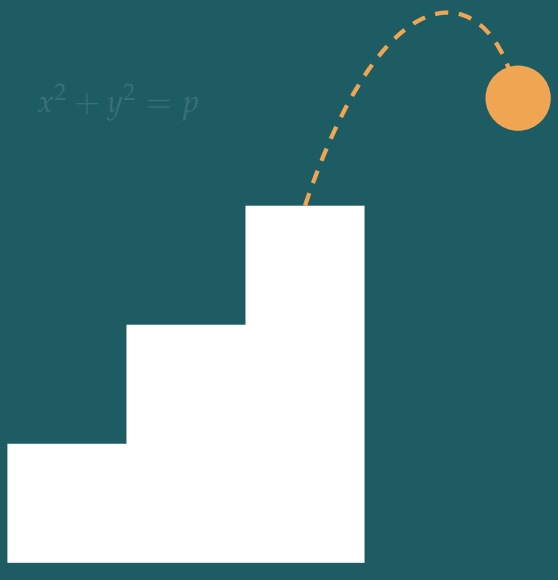
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2008 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

2008 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2008 年详细解答

2008 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 反证法 纯数学 . . . 3

第 2 题 代换积分法 纯数学 . . . 7

第 3 题 以重排法证明不等式 纯数学 . . . 10

第 4 题 凸函数 纯数学 . . . 13

第 5 题 多项式的界限 纯数学 . . . 17

第 6 题 反函数与面积 纯数学 . . . 21

第 7 题 旋转与距离 纯数学 . . . 25

第 8 题 克莱罗微分方程 纯数学 . . . 29

第 9 题 滚动圆环的动量 力学 . . . 33

第 10 题 抛体运动 力学 . . . 36

第 11 题 极限摩擦力平衡 力学 . . . 39

第 12 题 离散均匀分布变量的最大值 概率与统计 . . . 43

第 13 题 矩形排列 概率与统计 . . . 46

STEP II

第 1 题 复数迭代 纯数学 . . . 50

第 2 题 部分分式与级数 纯数学 . . . 54

第 3 题 不等式与反证法 纯数学 . . . 58

第 4 题 隐函数微分法 纯数学 . . . 62

第 5 题 代换积分法 纯数学 . . . 66

第 6 题 和化积公式 纯数学 . . . 70

第 7 题 一阶微分方程 纯数学 . . . 74

第 8 题 向量与角平分线 纯数学 . . . 78

第 9 题 抛体运动 力学 . . . 81

第 10 题 短墙壁的斜碰撞 力学 . . . 85
第 11 题 移动楔上的质点 力学 . . . 89
第 12 题 独立事件 概率与统计 . . . 93
第 13 题 不放回抽取 概率与统计 . . . 96

STEP III

第 1 题 非线性联立方程 纯数学 . . . 100
第 2 题 幂和 纯数学 . . . 103
第 3 题 椭圆的切线与焦点 纯数学 . . . 107
第 4 题 反双曲函数不等式 纯数学 . . . 111
第 5 题 递推关系 纯数学 . . . 115
第 6 题 一阶微分方程 纯数学 . . . 118
第 7 题 复数与几何 纯数学 . . . 121
第 8 题 级数与递推关系 纯数学 . . . 125
第 9 题 摩擦力与变力 力学 . . . 129
第 10 题 弹性绳与能量 力学 . . . 133
第 11 题 转动动力学 力学 . . . 137
第 12 题 生成函数 概率与统计 . . . 140
第 13 题 指示变量之和 概率与统计 . . . 144

详细解答

STEP I 2008，第 1 题

纯数学 · 反证法

> **STEP I 2008，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/08-s1-q1

关键思路

有理数（rational number）在加法、减法、乘法以及除以非零数的运算下是封闭（closed）的。本题几乎所有结论都源自这一个事实：如果某些数是有理数，那么用这些运算由它们构造出来的任何数也是有理数。在最后部分，假设四个数之中有两个是有理数，把它们巧妙地组合起来，从而得出一个题目已告诉我们是无理数（irrational number）的数。

解答

定义

若实数 x 可以写成 $x = \frac{m}{n}$，其中 m 和 n 为整数且 $n \neq 0$，则称它为*有理数*。若一个实数不是有理数，即不存在整数 m 和 $n \neq 0$ 使得 $x = \frac{m}{n}$，则称它为*无理数*。

命题 A

用反证法（proof by contradiction）：假设 pq 是无理数，但 p 和 q 都不是无理数。那么两者都是有理数，所以可以写成 $p = \frac{a}{b}$ 和 $q = \frac{c}{d}$，其中 a, b, c, d 为整数且 $b, d \neq 0$。于是

$$pq = \frac{ac}{bd},$$

其中 ac 和 bd 都是整数，而且 $bd \neq 0$，因为 b 和 d 都不是零。所以 pq 是有理数，这与我们的假设矛盾。因此 p 和 q 之中至少有一个是无理数。

命题 B

同样地，假设 $p+q$ 是无理数，但 $p=\frac{a}{b}$ 和 $q=\frac{c}{d}$ 都是有理数，其中 $b,d\neq 0$。于是

$$p+q=\frac{ad+bc}{bd},$$

这是两个整数之比，而且分母不为零，所以 $p+q$ 是有理数。这个矛盾说明 p 和 q 之中至少有一个是无理数。

命题 C

取 $p=\sqrt{2}$ 和 $q=-\sqrt{2}$。两者都是无理数，但 $p+q=0=\frac{0}{1}$ 是有理数，所以 C 不成立。

为完整起见，以下说明为何 $\sqrt{2}$ 是无理数。若 $\sqrt{2}=\frac{m}{n}$ 为最简分数（lowest terms），则 $m^2=2n^2$，所以 m^2 是偶数，因而 m 也是偶数（奇数的平方是奇数）。设 $m=2k$，便得 $n^2=2k^2$，所以 n 也是偶数，这与最简分数的假设矛盾。至于 $-\sqrt{2}$ 也是无理数，因为若 $-\sqrt{2}=\frac{m}{n}$，则 $\sqrt{2}=\frac{-m}{n}$。

最后部分

题目告诉我们 π、π^2 和 $\mathrm{e}\pi$ 都是无理数（我们只需要这三个事实）。以下两个简单的观察会反复用到：

- 若 r 和 s 是有理数，则 $r+s$、$r-s$、rs 也是有理数，而 r/s 在 $s\neq 0$ 时亦然（论证与 A 和 B 所用的完全相同）；
- 因此，若 r 是有理数，则 $\frac{1}{2}r$ 也是有理数，因为 $\frac{1}{2}$ 是有理数。

另外留意 $\pi+\mathrm{e}>0$ 和 $\pi-\mathrm{e}>0$，因为 $\pi>3>\mathrm{e}>0$；特别地，两者都不是零。

用反证法：假设以下四个数

$$\pi+\mathrm{e},\quad \pi-\mathrm{e},\quad \pi^2-\mathrm{e}^2,\quad \pi^2+\mathrm{e}^2$$

之中有两个是有理数。共有六种可能的配对，我们逐一排除。

1. *$\pi+\mathrm{e}$ 和 $\pi-\mathrm{e}$ 为有理数。*两者相加，2π 是有理数，所以 π 是有理数。矛盾。
2. *$\pi^2-\mathrm{e}^2$ 和 $\pi^2+\mathrm{e}^2$ 为有理数。*两者相加，$2\pi^2$ 是有理数，所以 π^2 是有理数。矛盾。
3. *$\pi+\mathrm{e}$ 和 $\pi^2+\mathrm{e}^2$ 为有理数。*把前者平方，$(\pi+\mathrm{e})^2$ 是有理数，而

$$2\mathrm{e}\pi=(\pi+\mathrm{e})^2-(\pi^2+\mathrm{e}^2),$$

所以 $\mathrm{e}\pi$ 是有理数。矛盾。

4. $\pi - \mathrm{e}$ 和 $\pi^2 + \mathrm{e}^2$ 为有理数。同理，

$$2\mathrm{e}\pi = (\pi^2 + \mathrm{e}^2) - (\pi - \mathrm{e})^2$$

是有理数，所以 $\mathrm{e}\pi$ 是有理数。矛盾。

5. $\pi + \mathrm{e}$ 和 $\pi^2 - \mathrm{e}^2$ 为有理数。由于 $\pi + \mathrm{e} \neq 0$，

$$\pi - \mathrm{e} = \frac{\pi^2 - \mathrm{e}^2}{\pi + \mathrm{e}}$$

是两个有理数之商，而且分母不为零，所以它是有理数。这样 $\pi + \mathrm{e}$ 和 $\pi - \mathrm{e}$ 都是有理数，而根据情况 1，这是不可能的。

6. $\pi - \mathrm{e}$ 和 $\pi^2 - \mathrm{e}^2$ 为有理数。由于 $\pi - \mathrm{e} \neq 0$，

$$\pi + \mathrm{e} = \frac{\pi^2 - \mathrm{e}^2}{\pi - \mathrm{e}}$$

是有理数，于是再次由情况 1 得出矛盾。

想一想 为甚么在情况 6 中需要 $\pi - \mathrm{e} \neq 0$?

因为不可以除以零：“有理数除以有理数仍是有理数”这一步，只在分母不为零时才成立。假如 $\pi = \mathrm{e}$，则 $\pi^2 - \mathrm{e}^2 = 0$ 是有理数，却不能告诉我们任何关于 $\pi + \mathrm{e}$ 的信息。这里 $\pi - \mathrm{e} \approx 0.42$，所以没有问题。

每一种配对都导致矛盾，所以四个数之中不可能有两个同时是有理数。因此其中至多有一个是有理数，这正是所要证明的。

验算

所用的恒等式很容易用数值验证：取 $\pi \approx 3.14159$ 和 $\mathrm{e} \approx 2.71828$，可得 $(\pi + \mathrm{e})(\pi - \mathrm{e}) \approx 5.85987 \times 0.42331 \approx 2.48055 \approx \pi^2 - \mathrm{e}^2$，以及 $(\pi + \mathrm{e})^2 - (\pi^2 + \mathrm{e}^2) \approx 17.0795 \approx 2\mathrm{e}\pi$。也数一数配对的数目：四个数共有 $\binom{4}{2} = 6$ 种配对，而六种都已处理。留意我们从未用过 e 和 e^2 是无理数这一点；根据对称性，在情况 1 和 2 中本来也可以用它们来代替 π 和 π^2。

常见错误

- 误证了 A 或 B 的逆命题（converse）。A 说的是“乘积为无理数 $\Rightarrow$ 其中一个因数为无理数”；它的逆命题并不成立（$\sqrt{2} \cdot \sqrt{2} = 2$）。
- 除以 $\pi \pm \mathrm{e}$ 时没有指出它不为零。
- 把命题 C 当作成立来使用，例如声称“π 和 e 是无理数，所以 $\pi + \mathrm{e}$ 是无理数”。事实上，$\pi + \mathrm{e}$ 是否无理数，至今仍是一个未解决的问题（open problem）。

STEP 之外

这里用到的封闭性质，说明有理数构成一个域（field）：在这样的集合里，你可以进行加、减、乘、除（除以零除外）而不会离开这个集合。域是大学一年级代数课程的核心对象，而“$\pi + \mathrm{e}$ 是否无理数?”这类问题则通向超越数论（transcendence theory），这个领域中已知的结果少得惊人。代数方面，可以读 Allenby 的 *Rings, Fields and Groups*（环、域与群）；至于有理数与无理数的数论，可以读 Davenport 的 *The Higher Arithmetic*（高等算术）。

详细解答

STEP I 2008，第 2 题

纯数学 · 代换积分法

> **STEP I 2008，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/08-s1-q2

关键思路

这个代换（substitution）的用意，是令平方根变成简单的 $t-x$。我们不必先解出 x 再对一个繁复的商求导，而是先平方消去根号，再对所得的关系式作隐函数求导（implicit differentiation）。所得的导数（derivative）恰好含有因子 $t-x$，可与被积函数（integrand）中的平方根互相约去，于是积分简化成 $\int \frac{1}{t+b}\,\mathrm{d}t$。最后一部分则是考验这个约简和除法在甚么情况下才算合理。

解答

导数

设 $R=\sqrt{x^2+2bx+c}$，则 $t-x=R$。两边平方，

$$t^2-2tx+x^2=x^2+2bx+c \quad\Longrightarrow\quad t^2-2tx-2bx-c=0.$$

把上式对 t 求导（视 x 为 t 的函数），

$$2t-2x-2t\frac{\mathrm{d}x}{\mathrm{d}t}-2b\frac{\mathrm{d}x}{\mathrm{d}t}=0, \qquad \text{所以} \qquad (t+b)\frac{\mathrm{d}x}{\mathrm{d}t}=t-x.$$

要除以 $t+b$，必须确定它不等于零。由于 $b^2<c$，

$$x^2+2bx+c=(x+b)^2+(c-b^2)>(x+b)^2,$$

所以 $R>|x+b|$，因此

$$t+b=(x+b)+R>(x+b)+|x+b|\geq 0.$$

故此对任何 x 都有 $t+b>0$，而且

$$\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{t-x}{t+b},$$

这正是所求的结果。（平方后的关系式也可以直接解出：$x=\dfrac{t^2-c}{2(t+b)}$，对此式求导，经过一些代数运算后会得到相同的答案。）

积分

二次式 $x^2+2bx+c$ 对所有 x 都是正数，所以被积函数处处有定义。作代换，并利用 $R=t-x$，

$$\begin{aligned}\int\frac{1}{\sqrt{x^2+2bx+c}}\,\mathrm{d}x&=\int\frac{1}{t-x}\,\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t=\int\frac{1}{t-x}\cdot\frac{t-x}{t+b}\,\mathrm{d}t\\&=\int\frac{1}{t+b}\,\mathrm{d}t=\ln(t+b)+C,\end{aligned}$$

其中因为 $t+b>0$，所以不需要加上绝对值（modulus）符号。约去 $t-x=R$ 是可以的，因为 $R>0$。换回 x，

$$\int\frac{1}{\sqrt{x^2+2bx+c}}\,\mathrm{d}x=\ln\left(x+b+\sqrt{x^2+2bx+c}\right)+C.$$

你可以用求导来验证：$x+b+R$ 的导数是 $1+\dfrac{x+b}{R}=\dfrac{R+x+b}{R}$，再除以 $x+b+R$ 便剩下 $\dfrac{1}{R}$。

$b^2=c$ 的情况

这时 $x^2+2bx+c=(x+b)^2$，所以 $\sqrt{x^2+2bx+c}=|x+b|$，被积函数是 $\dfrac{1}{|x+b|}$（在 $x\neq -b$ 时有定义）。

? **想一想** 为甚么 $\sqrt{(x+b)^2}$ 不能简单地写成 $x+b$?

平方根符号所指的是非负的那个根，所以 $\sqrt{(x+b)^2}=|x+b|$，当 $x+b<0$ 时它等于 $-(x+b)$。这个正负号正正决定了公式是否仍然成立。

若 $x+b>0$：直接计算，

$$\int\frac{1}{x+b}\,\mathrm{d}x=\ln(x+b)+C.$$

公式给出

$$\ln\left(x+b+(x+b)\right)=\ln 2+\ln(x+b),$$

它与直接计算的答案只相差常数 $\ln 2$。所以结果仍然成立。

若 $x+b<0$：直接计算，

$$\int \frac{1}{-(x+b)}\,\mathrm{d}x = -\ln\big(-(x+b)\big)+C = -\ln|x+b|+C.$$

但公式给出 $\ln\big(x+b+|x+b|\big)=\ln\big((x+b)-(x+b)\big)=\ln 0$，这是没有定义的。所以结果在这里不成立。

值得看看推导是在哪一步出错的。当 $x+b<0$ 时，对任何 x 都有 $t=x-(x+b)=-b$：t 是常数，根本不能用作新的变量，而 $t+b=0$ 正是我们用来作除数的那个量。严格不等式（strict inequality）$b^2<c$ 正是保证 $t+b>0$ 的条件。当 $b^2=c$ 时，只剩下非严格的不等式 $t+b\geq 0$，而等号恰好在 $x+b\leq 0$ 时成立。

验算

- 取 $b=0$、$c=1$，结果变成 $\int \frac{1}{\sqrt{x^2+1}}\,\mathrm{d}x=\ln\left(x+\sqrt{x^2+1}\right)+C=\operatorname{arsinh}x+C$，即标准结果。
- 数值上，随机选取 b 和 $c>b^2$，在随机选取的上下限之间以辛普森法则（Simpson's rule）计算积分，所得的值与对数公式之差吻合，误差小于 10^{-7}。至于 $b^2=c$ 而 $b=1.3$ 的情况，$\int_{0.5}^{4}\frac{\mathrm{d}x}{|x+b|}\approx 1.07992$，与公式给出的 $\ln\frac{5.3}{1.8}$ 相符。

常见错误

- 写成 $\ln|t+b|$ 而没有留意到 $t+b>0$，因而看不出 $b^2=c$ 时出了甚么问题。
- 在最后一部分把 $\sqrt{(x+b)^2}=x+b$ 当作成立，以致完全掩盖了 $x+b<0$ 时公式失效的情况。

STEP 之外

代换 $t=x+\sqrt{x^2+2bx+c}$ 是欧拉代换（*Euler's substitutions*）的其中一种，这类代换能把任何由 x 和 $\sqrt{ax^2+bx+c}$ 组成的有理函数（rational function）的积分，化为 t 的有理函数的积分。最后一部分说明了为甚么变量代换（change of variable）需要新变量平滑地、而且一对一（one-to-one）地依赖于旧变量：当 t 是常数时，这个方法便完全失效。Bartle 与 Sherbert 所著的 *Introduction to Real Analysis*（实分析入门）对变量代换定理有严谨的表述，并精确地列明了这些假设。

详细解答

STEP I 2008，第 3 题

纯数学 · 以重排法证明不等式

> **STEP I 2008，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/08-s1-q3

关键思路

不等式（inequality）$(*)$ 其实只是两个非负因子之积的另一种写法。一旦证明了它，其后每一个不等式都只在于如何选取 a、b、c、d，使 $(*)$ 的两边变成我们想要的两边，同时确保条件 $c \geq a$ 和 $d \geq b$ 确实成立。由于要证明的不等式都是对称（symmetric）的，我们可以先把各变量按大小排好次序，而正是这个次序提供了上述条件。

解答

证明 $(*)$

把所有项移到同一边，然后因式分解（factorise）：

$$ab + cd - bc - ad = a(b-d) - c(b-d) = (c-a)(d-b).$$

若 $c \geq a$ 且 $d \geq b$，两个因子皆为非负，故其积亦为非负。因此 $ab + cd \geq bc + ad$，而等号成立当且仅当 $c = a$ 或 $d = b$。

第 (i) 部分

两个变量。设 $x \geq y$。在 $(*)$ 中取

$$a = y, \quad b = y, \quad c = x, \quad d = x.$$

条件 $c \geq a$ 和 $d \geq b$ 所说的都是 $x \geq y$，这正是已知条件。于是 $ab + cd = y^2 + x^2$ 而 $bc + ad = xy + xy$，所以由 $(*)$ 得

$$x^2 + y^2 \geq 2xy.$$

三个变量。现在再假设 $x \geq z$ 且 $y \geq z$。取

$$a = z, \quad b = z, \quad c = x, \quad d = y.$$

这里 $c \geq a$ 即 $x \geq z$，$d \geq b$ 即 $y \geq z$，两者皆为已知。此时 $ab + cd = z^2 + xy$ 而 $bc + ad = zx + zy$，所以

$$z^2 + xy \geq xz + yz.$$

把这个不等式与 $x^2 + y^2 \geq 2xy$ 相加（由于 $x \geq y$，后者可以使用）：

$$x^2 + y^2 + z^2 + xy \geq 2xy + xz + yz,$$

再从两边减去 xy，得

$$x^2 + y^2 + z^2 \geq xy + yz + zx.$$

到目前为止，这个结果只在 $x \geq y \geq z$ 的情况下得到证明。

> **想一想** 一个只在 $x \geq y \geq z$ 时证明了的结果，怎样才能推出对所有 x、y、z 都成立的结果?

$x^2 + y^2 + z^2 \geq xy + yz + zx$ 的两边都是对称的：把任意两个变量对调，每一边都保持不变。给定任意实数 x、y、z，把它们由大至小排列为 $X \geq Y \geq Z$。根据刚才所证，$X^2 + Y^2 + Z^2 \geq XY + YZ + ZX$。但这与关于 x、y、z 的不等式是同一个不等式，只是各项的次序不同而已。所以这个不等式对所有实数 x、y 和 z 都成立。

（等号成立需要我们相加的两个不等式同时取等号，这就迫使 $x = y = z$。）

第 (ii) 部分

这个式子是三对数之和，每一对都是一个数加上它的倒数（reciprocal）：

$$\left(\frac{s}{t} + \frac{t}{s}\right) + \left(\frac{t}{r} + \frac{r}{t}\right) + \left(\frac{r}{s} + \frac{s}{r}\right).$$

因此只需证明对任意正数 p、q 都有 $\dfrac{p}{q} + \dfrac{q}{p} \geq 2$，这正是 $x^2 + y^2 \geq 2xy$ 的类比。

式子 $\frac{p}{q} + \frac{q}{p}$ 对 p 和 q 是对称的，所以我们可以适当命名，使得 $p \geq q > 0$。在 $(*)$ 中取

$$a = q, \quad b = \frac{1}{p}, \quad c = p, \quad d = \frac{1}{q}.$$

于是 $c \geq a$ 即 $p \geq q$，而 $d \geq b$ 即 $\frac{1}{q} \geq \frac{1}{p}$；由于 p 和 q 都是正数，后者同样可由 $p \geq q$ 推出。现在

$$ab + cd = \frac{q}{p} + \frac{p}{q}, \qquad bc + ad = \frac{p}{p} + \frac{q}{q} = 2,$$

所以由 $(*)$ 得 $\dfrac{p}{q} + \dfrac{q}{p} \geq 2$。

把这个结果应用于 (s,t)、(t,r) 和 (r,s) 三对数，再把三个结果相加，便得

$$\frac{s}{t}+\frac{t}{r}+\frac{r}{s}+\frac{t}{s}+\frac{r}{t}+\frac{s}{r}\geq 2+2+2=6$$

对所有正数 r、s 和 t 都成立，这正是所要证明的。等号成立当且仅当 $r=s=t$，因为每一对都需要其中两数相等。

验算

- 在第 (i) 部分试代入 $x=3$、$y=1$、$z=-2$：左边是 14，右边是 $3-2-6=-5$。把这三个值任意置换，两边都保持不变，正如对称性论证所需。
- 第 (i) 部分的另一个独立证明：$2(x^2+y^2+z^2-xy-yz-zx)=(x-y)^2+(y-z)^2+(z-x)^2\geq 0$。
- 取 $r=2$、$s=1$、$t=1$，第 (ii) 部分的式子等于 $1+\frac{1}{2}+2+1+2+\frac{1}{2}=7\geq 6$。用电脑对 20,000 个随机的精确有理数例子逐步检查，没有发现任何反例；而 $r=1.001$、$s=1$、$t=0.999$ 之类的数值所得的和仅仅略大于 6。

常见错误

- 使用 $(*)$ 时没有检查 $c\geq a$ 和 $d\geq b$。若代入方式不当，不等号的方向可能会相反。
- 在第 (ii) 部分，忘记了 $\frac{1}{q}\geq\frac{1}{p}$ 用到正数这个条件：对于负数，取倒数并不会以同样的方式把大小次序倒转。
- 只在 $x\geq y\geq z$ 的情况下证明第 (i) 部分，却没有解释为何凭对称性可得出一般情况。

STEP 之外

不等式 $(*)$ 是*排序不等式*（rearrangement inequality）在两项时的情况：若两列数都已按大小排好，则按相同次序配对时乘积之和最大，按相反次序配对时乘积之和最小。本题的两个结果，以及两项的算术—几何平均不等式（AM–GM），都可以由它直接推出；它亦与切比雪夫（Chebyshev）总和不等式密切相关。Siklos 的 *Advanced Problems in Mathematics* 有更多 STEP 风格的不等式练习，而 Bartle 与 Sherbert 的 *Introduction to Real Analysis* 则展示了这类关于大小次序的论证如何成为分析学（analysis）的基础。

详细解答

STEP I 2008，第 4 题

纯数学 · 凸函数

STEP I 2008，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/08-s1-q4

关键思路

两个函数都是指数函数（exponential），形如 $e^{h(x)}$，求导两次便得 $\left(h''+(h')^2\right)e^h$。指数因子恒为正，所以凸性（convexity）问题归结为 $h''+(h')^2$ 的正负，这是一个纯粹的三角不等式。在 (i) 中，这个不等式比较的恰好就是题目要求我们绘画的两条曲线。在 (ii) 中，它化简为 $\sin 2x$ 与 $\sin 2\alpha$ 的比较。

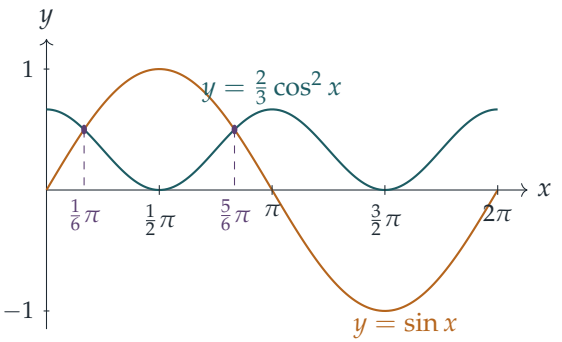
解答

第 (i) 部分

草图。曲线 $y=\frac{2}{3}\cos^2 x$ 介乎 0 与 $\frac{2}{3}$ 之间。它在 $x=0,\pi,2\pi$ 处取值 $\frac{2}{3}$，并在 $x=\frac{1}{2}\pi,\frac{3}{2}\pi$ 处与横轴相切，周期为 π。曲线 $y=\sin x$ 是常见的正弦波。要找出两者的交点，设 $s=\sin x$，并用 $\cos^2 x=1-s^2$：

$$\tfrac{2}{3}(1-s^2)=s \iff 2s^2+3s-2=0 \iff (2s-1)(s+2)=0.$$

由于 $s=-2$ 不可能成立，故 $\sin x=\frac{1}{2}$，所以两条曲线只在 $x=\frac{1}{6}\pi$ 和 $x=\frac{5}{6}\pi$ 相交。在 $(\pi,2\pi)$ 上，正弦曲线为负，而另一条曲线非负，所以该处没有交点。



f 的凸性。设 $f(x) = \mathrm{e}^{\frac{2}{3}\sin x}$，由链式法则（chain rule）得 $f'(x) = \frac{2}{3}\cos x\, f(x)$，再用乘积法则（product rule），

$$\begin{aligned} f''(x) &= -\tfrac{2}{3}\sin x\, f(x) + \tfrac{2}{3}\cos x\, f'(x) \\ &= f(x)\left(\tfrac{4}{9}\cos^2 x - \tfrac{2}{3}\sin x\right) \\ &= \tfrac{2}{3}f(x)\left(\tfrac{2}{3}\cos^2 x - \sin x\right). \end{aligned}$$

想一想 为甚么括号内的式子就能决定一切?

由于 $f(x)$ 是指数函数，它严格为正，所以 $f''(x) \geq 0$ 当且仅当 $\frac{2}{3}\cos^2 x \geq \sin x$。换言之，$f$ 为凸函数之处，恰好就是青绿色曲线位于琥珀色曲线之上或与之重合的地方。这就是题目先要求绘画草图的原因。

由草图及上面求得的交点可知，当 $0 < x \leq \frac{1}{6}\pi$ 以及 $\frac{5}{6}\pi \leq x < 2\pi$ 时，$\frac{2}{3}\cos^2 x \geq \sin x$。在 $\frac{1}{6}\pi$ 与 $\frac{5}{6}\pi$ 之间，正弦曲线在上方（例如在 $x = \frac{1}{2}\pi$ 处，两个值分别为 0 和 1）。所以 f 在

$$0 < x \leq \tfrac{1}{6}\pi \qquad \text{及} \qquad \tfrac{5}{6}\pi \leq x < 2\pi,$$

上是凸的，而且不能在更大的区间上是凸的。（端点 $\frac{1}{6}\pi$ 和 $\frac{5}{6}\pi$ 处 $f'' = 0$，是否把它们包括在内，视乎个人习惯。）

第 (ii) 部分

设 $g(x) = \mathrm{e}^{-k\tan x}$，则 $g'(x) = -k\sec^2 x\, g(x)$。利用 $\frac{\mathrm{d}}{\mathrm{d}x}\sec^2 x = 2\sec^2 x \tan x$ 及乘积法则，

$$\begin{aligned} g''(x) &= -2k\sec^2 x \tan x\, g(x) - k\sec^2 x\, g'(x) \\ &= g(x)\left(k^2\sec^4 x - 2k\sec^2 x \tan x\right) \\ &= k\sec^2 x\, g(x)\left(k\sec^2 x - 2\tan x\right). \end{aligned}$$

这里 $k = \sin 2\alpha > 0$，因为 $0 < 2\alpha < \frac{1}{2}\pi$；另外 $\sec^2 x > 0$，$g(x) > 0$。所以 $g''(x) \geq 0$ 当且仅当 $k \sec^2 x \geq 2\tan x$。两边乘以 $\cos^2 x > 0$，这就变成

$$k \geq 2\tan x \cos^2 x = 2\sin x \cos x = \sin 2x.$$

所以 g 为凸函数之处，恰好就是满足

$$\sin 2x \leq \sin 2\alpha.$$

的地方。

想一想 当 $0 < x < \frac{1}{2}\pi$ 时，$\sin 2x \leq \sin 2\alpha$ 在哪里成立?

当 x 取遍 $(0, \frac{1}{2}\pi)$ 时，$2x$ 取遍 $(0, \pi)$。在这个区间上，sin 先递增至 $\frac{1}{2}\pi$，然后递减，并且关于 $\frac{1}{2}\pi$ 对称。$\sin 2\alpha$ 这个值（其中 $0 < 2\alpha < \frac{1}{2}\pi$）在 $2x = 2\alpha$ 和 $2x = \pi - 2\alpha$ 处取得。所以 $\sin 2x \leq \sin 2\alpha$ 当且仅当 $2x \leq 2\alpha$ 或 $2x \geq \pi - 2\alpha$。即是

$$0 < x \leq \alpha \qquad \text{或} \qquad \tfrac{1}{2}\pi - \alpha \leq x < \tfrac{1}{2}\pi.$$

特别地，当 $0 < x < \alpha$ 时，我们有 $0 < 2x < 2\alpha < \frac{1}{2}\pi$，而 sin 在该处递增，所以 $\sin 2x < \sin 2\alpha = k$，且 $g''(x) > 0$。因此 g 在 $0 < x < \alpha$ 上是凸的，正如题目所求。

另一个区间是

$$\tfrac{1}{2}\pi - \alpha < x < \tfrac{1}{2}\pi.$$

它与第一个区间确实不同，因为由 $\alpha < \frac{1}{4}\pi$ 可得 $\frac{1}{2}\pi - \alpha > \frac{1}{4}\pi > \alpha$。在中间一段 $\alpha < x < \frac{1}{2}\pi - \alpha$ 上，g 不是凸的。

验算

- 在 $x = \frac{1}{6}\pi$ 处：$\frac{2}{3}\cos^2 x = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2} = \sin x$，所以该处 $f'' = 0$，这正是边界点应有的情况。
- 在 (ii) 中，问题在 $x \mapsto \frac{1}{2}\pi - x$ 之下是对称的，因为 $\sin 2x$ 保持不变。这解释了为甚么第二个区间是第一个区间的镜像。
- 数值上，取 $\alpha = 0.5$ ($k \approx 0.841$)，用有限差分（finite difference）计算 g 的二阶导数（second derivative），在 $x = 0.3$ 和 $x = 1.3$ 处为正，在 $x = 0.8$ 处为负。由于 $\frac{1}{2}\pi - \alpha \approx 1.071$，这与求得的区间吻合。用电脑测试大量随机的 α 和 x，结果与两部分的答案都一致。

常见错误

- 对 $e^{h(x)}$ 求导两次时遗漏了 $(h')^2$ 一项。二阶导数并不只是 $h''e^h$。

- 在 (i) 中漏掉了整个 $(\pi, 2\pi)$。在那里 $\sin x < 0 \leq \frac{2}{3}\cos^2 x$，所以 f 由 $\frac{5}{6}\pi$ 一直到 2π 都是凸的。
- 做到 $\sin 2x \leq \sin 2\alpha \Rightarrow x \leq \alpha$ 便停下来。由于 sin 在 $(0, \pi)$ 上并非单调（monotonic），这样会漏掉第二个区间。

STEP 之外

二阶导数判别法只是凸性的其中一面。真正的定义是几何的：每一条弦（chord）都位于图像之上或与图像重合。对可导函数而言，这等价于 f' 递增，并由此引出琴生不等式（Jensen's inequality），它是算术几何平均不等式（AM–GM）以及概率和最优化中许多结果的基础。Bartle 与 Sherbert 的 *Introduction to Real Analysis* 对凸函数有严谨的论述。Siklos 的 *Advanced Problems in Mathematics* 则提供更多 STEP 风格的练习。

详细解答

STEP I 2008，第 5 题

纯数学 · 多项式的界限

STEP I 2008，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/08-s1-q5

关键思路

当 $n = 1$ 时，多项式（polynomial）是一条直线，它在区间两个端点的值互相牵制：两者恰好相差 2，所以不可能同时都很小。第 (ii) 部分的窍门是先除以首项系数（leading coefficient），使切比雪夫（Chebyshev）定理可以直接应用于该曲线本身。定理于是迫使 $|y|$ 在 $[-1, 1]$ 内某处必须很大，而端点处的值却很小。一条曲线若升得比两个端点值都高（或降得比两者都低），就必定在中间某处转向。

解答

第 (i) 部分

$n = 1$ 的情况。这时 $\mathrm{p}(x) = x + a_0$，我们需要证明 $M \geq 2^0 = 1$。由于 M 不小于 $|\mathrm{p}|$ 在 $[-1, 1]$ 内任何一点的值，它自然不小于端点处的值：

$$M \geq |\mathrm{p}(1)| = |1 + a_0| \quad \text{及} \quad M \geq |\mathrm{p}(-1)| = |a_0 - 1|.$$

两式相加，并利用三角不等式（triangle inequality）$|u| + |v| \geq |u - v|$，

$$2M \geq |1 + a_0| + |a_0 - 1| \geq \big|(1 + a_0) - (a_0 - 1)\big| = 2,$$

所以 $M \geq 1 = 2^{1-n}$，这就证明了定理在 $n = 1$ 时成立。（当 $\mathrm{p}(x) = x$ 时等号成立，所以这个界不能再改进。）

(a) 对于 $\mathrm{p}(x) = x^2 - \frac{1}{2}$ 及 $|x| \leq 1$，我们有 $0 \leq x^2 \leq 1$，所以 $-\frac{1}{2} \leq \mathrm{p}(x) \leq \frac{1}{2}$。在 $x = \pm 1$ 处取得 $\frac{1}{2}$ 这个值（而在 $x = 0$ 处取得 $-\frac{1}{2}$），所以

$$M = \tfrac{1}{2} = 2^{1-2}.$$

定理成立，而且等号成立。

(b) 对于 $\mathrm{p}(x) = x^3 - x$，我们需要 $M \geq 2^{1-3} = \frac{1}{4}$。只要找到一个点使得 $|\mathrm{p}| \geq \frac{1}{4}$ 就足够了。驻点（stationary point）出现在 $3x^2 - 1 = 0$ 之处，即 $x = \pm\frac{1}{\sqrt{3}}$，两者都在 $[-1, 1]$ 内，而且

$$\mathrm{p}\left(-\tfrac{1}{\sqrt{3}}\right) = -\frac{1}{3\sqrt{3}} + \frac{1}{\sqrt{3}} = \frac{2}{3\sqrt{3}}.$$

由于 $\mathrm{p}(\pm 1) = 0$，这其实就是最大值：$M = \frac{2}{3\sqrt{3}} = \frac{2\sqrt{3}}{9} \approx 0.385$。若要不用计算机而与 $\frac{1}{4}$ 比较，可把两边平方：

$$\left(\frac{2}{3\sqrt{3}}\right)^2 = \frac{4}{27} > \frac{4}{64} = \frac{1}{16} = \left(\frac{1}{4}\right)^2,$$

所以 $M > \frac{1}{4}$，定理成立（这次还绰绰有余）。

第 (ii) 部分

设 $f(x) = 64x^5 + 25x^4 - 66x^3 - 24x^2 + 3x + 1$。切比雪夫定理针对的是首项系数为 1 的多项式，所以我们把它应用于

$$\mathrm{p}(x) = \tfrac{1}{64} f(x) = x^5 + \tfrac{25}{64}x^4 - \tfrac{66}{64}x^3 - \tfrac{24}{64}x^2 + \tfrac{3}{64}x + \tfrac{1}{64},$$

它具有所需的形式，其中 $n = 5$。定理给出 $\max_{|x|\leq 1} |\mathrm{p}(x)| \geq 2^{-4} = \frac{1}{16}$，再乘以 64，

$$\max_{|x|\leq 1} |f(x)| \geq 4.$$

所以存在某个 $x_0 \in [-1, 1]$ 使得 $|f(x_0)| \geq 4$。

现在看看端点：

$$f(1) = 64 + 25 - 66 - 24 + 3 + 1 = 3,$$
$$f(-1) = -64 + 25 + 66 - 24 - 3 + 1 = 1.$$

两者的模（modulus）都小于 4，所以 x_0 不是端点：$-1 < x_0 < 1$。

? **想一想** 我们知道 $|f|$ 在区间内部会变得很大。为甚么这样就必定有转向点?

