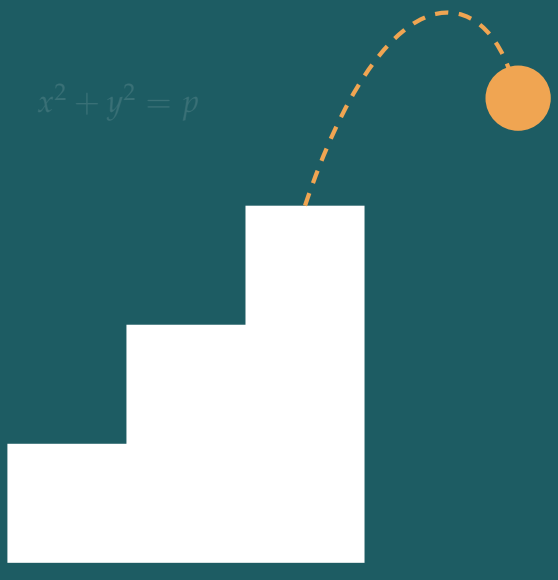
STEP 备试 · 大学数学

# Beyond STEP

## STEP I, II 及 III

2013 年详细解答



KEITH McNULTY

# Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP I, II 及 III

2013 年详细解答

Keith McNulty

**Beyond STEP：以历届 STEP 试题为门，走进大学数学**

**2013 年详细解答**

2013 年 STEP 试卷（STEP I, II 及 III）全部 39 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

# 目录

## STEP I


第 1 题 含平方根的方程 纯数学 3

第 2 题 下取整函数方程 纯数学 7

第 3 题 向量代数 纯数学 11

第 4 题 代换积分法 纯数学 15

第 5 题 圆锥曲线与旋转 纯数学 19

第 6 题 二项式系数与斐波那契数 纯数学 24

第 7 题 一阶微分方程 纯数学 28

第 8 题 复合函数 纯数学 32

第 9 题 抛体与碰撞 力学 36

第 10 题 恢复系数与摩擦力 力学 40

第 11 题 摩擦力与平衡 力学 44

第 12 题 不放回抽样 概率与统计 48

第 13 题 随数与期望值 概率与统计 52


## STEP II


第 1 题 对数曲线的弦 纯数学 55

第 2 题 归约公式 纯数学 59

第 3 题 三次方程的根 纯数学 63

第 4 题 弦中点的轨迹 纯数学 67

第 5 题 对称性与曲线描绘 纯数学 72

第 6 题 数列的收敛 纯数学 76

第 7 题 佩尔方程的解 纯数学 80

第 8 题 数积分与不等式 纯数学 83

第 9 题 平衡中的摩擦力 力学 87

第 10 题 抛体运动 力学 . . . . . 91
第 11 题 连续碰撞 力学 . . . . . 95
第 12 题 蓬松分布的矩 概率与统计 . . . . . 99
第 13 题 掷硬币中的游程 概率与统计 . . . . . 102

## STEP III

第 1 题 半角代换 纯数学 . . . . . 106
第 2 题 麦克劳林级数 纯数学 . . . . . 110
第 3 题 向量与球面 纯数学 . . . . . 113
第 4 题 单位根的乘积 纯数学 . . . . . 117
第 5 题 有理性与互质性 纯数学 . . . . . 121
第 6 题 复数模的不等式 纯数学 . . . . . 124
第 7 题 常微分方程的能量法 纯数学 . . . . . 128
第 8 题 圆锥曲线的极坐标方程 纯数学 . . . . . 131
第 9 题 浮力与简谐运动 力学 . . . . . 134
第 10 题 转动撞击 力学 . . . . . 137
第 11 题 弹簧与能量 力学 . . . . . 141
第 12 题 指示随机变量 概率与统计 . . . . . 145
第 13 题 矩与分布函数 概率与统计 . . . . . 149

详细解答

# STEP I 2013，第 1 题

纯数学 · 含平方根的方程

> **STEP I 2013，第 1 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s1-q1

## 关键思路

每条方程其实都是一条经过伪装的二次方程（quadratic equation）。把平方根（square root）记作 $y$，再把其余一切以 $y$ 表示，就得到一条关于 $y$ 的普通二次方程。唯一要留意的是平方根永不为负，所以 $y \geq 0$：每一次都正是这个条件把增根（spurious root）剔除。

## 解答

### 第 (i) 部分

设 $y = \sqrt{x}$，$y \geq 0$，则 $x = y^2$，方程变为

$$y^2 + 3y - \tfrac{1}{2} = 0 \quad \Longleftrightarrow \quad 2y^2 + 6y - 1 = 0.$$

由二次公式（quadratic formula），

$$y = \frac{-6 \pm \sqrt{36+8}}{4} = \frac{-3 \pm \sqrt{11}}{2}.$$

取负号会得到负值,而平方根不可能是负数。由于 $\sqrt{11} > 3$,取正号得 $y = \frac{1}{2}(\sqrt{11} - 3) > 0$，所以这是唯一的可能。于是

$$x = y^2 = \frac{11 - 6\sqrt{11} + 9}{4} = \frac{10 - 3\sqrt{11}}{2}.$$

一旦知道 $y \geq 0$，每一步都是可逆的（reversible），所以这确实是（唯一的）实根。

### 第 (ii)(a) 部分

这里开平方的是 $x+2$，所以设 $y=\sqrt{x+2}$，$y\geq 0$，得 $x=y^2-2$。方程变为

$$y^2-2+10y-22=0 \quad\Longleftrightarrow\quad y^2+10y-24=0 \quad\Longleftrightarrow\quad (y+12)(y-2)=0.$$

由于 $y\geq 0$，我们舍弃 $y=-12$，故 $y=2$，而 $x=2^2-2=2$。

验证：$2+10\sqrt{4}-22=0$。唯一的实根是 $x=2$。

> **想一想** 如果保留了 $y=-12$，会发生甚么事?

这会得到 $x=144-2=142$，而 $142+10\sqrt{144}-22=240\neq 0$。$y=-12$ 是相关方程 $x-10\sqrt{x+2}-22=0$ 的根；如果你粗心地把式子平方，你所解的其实就是这条方程。条件 $y\geq 0$ 正是把两者区分开来的关键。

### 第 (ii)(b) 部分

这次题目没有提示代换（substitution），所以我们要自己找一个。根号内的式子是

$$2x^2-8x-3=2(x^2-4x)-3,$$

而 $x^2-4x$ 亦出现在根号外。这就是两者之间的联系。

> **想一想** 这里 $y$ 应该取甚么：$x^2-4x$，还是整个平方根?

取 $y$ 为平方根，方程就会变成一条不含根号的二次方程。设 $y=\sqrt{2x^2-8x-3}$，其中 $y\geq 0$。则 $y^2=2(x^2-4x)-3$，所以

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

代入方程，得

$$\begin{aligned}\frac{y^2+3}{2}+y-9=0 \quad&\Longleftrightarrow\quad y^2+2y-15=0\\ &\Longleftrightarrow\quad (y+5)(y-3)=0.\end{aligned}$$

由于 $y\geq 0$，必须 $y=3$。于是

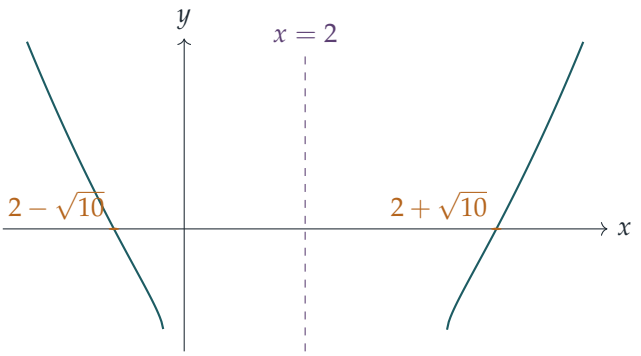
$$2x^2-8x-3=9 \quad\Longleftrightarrow\quad x^2-4x-6=0 \quad\Longleftrightarrow\quad x=2\pm\sqrt{10}.$$

这两个值都使 $2x^2-8x-3$ 等于 9，是正数，所以平方根有定义且等于 3。因此两者都是真正的根：实根为

$$x=2+\sqrt{10} \quad 及 \quad x=2-\sqrt{10}.$$

这两个根关于 $x = 2$ 对称（symmetric），这是理所当然的，因为方程只通过 $x^2 - 4x = (x-2)^2 - 4$ 依赖于 $x$。

左方的图像显示了这一点。它只在 $2x^2 - 8x - 3 \geq 0$ 时有定义，即 $x \leq 2 - \sqrt{5.5}$ 或 $x \geq 2 + \sqrt{5.5}$，而且在每一支上各穿过零一次。




## 验算

- 第 (i) 部分：$x = \frac{1}{2}(10 - 3\sqrt{11}) \approx 0.02506$，所以 $\sqrt{x} \approx 0.1583$，而 $x + 3\sqrt{x} \approx 0.02506 + 0.47494 = 0.5$。
- 第 (ii)(b) 部分：$x = 2 + \sqrt{10} \approx 5.1623$ 给出 $x^2 - 4x = 6$ 及 $2x^2 - 8x - 3 = 9$，所以左方为 $6 + 3 - 9 = 0$。被舍弃的值 $y = -5$ 需要 $x^2 - 4x = 14$，此时左方为 $14 + 5 - 9 = 10 \neq 0$。
- 在每个左方式子的整个定义域（domain）上作数值扫描，所找到的根恰好是上述各根，别无其他。

## 常见错误

- 保留了关于 $y$ 的二次方程的负根。平方根永不为负，所以那个根不会给出解。
- 在 (ii)(b) 一开始就把两边平方。这会得到一条四次方程（quartic equation），难度大得多，而且会引入额外的根，之后还得逐一检验。
- 在 (ii)(a) 忘记了 $x = y^2 - 2$，而不是 $y^2$。

## STEP 之外

引入新变量，把一条困难的方程化为熟悉的方程，是数学中最有用的习惯之一。在 (ii)(b) 中，方程只依赖于 $x^2 - 4x$，所以它是 $(x-2)^2$ 的函数。看出这类对称性（symmetry），并处理在该对称下保持不变的量，这种做法在伽罗瓦（Galois）理论（Galois theory）中会再次出现：在那里，人们通过找出根的、在对称变换下保持不变的表达式来解多项式方程（polynomial equation）。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是进入这些概念的一条平易途径。若想做更多这类题目，Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）非常出色（而且免费）。

详细解答

# STEP I 2013，第 2 题

纯数学 · 下取整函数方程

> **STEP I 2013，第 2 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s1-q2

## 关键思路

函数 $\lfloor x \rfloor$ 在每个区间（interval）$[k, k+1)$ 上都是常数，所以在该区间上 $f(x) = k/x$ 只是一段直角双曲线（rectangular hyperbola）。整个图像就是由这些曲线段串连而成，每个整数 $k$ 对应一段。于是解 $f(x) = c$ 便归结为：哪些曲线段能达到高度 $c$？在对应 $k$ 的那一段上，唯一的候选是 $x = k/c$，而它恰好在落入 $[k, k+1)$ 时才算数。

## 解答

### 第 (i) 部分

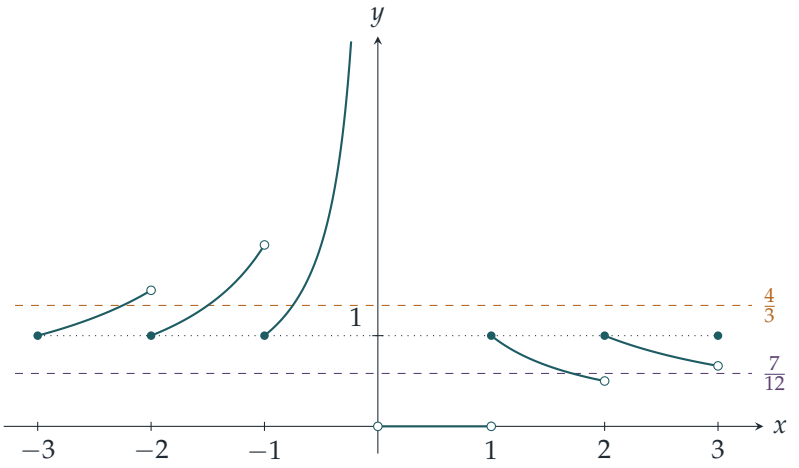
固定一个整数 $k$，考虑 $k \le x < k+1$（且 $x \neq 0$），此时 $f(x) = k/x$。

- $k = 0$，即 $0 < x < 1$：此时 $f(x) = 0$。
- $k \ge 1$：$f(x) = k/x$ 由 $f(k) = 1$ 递减（decreasing），趋向 $k/(k+1)$，但永远达不到。所以这一段的取值恰好是 $\left(\frac{k}{k+1}, 1\right]$。
- $k \le -1$：设 $k = -m$，其中 $m \ge 1$。此时 $x$ 和 $k$ 都是负数，所以 $f(x) = m/|x| > 0$。当 $x$ 由 $-m$ 增加至趋向 $-m+1$ 时，$|x|$ 递减，所以 $f$ 由 $f(-m) = 1$ 起递增。当 $m \ge 2$ 时，它趋向 $\frac{m}{m-1}$，但永远达不到，所以这一段的取值恰好是 $\left[1, \frac{m}{m-1}\right)$。当 $m = 1$ 时，当 $x \to 0^-$ 时 $f(x) = -1/x \to +\infty$，所以这一段的取值是 $[1, \infty)$。

在 $-3 \le x \le 3$ 范围内，各段如下：

| $x$ | $[-3,-2)$ | $[-2,-1)$ | $[-1,0)$ | $(0,1)$ | $[1,2)$ | $[2,3)$ |
|---|---|---|---|---|---|---|
| $f(x)$ | $-3/x$ | $-2/x$ | $-1/x$ | $0$ | $1/x$ | $2/x$ |
| 取值 | $[1, \frac{3}{2})$ | $[1,2)$ | $[1,\infty)$ | $0$ | $(\frac{1}{2}, 1]$ | $(\frac{2}{3}, 1]$ |

最后还有 $f(3)=1$。每一段都在其左端（整数端）从高度 1 开始，该点包含在内，而右端则是开端。




**第 (ii) 部分**

在 $\lfloor x\rfloor = k$ 的那一段上，方程 $f(x)=c$（其中 $c\neq 0$）即 $k/x=c$，所以 $x=k/c$。它恰好在 $k\le k/c<k+1$ 时才是真正的根（root）。因此对每个 $c$，我们只需利用第 (i) 部分得出的取值范围，判断哪些曲线段能达到高度 $c$。

$f(x)=\frac{7}{12}$。由于 $0<\frac{7}{12}<1$，只有 $k\ge 1$ 的曲线段可能有根，而且需要 $\frac{k}{k+1}<\frac{7}{12}$，即 $12k<7k+7$，即 $k<\frac{7}{5}$。所以只有 $k=1$ 可行，得 $x=1\big/\frac{7}{12}$：

$$x=\tfrac{12}{7}.$$

从图像上看，直线 $y=\frac{7}{12}$ 只与 $[1,2)$ 上的那一段相交，因为 $[2,3)$ 上的那一段一直高于 $\frac{2}{3}$。

$f(x)=\frac{17}{24}$。现在需要 $\frac{k}{k+1}<\frac{17}{24}$，即 $24k<17k+17$，即 $k<\frac{17}{7}$，所以 $k=1$ 或 $k=2$。根为 $x=k\big/\frac{17}{24}$：

$$x=\tfrac{24}{17}\quad\text{及}\quad x=\tfrac{48}{17}.$$

（确实 $2\le\frac{48}{17}\approx 2.82<3$。）

$f(x)=\frac{4}{3}$。由于 $\frac{4}{3}>1$，只有负数那边的曲线段可能有根。当 $\frac{4}{3}<\frac{m}{m-1}$（或 $m=1$）时，$k=-m$ 的那一段能达到 $\frac{4}{3}$，即 $4m-4<3m$，即 $m<4$。所以 $m=1,2,3$，根为 $x=-m\big/\frac{4}{3}=-\frac{3m}{4}$：

$$x=-\tfrac{3}{4},\quad x=-\tfrac{3}{2},\quad x=-\tfrac{9}{4}.$$

$m=4$ 的情况刚好不行：$[-4,-3)$ 上的那一段只取小于 $\frac{4}{3}$ 的值。

### 第 (iii) 部分

这里 $0 < \frac{9}{10} < 1$，所以根来自 $k \geq 1$ 且 $\frac{k}{k+1} < \frac{9}{10}$ 的曲线段，即 $10k < 9k + 9$，即 $k < 9$。所以 $k = 1, 2, \ldots, 8$ 各给出一个根 $x = \frac{10k}{9}$。最大的根对应 $k = 8$:

$$x = \tfrac{80}{9} \approx 8.89,$$

它确实落在 $[8, 9)$ 内。

### 最后部分

首先找出没有根或有无限多个根的 $c$ 值。

- $c < 0$: 没有根，因为当 $x > 0$ 时 $f(x) \geq 0$，而当 $x < 0$ 时 $f(x) > 0$。
- $c = 0$: 每个 $x \in (0, 1)$ 都是根，所以有无限多个根。
- $c = 1$: 每个非零整数都是根，所以同样有无限多个根。

> **想一想** 对于一般的 $c$，不画图的话，怎样数根的个数?

利用第 (i) 部分得出的取值范围: 每一段最多贡献一个根，所以只需数一数有多少段的取值范围包含 $c$。

**情况** $0 < c < 1$。 能达到 $c$ 的曲线段是 $k \geq 1$ 且 $\frac{k}{k+1} < c$ 的那些。由于 $\frac{k}{k+1}$ 随 $k$ 递增，这些段就是 $k = 1, 2, \ldots$，直至最后一个满足 $\frac{k}{k+1} < c$ 的 $k$。所以恰好有 $n$ 个根，当且仅当第 $n$ 段能达到 $c$ 而第 $(n+1)$ 段不能:

$$\frac{n}{n+1} < c \leq \frac{n+1}{n+2}.$$

(当 $0 < c \leq \frac{1}{2}$ 时，没有任何一段能达到 $c$，因此没有根。)

**情况** $c > 1$。 能达到 $c$ 的曲线段是 $k = -m$ 的那些，其中 $m = 1$ 或 $\frac{m}{m-1} > c$。由于 $\frac{m}{m-1} = 1 + \frac{1}{m-1}$ 随 $m$ 增加而递减，这些段就是 $m = 1, 2, \ldots$，直至最后一个满足 $\frac{m}{m-1} > c$ 的 $m$。所以恰好有 $n$ 个根，当且仅当 $m = n$ 的那一段能达到 $c$ 而 $m = n + 1$ 的那一段不能:

$$\frac{n+1}{n} \leq c < \frac{n}{n-1} \qquad (n \geq 2),$$

而当 $n = 1$ 时, 条件就是 $c \geq 2$ ($[-1, 0)$ 上的那一段必定能达到 $c$, 而当 $c \geq 2$ 时, $[-2, -1)$ 上的那一段不能)。

**答案**。 对 $n \geq 1$，方程 $f(x) = c$ 恰好有 $n$ 个根，当且仅当

$$\frac{n}{n+1} < c \leq \frac{n+1}{n+2} \qquad \text{或} \qquad \frac{n+1}{n} \leq c < \frac{n}{n-1},$$

其中当 $n = 1$ 时，第二个条件理解为 $c \geq 2$。留意等号放在哪里。两个等号都源自曲线段的开端：$c = \frac{n+1}{n+2}$ 不能由 $[n+1, n+2)$ 上的那一段达到，而 $c = \frac{n+1}{n}$ 不能由 $[-(n+1), -n)$ 上的那一段达到。所以这两个边界值仍然恰好给出 $n$ 个根。

## 验算

- 代回原式：$\lfloor \frac{12}{7} \rfloor = 1$ 且 $1/\frac{12}{7} = \frac{7}{12}$；$\lfloor \frac{48}{17} \rfloor = 2$ 且 $2/\frac{48}{17} = \frac{17}{24}$；$\lfloor -\frac{9}{4} \rfloor = -3$ 且 $-3/(-\frac{9}{4}) = \frac{4}{3}$；$\lfloor \frac{80}{9} \rfloor = 8$ 且 $8/\frac{80}{9} = \frac{9}{10}$。
- 一般答案与个别情况吻合：$\frac{2}{3} < \frac{17}{24} \leq \frac{3}{4}$ 给出 $n = 2$；$\frac{4}{3} \leq \frac{4}{3} < \frac{3}{2}$ 给出 $n = 3$；$\frac{8}{9} < \frac{9}{10} \leq \frac{9}{10}$ 给出 $n = 8$。后两个刚好落在范围的端点上，正好检验 $\leq$ 和 $<$ 的位置是否正确。
- 一个边界值：$c = 2$ 应该给出一个根。确实 $x = -\frac{1}{2}$ 可行，而 $x = -1$（来自 $k = -2$）有 $\lfloor -1 \rfloor = -1 \neq -2$。
- 试试 $c = \frac{1}{2}$：条件不给出任何 $n$；而确实对 $k \neq 0$，$x = k/c = 2k$ 永远不满足 $\lfloor 2k \rfloor = k$。

## 常见错误

- 忘记对负数 $x$ 而言，$\lfloor x \rfloor$ 是向下取整，所以 $\lfloor -2.25 \rfloor = -3$，而不是 $-2$。这样会完全漏掉 $c > 1$ 时的负数分支。
- 端点搞错：每一段包含其左端（整数端，高度为 1），但不包含右端。这决定了最终条件中出现的是 $\leq$ 还是 $<$。
- 漏掉 $c = 0$ 和 $c = 1$（有无限多个根），或漏掉 $c < 0$ 和 $0 < c \leq \frac{1}{2}$（没有根）。

## STEP 之外

像 $\lfloor x \rfloor / x$ 这样的函数，是“分段连续（continuous）但在孤立点上跳跃”的函数中最简单的例子。小心分辨哪些端点能够取到，正是大学第一个数学分析（analysis）课程处理上确界（supremum）和下确界（infimum）时所需的习惯。在这里，$f$ 在 $x > 0$ 上的 $\sup f$ 是 1，而且能取到；而在 $x \geq 1$ 上的下确界是 $\frac{1}{2}$，却取不到。同一个函数还说明了当 $x \to \infty$ 时 $\lfloor x \rfloor / x \to 1$，让你初尝解析数论（analytic number theory）中所用的估计方法。若想进一步了解上确界、极限（limit）和连续性，Abbott 的 *Understanding Analysis*（《理解分析》）是很好的读物。

详细解答

# STEP I 2013，第 3 题

纯数学 · 向量代数

> **STEP I 2013，第 3 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s1-q3

## 关键思路

这里的一切都是线性（linear）的，所以每次最干净利落的做法，都是把每一点写成 $\mathbf{x}$、$\mathbf{y}$ 和 $\mathbf{z}$ 的组合，然后相减。两点相异，当且仅当它们的位置向量（position vector）之差不为零；而这个差总会化成一个简单的因子乘以像 $\mathbf{x}-\mathbf{y}$ 这样的向量。从几何上看，$X*Y$ 是直线 $XY$ 上的一点，而当 $0<\lambda<1$ 时，它位于两点之间的线段（segment）上。

## 解答

在整个解答中，两点相同当且仅当它们的位置向量相等。

### 第 (i) 部分

按定义，$Y*X$ 的位置向量是 $\lambda\mathbf{y}+(1-\lambda)\mathbf{x}$，所以

$$\begin{aligned}\big(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\big)-\big(\lambda\mathbf{y}+(1-\lambda)\mathbf{x}\big) &= (2\lambda-1)\mathbf{x}-(2\lambda-1)\mathbf{y}\\ &= (2\lambda-1)(\mathbf{x}-\mathbf{y}).\end{aligned}$$

由于 $X$ 与 $Y$ 相异，$\mathbf{x}-\mathbf{y}\neq\mathbf{0}$，因此这个差为零当且仅当 $2\lambda-1=0$。故此，除非 $\lambda=\frac{1}{2}$（此时两者都是 $XY$ 的中点），否则 $X*Y$ 与 $Y*X$ 相异。

### 第 (ii) 部分

把两边分别展开：

$$\begin{aligned}(X*Y)*Z:&\quad \lambda\big(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\big)+(1-\lambda)\mathbf{z}=\lambda^2\mathbf{x}+\lambda(1-\lambda)\mathbf{y}+(1-\lambda)\mathbf{z},\\ X*(Y*Z):&\quad \lambda\mathbf{x}+(1-\lambda)\big(\lambda\mathbf{y}+(1-\lambda)\mathbf{z}\big)=\lambda\mathbf{x}+\lambda(1-\lambda)\mathbf{y}+(1-\lambda)^2\mathbf{z}.\end{aligned}$$

两者的 $\mathbf{y}$ 项相同，相减得

$$\begin{aligned}\text{差} &= (\lambda^2-\lambda)\mathbf{x} + \big((1-\lambda)-(1-\lambda)^2\big)\mathbf{z}\\ &= -\lambda(1-\lambda)\mathbf{x} + \lambda(1-\lambda)\mathbf{z}\\ &= \lambda(1-\lambda)(\mathbf{z}-\mathbf{x}).\end{aligned}$$

这个差不为零，当且仅当每个因子都不为零。所以 $(X*Y)*Z$ 与 $X*(Y*Z)$ 相异，当且仅当

$$\lambda \neq 0, \qquad \lambda \neq 1 \qquad \text{且} \qquad Z \neq X.$$

留意 $Y$ 完全不起作用。被排除的值也合情合理：当 $\lambda = 1$ 时，这个运算总是给出第一点；当 $\lambda = 0$ 时，总是给出第二点，所以两种加括号方式分别都给出 $X$ 或 $Z$。

**第 (iii) 部分**

为求简洁，记 $\mu = 1-\lambda$。右方为

$$\begin{aligned}(X*Z)*(Y*Z): \quad &\lambda(\lambda\mathbf{x}+\mu\mathbf{z}) + \mu(\lambda\mathbf{y}+\mu\mathbf{z})\\ &= \lambda^2\mathbf{x} + \lambda\mu\,\mathbf{y} + \mu(\lambda+\mu)\mathbf{z}\\ &= \lambda^2\mathbf{x} + \lambda\mu\,\mathbf{y} + \mu\,\mathbf{z},\end{aligned}$$

因为 $\lambda+\mu = 1$。这正是第 (ii) 部分中 $(X*Y)*Z$ 的展开式，所以对所有点 $X$、$Y$、$Z$，

$$(X*Y)*Z = (X*Z)*(Y*Z)$$

得证。

> **想一想** $X*(Y*Z)$ 的“对应结果”应该是怎样的?

在第 (iii) 部分中，外层的 $*Z$ 分配到左边的括号上。其镜像就是把 $X*$ 分配到右边的括号上，这提示我们试试 $(X*Y)*(X*Z)$。果然，

$$\begin{aligned}(X*Y)*(X*Z): \quad &\lambda(\lambda\mathbf{x}+\mu\mathbf{y}) + \mu(\lambda\mathbf{x}+\mu\mathbf{z})\\ &= \lambda(\lambda+\mu)\mathbf{x} + \lambda\mu\,\mathbf{y} + \mu^2\mathbf{z}\\ &= \lambda\mathbf{x} + \lambda\mu\,\mathbf{y} + \mu^2\mathbf{z},\end{aligned}$$

这正是第 (ii) 部分中 $X*(Y*Z)$ 的展开式。所以对任意点 $X$、$Y$、$Z$，

$$X*(Y*Z) = (X*Y)*(X*Z).$$

**第 (iv) 部分**

我们断言 $P_n$ 的位置向量为

$$\mathbf{p}_n = \lambda^n\mathbf{x} + (1-\lambda^n)\mathbf{y}.$$

当 $n=1$ 时，这就是 $X*Y$ 的定义。若它对 $n-1$ 成立，则

$$\begin{aligned}\mathbf{p}_n &= \lambda\mathbf{p}_{n-1}+(1-\lambda)\mathbf{y} = \lambda^n\mathbf{x}+\big(\lambda-\lambda^n+1-\lambda\big)\mathbf{y}\\ &= \lambda^n\mathbf{x}+(1-\lambda^n)\mathbf{y},\end{aligned}$$

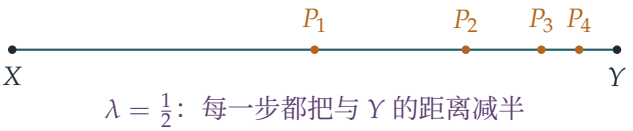
所以由归纳法（induction），此公式对所有 $n\geq 1$ 成立。

现在沿线段量度：

$$\mathbf{p}_n-\mathbf{x}=(1-\lambda^n)(\mathbf{y}-\mathbf{x}),\qquad \mathbf{y}-\mathbf{p}_n=\lambda^n(\mathbf{y}-\mathbf{x}).$$

两者都是同一个非零向量 $\mathbf{y}-\mathbf{x}$ 的正倍数，因为 $0<\lambda<1$ 给出 $0<\lambda^n<1$。所以 $P_n$ 严格位于 $X$ 与 $Y$ 之间，而且

$$XP_n : P_nY = (1-\lambda^n) : \lambda^n.$$




$\lambda=\frac{1}{2}$：每一步都把与 $Y$ 的距离减半

每一步都把余下与 $Y$ 的距离乘以 $\lambda$，所以这些点一步步向 $Y$ 推进，并且当 $n\to\infty$ 时 $P_n\to Y$。

## 验算

- 在第 (i) 部分，沿一条直线取 $\mathbf{x}=\mathbf{0}$、$\mathbf{y}=\mathbf{1}$ 及 $\lambda=\frac{1}{3}$：$X*Y$ 位于 $\frac{2}{3}$，$Y*X$ 位于 $\frac{1}{3}$，两者相差 $\frac{1}{3}=|2\lambda-1|$。
- 在第 (ii) 部分，当 $Z=X$ 时，无论 $\lambda$ 取何值，两种加括号方式都应相同；此时我们的差 $\lambda(1-\lambda)(\mathbf{z}-\mathbf{x})$ 确实为零，正如所料。
- 在第 (iv) 部分，取 $\lambda=\frac{1}{2}$ 及 $n=3$，比为 $\frac{7}{8}:\frac{1}{8}=7:1$，而 $P_3$ 确实位于由 $X$ 到 $Y$ 的 $\frac{7}{8}$ 处（距离减半三次）。取 $n=1$ 时，比 $(1-\lambda):\lambda$ 正是 $X*Y$ 的定义。

## 常见错误

- 在第 (ii) 部分忘记 $Z\neq X$：答案不只是关于 $\lambda$ 的条件。
- 在第 (iv) 部分把比的次序弄反。由于 $\mathbf{p}_n$ 在 $\mathbf{x}$ 上的权重是 $\lambda^n$，当 $\lambda^n$ 较大时它靠近 $X$，所以较短的一段 $XP_n$ 对应的是 $1-\lambda^n$。

## STEP 之外

点 $\lambda\mathbf{x}+(1-\lambda)\mathbf{y}$ 是一个凸组合(convex combination,或称仿射组合(affine combination)),它是凸集（convex set)、重心坐标（barycentric coordinates）以及电脑图像中线性插值（linear interpolation）的基本构件。第 (ii) 部分说明 $*$ 不满足结合律（associative），而第 (iii) 部分则说明它满足自分配律（self-distributive)；具有这种性质的运算在代数中以架（rack)、环结（quandle）和中介代数（medial algebra）等名称来研究。若想从线性代数的角度理解，可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》)。

详细解答

# STEP I 2013，第 4 题

纯数学 · 代换积分法

**STEP I 2013，第 4 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q4

## 关键思路

在第 (i) 部分，每个被积函数（integrand）都是“某个式子的函数，乘以该式子的导数（derivative）”：$\sec^2 x$ 是 $\tan x$ 的导数，而 $\sec x \tan x$ 是 $\sec x$ 的导数。在第 (ii) 部分，多出来的因子 $x$ 或 $x^2$ 提示我们要用分部积分法（integration by parts）。我们要积分的是三角函数那个因子，而第 (i) 部分刚好示范了该怎样积。

## 解答

### 第 (i) 部分

对第一个积分，作代换（substitution）$u = \tan x$，于是 $\mathrm{d}u = \sec^2 x \,\mathrm{d}x$。当 $x = 0$ 时 $u = 0$，当 $x = \frac{1}{4}\pi$ 时 $u = 1$。因此

$$\int_0^{\frac{1}{4}\pi} \tan^n x \, \sec^2 x \,\mathrm{d}x = \int_0^1 u^n \,\mathrm{d}u = \left[\frac{u^{n+1}}{n+1}\right]_0^1 = \frac{1}{n+1},$$

这里用了 $n > 0$，所以 $0^{n+1} = 0$。

? **想一想** 在第二个积分中，$\sec^n x \tan x$ 里面藏着哪个函数的导数?

由于 $\dfrac{\mathrm{d}}{\mathrm{d}x} \sec x = \sec x \tan x$，我们先分出一个 $\sec x$ 因子：

$$\sec^n x \tan x = \sec^{n-1} x \,(\sec x \tan x).$$

现在代换 $u = \sec x$，于是 $\mathrm{d}u = \sec x \tan x \,\mathrm{d}x$，而 $x = 0$ 时 $u = 1$，$x = \frac{1}{4}\pi$ 时 $u = \sqrt{2}$。

由于 $n > 0$，我们可以除以 $n$，得

$$\int_0^{\frac{1}{4}\pi} \sec^n x \tan x \,\mathrm{d}x = \int_1^{\sqrt{2}} u^{n-1} \,\mathrm{d}u = \left[\frac{u^n}{n}\right]_1^{\sqrt{2}} = \frac{(\sqrt{2})^n - 1}{n},$$

即所求。

**第 (ii) 部分**

**第一个积分**。记之为 $I_1$。我们用分部积分法，对 $x$ 求导，并对 $\sec^4 x \tan x$ 积分。用第 (i) 部分的代换 $u = \sec x$，可得原函数（antiderivative）

$$\int \sec^4 x \tan x \,\mathrm{d}x = \tfrac{1}{4} \sec^4 x + C.$$

（它在 $[0, \frac{1}{4}\pi]$ 上的定积分（definite integral）是 $\frac{1}{4}(4-1) = \frac{3}{4}$，即第 (i) 部分中 $n = 4$ 的情况。）所以

$$I_1 = \left[\tfrac{1}{4} x \sec^4 x\right]_0^{\frac{1}{4}\pi} - \frac{1}{4}\int_0^{\frac{1}{4}\pi} \sec^4 x \,\mathrm{d}x = \frac{\pi}{4}\cdot\frac{4}{4} - \frac{1}{4}\int_0^{\frac{1}{4}\pi} \sec^4 x \,\mathrm{d}x,$$

因为 $\sec^4 \frac{1}{4}\pi = (\sqrt{2})^4 = 4$。对余下的积分，写 $\sec^2 x = 1 + \tan^2 x$：

$$\int_0^{\frac{1}{4}\pi} \sec^4 x \,\mathrm{d}x = \int_0^{\frac{1}{4}\pi} \sec^2 x \,\mathrm{d}x + \int_0^{\frac{1}{4}\pi} \tan^2 x \sec^2 x \,\mathrm{d}x = 1 + \frac{1}{3} = \frac{4}{3},$$

其中第一个积分是 $[\tan x]_0^{\frac{1}{4}\pi} = 1$，第二个是第 (i) 部分中 $n = 2$ 的情况。因此

$$\int_0^{\frac{1}{4}\pi} x \sec^4 x \tan x \,\mathrm{d}x = \frac{\pi}{4} - \frac{1}{3}.$$

**第二个积分**。记之为 $I_2$。同样地，我们对 $x$ 的幂求导，并对 $\sec^2 x \tan x$ 积分。其中一个原函数是 $\frac{1}{2}\sec^2 x$，但 $\frac{1}{2}\tan^2 x$ 同样可用（两者相差常数 $\frac{1}{2}$），而且是更好的选择。

> ? **想一想** 为甚么 $\frac{1}{2}\tan^2 x$ 比 $\frac{1}{2}\sec^2 x$ 更好?

分部积分之后，我们会剩下 $x\tan^2 x$，而 $\tan^2 x = \sec^2 x - 1$ 可以拆成我们会积分的几部分。此外，$\tan^2 x$ 在 0 处为零，令边界项保持简洁。（用 $\frac{1}{2}\sec^2 x$ 最终也会得到相同答案。）

所以，由 $\tan \frac{1}{4}\pi = 1$，

$$I_2 = \left[\tfrac{1}{2} x^2 \tan^2 x\right]_0^{\frac{1}{4}\pi} - \int_0^{\frac{1}{4}\pi} x \tan^2 x \,\mathrm{d}x = \frac{\pi^2}{32} - \int_0^{\frac{1}{4}\pi} x \tan^2 x \,\mathrm{d}x.$$

现在

$$\int_0^{\frac{1}{4}\pi} x \tan^2 x \,\mathrm{d}x = \int_0^{\frac{1}{4}\pi} x \sec^2 x \,\mathrm{d}x - \int_0^{\frac{1}{4}\pi} x \,\mathrm{d}x = \int_0^{\frac{1}{4}\pi} x \sec^2 x \,\mathrm{d}x - \frac{\pi^2}{32}.$$

再用一次分部积分法，以 $\tan x$ 作为 $\sec^2 x$ 的原函数，并利用 $\int \tan x \,\mathrm{d}x = \ln \sec x$（由于在此区间上 $\sec x > 0$，此式成立）：

$$\begin{aligned}\int_0^{\frac{1}{4}\pi} x \sec^2 x \,\mathrm{d}x &= \Big[x \tan x\Big]_0^{\frac{1}{4}\pi} - \int_0^{\frac{1}{4}\pi} \tan x \,\mathrm{d}x \\ &= \frac{\pi}{4} - \Big[\ln \sec x\Big]_0^{\frac{1}{4}\pi} = \frac{\pi}{4} - \ln\sqrt{2} = \frac{\pi}{4} - \frac{1}{2}\ln 2.\end{aligned}$$

把各部分合起来，

$$I_2 = \frac{\pi^2}{32} - \left(\frac{\pi}{4} - \frac{1}{2}\ln 2 - \frac{\pi^2}{32}\right),$$

即

$$\int_0^{\frac{1}{4}\pi} x^2 \sec^2 x \tan x \,\mathrm{d}x = \frac{\pi^2}{16} - \frac{\pi}{4} + \frac{1}{2}\ln 2.$$

## 验算

- 第 (i) 部分取 $n = 1$：两条公式分别给出 $\int_0^{\frac{1}{4}\pi} \tan x \sec^2 x \,\mathrm{d}x = \frac{1}{2}$ 和 $\int_0^{\frac{1}{4}\pi} \sec x \tan x \,\mathrm{d}x = \sqrt{2} - 1$，可以直接由 $[\frac{1}{2}\tan^2 x]$ 和 $[\sec x]$ 验证。
- 正负与大小：第 (ii) 部分的两个被积函数在 $(0, \frac{1}{4}\pi)$ 上都是正数，答案 $\frac{\pi}{4} - \frac{1}{3} \approx 0.4521$ 和 $\frac{\pi^2}{16} - \frac{\pi}{4} + \frac{1}{2}\ln 2 \approx 0.1780$ 亦然。第二个较小，理应如此，因为在该区间上 $x^2 \sec^2 x \le x \sec^4 x$（因为 $x < 1 < \sec^2 x$）。
- 粗略上界：$x \sec^4 x \tan x \le \frac{\pi}{4} \sec^4 x \tan x$，所以 $I_1 \le \frac{\pi}{4} \cdot \frac{3}{4} \approx 0.589$。我们得到的 0.452 与此相符。用计算机以辛普森法则（Simpson's rule）分几个小区间计算，结果与两个值都很接近。

## 常见错误

- 在第 (i) 部分的第二个积分中，未先分出 $\sec x \tan x$ 就代换 $u = \sec x$，结果幂次留在 $n$ 而非 $n - 1$。
- 忘记 $\sec \frac{1}{4}\pi = \sqrt{2}$，因此 $\sec^4 \frac{1}{4}\pi = 4$ 而 $\sec^2 \frac{1}{4}\pi = 2$。
- 漏掉边界项（boundary term）$[\frac{1}{2}x^2 \tan^2 x]$，或把 $\int \tan x \,\mathrm{d}x$ 的正负号弄错（它是 $\ln \sec x = -\ln \cos x$）。

## STEP 之外

第 (i) 部分的积分是积分 $\sec^m x \tan^n x$ 这套标准理论的起步，其中递推公式（reduction formulae）让你逐步降低幂次。像第 (ii) 部分那样带多项式因子的分部积分，可推广为

“表格法”（tabular method），而在更后的阶段，更发展成把导数从一个函数转移到另一个函数的想法，这正是微分方程弱解（weak solutions）的基础。若想仔细了解代换法和分部积分法为何成立，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）；若想做更多同类型的题目，可试试 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）。

详细解答

# STEP I 2013，第 5 题

纯数学 · 圆锥曲线与旋转

**STEP I 2013，第 5 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q5

## 关键思路

唯一随 $k$ 改变的是 $xy$ 项，而它决定了曲线的类型。若没有 $xy$ 项，便可以对 $x$ 和 $y$ 分别配方（completing the square）。若二次部分可以因式分解（factorise），整条方程有可能分解成两条直线。而若二次部分是完全平方（perfect square）$(x+y)^2$，把坐标轴旋转（rotate）$45°$ 就能把 $x+y$ 变成单一个坐标，剩下一条熟悉的抛物线（parabola）。

另外留意，设 $x=0$ 或 $y=0$ 都会令 $xy$ 项消失，所以对任何 $k$，与坐标轴的截距（intercept）都相同：$x^2+3x=0$ 给出 $x=0$ 或 $x=-3$，而 $y^2+y=0$ 给出 $y=0$ 或 $y=-1$。每一条轨迹（locus）都经过 $(0,0)$、$(-3,0)$ 和 $(0,-1)$。

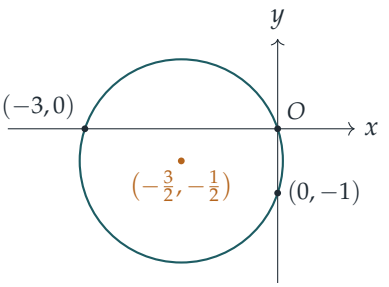
## 解答

### 第 (i) 部分

当 $k=0$ 时，对每个变量配方，得

$$\left(x+\tfrac{3}{2}\right)^2+\left(y+\tfrac{1}{2}\right)^2=\tfrac{9}{4}+\tfrac{1}{4}=\tfrac{5}{2}.$$

所以轨迹是圆心为 $\left(-\frac{3}{2},-\frac{1}{2}\right)$、半径为 $\sqrt{5/2}\approx 1.58$ 的圆。它与 $x$ 轴相交于 $(0,0)$ 和 $(-3,0)$，与 $y$ 轴相交于 $(0,0)$ 和 $(0,-1)$。（原点和 $(-3,0)$ 是一条弦（chord）的两个端点，所以圆心位于该弦的垂直平分线（perpendicular bisector）$x=-\frac{3}{2}$ 上，理应如此。）




**第 (ii) 部分**

当 $k=\frac{10}{3}$ 时，两边乘以 3：

$$3x^2+3y^2+10xy+9x+3y=0.$$

二次部分可因式分解为

$$3x^2+10xy+3y^2=(3x+y)(x+3y),$$

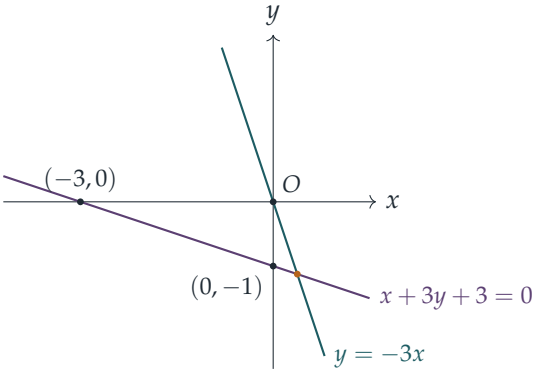
而一次部分是 $9x+3y=3(3x+y)$，同样含有因式 $3x+y$。所以

$$(3x+y)(x+3y)+3(3x+y)=(3x+y)(x+3y+3)=0.$$

乘积为零当且仅当其中一个因式为零，所以轨迹是一对直线

$$y=-3x \qquad \text{和} \qquad x+3y+3=0.$$

第一条直线经过原点（这是它与两条坐标轴唯一的交点）。第二条直线与 $x$ 轴相交于 $(-3,0)$，与 $y$ 轴相交于 $(0,-1)$。联立两条方程，得 $x-9x+3=0$，所以两线相交于 $\left(\frac{3}{8},-\frac{9}{8}\right)$。



**第 (iii) 部分**

当 $k=2$ 时，方程为

$$(x+y)^2+3x+y=0.$$

当 $\theta=45^\circ$ 时，$\cos\theta=\sin\theta=\frac{1}{\sqrt{2}}$，所以

$$X=\frac{x+y}{\sqrt{2}},\qquad Y=\frac{y-x}{\sqrt{2}}.$$

因此 $x+y=\sqrt{2}\,X$ 及 $y-x=\sqrt{2}\,Y$，把两式相加和相减，得

$$x=\frac{X-Y}{\sqrt{2}},\qquad y=\frac{X+Y}{\sqrt{2}}.$$

? **想一想** 为甚么我们需要以 $X$ 和 $Y$ 表示 $x$ 和 $y$，而不是反过来?

