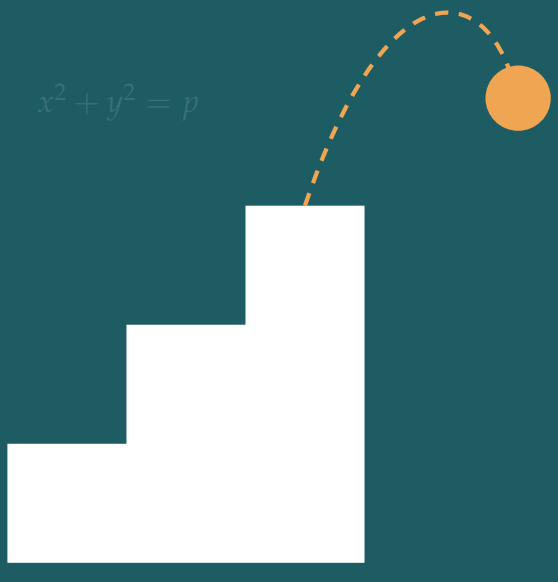
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2014 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

2014 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2014 年详细解答

2014 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 平方差 纯数学 . . . 3

第 2 题 对数函数的积分 纯数学 . . . 7

第 3 题 多项式方程与不等式 纯数学 . . . 11

第 4 题 相关变率 纯数学 . . . 15

第 5 题 三次不等式 纯数学 . . . 19

第 6 题 逻辑斯谛映射的递推关系 纯数学 . . . 23

第 7 题 三角形中的向量 纯数学 . . . 27

第 8 题 直线族的包络线 纯数学 . . . 31

第 9 题 受水平力作用的抛体 力学 . . . 34

第 10 题 碰撞与恢复系数 力学 . . . 38

第 11 题 连接滑轮系统 力学 . . . 41

第 12 题 离散随机变量 概率与统计 . . . 45

第 13 题 角分布 概率与统计 . . . 49

STEP II

第 1 题 三角形几何与二次方程 纯数学 . . . 53

第 2 题 积分不等式 纯数学 . . . 57

第 3 题 切线与包络线 纯数学 . . . 61

第 4 题 代换积分法 纯数学 . . . 65

第 5 题 齐次微分方程 纯数学 . . . 68

第 6 题 三角级数 纯数学 . . . 72

第 7 题 绝对值函数的图像 纯数学 . . . 76

第 8 题 最大二项式系数 纯数学 . . . 81

第 9 题 刚体平衡 力学 . . . 85

第 10 题 轨道的包络线 力学 89
第 11 题 圆周运动 力学 93
第 12 题 风险函数 概率与统计 98
第 13 题 离散随机变量 概率与统计 102

STEP III

第 1 题 对数级数展开 纯数学 106
第 2 题 双曲函数积分法 纯数学 110
第 3 题 到抛物线的距离 纯数学 114
第 4 题 积分不等式 纯数学 118
第 5 题 几何中的复数 纯数学 122
第 6 题 以凸性证明不等式 纯数学 126
第 7 题 向量与圆内接四边形 纯数学 130
第 8 题 级数的收敛 纯数学 134
第 9 题 有阻力的抛体 力学 138
第 10 题 耦合振动 力学 142
第 11 题 圆周运动 力学 146
第 12 题 对数正态分布 概率与统计 150
第 13 题 概率生成函数 概率与统计 154

详细解答

STEP I 2014，第 1 题

纯数学 · 平方差

> **STEP I 2014，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/14-s1-q1

关键思路

一切都建基于一个因式分解（factorisation）：$a^2 - b^2 = (a-b)(a+b)$。把 n 写成两个平方数之差（difference of two squares），等同于把它写成 $n = de$，其中 $d = a - b$、$e = a + b$；由于 d 与 e 相差偶数 $2b$，两者的奇偶性（parity）必须相同。看清这一点之后，每一部分都会变成关于奇偶性相同的因数对（factor pair）的问题。

解答

全文中，a 与 b 均为非负整数。若 $n = a^2 - b^2$ 且 $n \geq 1$，则 $a > b$，因此我们总可以把一种写法记作 (a, b)，其中 $a > b \geq 0$。

第 (i) 部分

$$3 = 2^2 - 1^2, \quad 5 = 3^2 - 2^2, \quad 8 = 3^2 - 1^2, \quad 12 = 4^2 - 2^2, \quad 16 = 5^2 - 3^2.$$

（对于 16，显而易见的 $4^2 - 0^2$ 并不符合要求，因为平方数必须非零。）

第 (ii) 部分

相邻平方数之差是奇数：$(m+1)^2 - m^2 = 2m + 1$。因此任何奇数 $2m + 1$（其中 $m \geq 0$）都可写成

$$2m + 1 = (m+1)^2 - m^2.$$

第 (iii) 部分

相隔二的平方数之差是 4 的倍数：$(k+1)^2-(k-1)^2=4k$。当 $k\geq 1$ 时，$k+1$ 与 $k-1$ 均为非负，所以

$$4k=(k+1)^2-(k-1)^2.$$

当 $k=0$ 时，直接有 $0=0^2-0^2$（或 1^2-1^2）。

第 (iv) 部分

每个平方数除以 4 的余数都是 0 或 1：$(2m)^2=4m^2$ 及 $(2m+1)^2=4(m^2+m)+1$。因此 a^2-b^2 的余数为

$$0-0=0,\quad 1-1=0,\quad 1-0=1\quad 或\quad 0-1\equiv 3\pmod 4,$$

而永远不会是 2。因此没有任何形如 $4k+2$ 的数是两个平方数之差。

（等价地说：若 $(a-b)(a+b)=4k+2$，则 $a-b$ 与 $a+b$ 的奇偶性相同。两者皆奇，乘积便是奇数；两者皆偶，乘积便可被 4 整除。两种情况都不会得到 $4k+2$。）

第 (v) 部分

设 $pq=a^2-b^2=(a-b)(a+b)$，其中 $a>b\geq 0$。令 $d=a-b$、$e=a+b$。那么 d 与 e 是正整数，且 $d\leq e$、$de=pq$。反之，任何满足 $d\leq e$ 且 d、e 奇偶性相同的分解 $pq=de$，都恰好给出一种写法，即

$$a=\frac{e+d}{2},\qquad b=\frac{e-d}{2},$$

而它们都是非负整数。所以 pq 的各种写法与这类因数对 (d,e) 一一对应（one-to-one correspondence）。

> **想一想** 为甚么这里的 d 与 e 总是奇偶性相同？又为甚么只有两对？

由于 p 与 q 是奇质数（odd prime），pq 是奇数，所以 pq 的每个因数（divisor）都是奇数，奇偶条件自动成立。取 $p\leq q$。根据唯一分解定理（unique factorisation），满足 $d\leq e$ 的因数对只有

$$(d,e)=(1,pq)\qquad 及\qquad (d,e)=(p,q).$$

（当 $p=q$ 时这仍然成立：p^2 的因数是 1、p 及 p^2。）由此得到

$$pq=\left(\frac{pq+1}{2}\right)^2-\left(\frac{pq-1}{2}\right)^2\qquad 及\qquad pq=\left(\frac{p+q}{2}\right)^2-\left(\frac{q-p}{2}\right)^2.$$

这两种写法是不同的，因为它们的 $d=a-b$ 值分别是 1 和 p，而 $p\geq 3$。所以 pq 恰好有两种方法写成两个平方数之差。例如 $15=8^2-7^2=4^2-1^2$。当 $p=q$ 时，第二种写法是 p^2-0^2，这在此处是容许的，因为只有第 (i) 部分要求平方数非零。

当 $q=2$ 时，结果不成立。此时 $pq=2p$，其中 p 是奇数，设 $p=2m+1$，则 $2p=4m+2$。根据第 (iv) 部分，$2p$ 根本无法以任何方式写成两个平方数之差。（因数对 $(1,2p)$ 与 $(2,p)$ 都是一奇一偶。）

第 (vi) 部分

与第 (v) 部分一样，675 的各种写法对应于满足 $d\le e$ 且 d、e 奇偶性相同的因数对 $675=de$。由于 675 是奇数，它的所有因数都是奇数，所以每个因数对都算数。

现在 $675=27\times25=3^3\times5^2$，它有 $(3+1)(2+1)=12$ 个因数。这些因数按 $d\leftrightarrow 675/d$ 两两配对，而且没有因数与自己配对，因为 675 不是完全平方数（perfect square）（$26^2=676$）。所以共有 $12/2=6$ 对：

(d,e)	$(1,675)$	$(3,225)$	$(5,135)$	$(9,75)$	$(15,45)$	$(25,27)$
(a,b)	$(338,337)$	$(114,111)$	$(70,65)$	$(42,33)$	$(30,15)$	$(26,1)$

所以 675 恰好有 **6** 种不同的方法写成两个平方数之差。

? **想一想** (a,b) 与 $(-a,b)$，或者 a^2-b^2 与 $-(b^2-a^2)$，应否算作不同的写法?

不应该。题目把一切限制于非负整数，而 $a^2-b^2=675>0$ 迫使 $a>b$。所以每种写法就是单一一对 (a,b)，其中 $a>b\ge0$，而这正是我们所数的。

验算

- 抽查表格：$26^2-1^2=676-1=675$ 及 $30^2-15^2=900-225=675$。
- 第 (ii)–(iv) 部分合起来说明：$n\ge1$ 是两个平方数之差，当且仅当 $n\not\equiv2\pmod4$。用电脑逐一检查不超过 2000 的每个 n，结果吻合；另外亦验证了对所有奇质数 $p\le q<200$，pq 恰好有两种表示法，而 $2p$ 则一种也没有。
- 对所有 b 进行搜索，找到 675 恰好有六种表示法，正是表中那六种。

常见错误

- 忘记奇偶条件：例如对 $2p$ 把 $(2,p)$ 这类因数对也算进去，这会给出非整数 $a=\frac{p+2}{2}$。
- 在第 (v) 部分漏掉 $p=q$ 的情况，或在那里排除 p^2-0^2。除第 (i) 部分外，零是容许的。
- 在第 (vi) 部分数了因数的个数（12），而不是因数对的数目（6）。

STEP 之外

这里的规律——表示法的数目由 n 的因数决定——是初等数论（elementary number theory）中的典型现象。一个著名的近亲是费马（Fermat）的二平方和定理：奇质数可写成 $x^2 + y^2$，当且仅当它是 1 mod 4；这个命题可以利用高斯整数（Gaussian integers）$a + bi$ 简洁地证明。想了解更多，可阅读 Davenport 的 *The Higher Arithmetic*（《高等算术》）。

详细解答

STEP I 2014，第 2 题

纯数学 · 对数函数的积分

> **STEP I 2014，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/14-s1-q2

关键思路

一切都建基于因式分解（factorising）$x^2-4=(x-2)(x+2)$，这样二次式的对数（logarithm）便可拆成 $\ln(2-x)+\ln(2+x)$，而这两部分都可以用第 (i) 部分（及其镜像）的结果即时积分。至于绝对值（modulus）符号，只是要求我们留意 $\ln|x^2-4|$ 在哪里为正、在哪里为负，而第 (ii) 部分的草图正好显示了这一点。

解答

第 (i) 部分

当 $x<2$ 时，$2-x>0$，因此所有式子都有定义。我们对右方求导，并对 $-(2-x)\ln(2-x)$ 使用乘积法则（product rule）。$-(2-x)$ 的导数（derivative）是 1，而根据链式法则（chain rule），$\ln(2-x)$ 的导数是 $-\frac{1}{2-x}$。所以

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left[-(2-x)\ln(2-x)+(2-x)\right] &= 1\cdot\ln(2-x)-(2-x)\cdot\left(-\frac{1}{2-x}\right)-1\\ &= \ln(2-x)+1-1=\ln(2-x),\end{aligned}$$

可见题中所给的原函数（antiderivative）正确。

（另一做法是代换（substitution）$u=2-x$，并利用 $\int \ln u\,\mathrm{d}u = u\ln u-u$。）

同理，当 $x>-2$ 时，

$$\int \ln(2+x)\,\mathrm{d}x=(2+x)\ln(2+x)-(2+x)+c,$$

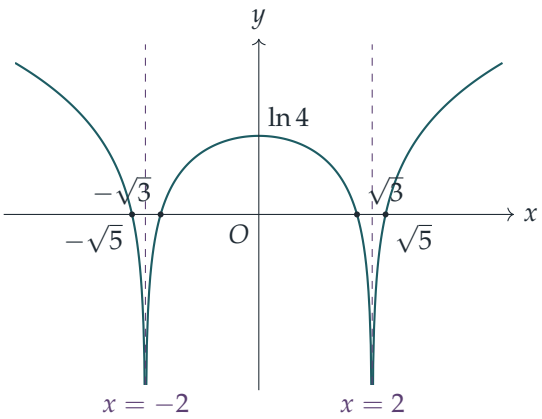
这个结果稍后亦会用到。

第 (ii) 部分

设 $f(x) = \ln|x^2 - 4|$，其定义域为 $x \neq \pm 2$。由于 $f(-x) = f(x)$，曲线关于 y 轴对称 (symmetric)。

- 当 $|x| < 2$ 时：$f(x) = \ln(4 - x^2)$，在 $x = 0$ 取得最大值，此时 $y = \ln 4$。当 $4 - x^2 = 1$，即 $x = \pm\sqrt{3}$ 时，函数值为零；当 $x \to \pm 2$ 时，函数趋向 $-\infty$。
- 当 $|x| > 2$ 时：$f(x) = \ln(x^2 - 4)$，在 $x = \pm 2$ 附近由 $-\infty$ 上升，在 $x^2 - 4 = 1$，即 $x = \pm\sqrt{5}$ 处与轴相交，之后缓慢增长，增长方式类似 $2\ln|x|$。

直线 $x = \pm 2$ 是铅垂渐近线（vertical asymptote）。



第 (iii) 部分

由正 x 轴、y 轴及 A 所围成的有限区域，就是 $x = 0$ 与 $x = \sqrt{3}$ 之间的拱形部分，在此范围内 $y = \ln(4 - x^2) \geq 0$。其面积为

$$\int_0^{\sqrt{3}} \ln(4 - x^2)\,\mathrm{d}x = \int_0^{\sqrt{3}} \ln(2 - x)\,\mathrm{d}x + \int_0^{\sqrt{3}} \ln(2 + x)\,\mathrm{d}x,$$

这里用了 $\ln(4 - x^2) = \ln(2 - x) + \ln(2 + x)$；由于两个因式在此范围内皆为正，等式成立。根据第 (i) 部分，

$$\begin{aligned}\int_0^{\sqrt{3}} \ln(2 - x)\,\mathrm{d}x &= \Big[-(2 - x)\ln(2 - x) + (2 - x)\Big]_0^{\sqrt{3}} \\ &= -(2 - \sqrt{3})\ln(2 - \sqrt{3}) + (2 - \sqrt{3}) + 2\ln 2 - 2 \\ &= -(2 - \sqrt{3})\ln(2 - \sqrt{3}) - \sqrt{3} + 2\ln 2,\end{aligned}$$

以及

$$\begin{aligned}\int_0^{\sqrt{3}} \ln(2+x)\,\mathrm{d}x &= \Big[(2+x)\ln(2+x)-(2+x)\Big]_0^{\sqrt{3}}\\ &= (2+\sqrt{3})\ln(2+\sqrt{3})-(2+\sqrt{3})-2\ln 2+2\\ &= (2+\sqrt{3})\ln(2+\sqrt{3})-\sqrt{3}-2\ln 2.\end{aligned}$$

想一想 怎样把 $\ln(2-\sqrt{3})$ 与 $\ln(2+\sqrt{3})$ 合并?

由于 $(2-\sqrt{3})(2+\sqrt{3})=4-3=1$，所以 $\ln(2-\sqrt{3})=-\ln(2+\sqrt{3})$。把两个积分相加，$\ln 2$ 项互相抵消，得

$$\begin{aligned}\int_0^{\sqrt{3}} \ln(4-x^2)\,\mathrm{d}x &= (2-\sqrt{3})\ln(2+\sqrt{3})+(2+\sqrt{3})\ln(2+\sqrt{3})-2\sqrt{3}\\ &= 4\ln(2+\sqrt{3})-2\sqrt{3},\end{aligned}$$

正如所求。

第 (iv) 部分

在 $|x|<2$ 的范围内，B 就是 A 在两条渐近线之间的部分，并把位于轴下方的部分向上反射。由草图可见，当 $|x|\le\sqrt{3}$ 时 $\ln(4-x^2)\ge 0$，而当 $\sqrt{3}<|x|<2$ 时 $\ln(4-x^2)<0$。利用对称性，所求面积为

$$2\left[\int_0^{\sqrt{3}} \ln(4-x^2)\,\mathrm{d}x-\int_{\sqrt{3}}^{2} \ln(4-x^2)\,\mathrm{d}x\right].$$

第二个积分是反常积分，因为被积函数在 $x=2$ 处趋向 $-\infty$。我们先积分至 $x=2-\varepsilon$，再令 $\varepsilon\to 0^+$。在第 (i) 部分 $\ln(2-x)$ 的原函数中，代入 $x=2-\varepsilon$ 得

$$-\varepsilon\ln\varepsilon+\varepsilon\to 0,$$

这是根据题中所给“当 $t\to 0$ 时 $t\ln t\to 0$”这一事实。因此上限在该处的贡献为 0，于是

$$\begin{aligned}\int_{\sqrt{3}}^{2} \ln(2-x)\,\mathrm{d}x &= 0-\Big[-(2-\sqrt{3})\ln(2-\sqrt{3})+(2-\sqrt{3})\Big]\\ &= -(2-\sqrt{3})\ln(2+\sqrt{3})-2+\sqrt{3},\end{aligned}$$

这里再次用了 $\ln(2-\sqrt{3})=-\ln(2+\sqrt{3})$。另一部分是普通积分:

$$\begin{aligned}\int_{\sqrt{3}}^{2} \ln(2+x)\,\mathrm{d}x &= (4\ln 4-4)-\Big[(2+\sqrt{3})\ln(2+\sqrt{3})-(2+\sqrt{3})\Big]\\ &= 8\ln 2-2+\sqrt{3}-(2+\sqrt{3})\ln(2+\sqrt{3}).\end{aligned}$$

相加，得

$$\int_{\sqrt{3}}^{2} \ln(4-x^2)\,\mathrm{d}x = 8\ln 2 - 4 + 2\sqrt{3} - 4\ln(2+\sqrt{3}),$$

其值约为 -0.259：是负数，与预期相符。因此面积为

$$\begin{aligned} &2\Big[\big(4\ln(2+\sqrt{3}) - 2\sqrt{3}\big) - \big(8\ln 2 - 4 + 2\sqrt{3} - 4\ln(2+\sqrt{3})\big)\Big] \\ &\quad = 16\ln(2+\sqrt{3}) - 16\ln 2 + 8 - 8\sqrt{3} \\ &\quad = 16\ln\Big(\frac{2+\sqrt{3}}{2}\Big) + 8 - 8\sqrt{3} \approx 4.125. \end{aligned}$$

验算

- 把 $[0,\sqrt{3}]$ 与 $[\sqrt{3},2]$ 上的两个积分相加，得 $\int_0^2 \ln(4-x^2)\,\mathrm{d}x = 8\ln 2 - 4 \approx 1.545$。这个结果亦可直接由第 (i) 部分一行得出，因为在 $x = 0$ 处，两个原函数分别给出 $-2\ln 2 + 2$ 和 $2\ln 2 - 2$，两者互相抵消。
- 第 (iii) 部分的面积是 $4\ln(2+\sqrt{3}) - 2\sqrt{3} \approx 1.804$。这比包围该拱形的长方形面积 $\sqrt{3} \times \ln 4 \approx 2.40$ 小，理应如此。
- 对 $\big|\ln|x^2-4|\big|$ 在 $(-2,2)$ 上进行数值积分（numerical integration），得 4.1246，与 $16\ln\big(\frac{2+\sqrt{3}}{2}\big) + 8 - 8\sqrt{3} = 4.12457\ldots$ 吻合。

常见错误

- 忘记曲线在 $x = \sqrt{3}$ 与 $x = 2$ 之间的部分位于轴下方，因而把有向积分加上去，而不是减去。
- 没有把 $\ln(2-\sqrt{3})$ 合并：正是恒等式 $(2-\sqrt{3})(2+\sqrt{3}) = 1$ 令答案变得简洁。
- 不加说明便直接把 $x = 2$ 代入 $(2-x)\ln(2-x)$：该式在此处并无定义，真正支持取值为 0 的，是极限（limit）$t\ln t \to 0$。

STEP 之外

第 (iv) 部分的积分是一个*反常积分*（improper integral）：被积函数无界，但面积仍然有限，因为对数爆增得极慢，以致 $t\ln t \to 0$。到了大学，你会把这一点严谨化，准确判断对于无界的 g，$\int_0^1 g(t)\,\mathrm{d}t$ 何时收敛（converges）（例如，$\int_0^1 t^{-p}\,\mathrm{d}t$ 收敛当且仅当 $p < 1$）。Abbott 的 *Understanding Analysis*（《理解分析》）对 $t\ln t \to 0$ 这类极限以及积分的定义都有细致的讨论。

详细解答

STEP I 2014，第 3 题

纯数学 · 多项式方程与不等式

> **STEP I 2014，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/14-s1-q3

关键思路

两个积分都是初等积分，所以条件不过是一条关于 a 和 b 的多项式方程。关键在于找出公因式 $b-a$（它不为零，所以可以约去），以及在第 (iii) 部分中，留意到剩下的式子具有足够的对称性，可以用和 $b+a$ 与差 $b-a$ 简洁地表示。

解答

计算两个积分后，条件为

$$\frac{b^3-a^3}{3}=\left(\frac{b^2-a^2}{2}\right)^2=\frac{(b-a)^2(b+a)^2}{4}. \qquad (*)$$

由于 $b^3-a^3=(b-a)(b^2+ab+a^2)$ 且 $b-a>0$，我们可以把 $(*)$ 除以 $b-a$，得到

$$4(b^2+ab+a^2)=3(b-a)(b+a)^2. \qquad (**)$$

第 (i) 部分

当 $a=0$ 时，$(**)$ 变成 $4b^2=3b^3$。由于 $b>0$，可以除以 b^2，得到

$$b=\tfrac{4}{3}.$$

第 (ii) 部分

当 $a=1$ 时，$(**)$ 变成

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

即 $3b^3+3b^2-3b-3-4b^2-4b-4=0$，亦即

$$3b^3-b^2-7b-7=0,$$

得证。

现设 $f(x)=3x^3-x^2-7x-7$，则

$$f'(x)=9x^2-2x-7=(9x+7)(x-1).$$

驻点（stationary point）位于 $x=-\frac{7}{9}$（局部极大值（local maximum），因为这个三次函数的首项系数为正）及 $x=1$（局部极小值（local minimum）），且

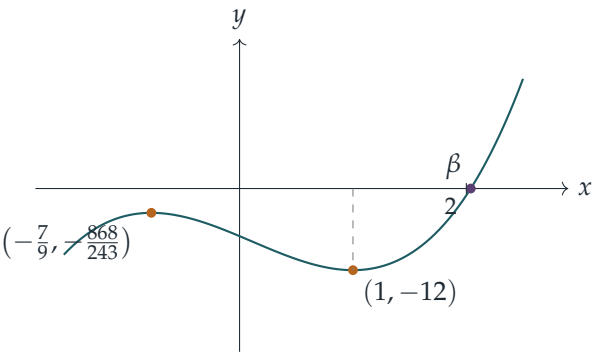
$$f(1)=3-1-7-7=-12,\qquad f\left(-\tfrac{7}{9}\right)=\frac{-343-147+1323-1701}{243}=-\frac{868}{243}.$$

想一想 知道局部极大值的数值，如何告诉我们有多少个根?

局部极大值是负数。当 $x\le 1$ 时，f 先递增至 $x=-\frac{7}{9}$，然后递减，所以在这范围内 $f(x)\le f(-\frac{7}{9})<0$：没有满足 $x\le 1$ 的根。当 $x>1$ 时，f 严格递增（f' 的两个因式皆为正），且 $f(x)\to\infty$，所以它恰好穿过零一次。因此恰好有一个实根，而

$$f(2)=24-4-14-7=-1<0,\qquad f(3)=81-9-21-7=44>0,$$

所以由连续性（continuity）可知，这个根位于 2 与 3 之间。



草图显示曲线 $y=3x^3-x^2-7x-7$：两个转向点（turning point）都位于 x 轴下方，所以曲线只穿过 x 轴一次，位于 $\beta\approx 2.04$。

第 (iii) 部分

设 $p=b+a$ 及 $q=b-a$，则 $p^2=b^2+2ab+a^2$ 及 $q^2=b^2-2ab+a^2$，所以

$$3p^2+q^2=4b^2+4ab+4a^2=4(b^2+ab+a^2).$$

因此 $(**)$ 的左方是 $3p^2+q^2$，而右方是 $3qp^2$。所以

$$3p^2+q^2=3p^2q,$$

得证。移项得 $q^2=3p^2(q-1)$。由于 $q>0$，左方为正，所以 $q\neq 1$，且

$$p^2=\frac{q^2}{3(q-1)}.$$

至于推论部分：

- 下界（*lower bound*）。$q^2>0$ 且 $p^2>0$（因为 $p=b+a>0$），所以 $3(q-1)=q^2/p^2>0$，即 $q>1$。
- 上界（*upper bound*）。由于 $a\geq 0$，我们有 $p=q+2a\geq q>0$，所以 $p^2\geq q^2$。因此

$$\frac{q^2}{3(q-1)}\geq q^2 \iff 1\geq 3(q-1) \iff q\leq \tfrac{4}{3},$$

  其中我们除以 $q^2>0$ 并乘以 $3(q-1)>0$，两者都不会改变不等号的方向。

因此 $1<b-a\leq \frac{4}{3}$，而等号成立若且唯若 $a=0$，这正是第 (i) 部分的情况。

验算

- 第 (i) 部分与第 (iii) 部分吻合：$a=0$ 给出 $q=b=\frac{4}{3}$，正是范围的上端。
- 第 (ii) 部分同样吻合：根为 $b\approx 2.0390$，所以 $q=b-1\approx 1.039$，位于 $(1,\frac{4}{3}]$ 之内。在 $[1,2.0390]$ 上计算两个积分，得到的数值相等（约为 2.4924），而且准确至多个小数位。
- 反过来，$(1,\frac{4}{3}]$ 中任何 q 都给出 $p=q/\sqrt{3(q-1)}\geq q$，从而得到合法的 $a=\frac{1}{2}(p-q)\geq 0$；所以这个界是最佳的（sharp）。

常见错误

- 在第 (ii) 部分忘记约去 $b-a$，结果得到一条带有多余根 $b=1$ 的四次方程。
- 只证明 $f(2)<0<f(3)$：这只证明了在 $(2,3)$ 内有一个根，却没有证明它是唯一的实根。你还需要考虑转向点。
- 在第 (iii) 部分，未确立 $q-1>0$ 便把不等式除以 $q-1$。

STEP 之外

积分形式的柯西－施瓦茨不等式（Cauchy–Schwarz inequality）给出 $\left(\int_a^b x\,\mathrm{d}x\right)^2 \leq (b-a)\int_a^b x^2\,\mathrm{d}x$，等号只在被积函数 x 为常数时成立，而它并非常数。结合题目给定的条件，这表示 $\int_a^b x^2\,\mathrm{d}x < (b-a)\int_a^b x^2\,\mathrm{d}x$，所以 $b-a>1$：一行便得出第 (iii) 部分的下界。柯西－施瓦茨不等式有多种形式，是大学一年级分析学中最常用的工具之一；Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）细致地建立了黎曼积分（Riemann integral）以及类似这样的不等式。

详细解答

STEP I 2014，第 4 题

纯数学 · 相关变率

STEP I 2014，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/14-s1-q4

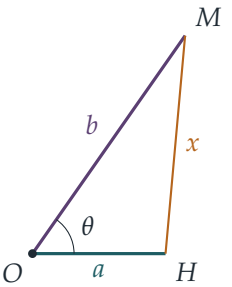
关键思路

两支指针与连接两针尖端的线段构成一个三角形，所以利用余弦定律（cosine rule）便可以用 θ 表示 x。由于时钟准确，θ 以恒定速率增加，因此 x 的增加率就是这个常数乘以 $\mathrm{d}x/\mathrm{d}\theta$。求这个量的最大值就变成一道普通的单变量微积分题目，而驻点（stationary point）方程原来是一条关于 $\cos\theta$ 的二次方程，而且可以整齐地因式分解。

解答

x 的增加率

设 O 为转轴，H 为时针尖端，M 为分针尖端，于是 $OH = a$、$OM = b$、$\angle HOM = \theta$ 及 $HM = x$。



由余弦定律，

$$x^2 = a^2 + b^2 - 2ab\cos\theta.$$

注意 $x \geq b - a > 0$，所以我们可以放心除以 x。对时间 t 求导，

$$2x\frac{\mathrm{d}x}{\mathrm{d}t} = 2ab\sin\theta\,\frac{\mathrm{d}\theta}{\mathrm{d}t}.$$

时钟准确，所以每支指针都以恒定的角速率（angular speed）转动，而两针之间的夹角以恒定速率 $\omega > 0$ 改变（即两个角速率之差；当 $0 \leq \theta < \pi$ 时，分针正在抛离时针，所以 θ 在增加）。因此

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \omega\, ab\,\frac{\sin\theta}{x}.$$

所以我们要找出 $f(\theta) = \dfrac{\sin\theta}{x}$ 在 $0 \leq \theta < \pi$ 上何处取得最大值，同时要记住 x 依赖于 θ，且 $\dfrac{\mathrm{d}x}{\mathrm{d}\theta} = \dfrac{ab\sin\theta}{x}$。

求最大值

由商法则（quotient rule），

$$f'(\theta) = \frac{x\cos\theta - \sin\theta\cdot\dfrac{ab\sin\theta}{x}}{x^2} = \frac{x^2\cos\theta - ab\sin^2\theta}{x^3}.$$

记 $c = \cos\theta$。利用 $x^2 = a^2 + b^2 - 2abc$ 及 $\sin^2\theta = 1 - c^2$，分子为

$$\begin{aligned} x^2\cos\theta - ab\sin^2\theta &= (a^2+b^2)c - 2abc^2 - ab + abc^2 \\ &= -\big(abc^2 - (a^2+b^2)c + ab\big) \\ &= -(ac-b)(bc-a) \\ &= (b-ac)(bc-a). \end{aligned}$$

? **想一想** 我们怎样知道这个驻点是极大点，而且是唯一的驻点?

由于 $c \leq 1 < b/a$，我们有 $ac < b$，所以因子 $b - ac$ 恒为正数。此外 $x^3 > 0$。所以 $f'(\theta)$ 与 $bc - a$ 同号，即与 $\cos\theta - \dfrac{a}{b}$ 同号：

- 当 $\cos\theta > a/b$（θ 较小）时，$f'(\theta) > 0$，f 递增；
- 当 $\cos\theta < a/b$（θ 较大）时，$f'(\theta) < 0$，f 递减。

由于 $\cos\theta$ 在 $[0, \pi)$ 上递减，且 $0 < a/b < 1$，在此范围内恰好有一个 θ 使 $\cos\theta = a/b$，而 f 在此之前上升，在此之后下降。所以这给出 x 增加率的全局最大值。此时

$$x^2 = a^2 + b^2 - 2ab\cdot\frac{a}{b} = b^2 - a^2,$$

所以当 $x = (b^2 - a^2)^{1/2}$ 时，x 的增加率最大，得证。

从几何角度看，$x^2 + a^2 = b^2$ 表示三角形在 H 处是直角：当连接两针尖端的线段与时针垂直时，增加率最大。

$b = 2a$ 的情况

此时 $\cos\theta = a/b = \frac{1}{2}$，所以最大增加率出现在 $\theta = \pi/3$ 时。

以正午后的分钟数量度 t。分针在 60 分钟内转过 2π，时针则在 720 分钟内转过 2π，两者都从竖直位置起顺时针转动。所以两针之间的夹角为

$$\theta = 2\pi\left(\frac{1}{60} - \frac{1}{720}\right)t = 2\pi\cdot\frac{11}{720}\,t = \frac{11\pi}{360}\,t,$$

此式在其值小于 π 时成立。由正午时 $\theta = 0$ 开始，θ 首次到达 $\pi/3$ 的时刻满足

$$\frac{11\pi}{360}\,t = \frac{\pi}{3} \quad\Longrightarrow\quad t = \frac{120}{11} = 10\tfrac{10}{11}\text{ 分钟.}$$

这大约是 10.91 分钟（10 分钟零大约 55 秒），比 11 分钟稍少，得证。

验算

- 当 $\theta = 0$ 时，增加率为 0（两针重叠，$x = b - a$ 处于最小值）；当 $\theta \to \pi$ 时，增加率再次趋向 0（此时 $x \to a + b$ 处于最大值）。所以最大值严格位于两者之间，正如我们所预期。
- 取 $a = 1$、$b = 2$，逐分钟追踪两针尖端的位置并估算它们之间距离的变化率，最大变化率出现在 $t \approx 10.909$ 分钟，此时 $x \approx 1.732 = \sqrt{3} = \sqrt{b^2 - a^2}$。
- 答案略少于 11，是因为假如时针静止不动，分针恰好需要 10 分钟到达 60°；但时针会慢慢向前移，所以需要多一点时间。120/11 与大家熟悉的“两针在 720/11 分钟后再次重合”属于同一类数字。

常见错误

- 求 x 本身的最大值，或者求 $\mathrm{d}x/\mathrm{d}\theta$ 的最大值时，在使用商法则那一步没有留意 x 也依赖于 θ。
- 只用分针的速率（每分钟 $2\pi/60$）作为 $\mathrm{d}\theta/\mathrm{d}t$，这样会得出恰好 10 分钟；其实两针之间的夹角是以两个速率的*差*增长的。
- 忘记舍去 $\cos\theta = b/a$ 这个根；由于 $b > a$，这个根是不可能的。

STEP 之外

这是一道相关变率（related rates）问题：某个量以已知速率改变，我们追踪一个在几何上与之相关的量如何随之变化。同样的想法——用随时间均匀变化的角度来描述位置——

正是研究行星运动的起点：两颗在圆形轨道上运行的行星之间的距离，表现与这里的 x 完全一样，而它们分离得最快的时刻也是用同样方法求得。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）以向量发展这类运动学。

详细解答

STEP I 2014，第 5 题

纯数学 · 三次不等式

> **STEP I 2014，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/14-s1-q5

关键思路

第 (i) 部分是一个单变量三次函数（cubic），它在 $x \geq 0$ 上的最小值恰好是零，并在 $x = a$ 处与轴相切。这一条不等式 $(x+2a)^3 \geq 27a^2x$，其实是经过伪装的算术—几何平均不等式（AM–GM inequality）：左边涉及和 $x+a+a$，右边则涉及积 $x \cdot a \cdot a$。第 (ii) 及 (iii) 部分的关键在于选择以甚么作为 x 和 a，并准确追踪等号何时成立。

解答

第 (i) 部分

求导数（derivative）：

$$f'(x) = 3(x+2a)^2 - 27a^2 = 3\big((x+2a)^2 - 9a^2\big) = 3(x-a)(x+5a),$$

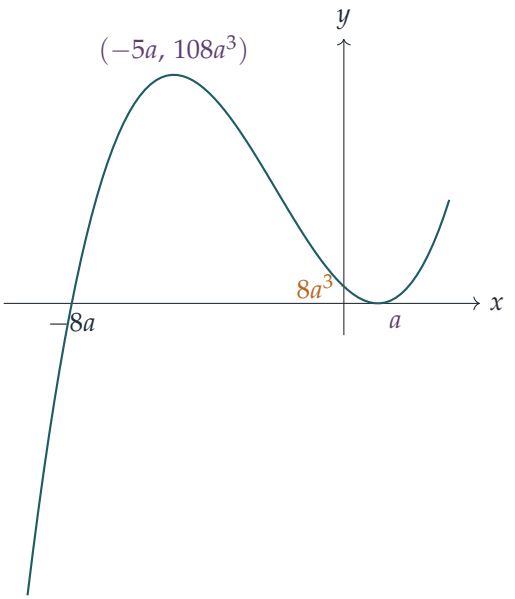
所以驻点（stationary points）位于 $x = a$ 及 $x = -5a$。在这两点，

$$f(a) = (3a)^3 - 27a^3 = 0, \qquad f(-5a) = (-3a)^3 + 135a^3 = 108a^3,$$

而 $f(0) = 8a^3$。由于 f 是首项系数为正的三次函数，$x = -5a$ 是局部极大点，$x = a$ 是局部极小点。因为极小值是 0，图像在 $x = a$ 处与 x 轴相切，所以 $(x-a)^2$ 是一个因式；再比较常数项，便得

$$f(x) = (x-a)^2(x+8a),$$

你可以把两边都展开成 $x^3 + 6ax^2 - 15a^2x + 8a^3$ 来验证。



当 $x \geq 0$ 时，图像由高度 $8a^3 \geq 0$ 开始，下降至 $x = a$ 处的 0，然后一直上升。所以 f 在 $x \geq 0$ 上的最小值是 0，并且

$$f(x) \geq 0 \quad \text{当} x \geq 0,$$

等号只在 $x = a$ 时成立。(因式分解后的形式直接说明了同一件事：当 $x \geq 0$ 时，$(x-a)^2 \geq 0$，而且除非 $x = a = 0$，否则 $x + 8a > 0$。) 若 $a = 0$，图像退化为 $f(x) = x^3$，它在 $x \geq 0$ 时仍然非负，等号只在 $x = 0 = a$ 时成立。在所有情况下，

$$(x+2a)^3 \geq 27a^2x \quad (x, a \geq 0), \qquad \text{等号成立当且仅当} x = a.$$

第 (ii) 部分

我们想处理 xy^2，所以在第 (i) 部分中取 $a = y \geq 0$。对于 $x \geq 0$、$y \geq 0$，

$$27xy^2 \leq (x+2y)^3 \leq 3^3 = 27,$$

第二步成立是因为 $0 \leq x + 2y \leq 3$，而立方是递增函数。因此在整个区域内 $xy^2 \leq 1$。

这个上界是可以达到的：等号需要 $x = y$（来自第 (i) 部分）及 $x + 2y = 3$，由此得 $x = y = 1$，而确实 $1 \cdot 1^2 = 1$。所以 xy^2 的最大值是 1，只在 $x = 1$、$y = 1$ 时取得。

第 (iii) 部分

? **想一想** 第 (i) 部分只涉及两个数 x 和 a。它怎能告诉我们关于三个数的事情?

