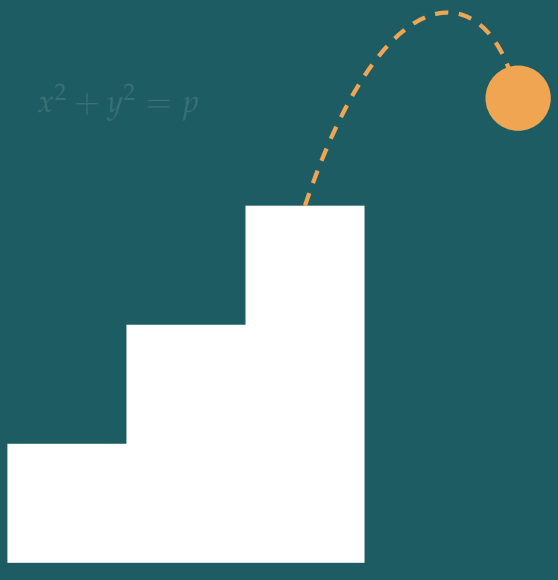
STEP 备试 · 大学数学

# Beyond STEP

## STEP II 及 III

2021 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**

2021 年详细解答

Keith McNulty

**Beyond STEP：以历届 STEP 试题为门，走进大学数学**

**2021 年详细解答**

2021 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 三角方程 纯数学 . . . 3

第 2 题 根的对称函数 纯数学 . . . 7

第 3 题 整数部分与小数部分 纯数学 . . . 11

第 4 题 Lambert W 函数 纯数学 . . . 15

第 5 题 一阶微分方程 纯数学 . . . 19

第 6 题 圆几何的近似 纯数学 . . . 23

第 7 题 旋转矩阵 纯数学 . . . 27

第 8 题 积分递推关系 纯数学 . . . 31

第 9 题 静力学 力学 . . . 35

第 10 题 惯性参考系 力学 . . . 39

第 11 题 条件概率的递推关系 概率与统计 . . . 43

第 12 题 赌徒破产问题 概率与统计 . . . 47


## STEP III


第 1 题 参数曲线的法线 纯数学 . . . 50

第 2 题 行列式与不等式 纯数学 . . . 54

第 3 题 积分不等式 纯数学 . . . 58

第 4 题 向量与投影 纯数学 . . . 61

第 5 题 极坐标曲线 纯数学 . . . 65

第 6 题 反函数与导数 纯数学 . . . 70

第 7 题 几何中的复数 纯数学 . . . 75

第 8 题 递推关系与极限 纯数学 . . . 79

第 9 题 斜碰撞与恢复系数 力学 . . . 83

第 10 题 缠绕绳的运动 力学 . . . 87

第 11 章 指数分布 概率与统计 . . . . . 91
第 12 章 和的方差 概率与统计 . . . . . 94

详细解答

# STEP II 2021，第 1 题

纯数学 · 三角方程

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

正弦与余弦的乘积不好处理，但它们的和却很容易处理，所以开首的恒等式（identity）让我们把乘积换成和。在 (i) 中，这把一个三重乘积化成一个关于 $\cos 2x$ 的三次式（cubic），而且有一个一眼可见的因式。在 (ii) 中，我们先消去正切（tangent），再把每个因子与另一类型的对应因子配对，整条方程便会化简成两个简单因式的乘积。唯一需要额外留意的，是正切无定义的那些点。

## 解答

### 恒等式

把展开式

$$\cos(A+B)=\cos A\cos B-\sin A\sin B,\qquad \cos(A-B)=\cos A\cos B+\sin A\sin B$$

相加，得 $\cos(A+B)+\cos(A-B)=2\cos A\cos B$。取 $A=3a$ 及 $B=a$，

$$\cos a\cos 3a=\tfrac{1}{2}(\cos 4a+\cos 2a),$$

此即所求。同样地，把 $\sin(A\pm B)=\sin A\cos B\pm\cos A\sin B$ 相加，得 $\sin(A+B)+\sin(A-B)=2\sin A\cos B$。取 $A=a$ 及 $B=3a$，并利用 $\sin(-2a)=-\sin 2a$，

$$\sin a\cos 3a=\tfrac{1}{2}(\sin 4a-\sin 2a).$$

（把角色对调，取 $A=3a$ 及 $B=a$，便得到配套的 $\cos a\sin 3a=\frac{1}{2}(\sin 4a+\sin 2a)$，稍后我们也会用到。）

**第 (i) 部分**

由第一条恒等式，$4\cos x\cos 2x\cos 3x = 2\cos 2x\,(\cos 4x + \cos 2x)$。设 $c = \cos 2x$，于是 $\cos 4x = 2c^2 - 1$。方程变成

$$2c\,(2c^2 + c - 1) = 1 \quad\Longleftrightarrow\quad 4c^3 + 2c^2 - 2c - 1 = 0.$$

把各项两两分组，$2c^2(2c+1) - (2c+1) = 0$，即

$$(2c+1)(2c^2 - 1) = 0,$$

所以 $\cos 2x = -\frac{1}{2}$ 或 $\cos 2x = \pm\frac{1}{\sqrt{2}}$。

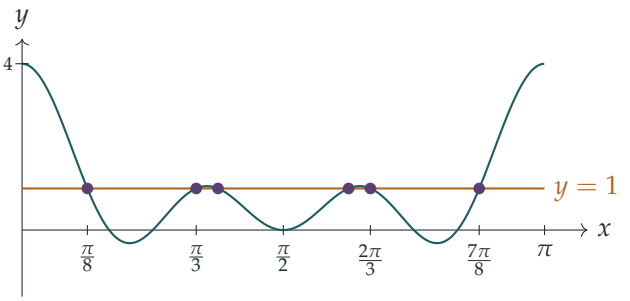
当 $0 \le x \le \pi$ 时，$0 \le 2x \le 2\pi$，而 $\cos 2x$ 每一个严格介乎 $-1$ 与 $1$ 之间的值，都在此范围内给出两个 $2x$ 的值：

$$\begin{aligned}
\cos 2x = -\tfrac{1}{2} &\implies 2x = \tfrac{2\pi}{3},\ \tfrac{4\pi}{3} &&\implies x = \tfrac{\pi}{3},\ \tfrac{2\pi}{3};\\
\cos 2x = \tfrac{1}{\sqrt{2}} &\implies 2x = \tfrac{\pi}{4},\ \tfrac{7\pi}{4} &&\implies x = \tfrac{\pi}{8},\ \tfrac{7\pi}{8};\\
\cos 2x = -\tfrac{1}{\sqrt{2}} &\implies 2x = \tfrac{3\pi}{4},\ \tfrac{5\pi}{4} &&\implies x = \tfrac{3\pi}{8},\ \tfrac{5\pi}{8}.
\end{aligned}$$

每一步都是可逆的，所以解为

$$x = \frac{\pi}{8},\ \frac{\pi}{3},\ \frac{3\pi}{8},\ \frac{5\pi}{8},\ \frac{2\pi}{3},\ \frac{7\pi}{8}.$$

图像证实曲线与直线 $y = 1$ 恰好相交六次。




**第 (ii) 部分**

设 $\tan x = \tan 2x \tan 3x \tan 4x$。要这式子有意义，四个正切都必须有定义，所以 $\cos x$、$\cos 2x$、$\cos 3x$ 及 $\cos 4x$ 全都不为零。因此我们可以两边同乘 $\cos x\cos 2x\cos 3x\cos 4x$，得

$$\sin x\cos 3x\cdot\cos 2x\cos 4x = \cos x\sin 3x\cdot\sin 2x\sin 4x.$$

? **想一想** 为甚么要把 $\sin x$ 与 $\cos 3x$ 配对，把 $\cos x$ 与 $\sin 3x$ 配对?

因为这正好是我们的恒等式能化成和的乘积，而且两个和都只涉及 $\sin 4x$ 与 $\sin 2x$，它们本来已在方程的其他地方出现。利用 $\sin x\cos 3x = \frac{1}{2}(\sin 4x - \sin 2x)$ 及 $\cos x \sin 3x = \frac{1}{2}(\sin 4x + \sin 2x)$，再乘以 2，

$$(\sin 4x - \sin 2x)\cos 2x \cos 4x = (\sin 4x + \sin 2x)\sin 2x \sin 4x.$$

现在把左边的 $\sin 2x \cos 2x$ 换成 $\frac{1}{2}\sin 4x$，使每一项都有因子 $\sin 4x$：

$$\sin 4x \cos 4x \cos 2x - \tfrac{1}{2}\sin 4x \cos 4x = \sin 4x \sin^2 2x + \sin 4x \sin 4x \sin 2x.$$

把所有项移到同一边，

$$\sin 4x \left[(\cos 4x \cos 2x - \sin 4x \sin 2x) - \left(\sin^2 2x + \tfrac{1}{2}\cos 4x\right)\right] = 0.$$

第一个括号是 $\cos(4x+2x) = \cos 6x$。至于第二个，$\sin^2 2x = \frac{1}{2}(1-\cos 4x)$，所以 $\sin^2 2x + \frac{1}{2}\cos 4x = \frac{1}{2}$。因此

$$\sin 4x \left(\cos 6x - \tfrac{1}{2}\right) = 0,$$

所以 $\cos 6x = \frac{1}{2}$ 或 $\sin 4x = 0$，此即所求。

接着求 $(\star)$ 在 $0 \le x \le \pi$ 内的解。

**想一想** $\sin 4x(\cos 6x - \frac{1}{2}) = 0$ 的每一个解都是 $(\star)$ 的解吗?

不一定：唯一可能不可逆的步骤是乘以那些余弦，而这一步恰好在四个正切都有定义时才可逆。所以 $(\star)$ 的解，就是 $\sin 4x(\cos 6x - \frac{1}{2}) = 0$ 中满足 $\cos x, \cos 2x, \cos 3x, \cos 4x \neq 0$ 的那些解。

情况 $\sin 4x = 0$。在 $[0,\pi]$ 内这给出 $x = 0,\ \frac{\pi}{4},\ \frac{\pi}{2},\ \frac{3\pi}{4},\ \pi$。在 $x = \frac{\pi}{4}$ 及 $x = \frac{3\pi}{4}$，$\tan 2x$ 无定义；在 $x = \frac{\pi}{2}$，$\tan x$ 无定义。在 $x = 0$ 及 $x = \pi$，每个正切都是 0，而 $(\star)$ 变成 $0 = 0$。所以 $x = 0$ 及 $x = \pi$ 是解。

情况 $\cos 6x = \frac{1}{2}$。此时 $6x = \pm\frac{\pi}{3} + 2k\pi$，所以 $x = \frac{(6k\pm 1)\pi}{18}$。由 $0 \le 6x \le 6\pi$ 得

$$x = \frac{\pi}{18},\ \frac{5\pi}{18},\ \frac{7\pi}{18},\ \frac{11\pi}{18},\ \frac{13\pi}{18},\ \frac{17\pi}{18}.$$

每一个都是 $\frac{m\pi}{18}$ 的形式，其中 $m$ 不是 3 的倍数。余弦 $\cos nx$ 只在 $nx$ 是 $\frac{\pi}{2} = \frac{9\pi}{18}$ 的奇数倍时才为零，这需要 $9 \mid nm$。由于 $3 \nmid m$，这就需要 $9 \mid n$，但对 $n \in \{1,2,3,4\}$ 而言这是不可能的。所以四个正切都有定义，这六个值全都是解。

总括而言，$(\star)$ 在 $0 \le x \le \pi$ 内有八个解：

$$x = 0,\ \frac{\pi}{18},\ \frac{5\pi}{18},\ \frac{7\pi}{18},\ \frac{11\pi}{18},\ \frac{13\pi}{18},\ \frac{17\pi}{18},\ \pi.$$

## 验算

- 在 (i) 中，当 $x = \frac{\pi}{8}$ 时：$\cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$，且 $\cos\frac{3\pi}{8}\cos\frac{\pi}{8} = \frac{1}{2}(\cos\frac{\pi}{2} + \cos\frac{\pi}{4}) = \frac{1}{2\sqrt{2}}$，所以乘积是 $4 \cdot \frac{1}{\sqrt{2}} \cdot \frac{1}{2\sqrt{2}} = 1$。
- 在 (ii) 中，当 $x = \frac{\pi}{18} = 10^\circ$ 时：$\tan 10^\circ \approx 0.17633$，而 $\tan 20^\circ \tan 30^\circ \tan 40^\circ \approx 0.36397 \times 0.57735 \times 0.83910 \approx 0.17633$。
- 用电脑在 $[0, \pi]$ 上扫描两条方程，分别恰好找到六个和八个解，位置正是上述各值。

## 常见错误

- 在 (ii) 中保留了 $x = \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$：它们满足 $\sin 4x = 0$，但会令 $(\star)$ 中某个正切无定义。
- 漏掉 $x = 0$ 及 $x = \pi$，或在每个周期内漏掉 $2x$（或 $6x$）的第二个值。
- 在化简过程中除以 $\sin 4x$，这会失去题目要你找出的那个因式。

## STEP 之外

积化和差公式（product-to-sum formulae）其实是 $\mathrm{e}^{\mathrm{i}A}\mathrm{e}^{\mathrm{i}B} = \mathrm{e}^{\mathrm{i}(A+B)}$ 在实变量世界中的投影：写成 $\cos\theta = \frac{1}{2}(\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta})$，便能把任何余弦的乘积化成和，而这正是通往傅里叶级数（Fourier series）的第一步——在傅里叶级数中，函数被展开成 $\cos nx$ 与 $\sin nx$ 的和。藏在 (ii) 中的惊人恒等式 $\tan 10^\circ = \tan 20^\circ \tan 30^\circ \tan 40^\circ$，正是把三角函数值看成单位根（roots of unity）时常会出现的那类惊喜。Needham 的 *Visual Complex Analysis*（《直观复分析》）把这个观点发挥得非常出色。

详细解答

# STEP II 2021，第 2 题

纯数学 · 根的对称函数

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

两个数的幂和（sums of powers）完全由两个量决定：它们的和与积。一旦知道 $a+b$ 和 $a^2+b^2$，积 $ab$ 便随之确定，$a^3+b^3$ 亦然；这就是第 (i) 部分中的约束条件。第 (iii) 部分运用同一事实，只是把 $c$ “移除”：从三个变量的和中分别减去 $c$、$c^2$、$c^3$，便剩下两个变量的和，而第 (i) 部分的约束条件就变成 $c$ 必须满足的一条三次方程（cubic equation）。

## 解答

### 第 (i) 部分

由 $(a+b)^2 = a^2+2ab+b^2$ 得 $p^2 = q+2ab$，所以

$$ab = \frac{p^2-q}{2}.$$

现在利用因式分解（factorisation）$a^3+b^3 = (a+b)^3 - 3ab(a+b)$：

$$\begin{aligned} r = a^3+b^3 &= p^3 - 3p\cdot\frac{p^2-q}{2} \\ &= \frac{2p^3-3p^3+3pq}{2} = \frac{3pq-p^3}{2}. \end{aligned}$$

因此 $3pq-p^3 = 2r$，得证。留意以上推导完全没有用到 $p$、$q$、$r$ 是实数（real）这一点：此恒等式对任何复数（complex）$a$ 和 $b$ 均成立。我们在第 (iii) 部分会用到这一点。

### 第 (ii) 部分

二次方程（quadratic）$2x^2-2px+(p^2-q)=0$ 总有两个根（可能相等，也可能不是实数），由二次公式（quadratic formula）给出。把它们记为 $a$ 和 $b$。两边除以 2，二次式为

$x^2 - px + \frac{1}{2}(p^2 - q)$，所以

$$a + b = p, \qquad ab = \frac{p^2 - q}{2}.$$

这直接给出 $(*)$ 中的第一条方程。然后

$$a^2 + b^2 = (a + b)^2 - 2ab = p^2 - (p^2 - q) = q,$$

这是第二条；而

$$a^3 + b^3 = (a + b)^3 - 3ab(a + b) = p^3 - \frac{3p(p^2 - q)}{2} = \frac{3pq - p^3}{2} = r,$$

最后一步用了假设 $3pq - p^3 = 2r$。所以 $(*)$ 中三条方程全部成立。

选取这条二次方程并非碰运气：第 (i) 部分已告诉我们 $ab$ 必须是甚么，而两根之和为 $p$、积为 $\frac{1}{2}(p^2 - q)$ 的二次式，正正就是 $x^2 - px + \frac{1}{2}(p^2 - q)$。

### 第 (iii) 部分

设

$$p = a + b = s - c, \qquad q = a^2 + b^2 = t - c^2, \qquad r = a^3 + b^3 = u - c^3.$$

> **想一想** 第 (i) 部分假设了 $p$、$q$、$r$ 是实数，但这里它们涉及 $c$，而 $c$ 可能是复数。我们仍然可以使用第 (i) 部分的结果吗?

可以：正如第 (i) 部分末尾所指出，那里的论证纯粹是代数运算，从未用到实数性。所以 $3pq - p^3 = 2r$，即

$$3(s - c)(t - c^2) - (s - c)^3 = 2(u - c^3).$$

展开，

$$3(s - c)(t - c^2) = 3st - 3sc^2 - 3tc + 3c^3,$$
$$(s - c)^3 = s^3 - 3s^2c + 3sc^2 - c^3,$$

所以方程变为

$$3st - 3sc^2 - 3tc + 3c^3 - s^3 + 3s^2c - 3sc^2 + c^3 = 2u - 2c^3.$$

按 $c$ 的幂次合并：

$$6c^3 - 6sc^2 + 3(s^2 - t)c + 3st - s^3 - 2u = 0.$$

所以 $c$ 是 $6x^3 - 6sx^2 + 3(s^2 - t)x + 3st - s^3 - 2u = 0$ 的一个根，得证。

系数只取决于 $s$、$t$、$u$，而它们对 $a$、$b$、$c$ 是对称（symmetric）的。以 $a$（或 $b$）代替 $c$ 的角色重复上述论证，可知 $a$ 和 $b$ 也是根。所以另外两个根就是 $a$ 和 $b$。

**想一想** 为何 $a$、$b$、$c$ 互不相同这一点很重要?

三次方程最多只有三个根。由于 $a$、$b$、$c$ 是三个不同的根，它们就是全部的根，而且每个都是单根（simple root），所以

$$6x^3 - 6sx^2 + 3(s^2 - t)x + 3st - s^3 - 2u = 6(x-a)(x-b)(x-c).$$

（假如例如 $a = b$，我们便只知道两个根，无法确定第三个。）比较常数项（constant term），得 $3st - s^3 - 2u = -6abc = -6v$，即

$$s^3 - 3st + 2u = 6v.$$

**第 (iv) 部分**

这里 $s = 3$、$t = 1$、$u = -3$、$v = 2$。首先检查一致性：第 (iii) 部分的条件为 $27 - 9 - 6 = 12 = 6 \times 2$，所以成立。第 (iii) 部分指出 $a$、$b$、$c$ 应是

$$6x^3 - 18x^2 + 3(9-1)x + 9 - 27 + 6 = 6x^3 - 18x^2 + 24x - 12 = 0,$$

的根，即 $x^3 - 3x^2 + 4x - 2 = 0$ 的根。由于 $x = 1$ 代入得 $1 - 3 + 4 - 2 = 0$，

$$x^3 - 3x^2 + 4x - 2 = (x-1)(x^2 - 2x + 2) = (x-1)\big((x-1)^2 + 1\big),$$

所以根为 $1$、$1+\mathrm{i}$、$1-\mathrm{i}$。取

$$a = 1, \qquad b = 1 + \mathrm{i}, \qquad c = 1 - \mathrm{i}.$$

第 (iii) 部分只告诉我们该往哪里找，所以现在必须验证：

- $a + b + c = 1 + (1+\mathrm{i}) + (1-\mathrm{i}) = 3$；
- $(1+\mathrm{i})^2 = 2\mathrm{i}$ 及 $(1-\mathrm{i})^2 = -2\mathrm{i}$，所以 $a^2 + b^2 + c^2 = 1 + 2\mathrm{i} - 2\mathrm{i} = 1$；
- $(1+\mathrm{i})^3 = 2\mathrm{i}(1+\mathrm{i}) = -2 + 2\mathrm{i}$ 及 $(1-\mathrm{i})^3 = -2\mathrm{i}(1-\mathrm{i}) = -2 - 2\mathrm{i}$，所以 $a^3 + b^3 + c^3 = 1 - 4 = -3$；
- $abc = 1 \cdot (1+\mathrm{i})(1-\mathrm{i}) = 1 \cdot 2 = 2$。

$(**)$ 中四条方程全部成立。（这三个数的任何排列都同样适用。）

## 验算

- 第 (i) 部分取 $a = 1$、$b = 2$：$p = 3$、$q = 5$、$r = 9$，而 $3pq - p^3 = 45 - 27 = 18 = 2r$。

- 第 (iii) 部分取 $a,b,c=1,2,3$：$s=6$、$t=14$、$u=36$、$v=6$，而 $s^3-3st+2u=216-252+72=36=6v$。三次式为 $6x^3-36x^2+66x-36=6(x-1)(x-2)(x-3)$，与预期相符。
- 第 (i)–(iii) 部分的恒等式亦已用电脑对大量随机复数 $a$、$b$、$c$ 作验证。

## 常见错误

- 在第 (iii) 部分直接说“$a$ 和 $b$ 是另外两个根”，却没有指出三次式的系数对 $a$、$b$、$c$ 是对称的。
- 由根之积推出 $s^3-3st+2u=6v$，却没有利用三数互不相同来确定 $a$、$b$、$c$ 就是全部的根。
- 在第 (iv) 部分只求出三次方程的根便停下：题目要求你明确验证四条方程。

## STEP 之外

这道题是牛顿恒等式（Newton's identities）的一个小特例。牛顿恒等式把幂和 $a^k+b^k+c^k$ 与初等对称函数（elementary symmetric functions）$a+b+c$、$ab+bc+ca$、$abc$ 联系起来。对称多项式基本定理（fundamental theorem of symmetric polynomials）指出，每个对称多项式都可以用初等对称多项式表示，而这正是伽罗瓦理论（Galois theory）的起点。如欲深入了解，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

# STEP II 2021，第 3 题

纯数学 · 整数部分与小数部分

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

每个实数都可以唯一地分拆成一个整数部分（integer part）和一个位于 $[0,1)$ 的小数部分（fractional part）。所以当一条方程说“一个整数加上一个 $[0,1)$ 中的数等于 4.9”时，两部分都被确定了：整数必定是 4，小数部分必定是 0.9。在 (ii) 和 (iii) 中，各方程互相纠缠，但把三条方程相加，就能让每个 $\lfloor t\rfloor + \{t\}$ 重新组合成 $t$。这样便得到一条简洁的线性方程（linear equation），再从它减去每条原方程，就能把未知数分离成正好是“整数加小数部分”的形式。

## 解答

以下一直会用到：若 $n$ 是整数，$0 \le f < 1$ 且 $n + f = c$，则 $n = \lfloor c\rfloor$ 及 $f = \{c\}$。这是因为 $n \le c < n+1$，从而 $n$ 被确定为不超过 $c$ 的最大整数。

### 第 (i) 部分

在第一条方程中，$\lfloor x\rfloor$ 是整数而 $0 \le \{y\} < 1$，所以

$$\lfloor x\rfloor = 4, \qquad \{y\} = 0.9.$$

在第二条方程中，$\lfloor y\rfloor$ 是整数而 $0 \le \{x\} < 1$，所以 $\lfloor y\rfloor = \lfloor -1.4\rfloor = -2$ 及 $\{x\} = -1.4 - (-2) = 0.6$。因此

$$x = 4 + 0.6 = 4.6, \qquad y = -2 + 0.9 = -1.1.$$

它们满足两条方程（$4 + 0.9 = 4.9$ 及 $0.6 - 2 = -1.4$），而我们已证明这是唯一的解。

**第 (ii) 部分**

把三条方程相加。$x$、$y$、$z$ 各自单独出现一次，另外又以 $\lfloor\cdot\rfloor+\{\cdot\}$ 的形式分拆在两条不同的方程中出现一次：

$$(x+y+z)+\big(\{x\}+\lfloor x\rfloor\big)+\big(\lfloor y\rfloor+\{y\}\big)+\big(\{z\}+\lfloor z\rfloor\big)=3.9+5.3+5,$$

即 $2(x+y+z)=14.2$，所以

$$x+y+z=7.1.$$

? **想一想** 既然已知 $x+y+z$，从中减去其中一条原方程后，还剩下甚么？

从 $x+y+z=7.1$ 减去第一条方程，$x$ 被消去，剩下 $y-\lfloor y\rfloor$ 和 $z-\{z\}$：

$$\{y\}+\lfloor z\rfloor=7.1-3.9=3.2,$$

这正是所求的结果。同样地，减去第二和第三条方程，得

$$\lfloor x\rfloor+\{z\}=7.1-5.3=1.8,\qquad \{x\}+\lfloor y\rfloor=7.1-5=2.1.$$

每条都是“整数加小数部分”的形式，所以

$$\begin{aligned}\lfloor z\rfloor&=3, & \{y\}&=0.2,\\ \lfloor x\rfloor&=1, & \{z\}&=0.8,\\ \lfloor y\rfloor&=2, & \{x\}&=0.1.\end{aligned}$$

因此 $x=1.1$、$y=2.2$、$z=3.8$。这是唯一可能的解，而且确实成立：

$$1.1+2+0.8=3.9,\qquad 0.1+2.2+3=5.3,\qquad 1+0.2+3.8=5.$$

**第 (iii) 部分**

同样的技巧依然奏效。把三条方程相加，并利用 $2\lfloor y\rfloor+2\{y\}=2y$，

$$2x+4y+2z=14.2,\qquad \text{所以}\qquad x+2y+z=7.1.$$

从这条方程依次减去每条原方程，得

$$2\{y\}+\lfloor z\rfloor=3.2,\qquad \lfloor x\rfloor+\{z\}=1.8,\qquad \{x\}+2\lfloor y\rfloor=2.1.$$

中间那条是熟悉的类型，所以 $\lfloor x\rfloor=1$ 及 $\{z\}=0.8$。

在第三条方程中，$2\lfloor y\rfloor$ 仍然是整数而 $0\le\{x\}<1$，所以先前的推理仍然适用：$2\lfloor y\rfloor=2$ 及 $\{x\}=0.1$。因此 $\lfloor y\rfloor=1$ 及 $x=1.1$。

**想一想** 第一条方程是否仍然确定唯一的答案?

不是。这里 $2\{y\}$ 位于 $[0,2)$ 而非 $[0,1)$，所以把 3.2 写成一个整数加上一个 $[0,2)$ 中的数，共有两种方法:

- $\lfloor z\rfloor = 3$，$2\{y\} = 0.2$，所以 $\{y\} = 0.1$;
- $\lfloor z\rfloor = 2$，$2\{y\} = 1.2$，所以 $\{y\} = 0.6$。

(若 $\lfloor z\rfloor \geq 4$，则 $2\{y\} < 0$；若 $\lfloor z\rfloor \leq 1$，则 $2\{y\} \geq 2.2$；两者皆不可能。) 配合 $\lfloor y\rfloor = 1$ 及 $\{z\} = 0.8$，得到两个候选解:

$$(x,y,z) = (1.1,\ 1.1,\ 3.8) \qquad 或 \qquad (x,y,z) = (1.1,\ 1.6,\ 2.8).$$

两者都满足原方程组:

$$\begin{aligned}(1.1,1.1,3.8):&\quad 1.1+2+0.8=3.9,\quad 0.1+2.2+3=5.3,\quad 1+0.2+3.8=5;\\ (1.1,1.6,2.8):&\quad 1.1+2+0.8=3.9,\quad 0.1+3.2+2=5.3,\quad 1+1.2+2.8=5.\end{aligned}$$

所以 (iii) 中的方程组恰好有这两个解。

## 验算

如上所示，把答案代回原方程：每条方程两边都相等。作为独立的验算，用电脑让 $x$ 和 $y$ 取遍 $[-10,10]$ 中所有 0.1 的倍数进行搜索，结果 (i) 恰有一个解，(ii) 恰有一个解，而 (iii) 恰有上述两个解。留意对于 (ii) 的解，$\{y\} + \lfloor z\rfloor = 0.2 + 3 = 3.2$，与预期相符。

