = 963$。所以 $x = 963$ 满足方程。任何解都必须有 $z = 9$，因此 $x = 963$，所以再没有其他值。 ■

验算

- 抽样验算第 (i) 部分：$249 - 15 = 234 = 9 \times 26$。
- 用电脑搜索所有 $x < 10^6$，发现 792 是 $x = 44\,\mathrm{d}(x)$ 的唯一解，而 963 是 $x = 107\,\mathrm{d}(\mathrm{d}(x))$ 的唯一解，与上述的上下界一致。
- 留意 $792 = 44 \times 18$ 及 $963 = 107 \times 9$：两者都是 9 的倍数，正如整除性论证所要求的。

常见错误

- 在第 (ii) 部分只得出 x 是 9 的倍数。更强的事实是 $\mathrm{d}(x)$ 也是 9 的倍数，因而 $396 \mid x$，正是这一点令搜索变得简短。
- 只检查了 $n = 5$ 便断言 $n \le 4$。你需要说明为何对所有更大的 n，10^{n-1} 都一直大于 $396n$；归纳法（或者观察到左边每次增大 10 倍，而右边最多只增大 $\frac{6}{5}$ 倍）正好提供了理由。
- 在第 (iii) 部分忘记了 $z > 0$，而正是这一点排除了 $z = 0$。

STEP 之外

$x \equiv \mathrm{d}(x) \pmod 9$ 这个事实，正是“弃九法”（casting out nines）背后的原理，那是一种古老的验算方法。这是在环（ring）$\mathbb{Z}/9\mathbb{Z}$ 中运算的第一个例子，在这个环里 10 的表现就像 1；同样的想法配合 $10 \equiv -1$，便得出判断能否被 11 整除的交错和检验法。把一个数与其各位数字的函数混在一起的方程，通常都正如这里般求解：用同余缩减可能性，再用增长估计令候选名单变得有限。Davenport 的 *The Higher Arithmetic*（《高等算术》）是一本很好的同余入门书。

详细解答

STEP II 2020，第 6 题

纯数学 · 矩阵的幂与迹

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

本题在 *Beyond STEP* 第 7.4 节回到起点的矩阵中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch07/finite-order.html

关键思路

每个实 2×2 矩阵（matrix）都满足它自己的特征方程（characteristic equation）：

$$\mathbf{M}^2 - t\,\mathbf{M} + \delta\,\mathbf{I} = \mathbf{0}, \qquad t = \operatorname{tr}(\mathbf{M}),\ \delta = \det(\mathbf{M}),$$

这就是凯莱–哈密顿定理（Cayley–Hamilton theorem）（把乘积展开，两行便可验证）。一旦把 $\mathbf{M}^2$ 写成 $\mathbf{M}$ 与 $\mathbf{I}$ 的组合，任何像 $\mathbf{M}^2 = \mathbf{I}$ 这样的方程都会变成“$t\mathbf{M}$ 是 $\mathbf{I}$ 的倍数”，这就迫使 $t = 0$，或者 $\mathbf{M}$ 本身是 $\mathbf{I}$ 的倍数。之后，第 (iii) 及 (iv) 部分只需把第 (ii) 部分应用到 $\mathbf{M}^2$ 和 $\mathbf{M}^4$（而非 $\mathbf{M}$）上便可得出。

全文记 $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$、$t = a + d$ 及 $\delta = ad - bc$。

解答

第 (i) 部分

把乘积展开，

$$\mathbf{M}^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & bc + d^2 \end{pmatrix} = (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} - (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

因为例如 $(a+d)a-(ad-bc)=a^2+bc$ 及 $(a+d)b=ab+bd$。所以

$$\mathbf{M}^2=t\mathbf{M}-\delta\mathbf{I}. \qquad (*)$$

对 $(*)$ 取迹（trace），并利用 $\operatorname{tr}(\mathbf{I})=2$，得

$$\operatorname{tr}(\mathbf{M}^2)=t\cdot t-2\delta=\operatorname{tr}(\mathbf{M})^2-2\det(\mathbf{M}),$$

这正是所求。（直接计算亦可：$a^2+2bc+d^2=(a+d)^2-2(ad-bc)$。）

第 (ii) 部分

第一个等价（equivalence），$(\Rightarrow)$。设 $\mathbf{M}^2=\mathbf{I}$ 且 $\mathbf{M}\neq\pm\mathbf{I}$。代入 $(*)$ 得

$$t\,\mathbf{M}=(1+\delta)\,\mathbf{I}.$$

若 $t\neq 0$，便可两边除以 t，得 $\mathbf{M}=k\mathbf{I}$，其中 $k=(1+\delta)/t$ 为实数；于是 $\mathbf{M}^2=k^2\mathbf{I}=\mathbf{I}$ 给出 $k=\pm1$，即 $\mathbf{M}=\pm\mathbf{I}$，但这已被排除。因此 $t=0$，从而 $(1+\delta)\mathbf{I}=\mathbf{0}$，所以 $\delta=-1$。

$(\Leftarrow)$。若 $t=0$ 且 $\delta=-1$，则 $(*)$ 给出 $\mathbf{M}^2=0\cdot\mathbf{M}+\mathbf{I}=\mathbf{I}$。此外 $\operatorname{tr}(\pm\mathbf{I})=\pm2\neq0$，所以 $\mathbf{M}\neq\pm\mathbf{I}$。

第二个等价，$(\Rightarrow)$。若 $\mathbf{M}^2=-\mathbf{I}$，则 $(*)$ 给出 $t\mathbf{M}=(\delta-1)\mathbf{I}$。若 $t\neq0$，则 $\mathbf{M}=k\mathbf{I}$，其中 k 为实数，而 $\mathbf{M}^2=k^2\mathbf{I}=-\mathbf{I}$ 便需要 $k^2=-1$，这对实数 k 是不可能的。所以 $t=0$，从而 $\delta=1$。

$(\Leftarrow)$。若 $t=0$ 且 $\delta=1$，则 $(*)$ 立即给出 $\mathbf{M}^2=-\mathbf{I}$。

? **想一想** 我们究竟在哪里用到 $\mathbf{M}$ 是实矩阵这一点?

只在第二个等价中用到：$k^2=-1$ 没有实数解。对于复矩阵，$\mathbf{M}=\mathrm{i}\mathbf{I}$ 满足 $\mathbf{M}^2=-\mathbf{I}$，但其迹为 $2\mathrm{i}\neq0$，所以结论便不成立。

第 (iii) 部分

$(\Leftarrow)$ 是显然的：若 $\mathbf{M}^2=\pm\mathbf{I}$，则 $\mathbf{M}^4=(\pm\mathbf{I})^2=\mathbf{I}$。

$(\Rightarrow)$。设 $\mathbf{N}=\mathbf{M}^2$，这是一个实 2×2 矩阵，且 $\mathbf{N}^2=\mathbf{I}$。要么 $\mathbf{N}=\pm\mathbf{I}$，这正是我们想要的；要么 $\mathbf{N}\neq\pm\mathbf{I}$，此时第 (ii) 部分告诉我们 $\det(\mathbf{N})=-1$。但

$$\det(\mathbf{N})=\det(\mathbf{M})^2\geq0,$$

所以第二种情况不可能出现。因此 $\mathbf{M}^2=\pm\mathbf{I}$。

$\mathbf{M}^4=-\mathbf{I}$ 的条件。同样设 $\mathbf{N}=\mathbf{M}^2$。把第 (ii) 部分的第二个等价应用到 $\mathbf{N}$ 上，

$$\mathbf{M}^4=-\mathbf{I}\iff\operatorname{tr}(\mathbf{N})=0\text{ 且 }\det(\mathbf{N})=1.$$

利用第 (i) 部分及 $\det(\mathbf{N}) = \delta^2$，这等于说

$$t^2 - 2\delta = 0 \quad 及 \quad \delta^2 = 1.$$

由 $\delta^2 = 1$ 得 $\delta = \pm 1$，而 $t^2 = 2\delta \geq 0$ 排除了 $\delta = -1$。所以 $\delta = 1$ 且 $t^2 = 2$。每一步都是等价的，因此

$$\boxed{\mathbf{M}^4 = -\mathbf{I} \iff \det(\mathbf{M}) = 1 \text{ 且 } \operatorname{tr}(\mathbf{M}) = \pm\sqrt{2}}.$$

第 (iv) 部分

把第 (iii) 部分应用到 $\mathbf{M}^2$ 上：$\mathbf{M}^8 = \mathbf{I} \iff \mathbf{M}^4 = \pm\mathbf{I}$，所以满足 $\mathbf{M}^8 = \mathbf{I}$ 的矩阵恰好是满足以下条件的矩阵：

$$\mathbf{M}^2 = \pm\mathbf{I} \qquad 或 \qquad \det(\mathbf{M}) = 1,\ \operatorname{tr}(\mathbf{M}) = \pm\sqrt{2}.$$

旋转（rotation）和反射（reflection）就是正交（orthogonal）矩阵：它们的列是互相垂直的单位向量（unit vector）（旋转的行列式（determinant）为 1，反射的行列式为 -1）。所以我们只需在上述矩阵中找一个列不是单位向量的矩阵。例如

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

的迹为 0、行列式为 -1，所以由第 (ii) 部分得 $\mathbf{M}^2 = \mathbf{I}$（验证：$\mathbf{M}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$），因而 $\mathbf{M}^8 = \mathbf{I}$。它的第二列 $(1, -1)$ 的长度为 $\sqrt{2} \neq 1$，所以它既不是旋转，也不是反射。（它是一个“斜反射”：它保持沿 x 轴的每个向量不变，并把沿 $(1, -2)$ 方向的每个向量反向，而这两个方向并不互相垂直。）

较有趣的那一类中的一个例子是 $\mathbf{M} = \begin{pmatrix} \sqrt{2} & 1 \\ -1 & 0 \end{pmatrix}$：行列式为 1、迹为 $\sqrt{2}$，所以 $\mathbf{M}^4 = -\mathbf{I}$ 及 $\mathbf{M}^8 = \mathbf{I}$，但它的第一列长度为 $\sqrt{3}$。它是一个“变形了的”45° 旋转：它的阶（order）恰好是 8，与该旋转相同，但不保持长度。

验算

- 转角为 45° 的旋转 $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ 的行列式为 1、迹为 $\sqrt{2}$，而转四次 45° 便是转半圈，即 $-\mathbf{I}$：与第 (iii) 部分一致。
- 把 $\mathbf{M} = \begin{pmatrix} \sqrt{2} & 1 \\ -1 & 0 \end{pmatrix}$ 平方，得 $\begin{pmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & -1 \end{pmatrix}$，其迹为 0、行列式为 1，所以由第 (ii) 部分，它的平方是 $-\mathbf{I}$。

- 以电脑搜索所有元素属于 $\{-3, -\frac{5}{2}, \ldots, 3\}$ 的矩阵，确切验证了第 (ii) 部分的两个等价，以及第 (iii) 部分的第一个等价。

常见错误

- 未先处理 $t = 0$ 的情况便除以 t；第 (ii) 部分的整个论证，正是按 t 是否为零来分情况讨论。
- 在第 (iii) 部分，忘记 $\det(\mathbf{M}^2) = \det(\mathbf{M})^2$ 不可能是负数；正是这一点排除了 $\mathbf{I}$ 的“其他”平方根（square root）。
- 在第 (iv) 部分以 45° 旋转作为例子：它虽然满足 $\mathbf{M}^8 = \mathbf{I}$，却正正是题目所排除的情况。

STEP 之外

恒等式 $(*)$ 是凯莱–哈密顿定理在 2×2 情况下的特例：每个方阵都满足它的特征多项式（characteristic polynomial）。像“哪些 $\mathbf{M}$ 满足 $\mathbf{M}^n = \mathbf{I}$?”这类问题，实质上是关于特征值（eigenvalue）的问题：在这里，特征值必须是八次单位根（roots of unity），而一个阶为 8 的实矩阵与某个旋转相似（similar），但本身不一定是旋转。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）正是本着这种精神，阐述特征值、相似及特征多项式。

详细解答

STEP II 2020，第 7 题

纯数学 · 复数变换

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

映射 $z \mapsto w = \dfrac{2}{z-2}$ 是先平移、再作反演（inversion），而反演会把不经过原点的直线映成经过原点的圆。要证明 w 位于某个特定的圆上，最干净利落的做法是先猜出圆心 c，再证明 $|w-c|$ 可化简为常数。把 $w-c$ 写成单一分式后，会发现分子与分母互为共轭（conjugate）（或几乎如此），正因如此，模（modulus）才是常数。

解答

以下全程使用 $|\alpha/\beta| = |\alpha|/|\beta|$，以及对实数 x、y 有 $|x+yi| = |x-yi| = |y+xi| = \sqrt{x^2+y^2}$。

第 (i) 部分

设 $z = 3+ti$，则 $z-2 = 1+ti$，所以

$$w - 1 = \frac{2}{1+ti} - 1 = \frac{2-(1+ti)}{1+ti} = \frac{1-ti}{1+ti}.$$

分子与分母互为共轭复数，因此模相同，都是 $\sqrt{1+t^2}$，而且永不为零。故此

$$|w-1| = \frac{\sqrt{1+t^2}}{\sqrt{1+t^2}} = 1,$$

与 t 无关。直线 $\mathrm{Re}(z) = 3$ 上每一个 z 都可写成 $3+ti$（t 为某实数），所以对每个这样的 z，点 w 与 1 的距离都是 1：它位于圆心为 1、半径为 1 的圆上。

直线 V。现在 $z = p+ti$，不妨记 $a = p-2$；由于 $p \neq 2$，a 是非零实数。于是 $z-2 = a+ti$，

永不为零。当 $p=3$ 时圆心是 $1=1/a$，所以一般情况下我们尝试以 $1/a$ 为圆心：

$$w-\frac{1}{a}=\frac{2}{a+ti}-\frac{1}{a}=\frac{2a-(a+ti)}{a(a+ti)}=\frac{a-ti}{a(a+ti)}.$$

同样地 $|a-ti|=|a+ti|$，所以

$$\left|w-\frac{1}{p-2}\right|=\frac{1}{|a|}=\frac{1}{|p-2|}.$$

因此 w 位于以下的圆上：

$$\text{圆心}\frac{1}{p-2},\qquad \text{半径}\frac{1}{|p-2|}.$$

半径中的绝对值符号不可省略：若 $p<2$，圆心位于负实轴上，但半径必须是正数。

想一想 如果题目没有先给出 $p=3$ 的情况，我们会怎样找出圆心 $1/(p-2)$?

点 $z=p$（即 $t=0$）给出 $w=2/a$；而当 $t\to\pm\infty$ 时，$w\to 0$。这两点是一条直径的两端（圆关于实轴对称，而它们正是圆上仅有的两个实数点），所以圆心是它们的中点 $1/a$。

至于虚部，把分子分母同乘以共轭：

$$w=\frac{2(a-ti)}{a^2+t^2},\qquad \text{所以}\qquad \operatorname{Im}(w)=\frac{-2t}{a^2+t^2}.$$

分母为正，所以 $\operatorname{Im}(w)>0$ 当且仅当 $t<0$。因此 V 上使 $\operatorname{Im}(w)>0$ 的点是

$$z=p+ti,\ \text{其中}t<0,\quad \text{即}V\text{上位于实轴下方的点}(\operatorname{Im}(z)<0).$$

第 (ii) 部分

现在 $z=s+qi$，其中 s 为实数。记 $u=s-2$，则 $z-2=u+qi$；由于 $q\neq 0$，它不为零。这里圆的“两端”是 $w=0$（当 $u\to\pm\infty$）及 $w=2/(qi)=-2i/q$（当 $u=0$），由此推测圆心为 $-i/q$。验证：

$$w+\frac{i}{q}=\frac{2}{u+qi}+\frac{i}{q}=\frac{2q+i(u+qi)}{q(u+qi)}=\frac{q+ui}{q(u+qi)}.$$

而 $|q+ui|=\sqrt{q^2+u^2}=|u+qi|$，所以

$$\left|w+\frac{i}{q}\right|=\frac{1}{|q|}.$$

因此 w 位于以下的圆上：

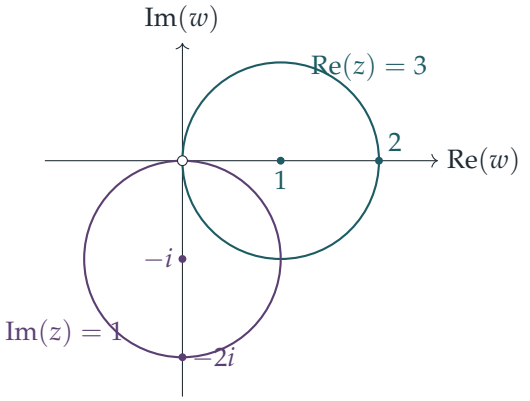
$$\text{圆心}-\frac{i}{q},\qquad \text{半径}\frac{1}{|q|}.$$

至于实部，

$$w=\frac{2(u-qi)}{u^2+q^2},\qquad \text{所以}\qquad \operatorname{Re}(w)=\frac{2u}{u^2+q^2},$$

分母为正。所以 $\mathrm{Re}(w) > 0$ 当且仅当 $u > 0$，即

$$z = s + qi，\text{其中} s > 2，\quad \text{即} H \text{ 中位于下述直线右方的点：}\ \mathrm{Re}(z) = 2.$$



上图显示 $\mathrm{Re}(z) = 3$ 的像（来自第 (i) 部分）及 $\mathrm{Im}(z) = 1$ 的像（第 (ii) 部分取 $q = 1$）。两个圆都经过原点，但 $w = 0$ 本身永远取不到，因为 $2/(z-2) \neq 0$；它对应于沿直线走到“无穷远处”的 z。

验算

- 在 V 的公式中取 $p = 3$：圆心 $1/(3-2) = 1$，半径 1，与第一个结果吻合。
- 代入数值：$z = 3 - i$ 给出 $w = 2/(1-i) = 1 + i$，有 $|w - 1| = 1$ 及 $\mathrm{Im}(w) = 1 > 0$，与 $t = -1 < 0$ 一致。对于 $q = 1$ 的 H，$z = 3 + i$ 给出 $w = 1 - i$，它与 $-i$ 的距离为 1，且实部为正，与 $3 > 2$ 相符。
- 当 $p \to 2$ 时，半径 $1/|p-2|$ 变大：此时直线 V 非常接近被映到无穷远的点 $z = 2$，所以它的像理应很大。用电脑对数以千计的随机 p、q 及每条直线上的点作检查，证实了两个圆及两个符号条件均正确。

常见错误

- 把半径写成 $1/(p-2)$ 或 $1/q$，而当 $p < 2$ 或 $q < 0$ 时它们是负数。
- 把 $\mathrm{Im}(w)$ 的符号弄反：除以 $z - 2$ 相当于关于实轴作反射，所以 $\mathrm{Im}(w)$ 与 $\mathrm{Im}(z)$ 的符号相反。

STEP 之外

映射 $z \mapsto 2/(z-2)$ 是一个默比乌斯变换（Möbius transformation），即形如 $z \mapsto (az+b)/(cz+d)$ 的映射。这类映射总是把“广义圆”（generalised circles，即圆和直线）映成广义圆；从共形映射（conformal mapping）到双曲几何（hyperbolic geometry），它们都是复分析（complex analysis）入门课程中最基本的几何工具。Needham 的 *Visual Complex Analysis*（《直观复分析》）有一章精彩地介绍它们，书中的图像令类似本题的结果一目了然。

详细解答

STEP II 2020，第 8 题

纯数学 · 多项式与积分

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

由于 $\mathrm{F}' = \mathrm{f}$，f 的每个根都是 F 的转向点（turning point），而每个面积条件都变成关于 F 在其中一个转向点的值的陈述。一旦知道 F 在 0 和 c 有二重根（double root），和式 $\mathrm{F}(x) + \mathrm{F}(c - x)$ 便化简为只有一个决定正负号的因子 $c - 2h$。把这个正负号与我们对 F 在 $[0, c]$ 上的最大值和最小值的认识作比较，便迫使图像具有对称性。

解答

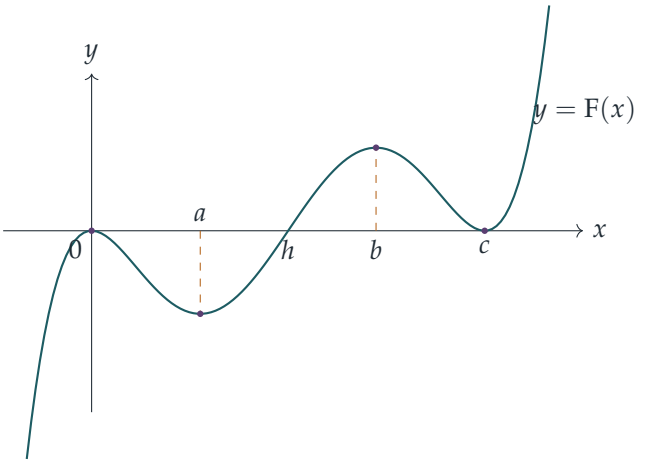
由于 f 是首一（monic）多项式，其根为 $0, a, b, c$，故 $\mathrm{f}(x) = x(x-a)(x-b)(x-c)$。因此在 $(0, a)$ 上 $\mathrm{f} < 0$，在 (a, b) 上 $\mathrm{f} > 0$，在 (b, c) 上 $\mathrm{f} < 0$，而当 $x < 0$ 及 $x > c$ 时 $\mathrm{f} > 0$。设 A 为 0 与 a 之间的面积。题设条件指出 b 与 c 之间的面积为 A，a 与 b 之间的面积为 $2A$，其中 $A > 0$。

第 (i) 部分

由于 $\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t$，在 $\mathrm{f} < 0$ 的区间上的积分等于该处面积的相反数。因此

$$\mathrm{F}(0) = 0, \qquad \mathrm{F}(a) = -A, \qquad \mathrm{F}(b) = -A + 2A = A, \qquad \mathrm{F}(c) = A - A = 0.$$

由于 $\mathrm{F}' = \mathrm{f}$，F 的转向点位于 $x = 0, a, b, c$。f 的正负号在 0 和 b 由 $+$ 变为 $-$，所以该处为极大点；在 a 和 c 由 $-$ 变为 $+$，所以该处为极小点。F 是五次多项式，其首项系数（leading coefficient）$\frac{1}{5}$ 为正，故当 $x \to \pm\infty$ 时 $\mathrm{F}(x) \to \pm\infty$。



曲线在 $(0,0)$ 有极大点，在 $x=a$ 有极小点且 $y=-A$，在 $x=b$ 有极大点且 $y=A$，并在 $(c,0)$ 有一个与 x 轴相切的极小点。

F 的形式。F 是首项系数为 $\frac{1}{5}$ 的 5 次多项式。在 $x=0$，有 $\mathrm{F}(0)=0$ 及 $\mathrm{F}'(0)=\mathrm{f}(0)=0$，所以 $x=0$ 是重数（multiplicity）至少为 2 的根。同理 $\mathrm{F}(c)=0=\mathrm{F}'(c)$，所以 c 是重数至少为 2 的根。这样便确定了五个一次因子中的四个：

$$\mathrm{F}(x)=\tfrac{1}{5}x^2(x-c)^2(x-h)$$

其中 h 为某实数。此外，$\mathrm{F}(a)=-A<0<A=\mathrm{F}(b)$，所以 F 在 a 与 b 之间（不含端点）有一个根。0 和 c 都不在 (a,b) 之内，所以这个根必定是 h。因此 $a<h<b$，特别地 $0<h<c$。

和式。注意 $(c-x)^2\big((c-x)-c\big)^2=x^2(x-c)^2$。因此

$$\begin{aligned}\mathrm{F}(x)+\mathrm{F}(c-x)&=\tfrac{1}{5}x^2(x-c)^2\big[(x-h)+(c-x-h)\big]\\&=\tfrac{1}{5}x^2(x-c)^2(c-2h).\end{aligned}$$

第 (ii) 部分

在 $[0,c]$ 上，F 在 $[0,a]$ 上严格递减，在 $[a,b]$ 上严格递增，在 $[b,c]$ 上严格递减。所以在 $[0,c]$ 上的最小值是 $\mathrm{F}(a)=-A$：唯一的其他候选值是端点值 $\mathrm{F}(c)=0>-A$。这个最小值只在 $x=a$ 取得。所以当 $0\le x\le c$ 时，

$$\mathrm{F}(b)+\mathrm{F}(x)\ \ge\ \mathrm{F}(b)+\mathrm{F}(a)=A-A=0,$$

而当 $x\ne a$ 时不等式严格成立。

想一想 取哪个 x 值，可以使 $\mathrm{F}(b)+\mathrm{F}(x)$ 看起来像第 (i) 部分的和式?

