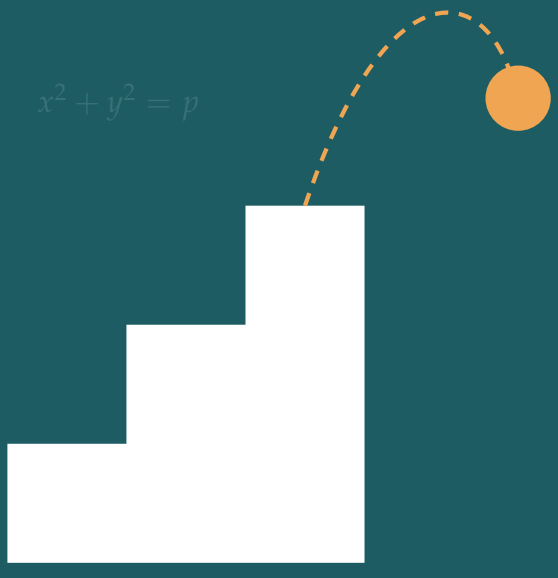
STEP 备试 · 大学数学

# Beyond STEP

STEP II 及 III

2023 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

2023 年详细解答

Keith McNulty

**Beyond STEP：以历届 STEP 试题为门，走进大学数学**

**2023 年详细解答**

2023 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 三角代换 纯数学 . . . 7

第 3 题 多项式与导数 纯数学 . . . 11

第 4 题 消去多项式中的根式 纯数学 . . . 15

第 5 题 递推关系与收敛 纯数学 . . . 19

第 6 题 斐波那契矩阵恒等式 纯数学 . . . 23

第 7 题 两个平方数之和 纯数学 . . . 27

第 8 题 三维向量 纯数学 . . . 31

第 9 题 斜坡上的连接质点 力学 . . . 35

第 10 题 抛体轨道的包络线 力学 . . . 39

第 11 题 混合分布 概率与统计 . . . 44

第 12 题 最大值的分布 概率与统计 . . . 48


## STEP III


第 1 题 圆锥曲线的切线 纯数学 . . . 52

第 2 题 极坐标曲线的面积 纯数学 . . . 56

第 3 题 三次方程的复数根 纯数学 . . . 60

第 4 题 切比雪夫多项式的系数 纯数学 . . . 65

第 5 题 数论 纯数学 . . . 69

第 6 题 双曲函数与单调性 纯数学 . . . 73

第 7 题 积分不等式 纯数学 . . . 77

第 8 题 分段微分方程 纯数学 . . . 81

第 9 题 以绳连接的两质点运动 力学 . . . 86

第 10 题 链的平衡 力学 . . . 90

第 11 章 条件期望值 概率与统计 . . . . . 94
第 12 章 组合概率 概率与统计 . . . . . 98

详细解答

# STEP II 2023，第 1 题

纯数学 · 代换积分法

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

代换（substitution）$x = 1/t$ 把 $\frac{1}{2}$ 和 2 互换，因此它把区间（interval）$[\frac{1}{2}, 2]$ 映射到自身（方向相反）。于是，在这个区间上的积分（integral）可以改写成在同一区间上、外表不同的另一个积分。有时新的被积函数（integrand）可以直接积分；有时它是原被积函数的“搭档”，把两者相加（或互相比较）便能替我们完成工作。

## 解答

### 第 (i) 部分

设 $x = 1/t$，则 $\mathrm{d}x = -\frac{1}{t^2}\,\mathrm{d}t$。由于 $a > 0$，$[a, b]$ 中每个 $x$ 都是正数，对应的 $t$ 亦然。于是

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

这里用到了 $\sqrt{t^2} = t$，因为 $t > 0$。积分限 $x = a$ 和 $x = b$ 变成 $t = a^{-1}$ 和 $t = b^{-1}$。因此

$$\int_a^b \frac{1}{(1 + x^2)^{3/2}}\,\mathrm{d}x = \int_{a^{-1}}^{b^{-1}} \frac{t^3}{(1 + t^2)^{3/2}} \cdot \left(-\frac{1}{t^2}\right)\mathrm{d}t = \int_{a^{-1}}^{b^{-1}} \frac{-t}{(1 + t^2)^{3/2}}\,\mathrm{d}t,$$

得证。

### 第 (ii) 部分

**(a)** 取 $a = \frac{1}{2}$、$b = 2$，新的积分限是 2 和 $\frac{1}{2}$。把上下限对调便可吸收负号：

$$\int_{1/2}^{2} \frac{1}{(1 + x^2)^{3/2}}\,\mathrm{d}x = \int_{1/2}^{2} \frac{t}{(1 + t^2)^{3/2}}\,\mathrm{d}t.$$

新的被积函数是 $1+t^2$ 某个幂的导数（derivative）（的倍数）：由于 $\frac{\mathrm{d}}{\mathrm{d}t}(1+t^2)^{-1/2} = -t(1+t^2)^{-3/2}$，

$$\int_{1/2}^{2} \frac{t}{(1+t^2)^{3/2}}\,\mathrm{d}t = \Big[-(1+t^2)^{-1/2}\Big]_{1/2}^{2} = -\frac{1}{\sqrt{5}} + \frac{1}{\sqrt{5/4}} = -\frac{1}{\sqrt{5}} + \frac{2}{\sqrt{5}} = \frac{1}{\sqrt{5}}.$$

**(b)** 被积函数是偶函数（even function），所以在 $[-2,2]$ 上的积分是在 $[0,2]$ 上积分的两倍。把后者拆成 $[0,\frac{1}{2}]\cup[\frac{1}{2},2]$；第二段就是 (a) 部分。

> **想一想** 第 (i) 部分要求 $a>0$。从 0 开始的那一段该怎样处理?

先在 $[a,\frac{1}{2}]$ 上处理，其中 $0<a<\frac{1}{2}$，这时第 (i) 部分适用；然后令 $a\to 0^+$。在第 (i) 部分取 $b=\frac{1}{2}$，得

$$\int_{a}^{1/2} \frac{1}{(1+x^2)^{3/2}}\,\mathrm{d}x = \int_{2}^{1/a} \frac{t}{(1+t^2)^{3/2}}\,\mathrm{d}t = \frac{1}{\sqrt{5}} - \frac{1}{\sqrt{1+a^{-2}}}.$$

当 $a\to 0^+$ 时，$a^{-2}\to\infty$，最后一项趋于 0。左边趋于在 $[0,\frac{1}{2}]$ 上的积分，因为被积函数在该处连续。所以这一段同样等于 $\frac{1}{\sqrt{5}}$，从而

$$\int_{-2}^{2} \frac{1}{(1+x^2)^{3/2}}\,\mathrm{d}x = 2\Big(\frac{1}{\sqrt{5}} + \frac{1}{\sqrt{5}}\Big) = \frac{4}{\sqrt{5}}.$$

**第 (iii) 部分**

**(a)** 记 $J=\int_{1/2}^{2} \frac{1}{(1+x^2)^2}\,\mathrm{d}x$，再次使用 $x=1/t$。此时 $(1+x^2)^2 = (1+t^2)^2/t^4$，所以

$$J = \int_{2}^{1/2} \frac{t^4}{(1+t^2)^2}\cdot\Big(-\frac{1}{t^2}\Big)\mathrm{d}t = \int_{1/2}^{2} \frac{t^2}{(1+t^2)^2}\,\mathrm{d}t,$$

把哑变量（dummy variable）$t$ 改名为 $x$ 后，这就是第一个等式。把这第二个积分记作 $K$，于是 $J=K$。把两个被积函数相加，

$$J+K = \int_{1/2}^{2} \frac{1+x^2}{(1+x^2)^2}\,\mathrm{d}x = \int_{1/2}^{2} \frac{1}{1+x^2}\,\mathrm{d}x,$$

由于 $J=K$，得 $J=\frac{1}{2}\int_{1/2}^{2} \frac{1}{1+x^2}\,\mathrm{d}x$，等式链便完成了。

因此

$$J = \tfrac{1}{2}(\arctan 2 - \arctan\tfrac{1}{2}).$$

这个结果可用以下两种方法之一化简。由于 $\arctan 2 + \arctan\frac{1}{2} = \frac{\pi}{2}$（两直角边为 1 和 2 的直角三角形中的两个余角（complementary angles））, 得 $J=\frac{\pi}{4} - \arctan\frac{1}{2}$。又或者，

用 tan 的减法公式，$\tan(\arctan 2 - \arctan\frac{1}{2}) = \dfrac{2-\frac{1}{2}}{1+1} = \frac{3}{4}$，而这个差位于 $(0, \frac{\pi}{2})$ 之内，所以

$$\int_{1/2}^{2} \frac{1}{(1+x^2)^2}\,\mathrm{d}x = \tfrac{1}{2}\arctan\tfrac{3}{4} \approx 0.3218.$$

**(b)** 设 $L$ 为所求的积分。在动用难求的原函数(antiderivative)之前,不妨再试一次 $x = 1/t$。被积函数变成

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

这里再次用到 $t > 0$。乘以 $\mathrm{d}x = -\frac{1}{t^2}\,\mathrm{d}t$，并注意积分限由 2 到 $\frac{1}{2}$，

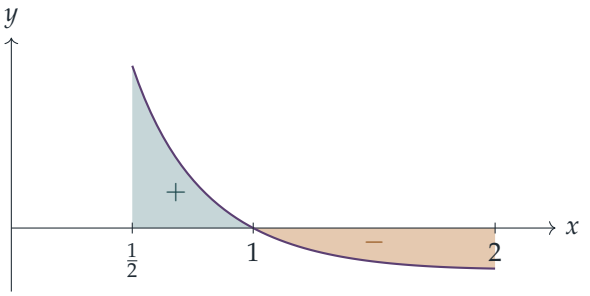
$$L = \int_{2}^{1/2} \frac{-(t-1)}{t\sqrt{1+t^2}}\,\mathrm{d}t = \int_{1/2}^{2} \frac{t-1}{t\sqrt{1+t^2}}\,\mathrm{d}t = -\int_{1/2}^{2} \frac{1-t}{t\sqrt{1+t^2}}\,\mathrm{d}t = -L.$$

**想一想** 代换把 $L$ 变成了 $-L$。这是矛盾吗?

不是：它正好告诉我们积分的值。由 $L = -L$ 得 $2L = 0$，所以

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

图像也吻合：$[\frac{1}{2}, 1]$ 上的正面积恰好抵消 $[1, 2]$ 上的负面积，因为 $x \mapsto 1/x$ 把这两段互换，而且由上面的推导可知，它把其中一段的面积变成另一段面积的相反数。




## 验算

- $(1+x^2)^{-3/2}$ 的一个原函数是 $x(1+x^2)^{-1/2}$（求导即可验证）。用它计算，(ii)(a) 得 $\frac{2}{\sqrt{5}} - \frac{1}{\sqrt{5}} = \frac{1}{\sqrt{5}} \approx 0.4472$，(ii)(b) 得 $\frac{4}{\sqrt{5}} \approx 1.7889$，与我们的答案一致。
- 在 (iii)(a) 中，$\frac{1}{2}\arctan\frac{3}{4} \approx 0.32175$ 与 $\frac{\pi}{4} - \arctan\frac{1}{2} \approx 0.32175$ 相符。对等式链中三个积分进行数值积分（numerical quadrature），都得到相同的值。
- 在 (iii)(b) 中，被积函数在 $x < 1$ 时为正，在 $x > 1$ 时为负，所以答案为零是合理的。数值积分在舍入误差范围内得到 0。

## 常见错误

- 漏掉 $\mathrm{d}x = -t^{-2}\,\mathrm{d}t$ 中的负号，或忘记对调积分上下限会把它抵消。
- 在 (ii)(b) 中直接以 $a = 0$ 套用第 (i) 部分，这需要积分限 $t = 1/0$。应该取极限，或者干脆用原函数。
- 写 $\sqrt{(1+t^2)/t^2} = \sqrt{1+t^2}/t$ 时没有指出这需要 $t > 0$。

## STEP 之外

映射 $x \mapsto 1/x$ 是区间 $[\frac{1}{2}, 2]$ 的一种对称性（symmetry），而利用积分的对称性是一个反复出现的技巧，例如藉 $x \mapsto \frac{\pi}{2} - x$ 计算 $\int_0^{\pi/2} \ln \sin x \,\mathrm{d}x$。第 (ii)(b) 部分让你初尝瑕积分（improper integral）：代换把一个有限的端点送到无穷远，我们必须小心地取极限。Abbott 的 *Understanding Analysis*（《理解分析》）严谨地处理积分的极限，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）则有更多 STEP 风格的积分练习。

详细解答

# STEP II 2023，第 2 题

纯数学 · 三角代换

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

每条等式的右方，其实都是一条熟悉的倍角公式（multiple-angle formula）换了个样子：当 $t = \tan\alpha$ 时，$\frac{2t}{1-t^2}$ 就是 $\tan 2\alpha$，$\frac{3t-t^3}{1-3t^2}$ 就是 $\tan 3\alpha$；而当 $t = \cos\alpha$ 时，$2t^2 - 1$ 就是 $\cos 2\alpha$。作代换（substitution）后，三条互相牵连的方程便化为一条只含一个角的三角方程，既容易求解，更重要的是容易计数。唯一要小心的，是确保这个代换恰好把每个解都捕捉一次。

## 解答

### 第 (i) 部分

每个实数都恰好对应 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 中的一个 $\alpha$，使该实数等于 $\tan\alpha$，所以把 $x$ 写成 $x = \tan\alpha$（其中 $\alpha$ 在该区间内）不会遗失任何解。第一条方程要有意义，须有 $x \neq \pm 1$，即 $\alpha \neq \pm\frac{1}{4}\pi$，因此 $2\alpha$ 不是 $\frac{1}{2}\pi$ 的奇数倍，可以使用二倍角公式（double-angle formula）：

$$y = \frac{2\tan\alpha}{1-\tan^2\alpha} = \tan 2\alpha.$$

对 $y = \tan 2\alpha$ 作完全相同的步骤（由于 $z$ 有定义，故 $y \neq \pm 1$），得 $z = \tan 4\alpha$；再对 $z$ 作同样步骤（由于 $z \neq \pm 1$），得 $x = \tan 8\alpha$。又因 $x = \tan\alpha$，故

$$\tan\alpha = \tan 8\alpha,$$

此即所求。两个角的正切相等，当且仅当两角相差 $\pi$ 的整数倍，所以对某整数 $k$ 有 $8\alpha = \alpha + k\pi$，即

$$\alpha = \frac{k\pi}{7}.$$

条件 $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$ 等价于 $-\frac{7}{2} < k < \frac{7}{2}$，所以 $k = -3, -2, \ldots, 3$。

反过来，这七个值每一个都给出真正的解：$2\alpha = \frac{2k\pi}{7}$ 和 $4\alpha = \frac{4k\pi}{7}$ 永远不会是 $\frac{1}{2}\pi$ 的奇数倍（否则 $4k$ 或 $8k$ 须等于 7 乘一个奇数，于是 $7 \mid k$，迫使 $k = 0$，但这时得到的是 0，并非奇数倍），所以 $x, y, z$ 都不等于 $\pm 1$，每个分母都不为零，上面的推导可以倒过来进行。不同的 $k$ 给出 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 中不同的 $\alpha$，因而给出不同的 $x$。所以恰好有七个解。

要以题目要求的形式写出这些解，可利用 $\tan(\theta - \pi) = \tan\theta$，把 $2\alpha$ 和 $4\alpha$ 减去 $\pi$ 的倍数，化到 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 之内。例如 $\tan\frac{4\pi}{7} = \tan(-\frac{3\pi}{7})$，$\tan\frac{8\pi}{7} = \tan\frac{\pi}{7}$。记 $t_m = \tan\frac{m\pi}{7}$，则解 $(x, y, z)$ 为：

| $k$ | $(x, y, z)$ |
|---|---|
| 0 | $(0, 0, 0)$ |
| 1 | $(t_1,\ t_2,\ t_{-3})$ |
| 2 | $(t_2,\ t_{-3},\ t_1)$ |
| 3 | $(t_3,\ t_{-1},\ t_{-2})$ |

| $k$ | $(x, y, z)$ |
|---|---|
| $-1$ | $(t_{-1},\ t_{-2},\ t_3)$ |
| $-2$ | $(t_{-2},\ t_3,\ t_{-1})$ |
| $-3$ | $(t_{-3},\ t_1,\ t_2)$ |

留意除 $(0, 0, 0)$ 外，这些解正是 $(t_1, t_2, t_{-3})$ 和 $(t_{-1}, t_{-2}, t_3)$ 的循环排列（cyclic rearrangement）。这是必然的，因为方程组在 $x \to y \to z \to x$ 之下保持不变。

### 第 (ii) 部分

先推导三倍角公式（triple-angle formula）：设 $t = \tan\alpha$，

$$\tan 3\alpha = \frac{\tan 2\alpha + \tan\alpha}{1 - \tan 2\alpha \tan\alpha} = \frac{\frac{2t}{1-t^2} + t}{1 - \frac{2t^2}{1-t^2}} = \frac{2t + t - t^3}{1 - t^2 - 2t^2} = \frac{3t - t^3}{1 - 3t^2}.$$

与前相同，设 $x = \tan\alpha$，其中 $-\frac{1}{2}\pi < \alpha < \frac{1}{2}\pi$。方程要求 $1 - 3x^2$、$1 - 3y^2$、$1 - 3z^2$ 皆不为零；在此前提下，公式依次给出 $y = \tan 3\alpha$、$z = \tan 9\alpha$ 和 $x = \tan 27\alpha$。于是 $\tan\alpha = \tan 27\alpha$，得 $26\alpha = k\pi$，即

$$\alpha = \frac{k\pi}{26}, \qquad -13 < k < 13.$$

这是 $\alpha$ 的 25 个候选值，各自给出不同的 $x$ 值。

**想一想** 这 25 个候选值当中，会不会有些令某个分母为零，因而并非真正的解？

这正是需要检查的地方。分母 $1 - 3\tan^2\beta$ 在 $\beta \equiv \pm\frac{1}{6}\pi \pmod{\pi}$ 时为零，而当 $\beta$ 是 $\frac{1}{2}\pi$ 的奇数倍时，正切本身没有定义。这里涉及的角是 $\beta = 3^j \cdot \frac{k\pi}{26}$，其中 $j = 0, 1, 2$。

- 若 $3^j k\pi/26 = \pm\frac{1}{6}\pi + m\pi$，则 $3^{j+1}k = \pm 13 + 78m$。左方是 3 的倍数，右方却不是（因为 78 是 3 的倍数而 13 不是）。不可能。
- 若 $3^j k\pi/26 = \frac{1}{2}(2m+1)\pi$，则 $3^j k = 13(2m+1)$，所以 13 整除 $k$，迫使 $k = 0$；但这时左方为 0，并非 13 的奇数倍。不可能。

所以每个候选值都是真正的解，方程组恰好有 25 个实数解。（区间所排除的 $k = \pm 13$ 对应 $\alpha = \pm\frac{1}{2}\pi$，此时 $\tan\alpha$ 不存在。）

**第 (iii)(a) 部分**

现在 $|x| \le 1$，所以可以对恰好一个 $\alpha \in [0, \pi]$ 写 $x = \cos\alpha$。二倍角公式 $\cos 2\alpha = 2\cos^2\alpha - 1$ 给出 $y = \cos 2\alpha$，继而 $z = \cos 4\alpha$ 及 $x = \cos 8\alpha$；这些自动满足 $|y| \le 1$ 和 $|z| \le 1$。因此三者皆在 $[-1, 1]$ 内的解，恰好对应满足

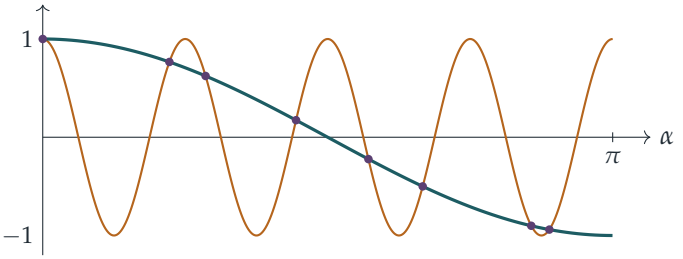
$$\cos 8\alpha = \cos\alpha.$$

的 $\alpha \in [0, \pi]$。两个角的余弦相等，当且仅当两角在相差 $2\pi$ 的整数倍的意义下相等或互为相反数，所以

$$8\alpha = \alpha + 2m\pi \quad \text{或} \quad 8\alpha = -\alpha + 2n\pi, \qquad \text{即} \qquad \alpha = \frac{2m\pi}{7} \quad \text{或} \quad \alpha = \frac{2n\pi}{9}.$$

在 $[0, \pi]$ 内，第一族给出 $m = 0, 1, 2, 3$，即 $0, \frac{2\pi}{7}, \frac{4\pi}{7}, \frac{6\pi}{7}$；第二族给出 $n = 0, 1, 2, 3, 4$，即 $0, \frac{2\pi}{9}, \frac{4\pi}{9}, \frac{6\pi}{9}, \frac{8\pi}{9}$。唯一重叠的是 $\alpha = 0$（公共值须满足 $\frac{m}{7} = \frac{n}{9}$，即 $9m = 7n$，迫使 $7 \mid m$）。于是剩下 $4 + 5 - 1 = 8$ 个不同的 $\alpha$ 值；由于余弦在 $[0, \pi]$ 上是一一对应（one-to-one）的，它们给出 8 个不同的 $x$ 值，每个 $x$ 决定 $y$ 和 $z$。所以满足 $|x|, |y|, |z| \le 1$ 的解恰好有 8 个。

下图显示在 $[0, \pi]$ 上 $\cos 8\alpha$（琥珀色）与 $\cos\alpha$（青绿色）的八个交点。




**第 (iii)(b) 部分**

现在取消限制。消去 $y$ 和 $z$：$y = 2x^2 - 1$ 是 $x$ 的二次式，所以 $z = 2y^2 - 1$ 是 $x$ 的 4 次多项式（polynomial），于是 $x = 2z^2 - 1$ 变成

$$x = 2\Big(2\big(2x^2 - 1\big)^2 - 1\Big)^2 - 1,$$

这是 $x$ 的 8 次多项式方程（右方的首项（leading term）是 $2 \cdot (2 \cdot 2^2)^2 x^8 = 128x^8$，没有任何项能把它抵消）。展开后为

$$128x^8 - 256x^6 + 160x^4 - 32x^2 - x + 1 = 0.$$

方程组的每个解 $(x,y,z)$ 中，$x$ 都是这个八次多项式（octic）的根，而 $x$ 决定 $y$ 和 $z$。8 次多项式最多有 8 个实根，而 (a) 部分已经找到 8 个不同的实根，即该部分求得的八个 $\alpha$ 值所对应的 $\cos\alpha$。所以这些就是全部的根，方程组的每个解都是 (a) 中八个解之一。特别地，每个解都满足 $|x|\le 1$、$|y|\le 1$、$|z|\le 1$，这正是要证明的。

> ? **想一想** 为甚么 (b) 需要用到次数，而不能单靠 (a) 的八个解?

(a) 部分只找出 $[-1,1]$ 内的解；单凭它，对例如 $x=5$ 的情况甚么也说不出来。排除其他一切可能的，是次数的上界：八次多项式最多只有八个根，而 (a) 已经把它们用尽。

## 验算

- 在 (i) 中取 $k=1$：$\tan\frac{\pi}{7}\approx 0.4816$，而 $\frac{2(0.4816)}{1-0.4816^2}\approx 1.2540=\tan\frac{2\pi}{7}$。再做两步便回到 0.4816。
- 以电脑在实数线上搜索 $f(f(f(x)))=x$ 的根，对 $f(t)=\frac{2t}{1-t^2}$ 恰好找到 7 个根，对 $f(t)=\frac{3t-t^3}{1-3t^2}$ 恰好找到 25 个，后者与 $\tan\frac{k\pi}{26}$（$|k|\le 12$）吻合。
- 在 (iii) 中，$x=1$ 给出 $y=z=1$，即 $\alpha=0$ 的解；而确实 $128-256+160-32-1+1=0$。在 $[-50,50]$ 上搜索该八次多项式的根，恰好找到 8 个，全部在 $[-1,1]$ 内。

## 常见错误

- 在 (ii) 中容许 $k=\pm 13$（或完全不限制 $\alpha$），得出 27 个解；$\alpha=\pm\frac{1}{2}\pi$ 并不对应任何实数 $x$。
- 在 (iii)(a) 中漏掉第二族 $8\alpha=-\alpha+2n\pi$，或把 $\alpha=0$ 计算了两次。
- 在 (i) 中没有化简 $\tan\frac{4\pi}{7}$ 之类的答案，而题目要求角度须在 $(-\frac{1}{2}\pi,\frac{1}{2}\pi)$ 之内。

## STEP 之外

满足 $T_n(\cos\alpha)=\cos n\alpha$ 的多项式 $T_n$ 称为切比雪夫多项式（Chebyshev polynomials）；第 (iii) 部分就是方程 $T_8(x)=x$，而上面的八次多项式就是 $T_8(x)-x$。反复迭代像 $x\mapsto 2x^2-1$ 这样的映射，并追问哪些点在三步后回到原处，正是动力系统（dynamical systems）理论的起点：这个映射在 $[-1,1]$ 上是“混沌”（chaotic）的，而角度代换把它变成简单的倍增 $\alpha\mapsto 2\alpha$。同样的技巧也解释了为何 (i) 和 (ii) 中的正切映射有那么多周期点（periodic points）。Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多同类题目。

详细解答

# STEP II 2023，第 3 题

纯数学 · 多项式与导数

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

对 $q$ 求导（differentiate），只会把和式中的每一项向后移一位，而最后一个导数为零，所以 $q' = q - p$。整道题都由这一个关系推动。之后三个证明 $q > 0$ 的方法，分别从三个角度解读这个关系：作为关于驻点（stationary point）的事实；作为 $\mathrm{e}^{-x}q(x)$ 的导数（即积分因子（integrating factor））；以及作为反复使用分部积分法（integration by parts）所得的积分公式。

## 解答

以下一律记 $p(x) = ax^n + \ldots$，其中 $a \neq 0$。对于 $k \geq 1$，$p^{(k)}$ 的次数小于 $n$（或为零），而 $p^{(n+1)} \equiv 0$。

### 第 (i) 部分

**(a)** 奇数次多项式（polynomial）在实数线的一端趋向 $+\infty$，在另一端趋向 $-\infty$。由于它是连续的，根据介值定理（intermediate value theorem），它必在某处取值 0。但对所有 $x$ 都有 $p(x) > 0$，所以 $p$ 没有实根，因此 $n$ 必为偶数。

此外 $a > 0$：若 $a < 0$，则当 $x \to \infty$ 时 $p(x) \to -\infty$，与 $p > 0$ 矛盾。在 $q = p + p' + \cdots + p^{(n)}$ 中，第一项以后的每一项次数都小于 $n$，所以 $q$ 的次数亦为 $n$，而首项系数（leading coefficient）为 $a > 0$。因此当 $x \to \infty$ 时 $q(x) \to +\infty$（而由于 $n$ 是偶数，当 $x \to -\infty$ 时亦然），所以 $q$ 必定取到正值。（若 $n = 0$，则 $q = p$ 是正的常数。）

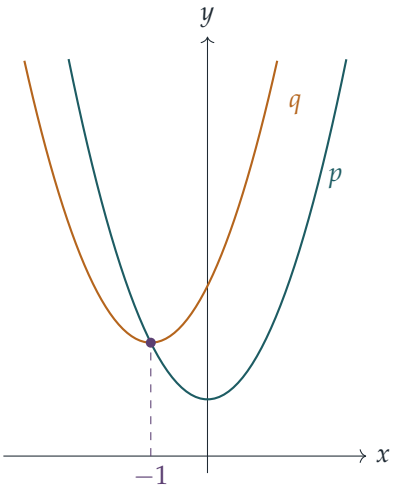
**(b)** 逐项求导，

$$q'(x) = \sum_{k=0}^{n} p^{(k+1)}(x) = \sum_{k=1}^{n+1} p^{(k)}(x) = \sum_{k=0}^{n} p^{(k)}(x) - p(x) + p^{(n+1)}(x) = q(x) - p(x),$$

因为对于 $n$ 次多项式，$p^{(n+1)} \equiv 0$。

### 第 (ii) 部分

**(a)** 在 $y = q(x)$ 的驻点 $x_0$ 处，有 $q'(x_0) = 0$，所以由 (i)(b) 得 $q(x_0) = p(x_0)$：两条曲线在该处相交。




（图中所示为 $p(x) = x^2 + 1$，此时 $q(x) = x^2 + 2x + 3$。$q$ 的最小值在 $x = -1$ 处，正正是两条曲线相交之处。）

现在证明 $q > 0$。若 $n = 0$，则 $q = p > 0$，所以设 $n \geq 2$。由 (i)(a)，当 $x \to \pm\infty$ 时 $q(x) \to +\infty$，所以可以选取 $R > 0$，使得只要 $|x| > R$，便有 $q(x) > q(0)$。

**想一想** 我们怎样知道 $q$ 确实有最低点，而且该点是驻点?

在闭区间 $[-R, R]$ 上，连续函数 $q$ 在某点 $x_0$ 取得最小值，且 $q(x_0) \leq q(0)$。在区间以外，$q$ 大于 $q(0)$，所以 $q(x_0)$ 是 $q$ 在整条实数线上的最小值。由于 $x_0$ 是 $\mathbb{R}$ 的内点（interior point），且 $q$ 可导（differentiable），故 $q'(x_0) = 0$。由第一个结果，$q(x_0) = p(x_0) > 0$。所以对每个 $x$，

$$q(x) \geq q(x_0) = p(x_0) > 0.$$

**(b)** 利用 (i)(b)，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathrm{e}^{-x} q(x)\right) = \mathrm{e}^{-x}\left(q'(x) - q(x)\right) = -\mathrm{e}^{-x} p(x) < 0$$

对所有 $x$ 成立，因为 $p>0$。所以 $\mathrm{e}^{-x}q(x)$ 是（严格）递减的。

至于 $q>0$：由于 $q$ 是多项式，当 $X\to\infty$ 时 $\mathrm{e}^{-X}q(X)\to 0$（指数函数增长得比幂函数快）。递减函数永远不会低于它在 $+\infty$ 的极限，所以对所有 $X$ 都有 $\mathrm{e}^{-X}q(X)\geq 0$。于是对任意 $x$，严格递减给出

$$\mathrm{e}^{-x}q(x)>\mathrm{e}^{-(x+1)}q(x+1)\geq 0,$$

再乘以 $\mathrm{e}^{x}>0$，便得 $q(x)>0$。

**(c)** 这里所有积分都收敛（converge），因为 $t$ 的多项式乘以 $\mathrm{e}^{-t}$ 在 $[0,\infty)$ 上可积。取 $u=p(x+t)$ 及 $\mathrm{d}v=\mathrm{e}^{-t}\,\mathrm{d}t$ 作分部积分（把 $x$ 视为常数）：

$$\begin{aligned}\int_0^\infty p(x+t)\mathrm{e}^{-t}\,\mathrm{d}t&=\Big[-p(x+t)\mathrm{e}^{-t}\Big]_0^\infty+\int_0^\infty p^{(1)}(x+t)\mathrm{e}^{-t}\,\mathrm{d}t\\&=p(x)+\int_0^\infty p^{(1)}(x+t)\mathrm{e}^{-t}\,\mathrm{d}t,\end{aligned}$$

因为当 $t\to\infty$ 时 $p(x+t)\mathrm{e}^{-t}\to 0$。这就是第一个结果。

上述计算除了 $p$ 是多项式之外，没有用到 $p$ 的任何其他性质，所以同样的结果对每一个导数都成立：记

$$I_k(x)=\int_0^\infty p^{(k)}(x+t)\mathrm{e}^{-t}\,\mathrm{d}t,\qquad\text{我们有}\qquad I_k(x)=p^{(k)}(x)+I_{k+1}(x)$$

对于 $k=0,1,\ldots,n$ 成立。此外 $I_{n+1}(x)=0$，因为 $p^{(n+1)}\equiv 0$。反复应用这个关系，

$$I_0(x)=p(x)+I_1(x)=p(x)+p^{(1)}(x)+I_2(x)=\cdots=\sum_{k=0}^{n}p^{(k)}(x)+I_{n+1}(x)=q(x),$$

这就是第二个结果。

> ? **想一想** 为甚么这条公式令 $q(x)>0$ 变得显而易见?

