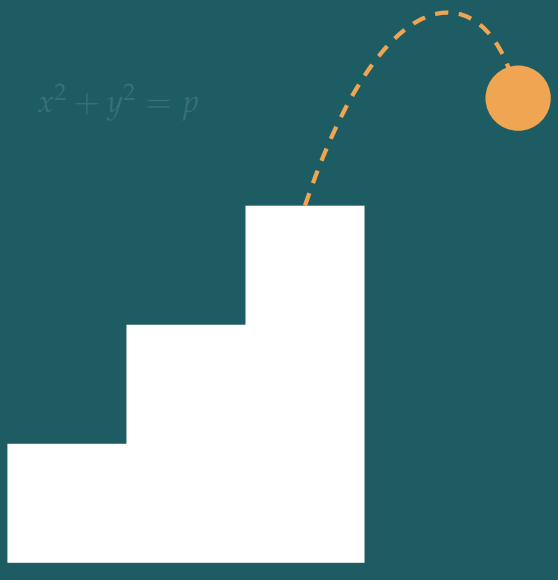
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2009 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

2009 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2009 年详细解答

2009 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 因数的计数 纯数学 . . . 3

第 2 题 三次曲线的切线 纯数学 . . . 6

第 3 题 实根的计数 纯数学 . . . 10

第 4 题 余弦公式 纯数学 . . . 14

第 5 题 约束最优化 纯数学 . . . 17

第 6 题 代换积分法 纯数学 . . . 21

第 7 题 分部积分法 纯数学 . . . 25

第 8 题 圆与切线 纯数学 . . . 28

第 9 题 相撞的抛体 力学 . . . 31

第 10 题 斜面上的连接质点 力学 . . . 35

第 11 题 碰撞与恢复系数 力学 . . . 38

第 12 题 概率与不等式 概率与统计 . . . 41

第 13 题 排列与游程 概率与统计 . . . 44

STEP II

第 1 题 曲线的对称性 纯数学 . . . 47

第 2 题 曲线描绘与近似 纯数学 . . . 51

第 3 题 三角恒等式与根式 纯数学 . . . 55

第 4 题 多项式的整除性 纯数学 . . . 59

第 5 题 嵌套平方根 纯数学 . . . 62

第 6 题 无穷级数的界限 纯数学 . . . 65

第 7 题 观察法积分 纯数学 . . . 69

第 8 题 向量几何 纯数学 . . . 73

第 9 题 质心 力学 . . . 76

第 10 题 碰撞与恢复系数 力学 80
第 11 题 功率与阻力 力学 84
第 12 题 正态分布 概率与统计 88
第 13 题 项分布的期望成本 概率与统计 92

STEP III

第 1 题 圆的坐标几何 纯数学 95
第 2 题 幂级数解 纯数学 98
第 3 题 极限与曲线描绘 纯数学 102
第 4 题 拉普拉斯变换 纯数学 106
第 5 题 根的对称函数 纯数学 109
第 6 题 托勒密定理 纯数学 112
第 7 题 逐次微分与数学归纳法 纯数学 116
第 8 题 极限与归约公式 纯数学 120
第 9 题 地体轨道的角度 力学 124
第 10 题 简谐运动 力学 128
第 11 题 变质量运动 力学 132
第 12 题 概率生成函数 概率与统计 136
第 13 题 中心极限定理 概率与统计 140

详细解答

STEP I 2009，第 1 题

纯数学 · 因数的计数

> **STEP I 2009，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/09-s1-q1

关键思路

$p^a q^b$（其中 p、q 为相异质数）的每一个因数（factor）都可写成 $p^i q^j$，其中 $0 \le i \le a$ 及 $0 \le j \le b$；而根据唯一分解定理（unique factorisation），不同的 (i, j) 会给出不同的因数。因此因数的数目是一个乘积 $(a+1)(b+1)$，而且无论有多少个质数，同样的结论都成立。计算真因数（proper factor）的数目只需再减去 2，于是两部分都变成把一个数写成乘积的问题。

解答

以 $d(N)$ 表示 N 的正因数的数目，包括 1 和 N 在内。若

$$N = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$$

其中 p_i 为相异质数且 $a_i \ge 1$，则 N 的正因数是 $p_1^{e_1} \cdots p_k^{e_k}$，其中 $0 \le e_i \le a_i$；每个 e_i 有 $a_i + 1$ 种选择，而根据唯一分解定理，不同的选择给出不同的因数。因此

$$d(N) = (a_1 + 1)(a_2 + 1) \cdots (a_k + 1).$$

当 $N > 1$ 时，因数 1 和 N 并不相同，所以真因数的数目是 $d(N) - 2$。

第 (i) 部分

对于 $3^2 \times 5^3$，我们得到 $d = (2+1)(3+1) = 12$，所以共有 $12 - 2 = 10$ 个真因数，符合题目所求。

现在考虑 $3^m \times 5^n$。要令这个数是整数，而且至少有一个真因数，我们需要 $m, n \ge 0$（3 或 5 的负数次幂无法被另一个质数抵销）。它的因数是 $3^i 5^j$，其中 $0 \le i \le m$、$0 \le j \le n$，

所以 $d=(m+1)(n+1)$（这亦涵盖 $m=0$ 或 $n=0$ 的情况）。我们需要

$$(m+1)(n+1)=12,$$

也就是把 12 分解成两个正整数的有序乘积。这些分解是

$$1\times 12,\quad 2\times 6,\quad 3\times 4,\quad 4\times 3,\quad 6\times 2,\quad 12\times 1,$$

从而得出

$$(m,n)=(0,11),\ (1,5),\ (2,3),\ (3,2),\ (5,1),\ (11,0).$$

同样根据唯一分解定理，每一组都给出不同的整数。其中一组 $(2,3)$ 就是题目给出的数，所以这种形式而又有 10 个真因数的**其他**整数恰好有 **5** 个：5^{11}、3×5^5、$3^3\times 5^2$、$3^5\times 5$ 和 3^{11}。

第 (ii) 部分

我们需要 $d(N)=426+2=428$，而

$$428=2^2\times 107,$$

其中 107 是质数（它不能被 2、3、5 或 7 整除，而且 $11^2>107$）。

想一想 N 的质因数分解（prime factorisation）可能有哪些形式?

各个 a_i+1 都至少是 2，而它们的乘积是 428。由于 107 是质数，当中恰好有一个是 107 的倍数，而它只能是 107、214 或 428；其余各数的乘积则分别是 4、2 或 1。不计次序的话，a_i+1 这组数的可能情况是

$$\{428\},\qquad \{214,\,2\},\qquad \{107,\,4\},\qquad \{107,\,2,\,2\},$$

所以指数组成的多重集（multiset）是

$$\{427\},\qquad \{213,\,1\},\qquad \{106,\,3\},\qquad \{106,\,1,\,1\}.$$

想一想 指数确定之后，应该把它们分配给哪些质数?

指数既定，要令 N 尽可能小，就应该使用最小的质数，并把最大的指数放在最小的质数上。使用最小的可用质数显然最好：把一个质数换成一个较小而未用过的质数，N 就会变小。另外，若 $p<q$，但 p 的指数是 a、q 的指数是 b，而 $a<b$，则把两个指数对调，相当于把 N 乘以

$$\frac{p^bq^a}{p^aq^b}=\left(\frac{p}{q}\right)^{b-a}<1,$$

所以对调之后 N 会变小。因此每种形式的最佳候选是

$$2^{427}, \qquad 2^{213}\times 3, \qquad 2^{106}\times 3^3, \qquad 2^{106}\times 3\times 5.$$

比较它们的大小：

- $2^{106}\times 3\times 5 = 2^{106}\times 15 < 2^{106}\times 27 = 2^{106}\times 3^3$；
- $2^{106}\times 27 < 2^{106}\times 2^5 = 2^{111}$，而这远小于 $2^{213}\times 3$ 和 2^{427}。

因此

$$N = 2^{106}\times 3\times 5.$$

验算一下：$d(N) = 107\times 2\times 2 = 428$，所以 N 有 426 个真因数。

验算

- 取一个较小的例子，直接列出 $3\times 5^5 = 9375$ 的因数：3^i5^j，其中 $i\in\{0,1\}$、$j\in\{0,\dots,5\}$，共有 12 个因数，所以有 10 个真因数。
- 用电脑对所有满足 $(a+1)(b+1)(c+1)(d+1) = 428$ 的指数组合 $2^a3^b5^c7^d$ 作一次简短的搜索，可确认最小的是 $2^{106}\times 15 \approx 1.22\times 10^{33}$，其次是 $2^{106}\times 27$。

常见错误

- 在第 (i) 部分答 6：题目问的是*其他*整数，所以题目给出的那一个不计算在内。
- 遗漏 $m = 0$ 或 $n = 0$ 的情况（即 5^{11} 和 3^{11}），它们同样符合题目要求的形式。
- 以为第 (ii) 部分的答案必须用上尽可能多的质数：107 是质数，这就迫使其中一个指数非常大，真正可以选择的只有因数 4 如何拆分。

STEP 之外

函数 $d(N)$ 是一个*积性*（multiplicative）函数：只要 m 和 n 互质（coprime），就有 $d(mn) = d(m)d(n)$，而这正是本题所用的乘积公式。因数数目比每一个较小的数都多的数，就是拉马努金（Ramanujan）的*高度合成数*（highly composite numbers），而第 (ii) 部分对调指数的论证，正是他刻划这类数的第一步。想阅读有关唯一分解定理和算术函数（arithmetic functions）的内容，Davenport 的 *The Higher Arithmetic*（一本初等数论经典入门书）是个好去处。

详细解答

STEP I 2009，第 2 题

纯数学 · 三次曲线的切线

STEP I 2009，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/09-s1-q2

本题在 *Beyond STEP* 第 1.9 节三次曲线的切线中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/tangents-cubic.html

关键思路

切线（tangent）在切点处与曲线相交两次：把切线方程代入曲线方程时，切点对应的是一个重根（repeated root）。一旦知道三次方程（cubic）的一个重根，余下的根便可藉因式分解直接得出，而这第三个交点的坐标原来是有理数。消去分母之后，曲线方程就化成一个立方和。

解答

在 $(-a, b)$ 处的切线

首先，$(-a, b)$ 确实在曲线上，因为 $x^3 + a^3 + b^3 = -a^3 + a^3 + b^3 = b^3 = y^3$。

把 $y^3 = x^3 + a^3 + b^3$ 对 x 作隐函数微分（implicit differentiation）（其中 a 和 b 为常数），得

$$3y^2 \frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2, \qquad \text{所以} \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{x^2}{y^2}$$

在 $y \neq 0$ 的地方均成立。在 $(-a, b)$ 处，由于 $b > 0$，斜率（gradient）为 a^2/b^2。因此切

线为

$$\begin{aligned} y - b &= \frac{a^2}{b^2}(x + a) \\ b^2y - b^3 &= a^2x + a^3 \\ b^2y - a^2x &= a^3 + b^3, \end{aligned}$$

正是所需的结果。

当 $a = 1, b = 2$ 时的交点

现在曲线是 $y^3 = x^3 + 9$，而在 $(-1, 2)$ 处的切线是 $4y - x = 9$，即 $y = \frac{1}{4}(x + 9)$。代入曲线方程：

$$\begin{aligned} \frac{(x+9)^3}{64} &= x^3 + 9 \\ x^3 + 27x^2 + 243x + 729 &= 64x^3 + 576 \\ 0 &= 63x^3 - 27x^2 - 243x - 153. \end{aligned}$$

每个系数都可被 9 整除，全式除以这个数后得

$$7x^3 - 3x^2 - 27x - 17 = 0,$$

这正是题目要求的三次方程。

求 p, q, r, s

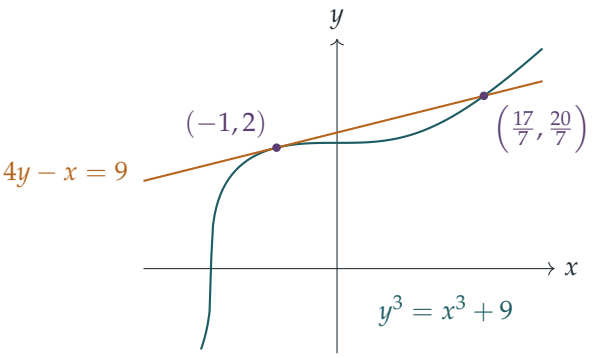
> ? **想一想** 我们已经知道这个三次方程的一个根。为甚么它应该是一个二重根?

直线在 $x = -1$ 处与曲线相切，所以 $x = -1$ 是重根：在切点附近，直线与曲线在一阶上吻合，因此两者之差在该处以二阶趋于零。故此 $(x + 1)^2 = x^2 + 2x + 1$ 是一个因式。比较 x^3 项和常数项，另一个因式必定是 $7x - 17$，我们可以验证：

$$(x + 1)^2(7x - 17) = 7x^3 + 14x^2 + 7x - 17x^2 - 34x - 17 = 7x^3 - 3x^2 - 27x - 17.$$

所以切线与曲线再次相交于 $x = \frac{17}{7}$ 处，而在该处

$$y = \tfrac{1}{4}\left(\tfrac{17}{7} + 9\right) = \tfrac{1}{4} \cdot \tfrac{80}{7} = \tfrac{20}{7}.$$



由于这一点在 $y^3 = x^3 + 9$ 上，

$$\left(\frac{20}{7}\right)^3 = \left(\frac{17}{7}\right)^3 + 9.$$

乘以 7^3 得 $20^3 = 17^3 + 9 \cdot 7^3$。最后一项并不是立方数，但 $9 = 1 + 8 = 1^3 + 2^3$，所以

$$9 \cdot 7^3 = 7^3 + 2^3 \cdot 7^3 = 7^3 + 14^3.$$

（原来的形式 $a^3 + b^3$（其中 $a = 1, b = 2$）就在这里重新出现。）因此

$$20^3 = 17^3 + 14^3 + 7^3,$$

我们可取 $p = 20, q = 17, r = 14, s = 7$（后三个数的次序可任意排列）。

验算

直接计算：$17^3 + 14^3 + 7^3 = 4913 + 2744 + 343 = 8000 = 20^3$。此外，三次方程的三个根是 $-1, -1, \frac{17}{7}$，其和为 $\frac{3}{7}$，其积为 $\frac{17}{7}$，与从系数读出的 $3/7$ 和 $17/7$ 相符。以数值方法计算 $y = \sqrt[3]{x^3 + a^3 + b^3}$ 在 $x = -a$ 处的导数，结果与 a^2/b^2 一致。

常见错误

- 没有利用二重根：试图从头解这个三次方程，而不是先除去因式 $(x + 1)^2$。
- 停在 $20^3 = 17^3 + 3087$ 这一步。我们需要的是立方相加的正整数，而 $3087 = 9 \cdot 7^3$ 必须拆成 $7^3 + 14^3$。

STEP 之外

这是处理三次曲线的经典*切线法*（tangent method）。一条直线与三次曲线相交于三点（按重数计算），所以两个有理点（rational point）（或如本题般，一个点计算两次）便

会产生第三个有理点。在椭圆曲线（elliptic curve）上，这种弦切作图法（chord-and-tangent construction）成为一个群运算法则（group law），是现代数论中一大领域的起点。Davenport 的 *The Higher Arithmetic* 讨论了这类丢番图方程（Diophantine equation），而 Silverman 与 Tate 的 *Rational Points on Elliptic Curves* 则完整地发展了当中的几何。

详细解答

STEP I 2009，第 3 题

纯数学 · 实根的计数

STEP I 2009，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/09-s1-q3

关键思路

把等式两边同时取适当的幂次来消去根号，但要留意这一步会失去甚么。平方根（square root）永不为负，所以只有同时要求 $x \geq 0$，平方这一步才可以逆推回去；实数立方根（cube root）则没有这种限制，因为每个实数都恰好有一个实数立方根。处理好这一点之后，每一次计数解的个数都归结为一个二次方程的根，而这些根的正负号可以从两根之和与两根之积看出来。

解答

第 (i) 部分：平方根方程

这里 $(a-x)^{\frac{1}{2}}$ 是指非负的平方根，所以任何解都满足 $x \geq 0$，而两边平方得 $x^2 = a - x$，即 $x^2 + x - a = 0$。反过来，若 $x \geq 0$ 且 $x^2 + x - a = 0$，则 $a - x = x^2 \geq 0$，并且 $(a-x)^{\frac{1}{2}} = \sqrt{x^2} = |x| = x$。因此

$$x = (a-x)^{\frac{1}{2}} \iff x^2 + x - a = 0 \text{ 且 } x \geq 0,$$

于是我们要计数的是二次方程 $x^2 + x - a = 0$ 的非负根的个数。设它的根为 α, β（当两根为实数时），则

$$\alpha + \beta = -1, \qquad \alpha\beta = -a.$$

情况 $a > 0$。判别式（discriminant）$1 + 4a$ 为正，所以有两个实根，而两根之积 $-a$ 为负。因此一根为正、另一根为负：恰好有一个解。

情况 $a = 0$。二次方程为 $x(x+1) = 0$，两根为 0 和 -1。只有 $x = 0$ 是非负的：恰好有一个解。

情况 $a<0$。若 $a<-\frac{1}{4}$，判别式为负，根本没有实根。若 $-\frac{1}{4}\le a<0$，两根为实数，两根之积 $-a$ 为正，所以两根同号，而两根之和为 -1，所以两根皆为负。无论哪一种情况都没有非负根，因此无解。

所以当 $a\ge 0$ 时方程有一个实数解，当 $a<0$ 时没有实数解，这正是所要证明的。

第 (i) 部分：立方根方程

每个实数 t 都恰好有一个实数立方根，而且 $x=t^{\frac{1}{3}} \iff x^3=t$，当中没有正负号的条件。所以原方程等价于

$$x^3-(1+a)x+a=0.$$

各项系数之和为 $1-(1+a)+a=0$，所以 $x=1$ 必定是一个根，把对应的因式除去后得

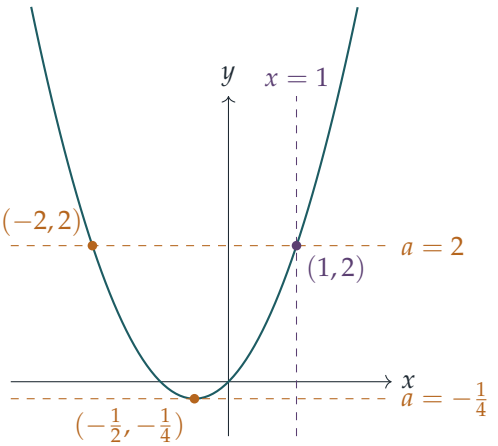
$$x^3-(1+a)x+a=(x-1)(x^2+x-a).$$

同一个二次式再次出现，而方程的解就是 $x=1$ 连同 $x^2+x-a=0$ 的实根，这次对正负号再没有限制。

> **想一想** 二次方程的根会不会与 $x=1$ 重合，或者两根彼此重合?

两种情况都会发生，而它们正是令解的个数减少的情况。当 $1+4a=0$，即 $a=-\frac{1}{4}$ 时，二次方程有重根（repeated root），该根为 $-\frac{1}{2}$。而当 $1+1-a=0$，即 $a=2$ 时，$x=1$ 是二次方程的根；这时 $x^2+x-2=(x-1)(x+2)$，所以两根为 1 和 -2。

把二次方程写成 $a=x^2+x$，它的实根就是高度为 a 的水平线与抛物线（parabola）$y=x^2+x$ 相交之处，而这条抛物线的最低点是 $(-\frac{1}{2},-\frac{1}{4})$。额外的解 $x=1$ 对应抛物线上的点 $(1,2)$。



综合各种情况，相异实数解的个数如下：

- $a < -\frac{1}{4}$：二次方程没有实根，所以只有一**个**解，即 $x = 1$；
- $a = -\frac{1}{4}$：解为 1 和 $-\frac{1}{2}$，所以有**两个**；
- $-\frac{1}{4} < a < 2$ 或 $a > 2$：二次方程有两个相异的根，两者都不等于 1，再加上 $x = 1$，所以有三**个**；
- $a = 2$：解为 1 和 -2，所以有**两个**。

第 (ii) 部分

与第 (i) 部分完全一样，平方根是非负的，所以

$$x = (b + x)^{\frac{1}{2}} \iff x^2 - x - b = 0 \text{ 且 } x \geq 0,$$

（反方向成立，是因为由 $x \geq 0$ 和 $b + x = x^2$ 可得 $(b + x)^{\frac{1}{2}} = |x| = x$）。现在 $x^2 - x - b = 0$ 的两根之和为 1，两根之积为 $-b$，而判别式为 $1 + 4b$。

情况 $b < -\frac{1}{4}$。没有实根，所以**没有**解。

情况 $b = -\frac{1}{4}$。二次方程为 $(x - \frac{1}{2})^2 = 0$，而 $x = \frac{1}{2} \geq 0$，所以有一**个**解。

情况 $-\frac{1}{4} < b < 0$。有两个相异实根，两根之积 $-b$ 为正，所以两根同号，而两根之和 1 为正，所以两根皆为正：有**两个**解。

情况 $b = 0$。二次方程为 $x(x - 1) = 0$，而 0 和 1 都是非负的：有**两个**解。（事实上 $0 = \sqrt{0}$ 而 $1 = \sqrt{1}$。）

情况 $b > 0$。有两个实根，两根之积为负，所以一正一负：有一**个**解。

总结：当 $b < -\frac{1}{4}$ 时无解；当 $b = -\frac{1}{4}$ 时有一个解；当 $-\frac{1}{4} < b \leq 0$ 时有两个解；当 $b > 0$ 时有一个解。

验算

- 从图像来看，第 (ii) 部分是问直线 $y = x$ 与半抛物线 $y = \sqrt{b + x}$ 在何处相交，而这条半抛物线由 $(-b, 0)$ 开始。当 $b = -\frac{1}{5}$ 时，解为 $\frac{1}{2} \pm \frac{1}{2\sqrt{5}}$，约为 0.276 和 0.724，两者都满足 $x = \sqrt{x - 0.2}$，准确至小数点后多个位。
- 在立方根方程中取 $a = 2$，$x = 1$ 和 $x = -2$ 都成立：$(3 \cdot (-2) - 2)^{\frac{1}{3}} = (-8)^{\frac{1}{3}} = -2$。
- 我们用电脑在细密的格点上计数了每个方程的根的个数，所取的 a 和 b 的值涵盖每个分界点（$-\frac{1}{4}$、0、2）的两侧，所得的个数与上面的分情况讨论一致。

常见错误

- 平方之后忘记了 $x \geq 0$：这样会得出第 (i) 部分对每个 $a > -\frac{1}{4}$ 都有两个解，这是错的。
- 对立方根方程也加上 $x \geq 0$ 的条件。负数的立方根是名正言顺的实数。
- 遗漏了 $a = 2$ 的情况：这时一次因式的根 $x = 1$ 同时也是二次式的根，所以三个解之中有两个合而为一。

STEP 之外

关键在于：在 $\mathbb{R}$ 上，平方不是可逆的运算，立方却是。$x \mapsto x^3$ 是 $\mathbb{R}$ 到自身的双射（bijection），而 $x \mapsto x^2$ 则不是，所以要先选定非负的分支（branch），才谈得上“那个”平方根。把形如 $x = f(x)$ 的方程看成不动点（fixed point）问题，并在参数变动时计数不动点的个数，就是分歧理论（bifurcation theory）的起点；在这套理论中，解会成对地诞生（如在 $a = -\frac{1}{4}$ 处），或者互换位置（如在 $a = 2$ 处）。若想细读有关函数、反函数（inverse）和连续性（continuity）的严谨论述，可参阅 Abbott 的 *Understanding Analysis*（实分析入门教材）。

详细解答

STEP I 2009，第 4 题

纯数学 · 余弦公式

STEP I 2009，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/09-s1-q4

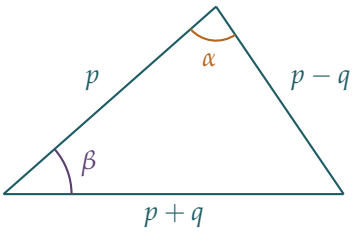
关键思路

在三角形中，最大的角对着最长的边，最小的角对着最短的边。因此，利用余弦定理（cosine rule）可以直接以 p 和 q 表示 $\cos\alpha$ 和 $\cos\beta$，而要证明的恒等式（identity）只是代数运算。至于第二部分，由 $\alpha = 2\beta$ 可用二倍角公式（double-angle formula）把 $\cos\alpha$ 以 $\cos\beta$ 表示，恒等式于是变成一条只含 $\cos\beta$ 的方程。

解答

恒等式

由于 $p - q < p < p + q$，最大的角 α 对着边 $p + q$，最小的角 β 对着边 $p - q$。



对 α 应用余弦定理，得

$$\cos\alpha = \frac{p^2 + (p-q)^2 - (p+q)^2}{2p(p-q)} = \frac{p^2 - 4pq}{2p(p-q)} = \frac{p-4q}{2(p-q)},$$

对 β 则得

$$\cos\beta = \frac{p^2+(p+q)^2-(p-q)^2}{2p(p+q)} = \frac{p^2+4pq}{2p(p+q)} = \frac{p+4q}{2(p+q)}.$$

这里我们约去了 p，由于 $p>0$，这样做是容许的。于是

$$1-\cos\alpha = \frac{2(p-q)-(p-4q)}{2(p-q)} = \frac{p+2q}{2(p-q)},$$
$$1-\cos\beta = \frac{2(p+q)-(p+4q)}{2(p+q)} = \frac{p-2q}{2(p+q)},$$

所以

$$4(1-\cos\alpha)(1-\cos\beta) = \frac{(p+2q)(p-2q)}{(p-q)(p+q)} = \frac{p^2-4q^2}{p^2-q^2}.$$

另一边则是

$$\begin{aligned}\cos\alpha+\cos\beta &= \frac{(p-4q)(p+q)+(p+4q)(p-q)}{2(p^2-q^2)}\\ &= \frac{(p^2-3pq-4q^2)+(p^2+3pq-4q^2)}{2(p^2-q^2)} = \frac{p^2-4q^2}{p^2-q^2}.\end{aligned}$$

两个表达式相等，所以 $4(1-\cos\alpha)(1-\cos\beta)=\cos\alpha+\cos\beta$，正是所要证明的结果。

（要令三角形存在，必须有 $(p-q)+p>p+q$，即 $p>2q$。这时恒等式两边都是正数，但上面的代数运算并不依赖这一点。）

$\alpha = 2\beta$ 的情况

设 $c=\cos\beta$。那么 $\cos\alpha=\cos 2\beta = 2c^2-1$，所以 $1-\cos\alpha = 2(1-c^2) = 2(1-c)(1+c)$。代入恒等式，得

$$8(1-c)^2(1+c) = 2c^2+c-1 = (2c-1)(c+1).$$

> ? **想一想** 我们可以把两边同时除以 $1+c$ 吗？

可以。β 是三角形的一只角，所以 $0<\beta<\pi$，而 $c=\cos\beta>-1$，即 $1+c\neq 0$。两边同时除以这个因子，得

$$8(1-c)^2 = 2c-1 \quad\Longrightarrow\quad 8c^2-18c+9=0 \quad\Longrightarrow\quad (4c-3)(2c-3)=0.$$

所以 $c=\frac{3}{4}$ 或 $c=\frac{3}{2}$。余弦值不可能大于 1，因此我们舍去 $\frac{3}{2}$。由此可知 $\cos\beta=\frac{3}{4}$，正是所要证明的结果。

三边之比

利用第一部分所得的 $\cos\beta$ 公式，

$$\frac{p+4q}{2(p+q)} = \frac{3}{4} \quad\Longrightarrow\quad 4p+16q = 6p+6q \quad\Longrightarrow\quad p = 5q.$$

这个结果满足 $p > 2q$，所以三角形确实存在。它的三边是 $p-q = 4q$、$p = 5q$ 和 $p+q = 6q$，因此三边之比是

$$4 : 5 : 6.$$

验算

在边长为 $4, 5, 6$ 的三角形中，由余弦定理得 $\cos\beta = \frac{25+36-16}{60} = \frac{3}{4}$ 和 $\cos\alpha = \frac{16+25-36}{40} = \frac{1}{8}$。另外 $2\cdot\left(\frac{3}{4}\right)^2 - 1 = \frac{1}{8}$，这就确认了 $\alpha = 2\beta$（两只角都在 $(0,\pi)$ 之内，而余弦函数在这个区间上是一对一（one-to-one）的）。以数值计算，$\beta \approx 0.7227$ 弧度（radians），$\alpha \approx 1.4455$ 弧度。在这个三角形中，恒等式变成 $4\cdot\frac{7}{8}\cdot\frac{1}{4} = \frac{7}{8} = \frac{1}{8}+\frac{3}{4}$。我们也用 10^5 个满足 $p > 2q$ 的随机三角形测试过这条恒等式，两边之差不超过 3×10^{-15}。

常见错误

- 把角与边配错。最大的角对着的是最长的边，即 $p+q$。
- 约去因子 $1+c$ 时没有说明它为何不等于零，或者保留了 $c = \frac{3}{2}$ 这个根。
- 答案只写 $p = 5q$，而没有给出三边之比 $4 : 5 : 6$。

STEP 之外

若不计缩放，边长为 $4, 5, 6$ 的三角形是唯一一个三边成等差数列（arithmetic progression）而最大角是最小角两倍的三角形。像“哪些三角形的边长是整数，而且各角成指定的比？”这类问题，很快就会引出丢番图方程（Diophantine equations），以及曲线上有理点（rational points）的参数化，情况与毕氏三元数（Pythagorean triples）十分相似。Davenport 的 *The Higher Arithmetic* 很适合用来了解这些想法如何发展，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多 STEP 风格的几何和三角学题目。

详细解答

STEP I 2009，第 5 题

纯数学 · 约束最优化

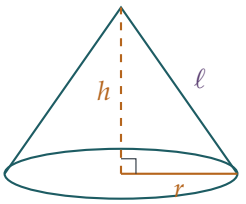
STEP I 2009，第 5 题

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/09-s1-q5

关键思路

这里有三个长度，但只有两个是自由的，因为 r、h 和 ℓ 由毕氏定理（Pythagoras' theorem）联系在一起：$\ell^2 = r^2 + h^2$。固定其中一个量，又会再减少一个自由度，所以在每一部分，我们都把要取驻值（stationary value）的量写成只含 r 的函数，然后求导。平方根会令代数运算变得繁复，因此改用 V^2 或 A^2 来计算会划算得多：它们取驻值的位置，与 V 或 A 取驻值的位置完全相同。



解答

在整个解答中，r、h 和 ℓ 均为正数，而且由横切面中的直角三角形可得

$$\ell^2 = r^2 + h^2.$$

由于 $V > 0$ 及 $A > 0$，V^2 的导数（derivative）是 $2V\dfrac{\mathrm{d}V}{\mathrm{d}r}$，它恰好在 $\dfrac{\mathrm{d}V}{\mathrm{d}r}$ 等于零时才等于零；A 的情况亦相同。因此我们可以改为求这些平方的驻点（stationary point）。

第 (i) 部分

当 A 固定时，$\ell = \dfrac{A}{\pi r}$，所以

$$h^2 = \ell^2 - r^2 = \frac{A^2}{\pi^2 r^2} - r^2.$$

（要得到一个真正的圆锥体，必须有 $h > 0$，即 $A > \pi r^2$。）于是

$$\begin{aligned} V^2 = \tfrac{1}{9}\pi^2 r^4 h^2 &= \tfrac{1}{9}\pi^2 r^4 \left(\frac{A^2}{\pi^2 r^2} - r^2\right) \\ &= \tfrac{1}{9}\left(A^2 r^2 - \pi^2 r^6\right). \end{aligned}$$

把 A 视为常数，对 r 求导，得

$$\frac{\mathrm{d}(V^2)}{\mathrm{d}r} = \tfrac{1}{9}\left(2A^2 r - 6\pi^2 r^5\right) = \tfrac{2}{9} r\left(A^2 - 3\pi^2 r^4\right).$$

由于 $r > 0$，上式恰好在以下情况等于零：

$$A^2 = 3\pi^2 r^4,$$

这就是第一个结果。取正平方根，得 $A = \sqrt{3}\,\pi r^2$，因此

$$\ell = \frac{A}{\pi r} = \frac{\sqrt{3}\,\pi r^2}{\pi r} = \sqrt{3}\, r,$$

正如题目所求。（留意 $A = \sqrt{3}\,\pi r^2 > \pi r^2$，所以这确实是一个圆锥体。）

当 r 递增并经过这个值时，V^2 的导数由正变负，所以这个驻值是极大值（maximum）：这就是在给定曲面面积（curved surface area）下体积最大的圆锥体。由 $\ell^2 = r^2 + h^2$ 还可得 $h^2 = 3r^2 - r^2 = 2r^2$，所以 $h = \sqrt{2}\, r$。

第 (ii) 部分

现在 V 是固定的，所以 $h = \dfrac{3V}{\pi r^2}$，而

$$\begin{aligned} A^2 = \pi^2 r^2 \ell^2 = \pi^2 r^2 (r^2 + h^2) &= \pi^2 r^4 + \pi^2 r^2 \cdot \frac{9V^2}{\pi^2 r^4} \\ &= \pi^2 r^4 + \frac{9V^2}{r^2}. \end{aligned}$$

把 V 视为常数，对 r 求导，得

$$\frac{\mathrm{d}(A^2)}{\mathrm{d}r} = 4\pi^2 r^3 - \frac{18V^2}{r^3},$$

它恰好在以下情况等于零：

$$V^2 = \tfrac{2}{9}\pi^2 r^6, \qquad \text{所以} \qquad V = \tfrac{\sqrt{2}}{3}\,\pi r^3$$

（因为 $V > 0$，故取正根）。

想一想 我们已求得以 r 表示的 V，但题目要求的是 h。怎样才能得到它?

再用一次体积公式：$V = \frac{1}{3}\pi r^2 h$。令它等于 $\frac{\sqrt{2}}{3}\pi r^3$，并约去 $\frac{1}{3}\pi r^2$，得

$$h = \sqrt{2}\, r.$$

这一次导数由负变正（当 r 很小时，$9V^2/r^2$ 这一项非常大；当 r 很大时，$\pi^2 r^4$ 这一项非常大），所以这是在给定体积下曲面面积最小的圆锥体。

想一想 两部分都得出相同的形状 $h = \sqrt{2}\, r$，这是巧合吗?