取 $x=c-b$，由于 $0<b<c$，它位于 $[0,c]$ 之内。由第 (i) 部分，

$$\mathrm{F}(b)+\mathrm{F}(c-b)=\tfrac{1}{5}b^2(b-c)^2(c-2h).$$

若 $c-b\neq a$，左方严格为正。由于 $\frac{1}{5}b^2(b-c)^2>0$，可推出 $c-2h>0$。所以 $c-b=a$ 或 $c>2h$，此即所求。

现在把同一论证倒过来用。在 $[0,c]$ 上，F 的最大值是 $\mathrm{F}(b)=A$，大于端点值 $\mathrm{F}(0)=\mathrm{F}(c)=0$，而且只在 $x=b$ 取得。所以当 $0\le x\le c$ 时，

$$\mathrm{F}(a)+\mathrm{F}(x)\le \mathrm{F}(a)+\mathrm{F}(b)=0,$$

而当 $x\neq b$ 时不等式严格成立。取 $x=c-a\in[0,c]$，得 $\frac{1}{5}a^2(c-a)^2(c-2h)\le 0$，若 $c-a\neq b$ 则严格成立。所以 $c-a=b$ 或 $c<2h$。

若 $c\neq a+b$，两个结论分别给出 $c>2h$ 和 $c<2h$，这是不可能的。因此 $c=a+b$。于是 $c-b=a$，所以

$$0=\mathrm{F}(b)+\mathrm{F}(a)=\mathrm{F}(b)+\mathrm{F}(c-b)=\tfrac{1}{5}b^2a^2(c-2h).$$

由于 $a,b\neq 0$，得 $c=2h$。

第 (iii) 部分

当 $h=\frac{1}{2}c$ 时，有 $\mathrm{F}(x)=\frac{1}{5}u\,(x-\frac{1}{2}c)$，其中 $u=x^2(x-c)^2$，因此 $u'=2x(x-c)(2x-c)$。用乘积法则（product rule）求导，

$$\begin{aligned}\mathrm{f}(x)=\mathrm{F}'(x)&=\tfrac{1}{5}\left[2x(x-c)(2x-c)\left(x-\tfrac{1}{2}c\right)+x^2(x-c)^2\right]\\&=\tfrac{1}{5}x(x-c)\left[(2x-c)^2+x(x-c)\right]\\&=\tfrac{1}{5}x(x-c)\left(5x^2-5cx+c^2\right)\\&=x(x-c)\left(x^2-cx+\tfrac{1}{5}c^2\right).\end{aligned}$$

（所以 $a,b=\frac{1}{2}c\mp\frac{c}{2\sqrt{5}}$，两者之和确实是 c。）

想一想 各根关于 $x=\frac{1}{2}c$ 对称。能否利用这一点令二阶导数（second derivative）变得容易计算?

可以。设 $y=x-\frac{1}{2}c$。则 $x(x-c)=y^2-\frac{1}{4}c^2$ 及 $x^2-cx+\frac{1}{5}c^2=y^2-\frac{1}{20}c^2$，所以

$$\mathrm{f}=\left(y^2-\tfrac{1}{4}c^2\right)\left(y^2-\tfrac{1}{20}c^2\right)=y^4-\tfrac{3}{10}c^2y^2+\tfrac{1}{80}c^4.$$

由于 $\mathrm{d}y/\mathrm{d}x = 1$，

$$\mathrm{f}''(x) = 12y^2 - \tfrac{3}{5}c^2,$$

它恰好在 $y^2 = \frac{1}{20}c^2$ 时为零。在这些点，f 的因子 $y^2 - \frac{1}{20}c^2$ 等于零，所以 $\mathrm{f} = 0$。每一点都是 f″（作为 y 的二次式）的单根（simple root），所以 f″ 在该处变号，它们确实是拐点（point of inflection）。所以两个拐点都位于 x 轴上。事实上，它们就是点 $(a, 0)$ 和 $(b, 0)$。

验算

- 取 $c = 2$，则 $a = 1 - \frac{1}{\sqrt{5}} \approx 0.5528$ 及 $b \approx 1.4472$。对 f 作数值积分，得 $[0, a]$、$[a, b]$ 和 $[b, c]$ 上的面积分别为 0.05724、0.11449 和 0.05724：外侧两个相等，而且各是中间那个的一半。
- 以 $a + b = c$ 及 $ab = \frac{1}{5}c^2$ 展开 $x(x-a)(x-b)(x-c)$，可重新得到第 (iii) 部分的式子。
- 取 $c = 2$，用有限差分（finite difference）估计 f″ 在 $x = a \approx 0.5528$ 的值，在舍入误差范围内为零，而且它在 $x = 0.54$ 与 $x = 0.56$ 之间变号。

常见错误

- 只记得 $\mathrm{F}(0) = \mathrm{F}(c) = 0$，忘记同时有 $\mathrm{F}'(0) = \mathrm{F}'(c) = 0$，因而遗漏了二重根。
- 没有证明 $0 < h < c$：决定 h 位置的，正是 F 在 a 与 b 之间的变号。
- 在第 (ii) 部分只用了第一个不等式。它只给出“$c = a + b$ 或 $c > 2h$”，单靠这一点并不足够。还需要用 $\mathrm{F}(a)$ 得出的反向不等式，才能完成论证。

STEP 之外

第 (ii) 部分的论证用到极值定理（extreme value theorem）：闭区间上的连续函数必定取得最大值和最小值，而在这里我们还确切知道它们在哪里取得。面积条件迫使这个四次多项式关于 $x = \frac{1}{2}c$ 对称，这说明了积分条件可以如何确定一个多项式的形状。Abbott 的 *Understanding Analysis*（《理解分析》）仔细论述了极值定理，以及联系 F 与 f 的微积分基本定理（fundamental theorem of calculus）。

详细解答

STEP II 2020，第 9 题

力学 · 相撞的抛体

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

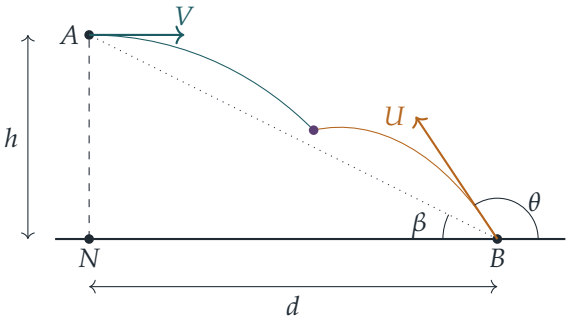
www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

把两粒质点放进同一个坐标系（coordinate system），写出各自在时刻 t 的位置。两者都受相同的向下加速度（acceleration）g，所以比较高度时，$-\frac{1}{2}gt^2$ 项会互相抵消：相对于对方而言，两粒质点都以恒定速度沿直线运动。这令碰撞时间变得非常简单，而之后每一部分都只是它的简短推论。

解答

把两粒质点都模型化为只受重力作用、自由运动的质点（point mass），不计空气阻力（air resistance）。取 N 为原点，x 轴沿 NB（故 $B=(d,0)$，$A=(0,h)$），y 轴铅垂向上。从 A 出发的质点向 B 运动，从 B 出发的质点向 N 运动，所以两者的水平速度方向相反；这正是题目所给结果所描述的设置。



抛出后时刻 t，从 A 出发的质点位于

$$x_A = Vt, \qquad y_A = h - \tfrac{1}{2}gt^2,$$

而从 B 出发的质点位于

$$x_B = d - Ut\cos\theta, \qquad y_B = Ut\sin\theta - \tfrac{1}{2}gt^2.$$

若两个坐标同时相等，它们便在时刻 T 碰撞（collision）。高度相等当且仅当

$$h - \tfrac{1}{2}gT^2 = UT\sin\theta - \tfrac{1}{2}gT^2 \quad\Longrightarrow\quad T = \frac{h}{U\sin\theta},$$

这是唯一可能的碰撞时间（由于 θ 是锐角，$\sin\theta > 0$）。在同一时刻，水平位置也必须相等：

$$VT = d - UT\cos\theta \quad\Longrightarrow\quad (V + U\cos\theta)\,\frac{h}{U\sin\theta} = d.$$

两边乘以 $U\sin\theta$，得 $Vh + Uh\cos\theta = Ud\sin\theta$，再除以 U，

$$d\sin\theta - h\cos\theta = \frac{Vh}{U},$$

此即所求。

第 (i) 部分

由直角三角形 ANB 得 $\tan\beta = h/d$，且 $0 < \beta < \frac{1}{2}\pi$。由于 V、h、U 都是正数，刚证得的结果给出 $d\sin\theta - h\cos\theta > 0$。除以 $d\cos\theta > 0$（因 θ 是锐角），

$$\tan\theta > \frac{h}{d} = \tan\beta.$$

两个角都落在 $(0, \frac{1}{2}\pi)$ 内，而 tan 在此区间上严格递增，所以 $\theta > \beta$。

第 (ii) 部分

> ? **想一想** 到目前为止的代数推导完全没有用到地面。我们还有甚么物理条件未加上?

碰撞必须在任何一粒质点到达地面之前发生，所以时刻 T 的共同高度必须至少为 0：

$$h - \tfrac{1}{2}gT^2 \geq 0 \quad\Longleftrightarrow\quad h \geq \frac{gh^2}{2U^2\sin^2\theta} \quad\Longleftrightarrow\quad U^2\sin^2\theta \geq \frac{gh}{2}.$$

（我们除以了 $h > 0$，并乘以了 $U^2\sin^2\theta > 0$。）由于 $U\sin\theta > 0$，取正平方根得

$$U\sin\theta \geq \sqrt{\frac{gh}{2}}.$$

等号对应两粒质点恰好在到达地面时相遇的情况。

第 (iii) 部分

记 $r = AB = \sqrt{d^2 + h^2}$，于是 $d = r\cos\beta$，$h = r\sin\beta$。那么

$$d\sin\theta - h\cos\theta = r(\sin\theta\cos\beta - \cos\theta\sin\beta) = r\sin(\theta - \beta),$$

而碰撞条件变成

$$r\sin(\theta-\beta) = \frac{Vr\sin\beta}{U} \quad\Longrightarrow\quad \frac{U}{V} = \frac{\sin\beta}{\sin(\theta-\beta)}.$$

由第 (i) 部分，$\theta - \beta > 0$；此外 $\theta - \beta < \theta < \frac{1}{2}\pi$。所以 $0 < \sin(\theta-\beta) < 1$，又因 $\sin\beta > 0$，

$$\frac{U}{V} = \frac{\sin\beta}{\sin(\theta-\beta)} > \sin\beta.$$

最后部分

碰撞点的高度是 $h - \frac{1}{2}gT^2$，所以

$$\text{碰撞点高于}\tfrac{1}{2}h \iff h - \tfrac{1}{2}gT^2 > \tfrac{1}{2}h \iff gT^2 < h.$$

从 B 出发的质点的铅垂速度是 $U\sin\theta - gt$，所以

$$\text{它正在上升，于时刻}T \iff U\sin\theta - gT > 0 \iff gT < U\sin\theta.$$

想一想 怎样才能把 $gT < U\sin\theta$ 与 $gT^2 < h$ 作比较?

利用碰撞时间本身：$T = h/(U\sin\theta)$ 表示 $U\sin\theta = h/T$。于是乘以 $T > 0$，

$$gT < U\sin\theta \iff gT < \frac{h}{T} \iff gT^2 < h.$$

两个条件都等价于 $gT^2 < h$，所以两粒质点在高于 $\frac{1}{2}h$ 处碰撞，当且仅当从 B 出发的质点在碰撞时仍在向上运动。

验算

- 取 $g = 9.8$，$h = 10$，$d = 20$，$U = 20$，$\theta = \frac{1}{4}\pi$。碰撞条件给出 $V = U(d\sin\theta - h\cos\theta)/h = 10\sqrt{2} \approx 14.14$，以及 $T = 1/\sqrt{2} \approx 0.707\,\text{s}$。此时两粒质点都位于 $x = 10$，$y = 10 - 4.9\times\frac{1}{2} = 7.55$，高于 $\frac{1}{2}h = 5$；从 B 出发的质点的铅垂速度为 $10\sqrt{2} - 9.8/\sqrt{2} \approx 7.21 > 0$，正在上升，与最后部分的预测一致。此外 $U/V = \sqrt{2} > \sin\beta = 1/\sqrt{5}$。
- 无重力的直觉：当 $g = 0$ 时，两粒质点沿直线运动，而条件 $U/V = \sin\beta/\sin(\theta-\beta)$ 正是由 A、B 和相遇点组成的三角形中的正弦定理（sine rule）。重力令两粒质点向下移动相同的距离，所以条件不变。

- 以电脑对数以千计随机选取的 h、d、U、θ 作验算，确认了碰撞、全部三条不等式，以及最后部分的等价关系。

常见错误

- 在第 (ii) 部分忘记地面：该不等式来自碰撞发生的高度至少为零，而不是来自碰撞方程本身的任何东西。
- 由 $\tan\theta > \tan\beta$ 推出 $\theta > \beta$，却没有说明两个角都是锐角，而 tan 在该范围内递增。
- 在第 (iii) 部分只写到 $\sin(\theta-\beta) \leq 1$ 便停下；严格不等式需要 $\theta - \beta < \frac{1}{2}\pi$。

STEP 之外

$-\frac{1}{2}gt^2$ 项的抵消，是转换到自由下落参考系（frame of reference）的一个例子：在一个以 g 加速的参考系中，重力消失，两粒质点都沿直线运动。这是伽利略（Galileo）变换和加速参考系等概念最简单的情形，这些概念正是大学第一门力学课程的开端，日后亦是爱因斯坦（Einstein）等效原理（equivalence principle）的出发点。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对非惯性参考系（non-inertial frame）有细致的讨论。

详细解答

STEP II 2020，第 10 题

力学 · 能量与平衡

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

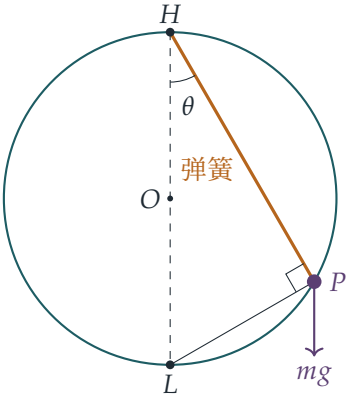
www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

由于 HL 是直径（diameter），利用“半圆上的圆周角是直角”（angle in a semicircle），弹簧长度和 P 的高度都可以直接用 θ 表示：$HP = 2a\cos\theta$，而 P 的高度比 H 低 $2a\cos^2\theta$。金属线（wire）是光滑的，所以它的反作用力（reaction）不做功。于是整道题只涉及一个函数，就是以 $c = \cos\theta$ 为自变量的势能（potential energy）。结果它是关于 c 的二次式（quadratic），配方后两部分便能一并解决。

解答

将 P 视为光滑金属线上的质点（particle），并把弹簧视为轻弹簧，遵从胡克定律（Hooke's law）（无论处于拉伸还是压缩状态）。设 O 为圆心。



几何关系。由于 HL 是直径，$\angle HPL = 90°$，因此在斜边为 $2a$ 的直角三角形 HPL 中，

$$HP = 2a\cos\theta.$$

由 H 到 P 的垂直落差为 $HP\cos\theta = 2a\cos^2\theta$。当 P 从 L 沿任一侧绕向 H 时，θ 的取值范围是 $(-\frac{\pi}{2}, \frac{\pi}{2})$，所以 $\cos\theta > 0$。

势能。取 H 所在的水平为重力势能(gravitational potential energy)的零点,并记 $c = \cos\theta$。弹簧的伸长量（extension）为 $2ac - l$，所以

$$\begin{aligned} V &= -2mga\,c^2 + \frac{\lambda}{2l}(2ac - l)^2 \\ &= \frac{2a(a\lambda - mgl)}{l}\,c^2 - 2a\lambda\,c + \frac{\lambda l}{2}. \end{aligned}$$

第 (i) 部分

光滑金属线的反作用力不做功, 因此 P 处于平衡（equilibrium）当且仅当 $\mathrm{d}V/\mathrm{d}\theta = 0$。(等价地说，沿切线方向的力互相平衡。）由连锁法则（chain rule），

$$\frac{\mathrm{d}V}{\mathrm{d}\theta} = -\sin\theta\,\frac{\mathrm{d}V}{\mathrm{d}c} = -\sin\theta\left(\frac{4a(a\lambda - mgl)}{l}\,c - 2a\lambda\right).$$

在 $\theta = \alpha$ 处（其中 $0 < \alpha < \frac{\pi}{2}$），$\sin\alpha \neq 0$，因此括号内的式子必须为零：

$$\frac{4a(a\lambda - mgl)}{l}\cos\alpha = 2a\lambda.$$

右方不为零，所以 $a\lambda \neq mgl$，故可以相除：

$$\cos\alpha = \frac{\lambda l}{2(a\lambda - mgl)},$$

此即所求。

要使这样的 α 存在，须有 $0 < \cos\alpha < 1$。

- $\cos\alpha > 0$：分子 λl 为正，故须 $a\lambda > mgl$。
- $\cos\alpha < 1$：两边乘以正数 $2(a\lambda - mgl)$，得 $\lambda l < 2a\lambda - 2mgl$，即 $\lambda(2a - l) > 2mgl$。

由于 $l < 2a$,第二个条件即 $\lambda > \dfrac{2mgl}{2a - l}$。这亦蕴含第一个条件,因为 $\dfrac{2mgl}{2a - l} > \dfrac{2mgl}{2a} = \dfrac{mgl}{a}$。所以 $\alpha > 0$ 的平衡位置存在，仅当 $\lambda > \dfrac{2mgl}{2a - l}$。（反之，若此条件成立，该公式给出一个真正的角 $\alpha \in (0, \frac{\pi}{2})$，而由对称性（symmetry），$-\alpha$ 亦是平衡位置。）

第 (ii) 部分

想一想 有没有一个巧妙的方法改写 V，使 $\cos\alpha$ 单独出现?

有：配方（completing the square）。记 $K = \dfrac{2a(a\lambda - mgl)}{l}$。由第 (i) 部分，$\dfrac{a\lambda - mgl}{l} = \dfrac{\lambda}{2\cos\alpha}$，所以

$$K = \frac{a\lambda}{\cos\alpha}, \qquad \frac{2a\lambda}{K} = 2\cos\alpha.$$

因此

$$V = K\left(c^2 - 2c\cos\alpha\right) + \frac{\lambda l}{2} = \frac{a\lambda}{\cos\alpha}(c - \cos\alpha)^2 + \text{常数}.$$

（作为验证，此式在 $c = \cos\alpha$ 处取最小值，正是第 (i) 部分求得的平衡位置。）按第 (i) 部分中平衡位置存在所需的条件，我们假设 $a\lambda > mgl$，于是 $\cos\alpha > 0$ 且 $K > 0$。

在 L 处，$\theta = 0$ 且 $c = 1$。对 L 与一般位置应用能量守恒（conservation of energy），得

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}mu^2 + \frac{a\lambda}{\cos\alpha}\left[(1 - \cos\alpha)^2 - (c - \cos\alpha)^2\right].$$

两边乘以 $\dfrac{\cos\alpha}{a\lambda} > 0$，

$$\frac{mv^2\cos\alpha}{2a\lambda} = D - (c - \cos\alpha)^2, \qquad \text{其中} D = (1 - \cos\alpha)^2 + \frac{mu^2}{2a\lambda}\cos\alpha. \qquad (*)$$

所以若 P 在 $\theta = \beta$ 处静止，则该处 $v = 0$，且

$$(\cos\alpha - \cos\beta)^2 = (1 - \cos\alpha)^2 + \frac{mu^2}{2a\lambda}\cos\alpha,$$

这就是所求的方程。

想一想 这个方程对 $\cos\beta$ 有两个根。哪一个才是真正的停止点？它又在甚么情况下位于 H 之前?