## 常见错误

- 算错负数的下取整 (floor)：$\lfloor -1.4\rfloor = -2$，而不是 $-1$，所以 (i) 中 $\{x\} = 0.6$。
- 在 (iii) 中误以为 $2\{y\}$ 是一个小数部分，因而只找到一个解。它其实位于 $[0,2)$，这正是第二个解出现的原因。
- 没有把候选解代回原方程检验。上面的推导只证明了它们是仅有的可能。

## STEP 之外

把 $x$ 写成 $x = \lfloor x\rfloor + \{x\}$，等同于在模 (modulo) 1 之下运算：小数部分就是 $x$ 在圆 $\mathbb{R}/\mathbb{Z}$ 上的像。下取整和小数部分在数论 (number theory) 中无处不在，例如计算整除 $n!$ 的质数 (prime) 最高次幂的勒让德 (Legendre) 公式，以及研究当 $\alpha$ 为无理数 (irrational

number）时，$n\alpha$ 的小数部分如何在 $[0,1)$ 上分布。Davenport 的 *The Higher Arithmetic*（《高等算术》）是这一系列概念的一本绝佳入门书。

详细解答

# STEP II 2021，第 4 题

纯数学 • Lambert W 函数

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 3.9 节创造一个新函数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/lambert-w.html

## 关键思路

函数 $g$ 是 $xe^x$ 的逆运算，所以整道题的重点就是把每条方程重新整理成 $ue^u = c$ 的形式，其中 $u$ 是某个新变量。一旦化成这个形式，第 (i) 部分的草图就告诉我们一切：若 $c < -\frac{1}{e}$，则无解；若 $-\frac{1}{e} < c < 0$，则有两个解，而 $g$ 永远只会选出满足 $u \geq -1$ 的那一个。至于对数（logarithm），则以代换 $x = e^u$ 处理。

## 解答

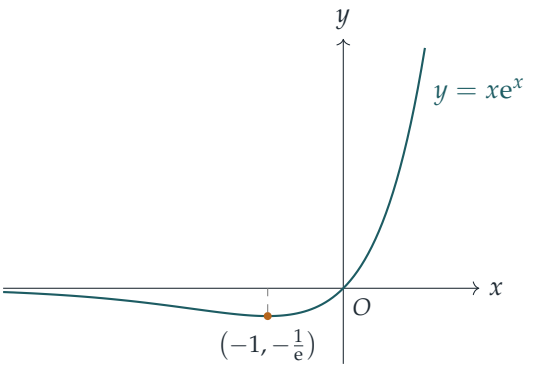
### 第 (i) 部分

设 $y = xe^x$。由乘积法则（product rule），

$$\frac{dy}{dx} = (1 + x)e^x,$$

由于 $e^x > 0$，这个导数（derivative）只在 $x = -1$ 处为零。当 $x < -1$ 时斜率（gradient）为负，当 $x > -1$ 时斜率为正，所以唯一的驻点（stationary point）是位于 $\left(-1, -\frac{1}{e}\right)$ 的极小点（minimum）。

曲线的正负号与 $x$ 相同，并且通过原点。当 $x \to \infty$ 时，$y \to \infty$，而且增长得很快。当 $x \to -\infty$ 时，$e^x$ 压倒 $x$，所以 $y \to 0$，并从下方趋近：负 $x$ 轴是一条渐近线（asymptote）。



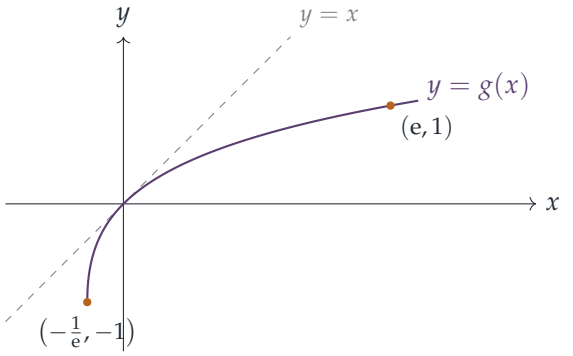

**第 (ii) 部分**

要 $f$ 有反函数（inverse function），它在 $[a,\infty)$ 上必须是一一对应（one-to-one）的。若 $a<-1$，区间包含了 $x=-1$ 附近极小点两侧的点；由于 $f$ 是连续的，并在 $(-\infty,-1]$ 上严格递减、在 $[-1,\infty)$ 上严格递增（strictly increasing），某个略大于 $-\frac{1}{\mathrm{e}}$ 的值会被取到两次。若 $a\geq -1$，则 $f$ 严格递增，因此是一一对应的。所以最小的可能值是

$$a=-1.$$

这时 $f$ 把 $[-1,\infty)$ 映射到 $[-\frac{1}{\mathrm{e}},\infty)$ 上，所以 $g$ 的定义域（domain）是 $[-\frac{1}{\mathrm{e}},\infty)$，值域（range）是 $[-1,\infty)$。

$g$ 的图像是第 (i) 部分右边那一支沿直线 $y=x$ 的反射。它从 $\left(-\frac{1}{\mathrm{e}},-1\right)$ 开始，在该处有一条铅垂切线（vertical tangent）（即极小点处水平切线的反射），然后通过原点，在原点的斜率是 1（因为 $f'(0)=1$），之后递增得越来越慢：$g(\mathrm{e})=1$，而当 $x$ 很大时，$g(x)$ 比 $\ln x$ 增长得更慢。




第 (iii) 部分会反复用到两个事实。对于 $c\geq -\frac{1}{\mathrm{e}}$，$u=g(c)$ 是方程 $u\mathrm{e}^u=c$ 满足 $u\geq -1$ 的那个解。另外，由第 (i) 部分的草图可知，方程 $u\mathrm{e}^u=c$

- 若 $c<-\frac{1}{\mathrm{e}}$，则没有实数解；

- 若 $-\frac{1}{\mathrm{e}} < c < 0$，则有两个实数解，分别是 $g(c) \in (-1, 0)$ 和另一个小于 $-1$ 的解；
- 若 $c \geq 0$，则恰有一个实数解 $g(c)$（因为当 $u < 0$ 时，左边是负数）。

**第 (iii) 部分**

**(a)** 两边乘以 $\mathrm{e}^x$（它永不为零），得

$$x\mathrm{e}^x = \tfrac{1}{5}.$$

由于 $\frac{1}{5} > 0$，恰有一个实根：

$$x = g\left(\tfrac{1}{5}\right).$$

**(b)** 要 $\ln x$ 有定义，须有 $x > 0$，所以设 $x = \mathrm{e}^u$，其中 $u = \ln x$ 为实数。方程变成

$$2u\mathrm{e}^u + 1 = 0 \iff u\mathrm{e}^u = -\tfrac{1}{2}.$$

但对所有 $u$ 都有 $u\mathrm{e}^u \geq -\frac{1}{\mathrm{e}}$，而由于 $\mathrm{e} > 2$，所以 $-\frac{1}{2} < -\frac{1}{\mathrm{e}}$。因此没有实根。

**(c)** 同样设 $x = \mathrm{e}^u$，

$$3u\mathrm{e}^u + 1 = 0 \iff u\mathrm{e}^u = -\tfrac{1}{3}.$$

由于 $\mathrm{e} < 3$，所以 $-\frac{1}{\mathrm{e}} < -\frac{1}{3} < 0$，因此 $u$ 有两个值，从而（因为 $u \mapsto \mathrm{e}^u$ 是一一对应的）$x$ 有两个实根。其中一个是

$$x = \mathrm{e}^{g(-1/3)}.$$

> ? **想一想** 这是较大的根还是较小的根?

$u$ 的另一个值小于 $-1$，而 $g(-\frac{1}{3}) > -1$。由于 $\mathrm{e}^u$ 随 $u$ 递增，根 $\mathrm{e}^{g(-1/3)}$ 是两个根中较大的一个。

**(d)** 同样有 $x > 0$。取指数，$x = 3\ln x \iff \mathrm{e}^{x/3} = x \iff x\mathrm{e}^{-x/3} = 1$。要得到 $u\mathrm{e}^u$ 的形式，前面的乘数必须与指数相配，所以设 $u = -\frac{x}{3}$，并两边乘以 $-\frac{1}{3}$：

$$\left(-\frac{x}{3}\right)\mathrm{e}^{-x/3} = -\frac{1}{3}.$$

与 (c) 一样，$-\frac{1}{\mathrm{e}} < -\frac{1}{3} < 0$，所以 $u$ 有两个值，两者都是负数，从而得到两个正根 $x = -3u$。其中一个根是

$$x = -3\,g\left(-\tfrac{1}{3}\right).$$

这次映射 $u \mapsto x = -3u$ 是递减的。$u$ 的另一个值小于 $-1$，所以另一个根大于 3，而我们求得的根满足 $x = -3g(-\frac{1}{3}) < 3$。因此我们求得的根是两者中较小的一个。

**第 (iv) 部分**

由于 $x > 0$，取对数：$x^x = 10 \iff x \ln x = \ln 10$。设 $x = \mathrm{e}^u$：

$$u\mathrm{e}^u = \ln 10.$$

由于 $\ln 10 > 0$，这方程只有一个解 $u = g(\ln 10)$，所以根是

$$x = \mathrm{e}^{g(\ln 10)}.$$

（等价地，由于当 $c \neq 0$ 时 $\mathrm{e}^{g(c)} = c/g(c)$，根也可写成 $\dfrac{\ln 10}{g(\ln 10)}$。）

## 验算

- 数值上，$g(-\frac{1}{3}) \approx -0.6191$。在 (c) 中，这给出 $x \approx 0.5384$，而 $3(0.5384)\ln(0.5384) \approx -1.000$；另一个根约为 0.2204，确实较小。
- 在 (d) 中，$x \approx 1.8572$，而 $3\ln 1.8572 \approx 1.8572$；另一个根约为 $4.536 > 3$，与预测相符。
- 在 (iv) 中，$g(\ln 10) \approx 0.9188$，所以 $x \approx 2.5062$，而 $2.5062^{2.5062} \approx 10.00$。
- 在 (b) 中，$2x \ln x + 1$ 的最小值是 $1 - \frac{2}{\mathrm{e}} \approx 0.264 > 0$（在 $x = \frac{1}{\mathrm{e}}$ 处取得），证实没有根。

## 常见错误

- 在第 (ii) 部分取 $a = 0$, 因为曲线在那里“看起来”是一一对应的: 其实转向点（turning point）在 $x = -1$，而不在原点。
- 在 (d) 中忘记整条式乘以 $-\frac{1}{3}$，以致前面的因子与指数不相配，无法应用 $g$。
- 以为经 $g$ 求得的根总是较大的那个：在 (c) 中确是如此，因为 $x = \mathrm{e}^u$ 随 $u$ 递增；但在 (d) 中并非如此，因为 $x = -3u$ 把次序倒转了。

## STEP 之外

函数 $g$ 是 *Lambert W 函数*（Lambert W function）的主支（principal branch），通常记作 $W_0$；当 $-\frac{1}{\mathrm{e}} < c < 0$ 时，$u\mathrm{e}^u = c$ 的另一个解来自第二个分支 $W_{-1}$。凡是未知数同时出现在指数函数之内和之外的地方，它都会出现：例如延迟微分方程（delay differential equations）、组合数学（combinatorics）（计算标号树（labelled trees）的数目）以及物理学。把反函数延拓越过它不再一一对应的那一点，正是复分析（complex analysis）中分支与分支点（branch points）理论的起点；Needham 的 *Visual Complex Analysis*（《直观复分析》）是初次接触这些概念的绝佳读物。

详细解答

# STEP II 2021，第 5 题

纯数学 · 一阶微分方程

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

第 (i) 部分是一条一阶（first-order）微分方程，它既不是可分离变量（separable）的，也不是一眼就能看出、形式友善的线性（linear）方程；但题目建议的代换（substitution）能把它化为可以直接积分的方程。第 (ii) 部分看似是几何题，可是一旦写出切线（tangent）方程，并要求它通过 $(1, t)$，得到的正是第 (i) 部分中 $a = 1$ 的方程。之后的草图可由导数（derivative）得出，而导数原来出奇地简单。

## 解答

### 第 (i) 部分

方程只在 $x \neq a$ 时才有意义，所以我们在 $x > a$ 或 $x < a$ 的某个区间上处理。设 $y = (x-a)u$，由乘积法则（product rule）得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = u + (x-a)\frac{\mathrm{d}u}{\mathrm{d}x}.$$

代入方程，

$$(x-a)u + (x-a)^2\frac{\mathrm{d}u}{\mathrm{d}x} = (x-a)u - x \quad \Longrightarrow \quad \frac{\mathrm{d}u}{\mathrm{d}x} = -\frac{x}{(x-a)^2}.$$

含 $u$ 的项互相抵消，这正是这个代换的用意所在。把 $x$ 写成 $x = (x-a) + a$，

$$\frac{\mathrm{d}u}{\mathrm{d}x} = -\frac{1}{x-a} - \frac{a}{(x-a)^2}, \qquad \text{故} \qquad u = -\ln|x-a| + \frac{a}{x-a} + c$$

其中 $c$ 为常数。两边乘以 $x-a$，

$$\boxed{y = a + (x-a)\big(c - \ln|x-a|\big)}$$

其中常数 $c$ 在 $x = a$ 两侧可以不同。

验证：$y' = c - \ln|x-a| - 1$，所以 $(x-a)y' = (y-a) - (x-a) = y - x$，符合要求。

**第 (ii) 部分**

在 $\big(t, \mathrm{f}(t)\big)$ 处的切线为 $y = \mathrm{f}(t) + \mathrm{f}'(t)(x-t)$。它通过 $(1,t)$ 当且仅当

$$t = \mathrm{f}(t) + \mathrm{f}'(t)(1-t), \qquad \text{即} \qquad (t-1)\,\mathrm{f}'(t) = \mathrm{f}(t) - t.$$

把 $t$ 改记为 $x$，这就是第 (i) 部分中 $a = 1$ 的方程，在 $x \neq 1$ 时成立。因此在 $x < 1$ 和 $x > 1$ 上分别有

$$\mathrm{f}(x) = 1 + (x-1)\big(c - \ln|x-1|\big),$$

而常数须在两侧各自确定。

> **想一想** 为甚么单凭一个条件，例如 $\mathrm{f}(0) = 0$，不能确定整条曲线在所有 $x$ 上的形状?

方程在 $x = 1$ 处不提供任何信息，所以 $x < 1$ 上的解与 $x > 1$ 上的解互不相关。这就是题目为每一小题分别给出条件的原因。

**(a)** 当 $x < 1$ 时，$|x-1| = 1 - x$。由条件 $\mathrm{f}(0) = 0$ 得 $0 = 1 + (-1)(c - \ln 1) = 1 - c$，故 $c = 1$，且

$$\boxed{\mathrm{f}(x) = 1 + (x-1)\big(1 - \ln(1-x)\big) \quad (x < 1).}$$

求导，$\mathrm{f}'(x) = 1 - \ln(1-x) + (x-1)\cdot\dfrac{1}{1-x} = -\ln(1-x)$。因此：

- 只有当 $1 - x = 1$ 时 $\mathrm{f}'(x) = 0$，所以唯一的驻点（stationary point）是 $(0,0)$。由于当 $x < 1$ 时 $\mathrm{f}''(x) = \dfrac{1}{1-x} > 0$，曲线是凸（convex）的，而 $(0,0)$ 是极小值（minimum）点。所以 $\mathrm{f}(x) \geq 0$，曲线在原点恰好与 $x$ 轴相切。
- 当 $x \to 1^-$ 时，设 $z = x - 1 \to 0$：利用 $z\ln|z| \to 0$，得 $\mathrm{f}(x) = 1 + z - z\ln|z| \to 1$。斜率（gradient）$-\ln(1-x) \to +\infty$，所以曲线趋向 $(1,1)$，并在该处变为铅垂。
- 当 $x \to -\infty$ 时，$x - 1$ 和 $1 - \ln(1-x)$ 都是绝对值很大的负数，所以 $\mathrm{f}(x) \to +\infty$。

一些有用的点：$\mathrm{f}(1-\mathrm{e}) = 1$ 及 $\mathrm{f}(-1) = 2\ln 2 - 1 \approx 0.39$。

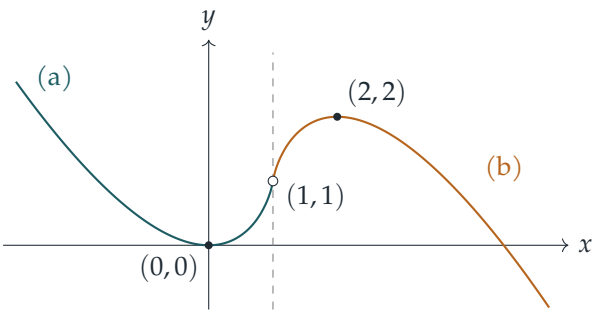
**(b)** 当 $x > 1$ 时，$|x-1| = x - 1$。由条件 $\mathrm{f}(2) = 2$ 得 $2 = 1 + 1\cdot(c - 0)$，所以同样有 $c = 1$，且

$$\mathrm{f}(x) = 1 + (x-1)\big(1 - \ln(x-1)\big) \quad (x > 1), \qquad \mathrm{f}'(x) = -\ln(x-1).$$

- 唯一的驻点在 $x = 2$，即 $(2,2)$。这里 $\mathrm{f}''(x) = -\dfrac{1}{x-1} < 0$，所以曲线是凹（concave）的，而 $(2,2)$ 是极大值（maximum）点。
- 当 $x \to 1^+$ 时，（由同一极限）$\mathrm{f}(x) \to 1$，且 $\mathrm{f}'(x) \to +\infty$：曲线以铅垂方向离开 $(1,1)$。
- 当 $x \to \infty$ 时，$1 - \ln(x-1) \to -\infty$，所以 $\mathrm{f}(x) \to -\infty$。曲线通过 $(1+\mathrm{e}, 1)$，并在 $x \approx 4.59$ 附近与 $x$ 轴相交一次。

**想一想** 做完 (a) 之后，有没有更快捷的方法看出 (b) 的形状?

有。设 $s = |x-1|$，则 (a) 中的分支为 $\mathrm{f}(1-s) = 1 - s(1-\ln s)$，而 (b) 中的分支为 $\mathrm{f}(1+s) = 1 + s(1-\ln s)$。所以 $\mathrm{f}(1+s) - 1 = -\big(\mathrm{f}(1-s) - 1\big)$：(b) 的分支就是 (a) 的分支绕 $(1,1)$ 旋转半周所得。极小值点 $(0,0)$ 被映射到极大值点 $(2,2)$。




曲线在 $x = 1$ 处没有定义（空心圆点），但两条分支都以铅垂切线趋近 $(1,1)$。

## 验算

- 直接检验切线性质。在 $t = 0$ 处，切线是水平的 $y = 0$，它确实通过 $(1,0)$。在 $t = 2$ 处，切线是 $y = 2$，它通过 $(1,2)$。
- 在 $t = -1$ 处：$\mathrm{f}(-1) \approx 0.386$，$\mathrm{f}'(-1) = -\ln 2 \approx -0.693$。切线在 $x = 1$ 处的高度为 $0.386 + 2(-0.693) \approx -1.000$，等于 $t$。
- 对称性验算：对每个 $s > 0$ 都有 $\mathrm{f}(1+s) + \mathrm{f}(1-s) = 2$，例如 $\mathrm{f}(4) + \mathrm{f}(-2) = (4 - 3\ln 3) + (3\ln 3 - 2) = 2$。

## 常见错误

- 忘记在 $\ln|x-a|$ 中加上绝对值（modulus），以致在第 (a) 部分无法处理 $x < 1$ 的情况。

- 以为 (a) 中求得的常数自动适用于 $x > 1$。这里碰巧得出相同的值，但必须用 $f(2) = 2$ 加以验证。
- 断定当 $x \to 1$ 时 $f(x) \to 0$ 或趋于无穷。$(x-1)\ln|x-1|$ 这一项趋向 0，所以 $f \to 1$。趋于无穷的是斜率。

## STEP 之外

把原点移到 $x = a$、$y = a$ 之后，方程 $(x-a)y' = y - x$ 就是齐次（homogeneous）方程：设 $X = x - a$ 及 $Y = y - a$，方程变为 $XY' = Y - X$。代换 $Y = Xu$ 正是解齐次方程的标准方法。使 $y'$ 的系数为零的点 $x = a$ 是一个奇点（singular point），而奇点两侧的解之间可以毫无关系。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）对这两个概念都有论述。

详细解答

# STEP II 2021，第 6 题

纯数学 · 圆几何的近似

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

一切都发生在同一个三角形之中：$A$、道路的圆心 $O$，以及骑车者所走圆弧的圆心 $O'$。由对称性可知，$O'$ 位于经过 $O$ 并平分 $\angle AOB$ 的直线上，而相切（tangency）的条件则确定了距离 $OO'$。在这个三角形中运用余弦定理（cosine rule）和正弦定理（sine rule），便可得出 $d$ 和 $\alpha - \theta$。至于最后一部分，远小于 1 的意思是我们只保留 $w/R$ 的一阶（first-order）项，并使用小角近似（small-angle approximation）。

## 解答

以 $O$ 为原点，并使对称轴成铅直方向，于是 $A = \big(-(R+w)\sin\alpha,\ (R+w)\cos\alpha\big)$，而 $B$ 是它的镜像。道路就是区域 $R \le r \le R+w$，其中 $r$ 是与 $O$ 的距离。

### 第 (i) 部分

弦（chord）$AB$ 上最接近 $O$ 的点是它的中点 $M$，因为 $OM \perp AB$（三角形 $OAB$ 是等腰（isosceles）三角形）。于是

$$OM = (R+w)\cos\alpha \le \tfrac{1}{2}(R+w) < \tfrac{1}{2}(R+R) = R,$$

这里先用了 $\cos\alpha \le \frac{1}{2}$（因为 $\frac{1}{3}\pi \le \alpha \le \frac{1}{2}\pi$），再用了 $w < R$。所以 $M$ 严格位于内圆之内，即不在道路上。任何由 $A$ 直线骑到 $B$ 的路线都会经过 $M$，因此不可能一直留在道路上。

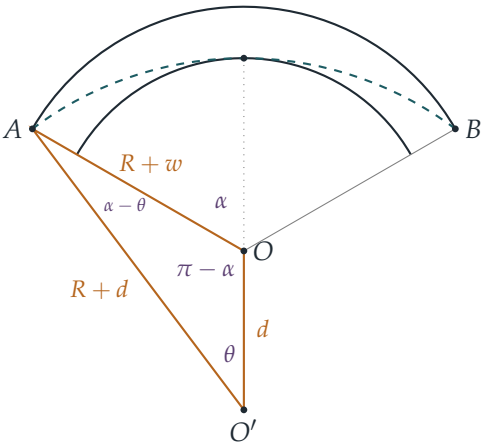
### 第 (ii) 部分

**$O'$ 在哪里？** 该圆经过 $A$ 和 $B$，所以其圆心与两点等距，位于 $AB$ 的垂直平分线（perpendicular bisector）上，也就是经过 $O$ 的对称轴。它与内圆相切，而且比内圆大（$R + d > R$），

所以内圆位于它之内，与它内切（internally tangent）。对于内切的两个圆，圆心之间的距离等于半径之差：

$$OO' = (R + d) - R = d.$$

路线向外、即朝道路的方向凸出，所以 $O'$ 位于 $O$ 背向道路的一侧，如图所示。于是在三角形 $AOO'$ 中，有 $OA = R + w$、$OO' = d$、$O'A = R + d$，而在 $O$ 的角是 $\pi - \alpha$（它与 $\angle AOM = \alpha$ 互补（supplementary））。在 $O'$ 的角是 $\frac{1}{2}\angle AO'B = \theta$，所以在 $A$ 的角是 $\pi - (\pi - \alpha) - \theta = \alpha - \theta$。




**(a)** 在三角形 $AOO'$ 中运用余弦定理，并利用 $\cos(\pi - \alpha) = -\cos\alpha$，得

$$\begin{aligned}(R + d)^2 &= (R + w)^2 + d^2 + 2(R + w)d\cos\alpha \\ R^2 + 2Rd &= R^2 + 2Rw + w^2 + 2(R + w)d\cos\alpha \\ 2d\big(R - (R + w)\cos\alpha\big) &= w(2R + w),\end{aligned}$$

所以

$$d = \frac{w(2R + w)}{2\big(R - (R + w)\cos\alpha\big)}.$$

由第 (i) 部分可知分母中的括号是正数，所以这是一个合理的正长度。

**(b)** 在同一个三角形中运用正弦定理，把角 $\alpha - \theta$（对边为 $OO' = d$）与角 $\pi - \alpha$（对边为 $O'A = R + d$）配对，得

$$\frac{\sin(\alpha - \theta)}{d} = \frac{\sin(\pi - \alpha)}{R + d} \quad \Longrightarrow \quad \sin(\alpha - \theta) = \frac{d\sin\alpha}{R + d}.$$

**第 (iii) 部分**

记 $\varepsilon = w/R$，于是 $0 < \varepsilon \ll 1$。把 (a) 的结果除以 $R$，

$$\frac{d}{R} = \frac{\varepsilon(2+\varepsilon)}{2\big(1-(1+\varepsilon)\cos\alpha\big)}.$$

由于 $\cos\alpha \le \frac{1}{2}$，分母至少是 $2\big(1-\frac{1}{2}(1+\varepsilon)\big) = 1-\varepsilon$，所以

$$0 < \frac{d}{R} \le \frac{\varepsilon(2+\varepsilon)}{1-\varepsilon} \approx 2\varepsilon,$$

这远小于 1。（准确至一阶，$d/R \approx \varepsilon/(1-\cos\alpha)$。）

? **想一想** 正弦值很小，只告诉我们该角接近 0 或接近 $\pi$。究竟是哪一个?