把 q 和 r 合在一起：设 $x = p$ 及 $2a = q + r$，所以 $a = \frac{1}{2}(q + r) \geq 0$。由第 (i) 部分得

$$(p + q + r)^3 \geq 27p\left(\frac{q + r}{2}\right)^2.$$

而 $\left(\frac{q+r}{2}\right)^2 - qr = \frac{(q-r)^2}{4} \geq 0$，且 $p \geq 0$，所以

$$(p + q + r)^3 \geq 27p\left(\frac{q + r}{2}\right)^2 \geq 27pqr,$$

这正是所求的不等式。

等号情况。若 $(p + q + r)^3 = 27pqr$，则这串不等式中的两个不等号都必须取等号。

- 第一个取等号，当且仅当 $x = a$，即 $p = \frac{1}{2}(q + r)$。
- 第二个取等号，当且仅当 $\frac{27}{4}p(q - r)^2 = 0$，即 $p = 0$ 或 $q = r$。

若 $p = 0$，第一个条件给出 $q + r = 0$，所以亦有 $q = r = 0$。若 $q = r$，第一个条件给出 $p = q$。无论哪种情况，都有 $p = q = r$。反过来，若 $p = q = r$，则两边都等于 $27p^3$。所以等号成立当且仅当

$$p = q = r.$$

验算

- 展开 $(x - a)^2(x + 8a)$ 可得回 $(x + 2a)^3 - 27a^2x$；取 $a = 1$，$f(0) = 8$、$f(1) = 0$、$f(3) = 4 \cdot 11 = 44$，全部非负。
- 在第 (ii) 部分，试试直线 $x + 2y = 3$ 上的其他点：$(x, y) = (2, \frac{1}{2})$ 给出 $\frac{1}{2}$，$(0.5, 1.25)$ 给出约 0.78，两者都小于 1。用电脑在三角形上的细密网格中搜索，亦得到最大值 1，位于 $(1, 1)$。
- 在第 (iii) 部分，$p = 1, q = 2, r = 3$ 给出 $216 \geq 162$；用电脑在较小的正整数中搜索，发现只有当 $p = q = r$ 时等号才成立。

常见错误

- 证明了 $f(a) = 0$，却没有解释为何这是 $x \geq 0$ 上的最小值：你需要用到图像的形状（或因式分解），包括端点值 $f(0) = 8a^3 \geq 0$。
- 在第 (ii) 部分，忘记检查上界 1 是否真的可以达到；单有上界并不等于最大值。
- 在第 (iii) 部分的等号情况中，只得出 $q = r$，而遗漏了另一可能 $p = 0$；这情况必须排除（或证明它迫使 $p = q = r = 0$）。

STEP 之外

第 (iii) 部分就是三项的算术—几何平均不等式 $\frac{1}{3}(p+q+r) \geq \sqrt[3]{pqr}$；而这里的方法——把两个变量合并为它们的平均值，再运用一条双变量不等式——正是 n 项算术—几何平均不等式各种经典证明的雏形。第 (ii) 部分是一个小型的约束最优化（constrained optimisation）问题，大学课程通常以拉格朗日（Lagrange）乘数法处理这类问题，不过像算术—几何平均不等式这类不等式，往往能更简洁地得出答案。如想看更多同类风格的内容，可参阅 Siklos，*Advanced Problems in Mathematics*。

详细解答

STEP I 2014，第 6 题

纯数学 · 逻辑斯谛映射的递推关系

> **STEP I 2014，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/14-s1-q6

本题在 *Beyond STEP* 第 8.7 节可以解出的混沌中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/solvable-chaos.html

关键思路

表达式 $4x(1-x)$ 看起来很别扭，但只要令 $x = \sin^2\theta$，便有 $1-x = \cos^2\theta$，而 $4\sin^2\theta\cos^2\theta = (2\sin\theta\cos\theta)^2$ 其实是乔装了的二倍角公式（double-angle formula）。递推关系（recurrence）的每一步只是把角度加倍。第 (ii) 部分的重点则在于认出：只要系数（coefficients）满足某一个条件，任何二次递推关系都可以通过线性变量代换（linear change of variable）化为这一个。

解答

第 (i) 部分

由 $u_0 = \sin^2\theta$ 得 $1-u_0 = \cos^2\theta$，所以

$$u_1 = 4\sin^2\theta\cos^2\theta = (2\sin\theta\cos\theta)^2 = \sin^2 2\theta.$$

这与 u_0 的形式完全相同，只是把 θ 换成 2θ，因此同样的计算给出

$$u_2 = 4\sin^2 2\theta\ \cos^2 2\theta = \sin^2 4\theta.$$

角度每一步都加倍，所以我们猜想（conjecture）

$$u_n = \sin^2(2^n\theta) \qquad (n \geq 0).$$

用归纳法（*induction*）证明。当 $n=0$ 时，式子即 $u_0=\sin^2\theta$，成立。假设对某个 $k\geq 0$ 有 $u_k=\sin^2(2^k\theta)$。那么 $1-u_k=\cos^2(2^k\theta)$，而

$$\begin{aligned} u_{k+1} &= 4\sin^2(2^k\theta)\cos^2(2^k\theta) \\ &= \big(2\sin(2^k\theta)\cos(2^k\theta)\big)^2 \\ &= \sin^2(2^{k+1}\theta), \end{aligned}$$

这里用了 $\sin 2A = 2\sin A\cos A$，其中 $A=2^k\theta$。因此 k 的结果蕴涵 $k+1$ 的结果，由归纳法，对所有 $n\geq 0$ 都有 $u_n=\sin^2(2^n\theta)$。

第 (ii) 部分

设 $v_n=\alpha u_n+\beta$，其中 $\alpha\neq 0$（需要 $\alpha\neq 0$，才能由 v_n 还原 u_n）。代入 v_n 的递推关系：

$$\alpha u_{n+1}+\beta = -p(\alpha u_n+\beta)^2+q(\alpha u_n+\beta)+r.$$

展开后除以 α，

$$u_{n+1} = -p\alpha\, u_n^2 + (q-2p\beta)\,u_n + \frac{-p\beta^2+(q-1)\beta+r}{\alpha}.$$

我们希望这就是 $u_{n+1}=4u_n-4u_n^2$，所以比较系数。

- u_n^2：$p\alpha=4$，所以 $\alpha=\dfrac{4}{p}$（由于 $p\neq 0$，这没有问题，而且此时 $\alpha\neq 0$ 亦成立）。
- u_n：$q-2p\beta=4$，所以 $\beta=\dfrac{q-4}{2p}$。
- 常数项：$-p\beta^2+(q-1)\beta+r=0$。

前两条确定了 α 和 β；第三条于是成为 p、q、r 须满足的条件。代入 $\beta=(q-4)/(2p)$，再整条乘以 $4p$：

$$\begin{aligned} & 0 = -p\cdot\frac{(q-4)^2}{4p^2}+(q-1)\frac{q-4}{2p}+r \\ \iff\quad & 0 = -(q-4)^2+2(q-1)(q-4)+4pr \\ & \quad = (q-4)\big(-(q-4)+2(q-1)\big)+4pr \\ & \quad = (q-4)(q+2)+4pr \\ & \quad = q^2-2q-8+4pr. \end{aligned}$$

所以若 $4pr=8+2q-q^2$，则取 $\alpha=4/p$ 及 $\beta=(q-4)/(2p)$ 时三个系数全部吻合，而数列 $u_n=(v_n-\beta)/\alpha$ 满足 $(*)$，正如所求。

该特定数列。这里 $p=1$、$q=2$、$r=2$。条件成立：$4pr=8$，而 $8+2q-q^2=8+4-4=8$。我们得到

$$\alpha=4,\qquad \beta=\frac{2-4}{2}=-1,\qquad \text{所以}\qquad v_n=4u_n-1.$$

想一想 第 (i) 部分需要 $u_0 = \sin^2\theta$。这里能否找到合适的 θ?

由于 $v_0 = 1$，我们需要 $u_0 = (1+1)/4 = \frac{1}{2}$，而 $\frac{1}{2} = \sin^2\frac{\pi}{4}$。所以 $\theta = \pi/4$ 可行，由第 (i) 部分得 $u_n = \sin^2(2^n\pi/4) = \sin^2(2^{n-2}\pi)$。因此

$$v_n = 4\sin^2\left(2^{n-2}\pi\right) - 1 = 1 - 2\cos\left(2^{n-1}\pi\right),$$

这里用了 $2\sin^2 A = 1 - \cos 2A$。逐项计算：

- $n = 0$：$v_0 = 4\sin^2\frac{\pi}{4} - 1 = 1$；
- $n = 1$：$v_1 = 4\sin^2\frac{\pi}{2} - 1 = 3$；
- $n \geq 2$：2^{n-2} 是整数，所以 $\sin(2^{n-2}\pi) = 0$，而 $v_n = -1$。

所以数列是

$$1,\ 3,\ -1,\ -1,\ -1,\ \ldots$$

即 $v_0 = 1$、$v_1 = 3$，而对所有 $n \geq 2$ 有 $v_n = -1$。

验算

直接迭代递推关系：$v_1 = -1+2+2 = 3$，$v_2 = -9+6+2 = -1$，以及 $v_3 = -1-2+2 = -1$；由于 -1 映射到自身，数列此后一直停留在 -1，与公式相符。至于第 (i) 部分，试取 $\theta = \pi/6$：$u_0 = \frac{1}{4}$，$u_1 = 4\cdot\frac{1}{4}\cdot\frac{3}{4} = \frac{3}{4} = \sin^2\frac{\pi}{3}$，以及 $u_2 = 4\cdot\frac{3}{4}\cdot\frac{1}{4} = \frac{3}{4} = \sin^2\frac{2\pi}{3}$，与预测一致。公式 $u_n = \sin^2(2^n\theta)$ 亦已用电脑就多个随机角度及首若干项验证过。

常见错误

- 写成 $u_n = \sin^2(2n\theta)$：角度每次加倍，所以增长方式是 2^n，而不是 $2n$。
- 忘记检查初始值 $u_0 = (v_0 - \beta)/\alpha$ 是否落在 $[0,1]$ 内，从而确实能写成 $\sin^2\theta$。
- 把常数项方程当作要解第三个未知数，而不是把它看成题目要你推导出的条件。

STEP 之外

递推关系 $u_{n+1} = 4u_n(1-u_n)$ 是取最极端参数时的逻辑斯谛映射（logistic map），是混沌（chaos）的标准例子：由于 $u_n = \sin^2(2^n\theta)$，把 θ/π 写成二进制（binary）便可看出，每一步只是把二进制数字移动一位，因此两个相近的初始值会以指数速度分开。第 (ii) 部分的代换是一种共轭（conjugacy），意思是两个动力系统（dynamical systems）在变换

坐标之后其实是同一个系统。Strogatz 的 *Nonlinear Dynamics and Chaos*（《非线性动力学与混沌》）是这些概念非常易读的入门书。

详细解答

STEP I 2014，第 7 题

纯数学 • 三角形中的向量

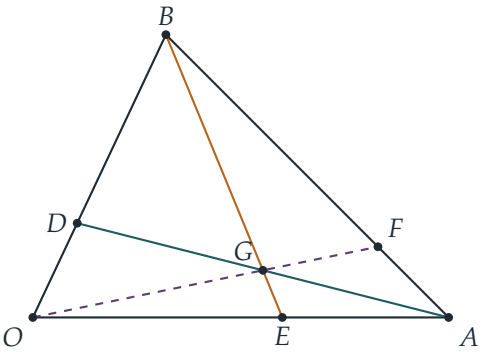
STEP I 2014，第 7 题

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/14-s1-q7

关键思路

把每一点都写成 **a** 与 **b** 的组合。经过两个已知点的直线上的点，有一条单参数的向量方程（vector equation），所以 G 可以用两种方式表示，每条直线各给出一种。由于 **a** 与 **b** 不平行（三角形并非退化），我们可以比较系数（comparing coefficients），从而得到关于两个参数的两条简单方程。



解答

首先确定 D 和 E 的位置。由于 $BD = r\,DO$，点 D 位于由 O 到 B 的路程的 $\frac{1}{1+r}$ 处；由于 $OE = s\,EA$，点 E 位于由 O 到 A 的路程的 $\frac{s}{1+s}$ 处。所以

$$\mathbf{d} = \frac{1}{1+r}\,\mathbf{b}, \qquad \mathbf{e} = \frac{s}{1+s}\,\mathbf{a}.$$

第 (i) 部分

直线 AD 上的一般点为

$$\mathbf{a}+\lambda(\mathbf{d}-\mathbf{a})=(1-\lambda)\,\mathbf{a}+\frac{\lambda}{1+r}\,\mathbf{b},$$

而直线 BE 上的一般点为

$$\mathbf{b}+\mu(\mathbf{e}-\mathbf{b})=\frac{\mu s}{1+s}\,\mathbf{a}+(1-\mu)\,\mathbf{b}.$$

在 G 处，两者相等。

? **想一想** 为甚么我们可以分别令 **a** 的系数相等及 **b** 的系数相等?

因为 O、A、B 构成一个真正的三角形，所以 **a** 与 **b** 都不是零向量，而且互不平行。若 $\alpha\mathbf{a}+\beta\mathbf{b}=\alpha'\mathbf{a}+\beta'\mathbf{b}$ 而 $\alpha\neq\alpha'$，则 **a** 便会是 **b** 的倍数，这是不可能的。所以

$$1-\lambda=\frac{\mu s}{1+s},\qquad \frac{\lambda}{1+r}=1-\mu.$$

由第二条方程得 $\lambda=(1+r)(1-\mu)$。代入第一条方程，

$$\begin{aligned}1-(1+r)(1-\mu)&=\frac{\mu s}{1+s}\\ \mu\left[(1+r)-\frac{s}{1+s}\right]&=r\\ \mu\,\frac{(1+r)(1+s)-s}{1+s}&=r\\ \mu\,\frac{1+r+rs}{1+s}&=r,\end{aligned}$$

所以

$$\mu=\frac{r(1+s)}{1+r+rs}.$$

由于 $r,s>0$，有 $1+r+rs>0$，所以这一步是容许的。把这个 μ 值代入 BE 的方程，**a** 的系数为

$$\frac{\mu s}{1+s}=\frac{rs}{1+r+rs},$$

而 **b** 的系数为

$$1-\mu=\frac{1+r+rs-r-rs}{1+r+rs}=\frac{1}{1+r+rs}.$$

因此

$$\mathbf{g}=\frac{rs}{1+r+rs}\,\mathbf{a}+\frac{1}{1+r+rs}\,\mathbf{b},$$

得证。

第 (ii) 部分

由于直线 OG 经过原点，其上每一点都是 $\mathbf{g}$ 的纯量倍数 $k\mathbf{g}$：

$$k\mathbf{g} = \frac{k}{1+r+rs}\left(rs\,\mathbf{a} + \mathbf{b}\right).$$

由于 AB 由点 $(1-\nu)\mathbf{a} + \nu\mathbf{b}$ 组成，点 $\alpha\mathbf{a} + \beta\mathbf{b}$ 位于直线 AB 上当且仅当 $\alpha + \beta = 1$。所以在 F 处我们需要 $k(rs+1) = 1 + r + rs$，于是

$$\mathbf{f} = \frac{rs}{1+rs}\,\mathbf{a} + \frac{1}{1+rs}\,\mathbf{b}.$$

另一方面，$AF = t\,FB$ 表示 F 位于由 A 到 B 的路程的 $\frac{t}{1+t}$ 处：

$$\mathbf{f} = \frac{1}{1+t}\,\mathbf{a} + \frac{t}{1+t}\,\mathbf{b}.$$

比较 $\mathbf{a}$ 的系数（理由与前面相同）得 $1 + t = \dfrac{1+rs}{rs}$，所以

$$t = \frac{1}{rs}.$$

$\mathbf{b}$ 的系数亦相符：$\dfrac{t}{1+t} = \dfrac{1/(rs)}{(1+rs)/(rs)} = \dfrac{1}{1+rs}$。

验算

- **中线**。当 $r = s = 1$ 时，D 和 E 是中点，所以 G 应该是重心（centroid）。公式给出 $\mathbf{g} = \frac{1}{3}\mathbf{a} + \frac{1}{3}\mathbf{b}$，这是正确的；而 $t = 1$，所以 F 是 AB 的中点。这是必然的，因为三条中线（median）共点（concurrent）。
- **塞瓦定理**。答案表明 $r \cdot s \cdot t = 1$，即 $\dfrac{BD}{DO} \cdot \dfrac{OE}{EA} \cdot \dfrac{AF}{FB} = 1$，这正是共点直线 AD、BE、OF 的塞瓦定理（Ceva's theorem）。
- **一个数值例子**。取 $\mathbf{a} = (5,0)$、$\mathbf{b} = (0,5)$、$r = 2$、$s = 3$。那么 $D = (0, \frac{5}{3})$、$E = (\frac{15}{4}, 0)$，而公式给出 $G = (\frac{30}{9}, \frac{5}{9}) = (\frac{10}{3}, \frac{5}{9})$。这点位于方程为 $x + 3y = 5$ 的 AD 上，亦位于方程为 $4x + 3y = 15$ 的 BE 上。直线 OG 与 $x + y = 5$ 相交于 $F = (\frac{30}{7}, \frac{5}{7})$，而 $\overrightarrow{AF} = (-\frac{5}{7}, \frac{5}{7})$、$\overrightarrow{FB} = (-\frac{30}{7}, \frac{30}{7})$，所以 $AF : FB = 1 : 6$，得 $t = \frac{1}{6} = \frac{1}{rs}$。这些公式亦已用电脑以精确算术，对二千个随机三角形及比例验证过。

常见错误

- 把比例弄反：$BD = r\,DO$ 使 D 位于由 O 到 B 的路程的 $\frac{1}{1+r}$ 处，而不是 $\frac{r}{1+r}$ 处。
- 比较系数时没有说明为何可行：这需要 $\mathbf{a}$ 与 $\mathbf{b}$ 不平行。
- 在第 (ii) 部分忘记把 $\mathbf{g}$ 重新缩放：G 本身并不在 AB 上，因为它的系数之和小于 1。

STEP 之外

像 **f** 的系数那样总和为 1 的系数，称为重心坐标（barycentric coordinates）。它们以各顶点上的权重来描述点，并使塞瓦定理和梅涅劳斯（Menelaus）定理等结果几乎水到渠成。用线性代数（linear algebra）的语言来说，比较系数之所以可行，是因为 **a** 与 **b** 线性独立（linearly independent），因而构成平面的一组基（basis）。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）对基和坐标有细致的阐述。

详细解答

STEP I 2014，第 8 题

纯数学 · 直线族的包络线

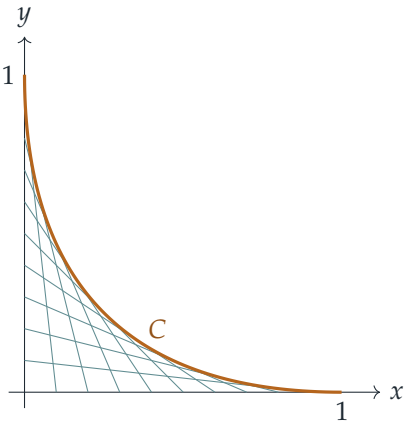
STEP I 2014，第 8 题

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/14-s1-q8

关键思路

把每条直线写成截距式（intercept form）$\frac{x}{a}+\frac{y}{1-a}=1$；这样，两条直线的交点便可从一对线性方程中直接求出，而且结果对 a 和 b 是对称的。令 $b\to a$，便得到一条直线与其“邻居”的交点，这些点描出一条曲线，而每条直线 L_a 都与这条曲线相切。曲线 C 就是这族直线的包络线（envelope），而第 (iii) 部分则从另一个方向证实这一点。



解答

经过 $(a,0)$ 和 $(0,1-a)$ 的直线，其截距（intercept）为 a 和 $1-a$，所以

$$L_a:\ \frac{x}{a}+\frac{y}{1-a}=1,\qquad \text{即}\qquad (1-a)x+ay=a(1-a).$$

第 (i) 部分

我们解

$$(1-a)x + ay = a(1-a),$$
$$(1-b)x + by = b(1-b).$$

为了消去 y，把第一条方程乘以 b，第二条乘以 a，然后相减：

$$\big[b(1-a) - a(1-b)\big]x = ab(1-a) - ab(1-b), \qquad \text{即} \qquad (b-a)x = ab(b-a).$$

由于 $a \neq b$，我们可以除以 $b-a$，得 $x = ab$。代入第一条方程，

$$ay = a(1-a) - (1-a)ab = a(1-a)(1-b),$$

又因 $a \neq 0$，得 $y = (1-a)(1-b)$。所以 L_a 与 L_b 相交于

$$\big(ab,\ (1-a)(1-b)\big).$$

（答案对 a 和 b 对称，这是理所当然的。你也可以验证它同样位于 L_b 上：$(1-b)ab + b(1-a)(1-b) = b(1-b)\big(a + 1 - a\big) = b(1-b)$。）

第 (ii) 部分

两个坐标都是 b 的多项式（polynomial），因此是连续（continuous）的，所以当 $b \to a$ 时的极限（limit）就是在 $b = a$ 处的值：

$$(x, y) = \big(a^2,\ (1-a)^2\big).$$

由于 $0 < a < 1$，我们有 $0 < x = a^2 < 1$，又因为 $a > 0$，取正平方根便得 $\sqrt{x} = a$。因此

$$y = (1-a)^2 = (1-\sqrt{x})^2,$$

所以极限点位于 C 上，得证。

第 (iii) 部分

在 C 上任取一点。它的 x 坐标位于 $(0,1)$ 之内，所以可以写成 t^2，其中 $t = \sqrt{x} \in (0,1)$，而该点就是 $\big(t^2, (1-t)^2\big)$。

用链式法则（chain rule）对 $y = (1-\sqrt{x})^2$ 求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2(1-\sqrt{x}) \cdot \left(-\frac{1}{2\sqrt{x}}\right) = -\frac{1-\sqrt{x}}{\sqrt{x}},$$

在 $x = t^2$ 处其值为 $-\frac{1-t}{t}$。该处的切线（tangent）为

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

想一想 我们怎样看出这条直线是其中一条 L_a?

一条直线 L_a 由它的两个截距完全确定，而这两个截距必须是 a 和 $1-a$。所以，求出切线的截距。

- 令 $y=0$：$-(1-t)^2 = -\frac{1-t}{t}(x-t^2)$。由于 $1-t \neq 0$，可以约去，得 $x-t^2 = t(1-t)$，所以 $x=t$。
- 令 $x=0$：$y=(1-t)^2+\frac{1-t}{t}t^2 = (1-t)^2+t(1-t) = 1-t$。

所以切线经过 $(t,0)$ 和 $(0,1-t)$，且 $0<t<1$：它正是 L_t。由于 C 上每一点都由某个 $t\in(0,1)$ 给出，C 的每条切线都具有 L_a 的形式（在切点处 $a=\sqrt{x}$）。证明完毕。

留意在 $\left(a^2,(1-a)^2\right)$ 处的切线是 L_a，而这一点正是第 (ii) 部分由 L_a 得出的点：每条直线都在它自己的极限交点处与 C 相切。

验算

- 取 $a=\frac{1}{2}$、$b=\frac{1}{3}$，第 (i) 部分给出 $\left(\frac{1}{6},\frac{1}{3}\right)$。在 $L_{1/2}$ 上：$\frac{1}{2}\cdot\frac{1}{6}+\frac{1}{2}\cdot\frac{1}{3}=\frac{1}{4}=\frac{1}{2}\cdot\frac{1}{2}$；在 $L_{1/3}$ 上：$\frac{2}{3}\cdot\frac{1}{6}+\frac{1}{3}\cdot\frac{1}{3}=\frac{2}{9}=\frac{1}{3}\cdot\frac{2}{3}$。
- 在 $x=\frac{1}{4}$ 处，曲线的 $y=\frac{1}{4}$，斜率为 -1，所以切线是 $x+y=\frac{1}{2}$，即经过 $(\frac{1}{2},0)$ 和 $(0,\frac{1}{2})$ 的 $L_{1/2}$。
- 曲线本身是抛物线（parabola）$\sqrt{x}+\sqrt{y}=1$ 的一部分，对 x 和 y 对称，与这族直线的对称性 $a\leftrightarrow 1-a$ 吻合。

常见错误

- 在第 (ii) 部分忘记解释为何 $\sqrt{x}=a$：这需要 $a>0$，而题目已给出此条件。
- 在第 (iii) 部分只证明了 L_a 与 C 相切。题目要求的是逆命题：从 C 的任意一条切线出发，证明它是某条 L_a。

STEP 之外

曲线 C 是直线族 L_a 的*包络线*：即与族中每一条直线都相切的曲线。一般而言，曲线族 $F(x,y,a)=0$ 的包络线，可以从 $F=0$ 和 $\partial F/\partial a=0$ 中消去 a 而求得，这正是第 (ii) 部分“令 $b\to a$”取极限的微积分版本。包络线在光学中以焦散线（caustics）的形式出现，例如咖啡杯中那条明亮的光线曲线。Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多同类题目。

详细解答

STEP I 2014，第 9 题

力学 • 受水平力作用的抛体

STEP I 2014，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：

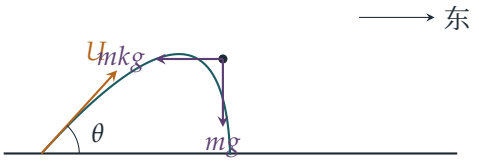
step.maths.org/questions/14-s1-q9

关键思路

两个力都是恒定的，所以每个方向上的加速度（acceleration）都是常数，我们可以把水平运动和竖直运动分开处理，做法和普通抛射体（projectile）完全一样。竖直运动完全不受向西的力影响。新的地方在于水平速度不断减少，所以质点会在某一刻停止向东移动，然后往西飘回去。整条轨迹的形状取决于这一刻何时出现，以及它与飞行最高点和落地时刻的先后关系。

解答

把质点视为在竖直平面内运动的质点（point mass），除了重量和题目给出的水平力之外不受其他力。取 x 向东、y 竖直向上，原点设在抛出点。



时间 T_H 与 T_L

在每个方向上应用牛顿第二定律（Newton's second law），再除以 m，得

$$\ddot{x} = -kg, \qquad \ddot{y} = -g.$$

初速度为 $(U\cos\theta,\ U\sin\theta)$，所以积分两次，得

$$\dot{x} = U\cos\theta - kgt, \qquad x = Ut\cos\theta - \tfrac{1}{2}kgt^2,$$
$$\dot{y} = U\sin\theta - gt, \qquad y = Ut\sin\theta - \tfrac{1}{2}gt^2.$$

当 $\dot{y} = 0$ 时到达最大高度，而质点在 $y = 0$ 且 $t > 0$ 时落地：

$$\boxed{T_H = \frac{U\sin\theta}{g}, \qquad T_L = \frac{2U\sin\theta}{g}.}$$

这些结果与没有向西的力时相同，因为那个力没有竖直分量。

T 的意义

题目给出的时间 $T = \dfrac{U\cos\theta}{kg}$ 正是 $\dot{x} = 0$ 的时刻。在 T 之前质点向东移动；在 T 之后它向西移动。所以，如果质点在时刻 T 仍在空中，那就是它位于最东的时刻。比较这三个时间：

$$\frac{T_H}{T} = \frac{U\sin\theta}{g}\cdot\frac{kg}{U\cos\theta} = k\tan\theta, \qquad \frac{T_L}{T} = 2k\tan\theta.$$

知道质点在哪里落地也会有帮助：

$$x(T_L) = T_L\left(U\cos\theta - \tfrac{1}{2}kgT_L\right) = \frac{2U^2\sin\theta}{g}(\cos\theta - k\sin\theta) = \frac{2U^2\sin\theta\cos\theta}{g}(1 - k\tan\theta).$$

若 $k\tan\theta < 1$，这个值为正（在起点以东）；若 $k\tan\theta > 1$，则为负（在起点以西）。

> ? **想一想** 轨迹不再对称，那么它实际上是甚么样子?

加速度 $(-kg, -g)$ 是常数，所以轨迹仍然是抛物线（parabola），只是它的对称轴倾斜了：这就是在大小为 $g\sqrt{1+k^2}$、指向西下方的“重力”作用下所得到的抛物线。飞行的最高点不再是抛物线的顶点，而下面三种情况讨论的，就是最东点（时刻 T）落在甚么位置。

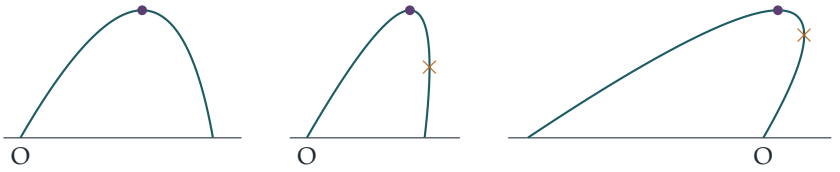
三种情况

情况 $k\tan\theta < \frac{1}{2}$。此时 $T_H < T_L < T$。质点在向东的速度耗尽之前已经落地，所以全程向东移动。轨迹看起来像普通的抛射轨迹，上升到最高点后下降，而由于水平速率不断减少，下降时比上升时更陡。它落在起点以东。

情况 $\frac{1}{2} < k\tan\theta < 1$。此时 $T_H < T < T_L$。质点在仍向东移动时经过最高点，下降时继续向东，直到时刻 T，此时它的速度竖直向下，之后便向西弯回，然后落地。由于 $k\tan\theta < 1$，它仍然落在起点以东，但在其最东点以西。

情况 $k\tan\theta > 1$。此时 $T < T_H$。质点仍在上升时，向东的运动已经停止，那一刻它正竖直向上移动。它在已经向西移动时才到达最大高度，然后下降并落在抛出点以西。

下面的草图展示了这三种情况，取 $\theta = 60°$，而 U 和 g 相同，分别有 $k\tan\theta \approx 0.35$、0.6 和 1.8。圆点标示时刻 T_H 的最高点，叉号标示时刻 T 的最东点，O 为抛出点。



$k\tan\theta < \frac{1}{2}$　　$\frac{1}{2} < k\tan\theta < 1$　　$k\tan\theta > 1$

$k\tan\theta = 1$ 的情况

此时 $T = T_H$，所以在飞行最高点时，速度的两个分量同时为零，质点瞬间静止。此外 $x(T_L) = 0$，所以它落回抛出点。

事实上，质点沿一条直线运动。当 $\tan\theta = 1/k$ 时，初速度 $U(\cos\theta, \sin\theta)$ 是 $(k, 1)$ 的正数倍，而加速度 $(-kg, -g) = -g(k, 1)$ 恰好指向相反方向。一个加速度恒定且与速度方向正好相反的质点会沿直线运动：它沿抛射线向上移动，以 $g\sqrt{1+k^2}$ 减速，在时刻 T_H 停下，然后沿同一直线滑回，在时刻 $T_L = 2T_H$ 落回起点。你可以从方程验证这一点：取 $k = \cot\theta$，

$$x\tan\theta = Ut\sin\theta - \tfrac{1}{2}gt^2 = y \quad \text{对所有} t,$$

所以质点始终不离开直线 $y = x\tan\theta$。

验算

- 令 $k \to 0$，则 $T \to \infty$，所以永远属于第一种情况，而 $x(T_L)$ 变成熟悉的射程 $U^2\sin 2\theta/g$。
- 落地位置 $\frac{2U^2\sin\theta\cos\theta}{g}(1 - k\tan\theta)$ 恰好在 $k\tan\theta = 1$ 时变号，与第二、第三种情况的分界吻合。
- 取 $U = g = 1$、$\theta = 60°$ 和 $k = 0.45$，质点落在 $x \approx 0.191$，但在时刻 $T \approx 1.11$ 到达 $x = \frac{\cos^2\theta}{2k} \approx 0.278$，该时刻介于 $T_H \approx 0.87$ 与 $T_L \approx 1.73$ 之间，所以它确实在落地前折返。对运动进行逐步数值积分（numerical integration），结果与三个时间及落地点全部吻合。

常见错误

- 让向西的力影响竖直运动：T_H 和 T_L 完全不涉及 k。
- 假设最高点就是轨迹的“中间”。有了水平力，轨迹是一条倾斜的抛物线，所以最高点和最东点并不相同。

- 把第二种情况画成质点落在起点以西：它虽然折返，但由于 $k\tan\theta < 1$，折返得不够远。

STEP 之外

任何方向上的恒定加速度总会产生抛物线。这里的等效重力只是倾斜了，把坐标轴旋转至与它对齐，问题便变回普通的抛射体问题。这种选取与受力相配合的坐标的想法，是大学一年级力学的核心内容；同样重要的还有均匀电场中的带电粒子，它们的行为正是如此。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）细致地发展了这些想法。

详细解答

STEP I 2014，第 10 题

力学 · 碰撞与恢复系数

STEP I 2014，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/14-s1-q10

关键思路

两部分的做法都一样：自由飞行阶段用能量守恒（conservation of energy），每次碰撞则用牛顿实验定律（Newton's experimental law）。第 (ii) 部分的窍门在于留意到：两球相撞的一刻，大球刚刚反向，而小球仍以全速下落。两者以很大的接近速率（closing speed）相遇，于是较重的下方球可以把较轻的上方球抛到远高于起始位置的高度。之后就只是一道含两条方程的碰撞题，而题目给出的数值令速率之比算得干净利落。

解答

把两球都模型化为沿铅垂方向运动的均匀刚体球（uniform rigid sphere），不计空气阻力，并以 g 表示重力加速度。每次碰撞都是瞬时的，所以碰撞期间重力不会改变任何动量（momentum）。

第 (i) 部分

当球心位于高度 R 时，球便接触地面，因此球已下落了距离 H。由能量守恒，球在碰撞前一刻的速率为

$$u = \sqrt{2gH}.$$

地面固定不动，所以由牛顿实验定律，反弹速率为向上的 eu。之后球上升至速率为零；再次由能量守恒，球再上升

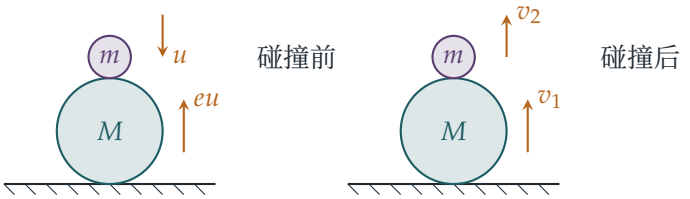
$$\frac{(eu)^2}{2g} = \frac{e^2 \cdot 2gH}{2g} = e^2 H.$$

球心在这段上升开始时位于高度 R，所以它到达离地高度 $R + He^2$，即为所求。

第 (ii) 部分

两球一同由静止开始自由下落，所以在每一时刻速率都相同。下方球的球心下落 H 后到达地面，因此在那一刻两球都以 $u = \sqrt{2gH}$ 向下运动。下方球反弹，如第 (i) 部分一样以向上速率 eu 离开地面；紧接着它与仍以 u 向下运动的上方球相撞。

取向上为正，并设两球相撞后一刻的速度分别为 v_1（质量 M 的球）和 v_2（质量 m 的球）。



由动量守恒（conservation of momentum）得

$$Meu - mu = Mv_1 + mv_2,$$

而由牛顿实验定律（分离速率（speed of separation）等于 e 乘以接近速率（speed of approach），其中接近速率为 $eu + u$）得

$$v_2 - v_1 = e(1+e)u.$$

把 $v_1 = v_2 - e(1+e)u$ 代入动量方程，

$$\begin{aligned}(M+m)v_2 &= Meu - mu + Me(1+e)u \\ &= \big(Me(2+e) - m\big)u,\end{aligned}$$

因此

$$v_2 = \frac{Me(2+e) - m}{M+m}\,u.$$

想一想 两球相撞时，小球的球心在哪里?

碰撞紧接在反弹之后发生，所以大球仍然接触地面：其球心位于高度 R，顶部位于 $2R$，而小球的球心位于 $2R + r$。小球由该处再上升 $v_2^2/(2g)$，因此

$$h = 2R + r + \frac{v_2^2}{2g}.$$

代入题目给出的数值，$2R + r = 0.45$，所以 $v_2^2/(2g) = 4.05$；又 $u^2/(2g) = H = 1.8$。两式相除，

$$\frac{v_2^2}{u^2} = \frac{4.05}{1.8} = \frac{9}{4}, \qquad \text{所以} \qquad v_2 = \tfrac{3}{2}u,$$

这里取正根，因为小球在碰撞后向上运动。

当 $e=\frac{2}{3}$ 时，$e(2+e)=\frac{2}{3}\cdot\frac{8}{3}=\frac{16}{9}$。设 $k=M/m$，并把分子和分母同除以 m，

$$\frac{\frac{16}{9}k-1}{k+1}=\frac{3}{2}\implies \tfrac{32}{9}k-2=3k+3\implies \tfrac{5}{9}k=5,$$

因此

$$\frac{M}{m}=9.$$

验算

- 当 $M/m=9$ 时，得 $v_2=\frac{3}{2}u$ 及 $v_1=v_2-\frac{10}{9}u=\frac{7}{18}u$。碰撞前的动量为 $9m\cdot\frac{2}{3}u-mu=5mu$，碰撞后为 $9m\cdot\frac{7}{18}u+m\cdot\frac{3}{2}u=5mu$。两者相符。
- 由于 $v_1<v_2$，两球在碰撞后分离，这是必然的。
- 以数值计算，$u=\sqrt{2\times 9.8\times 1.8}\approx 5.94\,\mathrm{m\,s^{-1}}$ 及 $v_2\approx 8.91\,\mathrm{m\,s^{-1}}$，以此速率可上升 $4.05\,\mathrm{m}$，得 $h=4.5$。
- 若 $M\gg m$，则 $v_2\to e(2+e)u$；当 $e=1$ 时为 $3u$，这正是完全弹性“叠球”下落中著名的高度九倍因子。

常见错误

- 只以 $v_2^2/(2g)$ 作为离地的最终高度，忘记碰撞发生时小球的球心位于 $2R+r$。
- 以 $e\cdot 2u$ 或 eu 作为分离速率：接近速率是 $eu+u$，因为两球正相向运动。
- 正负号错误：应固定向上为正，并把上方球在碰撞前的速度记为 $-u$。

STEP 之外

叠球下落是一个很好的示范，展示动量如何被“集中”传到一件较轻的物体上：若把多个球叠成一座塔，每个都比下面的一个轻，最顶的球可被抛到释放高度的许多倍。分析这类碰撞最简洁的方法是使用零动量参考系（zero-momentum frame，即质心参考系（centre-of-mass frame）），在这参考系中，每个球的速度只是反向并乘以 e。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对碰撞及质心参考系有细致的阐述。

详细解答

STEP I 2014，第 11 题

力学 · 连接滑轮系统

STEP I 2014，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/14-s1-q11

关键思路

对每个质点写出牛顿第二定律（Newton's second law），每条绳各用一个张力（tension），再由绳的约束条件（constraint）得出所有质点的加速度。在系统 II 中，关键在于滑轮（pulley）P_1 是轻的（light）：作用于它的合力必须为零（否则它的加速度会是无限大），所以上面那条绳的张力恰好是下面那条的两倍。确立了这一点之后，P_1 连同其下的两个质点对 M 的作用，就跟单一个质量为 4μ 的质点完全一样。

解答

全题中，质点视为点质量（point mass），绳视为轻而不可伸长（inextensible），滑轮视为光滑（smooth）且轻，因此每条绳上各处的张力相同。

第 (i) 部分

设绳的张力为 T。由于绳不可伸长，若 M 以 a_1 向下加速，则 m 以 a_1 向上加速。对每个质点应用牛顿第二定律，并以运动方向为正方向，得

$$M: \quad Mg - T = Ma_1,$$
$$m: \quad T - mg = ma_1.$$

两式相加，得 $(M-m)g = (M+m)a_1$，所以

$$a_1 = \frac{M-m}{M+m}\,g,$$

得证。代回 m 的方程，

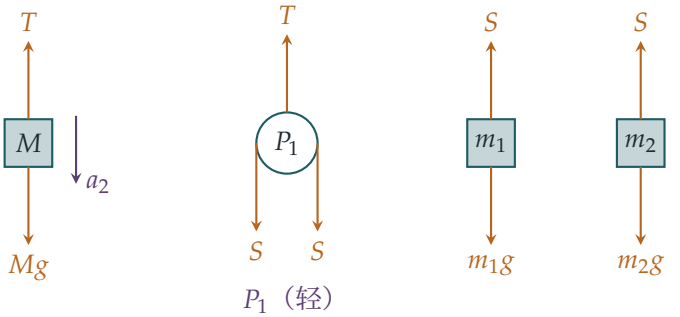
$$T = m(g+a_1) = m\cdot\frac{2M}{M+m}\,g = \frac{2Mmg}{M+m}.$$

绳的两段直线部分都从滑轮铅垂向下伸出，各以张力 T 拉动滑轮，所以张力对滑轮的作用力为

$$2T = \frac{4Mmg}{M+m} \quad ，方向铅垂向下。$$

第 (ii) 部分

设跨过固定滑轮 P 的绳的张力为 T，跨过 P_1 的绳的张力为 S。



轻滑轮。作用于 P_1 的力是向上的 T 和向下的 $2S$（它本身的重量为零）。由于它的质量为零，$T - 2S = 0 \cdot (\text{加速度}) = 0$，所以

$$T = 2S.$$

? **想一想** m_1 和 m_2 的加速度与 M 的加速度有甚么关系?