对每个固定的 $x$，被积函数（integrand）$p(x+t)\mathrm{e}^{-t}$ 对每个 $t\geq 0$ 都连续且严格为正，因为 $p>0$ 处处成立。连续且严格为正的函数在长度为正的区间上的积分严格为正。所以对所有 $x$，$q(x)=I_0(x)>0$。

## 验算

- 取 $p(x)=x^2+1$，得 $q(x)=x^2+2x+3=(x+1)^2+2$。确实 $q'=2x+2=q-p$，$q$ 的最小值为 2，在 $x=-1$ 处取得，而 $p(-1)=2$，正如 (ii)(a) 所预测。
- 对同一个 $p$，在 $x=0$ 处：$\int_0^\infty(t^2+1)\mathrm{e}^{-t}\,\mathrm{d}t=2!+1=3=q(0)$。
- 我们用电脑对 300 个次数不超过 6 的随机处处为正多项式作验证：$q'=q-p$ 准确成立；在 $[-20,20]$ 的网格上 $q>0$ 而 $\mathrm{e}^{-x}q(x)$ 递减；以及数值积分重现 $q(x)$ 至约十位有效数字。

## 常见错误

- 在 (ii)(a) 中，没有给出理由便假设最小值存在。这需要在两端都有 $q \to +\infty$，而这用到 (i)(a) 中的 $n$ 为偶数及 $a > 0$。
- 在 (ii)(b) 中，只由极限得出 $\mathrm{e}^{-x}q(x) \geq 0$ 便停下。正是严格递减把这个结论提升为 $q(x) > 0$。
- 忘记了边界项 $p(x+t)\mathrm{e}^{-t}$ 必须在 $t \to \infty$ 时趋于零，才可以把它舍去。

## STEP 之外

(ii)(b) 中的函数 $\mathrm{e}^{-x}$ 是一个积分因子，而 (ii)(c) 则明确地解出 $q' - q = -p$。公式 $\int_0^\infty p(x+t)\mathrm{e}^{-t}\,\mathrm{d}t$ 是一种拉普拉斯变换（Laplace transform）式的积分，而恒等式 $q = (1 + D + D^2 + \cdots)p = (1 - D)^{-1}p$（其中 $D$ 表示求导）则让人初尝线性微分方程（linear differential equations）的算子方法（operator methods）。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖积分因子及拉普拉斯变换；Abbott 的 *Understanding Analysis*（《理解分析》）则把 (ii)(a) 中的极值论证严格化。

详细解答

# STEP II 2023，第 4 题

纯数学 · 消去多项式中的根式

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

本题在 *Beyond STEP* 第 6.3 节以 $\sqrt{2}+\sqrt{3}+\sqrt{5}$ 为根的多项式中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch06/field-extensions.html

## 关键思路

要从方程中消去根式（surd），就把所有含该根式的项移到一边，其余的项移到另一边，然后平方（或立方）。每一部分用的都是同一招：把数写成“一个好处理的数加上一个根式”，把该根式单独分离出来，再升幂直至它消失为止。如果展开后仍有根式残留，就把含它的项集中到一边，再重复一次。

## 解答

### 第 (i) 部分

展开 $(x-\sqrt{2})^2=3$，得 $x^2-2\sqrt{2}\,x+2=3$，即

$$x^2-1=2\sqrt{2}\,x.$$

现在根式已单独在一边，于是两边平方：

$$x^4-2x^2+1=8x^2 \quad\Longrightarrow\quad x^4-10x^2+1=0,$$

得证。（平方只是单向推导，但这正是我们所需要的：每个满足第一条方程的 $x$ 都满足第二条方程。）

至于推论部分，取 $x=\sqrt{2}+\sqrt{3}$。那么 $x-\sqrt{2}=\sqrt{3}$，所以 $(x-\sqrt{2})^2=3$，而由刚才所证，$\mathrm{f}(\sqrt{2}+\sqrt{3})=0$。

**第 (ii) 部分**

设 $x = \sqrt{2}+\sqrt{3}+\sqrt{5}$ 及 $y = x-\sqrt{5} = \sqrt{2}+\sqrt{3}$。由第 (i) 部分，$y^4 - 10y^2 + 1 = 0$。现在把它改写成以 $x$ 表示。首先，

$$y^2 = (x-\sqrt{5})^2 = (x^2+5) - 2\sqrt{5}\,x,$$

因此

$$\begin{aligned} y^4 &= (x^2+5)^2 - 4\sqrt{5}\,x(x^2+5) + 20x^2 \\ &= x^4 + 30x^2 + 25 - 4\sqrt{5}\,x^3 - 20\sqrt{5}\,x. \end{aligned}$$

所以

$$\begin{aligned} 0 &= y^4 - 10y^2 + 1 \\ &= \left(x^4 + 30x^2 + 25 - 4\sqrt{5}\,x^3 - 20\sqrt{5}\,x\right) - 10\left(x^2 + 5 - 2\sqrt{5}\,x\right) + 1 \\ &= x^4 + 20x^2 - 24 - 4\sqrt{5}\,x^3. \end{aligned}$$

$\sqrt{5}\,x$ 的项恰好互相抵消，相当幸运。现在只剩下一个含根式的项，于是把它分离出来再平方：

$$x^4 + 20x^2 - 24 = 4\sqrt{5}\,x^3 \quad\Longrightarrow\quad (x^4 + 20x^2 - 24)^2 = 80x^6.$$

展开左边，

$$\begin{aligned} (x^4 + 20x^2 - 24)^2 &= x^8 + 40x^6 + 400x^4 - 48x^4 - 960x^2 + 576 \\ &= x^8 + 40x^6 + 352x^4 - 960x^2 + 576. \end{aligned}$$

减去 $80x^6$ 后，可取

$$\mathrm{g}(x) = x^8 - 40x^6 + 352x^4 - 960x^2 + 576,$$

它的次数（degree）为 8，具有整数系数（integer coefficients），并且 $\mathrm{g}(\sqrt{2}+\sqrt{3}+\sqrt{5}) = 0$。

> **想一想** 为甚么我们先利用第 (i) 部分处理 $\sqrt{2}+\sqrt{3}$，最后才消去 $\sqrt{5}$，而不是从头开始？

第 (i) 部分已经消去了三个根式中的两个，所以我们只需再处理一个。以平方消去一个根式会使次数加倍，由 4 变成 8，这正好是题目所要求的次数。若从头开始，就要同时应付三个根式。

**第 (iii) 部分**

设 $x = a+\sqrt{2}$，则 $a = x-\sqrt{2}$，且 $a^3 - 3a + 1 = 0$。展开，

$$(x-\sqrt{2})^3 = x^3 - 3\sqrt{2}\,x^2 + 6x - 2\sqrt{2},$$

所以

$$0 = (x-\sqrt{2})^3 - 3(x-\sqrt{2}) + 1 = x^3 + 3x + 1 - \sqrt{2}\,(3x^2 - 1).$$

把根式分离出来再平方：

$$x^3 + 3x + 1 = \sqrt{2}\,(3x^2-1) \quad \Longrightarrow \quad (x^3+3x+1)^2 = 2(3x^2-1)^2.$$

两边分别为

$$\begin{aligned}(x^3+3x+1)^2 &= x^6 + 6x^4 + 2x^3 + 9x^2 + 6x + 1,\\ 2(3x^2-1)^2 &= 18x^4 - 12x^2 + 2,\end{aligned}$$

所以可取

$$\mathrm{h}(x) = x^6 - 12x^4 + 2x^3 + 21x^2 + 6x - 1.$$

我们只用到 $a$ 满足 $t^3 - 3t + 1 = 0$ 这一事实，所以同样的推导亦适用于 $b$ 和 $c$。因此 $\mathrm{h}(a+\sqrt{2}) = \mathrm{h}(b+\sqrt{2}) = \mathrm{h}(c+\sqrt{2}) = 0$。

### 第 (iv) 部分

平方无法消去立方根（cube root），所以我们需要另一个恒等式（identity）。设 $p = \sqrt[3]{2}$、$q = \sqrt[3]{3}$ 及 $x = p + q$。由二项式展开（binomial expansion）得

$$x^3 = p^3 + q^3 + 3pq(p+q) = 5 + 3pq\,x.$$

? **想一想** $pq$ 仍然是无理数（irrational），怎样才能把它消去?

这里 $pq = \sqrt[3]{6}$ 是唯一剩下的无理量，而且它只以因子的形式出现。我们把它分离出来再立方，并利用 $(pq)^3 = 6$：

$$x^3 - 5 = 3pq\,x \quad \Longrightarrow \quad (x^3-5)^3 = 27\cdot 6\,x^3 = 162x^3.$$

由于 $(x^3-5)^3 = x^9 - 15x^6 + 75x^3 - 125$，可取

$$\mathrm{k}(x) = x^9 - 15x^6 - 87x^3 - 125,$$

它具有整数系数，且 $\mathrm{k}(\sqrt[3]{2} + \sqrt[3]{3}) = 0$。

## 验算

- 在第 (ii) 部分，$\sqrt{2}+\sqrt{3}+\sqrt{5} \approx 5.38233$。由 $x^2 \approx 28.9695$，得 $x^8 \approx 704\,310$ 及 $40x^6 \approx 972\,485$。再者 $352x^4 \approx 295\,410$ 及 $960x^2 \approx 27\,811$，而 g 的各项在舍入误差范围内互相抵消。

- 改变 $\sqrt{2}$、$\sqrt{3}$ 和 $\sqrt{5}$ 的正负号都不会改变 $\mathrm{g}(x)$，所以全部八个数 $\pm\sqrt{2}\pm\sqrt{3}\pm\sqrt{5}$ 都是根（root）。这解释了为何次数是 8。同理，h 有六个根 $a\pm\sqrt{2}$、$b\pm\sqrt{2}$、$c\pm\sqrt{2}$。
- $t^3-3t+1=0$ 的根是 $2\cos 40^\circ$、$2\cos 160^\circ$ 和 $2\cos 280^\circ$。例如 $a\approx 1.53209$ 给出 $a+\sqrt{2}\approx 2.94630$，而 $\mathrm{h}(2.94630)$ 在舍入误差范围内等于零。
- 在第 (iv) 部分，$\sqrt[3]{2}+\sqrt[3]{3}\approx 2.70217$，而 $\mathrm{k}(2.70217)$ 在舍入误差范围内等于零。以电脑验算，确认四个多项式（polynomial）的结果均准确至约 $10^{-10}$。

## 常见错误

- 未把根式分离出来就平方。若两边都留有 $\sqrt{5}$ 的项，它们便不会抵消。
- 在第 (iv) 部分尝试平方，或者把 $x$ 立方却没有留意到 $x^3=5+3pqx$。展开式 $(p+q)^3=p^3+q^3+3pq(p+q)$ 就是整个窍门所在。
- 在第 (iii) 部分展开三次式时弄错正负号，尤其是常数项 $-2\sqrt{2}+3\sqrt{2}=\sqrt{2}$。

## STEP 之外

若一个数是某个整数系数多项式的根，就称它为*代数数*（algebraic）。这道题以人手计算的方式说明了代数数之和仍是代数数。一般情况的证明要用到线性代数（linear algebra）：$\mathbb{Q}(\alpha,\beta)$ 是 $\mathbb{Q}$ 上的有限维向量空间（finite-dimensional vector space）。因此幂 $1,\alpha+\beta,(\alpha+\beta)^2,\ldots$ 最终必定线性相关（linearly dependent）。这里出现的次数 4、8、6 和 9，正是这些*域扩张*（field extensions）的维数。想进一步了解，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

# STEP II 2023，第 5 题

纯数学 · 递推关系与收敛

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

本题在 *Beyond STEP* 第 2.7 节六位还是六百位中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/newton-speed.html

## 关键思路

两条递推关系（recurrence）都是用来逼近 $\sqrt{2}$ 的“机器”，而题目关心的是误差缩小得有多快。每一部分的窍门，都在于用恰当的方式量度误差：在 (i) 中考虑 $x_n^2 - 2$，它每一步都会乘以一个大小不超过 $\frac{1}{4}$ 的因子（并且改变正负号）；在 (ii) 中考虑 $y_n - \sqrt{2}$，它每一步（大致上）会被平方。乘以一个常数只带来稳定的进展；平方则每次令正确的位数加倍，这正是 (ii) 能在同样十步之内达到 $10^{-600}$ 的原因。

## 解答

### 第 (i) 部分

**(a)** 我们可以把递推关系改写为

$$x_{n+1} = \frac{(x_n + 1) + 1}{x_n + 1} = 1 + \frac{1}{x_n + 1}.$$

已知 $x_0 = 1 \geq 1$。若 $x_n \geq 1$，则 $x_n + 1 > 0$，故 $\frac{1}{x_n+1} > 0$，从而 $x_{n+1} > 1$。由归纳法（induction），对所有 $n$ 都有 $x_n \geq 1$。

**(b)** 把各项通分至同一分母（common denominator）：

$$\begin{aligned} x_{n+1}^2 - 2 &= \frac{(x_n+2)^2 - 2(x_n+1)^2}{(x_n+1)^2} \\ &= \frac{x_n^2 + 4x_n + 4 - 2x_n^2 - 4x_n - 2}{(x_n+1)^2} \\ &= \frac{-(x_n^2-2)}{(x_n+1)^2}. \end{aligned}$$

分母是正数，所以 $x_{n+1}^2 - 2$ 与 $x_n^2 - 2$ 的正负号相反。（两者都永不为零：由归纳法，每个 $x_n$ 都是有理数（rational），而 $\sqrt{2}$ 是无理数（irrational）。）

取绝对值（modulus），并利用 (a) 得出的 $x_n + 1 \geq 2$，从而 $(x_n+1)^2 \geq 4$，

$$|x_{n+1}^2 - 2| = \frac{|x_n^2 - 2|}{(x_n+1)^2} \leq \tfrac{1}{4}|x_n^2 - 2|,$$

此即所求。

**(c)** 由于 $x_0^2 - 2 = -1$，把 (b) 应用十次，得

$$|x_{10}^2 - 2| \leq \left(\tfrac{1}{4}\right)^{10} |x_0^2 - 2| = \frac{1}{4^{10}} = \frac{1}{1\,048\,576} < 10^{-6}.$$

至于正负号：$x_0^2 - 2 < 0$，而正负号每一步都会翻转，所以对每个偶数 $n$，$x_n^2 - 2$ 都是负数。特别地，$x_{10}^2 - 2 < 0$，因此

$$-10^{-6} < x_{10}^2 - 2 < 0, \qquad \text{即} \qquad 2 - 10^{-6} \leq x_{10}^2 \leq 2.$$

**第 (ii) 部分**

**(a)** 首先，每个 $y_n$ 都是正数：$y_0 = 1 > 0$，而若 $y_n > 0$，则 $y_{n+1} = \frac{y_n^2+2}{2y_n} > 0$。所以递推关系永远不会出现除以零的情况。现在

$$y_{n+1} - \sqrt{2} = \frac{y_n^2 + 2 - 2\sqrt{2}\,y_n}{2y_n} = \frac{(y_n - \sqrt{2})^2}{2y_n},$$

因为 $(y_n - \sqrt{2})^2 = y_n^2 - 2\sqrt{2}\,y_n + 2$。这就是所求的恒等式（identity）。

至于推论部分：$y_0 = 1$。对 $n \geq 1$，把恒等式应用于 $n-1$，可见 $y_n - \sqrt{2}$ 是一个平方数除以正数 $2y_{n-1}$，所以 $y_n - \sqrt{2} \geq 0$，因此 $y_n \geq \sqrt{2} > 1$。故对所有 $n \geq 0$ 都有 $y_n \geq 1$。

**(b)** 记 $e_n = y_n - \sqrt{2}$ 及 $r = \dfrac{\sqrt{2}-1}{2}$。由 (a)，对 $n \geq 1$ 有 $e_n \geq 0$；又因为 $y_n \geq 1$，

$$e_{n+1} = \frac{e_n^2}{2y_n} \leq \frac{e_n^2}{2}.$$

我们用归纳法证明对 $n \geq 1$ 有 $e_n \leq 2r^{2^n}$。

基本情况（*base case*）$n = 1$。$y_1 = \frac{3}{2}$，所以 $e_1 = \frac{3}{2} - \sqrt{2}$，而

$$2r^2 = 2 \cdot \frac{(\sqrt{2}-1)^2}{4} = \frac{3-2\sqrt{2}}{2} = \tfrac{3}{2} - \sqrt{2}.$$

所以上界在此取等号。

归纳步骤（*inductive step*）。假设对某个 $n \geq 1$ 有 $0 \leq e_n \leq 2r^{2^n}$。

**想一想** 为甚么我们可以把 $e_n \leq 2r^{2^n}$ 的两边平方?

对非负数之间的不等式取平方，不等式仍然成立，而 $e_n \geq 0$ 正是 (a) 部分给我们的结果。（这也是上界由 $n = 1$ 而非 $n = 0$ 开始的原因：$e_0 = 1 - \sqrt{2}$ 是负数。）所以

$$e_{n+1} \leq \frac{e_n^2}{2} \leq \frac{\left(2r^{2^n}\right)^2}{2} = \frac{4r^{2^{n+1}}}{2} = 2r^{2^{n+1}},$$

归纳法至此完成。

**(c)** 下界 $y_{10} \geq \sqrt{2}$ 已在 (a) 证明。至于上界，由 $\sqrt{2} - 1 < \frac{1}{2}$ 得 $0 < r < \frac{1}{4}$，所以由 (b)

$$y_{10} - \sqrt{2} \leq 2r^{2^{10}} = 2r^{1024} < 2\left(\tfrac{1}{4}\right)^{1024} = 2 \cdot 2^{-2048} = 2^{-2047}.$$

要把它与 10 的幂比较，可利用 $2^{10} = 1024 > 10^3$：

$$2^{2047} > 2^{2040} = \left(2^{10}\right)^{204} > \left(10^3\right)^{204} = 10^{612} > 10^{600}.$$

因此 $y_{10} - \sqrt{2} < 2^{-2047} < 10^{-600}$，综合而言

$$\sqrt{2} \leq y_{10} \leq \sqrt{2} + 10^{-600}.$$

**想一想** (ii) 中惊人的速度从何而来?

在 (i) 中，误差每次最多乘以 $\frac{1}{4}$，所以十步只能多得大约六个小数位。在 (ii) 中，误差每次（大致上）被平方，所以指数 $2^n$ 不断加倍：十步就能把误差由大约 $\frac{1}{10}$ 降至低于 $10^{-600}$。

## 验算

- 首几项：$x_1 = \frac{3}{2}$，$x_2 = \frac{7}{5}$，$x_3 = \frac{17}{12}$，而 $x_n^2 - 2$ 依次为 $-1$，$\frac{1}{4}$，$-\frac{1}{25}$，$\frac{1}{144}$：正负交替，而且每次至少缩小 4 倍。
- 用分数精确计算，得 $x_{10}^2 \approx 2 - 3.0 \times 10^{-8}$，稳稳落在区间 $[2 - 10^{-6}, 2]$ 之内。
- $y_2 = \frac{17}{12}$，所以 $y_2 - \sqrt{2} \approx 0.00245$，对比上界 $2r^4 \approx 0.00368$。用电脑以精确分数验算，得 $y_{10} - \sqrt{2} \approx 10^{-783}$，远低于所保证的 $2^{-2047} \approx 10^{-616}$。

## 常见错误

- 在 (i)(c) 中忘记考虑正负号：单靠绝对值的上界只能得出 $2-10^{-6} \le x_{10}^2 \le 2+10^{-6}$；上界 2 需要用到正负号交替这一点。
- 未先确定 $e_n \ge 0$ 便把 $e_n \le 2r^{2^n}$ 平方。
- 停在 $2^{-2047}$，而没有真正把它与 $10^{-600}$ 比较；用不等式 $2^{10} > 10^3$ 是最简洁的做法。

## STEP 之外

(ii) 中的递推关系，就是把牛顿（Newton）法（Newton's method）应用于 $f(y) = y^2 - 2$；而 (ii)(a) 中的恒等式，正是在说牛顿法在单根（simple root）附近二次收敛（quadratic convergence），而 (i) 则是收敛率不超过 $\frac{1}{4}$ 的线性收敛（linear convergence）的例子。数 $x_n$ 亦是 $\sqrt{2}$ 的连分数（continued fraction）的渐近分数（convergents）。收敛阶（order of convergence）和不动点迭代（fixed-point iteration）都是大学第一个数值分析（numerical analysis）课程的标准课题，而背后的分析理论在 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）中有仔细的阐述。

详细解答

# STEP II 2023，第 6 题

纯数学 · 斐波那契矩阵恒等式

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

只要把斐波那契数（Fibonacci numbers）看成同一个 $2 \times 2$ 矩阵（matrix）各次幂的元素，关于它们的每一条恒等式（identity）就变成关于 $\mathbf{Q}$ 的幂的恒等式，而普通的代数运算（指数律（index laws）、行列式（determinant）、二项式定理（binomial theorem））便能完成所有工作。每一部分都依循同一个方法：找出同一个矩阵的两种写法，然后读出其中一个元素。几乎每次都是右上角的元素，而它正是 $F_n$ 本身。

## 解答

### 矩阵公式

当 $n = 1$ 时，左边是 $\begin{pmatrix} F_2 & F_1 \\ F_1 & F_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \mathbf{Q}$，因为 $F_2 = F_1 + F_0 = 1$。所以结果在 $n = 1$ 时成立。

假设结果对某个 $n \geq 1$ 成立。那么

$$\mathbf{Q}^{n+1} = \mathbf{Q}^n\mathbf{Q} = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} + F_n & F_{n+1} \\ F_n + F_{n-1} & F_n \end{pmatrix} = \begin{pmatrix} F_{n+2} & F_{n+1} \\ F_{n+1} & F_n \end{pmatrix},$$

这里用了两次递推关系（recurrence）。这正是把公式中的 $n$ 换成 $n+1$ 的结果，所以由归纳法（induction），公式对所有正整数 $n$ 成立。

**第 (i) 部分**

行列式具有乘法性，所以 $\det(\mathbf{Q}^n) = (\det \mathbf{Q})^n = (0-1)^n = (-1)^n$。从元素直接计算同一个行列式，

$$\det(\mathbf{Q}^n) = F_{n+1}F_{n-1} - F_n^2,$$

因此 $F_{n+1}F_{n-1} - F_n^2 = (-1)^n$，得证。（这就是卡西尼恒等式（Cassini's identity）。）

**第 (ii) 部分**

由指数律，$\mathbf{Q}^{m+n} = \mathbf{Q}^m\mathbf{Q}^n$：

$$\begin{pmatrix} F_{m+n+1} & F_{m+n} \\ F_{m+n} & F_{m+n-1} \end{pmatrix} = \begin{pmatrix} F_{m+1} & F_m \\ F_m & F_{m-1} \end{pmatrix}\begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}.$$

乘积的右上角元素是 $\mathbf{Q}^m$ 的第一行乘以 $\mathbf{Q}^n$ 的第二列，即 $F_{m+1}F_n + F_mF_{n-1}$。比较右上角元素，得

$$F_{m+n} = F_{m+1}F_n + F_mF_{n-1}.$$

**第 (iii) 部分**

直接计算，

$$\mathbf{Q}^2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \mathbf{I} + \mathbf{Q}.$$

（这是 $x^2 = 1 + x$ 的矩阵版本，而这条方程正是黄金比例（golden ratio）所满足的方程。）

在以下各部分中，注意对每个 $k \geq 1$，$\mathbf{Q}^k$ 的右上角元素都是 $F_k$；对 $k = 0$ 亦然，因为 $\mathbf{Q}^0 = \mathbf{I}$ 的右上角元素是 $0 = F_0$。取矩阵之和的右上角元素是线性（linear）运算，所以 $\sum c_k\mathbf{Q}^k$ 的右上角元素是 $\sum c_kF_k$。

**(a)** 利用 $\mathbf{Q}^2 = \mathbf{I} + \mathbf{Q}$ 及取 $\mathbf{A} = \mathbf{Q}$ 的二项式定理，

$$\mathbf{Q}^{2n} = (\mathbf{Q}^2)^n = (\mathbf{I} + \mathbf{Q})^n = \sum_{k=0}^{n} \binom{n}{k}\mathbf{Q}^k.$$

左边的右上角元素是 $F_{2n}$，右边的则是 $\sum_k \binom{n}{k}F_k$。因此 $F_{2n} = \sum_{k=0}^{n} \binom{n}{k}F_k$。

**(b)** 先用同样的方法求 $\mathbf{Q}^3$：

$$\mathbf{Q}^3 = \mathbf{Q}\,\mathbf{Q}^2 = \mathbf{Q}(\mathbf{I} + \mathbf{Q}) = \mathbf{Q} + \mathbf{Q}^2 = \mathbf{Q} + \mathbf{I} + \mathbf{Q} = \mathbf{I} + 2\mathbf{Q}.$$

然后应用取 $\mathbf{A} = 2\mathbf{Q}$ 的二项式定理：

$$\mathbf{Q}^{3n} = (\mathbf{I} + 2\mathbf{Q})^n = \sum_{k=0}^{n} \binom{n}{k}2^k\mathbf{Q}^k,$$

比较右上角元素，得 $F_{3n} = \sum_{k=0}^{n} \binom{n}{k} 2^k F_k$。

至于第二条恒等式，写成 $\mathbf{Q}^{3n} = \mathbf{Q}^n \mathbf{Q}^{2n}$，并利用 (a) 的展开式：

$$\mathbf{Q}^{3n} = \mathbf{Q}^n \sum_{k=0}^{n} \binom{n}{k} \mathbf{Q}^k = \sum_{k=0}^{n} \binom{n}{k} \mathbf{Q}^{n+k}.$$

$\mathbf{Q}^{n+k}$ 的右上角元素是 $F_{n+k}$，所以 $F_{3n} = \sum_{k=0}^{n} \binom{n}{k} F_{n+k}$。

**(c)**

? **想一想** 应该把哪个矩阵自乘 $n$ 次，才能产生符号 $(-1)^{n+k}$?