由 (b)，$0 < \sin(\alpha-\theta) = \dfrac{(d/R)\sin\alpha}{1+d/R} < \dfrac{d}{R}$，这是很小的。角 $\alpha-\theta$ 是三角形 $AOO'$ 的一个内角，所以是正数；又因为 $\theta > 0$，它小于 $\alpha \le \frac{1}{2}\pi$。在 $(0, \frac{1}{2}\pi)$ 上，正弦函数是递增的，且 $\sin x \ge \frac{2}{\pi}x$，所以 $\alpha-\theta \le \frac{\pi}{2}\sin(\alpha-\theta) < \frac{\pi}{2}\cdot\frac{d}{R}$，这同样远小于 1。

**第 (iv) 部分**

我的朋友沿外边缘骑行，那是一段半径为 $R+w$、在 $O$ 所对的角为 $2\alpha$ 的圆弧，所以他们的路线长度是 $L = 2\alpha(R+w)$。我的路线是半径为 $R+d$ 的圆弧，位于 $AB$ 背向 $O'$ 的一侧；它在 $O'$ 所对的角为 $2\theta$，所以长度是 $2\theta(R+d)$。因此

$$\begin{aligned} S &= 2\alpha(R+w) - 2\theta(R+d) \\ &= 2(R+d)(\alpha-\theta) + 2\alpha(R+w) - 2\alpha(R+d) \\ &= 2(R+d)(\alpha-\theta) + 2\alpha(w-d), \end{aligned}$$

正是所求（我们加上又减去了 $2\alpha(R+d)$）。

**近似**。现在我们准确至 $\varepsilon$ 的一阶，忽略相对于所保留各项而言为 $\varepsilon^2$ 阶的项。由 (iii)，$\alpha-\theta$ 很小，所以 $\alpha-\theta \approx \sin(\alpha-\theta)$，误差为 $(\alpha-\theta)^3$ 阶。由 (b)，

$$2(R+d)(\alpha-\theta) \approx 2(R+d)\cdot\frac{d\sin\alpha}{R+d} = 2d\sin\alpha.$$

因此

$$S \approx 2d\sin\alpha + 2\alpha w - 2\alpha d = 2\alpha w - 2d(\alpha-\sin\alpha).$$

现在利用 (iii) 中的 $d \approx \dfrac{w}{1-\cos\alpha}$:

$$S \approx 2w\left(\alpha - \frac{\alpha-\sin\alpha}{1-\cos\alpha}\right) = 2w\cdot\frac{\alpha-\alpha\cos\alpha-\alpha+\sin\alpha}{1-\cos\alpha} = \frac{2w(\sin\alpha-\alpha\cos\alpha)}{1-\cos\alpha}.$$

最后，$L = 2\alpha R(1+\varepsilon) \approx 2\alpha R$；由于 $S$ 本身已经是 $\varepsilon$ 的一阶量，把 $L$ 换成 $2\alpha R$ 只会在二阶上改变 $S/L$。所以

$$\frac{S}{L} \approx \left(\frac{\sin\alpha - \alpha\cos\alpha}{\alpha(1-\cos\alpha)}\right)\frac{w}{R},$$

这就是所求的比例。

## 验算

- 当 $\alpha = \frac{1}{2}\pi$ 时，该比例是 $\frac{2}{\pi}\cdot\frac{w}{R} \approx 0.637\,\frac{w}{R}$；当 $\alpha = \frac{1}{3}\pi$ 时，约为 $0.654\,\frac{w}{R}$。两者都是正数，这是必然的，因为 $\tan\alpha > \alpha$ 给出 $\sin\alpha > \alpha\cos\alpha$。
- 取 $R = 1$、$w = 0.01$、$\alpha = \frac{1}{2}\pi$：精确公式给出 $d \approx 0.01005$、$\alpha - \theta \approx 0.00995$ 及 $S/L \approx 0.006285$，而近似值是 $0.006366$。相对误差约为 1%，属于 $w/R$ 的量级，理应如此；当 $w$ 除以十时，误差亦缩小至十分之一。
- 一项电脑验算把 $A$ 和 $B$ 放在坐标上，直接求出相切的圆，并以数千组随机选取的 $R$、$w$ 和 $\alpha$ 证实了 $d$、$\sin(\alpha-\theta)$ 和 $S$ 的公式。

## 常见错误

- 把距离取为 $OO' = R + d + R$ 或 $OO' = w$：两圆是内切，所以圆心之间的距离是半径之差，即 $d$。
- 在三角形 $AOO'$ 中，对于在 $O$ 的角错用了 $\cos\alpha$ 而不是 $\cos(\pi-\alpha)$，这会令最后一项的正负号倒转。
- 在 (iii) 中只说 $\sin(\alpha-\theta)$ 很小：你还需要 $\alpha-\theta$ 位于 $(0, \frac{1}{2}\pi)$ 内，才可以断定角本身很小。

## STEP 之外

“远小于 1”是微扰理论（perturbation theory）的语言：以一个小参数 $\varepsilon$ 展开，保留首项，并确保误差确实是更高阶的。在大学里，这会借助泰勒定理（Taylor's theorem）和大 $O$ 记号（big-$O$ notation）严格化；在本题中，它会显示 $S/L$ 等于所给比例乘以 $\varepsilon$，再加上一个 $\varepsilon^2$ 阶的项。关于带余项的泰勒定理的严谨论述，可参看 Abbott 的 *Understanding Analysis*（《理解分析》）；如想做更多同类风格的题目，可参看 Siklos 的 *Advanced Problems in Mathematics*（《高等数学题集》）。

详细解答

# STEP II 2021，第 7 题

纯数学 · 旋转矩阵

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

$2 \times 2$ 旋转矩阵（rotation matrix）的形状非常固定：它必定是 $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$，因此它的每一列都是单位向量（unit vector）。要求 $\mathbf{R}+\mathbf{I}$ 保持这个形状，就能把 $\phi$ 完全确定下来。对于一般满足 $\mathbf{S}^3 = \mathbf{I}$ 的矩阵，有用的工具是行列式（determinant）（它具有乘性（multiplicative））以及关系式 $\mathbf{S}^2 = (a+d)\mathbf{S} - \mathbf{I}$；后者让我们可以把 $\mathbf{S}$ 的任何幂写成 $\mathbf{S}$ 与 $\mathbf{I}$ 的组合。

## 解答

### 第 (i) 部分

记 $c = \cos\phi$ 及 $s = \sin\phi$，于是

$$\mathbf{R} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}, \qquad \mathbf{R}+\mathbf{I} = \begin{pmatrix} 1+c & -s \\ s & 1+c \end{pmatrix}.$$

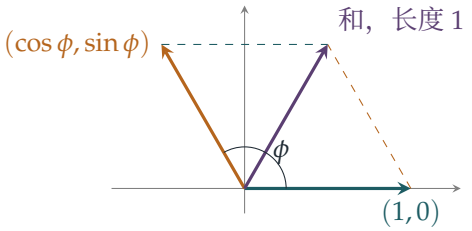
它本身已具有 $\begin{pmatrix} p & -q \\ q & p \end{pmatrix}$ 的形式，而这样的矩阵是旋转，当且仅当 $p^2+q^2=1$（此时对某个 $\theta$ 有 $(p,q) = (\cos\theta, \sin\theta)$）。因此我们需要

$$(1+c)^2 + s^2 = 1 \iff 2+2c = 1 \iff \cos\phi = -\tfrac{1}{2}.$$

在 $0^\circ \le \phi < 360^\circ$ 的范围内，得 $\phi = 120^\circ$ 或 $\phi = 240^\circ$。

从几何角度看，$\mathbf{R}+\mathbf{I}$ 的第一列是单位向量 $(1,0)$ 与 $(\cos\phi, \sin\phi)$ 之和，而我们要求这个和也是单位向量：三个向量构成一个等边三角形（equilateral triangle），如下图（$\phi = 120^\circ$）

所示。




无论哪种情况，$\mathbf{R}^3$ 都是转过 $3\phi$ 的旋转，即 360° 或 720°，是整数圈。因此 $\mathbf{R}^3 = \mathbf{I}$。

**第 (ii) 部分**

**行列式**。由于 det 具有乘性，$(\det \mathbf{S})^3 = \det(\mathbf{S}^3) = \det \mathbf{I} = 1$。矩阵是实矩阵，所以 $\det \mathbf{S}$ 是一个立方等于 1 的实数，而这样的实数只有 1。故 $\det \mathbf{S} = ad - bc = 1$。

**二次关系式**。展开得

$$\mathbf{S}^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & bc + d^2 \end{pmatrix}, \qquad (a+d)\mathbf{S} - \mathbf{I} = \begin{pmatrix} a^2 + ad - 1 & ab + bd \\ ac + dc & ad + d^2 - 1 \end{pmatrix}.$$

非对角元素（off-diagonal entries）相同，而对角元素亦相同，因为 $bc = ad - 1$（即 $\det \mathbf{S} = 1$）。因此 $\mathbf{S}^2 = (a+d)\mathbf{S} - \mathbf{I}$，得证。

**迹**。设 $t = a + d$（即迹（trace））。把关系式乘以 $\mathbf{S}$，再用一次该关系式：

$$\mathbf{S}^3 = t\mathbf{S}^2 - \mathbf{S} = t\big(t\mathbf{S} - \mathbf{I}\big) - \mathbf{S} = (t^2 - 1)\mathbf{S} - t\mathbf{I}.$$

令 $\mathbf{S}^3 = \mathbf{I}$，得

$$(t^2 - 1)\mathbf{S} = (t+1)\mathbf{I}.$$

? **想一想** 我们可以直接除以 $t^2 - 1$ 就完事吗?

不可以掉以轻心：我们必须另外考虑 $t^2 - 1 = 0$ 的情况，而事实上正正就是这个情况会出现。

- 若 $t^2 \neq 1$，则 $\mathbf{S} = \dfrac{t+1}{t^2-1}\mathbf{I} = k\mathbf{I}$，其中 $k$ 为实数。于是 $\mathbf{S}^3 = k^3\mathbf{I} = \mathbf{I}$ 迫使 $k^3 = 1$，所以 $k = 1$ 而 $\mathbf{S} = \mathbf{I}$，但这已被排除。
- 所以 $t^2 = 1$。方程变为 $\mathbf{0} = (t+1)\mathbf{I}$，故 $t = -1$（若 $t = 1$ 则会得出 $2\mathbf{I} = \mathbf{0}$，这是错的）。

因此 $a + d = -1$。

**第 (iii) 部分**

首先，$\mathbf{S} \neq \mathbf{I}$：否则 $\mathbf{S}+\mathbf{I}=2\mathbf{I}$，其行列式为 4，而且会把长度加倍，所以不是旋转。因此第 (ii) 部分适用，得 $a+d=-1$。

由于 $\mathbf{S}+\mathbf{I}$ 是旋转，设其转角为 $\theta$，则

$$\mathbf{S}=\begin{pmatrix}\cos\theta-1 & -\sin\theta\\ \sin\theta & \cos\theta-1\end{pmatrix}.$$

迹的条件给出 $2\cos\theta-2=-1$，所以 $\cos\theta=\frac{1}{2}$ 及 $\sin\theta=\pm\frac{\sqrt{3}}{2}$。于是

$$\mathbf{S}=\begin{pmatrix}-\frac{1}{2} & -\sin\theta\\ \sin\theta & -\frac{1}{2}\end{pmatrix},$$

它具有旋转的形式，且 $\left(-\frac{1}{2}\right)^2+\sin^2\theta=\frac{1}{4}+\frac{3}{4}=1$。所以 $\mathbf{S}$ 是旋转，其余弦为 $-\frac{1}{2}$，正弦为 $\pm\frac{\sqrt{3}}{2}$。

转角是 120°（当 $\theta=60°$）或 240°（当 $\theta=300°$）。两者确实都会出现：由第 (i) 部分，这两个旋转都使 $\mathbf{S}+\mathbf{I}$ 为旋转，且 $\mathbf{S}^3=\mathbf{I}$。（等价地，一旦知道 $\mathbf{S}$ 是旋转，第 (i) 部分本身已列出所有可能的转角。）

## 验算

- 当 $\phi=120°$，$\mathbf{R}+\mathbf{I}=\begin{pmatrix}\frac{1}{2} & -\frac{\sqrt{3}}{2}\\ \frac{\sqrt{3}}{2} & \frac{1}{2}\end{pmatrix}$，即转过 60° 的旋转；当 $\phi=240°$，它是转过 300° 的旋转。
- 矩阵 $\mathbf{S}=\begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix}$ 满足 $\mathbf{S}^3=\mathbf{I}$（乘出来便知），但不是旋转；它仍然有行列式 1 和迹 $-1$，与第 (ii) 部分的预测一致。它的 $\mathbf{S}+\mathbf{I}=\begin{pmatrix}1 & -1\\ 1 & 0\end{pmatrix}$ 不是旋转，与第 (iii) 部分相符。
- 用电脑检查了大量满足 $\mathbf{S}^3=\mathbf{I}$、$\mathbf{S}\neq\mathbf{I}$ 的随机实矩阵，每次都确认行列式为 1、迹为 $-1$，且 $\mathbf{S}^2$ 的关系式成立；在第 (iii) 部分扫描 $\theta$，只找到 $\theta=60°,300°$。

## 常见错误

- 除以 $t^2-1$ 而没有检查它会否为零；$t^2=1$ 的情况恰恰就是答案。
- 忘记在第 (iii) 部分中，必须先确立 $\mathbf{S}\neq\mathbf{I}$，才可以使用第 (ii) 部分。

- 以为 $\det(\mathbf{S}+\mathbf{I}) = 1$ 就足以令 $\mathbf{S}+\mathbf{I}$ 成为旋转；在这里它是自动成立的，所以不提供任何信息。

## STEP 之外

关系式 $\mathbf{S}^2 = (a+d)\mathbf{S} - (\det \mathbf{S})\mathbf{I}$ 是凯莱－哈密顿定理（Cayley–Hamilton theorem）在 $2\times 2$ 情况下的特例：每个方阵都满足它自己的特征多项式（characteristic polynomial）。第 (ii) 部分说明，一个阶为 3 的 $2\times 2$ 实矩阵，其特征值（eigenvalue）是两个非实数的三次单位根（cube roots of unity），而这样的矩阵必定相似（similar）于转过 $120^\circ$ 的旋转，但不一定等于它。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）对特征值和相似性有深入的发展。

详细解答

# STEP II 2021，第 8 题

纯数学 · 积分递推关系

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 2.9 节逐步逼近 $e$ 的分数中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/e-approximations.html

## 关键思路

函数 $t^n(1-t)^n$ 是单一个二次式 $u=t(1-t)$ 的幂，所以对它求导只不过是用两次链式法则（chain rule）。这个函数及其一阶导数（derivative）在 $t=0$ 和 $t=1$ 处都等于零，所以把它与 $\mathrm{e}^t$ 一起进行两次分部积分（integration by parts）时，所有边界项（boundary terms）都会消失，而第 (i) 部分的二阶导数就直接变成一条递推关系（recurrence）。这些积分被夹在 0 与一个极小的数之间，由此迫使分数 $-a_n/b_n$ 成为 e 的极佳近似。

## 解答

### 第 (i) 部分

设 $u=t(1-t)=t-t^2$，于是 $\dfrac{\mathrm{d}u}{\mathrm{d}t}=1-2t$，$\dfrac{\mathrm{d}^2u}{\mathrm{d}t^2}=-2$，且

$$(1-2t)^2=1-4t+4t^2=1-4u.$$

设 $f(t)=t^n(1-t)^n=u^n$，由链式法则得 $f'(t)=nu^{n-1}(1-2t)$。再用乘积法则（product rule）求导一次（对 $n\geq 2$ 成立，此时 $u^{n-2}$ 是多项式），

$$\begin{aligned}f''(t)&=n(n-1)u^{n-2}(1-2t)^2-2nu^{n-1}\\&=nu^{n-2}\big[(n-1)(1-4u)-2u\big]\\&=nu^{n-2}\big[(n-1)-(4n-2)u\big]\\&=nt^{n-2}(1-t)^{n-2}\big[(n-1)-2(2n-1)t(1-t)\big],\end{aligned}$$

这正是所求的结果。

**第 (ii) 部分**

仍设 $f(t)=t^n(1-t)^n$，其中 $n\geq 2$，于是 $n!\,T_n=\int_0^1 f(t)\,\mathrm{e}^t\,\mathrm{d}t$。进行两次分部积分，每次都对多项式求导、对 $\mathrm{e}^t$ 积分，

$$\int_0^1 f\,\mathrm{e}^t\,\mathrm{d}t=\Big[f\,\mathrm{e}^t\Big]_0^1-\Big[f'\,\mathrm{e}^t\Big]_0^1+\int_0^1 f''\,\mathrm{e}^t\,\mathrm{d}t.$$

? **想一想** 为甚么两个边界项都会消失?

因为 $f(t)=u^n$ 和 $f'(t)=nu^{n-1}(1-2t)$ 都含有因子 $u=t(1-t)$（这里用到 $n\geq 2$，故 $n-1\geq 1$），而在 $t=0$ 和 $t=1$ 处都有 $u=0$。所以

$$n!\,T_n=\int_0^1 f''(t)\,\mathrm{e}^t\,\mathrm{d}t.$$

现在代入第 (i) 部分的结果，把它写成 $f''=n(n-1)u^{n-2}-2n(2n-1)u^{n-1}$：

$$\begin{aligned}n!\,T_n&=n(n-1)\int_0^1 u^{n-2}\mathrm{e}^t\,\mathrm{d}t-2n(2n-1)\int_0^1 u^{n-1}\mathrm{e}^t\,\mathrm{d}t\\&=n(n-1)\,(n-2)!\,T_{n-2}-2n(2n-1)\,(n-1)!\,T_{n-1}\\&=n!\,T_{n-2}-2(2n-1)\,n!\,T_{n-1},\end{aligned}$$

这里用到 $\int_0^1 u^k\mathrm{e}^t\,\mathrm{d}t=k!\,T_k$。两边除以 $n!$ 得

$$T_n=T_{n-2}-2(2n-1)T_{n-1}\qquad(n\geq 2),$$

即为所求。

**第 (iii) 部分**

首先，$T_0=\int_0^1 \mathrm{e}^t\,\mathrm{d}t=\mathrm{e}-1$。

至于 $T_1=\int_0^1(t-t^2)\mathrm{e}^t\,\mathrm{d}t$，可以通过求导验证 $\int t\,\mathrm{e}^t\,\mathrm{d}t=(t-1)\mathrm{e}^t$ 及 $\int t^2\mathrm{e}^t\,\mathrm{d}t=(t^2-2t+2)\mathrm{e}^t$，所以

$$\int_0^1(t-t^2)\mathrm{e}^t\,\mathrm{d}t=\Big[(-t^2+3t-3)\mathrm{e}^t\Big]_0^1=-\mathrm{e}+3,$$

即 $T_1 = 3 - \mathrm{e}$。

因此 $T_0 = a_0 + b_0\mathrm{e}$，其中 $(a_0, b_0) = (-1, 1)$；$T_1 = a_1 + b_1\mathrm{e}$，其中 $(a_1, b_1) = (3, -1)$，全部都是整数。假设对某个 $n \geq 2$，$T_{n-2}$ 和 $T_{n-1}$ 都具有这种整数系数的形式。由第 (ii) 部分，

$$T_n = \big(a_{n-2} - 2(2n-1)a_{n-1}\big) + \big(b_{n-2} - 2(2n-1)b_{n-1}\big)\mathrm{e},$$

由于 $2(2n-1)$ 是整数，两个括号内都是整数。由（强）归纳法（strong induction），每个 $n \geq 0$ 的 $T_n$ 都可写成 $a_n + b_n\mathrm{e}$，其中 $a_n, b_n \in \mathbb{Z}$。

### 第 (iv) 部分

**上下界**。当 $0 < t < 1$ 时，有 $0 < t(1-t) \leq \frac{1}{4} < 1$，所以 $0 < t^n(1-t)^n \leq 1$，而且 $0 < \mathrm{e}^t < \mathrm{e}$。因此在开区间 $(0, 1)$ 上

$$0 < \frac{t^n(1-t)^n}{n!}\,\mathrm{e}^t < \frac{\mathrm{e}}{n!}.$$

被积函数（integrand）是连续的，所以把这些严格不等式在 $[0, 1]$（一个长度为 1 的区间）上积分后，不等式仍然是严格的：

$$0 < T_n < \frac{\mathrm{e}}{n!} \qquad (n \geq 0).$$

（当 $n = 0$ 时，左边的因子就是 1，严格性来自 $\mathrm{e}^t < \mathrm{e}$。）

**极限**。由于 $b_n \neq 0$，把 $T_n = a_n + b_n\mathrm{e}$ 除以 $b_n$：

$$-\frac{a_n}{b_n} - \mathrm{e} = -\frac{T_n}{b_n}, \qquad \text{所以} \qquad \left|-\frac{a_n}{b_n} - \mathrm{e}\right| = \frac{T_n}{|b_n|} \leq T_n < \frac{\mathrm{e}}{n!}.$$

这里用到 $b_n$ 是非零整数，所以 $|b_n| \geq 1$。当 $n \to \infty$ 时，$\mathrm{e}/n! \to 0$，所以 $-a_n/b_n \to \mathrm{e}$。

事实上，这个界非常宽松。由递推关系得 $(a_2, b_2) = (-19, 7)$ 及 $(a_3, b_3) = (193, -71)$，给出近似值 $\frac{19}{7} \approx 2.714$ 和 $\frac{193}{71} \approx 2.71831$，而 $\mathrm{e} \approx 2.71828$。

## 验算

- 在第 (i) 部分代入 $n = 2$：$t^2(1-t)^2 = t^2 - 2t^3 + t^4$ 的二阶导数是 $2 - 12t + 12t^2$，而公式给出 $2\big[1 - 6t(1-t)\big] = 2 - 12t + 12t^2$。
- $T_1 = 3 - \mathrm{e} \approx 0.2817$ 是正数且小于 $\mathrm{e}/1!$，正如第 (iv) 部分所要求。接着 $T_2 = T_0 - 6T_1 = 7\mathrm{e} - 19 \approx 0.0280$；用电脑计算积分可确认这个值，并确认 $n$ 直至 11 时递推关系和上下界都成立。
- 当 $n = 4, 5, 6$ 时，$-a_n/b_n$ 的值分别是 $\frac{2721}{1001}$、$\frac{49171}{18089}$、$\frac{1084483}{398959}$，与 $\mathrm{e}$ 分别相差约 $10^{-7}$、$3 \times 10^{-10}$ 和 $5 \times 10^{-13}$：比单凭 $\mathrm{e}/n!$ 所预示的快得多。

## 常见错误

- 忘记说明边界项为何消失，或没有留意这需要 $n \geq 2$：当 $n = 1$ 时，导数 $1 - 2t$ 在两端都不为零。
- 在第 (iv) 部分只用了 $b_n \neq 0$，而没有用 $|b_n| \geq 1$。一个非零实数可以非常小；正是因为 $b_n$ 是整数，$T_n / b_n$ 才不会变得无限大。

## STEP 之外

这正是埃尔米特（Hermite）在 1873 年证明 e 是无理数（irrational）、甚至是超越数（transcendental）的方法的核心：构造一些非零但趋向零的整数组合 $a_n + b_n \mathrm{e}$。若 $\mathrm{e} = p/q$，则 $q(a_n + b_n \mathrm{e}) = qa_n + b_n p$ 会是一个小于 $q\mathrm{e}/n!$ 的正整数，而当 $n$ 足够大时这是不可能的；同样的多项式 $t^n(1-t)^n$ 亦出现在 Beukers 证明 $\zeta(3)$ 是无理数的著名证明中。若想进一步了解数能被分数逼近到甚么程度，可阅读 Davenport 的 *The Higher Arithmetic*（《高等算术》）。

详细解答

# STEP II 2021，第 9 题

力学 · 静力学

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

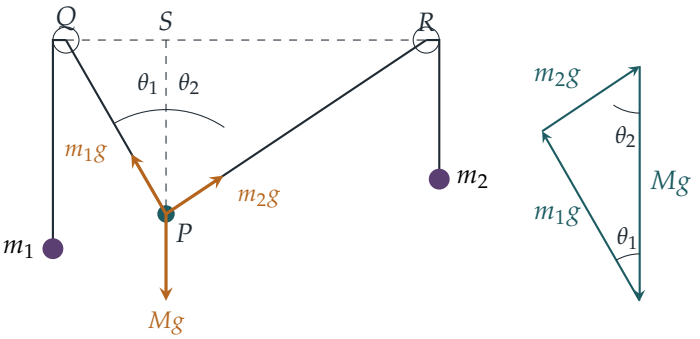
www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

由于滑轮（pulley）是光滑的，每段绳的张力（tension）都等于其远端所悬挂质点的重量，所以作用在 $P$ 上的三个力，大小之比为 $m_1 : m_2 : M$。这三个力首尾相接，组成一个力三角形（triangle of forces），三边分别为 $m_1 g$、$m_2 g$ 和 $Mg$，而各内角则由 $\theta_1$ 和 $\theta_2$ 构成。对这个三角形应用余弦定律（cosine rule），便能得出所有结果：题中的不等式来自要求该三角形存在且两个角均为锐角（acute），而所求的比则来自 $\tan\theta_1$ 和 $\tan\theta_2$。

## 解答

**模型**。各质点（particle）视为点质量，绳是轻而不可伸长的，滑轮细小且光滑。因此 $PQ$ 段的张力等于支撑 $m_1$ 的铅直段的张力，而 $m_1$ 处于平衡（equilibrium），故 $T_1 = m_1 g$。同理，$PR$ 段的张力为 $T_2 = m_2 g$。记 $h = SP$。



**第 (i) 部分**

在 $P$ 点沿水平和铅直方向分解（并约去 $g$）：

$$m_1 \sin\theta_1 = m_2 \sin\theta_2, \qquad m_1 \cos\theta_1 + m_2 \cos\theta_2 = M. \tag{$*$}$$

等价地说，把三个力首尾相接，会组成一个封闭三角形，三边为 $m_1 g$、$m_2 g$、$Mg$（见图右方）。边 $Mg$ 是铅直的，所以边 $Mg$ 与边 $m_1 g$ 之间的夹角为 $\theta_1$（与边 $m_2 g$ 相对），而 $Mg$ 与 $m_2 g$ 之间的夹角为 $\theta_2$（与 $m_1 g$ 相对）。第三个角与 $Mg$ 相对，大小为 $180^\circ - \theta_1 - \theta_2$。

如果你较喜欢用代数而非图形，同样的结果亦可直接由 $(*)$ 得出：写成 $m_1 \cos\theta_1 = M - m_2 \cos\theta_2$ 及 $m_1 \sin\theta_1 = m_2 \sin\theta_2$，两式平方后相加：

$$m_1^2 = M^2 + m_2^2 - 2Mm_2 \cos\theta_2, \quad \text{所以} \quad \cos\theta_2 = \frac{M^2 + m_2^2 - m_1^2}{2Mm_2},$$

再把两边的角色互换，得

$$\cos\theta_1 = \frac{M^2 + m_1^2 - m_2^2}{2Mm_1}.$$

**(a)** 力三角形是一个真正的（非退化的）三角形，因为 $\theta_1 + \theta_2$ 严格介于 $0^\circ$ 与 $180^\circ$ 之间。因此三角不等式（triangle inequality）是严格的，由此得 $M < m_1 + m_2$。

至于下界，我们利用 $\theta_2$ 是锐角这一事实，即 $\cos\theta_2 > 0$。由于 $2Mm_2 > 0$，上面的公式给出

$$M^2 + m_2^2 - m_1^2 > 0 \iff M^2 > m_1^2 - m_2^2 \iff M > \sqrt{m_1^2 - m_2^2},$$

其中最后一步成立，是因为 $M > 0$ 且 $m_1^2 - m_2^2 > 0$。合起来，

$$\sqrt{m_1^2 - m_2^2} < M < m_1 + m_2.$$

? **想一想** 为甚么重要的是关于 $\theta_2$ 的条件，而不是关于 $\theta_1$ 的条件?