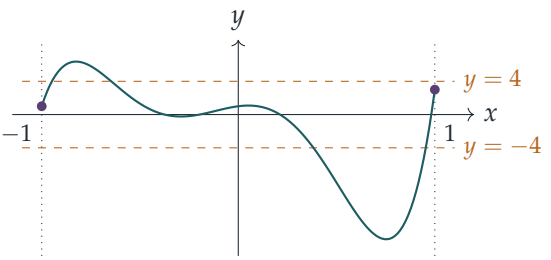
这里有两种情况。

- 若 $f(x_0) \geq 4$，则 $f(x_0)$ 比 $f(-1) = 1$ 和 $f(1) = 3$ 都大。连续函数（continuous function）f 在闭区间（closed interval）$[-1, 1]$ 上取得最大值；这个值至少是 $f(x_0) \geq 4$，所以它不是在任何一个端点取得的。因此它是在一个内点取得的，而该点就是 f 的局部极大点（local maximum）。

- 若 $f(x_0) \le -4$，则 $f(x_0)$ 低于两个端点值，同理 f 在 $[-1,1]$ 上的最小值是在一个内点取得的，而该点就是局部极小点（local minimum）。

无论哪种情况，f 都有一个局部极大点或极小点严格位于 $(-1,1)$ 之内。在该点 $f' = 0$，而由于 f 是非常数多项式，f' 只有有限个零点。所以 f' 在该点左边附近保持同一个符号，在右边附近也保持同一个符号。因为该点是局部极大点或极小点，这两个符号必定相反，所以它是一个真正的转向点（turning point）。因此曲线在 $-1 \le x \le 1$ 内至少有一个转向点，这正是所要证明的。

留意我们从头到尾都不必解四次方程 $f'(x) = 0$。这正是本题的用意。



草图显示曲线在 $[-1,1]$ 上的形状。两个端点值都位于虚线 $y = \pm 4$ 之间，但曲线却走出了这条带状区域。事实上它在区间内有四个转向点，最深的一个在 $x \approx 0.75$ 附近，该处 $y \approx -15$。

验算

- 当 $n = 1$ 时，试几个 a_0 的值便可看到 $\max(|1 + a_0|, |1 - a_0|)$ 永远不小于 1，而且只有在 $a_0 = 0$ 时才等于 1。
- 在 (b) 中，$\frac{2}{3\sqrt{3}} \approx 0.3849$，远高于 0.25。在细密的格点上对 $x^3 - x$ 取样，可确认这就是 $|x^3 - x|$ 的最大值。
- 在 (ii) 中，对 f 在 $[-1,1]$ 上作数值扫描，得到 $\max|f| \approx 15.0$（在 $x \approx 0.753$ 处，该处 $f \approx -15.0$），远高于定理所保证的 4。扫描亦发现 f' 在 $x \approx -0.825, -0.292, 0.052, 0.753$ 处变号。

常见错误

- 直接把定理应用于 f。它的首项系数是 64 而不是 1，所以必须先除以 64。所得的界变成 $64 \cdot 2^{-4} = 4$。
- 停在“某处 $|f| \ge 4$”这一步。单凭这一点并不能得出转向点：你还需要端点值 1 和 3，

才能说明极值是在区间内部取得的。

► 在 (b) 中，草率地比较 $\frac{2}{3\sqrt{3}}$ 与 $\frac{1}{4}$。把两个正数平方，便可作出准确的比较。

STEP 之外

切比雪夫定理中的等号恰好在 $2^{1-n}T_n(x)$ 时成立，其中 T_n 是由 $T_n(\cos\theta) = \cos n\theta$ 定义的切比雪夫多项式（Chebyshev polynomial）。(a) 部分就是 $\frac{1}{2}T_2$，而 $n = 1$ 的情况就是 $T_1(x) = x$。这些多项式是逼近理论（approximation theory）的起点：它们解释了为甚么在切比雪夫节点（Chebyshev nodes）上插值（interpolation）的效果最好，胜过在等距点上插值。第 (ii) 部分的论证用到极值定理（extreme value theorem，即闭区间上的连续函数必取得其上下界），这个定理会在大学第一门分析课中严格证明。Abbott 的 *Understanding Analysis*（一本分析学入门教科书）是阅读这个定理的好去处。

详细解答

STEP I 2008，第 6 题

纯数学 · 反函数与面积

STEP I 2008，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/08-s1-q6

关键思路

反函数（inverse function）的图像是原函数图像沿直线 $y = x$ 的反射（reflection），所以把 g 对 y 积分，得到的是一块横向量度的面积，即曲线 $y = \mathrm{f}(x)$ 与 y 轴之间的面积。积分上限 k 并非随意选取：它正好等于 $\mathrm{f}(\frac{1}{2})$。一旦看出这一点，两块面积便刚好拼成一个长方形。

解答

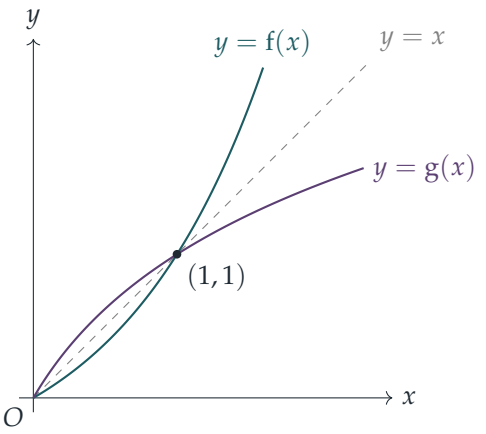
草图

函数 f 是递增（increasing）的，且 $\mathrm{f}(0) = 0$、$\mathrm{f}(1) = 1$。由于 $\mathrm{f}''(x) = \mathrm{e}^x/(\mathrm{e}-1) > 0$，它是凸（convex）的，而它在原点的斜率（gradient）为 $\mathrm{f}'(0) = 1/(\mathrm{e}-1) \approx 0.58 < 1$。因此在 $0 < x < 1$ 上，曲线位于直线 $y = x$ 之下，在 $(1,1)$ 与该直线相交，其后以指数方式增长。

要求出 g，可从 $y = \dfrac{\mathrm{e}^x - 1}{\mathrm{e}-1}$ 解出 x：

$$\mathrm{g}(x) = \ln\bigl(1 + (\mathrm{e}-1)x\bigr), \qquad x \geq 0,$$

这是因为 f 的值域（range）是 $[0, \infty)$。它的图像是 $y = \mathrm{f}(x)$ 沿 $y = x$ 的反射：它通过 $(0,0)$，在该点的斜率为 $\mathrm{e} - 1 \approx 1.72$，是凹（concave）的，在 $(1,1)$ 与 $y = x$ 相交，其后只以对数方式增长。



第一个积分

$$\int_0^{1/2} \mathrm{f}(x)\,\mathrm{d}x = \frac{1}{\mathrm{e}-1}\Big[\mathrm{e}^x - x\Big]_0^{1/2} = \frac{1}{\mathrm{e}-1}\Big(\sqrt{\mathrm{e}} - \tfrac{1}{2} - 1\Big) = \frac{\sqrt{\mathrm{e}} - \frac{3}{2}}{\mathrm{e}-1}.$$

第二个积分

首先留意

$$\mathrm{f}(\tfrac{1}{2}) = \frac{\sqrt{\mathrm{e}}-1}{\mathrm{e}-1} = \frac{\sqrt{\mathrm{e}}-1}{(\sqrt{\mathrm{e}}-1)(\sqrt{\mathrm{e}}+1)} = \frac{1}{\sqrt{\mathrm{e}}+1} = k,$$

这里用了平方差（difference of two squares）$\mathrm{e}-1 = (\sqrt{\mathrm{e}})^2 - 1$。换言之，$\mathrm{g}(k) = \frac{1}{2}$，而且 $1 + (\mathrm{e}-1)k = 1 + (\sqrt{\mathrm{e}}-1) = \sqrt{\mathrm{e}}$。

> ? **想一想** 怎样对 $\ln\big(1 + (\mathrm{e}-1)x\big)$ 积分?

作代换（substitution）$u = 1 + (\mathrm{e}-1)x$，于是 $\mathrm{d}x = \mathrm{d}u/(\mathrm{e}-1)$，积分上下限变为 $u = 1$ 和 $u = \sqrt{\mathrm{e}}$。然后利用 $\int \ln u\,\mathrm{d}u = u\ln u - u$（把被积函数看成 $\ln u$ 乘以 1，再用分部积分法（integration by parts））：

$$\begin{aligned}\int_0^k \mathrm{g}(x)\,\mathrm{d}x &= \frac{1}{\mathrm{e}-1}\int_1^{\sqrt{\mathrm{e}}} \ln u\,\mathrm{d}u = \frac{1}{\mathrm{e}-1}\Big[u\ln u - u\Big]_1^{\sqrt{\mathrm{e}}} \\ &= \frac{1}{\mathrm{e}-1}\Big(\tfrac{1}{2}\sqrt{\mathrm{e}} - \sqrt{\mathrm{e}} - (0-1)\Big) = \frac{1 - \frac{1}{2}\sqrt{\mathrm{e}}}{\mathrm{e}-1},\end{aligned}$$

其中用了 $\ln\sqrt{\mathrm{e}} = \frac{1}{2}$。

相加

$$\int_0^{1/2} \mathrm{f}(x)\,\mathrm{d}x + \int_0^k \mathrm{g}(x)\,\mathrm{d}x = \frac{\sqrt{\mathrm{e}} - \frac{3}{2} + 1 - \frac{1}{2}\sqrt{\mathrm{e}}}{\mathrm{e}-1} = \frac{\frac{1}{2}(\sqrt{\mathrm{e}}-1)}{(\sqrt{\mathrm{e}}-1)(\sqrt{\mathrm{e}}+1)} = \frac{1}{2(\sqrt{\mathrm{e}}+1)},$$

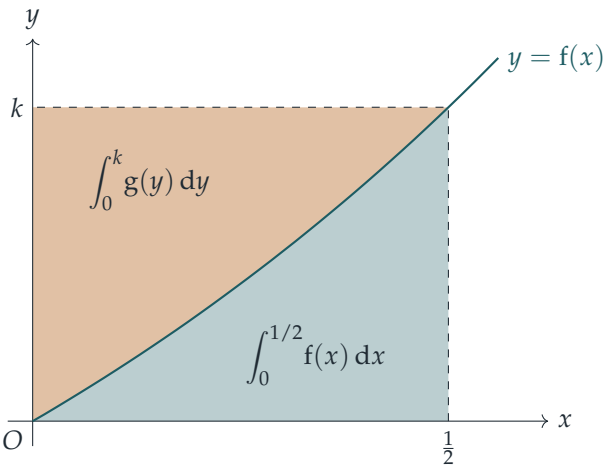
这正是所要求的结果。

解释这个结果

答案是 $\frac{1}{2}\times k$，即以 $(0,0)$ 和 $(\frac{1}{2},k)$ 为对角顶点的长方形的面积。原因如下。

> **想一想** 在 f 的图像上，$\int_0^k \mathrm{g}(x)\,\mathrm{d}x$ 在哪里?

沿 $y=x$ 反射不会改变面积。$y=\mathrm{g}(x)$ 之下、$0\le x\le k$ 的区域，经反射后变成曲线 $y=\mathrm{f}(x)$ 与 y 轴之间、$0\le y\le k$ 的区域。换句话说，它就是沿曲线 $y=\mathrm{f}(x)$ 的 $\int_0^k x\,\mathrm{d}y$：曲线*左边*的面积。由于 $\mathrm{f}(\frac{1}{2})=k$，曲线由长方形的顶点 $(0,0)$ 走到顶点 $(\frac{1}{2},k)$，把长方形分成两块：曲线之下的面积（第一个积分，蓝绿色）和曲线左边的面积（第二个积分，琥珀色）。



因此两个积分合起来正好是整个长方形，面积为 $\frac{1}{2}k=\dfrac{1}{2(\sqrt{\mathrm{e}}+1)}$。同一幅图还给出一个一般结论：对于满足 $\mathrm{f}(0)=0$ 的递增函数，

$$\int_0^a \mathrm{f}(x)\,\mathrm{d}x+\int_0^{\mathrm{f}(a)}\mathrm{f}^{-1}(y)\,\mathrm{d}y=a\,\mathrm{f}(a).$$

验算

以数值计算，第一个积分约为 0.08655，第二个约为 0.10222；对两者分别使用辛普森法则（Simpson's rule），结果与精确式子吻合至小数点后十二位，而两者之和 0.18877 等于 $1/\big(2(\sqrt{\mathrm{e}}+1)\big)$。再用草图作个合理性检查：两块面积都是正数，而且蓝绿色的一块较小，这是理所当然的，因为凸曲线下垂至长方形对角线之下。

常见错误

- 没有留意到 $k = \mathrm{f}(\frac{1}{2})$，因而错过了 $1 + (\mathrm{e} - 1)k = \sqrt{\mathrm{e}}$ 这个化简。
- 把 f 在原点附近画在 $y = x$ 之上：它在该处的斜率是 $1/(\mathrm{e} - 1) < 1$。
- 作图时，在另一幅图上把 $y = \mathrm{g}(x)$ 之下的面积涂上阴影，却没有说明它与 $y = \mathrm{f}(x)$ 旁边的区域有何关系；关键在于反射把其中一个区域变成另一个。

STEP 之外

恒等式 $\int_0^a \mathrm{f} + \int_0^{\mathrm{f}(a)} \mathrm{f}^{-1} = a\,\mathrm{f}(a)$ 是杨氏不等式（*Young's inequality*）中等号成立的情况：对任意 $b \geq 0$，$\int_0^a \mathrm{f} + \int_0^b \mathrm{f}^{-1} \geq ab$。取 $\mathrm{f}(x) = x^{p-1}$，便得到 $ab \leq a^p/p + b^q/q$，其中 $1/p + 1/q = 1$，这是分析学中赫尔德不等式（Hölder's inequality）的起点。想知道这类积分如何严格地建立起来，Abbott 的 *Understanding Analysis*（一本实分析入门书）是很好的读物。

详细解答

STEP I 2008，第 7 题

纯数学 · 旋转与距离

> **STEP I 2008，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/08-s1-q7

关键思路

要求一点到一条过原点直线的垂直距离（perpendicular distance），只要把平面旋转（rotation），令该直线落在 x 轴上，问题就变得容易：所求距离正是新 y 坐标的大小。第一部分的旋转公式，正好替三条边中的两条完成这件事。至于最后一部分，窍门是去掉绝对值符号（modulus signs）：三个有向距离（signed distance）之和永远是同一个常数，而取绝对值只会令总和变大。

解答

旋转公式

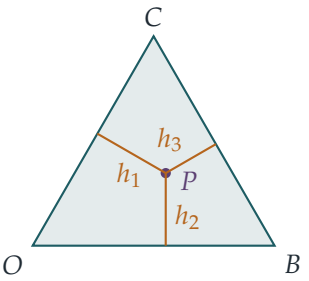
设 $x = r\cos\theta$，$y = r\sin\theta$，其中 $r = OP$，而 θ 是由正 x 轴量至 OP 的角。把 OP 顺时针（clockwise）旋转 60°，r 保持不变，角则变为 $\theta - 60^\circ$，所以新的 y 坐标是

$$\begin{aligned} r\sin(\theta - 60^\circ) &= r\sin\theta\cos 60^\circ - r\cos\theta\sin 60^\circ \\ &= \tfrac{1}{2}y - \tfrac{\sqrt{3}}{2}x = \tfrac{1}{2}\left(y - \sqrt{3}\,x\right), \end{aligned}$$

这正是所要证明的结果。若改为逆时针（anticlockwise）旋转，角变为 $\theta + 60^\circ$，同样展开，只是 $\sin 60^\circ$ 的符号相反，便得新的 y 坐标

$$\tfrac{1}{2}\left(y + \sqrt{3}\,x\right).$$

三个距离



h_1。直线 OC 与 x 轴的夹角为 $60°$。把整幅图绕 O 顺时针旋转 $60°$，OC 便落在 x 轴上，而距离保持不变。在旋转后的图中，P 到 x 轴的距离就是其新 y 坐标的绝对值，所以

$$h_1 = \tfrac{1}{2}\left| y - \sqrt{3}\,x \right|.$$

h_2。直线 OB 就是 x 轴，所以 $h_2 = |y|$。

h_3。现在绕 O 逆时针旋转 $60°$。这个旋转把 $B = (1,0)$ 送到 $(\frac{1}{2}, \frac{1}{2}\sqrt{3})$，并把 C（位于角 $60°$ 处）送到角 $120°$ 处的点，即 $(-\frac{1}{2}, \frac{1}{2}\sqrt{3})$。因此直线 BC 变成水平线 $y = \frac{1}{2}\sqrt{3}$，而 P 的 y 坐标则变为 $\frac{1}{2}(y + \sqrt{3}\,x)$。于是

$$h_3 = \left| \tfrac{1}{2}(y + \sqrt{3}\,x) - \tfrac{1}{2}\sqrt{3} \right| = \tfrac{1}{2}\left| \sqrt{3}\,x + y - \sqrt{3} \right|.$$

（换个做法：BC 的方程是 $\sqrt{3}\,x + y = \sqrt{3}$，由于 $\sqrt{3+1} = 2$，用惯常的距离公式亦得出相同答案。）

最后一部分

定义以下带正负号的量：

$$s_1 = \tfrac{1}{2}(\sqrt{3}\,x - y), \qquad s_2 = y, \qquad s_3 = \tfrac{1}{2}(\sqrt{3} - \sqrt{3}\,x - y),$$

使得对 $i = 1,2,3$ 均有 $h_i = |s_i|$。把三者相加，

$$s_1 + s_2 + s_3 = \tfrac{1}{2}\sqrt{3}\,x - \tfrac{1}{2}y + y + \tfrac{1}{2}\sqrt{3} - \tfrac{1}{2}\sqrt{3}\,x - \tfrac{1}{2}y = \tfrac{1}{2}\sqrt{3}$$

对每一点 (x,y) 都成立。

想一想 s_1、s_2、s_3 的正负号与三角形有甚么关系？

每个 s_i 都是一个线性式，并且在三角形的其中一条边上等于零。以位于三角形内部的形心（centroid）$(\frac{1}{2}, \frac{1}{6}\sqrt{3})$ 作测试，得 $s_1 = s_2 = s_3 = \frac{1}{6}\sqrt{3} > 0$。所以 $s_2 \geq 0$ 是 OB 包含 C

的一侧，$s_1 \geq 0$ 是 OC 包含 B 的一侧，而 $s_3 \geq 0$ 是 BC 包含 O 的一侧。闭三角形 OBC 正是这三个半平面（half-plane）的交集，所以

$$P \text{ 落在（含边界的）} OBC \iff s_1 \geq 0,\ s_2 \geq 0,\ s_3 \geq 0.$$

对任意实数 s，我们有 $|s| \geq s$，等号成立当且仅当（if and only if）$s \geq 0$。因此

$$h_1 + h_2 + h_3 = |s_1| + |s_2| + |s_3| \ \geq\ s_1 + s_2 + s_3 = \tfrac{1}{2}\sqrt{3},$$

等号成立当且仅当对每个 i 都有 $|s_i| = s_i$，亦即当且仅当三个 s_i 全部非负。根据上一段，这恰好在 P 位于三角形 OBC 的边上或内部时发生。这样便一次过证明了两个方向：

$$h_1 + h_2 + h_3 = \tfrac{1}{2}\sqrt{3} \iff P \text{ 落在（含边界的）三角形} OBC.$$

还要留意，在三角形外，这个和严格大于 $\frac{1}{2}\sqrt{3}$。

验算

- 在顶点上，例如 $P = B = (1,0)$：$h_1 = \frac{1}{2}\sqrt{3}$，$h_2 = 0$，$h_3 = 0$，和为 $\frac{1}{2}\sqrt{3}$。正确。
- 在形心处，每个距离都是 $\frac{1}{6}\sqrt{3}$，三个加起来是 $\frac{1}{2}\sqrt{3} \approx 0.8660$，即三角形的高。
- 在三角形外，取 $P = (2,0)$：$h_1 = \sqrt{3}$，$h_2 = 0$，$h_3 = \frac{1}{2}\sqrt{3}$，和为 $\frac{3}{2}\sqrt{3} > \frac{1}{2}\sqrt{3}$。
- 我们在 10^5 个随机点上，把三条公式与直接计算的点到直线距离作比较（吻合至 10^{-15}），并在 2×10^5 个点（包括许多位于边上的点）上测试了“当且仅当”的结论，没有发现任何例外。

常见错误

- 弄错旋转的正负号：顺时针表示角减小，得 $\theta - 60°$。
- 只证明一个方向。验证了在三角形内和为 $\frac{1}{2}\sqrt{3}$ 并不足够；你还必须证明在三角形外和不是 $\frac{1}{2}\sqrt{3}$。不等式 $|s| \geq s$ 可以一次过处理两个方向。
- 忘记 h_1、h_2、h_3 中的绝对值符号：距离永远不会是负数。

STEP 之外

这就是维维亚尼定理（Viviani's theorem）：在等边三角形（equilateral triangle）中，内部任意一点到三边的距离之和等于三角形的高。有向距离 s_1, s_2, s_3（相差一个比例常数）就是 P 的重心坐标（barycentric coordinates）或三线坐标（trilinear coordinates），它

们的和永远是常数，并且恰好在三角形内全部非负；它们是几何、电脑图形学和凸分析中的标准工具。旋转公式是 2×2 旋转矩阵（rotation matrix）的第二行，而 Strang 的 *Introduction to Linear Algebra*（线性代数入门）是认识如何把旋转看成线性映射（linear map）的好去处。

详细解答

STEP I 2008，第 8 题

纯数学 · 克莱罗微分方程

STEP I 2008，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/08-s1-q8

关键思路

两条方程都含有 y' 而不含 y''，而且都不属于可以直接分离变量（separate variables）或直接积分的类型。窍门是再求导一次：只含 y' 的项会互相抵消，余下的式子可以因式分解（factorise），其中一个因式是 y''。每个因式对应一种情况，而在每种情况中，我们都代回原来的方程，而不是进行积分，这样任意常数（arbitrary constant）便不费工夫地确定下来。

解答

以下一律记 $p = y'$，因此 $p' = y''$。

第 (i) 部分

把 $p^2 - xp + y = 0$ 对 x 求导，其中 xp 一项用积法则（product rule）：

$$2p\,y'' - \left(p + x\,y''\right) + p = 0 \quad \Longrightarrow \quad y''\,(2p - x) = 0.$$

两个 p 项互相抵消，所以在每一点上，或者 $y'' = 0$，或者 $2y' = x$，正如题目所求。

情况 $y'' = 0$。这时 y' 是常数 m，而 $y = mx + c$。代入 $(*)$：

$$m^2 - xm + (mx + c) = 0 \quad \Longrightarrow \quad c = -m^2.$$

所以曲线是直线 $y = mx - m^2$。

情况 $2y' = x$。把 $p = \frac{1}{2}x$ 代入 $(*)$：

$$\tfrac{1}{4}x^2 - \tfrac{1}{2}x^2 + y = 0 \quad \Longrightarrow \quad 4y = x^2.$$

想一想 为甚么不直接把 $y' = \frac{1}{2}x$ 积分，得出 $y = \frac{1}{4}x^2 + k$?

这样做也可以，但对 $(*)$ 求导时丢失了一些资料：任何 $y = \frac{1}{4}x^2 + k$ 都满足 $2y' = x$，却只有 $k = 0$ 才满足 $(*)$ 本身（例如 $k = 1$ 时，在 $x = 0$ 处得 $0 - 0 + 1 \neq 0$）。直接代回 $(*)$ 可一次过用尽所有资料，并且直接得出常数的值。第一种情况中 $c = -m^2$，也是同一道理。

所以曲线或者是直线 $y = mx + c$，其中 $c = -m^2$，或者是抛物线（parabola）$4y = x^2$。

这两个答案互有关连：直线 $y = mx - m^2$ 与 $4y = x^2$ 相交于 $x^2 - 4mx + 4m^2 = 0$ 之处，即 $(x - 2m)^2 = 0$，所以每条直线都是抛物线在 $(2m, m^2)$ 的切线（tangent）。这条抛物线就是这族直线的包络线（envelope）。

第 (ii) 部分

把 $(x^2 - 1)p^2 - 2xyp + y^2 - 1 = 0$ 对 x 求导：

$$2xp^2 + 2(x^2 - 1)p\,y'' - 2yp - 2xp^2 - 2xy\,y'' + 2yp = 0.$$

不含 y'' 的项两两抵消，余下

$$2y''\big[(x^2 - 1)p - xy\big] = 0.$$

情况 $y'' = 0$。这时 $y = mx + c$。留意到方程可以改写成

$$(y - xp)^2 = 1 + p^2,$$

会方便得多，因为 $(y - xp)^2 - p^2 - 1 = y^2 - 2xyp + (x^2 - 1)p^2 - 1$。对于这条直线，$y - xp = c$ 而 $p = m$，所以

$$c^2 = 1 + m^2, \qquad \text{即} \qquad y = mx \pm \sqrt{1 + m^2}.$$

情况 $(x^2 - 1)p = xy$。当 $x^2 \neq 1$ 时，有 $p = \dfrac{xy}{x^2 - 1}$，所以

$$y - xp = y - \frac{x^2 y}{x^2 - 1} = \frac{-y}{x^2 - 1}.$$

代入 $(y - xp)^2 = 1 + p^2$，再乘以 $(x^2 - 1)^2$：

$$\begin{aligned} y^2 &= (x^2 - 1)^2 + x^2 y^2 \\ y^2(1 - x^2) &= (1 - x^2)^2 \\ y^2 &= 1 - x^2, \end{aligned}$$

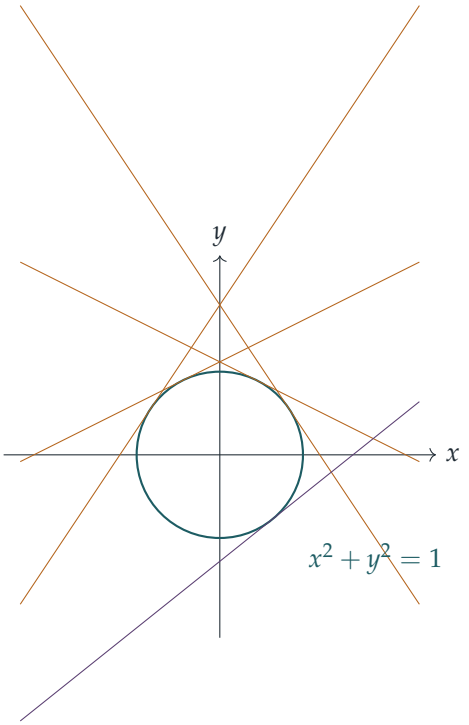
最后一步除以 $1-x^2 \neq 0$。所以曲线是圆 $x^2+y^2=1$。(在 $x=\pm 1$ 处, 条件 $(x^2-1)p=xy$ 迫使 $y=0$，得出点 $(\pm 1,0)$，它们也在同一个圆上。)

反过来，这个圆确实满足方程：在圆上 $p=-x/y$，所以 $y-xp=(y^2+x^2)/y=1/y$，而 $1+p^2=(x^2+y^2)/y^2=1/y^2$，两者吻合。

所以曲线或者是直线 $y=mx+c$,其中 $c^2=1+m^2$,或者是单位圆(unit circle)$x^2+y^2=1$。

想一想 直线 $y=mx\pm\sqrt{1+m^2}$ 与这个圆有甚么关系?

原点到 $y=mx+c$ 的距离是 $|c|/\sqrt{1+m^2}$，在这里刚好等于 1。所以这些直线正好就是单位圆的所有（非铅垂）切线，一如 (i) 中的直线是抛物线的切线。



验算

- 在 (i) 中取 $m=1$：$y=x-1$ 有 $p=1$，而 $1-x+x-1=0$。它与 $4y=x^2$ 相切于 $(2,1)$。
- 在 (ii) 中，圆上的点 $(\frac{3}{5},\frac{4}{5})$ 有 $p=-\frac{3}{4}$；于是 $(x^2-1)p^2-2xyp+y^2-1=-\frac{16}{25}\cdot\frac{9}{16}+\frac{18}{25}+\frac{16}{25}-1=0$。
- 数值方面, 我们沿随机选取的三次曲线检验了两个求导后的因式分解, 并确认直线、抛

物线和圆所得的残差（residual）均为 0（准确至 10^{-13}），而 $y = \frac{1}{4}x^2 + 1$ 和 $x^2 + y^2 = 4$ 则不然。

常见错误

- 把 $y' = \frac{1}{2}x$ 或 $y'' = 0$ 积分后留下一个任意常数：这个常数必须由原来的方程确定。
- 漏掉 $c = \pm\sqrt{1 + m^2}$ 中的 $\pm$：每个斜率（gradient）都有两条互相平行的切线。

STEP 之外

第 (i) 部分是一条*克莱罗方程*（Clairaut equation）$y = xp - f(p)$：它的通解（general solution）是直线族 $y = mx - f(m)$，此外还有一个*奇解*（singular solution），即该直线族的包络线，无论怎样选取常数都得不出它。第 (ii) 部分是同一现象，$y = xp \pm \sqrt{1 + p^2}$，包络线是那个圆。奇解说明，一阶方程（first-order equation）经过每一点的解不一定唯一；唯一性何时成立，正是大学第一门常微分方程（ODE）课程中存在与唯一性定理（existence and uniqueness theorems）所探讨的课题。Boyce 与 DiPrima 的 *Elementary Differential Equations*（常微分方程入门教科书）是阅读这方面内容的好去处。

详细解答

STEP I 2008，第 9 题

力学 · 滚动圆环的动量

STEP I 2008，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/08-s1-q9

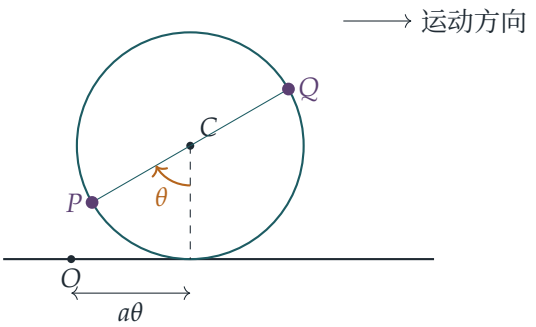
关键思路

纯滚动（rolling without slipping）把平移与转动连系起来：当圆环转过 θ 时，圆心移动的距离为 $a\theta$，即曾经接触桌面的那段环边的长度。因此每个质点的位置都是“圆心加上一条转动的半径”，求导便得出速度。至于最后一部分，留意外力没有水平分量，所以总水平动量（momentum）不会改变，而这个动量原来是 $\dot{\theta}$ 的倍数。

解答

P 的位置与速率

把 P 和 Q 视为轻圆环上的质点（particle）。取 P 的起始点为原点，x 轴沿滚动的直线（指向圆环前进的方向），y 轴垂直向上。设圆环沿正 x 方向滚动，因此在下图中它是顺时针转动的。



由于圆环没有滑动，滚到桌面上的那段环边弧长等于圆心移动的距离，所以圆心 C 位于

$(a\theta,\, a)$。起初 CP 指向正下方；顺时针转过 θ 之后，向量 $\overrightarrow{CP}$ 为 $(-a\sin\theta,\, -a\cos\theta)$。因此

$$\boxed{P = \big(a\theta - a\sin\theta,\; a - a\cos\theta\big)}$$

（这是一条摆线（cycloid））。利用链式法则（chain rule）对 t 求导，

$$\dot{x}_P = a\dot{\theta}(1-\cos\theta), \qquad \dot{y}_P = a\dot{\theta}\sin\theta,$$

所以

$$v_P^2 = a^2\dot{\theta}^2\big[(1-\cos\theta)^2 + \sin^2\theta\big] = 2a^2\dot{\theta}^2(1-\cos\theta) = 4a^2\dot{\theta}^2\sin^2\tfrac{1}{2}\theta,$$

其中用到 $1-\cos\theta = 2\sin^2\frac{1}{2}\theta$。因此 P 的速率为

$$v_P = 2a\left|\dot{\theta}\sin\tfrac{1}{2}\theta\right| \qquad \big(= 2a\dot{\theta}\sin\tfrac{1}{2}\theta,\ \text{当} 0 \le \theta \le 2\pi,\ \dot{\theta} \ge 0\big).$$

当 θ 是 2π 的倍数时，这个速率为零，而这正好是 P 成为接触点的时刻，与纯滚动的要求一致。

动能

Q 位于直径的另一端，所以 $\overrightarrow{CQ} = -\overrightarrow{CP}$，而

$$Q = \big(a\theta + a\sin\theta,\; a + a\cos\theta\big), \qquad \dot{x}_Q = a\dot{\theta}(1+\cos\theta), \quad \dot{y}_Q = -a\dot{\theta}\sin\theta.$$

同理，

$$v_Q^2 = a^2\dot{\theta}^2\big[(1+\cos\theta)^2 + \sin^2\theta\big] = 2a^2\dot{\theta}^2(1+\cos\theta) = 4a^2\dot{\theta}^2\cos^2\tfrac{1}{2}\theta.$$

圆环是轻的，所以动能（kinetic energy）只来自两个质点：

$$T = \tfrac{1}{2}mv_P^2 + \tfrac{1}{2}mv_Q^2 = 2ma^2\dot{\theta}^2\big(\sin^2\tfrac{1}{2}\theta + \cos^2\tfrac{1}{2}\theta\big) = 2ma^2\dot{\theta}^2,$$

正如所求。

速率恒定

外力包括两个质点的重量 mg 和桌面的反作用力（reaction），全部都是垂直的。因此系统所受的总水平力为零，而把牛顿第二定律（Newton's second law）应用于整个系统，可知其总水平动量恒定。（圆环是轻的，所以它不带动量。）

> ? **想一想** 质点也有垂直方向的运动，为甚么只考虑水平动量就足够?

因为水平动量恰好只取决于 $\dot{\theta}$。把水平速度相加，

$$m\dot{x}_P + m\dot{x}_Q = ma\dot{\theta}(1-\cos\theta) + ma\dot{\theta}(1+\cos\theta) = 2ma\dot{\theta}.$$

$\cos\theta$ 项互相抵消：系统的质心（centre of mass）是 C，它始终以速度 $a\dot{\theta}$ 水平移动。由于 $2ma\dot{\theta}$ 恒定，$\dot{\theta}$ 亦恒定，所以圆环的速率（即其圆心的速率 $a\dot{\theta}$）恒定。 □

或用能量方法。 P 和 Q 的高度之和为 $(a - a\cos\theta) + (a + a\cos\theta) = 2a$，所以势能（potential energy）$2mga$ 永不改变。反作用力作用于接触点，而纯滚动时接触点瞬时静止，所以反作用力不作功。因此 $T = 2ma^2\dot{\theta}^2$ 恒定，即 $|\dot{\theta}|$ 恒定；又因为 $\dot{\theta}$ 连续地变化，不可能不经过 0 而由一个符号跳到另一个符号，所以 $\dot{\theta}$ 本身恒定。动量的论证省却了这最后一步，这就是我们先用它的原因。

验算

- 当 $\theta = \pi$ 时，P 在最高点：公式给出 $P = (a\pi, 2a)$，速率为 $2a\dot{\theta}$，是圆心速率的两倍，这正是滚动车轮顶点的熟悉结果。
- 动能等于 $\frac{1}{2}(2m)(a\dot{\theta})^2 + \frac{1}{2}(2ma^2)\dot{\theta}^2$：质心的平移加上绕质心的转动，其中转动惯量（moment of inertia）为 $2ma^2$。两项各为 $ma^2\dot{\theta}^2$。
- 数值上，取 $a = 1.3$、$\dot{\theta} = 2.1$ 和 $\theta = 1$，以有限差分（finite differences）对位置求导，得 $v_P^2 + v_Q^2 = 4a^2\dot{\theta}^2 \approx 29.8$；而当 $\dot{\theta}$ 恒定时，P 和 Q 的水平加速度互相抵消。

常见错误

- 把圆心放在 (a, a)，或者完全忘记圆心会移动：$a\theta$ 这一项正是滚动带来的。
- 在 P 的位置中弄错 $\sin\theta$ 的正负号；应检查 P 起初的速度是否为零。
- 由动能恒定推断速率恒定，却没有说明为甚么 $\dot{\theta}$ 不能改变符号。

STEP 之外

这是一个双质点的例子，说明了一个普遍的事实：系统的总动量只会因外力而改变，内力（这里由圆环传递）则成对抵消。“验算”中动能的分拆就是柯尼希定理（König's theorem）：动能等于质心的动能加上相对于质心运动的动能,这是刚体动力学(rigid-body dynamics)的起点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对滚动运动和这些概念有细致的论述。

详细解答

STEP I 2008，第 10 题

力学 • 抛体运动

STEP I 2008，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/08-s1-q10

关键思路

重力只沿铅垂方向作用，所以水平速度 $V\cos\alpha$ 始终不变，射程（range）就是 $V\cos\alpha$ 乘以总飞行时间（time of flight）。飞行时间完全由铅垂方向的运动决定，而这段运动可分成三个阶段：向上穿过底层、在高度 h 以上的一段往返，以及向下穿过底层返回地面。每个阶段都是匀加速（constant acceleration）问题，而大部分工夫都可以交给对称性（symmetry）代劳。

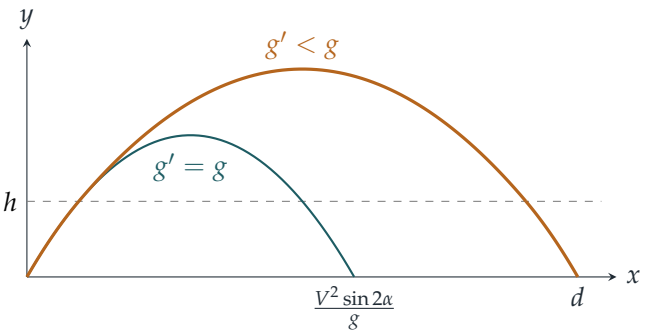
解答

我们把粒子（particle）看成一个质点，不计空气阻力，在平地上方的一个铅垂平面内运动。取 x 轴为水平方向，y 轴铅垂向上，原点设在抛射点。

草图

在高度 h 以下，两条轨迹（trajectory）服从完全相同的方程，所以在该处两者重合。条件 $V^2\sin^2\alpha > 2gh$ 表示初始的铅垂动能（kinetic energy）足以在加速度 g 之下升过高度 h，所以粒子确实会越过 $y = h$。

在 h 以上，$g' < g$ 的情况重力较弱，铅垂速率减少得较慢：粒子升得更高，在上面逗留得更久，描出一条较扁、较阔的抛物线（parabola）。速度在 $y = h$ 处连续（只有加速度出现跳跃），所以两段弧线平滑相接，没有尖角，但曲率（curvature）会突然改变。下降时，粒子重新进入底层，其速率和角度与离开时相同（由顶部弧线的对称性可知），而最后一段弧线正是第一段的镜像。因此 $g' < g$ 的路径落在更远的地方。



射程

初速度的铅垂分量（vertical component）为 $V\sin\alpha$，而当 $y<h$ 时加速度为 $-g$。

第 1 阶段：上升至高度 h。设 w 为粒子到达 $y=h$ 时的(向上)铅垂速率。由 $v^2=u^2+2as$，

$$w^2=V^2\sin^2\alpha-2gh=\sin^2\alpha\left(V^2-2gh\operatorname{cosec}^2\alpha\right)=V'^2\sin^2\alpha,$$

所以 $w=V'\sin\alpha$；这里取正根，因为 $\sin\alpha>0$，而且粒子仍在上升。（题目给出的条件正好保证 V' 是正实数。）由 $v=u+at$，所需时间为

$$t_1=\frac{V\sin\alpha-V'\sin\alpha}{g}=\frac{(V-V')\sin\alpha}{g}.$$

第 2 阶段：在高度 h 以上。此时粒子从 $y=h$ 出发，以速率 w 向上运动，加速度为 $-g'$。当 $wt-\frac{1}{2}g't^2=0$ 时，它回到 $y=h$，即经过时间

$$t_2=\frac{2w}{g'}=\frac{2V'\sin\alpha}{g'},$$

之后，而这时它正以铅垂速率 w 向下运动。

想一想 为甚么最后一个阶段所需的时间与第一个阶段相同?