上面那条绳不可伸长，所以若 M 以 a_2 向下加速，则 P_1 以 a_2 向上加速。下面那条绳只决定 m_1 和 m_2 相对于 P_1 的运动：若 m_1 相对于 P_1 以 b 向下加速，则 m_2 相对于 P_1 以 b 向上加速。再加上 P_1 本身的运动，实际的加速度为

$$m_1 : b - a_2 \text{，向下，} \qquad m_2 : b + a_2 \text{，向上.}$$

这里 b 是第二个未知量，稍后会把它消去。

运动方程。

$$\begin{aligned} M: &\quad Mg - 2S = Ma_2, \\ m_1: &\quad m_1 g - S = m_1(b - a_2), \\ m_2: &\quad S - m_2 g = m_2(b + a_2). \end{aligned}$$

把 m_1 的方程除以 m_1，m_2 的方程除以 m_2：

$$g - \frac{S}{m_1} + a_2 = b = \frac{S}{m_2} - g - a_2.$$

因此

$$2(g+a_2) = S\left(\frac{1}{m_1}+\frac{1}{m_2}\right) = S\,\frac{m_1+m_2}{m_1m_2} = \frac{S}{\mu},$$

所以 $S = 2\mu(g+a_2)$，而 $T = 2S = 4\mu(g+a_2)$。代入 M 的方程，

$$Mg - 4\mu g - 4\mu a_2 = Ma_2 \quad\Longrightarrow\quad (M+4\mu)a_2 = (M-4\mu)g,$$

因此

$$a_2 = \frac{M-4\mu}{M+4\mu}\,g,$$

得证。与第 (i) 部分比较：M 的表现跟它与单一个质量为 4μ 的质点悬挂在一起时完全相同。

a_1 与 a_2 何时相等？ 对 $x>0$ 设 $f(x) = \dfrac{M-x}{M+x}$，于是 $a_1 = f(m)\,g$，$a_2 = f(4\mu)\,g$。由于

$$f(x) = \frac{2M}{M+x} - 1$$

且 $M>0$，f 在 $x>0$ 上严格递减（strictly decreasing），因此是一对一（one-to-one）的。所以 $a_1 = a_2$ 当且仅当 $m = 4\mu$。由 $m = m_1+m_2$，这变成

$$m_1+m_2 = \frac{4m_1m_2}{m_1+m_2} \iff (m_1+m_2)^2 = 4m_1m_2 \iff (m_1-m_2)^2 = 0 \iff m_1 = m_2,$$

其中第一步是乘以 $m_1+m_2>0$，这一步是可逆的。这样便完成了双向的证明。

验算

- 若 $m_1 = m_2$，系统 II 的下半部分平衡，所以 $b=0$：P_1 及其下的质点作为一个质量为 m_1+m_2 的整体运动，系统 II 便化为系统 I。这与此情况下 $4\mu = m_1+m_2$ 相符。
- 若 $m_2 \to 0$，则 $\mu \to 0$ 且 $a_2 \to g$：没有重物牵制 M，因为 m_1 只是自由下落，不承受张力。合理。
- 第 (ii) 部分的张力 $T = 4\mu(g+a_2)$ 也可以从第 (i) 部分得出：若 P_1 固定不动，作用于它的力为 $4m_1m_2g/(m_1+m_2) = 4\mu g$。以 a_2 向上加速，只是把 g 换成 $g+a_2$。
- 数值上，取 $M=3$、$m_1=1$、$m_2=2$，得 $\mu = \frac{2}{3}$ 及 $a_2 = \frac{1}{17}g \approx 0.576\ \mathrm{m\,s^{-2}}$（取 $g = 9.8\ \mathrm{m\,s^{-2}}$）。直接解四条方程也得出相同的值。

常见错误

- 忘记 P_1 会移动，把 m_1 和 m_2 的加速度写成 $\pm b$，而不是 $b-a_2$ 和 $b+a_2$。

- 在轻滑轮两边使用 $T = S$。P_1 每一边都以 S 向下拉，所以它上方的绳必须承受 $2S$。
- 推出 $a_1 = a_2 \Rightarrow m = 4\mu$ 却没有说明理由：这需要 f 是一对一的。

STEP 之外

这个装置是阿特伍德机(Atwood machine)之中再套一部阿特伍德机,而 $\mu = m_1 m_2/(m_1 + m_2)$ 就是约化质量（reduced mass）——当把行星与其恒星这类二体问题（two-body problem）化为单体问题时，出现的正是同一个量。带有这类约束的系统，用拉格朗日力学（Lagrangian mechanics）处理最为简洁：它直接使用独立坐标（这里是 P_1 的高度，以及 m_1 相对于 P_1 的位置），完全不需要求张力。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）仔细讲解了滑轮系统，并进而讨论二体问题中的约化质量。

详细解答

STEP I 2014，第 12 题

概率与统计 · 离散随机变量

STEP I 2014，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/14-s1-q12

关键思路

平方可以消去绝对值(modulus)符号,因此 $\mathrm{E}(X^2)$ 只需要用到骰子点数 S 的 $\mathrm{E}(S)$ 和 $\mathrm{E}(S^2)$，再对硬币的两种结果取平均。整数条件随即化为一个关于 $(k-1)^2$ 的整除（divisibility）问题，于是只剩下极少数个位数候选值。一旦知道 k，X 只取五个值，而两轮游戏只需简短地枚举（enumeration）所有数对。

解答

求 $\mathrm{E}(X^2)$

设 S 为骰子显示的点数，则

$$\mathrm{P}(S=1)=\tfrac{1}{2},\qquad \mathrm{P}(S=2)=\tfrac{1}{3},\qquad \mathrm{P}(S=3)=\tfrac{1}{6}.$$

于是

$$\mathrm{E}(S)=\tfrac{1}{2}+\tfrac{2}{3}+\tfrac{3}{6}=\tfrac{5}{3},\qquad \mathrm{E}(S^2)=\tfrac{1}{2}+\tfrac{4}{3}+\tfrac{9}{6}=\tfrac{10}{3}.$$

硬币是公平（fair）的，且与骰子独立（independent）。由于 $|a|^2=a^2$，在掷得正面的条件下 $X^2=(kS-1)^2$，在掷得反面的条件下 $X^2=(k-S)^2$。所以

$$\begin{aligned}\mathrm{E}(X^2)&=\tfrac{1}{2}\,\mathrm{E}\left[(kS-1)^2\right]+\tfrac{1}{2}\,\mathrm{E}\left[(k-S)^2\right]\\&=\tfrac{1}{2}\left[k^2\mathrm{E}(S^2)-2k\mathrm{E}(S)+1\right]+\tfrac{1}{2}\left[k^2-2k\mathrm{E}(S)+\mathrm{E}(S^2)\right]\\&=\tfrac{1}{2}\left[\tfrac{10}{3}k^2-\tfrac{10}{3}k+1+k^2-\tfrac{10}{3}k+\tfrac{10}{3}\right]\\&=\frac{13k^2-20k+13}{6}.\end{aligned}$$

另一方面，

$$k+\frac{13(k-1)^2}{6}=\frac{6k+13k^2-26k+13}{6}=\frac{13k^2-20k+13}{6},$$

所以 $\mathrm{E}(X^2)=k+\frac{13}{6}(k-1)^2$，正如所求。

求 k

由于 k 是整数，$\mathrm{E}(X^2)$ 是整数若且唯若 6 整除 $13(k-1)^2$。因为 13 与 6 互质（coprime），这等同于 $6\mid(k-1)^2$，即 2 和 3 都整除 $(k-1)^2$。由于 2 和 3 是质数（prime），它们各自必须整除 $k-1$ 本身，所以 $6\mid k-1$。在 $1\le k\le 9$ 的范围内，只剩下 $k=1$ 或 $k=7$。

> **想一想** 两个候选值都令 $\mathrm{E}(X^2)$ 为整数。哪个条件排除了其中一个?

是关于 $\mathrm{E}(X)$ 的条件。若 $k=1$，两条规则都给出 $X=|S-1|=S-1$，所以 $\mathrm{E}(X)=\frac{5}{3}-1=\frac{2}{3}$，并非整数。

若 $k=7$，则对每一面都有 $7S-1>0$ 及 $7-S>0$，因此可以去掉绝对值符号：

$$\mathrm{E}(X)=\tfrac{1}{2}\left(7\cdot\tfrac{5}{3}-1\right)+\tfrac{1}{2}\left(7-\tfrac{5}{3}\right)=\tfrac{16}{3}+\tfrac{8}{3}=8,$$

而 $\mathrm{E}(X^2)=7+\frac{13}{6}\cdot 36=85$。两者都是正整数，所以 $k=7$。

X 的分布

当 $k=7$ 时，正面给出 $X=7S-1\in\{6,13,20\}$，反面给出 $X=7-S\in\{6,5,4\}$。每个（硬币，骰子）组合的概率为 $\frac{1}{2}\mathrm{P}(S=s)$，而数值 6 既来自正面且 $S=1$，也来自反面且 $S=1$：

x	4	5	6	13	20
$\mathrm{P}(X=x)$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{12}$

各概率之和为 1。

两轮游戏

设 X_1、X_2 为两轮的结果。由于每一轮都重新掷骰子和硬币，它们互相独立，且各自服从上述分布（distribution）。我们需要找出和大于 25 及和恰好等于 25 的数对。

> **想一想** 怎样才能不遗漏地列出总和大于 25 的所有数对?

按较大的值分类。若两个结果都不是 20，最大可能总和为 $13+13=26$，而 $13+6=19$ 已经太小，所以 $(13,13)$ 是唯一这样的数对。若其中一个结果是 20，另一个必须大于 5，

即为 6、13 或 20；若它等于 5，总和恰好是 25。因此

$$\begin{aligned}\mathrm{P}(X_1+X_2>25) &= \mathrm{P}(13,13)+2\,\mathrm{P}(20,6)+2\,\mathrm{P}(20,13)+\mathrm{P}(20,20)\\ &= \tfrac{1}{36}+2\cdot\tfrac{1}{12}\cdot\tfrac{1}{2}+2\cdot\tfrac{1}{12}\cdot\tfrac{1}{6}+\tfrac{1}{144}\\ &= \frac{4+12+4+1}{144}=\frac{7}{48},\end{aligned}$$

其中因子 2 计算两种次序。同理

$$\mathrm{P}(X_1+X_2=25)=2\,\mathrm{P}(20,5)=2\cdot\tfrac{1}{12}\cdot\tfrac{1}{6}=\tfrac{1}{36}.$$

所以赌客预期获付的金额（以英镑计）为

$$\frac{7w}{48}+\frac{1}{36}.$$

当这个金额少于 £1 的赌注时，游戏对赌场有利：

$$\frac{7w}{48}+\frac{1}{36}<1 \iff \frac{7w}{48}<\frac{35}{36} \iff w<\frac{35\cdot 48}{36\cdot 7}=\frac{20}{3}.$$

由于 w 是整数，最大可能值为 $w=6$。又因为 $\frac{20}{3}$ 不是整数，无论是否容许公平游戏（预期获付恰好 £1），答案都相同。

验算

- 由表可得 $\mathrm{E}(X)=\frac{4+10+36+26+20}{12}=8$ 及 $\mathrm{E}(X^2)=\frac{16+50+216+338+400}{12}=85$，与公式吻合。
- 当 $w=6$ 时，预期获付为 $\frac{7}{8}+\frac{1}{36}=\frac{65}{72}\approx$ £0.90，低于赌注；当 $w=7$ 时为 $\frac{151}{144}\approx$ £1.05，高于赌注。
- 以电脑模拟 2×10^5 局两轮游戏，约 14.5% 的局总和大于 25（$\frac{7}{48}\approx 14.6\%$），约 2.8% 的局总和恰好是 25（$\frac{1}{36}\approx 2.8\%$）。

常见错误

- 在分布中把 $X=6$ 列出两次，一次对应正面、一次对应反面，而没有把它们合并为 $\mathrm{P}(X=6)=\frac{1}{2}$。
- 对于像 $(20,6)$ 这样两数不相等的数对，忘记乘以代表两种次序的因子 2。
- 把“总和等于 25”当作赢得 £w 或当作输：赌客只是取回赌注。

STEP 之外

计算 $\mathrm{E}(X^2)$ 的过程是全期望定律（law of total expectation）$\mathrm{E}(Y) = \sum_i \mathrm{P}(A_i)\,\mathrm{E}(Y \mid A_i)$ 的一个例子，大学一年级的概率课程经常用到它。两轮游戏涉及独立随机变量（random variables）之和的分布，即卷积（convolution），而生成函数（generating functions）可以简洁地处理这类问题。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）涵盖了这两个概念。

详细解答

STEP I 2014，第 13 题

概率与统计 · 三角分布

STEP I 2014，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：

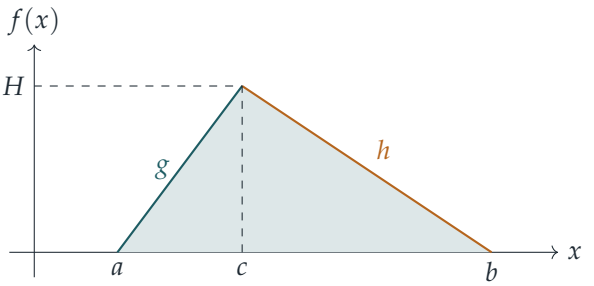
step.maths.org/questions/14-s1-q13

关键思路

密度函数（density function）的图像是一个三角形，底边由 a 延伸至 b，顶点（apex）位于 c 的正上方，而三角形的面积必须等于 1。这就确定了顶点的高度，之后每段线性部分都由两点完全确定。计算平均值（mean）和中位数（median）时，先平移变量，使每段线性部分都从零开始，积分就会保持简洁；至于中位数，首要问题只是它落在 c 的哪一边。

解答

全文假设 $a < c < b$，使两段都是真正存在的（题目的公式要有意义，$c - a$ 与 $b - c$ 必须不等于零）。



两段线性部分

设 $H = g(c) = h(c)$ 为顶点的高度。f 的图像下方的区域是一个底为 $b-a$、高为 H 的三角形，而总概率为 1，所以

$$\tfrac{1}{2}(b-a)H = 1, \qquad H = \frac{2}{b-a}.$$

由于 g 是线性函数且 $g(a) = 0$，故对某常数 k 有 $g(x) = k(x-a)$，而 $g(c) = H$ 给出 $k(c-a) = \dfrac{2}{b-a}$。因此

$$g(x) = \frac{2(x-a)}{(b-a)(c-a)},$$

得证。同理，$h(x) = m(b-x)$ 且 $h(c) = H$，所以

$$h(x) = \frac{2(b-x)}{(b-a)(b-c)}.$$

由于 $b-a$、$c-a$ 和 $b-c$ 全部为正，g 是递增的而 h 是递减的，符合要求。

第 (i) 部分

记 $\mathrm{E}(X) = I_1 + I_2$，其中 $I_1 = \int_a^c x\,g(x)\,\mathrm{d}x$，$I_2 = \int_c^b x\,h(x)\,\mathrm{d}x$。

在 I_1 中代入 $x = a + t$，于是 t 由 0 变到 $c-a$：

$$\begin{aligned} I_1 &= \frac{2}{(b-a)(c-a)} \int_0^{c-a} (a+t)\,t\,\mathrm{d}t = \frac{2}{(b-a)(c-a)} \left[\frac{a(c-a)^2}{2} + \frac{(c-a)^3}{3}\right] \\ &= \frac{(c-a)\big(3a + 2(c-a)\big)}{3(b-a)} = \frac{(c-a)(a+2c)}{3(b-a)}. \end{aligned}$$

在 I_2 中代入 $x = b - s$，于是 s 由 0 变到 $b-c$：

$$\begin{aligned} I_2 &= \frac{2}{(b-a)(b-c)} \int_0^{b-c} (b-s)\,s\,\mathrm{d}s = \frac{2}{(b-a)(b-c)} \left[\frac{b(b-c)^2}{2} - \frac{(b-c)^3}{3}\right] \\ &= \frac{(b-c)\big(3b - 2(b-c)\big)}{3(b-a)} = \frac{(b-c)(b+2c)}{3(b-a)}. \end{aligned}$$

两者相加，分子为

$$\begin{aligned} (c-a)(a+2c) + (b-c)(b+2c) &= \big(ac + 2c^2 - a^2 - 2ac\big) + \big(b^2 + 2bc - bc - 2c^2\big) \\ &= b^2 - a^2 + bc - ac \\ &= (b-a)(b+a+c). \end{aligned}$$

除以 $3(b-a)$，得

$$\mathrm{E}(X) = \tfrac{1}{3}(a+b+c),$$

得证。（这正是三角形形心（centroid）的 x 坐标，而这并非巧合：平均值就是密度函数下方区域的平衡点。）

第 (ii) 部分

首先求出顶点左方有多少概率。这就是左边那个三角形的面积：

$$\mathrm{P}(X \leq c) = \tfrac{1}{2}(c-a)H = \frac{c-a}{b-a}.$$

? **想一想** 我们怎样知道中位数在 c 的左边还是右边?

中位数 M 满足 $\mathrm{P}(X \leq M) = \frac{1}{2}$。若 $\mathrm{P}(X \leq c) \geq \frac{1}{2}$，中位数就在 c 处或其左方；否则就在其右方。而 $\frac{c-a}{b-a} \geq \frac{1}{2}$ 当且仅当 $c \geq \frac{1}{2}(a+b)$，即顶点位于 $[a,b]$ 的中点处或其右方。所以分情况的依据，就是比较 c 与 $\frac{1}{2}(a+b)$。

情况 $c \geq \frac{1}{2}(a+b)$。此时 $a < M \leq c$，而对于 $a \leq m \leq c$，由 a 到 m 的小三角形面积为

$$\mathrm{P}(X \leq m) = \int_a^m g(x)\,\mathrm{d}x = \frac{(m-a)^2}{(b-a)(c-a)}.$$

令它等于 $\frac{1}{2}$，并取正根（因为 $M > a$），得

$$M = a + \sqrt{\tfrac{1}{2}(b-a)(c-a)}.$$

情况 $c < \frac{1}{2}(a+b)$。这时中位数在 c 的右方，从右端入手会比较容易。对于 $c \leq m \leq b$，

$$\mathrm{P}(X \geq m) = \int_m^b h(x)\,\mathrm{d}x = \frac{(b-m)^2}{(b-a)(b-c)}.$$

令它等于 $\frac{1}{2}$，并利用 $M < b$，得

$$M = b - \sqrt{\tfrac{1}{2}(b-a)(b-c)}.$$

当 $c = \frac{1}{2}(a+b)$ 时，分布是对称的，两条公式都给出 $M = \frac{1}{2}(a+b)$（例如第一条给出 $a + \sqrt{\frac{1}{4}(b-a)^2} = a + \frac{1}{2}(b-a)$），所以两种情况能够衔接起来。

验算

- 取 $a=0$、$b=4$、$c=1$：$\mathrm{P}(X \leq c) = \frac{1}{4} < \frac{1}{2}$，所以属于第二种情况，$M = 4 - \sqrt{6} \approx 1.5505$。平均值为 $\frac{5}{3}$。用电脑模拟 200 000 个数值，得到样本中位数 1.5505 及样本平均值 1.668。
- 取 $a=0$、$b=4$、$c=3$：适用第一种情况，$M = \sqrt{6} \approx 2.449$，正好是上一个例子的镜像，理应如此。
- 中位数总是介乎众数（mode）c 与中点 $\frac{1}{2}(a+b)$ 之间。例如在第一种情况中 $M \leq c$，而由于当 $c \geq \frac{1}{2}(a+b)$ 时 $(b-a)(c-a) \geq \frac{1}{2}(b-a)^2$，所以 $M \geq \frac{1}{2}(a+b)$。

常见错误

- 求中位数时用错了密度函数的那一段：应先确定中位数落在 c 的哪一边。
- 取了负的平方根，令 M 落在 $[a, b]$ 之外。
- 不作平移，直接以 x 展开平均值的积分：这样做也可以，但代数运算繁重得多，容易出错。

STEP 之外

当 c 是中点时，三角分布（triangular distribution）就是两个独立均匀随机变量（uniform random variable）之和的分布；而在只知道最小值、最大值和最可能值的情况下（例如项目规划），三角分布亦广泛用作粗略的模型。计算独立随机变量之和的分布会引出卷积（convolution），这是大学一年级概率论的核心工具。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）是阅读这方面内容的好起点。

详细解答

STEP II 2014，第 1 题

纯数学 · 三角形几何与二次方程

STEP II 2014，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/14-s2-q1

关键思路

沿着锯齿形路线 $A \to P \to Q \to B$ 由 A 走到 B。三段路的长度都是 x，合起来必须刚好走过水平距离 1，而净上升高度为 0。分别沿水平及铅直方向分解（resolving），得到 $x\cos\theta$ 和 $x\sin\theta$，再利用 $\cos^2\theta + \sin^2\theta = 1$ 消去未知角。之后的一切，就只是一条关于 x 的二次方程（quadratic equation）及其判别式（discriminant）。

解答

把 A 放在原点，B 放在 $(1,0)$，而 C 在 AB 上方。由于 P 位于由 A 出发经过 C 的射线（ray）上，与 A 的距离为 x；而 Q 位于由 B 出发经过 C 的射线上，与 B 的距离为 x，所以

$$P = (x\cos\alpha,\ x\sin\alpha), \qquad Q = (1 - x\cos\beta,\ x\sin\beta).$$

第 (i) 部分

水平方向。$\overrightarrow{AP}$、$\overrightarrow{PQ}$ 和 $\overrightarrow{QB}$ 的水平分量（component）分别为 $x\cos\alpha$、$x\cos\theta$ 和 $x\cos\beta$，它们之和等于 $AB = 1$。因此

$$x\cos\theta = 1 - x\cos\alpha - x\cos\beta,$$

这就是第一个结果。（以坐标来说：这是 Q 与 P 的 x 坐标之差。）

铅直方向。P 的高度为 $x\sin\alpha$，Q 的高度为 $x\sin\beta$，所以

$$x\sin\theta = x\sin\beta - x\sin\alpha.$$

（这个值永不为负：由于三角形的内角和小于 180°，有 $\alpha \le \beta < 180^\circ - \alpha$，而在这个区间上 $\sin\beta \ge \sin\alpha$。所以 Q 永远不会低于 P。）

消去 θ。把两式平方后相加，并利用 $\cos^2\theta+\sin^2\theta=1$，得

$$x^2=\big(1-x(\cos\alpha+\cos\beta)\big)^2+x^2(\sin\beta-\sin\alpha)^2.$$

展开右方，x^2 的系数为

$$\begin{aligned}(\cos\alpha+\cos\beta)^2+(\sin\beta-\sin\alpha)^2&=(\cos^2\alpha+\sin^2\alpha)+(\cos^2\beta+\sin^2\beta)\\&\quad+2(\cos\alpha\cos\beta-\sin\alpha\sin\beta)\\&=2+2\cos(\alpha+\beta).\end{aligned}$$

所以 $x^2=1-2(\cos\alpha+\cos\beta)x+\big(2+2\cos(\alpha+\beta)\big)x^2$。把 x^2 移到右方，得

$$\big(1+2\cos(\alpha+\beta)\big)x^2-2(\cos\alpha+\cos\beta)x+1=0, \qquad (*)$$

此即所求。

P 和 Q 在两边的延长线上。若 P 位于 AC 的延长线（produced）上，它仍然在由 A 出发经过 C 的射线上，与 A 的距离为 x，所以其坐标仍是 $(x\cos\alpha, x\sin\alpha)$。同理，$Q$ 仍是 $(1-x\cos\beta, x\sin\beta)$。上面的推导并没有用到其他东西：由 P 到 Q 的位移为

$$\big(1-x\cos\alpha-x\cos\beta,\ x\sin\beta-x\sin\alpha\big),$$

而 $PQ^2=x^2$ 给出完全相同的方程 $(*)$。(此时 x 可以大于 AC 和 BC，而 $1-x\cos\alpha-x\cos\beta$ 可以是负数，所以 θ 可以是钝角。这并无影响，因为平方后正负号便消失了。)

第 (ii) 部分

方程 $(*)$ 是线性方程，当且仅当 $1+2\cos(\alpha+\beta)=0$，即 $\cos(\alpha+\beta)=-\frac{1}{2}$。由于 $0<\alpha+\beta<180^\circ$，唯一的可能是

$$\alpha+\beta=120^\circ.$$

否则 $(*)$ 是一条真正的二次方程，其判别式的四分之一为

$$\begin{aligned}\tfrac{1}{4}\Delta&=(\cos\alpha+\cos\beta)^2-\big(1+2\cos\alpha\cos\beta-2\sin\alpha\sin\beta\big)\\&=\cos^2\alpha+\cos^2\beta-1+2\sin\alpha\sin\beta\\&=1-\sin^2\alpha-\sin^2\beta+2\sin\alpha\sin\beta\\&=1-(\sin\beta-\sin\alpha)^2,\end{aligned}$$

这里用了 $\cos^2\alpha=1-\sin^2\alpha$ 及 $\cos^2\beta-1=-\sin^2\beta$。

> ? **想一想** 为甚么 $(\sin\beta-\sin\alpha)^2$ 必定小于 1?

两个角都严格介乎 0° 与 180° 之间，所以 $0<\sin\alpha\le 1$ 及 $0<\sin\beta\le 1$。两个属于 $(0,1]$ 的数之差严格介乎 -1 与 1 之间。因此 $(\sin\beta-\sin\alpha)^2<1$，故 $\Delta>0$，而 $(*)$ 有两个相异实根。

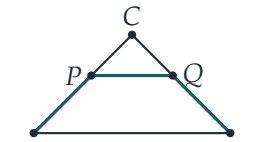
第 (iii) 部分

(a) $\alpha=\beta=45^\circ$。此时 $\cos(\alpha+\beta)=\cos 90^\circ=0$ 且 $\cos\alpha+\cos\beta=\sqrt{2}$，所以 $(*)$ 为

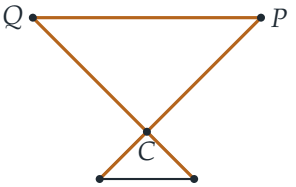
$$x^2-2\sqrt{2}\,x+1=0,\qquad x=\sqrt{2}\pm 1.$$

这个三角形是等腰直角三角形，$C=(\frac{1}{2},\frac{1}{2})$，且 $AC=BC=\frac{1}{\sqrt{2}}\approx 0.707$。

- $x=\sqrt{2}-1\approx 0.414<AC$：$P$ 和 Q 位于两边上，分别在 $P=(1-\frac{1}{\sqrt{2}},1-\frac{1}{\sqrt{2}})$ 和 $Q=(\frac{1}{\sqrt{2}},1-\frac{1}{\sqrt{2}})$。由对称性，$PQ$ 与 AB 平行，所以 $\theta=0$。验证：$PQ=\sqrt{2}-1=x$。
- $x=\sqrt{2}+1\approx 2.414>AC$：$P$ 和 Q 分别位于 AC 的延长线和 BC 的延长线上，在 $P=(1+\frac{1}{\sqrt{2}},1+\frac{1}{\sqrt{2}})$ 和 $Q=(-\frac{1}{\sqrt{2}},1+\frac{1}{\sqrt{2}})$。两条延长线已在 C 点相交而过，所以现在 P 在 Q 的*右方*。PQ 同样是水平的，长度为 $1+\sqrt{2}=x$。



$x=\sqrt{2}-1$

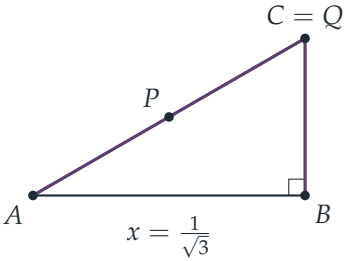


$x=\sqrt{2}+1$

(b) $\alpha=30^\circ$，$\beta=90^\circ$。现在 $\alpha+\beta=120^\circ$，所以这正是第 (ii) 部分中的线性情况。由于 $\cos\alpha+\cos\beta=\frac{\sqrt{3}}{2}$，$(*)$ 变成

$$-\sqrt{3}\,x+1=0,\qquad x=\frac{1}{\sqrt{3}}.$$

这个三角形的直角在 B，$BC=\tan 30^\circ=\frac{1}{\sqrt{3}}$，而 $AC=\frac{2}{\sqrt{3}}$。所以 $QB=BC$ 表示 Q 就是 C，而 $AP=\frac{1}{\sqrt{3}}=\frac{1}{2}AC$ 使 P 位于 AC 的中点。于是 $PQ=PC=\frac{1}{\sqrt{3}}=x$，正如所料。这是唯一的可能。



$x=\frac{1}{\sqrt{3}}$

第二个根去了哪里？若 $\alpha+\beta=120^\circ$，直线 AC 和 BC 相交成 60°。当 x 非常大时，沿这两条直线远处的点 P 和 Q 之间的距离便几乎刚好是 x，与 C 组成一个近乎等边的三角形。当 $\alpha+\beta\to 120^\circ$ 时，二次方程的其中一个根便趋向无穷大。

验算

- 在 (a) 中，$(\sqrt{2}-1)(\sqrt{2}+1)=1$，与两根之积 $1/(1+2\cos 90^\circ)=1$ 吻合。
- 对任意一个三角形，例如 $\alpha=40^\circ$、$\beta=70^\circ$，解 $(*)$，定出 P 和 Q 的位置并量度 PQ：对两个根而言，它都等于 x。我们已用电脑对许多随机三角形验证过这一点，误差约为 10^{-10}。
- 在 (b) 中，$x=1/\sqrt{3}\approx 0.577$ 刚好等于 BC，而 AC 的中点为 $(\frac{1}{2},\frac{1}{2\sqrt{3}})$，与 $C=(1,\frac{1}{\sqrt{3}})$ 的距离为 $\frac{1}{\sqrt{3}}$。

常见错误

- 把 $\cos\beta$ 项的正负号弄错：$\overrightarrow{QB}$ 是朝向 B 的，所以即使 Q 在 B 的左方，其水平分量仍是 $+x\cos\beta$。
- 在第 (ii) 部分中容许 $\alpha+\beta=240^\circ$。三角形的内角和小于 180°。
- 在 (b) 中，首项系数为零时仍试图使用二次公式，或断定方程无解。

STEP 之外

把一个封闭的向量多边形分解成分量，再用 $\cos^2+\sin^2=1$ 消去一个角，是运动学（kinematics）和连杆（linkage）问题中的标准技巧：我们的锯齿形路线就是一个简单的三杆连杆。当二次方程退化为线性方程时“逃往无穷远”的那个根，让我们初次窥见射影几何（projective geometry）。在射影几何中，平行线相交于无穷远点，而只要把这些点也计算在内，二次方程总有两个根。Coxeter 与 Greitzer 合著的 *Geometry Revisited*（《几何再探》）是在这个程度上踏入古典几何的绝佳入门书。

详细解答

STEP II 2014，第 2 题

纯数学 · 积分不等式

> **STEP II 2014，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/14-s2-q2

本题在 *Beyond STEP* 第 3.7 节 $\pi^2 \le 10$ 中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch03/wirtinger.html

关键思路

$(*)$ 两边都是积分，对任何明确给出的 f 都可以精确计算，所以每一个在两端都等于零的函数，都会把 $(*)$ 变成一条涉及 π 的数值不等式。真正需要选择的，只是怎样取自由常数，使 $f(0) = f(\pi) = 0$。若函数在某一端不等于零，它可以数值很大而变化很慢，这正是端点条件（end conditions）重要的原因。

解答

第 (i) 部分

取 $f(x) = \sin nx$，则 $f'(x) = n\cos nx$。利用 $\sin^2\theta = \frac{1}{2}(1 - \cos 2\theta)$ 及 $\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$，

$$\int_0^{\pi} \sin^2 nx \,\mathrm{d}x = \tfrac{1}{2}\Big[x - \tfrac{1}{2n}\sin 2nx\Big]_0^{\pi} = \frac{\pi}{2},$$
$$\int_0^{\pi} n^2\cos^2 nx \,\mathrm{d}x = \tfrac{n^2}{2}\Big[x + \tfrac{1}{2n}\sin 2nx\Big]_0^{\pi} = \frac{n^2\pi}{2},$$

这是因为对整数 n 有 $\sin 2n\pi = 0$。由于 $n \ge 1$，故 $n^2 \ge 1$，于是 $\frac{\pi}{2} \le \frac{n^2\pi}{2}$，$(*)$ 成立（当且仅当 $n = 1$ 时取等号）。

反例（counterexamples）。取 $f(x) = \pi - x$，它满足 $f(\pi) = 0$，但 $f(0) = \pi \ne 0$。于是

$f'(x) = -1$，而

$$\int_0^\pi (\pi - x)^2 \,\mathrm{d}x = \left[-\tfrac{1}{3}(\pi - x)^3\right]_0^\pi = \frac{\pi^3}{3}, \qquad \int_0^\pi 1 \,\mathrm{d}x = \pi.$$

由于 $\pi^2 > 3$，得 $\frac{\pi^3}{3} > \pi$，所以 $(*)$ 不成立。同理，$f(x) = x$ 满足 $f(0) = 0$，但 $f(\pi) = \pi \neq 0$，它给出相同的两个积分，$\frac{\pi^3}{3} > \pi$，所以 $(*)$ 再次不成立。因此，即使另一个端点条件成立，两个端点条件都不能省去。

第 (ii) 部分

二次函数 (quadratic)。我们需要 $f(0) = c = 0$ 及 $f(\pi) = a\pi^2 + b\pi = 0$，所以 $b = -a\pi$，且 $f(x) = ax(x - \pi)$。若取 $a = 0$，则 $f \equiv 0$，只得到毫无用处的不等式 $0 \leq 0$，所以取 $a \neq 0$。当 f 乘以 a 时，$(*)$ 两边都乘以 a^2，所以不妨取 $a = -1$：

$$f(x) = x(\pi - x), \qquad f'(x) = \pi - 2x.$$

代入 $x = \frac{\pi}{2} + u$，可使两个积分都变得对称：$f = \frac{\pi^2}{4} - u^2$，$f' = -2u$，而 u 的范围是 $[-\frac{\pi}{2}, \frac{\pi}{2}]$。记 $h = \frac{\pi}{2}$，

$$\int_0^\pi f^2 \,\mathrm{d}x = \int_{-h}^h \left(h^2 - u^2\right)^2 \mathrm{d}u = 2\left(h^5 - \tfrac{2}{3}h^5 + \tfrac{1}{5}h^5\right) = \frac{16h^5}{15} = \frac{\pi^5}{30},$$
$$\int_0^\pi f'^2 \,\mathrm{d}x = \int_{-h}^h 4u^2 \,\mathrm{d}u = \frac{8h^3}{3} = \frac{\pi^3}{3}.$$

所以 $(*)$ 给出 $\dfrac{\pi^5}{30} \leq \dfrac{\pi^3}{3}$，两边除以正数 $\frac{\pi^3}{30}$，得

$$\pi^2 \leq 10,$$

即为所求。

三角函数 (trigonometric function)。现在 $f(0) = q + r$，而 $f(\pi) = p \sin\frac{\pi}{2} + q\cos\frac{\pi}{2} + r = p + r$。两者同时为零，当且仅当 $p = q = -r$。同样地，$r = 0$ 会给出 $f \equiv 0$，而缩放并无影响，所以取 $r = -1$：

$$f(x) = \sin\tfrac{1}{2}x + \cos\tfrac{1}{2}x - 1, \qquad f'(x) = \tfrac{1}{2}\left(\cos\tfrac{1}{2}x - \sin\tfrac{1}{2}x\right).$$

? **想一想** 怎样把 f 平方，而不弄出一大堆交叉项（cross terms）?

把它组合成 $(s + c) - 1$，其中 $s = \sin\frac{1}{2}x$，$c = \cos\frac{1}{2}x$。于是 $(s + c)^2 = 1 + 2sc = 1 + \sin x$，所以

$$f^2 = (s + c)^2 - 2(s + c) + 1 = 2 + \sin x - 2\sin\tfrac{1}{2}x - 2\cos\tfrac{1}{2}x.$$

利用 $\int_0^\pi \sin x\,\mathrm{d}x = 2$、$\int_0^\pi \sin\frac{1}{2}x\,\mathrm{d}x = \left[-2\cos\frac{1}{2}x\right]_0^\pi = 2$ 及 $\int_0^\pi \cos\frac{1}{2}x\,\mathrm{d}x = \left[2\sin\frac{1}{2}x\right]_0^\pi = 2$,

$$\int_0^\pi f^2\,\mathrm{d}x = 2\pi + 2 - 4 - 4 = 2\pi - 6.$$

同样地，$(c-s)^2 = 1 - \sin x$，所以

$$\int_0^\pi f'^2\,\mathrm{d}x = \frac{1}{4}\int_0^\pi (1-\sin x)\,\mathrm{d}x = \frac{\pi-2}{4}.$$

于是 $(*)$ 表示 $2\pi - 6 \le \frac{1}{4}(\pi-2)$，即 $8\pi - 24 \le \pi - 2$，所以

$$7\pi \le 22, \qquad \text{即} \qquad \pi \le \frac{22}{7}.$$

比较。把第二条不等式平方（两边皆为正数），得

$$\pi^2 \le \frac{484}{49} < \frac{490}{49} = 10.$$

所以三角函数的选择给出较小的上界（upper bound）$\pi^2 \le \frac{484}{49} \approx 9.878$，因此是较好的估计。两者都是上界，所以较小的一个更接近真实值。

验算

- 真实值为 $\pi^2 \approx 9.8696$，低于 10 及 $\frac{484}{49} \approx 9.8776$，这是必然的。
- 以数值计算，对于三角函数 f，$(*)$ 两边分别为 $2\pi - 6 \approx 0.2832$ 及 $\frac{1}{4}(\pi-2) \approx 0.2854$：两者非常接近，所以界 $\pi \le \frac{22}{7}$ 才会如此精确。对于二次函数，两边分别为 $\frac{\pi^5}{30} \approx 10.20$ 及 $\frac{\pi^3}{3} \approx 10.34$。
- 在第 (i) 部分中，$n = 1$ 时取等号，所以 $\sin x$ 是 $(*)$ 的极端情况。三角函数 f 的形状与 $\sin x$ 很接近，而 $x(\pi - x)$ 则稍逊一点，这解释了为何某个界较好。

常见错误

- 选取像 $f = 1$ 这样同时破坏两个端点条件的反例。这样做是正确的，但 $\pi - x$ 和 x 说明得更多：每个条件本身都是必需的。
- 在三角函数的情况下计算 f^2 的积分时，忘记 $\int_0^\pi \sin x\,\mathrm{d}x = 2$（而不是 0）。
- 因为 10 是个“较整齐”的数，便说它是较好的估计。两者都是上界，所以较小的一个才较好。

STEP 之外

不等式 $(*)$ 就是 Wirtinger 不等式（Wirtinger's inequality），而最佳常数 1 来自 $\sin x$：它是在 $f(0) = f(\pi) = 0$ 条件下 $-f'' = \lambda f$ 的最小特征值（eigenvalue）。以选定的函数代入 $(*)$ 来测试，就是瑞利–里兹方法（Rayleigh–Ritz method），它总是给出该特征值的上界，也是工程师估算振动频率的基础。傅里叶级数（Fourier series）则能给出 $(*)$ 本身的一个简洁证明。有关背后的分析，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）；有关特征值方面，可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）。

详细解答

STEP II 2014，第 3 题

纯数学 · 切线与包络线

STEP II 2014，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/14-s2-q3

关键思路

在 P 点的切线（tangent）其实只是一条直线，斜率（gradient）为 y'，y 截距（y-intercept）为 $y - xy'$，因此第 (i) 部分把“距离恒定”这个性质化为一条微分方程（differential equation）。把这条方程微分后会出现因子 y''，而令乘积为零的两种方式，正好给出两族曲线：直线（$y'' = 0$）和一个圆。第 (iii) 部分则要求我们把这些曲线的片段平滑地拼接起来。

解答

第 (i) 部分

若 $m = 0$，直线为 $y = c$，距离显然是 c，所以设 $m \neq 0$。直线与 y 轴交于 $A = (0, c)$，并与 x 轴成锐角（acute angle）θ，其中 $\tan\theta = |m|$。设 N 为由 O 到该直线的垂足（foot of the perpendicular）。在直角三角形 ONA 中，直线与 y 轴在 A 点所成的角为 $\frac{\pi}{2} - \theta$，所以角 AON 为 θ，于是

$$ON = OA\cos\theta = \frac{c}{\sqrt{1+\tan^2\theta}} = \frac{c}{\sqrt{1+m^2}} = c(m^2+1)^{-\frac{1}{2}}.$$

（求 $x^2 + (mx + c)^2$ 的最小值也会得到同一结果：这是 x 的二次式（quadratic），在 $x = -\frac{mc}{1+m^2}$ 时取最小值 $\frac{c^2}{1+m^2}$。）

若 $c < 0$，以原点作反射（reflection），直线会变成 $y = mx + |c|$，而与 O 的距离不变，所以一般而言距离为 $|c|(1+m^2)^{-\frac{1}{2}}$。

第 (ii) 部分

在 $P=(x,y)$ 点，切线的斜率为 y' 并经过 P，所以以坐标 (X,Y) 表示，其方程为

$$Y = y'X + (y - xy').$$

这是第 (i) 部分那种形式的直线，其中 $m=y'$、$c=y-xy'$（按假设它不是铅垂线）。因此它与 O 的距离为

$$a = \frac{|y-xy'|}{\sqrt{1+(y')^2}},$$

两边平方得

$$(y-xy')^2 = a^2\big(1+(y')^2\big). \qquad (*)$$

微分。这里 a 是常数，而 $\frac{\mathrm{d}}{\mathrm{d}x}(y-xy') = y' - y' - xy'' = -xy''$。所以把 $(*)$ 微分得

$$2(y-xy')(-xy'') = 2a^2y'y'', \qquad \text{即} \qquad y''\big[\,x(y-xy') + a^2y'\,\big] = 0.$$

? **想一想** 乘积为零，怎样变成曲线的两种不同情况?