展开 $(\mathbf{Q} - \mathbf{I})^n$ 时，就会出现符号 $(-1)^{n+k} = (-1)^{n-k}$。为了配合我们可以使用的二项式定理的形式，写成 $\mathbf{Q} - \mathbf{I} = -(\mathbf{I} + (-\mathbf{Q}))$，于是

$$(\mathbf{Q} - \mathbf{I})^n = (-1)^n (\mathbf{I} + (-\mathbf{Q}))^n = (-1)^n \sum_{k=0}^{n} \binom{n}{k} (-1)^k \mathbf{Q}^k = \sum_{k=0}^{n} (-1)^{n+k} \binom{n}{k} \mathbf{Q}^k.$$

两边乘以 $\mathbf{Q}^n$，

$$\sum_{k=0}^{n} (-1)^{n+k} \binom{n}{k} \mathbf{Q}^{n+k} = \mathbf{Q}^n (\mathbf{Q} - \mathbf{I})^n = \big(\mathbf{Q}(\mathbf{Q} - \mathbf{I})\big)^n,$$

最后一步之所以成立，是因为 $\mathbf{Q}$ 与 $\mathbf{Q} - \mathbf{I}$ 可交换（commute）（两者都是 $\mathbf{Q}$ 的多项式（polynomial）），所以各因子可以重新组合。但由 (iii)，$\mathbf{Q}(\mathbf{Q} - \mathbf{I}) = \mathbf{Q}^2 - \mathbf{Q} = \mathbf{I}$，所以整个和等于 $\mathbf{I}^n = \mathbf{I}$。

$\mathbf{I}$ 的右上角元素是 0，而左边的右上角元素是 $\sum_k (-1)^{n+k} \binom{n}{k} F_{n+k}$。因此

$$\sum_{k=0}^{n} (-1)^{n+k} \binom{n}{k} F_{n+k} = 0.$$

额外收获：比较左上角元素，得 $\sum_k (-1)^{n+k} \binom{n}{k} F_{n+k+1} = 1$。

## 验算

- (i)，取 $n = 5$：$F_6 F_4 - F_5^2 = 8 \cdot 3 - 25 = -1 = (-1)^5$。
- (ii)，取 $m = 3$、$n = 4$：$F_4 F_4 + F_3 F_3 = 9 + 4 = 13 = F_7$。
- (a)，取 $n = 4$：$0 + 4 \cdot 1 + 6 \cdot 1 + 4 \cdot 2 + 1 \cdot 3 = 21 = F_8$。
- (b)，取 $n = 3$：$3 \cdot 2 \cdot 1 + 3 \cdot 4 \cdot 1 + 8 \cdot 2 = 34 = F_9$，另外 $F_3 + 3F_4 + 3F_5 + F_6 = 2 + 9 + 15 + 8 = 34$。
- (c)，取 $n = 3$：$-F_3 + 3F_4 - 3F_5 + F_6 = -2 + 9 - 15 + 8 = 0$。

这五条恒等式亦已用电脑验证，对所有不超过 59 的 $n$ 都成立（在 (ii) 中则对所有不超过 49 的 $m, n$ 成立）。

## 常见错误

- 在 (ii) 中比较错了元素，或者把行乘以行，会得到一条看似合理但错误的公式。$\mathbf{Q}^m\mathbf{Q}^n$ 的右上角元素用的是 $\mathbf{Q}^m$ 的第一行和 $\mathbf{Q}^n$ 的第二列。
- 对不可交换的矩阵使用 $(\mathbf{A}+\mathbf{B})^n$ 的二项式定理。这里每个矩阵都是 $\mathbf{Q}$ 的多项式，所以全部都可交换，但这一点值得明确指出。
- 遗漏 $k=0$ 的项：它贡献 $\mathbf{I}$，其右上角元素是 $F_0=0$。

## STEP 之外

关系式 $\mathbf{Q}^2=\mathbf{I}+\mathbf{Q}$ 就是 $\mathbf{Q}$ 的凯莱–哈密顿定理（Cayley–Hamilton theorem）：每个方阵都满足它自己的特征方程（characteristic equation），这里是 $\lambda^2-\lambda-1=0$。把特征值（eigenvalues）为 $\frac{1}{2}(1\pm\sqrt{5})$ 的 $\mathbf{Q}$ 对角化（diagonalising），便得到 $F_n$ 的比内（Binet）封闭公式；而矩阵的幂亦提供了一种以反复平方（repeated squaring）快速计算 $F_n$ 的方法。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）在介绍特征值时，正是用这种方式处理斐波那契数。

详细解答

# STEP II 2023，第 7 题

纯数学 · 两个平方数之和

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

本题在 *Beyond STEP* 第 1.6 节两个平方数之和中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/two-squares.html

## 关键思路

两个平方数之和（sum of two squares）$x^2+y^2$ 正是 $|x+y\mathrm{i}|^2$。这样一来，第 (i) 部分说的就是：两个“两平方数之和”的乘积仍然是两个平方数之和，而把对应的复数（complex number）相乘，就能知道是哪两个平方数。之后每一部分的做法都一样：把每个因数写成两个平方数之和（通常是找出一个相近的完全平方数），然后相乘。最后一部分则反其道而行，以除法代替乘法。

## 解答

### 第 (i) 部分

我们有 $zw=(a+b\mathrm{i})(c+d\mathrm{i})=(ac-bd)+(ad+bc)\mathrm{i}$，所以

$$\begin{aligned}|zw|^2&=(ac-bd)^2+(ad+bc)^2\\&=a^2c^2-2abcd+b^2d^2+a^2d^2+2abcd+b^2c^2\\&=a^2c^2+a^2d^2+b^2c^2+b^2d^2\\&=(a^2+b^2)(c^2+d^2)=|z|^2|w|^2.\end{aligned}$$

由于 $|zw|$ 和 $|z||w|$ 都是非负数，取平方根便得 $|zw|=|z||w|$，证毕。

以下我们会使用平方形式：若 $z$ 和 $w$ 的实部和虚部都是整数，则

$$|z|^2\,|w|^2 = |zw|^2 = (\operatorname{Re} zw)^2 + (\operatorname{Im} zw)^2.$$

**第 (ii) 部分**

这里 $|2+\mathrm{i}|^2 = 4+1 = 5$，而 $|10+11\mathrm{i}|^2 = 100+121 = 221$。两者的乘积是

$$(2+\mathrm{i})(10+11\mathrm{i}) = 20+22\mathrm{i}+10\mathrm{i}-11 = 9+32\mathrm{i},$$

所以由第 (i) 部分，$9^2+32^2 = 5\times 221$。（的确，$81+1024 = 1105$。）因此 $h=9$、$k=32$ 符合要求。

若改用 $10-11\mathrm{i}$（它的模（modulus）相同），则得 $(2+\mathrm{i})(10-11\mathrm{i}) = 31-12\mathrm{i}$，所以 $h=31$、$k=12$ 也符合要求。

**第 (iii) 部分**

这个数的末位是 5，所以 $8045 = 5\times 1609$，而如前所述 $5 = |2+\mathrm{i}|^2$。现在需要把 1609 写成两个平方数之和。它略大于 $40^2 = 1600$，而 $1609-1600 = 9 = 3^2$，所以 $1609 = |40+3\mathrm{i}|^2$。于是

$$(2+\mathrm{i})(40+3\mathrm{i}) = 80+6\mathrm{i}+40\mathrm{i}-3 = 77+46\mathrm{i},$$

由第 (i) 部分得

$$77^2+46^2 = 5\times 1609 = 8045.$$

所以 $m=77$、$n=46$。（若改用 $3+40\mathrm{i}$，则得 $-34+83\mathrm{i}$，所以 $34^2+83^2 = 8045$ 亦成立。）

**第 (iv) 部分**

这里多出的因数是 $36 = 6^2 = |6|^2$，即实数 6 的模的平方。把 $102+201\mathrm{i}$ 乘以 6，只是把实部和虚部同时放大：

$$(6\times 102)^2+(6\times 201)^2 = 36\,(102^2+201^2) = 36\times 50805.$$

所以 $p=612$、$q=1206$。

**第 (v) 部分**

思路是找出两种不同的方法，用高斯整数把 25 写成 $|\alpha|^2$，再把每一个都乘以 1002082 的同一个固定表示式。

首先，1002082 略大于 $1001^2 = 1002001$，而

$$1002082-1002001 = 81 = 9^2,$$

所以 $1002082 = 1001^2 + 9^2 = |1001 + 9\mathrm{i}|^2$。

其次，$25 = |5|^2 = |3 + 4\mathrm{i}|^2 = |3 - 4\mathrm{i}|^2$。把它们各自乘以 $1001 + 9\mathrm{i}$，每个都给出一对数：

$$\begin{aligned} 5(1001 + 9\mathrm{i}) &= 5005 + 45\mathrm{i}, \\ (3 + 4\mathrm{i})(1001 + 9\mathrm{i}) &= 3003 + 27\mathrm{i} + 4004\mathrm{i} - 36 = 2967 + 4031\mathrm{i}, \\ (3 - 4\mathrm{i})(1001 + 9\mathrm{i}) &= 3003 + 27\mathrm{i} - 4004\mathrm{i} + 36 = 3039 - 3977\mathrm{i}. \end{aligned}$$

由第 (i) 部分，每一个的模的平方都是 $25 \times 1002082$。把实部和虚部取绝对值，并把较小的放在前面，便得到三组不同且满足 $r < s$ 的数对：

$$(r, s) = (45, 5005), \qquad (2967, 4031), \qquad (3039, 3977).$$

**想一想** 为甚么不同时使用 $4 + 3\mathrm{i}$，得出第四组数对？

因为 $4 + 3\mathrm{i} = \mathrm{i}(3 - 4\mathrm{i})$，而乘以 i 只会把实部和虚部互换（不计正负号）。它重现的是 $(3039, 3977)$，只是次序倒转，所以并没有给出新的结果。真正不同的选择是 5、$3 + 4\mathrm{i}$ 和 $3 - 4\mathrm{i}$。

### 第 (vi) 部分

现在我们把方法倒过来用。同样先找一个相近的平方数：$1002037 - 1002001 = 36 = 6^2$，所以

$$1002037 = 1001^2 + 6^2 = |1001 + 6\mathrm{i}|^2,$$

而且 $109 = 10^2 + 3^2 = |10 + 3\mathrm{i}|^2$。如果我们能把 $1001 + 6\mathrm{i} = (10 + 3\mathrm{i})(t + u\mathrm{i})$ 中的 $t$、$u$ 取为整数，那么由第 (i) 部分得

$$1002037 = 109\,(t^2 + u^2),$$

再除以 109（利用 $109 \times 9193 = 1002037$）便得 $t^2 + u^2 = 9193$。

于是我们做除法，分子和分母同乘以共轭（conjugate）$10 - 3\mathrm{i}$：

$$\begin{aligned} \frac{1001 + 6\mathrm{i}}{10 + 3\mathrm{i}} &= \frac{(1001 + 6\mathrm{i})(10 - 3\mathrm{i})}{109} = \frac{10010 - 3003\mathrm{i} + 60\mathrm{i} + 18}{109} \\ &= \frac{10028 - 2943\mathrm{i}}{109} = 92 - 27\mathrm{i}, \end{aligned}$$

因为 $109 \times 92 = 10028$ 及 $109 \times 27 = 2943$。这个除法刚好整除，所以 $1001 + 6\mathrm{i} = (10 + 3\mathrm{i})(92 - 27\mathrm{i})$，因此

$$92^2 + 27^2 = 9193.$$

所以 $t = 92$、$u = 27$。验算一下：$8464 + 729 = 9193$。

**想一想** 如果除法不能整除，那怎么办？

那就试另一个共轭。改为除以 $10-3\mathrm{i}$，得到 $(9992+3063\mathrm{i})/109$，这并不是高斯整数（9992 不是 109 的倍数）。由于 109 是质数（prime），两个选择之中必有一个可行；这正是下文“STEP 之外”一节提到的高斯整数性质。至于拿 $1001-6\mathrm{i}$ 来除，并不会得到新结果，只会得到上述结果的共轭。

## 验算

每个答案都可以直接用算术验证：$9^2+32^2=1105=5\times221$；$77^2+46^2=5929+2116=8045$；$612^2+1206^2=1828980=36\times50805$；对于 (v)，$45^2+5005^2$、$2967^2+4031^2$ 和 $3039^2+3977^2$ 都等于 25052050；以及 $92^2+27^2=9193$。用电脑搜索可以证实这些结果，并发现 $25\times1002082$ 其实有六个满足 $0<r<s$ 的表示式（另外三个是 $(721,4953)$、$(2079,4553)$ 和 $(2395,4395)$）。它们来自 1002082 本身的另一个表示式，所以任何三组不同的数对都可接受。同样地，$9193=48^2+83^2$ 亦成立。

## 常见错误

- 在 (v) 中，以为 $(3+4\mathrm{i})$ 和 $(4+3\mathrm{i})$ 会给出不同的数对：其实它们给出的是同一对数，只是次序相反。
- 在 (vi) 中，除以错误的共轭而得到非整数，然后就放弃。应该试试另一个。
- 展开乘积时，在 $\mathrm{i}^2=-1$ 上弄错正负号：最后务必直接验算平方和。

## STEP 之外

所有 $a,b$ 为整数的数 $a+b\mathrm{i}$ 组成*高斯整数*（Gaussian integers），这是一个环（ring）；与普通整数一样，在其中分解为质数的方式是唯一的。若普通质数 $p\equiv1\pmod4$，则它可分裂为 $(x+y\mathrm{i})(x-y\mathrm{i})$，这就是费马（Fermat）的二平方和定理（two-squares theorem）。计算把这些因数分配的方法有多少种，便能准确解释一个数有多少个表示式，以及为何 1002082 的表示式不止我们用过的那一个。Davenport 的 *The Higher Arithmetic*（《高等算术》）对两个平方数之和有精彩的论述；Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则把高斯整数作为一个环来讨论。

详细解答

# STEP II 2023，第 8 题

纯数学 · 三维向量

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

先为三对对棱命名，第 (i) 部分就化为三条线性方程。之后的一切，都来自把每个长度条件转化为点积（dot product），所用的是 $|\mathbf{u}-\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2\,\mathbf{u}\cdot\mathbf{v}$。一旦三个点积 $\mathbf{a}\cdot\mathbf{b}$、$\mathbf{b}\cdot\mathbf{c}$、$\mathbf{c}\cdot\mathbf{a}$ 都以 $|\mathbf{a}|^2$、$|\mathbf{b}|^2$ 和 $|\mathbf{c}|^2$ 表示出来，第 (iii) 和 (iv) 部分就只是展开括号而已。

## 解答

### 第 (i) 部分

将四面体（tetrahedron）记为 $OABC$。它的六条棱（edge）组成三对没有公共顶点的棱：

$$OA \text{ 和} BC, \qquad OB \text{ 和} CA, \qquad OC \text{ 和} AB.$$

设 $OA = p$、$BC = p'$、$OB = q$、$CA = q'$、$OC = r$、$AB = r'$。四个面的周界（perimeter）分别为

$$OAB : p + q + r', \qquad OAC : p + r + q',$$
$$OBC : q + r + p', \qquad ABC : p' + q' + r'.$$

若四面体是等腰四面体（*isosceles tetrahedron*），则 $p = p'$、$q = q'$、$r = r'$，而四个周界都等于 $p + q + r$。

反过来，假设四个周界全部相等。设 $X = p - p'$、$Y = q - q'$、$Z = r - r'$。把前三个面

分别与 $ABC$ 比较：

$$(p+q+r')-(p'+q'+r')=X+Y=0,$$
$$(p+r+q')-(p'+q'+r')=X+Z=0,$$
$$(q+r+p')-(p'+q'+r')=Y+Z=0.$$

三式相加得 $2(X+Y+Z)=0$。再以 $X+Y+Z=0$ 依次减去每一条方程，得 $Z=0$、$Y=0$ 及 $X=0$。所以 $p=p'$、$q=q'$ 及 $r=r'$：四面体是等腰的。这就证明了两个方向。

**第 (ii) 部分**

由于 $\overrightarrow{BC}=\mathbf{c}-\mathbf{b}$，条件 $OA=BC$ 即是

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

移项后得 $2\,\mathbf{b}\cdot\mathbf{c}=|\mathbf{b}|^2+|\mathbf{c}|^2-|\mathbf{a}|^2$，正是所求。

对另外两对对棱（$OB=CA$ 及 $OC=AB$）作同样的论证，得

$$2\,\mathbf{a}\cdot\mathbf{c}=|\mathbf{a}|^2+|\mathbf{c}|^2-|\mathbf{b}|^2,\qquad 2\,\mathbf{a}\cdot\mathbf{b}=|\mathbf{a}|^2+|\mathbf{b}|^2-|\mathbf{c}|^2.$$

把这两式相加，$|\mathbf{b}|^2$ 和 $|\mathbf{c}|^2$ 项互相抵消：

$$2\,\mathbf{a}\cdot\mathbf{b}+2\,\mathbf{a}\cdot\mathbf{c}=2|\mathbf{a}|^2,\quad\text{所以}\quad\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=|\mathbf{a}|^2.$$

这三条点积公式对 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 是对称的，因此同理可得

$$\mathbf{b}\cdot(\mathbf{c}+\mathbf{a})=|\mathbf{b}|^2,\qquad\mathbf{c}\cdot(\mathbf{a}+\mathbf{b})=|\mathbf{c}|^2.$$

**第 (iii) 部分**

设 $S=|\mathbf{a}|^2+|\mathbf{b}|^2+|\mathbf{c}|^2$。把第 (ii) 部分中 $2\,\mathbf{b}\cdot\mathbf{c}$、$2\,\mathbf{c}\cdot\mathbf{a}$ 和 $2\,\mathbf{a}\cdot\mathbf{b}$ 的三条公式相加，每个长度的平方都以正号出现两次、以负号出现一次，所以

$$2(\mathbf{a}\cdot\mathbf{b}+\mathbf{b}\cdot\mathbf{c}+\mathbf{c}\cdot\mathbf{a})=S.$$

*与 $O$ 的距离。*

$$OG^2=\tfrac{1}{16}|\mathbf{a}+\mathbf{b}+\mathbf{c}|^2=\tfrac{1}{16}\big(S+2(\mathbf{a}\cdot\mathbf{b}+\mathbf{b}\cdot\mathbf{c}+\mathbf{c}\cdot\mathbf{a})\big)=\tfrac{1}{16}(2S)=\tfrac{1}{8}S.$$

*与 $A$ 的距离。*这里 $\overrightarrow{AG}=\frac{1}{4}(\mathbf{b}+\mathbf{c}-3\mathbf{a})$，而

$$\begin{aligned}|\mathbf{b}+\mathbf{c}-3\mathbf{a}|^2&=|\mathbf{b}+\mathbf{c}|^2-6\,\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})+9|\mathbf{a}|^2\\&=|\mathbf{b}|^2+|\mathbf{c}|^2+2\,\mathbf{b}\cdot\mathbf{c}-6|\mathbf{a}|^2+9|\mathbf{a}|^2\qquad\text{利用}\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=|\mathbf{a}|^2\\&=|\mathbf{b}|^2+|\mathbf{c}|^2+\big(|\mathbf{b}|^2+|\mathbf{c}|^2-|\mathbf{a}|^2\big)+3|\mathbf{a}|^2\\&=2S.\end{aligned}$$

所以同样有 $AG^2 = \frac{1}{16}(2S) = \frac{1}{8}S$。

*与 $B$ 和 $C$ 的距离。*我们用过的每一个关于 $\mathbf{a}$ 的事实，对 $\mathbf{b}$ 和 $\mathbf{c}$ 同样成立（见第 (ii) 部分末段），而 $S$ 是对称的。因此完全相同的计算给出 $BG^2 = CG^2 = \frac{1}{8}S$。

故此 $G$ 与四个顶点的距离都是 $\sqrt{S/8}$。

### 第 (iv) 部分

展开题目提示的向量：

$$\begin{aligned}|\mathbf{a}-\mathbf{b}-\mathbf{c}|^2 &= S - 2\,\mathbf{a}\cdot\mathbf{b} - 2\,\mathbf{a}\cdot\mathbf{c} + 2\,\mathbf{b}\cdot\mathbf{c}\\ &= S - 2|\mathbf{a}|^2 + 2\,\mathbf{b}\cdot\mathbf{c} && \text{由第 (ii) 部分}\\ &= \left(|\mathbf{b}|^2 + |\mathbf{c}|^2 - |\mathbf{a}|^2\right) + 2\,\mathbf{b}\cdot\mathbf{c} = 4\,\mathbf{b}\cdot\mathbf{c}.\end{aligned}$$

所以

$$\mathbf{b}\cdot\mathbf{c} = \tfrac{1}{4}|\mathbf{a}-\mathbf{b}-\mathbf{c}|^2 \geq 0,$$

因此角 $BOC$ 不是钝角（obtuse angle）。

> **想一想** 这只是十二个角之中的一个。怎样才能不必重复计算，就得到其余所有的角？

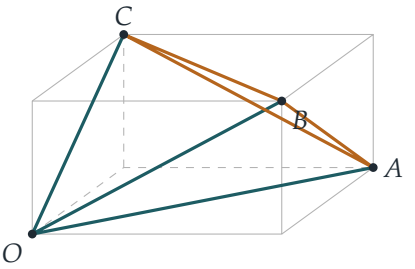
以 $\mathbf{b}-\mathbf{c}-\mathbf{a}$ 和 $\mathbf{c}-\mathbf{a}-\mathbf{b}$ 作同样的计算，得 $\mathbf{c}\cdot\mathbf{a} \geq 0$ 及 $\mathbf{a}\cdot\mathbf{b} \geq 0$，所以 $O$ 处没有钝角。“等腰”是四面体本身的性质，与我们怎样标记顶点无关，所以任何一个顶点都可以叫作 $O$。因此同一论证在 $A$、$B$ 和 $C$ 处同样适用。（另一做法：每个面的三边都是 $p$、$q$、$r$，所以四个面是全等（congruent）三角形，而它们的角正好就是 $O$ 处的三个角 $BOC$、$COA$、$AOB$。）故此在任何一个顶点处相交的两条棱之间的角都不是钝角。

*会不会是直角？*等号 $\mathbf{b}\cdot\mathbf{c} = 0$ 只在 $|\mathbf{a}-\mathbf{b}-\mathbf{c}| = 0$ 时成立，即 $\mathbf{a} = \mathbf{b}+\mathbf{c}$。但这样 $A$ 便位于经过 $O$、$B$、$C$ 的平面上（事实上 $OBAC$ 是长方形），$OABC$ 是平的，并不是四面体。每个角的情况都一样，所以**没有任何角可以是直角**：在真正的等腰四面体中，每个面角都是锐角（acute angle）。

## 验算

- 一个具体例子：取边长为 $x, y, z$ 的长方体（cuboid），选出它的四个顶点，使其中任何两个都不由长方体的棱相连，例如 $O = (0,0,0)$、$A = (x,y,0)$、$B = (x,0,z)$、$C = (0,y,z)$。对棱是相对两个面的对角线，所以彼此相等。取 $x = 1$、$y = 2$、$z = 3$，得 $\mathbf{b}\cdot\mathbf{c} = 9$ 及 $\mathbf{a}-\mathbf{b}-\mathbf{c} = (0,0,-6)$，而确实 $\frac{1}{4}\cdot 36 = 9$。此外 $G = (\frac{1}{2}, 1, \frac{3}{2})$ 是长方体的中心，与每个顶点的距离都是 $\sqrt{14}/2$，与 $\sqrt{S/8}$（其中 $S = 5+10+13 = 28$）吻合。

- 平面的情况：$\mathbf{b} = (3,0,0)$、$\mathbf{c} = (0,4,0)$、$\mathbf{a} = (3,4,0)$ 给出一个长方形，它的对“棱”（边和对角线）相等，而且有直角。这正是我们排除了的退化情况。
- 这些恒等式亦已用电脑在数千个随机的长方体例子上验证过。




## 常见错误

- 在第 (i) 部分只证明了一个方向。“当且仅当”（if and only if）需要两个方向，而有实质内容的是逆命题那一部分。
- 在第 (iv) 部分没有妥善排除直角。$\mathbf{b}\cdot\mathbf{c} = 0$ 迫使 $\mathbf{a} = \mathbf{b} + \mathbf{c}$，令四点共面（coplanar）。
- 只检查了 $O$ 处的角便停下来。要说明为甚么其余三个顶点也有同样结论。

## STEP 之外

上面的长方体图已道出全部真相：每个等腰四面体（又称双楔体（disphenoid））都可以从长方体选取相间的顶点得到。这样，第 (iii) 部分是说形心（centroid）就是长方体的中心，而第 (iv) 部分是说长方体的边长是正数。把长度条件转化为点积，进而转化为由内积（inner product）组成的格拉姆矩阵（Gram matrix），是大学一年级线性代数（linear algebra）和几何的标准工具。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）是了解这套方法如何发展的好去处。

详细解答

# STEP II 2023，第 9 题

力学 · 斜坡上的连接质点

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

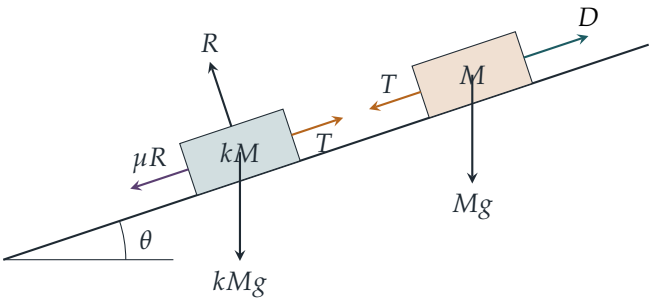
## 关键思路

在三种情况下，分别对货车和拖车写出牛顿（Newton）第二定律（Newton's second law）。下坡其实就是上一条倾角为 $-\alpha$ 的斜坡，所以一组方程就能涵盖全部三种情况。单凭货车的方程，几乎就能直接得出 (i) 的关系式。至于 (ii)，整道题归结为比较 $1-\cos\alpha$ 与 $\sin\alpha$ 的大小。

## 解答

我们把货车和拖车视为质点（particle），两者之间以一根沿运动方向放置的轻刚性杆连接。两者都沿货车的方向运动，所以拖车所受的摩擦力（friction）与运动方向相反，大小为 $\mu R$，其中 $R$ 是拖车所受的法向反作用力（normal reaction）。张力（tension）$T$ 把拖车向前拉，把货车向后拉。（若 $T$ 算出来是负数，就表示杆处于推力（thrust）状态；代数运算不变。）

把斜坡的倾角记为 $\theta$，规定 $\theta=\alpha$ 为上坡，$\theta=0$ 为平地，$\theta=-\alpha$ 为下坡。



设沿斜坡向上的加速度（acceleration）为 $a$。对每个物体沿斜坡方向分解力，并对拖车在垂直于斜坡的方向分解力（得 $R = kMg\cos\theta$），便有

$$\text{货车：} \qquad D - T - Mg\sin\theta = Ma, \tag{1}$$

$$\text{拖车：} \qquad T - kMg\sin\theta - \mu kMg\cos\theta = kMa. \tag{2}$$

把 (1) 与 (2) 相加，消去 $T$：

$$a = \frac{D}{(1+k)M} - g\sin\theta - \frac{\mu kg\cos\theta}{1+k}.$$

代入 (1)，得 $T = D - Mg\sin\theta - Ma$，所以

$$T = D - \frac{D}{1+k} + \frac{\mu kMg\cos\theta}{1+k} = \frac{kD + \mu kMg\cos\theta}{1+k}.$$

含 $\sin\theta$ 的项互相抵消，因此张力只通过 $\cos\theta$ 与 $\theta$ 有关。

### 第 (i) 部分

由于 $\cos(-\alpha) = \cos\alpha$，$\theta = \alpha$ 与 $\theta = -\alpha$ 两种情况的张力相同：

$$T_1 = T_3 = \frac{kD + \mu kMg\cos\alpha}{1+k}, \qquad T_2 = \frac{kD + \mu kMg}{1+k}.$$

要证明第二个结果，只需用到三种情况下的货车方程 (1)：

$$\begin{aligned} Ma_1 &= D - T_1 - Mg\sin\alpha, \\ Ma_2 &= D - T_2, \\ Ma_3 &= D - T_3 + Mg\sin\alpha. \end{aligned}$$

把第一和第三条方程相加，并利用 $T_3 = T_1$，得 $M(a_1 + a_3) = 2D - 2T_1$。再减去中间方程的两倍，

$$M(a_1 + a_3 - 2a_2) = (2D - 2T_1) - (2D - 2T_2) = 2(T_2 - T_1),$$

即为所求。

### 第 (ii)(a) 部分

第一个不等式。由 (i) 可知，$a_2 < \frac{1}{2}(a_1 + a_3)$ 等价于 $T_2 > T_1$（因为 $M > 0$）。由上面的公式，

$$T_2 - T_1 = \frac{\mu kMg\,(1 - \cos\alpha)}{1+k} > 0,$$

这是因为 $0 < \alpha < \frac{1}{2}\pi$ 给出 $\cos\alpha < 1$，而由于有摩擦力作用，$\mu > 0$。所以 $a_2 < \frac{1}{2}(a_1 + a_3)$。

第二个不等式。$\frac{1}{2}(a_1+a_3) < a_3$ 等价于 $a_1 < a_3$。从 $a$ 的公式可见，对于 $\theta = \pm\alpha$，含 $D$ 的项和摩擦力项都相同，所以

$$a_3 - a_1 = 2g\sin\alpha > 0.$$

（另一做法是把两条货车方程相减，并利用 $T_1 = T_3$，得 $M(a_3 - a_1) = 2Mg\sin\alpha$。）因此

$$a_2 < \tfrac{1}{2}(a_1+a_3) < a_3.$$

### 第 (ii)(b) 部分

在 $a$ 的公式中分别取 $\theta = 0$ 和 $\theta = \alpha$，

$$a_2 - a_1 = g\sin\alpha - \frac{\mu k g\,(1-\cos\alpha)}{1+k} = g\left[\sin\alpha - \frac{\mu k}{1+k}(1-\cos\alpha)\right].$$

**想一想** 上坡会损失 $g\sin\alpha$ 的加速度，但法向反作用力减小，又会省下一些摩擦力。哪一个效应较大?

省下的摩擦力是较小的一方。首先，$\dfrac{k}{1+k} < 1$ 且 $\mu < 1$，所以 $0 \le \dfrac{\mu k}{1+k} < 1$。其次，利用半角公式（half-angle formulae），

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

所以 $\dfrac{1-\cos\alpha}{\sin\alpha} = \tan\tfrac{1}{2}\alpha < 1$，因为 $0 < \tfrac{1}{2}\alpha < \tfrac{1}{4}\pi$。因此

$$\frac{\mu k}{1+k}(1-\cos\alpha) < 1-\cos\alpha < \sin\alpha,$$

其中第一个不等号是严格的，因为 $1-\cos\alpha > 0$。所以 $a_2 - a_1 > 0$，即 $a_1 < a_2$。条件 $\mu < 1$ 正是在这里用上的。若拖车非常重、$\mu$ 很大而斜坡又很陡，省下的摩擦力就可能超过 $g\sin\alpha$。