我们知道 $P$ 在曲线上，而要求的是 $Q$ 所满足的方程。所以必须把 $x$ 和 $y$ 代入 $P$ 的轨迹方程，这就需要以 $X$ 和 $Y$ 来表示它们。（由于这个映射是旋转，其逆映射（inverse）就只是反方向的旋转，所以这一步很容易。）

一次部分变成

$$3x+y=\frac{3X-3Y+X+Y}{\sqrt{2}}=\frac{4X-2Y}{\sqrt{2}}=2\sqrt{2}\,X-\sqrt{2}\,Y,$$

而 $(x+y)^2=2X^2$。所以 $Q$ 的轨迹是

$$2X^2+2\sqrt{2}\,X-\sqrt{2}\,Y=0\quad\Longleftrightarrow\quad\sqrt{2}\,Y=2X^2+2\sqrt{2}\,X=\left(\sqrt{2}\,X+1\right)^2-1,$$

正如所求。每一步都是可逆的，所以这正正就是 $Q$ 的轨迹。

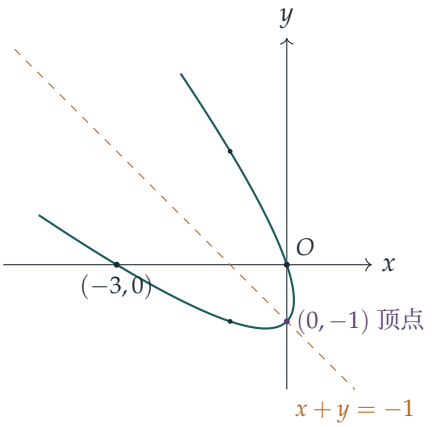
这是 $(X,Y)$ 平面上的一条抛物线，开口朝 $Y$ 增加的方向。它关于 $\sqrt{2}\,X+1=0$ 即 $X=-\frac{1}{\sqrt{2}}$ 这条直线对称，顶点（vertex）位于 $X=-\frac{1}{\sqrt{2}}$，$Y=-\frac{1}{\sqrt{2}}$。

现在把旋转还原。$Q$ 是由 $P$ 顺时针旋转 $45^\circ$ 得到的，所以 $P$ 的轨迹就是把这条抛物线逆时针旋转 $45^\circ$。以 $x$ 和 $y$ 表示：

- 对称轴（line of symmetry）$X=-\frac{1}{\sqrt{2}}$ 即 $\frac{x+y}{\sqrt{2}}=-\frac{1}{\sqrt{2}}$，也就是
$$x+y=-1;$$
- 顶点位于 $x=\frac{X-Y}{\sqrt{2}}=0$，$y=\frac{X+Y}{\sqrt{2}}=-1$，即 $(0,-1)$；
- 抛物线开口朝 $Y=\frac{y-x}{\sqrt{2}}$ 增加的方向，即 $(-1,1)$ 的方向：向左上方。

截距是 $(0,0)$、$(-3,0)$ 和 $(0,-1)$，正如开首所述；其中 $(0,-1)$ 是顶点。该处的切线（tangent）垂直于对称轴，所以斜率（gradient）为 1：曲线在顶点处穿过 $y$ 轴，而不是与之相切。作为对称性的验证，关于 $x+y=-1$ 的反射（reflection）把 $(a,b)$ 映到 $(-1-b,-1-a)$，所以原点映到 $(-1,-1)$，而 $(-3,0)$ 映到 $(-1,2)$；两点都满足 $(x+y)^2+3x+y=0$。

绘图时，设 $u=x+y$ 会有帮助：这样 $3x+y=-u^2$，解得 $x=-\frac{1}{2}(u^2+u)$，$y=\frac{1}{2}(u^2+3u)$。




$P$ 的轨迹是一条顶点为 $(0,-1)$、对称轴为 $x+y=-1$ 的抛物线。

## 验算

- 把截距点 $(0,0)$、$(-3,0)$、$(0,-1)$ 代入全部三条方程：无论 $k$ 为何值，结果都是 0。
- 在 (ii) 中，展开 $(3x+y)(x+3y+3)$ 再除以 3，便得回 $k=\frac{10}{3}$ 时的原方程。交点 $\left(\frac{3}{8},-\frac{9}{8}\right)$ 同时满足 $3x+y=0$ 和 $x+3y+3=0$。
- 在 (iii) 中，取 $u=1$：点 $(-1,2)$ 在 $P$ 的轨迹上。于是 $X=\frac{1}{\sqrt{2}}$，$Y=\frac{3}{\sqrt{2}}$，而 $\sqrt{2}\,Y=3=(1+1)^2-1$。

## 常见错误

- 在 (iii) 中，把 $X,Y$ 的公式代入方程，而不是以 $X,Y$ 表示 $x,y$。
- 把抛物线向错误方向旋转回去，得出对称轴为 $y-x=-1$，而非 $x+y=-1$。
- 在 (ii) 中，没有发现一次项含有公因式 $3x+y$，因而把轨迹当作其他类型的圆锥曲线。

## STEP 之外

每条方程 $ax^2 + bxy + cy^2 + dx + ey + f = 0$ 都是一条圆锥曲线（conic），其类型由二次部分的判别式（discriminant）$b^2 - 4ac$ 决定：负数时是椭圆（ellipse）（这里 $k = 0$ 时是圆），零时是抛物线（$k = 2$），正数时是双曲线（hyperbola），而双曲线可以退化（degenerate）成一对直线（$k = \frac{10}{3}$）。在大学一年级的线性代数（linear algebra）课程中，你会学到如何有系统地消去 $xy$ 项：把坐标轴旋转到二次型（quadratic form）所对应的对称矩阵（symmetric matrix）的特征向量（eigenvectors）方向。斯特朗（Strang）的 *Introduction to Linear Algebra*（《线性代数导论》）对此有清晰的解释。

详细解答

# STEP I 2013，第 6 题

纯数学 · 二项式系数与斐波那契数

**STEP I 2013，第 6 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q6

## 关键思路

帕斯卡法则（Pascal's rule）就是整道题的引擎。每个 $B_n$ 都是若干二项式系数（binomial coefficients）之和；若用帕斯卡法则把 $B_{n+2}$ 的每一项拆开，拆出来的各部分正好可以重新组合成 $B_{n+1}$ 和 $B_n$ 的各项。唯一要小心的是和式的两端：那里帕斯卡法则并不适用，落单的 1 要靠人手配对。由于 $B_n$ 对偶数和奇数 $n$ 的定义不同，我们把两种情况分开处理。

## 解答

### 帕斯卡法则

由于 $(1+x)^{n+1} = (1+x)(1+x)^n$，两边 $x^r$ 的系数相等。左边的系数是 $\binom{n+1}{r}$。右边则是

$$(1+x)\sum_{s=0}^{n}\binom{n}{s}x^s = \sum_{s=0}^{n}\binom{n}{s}x^s + \sum_{s=0}^{n}\binom{n}{s}x^{s+1},$$

当 $1 \le r \le n$ 时，第一个和式贡献 $\binom{n}{r}$（来自 $s=r$），第二个和式贡献 $\binom{n}{r-1}$（来自 $s=r-1$）。因此

$$\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r} \qquad (1 \le r \le n).$$

### 递推关系

先看清楚 $B_n$ 是甚么会有帮助：它把帕斯卡三角形（Pascal's triangle）的一条“浅对角线”加起来，即 $\binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots$，一直加到上方数字不小于下方数字的最后一项为止。把三角形靠左排列时，组成 $B_6 = 1+5+6+1 = 13$ 的各项落在一条直线上：

$$
\begin{array}{ccccccc}
1 & & & & & & \\
1 & 1 & & & & & \\
1 & 2 & 1 & & & & \\
1 & 3 & 3 & 1 & & & \\
1 & 4 & 6 & 4 & 1 & & \\
1 & 5 & 10 & 10 & 5 & 1 & \\
1 & 6 & 15 & 20 & 15 & 6 & 1
\end{array}
$$

（琥珀色带标出了 $\binom{6}{0}, \binom{5}{1}, \binom{4}{2}, \binom{3}{3}$。）

**偶数情况，** $n = 2m$（$m \geq 0$）。我们需要证明 $B_{2m+2} = B_{2m+1} + B_{2m}$。按定义

$$B_{2m+2} = \sum_{j=0}^{m+1} \binom{2m+2-j}{j}.$$

两端的项是 $j = 0$，给出 $\binom{2m+2}{0} = 1$；以及 $j = m+1$，给出 $\binom{m+1}{m+1} = 1$。对于 $1 \leq j \leq m$，我们以上方 $2m+1-j$、下方 $j$ 套用帕斯卡法则；这是容许的，因为 $j \leq m$ 推出 $j \leq 2m+1-j$。于是

$$B_{2m+2} = \underbrace{1 + \sum_{j=1}^{m} \binom{2m+1-j}{j}}_{\text{(a)}} + \underbrace{\sum_{j=1}^{m} \binom{2m+1-j}{j-1} + 1}_{\text{(b)}}.$$

在 (a) 中，那个 1 就是 $\binom{2m+1}{0}$，即缺少了的 $k = 0$ 项，所以 (a) 等于 $\sum_{k=0}^{m} \binom{2m+1-k}{k} = B_{2m+1}$。

> **想一想** 第二个和式应该变成甚么？多出来的那个 1 又该放到哪里？

在 (b) 中令 $i = j - 1$：和式变成 $\sum_{i=0}^{m-1} \binom{2m-i}{i}$，这正是缺了最后一项 $i = m$ 的 $B_{2m}$。那最后一项是 $\binom{m}{m} = 1$，恰好就是多出来的 1。所以 (b) $= B_{2m}$，从而 $B_{2m+2} = B_{2m+1} + B_{2m}$。（当 $m = 0$ 时，和式都是空和，式子变成 $2 = 1 + 1$。）

**奇数情况，** $n = 2m+1$（$m \geq 0$）。我们需要证明 $B_{2m+3} = B_{2m+2} + B_{2m+1}$。此时

$$B_{2m+3} = \sum_{k=0}^{m+1} \binom{2m+3-k}{k}.$$

只有 $k = 0$ 项 $1 = \binom{2m+2}{0}$ 保持不动。对于 $1 \leq k \leq m+1$，以上方 $2m+2-k$、下方 $k$ 套用帕斯卡法则是容许的，因为 $k \leq m+1$ 推出 $k \leq 2m+2-k$。因此

$$\begin{aligned}
B_{2m+3} &= \binom{2m+2}{0} + \sum_{k=1}^{m+1} \binom{2m+2-k}{k} + \sum_{k=1}^{m+1} \binom{2m+2-k}{k-1} \\
&= \sum_{j=0}^{m+1} \binom{2m+2-j}{j} + \sum_{i=0}^{m} \binom{2m+1-i}{i} \qquad (i = k-1) \\
&= B_{2m+2} + B_{2m+1}.
\end{aligned}$$

这一次没有剩下任何端项。

每个 $n \geq 0$ 都可写成 $2m$ 或 $2m+1$ 的形式，所以对所有 $n = 0, 1, 2, \ldots$ 都有 $B_{n+2} - B_{n+1} = B_n$，此即所求。

### 与斐波那契数的关系

直接计算可得 $B_0 = \binom{0}{0} = 1$ 及 $B_1 = \binom{1}{0} = 1$。斐波那契数（Fibonacci numbers）由 $F_1 = 1$、$F_2 = 1$ 开始。所以 $B_0 = F_1$、$B_1 = F_2$，而两个数列都服从同一条递推关系（recurrence）："每一项等于前两项之和"。由数学归纳法（induction）（若 $B_n = F_{n+1}$ 且 $B_{n+1} = F_{n+2}$，则 $B_{n+2} = F_{n+2} + F_{n+1} = F_{n+3}$），

$$B_n = F_{n+1} \qquad \text{对所有} n \geq 0.$$

数列 $B$ 就是把斐波那契数列移后一位：$1, 1, 2, 3, 5, 8, 13, \ldots$。

## 验算

- 直接由定义计算几项：$B_2 = 1 + 1 = 2$，$B_3 = 1 + 2 = 3$，$B_4 = 1 + 3 + 1 = 5$，$B_5 = 1 + 4 + 3 = 8$。这些就是 $F_3, \ldots, F_6$。
- 较大的一项：$B_9 = \binom{9}{0} + \binom{8}{1} + \binom{7}{2} + \binom{6}{3} + \binom{5}{4} = 1 + 8 + 21 + 20 + 5 = 55 = F_{10}$。
- 抽查帕斯卡法则：$\binom{5}{2} + \binom{5}{1} = 10 + 5 = 15 = \binom{6}{2}$。

## 常见错误

- 对 $\binom{m+1}{m+1}$ 这类端项套用帕斯卡法则，这样会得出 $\binom{m}{m+1}$，超出了法则已获证明的范围 $1 \leq r \leq n$。应先把端项分离出来。
- 只处理一种奇偶性（parity）。$B_{2m}$ 和 $B_{2m+1}$ 的定义不同，而递推关系把偶数下标的项与奇数下标的项联系起来，所以两种情况都必须处理。
- 弄错下标的位移：答案是 $B_n = F_{n+1}$，而不是 $B_n = F_n$（检查一下 $B_0 = 1 \neq F_0$）。

## STEP 之外

恒等式 $F_{n+1} = \sum_k \binom{n-k}{k}$ 有一个漂亮的计数证明：两边都在计算用正方形和骨牌（dominoes）铺满一条 $1 \times n$ 长条的方法数（$k$ 是骨牌的数目，而在 $n - k$ 块砖之中选出骨牌的位置，就得到 $\binom{n-k}{k}$）。以两种方法计算同一个集合来证明恒等式（counting in two ways），是大学组合学（combinatorics）的核心技巧；生成函数（generating function）的观点亦

然，在这观点下 $\sum F_n x^n = x/(1-x-x^2)$。如想深入了解，可以读 Wilf 的 *generatingfunctionology*（《生成函数学》），或 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

# STEP I 2013，第 7 题

纯数学 · 一阶微分方程

**STEP I 2013，第 7 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q7

## 关键思路

在 (i) 和 (ii) 中，右方只取决于比值 $y/x$，所以设 $y = ux$ 便可把方程化为可分离变量 (separation of variables) 的形式。第 (iii) 部分打破了这个模式，因为 $x^2/y$ 并非单单是 $y/x$ 的函数。解决方法是问：解“应该”表现得像 $x$ 的哪一个幂？平衡两项后得出 $y = u\,x^{3/2}$，于是一切又可以分离变量。

## 解答

### 第 (i) 部分

设 $y = ux$，由乘积法则（product rule）得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = u + x\dfrac{\mathrm{d}u}{\mathrm{d}x}$。由于 $x/y = 1/u$ 及 $y/x = u$，方程变为

$$u + x\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{u} + u \quad\Longrightarrow\quad x\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{u}.$$

$u$ 项互相抵消，这正是这个代换（substitution）奏效的原因。分离变量（由于 $x > 0$，可以除以 $x$），

$$\int u\,\mathrm{d}u = \int \frac{1}{x}\,\mathrm{d}x \quad\Longrightarrow\quad \tfrac{1}{2}u^2 = \ln x + C,$$

因为 $x > 0$，所以写 $\ln x$ 而非 $\ln|x|$。当 $x = 1$ 时 $y = 2$，所以 $u = y/x = 2$，得 $2 = 0 + C$。因此

$$u^2 = 4 + 2\ln x.$$

由于 $y > 0$ 及 $x > 0$，必有 $u > 0$，所以取正根 $u = \sqrt{4 + 2\ln x}$，从而

$$y = x\sqrt{4 + 2\ln x}.$$

这要求 $4+2\ln x>0$（我们需要 $y>0$，而不只是 $y\geq 0$），即 $\ln x>-2$，亦即 $x>\mathrm{e}^{-2}$。这就是所求的结果。

### 第 (ii) 部分

右方同样是 $y/x$ 的函数，所以用同一个代换 $y=ux$：

$$u+x\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{1}{u}+2u \quad\Longrightarrow\quad x\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{1}{u}+u=\frac{1+u^2}{u}.$$

这次 $u$ 项没有抵消，但变量仍然可以分离：

$$\int\frac{u}{1+u^2}\,\mathrm{d}u=\int\frac{1}{x}\,\mathrm{d}x \quad\Longrightarrow\quad \tfrac{1}{2}\ln(1+u^2)=\ln x+C,$$

其中左方的积分属 $\frac{1}{2}\int f'/f$ 的形式，而 $1+u^2>0$，无需加绝对值（modulus）。在 $x=1$ 时 $u=2$，所以 $C=\frac{1}{2}\ln 5$。乘以 2 再取指数，

$$1+u^2=5x^2, \qquad \text{所以} \qquad u=\sqrt{5x^2-1}$$

（因 $u>0$，取正根）。因此

$$y=x\sqrt{5x^2-1} \qquad \left(x>\tfrac{1}{\sqrt{5}}\right),$$

这个限制来自 $5x^2-1>0$ 及 $x>0$。

### 第 (iii) 部分

现在 $y=ux$ 给出 $x^2/y=x/u$，当中仍单独含有 $x$，所以方程不能分离变量。我们想找一个代换，使右方两项以相同方式缩放。

> **想一想** 若 $y$ 表现得像幂 $x^k$，$k$ 取甚么值时 $\dfrac{x^2}{y}$ 与 $\dfrac{2y}{x}$ 的大小相同?

我们需要 $x^{2-k}$ 与 $x^{k-1}$ 相配，所以 $2-k=k-1$，得 $k=\frac{3}{2}$。（此时左方的 $\mathrm{d}y/\mathrm{d}x$ 亦表现得像 $x^{1/2}$，所以三项一致。）这启示我们用代换 $y=u\,x^{3/2}$，此时

$$\frac{\mathrm{d}y}{\mathrm{d}x}=x^{3/2}\frac{\mathrm{d}u}{\mathrm{d}x}+\tfrac{3}{2}u\,x^{1/2}, \qquad \frac{x^2}{y}=\frac{x^{1/2}}{u}, \qquad \frac{2y}{x}=2u\,x^{1/2}.$$

把整条方程除以 $x^{1/2}$（它是正数），

$$x\frac{\mathrm{d}u}{\mathrm{d}x}+\tfrac{3}{2}u=\frac{1}{u}+2u \quad\Longrightarrow\quad x\frac{\mathrm{d}u}{\mathrm{d}x}=\frac{1}{u}+\frac{u}{2}=\frac{2+u^2}{2u}.$$

分离变量，

$$\int\frac{2u}{2+u^2}\,\mathrm{d}u=\int\frac{1}{x}\,\mathrm{d}x \quad\Longrightarrow\quad \ln(2+u^2)=\ln x+C.$$

在 $x=1$ 时 $u=y/x^{3/2}=2$，所以 $C=\ln 6$，因此 $2+u^2=6x$。取正根 $u=\sqrt{6x-2}$，所以

$$y=x^{3/2}\sqrt{6x-2}=\sqrt{6x^4-2x^3}\qquad \left(x>\tfrac{1}{3}\right),$$

其中需要 $x>\frac{1}{3}$ 才能使 $6x-2>0$。

**(iii) 的另一解法**。把方程乘以 $2y$，并设 $z=y^2$，得

$$\frac{\mathrm{d}z}{\mathrm{d}x}-\frac{4}{x}z=2x^2,$$

这是一条线性方程(linear equation)，积分因子(integrating factor)为 $x^{-4}$。于是 $\left(x^{-4}z\right)'=2x^{-2}$，利用 $z(1)=4$ 得 $x^{-4}z=-2/x+6$，所以 $y^2=6x^4-2x^3$，与前相同。同一技巧（$z=y^2$）亦可解 (i) 和 (ii)。

## 验算

- 每个答案在 $x=1$ 时都给出 $y=2$：$1\cdot\sqrt{4}$、$1\cdot\sqrt{4}$ 及 $\sqrt{6-2}$。
- 把 (ii) 的答案求导后代回原方程。由 $y^2=5x^4-x^2$ 得 $2yy'=20x^3-2x$，所以 $y'=(10x^3-x)/y$。另一方面 $\dfrac{x}{y}+\dfrac{2y}{x}=\dfrac{x^2+2y^2}{xy}=\dfrac{10x^4-x^2}{xy}$，两者相同。
- 在 $x=2$ 时，三个解分别给出 $2\sqrt{4+2\ln 2}\approx 4.642$、$2\sqrt{19}\approx 8.718$ 及 $4\sqrt{5}\approx 8.944$。正如预期，(ii) 中 $y/x$ 的系数较大，使 $y$ 增长得比 (i) 快。

## 常见错误

- 在 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=u+x\dfrac{\mathrm{d}u}{\mathrm{d}x}$ 中漏掉 $u$，以致关于 $u$ 的方程出错。
- 把初始条件（initial condition）套用在 $y$ 而非 $u$ 上：这里在 $x=1$ 时两者相同，但在 (iii) 中一般需要用 $u=y/x^{3/2}$。
- 遗漏定义域（domain）。平方根必须是实数且为正，因而得出 $x>\mathrm{e}^{-2}$、$x>1/\sqrt{5}$ 及 $x>\frac{1}{3}$。

## STEP 之外

右方只取决于 $y/x$ 的方程称为齐次（homogeneous）方程，而代换 $y=ux$ 总能把它们分离变量。第 (iii) 部分其实是一种经过伪装的缩放对称性（scaling symmetry）：在 $x\mapsto\lambda x$、$y\mapsto\lambda^{3/2}y$ 之下方程不变，而这类缩放对称性总会启示一个使方程可分离变量的代换。形如 $y'+P(x)y=Q(x)y^n$ 的方程，在 $z=y^{1-n}$ 之下会变成线性方程，称为伯努

利方程（Bernoulli equations）；这里三部分都属于此类，且 $n = -1$。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）在一阶方程一章中涵盖了这两类方程。

详细解答

# STEP I 2013，第 8 题

纯数学 · 复合函数

**STEP I 2013，第 8 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q8

## 关键思路

合成函数（composite function）pq 的意思是“先做 q，再做 p”，而且只有在 q 的输出落在 p 的定义域（domain）之内时才有定义。因此，定义域来自 q（有需要时再加以收窄），而值域（range）则要问 p 实际上接收到哪些值才能求得。代数运算并不繁重；得分关键在于小心处理 $\sqrt{x^2} = |x|$ 以及各个定义域。在第 (iii) 部分，恒等式 $(x + \sqrt{x^2 - 1})(x - \sqrt{x^2 - 1}) = 1$ 帮了大部分的忙。

## 解答

### 第 (i) 部分

- $\mathrm{cb}(x) = \mathrm{c}(\ln x) = 2\ln x$。定义域就是 b 的定义域，即 $x > 0$，而 c 接受任何实数。由于 $\ln x$ 取遍所有实数值，$2\ln x$ 亦然：值域是 $\mathbb{R}$。
- $\mathrm{ab}(x) = (\ln x)^2$，定义域为 $x > 0$。由于 $\ln x$ 取遍所有实数值，其平方便取遍 $[0, \infty)$ 中的每个值（在 $x = 1$ 时取 0）：值域是 $y \geq 0$。
- $\mathrm{da}(x) = \sqrt{x^2} = |x|$，而不是 $x$，因为平方根号表示非负根（non-negative root）。这里 $x^2 \geq 0$ 恒成立，所以定义域是整个 $\mathbb{R}$，值域是 $y \geq 0$。
- $\mathrm{ad}(x) = (\sqrt{x})^2 = x$，但只限于 $x \geq 0$，因为必须先由 d 作用。所以定义域是 $x \geq 0$，值域是 $y \geq 0$。

留意 da 和 ad 的值域相同，但定义域和公式都不同：它们是不同的函数。

**第 (ii) 部分**

对于 fg：对每个实数 $x$，$g(x) = \sqrt{x^2+1} \geq 1$，所以 $g(x)$ 总落在 f 的定义域内。于是

$$fg(x) = \sqrt{(x^2+1)-1} = \sqrt{x^2} = |x|, \qquad \text{定义域}\mathbb{R}, \quad \text{值域}y \geq 0.$$

对于 gf：f 先作用，所以需要 $|x| \geq 1$；而 g 接受任何输入。于是

$$gf(x) = \sqrt{(x^2-1)+1} = |x|, \qquad \text{定义域}|x| \geq 1, \quad \text{值域}y \geq 1.$$

两条公式同样一致，但函数却不同：gf 的定义域不包括 $-1 < x < 1$，因此其值域由 1 开始。

**第 (iii) 部分**

基本事实是

$$\left(x+\sqrt{x^2-1}\right)\left(x-\sqrt{x^2-1}\right) = x^2-(x^2-1) = 1,$$

所以当 $x \geq 1$ 时，$h(x)\,k(x) = 1$。

**h 的图像（$x \geq 1$）。**

- $h(1) = 1$。
- 当 $x > 1$ 时，$h'(x) = 1 + \dfrac{x}{\sqrt{x^2-1}} > 0$，所以 h 是递增（increasing）的；而当 $x \to 1^+$ 时 $h'(x) \to \infty$，所以曲线以垂直切线（vertical tangent）离开 $(1,1)$。
- 当 $x \to \infty$ 时，$h(x) \to \infty$。至于渐近线（asymptote），

  $$h(x) - 2x = \sqrt{x^2-1} - x = \frac{-1}{\sqrt{x^2-1}+x} \to 0^-,$$

  所以 $y = 2x$ 是一条渐近线，曲线从下方趋近它。

因此 h 由 1 开始连续递增，且无上界：其值域是 $[1,\infty)$。

**k 的图像，$x \geq 1$ 的分支。**

- $k(1) = 1$，而且由基本事实，$k(x) = 1/h(x)$。
- 由于 h 是正数并递增至无限大，k 是正数并递减，且 $k(x) \to 0^+$。所以 $y = 0$ 是一条渐近线，曲线从上方趋近它。
- 当 $x \to 1^+$ 时，$k'(x) = 1 - \dfrac{x}{\sqrt{x^2-1}} \to -\infty$：在 $(1,1)$ 处同样有一条垂直切线，这分支就在该点与 h 的图像相接。

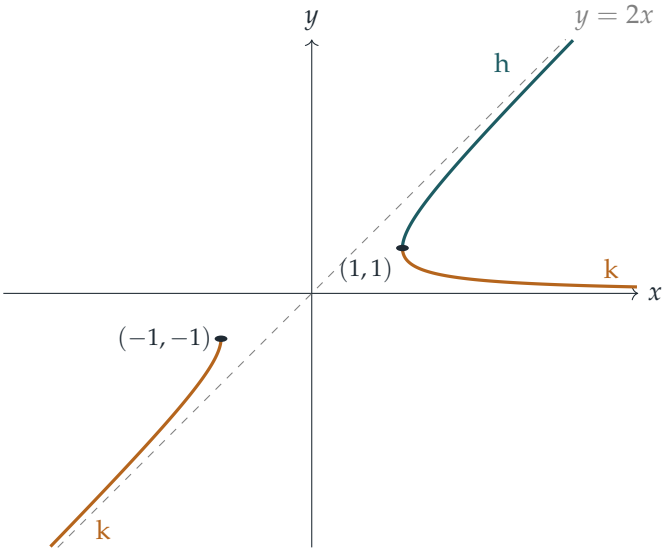
**k 的图像，$x \le -1$ 的分支。**

- $\mathrm{k}(-1) = -1$。
- 当 $x < -1$ 时有 $x < 0$，所以 $\mathrm{k}'(x) = 1 - \dfrac{x}{\sqrt{x^2-1}} > 1 > 0$：k 是递增的。当 $x \to -1^-$ 时，$\mathrm{k}'(x) \to +\infty$，所以在 $(-1,-1)$ 处有一条垂直切线。
- $x$ 和 $-\sqrt{x^2-1}$ 都是负数，所以当 $x \to -\infty$ 时 $\mathrm{k}(x) \to -\infty$。设 $x = -u$，其中 $u \ge 1$，

$$\mathrm{k}(x) - 2x = u - \sqrt{u^2-1} = \frac{1}{u + \sqrt{u^2-1}} \to 0^+,$$

所以 $y = 2x$ 亦是这分支的渐近线，曲线从上方趋近它。

这分支取遍 $(-\infty, -1]$ 中的每个值。




（有一个巧妙的方法可以一次看清这一切：设 $x = \cosh t$，其中 $t \ge 0$。那么 $\sqrt{x^2-1} = \sinh t$，所以 $\mathrm{h}(x) = e^t$ 而 $\mathrm{k}(x) = e^{-t}$；对于 $x = -\cosh t$，则 $\mathrm{k}(x) = -e^t$。）

**合成函数 kh。** h 的定义域是 $x \ge 1$，而每个值 $\mathrm{h}(x)$ 都满足 $\mathrm{h}(x) \ge 1$，落在 k 的定义域内。所以 kh 在整个 $x \ge 1$ 上都有定义。

> ? **想一想** kh 实际上用到 k 图像的哪一部分?

只用到 $x \ge 1$ 的分支，因为 h 从不输出小于 1 的值。此外，h 取遍 $[1,\infty)$ 中的每一个值，所以 kh 所取的值正好就是 k 在 $[1,\infty)$ 上的值。在这分支上，k 由 $\mathrm{k}(1) = 1$ 开始连续递减趋向 0，但永远达不到 0（因为 $\mathrm{k} = 1/\mathrm{h} > 0$）。因此

$$\mathrm{kh}: \quad \text{定义域} x \ge 1, \qquad \text{值域} 0 < y \le 1.$$

从双曲函数（hyperbolic function）的角度看，$\mathrm{kh}(\cosh t) = \mathrm{k}(e^t)$，这是 $t$ 的递减函数，在 $t = 0$ 时由 1 开始。

## 验算

- 代入数值：$\mathrm{h}(2) = 2 + \sqrt{3} \approx 3.732$，$\mathrm{k}(2) = 2 - \sqrt{3} \approx 0.268$；两者之积为 1。另外 $\mathrm{k}(-2) = -2 - \sqrt{3} \approx -3.732$，略高于 $2x = -4$，正如渐近线分析所预测。
- $\mathrm{kh}(1) = \mathrm{k}(1) = 1$，而 $\mathrm{kh}(2) = \mathrm{k}(3.732\ldots) = 3.732 - \sqrt{3.732^2 - 1} \approx 0.136$，落在 $(0, 1]$ 之内。
- 当 $x = 1000$ 时，$\mathrm{h}(x) - 2x \approx -0.0005$，$\mathrm{k}(x) \approx 0.0005$，与 $-1/(2x)$ 和 $1/(2x)$ 吻合。

## 常见错误

- 写成 $\sqrt{x^2} = x$。其实应是 $|x|$，这正是 da、fg 和 gf 都不是恒等函数（identity function）的原因。
- 因为最终公式 $|x|$ 处处有意义，便把 gf 的定义域写成 $\mathbb{R}$。合成函数的定义域由最先作用的函数决定。
- 遗漏了 k 在 $x \leq -1$ 的第二个分支，或遗漏了它的渐近线 $y = 2x$。

## STEP 之外

在大学里，函数是一个三元组：定义域、上域（codomain）和对应法则。两个法则相同但定义域不同的函数，例如这里的 fg 和 gf，确实是不同的函数；一旦讨论反函数（inverse function）、单射性（injectivity）和满射性（surjectivity），这种严谨就变得非常重要。函数 h 就是 $e^{\operatorname{arcosh} x}$，而 $\ln \mathrm{h}(x)$ 正是 $\operatorname{arcosh} x$ 的标准公式。Abbott 的 *Understanding Analysis*（《理解分析》）正是以这种细致的态度处理函数、连续性（continuity）和单调性（monotonicity）。

详细解答

# STEP I 2013，第 9 题

力学 · 抛体与碰撞

**STEP I 2013，第 9 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

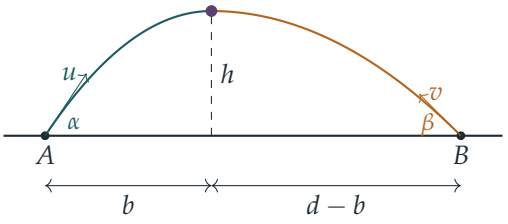
step.maths.org/questions/13-s1-q9

## 关键思路

抛射体（projectile）在飞行的最高点没有铅直速度，所以若在该处发生碰撞，之后落回地面所需的时间就等于上升所用的时间。要落回出发点，每个质点必须在同样的时间内走回原来的水平路程，这就迫使它的水平速度刚好反向。把这一点代入动量守恒（conservation of momentum），再利用两个质点在同一时刻到达最高点这个事实，其余结果便顺理成章。

## 解答

我们把 $A$ 和 $B$ 视为只受重力作用而自由运动的质点，不计空气阻力（air resistance），而碰撞时的冲量（impulse）沿连心线（line of centres）作用。由于这是直接碰撞（direct collision），而碰撞一刻两个质点都在水平运动，所以这个冲量是水平的。取 $x$ 轴沿平面、由 $A$ 的出发点指向 $B$ 的出发点，于是 $B$ 从 $x = d$ 出发。




### 在最高点相遇

$A$ 的铅直速度是 $u\sin\alpha - gt$，所以它在时间 $t = u\sin\alpha/g$ 到达最大高度。同理，$B$ 在 $t = v\sin\beta/g$ 时最高。两者同时抛出，并在各自处于最高点时相撞，因此这两个时间相

等：

$$T = \frac{u\sin\alpha}{g} = \frac{v\sin\beta}{g}, \qquad \text{所以} \qquad u\sin\alpha = v\sin\beta. \tag{1}$$

碰撞前一刻，$A$ 的速度是沿正 $x$ 方向的 $u\cos\alpha$，$B$ 的速度是沿负 $x$ 方向的 $v\cos\beta$；两者都没有铅直速度。

### 碰撞之后

> **想一想** 我们对恢复系数（coefficient of restitution）一无所知，又怎能对碰撞后的速度说出甚么来?

