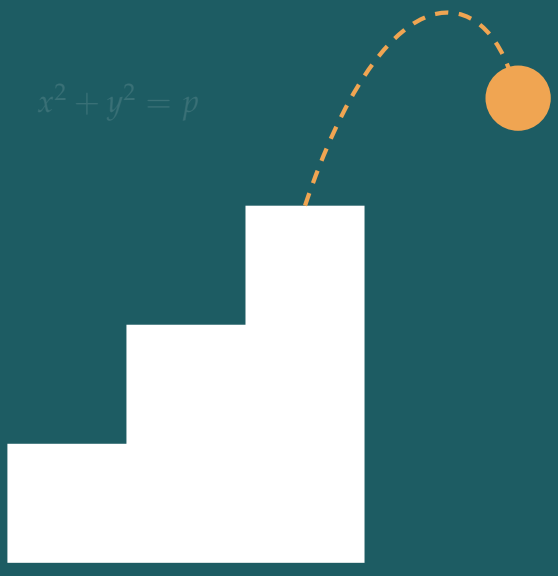
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2006 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

2006 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2006 年详细解答

2006 年 STEP 试卷（STEP I, II 及 III）全部 37 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 平方差 纯数学 . . . 3

第 2 题 面积与最优化 纯数学 . . . 6

第 3 题 多项式的根 纯数学 . . . 9

第 4 题 正多边形的面积 纯数学 . . . 13

第 5 题 代换积分法 纯数学 . . . 17

第 6 题 佩尔方程的变换 纯数学 . . . 20

第 7 题 含绝对值符号的积分 纯数学 . . . 24

第 8 题 三维向量 纯数学 . . . 28

第 9 题 连接质点 力学 . . . 31

第 10 题 抛体运动 力学 . . . 35

第 11 题 弹性碰撞 力学 . . . 39

第 12 题 条件概率 概率与统计 . . . 43

第 13 题 泊松混合分布 概率与统计 . . . 46

第 14 题 不放回抽样 概率与统计 . . . 49

STEP II

第 1 题 周期性递推关系 纯数学 . . . 52

第 3 题 根式与近似 纯数学 . . . 56

第 4 题 代换积分法 纯数学 . . . 59

第 5 题 下取整函数与面积 纯数学 . . . 63

第 7 题 椭圆的切线 纯数学 . . . 67

第 8 题 直线的向量几何 纯数学 . . . 71

第 9 题 摩擦力与倾倒 力学 . . . 75

第 10 题 碰撞与恢复系数 力学 . . . 79

第 14 题 连续随机变量 概率与统计 83

STEP III

第 1 题 曲线描绘 纯数学 87
第 2 题 利用对称性求定积分 纯数学 92
第 3 题 幂级数的系数 纯数学 96
第 4 题 函数方程 纯数学 100
第 5 题 复数与几何 纯数学 104
第 6 题 极坐标曲线与反射 纯数学 108
第 7 题 一阶微分方程 纯数学 112
第 8 题 公理化微分法 纯数学 116
第 9 题 能量守恒 力学 120
第 10 题 动量守恒 力学 124
第 11 题 连接系统中的碰撞 力学 128
第 12 题 动态近似与期望值 概率与统计 133
第 13 题 几何概率 概率与统计 136
第 14 题 方差与独立性 概率与统计 140

详细解答

STEP I 2006，第 1 题

纯数学 · 平方差

> **STEP I 2006，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/06-s1-q1

关键思路

若 $(n+m)^2 - 33\,127 = k^2$，则 $33\,127 = (n+m)^2 - k^2$ 是一个平方差（difference of two squares），因此可分解为 $(n+m-k)(n+m+k)$。这正是费马（Fermat）用来做因式分解（factorisation）的古老技巧：由刚好大于 $\sqrt{N}$ 的整数开始，逐步向上试，直至 $a^2 - N$ 是一个平方数为止。把这个论证倒过来看，33 127 的每一个分解都给出一个 m 的值，所以要数算 m 有多少个值，归根究底就是数算有多少个分解。

解答

求 n

由于 $180^2 = 32\,400$，33 127 的平方根略大于 180。我们求得

$$182^2 = 33\,124 < 33\,127 < 33\,489 = 183^2,$$

所以 $n = 182$。留意 $33\,127 = 182^2 + 3$，所以它本身并不是完全平方数（perfect square）。

一个较小的 m 值

我们依次试 $m = 1, 2, \ldots$：

$$183^2 - 33\,127 = 33\,489 - 33\,127 = 362, \quad \text{不是平方数}(19^2 = 361),$$
$$184^2 - 33\,127 = 33\,856 - 33\,127 = 729 = 27^2.$$

所以 $m = 2$ 可行，此时 $(n+m)^2 - 33\,127 = 27^2$。

分解 33 127

移项后运用平方差公式，

$$33\,127 = 184^2 - 27^2 = (184-27)(184+27) = 157 \times 211,$$

所以我们可取 $p = 157$ 和 $q = 211$，两者均大于 1。

恰好有两个 m 值

两个因数都是质数（prime）。对于 157：由于 $13^2 = 169 > 157$，我们只需测试质数 $2, 3, 5, 7, 11$，而当中没有一个能整除 157。对于 211：由于 $15^2 = 225 > 211$，我们测试 $2, 3, 5, 7, 11, 13$，而当中没有一个能整除 211。所以 $33\,127 = 157 \times 211$ 就是质因数分解（prime factorisation），而把 33 127 写成正整数之积 uv（其中 $u \le v$）的方法只有

$$1 \times 33\,127 \qquad 和 \qquad 157 \times 211.$$

现在假设 $(n+m)^2 - 33\,127 = k^2$，其中 $k \ge 0$。我们取 $a = n + m > 0$，这正是题目所设想的情况（由 n 开始向上走）。于是

$$(a-k)(a+k) = 33\,127.$$

这里 $a + k > 0$，所以 $a - k > 0$ 亦成立，而且 $a - k \le a + k$。因此 $(a-k,\, a+k)$ 必定是上面两对 (u, v) 的其中之一，于是

$$a = \frac{u+v}{2}, \qquad k = \frac{v-u}{2}.$$

由于 u 和 v 都是奇数，这两个数都是整数。

想一想 为甚么每一个分解恰好给出一个 m，又为甚么不会有其他 m 混进来？

这个对应是双向的。给定 a 和 k，数对 $(a-k, a+k)$ 就是一个分解；给定一个分解 (u, v)，上面的公式可唯一地还原出 a 和 k。所以 a 的值（因而 $m = a - 182$ 的值）与因数对一一对应（one-to-one correspondence），而因数对恰好有两个。

- $(u, v) = (157, 211)$ 给出 $a = 184$，$k = 27$，所以 $m = 184 - 182 = 2$，即我们之前找到的值。
- $(u, v) = (1, 33\,127)$ 给出 $a = 16\,564$，$k = 16\,563$，所以

$$m = 16\,564 - 182 = 16\,382.$$

因此恰好有两个 m 值，即 $m = 2$ 和 $m = 16\,382$。验算第二个值：$16\,564^2 - 16\,563^2 = 16\,564 + 16\,563 = 33\,127$。

（假如我们也容许 $n + m$ 为负数，则 a 可以是 -184 或 $-16\,564$，给出镜像的值 $m = -366$ 和 $m = -16\,746$。题目的原意显然是 $n + m$ 为正数。）

验算

用电脑直接搜索由 0 至 16 564 的每一个 m，结果发现只有当 $m = 2$ 和 $m = 16\,382$ 时，$(182 + m)^2 - 33\,127$ 才是完全平方数，而且 33 127 的因数恰好是 $1, 157, 211, 33\,127$。用手计算，$157 \times 211 = 157 \times 200 + 157 \times 11 = 31\,400 + 1\,727 = 33\,127$。

常见错误

- 忘记了平凡的分解 $1 \times 33\,127$，因而只找到一个 m 值。
- 未经检查就声称 157 和 211 是质数：“恰好有两个值”这个论证正是建基于此。
- 没有检查 u 和 v 的奇偶性（parity）是否相同。这里两者都是奇数，所以 a 和 k 都是整数；但若换成一个不是 4 的倍数的偶数，就没有任何分解可行。

STEP 之外

这就是费马因式分解法（Fermat's factorisation method），当 N 有两个接近 $\sqrt{N}$ 的因数时（正如本题），这方法很快见效。同一个想法，即改为寻找 $a^2 \equiv k^2 \pmod{N}$，而非要求 $a^2 - N = k^2$ 严格成立，正是二次筛法（quadratic sieve）的基础，而二次筛法是最快的通用因数分解算法之一。分解与平方差表示法之间的联系，正好让你初尝 Davenport 所著 *The Higher Arithmetic*（一本初等数论入门书）中的初等数论（elementary number theory）。

详细解答

STEP I 2006，第 2 题

纯数学 · 面积与最优化

STEP I 2006，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/06-s1-q2

关键思路

用一个数来确定系绳点的位置，即它与其中一个角的距离 x，然后跟着绳子看它怎样绕过谷仓。这样，吃草区域便分成一个半圆（semicircle）和四个四分之一圆（quarter-circle），它们的半径都是 x 的简单线性函数。因此 A 是 x 在一个闭区间（closed interval）上的二次函数（quadratic），题目的两个部分都归结为求它的最大值和最小值。

解答

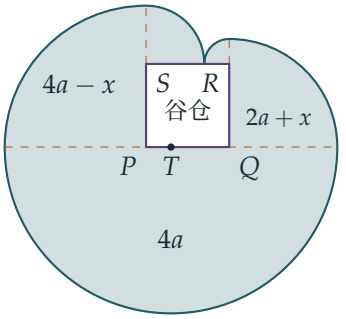
建立模型

把谷仓标记为 $PQRS$，山羊系在边 PQ 上的点 T，其中 $PT = x$，$TQ = 2a - x$，$0 \le x \le 2a$。根据对称性（symmetry），山羊系在哪一条边上并无分别。

若由 T 出发、避开谷仓而到达某一点的最短路径的长度不超过 $4a$，山羊便能到达该点。绳子拉紧时，这条路径除了在谷仓的角上转弯之外，都是直线。由 T 起绕着谷仓看：

- PQ **的前方**。直线 PQ 另一侧的每一点，都可以由 T 直接看见，所以山羊可吃到一个半径为 $4a$ 的半圆内的草。
- PS **和** QR **的旁边**。绳子绕过 P 之后，余下的长度是 $4a - x$，它沿着边 PS 扫出一个以 P 为圆心的四分之一圆。同样，在 Q 处余下 $4a - (2a - x) = 2a + x$，于是沿着 QR 得到一个以 Q 为圆心的四分之一圆。
- **谷仓的后方**。由于 $x \le 2a$，我们有 $4a - x \ge 2a$，所以绳子可到达 S，而且还余下 $(4a - x) - 2a = 2a - x$，这段绳子在边 SR 的后方扫出一个以 S 为圆心的四分之一圆。同理，绳子到达 R 时还余下 $(2a + x) - 2a = x$，于是得到一个以 R 为圆心、半径为 x

的四分之一圆。



想一想 谷仓后方的两块区域会否重叠，以致我们把某些草地重复计算了两次?