## 验算

- 取 $M = 1000\,\mathrm{kg}$、$k = 0.5$、$\mu = 0.4$、$\alpha = 30°$、$D = 5000\,\mathrm{N}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$。那么 $a_1 \approx -2.698$、$a_2 \approx 2.027$、$a_3 \approx 7.102\,\mathrm{m\,s^{-2}}$，而 $T_1 = T_3 \approx 2798\,\mathrm{N}$、$T_2 \approx 2973\,\mathrm{N}$。可见 $\frac{1}{2}(a_1+a_3) \approx 2.202$ 确实严格介乎 $a_2$ 与 $a_3$ 之间，而且 $a_1 < a_2$。用电脑对大量 $\mu < 1$ 的随机参数组合进行检验，所有恒等式和不等式均成立。
- 若 $\mu = 0$，则在任何斜坡上张力都是 $kD/(1+k)$，而 $a_2$ 恰好是 $a_1$ 与 $a_3$ 的平均值。这就是 (ii)(a) 需要有摩擦力的原因。
- 取 $\mu = 3$、$k = 100$ 及 $\alpha = 1.5$ 弧度，则 $a_1 > a_2$，所以 (b) 中的条件 $\mu < 1$ 确实不可缺少。

## 常见错误

- 在斜坡上把摩擦力写成 $\mu kMg$。斜坡上的法向反作用力是 $kMg\cos\alpha$，而第 (ii) 部分的全部内容正是建基于这个差别。
- 下坡时把重力分量的正负号弄错。把一切用 $\theta$ 表示，再令 $\theta = -\alpha$，就可避免这个错误。
- 在 (ii)(b) 中只论证 $\mu k/(1+k) < 1$，却没有比较 $1-\cos\alpha$ 与 $\sin\alpha$。两个事实缺一不可。

## STEP 之外

把一个系统拆分成若干物体，对每个物体应用牛顿定律，而内力则以大小相等、方向相反的形式成对出现——这是刚体（rigid body）力学和多质点力学的基本方法。把各方程相加，便得到质心（centre of mass）的运动，而内力则会消去。不等式 $1-\cos\alpha < \sin\alpha$ 也是一个利用半角恒等式比较函数（function）大小的小例子，这种技巧在整个分析学（analysis）中都很有用。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）对这些概念有细致的阐述。

详细解答

# STEP II 2023，第 10 题

力学 · 抛体轨道的包络线

**STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

一条轨迹与曲线 $E$ 相切，当且仅当求它们公共点的方程有重根，所以第 (i) 部分其实是一个判别式（discriminant）的计算。要求对每一个角度都出现这种情况，便迫使 $E$ 成为所有轨迹的包络线（envelope），即速度为 $u$ 的质点所能到达范围的边界。题目其余部分从三个角度审视这条边界：哪些点可以被击中、质点在某一固定时刻位于何处，以及沿着包络线本身滑行的第二粒质点。

## 解答

全题中，质点视为只受重力作用而自由运动的点，不计空气阻力，并记 $c = \cos\alpha$，$s = \sin\alpha$。

### 第 (i) 部分

设 $t$ 为抛出后经过的时间，则 $P_\alpha$ 一直留在平面 $y = 0$ 上，且

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

消去 $t = x/(uc)$，得到轨迹（trajectory）

$$z = x\tan\alpha - \frac{gx^2}{2u^2c^2}.$$

与 $E$ 的公共点满足 $A - Bx^2 = x\tan\alpha - \dfrac{gx^2}{2u^2c^2}$，即

$$\left(\frac{g}{2u^2c^2} - B\right)x^2 - x\tan\alpha + A = 0. \qquad (*)$$

若 $x^2$ 的系数为零，则 $(*)$ 是线性方程，而 $\tan\alpha \neq 0$，这只会给出一个穿越点而非相切，所以 $(*)$ 确实是二次方程。两曲线恰好相切一次，当且仅当 $(*)$ 有重根，即其判别式为零：

$$\tan^2\alpha = 4A\left(\frac{g}{2u^2c^2} - B\right).$$

两边乘以 $u^2c^2$：

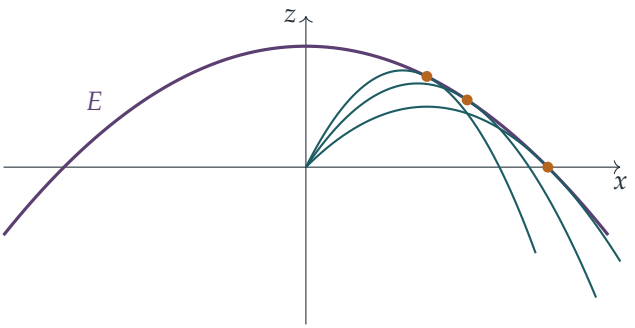
$$\begin{aligned} u^2s^2 &= 2gA - 4ABu^2c^2 \\ u^2 - u^2c^2 &= 2gA - 4ABu^2c^2 \qquad (\text{利用} s^2 = 1 - c^2) \\ u^2 - 2gA &= u^2(1 - 4AB)\cos^2\alpha, \end{aligned}$$

此即所求。

若此式对 $(0, \frac{1}{2}\pi)$ 中每一个 $\alpha$ 都成立，则左边是常数，而 $\cos^2\alpha$ 取遍 $(0,1)$ 中所有值。一个常数对所有这些值都等于 $k\cos^2\alpha$，便迫使 $k = 0$，从而该常数为 0。所以 $u^2 = 2gA$ 且 $4AB = 1$：

$$A = \frac{u^2}{2g}, \qquad B = \frac{g}{2u^2}.$$

反过来，这些值使恒等式对每一个 $\alpha$ 都成立。切点就是重根 $x = \dfrac{\tan\alpha}{2(g/(2u^2c^2) - B)} = \dfrac{u^2}{g\tan\alpha}$。




上图（取 $u^2/g = 2$ 绘制）显示三条轨迹，每条都在琥珀色的点与 $E$ 相切一次。

**第 (ii) 部分**

每粒质点都在包含其初速度的铅垂平面内运动，而整个情况关于 $z$ 轴对称。因此考虑一个与 $z$ 轴水平距离为 $r$、高度为 $z$ 的点，并在通过该点的铅垂平面内处理，设 $\theta \in (-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 为朝该点方向、高于水平线的抛射角。

若 $r = 0$，只有垂直向上或向下射出的质点能够到达；向上射出的那粒可到达不超过 $u^2/(2g)$ 的每一个高度，向下射出的那粒可到达每一个负高度，所以轴上可到达的点就是满足 $z \leq u^2/(2g)$ 的点。

若 $r > 0$，与第 (i) 部分完全一样，质点经过该点当且仅当

$$z = r\tan\theta - \frac{gr^2}{2u^2}(1+\tan^2\theta),$$

这里用了 $1/\cos^2\theta = 1+\tan^2\theta$。记 $T = \tan\theta$，它可取任何实数值，于是得到二次方程

$$\frac{gr^2}{2u^2}T^2 - rT + \left(z + \frac{gr^2}{2u^2}\right) = 0,$$

它有实根当且仅当

$$r^2 \geq \frac{2gr^2}{u^2}\left(z+\frac{gr^2}{2u^2}\right) \iff z \leq \frac{u^2}{2g} - \frac{gr^2}{2u^2},$$

这里除以了 $2gr^2/u^2 > 0$。

所以，由 $r^2 = x^2+y^2$，可被击中的点恰好是位于抛物面（paraboloid）

$$z = \frac{u^2}{2g} - \frac{g}{2u^2}(x^2+y^2),$$

之上或其下方的点；这个曲面正是把 $E$（取第 (i) 部分的值）绕 $z$ 轴旋转而得的。严格位于其下方的点可由平面内两个方向到达，位于其上的点恰好由一个方向到达（即重根，与第 (i) 部分的相切吻合），而位于其上方的点则无法到达。

### 第 (iii) 部分

设质点的初速度为 $u\mathbf{n}$，其中 $\mathbf{n}$ 是单位向量（unit vector）。经过时间 $t$ 后，其位置为

$$\mathbf{r} = ut\,\mathbf{n} - \tfrac{1}{2}gt^2\,\mathbf{k},$$

其中 $\mathbf{k}$ 是向上的单位向量。因此

$$\left|\mathbf{r} - (-\tfrac{1}{2}gt^2\,\mathbf{k})\right| = |ut\,\mathbf{n}| = ut.$$

所以在时刻 $t$，每粒质点都位于以 $(0,0,-\frac{1}{2}gt^2)$ 为球心、$ut$ 为半径的球面上。从物理角度看：重力使每粒质点获得相同的加速度，所以相对于一个从原点自由下落的点，它们全都以速率 $u$ 沿直线向外飞散。

### 第 (iv) 部分

设 $\tau$ 为 $Q$ 抛出后经过的时间，则 $Q$ 有 $y = 0$ 及

$$x = u\tau, \qquad z = A - \tfrac{1}{2}g\tau^2.$$

代入 $\tau = x/u$：

$$z = A - \frac{g}{2u^2}x^2 = A - Bx^2,$$

这里用了第 (i) 部分的 $B = g/(2u^2)$。所以 $Q$ 一直在 $E$ 上。

**第 (v) 部分**

假设 $Q$ 在爆炸后时刻 $T$ 抛出，并且 $Q$ 与 $P_\alpha$ 在爆炸后时刻 $t$ $(\geq T)$ 位于同一点。那么 $\tau = t - T$，而两者的 $y$ 坐标都是 0，所以需要 $x$ 坐标和 $z$ 坐标相等：

$$uct = u(t-T),$$
$$ust - \tfrac{1}{2}gt^2 = A - \tfrac{1}{2}g(t-T)^2.$$

第一式给出 $t - T = ct$，所以 $T = t(1-c)$。把 $t - T = ct$ 代入第二式，并利用 $1 - c^2 = s^2$：

$$ust - \tfrac{1}{2}gs^2t^2 = A = \frac{u^2}{2g}.$$

? **想一想** 怎样不用繁琐的公式来解出 $t$?

乘以 $2g$ 并把所有项移到一边：$g^2s^2t^2 - 2gus\,t + u^2 = 0$，这是一个完全平方

$$(gst - u)^2 = 0, \qquad \text{所以} \qquad t = \frac{u}{g\sin\alpha}.$$

出现重根正是意料之中：碰撞必须发生在 $P_\alpha$ 的轨迹与 $E$ 相切之处，因为那是两条路径唯一的公共点。事实上，此时 $P_\alpha$ 位于 $x = uct = u^2/(g\tan\alpha)$，正是第 (i) 部分求得的切点。因此

$$T = t(1-\cos\alpha) = \frac{u(1-\cos\alpha)}{g\sin\alpha},$$

此即所求。由于 $0 < 1 - \cos\alpha < 1$，得 $0 < T < t$，所以 $Q$ 确实是在爆炸之后、碰撞之前抛出的。

## 验算

- 在第 (i) 部分，取 $\alpha = \frac{1}{4}\pi$ 得切点 $x = u^2/g$，$z = 0$：这是在水平地面上的最大射程，可到达区域的边界理应经过此点。
- 在第 (ii) 部分，在轴 $r = 0$ 上，界限是 $z \leq u^2/(2g)$，即垂直向上射出的质点所达的最大高度。
- 在第 (v) 部分，取 $u = 7$、$g = 9.8$ 及 $\alpha = 0.5$，公式给出 $T \approx 0.182$；把两个运动逐步按时间推进，可确认两粒质点在 $t = u/(g\sin\alpha) \approx 1.49$ 相遇，而即使把 $Q$ 延迟 0.05 才抛出，它们便会错过。
- 当 $\alpha \to \frac{1}{2}\pi$ 时，$T \to u/g$ 且 $t \to u/g$：$P_\alpha$ 几乎垂直向上，到达 $E$ 的顶点 $(0, 0, A)$，到达时刻为 $u/g$，恰好是 $Q$ 从那里出发的时刻。

## 常见错误

- 未先把公共点方程整理成关于 $x$ 的单一二次方程便令判别式为零，结果在 $x^2$ 的系数中漏掉了 $-B$ 项。
- 在第 (ii) 部分只描述边界曲面：题目要求的是整个集合，即抛物面之上及其下方的实心区域。
- 在第 (v) 部分，两粒质点用了同一个时间变量：$Q$ 只运动了 $t-T$ 的时间。

## STEP 之外

曲线 $E$ 是这族轨迹的*包络线*：一条与族中每一成员都相切的曲线。一般而言，曲线族 $F(x,z,\alpha)=0$ 的包络线可通过在 $F=0$ 与 $\partial F/\partial\alpha=0$ 之间消去 $\alpha$ 求得，这就是本题所用重根条件的微积分版本。第 (ii) 部分的抛物面常被称为*安全抛物面*（paraboloid of safety），而第 (iii) 部分则是转换到自由下落参考系（freely falling frame）的一个例子，在该参考系中重力消失了。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对这两个概念都有阐述。

详细解答

# STEP II 2023，第 11 题

概率与统计 · 混合分布

> **STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/707102-2023-past-papers.zip

本题在 *Beyond STEP* 第 13.5 节问尴尬的问题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch13/awkward-questions.html

## 关键思路

$Y$ 是一个混合分布（mixture）：先随机选定一个分布，再从该分布中抽取数值。它的平均值（mean）是两个平均值的加权平均，但它的方差（variance）多了一项，因为两个平均值之间的差距本身也会带来变异。第 (ii) 和第 (iii) 部分都是先掷硬币，再掷骰子，但两者的表现截然不同。在 (ii) 中，每掷一次骰子都重新掷一次硬币，所以混合的效果会被平均掉。在 (iii) 中，一枚硬币决定一切，而那额外的一项最终会占主导地位。

## 解答

### 第 (i) 部分

选用哪个随机变量，与 $X_1$ 和 $X_2$ 的取值互相独立（independent），所以

$$\mathrm{P}(Y = x_i) = p\,a_i + q\,b_i \qquad (i = 1, \ldots, n).$$

因此

$$\mathrm{E}(Y) = \sum_i x_i(pa_i + qb_i) = p\sum_i x_i a_i + q\sum_i x_i b_i = p\mu_1 + q\mu_2.$$

同理，$\mathrm{E}(Y^2) = p\,\mathrm{E}(X_1^2) + q\,\mathrm{E}(X_2^2) = p(\sigma_1^2+\mu_1^2) + q(\sigma_2^2+\mu_2^2)$。所以

$$\begin{aligned}\mathrm{Var}(Y) &= p\sigma_1^2 + q\sigma_2^2 + p\mu_1^2 + q\mu_2^2 - (p\mu_1+q\mu_2)^2\\ &= p\sigma_1^2 + q\sigma_2^2 + p(1-p)\mu_1^2 + q(1-q)\mu_2^2 - 2pq\mu_1\mu_2\\ &= p\sigma_1^2 + q\sigma_2^2 + pq\mu_1^2 + pq\mu_2^2 - 2pq\mu_1\mu_2\\ &= p\sigma_1^2 + q\sigma_2^2 + pq(\mu_1-\mu_2)^2,\end{aligned}$$

当中用到 $1-p=q$ 及 $1-q=p$。这就是所求的结果。

**第 (ii) 部分**

设 $H$ 为硬币掷出正面的事件，$S$ 为骰子掷出六点的事件。硬币与骰子互相独立，所以

$$\mathrm{P}(B=1) = \mathrm{P}(H)\mathrm{P}(S) + \mathrm{P}(H^c)\mathrm{P}(S^c) = \tfrac{1}{2}\cdot\tfrac{1}{6} + \tfrac{1}{2}\cdot\tfrac{5}{6} = \tfrac{1}{2}.$$

所以每个 $B$ 取 0 或 1 的概率各为 $\frac{1}{2}$。$Z_1$ 是 $n$ 个独立的 $B$ 之中 1 的个数，而每个 $B$ 成功的概率同样是 $\frac{1}{2}$。这正是二项分布（binomial distribution）的设置，所以 $Z_1 \sim \mathrm{B}(n, \frac{1}{2})$，平均值为 $\frac{n}{2}$，方差为 $\frac{n}{4}$。

当 $n$ 很大时，我们以 $\mathrm{N}(\frac{n}{2}, \frac{n}{4})$ 近似 $Z_1$，其标准差（standard deviation）为 $\frac{\sqrt{n}}{2}$。与平均值相差不超过 10%，即 $|Z_1 - \frac{n}{2}| \le \frac{n}{20}$。设 $Z \sim \mathrm{N}(0,1)$，进行标准化（standardising），得

$$\mathrm{P}\big(|Z_1 - \tfrac{n}{2}| \le \tfrac{n}{20}\big) \approx \mathrm{P}\Big(|Z| \le \frac{n/20}{\sqrt{n}/2}\Big) = \mathrm{P}\Big(|Z| \le \frac{\sqrt{n}}{10}\Big) = 2\Phi\Big(\frac{\sqrt{n}}{10}\Big) - 1.$$

当 $n \to \infty$ 时，$\sqrt{n}/10 \to \infty$，所以 $\Phi(\sqrt{n}/10) \to 1$，而该概率趋向 $2-1=1$。（连续性校正（continuity correction）只会令界限改变 $\frac{1}{\sqrt{n}}$ 个标准差，并不影响极限。）

**第 (iii) 部分**

设 $X_1$ 为 $n$ 粒骰子中掷出六点的数目，$X_2$ 为不是六点的数目。则 $X_1 \sim \mathrm{B}(n, \frac{1}{6})$ 及 $X_2 \sim \mathrm{B}(n, \frac{5}{6})$。现在 $Z_2$ 以概率 $p = \frac{1}{2}$ 取 $X_1$ 的值，以概率 $q = \frac{1}{2}$ 取 $X_2$ 的值，而且硬币与骰子互相独立。这正是第 (i) 部分的情况，其中

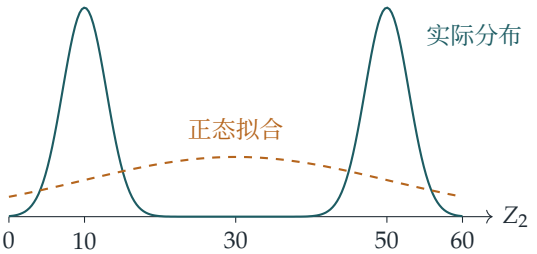
$$\mu_1 = \tfrac{n}{6}, \quad \mu_2 = \tfrac{5n}{6}, \quad \sigma_1^2 = \sigma_2^2 = n\cdot\tfrac{1}{6}\cdot\tfrac{5}{6} = \tfrac{5n}{36}.$$

所以

$$\begin{aligned}\mathrm{E}(Z_2) &= \tfrac{1}{2}\cdot\tfrac{n}{6} + \tfrac{1}{2}\cdot\tfrac{5n}{6} = \tfrac{n}{2},\\ \mathrm{Var}(Z_2) &= \tfrac{1}{2}\cdot\tfrac{5n}{36} + \tfrac{1}{2}\cdot\tfrac{5n}{36} + \tfrac{1}{4}\Big(\tfrac{4n}{6}\Big)^2 = \frac{5n}{36} + \frac{n^2}{9}.\end{aligned}$$

**想一想** 当 $n$ 很大时，$Z_2$ 的分布实际上是甚么样子?

$Z_2$ 有 $\frac{1}{2}$ 的概率接近 $\frac{n}{6}$，另有 $\frac{1}{2}$ 的概率接近 $\frac{5n}{6}$。这两个峰各自的标准差为 $\frac{\sqrt{5n}}{6}$，与两者之间的距离 $\frac{2n}{3}$ 相比极小。所以这个分布是双峰（bimodal）的：两个狭窄的峰，而在平均值 $\frac{n}{2}$ 附近几乎没有概率。正态分布（normal distribution）只在平均值处有单一个峰，所以在这里恰恰错了：它把最多的概率放在 $Z_2$ 几乎从不落入的地方。另外还有第二个问题。标准差约为 $\frac{n}{3}$，所以 $\mathrm{N}(\frac{n}{2}, \frac{n^2}{9})$ 给负值的概率约为 $\mathrm{P}(Z < -1.5) \approx 0.067$，给大于 $n$ 的值的概率也一样，但 $0 \le Z_2 \le n$ 总是成立。下图展示 $n = 60$ 时的两幅图像。




现在处理最后一部分。当 $0.45n \le Z_2 \le 0.55n$ 时，$Z_2$ 与其平均值相差不超过 10%。以硬币的结果为条件，

$$\mathrm{P}(0.45n \le Z_2 \le 0.55n) = \tfrac{1}{2}\,\mathrm{P}(0.45n \le X_1 \le 0.55n) + \tfrac{1}{2}\,\mathrm{P}(0.45n \le X_2 \le 0.55n).$$

对于第一项，$\mathrm{P}(0.45n \le X_1 \le 0.55n) \le \mathrm{P}(X_1 \ge 0.45n)$，而 $0.45n - \frac{n}{6} = \frac{17n}{60}$。利用正态近似（normal approximation）$X_1 \approx \mathrm{N}(\frac{n}{6}, \frac{5n}{36})$，其标准差为 $\frac{\sqrt{5n}}{6}$，得

$$\mathrm{P}(X_1 \ge 0.45n) \approx \mathrm{P}\Big(Z \ge \frac{17n/60}{\sqrt{5n}/6}\Big) = 1 - \Phi\Big(\frac{17\sqrt{n}}{10\sqrt{5}}\Big) \longrightarrow 1 - 1 = 0.$$

第二项是第一项的镜像：$X_2 = n - X_1$，所以 $0.45n \le X_2 \le 0.55n$ 当且仅当 $0.45n \le X_1 \le 0.55n$，因此第二项等于第一项。故此，$Z_2$ 与其平均值相差不超过 10% 的概率至多约为 $1 - \Phi\big(\frac{17\sqrt{n}}{10\sqrt{5}}\big)$，当 $n \to \infty$ 时趋向 0。

这一步其实并不依赖近似。切比雪夫（Chebyshev）不等式可准确地给出

$$\mathrm{P}\big(|X_1 - \tfrac{n}{6}| \ge \tfrac{17n}{60}\big) \le \frac{5n/36}{(17n/60)^2} = \frac{500}{289\,n} \longrightarrow 0.$$

## 验算

- 在 (i) 中，若 $\mu_1 = \mu_2$，额外的一项便会消失；若 $\sigma_1 = \sigma_2 = 0$，则 $Y$ 只取 $\mu_1$ 或 $\mu_2$ 两个值，方差为 $pq(\mu_1 - \mu_2)^2$，与经缩放的伯努利（Bernoulli）随机变量相同。

- 在 (ii) 中，$n = 400$ 时准确的二项概率为 0.9598，而正态近似值为 $2\Phi(2) - 1 = 0.9545$。$n = 1000$ 时，两者分别为 0.9986 和 0.9984。
- 在 (iii) 中，以电脑模拟 $n = 30$ 的 $Z_2$，得出平均值 15.00 及方差 104.2，而公式给出的是 15 和 104.17。当 $n = 100$ 时，与平均值相差不超过 10% 的准确概率约为 $3 \times 10^{-11}$。

## 常见错误

- 写出 $\mathrm{Var}(Y) = p\sigma_1^2 + q\sigma_2^2$，忘记了在两个平均值之间作选择本身就是变异的来源。
- 把 (iii) 中的 $Z_2$ 当作 (ii) 中的 $Z_1$ 处理。在 (ii) 中，每粒骰子都重新掷一次硬币；但在 (iii) 中，一次掷硬币便决定了全部 $n$ 粒骰子，所以各结果并不独立。
- 只用“方差很大”来解释拟合不佳。真正的问题在于形状：两个分开的峰，而平均值附近的概率极少。

## STEP 之外

第 (i) 部分是*全方差定律*（law of total variance）$\mathrm{Var}(Y) = \mathrm{E}\big(\mathrm{Var}(Y \mid C)\big) + \mathrm{Var}\big(\mathrm{E}(Y \mid C)\big)$ 最简单的情况，其中 $C$ 代表硬币。第 (ii) 部分是弱大数定律（weak law of large numbers），大学的入门课程会利用切比雪夫不等式严格证明它。第 (iii) 部分说明了为何这条定律需要独立性：只要一枚共用的硬币，便足以令它失效。像 $Z_2$ 这样的混合分布是统计建模的常用工具。Blitzstein 与 Hwang 所著的 *Introduction to Probability*（《概率导论》）以大量例子涵盖了以上三个概念。

详细解答

# STEP II 2023，第 12 题

概率与统计 · 最大值的分布

> **STEP II 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

若干个数的最大值不超过 $t$，当且仅当其中每一个都不超过 $t$，因此 $Y$ 的累积分布函数（cumulative distribution function）是单个 $X_i$ 的累积分布函数的幂。其余一切都源于使用累积分布函数而不是密度函数：中位数（median）可直接解方程 $\mathrm{P}(Y \le m) = \frac{1}{2}$ 得出，而均值（mean）则可通过分部积分（integration by parts）简洁地求得，使积分中出现的是分布函数而非密度函数。

## 解答

### 第 (i) 部分

记 $F(t) = \mathrm{P}(X_1 \le t)$。当 $0 \le t \le \pi$ 时，

$$F(t) = \int_0^t \tfrac{1}{2}\sin x \,\mathrm{d}x = \tfrac{1}{2}(1 - \cos t).$$

事件 $\{Y \le t\}$ 发生，当且仅当 $X_1 \le t$、$X_2 \le t$、…、$X_n \le t$ 全部发生：若最大者不超过 $t$，则其余各个亦不超过 $t$，反之亦然。各 $X_i$ 独立同分布（independent and identically distributed），故这个交集的概率等于各个概率之积：

$$\mathrm{P}(Y \le t) = \mathrm{P}(X_1 \le t)\,\mathrm{P}(X_2 \le t)\cdots\mathrm{P}(X_n \le t) = \big[\mathrm{P}(X_1 \le t)\big]^n.$$

因此，以 $G$ 记 $Y$ 的累积分布函数，

$$G(t) = \frac{(1-\cos t)^n}{2^n} \qquad (0 \le t \le \pi),$$

且当 $t<0$ 时 $G(t)=0$，当 $t>\pi$ 时 $G(t)=1$。用链式法则（chain rule）求导，得 $Y$ 的概率密度函数（probability density function）为

$$g(t)=\frac{n\,(1-\cos t)^{n-1}\sin t}{2^n}\quad \text{当} 0\le t\le \pi,$$

其他情况下为零。

**第 (ii) 部分**

中位数 $m=m(n)$ 满足 $G(m)=\frac{1}{2}$：

$$\left(\frac{1-\cos m}{2}\right)^n=\frac{1}{2}\iff \frac{1-\cos m}{2}=2^{-1/n}$$
$$\iff \cos m=1-2^{1-1/n},$$

由于 $(1-\cos m)/2\ge 0$，这里取正实数 $n$ 次方根。由于 $m$ 必须位于 $[0,\pi]$ 内，而余弦函数在此区间上是一一对应（one-to-one）的，

$$m(n)=\arccos\left(1-2^{1-1/n}\right).$$

当 $n$ 增大时，$1/n$ 减小，故 $2^{1-1/n}$ 增大（趋向 2），而 $1-2^{1-1/n}$ 减小（趋向 $-1$）。由于 arccos 是递减函数，$m(n)$ 随 $n$ 递增，且当 $n\to\infty$ 时 $m(n)\to\arccos(-1)=\pi$。（当 $n=1$ 时得 $m(1)=\arccos 0=\pi/2$，这正符合 $\frac{1}{2}\sin x$ 关于 $\pi/2$ 的对称性。）这是合理的：取越来越多个值的最大值，结果会被推向取值范围的顶端。

**第 (iii) 部分**

按定义，$\mu(n)=\int_0^\pi t\,g(t)\,\mathrm{d}t$。由于 $g=G'$，取 $u=t$ 及 $\mathrm{d}v=g(t)\,\mathrm{d}t$、$v=G(t)$ 进行分部积分：

$$\mu(n)=\Big[t\,G(t)\Big]_0^\pi-\int_0^\pi G(t)\,\mathrm{d}t$$
$$=\pi\cdot 1-0-\frac{1}{2^n}\int_0^\pi(1-\cos x)^n\,\mathrm{d}x,$$

其中用到 $G(\pi)=\left(\frac{1+1}{2}\right)^n=1$。这就是所求的公式。

> ? **想一想** 为甚么要用分部积分，而不是尝试直接计算 $\int t\,(1-\cos t)^{n-1}\sin t\,\mathrm{d}t$?