在三角形中，较长的边对着较大的角。由于 $m_1 > m_2$，角 $\theta_2$（与 $m_1 g$ 相对）大于 $\theta_1$（与 $m_2 g$ 相对）。所以只要 $\theta_2$ 是锐角，$\theta_1$ 自然也是锐角；真正起限制作用的只有 $\theta_2$。（事实上，由于 $m_1 > m_2$，对任何 $M > 0$ 都有 $\cos\theta_1 > 0$。）

**(b)** 在直角三角形 $QSP$ 和 $RSP$ 中，$SP = h$，

$$QS = h\tan\theta_1, \qquad SR = h\tan\theta_2.$$

因此

$$r = \frac{QS}{SR} = \frac{\tan\theta_1}{\tan\theta_2} = \frac{\sin\theta_1}{\sin\theta_2}\cdot\frac{\cos\theta_2}{\cos\theta_1}$$
$$= \frac{m_2}{m_1}\cdot\frac{M^2+m_2^2-m_1^2}{2Mm_2}\cdot\frac{2Mm_1}{M^2+m_1^2-m_2^2} \quad \text{（利用 } (*) \text{ 及上述余弦式）}$$
$$= \frac{M^2-m_1^2+m_2^2}{M^2-m_2^2+m_1^2},$$

这正是所求的比。由 (a) 部，分子分母两个式子均为正，所以 $r>0$；又因 $m_1>m_2$，故 $r<1$：点 $S$ 较接近 $Q$，与图中所示一致。

**第 (ii) 部分**

对力三角形中的边 $Mg$ 应用余弦定律，其对角为 $180^\circ-(\theta_1+\theta_2)$，得

$$M^2 = m_1^2+m_2^2+2m_1m_2\cos(\theta_1+\theta_2).$$

（用代数方法：把 $(*)$ 中两条方程平方后相加。）由 $M^2=m_1^2+m_2^2$，必然有 $\cos(\theta_1+\theta_2)=0$。由于 $\theta_1$ 和 $\theta_2$ 均为锐角，$0^\circ<\theta_1+\theta_2<180^\circ$，唯一的可能是

$$\theta_1+\theta_2 = 90^\circ.$$

所以绳在 $P$ 点成直角。（这个 $M$ 值确实落在 (a) 部的范围内，因此这个构型是存在的。）

此时余弦式得以简化：$\cos\theta_1=\dfrac{2m_1^2}{2Mm_1}=\dfrac{m_1}{M}$ 及 $\cos\theta_2=\dfrac{m_2}{M}$。由于 $\theta_2=90^\circ-\theta_1$，得 $\sin\theta_1=\cos\theta_2=m_2/M$，所以

$$\tan\theta_1=\frac{m_2}{m_1}, \qquad \tan\theta_2=\frac{1}{\tan\theta_1}=\frac{m_1}{m_2}.$$

因此

$$\frac{QR}{SP}=\frac{h\tan\theta_1+h\tan\theta_2}{h}=\frac{m_2}{m_1}+\frac{m_1}{m_2}=\frac{m_1^2+m_2^2}{m_1m_2},$$

即 $QR:SP=(m_1^2+m_2^2):m_1m_2$。

## 验算

- 取 $m_1=5$、$m_2=3$、$M=6$（在 $4<M<8$ 之内），以数值方法解 $(*)$，得 $\theta_1\approx29.93^\circ$ 及 $\theta_2\approx56.25^\circ$。于是 $\tan\theta_1/\tan\theta_2\approx0.3846=\frac{5}{13}$，与 $r=\frac{36-25+9}{36-9+25}=\frac{20}{52}$ 吻合。
- 在范围的两端：当 $M\to\sqrt{m_1^2-m_2^2}$ 时，$\theta_2\to90^\circ$ 且 $r\to0$（绳 $PR$ 变成水平）；当 $M\to m_1+m_2$ 时，两个角都趋向 0，而 $P$ 会下沉至远低于 $QR$ 的位置。
- 第 (ii) 部分取 $m_1=4$、$m_2=3$、$M=5$：$\tan\theta_1=\frac{3}{4}$，$\tan\theta_2=\frac{4}{3}$，且 $QR/SP=\frac{25}{12}$。验证 $(*)$：$4\cdot\frac{3}{5}=3\cdot\frac{4}{5}$ 及 $4\cdot\frac{4}{5}+3\cdot\frac{3}{5}=5$。

## 常见错误

- 误把力三角形中与 $Mg$ 相对的角当作 $\theta_1 + \theta_2$，而不是 $180^\circ - (\theta_1 + \theta_2)$，结果令余弦定律中的正负号反转。
- 使用非严格的三角不等式，或忘记下界需要用到两个角均为锐角这个条件。
- 把比倒转：$r = QS/SR$，而 $QS$ 是靠近较重质点的那一边。

## STEP 之外

力三角形是一个一般事实的最简单情况：处于平衡的各力之和为零向量，所以把它们首尾相接画出来，会组成一个封闭多边形。工程师利用这个想法（即“力多边形”（force polygon））来分析桁架结构；而在大学一年级的力学中，它以静力学（statics）的向量形式再次出现，并配合力矩（moment）亦须平衡的条件。如欲深入了解，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

# STEP II 2021，第 10 题

力学 · 非惯性参考系

> **STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 11.5 节加速火车中的重力中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/accelerating-frame.html

## 关键思路

坐标轴随列车移动，所以它们并不构成惯性参考系（inertial frame）。最干净利落的处理方法，是求出珠子的真实加速度：即珠子相对于列车的加速度，加上列车本身的加速度。这样牛顿第二定律（Newton's second law）便成立；又因为钢线光滑（smooth），反作用力（reaction）垂直于钢线，所以它沿珠子的路径不做功。把运动方程与珠子的（相对）速度取点积，便能消去未知的反作用力，剩下一个类似能量、而且守恒的量。

## 解答

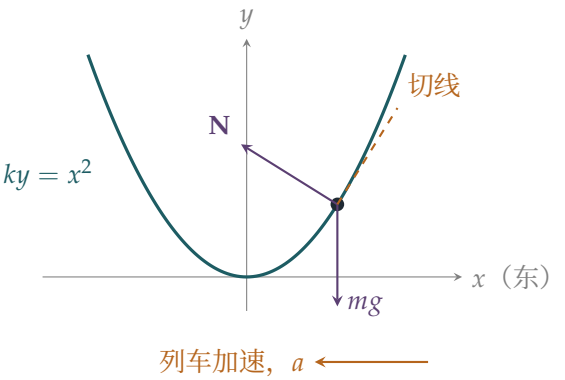
把珠子视为质量为 $m$ 的质点，并把钢线视为光滑，因此钢线对珠子的反作用力 $\mathbf{N}$ 恒垂直于钢线。设 $\mathbf{i}$ 和 $\mathbf{j}$ 分别为指向东方和向上的单位向量（unit vector）。

列车向西加速，所以它的加速度是 $-a\mathbf{i}$。珠子相对于列车的位置是 $x\mathbf{i}+y\mathbf{j}$，因此它的真实（地面参考系中的）加速度是

$$(\ddot{x}-a)\,\mathbf{i}+\ddot{y}\,\mathbf{j}.$$

作用于珠子的力只有重力 $-mg\mathbf{j}$ 和 $\mathbf{N}$，所以由牛顿第二定律得

$$m(\ddot{x}-a)\,\mathbf{i}+m\ddot{y}\,\mathbf{j}=\mathbf{N}-mg\,\mathbf{j}. \qquad (*)$$




**第 (i) 部分**

假设珠子相对于列车一直停留在原点。那么它便与列车有相同的加速度 $-a\mathbf{i}$，因此需要一个水平力 $-ma\mathbf{i}$，其中 $a > 0$。但在原点处，钢线 $y = x^2/k$ 的斜率为 $2x/k = 0$，即钢线在该处是水平的，而光滑钢线的反作用力垂直于钢线，也就是铅直方向。重力同样是铅直的。既然没有任何水平力作用，珠子便不可能随列车一起加速；它必定相对于列车移动。（珠子的真实水平加速度是 0，而列车的加速度是向西的 $a$，所以珠子开始沿钢线向东滑动，即朝 $x$ 的正方向移动。）

**第 (ii) 部分**

珠子沿钢线移动时，由于钢线固定在列车上，珠子相对于列车的速度 $\dot{x}\,\mathbf{i} + \dot{y}\,\mathbf{j}$ 与钢线相切。由于 $\mathbf{N}$ 垂直于钢线，

$$\mathbf{N} \cdot (\dot{x}\,\mathbf{i} + \dot{y}\,\mathbf{j}) = 0.$$

把 $(*)$ 改写为 $m(\ddot{x} - a)\,\mathbf{i} + m(\ddot{y} + g)\,\mathbf{j} = \mathbf{N}$，然后两边与 $\dot{x}\,\mathbf{i} + \dot{y}\,\mathbf{j}$ 取纯量积(scalar product)，再除以 $m$，得

$$\dot{x}(\ddot{x} - a) + \dot{y}(\ddot{y} + g) = 0,$$

此即所求。左边恰好是一个导数（exact derivative）：

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) - ax + gy\right) = \dot{x}\ddot{x} + \dot{y}\ddot{y} - a\dot{x} + g\dot{y} = 0,$$

所以在整个运动过程中，$\frac{1}{2}(\dot{x}^2 + \dot{y}^2) - ax + gy$ 是常数。

? **想一想** 既然反作用力在地面参考系中会做功，为甚么与相对速度取点积是合理的?

在地面参考系中，珠子的速度多了一项 $-at\,\mathbf{i}$，它并不沿着钢线方向，所以 $\mathbf{N}$ 在该参考系中会做功。但恒等式 $\mathbf{N} \cdot \mathbf{u} = 0$ 纯粹是几何事实：对任何与钢线相切的向量 $\mathbf{u}$ 都成立，而相对速度正是这样的向量。我们用到的就只有这一点。

### 第 (iii) 部分

开始时 $x = y = 0$ 且 $\dot{x} = \dot{y} = 0$，所以该常数为 0。设 $v$ 为相对于列车的速率，并利用 $y = x^2/k$，得

$$\tfrac{1}{2}v^2 = ax - gy = ax - \frac{gx^2}{k} = \frac{g}{k}\,x\Big(\frac{ak}{g} - x\Big). \qquad (\dagger)$$

由于 $v^2 \geq 0$，珠子被限制在 $0 \leq x \leq ak/g$ 之内，而且只有在这个区间的两个端点才有 $v = 0$。由第 (i) 部分，珠子开始时朝 $x$ 的正方向移动。当 $0 < x < ak/g$ 时，它的速率为正，所以它沿钢线一直朝同一方向移动，直至到达 $x = ak/g$，在该处瞬间静止。（珠子确实会在该处折返：$(*)$ 在该点的切向分量（tangential component）不为零，因为 $\frac{\mathrm{d}}{\mathrm{d}x}(ax - gx^2/k) = a - 2a = -a \neq 0$。）之后珠子在 $x = 0$ 与 $x = ak/g$ 之间来回振荡（oscillate）。

在这个区间上，$y = x^2/k$ 是 $x$ 的递增函数，所以最大高度在 $x = ak/g$ 处达到：

$$b = \frac{1}{k}\Big(\frac{ak}{g}\Big)^2 = \frac{a^2k}{g^2}.$$

### 第 (iv) 部分

由 $(\dagger)$，$v^2 = 2\big(ax - gx^2/k\big)$，这是 $x$ 的二次函数，其最大值出现在两根 0 与 $ak/g$ 的中点，所以相对于列车的速率在以下位置最大：

$$x = \frac{ak}{2g}.$$

在该处

$$v^2 = 2\Big(\frac{a^2k}{2g} - \frac{a^2k}{4g}\Big) = \frac{a^2k}{2g}, \qquad \text{所以} \qquad v_{\max} = a\sqrt{\frac{k}{2g}}.$$

## 验算

- 极限情况：若 $a = 0$，珠子只会停留在钢线的最低点；而的确 $b = 0$ 且 $v_{\max} = 0$。
- 有效重力（effective gravity）：在列车中，珠子感受到的有效重力是 $(a, -g)$，向东倾斜。速率最大的位置 $x = ak/(2g)$，恰好是钢线斜率 $2x/k$ 等于 $a/g$ 之处，因此该处的切线垂直于有效重力：这正是钢线在倾斜重力场中的“最低点”。折返点 $x = ak/g$ 与原点的距离是它的两倍，这与在一个关于 $x$ 为二次的势能中作对称振荡的情况一致。
- 量纲（dimensions）：$k$ 是长度，所以 $a^2k/g^2$ 是长度，而 $a\sqrt{k/g}$ 是速率。
- 数值验证：取 $a = 2$、$k = 1.5$ 及 $g = 9.8$（国际单位制），用电脑逐步推进运动，得到最大高度为 0.06247 m，最大速率为 0.5533 m s$^{-1}$，出现在 $x = 0.1531$ m，分别与 $a^2k/g^2$、$a\sqrt{k/(2g)}$ 和 $ak/(2g)$ 吻合。

## 常见错误

- 把 $\ddot{x}$ 当作珠子的真实水平加速度。$\ddot{x}$ 只是珠子相对于列车的加速度，必须再加上列车本身的加速度 $-a$。
- 弄错 $a$ 的正负号：列车向西加速，所以在列车中珠子被推向东，即朝 $x$ 的正方向。
- 以为速率在钢线的最低点 $x = 0$ 处最大。在列车中，有效的“向下”方向是倾斜的，所以速率最大的位置其实是 $x = ak/(2g)$。

## STEP 之外

第 (ii) 部分让你初尝非惯性参考系（non-inertial frame）中的力学：在列车的参考系中，$-ax$ 这一项就像一个向东、大小为 $ma$ 的均匀“虚拟力”（fictitious force）所对应的势能（potential energy），把它与重力合并，便得到倾斜的有效重力。到了大学，这套想法会推广至旋转参考系，届时会出现离心力（centrifugal force）和科里奥利力（Coriolis force）；亦会推广至拉格朗日力学（Lagrangian mechanics），在那里像 **N** 这样的约束力（constraint force）从一开始便被消去。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）有一章很好地讲解了非惯性系统和虚拟力。

详细解答

# STEP II 2021，第 11 题

概率与统计 · 条件概率的递推关系

**STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 12.5 节最后一位乘客的座位中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/airline-seat.html

## 关键思路

以 $T_1$ 坐在哪里为条件（condition on）。若 $T_1$ 坐了自己的座位 $S_1$，其他人都不受影响；若 $T_1$ 坐了 $S_k$，那么 $T_2, \ldots, T_{k-1}$ 都能安然入座，而 $T_k$ 就成了新的“随机选座者”。关键的观察是：此时 $T_k$ 的处境与当初的 $T_1$ 完全相同，只是列车变小了，而空着的座位 $S_1$ 则扮演 $T_k$ 自己座位的角色。这样问题便转化为一条递推关系（recurrence）。

## 解答

全文中，“$T_r$ 被挤走”指 $T_r$ 到达时发现 $S_r$ 已被人坐了。只有被挤走的乘客（或 $T_1$）才会随机选座。

### 第 (i) 部分

当 $n = 2$ 时，$T_1$ 各以概率（probability）$\frac{1}{2}$ 坐 $S_1$ 或 $S_2$，而 $T_2$ 能坐到 $S_2$ 恰好当 $T_1$ 坐了 $S_1$。所以

$$P_2 = \frac{1}{2}.$$

当 $n = 3$ 时，以 $T_1$ 的座位为条件，每种情况的概率均为 $\frac{1}{3}$：

- $S_1$：$T_2$ 和 $T_3$ 都发现自己的座位空着，所以 $T_3$ 坐到 $S_3$。
- $S_3$：$T_3$ 必定坐不到 $S_3$。

▶ $S_2$：$T_2$ 被挤走，在 $S_1$ 和 $S_3$ 之间随机选择；$T_3$ 坐到 $S_3$ 恰好当 $T_2$ 选了 $S_1$，概率为 $\frac{1}{2}$。

因此

$$P_3 = \tfrac{1}{3}\cdot 1 + \tfrac{1}{3}\cdot 0 + \tfrac{1}{3}\cdot\tfrac{1}{2} = \tfrac{1}{2}.$$

**第 (ii) 部分**

假设 $T_1$ 坐在 $S_k$，其中 $2 \le k \le n-1$。那么 $T_2, \ldots, T_{k-1}$ 都发现自己的座位空着并坐下。当 $T_k$ 到达时，空座位是

$$S_1,\ S_{k+1},\ S_{k+2},\ \ldots,\ S_n,$$

共 $n-k+1$ 个座位，而 $T_k$ 在其中均匀随机地选择。尚未到达的乘客是 $T_{k+1}, \ldots, T_n$，各人都想坐自己的座位，而尚未到达的人之中没有人被分配到 $S_1$。

? **想一想** 为甚么这恰好就是有 $n-k+1$ 个座位的原问题?

重新标记：把 $T_k$ 称为“第一位乘客”，并当作 $S_1$ 是 $T_k$ 获分配的座位。在有 $m$ 个座位的原问题中，第一位乘客从全部 $m$ 个座位中均匀地选择，之后每位乘客若自己的座位空着便坐下，否则从空座位中均匀地选择，而第一位乘客获分配的座位没有其他人想要。这里取 $m = n-k+1$，乘客 $T_k, T_{k+1}, \ldots, T_n$ 在座位 $S_1, S_{k+1}, \ldots, S_n$ 上正好遵循这些规则，而 $S_1$ 没有其他人想要。这些规则除了用来判断乘客自己的座位是否空着之外，从不用到座位的标签，所以两个过程的概率完全相同。在较小的问题中，$T_n$ 仍是最后一位乘客，所以

$$\mathrm{P}(T_n \text{ 坐在} S_n \mid T_1 \text{ 坐在} S_k) = P_{n-k+1}.$$

现在，对 $n \ge 3$，以 $T_1$ 的座位为条件，$S_1, \ldots, S_n$ 每个的概率均为 $\frac{1}{n}$。若是 $S_1$，$T_n$ 必定坐到 $S_n$；若是 $S_n$，$T_n$ 必定坐不到。所以由全概率公式（law of total probability），

$$P_n = \frac{1}{n}\cdot 1 + \frac{1}{n}\sum_{k=2}^{n-1} P_{n-k+1} + \frac{1}{n}\cdot 0 = \frac{1}{n}\left(1 + \sum_{r=2}^{n-1} P_r\right),$$

其中我们代入了 $r = n-k+1$，当 $k$ 由 2 变到 $n-1$ 时，它由 $n-1$ 递减至 2。这就是所求的结果。

**第 (iii) 部分**

我们断言对每个 $n \ge 2$ 都有 $P_n = \frac{1}{2}$。我们使用强归纳法（strong induction）。

基本情况。由第 (i) 部分得 $P_2 = \frac{1}{2}$。

归纳步骤。设 $n \ge 3$，并假设对 $r = 2, 3, \ldots, n-1$ 有 $P_r = \frac{1}{2}$。第 (ii) 部分中的和有 $n-2$ 项，每项都等于 $\frac{1}{2}$，所以

$$P_n = \frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{n}\cdot\frac{n}{2} = \frac{1}{2}.$$

因此对 $r = 2, \ldots, n$ 有 $P_r = \frac{1}{2}$，由归纳法，对所有 $n \geq 2$ 都有 $P_n = \frac{1}{2}$。（第 (i) 部分的 $P_3 = \frac{1}{2}$ 与此吻合。）

**第 (iv) 部分**

当 $n = 2$ 时，$T_{n-1} = T_1$，他以概率 $\frac{1}{2}$ 坐在 $S_1$。所以 $Q_2 = \frac{1}{2}$。

现在设 $n \geq 3$，于是 $T_{n-1}$ 不是 $T_1$，再次以 $T_1$ 的座位 $S_k$ 为条件：

- $k = 1$：每个人都坐到自己的座位，所以 $T_{n-1}$ 坐到 $S_{n-1}$。
- $k = n$：$T_2, \ldots, T_{n-1}$ 都发现自己的座位空着，所以 $T_{n-1}$ 同样坐到 $S_{n-1}$。
- $k = n-1$：$T_{n-1}$ 必定坐不到 $S_{n-1}$。
- $2 \leq k \leq n-2$：与第 (ii) 部分完全一样，余下的旅程就是以 $T_k$ 为首、有 $m = n-k+1$ 个座位的原问题。在这列较小的列车中，$T_{n-1}$ 是倒数第二位乘客，所以概率为 $Q_{n-k+1}$。这里 $m$ 取遍 $3, 4, \ldots, n-1$。

因此，对 $n \geq 3$，

$$Q_n = \frac{1}{n}\left(2 + \sum_{m=3}^{n-1} Q_m\right).$$

当 $n = 3$ 时和为空，故 $Q_3 = \frac{2}{3}$。若对 $3 \leq m \leq n-1$ 有 $Q_m = \frac{2}{3}$，则

$$Q_n = \frac{1}{n}\left(2 + \frac{2(n-3)}{3}\right) = \frac{1}{n} \cdot \frac{2n}{3} = \frac{2}{3},$$

所以由强归纳法，对每个 $n \geq 3$ 都有 $Q_n = \frac{2}{3}$。总结：

$$Q_2 = \tfrac{1}{2}, \qquad Q_n = \tfrac{2}{3} \quad (n \geq 3).$$

? **想一想** 有没有方法不用任何递推关系就看出 $\frac{1}{2}$ 和 $\frac{2}{3}$?

有的。每当有人随机选座时，$S_1$ 与每个后来乘客仍然空着的座位都受到同等对待，所以在过程结束之前，$S_1$ 和 $S_n$ 两者中先被坐的那个是哪一个，可能性相等；$T_n$ 成功恰好当 $S_1$ 先被坐，得 $\frac{1}{2}$。对于 $T_{n-1}$（其中 $n \geq 3$），考虑 $S_1, S_{n-1}, S_n$ 之中哪一个最先被随机选座者坐了；每个的可能性相等，而 $T_{n-1}$ 只有在那个是 $S_{n-1}$ 时才会被挤走，得 $1 - \frac{1}{3} = \frac{2}{3}$。

## 验算

- $n = 3$ 的情况可以用手列举：$(T_1, T_2, T_3)$ 的结果是 $(S_1, S_2, S_3)$，概率为 $\frac{1}{3}$；$(S_3, S_2, S_1)$，概率为 $\frac{1}{3}$；以及 $(S_2, S_1, S_3)$、$(S_2, S_3, S_1)$，概率各为 $\frac{1}{6}$。所以 $T_3$ 坐到 $S_3$ 的概率为 $\frac{1}{2}$，$T_2$ 坐到 $S_2$ 的概率为 $\frac{2}{3}$，与所求得的一致。

- 用电脑对 $n = 2, \ldots, 8$ 精确列举每一种可能的旅程，得出 $P_n = \frac{1}{2}$ 全部成立、对 $n \geq 3$ 有 $Q_n = \frac{2}{3}$，并证实第 (ii) 部分的条件概率（conditional probability）对每个 $k$ 都成立。以 $n = 20$ 模拟 $10^5$ 次旅程，得到 0.500 和 0.668。

## 常见错误

- 忘记 $T_1$ 本身可以选 $S_1$，这正是递推关系中“$1+$”的来源。
- 在第 (iv) 部分漏掉 $T_1$ 坐了 $S_n$ 的情况：这对 $T_{n-1}$ 是好消息，所以常数是 2 而不是 1。
- 在 $n = 2$ 时套用 $Q$ 的递推关系，但此时 $T_{n-1}$ 就是 $T_1$，答案是 $\frac{1}{2}$ 而不是 $\frac{2}{3}$。

## STEP 之外

这里的递推关系是一种首步分析（first-step analysis）：以第一次随机选择为条件，并认出同一问题的较小版本。同样的技巧可以求出马尔可夫链（Markov chain）的命中概率（hitting probability）和期望命中时间（expected hitting time），这是大学一年级概率论的核心课题。上面的对称性论证让人初尝一个选得好的“哪件事先发生?”问题如何取代冗长的计算。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）对这两种思路都有阐述。

详细解答

# STEP II 2021，第 12 题

概率与统计 · 赌徒破产问题

> **STEP II 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

每场比赛都会不断回到之前出现过的局面。因此，我们不去把无穷多种获胜方式逐一加起来，而是设 $P$ 为从该局面出发获胜的概率（probability），再按下一步发生的事情作条件分析（conditioning），得出一条 $P$ 同时出现在两边的方程。同样的技巧用在谨慎玩法上，就能把它化为第 (ii) 部分的比赛。

## 解答

### 第 (i) 部分

设 $P$ 为 A 赢得整场比赛的概率。按第一局的结果作条件分析。A 以概率 $p_{\mathrm{A}}$ 赢得该局，比赛随即结束。B 以概率 $p_{\mathrm{B}}$ 赢得该局，A 便输了。否则该局打和，概率为 $1-p_{\mathrm{A}}-p_{\mathrm{B}}$；由于各局相互独立（independent），比赛重新开始，所以 A 从该处获胜的机会仍然是 $P$。因此

$$P = p_{\mathrm{A}} + (1-p_{\mathrm{A}}-p_{\mathrm{B}})P \quad\Longrightarrow\quad (p_{\mathrm{A}}+p_{\mathrm{B}})P = p_{\mathrm{A}} \quad\Longrightarrow\quad P = \frac{p_{\mathrm{A}}}{p_{\mathrm{A}}+p_{\mathrm{B}}},$$

这里两边除以 $p_{\mathrm{A}}+p_{\mathrm{B}} > 0$。（等价地，A 在第 $k$ 局获胜的概率为 $(1-p_{\mathrm{A}}-p_{\mathrm{B}})^{k-1}p_{\mathrm{A}}$，而由于 $0 < p_{\mathrm{A}}+p_{\mathrm{B}} < 1$，这个几何级数（geometric series）收敛。）

### 第 (ii) 部分

设 $D$ 为进行若干局后 A 的胜局数减去 B 的胜局数。每一局使 $D$ 改变 $+1$ 或 $-1$，而比赛在首次出现 $|D| = 2$ 时结束。

*长度为偶数。*进行 $n$ 局后，若 A 胜 $a$ 局、B 胜 $b$ 局，则 $a+b=n$ 且 $D = a-b = n-2b$，

所以 $D$ 与 $n$ 的奇偶性（parity）相同。比赛在 $D = \pm 2$ 时结束，而这是偶数，所以 $n$ 是偶数。

概率。把各局两两成对来看。比赛开始时双方打平（$D = 0$），而由于比赛只能在偶数局后结束，每一对未能结束比赛的局之后，双方又会再次打平。从打平的局面出发，接下来两局是 AA 的概率为 $p^2$（A 胜），是 BB 的概率为 $q^2$（B 胜），是 AB 或 BA 的总概率为 $2pq$，此时比赛再次打平。设 $P$ 为 A 从打平局面出发获胜的概率，则

$$P = p^2 + 2pq\,P \quad \Longrightarrow \quad P = \frac{p^2}{1 - 2pq}.$$

由于 $p + q = 1$，有 $1 = (p+q)^2 = p^2 + 2pq + q^2$，所以 $1 - 2pq = p^2 + q^2$，且

$$P = \frac{p^2}{p^2 + q^2},$$

即为所求。这里 $p^2 + q^2 > 0$，所以除法没有问题。

### 第 (iii) 部分

谨慎玩法。筹码数目每轮增加或减少一个，由 1 开始，到 0（输）或 4（赢）时停止。它不能跳过 4，所以玩家获胜当且仅当筹码数目先到达 4 而不是先到达 0。

> **想一想** 从 2 个筹码出发，玩家需要做到甚么？我们以前见过这种情况吗？

从 2 个筹码出发，玩家要在输的轮数比赢的轮数多出两轮（到达 0）之前，先做到赢的轮数比输的轮数多出两轮（到达 4），才能获胜。这正是第 (ii) 部分的比赛，其中玩家是 A，碰运气的游戏是 B。所以从 2 个筹码出发，获胜的概率为 $\dfrac{p^2}{p^2 + q^2}$。

从 1 个筹码出发，第一轮必须赢，否则玩家跌至 0 而输掉。所以第一轮以概率 $p$ 使玩家到达 2 个筹码，而各轮相互独立，因此

$$P_{\text{cautious}} = p \cdot \frac{p^2}{p^2 + q^2} = \frac{p^3}{p^2 + q^2}.$$

（如果你偏好用方程：设 $P_k$ 为从 $k$ 个筹码出发获胜的概率，则 $P_1 = pP_2$、$P_2 = pP_3 + qP_1$ 及 $P_3 = p + qP_2$ 会给出相同的答案。）

大胆玩法。从 1 个筹码出发，数目变为 2 或 0；从 2 个出发，数目变为 4 或 0。所以玩家只有赢得头两轮才能获胜：

$$P_{\text{bold}} = p^2.$$

比较。两个概率都是正数，所以我们把它们的比与 1 作比较：

$$\frac{P_{\text{cautious}}}{P_{\text{bold}}} = \frac{p}{p^2 + q^2}.$$

由于 $p^2+q^2>0$，这个比大于 1 当且仅当

$$p>p^2+q^2 \iff p-p^2>q^2 \iff pq>q^2 \iff p>q,$$

最后一步除以 $q>0$（因为 $p<1$，所以成立）。而 $p>q=1-p$ 恰好在 $p>\frac{1}{2}$ 时成立。同理，这个比小于 1 恰好在 $p<q$，即 $p<\frac{1}{2}$ 时成立。所以当 $\frac{1}{2}<p<1$ 时谨慎玩法较好，当 $0<p<\frac{1}{2}$ 时大胆玩法较好。（在 $p=\frac{1}{2}$ 时两者都给出 $\frac{1}{4}$。）

## 验算

- 在 $p=\frac{1}{2}$ 时，谨慎玩法是一个公平的随机漫步（random walk），从 1 出发先到达 4 而不是先到达 0 的机会是 $\frac{1}{4}$。我们的公式给出 $\frac{1}{8}/\frac{1}{2}=\frac{1}{4}$，与大胆玩法的 $p^2=\frac{1}{4}$ 一致。
- 在 $p=0.7$ 时：谨慎玩法给出 $0.343/0.58\approx 0.591$，大胆玩法给出 0.49。在 $p=0.3$ 时：谨慎玩法给出 $0.027/0.58\approx 0.0466$，大胆玩法给出 0.09。两者都与逐轮追踪筹码分布的电脑验算相符。
- 在第 (ii) 部分，$p=\frac{1}{2}$ 给出 $\frac{1}{2}$，正如对称性所要求的。

## 常见错误

- 在第 (ii) 部分，把“AB”和“BA”当作同一个概率为 $pq$ 的结果，而不是 $2pq$。
- 忘记说明为何一对未结束比赛的局总会使比赛回到打平状态。理由是比赛不可能在奇数局后结束。
- 在比较时，没有说明有关的量是正数便展开相乘或相除。