以 S 和 R 为圆心的两个四分之一圆位于直线 SR 的同一侧。它们的半径 $2a-x$ 和 x 加起来等于 $2a=SR$，所以两个圆刚好在 SR 上与 S 相距 $2a-x$ 的那一点相切。因此它们并不重叠，而两者合起来，山羊可到达整条边 SR。其他各块区域也互不重叠：半圆位于直线 PQ 的一侧，四个四分之一圆则位于另一侧，各自占据谷仓旁边或后方的一个区域。此外，谷仓后方的一点能够到达，当且仅当它与 S 或 R 的距离不超过该处余下的绳长，所以没有遗漏任何地方。

面积

把半圆和四个四分之一圆加起来，

$$\begin{aligned} A &= \tfrac{1}{2}\pi(4a)^2 + \tfrac{1}{4}\pi\left[(4a-x)^2 + (2a+x)^2 + (2a-x)^2 + x^2\right] \\ &= 8\pi a^2 + \tfrac{1}{4}\pi\left(4x^2 - 8ax + 24a^2\right) \\ &= \pi\left(x^2 - 2ax + 14a^2\right) \\ &= \pi\left[(x-a)^2 + 13a^2\right]. \end{aligned}$$

这条公式对 $[0,2a]$ 内的每一个 x 都成立。在两个端点（$x=0$ 或 $x=2a$，即山羊系在角上），其中一个四分之一圆缩成一点，而半圆与半径为 $4a$ 的四分之一圆则合并成四分之三个圆，正如我们所预期。

最大值与最小值

在 $0 \le x \le 2a$ 上，我们有 $-a \le x-a \le a$，所以 $0 \le (x-a)^2 \le a^2$。因此

$$13\pi a^2 \le A \le 14\pi a^2,$$

这就证明了 $A \le 14\pi a^2$。右边的等号恰好在 $x = 0$ 或 $x = 2a$ 时成立，即山羊系在角上的时候。左边的界在 $x = a$ 时取得，所以 A 的最小值是

$$A_{\min} = 13\pi a^2,$$

在山羊系于一条边的中点时取得。

验算

- 在角上（$x = 0$），我们得到四分之三个半径为 $4a$ 的圆（$12\pi a^2$）、绕过邻边远端那个角的一个半径为 $2a$ 的四分之一圆（πa^2），以及谷仓后方一个半径为 $2a$ 的四分之一圆（πa^2）。合共 $14\pi a^2$，与公式相符。
- 我们取 $a = 1$，在一个细密的网格上估算吃草面积，逐点按它绕过谷仓的真正最短距离（一条直线，或在角上转弯的折线）来判断。系于中点时得 40.845，而 $13\pi \approx 40.841$；系于角上时得 43.988，而 $14\pi \approx 43.982$；其他系绳点的结果也同样吻合。

常见错误

- 忘记了谷仓后方的区域：绳子的长度足以在每一边绕过两个角。
- 担心谷仓后方的两块区域重叠，或在计算中硬加入重叠部分。它们的半径之和永远恰好是 $2a$，所以两个圆只是相切。

STEP 之外

拉紧的绳子所走的是一条*测地线*（geodesic），即在有障碍物的区域内的最短路径。在这里，它由若干直线段组成，只在谷仓的角上转弯，所以山羊能到达的点集，是在一种非欧几里得（non-Euclidean）的距离量度方式下的一个“球”。绕过障碍物的最短路径，通向变分法（calculus of variations），也通向计算机科学中的运动规划（motion planning）问题。Siklos 的 *Advanced Problems in Mathematics* 收录了更多这类题目，当中只要仔细分情况讨论，便可把几何问题化为简单的代数。

详细解答

STEP I 2006，第 3 题

纯数学 · 多项式的根

STEP I 2006，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/06-s1-q3

关键思路

着眼于图像，而不是公式。一条曲线若起初在轴的下方，最后又必须到达轴的上方，就一定会穿过轴；而 $x = 0$ 处的值（即常数项）告诉你曲线在原点处位于轴的哪一侧。对于三次函数（cubic），关键在于转向点（turning point）：曲线与轴相交三次，当且仅当它的局部极大值（local maximum）在轴的上方，而局部极小值（local minimum）在轴的下方。

解答

第 (i) 部分

设 $f(x) = x^2 + bx + c$。其图像是一条开口向上的抛物线（parabola），所以当 $x \to \pm\infty$ 时，$f(x) \to +\infty$。若 $c < 0$，则 $f(0) = c < 0$：曲线在 $x = 0$ 处位于轴的下方。

由于 f 连续（continuous），当 x 是绝对值很大的负数时取正值，而在 0 处取负值，所以它在 $(-\infty, 0)$ 内某处穿过轴。同理，它在 $(0, \infty)$ 内再穿过轴一次。这两个根（root）分别位于 0 的两侧，所以它们并不相同。二次方程（quadratic）最多有两个根，因此该方程恰好有两个相异实根，一负一正。所以 $c < 0$ 是*充分*条件（sufficient condition）。

它*并非必要*条件（necessary condition）。例如，$x^2 - 3x + 2 = (x - 1)(x - 2)$ 有相异实根 1 和 2，但 $c = 2 > 0$。（既必要又充分的条件是 $b^2 > 4c$；当 $c < 0$ 时，这个条件自动成立，因为 $b^2 \geq 0 > 4c$。）

第 (ii) 部分

我们断言：该方程有两个相异正根，当且仅当

$$b^2 > 4c, \qquad b < 0, \qquad c > 0.$$

必要性。假设两根为 $\alpha \neq \beta$，其中 $\alpha, \beta > 0$。那么 $x^2 + bx + c = (x - \alpha)(x - \beta)$，所以

$$c = \alpha\beta > 0, \qquad b = -(\alpha + \beta) < 0, \qquad b^2 - 4c = (\alpha + \beta)^2 - 4\alpha\beta = (\alpha - \beta)^2 > 0.$$

充分性。若 $b^2 > 4c$，则两根 α, β 是相异实数。两根之积为 $c > 0$，所以它们同号。两根之和为 $-b > 0$，所以它们同为正数。

> **想一想** 为甚么单靠 $b < 0$ 和 $c > 0$ 并不足够?

考虑 $x^2 - x + 1$：这里 $b = -1 < 0$ 而 $c = 1 > 0$，但 $b^2 - 4c = -3 < 0$，所以根本没有实根。从图像来看，$c > 0$ 使曲线在 $x = 0$ 处位于轴的上方，而 $b < 0$ 使顶点（vertex）位于 $x = -\frac{b}{2} > 0$。我们还需要顶点位于轴的下方：顶点的高度是 $c - \frac{b^2}{4}$，它为负数当且仅当 $b^2 > 4c$。三个条件缺一不可。

第 (iii) 部分

设 $g(x) = x^3 + px + q$，所以 $g'(x) = 3x^2 + p$。实系数三次方程连重数（multiplicity）计算共有三个根，而它的非实根以共轭复数（complex conjugate）成对出现。另外请注意，在两种情况下都有 $g(0) = q < 0$。

$p > 0$、$q < 0$ 的情况。 这里对所有 x 都有 $g'(x) = 3x^2 + p > 0$，所以 g 严格递增（strictly increasing），恰好有一个实根。由于 $g(0) = q < 0$ 而 g 递增，这个根是正数。余下的两个根是一对非实数的共轭复数。

所以有一个实根，它是正数而且是单根（simple root），另有两个非实数的共轭复根。

$p < 0$、$q < 0$ 的情况。 设 $p = -3a^2$，其中 $a > 0$，即 $a = \sqrt{-p/3}$。那么 $g'(x) = 3(x^2 - a^2)$，所以 g 在 $(-\infty, -a]$ 上递增，在 $[-a, a]$ 上递减，在 $[a, \infty)$ 上递增。在 $x = -a$ 处有局部极大值，在 $x = a$ 处有局部极小值，并且

$$g(-a) = -a^3 + 3a^3 + q = 2a^3 + q, \qquad g(a) = a^3 - 3a^3 + q = q - 2a^3.$$

由于 $q < 0$，局部极小值 $g(a)$ 是负数。在 $[0, a]$ 上，g 由 $g(0) = q < 0$ 开始递减，所以在该区间内没有根。在 $[a, \infty)$ 上，它由 $g(a) < 0$ 递增至 $+\infty$，所以恰好有一个大于 a 的根，这个根是正数。任何其他实根必定位于 $(-\infty, -a]$ 或 $(-a, 0)$ 内，所以是负数；而这样的根是否存在，只取决于局部极大值 $g(-a)$ 的正负。

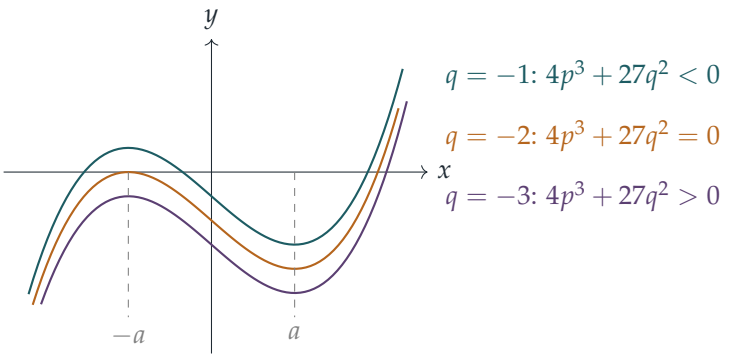
想一想 $g(-a)$ 的正负与 $4p^3+27q^2$ 有甚么关系?