在第 3 阶段，粒子从 $y=h$ 出发，以速率 w 向下运动，加速度为向下的 g，一直落到 $y=0$。这正是把第 1 阶段在时间上倒过来播放：匀加速运动是可逆的，而第 1 阶段是在同样的加速度下，由 $y=0$、速率 $V\sin\alpha$ 运动到 $y=h$、速率 w。所以粒子经过时间 $t_3=t_1$ 后到达地面，铅垂速率为 $V\sin\alpha$。（直接计算：铅垂速率以 g 的变率由 w 增至 $V\sin\alpha$，需时 $(V\sin\alpha-w)/g$。）

因此总飞行时间为

$$T=2t_1+t_2=2\sin\alpha\left(\frac{V-V'}{g}+\frac{V'}{g'}\right).$$

由于没有水平方向的力，水平速度自始至终保持为 $V\cos\alpha$，于是

$$\begin{aligned} d = V\cos\alpha \cdot T &= \left(\frac{V-V'}{g} + \frac{V'}{g'}\right) 2V\sin\alpha\cos\alpha \\ &= \left(\frac{V-V'}{g} + \frac{V'}{g'}\right) V\sin 2\alpha, \end{aligned}$$

其中用到 $2\sin\alpha\cos\alpha = \sin 2\alpha$。这就是所要求的结果。

验算

- 若 $g' = g$，括号内的式子变成 V/g，于是 $d = V^2\sin 2\alpha/g$，即我们熟悉的地球上的射程。
- 若 $g' < g$，则 $V'/g' > V'/g$，所以 d 大于均匀重力下的射程，与草图吻合。
- 若恰好有 $V^2\sin^2\alpha = 2gh$，则 $V' = 0$：粒子刚好触及 $y = h$，完全感受不到 g'，于是同样得到 $d = V^2\sin 2\alpha/g$。
- 数值上，取 $V = 10$、$\alpha = 60°$、$g = 10$、$h = 2$ 及 $g' = 5$，可得 $V' = \sqrt{140/3} \approx 6.83$ 及 $d \approx 14.576$；以数值方法逐步模拟这个运动，所得的落点相同，相对误差在 10^{-5} 以内。（草图所用的正是这些数值。）

常见错误

- 忘记了向下穿过底层的那一段，得出 $t_1 + t_2$ 而不是 $2t_1 + t_2$。
- 把高度 h 处的铅垂速率当作 V' 而不是 $V'\sin\alpha$：V' 是按 $\operatorname{cosec}\alpha$ 放大了的速率，并不是铅垂分量本身。
- 在两段弧线相接处画出尖角：速度是连续的，所以接合处是平滑的。

STEP 之外

重力随位置而变，其实是常态而不是例外：真实的重力随着与行星中心的距离按 $1/r^2$ 减弱，抛体的路径也因而变成椭圆（ellipse）的弧，而不是抛物线。把一个运动分成若干个力的规律都很简单的区域，再在边界上令位置和速度互相衔接，是一种标准技巧；在粒子的折射（refraction），以及量子力学（quantum mechanics）中的阶跃势（step potential）问题里，都会再次用到。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）既讲解匀加速运动的各种工具，也讨论平方反比（inverse-square）重力下的轨道运动。

详细解答

STEP I 2008，第 11 题

力学 · 极限摩擦力平衡

STEP I 2008，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/08-s1-q11

关键思路

由于 $\angle P_1OP_2$ 是直角，半径 OP_2 平行于 P_1 处的切线（tangent），而半径 OP_1 则平行于 P_2 处的切线。因此，只要沿这两条半径的方向分解（resolve）各力，每条方程便只涉及一个法向反作用力（normal reaction）和一个摩擦力（friction）。再加上一条对 O 取力矩（moment）的方程（两个法向反作用力对该点都没有力矩），并在两端代入极限摩擦（limiting friction）的条件 $F = \mu R$，便得到三条方程，从中可以消去 α 以外的未知量。最后一部分其实是乔装了的倍角公式（double-angle formulae）。

解答

模型与各力

把杆视为均匀（uniform）的刚性杆，质心（centre of mass）G 位于杆的中点；把两个环视为轻环，并设金属线圈的半径为 a。由于环是轻的，线圈作用于每个环的力会原封不动地传到杆上，所以可以当作杆在 P_1 和 P_2 两点直接与线圈接触。杆的每一端都受到一个沿半径方向的法向反作用力和一个沿切线方向的摩擦力。

取 O 为原点，x 轴向右，y 轴向上，并设 P_1 在线圈最低点的左方。由于 $\angle P_1OP_2 = 90^\circ$，而 P_1 与向下的铅垂线成角 α，所以 OP_2 在另一侧与该铅垂线成角 $90^\circ - \alpha$：

$$P_1 = a(-\sin\alpha, -\cos\alpha), \qquad P_2 = a(\cos\alpha, -\sin\alpha).$$

因为 $0 < \alpha < 45^\circ$，$\cos\alpha > \sin\alpha$，所以 P_1 确实比 P_2 低，而且两点都在 O 的下方。

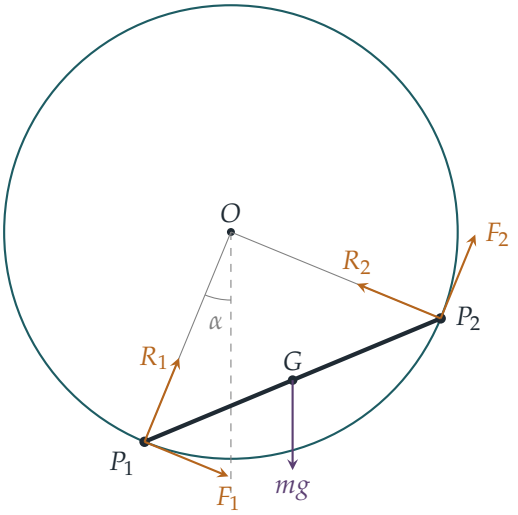
中点为 $G = \frac{a}{2}(\cos\alpha - \sin\alpha, -\cos\alpha - \sin\alpha)$，位于通过 O 的铅垂线的*右方*。因此重量对 O 的力矩是顺时针方向的，杆即将沿线圈顺时针滑动（P_2 向下移，P_1 向上移）。摩擦力

阻碍这个运动，所以它在两端都沿逆时针方向作用：在 P_1 处沿线圈向下，在 P_2 处沿线圈向上。

设 $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 分别为由 P_1 和 P_2 指向 O 的单位向量（unit vector）：

$$\mathbf{u}_1 = (\sin\alpha, \cos\alpha), \qquad \mathbf{u}_2 = (-\cos\alpha, \sin\alpha), \qquad \mathbf{u}_1 \cdot \mathbf{u}_2 = 0.$$

P_1 处的逆时针切线方向是 $-\mathbf{u}_2$，而 P_2 处的则是 $\mathbf{u}_1$。设法向反作用力（指向 O）为 R_1, R_2，摩擦力为 F_1, F_2。



平衡方程

重量 mg 铅垂向下作用，所以它沿 $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 方向的分量分别是 $-mg\cos\alpha$ 和 $-mg\sin\alpha$。

沿 $\mathbf{u}_1$ 方向（即沿 P_1O）分解：R_1 和 F_2 有贡献，而 R_2 和 F_1 则与这个方向垂直。所以

$$R_1 + F_2 = mg\cos\alpha. \tag{1}$$

沿 $\mathbf{u}_2$ 方向（沿 P_2O）分解：R_2 的贡献是 $+R_2$，而 F_1 指向 $-\mathbf{u}_2$ 方向，贡献是 $-F_1$。所以

$$R_2 - F_1 = mg\sin\alpha. \tag{2}$$

对 O 取力矩：两个法向反作用力都通过 O。每个摩擦力都沿切线方向，与 O 的距离为 a，产生逆时针力矩 aF_1 和 aF_2。重量的作用线与 O 的水平距离为 $\frac{a}{2}(\cos\alpha - \sin\alpha)$，力矩为顺时针方向。因此

$$F_1 + F_2 = \tfrac{1}{2}mg(\cos\alpha - \sin\alpha). \tag{3}$$

极限摩擦与 $\tan\alpha$

杆的两端都处于即将滑动的状态，所以 $F_1 = \mu R_1$ 及 $F_2 = \mu R_2$。记 $c = \cos\alpha$，$s = \sin\alpha$，方程 (1) 和 (2) 变成

$$R_1 + \mu R_2 = mgc, \qquad -\mu R_1 + R_2 = mgs,$$

其解为

$$R_1 = \frac{mg(c - \mu s)}{1 + \mu^2}, \qquad R_2 = \frac{mg(s + \mu c)}{1 + \mu^2}.$$

当 μ 很小时（这里必然如此，见下文），两者皆为正数，所以反作用力确实如图所示指向 O。现在 (3) 表示 $\mu(R_1 + R_2) = \frac{1}{2}mg(c - s)$，即

$$\begin{aligned} 2\mu\left(c - \mu s + s + \mu c\right) &= (1 + \mu^2)(c - s) \\ c\,(2\mu + 2\mu^2 - 1 - \mu^2) + s\,(2\mu - 2\mu^2 + 1 + \mu^2) &= 0 \\ s\,(1 + 2\mu - \mu^2) &= c\,(1 - 2\mu - \mu^2). \end{aligned}$$

两边除以 $c\,(1 + 2\mu - \mu^2)$（它不等于零，因为 $c > 0$，而 $\tan\alpha > 0$ 迫使 μ 必须很小，这一点稍后便会看到），得

$$\tan\alpha = \frac{1 - 2\mu - \mu^2}{1 + 2\mu - \mu^2},$$

这正是所求的结果。

想一想 我们怎样知道摩擦力是沿逆时针方向而不是顺时针方向作用？

重量对 O 的力矩是顺时针方向的，因为 G 在 O 的右方，而法向反作用力对 O 没有力矩。所以必须单靠摩擦力提供逆时针方向的力矩；若没有摩擦力，杆便会顺时针滑动。假如你把摩擦力画成相反方向，整个运算过程会完全一样，只是 μ 换成了 $-\mu$，结果得出 $\tan\alpha > 1$，与 $\alpha < 45^\circ$ 矛盾。

杆的倾角

由 P_1 到 P_2 的向量是 $a(\cos\alpha + \sin\alpha,\ \cos\alpha - \sin\alpha)$，它的两个分量都是正数。所以杆与水平线所成的锐角 θ 满足

$$\tan\theta = \frac{\cos\alpha - \sin\alpha}{\cos\alpha + \sin\alpha} = \frac{1 - \tan\alpha}{1 + \tan\alpha} = \tan(45^\circ - \alpha),$$

而由于 $0 < 45^\circ - \alpha < 45^\circ$，我们得到 $\theta = 45^\circ - \alpha$。

现在把 $\mu = \tan\lambda$ 代入 $\tan\alpha$ 的公式，并把分子和分母同乘以 $\cos^2\lambda$：

$$\tan\alpha = \frac{(\cos^2\lambda - \sin^2\lambda) - 2\sin\lambda\cos\lambda}{(\cos^2\lambda - \sin^2\lambda) + 2\sin\lambda\cos\lambda} = \frac{\cos 2\lambda - \sin 2\lambda}{\cos 2\lambda + \sin 2\lambda} = \frac{1 - \tan 2\lambda}{1 + \tan 2\lambda} = \tan(45^\circ - 2\lambda).$$

由于 $0 < \lambda < 22.5^\circ$，角 $45^\circ - 2\lambda$ 严格介乎 0° 与 45° 之间，α 亦然。正切函数在这个区间上是一对一（one-to-one）的，所以 $\alpha = 45^\circ - 2\lambda$。因此

$$\theta = 45^\circ - \alpha = 2\lambda,$$

这正是我们要证明的。（这也解释了条件 $\lambda < 22.5^\circ$，即 $\mu < \sqrt{2} - 1$ 的由来：它恰好就是令 $\tan\alpha > 0$ 的条件。）

验算

- 当 $\mu = 0$（光滑线圈）时，我们得到 $\tan\alpha = 1$，所以 $\alpha = 45^\circ$，杆呈水平，这是杆在光滑圆上能够静止的唯一位置。
- 当 $\mu = 0.2$ 时：$\tan\alpha = 0.56/1.36$，所以 $\alpha \approx 22.38^\circ$，而 $\theta = 45^\circ - \alpha \approx 22.62^\circ$，与 $2\arctan 0.2 \approx 22.62^\circ$ 吻合。
- 数值验证：随机选取 10^4 组 $\mu \in (0, \sqrt{2} - 1)$、$a$ 和 m 的值，把上述 R_1, R_2, F_1, F_2 代入完整的向量方程，所得的合力和力矩皆为零（误差约为 10^{-14}），并且 $R_1, R_2 > 0$ 及 $\theta = 2\lambda$。

常见错误

- 沿水平和铅垂方向分解。这样做是可行的，但每条方程都会包含全部四个力；沿两条互相垂直的半径分解则简洁得多。
- 把其中一端摩擦力的方向弄错。两个环沿圆周滑动的方向相同，所以两端的摩擦力都是逆时针方向的。
- 忘记了 $\theta = 45^\circ - \alpha$ 需要略加说明：弦垂直于 $\angle P_1OP_2$ 的角平分线。

STEP 之外

这题还有一个更巧妙的几何看法：在极限摩擦的情况下，每一端的总反作用力与半径成角 λ，而一个在三个力作用下处于平衡(equilibrium)的物体，这三个力必须共点(concurrent)，由此可直接推出 $\theta = 2\lambda$。这个“摩擦角”（angle of friction）的概念，以及摩擦力其实是不等式 $|F| \le \mu R$ 而不是等式这一事实，正是处理多个接触点的静力学问题的起点；在这类问题中，你必须问哪些接触点会先滑动。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）正是以这种严谨的态度处理摩擦力和刚体平衡。

详细解答

STEP I 2008，第 12 题

概率与统计 • 离散均匀分布变量的最大值

STEP I 2008，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/08-s1-q12

关键思路

X 其实就是两个分数中较大的一个。若直接求“较大者等于 r”的概率，便要分几种情况来数；但“较大者不超过 r”却很容易处理：这恰好在两个分数都不超过 r 时发生。因此我们先求累积分布（cumulative distribution）$\mathrm{P}(X \leq r)$，再取差。有了这个累积分布，中位数（median）也几乎立即可得。

解答

设 X_1 和 X_2 为两个独立（independent）的分数，并设 $X = \max(X_1, X_2)$（当 $X_1 = X_2$ 时，定义中的两种情况给出相同的值，所以 X 是明确定义的）。

第 (i) 部分

对于 $r = 0, 1, \ldots, N$，事件 $\{X \leq r\}$ 就是事件 $\{X_1 \leq r$ 且 $X_2 \leq r\}$。这两个事件各自的概率都是 r/N，而且由于 X_1 和 X_2 独立，它们亦互相独立，所以

$$\mathrm{P}(X \leq r) = \frac{r}{N} \cdot \frac{r}{N} = \frac{r^2}{N^2}.$$

这在 $r = 0$ 时同样成立，所得的值为 0。因此，对于 $r = 1, 2, \ldots, N$，

$$\mathrm{P}(X = r) = \mathrm{P}(X \leq r) - \mathrm{P}(X \leq r - 1) = \frac{r^2 - (r-1)^2}{N^2} = \frac{2r - 1}{N^2},$$

这正是所要证明的结果。（简单核对一下：$\sum_{r=1}^{N}(2r - 1) = N^2$，所以这些概率之和为 1。）

第 (ii) 部分

利用题目给出的两个求和公式，

$$\begin{aligned}\mu = \sum_{r=1}^{N} r\,\frac{2r-1}{N^2} &= \frac{1}{N^2}\left(2\sum_{r=1}^{N} r^2 - \sum_{r=1}^{N} r\right)\\ &= \frac{1}{N^2}\left(\frac{N(N+1)(2N+1)}{3} - \frac{N(N+1)}{2}\right)\\ &= \frac{N+1}{N}\cdot\frac{2(2N+1)-3}{6} = \frac{(N+1)(4N-1)}{6N}.\end{aligned}$$

当 $N = 100$ 时，

$$\mu = \frac{101\times 399}{600} = \frac{40\,299}{600} = 67.165,$$

与题目所求相符。

第 (iii) 部分

由第 (i) 部分，$\mathrm{P}(X \le m) = m^2/N^2$，而 $\mathrm{P}(X \ge m) = 1 - \mathrm{P}(X \le m-1) = 1 - (m-1)^2/N^2$。中位数须满足的两个条件于是变成

$$\mathrm{P}(X \le m) \ge \tfrac{1}{2} \iff m^2 \ge \tfrac{1}{2}N^2 \iff m \ge \frac{N}{\sqrt{2}},$$

$$\mathrm{P}(X \ge m) \ge \tfrac{1}{2} \iff (m-1)^2 \le \tfrac{1}{2}N^2 \iff m-1 \le \frac{N}{\sqrt{2}},$$

这里取了正平方根；由于 $m \ge 1$，m 和 $m-1$ 都是非负数，所以这一步是成立的。因此 m 必须是满足

$$\frac{N}{\sqrt{2}} \le m \le \frac{N}{\sqrt{2}} + 1.$$

的整数。

> **想一想** 长度为 1 的区间可以包含两个整数。会不会有两个中位数?

只有当区间的两个端点都是整数时才会，而这需要 $N/\sqrt{2}$ 是整数。但若 $N/\sqrt{2} = k$，便有 $\sqrt{2} = N/k$，是一个有理数（rational number），这是不可能的。所以两个不等式都是严格不等式，而长度为 1 的开区间（open interval）恰好包含一个整数，于是

$$m = \left\lceil \frac{N}{\sqrt{2}} \right\rceil,$$

即大于 $N/\sqrt{2}$ 的最小整数。

当 $N = 100$ 时，$N/\sqrt{2} = 50\sqrt{2} \approx 70.71$，所以 $m = 71$。直接验证：$71^2 = 5041 \ge 5000 = \frac{1}{2}\times 100^2$，而 $70^2 = 4900 \le 5000$，所以两个条件都成立。

第 (iv) 部分

由第 (iii) 部分，$m = N/\sqrt{2} + \theta$，其中 $0 < \theta < 1$，所以

$$\frac{m}{N} = \frac{1}{\sqrt{2}} + \frac{\theta}{N} \to \frac{1}{\sqrt{2}} \quad \text{当} N \to \infty.$$

由第 (ii) 部分，

$$\frac{\mu}{N} = \frac{(N+1)(4N-1)}{6N^2} = \frac{1}{6}\left(1 + \frac{1}{N}\right)\left(4 - \frac{1}{N}\right) \to \frac{4}{6} = \frac{2}{3}.$$

因此，当 N 很大时，

$$\frac{\mu}{m} = \frac{\mu/N}{m/N} \approx \frac{2/3}{1/\sqrt{2}} = \frac{2\sqrt{2}}{3}.$$

验算

- 取 $N = 2$：四对等可能的结果中，$X = 1$ 出现一次，$X = 2$ 出现三次，与第 (i) 部分的 $\frac{1}{4}$ 和 $\frac{3}{4}$ 相符；而 $\mu = \frac{7}{4}$，与 $\frac{3\times 7}{12}$ 相符。
- 对 $N \le 40$，用电脑列出全部 N^2 对结果，可确认第 (i) 和第 (ii) 部分的结果完全正确；对 $N \le 2000$，把每个整数逐一代入中位数的定义检验，每次都恰好找到一个中位数，并且等于 $\lceil N/\sqrt{2} \rceil$。
- 当 $N = 10^5$ 时，$\mu/m \approx 0.942812$，而 $2\sqrt{2}/3 \approx 0.942809$。

常见错误

- 把 $\mathrm{P}(X = r)$ 算成 $2 \times \frac{r}{N} \cdot \frac{1}{N}$，忘记了这样会把 $X_1 = X_2 = r$ 的情况数了两次；正确的数目是 $2(r-1) + 1 = 2r - 1$。
- 写成 $\mathrm{P}(X \ge m) = 1 - m^2/N^2$：这其实是 $\mathrm{P}(X > m)$，而不是 $\mathrm{P}(X \ge m)$。
- 在 $N = 100$ 时答 $m = 70$，原因是把 70.71 四舍五入至最接近的整数，而不是向上取整。

STEP 之外

X 是*顺序统计量*（order statistic）最简单的例子：样本的最大值。当 $N \to \infty$ 时，X/N 趋近一个在 $[0,1]$ 上的连续随机变量（continuous random variable），其密度函数为 $2x$，平均值是 $\frac{2}{3}$，中位数是 $1/\sqrt{2}$，这正是第 (iv) 部分出现该比值的原因。较大样本的顺序统计量会引出贝塔分布（beta distribution）。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门）以浅白易明的方式讲解顺序统计量。

详细解答

STEP I 2008，第 13 题

概率与统计 • 环形排列

STEP I 2008，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/08-s1-q13

关键思路

在圆桌上，只有各人之间的相对位置才重要，所以我们固定其中一人的座位，再把其余五人排在他周围。这样共有 $5! = 120$ 种等可能的排列（arrangement），而每个概率（probability）都是在 120 种之中计数所得的比例。必须相邻而坐的一对夫妇，可以黏合成一个“组块”，组块内部有两种次序。至于“恰好”和“没有”的情况，我们可以直接计数，也可以对“第 k 对夫妇相邻而坐”这些事件（event）运用容斥原理（inclusion–exclusion）。

解答

把三对夫妇记为 A, a; B, b; C, c。所有人沿桌边同时移动一个位置，各人之间的相对座次并没有改变。因此我们固定 A 的座位，再按顺时针方向把其余五人依次安排在余下的座位上。这样共有 $5! = 120$ 种排列，全部等可能。

假如我们改为替座位编号，便会有 $6! = 720$ 种排列。这时每种相对排列都出现 6 次，分别对应 A 可以坐的每一个座位，所以概率不变。

一个有用的一般结果：把 n 件物件排成一个圆圈，共有 $(n-1)!$ 种排法（先固定一件物件，再排列其余的）。

第 (i) 部分

如果每对夫妇都相邻而坐，就把每对夫妇黏合成一个组块。这样圆桌旁共有 3 个组块，排法有 $(3-1)! = 2$ 种。在每个组块内，丈夫可以坐在妻子的任何一边，所以有 $2^3 = 8$ 种选择。因此

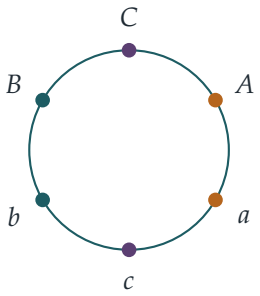
$$\mathrm{P}(\text{三对夫妇全部相邻}) = \frac{2 \times 8}{120} = \frac{16}{120} = \frac{2}{15},$$

正如所求。

第 (ii) 部分

首先选出哪一对夫妇不相邻，共有 3 种选择。假设是 C, c，即 A, a 和 B, b 各自相邻，但 C 和 c 不相邻。

把 Aa 和 Bb 黏合成组块。现在圆桌旁有 4 个单位：Aa, Bb, C, c，排法有 $3! = 6$ 种。固定 C，再看 c 可以落在这个 4 单位圆圈的其余三个位置中的哪一个。其中两个位置上 c 与 C 相邻，另一个位置上 c 则在正对面。所以在这 6 种排列中，C 和 c 分开的恰好有 $6/3 = 2$ 种。在这些排列中，C 与 c 之间的两个空隙各由一个组块填满，如图所示：



想一想 为甚么这比看起来简单?

当 C 与 c 相对而坐时，余下的自由度只有哪个组块坐在哪一边（2 种），以及每个组块内部的次序（$2 \times 2 = 4$ 种）。这样共有 $2 \times 4 = 8$ 种排列。把分开的那对夫妇的三种选择全部计算在内，

$$\mathrm{P}(\text{恰好两对夫妇相邻}) = \frac{3 \times 8}{120} = \frac{24}{120} = \frac{1}{5}.$$

第 (iii) 部分

设 E_k 为第 k 对夫妇相邻而坐的事件，其中 $k = 1, 2, 3$。我们在 120 种排列之中计数每一类排列的数目。

- $|E_1|$：把第 1 对夫妇黏合成一个组块。圆桌旁便剩下 5 个单位，所以有 $4! = 24$ 种排列，每种有 2 种内部次序，合共 48 种。
- $|E_1 \cap E_2|$：两个组块加上两个人，共 4 个单位。排列有 $3! = 6$ 种，内部次序有 $2^2 = 4$ 种，合共 24 种。
- $|E_1 \cap E_2 \cap E_3| = 16$，由第 (i) 部分得出。

根据对称性（symmetry），无论选取哪几对夫妇，以上数目都相同。由容斥原理，至少有

一对夫妇相邻的排列数目是

$$|E_1 \cup E_2 \cup E_3| = 3(48) - 3(24) + 16 = 144 - 72 + 16 = 88.$$

所以有 $120 - 88 = 32$ 种排列没有任何夫妇相邻，而

$$\mathrm{P}(\text{没有丈夫与自己的妻子相邻}) = \frac{32}{120} = \frac{4}{15}.$$

想一想 第 (ii) 部分的答案可以从同一组数目直接读出吗?

可以。每个 $|E_i \cap E_j|$ 所计数的，是恰好只有 i, j 这两对相邻的排列，再加上三对全部相邻的排列。所以恰好两对夫妇相邻的排列数目是 $3(24) - 3(16) = 24$，与第 (ii) 部分吻合。

验算

- 四种情况必须涵盖全部 120 种排列。恰好一对夫妇相邻的排列数目是 $3(48) - 2 \cdot 3(24) + 3(16) = 48$，而 $32 + 48 + 24 + 16 = 120$。
- 用电脑穷举全部 720 种有编号的座次，得出 $0, 1, 2, 3$ 对夫妇相邻的概率分别为 $\frac{4}{15}, \frac{2}{5}, \frac{1}{5}, \frac{2}{15}$。这些结果与上面的答案吻合。
- 合理性检查：某位妻子的两个邻座是从 5 个人中选出的，所以她丈夫坐在她旁边的概率是 $\frac{2}{5}$。这与 $|E_1|/120 = 48/120$ 一致。

常见错误

- 总数用了 $6!$ 种排列，却把符合条件的座次当作环形排列来计数，或者反过来。分子和分母必须采用同一种计数方式。
- 忘记每个组块内两人的次序所带来的因数 2。
- 在第 (ii) 部分，忘记剔除第三对夫妇也相邻的排列。若不剔除，$\frac{3\times 24}{120} = \frac{3}{5}$ 就大得太多了。

STEP 之外

这是*夫妇入座问题*（*ménage problem*）的缩影。经典版本还要求男女相间而坐，而它的一般答案（Touchard 公式）正是来自这一类容斥原理的运用。在我们这个不设限制的版本中，若有 n 对夫妇，相邻而坐的夫妇对数的期望值（expected value）是 $n \cdot \frac{2}{2n-1} \to 1$。没有任何夫妇相邻的概率趋向 e^{-1}，与著名的错排（derangement）极限一模一样。在男

女相间的 ménage 版本中，这个概率则趋向 e^{-2}。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率入门教材）对容斥原理和这类极限有很好的论述。

详细解答

STEP II 2008，第 1 题

纯数学 · 复数迭代

STEP II 2008，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/08-s2-q1

关键思路

$x^2 - y^2$ 和 $2xy$ 这两个式子，正是 $(x + \mathrm{i}y)^2$ 的实部（real part）和虚部（imaginary part）。因此，若把每个点写成复数（complex number）$z_n = x_n + \mathrm{i}y_n$，这个含两个分量的规则就化为单一方程 $z_{n+1} = z_n^2 + c$，当中只有一个复常数 c。这样，“常数”和“周期为 2”两个条件都变成只含一个复未知数的二次方程（quadratic equation），可以用二次公式（quadratic formula）求解。

解答

设 $z_n = x_n + \mathrm{i}y_n$ 及 $c = a + (b+2)\mathrm{i}$。由于

$$z_n^2 + c = (x_n^2 - y_n^2 + a) + \mathrm{i}\,(2x_n y_n + b + 2),$$

而两个复数相等，当且仅当它们的实部和虚部分别相等，所以该规则等价于

$$z_{n+1} = z_n^2 + c.$$

因为 a、b、x_n 和 y_n 都是实数，每个点 (x, y) 恰好对应一个复数，反之亦然。所以改用 z 来计算，不会有任何损失。

第 (i) 部分

当 $a = 1$、$b = -1$ 时，$c = 1 + \mathrm{i}$。数列（sequence）为常数数列，当且仅当 $z_2 = z_1$，因为此后每一步都只是重复同一个计算。所以我们需要

$$z^2 - z + (1 + \mathrm{i}) = 0, \qquad z = \frac{1 \pm \sqrt{1 - 4(1+\mathrm{i})}}{2} = \frac{1 \pm \sqrt{-3 - 4\mathrm{i}}}{2}.$$

要求 $-3-4\mathrm{i}$ 的一个平方根, 可设 $(p+q\mathrm{i})^2=-3-4\mathrm{i}$, 其中 p、q 为实数: 即 $p^2-q^2=-3$ 及 $pq=-2$。取 $p=1$、$q=-2$ 即可，因为 $(1-2\mathrm{i})^2=1-4\mathrm{i}-4=-3-4\mathrm{i}$。因此

$$z=\frac{1\pm(1-2\mathrm{i})}{2}=1-\mathrm{i}\quad\text{或}\quad\mathrm{i}.$$

所以数列为常数数列，当且仅当

$$(x_1,y_1)=(1,-1)\quad\text{或}\quad(x_1,y_1)=(0,1).$$

直接验证：$(1,-1)\mapsto(1-1+1,\,-2+1)=(1,-1)$ 及 $(0,1)\mapsto(0-1+1,\,0+1)=(0,1)$。

第 (ii) 部分

现在 $z_1=-1+\mathrm{i}$，所以 $z_1^2=1-2\mathrm{i}-1=-2\mathrm{i}$，而

$$z_2=c-2\mathrm{i}.$$

记 $w=z_2$。当 $z_3=z_1$ 但 $z_2\neq z_1$ 时，数列的周期（period）为 2。一旦 $z_3=z_1$，按照规则，数列便会以 $z_1,z_2,z_1,z_2,\ldots$ 的形式不断重复。利用 $c=w+2\mathrm{i}$,

$$z_3=w^2+c=w^2+w+2\mathrm{i},$$

所以 $z_3=z_1$ 变成

$$w^2+w+2\mathrm{i}=-1+\mathrm{i}\quad\Longleftrightarrow\quad w^2+w+(1+\mathrm{i})=0.$$

想一想 为甚么选 $w=z_2$ 作未知数，而不选 c?

无论用哪一个作未知数, 条件 $z_3=z_1$ 都是一个关于 c 的二次方程; 但若以 w 表示, 常数数列 $w=z_1$ 必定是其中一个根（root)。所以我们已经知道这个二次方程的一个根，而另一个根正是我们要找的。事实上 $w=-1+\mathrm{i}$ 确实满足方程：$(-1+\mathrm{i})^2+(-1+\mathrm{i})+1+\mathrm{i}=-2\mathrm{i}+2\mathrm{i}=0$。两根之和为 -1，所以另一个根是 $w=-\mathrm{i}$。（用二次公式也得到相同结果：$w=\frac{1}{2}\big(-1\pm\sqrt{-3-4\mathrm{i}}\big)=\frac{1}{2}\big(-1\pm(1-2\mathrm{i})\big)$，所用的平方根与 (i) 相同。）

- $w=-1+\mathrm{i}=z_1$ 给出 $c=-1+3\mathrm{i}$，即 $a=-1$、$b=1$。这时 $z_2=z_1$，数列是常数数列，所以它的周期是 1，而不是 2。
- $w=-\mathrm{i}$ 给出 $c=\mathrm{i}$，即 $a=0$ 及 $b+2=1$。

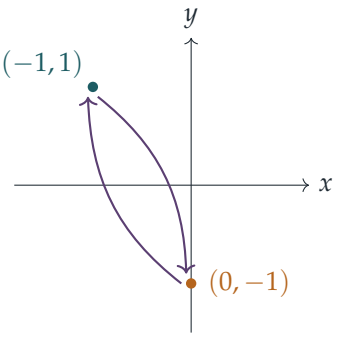
所以数列的周期为 2，当且仅当

$$a=0,\qquad b=-1.$$

用坐标验算：当 $a=0$、$b=-1$ 时，

$$(-1,1)\mapsto(1-1+0,\ -2+1)=(0,-1)\mapsto(0-1+0,\ 0+1)=(-1,1),$$

而这两点并不相同，所以周期恰好是 2。



验算

- 按原来的坐标规则迭代（iterate），可见当 $a=1$、$b=-1$ 时，$(1,-1)$ 和 $(0,1)$ 都固定不动；而当 $a=0$、$b=-1$ 时，$(-1,1)\to(0,-1)\to(-1,1)$。
- 平方根 $1-2\mathrm{i}$ 可以通过把它平方来验证；因式分解（factorisation）$w^2+w+1+\mathrm{i}=(w+1-\mathrm{i})(w+\mathrm{i})$ 同样可以验证：两根之积为 $(-1+\mathrm{i})(-\mathrm{i})=1+\mathrm{i}$。
- 以随机起始值进行数值求根搜索，在 (i) 中找不到其他不动点（fixed point），在 (ii) 中除了给出常数数列的 $(-1,1)$ 之外，也找不到其他满足 $z_3=z_1$ 的 (a,b)。

常见错误

- 忘记第二个分量中的“+2”，于是取了 $c=a+b\mathrm{i}$，而不是 $a+(b+2)\mathrm{i}$。
- 把 $a=-1$、$b=1$ 当作 (ii) 的第二个答案。该数列是常数数列，所以它的周期是 1。
- 只用实坐标计算，结果在联立二次方程中迷失方向。这样做是可行的，但复数形式简短得多。

STEP 之外

映射（map）$z\mapsto z^2+c$ 是复动力系统（complex dynamics）中最著名的例子。曼德博集合（Mandelbrot set）是使 0 的轨道（orbit）保持有界的所有 c 所组成的集合；它最大的两个部分，即主心脏线（main cardioid）和其左方的圆盘，分别由使该映射具有吸引不动

点（attracting fixed point）或吸引 2-循环（attracting 2-cycle）的 c 组成。在本题中，2-循环的方程同样有一个来自不动点的因式：一般而言 $(z^2+c)^2+c-z=(z^2-z+c)(z^2+z+c+1)$。想了解这背后的几何，可参阅 Needham 的 *Visual Complex Analysis*（《视觉复分析》），这是一本易读的入门书。

详细解答

STEP II 2008，第 2 题

纯数学 · 部分分式与级数

STEP II 2008，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/08-s2-q2

关键思路

部分分式（partial fractions）把一个难以处理的有理函数（rational function）拆成三项，而这三项的展开式我们都烂熟于心：一个等比级数（geometric series）、它的导数，以及一个以 $-x^2$ 为公比的等比级数。头两项给出的系数大致随 n 增长；第三项的贡献永远只是 $\pm\frac{1}{2}$，并以周期 4 重复，题目中“非此即彼”的两种情况正是由此而来。至于最后一部分，要找一个 x 的值，使函数变成题目所给的分数。

解答

部分分式

由于 $(1-x)^2$ 是重复的一次因式，而 $1+x^2$ 是不可约二次式（irreducible quadratic），我们设

$$\frac{1+x}{(1-x)^2(1+x^2)} = \frac{A}{1-x} + \frac{B}{(1-x)^2} + \frac{Cx+D}{1+x^2},$$

于是对所有 x，

$$1+x = A(1-x)(1+x^2) + B(1+x^2) + (Cx+D)(1-x)^2.$$

- 代入 $x=1$：$2=2B$，所以 $B=1$。
- 代入 $x=0$：$1=A+B+D$，所以 $A=-D$。
- 代入 $x=\mathrm{i}$（这是容许的，因为这是一个多项式恒等式）：$1+\mathrm{i}=(C\mathrm{i}+D)(1-\mathrm{i})^2=(C\mathrm{i}+D)(-2\mathrm{i})=2C-2D\mathrm{i}$。比较实部和虚部，得 $C=\frac{1}{2}$ 和 $D=-\frac{1}{2}$，因此 $A=\frac{1}{2}$。

作为验算，右边 x^3 的系数是 $-A+C=0$，与左边相符。所以

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

系数

当 $|x|<1$ 时，三个展开式全部收敛（converge），而且

$$\begin{aligned}\frac{1}{2(1-x)}&=\tfrac{1}{2}\sum_{n\geq 0}x^n, \qquad \frac{1}{(1-x)^2}=\sum_{n\geq 0}(n+1)x^n,\\ -\frac{1-x}{2(1+x^2)}&=-\tfrac{1}{2}(1-x)\left(1-x^2+x^4-x^6+\cdots\right)\\ &=-\tfrac{1}{2}\left(1-x-x^2+x^3+x^4-x^5-x^6+x^7+\cdots\right).\end{aligned}$$

最后一行括号内的系数是 $1,-1,-1,1$，以周期 4 重复：当 $n=2k$ 时，x^n 的系数是 $(-1)^k$；当 $n=2k+1$ 时，则是 $-(-1)^k$。所以，当 $n\equiv 0$ 或 $3\pmod 4$ 时，第三项对 x^n 的贡献是 $-\frac{1}{2}$；当 $n\equiv 1$ 或 $2\pmod 4$ 时，贡献是 $+\frac{1}{2}$。

> **想一想** 为甚么周期为 4 的规律只产生两种情况，而不是四种?

因为只有正负号才有影响：把三项的贡献相加，

$$a_n=\tfrac{1}{2}+(n+1)\mp\tfrac{1}{2}=\begin{cases}n+1, & n\equiv 0\text{或}3 \pmod 4,\\ n+2, & n\equiv 1\text{或}2 \pmod 4,\end{cases}$$

这就是所要求的结果。这个数列的首几项是 $1,3,4,4,5,7,8,8,9,11,12,12,\ldots$

小数近似值

> **想一想** x 取甚么值时，函数会等于 $\frac{11\,000}{8181}$?