## STEP 之外

谨慎玩法正是*赌徒破产*（gambler's ruin）问题。它是一个带有吸收壁（absorbing barriers）的随机漫步；当 $p\neq q$ 时，从 $k$ 出发、目标为 $N$ 的一般答案是 $\dfrac{1-(q/p)^k}{1-(q/p)^N}$。在不利的游戏中大胆玩法最好、在有利的游戏中保守玩法最好，这个结果可追溯至 Dubins 和 Savage 对最优赌博策略的研究。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）正是用这种“按第一步作条件分析”的方法来处理随机漫步和赌徒破产问题。

详细解答

# STEP III 2021，第 1 题

纯数学 · 参数曲线的法线

> **STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：
>
> www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

两部分都依循同一套步骤：先对参数方程（parametric equations）求导，再取 $\mathrm{d}y/\mathrm{d}x$ 的负倒数（negative reciprocal），得出法线（normal）的斜率（gradient），然后整理方程。整理的过程正是最有趣的地方。把整条方程乘开以消去分数，三角函数项便会化简成一条简短而对称的直线方程。一旦直线写成 $x\sin\phi + y\cos\phi = \ldots$ 或 $x\cos\phi + y\sin\phi = \ldots$ 的形式，它的几何性质（截距（intercepts）、与原点的距离）便一眼可见。

## 解答

### 第 (i) 部分

对 $t$ 求导，

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 12\cos^2 t\,\sin t, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = 12\cos t - 12\sin^2 t\cos t = 12\cos t\,(1-\sin^2 t) = 12\cos^3 t.$$

因此，当 $0 < t < \frac{1}{2}\pi$（此时 $\sin t$ 和 $\cos t$ 均不为零），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{12\cos^3 t}{12\cos^2 t\sin t} = \frac{\cos t}{\sin t}.$$

在 $t = \phi$ 处，切线（tangent）的斜率为 $\cot\phi$，所以法线的斜率为 $-\tan\phi = -\dfrac{\sin\phi}{\cos\phi}$。因此法线为

$$y - 12\sin\phi + 4\sin^3\phi = -\frac{\sin\phi}{\cos\phi}\left(x + 4\cos^3\phi\right).$$

乘以 $\cos\phi$，并把 $x$ 和 $y$ 项移到左方：

$$\begin{aligned} x\sin\phi + y\cos\phi &= 12\sin\phi\cos\phi - 4\sin^3\phi\cos\phi - 4\cos^3\phi\sin\phi \\ &= 12\sin\phi\cos\phi - 4\sin\phi\cos\phi\,(\sin^2\phi + \cos^2\phi) \\ &= 8\sin\phi\cos\phi. \end{aligned}$$

所以法线为

$$\boxed{x\sin\phi + y\cos\phi = 8\sin\phi\cos\phi} \qquad \text{或等价地} \qquad \frac{x}{\cos\phi} + \frac{y}{\sin\phi} = 8.$$

第二种形式显示法线与两轴相交于 $(8\cos\phi, 0)$ 和 $(0, 8\sin\phi)$。

**相切**。设 $P = (8\cos^3\phi,\ 8\sin^3\phi)$。我们需要证明三件事。

*$P$ 在曲线 $x^{2/3} + y^{2/3} = 4$ 上。*由于 $0 < \phi < \frac{1}{2}\pi$，两个坐标都是正数，所以其实数立方根（cube root）分别是 $2\cos\phi$ 和 $2\sin\phi$。因此 $x^{2/3} + y^{2/3} = 4\cos^2\phi + 4\sin^2\phi = 4$。

*$P$ 在法线上。*代入可得

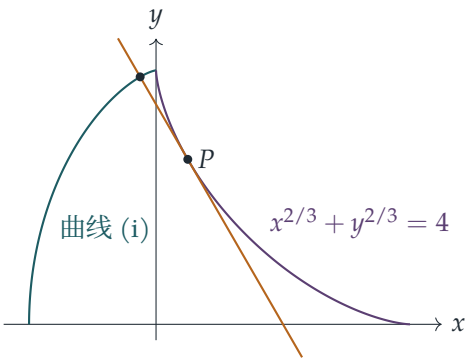
$$8\cos^3\phi\sin\phi + 8\sin^3\phi\cos\phi = 8\sin\phi\cos\phi\,(\cos^2\phi + \sin^2\phi) = 8\sin\phi\cos\phi.$$

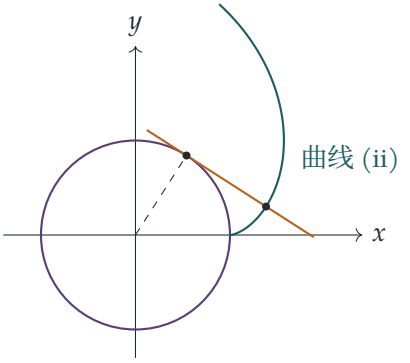
*两者在 $P$ 处的斜率相同。*对 $x^{2/3} + y^{2/3} = 4$ 进行隐函数微分（implicit differentiation）（当 $x, y > 0$ 时成立），

$$\tfrac{2}{3}x^{-1/3} + \tfrac{2}{3}y^{-1/3}\frac{\mathrm{d}y}{\mathrm{d}x} = 0 \quad\Longrightarrow\quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{y^{1/3}}{x^{1/3}} = -\frac{2\sin\phi}{2\cos\phi} = -\tan\phi \ \text{ 在}P,$$

这正好是法线的斜率。

一条穿过曲线上某点、且斜率与曲线在该点的斜率相同的直线，就是曲线在该点的切线。所以法线是 $x^{2/3} + y^{2/3} = 4$ 在 $P$ 处的切线，得证。




两幅图中，曲线（青色）上某点的法线（琥珀色）都与第二条曲线（紫红色）相切；两幅图都是就某个特定的 $\phi$ 值绘制的。

**第 (ii) 部分**

对 $t$ 求导（对 $t\sin t$ 和 $t\cos t$ 项使用乘积法则（product rule）），

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\sin t + \sin t + t\cos t = t\cos t, \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = \cos t - \cos t + t\sin t = t\sin t.$$

当 $0 < t < \frac{1}{2}\pi$ 时，$t\cos t \neq 0$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{t\sin t}{t\cos t} = \tan t$。在 $t = \phi$ 处，法线的斜率为 $-\cot\phi = -\dfrac{\cos\phi}{\sin\phi}$，其方程为

$$y - \sin\phi + \phi\cos\phi = -\frac{\cos\phi}{\sin\phi}\big(x - \cos\phi - \phi\sin\phi\big).$$

乘以 $\sin\phi$，

$$y\sin\phi - \sin^2\phi + \phi\sin\phi\cos\phi = -x\cos\phi + \cos^2\phi + \phi\sin\phi\cos\phi.$$

两边的 $\phi\sin\phi\cos\phi$ 项互相抵消，再利用 $\sin^2\phi + \cos^2\phi = 1$，剩下

$$\boxed{x\cos\phi + y\sin\phi = 1}.$$

**想一想** 怎样可以不求垂足（foot of the perpendicular）而求出原点到一条直线的距离？

点 $(x_0, y_0)$ 到直线 $ax + by = c$ 的距离是 $\dfrac{|ax_0 + by_0 - c|}{\sqrt{a^2 + b^2}}$。这里 $(x_0, y_0) = (0,0)$、$a = \cos\phi$、$b = \sin\phi$ 及 $c = 1$，所以距离为

$$\frac{|-1|}{\sqrt{\cos^2\phi + \sin^2\phi}} = 1.$$

你也可以直接看出来：$(\cos\phi, \sin\phi)$ 是垂直于该直线的单位向量（unit vector），而点 $(\cos\phi, \sin\phi)$ 本身亦满足该方程。所以从 $O$ 出发的垂足是 $(\cos\phi, \sin\phi)$，距离为 1。

**该曲线**。无论 $\phi$ 取何值，每条法线与原点的距离都是 1。一条与 $O$ 相距 1 的直线，只会在垂足处与以 $O$ 为圆心、半径为 1 的圆相交，而且在该处与半径垂直。因此它是该圆的切线。所以每条法线都是

$$\boxed{x^2 + y^2 = 1},$$

的切线，切点为 $(\cos\phi, \sin\phi)$。

## 验算

- 在 (i) 中取 $\phi = \frac{1}{3}\pi$。曲线上的点为 $(-\frac{1}{2},\ 6\sqrt{3} - \frac{3}{2}\sqrt{3}) = (-0.5,\ 7.794)$，而法线 $2x + \frac{2}{\sqrt{3}}y = 8$ 与两轴相交于 $(4,0)$ 和 $(0, 4\sqrt{3}) \approx (0, 6.928)$。点 $P = (1,\ 3\sqrt{3}) \approx (1,\ 5.196)$ 满足 $2 + 6 = 8$，所以它在法线上，而且 $1^{2/3} + (3\sqrt{3})^{2/3} = 1 + 3 = 4$。

- 在 (ii) 中，当 $\phi \to 0$ 时，曲线从单位圆上的 $(1,0)$ 开始，而该处的法线趋向铅垂线 $x = 1$，这确实是 $x^2 + y^2 = 1$ 的切线。
- 用电脑对一千个随机的 $\phi$ 值进行验算，在舍入误差范围内证实了所有法线方程、(i) 中的相切关系，以及 (ii) 中的单位距离。

## 常见错误

- 在 (i) 中把 $\mathrm{d}y/\mathrm{d}t$ 算错：$12\cos t - 12\sin^2 t \cos t$ 可化简为 $12\cos^3 t$，正是这个因式分解（factorisation）令斜率变得如此简单。
- 在 (i) 中只验证 $P$ 在法线上。一条穿过曲线上某点的直线不一定是该点的切线；斜率也必须相同。
- 在 (ii) 中只答“一个圆”，却没有给出它的方程、圆心和半径。

## STEP 之外

一条曲线所有法线的包络线（envelope）称为该曲线的*渐屈线*（evolute）。它是曲率中心（centres of curvature）的轨迹（locus）。因此第 (i) 部分显示星形线（astroid）$x^{2/3} + y^{2/3} = 4$ 是第一条曲线的渐屈线。在 (ii) 中，该曲线是单位圆的*渐开线*（involute）（即从线辘上拉紧解开的绳子末端所描出的路径），而它的渐屈线就是该圆。圆的渐开线是大多数齿轮齿形的轮廓。这些概念属于曲线的微分几何（differential geometry）。Needham 的 *Visual Complex Analysis*（《直观复分析》）以优美的几何方式探讨这类平面曲线的推理，是很好的参考读物。

详细解答

# STEP III 2021，第 2 题

纯数学 · 行列式与不等式

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

将 $x$、$y$、$z$ 各自的定义去分母后，都会得到一条关于 $a$、$b$、$c$ 的线性（linear）方程。三条这样的方程组成一个齐次线性方程组（homogeneous linear system），而它有非零解 $(a,b,c)$，所以系数矩阵（matrix of coefficients）必定是奇异（singular）的，即其行列式（determinant）为零。这个行列式正好就是我们想要的、关于 $x$、$y$、$z$ 的对称（symmetric）表达式。之后，只要把目标式改写成 $x$、$y$、$z$ 的式子，再运用平方和的基本性质和算术－几何平均不等式（AM–GM），就能得出那些不等式。

## 解答

### 第 (i) 部分

由于 $a$、$b$、$c$ 互不相同，所有分母都不为零。去分母，得

$$x(b-c)=a, \qquad y(c-a)=b, \qquad z(a-b)=c.$$

现在逐行展开矩阵乘积：

$$\begin{aligned} a-xb+xc &= a-x(b-c)=a-a=0,\\ ya+b-yc &= b-y(c-a)=b-b=0,\\ -za+zb+c &= c-z(a-b)=c-c=0, \end{aligned}$$

这正是所求的矩阵方程。

把这个 $3\times 3$ 矩阵记作 $M$。由于 $a$、$b$、$c$ 互不相同，它们不会全为零，所以 $M$ 把一个非零向量（vector）映射到零向量。

**想一想** 为甚么这样就迫使 $\det M = 0$?

若 $\det M \neq 0$, 则 $M^{-1}$ 存在; 在 $M\mathbf{v} = \mathbf{0}$ 两边左乘 $M^{-1}$, 便得 $\mathbf{v} = \mathbf{0}$, 与 $(a,b,c) \neq (0,0,0)$ 矛盾。所以 $\det M = 0$。沿第一行展开,

$$\begin{aligned}\det M &= 1\cdot(1+yz) + x(y-yz) + x(yz+z)\\ &= 1 + yz + xy - xyz + xyz + xz\\ &= 1 + xy + yz + zx.\end{aligned}$$

令它等于零, 即得 $yz + zx + xy = -1$。

至于不等式, 左边是 $x^2+y^2+z^2$。利用恒等式 (identity) $(x+y+z)^2 = x^2+y^2+z^2+2(xy+yz+zx)$,

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

因为实数 $x+y+z$ 的平方是非负的。这就是所求的结果。

**第 (ii) 部分**

这里的分母都是正数, 去分母后得 $x(b+c) = 2a$, 其余类似。重新整理成关于 $a$、$b$、$c$ 的线性方程:

$$\begin{pmatrix} -2 & x & x \\ y & -2 & y \\ z & z & -2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

(第一行即 $-2a + x(b+c) = 0$, 其余两行类似。) 由于 $a,b,c > 0$, 这个向量非零, 所以与第 (i) 部分完全一样, 行列式为零:

$$\begin{aligned}0 &= -2(4-yz) - x(-2y-yz) + x(yz+2z)\\ &= -8 + 2yz + 2xy + xyz + xyz + 2xz\\ &= -8 + 2(xy+yz+zx) + 2xyz.\end{aligned}$$

两边除以 2, 得 $xyz + yz + zx + xy = 4$。

**第一个不等式**。窍门是把左边每个括号除以右边对应的括号:

$$\frac{2a+b+c}{b+c} = \frac{2a}{b+c} + 1 = 1 + x,$$

另外两个也类似。由于 $(b+c)(c+a)(a+b) > 0$, 所求不等式等价于 $(1+x)(1+y)(1+z) > 5$。展开并利用刚证得的结果,

$$(1+x)(1+y)(1+z) = 1 + (x+y+z) + (xy+yz+zx) + xyz = 5 + (x+y+z).$$

由于 $x$、$y$、$z$ 都是正数，上式大于 5，第一个不等式得证。

**第二个不等式**。按同样的计算，现在需要证明 $x+y+z>2$。记 $S=x+y+z$。

> ? **想一想** 我们知道 $xy+yz+zx+xyz=4$。这怎样能迫使 $S$ 不会太小?

$xy+yz+zx$ 和 $xyz$ 都可以用 $S$ 给出上界：

- $S^2-3(xy+yz+zx)=\frac{1}{2}\left[(x-y)^2+(y-z)^2+(z-x)^2\right]\geq 0$，所以 $xy+yz+zx\leq \frac{1}{3}S^2$；
- 对正数 $x$、$y$、$z$ 应用算术—几何平均不等式，得 $xyz\leq\left(\frac{1}{3}S\right)^3=\frac{1}{27}S^3$。

两式相加，

$$4=xy+yz+zx+xyz\leq\frac{1}{3}S^2+\frac{1}{27}S^3.$$

当 $S>0$ 时，右边是 $S$ 的递增函数 (increasing function)，而当 $S=3$ 时它等于 $3+1=4$。所以若 $S<3$，右边便会小于 4，这是不可能的。因此 $S\geq 3$，而

$$\frac{(2a+b+c)(a+2b+c)(a+b+2c)}{(b+c)(c+a)(a+b)}=5+S\geq 8>7,$$

第二个不等式得证（事实上更得出较强的下界 8，当 $a=b=c$ 时取得）。

## 验算

- 在第 (ii) 部分取 $a=b=c=1$：左边为 $4^3=64$，而 $(b+c)(c+a)(a+b)=8$，所以比值为 8，与 $x=y=z=1$、$S=3$ 时的 $5+S$ 吻合。此外 $xyz+xy+yz+zx=1+3=4$。
- 在第 (i) 部分取 $(a,b,c)=(0,1,2)$：则 $x=0$、$y=\frac{1}{2}$、$z=-2$，而 $xy+yz+zx=-1$，符合要求；平方和为 $\frac{17}{4}\geq 2$。
- 用电脑对 20 000 组随机有理数三元组验算，两条恒等式都准确成立；第 (ii) 部分找到的最小比值约为 8.0001，而当 $a=b=1$、$c=10^{-6}$ 时比值约为 9.0，所以下界 8 确实是最小值，远高于 7。

## 常见错误

- 断言 $\det M=0$ 却不说明理由：关键在于 $(a,b,c)$ 是非零向量，这在第 (i) 部分需要“互不相同”，在第 (ii) 部分需要“正数”。
- 把 $2a+b+c$ 除以 $b+c$ 时得出 $1+\frac{1}{2}x$ 而不是 $1+x$：记住 $x$ 本身已经包含因数 2。
- 用 $xy+yz+zx<4$（一个上界）去尝试给 $S$ 找下界；不等式的方向必须反过来，如上文所示。

## STEP 之外

方阵把某个非零向量映射到零，当且仅当其行列式为零——这是线性代数（linear algebra）的核心：特征值（eigenvalues）正是通过 $\det(A-\lambda I)=0$ 求得的。第 (ii) 部分中的下界 $S \geq 3$ 其实就是改头换面的 Nesbitt 不等式（Nesbitt's inequality）$\frac{a}{b+c}+\frac{b}{c+a}+\frac{c}{a+b} \geq \frac{3}{2}$。线性代数方面，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）；若想多做这类不等式，Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题集》）是极佳的练习材料。

详细解答

# STEP III 2021，第 3 题

纯数学 · 积分不等式

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

对被升至 $n$ 次幂的那个式子求导。对于 $u = \sec x + \tan x$，其导数（derivative）是 $u \sec x$，这使 $u^{n-1}\frac{\mathrm{d}u}{\mathrm{d}x}$ 可以直接积分，所以要做的就是把 $I_{n+1} + I_{n-1}$ 的被积函数（integrand）改写成这个形式。至于不等式（inequality），则来自比较被积函数 $u^n = u^{n-1} \cdot u$ 与可直接积分的 $u^{n-1} \cdot u'$：若 $u < u'$，则积分（integral）小于 $\left[u^n/n\right]$。

## 解答

### 第 (i) 部分

设 $u = \sec x + \tan x$。则

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \sec x \tan x + \sec^2 x = \sec x\left(\sec x + \tan x\right) = u \sec x.$$

此外，利用 $\tan^2 x + 1 = \sec^2 x$，

$$u^2 + 1 = \sec^2 x + 2\sec x \tan x + \tan^2 x + 1 = 2\sec^2 x + 2\sec x \tan x = 2u \sec x = 2\frac{\mathrm{d}u}{\mathrm{d}x}.$$

因此对于 $n \geq 1$，

$$\begin{aligned} I_{n+1} + I_{n-1} &= \int_0^\beta u^{n-1}\left(u^2+1\right)\mathrm{d}x = 2\int_0^\beta u^{n-1}\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x = \frac{2}{n}\left[u^n\right]_0^\beta \\ &= \frac{2}{n}\left((\sec\beta + \tan\beta)^n - 1\right), \end{aligned}$$

因为在 $x = 0$ 时 $u = 1$。两边除以二，即得所求的恒等式（identity）。

? **想一想** 恒等式涉及 $I_{n+1}$ 和 $I_{n-1}$，但不等式却是关于 $I_n$ 的。怎样才能把中间那一项夹进去?

比较 $u^2+1$ 与 $2u$。我们有 $u^2+1-2u=(u-1)^2\geq 0$，而当 $0<x\leq\beta$ 时，$\sec x>1$ 且 $\tan x>0$，所以 $u>1$，不等式在该处严格成立。乘以 $u^{n-1}>0$，得在 $(0,\beta]$ 上 $u^{n+1}+u^{n-1}>2u^n$；积分后（被积函数连续，且不等式在一个长度为正的区间上严格成立）得

$$I_{n+1}+I_{n-1}>2I_n,\qquad \text{所以}\qquad I_n<\tfrac{1}{2}\left(I_{n+1}+I_{n-1}\right)=\frac{1}{n}\left((\sec\beta+\tan\beta)^n-1\right).$$

（等价地说：当 $x>0$ 时 $u'-u=u(\sec x-1)>0$，所以 $I_n=\int_0^\beta u^{n-1}u\,\mathrm{d}x<\int_0^\beta u^{n-1}u'\,\mathrm{d}x$。正是这个观点可以延用到第 (ii) 部分。）

**第 (ii) 部分**

记 $c=\cos\beta$，则 $0<c<1$，并设 $v=c\sec x+\tan x$。则

$$v'=\frac{\mathrm{d}v}{\mathrm{d}x}=c\sec x\tan x+\sec^2 x.$$

在端点处，$v(0)=c=\cos\beta$，而 $v(\beta)=1+\tan\beta$，正好是目标式中的两个数。因此上界就是 $\frac{1}{n}\left[v^n\right]_0^\beta=\int_0^\beta v^{n-1}v'\,\mathrm{d}x$，只需证明在整个 $[0,\beta]$ 上 $v<v'$ 便足够。

? **想一想** 在第 (i) 部分中，$u^2+1$ 恰好等于 $2u'$。对于 $v$，相应的关系是甚么?

重复第 (i) 部分的计算，

$$v^2+1=c^2\sec^2 x+2c\sec x\tan x+\tan^2 x+1=(1+c^2)\sec^2 x+2c\sec x\tan x,$$

所以

$$2v'-(v^2+1)=(1-c^2)\sec^2 x=\sin^2\beta\,\sec^2 x.$$

这不再是等式，但剩余项是正的。结合 $v^2+1-2v=(v-1)^2$，

$$2(v'-v)=(v-1)^2+\sin^2\beta\,\sec^2 x>0,$$

因为当 $0<\beta<\frac{\pi}{2}$ 时 $\sin\beta\neq 0$。故对每个 $x\in[0,\beta]$，$v<v'$。

最后，在 $[0,\beta]$ 上 $v>0$（因为 $c>0$、$\sec x>0$ 且 $\tan x\geq 0$），所以 $v^{n-1}>0$，从而 $v^n=v^{n-1}v<v^{n-1}v'$。在 $[0,\beta]$ 上积分，对于 $n\geq 1$，

$$J_n=\int_0^\beta v^n\,\mathrm{d}x<\int_0^\beta v^{n-1}\frac{\mathrm{d}v}{\mathrm{d}x}\,\mathrm{d}x=\frac{1}{n}\left[v^n\right]_0^\beta=\frac{1}{n}\left((1+\tan\beta)^n-\cos^n\beta\right),$$

得证。

## 验算

- 当 $n = 1$ 时，(i) 中的恒等式说 $\frac{1}{2}(I_2 + \beta) = \sec\beta + \tan\beta - 1$。取 $\beta = 1$，数值积分（numerical integration）得两边均为 2.40822…。
- 取 $\beta = 1$、$n = 3$：数值上 $I_3 \approx 8.774$，而上界为 12.863；$J_3 \approx 3.326$，而上界为 5.523。
- 在 (ii) 中形式上代入 $\cos\beta = 1$，$J_n$ 就变成 $I_n$，剩余项 $\sin^2\beta\sec^2 x$ 亦随之消失：第 (ii) 部分确实是第 (i) 部分的一个扰动（perturbation）。

## 常见错误

- 忘记检查正负：把 $v < v'$ 乘以 $v^{n-1}$ 需要 $v^{n-1} > 0$，这正是 $v > 0$ 之所以重要的原因。
- 只得出 $\leq$：严格不等式需要 $u > 1$（或 $\sin\beta \neq 0$）在一个区间上成立，而不仅仅在一点上成立。
- 计算 $v(\beta)$ 时粗心：$\sec\beta\cos\beta = 1$，$1 + \tan\beta$ 正是由此而来。

## STEP 之外

“若逐点有 $0 < f \leq g$，且在某处严格成立，则 $\int f < \int g$”这个论证，就是积分的单调性（monotonicity），分析课程会从黎曼积分（Riemann integral）的定义出发把它仔细证明。把一个难处理的被积函数与一个精确的导数作比较，是估计积分和级数（series）上下界的标准方法，它亦会在积分判别法（integral test）等估计中再次出现。Abbott 的 *Understanding Analysis*（《理解分析》）严谨地讲述了积分以及这类比较论证。

详细解答

# STEP III 2021，第 4 题

纯数学 · 向量与投影

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

这里的一切都靠纯量积（scalar product）完成：两个向量之间的夹角可从 $\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}||\mathbf{v}|\cos(\text{夹角})$ 读出；而投影（projection）是线性（linear）的，所以 $\frac{1}{2}(\mathbf{a}+\mathbf{b})$ 的投影就是 $\frac{1}{2}(\mathbf{a}_1+\mathbf{b}_1)$。第 (i) 部分证明两个长度相等的向量的平均会平分（bisect）它们之间的夹角。在第 (iii) 部分，投影 $\mathbf{a}_1$ 与 $\mathbf{b}_1$ 的长度通常不同，所以除非出现某种特殊情况，否则我们应预期平分不成立。条件可以漂亮地因式分解（factorise），正好对应两种特殊情况。

## 解答

### 第 (i) 部分

由于 $|\mathbf{a}| = |\mathbf{b}| = 1$ 且 $\mathbf{a}\cdot\mathbf{b} = \cos\theta$，

$$|\mathbf{m}|^2 = \tfrac{1}{4}\left(1 + 2\cos\theta + 1\right) = \tfrac{1}{2}(1+\cos\theta) = \cos^2\tfrac{\theta}{2},$$

所以 $|\mathbf{m}| = \cos\frac{\theta}{2}$，而因为 $0 < \theta < \pi$，它是正数。特别地，$\mathbf{m} \neq \mathbf{0}$。此外

$$\mathbf{m}\cdot\mathbf{a} = \tfrac{1}{2}(1+\cos\theta) = \cos^2\tfrac{\theta}{2} = \mathbf{m}\cdot\mathbf{b}.$$

因此 $\mathbf{m}$ 与 $\mathbf{a}$ 之间夹角的余弦是 $\cos^2\frac{\theta}{2}\big/\cos\frac{\theta}{2} = \cos\frac{\theta}{2}$，对 $\mathbf{b}$ 亦然。所以两个夹角都是 $\frac{\theta}{2}$。由于 $\mathbf{m}$ 位于 $\mathbf{a}$ 与 $\mathbf{b}$ 所在的平面上，而两个夹角之和为 $\theta$，故 $\mathbf{m}$ 平分 $\mathbf{a}$ 与 $\mathbf{b}$ 之间的夹角。

### 第 (ii) 部分

按定义，以 $\mathbf{c}$ 为单位法向量（unit normal），且 $\mathbf{a}\cdot\mathbf{c} = \cos\alpha$，

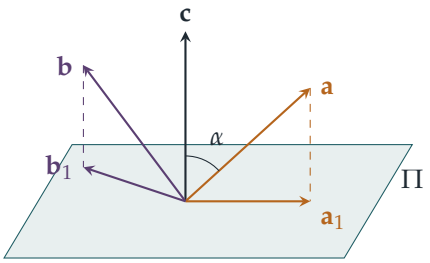
$$\mathbf{a}_1 = \mathbf{a} - \cos\alpha\,\mathbf{c}, \qquad \mathbf{b}_1 = \mathbf{b} - \cos\beta\,\mathbf{c}.$$

于是 $\mathbf{a}_1 \cdot \mathbf{c} = \cos\alpha - \cos\alpha\,(\mathbf{c}\cdot\mathbf{c}) = 0$，因为 $|\mathbf{c}| = 1$。同理 $\mathbf{b}_1 \cdot \mathbf{c} = 0$。

由于 $\mathbf{a}_1 \cdot \mathbf{c} = 0$，我们得到 $|\mathbf{a}_1|^2 = \mathbf{a}_1 \cdot (\mathbf{a} - \cos\alpha\,\mathbf{c}) = \mathbf{a}_1 \cdot \mathbf{a}$，而且

$$\mathbf{a}_1 \cdot \mathbf{a} = \mathbf{a}\cdot\mathbf{a} - \cos\alpha\,(\mathbf{c}\cdot\mathbf{a}) = 1 - \cos^2\alpha = \sin^2\alpha.$$

由于 $\alpha$ 是锐角，$\sin\alpha > 0$，所以 $|\mathbf{a}_1| = \sin\alpha$。同样地 $|\mathbf{b}_1| = \sin\beta$。两者皆非零，所以夹角 $\phi$ 有定义。