目标公式含有 $(1-\cos x)^n$，即 $2^nG(x)$，是分布函数而非密度函数。分部积分正是把 $t\,g(t)$ 换成 $G(t)$ 的一步，同时亦消去了麻烦的因子 $t$。

**(a)** 设 $h(x)=\frac{1}{2}(1-\cos x)$，则在 $[0,\pi]$ 上 $0\le h(x)\le 1$，且当 $0<x<\pi$ 时 $0<h(x)<1$。公式变为 $\mu(n)=\pi-\int_0^\pi h(x)^n\,\mathrm{d}x$，故

$$\mu(n+1)-\mu(n)=\int_0^\pi\left(h(x)^n-h(x)^{n+1}\right)\mathrm{d}x=\int_0^\pi h(x)^n\big(1-h(x)\big)\,\mathrm{d}x.$$

在 $0 < x < \pi$ 上，被积函数连续且严格为正，因为在该处 $h(x) > 0$ 且 $1 - h(x) > 0$。所以积分严格为正，对每个 $n \geq 1$ 都有 $\mu(n+1) > \mu(n)$：$\mu(n)$ 随 $n$ 递增。

**(b)** 取 $n = 2$，展开并利用 $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$：

$$\int_0^{\pi} (1 - \cos x)^2 \, \mathrm{d}x = \int_0^{\pi} \left(1 - 2\cos x + \tfrac{1}{2} + \tfrac{1}{2}\cos 2x\right) \mathrm{d}x = \tfrac{3}{2}\pi,$$

因为 $\cos x$ 和 $\cos 2x$ 在 $[0, \pi]$ 上的积分均为零。因此

$$\mu(2) = \pi - \tfrac{1}{4} \cdot \tfrac{3}{2}\pi = \frac{5\pi}{8}.$$

由第 (ii) 部分，$m(2) = \arccos\left(1 - \sqrt{2}\right)$。$5\pi/8$ 与 $m(2)$ 均位于 $[0, \pi]$ 内，而余弦函数在此区间上严格递减，故

$$\mu(2) < m(2) \iff \cos \tfrac{5\pi}{8} > \cos m(2) = 1 - \sqrt{2}.$$

**想一想** 怎样求得 $\cos \frac{5\pi}{8}$ 的准确值?

写成 $\cos \frac{5\pi}{8} = \cos\left(\frac{\pi}{2} + \frac{\pi}{8}\right) = -\sin \frac{\pi}{8}$，并利用半角公式（half-angle formula）$\sin^2 \frac{\pi}{8} = \frac{1}{2}\left(1 - \cos \frac{\pi}{4}\right) = \frac{2 - \sqrt{2}}{4}$。由于 $\sin \frac{\pi}{8} > 0$，

$$\cos \tfrac{5\pi}{8} = -\tfrac{1}{2}\sqrt{2 - \sqrt{2}}.$$

所以我们需要证明 $-\frac{1}{2}\sqrt{2 - \sqrt{2}} > 1 - \sqrt{2}$，即 $\sqrt{2} - 1 > \frac{1}{2}\sqrt{2 - \sqrt{2}}$。两边均为正数，故此式等价于两边平方后的不等式，再乘以 4：

$$\begin{aligned} 4\left(3 - 2\sqrt{2}\right) > 2 - \sqrt{2} &\iff 10 > 7\sqrt{2} \\ &\iff 100 > 98, \end{aligned}$$

其中最后一步是把两个正数平方。此式成立，而每一步都是等价变换，故 $\cos \frac{5\pi}{8} > 1 - \sqrt{2}$，因此 $\mu(2) < m(2)$。

## 验算

- 当 $n = 1$ 时，$Y = X_1$ 关于 $\pi/2$ 对称，而确实 $\mu(1) = \pi - \frac{1}{2}\int_0^{\pi}(1 - \cos x)\,\mathrm{d}x = \pi/2 = m(1)$。
- 数值上，$\mu(2) = 5\pi/8 \approx 1.9635$，$m(2) = \arccos(1 - \sqrt{2}) \approx 1.9979$，所以均值略低于中位数。对 $n$ 由 1 至 7 的每个值各以电脑模拟 $10^5$ 个最大值，结果与两条公式的误差均在 0.01 以内（例如 $\mu(3) \approx 2.1598$，$m(3) \approx 2.1986$），而且两个数列都递增。
- 当 $n = 2$ 时，$Y$ 的密度函数为 $\frac{1}{2}(1 - \cos t)\sin t$，属左偏（skewed to the left）分布（有一条伸向 0 的长尾），而左偏分布的均值通常低于中位数。

## 常见错误

- 忘记论证 $\mathrm{P}(Y \leq t) = [F(t)]^n$：既需要“全部都不超过 $t$”这个等价关系，也需要独立性。
- 做了分部积分，却忘记边界项 $\left[tG(t)\right]_0^{\pi} = \pi$，而开首的 $\pi$ 正是由此而来。
- 通过余弦比较角的大小时，没有指出余弦函数在 $[0, \pi]$ 上是递减的，结果把不等号方向弄反。

## STEP 之外

样本的最大值是顺序统计量（order statistic）的一个例子，而第 (i) 部分的技巧可以求出任何样本最大值的分布。恒等式 $\mathrm{E}[Y] = \pi - \int_0^{\pi} G(t)\,\mathrm{d}t$ 是非负随机变量的一般尾部积分公式（tail-integral formula）$\mathrm{E}[Y] = \int_0^{\infty} \mathrm{P}(Y > t)\,\mathrm{d}t$ 的特例，你在大学概率课程的初期就会遇到。这两个概念在 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）中均有深入讨论。

详细解答

# STEP III 2023，第 1 题

纯数学 · 圆锥曲线的切线

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

一条直线与圆相切（tangent），当且仅当圆心到该直线的垂直距离等于半径。因此，我们先用 $p$ 和 $q$ 写出弦（chord）$PQ$ 的方程，计算这个距离，然后看着相切条件化简成题目所给的关系式。之后，一切都归结为一条含两个参数的对称方程：最后一部分问的是，这个关系式对第三条边是否仍然成立，而这正是根之和与根之积（sums and products of roots）的用武之地。

## 解答

### 第 (i) 部分

$PQ$ 的斜率（gradient）是

$$\frac{aq^2 - ap^2}{2aq - 2ap} = \frac{p+q}{2},$$

由于两点相异，可以除以 $q - p \neq 0$。因此经过 $P$ 的直线是 $y - ap^2 = \frac{1}{2}(p+q)(x - 2ap)$，化简后得

$$(p+q)x - 2y - 2apq = 0.$$

圆心 $C(0, 3a)$ 到这条直线的距离是

$$\frac{|-6a - 2apq|}{\sqrt{(p+q)^2 + 4}} = \frac{2a|pq+3|}{\sqrt{(p+q)^2+4}}.$$

这条直线与圆相切，当且仅当此距离等于 $2a$，即（由于两边皆为正，可以平方，再约去 $4a^2 \neq 0$）当且仅当

$$(pq+3)^2 = (p+q)^2 + 4 \iff p^2q^2 + 6pq + 5 = (p+q)^2.$$

这正是 $(*)$，所以当 $(*)$ 成立时，直线 $PQ$ 是圆的切线。（每一步都是可逆的，所以逆命题（converse）亦成立，而我们在第 (iii) 部分会用到这一点。）

**第 (ii) 部分**

展开 $(*)$ 并按 $q$ 的幂次整理：

$$(1-p^2)q^2 - 4pq + (p^2-5) = 0.$$

由于 $p^2 \neq 1$，首项系数不为零，所以这确实是一条关于 $q$ 的二次方程。它的判别式（discriminant）是

$$\begin{aligned} 16p^2 - 4(1-p^2)(p^2-5) &= 16p^2 - 4(-p^4+6p^2-5) \\ &= 4(p^4-2p^2+5) \\ &= 4\big((p^2-1)^2+4\big) > 0, \end{aligned}$$

所以有两个相异实根 $q_1, q_2$。两者都不等于 $p$：把 $q=p$ 代入左边得 $-p^4-2p^2-5<0$。所以 $Q_1$ 和 $Q_2$ 确实是与 $P$ 相异的点。

由这条二次方程，

$$q_1+q_2 = \frac{4p}{1-p^2}, \qquad q_1q_2 = \frac{p^2-5}{1-p^2}.$$

**第 (iii) 部分**

取三角形 $PQ_1Q_2$，其中 $Q_i = (2aq_i, aq_i^2)$，而 $q_1, q_2$ 是第 (ii) 部分中的两个根。它的顶点全都在 $x^2 = 4ay$ 上，而且是三个相异点（如 (ii) 所证）；抛物线（parabola）上三个相异点必定不共线，因为一条直线与抛物线最多相交两次。所以这确实是一个三角形。

由第 (i) 部分，由于 $(p, q_1)$ 和 $(p, q_2)$ 满足 $(*)$，边 $PQ_1$ 和 $PQ_2$ 都与圆相切。

> ? **想一想** 如果我们不想明确解出 $q_1$ 和 $q_2$，怎样才能证明第三条边 $Q_1Q_2$ 也是切线?

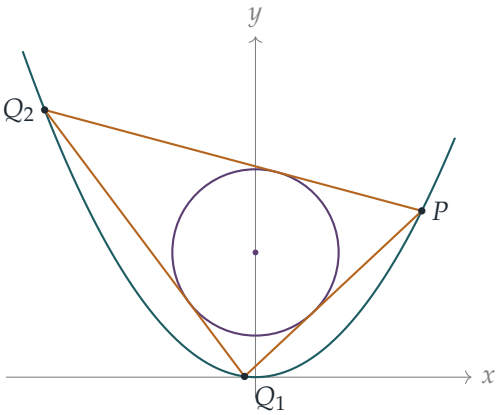
根据第 (i) 部分的可逆论证（把 $p, q$ 换成 $q_1, q_2$），$Q_1Q_2$ 是切线当且仅当

$$(q_1+q_2)^2 = q_1^2q_2^2 + 6q_1q_2 + 5,$$

而两边都只涉及 $s = q_1+q_2$ 和 $t = q_1q_2$，这两者我们已经知道。全式乘以 $(1-p^2)^2$。左边变成 $16p^2$。右边变成

$$\begin{aligned} &(p^2-5)^2 + 6(p^2-5)(1-p^2) + 5(1-p^2)^2 \\ &\quad = (p^4-10p^2+25) + (-6p^4+36p^2-30) + (5p^4-10p^2+5) \\ &\quad = 16p^2. \end{aligned}$$

两边相等，所以 $Q_1Q_2$ 也与圆相切，而 $PQ_1Q_2$ 就是所求的三角形。




图中显示 $a=1$、$p=2$ 的情况，此时 $q_{1,2}=\frac{1}{3}(-4\pm\sqrt{13})$。

## 验算

- 取 $a=1$、$p=2$：二次方程是 $3q^2+8q+1=0$，其中 $q_1+q_2=-\frac{8}{3}=\frac{4p}{1-p^2}$，$q_1q_2=\frac{1}{3}=\frac{p^2-5}{1-p^2}$。于是 $(q_1+q_2)^2=\frac{64}{9}$，而 $q_1^2q_2^2+6q_1q_2+5=\frac{1}{9}+2+5=\frac{64}{9}$，正如第 (iii) 部分所要求。
- 数值上，取 $q_1\approx-0.1315$、$q_2\approx-2.5352$，三条边所在直线到 $(0,3)$ 的距离都是 2.000；用电脑对数以千计的随机 $a$ 和 $p$ 作检验，结果吻合至约 $10^{-13}$。
- 被排除的情况 $p=\pm1$ 很有启发性：当 $p=1$ 时，二次方程退化为 $-4q-4=0$，只得出 $q=-1$。从 $P(2a,a)$ 出发那条“失踪”的切线是铅垂线 $x=2a$，它与圆相切，但与抛物线只相交于 $P$。

## 常见错误

- 在距离公式中忘记绝对值，或者平方时没有指出两边皆为正。
- 只用了第 (i) 部分的“若”方向：第 (iii) 部分需要逆命题，才能把 $Q_1Q_2$ 的相切转化为代数条件，所以要说明各步是可逆的。
- 没有检查 $q_1,q_2\neq p$，以确保三角形不是退化的。

## STEP 之外

这是庞斯莱闭合定理（Poncelet's porism）的一个例子：如果一个三角形可以内接于一条圆锥曲线（conic），同时外切于另一条圆锥曲线，那么这样的三角形可以从第一条圆锥曲线上的任何一点出发作出。这里的抛物线与圆正是这样的一对，而第 (iii) 部分用韦达定理（Vieta's formulas）直接证明了闭合性。一般情况的证明要用到射影几何（projective geometry），而最优雅的证明则用到椭圆曲线（elliptic curves）。Siklos 的 *Advanced Problems in Mathematics*（《数学高阶难题》）收录了更多这类参数化圆锥曲线几何的题目。

详细解答

# STEP III 2023，第 2 题

纯数学 · 极坐标曲线的面积

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

两条曲线都是由一条从 $\theta = 0$ 扫到 $\theta = \pi$ 的半径画出来的，所以整道题目的重点在于：在每个角度上，哪一条曲线离原点较远。比较 $k(1+\sin\theta)$ 与 $k+\cos\theta$，可见两者恰好在 $\theta = \alpha$ 处交换一次位置。于是 A 与 B 两个区域合起来，正是两条曲线所围区域的重叠部分；而所有面积都来自极坐标（polar coordinates）面积公式 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$，其中 $\sin^2$ 和 $\cos^2$ 用倍角公式（double angle formulae）处理。

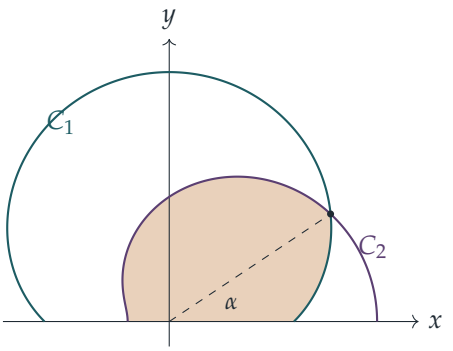
## 解答

### 第 (i) 部分

先算一些数值，有助作草图：

| | $\theta = 0$ | $\theta = \frac{\pi}{2}$ | $\theta = \pi$ |
|---|---|---|---|
| $C_1$: $r = k(1+\sin\theta)$ | $k$ | $2k$ | $k$ |
| $C_2$: $r = k+\cos\theta$ | $k+1$ | $k$ | $k-1$ |

由于 $k > 1$，两个半径在整个范围内都是正数。$C_1$ 是心脏线（cardioid）的上半部分，关于 $y$ 轴对称，最高到达 $(0, 2k)$。$C_2$ 是一个稍微压扁、“半径约为 $k$”的半圆，在 $\theta = 0$ 处向右推出 1 个单位，在 $\theta = \pi$ 处向内收入 1 个单位。下图取 $k = 1.5$；阴影部分是两条曲线所围区域的重叠部分。




在交点处两个半径相等，所以

$$k(1+\sin\alpha) = k+\cos\alpha \iff k\sin\alpha = \cos\alpha.$$

这里 $\cos\alpha \neq 0$（否则便需要 $k\sin\alpha = 0$，但此时 $\sin\alpha = \pm 1$），因此两边相除得 $\tan\alpha = \dfrac{1}{k}$，即所求。由于 $0 < \frac{1}{k} < 1$，在 $[0,\pi]$ 内恰好有一个这样的角，而且 $0 < \alpha < \frac{\pi}{4}$。

更有用的是，$k+\cos\theta - k(1+\sin\theta) = \cos\theta - k\sin\theta$，它在 $0 \leq \theta < \alpha$ 时为正，在 $\alpha < \theta \leq \pi$ 时为负（利用下面 $\sin\alpha$、$\cos\alpha$ 的值，它等于 $\sqrt{1+k^2}\,\sin(\alpha-\theta)$，而 $\alpha-\theta$ 分别落在 $(0,\alpha]$ 或 $[\alpha-\pi,0)$ 之内）。所以在 $\alpha$ 之前 $C_2$ 在 $C_1$ 之外，在 $\alpha$ 之后则在 $C_1$ 之内。

我们亦会用到由 $\tan\alpha = \frac{1}{k}$ 得出的

$$\sin\alpha = \frac{1}{\sqrt{1+k^2}}, \qquad \cos\alpha = \frac{k}{\sqrt{1+k^2}}.$$

**第 (ii) 部分**

区域 A 是在 $0 \leq \theta \leq \alpha$ 时，由伸到 $C_1$ 的半径扫出来的，所以其面积为

$$\begin{aligned}
\tfrac{1}{2}\int_0^\alpha k^2(1+\sin\theta)^2\,\mathrm{d}\theta &= \frac{k^2}{2}\int_0^\alpha \left(1+2\sin\theta+\tfrac{1}{2}-\tfrac{1}{2}\cos 2\theta\right)\mathrm{d}\theta \\
&= \frac{k^2}{2}\left[\tfrac{3}{2}\theta - 2\cos\theta - \tfrac{1}{4}\sin 2\theta\right]_0^\alpha \\
&= \frac{k^2}{2}\left(\tfrac{3}{2}\alpha - \tfrac{1}{2}\sin\alpha\cos\alpha - 2\cos\alpha + 2\right) \\
&= \frac{k^2}{4}(3\alpha - \sin\alpha\cos\alpha) + k^2(1-\cos\alpha),
\end{aligned}$$

这里用了 $\sin^2\theta = \frac{1}{2}(1-\cos 2\theta)$ 和 $\sin 2\alpha = 2\sin\alpha\cos\alpha$。这正是所求的表达式。

**第 (iii) 部分**

同样地，利用 $\cos^2\theta = \frac{1}{2}(1+\cos 2\theta)$，B 的面积为

$$\begin{aligned}\tfrac{1}{2}\int_\alpha^\pi (k+\cos\theta)^2\,\mathrm{d}\theta &= \tfrac{1}{2}\int_\alpha^\pi \left(k^2+\tfrac{1}{2}+2k\cos\theta+\tfrac{1}{2}\cos 2\theta\right)\mathrm{d}\theta\\ &= \tfrac{1}{2}\left[\left(k^2+\tfrac{1}{2}\right)\theta+2k\sin\theta+\tfrac{1}{4}\sin 2\theta\right]_\alpha^\pi\\ &= \frac{(2k^2+1)(\pi-\alpha)}{4}-k\sin\alpha-\tfrac{1}{4}\sin\alpha\cos\alpha.\end{aligned}$$

如果想消去 $\alpha$ 的三角函数，可用第 (i) 部分的数值：$k\sin\alpha = \dfrac{k}{\sqrt{1+k^2}}$，$\sin\alpha\cos\alpha = \dfrac{k}{1+k^2}$，所以

$$\text{B 的面积} = \frac{(2k^2+1)(\pi-\alpha)}{4}-\frac{k}{\sqrt{1+k^2}}-\frac{k}{4(1+k^2)}.$$

**第 (iv) 部分**

**想一想** A 与 B 合起来的区域是甚么形状？

由第 (i) 部分，当 $\theta<\alpha$ 时，A 的边界是位于内侧的曲线 $C_1$；当 $\theta>\alpha$ 时，B 的边界是位于内侧的曲线 $C_2$。所以 $R$ 是两个区域重叠部分的面积（即草图中的阴影部分）。这已经提示了答案：当 $k$ 很大时，$C_2$ 几乎完全位于 $C_1$ 之内，所以重叠部分应该差不多就是整个 $T$。

整个区域的面积可由同样的积分在 $[0,\pi]$ 上计算得出（$\sin\theta$ 和 $\sin 2\theta$ 项在两端都给出 0；$-2\cos\theta$ 给出 4）：

$$T = \tfrac{1}{2}\int_0^\pi (k+\cos\theta)^2\,\mathrm{d}\theta = \frac{\pi(2k^2+1)}{4},\qquad S = \tfrac{1}{2}\int_0^\pi k^2(1+\sin\theta)^2\,\mathrm{d}\theta = \frac{k^2(3\pi+8)}{4}.$$

当 $k\to\infty$ 时，$\alpha = \arctan\frac{1}{k}\to 0$。与其展开 A，不如为它找一个上界：在 $[0,\alpha]$ 上有 $r = k(1+\sin\theta)\le 2k$，所以 A 位于一个半径为 $2k$、圆心角为 $\alpha$ 的扇形（sector）之内；又因为 $\alpha<\tan\alpha = \frac{1}{k}$，

$$0<\text{A 的面积}\le \tfrac{1}{2}(2k)^2\alpha<2k.$$

因此 $\dfrac{\mathrm{A}}{T}<\dfrac{8k}{\pi(2k^2+1)}\to 0$。至于 B，把第 (iii) 部分的结果除以 $T$，

$$\frac{\mathrm{B}}{T} = \frac{\pi-\alpha}{\pi}-\frac{4\left(k\sin\alpha+\frac{1}{4}\sin\alpha\cos\alpha\right)}{\pi(2k^2+1)}\to 1-0=1,$$

因为 $\alpha\to 0$，而且 $0<k\sin\alpha+\frac{1}{4}\sin\alpha\cos\alpha<k+1$，与 $2k^2+1$ 相比微不足道。因此

$$\frac{R}{T} = \frac{\mathrm{A}}{T}+\frac{\mathrm{B}}{T}\to 1.$$

最后

$$\frac{R}{S}=\frac{R}{T}\cdot\frac{T}{S}=\frac{R}{T}\cdot\frac{\pi(2k^2+1)}{k^2(3\pi+8)}\to 1\cdot\frac{2\pi}{3\pi+8}=\frac{2\pi}{3\pi+8}.$$

## 验算

- 在 $\theta=0$ 处，$C_2$（$r=k+1$）在 $C_1$（$r=k$）之外；在 $\theta=\frac{\pi}{2}$ 处则在其内（$k<2k$），所以两条曲线必定在两者之间相交，与 $0<\alpha<\frac{\pi}{4}$ 一致。
- 取 $k=2$：$\alpha=\arctan\frac{1}{2}\approx 0.4636$，$\mathrm{A}\approx 1.4132$，$\mathrm{B}\approx 5.0309$，$T=\frac{9}{4}\pi\approx 7.0686$，所以 $R/T\approx 0.912$。直接对 $\frac{1}{2}r^2$ 作数值积分，结果与两条公式都吻合。
- 取 $k=50$，$R/T\approx 0.99987$，$R/S\approx 0.36061$，接近 $\frac{2\pi}{3\pi+8}\approx 0.36059$。
- 对极限作一个合理性检查：当 $k$ 极大时，$C_2$ 近似一个半径为 $k$ 的半圆（面积 $\frac{1}{2}\pi k^2$），而 $C_1$ 是一个固定的半心脏线放大 $k$ 倍，该半心脏线的面积为 $\frac{3\pi}{4}+2$；两者之比为 $\frac{2\pi}{3\pi+8}$。

## 常见错误

- 忘记极坐标面积公式中的因子 $\frac{1}{2}$，或把 $1+\sin\theta$ 的平方误写成 $1+\sin^2\theta$。
- 在第 (iv) 部分中，声称 $k^2(1-\cos\alpha)$“因为 $\alpha\to 0$ 所以趋于 0”：它乘了 $k^2$，事实上它趋于 $\frac{1}{2}$。真正重要的只是它相对于 $T$ 的大小。

## STEP 之外

求两个以极坐标形式给出的区域的重叠面积，是初尝大学一年级多变量微积分（multivariable calculus）中面积积分 $\iint r\,\mathrm{d}r\,\mathrm{d}\theta$ 的滋味；在那里，最难的部分总是描述区域本身，正如本题一样。第 (iv) 部分的极限论证（为某一项找上界，而不是精确计算它）正是实分析（real analysis）的风格；Abbott 的 *Understanding Analysis*（《理解分析》）是观摩这种风格如何被细致发展的好地方。

详细解答

# STEP III 2023，第 3 题

纯数学 · 三次方程的复数根

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

以 $a$ 为基准点量度，三点分别是 $sb\mathrm{i}$、$-sb\mathrm{i}$ 和 $b$。乘以 i 会把向量（vector）转一个直角，所以 $a$ 是底边的中点，而 $b$ 沿着对称轴方向。有了这个图像，第 (ii) 部分只需运用韦达定理（Vieta's formulas）：各根之和为零，由此可用 $a$ 表示 $b$，然后 $a$ 在 $p^3/q^2$ 中消去。第 (iii) 及 (iv) 部分把结果化为关于 $x = s^2 > 0$ 的单变量函数（function），从其图像可看出它最小可以是多少。

## 解答

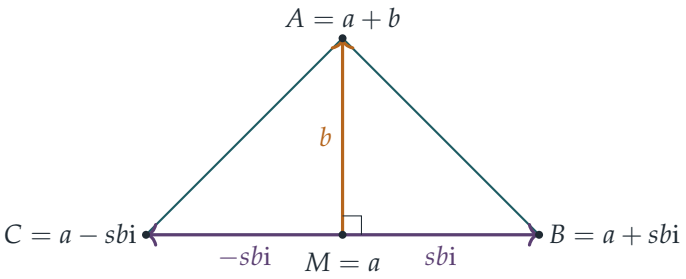
### 第 (i) 部分

由 $a + b$ 到另外两点的距离为

$$\big|(a + sb\mathrm{i}) - (a + b)\big| = |b|\,|s\mathrm{i} - 1| = |b|\sqrt{s^2 + 1},$$
$$\big|(a - sb\mathrm{i}) - (a + b)\big| = |b|\,|-s\mathrm{i} - 1| = |b|\sqrt{s^2 + 1}.$$

两者相等。我们还需要确认三点确实构成一个三角形。相对于 $a$，三点是 $sb\mathrm{i}$、$-sb\mathrm{i}$ 和 $b$。前两点互不相同（因为 $s > 0$ 且 $b \neq 0$），并且位于经过 $a$、方向为 $b\mathrm{i}$ 的直线上。由于 $b\mathrm{i}$ 是 $b$ 转了一个直角，而 $b \neq 0$，所以第三点 $a + b$ 不在该直线上。因此这三点构成一个等腰三角形（isosceles triangle），顶点为 $a + b$，底边由 $a - sb\mathrm{i}$ 到 $a + sb\mathrm{i}$，底边中点为 $a$。

**还原** $a$、$b$、$s$。设三角形的顶点为 $A$（即两条等边相交的顶点；若是等边三角形，任取一个顶点即可），其余两个顶点为 $B$、$C$。设 $M$ 为 $BC$ 的中点。




- 取 $a = M = \frac{1}{2}(B + C)$。
- 取 $b = A - M$。由于 $A$ 不在 $BC$ 上，这不等于零。
- 由于 $AB = AC$,$A$ 位于 $BC$ 的垂直平分线(perpendicular bisector)上,所以 $MB \perp MA$。因此 $B - M$ 是把 $b$ 转一个直角再伸缩所得：对某个实数 $\lambda \neq 0$，有 $B - M = \lambda b\mathrm{i}$。若 $\lambda < 0$，就把 $B$ 和 $C$ 的标记互换，这会改变 $\lambda$ 的正负号。于是 $s = \lambda > 0$ 为

$$s = \frac{MB}{MA} = \frac{\frac{1}{2}BC}{\text{三角形的高}}.$$

于是 $a + b = A$、$a + sb\mathrm{i} = B$ 及 $a - sb\mathrm{i} = M - (B - M) = C$，符合所求。

**第 (ii) 部分**

由第 (i) 部分，可把各根记为 $a + sb\mathrm{i}$、$a - sb\mathrm{i}$、$a + b$，其中 $b \neq 0$、$s > 0$。这个三次方程没有 $z^2$ 项，所以各根之和为零：

$$3a + b = 0, \qquad \text{所以} b = -3a,$$

而因为 $b \neq 0$，所以 $a \neq 0$。因此各根为

$$a(1 - 3s\mathrm{i}), \qquad a(1 + 3s\mathrm{i}), \qquad -2a.$$

前两个根之积为 $a^2(1 + 9s^2)$，和为 $2a$。所以由韦达定理，

$$p = a^2(1 + 9s^2) + (-2a)(2a) = a^2(9s^2 - 3) = 3a^2(3s^2 - 1),$$
$$q = -\big(\text{各根之积}\big) = -(-2a)\,a^2(1 + 9s^2) = 2a^3(9s^2 + 1).$$

由于 $a \neq 0$，所以 $q \neq 0$，可以相除：

$$\frac{p^3}{q^2} = \frac{27a^6(3s^2 - 1)^3}{4a^6(9s^2 + 1)^2} = \frac{27(3s^2 - 1)^3}{4(9s^2 + 1)^2},$$

其中 $s$ 是非零（事实上是正的）实数，符合所求。

**第 (iii) 部分**

设 $y = \dfrac{(3x-1)^3}{(9x+1)^2}$，定义域为 $x \neq -\frac{1}{9}$。由商法则（quotient rule），

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{9(3x-1)^2(9x+1)^2 - (3x-1)^3 \cdot 18(9x+1)}{(9x+1)^4} \\ &= \frac{9(3x-1)^2\big[(9x+1) - 2(3x-1)\big]}{(9x+1)^3} = \frac{27(3x-1)^2(x+1)}{(9x+1)^3}.\end{aligned}$$

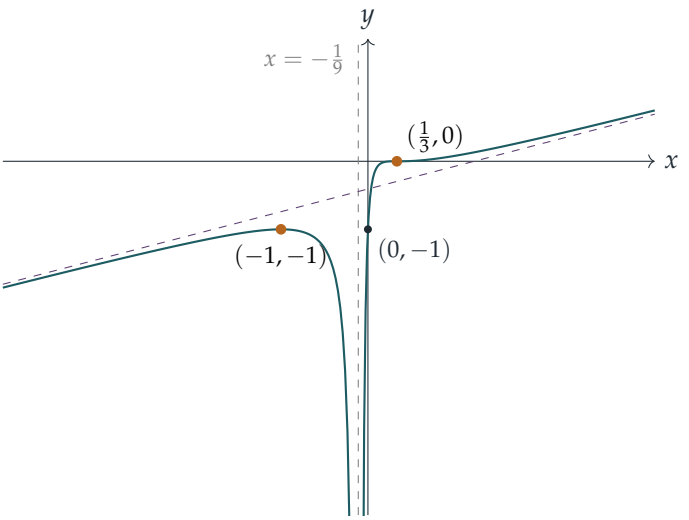
所以驻点（stationary point）位于 $x = -1$ 及 $x = \frac{1}{3}$。

- 在 $x = -1$：$y = (-4)^3/(-8)^2 = -1$。当 $x < -1$ 时，因子 $x+1$ 和 $(9x+1)^3$ 皆为负，所以 $y' > 0$。当 $-1 < x < -\frac{1}{9}$ 时，只有 $(9x+1)^3$ 为负，所以 $y' < 0$。因此 $(-1,-1)$ 是**局部极大点**（local maximum）。
- 在 $x = \frac{1}{3}$：$y = 0$。在其两侧（对 $x > -\frac{1}{9}$），除 $(3x-1)^2$ 外每个因子皆为正，而当 $x \neq \frac{1}{3}$ 时 $(3x-1)^2 > 0$。所以两侧皆有 $y' > 0$，因此 $(\frac{1}{3}, 0)$ 是**驻拐点**（stationary point of inflection）。

特别地，$y$ 在 $x > -\frac{1}{9}$ 上严格递增（strictly increasing），因为除单一点 $x = \frac{1}{3}$ 外，该处皆有 $y' > 0$。

其他特征：

- 垂直渐近线（vertical asymptote）$x = -\frac{1}{9}$。在该处分子趋向 $(-\frac{4}{3})^3 < 0$，分母趋向 $0^+$，所以从两侧皆有 $y \to -\infty$。
- 把 $27x^3 - 27x^2 + 9x - 1$ 除以 $81x^2 + 18x + 1$，得商式 $\frac{1}{3}x - \frac{11}{27}$，所以当 $x \to \pm\infty$ 时，$y = \frac{1}{3}x - \frac{11}{27}$ 是一条斜渐近线（oblique asymptote）。
- 曲线与坐标轴相交于 $(0,-1)$ 及 $(\frac{1}{3}, 0)$。