在每一点，要么 $y''=0$，要么方括号内的式子为零。在曲线上 $y''=0$ 的一段，曲线是直的；在 $y''\neq 0$ 的一段，方括号必须为零。我们逐一处理这两种情况。（我们假设 $a>0$；若 $a=0$，每条切线都经过 O，C 就只是一条经过原点的直线。）

情况一：$y''=0$。此时 $y'=m$ 为常数，而 $y=mx+k$，其中 m、k 为常数。代入 $(*)$（此时 $y-xy'=k$），得 $k^2=a^2(1+m^2)$，所以

$$y = mx \pm a(1+m^2)^{\frac{1}{2}}.$$

情况二：方括号为零。此时 $xy - x^2y' + a^2y' = 0$，所以

$$(x^2-a^2)\,y' = xy.$$

首先，$x^2\neq a^2$：若 $x=\pm a$，则 $xy=0$ 迫使 $y=0$，而 $(*)$ 变成 $a^2(y')^2 = a^2\big(1+(y')^2\big)$，这是不可能的。所以我们可以相除：

$$y' = \frac{xy}{x^2-a^2}, \qquad y - xy' = \frac{y(x^2-a^2) - x^2y}{x^2-a^2} = \frac{-a^2y}{x^2-a^2}.$$

把两者代入 $(*)$，再全式乘以 $(x^2-a^2)^2/a^2$：

$$\begin{aligned} & a^2y^2 = (x^2-a^2)^2 + x^2y^2 \\ \iff \quad & y^2(a^2-x^2) = (a^2-x^2)^2 \\ \iff \quad & (a^2-x^2)\,(x^2+y^2-a^2) = 0. \end{aligned}$$

由于 $a^2 - x^2 \neq 0$，我们得出 $x^2 + y^2 = a^2$。

所以，要么对某个 m 有 $y = mx \pm a(1+m^2)^{\frac{1}{2}}$，要么 $x^2 + y^2 = a^2$，正如所求。两者确实都具有该性质：由第 (i) 部分，这些直线与 O 的距离为 a；而以 O 为圆心、半径为 a 的圆，其每条切线与 O 的距离都是 a。

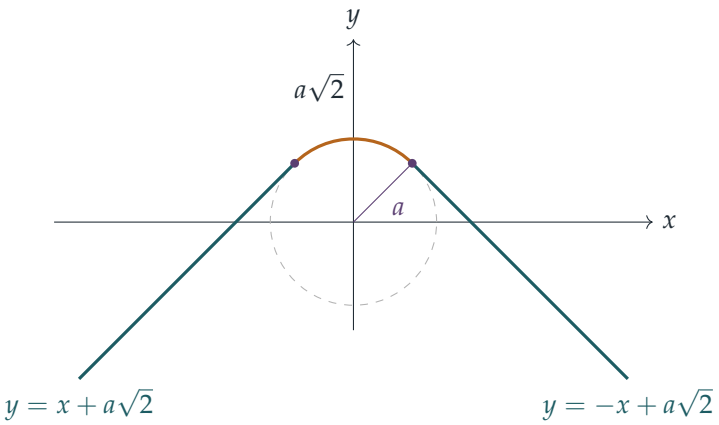
第 (iii) 部分

在直线段与圆弧相接之处，C 必须有切线，所以直线在该点的斜率必须与圆弧相同：每条直线都是圆在圆弧端点处的切线。此外，每条切线都必须不是铅垂线，所以圆弧不能到达 $(\pm a, 0)$ 两点；又因为 C 是 x 的函数（function）的图像，它只能用圆的上半部或只用下半部。

一个例子：取 $x^2 + y^2 = a^2$ 上由 $\left(-\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right)$ 经顶部到 $\left(\frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}\right)$ 的圆弧，两端再以该处的切线延伸：

$$y = \begin{cases} x + a\sqrt{2}, & x \leq -\frac{a}{\sqrt{2}}, \\ \sqrt{a^2 - x^2}, & |x| \leq \frac{a}{\sqrt{2}}, \\ -x + a\sqrt{2}, & x \geq \frac{a}{\sqrt{2}}. \end{cases}$$

这两条直线的 $m = \pm 1$，截距为 $a\sqrt{2} = a(1+m^2)^{\frac{1}{2}}$，符合情况一的要求。在 $x = \pm\frac{a}{\sqrt{2}}$ 处，圆弧的斜率为 $-\frac{x}{y} = \mp 1$，与直线吻合，所以斜率处处连续（continuous）且有限。



上半圆（或下半圆）上任何不包含 $(\pm a, 0)$ 的圆弧，两端以其切线延伸，都同样可行。

验算

- 在第 (i) 部分，$m = 0$ 给出距离 c；当 $|m| \to \infty$，直线变得陡峭，距离趋于 0，这对经过 $(0, c)$ 的直线来说理应如此。

- 在圆上，$y' = -x/y$ 给出 $y - xy' = a^2/y$，而 $(*)$ 两边都等于 a^4/y^2。
- 对于 $a = 1$ 的例子，我们用电脑在 $x = -5$ 至 $x = 5$ 之间的 1001 个点检查了 O 到切线的距离，每次都是 1；例如在 $x = 2$，切线为 $y = -x + \sqrt{2}$，距离为 $\sqrt{2}/\sqrt{2} = 1$。

常见错误

- 把微分后的方程除以 y''，因而遗漏了直线解。
- 忘记 a 是常数，或把 xy' 微分成 y'，而不是 $y' + xy''$。
- 在第 (iii) 部分，直线与圆弧在接驳处形成尖角，或用了整个半圆，而半圆两端的切线是铅垂的。

STEP 之外

这里的圆是直线族 $y = mx \pm a\sqrt{1 + m^2}$ 的*包络线*（envelope），而 $(*)$ 是一条克莱罗方程（Clairaut equation）$y = xy' + f(y')$：它的通解（general solution）是一族直线，而它的*奇解*（singular solution）就是这族直线的包络线。把 $(*)$ 视为 y' 的二次方程，其判别式（discriminant）为 $4a^2(x^2 + y^2 - a^2)$，所以两个斜率选择恰好在圆上重合；这就是为甚么解可以像第 (iii) 部分那样在圆上拼接，也是为甚么解的唯一性（uniqueness）不成立。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）讨论了一阶方程解的存在性（existence）与唯一性。

详细解答

STEP II 2014，第 4 题

纯数学 · 代换积分法

STEP II 2014，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/14-s2-q4

关键思路

代换（substitution）$u = 1/x$ 会把区间 $[1/b, b]$ 的两个端点互换，因此它把积分 I 变回同一范围上的积分。若新的被积函数（integrand）只是原来的被积函数变了号，便得到 $I = -I$。若新的被积函数与原来的密切相关，把 I 的两种形式相加，可能得到一个容易积分得多的被积函数。在最后一部分，同样的技巧（取 $u = a^2/x$，使 $[0, \infty)$ 映射到自身）会给出一条方程，从中可解出所求的积分。

解答

第 (i) 部分

记该积分为 I。令 $x = 1/u$，则 $\mathrm{d}x = -\frac{1}{u^2}\,\mathrm{d}u$ 及 $\ln x = -\ln u$。当 $x = 1/b$ 时 $u = b$；当 $x = b$ 时 $u = 1/b$。此外

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

所以

$$\begin{aligned} I &= \int_b^{1/b} \frac{\frac{1}{u}(-\ln u)}{(a^2u^2 + 1)(a^2 + u^2)/u^4}\left(-\frac{1}{u^2}\right)\mathrm{d}u \\ &= \int_b^{1/b} \frac{u \ln u}{(a^2 + u^2)(a^2u^2 + 1)}\,\mathrm{d}u = -\int_{1/b}^{b} \frac{u \ln u}{(a^2 + u^2)(a^2u^2 + 1)}\,\mathrm{d}u. \end{aligned}$$

最后这个积分其实就是 I，只是把哑变量（dummy variable）改了名。因此 $I = -I$，得 $I = 0$，即所求。（由于 $b > 0$，被积函数在 $[1/b, b]$ 上连续，所以 I 是一个有限的数，这一步是成立的。）

第 (ii) 部分

记这个积分为 I。同样的代换给出 $\frac{\mathrm{d}x}{x} = u \cdot \left(-\frac{1}{u^2}\right)\mathrm{d}u = -\frac{\mathrm{d}u}{u}$，而积分上下限也像之前一样互换：

$$I = \int_b^{1/b} \arctan\left(\frac{1}{u}\right)\left(-\frac{1}{u}\right)\mathrm{d}u = \int_{1/b}^{b} \frac{\arctan(1/u)}{u}\,\mathrm{d}u.$$

这一次并非单纯变号，所以我们把这个形式与原来的形式相加：

$$2I = \int_{1/b}^{b} \frac{\arctan u + \arctan(1/u)}{u}\,\mathrm{d}u.$$

想一想 当 $u > 0$ 时，$\arctan u + \arctan(1/u)$ 等于多少?

设 $\theta = \arctan u$，其中 $u > 0$，则 $0 < \theta < \frac{\pi}{2}$，且 $\tan(\frac{\pi}{2} - \theta) = \cot\theta = 1/u$。由于 $\frac{\pi}{2} - \theta$ 亦位于 $(0, \frac{\pi}{2})$ 内，它就是 $\arctan(1/u)$。所以对所有 $u > 0$，$\arctan u + \arctan(1/u) = \frac{\pi}{2}$。（当 $u < 0$ 时，两者之和为 $-\frac{\pi}{2}$，但 $[1/b, b]$ 内每个 u 都是正数。）因此

$$2I = \frac{\pi}{2}\int_{1/b}^{b} \frac{\mathrm{d}u}{u} = \frac{\pi}{2}\big(\ln b - \ln(1/b)\big) = \pi \ln b,$$

所以 $I = \dfrac{\pi \ln b}{2}$，即所求。无论 $b > 1$ 还是 $0 < b < 1$，此结果均成立。

第 (iii) 部分

设 $J = \displaystyle\int_0^{\infty} \frac{\mathrm{d}x}{(a^2 + x^2)^2}$。这个积分收敛（converge），因为被积函数连续，而且当 x 很大时不超过 $1/x^4$。我们想要一个代换，把 $(0, \infty)$ 映射到自身，并把 $a^2 + x^2$ 变成 $a^2 + u^2$ 的某个倍数，所以取 $u = a^2/x$，即 $k = a^2$。于是 $x = a^2/u$，$\mathrm{d}x = -\frac{a^2}{u^2}\,\mathrm{d}u$，$x \to 0^+$ 对应 $u \to \infty$，而且

$$a^2 + x^2 = a^2 + \frac{a^4}{u^2} = \frac{a^2(u^2 + a^2)}{u^2}.$$

因此

$$J = \int_{\infty}^{0} \frac{u^4}{a^4(a^2 + u^2)^2}\left(-\frac{a^2}{u^2}\right)\mathrm{d}u = \frac{1}{a^2}\int_0^{\infty} \frac{u^2}{(a^2 + u^2)^2}\,\mathrm{d}u.$$

现在把分子写成 $u^2 = (a^2 + u^2) - a^2$：

$$\frac{u^2}{(a^2 + u^2)^2} = \frac{1}{a^2 + u^2} - \frac{a^2}{(a^2 + u^2)^2}.$$

两部分在 $(0, \infty)$ 上的积分都收敛，所以可以把积分拆开。利用题目给出的结果，

$$J = \frac{1}{a^2}\left(\frac{\pi}{2a} - a^2 J\right) = \frac{\pi}{2a^3} - J.$$

因此 $2J = \dfrac{\pi}{2a^3}$，所以

$$\int_0^\infty \frac{\mathrm{d}x}{(a^2+x^2)^2} = \frac{\pi}{4a^3},$$

即所求。

验算

- 在 (ii) 中，$b = 1$ 给出长度为零的区间，而 $\frac{\pi}{2}\ln 1 = 0$ 确实成立。把 b 换成 $1/b$ 会使上下限互换，所以积分应该变号，而 $\ln(1/b) = -\ln b$ 与此吻合。
- 在 (iii) 中，答案应有正确的标度关系：令 $x = at$ 得 $J = a^{-3}\int_0^\infty (1+t^2)^{-2}\,\mathrm{d}t$，所以 J 必定是某常数乘以 a^{-3}。取 $a = 1$，代换 $x = \tan\theta$ 给出 $\int_0^{\pi/2}\cos^2\theta\,\mathrm{d}\theta = \frac{\pi}{4}$，与结果相符。
- 以数值计算，(ii) 中的积分在 $b = 2$ 时为 $1.08879\ldots = \frac{\pi}{2}\ln 2$。对于若干随机选取的 a 和 b，(i) 中的积分计算结果都是 0（准确至多个小数位）。

常见错误

- 忘记 $u = 1/x$ 同时会使上下限倒转。$\mathrm{d}x = -\mathrm{d}u/u^2$ 带来的负号和上下限的互换，两者都必须留意。
- 使用 $\arctan u + \arctan(1/u) = \frac{\pi}{2}$ 时没有说明它需要 $u > 0$。
- 在 (iii) 中，得到 $\frac{1}{a^2}\int_0^\infty \frac{u^2}{(a^2+u^2)^2}\,\mathrm{d}u$ 后就停下来，没有发现它可以改写成以 J 和题目给出的积分表示。

STEP 之外

利用定义域的对称性把一个积分与其自身联系起来，是一个常见技巧。它会在围道积分（contour integration）以及诸如 $\int_0^{\pi/2}\ln\sin x\,\mathrm{d}x$ 的积分中再次出现。(iii) 中的积分亦来自一个一般的积分族：把 $\int_0^\infty (\lambda+x^2)^{-1}\mathrm{d}x = \frac{\pi}{2}\lambda^{-1/2}$ 对 λ 求导，即可立刻得到同样的答案（“积分号下求导”（differentiation under the integral sign））。实分析（real analysis）课程会解释这类步骤，以及反常积分（improper integral）的处理，在甚么情况下是容许的。参见 Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）。

详细解答

STEP II 2014，第 5 题

纯数学 · 齐次微分方程

STEP II 2014，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/14-s2-q5

关键思路

两条方程的右方都是两个线性式之比。在 (i) 中，每一项都是一次项，因此斜率（gradient）只取决于 y/x，代入 $y = xu$ 便可把方程化为可分离变量（separable）的方程。在 (ii) 中，常数项破坏了这个性质，但只要把原点移到分子和分母所代表的两条直线的交点，便可消去常数项；之后就跟 (i) 完全一样。

解答

若 $y = xu$，由乘积法则（product rule）得

$$\frac{\mathrm{d}y}{\mathrm{d}x} = u + x\frac{\mathrm{d}u}{\mathrm{d}x}.$$

第 (i) 部分

把右方的分子和分母同除以 x，方程变为

$$u + x\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{2u+1}{u-2}, \qquad \text{所以} \qquad x\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{2u+1-u(u-2)}{u-2} = \frac{1+4u-u^2}{u-2}.$$

分离变量，得

$$\int \frac{u-2}{1+4u-u^2}\,\mathrm{d}u = \int \frac{1}{x}\,\mathrm{d}x.$$

? **想一想** 不用部分分式（partial fractions），怎样积分左方？

把分母求导：$\frac{\mathrm{d}}{\mathrm{d}u}(1+4u-u^2) = 4-2u = -2(u-2)$。因此分子是分母导数的 $-\frac{1}{2}$ 倍，于

是

$$-\tfrac{1}{2}\ln\left|1+4u-u^2\right| = \ln|x| + C.$$

两边乘以 -2 再取指数，得 $x^2\,|1+4u-u^2| = A$，其中 A 为正常数；即对某常数 K 有 $x^2(1+4u-u^2) = K$（沿着曲线，符号保持不变）。现代入 $u = y/x$，使得 $x^2(1+4y/x - y^2/x^2) = x^2+4xy-y^2$：

$$x^2+4xy-y^2 = K.$$

在 $(1,1)$ 处得 $K = 1+4-1 = 4$，所以解曲线为

$$\boxed{x^2+4xy-y^2 = 4}$$

（在 $(1,1)$ 处 $u=1$，此时 $1+4u-u^2 = 4 \neq 0$ 且 $u \neq 2$，所以我们除过的式子在起点处都不为零。）

第 (ii) 部分

直线 $x-2y-4=0$ 与 $2x+y-3=0$ 相交于 $y = 3-2x$ 且 $x-2(3-2x)-4=0$ 之处，即 $(2,-1)$。因此取 $a=2$、$b=-1$：$x = X+2$，$y = Y-1$。于是 $\mathrm{d}y/\mathrm{d}x = \mathrm{d}Y/\mathrm{d}X$，而且

$$x-2y-4 = (X+2)-2(Y-1)-4 = X-2Y,$$
$$2x+y-3 = 2(X+2)+(Y-1)-3 = 2X+Y,$$

所以常数项消失了：

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \frac{X-2Y}{2X+Y}.$$

这与 (i) 的形式相同，所以代入 $Y = XV$：

$$V + X\frac{\mathrm{d}V}{\mathrm{d}X} = \frac{1-2V}{2+V} \quad\Longrightarrow\quad X\frac{\mathrm{d}V}{\mathrm{d}X} = \frac{1-2V-V(2+V)}{2+V} = \frac{1-4V-V^2}{2+V}.$$

由于 $\frac{\mathrm{d}}{\mathrm{d}V}(1-4V-V^2) = -2(2+V)$，分子同样是分母导数的倍数。因此

$$\int \frac{2+V}{1-4V-V^2}\,\mathrm{d}V = \int \frac{1}{X}\,\mathrm{d}X \quad\Longrightarrow\quad -\tfrac{1}{2}\ln\left|1-4V-V^2\right| = \ln|X| + C,$$

与之前一样得 $X^2(1-4V-V^2) = K$，即

$$X^2-4XY-Y^2 = K.$$

点 $(1,1)$ 对应 $X=-1$、$Y=2$，得 $K = 1-4(-1)(2)-4 = 5$。换回 x 和 y，

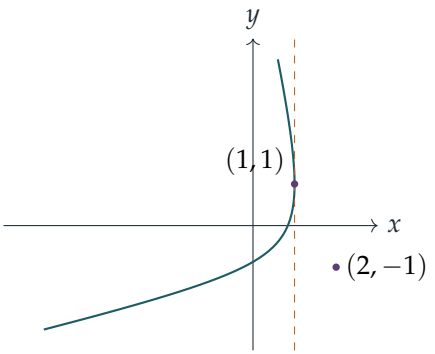
$$(x-2)^2-4(x-2)(y+1)-(y+1)^2 = 5,$$

展开得

$$(x^2-4x+4)-4(xy+x-2y-2)-(y^2+2y+1) = 5$$
$$\Longleftrightarrow\quad \boxed{x^2-4xy-y^2-8x+6y+6=0}$$

想一想 在 $(1,1)$ 处，分母 $2x+y-3$ 等于零。这个解还有意义吗?

有：该处的斜率为无穷大，所以曲线在 $(1,1)$ 处有一条铅垂切线（vertical tangent）。隐方程（implicit equation）完全可以描述这条曲线，只是它在该点附近并非某个函数 $y(x)$ 的图像而已。换句话说，x 是 y 的光滑函数。把二次方程对 x 求解，在经过 $(1,1)$ 的分支上得 $x=2y+4-\sqrt{5y^2+10y+10}$；由于 $5y^2+10y+10=5(y+1)^2+5>0$，它对所有 y 都有定义，并在 $y=1$ 时取得最大值 $x=1$。



图中显示经过 $(1,1)$ 的分支（蓝绿色），它在该点与铅垂线 $x=1$（虚线）相切；$(2,-1)$ 是双曲线（hyperbola）的中心。

验算

- 对答案进行隐函数求导(implicit differentiation)。在 (i) 中，由 $2x+4y+(4x-2y)y'=0$ 得 $y'=\dfrac{x+2y}{y-2x}$，符合要求。在 (ii) 中，由 $2x-4y-8+(6-4x-2y)y'=0$ 得 $y'=\dfrac{x-2y-4}{2x+y-3}$。
- 两条曲线都经过 $(1,1)$：$1+4-1=4$ 及 $1-4-1-8+6+6=0$。
- 用电脑从 $(1,1)$ 出发，以数值方法逐步求解每条微分方程（differential equation），方框内方程的左方一直保持等于 4 和 0，误差约为 10^{-13}。在 (i) 中，$(1,1)$ 处的斜率为 -3。

常见错误

- 在 $\mathrm{d}y/\mathrm{d}x=u+x\,\mathrm{d}u/\mathrm{d}x$ 中漏掉 u，结果得到错误的可分离方程。
- 在 (ii) 中用错坐标代入条件：点 $(1,1)$ 会变成 $(X,Y)=(-1,2)$，而不是 $(1,1)$。

- 对 u 贸然使用部分分式：$1+4u-u^2$ 的根是 $2\pm\sqrt{5}$，计算繁复，而被积函数其实已经是 f'/f 的形式。

STEP 之外

右方只取决于 y/x 的方程称为齐次方程（homogeneous equation），而 (ii) 中平移原点的做法，是处理线性函数之比的标准技巧。这里的两条方程同时也是恰当方程（exact equation）：把 (i) 写成 $(2y+x)\,\mathrm{d}x-(y-2x)\,\mathrm{d}y=0$，两个系数满足交叉偏导数检验，守恒量 $x^2+4xy-y^2$ 可以直接读出。恰当方程、积分因子（integrating factor）以及线性系统的相图（phase portrait）（这些双曲线的真正身份正是如此），可参阅 Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）。

详细解答

STEP II 2014，第 6 题

纯数学 · 三角级数

> **STEP II 2014，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/14-s2-q6

关键思路

把每个 $\cos rx$ 乘以 $2\sin\frac{1}{2}x$，就会变成两个正弦之差，于是整个和式可以伸缩相消（telescopes）。这个和式正正就是 $S_n'(x)$，这也是题目给出这个恒等式的原因：它告诉我们 S_n 在哪里有驻点（stationary point）。在这样的点上，新加的项 $\frac{1}{n}\sin nx$ 原来是非负的，所以加上它不会制造出负的极小值；再从 $S_1 = \sin x$ 开始用归纳法（induction），就可以完成证明。

解答

余弦之和

利用和差化积公式（factor formula）$\sin A - \sin B = 2\cos\frac{A+B}{2}\sin\frac{A-B}{2}$，取 $A = (r+\frac{1}{2})x$、$B = (r-\frac{1}{2})x$，得

$$\sin(r+\tfrac{1}{2})x - \sin(r-\tfrac{1}{2})x = 2\cos rx\,\sin\tfrac{1}{2}x.$$

由 $r = 1$ 加到 n，右边加起来是 $2\sin\frac{1}{2}x\,(\cos x + \cdots + \cos nx)$，而左边除了首项和末项之外全部互相抵消：

$$2\sin\tfrac{1}{2}x\sum_{r=1}^{n}\cos rx = \sin(n+\tfrac{1}{2})x - \sin\tfrac{1}{2}x.$$

两边除以 $2\sin\frac{1}{2}x \neq 0$，即得所求结果。留意逐项微分（differentiating term by term）可得

$$S_n'(x) = \sum_{r=1}^{n}\cos rx,$$

所以这个恒等式就是 S_n' 的闭形式（closed form）。

第 (i) 部分

这里 $S_2(x) = \sin x + \frac{1}{2}\sin 2x$，所以

$$S_2'(x) = \cos x + \cos 2x = 2\cos^2 x + \cos x - 1 = (2\cos x - 1)(\cos x + 1).$$

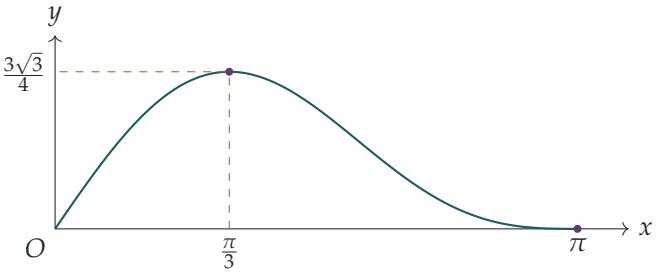
在 $[0, \pi]$ 上，当 $\cos x = \frac{1}{2}$，即 $x = \frac{\pi}{3}$，以及当 $\cos x = -1$，即 $x = \pi$ 时，它等于零。对应的值为

$$S_2(\tfrac{\pi}{3}) = \tfrac{\sqrt{3}}{2} + \tfrac{1}{2}\cdot\tfrac{\sqrt{3}}{2} = \tfrac{3\sqrt{3}}{4}, \qquad S_2(\pi) = 0.$$

由于 $\cos x + 1 \geq 0$，S_2' 的正负号与 $2\cos x - 1$ 相同：在 $[0, \frac{\pi}{3})$ 上为正，在 $(\frac{\pi}{3}, \pi)$ 上为负。所以 $x = \frac{\pi}{3}$ 是极大值（maximum），而 S_2 由此递减至 $x = \pi$ 时的 0，到达时斜率为零。在 π 附近，设 $x = \pi - t$,

$$S_2 = \sin t - \tfrac{1}{2}\sin 2t \approx \left(t - \tfrac{t^3}{6}\right) - \left(t - \tfrac{4t^3}{6}\right) = \tfrac{1}{2}t^3,$$

所以曲线到达横轴时会像三次曲线般变得平坦。此外 $S_2(0) = 0$，且 $S_2'(0) = 2$。



第 (ii) 部分

假设 $S_n'(x_0) = 0$, 其中 $0 < x_0 < \pi$。那么 $\sin\frac{1}{2}x_0 \neq 0$, 所以恒等式适用, 得 $\sin(n+\frac{1}{2})x_0 = \sin\frac{1}{2}x_0$。展开左边,

$$\sin nx_0\cos\tfrac{1}{2}x_0 + \cos nx_0\sin\tfrac{1}{2}x_0 = \sin\tfrac{1}{2}x_0 \iff \sin nx_0\cos\tfrac{1}{2}x_0 = (1-\cos nx_0)\sin\tfrac{1}{2}x_0.$$

由于 $0 < \frac{1}{2}x_0 < \frac{\pi}{2}$，有 $\cos\frac{1}{2}x_0 > 0$，两边除以它得

$$\sin nx_0 = (1-\cos nx_0)\tan\tfrac{1}{2}x_0,$$

此即所求。因此，对于 $n \geq 2$,

$$S_n(x_0) - S_{n-1}(x_0) = \frac{1}{n}\sin nx_0 = \frac{1}{n}(1-\cos nx_0)\tan\tfrac{1}{2}x_0 \geq 0,$$

因为 $1 - \cos nx_0 \geq 0$ 且 $\tan\frac{1}{2}x_0 > 0$。所以 $S_n(x_0) \geq S_{n-1}(x_0)$。

现在假设对所有 $x \in (0, \pi)$ 都有 $S_{n-1}(x) > 0$，并用反证法（proof by contradiction），假设存在某个 $c \in (0, \pi)$ 使得 $S_n(c) \leq 0$。

想一想 我们怎样知道有驻点可用?

S_n 在闭区间（closed interval）$[0,\pi]$ 上连续（continuous），所以它在那里取得最小值 m，而且 $m \le S_n(c) \le 0$。若 $m < 0$，由于 $S_n(0) = S_n(\pi) = 0$，最小值不在端点；若 $m = 0$，则 c 本身就是取得最小值的内点。无论哪种情况，都存在 $x_0 \in (0,\pi)$ 使得 $S_n(x_0) = m \le 0$；又由于 S_n 可微（differentiable），内部的最小值必定是驻点：$S_n'(x_0) = 0$。于是前面的结果给出

$$0 \ge S_n(x_0) \ge S_{n-1}(x_0) > 0,$$

矛盾。所以对所有 $x \in (0,\pi)$ 都有 $S_n(x) > 0$。

第 (iii) 部分

我们用归纳法证明当 $0 < x < \pi$ 时 $S_n(x) > 0$。当 $n = 1$ 时，在 $(0,\pi)$ 上 $S_1(x) = \sin x > 0$。若结论对 S_{n-1} 成立，则第 (ii) 部分说明它对 S_n 亦成立。所以结论对每个 $n \ge 1$ 都成立。在端点处，每一项 $\frac{1}{r}\sin rx$ 都是零，所以 $S_n(0) = S_n(\pi) = 0$。因此对所有 $n \ge 1$ 及 $0 \le x \le \pi$，都有 $S_n(x) \ge 0$。

验算

- 在余弦之和中代入 $n = 1$：$\dfrac{\sin\frac{3}{2}x - \sin\frac{1}{2}x}{2\sin\frac{1}{2}x}$ 应该等于 $\cos x$，而由和差化积公式，它确实如此。
- 第 (i) 部分与恒等式吻合：$S_2' = 0$ 即 $\sin\frac{5}{2}x = \sin\frac{1}{2}x$，它在 $(0,\pi]$ 内的解是 $x = \frac{\pi}{3}$ 和 $x = \pi$。
- 极大值 $\frac{3\sqrt{3}}{4} \approx 1.299$ 与在细密网格上直接计算 S_2 的结果相符。电脑验算找出了 $2 \le n \le 40$ 时 S_n 在 $(0,\pi)$ 内的每一个驻点，并在每一点确认了第 (ii) 部分的恒等式及 $S_n(x_0) \ge S_{n-1}(x_0)$；亦确认了对于 $n \le 60$，在区间内部 $S_n > 0$。

常见错误

- 论证“处处 $S_n \ge S_{n-1}$”。这是错的（例如在 $(\frac{\pi}{2},\pi)$ 上 $\sin 2x < 0$）；这个不等式只在 S_n 的驻点成立，所以才需要用最小值的论证。
- 找最小值时忘记端点：必须利用 $S_n(0) = S_n(\pi) = 0$，才能把非正的最小值定位在区间内部。
- 除以 $\cos\frac{1}{2}x_0$ 时没有说明它为何是正数。

STEP 之外

第 (iii) 部分的结果就是 Fejér–Jackson–Gronwall 不等式。当 $n \to \infty$ 时，$S_n(x)$ 在 $(0, \pi]$ 上收敛（converges）至 $\frac{1}{2}(\pi - x)$：$\sum \frac{1}{r} \sin rx$ 是锯齿波（sawtooth wave）的傅里叶级数（Fourier series），而在 $x = 0$ 附近，部分和会出现过冲（吉布斯现象，Gibbs phenomenon）。傅里叶级数，以及它们何时收敛这个微妙的问题，是大学第一门分析或数学方法课程的核心课题。T. W. Körner 的 *Fourier Analysis*（《傅里叶分析》）是一部精彩而涵盖广泛的著作。

详细解答

STEP II 2014，第 7 题

纯数学 · 绝对值函数的图像

STEP II 2014，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/14-s2-q7

关键思路

若干绝对值函数（modulus function）之和是分段线性（piecewise linear）的：只有在其中某个绝对值内的式子等于零的点，斜率（gradient）才会改变。因此 $|x-a|+|x-b|$ 的图像是一个“浴缸”：在 a 与 b 之间是平坦的底部，两旁是斜率为 ± 2 的缸壁。之后各条方程的解的个数，就归结为比较斜率：斜率为 2 的图像必定会追上斜率为 1 的图像，而两个斜率相同的图像要么重合，要么永不相交。

解答

全文设 $f(x)=|x-a|+|x-b|$，其中 $a<b$。

第 (i) 部分

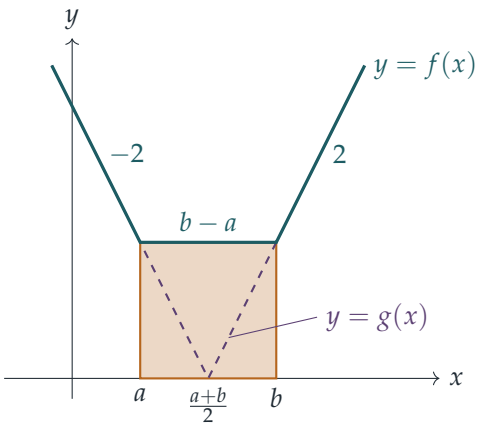
逐个区间去掉绝对值符号：

$$f(x)=\begin{cases} a+b-2x, & x\le a,\\ b-a, & a\le x\le b,\\ 2x-a-b, & x\ge b. \end{cases}$$

所以当 $x<a$ 时斜率为 -2，当 $a<x<b$ 时为 0，当 $x>b$ 时为 2。图像是连续（continuous）的，在 $[a,b]$ 上方有一段高度为 $b-a$ 的平坦底部。

$g(x)=|2x-a-b|=2|x-\frac{a+b}{2}|$ 的图像是一个 V 形，顶点在 $\left(\frac{a+b}{2},0\right)$，斜率为 ± 2。当 $x\ge b$ 时，$2x-a-b\ge b-a>0$，所以 $g(x)=2x-a-b=f(x)$；同理，当 $x\le a$ 时 $g(x)=a+b-2x=f(x)$。因此两个图像在 (a,b) 以外重合。在区间内，g 下降至 0，而

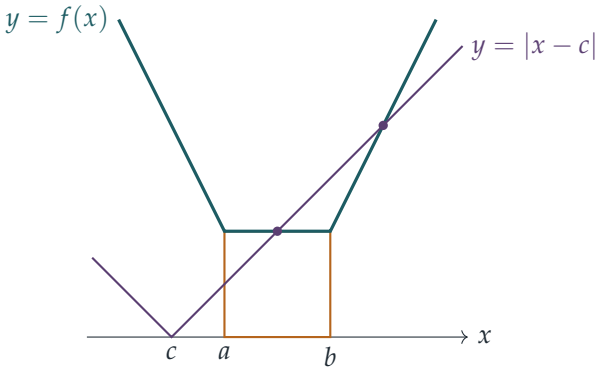
f 则保持在 $b-a$。（事实上 $f(x)=\max\{b-a,\,g(x)\}$。）



利用 $f(a)=f(b)=b-a$，可知该四边形的顶点为 $(a,0)$、$(b,0)$、$(b,b-a)$ 和 $(a,b-a)$。它的底边长 $b-a$，两条铅垂边长 $b-a$，而且所有角都是直角。所以它是一个边长为 $b-a$ 的**正方形**（square）。另外留意，由于 $g(a)=g(b)=b-a$，g 的 V 形从正方形底边的中点出发，向上穿过它的两个上方顶点。

第 (ii) 部分

现在比较 $y=f(x)$ 与 $y=|x-c|$；后者是一个顶点 $(c,0)$ 在 x 轴上、两臂斜率为 ±1 的 V 形。关键事实是 f 恒不小于 $b-a>0$，而且它的缸壁（斜率 ±2）比 V 形的两臂陡。草图显示 $c<a$ 的情况，此时 V 形与浴缸相交两次。



情况 $a<c<b$：无解。在 $[a,b]$ 上，距离 $|x-c|$ 至多为 $\max\{c-a,\,b-c\}<b-a=f(x)$，所以 V 形严格位于浴缸底部之下。当 $x\geq b$ 时，

$$f(x)-|x-c|=(2x-a-b)-(x-c)=x-b+(c-a)\geq c-a>0,$$

而同样的论证（或下文的对称性（symmetry））可处理 $x \le a$ 的情况。所以 V 形完全处于浴缸之内，方程无解。

情况 $c = a$ 或 $c = b$：一个解。取 $c = a$。在 $[a, b]$ 上，$|x - a| = x - a \le b - a$，等号只在 $x = b$ 时成立。当 $x > b$ 时，$f(x) - (x - a) = x - b > 0$；当 $x < a$ 时，$f(x) - (a - x) = b - x > 0$。所以唯一的解是 $x = b$：V 形的右臂恰好穿过正方形的上方顶点 $(b, b - a)$。同理，当 $c = b$ 时唯一的解是 $x = a$。

情况 $c < a$：两个解。左臂 $y = c - x$ 位于 $x \le c < a$ 的范围，而 f 在该处的斜率为 -2。在 $x = c$ 处有 $f(c) > 0 = c - c$，而向左移动时 f 上升得比左臂快，所以对所有 $x \le c$ 都有 $f(x) > c - x$。左臂不给出任何解。

至于右臂 $y = x - c$（$x \ge c$），设 $k(x) = f(x) - (x - c)$。把 f 的各段斜率减去 1，可知 k 在 (c, a) 上斜率为 -3，在 (a, b) 上为 -1，当 $x > b$ 时为 $+1$。所以 k 在 $[c, b]$ 上严格递减（strictly decreasing），在 $[b, \infty)$ 上严格递增。此外

$$k(c) = f(c) > 0, \qquad k(b) = (b - a) - (b - c) = c - a < 0, \qquad k(x) \to \infty \text{ 当} x \to \infty.$$

因此 k 在 (c, b) 内恰有一个零点，在 (b, ∞) 内也恰有一个零点。具体而言，它们是

$$x = a + b - c \quad \text{及} \quad x = \begin{cases} b - a + c, & 2a - b \le c < a, \\ \frac{1}{3}(a + b + c), & c < 2a - b, \end{cases}$$

其中第二个解分别落在浴缸的底部或其左壁上。

情况 $c > b$：两个解。关于直线 $x = \frac{a+b}{2}$ 的反射（reflection）$x \mapsto a + b - x$ 使 f 保持不变（它把 $|x - a|$ 与 $|x - b|$ 互换），并把 $|x - c|$ 变为 $|x - c'|$，其中 $c' = a + b - c < a$。所以这个情况是上一个情况的镜像。

总结：若 $a < c < b$，则有 0 个解；若 $c = a$ 或 $c = b$，则有 1 个解；若 $c < a$ 或 $c > b$，则有 2 个解。

第 (iii) 部分

设 $F(x) = |x - c| + |x - d|$。由第 (i) 部分，F 是第二个浴缸，在 $[c, d]$ 上方的底部高度为 $d - c$，缸壁斜率为 ± 2。由于 $d - c < b - a$，它的底部比 f 的更低也更窄。记

$$m = \tfrac{1}{2}(a + b), \qquad n = \tfrac{1}{2}(c + d)$$

为两个底部的中点，并设 $h(x) = f(x) - F(x)$。

在最左方（当 $x \le \min\{a, c\}$ 时），两个图像都在各自的左壁上，所以 $h(x) = (a + b - 2x) - (c + d - 2x) = 2(m - n)$。在最右方（当 $x \ge \max\{b, d\}$ 时），$h(x) = 2(n - m)$。在两者之间，h 是连续且分段线性的。

想一想 两个图像能否共有一整段线段，从而给出无限多个解?