不是。如果一个圆锥体在其面积之下体积最大，那么它在其体积之下面积也必然最小：任何体积相同而面积较小的圆锥体，都可以按比例放大至原来的面积，这样它的体积就会更大，而这是不可能的。由于题目只把形状确定到相差一个缩放比例，两部分必定选出相同的比 $h : r$。

验算

- 在第 (i) 部分取 $r = 1$：则 $A = \sqrt{3}\,\pi$，$\ell = \sqrt{3}$，$h = \sqrt{2}$，而 $\pi r \ell = \sqrt{3}\,\pi$，与应有的结果一致。把 A 保持在 $\sqrt{3}\,\pi$，将 r 稍为调整至 0.99 或 1.01，所得的体积都略小一些。
- 对随机选取的固定 A 以数值方法求 V 的最大值，以及对随机选取的固定 V 求 A 的最小值，每次都得出 $h/r \approx 1.41421$，即 $\sqrt{2}$。
- 量纲（dimensions）：A^2 和 $\pi^2 r^4$ 都是（长度）^{4}；V^2 和 $\pi^2 r^6$ 都是（长度）^{6}。

常见错误

- 忘记了 ℓ 取决于 r 和 h：在第 (i) 部分把 ℓ 当作常数，或在第 (ii) 部分把 h 当作常数，而事实上只有 A（或 V）是固定的。
- 保留着平方根直接对 V 求导；这样做是可行的，但很容易出错。先平方就完全不必用上链式法则（chain rule）。

STEP 之外

在固定某个量的情况下求另一个量的极值，是一个*约束优化*（constrained optimisation）问题。在大学里你会学到拉格朗日乘数法（Lagrange multipliers），它不必消去变量就能处理约束条件，并且解释了本题两部分之间的“对偶性”（duality）。若想找一本清晰的入

门书来学习多元微积分（multivariable calculus）和拉格朗日乘数法，可以看 Stewart 的 *Calculus*；Siklos 的 *Advanced Problems in Mathematics* 则有更多 STEP 风格的优化练习。

详细解答

STEP I 2009，第 6 题

纯数学 · 代换积分法

> **STEP I 2009，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/09-s1-q6

关键思路

积分上下限 $1/m$ 与 m 在 $x \mapsto 1/x$ 之下互相对调，所以最自然的做法就是试用这个代换（substitution）：它把区间映射到自身，并把幂 x^{-n} 换成幂 u^{n-1}。有了第 (ii) 部分的结果，我们便可以把第 (iii) 部分中难以处理的各项，化为第 (i) 部分那个容易处理的积分。在 (iii)(b) 中，区间不再对称，但同一个代换会把 $[1,2]$ 送到 $[\frac{1}{2},1]$，于是两半又可以重新拼合起来。

解答

第 (i) 部分

以 $x+1$ 除 x^2，得

$$\frac{x^2}{x+1} = x - 1 + \frac{1}{x+1},$$

由于在该区间上 $x+1>0$（因为 $m>0$），所以

$$\int_{1/m}^{m} \frac{x^2}{x+1}\,\mathrm{d}x = \left[\tfrac{1}{2}x^2 - x + \ln(x+1)\right]_{1/m}^{m}.$$

对数部分给出

$$\ln(m+1) - \ln\left(\frac{1}{m}+1\right) = \ln\frac{m+1}{(m+1)/m} = \ln m.$$

多项式部分给出

$$\begin{aligned}\tfrac{1}{2}\Big(m^2-\frac{1}{m^2}\Big)-\Big(m-\frac{1}{m}\Big) &= \Big(m-\frac{1}{m}\Big)\Big[\tfrac{1}{2}\Big(m+\frac{1}{m}\Big)-1\Big]\\ &= \frac{m^2-1}{m}\cdot\frac{m^2-2m+1}{2m}\\ &= \frac{(m-1)(m+1)(m-1)^2}{2m^2}=\frac{(m-1)^3(m+1)}{2m^2}.\end{aligned}$$

把两部分相加，即得所求结果。

第 (ii) 部分

设 $x=1/u$，则 $\mathrm{d}x=-\mathrm{d}u/u^2$。当 $x=1/m$ 时，$u=m$；当 $x=m$ 时，$u=1/m$。此外

$$x^n(x+1)=\frac{1}{u^n}\cdot\frac{1+u}{u}=\frac{u+1}{u^{n+1}}.$$

因此

$$\begin{aligned}\int_{1/m}^{m}\frac{1}{x^n(x+1)}\,\mathrm{d}x &= \int_{m}^{1/m}\frac{u^{n+1}}{u+1}\cdot\Big(-\frac{1}{u^2}\Big)\mathrm{d}u\\ &= \int_{1/m}^{m}\frac{u^{n-1}}{u+1}\,\mathrm{d}u,\end{aligned}$$

其中负号用来把上下限对调回来。这就是所求结果。

第 (iii) 部分

(a) 把被积函数（integrand）拆开：

$$\frac{x^5+3}{x^3(x+1)}=\frac{x^2}{x+1}+\frac{3}{x^3(x+1)}.$$

积分区间是 $[\frac{1}{2},2]$，即第 (ii) 部分的区间取 $m=2$。在第 (ii) 部分中取 $n=3$，得

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

这是第一项的三倍。由第 (i) 部分取 $m=2$，得

$$\int_{1/2}^{2}\frac{x^2}{x+1}\,\mathrm{d}x=\frac{1^3\cdot 3}{2\cdot 4}+\ln 2=\frac{3}{8}+\ln 2.$$

所以积分等于 $4\big(\frac{3}{8}+\ln 2\big)$，即

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

(b) 现在

$$\frac{x^5+x^3+1}{x^3(x+1)}=\frac{x^2}{x+1}+\frac{1}{x+1}+\frac{1}{x^3(x+1)}.$$

想一想 现在的区间是 $[1,2]$，并非 $[1/m, m]$ 的形式。第 (ii) 部分还能派上用场吗?

可以：代换 $x = 1/u$ 在任何由正数组成的区间上都适用，只是上下限会随之移动。计算与第 (ii) 部分完全相同，其中 $x = 1$ 对应 $u = 1$，$x = 2$ 对应 $u = \frac{1}{2}$，由此可得

$$\int_1^2 \frac{1}{x^3(x+1)}\,\mathrm{d}x = \int_{1/2}^1 \frac{u^2}{u+1}\,\mathrm{d}u.$$

它的被积函数与第一项相同，而区间恰好相邻，所以两者合起来就是第 (i) 部分取 $m = 2$ 时的积分：

$$\int_1^2 \frac{x^2}{x+1}\,\mathrm{d}x + \int_{1/2}^1 \frac{x^2}{x+1}\,\mathrm{d}x = \int_{1/2}^2 \frac{x^2}{x+1}\,\mathrm{d}x = \frac{3}{8} + \ln 2.$$

余下的一项是

$$\int_1^2 \frac{1}{x+1}\,\mathrm{d}x = \ln 3 - \ln 2.$$

相加得

$$\int_1^2 \frac{x^5+x^3+1}{x^3(x+1)}\,\mathrm{d}x = \frac{3}{8} + \ln 2 + \ln 3 - \ln 2 = \frac{3}{8} + \ln 3.$$

验算

- 在第 (i) 部分中，$m = 1$ 时积分区间只有一点，而右边确实是 $0 + \ln 1 = 0$。把 m 换成 $1/m$ 应该令积分变号（上下限对调），而公式的确会变号。
- 用辛普森法则（Simpson's rule）计算，(iii)(a) 得 4.272589，与 $\frac{3}{2} + 4\ln 2 \approx 4.272589$ 吻合；(iii)(b) 得 1.473612，与 $\frac{3}{8} + \ln 3 \approx 1.473612$ 吻合。
- (b) 的合理性检查：在 $[1,2]$ 上，被积函数由 1.5 下降至约 1.37，然后上升至约 1.71，所以答案接近 1.47 是合理的。

常见错误

- 忘记 $\mathrm{d}x = -\mathrm{d}u/u^2$ 中的负号，或忘记对调上下限；两者都必须做，而它们会互相抵消。
- 在 (iii)(b) 中，把第 (ii) 部分当作区间是 $[\frac{1}{2}, 2]$ 来套用。代换把 $[1,2]$ 移到 $[\frac{1}{2}, 1]$，而这正是各项能够拼合起来的原因。

STEP 之外

映射 $x \mapsto 1/x$ 是区间 $[1/m, m]$ 的一个对称（symmetry），而利用定义域的对称性来找出积分之间的关系，是分析学（analysis）中反复出现的技巧。例如，$\int_0^\infty \frac{\ln x}{1+x^2}\,\mathrm{d}x = 0$ 就是这样证明的。在复分析（complex analysis）中，同样的反演（inversion）$z \mapsto 1/z$ 会把单位圆的内部与外部对调，并成为一件基本工具。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）对反演有十分优美的几何阐述。

详细解答

STEP I 2009，第 7 题

纯数学 · 分部积分法

STEP I 2009，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/09-s1-q7

关键思路

开首的积分用的是经典的“分部积分两次，原来的积分又回来了”这一招：做完两步之后，原来的积分会在等式右边重新出现，于是我们把它当作方程中的未知数来求解。一旦对每一个整数 m 都得到 $\int_0^{2\pi} \mathrm{e}^x \cos mx \,\mathrm{d}x$ 的公式，其后各部分就只是按部就班的整理工作：把正弦与余弦的乘积化为若干个单一余弦之和，而每一项都可以立即积分。

解答

一般的积分

记 $I_m = \int_0^{2\pi} \mathrm{e}^x \cos mx \,\mathrm{d}x$，并为简便起见设 $E = \mathrm{e}^{2\pi} - 1$。取 $u = \cos mx$ 及 $\mathrm{d}v = \mathrm{e}^x \,\mathrm{d}x$ 作分部积分（integration by parts），

$$I_m = \Big[\mathrm{e}^x \cos mx\Big]_0^{2\pi} + m\int_0^{2\pi} \mathrm{e}^x \sin mx \,\mathrm{d}x.$$

由于 m 是整数，$\cos 2m\pi = 1$，所以方括号一项等于 $\mathrm{e}^{2\pi} - 1 = E$。再作一次分部积分，这次取 $u = \sin mx$，

$$\int_0^{2\pi} \mathrm{e}^x \sin mx \,\mathrm{d}x = \Big[\mathrm{e}^x \sin mx\Big]_0^{2\pi} - m\int_0^{2\pi} \mathrm{e}^x \cos mx \,\mathrm{d}x = 0 - mI_m,$$

这是因为 $\sin 2m\pi = \sin 0 = 0$，当中再次用到 m 是整数。把两式合起来，

$$I_m = E - m^2 I_m, \qquad \text{所以} \qquad (m^2+1)I_m = E.$$

? **想一想** 我们兜了一个圈，又得回 I_m。这是否一条死路？

不是：这正是我们想要的。等式 $I_m = E - m^2 I_m$ 是关于未知数 I_m 的一次方程（linear equation），而由于 $m^2 + 1 \neq 0$，我们可以把它除过去：

$$\int_0^{2\pi} \mathrm{e}^x \cos mx \,\mathrm{d}x = \frac{1}{m^2+1}\left(\mathrm{e}^{2\pi} - 1\right),$$

正如所求。（$m = 0$ 的情况亦包括在内：它只是说 $\int_0^{2\pi} \mathrm{e}^x \,\mathrm{d}x = E$。）留意整数这个条件的作用，是令边界项（boundary terms）变得简单；若 m 不是整数，这条公式并不成立。

第 (i) 部分

用复角公式（compound-angle formulae）展开，$\sin A \sin B$ 的项互相抵消：

$$\cos(A+B) + \cos(A-B) = 2\cos A \cos B.$$

取 $A = 6x$ 及 $B = x$，得 $\cos x \cos 6x = \frac{1}{2}(\cos 7x + \cos 5x)$。因此，在一般结果中取 $m = 7$ 及 $m = 5$，

$$\begin{aligned}\int_0^{2\pi} \mathrm{e}^x \cos x \cos 6x \,\mathrm{d}x &= \tfrac{1}{2}\left(I_7 + I_5\right) = \tfrac{1}{2}\left(\frac{1}{50} + \frac{1}{26}\right)E \\ &= \tfrac{1}{2} \cdot \frac{26+50}{1300}\,E = \frac{76}{2600}\,E = \frac{19}{650}\left(\mathrm{e}^{2\pi} - 1\right),\end{aligned}$$

这就是所求的结果。

第 (ii) 部分

这里需要做两次积化和差(product-to-sum)。把两条复角公式相减而不是相加,得 $\cos(A-B) - \cos(A+B) = 2\sin A \sin B$，所以取 $A = 4x$、$B = 2x$，

$$\sin 2x \sin 4x = \tfrac{1}{2}(\cos 2x - \cos 6x).$$

乘以 $\cos x$，并对每个乘积使用第 (i) 部分的恒等式，

$$\begin{aligned}\sin 2x \sin 4x \cos x &= \tfrac{1}{2}\cos 2x \cos x - \tfrac{1}{2}\cos 6x \cos x \\ &= \tfrac{1}{4}(\cos 3x + \cos x) - \tfrac{1}{4}(\cos 7x + \cos 5x).\end{aligned}$$

现在每一项都可用一般结果处理，分别取 $m = 3, 1, 7, 5$：

$$\begin{aligned}\int_0^{2\pi} \mathrm{e}^x \sin 2x \sin 4x \cos x \,\mathrm{d}x &= \tfrac{1}{4}\left(I_3 + I_1 - I_7 - I_5\right) \\ &= \tfrac{1}{4}\left(\frac{1}{10} + \frac{1}{2} - \frac{1}{50} - \frac{1}{26}\right)E \\ &= \tfrac{1}{4} \cdot \frac{65 + 325 - 13 - 25}{650}\,E = \tfrac{1}{4} \cdot \frac{352}{650}\,E.\end{aligned}$$

所以

$$\int_0^{2\pi} \mathrm{e}^x \sin 2x \sin 4x \cos x \,\mathrm{d}x = \frac{44}{325}\left(\mathrm{e}^{2\pi} - 1\right).$$

验算

- 对 $\dfrac{\mathrm{e}^x(\cos mx + m\sin mx)}{m^2+1}$ 求导，得 $\mathrm{e}^x\cos mx$，所以它是一个原函数（antiderivative）；在 2π 和 0 两处代入求值，便重新得到一般结果。
- 在第 (ii) 部分的恒等式中代入 $x=0$：左边是 0，右边是 $\frac{1}{4}(1+1-1-1)=0$。吻合。
- 以数值方法（辛普森法则，Simpson's rule）计算，第 (i) 部分的积分是 15.6236…，第 (ii) 部分的积分是 72.3619…，与 $\frac{19}{650}E$ 及 $\frac{44}{325}E$ 相符，其中 $E=\mathrm{e}^{2\pi}-1\approx 534.49$。

常见错误

- 第二次分部积分时弄错正负号：$\sin mx$ 的导数是 $+m\cos mx$，连同分部积分公式中的负号，得出的是 $-m^2 I_m$。
- 在每次积化和差时漏掉因子 $\frac{1}{2}$，以致第 (ii) 部分得到 $\frac{1}{2}$ 而不是 $\frac{1}{4}$。

STEP 之外

更利落的做法是写 $\cos mx = \operatorname{Re}\mathrm{e}^{\mathrm{i}mx}$，于是 $I_m = \operatorname{Re}\int_0^{2\pi}\mathrm{e}^{(1+\mathrm{i}m)x}\,\mathrm{d}x = \operatorname{Re}\dfrac{E}{1+\mathrm{i}m} = \dfrac{E}{1+m^2}$，完全不需要分部积分。像我们在后面各部分所做的那样，把一个函数拆成 $\cos mx$ 这些分量，正是傅里叶级数（Fourier series）的起点；傅里叶级数是大学一年级分析和微分方程（differential equations）中的核心工具。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）有一段易读的介绍，而 Needham 的 *Visual Complex Analysis*（以图像阐释复分析的名著）则说明了为何从指数函数的角度看问题会如此奏效。

详细解答

STEP I 2009，第 8 题

纯数学 · 圆与切线

STEP I 2009，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/09-s1-q8

关键思路

一个圆与一条直线相切（touch），当且仅当圆心到该直线的距离等于半径。各个圆 C 的圆心都位于一条通过原点的直线上，而这条直线平分（bisect）x 轴与 $3y = 4x$ 之间的夹角，所以每一个这样的圆都与这两条直线相切。因此 C 就是嵌在这个角内的整个圆族，而第 (ii) 部分的内切圆（incircle）必定是其中之一：我们只需选出同时令圆与第三条边相切的 t 值。

解答

第 (i) 部分

圆 C 的圆心为 $P = (2t, t)$，半径为 t。由于 $t > 0$，圆心位于 x 轴上方，与它的距离为 t，恰好等于半径。所以 C 与 $y = 0$ 相切，切点为 $(2t, 0)$。

直线 OP 的斜率（gradient）为 $\frac{t}{2t} = \frac{1}{2}$，所以 $\tan\alpha = \frac{1}{2}$，并且

$$\tan 2\alpha = \frac{2\tan\alpha}{1 - \tan^2\alpha} = \frac{1}{1 - \frac{1}{4}} = \frac{4}{3},$$

这正是所要证明的。

现在设 β 为 x 轴与直线 $3y = 4x$ 之间的锐角，该直线的斜率为 $\frac{4}{3}$。于是 $\tan\beta = \frac{4}{3} = \tan 2\alpha$。由于 $0 < \alpha < \frac{\pi}{4}$（因为 $\tan\alpha = \frac{1}{2} < 1$），角 2α 也是锐角，而 tan 在锐角范围内是一对一（one-to-one）的，所以 $\beta = 2\alpha$。

? **想一想** 为甚么 $\beta = 2\alpha$ 可以告诉我们 C 与 $3y = 4x$ 相切?

直线 OP 与 x 轴的夹角为 α，与直线 $3y = 4x$ 的夹角为 $\beta - \alpha = \alpha$，所以 OP 平分这两条直线之间的夹角。由 P 向两条直线分别作垂线，得到两个直角三角形，它们有相同的斜边（hypotenuse）OP，而且在 O 处的角同为 α，所以两个垂直距离（perpendicular distance）相等，都是 $OP\sin\alpha$。到 x 轴的距离为 t，所以由 P 到 $3y = 4x$ 的距离也是 t，即半径。因此 C 与 $3y = 4x$ 相切。

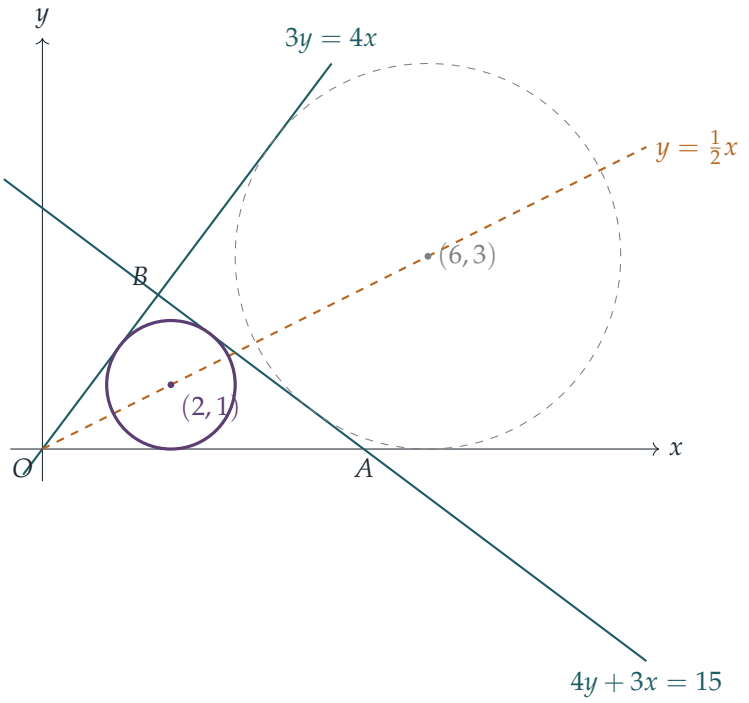
（作为覆核，由点到直线的垂直距离公式，可得由 $(2t, t)$ 到 $4x - 3y = 0$ 的距离为 $\frac{|8t-3t|}{5} = t$。）

第 (ii) 部分

首先找出这个三角形。直线 $y = 0$ 与 $3y = 4x$ 相交于 $O = (0,0)$；$y = 0$ 与 $4y + 3x = 15$ 相交于 $A = (5,0)$。至于第三个顶点，把 $y = \frac{4}{3}x$ 代入 $4y + 3x = 15$：

$$\tfrac{16}{3}x + 3x = 15 \quad \Longrightarrow \quad x = \tfrac{9}{5}, \qquad y = \tfrac{12}{5},$$

所以 $B = \left(\frac{9}{5}, \frac{12}{5}\right)$。$3y = 4x$ 与 $4y + 3x = 15$ 的斜率分别为 $\frac{4}{3}$ 和 $-\frac{3}{4}$，两者之积为 -1，所以三角形在 B 处有一个直角。它的三边为 $OA = 5$、$OB = 3$ 和 $AB = 4$。



内切圆位于三角形之内，所以在 O 附近，它嵌在正 x 轴与直线 $3y = 4x$ 之间的锐角内，并与这两条直线相切。它的圆心与这两条直线等距，所以圆心位于该角的平分线上；根

据第 (i) 部分，这条平分线就是直线 $y = \frac{1}{2}x$（即直线 OP）。若圆心为 $(2t, t)$，其中 $t > 0$，则它与 $y = 0$ 的距离为 t，而这必定就是半径。所以内切圆是第 (i) 部分的圆 C 其中之一。

余下只需令 C 与第三条边 $3x + 4y - 15 = 0$ 相切。由 $(2t, t)$ 到这条直线的距离必须等于 t：

$$\frac{|6t + 4t - 15|}{5} = t \quad \Longleftrightarrow \quad |10t - 15| = 5t.$$

要么 $10t - 15 = 5t$，得 $t = 3$；要么 $15 - 10t = 5t$，得 $t = 1$。

这两个圆都与三条直线相切，但只有一个位于三角形之内。当 $t = 3$ 时，圆心 $(6, 3)$ 满足 $3x + 4y = 30 > 15$，所以它位于 AB 的另一侧，与 O（满足 $3x + 4y = 0 < 15$）分处两边：这是一个*旁切圆*（excircle），即图中的虚线圆。当 $t = 1$ 时，圆心 $(2, 1)$ 满足 $3x + 4y = 10 < 15$，而且它在 $y = 0$ 之上、$3y = 4x$ 之下，所以它位于三角形之内。一个圆若圆心在三角形之内，并与三条边都相切，则整个圆位于三角形之内，所以内切圆的圆心为 $(2, 1)$，半径为 1：

$$(x - 2)^2 + (y - 1)^2 = 1.$$

验算

对于两条直角边为 a, b、斜边为 c 的直角三角形，内切圆半径（inradius）为 $\frac{1}{2}(a + b - c)$；这里 $\frac{1}{2}(3 + 4 - 5) = 1$，与所得结果一致。内切圆半径亦等于面积 / 半周长 $= 6/6 = 1$。在数值上，把内心（incentre）计算为各顶点以边长加权的平均值 $\frac{4O + 3A + 5B}{12}$，结果恰好是 $(2, 1)$，而由 $(2, 1)$ 到三条直线的距离全都是 1。

常见错误

- 解 $|10t - 15| = 5t$ 时忽略了绝对值，只求得 $t = 3$，而这是旁切圆，不是内切圆。
- 求得两个根，却没有解释为甚么 $t = 1$ 才是位于三角形之内的那一个。
- 在第 (i) 部分，只说 $\tan\beta = \tan 2\alpha$ 所以 $\beta = 2\alpha$，却没有指出两个角都是锐角。

STEP 之外

每个三角形都有一个内切圆和三个旁切圆，而这四个圆心正是内角平分线与外角平分线的交点；这里求得的 $t = 3$ 就是与 O 相对的旁切圆。这些圆引领我们进入古典几何中一个蕴藏丰富的宝库，包括九点圆（nine-point circle），以及指出九点圆与这四个圆都相切的费尔巴哈定理（Feuerbach's theorem）。考克斯特（Coxeter）的 *Introduction to Geometry*（几何学入门）是探索这个课题的经典之作；Siklos 的 *Advanced Problems in Mathematics*（高等数学难题）则提供更多 STEP 风格的练习。

详细解答

STEP I 2009，第 9 题

力学 · 相撞的抛体

STEP I 2009，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：

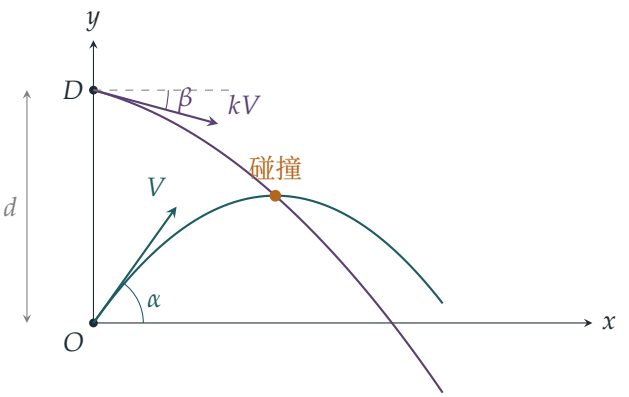
step.maths.org/questions/09-s1-q9

关键思路

两个只受重力作用的质点（particle）具有相同的加速度，所以比较两者位置时，$-\frac{1}{2}gt^2$ 项会互相抵消：相对于对方而言，它们是以恒定速度沿直线运动。碰撞（collision）的意思是在同一时刻 T，两者的水平位移相等，垂直位移也相等。水平方向的方程立即给出余弦关系式。垂直方向的方程含有 β，我们利用 $\sin^2\beta = 1 - \cos^2\beta$ 和余弦关系式把它消去。

解答

我们把 P 和 Q 当作在重力下自由运动的质点，不计空气阻力（air resistance）。要令两者相撞，它们必须在同一铅垂平面内、朝同一水平方向抛出。取 O 为原点，x 轴沿该水平方向，y 轴垂直向上，于是 D 的坐标为 $(0, d)$。



在时刻 t，两质点的位置为

$$\begin{aligned} P:&\quad x = Vt\cos\alpha, && y = Vt\sin\alpha - \tfrac{1}{2}gt^2,\\ Q:&\quad x = kVt\cos\beta, && y = d - kVt\sin\beta - \tfrac{1}{2}gt^2, \end{aligned}$$

其中 Q 的垂直速度带负号，是因为 β 是从水平线向下量度的。

余弦关系式与 T 的方程

碰撞时，$t = T > 0$，而且两者的坐标相同。水平方向上，

$$VT\cos\alpha = kVT\cos\beta,$$

两边除以 $VT \neq 0$，便得 $\cos\alpha = k\cos\beta$，正是所求。

垂直方向上，两边的 $-\frac{1}{2}gT^2$ 项互相抵消，余下

$$VT\sin\alpha = d - kVT\sin\beta, \qquad \text{所以} \qquad kVT\sin\beta = d - VT\sin\alpha.$$

想一想 怎样消去 β?

我们已知 $k\cos\beta$ 可用 α 表示，所以把方程平方，引入 $k^2\sin^2\beta$，而

$$k^2\sin^2\beta = k^2 - k^2\cos^2\beta = k^2 - \cos^2\alpha.$$

把垂直方向的方程平方再代入，

$$\begin{aligned} (k^2 - \cos^2\alpha)V^2T^2 &= d^2 - 2dVT\sin\alpha + V^2T^2\sin^2\alpha,\\ (k^2 - \cos^2\alpha - \sin^2\alpha)V^2T^2 + 2dVT\sin\alpha - d^2 &= 0,\\ (k^2 - 1)V^2T^2 + 2dVT\sin\alpha - d^2 &= 0, \end{aligned}$$

当中用了 $\sin^2\alpha + \cos^2\alpha = 1$。这就是所求的方程。

把它看成 T 的二次方程（quadratic），由于 $k > 1$，两根之积为 $-d^2/\big((k^2-1)V^2\big) < 0$。所以恰好有一个正根，而这个根就是碰撞的时刻。

在 P 的最高点碰撞

P 的垂直速度为 $V\sin\alpha - gt$，当 $t = V\sin\alpha/g$ 时为零。所以这里 $T = V\sin\alpha/g$。代入二次方程，

$$(k^2-1)\frac{V^4\sin^2\alpha}{g^2} + \frac{2dV^2\sin^2\alpha}{g} - d^2 = 0.$$

乘以 g^2 并合并含 $\sin^2\alpha$ 的项，得 $\sin^2\alpha\,\big((k^2-1)V^4 + 2gdV^2\big) = g^2d^2$，所以

$$\boxed{\sin^2\alpha = \frac{g^2d^2}{V^2\big((k^2-1)V^2 + 2gd\big)}}$$

由于 $k > 1$，括号内的式子为正，所以这个表达式有意义。

不等式

由于 $\sin^2\alpha \le 1$ 而分母为正，

$$\begin{aligned} g^2d^2 &\le (k^2-1)V^4 + 2gdV^2, \\ g^2d^2 - 2gdV^2 + V^4 &\le k^2V^4, \\ (gd - V^2)^2 &\le (kV^2)^2. \end{aligned}$$

$|gd - V^2|$ 和 kV^2 都是非负的，所以取平方根后不等式（inequality）的方向不变：$|gd - V^2| \le kV^2$。特别地，$gd - V^2 \le kV^2$，即

$$gd \le (1+k)V^2,$$

正是所求。等号成立需要 $\sin^2\alpha = 1$，即 P 垂直向上抛出。

验算

- 取 $g = 10$，$V = 10$，$k = 2$，$d = 10$。公式给出 $\sin^2\alpha = \frac{100\cdot 100}{100(300+200)} = \frac{1}{5}$，所以 $T = V\sin\alpha/g = 1/\sqrt{5}$。于是二次方程成为 $3\cdot 100\cdot\frac{1}{5} + 2\cdot 10\cdot 10\cdot\frac{1}{5} - 100 = 60 + 40 - 100 = 0$。取 $\cos\beta = \cos\alpha/2$，两个质点在时刻 T 的高度都是 1。这里 $gd = 100 \le 300 = (1+k)V^2$。
- 我们作了一次数值检验：随机选取数千组 V、k、α、T，逐步推算两个质点的运动，结果证实碰撞、余弦关系式、二次方程和 $\sin^2\alpha$ 的公式全部成立，误差只在舍入误差的范围之内。
- 极限情况：若 $\alpha \to \frac{1}{2}\pi$，则 $\cos\beta \to 0$，所以两者都沿垂直方向运动。P 在时间 V/g 内上升至高度 $V^2/(2g)$，而 Q 下降了 $kV^2/g + V^2/(2g)$，所以 $d = (1+k)V^2/g$：等号成立，与预测一致。

常见错误

- 弄错 Q 的垂直速度的正负号：它是从水平线向下抛出的，所以垂直速度为 $-kV\sin\beta$。
- 保留 $\frac{1}{2}gT^2$ 项：这些项会互相抵消，T 的方程里并没有 g。
- 对 $(gd - V^2)^2 \le k^2V^4$ 取平方根时，直接写成 $gd - V^2 \le kV^2$ 而没有提及绝对值（modulus）。结论是对的，但理由是 $x \le |x|$。

STEP 之外

重力互相抵消这一点，让我们初次体会到转换参考系（frame of reference）的威力。在一个以加速度 g 自由下落的参考系中，两个质点都以恒定速率沿直线运动，碰撞问题于是化为相对速度（relative velocity）的几何问题。这正是爱因斯坦（Einstein）等效原理（equivalence principle）背后的想法，许多抛体拦截问题也因此一句话就能解决。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）对相对运动和非惯性参考系（non-inertial frame）有细致的阐述。

详细解答

STEP I 2009，第 10 题

力学 · 斜面上的连接质点

STEP I 2009，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：

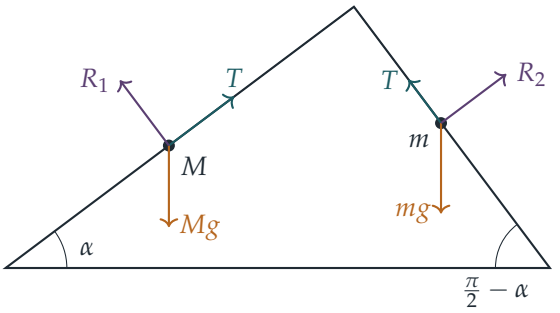
step.maths.org/questions/09-s1-q10

关键思路

由于绳子不可伸长（inextensible），滑轮又是光滑的，两个质点其实是作为一个系统一起运动：当 M 沿其斜面向下移动距离 s 时，m 便沿其斜面向上移动同样的距离 s，而且两边的张力（tension）相同。因此我们比较重力沿两个斜面的分量，再对每个质点应用牛顿第二定律（Newton's second law），便可求出共同的加速度。最后一部分则是关于单一变量 α 的普通微积分求极大值问题。

解答

我们把两个质点看作光滑斜面上的点质量，以一条轻而不可伸长的绳子绕过光滑滑轮相连，并假设系统由静止开始运动。设 T 为张力，R_1、R_2 分别为作用于 M 和 m 的法向反作用力（normal reaction）。



顶角为 $\pi-\alpha-(\frac{\pi}{2}-\alpha)=\frac{\pi}{2}$，所以两个斜面互相垂直。重力沿倾角为 α 的斜面向下的分量是 $Mg\sin\alpha$，而沿倾角为 $\frac{\pi}{2}-\alpha$ 的斜面向下的分量是 $mg\sin(\frac{\pi}{2}-\alpha)=mg\cos\alpha$。法向

反作用力垂直于斜面，所以对沿斜面的运动没有影响。

第 (i) 部分

取“M 沿其斜面向下、m 沿其斜面向上”为正方向，加速度为 a。沿各自的斜面分解力：

$$\begin{aligned} M: &\quad Mg\sin\alpha - T = Ma, \\ m: &\quad T - mg\cos\alpha = ma. \end{aligned}$$

两式相加消去 T：

$$a = \frac{g\left(M\sin\alpha - m\cos\alpha\right)}{M+m}.$$

由于当 $0 < \alpha < \frac{\pi}{2}$ 时 $\cos\alpha > 0$，

$$a > 0 \iff M\sin\alpha > m\cos\alpha \iff \tan\alpha > \frac{m}{M}.$$

所以若 $\tan\alpha > \frac{m}{M}$，由静止开始的系统会朝正方向加速：M 沿其斜面下滑，正如题目所求。

? **想一想** 我们怎样知道绳子一直保持拉紧，从而确定这些方程是正确的?

由第二条方程，$T = m(a + g\cos\alpha)$，当 $a > 0$ 时这是正数，所以绳子拉紧与方程并无矛盾。此外，绳子也不可能松弛：假如松弛了，每个质点都会沿自己的斜面向下滑，两者同时远离滑轮，这就需要一条更长的绳子。

第 (ii) 部分

现在 $\tan\alpha = \frac{2m}{M}$，所以 $m = \frac{1}{2}M\tan\alpha$，而且 $\tan\alpha > \frac{m}{M}$，故由第 (i) 部分可知，运动方向仍然是 M 向下。代入 $m = \frac{1}{2}M\tan\alpha$ 并约去 M：

$$\begin{aligned} a &= \frac{g\left(M\sin\alpha - \frac{1}{2}M\tan\alpha\cos\alpha\right)}{M + \frac{1}{2}M\tan\alpha} = \frac{g\left(\sin\alpha - \frac{1}{2}\sin\alpha\right)}{1 + \frac{1}{2}\tan\alpha} = \frac{\frac{1}{2}g\sin\alpha}{\frac{1}{2}(2+\tan\alpha)} \\ &= \frac{g\sin\alpha}{\tan\alpha + 2}, \end{aligned}$$

其中用到 $\tan\alpha\cos\alpha = \sin\alpha$。这就是所求的加速度大小。

求极大值。比值 m/M 通过 $\tan\alpha = 2m/M$ 决定了 α，所以我们要求以下函数的极大值：

$$f(\alpha) = \frac{\sin\alpha}{\tan\alpha + 2}, \qquad 0 < \alpha < \tfrac{\pi}{2}.$$

利用商法则（quotient rule），并由 $\frac{\mathrm{d}}{\mathrm{d}\alpha}\tan\alpha = \sec^2\alpha$，得

$$f'(\alpha) = \frac{\cos\alpha\left(\tan\alpha + 2\right) - \sin\alpha\sec^2\alpha}{(\tan\alpha+2)^2} = \frac{\sin\alpha + 2\cos\alpha - \dfrac{\sin\alpha}{\cos^2\alpha}}{(\tan\alpha+2)^2}.$$

把分子乘以 $\cos^2\alpha > 0$，并利用 $1-\cos^2\alpha = \sin^2\alpha$：

$$\cos^2\alpha\left(\sin\alpha + 2\cos\alpha\right) - \sin\alpha = 2\cos^3\alpha - \sin\alpha\left(1-\cos^2\alpha\right) = 2\cos^3\alpha - \sin^3\alpha.$$

所以 $f'(\alpha)$ 的正负号与 $2\cos^3\alpha - \sin^3\alpha = \cos^3\alpha\left(2-\tan^3\alpha\right)$ 相同。当 $\tan\alpha$ 由 0 增至 ∞ 时，这个式子在 $\tan^3\alpha < 2$ 时为正，一旦 $\tan^3\alpha > 2$ 便为负。因此 f 先递增后递减，其极大值出现于唯一的驻点（stationary point）

$$\tan^3\alpha = 2.$$

（你也可以这样看出这是极大值：在整个区间上 $f > 0$，而在两端 $f \to 0$。）配合 $\tan\alpha = \frac{2m}{M}$，这即是说

$$\frac{8m^3}{M^3} = 2, \qquad \text{即} \qquad 4m^3 = M^3,$$

正如题目所求。

验算

- 第 (i) 部分的极限情况：当 $m = 0$ 时得 $a = g\sin\alpha$，即光滑斜面上只有一个质点的情形；而当 $\tan\alpha = m/M$ 时得 $a = 0$，即平衡点。
- 在最优情况下，$\tan\alpha = 2^{1/3} \approx 1.26$，所以 $\alpha \approx 51.6^\circ$，而 $m/M = 4^{-1/3} \approx 0.63$。对 f 在 $(0, \frac{\pi}{2})$ 上作网格搜索，所得的极大值位于同一个 α，其值为 $f \approx 0.240$，所以最大加速度约为 $0.240g$。
- 以随机的质量和角度代入第 (i) 部分的一般公式，每当 $\tan\alpha = 2m/M$ 时，结果都与第 (ii) 部分的公式吻合，而且两条运动方程给出相同的正张力。

常见错误

- 对 m 使用 $mg\sin\alpha$：它所在斜面的倾角是 $\frac{\pi}{2} - \alpha$，所以分量应是 $mg\cos\alpha$。
- 固定 m 而让 M 变化（或反过来）去求极大值，却没有留意 a 只取决于 α。把一切都用 α 表示，问题便成为单变量问题。

STEP 之外

像这样由约束（constraint）把物体连系起来的系统，正是拉格朗日力学（Lagrangian mechanics）的起点：你只需选取一个坐标（这里是沿绳子移动的距离 s），张力和法向反作用力等约束力便会自动消失。大学一年级的课程会把这个想法发展成一套有系统的方法，适用于滑轮、摆以及许多其他情况。若想找一本在这个程度上清楚讲解牛顿力学的书，可以看 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）。

详细解答

STEP I 2009，第 11 题

力学 · 碰撞与恢复系数

STEP I 2009，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/09-s1-q11

关键思路

每一个一维碰撞都由两条方程支配：动量守恒（conservation of momentum）和牛顿恢复定律（Newton's law of restitution）。处理能量的窍门，是把每个质点的损失写成平方差 $\frac{1}{2}m(u-p)(u+p)$，因为动量守恒把因式 $u-p$ 和 $q-v$ 连系起来。一旦抽出公因式 $u-p$，余下的括号正好就是恢复定律所描述的东西。

解答

我们把两个质点模拟为在光滑桌面上沿一条直线运动，速度以 **i** 的方向量度。由于 $u > v$，质量为 m 的质点在后面，并会追上前面的质点。水平方向没有外力，所以动量守恒：

$$mu + Mv = mp + Mq, \qquad \text{所以} \qquad M(q-v) = m(u-p). \tag{1}$$

牛顿恢复定律指出，分离速率是接近速率的 e 倍：

$$q - p = e(u-v). \tag{2}$$



碰撞前（速度沿 **i** 方向），$u > v$

第 (i) 部分

动能（kinetic energy）的损失为

$$\Delta = \tfrac{1}{2}m(u^2-p^2) + \tfrac{1}{2}M(v^2-q^2) = \tfrac{1}{2}m(u-p)(u+p) - \tfrac{1}{2}M(q-v)(q+v).$$

根据 (1)，我们可以把 $M(q-v)$ 换成 $m(u-p)$，于是可抽出因式 $\frac{1}{2}m(u-p)$：

$$\begin{aligned}\Delta &= \tfrac{1}{2}m(u-p)\big[(u+p)-(q+v)\big]\\ &= \tfrac{1}{2}m(u-p)\big[(u-v)-(q-p)\big]\\ &= \tfrac{1}{2}m(u-p)\big[(u-v)-e(u-v)\big] \qquad \text{根据 (2)}\\ &= \tfrac{1}{2}m(u-p)(u-v)(1-e),\end{aligned}$$

这正是所求的结果。

想一想 一条能量损失的公式，怎样能告诉我们关于 p 的事情?