虚线是斜渐近线 $y = \frac{1}{3}x - \frac{11}{27}$。左支永不高于 $-1$。右支由 $-\infty$ 递增，经过每一个实数值。

### 第 (iv) 部分

由第 (ii) 部分，取该部分所得的实数 $s \neq 0$，并设 $x = s^2$，

$$\frac{p^3}{q^2} = \frac{27}{4} \cdot \frac{(3x-1)^3}{(9x+1)^2}.$$

由于 $s$ 是实数，右方是实数，因此 $p^3/q^2$ 亦是实数。

**想一想** 图像向下伸延至 $-\infty$，那么为甚么 $p^3/q^2$ 不能是很大的负数?

因为 $x = s^2$ 并非任意实数：$s$ 是非零实数，所以 $x > 0$。因此我们关心的 $x$ 值都位于右支上，而（由第 (iii) 部分）$y$ 在右支上严格递增。所以对 $x > 0$，

$$\frac{(3x-1)^3}{(9x+1)^2} > \frac{(0-1)^3}{(0+1)^2} = -1,$$

因此

$$\frac{p^3}{q^2} > -\frac{27}{4}.$$

这个不等式（inequality）是严格的，因为 $x = 0$（即 $s = 0$）被排除了：那会使三角形退化成一条线段。

## 验算

- 等边三角形：$z^3 - 1 = 0$ 的根构成一个等边三角形（equilateral triangle）。此时 $p = 0$，而公式给出 $p = 0$ 当且仅当 $s^2 = \frac{1}{3}$，即底边的一半等于高的 $\frac{1}{\sqrt{3}}$ 倍，这正是等边三角

形的形状。

- 一个具体例子：$a=1$、$s=1$ 给出根 $1\pm 3\mathrm{i}$ 及 $-2$，所以方程为 $z^3+6z+20=0$。于是 $p^3/q^2=216/400=\frac{27}{50}$，而公式给出 $\frac{27\cdot 8}{4\cdot 100}=\frac{27}{50}$。
- 当 $s\to 0$ 时，三角形压扁，趋向在 $a$ 处的二重根，而 $p^3/q^2\to -\frac{27}{4}$：$4p^3+27q^2=0$ 正是有重根（repeated root）的条件。当 $s=10^{-4}$ 时，其值约为 $-6.749998$。
- 以电脑检验了随机生成、形心（centroid）在原点的等腰三角形：$p^3/q^2$ 总是实数，且大于 $-6.75$。

## 常见错误

- 在第 (i) 部分把 $a$ 和 $b$ 当作实数。它们是复数，所以三角形可以朝任何角度摆放。
- 弄错 $q$ 的正负号：对首一三次多项式（monic cubic）而言，$q$ 是各根之积的负数。
- 在第 (iv) 部分使用整个图像，忘记了 $x=s^2>0$，而正是这一点使我们停留在右支上。

## STEP 之外

量 $-(4p^3+27q^2)$ 是这个三次多项式的判别式（discriminant），而 $p^3/q^2$ 是一个形状不变量（invariant）：当各根绕其形心旋转或伸缩时，它保持不变（把 $z$ 换成 $kz$，会使 $p$ 乘以 $k^2$、$q$ 乘以 $k^3$），所以它只记录了根所构成三角形的形状。这类不变量是在变量变换下对多项式进行分类的起点。Needham 的 *Visual Complex Analysis*（《直观复分析》）是一本出色的指南，教你像本题所要求那样，以几何方式思考复数。

详细解答

# STEP III 2023，第 4 题

纯数学 · 切比雪夫多项式的系数

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

把 $\cos\big((2n+1)\theta\big)$ 写成 $(\cos\theta+\mathrm{i}\sin\theta)^{2n+1}$ 的实部，就能把 p 化成一个明确的和式，而 p 的每个系数都变成二项式系数（binomial coefficient）之和。形如 $\sum\binom{N}{2r}$ 或 $\sum r\binom{N}{2r}$、只取偶数下标的和，可以把 $t=1$ 和 $t=-1$ 代入 $(1+t)^N$（或其导数（derivative）），再相加或相减求得，因为这样奇数下标的项会互相抵消。最后一部分则比较因式分解（factorisation）两边最高的几个系数。

## 解答

全文中记 $c=\cos\theta$ 及 $s=\sin\theta$。

### 第 (i) 部分

由棣美弗定理（de Moivre's theorem）及二项式定理（binomial theorem），

$$\cos\big((2n+1)\theta\big)+\mathrm{i}\sin\big((2n+1)\theta\big)=(c+\mathrm{i}s)^{2n+1}=\sum_{k=0}^{2n+1}\binom{2n+1}{k}c^{2n+1-k}(\mathrm{i}s)^k.$$

由于 $c$ 和 $s$ 是实数，实数项恰好是 $k$ 为偶数的那些项，设 $k=2r$，其中 $0\le r\le n$，此时 $(\mathrm{i}s)^{2r}=(-1)^r s^{2r}=(-s^2)^r=(c^2-1)^r$。取实部，得

$$\cos\big((2n+1)\theta\big)=\sum_{r=0}^{n}\binom{2n+1}{2r}\cos^{2n+1-2r}\theta\,(\cos^2\theta-1)^r,$$

即为所求。

**第 (ii) 部分**

由第 (i) 部分可知，对每个 $x = \cos\theta \in [-1,1]$，都有

$$\mathrm{p}(x) = 1 + \sum_{r=0}^{n} \binom{2n+1}{2r} x^{2n+1-2r}(x^2-1)^r$$

两边都是多项式（polynomial），且在无限多个点上相等，所以它们是同一个多项式。

第 $r$ 项的次数是 $(2n+1-2r)+2r = 2n+1$，首项系数（leading coefficient）为 $\binom{2n+1}{2r}$，而常数 1 不影响 $x^{2n+1}$。所以 p 中 $x^{2n+1}$ 的系数是

$$S_0 = \sum_{r=0}^{n} \binom{2n+1}{2r},$$

即偶数下标二项式系数之和。把 $t=1$ 和 $t=-1$ 代入 $(1+t)^{2n+1} = \sum_k \binom{2n+1}{k} t^k$，得

$$2^{2n+1} = \sum_k \binom{2n+1}{k}, \qquad 0 = \sum_k (-1)^k \binom{2n+1}{k}.$$

两式相加，$k$ 为奇数的项互相抵消，$k$ 为偶数的项则加倍：$2S_0 = 2^{2n+1}$。因此 $\mathrm{p}(x)$ 中 $x^{2n+1}$ 的系数是 $S_0 = 2^{2n}$。

**第 (iii) 部分**

在第 $r$ 项 $x^{2n+1-2r}(x^2-1)^r$ 中，要得到 $x^{2n-1}$，需要取 $(x^2-1)^r$ 的 $x^{2r-2}$ 项。该项是 $\binom{r}{1}(x^2)^{r-1}(-1) = -rx^{2r-2}$（当 $r=0$ 时此项不存在，与 $-r=0$ 一致）。由于 $n \geq 1$，有 $2n-1 \geq 1$，所以 p 中的常数 1 不起作用。因此 $x^{2n-1}$ 的系数是

$$-\sum_{r=0}^{n} r \binom{2n+1}{2r}.$$

? **想一想** 当每个偶数下标项都以 $r$ 加权时，怎样才能把它们挑出来?

微分会把下标作为因子拉下来。由 $(1+t)^{2n+1} = \sum_k \binom{2n+1}{k} t^k$，得

$$(2n+1)(1+t)^{2n} = \sum_{k=1}^{2n+1} k \binom{2n+1}{k} t^{k-1}.$$

在 $t=1$ 时，得 $\sum_k k\binom{2n+1}{k} = (2n+1)2^{2n}$。在 $t=-1$ 时，左边为 0（因为 $2n \geq 1$），所以

$$0 = \sum_k (-1)^{k-1} k \binom{2n+1}{k} = \sum_{k\text{ 奇数}} k \binom{2n+1}{k} - \sum_{k\text{ 偶数}} k \binom{2n+1}{k}.$$

所以 $k$ 为偶数和 $k$ 为奇数的两部分相等，各占总和的一半：

$$\sum_{k\text{ 偶数}} k \binom{2n+1}{k} = (2n+1)2^{2n-1}.$$

令 $k=2r$，这即是 $\sum_r 2r\binom{2n+1}{2r}=(2n+1)2^{2n-1}$，亦即 $\sum_r r\binom{2n+1}{2r}=(2n+1)2^{2n-2}$。因此 $\mathrm{p}(x)$ 中 $x^{2n-1}$ 的系数是 $-(2n+1)2^{2n-2}$，即为所求。

### 第 (iv) 部分

首项系数。由 (ii)，p 的次数是 $2n+1$，首项系数为 $2^{2n}$。若 q 的次数是 $m$、首项系数是 $a$，则 $(x+1)[\mathrm{q}(x)]^2$ 的次数是 $2m+1$，首项系数是 $a^2$。所以 $m=n$ 及 $a^2=2^{2n}$，又因 $a>0$，q 中 $x^n$ 的系数是 $2^n$。

关于 $p$ 的另一个事实。每一项 $x^{2n+1-2r}(x^2-1)^r$ 都只含 $x$ 的奇次幂，因为 $x^{2n+1-2r}$ 是奇次幂，而 $(x^2-1)^r$ 只有偶次幂。p 中唯一的偶次幂是常数 1。由于 $2n\geq 2$，p 中 $x^{2n}$ 的系数是 0。

比较系数。现设 $n\geq 2$，并写

$$\mathrm{q}(x)=2^n x^n+bx^{n-1}+cx^{n-2}+\cdots,$$

于是

$$[\mathrm{q}(x)]^2=2^{2n}x^{2n}+2^{n+1}b\,x^{2n-1}+\left(b^2+2^{n+1}c\right)x^{2n-2}+\cdots.$$

乘以 $x+1$ 后，$(x+1)[\mathrm{q}(x)]^2$ 中 $x^{2n}$ 和 $x^{2n-1}$ 的系数分别是

$$2^{n+1}b+2^{2n}\qquad 及\qquad \left(b^2+2^{n+1}c\right)+2^{n+1}b.$$

p 中 $x^{2n}$ 的系数是 0，所以 $b=-2^{2n}/2^{n+1}=-2^{n-1}$。于是 $b^2=2^{2n-2}$ 及 $2^{n+1}b=-2^{2n}$，再把 $x^{2n-1}$ 的系数与第 (iii) 部分的结果比较，得

$$2^{2n-2}+2^{n+1}c-2^{2n}=-(2n+1)2^{2n-2}.$$

整理得

$$2^{n+1}c=2^{2n-2}\big(-(2n+1)-1+4\big)=2^{2n-2}(2-2n)=2^{2n-1}(1-n),$$

所以 q 中 $x^{n-2}$ 的系数是

$$c=2^{n-2}(1-n),$$

即为所求。

## 验算

- 当 $n=1$：$\cos 3\theta=4\cos^3\theta-3\cos\theta$，所以 $\mathrm{p}(x)=4x^3-3x+1=(x+1)(2x-1)^2$。首项系数是 $2^2$，$x$ 的系数是 $-3=-3\cdot 2^0$，而 q 的首项系数是 $2^1$。
- 当 $n=2$：$\mathrm{p}(x)=16x^5-20x^3+5x+1=(x+1)(4x^2-2x-1)^2$。$x^3$ 的系数是 $-20=-5\cdot 2^2$，而 q 的常数项是 $-1=2^0(1-2)$。

- 当 $n=3$, 我们求得 $\mathrm{q}(x)=8x^3-4x^2-4x+1$, 其中 $x$ 的系数等于 $-4=2^1(1-3)$。我们亦用电脑验证过：对 $n=1,\ldots,12$, 精确展开 $\cos\big((2n+1)\theta\big)+1$, 再求 $\mathrm{p}(x)/(x+1)$ 的平方根，每条系数公式都成立。

## 常见错误

- 忘记 p 的 $x^{2n}$ 系数。它是 0，而正是这一点确定了 q 的 $x^{n-1}$ 系数 $b$，要求 $c$ 必须先知道它。
- 写成 $(\mathrm{i}s)^{2r}=s^{2r}$，漏掉了符号 $(-1)^r$。正是这个符号令结果是 $\cos^2\theta-1$ 而不是 $1-\cos^2\theta$。
- 在 (iii) 中只把 $t=1$ 代入导数，这样得到的是对所有 $k$ 的和，而不只是偶数项的和。

## STEP 之外

满足 $T_N(\cos\theta)=\cos N\theta$ 的多项式 $T_N$ 就是 $N$ 次的切比雪夫多项式（Chebyshev polynomial），而第 (ii) 和 (iii) 部分求出了它最高的两个系数。(iv) 中的因式分解反映了一个事实：$T_{2n+1}(x)+1$ 在 $x=-1$ 处有一个单根（simple root），而在其他每个使 $\cos\big((2n+1)\theta\big)=-1$ 的点处都有一个重根（double root）。切比雪夫多项式在逼近理论（approximation theory）中处于核心地位：在所有 $N$ 次首一多项式（monic polynomial）之中, $2^{1-N}T_N$ 在 $[-1,1]$ 上的最大绝对值最小。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）有更多同类题目。

详细解答

# STEP III 2023，第 5 题

纯数学 · 数论

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

每一部分都把一条整数方程化为一个乘积（product），令它等于一个我们完全清楚其因数（factor）的数：先是 49，然后是 $pq$，再然后是 $3p^2q$。一旦方程的一边是两个整数的乘积，每个因数的大小（哪一个较大、哪一个小于 $p$）就能把可能情况缩减至一两种。在第 (iii) 部分，我们能控制的因数是 $p+q-n$，而题目要求证明的不等式（inequality），正正就是阻止它与 $p$ 或 $q$ 有公因数的条件。

## 解答

### 第 (i) 部分

两边乘以 $7xy$（两者皆非零），得

$$7y+14x=2xy \quad \Longleftrightarrow \quad 2xy-14x-7y=0.$$

由于 $(2x-7)(y-7)=2xy-14x-7y+49$，两边加上 49 便得 $(2x-7)(y-7)=49$，此即所求。

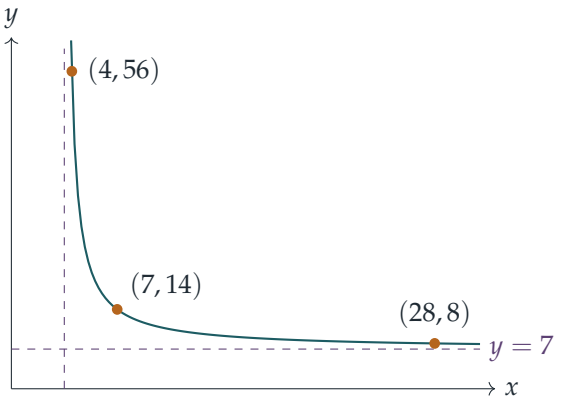
现设 $x$ 和 $y$ 为正整数。那么 $2x-7$ 和 $y-7$ 是乘积为 49 的整数，所以它们是 49 的一组因数对（factor pair）：

$$(2x-7,\ y-7)\in\{(1,49),\,(7,7),\,(49,1),\,(-1,-49),\,(-7,-7),\,(-49,-1)\}.$$

正数条件排除了负数的因数对：$x\geq 1$ 迫使 $2x-7\geq -5$，而 $y\geq 1$ 迫使 $y-7\geq -6$，所以两个因数都不可能是 $-7$ 或 $-49$，而每一组负数因数对都含有其中之一。正数的因数对给出

$$(x,y)=(4,56),\quad (7,14),\quad (28,8),$$

而每一组都确实满足原方程，例如 $\frac{1}{4}+\frac{2}{56}=\frac{7}{28}+\frac{1}{28}=\frac{2}{7}$。




这条曲线是一条直角双曲线（rectangular hyperbola），渐近线（asymptote）为 $x=\frac{7}{2}$ 和 $y=7$；三个解就是它正分支上的格点（lattice point）。

**第 (ii) 部分**

展开得 $(p+q)^2=p^2+2pq+q^2=n^2+pq$，所以

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

右方为正数，而 $p+q+n>0$，所以 $p+q-n$ 是一个正整数。$pq$ 的正因数是 $1,p,q,pq$（若 $p=q$ 则只有 $1,p,p^2$），而 $p+q+n$ 是其中之一。但 $n>0$ 表示 $p+q+n>p$ 及 $p+q+n>q$，所以 $p+q+n$ 既不是 $p$ 也不是 $q$；剩下的因数只有 $pq$。因此 $p+q+n=pq$，而另一个因数（co-factor）是

$$p+q-n=1, \qquad \text{即} \qquad p+q=n+1.$$

把 $n=p+q-1$ 代入 $p+q+n=pq$，得 $2p+2q-1=pq$，整理后得

$$pq-2p-2q+4=3 \iff (p-2)(q-2)=3.$$

质数（prime）满足 $p-2\geq 0$，所以两个因数皆非负，而乘积为 3，故 $\{p-2,q-2\}=\{1,3\}$。因此

$$(p,q)=(3,5) \quad \text{或} \quad (5,3),$$

而事实上 $9+15+25=49=7^2$，且 $n=7=3+5-1$。

**第 (iii) 部分**

这里 $n$ 是 $p^3+q^3+3pq^2$ 的正实数立方根（cube root）。

不等式。由于 $p,q>0$,

$$n^3=q^3+\left(p^3+3pq^2\right)>q^3,\qquad n^3=p^3+\left(q^3+3pq^2\right)>p^3,$$

而立方函数是严格递增（strictly increasing）的，所以 $n>q$ 及 $n>p$。整理后分别得 $p+q-n<p$ 及 $p+q-n<q$。我们亦需要下界（lower bound）：

$$(p+q)^3=p^3+3p^2q+3pq^2+q^3=n^3+3p^2q>n^3,$$

所以 $p+q>n$，即 $p+q-n>0$。

没有质数解。假设 $p$ 和 $q$ 是质数，而 $n$ 是整数。设 $d=p+q-n$，这是一个整数，且

$$0<d<p,\qquad 0<d<q.$$

由上面的展开式，$(p+q)^3-n^3=3p^2q$，再用立方差（difference of cubes）因式分解，

$$d\left((p+q)^2+(p+q)n+n^2\right)=3p^2q,$$

所以 $d$ 整除 $3p^2q$。

**想一想** 为甚么 $d$ 不能与 $p$ 或 $q$ 有公因数?

因为 $p$ 是质数，$d$ 与 $p$ 有公因数的唯一可能，是 $p$ 整除 $d$，而这需要 $d\geq p$。但 $d<p$。同理，$d<q$ 表示 $q$ 不整除 $d$。所以 $d$ 与 $p^2q$ 互质（coprime），而 $d\mid 3p^2q$ 便迫使 $d\mid 3$。这正是那些不等式的用处。因此 $d=1$ 或 $d=3$。设 $s=p+q$。

情况 $d=1$。此时 $n=s-1$，且

$$3p^2q=s^3-(s-1)^3=3s^2-3s+1,$$

这是不可能的：左方是 3 的倍数，而右方除以 3 的余数是 1。

情况 $d=3$。此时 $3=d<p$ 及 $3<q$，所以 $p$ 和 $q$ 都不是 3。这里 $n=s-3$，且

$$3p^2q=s^3-(s-3)^3=9s^2-27s+27,\qquad \text{所以}\qquad p^2q=3(s^2-3s+3).$$

因此 $3\mid p^2q$，而由于 3 是质数，它整除 $p$ 或 $q$，使其中一个等于 3。这与 $p,q>3$ 矛盾。

两种情况都不可能，所以不存在质数 $p$ 和 $q$ 使得 $p^3+q^3+3pq^2$ 是某整数的立方。 ■

## 验算

对所有正整数 $x,y<2000$ 进行穷举搜索（brute-force search），恰好找到第 (i) 部分的三组解。在所有质数 $p,q<1500$ 之中，$p^2+pq+q^2$ 只在 $\{p,q\}=\{3,5\}$ 时才是完全平方数（perfect square）（得 49），而 $p^3+q^3+3pq^2$ 从来不是完全立方数。随机抽取的正实数确认了第 (iii) 部分中的 $0<p+q-n<\min(p,q)$。第 (ii) 部分的快速人手验算：$(3+5+7)(3+5-7)=15\times 1=15=pq$。

## 常见错误

- 在第 (i) 部分忘记 49 的负数因数对。它们在这里不产生任何解，但必须把它们排除，而不是置之不理。
- 在第 (ii) 部分，未经论证便假设 $p+q-n$ 是较小的因数且为正数：两者都由 $n>0$ 及 $pq>0$ 推出，而且需要写明。
- 在第 (iii) 部分，停在 $d \mid 3$。$d=1$ 和 $d=3$ 两种情况仍然要排除，而第二种情况需要再次用到 $d<p,q$ 这个事实。

## STEP 之外

每一部分都用了同一个技巧：把一条丢番图方程（Diophantine equation）改写成若干整数的乘积，令它等于一个已知其因数分解的数，然后利用大小和互质性确定各个因数。第 (i) 部分属于埃及分数（Egyptian fraction）方程 $\frac{1}{x}+\frac{1}{y}=\frac{1}{n}$ 这一类，而第 (ii) 部分则关于含有一个 $120^\circ$ 角的整数边三角形，因为 $p^2+pq+q^2=n^2$ 正是该角的余弦定律（cosine rule）。Davenport 的 *The Higher Arithmetic*（《高等算术》）是这类初等数论（elementary number theory）的一本绝佳入门书。

详细解答

# STEP III 2023，第 6 题

纯数学 · 双曲函数与单调性

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

两个函数（function）都是 $\tan^{-1} x$ 与 $\tanh x$ 的某个倍数之差，所以它们的导数（derivative）都是 $\dfrac{1}{1+x^2}$ 与 $\operatorname{sech}^2 x$ 的某个倍数之差。因此一切都归结为比较 $\cosh^2 x$ 与 $1+x^2$：(i) 中的不等式（inequality）决定了 $\mathrm{f}'$ 的正负号；而证明比值 $\cosh^2 x/(1+x^2)$ 是递增的，就能准确告诉我们它可以等于 $\frac{1}{2}\pi$ 多少次，而这正是 $\mathrm{g}'$ 等于零的地方。

## 解答

### 第 (i) 部分

利用 $\cosh^2 x = \frac{1}{2}(1+\cosh 2x)$ 及 $\cosh 2x = \frac{1}{2}(\mathrm{e}^{2x}+\mathrm{e}^{-2x})$，由 $\mathrm{e}^x$ 的麦克劳林级数（Maclaurin series）可得（奇次幂互相抵消）

$$\cosh 2x = \sum_{n=0}^{\infty} \frac{(2x)^{2n}}{(2n)!} = 1 + 2x^2 + \frac{2}{3}x^4 + \cdots.$$

每一项都是非负的，所以 $\cosh 2x \geq 1 + 2x^2$，因此

$$\cosh^2 x = \tfrac{1}{2}(1+\cosh 2x) \geq \tfrac{1}{2}(2+2x^2) = 1 + x^2,$$

等号只在 $x = 0$ 时成立（当 $x \neq 0$ 时，$x^4$ 项严格为正）。

现在求导：

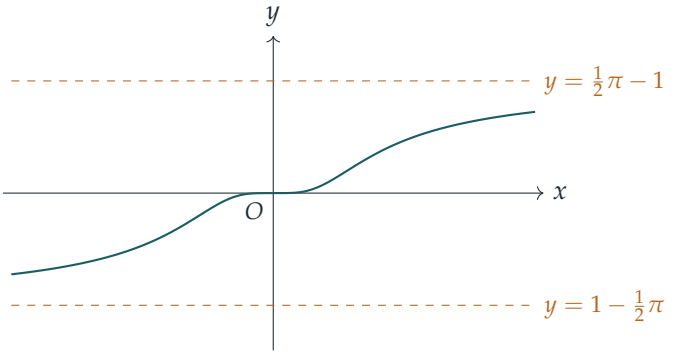
$$\mathrm{f}'(x) = \frac{1}{1+x^2} - \operatorname{sech}^2 x = \frac{1}{1+x^2} - \frac{1}{\cosh^2 x}.$$

由于 $\cosh^2 x \geq 1 + x^2 > 0$，取倒数得 $\dfrac{1}{\cosh^2 x} \leq \dfrac{1}{1+x^2}$，所以对所有 $x$ 都有 $\mathrm{f}'(x) \geq 0$，等号只在 $x = 0$ 处成立。一个除单一点外导数皆为正的函数是（严格）递增（increasing）

的，得证。

草图。图像上应标示的一些特征：

- f 是奇函数（odd function），因为 $\tan^{-1}$ 和 tanh 都是奇函数；而且 $\mathrm{f}(0) = 0$。
- $\mathrm{f}'(0) = 0$ 但 f 是递增的，所以原点是一个属于拐点（point of inflection）的驻点（stationary point）。事实上 $\tan^{-1} x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \cdots$ 而 $\tanh x = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \cdots$，所以在 0 附近 $\mathrm{f}(x) \approx \frac{1}{15}x^5$：图像在那里非常平坦。
- 当 $x \to \infty$ 时，$\tan^{-1} x \to \frac{1}{2}\pi$ 而 $\tanh x \to 1$，所以 $\mathrm{f}(x) \to \frac{1}{2}\pi - 1 \approx 0.571$，且从下方趋近；由奇性，当 $x \to -\infty$ 时 $\mathrm{f}(x) \to 1 - \frac{1}{2}\pi$。




**第 (ii)(a) 部分**

此时

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

它是连续（continuous）的偶函数（even function）。在 $x = 0$ 处，$\mathrm{g}'(0) = 1 - \frac{1}{2}\pi < 0$。在 $x = 2$ 处，

$$\mathrm{g}'(2) = \tfrac{1}{5} - \frac{\pi}{2\cosh^2 2},$$

而 $\cosh 2 > \frac{1}{2}\mathrm{e}^2 > 3.6$，所以 $\cosh^2 2 > 12.9$ 及 $\dfrac{\pi}{2\cosh^2 2} < \dfrac{1.6}{12.9} < \frac{1}{5}$；因此 $\mathrm{g}'(2) > 0$。由介值定理（intermediate value theorem），$\mathrm{g}'$ 在 $(0,2)$ 内有一个零点；又因 $\mathrm{g}'$ 是偶函数，它在 $(-2,0)$ 内还有另一个零点。所以 g 至少有两个驻点。

**第 (ii)(b) 部分**

设 $h(x) = (1+x^2)\sinh x - x\cosh x$。则

$$\begin{aligned} h'(x) &= 2x\sinh x + (1+x^2)\cosh x - \cosh x - x\sinh x \\ &= x\sinh x + x^2\cosh x. \end{aligned}$$

当 $x \geq 0$ 时，$x\sinh x$ 和 $x^2\cosh x$ 都是非负的（当 $x > 0$ 时更是正的），所以 $h$ 在 $[0,\infty)$ 上递增。由于 $h(0) = 0$，对 $x \geq 0$ 得 $h(x) \geq 0$；事实上，对 $x > 0$ 有 $h(x) > 0$。

**第 (ii)(c) 部分**

由商法则（quotient rule），

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\cosh^2 x}{1+x^2}\right) = \frac{2\cosh x\sinh x\,(1+x^2) - 2x\cosh^2 x}{(1+x^2)^2} = \frac{2\cosh x\,h(x)}{(1+x^2)^2}.$$

当 $x \geq 0$ 时，$\cosh x > 0$，而由 (b) 得 $h(x) \geq 0$，所以导数非负；它只在 $x = 0$ 处为零。因此当 $x \geq 0$ 时，$\cosh^2 x/(1+x^2)$ 是（严格）递增的。

**第 (ii)(d) 部分**

> **想一想** 关于 $\cosh^2 x/(1+x^2)$ 的命题，怎样可以用来数出 $\mathrm{g}'$ 的零点个数?