至于投影之间的夹角，再次利用 $\mathbf{a}_1 \cdot \mathbf{c} = 0$ 消去一项：

$$\mathbf{a}_1 \cdot \mathbf{b}_1 = \mathbf{a}_1 \cdot (\mathbf{b} - \cos\beta\,\mathbf{c}) = \mathbf{a}_1 \cdot \mathbf{b} = \mathbf{a}\cdot\mathbf{b} - \cos\alpha\,(\mathbf{c}\cdot\mathbf{b}) = \cos\theta - \cos\alpha\cos\beta.$$

除以 $|\mathbf{a}_1||\mathbf{b}_1| = \sin\alpha\sin\beta$，得

$$\cos\phi = \frac{\cos\theta - \cos\alpha\cos\beta}{\sin\alpha\sin\beta},$$

即为所求。

### 第 (iii) 部分

投影是线性的，所以

$$\mathbf{m}_1 = \mathbf{m} - (\mathbf{m}\cdot\mathbf{c})\,\mathbf{c} = \tfrac{1}{2}\big(\mathbf{a} - \cos\alpha\,\mathbf{c}\big) + \tfrac{1}{2}\big(\mathbf{b} - \cos\beta\,\mathbf{c}\big) = \tfrac{1}{2}(\mathbf{a}_1 + \mathbf{b}_1).$$

我们假设 $\mathbf{m}_1 \neq \mathbf{0}$，否则它没有方向，不可能平分任何东西。由于 $\mathbf{m}_1$ 是 $\mathbf{a}_1$ 与 $\mathbf{b}_1$ 的正系数组合，它位于两者所在的平面上，并且夹在两者之间。所以它平分该夹角，当且仅当它与 $\mathbf{a}_1$ 和 $\mathbf{b}_1$ 所成的角相等：

$$\frac{\mathbf{m}_1 \cdot \mathbf{a}_1}{|\mathbf{a}_1|} = \frac{\mathbf{m}_1 \cdot \mathbf{b}_1}{|\mathbf{b}_1|}.$$

（我们已约去公因子 $|\mathbf{m}_1|$，而 cos 在 $[0, \pi]$ 上是一一对应（one-to-one）的。）

> ? **想一想** 为何不直接把第 (i) 部分应用于 $\mathbf{a}_1$ 与 $\mathbf{b}_1$？

第 (i) 部分需要 $|\mathbf{a}| = |\mathbf{b}|$。这里 $|\mathbf{a}_1| = \sin\alpha$，$|\mathbf{b}_1| = \sin\beta$，两者通常不同。所以我们必须实际计算。

设 $P = \mathbf{a}_1 \cdot \mathbf{b}_1$，$s = \sin\alpha$，$t = \sin\beta$。则

$$\mathbf{m}_1 \cdot \mathbf{a}_1 = \tfrac{1}{2}(s^2 + P), \qquad \mathbf{m}_1 \cdot \mathbf{b}_1 = \tfrac{1}{2}(P + t^2).$$

两边乘以 $2st > 0$，条件变为

$$\begin{aligned} t(s^2 + P) = s(t^2 + P) &\iff st(s - t) - P(s - t) = 0 \\ &\iff (s - t)(st - P) = 0. \end{aligned}$$

所以 $\mathbf{m}_1$ 平分该夹角，当且仅当 $s = t$ 或 $P = st$。

- $s = t$ 即 $\sin\alpha = \sin\beta$。由于 sin 在 $(0, \frac{\pi}{2})$ 上严格递增，而两个角都是锐角，此式成立当且仅当 $\alpha = \beta$。
- 由第 (ii) 部分，$P = \cos\theta - \cos\alpha\cos\beta$。所以 $P = st$ 即

$$\cos\theta = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha - \beta).$$

每一步都是等价 (equivalence) 关系，所以 $\mathbf{m}_1$ 平分 $\mathbf{a}_1$ 与 $\mathbf{b}_1$ 之间的夹角，当且仅当 $\alpha = \beta$ 或 $\cos\theta = \cos(\alpha - \beta)$。

从几何角度看，第一种情况表示 $\mathbf{a}_1$ 与 $\mathbf{b}_1$ 长度相等，所以第 (i) 部分终究适用。第二种情况表示 $\cos\phi = 1$：两个投影指向同一方向（当 $\mathbf{a}$、$\mathbf{b}$ 与 $\mathbf{c}$ 共面（coplanar），且 $\mathbf{a}$ 与 $\mathbf{b}$ 位于 $\mathbf{c}$ 的同一侧时）。此时夹角为零，平分是平凡成立的。

## 验算

- 取 $\alpha = \beta = \frac{\pi}{2}$（本题不容许，但是一个有用的极限情况）。此时 $\mathbf{a}$ 与 $\mathbf{b}$ 本身已在平面上，公式给出 $\cos\phi = \cos\theta$，理应如此。
- 取 $\alpha = \frac{\pi}{3}$、$\beta = \frac{\pi}{4}$ 及 $\theta = \frac{\pi}{2}$，公式给出 $\cos\phi = -\frac{\sqrt{2}/4}{\sqrt{6}/4} = -1/\sqrt{3} \approx -0.577$。这与明确构造出这些向量后的结果一致。
- $\cos\phi$ 的公式及因式分解后的平分条件，都已用电脑在数千组随机单位向量三元组上验证。

## 常见错误

- 忘记 $|\mathbf{a}_1| \neq 1$。单位向量的投影并非单位向量，所以分母 $\sin\alpha\sin\beta$ 不可缺少。
- 以为 $\mathbf{a}_1, \mathbf{b}_1$ 的角平分线总是 $\mathbf{a}_1 + \mathbf{b}_1$。这只在长度相等时成立。一般而言，角平分线是 $\mathbf{a}_1/|\mathbf{a}_1| + \mathbf{b}_1/|\mathbf{b}_1|$。
- 除以 $\sin\alpha - \sin\beta$ 而不是因式分解，因而遗漏了 $\alpha = \beta$ 的情况。

## STEP 之外

投影到平面上是一个线性映射（*linear map*） $\mathbf{x} \mapsto \mathbf{x} - (\mathbf{x} \cdot \mathbf{n})\mathbf{n}$，其矩阵为 $I - \mathbf{n}\mathbf{n}^{\mathsf{T}}$。线性映射保持和，但不保持长度或角度，这正是本题结果如此的原因。$\cos\phi$ 的公式其实是乔装了的球面余弦定理（spherical cosine rule）。这里 $\phi$ 是以 **a**、**b** 及 **c** 为顶点的球面三角形（spherical triangle）在顶点 **c** 处的角，其三边为 $\alpha$、$\beta$ 及 $\theta$。关于投影与正交性（orthogonality），可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

# STEP III 2021，第 5 题

纯数学 · 极坐标曲线

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

两条曲线都以 $r$ 表示为 $\theta$ 的函数，所以它们在两个 $r$ 的表达式相等之处相交；若在该处两个 $\mathrm{d}r/\mathrm{d}\theta$ 的值亦相等，两曲线便在该处相切（touch）。把一切写成 $c=\cos\theta$ 的式子，相交条件便化为一个关于 $c$ 的二次方程（quadratic）。这个二次方程的重根（repeated root）是相切的一种情况，但并非唯一的情况：由于 $\cos\theta$ 本身在 $\theta=0$ 和 $\theta=\pi$ 处是平稳（stationary）的，所以在 $c=\pm1$ 处的单根（simple root）亦会给出 $\theta$ 的二重零点。

## 解答

全文设 $c=\cos\theta$，并记

$$f(\theta)=a+2\cos\theta,\qquad g(\theta)=2+\cos 2\theta=1+2\cos^2\theta.$$

注意对所有 $\theta$ 均有 $g(\theta)\geq 1$，所以第二条曲线永不经过原点 O。在 O 以外，一个 $r>0$ 的点有唯一的角 $\theta$（相差 $2\pi$ 的倍数除外）。因此两曲线只可能在 $r$ 和 $\theta$ 都相同的点相交。

### 第 (i) 部分

当 $f(\theta)=g(\theta)$ 时两曲线相交，即 $a+2c=1+2c^2$，亦即

$$2\cos^2\theta-2\cos\theta+1-a=0,$$

此即所求。为两者之差命名会方便一些：

$$h(\theta)=f(\theta)-g(\theta)=-(2c^2-2c+1-a).$$

何时相切？在公共点上，半径与切线（tangent）之间的夹角 $\psi$ 满足 $\tan\psi=r/\frac{\mathrm{d}r}{\mathrm{d}\theta}$。所以两条经过同一点 $(r,\theta)$ 的曲线在该点有相同切线，当且仅当它们的 $\mathrm{d}r/\mathrm{d}\theta$ 亦相同。因此

两曲线在 $h(\theta)=0$ 且 $h'(\theta)=0$ 之处相切。求导得

$$h'(\theta) = -2\sin\theta + 2\sin 2\theta = 2\sin\theta\,(2\cos\theta - 1).$$

所以 $h'(\theta)=0$ 当且仅当 $\cos\theta = \frac{1}{2}$ 或 $\sin\theta = 0$。

> **想一想** 关于 $c$ 的二次方程只在 $a=\frac{1}{2}$ 时有重根。那么为何还有其他 $a$ 值?

$c$ 的重根足以导致相切，但并非必要。映射 $\theta \mapsto \cos\theta$ 在 $\theta=0$ 和 $\theta=\pi$ 处是平稳的。所以只要 $c=\pm1$ 是根，$h$ 在该处便有 $\theta$ 的二重零点。条件 $h'=0$ 把两种可能都包括在内。

- $\cos\theta = \frac{1}{2}$：代入相交方程得 $\frac{1}{2} - 1 + 1 - a = 0$，所以 $a = \frac{1}{2}$。（等价地，判别式（discriminant）$4-8(1-a)$ 在 $a=\frac{1}{2}$ 时为零，此时重根为 $c=\frac{1}{2}$。）所以当 $a=\frac{1}{2}$ 时两曲线相切，此即所求。
- $\theta=0$，$c=1$：需要 $2-2+1-a=0$，所以 $a=1$。此时两曲线上皆有 $r=3$。
- $\theta=\pi$, $c=-1$: 需要 $2+2+1-a=0$, 所以 $a=5$。此时两曲线上皆有 $r=5-2=3$, 故 $r\geq 0$ 成立。

因此另外两个值是 $a=1$ 和 $a=5$。

### 第二条曲线的特征

三幅草图都包含 $r=2+\cos 2\theta$，所以我们只描述一次。它在 $\theta\to-\theta$ 和 $\theta\to\pi-\theta$ 下不变，所以它对两条坐标轴都对称（symmetric）。这里 $r$ 由 $\theta=0,\pi$ 处的最大值 3 变到 $\theta=\pm\frac{1}{2}\pi$ 处的最小值 1。它不是椭圆。在 $\theta=\frac{1}{2}\pi$ 附近，设 $s=\sin\theta$,

$$y = r\sin\theta = (3-2s^2)s = 3s-2s^3, \qquad \frac{\mathrm{d}y}{\mathrm{d}s} = 3-6s^2 = -3 \text{ 当} s=1.$$

所以当 $\theta$ 离开 $\frac{1}{2}\pi$（而 $s$ 降到 1 以下）时，$y$ 反而增大。曲线在 $(0,\pm1)$ 处有一个光滑的凹陷（dimple），而不是尖点（cusp），因为该处 $\mathrm{d}r/\mathrm{d}\theta = -2\sin 2\theta = 0$，切线是水平的。相比之下，$x=r\cos\theta = c+2c^3$ 关于 $c$ 递增，所以 $(\pm3,0)$ 两端只是圆滑的。

第一条曲线 $r=a+2\cos\theta$ 关于初始线（initial line）对称，因为它在 $\theta\to-\theta$ 下不变。它在 $\theta=0$ 处有 $r=a+2$，在 $\theta=\pm\frac{1}{2}\pi$ 处有 $r=a$，而（当 $a\geq 2$ 时）在 $\theta=\pi$ 处有 $r=a-2$。当 $a<2$ 时，限制 $r\geq 0$ 使曲线在 $\cos\theta=-\frac{1}{2}a$ 处被截断，曲线沿这两个方向到达 O。

### 第 (ii) 部分

当 $a=\frac{1}{2}$ 时，

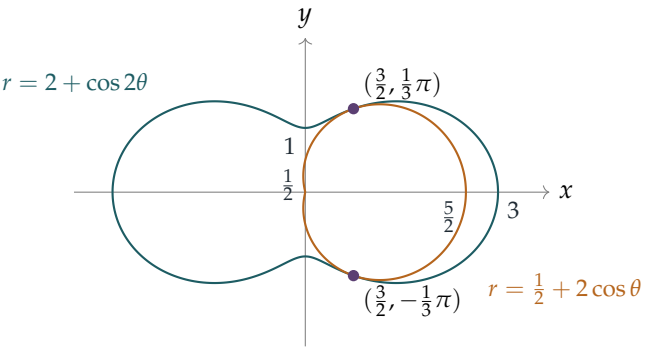
$$h(\theta) = -\left(2c^2-2c+\tfrac{1}{2}\right) = -2\left(c-\tfrac{1}{2}\right)^2 \leq 0.$$

所以第一条曲线处处位于第二条曲线之内，并只在 $\cos\theta = \frac{1}{2}$ 处与之相交。两者相切于

$$\theta = \pm\tfrac{1}{3}\pi, \qquad r = \tfrac{1}{2} + 1 = \tfrac{3}{2}.$$

（验算：$2 + \cos\frac{2}{3}\pi = \frac{3}{2}$，且在 $\theta = \frac{1}{3}\pi$ 处两个 $\mathrm{d}r/\mathrm{d}\theta$ 的值都等于 $-\sqrt{3}$。）

第一条曲线只在 $\cos\theta \geq -\frac{1}{4}$ 时存在，即 $|\theta| \leq \arccos(-\frac{1}{4}) \approx 104.5°$。它由初始线上的 $r = \frac{5}{2}$ 出发，经过 $\theta = \pm\frac{1}{2}\pi$ 处的 $r = \frac{1}{2}$，一直到达 O。它沿直线 $\theta = \pm\arccos(-\frac{1}{4})$ 到达 O，所以在原点处有一个角点（corner，即尖端），位置刚好在 $y$ 轴左侧。

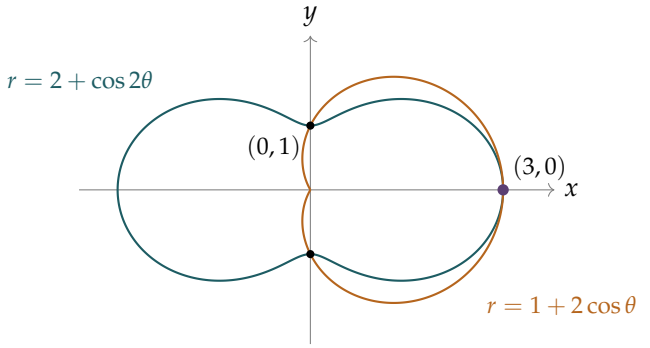



**第 (iii) 部分**

$a = 1$ 的情况。此时 $h(\theta) = -(2c^2 - 2c) = 2c(1-c)$。两曲线在 $c = 1$ 或 $c = 0$ 处相交。

- 它们在 $\theta = 0$、$r = 3$ 处相切，即点 $(3, 0)$。在 $\theta = 0$ 附近有 $0 < c < 1$，所以 $h > 0$，第一条曲线在切点两侧都略为位于第二条曲线之外。
- 在 $\theta = \pm\frac{1}{2}\pi$ 处，两曲线上皆有 $r = 1$：即点 $(0, \pm 1)$，也就是凹陷的底部。这里 $h$ 随 $c$ 变号，所以两曲线相交而穿过。第一条曲线在 $|\theta| < \frac{1}{2}\pi$ 时位于外侧，在 $|\theta| > \frac{1}{2}\pi$ 时位于内侧。

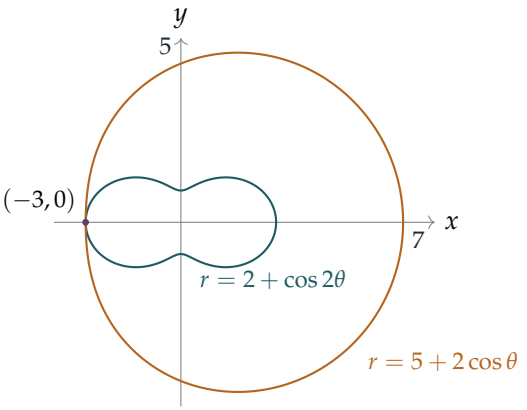
第一条曲线在 $\cos\theta \geq -\frac{1}{2}$ 时存在，即 $|\theta| \leq \frac{2}{3}\pi$。它沿直线 $\theta = \pm\frac{2}{3}\pi$ 到达 O，并在该处形成一个角点。



$a = 5$ 的情况。此时 $h(\theta) = -(2c^2 - 2c - 4) = 2(2 - c)(1 + c) \geq 0$，等号只在 $c = -1$ 时成立。所以第一条曲线完全位于第二条曲线之外，并只在

$$\theta = \pi, \qquad r = 3,$$

处与之相切，即点 $(-3, 0)$。由于 $a > 2$，现在处处有 $r > 0$，所以第一条曲线是一条闭合环线（一条没有凹陷的蚶线（limaçon），因为 $a \geq 4$）。它在 $\theta = 0$ 处有 $r = 7$，在 $\theta = \pm\frac{1}{2}\pi$ 处有 $r = 5$，在 $\theta = \pi$ 处有 $r = 3$。




## 验算

- 在每个切点，比较两个 $r$ 值和两个 $\mathrm{d}r/\mathrm{d}\theta$ 值。对于 $a = 1$ 在 $\theta = 0$ 处，两者皆 $r = 3$，两个导数皆为 0；$a = 5$ 在 $\theta = \pi$ 处亦同样成立。
- 以电脑在 $-3 \leq a < 9$ 范围内以 0.01 为步长扫描，只在 $a = \frac{1}{2}$、1 和 5 时找到 $h = h' = 0$（且 $r \geq 0$）。这与因式分解（factorisation）$h' = 2\sin\theta(2\cos\theta - 1)$ 吻合。
- 点 $(0, 1)$ 位于第二条曲线上，而附近 $\theta = \frac{1}{2}\pi + 0.1$ 处的点有 $y \approx 1.015 > 1$。这证实了凹陷的存在。

## 常见错误

- 只做到判别式便停下，因而漏掉 $a = 1$ 和 $a = 5$。关于 $c = \cos\theta$ 的单根，仍然可以是关于 $\theta$ 的二重根。
- 把 $r = 2 + \cos 2\theta$ 画成椭圆，或在 $(0, \pm 1)$ 处画成尖点而不是光滑的凹陷。
- 在 $a = \frac{1}{2}$ 或 $a = 1$ 时画出蚶线的内环。该内环上 $r < 0$，题目已将之排除。

## STEP 之外

相切实际上是关于重数（multiplicity）的命题。两条曲线在其方程之差沿曲线有重零点之处相切，而参数化（parametrisation）可以制造或隐藏这类零点，正如本题中的 $\cos\theta$ 一样。在大学第一门关于曲线的课程中，这会发展成接触阶（order of contact）的概念，而极坐标（polar）曲线的曲率（curvature）$\left(r^2+2r'^2-rr''\right)/\left(r^2+r'^2\right)^{3/2}$ 正好解释了 $2+\cos 2\theta$ 的凹陷从何而来。如想进一步了解参数曲线及其几何，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）；至于光滑性与相切背后的分析，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2021，第 6 题

纯数学 · 反函数与导数

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

两个函数看起来都令人望而生畏，但它们的导数（derivative）原来都非常简单。导数为 $\frac{1}{1+x^2}$ 的函数必定是 $\tan^{-1} x$ 加上一个常数，而导数为零的函数是常数，但只是在它有定义的每一个区间（interval）上各自成立。所以真正的工作是利用一个方便的数值或一个极限，求出每一段上各自的常数，并留意函数在无定义之点的跳跃。

## 解答

### 第 (i) 部分

设 $t = \tan\alpha$，则 $t > 0$；再设 $u = \dfrac{xt+1}{t-x}$，于是 $f_\alpha(x) = \tan^{-1} u$。由商法则（quotient rule），

$$\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{t(t-x)+(xt+1)}{(t-x)^2} = \frac{1+t^2}{(t-x)^2},$$

而且

$$1+u^2 = \frac{(t-x)^2+(xt+1)^2}{(t-x)^2} = \frac{t^2+x^2+x^2t^2+1}{(t-x)^2} = \frac{(1+t^2)(1+x^2)}{(t-x)^2}.$$

所以由链式法则（chain rule），

$$f'_\alpha(x) = \frac{1}{1+u^2}\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{1+x^2},$$

即所求。

? **想一想** 这是否表示 $f_\alpha(x) = \tan^{-1} x + c$，其中 $c$ 是单一个常数?

不是。函数在 $x = \tan\alpha$ 处无定义，所以定义域（domain）是两个分开的区间；在每一个区间上，$f_\alpha(x) - \tan^{-1} x$ 的导数为零，因此是常数，但两个常数不一定相同。

*当 $x < \tan\alpha$ 时*：在 $x = 0$ 处，我们有 $f_\alpha(0) = \tan^{-1}(\cot\alpha) = \frac{1}{2}\pi - \alpha$，因为 $0 < \frac{1}{2}\pi - \alpha < \frac{1}{2}\pi$。因此

$$f_\alpha(x) = \tan^{-1} x + \tfrac{1}{2}\pi - \alpha \qquad (x < \tan\alpha).$$

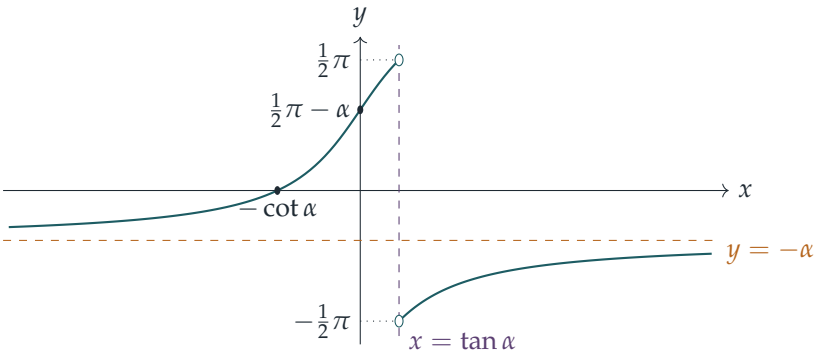
*当 $x > \tan\alpha$ 时*：当 $x \to \tan\alpha$ 而 $x$ 从上方趋近时，分子 $xt + 1$ 趋于 $1 + t^2 > 0$，而分母 $t - x$ 经负值趋于 0，所以 $u \to -\infty$，且 $f_\alpha(x) \to -\frac{1}{2}\pi$。由于 $\tan^{-1} x \to \alpha$，常数是 $-\frac{1}{2}\pi - \alpha$：

$$f_\alpha(x) = \tan^{-1} x - \tfrac{1}{2}\pi - \alpha \qquad (x > \tan\alpha).$$

（从下方用同样的极限论证，可得 $u \to +\infty$ 及数值 $\frac{1}{2}\pi$，与第一条公式一致。）

所以这个图像就是 $y = \tan^{-1} x$ 的图像分成两段，一段向上平移 $\frac{1}{2}\pi - \alpha$，另一段向下平移 $\frac{1}{2}\pi + \alpha$。它的特征如下：

- 在每一段上都是递增的，斜率（gradient）为 $\frac{1}{1+x^2}$；
- 当 $x \to \pm\infty$ 时，$y \to -\alpha$（左端从上方趋近，右端从下方趋近），所以 $y = -\alpha$ 是两端的渐近线（asymptote）；
- 左段由 $-\alpha$ 上升至 $x = \tan\alpha$ 处的 $\frac{1}{2}\pi$（不能达到），在 $\frac{1}{2}\pi - \alpha$ 与 $y$ 轴相交，并在 $\tan^{-1} x = \alpha - \frac{1}{2}\pi$ 之处，即 $x = -\cot\alpha$，与 $x$ 轴相交；
- 右段在 $x = \tan\alpha$ 处从略高于 $-\frac{1}{2}\pi$ 的位置开始，上升至 $-\alpha$；它整段位于 $x$ 轴下方；
- 在 $x = \tan\alpha$ 处有一个 $-\pi$ 的跳跃，而两段在该处的斜率都是 $\cos^2\alpha$。

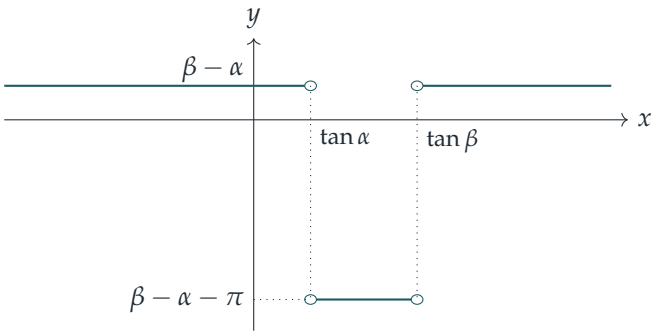



*两者之差*。由于 $0 < \alpha < \beta < \frac{1}{2}\pi$，我们有 $\tan\alpha < \tan\beta$，所以共有三个区间。利用每个函数的两条公式：

- 当 $x < \tan\alpha$ 时：$f_\alpha(x) - f_\beta(x) = (\frac{1}{2}\pi - \alpha) - (\frac{1}{2}\pi - \beta) = \beta - \alpha$；
- 当 $\tan\alpha < x < \tan\beta$ 时：$f_\alpha(x) - f_\beta(x) = (-\frac{1}{2}\pi - \alpha) - (\frac{1}{2}\pi - \beta) = \beta - \alpha - \pi$；

- 当 $x > \tan\beta$ 时：$f_\alpha(x) - f_\beta(x) = (-\frac{1}{2}\pi - \alpha) - (-\frac{1}{2}\pi - \beta) = \beta - \alpha$。

$\tan^{-1} x$ 项互相抵消，所以图像是一个阶梯函数（step function）：它是水平线 $y = \beta - \alpha$（位于 0 与 $\frac{1}{2}\pi$ 之间），但在 $\tan\alpha$ 与 $\tan\beta$ 之间的区间上则下降 $\pi$，变为 $y = \beta - \alpha - \pi$（位于 $-\pi$ 与 $-\frac{1}{2}\pi$ 之间）。函数在 $x = \tan\alpha$ 及 $x = \tan\beta$ 处无定义。




**第 (ii) 部分**

逐项求导。由于 $\dfrac{\mathrm{d}}{\mathrm{d}s}\tanh^{-1} s = \dfrac{1}{1-s^2}$，

$$\frac{\mathrm{d}}{\mathrm{d}x}\tanh^{-1}(\sin x) = \frac{\cos x}{1-\sin^2 x} = \frac{1}{\cos x} = \sec x,$$

又由于 $\dfrac{\mathrm{d}}{\mathrm{d}s}\sinh^{-1} s = \dfrac{1}{\sqrt{1+s^2}}$，

$$\frac{\mathrm{d}}{\mathrm{d}x}\sinh^{-1}(\tan x) = \frac{\sec^2 x}{\sqrt{1+\tan^2 x}} = \frac{\sec^2 x}{\sqrt{\sec^2 x}} = \frac{\sec^2 x}{|\sec x|} = |\sec x|.$$

这里的平方根取正值，所以才会出现绝对值（modulus）。因此 $g'(x) = \sec x - |\sec x|$。当 $\frac{1}{2}\pi < x < \frac{3}{2}\pi$ 时，$\cos x < 0$，所以 $|\sec x| = -\sec x$，从而

$$g'(x) = 2\sec x,$$

即所求。在另外两个区间 $0 \le x < \frac{1}{2}\pi$ 及 $\frac{3}{2}\pi < x \le 2\pi$ 上，$\cos x > 0$，所以同样的计算得出 $g'(x) = 0$。