碰撞不可能产生动能，所以 $\Delta \geq 0$。在乘积 $\frac{1}{2}m(u-p)(u-v)(1-e)$ 中，我们有 $m>0$、$u-v>0$（已知）和 $1-e>0$（因为 $e<1$）。所以 Δ 的正负号就是 $u-p$ 的正负号，而 $\Delta \geq 0$ 迫使 $u \geq p$。质量为 m 的质点不可能在 $\mathbf{i}$ 的方向上加速，这是合理的，因为 M 施加于它的冲量（impulse）是向后的。

如果你不想把 $\Delta \geq 0$ 当作物理事实来依赖，可以直接解 (1) 和 (2)：

$$p=\frac{mu+Mv-Me(u-v)}{m+M}, \qquad \text{所以} \qquad u-p=\frac{M(1+e)(u-v)}{m+M}>0$$

对 $e \geq 0$ 成立。这就确认了 $u \geq p$，因而 $\Delta \geq 0$。

第 (ii) 部分

每个质点损失相同而非零的能量，所以

$$\tfrac{1}{2}m(u-p)(u+p)=\tfrac{1}{2}M(v-q)(v+q)\neq 0.$$

根据 (1)，$M(v-q)=-m(u-p)$，所以右边是 $-\frac{1}{2}m(u-p)(v+q)$，于是

$$\tfrac{1}{2}m(u-p)(u+p)=-\tfrac{1}{2}m(u-p)(v+q).$$

左边非零，所以 $m(u-p)\neq 0$，我们可以除以 $\frac{1}{2}m(u-p)$：

$$u+p=-(v+q), \qquad \text{即} \qquad u+v+p+q=0. \tag{3}$$

现在 (2) 和 (3) 给出了 p 与 q 的和与差：

$$p+q=-(u+v), \qquad q-p=e(u-v),$$

所以

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

我们还未用上动量守恒的完整形式，而正是它决定了 e 的值。代入 $mp+Mq=mu+Mv$ 并乘以 2：

$$\begin{aligned}-(m+M)(u+v)+(M-m)e(u-v) &= 2mu+2Mv,\\ (M-m)e(u-v) &= (m+M)(u+v)+2mu+2Mv\\ &= (M+3m)u+(3M+m)v.\end{aligned}$$

由于 $u-v>0$，而且当 $m\neq M$ 时 $M-m\neq 0$，我们可以相除而得

$$e=\frac{(M+3m)u+(3M+m)v}{(M-m)(u-v)},$$

这正是所求的结果。

验算

- 取 $m=1$、$M=3$、$u=1$、$v=-\frac{1}{2}$。公式给出 $e=\frac{6-5}{2\times\frac{3}{2}}=\frac{1}{3}$。于是 $p=-\frac{1}{2}$ 及 $q=0$，而动量亦吻合：$1-\frac{3}{2}=-\frac{1}{2}+0$。每个质点损失 $\frac{3}{8}$ 的能量，而且 $u+v+p+q=1-\frac{1}{2}-\frac{1}{2}+0=0$。
- 当 $e=1$ 时，(i) 的公式给出零损失，这正符合完全弹性碰撞（perfectly elastic collision）应有的情况。
- 电脑以精确分数在 2000 个随机碰撞上检验了 (i)，并在 455 个满足 $0\leq e<1$ 的随机个案上检验了 (ii)。

常见错误

- 把恢复定律写反了：应该是 $q-p=e(u-v)$，即分离除以接近，而这里两者都以正数量度。
- 除以 $u-p$ 却没有说明它为何非零。题目中“非零的能量”这几个字，正是为此而设。

STEP 之外

(i) 中的因式分解，是在质心参考系（centre-of-mass frame）中考虑问题的一个漂亮结果。在这个参考系中，总动量为零，动能为 $\frac{1}{2}\mu$(相对速率)^2，其中约化质量（reduced mass）为 $\mu=\frac{mM}{m+M}$，而损失则为 $\frac{1}{2}\mu(1-e^2)(u-v)^2$。约化质量和质心参考系会在二体轨道和粒子散射中再次出现。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）对这些概念有细致的阐述。

详细解答

STEP I 2009，第 12 题

概率与统计 · 概率与不等式

STEP I 2009，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/09-s1-q12

关键思路

这里每个概率（probability）都是各袋糖果数目之积的和，再除以总数的某次幂，所以每次比较最终都是比较两个以 a, b, c 为变量的多项式（polynomial）。开首的不等式（inequality）$x^2 + y^2 \geq 2xy$ 正是为此而设的工具。在第 (ii) 部分，直接数算“恰好两颗同色”相当繁琐。先数算另外两种情况（全部同色、全部不同色），再从 1 中减去，会容易得多。

解答

不等式

对任意实数 x 和 y，

$$x^2 + y^2 - 2xy = (x - y)^2 \geq 0,$$

因为实数的平方永不为负。所以 $x^2 + y^2 \geq 2xy$，等号恰好在 $x = y$ 时成立。

第 (i) 部分

每个袋有 $a + b$ 颗糖果，而两次抽取来自不同的袋，所以互相独立（independent）。设 S 为两颗糖果同色的事件（event），D 为两颗糖果不同色的事件。那么

$$\mathrm{P}(S) = \mathrm{P}(\text{红、红}) + \mathrm{P}(\text{蓝、蓝}) = \frac{a}{a+b} \cdot \frac{b}{a+b} + \frac{b}{a+b} \cdot \frac{a}{a+b} = \frac{2ab}{(a+b)^2},$$

$$\mathrm{P}(D) = \mathrm{P}(\text{红、蓝}) + \mathrm{P}(\text{蓝、红}) = \frac{a}{a+b} \cdot \frac{a}{a+b} + \frac{b}{a+b} \cdot \frac{b}{a+b} = \frac{a^2 + b^2}{(a+b)^2}.$$

两个分母相等且为正数，而由上面的不等式可知 $2ab \leq a^2 + b^2$。因此 $\mathrm{P}(S) \leq \mathrm{P}(D)$，正如所求，等号只在 $a = b$ 时成立。

第 (ii) 部分

记 $n = a + b + c$，即每个袋中糖果的数目。各袋的内容如下：

	红	白	黄
第 1 袋	a	b	c
第 2 袋	b	c	a
第 3 袋	c	a	b

三次抽取互相独立，所以从各袋（按次序）抽出某一指定颜色的概率，就是三个相应数目之积除以 n^3。设 A 为三颗糖果全部同色的事件，T 为恰好两颗同色的事件，E 为三颗全部不同色的事件。这三个事件涵盖了所有可能，而且互不重叠。

全部同色。 沿表中每一直行向下读，

$$\mathrm{P}(A) = \frac{abc + bca + cab}{n^3} = \frac{3abc}{n^3}.$$

全部不同色。 从第 1、2、3 袋抽出的颜色，构成红、白、黄的 $3! = 6$ 种排列之一：

红白黄：$a \cdot c \cdot b$　红黄白：$a \cdot a \cdot a$　白红黄：$b \cdot b \cdot b$
白黄红：$b \cdot a \cdot c$　黄红白：$c \cdot b \cdot a$　黄白红：$c \cdot c \cdot c$

所以

$$\mathrm{P}(E) = \frac{a^3 + b^3 + c^3 + 3abc}{n^3}.$$

? **想一想** 怎样才能由这两个答案得出 $\mathrm{P}(T)$，而不必列出余下的 18 个结果?

由于 $\mathrm{P}(T) = 1 - \mathrm{P}(A) - \mathrm{P}(E)$，我们需要 $n^3 = (a + b + c)^3$ 的展开式。它包含三个立方项、六个形如 a^2b 而系数为 3 的项，以及系数为 6 的 abc：

$$(a + b + c)^3 = a^3 + b^3 + c^3 + 3\left(a^2b + b^2c + c^2a + ab^2 + bc^2 + ca^2\right) + 6abc.$$

因此

$$\begin{aligned}\mathrm{P}(T) &= \frac{n^3 - 3abc - (a^3 + b^3 + c^3 + 3abc)}{n^3} \\ &= \frac{3\left(a^2b + b^2c + c^2a + ab^2 + bc^2 + ca^2\right)}{(a + b + c)^3},\end{aligned}$$

这就是题目给出的结果。三颗糖果全部同色的概率是

$$\mathrm{P}(A) = \frac{3abc}{(a + b + c)^3}.$$

推论。 把六个项按公因子两两分组：

$$a^2b + b^2c + c^2a + ab^2 + bc^2 + ca^2 = a\left(b^2 + c^2\right) + b\left(c^2 + a^2\right) + c\left(a^2 + b^2\right).$$

由开首的不等式，$b^2 + c^2 \geq 2bc$；又因为 $a > 0$，乘上后不等号方向不变：$a(b^2 + c^2) \geq 2abc$。同理 $b(c^2 + a^2) \geq 2abc$ 及 $c(a^2 + b^2) \geq 2abc$。相加得

$$a^2b + b^2c + c^2a + ab^2 + bc^2 + ca^2 \geq 6abc.$$

乘以 $3/n^3 > 0$，

$$\mathrm{P}(T) \geq \frac{18abc}{n^3} = 6 \cdot \frac{3abc}{n^3} = 6\,\mathrm{P}(A),$$

正如所求。等号成立需要 $b = c$、$c = a$ 和 $a = b$ 同时成立，所以只在 $a = b = c$ 时成立。

验算

- 当 $a = b = c$ 时，每个袋中三种颜色各占三分之一，所以 $\mathrm{P}(A) = 3 \cdot \frac{1}{27} = \frac{1}{9}$ 及 $\mathrm{P}(E) = \frac{6}{27} = \frac{2}{9}$，余下 $\mathrm{P}(T) = \frac{2}{3} = 6 \cdot \frac{1}{9}$：这正是等号成立的情况。
- 当 $a, b, c = 1, 2, 3$ ($n = 6$) 时：$\mathrm{P}(A) = \frac{18}{216} = \frac{1}{12}$、$\mathrm{P}(E) = \frac{54}{216} = \frac{1}{4}$ 及 $\mathrm{P}(T) = \frac{144}{216} = \frac{2}{3}$。三者之和为 1，而 $\frac{2}{3}$ 是 $\frac{1}{12}$ 的 8 倍。
- 我们以精确穷举（enumeration）的方法，对 a, b, c 由 1 至 8 的每一组取值验证了全部三条公式，并以模拟（simulation）作检验（取 $a, b, c = 2, 3, 5$，200 000 次试验得出 0.659，而 $\mathrm{P}(T)$ 的精确值为 0.66）。

常见错误

- 把各袋的内容次序写错。第 2 袋有 a 颗黄色糖果，而不是 a 颗红色糖果；像上面那样列一个表就可避免这种错误。
- 以为 $\mathrm{P}(E)$ 中每种排列都给出 abc。其实只有三种如此；红黄白、白红黄和黄白红这三种排列给出的是 a^3、b^3 和 c^3。
- 把不等式乘以 a、b 或 c 时，忘记说明 $a, b, c > 0$。

STEP 之外

最后的不等式是 Muirhead 不等式（Muirhead's inequality）的一个特例，后者是比较单项式（monomial）对称和的一般法则；在这里它说明形如 a^2b 的各项之和，不小于形如 abc 的各项之和。第 (i) 部分是同一想法在两个变量时的情况，而两者都建基于算术—几何平均不等式（AM–GM inequality）。想进一步了解计数论证和互补事件（complementary events），可参阅 Blitzstein and Hwang, *Introduction to Probability*（概率论入门教科书）。

详细解答

STEP I 2009，第 13 题

概率与统计 · 排列与游程

STEP I 2009，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/09-s1-q13

关键思路

重要的只是女孩所占的位置，而不是哪个女孩或男孩是谁。因此，与其处理 $(n+3)!$ 种次序，不如直接选出女孩所占的三个位置所组成的集合：$\binom{n+3}{3}$ 个集合中每一个都是等可能的。要数出符合条件的集合，最简单的做法是先把 n 个男孩排成一行，再考虑女孩要放进男孩之间及两旁的哪些空隙。

解答

由于全部 $(n+3)!$ 种次序都是等可能的，而女孩的每一个三位置集合都对应相同数目（即 $3!\,n!$）的次序，所以女孩位置的集合在

$$\binom{n+3}{3} = \frac{(n+3)(n+2)(n+1)}{6}$$

个可能的集合上均匀分布（uniformly distributed）。在整个解答中，K 只可能是 1、2 或 3。

排成一行的 n 个男孩会产生 $n+1$ 个空隙（第一个男孩之前有一个，最后一个男孩之后有一个，相邻男孩之间共有 $n-1$ 个）：



把女孩放进空隙（同一个空隙可以放任意多个）便得到一行，而每一行都恰好以一种方式产生。同一个空隙内的女孩彼此相邻；不同空隙内的女孩之间则至少隔着一个男孩。

第 (i) 部分

$K=3$ 表示三个女孩全部连成一段，即她们都放进同一个空隙。空隙有 $n+1$ 种选法，所以

$$\mathrm{P}(K=3)=\frac{n+1}{\binom{n+3}{3}}=\frac{6(n+1)}{(n+3)(n+2)(n+1)}=\frac{6}{(n+2)(n+3)}.$$

（等价的说法：由三个连续位置组成的一段，可以从第 $1,2,\ldots,n+1$ 个位置中任何一个开始。）

第 (ii) 部分

$K=1$ 表示没有两个女孩相邻，即三个女孩分别在三个不同的空隙内。我们从 $n+1$ 个空隙中选出 3 个，共有 $\binom{n+1}{3}$ 种选法，然后在每个空隙放一个女孩。因此

$$\mathrm{P}(K=1)=\frac{\binom{n+1}{3}}{\binom{n+3}{3}}=\frac{(n+1)n(n-1)}{(n+3)(n+2)(n+1)}=\frac{n(n-1)}{(n+2)(n+3)},$$

正是所求。（当 $n=0$ 或 1 时，空隙不足，而公式亦正确地给出 0。）

> ? **想一想** 为甚么数目是 $\binom{n+1}{3}$，而不是 $\binom{n+1}{3}\times 3!$?

我们数的是位置的集合，而不是有名有姓的孩子的次序。选定使用哪三个空隙，女孩的位置便完全确定；至于哪个女孩坐在哪里，早已包含在“每个位置集合都是等可能的”这个论断之中。

第 (iii) 部分

首先我们需要求 $\mathrm{P}(K=2)$。由于 $K\in\{1,2,3\}$，

$$\begin{aligned}\mathrm{P}(K=2)&=1-\frac{6}{(n+2)(n+3)}-\frac{n(n-1)}{(n+2)(n+3)}\\&=\frac{(n^2+5n+6)-6-(n^2-n)}{(n+2)(n+3)}=\frac{6n}{(n+2)(n+3)}.\end{aligned}$$

作为验算，我们直接数一次：$K=2$ 表示有一对女孩在同一个空隙，而第三个女孩在另一个空隙。先为那一对选空隙，再为单独的一个选另一个空隙，共得 $(n+1)n$ 个集合，而 $(n+1)n/\binom{n+3}{3}=6n/\big((n+2)(n+3)\big)$，与上面的结果相同。

于是

$$\begin{aligned}\mathrm{E}(K) &= 1\cdot\frac{n(n-1)}{(n+2)(n+3)} + 2\cdot\frac{6n}{(n+2)(n+3)} + 3\cdot\frac{6}{(n+2)(n+3)}\\ &= \frac{n^2+11n+18}{(n+2)(n+3)} = \frac{(n+2)(n+9)}{(n+2)(n+3)}\\ &= \frac{n+9}{n+3}.\end{aligned}$$

验算

- $n=0$：只有女孩，所以必定有 $K=3$，而 $\frac{n+9}{n+3}=3$。
- $n=1$：可能的排列是 GGGB、GGBG、GBGG、BGGG，四者等可能，所以 $K=3$ 或 2 的概率各为 $\frac{1}{2}$，而 $\mathrm{E}(K)=\frac{5}{2}=\frac{10}{4}$。
- 当 $n\to\infty$ 时，$\mathrm{E}(K)\to 1$：男孩很多时，女孩几乎不会相邻。
- 我们以精确分数列举了 $n=0,\ldots,9$ 的所有位置集合(例如 $n=3$ 时得 $\mathrm{P}(K=1,2,3)=\frac{1}{5},\frac{3}{5},\frac{1}{5}$ 及 $\mathrm{E}(K)=2$)，并对 $n\le 4$ 列举了有标记的孩子的全部 $(n+3)!$ 种次序；所有结果都吻合。

常见错误

- 在同一个分数中，把有序排列的数目与位置集合的数目混在一起。选定一个样本空间（sample space），然后贯彻使用。
- 忘记两端的两个空隙，结果用了 $n-1$ 个空隙而不是 $n+1$ 个。

STEP 之外

“男孩造出空隙，女孩放进空隙”这个技巧，就是隔板法（stars and bars）的想法；而随机排列中最长连段（longest run）这个一般问题，则是一个经典课题：对于很长的随机序列，最长连段的长度大致按序列长度的对数增长。大学本科的概率课程会通过指示变量（indicator variables）和母函数（generating functions）来处理这些问题。Grimmett and Welsh 的 *Probability: An Introduction*（概率论入门教材）是学习如何发展这类计数论证的好去处。

详细解答

STEP II 2009，第 1 题

纯数学 · 曲线的对称性

STEP II 2009，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/09-s2-q1

关键思路

这一题几乎不用计算：靠对称性（symmetry）就能完成大部分工作。两条曲线对直线 $y = x$ 和 $y = -x$ 作反射（reflection）后都保持不变，所以只要有了一个交点，其余交点都可以由它反射得到。接着，由这四点的位置形状便可推出长方形（rectangle）。至于面积，我们根本不必求出 α 和 β 本身：答案原来只需要它们的一个组合，而这个组合可以直接由 $\alpha^4 + \beta^4 = u$ 和 $\alpha\beta = v$ 得到。

解答

对称轴

如果把曲线上任意一点对某条直线作反射，所得的点仍然在曲线上，我们就说该曲线对这条直线对称。我们需要的四个反射是

$$(x,y) \mapsto (-x,y), \quad (x,y) \mapsto (x,-y), \quad (x,y) \mapsto (y,x), \quad (x,y) \mapsto (-y,-x),$$

它们分别是对 $x = 0$、$y = 0$、$y = x$ 和 $y = -x$ 的反射。

曲线 $x^4 + y^4 = u$。表达式 $x^4 + y^4$ 在这四个映射下都保持不变，因为四次方与正负号无关，而且表达式对 x 和 y 是对称的。所以这条曲线有四条对称轴（lines of symmetry）：

$$x = 0, \qquad y = 0, \qquad y = x, \qquad y = -x.$$

曲线 $xy = v$。把 x 和 y 互换，xy 保持不变；$(x,y) \mapsto (-y,-x)$ 也是如此，因为 $(-y)(-x) = xy$。只改变其中一个坐标的正负号，会把 xy 变成 $-xy$，而由于 $v > 0$，点 $(-x,y)$ 便落在 $xy = -v$ 上，而不在我们的曲线上。所以这条等轴双曲线（rectangular hyperbola）恰好

有两条对称轴：

$$y = x, \qquad y = -x.$$

四个交点

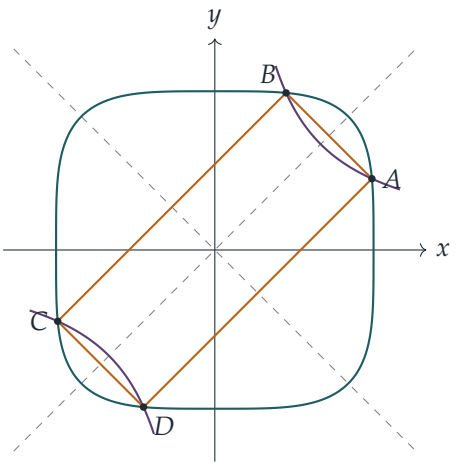
设 S 为对 $y = x$ 的反射，T 为对 $y = -x$ 的反射。两者都把两条曲线各自映射到自身，所以都把交点映射到交点。由 $A = (\alpha, \beta)$ 出发，我们得到

$$S(A) = (\beta, \alpha), \qquad T(A) = (-\beta, -\alpha), \qquad T(S(A)) = (-\alpha, -\beta).$$

因为 $\alpha > \beta > 0$，这四点互不相同。此外再没有其他交点。把 $y = v/x$ 代入四次曲线（quartic）的方程，得 $x^8 - ux^4 + v^4 = 0$，这是 x^4 的二次方程，最多有两个正根，所以 x 最多有四个实数值。

要按逆时针方向为这些点命名，可以看它们的极角（polar angle）。由于 $0 < \beta < \alpha$，点 A 的极角 θ 满足 $0 < \theta < 45°$。于是 (β, α) 的极角是 $90° - \theta$，$(-\alpha, -\beta)$ 的极角是 $180° + \theta$，而 $(-\beta, -\alpha)$ 的极角是 $270° - \theta$。这些角按此次序递增，所以

$$B = (\beta, \alpha), \qquad C = (-\alpha, -\beta), \qquad D = (-\beta, -\alpha).$$



（图中所示为 $u = 81$、$v = 4$ 的情况：四次曲线以蓝绿色表示，双曲线以紫红色表示。）

$ABCD$ 是长方形

对角线是 AC 和 BD。AC 的中点是 $\left(\frac{\alpha-\alpha}{2}, \frac{\beta-\beta}{2}\right) = (0, 0)$，同样地，$BD$ 的中点也是 $(0, 0)$。所以两条对角线互相平分，$ABCD$ 是平行四边形（parallelogram）。它们的长度是

$$AC = \sqrt{(2\alpha)^2 + (2\beta)^2} = 2\sqrt{\alpha^2 + \beta^2}, \qquad BD = \sqrt{(2\beta)^2 + (2\alpha)^2} = 2\sqrt{\alpha^2 + \beta^2},$$

两者相等。对角线相等的平行四边形是长方形，所以 $ABCD$ 是长方形。

想一想 为甚么对角线相等的平行四边形必定是长方形?

四个顶点与 O 的距离同为 $\sqrt{\alpha^2+\beta^2}$，所以它们都在以 O 为圆心的一个圆上，而 AC 和 BD 是这个圆的直径。根据半圆上的圆周角（angle in a semicircle）定理，$ABCD$ 的每一只角都是直角。如果想直接验证，则有 $\overrightarrow{AB}=(\beta-\alpha,\ \alpha-\beta)$ 和 $\overrightarrow{BC}=(-\alpha-\beta,\ -\alpha-\beta)$，其纯量积（scalar product）为

$$-(\beta-\alpha)(\alpha+\beta)-(\alpha-\beta)(\alpha+\beta)=(\alpha+\beta)\big[(\alpha-\beta)-(\alpha-\beta)\big]=0.$$

面积

边长是

$$AB=\sqrt{2(\alpha-\beta)^2}=\sqrt{2}\,(\alpha-\beta),\qquad BC=\sqrt{2(\alpha+\beta)^2}=\sqrt{2}\,(\alpha+\beta),$$

这里用了 $\alpha>\beta>0$ 来取正平方根。所以

$$\text{面积}=AB\cdot BC=2(\alpha-\beta)(\alpha+\beta)=2(\alpha^2-\beta^2).$$

现在利用 A 同时在两条曲线上这个事实，即 $\alpha^4+\beta^4=u$ 和 $\alpha\beta=v$。于是

$$(\alpha^2-\beta^2)^2=\alpha^4+\beta^4-2\alpha^2\beta^2=u-2v^2.$$

由于 $\alpha^2-\beta^2>0$，我们取正根，得

$$\boxed{\text{长方形面积}ABCD=2\sqrt{u-2v^2}}$$

这也说明，题目假设有四个不同的交点，就必然有 $u>2v^2$。如果 $u=2v^2$，则 $\alpha=\beta$，两条曲线只是相切。

当 $u=81$ 及 $v=4$ 时：$u-2v^2=81-32=49$，所以面积是 $2\times7=14$，与题目要求相符。

验算

当 $u=81$、$v=4$ 时，以数值方法解 $x^8-81x^4+256=0$，得 $\alpha\approx2.9690$ 及 $\beta=4/\alpha\approx1.3472$。于是 $2(\alpha^2-\beta^2)\approx14.000$，而对四个顶点使用鞋带公式（shoelace formula）也得到相同的面积。当 u 下降并趋向 $2v^2$ 时，每个象限内的两点在 $y=x$ 上合而为一，长方形压扁成一条线段，与面积 $2\sqrt{u-2v^2}\to0$ 吻合。

常见错误

- 把坐标轴当作 $xy=v$ 的对称轴。对坐标轴作反射，会把第一象限的分支送到第二象限，而那里 $xy<0$。

- 把各点的次序标错，例如取 $B = (-\alpha, -\beta)$。这样 $ABCD$ 便会连成一个交叉四边形。
- 试图明确地解出 α 和 β。面积只需要 $\alpha^2 - \beta^2$，而它可以直接由 u 和 v 得到。

STEP 之外

保持 $x^4 + y^4 = u$ 不变的四个反射和各个旋转，构成正方形的对称群（symmetry group），即阶为 8 的二面体群（dihedral group）。双曲线只保留了其中一个阶为 4 的子群（subgroup）（两个反射、半周旋转和恒等变换）。两条曲线共有的对称变换会把交点互相置换，而像这样借着群的作用来寻找对象，正是大学一年级群论（group theory）的一个主题。Allenby 的 *Rings, Fields and Groups*（环、域与群）是认识这些概念的好起点。

详细解答

STEP II 2009，第 2 题

纯数学 · 曲线描绘与近似

STEP II 2009，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/09-s2-q2

关键思路

由于 $a > 1$，函数 a^u 对 u 严格递增（strictly increasing），所以 y 恰好在 $\sin(\pi e^x)$ 最大之处取得最大值，在它最小之处取得最小值。因此，这条曲线的一切性质都来自内层函数 πe^x：它稳定地递增，依次经过正弦函数的波峰和波谷，而且随着 x 增大，这些点会愈来愈密集。至于面积，各梯形（trapezium）两端的高度相同，于是整个和式便化简为单单一个对数之差。

解答

第 (i) 部分

取对数（logarithm），得 $\ln y = \ln a \, \sin(\pi e^x)$。以隐函数求导（implicit differentiation），

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = \pi \ln a \, e^x \cos(\pi e^x), \qquad \text{所以} \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \pi \ln a \, e^x \cos(\pi e^x)\, a^{\sin(\pi e^x)}.$$

这里 $\ln a > 0$（因为 $a > 1$），$e^x > 0$ 而且 $a^{\sin(\pi e^x)} > 0$，所以导数恰好在 $\cos(\pi e^x) = 0$ 时为零，即 πe^x 是 $\frac{\pi}{2}$ 的奇数倍之时。由于 $\pi e^x > 0$，只有正的奇数倍才有可能：

$$e^x = n + \tfrac{1}{2}, \qquad x = \ln\left(n + \tfrac{1}{2}\right), \qquad n = 0, 1, 2, \ldots$$

在这些点上 $\sin(\pi e^x) = \sin\left((n + \frac{1}{2})\pi\right) = (-1)^n$，所以 $y = a^{(-1)^n}$。驻点（stationary point）为

$$\left(\ln\left(n + \tfrac{1}{2}\right),\, a\right) \ (n \text{ 为偶数}), \qquad \left(\ln\left(n + \tfrac{1}{2}\right),\, \tfrac{1}{a}\right) \ (n \text{ 为奇数}).$$

第一组是极大点，第二组是极小点。由于 a^u 递增，对任何 x 都有 $y \le a^1 = a$ 和 $y \ge a^{-1}$，所以 $y = a$ 的点是全局极大点（global maximum），$y = 1/a$ 的点是全局极小

点（global minimum）。具体来说，极大点位于 $x = \ln\frac{1}{2}, \ln\frac{5}{2}, \ln\frac{9}{2}, \ldots$，极小点位于 $x = \ln\frac{3}{2}, \ln\frac{7}{2}, \ln\frac{11}{2}, \ldots$

第 (ii) 部分

当 x 很小时，$e^x \approx 1 + x$，所以

$$\sin(\pi e^x) \approx \sin(\pi + \pi x) = -\sin(\pi x) \approx -\pi x,$$

这里先用了 $\sin(\pi + \theta) = -\sin\theta$，再用 $\sin t \approx t$，其中 $t = \pi x$ 很小。因此

$$y \approx a^{-\pi x} = e^{-\pi x \ln a} \approx 1 - \pi x \ln a,$$

最后一步再次用到 $e^t \approx 1 + t$，这次的小量是 $t = -\pi x \ln a$。这就是所求的近似式：曲线在 $x = 0$ 处穿过直线 $y = 1$，斜率（gradient）为 $-\pi \ln a$。

> ? **想一想** 为甚么不能把 $\sin t \approx t$ 直接代入 $\sin(\pi e^x)$?

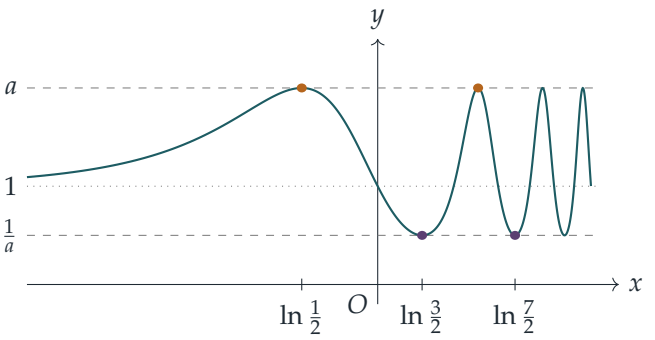
因为当 x 很小时，πe^x 接近的是 π，而不是 0。我们必须先用 $\sin(\pi + \theta) = -\sin\theta$，把小量 πx 移到正弦函数的自变量之中。

第 (iii) 部分

草图上应标示的一些特征：

- 当 $x \to -\infty$ 时，$e^x \to 0$，所以 $\sin(\pi e^x) \to 0$，而 $y \to 1$，并且是从上方趋近（当 $x < \ln\frac{1}{2}$ 时，我们有 $0 < \pi e^x < \frac{\pi}{2}$，所以正弦值为正，$y > 1$）。直线 $y = 1$ 是曲线在左方的渐近线（asymptote）。
- 曲线上升至位于 $x = \ln\frac{1}{2}$ 的第一个极大值 a，然后永远在 a 与 $1/a$ 之间振荡，并且始终在 x 轴上方。
- 每当 e^x 是正整数时，曲线便穿过 $y = 1$，即在 $x = 0, \ln 2, \ln 3, \ldots$ 处；而由第 (ii) 部分可知，在 $x = 0$ 处的斜率为 $-\pi \ln a$。
- 相邻驻点之间的距离为 $\ln\frac{n+3/2}{n+1/2} \to 0$，所以随着 x 增大，振荡愈来愈快。

下面的草图显示 $a = 2$ 的情况。



第 (iv) 部分

从左边数起：第 k 个极大点对应 $n = 2(k-1)$，所以它位于 $\mathrm{e}^x = 2k - \frac{3}{2}$。第 $(k+1)$ 个极大点位于 $\mathrm{e}^x = 2k + \frac{1}{2}$，而两者之间唯一的极小点位于 $\mathrm{e}^x = 2k - \frac{1}{2}$。记

$$x_1 = \ln\left(2k - \tfrac{3}{2}\right), \qquad x_2 = \ln\left(2k - \tfrac{1}{2}\right), \qquad x_3 = \ln\left(2k + \tfrac{1}{2}\right).$$

用直线把 (x_1, a) 连到 $(x_2, \frac{1}{a})$，再把 $(x_2, \frac{1}{a})$ 连到 (x_3, a)，则直线以下至 x 轴的区域分成两个梯形，每个梯形的两条平行边长度都是 a 和 $\frac{1}{a}$。所以面积的近似值为

$$\begin{aligned}\tfrac{1}{2}\left(a + \tfrac{1}{a}\right)(x_2 - x_1) + \tfrac{1}{2}\left(a + \tfrac{1}{a}\right)(x_3 - x_2) &= \frac{a^2+1}{2a}\,(x_3 - x_1) \\ &= \frac{a^2+1}{2a}\ln\frac{2k+\frac{1}{2}}{2k-\frac{3}{2}} = \frac{a^2+1}{2a}\ln\frac{k+\frac{1}{4}}{k-\frac{3}{4}}.\end{aligned}$$

最后，$\dfrac{k+\frac{1}{4}}{k-\frac{3}{4}} = 1 + \dfrac{1}{k-\frac{3}{4}}$，所以面积近似于

$$\left(\frac{a^2+1}{2a}\right)\ln\left(1 + \left(k - \tfrac{3}{4}\right)^{-1}\right),$$

正如题目所求。留意极小点的位置 x_2 完全消去了：只有两个极大点起作用。

验算

- 以数值方法求导，可确认在 $n = 0, \ldots, 7$ 时 $\mathrm{d}y/\mathrm{d}x = 0$ 于 $x = \ln(n + \frac{1}{2})$ 成立，而在 $[-6, \ln 8]$ 上再无其他零点；同时亦确认在 $x = 0$ 处的斜率为 $-\pi \ln a$（取 $a = 2$ 时，两者皆为 -2.1776）。
- 取 $a = 2$、$k = 1$ 时，公式给出 $\frac{5}{4}\ln 5 \approx 2.012$，而用辛普森法则（Simpson's rule）求得的真实面积是 1.954：两者相当接近，正是近似值应有的表现。梯形的估算略为偏高：曲线在极小点附近低于弦线的幅度，大于它在极大点附近高出弦线的幅度。
- 当 $k \to \infty$ 时，面积趋向 0，与振荡愈来愈窄的现象吻合。

常见错误

- 在第 (i) 部分把 $\frac{\pi}{2}$ 的负奇数倍也算进去：πe^x 永远是正数。
- 从 $k = 0$ 而不是 $k = 1$ 开始为极大点编号，结果答案会变成 $\ln\left(1 + (k + \frac{1}{4})^{-1}\right)$。
- 在第 (ii) 部分停在 $a^{-\pi x}$，没有做最后一步线性化 $e^t \approx 1 + t$。

STEP 之外

第 (ii) 部分其实是把较简单的展开式复合起来，从而得出一阶泰勒展开（Taylor expansion），这正是大学本科生处理复合函数展开的方法。当 k 很大时，一次振荡之下的真实面积大约是 $(x_3 - x_1)$ 乘以 $a^{\sin\theta}$ 在一个周期上的平均值，这个积分会引出修正贝塞尔函数（modified Bessel function）$I_0(\ln a)$；把它与 $\frac{1}{2}(a + \frac{1}{a})$ 比较，是一道很好的练习，可借此明白为甚么梯形法则（trapezium rule）对这类曲线会高估。若想看对近似与泰勒定理（Taylor's theorem）的严谨论述，可参阅 Abbott 的 *Understanding Analysis*（数学分析入门书）。

详细解答

STEP II 2009，第 3 题

纯数学 · 三角恒等式与根式

> **STEP II 2009，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/09-s2-q3

关键思路

恒等式（identity）$(*)$ 把你已知的角 x 的正切（tangent），化为角 $\frac{1}{4}\pi - \frac{1}{2}x$ 的正切，而后者的大小约为前者的一半。因此它就像一部半角（half-angle）机器。功夫在于选取一个起始角 x，使它恰好落在你想要的角上。之后的工作就是细心地进行根式（surd）运算：有理化（rationalising）、展开和平方。

解答

证明 $(*)$

设 $\theta = \frac{1}{4}\pi - \frac{1}{2}x$，则 $x = \frac{1}{2}\pi - 2\theta$。于是

$$\sin x = \cos 2\theta = 1 - 2\sin^2\theta, \qquad \cos x = \sin 2\theta = 2\sin\theta\cos\theta.$$

因此，只要两边均有定义（即 $\cos x \neq 0$），便有

$$\sec x - \tan x = \frac{1-\sin x}{\cos x} = \frac{2\sin^2\theta}{2\sin\theta\cos\theta} = \frac{\sin\theta}{\cos\theta} = \tan\left(\tfrac{1}{4}\pi - \tfrac{1}{2}x\right).$$

约去 $\sin\theta$ 是容许的，因为 $\cos x = 2\sin\theta\cos\theta \neq 0$ 必然使 $\sin\theta \neq 0$。这就证明了 $(*)$。

第 (i) 部分

我们要 $\frac{1}{4}\pi - \frac{1}{2}x = \frac{1}{8}\pi$，由此得 $x = \frac{1}{4}\pi$。由于 $\sec\frac{1}{4}\pi = \sqrt{2}$ 及 $\tan\frac{1}{4}\pi = 1$，

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

至于第二个结果，留意 $\frac{11}{24}\pi = \frac{3}{24}\pi + \frac{8}{24}\pi = \frac{1}{8}\pi + \frac{1}{3}\pi$，而我们知道 $\tan\frac{1}{3}\pi = \sqrt{3}$。由复角公式（addition formula）得

$$\tan\tfrac{11}{24}\pi = \frac{\tan\frac{1}{8}\pi + \tan\frac{1}{3}\pi}{1 - \tan\frac{1}{8}\pi\tan\frac{1}{3}\pi} = \frac{(\sqrt{2}-1)+\sqrt{3}}{1-(\sqrt{2}-1)\sqrt{3}} = \frac{\sqrt{3}+\sqrt{2}-1}{\sqrt{3}-\sqrt{6}+1},$$

这正是所求的结果。

第 (ii) 部分

与其把一个有三项的分母有理化，不如直接证明以下等式，这样更快：

$$(2+\sqrt{2}+\sqrt{3}+\sqrt{6})(\sqrt{3}-\sqrt{6}+1) = \sqrt{3}+\sqrt{2}-1.$$

把 $\sqrt{3}-\sqrt{6}+1$ 依次乘以第一个括号内的每一项，并利用 $\sqrt{12} = 2\sqrt{3}$ 及 $\sqrt{18} = 3\sqrt{2}$：

$$\begin{aligned}
2(\sqrt{3}-\sqrt{6}+1) &= 2\sqrt{3}-2\sqrt{6}+2,\\
\sqrt{2}(\sqrt{3}-\sqrt{6}+1) &= \sqrt{6}-2\sqrt{3}+\sqrt{2},\\
\sqrt{3}(\sqrt{3}-\sqrt{6}+1) &= 3-3\sqrt{2}+\sqrt{3},\\
\sqrt{6}(\sqrt{3}-\sqrt{6}+1) &= 3\sqrt{2}-6+\sqrt{6}.
\end{aligned}$$

现在把同类项合并。常数项得 $2+3-6=-1$。$\sqrt{2}$ 项得 $1-3+3=1$。$\sqrt{3}$ 项得 $2-2+1=1$。$\sqrt{6}$ 项得 $-2+1+1=0$。总和为 $\sqrt{3}+\sqrt{2}-1$。分母 $\sqrt{3}-\sqrt{6}+1 \approx 0.28$ 不等于零，所以我们可以除以它，于是

$$\frac{\sqrt{3}+\sqrt{2}-1}{\sqrt{3}-\sqrt{6}+1} = 2+\sqrt{2}+\sqrt{3}+\sqrt{6}.$$

? **想一想** 最后一部分应把哪个 x 值代入 $(*)$?