将 $\mathrm{g}'(x) = 0$ 两边乘以正数 $(1+x^2)\cosh^2 x$，可见

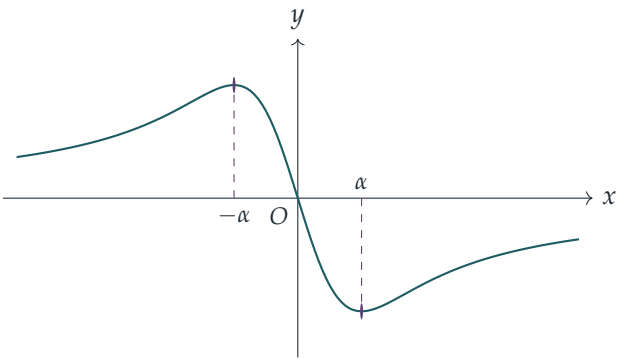
$$\mathrm{g}'(x) = 0 \iff \frac{\cosh^2 x}{1+x^2} = \frac{\pi}{2}.$$

记 $R(x) = \cosh^2 x/(1+x^2)$。在 $[0,\infty)$ 上，$R$ 连续，且由 (c) 知其严格递增，又 $R(0) = 1 < \frac{1}{2}\pi$，所以 $R$ 在该区间上最多只取值 $\frac{1}{2}\pi$ 一次；由 (a)，它确实取到此值，取值点为某个 $\alpha$，其中 $0 < \alpha < 2$。所以 $\mathrm{g}'$ 在 $[0,\infty)$ 内恰有一个零点。由于 $R$ 是偶函数，$x < 0$ 时的零点正是 $x > 0$ 时零点的相反数，因此恰好还有一个零点，位于 $-\alpha$。所以 g 恰有两个驻点，位于 $x = \pm\alpha$（数值上 $\alpha \approx 1.406$）。

**第 (ii)(e) 部分**

综合已知事实：

- g 是奇函数，$\mathrm{g}(0) = 0$，且 $\mathrm{g}'(0) = 1 - \frac{1}{2}\pi < 0$。
- 由 (d)，在 $(-\alpha,\alpha)$ 上 $\mathrm{g}' < 0$（因为那里 $R < \frac{1}{2}\pi$），而当 $|x| > \alpha$ 时 $\mathrm{g}' > 0$。所以在 $x = -\alpha$ 处有局部极大值（local maximum），在 $x = \alpha$ 处有局部极小值（local minimum），且 $\mathrm{g}(\alpha) = -\mathrm{g}(-\alpha) < 0$（约为 $-0.44$）。
- 当 $x \to \infty$ 时，$\mathrm{g}(x) \to \frac{1}{2}\pi - \frac{1}{2}\pi\cdot 1 = 0$。由于 g 在 $(\alpha,\infty)$ 上递增并趋向极限 0，它在那里保持为负：图像从下方趋近 $x$ 轴。由奇性，当 $x \to -\infty$ 时 $\mathrm{g}(x) \to 0$，且从上方趋近。
- 所以 g 唯一的零点在原点：当 $x < 0$ 时 $\mathrm{g} > 0$，当 $x > 0$ 时 $\mathrm{g} < 0$。




## 验算

- 在 $x = 1$ 处：$\cosh^2 1 \approx 2.381 \geq 2$，与 (i) 一致。
- $\mathrm{f}(30) \approx 0.537$，从下方趋近 $\frac{1}{2}\pi - 1 \approx 0.571$；而 $\mathrm{f}(0.1) \approx 6.6 \times 10^{-7}$，接近 $\frac{1}{15}(0.1)^5$。
- 数值求解 $\cosh^2 x = \frac{1}{2}\pi(1 + x^2)$ 得 $\alpha \approx 1.406$，正如 (a) 所预测般位于 $(0, 2)$ 内，且 $\mathrm{g}(\alpha) \approx -0.440$。在 $[-20, 20]$ 上对 $\mathrm{g}'$ 作精细扫描，恰好找到两次变号；而 $\mathrm{g}(20) \approx -0.050$ 是一个很小的负数，与草图相符。

## 常见错误

- 忘记 $\cosh^2 x \geq 1 + x^2$ 必须反转为 $\operatorname{sech}^2 x \leq \dfrac{1}{1 + x^2}$；这样做之所以成立，是因为两边都是正数。
- 在 (d) 中只证明了有一个正根，却忘记利用偶性（或指出 $x = 0$ 不是根）来数出那个负根。
- 画 g 的草图时，用了 $y = 0$ 以外的渐近线（asymptote），或在 $x$ 很大时再次穿过坐标轴；$(\alpha, \infty)$ 上的单调性（monotonicity）排除了这种情况。

## STEP 之外

(ii) 中的方法（把“$\mathrm{g}' = 0$ 有多少个解？”转化为“一个单调函数取某个值多少次？”）是分析学家数根的基本方法：连续性加上介值定理给出存在性，严格单调性则给出唯一性。这些想法会在第一门分析（analysis）课程中得到严谨处理；Abbott 的 *Understanding Analysis*（《理解分析》）是观看这些论证被仔细展开的绝佳之选。

详细解答

# STEP III 2023，第 7 题

纯数学 · 积分不等式

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

连续函数（continuous function）的平方的积分永不为负，而且只有当该函数恒等于零（identically zero）时，积分才等于零。每一部分都给我们一条涉及 $\int(\text{某式})^2$ 的等式，稍作整理后，就会变成“一个完全平方的积分等于零”。第 (i) 部分为第 (ii) 部分提供工具，第 (iii) 部分由分部积分（integration by parts）担当同样角色，而在第 (iv) 部分，我们要自己通过配方（completing the square）把平方找出来。

## 解答

### 第 (i) 部分

在左边的积分中代入 $x = u^2$，于是 $\mathrm{d}x = 2u\,\mathrm{d}u$。当 $x$ 由 0 变到 1 时，$u \geq 0$ 同样由 0 变到 1，而且 $\sqrt{x} = u$。因此

$$\int_0^1 \mathrm{f}(\sqrt{x})\,\mathrm{d}x = \int_0^1 \mathrm{f}(u)\,2u\,\mathrm{d}u = 2\int_0^1 x\,\mathrm{f}(x)\,\mathrm{d}x,$$

最后一步只是把哑变量（dummy variable）改名。这正是所求的结果。

### 第 (ii) 部分

在第 (i) 部分中取 $\mathrm{f} = \mathrm{g}$，已知条件就变成

$$\int_0^1 \big(\mathrm{g}(x)\big)^2\,\mathrm{d}x = 2\int_0^1 x\,\mathrm{g}(x)\,\mathrm{d}x - \frac{1}{3}.$$

现在展开目标积分，并利用 $\int_0^1 x^2\,\mathrm{d}x = \frac{1}{3}$：

$$\begin{aligned}\int_0^1 \big(\mathrm{g}(x)-x\big)^2\,\mathrm{d}x &= \int_0^1 \big(\mathrm{g}(x)\big)^2\,\mathrm{d}x - 2\int_0^1 x\,\mathrm{g}(x)\,\mathrm{d}x + \int_0^1 x^2\,\mathrm{d}x\\ &= -\frac{1}{3}+\frac{1}{3} = 0,\end{aligned}$$

得证。

> **想一想** 为甚么积分为零就能迫使 $\mathrm{g}(x)=x$ 处处成立，而不只是“大致上”成立？

设 $\phi(x)=\big(\mathrm{g}(x)-x\big)^2$。它是连续的（因为 g 连续），而且 $\phi(x)\ge 0$。假设对某个 $c\in[0,1]$ 有 $\phi(c)>0$。由连续性，存在一个包含 $c$、长度为某正数 $\delta$ 的区间 $I\subseteq[0,1]$，在其上 $\phi(x)>\frac{1}{2}\phi(c)$。由于在其他地方 $\phi\ge 0$，

$$\int_0^1 \phi(x)\,\mathrm{d}x \;\ge\; \int_I \phi(x)\,\mathrm{d}x \;\ge\; \tfrac{1}{2}\phi(c)\,\delta > 0,$$

这是矛盾。所以对 $0\le x\le 1$ 中的每一个 $x$，都有 $\phi(x)=0$，即 $\mathrm{g}(x)=x$。（连续性确实重要：一个只在单一点与 $x$ 不同的函数，积分仍然会是零。）

### 第 (iii) 部分

$2\mathrm{h}(1)-2\int_0^1 \mathrm{h}(x)\,\mathrm{d}x$ 这一项应该令我们想起分部积分。把 $1\cdot\mathrm{h}(x)$ 分部积分，取 $x$ 为 1 的原函数（antiderivative），

$$\int_0^1 \mathrm{h}(x)\,\mathrm{d}x = \Big[x\,\mathrm{h}(x)\Big]_0^1 - \int_0^1 x\,\mathrm{h}'(x)\,\mathrm{d}x = \mathrm{h}(1) - \int_0^1 x\,\mathrm{h}'(x)\,\mathrm{d}x.$$

所以 $2\mathrm{h}(1)-2\int_0^1 \mathrm{h}(x)\,\mathrm{d}x = 2\int_0^1 x\,\mathrm{h}'(x)\,\mathrm{d}x$，条件变为

$$\int_0^1 \big(\mathrm{h}'(x)\big)^2\,\mathrm{d}x = 2\int_0^1 x\,\mathrm{h}'(x)\,\mathrm{d}x - \frac{1}{3}.$$

这恰好就是第 (ii) 部分的条件（即我们经第 (i) 部分得到的形式），只是以 $\mathrm{h}'$ 代替 g。重复同一论证（这需要 $\mathrm{h}'$ 连续，正如题目的本意），

$$\int_0^1 \big(\mathrm{h}'(x)-x\big)^2\,\mathrm{d}x = 0 \quad\Longrightarrow\quad \mathrm{h}'(x)=x \text{ 对于} 0\le x\le 1.$$

积分得 $\mathrm{h}(x)=\frac{1}{2}x^2+C$，而 $\mathrm{h}(0)=0$ 给出 $C=0$。因此

$$\mathrm{h}(x)=\tfrac{1}{2}x^2.$$

### 第 (iv) 部分

首先留意 $a\neq 0$，因为条件中含有 $1/a$。对照前面各部分，我们寻找一个形如 $\mathrm{e}^{ax}\big(\mathrm{k}(x)-\mathrm{p}(x)\big)^2$ 的平方，使其交叉项（cross term）为 $-2\mathrm{k}(x)$。展开后交叉项为 $-2\mathrm{e}^{ax}\mathrm{p}(x)\mathrm{k}(x)$，

所以需要 $\mathrm{p}(x)=\mathrm{e}^{-ax}$：

$$\mathrm{e}^{ax}\big(\mathrm{k}(x)-\mathrm{e}^{-ax}\big)^2=\mathrm{e}^{ax}\big(\mathrm{k}(x)\big)^2-2\mathrm{k}(x)+\mathrm{e}^{-ax}.$$

在 $[0,1]$ 上积分，前两项用已知条件，最后一项用 $\int_0^1 \mathrm{e}^{-ax}\,\mathrm{d}x=\dfrac{1-\mathrm{e}^{-a}}{a}$：

$$\begin{aligned}\int_0^1 \mathrm{e}^{ax}\big(\mathrm{k}(x)-\mathrm{e}^{-ax}\big)^2\,\mathrm{d}x&=\frac{\mathrm{e}^{-a}}{a}-\frac{1}{a^2}-\frac{1}{4}+\frac{1-\mathrm{e}^{-a}}{a}\\&=\frac{1}{a}-\frac{1}{a^2}-\frac{1}{4}=-\Big(\frac{1}{a}-\frac{1}{2}\Big)^2=-\frac{(a-2)^2}{4a^2}.\end{aligned}$$

左边是一个非负连续函数的积分，所以 $\geq 0$；右边则 $\leq 0$。因此两边都必须为零，迫使 $(a-2)^2=0$，所以 $a=2$。

当 $a=2$ 时，连续非负函数 $\mathrm{e}^{2x}\big(\mathrm{k}(x)-\mathrm{e}^{-2x}\big)^2$ 的积分为零，所以按第 (ii) 部分的论证，它在 $[0,1]$ 上处处为零。由于 $\mathrm{e}^{2x}>0$，得

$$\mathrm{k}(x)=\mathrm{e}^{-2x},\qquad 0\leq x\leq 1.$$

## 验算

- 第 (i) 部分取 $\mathrm{f}(x)=x$：左边为 $\int_0^1\sqrt{x}\,\mathrm{d}x=\frac{2}{3}$，右边为 $2\int_0^1 x^2\,\mathrm{d}x=\frac{2}{3}$。
- 第 (iii) 部分：取 $\mathrm{h}=\frac{1}{2}x^2$，则 $\int_0^1 x^2\,\mathrm{d}x=\frac{1}{3}$，而 $2\cdot\frac{1}{2}-2\cdot\frac{1}{6}-\frac{1}{3}=\frac{1}{3}$。两者相符。
- 第 (iv) 部分：取 $a=2$ 及 $\mathrm{k}=\mathrm{e}^{-2x}$，左边为 $\int_0^1 \mathrm{e}^{-2x}\,\mathrm{d}x=\frac{1}{2}(1-\mathrm{e}^{-2})\approx 0.43233$，右边为 $(1-\mathrm{e}^{-2})+\frac{1}{2}\mathrm{e}^{-2}-\frac{1}{4}-\frac{1}{4}=\frac{1}{2}-\frac{1}{2}\mathrm{e}^{-2}$，两者相同。另外以电脑对随机的 $a$ 及随机多项式 k 验证恒等式“左减右 $=\int \mathrm{e}^{ax}(\mathrm{k}-\mathrm{e}^{-ax})^2+(a-2)^2/(4a^2)$”，结果亦相符。

## 常见错误

- 单凭积分为零就断定 $\mathrm{g}(x)=x$，而没有用到连续性和非负性：单是积分为零，甚么也说明不了。
- 在第 (iv) 部分，忘记说明 $a\neq 0$，或在配方之前就尝试估猜 k。

## STEP 之外

$\int_0^1 \mathrm{f}(x)\mathrm{g}(x)\,\mathrm{d}x$ 这个量的表现就像“向量” f 与 g 的点积 (dot product)，而 $\int_0^1(\mathrm{g}-x)^2\,\mathrm{d}x$ 则是 g 与 $x$ 之间距离的平方。每一部分都在说“这个距离是零”，而第 (iv) 部分中的权 $\mathrm{e}^{ax}$ 则给出一个加权内积（weighted inner product）。这个观点（内积空间（inner product space），以及范数（norm）为零的连续函数必为零这一事实）是大学一年级线性代数

（linear algebra）和分析（analysis）的核心内容；可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》），以及 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2023，第 8 题

纯数学 · 分段微分方程

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

绝对值（modulus）令这条方程变成非线性（non-linear），但只是轻微的非线性：在任何一段 $y'$ 保持同一正负号的区间上，$|y'|$ 不过就是 $y'$ 或 $-y'$，方程于是变成一条普通的常系数（constant-coefficient）方程。因此我们沿著解一直走，直到 $y'$ 变为零，然后转用另一条线性方程，并在接合点令 $y$ 和 $y'$ 相配。第 (iv) 部分是省力的窍门：把解反射（reflection），即 $x \mapsto c - x$，会令一阶导数项变号，而这正好就是两条线性方程之间的切换。

## 解答

### 第 (i) 部分

设 $\mathrm{f}(x) = x\mathrm{e}^{-x}$，则

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

因此对每个实数 $x$，

$$\mathrm{f}'' + 2\mathrm{f}' + \mathrm{f} = \big[(x-2) + 2(1-x) + x\big]\mathrm{e}^{-x} = 0$$

都成立。此外 $\mathrm{f}(0) = 0$ 及 $\mathrm{f}'(0) = 1 \cdot \mathrm{e}^{0} = 1$，符合要求。

由于 $\mathrm{e}^{-x} > 0$，$\mathrm{f}'(x)$ 的正负号与 $1 - x$ 的正负号相同，而当 $x \le 1$ 时，$1 - x$ 为 $\ge 0$。所以当 $x \le 1$ 时，$\mathrm{f}'(x) \ge 0$。

### 第 (ii) 部分

在 $x \le 1$ 上，我们有 $\mathrm{f}' \ge 0$，所以 $|\mathrm{f}'| = \mathrm{f}'$，而 (ii) 的方程与 (i) 的方程相同。于是 f 在 $x \le 1$ 上满足新方程及其初始条件，故

$$\mathrm{g}_1(x) = x\mathrm{e}^{-x}.$$

在接合点，$\mathrm{g}_1(1) = \mathrm{e}^{-1}$ 及 $\mathrm{g}_1'(1) = 0$。连续可微（continuously differentiable）意味著 $\mathrm{g}_2$ 必须以相同的值开始：$\mathrm{g}_2(1) = \mathrm{e}^{-1}$，$\mathrm{g}_2'(1) = 0$。

> **想一想** 紧接 $x = 1$ 之后，$y'$ 取甚么正负号?

在 $x = 1$ 处，方程给出 $y'' = -2|y'| - y = -\mathrm{e}^{-1} < 0$，所以 $y'$ 在经过 0 时递减，并在稍为越过 $x = 1$ 之处为负。在那里 $|y'| = -y'$，方程变成

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} - 2\frac{\mathrm{d}y}{\mathrm{d}x} + y = 0.$$

辅助方程（auxiliary equation）$m^2 - 2m + 1 = 0$ 有重根（repeated root）$m = 1$，所以 $\mathrm{g}_2(x) = (A + Bx)\mathrm{e}^x$ 及 $\mathrm{g}_2'(x) = (A + B + Bx)\mathrm{e}^x$。在 $x = 1$ 处的条件给出

$$(A + B)\mathrm{e} = \mathrm{e}^{-1}, \qquad (A + 2B)\mathrm{e} = 0,$$

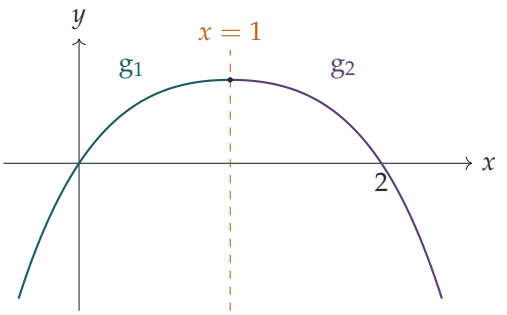
所以 $B = -\mathrm{e}^{-2}$ 及 $A = 2\mathrm{e}^{-2}$。因此

$$\mathrm{g}_2(x) = (2 - x)\mathrm{e}^{x-2}.$$

我们必须验证这与假设 $y' \le 0$ 一致：对所有 $x \ge 1$，$\mathrm{g}_2'(x) = (1 - x)\mathrm{e}^{x-2} \le 0$。所以在整个 $x \ge 1$ 上都有 $|\mathrm{g}_2'| = -\mathrm{g}_2'$，而 $\mathrm{g}_2$ 确实在整个 $x \ge 1$ 上满足这条含绝对值的方程。

### 第 (iii) 部分

在 $\mathrm{g}_1$ 中以 $2 - x$ 代替 $x$，得到 $(2 - x)\mathrm{e}^{-(2-x)} = (2 - x)\mathrm{e}^{x-2} = \mathrm{g}_2(x)$。所以 $\mathrm{g}_2(x) = \mathrm{g}_1(2 - x)$：曲线 $y = \mathrm{g}_2(x)$ 是 $y = \mathrm{g}_1(x)$ 关于直线 $x = 1$ 的反射。



**第 (iv) 部分**

设 $Y(x) = \mathrm{k}(c-x)$，并记 $u = c - x$。由链式法则（chain rule），配合 $\mathrm{d}u/\mathrm{d}x = -1$，

$$Y'(x) = -\mathrm{k}'(u), \qquad Y''(x) = \mathrm{k}''(u).$$

因此

$$Y'' - pY' + qY = \mathrm{k}''(u) + p\,\mathrm{k}'(u) + q\,\mathrm{k}(u),$$

而只要 $r \le u \le s$，右边便等于零，因为 k 在那里是原方程的解。而 $r \le c - x \le s$ 等价于

$$c - s \le x \le c - r,$$

所以 $y = \mathrm{k}(c-x)$ 在区间 $[c-s, c-r]$ 上满足 $y'' - py' + qy = 0$。

这解释了 (iii)：取 $p = 2$、$q = 1$、在 $(-\infty, 1]$ 上 $\mathrm{k} = \mathrm{g}_1$ 及 $c = 2$，反射后的函数 $\mathrm{g}_1(2-x)$ 在 $[1, \infty)$ 上是 $y'' - 2y' + y = 0$ 的解。

**第 (v) 部分**

设 $\mathrm{h}(x) = \mathrm{e}^{-x}\sin x$，则

$$\mathrm{h}'(x) = \mathrm{e}^{-x}(\cos x - \sin x) = \sqrt{2}\,\mathrm{e}^{-x}\cos\left(x + \tfrac{1}{4}\pi\right),$$

所以 $\mathrm{h}'(\frac{1}{4}\pi) = \sqrt{2}\,\mathrm{e}^{-\pi/4}\cos\frac{1}{2}\pi = 0$。（第二种形式亦显示，如题目所给，在 $[-\frac{3}{4}\pi, \frac{1}{4}\pi]$ 上 $\mathrm{h}' \ge 0$，以及 $\mathrm{h}'(-\frac{3}{4}\pi) = 0$，这一点稍后会用到。）

计划与第 (ii) 部分如出一辙。在 $y' \le 0$ 的区间上，方程是 $y'' - 2y' + 2y = 0$；在 $y' \ge 0$ 之处，方程是 $y'' + 2y' + 2y = 0$，而 h 满足这条方程。

**(a)** 在 $x = \frac{1}{4}\pi$ 处，我们有 $y = \mathrm{h}(\frac{1}{4}\pi) = \frac{1}{\sqrt{2}}\mathrm{e}^{-\pi/4} > 0$ 及 $y' = 0$，所以 $y'' = -2y < 0$，而 $y'$ 转为负。因此我们需要 $y'' - 2y' + 2y = 0$ 的一个解，而取 $p = q = 2$ 的第 (iv) 部分正好提供了一个：$y = \mathrm{h}(c-x)$ 在 $c - \frac{1}{4}\pi \le x \le c + \frac{3}{4}\pi$ 上成立。选取 $c = \frac{1}{2}\pi$，这个区间就恰好是 $[\frac{1}{4}\pi, \frac{5}{4}\pi]$，并且

$$y = \mathrm{h}\left(\tfrac{1}{2}\pi - x\right) = \mathrm{e}^{x-\pi/2}\sin\left(\tfrac{1}{2}\pi - x\right) = \mathrm{e}^{x-\pi/2}\cos x.$$

? **想一想** 我们怎样知道这是那个正确的解，而不只是某一个解?

它在接合点相配：$\mathrm{h}(\frac{1}{2}\pi - \frac{1}{4}\pi) = \mathrm{h}(\frac{1}{4}\pi)$，而它的导数 $-\mathrm{h}'(\frac{1}{2}\pi - x)$ 在 $x = \frac{1}{4}\pi$ 处等于 $-\mathrm{h}'(\frac{1}{4}\pi) = 0$。所以 $y$ 和 $y'$ 在那里连续。它亦是自洽的：对 $x \in [\frac{1}{4}\pi, \frac{5}{4}\pi]$，我们有 $\frac{1}{2}\pi - x \in [-\frac{3}{4}\pi, \frac{1}{4}\pi]$，而在那里 $\mathrm{h}' \ge 0$，所以 $y' = -\mathrm{h}'(\frac{1}{2}\pi - x) \le 0$，并且处处 $|y'| = -y'$。由于一条线性方程在某点给定 $y$ 和 $y'$ 后只有一个解，这就是解的延续。所以在 $\frac{1}{4}\pi \le x \le \frac{5}{4}\pi$ 上，

$$y = \mathrm{e}^{x-\pi/2}\cos x.$$

**(b)** 在 $x=\frac{5}{4}\pi$ 处：$y=\mathrm{e}^{3\pi/4}\cos\frac{5}{4}\pi=-\frac{1}{\sqrt{2}}\mathrm{e}^{3\pi/4}<0$，以及 $y'=-\mathrm{h}'(-\frac{3}{4}\pi)=0$。此时 $y''=-2y>0$，所以 $y'$ 转为正，方程变回 $y''+2y'+2y=0$。

这条方程是常系数的，所以把解平移后仍然是解。由于 $\sin(x-2\pi)=\sin x$，函数 $\mathrm{h}(x-2\pi)=\mathrm{e}^{2\pi-x}\sin x$ 是它的解；而在 $[\frac{5}{4}\pi,\frac{9}{4}\pi]$ 上，它的自变量 $x-2\pi$ 取遍 $[-\frac{3}{4}\pi,\frac{1}{4}\pi]$，在那里 $\mathrm{h}'\geq 0$，所以它与 $y'\geq 0$ 一致。验证在 $x=\frac{5}{4}\pi$ 处的接合：

$$\mathrm{e}^{2\pi-5\pi/4}\sin\tfrac{5}{4}\pi=-\tfrac{1}{\sqrt{2}}\mathrm{e}^{3\pi/4},\qquad \mathrm{h}'(\tfrac{5}{4}\pi-2\pi)=\mathrm{h}'(-\tfrac{3}{4}\pi)=0,$$

两者都相配。所以在 $\frac{5}{4}\pi\leq x\leq\frac{9}{4}\pi$ 上，

$$y=\mathrm{e}^{2\pi-x}\sin x.$$

在 $x=\frac{9}{4}\pi$ 处，这等于 $\mathrm{e}^{-\pi/4}\sin\frac{1}{4}\pi=\mathrm{h}(\frac{1}{4}\pi)$，且导数为零：解回到了它在 $\frac{1}{4}\pi$ 处的状态，所以对 $x\geq\frac{1}{4}\pi$，解是周期（periodic）为 $2\pi$ 的周期函数。

## 验算

- 各段按构造是光滑地接合的，但直接看一看也是值得的：在 $x=\frac{5}{4}\pi$ 处，(a) 和 (b) 都给出 $y=-\frac{1}{\sqrt{2}}\mathrm{e}^{3\pi/4}\approx -7.4605$，斜率为零。
- 从 $y(0)=0$、$y'(0)=1$ 开始，以数值方法逐步推进这些含绝对值的方程，可在 $[0,4]$ 上重现 $\mathrm{g}_1,\mathrm{g}_2$，并在 $[0,\frac{9}{4}\pi]$ 上重现 (v) 的三段，误差在 $10^{-9}$ 以内。
- 直接对 (a) 求导：$y'=\mathrm{e}^{x-\pi/2}(\cos x-\sin x)$，$y''=-2\mathrm{e}^{x-\pi/2}\sin x$，而 $y''-2y'+2y=0$，符合要求。

## 常见错误

- 过了转向点（turning point）后仍然保留 $|y'|=y'$：在 $x=1$ 之后，f 本身并不是解。
- 忘记验证新的一段在整个区间上都令 $y'$ 保持所假设的正负号。
- 在 (iv) 中把区间写成 $[r,s]$；反射会把它映射到 $[c-s,c-r]$。

## STEP 之外

像这样的方程，其右边除了在一个切换曲面（这里是 $y'=0$）两侧之外都是光滑的，称为分段光滑系统（piecewise-smooth systems）或切换系统（switching systems）；解是由线性的片段黏合而成的，而由于 $|y'|$ 是利普希茨（Lipschitz）连续的，存在性与唯一性（existence and uniqueness）仍然可由皮卡－林德洛夫定理（Picard–Lindelöf theorem）得出。(v) 中的 $2\pi$ 周期性是线性阻尼方程永远不可能产生的：取 $E=\frac{1}{2}y'^2+y^2$，我们得

到 $\mathrm{d}E/\mathrm{d}x = -2y'|y'|$，所以这个“阻尼”项在 $y$ 上升时移走能量，在 $y$ 下降时又把能量归还。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖线性理论及唯一性。

详细解答

# STEP III 2023，第 9 题

力学 • 以绳连接的两质点运动

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

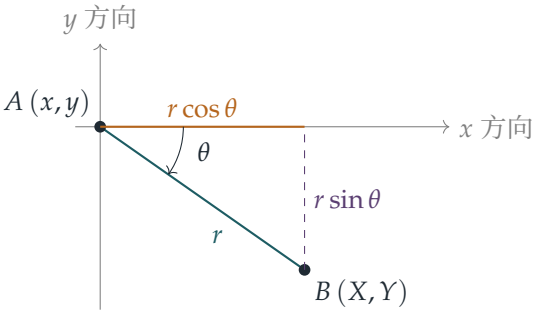
用 $A$ 的位置和绳子的方向 $\theta$ 来描述整个系统。张力（tension）只沿绳子方向作用，所以每个质点的加速度在垂直于绳子的方向上都没有分量。比较两个质点的垂直加速度，便可推出 $\ddot{\theta} = 0$：绳子以恒定的速率转动。之后再用动量守恒（conservation of momentum）就能完成解题。

## 解答

全题中，两个质点在光滑水平面上运动，绳子轻而不可伸长（inextensible），而且（按题意可假设）绳子一直保持拉紧，所以任何时刻都有 $AB = r$。

### 第 (i) 部分

向量 $\overrightarrow{AB}$ 的长度是 $r$。由于 $\theta$ 是从正 $x$ 轴顺时针量度的，当 $0 < \theta < \frac{\pi}{2}$ 时，这个向量指向 $x$ 方向的下方，如图所示。



从 $A$ 走到 $B$，我们在 $x$ 方向移动 $r\cos\theta$，在 $y$ 的负方向移动 $r\sin\theta$，所以

$$X = x + r\cos\theta, \qquad Y = y - r\sin\theta.$$

这两条式子对所有 $\theta$ 都成立，因为 $(\cos\theta, -\sin\theta)$ 正是从 $x$ 轴顺时针转 $\theta$ 角的单位向量（unit vector）。

**第 (ii) 部分**

对 $t$ 求导（$r$ 是常数，$\theta$ 是 $t$ 的函数）：

$$\dot{X} = \dot{x} - r\dot{\theta}\sin\theta, \qquad \dot{Y} = \dot{y} - r\dot{\theta}\cos\theta,$$
$$\ddot{X} = \ddot{x} - r\ddot{\theta}\sin\theta - r\dot{\theta}^2\cos\theta, \qquad \ddot{Y} = \ddot{y} - r\ddot{\theta}\cos\theta + r\dot{\theta}^2\sin\theta.$$

**第 (iii) 部分**

张力 $T$ 沿单位向量 $(\cos\theta, -\sin\theta)$ 把 $A$ 拉向 $B$，又沿 $(-\cos\theta, \sin\theta)$ 把 $B$ 拉向 $A$。根据牛顿第二定律（Newton's second law），

$$m\ddot{x} = T\cos\theta, \quad m\ddot{y} = -T\sin\theta, \qquad M\ddot{X} = -T\cos\theta, \quad M\ddot{Y} = T\sin\theta.$$

因此

$$m(\ddot{x}\sin\theta + \ddot{y}\cos\theta) = T\cos\theta\sin\theta - T\sin\theta\cos\theta = 0,$$

用完全相同的方法可得 $M(\ddot{X}\sin\theta + \ddot{Y}\cos\theta) = 0$。分别除以 $m$ 和 $M$，便得到所要求的两个结果。（从几何角度看：$(\sin\theta, \cos\theta)$ 垂直于绳子，而两个质点在这个方向上都没有加速度。）

? **想一想** 我们有两条方程，都是说“没有垂直方向的加速度”。它们怎样告诉我们关于 $\theta$ 的信息?