从分母可以猜到 $x=\frac{1}{10}$：这时 $(1-x)^2=0.81$ 而 $1+x^2=1.01$，所以 $(1-x)^2(1+x^2)=0.8181$，并且

$$\frac{1+\frac{1}{10}}{0.8181}=\frac{1.1}{0.8181}=\frac{11\,000}{8181}.$$

由于 $\frac{1}{10}<1$，级数展开成立，而且

$$\frac{11\,000}{8181}=\sum_{n\geq 0}\frac{a_n}{10^n}=1+0.3+0.04+0.004+0.0005+0.00007+\cdots.$$

只要 $a_n\leq 9$（即直至 $n=8$ 为止），各系数就正好是小数的各个数字：1.34457889。此后

的系数都是两位数，所以必须进位：

$$\begin{aligned}\sum_{n=0}^{8}\frac{a_n}{10^n} &= 1.344\,578\,89\\ +11\times10^{-9} &\to 1.344\,578\,901\\ +12\times10^{-10} &\to 1.344\,578\,902\,2\\ +12\times10^{-11} &\to 1.344\,578\,902\,32\\ +13\times10^{-12} &\to 1.344\,578\,902\,333.\end{aligned}$$

其后各项只会影响第十一位有效数字（significant figure）及以后的数字，所以准确至九位有效数字，

$$\frac{11\,000}{8181}\approx 1.344\,578\,90.$$

验算

- 在函数中代入 $x=0$ 得 $1=a_0$，而 $n=0$ 属于 $n+1$ 的情况。把 $(1+x)(1+2x+\cdots)(1-x^2+\cdots)$ 乘开，得 $1+3x+\cdots$，所以 $a_1=3=1+2$。
- 用长除法（long division）得 $11\,000/8181=1.344\,578\,902\,33\ldots$，与级数的结果一致。我们亦以精确的幂级数（power series）除法核对了 $n<60$ 的 a_n，并在多个随机选取的有理点上核对了部分分式恒等式。

常见错误

- 忘记进位：直接把 a_n 当作各位数字读出，会得到 $1.34457889\ldots$，第九位有效数字就错了。
- 把 $\frac{Cx+D}{1+x^2}$ 写成 $\frac{D}{1+x^2}$；不可约二次式的分子必须是一次式。
- 把 $1.344578902\ldots$ 舍入为 1.34457890 之后，却略去末尾的零；这个零在这里是一位有效数字。

STEP 之外

有理函数必定有幂级数展开式，其系数是 n 的多项式，再加上形如 r^n 的项和周期性的规律，分母的每个根各对应一项；这就是*线性递推关系*（linear recurrences）和*生成函数*（generating functions）的理论。在这一题，根 1（二重根）和 $\pm\mathrm{i}$ 分别带来线性增长和周期为 4 的起伏。这也解释了为甚么每个分数的小数展开最终都会循环。Wilf 的

generatingfunctionology 是一本生动有趣、可免费取得的入门书，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多同类风格的题目。

详细解答

STEP II 2008，第 3 题

纯数学 · 不等式与反证法

STEP II 2008，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/08-s2-q3

关键思路

从草图可见，当 $x \geq 0$ 时，这条三次曲线（cubic）永远不会跌到 x 轴之下；把“$y \geq 0$”重新整理，便得到不等式 $x^2(1-x) \leq 4/27$。至于反证法（proof by contradiction）的部分，留意 $bc(1-a)$、$ca(1-b)$、$ab(1-c)$ 这三个数逐个来看都不好处理，但它们的乘积可以重新组合成 $a^2(1-a)\cdot b^2(1-b)\cdot c^2(1-c)$，而这正好受第一个结果控制。第 (ii) 部分用的是同一个技巧，只是把三次式换成二次式（quadratic）。

解答

第 (i) 部分：曲线

对 $y = 27x^3 - 27x^2 + 4$ 求导，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 81x^2 - 54x = 27x(3x-2),$$

它在 $x = 0$ 和 $x = \frac{2}{3}$ 时为零。当 $x = 0$ 时，$y = 4$；当 $x = \frac{2}{3}$ 时，

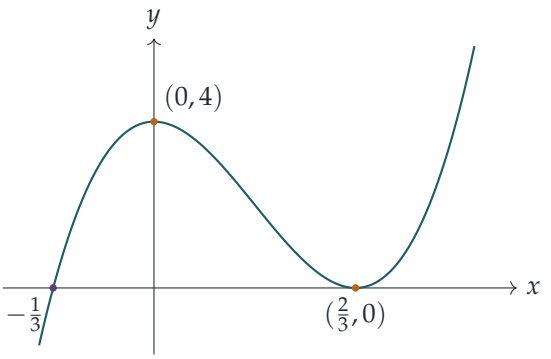
$$y = 27\cdot\tfrac{8}{27} - 27\cdot\tfrac{4}{9} + 4 = 8 - 12 + 4 = 0.$$

由于 $\mathrm{d}^2y/\mathrm{d}x^2 = 162x - 54$ 在 $x = 0$ 时为 $-54 < 0$，在 $x = \frac{2}{3}$ 时为 $54 > 0$，所以转向点（turning point）是极大点（maximum）$(0, 4)$ 和极小点（minimum）$(\frac{2}{3}, 0)$。

极小点位于 x 轴上，所以曲线在该处与横轴相切，而 $(3x-2)^2$ 是一个因式（factor）。比较 x^3 的系数和常数项，得

$$27x^3 - 27x^2 + 4 = (3x-2)^2(3x+1),$$

所以曲线与横轴相交的另一处只有 $x = -\frac{1}{3}$。曲线从 $-\infty$ 上升，在 $x = -\frac{1}{3}$ 穿过横轴，在 $(0,4)$ 达到顶点，然后下降至 $(\frac{2}{3},0)$ 与横轴相切，之后再度上升。



第 (i) 部分：不等式

当 $x \geq 0$ 时，我们有 $3x+1>0$ 和 $(3x-2)^2 \geq 0$，所以由因式分解（factorisation）可得 $27x^3 - 27x^2 + 4 \geq 0$。（从草图看也是一样：在 $x \geq 0$ 的范围内，曲线的最低点就是极小点 $(\frac{2}{3},0)$。）重新整理，

$$4 \geq 27x^2 - 27x^3 = 27x^2(1-x), \qquad \text{所以} \qquad x^2(1-x) \leq \frac{4}{27} \quad \text{对所有} x \geq 0,$$

而等号只在 $x = \frac{2}{3}$ 时成立。这就是所要求的结果。

第 (i) 部分：反证

用反证法，假设三个数全都大于 4/27：

$$bc(1-a) > \tfrac{4}{27}, \qquad ca(1-b) > \tfrac{4}{27}, \qquad ab(1-c) > \tfrac{4}{27}.$$

这里六个量全是正数，所以我们可以把三条不等式相乘：

$$bc(1-a)\cdot ca(1-b)\cdot ab(1-c) > \left(\tfrac{4}{27}\right)^3.$$

? **想一想** 第一个结果怎样才能对这些混合的乘积有所说明?

把左边重新组合，让每个字母与它自己的括号归在一起：

$$bc(1-a)\cdot ca(1-b)\cdot ab(1-c) = \left[a^2(1-a)\right]\left[b^2(1-b)\right]\left[c^2(1-c)\right].$$

a,b,c 每个都是非负数，所以根据第一个结果，每个方括号的值都介乎 0（因为 $1-a \geq 0$，其余类推）与 4/27 之间。把这三条非负数之间的不等式相乘，

$$\left[a^2(1-a)\right]\left[b^2(1-b)\right]\left[c^2(1-c)\right] \leq \left(\tfrac{4}{27}\right)^3.$$

这与上面的严格不等式矛盾。所以我们的假设不成立，即 $bc(1-a)$、$ca(1-b)$、$ab(1-c)$ 之中至少有一个小于或等于 $4/27$。

第 (ii) 部分

同一个想法对二次式也行得通。对任意实数 t，

$$t(1-t) = \tfrac{1}{4} - \left(t - \tfrac{1}{2}\right)^2 \leq \tfrac{1}{4},$$

这可由配方法(completing the square)得出。用反证法，假设 $p(1-q) > \frac{1}{4}$ 及 $q(1-p) > \frac{1}{4}$。两边都是正数，所以相乘可得

$$p(1-q)\cdot q(1-p) > \tfrac{1}{16}.$$

但左边可以重新组合成 $\left[p(1-p)\right]\left[q(1-q)\right]$，而由于 $0 < p, q < 1$，每个方括号的值都介乎 0 与 $\frac{1}{4}$ 之间，所以乘积至多为 $\frac{1}{16}$。这是矛盾的，所以 $p(1-q)$ 和 $q(1-p)$ 之中至少有一个小于或等于 $\frac{1}{4}$。

验算

- 等号成立的情况：$a = b = c = \frac{2}{3}$ 使 (i) 中的三个数全都等于 $\frac{4}{9}\cdot\frac{1}{3} = \frac{4}{27}$，而 $p = q = \frac{1}{2}$ 使 (ii) 中的两个数都等于 $\frac{1}{4}$。所以 $4/27$ 和 $1/4$ 这两个界不能再改进。
- 展开 $(3x-2)^2(3x+1) = (9x^2 - 12x + 4)(3x+1) = 27x^3 - 27x^2 + 4$，确认因式分解无误。
- 以数值计算，$x^2(1-x)$ 在 $[0,3]$ 上的最大值是 $0.148148\ldots = 4/27$。在 $(0,1)$ 内随机抽取一百万组三元组 (a,b,c)，三个乘积中最小的一个从未超过 0.1468；随机抽取一百万对数值，$p(1-q)$、$q(1-p)$ 中较小的一个从未超过 0.24998。

常见错误

- 把不等式相乘，却没有说明所有量都是正数。“$A > \alpha$、$B > \beta$ 蕴含 $AB > \alpha\beta$”这一步对负数并不成立，所以 $0 < a, b, c < 1$ 这个假设确实是必需的。
- 在 $x < 0$ 时使用 $x^2(1-x) \leq 4/27$ 这个界，但它在该范围并不成立（试试 $x = -1$）。
- 只证明某一个特定的数（例如 $bc(1-a)$）很小。它不一定小：当 a 接近 0 而 $b = c$ 接近 1 时，它接近 1。

STEP 之外

把若干个量相乘,使一个对称的界得以派上用场,这种做法与算术—几何平均不等式(AM–GM inequality)十分相近,而后者可以直接给出第一个结果:对 $0 \le x \le 1$,把算术—几何平均不等式应用于 $\frac{x}{2}, \frac{x}{2}, 1-x$(三者之和为 1),便得 $\frac{x^2}{4}(1-x) \le \left(\frac{1}{3}\right)^3$。这个论证也带有鸽巢原理(pigeonhole principle)的味道:如果乘积很小,就不可能每个因子都很大。这些关于估界和求最优值的想法,会在大学一年级的分析(analysis)课程中进一步发展;Abbott 的 *Understanding Analysis*(分析学入门读物)是个平易近人的起点,而 Siklos 的 *Advanced Problems in Mathematics*(进阶数学难题集)则有更多这类风格的 STEP 题目。

详细解答

STEP II 2008，第 4 题

纯数学 · 隐函数微分法

> **STEP II 2008，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/08-s2-q4

关键思路

处理两条直线之间的夹角，用向量（vector）最为方便：$F(x,y)=1$ 的法线（normal）沿 ∇F 的方向，而对于平面上两个向量 $\mathbf{p}$ 和 $\mathbf{n}$，两者夹角的 tan 等于 |叉积| 除以点积（dot product）。在这一题，点积原来恰好是曲线方程的左方，即 1。之后，θ 取驻值（stationary value）当且仅当 y^2-x^2 取驻值，余下的就是关于 x^2+y^2 和 xy 的代数运算。

解答

斜率

把 $x^2+y^2+2axy=1$ 对 x 作隐函数微分（implicit differentiation），

$$2x+2y\frac{\mathrm{d}y}{\mathrm{d}x}+2a\left(y+x\frac{\mathrm{d}y}{\mathrm{d}x}\right)=0 \quad\Longrightarrow\quad (ax+y)\frac{\mathrm{d}y}{\mathrm{d}x}=-(x+ay).$$

因此，只要 $ax+y\neq 0$，斜率（gradient）就是 $-\dfrac{x+ay}{ax+y}$，正如题目所求。

角 θ

切线（tangent）的方向为 $(ax+y,\ -(x+ay))$，所以法线的方向为

$$\mathbf{n}=(x+ay,\ ax+y),$$

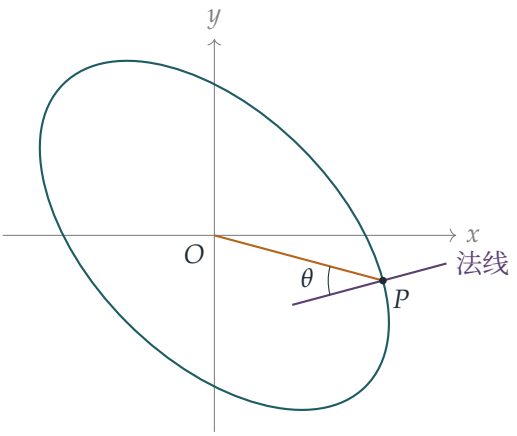
即与它垂直的向量。（这就是 $F = x^2 + y^2 + 2axy$ 的 $\frac{1}{2}\nabla F$，即使在 $ax + y = 0$ 的地方也有意义。）向量 $\overrightarrow{OP}$ 是 $\mathbf{p} = (x, y)$。设 ϕ 为 $\mathbf{p}$ 与 $\mathbf{n}$ 之间的夹角，则

$$|\mathbf{p}||\mathbf{n}|\cos\phi = \mathbf{p}\cdot\mathbf{n} = x(x+ay) + y(ax+y) = x^2 + y^2 + 2axy = 1,$$
$$|\mathbf{p}||\mathbf{n}|\sin\phi = |x(ax+y) - y(x+ay)| = |ax^2 - ay^2| = a|y^2 - x^2|,$$

其中第二行是二维叉积（cross product）的大小（即由 $\mathbf{p}$ 和 $\mathbf{n}$ 所张成的平行四边形的面积），而 $a > 0$。

由于点积为 $+1 > 0$，ϕ 本身已是锐角（或零），所以 $\phi = \theta$。把两行相除，得

$$\tan\theta = a\,|y^2 - x^2|.$$



上面的草图（取 $a = \frac{1}{2}$）显示了这条曲线：由于 $0 < a < 1$，它是一个椭圆（ellipse），两轴分别沿 $y = \pm x$。图中所画的点 P 是 θ 最大的其中一点。

第 (i) 部分

函数 $t \mapsto \arctan(t)$ 严格递增(strictly increasing),所以凡在 $f \neq 0$ 之处,θ 都是 $f = y^2 - x^2$ 的光滑递增函数。

想一想 在 $y^2 = x^2$ 的点，绝对值（modulus）出现尖角，那又如何?

在这些点 $\theta = 0$。假如 θ 在这样的一点导数（derivative）为零，那么 $a|f|$ 除以 x 的改变量会趋向 0，这迫使 $\mathrm{d}f/\mathrm{d}x = 0$ 同样成立。所以在任何情况下都有

$$\frac{\mathrm{d}\theta}{\mathrm{d}x} = 0 \implies \frac{\mathrm{d}}{\mathrm{d}x}(y^2 - x^2) = 0 \implies y\frac{\mathrm{d}y}{\mathrm{d}x} = x.$$

这里我们用到斜率公式，所以需要 $ax + y \neq 0$（下文会确认这个条件必然成立）。代入后再乘以 $ax + y$:

$$-y(x+ay) = x(ax+y) \implies -xy - ay^2 = ax^2 + xy \implies a(x^2 + y^2) + 2xy = 0.$$

这就是 (i)。

顺带一提，假如 $ax+y=0$，则由 (i) 会得出 $ax^2(1+a^2)-2ax^2=-a(1-a^2)x^2=0$，于是 $x=y=0$，但这一点不在曲线上。所以这类点永远不会造成麻烦。

第 (ii) 部分

记 $S=x^2+y^2$。由 (i)，$2xy=-aS$。把它代入曲线方程，

$$S+a(-aS)=1 \quad\Longrightarrow\quad S=\frac{1}{1-a^2}.$$

因此

$$x^2+y^2+2xy=S-aS=\frac{1-a}{1-a^2}=\frac{1}{1+a},$$

即 $(1+a)(x^2+y^2+2xy)=1$，正如题目所求。

第 (iii) 部分

我们需要求 $|y^2-x^2|$。利用 $(y^2-x^2)^2=(x^2+y^2)^2-4x^2y^2=S^2-(2xy)^2$，

$$(y^2-x^2)^2=S^2-a^2S^2=(1-a^2)\cdot\frac{1}{(1-a^2)^2}=\frac{1}{1-a^2}.$$

取正平方根（因 $0<a<1$，此举成立），得 $|y^2-x^2|=1/\sqrt{1-a^2}$，所以

$$\tan\theta=\frac{a}{\sqrt{1-a^2}}.$$

验算

- 旋转坐标轴：设 $u=(x+y)/\sqrt{2}$ 及 $v=(x-y)/\sqrt{2}$，曲线便化为 $(1+a)u^2+(1-a)v^2=1$，而 $y^2-x^2=-2uv$。由算术平均—几何平均不等式（AM–GM），$1\geq 2\sqrt{(1-a^2)}\,|uv|$，所以 $\tan\theta=2a|uv|\leq a/\sqrt{1-a^2}$，等号在 $(1+a)u^2=\frac{1}{2}$ 时成立。由于 $x^2+y^2+2xy=2u^2$，这个等号条件正正就是 (ii)。所以该驻值是 θ 的最大值。
- 取 $a=\frac{1}{2}$：$P\approx(1.1154,\,-0.2989)$ 满足曲线方程和 (i)，而 $\tan\theta=\frac{1}{2}|y^2-x^2|\approx 0.5774=1/\sqrt{3}$，所以 $\theta=30^\circ$。对若干个 a 值，沿椭圆一周以数值方法扫描 $\tan\theta$，所得的最大值正是这些数值。
- 当 $a\to 0$ 时，曲线变成单位圆，其法线通过 O，所以 $\theta\to 0$，与 $\tan\theta\to 0$ 吻合。

常见错误

- 以 OP 和法线的斜率套用 $\tan(\alpha-\beta)$：这方法行得通，但要小心处理 $x=0$ 的情况和正负号，而向量的做法可同时避开这两个问题。

- 忘记了为何所求得的角是*锐角*：关键在于点积为 $+1$。
- 以平方求 $|y^2 - x^2|$，其后却忘记取正根。

STEP 之外

把 $x^2 + y^2 + 2axy$ 写成二次型（quadratic form）$\mathbf{x}^{\mathsf{T}} A\mathbf{x}$，其中 $A = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$，则椭圆的两轴就是 A 的特征向量（eigenvector）$(1,1)$ 和 $(1,-1)$，而特征值（eigenvalue）$1 \pm a$ 决定了半轴的长度。对称矩阵（symmetric matrix）的这种对角化（diagonalisation）是大学一年级线性代数的核心概念；参看 Strang 的 *Introduction to Linear Algebra*（线性代数入门）。

详细解答

STEP II 2008，第 5 题

纯数学 · 代换积分法

STEP II 2008，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/08-s2-q5

关键思路

两个积分（integral）都可以靠看出导数（derivative）来解决。第一个积分中，分子 $\sin 2x = 2\sin x\cos x$ 恰好是分母 $1+\sin^2 x$ 的导数，所以一眼就能看出积分结果是一个对数（logarithm）。第二个积分中，分子只有一个 $\sin x$，这提示我们设 $c = \cos x$，并先把分母写成 $1+\sin^2 x = 2 - \cos^2 x$，余下的就是一个标准的部分分式（partial fractions）积分。于是，比较两个答案就变成比较两个含有 $\sqrt{2}$ 的数，而最后一部分的不等式（inequality）正是为了解决这个问题而设计的。

解答

两个积分

把两个积分分别记作 I_1（分子为 $\sin 2x$）和 I_2（分子为 $\sin x$）。

先看 I_1。由于 $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(1+\sin^2 x\right) = 2\sin x\cos x = \sin 2x$，被积函数（integrand）具有 $f'(x)/f(x)$ 的形式，其中 $f(x) = 1+\sin^2 x > 0$。因此

$$I_1 = \Big[\ln\left(1+\sin^2 x\right)\Big]_0^{\pi/2} = \ln 2 - \ln 1 = \ln 2.$$

再看 I_2。写成 $1+\sin^2 x = 2-\cos^2 x$，并作代换（substitution）$c = \cos x$，于是 $\mathrm{d}c = -\sin x\,\mathrm{d}x$；当 x 由 0 增至 $\frac{1}{2}\pi$ 时，c 由 1 减至 0。负号把积分上下限（limits）对调：

$$I_2 = \int_0^1 \frac{\mathrm{d}c}{2-c^2}.$$

利用部分分式，

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

而当 $0 \le c \le 1$ 时，两个分母都保持为正，所以

$$I_2 = \frac{1}{2\sqrt{2}}\Big[\ln(\sqrt{2}+c) - \ln(\sqrt{2}-c)\Big]_0^1 = \frac{1}{2\sqrt{2}} \ln \frac{\sqrt{2}+1}{\sqrt{2}-1}.$$

这个结果可以化简得很漂亮：$(\sqrt{2}+1)(\sqrt{2}-1) = 1$，所以 $\dfrac{\sqrt{2}+1}{\sqrt{2}-1} = (\sqrt{2}+1)^2$，因此

$$I_2 = \frac{1}{2\sqrt{2}} \cdot 2\ln(1+\sqrt{2}) = \frac{\ln(1+\sqrt{2})}{\sqrt{2}}.$$

不等式

根据二项式定理（binomial theorem），并利用 $(\sqrt{2})^2 = 2$、$(\sqrt{2})^3 = 2\sqrt{2}$、$(\sqrt{2})^4 = 4$ 和 $(\sqrt{2})^5 = 4\sqrt{2}$，

$$\begin{aligned}(1+\sqrt{2})^5 &= 1 + 5\sqrt{2} + 10\cdot 2 + 10\cdot 2\sqrt{2} + 5\cdot 4 + 4\sqrt{2} \\ &= 41 + 29\sqrt{2}.\end{aligned}$$

现在 $\sqrt{2} < 2$（因为 $2 < 4$），所以 $29\sqrt{2} < 58$，从而 $(1+\sqrt{2})^5 < 41 + 58 = 99$，这正是所要证明的。

其次，$1.4^2 = 1.96 < 2$，而 1.4 和 $\sqrt{2}$ 都是正数，平方不改变大小次序，所以 $\sqrt{2} > 1.4$。

想一想 $(1+\sqrt{2})^5 < 99$ 和 $\sqrt{2} > 1.4$ 怎么可能告诉我们任何关于 $2^{\sqrt{2}}$ 的事情?

写成 $1.4 = \frac{7}{5}$。那么 $2^{7/5}$ 就是 $2^7 = 128$ 的五次方根，而 $1+\sqrt{2}$ 的五次方正是我们刚刚估计过上界的量。准确地说：由于 $t \mapsto 2^t$ 是递增（increasing）函数，而且 $\sqrt{2} > \frac{7}{5}$，

$$2^{\sqrt{2}} > 2^{7/5}.$$

另外 $(1+\sqrt{2})^5 < 99 < 128 = \left(2^{7/5}\right)^5$，而 $y \mapsto y^5$ 在 $y > 0$ 时递增，取正五次方根便得 $1+\sqrt{2} < 2^{7/5}$。两者结合，

$$2^{\sqrt{2}} > 2^{7/5} > 1 + \sqrt{2}.$$

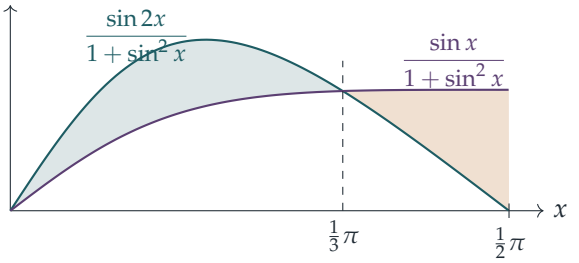
哪个积分较大?

函数 ln 是递增的，所以由最后一个不等式可得 $\ln\left(2^{\sqrt{2}}\right) > \ln(1+\sqrt{2})$，即 $\sqrt{2}\ln 2 > \ln(1+\sqrt{2})$。两边除以 $\sqrt{2} > 0$，

$$I_1 = \ln 2 > \frac{\ln(1+\sqrt{2})}{\sqrt{2}} = I_2.$$

所以分子为 $\sin 2x$ 的那个积分较大。

这个大小关系并不能直接从被积函数看出来。由于 $\sin 2x - \sin x = \sin x\,(2\cos x - 1)$，第一个被积函数在 $0 < x < \frac{1}{3}\pi$ 时较大，而在 $\frac{1}{3}\pi < x < \frac{1}{2}\pi$ 时较小：



问题在于青绿色区域的面积是否大于琥珀色区域的面积。答案是肯定的，但只多出约 0.07。这就是为甚么要对 $\sqrt{2}$ 作一番仔细的运算。

验算

- 对原来的被积函数使用辛普森法则（Simpson's rule），得 $I_1 \approx 0.693147$ 和 $I_2 \approx 0.623225$，与 $\ln 2$ 和 $\ln(1+\sqrt{2})/\sqrt{2}$ 吻合至小数点后十位。
- $(1+\sqrt{2})^5 \approx 82.01$，而 $41 + 29\sqrt{2} \approx 82.01$；远低于 99。
- $2^{\sqrt{2}} \approx 2.665$、$2^{1.4} \approx 2.639$ 和 $1+\sqrt{2} \approx 2.414$，次序与所断言的一致。
- 对 I_2 作合理性检查：被积函数介乎 $\frac{1}{2}\sin x$ 与 $\sin x$ 之间，所以 $\frac{1}{2} < I_2 < 1$，而 0.62 确实在此范围内。

常见错误

- 忘记 $\mathrm{d}c = -\sin x\,\mathrm{d}x$ 会把积分上下限对调，结果为一个正的被积函数算出负值。
- 把 I_2 停留在 $\frac{1}{2\sqrt{2}}\ln\frac{\sqrt{2}+1}{\sqrt{2}-1}$ 的形式：若没有看出对数内的分式就是 $(\sqrt{2}+1)^2$，它与最后一部分中 $1+\sqrt{2}$ 的联系便隐而不见。
- 单凭计算机就断言 $2^{\sqrt{2}} > 1+\sqrt{2}$。题目要求的是经由 $2^{7/5}$ 的那一串不等式，并须说明 2^t 和 y^5 的单调性（monotonicity）。

STEP 之外

$2^{\sqrt{2}}$ 究竟是甚么意思？在大学里，它被定义为 $e^{\sqrt{2}\ln 2}$，而我们用过的那个事实，即 2^t 不论 t 是有理数还是无理数都随 t 递增，是由指数函数（exponential function）的性质推出的，这些性质会在大学的第一门分析课中严格证明。（盖尔范德（Gelfond）与施奈德（Schneider）的一条著名定理表明，$2^{\sqrt{2}}$ 其实是超越数（transcendental number）。）至于 e^x 和 $\ln x$ 的严格构造，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP II 2008，第 6 题

纯数学 · 和化积公式

STEP II 2008，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/08-s2-q6

关键思路

一个余弦与一个正弦相加，而两者的辐角又不相同，处理起来并不方便；但两个余弦之和可以用和化积公式（sum-to-product formula）分解成乘积。先把正弦改写成余弦；这样 f 的零点（zeros）就是两个简单因子的零点，而两个因子同时为零的那一点，正是曲线与轴相切之处。至于最后一部分，要记得余弦和正弦各自最大只能是 1。

解答

以下一律记

$$A = 2x + \frac{\pi}{3}, \qquad \sin\left(\frac{3x}{2} - \frac{\pi}{4}\right) = \cos\left(\frac{\pi}{2} - \frac{3x}{2} + \frac{\pi}{4}\right) = \cos B, \quad \text{其中} B = \frac{3\pi}{4} - \frac{3x}{2},$$

于是 $f(x) = \cos A + \cos B$。

第 (i) 部分

$\cos(2x + \frac{\pi}{3})$ 这一项的周期（period）是 π，而 $\sin(\frac{3x}{2} - \frac{\pi}{4})$ 的周期是 $\frac{2\pi}{3/2} = \frac{4\pi}{3}$。公共周期必须是两者各自的整数倍：$p\pi = q \cdot \frac{4\pi}{3}$ 即 $3p = 4q$，其最小的正数解是 $p = 4, q = 3$。所以对所有 x 都有 $f(x + 4\pi) = f(x)$，而 f 的周期是

$$4\pi.$$

更小的数都不行：我们会在第 (iii) 部分看到，$f(x) = 2$ 当且仅当 $x = \frac{11\pi}{6} + 4t\pi$（$t$ 为整数），所以 f 取得最大值的各点彼此相隔 4π，任何周期都必定是 4π 的倍数。

第 (ii) 部分

利用 $\cos A+\cos B=2\cos\frac{A+B}{2}\cos\frac{A-B}{2}$，其中

$$\frac{A+B}{2}=\frac{1}{2}\left(\frac{x}{2}+\frac{13\pi}{12}\right)=\frac{x}{4}+\frac{13\pi}{24},$$
$$\frac{A-B}{2}=\frac{1}{2}\left(\frac{7x}{2}-\frac{5\pi}{12}\right)=\frac{7x}{4}-\frac{5\pi}{24},$$

我们得到

$$f(x)=2\cos\left(\frac{x}{4}+\frac{13\pi}{24}\right)\cos\left(\frac{7x}{4}-\frac{5\pi}{24}\right).$$

所以 $f(x)=0$ 当且仅当两个余弦之中有一个为零。

第一个因子。由 $\frac{x}{4}+\frac{13\pi}{24}=\frac{\pi}{2}+k\pi$ 得 $x=-\frac{\pi}{6}+4k\pi$。落在 $[-\pi,\pi]$ 内的只有 $x=-\frac{\pi}{6}$。

第二个因子。由 $\frac{7x}{4}-\frac{5\pi}{24}=\frac{\pi}{2}+k\pi$ 得 $\frac{7x}{4}=\frac{17\pi}{24}+k\pi$，所以

$$x=\frac{(17+24k)\pi}{42}.$$

我们需要 $-42\le 17+24k\le 42$，所以 $k=-2,-1,0,1$，得出 $x=-\frac{31\pi}{42}$，$-\frac{7\pi}{42}=-\frac{\pi}{6}$，$\frac{17\pi}{42}$，$\frac{41\pi}{42}$。

综合起来，f 在 $-\pi\le x\le\pi$ 内的零点是

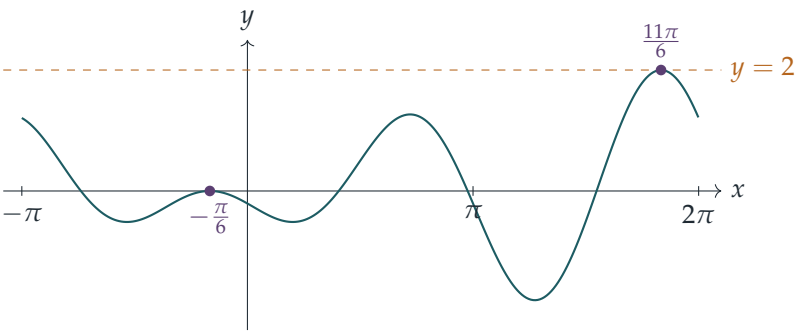
$$x=-\frac{31\pi}{42},\quad -\frac{\pi}{6},\quad \frac{17\pi}{42},\quad \frac{41\pi}{42}.$$

想一想 为甚么在两张清单中都出现的值会与别不同?

在 $x=-\frac{\pi}{6}$ 处，两个余弦因子都是零。每个因子都是一个穿过零点的余弦，所以各自在该处变号（其导数（derivative）不为零，因为在 cos 为 0 之处 sin 是 ±1）。因此在 $x=-\frac{\pi}{6}$ 附近，f 的表现就像一个常数乘以 $(x+\frac{\pi}{6})^2$：两次变号相乘，结果是没有变号。事实上，当 x 递增经过 $-\frac{\pi}{6}$ 时，第一个因子由正变负（其辐角向上穿过 $\frac{\pi}{2}$），第二个因子由负变正（其辐角向上穿过 $-\frac{\pi}{2}$），所以 f 在两侧都是负数。因此曲线与 x-轴相切而不穿过它之处是

$$x=-\frac{\pi}{6}.$$

在其余三个零点，只有一个因子为零，所以 f 变号，曲线穿过轴。



第 (iii) 部分

由于 $\cos A \le 1$ 及 $\cos B \le 1$，我们有 $f(x) \le 2$，等号成立当且仅当 $\cos A = 1$ 并且 $\cos B = 1$。以原来的函数表示：

$$\cos\left(2x + \frac{\pi}{3}\right) = 1 \iff 2x + \frac{\pi}{3} = 2m\pi \iff x = -\frac{\pi}{6} + m\pi,$$
$$\sin\left(\frac{3x}{2} - \frac{\pi}{4}\right) = 1 \iff \frac{3x}{2} - \frac{\pi}{4} = \frac{\pi}{2} + 2n\pi \iff x = \frac{\pi}{2} + \frac{4n\pi}{3},$$

其中 m, n 为整数。两者同时成立的条件是 $-\frac{\pi}{6} + m\pi = \frac{\pi}{2} + \frac{4n\pi}{3}$；两边乘以 $\frac{6}{\pi}$，即 $6m - 1 = 3 + 8n$，也就是

$$3m - 4n = 2.$$

其中一个解是 $m = 2, n = 1$，而通解是 $m = 2 + 4t, n = 1 + 3t$（因为 3 与 4 互质 (coprime)）。所以 $f(x) = 2$ 当且仅当

$$x = -\frac{\pi}{6} + (2 + 4t)\pi = \frac{11\pi}{6} + 4t\pi, \qquad t \in \mathbb{Z}.$$

在 $0 \le x \le 2\pi$ 内，这样的值只有

$$x = \frac{11\pi}{6}.$$

验算：$2 \cdot \frac{11\pi}{6} + \frac{\pi}{3} = 4\pi$，而 $\frac{3}{2} \cdot \frac{11\pi}{6} - \frac{\pi}{4} = \frac{5\pi}{2}$，所以 $f = \cos 4\pi + \sin\frac{5\pi}{2} = 1 + 1 = 2$。

验算

- 在 10^4 个随机点上，因式分解后的形式与原来的 f 一致（误差在 10^{-12} 之内）。
- 以数值方法扫描 $[-\pi, \pi]$，只在 $-\frac{31\pi}{42}$、$\frac{17\pi}{42}$ 和 $\frac{41\pi}{42}$ 附近发现变号；在 $-\frac{\pi}{6}$ 处，我们发现两侧都有 $f(-\frac{\pi}{6} \pm 10^{-3}) \approx -8.7 \times 10^{-7}$，证实这是一个相切而不变号的零点。
- f 在 $[0, 2\pi]$ 上的最大值是 2，在 $x \approx 1.8333\pi = \frac{11\pi}{6}$ 处取得。

常见错误

- 把 $-\frac{\pi}{6}$ 列出两次，或者没有留意到它同时来自两个因子，而这恰恰就是找出相切点的线索。
- 分别解出 $\cos A = 1$ 和 $\cos B = 1$ 之后就停下来：两个条件必须在同一个 x 上成立，这就需要整数方程 $3m - 4n = 2$。

STEP 之外

第 (iii) 部分其实是在问两个周期运动何时会同步：条件 $3m - 4n = 2$ 是一条线性丢番图方程（linear Diophantine equation），它有解是因为 $\gcd(3,4)$ 整除 2。同一个想法若同时应用于多个频率，就会引向傅里叶级数（Fourier series），以及这个问题：周期函数之和在甚么情况下本身也是周期函数（只有当各周期之比为有理数时才是）。想了解背后的数论，可以读 Davenport 的 *The Higher Arithmetic*（一本数论入门经典）。

详细解答

STEP II 2008，第 7 题

纯数学 · 一阶微分方程

STEP II 2008，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/08-s2-q7

关键思路

左边的 $\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x}$ 正是 $\ln y$ 的导数（derivative），所以处理这个代换（substitution）的最佳方法是取对数（logarithm）：乘积 $y = u\,(1+x^2)^{1/2}$ 会变成和，而右边的 $\frac{x}{1+x^2}$ 一项恰好是 $\ln(1+x^2)^{1/2}$ 的导数。这一项消去后，剩下的是一条关于 u 的可分离变量方程（separable equation）。第 (ii) 部分是同一个想法，只是换成立方根；把两个答案并排一看，第 (iii) 部分的规律便一目了然。

解答

第 (i) 部分

在 $x = 0$ 处有 $y = 1 \neq 0$，而方程中含有 $1/y$，所以我们在 $x = 0$ 附近一个使 $y \neq 0$ 的区间上进行讨论。由于 $(1+x^2)^{1/2} > 0$，由代换可知在该区间上亦有 $u \neq 0$，并且 $u(0) = 1$。

对 $|y| = |u|\,(1+x^2)^{1/2}$ 取对数，然后求导，得

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

与原来的微分方程（differential equation）比较，$\frac{x}{1+x^2}$ 各项互相消去，剩下

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

这里两边除以 u，由于 $u \neq 0$，这样做是容许的。

? **想一想** 怎样才能干净利落地求 $x(1+x^2)^{1/2}$ 的积分？

因子 x 是 $1+x^2$ 的导数的一半，所以由逆向链式法则（reverse chain rule）得 $\int x(1+x^2)^{1/2}\,\mathrm{d}x = \frac{1}{2}\cdot\frac{2}{3}(1+x^2)^{3/2} = \frac{1}{3}(1+x^2)^{3/2}$，再加上一个常数。（不妨求导验证一下。）

两边对 x 积分，得

$$-\frac{1}{u} = \tfrac{1}{3}(1+x^2)^{3/2} + C.$$

当 $x=0$ 时，$u=1$，所以 $-1 = \frac{1}{3}+C$，即 $C=-\frac{4}{3}$。因此

$$\frac{1}{u} = \frac{4-(1+x^2)^{3/2}}{3}, \qquad u = \frac{3}{4-(1+x^2)^{3/2}},$$

于是

$$\boxed{y = \frac{3\,(1+x^2)^{1/2}}{4-(1+x^2)^{3/2}}}$$

这个解在包含 0 而分母为正的区间上成立，亦即 $(1+x^2)^{3/2}<4$，即 $|x|<\sqrt{2^{4/3}-1}\approx 1.233$；当 $|x|$ 趋近这个值时，$y\to\infty$。