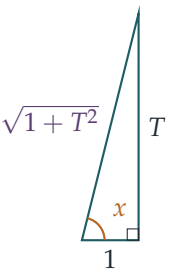
我们需要 $\frac{1}{4}\pi - \frac{1}{2}x = \frac{1}{48}\pi$，所以 $\frac{1}{2}x = \frac{11}{48}\pi$，即 $x = \frac{11}{24}\pi$。这正好是我们在 (i) 和 (ii) 中求得其正切的那个角。

第 (iii) 部分

设 $T = \tan\frac{11}{24}\pi = 2+\sqrt{2}+\sqrt{3}+\sqrt{6}$。在 $(*)$ 中取 $x = \frac{11}{24}\pi$，得

$$\tan\tfrac{1}{48}\pi = \sec\tfrac{11}{24}\pi - T.$$

至于正割（secant），我们有 $\sec^2 x = 1+\tan^2 x$。由于 $0 < \frac{11}{24}\pi < \frac{1}{2}\pi$，余弦为正，所以取正根，$\sec\frac{11}{24}\pi = \sqrt{1+T^2}$。下面的三角形显示的是同一件事。



把 T 平方，四项各自的平方之和为 $4+2+3+6=15$。交叉项之和为

$$2\left(2\sqrt{2}+2\sqrt{3}+2\sqrt{6}+\sqrt{6}+\sqrt{12}+\sqrt{18}\right)=2\left(5\sqrt{2}+4\sqrt{3}+3\sqrt{6}\right)=10\sqrt{2}+8\sqrt{3}+6\sqrt{6}.$$

所以 $1+T^2=16+10\sqrt{2}+8\sqrt{3}+6\sqrt{6}$，因此

$$\tan\tfrac{1}{48}\pi=\sqrt{16+10\sqrt{2}+8\sqrt{3}+6\sqrt{6}}-2-\sqrt{2}-\sqrt{3}-\sqrt{6},$$

正如所求。

验算

- $\sqrt{2}-1\approx 0.41421$，与 $\tan 22.5^\circ$ 相符。
- $2+\sqrt{2}+\sqrt{3}+\sqrt{6}\approx 7.59575$，与 $\tan 82.5^\circ$ 相符。(i) 中的分数也给出相同的值。
- 最后的表达式为 $\sqrt{57.45\ldots}-7.59575\approx 0.0655435$，与 $\tan 3.75^\circ\approx 0.0655435$ 相符。
- 快速检验 $(*)$：当 $x=0$ 时两边均等于 1，而当 $x=\frac{1}{3}\pi$ 时两边均等于 $2-\sqrt{3}=\tan\frac{1}{12}\pi$。

常见错误

- 解 $\frac{1}{4}\pi-\frac{1}{2}x=\frac{1}{48}\pi$ 时出错，误用 $x=\frac{11}{48}\pi$。那个 $\frac{1}{2}$ 必须消去。
- 把四项的根式平方时遗漏交叉项。交叉项共有六个，每个都要乘以二。
- 直接取 $\sec x=\sqrt{1+\tan^2 x}$，却没有说明为何正根才是正确的。

STEP 之外

像 $\tan\frac{\pi}{48}$ 这类精确值，连同当中层层相套的平方根，就是这样靠反复把角减半而逐步构造出来的。这正是一个古典问题在代数上的一面：哪些正多边形可以用尺规作图（ruler and compass construction）作出？高斯（Gauss）证明了答案取决于 $\cos\frac{2\pi}{n}$ 能否只用平方根来表示，而伽罗华理论（Galois theory）则解释了个中原因。域（field）和域扩张

（field extension）是讲述这个故事的语言，若想初步认识，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 2009，第 4 题

纯数学 · 多项式的整除性

STEP II 2009，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/09-s2-q4

关键思路

处理“可被 $(x-1)^5$ 整除（divisible）”这个条件，最好的方法是把它直接写出来：对某个多项式（polynomial）q，有 $\mathrm{p}(x) = 1 + (x-1)^5\mathrm{q}(x)$。求导（differentiate）后可见，$\mathrm{p}(x) - 1$ 的重因式（repeated factor）会保留在 $\mathrm{p}'(x)$ 之中，只是幂次低一次。在最后一部分，这几乎完全确定了 p'，因此我们只需对 p' 积分（integrate）便可求出 p，而不必与十个未知系数（coefficient）苦苦纠缠。

解答

由于 $\mathrm{p}(x) - 1$ 可被 $(x-1)^5$ 整除，而且次数（degree）为 9，我们可以写成

$$\mathrm{p}(x) = 1 + (x-1)^5\,\mathrm{q}(x),$$

其中 q 是一个 4 次多项式。

第 (i) 部分

代入 $x = 1$，第二项便会消失，所以

$$\mathrm{p}(1) = 1.$$

第 (ii) 部分

利用乘积法则（product rule）求导，

$$\begin{aligned} \mathrm{p}'(x) &= 5(x-1)^4\,\mathrm{q}(x) + (x-1)^5\,\mathrm{q}'(x) \\ &= (x-1)^4\big(5\mathrm{q}(x) + (x-1)\mathrm{q}'(x)\big). \end{aligned}$$

括号内是一个多项式，所以 $\mathrm{p}'(x)$ 可被 $(x-1)^4$ 整除，正如题目所求。

第 (iii) 部分

把 $x-1$ 换成 $x+1$，完全相同的论证依然适用：写成 $\mathrm{p}(x) = -1 + (x+1)^5\,\mathrm{r}(x)$，便得 $\mathrm{p}(-1) = -1$，并且可知 $\mathrm{p}'(x)$ 可被 $(x+1)^4$ 整除。

因此 $\mathrm{p}'(x)$ 同时可被 $(x-1)^4$ 和 $(x+1)^4$ 整除。

> ? **想一想** 为甚么由此可知 $\mathrm{p}'(x)$ 可被两者之积 $(x-1)^4(x+1)^4$ 整除?

设 $\mathrm{p}'(x) = (x-1)^4\mathrm{s}(x)$。$(x+1)^4$ 的每一个根（root）都是 -1，而它并不是 $(x-1)^4$ 的根，所以因式 $(x+1)^4$ 必定整除 $\mathrm{s}(x)$。（换句话说，一次因式（linear factor）$x+1$ 在 p' 中出现了四次，而这四个因式没有一个可以被 $(x-1)^4$ 用掉。）因此 $(x-1)^4(x+1)^4 = (x^2-1)^4$ 整除 $\mathrm{p}'(x)$。

现在 p' 的次数是 8，与 $(x^2-1)^4$ 的次数相同，所以商（quotient）是一个非零常数：

$$\mathrm{p}'(x) = k(x^2-1)^4 = k\left(x^8 - 4x^6 + 6x^4 - 4x^2 + 1\right).$$

积分后得

$$\mathrm{p}(x) = k\,\mathrm{F}(x) + c, \qquad \mathrm{F}(x) = \frac{x^9}{9} - \frac{4x^7}{7} + \frac{6x^5}{5} - \frac{4x^3}{3} + x.$$

留意 F 是奇函数（odd function），所以 $\mathrm{F}(-1) = -\mathrm{F}(1)$。由已知的两个函数值可得

$$k\mathrm{F}(1) + c = \mathrm{p}(1) = 1, \qquad -k\mathrm{F}(1) + c = \mathrm{p}(-1) = -1.$$

两式相加，得 $c = 0$；两式相减，得 $k\mathrm{F}(1) = 1$。以 315 为公分母，

$$\mathrm{F}(1) = \frac{35 - 180 + 378 - 420 + 315}{315} = \frac{128}{315}, \qquad \text{所以} \qquad k = \frac{315}{128}.$$

因此

$$\mathrm{p}(x) = \frac{1}{128}\left(35x^9 - 180x^7 + 378x^5 - 420x^3 + 315x\right).$$

我们还应该验证这个多项式确实符合要求，因为到目前为止，我们只证明了任何符合条件的 p 都必定是这个形式。它的次数是 9。此外，$\mathrm{p}(1) - 1 = 0$，而且 $\mathrm{p}'(x) = \frac{315}{128}(x-1)^4(x+1)^4$，所以 $\mathrm{p}(x) - 1$ 的首四阶导数（derivative）在 $x = 1$ 处全部为零；因此 $x = 1$

是 $\mathrm{p}(x)-1$ 的根，其重数（multiplicity）至少为 5，即 $(x-1)^5$ 整除 $\mathrm{p}(x)-1$。由于 p 是奇函数，$\mathrm{p}(x)+1=-\big(\mathrm{p}(-x)-1\big)$，而后者可被 $(-x-1)^5=-(x+1)^5$ 整除。所以两个条件都成立，而这个 p 就是唯一的答案。

验算

- 系数 $35,-180,378,-420,315$ 之和为 128，由此确认 $\mathrm{p}(1)=1$。
- 以准确的分数运算，把 $\mathrm{p}(x)-1$ 反复除以 $x-1$，余数（remainder）恰好有五次为零；把 $\mathrm{p}(x)+1$ 反复除以 $x+1$，结果亦然。
- p 是奇函数，这是必然的：若 p 符合条件，则 $-\mathrm{p}(-x)$ 亦符合条件，而答案是唯一的。

常见错误

- 试图由十条联立方程（simultaneous equations）直接求出 p 的全部十个系数。这在原则上可行，但对 p' 积分要快得多。
- 忘记积分常数（constant of integration），或者没有同时利用 $\mathrm{p}(1)=1$ 和 $\mathrm{p}(-1)=-1$ 来确定积分常数和 k。

STEP 之外

第 (ii) 部分背后的事实是：a 是 f 的 m 重根，当且仅当 $\mathrm{f}(a)=\mathrm{f}'(a)=\cdots=\mathrm{f}^{(m-1)}(a)=0$。像这样在若干点上指定函数值和导数值，称为埃尔米特插值（*Hermite interpolation*）；而这里求得的多项式是一条“smoothstep”曲线，由 -1 上升至 1，并且在两端尽可能平坦，常用于电脑图像（computer graphics）。“互质（coprime）的因式各自整除，则其乘积亦整除”这个论证属于环论（ring theory）：域（field）上的多项式具有唯一分解（unique factorisation）性质，与整数一样。想进一步了解，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）。

详细解答

STEP II 2009，第 5 题

纯数学 · 嵌套平方根

> **STEP II 2009，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/09-s2-q5

关键思路

开首的展开式其实就是整道题的化身：它告诉我们 $x+2\sqrt{x-1}$ 是一个完全平方（perfect square），而用同样的手法可知 $x-2\sqrt{x-1}$ 也是。只要把嵌套根式（nested roots）拆开，每个被积函数（integrand）都会化简成初等的式子。唯一要留神的是 $\sqrt{a^2}=|a|$，而不是 a，所以我们必须留意正负号，而这正是第 (ii) 部分有趣之处。

解答

展开式

当 $x \geq 1$ 时，

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

同理，$(\sqrt{x-1}-1)^2 = x-2\sqrt{x-1}$。取非负平方根，并记住 $\sqrt{a^2}=|a|$，

$$\sqrt{x+2\sqrt{x-1}} = \sqrt{x-1}+1, \qquad \sqrt{x-2\sqrt{x-1}} = \left|\sqrt{x-1}-1\right|.$$

第一式不需要绝对值（modulus），因为 $\sqrt{x-1}+1>0$。第二式则需要：$\sqrt{x-1}-1$ 在 $1 \leq x < 2$ 时为负，在 $x \geq 2$ 时为非负。

第 (i) 部分

在 $[5,10]$ 上，我们有 $\sqrt{x-1} \geq 2 > 1$，所以 $\left|\sqrt{x-1}-1\right| = \sqrt{x-1}-1$。因此分子是

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

而被积函数就只是 2。故此

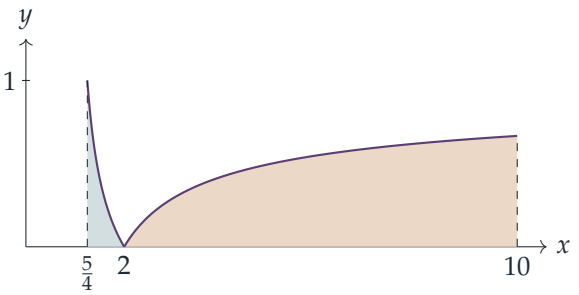
$$\int_5^{10} \frac{\sqrt{x+2\sqrt{x-1}}+\sqrt{x-2\sqrt{x-1}}}{\sqrt{x-1}}\,\mathrm{d}x = \int_5^{10} 2\,\mathrm{d}x = 10.$$

第 (ii) 部分

由展开式可得

$$y = \frac{\left|\sqrt{x-1}-1\right|}{\sqrt{x-1}} = \left|1-\frac{1}{\sqrt{x-1}}\right|.$$

所以 $y \geq 0$ 恒成立，而曲线在 $x=2$ 处与 x-轴相切，该处 $\sqrt{x-1}=1$。



想一想 为甚么我们不能直接去掉平方根，在整个范围上对 $1-\dfrac{1}{\sqrt{x-1}}$ 积分?

因为 $\sqrt{x-2\sqrt{x-1}}$ 永不为负，但 $\sqrt{x-1}-1$ 在 $x<2$ 时为负。忽略绝对值的话，就会把 $[\frac{5}{4},2]$ 上的那一块当作负面积计算。所以我们在 $x=2$ 处把积分分开。

在 $[\frac{5}{4},2]$ 上，$\sqrt{x-1}\leq 1$，所以 $y=\dfrac{1}{\sqrt{x-1}}-1$，而

$$\int_{5/4}^{2}\left(\frac{1}{\sqrt{x-1}}-1\right)\mathrm{d}x = \left[2\sqrt{x-1}-x\right]_{5/4}^{2} = (2-2)-\left(1-\tfrac{5}{4}\right) = \tfrac{1}{4}.$$

在 $[2,10]$ 上，$\sqrt{x-1}\geq 1$，所以 $y=1-\dfrac{1}{\sqrt{x-1}}$，而

$$\int_2^{10}\left(1-\frac{1}{\sqrt{x-1}}\right)\mathrm{d}x = \left[x-2\sqrt{x-1}\right]_2^{10} = (10-6)-(2-2) = 4.$$

因此总面积是 $\frac{1}{4}+4=\frac{17}{4}$。

第 (iii) 部分

分子中第二个根式的形状有所不同，但只要以 $x+1$ 取代 $x-1$，同样的手法仍然奏效：

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

当 $x \geq \frac{5}{4}$ 时，我们有 $\sqrt{x+1} > 1$，所以 $\sqrt{x-2\sqrt{x+1}+2} = \sqrt{x+1}-1$，不需要绝对值。写成 $\sqrt{x^2-1} = \sqrt{x-1}\,\sqrt{x+1}$（由于当 $x > 1$ 时两个因子皆为正，此式成立），被积函数便变成

$$\frac{(\sqrt{x-1}+1)+(\sqrt{x+1}-1)}{\sqrt{x-1}\,\sqrt{x+1}} = \frac{\sqrt{x-1}+\sqrt{x+1}}{\sqrt{x-1}\,\sqrt{x+1}} = \frac{1}{\sqrt{x+1}} + \frac{1}{\sqrt{x-1}}.$$

两个 ± 1 项互相抵消，这正是分子看来有点古怪的用意所在。故此

$$\begin{aligned}\int_{5/4}^{10}\Big(\frac{1}{\sqrt{x+1}} + \frac{1}{\sqrt{x-1}}\Big)\mathrm{d}x &= \Big[2\sqrt{x+1}+2\sqrt{x-1}\Big]_{5/4}^{10} \\ &= \big(2\sqrt{11}+6\big) - \big(3+1\big) \\ &= 2\sqrt{11}+2,\end{aligned}$$

其中在下限处用到 $\sqrt{9/4} = \frac{3}{2}$ 和 $\sqrt{1/4} = \frac{1}{2}$。

验算

- 在 $x = 5$ 处：$\sqrt{5+4} = 3$ 而 $\sqrt{5-4} = 1$，所以 (i) 的被积函数是 $(3+1)/2 = 2$，与所说相符。
- 在 (ii) 中，曲线在 $x = \frac{5}{4}$ 处的高度是 $|1-2| = 1$，在 $x = 10$ 处的高度是 $1 - \frac{1}{3} = \frac{2}{3}$，与草图吻合；第一块是面积为 $\frac{1}{4}$ 的狭长小片，这是合理的。
- 对原来的被积函数使用辛普森法则（Simpson's rule），得出 10.000、4.250 和 8.6332，与 10、$\frac{17}{4}$ 和 $2\sqrt{11}+2 \approx 8.6332$ 一致。

常见错误

- 对所有 x 都写 $\sqrt{x-2\sqrt{x-1}} = \sqrt{x-1}-1$。在 (ii) 中这会得出 $4 - \frac{1}{4} = \frac{15}{4}$，而不是 $\frac{17}{4}$。
- 在 (iii) 中，没有察觉 $x-2\sqrt{x+1}+2$ 也是完全平方，反而尝试用代换法（substitution）。

STEP 之外

恒等式 $\sqrt{a^2} = |a|$ 是分析学（analysis）中一个重大主题的小例子：以公式定义的函数可以暗藏尖角和正负号的转变，在积分之前，你必须先看清楚式子在哪里变号。把 $\sqrt{x+2\sqrt{x-1}}$ 这类根式解除嵌套（denesting），背后也有出奇地深刻的代数理论，与域扩张（field extensions）有关。在分析学方面，Abbott 的 *Understanding Analysis*（《理解分析》）是很好的下一步。

详细解答

STEP II 2009，第 6 题

纯数学 · 无穷级数的界限

STEP II 2009，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/09-s2-q6

关键思路

我们无法准确求出 $\sum 1/F_i$ 的和，但可以把它夹在两个界之间。递推关系（recurrence）给出两个简单的比较：每个斐波那契数（Fibonacci number）都小于前一项的两倍，又大于前两项那一个的两倍。第一个比较用一个等比级数（geometric series）从下方界定级数的尾部（tail）。第二个比较则把尾部与它自身平移后的版本联系起来，从而由上方界定它。在使用任何一个界之前先准确地加起若干项，就可以把估计收紧到我们想要的程度。

解答

把相邻两项逐对相加，得

$$F_3 = 2, \quad F_4 = 3, \quad F_5 = 5, \quad F_6 = 8, \quad F_7 = 13, \quad F_8 = 21, \quad F_9 = 34, \quad F_{10} = 55.$$

所有 F_n 均为正数，而且当 $n \geq 2$ 时 $F_{n+1} = F_n + F_{n-1} > F_n$，所以数列由 F_2 起严格递增（strictly increasing）。

级数收敛。 当 $i \geq 4$ 时，$F_i = F_{i-1} + F_{i-2} > 2F_{i-2}$，因为当 $i - 2 \geq 2$ 时 $F_{i-1} > F_{i-2}$。由于 $F_1 = 1 \geq 2^{-1/2}$ 和 $F_2 = 1 \geq 2^0$，而 $F_3 = 2 \geq 2^{1/2}$，以每次两步的归纳法（induction）可得对每个 i 都有 $F_i \geq 2^{(i-2)/2}$。所以 $0 < 1/F_i \leq (1/\sqrt{2})^{i-2}$，于是与公比为 $1/\sqrt{2} < 1$ 的等比级数比较，可知 S 收敛（converges）。这样我们就可以放心地对下面的和式进行运算。我们还会反复用到一个事实：若对每个 i 都有 $a_i \leq b_i$，并且至少有一个 i 使 $a_i < b_i$，则对收敛级数而言 $\sum a_i < \sum b_i$。

第 (i) 部分

下界。 当 $i \geq 4$ 时，$F_{i-2} < F_{i-1}$，所以

$$F_i = F_{i-1} + F_{i-2} < 2F_{i-1} \qquad \Longrightarrow \qquad \frac{1}{F_i} > \frac{1}{2F_{i-1}} \qquad (i \geq 4),$$

这就是第一个结果。（当 $i = 3$ 时等号成立，即 $F_3 = 2F_2$，所以这个不等式要由 $i = 4$ 开始。）

反复应用这个结果，当 $i \geq 4$ 时，

$$\frac{1}{F_i} > \frac{1}{2F_{i-1}} > \frac{1}{2^2 F_{i-2}} > \cdots > \frac{1}{2^{i-3}F_3} = \frac{1}{2^{i-2}}.$$

对 $i \geq 4$ 求和，

$$\sum_{i=4}^{\infty} \frac{1}{F_i} > \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{2},$$

所以

$$S = 1 + 1 + \tfrac{1}{2} + \sum_{i=4}^{\infty} \frac{1}{F_i} > \tfrac{5}{2} + \tfrac{1}{2} = 3.$$

上界。 现在使用另一个比较：由上面的 $F_i > 2F_{i-2}$，当 $i \geq 4$ 时 $1/F_i < 1/(2F_{i-2})$。我们把它应用于由 $i = 5$ 起的每一项，并设 $j = i - 2$：

$$\sum_{i=5}^{\infty} \frac{1}{F_i} < \frac{1}{2} \sum_{i=5}^{\infty} \frac{1}{F_{i-2}} = \frac{1}{2} \sum_{j=3}^{\infty} \frac{1}{F_j} = \frac{1}{2}(S - 2).$$

左方是 S 减去它的首四项，即 $1 + 1 + \frac{1}{2} + \frac{1}{3} = \frac{17}{6}$。因此

$$S - \tfrac{17}{6} < \tfrac{1}{2}S - 1 \qquad \Longrightarrow \qquad \tfrac{1}{2}S < \tfrac{11}{6} \qquad \Longrightarrow \qquad S < \tfrac{11}{3} = 3\tfrac{2}{3}.$$

> ? **想一想** 为甚么上界的比较要由 $i = 5$ 开始，而不是由 $i = 4$ 开始?

由 $i = 4$ 开始也行得通，但只会得到 $S - \frac{5}{2} < \frac{1}{2}(S - 1)$，即 $S < 4$，这并不足够。在估计之前准确地加起的项愈多，因子 $\frac{1}{2}$ 令我们损失的就愈少。第 (ii) 部分正是靠这个想法来推动。

第 (ii) 部分

记部分和（partial sum）为 $S_k = \sum_{i=1}^{k} 1/F_i$，尾部为 $T_k = \sum_{i=k}^{\infty} 1/F_i$，于是 $S = S_{k-1} + T_k$。第 (i) 部分用了两个一般的界，两者对 $k \geq 3$ 都成立。

下界：与之前完全一样，当 $i > k$ 时 $1/F_i > 1/(2^{i-k}F_k)$，所以

$$T_{k+1} > \frac{1}{F_k}\left(\tfrac{1}{2} + \tfrac{1}{4} + \cdots\right) = \frac{1}{F_k}, \qquad \text{所以} \qquad S > S_k + \frac{1}{F_k}.$$

上界：当 $i \geq k+2 \geq 5$ 时 $1/F_i < 1/(2F_{i-2})$，由此得 $T_{k+2} < \frac{1}{2}T_k$，而 $T_{k+2} = T_k - \frac{1}{F_k} - \frac{1}{F_{k+1}}$，所以

$$T_k < 2\Big(\frac{1}{F_k} + \frac{1}{F_{k+1}}\Big), \qquad \text{所以} \qquad S < S_{k-1} + 2\Big(\frac{1}{F_k} + \frac{1}{F_{k+1}}\Big).$$

取 $k = 3$ 时，这两个界就是第 (i) 部分的 3 和 $3\frac{2}{3}$。现在在两者中都取 $k = 5$。

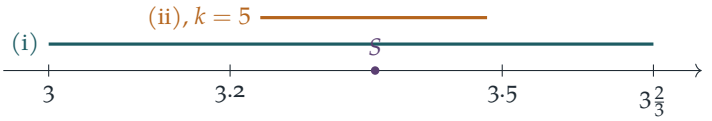
对于下界，$S_5 = 1 + 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} = \frac{91}{30}$，所以

$$S > \frac{91}{30} + \frac{1}{5} = \frac{97}{30} > \frac{96}{30} = 3.2.$$

对于上界，$S_4 = \frac{17}{6}$，而 $F_5 = 5$，$F_6 = 8$，所以

$$S < \frac{17}{6} + 2\Big(\frac{1}{5} + \frac{1}{8}\Big) = \frac{17}{6} + \frac{13}{20} = \frac{170 + 39}{60} = \frac{209}{60} < \frac{210}{60} = 3.5.$$

合起来便得 $3.2 < S < 3.5$，正如所求。（$k = 4$ 并不足够：它只给出 $\frac{19}{6} < S < \frac{107}{30}$，约为 $3.17 < S < 3.57$。）



验算

- 以数值方法把首 200 项加起来，得 $S \approx 3.35989$。这个值落在两个区间之内，而且 $k = 3, \ldots, 6$ 的界全都把它夹在中间：$3 < \frac{19}{6} < \frac{97}{30} < \frac{197}{60} < S < \frac{2681}{780} < \frac{209}{60} < \frac{107}{30} < \frac{11}{3}$。
- 用实际数值检验这些比较：$F_7 = 13 < 2 \times 8$，而 $13 > 2 \times 5$。

常见错误

- 在 $i = 3$ 时使用 $1/F_i > 1/(2F_{i-1})$，但该处其实是等式。这会破坏不等式的严格性，或者得出错误的等比级数和。
- 未先确定级数收敛，就把 S 当作一个数来运算（“$S - \frac{17}{6} < \frac{1}{2}S - 1$”）。
- 因为由 $i = 4$ 开始的比较只给出 $S < 4$，就放弃求上界。应在估计之前先准确地加起更多的项。

STEP 之外

这个值是 $S \approx 3.3598856\ldots$，称为*斐波那契倒数常数*（reciprocal Fibonacci constant）。现已知道它是无理数（irrational；这是 André-Jeannin 在 1989 年的结果），但至今未有

已知的闭合形式（closed form）。这里的比较其实是乔装了的比值审敛法（ratio test）。F_{i+1}/F_i 趋向黄金比（golden ratio）$\varphi = \frac{1}{2}(1+\sqrt{5})$，所以尾部的表现就像一个公比为 $1/\varphi \approx 0.618$ 的等比级数，介乎我们那两个粗略的公比 $\frac{1}{2}$ 与 $1/\sqrt{2}$ 之间。这类比较审敛法（comparison test）和尾部估计，是大学第一门分析（analysis）课程的核心内容。Abbott 的 *Understanding Analysis*（《理解分析》）是正式认识这些内容的绝佳起点。

详细解答

STEP II 2009，第 7 题

纯数学 · 观察法积分

> **STEP II 2009，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/09-s2-q7

关键思路

这类积分乍看无从入手，直到你发现它们恰好是某个函数的导数（derivative）。第一个结果告诉我们 $(x-a)^n \mathrm{e}^{bx}\sqrt{1+x^2}$ 的导数是甚么模样：$(x-a)$ 的幂、指数函数（exponential）、一个三次多项式（cubic），全部除以 $\sqrt{1+x^2}$。所以每个积分（integral）其实都是一道配对练习：读出 n、a 和 b，再检查三次多项式是否吻合。当分子中的多项式是四次（quartic）而非三次时，我们就要弄清楚多出来的那一次从何而来。

解答

导数

对三个因式的乘积求导，并写出 $\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{1+x^2} = \frac{x}{\sqrt{1+x^2}}$：

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x} &= n(x-a)^{n-1}\mathrm{e}^{bx}\sqrt{1+x^2} + b(x-a)^n\mathrm{e}^{bx}\sqrt{1+x^2} + (x-a)^n\mathrm{e}^{bx}\frac{x}{\sqrt{1+x^2}} \\ &= \frac{(x-a)^{n-1}\mathrm{e}^{bx}}{\sqrt{1+x^2}}\Big[\,n(1+x^2) + b(x-a)(1+x^2) + x(x-a)\,\Big],\end{aligned}$$

这里我们抽出了公因式 $(x-a)^{n-1}\mathrm{e}^{bx}/\sqrt{1+x^2}$（把首两项各乘以 $\sqrt{1+x^2}$，使它们置于公分母之上）。展开括号，得

$$q(x) = bx^3 + (n+1-ab)x^2 + (b-a)x + (n-ab).$$

由于 $b \neq 0$，x^3 的系数不为零，所以 q 确实是三次多项式。这就是所要求的结果。我们会以下列形式使用它：

$$\int \frac{(x-a)^{n-1}\mathrm{e}^{bx}\,q(x)}{\sqrt{1+x^2}}\,\mathrm{d}x = (x-a)^n\mathrm{e}^{bx}\sqrt{1+x^2} + C. \qquad (*)$$

第 (i) 部分

配对各个因式，我们需要 $a=4$、$b=4$ 和 $n-1=14$，所以 $n=15$。代入这些数值，

$$q(x)=4x^3+(16-16)x^2+(4-4)x+(15-16)=4x^3-1,$$

这正是被积函数（integrand）中的三次多项式。因此由 $(*)$ 得

$$\int \frac{(x-4)^{14}\mathrm{e}^{4x}(4x^3-1)}{\sqrt{1+x^2}}\,\mathrm{d}x=(x-4)^{15}\mathrm{e}^{4x}\sqrt{1+x^2}+C.$$

第 (ii) 部分

这次的多项式是四次的，所以它不可能直接就是 q。

> ? **想一想** 一个只会产生三次多项式的公式，怎样才能容纳一个四次多项式?

$(*)$ 中 $(x-a)$ 的幂可以由我们选择。如果这个四次多项式有因式 $(x-1)$，我们就可以把它移过去与 $(x-1)^{21}$ 合并，余下一个三次多项式。试代 $x=1$：$12-1-11=0$，所以 $(x-1)$ 确实是因式，除去后得

$$12x^4-x^2-11=(x-1)(12x^3+12x^2+11x+11).$$

（你也可以看出它等于 $(x^2-1)(12x^2+11)$。）因此被积函数是

$$\frac{(x-1)^{22}\mathrm{e}^{12x}(12x^3+12x^2+11x+11)}{\sqrt{1+x^2}}.$$

取 $a=1$、$b=12$、$n=23$：

$$q(x)=12x^3+(24-12)x^2+(12-1)x+(23-12)=12x^3+12x^2+11x+11,$$

两者吻合。故此

$$\int \frac{(x-1)^{21}\mathrm{e}^{12x}(12x^4-x^2-11)}{\sqrt{1+x^2}}\,\mathrm{d}x=(x-1)^{23}\mathrm{e}^{12x}\sqrt{1+x^2}+C.$$

第 (iii) 部分

多项式同样是四次的，但这次代入 $x=2$ 得 $64+8-2=70\neq 0$，所以 $(x-2)$ 不是因式，第 (ii) 部分的技巧行不通。

> ? **想一想** 如果 $(x-2)$ 不是因式，多出来的那一次从何而来?