关键在于“返回原点”这个条件。冲量是水平的，所以碰撞后一刻 $A$ 的铅直速度仍然是零，位置在平面上方高度 $h$ 处。（在铅直方向上）由静止下落 $h$ 所需的时间，恰好等于上升 $h$ 所用的时间，即 $T$。在这段时间内，$A$ 必须在水平方向由 $x = b$ 回到 $x = 0$，其中 $b = u\cos\alpha\, T$ 是它上升途中走过的水平距离。所以它新的水平速度是

$$-\frac{b}{T} = -u\cos\alpha :$$

$A$ 的水平速度刚好反向，同理 $B$ 的也是如此。（事实上，每个质点都沿自己原来的路径倒退回去。）

现在，$x$ 方向的动量守恒给出

$$mu\cos\alpha - Mv\cos\beta = -mu\cos\alpha + Mv\cos\beta, \qquad \text{所以} \qquad mu\cos\alpha = Mv\cos\beta. \tag{2}$$

换言之，碰撞前后的总动量都是零。以 (1) 除 (2)——这是容许的，因为 $\alpha$ 和 $\beta$ 都严格介于 0 与 $\frac{1}{2}\pi$ 之间，所以 (1) 的两边都不为零——得

$$m\cot\alpha = M\cot\beta,$$

此即所求。

### 碰撞发生的位置

在时间 $T$ 内，$A$ 走过水平距离 $b = u\cos\alpha\, T$，$B$ 则走过 $v\cos\beta\, T$。两者合起来刚好走完两者之间的距离，所以

$$u\cos\alpha\, T + v\cos\beta\, T = d.$$

由 (2)，$v\cos\beta = \dfrac{m}{M}u\cos\alpha$，所以

$$b\left(1 + \frac{m}{M}\right) = d \qquad \Longrightarrow \qquad b = \frac{Md}{m+M},$$

此即所求。（$B$ 越重，它的水平速度就越慢，相遇点便离 $A$ 的出发点越远。）

### 碰撞的高度

对 $A$ 在铅直方向使用 $v^2 = u^2 + 2as$，得最大高度为

$$h = \frac{u^2 \sin^2 \alpha}{2g}, \qquad \text{而} \qquad b = u\cos\alpha \cdot \frac{u \sin\alpha}{g} = \frac{u^2 \sin\alpha\cos\alpha}{g}.$$

两式相除，

$$\frac{h}{b} = \frac{\sin\alpha}{2\cos\alpha} = \tfrac{1}{2}\tan\alpha, \qquad \text{所以} \qquad h = \tfrac{1}{2} b \tan\alpha.$$

## 验算

- 对称性：当 $m = M$ 时，得 $\alpha = \beta$ 及 $b = \frac{1}{2}d$，所以两个质点在正中间相遇，理应如此。
- 高度公式表明，由 $A$ 的出发点到其路径最高点的连线，斜率是 $\frac{1}{2}\tan\alpha$，即抛射时斜率的一半：这正是一个熟悉的事实——抛物线（parabola）在起点的切线与对称轴（axis of symmetry）相交之处，高度是顶点高度的两倍。对 $B$ 作同样的计算，得 $h = \frac{1}{2}(d - b)\tan\beta$；这与 $\frac{1}{2}b\tan\alpha$ 吻合，原因正是 $m\cot\alpha = M\cot\beta$。
- 数值验证：取 $m = 1$、$M = 2$、$\alpha = 45°$、$u = 10\,\mathrm{m\,s^{-1}}$、$g = 9.8\,\mathrm{m\,s^{-2}}$。于是 $\cot\beta = \frac{1}{2}$，$v = \sqrt{50}/\sin\beta \approx 7.91\,\mathrm{m\,s^{-1}}$，$d \approx 7.65\,\mathrm{m}$，从而 $b = \frac{2}{3}d \approx 5.10\,\mathrm{m}$，$h = \frac{1}{2}b \approx 2.55\,\mathrm{m}$。用电脑把两个运动逐步向前推算，并在碰撞时把水平速度反向，结果证实两个质点确实在此点、各自处于最高点时相遇，而且都落回各自的出发点。

## 常见错误

- 试图使用恢复系数：题目没有给出，也根本不需要，因为返回原点的条件已经确定了碰撞后的速度。
- 忘记两个质点必须同时到达最高点，而正是这一点给出 (1)；没有它，就无法得出余切（cotangent）。
- 动量方程中的正负号错误：碰撞前 $B$ 是沿负方向运动的。

## STEP 之外

返回原点的条件其实就是时间反演（time reversal）：在重力作用下，牛顿（Newton）的运动方程即使把时钟倒转，形式也不变，所以把质点的速度反向，它就会沿自己原来的路径返回。这里总动量为零，意味着我们身处质心参考系（centre-of-mass frame）之中；在这个参考系里，弹性碰撞（elastic collision）只会把两个速度反向，与这里发生的情况完

全一样（所以这次碰撞是完全弹性的）。质心参考系和守恒定律在 Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）中有精彩的探讨。

详细解答

# STEP I 2013，第 10 题

力学 · 恢复系数与摩擦力

**STEP I 2013，第 10 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q10

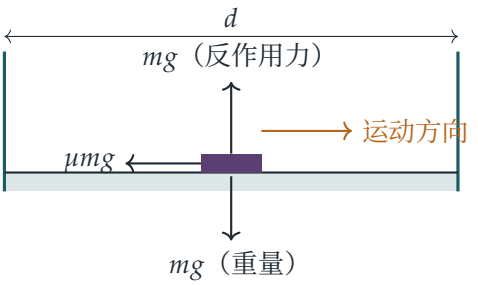
## 关键思路

在两次碰撞之间，唯一的水平力是摩擦力（friction），所以冰球的减速度（deceleration）恒为 $\mu g$，而等加速度公式 $v^2 = u^2 + 2as$ 把冰球离开一道挡板时的速率与到达下一道挡板时的速率联系起来；然后牛顿（Newton）碰撞定律（law of restitution）把到达时的速率乘以 $r$。结果得到 $v_i^2$ 的一条线性递推关系（linear recurrence）。要解线性递推关系 $w_{i+1} = aw_i + b$，只需平移到它的不动点（fixed point），之后它就只是一个等比数列（geometric sequence）。

## 解答

### 递推关系

把冰球视为一个质点，沿垂直于挡板的直线运动。在两次碰撞之间冰球在冰面上滑行时，法向反作用力（normal reaction）为 $mg$（没有铅垂方向的加速度）；由于冰球正在滑动，摩擦力取其极限值（limiting value）$\mu mg$，方向与运动方向相反。



由牛顿第二定律，减速度为常数 $\mu g$。因此，对于一趟长度为 $d$、起始速率为 $v_i$（即第 $i$ 次碰撞刚结束时的速率）的横越（traverse），冰球到达另一道挡板时的速率 $u$ 满足

$$u^2 = v_i^2 - 2\mu g d.$$

挡板是固定的，所以牛顿碰撞定律指出冰球离开时的速率是其接近速率的 $r$ 倍：$v_{i+1} = ru$。两边平方，

$$v_{i+1}^2 = r^2u^2 = r^2v_i^2 - 2r^2\mu g d, \qquad \text{所以} \qquad v_{i+1}^2 - r^2v_i^2 = -2r^2\mu g d,$$

这正是所求的结果。若把起始速率记作 $v_0 = v$，同样的论证也适用于第一趟横越，所以这条递推关系对所有 $i \geq 0$ 都成立。

**横越的次数**

两边除以 $2\mu g d$，并令 $w_i = \dfrac{v_i^2}{2\mu g d}$，于是 $w_0 = k$ 及

$$w_{i+1} = r^2(w_i - 1).$$

“经过 $n$ 趟横越后在挡板旁静止”的意思是：冰球在第 $n$ 趟横越结束时以零速率到达挡板。这时它离开的速率也是零，所以 $v_n = 0$，即 $w_n = 0$（而对于 $i < n$，$w_i > 0$）。

**想一想** 怎样解 $w_{i+1} = r^2(w_i - 1)$?

先找常数解 $w_i = c$：需要 $c = r^2(c-1)$，所以 $c = -\dfrac{r^2}{1-r^2}$；由于 $r \neq 1$，这是有意义的。把递推关系减去 $c = r^2(c-1)$，得

$$w_{i+1} - c = r^2(w_i - c),$$

所以 $w_i - c$ 是公比为 $r^2$ 的等比数列：

$$w_i = c + r^{2i}(k - c).$$

令 $w_n = 0$，得 $r^{2n}(k-c) = -c$，即

$$r^{2n}\left(k + \frac{r^2}{1-r^2}\right) = \frac{r^2}{1-r^2} \quad \Longrightarrow \quad r^{2n-2}\left(k(1-r^2) + r^2\right) = 1.$$

两边乘以 $r^{-2n}$，整理成

$$r^{-2n} = 1 + k\left(r^{-2} - 1\right), \qquad \text{所以} \qquad \boxed{n = \frac{\ln\left(1 + k(r^{-2} - 1)\right)}{2\ln(1/r)}}$$

（当 $0 < r < 1$ 时，$\ln(1/r) > 0$；又因为 $k > 0$ 且 $r^{-2} > 1$，对数的自变量大于 1，所以 $n > 0$，理应如此。）

### $r = \mathrm{e}^{-1}$ 的情况

此时 $r^{-2} = \mathrm{e}^2$ 而 $\ln(1/r) = 1$，所以公式变成

$$n = \tfrac{1}{2}\ln\left(1 + k(\mathrm{e}^2 - 1)\right),$$

这正是所求的结果。

### $r = 1$ 的情况

递推关系变成 $w_{i+1} = w_i - 1$：由于在挡板处没有损失，每趟横越恰好消耗 $v^2$ 中的 $2\mu g d$。所以 $w_i = k - i$，而由 $w_n = 0$ 得

$$n = k = \frac{v^2}{2\mu g d}.$$

这其实就是能量守恒的考虑：初始动能 $\frac{1}{2}mv^2$ 在总距离 $nd$ 上全部被摩擦力耗散，所以 $\frac{1}{2}mv^2 = \mu mg\, nd$。

## 验算

- 取 $r = \mathrm{e}^{-1}$，从 $w_3 = 0$ 把递推关系倒过来推：$w_2 = 1$，$w_1 = \mathrm{e}^2 + 1$，$w_0 = \mathrm{e}^4 + \mathrm{e}^2 + 1 \approx 62.99$。公式给出 $\frac{1}{2}\ln\left(1 + (\mathrm{e}^4 + \mathrm{e}^2 + 1)(\mathrm{e}^2 - 1)\right) = \frac{1}{2}\ln \mathrm{e}^6 = 3$，理应如此。
- 当 $r \to 1$ 时，$r^{-2} - 1 \approx 2(1 - r)$ 且 $\ln(1/r) \approx 1 - r$，所以一般公式趋向 $\dfrac{2k(1-r)}{2(1-r)} = k$，与 $r = 1$ 时的答案吻合。
- $k$ 是无量纲的（$v^2$ 和 $gd$ 都是速率的平方），而 $n$ 也必须是无量纲的。
- 对随机选取的 $\mu$、$d$、$r$ 和 $v$ 以数值方法逐步模拟运动，结果以很高的精确度证实了这条递推关系。

## 常见错误

- 对 $v_i^2$ 应用碰撞定律时用了因子 $r$ 而不是 $r^2$：碰撞定律讲的是速率，所以速率的平方要乘以 $r^2$，摩擦力那一项也一样。
- 忘记 $n$ 数的是横越的趟数，而不是反弹速率非零的碰撞次数：条件是 $v_n = 0$，而 $v_0 = v$。

## STEP 之外

递推关系 $w_{i+1} = aw_i + b$ 是最简单的线性差分方程（linear difference equation），而减去不动点这个技巧，正是你将来在线性微分方程中会遇到的“余函数加特解”（complementary

function plus particular solution）方法。这类弹跳问题（经典例子是皮球在地面上弹跳）也是能量耗散的好例子；在没有摩擦力的情况下，它们还展示了在有限时间内发生无限多次事件的现象。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对碰撞和摩擦力有更深入的论述。

详细解答

# STEP I 2013，第 11 题

力学 · 摩擦力与平衡

**STEP I 2013，第 11 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q11

## 关键思路

两条绳在相对的两侧分别与水平成 $\frac{1}{2}\pi - \alpha$ 和 $\alpha$ 的角，因此两绳互相垂直（perpendicular）。两个互相垂直的力 $T\cos\beta$ 和 $T\sin\beta$ 合成一个大小为 $T$ 的单一力，方向由第一个力转过 $\beta$。把两条绳换成一个与水平成 $\alpha+\beta$ 角、大小为 $T$ 的拉力之后，两部分都变成简短的练习：第 (i) 部分是极限摩擦（limiting friction），第 (ii) 部分则要察觉物块会被整个提离表面。

## 解答

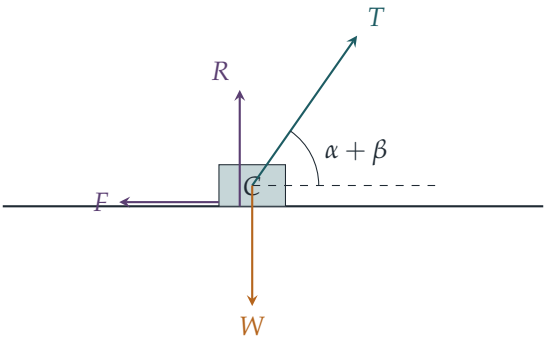
我们把 $C$ 视为质点（particle），并把绳视为轻绳。取 $x$ 轴为水平并指向 $B$，$y$ 轴为竖直向上。

### 张力的合力

绳 $BC$ 指向正 $x$ 轴上方 $\alpha$ 角的方向，所以其单位向量（unit vector）为 $(\cos\alpha, \sin\alpha)$。绳 $AC$ 与负 $x$ 轴成 $\frac{1}{2}\pi - \alpha$ 角，所以它与正 $x$ 轴成 $\pi - (\frac{1}{2}\pi - \alpha) = \frac{1}{2}\pi + \alpha$ 角，单位向量为 $(-\sin\alpha, \cos\alpha)$。两者互相垂直。因此两绳的总拉力为

$$\begin{aligned} T\cos\beta\,(\cos\alpha, \sin\alpha) + T\sin\beta\,(-\sin\alpha, \cos\alpha) &= T\big(\cos\alpha\cos\beta - \sin\alpha\sin\beta, \\ &\qquad \sin\alpha\cos\beta + \cos\alpha\sin\beta\big) \\ &= T\big(\cos(\alpha+\beta),\ \sin(\alpha+\beta)\big). \end{aligned}$$

所以两绳的作用恰好等同一个大小为 $T$、在水平之上成 $\alpha+\beta$ 角、拉向 $B$ 一侧的单一力。由于 $0 < \alpha + \beta < \frac{1}{2}\pi$，两个分量都是正数。




**第 (i) 部分**

设 $R$ 为法向反作用力（normal reaction），$F$ 为摩擦力；若物块静止，摩擦力沿水平方向背离 $B$。平衡时，

$$\begin{aligned} &\text{竖直方向：} \quad R + T\sin(\alpha+\beta) = W, \qquad \text{所以} R = W - T\sin(\alpha+\beta),\\ &\text{水平方向：} \quad F = T\cos(\alpha+\beta). \end{aligned}$$

条件 $W > T\sin(\alpha+\beta)$ 使 $R > 0$，所以物块确实被压在表面上，可以产生摩擦。只要所需的摩擦力不超过其极限值，即 $F \le \mu R$，物块便保持静止：

$$T\cos(\alpha+\beta) \le \mu\big(W - T\sin(\alpha+\beta)\big).$$

现在代入 $\mu = \tan\lambda = \sin\lambda/\cos\lambda$，并在两边乘以 $\cos\lambda$；由于 $\lambda$ 是锐角，这个数是正数：

$$T\cos(\alpha+\beta)\cos\lambda \le W\sin\lambda - T\sin(\alpha+\beta)\sin\lambda.$$

移项并利用 $\cos(P-Q) = \cos P\cos Q + \sin P\sin Q$，得

$$W\sin\lambda \ge T\big(\cos(\alpha+\beta)\cos\lambda + \sin(\alpha+\beta)\sin\lambda\big) = T\cos(\alpha+\beta-\lambda),$$

这正是所求的条件。（每一步都可逆，所以事实上物块保持静止当且仅当此式成立。）

? **想一想** 为甚么我们需要条件 $W > T\sin(\alpha+\beta)$ 呢?

若没有这个条件，竖直方向的方程会给出 $R \le 0$。表面只能推、不能拉，所以 $R$ 为负是不可能的：物块会被提离表面（或仅仅刚好接触），也就没有摩擦力把它固定。这正是第 (ii) 部分的情况。

**第 (ii) 部分**

现在 $W = T\tan\phi$ 而 $\alpha+\beta = 2\phi$，所以 $0 < \phi < \frac{1}{4}\pi$。比较重量与两绳向上的拉力：

$$\begin{aligned} T\sin(\alpha+\beta) - W &= T\left(2\sin\phi\cos\phi - \tan\phi\right) \\ &= T\tan\phi\left(2\cos^2\phi - 1\right) \\ &= T\tan\phi\cos 2\phi > 0, \end{aligned}$$

因为 $0 < \phi < \frac{1}{4}\pi$ 使 $\tan\phi$ 和 $\cos 2\phi$ 都是正数。所以两绳向上的拉力大于重力向下的拉力。若物块留在表面上，我们便需要 $R = W - T\sin(\alpha+\beta) < 0$，这是不可能的。因此物块立即离开表面：$R = 0$，而既然没有接触，也就没有摩擦力。

此时仅有的力是与水平成 $2\phi$ 角的拉力 $T$ 和重量。合力（resultant force）为

$$\left(T\cos 2\phi,\ T\sin 2\phi - T\tan\phi\right) = \left(T\cos 2\phi,\ T\tan\phi\cos 2\phi\right) = \frac{T\cos 2\phi}{\cos\phi}\left(\cos\phi,\ \sin\phi\right),$$

其中竖直分量用了上面的计算。其大小 $T\cos 2\phi/\cos\phi$ 是正数，所以不为零。物块由静止开始，所以其初始加速度，从而它开始运动的方向，都沿着合力的方向。该方向为 $(\cos\phi, \sin\phi)$，即在水平之上成 $\phi$ 角（指向 $B$），正如所求。

## 验算

- 取 $T = 10\,\mathrm{N}$、$\alpha = 20°$、$\beta = 40°$，于是 $\phi = 30°$，$W = 10\tan 30° \approx 5.77\,\mathrm{N}$。向上的拉力为 $10\sin 60° \approx 8.66\,\mathrm{N} > W$，合力为 $(10\cos 60°,\ 8.66 - 5.77) = (5,\ 2.89)$，其角度为 $\arctan(2.89/5) = 30° = \phi$。
- 在第 (i) 部分中，$\lambda = \frac{1}{2}\pi$（无限粗糙）给出 $W \geq T\sin(\alpha+\beta)$，这本来已经成立，所以物块永不滑动；$\lambda \to 0$（光滑）给出 $0 \geq T\cos(\alpha+\beta)$，这不成立，所以光滑表面无法使物块静止。两者都符合我们的预期。
- 第 (i) 部分的条件亦已用电脑，就 $\alpha, \beta, T, W, \mu$ 的大量随机取值，与直接检验 $F \leq \mu R$ 的结果互相核对。

## 常见错误

- 把每条绳分别分解，结果在代数中迷失，而没有察觉两绳互相垂直、应先把它们合成。
- 在第 (ii) 部分写出 $R = W - T\sin(\alpha+\beta)$ 后继续处理摩擦力，却没有察觉 $R$ 会是负数。
- 忘记说明由静止开始的质点最初沿合力方向运动。

## STEP 之外

$\lambda$ 即摩擦角（angle of friction），其意思是法向反作用力与极限摩擦力合成一个与法线成 $\lambda$ 角的总反作用力；平衡可能出现，恰恰在总反作用力能落在这个“摩擦锥”（cone of friction）之内的时候。第 (i) 部分正是这个命题的变装版本，而同一幅图像亦能处理三维中的摩擦以及刚体问题。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对摩擦、约束和接触力有细致的论述。

详细解答

# STEP I 2013，第 12 题

概率与统计 • 不放回抽样

> **STEP I 2013，第 12 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s1-q12

## 关键思路

“随机”选取 $k$ 粒药片，意思是容器中剩下的药片里，每一组 $k$ 粒药片被选中的机会均等，所以每天的概率（probability）都是两个计数之比：有利的选法数目除以 $\binom{\text{剩下的药片}}{k}$。至于最后一天的问题，窍门在于对称性（symmetry）：留到最后一天的药片，与第一天选出的药片同样随机。在 (iii) 中，各天之间环环相扣，十分顺畅，因为在“好”的一天之后，容器的状态恰好就像一个较短的假期。

## 解答

在整个解答中，把药片视为可区分的（例如编上号码），尽管它们看起来一模一样。这样做不会改变任何概率，而且令从 $N$ 粒中选 $k$ 粒的全部 $\binom{N}{k}$ 种选法机会均等。

### 第 (i) 部分

当 $k = 3$ 及 $n = 3$ 时，共有 9 粒药片，每种 3 粒。从中选 3 粒的方法数为 $\binom{9}{3} = 84$。要每种各得一粒，就要从 3 粒第 1 种药片中选一粒、从 3 粒第 2 种中选一粒、从 3 粒第 3 种中选一粒，所以共有 $3^3 = 27$ 种方法。因此

$$\mathrm{P}(\text{第 1 天每种各一粒}) = \frac{27}{84} = \frac{9}{28},$$

正如所求。

> ? **想一想** 为甚么求最后一天的概率时，我不需要逐一分析第 1 天和第 2 天发生了甚么？

在整个假期中，九粒药片被随机分成三组，每组三粒：第 1 天的一组、第 2 天的一组和第

3 天的一组（即剩下的）。无论按 1、2、3 的次序还是按 3、2、1 的次序选出各组，所得的分布（distribution）都相同：把药片分成三个有标记的组，每一种分法的机会都均等。特别地，留给第 3 天的那组药片是一组均匀随机（uniformly random）的三粒药片，与第 1 天选出的那组完全一样。所以最后一天剩下每种各一粒药片的概率同样是

$$\frac{9}{28}.$$

**第 (ii) 部分**

现在共有 $3n$ 粒药片，每种 $n$ 粒。共有 $\binom{3n}{3}$ 种机会均等的选法，其中 $n^3$ 种包含每种各一粒。所以

$$\begin{aligned}\mathrm{P}(\text{第 1 天每种各一粒}) &= \frac{n^3}{\binom{3n}{3}} = \frac{6n^3}{3n(3n-1)(3n-2)} \\ &= \frac{2n^2}{(3n-1)(3n-2)}.\end{aligned}$$

当 $n=3$ 时，此值为 $\frac{18}{8\cdot 7}=\frac{9}{28}$，与 (i) 吻合。

**第 (iii) 部分**

现在 $k=2$，共有 $2n$ 粒药片，两种药片各 $n$ 粒。对 $m=n,n-1,\ldots,1$，设 $G_m$ 为以下事件：在某天开始时容器中每种药片各有 $m$ 粒，而我在当天选到每种各一粒（所以第 1 天时 $m=n$）。

若容器中每种各有 $m$ 粒，则选两粒药片共有 $\binom{2m}{2}=m(2m-1)$ 种选法，其中 $m\cdot m$ 种是每种各一粒。所以

$$\mathrm{P}(\text{每种各一粒} \mid m \text{ 为每种剩余粒数}) = \frac{m^2}{m(2m-1)} = \frac{m}{2m-1}.$$

若到目前为止每一天都是“好”的，那么每天都恰好取走了每种各一粒，所以下一天开始时每种各有 $m-1$ 粒。把每天的条件概率（conditional probability）逐一相乘，

$$\mathrm{P}(\text{每天都是每种各一粒}) = \prod_{m=1}^{n}\frac{m}{2m-1} = \frac{n!}{1\times 3\times 5\times\cdots\times(2n-1)}.$$

为了整理奇数之积，把分子和分母同乘以偶数之积 $2\times 4\times\cdots\times(2n)=2^n\,n!$：

$$1\times 3\times\cdots\times(2n-1) = \frac{(2n)!}{2^n\,n!}.$$

因此

$$\boxed{\mathrm{P}(\text{每天都是每种各一粒}) = \frac{2^n\,(n!)^2}{(2n)!}}$$

（验算一下：$n = 1$ 时得 1，这是正确的：每种只有一粒时，我必定两粒都取。）

现在对 $n!$ 及 $(2n)!$ 使用斯特林（Stirling）近似（Stirling's approximation）：

$$(n!)^2 \approx 2n\pi \left(\frac{n}{\mathrm{e}}\right)^{2n},$$
$$(2n)! \approx \sqrt{4n\pi}\left(\frac{2n}{\mathrm{e}}\right)^{2n} = 2\sqrt{n\pi}\, 2^{2n}\left(\frac{n}{\mathrm{e}}\right)^{2n}.$$

因子 $(n/\mathrm{e})^{2n}$ 互相抵消，剩下

$$\frac{2^n\,(n!)^2}{(2n)!} \approx \frac{2^n \cdot 2n\pi}{2\sqrt{n\pi}\, 2^{2n}} = 2^{-n}\,\frac{n\pi}{\sqrt{n\pi}} = 2^{-n}\sqrt{n\pi},$$

这正是所求的近似式。

## 验算

- 用电脑穷举所有抽取药片的方式，证实了 (i) 中第 1 天和第 3 天的概率都是 $\frac{9}{28}$，(ii) 的公式在 $n \le 7$ 时成立，以及 (iii) 的公式在 $n \le 4$ 时成立。
- 当 $n = 10$ 时，(iii) 的准确概率为 $\frac{2^{10}(10!)^2}{20!} \approx 0.005542$，而 $2^{-10}\sqrt{10\pi} \approx 0.005474$：偏低约 1.3%。随着 $n$ 增大，两者之比趋向 1（$n = 50$ 时约为 1.0025），这与斯特林近似的预期一致。
- (iii) 的另一种计数方法：把 $2n$ 粒药片分成 $n$ 对并排成有序列，共有 $(2n)!/2^n$ 种机会均等的分法；好的分法是每天把第 1 种药片与第 2 种药片配对，共有 $n! \times n!$ 种。两者之比同样是 $2^n(n!)^2/(2n)!$。

## 常见错误

- 每次抽一粒药片，写出 $\frac{3}{9} \times \frac{3}{8} \times \frac{3}{7}$，但这是按此次序抽到第 1、2、3 种的机会。三种药片出现的次序共有 3! 种，所以应得 $6 \times \frac{27}{504} = \frac{9}{28}$。（又或者：第一粒可以是任何一种，所以是 $1 \times \frac{6}{8} \times \frac{3}{7} = \frac{9}{28}$。）
- 求最后一天的概率时，尝试对第 1 天发生的所有情况作条件分析；这方法可行，但十分冗长，而对称性能立即给出答案。
- 对 $(2n)!$ 用斯特林近似时，要把所有 $n$ 都换成 $2n$，包括平方根之内：$\sqrt{2(2n)\pi} = 2\sqrt{n\pi}$，而不是 $\sqrt{2n\pi}$。

## STEP 之外

(iii) 的答案是 $2^n/\binom{2n}{n}$，而估计式 $\binom{2n}{n} \approx 4^n/\sqrt{n\pi}$ 在概率论中随处可见：它给出简单随机游走（simple random walk）在 $2n$ 步后回到起点的机会，并且是证明一维或二维的这类游走以概率 1 回到原点的关键。(i) 中的对称性论证，就是可交换性（exchangeability）的概念：随机抽取的次序不会改变它们的联合分布（joint distribution）。这两个概念在 Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）以及 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）中都有深入发展。

详细解答

# STEP I 2013，第 13 题

概率与统计 · 计数与期望值

**STEP I 2013，第 13 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s1-q13

## 关键思路

数字 $1,\ldots,52$ 可分成 26 个互不相交的“伙伴对” $\{k, 53-k\}$，而被弃掉的一对，其实就是我们把两个成员都选中了的伙伴对。因此，要数出恰好有 $k$ 个完整对的选法，我们先选定哪些对是完整的，再用来自不同对的数字填满其余位置，每次可自由选取该对中的哪一个成员。至于期望值（expectation），对这 26 个对运用线性性质（linearity）便可迅速验证结果。

## 解答

### 数算选法

把这些整数写成 26 个对 $\{1,52\},\{2,51\},\ldots,\{26,27\}$。由 1 至 52 中的两个整数之和为 53，当且仅当它们组成其中一个对，而每个整数恰好属于一个对。所以 $X$ 就是两个成员都在所选七个数之中的对的数目。由于 $2X \le 7$，$X$ 的取值为 $0,1,2,3$。

所有 $\binom{52}{7}$ 种选法的可能性相等，因此我们数算 $X=k$ 的选法数目。

- 选出 $k$ 个完整对：共 $\binom{26}{k}$ 种方法。这些对用去七个数中的 $2k$ 个。
- 其余 $7-2k$ 个数必须来自余下 $26-k$ 个对中 $7-2k$ 个不同的对（若有两个来自同一对，便会多出一个完整对）：共 $\binom{26-k}{7-2k}$ 种方法。
- 从每个这样的对中取其两个成员之一：共 $2^{7-2k}$ 种方法。

每个 $X=k$ 的选法以这种方式恰好出现一次，所以

$$\mathrm{P}(X=k)=\frac{\binom{26}{k}\binom{26-k}{7-2k}2^{7-2k}}{\binom{52}{7}},\qquad k=0,1,2,3.$$

**具体数值**

$$\begin{aligned}
X=0:&\quad \tbinom{26}{7}2^7 = 657\,800\times 128 = 84\,198\,400,\\
X=1:&\quad 26\tbinom{25}{5}2^5 = 26\times 53\,130\times 32 = 44\,204\,160,\\
X=2:&\quad \tbinom{26}{2}\tbinom{24}{3}2^3 = 325\times 2024\times 8 = 5\,262\,400,\\
X=3:&\quad \tbinom{26}{3}\tbinom{23}{1}2 = 2600\times 23\times 2 = 119\,600,
\end{aligned}$$

而 $\binom{52}{7} = 133\,784\,560$。这四个数目加起来确实等于 $\binom{52}{7}$，这是一个很好的验证，说明没有遗漏或重复计算。

> ? **想一想** 分数中有九位数时，我们该如何化简?

题目中的附注就是提示：$5593 = 7\times 17\times 47$ 应该是分母。事实上 $\binom{52}{7} = 23\,920\times 5593$，而这四个数目每一个都是 23 920 的倍数：

$$84\,198\,400 = 3520\times 23\,920, \qquad 44\,204\,160 = 1848\times 23\,920,$$
$$5\,262\,400 = 220\times 23\,920, \qquad 119\,600 = 5\times 23\,920.$$

所以 $X$ 的分布（distribution）为

| $k$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\mathrm{P}(X=k)$ | $\frac{3520}{5593}$ | $\frac{1848}{5593}$ | $\frac{220}{5593}$ | $\frac{5}{5593}$ |

（第二个的最简形式为 $\frac{264}{799}$），而 $3520+1848+220+5 = 5593$，理应如此。

**期望值**

$$\mathrm{E}(X) = \frac{0\times 3520 + 1\times 1848 + 2\times 220 + 3\times 5}{5593} = \frac{2303}{5593}.$$

由于 $2303 = 7^2\times 47$ 而 $5593 = 7\times 17\times 47$，所以

$$\mathrm{E}(X) = \frac{7^2\times 47}{7\times 17\times 47} = \frac{7}{17},$$

正如所求。

**求期望值的捷径**

对于 $j = 1,\ldots,26$，若第 $j$ 个对的两个成员都被选中，则设 $I_j = 1$，否则为 0，于是 $X = I_1 + \cdots + I_{26}$。选中某一指定对的两个成员后，须从其余 50 个数中再选 5 个，所以

$$\mathrm{P}(I_j = 1) = \frac{\binom{50}{5}}{\binom{52}{7}} = \frac{7\times 6}{52\times 51} = \frac{7}{442}.$$

由期望值的线性性质（这不需要独立性（independence）；事实上这些 $I_j$ 并不独立），

$$\mathrm{E}(X) = 26 \times \frac{7}{442} = \frac{7}{17}.$$

## 验算

- 四个概率之和为 1，而两种方法都给出相同的 $\mathrm{E}(X) = \frac{7}{17} \approx 0.412$。
- 以电脑模拟 300 000 次随机选取，得到 $X = 0, 1, 2, 3$ 的相对频率约为 0.630、0.329、0.039、0.001，而准确值为 0.6294、0.3304、0.0393、0.0009。
- 合理性检查：$\mathrm{P}(X = 3)$ 非常小，这合乎预期，因为只有七个数却要有三个完整对。

## 常见错误

- 把 $X = 0$ 的选法数目算成 $\binom{52}{7}$ 减去某个粗略的数，或只算 $\binom{26}{7}$，忘记了每个对取哪一个成员所带来的因子 $2^7$。
- 对于 $X = 1$，把其余五个数的选法算成 $\binom{50}{5}$：这容许出现第二个完整对，因而多算了。

## STEP 之外

把一个计数写成指示变量（indicator variable）之和，再运用期望值的线性性质，是大学一年级概率论中最有用的技巧之一：即使分布根本无法求出，它仍能给出期望值。本题的完整分布尚可求得，但例如要求随机排列（permutation）的不动点（fixed point）的期望数目，指示变量便是最自然的切入点。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）一书反复运用这个想法。

详细解答

# STEP II 2013，第 1 题

纯数学 · 对数曲线的弦

**STEP II 2013，第 1 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q1

本题在 *Beyond STEP* 第 3.5 节 $\pi^e$ 还是 $e^\pi$？中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/pi-e-or-e-pi.html

## 关键思路

这里每一条不等式，都是比较 $p^q$ 与 $q^p$。取对数（logarithm）再除以 $pq$，就变成比较 $\frac{\ln p}{p}$ 与 $\frac{\ln q}{q}$；而 $\frac{\ln x}{x}$ 正是由原点连到曲线上点 $(x, \ln x)$ 的直线的斜率（gradient）。因此整道题都围绕与 $y = \ln x$ 相交的直线和弦（chord）。我们需要的事实是：这条曲线是凹（concave）的（即向下弯曲），而且由原点出发的切线（tangent）在 $x = \mathrm{e}$ 处与曲线相切。

## 解答

### 第 (i) 部分

设 $y = mx$ 在 $x = x_0$ 处与 $y = \ln x$ 相切。两者斜率相同，而且曲线相交，所以

$$m = \frac{1}{x_0}, \qquad mx_0 = \ln x_0.$$

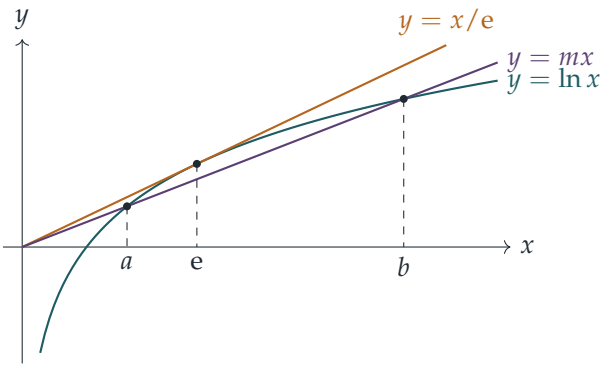
由第一条方程得 $mx_0 = 1$，所以 $\ln x_0 = 1$，$x_0 = \mathrm{e}$，以及

$$m = \frac{1}{\mathrm{e}}.$$

现在设 $y = mx$ 与曲线相交于 $x = a$ 及 $x = b$。那么 $\ln a = ma$ 及 $\ln b = mb$，所以（因为 $a, b > 0$）

$$\frac{\ln a}{a} = m = \frac{\ln b}{b} \implies b \ln a = a \ln b \implies \ln\left(a^b\right) = \ln(b^a),$$

由于 ln 是一一对应（one-to-one）的，故 $a^b = b^a$，得证。




关于草图：若 $m \le 0$，由于此时 $\ln x - mx$ 严格递增，直线与曲线最多只相交一次。若 $m \ge \frac{1}{\mathrm{e}}$，则对 $x > 0$ 而言，直线就是那条切线或位于切线之上；而凹曲线位于其切线 $y = x/\mathrm{e}$ 之下，只在 e 处相切，所以同样不会有两个交点。因此 $0 < m < \frac{1}{\mathrm{e}}$，直线 $y = mx$ 位于切线之下，如图所示。在 $x = \mathrm{e}$ 处，曲线的高度是 1，直线的高度是 $m\mathrm{e} < 1$，所以该处曲线在直线之上。由于曲线是凹的而直线是直的，$\ln x - mx$ 是凹函数。因此它取正值的集合是单一个区间，其两端正是两个交点 $a$ 和 $b$。这个区间包含 e，所以 $a < \mathrm{e} < b$。

**第 (ii) 部分**

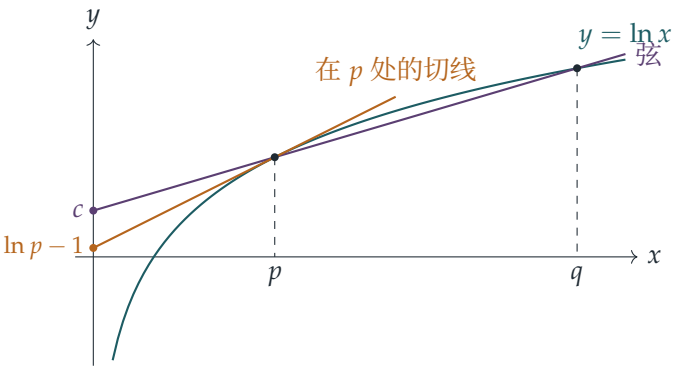
这里 $\ln p = mp + c$ 及 $\ln q = mq + c$，其中 $0 < p < q$。分别除以 $p$ 和 $q$，得

$$\frac{\ln p}{p} = m + \frac{c}{p}, \qquad \frac{\ln q}{q} = m + \frac{c}{q}.$$

由于 $c > 0$ 且 $\frac{1}{p} > \frac{1}{q}$，我们有 $\frac{c}{p} > \frac{c}{q}$，所以 $\frac{\ln p}{p} > \frac{\ln q}{q}$。乘以 $pq > 0$ 得 $q \ln p > p \ln q$，即 $\ln(p^q) > \ln(q^p)$，因此 $p^q > q^p$。

从图像上看，由原点连到 $(p, \ln p)$ 和 $(q, \ln q)$ 的直线，斜率分别是 $\frac{\ln p}{p}$ 和 $\frac{\ln q}{q}$。若弦在原点上方与 $y$ 轴相交，则较近的那一点与 $O$ 所连的直线较陡。（$c$ 的正负决定不等式的方向。我们在第 (iv) 部分会再用到这一点。）

### 第 (iii) 部分




设经过 $P = (p, \ln p)$ 和 $Q = (q, \ln q)$ 的弦的斜率为 $s$，$y$ 截距（intercept）为 $c$，所以 $c = \ln p - sp$。由于曲线是凹的，由 $P$ 连到更右方一点的弦，比 $P$ 处的切线平缓，所以

$$s < \frac{1}{p}.$$

（又或者，由中值定理（mean value theorem），对某个 $\xi \in (p, q)$ 有 $s = \frac{1}{\xi}$，而 $\frac{1}{\xi} < \frac{1}{p}$。）$P$ 处的切线是 $y = \ln p + \frac{1}{p}(x - p)$，它与 $y$ 轴相交于 $\ln p - 1$。弦和切线都经过 $P$，而弦较平缓。所以由 $P$ 向左移到 $x = 0$ 的过程中，弦一直在切线之上：

$$c = \ln p - sp > \ln p - 1 \geq 0,$$

最后一步用了 $p \geq \mathrm{e}$。所以弦在 $y$ 取正值之处与 $y$ 轴相交。

? **想一想** 这样如何判断 $\pi^{\mathrm{e}}$ 与 $\mathrm{e}^{\pi}$ 的大小?

取 $p = \mathrm{e}$ 和 $q = \pi$，它们满足 $\mathrm{e} \leq p < q$。这条弦的 $c > 0$，所以可应用第 (ii) 部分，得 $p^q > q^p$：

$$\mathrm{e}^{\pi} > \pi^{\mathrm{e}}.$$

### 第 (iv) 部分

现在经过 $(p, \ln p)$ 和 $(q, \ln q)$ 的弦的斜率是 $\frac{1}{\mathrm{e}}$，所以它是 $y = \frac{x}{\mathrm{e}} + c$，其中

$$c = \ln p - \frac{p}{\mathrm{e}}.$$

直线 $y = \frac{x}{\mathrm{e}}$ 就是第 (i) 部分中在 $x = \mathrm{e}$ 处的切线，而凹曲线位于它之上或之下——准确地说是在它上面或下方，即 $\ln x \leq \frac{x}{\mathrm{e}}$，等号只在 $x = \mathrm{e}$ 时成立。所以 $c \leq 0$，而 $c = 0$ 只在 $p = \mathrm{e}$ 时出现。但 $c = 0$ 会令弦就是切线本身，而切线只在 e 处与曲线相交，并非在两点 $p < q$ 相交。因此

$$c < 0.$$

这条弦与切线平行，并位于切线之下。现在把第 (ii) 部分的论证反转符号再做一次：

$$\frac{\ln p}{p} = \frac{1}{\mathrm{e}} + \frac{c}{p}, \qquad \frac{\ln q}{q} = \frac{1}{\mathrm{e}} + \frac{c}{q}.$$

由于 $c < 0$ 且 $\frac{1}{p} > \frac{1}{q}$，得 $\frac{c}{p} < \frac{c}{q}$，所以 $\frac{\ln p}{p} < \frac{\ln q}{q}$。乘以 $pq > 0$ 得 $q \ln p < p \ln q$，因此

$$q^p > p^q.$$

（与第 (i) 部分一样，凹性在这里也迫使 $p < \mathrm{e} < q$，但论证并不需要这一点。）

## 验算

- $\mathrm{e}^\pi \approx 23.14$ 而 $\pi^\mathrm{e} \approx 22.46$，与第 (iii) 部分一致。
- 第 (i) 部分：$a = 2$、$b = 4$ 这一对满足 $2^4 = 4^2$，而且确实 $\frac{\ln 2}{2} = \frac{\ln 4}{4}$ 及 $2 < \mathrm{e} < 4$。
- 第 (iv) 部分：取 $p = 1$，由弦的斜率 $\frac{\ln q}{q-1} = \frac{1}{\mathrm{e}}$ 得 $q \approx 5.76$。于是 $q^p \approx 5.76 > 1 = p^q$。
- 第 (ii) 和第 (iii) 部分亦已用电脑对大量随机的弦作检验确认。

## 常见错误

- 把 $q \ln p > p \ln q$ 的方向弄反。乘以 $pq$ 之前，先清楚写出 $\frac{\ln p}{p} > \frac{\ln q}{q}$。
- 在第 (iii) 部分，只写到“曲线是凹的”便停下，没有解释为何截距是正数。$p$ 处的切线与轴相交于 $\ln p - 1 \geq 0$，而弦在它之上穿过。
- 在第 (iv) 部分，忘记排除 $c = 0$，而这是得到严格不等式所必需的。

## STEP 之外

函数 $\frac{\ln x}{x}$ 在 e 处取得最大值，这解释了著名的一对 $2^4 = 4^2$。方程 $x^y = y^x$ 在 $x \neq y$ 时的正数解构成一条曲线，可以参数化（parametrise）为 $x = t^{1/(t-1)}$、$y = t^{t/(t-1)}$。更一般地说，“凹函数的弦位于曲线之下、切线之上”是凸性（convexity）的基本不等式。它是延森不等式（Jensen's inequality）和算术–几何平均不等式（AM–GM inequality）的基础。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）对中值定理和凸函数有仔细的论述。

详细解答

# STEP II 2013，第 2 题

纯数学 · 归约公式

**STEP II 2013，第 2 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q2

## 关键思路

被积函数（integrand）$x^n(1-x)^n$ 在反射 $x \mapsto 1-x$ 下对称（symmetric），而这种对称性让我们可以毫无代价地互换 $x$ 与 $1-x$ 的角色。接着用分部积分（integration by parts）把 $x$ 的一个幂换成 $1-x$ 的一个幂，再利用恒等式 $x+(1-x)=1$ 把乘积拆回我们已知的积分。两者结合便得到一条递推公式（reduction formula）；对整数 $n$，它可以逐步展开成阶乘（factorial），而对半整数（half-integer）也同样适用。

## 解答

### 第 (i) 部分

对 $n \geq 1$，记 $J_n = \int_0^1 x^{n-1}(1-x)^n \,\mathrm{d}x$。代换（substitute）$u = 1-x$，于是 $\mathrm{d}x = -\mathrm{d}u$，而积分限 $0, 1$ 变成 $1, 0$：

$$J_n = \int_1^0 (1-u)^{n-1}u^n\,(-\mathrm{d}u) = \int_0^1 u^n(1-u)^{n-1}\,\mathrm{d}u,$$

这就是所要求的等式（哑变量（dummy variable）用甚么名字并不重要）。

把 $J_n$ 的两个相等形式相加，

$$\begin{aligned} 2J_n &= \int_0^1 \left[x^{n-1}(1-x)^n + x^n(1-x)^{n-1}\right]\mathrm{d}x \\ &= \int_0^1 x^{n-1}(1-x)^{n-1}\left[(1-x)+x\right]\mathrm{d}x = I_{n-1}. \end{aligned}$$

接着对 $I_n$ 作分部积分，取 $u = x^n$ 及 $\mathrm{d}v = (1-x)^n\,\mathrm{d}x$，故 $v = -\dfrac{(1-x)^{n+1}}{n+1}$：

$$I_n = \Big[-\frac{x^n(1-x)^{n+1}}{n+1}\Big]_0^1 + \frac{n}{n+1}\int_0^1 x^{n-1}(1-x)^{n+1}\,\mathrm{d}x.$$

边界项（boundary term）在两端都等于零：在 $x=1$ 处是因为有因子 $(1-x)^{n+1}$，在 $x=0$ 处则是因为有因子 $x^n$ 且 $n \geq 1$。由此得到题目所给的公式

$$I_n = \frac{n}{n+1}\int_0^1 x^{n-1}(1-x)^{n+1}\,\mathrm{d}x.$$

**想一想** 怎样才能从这个积分回到 $I_n$ 和 $I_{n-1}$?

分出一个因子 $1-x$，并照字面写成 $1-x$：

$$x^{n-1}(1-x)^{n+1} = x^{n-1}(1-x)^n\,(1-x) = x^{n-1}(1-x)^n - x^n(1-x)^n.$$

积分后，利用上面的结果，右边等于 $J_n - I_n = \frac{1}{2}I_{n-1} - I_n$。所以

$$I_n = \frac{n}{n+1}\Big(\tfrac{1}{2}I_{n-1} - I_n\Big) \quad\Longrightarrow\quad I_n \cdot \frac{2n+1}{n+1} = \frac{n}{2(n+1)}I_{n-1},$$

再乘以 $\dfrac{n+1}{2n+1}$，

$$I_n = \frac{n}{2(2n+1)}\,I_{n-1},$$

正是所求。留意这个论证完全没有用到 $n$ 是整数：只需要 $n \geq 1$（为了处理边界项，以及确保每个被积函数在 $[0,1]$ 上连续）。我们会在第 (iii) 部分用到这一点。

### 第 (ii) 部分

我们用归纳法（induction）证明 $I_n = \dfrac{(n!)^2}{(2n+1)!}$，从 $n=0$ 开始：$I_0 = \int_0^1 1\,\mathrm{d}x = 1 = \dfrac{(0!)^2}{1!}$。假设公式对 $n-1$ 成立，其中 $n \geq 1$。那么由第 (i) 部分，

$$\begin{aligned} I_n &= \frac{n}{2(2n+1)} \cdot \frac{\big((n-1)!\big)^2}{(2n-1)!} = \frac{n^2\big((n-1)!\big)^2}{2n(2n+1)(2n-1)!} \\ &= \frac{(n!)^2}{(2n+1)!}, \end{aligned}$$

这里我们把分子和分母同乘以 $n$，并用了 $2n(2n+1)(2n-1)! = (2n+1)!$。归纳至此完成，所以公式对每个正整数 $n$（以及 $n=0$）都成立。