只有在两者斜率相同的地方才有可能。两个底部都是平的，但它们处于不同高度 $b-a \neq d-c$，所以不会重叠。除此之外，斜率只在两个图像都位于左壁（$x < \min\{a,c\}$）或都位于右壁（$x > \max\{b,d\}$）时才相同，而在这些地方 $h = \pm 2(m-n)$。所以两个图像共有一段线段，当且仅当 $m = n$。

情况 $a+b=c+d$：无限多个解。此时当 $x \leq \min\{a,c\}$ 及 $x \geq \max\{b,d\}$ 时 $h = 0$。由于 $d-c < b-a$ 而两个中点相同，有 $a < c < d < b$：小浴缸正好位于大浴缸的中央。当 $c \leq x \leq b$ 时，F 的图像在其底部或右壁上，所以 $F(x) \leq 2b-c-d = b-a = f(x)$，等号只在 $x = b$ 时成立。由对称性，在 $[a,d]$ 上亦然。所以解集恰好是 $x \leq a$ 连同 $x \geq b$。

情况 $a+b < c+d$：恰有一个解。此时在最左方 $h = 2(m-n) < 0$，在最右方 $h = 2(n-m) > 0$。h 在哪里会递减？这需要 f 的斜率小于 F 的斜率，有两种可能。

- f 在其左壁上而 F 不在，这需要 $c < x < a$。于是 $c < a$，又由于 $d-c < b-a$，亦有 $d < b$，从而 $c+d < a+b$，但这在此情况下不成立。
- f 在其底部而 F 在其右壁上，这需要 $d < x < b$。

所以 h 在 $(-\infty, d]$ 上是非递减的。在 $[d,\infty)$ 上，h 在 $[d,b]$ 上非递增（若 $d < b$），然后当 $x \geq \max\{b,d\}$ 时恒等于 $2(n-m)$。所以对所有 $x \geq d$ 都有 $h(x) \geq 2(n-m) > 0$。在 $(-\infty, d]$ 上，h 从值 $2(m-n) < 0$ 不递减地上升至 $h(d) > 0$，而且（由上面方框的讨论）它不会在任何区间上恒等于零。所以 h 在该处恰有一个零点。总共恰有一个解。

情况 $a+b > c+d$：恰有一个解，可由反射 $x \mapsto -x$ 得出；这个反射把 a,b,c,d 换成 $-b,-a,-d,-c$，并使两个和之间的不等式反向。

总结：在 $a < b$、$c < d$ 及 $d-c < b-a$ 的条件下，若 $a+b = c+d$（两个区间的中点相同），则有无限多个解，即所有 $x \leq a$ 及所有 $x \geq b$。否则恰有一个解。

验算

- 第 (ii) 部分取 $a = 1$、$b = 3$、$c = 0$：解 $x = a+b-c = 4$ 及 $x = b-a+c = 2$ 分别给出 $|3|+|1| = 4$ 及 $1+1 = 2$，符合要求。取 $c = -5$ 时，第二个解是 $\frac{1}{3}(1+3-5) = -\frac{1}{3}$，而确实 $\frac{4}{3}+\frac{10}{3} = \frac{14}{3} = |-\frac{1}{3}+5|$。
- 第 (iii) 部分取 $a = 0$、$b = 10$、$c = 6$、$d = 7$：h 在最左方为 -3，在 $[6,7]$ 上升至 9，然后回落至 3 并保持不变。它唯一的零点是 $x = \frac{3}{2}$，此时两边都等于 10。
- 以电脑对数以万计随机选取的整数 a,b,c,d 作检验，每种情况下得到的解的个数都与上述完全一致。

常见错误

- 在 (ii) 中遗漏边界情况 $c = a$ 及 $c = b$；此时 V 形刚好碰到正方形的一个顶点。
- 在 (iii) 中假设 h 是单调（monotonic）的。其实不然：当小浴缸偏离中央地位于大浴缸内时，h 在 (d, b) 上递减。令解的个数不受影响的是，h 在该处已经是正数。
- 在 (iii) 中遗漏中点重合时的无限多解情况。

STEP 之外

函数 $|x - a| + |x - b|$ 是 x 到两个点的距离总和。把这类距离之和最小化，正是统计学中中位数（median）以及 ℓ^1（“计程车”）范数（norm）背后的概念。像 f 这样的函数是凸（convex）的，而这里关于斜率逐段递增的论证，正是凸分析（convex analysis）的起点；凸分析是线性规划（linear programming）和最优化（optimisation）的基础。Bartle 与 Sherbert 所著的 *Introduction to Real Analysis*（《实分析导论》）把这里用到的连续性和单调性论证严格化。

详细解答

STEP II 2014，第 8 题

纯数学 · 最大二项式系数

STEP II 2014，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/14-s2-q8

关键思路

要找出一个正项数列（sequence）在哪里达到最大值，可以把每一项与下一项比较：考虑比值（ratio）c_{r+1}/c_r。对二项式展开（binomial expansion）而言，这个比值是 r 的一个简单分式，而且稳定递减，所以当它大于 1 时系数（coefficient）上升，一旦小于 1 系数便下降。最大的系数就在比值越过 1 的位置，而之后各部分都只需读出某个式子的整数部分（integer part）。

解答

系数及其比值

由二项式定理（binomial theorem），

$$c_r = \binom{n}{r} a^{n-r} b^r, \qquad 0 \le r \le n.$$

由于 $a, b > 0$，这些系数全部是正数。对 $0 \le r \le n-1$，

$$\frac{c_{r+1}}{c_r} = \frac{\binom{n}{r+1}}{\binom{n}{r}} \cdot \frac{b}{a} = \frac{(n-r)\,b}{(r+1)\,a},$$

这里用了 $\binom{n}{r+1} = \binom{n}{r}\frac{n-r}{r+1}$。记

$$L = \frac{b(n+1)}{a+b}.$$

因为 $(r+1)a > 0$，

$$c_{r+1} \ge c_r \iff (n-r)b \ge (r+1)a \iff nb - a \ge r(a+b)$$
$$\iff r \le \frac{nb-a}{a+b} = \frac{b(n+1)}{a+b} - 1 = L - 1,$$

而把各处换成 $=$ 后，同一串推导依然成立。因此

$$c_{r+1} > c_r \text{ 若} r < L-1, \qquad c_{r+1} = c_r \text{ 若} r = L-1, \qquad c_{r+1} < c_r \text{ 若} r > L-1. \qquad (*)$$

m 的上下界

上界（*upper bound*）。若 $m = 0$，由于 $L > 0$，有 $m \le L$。若 $m \ge 1$，由于 c_m 最大，有 $c_m \ge c_{m-1}$，再由 $(*)$ 取 $r = m-1$，得 $m - 1 \le L - 1$，即 $m \le L$。

下界（*lower bound*）。若 $m = n$，由于 $L < n+1$（因为 $\frac{b}{a+b} < 1$），有 $m \ge L - 1$。若 $m \le n-1$，则 $c_m \ge c_{m+1}$，由 $(*)$ 取 $r = m$，便迫使 $m \ge L - 1$。

所以在所有情况下都有

$$\frac{b(n+1)}{a+b} - 1 \le m \le \frac{b(n+1)}{a+b},$$

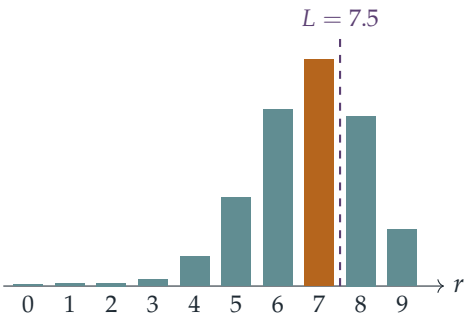
正是所求。

> **想一想** 为甚么 m 不可能有三个值，或者有两个不相邻的值?

区间 $[L-1, L]$ 的长度是 1。若 L 不是整数，这个区间恰好包含一个整数 $\lfloor L \rfloor$，所以 m 是唯一的。若 L 是整数，区间恰好包含两个相邻整数 $L-1$ 和 L，所以 m 是其中之一。（事实上，由 $(*)$ 取 $r = L-1$ 可见 $c_{L-1} = c_L$，所以两者确实都是最大的。）无论哪种情况，唯一的那个值，或两个值中较大者，都是

$$G(n,a,b) = \left\lfloor \frac{b(n+1)}{a+b} \right\rfloor.$$

从此以后，我们只需要这条公式。



按比例绘画的 $(1+3x)^9$ 各项系数：当 $r < L - 1 = 6.5$ 时系数上升，之后下降，在 $r = 7$ 处达到最大。

第 (i) 部分

对 $(n,a,b) = (9,1,3)$，$L = \frac{3\times 10}{4} = 7.5$，所以 $G(9,1,3) = 7$。

对 $(n,a,b)=(9,2,3)$，$L=\frac{3\times10}{5}=6$ 是整数，所以最大系数同时出现在 $r=5$ 和 $r=6$，而 $G(9,2,3)=6$。（验证：$c_5=126\cdot2^4\cdot3^5=489\,888=84\cdot2^3\cdot3^6=c_6$。）

第 (ii) 部分

当 $a=b$ 时，$L=\frac{n+1}{2}$。

- $n=2k$：$L=k+\frac{1}{2}$，所以 $G(2k,a,a)=k$（即中间的系数，是唯一的）。
- $n=2k-1$：$L=k$ 是整数，所以中间两个系数 $r=k-1$ 和 $r=k$ 相等，而 $G(2k-1,a,a)=k$。

这与帕斯卡三角形（Pascal's triangle）各行的对称性（symmetry）吻合。

第 (iii) 部分

固定 n 和 b 时，$L=\frac{b(n+1)}{a+b}$ 是 a 的递减函数（decreasing function），所以 $\lfloor L\rfloor$ 是 a 的非递增函数（non-increasing function）。因此在正整数范围内，它在 $a=1$ 时最大，此时 $G(n,1,b)=\left\lfloor\frac{b(n+1)}{b+1}\right\rfloor$。（其他 a 值可能与之相等，但没有一个比它更大。）

第 (iv) 部分

现在

$$\frac{b(n+1)}{b+1}=(n+1)-\frac{n+1}{b+1}.$$

由于 $\frac{n+1}{b+1}>0$，此式小于 $n+1$，所以 $G(n,1,b)\le n$。（这亦可由 $m\le n$ 得出。）此外，

$$G(n,1,b)=n\iff(n+1)-\frac{n+1}{b+1}\ge n\iff\frac{n+1}{b+1}\le1\iff b\ge n.$$

所以 $G(n,1,b)$ 的最大可能值是 n，而且恰好在 $b\ge n$ 时达到。当 $b=n$ 时，最后两个系数相等，因为 $c_{n-1}=n\cdot b^{n-1}=b^n=c_n$；当 $b>n$ 时，最后一个系数 b^n 是唯一的最大者。

验算

- $(1+3x)^9$ 的系数是 $1,27,324,2268,10206,30618,61236,78732,59049,19683$：最大者在 $r=7$，与 (i) 的结果一致。
- 对 $(1+x)^4$ $(n=2k,\ k=2)$，系数 $1,4,6,4,1$ 在 $r=2=k$ 处最大；对 $(1+x)^3$ $(k=2)$，系数是 $1,3,3,1$，在 $r=1,2$ 处相等，较大的值是 $2=k$。
- 在 (iv) 中取 $n=3$：$(1+2x)^3$ 的系数是 $1,6,12,8$，在 $r=2$ 处最大；$(1+3x)^3$ 的系数是 $1,9,27,27$，所以当 $b=3$ 时便达到 $G=3$。

- 用电脑对所有 $n < 30$ 及 $a, b < 15$ 验证 $G(n,a,b) = \lfloor b(n+1)/(a+b) \rfloor$，每一种情况都吻合。

常见错误

- 证明上下界时忘记处理端点情况 $m = 0$ 和 $m = n$：此时并没有 c_{-1} 或 c_{n+1} 可供比较。
- 当 L 是整数时，取了两个相等值中较小的一个：G 定义为较大者，所以 $G(9,2,3) = 6$，而不是 5。
- 在 (iv) 中只答 $b = n$：其实每个 $b \geq n$ 都给出 $G = n$。

STEP 之外

这里所用的比值检验，即比较相邻两项，与判断级数（series）收敛（convergence）的比值检验法（ratio test）是同一个想法；它亦可找出二项分布（binomial distribution）$\mathrm{B}(n,p)$（其中 $p = \frac{b}{a+b}$）的众数（mode）：众数是 $\lfloor (n+1)p \rfloor$。像二项式系数这样先升后降的数列称为单峰（unimodal）数列，而证明这一点的常用工具是对数凹性（log-concavity）（$c_r^2 \geq c_{r-1}c_{r+1}$）。概率方面的内容，可参阅 Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）。

详细解答

STEP II 2014，第 9 题

力学 · 刚体平衡

STEP II 2014，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/14-s2-q9

关键思路

未知的力有三个（张力（tension）、法向反作用力（normal reaction）和 A 处的摩擦力（friction）），而方程也有三条：水平方向分解、铅垂方向分解，以及取力矩（moments）。关于角 A 取力矩是聪明的选择，因为墙壁施加的两个接触力都作用于该点，因而在方程中消去，只剩下一条把张力和重量联系起来的方程。得到这条方程后，摩擦力的方向就只取决于单靠绳子所支撑的力比重量大还是小。

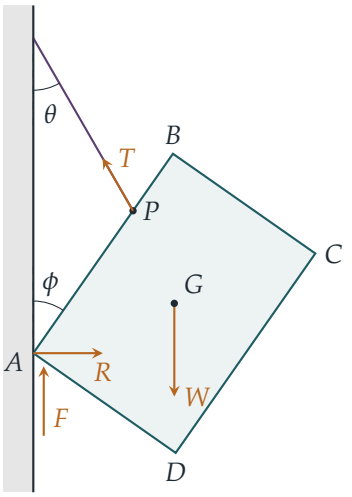
解答

我们把薄板（lamina）模型化为重量 W 的均匀刚体（uniform rigid body），其质心（centre of mass）G 位于长方形的中心；绳子则视为轻而不可伸长（light and inextensible），所以沿绳的张力 T 处处相同。在 A 处，墙壁施加一个法向反作用力 R（水平方向，背离墙壁）及一个摩擦力 F（铅垂方向）。我们规定 F 以向上为正，所以若 F 为负，即表示摩擦力实际上向下作用。

设坐标轴，以 A 为原点，x 轴水平背离墙壁，y 轴铅垂向上。边 AB 与向上的铅垂线成角 ϕ，所以沿 AB 的单位向量是 $(\sin\phi, \cos\phi)$。边 AD 与之垂直，并指向背离墙壁的方向，所以其单位向量是 $(\cos\phi, -\sin\phi)$。因此

$$G = b(\sin\phi, \cos\phi) + a(\cos\phi, -\sin\phi), \qquad P = d(\sin\phi, \cos\phi),$$

其中 P 是系绳之点。绳子由 P 向上连到墙壁，与向下的铅垂线成角 θ，所以作用于薄板的张力是 $T(-\sin\theta, \cos\theta)$。



第 (i) 部分

沿水平及铅垂方向分解，

$$R = T\sin\theta,$$
$$F + T\cos\theta = W.$$

关于 A 取力矩时，R 和 F 没有贡献。一个作用于 (x, y) 的力 (f_x, f_y) 的逆时针力矩是 $xf_y - yf_x$。对于作用于 P 的张力，这是

$$d\sin\phi \cdot T\cos\theta - d\cos\phi \cdot (-T\sin\theta) = Td\sin(\theta+\phi),$$

而对于作用于 G 的重量 $(0, -W)$，则是 $-W(a\cos\phi + b\sin\phi)$。所以

$$Td\sin(\theta+\phi) = W(a\cos\phi + b\sin\phi). \tag{1}$$

量 $a\cos\phi + b\sin\phi$ 其实就是 G 与墙壁的水平距离，这是看出重量力矩的快捷方法。

在摩擦力向上的极限平衡（limiting equilibrium）中，$F = \mu R = \mu T\sin\theta$，所以铅垂方向的方程变成

$$T(\cos\theta + \mu\sin\theta) = W.$$

把 W 代入 (1)，并约去 T（由于绳子拉紧，它是正数），得

$$d\sin(\theta+\phi) = (\cos\theta + \mu\sin\theta)(a\cos\phi + b\sin\phi),$$

这就是 $(*)$。

第 (ii) 部分

若摩擦力达到极限并向下作用, 则 $F = -\mu R = -\mu T \sin\theta$, 而铅垂方向的方程是 $T(\cos\theta - \mu\sin\theta) = W$。其余一切不变，所以 $(*)$ 变成

$$d\sin(\theta+\phi) = (\cos\theta - \mu\sin\theta)(a\cos\phi + b\sin\phi);$$

即是把 μ 换成 $-\mu$。

第 (iii) 部分

现在我们不假设摩擦力达到极限。方程 (1) 在任何平衡状态下都成立，并给出

$$T = \frac{W(a\cos\phi + b\sin\phi)}{d\sin(\theta+\phi)},$$

而由于 θ 和 ϕ 都是锐角，所以 $\sin(\theta+\phi) > 0$。铅垂方向的方程于是给出

$$F = W - T\cos\theta = W\left(1 - \frac{\cos\theta\,(a\cos\phi + b\sin\phi)}{d\sin(\theta+\phi)}\right).$$

想一想 怎样把“$F > 0$”转化为只涉及正切的条件?

摩擦力向上作用，当且仅当 $F > 0$，即是当且仅当

$$d\sin(\theta+\phi) > \cos\theta\,(a\cos\phi + b\sin\phi),$$

因为 $d\sin(\theta+\phi) > 0$。展开 $\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$，并把两边除以 $\cos\theta\cos\phi$；由于两个角都是锐角，这个数是正数，所以不等式方向不变：

$$d(\tan\theta + \tan\phi) > a + b\tan\phi.$$

由于 $\tan\theta + \tan\phi > 0$，充分必要条件是

$$d > \frac{a + b\tan\phi}{\tan\theta + \tan\phi}.$$

(当 d 等于这个值时，完全没有摩擦力；当 d 较小时，摩擦力向下作用。)

从物理上看，绳子系在边 AB 上，所以 $d \le 2b$。若条件成立，则

$$\frac{a + b\tan\phi}{\tan\theta + \tan\phi} < d \le 2b,$$

所以 $a + b\tan\phi < 2b\tan\theta + 2b\tan\phi$，即 $a < b(2\tan\theta + \tan\phi)$。取逆否命题 (contrapositive)：若 $b(2\tan\theta + \tan\phi) < a$，则系绳点无论放在任何容许的位置，都不能使摩擦力向上作用，所以该条件不可能成立。证明完毕。

验算

- 当 $\mu = 0$ 时墙壁光滑，$(*)$ 与其第 (ii) 部分的版本重合，理应如此。两者都化为第 (iii) 部分的临界情况 $d\sin(\theta+\phi) = \cos\theta\,(a\cos\phi + b\sin\phi)$。
- 薄板的 AB 铅垂时（$\phi \to 0$）：临界情况变成 $d\tan\theta = a$。这表示绳子的作用线（line of action）通过经过 G 的铅垂线，该线与墙壁的水平距离为 a。这时单靠绳子和重量便能互相平衡，墙壁只需施加水平推力。
- 数值上，取 $a = 1$、$b = 1.5$、$\theta = 30°$、$\phi = 40°$：临界值是 $d \approx (1 + 1.259)/(0.577 + 0.839) \approx 1.595 < 3 = 2b$。改为关于 G 而非 A 取力矩，直接解三条平衡方程，得出在略高于此值时 $F > 0$，略低于此值时 $F < 0$。用电脑检验数千个随机配置，证实了 $(*)$、其第 (ii) 部分的形式，以及 (iii) 中的条件。

常见错误

- 关于 P 或 G 取力矩：这样做可行，但会把 R 和 F 带进力矩方程，令代数运算繁复得多。
- 弄错 G 的位置：D 位于背离墙壁的一侧，所以 AD 沿 $(\cos\phi, -\sin\phi)$ 方向，而 G 与墙壁的水平距离是 $a\cos\phi + b\sin\phi$。
- 在最后一部分忘记约束条件 $d \le 2b$；这是 AB 的长度唯一出现的地方。

STEP 之外

这是一个小型的静定（statically determinate）问题：平面上的刚体有三个未知的力和三条独立的平衡方程。工程师首先就会核对这个数目；若支撑点更多，单靠静力学便无法求出各力，还需要知道物体如何变形。关于未知力的作用点取力矩这个技巧，是巧妙选取力矩原点的平面版本。如欲深入了解，可参阅 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP II 2014，第 10 题

力学 · 轨道的包络线

STEP II 2014，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/14-s2-q10

本题在 *Beyond STEP* 第 11.9 节安全抛物线中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/safety-parabola.html

关键思路

消去时间后，轨迹（trajectory）方程是一个关于 $\lambda = \tan\theta$ 的二次方程（*quadratic*），因为 $\sec^2\theta = 1 + \lambda^2$。固定一点 (x, y)，然后问：是否存在某个发射角，使轨迹经过这一点？这等同于问这个关于 λ 的二次方程是否有实根，所以判别式（discriminant）就给出可到达区域的边界。至于最后一部分，边界曲线原来是一条焦点（focus）在 O 的抛物线，这使得计算与 O 的距离变得非常容易。

解答

轨迹

把抛射体（projectile）视为在重力作用下自由运动的质点（particle），不计空气阻力。经过时间 t 后，

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

由于 $\cos\theta \neq 0$（该角小于直角），$t = x/(u\cos\theta)$，代入得

$$y = x\tan\theta - \frac{gx^2}{2u^2\cos^2\theta}.$$

利用 $\sec^2\theta = 1 + \tan^2\theta = 1 + \lambda^2$，得笛卡儿方程（Cartesian equation）

$$\boxed{y = \lambda x - \frac{gx^2}{2u^2}\left(1 + \lambda^2\right).}$$

给定 x 上方的最大高度 Y

固定 u，并固定 $x > 0$。此时 y 只是 λ 的函数，可以改写为

$$\frac{gx^2}{2u^2}\lambda^2 - x\lambda + \left(\frac{gx^2}{2u^2} + y\right) = 0.$$

想一想 为甚么要把它看成关于 λ 的方程，而不是关于 y 的方程?

某个值 y 会在这个 x 的上方出现，当且仅当有某个发射角能产生它，即这个二次方程有实根 λ。因此其判别式必须非负：

$$x^2 - 4 \cdot \frac{gx^2}{2u^2}\left(\frac{gx^2}{2u^2} + y\right) \geq 0.$$

除以 $x^2 > 0$，再乘以 $u^2/(2g) > 0$，这不会改变不等式的方向，得

$$\frac{u^2}{2g} \geq \frac{gx^2}{2u^2} + y \iff y \leq \frac{u^2}{2g} - \frac{gx^2}{2u^2}.$$

所以

$$\boxed{Y = \frac{u^2}{2g} - \frac{gx^2}{2u^2}}$$

而且没有任何轨迹在这个 x 的上方高于 Y。

我们还需要证明*不超过* Y 的每一个值都能以真正的发射角（即 $\lambda > 0$）达到。作为 λ 的函数，y 是开口向下的二次函数。其顶点（vertex）在 $\lambda_0 = u^2/(gx) > 0$，此处 $y = Y$（代入验证即可）。当 $\lambda \geq \lambda_0$ 时，y 连续（continuous）且递减，并且当 $\lambda \to \infty$ 时 $y \to -\infty$。所以由连续性，每个值 $y \leq Y$ 都会在某个 $\lambda \geq \lambda_0 > 0$ 取得，即在某个 $\theta \in (0, \frac{\pi}{2})$ 取得。因此 y 的取值正是整个区间 $y \leq Y$，符合所求。（当 $x = 0$ 时，我们必须垂直向上抛。此时 y 由 0 变到 $u^2/(2g)$，同样等于 $Y(0)$。）

若容许向任一方向抛射，则 $x < 0$ 的情况是镜像，同一公式依然成立，因为 Y 只取决于 x^2。

图像与可到达区域

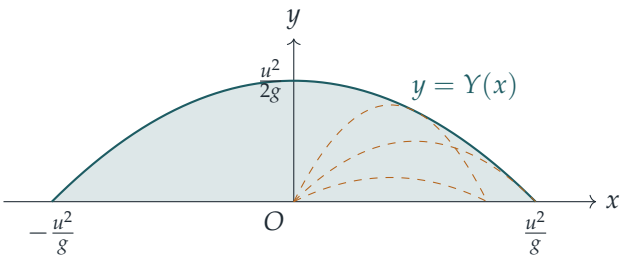
Y 的图像是一条开口向下的抛物线。其顶点在 $(0,\, u^2/2g)$，即可能达到的最大高度；它与地面相交于 $x = \pm u^2/g$，即可能达到的最大射程。每一条轨迹都位于这条曲线之下，并在 $\lambda = \lambda_0$ 处与它相切一次。因此它常被称为*包络抛物线*（bounding parabola）或*安全抛物线*（parabola of safety）。

质点一碰到地面便停下。取一条 $\lambda > 0$ 的轨迹上一点，其中 $y \geq 0$ 且 $x > 0$。在 O 与该点之间的所有 x，这条轨迹都在地面之上，所以质点确实会在落地前到达该点。因此可到

达的集合是区域

$$0 \le y \le \frac{u^2}{2g} - \frac{gx^2}{2u^2}, \qquad |x| \le \frac{u^2}{g},$$

即下图的阴影部分。图中以 u^2/g 为单位，虚线是三条示例轨迹。



与 O 的最远距离

任何可到达的点 P 都位于由 O 到点 Q 的线段上，其中 Q 是射线 OP 与边界的交点。该区域以包络抛物线和地面为边界，而且沿每条射线都包含 O 与边界之间的所有点。所以 $OP \le OQ$，我们只需在边界上求距离的最大值。在地面上，最远的点距离为 u^2/g。在曲线部分，$y = Y(x)$，所以

$$\begin{aligned} OQ^2 = x^2 + Y^2 &= x^2 + \left(\frac{u^2}{2g}\right)^2 - 2 \cdot \frac{u^2}{2g} \cdot \frac{gx^2}{2u^2} + \left(\frac{gx^2}{2u^2}\right)^2 \\ &= \left(\frac{u^2}{2g}\right)^2 + \frac{x^2}{2} + \left(\frac{gx^2}{2u^2}\right)^2 = \left(\frac{u^2}{2g} + \frac{gx^2}{2u^2}\right)^2. \end{aligned}$$

所以 $OQ = \dfrac{u^2}{2g} + \dfrac{gx^2}{2u^2} = \dfrac{u^2}{g} - Y$。换句话说，与 O 的距离等于向下到直线 $y = u^2/g$ 的距离：O 是包络抛物线的焦点，而 $y = u^2/g$ 是它的准线（directrix）。当 Y 最小时这个距离最大，而在曲线的可到达部分上，Y 的最小值是 $Y = 0$，位于 $x = \pm u^2/g$。因此抛射体能到达的与 O 的最远距离是

$$\boxed{\frac{u^2}{g}},$$

这正是最大水平射程，在 $\theta = \frac{\pi}{4}$ 时达到。边界上每个高度 $y > 0$ 的点都较近，恰好近 y。

验算

- 代入 $x = 0$：$Y = u^2/(2g)$，即垂直向上抛时达到的高度（由 $v^2 = u^2 - 2gh$ 得出）。代入 $Y = 0$：$x = u^2/g$，即熟悉的最大射程 $u^2 \sin 2\theta/g$（在 $\theta = \frac{\pi}{4}$ 时）。
- 取 $u^2/g = 10$（例如 $u = \sqrt{98}\,\mathrm{m\,s^{-1}}, g = 9.8$）。在 $x = 4$ 时，公式给出 $Y = 5 - 0.8 = 4.2$，在 $\lambda_0 = 10/4 = 2.5$ 时达到：确实 $2.5 \cdot 4 - \frac{1}{20} \cdot 16 \cdot 7.25 = 10 - 5.8 = 4.2$。

- 以电脑对大量发射角和时间进行搜索，取 $u = 10\ \mathrm{m\,s^{-1}}$，找到与 O 最远的点距离为 $10.204\ \mathrm{m} = u^2/g$，位于地面上。

常见错误

- 只做到判别式便停下。判别式只显示 $y \le Y$ 是必要的，但题目还要求每个这样的值都能以位于地面之上的真正发射角达到。
- 把 $y = Y$ 之下的整个区域都涂上阴影，包括地面以下的点。质点落地便停下，所以可到达的集合还需满足 $y \ge 0$。
- 以为最远的点是安全抛物线的顶点。该点与 O 的距离只有 $u^2/(2g)$，是射程的一半。

STEP 之外

包络抛物线是这族轨迹的*包络线*（envelope）：即与族中每一条曲线都相切的曲线。求包络线的一个标准方法，是在 $F(x, y, \lambda) = 0$ 与 $\partial F/\partial\lambda = 0$ 之间消去参数，而这正是我们所用的重根条件。这里还有一个与能量的联系：每条轨迹都有同一条准线，即直线 $y = u^2/(2g)$，那正是全部动能（kinetic energy）都转化为势能（potential energy）时的高度。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）以大学一年级程度讲解抛射运动与能量。

详细解答

STEP II 2014，第 11 题

力学 · 圆周运动

STEP II 2014，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/14-s2-q11

关键思路

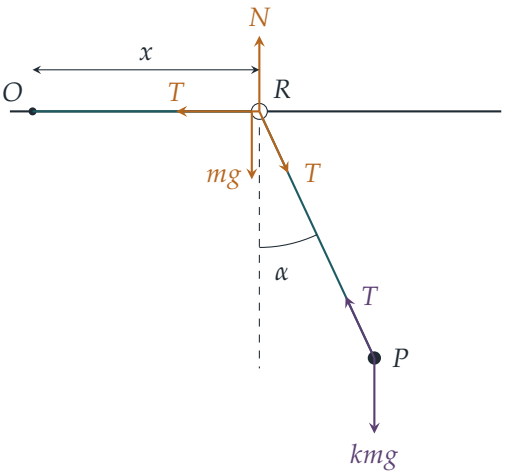
整个问题只由一个变量 x 主导。由于绳子不可伸长（inextensible），而 α 固定不变，P 的位置是 x 的一个简单函数（function），因此 P 的两个加速度分量都是 $\ddot{x}$ 的倍数。这样便得到三条运动方程（equations of motion）（两条关于 P，一条关于环的水平方向），但未知数只有两个：$\ddot{x}$ 和 T。超定方程组（overdetermined system）只有在某个特殊条件成立时才相容，而正是这个相容条件（consistency condition）把 α 和 k 联系起来。

解答

模型。环和 P 均视为质点（particle），绳子轻而不可伸长，环和轨道均光滑（smooth）。由于环是光滑的，绳子两段的张力（tension）T 相同。取 O 为原点，x 轴沿轨道、方向由 O 指向环，y 轴竖直向上。

P 的坐标。绳子由 O 到环的一段沿轨道，长度为 x，所以由环到 P 的一段长度为 $L-x$。这一段与向下的铅垂线成角 α，而 P 悬在环远离 O 的一侧（我们会在第 (i) 部分末尾说明为何如此选取）。因此

$$P = \big(x + (L-x)\sin\alpha,\ -(L-x)\cos\alpha\big).$$



第 (i) 部分

在 α 为常数的情况下，把 P 的坐标对时间微分两次，得

$$\ddot{x}_P = \ddot{x}\,(1-\sin\alpha), \qquad \ddot{y}_P = \ddot{x}\cos\alpha.$$

作用于 P 的力有向下的重量 kmg，以及沿绳子指向环的张力 T；张力的水平分量为 $-T\sin\alpha$，竖直分量为 $+T\cos\alpha$。

P，竖直方向： $km\,\ddot{y}_P = T\cos\alpha - kmg$，即

$$km\ddot{x}\cos\alpha = T\cos\alpha - kmg,$$

与所求相符。

P，水平方向：

$$km\ddot{x}\,(1-\sin\alpha) = -T\sin\alpha. \tag{1}$$

环，水平方向。环两边都受到绳子的作用：沿轨道的一段以力 T 把环拉向 O，而往下连到 P 的一段则以水平分量 $T\sin\alpha$ 把环拉离 O。环的重量和轨道的法向反作用力（normal reaction）都是竖直的（轨道光滑）。因此

$$m\ddot{x} = T\sin\alpha - T = -T(1-\sin\alpha). \tag{2}$$

想一想 为甚么 P 必须在环远离 O 的一侧?

假如 P 悬在靠近 O 的一侧，它的 x 坐标便是 $x-(L-x)\sin\alpha$，两条水平方程会变成 $km\ddot{x}(1+\sin\alpha) = T\sin\alpha$ 和 $m\ddot{x} = -T(1+\sin\alpha)$。第一条使 $\ddot{x}\geq 0$，第二条使 $\ddot{x}\leq 0$，于是 $\ddot{x}=0$ 且 $T=0$；但这样竖直方程便给出 $0=-kmg$，这是不可能的。所以那种位形不可能在 α 不变的情况下运动，上面的坐标才是正确的。

第 (ii) 部分

首先，$T \neq 0$：若 $T = 0$，则 (2) 给出 $\ddot{x} = 0$，而竖直方程给出 $0 = -kmg$，矛盾。

由 (2)，$m\ddot{x} = -T(1 - \sin\alpha)$。代入 (1)：

$$k\big(-T(1-\sin\alpha)\big)(1-\sin\alpha) = -T\sin\alpha.$$

两边除以 $-T \neq 0$,

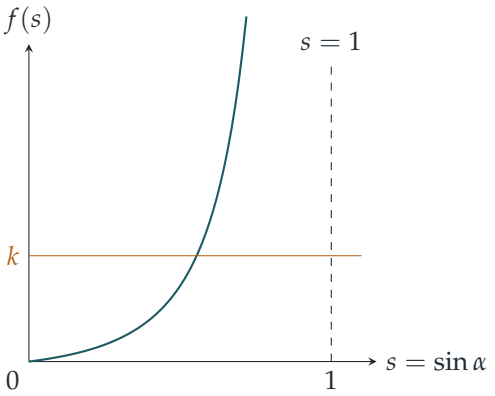
$$\frac{\sin\alpha}{(1-\sin\alpha)^2} = k,$$

与所求相符。（由于 α 是锐角（acute angle），$0 < \sin\alpha < 1$，所以除以 $(1-\sin\alpha)^2$ 是没有问题的。）

每一个 k 都可以出现。设 $s = \sin\alpha$，并在 $0 < s < 1$ 上定义 $f(s) = \dfrac{s}{(1-s)^2}$，这恰好涵盖所有 $\alpha \neq 0$ 的锐角 α。由商法则（quotient rule），

$$f'(s) = \frac{(1-s)^2 + 2s(1-s)}{(1-s)^4} = \frac{1+s}{(1-s)^3} > 0,$$

所以 f 严格递增（strictly increasing）。此外，当 $s \to 0^+$ 时 $f(s) \to 0$，当 $s \to 1^-$ 时 $f(s) \to \infty$，而且 f 连续（continuous）。因此 f 恰好取每个正值一次，即对每个 $k > 0$，恰有一个锐角 α 满足该条件。



无论水平线 $y = k$ 的高度是多少，它都恰好与曲线相交一次。

第 (iii) 部分

现在利用 (2) 和竖直方程消去 T。由 (2)，$T = -\dfrac{m\ddot{x}}{1-\sin\alpha}$。代入竖直方程并除以 m：

$$k\ddot{x}\cos\alpha = -\frac{\ddot{x}\cos\alpha}{1-\sin\alpha} - kg \quad\Longrightarrow\quad \ddot{x}\cos\alpha\left(k + \frac{1}{1-\sin\alpha}\right) = -kg.$$

再利用第 (ii) 部分替换括号内的 k：

$$k+\frac{1}{1-\sin\alpha}=\frac{\sin\alpha+(1-\sin\alpha)}{(1-\sin\alpha)^2}=\frac{1}{(1-\sin\alpha)^2}.$$

因此

$$\ddot{x}=-\frac{kg(1-\sin\alpha)^2}{\cos\alpha}=-\frac{g\sin\alpha}{\cos\alpha}=-g\tan\alpha,$$

这里再一次用了 $k(1-\sin\alpha)^2=\sin\alpha$。这就是所求的结果。

额外一提，$T=-\dfrac{m\ddot{x}}{1-\sin\alpha}=\dfrac{mg\tan\alpha}{1-\sin\alpha}>0$，所以绳子确实是拉紧的，符合模型的要求。

验算

- 取 $k=1$，条件 $s=(1-s)^2$ 给出 $\sin\alpha=\frac{1}{2}(3-\sqrt{5})\approx 0.382$，所以 $\alpha\approx 22.46^\circ$，$\ddot{x}\approx -0.413g$。把这些代入 (1)、(2) 和竖直方程，三条都成立。
- 我们以数值方法模拟了完整的双变量运动（以 x 和绳子的角度为变量），从这个角度开始并让环在运动中：角度在 10^{-10} rad 的误差内保持不变，而环的减速度恰好是 $g\tan\alpha$。若从稍为不同的角度开始，角度会立即开始改变，可见 α 这个特殊值确实关键。
- 极限情况：当 $k\to 0$（质点非常轻），$\alpha\to 0$ 且 $\ddot{x}\to 0$；当 $k\to\infty$，$\alpha\to 90^\circ$，减速度无限增大，因为轻环正被一个重质点拖动。
- 正负号：$\ddot{x}<0$, 所以环的加速度总是指向 O。这是合理的：环所受的水平合力 $-T(1-\sin\alpha)$ 指向 O。

常见错误

- 忘记环两边都受到绳子的作用；正是沿轨道指向 O 的拉力令 (2) 成立。
- 把 P 的水平加速度写成 $\ddot{x}$；悬垂部分的长度 $L-x$ 也在改变，这会贡献 $-\ddot{x}\sin\alpha$ 一项。
- 微分时把 α 当作变量。这里它是常数，而这正是问题得以处理的原因。

STEP 之外

这是一个有两个自由度（degrees of freedom）的系统（环的位置和绳子的角度），题目要找的是其中一个自由度保持不变的特殊解，颇有相对平衡（relative equilibrium）的意味。在大学一年级的课程中，你会以 x 和角度写出拉格朗日量（Lagrangian），并发现只有在这里求得的那一个 α 值下，角度不变的运动才有可能；接下来自然要问的是，这种

运动是否稳定（stable）。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对这类约束系统（constrained systems）有很好的讲解。

详细解答

STEP II 2014，第 12 题

概率与统计 · 风险函数

> **STEP II 2014，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/14-s2-q12

关键思路

风险函数（hazard function）是对仍然存活者而言的死亡率。后面各部分全都源于一个观察：由于 $F'(t) = f(t)$，

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln\bigl(1 - F(t)\bigr).$$

所以知道 h，就等于知道 $\ln(1 - F)$ 的导数（derivative）。先积分，再利用寿命开始时 $F = 0$ 这一事实，最后对 F 求导便得到 f。

解答

第 (i) 部分

设 A 为事件 $T \geq t$（苍蝇在时间 t 仍然存活），B 为事件 $t \leq T \leq t + \delta t$（苍蝇在接下来的 δt 内死亡）。由于 $B \cap A = B$（对连续随机变量（continuous random variable）而言 $\mathrm{P}(T = t) = 0$，所以端点是否包括在内并无影响），

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(B)}{\mathrm{P}(A)} = \frac{F(t + \delta t) - F(t)}{1 - F(t)}.$$

当 δt 很小时，$F(t + \delta t) - F(t) \approx F'(t)\,\delta t = f(t)\,\delta t$，因此

$$\mathrm{P}(B \mid A) \approx \frac{f(t)\,\delta t}{1 - F(t)} = h(t)\,\delta t,$$

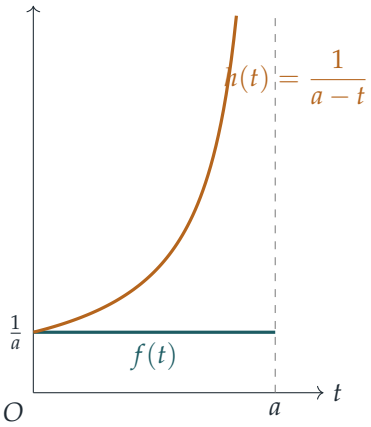
此即所求。（我们需要 $F(t) < 1$，以免以一个不可能事件作为条件；而这正是 h 有定义的范围。）

第 (ii) 部分

这里当 $0 < t < a$ 时 $f(t) = F'(t) = \frac{1}{a}$：$T$ 在 $(0, a)$ 上服从均匀分布（uniform distribution）。于是

$$h(t) = \frac{1/a}{1 - t/a} = \frac{1}{a - t}, \qquad 0 < t < a.$$

概率密度是平的，但风险函数由 $h(0) = \frac{1}{a}$ 开始递增，且当 $t \to a^-$ 时 $h(t) \to \infty$。这是合理的：一只存活至接近时间 a 的苍蝇，必定会在余下极短的时间内死亡。



（图中取 $a = 2$。）

第 (iii) 部分

利用关键思路，当 $t > a$ 时 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln\big(1 - F(t)\big) = \dfrac{1}{t}$。由 a 积分至 t，并因 T 只取大于 a 的值而有 $F(a) = 0$，得

$$-\ln\big(1 - F(t)\big) = \ln t - \ln a = \ln\frac{t}{a}, \qquad \text{所以} \qquad 1 - F(t) = \frac{a}{t}.$$

因此 $F(t) = 1 - \dfrac{a}{t}$，且

$$f(t) = \frac{a}{t^2}, \qquad t > a$$

（而当 $t \le a$ 时 $f(t) = 0$）。验算一下：$\int_a^\infty at^{-2}\,\mathrm{d}t = 1$，且 $f/(1 - F) = (a/t^2)/(a/t) = 1/t$。

第 (iv) 部分

“若”的方向。设当 $t > b$ 时 $f(t) = k\mathrm{e}^{-k(t-b)}$，其中 $k > 0$。由于 $\int_b^\infty k\mathrm{e}^{-k(t-b)}\,\mathrm{d}t = 1$，全部概率都落在 $t > b$，所以当 $t \le b$ 时 $f(t) = 0$ 及 $F(t) = 0$。因此当 $t \le b$ 时 $h(t) = 0/1 = 0$。当 $t > b$ 时，

$$F(t) = \int_b^t k\mathrm{e}^{-k(s-b)}\,\mathrm{d}s = 1 - \mathrm{e}^{-k(t-b)}, \qquad h(t) = \frac{k\mathrm{e}^{-k(t-b)}}{\mathrm{e}^{-k(t-b)}} = k,$$

这是一个常数。

“仅当”的方向。设当 $t \le b$ 时 $h(t) = 0$，而当 $t > b$ 时 $h(t) = k$（常数）。

> ? **想一想** 当 $t \le b$ 时 $h(t) = 0$，这告诉我们 $F(b)$ 是甚么？

在 $h(t) = 0$ 之处，有 $f(t) = h(t)\big(1 - F(t)\big) = 0$。所以对所有 $t \le b$，f 都等于零，从而当 $t \le b$ 时 $F(t) = \int_{-\infty}^{t} f = 0$；特别地 $F(b) = 0$。（这亦确认了该处 $F < 1$，所以 h 在 $t \le b$ 上确实有定义。）

当 $t > b$ 时，$-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln\big(1 - F(t)\big) = k$。由 b 积分至 t，并利用 $F(b) = 0$，

$$-\ln\big(1 - F(t)\big) = k(t - b), \qquad F(t) = 1 - \mathrm{e}^{-k(t-b)},$$

求导后得当 $t > b$ 时 $f(t) = k\mathrm{e}^{-k(t-b)}$。最后，$k$ 必须是正数：因为 $f \ge 0$，所以 $k \ge 0$；而若 $k = 0$，则 F 会永远停留在 0，于是 $F(t) \not\to 1$，T 便不是一个真正的随机变量。至此两个方向都已证明。

第 (v) 部分

同样地，当 $t > 0$ 时 $-\dfrac{\mathrm{d}}{\mathrm{d}t}\ln\big(1 - F(t)\big) = \dfrac{\lambda}{\theta^{\lambda}}t^{\lambda-1}$。由 0 积分至 t，因 $T > 0$ 而有 $F(0) = 0$，并注意 $\lambda > 0$ 使得当 $t \to 0$ 时 $t^{\lambda} \to 0$，

$$-\ln\big(1 - F(t)\big) = \frac{\lambda}{\theta^{\lambda}} \cdot \frac{t^{\lambda}}{\lambda} = \left(\frac{t}{\theta}\right)^{\lambda}.$$

所以 $F(t) = 1 - \exp\big(-(t/\theta)^{\lambda}\big)$，求导得

$$f(t) = \frac{\lambda}{\theta^{\lambda}}\, t^{\lambda-1} \exp\!\left(-\left(\frac{t}{\theta}\right)^{\lambda}\right), \qquad t > 0.$$

这是一个真正的概率密度函数（probability density function）：它是正的，而且当 $t \to \infty$ 时 $F(t) \to 1$。

验算

- 每个答案都可以直接计算 $f/(1 - F)$ 来验证。例如在 (v) 中，指数因子互相抵消，剩下 $\lambda\theta^{-\lambda}t^{\lambda-1}$。
- 在第 (v) 部分取 $\lambda = 1$，得常数风险 $1/\theta$ 及 $f(t) = \frac{1}{\theta}\mathrm{e}^{-t/\theta}$，这正是第 (iv) 部分取 $b = 0$、$k = 1/\theta$ 的情况。
- 数值上，取 $\lambda = \frac{1}{2}$、$\theta = 2$，将 f 在 $(0, \infty)$ 上积分得 1.0000，而 $F(1) = 1 - \mathrm{e}^{-1/\sqrt{2}} \approx 0.5069$ 与直接将 f 由 0 积分至 1 的结果吻合。

常见错误

- 忘记寿命开始时 $F = 0$ 这个条件，以致 F 中留下一个任意的积分常数。
- 在 (iv) 中只证明了“当且仅当”（if and only if）的其中一个方向，或者没有利用当 $t \leq b$ 时 $h = 0$ 来推出 $F(b) = 0$。
- 漏掉负号：$\frac{\mathrm{d}}{\mathrm{d}t} \ln(1 - F) = -f/(1 - F)$。

STEP 之外

关系式 $1 - F(t) = \exp\left(-\int_0^t h(s)\,\mathrm{d}s\right)$ 是*存活分析*（survival analysis）及可靠性理论（reliability theory）的基础。(v) 中的分布是韦伯分布（Weibull distribution）：$\lambda < 1$ 给出递减的风险（早期故障），$\lambda = 1$ 给出具无记忆性（memoryless）的指数分布（exponential distribution），而 $\lambda > 1$ 则给出递增的风险（损耗老化）。存活分析的应用范围极广，由医学临床试验到工程学都有。关于其背后的概率论，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）。

详细解答

STEP II 2014，第 13 题

概率与统计 • 离散随机变量

STEP II 2014，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/14-s2-q13

关键思路

这其实是乔装了的生日问题（birthday problem）。事件 $X \geq k$ 不过是说首 $k-1$ 个整数互不相同，而 $X = r$ 则是说首 $r-1$ 个互不相同，而第 r 个与其中一个相同。两个概率（probability）都是一些简单因子的乘积。至于那两条奇特的恒等式，则来自以两种方法计算同一个量：第 (i) 部分是总概率等于 1，第 (iv) 部分是 $\mathrm{E}(X)$ 的两条公式相等。

解答

全文中记

$$Q_m = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right) \qquad (m \geq 1),$$

其中 $Q_1 = 1$（空积（empty product））。由于每个整数都与之前的整数独立（independent），并在 $\{1, \ldots, n\}$ 上均匀分布（uniform），Q_m 就是 $I_1, \ldots, I_m$ 互不相同的概率：已知首 j 个互不相同，下一个避开它们全部的概率是 $1 - \frac{j}{n}$。

$\mathrm{P}(X = 4)$ 的值

我们需要 I_1, I_2, I_3 互不相同，然后 I_4 等于这三个之中的一个：

$$\mathrm{P}(X = 4) = 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \frac{3}{n} = \frac{3(n-1)(n-2)}{n^3}.$$

第 (i) 部分

同样地，$X = r$ 表示 $I_1, \ldots, I_{r-1}$ 互不相同（概率为 Q_{r-1}），而 I_r 是这 $r-1$ 个值之一（概率为 $\frac{r-1}{n}$，且独立）。所以

$$\mathrm{P}(X = r) = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{r-2}{n}\right)\frac{r-1}{n} = Q_{r-1}\,\frac{r-1}{n}, \qquad 2 \le r \le n+1.$$

想一想 为甚么这些概率加起来等于 1?