第 (ii) 部分

这次多出来的一项是 $\frac{x^2}{1+x^3}$，它是 $\frac{1}{3}\ln(1+x^3) = \ln(1+x^3)^{1/3}$ 的导数。这提示我们设 $y = u\,(1+x^3)^{1/3}$，并在 $x=0$ 附近（该处 $1+x^3>0$）进行讨论。与上面完全一样，

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

所以方程变成

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

这次 x^2 是 $1+x^3$ 的导数的三分之一，所以

$$-\frac{1}{u} = \tfrac{1}{3}\cdot\tfrac{3}{4}(1+x^3)^{4/3} + C = \tfrac{1}{4}(1+x^3)^{4/3} + C.$$

由 $u=1$ 于 $x=0$，得 $-1=\frac{1}{4}+C$，所以 $C=-\frac{5}{4}$。于是 $\frac{1}{u} = \frac{5-(1+x^3)^{4/3}}{4}$，并且

$$\boxed{y = \frac{4\,(1+x^3)^{1/3}}{5-(1+x^3)^{4/3}}}$$

这个解在 0 附近成立，范围是 $-1<x<\left(5^{3/4}-1\right)^{1/3}\approx 1.33$。

第 (iii) 部分

比较两个答案：分子出现的是方根 $(1+x^n)^{1/n}$，分母出现的是幂 $(1+x^n)^{(n+1)/n}$，而常数分别是 $n+1$ 和 $n+2$。因此我们猜想（conjecture）

$$\boxed{y=\frac{(n+1)\,(1+x^n)^{1/n}}{n+2-(1+x^n)^{(n+1)/n}}}$$

当 $n=2$ 和 $n=3$ 时，这个式子正好给出 (i) 和 (ii) 的答案。（题目没有要求证明，但用同样的代换 $y=u(1+x^n)^{1/n}$ 便可证明：$\int x^{n-1}(1+x^n)^{1/n}\,\mathrm{d}x=\frac{1}{n+1}(1+x^n)^{(n+1)/n}$，而 $x=0$ 处的条件给出常数 $-\frac{n+2}{n+1}$。）

验算

- 当 $x=0$ 时，每条公式都给出 $\frac{n+1}{n+2-1}=1$，符合要求。
- 在 (i) 中 $x=0$ 附近，由方程可得 $y'(0)=0$，而 $(1+x^2)^{1/2}$ 和 $(1+x^2)^{3/2}$ 在 0 处的斜率确实都是零。
- 以数值方法解这些方程（由 $x=0$ 出发，用龙格－库塔（Runge–Kutta）法逐步计算），所得结果与公式相符，误差小于 10^{-8}，已对 $n=2,3,4,5,7$ 在多个点上作过检验。例如在 (i) 中当 $x=1$ 时，$y=\frac{3\sqrt{2}}{4-2\sqrt{2}}\approx 3.6213$；在 (ii) 中当 $x=1$ 时，$y=\frac{4\cdot 2^{1/3}}{5-2^{4/3}}\approx 2.0320$。

常见错误

- 代入 $y=u(1+x^2)^{1/2}$ 后直接用积法则（product rule）求导：这样做是可行的，但式子很繁琐。先把全式除以 y（等价于取对数），相消的情况便一目了然。
- 忘记 $\int u^{-2}\,\mathrm{d}u=-u^{-1}$ 中的负号，以致常数出错。
- 在错误的阶段确定常数：应在移项整理之前代入 $u(0)=1$（而不是在关于 u 的方程中代入 $y(0)=1$）。

STEP 之外

这条方程是一条*伯努利方程*（*Bernoulli equation*）：$y'=P(x)y+Q(x)y^2$，通用的技巧是代换 $v=1/y$，把它化为线性（linear）方程。本题所给的代换则分两步完成同一件事：先用类似积分因子（integrating factor）的 $(1+x^n)^{1/n}$ 消去含 P 的一项，然后分离变量。还要留意，虽然方程看来平平无奇，它的解却在有限的 x 处趋向无穷；这种“有限时间爆破（finite-time blow-up）”是含有 y^2 项的方程的典型现象，在常微分方程（ODE）的

入门课程中会学到。Boyce 与 DiPrima 的 *Elementary Differential Equations*（常微分方程入门教科书）涵盖了伯努利方程和解的存在性。

详细解答

STEP II 2008，第 8 题

纯数学 · 向量与角平分线

STEP II 2008，第 8 题

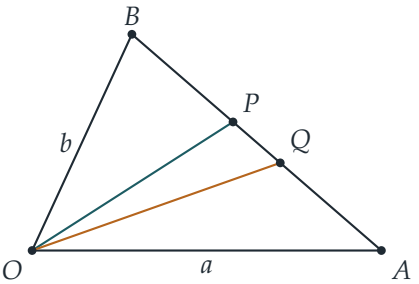
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/08-s2-q8

关键思路

沿 $\angle AOB$ 的角平分线（angle bisector）方向的向量，就是沿 OA 和 OB 的单位向量（unit vector）之和：把两个长度相等的向量相加，得到的是菱形（rhombus）的对角线，而菱形的对角线平分其顶角。把这个向量与 P 的位置向量（position vector）对照，便可定出 λ。之后，Q 不过是把 A 与 B 的角色互换后的 P，而最后的恒等式只需用纯量积（scalar product）直接计算，当中棘手的一项 $\mathbf{a} \cdot \mathbf{b}$ 会互相抵销。

解答



图中 $a > b$，所以 P（位于角平分线上，以蓝绿色标示）较接近 B，而 Q 与 B 的距离，等于 P 与 A 的距离。

P 的位置向量

由于 $AP : PB = (1-\lambda) : \lambda$，而这两段合起来就是 AB，所以 $AP = (1-\lambda)AB$，因此 $\overrightarrow{AP} = (1-\lambda)(\mathbf{b}-\mathbf{a})$，并且

$$\overrightarrow{OP} = \mathbf{a} + (1-\lambda)(\mathbf{b}-\mathbf{a}) = \lambda\mathbf{a} + (1-\lambda)\mathbf{b}.$$

（验证：$\lambda = 1$ 时得 A，$\lambda = 0$ 时得 B，与该比例相符。）

求 λ

向量 $\mathbf{a}/a$ 和 $\mathbf{b}/b$ 分别是沿 OA 和 OB 的单位向量。两者之和是一个菱形的对角线，该菱形的边分别沿 OA 和 OB，所以它指向 $\angle AOB$ 的角平分线方向。由于 P 位于这条角平分线上，

$$\lambda\mathbf{a} + (1-\lambda)\mathbf{b} = k\left(\frac{\mathbf{a}}{a} + \frac{\mathbf{b}}{b}\right)$$

其中 k 为某个纯量。

> ? **想一想** 为甚么可以分别比较 $\mathbf{a}$ 和 $\mathbf{b}$ 的系数?

因为 O、A、B 不共线（collinear），$\mathbf{a}$ 和 $\mathbf{b}$ 均非零向量，而且互不平行，所以 $\mathbf{a}$ 和 $\mathbf{b}$ 线性独立（linearly independent）：若 $\alpha\mathbf{a} + \beta\mathbf{b} = \mathbf{0}$，则 $\alpha = \beta = 0$。因此

$$\lambda = \frac{k}{a}, \qquad 1-\lambda = \frac{k}{b}.$$

两式相加，得 $1 = k\left(\frac{1}{a} + \frac{1}{b}\right)$，所以 $k = \dfrac{ab}{a+b}$，并且

$$\boxed{\lambda = \frac{b}{a+b}}, \qquad 1-\lambda = \frac{a}{a+b}.$$

这其实就是角平分线定理（angle bisector theorem）的另一种面貌：$AP : PB = (1-\lambda) : \lambda = a : b = OA : OB$。所以

$$\overrightarrow{OP} = \frac{b\,\mathbf{a} + a\,\mathbf{b}}{a+b}.$$

点 Q 与恒等式

Q 在 AB 上，且 $BQ = AP = (1-\lambda)AB$，所以沿用求 P 时的推论，只是改由 B 而非 A 量起，

$$\overrightarrow{OQ} = \mathbf{b} + (1-\lambda)(\mathbf{a}-\mathbf{b}) = (1-\lambda)\mathbf{a} + \lambda\mathbf{b} = \frac{a\,\mathbf{a} + b\,\mathbf{b}}{a+b}.$$

设 $c = \mathbf{a}\cdot\mathbf{b}$，并利用 $\mathbf{a}\cdot\mathbf{a} = a^2$、$\mathbf{b}\cdot\mathbf{b} = b^2$。于是

$$(a+b)^2\,OQ^2 = (a\mathbf{a} + b\mathbf{b})\cdot(a\mathbf{a} + b\mathbf{b}) = a^2\cdot a^2 + 2ab\,c + b^2\cdot b^2 = a^4 + b^4 + 2abc,$$
$$(a+b)^2\,OP^2 = (b\mathbf{a} + a\mathbf{b})\cdot(b\mathbf{a} + a\mathbf{b}) = b^2a^2 + 2ab\,c + a^2b^2 = 2a^2b^2 + 2abc.$$

两式相减，含 c 的项互相抵销：

$$(a+b)^2\left(OQ^2-OP^2\right)=a^4-2a^2b^2+b^4=(a^2-b^2)^2=(a-b)^2(a+b)^2.$$

两边除以 $(a+b)^2\neq 0$，得

$$OQ^2-OP^2=(a-b)^2=(b-a)^2,$$

这正是所求。

验算

取 $\mathbf{a}=(3,0)$ 和 $\mathbf{b}=(0,4)$，则 $a=3$、$b=4$，而 $\angle AOB$ 是直角。这时 $\lambda=\frac{4}{7}$，$\overrightarrow{OP}=\left(\frac{12}{7},\frac{12}{7}\right)$，它位于 $y=x$ 上，即该直角的角平分线。另外 $\overrightarrow{OQ}=\left(\frac{9}{7},\frac{16}{7}\right)$，所以 $OQ^2-OP^2=\frac{337}{49}-\frac{288}{49}=1=(4-3)^2$。若 $a=b$，三角形是等腰三角形，$P=Q$ 是 AB 的中点，两者之差为 0，正如所料。以二维和三维空间中 20,000 个随机三角形作数值测试，所有结果均吻合，误差在 10^{-13} 以内。

常见错误

- 写成 $\overrightarrow{OP}=(1-\lambda)\mathbf{a}+\lambda\mathbf{b}$：比例中较大的一份 $1-\lambda$ 是与 A 的距离，所以它应该配 $\mathbf{b}$。可用 $\lambda=1$ 检验。
- 用 $\mathbf{a}+\mathbf{b}$ 作为角平分线的方向。这只在 $a=b$ 时才成立；必须使用单位向量。
- 把 $|\mathbf{a}+\mathbf{b}|^2$ 展开成 a^2+b^2，漏掉了交叉项（cross term）$2\,\mathbf{a}\cdot\mathbf{b}$。本题中这一项会抵销，但前提是两个展开式中都有它。

STEP 之外

最后的计算，其实是长度在线性映射（linear map）下如何变换的一个特例：OP^2 和 OQ^2 是二次型（quadratic form）$\mathbf{x}\mapsto\mathbf{x}\cdot\mathbf{x}$ 的值，而交叉项则由 $\mathbf{a}$ 和 $\mathbf{b}$ 的格拉姆矩阵（Gram matrix）决定。单位向量相加得出角平分线这一事实亦可推广：在任何内积空间（inner-product space）中，要找出两个向量“正中间”的方向，做法就是先归一化（normalise）再相加。Strang 的 *Introduction to Linear Algebra*（线性代数入门）正是本着这种精神阐述纯量积、长度和角度。

详细解答

STEP II 2008，第 9 题

力学 · 抛体运动

STEP II 2008，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：

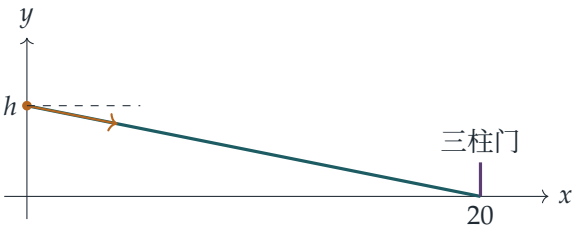
step.maths.org/questions/08-s2-q9

关键思路

水平方向的运动决定了以 α 表示的飞行时间（time of flight），把这个时间代入垂直方向的运动，便得到一条连系 α 与 h 的方程。利用 $\sec^2\alpha = 1 + \tan^2\alpha$，可把它化为关于 $\tan\alpha$ 的二次方程（quadratic），这是处理“以哪个角度才能击中目标?”一类问题的标准技巧。之后，凭着对板球的常识选出合适的根，再以小角近似（small-angle approximation）完成余下的工作。

解答

我们把球看成一个质点（particle），只受重力作用而自由运动，不计空气阻力，也不计旋转。取抛射点正下方地面上的一点为原点，x 轴沿水平方向指向三柱门，y 轴垂直向上，于是球从 $(0, h)$ 出发，速率为 $40\,\mathrm{m\,s^{-1}}$，方向与水平线成仰角 α（α 为负数表示球是向下投出的）。



第 (i) 部分

水平方向没有力的作用，所以 $x = 40t\cos\alpha$。球到达 $x = 20$ 的时刻为

$$t = \frac{20}{40\cos\alpha} = \frac{1}{2\cos\alpha}.$$

垂直方向的加速度为 $-10\,\mathrm{m\,s^{-2}}$，所以

$$y = h + 40t\sin\alpha - 5t^2.$$

此时球位于地面，即 $y = 0$，所以

$$0 = h + 40\sin\alpha\cdot\frac{1}{2\cos\alpha} - \frac{5}{4\cos^2\alpha} = h + 20\tan\alpha - \tfrac{5}{4}\sec^2\alpha.$$

现在利用 $\sec^2\alpha = 1 + \tan^2\alpha$，并记 $T = \tan\alpha$：

$$h + 20T - \tfrac{5}{4}(1 + T^2) = 0 \quad\Longleftrightarrow\quad 5T^2 - 80T + 5 - 4h = 0.$$

由二次公式（quadratic formula），

$$T = \frac{80 \pm \sqrt{6400 - 20(5-4h)}}{10} = \frac{80 \pm \sqrt{6300 + 80h}}{10},$$

所以两个可能的值是

$$\tan\alpha = 8 \pm \sqrt{63 + \tfrac{4}{5}h}.$$

两者皆为实数，因为 $h \geq 0$。

想一想 快速投球手实际上会用哪一个根?

由于 $\sqrt{63} \approx 7.9$，较大的根给出 $\tan\alpha \approx 16$，即 $\alpha \approx 86°$：这是一记冲天高吊球，会升高约 80 m，并要大约 8 s 才落到三柱门上。快速投球手不会这样投球。较小的根，

$$\tan\alpha = 8 - \sqrt{63 + \tfrac{4}{5}h},$$

才是合适的选择。对于合乎现实的出手高度（约两米左右），$63 + \frac{4}{5}h$ 接近 64，所以 $\tan\alpha$ 接近 0。例如，只要 h 介乎 0 与 3 之间，便有 $|\tan\alpha| < 0.09$。因此 α 是一个小角，而 $\cos\alpha \approx 1$，球到达三柱门所需的时间为

$$t = \frac{1}{2\cos\alpha} \approx \tfrac{1}{2}\,\mathrm{s},$$

正如题目所求。（事实上 $t = \frac{1}{2}\sqrt{1 + \tan^2\alpha}$，只要 $|\tan\alpha| < 0.1$，它与 $\frac{1}{2}$ 的差距便不超过 0.5%。）

第 (ii) 部分

球向水平线以下投出，当且仅当 $\tan\alpha < 0$。较大的根恒为正数，所以我们采用较小的根：

$$8 - \sqrt{63 + \tfrac{4}{5}h} < 0 \iff \sqrt{63 + \tfrac{4}{5}h} > 8 \iff 63 + \tfrac{4}{5}h > 64 \iff h > \tfrac{5}{4},$$

其中平方这一步可逆，因为两边皆为正数。所以当 $h > 1.25$ 时，投球手是向水平线以下投出球的。当 $h = 1.25$ 恰好成立时，球沿水平方向投出。

第 (iii) 部分

当 $h = 2.5$ 时，由第 (ii) 部分可知 α 是一个小的负角。取 $\cos\alpha \approx 1$ 和 $\sin\alpha \approx \alpha$（以弧度（radians）计），便有 $\tan\alpha \approx \alpha$ 和 $\sec^2\alpha \approx 1$，所以第 (i) 部分的方程 $0 = h + 20\tan\alpha - \frac{5}{4}\sec^2\alpha$ 变成

$$0 \approx 2.5 + 20\alpha - 1.25, \qquad \text{所以} \qquad \alpha \approx -\frac{1.25}{20} = -0.0625 \text{ rad}.$$

换算成度数，

$$\alpha \approx -0.0625 \times \frac{180}{\pi} \approx -3.58^\circ,$$

所以 $\alpha \approx -3.6^\circ$。球出手时的方向在水平线以下约 3.6°。

另一个看法：利用二项式近似（binomial approximation），$\tan\alpha = 8 - \sqrt{65} = 8 - 8\sqrt{1 + \frac{1}{64}} \approx 8 - 8\left(1 + \frac{1}{128}\right) = -\frac{1}{16}$。这给出相同的答案。

验算

- *精确值。* $\tan\alpha = 8 - \sqrt{65} = -0.06226$，所以 $\alpha = -3.563^\circ$，准确至两位有效数字是 -3.6°，与近似结果吻合。
- *模拟。* 从 $(0, 2.5)$ 出发，以 $40\,\text{m}\,\text{s}^{-1}$ 和 -3.563° 逐步作数值计算，球在 $0.501\,\text{s}$ 后到达 $x = 20$，此时高度为 $0.0000\,\text{m}$。
- *临界情况。* 当 $h = 1.25$ 时，较小的根是 $8 - \sqrt{64} = 0$：球沿水平方向投出，在半秒的飞行中下坠 $5(0.5)^2 = 1.25\,\text{m}$，刚好落在三柱门处。

常见错误

- 忘记 $\sec^2\alpha = 1 + \tan^2\alpha$，而试图分别以 $\sin\alpha$ 和 $\cos\alpha$ 求解，结果变得十分繁复。
- 以弧度给出角度，或在度数制下使用小角近似：$\sin\theta \approx \theta$ 只在 θ 以弧度计时才成立。
- 在第 (iii) 部分漏掉负号。正如第 (ii) 部分所预示，这个角是在水平线以下的。

STEP 之外

两个根的结构反映了一个普遍事实：以固定速率抛出的抛体，可沿两条轨迹到达其“包络线”（envelope，即安全抛物线（parabola of safety））以内的任何一点。在包络线上，两条轨迹合而为一，而包络线以外的点则完全无法到达。真实的板球还会受到阻力（drag）以及由旋转产生的马格努斯效应（Magnus effect）影响，这会引出非线性微分方程（nonlinear differential equations），通常要用数值方法求解。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）以大学一年级的程度讨论抛体运动和有阻力的运动。

详细解答

STEP II 2008，第 10 题

力学 · 与墙壁的斜碰撞

STEP II 2008，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/08-s2-q10

关键思路

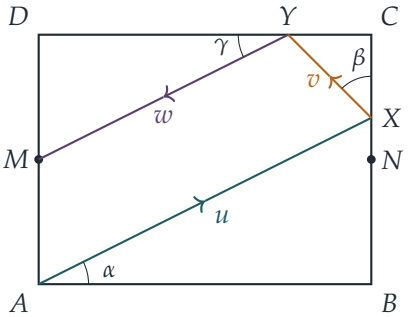
在光滑（smooth）的桌边上，冲量（impulse）垂直于桌边，所以沿桌边方向的速度分量（component）不变，只有垂直撞向桌边的分量会乘以 e。把速度分解成平行于 AB 和平行于 AD 的两个分量，一切便一目了然：每次反弹只会改变两个分量的其中一个。两次反弹之后，两个分量都各乘了 e，最终的方向和动能（kinetic energy）便同时确定下来。至于第 (ii) 部分的几何，只需把各段水平距离加起来。

解答

我们把球看作质点（particle），把桌边看作光滑的墙。取 A 为原点，x 轴沿 AB，y 轴沿 AD，于是

$$B = (a, 0), \quad C = (a, 2b), \quad D = (0, 2b), \quad M = (0, b), \quad N = (a, b).$$

为简便起见，记 $t = \tan\alpha$。



第 (i) 部分

最初的速度是 $(u\cos\alpha,\ u\sin\alpha)$。

*在 X 的碰撞。*边 BC 平行于 y 轴。由于它是光滑的，球所受的冲量沿 x 方向，所以 y 分量 $u\sin\alpha$ 不变。根据牛顿恢复定律（Newton's law of restitution），x 分量会反向并乘以 e。碰撞后的速度是

$$(-eu\cos\alpha,\ u\sin\alpha).$$

角 $\beta = \angle CXY$ 是从 XC 量起的，而这个方向指向 y 轴的正方向。所以

$$\tan\beta = \frac{eu\cos\alpha}{u\sin\alpha} = \frac{e}{\tan\alpha}, \qquad \text{即} \qquad \tan\alpha\tan\beta = e,$$

这正是所要证明的。

*在 Y 的碰撞。*边 DC 平行于 x 轴，所以这次是 x 分量不变，而 y 分量反向并乘以 e。碰撞后的速度是

$$(-eu\cos\alpha,\ -eu\sin\alpha) = -e\,(u\cos\alpha,\ u\sin\alpha).$$

这是初速度的负数倍，所以路径的最后一段平行于 AX，而且方向相反。角 $\gamma = \angle DYM$ 是从 YD 量起的，而这个方向指向 x 轴的负方向；方向 $(-\cos\alpha, -\sin\alpha)$ 与它的夹角为 α。因此

$$\gamma = \alpha.$$

第 (ii) 部分

我们沿着路径追踪 x 坐标。

*由 A 到 X。*球横越宽度 a 的同时上升了 $a\tan\alpha$，所以 $X = (a,\ at)$。

*由 X 到 Y。*球再上升 $2b - at$，而在这一段，它每上升一个单位，就在 x 方向移动 $\tan\beta$ 个单位（因为 β 是从竖直的边量起的）。利用第 (i) 部分的 $\tan\beta = e/t$，球向左移动的距离是 $(2b - at)\,e/t$，所以

$$x_Y = a - \frac{e(2b - at)}{t}.$$

*由 Y 到 M。*球由高度 $2b$ 下降到高度 b，即下降了 b，路径在水平线之下成角 $\gamma = \alpha$。所以走过的水平距离是 $b/\tan\alpha$，而这段距离必须刚好把球带到 $x = 0$：

$$x_Y = \frac{b}{t}.$$

想一想 为甚么只追踪 x 坐标就足够?

每一段直线路径都从上一段的终点开始，而它的方向已由第 (i) 部分得知。只要知道球在某一段上升或下降了多少，角度就决定了它横向移动多远。各段端点的高度都是固定的

（先是 o，然后在 DC 处是 $2b$，再在 M 处是 b），所以余下的唯一条件，就是各段横向距离加起来要吻合。

令 x_Y 的两个表达式相等，并乘以 t：

$$\begin{aligned} b &= at - e(2b - at) \\ &= at(1+e) - 2be, \end{aligned}$$

所以 $at(1+e) = b(1+2e)$，由此得

$$\tan\alpha = \frac{(1+2e)b}{(1+e)a},$$

这正是所要证明的。

这一击总是可行的。对任何 $e > 0$，这个 $\tan\alpha$ 的值都是正数，所以它确定了 $(0, \frac{\pi}{2})$ 内的一个角 α。我们还要确认各个反弹点确实落在应该落在的边上。

- X 的高度是 $at = \dfrac{1+2e}{1+e}\,b$。由于当 $e > 0$ 时 $1+e < 1+2e < 2+2e$，这个高度严格介乎 b 与 $2b$ 之间。所以 X 确实在 BC 上，位于 C 之下，并且在球袋 N 之上，球不会在途中掉进 N。
- Y 位于 $x_Y = b/t = \dfrac{1+e}{1+2e}\,a$，当 $e > 0$ 时这个值严格介乎 0 与 a 之间。所以 Y 确实在 DC 上，位于 D 与 C 之间。

最后一段则由 Y 沿直线走到 M，全程都在球桌之内。因此，无论 e 取甚么值，路径的每一段都是可行的。

第 (iii) 部分

由第 (i) 部分可知，末速度是初速度的 $-e$ 倍，所以 $w = eu$。（这也可以直接看出：每个分量都恰好乘了一次 e。）动能由 $\frac{1}{2}mu^2$ 减至 $\frac{1}{2}me^2u^2$，其中 m 是球的质量。因此损失的比例是

$$\frac{\frac{1}{2}mu^2 - \frac{1}{2}me^2u^2}{\frac{1}{2}mu^2} = 1 - e^2.$$

这个结果与 a、b 和各个角度都无关。（若 $e > 1$，这个“损失”是负数，即桌边反而给球增添能量。这并不符合物理实际，但代数运算依然成立。）

验算

- 取 $a = 4$、$b = 1.5$ 和 $e = \frac{1}{2}$（即图中的情况），公式给出 $\tan\alpha = \frac{1}{2}$。于是 $X = (4, 2)$，$Y = (3, 3)$，而由 Y 出发、斜率（gradient）为 $\frac{1}{2}$ 的直线在水平距离 3 之内下降 1.5，刚好落在 $M = (0, 1.5)$。损失的动能比例是 $1 - \frac{1}{4} = \frac{3}{4}$。

- 当 $e = 1$（完全弹性，perfectly elastic）时，$\tan\alpha = 3b/(2a)$。把球桌先对 BC 反射、再对 DC 反射而“展开”：路径变成由 A 到 M 的像的一条直线，横向走了 $2a$，向上走了 $3b$。这给出相同的斜率 $3b/(2a)$。没有能量损失，与 $1 - e^2 = 0$ 吻合。
- 我们对 a、b、e 和 u 随机取了 20,000 组值来模拟反弹，结果证实了 $\tan\alpha\tan\beta = e$、$\gamma = \alpha$、球到达 M，以及比例 $1 - e^2$，误差全部在 10^{-13} 之内。

常见错误

- 在碰撞时把 e 乘到整个速率上。其实只有垂直于桌边的分量会改变，沿光滑桌边的分量丝毫不变。
- 从 BC 的法线（normal）而不是从 XC 量度 β，结果把 $\tan\beta$ 变成了它的倒数。
- 忘记验证 X 在 N 之上，以及 Y 在 D 与 C 之间。这项验证正是“这一击可行”的内容。

STEP 之外

当 $e = 1$ 时，把球桌反射来“展开”桌球路径，可以把路径变成一条直线。这是数学桌球（mathematical billiards）的起点，它与动力系统（dynamical systems）相关，而对于有理球桌，更与曲面上的几何相关。当 $e < 1$ 时，反射不再像镜面反射，但把速度分解成分量后仍然容易处理。关于冲量与恢复的力学，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）。想做更多这类题目，可以试试 Siklos 的 *Advanced Problems in Mathematics*（高等数学难题集）。

详细解答

STEP II 2008，第 11 题

力学 · 移动楔上的质点

STEP II 2008，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：

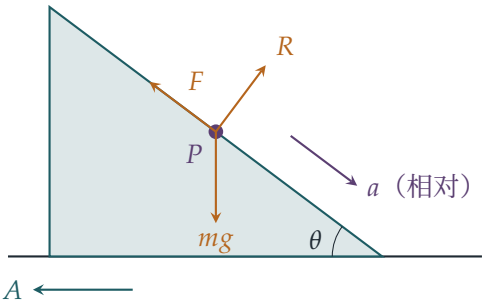
step.maths.org/questions/08-s2-q11

关键思路

系统以外没有任何东西沿水平方向施力（地面是光滑的），所以质点（particle）与楔块（wedge）的水平动量（momentum）一直保持为零；楔块的加速度几乎不费工夫就得到了。其余一切，都来自把质点的*真实*加速度写成“它相对于楔块的加速度，加上楔块的加速度”，然后沿斜面方向和垂直于斜面方向分解（resolve）质点所受的力。静止的观察者所见的路径，是一条沿真实加速度方向的直线，因为一切都由静止开始，而且各加速度都是常数。

解答

我们把 P 视为质点，把楔块视为刚体（rigid body）。地面光滑；斜面粗糙，摩擦系数（coefficient of friction）为 μ。把铅垂面放在左边，于是斜面向右下方倾斜。当 P 向右下方滑动时，楔块被推向左方。设楔块的加速度为向*左*的 A，并设 P *相对于楔块*沿斜面向下的加速度为 a。设 R 为法向反作用力（normal reaction），F 为作用于 P 的摩擦力（friction）；当 P 向下滑动时，摩擦力沿斜面向上。



取 x 轴向右、y 轴向上。沿斜面向下的单位向量（unit vector）是 $(\cos\theta, -\sin\theta)$，所以 P 的真实加速度是

$$\big(a\cos\theta - A,\ -a\sin\theta\big).$$

第 (i) 部分：楔块的加速度

作用于“P 加楔块”这个系统的外力，只有重力和光滑地面的法向反作用力，两者都是铅垂的。所以总水平动量是常数，而由于一切都由静止开始，这个常数是零。因此 $m\times$（加速度）的水平分量之和为零：

$$m(a\cos\theta - A) + km(-A) = 0 \quad\Longrightarrow\quad A = \frac{a\cos\theta}{k+1},$$

这正是所要求的结果。（P 与楔块之间的力 R 和 F 是内力，根据牛顿第三定律（Newton's third law）互相抵消。）

第 (i) 部分：45° 路径

? **想一想** 为甚么观察到的路径必定是一条沿真实加速度方向的直线?

下文的运动方程（equations of motion）确定了 a 和 A 都是常数，所以 P 的真实加速度是一个常向量。由静止开始，它的速度总是该向量的 t 倍，位移则是该向量的 $\frac{1}{2}t^2$ 倍，所以它沿着该方向的一条固定直线运动。

P 的加速度的水平分量是

$$a\cos\theta - \frac{a\cos\theta}{k+1} = \frac{k}{k+1}\,a\cos\theta,$$

而铅垂分量的大小是 $a\sin\theta$。路径是一条 45° 的直线，即表示这两者相等：

$$a\sin\theta = \frac{k}{k+1}\,a\cos\theta \quad\Longrightarrow\quad \tan\theta = \frac{k}{k+1},$$

其中两边除以了 $a\cos\theta \neq 0$（因为 P 确实在滑动，$a > 0$）。这正是所要求的结果。

第 (i) 部分：$k = 3$ 的情况

现在 $\tan\theta = \frac{3}{4}$，所以 $\sin\theta = \frac{3}{5}$，$\cos\theta = \frac{4}{5}$，而 $A = \frac{a\cdot 4/5}{4} = \frac{a}{5}$。

我们对 P 分解各力。它的真实加速度沿斜面向下的分量是 $a - A\cos\theta$，沿向外法线方向的分量是 $-A\sin\theta$（楔块向左的加速度有这两个分量；相对加速度则完全沿着斜面）。

垂直于斜面方向：

$$R - mg\cos\theta = -mA\sin\theta \quad\Longrightarrow\quad R = m\Big(\tfrac{4}{5}g - \tfrac{3}{25}a\Big).$$

留意 R 比 $mg\cos\theta$ 小：楔块正在离 P 而去。

沿斜面向下方向，由于 P 正在滑动，所以 $F=\mu R$：

$$mg\sin\theta-\mu R=m(a-A\cos\theta)=m\left(a-\tfrac{4}{25}a\right)=\tfrac{21}{25}ma.$$

代入 R 并乘以 $25/m$：

$$15g-20\mu g+3\mu a=21a\quad\Longrightarrow\quad a=\frac{5g(3-4\mu)}{3(7-\mu)}.$$

这个值恰好在 $\mu<\frac{3}{4}=\tan\theta$ 时为正，而 P 要滑动就必须如此；这正好带我们进入第 (ii) 部分。

第 (ii) 部分

我们断言：若 $\tan\theta\le\mu$，则 P 留在原处，楔块也不会移动。

静止是可能的。若两个物体都不动，则对 P 而言，我们需要 $R=mg\cos\theta$，以及沿斜面向上的摩擦力 $F=mg\sin\theta$。只要 $F\le\mu R$，即 $\sin\theta\le\mu\cos\theta$，亦即 $\tan\theta\le\mu$，粗糙的接触面就能提供这个力，而这正是我们的条件。楔块所受的则是方向相反的 R 和 F；它们的水平分量分别是向左的 $R\sin\theta=mg\sin\theta\cos\theta$ 和向右的 $F\cos\theta=mg\sin\theta\cos\theta$，两者互相抵消。所以楔块也处于平衡（equilibrium），而地面只提供一个铅垂的力。

> ? **想一想** 即使静止是可能的，P 会不会仍然滑动?

滑动是不可能的。假设 P 确实向下滑动，且 $F=\mu R$，其中 k 和 θ 为一般的值。设 $s=\sin\theta-\mu\cos\theta$。楔块的水平方程是 $kmA=R\sin\theta-\mu R\cos\theta=Rs$。把它与法向方程 $R=m(g\cos\theta-A\sin\theta)$ 结合，得

$$R=\frac{kmg\cos\theta}{k+s\sin\theta},\qquad a=\frac{(k+1)\,gs}{k+s\sin\theta}.$$

要保持接触，我们需要 $R>0$，这迫使 $k+s\sin\theta>0$；于是 a 与 s 同号。但 $\tan\theta\le\mu$ 表示 $s\le0$，所以 $a\le0$，这与向下滑动矛盾。也没有任何东西能使 P 沿斜面向上滑动，因为只有重力才能引起相对运动，而重力的分量是沿斜面向下的。所以当 $\tan\theta\le\mu$ 时，P 在楔块上保持静止，楔块也在地面上保持静止。（当 $\tan\theta=\mu$ 时，摩擦力达到极限值（limiting friction），但 P 仍然不动。）

验算

- 取 $\mu=0$ 和 $k=3$，我们得到 $a=\frac{5g}{7}$ 和 $A=\frac{g}{7}$。于是 P 的真实加速度的水平分量是 $\frac{4}{5}\cdot\frac{5g}{7}-\frac{g}{7}=\frac{3g}{7}$，铅垂分量是 $\frac{3}{5}\cdot\frac{5g}{7}=\frac{3g}{7}$，方向确实是 45°。

- 当 $k \to \infty$（楔块无法移动）时，$A \to 0$，而一般公式给出 $a \to g(\sin\theta - \mu\cos\theta)$，即大家熟悉的固定斜面结果。
- 数值上，取 $g = 9.8$、$k = 3$ 和 $\mu = 0.2$，直接解三条运动方程得 $a \approx 5.284\ \mathrm{m\,s^{-2}}$，与 $\frac{5g(3-0.8)}{3(6.8)}$ 吻合。把运动随时间逐步推算，水平动量一直保持为零，所得路径的倾角恰好是 45°。

常见错误

- 取 $R = mg\cos\theta$。这只在楔块固定时才成立；这里楔块加速离开，所以 R 较小。
- 分解时只用相对加速度 a，忘记了楔块的加速度。牛顿第二定律需要的是真实加速度。
- 在第 (ii) 部分，即使 P 没有滑动，仍假设摩擦力是 μR。摩擦力最多是 μR，实际取值视乎需要而定。

STEP 之外

第 (i) 部分最巧妙的一步是一条守恒定律（conservation law）：由于外力没有水平分量，动量的其中一个分量守恒。在大学一年级的力学课程中，这会发展成一个把对称性（symmetry）与守恒定律联系起来的一般原理：在拉格朗日力学（Lagrangian mechanics）中，若拉格朗日量（Lagrangian）不依赖某个坐标（这里是整个系统的水平位置），就会自动得出一个守恒的动量。光滑楔块与物块正是那里标准的入门例子。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）详细讲解了正是这一类的问题。

详细解答

STEP II 2008，第 12 题

概率与统计 · 独立事件

STEP II 2008，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/08-s2-q12

关键思路

一个计数的期望值（expectation），等于构成它的各个事件的概率之和，所以第一部分只是一条关于 p 的二次方程。至于第二部分，法官们在心中掷硬币时根本没有看被告一眼，所以国王的宣判与真相独立（independent）。把宣判正确的概率用 t 表示，再围绕 $\frac{1}{2}$ 重新整理，答案一行便出来。

解答

求 p

设 A、B、C 分别为驴子、河狸和半人马裁定有罪的事件，并设 I_A、I_B、I_C 为它们的指示变量（indicator variable）（事件发生时等于 1，否则等于 0）。那么 $X = I_A + I_B + I_C$，而由于 $\mathrm{E}(I_A) = \mathrm{P}(A)$，其余类推，由期望值的线性性质（linearity of expectation）可得

$$\mathrm{E}(X) = p + \frac{p}{3} + p^2.$$

（线性性质并不需要独立性；我们稍后才会用到独立性。）令上式等于 $\frac{4}{3}$，再乘以 3：

$$3p^2 + 4p - 4 = 0 \quad \Longrightarrow \quad (3p - 2)(p + 2) = 0.$$

概率不可能是负数，所以 $p + 2 \neq 0$，因此

$$p = \frac{2}{3}.$$

于是三位法官说“有罪”的概率分别为

$$\mathrm{P}(A) = \frac{2}{3}, \qquad \mathrm{P}(B) = \frac{2}{9}, \qquad \mathrm{P}(C) = \frac{4}{9},$$

全部都在 $[0, 1]$ 之内，正如所应有的。

国王说“有罪”的概率

设 K 为最少两位法官裁定有罪的事件，即 K 是国王宣判被告有罪的事件。把上面三个概率记作 a、b、c。恰好两个有罪裁决有三种情况，三个全是有罪则有一种情况；利用法官之间的独立性，

$$\begin{aligned}\mathrm{P}(K) &= ab(1-c) + ac(1-b) + bc(1-a) + abc \\ &= ab + ac + bc - 2abc.\end{aligned}$$

这里 $ab = \frac{4}{27}$，$ac = \frac{8}{27}$，$bc = \frac{8}{81}$，$abc = \frac{16}{243}$，所以

$$\mathrm{P}(K) = \frac{36}{243} + \frac{72}{243} + \frac{24}{243} - \frac{32}{243} = \frac{100}{243}.$$

为简便起见，把它记作 g。

求 t

设 D 为被告确实有罪的事件，所以 $\mathrm{P}(D) = t$。当 D 与 K 同时发生，或两者皆不发生时，国王的宣判便是正确的。

> ? **想一想** 为甚么我们可以把 $\mathrm{P}(D)$ 乘以 $\mathrm{P}(K)$?