### 第 (iii) 部分

取 $x = \sin^2\theta$，则 $\mathrm{d}x = 2\sin\theta\cos\theta\,\mathrm{d}\theta$，而 $\theta$ 由 0 变到 $\frac{\pi}{2}$。在这个区间上 $\sin\theta$ 和 $\cos\theta$ 都非负，所以

$$\big(x(1-x)\big)^{1/2} = \big(\sin^2\theta\cos^2\theta\big)^{1/2} = \sin\theta\cos\theta.$$

**想一想** 为甚么这里需要 $\sin\theta\cos\theta \geq 0$?

因为 $\sqrt{t^2} = |t|$，而不是 $t$；在 $[0, \frac{\pi}{2}]$ 上，绝对值（modulus）符号可以直接去掉。因此

$$\begin{aligned} I_{\frac{1}{2}} &= \int_0^{\pi/2} \sin\theta\cos\theta\cdot 2\sin\theta\cos\theta\,\mathrm{d}\theta = \int_0^{\pi/2} \tfrac{1}{2}\sin^2 2\theta\,\mathrm{d}\theta \\ &= \int_0^{\pi/2} \tfrac{1}{4}\big(1-\cos 4\theta\big)\mathrm{d}\theta = \tfrac{1}{4}\Big[\theta - \tfrac{1}{4}\sin 4\theta\Big]_0^{\pi/2} = \frac{\pi}{8}. \end{aligned}$$

至于 $I_{\frac{3}{2}}$，用第 (i) 部分的递推公式并取 $n = \frac{3}{2}$；这是容许的，因为 $\frac{3}{2} \geq 1$，而推导过程从来不需要 $n$ 是整数：

$$I_{\frac{3}{2}} = \frac{3/2}{2\cdot 4} I_{\frac{1}{2}} = \frac{3}{16}\cdot\frac{\pi}{8} = \frac{3\pi}{128}.$$

作为交叉验证，同一代换可直接得到

$$I_{\frac{3}{2}} = \int_0^{\pi/2} 2\sin^4\theta\cos^4\theta\,\mathrm{d}\theta = \frac{1}{8}\int_0^{\pi/2} \sin^4 2\theta\,\mathrm{d}\theta = \frac{1}{8}\cdot\frac{3\pi}{16} = \frac{3\pi}{128},$$

其中用了 $\sin^4\phi = \frac{3}{8} - \frac{1}{2}\cos 2\phi + \frac{1}{8}\cos 4\phi$。

## 验算

- $I_1 = \int_0^1 (x - x^2)\,\mathrm{d}x = \frac{1}{6}$，而公式给出 $\frac{1}{3!} = \frac{1}{6}$。同样，$I_2 = \int_0^1 (x^2 - 2x^3 + x^4)\,\mathrm{d}x = \frac{1}{3} - \frac{1}{2} + \frac{1}{5} = \frac{1}{30} = \frac{4}{5!}$。
- $I_{\frac{1}{2}}$ 是半径为 $\frac{1}{2}$ 的半圆（semicircle）$y = \sqrt{\frac{1}{4} - (x - \frac{1}{2})^2}$ 下方的面积，即 $\frac{1}{2}\pi\big(\frac{1}{2}\big)^2 = \frac{\pi}{8}$。
- 数值上，$I_{\frac{3}{2}} \approx 0.073631$，与 $3\pi/128$ 吻合；递推公式亦已用电脑对非整数 $n$（例如 $n = 2.7$）验证过。

## 常见错误

- 忘记检查分部积分中的边界项是否等于零，尤其是在 $x = 0$ 处。
- 写 $\sqrt{\sin^2\theta\cos^2\theta} = \sin\theta\cos\theta$ 时没有指出在 $[0, \frac{\pi}{2}]$ 上的正负号。
- 用第 (ii) 部分的阶乘公式计算 $I_{\frac{3}{2}}$，但该公式只对整数证明过；用递推公式才是稳妥的做法。

## STEP 之外

$I_n$ 是贝塔函数（beta function）$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x$ 的特例，这里 $I_n = B(n+1, n+1)$，而第 (ii) 部分正是 $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ 的整数情况。把 $n = \frac{1}{2}$ 代入这个恒等式，便可得出 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$，这就是第 (iii) 部分出现 $\pi$ 的更深层原因。这类积分是统计学中贝塔分布（beta distribution）以及沃利斯（Wallis）的 $\pi$ 乘积公式背后的基础；Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多同类风格的递推公式题目。

详细解答

# STEP II 2013，第 3 题

纯数学 · 三次方程的根

**STEP II 2013，第 3 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q3

## 关键思路

一切都源于首项系数为正、且有三个相异实根的三次函数（cubic）的形状：曲线从 $-\infty$ 上升，向上穿过横轴，回落，然后再次向上穿过横轴。数值 $f(0)=c$ 告诉我们 0 相对于各根的位置，而转向点（turning points）（即二次方程 $f'(x)=0$ 的根）则可以通过它们的和与积来确定。最后一部分提醒我们：单靠转向点的条件，并不能保证曲线真的与横轴相交三次。

## 解答

全文中设

$$f(x)=x^3+3ax^2+3bx+c, \qquad f'(x)=3\big(x^2+2ax+b\big).$$

当 $f(x)=0$ 有三个相异实根 $\alpha<\beta<\gamma$ 时，我们可以写成 $f(x)=(x-\alpha)(x-\beta)(x-\gamma)$，于是从三个因式的正负号可读出

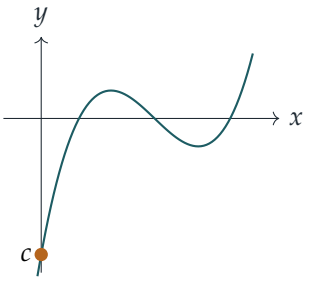
$$f(x)<0 \text{ 于 } (-\infty,\alpha)\cup(\beta,\gamma), \qquad f(x)>0 \text{ 于 } (\alpha,\beta)\cup(\gamma,\infty).$$

此外，由罗尔定理（Rolle's theorem），$f'$ 在 $(\alpha,\beta)$ 内有一个零点，在 $(\beta,\gamma)$ 内有另一个零点。由于 $f'$ 是二次函数，它再没有其他零点，因此转向点位于 $p<q$，且

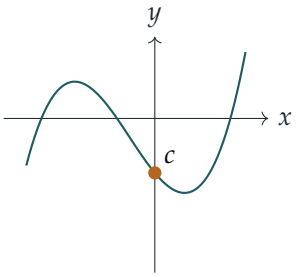
$$\alpha<p<\beta<q<\gamma.$$

### 第 (i) 部分

曲线 $y=f(x)$ 与 $y$ 轴相交于 $(0,c)$；由于 $c<0$，这点位于 $x$ 轴下方。根据上面的正负号规律，$f(0)<0$ 表示 $0<\alpha$ 或 $\beta<0<\gamma$。这就是下面的两幅图。



三个根皆为正



恰有一个根为正

若 $0 < \alpha$，则三个根皆为正。若 $\beta < 0 < \gamma$，则 $\alpha < \beta < 0$，因此恰有一个根（$\gamma$）为正。由于 $f(0) = c \neq 0$，没有根等于零。所以要么恰有一个根为正，要么三个根皆为正，得证。

### 第 (ii) 部分

现在 $0 < \alpha < \beta < \gamma$。

$c$ 的正负号。由于 $0 < \alpha$，点 0 位于 $(-\infty, \alpha)$ 内，而在该区间上 $f < 0$。所以 $c = f(0) < 0$。

转向点。方程 $x^2 + 2ax + b = 0$ 有两个相异实根 $p$ 和 $q$，因此其判别式（discriminant）为正：

$$4a^2 - 4b > 0 \quad \Longrightarrow \quad a^2 > b.$$

由于 $p > \alpha > 0$ 且 $q > p$，两个转向点皆为正。对于二次式 $x^2 + 2ax + b$，两根之和与积分别为

$$p + q = -2a, \qquad pq = b.$$

由于 $p, q > 0$，得 $-2a > 0$，所以 $a < 0$，且 $b = pq > 0$。综合以上，

$$a^2 > b > 0, \qquad a < 0, \qquad c < 0,$$

这就是 $(*)$。

（你也可以直接从根得出 $a < 0$、$b > 0$ 和 $c < 0$：$\alpha + \beta + \gamma = -3a$、$\alpha\beta + \beta\gamma + \gamma\alpha = 3b$ 和 $\alpha\beta\gamma = -c$ 皆为正。但 $a^2 > b$ 确实需要用到转向点。）

### 第 (iii) 部分

现在 $c = f(0) > 0$，所以 0 位于 $(\alpha, \beta)$ 或 $(\gamma, \infty)$ 内，即 $f > 0$ 的两个区间。这给出两幅可能的图：要么 $\alpha < 0 < \beta < \gamma$（一个负根、两个正根），要么 $\alpha < \beta < \gamma < 0$（三个根皆为负）。

? **想一想** 条件 $ab < 0$ 怎样排除其中一种情况?

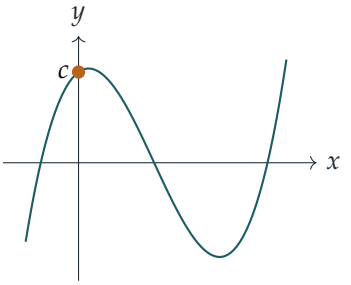
再看看转向点。若三个根皆为负，则 $p<q<\gamma<0$，所以两个转向点皆为负。于是

$$-2a=p+q<0 \quad 及 \quad b=pq>0,$$

所以 $a>0$ 且 $b>0$，得 $ab>0$。这与 $ab<0$ 矛盾。因此第二幅图不可能出现，故

$$\alpha<0<\beta<\gamma:$$

一个根为负，另外两个根为正。




这种情况确实可以出现。例如，根 $-1$、2 和 5 给出 $x^3-6x^2+3x+10$，所以 $a=-2$、$b=1$、$c=10$，其中 $ab=-2<0$ 且 $c>0$；这就是上面所绘的曲线。

**第 (iv) 部分**

**想一想** 除了 $(*)$ 之外，三次函数还需要甚么额外条件才有三个实根?

条件 $(*)$ 只控制转向点和 $f(0)$ 的正负号。要有三个实根，局部极大值（local maximum）还必须位于横轴上方。所以我们应选取 $a$ 和 $b$ 使两个转向点皆为正，然后令 $c$ 足够负，把局部极大值拉到横轴下方。

取 $a=-1$ 及 $b=\frac{3}{4}$，使得

$$f'(x)=3\left(x^2-2x+\tfrac{3}{4}\right)=3\left(x-\tfrac{1}{2}\right)\left(x-\tfrac{3}{2}\right).$$

局部极大值位于 $x=\frac{1}{2}$，该处

$$f\left(\tfrac{1}{2}\right)=\tfrac{1}{8}-\tfrac{3}{4}+\tfrac{9}{8}+c=\tfrac{1}{2}+c.$$

选取 $c=-1$。方程于是为

$$x^3-3x^2+\tfrac{9}{4}x-1=0 \qquad \left(a=-1,\ b=\tfrac{3}{4},\ c=-1\right),$$

且 $a^2=1>\frac{3}{4}=b>0$、$a<0$、$c<0$，所以 $(*)$ 成立。

现在数一数根的个数。$f$ 在 $(-\infty, \frac{1}{2}]$ 上递增，所以在该处 $f(x) \le f(\frac{1}{2}) = -\frac{1}{2} < 0$。在 $[\frac{1}{2}, \frac{3}{2}]$ 上，$f$ 递减，所以一直低于 $-\frac{1}{2}$。在 $[\frac{3}{2}, \infty)$ 上，$f$ 从 $f(\frac{3}{2}) = -1$ 递增至 $+\infty$，所以恰好穿过横轴一次。因此恰有一个实根（由于 $f(2) = -\frac{1}{2}$ 及 $f(3) = \frac{23}{4}$，它位于 2 与 3 之间）。

## 验算

- 三次式 $(x-1)(x-2)(x-3) = x^3 - 6x^2 + 11x - 6$ 的 $a = -2$、$b = \frac{11}{3}$、$c = -6$，而的确 $a^2 = 4 > \frac{11}{3} > 0$，正如第 (ii) 部分所预测。
- 把曲线反射，即 $x \mapsto -x$，会把根 $\alpha, \beta, \gamma$ 变为 $-\alpha, -\beta, -\gamma$，并改变 $a$ 和 $c$ 的正负号，但不改变 $b$。所以 $ab$ 和 $c$ 都改变正负号，而第 (i) 部分的镜像版本说明：当 $c > 0$ 时，要么恰有一个根为负，要么三个根皆为负——这正是我们在第 (iii) 部分开始时的两幅图。
- 在第 (iv) 部分中，唯一的实根是 $x \approx 2.178$。以电脑检查数以千计随机生成、具三个相异实根的三次式，结果亦与第 (i)–(iii) 部分吻合。

## 常见错误

- 试图单从根的正负号证明 $a^2 > b$。这个不等式其实来自转向点是相异实数这一事实。
- 以为 $(*)$ 等价于有三个正根。第 (iv) 部分说明它只是必要条件（necessary），而非充分条件（sufficient）。
- 在第 (iv) 部分中，选取的 $c$ 使局部极小值为负，但局部极大值仍为正。这样会得到三个根，而不是一个。

## STEP 之外

这里的正负号论证是一套一般理论的起点：笛卡儿符号法则（Descartes' rule of signs）以系数中正负号改变的次数作为正根个数的上界，而斯图姆定理（Sturm's theorem）则能准确计算某区间内实根的个数。对于三次函数，第 (iv) 部分所缺少的条件是判别式的正负号；判别式（相差一个正因子）就是两个转向点处函数值的乘积 $-f(p)f(q)$。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）以大学一年级程度讨论多项式及其根。

详细解答

# STEP II 2013，第 4 题

纯数学 · 弦中点的轨迹

**STEP II 2013，第 4 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q4

## 关键思路

求弦（chord）的中点（midpoint）时，根本不必求出弦的两个端点：$P$ 和 $Q$ 的 $x$ 坐标是某个二次方程（quadratic equation）的根，而它们的平均值就是首两项系数之比的一半再取负号。之后，每个轨迹（locus）都只是消去一个参数（parameter）的问题。每个答案都可以用几何解释：由圆心到弦中点的直线与该弦垂直（perpendicular）。

## 解答

### $M$ 的坐标

直线为 $y = b(x - a)$。代入 $x^2 + y^2 = 1$，得

$$(1 + b^2)x^2 - 2ab^2x + a^2b^2 - 1 = 0.$$

它的根是 $P$ 和 $Q$ 的 $x$ 坐标 $x_P$、$x_Q$，所以由根之和（sum of the roots）得

$$x_M = \tfrac{1}{2}(x_P + x_Q) = \frac{ab^2}{1 + b^2}.$$

由于 $M$ 在该直线上，$y_M = b(x_M - a) = b \cdot \dfrac{ab^2 - a - ab^2}{1 + b^2}$，所以

$$M = \left(\frac{ab^2}{1 + b^2},\ -\frac{ab}{1 + b^2}\right).$$

要 $P$ 和 $Q$ 存在且互不相同，判别式（discriminant）必须为正：

$$4a^2b^4 - 4(1 + b^2)(a^2b^2 - 1) = 4\left(1 + b^2 - a^2b^2\right) > 0 \iff a^2b^2 < 1 + b^2.$$

（这表示由 $O$ 到该直线的距离 $|ab|/\sqrt{1 + b^2}$ 小于 1。）每一部分都需要用到这个条件。

**第 (i) 部分**

固定 $b > 0$，则

$$-by_M = -b \cdot \left(-\frac{ab}{1+b^2}\right) = \frac{ab^2}{1+b^2} = x_M,$$

所以轨迹上每一点都满足 $x = -by$，正如所求。这是通过 $O$、斜率（gradient）为 $-1/b$ 的直线，与弦垂直，与关键思路的预期完全一致。

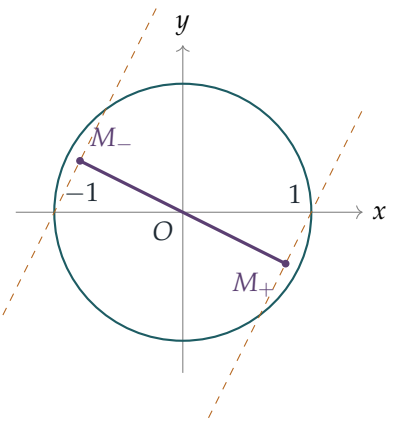
若 $-1 \le a \le 1$，则 $a^2b^2 \le b^2 < 1 + b^2$，所以直线必与圆相交于两点，对每个这样的 $a$，$M$ 都有定义。$M$ 的两个坐标都是 $a$ 乘以一个常数，所以当 $a$ 取遍 $[-1,1]$ 时，点 $M$ 沿直线 $x = -by$ 由 $a = -1$ 对应的点稳定地移动到 $a = 1$ 对应的点。因此轨迹是连接

$$M_\pm = \pm\left(\frac{b^2}{1+b^2}, -\frac{b}{1+b^2}\right),$$

的线段，其长度为

$$2\,\frac{\sqrt{b^4+b^2}}{1+b^2} = \frac{2b\sqrt{1+b^2}}{1+b^2} = \frac{2b}{(1+b^2)^{1/2}},$$

这里利用了 $b > 0$ 来开 $b^2$ 的平方根。这就是所求的长度。线段以 $O$ 为中点；其端点 $M_+$ 是通过 $(1,0)$ 的弦的中点，而 $M_-$ 则是通过 $(-1,0)$ 的弦的中点。下图显示 $b = 2$ 的情况，此时端点为 $\pm(\frac{4}{5}, -\frac{2}{5})$。




虚线是两条极端的弦，分别通过 $(\pm 1, 0)$，斜率为 2。

**第 (ii)(a) 部分**

现在固定 $a$，且 $0 < a < 1$。于是 $a^2b^2 < b^2 < 1 + b^2$，所以每条通过 $A = (a, 0)$ 的直线都与圆相交两次。由 $M$ 的公式，

$$x^2 + y^2 = \frac{a^2b^4 + a^2b^2}{(1+b^2)^2} = \frac{a^2b^2}{1+b^2} = a \cdot \frac{ab^2}{1+b^2} = ax.$$

所以 $M$ 在

$$x^2+y^2-ax=0,\quad \text{即}\quad \left(x-\tfrac{a}{2}\right)^2+y^2=\left(\tfrac{a}{2}\right)^2,$$

上，这是圆心为 $(\frac{a}{2},0)$、半径为 $\frac{a}{2}$ 的圆，以 $OA$ 为直径（diameter）。

**想一想** 这个圆是否整个都被描出?

并不完全。由于 $x=ab^2/(1+b^2)=a-a/(1+b^2)$，总有 $0\le x<a$，所以点 $A=(a,0)$ 本身永远不会到达；要到达它，需要通过 $A$ 的铅垂线，而铅垂线没有斜率。其他每一点都能到达：设 $b=\tan\theta$，其中 $-\frac{\pi}{2}<\theta<\frac{\pi}{2}$，

$$M=\left(a\sin^2\theta,\ -a\sin\theta\cos\theta\right)=\left(\tfrac{a}{2}-\tfrac{a}{2}\cos 2\theta,\ -\tfrac{a}{2}\sin 2\theta\right),$$

当 $2\theta$ 取遍 $(-\pi,\pi)$ 时，这恰好走遍整个圆一次，只差 $2\theta=\pm\pi$ 对应的点，即 $A$。所以轨迹是以 $OA$ 为直径的圆，但除去单独一点 $A$。从几何角度看：$OM$ 与弦 $AM$ 垂直，所以 $\angle OMA$ 是直角，$M$ 在以 $OA$ 为直径的圆上（当 $b=0$ 时 $M=O$）。这个圆位于单位圆之内。

### 第 (ii)(b) 部分

现在 $a=1/b$，其中 $0<b\le 1$。条件 $a^2b^2<1+b^2$ 变成 $1<1+b^2$，此式成立。直线为 $y=b(x-1/b)=bx-1$，所以它必定通过单位圆上的点 $B=(0,-1)$；因此 $B$ 必定是 $P$、$Q$ 其中之一。公式给出

$$M=\left(\frac{b}{1+b^2},\ -\frac{1}{1+b^2}\right),\qquad x^2+y^2=\frac{b^2+1}{(1+b^2)^2}=\frac{1}{1+b^2}=-y.$$

所以 $M$ 在

$$x^2+y^2+y=0,\quad \text{即}\quad x^2+\left(y+\tfrac{1}{2}\right)^2=\left(\tfrac{1}{2}\right)^2,$$

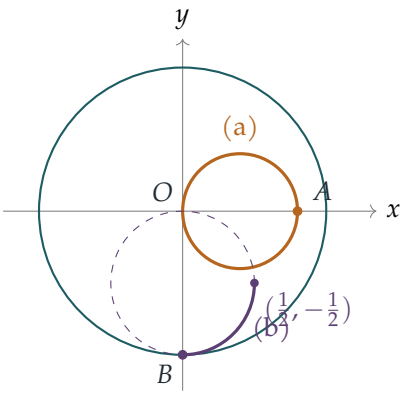
上，这是圆心为 $(0,-\frac{1}{2})$、半径为 $\frac{1}{2}$ 的圆，以 $OB$ 为直径（理由与 (a) 中的直角相同）。

这次只描出圆的一部分。设 $b=\tan\theta$，$0<\theta\le\frac{\pi}{4}$，

$$M=\left(\tfrac{1}{2}\sin 2\theta,\ -\tfrac{1}{2}-\tfrac{1}{2}\cos 2\theta\right),$$

而 $2\theta$ 取遍 $(0,\frac{\pi}{2}]$。所以轨迹是四分之一个圆：由 $B=(0,-1)$（不包括在内，因为它需要 $b=0$）沿逆时针方向到 $(\frac{1}{2},-\frac{1}{2})$（包括在内，由 $b=1$、$a=1$ 得到）的圆弧。在这段弧上，$0<x\le\frac{1}{2}$ 且 $-1<y\le-\frac{1}{2}$。

下图连同单位圆显示两个轨迹，(a) 中取 $a=0.8$。空心圆点标示 (a) 中缺少的点 $A$ 和 (b) 中缺少的端点 $B$。




在 (b) 中，实线弧是轨迹；小圆其余的虚线部分不会到达。

## 验算

- 当 $a = 0$ 时，直线通过 $O$，所以弦是直径，$M$ 应该是 $O$：公式给出 $(0, 0)$。
- 在 (i) 中取 $b = 1$：通过 $(1, 0)$、斜率为 1 的弦连接 $(1, 0)$ 和 $(0, -1)$，中点为 $(\frac{1}{2}, -\frac{1}{2})$，与 $M_+$ 吻合。此时线段长度为 $\sqrt{2} = 2/\sqrt{2}$，与公式一致。
- 同一条弦正是 (ii)(b) 中圆弧在 $b = 1$ 的端点，而 $(\frac{1}{2}, -\frac{1}{2})$ 满足 $x^2 + y^2 + y = \frac{1}{4} + \frac{1}{4} - \frac{1}{2} = 0$。
- 在 (ii)(a) 中，$b = 1$ 给出 $M = (\frac{a}{2}, -\frac{a}{2})$，即以 $OA$ 为直径的圆的最低点，对于倾角 45° 的弦，理应如此。

## 常见错误

- 用二次方程公式明确求出 $P$ 和 $Q$。这样做可行，但用根之和一行就能得出 $M$。
- 在 (ii)(a) 或 (ii)(b) 中声称轨迹是整个圆。必须检查参数实际取哪些值，以及这些值对应哪些点。
- 忘记检查直线是否真的与圆相交；在 (i) 和 (ii) 的两种情况中，它都必定相交，但这一点需要说明。

## STEP 之外

平行弦的中点位于一条通过中心的直线上，这个事实对任何圆锥曲线 (conic) 都成立：对椭圆而言，这条直线称为*共轭直径*（conjugate diameter），这也让我们初步窥见射影几

何中极点与极线（poles and polars）的概念。第 (ii) 部分中的圆，其实就是经典定理“半圆上的圆周角是直角”反过来运用。若想进一步了解轨迹和以参数化（parametrisation）作为解题工具，Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）是很好的下一步。

详细解答

# STEP II 2013，第 5 题

纯数学 · 对称性与曲线描绘

> **STEP II 2013，第 5 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s2-q5

## 关键思路

如果一个函数在 $x \mapsto 1-x$ 或 $x \mapsto 1/x$ 这类代换下保持不变，那么用链式法则（chain rule）对该恒等式求导，便能把 $x$ 处的斜率与像点处的斜率联系起来。在代换的不动点（fixed point）上，斜率必须等于自身的相反数，因此为零。至于第 (iii) 部分的方程，对称性提示我们把一切以 $u = x^2 - x$ 表示，而它本身在 $x \mapsto 1-x$ 下不变；这样一来，处理关于 $u$ 的三次方程，远比处理关于 $x$ 的六次方程容易。

## 解答

### 第 (i) 部分

把 $\mathrm{f}(x) = \mathrm{f}(1-x)$ 对 $x$ 求导，右边使用链式法则，得

$$\mathrm{f}'(x) = -\mathrm{f}'(1-x) \qquad \text{对所有} x.$$

代入 $x = \frac{1}{2}$，得 $\mathrm{f}'(\frac{1}{2}) = -\mathrm{f}'(\frac{1}{2})$，因此 $\mathrm{f}'(\frac{1}{2}) = 0$。

现在再假设当 $x \neq 0$ 时 $\mathrm{f}(x) = \mathrm{f}(1/x)$。由于 $\frac{\mathrm{d}}{\mathrm{d}x}(\frac{1}{x}) = -\frac{1}{x^2}$，求导得

$$\mathrm{f}'(x) = -\frac{1}{x^2}\,\mathrm{f}'\left(\frac{1}{x}\right) \qquad (x \neq 0).$$

在 $x = -1$（它是自身的倒数）处，此式变为 $\mathrm{f}'(-1) = -\mathrm{f}'(-1)$，因此 $\mathrm{f}'(-1) = 0$。

在 $x = 2$ 处，由第一个结果，此式给出 $\mathrm{f}'(2) = -\frac{1}{4}\mathrm{f}'(\frac{1}{2}) = 0$。（同样地，在第一个恒等式中取 $x = 2$，可得 $\mathrm{f}'(2) = -\mathrm{f}'(-1) = 0$。）

**第 (ii) 部分**

以 $1/x$ 代替 $x$。分子变为

$$\left(\frac{1}{x^2}-\frac{1}{x}+1\right)^3=\frac{(1-x+x^2)^3}{x^6},$$

而分母变为

$$\left(\frac{1}{x^2}-\frac{1}{x}\right)^2=\frac{(1-x)^2}{x^4}.$$

相除，并利用 $x^2(1-x)^2=(x^2-x)^2$，得

$$\mathrm{f}\left(\frac{1}{x}\right)=\frac{(x^2-x+1)^3}{x^6}\cdot\frac{x^4}{(1-x)^2}=\frac{(x^2-x+1)^3}{x^2(x-1)^2}=\mathrm{f}(x).$$

以 $1-x$ 代替 $x$：由于 $(1-x)^2-(1-x)=(1-x)(-x)=x^2-x$，所以 $x^2-x$ 不变，从而 $x^2-x+1$ 亦不变。因此 $\mathrm{f}(1-x)=\mathrm{f}(x)$。

**草图**。由第 (i) 部分，$x=-1$、$\frac{1}{2}$ 和 2 是驻点（stationary point）；由于驻点恰好有三个，这些就是全部。它们的函数值为

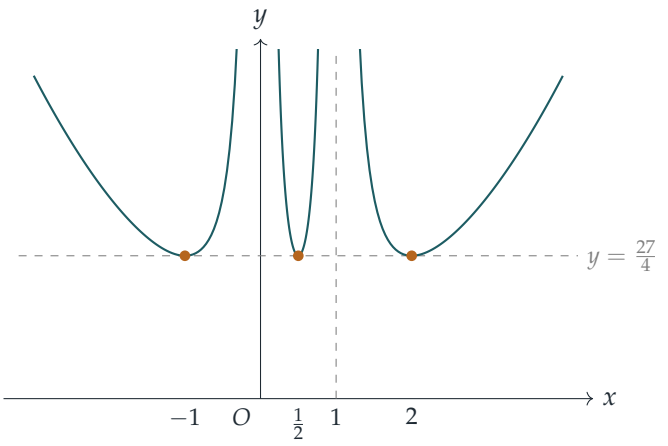
$$\mathrm{f}(-1)=\frac{3^3}{2^2}=\frac{27}{4},\qquad \mathrm{f}(\tfrac{1}{2})=\frac{(3/4)^3}{(1/4)^2}=\frac{27}{4},\qquad \mathrm{f}(2)=\frac{3^3}{2^2}=\frac{27}{4}.$$

三者相同，这是必然的：$\mathrm{f}(2)=\mathrm{f}(\frac{1}{2})$ 且 $\mathrm{f}(2)=\mathrm{f}(1-2)=\mathrm{f}(-1)$。

其他特征：

- $x^2-x+1=(x-\frac{1}{2})^2+\frac{3}{4}>0$，所以处处有 $\mathrm{f}(x)>0$；曲线永不与 $x$ 轴相交。
- 当 $x\to 0$ 或 $x\to 1$ 时，分子趋向 1，分母趋向 $0^+$，所以从两侧皆有 $\mathrm{f}(x)\to+\infty$：$x=0$ 和 $x=1$ 是铅直渐近线（vertical asymptote）。
- 当 $|x|$ 很大时，$\mathrm{f}(x)\approx x^6/x^4=x^2$，所以当 $x\to\pm\infty$ 时 $\mathrm{f}(x)\to+\infty$，大致像一条抛物线。

因此在 $(-\infty,0)$、$(0,1)$、$(1,\infty)$ 这三个区间的每一个上，函数都是连续的，并在两端都趋向 $+\infty$。所以每个区间都必定包含一个最小值点，而它是驻点；每个区间恰好包含 $-1$、$\frac{1}{2}$、2 其中之一，所以该点就是最小值点，且没有其他转向点（turning point）。每一支都呈 U 形，最低值为 $\frac{27}{4}$。




**第 (iii) 部分**

**方程** $\mathrm{f}(x) = \frac{27}{4}$。从草图可见，$\frac{27}{4}$ 是每一支上的最小值，而且只在三个驻点处取得。所以根为

$$x = -1, \qquad x = \tfrac{1}{2}, \qquad x = 2,$$

而对定义域内其他所有 $x$ 都有 $\mathrm{f}(x) > \frac{27}{4}$，即对于

$$x < -1, \quad -1 < x < 0, \quad 0 < x < \tfrac{1}{2}, \quad \tfrac{1}{2} < x < 1, \quad 1 < x < 2, \quad x > 2.$$

我们可以用代数方法确认这一点。设 $u = x^2 - x$（故在定义域内 $u \neq 0$）；则 $\mathrm{f}(x) = \dfrac{(u+1)^3}{u^2}$，且

$$\mathrm{f}(x) - \frac{27}{4} = \frac{4u^3 - 15u^2 + 12u + 4}{4u^2} = \frac{(u-2)^2(4u+1)}{4u^2}.$$

由于 $4u + 1 = 4x^2 - 4x + 1 = (2x-1)^2 \geq 0$，此式永不为负，而它为零当且仅当 $u = 2$（即 $x^2 - x - 2 = 0$，所以 $x = 2$ 或 $-1$）或 $x = \frac{1}{2}$。

? **想一想** 怎样才能不解六次方程，就猜出 $\mathrm{f}(x) = \frac{343}{36}$ 的一个根?

**方程** $\mathrm{f}(x) = \frac{343}{36}$。注意 $\frac{343}{36} = \frac{7^3}{6^2}$，所以我们寻找满足 $x^2 - x + 1 = 7$ 和 $x^2 - x = 6$ 的 $x$。$x = 3$ 两者皆满足，所以 $\mathrm{f}(3) = \frac{343}{36}$。对称性现在免费送上更多的根：对 3 反复施行 $x \mapsto 1 - x$ 和 $x \mapsto 1/x$，得到

$$3, \quad 1 - 3 = -2, \quad \tfrac{1}{3}, \quad 1 - \tfrac{1}{3} = \tfrac{2}{3}, \quad \tfrac{3}{2}, \quad 1 - \tfrac{3}{2} = -\tfrac{1}{2}.$$

这是六个不同的数，而去分母后，$\mathrm{f}(x) = \frac{343}{36}$ 会变成一个六次多项式方程，所以这些就是全部的根：

$$x = -2, -\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{3}{2}, 3.$$

（以 $u$ 表示，方程为 $36u^3 - 235u^2 + 108u + 36 = 0$，可因式分解为 $(u-6)(9u+2)(4u-3) = 0$，得出数对 $\{3,-2\}$、$\{\frac{1}{3},\frac{2}{3}\}$ 和 $\{\frac{3}{2},-\frac{1}{2}\}$。）

由于 $\frac{343}{36} > \frac{27}{4}$，水平线 $y = \frac{343}{36}$ 与每一支 U 形曲线各相交两次，在最小值两侧各一次，而在这两个交点以外，曲线位于直线上方。$(-\infty,0)$ 上的一支交于 $-2$ 和 $-\frac{1}{2}$，$(0,1)$ 上的一支交于 $\frac{1}{3}$ 和 $\frac{2}{3}$，$(1,\infty)$ 上的一支交于 $\frac{3}{2}$ 和 3。因此 $\mathrm{f}(x) > \frac{343}{36}$ 当且仅当

$$x < -2, \quad -\tfrac{1}{2} < x < 0, \quad 0 < x < \tfrac{1}{3}, \quad \tfrac{2}{3} < x < 1, \quad 1 < x < \tfrac{3}{2}, \quad x > 3.$$

## 验算

- 试代一个值：$\mathrm{f}(-\frac{1}{2}) = \dfrac{(7/4)^3}{(3/4)^2} = \dfrac{343/64}{9/16} = \dfrac{343}{36}$，与所述相符。
- 取所述区间内的一点：$\mathrm{f}(4) = \frac{13^3}{12^2} \approx 15.3 > \frac{343}{36} \approx 9.53$，而 $\mathrm{f}(2) = 6.75$ 位于下方，与 $(\frac{3}{2},3)$ 被排除一致。
- 六个根既可配成和为 1 的数对，又可配成互为倒数的数对，正如两个对称性所要求；以电脑在 $10^5$ 个随机点上检验，确认了这些不等式。

## 常见错误

- 对 $\mathrm{f}(1/x)$ 求导时，忘记链式法则带来的因子 $-1/x^2$。这在 $x = -1$ 处无关重要，但要明白为何 2 处的斜率为零，这个因子不可或缺。
- 把范围写成 $x < -2$ 或 $x > 3$ 之类的单一区间，忘记了 $x = 0$ 和 $x = 1$ 处的渐近线把解集分割开来；必须排除 $x = 0$ 和 $x = 1$。
- 把驻点当作极大值点。每一支在两端都趋向 $+\infty$，所以它们是极小值点。

## STEP 之外

由 $x \mapsto 1 - x$ 和 $x \mapsto 1/x$ 生成的六个映射构成一个群（group）（同构（isomorphic）于三角形的对称群），而 f 在相差一个常数因子的意义下，就是交比（cross-ratio）的经典 $j$ 不变量（j-invariant）：即在全部六个代换下都保持不变的函数。在一组代换下不变的函数，是伽罗瓦理论（Galois theory）和椭圆曲线（elliptic curve）理论的核心主题。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是认识这类变换群的好起点。

详细解答

# STEP II 2013，第 6 题

纯数学 · 数列的收敛

**STEP II 2013，第 6 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q6

本题在 *Beyond STEP* 第 2.2 节作为极限的黄金比例中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/golden-limit.html

## 关键思路

各项 $1, 2, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \ldots$ 交替地跳到其最终极限（limit）的上方和下方，所以整个数列（sequence）并非单调（monotonic），不能直接套用该定理。然而，每隔一项取一项，便得到两个单调数列。第 (i) 部分正是显示这一点的工具：它说明由于分母为正，步长 $u_{n+2} - u_n$ 与上一步 $u_n - u_{n-2}$ 同号。

## 解答

首先注意每个 $u_n$ 都是正数：$u_1 = 1 > 0$，而若 $u_n > 0$，则 $u_{n+1} = 1 + 1/u_n > 1 > 0$。因此我们从不会除以零，而且像 $1 + u_n$ 这样的因子都是正数。

### 第 (i) 部分

两次使用 $(*)$，

$$u_{n+2} - u_n = \frac{1}{u_{n+1}} - \frac{1}{u_{n-1}} = \frac{u_{n-1} - u_{n+1}}{u_{n+1}u_{n-1}}, \qquad u_{n+1} - u_{n-1} = \frac{1}{u_n} - \frac{1}{u_{n-2}} = \frac{u_{n-2} - u_n}{u_n u_{n-2}}.$$

（第二式需要 $u_{n-2}$ 存在，这就是 $n \geq 3$ 的原因。）把第二式代入第一式，两个负号互相抵消：

$$u_{n+2} - u_n = \frac{u_n - u_{n-2}}{(u_n u_{n+1})(u_{n-2}u_{n-1})}.$$

最后，把 $(*)$ 乘以 $u_n$ 得 $u_nu_{n+1} = u_n + 1$，同理 $u_{n-2}u_{n-1} = u_{n-2} + 1$。因此

$$u_{n+2} - u_n = \frac{u_n - u_{n-2}}{(1+u_n)(1+u_{n-2})},$$

得证。

**第 (ii) 部分**

我们使用归纳法（induction）。由于 $u_1 = 1$，命题 $1 \le u_1 \le 2$ 成立。假设 $1 \le u_n \le 2$。对正数取倒数（reciprocal）会令不等式反向，所以 $\frac{1}{2} \le 1/u_n \le 1$，因此

$$\tfrac{3}{2} \le u_{n+1} = 1 + \frac{1}{u_n} \le 2,$$

这当然给出 $1 \le u_{n+1} \le 2$。由归纳法，对所有 $n$ 均有 $1 \le u_n \le 2$。

**第 (iii) 部分**

奇数项。我们有 $u_1 = 1$ 及 $u_3 = \frac{3}{2}$，所以 $u_3 - u_1 > 0$。对于奇数 $n \ge 3$，第 (i) 部分说明 $u_{n+2} - u_n$ 等于 $u_n - u_{n-2}$ 乘以正数 $1/\big((1+u_n)(1+u_{n-2})\big)$，所以两者同号。由归纳法，每一步 $u_{n+2} - u_n$（$n$ 为奇数）都是正数：数列 $u_1, u_3, u_5, \ldots$ 是递增（increasing）的。由 (ii)，它有上界（bounded above）2，所以由该定理，它趋向某个极限 $L$；由于其所有项都在 $[1,2]$ 内，故 $1 \le L \le 2$。

偶数项。这里 $u_2 = 2$ 及 $u_4 = \frac{5}{3}$，所以 $u_4 - u_2 < 0$，同样的论证显示 $u_2, u_4, u_6, \ldots$ 是递减（decreasing）的。它有下界 1。

> ? **想一想** 该定理是关于有上界的递增数列的。我们怎样把它用于递减数列?

把它应用于各项的相反数：$-u_2, -u_4, -u_6, \ldots$ 是递增的，且有上界 $-1$，所以它趋向某个极限，因此 $u_2, u_4, \ldots$ 亦然；记其极限为 $M$。同样有 $1 \le M \le 2$。

求极限。使用 $(*)$ 两步，便得到一条联系同奇偶性（parity）各项的递推关系（recurrence）：

$$u_{n+2} = 1 + \frac{1}{1 + \dfrac{1}{u_n}} = 1 + \frac{u_n}{u_n + 1} = \frac{2u_n + 1}{u_n + 1}.$$

> ? **想一想** 为甚么我们可以直接把 $u_{n+2}$ 和 $u_n$ 换成 $L$?