当 P 从 L 升向 H 时，c 由 1 递减至 0。由 $(*)$，只要 $(c - \cos\alpha)^2 < D$，P 便会继续运动；所以质点首次停下之处，就是 c 从 1 往下减少时碰到区间 $(\cos\alpha - \sqrt{D},\ \cos\alpha + \sqrt{D})$ 较近一端的位置。现在 $D \geq (1 - \cos\alpha)^2$（若 $u > 0$ 则严格不等），所以 $\cos\alpha + \sqrt{D} > 1$ 且 $\cos\alpha - \sqrt{D} < 1$。因此 $c = 1$ 位于区间之内，而停止时的值为

$$\cos\beta = \cos\alpha - \sqrt{D}.$$

P 在到达 H（该处 $c=0$）之前停下，当且仅当此值为正，即 $\sqrt{D}<\cos\alpha$，亦即 $D<\cos^2\alpha$：

$$\begin{aligned}&(1-\cos\alpha)^2+\frac{mu^2}{2a\lambda}\cos\alpha<\cos^2\alpha\\ \iff\quad &1-2\cos\alpha+\frac{mu^2}{2a\lambda}\cos\alpha<0\\ \iff\quad &\frac{mu^2}{2a\lambda}<2-\sec\alpha,\end{aligned}$$

最后一步两边除以 $\cos\alpha>0$。因此，P 在到达 H 之前静止，仅当

$$u^2<\frac{2a\lambda}{m}(2-\sec\alpha).$$

特别地，这要求 $\sec\alpha<2$，即 $\alpha<\frac{\pi}{3}$。否则无论 P 在 L 处的速率为何，都会到达 H。

验算

- 取 $m=1$、$g=10$、$a=1$、$l=1$、$\lambda=40$（采用一致的单位）。则 $\lambda>2mgl/(2a-l)=20$，且 $\cos\alpha=40/60=\frac{2}{3}$。临界值为 $u^2=80(2-\frac{3}{2})=40$。当 $u^2=40$ 时，$D=\frac{1}{9}+\frac{40}{80}\cdot\frac{2}{3}=\frac{4}{9}$，所以 $\cos\beta=\frac{2}{3}-\frac{2}{3}=0$。质点恰好到达 H，正如预期。
- 若 $u=0$，则 $D=(1-\cos\alpha)^2$ 且 $\cos\beta=1$。在 L 处静止的质点会一直留在那里；这是正确的，因为 L 亦是平衡位置。
- 以电脑对运动方程进行数值积分，并测试大量随机参数组合，结果显示：恰好在 u^2 低于临界值时，质点停在预测的 $\cos\beta$ 处。

常见错误

- 误取 $HP=2a\sin\theta$，或从 O 量度高度时用错了角。记住圆心角是 2θ，而不是 θ。
- 在第 (i) 部分只检查 $\cos\alpha>0$。其实是条件 $\cos\alpha<1$ 给出 λ 的界限。
- 选取根 $\cos\alpha+\sqrt{D}$ 作为 $\cos\beta$。它大于 1，所以不是金属线上的位置。

STEP 之外

把能量写成 $K(c-\cos\alpha)^2+$ 常数，正是大学本科力学中一个关键思想的一维版本。对于只有一个自由度（degree of freedom）的系统，平衡位置就是势能的驻点（stationary points），其稳定性（stability）取决于这些点是极小值还是极大值，而容许运动的区域就是能量超过势能的地方。在这里，当 $\alpha>0$ 存在时，L 会变成 V 的极大值，因而是不稳定平衡。弹簧令金属线的底部变得不稳定！Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）正是以这种精神介绍能量图与稳定性。

详细解答

STEP II 2020，第 11 题

概率与统计 • 掷硬币中的游程

> **STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

这个游戏几乎没有“记忆”：任何时刻，重要的只是最近一次掷出哪一面，以及目前连续（run）的长度。特别地，一旦掷出的面改变，游戏所处的局面就与“这新的一面是第一次投掷”完全相同。因此我们以第一次投掷为条件（conditioning），把“已知第一掷为正面时 A 获胜的概率（probability）”用“已知第一掷为反面时 A 获胜的概率”表示，反之亦然。这样便得到两条联立线性方程（simultaneous linear equations）。

解答

全文假设 $0 < p < 1$。（若 $p = 0$ 或 $p = 1$，每次掷出的都是同一面，游戏会在第二或第三次投掷时结束，答案是显然的。）记

$$\alpha = \mathrm{P}(\text{A 获胜} \mid \text{第一掷为 H}), \qquad \beta = \mathrm{P}(\text{A 获胜} \mid \text{第一掷为 T}).$$

各次投掷是独立（independent）的，所以任何一次投掷之后发生的事，只取决于之后尚未进行的投掷。

第 (i) 部分

游戏必定结束。掷了 n 次后游戏仍未结束，当且仅当到目前为止没有两次相邻投掷结果相同，即首 n 次投掷交替出现：$HTHT\ldots$ 或 $THTH\ldots$。当 $n = 2m$ 时，这两个序列各含 m 个正面和 m 个反面，所以

$$\mathrm{P}(\text{掷 } 2m \text{ 次后仍未结束}) = 2p^m q^m = 2(pq)^m \leq 2\left(\tfrac{1}{4}\right)^m,$$

因为 $pq = p(1-p) = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leq \frac{1}{4}$。事件“游戏永不结束”包含于上述每一个事件之中，因此对每一个 m，它的概率至多为 $2\left(\frac{1}{4}\right)^m$；令 $m \to \infty$，即知这概率为 0。

已知第一掷为正面。考虑第二次投掷。

- 以概率 p 掷出正面，A 立即获胜。
- 以概率 q 掷出反面。这时之前的正面已无关重要：B 还需要再多一个反面，A 需要连续两个正面，情况与这个反面是第一次投掷完全一样。所以此后 A 获胜的概率为 β。

同理，若第一掷为反面，则第二掷为反面时 B 立即获胜；第二掷为正面时，我们便处于“第一掷为正面”的局面。因此

$$\alpha = p + q\beta, \qquad \beta = p\alpha.$$

把第二条方程代入第一条，得 $\alpha = p + pq\,\alpha$，所以

$$\alpha = \frac{p}{1-pq},$$

正如所求（$1 - pq \geq \frac{3}{4} > 0$）。于是 $\beta = \dfrac{p^2}{1-pq}$。

整体概率。以第一次投掷为条件，

$$\mathrm{P}(\text{A 获胜}) = p\alpha + q\beta = \frac{p^2 + p^2 q}{1-pq} = \frac{p^2(1+q)}{1-pq}.$$

第 (ii) 部分

现在 α 和 β 指的是需要连续三次的游戏。假设第一掷为正面。

- HH 之后再掷出 H，概率为 p^2：A 获胜。
- HHT，概率为 pq；或 HT，概率为 q：自开局以来第一个反面刚刚出现，之前的正面不再计算，所以我们处于“第一掷为反面”的局面，A 获胜的概率为 β。

这些情况涵盖了所有可能，所以

$$\alpha = p^2 + (q + pq)\,\beta,$$

这正是题目所给的结果。把两面的角色互换（在“第一掷为反面”的局面下，TTT 令 B 获胜，对 A 的概率贡献为 0），得

$$\beta = (p + pq)\,\alpha.$$

想一想 因子 $q+pq$ 和 $p+pq$ 怎样变成 $1-p^2$ 和 $1-q^2$?

由于 $p+q=1$，我们有 $q+pq=q(1+p)=(1-p)(1+p)=1-p^2$，同理 $p+pq=p(1+q)=1-q^2$。这也可以直接理解：$1-p^2$ 是开局掷出正面后，接着两次投掷不全是正面的概率，而这恰恰就是连续被打断的情况。所以方程组为

$$\alpha = p^2 + (1-p^2)\beta, \qquad \beta = (1-q^2)\alpha,$$

代入后得 $\alpha\left[1-(1-p^2)(1-q^2)\right]=p^2$，即

$$\alpha = \frac{p^2}{1-(1-p^2)(1-q^2)}.$$

（分母为正，因为 $0<1-p^2<1$ 且 $0<1-q^2<1$。）最后，

$$\mathrm{P}(\text{A 获胜}) = p\alpha + q\beta = \alpha\big(p+q(1-q^2)\big) = \alpha\left(p+q-q^3\right) = \alpha\left(1-q^3\right),$$

所以

$$\mathrm{P}(\text{A 获胜}) = \frac{p^2(1-q^3)}{1-(1-p^2)(1-q^2)},$$

正如所求。

第 (iii) 部分

同样的论证适用于长度为 a 和 b 的连续。已知第一掷为正面，要么接着的 $a-1$ 次投掷全是正面（概率 p^{a-1}），A 获胜；要么其中出现反面（概率 $1-p^{a-1}$）。在后一情况下，游戏就从那个反面重新开始，令我们处于“第一掷为反面”的局面。同理，已知第一掷为反面，B 以概率 q^{b-1} 获胜，否则在 B 的连续完成之前会出现正面。所以

$$\alpha = p^{a-1} + \left(1-p^{a-1}\right)\beta, \qquad \beta = \left(1-q^{b-1}\right)\alpha,$$

因此

$$\alpha = \frac{p^{a-1}}{1-\left(1-p^{a-1}\right)\left(1-q^{b-1}\right)}.$$

分母同样为正，因为两个括号的值都严格介乎 0 与 1 之间。于是

$$p\alpha + q\beta = \alpha\left(p+q-q^b\right) = \alpha\left(1-q^b\right),$$

所以

$$\boxed{\mathrm{P}(\text{A 获胜}) = \frac{p^{a-1}\left(1-q^b\right)}{1-\left(1-p^{a-1}\right)\left(1-q^{b-1}\right)}}$$

如果你喜欢，分母可展开为 $p^{a-1}+q^{b-1}-p^{a-1}q^{b-1}$。

$a = b = 2$ **的情况**。公式变成

$$\frac{p(1-q^2)}{1-(1-p)(1-q)} = \frac{p(1-q)(1+q)}{1-pq} = \frac{p^2(1+q)}{1-pq},$$

其中分母用了 $1-p=q$、$1-q=p$，分子用了 $1-q=p$。这与第 (i) 部分一致。代入 $a=b=3$ 则恰好还原第 (ii) 部分的结果。

验算

- 公平硬币，第 (i) 部分。当 $p=q=\frac{1}{2}$ 时，答案为 $\frac{1}{4}\cdot\frac{3}{2}/\frac{3}{4}=\frac{1}{2}$，与对称性的要求相符。在 (iii) 中，只要 $a=b$ 且 $p=\frac{1}{2}$，结果也是如此。
- 极端情况。当 $p\to 1$ 时，(iii) 的答案趋向 1；当 $p\to 0$ 时，它趋向 0，理应如此。
- 数值。当 $p=0.4$ 时，第 (i) 部分的答案为 $0.16\times 1.6/0.76\approx 0.3368$，第 (ii) 部分的答案约为 ≈ 0.2713。以电脑模拟 200 000 局游戏，得到 0.336 和 0.271。当 $p=0.55$、$a=4$、$b=2$ 时，公式给出 0.2450，同样与模拟结果吻合。

常见错误

- 忘记了一旦掷出的面改变，之前的连续便完全不算数。正因如此，改变之后的局面才与“该面为第一掷”相同。
- 在 (ii) 中漏掉 HHT 的情况：连续可以在第二次或第三次投掷时被打断。
- 把 α 当作“A 获胜的概率”的最终答案。你仍须对第一次投掷取平均，即 $p\alpha+q\beta$。

STEP 之外

这是对吸收马尔可夫链（absorbing Markov chain）的首步分析（first-step analysis）。状态是“目前的面及连续的长度”，而吸收概率满足以下一步为条件所得的线性方程。同样的技巧可以求出掷硬币时各种图案的期望等待时间，并引申至彭尼游戏（Penney's game）：在该游戏中，图案的先后次序会产生一些令人意外的非传递（non-transitive）结果。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）以合适的程度讲解了马尔可夫链与吸收。

详细解答

STEP II 2020，第 12 题

概率与统计 · 不公平骰子的不等式

STEP II 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

各概率（probability）之和必须等于 1，所以各偏差（bias）满足 $\sum \varepsilon_i = 0$。把这一点写下来之后，展开 $\sum(\frac{1}{n}+\varepsilon_i)^2$ 时所有一次项都会消失，剩下的式子一眼就看得出是非负的。第 (ii) 部分其实是第 (i) 部分换了个样子：把一组正数各自除以它们的总和，就得到某颗骰子（die）的各个概率。

解答

设 $p_i = \mathrm{P}(X=i) = \frac{1}{n} + \varepsilon_i$。由于 $\sum_{i=1}^{n} p_i = 1$，

$$1 = n\cdot\frac{1}{n} + \sum_{i=1}^{n}\varepsilon_i \quad\Longrightarrow\quad \sum_{i=1}^{n}\varepsilon_i = 0.$$

另外每个 $p_i \geq 0$，所以对每个 i 都有 $\varepsilon_i \geq -\frac{1}{n}$。以下所有求和都是对 $i = 1,\ldots,n$ 进行。

第 (i) 部分

两次掷骰是独立（independent）的，所以两次都掷出 i 的概率是 p_i^2，而对不同的 i，这些事件互斥（disjoint）。因此

$$\begin{aligned}\mathrm{P}(\text{两次点数相同}) &= \sum p_i^2 = \sum\left(\frac{1}{n^2} + \frac{2\varepsilon_i}{n} + \varepsilon_i^2\right)\\ &= \frac{1}{n} + \frac{2}{n}\sum\varepsilon_i + \sum\varepsilon_i^2 = \frac{1}{n} + \sum_{i=1}^{n}\varepsilon_i^2.\end{aligned}$$

对公平的骰子，每个 $\varepsilon_i = 0$，概率为 $\frac{1}{n}$。对有偏差的骰子，ε_i 不全为零，所以 $\sum\varepsilon_i^2 > 0$。因此，有偏差的骰子在连续两次掷骰中掷出相同点数的可能性，高于公平的骰子。

第 (ii) 部分

先改写这个二重和。它把所有 $j < i$ 的 $x_i x_j$ 加起来，也就是说，每一对不同下标组成的无序对恰好加一次。把 $S = \sum x_i$ 平方，会得到每个 x_i^2 各一次，以及每个这样的无序对各两次，所以

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i x_j = \frac{1}{2}\left(S^2 - \sum x_i^2\right).$$

想一想 当 x_i 只是任意的正数时，骰子从何而来?

设 $p_i = x_i / S$。这些数都是正数，而且总和为 1，所以它们是某颗 n 面骰子的各个概率。由第 (i) 部分，$\sum p_i^2 = \frac{1}{n} + \sum \varepsilon_i^2 \geq \frac{1}{n}$（等号成立当且仅当骰子是公平的，即所有 x_i 相等）。两边乘以 $S^2 > 0$，

$$\sum x_i^2 \geq \frac{S^2}{n}.$$

因此

$$\sum_{i=2}^{n} \sum_{j=1}^{i-1} x_i x_j = \frac{1}{2}\left(S^2 - \sum x_i^2\right) \leq \frac{1}{2}\left(S^2 - \frac{S^2}{n}\right) = \frac{n-1}{2n}\left(\sum_{i=1}^{n} x_i\right)^2,$$

这正是所求的不等式（inequality）。当所有 x_i 相等时，等号成立。

第 (iii) 部分

现在概率是 $\sum p_i^3$。展开，并再次利用 $\sum \varepsilon_i = 0$，

$$\begin{aligned}\sum p_i^3 &= \sum \left(\frac{1}{n^3} + \frac{3\varepsilon_i}{n^2} + \frac{3\varepsilon_i^2}{n} + \varepsilon_i^3\right) \\ &= \frac{1}{n^2} + \sum_{i=1}^{n}\left(\frac{3\varepsilon_i^2}{n} + \varepsilon_i^3\right) = \frac{1}{n^2} + \sum_{i=1}^{n} \varepsilon_i^2 \left(\frac{3}{n} + \varepsilon_i\right).\end{aligned}$$

对公平的骰子，这个值是 $\frac{1}{n^2}$。

想一想 立方项 ε_i^3 可以是负数。它们会不会盖过平方项?

不会，因为偏差不可能负得太多：由 $p_i \geq 0$ 得 $\varepsilon_i \geq -\frac{1}{n}$，所以

$$\frac{3}{n} + \varepsilon_i \geq \frac{2}{n} > 0.$$

因此每一项 $\varepsilon_i^2\left(\frac{3}{n} + \varepsilon_i\right)$ 都 ≥ 0，而且只要 $\varepsilon_i \neq 0$ 就严格大于零；对有偏差的骰子，至少有一个 i 满足这个条件。所以

$$\sum p_i^3 > \frac{1}{n^2},$$

即有偏差的骰子在连续三次掷骰中掷出相同点数的可能性，同样高于公平的骰子。

验算

- 取 $n = 4$，各概率为 $\frac{1}{2}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$，于是 $\varepsilon = \frac{1}{4}, -\frac{1}{12}, -\frac{1}{12}, -\frac{1}{12}$。直接计算得 $\sum p_i^2 = \frac{1}{4} + \frac{3}{36} = \frac{1}{3}$，而 $\frac{1}{4} + \sum \varepsilon_i^2 = \frac{1}{4} + \frac{1}{16} + \frac{3}{144} = \frac{1}{3}$，比公平骰子的 $\frac{1}{4}$ 大。掷三次时，$\sum p_i^3 = \frac{1}{8} + \frac{3}{216} = \frac{5}{36} \approx 0.139$，远高于公平骰子的 $\frac{1}{16} = 0.0625$。
- 最极端的骰子总是掷出 1：这时两个概率都是 1，即最大可能值。
- 在 (ii) 中，当 $n = 2$ 时不等式变成 $x_1 x_2 \le \frac{1}{4}(x_1 + x_2)^2$，这正是算术–几何平均不等式（AM–GM）。(i)、(ii) 和 (iii) 中的不等式亦已用电脑在数以千计的随机例子上验证过。

常见错误

- 忘记 $\sum \varepsilon_i = 0$，结果留下一些正负号未知的一次项。
- 在 (iii) 中，没有用 $\varepsilon_i \ge -\frac{1}{n}$，就断言 $\sum \varepsilon_i^3$ “很小”或可正可负；正是这个界令论证成立。
- 在 (ii) 中，把第 (i) 部分直接用在 x_i 本身，而不是用在 $x_i / \sum x_j$ 上；x_i 的总和不一定是 1。

STEP 之外

$\sum p_i^2$ 这个量称为一个分布的*碰撞概率*（collision probability），第 (i) 部分说明它在均匀分布（uniform distribution）时最小；第 (iii) 部分则是关于 $\sum p_i^3$ 的同一结论。两者都是函数 $t \mapsto t^k$ 的凸性（convexity）——即延森不等式（Jensen's inequality）——或幂平均不等式（power-mean inequality）的特例；而 $\sum p_i^k$ 的对数则给出信息理论（information theory）和密码学（cryptography）中使用的雷尼熵（Rényi entropy）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）是在概率框架下认识这些概念的好起点。

详细解答

STEP III 2020，第 1 题

纯数学 · 归约公式

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

第 (i) 部分是一条递推公式（reduction formula），所以自然的第一步是分部积分法（integration by parts）：对 $\cos x$ 的幂求导，并对 $\cos bx$ 积分。这样会剩下乘积 $\sin x \sin bx$，而 $\cos(b-1)x$ 的复角公式（compound-angle formula）可以把它变回原本类型的积分。第 (ii) 部分不需要任何新工具：递推关系（recurrence relation）把 a 和 b 同时减一，使差 $b-a=2m+1$ 保持不变，因此可以顺利地由 n 推进到 $n+1$。

解答

第 (i) 部分

设 $a, b \geq 1$。取 $u=\cos^a x$ 及 $\mathrm{d}v=\cos bx\,\mathrm{d}x$ 进行分部积分，于是 $\mathrm{d}u=-a\cos^{a-1}x\sin x\,\mathrm{d}x$，而 $v=\frac{1}{b}\sin bx$（由于 $b\neq 0$，这样做没有问题）：

$$\mathrm{I}(a,b)=\left[\frac{\cos^a x\,\sin bx}{b}\right]_0^{\pi/2}+\frac{a}{b}\int_0^{\pi/2}\cos^{a-1}x\,\sin x\,\sin bx\,\mathrm{d}x.$$

边界项在两端都等于零：在 $x=\frac{\pi}{2}$ 处是因为 $\cos^a\frac{\pi}{2}=0$（这里用到 $a\geq 1$），而在 $x=0$ 处是因为 $\sin 0=0$。

想一想 怎样才能把 $\sin x\sin bx$ 变回 x 的倍数的余弦？

我们希望出现 $\mathrm{I}(a-1,b-1)$，所以考虑 $\cos(b-1)x$：

$$\cos(b-1)x=\cos bx\cos x+\sin bx\sin x \quad\Longrightarrow\quad \sin x\sin bx=\cos(b-1)x-\cos x\cos bx.$$

代入后得

$$\begin{aligned}\mathrm{I}(a,b) &= \frac{a}{b}\int_0^{\pi/2} \cos^{a-1}x\,\cos(b-1)x\,\mathrm{d}x - \frac{a}{b}\int_0^{\pi/2}\cos^a x\,\cos bx\,\mathrm{d}x\\ &= \frac{a}{b}\,\mathrm{I}(a-1,b-1) - \frac{a}{b}\,\mathrm{I}(a,b).\end{aligned}$$

两边乘以 b，再把含 $\mathrm{I}(a,b)$ 的项合并，得 $(a+b)\,\mathrm{I}(a,b) = a\,\mathrm{I}(a-1,b-1)$；由于 $a+b>0$，

$$\mathrm{I}(a,b) = \frac{a}{a+b}\,\mathrm{I}(a-1,b-1),$$

此即所求。

第 (ii) 部分

左边的积分是 $\mathrm{I}(n,\,n+2m+1)$。记右边为

$$R(n,m) = (-1)^m\,\frac{2^n\,n!\,(2m)!\,(n+m)!}{m!\,(2n+2m+1)!}$$

。固定任意非负整数 m；我们对 n 用数学归纳法（induction）证明对所有 $n\ge 0$，$\mathrm{I}(n,n+2m+1) = R(n,m)$。

基本情况， $n=0$。

$$\mathrm{I}(0,2m+1) = \int_0^{\pi/2}\cos(2m+1)x\,\mathrm{d}x = \frac{\sin\!\left((2m+1)\frac{\pi}{2}\right)}{2m+1} = \frac{(-1)^m}{2m+1},$$

因为 $\sin\!\left(m\pi+\frac{\pi}{2}\right) = \cos m\pi = (-1)^m$。另一方面，

$$R(0,m) = (-1)^m\,\frac{1\cdot 1\cdot(2m)!\,m!}{m!\,(2m+1)!} = \frac{(-1)^m}{2m+1}.$$

所以结果在 $n=0$ 时成立。

归纳步骤。假设对某个 $k\ge 0$，$\mathrm{I}(k,k+2m+1) = R(k,m)$。在第 (i) 部分中取 $a = k+1$ 及 $b = k+2m+2$，两者都是正整数，于是 $a-1=k$、$b-1 = k+2m+1$ 而 $a+b = 2k+2m+3$：

$$\mathrm{I}(k+1,\,k+2m+2) = \frac{k+1}{2k+2m+3}\,\mathrm{I}(k,\,k+2m+1) = \frac{k+1}{2k+2m+3}\,R(k,m).$$

留意 $k+2m+2 = (k+1)+2m+1$，所以左边正正是 $n=k+1$ 且 m 不变时的积分。

> ? **想一想** 按公式，$R(k+1,m)/R(k,m)$ 应该等于甚么?