X 必定存在，并且介乎 2 与 $n+1$ 之间：第一个整数不可能是重复，所以 $X \ge 2$；而 $I_1, \ldots, I_{n+1}$ 共有 $n+1$ 个整数，却只取 n 个值，所以由抽屉原理（pigeonhole principle），到第 $n+1$ 个时必有某个值出现了两次，故 $X \le n+1$。（数列永不重复的机会是零，所以没有遗漏任何情况。）因此

$$\sum_{r=2}^{n+1} \mathrm{P}(X = r) = 1.$$

把 $r = 2, 3, 4, \ldots$ 的项逐一写出，得

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

这正是所求的恒等式。级数（series）在 $r = n+1$ 终止，之后的任何项都是零，因为它们含有因子 $1 - \frac{n}{n}$。证毕。令 $m = r - 1$，上式可写成

$$\sum_{m=1}^{n} \frac{m}{n}\,Q_m = 1. \qquad (*)$$

第 (ii) 部分

由期望值（expectation）的定义，

$$\mathrm{E}(X) = \sum_{r=2}^{n+1} r\,\frac{r-1}{n}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{r-2}{n}\right) = \sum_{m=1}^{n} \frac{m(m+1)}{n}\,Q_m.$$

第 (iii) 部分

$X \ge k$ 恰好在 $I_1, \ldots, I_{k-1}$ 之中没有重复时发生，即这 $k-1$ 个整数互不相同。所以

$$\mathrm{P}(X \ge k) = Q_{k-1} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-2}{n}\right), \qquad 2 \le k \le n+2,$$

而 $\mathrm{P}(X \ge 1) = \mathrm{P}(X \ge 2) = 1$。（当 $k = n+2$ 时，乘积含有 $1 - \frac{n}{n} = 0$，理应如此。）

第 (iv) 部分

记 $p_r = \mathrm{P}(Y = r)$。由于 $r = \sum_{k=1}^{r} 1$，所以

$$\mathrm{E}(Y) = \sum_{r=1}^{N} r\, p_r = \sum_{r=1}^{N} \sum_{k=1}^{r} p_r = \sum_{k=1}^{N} \sum_{r=k}^{N} p_r = \sum_{k=1}^{N} \mathrm{P}(Y \geq k).$$

由于是有限和，交换求和次序并无问题：两个二重和都遍历所有满足 $1 \leq k \leq r \leq N$ 的数对 (k, r)。证毕。

> **想一想** 我们应该比较 $\mathrm{E}(X)$ 的哪两个表达式?

把这结果应用于 $Y = X$，它在 $\{1, \ldots, N\}$ 中取值，其中 $N = n + 1$（取值 1 的概率只是 0 而已）。利用第 (iii) 部分，

$$\mathrm{E}(X) = \sum_{k=1}^{n+1} \mathrm{P}(X \geq k) = 1 + \sum_{k=2}^{n+1} Q_{k-1} = 1 + \sum_{m=1}^{n} Q_m.$$

现在令这式等于第 (ii) 部分的结果，把 $\frac{m(m+1)}{n} = \frac{m^2}{n} + \frac{m}{n}$ 拆开，并利用 $(*)$：

$$\sum_{m=1}^{n} \frac{m(m+1)}{n} Q_m = \sum_{m=1}^{n} \frac{m^2}{n} Q_m + \underbrace{\sum_{m=1}^{n} \frac{m}{n} Q_m}_{=1 \text{ 由 } (*)}.$$

所以 $1 + \sum_{m=1}^{n} Q_m = \sum_{m=1}^{n} \frac{m^2}{n} Q_m + 1$，即

$$\sum_{m=1}^{n} Q_m \left(1 - \frac{m^2}{n}\right) = 0.$$

$m = 1, 2, 3$ 的项分别是 $\left(1 - \frac{1^2}{n}\right)$、$\left(1 - \frac{1}{n}\right)\left(1 - \frac{2^2}{n}\right)$ 和 $\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\left(1 - \frac{3^2}{n}\right)$，所以这正是所求的恒等式。与第 (i) 部分一样，$m = n$ 之后的每一项都含有因子 $1 - \frac{n}{n} = 0$。证毕。

验算

- 当 $n = 1$ 时，$X = 2$ 必然成立。第 (i) 部分给出 $\frac{1}{1} = 1$，而最后的和只有一项 $1 - 1 = 0$。
- 当 $n = 4$ 时，最后的和是 $\frac{3}{4} + \frac{3}{4} \cdot 0 + \frac{3}{4} \cdot \frac{1}{2} \cdot \left(-\frac{5}{4}\right) + \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot (-3) = \frac{24 - 15 - 9}{32} = 0$，而两条公式都给出 $\mathrm{E}(X) = \frac{103}{32}$。
- 当 $n = 365$（生日）时，$\mathrm{E}(X) = 1 + \sum Q_m \approx 24.62$。以电脑模拟 10^5 个数列，得平均值 24.61；而对 $n \leq 5$ 作完全穷举（exact enumeration），亦证实了上述每一条公式。

常见错误

- 写 $\mathrm{P}(X = r)$ 时用了 $r-1$ 个形如 $1-\frac{j}{n}$ 的因子，而不是 $r-2$ 个。只有第二至第 $(r-1)$ 个整数须避开之前的值。
- 在 $\sum \mathrm{P}(X \geq k)$ 中漏掉 $k = 1$ 的项。这一项贡献 1，正好与 $(*)$ 互相抵消。

STEP 之外

尾和公式（tail-sum formula）$\mathrm{E}(Y) = \sum_k \mathrm{P}(Y \geq k)$ 对任何取非负整数值的随机变量（random variable）都成立，而它的连续版本 $\mathrm{E}(T) = \int_0^\infty \mathrm{P}(T > t)\,\mathrm{d}t$ 是大学一年级概率论的标准工具。当 n 很大时，$\mathrm{E}(X) \approx \sqrt{\pi n/2}$，这解释了为何杂凑函数（hash function）的碰撞（collision）只需约 $\sqrt{n}$ 次尝试便会出现（即“生日攻击”（birthday attack））。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）涵盖了这两个概念。

详细解答

STEP III 2014，第 1 题

纯数学 · 对数级数展开

STEP III 2014，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/14-s3-q1

关键思路

三次多项式（cubic）$1+qx^2+rx^3$ 可因式分解为 $(1+ax)(1+bx)(1+cx)$，因此它的对数（logarithm）可拆成三个对数之和，而每个对数的级数（series）我们都已知道。这样便能以 a、b、c 表示 x^n 的系数（coefficient）。我们亦可把同一个对数直接展开为 $\ln(1+u)$，其中 $u=qx^2+rx^3$，从而以 q 和 r 表示各系数。比较这两个展开式，便可把关于 q 和 r 的事实转化为各 S_n 之间的恒等式（identity）。

解答

展开得

$$(1+ax)(1+bx)(1+cx)=1+(a+b+c)x+(ab+bc+ca)x^2+abc\,x^3.$$

由于 $a+b+c=0$，x 项消失。两个对所有实数 x 都相等的多项式，其系数必然相等，因此把 x^2 和 x^3 的系数与 $1+qx^2+rx^3$ 比较，得

$$q=bc+ca+ab,\qquad r=abc,$$

此即所求。

第 (i) 部分

当 $|x|$ 足够小（具体而言，$|x|<1/\max(|a|,|b|,|c|)$；若 $a=b=c=0$，则任何 x 皆可），每个因子都是正数，而每个对数都有收敛（convergent）的级数。利用 $|t|<1$ 时

$\ln(1+t) = \sum_{n\geq 1}(-1)^{n+1}t^n/n$，

$$\begin{aligned}\ln(1+qx^2+rx^3) &= \ln(1+ax)+\ln(1+bx)+\ln(1+cx)\\ &= \sum_{n\geq 1}(-1)^{n+1}\frac{a^nx^n}{n}+\sum_{n\geq 1}(-1)^{n+1}\frac{b^nx^n}{n}+\sum_{n\geq 1}(-1)^{n+1}\frac{c^nx^n}{n}\\ &= \sum_{n\geq 1}(-1)^{n+1}\frac{a^n+b^n+c^n}{n}x^n.\end{aligned}$$

所以 x^n 的系数是 $(-1)^{n+1}S_n$，此即所求。（当 $n=1$ 时，这是 $S_1 = a+b+c = 0$，与没有 x 项的事实吻合。）

第 (ii) 部分

现在用第二种方法展开同一个函数。设 $u = qx^2+rx^3$，则

$$\ln(1+u) = u - \frac{u^2}{2}+\frac{u^3}{3}-\frac{u^4}{4}+\cdots.$$

u 的每一项次数至少为 2，所以 u^k 只涉及 x^{2k} 至 x^{3k} 的幂。要求 x^n 的系数，只需考虑满足 $2k\leq n\leq 3k$ 的 u^k：

$$\begin{aligned}u &= qx^2+rx^3,\\ u^2 &= q^2x^4+2qr\,x^5+r^2x^6.\end{aligned}$$

因此：

- x^2：只有 u 有贡献，得 q；
- x^3：只有 u 有贡献，得 r；
- x^5：只有 $-\frac{1}{2}u^2$ 有贡献，得 $-\frac{1}{2}(2qr) = -qr$。

与第 (i) 部分比较，该处的系数分别是 $-S_2$、S_3 和 S_5，故

$$S_2 = -q, \qquad S_3 = r, \qquad S_5 = -qr.$$

因此 $S_2S_3 = (-q)(r) = -qr = S_5$，此即所求。

想一想 为甚么我们可以比较同一个函数的两个不同级数的系数?

两个展开式都是 x 的幂级数（power series），并在 0 附近的某个区间上收敛至 $\ln(1+qx^2+rx^3)$。在 0 附近收敛的幂级数，其系数是唯一的（它们就是函数在 0 处的导数除以 $n!$，正如麦克劳林定理（Maclaurin's theorem）所述），因此两组系数必然相同。

第 (iii) 部分

对于 x^7，我们需要 $2k \le 7 \le 3k$，这迫使 $k = 3$。用二项式定理（binomial theorem）展开 $u^3 = (qx^2 + rx^3)^3$，

$$u^3 = q^3x^6 + 3q^2r\,x^7 + 3qr^2x^8 + r^3x^9,$$

所以 $\ln(1+u)$ 中 x^7 的系数是 $\frac{1}{3} \cdot 3q^2r = q^2r$。由第 (i) 部分，该系数是 $(-1)^8 S_7 = S_7$，所以 $S_7 = q^2r$。利用第 (ii) 部分，

$$S_2S_5 = (-q)(-qr) = q^2r = S_7,$$

此即所求。

第 (iv) 部分

对于 x^9，我们需要 $2k \le 9 \le 3k$，所以 $k = 3$ 或 $k = 4$。

- 来自 $\frac{1}{3}u^3$：u^3 的 x^9 项是 r^3x^9，得 $\frac{1}{3}r^3$。
- 来自 $-\frac{1}{4}u^4$：在 $(qx^2 + rx^3)^4$ 中，x^9 项需要三个 qx^2 因子和一个 rx^3 因子，得 $\binom{4}{1}q^3r\,x^9 = 4q^3r\,x^9$。这贡献 $-q^3r$。

所以 x^9 的系数是 $\frac{1}{3}r^3 - q^3r$，而由第 (i) 部分，它等于 $(-1)^{10}S_9 = S_9$。另一方面，

$$S_2S_7 = (-q)(q^2r) = -q^3r.$$

因此

$$S_9 - S_2S_7 = \tfrac{1}{3}r^3 = \tfrac{1}{3}(abc)^3,$$

而命题 $S_2S_7 = S_9$ 成立当且仅当 $abc = 0$。一般而言它并不成立。

举一个具体的反例（counterexample），取 $a = b = 1$、$c = -2$（于是 $a + b + c = 0$ 且 $abc = -2 \ne 0$）。则

$$S_2 = \frac{1+1+4}{2} = 3, \qquad S_7 = \frac{1+1-128}{7} = -18, \qquad S_9 = \frac{1+1-512}{9} = -\frac{170}{3},$$

所以 $S_2S_7 = -54 \ne -\frac{170}{3} = S_9$。命题不成立。

想一想 第 (ii) 和 (iii) 部分都行得通，为甚么规律到了 x^9 就失效?

对于 x^5 和 x^7，只有一个 u 的幂有贡献，所以每个系数都是 q 和 r 的单一乘积，而这乘积恰好能分解为适当的前面几个 S 的乘积。到了 x^9，有两个 u 的幂作出贡献，而来自 u^3 的额外 r^3 项破坏了这种分解。

验算

- 取 $a=b=1$、$c=-2$，则 $q=-3$、$r=-2$。于是 $S_2=3=-q$，$S_3=(1+1-8)/3=-2=r$，$S_5=(1+1-32)/5=-6=-qr$，所以 $S_2S_3=S_5$。此外 $S_7=-18=q^2r$，而 $S_9-S_2S_7=-\frac{170}{3}+54=-\frac{8}{3}=\frac{1}{3}r^3$，与预测完全一致。
- 若 $c=0$（于是 $b=-a$），所有奇数下标的 S_n 都等于零，(iv) 中命题的两边都是 0，与命题在 $abc=0$ 时成立一致。
- (ii) 至 (iv) 的恒等式以及 $S_9-S_2S_7$ 的公式，亦已用电脑以精确算术，对 200 组随机选取的有理数 a 和 b 验证。

常见错误

- 遗漏 u^4 对 x^9 的贡献：很容易只考虑到 u^3 便停下，得出错误的系数 $\frac{1}{3}r^3$。
- $(-1)^{n+1}$ 引起的符号错误：x^2 的系数是 $-S_2$，所以 $S_2=-q$，而不是 q。
- 忘记二项式系数：u^3 的 x^7 项是 $3q^2r$，而不是 q^2r。

STEP 之外

各 S_n（除了因子 $1/n$ 之外）就是三次多项式各根的幂和（power sums），而 q、r 则是它的基本对称多项式（elementary symmetric polynomials）。取对数再比较系数，是牛顿恒等式（Newton's identities）的一个生成函数（generating function）证明；牛顿恒等式把每个幂和以基本对称多项式表示，而其背后的定理是：每个对称多项式都可以用基本对称多项式表示。关于对称多项式和多项式的根，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）；若想多做以级数为题材、STEP 风格的练习，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP III 2014，第 2 题

纯数学 · 双曲函数积分法

STEP III 2014，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/14-s3-q2

关键思路

两个被积函数（integrand）的分母都是 $\cosh 2x$，而倍角公式（double-angle formula）既可以用 $\cosh^2 x$ 表示，也可以用 $\sinh^2 x$ 表示。选取与分子配合的那个版本，每个积分便会化成标准的有理函数积分。至于最后一部分，窍门是回到指数函数：令 $u = \mathrm{e}^x$，两个被积函数都会变成分母为 $1 + u^4$ 的 u 的有理函数（rational function），而把它们适当组合，恰好得到 $1/(1+u^4)$。

解答

第 (i) 部分

我们利用 $\cosh 2x = 2\cosh^2 x - 1$。令 $u = \cosh x$，则 $\mathrm{d}u = \sinh x \,\mathrm{d}x$，所以

$$\int \frac{\sinh x}{\cosh 2x}\,\mathrm{d}x = \int \frac{\mathrm{d}u}{2u^2 - 1}.$$

分母可因式分解为 $(\sqrt{2}u - 1)(\sqrt{2}u + 1)$，而两个因式相差 2，所以

$$\frac{1}{2u^2-1} = \frac{1}{2}\left(\frac{1}{\sqrt{2}u - 1} - \frac{1}{\sqrt{2}u + 1}\right).$$

每一项积分后都是对数，并因 u 的系数而多出因子 $\frac{1}{\sqrt{2}}$：

$$\begin{aligned}\int \frac{\mathrm{d}u}{2u^2-1} &= \frac{1}{2\sqrt{2}}\Big(\ln|\sqrt{2}u - 1| - \ln|\sqrt{2}u + 1|\Big) + C \\ &= \frac{1}{2\sqrt{2}} \ln\left|\frac{\sqrt{2}\cosh x - 1}{\sqrt{2}\cosh x + 1}\right| + C,\end{aligned}$$

这就是所求结果。（事实上 $\cosh x \geq 1$ 使得 $\sqrt{2}\cosh x - 1 > 0$，所以对实数 x 而言，绝对值符号并非必要。）

第 (ii) 部分

现在分子是 $\cosh x$，即 $\sinh x$ 的导数（derivative），所以相应的代换（substitution）是 $u = \sinh x$，$\mathrm{d}u = \cosh x\,\mathrm{d}x$。写成 $\cosh 2x = 1 + 2\sinh^2 x$，得

$$\begin{aligned}\int \frac{\cosh x}{\cosh 2x}\,\mathrm{d}x &= \int \frac{\mathrm{d}u}{1+2u^2} = \frac{1}{\sqrt{2}}\arctan\left(\sqrt{2}\,u\right) + C \\ &= \frac{1}{\sqrt{2}}\arctan\left(\sqrt{2}\sinh x\right) + C,\end{aligned}$$

这里用了 $\int \mathrm{d}u/(1+a^2u^2) = \frac{1}{a}\arctan(au)$，其中 $a = \sqrt{2}$。

第 (iii) 部分

乍看之下，$\int_0^1 \frac{\mathrm{d}u}{1+u^4}$ 里完全没有 cosh。

> **想一想** 如果把所有东西都用 $u = \mathrm{e}^x$ 表示，$\cosh x/\cosh 2x$ 会变成甚么样子？

由于 $\cosh x = \frac{1}{2}(u + u^{-1})$ 及 $\cosh 2x = \frac{1}{2}(u^2 + u^{-2})$，分子分母同乘 u^2，得

$$\frac{\cosh x}{\cosh 2x} = \frac{u + u^{-1}}{u^2 + u^{-2}} = \frac{u(u^2+1)}{u^4+1}, \qquad \frac{\sinh x}{\cosh 2x} = \frac{u(u^2-1)}{u^4+1}.$$

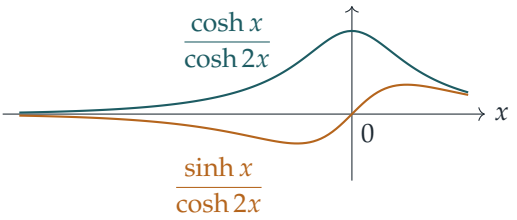
此外 $\mathrm{d}x = \mathrm{d}u/u$，所以多出来的因子 u 会约去：

$$\frac{\cosh x}{\cosh 2x}\,\mathrm{d}x = \frac{1+u^2}{1+u^4}\,\mathrm{d}u, \qquad \frac{\sinh x}{\cosh 2x}\,\mathrm{d}x = \frac{u^2-1}{1+u^4}\,\mathrm{d}u.$$

第一式减去第二式，剩下 $\dfrac{2}{1+u^4}\,\mathrm{d}u$。最后，$u$ 由 0 变到 1，即 $x = \ln u$ 由 $-\infty$ 变到 0。因此

$$\int_0^1 \frac{\mathrm{d}u}{1+u^4} = \frac{1}{2}\int_{-\infty}^0 \frac{\cosh x}{\cosh 2x}\,\mathrm{d}x - \frac{1}{2}\int_{-\infty}^0 \frac{\sinh x}{\cosh 2x}\,\mathrm{d}x.$$

右边两个积分都收敛（converge），因为当 $x \to -\infty$ 时，每个被积函数都是 $O(\mathrm{e}^x)$。（左边是普通的积分，所以这其实只是一次变量代换，无穷的那一端来自 $u \to 0$。）



cosh 的积分。由第 (ii) 部分，当 $x \to -\infty$ 时 $\sinh x \to -\infty$，所以 $\arctan(\sqrt{2}\sinh x) \to -\frac{\pi}{2}$。因此

$$\int_{-\infty}^{0} \frac{\cosh x}{\cosh 2x}\,\mathrm{d}x = \frac{1}{\sqrt{2}}\left(\arctan 0 - \left(-\tfrac{\pi}{2}\right)\right) = \frac{\pi}{2\sqrt{2}}.$$

sinh 的积分。由第 (i) 部分，当 $x \to -\infty$ 时 $\cosh x \to \infty$，所以对数内的分数趋向 1，对数则趋向 0。在 $x = 0$ 处，$\cosh 0 = 1$，所以

$$\int_{-\infty}^{0} \frac{\sinh x}{\cosh 2x}\,\mathrm{d}x = \frac{1}{2\sqrt{2}} \ln \frac{\sqrt{2}-1}{\sqrt{2}+1}.$$

有理化（rationalising）后，$\dfrac{\sqrt{2}-1}{\sqrt{2}+1} = (\sqrt{2}-1)^2$，而 $\sqrt{2}-1 = \dfrac{1}{\sqrt{2}+1}$，所以

$$\int_{-\infty}^{0} \frac{\sinh x}{\cosh 2x}\,\mathrm{d}x = \frac{1}{\sqrt{2}} \ln(\sqrt{2}-1) = -\frac{1}{\sqrt{2}} \ln(\sqrt{2}+1).$$

这个值是负数，理应如此：当 $x < 0$ 时 $\sinh x < 0$（见琥珀色曲线）。

综合起来。

$$\int_0^1 \frac{\mathrm{d}u}{1+u^4} = \frac{1}{2}\left(\frac{\pi}{2\sqrt{2}} + \frac{1}{\sqrt{2}}\ln(\sqrt{2}+1)\right) = \frac{\pi + 2\ln(\sqrt{2}+1)}{4\sqrt{2}},$$

正如所求。

验算

- 把 (i) 和 (ii) 的答案求导，应还原为被积函数；就 (ii) 而言，$\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{\sqrt{2}}\arctan(\sqrt{2}\sinh x) = \dfrac{\cosh x}{1+2\sinh^2 x}$。
- 最终数值为 ≈ 0.86697，与用辛普森法则（Simpson's rule）计算 $\int_0^1 \mathrm{d}u/(1+u^4)$ 的结果相符。它合理地介乎 $\frac{\pi}{4} \approx 0.785$（较小的函数 $1/(1+u^2)$ 的积分）与 1 之间。
- 另外，两个无穷积分亦已用数值方法分别核实：$\pi/(2\sqrt{2}) \approx 1.1107$ 及 $-\ln(1+\sqrt{2})/\sqrt{2} \approx -0.6232$。

常见错误

- 在第 (ii) 部分也用 $\cosh 2x = 2\cosh^2 x - 1$：当 $u = \sinh x$ 时，你需要的是 $\sinh^2$ 的形式，即 $1 + 2\sinh^2 x$。
- 忘记当 $u = \mathrm{e}^x$ 时 $\mathrm{d}x = \mathrm{d}u/u$；正是这个因子约去了分子中多余的 u。
- 弄错积分上下限：$u \in [0,1]$ 对应的是 $x \in (-\infty, 0]$，而不是 $[0,1]$。

STEP 之外

同一个积分也可以直接用部分分式（partial fractions）来做，因为 $1+u^4=(u^2+\sqrt{2}u+1)(u^2-\sqrt{2}u+1)$；而利用 $\frac{1+u^2}{1+u^4}$ 与 $\frac{1-u^2}{1+u^4}$ 的组合，是一条省去大部分代数运算的巧妙捷径。在大学里，像 $\int_0^\infty \mathrm{d}u/(1+u^4)=\pi/(2\sqrt{2})$ 这类积分，用围道积分（contour integration）和留数定理（residue theorem）一行就能完成，而反常积分（improper integral）则会在分析（analysis）课程中以完全严谨的方式处理。Needham 的 *Visual Complex Analysis*（《直观复分析》）是领略复分析这一面的绝佳读物。

详细解答

STEP III 2014，第 3 题

纯数学 • 到抛物线的距离

STEP III 2014，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/14-s3-q3

关键思路

抛物线（parabola）上每一点都可以写成 $(at^2, 2at)$，因此我们需要的每个距离都变成单一变量 t 的函数；而在两个部分中，这个函数原来都是二次函数（quadratic）（对直线而言是 t 的二次函数，对点而言是 t^2 的二次函数）。求二次函数的最小值只不过是配方（completing the square），但我们必须留意变量可以取值的范围：t^2 不可以是负数，而正是这个限制把第 (ii) 部分分成不同情况。至于圆的部分则唾手可得，因为一点到一个圆的距离，就是该点到圆心的距离减去半径。

解答

第 (i) 部分

把 L 写成 $mx - y + c = 0$。抛物线上的点 $P = (at^2, 2at)$ 到 L 的垂直距离（perpendicular distance）是

$$\frac{|mat^2 - 2at + c|}{\sqrt{m^2 + 1}}.$$

设 $f(t) = mat^2 - 2at + c$。由于 $ma > 0$，配方得

$$f(t) = ma\left(t - \frac{1}{m}\right)^2 + c - \frac{a}{m} = ma\left(t - \frac{1}{m}\right)^2 + \frac{mc - a}{m},$$

所以 f 的最小值是 $\dfrac{mc - a}{m}$，在 $t = 1/m$ 时取得。

情况 $mc > a$。此时对每个 t 都有 $f(t) \geq \frac{mc-a}{m} > 0$，因此整条抛物线位于 L 的同一侧，永不与之相交。L 与抛物线之间的最短距离，就是抛物线上的点到 L 的最小距离，即

$|f(t)|/\sqrt{m^2+1} = f(t)/\sqrt{m^2+1}$ 的最小值：

$$\frac{mc-a}{m\sqrt{m^2+1}},$$

正如所求。最近的点是 $t = 1/m$，该处抛物线切线（tangent）的斜率为 $\mathrm{d}y/\mathrm{d}x = 1/t = m$：切线与 L 平行，与几何直觉一致。

情况 $mc \le a$。此时 $f(1/m) \le 0$，而当 $t \to \infty$ 时 $f(t) \to +\infty$。由于 f 是连续（continuous）的，存在某个 t 使 $f(t) = 0$，所以 L 与抛物线相交，最短距离为 0。

第 (ii) 部分：点 $(p, 0)$

由 $C = (p, 0)$ 到 $P = (at^2, 2at)$ 的距离的平方是

$$(at^2 - p)^2 + 4a^2t^2.$$

这只取决于 $u = t^2$，而 u 可取 $[0, \infty)$ 中的每一个值。因此我们要求

$$g(u) = (au - p)^2 + 4a^2u = a^2u^2 - 2a(p-2a)u + p^2 \qquad (u \ge 0).$$

的最小值。配方得

$$g(u) = a^2\left(u - \frac{p-2a}{a}\right)^2 + p^2 - (p-2a)^2 = a^2\left(u - \frac{p-2a}{a}\right)^2 + 4a(p-a).$$

> **想一想** 不受限制时，g 的最小值在 $u = (p-2a)/a$ 取得。这个值是否总是可以取到？

不是：$u = t^2$ 必须是非负数，而只有当 $p \ge 2a$ 时才有 $(p-2a)/a \ge 0$。

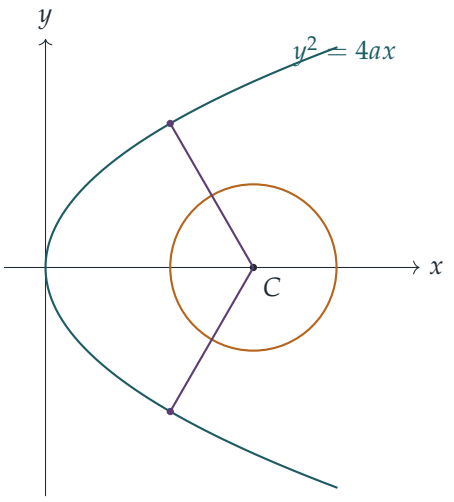
情况 $p/a > 2$。顶点（vertex）$u = p/a - 2$ 是正数，因此可以取到，g 的最小值是 $4a(p-a)$；由于 $p > 2a > a$，这个值是正数。最短距离是

$$2\sqrt{a(p-a)},$$

在两点 $\left(p-2a,\ \pm 2\sqrt{a(p-2a)}\right)$ 取得。

情况 $p/a \le 2$。顶点位于 $u \le 0$，因此 g 在 $u \ge 0$ 上递增，其在该范围内的最小值是 $g(0) = p^2$。最短距离是 p，在抛物线的顶点 $(0, 0)$ 取得。

两个答案在 $p = 2a$ 时吻合，两者都给出 $2a$。



图中所示为 $a = 1$、$p = 5$ 的情况：抛物线上最接近 C 的点（紫色）距离为 $2\sqrt{1 \cdot 4} = 4$，而以 C 为圆心、半径为 $b = 2$ 的圆（琥珀色）与抛物线保持 2 的距离。

第 (ii) 部分：圆

设 d 为刚才求得的由 $C = (p, 0)$ 到抛物线的距离，即

$$d = \begin{cases} p & \text{若} p \le 2a, \\ 2\sqrt{a(p-a)} & \text{若} p > 2a. \end{cases}$$

该圆的圆心为 C，半径为 b。

若 $d > b$。在抛物线上任取一点 P，在圆上任取一点 Q。由三角不等式（triangle inequality），

$$|PQ| \ge |PC| - |CQ| = |PC| - b \ge d - b > 0.$$

当 P 是抛物线上最接近 C 的点，而 Q 是线段 CP 与圆的交点时（由于 $|CP| = d > b$，线段确实与圆相交），等号成立。因此最短距离是 $d - b$。

若 $d \le b$。当 P 沿抛物线移动时，$|PC|$ 由其最小值 $d \le b$ 连续地变化至任意大的值，因此在某处有 $|PC| = b$。该点同时位于两条曲线上，最短距离为 0。

综合以上，圆与抛物线之间的最短距离是

$$\begin{cases} p - b & \text{若} p \le 2a \text{ 且} b < p, \\ 2\sqrt{a(p-a)} - b & \text{若} p > 2a \text{ 且} b < 2\sqrt{a(p-a)}, \\ 0 & \text{其他情况（即若} b \ge d\text{）}. \end{cases}$$

验算

- 取 $m=1$、$c=2$、$a=1$，则 $mc=2>a$。公式给出 $\frac{1}{\sqrt{2}}\approx 0.7071$；直接计算，$t=1$ 的点是 $(1,2)$，它到 $x-y+2=0$ 的距离是 $|1-2+2|/\sqrt{2}=1/\sqrt{2}$。若改取 $c=\frac{1}{2}$（使 $mc<a$），直线 $y=x+\frac{1}{2}$ 与 $y^2=4x$ 相交于满足 $x^2-3x+\frac{1}{4}=0$ 之处，此方程有实根，确认距离为 0。
- 取 $a=1$、$p=5$：点 $(3,\pm 2\sqrt{3})$ 与 $(5,0)$ 的距离为 $\sqrt{4+12}=4=2\sqrt{1\cdot 4}$，而以电脑在抛物线上搜索，找不到更近的点。
- 当 $p\to 0$ 时距离 $p\to 0$，这是正确的，因为 $(0,0)$ 在抛物线上。当 p 很大时，$2\sqrt{a(p-a)}<p$，这是必然的，因为顶点的距离是 p。
- 所有结果都经电脑以数百组随机选取的 m、c、a、p 和 b 搜索验证。

常见错误

- 把距离平方对 t 的导数（derivative）设为零时，遗漏了根 $t=0$，或忘记当 $p<2a$ 时 $t^2=p/a-2$ 无解：分情况讨论正是第 (ii) 部分的重点所在。
- 在 $mc<a$ 时仍给出 $(mc-a)/(m\sqrt{m^2+1})$：这个值是负数，而且直线其实与抛物线相交。
- 在圆的部分写出 $d-b$ 而不检查其正负；当 $b\geq d$ 时两条曲线相交，答案是 0。

STEP 之外

第 (ii) 部分中的最近点，正是 CP 为抛物线法线（normal）之处，而在 $p=2a$ 处的转变并非偶然：$(2a,0)$ 是抛物线在顶点处的曲率中心（centre of curvature）。在它之外的点，在对称轴以外有两个最近点；这类转变点的轨迹称为渐屈线（evolute），是曲线微分几何（differential geometry）中的经典对象。在受限制的范围内求函数的最小值，而答案可能位于边界而非驻点（stationary point），正是约束优化（constrained optimisation）的起点。如想做更多同类型的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

STEP III 2014，第 4 题

纯数学 · 积分不等式

STEP III 2014，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/14-s3-q4

本题在 *Beyond STEP* 第 4.7 节变分法中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/variations.html

关键思路

把 I_1 中的平方展开，再与 I 比较：两者之差是一个恰当导数（exact derivative）$\frac{\mathrm{d}}{\mathrm{d}x}\left(y^2\tan x\right)$ 的积分，而由于 $y(1)=0$ 及 $\tan 0=0$，这个积分等于零。一旦把 I 写成某个平方的积分，它非负就显而易见了。在第 (ii) 部分，只要选取 b 使 $\sec^2$ 项恰好产生 a^2y^2，同一技巧依然奏效；唯一可能出问题的地方，是 $\tan bx$ 在 $[0,1]$ 内趋于无穷。

解答

如题目所意，我们全程假设 y 在 $[0,1]$ 上有连续导数（continuous derivative）。

第 (i) 部分

由于 $0\le x\le 1<\frac{\pi}{2}$，函数 $\tan x$ 在 $[0,1]$ 上连续，所以 I_1 有意义。展开平方，

$$I_1=\int_0^1\left((y')^2+2yy'\tan x+y^2\tan^2 x\right)\mathrm{d}x,$$

因此

$$\begin{aligned}I-I_1&=-\int_0^1\left(y^2+2yy'\tan x+y^2\tan^2 x\right)\mathrm{d}x\\&=-\int_0^1\left(2yy'\tan x+y^2\sec^2 x\right)\mathrm{d}x,\end{aligned}$$

这里用了 $1+\tan^2 x = \sec^2 x$。由乘积法则（product rule），$\frac{\mathrm{d}}{\mathrm{d}x}\left(y^2\tan x\right) = 2yy'\tan x + y^2\sec^2 x$，正好就是被积函数。故此

$$I - I_1 = -\Big[y^2\tan x\Big]_0^1 = -\left(y(1)^2\tan 1 - y(0)^2\tan 0\right) = 0,$$

因为 $y(1) = 0$ 且 $\tan 0 = 0$。所以 $I = I_1$，得证。

由于 I_1 是一个平方在长度为正的区间上的积分，$I_1 \geq 0$，因此 $I \geq 0$。

现在假设 $I = 0$。那么 $I_1 = 0$，而被积函数 $(y' + y\tan x)^2$ 连续且非负，所以它必须在 $[0,1]$ 上处处为零（若它在某一点为正，便会在一个小区间上为正，令积分为正）。因此对 $0 \leq x \leq 1$ 有 $y' + y\tan x = 0$。

? **想一想** 怎样在不除以 y 的情况下解 $y' + y\tan x = 0$？毕竟 y 可能在某处为零。

使用积分因子（integrating factor）。由于在 $[0,1]$ 上 $\cos x > 0$，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{y}{\cos x}\right) = \frac{y'\cos x + y\sin x}{\cos^2 x} = \frac{y' + y\tan x}{\cos x} = 0,$$

所以对某常数 C 有 $y = C\cos x$。条件 $y(1) = 0$ 给出 $C\cos 1 = 0$，而 $\cos 1 \neq 0$，所以 $C = 0$。故此，仅当对 $[0,1]$ 中所有 x 都有 $y = 0$ 时，才有 $I = 0$。

第 (ii) 部分

设 $K = \int_0^1 (y' + ay\tan bx)^2\,\mathrm{d}x$，其中 $b > 0$ 有待选取。展开，

$$K = \int_0^1 \left((y')^2 + 2ayy'\tan bx + a^2y^2\tan^2 bx\right)\mathrm{d}x.$$