法官们的决定是“完全随机”的：他们的裁决并不取决于证据，所以 K 与 D 独立。因此

$$\mathrm{P}(\text{正确}) = \mathrm{P}(D)\,\mathrm{P}(K) + \mathrm{P}(D')\,\mathrm{P}(K') = tg + (1-t)(1-g).$$

现在减去 $\frac{1}{2}$ 并因式分解（factorise）：

$$\begin{aligned}tg + (1-t)(1-g) - \tfrac{1}{2} &= 2tg - t - g + \tfrac{1}{2} \\ &= (2t-1)\left(g - \tfrac{1}{2}\right).\end{aligned}$$

所以国王宣判正确的概率为 $\frac{1}{2}$，当且仅当 $(2t-1)(g-\frac{1}{2}) = 0$。由于 $g = \frac{100}{243} \neq \frac{1}{2}$，第二个因式不等于零，故必有

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

不用因式分解也可以得到同样的结果：由 $\frac{100}{243}t + \frac{143}{243}(1-t) = \frac{1}{2}$ 得 $143 - 43t = \frac{243}{2}$，所以 $43t = \frac{43}{2}$，即 $t = \frac{1}{2}$。

留意 p 的值在这里几乎无关重要：不论法官们的概率是多少，只要 $\mathrm{P}(K) \neq \frac{1}{2}$，一个无视证据的法庭有一半机会判对，当且仅当有罪与无辜的被告同样常见。

验算

► 取 $p = \frac{2}{3}$：$\frac{2}{3} + \frac{2}{9} + \frac{4}{9} = \frac{12}{9} = \frac{4}{3}$，符合要求。

- 用分数精确地列举全部八种裁决组合，得 $\mathrm{P}(K) = \frac{100}{243} \approx 0.412$ 及 $\mathrm{E}(X) = \frac{4}{3}$。
- 取 $t = \frac{1}{2}$ 进行 200 000 次模拟试验，国王宣判正确的个案所占比例为 0.5007。
- 检查公式是否合理：若 $t = 1$（人人有罪），国王判对的概率为 g；若 $t = 0$，则为 $1 - g$；这个概率是 t 的线性函数，所以它在中点处经过 $\frac{1}{2}$。

常见错误

- 把“最少两个”当作 $ab + ac + bc$。这样会把三个全中的情况数了三次；正确的表达式要减去 $2abc$（或者直接把四个互斥的情况相加）。
- 忘记舍去 $p = -2$，或者没有检查 p^2 和 $p/3$ 是否有效的概率。
- 把“国王宣判正确”只当作 $\mathrm{P}(K)$，忘记了那些获正确判定无罪的无辜被告。

STEP 之外

第一部分是一个强大技巧的小例子：把一个计数写成指示变量之和，于是它的期望值就只是概率之和，即使各事件并不独立亦然。大学本科的概率论经常使用这个技巧，例如用来求随机排列（random permutation）的不动点（fixed point）的期望数目。第二部分讲的是一个输出与真相独立的分类器（classifier），这正是思考检测中的灵敏度（sensitivity）、特异度（specificity）和贝叶斯定理（Bayes′ theorem）的起点。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门）大量运用指示变量，是很好的下一步读物。

详细解答

STEP II 2008，第 13 题

概率与统计 • 不放回抽取

STEP II 2008，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/08-s2-q13

关键思路

不必追踪每一枚筹码：只追踪黑色的就够了。每枚黑色筹码的去向由两次抽取决定，而每次抽取时，某一枚指定筹码被抽中的机会就是（抽出的数目）/（袋中的数目）。按第一次抽取的结果分情况讨论，两部分都会化成简短的和，而且化简的程度远超你的预期。

解答

在整个解答中，“随机抽出”是指袋中每一组 k 枚筹码被抽出的可能性均等。我们反复用到的事实是：若从一个装有 m 枚筹码的袋中随机抽出 k 枚，则根据对称性（symmetry），某一枚指定的筹码在其中的概率为 k/m（这 m 枚筹码各自被抽中的可能性均等，而平均有 k 枚被抽中）。

第一次转移之后，袋 P 有 $n-k$ 枚筹码，袋 Q 有 $n+k$ 枚。第二次转移时，从这 $n+k$ 枚中抽出 k 枚，移到 P。

第 (i) 部分

设 B 为那枚黑色筹码，它起初在 P 中。按 B 在第一次转移时有没有移动来分情况。

- B 没有从 P 中被抽出：概率为 $\dfrac{n-k}{n}$。这时 B 自始至终留在 P 中，因为第二次转移只会从 Q 取出筹码。
- B 从 P 中被抽出：概率为 $\dfrac{k}{n}$。这时 B 是 Q 中 $n+k$ 枚筹码之一，而它回到 P 当且仅当它在被抽回的 k 枚之中，其概率为 $\dfrac{k}{n+k}$。

这两种情况互斥（disjoint），并且涵盖了所有可能，所以

$$\begin{aligned}\mathrm{P}(B\text{ 最终在}P) &= \frac{n-k}{n} + \frac{k}{n}\cdot\frac{k}{n+k} = \frac{(n-k)(n+k)+k^2}{n(n+k)}\\ &= \frac{n^2}{n(n+k)} = \frac{n}{n+k}.\end{aligned}$$

这个概率随 k 严格递减，所以在 $0 \le k \le n$ 的范围内，它只在 $k=0$ 时取得最大值，此时其值为 1。（这并不意外：如果甚么都不移动，黑色筹码必定留在原处。）

第 (ii) 部分

现在设 B_1 为起初在 P 中的黑色筹码，B_2 为起初在 Q 中的那一枚。设 S 为两者最终在同一个袋中的事件。同样按 B_1 在第一次转移时有没有移动来分情况。

情况 1：B_1 留在 P 中（概率为 $\frac{n-k}{n}$）。这时无论之后发生甚么，B_1 最终都在 P 中，所以 S 发生当且仅当 B_2 在从 $n+k$ 枚筹码中抽出的 k 枚之中，这些筹码都在 Q 内：

$$\mathrm{P}(S \mid \text{情况 1}) = \frac{k}{n+k}.$$

情况 2：B_1 移到 Q（概率为 $\frac{k}{n}$）。这时两枚黑色筹码都在 Q 的 $n+k$ 枚筹码之中，而 S 发生的条件是第二次抽取把它们两枚都取走，或者两枚都不取走。

> **想一想** 从 $n+k$ 枚筹码中随机抽出一个 k 元子集，它同时包含两枚指定筹码，或者两枚都不包含的概率是多少？

计数子集的数目。共有 $\binom{n+k}{k}$ 种可能性均等的选法。同时包含两枚黑色筹码的选法，是从余下的 $n+k-2$ 枚中选出另外 $k-2$ 枚；两枚都不包含的选法，则是从 $n+k-2$ 枚白色筹码中选出全部 k 枚。相除可得

$$\begin{aligned}\mathrm{P}(\text{两枚都有}) &= \binom{n+k-2}{k-2}\Big/\binom{n+k}{k} = \frac{k(k-1)}{(n+k)(n+k-1)},\\ \mathrm{P}(\text{两枚都没有}) &= \binom{n+k-2}{k}\Big/\binom{n+k}{k} = \frac{n(n-1)}{(n+k)(n+k-1)}.\end{aligned}$$

（这些公式在 $k=0,1$ 时同样成立，此时 $\mathrm{P}(\text{两枚都有}) = 0$。）所以

$$\mathrm{P}(S \mid \text{情况 2}) = \frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)}.$$

把各情况合并。根据全概率定律（law of total probability），

$$\begin{aligned}\mathrm{P}(S) &= \frac{n-k}{n}\cdot\frac{k}{n+k} + \frac{k}{n}\cdot\frac{k(k-1)+n(n-1)}{(n+k)(n+k-1)}\\ &= \frac{k\big[(n-k)(n+k-1)+k^2-k+n^2-n\big]}{n(n+k)(n+k-1)}.\end{aligned}$$

在方括号内，$(n-k)(n+k-1) = n^2 - k^2 - n + k$，所以方括号等于 $2n^2 - 2n = 2n(n-1)$。因此

$$\boxed{\mathrm{P}(S) = \frac{2k(n-1)}{(n+k)(n+k-1)}}$$

求最大值。当 $k = 0$ 时概率为 0，所以取 $1 \le k \le n$，并求其倒数的最小值。撇开正常数 $2(n-1)$ 不计，

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

所以我们需要在整数范围内求 $g(k) = k + \dfrac{n(n-1)}{k}$ 的最小值。比较相邻两项：

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

当 $k(k+1) < n(n-1)$，即 $k < n-1$ 时，这个差为负数；而当 $k = n-1$ 时，它恰好为零，因为此时 $k(k+1) = (n-1)n$。所以 g 由 $k = 1$ 到 $k = n-1$ 严格递减，并且 $g(n) = g(n-1)$。因此概率取得最大值之处为

$$k = n-1 \quad \text{及} \quad k = n,$$

由于 $n \ge 2$，两者都在容许的范围内。当 $k = n-1$ 时，最大值为

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

而 $k = n$ 给出相同的值，$\dfrac{2n(n-1)}{2n(2n-1)} = \dfrac{n-1}{2n-1}$。（把 g 看成连续函数时，其最小值点为 $\sqrt{n(n-1)}$，严格位于 $n-1$ 与 n 之间；这两个整数处的值是完全相等，而非近似相等。）

验算

- 第 (i) 部分取 $k = n$：得 $\frac{1}{2}$。这时全部 $2n$ 枚筹码都经过 Q，其中 n 枚返回，所以黑色筹码最终落在两个袋中的可能性均等。吻合。
- 第 (ii) 部分取 $k = 0$ 得 0，这是正确的，因为两枚黑色筹码永远不会相遇。取 $n = 2$、$k = 1$ 得 $\frac{2}{3\cdot 2} = \frac{1}{3}$，你可以亲手验证。
- 我们对 $2 \le n \le 7$ 及每一个 k，以分数精确地列举了每一对可能的抽取结果；两条公式在所有情况下都吻合，而 (ii) 中取得最大值之处总是 $\{n-1, n\}$，例如最大值为 $\frac{3}{7} \approx 0.4286$，对应 $n = 4$。

常见错误

- 在第二次抽取时用了 k/n 而不是 $k/(n+k)$：袋 Q 已增至 $n+k$ 枚筹码。

- 在第 (ii) 部分的情况 2 中，忘记了“两枚都没有被抽出”的可能：如果两枚黑色筹码都留在 Q 中，它们仍然在一起。
- 用微积分求最大值，再把 $\sqrt{n(n-1)}$ 舍入为单一个整数。离散的比较显示，有两处同样取得最大值。

STEP 之外

两部分背后的工具是超几何分布（hypergeometric distribution）：在不放回的随机抽样中，样本内带记号物件的数目。先以第一阶段的结果为条件（conditioning），再把它用于第二阶段，这就是全概率定律；同样的两阶段推理，也是容器之间混合与扩散模型的基础，例如埃伦费斯特瓮模型（Ehrenfest urn）。Blitzstein and Hwang, *Introduction to Probability*（概率论入门教材）正是以这种方式处理超几何分布与条件概率。

详细解答

STEP III 2008，第 1 题

纯数学 · 非线性联立方程

STEP III 2008，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/08-s3-q1

关键思路

设 $s_k = ax^k + by^k$，则题目的方程给出了 s_0, s_1, s_2, s_3。由于 x 和 y 是 $t^2 - pt + q = 0$ 的两个根，其中 $p = x + y$、$q = xy$，这些幂和（power sums）满足一个递推关系（recurrence）：每个 s_k 等于 p 乘以前一项，再减去 q 乘以再前一项。这正是提示中“乘以 $x + y$”所得到的结果。于是四条非线性方程化为关于 p 和 q 的两条线性方程。之后解一条二次方程便得 x 和 y，再由首两条方程求出 a 和 b。

解答

设 $s_k = ax^k + by^k$，其中 $k = 0, 1, 2, 3$，则 $s_0 = 1$、$s_1 = \frac{1}{3}$、$s_2 = \frac{1}{5}$、$s_3 = \frac{1}{7}$。令 $p = x + y$ 及 $q = xy$。

幂和的递推关系

对任意 a, b, x, y 及任意 $k \geq 1$，

$$(x + y)(ax^k + by^k) = ax^{k+1} + by^{k+1} + xy\,(ax^{k-1} + by^{k-1}),$$

因为交叉项是 $axyx^{k-1}$ 和 $bxy\,y^{k-1}$。用我们的记号，即

$$s_{k+1} = p\,s_k - q\,s_{k-1}.$$

这是一条恒等式（identity），所以对方程组的每一个解都成立。

求 $x+y$ 和 xy

取 $k=1$ 和 $k=2$，并代入已知数值：

$$\begin{aligned}\tfrac{1}{5} &= \tfrac{1}{3}p - q,\\ \tfrac{1}{7} &= \tfrac{1}{5}p - \tfrac{1}{3}q.\end{aligned}$$

由第一式得 $q=\frac{1}{3}p-\frac{1}{5}$。代入第二式，

$$\tfrac{1}{7}=\tfrac{1}{5}p-\tfrac{1}{9}p+\tfrac{1}{15} \quad\Longrightarrow\quad \tfrac{4}{45}p=\tfrac{1}{7}-\tfrac{1}{15}=\tfrac{8}{105} \quad\Longrightarrow\quad p=\tfrac{6}{7},$$

于是 $q=\frac{2}{7}-\frac{1}{5}=\frac{3}{35}$。这个线性方程组的行列式（determinant）不为零，所以这些数值是唯一确定的：每一个解都满足 $x+y=\frac{6}{7}$ 及 $xy=\frac{3}{35}$。

求 x 和 y

因此 x 和 y 是以下方程的两个根：

$$t^2-\tfrac{6}{7}t+\tfrac{3}{35}=0, \qquad \text{即} \qquad 35t^2-30t+3=0,$$

由此得

$$t=\frac{30\pm\sqrt{900-420}}{70}=\frac{30\pm4\sqrt{30}}{70}=\frac{15\pm2\sqrt{30}}{35}.$$

判别式（discriminant）为正，所以 $x\neq y$，而 $\{x,y\}$ 就是这两个数，次序可以任选其一。

? **想一想** 知道了 x 和 y 之后，能否肯定 a 和 b 也随之确定?

可以。x 和 y 已知之后，首两条方程 $a+b=1$ 和 $ax+by=\frac{1}{3}$ 是关于 a 和 b 的线性方程，其行列式为 $y-x$。由于 $x\neq y$，这个行列式不为零，所以恰好有一对 (a,b)。

求 a 和 b

取 $x=\frac{15+2\sqrt{30}}{35}$ 及 $y=\frac{15-2\sqrt{30}}{35}$，则 $x-y=\frac{4\sqrt{30}}{35}$。从 $a+b=1$ 和 $ax+by=\frac{1}{3}$ 消去 b，得 $a(x-y)=\frac{1}{3}-y$。现在

$$\tfrac{1}{3}-y=\frac{35-3(15-2\sqrt{30})}{105}=\frac{6\sqrt{30}-10}{105},$$

所以

$$a=\frac{6\sqrt{30}-10}{105}\cdot\frac{35}{4\sqrt{30}}=\frac{6\sqrt{30}-10}{12\sqrt{30}}=\frac{1}{2}-\frac{5}{6\sqrt{30}}=\frac{1}{2}-\frac{\sqrt{30}}{36},$$

而 $b=1-a=\frac{1}{2}+\frac{\sqrt{30}}{36}$。

这些确实是解

我们至此只证明了任何解都必定是这个样子，所以仍须验证它确实成立。按构造方式，$s_0 = 1$ 及 $s_1 = \frac{1}{3}$。由于 p 和 q 满足上面两条线性方程，递推关系给出 $s_2 = ps_1 - qs_0 = \frac{1}{5}$，继而 $s_3 = ps_2 - qs_1 = \frac{1}{7}$。所以四条方程全部成立。

把 x 和 y 的角色互换，a 和 b 也随之互换。因此恰好有两个解：

$$(a, b, x, y) = \left(\frac{18 - \sqrt{30}}{36}, \frac{18 + \sqrt{30}}{36}, \frac{15 + 2\sqrt{30}}{35}, \frac{15 - 2\sqrt{30}}{35} \right)$$

以及把 $a \leftrightarrow b$ 和 $x \leftrightarrow y$ 同时互换所得的解。在两个解中，x, y 之中较大者都与较小的系数配对。

验算

数值上，$x \approx 0.741556$、$y \approx 0.115587$、$a \approx 0.347855$ 及 $b \approx 0.652145$。于是 $ax + by \approx 0.333333$、$ax^2 + by^2 \approx 0.200000$ 及 $ax^3 + by^3 \approx 0.142857$，与 $\frac{1}{3}$、$\frac{1}{5}$ 和 $\frac{1}{7}$ 吻合至机器精度。再作一个简单的合理性检查：x 和 y 都在 $(0, 1)$ 之内，而 a 和 b 都是正数。这与右边各数是 t^{2k} 在 $[0, 1]$ 上的平均值一致（见下文）。

常见错误

- 解出了 x 和 y，却忘记配对是有关系的：a 与 x 相配，所以不能只互换 x 和 y 而不同时互换 a 和 b。
- 找到必要条件（necessary conditions）便停下来。题目说的是“求所有解”，所以既要证明每一个解都是这个形式，也要证明这些数值确实满足全部四条方程。

STEP 之外

$1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}$ 这些数就是 $\frac{1}{2}\int_{-1}^{1} t^{2k}\,\mathrm{d}t$。所以这个方程组其实是在找一条法则 $\frac{1}{2}\int_{-1}^{1} f \approx \frac{a}{2}\big(f(\sqrt{x}) + f(-\sqrt{x})\big) + \frac{b}{2}\big(f(\sqrt{y}) + f(-\sqrt{y})\big)$，使它对次数不超过 7 的多项式都精确成立。这就是四点高斯—勒让德求积法（*Gauss–Legendre quadrature*）：它的节点（nodes）是 $\pm\sqrt{x}$ 和 $\pm\sqrt{y}$，而 $a \approx 0.3479$ 和 $b \approx 0.6521$ 就是它著名的权（weights）。像我们这里所做的那样，由幂和找出递推关系，也是信号处理中 Prony 方法（Prony's method）背后的想法。想进一步了解，可参阅 Süli 与 Mayers 的 *An Introduction to Numerical Analysis*（数值分析入门）。

详细解答

STEP III 2008，第 2 题

纯数学 · 幂和

STEP III 2008，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/08-s3-q2

关键思路

$(r+1)^k - r^k$ 的和可以裂项相消（telescoping），最后只剩下一项。但我们也可以用二项式定理（binomial theorem）把每个括号展开，这样同一个和就变成较低次幂和 $S_j(n)$ 的组合。令两者相等，便得到一条递推关系（recurrence），可由前面各个幂和逐一建立 S_{k-1}；而要看出第 (ii) 部分的多项式结构，所需的也正是这条递推关系。

解答

第 (i) 部分

首先是裂项相消。把和式逐项写出，

$$\sum_{r=0}^{n}\left[(r+1)^k - r^k\right] = (1^k - 0^k) + (2^k - 1^k) + \cdots + \left((n+1)^k - n^k\right) = (n+1)^k,$$

因为中间各项全部互相抵消，而且 $0^k = 0$（因为 $k \geq 1$）。

其次是二项式定理。对每个 r，

$$(r+1)^k - r^k = \sum_{j=0}^{k-1}\binom{k}{j}r^j = 1 + \binom{k}{1}r + \binom{k}{2}r^2 + \cdots + \binom{k}{k-1}r^{k-1},$$

因为 r^k 项互相抵消。对 $r = 0, 1, \ldots, n$ 求和，常数 1 贡献 $n+1$（r 共有 $n+1$ 个值），而每个 r^j（$j \geq 1$）则贡献 $S_j(n)$。所以

$$(n+1)^k = (n+1) + \binom{k}{1}S_1(n) + \binom{k}{2}S_2(n) + \cdots + \binom{k}{k-2}S_{k-2}(n) + \binom{k}{k-1}S_{k-1}(n).$$

最后一个系数是 $\binom{k}{k-1} = k$。至于其余各项，我们利用对称性 $\binom{k}{j} = \binom{k}{k-j}$，可知 $S_{k-j}(n)$ 的系数是 $\binom{k}{k-j} = \binom{k}{j}$；特别地，$S_1(n)$ 的系数是 $\binom{k}{1} = \binom{k}{k-1}$。把 $kS_{k-1}(n)$ 留在一边，其余各项全部移到另一边，得

$$kS_{k-1}(n) = (n+1)^k - (n+1) - \binom{k}{2}S_{k-2}(n) - \binom{k}{3}S_{k-3}(n) - \cdots - \binom{k}{k-1}S_1(n),$$

这正是所求的 $(*)$。（这里需要 $k \geq 2$，S_{k-1} 才是我们所定义的和之一。）

想一想 为甚么我们把 $j = 0$ 的一项分开处理，而不把它写成 $S_0(n)$?

S_k 只对正数 k 有定义，而且在 $r = 0$ 时 “0^0” 的意义并不明确。把常数抽出来，直接数出它的总和是 $n+1$，就能同时避开这两个问题，也解释了 $(*)$ 中那个单独出现的 $-(n+1)$。

$S_3(n)$。在 $(*)$ 中取 $k = 4$，这时 $\binom{4}{2} = 6$ 及 $\binom{4}{3} = 4$：

$$\begin{aligned} 4S_3(n) &= (n+1)^4 - (n+1) - 6S_2(n) - 4S_1(n) \\ &= (n+1)^4 - (n+1) - n(n+1)(2n+1) - 2n(n+1). \end{aligned}$$

每一项都有因式 $n+1$，所以

$$\begin{aligned} 4S_3(n) &= (n+1)\left[(n+1)^3 - 1 - n(2n+1) - 2n\right] \\ &= (n+1)\left[n^3 + 3n^2 + 3n - 2n^2 - n - 2n\right] \\ &= (n+1)(n^3 + n^2) = n^2(n+1)^2. \end{aligned}$$

因此

$$S_3(n) = \tfrac{1}{4}n^2(n+1)^2 = \big(S_1(n)\big)^2.$$

$S_4(n)$。取 $k = 5$，这时 $\binom{5}{2} = \binom{5}{3} = 10$ 及 $\binom{5}{4} = 5$：

$$5S_4(n) = (n+1)^5 - (n+1) - 10S_3(n) - 10S_2(n) - 5S_1(n).$$

我们再次抽出因式 $n+1$，并代入 S_3、S_2 和 S_1 的表达式：

$$\begin{aligned} 5S_4(n) &= (n+1)\left[(n+1)^4 - 1 - \tfrac{5}{2}n^2(n+1) - \tfrac{5}{3}n(2n+1) - \tfrac{5}{2}n\right] \\ &= (n+1)\Big[n^4 + 4n^3 + 6n^2 + 4n - \tfrac{5}{2}n^3 - \tfrac{5}{2}n^2 \\ &\qquad - \tfrac{10}{3}n^2 - \tfrac{5}{3}n - \tfrac{5}{2}n\Big] \\ &= (n+1)\left[n^4 + \tfrac{3}{2}n^3 + \tfrac{1}{6}n^2 - \tfrac{1}{6}n\right] \\ &= \tfrac{1}{6}n(n+1)\left(6n^3 + 9n^2 + n - 1\right). \end{aligned}$$

由于 S_4 应该像 S_2 一样含有因式 $2n+1$，我们把 $n=-\frac{1}{2}$ 代入这个三次式试一试：$-\frac{3}{4}+\frac{9}{4}-\frac{1}{2}-1=0$。所以 $2n+1$ 是因式，而 $6n^3+9n^2+n-1=(2n+1)(3n^2+3n-1)$。于是

$$S_4(n)=\tfrac{1}{30}\,n(n+1)(2n+1)(3n^2+3n-1).$$

第 (ii) 部分

我们对 k 使用强归纳法 (strong induction)。命题在 $k=1$ 时成立, 因为 $S_1(n)=\frac{1}{2}n^2+\frac{1}{2}n$ 的次数是 2（在 $k=2$ 时亦成立）。

假设对某个 $k\geq 2$，$S_1(n),\ldots,S_{k-1}(n)$ 每一个都是 n 的多项式（polynomial），其中 S_j 的次数是 $j+1$。在 $(*)$ 中以 $k+1$ 代替 k：

$$(k+1)\,S_k(n)=(n+1)^{k+1}-(n+1)-\binom{k+1}{2}S_{k-1}(n)-\cdots-\binom{k+1}{k}S_1(n).$$

在右边：

- $(n+1)^{k+1}$ 是次数为 $k+1$ 的多项式，首项系数（leading coefficient）是 1；
- $n+1$ 的次数是 1；
- 根据归纳假设，每个 $S_j(n)$（$1\leq j\leq k-1$）都是次数为 $j+1\leq k$ 的多项式。

所以右边是一个多项式，它的 n^{k+1} 项只来自 $(n+1)^{k+1}$，不可能被抵消。除以 $k+1\neq 0$ 后，可知 $S_k(n)$ 是次数恰好为 $k+1$ 的多项式（首项系数是 $\frac{1}{k+1}$）。归纳至此完成。

想一想 多项式 $P(n)$ 的“常数项”和“各系数之和”分别对应甚么?

它们就是 $P(0)$ 和 $P(1)$。设 $S_k(n)=P(n)=a_{k+1}n^{k+1}+\cdots+a_1n+a_0$; 对每个整数 $n\geq 0$, 它都与和式 $\sum_{r=0}^{n}r^k$ 相等，所以我们可以代入 $n=0$ 和 $n=1$ 求值。

- 常数项：$a_0=P(0)=S_k(0)=0^k=0$，因为 $k\geq 1$。
- 各系数之和：$a_{k+1}+\cdots+a_1+a_0=P(1)=S_k(1)=0^k+1^k=1$。

两个结果均已证明。

验算

- 较小的数值：$S_3(3)=1+8+27=36=\frac{1}{4}\cdot 9\cdot 16$，以及 $S_4(3)=1+16+81=98=\frac{1}{30}\cdot 3\cdot 4\cdot 7\cdot 35$。
- 用我们的答案检验第 (ii) 部分：$S_4(n)=\frac{1}{5}n^5+\frac{1}{2}n^4+\frac{1}{3}n^3-\frac{1}{30}n$ 的常数项是 0，各系数之和是 $\frac{6+15+10-1}{30}=1$；它的首项系数是 $\frac{1}{5}=\frac{1}{k+1}$，与归纳法所预测的一致。

- 我们以直接求和的方法，对 $2 \le k \le 9$ 及 $0 \le n \le 30$ 准确地验证了 $(*)$，并对 $0 \le n < 200$ 验证了两个闭合形式（closed form）。

常见错误

- 忘记了 $\sum 1$ 中 $r = 0$ 的一项：它的贡献是 $n+1$，而不是 n。
- 没有先抽出公因式 $n+1$，而是把所有项全部展开，令 S_4 的代数运算繁复得多。
- 在第 (ii) 部分只说次数“至多是 $k+1$”。题目要求的是恰好 $k+1$，所以要指出 n^{k+1} 项不可能被抵消。

STEP 之外

这些多项式的系数由伯努利数（Bernoulli numbers）决定，而福尔哈伯公式（Faulhaber's formula）把 $S_k(n)$ 写成 n 的各次幂的组合，其系数是二项式系数（binomial coefficients）乘以伯努利数。$S_k(n)$ 是次数为 $k+1$ 的多项式，首项是 $\frac{n^{k+1}}{k+1}$，这一事实正是 $\int_0^x t^k \,\mathrm{d}t = \frac{x^{k+1}}{k+1}$ 的离散类比，而有限差分学（calculus of finite differences）把这个想法严格化。想做更多同类题目，可试试 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP III 2008，第 3 题

纯数学 · 椭圆的切线与焦点

STEP III 2008，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：

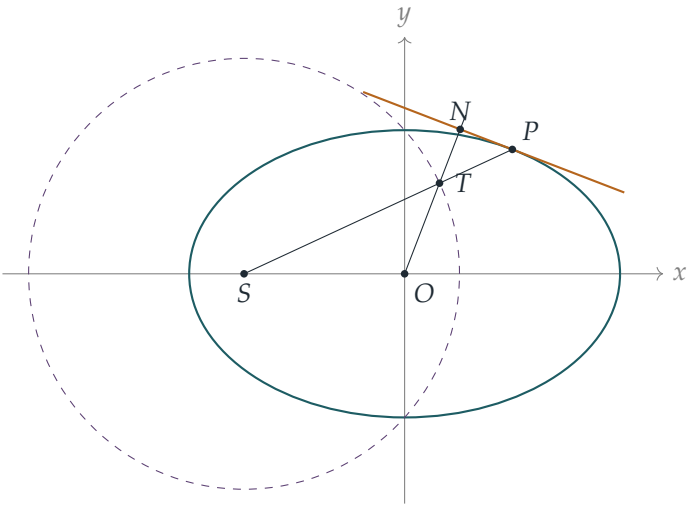
step.maths.org/questions/08-s3-q3

关键思路

整道题只涉及两条直线，所以要做的就是把它们的方程干净利落地写出来，然后求交点。直线 ON 垂直于切线（tangent），所以它的方向就是切线的法向量（normal vector），可以直接从切线方程读出。由 S 出发把 SP 参数化（parametrise），求交点便只需一行计算；而求得的参数原来正是处理圆时所需要的：它告诉我们点 T 位于 SP 上几分之几的位置。

解答

全题中 $0 < e < 1$：我们有 $e^2 = 1 - b^2/a^2$，由于 $a > b > 0$，此值严格介乎 0 与 1 之间。我们亦假设 $\sin\theta \neq 0$。若 $\sin\theta = 0$，则 $P = (\pm a, 0)$，切线为 $x = \pm a$，$N = P$，而 ON 和 SP 都是 x 轴，所以 T 没有定义。



（图中取 $a = 3$，$b = 2$，$\theta = 60°$。椭圆（ellipse）为蓝绿色，在 P 的切线为琥珀色，虚线圆的圆心为 S，半径为 a。）

T 的 y 坐标

对 $x^2/a^2 + y^2/b^2 = 1$ 作隐函数微分（implicit differentiation），或直接引用标准结果，可知在 (x_1, y_1) 的切线为 $xx_1/a^2 + yy_1/b^2 = 1$。在 P 处，这就是

$$\frac{x\cos\theta}{a} + \frac{y\sin\theta}{b} = 1.$$

直线 $px + qy = 1$ 的法向量为 (p, q)，所以切线的法向量为 $\left(\frac{\cos\theta}{a}, \frac{\sin\theta}{b}\right)$，或者乘以 ab 后，写成 $(b\cos\theta, a\sin\theta)$。直线 ON 沿这个方向通过 O（它垂直于切线），所以线上的点为 $\mu(b\cos\theta, a\sin\theta)$，其方程为

$$a\sin\theta\, x = b\cos\theta\, y. \tag{1}$$

现在由 S 出发把 SP 参数化：线上的点为

$$\big(-ea + \lambda a(\cos\theta + e),\ \lambda b\sin\theta\big),$$

其中 $\lambda = 0$ 对应 S，$\lambda = 1$ 对应 P。代入 (1)，并约去因子 $\sin\theta \neq 0$，得

$$\begin{aligned} a\big(-ea + \lambda a(\cos\theta + e)\big) &= \lambda b^2\cos\theta \\ \lambda\big(a^2\cos\theta + a^2e - b^2\cos\theta\big) &= ea^2. \end{aligned}$$

焦点（focus）条件正是在这里派上用场：$a^2 - b^2 = a^2e^2$，所以括号内的式子是 $a^2e^2\cos\theta + a^2e = a^2e(1 + e\cos\theta)$。除以 $a^2e \neq 0$，得

$$\lambda = \frac{1}{1 + e\cos\theta},$$

这一步是合理的，因为 $1+e\cos\theta \geq 1-e > 0$。因此 T 的 y 坐标为

$$y_T = \lambda b \sin\theta = \frac{b\sin\theta}{1+e\cos\theta},$$

正如所求。（留待稍后参考：由 (1) 可得 $x_T = \dfrac{b\cos\theta}{a\sin\theta}\, y_T = \dfrac{a(1-e^2)\cos\theta}{1+e\cos\theta}$。）

T 位于以 S 为圆心、半径为 a 的圆上

想一想 我们已经知道线段 ST 占 SP 的几分之几。还需要知道甚么?

由于 $T = S + \lambda(P-S)$ 且 $\lambda > 0$，我们有 $ST = \lambda\, SP$。所以只需求出长度 SP，即 P 的焦半径（focal distance）：

$$\begin{aligned} SP^2 &= (a\cos\theta + ea)^2 + b^2\sin^2\theta \\ &= a^2\left[\cos^2\theta + 2e\cos\theta + e^2 + (1-e^2)\sin^2\theta\right] \\ &= a^2\left[(\cos^2\theta+\sin^2\theta) + 2e\cos\theta + e^2(1-\sin^2\theta)\right] \\ &= a^2\left[1 + 2e\cos\theta + e^2\cos^2\theta\right] = a^2(1+e\cos\theta)^2, \end{aligned}$$

其中第二行用了 $b^2 = a^2(1-e^2)$。由于 $a > 0$ 且 $1+e\cos\theta > 0$，由此得 $SP = a(1+e\cos\theta)$。于是

$$ST = \lambda\, SP = \frac{a(1+e\cos\theta)}{1+e\cos\theta} = a,$$

所以 T 位于以 S 为圆心、半径为 a 的圆上。

（如果你宁愿直接用坐标计算：$x_T + ea = \dfrac{a(\cos\theta+e)}{1+e\cos\theta}$，而 $(x_T+ea)^2 + y_T^2$ 可化简为 a^2，所用的代数运算与上面计算 SP^2 时完全相同。）

验算

- 当 $\theta = \frac{\pi}{2}$ 时，$P = (0, b)$，切线为 $y = b$，所以 ON 是 y 轴，而 $T = (0, y_T)$，由公式得 $y_T = b$。于是 $ST^2 = e^2a^2 + b^2 = a^2$，与预期相符。
- 取 $a = 3$，$b = 2$，$\theta = 60^\circ$（即图中的情况），以数值方法求两条直线的交点，得 $T \approx (0.4857,\ 1.2618)$；公式则给出 $y_T = \sqrt{3}/(1+\sqrt{5}/6) \approx 1.2618$，而 $ST \approx 3.0000$。
- 在趋向圆的极限 $e \to 0$ 之下，$S \to O$ 且 $T \to P$，而后者确实位于以 O 为圆心、半径为 a 的圆上。

常见错误

- 把 ON 当成平行于切线，而不是垂直于切线：ON 的方向是法向量 $(b\cos\theta, a\sin\theta)$，而不是 $(-a\sin\theta, b\cos\theta)$。
- 忘记把 $a^2 - b^2$ 换成 a^2e^2，结果得到一个无法化简的 λ 表达式。
- 除以 $\sin\theta$ 而不加说明；$\sin\theta = 0$ 的情况确实不成立。

STEP 之外

结果 $SP = a(1 + e\cos\theta) = a + ex_P$ 其实是焦点—准线性质（focus–directrix property）的另一种面貌：到焦点的距离等于到准线 $x = -a/e$ 的距离的 e 倍。以焦点为极点，用极坐标（polar coordinates）把圆锥曲线（conic）写成 $r = \ell/(1 + e\cos\phi)$，正是大学一年级力学课描述开普勒（Kepler）轨道的方式，而同一个分母 $1 + e\cos\theta$ 亦在这里出现了。关于圆锥曲线的几何，以及更多 STEP 风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*；关于轨道，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*。

详细解答

STEP III 2008，第 4 题

纯数学 · 反双曲函数不等式

STEP III 2008，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/08-s3-q4

关键思路

设 $x = \cosh y$。半角公式（half-angle formula）把 $\tanh(y/2)$ 化为 x 的代数函数，于是一个关于 tanh 的显浅不等式（inequality），就变成一个关于 arcosh 的不显浅不等式。之后，题目其实是在运转一部小机器：把不等式由 1 积分至 x，整理一下，便得到一个更强的不等式。把这部机器运转两次，先得出 (ii)，再得出 (iii)。

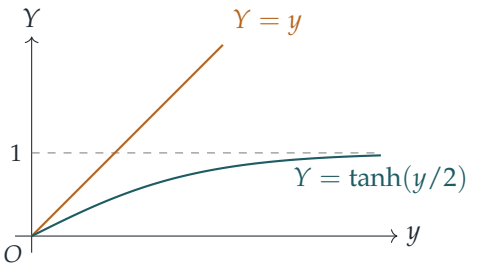
解答

以下一律记 $A(x) = \operatorname{arcosh} x$ 及 $S(x) = \sqrt{x^2-1}$，其中 $x \geq 1$。我们会用到以下两个导数（derivative）：

$$\frac{\mathrm{d}}{\mathrm{d}x} \operatorname{arcosh} x = \frac{1}{\sqrt{x^2-1}}, \qquad \frac{\mathrm{d}}{\mathrm{d}x}\sqrt{x^2-1} = \frac{x}{\sqrt{x^2-1}} \qquad (x > 1).$$

第 (i) 部分

曲线 $Y = \tanh(y/2)$ 通过原点，在原点的斜率（gradient）为 $\frac{1}{2}$，在 $y > 0$ 时是凹（concave）的，而且永远达不到 1；而直线 $Y = y$ 的斜率为 1：



对所有 $y>0$，直线都在曲线之上。要严格说明这一点，设 $g(t)=t-\tanh t$。则 $g(0)=0$，而且

$$g'(t)=1-\operatorname{sech}^2 t=\tanh^2 t>0 \quad \text{对于} t>0,$$

所以 g 严格递增（strictly increasing），而当 $t>0$ 时 $\tanh t<t$。取 $t=y/2$，得

$$\tanh(y/2)<\tfrac{1}{2}y<y \qquad (y>0).$$

现在设 $x>1$，并令 $y=\operatorname{arcosh} x$，于是 $y>0$、$\cosh y=x$ 及 $\sinh y=+\sqrt{x^2-1}$（因为 $y>0$，所以取正值）。利用 $\sinh y=2\sinh\frac{y}{2}\cosh\frac{y}{2}$ 及 $1+\cosh y=2\cosh^2\frac{y}{2}$，

$$\tanh\frac{y}{2}=\frac{\sinh y}{1+\cosh y}=\frac{\sqrt{x^2-1}}{x+1}=\frac{x^2-1}{(x+1)\sqrt{x^2-1}}=\frac{x-1}{\sqrt{x^2-1}}.$$

因此 $y>\tanh(y/2)$ 变成

$$\operatorname{arcosh} x>\frac{x-1}{\sqrt{x^2-1}} \qquad (x>1),$$

这就是 $(*)$。

第 (ii) 部分

固定 $x>1$，以 t 为变量，把 $(*)$ 在 $[1,x]$ 上积分。

> **想一想** $(*)$ 的右边，分母含有 $\sqrt{t^2-1}$。由 $t=1$ 开始积分是否稳妥？积分之后，不等式是否仍然是严格的？

两个问题的答案都是肯定的。当 $t>1$ 时，$\dfrac{t-1}{\sqrt{t^2-1}}=\sqrt{\dfrac{t-1}{t+1}}$，而当 $t\to 1$ 时它趋向 0，所以两边在 $[1,x]$ 上都是连续（continuous）的。两者之差是连续的，并且在 $(1,x]$ 上严格为正，所以它在 $[1,x]$ 上的积分严格为正。