把两个转向点的函数值相乘:

$$g(-a)\,g(a)=(q+2a^3)(q-2a^3)=q^2-4a^6=q^2-4\Big(-\frac{p}{3}\Big)^3=\frac{4p^3+27q^2}{27}.$$

由于 $g(a)<0$，$g(-a)$ 的正负与 $4p^3+27q^2$ 的正负相反。共有三种情况。

- $4p^3+27q^2>0$。此时 $g(-a)<0$，所以在整个 $(-\infty,a]$ 上都有 $g<0$（它先递增至负值 $g(-a)$，然后递减）。唯一的实根就是那个正根。与之前一样，有一个正实根和两个非实数的共轭复根。
- $4p^3+27q^2=0$。此时 $g(-a)=0$ 且 $g'(-a)=0$，所以 $x=-a$ 是重根（repeated root）。由于各根之和为 0（方程没有 x^2 项），第三个根是 $2a$。三个根是 $-a$、$-a$、$2a$：一个负的二重根和一个正的单根，全部都是实数。
- $4p^3+27q^2<0$。此时 $g(-a)>0$。由于当 $x\to-\infty$ 时 $g(x)\to-\infty$，在 $(-\infty,-a)$ 内有一个根。由于 $g(-a)>0>g(0)$，在 $(-a,0)$ 内还有一个根。连同大于 a 的那个根，共有三个相异实根：两负一正。



草图显示了 $q=-1$、-2、-3 时的 $y=x^3-3x+q$（即 $p=-3$，$a=1$）：把局部极大值提升至越过轴，一个实根就变成三个。

在 $q<0$ 的每一种情况下，都恰好有一个正根；这与各根之积为 $-q>0$ 一致。

验算

- 取 $p=-3$、$q=-1$，得 $4p^3+27q^2=-108+27=-81<0$，而数值计算所得的根约为 -1.532、-0.347 和 1.879：两负一正，与预测相符。
- 取 $p=-3$、$q=-2$：$4p^3+27q^2=0$，而 $x^3-3x-2=(x+1)^2(x-2)$，得出根 $-1,-1,2$。

- 在每种情况下，对数千组随机的 (p,q) 使用数值求根程序，并以随机的 (b,c) 检验 (ii) 中的条件，结果与上述每一项结论都吻合。

常见错误

- 在 (ii) 中只给出 $b<0$ 和 $c>0$（这两个条件只保证如果根是实数，它们便是正数），而忘记了 $b^2>4c$。
- 在 (ii) 中只证明必要性，或只证明充分性。“必要且充分”需要两个方向都证明。
- 在 (iii) 中忘记了边界情况 $4p^3+27q^2=0$，此时有重根。

STEP 之外

$-(4p^3+27q^2)$ 这个量是三次式 x^3+px+q 的判别式（discriminant），与 b^2-4ac 完全对应。它等于各根两两之差的平方之积，所以它为零当且仅当有重根，而它为正数当且仅当三个根都是实数且互不相同。判别式，以及根与系数之间的关系，通向伽罗瓦理论（Galois theory）以及“哪些方程可用根式求解（solvable by radicals）”这个问题。Allenby 的 *Rings, Fields and Groups*（环、域与群）是很好的入门读物。

详细解答

STEP I 2006，第 4 题

纯数学 · 正多边形的面积

STEP I 2006，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：

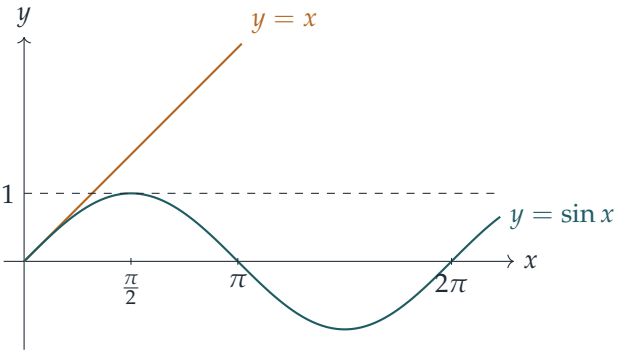
step.maths.org/questions/06-s1-q4

关键思路

直线 $y = x$ 是 $y = \sin x$ 在原点的切线（tangent），而正弦曲线向下弯离这条直线。单凭这一幅图，就同时得到不等式和近似式。至于多边形，把它分割成 n 个在中心相接的全等（congruent）等腰三角形；所有的量于是都变成半角 π/n 的三角运算，而不等式 $\sin x < x$ 原来正好决定了导数（derivative）的正负号。

解答

两个图像



两条曲线都通过原点，而且在该处的斜率（gradient）都是 1，因为在 $x = 0$ 时 $\frac{\mathrm{d}}{\mathrm{d}x}\sin x = \cos x = 1$。所以 $y = x$ 是正弦曲线在 O 的切线。

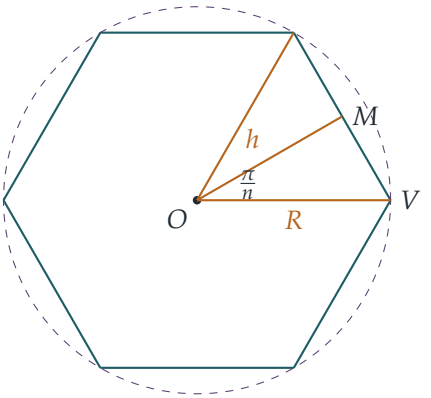
(i) 当 $0 < x \leq \frac{\pi}{2}$ 时，正弦曲线是凹的（concave）（$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sin x = -\sin x < 0$），所以它位于其在 O 的切线之下，由此得 $\sin x < x$。（等价的说法：$g(x) = x - \sin x$ 满足 $g(0) = 0$，且当

$0 < x < 2\pi$ 时 $g'(x) = 1 - \cos x > 0$，所以 g 在该区间上严格递增（strictly increasing）。）当 $x > \frac{\pi}{2}$ 时，我们有 $\sin x \le 1 < \frac{\pi}{2} < x$。所以对所有 $x > 0$ 都有 $x > \sin x$。

(ii) 由于这条直线是在 O 的切线，两个图像在原点附近几乎无法分辨：当 x 很小时，高度 $\sin x$ 与高度 x 非常接近，所以 $\dfrac{\sin x}{x} \approx 1$。（更精确的说法：由于当 $x > 0$ 时 $\sin x \ge x - \frac{x^3}{6}$，我们有 $1 - \frac{x^2}{6} \le \frac{\sin x}{x} < 1$。）

多边形的面积

把多边形的中心 O 与各顶点连起来。这样就把多边形分割成 n 个全等的等腰三角形，每个三角形的底等于一条边 $s = P/n$，在 O 的顶角为 $2\pi/n$。从 O 作垂线至一条边的中点 M；这条垂线平分顶角。



在直角三角形 OMV 中（其中 V 是一个顶点），在 O 的角是 π/n，且 $MV = \frac{1}{2}s$。所以由中心到一条边的距离 $h = OM$ 是

$$h = \frac{s/2}{\tan(\pi/n)} = \frac{P}{2n\tan(\pi/n)}.$$

这 n 个三角形每个的面积都是 $\frac{1}{2}sh$，所以多边形的面积是

$$A = n \cdot \tfrac{1}{2} \cdot \frac{P}{n} \cdot \frac{P}{2n\tan(\pi/n)} = \frac{P^2}{4n\tan(\pi/n)},$$

正如所求。

面积随 n 增加

固定 P，把面积写成 $A = \dfrac{P^2}{4f(n)}$，其中 $f(n) = n\tan\dfrac{\pi}{n}$。由于当 $n \ge 3$ 时 $f(n) > 0$，A 递增恰好等同于 f 递减。根据积法则（product rule）和链式法则（chain rule），

$$f'(n) = \tan\frac{\pi}{n} + n\sec^2\frac{\pi}{n} \cdot \left(-\frac{\pi}{n^2}\right) = \tan\theta - \theta\sec^2\theta, \qquad \text{其中} \theta = \frac{\pi}{n}.$$

想一想 我们怎样判断 $\tan\theta - \theta\sec^2\theta$ 的正负号?

抽出正因子 $\sec^2\theta$，并利用 $\sin\theta\cos\theta = \frac{1}{2}\sin 2\theta$:

$$f'(n) = \sec^2\theta\left(\sin\theta\cos\theta - \theta\right) = \tfrac{1}{2}\sec^2\theta\left(\sin 2\theta - 2\theta\right).$$

当 $n \geq 3$ 时, 我们有 $0 < \theta \leq \frac{\pi}{3}$, 所以 $2\theta > 0$, 由第 (i) 部分得 $\sin 2\theta < 2\theta$。因此 $f'(n) < 0$, 并且

$$\frac{\mathrm{d}A}{\mathrm{d}n} = -\frac{P^2 f'(n)}{4f(n)^2} > 0.$$

所以，在周界（perimeter）固定的情况下，面积随 n 增加而增加。

与外接圆比较

包含这个多边形的最小的圆是外接圆（circumcircle），它以 O 为圆心，并通过所有顶点。要看出更小的圆都不行，设各顶点为 $V_1, \ldots, V_n$，其中 $\overrightarrow{OV_k} = \mathbf{v}_k$，$|\mathbf{v}_k| = R$。由对称性可知 $\sum \mathbf{v}_k = \mathbf{0}$，所以对任意一点 C，设 $\overrightarrow{OC} = \mathbf{c}$，则有

$$\frac{1}{n}\sum_{k=1}^{n}|\mathbf{v}_k - \mathbf{c}|^2 = R^2 - \frac{2}{n}\,\mathbf{c}\cdot\sum_{k=1}^{n}\mathbf{v}_k + |\mathbf{c}|^2 = R^2 + |\mathbf{c}|^2 \geq R^2.$$

所以必有某个顶点与 C 的距离至少是 R，而以 C 为圆心并包含这个多边形的圆，半径至少是 R。外接圆半径（circumradius）是 $R = OV$，而由同一个直角三角形可得

$$\tfrac{1}{2}s = R\sin\theta \quad\Longrightarrow\quad R = \frac{P}{2n\sin\theta}.$$

所以两个面积之比是

$$\frac{A}{\pi R^2} = \frac{P^2}{4n\tan\theta}\cdot\frac{4n^2\sin^2\theta}{\pi P^2} = \frac{n}{\pi}\cdot\frac{\sin^2\theta}{\tan\theta} = \frac{\sin\theta\cos\theta}{\theta} = \frac{\sin 2\theta}{2\theta},$$