沿着奇数 $n$，$u_n$ 和 $u_{n+2}$ 都趋向 $L$。右方是 $u_n$ 的连续函数（continuous function）（其分母至少为 2），所以它趋向 $(2L+1)/(L+1)$。极限是唯一的，所以

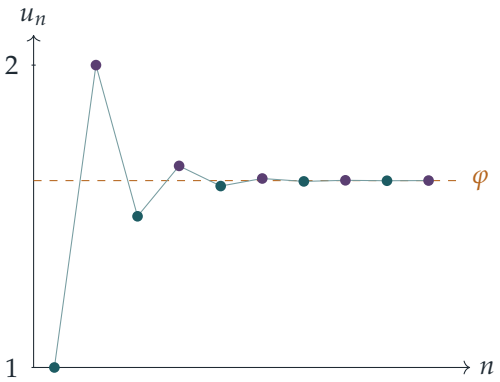
$$L = \frac{2L+1}{L+1} \quad \Longrightarrow \quad L^2 + L = 2L + 1 \quad \Longrightarrow \quad L^2 - L - 1 = 0,$$

于是 $L = \frac{1}{2}(1 \pm \sqrt{5})$。由于 $L \geq 1$，必须有 $L = \frac{1}{2}(1 + \sqrt{5})$。$M$ 满足完全相同的方程，且 $M \geq 1$，所以

$$\lim_{k\to\infty} u_{2k-1} = \lim_{k\to\infty} u_{2k} = \tfrac{1}{2}(1 + \sqrt{5}).$$

这两个极限本来有可能不同（那样整个数列便会振荡），但该方程只有一个至少为 1 的根。

整个数列。记 $\varphi = \frac{1}{2}(1 + \sqrt{5})$，并设 $\varepsilon > 0$。存在 $K_1$ 使得对所有奇数 $n > K_1$ 均有 $|u_n - \varphi| < \varepsilon$，亦存在 $K_2$ 使得对所有偶数 $n > K_2$ 同样成立。那么对每个 $n > \max(K_1, K_2)$ 均有 $|u_n - \varphi| < \varepsilon$，所以 $u_n \to \varphi$。




奇数项（青绿色）向 $\varphi$ 攀升，偶数项（梅紫色）则向它下降。

**由 $u_1 = 3$ 开始**

结论不变。现在 $u_1 = 3$，$u_2 = \frac{4}{3}$，$u_3 = \frac{7}{4}$，$u_4 = \frac{11}{7}$。论证只需稍作调整。

- 第 (i) 部分只用到 $(*)$ 及各项为正，这仍然成立，所以该恒等式（identity）依然成立。
- 第 (ii) 部分在 $n = 1$ 时不成立，但 $1 \leq u_2 = \frac{4}{3} \leq 2$，而归纳步骤不变，所以对所有 $n \geq 2$ 均有 $1 \leq u_n \leq 2$。另外 $u_1 = 3 \geq 1$，所以每一项都至少为 1。
- 这次 $u_3 - u_1 = -\frac{5}{4} < 0$ 而 $u_4 - u_2 = \frac{5}{21} > 0$。所以奇数项递减且有下界 1，偶数项递增且有上界 2：角色互换了，但两者仍然收敛（converge）。
- 递推关系 $u_{n+2} = (2u_n + 1)/(u_n + 1)$ 仍然成立，而两个极限都至少为 1，所以两者同样都是 $\varphi$。

所以该数列仍然趋向 $\frac{1}{2}(1 + \sqrt{5})$。

## 验算

当 $u_1 = 1$ 时，各项为 $1, 2, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \frac{13}{8}, \ldots$，即相邻斐波那契数（Fibonacci numbers）之比，而 $u_{10} = \frac{89}{55} \approx 1.61818$，已经非常接近 $\varphi \approx 1.61803$。当 $u_1 = 3$ 时，各项为 $3, \frac{4}{3}, \frac{7}{4}, \frac{11}{7}, \frac{18}{11}, \ldots$（卢卡斯数（Lucas numbers）之比），它们同样趋近 1.61803。你亦可以用 $n = 3$ 验算 (i) 中的恒等式：$u_5 - u_3 = \frac{1}{10}$，而 $\dfrac{u_3 - u_1}{(1+u_3)(1+u_1)} = \dfrac{1/2}{5} = \frac{1}{10}$。

## 常见错误

- 假设整个数列收敛，然后直接解 $L = 1 + 1/L$。这样得出的数值是对的，但只有在证明了收敛之后才成立，而这正是本题的重点。
- 忘记了根 $\frac{1}{2}(1 - \sqrt{5})$：必须把它排除，这里是利用 $L \geq 1$。
- 把该定理应用于递减的偶数项子数列（subsequence），却没有说明怎样做（例如通过 $-u_{2k}$）。

## STEP 之外

这是大学一年级分析学（analysis）中一个标准论证的范本：单调收敛定理（monotone convergence theorem）、子数列，以及“若奇数项子数列与偶数项子数列收敛到同一极限，则整个数列亦收敛到该极限”这一事实。事实上，$u_n$ 是连分数（continued fraction）$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}$ 的第 $n$ 个渐近分数（convergent），而映射 $x \mapsto 1 + 1/x$ 在 $\varphi$ 附近是一个压缩映射（contraction），这就是任何正数起点都会趋向 $\varphi$ 的原因。Abbott 的 *Understanding Analysis*（《理解分析》）仔细讨论了单调数列和子数列，而 Davenport 的 *The Higher Arithmetic*（《高等算术》）则涵盖了连分数。

详细解答

# STEP II 2013，第 7 题

纯数学 · 佩尔方程的解

**STEP II 2013，第 7 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q7

## 关键思路

形如 $x^2 - 2y^2 = 1$ 的方程称为佩尔方程（Pell equation）：只要找到一组解，一个固定的线性映射（linear map）便能产生无限多组解。第 (i) 和 (ii) 部分把这些解与同时是平方数的三角形数（triangular number）联系起来。第 (iii) 部分的技巧是把 $a^2$ 移到另一边，再把 $c^4 - a^2$ 按平方差（difference of two squares）因式分解。由于 $b$ 是质数（prime），把 $b^3$ 拆成两个因数的方法寥寥可数。最后，第 (ii) 部分为最后一部分提供一个非质数的 $b$。

## 解答

### 第 (i) 部分

显然的解是 $x = 1$、$y = 0$，因为 $1^2 - 2 \cdot 0^2 = 1$。

现设 $p^2 - 2q^2 = 1$。那么

$$\begin{aligned}(3p + 4q)^2 - 2(2p + 3q)^2 &= (9p^2 + 24pq + 16q^2) - 2(4p^2 + 12pq + 9q^2)\\ &= p^2 - 2q^2 = 1,\end{aligned}$$

因为含 $pq$ 的项互相抵消。所以 $x = 3p + 4q$、$y = 2p + 3q$ 也是一组解，正如所求。

把这个映射作用于 $(p, q) = (1, 0)$ 得 $(3, 2)$，再作用于 $(3, 2)$ 得 $(3 \cdot 3 + 4 \cdot 2,\ 2 \cdot 3 + 3 \cdot 2) = (17, 12)$。验证：$9 - 8 = 1$ 及 $289 - 288 = 1$。因此，两组 $x$ 为正奇数、$y$ 为正偶数的解是

$$(x, y) = (3, 2) \quad \text{及} \quad (x, y) = (17, 12).$$

（事实上，这个映射总会保持 $x$ 为奇数，因为 $3p + 4q \equiv p \pmod 2$；而一旦 $y$ 是偶数，它亦会保持 $y$ 为偶数，因为 $2p + 3q \equiv q \pmod 2$。所以这条链中往后的每一组解都具有正确的奇偶性（parity）：$(99, 70)$、$(577, 408)$，如此类推。）

### 第 (ii) 部分

设 $x$ 为奇数、$y$ 为偶数，并写 $x=2m+1$ 及 $y=2n$，其中 $m$、$n$ 为整数。那么 $(*)$ 变成

$$(2m+1)^2-2(2n)^2=1 \iff 4m^2+4m=8n^2 \iff n^2=\tfrac{1}{2}m(m+1),$$

最后一步是除以 8。这正是所求的形式。（每一步都是可逆的，所以对这样的 $x$ 和 $y$，两条方程是等价的。）

所以 $n^2$ 是三角形数 $1+2+\cdots+m$。由第 (i) 部分，$(3,2)$ 给出 $m=1$、$n=1$（即 $1=1$），而 $(17,12)$ 给出 $m=8$、$n=6$（即 $36=\frac{1}{2}\cdot 8\cdot 9$）。我们会在第 (iii) 部分用到后者。

### 第 (iii) 部分

移项并按平方差因式分解，

$$b^3=c^4-a^2=(c^2-a)(c^2+a).$$

由于 $b^3>0$ 及 $c^2+a>0$，因数 $c^2-a$ 也是正整数。此外 $a>0$，所以 $c^2-a<c^2+a$。

> ? **想一想** 把 $b^3$ 写成两个正整数之积，共有多少种方法?

因为 $b$ 是质数，$b^3$ 的正因数只有 1、$b$、$b^2$ 和 $b^3$。所以把 $b^3$ 分解成较小因数乘以较大因数的方式只有 $1\times b^3$ 和 $b\times b^2$。此外别无其他，而且两个因数永远不可能相等，因为 $b^3$ 不是完全平方数（perfect square）。

情况 1：$c^2-a=1$ 及 $c^2+a=b^3$。两式相加及相减，得

$$c^2=\tfrac{1}{2}(b^3+1), \qquad a=\tfrac{1}{2}(b^3-1).$$

情况 2：$c^2-a=b$ 及 $c^2+a=b^2$。两式相加及相减，得

$$c^2=\tfrac{1}{2}b(b+1), \qquad a=\tfrac{1}{2}b(b-1).$$

（题目只要求这些公式，但值得知道哪些情况真的会出现。情况 2 永远不会出现。当 $b=2$ 时，它需要 $c^2=3$。若 $b$ 是奇质数，则 $b$ 与 $\frac{1}{2}(b+1)$ 互质（coprime），因为 $1\le\frac{1}{2}(b+1)<b$；而它们的积是平方数 $c^2$，所以两者都必须是平方数。但质数永远不是平方数。情况 1 则确实会出现：$b=23$ 给出 $\frac{1}{2}(23^3+1)=6084=78^2$，所以 $a=6083$、$b=23$、$c=78$ 是一组解。）

### 最后部分

情况 2 对质数不成立，但若我们放弃 $b$ 必须是质数的要求，它的公式仍然能给出解。每当 $\frac{1}{2}b(b+1)$ 是完全平方数 $c^2$ 时，令 $a=\frac{1}{2}b(b-1)$。那么

$$a^2+b^3=(c^2-b)^2+b^3=c^4-2bc^2+b^2+b^3=c^4-b^2(b+1)+b^2+b^3=c^4,$$

这里用了 $2c^2 = b(b+1)$。第 (ii) 部分正好告诉我们 $\frac{1}{2}b(b+1)$ 何时是平方数：$b = m$，其中 $n^2 = \frac{1}{2}m(m+1)$。解 $m = 1$ 给出 $a = 0$，这是不容许的，但 $m = 8$、$n = 6$ 可行。取 $b = 8$、$c = 6$、$a = \frac{1}{2}\cdot 8\cdot 7 = 28$：

$$28^2 + 8^3 = 784 + 512 = 1296 = 6^4.$$

所以 $a = 28$、$b = 8$、$c = 6$ 是一组解，其中 $b = 2^3$ 不是质数。

## 验算

- 链中的下一组解 $(99, 70)$ 满足 $9801 - 2\cdot 4900 = 1$，它给出 $m = 49$、$n = 35$，且 $\frac{1}{2}\cdot 49\cdot 50 = 1225 = 35^2$。
- $6^4 = 1296$ 及 $28^2 + 8^3 = 784 + 512 = 1296$。
- 以电脑搜索所有 $c < 400$，发现 $a = 6083$、$b = 23$、$c = 78$ 是 $b$ 为质数时的唯一解。它属于情况 1 的解：$78^4 - 6083^2 = (78^2 - 6083)(78^2 + 6083) = 1 \times 12167 = 23^3$。

## 常见错误

- 忘记 $c^2 - a$ 必须是较小的因数，因而多出一些 $a$ 为负数的“情况”。
- 漏掉分解 $1 \times b^3$，只找到 $b \times b^2$ 的情况，或者反过来。
- 在最后部分使用由 $(3, 2)$ 得出的 $m = 1$，这会给出 $a = 0$；但题目要求 $a$ 为正数。

## STEP 之外

方程 $x^2 - 2y^2 = 1$ 就是佩尔方程。它的解恰好就是幂 $(3 + 2\sqrt{2})^k = x + y\sqrt{2}$，而第 (i) 部分的映射只不过是乘以 $3 + 2\sqrt{2}$。这是环（ring）如 $\mathbb{Z}[\sqrt{2}]$ 中的单位（unit）理论以及连分数（continued fraction）理论的起点。像 $a^2 + b^3 = c^4$ 这样的方程属于广义费马方程（generalised Fermat equation）的研究范畴。若想读一本精彩的入门书，涵盖以上种种，可阅读 Davenport 的 *The Higher Arithmetic*（《高等算术》）。

详细解答

# STEP II 2013，第 8 题

纯数学 · 微积分与不等式

**STEP II 2013，第 8 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q8

## 关键思路

由于 f 是递减（decreasing）函数，这个区域是一道越往右越低的“楼梯”，所以最佳的矩形会位于左下角：底边在直线 $y = \mathrm{f}(t)$ 上，左边贴着 $y$ 轴，而右上角落在曲线上。这样第 (i) 部分就化为求单变量函数的最大值（maximum）。第 (ii) 部分比较两个面积，即整个区域与其内的矩形；而第 (iii) 部分只是把 $\mathrm{f}(x) = 1/(1+x)$ 代入，把一切具体算出来。

## 解答

### 第 (i) 部分

该区域是 $\{(x,y) : 0 \le x \le t,\ \mathrm{f}(t) \le y \le \mathrm{f}(x)\}$：由于 f 递减，当 $x < t$ 时曲线位于直线 $y = \mathrm{f}(t)$ 之上。

取区域内任意一个矩形 $[a,b] \times [c,d]$。它的底边必须满足 $c \ge \mathrm{f}(t)$，左边满足 $a \ge 0$，而右上角 $(b,d)$ 必须在区域内，所以 $b \le t$ 及 $d \le \mathrm{f}(b)$。因此其面积至多为

$$b\big(\mathrm{f}(b) - \mathrm{f}(t)\big),$$

而矩形 $[0,b] \times [\mathrm{f}(t), \mathrm{f}(b)]$ 正好达到这个面积，并且确实放得进区域：对于 $0 \le x \le b$，有 $\mathrm{f}(x) \ge \mathrm{f}(b)$，所以该矩形位于曲线之下。于是

$$A_0(t) = \max_{0 \le x \le t} R(x), \qquad R(x) = x\big(\mathrm{f}(x) - \mathrm{f}(t)\big).$$

由于 $R(0) = R(t) = 0$，而对于 $0 < x < t$ 有 $R(x) > 0$，所以当 $t > 0$ 时，最大值在某个内点 $x_0 \in (0,t)$ 取得，而该点满足 $R'(x_0) = 0$：

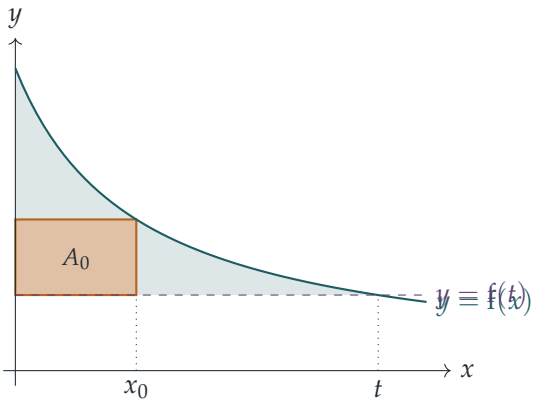
$$R'(x) = \mathrm{f}(x) + x\mathrm{f}'(x) - \mathrm{f}(t) = 0 \quad \Longrightarrow \quad x_0\mathrm{f}'(x_0) + \mathrm{f}(x_0) = \mathrm{f}(t).$$

因此 $A_0(t) = x_0\big(\mathrm{f}(x_0) - \mathrm{f}(t)\big)$，其中 $x_0$ 满足上述条件，得证。（当 $t = 0$ 时，区域没有宽度，$A_0(0) = 0$，而 $x_0 = 0$ 显然满足该条件。）

**第 (ii) 部分**

记 $t\,\mathrm{g}(t) = \displaystyle\int_0^t \mathrm{f}(x)\,\mathrm{d}x$。两边求导，左边用乘积法则（product rule），右边用微积分基本定理（fundamental theorem of calculus），得

$$\mathrm{g}(t) + t\,\mathrm{g}'(t) = \mathrm{f}(t) \quad\Longrightarrow\quad t\,\mathrm{g}'(t) = \mathrm{f}(t) - \mathrm{g}(t).$$




积分 $\int_0^t \big(\mathrm{f}(x) - \mathrm{f}(t)\big)\,\mathrm{d}x$ 正是阴影区域的面积，即 $0 \le x \le t$ 范围内曲线与直线 $y = \mathrm{f}(t)$ 之间的面积。第 (i) 部分的最佳矩形位于这个区域之内，但并未填满它：当 $x_0 < x < t$ 时，矩形没有任何贡献，但由于 f 严格递减（strictly decreasing），区域在该处的高度为 $\mathrm{f}(x) - \mathrm{f}(t) > 0$。所以那一窄条的面积是正数，因此

$$\int_0^t \big(\mathrm{f}(x) - \mathrm{f}(t)\big)\,\mathrm{d}x > A_0(t).$$

**想一想** 怎样把左边化成与 $\mathrm{g}'$ 有关的式子？

把积分拆开，利用 g 的定义，再用刚证明的结果：

$$\begin{aligned}\int_0^t \big(\mathrm{f}(x) - \mathrm{f}(t)\big)\,\mathrm{d}x &= t\,\mathrm{g}(t) - t\,\mathrm{f}(t)\\ &= -t\big(\mathrm{f}(t) - \mathrm{g}(t)\big)\\ &= -t\cdot t\,\mathrm{g}'(t) = -t^2\mathrm{g}'(t).\end{aligned}$$

因此 $-t^2\mathrm{g}'(t) > A_0(t)$。

### 第 (iii) 部分

这里 $\mathrm{f}(x)=\dfrac{1}{1+x}$ 在 $x\geq 0$ 时为正且严格递减，所以以上所有结果都适用。

*左边*。由第 (ii) 部分最后的计算，

$$-t^2\mathrm{g}'(t)=\int_0^t\frac{\mathrm{d}x}{1+x}-\frac{t}{1+t}=\ln(1+t)-\frac{t}{1+t}.$$

*矩形*。由 $\mathrm{f}'(x)=-\dfrac{1}{(1+x)^2}$，

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

所以 $x_0$ 的条件是 $(1+x_0)^2=1+t$。记 $s=\sqrt{1+t}$，则 $s>1$ 且 $x_0=s-1$（取正根 (root)；留意 $0<x_0<t$）。于是 $1+t=s^2$，而

$$A_0(t)=(s-1)\left(\frac{1}{s}-\frac{1}{s^2}\right)=\frac{(s-1)^2}{s^2}=\left(1-\frac{1}{s}\right)^2.$$

> ? **想一想** 两边看起来都很繁复：自然的变量是甚么？

以 $s$ 表示，一切都简洁得多。我们有 $\ln(1+t)=2\ln s$ 及 $\dfrac{t}{1+t}=1-\dfrac{1}{s^2}$，所以 $-t^2\mathrm{g}'(t)>A_0(t)$ 变成

$$\begin{aligned}2\ln s-1+\frac{1}{s^2}&>1-\frac{2}{s}+\frac{1}{s^2}\\ \iff\quad 2\ln s&>2-\frac{2}{s}\\ \iff\quad \ln s&>1-\frac{1}{s}.\end{aligned}$$

由于 $s=\sqrt{1+t}$ 及 $\ln s=\ln\sqrt{1+t}$，上式即

$$\ln\sqrt{1+t}>1-\frac{1}{\sqrt{1+t}}\qquad(t>0),$$

得证。

## 验算

- 取 $t=3$：$s=2$，$x_0=1$，$A_0=\frac{1}{4}$，而区域面积为 $\ln 4-\frac{3}{4}\approx 0.636>0.25$。最终的不等式为 $\ln 2\approx 0.693>0.5$。
- 以电脑对多个 $t$ 值直接搜索矩形面积 $x\big(\mathrm{f}(x)-\mathrm{f}(t)\big)$，确认最大值为 $(1-1/\sqrt{1+t})^2$，在 $x=\sqrt{1+t}-1$ 取得。

- 当 $t \to 0^+$ 时，最终不等式两边都趋向 0；展开可得 $\ln s \approx (s-1) - \frac{1}{2}(s-1)^2$ 及 $1 - \frac{1}{s} \approx (s-1) - (s-1)^2$，所以左边大约多出 $\frac{1}{2}(s-1)^2$，正合预期。

## 常见错误

- 不加说明便假设最佳矩形的角落在曲线上：这需要用到递减的性质，如第 (i) 部分所示。
- 忘记解释为何面积不等式是严格的：矩形漏掉了区域中 $x_0 < x < t$ 的部分。
- 取了负根 $1 + x_0 = -\sqrt{1+t}$，这会得出 $x_0 < 0$。

## STEP 之外

最终的不等式，即对 $s > 1$ 有 $\ln s > 1 - 1/s$，是经典界限 $1 - \frac{1}{s} < \ln s < s - 1$ 的其中一半；这组界限来自把 $\int_1^s \mathrm{d}u/u$ 与矩形作比较，而本题则以一条更巧妙的几何途径得到它。把积分与矩形面积作比较，正是黎曼（Riemann）积分的核心；大学一年级的分析课程会由上和与下和（upper and lower sums）出发，仔细建立这套理论。Abbott 的 *Understanding Analysis*（《理解分析》）对此有非常易读的阐述。

详细解答

# STEP II 2013，第 9 题

力学 • 平衡中的摩擦力

**STEP II 2013，第 9 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q9

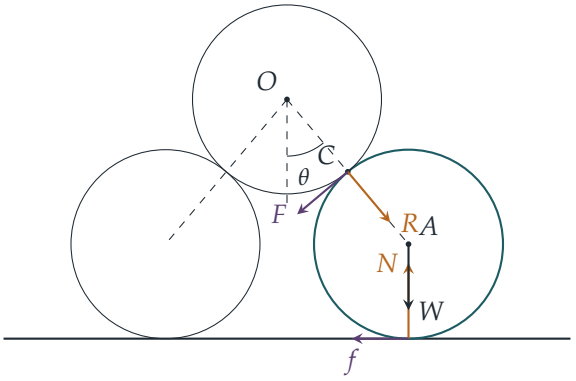
## 关键思路

大部分工作都由对称性（symmetry）完成。把三个圆盘当作一个整体系统，立即就能得出地面的反作用力。然后对其中一个底部圆盘的圆心取力矩（moments）：地面的摩擦力（friction）和顶部圆盘施加的摩擦力都沿切线（tangent）方向作用，因此两者的力臂（lever arm）相同（都是半径），大小必然相等。知道这一点后，分别对底部圆盘和顶部圆盘分解力（resolving forces），便得到两条方程，可解出两个未知数。

## 解答

我们把各圆盘视为全同的均匀刚体（rigid body）圆盘，半径为 $r$，重量各为 $W$，全部位于同一铅直平面内。设 $O$ 为顶部圆盘的圆心，$A$ 为右边底部圆盘的圆心，则 $OA = 2r$，且 $OA$ 与向下的铅直线成角 $\theta$。这两个圆盘的接触点 $C$ 是 $OA$ 的中点。

由左右对称可知，左边底部圆盘所受的力与右边圆盘所受的力互为镜像，所以我们只需处理右边的圆盘。在 $C$ 点，右边圆盘受到沿 $OA$ 方向的法向反作用力（normal reaction）$R$（把它推离 $O$），以及沿公切线方向的摩擦力 $F$。在地面上，它受到法向反作用力 $N$ 和摩擦力 $f$。




图中显示右边底部圆盘所受的力。根据牛顿（Newton）第三定律（Newton's third law），顶部圆盘在 $C$ 点受到方向相反的 $R$ 和 $F$，因此摩擦力沿切线把顶部圆盘向上及向外推。

### 第 (i) 部分

把三个圆盘合起来视为一个系统。圆盘之间的力是内力，会互相抵消，所以仅有的外力是三个重量和地面施加的力。由对称性，两个地面反作用力相等，而地面的摩擦力是水平的，因此铅直方向分解力得

$$2N = 3W \quad \Longrightarrow \quad N = \tfrac{3}{2}W,$$

得证。

### 第 (ii) 部分

**想一想** 底部圆盘上似乎有三个未知数 $R$、$F$ 和 $f$。第三条方程从何而来?

来自力矩。对右边底部圆盘关于 $A$ 取力矩。重量 $W$、法向反作用力 $N$ 和法向反作用力 $R$ 都通过 $A$，所以没有力矩。摩擦力 $f$ 沿最低点的切线作用，$F$ 沿 $C$ 点的切线作用，所以两者的力臂都是 $r$。它们必须使圆盘向相反方向转动，因此

$$fr = Fr \quad \Longrightarrow \quad f = F.$$

（这也确定了图中各力的方向。圆盘被向外推，所以 $f$ 指向内，于是底部圆盘上的 $F$ 必须沿切线指向下方及内侧，才能平衡 $f$ 的力矩。）

由 $O$ 到 $A$ 的单位向量是 $(\sin\theta, -\cos\theta)$，而底部圆盘上 $F$ 的切线方向是 $(-\cos\theta, -\sin\theta)$。对底部圆盘水平方向分解力：

$$R\sin\theta - F\cos\theta - f = 0 \quad \Longrightarrow \quad R\sin\theta = F(1+\cos\theta),$$

其中用了 $f = F$。

对顶部圆盘而言，两个接触点提供相等的向上分量，各为 $R\cos\theta + F\sin\theta$。铅直方向分解力：

$$2(R\cos\theta + F\sin\theta) = W.$$

代入 $F = \dfrac{R\sin\theta}{1+\cos\theta}$：

$$W = 2R\left(\cos\theta + \frac{\sin^2\theta}{1+\cos\theta}\right) = 2R\cdot\frac{\cos\theta+\cos^2\theta+\sin^2\theta}{1+\cos\theta} = 2R\cdot\frac{1+\cos\theta}{1+\cos\theta} = 2R.$$

所以两个互相接触的圆盘之间的法向反作用力是

$$R = \tfrac{1}{2}W,$$

且与 $\theta$ 无关。于是

$$F = \frac{R\sin\theta}{1+\cos\theta} = \frac{W\sin\theta}{2(1+\cos\theta)},$$

这就是所求的摩擦力。

作为一致性检查，对底部圆盘铅直方向分解力得 $N = W + R\cos\theta + F\sin\theta = W + \frac{1}{2}W = \frac{3}{2}W$，与第 (i) 部分吻合。

**第 (iii) 部分**

利用半角公式（half-angle formulae）$\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$ 及 $1+\cos\theta = 2\cos^2\frac{\theta}{2}$，

$$\frac{F}{R} = \frac{\sin\theta}{1+\cos\theta} = \tan\tfrac{\theta}{2}.$$

要达到平衡（equilibrium），每个接触点的摩擦力都不能超过其极限值（limiting value）。在圆盘之间，这表示 $F \le \mu R$，即

$$\tan\tfrac{\theta}{2} \le \mu.$$

（在地面上，我们需要 $f \le \mu N$，即 $\frac{1}{2}W\tan\frac{\theta}{2} \le \frac{3}{2}\mu W$。这个条件较弱，所以起决定作用的是圆盘之间的接触。）

> ? **想一想** 是甚么阻止 $\theta$ 任意地小，从而令 $\tan\frac{\theta}{2}$ 变得极小?

两个底部圆盘并不接触。它们的圆心相距 $2\cdot 2r\sin\theta = 4r\sin\theta$，而这必须大于 $2r$，所以 $\sin\theta > \frac{1}{2}$。由于 $0 < \theta < \frac{\pi}{2}$，这表示 $\theta > \frac{\pi}{6}$，所以 $\frac{\theta}{2} > \frac{\pi}{12}$。由于 tan 在 $(0, \frac{\pi}{2})$ 上递增，

$$\tan\tfrac{\theta}{2} > \tan\tfrac{\pi}{12} = \tan(45^\circ - 30^\circ) = \frac{1-\frac{1}{\sqrt{3}}}{1+\frac{1}{\sqrt{3}}} = \frac{\sqrt{3}-1}{\sqrt{3}+1} = \frac{(\sqrt{3}-1)^2}{2} = 2-\sqrt{3}.$$

所以平衡要求 $\mu \ge \tan\frac{\theta}{2} > 2-\sqrt{3}$。若 $\mu < 2-\sqrt{3}$，则没有任何容许的 $\theta$ 值能满足此条件，平衡不可能出现。

## 验算

- 当 $\theta \to 0$ 时，顶部圆盘正正坐在其他圆盘之上，而 $F \to 0$：不需要摩擦力，与预期一致。
- 取任何样本角度，例如 $\theta = 1$ 弧度，代入 $R = \frac{1}{2}W$、$F = f = \frac{1}{2}W\tan\frac{1}{2}$ 及 $N = \frac{3}{2}W$，每个圆盘的两个力分量和力矩都平衡。
- $2 - \sqrt{3} \approx 0.2679$，与 $\tan 15^\circ \approx 0.2679$ 吻合。对略大于 $30^\circ$ 的 $\theta$ 作数值扫描，证实 $\mu = 0.2679$ 时平衡不可能，而 $\mu = 0.268$ 时平衡可能。

## 常见错误

- 假设圆盘之间的摩擦力为零，或完全忘记它。没有它的话，底部圆盘的力矩方程会迫使 $f = 0$，那就没有东西能阻止底部圆盘向两边滑开。
- 遗漏了底部圆盘互不接触这个条件。它是唯一给 $\theta$ 设下限的条件，因此 $2 - \sqrt{3}$ 正是由此而来。
- 弄错 $F$ 的方向。应利用关于 $A$ 的力矩方程来确定方向，而不是靠猜。

## STEP 之外

这个问题之所以是静定（statically determinate）的，全靠对称性和力矩方程；在许多由粗糙物体堆叠而成的情况中，单凭平衡条件无法求出摩擦力，工程师需要为物体的形变建立模型。简洁的结果 $F/R = \tan\frac{\theta}{2}$ 是摩擦角（friction angle）概念的一个例子：当且仅当总接触力与法线所成的角不超过 $\arctan\mu$ 时，平衡才可能出现。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）在大学一年级程度上仔细讨论了摩擦力和刚体静力学。

详细解答

# STEP II 2013，第 10 题

力学 • 抛体运动

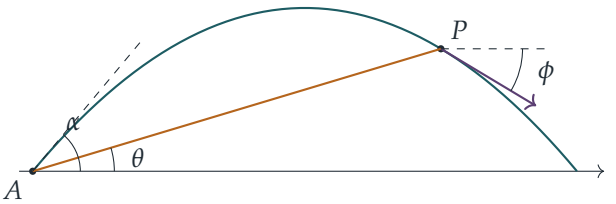
**STEP II 2013，第 10 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q10

## 关键思路

写出时间 $t$ 时的位置和速度，便会发现 $\tan\theta$ 和 $\tan\phi$ 都是 $t$ 的线性（linear）函数。(i) 中的恒等式（identity）随即得出，而它把 $B$ 和 $C$ 都转化为关于角的正切的陈述。由于两个角都随时间稳定地减小，“哪一点先到”不过是比较两个数的大小，而 (i) 让我们只需在 $B$ 这一点上作比较。




在一般的点 $P$，$\theta$ 是从 $A$ 看 $P$ 的仰角（angle of elevation），而 $\phi$ 是速度与水平方向所成的角（图中为负值，因为质点正在下落）。

## 解答

我们把质点模型化为只受重力 $g$ 作用而自由运动，不计空气阻力（air resistance）。设质点在时间 $t = 0$ 以速率 $u$ 抛出，其中 $0 < \alpha < 90°$。以 $A$ 为原点，在时间 $t$ 时

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

当 $y = 0$ 再次成立时质点落地，此时 $t = 2u\sin\alpha/g$。

**第 (i) 部分**

当 $t > 0$ 时，从 $A$ 看质点的仰角满足

$$\tan\theta = \frac{y}{x} = \frac{ut\sin\alpha - \frac{1}{2}gt^2}{ut\cos\alpha} = \tan\alpha - \frac{gt}{2u\cos\alpha},$$

而运动方向满足

$$\tan\phi = \frac{\dot{y}}{\dot{x}} = \frac{u\sin\alpha - gt}{u\cos\alpha} = \tan\alpha - \frac{gt}{u\cos\alpha}.$$

把第一式乘以 2，再利用第二式，

$$2\tan\theta = 2\tan\alpha - \frac{gt}{u\cos\alpha} = \tan\alpha + \left(\tan\alpha - \frac{gt}{u\cos\alpha}\right) = \tan\alpha + \tan\phi,$$

得证。

值得记下稍后会用到的结果：记 $k = \dfrac{g}{u\cos\alpha} > 0$，

$$\tan\theta = \tan\alpha - \tfrac{1}{2}kt, \qquad \tan\phi = \tan\alpha - kt, \tag{$*$}$$

因此 $\tan\theta$ 和 $\tan\phi$ 都随时间严格递减。在整个飞行过程中，$\theta$ 由 $\alpha$ 减至 0，$\phi$ 由 $\alpha$ 减至 $-\alpha$，所以 $\theta = \frac{1}{2}\alpha$ 和 $\phi = -\frac{1}{2}\alpha$ 各恰好出现一次：$B$ 和 $C$ 的定义是明确的。

**第 (ii) 部分**

记 $T = \tan\frac{1}{2}\alpha$；由于 $0 < \frac{1}{2}\alpha < 45°$，我们有 $0 < T < 1$。在 $B$，$\tan\theta = T$。在 $C$，$\tan\phi = \tan(-\frac{1}{2}\alpha) = -T$。若 $B$ 和 $C$ 是同一点，则两式在该点同时成立，由 (i) 得

$$2T = \tan\alpha - T, \qquad \text{即} \qquad \tan\alpha = 3T.$$

由二倍角公式（double-angle formula）$\tan\alpha = \dfrac{2T}{1-T^2}$，故

$$\frac{2T}{1-T^2} = 3T \implies 2 = 3(1-T^2) \implies T^2 = \tfrac{1}{3},$$

其中我们除以了 $T$，这是容许的，因为 $T > 0$。取正根，$\tan\frac{1}{2}\alpha = \frac{1}{\sqrt{3}}$，所以 $\frac{1}{2}\alpha = 30°$，且

$$\alpha = 60°.$$

**第 (iii) 部分**

> ? **想一想** 不计算两个时间，怎样判断哪一点先到达?

由 $(*)$，$\phi$ 在整个飞行过程中递减。因此 $C$ 在 $B$ 之后到达，当且仅当质点位于 $B$ 的那一刻，$\phi$ 尚未减至 $-\frac{1}{2}\alpha$；即当 $\tan\phi_B > -T$ 时，其中 $\phi_B$ 是质点在 $B$ 的运动方向。

在 $B$，由第 (i) 部分得 $\tan\phi_B = 2\tan\theta - \tan\alpha = 2T - \tan\alpha$。因此

$$B \text{ 先到达} \iff 2T - \tan\alpha > -T \iff \tan\alpha < 3T.$$

再次利用 $\tan\alpha = \dfrac{2T}{1-T^2}$，并除以 $T > 0$、乘以 $1 - T^2 > 0$（这样不等式的方向不变），

$$\frac{2T}{1-T^2} < 3T \iff 2 < 3 - 3T^2 \iff T^2 < \tfrac{1}{3} \iff \tfrac{1}{2}\alpha < 30^\circ,$$

最后一步成立是因为 tan 在 $(0^\circ, 45^\circ)$ 上递增。每一步都可逆，所以

$$\alpha < 60^\circ \iff B \text{ 比 } C \text{ 先到达.}$$

所以，既然 $\alpha < 60^\circ$，质点会先到达 $B$。

若你偏好明确写出时间：由 $(*)$，$kt_B = 2(\tan\alpha - T)$ 及 $kt_C = \tan\alpha + T$，而 $t_B < t_C$ 同样等价于 $\tan\alpha < 3T$。

## 验算

- 在 $t = 0$ 时，$(*)$ 中两条公式都给出 $\tan\alpha$，理应如此：刚抛出时质点沿抛射方向的直线运动。落地时 $kt = 2\tan\alpha$，得 $\tan\theta = 0$ 及 $\tan\phi = -\tan\alpha$，这正是轨迹为人熟知的对称性。
- 当 $\alpha = 60^\circ$ 时，$\tan\alpha = \sqrt{3} = 3\tan 30^\circ$，与 (ii) 一致。
- 取 $u = 10\,\mathrm{m\,s^{-1}}$、$g = 9.8\,\mathrm{m\,s^{-2}}$ 及 $\alpha = 45^\circ$，公式给出 $t_B \approx 0.845\,\mathrm{s}$ 及 $t_C \approx 1.020\,\mathrm{s}$，所以 $B$ 先到；以电脑逐步数值模拟运动，得出相同的时间，并确认当 $\alpha = 61^\circ$、$75^\circ$ 及 $85^\circ$ 时 $C$ 先到。

## 常见错误

- 把 $\theta$（质点相对于 $A$ 的方向）当作 $\phi$（质点速度的方向）。两者只在抛出的一刻相同。
- 在 $C$ 弄错符号：该处 $\tan\phi = -\tan\frac{1}{2}\alpha$，而不是 $+\tan\frac{1}{2}\alpha$。
- 在 (iii) 中直接断言答案，却没有解释为何“$\tan\phi_B > -T$”表示 $B$ 先到。支持这一点的正是 $\phi$ 的单调递减（monotonic decrease）。

## STEP 之外

(i) 中的恒等式说明：从 $A$ 到抛物线（parabola）上任一点的弦（chord），其斜率等于两端点切线斜率的平均值，这是二次函数的特殊性质（相当于均值定理（mean value theorem），

而且中间点被准确定在正中)。在大学一年级的力学课程中, 研究斜面 (inclined plane) 上的抛射体 (projectile) 问题时会遇到同样的想法, 那里从抛出点出发的直线所成的角, 所扮演的角色正与这里的 $\theta$ 完全相同。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics* (《力学导论》) 是很好的参考书，可看到抛射体问题以这种严谨方式处理。

详细解答

# STEP II 2013，第 11 题

力学 · 连续碰撞

**STEP II 2013，第 11 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

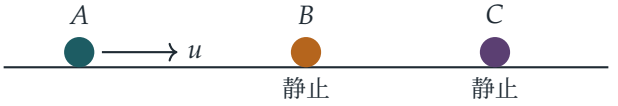
step.maths.org/questions/13-s2-q11

## 关键思路

这里每一次碰撞（collision）都发生在两个质量相等的质点之间，所以每次都是同一个小计算：动量守恒（conservation of momentum）加上牛顿恢复定律（Newton's law of restitution）。先做一次一般情况，之后重复套用即可。会否再发生碰撞，只取决于后面的质点是否比前面的质点移动得快，因此整道题归结为比较速度。

## 解答

把这些质点视为在光滑（smooth）平面上、质量均为 $m$ 的相同质点，因此在两次碰撞之间没有水平力作用，每个质点都以等速移动。沿直线依次称它们为 $A$、$B$、$C$，其中 $A$ 是被射向 $B$ 和 $C$ 的那个质点。取 $A$ 初始运动的方向为正方向。




### 一般的碰撞

假设一个速度为 $p$ 的质点追上前方一个速度为 $q < p$ 的相同质点，碰撞后两者速度分别为 $p'$ 和 $q'$。由动量守恒（约去 $m$）及恢复定律得

$$p' + q' = p + q, \qquad q' - p' = e(p - q).$$

两式相加及相减，

$$p' = \tfrac{1}{2}\big[(p+q) - e(p-q)\big], \qquad q' = \tfrac{1}{2}\big[(p+q) + e(p-q)\big]. \tag{$*$}$$

由于 $e>0$ 且 $p>q$，我们总有 $q'>p'$：碰撞后前方的质点比后方的移动得快，所以除非有其他因素改变其中一个的速度，这两个质点不可能再相遇。

**第 (i) 部分**

**第一次碰撞**（$A$ 撞 $B$）：在 $(*)$ 中代入 $p=u$、$q=0$：

$$v_A=\tfrac{1}{2}u(1-e),\qquad v_B=\tfrac{1}{2}u(1+e).$$

此时 $v_B>v_A$，所以 $A$ 追不上 $B$；但 $B$ 正向静止的 $C$ 移动，所以第二次碰撞是 $B$ 撞 $C$。

**第二次碰撞**（$B$ 撞 $C$）：代入 $p=\tfrac{1}{2}u(1+e)$、$q=0$：

$$v_B=\tfrac{1}{2}\cdot\tfrac{1}{2}u(1+e)(1-e)=\tfrac{1}{4}u(1-e^2),\qquad v_C=\tfrac{1}{4}u(1+e)^2.$$

$A$ 没有受到影响，所以仍然是 $v_A=\tfrac{1}{2}u(1-e)$。三个速度都是正数（因为 $0<e<1$），所以速率分别为 $\tfrac{1}{2}u(1-e)$、$\tfrac{1}{4}u(1-e^2)$、$\tfrac{1}{4}u(1+e)^2$，正如所求。

**第三次碰撞**。$C$ 比 $B$ 快（由 $(*)$），所以 $B$ 和 $C$ 会分开。唯一可能的下一次碰撞是 $A$ 追上 $B$，而这恰好在 $v_A>v_B$ 时发生：

$$v_A-v_B=\tfrac{1}{4}u(1-e)\big[2-(1+e)\big]=\tfrac{1}{4}u(1-e)^2>0,$$

因为 $e\neq 1$。所以 $A$ 在 $B$ 后方且移动得比 $B$ 快，对于 $0<e<1$ 中的每一个 $e$，第三次碰撞（$A$ 撞 $B$）都会发生。

**第 (ii) 部分**

**第三次碰撞**（$A$ 撞 $B$）：此时 $p=\tfrac{1}{2}u(1-e)$、$q=\tfrac{1}{4}u(1-e^2)$，所以

$$p+q=\tfrac{1}{4}u(1-e)\big[2+(1+e)\big]=\tfrac{1}{4}u(1-e)(3+e),\qquad p-q=\tfrac{1}{4}u(1-e)^2.$$

由 $(*)$，$B$ 的新速度为

$$\begin{aligned}v_B&=\tfrac{1}{2}\Big[\tfrac{1}{4}u(1-e)(3+e)+\tfrac{1}{4}eu(1-e)^2\Big]\\&=\tfrac{1}{8}u(1-e)\big[3+e+e-e^2\big]\\&=\tfrac{1}{8}u(1-e)(3+2e-e^2)=\tfrac{1}{8}u(1-e)(1+e)(3-e),\end{aligned}$$

而（为求完整）$v_A=\tfrac{1}{8}u(1-e)\big[3+e-e(1-e)\big]=\tfrac{1}{8}u(1-e)(3+e^2)$，它小于 $v_B$。

**想一想** 第三次碰撞之后，哪些质点对有可能再次碰撞？

$A$ 和 $B$ 刚刚碰撞过，所以由 $(*)$，$B$ 现在比 $A$ 快，两者正在分开。$C$ 仍以 $\tfrac{1}{4}u(1+e)^2$ 移动，而 $C$ 前方再没有其他质点。所以第四次碰撞发生若且唯若 $B$ 追上 $C$，即若且唯若 $v_B>v_C$：

两边除以 $\frac{1}{8}u(1+e)>0$，

$$\begin{aligned}\tfrac{1}{8}u(1-e)(1+e)(3-e)>\tfrac{1}{4}u(1+e)^2 &\iff (1-e)(3-e)>2(1+e)\\ &\iff e^2-6e+1>0.\end{aligned}$$

二次式 $e^2-6e+1$ 的根为 $e=3\pm2\sqrt{2}$，它在两根以外取正值。较大的根 $3+2\sqrt{2}$ 大于 1，而 $3-2\sqrt{2}\approx0.172$ 位于 $(0,1)$ 之内。所以，对于 $0<e<1$，

$$\text{发生第四次碰撞} \iff e<3-2\sqrt{2}.$$

当 $e=3-2\sqrt{2}$ 时，$B$ 和 $C$ 以相同速度移动，永不相遇，所以不等式（inequality）是严格的。

## 验算

- 动量：第二次碰撞后总动量为 $\frac{1}{4}u\left[2(1-e)+(1-e^2)+(1+e)^2\right]=\frac{1}{4}u\cdot4=u$，与预期相符。
- 极限情况：当 $e\to1$ 时，两次碰撞后的速度趋向 0、0、$u$（即大家熟悉的“牛顿摆（Newton's cradle）”式速度交换），而 $v_A-v_B\to0$，所以第三次碰撞仅仅勉强发生；临界值 $3-2\sqrt{2}$ 远小于 1，与此一致。
- 数值上，取 $e=0.1$，第三次碰撞后的速度为 $v_B\approx0.3589u$ 和 $v_C=0.3025u$，所以 $B$ 确实追上 $C$；取 $e=0.2$，$v_B=0.336u<v_C=0.36u$，则追不上。以电脑模拟这个运动，证实了对每个 $e>0.1716$ 恰好有三次碰撞，而低于此值则至少有四次。

## 常见错误

- 忘记说明下一次碰撞的是哪一对质点。每次都要检查刚碰撞过的两个质点正在分开，而余下的一对正在接近。
- 恢复定律中的符号弄错：应为（分离速率）$=e\times$（接近速率），所以 $q'-p'=e(p-q)$，左边先写前方质点的速度。
- 保留了根 $3+2\sqrt{2}$，但它在容许范围 $0<e<1$ 之外。

## STEP 之外

规则 $(*)$ 是一个线性映射（linear map），把碰撞前的一对速度映射到碰撞后的一对速度，所以一连串碰撞就是矩阵（matrix）的乘积；问总共会发生多少次碰撞是一个经典问题，对于弹性（elastic）桌球系统，把运动“展开”成楔形区域中的一条直线，便可得到一个优

美的几何答案。若想看以大学程度的严谨处理碰撞、动量及参考系（frames of reference），可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

# STEP II 2013，第 12 题

概率与统计 · 泊松分布的矩

**STEP II 2013，第 12 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s2-q12

## 关键思路

$X$ 保留 $U$ 的奇数值，$Y$ 则保留偶数值，因此每个期望值（expectation）都是只取奇数项或只取偶数项的泊松（Poisson）和。求 $\mathrm{E}(U)$ 和 $\mathrm{E}(U(U-1))$ 的技巧依然适用：把 $k$ 或 $k(k-1)$ 与 $k!$ 约去，再提出 $\lambda$ 或 $\lambda^2$。$k$ 每减少一，奇偶便互换一次，这正是 $\alpha$ 与 $\beta$ 互换位置的原因。另外留意 $\alpha+\beta=\mathrm{e}^{\lambda}$，所以因子 $\mathrm{e}^{-\lambda}$ 就是 $1/(\alpha+\beta)$。

## 解答

全文中，对 $k=0,1,2,\ldots$，$\mathrm{P}(U=k)=\mathrm{e}^{-\lambda}\lambda^k/k!$，其中 $\lambda>0$。把两个级数（series）逐项相加，得

$$\alpha+\beta=\sum_{k\geq 0}\frac{\lambda^k}{k!}=\mathrm{e}^{\lambda},\qquad \text{所以}\qquad \mathrm{e}^{-\lambda}=\frac{1}{\alpha+\beta}.$$

（事实上 $\alpha=\cosh\lambda$，而 $\beta=\sinh\lambda$。）

### 第 (i) 部分

除非 $U$ 是奇数，否则 $X$ 为 0，所以只有奇数 $k$ 有贡献：

$$\mathrm{E}(X)=\sum_{k\text{ 为奇数}}k\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}=\mathrm{e}^{-\lambda}\lambda\sum_{k\text{ 为奇数}}\frac{\lambda^{k-1}}{(k-1)!}=\mathrm{e}^{-\lambda}\lambda\left(1+\frac{\lambda^2}{2!}+\frac{\lambda^4}{4!}+\cdots\right),$$

因为 $k-1$ 取遍 $0,2,4,\ldots$。所以

$$\mathrm{E}(X)=\frac{\lambda\alpha}{\alpha+\beta}.$$

同理，$Y$ 取的是偶数 $k \geq 2$（$k = 0$ 那一项本来就是零），此时 $k-1$ 取遍 $1, 3, 5, \ldots$：

$$\mathrm{E}(Y) = \mathrm{e}^{-\lambda}\lambda \sum_{k \text{ 为偶数}, k\geq 2} \frac{\lambda^{k-1}}{(k-1)!} = \frac{\lambda\beta}{\alpha+\beta}.$$

验证一下：$\mathrm{E}(X) + \mathrm{E}(Y) = \lambda = \mathrm{E}(U)$，理应如此，因为 $X + Y = U$。

**第 (ii) 部分**

> **想一想** 当 $k^2$ 不能与 $k!$ 干净地约去时，怎样求 $k^2\lambda^k/k!$ 的和?