我们希望出现的导数是

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(ay^2\tan bx\right) = 2ayy'\tan bx + aby^2\sec^2 bx = 2ayy'\tan bx + aby^2 + aby^2\tan^2 bx.$$

把它与 K 的被积函数比较，$\tan^2$ 项恰好吻合当且仅当 $ab = a^2$，即 $b = a$。作此选择后，

$$(y' + ay\tan ax)^2 = (y')^2 - a^2y^2 + \frac{\mathrm{d}}{\mathrm{d}x}\left(ay^2\tan ax\right),$$

于是

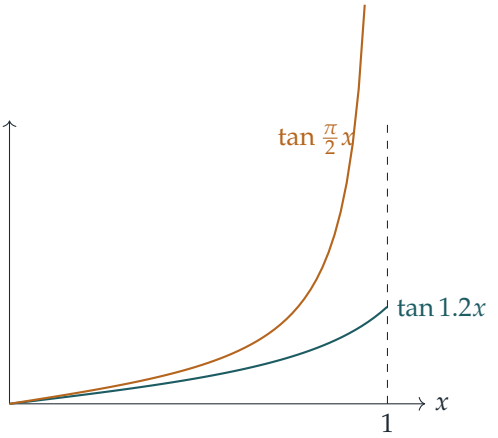
$$K = J + \Big[ay^2\tan ax\Big]_0^1 = J + a\,y(1)^2\tan a = J,$$

因为 $y(1) = 0$ 且 $\tan 0 = 0$。

这个论证需要 $\tan ax$ 在整个 $[0,1]$ 上有定义且连续，这样 K 才存在，微积分基本定理（fundamental theorem of calculus）才适用。由于 $a > 0$，ax 取遍 $[0,a]$，而 tan 在其上连续当且仅当 $a < \frac{\pi}{2}$。所以证明在以下情况成立：

$$a < k = \frac{\pi}{2}.$$

对这样的 a，我们有 $J = K \geq 0$，因为 K 是一个平方的积分。（事实上，把第 (i) 部分论证中的 $\cos x$ 换成 $\cos ax$，可知 $J = 0$ 会迫使 $y = C\cos ax$ 且 $C\cos a = 0$，所以 $y \equiv 0$；对于不恒为零的 y，我们甚至得到 $J > 0$。）



上图说明为何 $k = \frac{\pi}{2}$：当 $a < \frac{\pi}{2}$ 时，权函数 $\tan ax$ 在 $[0,1]$ 上保持有界；但当 $a = \frac{\pi}{2}$ 时，它在端点 $x = 1$ 处有一条渐近线（asymptote）。

$a = \frac{\pi}{2}$ 的情况

证明本身提示了该往哪里找：$K = 0$ 需要 $y' + ay\tan ax = 0$，其解为 $y = C\cos ax$。当 $a < \frac{\pi}{2}$ 时，这些解不满足条件 $y(1) = 0$；但当 $a = \frac{\pi}{2}$ 时，它们满足。所以试取

$$y = \cos\tfrac{\pi x}{2}, \qquad y' = -\tfrac{\pi}{2}\sin\tfrac{\pi x}{2},$$

它并非处处为零，且 $y(1) = \cos\frac{\pi}{2} = 0$。于是

$$\begin{aligned} J &= \frac{\pi^2}{4}\int_0^1 \left(\sin^2\tfrac{\pi x}{2} - \cos^2\tfrac{\pi x}{2}\right)\mathrm{d}x = -\frac{\pi^2}{4}\int_0^1 \cos \pi x\,\mathrm{d}x \\ &= -\frac{\pi^2}{4}\left[\frac{\sin \pi x}{\pi}\right]_0^1 = 0. \end{aligned}$$

所以当 $a = k$ 时，$y = \cos\frac{\pi x}{2}$（或它的任何非零倍数）给出 $J = 0$。

验算

- 在第 (i) 部分取 $y = 1 - x$：$I = \int_0^1 \big(1 - (1-x)^2\big)\mathrm{d}x = 1 - \frac{1}{3} = \frac{2}{3} > 0$，而以数值方法计算 I_1 也得到 0.6667。
- 对许多随机选取的 $(1-x)$ 的三次多项式倍数，以及 $(0, \frac{\pi}{2})$ 中随机选取的 a，电脑验算以高精度确认了 $J = K > 0$。

- 界限 $k=\frac{\pi}{2}$ 是最佳的：取 $y=\cos\frac{\pi x}{2}$，可得 $J=\frac{1}{2}\left(\frac{\pi^2}{4}-a^2\right)$，对每个 $a>\frac{\pi}{2}$ 都是负数（例如当 $a=2$ 时 $J\approx -0.766$）。

常见错误

- 忘记 $x=0$ 处的边界项为零是因为 $\tan 0=0$，而不是因为 $y(0)$ 满足甚么条件；题目对 $y(0)$ 没有任何假设。
- 由 $I_1=0$ 推出 $y'+y\tan x=0$ 时没有说明理由：关键在于被积函数的连续性。
- 选取 b 时没有检查 $\tan bx$ 在 $[0,1]$ 上是否保持有限；条件 $a<\frac{\pi}{2}$ 正是由此而来。

STEP 之外

对于 $y(1)=0$ 的不等式 $\int_0^1 (y')^2\,\mathrm{d}x\geq\frac{\pi^2}{4}\int_0^1 y^2\,\mathrm{d}x$，是 Wirtinger 不等式（Wirtinger's inequality，亦称庞加莱（Poincaré）不等式）的一种形式，而常数 $\frac{\pi^2}{4}$ 是 $-y''=\lambda y$ 在 $y'(0)=0$、$y(1)=0$ 条件下的最小特征值（eigenvalue）；极值函数 $\cos\frac{\pi x}{2}$ 就是对应的特征函数（eigenfunction）。在变分法（calculus of variations）中，借助一个精心选取的导数来配方，正是勒让德（Legendre）和雅可比（Jacobi）判别法背后的想法，用以判断一条驻定路径（stationary path）是否真的是极小值。想找更多同类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*；至于背后的特征值问题，可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*。

详细解答

STEP III 2014，第 5 题

纯数学 · 几何中的复数

STEP III 2014，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/14-s3-q5

关键思路

在复平面（complex plane）上，“平行且长度相等”其实就是“作为位移向量（displacement vector）相等”，而逆时针（anticlockwise）旋转四分之一圈就是乘以 i。因此，本题的每一个几何条件都可以化为关于顶点的一条简短线性方程。至于建在各边上的正方形，我们先走到该边的中点（midpoint），再沿垂直于该边的方向向外走半条边长，就得到正方形的中心。

解答

平行四边形的判别条件

首先设 $a+c=b+d$。那么

$$b-a=c-d \qquad 及 \qquad d-a=c-b,$$

所以位移 $\overrightarrow{AB}$ 等于 $\overrightarrow{DC}$，而 $\overrightarrow{AD}$ 等于 $\overrightarrow{BC}$。相等的位移互相平行且长度相等，故两组对边均平行且相等：$ABCD$ 是平行四边形（parallelogram）。

反之，设 $ABCD$ 是平行四边形。两条平行且等长的线段，其位移必定相等或相反，所以

$$b-a=\varepsilon(c-d), \qquad d-a=\eta(c-b), \qquad \varepsilon,\eta\in\{1,-1\}.$$

若 $\varepsilon=1$ 或 $\eta=1$，相应的方程移项后即得 $a+c=b+d$。余下唯一的情况是 $\varepsilon=\eta=-1$，即 $b-a=d-c$ 及 $d-a=b-c$。以第一式减去第二式，得 $b-d=d-b$，故 $b=d$，这与顶点互不相同矛盾。因此 $a+c=b+d$，等价关系得证。

正方形的判别条件

设 $ABCD$ 是平行四边形，并设 $m = \frac{1}{2}(a+c) = \frac{1}{2}(b+d)$ 为两条对角线（diagonal）的公共中点（即其中心）。那么

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

若 $ABCD$ 是正方形（*square*），则其对角线长度相等，并互相垂直平分，所以 B 是由 A 绕 m 旋转四分之一圈得到的。由于顶点按逆时针方向标记，这个旋转是逆时针的。所以 $b-m = \mathrm{i}(a-m)$，即 $\frac{1}{2}(b-d) = \frac{1}{2}\mathrm{i}(a-c)$，也就是 $\mathrm{i}(a-c) = b-d$。

反之，若 $\mathrm{i}(a-c) = b-d$，记 $w = a-m \neq 0$。那么

$$a-m = w, \quad b-m = \mathrm{i}w, \quad c-m = -w, \quad d-m = -\mathrm{i}w,$$

所以四个顶点就是 w 绕 m 依次旋转 90° 所得。它们是一个正方形的顶点（四条边长均为 $|w|\sqrt{2}$，而且每条边都是前一条边乘以 i）。第二个等价关系得证。

第 (i) 部分

顶点 P, Q, R, S 按逆时针排列，所以当我们由 P 走向 Q 时，$PQRS$ 的内部在我们的左方。因此建在 PQ 外侧的正方形位于我们的右方。

从 PQ 的中点 $\frac{1}{2}(p+q)$ 出发，沿垂直于 PQ 的方向向右移动 $\frac{1}{2}|q-p|$（半条边长），便到达中心 X。把 $q-p$ 乘以 $-\mathrm{i}$ 会使它顺时针旋转 90°，即转向右方，所以这一步是 $-\frac{1}{2}\mathrm{i}(q-p)$。因此

$$x = \tfrac{1}{2}(p+q) - \tfrac{1}{2}\mathrm{i}(q-p) = \tfrac{1}{2}\big(p(1+\mathrm{i}) + q(1-\mathrm{i})\big),$$

即为所求。

想一想 为甚么是 $-\mathrm{i}$ 而不是 $+\mathrm{i}$?

乘以 $+\mathrm{i}$ 会把 $\overrightarrow{PQ}$ 转向左方，指向四边形内部，得到的是建在内侧的正方形的中心。由于每个内角都小于 180°，内部完全位于每条边的左方，所以右方确实是外侧。

第 (ii) 部分

对每条边依次应用第 (i) 部分（同样的推理对每条边都适用，因为每条边都是在内部位于左方的情况下走过的），

$$2x = p(1+\mathrm{i}) + q(1-\mathrm{i}), \qquad 2y = q(1+\mathrm{i}) + r(1-\mathrm{i}),$$
$$2z = r(1+\mathrm{i}) + s(1-\mathrm{i}), \qquad 2t = s(1+\mathrm{i}) + p(1-\mathrm{i}).$$

我们比较两个判别条件中的表达式。

对顶点之和。

$$\begin{aligned}2(x+z)-2(y+t) &= (p+r)(1+\mathrm{i})+(q+s)(1-\mathrm{i})-(q+s)(1+\mathrm{i})-(p+r)(1-\mathrm{i})\\ &= 2\mathrm{i}\,(p+r-q-s),\end{aligned}$$

所以

$$x+z-(y+t)=\mathrm{i}\big(p+r-(q+s)\big). \tag{1}$$

对角线。

$$2\mathrm{i}(x-z)=\mathrm{i}(p-r)(1+\mathrm{i})+\mathrm{i}(q-s)(1-\mathrm{i})=(p-r)(\mathrm{i}-1)+(q-s)(1+\mathrm{i}),$$
$$2(y-t)=(q-s)(1+\mathrm{i})+(r-p)(1-\mathrm{i})=(q-s)(1+\mathrm{i})+(p-r)(\mathrm{i}-1).$$

两者完全相同，所以

$$\mathrm{i}(x-z)=y-t \tag{2}$$

对每一个这样的四边形 $PQRS$ 都成立：线段 XZ 与 YT 总是长度相等且互相垂直。

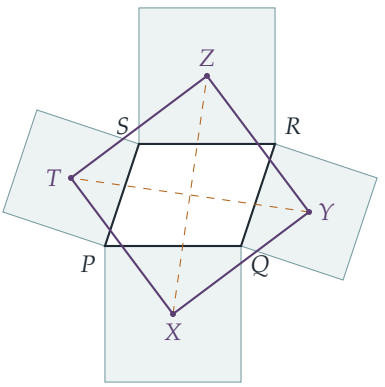
想一想 既然 (2) 总是成立，那么决定 $XYZT$ 是否正方形的是甚么？

只有平行四边形条件。现在我们把各部分结合起来。

若 $PQRS$ 是平行四边形，则 $p+r=q+s$，所以由 (1) 得 $x+z=y+t$，根据第一个结果，$XYZT$ 是平行四边形。再结合 (2)，由正方形的判别条件可知 $XYZT$ 是正方形。（该判别条件中“反之”的一半不需要任何关于标记方向的假设；而且 $X\neq Z$，因为若 $x=z$，则 $r-p=\mathrm{i}(s-q)$，即 $\overrightarrow{PR}$ 是 $\overrightarrow{QS}$ 逆时针旋转四分之一圈所得；但在按逆时针标记的凸四边形（convex quadrilateral）中，$\overrightarrow{PR}$ 总是由 $\overrightarrow{QS}$ 顺时针旋转一个介乎 0° 与 180° 之间的角，再经缩放而得。）

若 $XYZT$ 是正方形，则它特别地是平行四边形，所以根据第一个结果，$x+z=y+t$。由 (1) 得 $\mathrm{i}(p+r-q-s)=0$，所以 $p+r=q+s$，$PQRS$ 是平行四边形。

因此，$XYZT$ 是正方形当且仅当 $PQRS$ 是平行四边形。



验算

取平行四边形 $p = 0$、$q = 4$、$r = 5 + 3\mathrm{i}$、$s = 1 + 3\mathrm{i}$（如上图）。公式给出 $x = 2 - 2\mathrm{i}$、$y = 6 + \mathrm{i}$、$z = 3 + 5\mathrm{i}$、$t = -1 + 2\mathrm{i}$。$XYZT$ 的各边为 $4 + 3\mathrm{i}$、$-3 + 4\mathrm{i}$、$-4 - 3\mathrm{i}$、$3 - 4\mathrm{i}$，每条都是前一条的 i 倍：这是边长为 5 的正方形。作为对第 (i) 部分的验算，$x = 2 - 2\mathrm{i}$ 确实是 PQ 下方、顶点为 0、4、$4 - 4\mathrm{i}$、$-4\mathrm{i}$ 的正方形的中心。以电脑对数以千计的随机凸四边形进行检验，每次都证实恒等式（identity）(2) 成立，而且 $XYZT$ 是正方形恰好当 $PQRS$ 是平行四边形。

常见错误

- 在第 (i) 部分使用 $+\mathrm{i}$，结果把正方形放在四边形内侧。
- 在平行四边形的判别条件中，忘记“平行且相等”也容许 $b - a = -(c - d)$，而这种情况必须排除。
- 对“当且仅当”只证明了一个方向。在第 (ii) 部分，“正方形 $\Rightarrow$ 平行四边形”这个方向只需用到 (1)。

STEP 之外

恒等式 (2) 就是范奥贝尔定理（van Aubel's theorem）：对任何四边形，连接相对两个外接正方形中心的线段长度相等且互相垂直。这是一个普遍原理的小例子：由旋转和中点构成的几何作图，都是作用于顶点上的线性映射（linear map），因此有关它们的定理都可化为代数恒等式（关于等边三角形的拿破仑定理（Napoleon's theorem）是另一个例子）。Needham 的 *Visual Complex Analysis*（《可视化复分析》）以优美的方式发展了这种看待复数乘法的几何观点。

详细解答

STEP III 2014，第 6 题

纯数学 · 以凸性证明不等式

STEP III 2014，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/14-s3-q6

关键思路

一个函数若在零点处取值为零、斜率（slope）也为零，而且图像向上弯曲，那么它必然升到零以上。把题目给出的事实积分两次，就能把这幅图像变成严谨的证明。之后题中每条不等式，都只需把所有项移到同一边，得出一个满足 $f(0) = f'(0) = 0$ 的函数 f，再验证 f'' 为正即可。真正要花心思的是选择恰当的整理方式，使 f'' 的形式简单。

解答

开首的结果

设对 $0 < t < x_0$ 有 $f''(t) > 0$，且 $f(0) = f'(0) = 0$。取定 x，使 $0 < x < x_0$。

取 $h = f''$。由于对 $0 < t < x$ 有 $f''(t) > 0$，由题目给出的结果得

$$f'(x) = f'(x) - f'(0) = \int_0^x f''(t)\,\mathrm{d}t > 0.$$

这对 $(0, x_0)$ 内每一个 x 都成立，所以对 $0 < t < x_0$ 有 $f'(t) > 0$。

现在再次应用题目给出的结果，这次取 $h = f'$。对 $0 < x < x_0$，我们有：对 $0 < t < x$，$f'(t) > 0$，所以

$$f(x) = f(x) - f(0) = \int_0^x f'(t)\,\mathrm{d}t > 0.$$

因此对 $0 < t < x_0$ 有 $f(t) > 0$，得证。下文会一直以 $x_0 = \frac{1}{2}\pi$ 应用这个结果。

第 (i) 部分

设 $f(x) = 1 - \cos x \cosh x$。则 $f(0) = 1 - 1 = 0$，而由乘积法则（product rule）得

$$f'(x) = \sin x \cosh x - \cos x \sinh x, \qquad f'(0) = 0 - 0 = 0.$$

再求导一次，

$$\begin{aligned} f''(x) &= (\cos x \cosh x + \sin x \sinh x) - (-\sin x \sinh x + \cos x \cosh x) \\ &= 2 \sin x \sinh x. \end{aligned}$$

当 $0 < x < \frac{1}{2}\pi$ 时，$\sin x > 0$ 及 $\sinh x > 0$，所以 $f''(x) > 0$。由开首的结果，$f(x) > 0$，即

$$\cos x \cosh x < 1 \qquad \text{对} 0 < x < \tfrac{1}{2}\pi.$$

第 (ii) 部分

想一想 应该把开首的结果应用到哪个函数上？直接用 $\frac{\sin x}{x} - \frac{1}{\cosh x}$ 看来会很繁复。

先消去分数。当 $0 < x < \frac{1}{2}\pi$ 时，x、$\sin x$、$\sinh x$ 和 $\cosh x$ 全部为正，所以两边乘以正数不会改变不等号的方向。于是

$$\frac{1}{\cosh x} < \frac{\sin x}{x} \iff x < \sin x \cosh x, \qquad \frac{\sin x}{x} < \frac{x}{\sinh x} \iff \sin x \sinh x < x^2.$$

（第一条两边乘以 $x \cosh x > 0$，第二条两边乘以 $x \sinh x > 0$。）乘积 $\sin x \cosh x$ 和 $\sin x \sinh x$ 求导后形式整齐，这正是选用这个形式的原因。

左边的不等式。设 $g(x) = \sin x \cosh x - x$。则 $g(0) = 0$，且

$$g'(x) = \cos x \cosh x + \sin x \sinh x - 1, \qquad g'(0) = 1 + 0 - 1 = 0.$$

再求导一次，$\sin x \cosh x$ 项互相抵消：

$$\begin{aligned} g''(x) &= (-\sin x \cosh x + \cos x \sinh x) + (\cos x \sinh x + \sin x \cosh x) \\ &= 2 \cos x \sinh x, \end{aligned}$$

当 $0 < x < \frac{1}{2}\pi$ 时它为正，因为在该区间内 $\cos x > 0$ 及 $\sinh x > 0$。（上限 $\frac{1}{2}\pi$ 正是在这里起作用。）由开首的结果，$g(x) > 0$，所以 $x < \sin x \cosh x$，从而 $\frac{1}{\cosh x} < \frac{\sin x}{x}$。

右边的不等式。设 $k(x) = x^2 - \sin x \sinh x$。则 $k(0) = 0$，且

$$k'(x) = 2x - \cos x \sinh x - \sin x \cosh x, \qquad k'(0) = 0.$$

再求导一次，$\sin x \sinh x$ 项互相抵消：

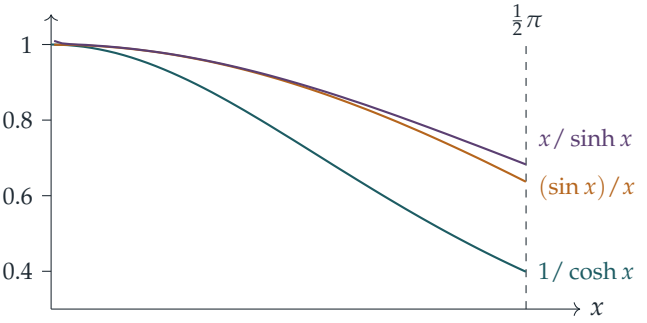
$$\begin{aligned} k''(x) &= 2 - (-\sin x \sinh x + \cos x \cosh x) - (\cos x \cosh x + \sin x \sinh x) \\ &= 2 - 2\cos x \cosh x. \end{aligned}$$

由第 (i) 部分，当 $0 < x < \frac{1}{2}\pi$ 时 $\cos x \cosh x < 1$，所以在该区间内 $k''(x) > 0$。由开首的结果，$k(x) > 0$，所以 $\sin x \sinh x < x^2$，从而 $\frac{\sin x}{x} < \frac{x}{\sinh x}$。

两者合起来，对 $0 < x < \frac{1}{2}\pi$，

$$\frac{1}{\cosh x} < \frac{\sin x}{x} < \frac{x}{\sinh x},$$

得证。



验算

- 级数（series）：$\cos x \cosh x = 1 - \frac{1}{6}x^4 + \cdots$，所以当 x 很小时它确实略低于 1，与 (i) 一致。同样，$\frac{1}{\cosh x} \approx 1 - \frac{1}{2}x^2$、$\frac{\sin x}{x} \approx 1 - \frac{1}{6}x^2$ 及 $\frac{x}{\sinh x} \approx 1 - \frac{1}{6}x^2$，所以 x 很小时左边的不等式显然成立，而右边的不等式则由下一项决定（$\frac{\sin x}{x}$ 的 x^4 系数是 $\frac{1}{120}$，而 $\frac{x}{\sinh x}$ 的则是较大的 $\frac{7}{360}$）。
- 在 $x = 1$：$\cos 1 \cosh 1 \approx 0.834 < 1$，且 $\frac{1}{\cosh 1} \approx 0.648 < \sin 1 \approx 0.841 < \frac{1}{\sinh 1} \approx 0.851$。
- 在 $x = \frac{1}{2}\pi$ 附近，取 $x = 1.57$：三个值约为 0.399、0.637 和 0.683，而 $\cos x \cosh x \approx 0.002$。

常见错误

- 直接把开首的结果应用到 $\frac{\sin x}{x} - \frac{1}{\cosh x}$ 本身：它在 0 处根本没有定义，而且其导数（derivative）十分难处理。应先消去分母，并检查每个乘上的因子都是正数。
- 只检查 $f(0) = 0$ 而忘记检查 $f'(0) = 0$；缺少后者，开首的结果便不适用。
- 没有察觉第二条不等式需要用到第 (i) 部分：$k'' = 2 - 2\cos x \cosh x$ 为正，正是因为 (i)。

STEP 之外

开首的结果其实是均值定理（mean value theorem）论证的一种形式：导数为正的函数是递增（increasing）的，这里用了两次。它说明凸函数（convex function）位于其切线（tangent line）之上，这里的切线就是原点处的 $y = 0$；而这种“比较二阶导数”的技巧，是证明初等函数之间不等式的标准方法。Abbott 的 *Understanding Analysis*（《理解分析》）对均值定理及其推论有细致的发展。

详细解答

STEP III 2014，第 7 题

纯数学 · 向量与圆内接四边形

STEP III 2014，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：

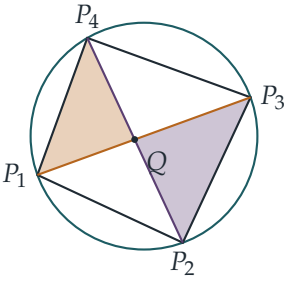
step.maths.org/questions/14-s3-q7

关键思路

第 (i) 部分是相交弦定理（intersecting chords theorem），源自同弓形内的圆周角相等（angles in the same segment）。向量部分则从另一途径再次得出同一事实。平面上的四点是“仿射相关”（affinely dependent）的，而该相关关系中的系数确定了两条对角线（diagonal）的交点位置，因此它们以已知比例分割每条对角线。把这些比例代入 (i)，便得出最终的关系式。

解答

整个解答都会用到这个构形的两个事实。第一，P_i 之中任何三点都不共线（collinear），因为一条直线与一个圆最多只相交于两点。第二，由于各顶点是沿圆周依次标记的，P_1 和 P_3 分别位于弦（chord）P_2P_4 所分出的两段不同的弧（arc）上。所以它们位于直线 P_2P_4 的两侧，而两条对角线相交于一点 Q，且 Q 严格位于线段 P_1P_3 和 P_2P_4 的内部。



第 (i) 部分

在三角形 P_1QP_4 和 P_2QP_3 中：

- $\angle P_1QP_4 = \angle P_2QP_3$，因为它们是在 Q 处的对顶角（vertically opposite angles）；
- $\angle QP_1P_4 = \angle P_3P_1P_4$ 及 $\angle QP_2P_3 = \angle P_4P_2P_3$。这两角相等，因为它们都是弦 P_3P_4 在同一段弧上的点 P_1、P_2 所对的角（该弧由 P_4 经 P_1 和 P_2 到 P_3）。

所以这两个三角形相似（similar），对应关系为 $P_1 \leftrightarrow P_2$、$Q \leftrightarrow Q$、$P_4 \leftrightarrow P_3$。对应边成比例：

$$\frac{P_1Q}{P_2Q} = \frac{QP_4}{QP_3}, \qquad \text{所以} \qquad (P_1Q)(QP_3) = (P_2Q)(QP_4),$$

此即所求。

第 (ii) 部分

圆内接四边形(cyclic quadrilateral)位于一个平面上,所以三个向量 $\mathbf{p}_2 - \mathbf{p}_1$、$\mathbf{p}_3 - \mathbf{p}_1$、$\mathbf{p}_4 - \mathbf{p}_1$ 都位于一个二维空间中。二维空间中的三个向量必定线性相关（linearly dependent），所以存在不全为零的数 b_2, b_3, b_4，使得

$$b_2(\mathbf{p}_2 - \mathbf{p}_1) + b_3(\mathbf{p}_3 - \mathbf{p}_1) + b_4(\mathbf{p}_4 - \mathbf{p}_1) = \mathbf{0}.$$

（具体来说：由于 P_1, P_2, P_3 不共线，$\mathbf{p}_2 - \mathbf{p}_1$ 与 $\mathbf{p}_3 - \mathbf{p}_1$ 不平行，因此它们张成（span）整个平面，而 $\mathbf{p}_4 - \mathbf{p}_1$ 是它们的线性组合。于是可取 $b_4 = -1$。）现设 $a_1 = -(b_2 + b_3 + b_4)$，并对 $i = 2, 3, 4$ 设 $a_i = b_i$。那么 $\sum a_i = 0$，各 a_i 不全为零，而且

$$\sum_{i=1}^{4} a_i\mathbf{p}_i = a_1\mathbf{p}_1 + b_2\mathbf{p}_2 + b_3\mathbf{p}_3 + b_4\mathbf{p}_4 = b_2(\mathbf{p}_2 - \mathbf{p}_1) + b_3(\mathbf{p}_3 - \mathbf{p}_1) + b_4(\mathbf{p}_4 - \mathbf{p}_1) = \mathbf{0},$$

所以 $(*)$ 成立。

第 (iii) 部分

$a_1 + a_3 \neq 0$。假设相反，即 $a_1 + a_3 = 0$。那么由 $(*)$ 的第一条方程，亦有 $a_2 + a_4 = 0$，所以 $a_3 = -a_1$ 及 $a_4 = -a_2$。第二条方程变成

$$a_1(\mathbf{p}_1 - \mathbf{p}_3) + a_2(\mathbf{p}_2 - \mathbf{p}_4) = \mathbf{0}.$$

直线 P_1P_3 与 P_2P_4 是不同的直线（否则四点全部共线），而且它们相交于 Q，所以两者不平行。因此 $\mathbf{p}_1 - \mathbf{p}_3$ 与 $\mathbf{p}_2 - \mathbf{p}_4$ 均非零且不平行，这迫使 $a_1 = a_2 = 0$。于是四个 a_i 全为零，产生矛盾。所以 $a_1 + a_3 \neq 0$，同理 $a_2 + a_4 = -(a_1 + a_3) \neq 0$。

交点。由 $(*)$ 得 $a_1\mathbf{p}_1 + a_3\mathbf{p}_3 = -(a_2\mathbf{p}_2 + a_4\mathbf{p}_4)$ 及 $a_1 + a_3 = -(a_2 + a_4)$。两式相除，

$$\mathbf{r} = \frac{a_1\mathbf{p}_1 + a_3\mathbf{p}_3}{a_1 + a_3} = \frac{a_2\mathbf{p}_2 + a_4\mathbf{p}_4}{a_2 + a_4}.$$

第一个形式显示 $\mathbf{r} = \mathbf{p}_1 + \dfrac{a_3}{a_1+a_3}(\mathbf{p}_3 - \mathbf{p}_1)$，是直线 P_1P_3 上的一点。同样地，第二个形式显示它是直线 P_2P_4 上的一点。两条不平行的直线只相交于一点，所以 $\mathbf{r}$ 就是 Q 的位置向量（position vector）。

> **想一想** 怎样把这条 Q 的公式转化为长度 P_1Q 和 QP_3?

相减即可。利用 $\mathbf{q} = \mathbf{r}$,

$$\mathbf{q} - \mathbf{p}_1 = \frac{a_3(\mathbf{p}_3 - \mathbf{p}_1)}{a_1+a_3}, \qquad \mathbf{p}_3 - \mathbf{q} = \frac{a_1(\mathbf{p}_3 - \mathbf{p}_1)}{a_1+a_3}.$$

取长度得 $P_1Q = \dfrac{|a_3|}{|a_1+a_3|}P_1P_3$ 及 $QP_3 = \dfrac{|a_1|}{|a_1+a_3|}P_1P_3$。同样地，由 $\mathbf{r}$ 的第二个形式得 $P_2Q = \dfrac{|a_4|}{|a_2+a_4|}P_2P_4$ 及 $QP_4 = \dfrac{|a_2|}{|a_2+a_4|}P_2P_4$。

正负号。由于 Q 严格位于 P_1 与 P_3 之间，分数 $\frac{a_3}{a_1+a_3}$ 严格介乎 0 与 1 之间。于是 $\frac{a_1}{a_1+a_3} = 1 - \frac{a_3}{a_1+a_3}$ 亦严格介乎 0 与 1 之间。两个正数的积为正，所以

$$\frac{a_1a_3}{(a_1+a_3)^2} > 0, \qquad \text{即} a_1a_3 > 0.$$

对 P_2P_4 作同样的论证，得 $a_2a_4 > 0$。所以 $|a_1a_3| = a_1a_3$ 及 $|a_2a_4| = a_2a_4$。

推导结论。代入第 (i) 部分:

$$\frac{a_1a_3}{(a_1+a_3)^2}(P_1P_3)^2 = (P_1Q)(QP_3) = (P_2Q)(QP_4) = \frac{a_2a_4}{(a_2+a_4)^2}(P_2P_4)^2.$$

由于 $(a_2+a_4)^2 = (a_1+a_3)^2 \neq 0$，两边乘以它，得

$$a_1a_3(P_1P_3)^2 = a_2a_4(P_2P_4)^2,$$

此即所求结果。

验算

- 取顶点为 $(1,0),(0,1),(-1,0),(0,-1)$ 的正方形。那么 $\mathbf{p}_1 + \mathbf{p}_3 = \mathbf{0} = \mathbf{p}_2 + \mathbf{p}_4$，所以 $a = (1,-1,1,-1)$ 满足 $(*)$。公式给出 Q 位于原点，且 $a_1a_3(P_1P_3)^2 = 4 = a_2a_4(P_2P_4)^2$。
- 对于不对称的情况，如图所示，取单位圆（unit circle）上角度为 $200°, 290°, 20°, 120°$ 的点 $P_1, \ldots, P_4$。(iii) 的全部三项结论，包括 $a_1a_3 > 0$ 和 $a_2a_4 > 0$，都已用电脑在这个例子及数以千计的随机圆内接四边形上验证，误差在 10^{-10} 以内。
- 把所有 a_i 乘以同一因子 λ 后，答案依然成立，这是理所当然的：最终关系式两边都乘以 λ^2。

常见错误

- 在 (i) 中配错了角：在 P_1 和 P_2 处的两个相等的角都是弦 P_3P_4 所对的角，这决定了哪些边互相对应。
- 在 (iii) 中未加说明便去掉绝对值（absolute value）符号。带绝对值的关系式很容易得出；真正的功夫在于证明 $a_1a_3 > 0$ 和 $a_2a_4 > 0$，这要用到 Q 位于两条对角线内部这一事实。
- 证明 $a_1 + a_3 \neq 0$ 时，忘记说明两条对角线为何不平行。

STEP 之外

量 $(P_1Q)(QP_3)$ 称为点 Q 关于该圆的幂（power）；对于通过 Q 的每一条弦，它的值都相同。第 (ii) 部分说的是平面上的四点是仿射相关的。数 a_i 是一种类似重心坐标（barycentric coordinates）的东西，而 Q 的公式让我们初步看到仿射几何（affine geometry）如何在不用任何坐标的情况下处理点、直线和比例。斯特朗（Strang）的 *Introduction to Linear Algebra*（《线性代数导论》）讲解了线性相关，以及 (ii) 背后的维数计算。

详细解答

STEP III 2014，第 8 题

纯数学 · 级数的收敛

STEP III 2014，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/14-s3-q8

本题在 *Beyond STEP* 第 2.3 节一个发散级数与它的收敛近亲中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/kempner.html

关键思路

把正整数切成一段段的区间 $[k^n, k^{n+1})$。每段中的每一项都介乎该段起点的值与下一段起点的值之间，所以每段之和都被夹在“项数乘以这两个值”之间。把各段加起来，一个难以处理的级数（series）便变成几何级数（geometric series），或者变成每段都贡献一个常数的级数，于是我们一眼就能看出它是否发散（diverges）。同样的计数方法也适用于不含数字 2 的整数，只要我们数清楚每段里有多少个这样的整数便可。

解答

分段不等式

区间 $k^n \le r \le k^{n+1} - 1$ 共包含

$$(k^{n+1} - 1) - k^n + 1 = k^{n+1} - k^n = k^n(k-1)$$

个整数。由于 f 是严格递减（strictly decreasing）的，对该段中的每个 r，

$$\mathrm{f}(k^{n+1}) < \mathrm{f}(r) \le \mathrm{f}(k^n),$$

第一个不等式成立是因为 $r < k^{n+1}$，第二个成立是因为 $r \geq k^n$（等号只在 $r = k^n$ 时成立）。把这些不等式对 r 的 $k^n(k-1)$ 个值相加，得

$$k^n(k-1)\,\mathrm{f}(k^{n+1}) \leq \sum_{r=k^n}^{k^{n+1}-1} \mathrm{f}(r) \leq k^n(k-1)\,\mathrm{f}(k^n),$$

即为所求。（事实上左边的不等式是严格的。）



$k = 2$ 时的各段，并标示每段的项数。

第 (i) 部分

取 $k = 2$ 及 $\mathrm{f}(r) = \frac{1}{r}$，这是严格递减的。分段不等式变成

$$\frac{2^n}{2^{n+1}} \leq \sum_{r=2^n}^{2^{n+1}-1} \frac{1}{r} \leq \frac{2^n}{2^n}, \qquad \text{即} \qquad \frac{1}{2} \leq \sum_{r=2^n}^{2^{n+1}-1} \frac{1}{r} \leq 1.$$

$n = 0, 1, \ldots, N$ 的各段恰好拼合起来覆盖 $1 \leq r \leq 2^{N+1} - 1$，因为每段都在下一段开始之前刚好结束。把这 $N+1$ 个不等式相加，

$$\frac{N+1}{2} \leq \sum_{r=1}^{2^{N+1}-1} \frac{1}{r} \leq N+1.$$

至于推论：部分和（partial sums）$H_m = \sum_{r=1}^{m} \frac{1}{r}$ 随 m 递增，而 $H_{2^{N+1}-1} \geq \frac{N+1}{2}$，只要取 N 足够大，它便可以大于任何我们指定的数。因此部分和无上界，$\sum \frac{1}{r}$ 不收敛（converge）。

第 (ii) 部分

取 $k = 2$ 及 $\mathrm{f}(r) = \frac{1}{r^3}$，同样是严格递减的。这次我们只需要右边的不等式：

$$\sum_{r=2^n}^{2^{n+1}-1} \frac{1}{r^3} \leq \frac{2^n}{2^{3n}} = \frac{1}{4^n}.$$

如前对 $n = 0, 1, \ldots, N$ 相加，

$$\sum_{r=1}^{2^{N+1}-1} \frac{1}{r^3} \leq \sum_{n=0}^{N} \frac{1}{4^n} < \sum_{n=0}^{\infty} \frac{1}{4^n} = \frac{1}{1-\frac{1}{4}} = \frac{4}{3}.$$

想一想 这只为部分的部分和给出了上界。为甚么它能为整个无穷级数给出上界？

所有项都是正数，所以部分和递增。任何部分和 $\sum_{r=1}^{m} r^{-3}$ 都不大于满足 $2^{N+1}-1 \geq m$ 的那个部分和，因此小于 $\frac{4}{3}$。一个有上界（bounded above）的递增数列必定收敛，而且其极限不大于该上界。因此

$$\sum_{r=1}^{\infty} \frac{1}{r^3} \leq \frac{4}{3} = 1\tfrac{1}{3}.$$

第 (iii) 部分

计算 $S(1000)$ 的元素个数。把 0 至 999 的每个整数写成三位数字串，不足的在前面补零（例如 7 写成 007）。补上的零不会引入 2，所以一个整数不含 2，当且仅当其数字串不含 2。三个位置中每一位都可以是 $0,1,3,4,\ldots,9$ 这九个数字之一，共得 9^3 个数字串，各自对应不同的整数。其中一个是 000，它不是正数，所以 $S(1000)$ 含有 9^3-1 个数。

分段计数。以 $n+1$ 个位置作同样的论证，可知 $S(10^{n+1})$ 中有 $9^{n+1}-1$ 个整数。所以由 $(n+1)$ 位数组成的区间 $10^n \leq r \leq 10^{n+1}-1$ 含有

$$(9^{n+1}-1)-(9^n-1) = 8\cdot 9^n$$

个不含 2 的整数。（等价地说，首位数字不能是 0 或 2，所以有 8 个选择，其余 n 位各有 9 个选择。）

估计 $\sigma(n)$ 的上界。这就是 $k=10$ 的分段论证，只不过我们只计算不含 2 的整数。该段中每个 r 都满足 $\frac{1}{r} \leq \frac{1}{10^n}$，所以这一段对总和的贡献最多是

$$8\cdot 9^n \cdot \frac{1}{10^n} = 8\left(\tfrac{9}{10}\right)^n.$$

现在固定 n，并选取 M 使得 $n \leq 10^{M+1}$。$S(n)$ 的每个元素都落在 $m=0,1,\ldots,M$ 的某一段之中，而且所有倒数都是正数。所以

$$\sigma(n) \leq \sum_{m=0}^{M} 8\left(\tfrac{9}{10}\right)^m < \sum_{m=0}^{\infty} 8\left(\tfrac{9}{10}\right)^m = \frac{8}{1-\frac{9}{10}} = 80.$$

第二个不等式是严格的，因为我们略去了 $m>M$ 的正项。因此对所有 n，$\sigma(n)<80$。

验算

- 第 (ii) 部分：$\sum r^{-3} = 1.2020\ldots$（阿培里常数（Apéry's constant)），远低于 $1\frac{1}{3}$。
- 第 (i) 部分取 $N=2$：$\sum_{r=1}^{7} \frac{1}{r} \approx 2.593$，介乎 1.5 与 3 之间。
- 第 (iii) 部分：直接计数可证实 1000 以下有 $728 = 9^3-1$ 个数，而每段有 $8\cdot 9^n$ 个。用电脑直接求和得 $\sigma(10^7) \approx 10.28$，远低于 80。这个上界相当宽松，但题目要求的也就只是这么多。

常见错误

- 数错一段的项数。该段由 k^n 至 $k^{n+1}-1$，共 $k^{n+1}-k^n$ 项，而不是 $k^{n+1}-k^n-1$ 项。
- 在 (iii) 中把第 n 段的全部 9^n 个不含 2 的数字串都算进去。首位数字不能是 0，所以只有 $8\cdot 9^n$ 个。若用 9^{n+1}，所得上界是 90，并不足够。
- 在 (ii) 中只得出至 $2^{N+1}-1$ 为止的部分和的上界便停下，没有说明为甚么整个无穷级数也有上界。