当中用到 $n/\pi = 1/\theta$。当 n 很大时，$2\theta = 2\pi/n$ 很小，所以由第 (ii) 部分可知这个比值约等于 1。

验算

对于正方形（$n = 4$，$\theta = \pi/4$），公式给出 $A = P^2/16$，这是正确的，因为每条边长 $P/4$；面积比是 $\sin(\pi/2)/(\pi/2) = 2/\pi \approx 0.637$，与单位圆内边长为 $\sqrt{2}$ 的正方形吻合。对于六边形，这个比值是 $3\sqrt{3}/(2\pi) \approx 0.827$。我们用鞋带公式（shoelace formula）核对了 $n = 3, \ldots, 39$ 时的面积公式，用数值导数核对了 $f'(n) < 0$，并求得这个比值在 $n = 100$ 时是 0.99934，在 $n = 1000$ 时是 0.9999934。

常见错误

- 在直角三角形中用了整个顶角 $2\pi/n$，而不是半角 π/n。
- 对 $\tan(\pi/n)$ 求导时，漏掉链式法则带来的因子 $-\pi/n^2$。
- 声称“因为 $\tan\theta \approx \theta$”，所以 $\tan\theta < \theta\sec^2\theta$：近似式不能决定正负号。能决定正负号的是精确的不等式 $\sin 2\theta < 2\theta$。

STEP 之外

极限（limit）$\sin x/x \to 1$ 正是令 $\sin x$ 的导数等于 $\cos x$ 的事实，而大学的第一门分析（analysis）课程会从 sin 的定义出发严格地证明它，而不是靠一幅图。这个单调性（monotonicity）结果，让我们从离散的角度一窥等周不等式（isoperimetric inequality）：在所有长度给定的闭曲线之中，圆所围的面积最大，而多边形的面积会递增并趋向圆的面积值 $P^2/(4\pi)$。如想看极限和导数的严格论述，可以读 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP I 2006，第 5 题

纯数学 · 代换积分法

STEP I 2006，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/06-s1-q5

关键思路

两个被积函数（integrand）都含有一个棘手的平方根，而要消去线性式的平方根，最干净利落的方法就是把整个根式设为新变量。根式一旦消失，剩下的便是有理函数（rational function），而有理函数总可以用部分分式（partial fractions）来积分。在第 (ii) 部分，最好先把指数函数除去，使平方根变成某个线性式的根，然后同一招便能派上用场。

解答

第 (i) 部分

由于 $x > 4$，我们有 $2x+1 > 9$，故可取 $u = \sqrt{2x+1} > 3$。由 $u^2 = 2x+1$ 得

$$2u\,\mathrm{d}u = 2\,\mathrm{d}x, \qquad \text{所以} \qquad \mathrm{d}x = u\,\mathrm{d}u,$$

以及

$$x - 4 = \frac{u^2-1}{2} - 4 = \frac{u^2-9}{2}.$$

代入后，

$$\int \frac{3}{(x-4)\sqrt{2x+1}}\,\mathrm{d}x = \int \frac{3}{\frac{1}{2}(u^2-9)\,u}\,u\,\mathrm{d}u = \int \frac{6}{u^2-9}\,\mathrm{d}u.$$

来自 $\mathrm{d}x$ 的因子 u 恰好把平方根约去，这正是这个代换（substitution）奏效的原因。现在 $u^2 - 9 = (u-3)(u+3)$，而

$$\frac{6}{(u-3)(u+3)} = \frac{1}{u-3} - \frac{1}{u+3},$$

你可以把右边通分来验证：分子是 $(u+3)-(u-3)=6$。因此

$$\int \frac{6}{u^2-9}\,\mathrm{d}u = \ln(u-3)-\ln(u+3)+K = \ln\left(\frac{u-3}{u+3}\right)+K,$$

这里无需加上绝对值符号，因为 $u>3$ 令 $u-3$ 和 $u+3$ 均为正数。把 $u=\sqrt{2x+1}$ 代回，得

$$\int \frac{3}{(x-4)\sqrt{2x+1}}\,\mathrm{d}x = \ln\left(\frac{\sqrt{2x+1}-3}{\sqrt{2x+1}+3}\right)+K,$$

正是所求。

第 (ii) 部分

设 I 为该积分。仿照第 (i) 部分，最自然的代换就是把平方根设为新变量：

$$w=\sqrt{\mathrm{e}^x+1}, \qquad w^2=\mathrm{e}^x+1.$$

于是 $2w\,\mathrm{d}w=\mathrm{e}^x\,\mathrm{d}x$，所以 $\mathrm{d}x=\dfrac{2w}{\mathrm{e}^x}\,\mathrm{d}w=\dfrac{2w}{w^2-1}\,\mathrm{d}w$。积分上下限变为

$$x=\ln 3:\ w=\sqrt{3+1}=2, \qquad x=\ln 8:\ w=\sqrt{8+1}=3.$$

因此

$$I=\int_2^3 \frac{2}{(w^2-1)\,w}\cdot\frac{2w}{w^2-1}\,\mathrm{d}w=\int_2^3\frac{4}{(w^2-1)^2}\,\mathrm{d}w.$$

? **想一想** 怎样才能不经冗长的计算，就把 $\dfrac{4}{(w^2-1)^2}$ 分解成部分分式?

从简单的分解式 $\dfrac{2}{w^2-1}=\dfrac{1}{w-1}-\dfrac{1}{w+1}$ 出发，把它平方：

$$\begin{aligned}\frac{4}{(w^2-1)^2} &= \frac{1}{(w-1)^2}-\frac{2}{(w-1)(w+1)}+\frac{1}{(w+1)^2}\\ &= \frac{1}{(w-1)^2}+\frac{1}{(w+1)^2}-\frac{1}{w-1}+\frac{1}{w+1},\end{aligned}$$

其中对中间一项再用了一次那个简单的分解式。现在每一项都可以一眼看出积分：

$$\int\frac{4}{(w^2-1)^2}\,\mathrm{d}w=-\frac{1}{w-1}-\frac{1}{w+1}-\ln\left(\frac{w-1}{w+1}\right)+C,$$

由于在积分范围内 $w\geq 2$，所以无需加上绝对值符号。代入上下限计算，

$$\begin{aligned}&\text{当}w=3: \quad -\tfrac{1}{2}-\tfrac{1}{4}-\ln\tfrac{2}{4}=-\tfrac{3}{4}+\ln 2,\\ &\text{当}w=2: \quad -1-\tfrac{1}{3}-\ln\tfrac{1}{3}=-\tfrac{4}{3}+\ln 3.\end{aligned}$$

所以

$$I = \left(-\tfrac{3}{4} + \ln 2\right) - \left(-\tfrac{4}{3} + \ln 3\right) = \frac{7}{12} + \ln\frac{2}{3},$$

这就是要证明的结果。

你同样可以分两步来做：先设 $v = \mathrm{e}^x$（得出 $\int_3^8 \frac{2}{v^2\sqrt{v+1}}\,\mathrm{d}v$），然后设 $w^2 = v+1$；上面的单一代换则一次过完成了这两步。

验算

- 对 (i) 的答案求导：设 $u = \sqrt{2x+1}$，则 $\frac{\mathrm{d}u}{\mathrm{d}x} = \frac{1}{u}$，而 $\frac{\mathrm{d}}{\mathrm{d}u}\ln\frac{u-3}{u+3} = \frac{6}{u^2-9}$，所以导数为 $\frac{6}{u(u^2-9)} = \frac{3}{(x-4)\sqrt{2x+1}}$。
- 数值上，$\frac{7}{12} + \ln\frac{2}{3} \approx 0.58333 - 0.40547 = 0.17787$，而在 $[\ln 3, \ln 8]$ 上对原来的被积函数应用辛普森法则（Simpson's rule），得出 0.177868，两者有六位数字吻合。
- 就正负号和大小作一个合理性检查：(ii) 的被积函数为正，并在长度为 $\ln\frac{8}{3} \approx 0.98$ 的区间上由约 0.33 下降至约 0.08，所以答案略小于 0.2 正合预期。

常见错误

- 在 (ii) 中变换变量时忘记更改积分上下限，或把 x 的上下限代入以 w 表示的式子。
- 把 $\frac{4}{(w^2-1)^2}$ 只写成 $\frac{A}{(w-1)^2} + \frac{B}{(w+1)^2}$：重复因子还需要线性项 $\frac{1}{w-1}$ 和 $\frac{1}{w+1}$。
- 在 (i) 中得出 $\mathrm{d}x = \mathrm{d}u$ 而不是 $\mathrm{d}x = u\,\mathrm{d}u$，结果留下一个多余的平方根。

STEP 之外

两部分的答案都由对数和有理项组成，原因在于一条一般性的定理：每个有理函数都有初等原函数（elementary antiderivative），可藉部分分式由有理函数、对数和反正切函数（inverse tangent）构成；而任何关于 x 和 $\sqrt{ax+b}$ 的有理被积函数，经代换 $u^2 = ax+b$ 后都会变成有理函数。把这个想法推广到二次式的根，便引出欧拉代换（Euler's substitutions）；而追问哪些积分不能以初等形式求出，则引出刘维尔（Liouville）的有限形式积分理论（integration in finite terms）。若想细致地了解积分及其背后的道理，可阅读 Abbott 的 *Understanding Analysis*（一本实分析入门书）。

详细解答

STEP I 2006，第 6 题

纯数学 · 佩尔方程的变换

STEP I 2006，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/06-s1-q6

关键思路

每一部分都在问：某个线性坐标变换（linear change of coordinates）会否保持一个二次式（quadratic expression）不变。直接的做法是代入、展开，然后比较 a^2、ab 和 b^2 的系数（coefficients）。背后其实有一幅更简洁的图像：$x^2 - 2y^2$ 是数 $x + y\sqrt{2}$ 的“大小”，而乘以 $3 + 2\sqrt{2}$（其大小为 $9 - 8 = 1$）不会改变这个大小。看通这一点，你就确切知道 (iii) 中的映射（map）该如何构造。