把 $k^2 = k(k-1) + k$ 写出来。$k(k-1)$ 部分约去 $k!$ 的两个因子，而 $k$ 部分约去一个。对奇数 $k$，$k(k-1)$ 的和由 $k = 3$ 开始（在 $k = 1$ 时该项为零），而 $k-2$ 取遍 $1, 3, 5, \ldots$：

$$\sum_{k \text{ 为奇数}} k(k-1)\frac{\lambda^k}{k!} = \lambda^2 \sum_{k \text{ 为奇数}, k\geq 3} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2\beta, \qquad \sum_{k \text{ 为奇数}} k\frac{\lambda^k}{k!} = \lambda\alpha$$

（第二个和就是第 (i) 部分中的那个）。因此

$$\mathrm{E}(X^2) = \mathrm{e}^{-\lambda}\left(\lambda\alpha + \lambda^2\beta\right) = \frac{\lambda\alpha + \lambda^2\beta}{\alpha+\beta},$$

再减去 $\mathrm{E}(X)^2$，得

$$\mathrm{Var}(X) = \frac{\lambda\alpha + \lambda^2\beta}{\alpha+\beta} - \frac{\lambda^2\alpha^2}{(\alpha+\beta)^2},$$

正如所求。

对 $Y$ 而言，奇偶性（parity）互换。对偶数 $k$，$k(k-1)$ 的和由 $k = 2$ 开始，$k-2$ 取遍 $0, 2, 4, \ldots$，结果是 $\lambda^2\alpha$。$k$ 的和则如第 (i) 部分一样给出 $\lambda\beta$。所以

$$\mathrm{Var}(Y) = \frac{\lambda\beta + \lambda^2\alpha}{\alpha+\beta} - \frac{\lambda^2\beta^2}{(\alpha+\beta)^2}.$$

**最后一问。** 由于 $X + Y = U$，我们有 $\mathrm{Var}(X+Y) = \mathrm{Var}(U) = \lambda$。把两个方差（variance）相加，

$$\begin{aligned}\mathrm{Var}(X) + \mathrm{Var}(Y) &= \frac{\lambda(\alpha+\beta) + \lambda^2(\alpha+\beta)}{\alpha+\beta} - \frac{\lambda^2(\alpha^2+\beta^2)}{(\alpha+\beta)^2} \\ &= \lambda + \lambda^2\frac{(\alpha+\beta)^2 - \alpha^2 - \beta^2}{(\alpha+\beta)^2} = \lambda + \frac{2\lambda^2\alpha\beta}{(\alpha+\beta)^2}.\end{aligned}$$

> **想一想** 为甚么我们应该预期两边不相等?

一般而言，$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$。此处 $XY = 0$ 恒成立，因为 $X$、$Y$ 之中最多只有一个非零。所以 $\mathrm{Cov}(X, Y) = -\mathrm{E}(X)\mathrm{E}(Y) = -\lambda^2\alpha\beta/(\alpha+\beta)^2$，与上面

的计算吻合。要等式成立，便需要 $\lambda^2\alpha\beta = 0$。然而对 $\lambda > 0$，有 $\alpha \geq 1$ 及 $\beta \geq \lambda > 0$，所以 $\lambda^2\alpha\beta > 0$。因此**不存在非零的** $\lambda$ **值**使 $\mathrm{Var}(X) + \mathrm{Var}(Y) = \mathrm{Var}(X+Y)$：左边总是严格较大。事实上 $\alpha = \cosh\lambda$，$\beta = \sinh\lambda$，所以多出的部分是 $\lambda^2 \sinh 2\lambda / \mathrm{e}^{2\lambda} = \frac{1}{2}\lambda^2\left(1 - \mathrm{e}^{-4\lambda}\right)$。

## 验算

- 当 $\lambda \to 0$ 时，$U$ 几乎总是 0 或 1，所以 $Y$ 几乎总是 0。的确，$\mathrm{E}(Y) \approx \lambda^2 \to 0$，而 $\mathrm{E}(X) \approx \lambda$。
- 当 $\lambda$ 很大时，$\alpha \approx \beta \approx \frac{1}{2}\mathrm{e}^{\lambda}$，所以 $\mathrm{E}(X) \approx \mathrm{E}(Y) \approx \frac{1}{2}\lambda$：奇数值与偶数值变得同样可能出现。
- 在 $\lambda = 1$ 时，直接把泊松概率相加，得 $\mathrm{Var}(X) \approx 0.6778$ 及 $\mathrm{Var}(Y) \approx 0.8131$，所以两者之和比 $\mathrm{Var}(U) = 1$ 多出约 $0.4908 = \frac{1}{2}(1 - \mathrm{e}^{-4})$。两条公式都吻合。

## 常见错误

- 约去之后弄错奇偶：若 $k$ 是奇数，则 $k-1$ 是偶数（给出 $\alpha$），而 $k-2$ 是奇数（给出 $\beta$）。
- 忘记 $\mathrm{e}^{-\lambda} = 1/(\alpha+\beta)$，以致答案不是题目要求的形式。
- 假设方差可以直接相加。$X$ 和 $Y$ 远非独立（independent）：知道 $X \neq 0$ 便迫使 $Y = 0$。

## STEP 之外

把一个级数分拆成偶数部分和奇数部分，就是分解式 $\mathrm{e}^{\lambda} = \cosh\lambda + \sinh\lambda$。借助概率母函数（probability generating functions），这变成例行手续：$G(s)$ 的偶数部分是 $\frac{1}{2}\big(G(s) + G(-s)\big)$。最后用到的协方差（covariance）恒等式，是大学概率课程最早接触的工具之一。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）同时涵盖母函数和协方差。

详细解答

# STEP II 2013，第 13 题

概率与统计 • 掷硬币中的游程

> **STEP II 2013，第 13 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s2-q13

## 关键思路

事件 $S = n$ 完全确定了首 $n+1$ 次掷币的结果：先是 $n$ 次相同的结果，然后一次打断这段连串（run）的结果。$A = n$ 亦是如此：先是 $n$ 次交替（alternating）出现的结果，然后一次重复。因此每个概率（probability）都是两个乘积之和，各对应第一次掷币的一种可能结果。比较 $S$ 与 $A$ 便归结为将四项之差因式分解（factorisation）；若我们把 $p$ 和 $q$ 当作两个独立的字母，而不是以 $1-p$ 代替 $q$，因式分解会较容易看出来。

## 解答

全文中，各次掷币互相独立（independent），每次得正面的概率为 $p$，得反面的概率为 $q$，且 $p, q > 0$，$p \neq q$。

### 一般公式

$S = n$ 发生，当且仅当第 $1, \ldots, n$ 次掷币全部与第一次相同，而第 $n+1$ 次不同。若第一次是正面，结果便是 $\mathrm{H}^n\mathrm{T}$，概率为 $p^n q$；若第一次是反面，结果便是 $\mathrm{T}^n\mathrm{H}$，概率为 $q^n p$。这两种情况互斥（disjoint），所以

$$\mathrm{P}(S = n) = p^n q + q^n p = pq\left(p^{n-1} + q^{n-1}\right).$$

$A = n$ 发生，当且仅当第 $1, \ldots, n$ 次掷币交替出现，而第 $n+1$ 次与第 $n$ 次相同，从而结束交替。

▶ 当 $n = 2m$（偶数）：连串 $\mathrm{HTHT}\cdots\mathrm{HT}$ 有 $m$ 个正面和 $m$ 个反面，并以 T 结尾，所以

下一次必须是 T，得 $p^m q^m \cdot q$。以反面开始则得 $q^m p^m \cdot p$。因此

$$\mathrm{P}(A = 2m) = p^m q^m (p + q).$$

- 当 $n = 2m + 1$（奇数）：连串 HTH$\cdots$H 有 $m + 1$ 个正面和 $m$ 个反面，并以 H 结尾，所以下一次必须是 H，得 $p^{m+2} q^m$。以反面开始则得 $q^{m+2} p^m$。因此

$$\mathrm{P}(A = 2m + 1) = p^m q^m \left(p^2 + q^2\right).$$

**第 (i) 部分**

$A = 1$ 表示第二次掷币重复第一次的结果：HH 或 TT。所以 $\mathrm{P}(A = 1) = p^2 + q^2$（即上面 $m = 0$ 的情况）。$S = 1$ 表示第二次掷币与第一次不同：HT 或 TH，所以

$$\mathrm{P}(S = 1) = pq + qp = 2pq.$$

于是

$$\mathrm{P}(A = 1) - \mathrm{P}(S = 1) = p^2 - 2pq + q^2 = (p - q)^2 > 0,$$

由于 $p \neq q$，不等号是严格的。所以 $\mathrm{P}(S = 1) < \mathrm{P}(A = 1)$。

**第 (ii) 部分**

由公式，

$$\mathrm{P}(S = 2) = p^2 q + q^2 p = pq(p + q) = \mathrm{P}(A = 2),$$

此即所求。至于长度 3，

$$\mathrm{P}(S = 3) = p^3 q + q^3 p = pq\left(p^2 + q^2\right) = \mathrm{P}(A = 3),$$

所以这两者亦相等。

**第 (iii) 部分**

**偶数长度**。设 $n > 1$。将两个概率都除以 $pq > 0$，

$$\frac{\mathrm{P}(S = 2n)}{pq} = p^{2n-1} + q^{2n-1}, \qquad \frac{\mathrm{P}(A = 2n)}{pq} = p^{n-1} q^{n-1}(p + q) = p^n q^{n-1} + p^{n-1} q^n.$$

? **想一想** 怎样才能判断 $p^{2n-1} + q^{2n-1} - p^n q^{n-1} - p^{n-1} q^n$ 的正负号?

将每个纯幂与含同一字母最高次幂的交叉项配对：

$$\begin{aligned} p^{2n-1} + q^{2n-1} - p^n q^{n-1} - p^{n-1} q^n &= p^n\left(p^{n-1} - q^{n-1}\right) - q^n\left(p^{n-1} - q^{n-1}\right) \\ &= \left(p^n - q^n\right)\left(p^{n-1} - q^{n-1}\right). \end{aligned}$$

由于 $p,q>0$ 且 $p\neq q$，对任何 $k\geq 1$，$p^k-q^k$ 均非零，且与 $p-q$ 同号（因为 $x\mapsto x^k$ 在 $x>0$ 上严格递增（strictly increasing））。这里 $n\geq 2$，所以 $n$ 和 $n-1$ 都至少是 1，两个括号均非零且同号。它们的乘积为正，再乘回 $pq>0$，得

$$\mathrm{P}(S=2n)>\mathrm{P}(A=2n)\qquad(n>1).$$

（当 $n=1$ 时，第二个括号是 $p^0-q^0=0$，这正是第 (ii) 部分的等式。）

**奇数长度**。同样地，仍设 $n>1$，

$$\frac{\mathrm{P}(S=2n+1)}{pq}=p^{2n}+q^{2n},$$

$$\frac{\mathrm{P}(A=2n+1)}{pq}=p^{n-1}q^{n-1}\left(p^2+q^2\right)=p^{n+1}q^{n-1}+p^{n-1}q^{n+1},$$

而

$$\begin{aligned}p^{2n}+q^{2n}-p^{n+1}q^{n-1}-p^{n-1}q^{n+1}&=p^{n+1}\left(p^{n-1}-q^{n-1}\right)-q^{n+1}\left(p^{n-1}-q^{n-1}\right)\\&=\left(p^{n+1}-q^{n+1}\right)\left(p^{n-1}-q^{n-1}\right).\end{aligned}$$

根据同样的论证，两个括号均非零且与 $p-q$ 同号，所以

$$\mathrm{P}(S=2n+1)>\mathrm{P}(A=2n+1)\qquad(n>1).$$

同样，因子 $p^{n-1}-q^{n-1}$ 在 $n=1$ 时为零，与 $\mathrm{P}(S=3)=\mathrm{P}(A=3)$ 吻合。

总括而言：长度为 1 的相同连串比交替连串出现的机会小，长度 2 和 3 的机会相等，而由 4 起的每个长度，相同连串出现的机会都较大。

## 验算

- 取 $p=\frac{1}{3}$，$q=\frac{2}{3}$。则 $\mathrm{P}(S=4)=pq(p^3+q^3)=\frac{2}{9}\cdot\frac{1}{3}=\frac{2}{27}\approx 0.074$，而 $\mathrm{P}(A=4)=p^2q^2=\frac{4}{81}\approx 0.049$，所以 $S=4$ 的机会确实较大。
- 将 $\mathrm{P}(S=n)$ 对所有 $n\geq 1$ 求和，得 $q\cdot\frac{p}{1-p}+p\cdot\frac{q}{1-q}=p+q=1$，理应如此。
- 每条公式都已用电脑核实：对多个 $p$ 值，以精确分数列举所有 14 次掷币的序列，验证长度至 13 的情况。

## 常见错误

- 忘记那次结束连串的掷币。$S=n$ 需要第 $n+1$ 次掷币不同；缺了这一步，你计算的其实是 $\mathrm{P}(S\geq n)$。

- 交替连串的结尾字母弄错：奇数长度的连串以开头的同一字母结尾，偶数长度的则以另一字母结尾。
- 在第 (iii) 部分代入 $q = 1 - p$：这样各式会变成 $p$ 的多项式，因式分解难看得多。

## STEP 之外

$S$ 是一个几何型随机变量（geometric-type random variable）：给定第一次掷币的结果，连串长度就是等到第一次改变的等候时间（waiting time）。这类连串是统计学中游程检定（runs test）的基础，该检定用来判断一串结果看来是否随机；而生成函数（generating functions）则提供一个简洁的方法，求出最长连串的完整分布。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）以合适的程度讲解生成函数和等候时间。

详细解答

# STEP III 2013，第 1 题

纯数学 · 半角代换

**STEP III 2013，第 1 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q1

## 关键思路

半角代换（half-angle substitution）$t = \tan\frac{1}{2}x$ 可把任何 $\sin x$ 的有理函数（rational function）化为 $t$ 的有理函数，因此第一个积分会变成 $\int \mathrm{d}t/(\text{二次式})$，再用配方法（completing the square）即可完成。要把答案写成题目要求的形式，则需要用到 $\tan(A-B)$ 公式。至于 $I_n$，提示的设计正是要令那个麻烦的分母约去：$\sin^{n+1}x + 2\sin^n x = \sin^n x\,(2+\sin x)$。由此得到一条递推关系（recurrence），从 $I_0$ 开始逐级往上推；而 $I_0$ 是唯一真正需要用到第一部分结果的积分。

## 解答

### 代换

设 $t = \tan\frac{1}{2}x$，由连锁法则（chain rule）得

$$\frac{\mathrm{d}t}{\mathrm{d}x} = \tfrac{1}{2}\sec^2\tfrac{1}{2}x = \tfrac{1}{2}\left(1+\tan^2\tfrac{1}{2}x\right) = \tfrac{1}{2}(1+t^2).$$

至于正弦，先用二倍角公式（double-angle formula），再写成 $\cos^2\frac{1}{2}x = 1/\sec^2\frac{1}{2}x$：

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

两个结果均已证明。

### 第一个积分

当 $x$ 由 0 变到 $\frac{1}{2}\pi$ 时，$t$ 由 $\tan 0 = 0$ 变到 $\tan\frac{1}{4}\pi = 1$，而且 $\mathrm{d}x = \dfrac{2\,\mathrm{d}t}{1+t^2}$。此外

$$1 + a\sin x = \frac{1 + t^2 + 2at}{1 + t^2},$$

因此因子 $1 + t^2$ 互相约去，得

$$\int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{1 + a\sin x} = \int_0^1 \frac{2\,\mathrm{d}t}{t^2 + 2at + 1} = \int_0^1 \frac{2\,\mathrm{d}t}{(t+a)^2 + (1-a^2)}.$$

由于 $0 < a < 1$，常数 $1 - a^2$ 为正；记 $r = \sqrt{1-a^2}$。利用 $\int \mathrm{d}u/(u^2 + r^2) = \frac{1}{r}\arctan\frac{u}{r}$，

$$\int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{1 + a\sin x} = \frac{2}{r}\Big[\arctan\frac{t+a}{r}\Big]_0^1 = \frac{2}{r}(\alpha - \beta),$$

其中

$$\tan\alpha = \frac{1+a}{r} = \frac{1+a}{\sqrt{1-a}\sqrt{1+a}} = \frac{\sqrt{1+a}}{\sqrt{1-a}}, \qquad \tan\beta = \frac{a}{r},$$

且 $\alpha, \beta \in (0, \frac{1}{2}\pi)$。

> **想一想** 这看来不像我们想要的答案。怎样才能把两个反正切之差化为一个反正切?

取差的正切。我们有

$$\tan\alpha - \tan\beta = \frac{(1+a) - a}{r} = \frac{1}{r}, \qquad 1 + \tan\alpha\tan\beta = 1 + \frac{a(1+a)}{1-a^2} = 1 + \frac{a}{1-a} = \frac{1}{1-a},$$

所以

$$\tan(\alpha - \beta) = \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta} = \frac{1-a}{r} = \frac{1-a}{\sqrt{1-a}\sqrt{1+a}} = \frac{\sqrt{1-a}}{\sqrt{1+a}}.$$

在取反函数之前，我们需要 $\alpha - \beta$ 位于 arctan 的值域 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 之内。事实确是如此：两个角都在 $(0, \frac{1}{2}\pi)$ 内，而且 $\tan\alpha > \tan\beta$（两者之差为 $1/r > 0$），所以 $0 < \alpha - \beta < \frac{1}{2}\pi$。因此 $\alpha - \beta = \arctan\dfrac{\sqrt{1-a}}{\sqrt{1+a}}$，并且

$$\int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{1 + a\sin x} = \frac{2}{\sqrt{1-a^2}}\arctan\frac{\sqrt{1-a}}{\sqrt{1+a}},$$

此即所求。

### 计算 $I_3$

按照提示，

$$I_{n+1} + 2I_n = \int_0^{\frac{1}{2}\pi} \frac{\sin^n x\,(\sin x + 2)}{2 + \sin x}\,\mathrm{d}x = \int_0^{\frac{1}{2}\pi} \sin^n x\,\mathrm{d}x.$$

我们需要右方在 $n = 0, 1, 2$ 时的值：

$$\int_0^{\frac{1}{2}\pi} 1\,\mathrm{d}x = \tfrac{1}{2}\pi, \qquad \int_0^{\frac{1}{2}\pi} \sin x\,\mathrm{d}x = 1, \qquad \int_0^{\frac{1}{2}\pi} \sin^2 x\,\mathrm{d}x = \int_0^{\frac{1}{2}\pi} \tfrac{1}{2}(1-\cos 2x)\,\mathrm{d}x = \tfrac{1}{4}\pi.$$

**想一想** 递推关系把每个 $I_n$ 与下一个联系起来，但它从哪里开始？

我们需要另行求出 $I_0$，这正是第一部分派上用场的地方。提出因子 2 并取 $a = \frac{1}{2}$，

$$\begin{aligned} I_0 = \int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{2+\sin x} &= \frac{1}{2}\int_0^{\frac{1}{2}\pi} \frac{\mathrm{d}x}{1+\frac{1}{2}\sin x} = \frac{1}{2}\cdot\frac{2}{\sqrt{3/4}} \arctan\frac{\sqrt{1/2}}{\sqrt{3/2}} \\ &= \frac{2}{\sqrt{3}} \arctan\frac{1}{\sqrt{3}} = \frac{2}{\sqrt{3}}\cdot\frac{\pi}{6} = \frac{\pi}{3\sqrt{3}}. \end{aligned}$$

现在沿递推关系 $I_{n+1} = \int_0^{\frac{1}{2}\pi} \sin^n x\,\mathrm{d}x - 2I_n$ 逐级往上推：

$$\begin{aligned} I_1 &= \tfrac{1}{2}\pi - 2I_0 = \frac{\pi}{2} - \frac{2\pi}{3\sqrt{3}}, \\ I_2 &= 1 - 2I_1 = 1 - \pi + \frac{4\pi}{3\sqrt{3}}, \\ I_3 &= \tfrac{1}{4}\pi - 2I_2 = \frac{\pi}{4} - 2 + 2\pi - \frac{8\pi}{3\sqrt{3}}. \end{aligned}$$

因此

$$I_3 = \frac{9\pi}{4} - \frac{8\sqrt{3}\,\pi}{9} - 2.$$

## 验算

- 当 $a \to 0$ 时，第一个结果应给出 $\int_0^{\frac{1}{2}\pi} 1\,\mathrm{d}x = \frac{1}{2}\pi$；它的确趋向 $2\arctan 1 = \frac{1}{2}\pi$。
- 由于在 $[0, \frac{1}{2}\pi]$ 上 $0 \le \sin^3 x/(2+\sin x) \le \frac{1}{3}$，我们需要 $0 < I_3 < \frac{1}{6}\pi \approx 0.52$。我们的答案是 $I_3 \approx 0.2318$，而 $I_3$ 的数值积分结果与之在许多位小数上吻合。同样，$I_0 = \pi/(3\sqrt{3}) \approx 0.6046$，它理应介乎 $\frac{\pi}{6}$ 与 $\frac{\pi}{4}$ 之间，事实亦然。
- 各值递减，$I_0 > I_1 > I_2 > I_3 > 0$，这是应有的结果，因为在该区间上 $0 \le \sin x \le 1$。

## 常见错误

- 忘记把积分上下限改为 $t = 0$ 和 $t = 1$，或忘记 $\mathrm{d}x$ 带来的因子 $2/(1+t^2)$。
- 用 $\tan(\alpha - \beta)$ 推断 $\alpha - \beta = \arctan(\dots)$，却没有检查 $\alpha - \beta$ 是否位于 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 之内。

- 以 $a = 2$ 把第一个结果套用于 $I_0$：该结果只在 $0 < a < 1$ 时成立，所以必须先提出因子 2。

## STEP 之外

$t = \tan\frac{1}{2}x$ 代换（常称为魏尔斯特拉斯代换（Weierstrass substitution））说明了每个 $\sin x$ 与 $\cos x$ 的有理函数都有初等的原函数（antiderivative），因为部分分式（partial fractions）可以处理任何 $t$ 的有理函数。从几何角度看，它就是单位圆的球极投影参数化（stereographic parametrisation），$(\cos x, \sin x) = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)$，而由有理数 $t$ 出发，它亦能产生所有毕氏三元数组（Pythagorean triple）。像 $\int_0^{2\pi} \mathrm{d}x/(1 + a\sin x)$ 这类积分，亦是复分析（complex analysis）中围道积分（contour integration）的经典入门应用；Needham 的 *Visual Complex Analysis*（《直观复分析》）是理解个中原因的绝佳读物。

详细解答

# STEP III 2013，第 2 题

纯数学 · 麦克劳林级数

**STEP III 2013，第 2 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q2

## 关键思路

先写成 $y\sqrt{1-x^2} = \arcsin x$ 以消去平方根，一阶导数（derivative）便轻而易举。再用莱布尼兹法则（Leibniz's rule）对这条一阶方程反复求导，就得到高阶关系式。在 $x = 0$ 处大部分项都消失，剩下一条隔两步的导数递推关系（recurrence relation），由此逐一构造出麦克劳林（Maclaurin）系数。最后的和式看起来正是该级数在一个精心选取的 $x$ 值处的取值。

## 解答

### 一阶方程

两边乘开得 $y\sqrt{1-x^2} = \arcsin x$，然后两边对 $x$ 求导，左边用乘积法则（product rule）：

$$\sqrt{1-x^2}\,\frac{\mathrm{d}y}{\mathrm{d}x} - \frac{xy}{\sqrt{1-x^2}} = \frac{1}{\sqrt{1-x^2}}.$$

全式乘以 $\sqrt{1-x^2}$（当 $|x| < 1$ 时不为零），得

$$(1-x^2)\frac{\mathrm{d}y}{\mathrm{d}x} - xy - 1 = 0,$$

即为所求。

### 高阶关系式

以 $y^{(k)}$ 表示 $\dfrac{\mathrm{d}^k y}{\mathrm{d}x^k}$。将一阶方程求导 $n+1$ 次。常数 1 消失，而两个乘积则由莱布尼兹法则处理。由于 $1-x^2$ 的各阶导数依次为 $-2x$、$-2$，之后是 0，而 $x$ 的导数依次为 1，之后

是 0,

$$\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}}\Big[(1-x^2)y'\Big] = (1-x^2)y^{(n+2)} + \binom{n+1}{1}(-2x)\,y^{(n+1)} + \binom{n+1}{2}(-2)\,y^{(n)}$$
$$= (1-x^2)y^{(n+2)} - 2(n+1)x\,y^{(n+1)} - n(n+1)\,y^{(n)},$$

$$\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}}\Big[xy\Big] = x\,y^{(n+1)} + (n+1)\,y^{(n)}.$$

第一式减去第二式并合并同类项,

$$(1-x^2)y^{(n+2)} - \big(2(n+1)+1\big)x\,y^{(n+1)} - \big(n(n+1)+(n+1)\big)y^{(n)} = 0,$$

即

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

以上推导并不需要 $n \geq 1$: 只求导一次($n=0$)也得到同样的模式 $(1-x^2)y'' - 3xy' - y = 0$，而我们稍后也会用到这个情况。

### 麦克劳林级数

代入 $x = 0$。中间一项带有因子 $x$，因此消失，得

$$y^{(n+2)}(0) = (n+1)^2\,y^{(n)}(0) \qquad (n \geq 0).$$

这条递推关系每次跨两步，所以需要两个初始值。由定义，$y(0) = \arcsin 0 = 0$；而一阶方程在 $x = 0$ 处给出 $y'(0) - 0 - 1 = 0$，所以 $y'(0) = 1$。

**想一想** 怎样只凭这两个值得出每一阶导数?

递推关系只把 $y^{(n)}(0)$ 与 $y^{(n+2)}(0)$ 联系起来，所以偶数阶导数组成一条由 $y(0)$ 开始的链，奇数阶导数则组成另一条由 $y'(0)$ 开始的链。

偶次幂。由于 $y(0) = 0$，递推关系给出 $y''(0) = 1^2 \cdot 0 = 0$，然后 $y^{(4)}(0) = 0$，如此类推: 对每个 $k$ 都有 $y^{(2k)}(0) = 0$。因此 $x$ 的每个偶次幂的系数都是零。(这与 $y$ 是奇函数（odd function）相符，因为它是奇函数 $\arcsin x$ 除以偶函数 $\sqrt{1-x^2}$ 的商。)

奇次幂。以 $n = 1, 3, \ldots, 2k-1$ 代入递推关系,

$$y^{(2k+1)}(0) = (2k)^2(2k-2)^2 \cdots 4^2 \cdot 2^2 \cdot y'(0) = \big(2 \cdot 4 \cdots 2k\big)^2 = \big(2^k\,k!\big)^2,$$

这对 $k = 0$ 亦成立。除以 $(2k+1)!$ 便得到麦克劳林系数:

$$\frac{\arcsin x}{\sqrt{1-x^2}} = \sum_{k=0}^{\infty} \frac{2^{2k}(k!)^2}{(2k+1)!}\,x^{2k+1} = x + \frac{2^2}{3!}x^3 + \frac{2^2 \cdot 4^2}{5!}x^5 + \cdots,$$

所以一般奇次项为 $\dfrac{2^{2k}(k!)^2}{(2k+1)!}x^{2k+1}$，而一般偶次项为 $0 \cdot x^{2k}$。

### 无穷级数之和

和式的第 $n$ 项是 $\dfrac{(n!)^2}{(2n+1)!}$，其中 $n=0$ 给出首项 1。与级数比较，我们想消去因子 $2^{2n}$，所以选取 $x$ 使 $2^{2n}x^{2n+1}$ 为 1 的常数倍：$x=\frac{1}{2}$ 给出 $2^{2n}\left(\frac{1}{2}\right)^{2n+1}=\frac{1}{2}$。因此

$$\frac{\arcsin\frac{1}{2}}{\sqrt{1-\frac{1}{4}}}=\frac{1}{2}\sum_{n=0}^{\infty}\frac{(n!)^2}{(2n+1)!}.$$

左边等于 $\dfrac{\pi/6}{\sqrt{3}/2}=\dfrac{\pi}{3\sqrt{3}}$，所以

$$1+\frac{1}{3!}+\frac{2^2}{5!}+\frac{2^2\times 3^2}{7!}+\cdots=\frac{2\pi}{3\sqrt{3}}=\frac{2\sqrt{3}\,\pi}{9}.$$

## 验算

- 将已知级数 $\arcsin x=x+\frac{1}{6}x^3+\cdots$ 乘以 $(1-x^2)^{-1/2}=1+\frac{1}{2}x^2+\cdots$，得 $x+\left(\frac{1}{6}+\frac{1}{2}\right)x^3+\cdots=x+\frac{2}{3}x^3+\cdots$，而 $\frac{2^2}{3!}=\frac{2}{3}$ 果然吻合。
- 和式的首五项为 $1+0.16667+0.03333+0.00714+0.00159\approx 1.2087$，已经接近 $\frac{2\pi}{3\sqrt{3}}\approx 1.20920$；各项每次大约缩小 4 倍，所以余项很小。

## 常见错误

- 在莱布尼兹法则中漏掉 $y^{(n)}$ 项上的二项式系数（binomial coefficient）$\binom{n+1}{2}$，以致得不出 $(n+1)^2$。
- 只对 $n\geq 1$ 使用递推关系，因而始终没有证明 $y''(0)=0$；偶数链需要 $n=0$（求导一次）的情况。
- 在最后一部分代入 $x=1$ 或 $x=\frac{1}{4}$：因子 $2^{2n}$ 必须由 $x^{2n}$ 抵消，这迫使 $x=\frac{1}{2}$。

## STEP 之外

这里我们“忽略了收敛性的问题”，但大学第一门分析课会把这点严格化：由比值审敛法（ratio test）可知该级数的收敛半径（radius of convergence）为 1，而带余项的泰勒定理（Taylor's theorem with remainder）则证明在该区间内级数确实等于该函数。系数 $\frac{2^{2k}(k!)^2}{(2k+1)!}$ 与中心二项式系数（central binomial coefficients）关系密切，而同一个函数积分一次后便可生成 $(\arcsin x)^2$ 的级数。关于幂级数背后的分析，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

# STEP III 2013，第 3 题

纯数学 · 向量与球面

> **STEP III 2013，第 3 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s3-q3

## 关键思路

一切都源于两个事实：每个 $\mathbf{p}_i$ 都是单位向量（unit vector），而且四者之和为 $\mathbf{0}$。把长度的平方展开为点积（dot product），就能把 $(XP_i)^2$ 化成关于 $\mathbf{x}\cdot\mathbf{p}_i$ 的线性（linear）式子；由于 $\sum\mathbf{p}_i=\mathbf{0}$，线性项之和便会全部消去。至于四次方，会剩下一个二次项，因此需要实际的坐标，而这正是第 (ii) 部分所提供的。

## 解答

### 点积 $\mathbf{p}_i\cdot\mathbf{p}_j$

由于这是正四面体（regular tetrahedron），六条棱的长度全部相等。对 $i\neq j$，

$$(P_iP_j)^2=(\mathbf{p}_i-\mathbf{p}_j)\cdot(\mathbf{p}_i-\mathbf{p}_j)=|\mathbf{p}_i|^2+|\mathbf{p}_j|^2-2\,\mathbf{p}_i\cdot\mathbf{p}_j=2-2\,\mathbf{p}_i\cdot\mathbf{p}_j,$$

这里用了 $|\mathbf{p}_i|=1$（每个顶点都在单位球面（unit sphere）上）。因此，棱长相等迫使每个 $\mathbf{p}_i\cdot\mathbf{p}_j$（$i\neq j$）都取同一个值；记这个值为 $c$。

现在把 $\mathbf{p}_1+\mathbf{p}_2+\mathbf{p}_3+\mathbf{p}_4=\mathbf{0}$ 与 $\mathbf{p}_1$ 作点积：

$$\mathbf{p}_1\cdot\mathbf{p}_1+\mathbf{p}_1\cdot\mathbf{p}_2+\mathbf{p}_1\cdot\mathbf{p}_3+\mathbf{p}_1\cdot\mathbf{p}_4=0\quad\Longrightarrow\quad 1+3c=0,$$

所以 $c=-\frac{1}{3}$。即对所有 $i\neq j$，都有 $\mathbf{p}_i\cdot\mathbf{p}_j=-\frac{1}{3}$，得证。

### 第 (i) 部分

由于 $|\mathbf{x}|=1$ 亦成立，

$$(XP_i)^2=(\mathbf{p}_i-\mathbf{x})\cdot(\mathbf{p}_i-\mathbf{x})=|\mathbf{p}_i|^2-2\,\mathbf{x}\cdot\mathbf{p}_i+|\mathbf{x}|^2=2-2\,\mathbf{x}\cdot\mathbf{p}_i.$$

对 $i$ 求和，并利用点积的分配律（distributive law），

$$\sum_{i=1}^{4}(XP_i)^2 = 8 - 2\mathbf{x}\cdot\left(\mathbf{p}_1+\mathbf{p}_2+\mathbf{p}_3+\mathbf{p}_4\right) = 8 - 2\mathbf{x}\cdot\mathbf{0} = 8,$$

得证。

**第 (ii) 部分**

设 $\mathbf{p}_1 = (0,0,1)$ 及 $\mathbf{p}_2 = (a,0,b)$，条件 $\mathbf{p}_1\cdot\mathbf{p}_2 = -\frac{1}{3}$ 立即给出 $b = -\frac{1}{3}$。再由 $|\mathbf{p}_2| = 1$ 得

$$a^2 = 1 - b^2 = \tfrac{8}{9}, \qquad \text{所以} \qquad a = \frac{2\sqrt{2}}{3},$$

由于 $a > 0$，取正根。

设 $\mathbf{p}_3 = (u,v,w)$。三个条件把它确定下来：

$$\mathbf{p}_1\cdot\mathbf{p}_3 = -\tfrac{1}{3} \implies w = -\tfrac{1}{3},$$

$$\mathbf{p}_2\cdot\mathbf{p}_3 = -\tfrac{1}{3} \implies \tfrac{2\sqrt{2}}{3}u + \tfrac{1}{9} = -\tfrac{1}{3} \implies u = -\tfrac{4}{9}\cdot\tfrac{3}{2\sqrt{2}} = -\frac{\sqrt{2}}{3},$$

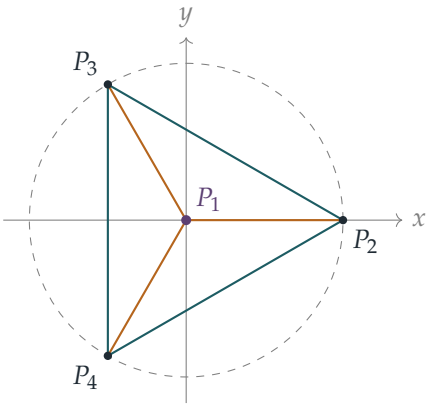
$$|\mathbf{p}_3| = 1 \implies v^2 = 1 - \tfrac{2}{9} - \tfrac{1}{9} = \tfrac{6}{9} \implies v = \pm\frac{\sqrt{6}}{3}.$$

$\mathbf{p}_4$ 满足完全相同的方程，所以 $P_3$ 和 $P_4$ 就是由此得出的两个点（它们必须是不同的点，所以各取一个符号）。按其中一种方式标记，

$$P_3 = \left(-\frac{\sqrt{2}}{3}, \frac{\sqrt{6}}{3}, -\frac{1}{3}\right), \qquad P_4 = \left(-\frac{\sqrt{2}}{3}, -\frac{\sqrt{6}}{3}, -\frac{1}{3}\right).$$

（$P_3$ 和 $P_4$ 的标记可以互换。）验证一下：四个位置向量（position vector）之和确实为 $\mathbf{0}$，而且 $\mathbf{p}_3\cdot\mathbf{p}_4 = \frac{2}{9} - \frac{6}{9} + \frac{1}{9} = -\frac{1}{3}$。

从上方俯视，$P_2$、$P_3$、$P_4$ 在平面 $z = -\frac{1}{3}$ 上组成一个等边三角形（equilateral triangle），而 $P_1$ 位于其中心的正上方：



**第 (iii) 部分**

由第 (i) 部分，$(XP_i)^2 = 2(1 - \mathbf{x}\cdot\mathbf{p}_i)$，两边平方，

$$(XP_i)^4 = 4(1-\mathbf{x}\cdot\mathbf{p}_i)^2 \qquad \Longrightarrow \qquad \sum_{i=1}^{4}(XP_i)^4 = 4\sum_{i=1}^{4}(1-\mathbf{x}\cdot\mathbf{p}_i)^2,$$

得证。展开括号，并再次利用 $\sum \mathbf{x}\cdot\mathbf{p}_i = 0$，

$$\sum_{i=1}^{4}(XP_i)^4 = 4\Big(4 - 2\sum_{i=1}^{4}\mathbf{x}\cdot\mathbf{p}_i + \sum_{i=1}^{4}(\mathbf{x}\cdot\mathbf{p}_i)^2\Big) = 16 + 4\sum_{i=1}^{4}(\mathbf{x}\cdot\mathbf{p}_i)^2.$$

? **想一想** 线性项又一次消失了。为甚么同一招不能把 $\sum(\mathbf{x}\cdot\mathbf{p}_i)^2$ 也消掉?

平方和对 $\mathbf{p}_i$ 而言并非线性，所以知道 $\sum \mathbf{p}_i = \mathbf{0}$ 并不能直接告诉我们它的值。这正是需要用到坐标的地方。设 $\mathbf{x} = (x,y,z)$，并用第 (ii) 部分的顶点，

$$\mathbf{x}\cdot\mathbf{p}_1 = z, \qquad \mathbf{x}\cdot\mathbf{p}_2 = \tfrac{1}{3}(2\sqrt{2}\,x - z),$$
$$\mathbf{x}\cdot\mathbf{p}_3 = \tfrac{1}{3}(-\sqrt{2}\,x + \sqrt{6}\,y - z), \qquad \mathbf{x}\cdot\mathbf{p}_4 = \tfrac{1}{3}(-\sqrt{2}\,x - \sqrt{6}\,y - z).$$

平方后相加，合并同类项：

- $x^2$：$\frac{1}{9}(8+2+2) = \frac{4}{3}$；　$y^2$：$\frac{1}{9}(6+6) = \frac{4}{3}$；　$z^2$：$1 + \frac{1}{9}(1+1+1) = \frac{4}{3}$。
- $xy$ 和 $yz$：这两项只出现在 $\mathbf{x}\cdot\mathbf{p}_3$ 和 $\mathbf{x}\cdot\mathbf{p}_4$ 之中，而这两者只在 $y$ 项的符号上有分别，所以互相抵消。
- $xz$：$\frac{1}{9}(-4\sqrt{2} + 2\sqrt{2} + 2\sqrt{2}) = 0$。

因此

$$\sum_{i=1}^{4}(\mathbf{x}\cdot\mathbf{p}_i)^2 = \tfrac{4}{3}(x^2+y^2+z^2) = \tfrac{4}{3},$$

因为 $X$ 在单位球面上。所以

$$\sum_{i=1}^{4}(XP_i)^4 = 16 + 4\cdot\tfrac{4}{3} = \frac{64}{3},$$

这与 $X$ 的位置无关，得证。

? **想一想** 我们用了特定的坐标。这个结论对内接于球面的每一个正四面体都成立吗?