数一数次数。被积函数的分子次数是 $6+4=10$，但 $(*)$ 配上 $(x-2)^6$ 只能给出次数 $6+3=9$。答案中若多一个因式 x，次数便会增加一，所以不妨试试对 xy 求导，其中 y

是 $a=2$、$b=4$、$n=7$ 时的函数：

$$y=(x-2)^7\mathrm{e}^{4x}\sqrt{1+x^2},\qquad q(x)=4x^3+(8-8)x^2+(4-2)x+(7-8)=4x^3+2x-1.$$

由积法则（product rule），$(xy)'=y+xy'$。把 y 也置于分母 $\sqrt{1+x^2}$ 之上，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(xy)&=\frac{(x-2)^7\mathrm{e}^{4x}(1+x^2)}{\sqrt{1+x^2}}+\frac{x(x-2)^6\mathrm{e}^{4x}(4x^3+2x-1)}{\sqrt{1+x^2}}\\&=\frac{(x-2)^6\mathrm{e}^{4x}}{\sqrt{1+x^2}}\Big[(x-2)(1+x^2)+4x^4+2x^2-x\Big].\end{aligned}$$

由于 $(x-2)(1+x^2)=x^3-2x^2+x-2$，括号内的式子是

$$x^3-2x^2+x-2+4x^4+2x^2-x=4x^4+x^3-2,$$

正是我们想要的四次多项式。因此

$$\int\frac{(x-2)^6\mathrm{e}^{4x}(4x^4+x^3-2)}{\sqrt{1+x^2}}\,\mathrm{d}x=x(x-2)^7\mathrm{e}^{4x}\sqrt{1+x^2}+C.$$

验算

- 对每个答案求导再作比较：这是最确实的验算方法，而每一次都只是再用一次 $(*)$（第 (iii) 部分另加积法则）。
- 快速数次数：对一个多项式次数为 d 的答案求导（不计平方根和指数函数），所得分子的次数是 $d+2$。在 (i) 中，$15+2=14+3$；在 (ii) 中，$23+2=21+4$；在 (iii) 中，$8+2=6+4$。
- 数值上，以森逊法则（Simpson's rule）计算 (iii) 的积分在 $[-1,1.2]$ 上的值，得 $-104.41424\ldots$，与 $F(1.2)-F(-1)$ 吻合至约 13 位有效数字，其中 $F(x)=x(x-2)^7\mathrm{e}^{4x}\sqrt{1+x^2}$；(i) 和 (ii) 同样吻合得很好。

常见错误

- 忘记 n 比被积函数中的幂大一：在 (i) 中，答案含 $(x-4)^{15}$，而不是 $(x-4)^{14}$。
- 在 (ii) 中，试图直接配对四次多项式，然后断定无解，而没有先把因式 $(x-1)$ 并入。
- 漏写积分常数（constant of integration）。

STEP 之外

这里的方法是“辨认式积分”（integration by recognition）：猜出原函数（antiderivative）的形状，对它求导，然后比较系数。把这个想法写成算法，就成了 Risch 算法（Risch algo-

rithm)；电脑代数系统（computer algebra system）用它来判定一个初等函数（elementary function）是否有初等原函数，若有的话便把它求出来。一个相关的想法，是猜一个带有未知系数、具指定形式的解，这就是求解线性微分方程（linear differential equation）的待定系数法（method of undetermined coefficients），见 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）。想找更多 STEP 风格的积分，可参阅 Siklos 的 *Advanced Problems in Mathematics*（进阶数学难题集）。

详细解答

STEP II 2009，第 8 题

纯数学 · 向量几何

STEP II 2009，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/09-s2-q8

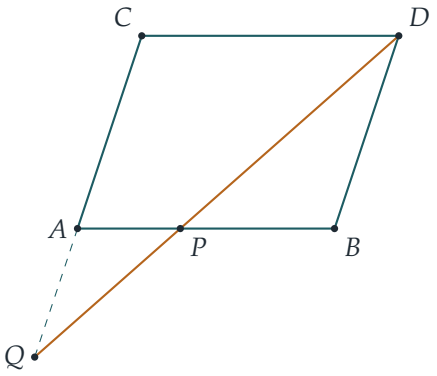
关键思路

把每个定义改写成“从某一点出发，沿着一条边移动若干倍”。于是 $\mathbf{p}-\mathbf{b}=\lambda(\mathbf{a}-\mathbf{b})$ 及 $\mathbf{q}-\mathbf{c}=\mu(\mathbf{a}-\mathbf{c})$，所以长度 BP 和 CQ 不过是 λ AB 和 μ AC，而长度条件就化为一条十分简单的方程。要证明三点共线 (collinear)，只须证明连接它们的其中两个向量 (vector) 平行。

解答

图形

由于 $\mathbf{p}=\mathbf{b}+\lambda(\mathbf{a}-\mathbf{b})$ 且 $0<\lambda<1$，点 P 位于线段 AB 上，严格介乎 A 与 B 之间。由于 $\mathbf{q}=\mathbf{c}+\mu(\mathbf{a}-\mathbf{c})$ 且 $\mu>1$，点 Q 位于直线 CA 上，并在 A 之外：从 C 出发，走到 A，然后继续向前走。下图（以 $\lambda=0.6$ 绘画）亦显示了下文出现的点 D。



求 μ

由上面的改写可得

$$BP = |\mathbf{p}-\mathbf{b}| = \lambda|\mathbf{a}-\mathbf{b}| = \lambda\, AB, \qquad CQ = |\mathbf{q}-\mathbf{c}| = \mu|\mathbf{a}-\mathbf{c}| = \mu\, AC,$$

其中用到 $\lambda > 0$ 和 $\mu > 0$，才可把它们从模（modulus）中抽出。条件 $CQ \times BP = AB \times AC$ 变成 $\lambda\mu\, AB \cdot AC = AB \cdot AC$。由于 A、B、C 是相异的点，$AB \cdot AC \neq 0$，所以

$$\mu = \frac{1}{\lambda}.$$

这确实满足 $\mu > 1$，因为 $0 < \lambda < 1$。

定点 D

设 $\mathbf{d} = -\mathbf{a} + \mathbf{b} + \mathbf{c}$。我们比较 $\overrightarrow{PQ}$ 与 $\overrightarrow{PD}$。首先，

$$\begin{aligned}\mathbf{d} - \mathbf{p} &= -\mathbf{a} + \mathbf{b} + \mathbf{c} - \lambda\mathbf{a} - (1-\lambda)\mathbf{b} \\ &= -(1+\lambda)\mathbf{a} + \lambda\mathbf{b} + \mathbf{c}.\end{aligned}$$

其次，代入 $\mu = 1/\lambda$，

$$\begin{aligned}\mathbf{q} - \mathbf{p} &= \left(\frac{1}{\lambda} - \lambda\right)\mathbf{a} + \left(1 - \frac{1}{\lambda}\right)\mathbf{c} - (1-\lambda)\mathbf{b} \\ &= \frac{1-\lambda}{\lambda}\Big[(1+\lambda)\mathbf{a} - \lambda\mathbf{b} - \mathbf{c}\Big],\end{aligned}$$

其中用到 $\frac{1}{\lambda} - \lambda = \frac{(1-\lambda)(1+\lambda)}{\lambda}$ 及 $1 - \frac{1}{\lambda} = -\frac{1-\lambda}{\lambda}$。

> ? **想一想** 怎样看出这两个向量平行?

方括号内的式子正好是 $-(\mathbf{d}-\mathbf{p})$。所以

$$\mathbf{q} - \mathbf{p} = -\frac{1-\lambda}{\lambda}(\mathbf{d}-\mathbf{p}), \qquad \text{即} \qquad \mathbf{d} = \mathbf{p} - \frac{\lambda}{1-\lambda}(\mathbf{q}-\mathbf{p}).$$

因为 $\lambda \neq 1$，这个系数是有限的；又因为 Q 不在直线 AB 上，所以 $\mathbf{q} \neq \mathbf{p}$。因此 D 位于直线 PQ 上。向量 $\mathbf{d}$ 只涉及 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$，所以对任何 λ 而言这都是同一点，正如题目所求。由于系数 $-\lambda/(1-\lambda)$ 是负数，P 位于 Q 与 D 之间，与图中所见一致。

四边形 $ABDC$

我们有

$$\mathbf{d} - \mathbf{c} = \mathbf{b} - \mathbf{a} \quad \text{及} \quad \mathbf{d} - \mathbf{b} = \mathbf{c} - \mathbf{a},$$

所以 $\overrightarrow{CD} = \overrightarrow{AB}$ 及 $\overrightarrow{BD} = \overrightarrow{AC}$。对边相等且平行，所以 $ABDC$ 是**平行四边形**（parallelogram），而 D 是与 A 相对的顶点。（由于 A、B、C 不共线，这个平行四边形并非退化。）

验算

- 以几何方法覆核：由于 $BD \parallel AC$，三角形 PAQ 与 PBD 相似，所以 $AQ/BD = AP/BP$。这里 $AQ = (\mu-1)AC$、$BD = AC$、$AP = (1-\lambda)AB$ 及 $BP = \lambda AB$，得出 $\mu - 1 = (1-\lambda)/\lambda$，亦即再次得到 $\mu = 1/\lambda$。
- 取 $A = (0,0)$、$B = (4,0)$、$C = (1,3)$ 及 $\lambda = 0.6$（即图中的情况），可得 $P = (1.6,0)$、$Q = (-\frac{2}{3}, -2)$ 及 $D = (5,3)$。QP 与 PD 的斜率同为 $\frac{15}{17}$。
- 我们亦测试了 20,000 个随机抽取的三维三角形。我们以数值方法从长度条件解出 μ，每一次的结果都与 $1/\lambda$ 吻合，误差在 10^{-13} 以内，而且 P、Q、D 共线。

常见错误

- 写成 $BP = (1-\lambda)AB$。在 $\mathbf{p}$ 中 $\mathbf{b}$ 的系数是 $1-\lambda$，但量度与 B 的距离的，却是 A 处的权重。
- 没有检查顶点的次序，就说 $ABDC$ 是“平行四边形”。这里的次序是 $A \to B \to D \to C$，所以 AD 和 BC 是对角线。

STEP 之外

像 $\lambda\mathbf{a} + (1-\lambda)\mathbf{b}$ 这类系数之和为 1 的式子，称为*仿射组合*（affine combination）。它们是描述点（而非向量）的自然语言，而共线是一种仿射性质：它不取决于原点的选择。由此可以进一步认识重心坐标（barycentric coordinates），以及梅涅劳斯定理（Menelaus' theorem）等经典结果；该定理准确描述了一条直线在甚么情况下与三角形的三条边相交。Strang 的 *Introduction to Linear Algebra* 详细阐述向量方面的内容，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多 STEP 风格的向量几何题。

详细解答

STEP II 2009，第 9 题

力学 • 质心

STEP II 2009，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/09-s2-q9

关键思路

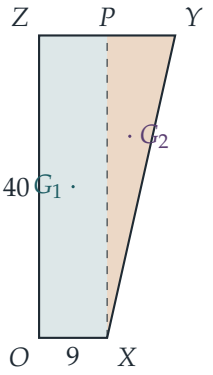
我们需要的每一个距离，都是垂直于同一个固定平面量度的：第 (i) 部分是边 OZ，第 (ii) 部分是水箱的背面，而背面包含 OZ。因此我们由始至终只需要一个坐标。把每个图形分割成质心（centre of mass）已知的若干部分（长方形和一个三角形），对该直线或平面取矩（take moments），再除以总质量。水箱的计算方法相同，只是高了一个层次：它的组成部分是两块梯形（trapezium）和三块长方形的金属薄片，然后再加上一块液体。

解答

第 (i) 部分

取 O 为原点，x 轴沿 OX，y 轴沿 OZ。于是 $X = (9,0)$，$Z = (0,40)$，而且由于 YZ 平行于 OX 且长度为 18，故 $Y = (18,40)$。验算一下：$XY^2 = 9^2 + 40^2 = 1681 = 41^2$，与题目相符。与 OZ 的距离就是 x 坐标。

由 X 作一条平行于 OZ 的直线，与 ZY 相交于 $P = (9,40)$。这条线把薄片（lamina）分割成长方形 $OXPZ$ 和直角三角形 XYP。



设薄片每 cm^2 的质量为 1，则质量等于面积。

- 长方形 $OXPZ$：面积为 $9 \times 40 = 360$，中心 G_1 位于 $x = 4.5$。
- 三角形 XYP：面积为 $\frac{1}{2} \times 9 \times 40 = 180$。均匀（uniform）三角形的质心位于三个顶点的平均位置，所以 G_2 的 $x = \frac{1}{3}(9 + 18 + 9) = 12$。

整块薄片的面积为 540。对 OZ 取矩，

$$540\,\bar{x} = 360 \times 4.5 + 180 \times 12 = 1620 + 2160 = 3780,$$

所以 $\bar{x} = 7$。质心与 OZ 的距离为 7 cm，正如所求。

第 (ii) 部分

现在设金属薄片每 cm^2 的质量为 1。背面 $OTWZ$ 位于平面 $x = 0$ 上，而由于阔度 d 的方向垂直于两个端面，水箱上每一点的 x 坐标都与第 (i) 部分相同。因此我们对背面取矩。每个长方形的面都是均匀的，所以其质心位于其中心，而该中心的 x 坐标与梯形对应边的中点相同。

组成部分	面积（cm^2）	中心的 x（cm）	矩
两个端面	$2 \times 540 = 1080$	7	7560
底面 $OXUT$	$9d$	4.5	$40.5d$
背面 $OTWZ$	$40d$	0	0
正面 $XUVY$	$41d$	$\frac{1}{2}(9 + 18) = 13.5$	$553.5d$
总计	$1080 + 90d$		$7560 + 594d$

? **想一想** 倾斜的正面，中心在哪里？

正面是一个长方形，其中一对边是 XU 和 YV，分别位于 $x = 9$ 和 $x = 18$。它的中心在两者的正中间，即 $x = 13.5$。计算它的面积时，用的是斜边长度 $XY = 41$，而不是水平阔度 9。

所以空水箱的

$$\bar{x} = \frac{7560 + 594d}{1080 + 90d}.$$

每个系数都可被 18 整除：$7560 = 18 \times 420$，$594 = 18 \times 33$，$1080 = 18 \times 60$，$90 = 18 \times 5$。因此

$$\bar{x} = \frac{420 + 33d}{60 + 5d} = \frac{3(140 + 11d)}{5(12 + d)} \text{ cm},$$

这就是所求的与背面的距离。

注满的水箱

液体注满一个棱柱（prism），其横切面（cross-section）是第 (i) 部分的梯形，长度为 d。它的体积是 $540d$，所以按我们的单位，它的质量是 $540dk$（每单位体积的质量是金属每单位面积质量的 k 倍，而后者我们已取为 1）。每个横切面都是同一个梯形，所以液体的质心位于 $x = 7$，与第 (i) 部分相同。

当 $d = 20$ 时，水箱的质量为 $1080 + 1800 = 2880$，对背面的矩为 $7560 + 594 \times 20 = 19440$。液体的质量为 $10800k$，矩为 $75600k$。所以注满的水箱有

$$\bar{x} = \frac{19440 + 75600k}{2880 + 10800k} = \frac{27 + 105k}{4 + 15k} \text{ cm},$$

其中分子和分母已同时除以 720。这就是注满的水箱的质心与背面的距离。

验算

- 当 $k = 0$（没有液体，或液体的密度可忽略不计）时，答案是 $\frac{27}{4} = 6.75$，而空水箱的公式在 $d = 20$ 时同样得出 $\frac{3 \times 360}{5 \times 32} = 6.75$。
- 当 $k \to \infty$ 时，液体占主导地位，$\bar{x} \to \frac{105}{15} = 7$，即液体本身的质心。
- 答案总是介乎 6.75 与 7 之间，因为它是两者的加权平均（weighted average），理应如此。例如 $k = 1$ 时得出 $\frac{132}{19} \approx 6.947$。
- 数值验算（以鞋带公式（shoelace formula）求梯形的形心（centroid）、以蒙地卡罗（Monte Carlo）方法估算得出 $\bar{x} \approx 6.9994$，以及用精确分数计算两条水箱公式）与以上所有结果吻合。

常见错误

- 计算正面的面积时用了 $9d$ 而不是 $41d$：这个面是倾斜的，所以它的阔度是斜边长度 XY。
- 因为背面的矩为零而把它漏掉。它仍然为总质量贡献 $40d$。
- 只计算一个端面，或者给水箱加上盖。

STEP 之外

这里我们把图形分割成中心已知的若干部分，从而求出质心。在大学里，薄片或立体的质心是用积分来定义的，例如 $\bar{x} = \frac{1}{M}\iint x\,\rho\,\mathrm{d}A$，而“分割成若干部分”的方法，不过是“把区域切成几部分后，积分可以逐部分相加”这个事实而已。同样的积分还可给出转动惯量（moment of inertia），它决定物体如何转动。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对这两个概念都有严谨细致的阐述。

详细解答

STEP II 2009，第 10 题

力学 · 碰撞与恢复系数

STEP II 2009，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/09-s2-q10

关键思路

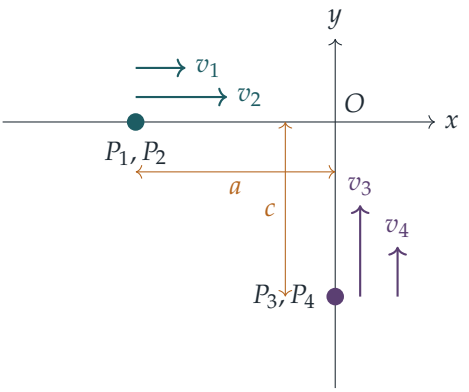
全题共有四次碰撞（collision），但只有头两次需要用到力学：它们决定了碰撞刚结束时的四个速率。其余全是时间上的计算。两个质点（particle）沿不同的轴运动而“在 O 碰撞”，其实就是在同一瞬间到达 O，所以其后每一次碰撞都给出一条形如“距离 / 速率 = 距离 / 速率”的方程。把其中一条时间方程除以另一条，未知的距离便会消去，剩下一个关于质量的条件。

解答

设置

我们把四件物体视为在光滑轴上运动的质点，所以在两次碰撞之间，每个质点都以恒定速度运动。设 P_2 起初位于负 x 轴上与 O 相距 a 之处，P_3 起初位于负 y 轴上与 O 相距 c 之处。我们须证明 $a = c$。

取 $t = 0$ 为头两次碰撞（同时发生）的一刻。这两次碰撞发生在 P_2 和 P_3 原本所在的位置，所以在 $t = 0$ 时，质点 P_1 和 P_2 与 O 的距离都是 a，而 P_3 和 P_4 与 O 的距离都是 c。



头两次碰撞

假设一个质量为 M、以速率 u 运动的质点正面撞向一个质量为 m 的静止质点，恢复系数（coefficient of restitution）为 e，而碰撞后两者沿原来运动方向的速度分别为 V 和 w。由动量守恒（conservation of momentum）和牛顿实验定律（Newton's experimental law）可得

$$Mu = MV + mw, \qquad w - V = eu,$$

所以

$$w = \frac{M(1+e)u}{M+m}, \qquad V = \frac{(M-em)u}{M+m}.$$

把这个结果应用于 P_1 撞向 P_2，以及 P_4 撞向 P_3，可知在 $t = 0$ 刚过后，各质点朝向 O 的速度为

$$v_1 = \frac{(m_1 - em_2)u}{m_1 + m_2}, \qquad v_2 = \frac{m_1(1+e)u}{m_1 + m_2},$$
$$v_4 = \frac{(m_4 - em_3)u}{m_3 + m_4}, \qquad v_3 = \frac{m_4(1+e)u}{m_3 + m_4}.$$

显然 $v_2, v_3 > 0$。由于 P_1 和 P_4 其后确实到达 O，它们必定仍在朝着它运动，所以同样有 $v_1 > 0$ 和 $v_4 > 0$。

利用其后的碰撞

想一想 两个质点分别在互相垂直的轴上，怎样才能“碰撞”？

两条轴唯一的公共点是 O。所以“P_2 和 P_3 在 O 碰撞”的意思不过是它们同时到达 O，P_1 和 P_4 亦然。由 $t = 0$ 至到达 O 为止，每个质点都以恒定速率运动（题目告诉我们，任何

一个质点所涉及的下一次碰撞都在 O 发生），所以到达时间就是距离除以速率：

$$\frac{a}{v_2} = \frac{c}{v_3} \qquad (P_2 \text{ 与 } P_3 \text{ 相遇于 } O), \tag{1}$$

$$\frac{a}{v_1} = \frac{c}{v_4} \qquad (P_1 \text{ 与 } P_4 \text{ 相遇于 } O). \tag{2}$$

把 (1) 除以 (2)，距离便会约去：

$$\frac{v_1}{v_2} = \frac{v_4}{v_3}.$$

根据上面的公式，每个比之中的因子 u 和 $m_1 + m_2$（或 $m_3 + m_4$）都会约去，所以上式即是

$$\frac{m_1 - em_2}{m_1(1+e)} = \frac{m_4 - em_3}{m_4(1+e)},$$

两边同乘以 $m_1 m_4(1+e)$，得

$$m_4(m_1 - em_2) = m_1(m_4 - em_3), \qquad \text{所以} \qquad e\,m_2 m_4 = e\,m_1 m_3.$$

由于 $e > 0$，我们可以除以 e，于是 $m_1 m_3 = m_2 m_4$，即

$$\frac{m_1}{m_2} = \frac{m_4}{m_3} = k, \text{ 姑且如此记之。}$$

现在回到 v_2 和 v_3。把分子和分母分别同除以 m_2 和 m_3，得

$$v_2 = \frac{k(1+e)u}{k+1} = v_3.$$

所以 P_2 和 P_3 在头两次碰撞后以相等的速率离开，于是由 (1) 得 $a = c$。因此 P_2 和 P_3 起初与 O 等距，这正是所要证明的。

验算

- 取 $m_1 = 2$、$m_2 = 1$、$m_3 = 3$、$m_4 = 6$（所以 $k = 2$）、$e = \frac{1}{2}$、$u = 1$ 及 $a = c = 1.7$。那么 $v_2 = v_3 = 1$ 而 $v_1 = v_4 = \frac{1}{2}$：P_2 和 P_3 在 $t = 1.7$ 时一同到达 O，而 P_1 和 P_4 在 $t = 3.4$ 时一同到达，正如题目所说是在较后的时间。逐步模拟的结果与此吻合。
- 留意 $v_2 - v_1 = eu > 0$，所以 P_1 确实在 P_2 之后才到达，与“一段时间之后”相符。
- 若 $e = 0$，这个论证便会失效，而这是理所当然的：此时 $v_1 = v_2$ 且 $v_3 = v_4$，所以 (2) 与 (1) 完全相同，任何满足 $a/c = v_2/v_3$ 的 a 和 c 都可以。这正是题目坚持 $e > 0$ 的原因。

常见错误

- 从 P_1 的起点量度它的路程。在 $t = 0$ 之后，它是从 P_2 原本所在的位置出发，与 O 相距 a，这就是为甚么同一个 a 会同时出现在 (1) 和 (2) 之中。

- 不加说明便除以 e。$e = 0$ 的情况确实有所不同。
- 试图分析 P_2 与 P_3 在 O 的碰撞。我们完全不需要知道这次碰撞的结果，只需要知道它们同时到达那里。

STEP 之外

质量条件 $m_1 m_3 = m_2 m_4$ 说明这两次碰撞是彼此按比例缩放的翻版：正面碰撞后的速度只取决于质量比和 e，这让我们初次窥见量纲分析（dimensional analysis）与相似性（similarity）。在大学里，碰撞问题放在质心参考系（centre-of-mass frame）中处理最为简洁，在那里恢复系数只是把相对速度（relative velocity）按比例缩放。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）正是以这种方式阐述动量与碰撞。

详细解答

STEP II 2009，第 11 题

力学 · 功率与阻力

STEP II 2009，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/09-s2-q11

关键思路

把整列火车视为一个质量为 $(n+1)M$、总阻力为 $(n+1)R$ 的物体。在功率（power）为 P 时，速率为 v 时的驱动力是 P/v，所以由牛顿第二定律（Newton's second law）可立即得出加速度。要求出达到某个速率所需的时间，可把加速度写成 $\mathrm{d}v/\mathrm{d}t$，分离变量（separate the variables）后积分。最后一部分是能量收支：引擎所作的功转化为动能（kinetic energy），以及克服阻力所作的功。

解答

把引擎和各节车卡看成一起运动的质点（particle）（车钩保持拉紧且为刚性），于是整列火车是单一物体。设

$$m = (n+1)M, \qquad k = (n+1)R$$

分别为它的总质量和总阻力。

加速度

功率等于力乘以速率，所以在最大功率下、速率为 v 时，引擎的牵引力（tractive force）为 P/v。沿路轨方向分解，对整列火车应用牛顿第二定律，得

$$ma = \frac{P}{v} - k, \qquad \text{所以} \qquad a = \frac{P-(n+1)Rv}{(n+1)Mv}.$$

达到速率 V 所需的时间

设 $a = \dfrac{\mathrm{d}v}{\mathrm{d}t}$，然后分离变量：

$$\frac{\mathrm{d}t}{\mathrm{d}v} = \frac{mv}{P - kv}, \qquad T = m\int_0^V \frac{v}{P-kv}\,\mathrm{d}v.$$

要计算这个积分，可把分子用分母表示，即 $v = \frac{1}{k}\big(P - (P - kv)\big)$，从而把分式拆开：

$$\frac{v}{P-kv} = -\frac{1}{k} + \frac{P}{k}\cdot\frac{1}{P-kv}.$$

因此

$$T = m\Big[-\frac{v}{k} - \frac{P}{k^2}\ln(P-kv)\Big]_0^V = -\frac{mV}{k} - \frac{mP}{k^2}\ln\Big(1 - \frac{kV}{P}\Big).$$

由于 $m/k = M/R$ 及 $m/k^2 = M/\big((n+1)R^2\big)$，

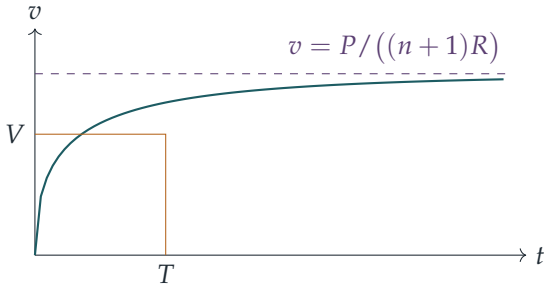
$$T = -\frac{MV}{R} - \frac{MP}{(n+1)R^2}\ln\Big(1 - \frac{(n+1)RV}{P}\Big).$$

想一想 为甚么 P 必须大于 $(n+1)RV$，而这不仅仅是令对数有意义的条件?

从物理上看，当 $v < P/k$ 时，加速度 $(P-kv)/(mv)$ 为正，而在 $v = P/k$ 时加速度会变为零。火车由 $v = 0 < P/k$ 出发并加速，但随着 v 趋近 P/k，加速度逐渐消失。公式显示火车永远到不了这个速率：当 $V \to P/k$（由下方趋近）时，$\ln(1 - kV/P) \to -\infty$，所以 $T \to \infty$。因此 P/k 是终端速率（terminal speed），火车会不断趋近它，却无法在有限时间内达到，而等于或高于它的速率则永远达不到。所以，只有当 $V < P/k$ 时才能达到速率 V，即

$$P > (n+1)RV,$$

这正是所需的结果。（等价地说，$P \le kV$ 意味着引擎即使全力输出，也无法在速率 V 时抵消阻力。）



（此草图只属定性描绘：起初驱动力 P/v 很大，速率急升，其后逐渐趋平，接近终端速率。）

第 (i) 部分

设 $x = \dfrac{(n+1)RV}{P} = \dfrac{kV}{P}$，它是一个很小的量。于是 $\ln(1-x) \approx -x - \frac{1}{2}x^2$，再利用 $V = Px/k$，得

$$T = -\frac{mV}{k} - \frac{mP}{k^2}\ln(1-x) \approx -\frac{mP}{k^2}x + \frac{mP}{k^2}\left(x + \tfrac{1}{2}x^2\right) = \frac{mP}{2k^2}x^2 = \frac{mP}{2k^2}\cdot\frac{k^2V^2}{P^2} = \frac{mV^2}{2P}.$$

一阶项刚好完全抵消，这就是我们需要 x^2 项的原因。因此

$$PT \approx \tfrac{1}{2}(n+1)MV^2.$$

诠释。PT 是引擎在时间 T 内所作的功，而 $\frac{1}{2}(n+1)MV^2$ 是火车所获得的动能。所以当 $(n+1)RV$ 远小于 P 时（阻力乘以速率远小于引擎的功率），引擎所作的功几乎全部转化为动能，只有极少耗于阻力：火车的表现就如同没有阻力一样。

第 (ii) 部分

在时间 T 内，引擎以恒定功率 P 运作，所作的功为 PT。在距离 X 内，恒定的总阻力 $(n+1)R$ 所作的功为 $-(n+1)RX$；换言之，火车克服阻力作了 $(n+1)RX$ 的功。火车由静止出发，最后速率为 V，而路轨是水平的，所以势能（potential energy）没有改变。根据功能原理（work–energy principle），

$$PT = \tfrac{1}{2}(n+1)MV^2 + (n+1)RX.$$

移项整理，得

$$X = \frac{2PT - (n+1)MV^2}{2(n+1)R},$$

这正是所需的结果。这与第 (i) 部分吻合：若阻力可以忽略，$(n+1)RX$ 一项便会消失，剩下 $PT \approx \frac{1}{2}(n+1)MV^2$。

验算

- 取 $n = 3$、$M = 2$、$R = 1$、$P = 40$、$V = 6$（于是 $(n+1)RV = 24 < 40$）。公式给出 $T \approx 6.3258$，继而 $X \approx 27.258$。以数值方法把运动方程逐步向前推算，所得数值准确至四位有效数字都相同。
- 量纲（dimensions）：MV/R 是（质量）（速率）/（力），即时间，MP/R^2 亦然；对数则是无量纲的。
- 当 $V \to P/\big((n+1)R\big)$ 时，T 无止境地增大（以上述数值计算，在终端速率的 90%、99%、99.9% 时，$T \approx 28$、72、118），这印证了该条件。
- 当 x 很小时，比值 $PT/\frac{1}{2}(n+1)MV^2$ 约为 $1 + \frac{2}{3}x$，趋向 1，与第 (i) 部分的预测一致。

常见错误

- 只用引擎本身的质量和阻力，而不是整列火车的：引擎要拉动全部 $n+1$ 节车辆。
- 题目要求的是时间，却写成 $a = \mathrm{d}v/\mathrm{d}x$，或使用 $v\,\mathrm{d}v/\mathrm{d}x$。这条路得出的是距离，而不是时间。
- 在第 (i) 部分只把对数展开至一阶：x 项互相抵消，结果得到 $T \approx 0$，答案就此丢失。

STEP 之外

方程 $m\,\mathrm{d}v/\mathrm{d}t = P/v - k$ 是一阶非线性常微分方程（ODE），而且恰好是可分离变量的；终端速率 P/k 则是一个稳定平衡点（stable equilibrium），解会不断趋近它，却无法在有限时间内到达。这种现象在光滑常微分方程的平衡点附近十分典型，由解的唯一性（uniqueness of solutions）即可看出。第 (ii) 部分就是功能定理（work–energy theorem），它在大学力学中会成为贯穿全局的核心原理。想进一步了解，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门），至于常微分方程方面，可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）。

详细解答

STEP II 2009，第 12 题

概率与统计 · 半正态分布

STEP II 2009，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/09-s2-q12

关键思路

指数 $-2x^2$ 其实是一个乔装了的正态分布密度函数（normal density）。把它写成 $-2x^2 = -\dfrac{x^2}{2\sigma^2}$，便得 $\sigma = \frac{1}{2}$，所以 f 恰好是 $Y \sim \mathrm{N}(0, \frac{1}{4})$ 的密度函数的两倍，并在 0 处截断。换句话说，X 与 $|Y|$ 有相同的分布。一旦看出这一点，归一化常数（normalising constant）、$\mathrm{E}(X^2)$ 和中位数（median）全都可以直接由正态分布的性质得出。只有 $\mathrm{E}(X)$ 需要计算一个（非常容易的）积分。

解答

以下一律设 $Y \sim \mathrm{N}(0, \frac{1}{4})$ 及 $Z \sim \mathrm{N}(0,1)$，于是 Y 与 $\frac{1}{2}Z$ 有相同的分布，而它的密度函数为

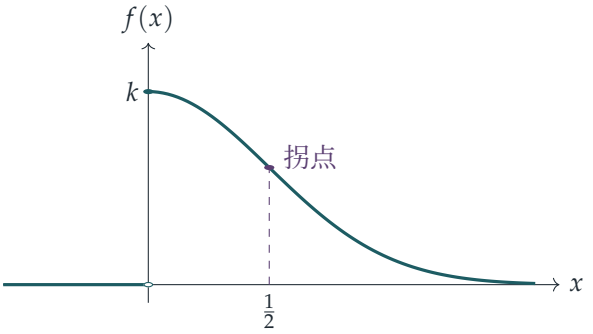
$$g(y) = \frac{1}{\frac{1}{2}\sqrt{2\pi}}\,\mathrm{e}^{-y^2/(2\cdot\frac{1}{4})} = \frac{2}{\sqrt{2\pi}}\,\mathrm{e}^{-2y^2}.$$

第 (i) 部分

当 $x \geq 0$ 时，$f'(x) = -4kx\,\mathrm{e}^{-2x^2} \leq 0$，所以 f 由 $x = 0$ 处的最大值 k 开始递减，而该处的斜率为零。另外

$$f''(x) = k(16x^2 - 4)\,\mathrm{e}^{-2x^2},$$

它在 $x = \frac{1}{2}$ 处变号，所以在 $x = \frac{1}{2}$ 处有一个拐点（point of inflection）。当 $x \to \infty$ 时，$f(x) \to 0$，而且趋近得非常快；当 $x < 0$ 时，$f(x) = 0$。图像是钟形曲线的右半部分（由第 (ii) 部分可知 $k \approx 1.60$），并在 $x = 0$ 处由 0 跳升至 k。



第 (ii) 部分

总概率必须等于 1。由于 g 关于 0 对称，而且在整条实数线上的积分为 1，所以它在 $[0,\infty)$ 上的积分为 $\frac{1}{2}$：

$$\int_0^\infty \frac{2}{\sqrt{2\pi}}\,\mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{2} \quad\Longrightarrow\quad \int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{\sqrt{2\pi}}{4}.$$

因此 $k\cdot\dfrac{\sqrt{2\pi}}{4} = 1$，得

$$k = \frac{4}{\sqrt{2\pi}} = 2\sqrt{\frac{2}{\pi}} \approx 1.596.$$

留意当 $x \geq 0$ 时 $f(x) = 2g(x)$，这印证了 X 的分布与 $|Y|$ 相同。

第 (iii) 部分

求平均值时，因子 x 令积分变得十分简单，因为 $x\,\mathrm{e}^{-2x^2}$ 是 e^{-2x^2} 的导数的常数倍：

$$\mathrm{E}(X) = k\int_0^\infty x\,\mathrm{e}^{-2x^2}\,\mathrm{d}x = k\Big[-\tfrac{1}{4}\mathrm{e}^{-2x^2}\Big]_0^\infty = \frac{k}{4} = \frac{1}{\sqrt{2\pi}} \approx 0.399.$$

想一想 怎样才能不用冗长的分部积分法（integration by parts）而求出 $\displaystyle\int_0^\infty x^2\,\mathrm{e}^{-2x^2}\,\mathrm{d}x$?