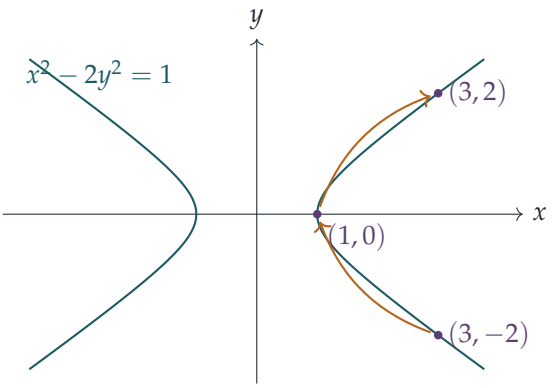
解答

第 (i) 部分

设 $a^2 - 2b^2 = 1$。则

$$\begin{aligned}(3a + 4b)^2 - 2(2a + 3b)^2 &= (9a^2 + 24ab + 16b^2) - 2(4a^2 + 12ab + 9b^2) \\ &= a^2 - 2b^2 \\ &= 1,\end{aligned}$$

所以 $(3a + 4b,\ 2a + 3b)$ 位于 $x^2 - 2y^2 = 1$ 上，正如所求。事实上，头两行是一条恒等式（identity），对任何 a 和 b 都成立：这个映射在每一处都保持 $x^2 - 2y^2$ 不变，而不仅是在曲线上。



这个映射把点沿着右边的分支推移：$(3,-2) \mapsto (1,0) \mapsto (3,2) \mapsto (17,12) \mapsto \cdots$。

第 (ii) 部分

设 $Ma^2 - Nb^2 = 1$。展开得

$$M(5a+6b)^2 - N(4a+5b)^2 = (25M-16N)a^2 + (60M-40N)ab + (36M-25N)b^2.$$

像点位于曲线上，当且仅当上式等于 $1 = Ma^2 - Nb^2$。减去 $Ma^2 - Nb^2$ 后，条件变成

$$\begin{aligned} 0 &= (24M-16N)a^2 + (60M-40N)ab + (36M-24N)b^2 \\ &= 4(3M-2N)(2a^2+5ab+3b^2) \\ &= 4(3M-2N)(2a+3b)(a+b). \end{aligned}$$

? **想一想** 我们真的需要 $3M = 2N$ 吗？会不会是其他因式等于零？

这个条件必须对曲线上每一点都成立。当 $M, N > 0$ 时，曲线 $Mx^2 - Ny^2 = 1$ 是一条双曲线（hyperbola），上面有无穷多个点。直线 $2x + 3y = 0$ 和 $x + y = 0$ 各自与它最多相交于两点，所以曲线上绝大部分的点都满足 $(2a+3b)(a+b) \neq 0$。因此我们需要 $3M - 2N = 0$。反过来说，若 $3M = 2N$，条件在每一点都成立。

所以 $3M = 2N$。于是 $2 \mid 3M$，故 M 是偶数。最小的选择是 $M = 2$，由此得 $N = 3$。最后验算一次：

$$2(5a+6b)^2 - 3(4a+5b)^2 = (50-48)a^2 + (120-120)ab + (72-75)b^2 = 2a^2 - 3b^2.$$

最小的正整数是 $M = 2$ 和 $N = 3$。

第 (iii) 部分

我们要找正整数，使得

$$(Pa+Qb)^2 - 3(Ra+Sb)^2 = a^2 - 3b^2 \quad \text{对任意} a, b,$$

这样必定能令曲线上每一点的像都位于曲线上。比较系数：

$$P^2 - 3R^2 = 1, \qquad PQ = 3RS, \qquad Q^2 - 3S^2 = -3.$$

第一条方程要求在 $x^2 - 3y^2 = 1$ 上找一个正整数点 (P, R)。最小的是 $(2,1)$，所以试取 $P = 2$，$R = 1$。第二条方程随即给出 $Q = \frac{3}{2}S$，而第三条变成 $\frac{9}{4}S^2 - 3S^2 = -3$，即 $S^2 = 4$。所以 $S = 2$ 而 $Q = 3$：

$$P = 2, \quad Q = 3, \quad R = 1, \quad S = 2.$$

验算：$(2a + 3b)^2 - 3(a + 2b)^2 = 4a^2 + 12ab + 9b^2 - 3a^2 - 12ab - 12b^2 = a^2 - 3b^2 = 1$。所以 $(2a + 3b,\ a + 2b)$ 位于 $x^2 - 3y^2 = 1$ 上。

想一想 这样的映射是从哪里来的?

写出 $(a + b\sqrt{3})(2 + \sqrt{3}) = (2a + 3b) + (a + 2b)\sqrt{3}$。$x + y\sqrt{3}$ 的范数（norm）$x^2 - 3y^2$ 具有积性（multiplicative），而 $2 + \sqrt{3}$ 的范数是 $4 - 3 = 1$，所以乘以它会保持范数不变。第 (i) 部分的原理相同，用的是 $3 + 2\sqrt{2}$。至于 (ii)，$(a\sqrt{2} + b\sqrt{3})(5 + 2\sqrt{6}) = (5a + 6b)\sqrt{2} + (4a + 5b)\sqrt{3}$，这就是 $2x^2 - 3y^2$ 这个形式在那里出现的原因。(iii) 还有其他正确答案：乘以 $(2 + \sqrt{3})^2 = 7 + 4\sqrt{3}$ 会得出 $(7a + 12b,\ 4a + 7b)$。

验算

- 从 $x^2 - 3y^2 = 1$ 上的 $(1,0)$ 出发。重复施行 (iii) 的映射得出 $(2,1)$、$(7,4)$、$(26,15)$、$(97,56)$，每一个都满足 $x^2 - 3y^2 = 1$（例如 $26^2 - 3 \cdot 15^2 = 676 - 675 = 1$）。
- 对于 (ii)，$(1/\sqrt{2}, 0)$ 在 $2x^2 - 3y^2 = 1$ 上，并映射到 $(5/\sqrt{2},\ 4/\sqrt{2})$。于是 $2 \cdot \frac{25}{2} - 3 \cdot \frac{16}{2} = 25 - 24 = 1$。
- 在 $1 \le M, N < 30$ 的范围内作穷举搜索，并以每条双曲线上的实际点来测试，证实 $(2,3)$ 是最小的可行数对。

常见错误

- 在 (ii) 中，只令 ab 的系数为零，却没有检查 a^2 和 b^2 的系数是否随之等于 M 和 $-N$。在这题中它们确实如此，但这一点必须写出来证明。
- 没有给出理由就说“恒等式必须对所有 a, b 成立”，而题目其实只给出曲线上的点。你需要一个论证，例如上面的因式分解（factorisation），来排除其他因式等于零的可能。

STEP 之外

这些都是佩尔方程（Pell equations）$x^2 - dy^2 = 1$。它们的整数解由环（ring）$\mathbb{Z}[\sqrt{d}]$ 中一个基本单位（fundamental unit）$x_1 + y_1\sqrt{d}$ 的幂生成，而这里的映射正是乘以这类单位。用矩阵的语言来说，每个映射都是一个保持某个二次型（quadratic form）不变的 2×2 整数矩阵，即某个“整正交群”（integral orthogonal group）的元素。Davenport 的 *The Higher Arithmetic*（《高等算术》）有精彩的一整章，以连分数（continued fractions）处理佩尔方程。

详细解答

STEP I 2006，第 7 题

纯数学 · 含绝对值符号的积分

STEP I 2006，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/06-s1-q7

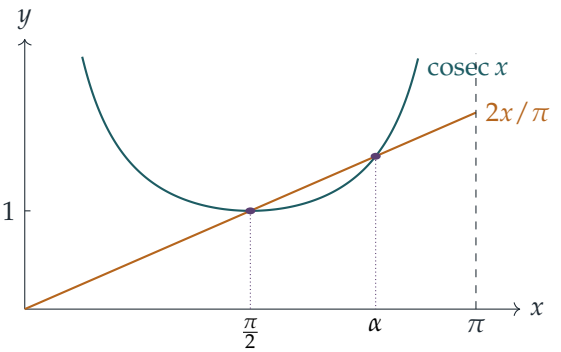
关键思路

把 $x\sin x = \pi/2$ 两边除以 $\frac{\pi}{2}\sin x$（它在 $(0,\pi)$ 上为正），方程便化为 $\operatorname{cosec} x = 2x/\pi$，所以方程的根（root）就是一条直线与一条 U 形曲线相交之处，而一条直线与一条凸（convex）曲线最多只能相交两次。至于积分，处理绝对值（modulus）的全部窍门，就是找出绝对值符号内的式子在何处变号，在该处把积分范围分开，然后以正确的正负号对每一段分别积分。

解答

第 (i) 部分：草图与根

在 $0 < x < \pi$ 上，$\operatorname{cosec} x$ 在两端均趋向 $+\infty$，关于 $x = \pi/2$ 对称，并在该处取得最小值 1。直线 $y = 2x/\pi$ 由 $(0,0)$ 伸延至 $(\pi,2)$，并通过 cosec 曲线的最低点 $(\pi/2,1)$。



由于在 $(0,\pi)$ 上 $\sin x > 0$,

$$x\sin x = \frac{\pi}{2} \iff \frac{1}{\sin x} = \frac{2x}{\pi},$$

所以方程的根就是草图中各交点的 x 坐标。其中一个根一眼便可看出：$x = \pi/2$，因为 $\frac{\pi}{2}\sin\frac{\pi}{2} = \frac{\pi}{2}$。

为了确定恰好有两个根，设 $h(x) = \operatorname{cosec} x - 2x/\pi$。那么

$$h''(x) = \operatorname{cosec} x\,(\operatorname{cosec}^2 x + \cot^2 x) > 0 \quad \text{于}(0,\pi),$$

所以 h 是严格凸（strictly convex）的，而严格凸函数最多只有两个零点。（假如它有三个零点 $a < b < c$，那么 b 严格地位于 a 与 c 之间，但 $h(b)$ 却落在连接 $(a,0)$ 与 $(c,0)$ 的弦（chord）上，而不是严格地在弦的下方。）现在 $h'(\pi/2) = -\cot\frac{\pi}{2}\operatorname{cosec}\frac{\pi}{2} - \frac{2}{\pi} = -\frac{2}{\pi} < 0$，所以 h 在 $\pi/2$ 的右侧附近为负，而当 $x \to \pi^-$ 时 $h(x) \to +\infty$。根据连续性（continuity），h 有一个零点 α，满足 $\pi/2 < \alpha < \pi$。因此方程恰好有两个根，即 $\pi/2$ 和 α，而 α 是较大的一个。

第 (i) 部分：积分

由草图（以及上述关于凸性的论证）可知，cosec 曲线在两个根之间位于直线下方，而在 α 之后则位于直线上方。把 $\operatorname{cosec} x \leqslant 2x/\pi$ 乘以正数 $\frac{\pi}{2}\sin x$，即得

$$x\sin x - \frac{\pi}{2} \geq 0 \text{ 于 } \left[\tfrac{\pi}{2},\alpha\right], \qquad x\sin x - \frac{\pi}{2} \leq 0 \text{ 于} [\alpha,\pi].$$

想一想 我们应怎样处理绝对值符号?