在 R 中由 $n=k$ 变到 $n=k+1$ 时，因子 $(-1)^m(2m)!/m!$ 不变，而

$$\begin{aligned}\frac{R(k+1,m)}{R(k,m)} &= \frac{2^{k+1}}{2^k}\cdot\frac{(k+1)!}{k!}\cdot\frac{(k+m+1)!}{(k+m)!}\cdot\frac{(2k+2m+1)!}{(2k+2m+3)!}\\ &= \frac{2(k+1)(k+m+1)}{(2k+2m+3)(2k+2m+2)} = \frac{k+1}{2k+2m+3},\end{aligned}$$

这里用到 $2k+2m+2=2(k+m+1)$。因此

$$\mathrm{I}(k+1,\,k+2m+2)=\frac{k+1}{2k+2m+3}\,R(k,m)=R(k+1,m),$$

这就是 $n=k+1$ 时的命题。

由于结果在 $n=0$ 时成立，而且它在 $n=k$ 时成立可推出它在 $n=k+1$ 时成立，根据数学归纳法，它对所有非负整数 n 都成立。由于 m 是任意的，它对所有非负整数 m 也成立：

$$\int_0^{\pi/2}\cos^n x\,\cos(n+2m+1)x\,\mathrm{d}x=(-1)^m\,\frac{2^n\,n!\,(2m)!\,(n+m)!}{m!\,(2n+2m+1)!}.$$

验算

- 取 $n=1$、$m=0$：公式给出 $\frac{2\cdot1\cdot1\cdot1}{1\cdot3!}=\frac{1}{3}$。直接计算，$\cos x\cos 2x=\frac{1}{2}(\cos 3x+\cos x)$ 在 $[0,\frac{\pi}{2}]$ 上的积分是 $\frac{1}{2}\left(-\frac{1}{3}+1\right)=\frac{1}{3}$。
- 取 $n=1$、$m=1$：公式给出 $-\frac{2\cdot2\cdot2}{5!}=-\frac{1}{15}$，而以数值方法计算 $\int_0^{\pi/2}\cos x\cos 5x\,\mathrm{d}x$ 得 -0.066667。
- 第 (i) 部分的递推关系和第 (ii) 部分的闭合形式，亦已用电脑就所有 $n\le 6$、$m\le 5$ 验证过。

常见错误

- 忘记说明边界项为何等于零；上限处需要 $a\ge 1$。
- 对 n 进行归纳，却让 m 跟着改变：第 (i) 部分把 a 和 b 一同减一，所以 m 保持不变。
- 在 $\sin\left((2m+1)\frac{\pi}{2}\right)$ 中弄错正负号；它依次为 $1,-1,1,\ldots$，因而给出 $(-1)^m$。

STEP 之外

这类递推公式是递推关系在积分上的对应，而闭合形式由阶乘（factorial）的比构成，方式与沃利斯乘积（Wallis's product）及 Beta 函数（Beta function）$\int_0^{\pi/2}\cos^p x\sin^q x\,\mathrm{d}x$ 相同；这些内容在大学分析或数学方法课程的早期便会遇到。如想做更多同类题目，可参阅 Siklos，*Advanced Problems in Mathematics*（《数学进阶题》）。

详细解答

STEP III 2020，第 2 题

纯数学 · 双曲函数曲线描绘

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

进行隐函数求导（implicit differentiation），并反复利用 $\cosh^2 = 1 + \sinh^2$ 把一切化回 sinh 的形式：这样二阶导数的条件便可以因式分解，而其中一个因子正是曲线本身的方程。至于直线 $x + y = a$，代入 $y = a - x$ 并把所有项写成指数形式；要有实数交点，判别式（discriminant）必须非负，由此便得出 $\cosh a$ 的界限。

解答

第 (i) 部分

由于 sinh 是 $\mathbb{R}$ 上的双射（bijection），对每个 x，恰好有一个 y 使 $\sinh y = 2k - \sinh x$。因此 C 是函数 $y = \operatorname{arsinh}(2k - \sinh x)$ 的图像，对所有实数 x 均有定义，我们可以进行隐函数求导：

$$\cosh x + \cosh y \frac{\mathrm{d}y}{\mathrm{d}x} = 0 \quad \Longrightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\cosh x}{\cosh y}.$$

由于 $\cosh x \geq 1$，斜率（gradient）永不为零（它总是负数），所以 C 没有驻点（stationary point）。

再求导一次，记 $y' = \mathrm{d}y/\mathrm{d}x$，得

$$\sinh x + \sinh y \, (y')^2 + \cosh y \, y'' = 0.$$

代入 $y' = -\cosh x / \cosh y$，并在两边乘以 $\cosh^2 y$：

$$\begin{aligned}\cosh^3 y\, y'' &= -\left(\sinh x \cosh^2 y + \sinh y \cosh^2 x\right)\\ &= -\left(\sinh x\,(1+\sinh^2 y) + \sinh y\,(1+\sinh^2 x)\right)\\ &= -(\sinh x + \sinh y) - \sinh x \sinh y\,(\sinh y + \sinh x)\\ &= -(\sinh x + \sinh y)(1 + \sinh x \sinh y).\end{aligned}$$

在 C 上，第一个因子等于 $2k$，所以

$$y'' = -\frac{2k\,(1+\sinh x \sinh y)}{\cosh^3 y}.$$

由于 $k > 0$ 且 $\cosh y > 0$，可知 $y'' = 0$ 当且仅当 $1 + \sinh x \sinh y = 0$，此即所求。

想一想 单凭 $y'' = 0$ 足以断定该点是拐点吗?

不足够：我们还需要 y'' 变号。设 $s = \sinh x$ 及 $t = \sinh y$。在 $y'' = 0$ 的点上，我们有 $s + t = 2k$ 及 $st = -1$，所以 s 和 t 是 $u^2 - 2ku - 1 = 0$ 的两个根：

$$u = k \pm \sqrt{k^2+1}.$$

一个根为正，另一个为负；s 可取其中任一个值，t 则取另一个。沿着曲线，$1 + st = 1 + s(2k - s) = -(s^2 - 2ks - 1)$ 是 s 的二次式，有两个单根，所以在每个根处都会变号。由于 $s = \sinh x$ 是 x 的递增函数，y'' 在这两点的确会变号。因此拐点（point of inflection）为

$$\left(\operatorname{arsinh}\left(k+\sqrt{k^2+1}\right),\ \operatorname{arsinh}\left(k-\sqrt{k^2+1}\right)\right) \quad 及$$
$$\left(\operatorname{arsinh}\left(k-\sqrt{k^2+1}\right),\ \operatorname{arsinh}\left(k+\sqrt{k^2+1}\right)\right).$$

它们互为关于 $y = x$ 的反射，这正合预期，因为 C 的方程对 x 和 y 是对称的。

第 (ii) 部分

在直线上，$y = a - x$。利用 $2\sinh u = \mathrm{e}^u - \mathrm{e}^{-u}$，把 C 的方程乘以 2 后变成

$$\mathrm{e}^x - \mathrm{e}^{-x} + \mathrm{e}^{a-x} - \mathrm{e}^{x-a} = 4k.$$

乘以 e^x 得 $\mathrm{e}^{2x} - 1 + \mathrm{e}^a - \mathrm{e}^{2x-a} = 4k\mathrm{e}^x$，即

$$\mathrm{e}^{2x}(1 - \mathrm{e}^{-a}) - 4k\mathrm{e}^x + (\mathrm{e}^a - 1) = 0,$$

此即所求的方程。

想一想 这样的点存在，对这个方程意味着甚么？

它是 $X = \mathrm{e}^x$ 的二次方程（只要首项系数不为零），而且必须有实根。

首先，$a \neq 0$：若 $a = 0$，方程会退化为 $-4k\mathrm{e}^x = 0$，由于 $k > 0$，这是不可能的。因此 $\cosh a > 1$。

当 $a \neq 0$ 时，首项系数 $1 - \mathrm{e}^{-a}$ 不为零，所以该二次方程有实根，仅当其判别式非负：

$$16k^2 - 4(1 - \mathrm{e}^{-a})(\mathrm{e}^a - 1) \geq 0.$$

而 $(1 - \mathrm{e}^{-a})(\mathrm{e}^a - 1) = \mathrm{e}^a - 2 + \mathrm{e}^{-a} = 2\cosh a - 2$，所以条件为 $16k^2 \geq 8(\cosh a - 1)$，即 $\cosh a \leq 2k^2 + 1$。合起来，

$$1 < \cosh a \leq 2k^2 + 1.$$

所以 C 上每一点的 $x + y$ 都介乎 $-2\operatorname{arsinh} k$ 与 $2\operatorname{arsinh} k$ 之间（因为 $\cosh(2\operatorname{arsinh} k) = 1 + 2\sinh^2(\operatorname{arsinh} k) = 1 + 2k^2$），而且 $x + y \neq 0$。

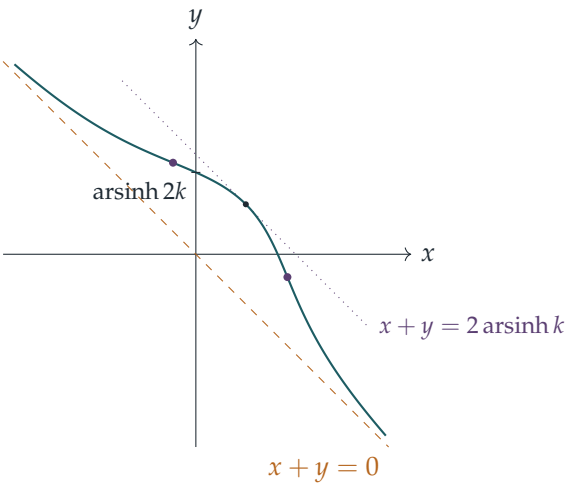
事实上，在 C 上 $x + y > 0$。若 $a < 0$，则 $1 - \mathrm{e}^{-a} < 0$ 且 $\mathrm{e}^a - 1 < 0$，所以该二次方程的三个系数全是负数，因而没有正根 X。（等价地，$\sinh x + \sinh y = 2\sinh \frac{x+y}{2} \cosh \frac{x-y}{2}$，其正负号与 $x + y$ 相同。）最大值 $x + y = 2\operatorname{arsinh} k$ 出现在 $x = y = \operatorname{arsinh} k$，此时判别式为零，直线刚好与 C 相切。

第 (iii) 部分

综合已知的结果：

- C 关于直线 $y = x$ 对称，并只在 $(\operatorname{arsinh} k, \operatorname{arsinh} k)$ 与之相交，该处的斜率为 -1。
- y 对所有 x 均有定义，且斜率总是负数。曲线与坐标轴相交于 $(\operatorname{arsinh} 2k, 0)$ 及 $(0, \operatorname{arsinh} 2k)$。
- $0 < x + y \leq 2\operatorname{arsinh} k$，所以 C 位于直线 $x + y = 0$ 上方、直线 $x + y = 2\operatorname{arsinh} k$ 上或其下方的带状区域内，而后者在 C 位于 $y = x$ 上的那一点与 C 相切。
- 当 $x \to +\infty$，$\sinh y = 2k - \sinh x \to -\infty$，且 $\mathrm{e}^{-y} \approx \mathrm{e}^x - 4k$，所以 $x + y \approx -\ln(1 - 4k\mathrm{e}^{-x}) \approx 4k\mathrm{e}^{-x} \to 0^+$。由对称性，当 $x \to -\infty$ 时情况相同。因此 $x + y = 0$ 是曲线两端的渐近线（asymptote），曲线从上方趋近它。
- 由 y'' 的公式可知，在两个拐点之间，即 $1 + \sinh x \sinh y > 0$ 之处（例如 $y = x$ 上的那一点），曲线是凹的（concave，位于其切线下方）；在两个拐点以外，曲线是凸的（convex）。

下图显示 $k = 1$ 的情况，并标出了拐点。



验算

- 当 $k = 1$，拐点位于 $\left(\operatorname{arsinh}(1+\sqrt{2}), \operatorname{arsinh}(1-\sqrt{2})\right) \approx (1.615, -0.403)$ 及其镜像；在两侧计算 $y = \operatorname{arsinh}(2 - \sinh x)$ 的二阶差分，确认 y'' 在该处变号；在 $-8 \le x \le 8$ 范围内扫描，找不到 y'' 的其他零点。
- 公式 $y'' = -2k(1 + \sinh x \sinh y)/\cosh^3 y$ 在许多随机点、对多个 k 值，都与数值二阶导数吻合。
- 当 $k = 1$ 及 $x = 6$，$x + y \approx 0.00996$，接近预测值 $4\mathrm{e}^{-6} \approx 0.00992$：一如所述，曲线从上方趋近渐近线。
- 在 $x = y = \operatorname{arsinh} k$，$\cosh(x+y) = 1 + 2\sinh^2 x = 1 + 2k^2$：第 (ii) 部分的上界可以达到。

常见错误

- 只得出 $y'' = 0$ 便停下，没有检查 y'' 是否变号。这里它确实变号，但必须说明。
- 忘记排除 $a = 0$，而正是这一点令左边的不等式成为严格不等式。
- 只写出一个拐点。$u^2 - 2ku - 1 = 0$ 的任一个根都可以是 $\sinh x$，因此得出两个关于 $y = x$ 互为镜像的点。

STEP 之外

旋转后的坐标 $u = x + y$、$v = x - y$ 令这条曲线一目了然：恒等式 $\sinh x + \sinh y = 2\sinh\frac{u}{2}\cosh\frac{v}{2}$ 把 C 化为 $\sinh\frac{u}{2} = k\,\mathrm{sech}\frac{v}{2}$，即在渐近线 $u = 0$ 上方的单一隆起。选取与对称性相配合的坐标，是大学一年级微积分和线性代数中贯穿始终的主题（把二次型（quadratic form）对角化正是同一个想法）。隐函数曲线，以及方程 $F(x, y) = 0$ 在甚么情况下真正把 y 定义为 x 的光滑函数，是隐函数定理（implicit function theorem）的研究对象；Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）对其背后的单变量分析有严谨的阐述。

详细解答

STEP III 2020，第 3 题

纯数学 · 几何中的复数

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

在复数平面（complex plane）上，以 a 为中心、转角为 θ 的旋转（rotation）不过就是 $z \mapsto a + (z-a)\mathrm{e}^{\mathrm{i}\theta}$，因此 G_{AB} 是 a 和 b 的一个明确的线性组合（linear combination）。一旦第 (i) 部分给出这条公式，几何问题便化为代数问题：平行四边形（parallelogram）是顶点所满足的一个线性条件，而等边三角形（equilateral triangle）则是某一边等于另一边转过一个固定角度。整道题目只建基于关于 ω 的幂的几个事实。

解答

全文中，$\omega = \mathrm{e}^{\mathrm{i}\pi/6} = \frac{\sqrt{3}}{2} + \frac{1}{2}\mathrm{i}$，故 $\omega^* = \mathrm{e}^{-\mathrm{i}\pi/6}$，并记

$$\rho = \mathrm{e}^{-\mathrm{i}\pi/3} = \tfrac{1}{2} - \tfrac{\sqrt{3}}{2}\mathrm{i}.$$

我们会用到以下三个事实：

$$\omega^2 = \mathrm{e}^{\mathrm{i}\pi/3}, \qquad \omega^3 = \mathrm{i}, \qquad \omega - \omega^* = 2\mathrm{i}\sin\tfrac{\pi}{6} = \mathrm{i}.$$

第 (i) 部分

顺时针(clockwise)旋转 $\frac{1}{3}\pi$ 相当于乘以 ρ。以 A 为中心旋转，意思是把位移(displacement) $b-a$ 旋转，再加回到 a 上，所以

$$k = a + (b-a)\rho.$$

因此 ABK 的形心（centroid）为

$$g_{ab} = \tfrac{1}{3}(a+b+k) = \tfrac{1}{3}\left[(2-\rho)\,a + (1+\rho)\,b\right].$$

接着整理两个系数：

$$2-\rho=\tfrac{3}{2}+\tfrac{\sqrt{3}}{2}\mathrm{i}=\sqrt{3}\left(\tfrac{\sqrt{3}}{2}+\tfrac{1}{2}\mathrm{i}\right)=\sqrt{3}\,\omega,$$
$$1+\rho=\tfrac{3}{2}-\tfrac{\sqrt{3}}{2}\mathrm{i}=\sqrt{3}\left(\tfrac{\sqrt{3}}{2}-\tfrac{1}{2}\mathrm{i}\right)=\sqrt{3}\,\omega^*.$$

于是

$$g_{ab}=\tfrac{1}{3}\left(\sqrt{3}\,\omega a+\sqrt{3}\,\omega^* b\right)=\frac{1}{\sqrt{3}}\left(\omega a+\omega^* b\right),$$

此即所求。

第 (ii) 部分

四边形（quadrilateral）$PQRS$（顶点依此次序排列）是平行四边形，当且仅当 $\overrightarrow{PQ}=\overrightarrow{SR}$，即 $q-p=r-s$，亦即

$$p-q+r-s=0.$$

（这等同于说对角线（diagonal）PR 与 QS 有相同的中点（midpoint）。）所以 Q_1 是平行四边形当且仅当 $a-b+c-d=0$，而 Q_2 是平行四边形当且仅当

$$g_{ab}-g_{bc}+g_{cd}-g_{da}=0.$$

对四个点各自运用第 (i) 部分的结果，

$$\begin{aligned}\sqrt{3}\left(g_{ab}-g_{bc}+g_{cd}-g_{da}\right)&=\left(\omega a+\omega^* b\right)-\left(\omega b+\omega^* c\right)+\left(\omega c+\omega^* d\right)-\left(\omega d+\omega^* a\right)\\&=\left(\omega-\omega^*\right)a-\left(\omega-\omega^*\right)b+\left(\omega-\omega^*\right)c-\left(\omega-\omega^*\right)d\\&=\mathrm{i}\,(a-b+c-d).\end{aligned}$$

因此

$$g_{ab}-g_{bc}+g_{cd}-g_{da}=\frac{\mathrm{i}}{\sqrt{3}}\,(a-b+c-d),$$

由于 $\mathrm{i}/\sqrt{3}\neq 0$，一边为零当且仅当另一边为零。所以 Q_1 是平行四边形当且仅当 Q_2 是平行四边形。

想一想 Q_2 有没有可能满足平行四边形条件，却退化成有两个顶点重合？

有的，在一个特殊情况下会发生。若 ABC 是一个按顺时针标记的等边三角形，那么把 B 绕 A 顺时针旋转便会落在 C 上，因此 G_{AB} 和 G_{BC} 都是 ABC 的形心。把它补成菱形（rhombus）$ABCD$，Q_2 就压扁成一条线段。此时条件 $p-q+r-s=0$ 仍然成立，所以这个等价关系实际上是关于这个向量条件的陈述，而在这类题目中，“平行四边形”指的正是这个条件。

第 (iii) 部分

我们证明 T_2 的某一边（作为向量）等于另一边转过一个固定角度。由第 (i) 部分，

$$\sqrt{3}\,(g_{bc}-g_{ab}) = (\omega b+\omega^* c)-(\omega a+\omega^* b) = -\omega a+\mathrm{i}\,b+\omega^* c,$$
$$\sqrt{3}\,(g_{ca}-g_{bc}) = (\omega c+\omega^* a)-(\omega b+\omega^* c) = \omega^* a-\omega b+\mathrm{i}\,c,$$

每一行都用了 $\omega-\omega^*=\mathrm{i}$。

> **想一想** 哪个数 λ 能把第二个括号变成第一个?