再次利用正态分布。函数 $x^2g(x)$ 是偶函数（even function），所以

$$\mathrm{E}(X^2) = \int_0^\infty x^2\cdot 2g(x)\,\mathrm{d}x = \int_{-\infty}^\infty x^2 g(x)\,\mathrm{d}x = \mathrm{E}(Y^2) = \mathrm{Var}(Y) = \frac{1}{4},$$

因为 $\mathrm{E}(Y) = 0$。（取 $u = x$、$\mathrm{d}v = x\mathrm{e}^{-2x^2}\,\mathrm{d}x$ 作分部积分，亦得相同结果：$\int_0^\infty x^2\mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{4}\int_0^\infty \mathrm{e}^{-2x^2}\,\mathrm{d}x = \frac{1}{4k}$。）因此

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = \frac{1}{4} - \frac{1}{2\pi} = \frac{\pi-2}{4\pi} \approx 0.0908.$$

第 (iv) 部分

中位数 m 满足 $\mathrm{P}(X \le m) = \frac{1}{2}$。由于 X 的分布与 $|Y|$ 相同，而 Y 的分布与 $\frac{1}{2}Z$ 相同，

$$\mathrm{P}(X \le m) = \mathrm{P}(|Z| \le 2m) = 2\Phi(2m) - 1,$$

这里用到了标准正态分布（standard normal distribution）的对称性。（直接计算：$\mathrm{P}(X \le m) = \int_0^m 2g(x)\,\mathrm{d}x = 2\big(\mathrm{P}(Y \le m) - \frac{1}{2}\big)$。）令上式等于 $\frac{1}{2}$，得

$$\Phi(2m) = \frac{3}{4}.$$

查表得 $\Phi^{-1}(0.75) = 0.6745$，所以 $2m = 0.6745$，而

$$m \approx 0.337 \text{ （准确至三位有效数字）}.$$

验算

取 $k = 2\sqrt{2/\pi}$，以数值方法（辛普森法则，Simpson's rule）在 $[0, 10]$ 上对 f 积分，得总概率为 1.0000、$\mathrm{E}(X) = 0.39894$ 及 $\mathrm{Var}(X) = 0.090845$，与 $1/\sqrt{2\pi}$ 和 $(\pi - 2)/(4\pi)$ 吻合。以数值方法解 $\int_0^m f(x)\,\mathrm{d}x = \frac{1}{2}$，得 $m = 0.33724$。以 200 000 个标准差（standard deviation）为 $\frac{1}{2}$ 的正态随机数模拟 $|Y|$，得平均值 0.399 及方差（variance）0.091，并有 50.0% 的数值小于 0.337。合理性检查：中位数 0.337 小于平均值 0.399，对于一个右尾较长的分布来说，理应如此。

常见错误

- 从指数误取 $\sigma^2 = 2$ 或 $\sigma = \frac{1}{4}$：应把 $2x^2$ 与 $x^2/(2\sigma^2)$ 对照，得 $\sigma^2 = \frac{1}{4}$。
- 忘记因子 2：在半条实数线上，正态密度函数的积分只有 $\frac{1}{2}$。
- 解 $\Phi(2m) = \frac{1}{2}$，结果得出 $m = 0$。X 在 m 以下的那一半概率，对应的是 Y 中间的那一半概率，所以条件应为 $\Phi(2m) = \frac{3}{4}$。

STEP 之外

X 是一个半正态（half-normal）随机变量（random variable），而当 Y 服从正态分布时，$|Y|$ 就是随机变量的变换（transformation of a random variable）最简单的例子。一般而言，若 $X = h(Y)$，则 X 的密度函数可由 Y 的密度函数对所有原像（preimage）求和而得，这正是此处因子 2 的由来。同样的想法可以证明 Z^2 服从自由度（degree of freedom）为一的卡方分布（chi-squared distribution），而这正是许多统计学内容的起点。Grimmett

and Welsh, *Probability: An Introduction*（概率入门教材）对随机变量的函数有仔细的论述。

详细解答

STEP II 2009，第 13 题

概率与统计 · 二项分布的期望成本

STEP II 2009，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/09-s2-q13

关键思路

发生故障的引擎数目服从二项分布 (binomial distribution)，所以维修费用是一个二项随机变量(random variable)的简单函数，而其期望值(expectation)只是一个简短的和。要得出简洁的形式 $(*)$，窍门是不要把“全部更换”的情况逐一相加：改用补集 (complement)，然后利用因式分解 $1-q^n = p(1+q+\cdots+q^{n-1})$ 抽出因子 p。

解答

我们按照题目的原意，假设各引擎是否发生故障互相独立（independent），每个引擎发生故障的概率均为 p，并记 $q = 1-p$。所有费用一律以 K 为单位计算，直到最后为止。

Andover 型

设 X 为 Andover 型引擎中发生故障的数目，则 $X \sim \mathrm{B}(4, p)$；又设 C_A 为维修费用。于是

$$C_A = \begin{cases} 0 & X = 0, \\ K & X = 1, \\ 4K & X \geq 2. \end{cases}$$

现在 $\mathrm{P}(X=0) = q^4$ 及 $\mathrm{P}(X=1) = 4pq^3$，所以 $\mathrm{P}(X \geq 2) = 1 - q^4 - 4pq^3$。因此

$$\begin{aligned} \mathrm{E}(C_A) &= K \cdot 4pq^3 + 4K\left(1 - q^4 - 4pq^3\right) \\ &= 4K(1-q^4) - 12Kpq^3. \end{aligned}$$

由于 $1-q^4=(1-q)(1+q+q^2+q^3)=p(1+q+q^2+q^3)$，

$$\begin{aligned}\mathrm{E}(C_A)&=4Kp\big(1+q+q^2+q^3\big)-4Kp\cdot 3q^3\\&=4Kp\big(1+q+q^2-2q^3\big),\end{aligned}$$

这正是所要求的 $(*)$。

Basingstoke 型

设 $Y\sim \mathrm{B}(6,p)$ 为 Basingstoke 型引擎中发生故障的数目，C_B 为维修费用。则当 $Y\le 2$ 时 $C_B=YK$，而当 $Y\ge 3$ 时 $C_B=6K$。由

$$\mathrm{P}(Y=0)=q^6,\qquad \mathrm{P}(Y=1)=6pq^5,\qquad \mathrm{P}(Y=2)=15p^2q^4,$$

可得

$$\begin{aligned}\mathrm{E}(C_B)&=K\cdot 6pq^5+2K\cdot 15p^2q^4+6K\big(1-q^6-6pq^5-15p^2q^4\big)\\&=6K(1-q^6)-30Kpq^5-60Kp^2q^4.\end{aligned}$$

利用 $1-q^6=p(1+q+q^2+q^3+q^4+q^5)$ 并抽出 $6Kp$，

$$\mathrm{E}(C_B)=6Kp\big(1+q+q^2+q^3+q^4+q^5-5q^5-10pq^4\big).$$

> **想一想** 括号内仍然有一个 p。怎样才能把它化成“与 $(*)$ 相似的形式”?

形式 $(*)$ 是因子 p 乘以一个只含 q 的多项式（polynomial），所以我们把余下的 p 换成 $1-q$：$10pq^4=10q^4-10q^5$。合并各项后，

$$\boxed{\mathrm{E}(C_B)=6Kp\big(1+q+q^2+q^3-9q^4+6q^5\big).}$$

Andover 型的费用何时是 Basingstoke 型费用的 $\frac{2}{3}$?

我们需要 $\mathrm{E}(C_A)=\frac{2}{3}\mathrm{E}(C_B)$，即

$$4Kp\big(1+q+q^2-2q^3\big)=4Kp\big(1+q+q^2+q^3-9q^4+6q^5\big).$$

两边都有因子 $4Kp$，而 $K>0$，所以此式成立当且仅当 $p=0$ 或

$$-2q^3=q^3-9q^4+6q^5\iff 3q^3-9q^4+6q^5=0\iff 3q^3(1-q)(1-2q)=0.$$

所以 $q=0$、$q=1$ 或 $q=\frac{1}{2}$，即 $p=1$、$p=0$ 或 $p=\frac{1}{2}$。三个值都在 $[0,1]$ 之内，故所求的值为

$$p=0,\qquad p=\tfrac{1}{2},\qquad p=1.$$

$p=0$ 这个值最乏味：引擎永不发生故障，两个期望费用都是零。当 $p=1$ 时，每个引擎都发生故障，费用分别为 $4K$ 和 $6K$。真正有趣的值是 $p=\frac{1}{2}$。

验算

当 $p = \frac{1}{2}$ 时，$(*)$ 给出 $\mathrm{E}(C_A) = 2K\left(1 + \frac{1}{2} + \frac{1}{4} - \frac{1}{4}\right) = 3K$，而 Basingstoke 型的公式给出 $\mathrm{E}(C_B) = 3K\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} - \frac{9}{16} + \frac{6}{32}\right) = \frac{9}{2}K$，果然 $3 = \frac{2}{3} \cdot \frac{9}{2}$。我们亦取了 200 个有理数 p 值，把两条公式与直接按二项分布逐一列举所得的结果作精确比对，并以模拟方式验证：当 $p = 0.3$ 时，2×10^5 次模拟发射所得的平均费用为 $1.811K$ 和 $2.483K$，而公式所得的则是 $1.805K$ 和 $2.485K$。

常见错误

- 当有两个或以上 Andover 型引擎发生故障时，把费用算作 XK 而不是 $4K$：一旦超过门槛，*所有*引擎都要更换，包括完好的那些。
- 在 Basingstoke 型的括号内留下一个 p，以致答案不符合题目要求的形式。
- 在最后的方程中除以 p 而漏掉 $p = 0$，或除以 q^3 而漏掉 $p = 1$。

STEP 之外

这是在门槛策略（threshold policy）下计算期望费用的一个小例子，而这类计算正是可靠性理论（reliability theory）和运筹学（operational research）的家常便饭：管理者要选定门槛，在更换完好零件的费用与更频密维修的费用之间取得平衡。我们所用的尾和（tail sum）处理手法，即把 $\mathrm{P}(X \geq k)$ 写成一减去首几项，与得出非负整数值变量的公式 $\mathrm{E}(X) = \sum_{k \geq 1} \mathrm{P}(X \geq k)$ 背后的想法如出一辙。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）对这些想法有严谨细致的阐述。

详细解答

STEP III 2009，第 1 题

纯数学 · 圆的坐标几何

STEP III 2009，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：

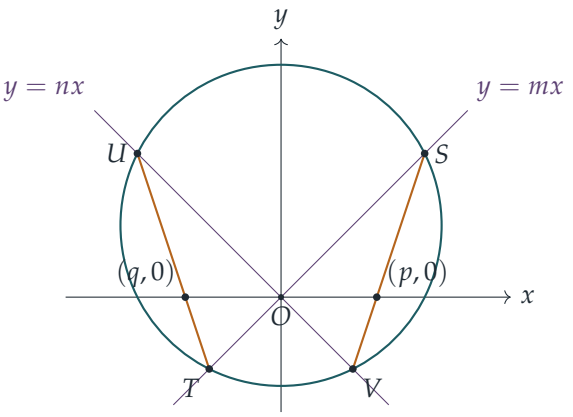
step.maths.org/questions/09-s3-q1

关键思路

四个点全部位于通过原点的直线上，所以每个点都可以用一个数来描述：它的 x 坐标。把圆与直线 $y = mx$ 相交，会得到一条二次方程（quadratic），它的两个根就是 s 和 t；我们始终不需要求出这两个根本身，只需要它们的和与积。最后的恒等式（identity），来自把 $p + q$ 的分子适当分组，使当中只出现 st、$s + t$、uv 和 $u + v$。

解答

在整个解答中，S 和 T 位于直线 $y = mx$ 上，U 和 V 位于 $y = nx$ 上，两条直线都通过原点 O。我们假设 $m \neq n$（否则四个点全在同一条直线上），并假设 SV 和 UT 都不是水平线，使它们确实与 $y = 0$ 相交；即 $ms \neq nv$ 及 $mt \neq nu$。



截距 p 和 q

通过 $S(s, ms)$ 和 $V(v, nv)$ 的直线，斜率（gradient）为 $\dfrac{nv - ms}{v - s}$，所以线上的点满足

$$y - ms = \frac{nv - ms}{v - s}(x - s).$$

代入 $y = 0$ 和 $x = p$，得 $p - s = -\dfrac{ms(v - s)}{nv - ms}$，所以

$$p = \frac{s(nv - ms) - ms(v - s)}{nv - ms} = \frac{nsv - msv}{nv - ms} = \frac{(m - n)sv}{ms - nv},$$

这正是所要求的结果。（若 $v = s$，直线 SV 是铅垂线，所以 $p = s$；公式仍然给出 $\frac{(m-n)s^2}{(m-n)s} = s$。）

直线 UT 把 $y = mx$ 上的 $T(t, mt)$ 与 $y = nx$ 上的 $U(u, nu)$ 连起来，情况与 SV 连接 S 和 V 完全一样。把 s 换成 t，v 换成 u，得

$$q = \frac{(m - n)tu}{mt - nu}.$$

s 和 t 所满足的二次方程

点 (x, mx) 位于 $x^2 + (y - c)^2 = r^2$ 上，当且仅当 $x^2 + (mx - c)^2 = r^2$，即

$$(1 + m^2)x^2 - 2mc\,x + c^2 - r^2 = 0.$$

s 和 t 都满足这条方程。取 S 和 T 为该直线与圆的两个（相异）交点，则 s 和 t 就是方程的两个根，所以根据根与系数的关系（relations between roots and coefficients），

$$st = \frac{c^2 - r^2}{1 + m^2}, \qquad s + t = \frac{2mc}{1 + m^2}.$$

证明 $p + q = 0$

对直线 $y = nx$ 作同样的论证，可知 u 和 v 是 $(1 + n^2)x^2 - 2nc\,x + c^2 - r^2 = 0$ 的两个根，所以

$$uv = \frac{c^2 - r^2}{1 + n^2}, \qquad u + v = \frac{2nc}{1 + n^2}.$$

现在把两个截距通分相加：

$$p + q = (m - n)\,\frac{sv(mt - nu) + tu(ms - nv)}{(ms - nv)(mt - nu)}.$$

? **想一想** 分子同时涉及 s, t, u, v 四个量，我们该怎样把它驯服？

把它展开，再按系数 m 或 n 把各项配对：

$$\begin{aligned} sv(mt-nu)+tu(ms-nv) &= mstv-nsuv+mstu-ntuv \\ &= mst(u+v)-nuv(s+t). \end{aligned}$$

现在每一项都是我们已知的对称函数（symmetric function）。代入后，

$$\begin{aligned} mst(u+v)-nuv(s+t) &= m\cdot\frac{c^2-r^2}{1+m^2}\cdot\frac{2nc}{1+n^2}-n\cdot\frac{c^2-r^2}{1+n^2}\cdot\frac{2mc}{1+m^2} \\ &= \frac{2mnc(c^2-r^2)}{(1+m^2)(1+n^2)}-\frac{2mnc(c^2-r^2)}{(1+m^2)(1+n^2)}=0. \end{aligned}$$

根据我们的假设，分母 $(ms-nv)(mt-nu)$ 不等于零，所以 $p+q=0$，证明完毕。

验算

如图中所示，取 $c=1$、$r^2=5$、$m=1$、$n=-1$。两条二次方程分别是 $x^2-x-2=0$ 和 $x^2+x-2=0$，所以 $s=2$、$t=-1$，而 $u=-2$、$v=1$。于是 $st=-2=\frac{1-5}{2}$，$s+t=1=\frac{2}{2}$，与预期一致。由 $S(2,2)$ 到 $V(1,-1)$ 的直线斜率为 3，与 $y=0$ 相交于 $p=\frac{4}{3}$，与 $\frac{2\cdot2\cdot1}{2+1}$ 吻合；由 $U(-2,2)$ 到 $T(-1,-1)$ 的直线则相交于 $q=-\frac{4}{3}$。以电脑随机选取 16 000 组圆和直线作检验，每一次 $|p+q|$ 都小于 10^{-12}。

常见错误

- 把点配错：q 来自 U 和 T，所以应作的代换是 $s\to t$、$v\to u$；若改为对调 $s\leftrightarrow v$，得到的仍然是 p。
- 试图直接解二次方程，把 s,t,u,v 明确求出来。平方根会令代数运算无从入手；正是 $mst(u+v)-nuv(s+t)$ 这个分组，才令整个式子化简为零。

STEP 之外

直线 $y=0$ 与圆相交所成的弦（chord），中点是 O，因为圆心 $(0,c)$ 在 y 轴上。所以这道题其实证明了蝴蝶定理（butterfly theorem）：若两条弦 ST 和 UV 都通过某条弦的中点 O，则 SV 和 UT 与该弦的交点到 O 的距离相等。我们的证明从没用过 $c^2<r^2$，所以即使 O 在圆外，证明依然成立；这暗示了这个结果真正的归宿是射影几何（projective geometry），在那里它是笛沙格（Desargues）对合定理（involution theorem）的一个特例。Coxeter 与 Greitzer 合著的 *Geometry Revisited*（《几何学的新探索》）是进一步阅读的上佳之选。

详细解答

STEP III 2009，第 2 题

纯数学 · 幂级数解

STEP III 2009，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/09-s3-q2

关键思路

若 y 是幂级数（power series），则 y'、y'' 和 x^3y 也是幂级数，而微分方程（differential equation）所说的，就是把它们合并而成的一个幂级数恒等于零。一个对所有 x 都等于零的幂级数，每个系数（coefficient）都必定是零，于是微分方程就化为一连串联系各个 a_n 的简单方程。真正要下工夫的是整理工作：在比较系数之前，先把求和指标平移，使每个级数都以 x 的同一个幂写出。

解答

第 (i) 部分

逐项求导（term-by-term differentiation）（题目已说明所有级数均收敛，所以我们视此做法为可行），

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(a_n x^n\right) = n a_n x^{n-1}.$$

$n = 0$ 的一项 a_0 是常数，其导数为 0，所以求和由 $n = 1$ 开始：

$$y' = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

这就是所要求的结果。再做一次，y' 中 $n = 1$ 的一项 a_1 是常数，所以

$$y'' = \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2}.$$

逐项写出至 a_3：

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots,$$
$$y' = a_1 + 2a_2x + 3a_3x^2 + \cdots,$$
$$y'' = 2a_2 + 6a_3x + \cdots.$$

第 (ii) 部分

我们把 $xy'' - y' + 4x^3y$ 的每一项都写成以 x^{n-1} 表示的级数。首先，

$$xy'' = \sum_{n=2}^{\infty} n(n-1)a_n x^{n-1}.$$

这个和式中 $n = 1$ 的一项本应是 $1 \cdot 0 \cdot a_1 = 0$，所以也可以让它由 $n = 1$ 开始。于是

$$xy'' - y' = \sum_{n=1}^{\infty} \big[n(n-1) - n\big]a_n x^{n-1} = \sum_{n=1}^{\infty} n(n-2)a_n x^{n-1}.$$

至于最后一项，$4x^3y = \sum_{m=0}^{\infty} 4a_m x^{m+3}$，代入 $m = n - 4$ 便可令幂次变成 x^{n-1}：

$$4x^3y = \sum_{n=4}^{\infty} 4a_{n-4}\, x^{n-1}.$$

想一想 为甚么我们可以比较系数?

方程说的是：在级数收敛的区间内，对每一个 x 都有 $xy'' - y' + 4x^3y = 0$，所以合并后的幂级数在该区间上是零函数。一个收敛的幂级数若在某区间上恒等于零，则它的所有系数必定都是零（代入 $x = 0$ 得出常数项，除以 x，然后重复这个步骤）。因此 x 的每个幂的系数都必须等于零。

各系数如下：

- x^0（$n = 1$）：$1 \cdot (-1)\, a_1 = 0$，所以 $a_1 = 0$，正如所求。
- x^1（$n = 2$）：$2 \cdot 0 \cdot a_2 = 0$，此式自动成立，所以 a_2 可任意取值。
- x^2（$n = 3$）：$3 \cdot 1 \cdot a_3 = 0$，所以 $a_3 = 0$。
- x^{n-1}，其中 $n \geq 4$：$n(n-2)a_n + 4a_{n-4} = 0$。

由于当 $n \geq 4$ 时 $n(n-2) \neq 0$，我们可以除以它，从而得到

$$a_n = -\frac{4}{n(n-2)}\, a_{n-4} \qquad (n \geq 4),$$

正如所求。（a_0 也可任意取值，因为没有任何方程单独涉及它。）

$a_0 = 1$、$a_2 = 0$ **的情况**。 这个递推关系（recurrence relation）只把 a_n 与 a_{n-4} 联系起来，所以各系数按 n 模（modulo）4 的余数分成四条链。由 $a_1 = 0$、$a_2 = 0$ 和 $a_3 = 0$ 开始的三条链全部为零，所以只有 a_{4k} 可能不等于零。代入 $n = 4k$，其中 $k \geq 1$，

$$a_{4k} = -\frac{4}{4k(4k-2)}\, a_{4k-4} = -\frac{1}{2k(2k-1)}\, a_{4k-4}.$$

由 $a_0 = 1$ 出发，每一步都乘以 -1，并除以阶乘（factorial）中接下来的两个因数，所以

$$a_{4k} = \frac{(-1)^k}{(2k)!}.$$

（验算：$a_4 = -\frac{1}{2}$，$a_8 = -\frac{1}{12} \cdot (-\frac{1}{2}) = \frac{1}{24}$。）因此

$$y = \sum_{k=0}^{\infty} \frac{(-1)^k x^{4k}}{(2k)!} = \sum_{k=0}^{\infty} \frac{(-1)^k (x^2)^{2k}}{(2k)!} = \cos(x^2),$$

这里用到马克劳林级数（Maclaurin series）$\cos u = \sum_k (-1)^k u^{2k}/(2k)!$，并取 $u = x^2$。这就是所要求的结果。

$a_0 = 0$、$a_2 = 1$ **的情况**。 这次由 a_0、a_1 和 a_3 开始的三条链全部为零，只有 a_{4k+2} 保留下来。代入 $n = 4k + 2$，其中 $k \geq 1$，

$$a_{4k+2} = -\frac{4}{(4k+2)\,4k}\, a_{4k-2} = -\frac{1}{(2k+1)\,2k}\, a_{4k-2},$$

再由 $a_2 = 1$ 得 $a_{4k+2} = (-1)^k/(2k+1)!$。所以

$$y = \sum_{k=0}^{\infty} \frac{(-1)^k x^{4k+2}}{(2k+1)!} = \sum_{k=0}^{\infty} \frac{(-1)^k (x^2)^{2k+1}}{(2k+1)!} = \sin(x^2).$$

验算

直接代入。对于 $y = \sin(x^2)$，有 $y' = 2x\cos(x^2)$ 和 $y'' = 2\cos(x^2) - 4x^2\sin(x^2)$，所以

$$xy'' - y' + 4x^3 y = 2x\cos(x^2) - 4x^3\sin(x^2) - 2x\cos(x^2) + 4x^3\sin(x^2) = 0.$$

$\cos(x^2)$ 的情况亦同样成立。数值上，余弦级数在 $x = 1.2$ 处的首 4o 项之和为 0.1304237…，与 $\cos(1.44)$ 吻合。此外，两个答案在 $x = 0$ 处都合理：$\cos 0 = 1 = a_0$，而在 0 附近 $\sin(x^2) \approx x^2$，所以 $a_2 = 1$。

常见错误

- 未平移指标就比较系数，以致把一个级数中的 a_n 与另一个级数中的 a_n 配对，尽管它们所乘的是 x 的不同幂。

- 忘记 $a_3 = 0$，而这正是消去 $a_3, a_7, a_{11}, \ldots$ 这条链所必需的。
- 除以 $n(n-2)$ 时没有指出它在 $n \geq 4$ 时不等于零。当 $n = 2$ 时它等于零，这正是 a_2 可任意取值的原因。

STEP 之外

这是弗罗贝尼乌斯方法（method of Frobenius）最简单的形式。点 $x = 0$ 是这个方程的正则奇点（regular singular point），因为 y'' 的系数在该处等于零；一般理论解释了为甚么单凭一个普通的幂级数在这里行得通，而在别的情况下有时却会失败（指标根（indicial roots）0 和 2 相差一个整数）。同样的方法可以得出贝塞尔函数（Bessel functions）和勒让德函数（Legendre functions）。要严格论证逐项求导的合理性，需要用到一致收敛（uniform convergence），那是大学一年级分析课的内容。关于这两个课题，可参阅 Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材），以及 Abbott 的 *Understanding Analysis*（数学分析入门教材）。

详细解答

STEP III 2009，第 3 题

纯数学 · 极限与曲线描绘

STEP III 2009，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/09-s3-q3

关键思路

这个函数在 $t = 0$ 处没有定义，纯粹是因为出现了 $0/0$，所以幂级数（power series）可以告诉我们它在该处“应该”取甚么值。第 (ii) 部分只需把各项通分，写在公分母（common denominator）之上；当中出现的偶函数（even function）其实就是乔装了的 $\frac{t}{2}\coth\frac{t}{2}$。至于第 (iii) 部分，$f'(t)$ 的正负完全取决于它的分子，而这个分子恰好就是 $\mathrm{e}^t(1-t)-1$，这正是题目给出该不等式的原因。

解答

第 (i) 部分

利用 $\mathrm{e}^t = 1 + t + \frac{1}{2}t^2 + \frac{1}{6}t^3 + \cdots$，当 $t \neq 0$ 时可以约去一个因子 t：

$$f(t) = \frac{t}{t + \frac{1}{2}t^2 + \frac{1}{6}t^3 + \cdots} = \frac{1}{1 + \frac{1}{2}t + \frac{1}{6}t^2 + \cdots} \longrightarrow 1 \quad \text{当} t \to 0,$$

因为分母趋向 1。所以 $\lim_{t\to 0} f(t) = 1$。

由商法则（quotient rule），

$$f'(t) = \frac{(\mathrm{e}^t - 1) - t\,\mathrm{e}^t}{(\mathrm{e}^t - 1)^2}.$$

分子和分母都趋向 0，所以把两者分别展开。分子是

$$\begin{aligned}\mathrm{e}^t - 1 - t\,\mathrm{e}^t &= \left(t + \tfrac{1}{2}t^2 + \tfrac{1}{6}t^3 + \cdots\right) - \left(t + t^2 + \tfrac{1}{2}t^3 + \cdots\right) \\ &= -\tfrac{1}{2}t^2 - \tfrac{1}{3}t^3 + \cdots,\end{aligned}$$

而分母是 $\left(t+\frac{1}{2}t^2+\cdots\right)^2 = t^2\left(1+\frac{1}{2}t+\cdots\right)^2$。把分子和分母同除以 t^2，

$$f'(t) = \frac{-\frac{1}{2}-\frac{1}{3}t+\cdots}{\left(1+\frac{1}{2}t+\cdots\right)^2} \longrightarrow -\frac{1}{2} \quad \text{当} t \to 0.$$

第 (ii) 部分

设 $g(t) = f(t) + \frac{1}{2}t$。通分后，

$$g(t) = \frac{t}{\mathrm{e}^t-1} + \frac{t}{2} = \frac{2t+t(\mathrm{e}^t-1)}{2(\mathrm{e}^t-1)} = \frac{t\,(\mathrm{e}^t+1)}{2(\mathrm{e}^t-1)}.$$

现在以 $-t$ 代替 t，再把分子和分母同乘以 e^t：

$$g(-t) = \frac{-t\,(\mathrm{e}^{-t}+1)}{2(\mathrm{e}^{-t}-1)} = \frac{-t\,(1+\mathrm{e}^t)}{2(1-\mathrm{e}^t)} = \frac{t\,(\mathrm{e}^t+1)}{2(\mathrm{e}^t-1)} = g(t).$$

所以 $f(t)+\frac{1}{2}t$ 是偶函数，正如题目所求。（把分子和分母同除以 $\mathrm{e}^{t/2}$，可见 $g(t) = \frac{t}{2}\coth\frac{t}{2}$，是两个奇函数（odd function）之积，这是另一个看出此结果的方法。）

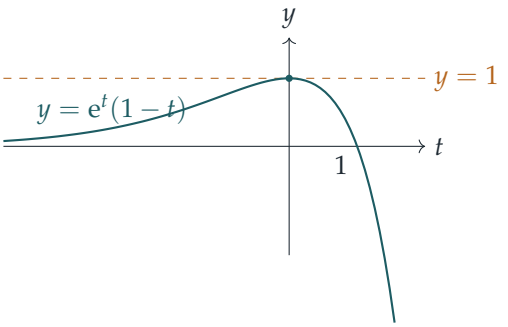
这对绘画草图有一个有用的推论：$f(t)-f(-t) = -t$，所以 f 的图像就是一个偶函数的图像被直线 $y = -\frac{1}{2}t$ 倾斜而成。

第 (iii) 部分

设 $h(t) = \mathrm{e}^t(1-t)$。则

$$h'(t) = \mathrm{e}^t(1-t) - \mathrm{e}^t = -t\,\mathrm{e}^t,$$

它在 $t<0$ 时为正，在 $t>0$ 时为负。所以 h 一直递增至 $t=0$，其后递减：它唯一的转向点（turning point）是极大点，$h(0)=1$。此外，当 $t\to-\infty$ 时 $h(t)\to 0$，而当 $t>1$ 时 $h(t)<0$。



曲线在 $t=0$ 处与 $y=1$ 相切，在其余各处都严格位于其下方，所以

$$\mathrm{e}^t(1-t) \le 1, \quad \text{等号成立当且仅当} t = 0.$$

想一想 这个不等式怎样告诉我们有关 $f'(t)$ 的信息?

再看一次第 (i) 部分中 $f'(t)$ 的分子:

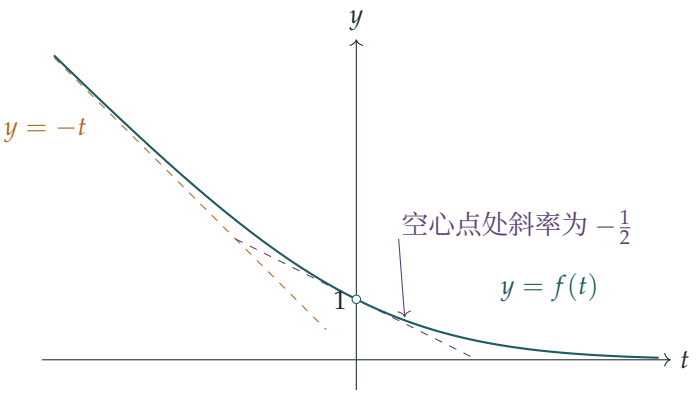
$$e^t - 1 - te^t = e^t(1-t) - 1.$$

根据上面的严格不等式，当 $t \neq 0$ 时它严格为负。分母 $(e^t-1)^2$ 在 $t \neq 0$ 时严格为正。因此对所有 $t \neq 0$ 都有 $f'(t) < 0$；特别地，$f'(t) \neq 0$，所以 f 没有驻点（stationary point），并且在 0 的两侧都是递减的。

f 的草图

把各项事实整理如下。

- 对所有 $t \neq 0$ 都有 $f(t) > 0$，因为 t 与 $e^t - 1$ 同号。
- 在 0 附近，图像以斜率 $-\frac{1}{2}$ 趋近 $(0,1)$（第 (i) 部分）；该点本身并不在图像上，因为 $f(0)$ 没有定义。
- f 处处递减（第 (iii) 部分）。
- 当 $t \to \infty$ 时，$f(t) = t/(e^t-1) \to 0^+$，因为指数函数的增长胜过 t。
- 当 $t \to -\infty$ 时，$f(t) + t = \dfrac{te^t}{e^t-1} \to 0$，所以 $y = -t$ 是一条斜渐近线（oblique asymptote）。这也可由第 (ii) 部分得出：$f(t) = f(-t) - t$ 而 $f(-t) \to 0$。



验算

- 第 (ii) 部分预测 $f(-1) - f(1) = 1$。以数值计算，$f(1) = 1/(e-1) \approx 0.58198$ 而 $f(-1) = e/(e-1) \approx 1.58198$，两者之差恰好是 1。

- 把 $t = \pm 10^{-4}$ 代入导数公式，得 -0.49998 和 -0.50002，与极限 $-\frac{1}{2}$ 吻合。
- 由 coth 的形式可得，当 t 很小时 $g(t) \approx 1 + \frac{1}{12}t^2$，所以 $f(t) \approx 1 - \frac{1}{2}t + \frac{1}{12}t^2$，与第 (i) 部分的两个极限都相符。

常见错误

- 对 $f'(t)$ 用了一次洛必达法则（L'Hôpital's rule）后，遇到另一个 $0/0$ 便停下来。把分子和分母都展开至 t^2 阶，就可避免这个问题。
- 只证明了 $\mathrm{e}^t(1-t) \leq 1$，便声称 $f'(t) \neq 0$。你需要该不等式在 $t \neq 0$ 时是严格的，而这是由 $t = 0$ 是 h 唯一的驻点得出的。
- 绘画草图时把 $(0,1)$ 画成实心点：f 在该处并无定义。

STEP 之外

展开式 $\dfrac{t}{\mathrm{e}^t - 1} = \sum_{n \geq 0} B_n \dfrac{t^n}{n!}$ 定义了伯努利数（Bernoulli numbers）B_n。第 (i) 部分表明 $B_0 = 1$ 和 $B_1 = -\frac{1}{2}$，而第 (ii) 部分表明对每个奇数 $n \geq 3$ 都有 $B_n = 0$。这些数出现在欧拉—麦克劳林求和公式（Euler–Maclaurin summation formula）、幂和 $1^k + 2^k + \cdots + n^k$，以及欧拉（Euler）关于 $\sum 1/n^{2k}$ 的公式之中。把像 $t = 0$ 处那样的“洞”补上，正是可去奇点（removable singularity）的概念；若想看极限与连续性的严谨论述，可阅读 Abbott 的 *Understanding Analysis*。

详细解答

STEP III 2009，第 4 题

纯数学 · 拉普拉斯变换

STEP III 2009，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/09-s3-q4

关键思路

第 (i) 至 (iii) 部分各自只需对定义所用的积分（integral）施以一行微积分运算：(i) 只是把两个指数函数合并，(ii) 是代换法（substitution），而 (iii) 是分部积分法（integration by parts）。第 (iv) 部分可以连续两次分部积分来完成，但更简洁的做法是使用 (iii) 两次，让 sin 和 cos 的变换互相确定对方。至于最后一部分，可把目标 $\mathrm{e}^{-pt}\cos qt$ 看成分三步构造而成：$\sin t \to \cos t \to \cos qt \to \mathrm{e}^{-pt}\cos qt$，每一步都是四个结果的其中之一。

解答

以 $\mathcal{L}[g](s)$ 表示函数 g 的拉普拉斯变换（Laplace transform），于是 $F = \mathcal{L}[f]$。“合适”是指有关积分收敛（converge）；特别地，我们假设对 $s > 0$，当 $t \to \infty$ 时 $\mathrm{e}^{-st}f(t) \to 0$。

第 (i) 部分

把指数函数合并，

$$\mathcal{L}\left[\mathrm{e}^{-bt}f(t)\right](s) = \int_0^\infty \mathrm{e}^{-st}\mathrm{e}^{-bt}f(t)\,\mathrm{d}t = \int_0^\infty \mathrm{e}^{-(s+b)t}f(t)\,\mathrm{d}t = F(s+b),$$

这正是 F 的定义在 $s+b$ 处的取值。由于 $s+b > 0$，这一步是成立的。

第 (ii) 部分

代入 $u = at$，于是 $t = u/a$ 及 $\mathrm{d}t = \mathrm{d}u/a$。由于 $a > 0$，积分上下限 0 和 ∞ 保持不变，所以

$$\mathcal{L}\left[f(at)\right](s) = \int_0^\infty \mathrm{e}^{-st}f(at)\,\mathrm{d}t = \frac{1}{a}\int_0^\infty \mathrm{e}^{-(s/a)u}f(u)\,\mathrm{d}u = \frac{1}{a}F\left(\frac{s}{a}\right),$$

这里同样用到 $s/a > 0$。

第 (iii) 部分

使用分部积分法，对 e^{-st} 求导，对 $f'(t)$ 积分：

$$\begin{aligned}\mathcal{L}[f'](s) &= \Big[\mathrm{e}^{-st}f(t)\Big]_0^\infty - \int_0^\infty (-s)\mathrm{e}^{-st}f(t)\,\mathrm{d}t \\ &= \big(0 - f(0)\big) + s\int_0^\infty \mathrm{e}^{-st}f(t)\,\mathrm{d}t \\ &= sF(s) - f(0),\end{aligned}$$

其中上限处的项为零，因为当 $t \to \infty$ 时 $\mathrm{e}^{-st}f(t) \to 0$（这是 f 为合适函数的条件之一）。

第 (iv) 部分

设 $S(s) = \mathcal{L}[\sin t](s)$ 及 $C(s) = \mathcal{L}[\cos t](s)$。对 $s > 0$, 两个积分都收敛, 因为 $|\mathrm{e}^{-st}\sin t| \le \mathrm{e}^{-st}$，cos 的情况亦类似；而且 $\mathrm{e}^{-st}\sin t$ 和 $\mathrm{e}^{-st}\cos t$ 都趋向 0，所以 (iii) 对两者均适用。

取 $f = \sin$ 应用 (iii)，其导数（derivative）为 cos，且 $\sin 0 = 0$:

$$C(s) = sS(s).$$

取 $f = \cos$ 应用 (iii)，其导数为 $-\sin$，且 $\cos 0 = 1$:

$$-S(s) = sC(s) - 1.$$

（$-\sin t$ 的变换是 $-S(s)$，因为常数因子可以直接提到积分号外。）把第一条方程代入第二条，

$$-S(s) = s^2S(s) - 1 \quad\Longrightarrow\quad (s^2+1)S(s) = 1 \quad\Longrightarrow\quad S(s) = \frac{1}{s^2+1},$$

这正是所需的结果。顺带可得

$$C(s) = \frac{s}{s^2+1}.$$

最后一部分

> **想一想** 怎样才能不作任何新的积分，就把 $\sin t$ 变成 $\cos qt$?