把第二条减去第一条，再把第 (ii) 部分中 $\ddot{X} - \ddot{x}$ 和 $\ddot{Y} - \ddot{y}$ 的表达式代入：

$$\begin{aligned} 0 &= (\ddot{X} - \ddot{x})\sin\theta + (\ddot{Y} - \ddot{y})\cos\theta \\ &= \left(-r\ddot{\theta}\sin\theta - r\dot{\theta}^2\cos\theta\right)\sin\theta + \left(-r\ddot{\theta}\cos\theta + r\dot{\theta}^2\sin\theta\right)\cos\theta \\ &= -r\ddot{\theta}\left(\sin^2\theta + \cos^2\theta\right) = -r\ddot{\theta}. \end{aligned}$$

由于 $r \neq 0$，所以 $\ddot{\theta} = 0$，即 $\dot{\theta}$ 是常数。

要求出这个常数，可用初始条件。在 $t = 0$ 时，$B$ 相对于 $A$ 位于 $(r, 0)$，所以 $\theta(0) = 0$。此外 $\dot{y}(0) = u$ 而 $\dot{Y}(0) = 0$，于是第 (ii) 部分中 $\dot{Y}$ 的表达式给出

$$0 = u - r\dot{\theta}(0)\cos 0 \quad \Longrightarrow \quad \dot{\theta}(0) = \frac{u}{r}.$$

（$\dot{X}$ 的方程与此一致：$0 = 0 - r\dot{\theta}(0)\sin 0$。）所以对所有 $t$ 都有 $\dot{\theta} = u/r$；利用 $\theta(0) = 0$ 积分，得

$$\theta = \frac{ut}{r},$$

此即所求。

**第 (iv) 部分**

把第 (iii) 部分的运动方程两两相加，

$$m\ddot{x} + M\ddot{X} = T\cos\theta - T\cos\theta = 0, \qquad m\ddot{y} + M\ddot{Y} = -T\sin\theta + T\sin\theta = 0.$$

这就是动量守恒：张力是系统的内力。把第二条方程积分一次，可知 $m\dot{y} + M\dot{Y}$ 是常数，等于其初始值 $mu + M\cdot 0 = mu$。再积分一次，并利用 $y(0) = Y(0) = 0$，得

$$my + MY = mut.$$

**第 (v) 部分**

代入第 (i) 部分的 $Y = y - r\sin\theta$ 和第 (iii) 部分的 $\theta = ut/r$：

$$my + M\left(y - r\sin\frac{ut}{r}\right) = mut \quad\Longrightarrow\quad y = \frac{1}{m+M}\left(mut + Mr\sin\frac{ut}{r}\right),$$

此即所求结果。

**第 (vi) 部分**

把第 (v) 部分的结果求导，

$$\dot{y} = \frac{1}{m+M}\left(mu + Mu\cos\frac{ut}{r}\right) = \frac{u}{m+M}\left(m + M\cos\frac{ut}{r}\right).$$

在时刻 $t = \pi r/u$，即绳子转了半圈时，$\cos(ut/r) = -1$，而

$$\dot{y} = \frac{u(m-M)}{m+M},$$

当 $M > m$ 时这是负数（因为 $u > 0$）。所以在那一刻，$A$ 的速度的 $y$ 分量是负的；由连续性（continuity）可知，在该时刻附近的一段时间区间内亦然。

## 验算

- 在 $t = 0$ 时，由第 (v) 部分的公式得 $\dot{y} = u(m+M)/(m+M) = u$，与 $A$ 的初速度相符。

- 在 $x$ 方向用同样方法可得 $mx + MX = Mr$，所以 $x = \dfrac{Mr}{m+M}(1-\cos\theta)$。结合第 (v) 部分，质心（centre of mass）在 $y$ 方向以速率 $mu/(m+M)$ 作直线运动，而两个质点以角速度（angular speed）$u/r$ 绕质心转动，这正是我们熟悉的图像。于是张力是常数 $\dfrac{mM}{m+M}\dfrac{u^2}{r} > 0$，所以绳子确实一直保持拉紧。
- 若 $M \to \infty$，$B$ 保持不动，第 (v) 部分给出 $y = r\sin(ut/r)$：$A$ 绕 $B$ 作半径为 $r$ 的圆周运动，理应如此。
- 取 $m = 1$、$M = 3$、$u = r = 1$，用电脑模拟两个以极硬弹簧相连的质点，在 $t = \pi$ 时得到 $y \approx 0.7854$ 和 $\dot{y} \approx -0.5000$，与 $\frac{\pi}{4}$ 及 $\frac{1-3}{1+3} = -\frac{1}{2}$ 吻合。

## 常见错误

- 因顺时针的约定而弄错正负号：$Y = y - r\sin\theta$，所以 $\dot{Y}$ 中的项是 $-r\dot{\theta}\cos\theta$，而不是 $+$。
- 忘记交代常数 $u/r$ 的来由：它来自 $\dot{Y}(0) = 0$，而不单单来自 $A$ 的速率。
- 在第 (vi) 部分中只声称 $\dot{y} < 0$“最终”会成立，却没有给出具体时刻；$t = \pi r/u$ 便是一个确定的时刻。

## STEP 之外

这是二体问题（two-body problem）最简单的形式：运动分解为质心的匀速运动，加上相对于质心的运动，而后者由约化质量（reduced mass）$mM/(m+M)$ 支配。同样的分解方法是行星轨道和双原子分子（diatomic molecule）振动的基础。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）仔细地发展了质心坐标和约化质量的理论。

详细解答

# STEP III 2023，第 10 题

力学 · 梁的平衡

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

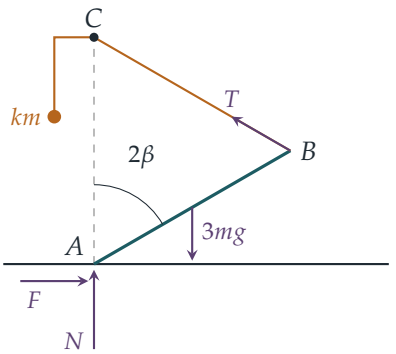
www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

滑轮（pulley）恰好位于 $A$ 正上方一条横梁长度之处，因此 $A$、$B$、$C$ 组成一个等腰三角形（isosceles triangle），其中 $AB = AC = 2h$。这个几何关系让我们不费吹灰之力便得到绳子的方向，以及绳子与 $A$ 的垂直距离。之后，一切都按处理平衡（equilibrium）中的杆的惯常步骤进行：对 $A$ 取力矩（moments）（这样便同时消去两个未知的接触力），分解（resolve）各力，再加上条件 $F \leq \mu N$。

## 解答

我们把横梁视为均匀杆（uniform rod），两个质点（particle）视为点质量，绳子视为轻而不可伸长（light and inextensible），并跨过一个光滑（smooth）滑轮，因此整条绳的张力（tension）相同。由于悬挂的质点处于平衡，张力为 $T = kmg$。




设 $A$ 为原点，$x$ 轴水平（指向 $B$ 所在的一侧），$y$ 轴铅垂向上。于是 $C = (0, 2h)$，而

$B = 2h(\sin 2\beta, \cos 2\beta)$。

*绳子的方向*。由 $B$ 指向 $C$ 的向量（vector）为

$$2h\big(-\sin 2\beta,\ 1-\cos 2\beta\big) = 4h\sin\beta\,(-\cos\beta,\ \sin\beta),$$

这里用了 $1-\cos 2\beta = 2\sin^2\beta$。因此，作用于横梁的张力有沿 $-x$ 方向（即指回直线 $AC$）的水平分量 $T\cos\beta$，以及向上的铅垂分量 $T\sin\beta$。从几何角度看：在等腰三角形 $ABC$ 中，两个底角均为 $90^\circ - \beta$，所以 $BC$ 与水平线所成的角为 $\beta$。

*力臂*（*moment arm*）。由 $A$ 到 $BC$ 的垂线平分等腰三角形的顶角 $2\beta$，所以其长度为 $2h\cos\beta$。

**第 (i) 部分**

*对 $A$ 取力矩*。重量 $3mg$ 作用于 $AB$ 的中点，与 $A$ 的水平距离为 $h\sin 2\beta$。令它与张力的力矩相等：

$$kmg\cdot 2h\cos\beta = 3mg\cdot h\sin 2\beta = 6mgh\sin\beta\cos\beta.$$

由于 $0 < 2\beta < 180^\circ$，故 $\cos\beta \neq 0$，可以两边除以它：

$$\boxed{k = 3\sin\beta.}$$

（这与 $k < 3$ 一致。）

*分解各力*。铅垂方向：$N + T\sin\beta = 3mg$，所以

$$N = 3mg - 3mg\sin^2\beta = 3mg\cos^2\beta.$$

水平方向：摩擦力（friction）必须抵消绳子的水平拉力，因此它沿 $+x$ 方向作用，且

$$F = T\cos\beta = 3mg\sin\beta\cos\beta.$$

故 $F/N = \tan\beta$。

*摩擦条件*。平衡要求 $F \le \mu N$，即 $\tan\beta \le \mu$。而

$$k^2 = 9\sin^2\beta = \frac{9\tan^2\beta}{1+\tan^2\beta}.$$

函数（function）$t \mapsto \dfrac{t}{1+t} = 1 - \dfrac{1}{1+t}$ 在 $t \ge 0$ 时递增，且 $0 \le \tan^2\beta \le \mu^2$，所以

$$k^2 \le \frac{9\mu^2}{\mu^2+1},$$

此即所求。

**第 (ii) 部分**

现在 $T = 2mg$，而附加的质量为 $m$ 的质点与 $A$ 的水平距离为 $xh\sin 2\beta$。绳子的几何关系不变。

*对 $A$ 取力矩：*

$$2mg \cdot 2h\cos\beta = 3mg\,h\sin 2\beta + mg\,xh\sin 2\beta = 2mgh(3+x)\sin\beta\cos\beta,$$

两边除以 $2mgh\cos\beta \neq 0$，得

$$\sin\beta = \frac{2}{3+x}.$$

当 $0 \le x \le 2$ 时，此值介乎 $\frac{2}{5}$ 与 $\frac{2}{3}$ 之间，所以这样的 $\beta$ 存在。

*分解各力：*

$$N = 4mg - 2mg\sin\beta, \qquad F = 2mg\cos\beta,$$

所以

$$\frac{F^2}{N^2} = \frac{4\cos^2\beta}{(4-2\sin\beta)^2} = \frac{\cos^2\beta}{(2-\sin\beta)^2}.$$

代入 $\sin\beta = 2/(3+x)$：

$$\cos^2\beta = 1 - \frac{4}{(3+x)^2} = \frac{(3+x-2)(3+x+2)}{(3+x)^2} = \frac{(x+1)(x+5)}{(3+x)^2},$$

$$2-\sin\beta = \frac{2(3+x)-2}{3+x} = \frac{2(x+2)}{3+x}.$$

$(3+x)^2$ 的因子互相消去，剩下

$$\frac{F^2}{N^2} = \frac{(x+1)(x+5)}{4(x+2)^2} = \frac{x^2+6x+5}{4(x+2)^2},$$

此即所求结果。

**想一想** 怎样把“无论 $x$ 取何值”转化为关于 $\mu$ 的条件?

对每个 $x$，我们都需要 $F \le \mu N$，即 $\mu \ge F/N$。要涵盖所有 $x \in [0,2]$，$\mu$ 必须至少等于 $F/N$ 在该区间上的最大值。因此我们要求 $F^2/N^2$ 的最大值。

按照提示，

$$\frac{1}{3} - \frac{F^2}{N^2} = \frac{4(x+2)^2 - 3(x^2+6x+5)}{12(x+2)^2} = \frac{x^2-2x+1}{12(x+2)^2} = \frac{(x-1)^2}{12(x+2)^2} \ge 0,$$

等号当且仅当 $x = 1$ 时成立，而这个值在容许范围之内。所以对所有 $x \in [0,2]$，$F^2/N^2 \le \frac{1}{3}$，且此界在 $x = 1$ 时达到。因此 $F/N$ 的最大值为 $1/\sqrt{3}$，而横梁对每个 $x$ 都能保持平衡，当且仅当 $\mu \ge 1/\sqrt{3}$。所求的最小值为

$$\boxed{\mu = \frac{1}{\sqrt{3}}}.$$

## 验算

- 在第 (i) 部分，试考虑横梁水平放置的情况，即 $2\beta = 90°$。此时 $B = (2h, 0)$，绳子 $BC$ 与水平成 $45°$，它与 $A$ 的距离为 $\sqrt{2}\,h$。取力矩得 $T\sqrt{2}\,h = 3mgh$，所以 $T = 3mg/\sqrt{2}$，与 $k = 3\sin 45°$ 吻合。此外 $N = F = \frac{3}{2}mg$，所以 $F/N = 1 = \tan 45°$。
- 在第 (ii) 部分取 $x = 0$，公式给出 $5/16$；直接计算则有 $\sin\beta = \frac{2}{3}$、$\cos^2\beta = \frac{5}{9}$、$(2 - \frac{2}{3})^2 = \frac{16}{9}$，同样得到 $\frac{5}{16}$。
- 取 $x = 1$：$\sin\beta = \frac{1}{2}$，所以 $\beta = 30°$，横梁与铅垂线成 $60°$；$F/N = \cos 30°/(2 - \frac{1}{2}) = 1/\sqrt{3} \approx 0.5774$。以电脑对 $x$ 在 $[0, 2]$ 范围内的力矩及分解后的力进行验算，确认了公式及这个最大值。

## 常见错误

- 忘记绳子是沿 $BC$ 拉，而不是铅垂地拉：滑轮在 $C$，所以张力有水平分量，而摩擦力要抵抗的正是这个分量。
- 取了 $F/N$ 的最小值（或其在端点的值），而不是它在 $0 \le x \le 2$ 上的最大值。最坏情况出现在区间内部，即 $x = 1$ 处。
- 混淆力臂：重量的力臂是水平距离（$h\sin 2\beta$），而张力的力臂是到 $BC$ 的垂直距离（$2h\cos\beta$）。

## STEP 之外

取力矩与分解各力，正是刚体（rigid body）平衡条件的两半：合力为零，以及合力矩（torque）为零。到了大学，这些条件会写成向量形式 $\sum \mathbf{F} = \mathbf{0}$ 及 $\sum \mathbf{r} \times \mathbf{F} = \mathbf{0}$；而第 (ii) 部分那种“对参数取最坏情况”的推理，正好让你初尝工程师如何为安全余量作设计。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）以这种向量语言，细致地发展了静力学（statics）与力矩的理论。

详细解答

# STEP III 2023，第 11 题

概率与统计 · 条件期望值

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

只要把 $k+1=k+1$ 写开，并把 $k$ 与 $k!$ 约去，开首的级数（series）便分拆成两个熟悉的指数级数（exponential series）。在第 (i) 部分，$D$ 的分布（distribution）涉及一个求和，遍及所有可能产生 $D=d$ 的 $Y$ 值；交换二重和（double sum）的求和次序后，内层的和就变成简单的 $1+2+\cdots+k$，而随之出现的因子 $\frac{1}{2}(k+1)$ 正好可以用开首的级数处理。第 (ii) 部分则只需以骰子的结果作条件（conditioning）。

## 解答

### 级数

当 $k \geq 1$ 时，$\dfrac{k}{k!}=\dfrac{1}{(k-1)!}$，因此把分子拆开，得

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{k+1}{k!}x^k &= \sum_{k=1}^{\infty}\frac{x^k}{(k-1)!}+\sum_{k=1}^{\infty}\frac{x^k}{k!}\\ &= x\sum_{j=0}^{\infty}\frac{x^j}{j!}+\left(\sum_{k=0}^{\infty}\frac{x^k}{k!}-1\right)\\ &= xe^x+e^x-1=(x+1)e^x-1,\end{aligned}$$

其中第一个和用了 $j=k-1$。两个级数对每个实数 $x$ 都收敛（converge），所以这样拆开是合理的。这就是所求的恒等式。

### 第 (i)(a) 部分

$D=0$ 当且仅当 $Y=0$（若 $Y=k\geq 1$，骰子显示的数至少是 1），所以

$$\mathrm{P}(D=0)=\mathrm{P}(Y=0)=\mathrm{e}^{-n}.$$

### 第 (i)(b) 部分

固定 $d\geq 1$。骰子只有在至少有 $d$ 面时才可能显示 $d$，即只有当 $Y=k$ 且 $k\geq d$ 时才可能；而在 $Y=k\geq d$ 的条件下，公正的骰子显示 $d$ 的概率是 $\frac{1}{k}$。根据全概率定律（law of total probability），

$$\mathrm{P}(D=d)=\sum_{k=d}^{\infty}\mathrm{P}(D=d\mid Y=k)\,\mathrm{P}(Y=k)=\sum_{k=d}^{\infty}\frac{1}{k}\cdot\frac{n^k}{k!}\mathrm{e}^{-n}.$$

按定义，$\mathrm{E}(D)=\sum_{d=0}^{\infty}d\,\mathrm{P}(D=d)$，而 $d=0$ 的项为零，所以

$$\mathrm{E}(D)=\sum_{d=1}^{\infty}\left[d\sum_{k=d}^{\infty}\left(\frac{1}{k}\cdot\frac{n^k}{k!}\mathrm{e}^{-n}\right)\right],$$

正如所求。

> **想一想** 一个外层对 $d$ 求和、内层对 $k\geq d$ 求和的式子，怎样才能变成外层对 $k$ 求和?

这个二重和遍及所有满足 $1\leq d\leq k$ 的整数对 $(d,k)$。每一项 $\frac{d}{k}\cdot\frac{n^k}{k!}\mathrm{e}^{-n}$ 都是非负的，所以我们可以按任何次序把各项相加，总和不变。改为按 $k$ 而不是按 $d$ 分组：对每个 $k\geq 1$，$d$ 由 1 取到 $k$。因此

$$\mathrm{E}(D)=\sum_{k=1}^{\infty}\left(\frac{1}{k}\cdot\frac{n^k}{k!}\mathrm{e}^{-n}\sum_{d=1}^{k}d\right),$$

这就是所求的第二个形式。

### 第 (i)(c) 部分

利用 $\sum_{d=1}^{k}d=\frac{1}{2}k(k+1)$，因子 $\frac{1}{k}$ 被约去：

$$\begin{aligned}\mathrm{E}(D)&=\sum_{k=1}^{\infty}\frac{1}{k}\cdot\frac{n^k}{k!}\mathrm{e}^{-n}\cdot\frac{k(k+1)}{2}=\frac{\mathrm{e}^{-n}}{2}\sum_{k=1}^{\infty}\frac{k+1}{k!}n^k\\&=\frac{\mathrm{e}^{-n}}{2}\Big((n+1)\mathrm{e}^{n}-1\Big)=\tfrac{1}{2}\left(n+1-\mathrm{e}^{-n}\right),\end{aligned}$$

这里用了开首的恒等式，并取 $x=n$。这就是所求的结果。

### 第 (ii)(a) 部分

设 $K$ 为 $n$ 面骰子显示的数，则对 $k = 1, \ldots, n$ 有 $\mathrm{P}(K = k) = \frac{1}{n}$；而在 $K = k$ 的条件下，$Z = X_k$，它服从均值为 $k$ 的泊松分布（Poisson distribution）。所以 $\mathrm{P}(Z = 0 \mid K = k) = \mathrm{e}^{-k}$，再根据全概率定律，

$$\mathrm{P}(Z = 0) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{e}^{-k}.$$

这是首项为 $\mathrm{e}^{-1}$、公比为 $\mathrm{e}^{-1}$ 的几何级数（geometric series）：

$$\mathrm{P}(Z = 0) = \frac{1}{n} \cdot \frac{\mathrm{e}^{-1}\left(1 - \mathrm{e}^{-n}\right)}{1 - \mathrm{e}^{-1}} = \frac{1 - \mathrm{e}^{-n}}{n(\mathrm{e} - 1)},$$

最后一步是把分子和分母同乘以 $\mathrm{e}$。

### 第 (ii)(b) 部分

再次以骰子的结果作条件，$\mathrm{E}(Z \mid K = k) = \mathrm{E}(X_k) = k$，所以

$$\mathrm{E}(Z) = \sum_{k=1}^{n} \mathrm{E}(Z \mid K = k)\,\mathrm{P}(K = k) = \frac{1}{n} \sum_{k=1}^{n} k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

与第 (i)(c) 部分比较，

$$\mathrm{E}(Z) - \mathrm{E}(D) = \frac{n+1}{2} - \frac{n+1-\mathrm{e}^{-n}}{2} = \tfrac{1}{2}\mathrm{e}^{-n} > 0,$$

所以 $\mathrm{E}(Z) > \mathrm{E}(D)$，正如所求。

有趣的是，如果没有零的情况，两者便会相等：一颗面数 $Y$ 服从泊松分布的公正骰子，平均显示 $\frac{1}{2}(Y+1)$，而它的均值是 $\frac{1}{2}(n+1)$。唯一的损失在于 $Y = 0$ 时得到 $D = 0$，而不是公式 $\frac{1}{2}(Y+1)$ 所给出的 $\frac{1}{2}$；这情况发生的概率是 $\mathrm{e}^{-n}$，恰好令期望值（expectation）少了 $\frac{1}{2}\mathrm{e}^{-n}$。

## 验算

- 取 $x = 0$，开首的恒等式变成 $0 = 1 \cdot 1 - 1$；取 $x = 1$，则得 $\sum_{k \geq 1} \frac{k+1}{k!} = 2\mathrm{e} - 1 \approx 4.4366$，部分和亦证实了这一点。
- 当 $n = 3$ 时：$\mathrm{E}(D) = \frac{1}{2}(4 - \mathrm{e}^{-3}) \approx 1.9751$，$\mathrm{E}(Z) = 2$，以及 $\mathrm{P}(Z = 0) \approx 0.1843$。以电脑把每个实验各模拟 200 000 次，得到 1.970、2.002 和 0.185。
- 当 $n = 1$ 时，第 (ii) 部分的骰子必定显示 1，所以 $\mathrm{P}(Z = 0) = \mathrm{e}^{-1}$；公式给出 $\frac{1-\mathrm{e}^{-1}}{\mathrm{e}-1} = \mathrm{e}^{-1}$。

## 常见错误

- 在 (i)(b) 中把内层的和由 $k = 1$ 而不是由 $k = d$ 开始：$k$ 面骰子不可能显示比 $k$ 大的数。
- 忘记开首恒等式中的 $-1$，它来自 $\mathrm{e}^x$ 中缺少的 $k = 0$ 项；正是它产生了 (i)(c) 中的 $\mathrm{e}^{-n}$。

## STEP 之外

两个部分都是塔性质（tower law）$\mathrm{E}(Z) = \mathrm{E}\big(\mathrm{E}(Z \mid K)\big)$ 的实例：在第 (i) 部分，它一行便给出 $\mathrm{E}(D) = \mathrm{E}\big(\frac{1}{2}(Y+1)\mathbf{1}_{Y\geq 1}\big)$，而我们亲手完成的交换求和次序，正是这条定律背后所隐藏的步骤。可以自由重排非负项二重级数的次序，是一条定理（其一般形式为 Tonelli 定理），在大学一年级的分析（analysis）课程中会遇到。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）讨论了条件期望（conditional expectation）以及由它建构的复合分布（compound distributions）。

详细解答

# STEP III 2023，第 12 题

概率与统计 · 组合概率

**STEP III 2023，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 2023 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/707102-2023-past-papers.zip

## 关键思路

虽然同一对袜子的两只看起来一样，但抽出的是从 $2n$ 只*实际存在的*袜子中均匀随机（uniformly random）选取的 $2k$ 只所组成的集合，所以我们为每只袜子编上标号，然后数集合的个数。这样一来，一切都归结为分阶段构造一个选法：先选出哪些对是完整出现的，再选出哪些其他的对只贡献一只袜子，最后选出每一个这样的对中取哪一只。第 (iii) 部分用的是熟悉的技巧 $r\binom{n}{r} = n\binom{n-1}{r-1}$，它把一个加权和转化为一个概率（probability）之和。

## 解答

为袜子编上标号，使全部 $2n$ 只互不相同。共有 $\binom{2n}{2k}$ 种机会均等的选法，而下面每个概率都等于（有利选法的数目）$/\binom{2n}{2k}$。

### 第 (i) 部分

一个没有完整一对的选法，从每一对中最多只含一只袜子，所以它的 $2k$ 只袜子来自 $2k$ 个不同的对。要构造这样的选法，先选出有代表出现的 $2k$ 对，有 $\binom{n}{2k}$ 种方法（因为 $2k \le n$，这是可行的），然后从每一对的两只中选一只，有 $2^{2k}$ 种方法。每一种这样的选择都给出不同的集合，而每个没有成对的集合都以这种方式产生，所以

$$\mathrm{P}(\text{没有成对}) = \frac{\binom{n}{2k}\,2^{2k}}{\binom{2n}{2k}}.$$

**第 (ii) 部分**

固定 $r$，其中 $0 \le r \le k$。一个恰好有 $r$ 对完整袜子的选法由以下部分组成：

- 整对取出的 $r$ 对（共 $2r$ 只袜子），以及
- 另外 $2k-2r = 2(k-r)$ 只袜子，全部来自不同的对（否则成对的数目会多于 $r$），而且都不来自已用的 $r$ 对。

因此，我们以 $\binom{n}{r}$ 种方法选出 $r$ 个完整的对；从余下的 $n-r$ 对中，以 $\binom{n-r}{2(k-r)}$ 种方法选出贡献一只袜子的 $2(k-r)$ 对；再从这些对中各取一只，有 $2^{2(k-r)}$ 种方法。同样地，每个恰好有 $r$ 对的集合都恰好被产生一次，所以

$$\mathrm{P}(X_{n,k} = r) = \frac{\binom{n}{r}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n}{2k}},$$

此即所求。（当 $r = 0$ 时，这就是第 (i) 部分。）

**第 (iii) 部分**

设 $1 \le r \le k$。由于

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

由第 (ii) 部分得

$$r\,\mathrm{P}(X_{n,k} = r) = \frac{n\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n}{2k}}.$$

现在把第 (ii) 部分应用于 $n-1$ 对袜子、抽出 $2(k-1)$ 只、找到 $r-1$ 对的情况。这是容许的，因为由 $2k \le n$ 可推出 $2(k-1) \le n-1$，而且 $0 \le r-1 \le k-1$。由于 $(n-1)-(r-1) = n-r$ 及 $(k-1)-(r-1) = k-r$，

$$\mathrm{P}(X_{n-1,k-1} = r-1) = \frac{\binom{n-1}{r-1}\binom{n-r}{2(k-r)}\,2^{2(k-r)}}{\binom{2n-2}{2k-2}}.$$

两者的分子除了因子 $n$ 之外完全相同，所以

$$\begin{aligned}\frac{r\,\mathrm{P}(X_{n,k} = r)}{\mathrm{P}(X_{n-1,k-1} = r-1)} &= \frac{n\binom{2n-2}{2k-2}}{\binom{2n}{2k}} = n \cdot \frac{(2n-2)!}{(2k-2)!\,(2n-2k)!} \cdot \frac{(2k)!\,(2n-2k)!}{(2n)!} \\ &= n \cdot \frac{2k(2k-1)}{2n(2n-1)} = \frac{k(2k-1)}{2n-1}.\end{aligned}$$

（上面的除法是安全的：$n-r \ge 2k-r \ge 2(k-r)$，所以这里每个二项式系数（binomial coefficient）都是正数。）两边相乘即得所求结果。

**想一想** 这个恒等式对求 $\mathrm{E}(X_{n,k})$ 有甚么帮助?

期望值（expectation）正好是左边各项之和，而右边各项加起来是一个完整的概率分布（probability distribution）。期望值中 $r = 0$ 的一项为零，所以

$$\begin{aligned}\mathrm{E}(X_{n,k}) = \sum_{r=1}^{k} r\,\mathrm{P}(X_{n,k} = r) &= \frac{k(2k-1)}{2n-1} \sum_{r=1}^{k} \mathrm{P}(X_{n-1,k-1} = r-1) \\ &= \frac{k(2k-1)}{2n-1} \sum_{s=0}^{k-1} \mathrm{P}(X_{n-1,k-1} = s).\end{aligned}$$

变量 $X_{n-1,k-1}$ 数的是 $2(k-1)$ 只袜子中的对数，所以它只取值 $0, 1, \ldots, k-1$，因此最后的和等于 1。故此

$$\mathrm{E}(X_{n,k}) = \frac{k(2k-1)}{2n-1}.$$

## 验算

- 当 $k = 1$（两只袜子）时，答案是 $\frac{1}{2n-1}$：无论先抽到哪一只，余下的 $2n-1$ 只中恰好有一只是它的配对。
- 用指示变量（indicator variables）可以直接求出平均值：第 $i$ 对完整出现的概率是 $\binom{2n-2}{2k-2}/\binom{2n}{2k} = \frac{2k(2k-1)}{2n(2n-1)}$，乘以 $n$ 对，同样得到 $\frac{k(2k-1)}{2n-1}$。
- 当 $n = 6$、$k = 3$ 时：第 (i) 部分给出 $\frac{64}{924} = \frac{16}{231}$，而平均值是 $\frac{15}{11}$。以电脑对所有 $n \le 8$ 枚举全部选法，证实了 (ii) 中的分布、(iii) 中的恒等式以及平均值。

## 常见错误

- 把同一对的两只袜子当作真正相同，然后去数“颜色组合”：这些组合并非机会均等，所以必须对有标号的袜子计数。
- 忘记了因子 $2^{2(k-r)}$，它对应每个单只出现的对中取哪一只。
- 在 (iii) 中没有留意到 $\sum_s \mathrm{P}(X_{n-1,k-1} = s) = 1$，而尝试硬算这个和。

## STEP 之外

第二项验算用的是期望值的线性性（linearity of expectation）配合指示变量，它无需知道分布就能求出平均值；即使“第 $i$ 对完整出现”这些事件并不独立，这方法依然成立。(iii) 中的恒等式是一个尺寸偏倚（size-biasing）关系：以 $r$ 为分布加权，会得到一个较小问题的平移版本；这个技巧同样可以求出二项分布（binomial distribution）和超几何

分布（hypergeometric distribution）的平均值。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）全书都采用这类指示变量和故事式证明。

# 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

**2023 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。