比较 a 的系数，必须有 $\lambda\omega^*=-\omega$，所以 $\lambda=-\omega/\omega^*=-\omega^2=-\mathrm{e}^{\mathrm{i}\pi/3}=\mathrm{e}^{-2\pi\mathrm{i}/3}$。同一个 λ 对 b 和 c 亦适用：

$$\lambda\cdot(-\omega)=\omega^3=\mathrm{i},$$
$$\lambda\cdot\mathrm{i}=-\omega^2\cdot\omega^3=-\omega^5=-\mathrm{e}^{5\pi\mathrm{i}/6}=\mathrm{e}^{-\mathrm{i}\pi/6}=\omega^*.$$

因此，对所有 a、b、c，

$$g_{bc}-g_{ab}=\mathrm{e}^{-2\pi\mathrm{i}/3}\,(g_{ca}-g_{bc}).$$

取模（modulus），由于 $\left|\mathrm{e}^{-2\pi\mathrm{i}/3}\right|=1$，得 $G_{AB}G_{BC}=G_{BC}G_{CA}$。此外，沿 T_2 行进的方向在 G_{BC} 处转过 $\frac{2}{3}\pi$，所以 T_2 在 G_{BC} 处的内角（interior angle）为 $\pi-\frac{2}{3}\pi=\frac{1}{3}\pi$。一个三角形若有两条相等的边夹着 $\frac{1}{3}\pi$ 的角，便是等腰（isosceles）三角形，其底角为 $\frac{1}{2}(\pi-\frac{1}{3}\pi)=\frac{1}{3}\pi$，所以三个角都是 $\frac{1}{3}\pi$，T_2 是等边三角形。

唯一的例外是 $g_{ca}=g_{bc}$ 的情况，此时上述关系亦给出 $g_{ab}=g_{bc}$，所以 T_2 缩成一点。把 $\omega^* a-\omega b+\mathrm{i}c=0$ 乘以 ω，得 $a-\mathrm{e}^{\mathrm{i}\pi/3}b+\mathrm{e}^{2\pi\mathrm{i}/3}c=0$，这恰好在 ABC 本身是按顺时针标记的等边三角形时发生（正如第 (ii) 部分的备注所述）。其他任何三角形 T_1（包括退化（degenerate）三角形）都给出一个真正的等边三角形 T_2。

验算

- 取 $a=0$、$b=1$。则 $k=\rho$，而 $0,1,\rho$ 的形心为 $\frac{1}{3}(1+\rho)=\frac{1}{2}-\frac{\sqrt{3}}{6}\mathrm{i}$。公式给出 $\omega^*/\sqrt{3}=\frac{1}{2}-\frac{1}{2\sqrt{3}}\mathrm{i}$，是同一个数。
- 取 $a=0$、$b=1$、$c=\mathrm{i}$（一个直角三角形），T_2 三个点两两之间的距离都约为 1.1154。以电脑对 10^4 个随机三角形和四边形进行检验，在舍入误差范围内证实了第 (ii) 和第 (iii) 部分的恒等式。

常见错误

- 把 b 绕原点而非绕 A 旋转，写成 $k=b\rho$。

- 用 $e^{+i\pi/3}$ 表示顺时针旋转。顺时针是负方向。
- 在第 (ii) 部分只证明了一个方向。正是因子 $i/\sqrt{3}$ 不为零，才得出“当且仅当”。

STEP 之外

第 (iii) 部分就是拿破仑定理（Napoleon's theorem）。在任意三角形的各边上作等边三角形（这里当 ABC 按逆时针标记时是向外作），它们的形心构成一个等边三角形。复数证明之所以奏效，是因为旋转和形心都是平面的线性映射（linear map）。这正是研究几何中变换群（transformation group）的起点。Needham 的 *Visual Complex Analysis*（《直观复分析》）对复数乘法的这种几何观点有十分精彩的论述。

详细解答

STEP III 2020，第 4 题

纯数学 · 反射矩阵与旋转矩阵

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

开首那个结果，其实就是整道题目的化身。减去 $(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$ 会把一点直接垂直移到平面上，所以减去*两倍*这个量，就会把它移到平面另一边同样距离之处，这正是反射（reflection）。一旦把反射写成 $\mathbf{x}\mapsto\mathbf{x}-2(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$，它的矩阵（matrix）、绕 $\mathbf{n}$ 的半转（half-turn），以及两者的乘积，都只需很少计算便可得出。

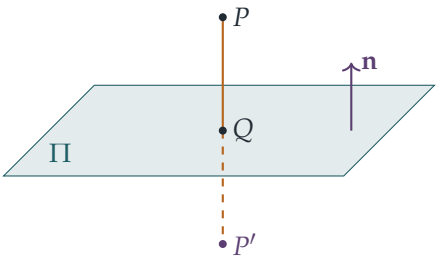
解答

开首的结果

设 $\mathbf{q}=\mathbf{x}-(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$。由于 $\mathbf{n}\cdot\mathbf{n}=1$，

$$\mathbf{q}\cdot\mathbf{n}=\mathbf{x}\cdot\mathbf{n}-(\mathbf{x}\cdot\mathbf{n})(\mathbf{n}\cdot\mathbf{n})=0,$$

所以 Q 位于 Π 上。此外，$\overrightarrow{PQ}=\mathbf{q}-\mathbf{x}=-(\mathbf{x}\cdot\mathbf{n})\mathbf{n}$，是法向量（normal）$\mathbf{n}$ 的倍数。由于 P 不在 Π 上（因此 $\mathbf{x}\cdot\mathbf{n}\neq 0$），这个向量不是零向量。故此 PQ 垂直于 Π，得证。



这幅图也显示了我们的方向：P 在 Π 中的镜像 P' 位于 Q 另一边，与 Q 的距离等于 P 在

Q 上方的距离，所以它的位置向量（position vector）是

$$\mathbf{x} - 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n}.$$

（若 P 在 Π 上，这条公式给出 P 本身，同样正确。）

第 (i) 部分

取 $\mathbf{n} = (a,b,c)^{\mathsf{T}}$，由于 $a^2+b^2+c^2 = 1$，它是单位法向量（unit normal）。对于 $\mathbf{i}$，有 $\mathbf{i}\cdot\mathbf{n} = a$，所以

$$T(\mathbf{i}) = \begin{pmatrix}1\\0\\0\end{pmatrix} - 2a\begin{pmatrix}a\\b\\c\end{pmatrix} = \begin{pmatrix}1-2a^2\\-2ab\\-2ac\end{pmatrix} = \begin{pmatrix}b^2+c^2-a^2\\-2ab\\-2ac\end{pmatrix},$$

其中第一个分量用了 $1 = a^2+b^2+c^2$。这就是所求的像。同理，利用 $\mathbf{j}\cdot\mathbf{n} = b$ 及 $\mathbf{k}\cdot\mathbf{n} = c$，

$$T(\mathbf{j}) = \begin{pmatrix}-2ab\\a^2+c^2-b^2\\-2bc\end{pmatrix}, \qquad T(\mathbf{k}) = \begin{pmatrix}-2ac\\-2bc\\a^2+b^2-c^2\end{pmatrix}.$$

$\mathbf{M}$ 的各行（column）就是 $\mathbf{i},\mathbf{j},\mathbf{k}$ 的像，所以

$$\mathbf{M} = \begin{pmatrix}b^2+c^2-a^2 & -2ab & -2ac\\ -2ab & a^2+c^2-b^2 & -2bc\\ -2ac & -2bc & a^2+b^2-c^2\end{pmatrix} = \mathbf{I} - 2\mathbf{n}\mathbf{n}^{\mathsf{T}}.$$

第 (ii) 部分

把已知矩阵与 $\mathbf{M}$ 对比，并把各对角元素写成 $1-2a^2$、$1-2b^2$、$1-2c^2$：

$$1-2a^2 = 0.64, \quad 1-2b^2 = 0.36, \quad 1-2c^2 = 0 \implies a^2 = 0.18, \quad b^2 = 0.32, \quad c^2 = 0.5.$$

所以 $|a| = \frac{3}{5\sqrt{2}}$、$|b| = \frac{4}{5\sqrt{2}}$ 及 $|c| = \frac{5}{5\sqrt{2}}$。

? **想一想** 对角线只给出 a^2,b^2,c^2。我们怎样确定正负号?

非对角元素给出各乘积：

$$-2ab = 0.48, \quad -2ac = 0.6, \quad -2bc = -0.8 \implies ab = -0.24, \quad ac = -0.3, \quad bc = 0.4.$$

所以 b 与 c 同号，而 a 与它们异号。把 (a,b,c) 换成 $(-a,-b,-c)$ 不会改变任何东西（仍是同一个平面），所以可取 $c>0$，得

$$(a,b,c) = \tfrac{1}{5\sqrt{2}}(-3,\,4,\,5).$$

这些值给出 $ab = -\frac{12}{50}$、$ac = -\frac{15}{50}$ 及 $bc = \frac{20}{50}$，因此全部六个独立元素都吻合。两边乘以 $-5\sqrt{2}$，平面为

$$3x - 4y - 5z = 0.$$

第 (iii) 部分

把任何向量分拆为 $\mathbf{x} = (\mathbf{x}\cdot\mathbf{n})\mathbf{n} + \big(\mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}\big)$，即沿轴的部分加上垂直于轴的部分（由开首的结果可知）。绕轴旋转 π 会保持第一部分不变，并把第二部分反向，所以

$$\mathbf{x} \mapsto (\mathbf{x}\cdot\mathbf{n})\mathbf{n} - \big(\mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}\big) = 2(\mathbf{x}\cdot\mathbf{n})\mathbf{n} - \mathbf{x}.$$

这正是 $-T(\mathbf{x})$。因此 $\mathbf{N} = 2\mathbf{n}\mathbf{n}^{\mathsf{T}} - \mathbf{I} = -\mathbf{M}$，即

$$\mathbf{N} = \begin{pmatrix} a^2 - b^2 - c^2 & 2ab & 2ac \\ 2ab & b^2 - a^2 - c^2 & 2bc \\ 2ac & 2bc & c^2 - a^2 - b^2 \end{pmatrix}.$$

第 (iv) 部分

> ? **想一想** 这里我们需要把两个以 a、b、c 表示的 3×3 矩阵相乘吗?

不需要。由 (iii)，$\mathbf{NM} = -\mathbf{M}^2$。在同一平面中反射两次，会把每一点带回原处，所以 $\mathbf{M}^2 = \mathbf{I}$。（从代数上看，由于 $\mathbf{n}^{\mathsf{T}}\mathbf{n} = 1$，$(\mathbf{I} - 2\mathbf{n}\mathbf{n}^{\mathsf{T}})^2 = \mathbf{I} - 4\mathbf{n}\mathbf{n}^{\mathsf{T}} + 4\mathbf{n}(\mathbf{n}^{\mathsf{T}}\mathbf{n})\mathbf{n}^{\mathsf{T}} = \mathbf{I}$。）因此

$$\mathbf{NM} = -\mathbf{I},$$

它把每个 $\mathbf{x}$ 映射到 $-\mathbf{x}$。这是关于原点的反射，又称为中心反演（central inversion），或以原点为中心、比例因子为 -1 的放大（enlargement）。留意它完全不依赖 a、b 或 c。

验算

- $\mathbf{M}$ 的每一行都应是单位向量。以第一行为例，$(b^2+c^2-a^2)^2 + 4a^2b^2 + 4a^2c^2 = (b^2+c^2+a^2)^2 = 1$。
- 在 (ii) 中，法向量 $(3,-4,-5)$ 应被映射到其负向量。已知矩阵把它映射到 $(1.92 - 1.92 - 3,\ 1.44 - 1.44 + 4,\ 1.8 + 3.2) = (-3,4,5)$。正确。
- 取 $\mathbf{n} = \mathbf{k}$，则 $\mathbf{M} = \operatorname{diag}(1,1,-1)$，即在 $z = 0$ 中的反射，而 $\mathbf{N} = \operatorname{diag}(-1,-1,1)$，即绕 z 轴的半转。两者的乘积显然是 $-\mathbf{I}$。
- 用电脑以 1000 个随机单位向量 $\mathbf{n}$ 验算，确认了 $\mathbf{M}$ 的公式、$\mathbf{N}$ 保持 $\mathbf{n}$ 不变并把垂直向量反向，以及 $\mathbf{NM} = -\mathbf{I}$。(ii) 中的矩阵亦在 10^{-15} 的误差内重现。

常见错误

- 把 $\mathbf{x} - (\mathbf{x}\cdot\mathbf{n})\mathbf{n}$ 当作反射。这其实是垂足（foot of the perpendicular），漏了因子 2。
- 在 (ii) 中，从平方根独立地选取 a, b, c 的正负号。只有非对角元素才能确定它们之间的相对正负号。
- 说 $\mathbf{NM}$ 是“恒等矩阵”。反射的行列式（determinant）是 -1，旋转的行列式是 $+1$，所以乘积的行列式是 -1。在三维中 $\det(-\mathbf{I}) = -1$，两者一致。

STEP 之外

矩阵 $\mathbf{I} - 2\mathbf{n}\mathbf{n}^{\mathsf{T}}$ 称为豪斯霍尔德反射（*Householder reflection*）。数值线性代数经常用它把矩阵化为三角形式，而且 n 维中的每个正交矩阵（orthogonal matrix）都是最多 n 个这类反射的乘积。第 (iv) 部分是这个结构的一个小例子：在 $\mathbb{R}^3$ 中，半转与在垂直平面中的反射复合，得出 $-\mathbf{I}$。关于正交矩阵、投影和反射，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

STEP III 2020，第 5 题

纯数学 · 部分分式

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

开首的恒等式其实就是变了装的有限等比级数（geometric series），而正是它把 $(1-(x/k)^n)/(x-k)$ 化成一个明确的多项式。这样便可得到 $1/(x^n(x-k))$ 完整的部分分式分解（partial-fraction decomposition），而无须进行繁重的系数比较。接着求导，便得出含重复因子（repeated factor）$(x-k)^2$ 的分解；最后的积分只是其中一个特例，当中每一项都可以直接积分。

解答

开首的恒等式

展开右方，并把它拆成两个和：

$$\begin{aligned}(x-y)\sum_{r=1}^{n}x^{n-r}y^{r-1} &= \sum_{r=1}^{n}x^{n-r+1}y^{r-1}-\sum_{r=1}^{n}x^{n-r}y^{r}\\ &= \left(x^n+x^{n-1}y+\cdots+xy^{n-1}\right)-\left(x^{n-1}y+\cdots+xy^{n-1}+y^n\right)\\ &= x^n-y^n,\end{aligned}$$

这是因为除了 x^n（来自第一个和中 $r=1$ 的项）和 y^n（来自第二个和中 $r=n$ 的项）之外，每一项都在两个括号中各出现一次，因而互相抵消。

第 (i)(a) 部分

把 $\mathrm{F}(x)=\dfrac{A}{x-k}+\dfrac{\mathrm{f}(x)}{x^n}$ 两边乘以 $x^n(x-k)$，得

$$1=Ax^n+\mathrm{f}(x)(x-k).$$

这是多项式之间的恒等式，对所有 $x \neq 0, k$ 成立，因此对所有 x 都成立（在无限多个点上相等的两个多项式必然相同）。代入 $x = k$ 得 $1 = Ak^n$，所以 $A = 1/k^n$（这是容许的，因为 $k \neq 0$）。于是

$$\mathrm{f}(x)(x-k) = 1 - \frac{x^n}{k^n}, \qquad \text{所以} \qquad \mathrm{f}(x) = \frac{1}{x-k}\left(1 - \left(\frac{x}{k}\right)^n\right)$$

对 $x \neq k$ 成立，正如所求。

想一想 怎样才能把它化成分母不含 $x-k$ 的形式?

在开首的恒等式中，以 1 代替 x，以 x/k 代替 y:

$$1 - \left(\frac{x}{k}\right)^n = \left(1 - \frac{x}{k}\right)\sum_{r=1}^{n}\left(\frac{x}{k}\right)^{r-1} = -\frac{x-k}{k}\sum_{r=1}^{n}\frac{x^{r-1}}{k^{r-1}}.$$

除以 $x-k$,

$$\mathrm{f}(x) = -\frac{1}{k}\sum_{r=1}^{n}\frac{x^{r-1}}{k^{r-1}}, \qquad \text{所以} \qquad \frac{\mathrm{f}(x)}{x^n} = -\frac{1}{k}\sum_{r=1}^{n}\frac{1}{k^{r-1}x^{n+1-r}}.$$

（这亦证实了 f 确实是多项式。）现在以 $s = n+1-r$ 重新编号：当 r 取遍 $1, \ldots, n$ 时，它亦同样取遍这些值；此时 $r-1 = n-s$，而这个和变成 $\sum_{s=1}^{n} 1/(k^{n-s}x^s)$。因此

$$\mathrm{F}(x) = \frac{1}{k^n(x-k)} - \frac{1}{k}\sum_{r=1}^{n}\frac{1}{k^{n-r}x^r}.$$

第 (i)(b) 部分

由于 $x^n\mathrm{F}(x) = \dfrac{1}{x-k}$，用乘积法则（product rule）对两边求导，得

$$nx^{n-1}\mathrm{F}(x) + x^n\mathrm{F}'(x) = -\frac{1}{(x-k)^2}.$$

除以 $-x^n$（这没有问题，因为 $x \neq 0$）:

$$\frac{1}{x^n(x-k)^2} = -\frac{n}{x}\mathrm{F}(x) - \mathrm{F}'(x).$$

由 (a) 部分,

$$-\frac{n}{x}\mathrm{F}(x) = -\frac{n}{xk^n(x-k)} + \sum_{r=1}^{n}\frac{n}{k^{n+1-r}x^{r+1}},$$

而把 (a) 部分的结果逐项求导，得

$$-\mathrm{F}'(x) = \frac{1}{k^n(x-k)^2} - \frac{1}{k}\sum_{r=1}^{n}\frac{r}{k^{n-r}x^{r+1}} = \frac{1}{k^n(x-k)^2} - \sum_{r=1}^{n}\frac{r}{k^{n+1-r}x^{r+1}}.$$

两式相加，两个和合并后得

$$\frac{1}{x^n(x-k)^2} = \frac{1}{k^n(x-k)^2} - \frac{n}{xk^n(x-k)} + \sum_{r=1}^{n}\frac{n-r}{k^{n+1-r}x^{r+1}},$$

这就是所求的结果。

第 (ii) 部分

在 (b) 部分中取 $n=3$ 及 $k=1$。和式中 $r=3$ 的项为零，于是

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

中间一项还需再拆一次：$\dfrac{1}{x(x-1)}=\dfrac{1}{x-1}-\dfrac{1}{x}$。在 $[2,N]$ 上所有项都连续，且 $x-1>0$，因此在该区间上的一个原函数（antiderivative）是

$$G(x)=-\frac{1}{x-1}-3\ln\left(\frac{x-1}{x}\right)-\frac{2}{x}-\frac{1}{2x^2},$$

而

$$\int_2^N\frac{1}{x^3(x-1)^2}\,\mathrm{d}x=G(N)-G(2).$$

? **想一想** 为甚么必须把两个对数放在一起?

分开来看，$-3\ln(N-1)$ 和 $+3\ln N$ 各自趋向无限大，所以不能逐项取极限。合并后，当 $N\to\infty$ 时 $\ln\big((N-1)/N\big)=\ln\big(1-\frac{1}{N}\big)\to\ln 1=0$，因为 ln 在 1 处连续。$G(N)$ 的其余三项显然都趋向 0，所以 $G(N)\to 0$。

最后，

$$G(2)=-1-3\ln\tfrac{1}{2}-1-\tfrac{1}{8}=3\ln 2-\tfrac{17}{8},$$

所以

$$\lim_{N\to\infty}\int_2^N\frac{1}{x^3(x-1)^2}\,\mathrm{d}x=\frac{17}{8}-3\ln 2.$$

验算

- 由于被积函数为正，答案必须是正数：确实 $\frac{17}{8}-3\ln 2\approx 2.1250-2.0794=0.0456$。
- 答案亦应该很小：被积函数在 $x=2$ 时为 $\frac{1}{8}$，之后大约以 x^{-5} 的速度递减。粗略地以 $\frac{1}{8}(2/x)^5$ 作上界，可知积分最多为 $\int_2^\infty 4x^{-5}\,\mathrm{d}x=\frac{1}{16}$。数值积分亦证实其值为 $0.045558\ldots$，准确至多个小数位。
- 在 (b) 部分中试取 $n=1$。和式为零（它唯一的一项有 $n-r=0$），而右方为 $\dfrac{1}{k(x-k)^2}-\dfrac{1}{xk(x-k)}=\dfrac{x-(x-k)}{xk(x-k)^2}=\dfrac{1}{x(x-k)^2}$，与预期一致。

常见错误

- 把 $1-x/k$ 写成 $-(x-k)/k$ 时弄错正负号。

- 在 (a) 部分忘记把和式重新编号，结果 k 和 x 的幂次与目标不符。
- 把 $\ln\big((x-1)/x\big)$ 拆成两个对数，然后声称积分发散，或者写出“$\infty-\infty=0$”。

STEP 之外

你所建立的，正是对形如 x^n 和 $(x-k)^m$ 的因子的一般部分分式分解；对 x 求导（或对根 k 求导）是产生重复因子项的标准方法。判断无限区间上的积分是否收敛（如第 (ii) 部分），是大学第一门分析课程中瑕积分（improper integrals）的课题。Abbott 的 *Understanding Analysis*（《理解分析》）对这类极限有细致的处理。

详细解答

STEP III 2020，第 6 题

纯数学 · 极坐标曲线与面积

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

三个部分都建基于同一个函数 $\cos\theta + \sqrt{\cos 2\theta}$。$C_2$ 的方程是关于 r 的二次方程（quadratic），其两个根为 $\cos\theta \pm \sqrt{\cos 2\theta}$。因此 C_2 就是 C_1（较大的根）再加上第二条位于内侧的分支（较小的根），而两条分支在判别式（discriminant）$\cos 2\theta$ 等于零之处相接。掌握了这一点，两条分支之间的面积便化为单一个积分，因为 $r_+^2 - r_-^2$ 可以因式分解。

解答

第 (i) 部分

设 $y = \cos x + \sqrt{\cos 2x}$。它恰好在 $\cos 2x \geq 0$ 时为实数，而在给定区间上，这正是整个 $[-\frac{1}{4}\pi, \frac{1}{4}\pi]$。此函数是偶函数（even function），所以图像对称于 y 轴。