求导把 sin 变成 cos（这是 (iii)），把 t 重新缩放会把 $\cos t$ 变成 $\cos qt$（这是 (ii)），而乘以一个指数函数就是 (i)。所以我们分三步进行。

第 1 步：$\cos t$。由 (iii)，取 $f(t) = \sin t$，并以 (iv) 的结果作为 F,

$$\mathcal{L}[\cos t](s) = s\cdot\frac{1}{s^2+1} - \sin 0 = \frac{s}{s^2+1}.$$

（这就是第 (iv) 部分顺带得到的结果。）

第 2 步：$\cos qt$。由 (ii)，取 $f(t)=\cos t$ 及 $a=q>0$，

$$\mathcal{L}[\cos qt](s)=\frac{1}{q}\cdot\frac{s/q}{(s/q)^2+1}=\frac{1}{q}\cdot\frac{sq}{s^2+q^2}=\frac{s}{s^2+q^2}.$$

第 3 步：$\mathrm{e}^{-pt}\cos qt$。由 (i)，取 $f(t)=\cos qt$ 及 $b=p>0$，我们把 s 换成 $s+p$：

$$\boxed{\mathcal{L}\left[\mathrm{e}^{-pt}\cos qt\right](s)=\frac{s+p}{(s+p)^2+q^2}}\qquad(s>0).$$

验算

- 令 $p\to 0$ 及 $q=1$，得 $s/(s^2+1)$，即第 1 步中 $\cos t$ 的变换。令 $q\to 0$，得 $1/(s+p)$，而事实上 $\int_0^\infty \mathrm{e}^{-(s+p)t}\,\mathrm{d}t=1/(s+p)$。
- 当 $s\to\infty$ 时，$sF(s)\to 1=\cos 0$，这是必然的：当 s 很大时，权重 $s\mathrm{e}^{-st}$ 集中在 $t=0$ 附近，所以 $sF(s)$ 趋向函数在 0 处的值。
- 数值上，取 $s=1$、$p=\frac{1}{2}$、$q=2$，公式给出 $1.5/6.25=0.24$，而对该积分使用辛普森法则（Simpson's rule）所得的结果与此吻合。电脑亦在随机选取的参数值下验证了 (i) 至 (iv) 及最终答案，误差在 10^{-12} 以内。

常见错误

- 在 (ii) 中，遗漏了由 $\mathrm{d}t=\mathrm{d}u/a$ 而来的因子 $1/a$，或把 $F(s/a)$ 写成 $F(as)$。
- 在第 3 步中，只平移了部分的 s：每一个 s 都要变成 $s+p$，所以分子是 $s+p$，而不是 s。
- 在 (iii) 中弄错 $f(0)$ 的正负号：边界项是 $0-f(0)$。

STEP 之外

这些法则正是以拉普拉斯变换法求解微分方程（differential equation）的核心：(iii) 把求导变成乘以 s，于是一条常系数线性常微分方程（linear ODE with constant coefficients）就变成关于 $F(s)$ 的代数方程，你把它解出来，然后在一个由本题这类结果编成的表中“查表”即可。平移法则 (i) 正好解释了为甚么阻尼振荡（damped oscillation）$\mathrm{e}^{-pt}\cos qt$ 会表现为位于 $s=-p\pm \mathrm{i}q$ 的极点（pole）。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（常微分方程入门教材）有一整章讲述这个方法。

详细解答

STEP III 2009，第 5 题

纯数学 · 根的对称函数

STEP III 2009，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/09-s3-q5

关键思路

这一题的一切都围绕着三个基本对称函数（elementary symmetric functions）

$$e_1 = x + y + z, \qquad e_2 = yz + zx + xy, \qquad e_3 = xyz,$$

因为一旦知道它们的值，x、y 和 z 就是一条已知三次方程（cubic）$t^3 - e_1 t^2 + e_2 t - e_3 = 0$ 的根。每一项新资料，都是把两个已知的幂和（power sums）相乘，再看看出现了哪些交叉项而得来的。递推关系（recurrence）则可由这条三次方程直接得出：x、y、z 各自满足该方程，所以它们的各次幂也同样满足。

解答

设 $S_n = x^n + y^n + z^n$，则已知 $S_1 = 1$、$S_2 = 2$、$S_3 = 3$。

$yz + zx + xy$ 的值

把 S_1 平方，

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

所以 $1 = 2 + 2(yz + zx + xy)$，由此得

$$yz + zx + xy = -\tfrac{1}{2},$$

正如所求。

$\sum x^2y$ 的值，然后是 xyz

设 $T = x^2y + x^2z + y^2z + y^2x + z^2x + z^2y$，即所有形如（一个变量的平方）×（另一个不同的变量）的六项之和。把 S_2 乘以 S_1 会得到九项：三个立方项和 T 的六项：

$$(x^2 + y^2 + z^2)(x + y + z) = (x^3 + y^3 + z^3) + T.$$

因此 $2 \times 1 = 3 + T$，于是

$$T = -1,$$

这就是第二个结果。

> ? **想一想** 已知的量当中，哪两个相乘会同时包含 T 和 xyz?

试试 $(x + y + z)(yz + zx + xy)$。展开后得到九项。其中六项的形状是 $x \cdot xy = x^2y$，而这些正好就是 T 的六项。其余三项是 $x \cdot yz$、$y \cdot zx$ 和 $z \cdot xy$，每一项都等于 xyz。所以

$$(x + y + z)(yz + zx + xy) = T + 3xyz.$$

代入已知的值，$1 \times \left(-\frac{1}{2}\right) = -1 + 3xyz$，所以 $3xyz = \frac{1}{2}$，而

$$xyz = \frac{1}{6},$$

正如所求。

递推关系

现在我们知道 $e_1 = 1$、$e_2 = -\frac{1}{2}$ 和 $e_3 = \frac{1}{6}$。由于

$$(t - x)(t - y)(t - z) = t^3 - e_1t^2 + e_2t - e_3 = t^3 - t^2 - \tfrac{1}{2}t - \tfrac{1}{6},$$

x、y、z 各自都是这条三次方程的根。特别地，

$$x^3 = x^2 + \tfrac{1}{2}x + \tfrac{1}{6},$$

y 和 z 亦同样如此。乘以 x^{n-2} 得

$$x^{n+1} = x^n + \tfrac{1}{2}x^{n-1} + \tfrac{1}{6}x^{n-2},$$

把 x 换成 y 或 z 也有同样的式子。把三条等式相加，

$$S_{n+1} = S_n + \tfrac{1}{2}S_{n-1} + \tfrac{1}{6}S_{n-2},$$

所以

$$a = 1, \qquad b = \tfrac{1}{2}, \qquad c = \tfrac{1}{6}.$$

想一想 “对所有 n”这个条件，在负数次幂上会否出问题?

当 $n \geq 2$ 时，乘以 x^{n-2} 并无问题。对于较小的 n，我们需要用到 x、y 和 z 的负数次幂，而这些幂是存在的，因为三者都不是零：它们的积是 $\frac{1}{6} \neq 0$。所以这个关系对每一个整数 n 都成立。(即使 x、y、z 并非全是实数，它仍然成立，而事实上它们的确不全是实数：这条三次方程有一个实根，约为 1.431，另有两个互为共轭的复数根（complex conjugate roots)。)

验算

- 取 $n = 2$ 时，递推关系应该重现已知的数据，当中用到 $S_0 = 3$：果然 $S_2 + \frac{1}{2}S_1 + \frac{1}{6}S_0 = 2 + \frac{1}{2} + \frac{1}{2} = 3 = S_3$。
- 递推关系预测 $S_4 = 3 + 1 + \frac{1}{6} = \frac{25}{6}$ 及 $S_5 = \frac{25}{6} + \frac{3}{2} + \frac{1}{3} = 6$。以数值方法解这条三次方程（根为 1.43085 和 $-0.21542 \pm 0.26471\,\mathrm{i}$)，再把四次幂和五次幂分别相加，得到 4.16667 和 6.00000，两者吻合。这些根在 n 为负数时同样满足递推关系。

常见错误

- 弄错三次方程中的正负号：它是 $t^3 - e_1t^2 + e_2t - e_3$，所以这里 $b = -e_2 = +\frac{1}{2}$ 而 $c = e_3 = \frac{1}{6}$。
- 在 $n = 2$ 检验递推关系时，忘记了 $S_0 = 3$，而不是 1。

STEP 之外

这些就是牛顿恒等式（Newton's identities)，它们把幂和与基本对称函数联系起来，适用于任意数目的变量。在它们背后的是对称多项式基本定理（fundamental theorem of symmetric polynomials)：每一个以 $x_1, \ldots, x_k$ 为变量的对称多项式，都是 $e_1, \ldots, e_k$ 的多项式。S_n 的递推关系是一个线性递推关系（linear recurrence)，其特征方程（characteristic equation）正是那条三次方程本身；同一个想法可用来解线性差分方程（linear difference equations)，并给出斐波那契（Fibonacci）数列的通项公式。Allenby 的 *Rings, Fields and Groups*（环、域与群）正是本着这种精神来阐述多项式及其根。

详细解答

STEP III 2009，第 6 题

纯数学 · 托勒密定理

STEP III 2009，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/09-s3-q6

关键思路

单位圆（unit circle）上两点之间的距离只取决于两点之间的夹角，而求这个距离的窍门是抽出“平均”角：$e^{i\beta} - e^{i\alpha}$ 等于 $e^{i(\alpha+\beta)/2}$ 乘以一个纯虚数（purely imaginary number）。一旦把每条弦（chord）都写成半角差的正弦，那条大恒等式就变成一条三角恒等式（trigonometric identity），用积化和差公式（product-to-sum formulae）两三行便可解决。这四点于是构成一个内接于圆的四边形，而该恒等式正是关于这种四边形的一条著名定理。

解答

弦长

抽出因子 $e^{i(\alpha+\beta)/2}$：

$$\begin{aligned} e^{i\beta} - e^{i\alpha} &= e^{i(\alpha+\beta)/2}\left(e^{i(\beta-\alpha)/2} - e^{-i(\beta-\alpha)/2}\right) \\ &= e^{i(\alpha+\beta)/2} \cdot 2i \sin \tfrac{1}{2}(\beta-\alpha). \end{aligned}$$

由于 $\left|e^{i(\alpha+\beta)/2}\right| = 1$ 及 $|2i| = 2$，

$$\left|e^{i\beta} - e^{i\alpha}\right| = 2\left|\sin \tfrac{1}{2}(\beta-\alpha)\right|.$$

最后，由 $0 < \alpha < \beta < 2\pi$ 得 $0 < \frac{1}{2}(\beta-\alpha) < \pi$，所以正弦值为正，模（modulus）的符号可以除去：$\left|e^{i\beta} - e^{i\alpha}\right| = 2\sin \frac{1}{2}(\beta-\alpha)$，此即所求。

? **想一想** 为甚么题目要设 $\alpha < \beta$ 这个条件?

若没有这个条件，$\frac{1}{2}(\beta-\alpha)$ 便可能落在 $(-\pi, 0)$ 之内，正弦值为负，公式就会给出负的长度。

正是因为各角按大小排列，我们才可以除去模的符号。另外留意 $|e^{i\alpha}-e^{i\beta}|=|e^{i\beta}-e^{i\alpha}|$，所以模内两项的次序并不重要。

恒等式

因为 $0<\alpha<\beta<\gamma<\delta<2\pi$，恒等式中出现的每一对角都是由小至大排列的，所以第一个结果适用于全部六个模。两边同除以公因子 $2\cdot 2=4$ 后，我们需要证明

$$\sin\tfrac{1}{2}(\beta-\alpha)\sin\tfrac{1}{2}(\delta-\gamma)+\sin\tfrac{1}{2}(\gamma-\beta)\sin\tfrac{1}{2}(\delta-\alpha)=\sin\tfrac{1}{2}(\gamma-\alpha)\sin\tfrac{1}{2}(\delta-\beta).$$

最自然的变量是相邻两点之间的三个间距。设

$$x=\tfrac{1}{2}(\beta-\alpha),\qquad y=\tfrac{1}{2}(\gamma-\beta),\qquad z=\tfrac{1}{2}(\delta-\gamma),$$

则 $\frac{1}{2}(\gamma-\alpha)=x+y$、$\frac{1}{2}(\delta-\beta)=y+z$ 及 $\frac{1}{2}(\delta-\alpha)=x+y+z$。要证明的是

$$\sin x\sin z+\sin y\sin(x+y+z)=\sin(x+y)\sin(y+z).$$

对每个乘积使用 $2\sin P\sin Q=\cos(P-Q)-\cos(P+Q)$：

$$\begin{aligned}2\sin x\sin z&=\cos(x-z)-\cos(x+z),\\ 2\sin y\sin(x+y+z)&=\cos(x+z)-\cos(x+2y+z),\\ 2\sin(x+y)\sin(y+z)&=\cos(x-z)-\cos(x+2y+z).\end{aligned}$$

把首两行相加，$\cos(x+z)$ 项互相抵消，刚好得到第三行。因此这条三角恒等式成立（事实上对所有 x,y,z 都成立），再乘回 4 并使用弦长公式，便得

$$\left|e^{i\alpha}-e^{i\beta}\right|\left|e^{i\gamma}-e^{i\delta}\right|+\left|e^{i\beta}-e^{i\gamma}\right|\left|e^{i\alpha}-e^{i\delta}\right|=\left|e^{i\alpha}-e^{i\gamma}\right|\left|e^{i\beta}-e^{i\delta}\right|,$$

此即所求。

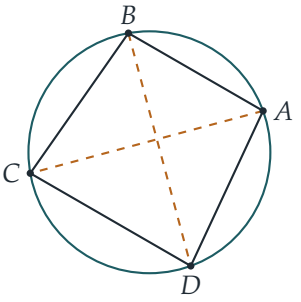
> **想一想** 这条三角恒等式对所有 x、y、z 都成立。那么我们在哪里用到 $\alpha<\beta<\gamma<\delta$ 这个次序?

就在把每个模换成 $2\sin$(半角差) 而*不加*模符号的那一步。假如各点的次序乱了，其中一些正弦值便会是负数，各长度之间的恒等式一般就不成立。例如取角度为 $0.3, 2, 1, 4$（即 $\gamma<\beta$），左边约为 4.84，右边约为 1.15。

诠释

设 A,B,C,D 为点 $e^{i\alpha},e^{i\beta},e^{i\gamma},e^{i\delta}$。它们都在单位圆上，而各角的大小次序表示沿逆时针方向走时，我们会依次遇上 A,B,C,D。所以 $ABCD$ 是一个（凸的）圆内接四边形（cyclic quadrilateral），边为 AB,BC,CD,DA，对角线为 AC、BD，而恒等式说的是

$$AB\cdot CD+BC\cdot AD=AC\cdot BD.$$



任何圆都可以由单位圆经缩放和平移得出，而等式两边按相同的倍数（即缩放因子的平方）缩放，所以这个结果对每个圆内接四边形都成立：*在圆内接四边形中，两组对边乘积之和等于两条对角线的乘积*。这就是托勒密定理（Ptolemy's theorem）。

验算

- 内接于单位圆的正方形（$\alpha, \beta, \gamma, \delta = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$）边长为 $\sqrt{2}$，对角线长为 2：$\sqrt{2} \cdot \sqrt{2} + \sqrt{2} \cdot \sqrt{2} = 4 = 2 \cdot 2$。
- 边长为 a, b 的长方形给出 $a^2 + b^2 =$ (对角线)^2：托勒密定理包含了毕氏定理（Pythagoras' theorem）。
- 数值上，随机抽取 10^5 组按大小排列的四个角，两边之差都不超过 2×10^{-15}。

常见错误

- 写下 $|e^{i\beta} - e^{i\alpha}| = 2 \sin \frac{1}{2}(\beta - \alpha)$ 而没有检查正弦值是否为正。
- 用复角公式（compound-angle formulae）把 $\sin \frac{1}{2}(\delta - \alpha)$ 等各项按原来的四个角展开：这样做是行得通的，但比使用三个间距 x, y, z 繁复得多。
- 忘记恒等式要求各顶点按环绕圆周的次序排列：右边的是两条对角线，而不是随便两条弦。

STEP 之外

对于处于一般位置的四点，同样的量只满足托勒密不等式（Ptolemy's inequality）$AB \cdot CD + BC \cdot AD \geq AC \cdot BD$，等号成立当且仅当四点按该次序共圆（concyclic）；有一个巧妙的证明，用的是复数恒等式 $(a-b)(c-d) + (a-d)(b-c) = (a-c)(b-d)$ 和三角不等式（triangle inequality）。这背后是交比（cross-ratio），它是莫比乌斯变换（Möbius

transformations）下的不变量，而莫比乌斯变换把圆映射到圆，是复几何的核心。Needham 的 *Visual Complex Analysis*（以图像直观讲解复分析）把这些想法阐述得十分精彩。

详细解答

STEP III 2009，第 7 题

纯数学 · 逐次微分与数学归纳法

> **STEP III 2009，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/09-s3-q7

关键思路

一切都源自一条看似平平无奇的恒等式（identity）：$(1+x^2)f_0(x)=1$。把它反复求导，就会得出相邻各阶导数（derivative）之间的关系式；而对某个 n 成立的关系式，再求导一次便得到下一条，这正好就是归纳步骤的模样。在第 (ii) 部分，乘以 $1+x^2$ 的适当次幂便可消去所有分母，再用第二次归纳法（induction）追踪首项，从而确定次数。

解答

第 (i) 部分

把 $f_n(x)$ 简记为 f_n，并以 $P(n)$ 表示以下命题：

$$(1+x^2)f_{n+1}+2(n+1)xf_n+n(n+1)f_{n-1}=0.$$

基础情况（*base case*）。由于对所有 x 都有 $(1+x^2)f_0=1$，求导（并利用 $f_0'=f_1$）得

$$(1+x^2)f_1+2xf_0=0.$$

再求导一次，对每一项使用积法则（product rule），

$$(1+x^2)f_2+2xf_1+2xf_1+2f_0=0, \qquad \text{即} \qquad (1+x^2)f_2+4xf_1+2f_0=0,$$

这就是 $P(1)$。

归纳步骤（*inductive step*）。假设对某个 $n\geq 1$，$P(n)$ 对所有 x 成立。由于它是关于 x 的

恒等式，我们可以对它求导。利用 $f_k' = f_{k+1}$，

$$\begin{aligned} 0 &= (1+x^2)f_{n+2} + 2xf_{n+1} + 2(n+1)f_n + 2(n+1)xf_{n+1} + n(n+1)f_n \\ &= (1+x^2)f_{n+2} + 2(n+2)xf_{n+1} + (n+1)(n+2)f_n, \end{aligned}$$

因为 $2+2(n+1) = 2(n+2)$ 及 $2(n+1)+n(n+1) = (n+1)(n+2)$。这就是 $P(n+1)$。因此根据归纳法，$P(n)$ 对所有 $n \geq 1$ 成立，证明完毕。

（等价地，也可以用莱布尼兹法则（Leibniz's rule）把 $(1+x^2)f_0 = 1$ 求导刚好 $n+1$ 次：由于 $1+x^2$ 的三阶导数为零，只有三项会留下来，而它们正是 $P(n)$ 的三项。）

第 (ii) 部分

首三个。我们有 $P_0 = (1+x^2)f_0 = 1$。其次，

$$f_1 = \frac{-2x}{(1+x^2)^2}, \qquad \text{所以} \qquad P_1 = (1+x^2)^2 f_1 = -2x.$$

用商法则（quotient rule）对 f_1 求导，

$$f_2 = \frac{-2(1+x^2)^2 + 2x\cdot 2(1+x^2)\cdot 2x}{(1+x^2)^4} = \frac{-2(1+x^2)+8x^2}{(1+x^2)^3} = \frac{6x^2-2}{(1+x^2)^3},$$

所以 $P_2 = (1+x^2)^3 f_2 = 6x^2 - 2$。（在第 (i) 部分取 $n=1$ 会得到相同的结果：$(1+x^2)f_2 = -4xf_1 - 2f_0$。）

递归关系（recurrence）。用积法则对 $P_n = (1+x^2)^{n+1} f_n$ 求导：

$$P_n' = 2(n+1)x(1+x^2)^n f_n + (1+x^2)^{n+1} f_{n+1}.$$

全式乘以 $1+x^2$，每一项都会变成一个 P：

$$(1+x^2)P_n' = 2(n+1)x\,(1+x^2)^{n+1} f_n + (1+x^2)^{n+2} f_{n+1} = 2(n+1)xP_n + P_{n+1}.$$

移项后，

$$P_{n+1} - (1+x^2)P_n' + 2(n+1)xP_n = 0$$

对每个 $n \geq 0$ 成立，证明完毕。这里不需要用归纳法；这是对每个 n 都成立的直接计算。

> **想一想** 我们怎样知道次数恰好是 n，而不会因为最高次项互相抵消而变得更小？

我们证明一个稍强的命题，让它把所需的资料一路带过整个归纳过程：P_n 是一个 n 次多项式（*polynomial*），其首项系数（*leading coefficient*）$a_n \neq 0$。$n = 0$ 的情况成立，因为 $P_0 = 1$。

假设命题对某个 $n \geq 0$ 成立，即 $P_n = a_n x^n + (\text{较低次幂})$，其中 $a_n \neq 0$。根据递归关系，

$$P_{n+1} = (1+x^2)P_n' - 2(n+1)xP_n.$$

右边是一个多项式，因为 P_n 是多项式，所以 P_n' 也是。现在 $P_n' = na_n x^{n-1} + \cdots$（当 $n = 0$ 时这是 0，与 $na_n = 0$ 吻合），所以 $(1+x^2)P_n'$ 的次数至多为 $n+1$，而 x^{n+1} 的系数是 na_n。另外，$2(n+1)xP_n$ 的次数为 $n+1$，首项系数是 $2(n+1)a_n$。因此 P_{n+1} 的次数至多为 $n+1$，而它的 x^{n+1} 系数是

$$a_{n+1} = na_n - 2(n+1)a_n = -(n+2)a_n.$$

这个数不等于零，因为 $a_n \neq 0$ 而 $n+2 > 0$。所以最高次项不可能抵消，P_{n+1} 的次数恰好是 $n+1$。根据归纳法，对每个 $n \geq 0$，P_n 都是 n 次多项式。

额外的收获是：由 $a_{n+1} = -(n+2)a_n$ 及 $a_0 = 1$ 可得 $a_n = (-1)^n(n+1)!$，与上面的 $a_1 = -2$ 和 $a_2 = 6$ 吻合。

验算

- 第 (i) 部分，取 $x = 0$、$n = 1$：由 P_n 可得 $f_0(0) = 1$、$f_1(0) = 0$、$f_2(0) = -2$，而 $f_2 + 0 + 2f_0 = -2 + 2 = 0$。
- 递归关系，取 $n = 1$：$(1+x^2)P_1' - 4xP_1 = -2 - 2x^2 + 8x^2 = 6x^2 - 2 = P_2$。
- 数值上，在 $x = 0.7$ 处，中心差分（central difference）给出 $f_0''(0.7) \approx 0.284164$，而 $(6x^2-2)/(1+x^2)^3 = 0.284164$。我们亦以有理数运算，在 200 个随机有理点上精确验证了第 (i) 部分和递归关系，直至 $n = 12$。

常见错误

- 所证明的基础情况与题目指定的范围不符：第 (i) 部分的恒等式由 $n = 1$ 开始，所以基础情况必须是 $n = 1$，这需要用到 f_2。
- 只说“次数每次增加一”，却没有检查 $(1+x^2)P_n'$ 与 $2(n+1)xP_n$ 的 x^{n+1} 项不会互相抵消。它们的确不会抵消，因为 $n \neq 2(n+1)$，但这一点必须写出来。

STEP 之外

利用复数（complex numbers）把它分解成部分分式（partial fractions），$\frac{1}{1+x^2} = \frac{1}{2\mathrm{i}}\left(\frac{1}{x-\mathrm{i}} - \frac{1}{x+\mathrm{i}}\right)$，每一阶导数都变得容易计算，并得出闭合形式（closed form）$P_n(x) = (-1)^n\, n!\, \operatorname{Im}(x+\mathrm{i})^{n+1}$（不妨用 $n = 1, 2$ 试试）。再设 $x = \cot\theta$，便可看出 P_n 的零

点是 $\cot\frac{k\pi}{n+1}$，其中 $k = 1,\ldots,n$：全部都是实数而且互不相同，就像经典正交多项式（orthogonal polynomials）的零点一样，而那些多项式同样是由反复对一个简单函数求导而产生的。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是个绝佳的去处，可以从中看到复数为何能把这类计算变得服服贴贴。

详细解答

STEP III 2009，第 8 题

纯数学 · 极限与归约公式

STEP III 2009，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/09-s3-q8

本题在 *Beyond STEP* 第 4.5 节大二学生之梦中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/sophomores-dream.html

关键思路

每一部分都围绕着对数（logarithm）的“弱”：在 0 附近，$\ln x$ 趋向无穷的速度极慢，以致 x 的任何正数次幂都能压倒 $\ln x$ 的任何次幂。代换（substitution）$x = \mathrm{e}^{-t}$ 把这件事化为题目给出的事实：指数函数胜过幂函数。分部积分（integration by parts）时，正是这个极限（limit）令边界项（boundary term）消失。在最后一部分，把指数函数的级数（series）用于 $x^x = \mathrm{e}^{x\ln x}$，得出的恰好就是第 (ii) 部分的积分。

解答

第 (i) 部分

由于 $\ln x$ 只在 $x > 0$ 时有定义，这个极限的意思是 $x \to 0^+$。设 $x = \mathrm{e}^{-t}$，则 $\ln x = -t$，而 $x \to 0^+$ 恰好相当于 $t \to \infty$。于是

$$x^m (\ln x)^n = \mathrm{e}^{-mt}(-t)^n = (-1)^n \mathrm{e}^{-mt} t^n.$$

因子 $(-1)^n$ 只是一个固定的正负号，而根据题目给出的结果，当 $t \to \infty$ 时 $\mathrm{e}^{-mt}t^n \to 0$。因此

$$\lim_{x\to 0} x^m (\ln x)^n = 0.$$

至于第二个结果，取 $m = n = 1$：由此得当 $x \to 0$ 时 $x\ln x \to 0$。由于指数函数是连续

（continuous）的，

$$x^x = \mathrm{e}^{x\ln x} \to \mathrm{e}^0 = 1 \qquad \text{当} x \to 0,$$

这正是所要证明的。

第 (ii) 部分

首先，这个积分是有意义的。被积函数（integrand）$x^m(\ln x)^n$ 在 $(0,1]$ 上连续，而且根据第 (i) 部分，当 $x \to 0$ 时它趋向 0。所以它在 $(0,1]$ 上有界（bounded），I_n 是一个普通的（收敛（convergent）的）积分。

对 I_{n+1} 作分部积分，取 $u = (\ln x)^{n+1}$ 和 $\dfrac{\mathrm{d}v}{\mathrm{d}x} = x^m$，于是 $\dfrac{\mathrm{d}u}{\mathrm{d}x} = \dfrac{(n+1)(\ln x)^n}{x}$ 及 $v = \dfrac{x^{m+1}}{m+1}$：

$$\begin{aligned} I_{n+1} &= \left[\frac{x^{m+1}(\ln x)^{n+1}}{m+1}\right]_0^1 - \int_0^1 \frac{x^{m+1}}{m+1}\cdot\frac{(n+1)(\ln x)^n}{x}\,\mathrm{d}x \\ &= \left[\frac{x^{m+1}(\ln x)^{n+1}}{m+1}\right]_0^1 - \frac{n+1}{m+1}\int_0^1 x^m(\ln x)^n\,\mathrm{d}x. \end{aligned}$$

想一想 在 $x = 0$ 处，$\ln x$ 根本没有定义，那么该处的边界项是甚么？

下限其实是指 $x \to 0^+$ 时的极限。根据第 (i) 部分，以 $m+1$ 和 $n+1$ 代替 m 和 n（两者都符合条件，因为 $m+1 \geq 1$ 及 $n+1 \geq 0$），可得 $x^{m+1}(\ln x)^{n+1} \to 0$。在上限处，$\ln 1 = 0$ 而 $n+1 \geq 1$，所以该项也是 0。因此方括号内的一项为零，于是

$$I_{n+1} = -\frac{n+1}{m+1}\,I_n,$$

这正是所要证明的。

要计算 I_n，由

$$I_0 = \int_0^1 x^m\,\mathrm{d}x = \frac{1}{m+1}.$$

出发。递推关系（recurrence）的每一步都乘以 $-\dfrac{k}{m+1}$，其中 $k = 1, 2, \ldots, n$，所以

$$I_n = \left(-\frac{1}{m+1}\right)\left(-\frac{2}{m+1}\right)\cdots\left(-\frac{n}{m+1}\right)\frac{1}{m+1} = \frac{(-1)^n\,n!}{(m+1)^{n+1}}.$$

（用一行归纳法（induction）便可确认：若结果对 n 成立，则由递推关系得 $-\frac{n+1}{m+1}\cdot\frac{(-1)^n n!}{(m+1)^{n+1}} = \frac{(-1)^{n+1}(n+1)!}{(m+1)^{n+2}}$。）

第 (iii) 部分

对于 $0 < x \leq 1$，写 $x^x = \mathrm{e}^{x \ln x}$，并使用指数级数：

$$x^x = \sum_{n=0}^{\infty} \frac{(x \ln x)^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n (\ln x)^n}{n!}.$$

（函数在 $x = 0$ 处的值不影响积分；根据第 (i) 部分，函数 x^x 在该处趋向 1，所以它有界，这个积分是普通的积分。）

现在逐项积分（integrate term by term）。$n = 0$ 的一项是 $\int_0^1 1 \,\mathrm{d}x = 1$。对于 $n \geq 1$，第 n 项就是第 (ii) 部分的 I_n，其中 $m = n$（符合条件，因为此时 m 是正整数）：

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

这条公式在 $n = 0$ 时同样给出 1。所以

$$\int_0^1 x^x \,\mathrm{d}x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)^{n+1}} = 1 - \left(\tfrac{1}{2}\right)^2 + \left(\tfrac{1}{3}\right)^3 - \left(\tfrac{1}{4}\right)^4 + \cdots,$$

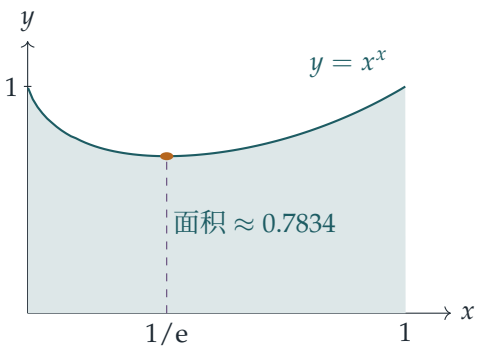
这就是所求的结果。

想一想 我们可以对无穷级数逐项积分吗？

并非必然可以，但在这里是稳妥的，因为 $x \ln x$ 始终很小。它的导数是 $\ln x + 1$，所以在 $(0, 1]$ 上它在 $x = 1/\mathrm{e}$ 处取得最小值，而且 $|x \ln x| \leq 1/\mathrm{e} < 1$。设 $u = x \ln x$，若在 u^N 项之后截断级数，则余项（remainder）满足

$$\left| \sum_{n>N} \frac{u^n}{n!} \right| \leq \sum_{n>N} \frac{|u|^n}{n!} \leq \sum_{n>N} \frac{1}{n!} \leq \frac{2}{(N+1)!}$$

对 $(0, 1]$ 内的每一个 x 都成立。在长度为 1 的区间上积分，截断后逐项积分的误差最多是 $2/(N+1)!$，而它趋向 0。所以积分等于级数的和。



验算

- 在第 (ii) 部分取 $m = 1$、$n = 1$：$\int_0^1 x \ln x \,\mathrm{d}x = -\frac{1}{4}$，与直接分部积分的结果一致（$[\frac{1}{2}x^2 \ln x - \frac{1}{4}x^2]_0^1 = -\frac{1}{4}$）。
- 数值上，用中点法则（midpoint rule）可得 $\int_0^1 x^x \,\mathrm{d}x \approx 0.7834305$，而部分和（partial sums）$1 - \frac{1}{4} + \frac{1}{27} - \frac{1}{256} + \frac{1}{3125} - \cdots$ 只需寥寥数项便稳定于 0.7834305，因为各项缩小得极快。
- 图像位于其最小值 $\mathrm{e}^{-1/\mathrm{e}} \approx 0.69$ 与 1 之间，所以面积必定介乎 0.69 与 1 之间，事实亦确是如此。

常见错误

- 不加说明便略去 0 处的边界项。这正是第 (i) 部分的全部用意，而且需要 $m + 1 \geq 1$。
- 弄错 I_n 的正负号：在 $(0, 1)$ 上 $\ln x < 0$，所以 I_n 的正负号与 $(-1)^n$ 相同。
- 在第 (iii) 部分忘记 $n = 0$ 的一项，而首项的 1 正是由此而来。

STEP 之外

最后这条恒等式称为“二年级生之梦”（sophomore's dream），由约翰·伯努利（Johann Bernoulli）于 1697 年发现。我们需要论证的那一步，即交换积分与无穷求和的次序，是大学一年级分析课的核心课题：当级数一致收敛（converges uniformly）时，这样做是稳妥的，而我们的余项估计所证明的正是这一点（它是魏尔斯特拉斯 M 判别法（Weierstrass M-test）的一个例子）。Abbott 的 *Understanding Analysis* 对一致收敛以及逐项积分何时可行，有十分易读的讲解。

详细解答

STEP III 2009，第 9 题

力学 • 抛体轨道的角度

STEP III 2009，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：

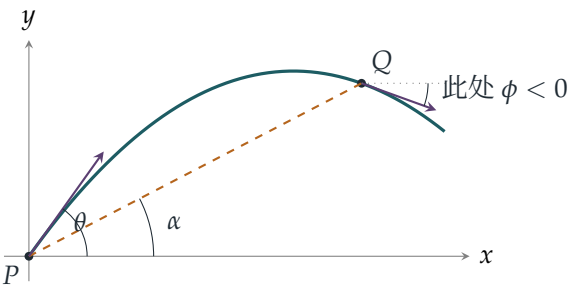
step.maths.org/questions/09-s3-q9

关键思路

在抛体运动（projectile motion）中，水平速度始终不变，而垂直速度则以恒定的变率改变。因此，由 P 到 Q 之间的平均垂直速度，就是它在两端的值的平均数，而弦（chord）PQ 的斜率（gradient）就是两端斜率的平均数。这样一行就得出第 (i) 部分。至于第 (ii) 部分，轨迹（trajectory）方程是 $\tan\theta$ 的二次方程（quadratic）（只要把 $\sec^2\theta = 1 + \tan^2\theta$ 代入），余下的工作交给根之和与根之积（sums and products of roots）便可。

解答

我们把质点（particle）视为在恒定重力 g 之下自由运动，不计空气阻力（air resistance）。取 P 为原点，x 轴沿水平正方向，y 轴垂直向上。设 u 为抛射速率，并设 $Q = (X, Y)$。我们假设 $X > 0$：Q 位于 P 的正方向一侧，所以水平速度 $u\cos\theta$ 为正，而所有角 $\theta, \phi, \theta', \phi'$ 都在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 之内。（假如 Q 在 P 的正上方或正下方，在 P 和 Q 的角便都是 $\pm\frac{\pi}{2}$，题目就没有任何内容可言。）于是 $\tan\alpha = Y/X$。



第 (i) 部分

水平速度恒为 $u\cos\theta$。垂直速度在 P 是 $u\sin\theta$，在 Q 设为 v_Q，并且以恒定的变率 g 递减，所以它是时间的线性函数（linear function）。因此，若由 P 飞行到 Q 需时 T，则

$$X = (u\cos\theta)\,T, \qquad Y = \frac{u\sin\theta + v_Q}{2}\,T,$$

因为线性函数在一个区间上的平均值，就是它两个端点值的平均数。在每一点，运动方向就是速度的方向，所以 $\tan\theta = \dfrac{u\sin\theta}{u\cos\theta}$ 及 $\tan\phi = \dfrac{v_Q}{u\cos\theta}$。两式相除，得

$$\tan\alpha = \frac{Y}{X} = \frac{1}{2}\left(\frac{u\sin\theta}{u\cos\theta} + \frac{v_Q}{u\cos\theta}\right) = \tfrac{1}{2}(\tan\theta + \tan\phi),$$

即 $\tan\theta + \tan\phi = 2\tan\alpha$，正是所求。

第 (ii) 部分

二次方程。从 $x = ut\cos\theta$、$y = ut\sin\theta - \frac{1}{2}gt^2$ 中消去 t，得轨迹

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

记 $k = \dfrac{gX^2}{2u^2}$，它是正数。由于路径通过 (X, Y)，$T = \tan\theta$ 满足

$$kT^2 - XT + (Y + k) = 0.$$

第二条轨迹有相同的 u、X 和 Y，所以 $\tan\theta'$ 满足同一条方程。两条轨迹并不相同，所以 $\theta \neq \theta'$；又因为 tan 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上是一对一（one-to-one）的，$\tan\theta$ 和 $\tan\theta'$ 就是那两个根。因此

$$\tan\theta + \tan\theta' = \frac{X}{k}, \qquad \tan\theta\tan\theta' = \frac{Y + k}{k} = 1 + \frac{Y}{k}.$$

第一个结果。由复角公式（compound-angle formula），

$$\tan(\theta + \theta') = \frac{\tan\theta + \tan\theta'}{1 - \tan\theta\tan\theta'} = \frac{X/k}{-Y/k} = -\frac{X}{Y} = -\cot\alpha.$$

这里需要 $Y \neq 0$。若 $Y = 0$（于是 $\alpha = 0$，而 $\cot\alpha$ 无定义），两根之积为 1，这表示 $\theta + \theta' = \frac{\pi}{2}$，也就是大家熟悉的情况：在水平地面上，射程相同的一对抛射角互为余角（complementary angles）。从这个意义上说，结果依然成立。

第二个结果。记 $t = \tan\alpha = Y/X$ 及 $s = \tan\theta + \tan\theta' = X/k$。于是 $Y/k = ts$，所以 $\tan\theta\tan\theta' = 1 + ts$。把第 (i) 部分分别应用于两条轨迹，得

$$\tan\phi = 2t - \tan\theta, \qquad \tan\phi' = 2t - \tan\theta'.$$

所以

$$\tan\phi+\tan\phi'=4t-s,$$
$$1-\tan\phi\tan\phi'=1-\left(4t^2-2ts+\tan\theta\tan\theta'\right)=1-4t^2+2ts-1-ts=t(s-4t).$$

这两个量不可能同时为零：若 $\tan\phi'=-\tan\phi$，则 $\tan\phi\tan\phi'=-\tan^2\phi\le 0\neq 1$。当 $Y\neq 0$（于是 $t\neq 0$）且 $s\neq 4t$ 时，

$$\tan(\phi+\phi')=\frac{4t-s}{t(s-4t)}=-\frac{1}{t}=-\cot\alpha=\tan(\theta+\theta').$$

在其余情况下，要么 $1-\tan\phi\tan\phi'$ 和 $1-\tan\theta\tan\theta'$ 同时为零（$t=0$），要么 $\tan(\phi+\phi')$ 和 $\tan(\theta+\theta')$ 同为无穷大，所以在任何情况下，$\phi+\phi'$ 与 $\theta+\theta'$ 要么正切相同，要么同为 $\frac{\pi}{2}$ 的奇数倍。无论如何，

$$\theta+\theta'-(\phi+\phi')=n\pi\quad\text{对某个整数}n.$$

想一想 正切相等，只能把角确定到相差 π 的某个倍数。我们怎样知道这个倍数刚好是一个 π?