我们在变号之处 α 把积分分开，而在第二段上则对该式子的相反数积分。

运用分部积分法（integration by parts），取 $u = x$，$\mathrm{d}v = \sin x\,\mathrm{d}x$,

$$\int x\sin x\,\mathrm{d}x = -x\cos x + \int \cos x\,\mathrm{d}x = \sin x - x\cos x.$$

所以 $x\sin x - \pi/2$ 的一个原函数（antiderivative）是

$$F(x) = \sin x - x\cos x - \frac{\pi x}{2},$$

而所求的积分是

$$I = \big[F(x)\big]_{\pi/2}^{\alpha} - \big[F(x)\big]_{\alpha}^{\pi} = 2F(\alpha) - F\left(\tfrac{\pi}{2}\right) - F(\pi).$$

这三个值分别是

$$F(\alpha) = \sin\alpha - \alpha\cos\alpha - \frac{\alpha\pi}{2},$$

$$F\left(\tfrac{\pi}{2}\right) = 1 - 0 - \frac{\pi^2}{4},$$

$$F(\pi) = 0 - \pi(-1) - \frac{\pi^2}{2} = \pi - \frac{\pi^2}{2}.$$

因此

$$I = 2\sin\alpha - 2\alpha\cos\alpha - \alpha\pi - 1 + \frac{\pi^2}{4} - \pi + \frac{\pi^2}{2}$$
$$= 2\sin\alpha + \frac{3\pi^2}{4} - \alpha\pi - \pi - 2\alpha\cos\alpha - 1,$$

这正是所要求的结果。

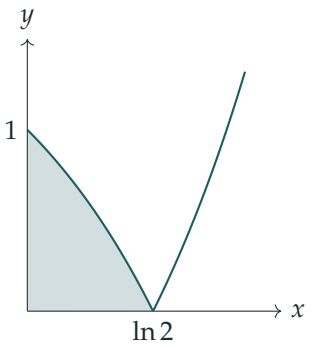
第 (ii) 部分

> **想一想** 两个绝对值符号之中，哪一个在这里才真正起作用?

在正 x 轴上，$x \geq 0$，所以 $\mathrm{e}^x \geq 1$，内层的绝对值符号并无影响：$|\mathrm{e}^x - 1| = \mathrm{e}^x - 1$。因此曲线就是

$$y = |\mathrm{e}^x - 2| \qquad (x \geq 0).$$

当 $0 \leq x \leq \ln 2$ 时，它等于 $2 - \mathrm{e}^x$，由 $x = 0$ 处的 $y = 1$ 下降至 $x = \ln 2$ 处的 $y = 0$；当 $x \geq \ln 2$ 时，它等于 $\mathrm{e}^x - 2$，无止境地上升。



所以曲线只在 $x = \ln 2$ 处触及 x 轴，而由正 x 轴、y 轴和曲线所围成的区域，就只有 $0 \leq x \leq \ln 2$ 上的阴影部分（过了 $\ln 2$ 之后，曲线一直向上攀升，再围不成任何封闭区域）。它的面积是

$$\int_0^{\ln 2} (2 - \mathrm{e}^x)\,\mathrm{d}x = \Big[2x - \mathrm{e}^x\Big]_0^{\ln 2} = (2\ln 2 - 2) - (0 - 1)$$
$$= 2\ln 2 - 1 = \ln 4 - 1,$$

正如所求。

验算

用二分法（bisection）可得 $\alpha \approx 2.4433$，代入 (i) 的公式后得出约 0.61343；把森逊法则（Simpson's rule）直接应用于 $[\pi/2, \pi]$ 上的 $|x\sin x - \pi/2|$，所得的值在小数点后 14 位

以内完全相同。在 $(0, \pi)$ 上作格点搜索（grid search），发现 $x \sin x - \pi/2$ 恰好变号两次，分别在 1.5708 和 2.4433 附近。在 (ii) 中，$\ln 4 - 1 \approx 0.38629$，同样与数值积分的结果吻合。作为对 (i) 中正负号的合理性检查，答案必须为正，而它确实为正。

常见错误

- 没有在 α 处分开，便直接在整个 $[\pi/2, \pi]$ 上对 $x \sin x - \pi/2$ 积分：这样得出的是有向面积（signed area）（这里是负数），而不是绝对值的积分。
- 遗漏了 $x = \pi/2$ 这个根，因为直线刚好通过 cosec 曲线的最低点，图形看起来可能像是相切。其实并非相切：曲线在该处的斜率是 0，而直线的斜率是 $2/\pi$，所以两者是相交而过的。
- 在 (ii) 中，把两个绝对值符号都保留下来，结果分出过多的情况。应先利用 $x \geq 0$ 确定 $\mathrm{e}^x - 1$ 的正负号。

STEP 之外

根 α 没有闭合形式（closed form），但我们仍然可以用它把积分准确地表达出来：这是分析学（analysis）中常见的情况，即一个量是由一条方程而非一条公式所确定的。严格凸函数与一条直线最多相交两次，这只是凸性众多精妙推论之一；凸性也是延森不等式（Jensen's inequality）以及最优化（optimisation）中许多理论的基础。Bartle 与 Sherbert 合著的 *Introduction to Real Analysis*（实分析入门）对连续性、介值定理（intermediate value theorem）和凸函数都有严谨细致的论述。

详细解答

STEP I 2006，第 8 题

纯数学 · 三维向量

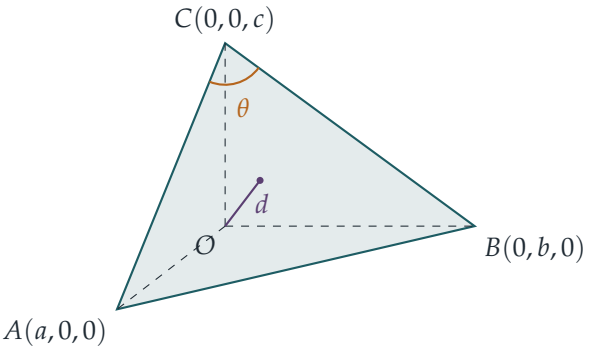
STEP I 2006，第 8 题

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/06-s1-q8

关键思路

四面体（tetrahedron）$OABC$ 在 O 处有三条互相垂直的棱，所以若以 OAB 为底，体积便很容易求出。同一个立体亦可以看成以 ABC 为底，这时的高正好就是我们要求的距离 d。用两种方法计算体积，便得到一条关于 d 的方程，前提是我们知道 ABC 的面积。这个面积可由 C 处的角求得，而该角则由纯量积（scalar product）给出。



解答

第 (i) 部分

取位于平面 $z = 0$ 上的三角形 OAB 为底。由于 OA 和 OB 分别沿 x 轴和 y 轴，角 AOB 是直角，所以

$$\text{面积}OAB = \tfrac{1}{2}ab.$$

高是 C 到平面 $z=0$ 的距离，即 c，因为 OC 沿 z 轴，故与该平面垂直。因此

$$V = \tfrac{1}{3} \times \tfrac{1}{2}ab \times c = \frac{abc}{6}.$$

第 (ii) 部分

角 ACB 是向量（vector）$\overrightarrow{CA}$ 与 $\overrightarrow{CB}$ 之间的夹角：

$$\overrightarrow{CA} = \begin{pmatrix} a \\ 0 \\ -c \end{pmatrix}, \qquad \overrightarrow{CB} = \begin{pmatrix} 0 \\ b \\ -c \end{pmatrix}.$$

它们的纯量积是 $0+0+c^2=c^2$，而长度分别是 $\sqrt{a^2+c^2}$ 和 $\sqrt{b^2+c^2}$。所以

$$\cos\theta = \frac{\overrightarrow{CA}\cdot\overrightarrow{CB}}{|\overrightarrow{CA}|\,|\overrightarrow{CB}|} = \frac{c^2}{\sqrt{(a^2+c^2)(b^2+c^2)}},$$

这就是所要求的结果。

要求面积，我们需要 $\sin\theta$。由于 $0<\theta<\pi$，$\sin\theta>0$，所以

$$\begin{aligned}\sin\theta = \sqrt{1-\cos^2\theta} &= \sqrt{\frac{(a^2+c^2)(b^2+c^2)-c^4}{(a^2+c^2)(b^2+c^2)}} \\ &= \frac{\sqrt{a^2b^2+b^2c^2+c^2a^2}}{\sqrt{(a^2+c^2)(b^2+c^2)}},\end{aligned}$$

因为展开 $(a^2+c^2)(b^2+c^2)$ 后，c^4 项互相抵消。

> ? **想一想** 哪一条三角形面积公式切合我们现有的资料?