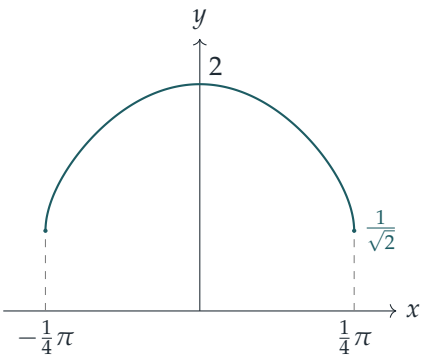
数值。在 $x = 0$ 时，$y = 2$。在 $x = \pm\frac{1}{4}\pi$ 时，$\cos 2x = 0$，而 $y = \cos\frac{1}{4}\pi = \frac{1}{\sqrt{2}}$。

斜率。当 $0 < x < \frac{1}{4}\pi$ 时，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\sin x - \frac{\sin 2x}{\sqrt{\cos 2x}} < 0,$$

因为在该范围内两项皆为负。所以当 x 由 0 增至 $\frac{1}{4}\pi$ 时，y 由 2 递减至 $\frac{1}{\sqrt{2}}$，最大值位于 $(0, 2)$。当 $x \to \frac{1}{4}\pi^-$ 时，$\sqrt{\cos 2x} \to 0^+$ 而 $\sin 2x \to 1$，故 $\mathrm{d}y/\mathrm{d}x \to -\infty$：曲线以铅垂切线（tangent）到达端点。

$x = 0$ 附近。利用 $\cos x \approx 1 - \frac{1}{2}x^2$ 及 $\sqrt{\cos 2x} \approx \sqrt{1 - 2x^2} \approx 1 - x^2$，得 $y \approx 2 - \frac{3}{2}x^2$，是一个平滑的极大点。



第 (ii) 部分

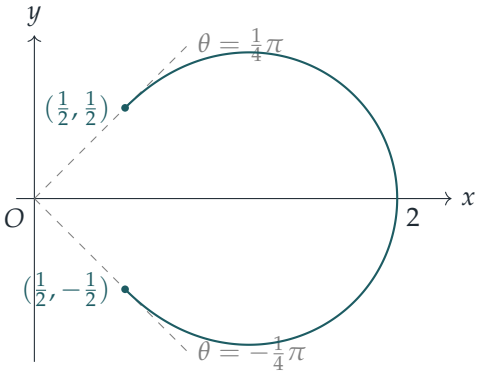
现在同一个函数给出在方向 θ 上与原点的距离 r。C_1 的端点位于 $r = \frac{1}{\sqrt{2}}$、$\theta = \pm\frac{1}{4}\pi$，即直角坐标点 $(\frac{1}{2}, \pm\frac{1}{2})$。在 $\theta = 0$ 时，曲线位于 $(2,0)$，是离 O 最远的点。由于 r 是 θ 的偶函数，C_1 对称于初始线（initial line）。

> **想一想** 对极坐标曲线（polar curve）而言，$\mathrm{d}r/\mathrm{d}\theta$ 为无限大代表甚么？

由第 (i) 部分可知，当 $\theta \to \frac{1}{4}\pi^-$ 时，$\mathrm{d}r/\mathrm{d}\theta \to -\infty$。在该端附近，$r$ 变化很大而 θ 几乎不变，所以曲线是沿着径向直线 $\theta = \frac{1}{4}\pi$（即直线 $y = x$）移动。由于 r 递减，曲线是沿 $y = x$ 朝向原点到达 $(\frac{1}{2}, \frac{1}{2})$ 的。特别地，高度 $y = r\sin\theta$ 必须在到达端点之前取得最大值。事实上

$$\frac{\mathrm{d}y}{\mathrm{d}\theta} = \frac{\mathrm{d}r}{\mathrm{d}\theta}\sin\theta + r\cos\theta$$

在 $\theta = 0$ 附近为正，而当 $\theta \to \frac{1}{4}\pi$ 时趋向 $-\infty$。（最大高度约为 0.81，出现在 $\theta \approx 34°$。）因此 C_1 是一道经过 $(2,0)$ 的拱形，向上鼓出超过 $y = \frac{1}{2}$，再回落至 $(\frac{1}{2}, \frac{1}{2})$，并在该处与直线 $y = x$ 相切；轴下方则是其镜像。



第 (iii) 部分

*两条分支。*把方程视为关于 r 的二次方程：

$$r^2 - 2r\cos\theta + \sin^2\theta = 0 \iff r = \cos\theta \pm \sqrt{\cos^2\theta - \sin^2\theta} = \cos\theta \pm \sqrt{\cos 2\theta}.$$

记 $r_\pm = \cos\theta \pm \sqrt{\cos 2\theta}$。外侧分支 r_+ 正是 C_1。内侧分支满足 $r_- \geq 0$，因为 $\cos^2\theta \geq \cos^2\theta - \sin^2\theta = \cos 2\theta$，且只在 $\sin\theta = 0$ 时取等号。所以 r_- 是一条真正的极坐标曲线，而且只在 $\theta = 0$ 时到达原点。

*在 $\theta = \pm\frac{1}{4}\pi$ 处。*此时 $\cos 2\theta = 0$，所以二次方程有重根（repeated root）：

$$r = \cos\tfrac{1}{4}\pi = \frac{1}{\sqrt{2}}.$$

两条分支都经过 $(\frac{1}{2}, \pm\frac{1}{2})$。

*r 很小时。*若 r 很小，则 r^2 相对于 $2r\cos\theta$ 可以忽略，而 θ 也必须很小，因为 $\sin^2\theta = 2r\cos\theta - r^2$ 很小。于是 $2r\cos\theta \approx \sin^2\theta$，再利用 $\cos\theta \approx 1$、$\sin\theta \approx \theta$，

$$r \approx \frac{\sin^2\theta}{2\cos\theta} \approx \tfrac{1}{2}\theta^2,$$

即为所求。（等价地，$r_- \approx (1 - \frac{1}{2}\theta^2) - (1 - \theta^2) = \frac{1}{2}\theta^2$。）

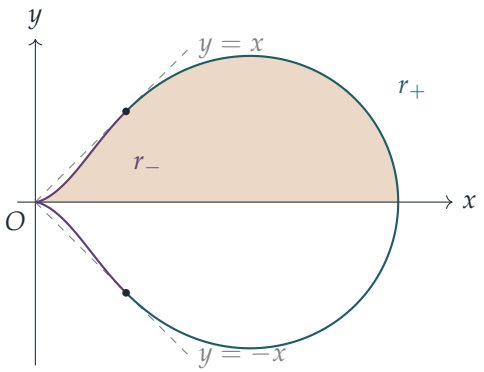
*$r = 0$ 附近的形态。*此时 $x = r\cos\theta \approx \frac{1}{2}\theta^2$，$y = r\sin\theta \approx \frac{1}{2}\theta^3$，所以 $y^2 \approx 2x^3$。因此内侧分支从 $x > 0$ 的一侧、自上方及下方进入原点，并与初始线相切，在 O 处形成一个指向左方的尖点（cusp）。

*$\theta = \pm\frac{1}{4}\pi$ 附近的形态。*对方程进行隐函数微分（implicit differentiation），

$$(2r - 2\cos\theta)\,\mathrm{d}r + (2r\sin\theta + 2\sin\theta\cos\theta)\,\mathrm{d}\theta = 0 \quad\Longrightarrow\quad \frac{\mathrm{d}\theta}{\mathrm{d}r} = \frac{\cos\theta - r}{\sin\theta\,(r + \cos\theta)}.$$

在 $r = \frac{1}{\sqrt{2}}$、$\theta = \frac{1}{4}\pi$ 处，分子为 0 而分母不为零，所以 $\mathrm{d}\theta/\mathrm{d}r = 0$。此外，在整条曲线上 $\theta \leq \frac{1}{4}\pi$。因此 θ 在该处取得最大值，而曲线与径向直线 $y = x$ 相切。外侧分支沿 $y = x$ 朝 O 而来，内侧分支沿 $y = x$ 背离 O 而去，两者在 $(\frac{1}{2}, \frac{1}{2})$ 平滑地相接，并没有角点。由对称性，$(\frac{1}{2}, -\frac{1}{2})$ 处亦然。

所以 C_2 是一条闭曲线，对称于初始线，形如一滴泪珠，其尖端就是 O 处的尖点。它的外缘是 C_1，内缘则由 $(\frac{1}{2}, \frac{1}{2})$ 向下进入 O 处的尖点，再出来到达 $(\frac{1}{2}, -\frac{1}{2})$。



（绘图时使用 $\sin\theta = \frac{1}{\sqrt{2}}\sin u$，这使得 $\sqrt{\cos 2\theta} = \cos u$。让 u 走一整圈，便能描出整条 C_2，而不会出现无限大的斜率。）

*面积。*当 $0 \le \theta \le \frac{1}{4}\pi$ 时，角度为 θ 的射线恰好在 $r_- \le r \le r_+$ 时位于 C_2 之内。所以阴影部分的面积是两个极坐标扇形（sector）之差：

$$\begin{aligned} A &= \int_0^{\pi/4} \tfrac{1}{2}\left(r_+^2 - r_-^2\right) \mathrm{d}\theta = \int_0^{\pi/4} \tfrac{1}{2}(r_+ - r_-)(r_+ + r_-)\,\mathrm{d}\theta \\ &= \int_0^{\pi/4} \tfrac{1}{2}\cdot 2\sqrt{\cos 2\theta}\cdot 2\cos\theta\,\mathrm{d}\theta = 2\int_0^{\pi/4} \cos\theta\sqrt{1-2\sin^2\theta}\,\mathrm{d}\theta. \end{aligned}$$

作代换（substitution）$\sqrt{2}\sin\theta = \sin u$，于是 $\sqrt{2}\cos\theta\,\mathrm{d}\theta = \cos u\,\mathrm{d}u$，而当 θ 由 0 变至 $\frac{1}{4}\pi$ 时，u 由 0 变至 $\frac{1}{2}\pi$：

$$A = 2\int_0^{\pi/2} \cos u\cdot\frac{\cos u}{\sqrt{2}}\,\mathrm{d}u = \sqrt{2}\int_0^{\pi/2}\cos^2 u\,\mathrm{d}u = \sqrt{2}\cdot\frac{\pi}{4} = \frac{\pi}{2\sqrt{2}},$$

即为所求结果。（等价地，代换 $s = \sqrt{2}\sin\theta$ 把积分化为 $\sqrt{2}\int_0^1\sqrt{1-s^2}\,\mathrm{d}s$，即单位圆盘（unit disc）四分之一面积的 $\sqrt{2}$ 倍。）

验算

- 在 $\theta = 0$ 时，$r_\pm = 1 \pm 1$，得出点 $(2,0)$ 和 O，与草图吻合。
- 在 $\theta = 0.01$ 时，$r_- = \cos 0.01 - \sqrt{\cos 0.02} \approx 5.0002\times 10^{-5}$，对比 $\frac{1}{2}\theta^2 = 5\times 10^{-5}$。
- 数值上，$\pi/(2\sqrt{2}) \approx 1.11072$，而直接在 $[0, \frac{1}{4}\pi]$ 上对 $\frac{1}{2}(r_+^2 - r_-^2)$ 作数值积分，得出的值准确至小数点后六位相同。
- 粗略估界：C_2 的上半部位于矩形 $[0,2]\times[0,0.81]$ 之内，其面积约为 1.62；而它包含以 $(\frac{1}{2},0)$、$(2,0)$、$(\frac{1}{2},\frac{1}{2})$ 为顶点的三角形，面积为 0.375。数值 1.11 合理地落在两者之间。

常见错误

- 在 $(\frac{1}{2}, \frac{1}{2})$ 处把 C_1 画成有角点，或画成曲线穿过 $y = x$：$\mathrm{d}r/\mathrm{d}\theta$ 为无限大，表示曲线在该处与径向直线相切。
- 在 O 处画成平滑的圆头而非尖点：$y^2 \approx 2x^3$ 正是典型的尖点形状。
- 只计算 $\int \frac{1}{2} r_+^2 \, \mathrm{d}\theta$，这是直至 C_1 的面积，连内侧分支之内的空心区域也包括在内。

STEP 之外

原点附近的曲线 $y^2 = 2x^3$ 是一条半立方抛物线（semicubical parabola），是尖点的标准例子。每当一个平滑的参数化（parametrisation）在某一时刻速度为零，尖点便会出现：这里 $(x, y) \approx (\frac{1}{2}\theta^2, \frac{1}{2}\theta^3)$ 在 $\theta = 0$ 处的导数为零。为代数曲线上这类奇点分类，正是代数几何（algebraic geometry）的起点。极坐标曲线及其面积亦自然地引向格林定理（Green's theorem），它把一个封闭区域的面积表示为沿其边界的线积分（line integral）。若想从几何角度认识这类曲线，可以参阅 Needham 的 *Visual Complex Analysis*（《直观复分析》）。

详细解答

STEP III 2020，第 7 题

纯数学 · 二阶微分方程

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

第 (i) 部分说明，二阶线性方程（second-order linear equation）有时可以因式分解（factorise）成两个一阶线性方程，而每一个都可以用积分因子（integrating factor）求解。第 (ii) 部分的工作是找出这个分解：比较系数（matching coefficients）会得出一条关于 g 的非线性（non-linear）一阶方程，而一个幂函数形式（power-law）的猜测就能解出它。之后只需依次解两条一阶方程，并在过程中代入初始条件（initial conditions）。

解答

第 (i) 部分

由第二条方程可以把 u 用 y 表示：

$$u = \frac{\mathrm{d}y}{\mathrm{d}x} + \mathrm{g}(x)\, y.$$

对其求导，并对 $\mathrm{g}(x)y$ 使用乘积法则（product rule），得

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{\mathrm{d}^2y}{\mathrm{d}x^2} + \mathrm{g}'(x)\, y + \mathrm{g}(x)\frac{\mathrm{d}y}{\mathrm{d}x}.$$

把两者代入 $\dfrac{\mathrm{d}u}{\mathrm{d}x} + \mathrm{f}(x)u = \mathrm{h}(x)$：

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + \mathrm{g}'(x)\, y + \mathrm{g}(x)\frac{\mathrm{d}y}{\mathrm{d}x} + \mathrm{f}(x)\frac{\mathrm{d}y}{\mathrm{d}x} + \mathrm{f}(x)\mathrm{g}(x)\, y = \mathrm{h}(x),$$

再把含 $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ 的项和含 y 的项分别合并，便正好得到方程 (1)。

第 (ii) 部分

g 的方程。 比较 (2) 与 (1)，我们需要

$$\mathrm{g}(x)+\mathrm{f}(x)=1+\frac{4}{x},\qquad \mathrm{g}'(x)+\mathrm{f}(x)\mathrm{g}(x)=\frac{2}{x}+\frac{2}{x^2},\qquad \mathrm{h}(x)=4x+12.$$

由第一式得 $\mathrm{f}=1+\frac{4}{x}-\mathrm{g}$；代入第二式即可消去 f:

$$\mathrm{g}'(x)+\Big(1+\frac{4}{x}\Big)\mathrm{g}(x)-\big(\mathrm{g}(x)\big)^2=\frac{2}{x}+\frac{2}{x^2}.$$

这就是所求的关于 g 的一阶方程。

幂函数形式的解。 设 $\mathrm{g}(x)=kx^n$，则左边为

$$knx^{n-1}+kx^n+4kx^{n-1}-k^2x^{2n}.$$

? **想一想** 在未知道 k 之前，应如何选取 n?

留意 x 的幂次。右边只含 x^{-1} 和 x^{-2}，所以 x^n、x^{n-1} 和 x^{2n} 这些幂次都只能产生这两项。取 $n=-1$ 正好做到：$x^n=x^{-1}$，而 x^{n-1} 和 x^{2n} 都是 x^{-2}。左边于是变成

$$\frac{k}{x}+\frac{-k+4k-k^2}{x^2}=\frac{k}{x}+\frac{3k-k^2}{x^2}.$$

比较 x^{-1} 项得 $k=2$，此时 $3k-k^2=6-4=2$，与 x^{-2} 项亦吻合。因此

$$n=-1,\quad k=2:\qquad \mathrm{g}(x)=\frac{2}{x},\qquad \mathrm{f}(x)=1+\frac{4}{x}-\frac{2}{x}=1+\frac{2}{x}.$$

解方程 (2)。 根据第 (i) 部分，(2) 现在等价于以下一对方程:

$$\frac{\mathrm{d}u}{\mathrm{d}x}+\Big(1+\frac{2}{x}\Big)u=4x+12,\qquad \frac{\mathrm{d}y}{\mathrm{d}x}+\frac{2}{x}y=u.$$

由于条件是在 $x=1$ 处给出，我们在 $x>0$ 上求解。

*关于 u 的方程。*积分因子为 $\exp\int\big(1+\frac{2}{x}\big)\mathrm{d}x=\mathrm{e}^{x+2\ln x}=x^2\mathrm{e}^x$，因此

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(x^2\mathrm{e}^xu\big)=\big(4x^3+12x^2\big)\,\mathrm{e}^x.$$

右边正好是 $4x^3\mathrm{e}^x$ 的导数（乘积法则），所以

$$x^2\mathrm{e}^xu=4x^3\mathrm{e}^x+C,\qquad u=4x+\frac{C\mathrm{e}^{-x}}{x^2}.$$

u 在 $x=1$ 的值由初始条件得出：$u=y'+\frac{2}{x}y=-3+2\times5=7$。于是 $4+C\mathrm{e}^{-1}=7$，得 $C=3\mathrm{e}$，以及

$$u=4x+\frac{3\mathrm{e}^{1-x}}{x^2}.$$

关于 y 的方程。积分因子为 $\mathrm{e}^{2\ln x} = x^2$，因此

$$\frac{\mathrm{d}}{\mathrm{d}x}(x^2 y) = x^2 u = 4x^3 + 3\mathrm{e}^{1-x}, \qquad \text{因此} \qquad x^2 y = x^4 - 3\mathrm{e}^{1-x} + D.$$

在 $x = 1$ 处：$5 = 1 - 3 + D$，所以 $D = 7$。因此

$$\boxed{y = x^2 + \frac{7 - 3\mathrm{e}^{1-x}}{x^2}}.$$

这里用了 $y(1) = 5$，而求 C 时用了 $y'(1) = -3$，所以两个条件都成立。

验算

- 在 $x = 1$ 处：$y = 1 + 7 - 3 = 5$。求导得 $y' = 2x - 14x^{-3} + 3\mathrm{e}^{1-x}x^{-2} + 6\mathrm{e}^{1-x}x^{-3}$，在 $x = 1$ 处为 $2 - 14 + 3 + 6 = -3$。
- 特解部分（particular part）$y = x^2$ 本身应满足 (2)：$2 + (1 + \frac{4}{x})2x + (\frac{2}{x} + \frac{2}{x^2})x^2 = 2 + 2x + 8 + 2x + 2 = 4x + 12$。确实如此。
- 用电脑把方框中的 y 在 $x = 0.3$ 至 $x = 5$ 之间的许多点代入 (2)，两边在数值微分误差（约 10^{-5}）范围内一致。

常见错误

- 直接把 $y'(1) = -3$ 当作 u 的初始值。u 的初始值是 $y' + \mathrm{g}y = 7$，而不是 -3。
- 在第 (i) 部分漏掉 $\mathrm{g}'(x)$ 项：g 依赖于 x，所以 $\mathrm{g}y$ 必须用乘积法则求导。
- 在两个步骤之一漏掉积分常数。总共有两个常数，正好对应两个初始条件。

STEP 之外

这就是把微分算子（differential operator）因式分解的想法：$(D + \mathrm{f})(D + \mathrm{g})y = \mathrm{h}$，其中 D 表示 $\mathrm{d}/\mathrm{d}x$。由于 g 依赖于 x，这两个因子不可交换（commute），这正是交叉项 g' 出现的原因。我们为 g 找到的方程是一条*黎卡提方程*（Riccati equation），以黎卡提（Riccati）命名。黎卡提方程是非线性的，但只要知道一个解，就足以把线性二阶问题化为两个一阶问题，正如本题所示。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖积分因子、降阶法（reduction of order）及黎卡提方程。

详细解答

STEP III 2020，第 8 题

纯数学 · 数列与递降法

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

本题在 *Beyond STEP* 第 1.8 节 *计数有理数* 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/counting-rationals.html

关键思路

每个下标（subscript）较大的项，都由下标约为其一半的项构成。因此，任何关于数对 (u_n, u_{n+1}) 的命题，都可以推回到下标较小的数对；以*递降法*（descent）论证（等价于选取最小的反例（counterexample）），便能处理每一部分。先写出首几项，并观察每一对相邻项如何产生两对新的项，在证明之前就能看清规律。

解答

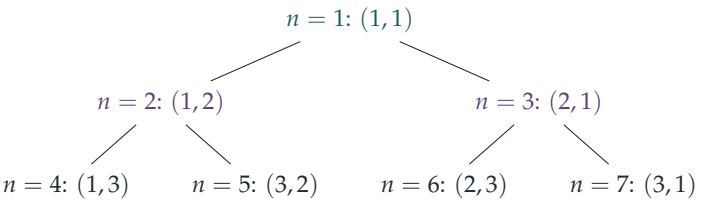
首十六项为

$$1,\ 1,\ 2,\ 1,\ 3,\ 2,\ 3,\ 1,\ 4,\ 3,\ 5,\ 2,\ 5,\ 3,\ 4,\ 1.$$

每一项都是正整数：u_1 显然如此，而之后每一项要么是较前的某一项，要么是两个较前的项之和，所以由（强）归纳法（induction）可得此结论。

值得记下一个数对如何产生后面的数对。若 $(u_k, u_{k+1}) = (a, b)$，则直接由定义得

$$(u_{2k}, u_{2k+1}) = (a,\ a+b), \qquad (u_{2k+1}, u_{2k+2}) = (a+b,\ b). \tag{$*$}$$



位置 n 的数对是 (u_n, u_{n+1})；由 $(*)$，它的左子节点位于 $2n$，右子节点位于 $2n+1$。每个位于 $n \geq 2$ 的数对，都恰好以这种方式由位置 $\lfloor n/2 \rfloor$ 的数对产生。

第 (i) 部分

取一对相邻项 (u_n, u_{n+1})，其中 $n \geq 2$，并设 $k = \lfloor n/2 \rfloor \geq 1$。

- 若 $n = 2k$, 该数对为 $(u_k,\ u_k + u_{k+1})$。由于 $u_{k+1} \geq 1$, 奇数下标的项 $u_{2k+1} = u_k + u_{k+1}$ 大于偶数下标的项 $u_{2k} = u_k$。
- 若 $n = 2k+1$，该数对为 $(u_k + u_{k+1},\ u_{k+1})$。此时由于 $u_k \geq 1$，$u_{2k+1} = u_k + u_{k+1}$ 大于 $u_{2k+2} = u_{k+1}$。

因此，除了第一对 $(u_1, u_2) = (1,1)$ 之外，每一对中奇数下标的项都严格较大。特别地，只有第一对的两项相等。

第 (ii) 部分

假设 $d > 1$ 同时整除 u_n 和 u_{n+1}。由于 $u_1 = u_2 = 1$, 这不可能在 $n = 1$ 时发生, 故 $n \geq 2$; 设 $k = \lfloor n/2 \rfloor$，则 $1 \leq k < n$。

- 若 $n = 2k$：$d \mid u_{2k} = u_k$ 且 $d \mid u_{2k+1} = u_k + u_{k+1}$，所以 d 整除两者之差 u_{k+1}。
- 若 $n = 2k+1$：$d \mid u_{2k+2} = u_{k+1}$ 且 $d \mid u_{2k+1} = u_k + u_{k+1}$，所以 $d \mid u_k$。

无论哪种情况，d 都同时整除 u_k 和 u_{k+1}，即较前的一对相邻项，这正是需要证明的。

现在假设某一对相邻项有公因数 $d > 1$, 并在这样的数对中选取 n 最小的一对 (u_n, u_{n+1})。如上所述 $n \neq 1$，而刚才的论证会给出位于 $k < n$、具有相同公因数的数对，与 n 的最小性矛盾。因此任意两个相邻项都互质（co-prime）。

第 (iii) 部分

假设某个有序数对 (a, b) 在两个不同位置出现，即 $(u_m, u_{m+1}) = (u_n, u_{n+1}) = (a, b)$，其中 $m < n$。在所有这类重复中，选取 m 最小的一个。

若 $m = 1$，则 $a = b = 1$；但 $n \geq 2$，由第 (i) 部分，位置 n 的数对两项不相等。所以

$m \geq 2$，再由第 (i) 部分亦有 $a \neq b$。

想一想 单凭 a 和 b，怎样判断位置是奇数还是偶数?