左边用分部积分法（integration by parts），取 $u=\operatorname{arcosh} t$：

$$\int_1^x \operatorname{arcosh} t\,\mathrm{d}t=\Big[t\operatorname{arcosh} t\Big]_1^x-\int_1^x\frac{t}{\sqrt{t^2-1}}\,\mathrm{d}t=x\,A(x)-S(x),$$

因为 $\operatorname{arcosh} 1=0$ 及 $\sqrt{1^2-1}=0$。

右边可拆成两个标准积分：

$$\int_1^x\frac{t-1}{\sqrt{t^2-1}}\,\mathrm{d}t=\int_1^x\frac{t}{\sqrt{t^2-1}}\,\mathrm{d}t-\int_1^x\frac{1}{\sqrt{t^2-1}}\,\mathrm{d}t=S(x)-A(x).$$

因此

$$x\,A(x)-S(x)>S(x)-A(x) \quad\Longrightarrow\quad (x+1)A(x)>2S(x).$$

两边除以 $x+1>0$，并如第 (i) 部分般化简，

$$\operatorname{arcosh} x > \frac{2\sqrt{x^2-1}}{x+1} = 2\,\frac{x-1}{\sqrt{x^2-1}} \qquad (x>1),$$

正是所求。这个下界刚好是 $(*)$ 中下界的两倍。

第 (iii) 部分

第 (ii) 部分说明，当 $t>1$ 时 $\operatorname{arcosh} t > 2\cdot\dfrac{t-1}{\sqrt{t^2-1}}$：右边与 $(*)$ 的右边相同，只是乘大了一倍。所以我们把同一部机器再运转一次。在 $[1,x]$ 上积分（理据与之前相同），并利用已求得的两个积分，

$$x\,A(x) - S(x) > 2\big(S(x)-A(x)\big) \quad\Longrightarrow\quad (x+2)A(x) > 3S(x).$$

由于 $x+2>0$，

$$\operatorname{arcosh} x > 3\,\frac{\sqrt{x^2-1}}{x+2} \qquad (x>1),$$

这就是所求的结果。

想一想 为甚么到此为止？可以再积分一次吗？

这部机器要求右边是 $\dfrac{t-1}{\sqrt{t^2-1}}$ 的倍数，因为我们知道它的积分。新的下界 $3\sqrt{t^2-1}/(t+2)$ 并非这个形式，所以这个技巧到此便用尽了。

作为独立的验证（即题目所谓“或用其他方法”的途径），设 $f(x)=A(x)-3S(x)/(x+2)$。则 $f(1)=0$，而根据商法则（quotient rule），

$$f'(x) = \frac{(x+2)^2 - 3x(x+2) + 3(x^2-1)}{(x+2)^2\sqrt{x^2-1}} = \frac{(x-1)^2}{(x+2)^2\sqrt{x^2-1}} > 0 \quad (x>1),$$

所以 f 严格递增，而当 $x>1$ 时 $f(x)>0$，与之前的结果一致。

验算

在 $x=2$：$\operatorname{arcosh} 2 = \ln(2+\sqrt{3}) \approx 1.3170$，而三个下界分别是 $1/\sqrt{3}\approx 0.577$、$2/\sqrt{3}\approx 1.155$ 及 $3\sqrt{3}/4 \approx 1.299$。每个下界都比前一个好，而且全部低于 1.3170。在 $x=1$ 附近，这些下界变得非常精确：在 $x=1.001$，$\operatorname{arcosh} x \approx 0.04471763$，(ii) 的下界只差约 7×10^{-6}，而 (iii) 的下界更只差约 10^{-9}，因为每个下界与 $\operatorname{arcosh} x$ 吻合至 $x-1$ 的更高阶。我们亦以数值方法检验了那两个积分（在 $x=1.5, 2, 5$ 用中点求积法（midpoint quadrature）），并在 10 000 个随机选取的 $x>1$ 检验了全部三个不等式，均没有发现任何出入。

常见错误

- 取 $\sinh(\operatorname{arcosh} x) = \pm\sqrt{x^2-1}$ 而没有确定正负号。因为 $\operatorname{arcosh} x > 0$，所以应取正根。
- 由 0 而不是由 1 开始积分：arcosh 只在 $x \geq 1$ 时有定义，而正是下限 1 令边界项（boundary terms）消失。
- 忘记说明为何积分后的不等式仍然是严格的。

STEP 之外

下界愈来愈好，是因为每积分一次，误差在 $x = 1$ 处消失的阶数便高一阶；这是一系列愈来愈精确的有理型近似的开端，它们是帕德逼近（Padé approximants）的近亲。全题贯彻使用的原理（若 $f(a) = 0$ 且 $f' > 0$，或其积分形式成立，则 $f > 0$）其实是中值定理（mean value theorem）的化身，也是大量分析学结果的基础。Abbott 的 *Understanding Analysis* 对中值定理和积分有严谨的论述。

详细解答

STEP III 2008，第 5 题

纯数学 · 递推关系

STEP III 2008，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/08-s3-q5

关键思路

$\mathrm{T}_n^2 - \mathrm{T}_{n-1}\mathrm{T}_{n+1}$ 这个量是递推关系（recurrence）的一个不变量（invariant）：当 n 增加一时，前后两项之差会互相抵销，因为递推关系容许我们把 $\mathrm{T}_{n+1} + \mathrm{T}_{n-1}$ 换成 $2x\mathrm{T}_n$。当不变量为零时，相邻两项之比全部相等，所以数列是等比数列（geometric sequence）。再把等比数列代入递推关系，便可确定公比（common ratio）是某个二次方程（quadratic）的根。

解答

以下 x 始终固定，并把 $\mathrm{T}_n(x)$ 记作 T_n，把 $\mathrm{f}(x)$ 记作 f。递推关系 $(*)$ 即是

$$\mathrm{T}_{n+1} + \mathrm{T}_{n-1} = 2x\,\mathrm{T}_n \qquad (n \geq 1).$$

不变量

对 $n \geq 1$，设 $D_n = \mathrm{T}_n^2 - \mathrm{T}_{n-1}\mathrm{T}_{n+1}$。我们用数学归纳法（induction）证明对所有 $n \geq 1$ 都有 $D_n = f$。

基本情况。$D_1 = \mathrm{T}_1^2 - \mathrm{T}_0\mathrm{T}_2$，按定义这就是 f。

归纳步骤。假设对某个 $k \geq 1$ 有 $D_k = f$。则

$$\begin{aligned} D_{k+1} - D_k &= \mathrm{T}_{k+1}^2 - \mathrm{T}_k\mathrm{T}_{k+2} - \mathrm{T}_k^2 + \mathrm{T}_{k-1}\mathrm{T}_{k+1} \\ &= \mathrm{T}_{k+1}\big(\mathrm{T}_{k+1} + \mathrm{T}_{k-1}\big) - \mathrm{T}_k\big(\mathrm{T}_{k+2} + \mathrm{T}_k\big) \\ &= \mathrm{T}_{k+1} \cdot 2x\mathrm{T}_k - \mathrm{T}_k \cdot 2x\mathrm{T}_{k+1} \\ &= 0, \end{aligned}$$

其中第三行对第一个括号用了 $(*)$ 取 $n=k$，对第二个括号则取 $n=k+1$（两者皆可，因为 $k\geq 1$）。因此 $D_{k+1}=D_k=f$，由归纳法可知

$$\mathrm{T}_n^2-\mathrm{T}_{n-1}\mathrm{T}_{n+1}=f\quad \text{对所有}n\geq 1,$$

这正是所要证明的。

$f\equiv 0$ 的情形

首先看看 $f=0$ 对 $r=\mathrm{T}_1/\mathrm{T}_0$ 有甚么含意。由 $(*)$ 取 $n=1$，得 $\mathrm{T}_2=2x\mathrm{T}_1-\mathrm{T}_0$，所以

$$f=\mathrm{T}_1^2-\mathrm{T}_0(2x\mathrm{T}_1-\mathrm{T}_0)=\mathrm{T}_0^2\big(r^2-2xr+1\big),$$

这里用了 $\mathrm{T}_1=r\mathrm{T}_0$。于是，由于 $\mathrm{T}_0\neq 0$，

$$f=0\iff r^2=2xr-1. \tag{†}$$

> **想一想** 不变量告诉我们 $\mathrm{T}_n^2=\mathrm{T}_{n-1}\mathrm{T}_{n+1}$，所以各个比值相等。为甚么不直接相除?

要除以 T_{n-1}，就必须先确定它不是零，这需要额外的论证。较简洁的做法是从“比值相等”这个想法猜出答案，然后利用 (†)，直接以递推关系本身来证明。

断言。对所有 $n\geq 0$，$\mathrm{T}_n=\mathrm{T}_0\,r^n$。

证明。我们使用强归纳法（strong induction），每次用上相邻的两个情况。断言在 $n=0$ 时显然成立；在 $n=1$ 时亦成立，因为按 r 的定义有 $\mathrm{T}_1=r\mathrm{T}_0$。假设断言对 $n=k-1$ 和 $n=k$ 成立，其中 $k\geq 1$。则由 $(*)$，

$$\begin{aligned}\mathrm{T}_{k+1}&=2x\mathrm{T}_k-\mathrm{T}_{k-1}=2x\mathrm{T}_0r^k-\mathrm{T}_0r^{k-1}\\&=\mathrm{T}_0r^{k-1}\big(2xr-1\big)=\mathrm{T}_0r^{k-1}\cdot r^2=\mathrm{T}_0r^{k+1},\end{aligned}$$

其中第四步用了 (†)。所以断言对 $n=k+1$ 成立，由归纳法可知它对所有 $n\geq 0$ 成立：

$$\boxed{\mathrm{T}_n(x)=\mathrm{T}_0(x)\,\big(\mathrm{r}(x)\big)^n}$$

$\mathrm{r}(x)$ 的可能形式

由 (†)，r 满足 $r^2-2xr+1=0$，所以由二次公式（quadratic formula）得

$$\mathrm{r}(x)=x+\sqrt{x^2-1}\qquad 或\qquad \mathrm{r}(x)=x-\sqrt{x^2-1}.$$

反过来，无论取哪一个都满足 (†)，所以 $f\equiv 0$，而数列 T_0r^n 确实符合 $(*)$。两根之积为 1，所以它们互为倒数。当 $|x|<1$ 时两根是复数，即 $x\pm\mathrm{i}\sqrt{1-x^2}$；设 $x=\cos\theta$，它们便成为 $\mathrm{e}^{\pm\mathrm{i}\theta}$，而 $\mathrm{T}_n=\mathrm{T}_0\mathrm{e}^{\pm\mathrm{i}n\theta}$。

验算

- 切比雪夫多项式（Chebyshev polynomials）$\mathrm{T}_0 = 1$、$\mathrm{T}_1 = x$、$\mathrm{T}_2 = 2x^2 - 1$ 给出 $f = x^2 - (2x^2 - 1) = 1 - x^2$，而事实上对每个 n 都有 $\cos^2 n\theta - \cos(n-1)\theta\cos(n+1)\theta = \sin^2\theta$。取 $x = \frac{3}{7}$，D_n 的首十四个值全部等于 $\frac{40}{49}$。
- 取 $r = 2$, 则 $x = \frac{1}{2}(2 + \frac{1}{2}) = \frac{5}{4}$, 并取 $\mathrm{T}_0 = 1$、$\mathrm{T}_1 = 2$。由递推关系得 $\mathrm{T}_2 = \frac{5}{2}\cdot 2 - 1 = 4$ 及 $\mathrm{T}_3 = \frac{5}{2}\cdot 4 - 2 = 8$：正如所料，都是 2 的幂。此外，$x - \sqrt{x^2-1} = \frac{5}{4} - \frac{3}{4} = \frac{1}{2}$ 是另一个根。

常见错误

- 把 $(*)$ 用于 $n = 0$：递推关系由 $n = 1$ 开始，所以基本情况必须是 D_1，而不是 D_0（后者根本没有定义）。
- 把 $\mathrm{T}_n^2 = \mathrm{T}_{n-1}\mathrm{T}_{n+1}$ 除以 T_{n-1}，却没有检查它是否非零。

STEP 之外

常系数线性递推关系的解具有 $\alpha r_1^n + \beta r_2^n$ 的形式，其中 r_1, r_2 是其特征方程（characteristic equation）的根，在这里特征方程是 $r^2 - 2xr + 1 = 0$。条件 $f = 0$ 所挑出的，恰好是只剩下一个等比项的情况；一般而言，f 是 $\alpha\beta$ 的倍数。不变量 D_n 可说是二阶微分方程中朗斯基行列式（Wronskian）的离散版本，而对于 $y'' + q(x)y = 0$，朗斯基行列式是常数。Boyce 与 DiPrima 的 *Elementary Differential Equations* 详细阐述了朗斯基行列式，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多 STEP 风格的递推关系题目。

详细解答

STEP III 2008，第 6 题

纯数学 · 一阶微分方程

STEP III 2008，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/08-s3-q6

关键思路

每条方程都直接以 x 和斜率 p 表示 y，但并未解出 p，所以不能直接积分。对 x 求导可以消去 y（因为 $\mathrm{d}y/\mathrm{d}x = p$），余下一条关于 x、p 和 $\mathrm{d}p/\mathrm{d}x$ 的关系式。把这条关系式上下倒转，便得到一条以 x 为 p 的函数的线性一阶微分方程（linear first-order differential equation），可用积分因子（integrating factor）求解。然后把 $x(p)$ 代回原方程，便可求得 y。

解答

第 (i) 部分

把 $y = p^2 + 2xp$ 对 x 求导，对 p^2 用链式法则（chain rule），对 $2xp$ 用积法则（product rule），并以 p 取代 $\mathrm{d}y/\mathrm{d}x$：

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

? **想一想** 我们可以把 $\mathrm{d}p/\mathrm{d}x$ 倒转而得出 $\mathrm{d}x/\mathrm{d}p$ 吗?

我们需要 $\mathrm{d}p/\mathrm{d}x \neq 0$。上面的方程显示，凡是 $p \neq 0$ 之处，$\mathrm{d}p/\mathrm{d}x$ 和 $2p + 2x$ 都不等于零。（反正我们要除以 p，所以整个过程都需要 $p \neq 0$。）因此可以写成

$$\frac{\mathrm{d}x}{\mathrm{d}p} = -\frac{2p + 2x}{p} = -2 - \frac{2x}{p},$$

正是题目所求。

整理后，这是一条对 x 为线性的方程：

$$\frac{\mathrm{d}x}{\mathrm{d}p}+\frac{2}{p}x=-2.$$

积分因子是 $\mathrm{e}^{\int 2/p\,\mathrm{d}p}=p^2$，全式乘以它，得

$$\frac{\mathrm{d}}{\mathrm{d}p}(p^2x)=-2p^2 \quad\Longrightarrow\quad p^2x=-\tfrac{2}{3}p^3+A \quad\Longrightarrow\quad x=-\tfrac{2}{3}p+Ap^{-2},$$

其中 A 为任意常数。这就是题目给出的结果。

现在利用当 $x=2$ 时 $p=-3$：

$$2=-\tfrac{2}{3}(-3)+\frac{A}{9}=2+\frac{A}{9},$$

所以 $A=0$，而 $x=-\frac{2}{3}p$，即 $p=-\frac{3}{2}x$。代入原方程，

$$y=p^2+2xp=\tfrac{9}{4}x^2+2x\left(-\tfrac{3}{2}x\right)=\tfrac{9}{4}x^2-3x^2,$$

所以

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

验算一下：$\mathrm{d}y/\mathrm{d}x=-\frac{3}{2}x$，确实就是 p，而且在 $x=2$ 时等于 -3。

第 (ii) 部分

这里出现了 $\ln p$, 所以 $p>0$。把 $y=2xp+p\ln p$ 对 x 求导, 并利用 $\frac{\mathrm{d}}{\mathrm{d}p}(p\ln p)=\ln p+1$：

$$p=2p+2x\frac{\mathrm{d}p}{\mathrm{d}x}+(\ln p+1)\frac{\mathrm{d}p}{\mathrm{d}x} \quad\Longrightarrow\quad -p=(2x+\ln p+1)\frac{\mathrm{d}p}{\mathrm{d}x}.$$

与之前一样，$p\neq 0$ 令 $\mathrm{d}p/\mathrm{d}x\neq 0$，所以

$$\frac{\mathrm{d}x}{\mathrm{d}p}=-\frac{2x+\ln p+1}{p} \quad\Longrightarrow\quad \frac{\mathrm{d}x}{\mathrm{d}p}+\frac{2}{p}x=-\frac{\ln p+1}{p}.$$

积分因子同样是 p^2：

$$\frac{\mathrm{d}}{\mathrm{d}p}(p^2x)=-p\ln p-p.$$

用分部积分法（integration by parts），取 $u=\ln p$，$\mathrm{d}v=p\,\mathrm{d}p$，

$$\int p\ln p\,\mathrm{d}p=\tfrac{1}{2}p^2\ln p-\int\tfrac{1}{2}p\,\mathrm{d}p=\tfrac{1}{2}p^2\ln p-\tfrac{1}{4}p^2,$$

所以

$$\begin{aligned}p^2x&=-\tfrac{1}{2}p^2\ln p+\tfrac{1}{4}p^2-\tfrac{1}{2}p^2+B\\&=-\tfrac{1}{2}p^2\ln p-\tfrac{1}{4}p^2+B,\end{aligned}$$

而

$$x = -\tfrac{1}{2}\ln p - \tfrac{1}{4} + Bp^{-2}.$$

利用当 $x = -\frac{1}{4}$ 时 $p = 1$：$-\frac{1}{4} = 0 - \frac{1}{4} + B$，所以 $B = 0$，而

$$x = -\tfrac{1}{2}\ln p - \tfrac{1}{4},$$

这正是题目要求我们证明的结果。

要求出 y，留意上式即 $\ln p = -2x - \frac{1}{2}$。把它代入原方程，比起把每一个 p 都代换掉要快：

$$y = 2xp + p\left(-2x - \tfrac{1}{2}\right) = -\tfrac{1}{2}p,$$

而 $p = \mathrm{e}^{-2x-1/2}$，所以

$$y = -\tfrac{1}{2}\mathrm{e}^{-2x-\frac{1}{2}}.$$

验算：$\mathrm{d}y/\mathrm{d}x = \mathrm{e}^{-2x-\frac{1}{2}} = p$，理应如此；而在 $x = -\frac{1}{4}$ 时，其值为 $\mathrm{e}^0 = 1$。

验算

在每一部分，把最后求得的 y 求导，确认导数确实就是你所用的 p：$y = -\frac{3}{4}x^2$ 给出 $p = -\frac{3}{2}x$，而 $p^2 + 2xp = \frac{9}{4}x^2 - 3x^2 = y$。至于第 (ii) 部分，在 $x = 0$ 时，$p = \mathrm{e}^{-1/2} \approx 0.6065$，而 $2xp + p\ln p = 0.6065 \times (-0.5) \approx -0.3033 = y(0)$。数值验算亦确认了，对随机选取的 A 和 B 值，一般的参数曲线 $(x(p), y(p))$ 的斜率 $\mathrm{d}y/\mathrm{d}x$ 都等于 p。

常见错误

- 忘记左边的 $\mathrm{d}y/\mathrm{d}x$ 就是 p，因而看不出相消之后余下 $-p$。
- 未用初始条件（initial condition）便把积分常数丢掉。这里的常数碰巧等于零，但你要把它求出来之后才知道。
- 在第 (ii) 部分，把 $p\ln p$ 对 x 求导时只写成 $\ln p$，遗漏了因子 $\mathrm{d}p/\mathrm{d}x$ 和额外的 $+1$。

STEP 之外

形如 $y = xf(p) + g(p)$ 的方程称为达朗贝尔（d'Alembert）方程，又称拉格朗日（Lagrange）方程；其特殊情况 $y = xp + g(p)$ 就是克莱罗（Clairaut）方程。把 p 当作自变量（independent variable）的技巧对这些方程全部适用。它们还可能有奇异解（singular solutions），出现在我们曾除以 $\mathrm{d}p/\mathrm{d}x$ 的地方，是通解（general solution）族的包络线（envelope）。Boyce 与 DiPrima 所著的 *Elementary Differential Equations*（初等微分方程）涵盖一阶方程及其解法。

详细解答

STEP III 2008，第 7 题

纯数学 · 复数与几何

STEP III 2008，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/08-s3-q7

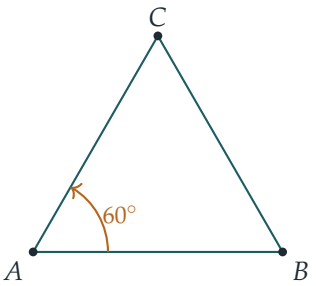
关键思路

在阿根图(Argand diagram)上,乘以 $e^{i\pi/3}$ 相当于逆时针旋转 60°。等边三角形(equilateral triangle)正好就是把一条边绕着一个顶点旋转 60° 而成,所以开首的关系式把“等边”这个条件化成了 a、b、c 的线性方程。之后两部分都只是按部就班的整理功夫：把每个新的点用原有的点表示出来，再验证我们想要的条件——若要证明是平行四边形（parallelogram），就验证 $p+r=q+s$；若要证明是等边三角形，就验证开首的关系式本身。

解答

开首的关系式

设 $\omega = e^{i\pi/3} = \frac{1}{2} + \frac{\sqrt{3}}{2}i$。由于 ABC 是等边三角形，AC 与 AB 等长，并与它成 60° 角。因为三角形的顶点是按逆时针方向排列的，所以由 $\overrightarrow{AB}$ 转到 $\overrightarrow{AC}$ 是逆时针方向。因此 $\overrightarrow{AC}$ 就是 $\overrightarrow{AB}$ 旋转 $+60°$ 所得：



$$c-a=\omega(b-a)=\tfrac{1}{2}(b-a)+\tfrac{\sqrt{3}}{2}\mathrm{i}(b-a),$$
$$2c=2a+(b-a)+\mathrm{i}\sqrt{3}(b-a)=(a+b)+\mathrm{i}\sqrt{3}(b-a),$$

这就是所要求的结果。

若三角形的顶点按顺时针方向排列，$\overrightarrow{AC}$ 就是 $\overrightarrow{AB}$ 旋转 -60° 所得，所以 $c-a=\bar{\omega}(b-a)$，其中 $\bar{\omega}=\frac{1}{2}-\frac{\sqrt{3}}{2}\mathrm{i}$。同样的代数运算给出

$$2c=(a+b)-\mathrm{i}\sqrt{3}(b-a).$$

反过来说，任何满足其中一条关系式的 c 都会令 ABC 成为等边三角形，因为每条关系式都等价于 $c-a=\omega^{\pm1}(b-a)$。所以要证明一个三角形是等边三角形，只需验证其中一条关系式便可。

第 (i) 部分

四个顶点由我们自行选择，而每一个都可以放在所属边的任何一侧。利用开首的结果，并让这些顶点担当 c 的角色，取

$$2p=(d+e)+\mathrm{i}\sqrt{3}(e-d) \qquad (D,E,P \text{ 逆时针}),$$
$$2q=(e+f)-\mathrm{i}\sqrt{3}(f-e) \qquad (E,F,Q \text{ 顺时针}),$$
$$2r=(f+g)+\mathrm{i}\sqrt{3}(g-f) \qquad (F,G,R \text{ 逆时针}),$$
$$2s=(g+d)-\mathrm{i}\sqrt{3}(d-g) \qquad (G,D,S \text{ 顺时针}).$$

根据开首的结果，PDE、QEF、RFG、SGD 于是全都是等边三角形。

> **想一想** 为甚么要让方向交替变换，而不是把每个三角形都放在所属边的同一侧？

不妨试试看。$PQRS$ 是平行四边形，当且仅当 $\overrightarrow{PQ}=\overrightarrow{SR}$，即 $q-p=r-s$，亦即 $p+r=q+s$。（这表示对角线 PR 与 QS 有相同的中点。）按上面的选择，

$$2(p+r)=(d+e+f+g)+\mathrm{i}\sqrt{3}\,(e-d+g-f),$$
$$2(q+s)=(d+e+f+g)-\mathrm{i}\sqrt{3}\,(f-e+d-g)=(d+e+f+g)+\mathrm{i}\sqrt{3}\,(e-d+g-f).$$

两者相等，所以 $p+r=q+s$，$PQRS$ 是平行四边形。假如我们全部都用“+”号，虚部就会是 $\mathrm{i}\sqrt{3}(e-d+g-f)$ 和 $\mathrm{i}\sqrt{3}(f-e+d-g)$。这两者只在 $d+f=e+g$ 时才相等，也就是只在 $DEFG$ 本身已经是平行四边形时才相等。正是正负号交替，才令含 $\sqrt{3}$ 的项对任何四边形都能吻合。

（对某些四边形来说，这四点可能刚好共线，得出一个退化的平行四边形。在这种情况下，条件 $\overrightarrow{PQ}=\overrightarrow{SR}$ 仍然成立。）

第 (ii) 部分

若 LMN 的顶点按顺时针方向排列, 可把整个图形对实轴作反射（即取共轭复数（complex conjugate））。反射保持等边三角形和形心（centroid）不变，也保持边上的三角形是否向外作这一性质，所以只需处理 LMN 按逆时针方向排列的情况。

这时 N 位于有向边 $L \to M$ 的左方，而 LM 上向外作的三角形 LMX 的顶点 X 则在右方，所以 L, M, X 是按顺时针方向排列的：

$$2x = (l+m) - \mathrm{i}\sqrt{3}(m-l).$$

LMX 的形心是

$$u = \tfrac{1}{3}(l+m+x) = \tfrac{1}{3}(l+m) + \tfrac{1}{6}(l+m) - \tfrac{\mathrm{i}\sqrt{3}}{6}(m-l) = \frac{l+m}{2} + \frac{\mathrm{i}(l-m)}{2\sqrt{3}},$$

其中用到 $\frac{\sqrt{3}}{6} = \frac{1}{2\sqrt{3}}$。作轮换 $l \to m \to n \to l$（这样会保持逆时针的次序），得

$$v = \frac{m+n}{2} + \frac{\mathrm{i}(m-n)}{2\sqrt{3}}, \qquad w = \frac{n+l}{2} + \frac{\mathrm{i}(n-l)}{2\sqrt{3}}.$$

我们现在证明 U, V, W 满足逆时针的关系式 $2w = (u+v) + \mathrm{i}\sqrt{3}(v-u)$。首先，

$$u+v = \tfrac{1}{2}l + m + \tfrac{1}{2}n + \frac{\mathrm{i}(l-n)}{2\sqrt{3}},$$
$$v-u = \tfrac{1}{2}(n-l) + \frac{\mathrm{i}(2m-l-n)}{2\sqrt{3}},$$
$$\mathrm{i}\sqrt{3}(v-u) = \frac{\mathrm{i}\sqrt{3}}{2}(n-l) - \tfrac{1}{2}(2m-l-n).$$

想一想 m 是怎样消失的?

把第一行和第三行相加。实部给出 $\frac{1}{2}l + m + \frac{1}{2}n - m + \frac{1}{2}l + \frac{1}{2}n = l + n$，所以 m 在实部中互相抵消，而它从来没有在虚部中出现过。虚部合并成

$$(n-l)\Big(\frac{\sqrt{3}}{2} - \frac{1}{2\sqrt{3}}\Big)\mathrm{i} = (n-l)\,\frac{3-1}{2\sqrt{3}}\,\mathrm{i} = \frac{\mathrm{i}(n-l)}{\sqrt{3}}.$$

因此

$$(u+v) + \mathrm{i}\sqrt{3}(v-u) = (n+l) + \frac{\mathrm{i}(n-l)}{\sqrt{3}} = 2w.$$

根据开首的结果（及其逆命题），UVW 是一个顶点按逆时针方向排列的等边三角形。这就是拿破仑定理（Napoleon's theorem）。

验算

- 开首的关系式：$a = 0$，$b = 2$ 给出 $c = 1 + \mathrm{i}\sqrt{3}$，顶点在实轴上方，与逆时针排列的三角形应有的情况相符。
- 第 (ii) 部分，取 $l = 0$，$m = 6$，$n = 6\mathrm{i}$：由公式得 $u = 3 - \mathrm{i}\sqrt{3}$，$v = (3 + \sqrt{3})(1 + \mathrm{i})$ 和 $w = -\sqrt{3} + 3\mathrm{i}$。于是 UVW 三条边的长度平方都是 $24 + 12\sqrt{3}$。
- 我们用 10^4 个随机构形，以数值方法检验了开首的关系式、平行四边形条件和拿破仑定理。我们亦确认了，在第 (i) 部分若每条边都用同一方向，一般而言并不成立。

常见错误

- 在第 (ii) 部分弄错向外的方向。对于按逆时针方向排列的 LMN，LM 上向外的顶点相对于 L, M 是顺时针方向的。用错了正负号，得到的就会是内拿破仑三角形。（它同样是等边三角形，但并不是题目所问的。）
- 在第 (i) 部分，试图令四个三角形全部指向同一方。这只在 $DEFG$ 本身是平行四边形时才行得通。
- 用 $p + q = r + s$ 来检验平行四边形。应该是相对的顶点配成一对：条件是 $p + r = q + s$。

STEP 之外

背后的想法是：映射 $z \mapsto \alpha z + \beta$ 正好就是平面上的旋转、放大和平移，所以几何条件会变成 $\mathbb{C}$ 上的线性方程。开首的关系式有一个简洁的改写方式：ABC 是等边三角形，当且仅当 $a + \omega^2 b + \omega^4 c = 0$ 或 $a + \omega^4 b + \omega^2 c = 0$，而拿破仑定理只需几行代数运算便随之而得。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）把复数乘法的这种几何观点发挥得淋漓尽致。

详细解答

STEP III 2008，第 8 题

纯数学 · 级数与递推关系

STEP III 2008，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/08-s3-q8

关键思路

系数满足线性递推关系（linear recurrence），意味着只要把级数乘以适当的多项式，几乎所有项都会互相抵消。这个多项式可以直接从递推关系读出来：$a_{r+1}+pa_r=0$ 提示我们用 $1+px$，而 $a_{r+2}+pa_{r+1}+qa_r=0$ 则提示我们用 $1+px+qx^2$。余下的只是一个很短的多项式，所以级数之和是一个有理函数（rational function）。对于有限项之和，同样的技巧会在末端留下几个额外的项。另一个做法是用部分分式（partial fractions）把级数拆成几个等比级数（geometric series），而等比级数的和我们是懂得求的。

解答

第 (i) 部分

这里级数由 x 项开始，所以 $a_1=\frac{1}{3}$，$a_2=\frac{1}{6}$，$a_3=\frac{1}{12}$。每个系数都是前一个的一半，所以 $a_{r+1}-\frac{1}{2}a_r=0$，即

$$p=-\tfrac{1}{2}.$$

（这里没有常数项。递推关系只把 $a_1,a_2,a_3,\ldots$ 联系起来，对 $r=0$ 的情况并无任何说明。）解这个递推关系，得 $a_r=\frac{1}{3}\left(\frac{1}{2}\right)^{r-1}$。

无穷项之和（sum to infinity）。 乘以 $1-\frac{1}{2}x$：

$$\left(1-\tfrac{1}{2}x\right)S=a_1x+\sum_{r\geq 1}\left(a_{r+1}-\tfrac{1}{2}a_r\right)x^{r+1}=\tfrac{1}{3}x,$$

因为根据递推关系，每个括号都等于零。因此

$$S = \frac{\frac{1}{3}x}{1-\frac{1}{2}x} = \frac{2x}{3(2-x)}.$$

这是一个公比为 $\frac{1}{2}x$ 的等比级数，所以无穷项之和恰好在 $|x|<2$ 时存在。

首 $n+1$ 项。 这些就是 $x, x^2, \ldots, x^{n+1}$ 各项。设 $S_{n+1} = \sum_{r=1}^{n+1} a_r x^r$。重复同样的乘法，中间各项仍然全部抵消，但最高次的一项 $-\frac{1}{2}a_{n+1}x^{n+2}$ 却没有对应的项与它抵消：

$$\left(1-\tfrac{1}{2}x\right)S_{n+1} = \tfrac{1}{3}x - \tfrac{1}{2}a_{n+1}x^{n+2} = \tfrac{1}{3}x - \tfrac{1}{3}\left(\tfrac{1}{2}\right)^{n+1}x^{n+2} = \tfrac{1}{3}x\left(1-\left(\tfrac{x}{2}\right)^{n+1}\right).$$

所以，当 $x \neq 2$ 时，

$$S_{n+1} = \frac{2x}{3(2-x)}\left(1-\left(\frac{x}{2}\right)^{n+1}\right).$$

当 $x=2$ 时，每一项都等于 $\frac{2}{3}$，所以 $S_{n+1} = \frac{2}{3}(n+1)$。

第 (ii) 部分

现在 $a_0=2$，$a_1=8$，$a_2=18$，$a_3=37$。把 $r=0$ 和 $r=1$ 代入递推关系，得到两条方程：

$$18+8p+2q=0, \qquad 37+18p+8q=0.$$

由第一条方程得 $q=-9-4p$。代入第二条方程，得 $37+18p-72-32p=0$，所以 $p=-\frac{5}{2}$，从而 $q=1$。递推关系是

$$a_{r+2} - \tfrac{5}{2}a_{r+1} + a_r = 0.$$

无穷项之和。 乘以 $1+px+qx^2 = 1-\frac{5}{2}x+x^2$。对每个 $r \geq 0$，乘积中 x^{r+2} 的系数是 $a_{r+2}-\frac{5}{2}a_{r+1}+a_r = 0$。只有 x^0 和 x^1 的系数留下来：

$$\left(1-\tfrac{5}{2}x+x^2\right)T = a_0 + \left(a_1 - \tfrac{5}{2}a_0\right)x = 2+3x.$$

把分子和分母同时加倍，并把 $2-5x+2x^2 = (1-2x)(2-x)$ 因式分解，得

$$T = \frac{4+6x}{(1-2x)(2-x)}.$$

部分分式。 设 $\dfrac{4+6x}{(1-2x)(2-x)} = \dfrac{A}{1-2x} + \dfrac{B}{2-x}$。用遮盖法（cover-up），在 $x=\frac{1}{2}$ 处得 $A = \dfrac{7}{3/2} = \frac{14}{3}$，在 $x=2$ 处得 $B = \dfrac{16}{-3} = -\frac{16}{3}$。所以

$$T = \frac{14}{3}\cdot\frac{1}{1-2x} - \frac{8}{3}\cdot\frac{1}{1-\frac{1}{2}x}.$$

把每一项展开成等比级数，得

$$a_r = \tfrac{14}{3}\,2^r - \tfrac{8}{3}\left(\tfrac{1}{2}\right)^r.$$

验算：$r = 0, 1, 2, 3$ 分别给出 2、$\frac{28}{3} - \frac{4}{3} = 8$、$\frac{56}{3} - \frac{2}{3} = 18$ 和 $\frac{112}{3} - \frac{1}{3} = 37$，与已知的值相符。

想一想 无穷项之和究竟对哪些 x 才存在?