我们利用各角的大小。在每条轨迹上，斜率是 $\tan\theta-\dfrac{gx}{u^2\cos^2\theta}$，它随 x 增加而严格递减。由于 $X>0$，我们得到 $\tan\phi<\tan\theta$，所以 $\phi<\theta$；同理 $\phi'<\theta'$。四个角全都在 $(-\frac{\pi}{2},\frac{\pi}{2})$ 之内，所以 $0<\theta-\phi<\pi$ 及 $0<\theta'-\phi'<\pi$。两式相加，

$$0<n\pi=(\theta+\theta')-(\phi+\phi')<2\pi,$$

所以 $n=1$，而 $\theta+\theta'=\pi+\phi+\phi'$，正是所求。

验算

- 取 $X=1$、$Y=\frac{1}{2}$ 及 $u^2=4g$，则 $k=\frac{1}{8}$。二次方程变成 $T^2-8T+5=0$，其根为 $4\pm\sqrt{11}$。于是 $\tan(\theta+\theta')=8/(1-5)=-2=-\cot\alpha$。以数值计算，$\theta+\theta'\approx 2.0344$ 而 $\pi+\phi+\phi'\approx 2.0344$，其中 $\tan\phi=1-\tan\theta$ 及 $\tan\phi'=1-\tan\theta'$。
- 水平地面（$Y=0$）：$\theta+\theta'=\frac{\pi}{2}$，而根据对称性，$\phi=-\theta$、$\phi'=-\theta'$。所以 $\pi+\phi+\phi'=\pi-\frac{\pi}{2}=\frac{\pi}{2}$，与预期相符。
- 我们以数值方法逐步模拟运动，确认了第 (i) 部分；又随机选取 20,000 组容许两条轨迹存在的 u、X 和 Y，检验了第 (ii) 部分的两个结果。

常见错误

- 单凭正切相等就断定 $\theta+\theta'=\pi+\phi+\phi'$。正切相等只能把角确定到相差 π 的某个倍数，你需要不等式 $\phi<\theta$ 和 $\phi'<\theta'$ 才能判定是哪一个倍数。

- 忘记说明为甚么 $\tan\theta$ 和 $\tan\theta'$ 是同一条二次方程的两个不同的根。它们不同，是因为两条轨迹不同，而 tan 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上是一对一的。

STEP 之外

第 (ii) 部分可以从可逆性（reversibility）的角度作一个颇为有趣的解读。把任何一条路径倒过来走，就得到一个由 Q 到 P、以角 $\phi + \pi$ 抛出的抛体。所以这四个角是由两端之间的一种对称性联系起来的，而关系式 $\theta + \theta' = \pi + \phi + \phi'$ 正好把它表达出来。牛顿（Newton）定律（在没有摩擦力或阻力的情况下）的时间反演对称性（time-reversal symmetry），是贯穿整个大学力学的主题。若想看大学一年级程度、对抛体以至更多课题的透彻论述，可参阅 Kleppner and Kolenkow, *An Introduction to Mechanics*（力学导论）。

详细解答

STEP III 2009，第 10 题

力学 • 简谐运动

STEP III 2009，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/09-s3-q10

关键思路

题目给出平衡（equilibrium）时的压缩量 d，是为了让我们无须另外设置弹簧的劲度（stiffness）也能知道它的大小：平衡时弹簧的力与重量互相抵消，所以 $k/m = g/d$。之后，由牛顿第二定律（Newton's second law）可知，运动是以平衡位置为中心的简谐运动（simple harmonic motion），而下落高度 h 只通过初速度影响结果。当弹簧再也无法推动质点时，质点便离开弹簧，这发生在压缩量回到零的一刻。

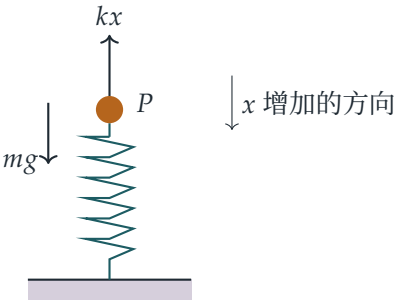
解答

微分方程

把 P 视为质量为 m 的质点（particle），弹簧视为轻弹簧（light spring），服从胡克定律（Hooke's law），劲度为 k，并忽略空气阻力。当 P 静止在弹簧上时，作用于它的力是向下的重量 mg 和向上的弹簧力 kd，所以

$$kd = mg, \qquad \text{即} \qquad \frac{k}{m} = \frac{g}{d}.$$

现在考虑质点从高处落下的情况。取 x 向下为正（它就是压缩量），并以碰到弹簧的一刻为 $t = 0$。当 P 与弹簧接触时，作用于它的力如图所示。



由于弹簧是轻的，它施于 P 的力恰好是向上的 kx（轻弹簧会把两端所受的力原封不动地传递过去）。取向下为正，由牛顿第二定律得 $m\ddot{x} = mg - kx$，所以

$$\ddot{x} + \frac{g}{d}x = g, \qquad 0 \le t \le T.$$

解方程

这是以 $x = d$ 为中心的简谐运动，其中

$$\omega = \sqrt{g/d}.$$

常数 $x = d$ 是一个特解（particular solution），而余函数（complementary function）是 $B\cos\omega t + C\sin\omega t$，所以通解（general solution）是

$$x = d + B\cos\omega t + C\sin\omega t.$$

初始条件（initial conditions）来自下落的过程。在 $t = 0$ 时质点刚刚碰到弹簧，所以 $x(0) = 0$；而 P 已自由下落了高度 h，所以它的速率是 $\sqrt{2gh}$，方向向下，亦即 x 增加的方向。因此

$$\begin{aligned} x(0) = d + B = 0 \qquad &\Rightarrow \quad B = -d, \\ \dot{x}(0) = C\omega = \sqrt{2gh} \qquad &\Rightarrow \quad C = \sqrt{2gh}\,\sqrt{d/g} = \sqrt{2hd}. \end{aligned}$$

所以

$$x = d - d\cos\omega t + \sqrt{2hd}\,\sin\omega t, \qquad A = d, \quad B = -d, \quad C = \sqrt{2hd}, \quad \omega = \sqrt{g/d}.$$

分离的时刻

? **想一想** P 究竟在甚么时候失去接触?

弹簧并没有系在 P 上，所以它只能推，不能拉。作用于 P 的接触力就是弹簧力 kx，当 $x > 0$ 时它保持为正，而当 x 回到 0 时它降至零，此时 P 正在向上运动。假如运动越过这

一点后仍按原来的方程进行，弹簧便要提供向下的拉力，但它做不到。所以 T 是使 $x = 0$ 的第一个正的时刻。

要求出这个时刻，可把振动的部分写成单一个余弦。设 $R = \sqrt{d^2 + 2hd}$，并以下式定义锐角 φ：

$$\cos\varphi = \frac{d}{R}, \qquad \sin\varphi = \frac{\sqrt{2hd}}{R},$$

于是 $\tan\varphi = \sqrt{2h/d}$，而 $\varphi = \arctan\sqrt{2h/d}$。那么 $d\cos\omega t - \sqrt{2hd}\sin\omega t = R\cos(\omega t + \varphi)$，并且

$$x = d - R\cos(\omega t + \varphi).$$

所以 $x = 0$ 当且仅当 $\cos(\omega t + \varphi) = d/R = \cos\varphi$。当 t 由 0 开始增加时，角 $\omega t + \varphi$ 由 φ 出发，而在区间 $\varphi < \omega t + \varphi < 2\pi - \varphi$ 上有 $\cos(\omega t + \varphi) < \cos\varphi$，所以在该处 $x > 0$。余弦值第一次回到 $\cos\varphi$ 是在

$$\omega T + \varphi = 2\pi - \varphi, \qquad \text{即} \qquad \omega T = 2\pi - 2\varphi.$$

除以 $\omega = \sqrt{g/d}$，得

$$T = \sqrt{d/g}\left(2\pi - 2\arctan\sqrt{2h/d}\right),$$

这正是所求的结果。

验算

- 若 $h = 0$，质点是以零速率放在处于自然长度（natural length）的弹簧上，此时 $\varphi = 0$，而 $T = 2\pi\sqrt{d/g}$：刚好是一个完整的周期（period），之后 P 以零速率回到最高点。这完全正确。
- 当 $h \to \infty$ 时，$\varphi \to \frac{\pi}{2}$ 而 $\omega T \to \pi$：速度极快的质点只在弹簧上停留半个周期，因为与振幅（amplitude）相比，平衡位置的偏移 d 变得微不足道。
- 根据能量守恒（energy conservation），P 离开时的速率应为 $\sqrt{2gh}$。事实上 $\dot{x}(T) = R\omega\sin(2\pi - \varphi) = -R\omega\sin\varphi = -\sqrt{2gh}$。
- 数值上，取 $d = 0.1\,\mathrm{m}$、$h = 0.2\,\mathrm{m}$ 和 $g = 9.8\,\mathrm{m\,s^{-2}}$ 逐步模拟这个运动，得出接触时间为 0.411 s，与公式吻合。随机选取的 d 和 h 值亦吻合至小数点后六位。

常见错误

- 从平衡位置量度位移，却忘记了题目中的 x 是压缩量，结果得出 $A = 0$。
- 把初速率取为 $\sqrt{2gh}$，但正负号弄错了：向下才是 x 增加的方向。

- 假设分离发生在平衡位置 $x = d$，或发生在半个周期之时。其实弹簧会一直推着质点，直至回复自然长度为止。

STEP 之外

这是单边约束（unilateral constraint）最简单的例子：弹簧能推而不能拉。运动在不同阶段服从不同的微分方程，而各阶段之间靠位置和速度的吻合来衔接，正如本题中自由下落阶段与弹簧阶段的衔接一样。同样的想法贯穿了碰撞和弹跳系统的理论。想进一步了解振子（oscillator），以及上文验算中所用的那类能量方法，可参阅 Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP III 2009，第 11 题

力学 • 变质量运动

STEP III 2009，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/09-s3-q11

关键思路

当质量会改变时，$F = ma$ 并不是正确的定律：我们必须使用“力 = 动量（momentum）的变化率”。尘埃本身是静止的，所以它没有带来任何动量，而彗星的动量 $M(1+bx)v$ 只会因外力而改变。这样马上便得到一个首次积分（first integral）。再写成 $v = \mathrm{d}x/\mathrm{d}t$，它就变成一条可分离变量方程（separable equation），其解是一条关于 x 的二次方程。

解答

我们把彗星视为一个质点（particle）。它所收集的尘埃原本是静止的，所以不会为彗星带来动量，而作用于整个系统（彗星加上已收集的尘埃）的唯一的力，就是沿运动方向的引力 Mf。

第 (i) 部分

当 $f = 0$ 时没有外力，所以动量守恒（conserved）。进入尘埃云时动量为 MV；在时间 t 时动量为 $M(1+bx)v$。因此

$$(1+bx)\,v = V.$$

现在代入 $v = \dfrac{\mathrm{d}x}{\mathrm{d}t}$ 并分离变量：

$$\int_0^x (1+bu)\,\mathrm{d}u = \int_0^t V\,\mathrm{d}s \quad\Longrightarrow\quad x + \tfrac{1}{2}bx^2 = Vt,$$

其中用到了 $x = 0$ 当 $t = 0$。所以 $bx^2 + 2x - 2Vt = 0$，而由于 $x \geq 0$，我们要取平方根为正的那个根：

$$x = \frac{\sqrt{1+2bVt} - 1}{b}.$$

第 (ii) 部分

现在外力是 Mf，所以

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[M(1+bx)v\right] = Mf.$$

由于 M 和 f 都是常数，由 0 积分至 t 可得 $M(1+bx)v - MV = Mft$（在 $t=0$ 时，$x=0$ 且 $v=V$）。因此

$$v = \frac{ft+V}{1+bx},$$

这正是所求。

? **想一想** 为甚么我们不能直接写 $M(1+bx)\dfrac{\mathrm{d}v}{\mathrm{d}t} = Mf$?

那条方程忽略了把每一粒新尘埃由静止加速至速率 v 所需的动量。积法则（product rule）显示出两者的分别：动量的变化率是 $M(1+bx)\dot{v} + Mb\dot{x}v$，而第二项正正就是这个“追上去”的代价。

与第 (i) 部分一样，写成 $v = \mathrm{d}x/\mathrm{d}t$ 并分离变量：

$$\int_0^x (1+bu)\,\mathrm{d}u = \int_0^t (fs+V)\,\mathrm{d}s \quad\Longrightarrow\quad x + \tfrac{1}{2}bx^2 = \tfrac{1}{2}ft^2 + Vt.$$

所以 $bx^2 + 2x - (ft^2 + 2Vt) = 0$，再次取非负根，得

$$x = \frac{\sqrt{1+bft^2+2bVt}-1}{b}.$$

（设 $f=0$ 便回到第 (i) 部分的结果。）

恒定速率。彗星以恒定速率 V 运动，当且仅当对所有 t 都有 $x = Vt$，即 $\sqrt{1+bft^2+2bVt} = 1+bVt$。两边平方，这要求

$$1+bft^2+2bVt = 1+2bVt+b^2V^2t^2 \quad \text{对所有}t,$$

而这恰好在 $f = bV^2$ 时成立。此时 $v = \dfrac{bV^2t+V}{1+bVt} = V$，所以在这个选择下彗星确实以恒定速率运动：引力刚好提供了加速它所扫起的尘埃所需的动量。

极限速率。一般而言，利用由 x 的公式得出的 $1+bx = \sqrt{1+bft^2+2bVt}$，

$$v = \frac{ft+V}{\sqrt{1+2bVt+bft^2}} = \frac{f+V/t}{\sqrt{bf+2bV/t+1/t^2}} \longrightarrow \frac{f}{\sqrt{bf}} = \sqrt{\frac{f}{b}} \quad \text{当}t\to\infty,$$

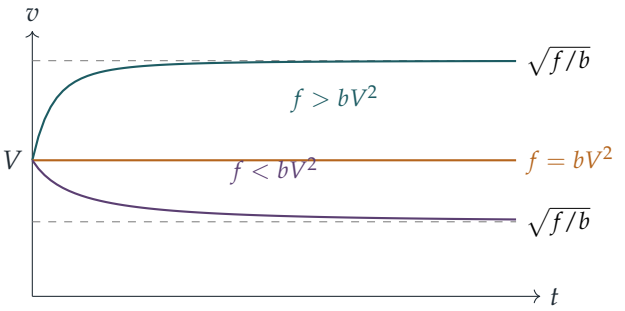
其中我们把分子和分母同除以 $t>0$，并用到了 $bf>0$（我们需要 $f>0$，因为力是沿运动方向作用的）。所以速率趋向常数 $\sqrt{f/b}$。

想一想 这与恒定速率的情况一致吗?

一致：若 $f = bV^2$，则 $\sqrt{f/b} = V$，即彗星一直以来的速率。事实上，对 v^2 求导可得

$$\frac{\mathrm{d}(v^2)}{\mathrm{d}t} = \frac{2(ft+V)(f-bV^2)}{(1+2bVt+bft^2)^2},$$

所以若 $f > bV^2$，速率会稳步上升趋向 $\sqrt{f/b}$；若 $f < bV^2$，速率会稳步下降趋向它；若 $f = bV^2$，速率则保持为 V。



验算

- 单位：b 的单位是 1/长度，而 f 是加速度，所以 f/b 的单位是速率的平方，$\sqrt{f/b}$ 则是速率；bft^2 和 bVt 都是无量纲（dimensionless）的，在平方根内也必须如此。
- 取 $b = 1$、$V = 1$、$f = 3$，以数值方法逐步计算方程 $\dot{x} = p/(1+bx)$、$\dot{p} = f$，结果与 x 和 v 的公式吻合至小数点后八位，而在 $t = 10^6$ 时速率为 $1.7320508\ldots \approx \sqrt{3}$。
- 当 $b \to 0$（没有尘埃）时，$x = \left(\sqrt{1+bft^2+2bVt} - 1\right)/b \to \frac{1}{2}ft^2 + Vt$，即常见的匀加速结果。

常见错误

- 对质量正在改变的物体使用 (质量) $\times\, \dot{v} =$ 力。
- 取了二次方程的负根，结果得出 $x < 0$。
- 对动量方程积分时，忘记了积分常数（constant of integration）MV。

STEP 之外

变质量问题是大学一年级的经典课题：同样的动量帐目可推出火箭方程（rocket equation），而雨点穿过云层下坠时不断增大，所引出的方程也与这里的十分相似。极限速率 $\sqrt{f/b}$ 是一个终端速度（terminal velocity），此时力刚好抵销动量耗用于新增质量的速率。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对变质量系统有仔细的论述。

详细解答

STEP III 2009，第 12 题

概率与统计 • 概率生成函数

STEP III 2009，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/09-s3-q12

关键思路

在第 (i) 部分，每掷一轮，硬币数目的期望值（expectation）便减半，但掷的次数本身是随机的，所以我们以它为条件（conditioning）：已知上一轮的数目，下一轮的数目便服从二项分布（binomial distribution），其平均值很容易求得。这样各个期望值便构成一个等比数列（geometric sequence）。在第 (ii) 部分，总数是若干个独立同分布（independent, identically distributed）部分之和，而概率生成函数（probability generating function）可把“独立变量之和”化为“生成函数之积”。

解答

第 (i) 部分

每次掷出正面的概率为 $\frac{1}{2}$，且各次独立，所以 $X_1 \sim \mathrm{B}(k, \frac{1}{2})$，并且

$$\mathrm{E}(X_1) = \tfrac{1}{2}k.$$

已知 $X_1 = x_1$ 时，Albert 会重新掷 x_1 次，各次互相独立，所以 $X_2 \mid X_1 = x_1 \sim \mathrm{B}(x_1, \frac{1}{2})$，并且

$$\mathrm{E}(X_2 \mid X_1 = x_1) = \tfrac{1}{2}x_1.$$

（若 $x_1 = 0$，他一次也不掷，$X_2 = 0$，与上式吻合。）对 X_1 所有可能的取值求平均，由全期望定律（law of total expectation）得

$$\mathrm{E}(X_2) = \sum_{x_1=0}^{k} \mathrm{E}(X_2 \mid X_1 = x_1)\,\mathrm{P}(X_1 = x_1) = \tfrac{1}{2}\sum_{x_1=0}^{k} x_1\,\mathrm{P}(X_1 = x_1) = \tfrac{1}{2}\,\mathrm{E}(X_1) = \tfrac{1}{4}k,$$

正如所求。

想一想 为甚么不能直接说“X_2 服从 $\mathrm{B}(\frac{1}{2}k, \frac{1}{2})$”?

因为第二轮掷的次数是随机变量（random variable）X_1，而不是数值 $\frac{1}{2}k$；事实上 $\frac{1}{2}k$ 甚至不一定是整数。以 X_1 为条件，掷的次数才得以固定，这时才可以使用二项分布的平均值。

完全相同的论证适用于每一个阶段：已知 $X_i = x_i$ 时，$X_{i+1} \sim \mathrm{B}(x_i, \frac{1}{2})$，所以 $\mathrm{E}(X_{i+1}) = \frac{1}{2}\mathrm{E}(X_i)$。由归纳法（induction）得

$$\mathrm{E}(X_i) = \frac{k}{2^i} \qquad (i \geq 1),$$

再把这个等比级数（geometric series，公比为 $\frac{1}{2}$）求和，得

$$\sum_{i=1}^{\infty} \mathrm{E}(X_i) = k\left(\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \cdots\right) = k.$$

第 (ii) 部分

就一枚硬币而言，$Y_i = n$ 表示先掷出 n 次正面，然后掷出一次反面，所以

$$\mathrm{P}(Y_i = n) = \left(\tfrac{1}{2}\right)^n \cdot \tfrac{1}{2} = \left(\tfrac{1}{2}\right)^{n+1}, \qquad n = 0, 1, 2, \ldots$$

它的概率生成函数是一个等比级数，在 $|t| < 2$ 时收敛（convergent）：

$$G_{Y_i}(t) = \sum_{n=0}^{\infty} \left(\tfrac{1}{2}\right)^{n+1} t^n = \frac{\frac{1}{2}}{1 - \frac{1}{2}t} = \frac{1}{2 - t}.$$

这 k 枚硬币是分开掷的，所以 $Y_1, \ldots, Y_k$ 互相独立，而独立变量之和的生成函数等于各自生成函数之积。因此

$$G_Y(t) = \mathrm{E}\big(t^{Y_1 + \cdots + Y_k}\big) = \prod_{i=1}^{k} \mathrm{E}\big(t^{Y_i}\big) = (2 - t)^{-k}.$$

平均值与方差（variance）。 求导数，得

$$G_Y'(t) = k(2 - t)^{-k-1}, \qquad G_Y''(t) = k(k+1)(2 - t)^{-k-2},$$

所以 $G_Y'(1) = k$ 及 $G_Y''(1) = k(k+1)$。因此

$$\mathrm{E}(Y) = G_Y'(1) = k,$$

而且，由于 $G_Y''(1) = \mathrm{E}\big(Y(Y-1)\big)$，

$$\mathrm{Var}(Y) = G_Y''(1) + G_Y'(1) - \big(G_Y'(1)\big)^2 = k(k+1) + k - k^2 = 2k.$$

分布。 $\mathrm{P}(Y=r)$ 是 $G_Y(t)$ 中 t^r 的系数。写成

$$G_Y(t) = 2^{-k}\left(1-\tfrac{1}{2}t\right)^{-k},$$

并使用负指数的二项级数（binomial series，在 $|t|<2$ 时成立）：

$$(1-x)^{-k} = \sum_{r=0}^{\infty}\binom{k+r-1}{r}x^r.$$

（其系数为 $\dfrac{k(k+1)\cdots(k+r-1)}{r!}$，即 $\binom{k+r-1}{r}$。）取 $x=\frac{1}{2}t$，得

$$\mathrm{P}(Y=r) = \binom{k+r-1}{r}\frac{1}{2^{k+r}}, \qquad r=0,1,2,\ldots$$

想一想 有没有方法可以直接看出这条公式?

有。设想把所有掷币结果排成一长列：$Y=r$ 表示第 k 次反面出现在第 $k+r$ 次掷币。这样一个由 $k+r$ 次掷币组成的特定序列，概率为 $2^{-(k+r)}$，而最后一次必定是反面，所以我们只需选择首 $k+r-1$ 次掷币中哪 r 次是正面：共有 $\binom{k+r-1}{r}$ 种方法。这就是负二项分布（negative binomial distribution）。

验算

- 当 $k=1$ 时，$\mathrm{P}(Y=r)=2^{-(r+1)}$，即单一硬币的几何分布（geometric distribution），而 $\mathrm{E}(Y)=1$、$\mathrm{Var}(Y)=2$，正是 $p=\frac{1}{2}$、由 0 起计的几何变量为人熟悉的数值。
- $G_Y(1)=1$，这是每个生成函数都必须满足的。
- 当 $k=2$ 时，$\mathrm{P}(Y=0)=\frac{1}{4}$（两枚硬币第一次都掷出反面），而 $\mathrm{P}(Y=1)=\binom{2}{1}\frac{1}{8}=\frac{1}{4}$（一枚硬币先正面后反面，另一枚掷出反面：有两种次序，每种的概率为 $\frac{1}{8}$）。
- 以分数作精确计算，已核实对于 $k\le 7$ 及 $i<40$ 有 $\mathrm{E}(X_i)=k/2^i$（所以部分和趋近 k），并且对于 $k\le 6$，$\mathrm{P}(Y=r)$ 的公式、平均值 k 及方差 $2k$ 均正确。

常见错误

- 取 $\mathrm{P}(Y_i=n)=(\frac{1}{2})^n$，忘记了最后的那次反面；这样各概率之和便是 2。
- 使用 $\mathrm{Var}(Y)=G''(1)-\big(G'(1)\big)^2$，漏了 $G'(1)$ 一项，结果得出 k 而不是 $2k$。
- 把系数写成 $\binom{k+r}{r}$：最后一次掷币已固定为反面，所以只有 $k+r-1$ 次掷币可以自由选择。

STEP 之外

第 (i) 部分是一个分支过程（branching process）：每个“个体”（一次正面）在下一代有随机数目的“后代”，这里后代数目服从 $\mathrm{B}(1, \frac{1}{2})$，而由于后代数目的平均值小于 1，这个族群终会灭绝。取条件的那一步 $\mathrm{E}(X_{i+1}) = \mathrm{E}\big(\mathrm{E}(X_{i+1} \mid X_i)\big)$ 就是塔性质（tower property），而在大学一年级的课程中，会证明 X_i 的生成函数是后代生成函数的 i 重复合（composition）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教科书）对生成函数和分支过程均有论述。

详细解答

STEP III 2009，第 13 题

概率与统计 • 中心极限定理

> **STEP III 2009，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/09-s3-q13

关键思路

一切都是单一角度 θ 的函数：$X = \cos\theta$ 及 $Y = \sin\theta$。因此，X 的分布可由累积分布函数（cumulative distribution function）求得（找出圆上哪一段弧使 $\cos\theta \le x$），而平均值、方差（variance）和协方差（covariance）不过是三角函数在一整圈上的平均。在第 (ii) 部分，$\overline{X}$ 是许多独立同分布（independent, identically distributed）变量的平均，这正是中心极限定理（central limit theorem）适用的情况。

解答

第 (i) 部分

由于 P 在单位圆上，$X = \cos\theta$ 及 $Y = \sin\theta$，其中 θ 在 $[0, 2\pi)$ 上的密度为 $\frac{1}{2\pi}$。

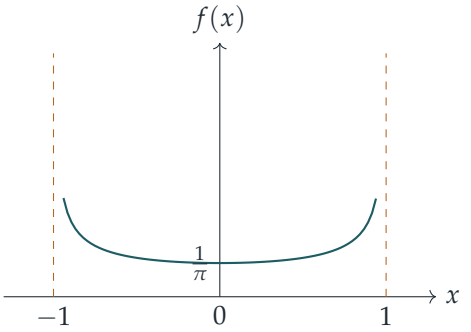
X 的密度。 显然 $-1 \le X \le 1$。取 $-1 < x < 1$，并设 $\alpha = \arccos x \in (0, \pi)$。在 $[0, 2\pi)$ 上，$\cos\theta > x$ 当且仅当 $\theta < \alpha$ 或 $\theta > 2\pi - \alpha$，这是一段总长度为 2α 的弧。因此

$$F(x) = \mathrm{P}(X \le x) = 1 - \frac{2\alpha}{2\pi} = 1 - \frac{\arccos x}{\pi}.$$

利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\arccos x = -\dfrac{1}{\sqrt{1-x^2}}$ 求导，得

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad (-1 < x < 1),$$

而在其他情况下 $f(x) = 0$。（简单验算一下：$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{1}{\pi}\left[\arcsin x\right]_{-1}^{1} = 1$。）



密度在 $x = 0$ 处最小，在 $x = \pm 1$ 处趋向无穷大：点 P 在水平直径两端附近的水平移动最慢，所以 X 大部分时间都停留在那里。

方差。 由对称性（f 是偶函数），$\mathrm{E}(X) = 0$。直接用 θ 计算，

$$\mathrm{E}(X^2) = \frac{1}{2\pi}\int_0^{2\pi} \cos^2\theta \,\mathrm{d}\theta = \frac{1}{2\pi}\int_0^{2\pi} \tfrac{1}{2}(1+\cos 2\theta)\,\mathrm{d}\theta = \frac{1}{2},$$

所以

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = \tfrac{1}{2}.$$

同理（或因为 $\theta - \frac{\pi}{2}$ 同样在一整圈上均匀分布），$\mathrm{E}(Y) = 0$ 及 $\mathrm{Var}(Y) = \frac{1}{2}$。

不相关（uncorrelated）。 我们有

$$\mathrm{E}(XY) = \frac{1}{2\pi}\int_0^{2\pi} \cos\theta \sin\theta \,\mathrm{d}\theta = \frac{1}{4\pi}\int_0^{2\pi} \sin 2\theta \,\mathrm{d}\theta = 0,$$

所以 $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 0 - 0 = 0$。因此 X 与 Y 不相关。

并非独立。 它们肯定不是独立的，因为 $X^2 + Y^2 = 1$：一旦知道 X, Y 就只能是 $\pm\sqrt{1-X^2}$ 其中之一。举一个具体事件为例：$\mathrm{P}(X > 0.8) > 0$ 及 $\mathrm{P}(Y > 0.8) > 0$，但 $\mathrm{P}(X > 0.8 \text{ 及 } Y > 0.8) = 0$，因为那样的话 $X^2 + Y^2 > 1.28$。若两者独立，这个联合概率便须等于那个（正的）乘积。相关系数为零只表示没有线性关系；这里的关系是一个圆，它足够对称，以致完全没有线性趋势。

第 (ii) 部分

每一对 (X_i, Y_i) 都服从第 (i) 部分的分布，而由于各点是独立选取的，下标不同的各对互相独立。

$\overline{X}$ 与 $\overline{Y}$ 不相关。 协方差是双线性（bilinear）的，所以

$$\operatorname{Cov}(\overline{X},\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Cov}(X_i,Y_j).$$

当 $i \neq j$ 时，X_i 只取决于 P_i，Y_j 只取决于 P_j，而这两点互相独立，所以 $\operatorname{Cov}(X_i,Y_j)=0$。当 $i=j$ 时，由第 (i) 部分得 $\operatorname{Cov}(X_i,Y_i)=0$。每一项都等于零，所以 $\operatorname{Cov}(\overline{X},\overline{Y})=0$：$\overline{X}$ 与 $\overline{Y}$ 不相关。

近似概率。 各 X_i 独立同分布，平均值为 0，方差为 $\frac{1}{2}$。所以

$$\mathrm{E}(\overline{X})=0, \qquad \operatorname{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{1}{n^2}\cdot\frac{n}{2} = \frac{1}{2n},$$

这里可以把方差相加，是因为各 X_i 互相独立。

> **想一想** X 一点也不像正态分布（normal distribution）（看看它的密度）。为甚么我们可以用正态近似?

中心极限定理只要求各项独立同分布且方差有限，不论其分布形状如何。所以当 n 很大时，$\overline{X}$ 近似服从 $\mathrm{N}\left(0,\frac{1}{2n}\right)$，标准差（standard deviation）为 $\sigma = \dfrac{1}{\sqrt{2n}}$。现在

$$\sqrt{\frac{2}{n}} = \frac{2}{\sqrt{2n}} = 2\sigma,$$

所以设 $Z \sim \mathrm{N}(0,1)$，则

$$\begin{aligned}\mathrm{P}\left(|\overline{X}| \le \sqrt{\tfrac{2}{n}}\right) = \mathrm{P}\left(\frac{|\overline{X}|}{\sigma} \le 2\right) &\approx \mathrm{P}(|Z|\le 2)\\ &= 2\Phi(2)-1 = 2(0.9772)-1 \approx 0.954 \approx 0.95,\end{aligned}$$

正如题目所求。

验算

- 方差 $\frac{1}{2}$ 是合理的：$X^2+Y^2=1$，而由对称性 $\mathrm{E}(X^2)=\mathrm{E}(Y^2)$，所以两者都必定是 $\frac{1}{2}$。根本不用积分！
- 用 4×10^5 个随机角度进行模拟，结果在多个 x 值上与 $F(x)=1-\frac{1}{\pi}\arccos x$ 吻合至约小数点后三位，并得出 $\operatorname{Var}(X)\approx 0.5007$ 及 $\operatorname{Cov}(X,Y)\approx 0.001$。
- 取 $n=50$ 并重复 40 000 次，模拟所得的 $\operatorname{Var}(\overline{X})$ 为 0.00997（理论值 0.01），而 $|\overline{X}| \le \sqrt{2/50}=0.2$ 所占的比例为 0.954，对比 $\mathrm{P}(|Z|\le 2)=0.9545$。

常见错误

- 对 $1-\frac{1}{\pi}\arccos x$ 求导时弄错正负号，得出一个负的“密度”。密度永远不可能是负数，所以要检查正负号。
- 由协方差为零推断两者独立。独立可推出协方差为零，但反过来并不成立。
- 使用 $\mathrm{Var}(\overline{X})=\frac{1}{2}$ 或 $\frac{1}{2}\cdot\frac{1}{\sqrt{n}}$：$n$ 个独立副本的平均值，其方差是单个方差除以 n；按 $1/\sqrt{n}$ 的比例变化的是标准差。

STEP 之外

密度 $1/(\pi\sqrt{1-x^2})$ 就是反正弦分布（arcsine distribution），它出现的频繁程度出人意表，例如它是随机游走（random walk）处于正值的时间所占比例的极限分布。这道题的其余部分，谈的是不相关与独立之间的差距，以及中心极限定理；在大学一年级的概率课程中，你会用矩母函数（moment generating function）或特征函数（characteristic function）来证明这个定理。事实上，$(\overline{X},\overline{Y})$ 近似服从二元正态分布（bivariate normal distribution），而对于联合正态的变量，相关系数为零确实可推出独立。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教材）对以上各点都有很好的讲解。

Beyond STEP

STEP 是大多数中学生一生中应考过最艰深的数学考试，也是最出色的考试之一：题目简短、陌生而深刻，当中最精彩的，其实是乔装了的大学数学小品。

本书精选最有趣的 STEP 试题，由 1987 年首次开考之前的样本试卷（specimen papers），到 2025 年的试卷，并以每一道题目作为一扇门。每道题目都会仔细解答，沿途提出优秀学生会问的问题，然后带你穿过这扇门：这个想法从何而来，到了大学会演变成甚么，以及接下来该读些甚么。

本书为正在准备 STEP、有天分又有志投身数学的学生而写。他们想要的不止是模拟试卷：他们想开始像数学系本科生那样思考，并一探前方有甚么正等着自己。

- 共 13 章，涵盖纯数学（pure mathematics）、力学（mechanics）、概率（probability）与统计（statistics）
- 136 道 STEP 历届试题，每题均附引导式解答
- 每道题目背后的大学数学
- 附答案的延伸练习，以及延伸阅读

2009 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。