第 (i) 部分告诉我们，数对中较大的一项具有奇数下标，所以位置的奇偶性（parity）由数对本身决定：

- 若 $a < b$，第二项具有奇数下标，所以 m 和 n 都是偶数：$m = 2j$，$n = 2k$，其中 $j < k$。于是 $u_j = u_{2j} = a$ 且 $u_{j+1} = u_{2j+1} - u_j = b - a$；同理 $(u_k, u_{k+1}) = (a, b-a)$。
- 若 $a > b$，则 m 和 n 都是奇数：$m = 2j+1$，$n = 2k+1$，其中 $1 \leq j < k$。于是 $u_{j+1} = u_{2j+2} = b$ 且 $u_j = u_{2j+1} - u_{j+1} = a - b$；同理 $(u_k, u_{k+1}) = (a-b, b)$。

两种情况下，都有同一个有序数对出现在两个不同位置 j 和 k，而 $j \leq m/2 < m$。这与 m 的选取矛盾。所以没有任何正整数有序数对会在两个不同位置相邻出现。

第 (iv) 部分

设 $a > b$，a、b 互质，并假设 (a, b) 不是数列中的一对相邻项。那么 $a - b$ 和 b 都是正整数，而且互质：$a - b$ 与 b 的任何公因数也整除 $(a-b) + b = a$，所以它是 a 与 b 的公因数，必然是 1。

若在某处 $a - b$ 之后紧接 b，例如 $(u_k, u_{k+1}) = (a-b, b)$，则由 $(*)$

$$(u_{2k+1}, u_{2k+2}) = \big((a-b) + b,\ b\big) = (a, b),$$

与对 (a, b) 的假设矛盾。所以 $(a-b, b)$ 也是一对不以相邻项出现的互质数对，而它的和为 $a < a + b$。

$a < b$ 时的相应结果：*a 和 $b - a$ 是互质的正整数，它们不会以 $b - a$ 紧接 a 的方式相邻出现，而且它们的和 b 小于 $a + b$*。证明相同：a 与 $b - a$ 的公因数整除 $a + (b - a) = b$；而若 $(u_k, u_{k+1}) = (a, b-a)$，则 $(u_{2k}, u_{2k+1}) = (a,\ a + (b-a)) = (a, b)$。

第 (v) 部分

由于每一项都是正整数，$f(n) = u_n / u_{n+1}$ 是正有理数（rational），所以值域（range）包含于正有理数之中。我们证明每个正有理数都会被取到，而且只取到一次。

每个正有理数都在值域之内。把一个正有理数写成最简分数（lowest terms）p/q，其中 p、q 是互质的正整数。我们断言每个互质数对 (p, q) 都会以 (u_n, u_{n+1}) 的形式在某个 n 出现。若不然，在没有出现的互质数对中，选取 $a + b$ 最小的一对 (a, b)。不可能有 $a = b$：此时互质迫使 $a = b = 1$，而 $(1, 1)$ 在 $n = 1$ 出现。所以 $a > b$ 或 $a < b$，而第 (iv) 部分给出一对没有出现、和更小的互质数对。这与最小性矛盾，所以每个互质数对 (p, q) 都会出现，设在位置 n，于是 $f(n) = p/q$。

f 是一对一（*one-to-one*）的。假设 $f(m) = f(n)$。由第 (ii) 部分，u_m/u_{m+1} 和 u_n/u_{n+1} 都是分母为正的最简分数。一个正有理数只有一种这样的表示，所以 $u_m = u_n$ 且 $u_{m+1} = u_{n+1}$。由第 (iii) 部分，同一个有序数对不能在两个不同位置出现，所以 $m = n$。

因此 f 是由正整数到正有理数的双射（bijection），故有反函数（inverse）f^{-1}：给定最简分数 p/q，$f^{-1}(p/q)$ 就是 p 之后紧接 q 的唯一位置。例如 3 只在位置 10 紧接 5，所以 $f^{-1}(\frac{3}{5}) = 10$。

验算

- 所列各项给出数对 $(1,1), (1,2), (2,1), (1,3), (3,2), (2,3), (3,1), (1,4), \ldots$：第一对之后，奇数下标的项都较大；所有数对都互质；没有数对重复。
- 对 3/5 执行第 (iv) 部分的递降：$(3,5) \to (3,2) \to (1,2) \to (1,1)$。再用 $(*)$ 往回推：位置 1 的 $(1,1)$ 给出位置 2 的 $(1,2)$、位置 5 的 $(3,2)$ 和位置 10 的 $(3,5)$；确实 $u_{10} = 3$，$u_{11} = 5$。
- 用电脑检查首二百万项，证实了第 (i)–(iii) 部分，并发现每个满足 $p + q \leq 21$ 的互质数对 (p,q) 都恰好出现一次。

常见错误

- 在第 (ii) 和第 (iii) 部分忘记另行排除第一对 $(1,1)$：它是每次递降的基本情况。
- 在第 (iii) 部分，未加论证便假设两个位置的奇偶性相同；正是第 (i) 部分迫使这一点成立。
- 在第 (v) 部分只证明值域是全体正有理数。要有反函数，还需要 f 是一对一的，而这要同时用到第 (ii) 和第 (iii) 部分。

STEP 之外

这就是 Stern 双原子数列（Stern's diatomic sequence），而列出 u_n/u_{n+1} 便得到有理数的 Calkin–Wilf 枚举（Calkin–Wilf enumeration）：一个 $\mathbb{N}$ 与 $\mathbb{Q}^+$ 之间的显式双射，具体地说明有理数是可数的（countable）。第 (iv) 和第 (v) 部分的递降，正是欧几里得算法（Euclid's algorithm）的减法形式；而树中由 $(1,1)$ 到 (p,q) 的路径，编码了 p/q 的连分数（continued fraction）。Davenport 的 *The Higher Arithmetic*（《高等算术》）对欧几里得算法和连分数有精彩的论述。

详细解答

STEP III 2020，第 9 题

力学 · 刚体平衡

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

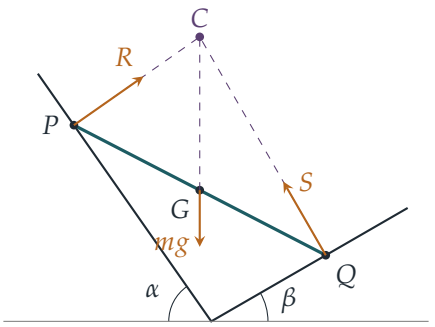
www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

杆只受三个力作用：作用于中点的重力，以及两端各一个接触力。一个物体在三个互不平行的力作用下处于平衡（equilibrium）时，三力的作用线（lines of action）必交于同一点，因此整道题就变成几何问题：用两种不同方法求出该交点的高度，然后令两者相等。在第 (ii) 部分，Q 处的法向反作用力（normal reaction）与极限摩擦力（limiting friction）合成为一个总反作用力，它偏离法线的角度为摩擦角（angle of friction）λ，其中 $\tan\lambda = \mu$；同样的几何论证仍然适用，只需把 β 换成 $\beta + \lambda$。

解答

把杆视为均匀（uniform）的刚体（rigid body），长度为 $2l$，中点为 G，并假设两个平面固定不动。设 Π_1 位于两平面交线的左方，Π_2 位于右方，于是 P 在左方且较高，Q 在右方且较低。



第 (i) 部分

两个平面都是光滑（smooth）的，所以 P 处的反作用力 R 垂直于 Π_1，Q 处的反作用力 S 垂直于 Π_2。垂直于一个倾角为 α 的平面的直线，与铅垂线的夹角为 α；因此 R 的作用线从 P 向上延伸并向 Q 一方倾斜，与向上铅垂线成 α 角；而 S 的作用线从 Q 向上延伸并向 P 一方倾斜，与向上铅垂线成 β 角。

杆在三个力作用下平衡：R、S，以及经过 G 铅垂向下的重力 mg。对 R 与 S 作用线的交点 C 取力矩（moments），R 和 S 的力矩均为零，所以重力的力矩也必须为零：重力的作用线经过 C。因此 C 位于 G 的正上方。

> ? **想一想** 怎样把“C 位于 G 的正上方”转化为一条方程?

沿两条途径计算 C 高出 Q 的高度。由于 G 是中点，而杆与水平线成 θ 角，G 与两端的水平距离均为 $l\cos\theta$，而 P 比 Q 高 $2l\sin\theta$。

- 沿 R 的作用线：沿一条与铅垂线成 α 角的直线移动水平距离 $l\cos\theta$，高度上升 $l\cos\theta\cot\alpha$。所以 C 高出 Q 的高度为 $2l\sin\theta + l\cos\theta\cot\alpha$。
- 沿 S 的作用线：同理，C 高出 Q 的高度为 $l\cos\theta\cot\beta$。

令两者相等，再除以 $l\cos\theta \neq 0$，得

$$2\tan\theta + \cot\alpha = \cot\beta, \qquad \text{即} \qquad 2\tan\theta = \cot\beta - \cot\alpha,$$

此即所求。由于 $\beta < \alpha$（两者皆为锐角），$\cot\beta > \cot\alpha$，所以 $\theta > 0$，与 P 高于 Q 一致。

第 (ii) 部分

现在 Π_1 仍然光滑，所以 P 处的反作用力与之前完全相同。在 Q 处则有法向反作用力 N，以及沿 Π_2 的摩擦力 F。

*摩擦力的方向。*杆是刚体，所以 P 与 Q 的速度沿杆方向的分量相等。若 P 开始以速率 u 沿 Π_1 向下滑动，而 Q 以（带正负号的）速率 w 沿 Π_2 向上移动，则沿杆从 P 指向 Q 的方向（此方向在水平线下方 ϕ 角）分解，得

$$u\cos(\alpha - \phi) = w\cos(\beta + \phi).$$

这里 $\cos(\alpha - \phi) > 0$，而下文会证明 $\phi + \beta < \frac{1}{2}\pi$，所以 $w > 0$：Q 即将沿 Π_2 向上移动。摩擦力阻碍这个运动，所以 F 沿 Π_2 向下作用；又因平衡处于极限状态，$F = \mu N$。

*总反作用力。*把 N 与 F 合成为单一个力 S'。它与法线成 λ 角，其中 $\tan\lambda = F/N = \mu$；由于 F 沿平面向下（指向两平面的交线），它令反作用力进一步向 P 一方倾斜。所以 S' 的作用线从 Q 向上延伸，与向上铅垂线成 $\beta + \lambda$ 角，并向 P 一方倾斜。（若 $\beta + \lambda > \frac{1}{2}\pi$，这条作用线实际上是向下的；下面的高度公式仍然成立，只是余切值为负。）

杆再次在三个力作用下平衡，所以三力的作用线交于 G 正上方的一点，而第 (i) 部分的论证可以逐字照搬，只需把 θ 换成 ϕ，β 换成 $\beta+\lambda$：

$$2\tan\phi = \cot(\beta+\lambda) - \cot\alpha.$$

由此可见 $2\tan\phi < \cot\beta$，所以 $\tan\phi < \cot\beta = \tan(\frac{1}{2}\pi - \beta)$，即 $\phi + \beta < \frac{1}{2}\pi$，这就证实了我们对摩擦力方向的假设。

与 $2\tan\theta = \cot\beta - \cot\alpha$ 相减，得

$$2(\tan\theta - \tan\phi) = \cot\beta - \cot(\beta+\lambda) = \frac{\sin(\beta+\lambda)\cos\beta - \cos(\beta+\lambda)\sin\beta}{\sin\beta\,\sin(\beta+\lambda)}$$
$$= \frac{\sin\lambda}{\sin\beta\,(\sin\beta\cos\lambda + \cos\beta\sin\lambda)}.$$

分子分母同除以 $\cos\lambda$，利用 $\tan\lambda = \mu$，再提出因子 $\cos\beta$：

$$2(\tan\theta - \tan\phi) = \frac{\mu}{\sin\beta\,(\sin\beta + \mu\cos\beta)} = \frac{\mu}{\sin\beta\cos\beta\,(\tan\beta+\mu)} = \frac{2\mu}{(\mu+\tan\beta)\sin 2\beta}.$$

两边除以 2，得

$$\tan\theta - \tan\phi = \frac{\mu}{(\mu+\tan\beta)\sin 2\beta},$$

此即所求结果。

验算

- 当 $\mu = 0$ 时右方为 0，所以 $\phi = \theta$：第 (ii) 部分退化为第 (i) 部分，理应如此。
- 右方为正，所以 $\phi < \theta$：摩擦力令杆可以比在光滑平面上更平地静止，这是合理的，因为 P 正是即将向下滑动的一端。
- 取 $\alpha = 60°$，$\beta = 30°$。则 $2\tan\theta = \sqrt{3} - 1/\sqrt{3} = 2/\sqrt{3}$，所以 $\theta = 30°$。取 $\mu = 0.5$，公式给出 $\tan\phi = 0.5774 - 0.5/(1.0774\times 0.8660) \approx 0.0415$，所以 $\phi \approx 2.4°$。对这些数值直接解水平分解、铅垂分解及对 G 取力矩所得的方程（以电脑验算，并对数千组随机角度及摩擦系数重复进行），得到相同的 ϕ。
- 任何 μ 都可以借助 $2\tan\phi = \cot(\beta+\lambda) - \cot\alpha$ 迅速验算：这里 $\lambda \approx 26.57°$，$\cot 56.57° \approx 0.6606$，而 $(0.6606 - 0.5774)/2 \approx 0.0416$，在舍入误差范围内吻合。

常见错误

- 把摩擦力设为沿 Π_2 向上。当 P 向下滑动时，杆长不变迫使 Q 沿其所在平面向上移动，所以 Q 处的摩擦力沿该平面向下。
- 从水平线而非铅垂线量度反作用力的角度：倾角为 α 的平面的法线与铅垂线成 α 角。

STEP 之外

这里用到的三力共点（concurrency）原理，是刚体平衡一般条件的特例：合力为零，以及对任何一点的合力矩为零。把摩擦力与法向反作用力合成为半顶角为 λ 的“摩擦锥”（cone of friction），正是工程师分析梯子、楔子和螺丝稳定性的方法；而用速度判断摩擦力方向的论证，就是瞬时转动中心（instantaneous centre of rotation）的概念。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）正是以这种精神处理刚体静力学和摩擦。

详细解答

STEP III 2020，第 10 题

力学 · 简谐运动

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

质点悬挂于弹簧下时，会绕其平衡位置（equilibrium position）作简谐运动（simple harmonic motion），由此可用 k、g 和 a 表示 Ω。在平台上时，质点的加速度就是平台的加速度，而平台的加速度可用 x 表示，因此牛顿第二定律（Newton's second law）直接给出反作用力（reaction）。要保持接触，反作用力必须时刻非负；由于它原来是 x 的线性（linear）函数，我们只需在平台行程的两端检查即可，而哪一端是关键，则取决于 ω 比 Ω 小还是大。

解答

把质点视为质点模型（point mass），弹簧视为轻弹簧，且无论伸长或压缩（compression）均遵从胡克定律（Hooke's law）。全文中，作用于质点的力以向上为正。

悬挂的质点。 设 y 为质点在 A 下方的距离，则伸长量（extension）为 $y-a$，张力（tension）为 $kmg(y-a)/a$，向上作用于质点。以向下为正应用牛顿第二定律，得

$$m\ddot{y} = mg - \frac{kmg}{a}(y-a) \qquad \Longrightarrow \qquad \ddot{y} = -\frac{kg}{a}\left(y - a - \frac{a}{k}\right).$$

令 $z = y - a(1+\frac{1}{k})$，上式变为 $\ddot{z} = -\frac{kg}{a}z$，即绕平衡点 $y = a(1+\frac{1}{k})$ 的简谐运动。质点从静止释放时并不在此点（任何不等于 a/k 的伸长量都给出 $z \neq 0$），因此它以角频率（angular frequency）$\sqrt{kg/a}$ 作竖直振动，周期（period）为 $2\pi\sqrt{a/(kg)}$。令此等于 $2\pi/\Omega$，得 $\Omega^2 = kg/a$，即

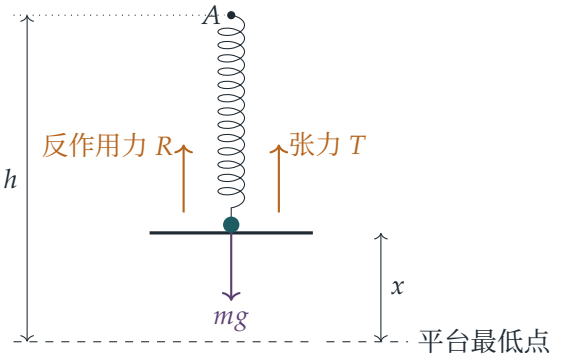
$$kg = a\Omega^2.$$

下文会以 $g/\Omega^2 = a/k$ 的形式使用此式。

平台。 平台作振幅（amplitude）为 b、角频率为 ω 的简谐运动，因此它在其最低点上方的高度 x 的范围为 $0 \le x \le 2b$，中心为 $x = b$。所以其加速度为

$$\ddot{x} = -\omega^2(x-b) = \omega^2(b-x).$$

（例如，$x = b(1-\cos\omega t)$ 给出 $\ddot{x} = \omega^2 b\cos\omega t = \omega^2(b-x)$。）



反作用力。 质点与平台接触期间随平台一起运动，所以它在平台最低点上方的高度为 x，加速度为 $\ddot{x}$。它在 A 下方的距离为 $h-x$，因此弹簧的伸长量为 $h-x-a$，张力为

$$T = \frac{kmg}{a}(h-x-a) = m\Omega^2(h-x-a),$$

这里用了 $kg/a = \Omega^2$。（若 $h-x<a$，弹簧被压缩，T 为负，即推力（thrust），上式依然成立。）作用于质点的力为向上的 R 和 T 以及向下的 mg，所以

$$R + T - mg = m\ddot{x} = m\omega^2(b-x),$$

因此

$$R = mg - m\Omega^2(h-x-a) + m\omega^2(b-x) = mg + m\Omega^2(a+x-h) + m\omega^2(b-x),$$

此即所求。

保持接触。 质点始终留在平台上，当且仅当对 $[0,2b]$ 内每一个 x 都有 $R \ge 0$。把含 x 的项归并，

$$\frac{R}{m} = g + \Omega^2(a-h) + \omega^2 b + (\Omega^2 - \omega^2)\,x.$$

想一想 为甚么只需看 x 的一个值就足够?