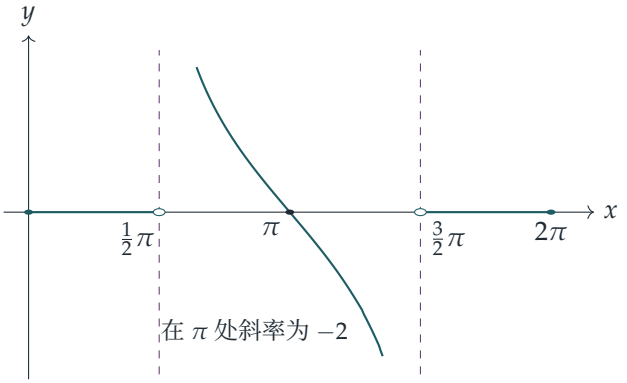
*外侧的区间*。在这里 $g$ 在每个区间上都是常数。由于 $g(0) = \tanh^{-1} 0 - \sinh^{-1} 0 = 0$，同样 $g(2\pi) = 0$，我们得到当 $0 \le x < \frac{1}{2}\pi$ 及 $\frac{3}{2}\pi < x \le 2\pi$ 时，$g(x) = 0$。

*中间的区间*。在这里 $g'(x) = 2\sec x < 0$，所以 $g$ 严格递减（strictly decreasing）。在中点处，$g(\pi) = \tanh^{-1} 0 - \sinh^{-1} 0 = 0$，且 $g'(\pi) = 2\sec\pi = -2$。在两端：

- 当 $x \to \frac{1}{2}\pi^+$ 时，$\sin x \to 1$，所以 $\tanh^{-1}(\sin x) \to +\infty$，而 $\tan x \to -\infty$，所以 $-\sinh^{-1}(\tan x) \to +\infty$；因此 $g(x) \to +\infty$；

► 当 $x \to \frac{3}{2}\pi^-$ 时，$\sin x \to -1$ 且 $\tan x \to +\infty$，所以两项都令 $g(x) \to -\infty$。

此外 $g''(x) = 2\sec x \tan x$。当 $\frac{1}{2}\pi < x < \pi$ 时，$\sec x$ 和 $\tan x$ 都是负数，所以 $g'' > 0$；当 $\pi < x < \frac{3}{2}\pi$ 时，$\sec x < 0 < \tan x$，所以 $g'' < 0$。因此 $(\pi, 0)$ 是一个拐点（point of inflection），该处斜率为 $-2$，而曲线关于这一点旋转对称（rotational symmetry）（事实上 $g'(2\pi - x) = g'(x)$ 且 $g(\pi) = 0$）。其实，由于 $g$ 与 $2\tanh^{-1}(\sin x)$ 在这个区间上有相同的导数，而且两者在 $\pi$ 处都等于零，所以在该区间上 $g(x) = 2\tanh^{-1}(\sin x) = 2\ln|\sec x + \tan x|$。




## 验算

► 取 $\alpha = 0.6$：$f_\alpha(0) = \tan^{-1}(\cot 0.6) \approx 0.9708 = \frac{1}{2}\pi - 0.6$，而 $f_\alpha(\pm 10^9) \approx -0.6$，与渐近线的预测一致。

► $g(1) = \tanh^{-1}(\sin 1) - \sinh^{-1}(\tan 1) = 1.2261 - 1.2261 = 0$，证实了平坦的部分。用电脑检验数以千计的随机点，结果与上述所有公式的误差都在 $10^{-6}$ 以内。

## 常见错误

► 写成 $f_\alpha(x) = \tan^{-1} x + c$，处处只用一个常数。导数为零只能在一个区间上推出常数。

► 写成 $\sqrt{\sec^2 x} = \sec x$。它其实是 $|\sec x|$，而正是这个符号，令 $g$ 在两个区间上为零，在第三个区间上却不是。

## STEP 之外

在一个区间上导数为零就迫使函数为常数，这是中值定理（Mean Value Theorem）的一个推论；而在不是区间的集合上，这个结论并不成立，两个部分都显示了这一点。第 (i) 部

分其实就是 $\tan^{-1}$ 的加法公式（addition formula）连同 $\pm\pi$ 的修正，这种分支（branch）的记账工作在复分析（complex analysis）中十分重要。Abbott 的 *Understanding Analysis* 对中值定理及其推论有细致的论述。

详细解答

# STEP III 2021，第 7 题

纯数学 · 几何中的复数

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

把单位圆（unit circle）上的两点相加或相减，即 $\mathrm{e}^{\mathrm{i}\theta} \pm \mathrm{e}^{\mathrm{i}\phi}$ 时，先抽出“平均”因子 $\mathrm{e}^{\mathrm{i}(\theta+\phi)/2}$。剩下的部分就是纯粹的余弦或纯粹的正弦，第 (i) 部分正是靠这一点得以成立。第 (i) 部分的结论是 $(a+b)/(a-b)$ 为纯虚数（purely imaginary），而在复平面（complex plane）上，“比值为纯虚数”就代表“互相垂直（perpendicular）”。之后的一切，不过是对 $a$、$b$、$c$、$d$ 的各种和作整理而已。

## 解答

### 第 (i) 部分

设 $\alpha = \frac{1}{2}(\theta+\phi)$ 及 $\beta = \frac{1}{2}(\theta-\phi)$，于是 $\theta = \alpha+\beta$ 而 $\phi = \alpha-\beta$。那么

$$\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{\mathrm{i}\phi} = \mathrm{e}^{\mathrm{i}\alpha}\left(\mathrm{e}^{\mathrm{i}\beta} + \mathrm{e}^{-\mathrm{i}\beta}\right) = 2\cos\beta\,\mathrm{e}^{\mathrm{i}\alpha},$$
$$\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi} = \mathrm{e}^{\mathrm{i}\alpha}\left(\mathrm{e}^{\mathrm{i}\beta} - \mathrm{e}^{-\mathrm{i}\beta}\right) = 2\mathrm{i}\sin\beta\,\mathrm{e}^{\mathrm{i}\alpha}.$$

条件 $\theta-\phi \neq 2n\pi$ 正好表示 $\beta$ 不是 $\pi$ 的倍数，因此 $\sin\beta \neq 0$，分母不为零。两式相除，得

$$z = \frac{2\cos\beta}{2\mathrm{i}\sin\beta} = -\mathrm{i}\cot\beta = -\mathrm{i}\cot\left(\tfrac{1}{2}(\theta-\phi)\right) = \mathrm{i}\cot\left(\tfrac{1}{2}(\phi-\theta)\right),$$

这里用到 $1/\mathrm{i} = -\mathrm{i}$，以及 cot 是奇函数（odd function）这一事实。这就是所求的结果。

由于 $\cot\left(\frac{1}{2}(\phi-\theta)\right)$ 是实数，$z$ 位于虚轴上：

$$|z| = \left|\cot\left(\tfrac{1}{2}(\phi-\theta)\right)\right|, \qquad \arg z = \begin{cases} \frac{1}{2}\pi & \text{若 } \cot\left(\frac{1}{2}(\phi-\theta)\right) > 0, \\ -\frac{1}{2}\pi & \text{若 } \cot\left(\frac{1}{2}(\phi-\theta)\right) < 0. \end{cases}$$

若 $\cot\left(\frac{1}{2}(\phi-\theta)\right)=0$，即 $\phi-\theta$ 是 $\pi$ 的奇数倍，则 $z=0$，其辐角（argument）没有定义。（这种情况出现在两点位于直径两端时，此时两者之和为零。）

### 第 (ii) 部分

由于 $A$ 和 $B$ 都在以原点为圆心的单位圆上，可记 $a=\mathrm{e}^{\mathrm{i}\theta}$ 及 $b=\mathrm{e}^{\mathrm{i}\phi}$。两点相异，所以 $\theta-\phi$ 不是 $2\pi$ 的倍数，第 (i) 部分适用：

$$\frac{x}{a-b}=\frac{a+b}{a-b}=\mathrm{i}k,\qquad k=\cot\left(\tfrac{1}{2}(\phi-\theta)\right)\in\mathbb{R}.$$

因为 $x=a+b\neq 0$,所以 $k\neq 0$。于是 $x=\mathrm{i}k\,(a-b)$：向量 $\overrightarrow{OX}$（由 $x$ 表示）就是把向量 $\overrightarrow{BA}$（由 $a-b$ 表示）旋转一个直角，再乘以非零实数 $k$ 所得。等价地说，$\arg x-\arg(a-b)=\pm\frac{1}{2}\pi$。因此 $OX$ 垂直于 $AB$。

> **想一想** 为甚么需要 $a+b\neq 0$?

若 $a+b=0$，则 $X$ 就是原点本身。这时根本没有直线 $OX$，也就谈不上垂直。在代数上，这对应于 $k=0$ 的情况，此时 $x$ 的辐角没有定义。

### 第 (iii) 部分

设 $h=a+b+c$，则

$$h-a=b+c.$$

复数 $h-a$ 表示向量 $\overrightarrow{AH}$。分两种情况。

- 若 $b+c=0$，则 $h-a=0$，故 $h=a$。
- 若 $b+c\neq 0$，则对单位圆上的相异两点 $B$ 和 $C$ 应用第 (ii) 部分，可知由 $b+c$ 表示的向量垂直于 $BC$。这个向量正是 $\overrightarrow{AH}$，所以 $AH$ 垂直于 $BC$。（此时 $H\neq A$，故 $AH$ 确实是一条直线。）

所以 $h=a$ 或 $AH\perp BC$，即为所求。由对称性，$BH$ 与 $CA$、$CH$ 与 $AB$ 亦有同样结论，所以 $H$ 确实位于三条高（altitude）上。$h=a$ 的情况就是 $BC$ 为直径的时候：三角形在 $A$ 处是直角，其垂心（orthocentre）就是顶点 $A$。

### 第 (iv) 部分

根据垂心的定义，

$$p=a+b+c,\qquad q=b+c+d,\qquad r=c+d+a,\qquad s=d+a+b.$$

$AQ$ 的中点是

$$\tfrac{1}{2}(a+q)=\tfrac{1}{2}(a+b+c+d).$$

这个式子对 $a,b,c,d$ 是对称的，所以它同时也是 $BR$、$CS$ 及 $DP$ 的中点：

$$\tfrac{1}{2}(b+r)=\tfrac{1}{2}(c+s)=\tfrac{1}{2}(d+p)=\tfrac{1}{2}(a+b+c+d).$$

设 $M$ 为由 $m=\frac{1}{2}(a+b+c+d)$ 表示的点。那么

$$q=2m-a,\qquad r=2m-b,\qquad s=2m-c,\qquad p=2m-d.$$

所以单一个映射

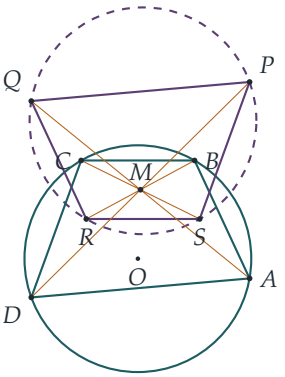
$$T:\ w\longmapsto 2m-w$$

把 $A\mapsto Q$、$B\mapsto R$、$C\mapsto S$ 及 $D\mapsto P$。因此它把四边形 $ABCD$ 逐个顶点地映射到 $QRSP$。

要描述 $T$，可把它写成 $T(w)-m=-(w-m)$。相对于 $M$ 的位移被反向，也就是以 $M$ 为中心乘以 $\mathrm{e}^{\mathrm{i}\pi}$。所以

$T$ 是绕复数为 $\frac{1}{2}(a+b+c+d)$ 的点 $M$ 旋转（*rotation*）180°（即半周旋转）。

等价地说，$T$ 是关于点 $M$ 的点反射（point reflection）。特别地，$QRSP$ 与 $ABCD$ 全等（congruent），而且由于旋转保持定向，$Q,R,S,P$ 同样按逆时针方向排列。四个垂心都位于以复数 $a+b+c+d$ 所表示的点为圆心的单位圆上，这个圆就是经过 $A,B,C,D$ 的圆的像。




## 验算

- 第 (i) 部分：取 $\theta=0$、$\phi=\frac{1}{2}\pi$，得 $z=(1+\mathrm{i})/(1-\mathrm{i})=\mathrm{i}$，而 $\mathrm{i}\cot\frac{1}{4}\pi=\mathrm{i}$。所以 $|z|=1$ 且 $\arg z=\frac{1}{2}\pi$。
- 第 (iv) 部分：取 $a=1$、$b=\mathrm{i}$、$c=-1$、$d=\frac{3}{5}-\frac{4}{5}\mathrm{i}$。则 $m=0.3+0.1\mathrm{i}$，而 $q=b+c+d=-0.4+0.2\mathrm{i}$，正如所言等于 $2m-a$。对于正方形 $(a,b,c,d=1,\mathrm{i},-1,-\mathrm{i})$，得 $m=0$，而 $Q,R,S,P$ 恰好就是 $C,D,A,B$：半周旋转的中心是 $O$。
- 四个部分亦已用电脑在圆上数以千计的随机点上验证过。

## 常见错误

- 只答 $\arg z = \frac{1}{2}\pi$。余切的正负号决定答案是 $\pm\frac{1}{2}\pi$ 中的哪一个，而 $z = 0$ 则没有辐角。
- 在第 (iv) 部分把顶点的对应次序弄错。$A$ 对应的是 $Q$，不是 $P$：与每个顶点相对的垂心，是*不包含*该顶点的那个三角形的垂心。
- 把映射描述为“一个旋转”却没有说明中心和角度，或描述为“比例因子为 $-1$ 的放大（enlargement）”却没有说明中心。两种描述都可以，但都必须包含齐两项资料。

## STEP 之外

简洁的公式 $h = a + b + c$（外心（circumcentre）位于原点时）是一整套复数几何手法的起点。例如，形心（centroid）是 $\frac{1}{3}(a + b + c)$，所以 $O$、形心与 $H$ 共线，这条直线就是欧拉线（Euler line）。第 (iv) 部分是同一故事在圆内接四边形（cyclic quadrilateral）上的一个小片段。Tristan Needham 的 *Visual Complex Analysis*（《直观复分析》）全书都把复数乘法看作旋转与缩放，是极佳的下一步读物。

详细解答

# STEP III 2021，第 8 题

纯数学 · 递推关系与极限

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 2.8 节平方再减二中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/squaring-minus-two.html

## 关键思路

一切都建基于把递推关系（recurrence）因式分解（factorise）。两边减去 2，得 $x_{n+1}-2=(x_n-2)(x_n+2)$，所以一旦各项都大于 2，它们与 2 的距离在每一步都会被乘以一个大于 4 的数。由此得到第 (i) 部分，第 (ii) 部分亦随之而来。至于第 (iii) 部分，从 $x_{n+1}^2$ 减去 4 同样可以漂亮地因式分解，而正正是这一点令乘积 $x_1x_2\cdots x_n$ 可以逐项相消（telescope）。

## 解答

### 第 (i) 部分

我们对 $n$ 使用数学归纳法（induction）。当 $n=1$ 时，命题变成 $x_1\geq 2+4^0(a-2)=a$，这是成立的，而且取等号（注意 $4^0=1$）。

假设对某个 $n\geq 1$ 有 $x_n\geq 2+4^{n-1}(a-2)$。由于 $a>2$，右方大于 2，所以 $x_n>2$。于是

$$x_{n+1}-2=x_n^2-4=(x_n-2)(x_n+2).$$

这里 $x_n-2>0$ 而且 $x_n+2>4$，所以 $(x_n-2)(x_n+2)\geq 4(x_n-2)$。利用归纳假设（inductive hypothesis）$x_n-2\geq 4^{n-1}(a-2)$，

$$x_{n+1}-2\geq 4(x_n-2)\geq 4\cdot 4^{n-1}(a-2)=4^n(a-2),$$

这正是 $n+1$ 时的命题。由归纳法，对所有 $n\geq 1$ 都有 $x_n\geq 2+4^{n-1}(a-2)$。

我们从来没有把不等式平方。把 $x_n \geq$ (界) 两边平方，只有在两边都非负时才成立，而因式分解的形式完全避开了这个顾虑。

### 第 (ii) 部分

若 $|a| > 2$。此时 $x_2 = a^2 - 2 > 4 - 2 = 2$。数列（sequence）$x_2, x_3, \ldots$ 服从同一个递推关系，而且首项大于 2，所以以 $x_2$ 代替 $a$（把下标平移一位）应用第 (i) 部分，得

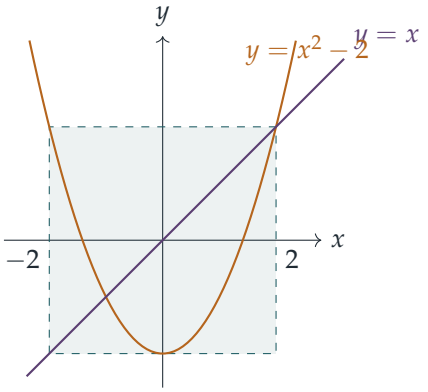
$$x_n \geq 2 + 4^{n-2}(x_2 - 2) \qquad (n \geq 2).$$

由于 $x_2 - 2 > 0$ 而且 $4^{n-2} \to \infty$，得 $x_n \to \infty$。

若 $|a| \leq 2$。我们用归纳法证明对每个 $n$ 都有 $-2 \leq x_n \leq 2$。这在 $n = 1$ 时成立。若 $-2 \leq x_n \leq 2$，则 $0 \leq x_n^2 \leq 4$，所以 $-2 \leq x_n^2 - 2 \leq 2$，即 $-2 \leq x_{n+1} \leq 2$。因此这个数列有界（bounded），不可能趋向无限大。

> **想一想** 在 $|a| \leq 2$ 的情况下，只证明对所有 $n$ 都有 $x_n \leq 2$ 足够吗?

不足够。单有上界并不够，因为若某一项小于 $-2$，它的平方便会大于 4，于是下一项就会大于 2。这就是为甚么归纳过程必须同时带着两个界。图像说明了为何这样可行：抛物线（parabola）$y = x^2 - 2$ 把区间 $[-2, 2]$ 映射到它自身上。




综合两种情况，$x_n \to \infty$ 当且仅当 $|a| > 2$。

### 第 (iii) 部分

推动这部分一切的因式分解是

$$x_{n+1}^2 - 4 = (x_n^2 - 2)^2 - 4 = x_n^4 - 4x_n^2 = x_n^2\,(x_n^2 - 4).$$

由于 $a > 2$，第 (i) 部分给出对所有 $n$ 都有 $x_n > 2$，所以 $x_n > 0$ 而且 $x_n^2 - 4 > 0$。取正平方根，

$$\sqrt{x_{n+1}^2 - 4} = x_n \sqrt{x_n^2 - 4}.$$

重复应用这个关系（严格来说，是对 $n$ 用归纳法），

$$\sqrt{x_{n+1}^2 - 4} = x_n \sqrt{x_n^2 - 4} = x_n x_{n-1} \sqrt{x_{n-1}^2 - 4} = \cdots = x_1 x_2 \cdots x_n \sqrt{a^2 - 4}.$$

因此，对每个 $n \geq 1$，

$$\frac{\sqrt{x_{n+1}^2 - 4}}{x_{n+1}} = \frac{\sqrt{a^2 - 4}\, x_1 x_2 \cdots x_n}{x_{n+1}},$$

与 $y_n$ 的定义比较，可见所求的公式对所有 $n$ 成立，当且仅当 $A = \sqrt{a^2 - 4}$。

*求 $a$。* $A = \sqrt{a^2 - 4}$ 这个值是必然的。在 $n = 1$ 时，所求的公式变成 $Aa/x_2 = a\sqrt{a^2 - 4}/x_2$，由于 $a > 0$ 而且 $x_2 > 0$，得 $A = \sqrt{a^2 - 4}$。反过来，在 $a > 2$ 之下 $A = \sqrt{a^2 - 4}$ 表示 $A^2 = a^2 - 4$，所以

$$a = \sqrt{A^2 + 4}.$$

这个值确实满足 $a > 2$，因为 $A > 0$ 给出 $A^2 + 4 > 4$。取这个 $a$，便有 $\sqrt{a^2 - 4} = A$，而上面列出的恒等式（identity）就变成对所有 $n \geq 1$ 都有 $y_n = \sqrt{x_{n+1}^2 - 4} \,/\, x_{n+1}$，正如所求。

*收敛性。* 取这个 $a$ 值，由于 $x_{n+1} > 0$，

$$y_n = \frac{\sqrt{x_{n+1}^2 - 4}}{x_{n+1}} = \sqrt{1 - \frac{4}{x_{n+1}^2}}.$$

由于 $a > 2$，第 (ii) 部分给出 $x_{n+1} \to \infty$，所以 $4/x_{n+1}^2 \to 0$，因而 $y_n \to 1$。第二个数列收敛（converges），其极限（limit）是 1。

## 验算

- 取 $A = 1$，则 $a = \sqrt{5} \approx 2.236$。于是 $x_2 = 3$、$x_3 = 7$，而且 $y_1 = \sqrt{5}/3 \approx 0.745$。由公式得 $\sqrt{9 - 4}/3 = \sqrt{5}/3$，两者吻合。接着，$y_2 = \sqrt{5} \cdot 3/7 \approx 0.958$，而 $\sqrt{49 - 4}/7 = 3\sqrt{5}/7$，同样吻合。$y_3 \approx 0.99909$ 和 $y_4 \approx 0.9999996$ 这两项正飞快地趋向 1。
- 以 $t > 1$ 作代换 $a = t + 1/t$，可得 $x_n = t^{2^{n-1}} + t^{-2^{n-1}}$。这显示了增长实际上有多快（双重指数（doubly exponential）增长，远快于第 (i) 部分的 $4^{n-1}$），而且与上面每个结果都一致。
- 第 (i) 部分的界和第 (ii) 部分的有界性，亦已用电脑以大量随机起始值测试过。

## 常见错误

- 未先检查两边是否非负，就把归纳假设中的不等式平方。
- 在第 (ii) 部分，当 $|a| \leq 2$ 时只证明 $x_n \leq 2$。你还需要 $x_n \geq -2$，否则下一项可能跳到 2 以上。
- 倒推求出 $a$，却从未确认这个值对所有 $n$ 都适用，并且满足 $a > 2$。

## STEP 之外

映射（map）$x \mapsto x^2 - 2$ 是二次映射族（quadratic family）$x \mapsto x^2 + c$ 中最著名的成员。在 $[-2, 2]$ 上，代入 $x = 2\cos\theta$ 会把它变成倍角映射 $\theta \mapsto 2\theta$，这是混沌动力学（chaotic dynamics）的标准入门例子。在这个区间以外，每条轨道（orbit）都会逃逸到无限大。当 $a > 2$ 时，同样的想法可用 $x = 2\cosh u$ 实现，而第 (iii) 部分的乘积就是经典的 $\prod \cos(\theta/2^k)$ 逐项相消的双曲（hyperbolic）版本。想了解这里的极限和归纳法背后的分析学，可阅读 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2021，第 9 题

力学 · 斜碰撞与恢复系数

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

## 关键思路

几何部分是在两个角落三角形 $CPQ$ 和 $BRP$ 中各用一次正弦定理（sine rule）；把两式相乘时，$QR$ 离 $CB$ 的共同高度会互相抵消。至于反弹，光滑（smooth）的墙会保留沿墙方向的速度分量，并把垂直于墙的分量乘以 $e$，因此路径在反弹前后与墙所成的角满足 $\tan(\text{反弹后}) = e\tan(\text{反弹前})$。每次反弹各给出第一个结果中的一个括号。

## 解答

### 几何恒等式

设 $h$ 为直线 $QR$ 与 $BC$ 的距离。由于 $QR \parallel CB$，三角形 $AQR$ 是等边三角形，所以 $CQ = BR = h/\sin 60^\circ = 2h/\sqrt{3}$。

在三角形 $CPQ$ 中，$C$ 处的角是 $60^\circ$，$P$ 处的角是 $\theta$，所以 $Q$ 处的角是 $120^\circ - \theta$。由正弦定理得

$$\frac{CP}{\sin(120^\circ - \theta)} = \frac{CQ}{\sin\theta} \quad \Longrightarrow \quad (a-x)\sin\theta = \frac{2h}{\sqrt{3}}\left(\tfrac{\sqrt{3}}{2}\cos\theta + \tfrac{1}{2}\sin\theta\right).$$

两边除以 $\sin\theta$（因 $0 < \theta < 180^\circ$，故非零）并整理，得

$$\sqrt{3}\cot\theta + 1 = \frac{\sqrt{3}\,(a-x)}{h}.$$

在三角形 $BRP$ 中，$B$ 处的角是 $60^\circ$，$R$ 处的角是 $\phi$，$P$ 处的角是 $120^\circ - \phi$。由正弦定理得

$$\frac{BP}{\sin\phi} = \frac{BR}{\sin(120^\circ - \phi)} \quad \Longrightarrow \quad x\left(\tfrac{\sqrt{3}}{2}\cos\phi + \tfrac{1}{2}\sin\phi\right) = \frac{2h}{\sqrt{3}}\sin\phi,$$

因此

$$\sqrt{3}\cot\phi + 1 = \frac{4h}{\sqrt{3}\,x}.$$

把两个结果相乘，$h$ 和 $\sqrt{3}$ 都抵消掉：

$$(\sqrt{3}\cot\phi + 1)(\sqrt{3}\cot\theta + 1) = \frac{4(a-x)}{x},$$

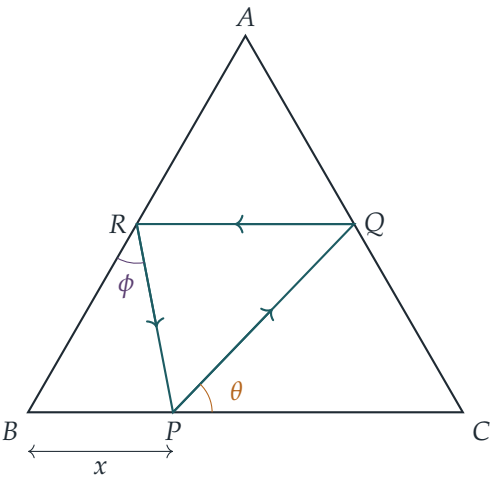
两边乘以 $x$ 即得所求的恒等式。

**反弹**

把球视为在光滑桌面上运动的质点，因此它在两次碰撞之间沿直线行进；框架亦是光滑的，所以每次碰撞只产生垂直于墙的冲量（impulse）。设球以速率 $u$、与墙成角 $\alpha$ 撞向墙，并以速率 $v$、与墙成角 $\beta$ 离开。沿墙的分量不变，垂直于墙的分量则反向并乘以 $e$：

$$v\cos\beta = u\cos\alpha, \qquad v\sin\beta = eu\sin\alpha \quad\Longrightarrow\quad \tan\beta = e\tan\alpha.$$

球沿墙的运动方向亦保持不变，所以从该方向量度时，$\alpha$ 和 $\beta$ 都是锐角。（我们需要 $e > 0$，否则第一次碰撞后球便会沿着墙滑行。）




在 $Q$ 点。球沿 $PQ$ 到达，与墙 $CA$ 成角 $\angle PQC = 120^\circ - \theta$，朝 $A$ 方向移动。它沿 $QR$ 离开，与墙成角 $\angle AQR = 60^\circ$（因 $QR \parallel CB$，两者为同位角（corresponding angles））。所以

$$\tan 60^\circ = e\tan(120^\circ - \theta).$$

利用差角公式及 $\tan 120^\circ = -\sqrt{3}$，再把分子分母同乘以 $\cot\theta$，得

$$\tan(120^\circ - \theta) = \frac{-\sqrt{3} - \tan\theta}{1 - \sqrt{3}\tan\theta} = \frac{\sqrt{3}\cot\theta + 1}{\sqrt{3} - \cot\theta}.$$

记 $c=\sqrt{3}\cot\theta$。反弹条件 $\sqrt{3}(\sqrt{3}-\cot\theta)=e(\sqrt{3}\cot\theta+1)$ 变成 $3-c=e(c+1)$，即 $4-(c+1)=e(c+1)$，所以

$$\sqrt{3}\cot\theta+1=\frac{4}{1+e}.$$

在 $R$ 点。球沿 $QR$ 到达，与墙 $AB$ 成角 $\angle ARQ=60^\circ$，朝 $B$ 方向移动，然后沿 $RP$ 离开，所成角为 $\angle BRP=\phi$。所以

$$\tan\phi=e\tan 60^\circ=\sqrt{3}e,\qquad \text{所以}\qquad \sqrt{3}\cot\phi=\frac{1}{e},\qquad \sqrt{3}\cot\phi+1=\frac{1+e}{e}.$$