我们知道两条边 CA 和 CB，以及它们之间的夹角 θ，所以使用 $\frac{1}{2}CA\cdot CB\sin\theta$。那些麻烦的平方根刚好约去：

$$\text{面积}ABC = \tfrac{1}{2}\sqrt{(a^2+c^2)(b^2+c^2)}\cdot\frac{\sqrt{a^2b^2+b^2c^2+c^2a^2}}{\sqrt{(a^2+c^2)(b^2+c^2)}} = \tfrac{1}{2}\sqrt{a^2b^2+b^2c^2+c^2a^2}.$$

令人欣喜的是，答案对 a、b、c 是对称的，而这也是必然的：三角形 ABC 并不在乎我们把哪一个顶点称为 C。

距离 d

现在把 ABC 看成同一个四面体的底。它的高是对面的顶点 O 到 ABC 所在平面的垂直距离，即 d。由于体积与我们选哪一个面作底无关，由第 (i) 部分和第 (ii) 部分可得

$$\frac{abc}{6} = \tfrac{1}{3}\times\tfrac{1}{2}\sqrt{a^2b^2+b^2c^2+c^2a^2}\times d, \qquad \text{所以} \qquad d = \frac{abc}{\sqrt{a^2b^2+b^2c^2+c^2a^2}}.$$

两边平方后取倒数（各项均为正数），

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

其中分子的每一项都除以 $a^2b^2c^2$。这就是所要求的结果。

验算

- 取 $a = b = c = 1$。这时 ABC 是边长为 $\sqrt{2}$ 的等边三角形，面积为 $\frac{\sqrt{3}}{4} \cdot 2 = \frac{\sqrt{3}}{2}$，与 $\frac{1}{2}\sqrt{3}$ 吻合。该平面是 $x + y + z = 1$，与 O 的距离为 $1/\sqrt{3}$，而确实有 $1/d^2 = 3$。
- 取 $a = 1$、$b = c = 2$，由公式得体积为 $\frac{2}{3}$，面积为 $\sqrt{6} \approx 2.449$，以及 $d = 2/\sqrt{6} \approx 0.8165$，所以 $1/d^2 = \frac{3}{2} = 1 + \frac{1}{4} + \frac{1}{4}$。
- 数值验证：随机选取 10 000 组 (a, b, c)，体积（用纯量三重积（scalar triple product）计算）、面积（用向量积（cross product）计算）和 d（直接找出垂足的位置来计算）全部与公式吻合，误差在 10^{-12} 以内。

常见错误

- 把 $\overrightarrow{AC}$ 与 $\overrightarrow{CB}$ 配搭使用：方向混乱会令纯量积的正负号颠倒，得出的是 θ 的补角。
- 忘记了以 ABC 为底时，高是由 O 到该平面的垂线，而不是任何一条棱的长度。

STEP 之外

通过 A、B、C 的平面是 $\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$，其法向量（normal）为 $(\frac{1}{a}, \frac{1}{b}, \frac{1}{c})$，利用一般的点到平面距离公式，一行便可求出 d。面积的结果是三维的毕氏定理（Pythagoras' theorem），即德古阿定理（de Gua's theorem）：ABC 面积的平方，等于在 O 处三个直角面的面积平方之和，因为 $\frac{1}{4}(a^2b^2 + b^2c^2 + c^2a^2)$ 正好就是这个和。这两个想法都源自向量积和行列式（determinant），在大学第一门线性代数（linear algebra）课程的初段便会学到；Strang 的 *Introduction to Linear Algebra*（线性代数入门教科书）是认识它们的好地方。

详细解答

STEP I 2006，第 9 题

力学 · 连接质点

STEP I 2006，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/06-s1-q9

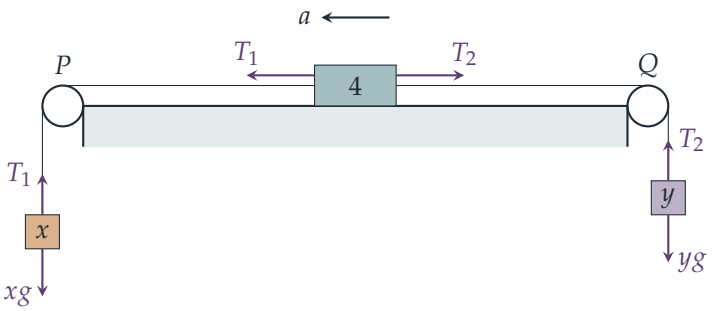
关键思路

整个运动分为两个阶段，每个阶段的加速度都是常数。两条绳子都连着的时候，三个物体组成的系统作为一个整体运动，由两个悬挂重物的重量之差驱动。绳子剪断之后，只有较轻的质点拉着方块，而且拉力与运动方向相反，所以方块作匀减速运动。剪断一刻的速率把两个阶段连接起来，而约束条件（constraint）$x + y = 6$ 让我们可以把一切都用 y 表示。

解答

模型

我们把方块和两个悬挂的物体都视为质点（particle），绳子视为轻而不可伸长（light and inextensible），桌面和滑轮（pulley）视为光滑（smooth）。由于 $x > y$，方块向 P 移动，质量为 x 的质点下降，质量为 y 的质点上升。设跨过 P 和 Q 的绳子的张力（tension）分别为 T_1 和 T_2。



第一阶段：两条绳子都连着

设共同的加速度为 a_1（方块向 P，质点 x 向下，质点 y 向上）。对每个物体应用牛顿第二定律（Newton's second law），得

$$\begin{aligned} xg - T_1 &= xa_1, \\ T_1 - T_2 &= 4a_1, \\ T_2 - yg &= ya_1. \end{aligned}$$

三式相加便消去张力：

$$(x-y)g = (x+y+4)a_1 = 10a_1, \qquad \text{所以} \qquad a_1 = \frac{(x-y)g}{10} = \frac{(6-2y)g}{10},$$

其中用了 $x = 6 - y$。这个值是正数，因为 $y < x$ 表示 $y < 3$。

方块由静止出发，以恒定加速度运动，在时间 t_1 内走过距离 d，并达到速率 V，其中

$$d = \tfrac{1}{2}a_1 t_1^2, \qquad V^2 = 2a_1 d.$$

因此

$$t_1 = \sqrt{\frac{2d}{a_1}} = \sqrt{\frac{20d}{(6-2y)g}} = \sqrt{\frac{d}{5g}}\,\frac{10}{\sqrt{6-2y}}, \qquad V = \sqrt{\frac{(6-2y)gd}{5}}.$$

第二阶段：剪断之后

现在 $T_1 = 0$。方块仍然以速率 V 向 P 移动，质量为 y 的质点也仍然以速率 V 上升。只有跨过 Q 的绳子作用于方块，并把它拉回 Q 的方向。设减速度（deceleration）为 a_2。于是

$$\begin{aligned} T_2 &= 4a_2 \quad \text{（方块）}, \\ yg - T_2 &= ya_2 \quad \text{（质点 } y\text{，上升时减速）}, \end{aligned}$$

所以

$$a_2 = \frac{yg}{4+y}.$$

? **想一想** 我们怎样知道跨过 Q 的绳子在剪断之后仍然保持拉紧?

假如绳子松弛了，质点便会以 g 减速，而方块则会保持原有速率。绳子拉紧时，张力为 $T_2 = 4a_2 = 4yg/(4+y) > 0$，并无矛盾，而且共同的减速度 a_2 小于 g。所以绳子确实保持拉紧，方块和质点一起运动，直至停下为止。

方块再经过以下时间便会静止：

$$t_2 = \frac{V}{a_2} = \sqrt{\frac{(6-2y)gd}{5}} \cdot \frac{4+y}{yg}$$
$$= \sqrt{\frac{d}{5g}}\sqrt{6-2y}\left(1+\frac{4}{y}\right),$$

其中我们用了 $\sqrt{(6-2y)gd/5}\,/\,g = \sqrt{d/(5g)}\sqrt{6-2y}$ 和 $(4+y)/y = 1+4/y$。

总时间

方块在第二阶段结束时首次回复静止，所以总时间为

$$t_1 + t_2 = \sqrt{\frac{d}{5g}}\left[\frac{10}{\sqrt{6-2y}} + \left(1+\frac{4}{y}\right)\sqrt{6-2y}\right] = \sqrt{\frac{d}{5g}}\,\mathrm{f}(y),$$

正是所求。

f 的驻值

符合物理情况的范围是 $0 < y < 3$。设 $u = 6-2y$，则 $\mathrm{d}u/\mathrm{d}y = -2$，而

$$\mathrm{f}(y) = 10u^{-1/2} + \left(1+\frac{4}{y}\right)u^{1/2}.$$

求导，并对第二项使用积法则（product rule），得

$$\mathrm{f}'(y) = 10\cdot\left(-\tfrac{1}{2}\right)u^{-3/2}(-2) - \frac{4}{y^2}u^{1/2} + \left(1+\frac{4}{y}\right)\tfrac{1}{2}u^{-1/2}(-2)$$
$$= \frac{10}{u^{3/2}} - \frac{4u^{1/2}}{y^2} - \frac{y+4}{y\,u^{1/2}}.$$

两边乘以 $y^2u^{3/2}$（在 $0<y<3$ 时为正数），条件 $\mathrm{f}'(y)=0$ 便化为

$$10y^2 - 4u^2 - (y^2+4y)u = 0.$$

代入 $u = 6-2y$：

$$10y^2 - 4(36-24y+4y^2) - (6y^2 - 2y^3 + 24y - 8y^2) = 0$$
$$2y^3 - 4y^2 + 72y - 144 = 0$$
$$2(y-2)(y^2+36) = 0.$$

由于 $y^2+36>0$，唯一的解是

$$y = 2,$$

它位于 $0<y<3$ 之内（并给出 $x=4$）。

因为当 $y\to 0^+$ 和 $y\to 3^-$ 时 $\mathrm{f}(y)\to\infty$，这个唯一的驻点（stationary point）就是极小值：$\mathrm{f}(2) = 10/\sqrt{2} + 3\sqrt{2} = 8\sqrt{2}$，所以回复静止所需的最短时间是 $8\sqrt{2d/(5g)}$。

验算

- 以数值方法逐步模拟运动（时间步长 $2\times10^{-6}\,\mathrm{s}$），随机选取 $y\in(0,3)$ 和 d，所得结果与 $\sqrt{d/(5g)}\,\mathrm{f}(y)$ 吻合至五位有效数字；例如取 $y=1.671$、$d=0.581$，并用 $g=9.8$，两种方法都得出 $1.27096\,\mathrm{s}$。
- f 的数值导数在 $(0,3)$ 上恰好变号一次，就在 $y=2$，而 $\mathrm{f}(2)=11.3137\ldots=8\sqrt{2}$。
- 极限情况合乎情理：当 $y\to3$ 时，两边的质量几乎平衡，$a_1\to0$，第一阶段需时无限长；当 $y\to0$ 时，剪断之后没有任何东西令方块减速，所以第二阶段需时无限长。

常见错误

- 在第二阶段使用质量 $4+x+y$：绳子一经剪断，质量为 x 的质点便不再属于这个系统。
- 忘记质量为 y 的质点在剪断之后仍以速率 V 向上运动，因而把第二阶段当作由静止开始。
- 对 $\sqrt{6-2y}$ 求导时漏掉 $\mathrm{d}u/\mathrm{d}y=-2$。