R 是 x 的线性函数，所以在区间 $0 \le x \le 2b$ 上，其最小值出现在其中一端：若斜率 $\Omega^2 - \omega^2$ 为正，则在 $x = 0$；若为负，则在 $x = 2b$。平台在每个周期中都会到达两端，所以“$R \ge 0$ 时刻成立”等价于“$R \ge 0$ 在相应的一端成立”。

情况 $\omega < \Omega$。最小值在 $x = 0$（运动的最低点），所以需要

$$g + \Omega^2(a-h) + \omega^2 b \geq 0 \iff h \leq a + \frac{g}{\Omega^2} + \frac{\omega^2 b}{\Omega^2} = a\left(1+\frac{1}{k}\right) + \frac{\omega^2 b}{\Omega^2},$$

这里除以 $\Omega^2 > 0$ 并用了 $g/\Omega^2 = a/k$。这就是所求的不等式。

情况 $\omega > \Omega$。此时最小值在 $x = 2b$（运动的最高点），所以需要

$$g + \Omega^2(a-h) + \omega^2 b + 2b(\Omega^2 - \omega^2) \geq 0 \iff h \leq a\left(1+\frac{1}{k}\right) + b\left(2 - \frac{\omega^2}{\Omega^2}\right).$$

对所有 ω 都成立的界。 假设质点保持接触。

- 若 $\omega < \Omega$, 则 $\omega^2 b/\Omega^2 < b$（因为 $b > 0$), 所以 $h \leq a(1+\frac{1}{k}) + \omega^2 b/\Omega^2 < a(1+\frac{1}{k}) + b$。
- 若 $\omega > \Omega$，则 $2 - \omega^2/\Omega^2 < 1$，所以 $h \leq a(1+\frac{1}{k}) + b(2-\omega^2/\Omega^2) < a(1+\frac{1}{k}) + b$。
- 若 $\omega = \Omega$，R 中含 x 的项互相抵消，$R/m = g + \Omega^2(a-h) + \Omega^2 b$ 为常数；$R \geq 0$ 恰好给出 $h \leq a(1+\frac{1}{k}) + b$。

所以无论哪种情况,

$$h \leq a\left(1+\frac{1}{k}\right) + b$$

都是必要条件，与 ω 的值无关。由于 $h - b$ 是平台运动中心在 A 下方的深度，而 $a(1+\frac{1}{k})$ 是悬挂平衡位置的深度，这个条件的意思是：平台振动的中心点不能低于质点本身的平衡位置。

验算

- 取 $a = 1$、$k = 2$、$g = 9.8$, 以数值方法逐步模拟自由运动, 所得周期给出 $a\Omega^2 \approx 19.600$, 与 $kg = 19.6$ 吻合。
- 若平台静止（$b = 0$），两个界都变成 $h \leq a(1+\frac{1}{k})$：平台不能低于平衡点，否则弹簧会把质点拉离平台。这完全正确。
- 在 $\omega = \Omega$ 时，两种情况的界一致，都给出 $a(1+\frac{1}{k}) + b$，按连续性（continuity）理应如此。

常见错误

- 只检查运动的最高点。对于缓慢的平台（$\omega < \Omega$），质点最危险的位置是最低点，那里弹簧向上拉得最用力。
- 忘记弹簧可以被压缩：张力公式 $kmg(h-x-a)/a$ 无论正负都适用，所以不必另行分情况讨论。

STEP 之外

这其实是一个受迫振子（forced oscillator）：弹簧试图令质点以其固有频率（natural frequency）Ω 振动，而平台则强加频率 ω，反作用力正是调和两者所需的力。大学的振动课程会研究受驱系统、共振（resonance），以及响应如何随 ω/Ω 变化。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对这些概念有细致的论述。

详细解答

STEP III 2020，第 11 题

概率与统计 · 随机变量的函数

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

严格递减(strictly decreasing)函数会把不等式反向;又因为 f 是自己的反函数(inverse),我们可以把它同时作用于 $f(X) \le y$ 的两边,得到一个只关于 X 的事件。期望值(expectation)则由分部积分(integration by parts)求得,这一步把 f' 换成 f。在第 (ii) 部分,Z 与 X 之间的关系对两个变量是对称的,所以 $Z = f(X)$,其中 f 是自反函数(self-inverse function),于是可以套用第 (i) 部分。所求的不等式其实只是“方差(variance)为正”这个事实。

解答

第 (i) 部分

由于 f^{-1} 在整个 $[a,b]$ 上都有定义,f 是 $[a,b]$ 到自身的双射(bijection),所以 Y 的取值在 $[a,b]$ 之内。固定 $y \in [a,b]$。因为 f 严格递减,把它作用于不等式的两边会使不等式反向,而且 $f(f(x)) = x$。因此,对于 $x \in [a,b]$,

$$f(x) \le y \iff f(f(x)) \ge f(y) \iff x \ge f(y).$$

(反方向的蕴涵同理成立,只需把 f 作用于 $x \ge f(y)$。)由于 X 在 $[a,b]$ 上均匀分布(uniform),且 $f(y) \in [a,b]$,

$$P(Y \le y) = P\big(X \ge f(y)\big) = \frac{b - f(y)}{b - a},$$

此即所求。

求导后,Y 的概率密度函数(probability density function)为

$$g(y) = -\frac{f'(y)}{b-a} \quad \text{当} a \le y \le b,$$

其余情况下为零。由于 f 递减，这个函数非负。接着，取 $u = y^2$ 及 $\mathrm{d}v = -\mathrm{f}'(y)\,\mathrm{d}y$ 进行分部积分，

$$\begin{aligned}\mathrm{E}(Y^2) &= \frac{1}{b-a}\int_a^b y^2\big(-\mathrm{f}'(y)\big)\,\mathrm{d}y \\ &= \frac{1}{b-a}\Big[-y^2\mathrm{f}(y)\Big]_a^b + \int_a^b \frac{2y\,\mathrm{f}(y)}{b-a}\,\mathrm{d}y \\ &= \frac{-b^2a + a^2b}{b-a} + \int_a^b \frac{2y\,\mathrm{f}(y)}{b-a}\,\mathrm{d}y,\end{aligned}$$

其中用到 $\mathrm{f}(b) = a$ 及 $\mathrm{f}(a) = b$。第一项等于 $-ab(b-a)/(b-a) = -ab$。把哑变量（dummy variable）y 改记为 x，得

$$\mathrm{E}(Y^2) = -ab + \int_a^b \frac{2x\,\mathrm{f}(x)}{b-a}\,\mathrm{d}x.$$

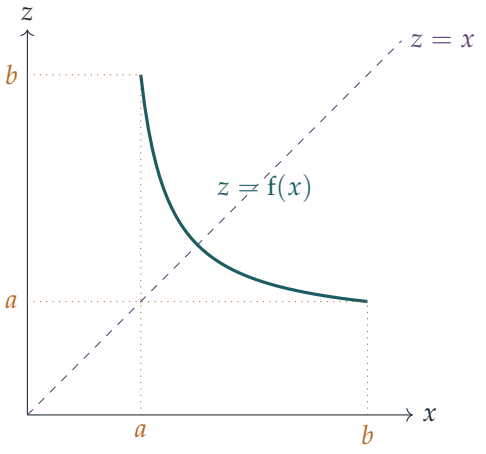
第 (ii) 部分

注意由 $\frac{1}{c} = \frac{1}{a} + \frac{1}{b}$ 得 $c = \dfrac{ab}{a+b}$，又因 $\frac{1}{c} > \frac{1}{a}$，所以 $c < a$。解出 Z：

$$\frac{1}{Z} = \frac{1}{c} - \frac{1}{X} = \frac{X-c}{cX}, \qquad \text{所以} \qquad Z = \mathrm{f}(X)\text{，其中 } \mathrm{f}(x) = \frac{cx}{x-c}.$$

由于 $x \geq a > c$，分母在 $[a,b]$ 上为正。我们验证 f 满足第 (i) 部分的三个条件。

- $\frac{1}{\mathrm{f}(a)} = \frac{1}{c} - \frac{1}{a} = \frac{1}{b}$，所以 $\mathrm{f}(a) = b$。同理 $\frac{1}{\mathrm{f}(b)} = \frac{1}{c} - \frac{1}{b} = \frac{1}{a}$，所以 $\mathrm{f}(b) = a$。
- $\frac{1}{\mathrm{f}(x)} = \frac{1}{c} - \frac{1}{x}$ 在 $[a,b]$ 上为正且严格递增，所以 $\mathrm{f}(x)$ 在该区间上严格递减。
- 关系式 $\frac{1}{z} + \frac{1}{x} = \frac{1}{c}$ 对 x 与 z 是对称的。所以 $z = \mathrm{f}(x)$ 当且仅当 $x = \mathrm{f}(z)$，即 $\mathrm{f} = \mathrm{f}^{-1}$。



f 的图像由 (a,b) 延伸至 (b,a)，并且关于直线 $z = x$ 对称——自反函数理应如此。

想一想 为甚么除了 $\mathrm{E}(Z^2)$ 之外，还要求 $\mathrm{E}(Z)$?

所求的不等式必须来自 $\mathrm{Var}(Z) > 0$，而方差需要用到这两个矩（moment）。对数会在每个矩中出现，其一次项会互相抵消，只剩下平方项。

对两个积分，先做除法：

$$\mathrm{f}(x) = c + \frac{c^2}{x-c}, \qquad x\,\mathrm{f}(x) = \frac{cx^2}{x-c} = c\Big(x + c + \frac{c^2}{x-c}\Big).$$

记 $L = \ln\Big(\dfrac{b-c}{a-c}\Big)$；由于 $b-c > a-c > 0$，L 为正。于是

$$\mathrm{E}(Z) = \frac{1}{b-a}\int_a^b \Big(c + \frac{c^2}{x-c}\Big)\mathrm{d}x = c + \frac{c^2 L}{b-a}.$$

由第 (i) 部分，

$$\begin{aligned}\mathrm{E}(Z^2) &= -ab + \frac{2c}{b-a}\int_a^b \Big(x + c + \frac{c^2}{x-c}\Big)\mathrm{d}x \\ &= -ab + \frac{2c}{b-a}\Big(\frac{b^2-a^2}{2} + c(b-a) + c^2 L\Big) \\ &= -ab + c(a+b) + 2c^2 + \frac{2c^3 L}{b-a}.\end{aligned}$$

由于 $c(a+b) = ab$，首两项互相抵消。减去 $\mathrm{E}(Z)^2 = c^2 + \dfrac{2c^3L}{b-a} + \dfrac{c^4L^2}{(b-a)^2}$ 后，含 L 的一次项抵消，得

$$\mathrm{Var}(Z) = c^2 - \frac{c^4L^2}{(b-a)^2}.$$

Z 并非常数：由第 (i) 部分，其分布函数（distribution function）$\big(b - \mathrm{f}(y)\big)/(b-a)$ 在 $[a,b]$ 上由 0 严格递增至 1。所以 $\mathrm{Var}(Z) > 0$，由此得

$$L^2 < \frac{(b-a)^2}{c^2}.$$

L 与 $(b-a)/c$ 均为正，所以可以两边开平方根：

$$\ln\Big(\frac{b-c}{a-c}\Big) < \frac{b-a}{c}.$$

验算

- 在第 (i) 部分，试用最简单的自反函数例子 $\mathrm{f}(x) = a + b - x$。此时 Y 在 $[a,b]$ 上均匀分布，而公式给出 $\mathrm{E}(Y^2) = -ab + \frac{2}{b-a}\int_a^b x(a+b-x)\,\mathrm{d}x = \frac{1}{3}(a^2+ab+b^2)$，这正是均匀分布随机变量的二阶矩。

- 取 $a=1$、$b=3$，则 $c=\frac{3}{4}$，$L=\ln 9\approx 2.197$，而 $(b-a)/c=\frac{8}{3}\approx 2.667$。不等式成立，且 $\mathrm{Var}(Z)=\frac{9}{16}-\frac{81}{256}\cdot\frac{(\ln 9)^2}{4}\approx 0.180$。用电脑对 f 和 f^2 作数值积分，得到相同的方差。
- 当 $b\to a$ 时，不等式两边都趋于 0，而 $\mathrm{Var}(Z)\to 0$，理应如此：Z 被挤压在 $[a,b]$ 之内。

常见错误

- 忘记递减函数会把不等式反向，因而写成 $\mathrm{P}(Y\le y)=\mathrm{P}(X\le \mathrm{f}(y))$。
- 漏掉密度 $-\mathrm{f}'(y)/(b-a)$ 中的负号。密度不可能是负数。
- 没有说明为何 $\mathrm{Var}(Z)$ 严格为正，或者开平方根时没有指出两边均为正。

STEP 之外

第 (i) 部分是密度变换法则的特例：若 $Y=\mathrm{f}(X)$ 且 f 单调（monotonic），则 $g_Y(y)=g_X\big(\mathrm{f}^{-1}(y)\big)\,\big|(\mathrm{f}^{-1})'(y)\big|$。由 $\mathrm{Var}(Z)>0$ 得出严格不等式，是一个强大技巧的例子：$\mathrm{E}(Z)^2\le \mathrm{E}(Z^2)$ 其实是乔装了的詹森不等式（Jensen's inequality）（或柯西–施瓦茨不等式（Cauchy–Schwarz inequality）），只要巧妙地选取随机变量，就能证明许多经典不等式。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）涵盖了这两个概念。

详细解答

STEP III 2020，第 12 题

概率与统计 · 几何分布的独立性

STEP III 2020，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 2020 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696589-2020-past-papers.zip

关键思路

所有随机变量都是独立几何分布随机变量（geometric random variables）对 (X, Y) 的函数，而且 $\mathrm{P}(X = x,\, Y = y) = p^2 q^{x+y-2}$。处理和与差时，最好沿着 (x, y) 网格中的一条直线把这个联合概率（joint probability）加起来；处理最大值与最小值时，则最容易通过累积概率（cumulative probability），例如 $\mathrm{P}(T \le t)$。至于独立性（independence），可留意某个变量是否有一个取值能完全确定另一个变量。

解答

由于 A 一直掷到第一次出现正面为止，当首 $k-1$ 次都是反面而第 k 次是正面时，X 取值 $k \ge 1$：

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

Y 亦然。两位参与者各自独立掷币，所以 X 与 Y 互相独立，并且

$$\mathrm{P}(X = x,\, Y = y) = p^2 q^{x+y-2} \qquad (x, y \ge 1).$$

有两个事实我们会反复用到：$\mathrm{P}(X > k) = q^k$（首 k 次全是反面），所以 $\mathrm{P}(X \le k) = 1 - q^k$；以及 $1 - q^2 = (1-q)(1+q) = p(1+q)$。

第 (i) 部分

和。S 的取值为 $s \ge 2$。要 $S = s$，需要 $X = x$ 及 $Y = s - x$，其中 $1 \le x \le s-1$。这 $s-1$ 个结果每个的概率都是 $p^2 q^{s-2}$，因此

$$\mathrm{P}(S = s) = (s-1) p^2 q^{s-2}, \qquad s = 2, 3, 4, \ldots$$

最大值。$T \le t$ 当且仅当 $X \le t$ 与 $Y \le t$ 同时成立，所以由独立性得 $\mathrm{P}(T \le t) = (1-q^t)^2$。因此，对 $t \ge 1$，

$$\begin{aligned}\mathrm{P}(T=t) &= \mathrm{P}(T \le t) - \mathrm{P}(T \le t-1) = (1-q^t)^2 - (1-q^{t-1})^2 \\ &= \left[(1-q^t) - (1-q^{t-1})\right]\left[(1-q^t) + (1-q^{t-1})\right] \\ &= (q^{t-1}-q^t)(2-q^{t-1}-q^t) \\ &= pq^{t-1}(2-q^{t-1}-q^t),\end{aligned}$$

这里用了 $q^{t-1}-q^t = q^{t-1}(1-q) = pq^{t-1}$。这就是所求的结果。（当 $t=1$ 时，公式给出 $p(1-q)=p^2$，是正确的。）

第 (ii) 部分

差。要 $U=0$，需要对某个 $k \ge 1$ 有 $X=Y=k$：

$$\mathrm{P}(U=0) = \sum_{k=1}^{\infty} p^2 q^{2k-2} = \frac{p^2}{1-q^2} = \frac{p^2}{p(1+q)} = \frac{p}{1+q}.$$

对 $u \ge 1$，要么 $Y = X+u$，要么 $X = Y+u$；这两种情况互斥，而且由对称性可知概率相等，所以

$$\mathrm{P}(U=u) = 2\sum_{k=1}^{\infty} \mathrm{P}(X=k,\, Y=k+u) = 2\sum_{k=1}^{\infty} p^2 q^{2k+u-2} = \frac{2p^2q^u}{1-q^2} = \frac{2pq^u}{1+q}.$$

最小值。$W > w$ 当且仅当 $X > w$ 与 $Y > w$ 同时成立，所以 $\mathrm{P}(W > w) = q^{2w}$，而对 $w \ge 1$，

$$\mathrm{P}(W=w) = \mathrm{P}(W > w-1) - \mathrm{P}(W > w) = q^{2w-2} - q^{2w} = (1-q^2)\,q^{2(w-1)}.$$

所以 W 本身也服从几何分布，其“成功”概率为 $1-q^2$：A 和 B 各掷一次为一轮，每轮至少出现一次正面的概率是 $1-q^2$。

第 (iii) 部分

若 $S=2$，由于 $X, Y \ge 1$，必然有 $X=Y=1$，于是 $T=1$。所以 $S=2$ 与 $T=3$ 不可能同时发生：

$$\mathrm{P}(S=2 \text{ 且 } T=3) = 0.$$

另一方面，$\mathrm{P}(S=2) = p^2 > 0$，而 $\mathrm{P}(T=3) = pq^2(2-q^2-q^3) > 0$，因为 $0<q<1$ 给出 $q^2+q^3<2$。因此 $\mathrm{P}(S=2)\times\mathrm{P}(T=3) > 0$，两边并不相等。

第 (iv) 部分

U 与 W 互相独立。我们须证明对每个 $u \ge 0$ 及 $w \ge 1$，都有 $\mathrm{P}(U=u, W=w) = \mathrm{P}(U=u)\,\mathrm{P}(W=w)$。知道 U 和 W 两者，就知道较小的值是 w，较大的值是 $w+u$。

想一想 哪些数对 (X,Y) 会给出 $U=u$ 及 $W=w$?

若 $u=0$，唯一的数对是 (w,w)。若 $u\geq 1$，则恰好有两个：$(w,\,w+u)$ 和 $(w+u,\,w)$，每个的概率都是 p^2q^{2w+u-2}。

当 $u=0$，$\mathrm{P}(U=0,\,W=w)=p^2q^{2w-2}$，而

$$\mathrm{P}(U=0)\,\mathrm{P}(W=w)=\frac{p}{1+q}\cdot p(1+q)\,q^{2w-2}=p^2q^{2w-2}.$$

当 $u\geq 1$，$\mathrm{P}(U=u,\,W=w)=2p^2q^{2w+u-2}$，而

$$\mathrm{P}(U=u)\,\mathrm{P}(W=w)=\frac{2pq^u}{1+q}\cdot p(1+q)\,q^{2w-2}=2p^2q^{2w+u-2}.$$

两者在所有情况下都相等，所以 U 与 W 互相独立。

其他各对都不独立。余下的各对是 (S,T)、(S,U)、(S,W)、(T,U) 及 (T,W)。要证明两个变量不独立，只需找到一对取值使得 $\mathrm{P}(A=a,\,B=b)\neq\mathrm{P}(A=a)\,\mathrm{P}(B=b)$ 便足够。

想一想 S 或 T 的哪一个取值能确定其余一切?

事件 $S=2$ 与事件 $T=1$ 是同一个事件 $\{X=Y=1\}$，其概率为 $p^2>0$；在此事件上 $T=1$、$U=0$ 及 $W=1$。由于 $0<p,q<1$，根据第 (i) 及 (ii) 部分的公式，下列清单中 T、U、W 的每个取值都有正概率。所以:

- (S,T)：$\mathrm{P}(S=2,\,T=3)=0\neq\mathrm{P}(S=2)\,\mathrm{P}(T=3)$，见第 (iii) 部分；
- (S,U)：$\mathrm{P}(S=2,\,U=1)=0$，但 $\mathrm{P}(S=2)\,\mathrm{P}(U=1)>0$；
- (S,W)：$\mathrm{P}(S=2,\,W=2)=0$，但 $\mathrm{P}(S=2)\,\mathrm{P}(W=2)>0$；
- (T,U)：$\mathrm{P}(T=1,\,U=1)=0$，但 $\mathrm{P}(T=1)\,\mathrm{P}(U=1)>0$；
- (T,W)：$\mathrm{P}(T=1,\,W=2)=0$，但 $\mathrm{P}(T=1)\,\mathrm{P}(W=2)>0$。

所以在 S、T、U、W 之中，只有 U 与 W 这一对是互相独立的。

验算

- 每个分布的总和都是 1: 例如 $\sum_{t\geq 1}\mathrm{P}(T=t)$ 经裂项相消 (telescoping) 后等于 $\lim(1-q^t)^2=1$，而 $\frac{p}{1+q}+\sum_{u\geq 1}\frac{2pq^u}{1+q}=\frac{p}{1+q}\left(1+\frac{2q}{p}\right)=\frac{p+2q}{1+q}=1$。
- 取 $p=q=\frac{1}{2}$: $\mathrm{P}(T=2)=\frac{1}{4}\left(2-\frac{1}{2}-\frac{1}{4}\right)=\frac{5}{16}$, 与逐一列出三个结果 $(1,2),(2,1),(2,2)$ 的概率 $\frac{1}{8}+\frac{1}{8}+\frac{1}{16}$ 吻合。
- 取 $p=q=\frac{1}{2}$，$\mathrm{P}(U=1,\,W=1)=\mathrm{P}\{(1,2),(2,1)\}=\frac{1}{4}$，而 $\mathrm{P}(U=1)\,\mathrm{P}(W=1)=\frac{1}{3}\times\frac{3}{4}=\frac{1}{4}$，正符合独立性的要求。

- 所有公式以及 U 与 W 的独立性，都已用电脑对 $p = 0.3$、0.55 及 0.8 把联合分布加总而得到验证。

常见错误

- 在 $u \geq 1$ 时漏掉 $\mathrm{P}(U = u)$ 中的因子 2，或在 $u = 0$ 时错误地加上它——此时 $X = Y$ 只算一次。
- 尝试直接把各种情况相加来求 $\mathrm{P}(T = t)$，结果把 $X = Y = t$ 重复计算；用累积概率的方法可避免这个问题。
- 以为检查一对取值就能证明独立：证明独立需要对*所有* u 和 w 成立，而只要有一处不成立便足以说明不独立。

STEP 之外

U 与 W 互相独立，正是无记忆性（memoryless property）的离散形式：一旦第一次正面在时间 W 出现，另一位参与者余下的等待时间就是一个全新的几何分布随机变量，与之前等了多久无关。在连续的情况下，对应的结果是：对独立的指数分布（exponential distribution）随机变量而言，最小值与“最大值和最小值之差”互相独立；这是泊松过程（Poisson process）和排队论（queueing theory）的基础。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）对这些概念有深入的发展。

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

2020 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。