成立。任何这样的四面体都可以绕 $O$ 旋转（rotation），使 $P_1$ 转到 $(0,0,1)$，而 $P_2$ 转到半平面 $y = 0$、$x > 0$ 之内；旋转保持长度不变，而当 $X$ 走遍整个球面时，它旋转后的像亦走遍整个球面。所以第 (ii) 部分的坐标并不失一般性。

## 验算

- 取 $X = P_1$。各距离为 0 及三条棱，每条长 $\sqrt{2-2(-\frac{1}{3})} = \sqrt{8/3}$。于是 $\sum (XP_i)^2 = 3 \cdot \frac{8}{3} = 8$，而 $\sum (XP_i)^4 = 3 \cdot \frac{64}{9} = \frac{64}{3}$，与两个结果都吻合。
- 取 $X = (0, 0, -1)$，即 $P_1$ 的对径点。此时 $XP_1 = 2$，而其余三个顶点有 $\mathbf{x} \cdot \mathbf{p}_i = \frac{1}{3}$，所以 $(XP_i)^2 = \frac{4}{3}$。这给出 $4 + 3 \cdot \frac{4}{3} = 8$ 及 $16 + 3 \cdot \frac{16}{9} = \frac{64}{3} \approx 21.33$。
- 用电脑在球面上数以千计的随机点上验算，每次都得到 8 和 21.333…。

## 常见错误

- 断言所有 $\mathbf{p}_i \cdot \mathbf{p}_j$ 都相等，却没有说明原因：原因是棱长相等，通过 $(P_iP_j)^2 = 2 - 2\,\mathbf{p}_i \cdot \mathbf{p}_j$ 得出。
- 忘记 $|\mathbf{x}| = 1$；推导 $(XP_i)^2 = 2 - 2\,\mathbf{x} \cdot \mathbf{p}_i$ 和最后一步 $x^2 + y^2 + z^2 = 1$ 都需要用到这一点。
- 第 (iii) 部分交叉项（cross terms）的计算错误：那里一个符号出错，就会留下对 $X$ 的虚假依赖。

## STEP 之外

第 (iii) 部分的核心是恒等式 $\sum_i (\mathbf{x} \cdot \mathbf{p}_i)^2 = \frac{4}{3}|\mathbf{x}|^2$，它对每一个向量 $\mathbf{x}$ 都成立。用矩阵（matrix）的语言来说，就是 $\sum_i \mathbf{p}_i \mathbf{p}_i^{\mathsf{T}} = \frac{4}{3} I$：这四个顶点向量的表现犹如一组超完备的标准正交基（over-complete orthonormal basis），称为紧框架（tight frame）。这一点由对称性必然成立：该矩阵与四面体的每一个旋转都可交换，因此必定是 $I$ 的倍数，而它的迹（trace）是 4。球面上的点集，若其低次平均值与球面本身的平均值相同，便称为球面设计（spherical designs）；它们应用于数值积分（numerical integration）和编码理论（coding theory）。关于背后的线性代数（外积（outer product）、对称矩阵（symmetric matrix）和特征值（eigenvalue）），可参阅斯特朗（Strang）的 *Introduction to Linear Algebra*（《线性代数导论》）。

详细解答

# STEP III 2013，第 4 题

纯数学 · 单位根的乘积

**STEP III 2013，第 4 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q4

本题在 *Beyond STEP* 第 5.3 节因式分解 $z^{2n}+1$ 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/z2n-plus-one.html

## 关键思路

$z^{2n}=-1$ 的根以共轭（conjugate）对 $\mathrm{e}^{\pm \mathrm{i}\theta}$ 的形式出现，而每一对相乘都得出一个实系数二次式，由此便得到所需的因式分解（factorisation）。接着，只要选取一个恰当的 $z$ 值，就能把这个因式分解变成一个关于余弦的命题：当 $z=\mathrm{i}$ 时，每个二次式中 $z^2+1$ 的部分都等于零，只剩下 $-2z\cos\theta$ 这一项。在第 (ii) 部分，同样的代入会令两边都等于零，所以我们要先约去在 $z=\mathrm{i}$ 时等于零的那个因式。

## 解答

### 因式分解

展开，

$$\begin{aligned}(z-\mathrm{e}^{\mathrm{i}\theta})(z-\mathrm{e}^{-\mathrm{i}\theta}) &= z^2 - z\left(\mathrm{e}^{\mathrm{i}\theta}+\mathrm{e}^{-\mathrm{i}\theta}\right)+\mathrm{e}^{\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}\theta}\\ &= z^2-2z\cos\theta+1,\end{aligned}$$

因为 $\mathrm{e}^{\mathrm{i}\theta}+\mathrm{e}^{-\mathrm{i}\theta}=2\cos\theta$。

现在，$\mathrm{e}^{\mathrm{i}\theta}$ 满足 $z^{2n}=-1=\mathrm{e}^{\mathrm{i}\pi}$ 当且仅当 $2n\theta$ 是 $\pi$ 的奇数倍，即对某整数 $k$，$\theta=\frac{(2k-1)\pi}{2n}$。在 $(-\pi,\pi]$ 内的值是

$$\theta=\pm\frac{\pi}{2n},\ \pm\frac{3\pi}{2n},\ \ldots,\ \pm\frac{(2n-1)\pi}{2n},$$

因为 $\frac{(2n-1)\pi}{2n} < \pi$，而 $\frac{(2n+1)\pi}{2n} > \pi$。这些是 $2n$ 个互不相同的根，所以它们就是 $-1$ 的全部 $(2n)$ 次方根：

$$\mathrm{e}^{\pm \mathrm{i}\theta_k}, \qquad \theta_k = \frac{(2k-1)\pi}{2n}, \quad k = 1, 2, \ldots, n.$$

由于 $z^{2n}+1$ 是 $2n$ 次的首一（monic）多项式（polynomial），而且有这 $2n$ 个互不相同的根，

$$z^{2n}+1 = \prod_{k=1}^{n} (z - \mathrm{e}^{\mathrm{i}\theta_k})(z - \mathrm{e}^{-\mathrm{i}\theta_k}) = \prod_{k=1}^{n} \left(z^2 - 2z\cos\theta_k + 1\right),$$

这里对每一对共轭根都用了第一个结果。这就是所求的恒等式。

### 第 (i) 部分

代入 $z = \mathrm{i}$。每个因式变成

$$\mathrm{i}^2 - 2\mathrm{i}\cos\theta_k + 1 = -2\mathrm{i}\cos\theta_k,$$

而由于 $n$ 是偶数，左边是 $\mathrm{i}^{2n}+1 = (-1)^n + 1 = 2$。因此

$$2 = (-2\mathrm{i})^n \prod_{k=1}^{n} \cos\theta_k.$$

由于 $n$ 是偶数，$(-2)^n = 2^n$ 而 $\mathrm{i}^n = (\mathrm{i}^2)^{n/2} = (-1)^{\frac{1}{2}n}$，所以 $(-2\mathrm{i})^n = 2^n(-1)^{\frac{1}{2}n}$。因此

$$\prod_{k=1}^{n} \cos\theta_k = \frac{2}{2^n(-1)^{\frac{1}{2}n}} = (-1)^{\frac{1}{2}n}\, 2^{1-n},$$

这里用了 $1/(-1)^{\frac{1}{2}n} = (-1)^{\frac{1}{2}n}$。左边的乘积正是 $\cos\frac{\pi}{2n}\cos\frac{3\pi}{2n}\cdots\cos\frac{(2n-1)\pi}{2n}$，得证。

### 第 (ii) 部分

> **想一想** 为甚么不直接把 $z = \mathrm{i}$ 再代入因式分解?

当 $n$ 是奇数时，$\mathrm{i}^{2n}+1 = (-1)^n + 1 = 0$，所以左边等于零。在右边，中间的值 $k = m$（其中 $m = \frac{1}{2}(n+1)$）给出 $\theta_m = \frac{n\pi}{2n} = \frac{\pi}{2}$，所以该因式是 $z^2+1$，它在 $z = \mathrm{i}$ 时同样等于零。我们只得到 $0 = 0$，甚么也得不到。解决办法是先从两边约去因式 $z^2+1$。

利用题目给出的事实和因式分解，

$$(1+z^2)\left(1 - z^2 + z^4 - \cdots + z^{2n-2}\right) = (z^2+1) \prod_{k \neq m} \left(z^2 - 2z\cos\theta_k + 1\right).$$

这些都是多项式，而 $z^2+1$ 不是零多项式，所以我们可以把它约去（多项式除法的结果是唯一的）：

$$1 - z^2 + z^4 - \cdots + z^{2n-2} = \prod_{k \neq m} \left(z^2 - 2z\cos\theta_k + 1\right),$$

这是一个多项式恒等式，对每个 $z$ 都成立，包括 $z = \mathrm{i}$。

现在代入 $z = \mathrm{i}$。在左边，$(-1)^j z^{2j}$ 这一项变成 $(-1)^j(-1)^j = 1$，而共有 $n$ 项（$j = 0, 1, \ldots, n-1$），所以左边等于 $n$。右边有 $n-1$ 个因式，每个都等于 $-2\mathrm{i}\cos\theta_k$：

$$n = (-2\mathrm{i})^{n-1} \prod_{k \neq m} \cos\theta_k.$$

由于 $n-1$ 是偶数，与第 (i) 部分完全一样，$(-2\mathrm{i})^{n-1} = 2^{n-1}(-1)^{\frac{1}{2}(n-1)}$。

要把乘积化成平方，可把 $k$ 与 $n+1-k$ 配对。由于

$$\theta_{n+1-k} = \frac{(2n+1-2k)\pi}{2n} = \pi - \theta_k,$$

我们有 $\cos\theta_{n+1-k} = -\cos\theta_k$。所有 $k \neq m$ 的值可分成 $\frac{1}{2}(n-1)$ 对 $\{k, n+1-k\}$，其中 $1 \leq k \leq \frac{1}{2}(n-1)$，所以

$$\prod_{k \neq m} \cos\theta_k = \prod_{k=1}^{(n-1)/2} \left(-\cos^2\theta_k\right) = (-1)^{\frac{1}{2}(n-1)}\, P, \qquad P = \prod_{k=1}^{(n-1)/2} \cos^2\theta_k.$$

综合起来，两个符号 $(-1)^{\frac{1}{2}(n-1)}$ 相乘得 $+1$：

$$n = 2^{n-1}(-1)^{\frac{1}{2}(n-1)}(-1)^{\frac{1}{2}(n-1)} P = 2^{n-1} P,$$

所以 $P = n\,2^{1-n}$。$P$ 中最后一个因式对应 $k = \frac{1}{2}(n-1)$，即 $\theta_k = \frac{(n-2)\pi}{2n}$，所以 $P$ 正是

$$\cos^2\frac{\pi}{2n}\cos^2\frac{3\pi}{2n}\cdots\cos^2\frac{(n-2)\pi}{2n} = n\,2^{1-n},$$

得证。（当 $n = 1$ 时，乘积是空乘积，等于 $1 = 1 \cdot 2^0$，结果一致。）

## 验算

- 第 (i) 部分取 $n = 2$：$\cos\frac{\pi}{4}\cos\frac{3\pi}{4} = -\frac{1}{2} = (-1)^1 2^{-1}$。当 $n = 4$ 时，乘积是 $\frac{1}{8}$，与 $(-1)^2 2^{-3}$ 吻合。
- 第 (ii) 部分取 $n = 3$：$\cos^2\frac{\pi}{6} = \frac{3}{4} = 3 \cdot 2^{-2}$。当 $n = 5$ 时，$\cos^2\frac{\pi}{10}\cos^2\frac{3\pi}{10} = 0.3125 = \frac{5}{16}$。
- 两条恒等式亦已用电脑对所有不超过 39 的 $n$ 验证过，而因式分解也在随机选取的复数 $z$ 上测试过。

## 常见错误

- 在第 (ii) 部分直接把 $z = \mathrm{i}$ 代入因式分解，结果只得到 $0 = 0$。

- 弄错符号：在第 (i) 部分，当 $n \equiv 2 \pmod 4$ 时答案确实是负数，因为对于大于 $\frac{\pi}{2}$ 的那 $\frac{1}{2}n$ 个角，$\cos\theta_k < 0$。
- 忘记 $\cos\theta_{n+1-k} = -\cos\theta_k$，因而在组成平方时弄错符号。

## STEP 之外

同样的技巧——把 $z^N \pm 1$ 按其单位根（roots of unity）作因式分解，再在一个巧妙的点上求值——可以得出一整族乘积公式，例如 $\prod_{k=1}^{n-1} \sin\frac{k\pi}{n} = \frac{n}{2^{n-1}}$（试在 $\frac{z^n-1}{z-1}$ 中代入 $z=1$）。在这类有限乘积中令 $n \to \infty$，便会得到欧拉（Euler）的无穷乘积（infinite product）$\sin\pi x = \pi x \prod_{k\geq 1}\left(1-\frac{x^2}{k^2}\right)$，让你初尝复分析（complex analysis）如何由函数的零点把函数建构出来。Needham 的 *Visual Complex Analysis*（《直观复分析》）是从几何角度欣赏单位根及这些概念的绝佳读物。

详细解答

# STEP III 2013，第 5 题

纯数学 · 有理性与互质性

**STEP III 2013，第 5 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q5

## 关键思路

每一步都只靠一件工具：如果一个数整除一个乘积，而且与其中一个因数互质（coprime），那么它必定整除另一个因数。把有理数（rational number）写成最简分数（lowest terms），消去分母，然后留意每个因数必须落在哪一边。在第 (ii) 部分，这个方法把分母 $b$ 夹在两个事实之间，而除非 $b$ 很小，否则这两个事实不可能同时成立。

我们会反复用到一个小事实：*若 $p$ 与 $q$ 互质，则对任何正整数 $s$、$t$，$p^s$ 与 $q^t$ 亦互质*。事实上，若有一个质数（prime）同时整除 $p^s$ 和 $q^t$，它就会同时整除 $p$ 和 $q$（整除一个乘积的质数必整除其中一个因数，这正是题目所给结果在 $a$ 为质数时的情况），与互质的假设矛盾。

## 解答

### 第 (i) 部分

由 $q^nN = p^n$ 可见 $p^n$ 整除 $q^nN$。由于 $p$ 与 $q$ 互质，$p^n$ 与 $q^n$ 亦互质，于是由题目所给的结果（取 $a = p^n$、$b = q^n$、$c = N$）得知 $p^n$ 整除 $N$。因此对某个正整数 $k$，有 $N = kp^n$，即为所求。

代回原式，得 $q^nkp^n = p^n$，所以 $q^nk = 1$。两者皆为正整数，故 $k = 1$ 且 $q^n = 1$，从而

$$q = 1 \quad (\text{且} N = p^n).$$

现在取任意正整数 $n$ 和 $N$，并假设 $\sqrt[n]{N}$ 是有理数。它是正数，所以可以写成 $\sqrt[n]{N} = p/q$，其中 $p$、$q$ 为正整数，且分数为最简形式，即 $p$ 与 $q$ 互质。两边取 $n$ 次方，得 $q^nN = p^n$，而上面的推导给出 $q = 1$。因此 $\sqrt[n]{N} = p$ 是正整数。所以 $\sqrt[n]{N}$ 要么是正整数，要么是无

理数（irrational）。

### 第 (ii) 部分

已知 $a^a d^b = b^a c^b$，且 $\gcd(a,b) = \gcd(c,d) = 1$。

**证明** $d^b = b^a$。由右边可见 $b^a$ 整除 $a^a d^b$。由于 $a$ 与 $b$ 互质，$b^a$ 与 $a^a$ 互质，所以 $b^a$ 整除 $d^b$。同理，$d^b$ 整除 $b^a c^b$，而 $d^b$ 与 $c^b$ 互质，所以 $d^b$ 整除 $b^a$。两个互相整除的正整数必定相等：

$$d^b = b^a,$$

再在原方程中约去，亦得 $a^a = c^b$。

若 $p$ 是 $d$ 的一个质因数（prime factor），则 $p$ 整除 $d^b = b^a$，而整除 $b$ 的某个幂的质数必整除 $b$。所以 $p$ 是 $b$ 的一个质因数。

**$p$ 的幂**。设 $p^m$ 和 $p^n$ 分别是整除 $d$ 和 $b$ 的 $p$ 的最高次幂，由上一段可知 $m, n \geq 1$。由质因数分解唯一性（unique factorisation），整除 $d^b$ 的 $p$ 的最高次幂是 $p^{mb}$，而整除 $b^a$ 的则是 $p^{na}$。由于 $d^b = b^a$，两者相等，所以 $mb = na$，且

$$b = \frac{na}{m}.$$

> **想一想** $b = na/m$ 怎样给出 $p^n$ 的一个上界?

把它改写成 $mb = na$：于是 $b$ 整除 $na$。但 $b$ 与 $a$ 互质，所以由题目所给的结果，$b$ 整除 $n$，特别地 $b \leq n$。另外 $p^n$ 整除 $b$，所以 $p^n \leq b$。合起来，

$$p^n \leq b \leq n.$$

但 $p \geq 2$ 且 $n > 0$，所以由题目给出的事实 $y^x > x$（取 $y = p$、$x = n$）得 $p^n > n$。这个矛盾说明我们的假设不可能成立：$d$ 没有任何质因数。因此 $d = 1$，于是 $b^a = d^b = 1$，从而

$$b = 1.$$

### 最后部分

设 $r$ 是正有理数，且 $r^r$ 是有理数。把 $r = a/b$ 写成最简形式（$a, b$ 为正整数且互质），并把 $r^r = c/d$ 写成最简形式（由于 $r^r > 0$，$c, d$ 为正整数，且互质）。那么

$$\left(\frac{a}{b}\right)^{a/b} = \frac{c}{d} \quad \Longrightarrow \quad \left(\frac{a}{b}\right)^a = \left(\frac{c}{d}\right)^b \quad \Longrightarrow \quad a^a d^b = b^a c^b,$$

这里把两边取 $b$ 次方，再消去分母。这正好是第 (ii) 部分的情况，所以 $b = 1$，而 $r = a$ 是正整数。

## 验算

- 整数 $r$ 确实符合条件：$r=2$ 给出 $r^r=4$；而结论 $c^b=a^a$ 在 $b=1$ 时就是 $c=a^a$，理应如此。
- 一个反面例子：$r=\frac{1}{2}$ 给出 $r^r=1/\sqrt{2}$，是无理数，与结果一致。在第 (i) 部分，$\sqrt[3]{27}=3$ 而 $\sqrt[3]{24}$ 是无理数：两种情况都会出现。
- 用电脑搜索所有互质的 $a,b\le 24$、$c<200$、$d<60$，只找到 $b=d=1$ 时 $a^ad^b=b^ac^b$ 的解（即 $a=1,2,3$，且 $c=a^a$）。

## 常见错误

- 忘记交代为何 $p^n$ 与 $q^n$ 互质：题目所给的结果是关于 $a$ 与 $b$ 互质的，所以你需要那些幂互质。
- 由 $b=na/m$ 直接得出 $p^n\le n$，却漏了 $b\mid n$ 这一步：正是 $a$ 与 $b$ 互质，才把 $mb=na$ 变成 $b\mid n$。
- 写了 $(a/b)^{a/b}=c/d$，却没有坚持 $c/d$ 是最简形式；没有 $\gcd(c,d)=1$，$d^b\mid b^a$ 这一步就不成立。

## STEP 之外

题目所给的“互质因数”结果就是欧几里得（Euclid）引理（Euclid's lemma），通常由贝祖（Bézout）等式（Bézout's identity）证明，它也是质因数分解唯一性背后的推动力。第一部分是 $\sqrt{2}$ 为无理数这个结果的一般化版本；第二部分则让你初尝超越数论（transcendence theory）的滋味：在这个理论中，格尔丰德–施奈德定理（Gelfond–Schneider theorem）证明了，例如 $\sqrt{2}^{\sqrt{2}}$ 甚至不是代数数（algebraic number）。Davenport 的 *The Higher Arithmetic*（《高等算术》）把这些基础发展得非常优美。

详细解答

# STEP III 2013，第 6 题

纯数学 · 复数模的不等式

**STEP III 2013，第 6 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q6

本题在 *Beyond STEP* 第 5.8 节圆盘中隐藏的距离中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch05/poincare-disc.html

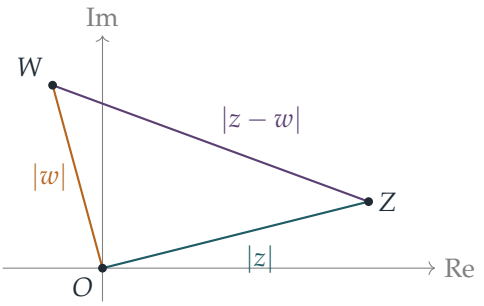
## 关键思路

一切都建基于把模的平方写成一个数乘以它的共轭（conjugate），即 $|u|^2 = uu^*$。用这个方法展开 $|z-w|^2$ 和 $|1-zw^*|^2$，两者都会出现相同的交叉项（cross terms）$zw^* + z^*w$，这正是同一个量 $E$ 在两式中都出现的原因。一旦把两个模（modulus）都写成“某数的平方减去 $E$”的形式，最后的不等式就变成比较 $\frac{A-E}{B-E}$ 与 $\frac{A}{B}$ 这两个分数，而这归结为 $B-A$ 的正负号。

## 解答

### 三角不等式

在阿冈图（Argand diagram）中标出原点 $O$，以及代表 $z$ 和 $w$ 的点 $Z$ 和 $W$。距离 $OZ$ 和 $OW$ 分别是 $|z|$ 和 $|w|$，而距离 $WZ$ 是 $|z-w|$。




线段 $WZ$ 是由 $W$ 到 $Z$ 的最短路径，所以它不会比路径 $W \to O \to Z$ 长：

$$|z-w| = WZ \le WO + OZ = |w| + |z|.$$

等号成立若且唯若 $O$ 位于线段 $WZ$ 上，即 $z$ 和 $w$ 指向相反方向（或其中一个为 0）。若 $O$、$Z$、$W$ 共线（collinear）但 $O$ 不在两者之间，则这个“三角形”是扁平的，但最短路径的论证仍然适用。

**第 (i) 部分**

利用 $|u|^2 = uu^*$ 及 $(z-w)^* = z^* - w^*$，

$$\begin{aligned} |z-w|^2 &= (z-w)(z^*-w^*) \\ &= zz^* + ww^* - zw^* - z^*w \\ &= |z|^2 + |w|^2 - (zw^* + z^*w). \end{aligned}$$

由于 $|zw| = |z||w|$，我们有 $(|z|+|w|)^2 = |z|^2 + |w|^2 + 2|zw|$，所以

$$(|z|+|w|)^2 - E = |z|^2 + |w|^2 + 2|zw| - zw^* - z^*w - 2|zw| = |z-w|^2,$$

得证。移项可得

$$E = (|z|+|w|)^2 - |z-w|^2.$$

右方是两个实数的平方之差，所以 $E$ 是实数。根据三角不等式（triangle inequality），$0 \le |z-w| \le |z| + |w|$，而把非负的量平方不会改变不等号方向，所以 $E \ge 0$。

（直接计算：$z^*w$ 是 $zw^*$ 的共轭，所以 $E = 2\,\mathrm{Re}(zw^*) + 2|zw^*|$，这是实数，而且由于 $\mathrm{Re}\,u \ge -|u|$，它至少为 0。）

### 第 (ii) 部分

同样地，利用 $(zw^*)^* = z^*w$，

$$\begin{aligned}|1 - zw^*|^2 &= (1 - zw^*)(1 - z^*w)\\ &= 1 - zw^* - z^*w + zz^*ww^*\\ &= 1 + |zw|^2 - (zw^* + z^*w)\\ &= \left(1 + 2|zw| + |zw|^2\right) - (zw^* + z^*w + 2|zw|)\\ &= (1 + |zw|)^2 - E,\end{aligned}$$

得证。

### 最后的不等式

设

$$A = (|z| + |w|)^2, \qquad B = (1 + |zw|)^2,$$

于是第 (i) 及 (ii) 部分表明 $|z - w|^2 = A - E$ 及 $|1 - zw^*|^2 = B - E$。

首先，左方是有意义的：若 $|z| > 1$ 且 $|w| > 1$，则 $|zw^*| = |z||w| > 1$，所以 $zw^* \neq 1$，且 $B - E = |1 - zw^*|^2 > 0$。此外 $B > 0$。

所求不等式的两边都是非负的，所以它等价（equivalent）于两边平方后的不等式

$$\frac{A - E}{B - E} \leq \frac{A}{B}.$$

两边同乘以正数 $B(B - E)$，这等价于

$$(A - E)B \leq A(B - E) \iff -EB \leq -EA \iff E(B - A) \geq 0.$$

**想一想** $B - A$ 的正负号是甚么？关于 $|z|$ 和 $|w|$ 的条件在哪里发挥作用？

设 $a = |z|$ 及 $b = |w|$，

$$B - A = (1 + ab)^2 - (a + b)^2 = 1 + a^2b^2 - a^2 - b^2 = (1 - a^2)(1 - b^2).$$

若 $a > 1$ 且 $b > 1$，两个括号都是负数，所以 $B - A > 0$。由于第 (i) 部分已证 $E \geq 0$，故 $E(B - A) \geq 0$。上面每一步都是等价的，所以

$$\frac{|z - w|}{|1 - zw^*|} \leq \frac{|z| + |w|}{1 + |zw|}.$$

等号成立若且唯若 $E = 0$，即 $z$ 和 $w$ 指向相反方向。

**当 $|z|<1$ 且 $|w|<1$ 时**

是的，不等式仍然成立。此时 $|zw^*| = ab < 1$，所以同样有 $zw^* \neq 1$ 及 $B - E > 0$，整条等价推导链原封不动地成立。这次 $(1-a^2)(1-b^2)$ 中两个括号都是正数，所以同样有 $B - A > 0$，得到 $E(B-A) \geq 0$，从而得出该不等式。

关键只在于 $1-a^2$ 与 $1-b^2$ 同号。若一个模大于 1 而另一个小于 1，则 $B - A < 0$，不等式可能不成立：取 $z = 2$、$w = 0.4$，左方为 $1.6/0.2 = 8$，而右方为 $2.4/1.8 = \frac{4}{3}$。

## 验算

- 取 $z = 2$、$w = 2\mathrm{i}$。则 $|z-w| = 2\sqrt{2}$，而 $1 - zw^* = 1 - 2(-2\mathrm{i}) = 1 + 4\mathrm{i}$，所以左方为 $2\sqrt{2}/\sqrt{17} \approx 0.686$，右方为 $4/5 = 0.8$。把两者都乘以 $\frac{1}{4}$（$z = \frac{1}{2}$、$w = \frac{1}{2}\mathrm{i}$）会得到同样的两个数，$0.686 \leq 0.8$，这说明了 $|z|, |w| < 1$ 的情况。
- 等号情况：$z = 2 + \mathrm{i}$、$w = -1.5z$ 给出 $E = 0$，两边准确至四位小数都等于 0.6577。
- 用电脑对数十万组随机数对进行检验，证实了两条恒等式、$E \geq 0$，以及两种情况下的不等式。

## 常见错误

- 写成 $|z-w|^2 = |z|^2 + |w|^2 - 2zw^*$：交叉项是 $zw^* + z^*w$，两者互为共轭，但一般而言并不相等。
- 交叉相乘前没有检查 $|1 - zw^*| \neq 0$；这里 $|zw| \neq 1$ 保证了这一点。
- 因为推导时用了 $|z|, |w| > 1$ 而在最后一部分回答“否”：因式分解（factorisation）$(1-a^2)(1-b^2)$ 显示，关键只在于正负号是否一致。

## STEP 之外

由于 $|1 - zw^*| = |1 - z^*w|$，左方是 $\dfrac{z-w}{1-z^*w}$ 的模。对于固定的 $z$（其中 $|z| < 1$），映射 $w \mapsto \dfrac{z-w}{1-z^*w}$ 是一个把单位圆盘（unit disc）映到自身的莫比乌斯变换（Möbius transformation），而 $\left|\frac{z-w}{1-z^*w}\right|$ 是 $z$ 与 $w$ 之间的伪双曲距离（pseudo-hyperbolic distance）。本题的不等式表明，这个距离不超过 $z$ 和 $w$ 各自到 0 的距离的“双曲和”：这是三角不等式的一种加强形式，也是双曲几何（hyperbolic geometry）的基础之一。Needham 的 *Visual Complex Analysis*（《直观复分析》）以精美的图像解说莫比乌斯变换和双曲圆盘。

详细解答

# STEP III 2013，第 7 题

纯数学 · 常微分方程的能量法

**STEP III 2013，第 7 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q7

本题在 *Beyond STEP* 第 8.2 节*不求解而估界*中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/bounding-without-solving.html

## 关键思路

每条方程都像一粒系在弹簧上的质点：一个只依赖于未知函数的“力”项，再加上（在 (ii) 和 (iii) 中）一个与导数（derivative）成正比的摩擦项。将方程乘以导数，就可以把它化成某个“能量”的导数。没有摩擦时，能量守恒；若摩擦从不朝错误方向推动，能量就只会减少，因此永远不会超过其初始值。第 (iii) 部分要求我们自行构造能量，并证明摩擦系数从不为负。

## 解答

### 第 (i) 部分

记 $y' = \frac{\mathrm{d}y}{\mathrm{d}x}$。由链式法则（chain rule），

$$\begin{aligned}\frac{\mathrm{dE}}{\mathrm{d}x} &= 2y'y'' + 2y^3y' \\ &= 2y'\left(y'' + y^3\right) = 0,\end{aligned}$$

这里用了微分方程（differential equation）。因此 E 是常数，其值为 $\mathrm{E}(0) = 0^2 + \frac{1}{2}\cdot 1^4 = \frac{1}{2}$。故对每个 $x$，

$$\tfrac{1}{2}y^4 = \tfrac{1}{2} - (y')^2 \le \tfrac{1}{2},$$

因为 $(y')^2 \ge 0$。所以 $y^4 \le 1$，即对所有 $x$ 都有 $|y(x)| \le 1$。

**第 (ii) 部分**

用同样方法求导，再把方程写成 $v'' + \sinh v = -xv'$ 的形式代入，

$$\begin{aligned}\frac{\mathrm{dE}}{\mathrm{d}x} &= 2v'v'' + 2\sinh v\, v' \\ &= 2v'\left(v'' + \sinh v\right) \\ &= -2x\,(v')^2.\end{aligned}$$

当 $x \geq 0$ 时，$x$ 与 $(v')^2$ 均非负，所以 $\frac{\mathrm{dE}}{\mathrm{d}x} \leq 0$，正如所求。

因此 E 在 $x \geq 0$ 上非递增（non-increasing），并在该范围内有 $\mathrm{E}(x) \leq \mathrm{E}(0)$。在 $x = 0$ 处，$\cosh(\ln 3) = \frac{1}{2}\left(3 + \frac{1}{3}\right) = \frac{5}{3}$，所以 $\mathrm{E}(0) = 0 + 2 \cdot \frac{5}{3} = \frac{10}{3}$。因此，当 $x \geq 0$ 时，

$$2\cosh v(x) = \mathrm{E}(x) - (v')^2 \leq \mathrm{E}(x) \leq \tfrac{10}{3},$$

故 $\cosh v(x) \leq \frac{5}{3}$。

**第 (iii) 部分**

仿照 (ii)，我们寻找形如 $\mathrm{E} = (w')^2 + F(w)$ 的能量。于是

$$\frac{\mathrm{dE}}{\mathrm{d}x} = 2w'w'' + F'(w)\,w' = w'\left(2w'' + F'(w)\right),$$

我们希望借助方程，使括号内变成 $w'$ 的倍数。只要 $F'(w) = 2(w\cosh w + 2\sinh w)$，就能做到这一点。

**想一想** 哪个 $w$ 的函数的导数是 $w\cosh w + 2\sinh w$?

用分部积分（integration by parts），$\int w\cosh w\,\mathrm{d}w = w\sinh w - \cosh w$，而 $\int 2\sinh w\,\mathrm{d}w = 2\cosh w$。两者相加，

$$\frac{\mathrm{d}}{\mathrm{d}w}\left(w\sinh w + \cosh w\right) = \sinh w + w\cosh w + \sinh w = w\cosh w + 2\sinh w.$$

所以我们取

$$\mathrm{E}(x) = \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2 + 2\left(w\sinh w + \cosh w\right).$$

记 $g(x) = 5\cosh x - 4\sinh x - 3$，方程即为 $w'' + (w\cosh w + 2\sinh w) = -g(x)\,w'$，所以

$$\frac{\mathrm{dE}}{\mathrm{d}x} = 2w'\left(w'' + w\cosh w + 2\sinh w\right) = -2g(x)\,(w')^2.$$

**想一想** $g(x)$ 会不会是负数?

用指数函数表示，

$$\begin{aligned} g(x) &= \tfrac{5}{2}\left(\mathrm{e}^{x}+\mathrm{e}^{-x}\right)-2\left(\mathrm{e}^{x}-\mathrm{e}^{-x}\right)-3 \\ &= \tfrac{1}{2}\mathrm{e}^{x}-3+\tfrac{9}{2}\mathrm{e}^{-x} \\ &= \tfrac{1}{2}\mathrm{e}^{-x}\left(\mathrm{e}^{2x}-6\mathrm{e}^{x}+9\right)=\tfrac{1}{2}\mathrm{e}^{-x}\left(\mathrm{e}^{x}-3\right)^{2} \;\geq\; 0 \end{aligned}$$

对每个 $x$ 都成立（只在 $x=\ln 3$ 处等于零）。因此 $\frac{\mathrm{dE}}{\mathrm{d}x}\leq 0$，并且当 $x\geq 0$ 时 $\mathrm{E}(x)\leq \mathrm{E}(0)$。在 $x=0$ 处，$w=0$ 且 $w'=\frac{1}{\sqrt{2}}$，所以 $\mathrm{E}(0)=\frac{1}{2}+2(0+1)=\frac{5}{2}$。因此，当 $x\geq 0$ 时，

$$2\left(w\sinh w+\cosh w\right)\leq \mathrm{E}(x)\leq \tfrac{5}{2}, \qquad \text{所以} \qquad w\sinh w+\cosh w\leq \tfrac{5}{4}.$$

最后，对每个实数 $w$ 都有 $w\sinh w\geq 0$，因为 $w$ 与 $\sinh w$ 总是同号。略去这一项，便得到当 $x\geq 0$ 时 $\cosh w(x)\leq \frac{5}{4}$，正如所求。

## 验算

- 在 (ii) 中，这个界在起点处是紧的（sharp）：$\cosh v(0)=\cosh(\ln 3)=\frac{5}{3}$。
- 在 (iii) 中，$g(0)=5-0-3=2$，与因式分解形式所得的 $\frac{1}{2}(1-3)^2=2$ 一致。
- 对每条方程求数值解，都印证了上述图像：在 (i) 中 E 一直保持 0.5；在 (ii) 中 E 从 $\frac{10}{3}$ 开始下降；在 (iii) 中 E 从 2.5 开始下降，而 $\cosh w$ 从不超过约 1.03，远在界 1.25 之内。

## 常见错误

- 把 $F(w)$ 对 $x$ 求导时，忘记链式法则带来的因子 $w'$。
- 在 (iii) 中，停在 $w\sinh w+\cosh w\leq\frac{5}{4}$，却没有解释为何 $w\sinh w\geq 0$ 使我们可以略去该项。
- 声称 $g(x)>0$：它在 $x=\ln 3$ 处等于零，所以只有 $\leq$ 成立（而这已足够我们所需）。

## STEP 之外

这些能量函数就是*李雅普诺夫函数*（*Lyapunov functions*）：沿着解从不增加的量，可用来在不求解的情况下证明一个系统保持有界或逐渐趋于稳定。同一个想法——能量守恒加上阻尼造成的耗散——正是非线性振子（nonlinear oscillators）稳定性理论的基础。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）在讲述非线性系统与稳定性的一章中介绍了李雅普诺夫方法。

详细解答

# STEP III 2013，第 8 题

纯数学 · 圆锥曲线的极坐标方程

**STEP III 2013，第 8 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q8

## 关键思路

关系式 $r = ks$（其中 $k < 1$）表明 $E$ 是一个以 $O$ 为焦点（focus）、以 $D$ 为准线（directrix）的椭圆（ellipse）；若采用以 $O$ 为中心的极坐标（polar coordinates），它便化为一条以 $\theta$ 表示 $r$ 的非常简单的方程。一条经过 $O$ 的弦（chord）其实就是两条方向相反的半径，因此 $1/l$ 会化为一个关于 $\cos^2\theta$ 的简单式子。接着，$n$ 条等距分布的直线把总和转化为一些 $\cos 2\theta$ 项之和，而这正是开首那个求和所处理的。

## 解答

### 开首的求和

各项组成一个等比数列（geometric progression），首项为 $\mathrm{e}^{2\mathrm{i}\alpha}$，公比（common ratio）为 $w = \mathrm{e}^{2\pi\mathrm{i}/n}$。由于 $n \geq 2$，角 $2\pi/n$ 位于 $(0, \pi]$ 内，所以 $w \neq 1$，我们可以使用一般的公式：

$$\sum_{r=0}^{n-1} \mathrm{e}^{2\mathrm{i}(\alpha + r\pi/n)} = \mathrm{e}^{2\mathrm{i}\alpha}\,\frac{1-w^n}{1-w} = \mathrm{e}^{2\mathrm{i}\alpha}\,\frac{1-\mathrm{e}^{2\pi\mathrm{i}}}{1-w} = 0.$$

取实部（real part），得 $\sum_{r=0}^{n-1} \cos\left(2\alpha + 2r\pi/n\right) = 0$；我们将会用到的正是这个形式。

### 距离 $s$

如图所示，从 $O$ 到 $D$ 的垂线起量度 $\theta$。该垂线的垂足与 $O$ 的距离为 $d$，而 $P$ 在同一方向上与 $O$ 的距离为 $r\cos\theta$，因此当 $P$ 与 $O$ 位于 $D$ 的同一侧时，

$$s = d - r\cos\theta.$$

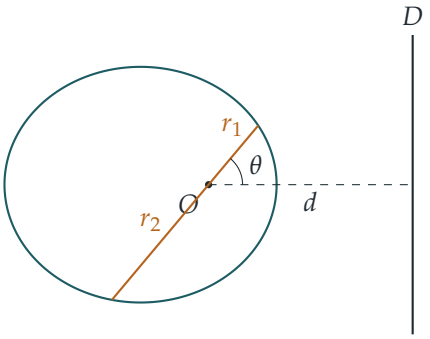
（这对所有 $\theta$ 均成立：当 $\cos\theta < 0$ 时，$P$ 比 $O$ 离 $D$ 更远，而公式亦正确地给出 $s > d$。）

### $E$ 的极坐标方程

$E$ 上每一点都与 $O$ 位于 $D$ 的同一侧。事实上，若 $P$ 位于另一侧，则有 $s = r\cos\theta - d$，而 $r = ks$ 会给出 $r(1 - k\cos\theta) = -kd < 0$；这是不可能的，因为当 $0 < k < 1$ 时 $1 - k\cos\theta > 0$。因此我们可以用 $s = d - r\cos\theta$，于是 $r = ks$ 变为

$$r = k(d - r\cos\theta) \quad \Longrightarrow \quad r = \frac{kd}{1 + k\cos\theta}.$$

由于 $k < 1$，分母对所有 $\theta$ 均为正，所以从 $O$ 出发的每条射线与 $E$ 恰好相交一次。




### 经过 $O$ 的弦的长度

一条经过 $O$、与垂线成角 $\theta$ 的直线，与 $E$ 相交于方向 $\theta$ 及 $\theta + \pi$ 上的两点，而这两点位于 $O$ 的两侧。由于 $\cos(\theta + \pi) = -\cos\theta$，弦长为

$$l = \frac{kd}{1 + k\cos\theta} + \frac{kd}{1 - k\cos\theta} = \frac{2kd}{1 - k^2\cos^2\theta}, \qquad \text{因此} \qquad \frac{1}{l} = \frac{1 - k^2\cos^2\theta}{2kd}.$$

留意这只取决于直线本身，而与我们把它的哪一个方向称为 $\theta$ 无关。

**想一想** 我们应如何把 $\cos^2\theta$ 对 $n$ 条直线求和?

利用倍角公式（double-angle formula）$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$：那些 $\cos 2\theta$ 项正好就是开首求和的实部。

### 对 $n$ 条直线求和

由于相邻直线之间的夹角为 $\pi/n$，对某个固定的 $\alpha$，这 $n$ 条直线所成的角为 $\theta_j = \alpha + j\pi/n$ $(j = 1, \ldots, n)$。这 $n$ 个 $2\theta_j$ 的值是 $2\alpha + 2j\pi/n$；把 $j = n$ 换成 $j = 0$（这只会令 $2\theta_j$ 改变 $2\pi$）后，它们正是开首求和中的各个角。因此

$$\sum_{j=1}^{n} \cos^2\theta_j = \frac{n}{2} + \frac{1}{2}\sum_{j=1}^{n} \cos 2\theta_j = \frac{n}{2} + 0 = \frac{n}{2},$$

从而

$$\sum_{j=1}^{n}\frac{1}{l_j}=\frac{1}{2kd}\left(n-k^2\cdot\frac{n}{2}\right)=\frac{(2-k^2)n}{4kd},$$

这就是所求的结果。条件 $n\geq 2$ 只用过一次，就是用来确保开首求和中 $w\neq 1$。

## 验算

- 取 $n=2$、$k=\frac{1}{2}$、$d=1$，两条直线分别沿着 $OD$ 及垂直于 $OD$。沿 $OD$ 的弦是长轴（major axis），长度为 $\frac{1/2}{3/2}+\frac{1/2}{1/2}=\frac{4}{3}$；垂直于它的弦（即正焦弦（latus rectum））长度为 $2kd=1$。所以总和为 $\frac{3}{4}+1=\frac{7}{4}$，而公式给出 $\frac{(7/4)\cdot 2}{2}=\frac{7}{4}$。
- 当 $n=1$ 时，结果并不成立，这是理所当然的：单一条直线的 $1/l$ 取决于它的方向。而 $n=1$ 正是 $w=1$ 的情况。
- 我们用电脑对随机的 $d$、$k$、$n$ 及 $\alpha$ 直接从 $r=ks$ 求出每条弦作验算，结果与公式吻合；例如 $d=1$、$k=\frac{1}{2}$、$n=3$ 时，对任何 $\alpha$ 都得到 $2.625=\frac{21}{8}$。

## 常见错误

- 写成 $s=d+r\cos\theta$：在图中 $\theta$ 是朝 $D$ 的方向量度的，所以沿 $OP$ 向外移动会令你更接近 $D$。
- 把弦的另一半再次当作 $kd/(1+k\cos\theta)$，忘记了相反方向的角是 $\theta+\pi$。
- 直接把角 $\alpha+j\pi/n$ 求和，而不是对 $2\theta_j$ 求和：均匀地绕圆周一圈的是倍角，这正是开首求和中出现因子 2 的原因。

## STEP 之外

焦点－准线定义使每一条圆锥曲线（conic）都有极坐标方程 $r=\ell/(1+e\cos\theta)$，其中 $e<1$、$e=1$ 及 $e>1$ 分别对应椭圆、抛物线（parabola）及双曲线（hyperbola）；当我们解牛顿（Newton）的行星运动方程时，开普勒（Kepler）的轨道正是以这个形式出现。任意两条互相垂直的焦点弦（focal chord）的 $1/l_1+1/l_2$ 都相同，这是一个经典性质，而本题的结果正是它的推广。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）便是用这种方法推导轨道方程的。

详细解答

# STEP III 2013，第 9 题

力学 · 浮力与简谐运动

**STEP III 2013，第 9 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q9

## 关键思路

球体浸没在液体中的部分是一个球冠（spherical cap），其体积可以把它切成一片片圆盘求得，这就是旋转体体积（volume of revolution）的计算。之后，题目就是应用牛顿第二定律（Newton's second law），除了重力之外，唯一的力就是浮力（buoyancy）。对于小幅振动，我们设 $x = x_0 + y$，把右边的三次式在平衡位置附近线性化（linearise），再从形如 $\ddot{y} = -\omega^2 y$ 的方程读出 $\omega^2$。

## 解答

### 排开的体积

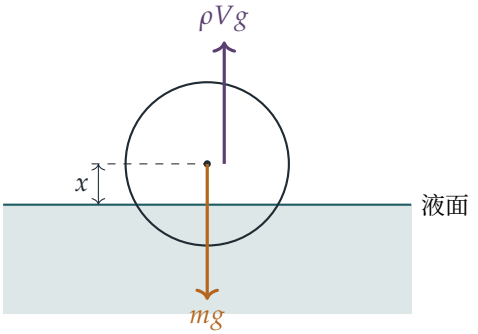
取球心为原点，$z$ 向上量度。液面就是平面 $z = -x$，浸没部分就是球体中 $-R \le z \le -x$ 的部分。这对 $-R < x < R$ 有意义，因此涵盖了题目所述的情况。在高度 $z$ 处，截面是半径为 $\sqrt{R^2 - z^2}$ 的圆盘，所以

$$\begin{aligned} V &= \int_{-R}^{-x} \pi \left(R^2 - z^2\right) \mathrm{d}z = \pi \left[R^2 z - \tfrac{1}{3} z^3\right]_{-R}^{-x} \\ &= \pi \left(-R^2 x + \tfrac{1}{3} x^3 + R^3 - \tfrac{1}{3} R^3\right) = \frac{\pi}{3} \left(2R^3 - 3R^2 x + x^3\right), \end{aligned}$$

这就是所求的体积。

### 运动方程

把球体视为作铅直运动的刚体（rigid body）。由于水缸很大，球体上下浮动时液面不会移动，所以 $\ddot{x}$ 就是球体向上的加速度。球体的质量为 $m = \frac{4}{3}\pi R^3 \rho_s$。作用在球体上的力有两个：向下的重量 $mg$，以及向上的浮力 $\rho V g$。




沿铅直方向分解，以向上为正，

$$m\ddot{x} = \rho Vg - mg \quad \Longrightarrow \quad m\left(g + \ddot{x}\right) = \rho gV.$$

代入 $m$ 和 $V$：

$$\tfrac{4}{3}\pi R^3 \rho_{\mathrm{s}}\left(g + \ddot{x}\right) = \tfrac{\pi}{3}\left(2R^3 - 3R^2 x + x^3\right)\rho g,$$

两边乘以 $3/\pi$，得

$$4R^3 \rho_{\mathrm{s}}\left(g + \ddot{x}\right) = \left(2R^3 - 3R^2 x + x^3\right)\rho g,$$

此即所求。

### 平衡时的密度

在平衡（equilibrium）时 $\ddot{x} = 0$。代入 $x = \frac{1}{2}R$：

$$4R^3 \rho_{\mathrm{s}} g = \left(2R^3 - \tfrac{3}{2}R^3 + \tfrac{1}{8}R^3\right)\rho g = \tfrac{5}{8}R^3 \rho g,$$

所以

$$\rho_{\mathrm{s}} = \frac{5}{32}\rho.$$

这个值小于 $\frac{1}{2}\rho$，理应如此：球心在液面之上，所以浸没在液体中的部分少于球体的一半。

### 小幅振动

设 $f(x) = 2R^3 - 3R^2 x + x^3$，并设 $x = \frac{1}{2}R + y$，其中 $y$ 很小。于是 $\ddot{x} = \ddot{y}$。

> ? **想一想** 怎样把这个三次式变成关于 $y$ 的线性式?