求 $x$。把两个括号代入恒等式，

$$\frac{1+e}{e}\cdot\frac{4}{1+e}\cdot x=4(a-x)\quad\Longrightarrow\quad x=e(a-x)\quad\Longrightarrow\quad x=\frac{ae}{1+e},$$

正如所求。

### 再绕一圈

**想一想** 在 $P$ 点必须发生甚么，球才会再一次沿 $PQ$ 出发?

球沿 $RP$ 到达 $P$。在三角形 $BRP$ 中，这条路径与 $BC$ 成角 $\angle RPB=120^\circ-\phi$，所以只有当这个角是锐角时，球到达时才是朝 $C$ 方向移动。要重走 $PQ$，球必须以与 $PC$ 成角 $\theta$ 离开，因此我们需要

$$\tan\theta=e\tan(120^\circ-\phi).$$

由在 $Q$ 点的计算，$\sqrt{3}\cot\theta=\frac{4}{1+e}-1=\frac{3-e}{1+e}$，所以 $\tan\theta=\dfrac{\sqrt{3}(1+e)}{3-e}$，此值为正：$PQ$ 确实是朝 $C$ 那一侧前进。

把在 $Q$ 点用过的同一公式中的 $\theta$ 换成 $\phi$，并用 $\sqrt{3}\cot\phi=\frac{1}{e}$，得

$$\tan(120^\circ-\phi)=\frac{\sqrt{3}\,(\sqrt{3}\cot\phi+1)}{3-\sqrt{3}\cot\phi}=\frac{\sqrt{3}\,(1+e)/e}{(3e-1)/e}=\frac{\sqrt{3}(1+e)}{3e-1}.$$

若 $e\le\frac{1}{3}$，则 $120^\circ-\phi\ge 90^\circ$，所以碰撞后球会垂直离开 $BC$，或折返朝 $B$ 方向移动。这样它便不可能沿着向 $C$ 倾斜的 $PQ$ 前进。所以 $e>\frac{1}{3}$，而条件变为

$$\frac{\sqrt{3}(1+e)}{3-e}=\frac{\sqrt{3}\,e(1+e)}{3e-1}\quad\Longleftrightarrow\quad 3e-1=3e-e^2\quad\Longleftrightarrow\quad e^2=1.$$

由于 $e>0$，得 $e=1$，正如所求。球的速率在每次碰撞时都会改变（除非 $e=1$），但速率在这里毫不相干：重要的只是方向。

## 验算

- 当 $e = 1$：$x = \frac{a}{2}$ 且 $\tan\theta = \tan\phi = \sqrt{3}$，所以 $\theta = \phi = 60°$，而 $PQR$ 是连接各边中点的三角形。由对称性，这是一条闭合的完全弹性轨道，与预期相符。
- 取 $a = 1$、$e = \frac{1}{2}$：$x = \frac{1}{3}$ 且 $\theta \approx 46.10°$。用电脑逐步模拟球经过三次碰撞，它会准确地回到 $P$。下一次碰撞后，它以与 $BC$ 约成 $68.95°$ 的角离开，而不是 $46.10°$，所以不会再绕 $PQR$ 一圈。
- 当 $e \to 0$，$x \to 0$：$P$ 滑向角落 $B$。这与球在每道墙上几乎失去全部垂直于墙的速度相吻合。

## 常见错误

- 把 $e$ 乘到整个速度上，或乘到沿墙的分量上。只有垂直于光滑墙的分量会改变。
- 从法线量度入射角，却使用 $\tan\beta = e\tan\alpha$；这条式只适用于从墙量度的角。若从法线量度，定律应为 $\cot\beta = e\cot\alpha$。
- 在最后一部分，忘记 $120° - \phi$ 可能是钝角，或忘记必须舍去 $e = -1$。

## STEP 之外

球在一个区域内反弹的路径，是数学撞球（mathematical billiards）的研究对象。当 $e = 1$ 时，连接各边中点的三角形就是法尼亚诺（Fagnano）轨道。在任何锐角三角形中，这条轨道连接三条高的垂足，而且它是周长最短的内接三角形（inscribed triangle）。当 $e < 1$ 时，每次反弹都会使路径向墙压平；正如我们在这里发现的，球无法重复一条闭合路线。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对冲量和碰撞有仔细的讨论。至于撞球问题本身，Tabachnikov 的 *Geometry and Billiards*（《几何与撞球》）是很好的下一步。

详细解答

# STEP III 2021，第 10 题

力学 · 缠绕绳的运动

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

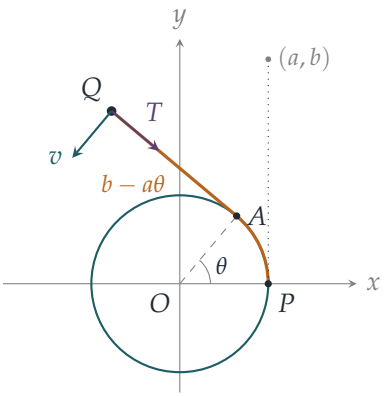
## 关键思路

用绳子直线部分转过的角度 $\theta$ 来描述一切。绳子沿切线（tangent）离开圆柱（cylinder），所以质点的位置等于切点加上一个长度为 $b - a\theta$ 的切向量。求导后可见质点的速度永远与绳子垂直，因此张力（tension）不做功（work），速率始终不变。在每一瞬间，质点都在一个以绳子离开圆柱之点为圆心的圆上运动，而张力提供所需的向心力（centripetal force）。

## 解答

我们把质点视为在光滑（smooth）水平桌面上运动的一点，因此它所受的唯一水平力就是轻而不可伸长的绳（light inextensible string）中的张力 $T$（重量由桌面的法向反作用力（normal reaction）平衡）。整个过程中绳子都保持拉紧。

要令绳子绕到圆柱上，冲量（impulse）必须使质点从 $(a, b)$ 沿 $x$ 负方向出发，于是绳子按逆时针（anticlockwise）方向转动。设 $A$ 为时刻 $t$ 时绳子直线部分离开圆柱之点。




**第 (i) 部分**

由于通往 $P$ 的半径已转过 $\theta$，点 $A$ 为 $(a\cos\theta, a\sin\theta)$。绕在圆柱上的绳长就是弧 $PA$，长度为 $a\theta$，所以直线部分的长度为 $b-a\theta$。它沿着 $A$ 点的切线、朝 $\theta$ 增大的方向伸出，即单位向量（unit vector）$(-\sin\theta, \cos\theta)$ 的方向。（当 $\theta = 0$ 时这向量是 $(0,1)$，与初始位置 $(a,b)$ 吻合。）因此质点位于

$$x = a\cos\theta - (b-a\theta)\sin\theta, \qquad y = a\sin\theta + (b-a\theta)\cos\theta.$$

利用积法则（product rule）及 $\frac{\mathrm{d}}{\mathrm{d}t}(b-a\theta) = -a\dot\theta$，对 $t$ 求导，得

$$\dot x = -a\dot\theta\sin\theta + a\dot\theta\sin\theta - (b-a\theta)\dot\theta\cos\theta = -(b-a\theta)\dot\theta\cos\theta,$$
$$\dot y = a\dot\theta\cos\theta - a\dot\theta\cos\theta - (b-a\theta)\dot\theta\sin\theta = -(b-a\theta)\dot\theta\sin\theta.$$

因此 $\dot x^2 + \dot y^2 = (b-a\theta)^2\dot\theta^2$；又因 $b-a\theta > 0$ 且 $\dot\theta > 0$（绳子正在绕上去），速率为

$$v = (b-a\theta)\dot\theta.$$

**第 (ii) 部分**

> ? **想一想** 为甚么速率应该一直等于 $u$？绳子明明一直在拉着质点。

由第 (i) 部分，速度为 $-(b-a\theta)\dot\theta\,(\cos\theta, \sin\theta)$，与半径 $OA$ 平行，因此与方向为 $(-\sin\theta, \cos\theta)$ 的绳子垂直。张力沿绳子作用，所以它做功的速率 $\mathbf{T}\cdot\mathbf{v}$ 在每一瞬间都是零。没有其他水平力作用，所以动能（kinetic energy）以至速率都保持不变。初始速率为 $u$，故

$$\dot\theta\,(b-a\theta) = u,$$

即为所求。

**第 (iii) 部分**

由 (ii)，速度为 $\mathbf{v} = -u(\cos\theta, \sin\theta)$，其中 $u$ 为常数，所以加速度为

$$\dot{\mathbf{v}} = -u\dot{\theta}(-\sin\theta, \cos\theta).$$

它的大小为 $u\dot{\theta}$，方向由质点指向 $A$，恰好沿着绳子。（等价地说：在每一瞬间，质点以速率 $u$ 绕 $A$ 在半径为 $b - a\theta$ 的圆上运动，所以它的加速度是指向 $A$ 的 $u^2/(b - a\theta)$。）沿绳子方向应用牛顿第二定律（Newton's second law），得

$$T = mu\dot{\theta} = \frac{mu^2}{b - a\theta}.$$

余下要做的是把 $b - a\theta$ 以 $t$ 表示。将 (ii) 对 $t$ 积分，并利用 $t = 0$ 时 $\theta = 0$，得

$$b\theta - \tfrac{1}{2}a\theta^2 = ut.$$

于是

$$(b - a\theta)^2 = b^2 - 2a\left(b\theta - \tfrac{1}{2}a\theta^2\right) = b^2 - 2aut,$$

又因 $b - a\theta > 0$，我们取正根，$b - a\theta = \sqrt{b^2 - 2aut}$。因此

$$T = \frac{mu^2}{\sqrt{b^2 - 2aut}},$$

即为所求。

## 验算

- 当 $t = 0$ 时张力为 $mu^2/b$：质点起初以速率 $u$ 绕 $P$ 在半径为 $b$ 的圆上运动，与预期一致。
- 当 $a\theta = b$ 时绳子完全绕上，这发生在 $t = b^2/(2au)$；此时 $b^2 - 2aut = 0$，张力无限增大，因为在速率不变的情况下，瞬时圆的半径缩小至零。
- 量纲（dimensions）：$b^2$ 和 $aut$ 都是面积，而 $mu^2$/长度 是力。
- 我们用电脑验证了结果：对 $a$、$b$、$u$、$m$ 的多组随机取值，以数值方法逐步模拟运动，并以数值方法对位置求导；速率保持为 $u$，而量得沿绳方向的力与 $mu^2/\sqrt{b^2 - 2aut}$ 相符，相对误差约为 $3 \times 10^{-5}$。

## 常见错误

- 假设绕 $O$ 的角动量（angular momentum）守恒。其实并不守恒：张力沿圆柱的切线作用，而这条切线不经过 $O$，所以张力对 $O$ 有力矩（moment）。这里守恒的量是能量。

- 把直线部分的长度当作 $b$ 而非 $b - a\theta$，忘记了已经绕在圆柱上的那段绳子。
- 在计算张力时，把质点路径的半径 $OQ$，而非直线长度 $AQ$，当作曲率半径（radius of curvature）。

## STEP 之外

绳子从一条曲线上解开（或绕上去）时，绳端描出的路径称为渐开线（involute）；圆的渐开线是大多数齿轮齿形的形状，因为它能令互相啮合的齿轮平稳地传递运动。这道题也展示了大学一年级力学中会遇到的一个普遍原理：与速度垂直的约束力（constraint force）不做功，所以即使动量（momentum）和角动量不守恒，能量仍可以守恒。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对这类问题有很好的讨论。

详细解答

# STEP III 2021，第 11 题

概率与统计 · 指数分布

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 12.18 节无记忆性中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/memoryless.html

## 关键思路

所有关于 $Y$ 和 $Z$ 的问题，其实都是在问 $X$ 落在哪一个区间（interval）。事件 $Y = n$ 就是 $n \le X < n+1$，而事件 $Z < z$ 可拆成一系列互斥（disjoint）的部分 $n \le X < n+z$，每个 $n$ 对应一部分。指数分布（exponential distribution）的尾概率（tail probability）$\mathrm{P}(X \ge x) = \mathrm{e}^{-\lambda x}$ 会把每一部分变成一个几何级数（geometric series）中的项。

## 解答

全文记 $q = \mathrm{e}^{-\lambda}$；由于 $\lambda > 0$，故 $0 < q < 1$。对 $x \ge 0$，

$$\mathrm{P}(X \ge x) = \int_x^\infty \lambda \mathrm{e}^{-\lambda t}\,\mathrm{d}t = \mathrm{e}^{-\lambda x},$$

又因 $X$ 是连续（continuous）随机变量，对 $0 \le a < b$ 有 $\mathrm{P}(a \le X < b) = \mathrm{P}(a < X < b) = \mathrm{e}^{-\lambda a} - \mathrm{e}^{-\lambda b}$。

### 第 (i) 部分

$Y$ 是 $X$ 的整数部分（integer part），所以 $Y = n$ 当且仅当 $n \le X < n+1$。因此

$$\mathrm{P}(Y = n) = \mathrm{e}^{-n\lambda} - \mathrm{e}^{-(n+1)\lambda} = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda},$$

即为所求。（所以 $Y$ 服从由 0 开始、公比为 $q$ 的几何分布（geometric distribution）。）

**第 (ii) 部分**

$Z = X - Y$ 是 $X$ 的小数部分（fractional part），所以 $0 \le Z < 1$。取 $0 \le z \le 1$。事件 $Z < z$ 发生，当且仅当对整数 $n = Y$ 有 $n \le X < n + z$。对 $n = 0, 1, 2, \ldots$，这些事件两两互斥，所以

$$\begin{aligned} \mathrm{P}(Z < z) &= \sum_{n=0}^{\infty} \mathrm{P}(n \le X < n + z) = \sum_{n=0}^{\infty} \left(\mathrm{e}^{-n\lambda} - \mathrm{e}^{-n\lambda - \lambda z}\right) \\ &= (1 - \mathrm{e}^{-\lambda z}) \sum_{n=0}^{\infty} q^n = \frac{1 - \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}}, \end{aligned}$$

这里把几何级数求和，而因 $0 < q < 1$，该级数收敛（converges）。这就是所求的结果。

**第 (iii) 部分**

把 (ii) 的分布函数（distribution function）求导，得 $Z$ 的概率密度（probability density）为

$$g(z) = \frac{\lambda \mathrm{e}^{-\lambda z}}{1 - \mathrm{e}^{-\lambda}} \qquad (0 \le z \le 1),$$

其余情况 $g(z) = 0$。取 $u = z$ 及 $\mathrm{d}v = \lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z$ 进行分部积分（integration by parts），

$$\int_0^1 z\lambda \mathrm{e}^{-\lambda z}\,\mathrm{d}z = \left[-z\mathrm{e}^{-\lambda z}\right]_0^1 + \int_0^1 \mathrm{e}^{-\lambda z}\,\mathrm{d}z = -\mathrm{e}^{-\lambda} + \frac{1 - \mathrm{e}^{-\lambda}}{\lambda}.$$

除以 $1 - \mathrm{e}^{-\lambda}$，

$$\mathrm{E}(Z) = \frac{1}{\lambda} - \frac{\mathrm{e}^{-\lambda}}{1 - \mathrm{e}^{-\lambda}} = \frac{1}{\lambda} - \frac{1}{\mathrm{e}^{\lambda} - 1}.$$

**第 (iv) 部分**

设 $0 \le z_1 < z_2 \le 1$。事件 $\{Y = n \text{ 及} z_1 < Z < z_2\}$ 恰好就是 $n + z_1 < X < n + z_2$，所以

$$\mathrm{P}(Y = n \text{ 及} z_1 < Z < z_2) = \mathrm{e}^{-\lambda(n+z_1)} - \mathrm{e}^{-\lambda(n+z_2)} = \mathrm{e}^{-n\lambda}\left(\mathrm{e}^{-\lambda z_1} - \mathrm{e}^{-\lambda z_2}\right).$$

**想一想** 要判断是否独立，我们需要拿这个结果跟甚么比较?

跟乘积 $\mathrm{P}(Y = n)\,\mathrm{P}(z_1 < Z < z_2)$ 比较。由于 $Z$ 是连续的，$\mathrm{P}(Z = z_1) = 0$，故由 (ii)

$$\mathrm{P}(z_1 < Z < z_2) = \mathrm{P}(Z < z_2) - \mathrm{P}(Z < z_1) = \frac{\mathrm{e}^{-\lambda z_1} - \mathrm{e}^{-\lambda z_2}}{1 - \mathrm{e}^{-\lambda}}.$$

乘以 (i) 中的 $\mathrm{P}(Y = n) = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-n\lambda}$，因子 $1 - \mathrm{e}^{-\lambda}$ 互相抵消，正好得到上面的联合概率（joint probability）。所以

$$\mathrm{P}(Y = n \text{ 及} z_1 < Z < z_2) = \mathrm{P}(Y = n)\,\mathrm{P}(z_1 < Z < z_2)$$

对每个非负整数 $n$ 及每个区间 $(z_1, z_2) \subseteq [0,1]$ 都成立。由于 $Y$ 只取非负整数值，而 $Z$ 只取 $[0,1)$ 中的值，这类事件决定了整个联合分布。因此 $Y$ 和 $Z$ 是独立（independent）的：知道 $X$ 落在哪一个单位区间，并不会告诉我们它落在该区间内的哪个位置。

## 验算

- 在 (ii) 中代入 $z = 1$，得 $\mathrm{P}(Z < 1) = 1$，理应如此；代入 $z = 0$ 则得 0。
- 当 $\lambda \to 0$ 时，$X$ 的密度在每个单位区间上几乎是平的，所以 $Z$ 应接近 $[0,1]$ 上的均匀分布（uniform distribution）。事实上，由 $\mathrm{e}^{\lambda} - 1 = \lambda + \frac{1}{2}\lambda^2 + \cdots$ 得 $\frac{1}{\mathrm{e}^{\lambda}-1} = \frac{1}{\lambda} - \frac{1}{2} + O(\lambda)$，所以 $\mathrm{E}(Z) \to \frac{1}{2}$。当 $\lambda \to \infty$ 时，$X$ 几乎总是非常小，而 $\mathrm{E}(Z) \to 0$。
- 当 $\lambda = 1$ 时，$\mathrm{E}(Z) = 1 - \frac{1}{\mathrm{e}-1} \approx 0.4180$，这与以电脑模拟 200 000 个 $X$ 值的结果吻合；(iv) 中模拟所得的联合频率亦与边际分布（marginal distributions）的乘积吻合。

## 常见错误

- 写成 $\mathrm{P}(Z < z) = \mathrm{P}(X < z)$，忘记了 $Z < z$ 可以在每一个单位区间发生，而不只是第一个。
- 只凭一组 $n$、$z_1$、$z_2$ 的取值就断定独立：因式分解必须对所有取值都成立。

## STEP 之外

(iv) 中的独立性，其实是指数分布无记忆性（memoryless property）的一个影子：已知 $X \geq n$，无论 $n$ 是多少，超出部分 $X - n$ 仍然服从相同参数的指数分布。这正是指数分布用来描述泊松过程（Poisson process）中等候时间的原因，也是它的整数部分服从几何分布——即离散的无记忆分布——的原因。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（《概率导论》）对这两种分布及泊松过程都有详细阐述。

详细解答

# STEP III 2021，第 12 题

概率与统计 · 和的方差

**STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

www.ocr.org.uk/Images/696588-2021-past-papers.zip

本题在 *Beyond STEP* 第 12.4 节两两独立，却非相互独立中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/pairwise-independence.html

## 关键思路

把全队的得分写成一些指示型变量（indicator variable）之和，每对球员对应一个变量，然后展开这个和的方差（variance）。一切都取决于两个不同的球员对之间如何相互影响。如果两对没有共同的球员，它们显然是独立（independent）的，所以真正要处理的是共有一名球员的两对。在 (i) 中，这样的两对原来是独立的，所以交叉项消失。在 (iii) 中则不然，而 (ii) 中的恒等式正是计算它们贡献的工具。

## 解答

### 第 (i) 部分

每颗骰子都是公正的，各球员独立掷骰。不论两名指定球员中的第一人掷出甚么，二人掷出相同点数的概率都是 $\frac{1}{6}$。所以

$$\mathrm{P}(X_{12}=1)=\mathrm{P}(X_{23}=1)=\frac{1}{6}.$$

$X_{12}=1$ 和 $X_{23}=1$ 同时发生，当且仅当球员 1、2、3 都掷出相同点数，所以

$$\mathrm{P}(X_{12}=1,\ X_{23}=1)=\frac{1}{6}\cdot\frac{1}{6}=\frac{1}{36}=\mathrm{P}(X_{12}=1)\,\mathrm{P}(X_{23}=1).$$

从边际概率（marginal probability）中减去这个值，便得到其余三格。例如 $\mathrm{P}(X_{12}=1,X_{23}=0)=\frac{1}{6}-\frac{1}{36}=\frac{1}{6}\cdot\frac{5}{6}$；同理 $\mathrm{P}(X_{12}=0,X_{23}=1)=\frac{5}{6}\cdot\frac{1}{6}$，以及 $\mathrm{P}(X_{12}=0,X_{23}=$

$0) = 1 - \frac{1}{6} - \frac{1}{6} + \frac{1}{36} = \frac{5}{6} \cdot \frac{5}{6}$。所以联合分布（joint distribution）等于边际分布的乘积，$X_{12}$ 与 $X_{23}$ 是独立的。

**想一想** 为甚么知道 $X_{12} = 1$ 对 $X_{23}$ 毫无信息?

已知球员 1 与 2 点数相同，球员 2 的点数在 $\{1, \ldots, 6\}$ 上仍然是均匀分布（uniform）的，而球员 3 与之相同的概率是 $\frac{1}{6}$，跟之前完全一样。由于每一面出现的机会均等，共有的那名球员并不带来任何信息。

设 $T = \sum_{i<j} X_{ij}$ 为总得分，它是 $N = \binom{n}{2}$ 项之和。每个 $X_{ij}$ 都是参数为 $\frac{1}{6}$ 的伯努利（Bernoulli）变量，所以 $\mathrm{E}(X_{ij}) = \frac{1}{6}$，$\mathrm{Var}(X_{ij}) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$。因此

$$\mathrm{E}(T) = \frac{N}{6} = \frac{n(n-1)}{12}.$$

至于方差，任取两个不同的球员对。如果它们没有共同球员，它们取决于互不相交的两组骰子，所以是独立的。如果它们共有一名球员，那么把球员重新编号后，由上面的论证可知它们也是独立的。所以各 $X_{ij}$ 是成对独立（pairwise independent）的，每个协方差（covariance）都是零，和的方差等于方差之和：

$$\mathrm{Var}(T) = N \cdot \frac{5}{36} = \frac{5n(n-1)}{72}.$$

**第 (ii) 部分**

设 $S = Y_1 + \cdots + Y_m$。由线性（linearity）得 $\mathrm{E}(S) = 0$，所以 $\mathrm{Var}(S) = \mathrm{E}(S^2) - 0^2 = \mathrm{E}(S^2)$。展开平方，每个 $Y_i^2$ 出现一次，而每个 $i < j$ 的乘积 $Y_i Y_j$ 出现两次：

$$S^2 = \sum_{i=1}^{m} Y_i^2 + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} Y_i Y_j.$$

取期望值（expectation）并利用线性，便得到所求的公式。

**第 (iii) 部分**

对两名球员而言，$\mathrm{P}(\text{掷出相同的偶数}) = \frac{3}{36} = \frac{1}{12}$，同样 $\mathrm{P}(\text{掷出相同的奇数}) = \frac{1}{12}$。所以

$$\mathrm{P}(Z_{ij} = 1) = \mathrm{P}(Z_{ij} = -1) = \tfrac{1}{12}, \qquad \mathrm{P}(Z_{ij} = 0) = \tfrac{5}{6}.$$

*不独立*。如果 $Z_{12} = 1$ 且 $Z_{23} = -1$，球员 2 的点数便要既是偶数又是奇数。所以

$$\mathrm{P}(Z_{12} = 1,\ Z_{23} = -1) = 0 \neq \tfrac{1}{12} \cdot \tfrac{1}{12} = \mathrm{P}(Z_{12} = 1)\,\mathrm{P}(Z_{23} = -1),$$

因此 $Z_{12}$ 与 $Z_{23}$ 并不独立。

*平均值*。由对称性，$\mathrm{E}(Z_{ij}) = \frac{1}{12} - \frac{1}{12} = 0$，所以总得分 $U = \sum_{i<j} Z_{ij}$ 的平均值为 0。

方差。每个 $Z_{ij}$ 的平均值都是零，所以可以对这 $N=\binom{n}{2}$ 个变量 $Z_{ij}$ 应用 (ii)。我们需要两类期望值。

- 平方项：$Z_{ij}^2$ 等于 1 当且仅当球员 $i$ 与 $j$ 点数相同，所以 $\mathrm{E}(Z_{ij}^2)=\frac{1}{6}$。
- 两个不同球员对的乘积。如果两对没有共同球员，两个变量独立且平均值为零，所以 $\mathrm{E}(Z_{ij}Z_{kl})=\mathrm{E}(Z_{ij})\,\mathrm{E}(Z_{kl})=0$。如果它们共有一名球员，例如 $Z_{12}Z_{23}$，那么只有当三名球员都掷出相同点数时乘积才非零。这时两个因子同为 $+1$（偶数）或同为 $-1$（奇数），所以乘积是 1。因此 $\mathrm{E}(Z_{12}Z_{23})=\frac{1}{36}$。

**想一想** 有多少个无序的“球员对之对”恰好共有一名球员?

先以 $n$ 种方法选出共有的球员，再从其余 $n-1$ 名球员中以 $\binom{n-1}{2}$ 种方法选出他的两个搭档。这样得到 $n\binom{n-1}{2}=\frac{1}{2}n(n-1)(n-2)$ 个这样的“球员对之对”，每个只计算一次，正是 (ii) 中双重求和所需要的。

综合起来，

$$\begin{aligned}\mathrm{Var}(U)&=\binom{n}{2}\cdot\frac{1}{6}+2\cdot\frac{n(n-1)(n-2)}{2}\cdot\frac{1}{36}\\&=\frac{n(n-1)}{36}\big(3+(n-2)\big)\\&=\frac{n(n-1)(n+1)}{36}=\frac{1}{36}n(n^2-1),\end{aligned}$$

得证。

## 验算

- 以例子中的掷骰结果 $3,3,3,3,6,6,2$ 为例，配对计数得 $6+1=7$，带符号计数得 $-6+1=-5$，与题目一致。
- 当 $n=2$ 时只有一对，所以 $\mathrm{Var}(U)=\mathrm{E}(Z_{12}^2)=\frac{1}{6}$，而公式给出 $\frac{2\cdot3}{36}=\frac{1}{6}$。
- 当 $n=3$ 时，若三颗骰子点数全部相同，则 $U=\pm3$（各自概率为 $\frac{1}{72}$）。否则恰好有一对点数相同，$U=\pm1$，总概率为 $\frac{15}{36}$。所以 $\mathrm{E}(U^2)=9\cdot\frac{1}{36}+\frac{15}{36}=\frac{2}{3}$，与 $\frac{3\cdot8}{36}$ 相符。以电脑精确列举 $n\le6$ 的所有结果，证实了两个平均值和两个方差；例如 $n=6$ 时，(i) 得 $\frac{25}{12}$，(iii) 得 $\frac{35}{6}$。

## 常见错误

- 只检查了一格便声称 $X_{12}$ 与 $X_{23}$ 独立，却没有说明为何其他格也随之成立。对于只取两个值的变量，其他格确实随之成立，但要把这一点说出来。

- 在 (iii) 中，忘记了共有一名球员的两对不论共同点数是偶数还是奇数，贡献都是 $+\frac{1}{36}$。负号在乘积中互相抵消。
- 数错有重叠的“球员对之对”的数目，例如用了 $n(n-1)(n-2)$（有序）之后再乘以 2。

## STEP 之外

第 (i) 部分是一个经典例子：变量成对独立，却并非相互独立（mutually independent）：若 $X_{12} = X_{23} = 1$，则必然 $X_{13} = 1$。计算方差只需要成对独立，这个想法应用于随机化算法（randomised algorithms）和生日问题（birthday problem）之中。第 (iii) 部分展示了当各项相依时，利用协方差计算和的方差的一般方法。两者在 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中均有详细阐述。

# 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

**2021 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。