两个等比级数都必须收敛（converge），所以需要 $|2x| < 1$ 和 $|\frac{1}{2}x| < 1$。第一个条件较为严格，所以我们需要 $|x| < \frac{1}{2}$。

首 $n+1$ 项。 这些就是由 x^0 到 x^n 的各项。把每个等比级数求和至 $n+1$ 项，当 $x \neq \frac{1}{2}, 2$ 时，得

$$\sum_{r=0}^{n} a_r x^r = \frac{14}{3}\cdot\frac{1-(2x)^{n+1}}{1-2x} - \frac{16}{3}\cdot\frac{1-\left(\frac{x}{2}\right)^{n+1}}{2-x}.$$

（第二个分式就是 $\frac{8}{3}\cdot\frac{1-(x/2)^{n+1}}{1-x/2}$，只是分子和分母都乘以 2。）若 $x = \frac{1}{2}$ 或 $x = 2$，对应的等比级数之和就只是 $n+1$ 个 1 相加。

验算

- 在 $|x| < \frac{1}{2}$ 的情况下令 $n \to \infty$：幂 $(2x)^{n+1}$ 和 $(x/2)^{n+1}$ 趋向零，于是我们重新得到 T 的部分分式形式。在第 (i) 部分，当 $|x| < 2$ 时情况相同。
- 取 $n = 0$，第 (i) 部分的公式给出 $\frac{2x}{3(2-x)}\cdot\frac{2-x}{2} = \frac{1}{3}x$。第 (ii) 部分的公式给出 $\frac{14}{3} - \frac{16}{3}\cdot\frac{1}{2} = 2$。两者都是首项，正如所料。
- 数值上，在 $x = 0.3$ 处有 $T = \dfrac{5.8}{0.4 \times 1.7} \approx 8.5294$，而 $2 + 8x + 18x^2 + 37x^3 + \cdots$ 的部分和（partial sums）亦趋近这个值。

常见错误

- 在第 (i) 部分，把首项当作 $a_0 x^0$。级数由 x^1 开始，所以首 $n+1$ 项一直去到 x^{n+1}。
- 把有限项之和乘以 $1 + px$ 时，忘记了末端剩下的项。正是这些项产生了 $(\frac{x}{2})^{n+1}$ 这个因子。
- 对 T 写出 $|x| < 2$。决定收敛半径（radius of convergence）的，是较接近原点的根 $x = \frac{1}{2}$。

STEP 之外

幂级数（power series）$\sum a_r x^r$ 是数列 (a_r) 的生成函数（generating function）。常系数线性递推关系必定给出有理的生成函数，而这个函数的极点（poles）既决定 a_r 的闭合形式（closed form），也决定收敛半径。生成函数是组合数学（combinatorics）和概率论中的标准工具，而幂级数的收敛半径则会在大学一年级的分析课程中严谨地建立起来。Abbott 的 *Understanding Analysis*（一本分析学入门教科书）对收敛这方面有很好的讲解。

详细解答

STEP III 2008，第 9 题

力学 · 摩擦力与变力

STEP III 2008，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：

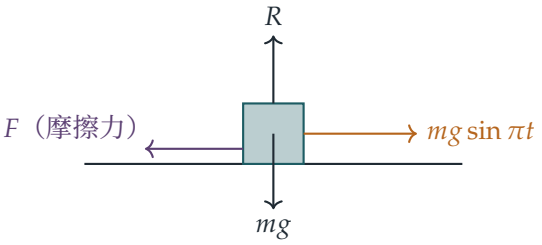
step.maths.org/questions/08-s3-q9

关键思路

摩擦力有两种状态。质点（particle）静止时，静摩擦力（static friction）恰好与外加的力相等，上限为 μmg，因此加速度为零。质点一旦滑动，摩擦力的大小固定为 μmg，方向与速度相反。所以加速度的图像由若干段平坦的零值线段和平移后的正弦曲线连接而成。“静止下来”的意思是：速度（即自运动开始起加速度图像下的面积）回到零。

解答

我们把物体看成粗糙水平面上的质点。法向反作用力（normal reaction）为 $R = mg$，所以静止时摩擦力的大小至多为 μmg，滑动时则恰好为 μmg。取 $\sin \pi t > 0$ 时外力的方向为正方向，并设 a 和 v 分别为加速度和速度。



开始运动

沿铅垂方向分解，得 $R = mg$。质点静止时，水平方向的平衡需要大小为 $mg|\sin \pi t|$ 的摩擦力，而只要 $mg|\sin \pi t| \leq \mu mg$，这是可以做到的。当外力首次超过极限摩擦力（limiting

friction）时，质点便首次开始运动，所以（当 $\mu < 1$ 时）

$$\sin \pi T_0 = \mu, \qquad \text{即} \qquad T_0 = \frac{1}{\pi} \arcsin \mu, \quad 0 \le T_0 < \tfrac{1}{2}.$$

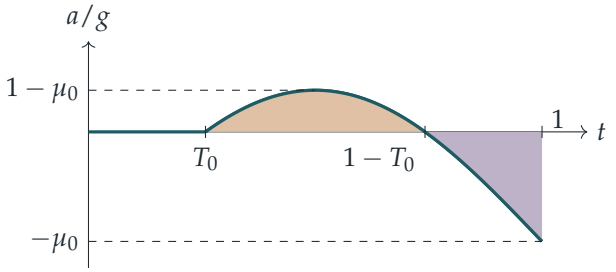
（若 $\mu \ge 1$，质点永远不会运动。）

第 (i) 部分

当 $0 \le t \le T_0$ 时质点静止，所以 $a = 0$。当 $t > T_0$ 时，质点沿正方向滑动，摩擦力 $\mu_0 mg$ 指向后方，由牛顿（Newton）第二定律得

$$a = g(\sin \pi t - \mu_0) \qquad (\text{只要} v > 0).$$

这个加速度在 $t = T_0$ 时由 0 开始，在 $t = \frac{1}{2}$ 时升至 $g(1 - \mu_0)$，在 $t = 1 - T_0$ 时回落至 0（由对称性 $\sin \pi(1 - t) = \sin \pi t$ 可知），此后一直为负，在 $t = 1$ 时达到 $-\mu_0 g$。



速度就是这个图像由 T_0 起计的面积：曲线在横轴之上时（琥珀色区域）速度增加，过了 $1 - T_0$ 之后（紫红色区域）速度减少。在两块面积互相抵销之前，速度一直为正，而抵销的一刻就是质点停下来的时刻。所以我们需要

$$v(1) = \int_{T_0}^{1} g(\sin \pi t - \mu_0)\, \mathrm{d}t = 0.$$

现在

$$\begin{aligned} \int_{T_0}^{1} (\sin \pi t - \mu_0)\, \mathrm{d}t &= \Big[-\frac{\cos \pi t}{\pi} \Big]_{T_0}^{1} - \mu_0 (1 - T_0) \\ &= \frac{1 + \cos \pi T_0}{\pi} - \mu_0 + \mu_0 T_0. \end{aligned}$$

令上式等于零，再乘以 π：

$$1 + \cos \pi T_0 - \mu_0 \pi + \mu_0 \pi T_0 = 0.$$

? **想一想** 用 μ_0 表示 $\cos \pi T_0$ 时，它应取哪个符号？

由于 $0 \le \pi T_0 < \frac{\pi}{2}$，余弦为正，所以 $\cos \pi T_0 = \sqrt{1 - \sin^2 \pi T_0} = (1 - \mu_0^2)^{1/2}$。另外 $\pi T_0 = \arcsin \mu_0$。把两者代入，得

$$1 + \left(1 - \mu_0^2\right)^{\frac{1}{2}} - \mu_0 \pi + \mu_0 \arcsin \mu_0 = 0,$$

这正是所要证明的。（数值上 $\mu_0 \approx 0.7246$，$T_0 \approx 0.258$。）

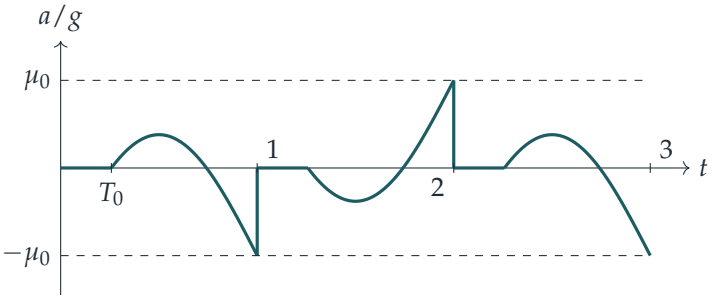
第 (ii) 部分

在 $t = 1$ 时，质点静止，外力为 $mg \sin \pi = 0$。当 $1 < t < 2$ 时，外力指向负方向，大小为 $mg|\sin \pi t|$。在 $|\sin \pi t| = \mu_0$ 之前，这个力小于 $\mu_0 mg$，而该等式首次成立是在 $t = 1 + T_0$（因为 $\sin \pi(1 + T_0) = -\sin \pi T_0$）。所以当 $1 \le t \le 1 + T_0$ 时，质点保持静止，$a = 0$。

此后质点沿负方向滑动，摩擦力指向正方向，于是

$$a = g(\sin \pi t + \mu_0).$$

由于 $\sin \pi(t + 1) = -\sin \pi t$，这恰好是第 (i) 部分中时刻 $t - 1$ 的加速度的相反数。因此 $[1, 2]$ 上的运动是 $[0, 1]$ 上运动的镜像：速度是先前速度的相反数，所以质点在 $t = 2$ 时再次静止，并且往回走了相同的距离。这时质点静止而外力为零，与 $t = 0$ 时的情况完全相同，所以在 $[2, 3]$ 上，$[0, 1]$ 的运动会重演一次。



在 $t = 1$ 和 $t = 2$ 处的铅垂线段表示加速度的跳跃：在质点停下的一瞬间，动摩擦力（kinetic friction，大小为 $\mu_0 mg$）被静摩擦力取代，而静摩擦力只需平衡大小为零的外力。

*$\mu = \mu_0$ 时的运动。*质点一直静止至 $t = T_0$，然后向前滑动，在 $t = 1$ 时停下，等到 $t = 1 + T_0$，再向后滑动，在 $t = 2$ 时停在出发点，如此类推。它在同一段地面上来回往返，周期为 2，在两端各停留 T_0 的时间。（再积分一次，可得这段路的长度为 $g\left[\mu_0/\pi^2 - (1 - T_0)(1 - \sqrt{1 - \mu_0^2})/(2\pi)\right] \approx 0.0367g$，不过题目并没有要求这个结果。）

*$\mu = 0$ 时的运动。*这时 $T_0 = 0$，而且在任何时刻都有 $a = g \sin \pi t$，所以

$$v = \frac{g}{\pi}(1 - \cos \pi t) \ge 0, \qquad x = \frac{g}{\pi}\left(t - \frac{\sin \pi t}{\pi}\right).$$

速度永不为负，并且只在 $t = 0, 2, 4, \ldots$ 时为零。所以质点永不回头：它一阵一阵地向前冲，在偶数时刻瞬间静止，每个长度为 2 的周期前进 $2g/\pi$，平均速率为 g/π。在粗糙的情况下，令质点在 $t = 1$ 停下来的正是摩擦力；没有摩擦力时，外力向后的那一半只会把质点减速至速度为零。

验算

用二分法（bisection）解第 (i) 部分的方程，得 $\mu_0 \approx 0.72461$，因而 $T_0 \approx 0.25798$。取 $g = 9.8$，对 $0 \le t \le 3$ 的运动作数值模拟，并在静摩擦力与动摩擦力之间切换，结果显示质点在 $t \approx 0.258$ 时起动，走了 0.3598 之后在 $t = 1.000$ 时停下，在 $t \approx 1.258$ 时开始往回走，并在 $t = 2.000$ 时停在出发点，与 $0.0367g$ 吻合。当 $\mu = 0$ 时，模拟结果与 $x = g(t - \sin \pi t/\pi)/\pi$ 一致。对方程作一个简单的合理性检查：在 $\mu_0 = 0$ 时左边为 $2 > 0$，在 $\mu_0 = 1$ 时左边为 $1 - \pi/2 < 0$，所以在 $(0, 1)$ 内有根。

常见错误

- 由 $t = 0$ 起就取 $a = g(\sin \pi t - \mu)$：在 T_0 之前质点静止，摩擦力是静摩擦力，所以 $a = 0$。
- 在第 (ii) 部分中，质点反向之后仍让摩擦力保持原来的方向。摩擦力总是与速度方向相反。
- 以为质点在 $t = 1 - T_0$ 时停下，但那时为零的是加速度，而不是速度。

STEP 之外

摩擦力在“黏住”与“滑动”之间切换，令运动方程变成分段（piecewise）的形式，加速度亦变得不连续。这就是门铰吱吱作响和小提琴弓拉琴弦背后的黏滑（stick–slip）现象，也是非光滑动力系统（non-smooth dynamical system）的一个入门例子。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）以合适的程度讨论了摩擦力模型和受迫运动。

详细解答

STEP III 2008，第 10 题

力学 · 弹性绳与能量

STEP III 2008，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/08-s3-q10

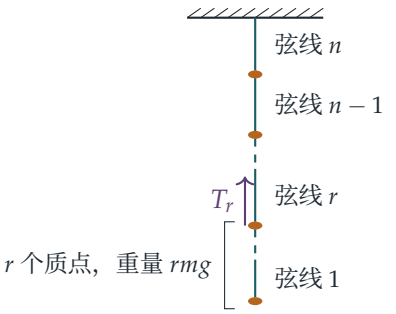
关键思路

每条短弦线都要承托悬挂在它下方的一切，所以它的张力（tension）很容易写出；再由胡克定律（Hooke's law）得出它的伸长量（extension）和所储存的能量，而总和就是 r 与 r^2 的标准求和。绳子的情况其实是同一幅图画，只是质量分散开来：把它切成 n 等份，把每一份的质量当作位于其下端的一个质点（particle），再让 $n \to \infty$。求和变成积分，而长度中的因子 $\frac{1}{2}$ 之所以出现，是因为平均而言，绳子的每一小段承托著绳子一半的重量。

解答

挂著质点的弦线

把各条短弦线视为遵从胡克定律的轻弹性弦线（light elastic string），并把各个质量视为质点。按题目所述，由底部起为弦线编号，于是弦线 r 下方有 r 个质点：它自己的质点，以及弦线 $1, \ldots, r-1$ 的质点。



整个系统处于静止状态。考虑弦线 r 顶端以下的一切：弦线 r、弦线 $1, \ldots, r-1$，以及挂

在它们上面的 r 个质点。作用于这部分的外力只有向下的重量 rmg 和弦线 r 中向上的张力 T_r（弦线是轻的）。所以

$$T_r = rmg.$$

根据胡克定律，弦线 r 的伸长量为

$$x_r = \frac{T_r\,\ell}{\lambda} = \frac{rmg\ell}{\lambda}.$$

所以长弦线的长度为

$$\sum_{r=1}^{n}(\ell + x_r) = n\ell + \frac{mg\ell}{\lambda}\sum_{r=1}^{n} r = n\ell + \frac{n(n+1)mg\ell}{2\lambda}.$$

自然长度（natural length）为 ℓ 的弦线伸长了 x 时，其弹性能（elastic energy）为 $\lambda x^2/(2\ell)$。对弦线 r 而言，这等于

$$\frac{\lambda}{2\ell}\left(\frac{rmg\ell}{\lambda}\right)^2 = \frac{r^2m^2g^2\ell}{2\lambda},$$

再利用 $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$，总弹性能为

$$\frac{m^2g^2\ell}{2\lambda}\cdot\frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)(2n+1)\,m^2g^2\ell}{12\lambda}.$$

有重量的绳子

> **想一想** 对于质量连续分布的绳子，第一部分能帮上甚么忙?

把绳子切成 n 等份。每一份的自然长度为 $\ell = L_0/n$，质量为 $m = M/n$，而且每一份的弹性模量（modulus）都与整条绳子相同，同为 λ（模量与小段的长度无关）。如果把每一份的质量集中在其下端，得到的正正就是上面那条长弦线。当 $n \to \infty$ 时，各小段愈来愈短，这个集中质量的模型便趋近真实的绳子。

代入长度的公式，

$$n\ell + \frac{n(n+1)mg\ell}{2\lambda} = L_0 + \frac{n(n+1)}{2\lambda}\cdot\frac{Mg}{n}\cdot\frac{L_0}{n} = L_0 + \frac{MgL_0}{2\lambda}\left(1+\frac{1}{n}\right) \longrightarrow L_0\left(1+\frac{Mg}{2\lambda}\right)$$

当 $n \to \infty$。所以

$$L = L_0\left(1+\frac{Mg}{2\lambda}\right),$$

正如所求。

你也可以直接计算。设 s 为某一点下方绳子的自然长度，则该点所承托的重量为 Mgs/L_0。在该处取自然长度为 δs 的一小段，其张力为 $T = Mgs/L_0$，所以它伸长了 $T\,\delta s/\lambda$。全部加起来，

$$L - L_0 = \int_0^{L_0}\frac{Mgs}{\lambda L_0}\,\mathrm{d}s = \frac{MgL_0}{2\lambda}.$$

至于能量，作同样的代入可得

$$\frac{n(n+1)(2n+1)}{12\lambda}\cdot\frac{M^2g^2}{n^2}\cdot\frac{L_0}{n}=\frac{M^2g^2L_0}{12\lambda}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)\longrightarrow\frac{M^2g^2L_0}{6\lambda}.$$

（直接计算：该小段储存的能量为 $T^2\,\delta s/(2\lambda)$，而 $\int_0^{L_0}\frac{M^2g^2s^2}{2\lambda L_0^2}\,\mathrm{d}s=\frac{M^2g^2L_0}{6\lambda}$。）

要把结果以 L、L_0 和 λ 表示，可利用长度的公式消去 Mg：

$$L-L_0=\frac{MgL_0}{2\lambda}\quad\Longrightarrow\quad Mg=\frac{2\lambda(L-L_0)}{L_0}.$$

因此绳子所储存的弹性能为

$$\frac{L_0}{6\lambda}\cdot\frac{4\lambda^2(L-L_0)^2}{L_0^2}=\frac{2\lambda(L-L_0)^2}{3L_0}.$$

验算

- 取 $n=1$ 时，第一部分给出长度 $\ell+mg\ell/\lambda$ 和能量 $m^2g^2\ell/(2\lambda)$：这正是一条弦线挂著一个质点的情况，理应如此。
- 试比较一条总伸长量同为 $e=L-L_0$、但张力均匀的绳子：它所储存的能量是 $\lambda e^2/(2L_0)$。我们的答案是这个值的 $\frac{4}{3}$ 倍，这是合理的，因为拉伸集中在顶部附近，而能量随局部应变（strain）的平方增长。
- 数值上，取 $M=3$、$L_0=2$、$\lambda=50$、$g=9.8$：公式给出 $L=2.588$ 和能量 5.7624；取 $n=100\,000$ 的离散链则给出 2.58801 和 5.76249，而 $2\lambda(L-L_0)^2/(3L_0)=5.7624$。

常见错误

- 把弦线 r 的张力当作 $(n-r+1)mg$：按题目的编号，弦线 n 在最顶，所以弦线 r 承托的是 r 个质点。
- 把每一小段绳子的模量取为 λ/n。模量是材料和横截面的性质，与长度无关；随长度改变的是劲度（stiffness）λ/ℓ。
- 题目要求以 L、L_0 和 λ 表示能量，却把答案留在以 M 和 g 表示的形式。

STEP 之外

由 n 个集中质量过渡到连续的绳子，是连续介质力学（continuum mechanics）最基本的一步：张力变成函数 $T(s)$，胡克定律变成局部应力（stress）与应变之间的关系，而能量变

成能量密度的积分。同样是这条由质量和弹簧组成的链，一旦让它振动起来，取极限后便得到弦线的波动方程（wave equation）。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）是了解这些想法如何展开的好去处。

详细解答

STEP III 2008，第 11 题

力学 · 转动动力学

STEP III 2008，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：

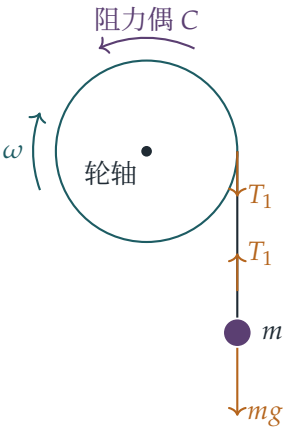
step.maths.org/questions/08-s3-q11

关键思路

力偶（couple）恒定，即角减速度（angular deceleration）恒定，所以绳子脱离之后，轮子服从转动版本的“suvat”方程，而第一个结果不过是“距离 = 平均速率 × 时间”。同一阶段亦让我们以 I 表示阻力偶的大小。至于下降阶段，能量方程是最自然的工具：我们知道各部分移动了多远，而克服恒定力偶所作的功就是力偶 × 角度。

解答

模型。轮子是绕固定光滑轮轴转动的刚体（rigid body），唯一例外是受到一个大小恒为 C 的阻力偶（resisting couple），它在整个运动过程中都起作用（质点下降期间如是，其后亦如是）。绳子是轻而不可伸长的，并且不打滑，所以绳子仍然连着时，质点的速率为 $v = r\omega$，其中 ω 是轮子的角速率（angular speed）。



绳子从轮子的右边离开，所以质点下落时轮子顺时针转动，而力偶 C 则逆时针作用，阻碍运动。（我们其实用不着绳子的张力（tension）T_1，因为我们用的是能量方法。）

绳子脱离之后

此时轮子所受的力偶只有 C，所以对轮轴取矩（taking moments），

$$I\dot{\omega} = -C,$$

而角减速度 C/I 是常数。因此角速率在时间 T 内由 ω_0 线性地下降至 0，所以

$$\omega_0 = \frac{C}{I}\,T, \qquad \text{即} \qquad C = \frac{I\omega_0}{T}. \tag{1}$$

角速率既然线性下降，平均角速率就是 $\frac{1}{2}(\omega_0 + 0)$，而转过的角度等于平均角速率乘以时间。这个角度是 n_2 转，即 $2\pi n_2$ 弧度（radians），所以

$$\tfrac{1}{2}\omega_0 T = 2\pi n_2,$$

这就是第一个结果。（它是 $s = \frac{1}{2}(u+v)t$ 的转动版本。）由此亦得

$$\omega_0 = \frac{4\pi n_2}{T}, \qquad \text{于是由 (1) 得} \qquad C = \frac{4\pi n_2 I}{T^2}. \tag{2}$$

质点下降期间

轮子转过 n_1 转（即 $2\pi n_1$ 弧度）期间，绳子松开的长度为 $2\pi n_1 r$，所以质点下落的距离为 $2\pi n_1 r$。在脱离的一刻，轮子的角速率为 ω_0，而由于绳子没有打滑，质点的速率为 $r\omega_0$。

> ? **想一想** 既然有阻力，为甚么这里仍可以用能量?

能量并不守恒，但功能原理（work–energy principle）依然成立：势能（potential energy）的损失等于动能（kinetic energy）的增加，再加上克服阻力所作的功。转过角度 θ 时，克服恒定力偶 C 所作的功为 $C\theta$，正如克服恒力所作的功是力 × 距离。张力对整个系统不作净功，因为绳子轻而不可伸长，并且不打滑，所以它对轮子所作的功恰好等于它从质点取走的功。

所以，由释放至脱离，

$$mg\,(2\pi n_1 r) = \tfrac{1}{2}m(r\omega_0)^2 + \tfrac{1}{2}I\omega_0^2 + C\,(2\pi n_1).$$

现在把 (2) 中的 ω_0 和 C 代入：

$$2\pi mgrn_1 = \tfrac{1}{2}\left(mr^2 + I\right)\frac{16\pi^2 n_2^2}{T^2} + 2\pi n_1 \cdot \frac{4\pi n_2 I}{T^2}.$$

全式乘以 $T^2/(2\pi)$，

$$\begin{aligned} mgrn_1T^2 &= 4\pi n_2^2\left(mr^2+I\right)+4\pi n_1n_2I \\ &= 4\pi mr^2n_2^2+4\pi n_2(n_1+n_2)\,I. \end{aligned}$$

移项，再除以 $4\pi n_2(n_1+n_2)$（此数为正），

$$I=\frac{mgrn_1T^2-4\pi mr^2n_2^2}{4\pi n_2(n_1+n_2)},$$

正是所求。

验算

- 量纲。分子的两项 $mgrT^2$ 和 mr^2 都具有转动惯量（moment of inertia）的单位（$\mathrm{kg\,m^2}$），而 n_1、n_2 是纯数。
- 数值验算。我们取 $I=0.5\,\mathrm{kg\,m^2}$、$m=1\,\mathrm{kg}$、$r=0.2\,\mathrm{m}$、$C=0.5\,\mathrm{N\,m}$、$g=9.8\,\mathrm{m\,s^{-2}}$ 和 $n_1=5$，以数值方法逐步计算这个运动。结果是 $\omega_0\approx 13.03\,\mathrm{rad\,s^{-1}}$、$T\approx 13.03\,\mathrm{s}$ 和 $n_2\approx 13.52$。于是 $\frac{1}{2}\omega_0T\approx 84.94\approx 2\pi n_2$，而公式给出 $I=0.5000$。
- 没有阻力。若 $C\to 0$，轮子永不停下，所以 T 和 n_2 都趋向无限大；公式的表现仍然合理，因为分子的增长如 T^2，而分母的增长如 $n_2^2\propto T^2$。

常见错误

- 漏掉下降期间克服阻力所作的功。题目说明力偶是恒定的，而且它自始至终都起作用，并非只在绳子脱离之后。少了这一项，公式就会出错：在上面的例子中，会得出 $I\approx 0.39$ 而不是 0.5。
- 忘记质点的动能 $\frac{1}{2}mr^2\omega_0^2$；$-4\pi mr^2n_2^2$ 这一项正是由此而来。
- 混淆转数与弧度：n 转等于 $2\pi n$ 弧度。

STEP 之外

这是一个巧妙的实验方法：只要量度轮子由转动至停下所需的时间，并数算转数，就可以同时测出它的转动惯量和轴承的摩擦，即使事前两者皆不知道。大学本科的力学会把这些内容发展成完整的刚体转动理论，其中角动量（angular momentum）和力矩（torque）分别扮演动量和力的角色，而对于绕任意轴旋转的物体，转动惯量则变成一个张量（tensor）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）对此有细致的阐述，并附有许多正是这种风格的习题。

详细解答

STEP III 2008，第 12 题

概率与统计 · 矩生成函数

STEP III 2008，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/08-s3-q12

本题在 *Beyond STEP* 第 12.24 节中心极限定理中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/clt.html

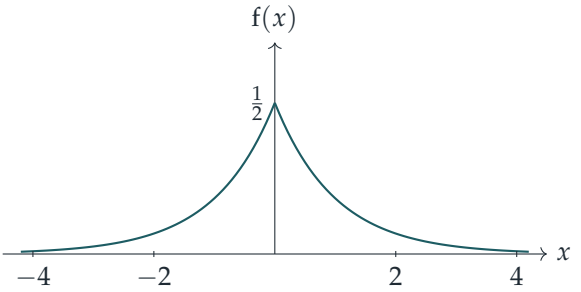
关键思路

密度函数中的绝对值（absolute value）把每个积分分成 $x < 0$ 和 $x > 0$ 两半，每一半都是普通的指数积分。此后整道题只靠矩母函数（moment generating function，简称 m.g.f.）的两个性质：独立（independent）变量之和的矩母函数，等于各自矩母函数之积；以及 $\mathrm{M}_{aY}(\theta) = \mathrm{M}_Y(a\theta)$。取对数后，$n$ 次方就变成 n 乘以一个很小的量，这正是麦克劳林级数（Maclaurin series）最擅长处理的情况。这就是中心极限定理（central limit theorem），而且在这个例子中，每一步都可以亲手算出来。

解答

略图

密度函数是偶函数，在 $x = 0$ 处取得最大值 $\frac{1}{2}$，该处有一个尖角（左、右导数分别为 $+\frac{1}{2}$ 和 $-\frac{1}{2}$），并在两侧以指数形式递减至 0。



矩母函数

根据定义，$M_X(\theta) = E(e^{\theta X})$。在 0 处把积分分开，

$$\begin{aligned} M_X(\theta) &= \tfrac{1}{2}\int_{-\infty}^{0} e^{(1+\theta)x}\,dx + \tfrac{1}{2}\int_{0}^{\infty} e^{-(1-\theta)x}\,dx \\ &= \frac{1}{2(1+\theta)} + \frac{1}{2(1-\theta)} = \frac{(1-\theta)+(1+\theta)}{2(1-\theta^2)} = \frac{1}{1-\theta^2}. \end{aligned}$$

第一个积分只在 $1+\theta > 0$ 时收敛（converge），第二个只在 $1-\theta > 0$ 时收敛，所以上式对 $-1 < \theta < 1$ 成立，而我们只需要 $\theta = 0$ 附近的情况，这已经足够。这就是所求的结果。

至于方差（variance），当 $|\theta| < 1$ 时，把它展开为等比级数（geometric series）：

$$M_X(\theta) = 1 + \theta^2 + \theta^4 + \cdots.$$

与 $M_X(\theta) = 1 + E(X)\,\theta + \frac{1}{2}E(X^2)\,\theta^2 + \cdots$ 比较，可得 $E(X) = 0$ 和 $\frac{1}{2}E(X^2) = 1$。因此

$$\mathrm{Var}(X) = E(X^2) - \big(E(X)\big)^2 = 2.$$

（求导两次再代入 $\theta = 0$ 也得到相同结果，但用级数较快。）

青蛙：$Y/\sqrt{2n}$ 的矩母函数

由于各次跳跃 $X_1, \ldots, X_n$ 互相独立，$e^{\theta X_1}, \ldots, e^{\theta X_n}$ 也互相独立，所以它们乘积的期望值（expectation）等于各期望值之积：

$$M_Y(\theta) = E\big(e^{\theta X_1}\cdots e^{\theta X_n}\big) = \prod_{i=1}^{n} M_{X_i}(\theta) = (1-\theta^2)^{-n}.$$

设 $Z = Y/\sqrt{2n}$，则 $E(e^{\theta Z}) = E\big(e^{(\theta/\sqrt{2n})Y}\big) = M_Y\big(\theta/\sqrt{2n}\big)$，所以

$$M_Z(\theta) = \left(1 - \frac{\theta^2}{2n}\right)^{-n}, \qquad |\theta| < \sqrt{2n}.$$

以 $\sqrt{2n}$ 作缩放是很自然的：$\mathrm{Var}(Y) = 2n$，所以 Z 的平均值为 0，方差为 1。

极限

固定 θ。只要 $n > \frac{1}{2}\theta^2$，便有 $0 \le \theta^2/(2n) < 1$，于是可以取对数，并使用对 $|u| < 1$ 成立的 $\ln(1-u) = -u - \frac{1}{2}u^2 - \frac{1}{3}u^3 - \cdots$：

$$\begin{aligned}\ln \mathrm{M}_Z(\theta) &= -n\ln\left(1 - \frac{\theta^2}{2n}\right) = n\left(\frac{\theta^2}{2n} + \frac{\theta^4}{8n^2} + \frac{\theta^6}{24n^3} + \cdots\right)\\ &= \frac{\theta^2}{2} + \frac{\theta^4}{8n} + \frac{\theta^6}{24n^2} + \cdots.\end{aligned}$$

? **想一想** 我们怎样知道，当 $n \to \infty$ 时，这个级数的尾部真的会消失?

设 $u = \theta^2/(2n)$，首项之后的所有项之和为 $n\left(\frac{1}{2}u^2 + \frac{1}{3}u^3 + \cdots\right) \le n \cdot \frac{1}{2}u^2(1 + u + u^2 + \cdots) = \dfrac{nu^2}{2(1-u)} = \dfrac{\theta^4}{8n(1-u)}$，由于 $u \to 0$，这个量趋向 0。所以 $\ln \mathrm{M}_Z(\theta) \to \frac{1}{2}\theta^2$，而 exp 是连续（continuous）函数，故

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) \to \exp\left(\tfrac{1}{2}\theta^2\right) \quad \text{当} n \to \infty,$$

这正是所要证明的。

最少跳跃次数

由于这个极限正是标准正态分布（standard Normal）的矩母函数，当 n 很大时，我们把 $Z = Y/\sqrt{2n}$ 视为近似服从 $\mathrm{N}(0,1)$。“落在距离起点 25 cm 或以上的位置”是指 $|Y| \ge 25$，两个方向都计算在内，所以

$$\mathrm{P}(|Y| \ge 25) = \mathrm{P}\left(|Z| \ge \frac{25}{\sqrt{2n}}\right) \approx 2\left(1 - \Phi\left(\frac{25}{\sqrt{2n}}\right)\right).$$

这个概率随 n 增加而上升（青蛙散布得更远），所以我们要找使它达到 5% 的最小 n。由于 $\Phi(1.96) = 0.975$，我们需要

$$\frac{25}{\sqrt{2n}} \le 1.96 \iff 2n \ge \left(\frac{25}{1.96}\right)^2 \approx 162.7 \iff n \ge 81.3.$$

所以最少跳跃次数是 $n = 82$。（当 $n = 81$ 时，近似概率约为 4.95%；当 $n = 82$ 时，约为 5.09%。）

验算

- $\mathrm{M}_X(0) = 1$，任何矩母函数都必须如此；而且 M_X 是偶函数，与略图的对称性吻合。
- 以数值积分（numerical quadrature）计算 $\int x^2 \mathrm{f}(x)\,\mathrm{d}x$ 得 2.0000，计算 $\int \mathrm{e}^{0.3x}\mathrm{f}(x)\,\mathrm{d}x$ 得 $1.0989 = 1/0.91$。

- 取 $\theta = 1$，$(1 - \frac{1}{2n})^{-n}$ 在 $n = 10$ 时为 1.6702，在 $n = 100$ 时为 1.6508，逐渐逼近 $e^{1/2} = 1.6487$。
- 模拟 10^5 只青蛙，每只作 82 次拉普拉斯（Laplace）跳跃，结果约有 5.1% 的青蛙最后落在距离起点至少 25 cm 的位置。

常见错误

- 使用单尾（one-tailed）数值 1.645：青蛙可以落在任何一侧 25 cm 以外，所以两边尾部都要计算。
- 写成 $M_{Y/\sqrt{2n}}(\theta) = M_Y(\theta)/\sqrt{2n}$。把变量缩放，缩放的是其矩母函数的自变量。
- 把 81.3 四舍五入至最接近的整数而答 $n = 81$：当 $n = 81$ 时，概率仅略低于 5%。

STEP 之外

你刚才证明的，就是步长服从拉普拉斯分布时的中心极限定理，而一般情况的证明也循同一路线：把 $\ln M(\theta/\sqrt{n})$ 展开至二阶，再证明余下部分会消失。在大学一年级的课程中，你会遇到“连续性定理”（continuity theorem），它说明为何可以由矩母函数（或特征函数（characteristic function），它必定存在）的收敛推出概率的收敛，这一步我们在这里只是姑且相信。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门）对两者都有论述。

详细解答

STEP III 2008，第 13 题

概率与统计 · 指示变量之和

STEP III 2008，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/08-s3-q13

本题在 *Beyond STEP* 第 12.10 节把绳头系起来中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/loose-ends.html

关键思路

无论某一步发生甚么事，绳子的段数总是刚好减少一：要么有一个绳圈退出游戏，要么两段绳子合成一段。因此每一步之后，不管之前发生过甚么，我们都回到同样的处境，只是少了一段绳子。这样就可以把绳圈的总数写成一些独立（independent）指示变量（indicator variable）之和，于是平均值和方差（variance）都只是简单的求和。

解答

第一步

有 n 段绳子时，共有 $2n$ 个自由端，所以选取两个端点的方法有

$$\binom{2n}{2} = n(2n-1)$$

种，而且各种选法的可能性相等。当且仅当两个端点属于同一段绳子时，才会形成绳圈，而这样的配对共有 n 个（每段绳子一个）。因此

$$\mathrm{P}(\text{第一步形成绳圈}) = \frac{n}{n(2n-1)} = \frac{1}{2n-1}.$$

换个说法：无论我们先选哪一个端点，余下的 $2n-1$ 个端点之中，恰好有一个是它的同伴。设 X_n 在第一步形成绳圈时取值 1，否则取值 0，则第一步所形成绳圈数目的期望值

（expectation）是

$$\mathrm{E}(X_n) = \frac{1}{2n-1}.$$

绳圈数目的期望值

第一步之后，剩下 $n-1$ 段绳子，每段各有两个自由端。两种情况下都是如此：形成绳圈会令一段绳子退出游戏，而把两段不同的绳子接起来则令两段变成一段。绳子的长短无关重要，要紧的只是它们的端点。

> **想一想** 现在有些绳子比其他的长，又或者之前有没有形成过绳圈，这些有影响吗？

没有影响。唯一会影响选中哪一对端点的，就是还剩下多少段绳子（因而还剩下多少个端点）。所以当剩下 k 段绳子时，不论之前各步的结果如何，这一步形成绳圈的概率都是 $\frac{1}{2k-1}$。

以每一步开始前的绳子段数 $k = n, n-1, \ldots, 1$ 来标记各步，并设 X_k 为该步形成绳圈的指示变量。于是 $\mathrm{P}(X_k = 1) = \frac{1}{2k-1}$，而整个过程在 $k = 1$ 那一步之后结束（这一步必定形成绳圈，即 $X_1 = 1$）。绳圈的总数是

$$R = X_1 + X_2 + \cdots + X_n,$$

所以由期望值的线性性质（linearity of expectation）可得

$$\mathrm{E}(R) = \sum_{k=1}^{n} \frac{1}{2k-1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}.$$

方差

求方差时，单靠线性性质并不足够：我们还需要各个 X_k 互相独立。它们确实如此。无论之前各步的结果是甚么，有 k 段绳子的那一步形成绳圈的概率仍然是 $\frac{1}{2k-1}$。所以在已知之前所有 X 的条件下，X_k 的条件分布（conditional distribution）与它的无条件分布相同，因此各个 X_k 互相独立。

每个 X_k 都服从伯努利分布（Bernoulli distribution），其中 $p_k = \frac{1}{2k-1}$，所以 $\mathrm{Var}(X_k) = p_k(1-p_k)$；而对于独立变量，方差可以相加：

$$\begin{aligned}\mathrm{Var}(R) &= \sum_{k=1}^{n} \frac{1}{2k-1}\left(1 - \frac{1}{2k-1}\right) = \sum_{k=1}^{n} \frac{2k-2}{(2k-1)^2} \\ &= \sum_{k=1}^{n} \frac{1}{2k-1} - \sum_{k=1}^{n} \frac{1}{(2k-1)^2}.\end{aligned}$$

$n = 40\,000$ 的情况

记 $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$。直至 $2n-1$ 为止的奇数倒数之和，等于直至 $2n$ 为止的所有倒数之和，减去其中偶数的倒数：

$$\sum_{k=1}^{n} \frac{1}{2k-1} = H_{2n} - \left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2n}\right) = H_{2n} - \tfrac{1}{2}H_n.$$

利用 m 很大时 $H_m \approx \ln m$，

$$\mathrm{E}(R) \approx \ln(2n) - \tfrac{1}{2}\ln n = \tfrac{1}{2}\left(2\ln 2n - \ln n\right) = \tfrac{1}{2}\ln(4n).$$

当 $n = 40\,000$ 时，$4n = 160\,000 = 20^4$，所以

$$\mathrm{E}(R) \approx \tfrac{1}{2}\ln\left(20^4\right) = 2\ln 20 \approx 6.$$

因此，由 40 000 段绳子开始，我们预期只会得到大约六个绳圈。

验算

- $n = 1$：只有一段绳子，必定得到一个绳圈。公式给出 $\mathrm{E}(R) = 1$ 和 $\mathrm{Var}(R) = 0$，正如所料。
- $n = 2$：第一步形成绳圈的概率是 $\frac{1}{3}$，而最后一步必定形成绳圈，所以 R 取值 2 的概率是 $\frac{1}{3}$，否则取值 1。由此得平均值为 $\frac{4}{3}$，方差为 $\frac{2}{9}$，与公式给出的 $1 + \frac{1}{3}$ 和 $\frac{2}{9}$ 吻合。
- 对 $n \le 5$ 精确地列举所有可能的选择序列，并对 $n = 6$ 进行模拟，结果都与两条公式一致。当 $n = 40\,000$ 时，精确的和是 $\mathrm{E}(R) \approx 6.28$。比我们的估计多出的 0.29，来自 $H_m \approx \ln m$ 这个粗略的近似（见下文）。

常见错误

- 把形成绳圈的概率当作 $\frac{1}{n}$（选两段绳子），而不是 $\frac{1}{2n-1}$（选两个端点）。
- 没有先说明各个 X_k 互相独立，就把方差相加。求平均值时有线性性质已经足够，求方差则不然。
- 没有把偶数项分拆出来，就以 $\ln n$ 或 $\frac{1}{2}\ln n$ 来近似 $\sum \frac{1}{2k-1}$。

STEP 之外

更精确的估计是 $H_m = \ln m + \gamma + o(1)$，其中 $\gamma \approx 0.5772$ 是欧拉－马斯刻若尼常数（Euler–Mascheroni constant）。由此可得 $\mathrm{E}(R) \approx \frac{1}{2}\ln(4n) + \frac{1}{2}\gamma$，以及 $\mathrm{Var}(R) \approx \frac{1}{2}\ln(4n) + \frac{1}{2}\gamma -$

$\frac{\pi^2}{8}$，因为 $\sum 1/(2k-1)^2 = \pi^2/8$。同样的“概率为 $1/k$ 的独立指示变量之和”结构，也可以用来数算随机置换（random permutation）的循环（cycle）数目，以及随机数列中破纪录（record）的次数；两种情况下，这个数目的增长都与 $\ln n$ 相若。想进一步了解指示变量和这些随机结构，Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）是个很好的起点。

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

2008 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。