我们把 $f$ 在平衡位置附近展开，只保留一阶项：$f(\frac{1}{2}R + y) = f(\frac{1}{2}R) + f'(\frac{1}{2}R)\,y + O(y^2)$。这里

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

而由前面的计算，$f(\frac{1}{2}R) = \frac{5}{8}R^3$。利用 $\rho_s = \frac{5}{32}\rho$，运动方程的左边为 $4R^3 \cdot \frac{5}{32}\rho\,(g+\ddot{y}) = \frac{5}{8}R^3\rho\,(g+\ddot{y})$。因此，准确至一阶，

$$\tfrac{5}{8}R^3\rho\,(g+\ddot{y}) = \left(\tfrac{5}{8}R^3 - \tfrac{9}{4}R^2 y\right)\rho g.$$

两边的 $\frac{5}{8}R^3\rho g$ 项互相抵消（这正是平衡条件），剩下

$$\tfrac{5}{8}R^3\ddot{y} = -\tfrac{9}{4}R^2 g\, y \quad\Longrightarrow\quad \ddot{y} = -\frac{18g}{5R}\,y.$$

这是简谐运动（simple harmonic motion），其中 $\omega^2 = \dfrac{18g}{5R}$，所以小幅振动的周期（period）为

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{5R}{18g}}.$$

## 验算

- 体积：当 $x = R$ 时，公式给出 0（球体刚好接触液面）；当 $x = 0$ 时给出 $\frac{2}{3}\pi R^3$（半个球体）；当 $x = -R$ 时给出 $\frac{4}{3}\pi R^3$（完全浸没）。
- 量纲（dimensions）：$\omega^2 = 18g/(5R)$ 的单位是 $\mathrm{s}^{-2}$，理应如此。
- 数值：取 $R = 1\,\mathrm{m}$ 及 $g = 9.81\,\mathrm{m\,s^{-2}}$，周期约为 $1.06\,\mathrm{s}$。用电脑从一个小位移出发，逐步推进完整的非线性方程，得到的周期与此相同，误差小于万分之一（$10^4$ 分之一）。

## 常见错误

- 把浸没球冠的高度当作 $x$ 而不是 $R - x$：$x$ 是球心在液面之上的高度。
- 忘记回复力来自 $f'(\frac{1}{2}R)$，而它是负数。这里出现正负号错误会得到 $\ddot{y} = +\omega^2 y$，那根本不是振动。
- 漏掉左边球体本身的质量：与 $\ddot{x}$ 相乘的是 $\rho_s$，而不是 $\rho$。（现实中的浮体还会带动周围部分液体一起运动，即所谓“附加质量”（added mass），但题目叫我们忽略这一点。）

## STEP 之外

这里所用的方法——把力在平衡位置附近展开并保留线性项——是求任何系统小幅振动的标准方法。线性项的正负号决定稳定性，其大小则决定频率。在大学一年级的课程中，这会发展为 $\ddot{x} = F(x)$ 的线性化，然后在有多个自由度（degrees of freedom）时，进一步研究简正模态（normal modes）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）以这个程度讲解了平衡位置附近的小幅振动及流体浮力。

详细解答

# STEP III 2013，第 10 题

力学 · 转动撞击

**STEP III 2013，第 10 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

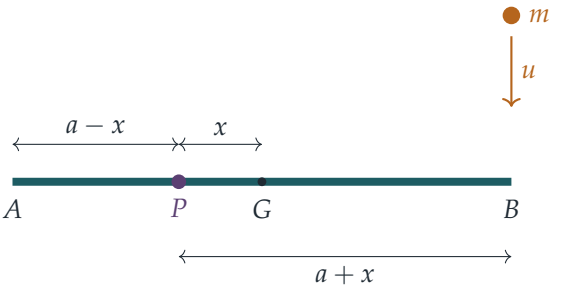
step.maths.org/questions/13-s3-q10

## 关键思路

只要留意到杆的中点与 $P$ 相距多远，转动惯量（moment of inertia）便可直接由平行轴定理（parallel axis theorem）得出。至于撞击，转轴会对杆施加一个大小未知的冲量（impulse），因此线动量（linear momentum）并不守恒；但该冲量作用于 $P$，所以它对 $P$ 的力矩为零。对 $P$ 的角动量守恒（conservation of angular momentum），再加上在 $B$ 处应用牛顿（Newton）实验定律（Newton's experimental law），便得到两条方程，足以求出两个未知量。

## 解答

我们把杆视为均匀刚性杆，把质点视为质点，并假设撞击在瞬间完成，因此唯一的水平冲量来自 $B$ 处和转轴处（重力等有限大小的力在撞击的一瞬间不产生冲量）。



### 转动惯量

设 $G$ 为杆的中点，即其质心（centre of mass）。由于 $AG = a$ 及 $AP = a - x$，距离 $PG$ 为 $x$（当 $x > 0$ 时，$G$ 位于 $P$ 靠 $B$ 的一侧）。均匀杆对通过其中心的垂直轴的转动惯量为 $\frac{1}{3}Ma^2$（杆长为 $2a$）。由平行轴定理，

$$I_P = \tfrac{1}{3}Ma^2 + Mx^2 = \tfrac{1}{3}M(a^2 + 3x^2),$$

与所求相符。

如果你喜欢从基本原理出发：线密度（line density）为 $M/(2a)$，以 $s$ 表示沿杆从 $P$ 量起的距离，则杆由 $s = -(a - x)$ 延伸至 $s = a + x$，而

$$I_P = \frac{M}{2a}\int_{-(a-x)}^{a+x} s^2\,\mathrm{d}s = \frac{M}{6a}\left[(a+x)^3 + (a-x)^3\right] = \frac{M}{6a}(2a^3 + 6ax^2),$$

结果相同。

### 撞击

由 $P$ 到 $B$ 的距离为 $2a - (a - x) = a + x$。设 $v$ 为质点在撞击后瞬间的速度（以原来的运动方向为正），$\omega$ 为杆的角速度（angular velocity），于是 $B$ 此时以速率 $(a + x)\omega$ 沿同一方向运动。

> **想一想** 为甚么这里不能用线动量守恒?

因为转轴是固定的，它在碰撞期间对杆施加一个大小未知的冲量。该冲量作用于 $P$，所以它对 $P$ 的力矩为零。质点与杆之间的冲量是杆与质点所组成系统的内部冲量。因此，系统对 $P$ 的总角动量守恒：

$$mu(a + x) = mv(a + x) + I_P\,\omega. \tag{1}$$

在 $B$ 处应用牛顿实验定律（分离速率等于 $e$ 乘以接近速率），得

$$(a + x)\omega - v = eu. \tag{2}$$

由 (2)，$v = (a + x)\omega - eu$。代入 (1)，

$$\begin{aligned} mu(a + x) &= m(a + x)\left[(a + x)\omega - eu\right] + I_P\,\omega \\ mu(1 + e)(a + x) &= \left[m(a + x)^2 + \tfrac{1}{3}M(a^2 + 3x^2)\right]\omega. \end{aligned}$$

两边乘以 3，

$$\omega = \frac{3mu(1 + e)(a + x)}{M(a^2 + 3x^2) + 3m(a + x)^2},$$

即题目所给的结果。

**$m = 2M$ 的情况**

代入 $m = 2M$ 并约去 $M$：

$$\omega = \frac{6u(1+e)(a+x)}{a^2+3x^2+6(a+x)^2}.$$

自然出现的变量是力臂（lever arm）$s = a + x$，即距离 $PB$，它的取值范围是 $[0, 2a]$。以 $x = s - a$ 代入，分母为

$$a^2 + 3(s-a)^2 + 6s^2 = 9s^2 - 6as + 4a^2,$$

所以

$$\omega = \frac{6u(1+e)\,s}{9s^2 - 6as + 4a^2}.$$

当 $s = 0$（转轴位于 $B$）时 $\omega = 0$，所以最大值出现在 $s > 0$ 处，此时可把分子分母同除以 $s$：

$$\omega = \frac{6u(1+e)}{9s + \dfrac{4a^2}{s} - 6a}.$$

我们要令分母尽可能小。由于 $s > 0$，算术—几何平均不等式（AM–GM inequality）给出

$$9s + \frac{4a^2}{s} \;\geq\; 2\sqrt{9s \cdot \frac{4a^2}{s}} = 12a,$$

等号成立当且仅当 $9s = 4a^2/s$，即 $s = \frac{2}{3}a$，而这确实落在 $(0, 2a]$ 之内。因此分母至少为 $12a - 6a = 6a$，且只在 $s = \frac{2}{3}a$ 时取等号，故

$$\omega \leq \frac{6u(1+e)}{6a} = \frac{u(1+e)}{a}.$$

所以最大角速度为 $u(1+e)/a$，在 $a + x = \frac{2}{3}a$ 时达到，即

$$x = -\tfrac{1}{3}a.$$

负号只是表示 $P$ 位于中点靠 $B$ 的一侧：$P$ 与 $A$ 相距 $\frac{4}{3}a$，与 $B$ 相距 $\frac{2}{3}a$。

## 验算

- 当 $x = a$（转轴位于 $A$）且 $e = 1$ 时，公式给出 $\omega = 3mu/\big(a(M+3m)\big)$；令 $M \to 0$，得 $\omega = u/a$，于是 $B$ 以 $2u$ 运动，这正是极轻的“靶子”受弹性撞击时大家熟悉的结果。
- 在最佳情况下取 $a = 1$、$u = 1$、$e = \frac{1}{2}$，杆以 $\omega = 1.5$ 转动，所以 $B$ 以 1 运动，而质点在撞击后的速度为 $v = 1 - \frac{1}{2} = \frac{1}{2}$。此时对 $P$ 的角动量为 $2M \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{3}M\big(1 + \frac{1}{3}\big) \cdot 1.5 = \frac{4}{3}M$，与撞击前的 $2M \cdot 1 \cdot \frac{2}{3}$ 相符。
- 以电脑在 $x \in [-a, a]$ 上作细密搜索，确认最大值出现在 $x = -a/3$。

## 常见错误

- 使用线动量守恒：固定的转轴会产生冲量，所以只有对 $P$ 的角动量守恒。
- 把距离 $PB$ 当作 $a - x$ 或 $2a - x$，而不是 $a + x$。
- 因 $x = -\frac{1}{3}a$ 是负数而舍弃它：$x$ 是带正负号的位置，任何 $[-a, a]$ 内的 $x$ 都会使 $P$ 落在杆上。

## STEP 之外

选择适当的点来取力矩，使未知的反作用力消去，是刚体动力学（rigid-body dynamics）中最有用的技巧之一；它在“撞击中心”（centre of percussion）中再次出现——那是球棒或杆上的一点，打击该点时，支点处不会产生冲量。大学一年级的力学会把这发展成角动量和惯性张量（inertia tensor）的一般理论。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对冲击性转动和撞击中心有非常清晰的论述。

详细解答

# STEP III 2013，第 11 题

力学 · 弹簧与能量

> **STEP III 2013，第 11 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s3-q11

## 关键思路

一切都由对称性（symmetry）决定：三角形保持水平，其中心一直位于悬挂点的正下方，而且只能上下移动。因此只需考虑一个质点和一条弹簧，便能掌握整个情况。经过一条弹簧的铅垂平面上的几何关系，给出弹簧长度关于 $\theta$ 的表达式。沿铅垂方向分解力可得平衡（equilibrium）条件，而能量守恒（conservation of energy）（弹性势能、重力势能和动能）则给出速率。

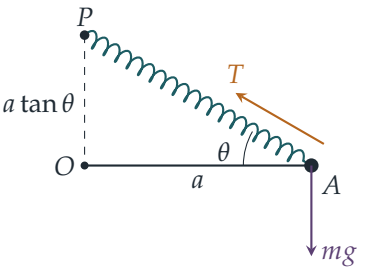
## 解答

### 模型与几何

把各质量视为质点（particle），各杆视为轻而刚性，各弹簧视为轻弹簧并遵守胡克定律（Hooke's law）。设 $P$ 为悬挂点，$O$ 为三角形的中心，位于 $P$ 的正下方。边长为 $s$ 的等边三角形，其中心到顶点的距离为 $s/\sqrt{3}$，所以每个顶点与 $O$ 的水平距离为

$$OA = \frac{\sqrt{3}a}{\sqrt{3}} = a$$

取一个顶点 $A$，考虑经过 $P$、$O$ 和 $A$ 的铅垂平面：




弹簧 $PA$ 与水平线 $OA$ 成角 $\theta$，所以弹簧长度为 $a/\cos\theta$，而 $A$ 位于 $P$ 下方 $a\tan\theta$ 的深度。

### 张力

每条弹簧的伸长量（extension）为

$$x = \frac{a}{\cos\theta} - a = \frac{a(1-\cos\theta)}{\cos\theta},$$

因此由胡克定律，配合弹性模量（modulus）$kmg$ 和自然长度（natural length）$a$，得

$$T = \frac{kmg}{a}\,x = \frac{kmg(1-\cos\theta)}{\cos\theta},$$

即所求结果。

### $\theta = \frac{1}{6}\pi$ 时的平衡

作用于 $A$ 处质点的力包括其重量、沿 $AP$ 方向的弹簧张力（tension），以及两条杆施加的力。各杆位于水平面内，所以杆的作用力是水平的。

> **想一想** 为甚么可以不理会杆的作用力?

我们沿铅垂方向分解，而杆的作用力没有铅垂分量。（在水平方向上，张力分量 $T\cos\theta$ 由杆的作用力平衡，这正是需要杆的原因，但我们用不着这条方程。）在 $A$ 处沿铅垂方向分解:

$$T\sin\theta = mg \quad\Longrightarrow\quad k(1-\cos\theta)\tan\theta = 1.$$

当 $\theta = \frac{1}{6}\pi$ 时，$\cos\theta = \frac{\sqrt{3}}{2}$，$\tan\theta = \frac{1}{\sqrt{3}}$，所以

$$k\cdot\frac{2-\sqrt{3}}{2}\cdot\frac{1}{\sqrt{3}} = 1 \quad\Longrightarrow\quad k = \frac{2\sqrt{3}}{2-\sqrt{3}} = \frac{2\sqrt{3}(2+\sqrt{3})}{(2-\sqrt{3})(2+\sqrt{3})} = 4\sqrt{3}+6,$$

因为 $(2-\sqrt{3})(2+\sqrt{3}) = 1$。这就是所求的值。

**在平衡位置时的速率**

由对称性，三角形保持水平，中心位于 $P$ 的正下方，所以三个质点都以相同速率作铅垂运动。除重力和弹簧外，没有其他外力做功，所以能量守恒。伸长量为 $x$ 的弹簧所储存的弹性势能为 $kmg\,x^2/(2a)$。

起始位置（$\theta=\frac{1}{3}\pi$，静止）。此时 $\cos\theta=\frac{1}{2}$，所以伸长量为 $x_1=a$，而在 $P$ 下方的深度为 $a\tan\frac{1}{3}\pi=\sqrt{3}a$。

平衡位置（$\theta=\frac{1}{6}\pi$）。伸长量为

$$x_2=a\left(\frac{2}{\sqrt{3}}-1\right)=\frac{a(2-\sqrt{3})}{\sqrt{3}},\qquad x_2^2=\frac{a^2(7-4\sqrt{3})}{3},$$

深度为 $a\tan\frac{1}{6}\pi=\frac{a}{\sqrt{3}}$。所以每个质点上升了

$$\sqrt{3}a-\frac{a}{\sqrt{3}}=\frac{2a}{\sqrt{3}}.$$

三条弹簧释放的弹性势能：

$$\begin{aligned}3\cdot\frac{kmg}{2a}\left(x_1^2-x_2^2\right)&=\frac{3kmg}{2a}\cdot\frac{a^2\left(3-7+4\sqrt{3}\right)}{3}\\&=2kmga(\sqrt{3}-1).\end{aligned}$$

由于 $k=4\sqrt{3}+6$，$k(\sqrt{3}-1)=12-4\sqrt{3}+6\sqrt{3}-6=6+2\sqrt{3}$，所以释放的弹性势能为 $4mga(3+\sqrt{3})$。

三个质点增加的重力势能：$3mg\cdot\dfrac{2a}{\sqrt{3}}=2\sqrt{3}\,mga$。

两者之差转化为三个质点的动能，每个质点的速率均为 $V$：

$$\begin{aligned}\tfrac{3}{2}mV^2&=4mga(3+\sqrt{3})-2\sqrt{3}\,mga=mga\left(12+2\sqrt{3}\right),\\V^2&=\frac{2ag}{3}\left(12+2\sqrt{3}\right)=\frac{4ag}{3}\left(6+\sqrt{3}\right),\end{aligned}$$

即所求结果。

## 验算

- 正负号：在 $\theta=\frac{1}{3}\pi$ 时，每条弹簧的铅垂拉力为 $T\sin\theta=kmg\cdot\frac{\sqrt{3}}{2}\approx 11.2\,mg>mg$，所以三角形起初确实是向上朝平衡位置移动，与能量方程的假设一致。
- 答案为正数，而且具有速率平方的单位（$ag$）。
- 取 $a=1.3\,\mathrm{m}$ 及 $g=9.81\,\mathrm{m\,s^{-2}}$，由 $\theta=\frac{1}{3}\pi$ 的静止状态出发，以数值方法逐步推进运动方程 $\ddot{y}=g-\frac{T}{m}\sin\theta$，在 $\theta=\frac{1}{6}\pi$ 时得 $V^2\approx 131.48\,\mathrm{m^2\,s^{-2}}$，与 $\frac{4ag}{3}(6+\sqrt{3})$ 吻合。

## 常见错误

- 把中心到顶点的距离误当作 $\sqrt{3}a$ 或 $\frac{\sqrt{3}}{2}a$，而不是 $a$。
- 忘记共有三条弹簧和三个质点：因数 3 同时出现在弹性势能、重力势能和动能各项中。
- 弄错运动方向：质点在两个位置之间是上升的，所以重力势能是增加，而不是减少。

## STEP 之外

在这里，三角形只有一个自由度（degree of freedom），因为对称性排除了所有其他运动。选取单一坐标（这里是 $\theta$，或深度 $y$），并以它表示能量，正是拉格朗日力学（Lagrangian mechanics）的起点。把势能在平衡位置附近展开，还可以求得微小铅垂振动的周期。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）便是本着这种精神，发展能量方法和小振动（small oscillations）的理论。

详细解答

# STEP III 2013，第 12 题

概率与统计 · 指示随机变量

> **STEP III 2013，第 12 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/13-s3-q12

## 关键思路

每个 $X_k$ 恰好在一段新的 $A$ 从位置 $k$ 开始时等于 1，所以 $S$ 数的是这一行中 $A$ 的段数（连串，runs）。这些变量都是指示变量（indicator variables），所以每个期望值（expectation）都只是若干指定位置放着指定字母的概率，而这个概率只取决于涉及多少个位置，与它们在哪里无关。至于方差（variance），我们把 $S^2$ 展开，并把乘积 $X_iX_j$ 分类：下标相等、下标相邻（结果是不可能发生的）以及其余情况。

## 解答

全文中，所有 $\binom{n}{a}$ 种排列的可能性均等。由此得到一个有用的结论：对任意 $m$ 个不同的位置，它们放着某个指定图样（其中含 $p$ 个 $A$ 和 $q$ 个 $B$，$p+q=m$）的概率，可以逐个位置填入来求得，例如

$$\Pr(\text{位置}r,s\text{ 分别为}B,A)=\frac{b}{n}\cdot\frac{a}{n-1},$$

不论不同的位置 $r$ 和 $s$ 是哪两个。由于每个 $X_k$ 只取值 0 和 1，所以 $\mathrm{E}(X_k)=\Pr(X_k=1)$；同理 $\mathrm{E}(X_iX_j)=\Pr(X_i=1\text{ 且}X_j=1)$。

### 第 (i) 部分

当 $i=1$ 时，我们需要第一个字母是 $A$，所以

$$\mathrm{E}(X_1)=\frac{a}{n}.$$

当 $2\le i\le n$ 时，我们需要位置 $i-1$ 是 $B$，而位置 $i$ 是 $A$，所以

$$\mathrm{E}(X_i)=\frac{b}{n}\cdot\frac{a}{n-1}=\frac{ab}{n(n-1)}.$$

根据期望值的线性性质（linearity of expectation），对 $n-1$ 个下标 $i \geq 2$ 求和，

$$\mathrm{E}(S) = \frac{a}{n} + (n-1)\cdot\frac{ab}{n(n-1)} = \frac{a}{n} + \frac{ab}{n} = \frac{a(b+1)}{n},$$

即为所求。

**第 (ii) 部分**

**(a)** 设 $j \geq 3$。那么 $X_1X_j = 1$ 恰好在位置 1 是 $A$、位置 $j-1$ 是 $B$ 且位置 $j$ 是 $A$ 时成立。由于 $j-1 \geq 2$，这是三个不同的位置，放着两个 $A$ 和一个 $B$。逐个填入，

$$\mathrm{E}(X_1X_j) = \frac{a}{n}\cdot\frac{b}{n-1}\cdot\frac{a-1}{n-2} = \frac{a(a-1)b}{n(n-1)(n-2)}.$$

**(b)** 现在取 $2 \leq i$ 及 $j \geq i+2$。那么 $X_iX_j = 1$ 恰好在位置 $i-1$, $i$, $j-1$, $j$ 分别是 $B, A, B, A$ 时成立。由于 $j-1 \geq i+1 > i$，这四个位置互不相同，放着两个 $A$ 和两个 $B$，所以

$$\mathrm{E}(X_iX_j) = \frac{b}{n}\cdot\frac{a}{n-1}\cdot\frac{b-1}{n-2}\cdot\frac{a-1}{n-3} = \frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)}.$$

（这里因为 $a, b \geq 2$，所以 $n \geq 4$，分母不会为零。）这个值对双重和中的每一对都相同，所以我们只需数出有多少对。对给定的 $i$，$j$ 由 $i+2$ 取到 $n$，共 $n-i-1$ 个值；当 $i$ 由 2 取到 $n-2$ 时，这个数依次取值 $n-3$, $n-4$, …, 1。所以对数为

$$1 + 2 + \cdots + (n-3) = \tfrac{1}{2}(n-3)(n-2),$$

而

$$\sum_{i=2}^{n-2}\left(\sum_{j=i+2}^{n}\mathrm{E}(X_iX_j)\right) = \frac{(n-2)(n-3)}{2}\cdot\frac{a(a-1)b(b-1)}{n(n-1)(n-2)(n-3)} = \frac{a(a-1)b(b-1)}{2n(n-1)}.$$

**(c)** 我们利用 $\mathrm{Var}(S) = \mathrm{E}(S^2) - \big(\mathrm{E}(S)\big)^2$，并展开

$$S^2 = \sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_iX_j.$$

由于 $X_i$ 取值 0 或 1，所以 $X_i^2 = X_i$，因此第一个和的期望值是 $\mathrm{E}(S)$。

**想一想** 那么 (b) 部分略去的 $j = i+1$ 的那些对又如何？

它们没有任何贡献：$X_iX_{i+1} = 0$ 恒成立。当 $i = 1$ 时，$X_1 = 1$ 要求位置 1 是 $A$，但 $X_2 = 1$ 要求该位置是 $B$。当 $i \geq 2$ 时，$X_i = 1$ 要求位置 $i$ 是 $A$，但 $X_{i+1} = 1$ 要求该位置是 $B$。（两段 $A$ 不可能在相邻的位置开始。）

所以真正有关的 $i<j$ 对，就是那些 $j\geq i+2$ 的。其中 $i=1$ 的有 $j=3,\ldots,n$，共 $n-2$ 对，每一对都由 (a) 部分给出；$i\geq 2$ 的恰好就是 (b) 部分的双重和。因此

$$\begin{aligned}\sum_{i<j}\mathrm{E}(X_iX_j) &= (n-2)\cdot\frac{a(a-1)b}{n(n-1)(n-2)}+\frac{a(a-1)b(b-1)}{2n(n-1)}\\ &= \frac{a(a-1)b}{n(n-1)}\left(1+\frac{b-1}{2}\right)=\frac{a(a-1)b(b+1)}{2n(n-1)},\end{aligned}$$

于是

$$\mathrm{E}(S^2)=\frac{a(b+1)}{n}+\frac{a(a-1)b(b+1)}{n(n-1)}.$$

减去 $\big(\mathrm{E}(S)\big)^2=a^2(b+1)^2/n^2$，并抽出公因式 $\dfrac{a(b+1)}{n^2(n-1)}$，

$$\mathrm{Var}(S)=\frac{a(b+1)}{n^2(n-1)}\Big[n(n-1)+n(a-1)b-a(b+1)(n-1)\Big].$$

展开括号，$nab$ 与 $-nab$ 两项互相抵消，剩下

$$\begin{aligned}n^2-n-nb-na+ab+a &= n(n-a-b)-n+ab+a\\ &= -n+ab+a \qquad \text{因为}n=a+b\\ &= ab-b=b(a-1).\end{aligned}$$

所以

$$\mathrm{Var}(S)=\frac{a(b+1)}{n^2(n-1)}\cdot b(a-1)=\frac{a(a-1)b(b+1)}{n^2(n-1)},$$

即为所求。

## 验算

取 $a=3$、$b=2$，于是 $n=5$，共有 10 种排列。数出 $A$ 的段数：排列 $AAABB$、$BAAAB$、$BBAAA$ 给出 $S=1$；$ABABA$ 给出 $S=3$；其余六种给出 $S=2$。所以 $\mathrm{E}(S)=\frac{18}{10}=\frac{9}{5}$，$\mathrm{E}(S^2)=\frac{36}{10}$，从而 $\mathrm{Var}(S)=\frac{18}{5}-\frac{81}{25}=\frac{9}{25}$。公式给出 $\frac{3\cdot3}{5}=\frac{9}{5}$ 及 $\frac{3\cdot2\cdot2\cdot3}{25\cdot4}=\frac{9}{25}$。以电脑穷举验证，对所有 $2\leq a,b\leq 6$，题中每条公式都完全正确。

这些公式在题目所述范围以外也表现合理。当 $a=2$、$b=1$ 时，排列 $AAB$、$ABA$、$BAA$ 给出 $S=1,2,1$，所以 $\mathrm{E}(S)=\frac{4}{3}$，$\mathrm{Var}(S)=2-\frac{16}{9}=\frac{2}{9}$，与 $\frac{2\cdot2}{3}$ 及 $\frac{2\cdot1\cdot1\cdot2}{9\cdot2}$ 吻合。而若 $a=1$，方差公式给出 0，这理应如此，因为此时永远恰好只有一段。

## 常见错误

- 把 $X_1$ 当作与其他变量一样处理：并没有位置 0，所以 $X_1$ 只涉及一个字母，$\mathrm{E}(X_1)=a/n$，而不是 $ab/(n(n-1))$。

- 忘记 $X_i$ 与 $X_{i+1}$ 不可能同时为 1，或用一条非零的公式去计算 $\mathrm{E}(X_iX_{i+1})$。
- 数错 (b) 部分的对数：应有 $\frac{1}{2}(n-2)(n-3)$ 对，而不是 $\frac{1}{2}n(n-1)$ 对。

## STEP 之外

把一个计数写成指示变量之和，再用线性性质求平均值，并由两两乘积求方差，是大学概率论中最有用的技巧之一；分析连串、配对、随机排列（random permutations）的不动点（fixed points）以及随机图（random graphs），都是用这个方法。这里的 $S$ 本质上就是 Wald–Wolfowitz 连串检定（runs test）的统计量，用来检测一个序列是否并非随机。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）载有许多运用指示变量方法的例子。

详细解答

# STEP III 2013，第 13 题

概率与统计 · 矩与分布函数

**STEP III 2013，第 13 题**

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/13-s3-q13

## 关键思路

由于 $\mathrm{F}' = \mathrm{f}$，表达式 $2\mathrm{F}(x)\mathrm{f}(x)$ 正是 $\big(\mathrm{F}(x)\big)^2$ 的导数（derivative）。因此第 (i)(b) 部分只需一次分部积分（integration by parts）；而只要适当选取 g，关于 $Y$ 的每个问题都能化为关于 $X$ 的问题。至于各个界（bound），则来自把 f 或 F 换成第 (i)(a) 部分中较大的量 $M$ 或 $Mx$，而每次替换都朝着令积分变大的方向进行。

## 解答

全文中，$\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t$，故 $\mathrm{F}'(x) = \mathrm{f}(x)$；又由于 $X$ 取值于 $[0,1]$，有 $\mathrm{F}(0) = 0$ 及 $\mathrm{F}(1) = 1$。

### 第 (i) 部分

**(a)** 对 $0 \le x \le 1$，由于对每个 $t$ 都有 $0 \le \mathrm{f}(t) \le M$，

$$0 \le \int_0^x \mathrm{f}(t)\,\mathrm{d}t \le \int_0^x M\,\mathrm{d}t = Mx,$$

即 $0 \le \mathrm{F}(x) \le Mx$。

**(b)** 由链式法则(chain rule)，$\dfrac{\mathrm{d}}{\mathrm{d}x}\big(\mathrm{F}(x)\big)^2 = 2\mathrm{F}(x)\mathrm{f}(x)$。取 $u = \mathrm{g}(x)$ 及 $\mathrm{d}v = 2\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x$ 进行分部积分，于是 $v = \big(\mathrm{F}(x)\big)^2$，

$$\begin{aligned}\int_0^1 2\mathrm{g}(x)\mathrm{F}(x)\mathrm{f}(x)\,\mathrm{d}x &= \Big[\mathrm{g}(x)\big(\mathrm{F}(x)\big)^2\Big]_0^1 - \int_0^1 \mathrm{g}'(x)\big(\mathrm{F}(x)\big)^2\,\mathrm{d}x \\ &= \mathrm{g}(1) - \int_0^1 \mathrm{g}'(x)\big(\mathrm{F}(x)\big)^2\,\mathrm{d}x,\end{aligned}$$

因为 $\mathrm{F}(1) = 1$ 且 $\mathrm{F}(0) = 0$。（这里需要 g 可微（differentiable），题目已隐含此假设。）

### 第 (ii) 部分

**(a)** $Y$ 的概率密度函数（probability density function）的积分必须等于 1。在 (i)(b) 中取 $\mathrm{g}(x) = 1$，使 $\mathrm{g}' = 0$，得

$$\int_0^1 2\mathrm{F}(y)\mathrm{f}(y)\,\mathrm{d}y = 1,$$

故 $k = 2$。（亦可直接计算：$\int_0^1 k\mathrm{F}\mathrm{f}\,\mathrm{d}y = \frac{k}{2}\big[\mathrm{F}^2\big]_0^1 = \frac{k}{2}$。）另请注意 $k\mathrm{F}\mathrm{f} \geq 0$，所以这确实是一个概率密度函数。

**(b)** 我们有

$$\mathrm{E}(Y^n) = \int_0^1 2y^n\,\mathrm{F}(y)\mathrm{f}(y)\,\mathrm{d}y.$$

上界。由 (i)(a)，$\mathrm{F}(y) \leq My$，而其余因子 $2y^n\mathrm{f}(y)$ 均非负，所以

$$\mathrm{E}(Y^n) \leq \int_0^1 2y^n \cdot My \cdot \mathrm{f}(y)\,\mathrm{d}y = 2M\int_0^1 y^{n+1}\mathrm{f}(y)\,\mathrm{d}y = 2M\mu_{n+1}.$$

下界。现在在 (i)(b) 中取 $\mathrm{g}(y) = y^n$，于是 $\mathrm{g}(1) = 1$ 及 $\mathrm{g}'(y) = ny^{n-1}$：

$$\mathrm{E}(Y^n) = 1 - \int_0^1 ny^{n-1}\big(\mathrm{F}(y)\big)^2\,\mathrm{d}y.$$

（当 $n = 0$ 时，该积分就是 0。）要求 $\mathrm{E}(Y^n)$ 的下界，我们需要求这个积分的上界。

> **想一想** $\big(\mathrm{F}(y)\big)^2$ 的两个因子中，应该把哪一个换成 $My$?

只换其中一个。由于 $0 \leq \mathrm{F}(y) \leq My$，有 $\big(\mathrm{F}(y)\big)^2 \leq My\,\mathrm{F}(y)$，加上 $ny^{n-1} \geq 0$，所以

$$\int_0^1 ny^{n-1}\big(\mathrm{F}(y)\big)^2\,\mathrm{d}y \leq nM\int_0^1 y^n\mathrm{F}(y)\,\mathrm{d}y.$$

若把两个因子都换掉，就只剩下 $\int y^{n+1}\,\mathrm{d}y$，对 $X$ 毫无信息；保留一个 F，便可再做一次分部积分，得出 $X$ 的一个矩（moment）。取 $u = \mathrm{F}(y)$ 及 $\mathrm{d}v = y^n\,\mathrm{d}y$ 进行分部积分，

$$\int_0^1 y^n\mathrm{F}(y)\,\mathrm{d}y = \Big[\frac{y^{n+1}}{n+1}\mathrm{F}(y)\Big]_0^1 - \int_0^1 \frac{y^{n+1}}{n+1}\mathrm{f}(y)\,\mathrm{d}y = \frac{1-\mu_{n+1}}{n+1}.$$

综合以上，

$$\mathrm{E}(Y^n) \geq 1 - \frac{nM}{n+1}\left(1-\mu_{n+1}\right) = 1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1},$$

连同上界，便是所求的一对不等式（inequality）。

**(c)** 把 (b) 的两端结合起来，对 $n \geq 0$，

$$1 + \frac{nM}{n+1}\mu_{n+1} - \frac{nM}{n+1} \leq 2M\mu_{n+1},$$

再把含 $\mu_{n+1}$ 的项移到右边，

$$1 - \frac{nM}{n+1} \leq \left(2M - \frac{nM}{n+1}\right)\mu_{n+1} = \frac{(n+2)M}{n+1}\,\mu_{n+1}.$$

系数 $\frac{(n+2)M}{n+1}$ 为正（$M>0$，因为 f 的积分等于 1），所以可以除以它而不改变不等号方向：

$$\mu_{n+1} \geq \frac{n+1}{(n+2)M} - \frac{n}{n+2}.$$

此式对每个 $n \geq 0$ 成立。令 $m = n+1$，则 $m \geq 1$，上式变为

$$\mu_m \geq \frac{m}{(m+1)M} - \frac{m-1}{m+1} \qquad (m \geq 1),$$

这正是所求结果（只需把 $m$ 改名为 $n$）。

## 验算

- 均匀分布（uniform distribution）的 $X$：$\mathrm{f}=1$，$M=1$，$\mu_n = \frac{1}{n+1}$。(c) 中的界为 $\frac{n}{n+1} - \frac{n-1}{n+1} = \frac{1}{n+1}$，所以等号成立。此时 $Y$ 的概率密度函数为 $2y$，且 $\mathrm{E}(Y^n) = \frac{2}{n+2}$，而 (b) 中的两个界亦都等于 $\frac{2}{n+2}$。这是合理的：我们用过的每一个不等式 $\mathrm{F}(y) \leq My$，对均匀分布而言都是等式。
- $\mathrm{f}(x) = 3x^2$，$M=3$，$n=2$：$\mu_2 = \frac{3}{5}$，而界为 $\frac{2}{9} - \frac{1}{3} < 0$，轻松成立。对这个概率密度函数，$\mathrm{E}(Y^2) = \frac{3}{4}$ 及 $\mu_3 = \frac{1}{2}$，所以 (b) 中的界为 $1 + 2\cdot\frac{1}{2} - 2 = 0$ 及 $2\cdot 3\cdot\frac{1}{2} = 3$，而 $\frac{3}{4}$ 确实介乎两者之间。
- 我们亦以数值积分，对多个概率密度函数及多种 g 的选择，验证了 (i)(b) 中的恒等式及所有不等式。

## 常见错误

- 以 $M^2y^2$ 作为 $\left(\mathrm{F}\right)^2$ 的上界：这虽然正确，却毫无用处，因为它丢失了关于 $X$ 的所有信息。

- 忘记 $\mathrm{E}(Y^n)$ 的表达式中积分前面带有负号，所以要求 $\mathrm{E}(Y^n)$ 的下界，就需要积分的上界。
- 得出 $\mu_{n+1}$ 的界后便停下，没有移动下标，以得出题目所述对 $n \geq 1$ 的形式。

## STEP 之外

概率密度函数 $2\mathrm{F}(y)\mathrm{f}(y)$ 并非偶然：它是 $X$ 的两个独立副本的 $\max(X_1, X_2)$ 的概率密度函数，因为 $\mathrm{P}(\max \leq y) = \big(\mathrm{F}(y)\big)^2$。更一般地，$n$ 个独立观测值中第 $r$ 小者的概率密度函数，由 F 和 $1 - \mathrm{F}$ 的幂组成；这些*顺序统计量*（order statistics）是无母数统计（non-parametric statistics）的核心。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）对顺序统计量有清晰的讲解。

# 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

**2013 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。