STEP 之外

这就是柯西（Cauchy）凝聚判别法（Cauchy's condensation test）：对于递减的正数数列，$\sum a_r$ 收敛当且仅当 $\sum 2^n a_{2^n}$ 收敛。由此可立即看出 $\sum r^{-p}$ 收敛当且仅当 $p>1$。第 (iii) 部分是 Kempner 级数（Kempner series）：把所有含某个指定数字的整数删去，原本发散的调和级数（harmonic series）便变成收敛，因为几乎所有很大的整数都包含每一个数字。Abbott 的 *Understanding Analysis*（《理解分析》）对凝聚判别法及各种比较判别法（comparison tests）有细致的论述。

详细解答

STEP III 2014，第 9 题

力学 · 有阻力的抛体

STEP III 2014，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/14-s3-q9

关键思路

题目已给出位置，所以速度（velocity）只需求导一次便得到；要验证运动方程（equation of motion），只需把 $\mathbf{v}$ 和 $\dot{\mathbf{v}}$ 代入 $\dot{\mathbf{v}} = \mathbf{g} - k\mathbf{v}$。之后的一切都围绕分量（components）展开：$\mathbf{r}$ 的 $\mathbf{j}$ 分量决定 T，而 $\mathbf{v}$ 在时间 T 时两个分量之比给出 $\tan\beta$。最后的不等式乍看很繁复，但只要令 $x = kT$ 并消去分母，它便正好化简为提示中的 $\sinh x > x$。

解答

速度，以及验证运动

我们把物体视为质点（particle）。牛顿（Newton）第二定律（Newton's second law）给出 $m\dot{\mathbf{v}} = m\mathbf{g} - mk\mathbf{v}$，所以运动方程是

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{g} - k\mathbf{v},$$

而初始条件（initial conditions）是在 $t = 0$ 时 $\mathbf{r} = \mathbf{0}$ 及 $\mathbf{v} = \mathbf{u}$。

这里 $\mathbf{g}$ 和 $\mathbf{u}$ 都是常向量，所以我们把给定的 $\mathbf{r}$ 逐个系数求导：

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{k - k\mathrm{e}^{-kt}}{k^2}\,\mathbf{g} + \frac{k\mathrm{e}^{-kt}}{k}\,\mathbf{u} = \frac{1 - \mathrm{e}^{-kt}}{k}\,\mathbf{g} + \mathrm{e}^{-kt}\,\mathbf{u}.$$

再求导一次得 $\dot{\mathbf{v}} = \mathrm{e}^{-kt}\mathbf{g} - k\mathrm{e}^{-kt}\mathbf{u}$。另一方面，

$$\mathbf{g} - k\mathbf{v} = \mathbf{g} - \left(1 - \mathrm{e}^{-kt}\right)\mathbf{g} - k\mathrm{e}^{-kt}\mathbf{u} = \mathrm{e}^{-kt}\mathbf{g} - k\mathrm{e}^{-kt}\mathbf{u},$$

两者相同，所以运动方程成立。在 $t = 0$ 时 $\mathrm{e}^0 = 1$，所以 $\mathbf{r} = \frac{0-1+1}{k^2}\mathbf{g} + 0\,\mathbf{u} = \mathbf{0}$ 及 $\mathbf{v} = 0\,\mathbf{g} + \mathbf{u} = \mathbf{u}$。两个初始条件都成立。

飞行时间

现在取 $\mathbf{u} = u\cos\alpha\,\mathbf{i} + u\sin\alpha\,\mathbf{j}$ 及 $\mathbf{g} = -g\,\mathbf{j}$。$\mathbf{r}$ 的 $\mathbf{j}$ 分量是

$$\mathbf{r}\cdot\mathbf{j} = -\frac{kt - 1 + \mathrm{e}^{-kt}}{k^2}\,g + \frac{1-\mathrm{e}^{-kt}}{k}\,u\sin\alpha.$$

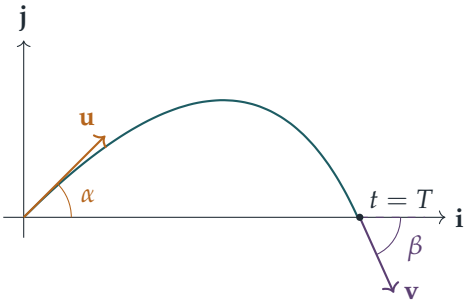
令它在 $t = T$ 时等于零，并乘以 k^2：

$$uk\sin\alpha\left(1-\mathrm{e}^{-kT}\right) = \left(kT - 1 + \mathrm{e}^{-kT}\right)g.$$

由于 $T > 0$ 及 $k > 0$，我们有 $1 - \mathrm{e}^{-kT} > 0$，所以可以除以它。把 $kT - 1 + \mathrm{e}^{-kT} = kT - \left(1 - \mathrm{e}^{-kT}\right)$ 代入，得

$$uk\sin\alpha = \left(\frac{kT}{1-\mathrm{e}^{-kT}} - 1\right)g,$$

即所求。



时间 T 时的角度

在 $t = T$ 时，$\mathbf{v} = \frac{1-\mathrm{e}^{-kT}}{k}\mathbf{g} + \mathrm{e}^{-kT}\mathbf{u}$ 的分量是

$$v_x = \mathrm{e}^{-kT}u\cos\alpha, \qquad v_y = -\frac{1-\mathrm{e}^{-kT}}{k}\,g + \mathrm{e}^{-kT}u\sin\alpha.$$

由于 $0 < \alpha < 90°$，所以 $v_x > 0$。我们预期 $v_y < 0$，因为质点正在下降，这一点稍后会证明。因此 $\mathbf{v}$ 与 $\mathbf{i}$ 之间的锐角（acute angle）满足 $\tan\beta = -v_y/v_x$。全式除以 v_x：

$$\tan\beta = \frac{\left(1-\mathrm{e}^{-kT}\right)g}{k\,\mathrm{e}^{-kT}u\cos\alpha} - \frac{\mathrm{e}^{-kT}u\sin\alpha}{\mathrm{e}^{-kT}u\cos\alpha} = \frac{\left(\mathrm{e}^{kT}-1\right)g}{uk\cos\alpha} - \tan\alpha,$$

这里用了 $\left(1-\mathrm{e}^{-kT}\right)/\mathrm{e}^{-kT} = \mathrm{e}^{kT} - 1$。

证明 $\tan\beta > \tan\alpha$

由上一个结果，$\tan\beta > \tan\alpha$ 等价于

$$\frac{\left(\mathrm{e}^{kT}-1\right)g}{uk\cos\alpha} > 2\tan\alpha \iff \left(\mathrm{e}^{kT}-1\right)g > 2uk\sin\alpha,$$

这是乘以 $uk\cos\alpha > 0$ 所得。现在用飞行时间的结果代替 $uk\sin\alpha$，这样 u 和 α 便完全消去。

> **想一想** 怎样把它变成与 $\sinh kT$ 有关的式子?

令 $x = kT > 0$。那么 $\frac{x}{1-\mathrm{e}^{-x}} = \frac{x\mathrm{e}^{x}}{\mathrm{e}^{x}-1}$，除以 $g > 0$ 之后，所需的不等式变为

$$\mathrm{e}^{x} - 1 > \frac{2x\mathrm{e}^{x}}{\mathrm{e}^{x}-1} - 2.$$

乘以 $\mathrm{e}^{x} - 1 > 0$（因此不等号方向不变）并展开：

$$\begin{aligned}
&\iff \mathrm{e}^{2x} - 2\mathrm{e}^{x} + 1 > 2x\mathrm{e}^{x} - 2\mathrm{e}^{x} + 2 \\
&\iff \mathrm{e}^{2x} - 1 > 2x\mathrm{e}^{x} \\
&\iff \mathrm{e}^{x} - \mathrm{e}^{-x} > 2x \qquad (\text{除以} \mathrm{e}^{x} > 0) \\
&\iff \sinh x > x.
\end{aligned}$$

每一步都可逆，而且当 $kT > 0$ 时 $\sinh kT > kT$（题目容许我们假设这一点）。所以 $\tan\beta > \tan\alpha$。

这也确定了 v_y 的正负：我们已证明 $-v_y/v_x > \tan\alpha > 0$ 且 $v_x > 0$，所以 $v_y < 0$。质点在时间 T 时确实正在下降，$\tan\beta$ 的公式因而成立。

最后，α 和 β 都在 $(0°, 90°)$ 之内，而 tan 在此区间上严格递增（strictly increasing）。所以由 $\tan\beta > \tan\alpha$ 得 $\beta > \alpha$。有阻力（resistance）时，质点落下的路径比上升时更陡。

验算

- 当 $k \to 0$ 时，利用 $\mathrm{e}^{-kt} \approx 1 - kt + \frac{1}{2}k^2t^2$，得 $\mathbf{r} \to \frac{1}{2}t^2\mathbf{g} + t\mathbf{u}$，这正是一般的抛射体（projectile）运动。此外，当 x 很小时 $\frac{x}{1-\mathrm{e}^{-x}} - 1 \approx \frac{x}{2}$，所以飞行时间的结果给出 $T \to 2u\sin\alpha/g$。同理 $\tan\beta \to \frac{gT}{u\cos\alpha} - \tan\alpha = \tan\alpha$，即对称的抛物线。
- 取 $k = 1\,\mathrm{s}^{-1}$、$g = 10\,\mathrm{m\,s}^{-2}$、$u = 10\,\mathrm{m\,s}^{-1}$ 及 $\alpha = 45°$。以数值方法解 $\mathbf{r}\cdot\mathbf{j} = 0$ 得 $T \approx 1.185\,\mathrm{s}$，比没有阻力时的 $1.414\,\mathrm{s}$ 短。落地角为 $\beta \approx 65.7° > 45°$，两条公式都吻合至小数点后很多位。用电脑对 $\dot{\mathbf{v}} = \mathbf{g} - k\mathbf{v}$ 作逐步数值积分（numerical integration），也能重现给定的 $\mathbf{r}$ 和 $\mathbf{v}$。

常见错误

- 取 $\tan\beta = v_y/v_x$ 而漏掉负号：落地时 v_y 是负数，而 β 是锐角。
- 除以或乘以 $1 - \mathrm{e}^{-kT}$ 或 $\mathrm{e}^{kT} - 1$ 时没有说明它是正数。正是这一点保证不等号方向正确。

► 停在 $\tan\beta > \tan\alpha$：要推出 $\beta > \alpha$，需要用到 tan 在锐角范围内递增这个事实。

STEP 之外

给定的 $\mathbf{r}$ 是线性微分方程（linear differential equation）$\ddot{\mathbf{r}} + k\dot{\mathbf{r}} = \mathbf{g}$ 的解，你可以用积分因子（integrating factor）e^{kt} 直接解出它。线性阻力（linear drag）是唯一一种能使 $\mathbf{i}$ 方向与 $\mathbf{j}$ 方向的运动分离的阻力定律。若采用更贴近现实的二次阻力（quadratic drag）$-mk|\mathbf{v}|\mathbf{v}$，两个方向的运动便会耦合（coupled），而且不存在闭式解（closed form）。至于轨迹不对称、下降时较陡这一特点，在两种情况下都成立。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）正是以这种思路讨论受阻力作用的运动。

详细解答

STEP III 2014，第 10 题

力学 · 耦合振动

STEP III 2014，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/14-s3-q10

本题在 *Beyond STEP* 第 11.2 节简正模式中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/normal-modes.html

关键思路

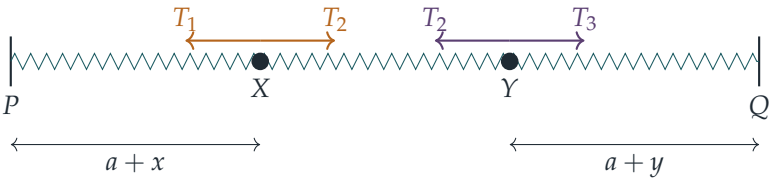
把所有东西放在同一条直线上，只用一个坐标，以 x 和 y 表示每条弹簧的伸长量（extension），然后由牛顿（Newton）第二定律（Newton's second law）得出两条耦合方程（coupled equations）。具有这种对称性的耦合方程，只要相加和相减就可以分离：$x-y$ 和 $x+y$ 各自满足一条简单的简谐方程，而且频率不同。Y 会否回到原处，就取决于这两个频率之比是否有理数（rational）。

解答

运动方程

把 X 和 Y 视为质点，弹簧视为轻弹簧并遵守胡克定律（Hooke's law）；桌面光滑，所以唯一的水平力就是弹簧的张力（tension）。所有力都沿直线 PQ 作用，运动亦一直在这条直线上。

沿直线从 P 起，以指向 Q 的方向量度位置，于是 P 位于 0，Q 位于 $3a$。这样 X 位于 $a+x$，而 Y 位于 $3a-(a+y)=2a-y$。



三条弹簧的长度和张力如下（由胡克定律，张力 $=\lambda\times$ 伸长量$/a$，负值表示推力（thrust））：

$$\begin{aligned} &PX:\ \text{长度}a+x, && T_1=\frac{\lambda x}{a},\\ &XY:\ \text{长度}(2a-y)-(a+x)=a-x-y, && T_2=-\frac{\lambda(x+y)}{a},\\ &YQ:\ \text{长度}a+y, && T_3=\frac{\lambda y}{a}. \end{aligned}$$

受张力的弹簧会把两端互相拉近。所以 X 受到指向 P 的 T_1 和指向 Y 的 T_2；取 $P\to Q$ 方向为正，

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=-T_1+T_2=-\frac{\lambda}{a}x-\frac{\lambda}{a}(x+y)=-\frac{\lambda}{a}(2x+y),$$

正是所求。（正负号会自动处理好：若 $T_2<0$，弹簧 XY 受压缩，会把 X 推向 P。）

至于 Y，它的位置是 $2a-y$，所以它在正方向的加速度是 $-\ddot{y}$。它受到指向 X 的 T_2（负方向）和指向 Q 的 T_3（正方向）：

$$-m\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=-T_2+T_3=\frac{\lambda}{a}(x+y)+\frac{\lambda}{a}y,\qquad\text{所以}\qquad m\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=-\frac{\lambda}{a}(x+2y).$$

这就是把第一条方程中的 x 和 y 互换，正如装置的对称性（对 PQ 的中点作反射）所要求的一样。

组合 $x-y$

把两条方程相减，

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}(x-y)=-\frac{\lambda}{a}\left[(2x+y)-(x+2y)\right]=-\frac{\lambda}{a}(x-y).$$

设 $\omega^2=\lambda/(ma)$，上式就是 $\dfrac{\mathrm{d}^2}{\mathrm{d}t^2}(x-y)=-\omega^2(x-y)$，即简谐运动（simple harmonic motion），所以

$$x-y=A\cos\omega t+B\sin\omega t.$$

起初 $XP=a$ 而 $YQ=\frac{1}{2}a$，所以在 $t=0$ 时 $x=0$、$y=-\frac{1}{2}a$，得 $x-y=\frac{1}{2}a$，因此 $A=\frac{1}{2}a$。两质点由静止释放，所以在 $t=0$ 时 $\dot{x}=\dot{y}=0$；求导后得 $\omega B=0$，所以 $B=0$。因此

$$x-y=\tfrac{1}{2}a\cos\omega t.$$

组合 $x+y$

改为把两条方程相加，

$$m\frac{\mathrm{d}^2}{\mathrm{d}t^2}(x+y)=-\frac{\lambda}{a}(3x+3y),\qquad \text{所以}\qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2}(x+y)=-3\omega^2(x+y).$$

这是角频率（angular frequency）为 $\sqrt{3}\,\omega$ 的简谐运动，所以 $x+y=C\cos\sqrt{3}\,\omega t+D\sin\sqrt{3}\,\omega t$。在 $t=0$ 时，$x+y=-\frac{1}{2}a$ 且其导数为 0，所以 $C=-\frac{1}{2}a$、$D=0$：

$$x+y=-\tfrac{1}{2}a\cos\sqrt{3}\,\omega t.$$

Y 永不回到原处

取两个结果之差的一半，得

$$y=\tfrac{1}{2}\big[(x+y)-(x-y)\big]=-\tfrac{1}{4}a\big(\cos\omega t+\cos\sqrt{3}\,\omega t\big).$$

（同理 $x=\frac{1}{4}a(\cos\omega t-\cos\sqrt{3}\,\omega t)$。）$Y$ 回到初始位置，当且仅当 $y=-\frac{1}{2}a$，即

$$\cos\omega t+\cos\sqrt{3}\,\omega t=2.$$

> ? **想一想** 两个余弦函数都是周期函数，那么它们终究总会再次对齐吧?

每个余弦值最多是 1，所以两者之和等于 2 只有在某一时刻 $t>0$ 两者同时等于 1 时才会发生。这需要

$$\omega t=2\pi n\quad\text{及}\quad\sqrt{3}\,\omega t=2\pi k$$

其中 n,k 为整数，且因 $t>0$ 而有 $n\geq 1$。两式相除，$\sqrt{3}=k/n$ 便会是有理数，但它并不是。所以不存在这样的 $t>0$，Y 永远不会回到初始位置。两个周期 $2\pi/\omega$ 和 $2\pi/(\sqrt{3}\,\omega)$ 是不可公度（incommensurable）的，所以两个振动永远不会再次同步，尽管它们可以任意接近同步。

验算

- 在 $t=0$ 时，公式给出 $x=\frac{1}{4}a(1-1)=0$ 及 $y=-\frac{1}{4}a(1+1)=-\frac{1}{2}a$，与起始位置吻合，而且两个导数在该处都等于零。
- 弹簧 XY 的长度是 $a-x-y=a+\frac{1}{2}a\cos\sqrt{3}\,\omega t$，一直介乎 $\frac{1}{2}a$ 与 $\frac{3}{2}a$ 之间，所以两质点永不相撞；同样 $PX=a+x\geq\frac{1}{2}a$ 及 $YQ=a+y\geq\frac{1}{2}a$。整个过程中模型在物理上都合理。
- 取 $m=a=1$、$\lambda=2$ 以数值方法逐步模拟运动，直至 $t=20$，结果与 x 和 y 的公式相差少于 10^{-10}。以电脑在 $0<\omega t\leq 20\,000$ 范围内搜索，发现 $\cos\omega t+\cos\sqrt{3}\,\omega t$ 最接近 2 时可达 1.999995，但从未到达 2。

常见错误

- 弄错 Y 方程的正负号：y 是从 Q 起量度的，所以 y 增加会使 Y 向负方向移动。
- 把 XY 的长度写成 $a+x+y$ 而不是 $a-x-y$；x 和 y 增加都会令中间的弹簧变短。
- 只论证两个周期不同。单是周期不同并不足够（周期 1 和 2 在 $t=2$ 时会再次对齐）；关键在于两者之比 $\sqrt{3}$ 是无理数（irrational）。

STEP 之外

组合 $x-y$ 和 $x+y$ 是这个系统的简正模（normal modes）：它们是矩阵 $\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix}$ 的特征向量（eigenvectors），对应的特征值（eigenvalues）1 和 3 给出频率的平方 ω^2 和 $3\omega^2$。耦合系统的每个微小振动都可以这样分解为简正模，而只有当所有频率之比都是有理数时，运动才是周期的；否则运动是准周期（quasi-periodic）的，永不重复，但会任意接近其起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）讲述耦合振子，而 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）则解释背后的特征向量方法。

详细解答

STEP III 2014，第 11 题

力学 • 圆周运动

STEP III 2014，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/14-s3-q11

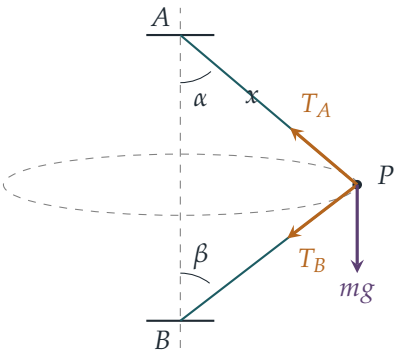
关键思路

质点受三个力作用（两条绳的张力（tension）及其重量），三者合起来必须恰好提供圆周运动所需的向心力（centripetal force）。分别沿铅垂及水平方向分解，便得到关于两个张力的两条线性方程。绳只能拉不能推，故 $T_B \geq 0$，由此得出所求的不等式。在临界情况下，下方的绳变得松弛，而在三角形 ABP 中应用余弦定理（cosine rule），便可知道 α 可以取哪些角度。

解答

建立模型

把 P 视为质点，两条绳视为轻而不可伸长。P 在一个以直线 AB 上一点为圆心的水平圆上运动，半径为 $r = x\sin\alpha$，所以其加速度为 $\omega^2 x\sin\alpha$，方向水平指向轴线。张力 T_A 沿 PA 作用，与向上的铅垂线成角 α。张力 T_B 沿 PB 作用，与向下的铅垂线成角 β。



求张力

沿铅垂方向（铅垂方向没有加速度）及沿水平指向轴线的方向分解：

$$T_A \cos\alpha - T_B \cos\beta = mg, \tag{1}$$

$$T_A \sin\alpha + T_B \sin\beta = m\omega^2 x \sin\alpha. \tag{2}$$

即使 β 是钝角，即 P 位于 B 之下，这两条方程仍然成立，因为此时 $\cos\beta < 0$，T_B 的铅垂分量指向上方，正合实际情况。

要消去 T_B，把 (1) 乘以 $\sin\beta$，(2) 乘以 $\cos\beta$，然后相加。利用 $\sin\alpha\cos\beta + \cos\alpha\sin\beta = \sin(\alpha+\beta)$，得

$$T_A = \frac{m\left(g\sin\beta + \omega^2 x \sin\alpha\cos\beta\right)}{\sin(\alpha+\beta)}.$$

要消去 T_A，把 (2) 乘以 $\cos\alpha$，(1) 乘以 $\sin\alpha$，然后相减：

$$T_B = \frac{m\sin\alpha\left(\omega^2 x\cos\alpha - g\right)}{\sin(\alpha+\beta)}.$$

这里 α 和 β 是三角形 ABP 的两个内角，所以 $0 < \alpha+\beta < \pi$，从而 $\sin(\alpha+\beta) > 0$。此外 $\sin\alpha > 0$。拉紧的绳满足 $T_B \geq 0$，因此

$$\omega^2 x\cos\alpha \geq g,$$

即所求。

情况 $\omega^2 x\cos\alpha = g$

此时 $T_B = 0$：下方的绳刚好拉直，但没有张力。于是方程 (1) 给出

$$T_A = \frac{mg}{\cos\alpha},$$

这与 (2) 一致，因为此时 (2) 给出 $T_A = m\omega^2 x$。另外注意 $\cos\alpha = g/(\omega^2 x) > 0$，所以 α 是锐角。

> ? **想一想** 一条只涉及 h、d 和 α 的不等式，究竟从何而来?

它来自几何，而不是力学。在三角形 ABP 的顶点 A 处应用余弦定理：

$$d^2 = h^2 + x^2 - 2hx\cos\alpha \quad\Longleftrightarrow\quad x^2 - (2h\cos\alpha)\,x + (h^2 - d^2) = 0.$$

长度 x 是这个二次方程（quadratic equation）的实根，所以其判别式（discriminant）非负：

$$4h^2\cos^2\alpha - 4(h^2 - d^2) \geq 0 \quad\Longleftrightarrow\quad h^2\cos^2\alpha \geq h^2 - d^2.$$

由于 α 是锐角且 $h > d$，$h\cos\alpha$ 和 $\sqrt{h^2-d^2}$ 均为正数，因此可以两边开平方，得

$$h\cos\alpha \geq \sqrt{h^2 - d^2}.$$

T_A 的上下界

质点确实不在轴线上，所以 $0 < \alpha < \frac{\pi}{2}$，从而 $0 < \cos\alpha < 1$。因此 $T_A = mg/\cos\alpha > mg$。至于另一个界，由上一个结果得 $\cos\alpha \geq \sqrt{h^2 - d^2}/h$，所以

$$mg < T_A = \frac{mg}{\cos\alpha} \leq \frac{mgh}{\sqrt{h^2 - d^2}}.$$

何时取得上界

等号成立需要 $h\cos\alpha = \sqrt{h^2 - d^2}$，即判别式为零，此时二次方程有重根

$$x = h\cos\alpha = \sqrt{h^2 - d^2}.$$

于是 $x^2 + d^2 = h^2$，即 $AP^2 + BP^2 = AB^2$。根据毕氏定理的逆定理（converse of Pythagoras' theorem），角 APB 是直角：两条绳互相垂直，而 P 位于以 AB 为直径的圆上。换一个说法，AP 是以 B 为圆心、半径为 d 的圆的切线（tangent），这正是 α 在此位置取得最大可能值的原因。在这个位置，$\sin\alpha = d/h$，而 P 位于 B 之上 d^2/h 的高度。对于不等式容许的其他任何锐角 α，都有两个可能的三角形（两个 x 值），而两者在这个位置合而为一。

验算

- 若 ω 非常大，则 $T_B \approx m\omega^2 x\sin\alpha/\sin(\alpha+\beta)$ 以 ω^2 的量级增长，因为此时两条绳都必须提供很大的向心力。
- 在极端位置取 $h = 3$、$d = 2$：$\cos\alpha = \sqrt{5}/3$，$x = \sqrt{5}$，且 $x^2 + d^2 = 5 + 4 = 9 = h^2$。于是 $T_A/(mg) = 3/\sqrt{5} \approx 1.342$，等于 $h/\sqrt{h^2 - d^2}$。
- 以电脑检查多个随机构形，把 T_A 和 T_B 的公式代回向量形式的运动方程，所得残差均为零。

常见错误

- 把圆的半径当成 x，而不是 $x\sin\alpha$。
- 把 T_B 的铅垂方向弄反。由于 β 是从 BA 量起，绳 PB 把 P 向下拉（当 β 是锐角时）。
- 没有先确认两边都是正数，便把不等式两边平方或开平方。

STEP 之外

临界情况 $T_B = 0$ 是一个约束变成*非作用*（inactive）的例子：绳是一种单边约束（one-sided constraint），只能拉、不能推。这类带有不等式约束的问题，在力学和最优化中会

引出带符号条件的拉格朗日（Lagrange）乘数（Lagrange multipliers）。圆锥摆（conical pendulum）及其相关问题，在 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）中有很好的论述。

详细解答

STEP III 2014，第 12 题

概率与统计 • 对数正态分布

> **STEP III 2014，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/14-s3-q12

关键思路

由于 $y = \mathrm{e}^x$ 是严格递增（strictly increasing）的，只要取对数（logarithm），关于 Y 的命题便可化为关于 X 的命题：$Y \le y$ 当且仅当 $X \le \ln y$。这样便能立即得出 Y 的中位数（median）和密度。至于平均值（mean），技巧在于把指数部分配方（complete the square），使被积函数变成另一个正态分布的密度函数，而我们早已知道它的积分等于 1。

解答

第 (i) 部分

由于 e^x 严格递增，事件 $\{Y \le \mathrm{e}^{x_m}\}$ 与事件 $\{X \le x_m\}$ 是同一个事件。因此

$$\mathrm{P}\left(Y \le \mathrm{e}^{x_m}\right) = \mathrm{P}(X \le x_m) = \tfrac{1}{2},$$

所以 Y 的中位数是

$$y_m = \mathrm{e}^{x_m}.$$

（中位数可以这样经过任何递增函数而保持对应；平均值却不可以，第 (iv) 部分会说明这一点。）

第 (ii) 部分

Y 只取正值。对于 $y > 0$，设 G 为 Y 的累积分布函数（cumulative distribution function）。利用与前面相同的等价关系，

$$G(y) = \mathrm{P}(Y \le y) = \mathrm{P}(X \le \ln y) = F(\ln y).$$

利用连锁法则（chain rule）求导，并使用 $F' = f$，得

$$g(y) = G'(y) = f(\ln y) \cdot \frac{1}{y} = \frac{f(\ln y)}{y}, \qquad y > 0,$$

此即所求（而当 $y \le 0$ 时 $g(y) = 0$）。

众数（mode）λ 是 g 取最大值之处。它位于开区间 $(0, \infty)$ 内，而 g 在该区间上可导，所以 $g'(\lambda) = 0$。由商法则（quotient rule），

$$g'(y) = \frac{f'(\ln y) \cdot \frac{1}{y} \cdot y - f(\ln y)}{y^2} = \frac{f'(\ln y) - f(\ln y)}{y^2}.$$

由于 $y^2 \neq 0$，令 $g'(\lambda) = 0$ 便得

$$f'(\ln \lambda) = f(\ln \lambda),$$

这正是所求的条件。

第 (iii) 部分

函数

$$\frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu-\sigma^2)^2/(2\sigma^2)}$$

正好是一个 $\mathrm{N}(\mu + \sigma^2, \sigma^2)$ 随机变量（random variable）的概率密度函数（probability density function）：它就是题目给出的正态密度，只是把 μ 换成 $\mu + \sigma^2$。任何概率密度函数在整条实数线上的积分都等于 1，这便解释了题目所述的积分。

现在

$$\mathrm{E}(Y) = \mathrm{E}\big(\mathrm{e}^X\big) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^x \, \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \, \mathrm{d}x.$$

我们把合并后的指数配方。展开 $(x - \mu - \sigma^2)^2 = (x-\mu)^2 - 2\sigma^2 x + 2\mu\sigma^2 + \sigma^4$，得

$$\begin{aligned} x - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{(x-\mu)^2 - 2\sigma^2 x}{2\sigma^2} \\ &= -\frac{(x-\mu-\sigma^2)^2 - 2\mu\sigma^2 - \sigma^4}{2\sigma^2} \\ &= -\frac{(x-\mu-\sigma^2)^2}{2\sigma^2} + \mu + \tfrac{1}{2}\sigma^2. \end{aligned}$$

常数 $\mathrm{e}^{\mu + \frac{1}{2}\sigma^2}$ 可提到积分号外，所以

$$\mathrm{E}(Y) = \mathrm{e}^{\mu+\frac{1}{2}\sigma^2} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-(x-\mu-\sigma^2)^2/(2\sigma^2)} \, \mathrm{d}x = \mathrm{e}^{\mu+\frac{1}{2}\sigma^2},$$

这里用到了上面解释过的积分。这就是所求的结果。

第 (iv) 部分

我们把三个量都明确求出来。

中位数。正态密度关于 μ 对称，所以 $x_m = \mu$，再由第 (i) 部分得 $y_m = \mathrm{e}^{\mu}$。

众数。对正态密度求导，

$$f'(x) = -\frac{x-\mu}{\sigma^2} f(x).$$

? **想一想** 当 f 永不为零时，我们该如何运用第 (ii) 部分的条件?

由于对每个 x 都有 $f(x) > 0$，我们可以把条件 $f'(\ln\lambda) = f(\ln\lambda)$ 除以 $f(\ln\lambda)$，得到

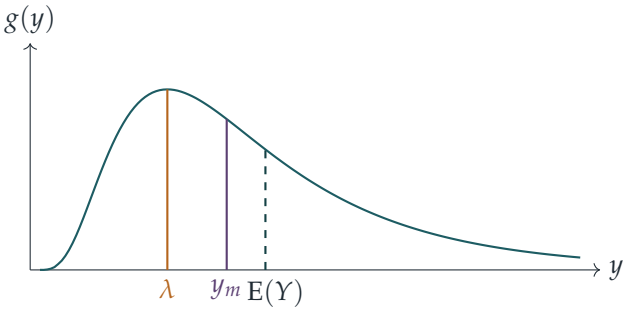
$$-\frac{\ln\lambda - \mu}{\sigma^2} = 1 \quad\Longrightarrow\quad \ln\lambda = \mu - \sigma^2 \quad\Longrightarrow\quad \lambda = \mathrm{e}^{\mu-\sigma^2}.$$

这是 g 唯一的驻点（stationary point），而题目已告诉我们 Y 有唯一的众数，所以它就是众数。

比较。由于 $\sigma > 0$，我们有 $\mu - \sigma^2 < \mu < \mu + \frac{1}{2}\sigma^2$，而 e^t 严格递增，所以

$$\lambda = \mathrm{e}^{\mu-\sigma^2} \;<\; y_m = \mathrm{e}^{\mu} \;<\; \mathrm{E}(Y) = \mathrm{e}^{\mu+\frac{1}{2}\sigma^2},$$

此即所求。这些严格不等式（inequality）依赖于 $\sigma^2 > 0$，而这对任何真正的正态分布（normal distribution）都成立。



上图显示 $\mu = 0$、$\sigma = 0.6$ 时 Y 的密度：右边长长的尾部把平均值拉到最右，中位数居中，而峰值则在两者的左边。

验算

- 当 $\sigma \to 0$ 时，Y 集中于 e^{μ}，三个值都趋向 e^{μ}，理应如此。
- 取 $\mu = 0$、$\sigma = 1$：$\lambda = \mathrm{e}^{-1} \approx 0.368$，$y_m = 1$，而 $\mathrm{E}(Y) = \mathrm{e}^{1/2} \approx 1.649$。对 $\mathrm{e}^x f(x)$ 作数值积分，以及以数值方法搜索 g 的最大值，结果都与这些数值吻合。

- 比值 $\mathrm{E}(Y)/\lambda = \mathrm{e}^{3\sigma^2/2}$ 只取决于 σ，这是合理的：改变 μ 只是把 Y 乘以因子 e^{μ} 作缩放。

常见错误

- 把 Y 的密度写成 $f(\ln y)$，忘记了连锁法则带来的因子 $1/y$。缺少这个因子，众数便会算成 e^{μ}，与中位数相同。
- 误以为 $\mathrm{E}(\mathrm{e}^{X}) = \mathrm{e}^{\mathrm{E}(X)}$。这对中位数成立，但对平均值不成立。
- 配方时出现正负号错误：要小心展开 $(x-\mu-\sigma^2)^2$，并检查剩下的常数。

STEP 之外

这里的 Y 服从对数正态（lognormal）分布，常用来为收入、股价和粒子大小建立模型。不等式 $\mathrm{E}(\mathrm{e}^{X}) \geq \mathrm{e}^{\mathrm{E}(X)}$ 是凸函数（convex function）的琴生不等式（Jensen's inequality）的一个特例，而第 (iii) 部分的计算其实就是正态分布的矩母函数（moment generating function）$\mathrm{E}(\mathrm{e}^{tX}) = \mathrm{e}^{\mu t + \frac{1}{2}\sigma^2 t^2}$ 在 $t=1$ 时的值。Blitzstein 与 Hwang 所著的 *Introduction to Probability*（《概率导论》）涵盖了随机变量的变换、矩母函数以及对数正态分布。

详细解答

STEP III 2014，第 13 题

概率与统计 · 概率生成函数

> **STEP III 2014，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/14-s3-q13

关键思路

就第一回合的结果作条件分析。这个游戏没有记忆：在任何一个游戏得以继续的回合之后，接下来的进程就是同一个游戏的全新副本，因此从那一刻起的额外得分与 N 本身具有相同的分布（distribution）。这样便得到一条以 G 自身表示 G 的方程，我们解出这条方程，再把结果展开成几何级数（geometric series）。

解答

记 $p_n = P(N = n)$，于是 $G(t) = \sum_{n\geq 0} p_n t^n$。设 E_1、E_2、E_3 分别为第一回合结果是 1、2 或 3 的事件，因此 $P(E_1) = a$、$P(E_2) = b$、$P(E_3) = c$。

第 (i) 部分

在 E_1 发生的条件下，游戏以得分 0 结束，所以 $P(N = 0 \mid E_1) = 1$，而当 $n \geq 1$ 时 $P(N = n \mid E_1) = 0$。因此条件生成函数（conditional generating function）为

$$\sum_{n\geq 0} P(N = n \mid E_1)\, t^n = 1 \cdot t^0 = 1.$$

第 (ii) 部分

在 E_2 发生的条件下，第一回合后得分仍是 0，游戏继续。由于每个回合的概率相同，而且不受之前的结果影响，从第二回合起的各回合便构成一个从得分 0 开始的全新游戏。设 N' 为这个新游戏所产生的得分，它与 N 同分布。最终得分为 $0 + N' = N'$，所以

$$P(N = n \mid E_2) = P(N' = n) = p_n \quad \text{对所有} n \geq 0,$$

而条件生成函数为 $\sum p_n t^n = G(t)$。

第 (iii) 部分

同理，在 E_3 发生的条件下，第一回合后得分为 1，其余部分是一个得分为 N' 的全新游戏，所以最终得分为 $1+N'$。因此 $P(N=0 \mid E_3)=0$，而当 $n \geq 1$ 时 $P(N=n \mid E_3)=p_{n-1}$；条件生成函数为 $\sum_{n\geq 1} p_{n-1}t^n = tG(t)$。

根据全概率定律（law of total probability），

$$p_n = a\,P(N=n \mid E_1) + b\,P(N=n \mid E_2) + c\,P(N=n \mid E_3).$$

$G(t)$ 中 t^n 的系数是 p_n。在右边，1、$G(t)$ 和 $tG(t)$ 中 t^n 的系数恰好就是第 (i)–(iii) 部分所得的 $P(N=n \mid E_1)$、$P(N=n \mid E_2)$ 和 $P(N=n \mid E_3)$。所以对所有 $n \geq 0$，$G(t)$ 与 $a+bG(t)+ctG(t)$ 中 t^n 的系数都相等，这就证明了

$$G(t) = a + bG(t) + ctG(t).$$

想一想 我们可以除以 $1-b-ct$ 吗？把它展开成几何级数又是否合理？

移项得 $G(t)(1-b-ct)=a$。由于 $1-b=a+c>c>0$，只要 $|t|<(1-b)/c$，便有 $1-b-ct\neq 0$；又因为 $(1-b)/c = 1+a/c > 1$，这个范围包含 $|t|\leq 1$。对于这样的 t，

$$G(t) = \frac{a}{1-b-ct} = \frac{a}{1-b}\cdot\frac{1}{1-\dfrac{ct}{1-b}} = \frac{a}{1-b}\sum_{n\geq 0}\left(\frac{c}{1-b}\right)^n t^n,$$

因为 $|ct/(1-b)|<1$。读出 t^n 的系数，得

$$P(N=n) = \frac{ac^n}{(1-b)^{n+1}}, \qquad n \geq 0.$$

（等价地，比较系数可得 $p_0 = a+bp_0$，以及当 $n\geq 1$ 时 $p_n = bp_n + cp_{n-1}$，即 $p_0 = a/(1-b)$ 及 $p_n = \frac{c}{1-b}p_{n-1}$，同样得出上述公式。）

第 (iv) 部分

平均值（mean）为 $\mu = G'(1)$。对 $G(t) = a(1-b-ct)^{-1}$ 求导，

$$G'(t) = \frac{ac}{(1-b-ct)^2}, \qquad \text{因此} \qquad \mu = G'(1) = \frac{ac}{(1-b-c)^2} = \frac{ac}{a^2} = \frac{c}{a},$$

这里用了 $1-b-c=a$。于是 $1+\mu = \dfrac{a+c}{a} = \dfrac{1-b}{a}$，所以

$$\frac{\mu^n}{(1+\mu)^{n+1}} = \frac{c^n}{a^n}\cdot\frac{a^{n+1}}{(1-b)^{n+1}} = \frac{ac^n}{(1-b)^{n+1}} = P(N=n),$$

得证。

验算

- 概率总和为 1：$\sum_n p_n = G(1) = a/(1-b-c) = a/a = 1$。
- 直接论证：忽略那些“白费”的结果 2 回合。余下的每个回合，以概率 $a/(a+c)$ 结束游戏，或以概率 $c/(a+c)$ 得一分，所以 N 是第一次结束之前所得的分数，而 $P(N=n) = \left(\frac{c}{a+c}\right)^n \frac{a}{a+c}$；由于 $a+c = 1-b$，这与上述公式相同。
- 取 $a = \frac{1}{5}$、$b = \frac{3}{10}$、$c = \frac{1}{2}$，得 $\mu = \frac{5}{2}$，以及 $P(N=0) = \frac{2}{7} \approx 0.2857$、$P(N=1) = \frac{10}{49} \approx 0.2041$。以电脑模拟 2×10^5 局游戏，得到 0.2855、0.2038，平均得分为 2.498。

常见错误

- 把情况 3 的条件生成函数写成 $G(t)$ 而不是 $tG(t)$：已经得到的那一分会令 t 的每个幂次都提高一次。
- 忘记说明为何第一回合之后的游戏是原游戏的副本：这有赖每个回合的概率都相同，而且与过去独立。

STEP 之外

就第一步作条件分析，是马可夫链（Markov chain）理论的基本工具，在该理论中称为首步分析（first-step analysis）；而这里的答案其实是一个经过伪装的几何分布（geometric distribution）。第 (iv) 部分说明，在 $\{0,1,2,\dots\}$ 上的几何分布单凭其平均值便可确定。同样的“为 G 建立方程”技巧也可处理分支过程（branching process），并由此导出关于一个姓氏最终消失的概率的著名结果。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）对生成函数和分支过程的讲解非常易读。

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

2014 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。