STEP 之外

把运动分成若干阶段，再以速度的连续性把它们连接起来，正是分段定义的动力学（piecewise-defined dynamics）背后的基本想法；这个想法会在冲力（impulsive forces）以及控制问题中再次出现，后者研究的是某个力何时开启、何时关闭。问哪一个 y 值令时间最短，则是初尝力学中的最优化（optimisation），日后会发展成变分法（calculus of variations），最速降线（brachistochrone）就是经典例子。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对连接体和约束力有仔细的讨论。

详细解答

STEP I 2006，第 10 题

力学 · 抛体运动

STEP I 2006，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/06-s1-q10

关键思路

轨迹（trajectory）是一条抛物线（parabola），屋顶是一条直线，所以“刚好碰到”的意思就是这条直线是一条*切线*（tangent）：把其中一条方程代入另一条，会得到一条关于 x 的二次方程（quadratic），而且它有重根（repeated root）。至于正负号的问题，只要问这个重根位于何处便可解决。在最后一部分，两个质点只可能在 P 碰到屋顶的那一点相遇，所以 Q 必须恰好在 P 到达该点的那一刻到达；比较两者在该时刻的水平位置，便得出结果。

解答

轨迹

把 P 看成一个只受重力作用而自由运动的质点（particle），不计空气阻力。它的初速度在水平和垂直方向的分量都是 $V\cos 45^\circ = V/\sqrt{2}$，所以在时间 t

$$x = \frac{Vt}{\sqrt{2}}, \qquad y = \frac{Vt}{\sqrt{2}} - \tfrac{1}{2}gt^2.$$

利用 $t = \sqrt{2}\,x/V$ 消去时间，得 $\frac{1}{2}gt^2 = gx^2/V^2$，所以轨迹是

$$y = x - \frac{gx^2}{V^2}.$$

碰到屋顶

P 与屋顶相遇之处满足 $x - gx^2/V^2 = x\tan\alpha + b$，即

$$\frac{g}{V^2}\,x^2 + (\tan\alpha - 1)\,x + b = 0. \qquad (*)$$

相切（而不是穿过）表示 $(*)$ 有重根，所以它的判别式（discriminant）等于零：

$$(\tan\alpha - 1)^2 = \frac{4gb}{V^2} \quad \Longrightarrow \quad V(\tan\alpha - 1) = \pm 2\sqrt{bg}.$$

想一想 两个正负号都满足判别式的条件。怎样分辨该取哪一个？

$(*)$ 的重根是

$$x_0 = -\frac{(\tan\alpha - 1)V^2}{2g} = \frac{(1 - \tan\alpha)V^2}{2g}.$$

质点沿 x 递增的方向运动（因为 $x = Vt/\sqrt{2}$），所以它只可能在 $x_0 > 0$ 之处碰到屋顶。这就迫使 $\tan\alpha - 1 < 0$，而由于 $V > 0$，左边的 $V(\tan\alpha - 1)$ 是负数。因此

$$V(-1 + \tan\alpha) = -2\sqrt{bg},$$

这正是所要证明的。（若取正根，切点便会在 $x_0 < 0$ 之处，即位于抛物线向后延伸的部分，而 P 从不经过那里。）从几何上看，屋顶的斜度必须小于初始运动方向的斜度，即 $\alpha < 45°$；否则屋顶一直在直线 $y = x$ 之上，而抛物线在 $x > 0$ 时位于该直线之下。

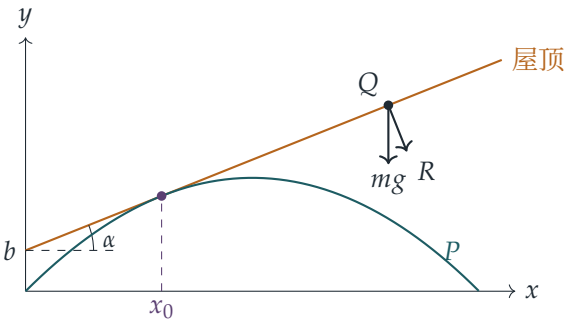
相切的时刻

由于 $x = Vt/\sqrt{2}$，相切的时刻是

$$T = \frac{\sqrt{2}\,x_0}{V} = \frac{\sqrt{2}\,(1 - \tan\alpha)V}{2g} = \frac{V(1 - \tan\alpha)}{\sqrt{2}\,g}.$$

路轨上的质点

把 Q 看成一个在光滑（smooth）路轨上的质点，路轨沿直线 $y = x\tan\alpha + b$；它从 $(0, b)$ 沿路轨向 x 递增的方向射出（必须如此，才能到达 P，因为后者的 x 永远是正数）。作用于 Q 的力有它的重量 mg，以及路轨对它的法向反作用力（normal reaction）R，后者垂直于路轨。下图显示了这些力，以及 P 的路径在 $x = x_0$ 处与屋顶相切的情况。



由于没有摩擦力，沿路轨向斜上方分解，得

$$m\ddot{s} = -mg\sin\alpha,$$

其中 s 是 Q 沿路轨与 $(0,b)$ 的距离。以 $s=0$ 及 $\dot{s}=U$ 作为 $t=0$ 时的初始条件，

$$s = Ut - \tfrac{1}{2}gt^2\sin\alpha, \qquad x_Q = s\cos\alpha.$$

P 除了在切点之外从不在屋顶上，而 Q 则一直在屋顶上，所以两个质点要相碰，唯一的可能是 Q 在时间 T 位于该点。那时两个质点都在屋顶所在的直线上，所以只需它们的 x 坐标在时间 T 相等：

$$\left(UT - \tfrac{1}{2}gT^2\sin\alpha\right)\cos\alpha = \frac{VT}{\sqrt{2}}.$$

两边除以 $T>0$，并代入 $gT = V(1-\tan\alpha)/\sqrt{2}$：

$$U\cos\alpha = \frac{V}{\sqrt{2}} + \tfrac{1}{2}\sin\alpha\cos\alpha\cdot\frac{V(1-\tan\alpha)}{\sqrt{2}}.$$

两边乘以 $2\sqrt{2}$，并利用 $\sin\alpha\cos\alpha\tan\alpha = \sin^2\alpha$，

$$\begin{aligned} 2\sqrt{2}\,U\cos\alpha &= 2V + V\sin\alpha\cos\alpha\,(1-\tan\alpha) \\ &= V\left(2 + \sin\alpha\cos\alpha - \sin^2\alpha\right), \end{aligned}$$

这就是所要求的结果。

验算

- 单位：$V(1-\tan\alpha)/g$ 是速率除以加速度，所以是时间；而 $\sqrt{bg}$ 是速率，与 $V(\tan\alpha-1)$ 一样。
- 极限情况 $\alpha\to 0$（屋顶水平）：$b = V^2/(4g)$ 正是大家熟悉的最大高度 $(V\sin 45^\circ)^2/(2g)$，在 $T = V/(\sqrt{2}g)$ 时到达；而 Q 沿路轨方向不受力，所以它的速率必须与 P 的水平速率相同，即 $U = V/\sqrt{2}$。公式给出的正是这个结果。
- 数值验算：取 $g=9.8$、$V=10$ 及 $\alpha = 0.3$ 弧度，切点位于 $x_0 \approx 3.524$。让 Q 以推导所得的 U 沿路轨逐步前进，对随机选取的 V 和 $\alpha < \pi/4$，它在时间 T 的位置与 P 的位置相差不超过 10^{-6}，而 P 在其他所有时刻都在屋顶之下。

常见错误

- 把 Q 的加速度写成 $-g$ 或 $-g\cos\alpha$：只有重力沿路轨方向的分量 $g\sin\alpha$ 才沿路轨起作用。

- 把 U 当作水平速率：U 是沿路轨的速率，所以 Q 的水平位移是 $s\cos\alpha$。
- 只用“速率是正数”来解释为何取负根。真正起决定作用的是 x_0 的正负号（等价于 $\tan\alpha < 1$）。

STEP 之外

相切条件是一个普遍想法的特殊情况：一族曲线所能到达的区域，其边界就是这族曲线的包络线（envelope），而求包络线，归根究底就是问一条二次方程何时有重根。对于以固定速率向各个方向抛出的抛体（projectile），包络线就是所谓的“安全抛物线”（parabola of safety）。Q 的运动是最简单的约束运动（constrained motion）；大学一年级的课程会用拉格朗日力学（Lagrangian mechanics）有系统地处理约束。Kleppner and Kolenkow, *An Introduction to Mechanics*（力学入门教科书）对抛体和约束运动都有细致的讲解。

详细解答

STEP I 2006，第 11 题

力学 · 弹性碰撞

STEP I 2006，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/06-s1-q11

关键思路

我们无法逐次追踪每一次碰撞：我们连 $A_1, \ldots, A_{n-2}$ 的质量也不知道。因此我们只考虑最终状态。有两项事实限制着最终状态。第一，动量（momentum）和动能（kinetic energy）在结束时与开始时相同。第二，所有碰撞结束之后，没有任何粒子能追上它前面的粒子，所以最终速度沿着槽（弱）递增。于是每一部分都只需一段简短的论证，说明这两项事实容许哪些最终状态。

解答

建立模型

我们把各粒子视为在一条直线上无摩擦地运动的质点，而每次碰撞都是完全弹性（perfectly elastic）的。取 A_0 初始运动的方向为正方向，并设 v_k 为 A_k 的最终速度（$k = 0, 1, \ldots, n$）。以 m_k 表示 A_k 的质量，所以 $m_0 = m_{n-1} = m$ 而 $m_n = \lambda m$。由守恒（conservation）定律得

$$\sum_{k=0}^{n} m_k v_k = mu, \tag{M}$$

$$\sum_{k=0}^{n} m_k v_k^2 = mu^2. \tag{E}$$

粒子在槽内不能互相越过，所以它们一直保持 $A_0, A_1, \ldots, A_n$ 的次序。假如最终对某个 k 有 $v_k > v_{k+1}$，那么 A_k 便会追上 A_{k+1}，于是还会再有一次碰撞。因此在所有碰撞结束之后

$$v_0 \leq v_1 \leq \cdots \leq v_n. \tag{O}$$

特别地，向前运动的粒子前面不能有静止的粒子，而向后运动的粒子后面也不能有静止的粒子。

第 (i) 部分

假设最终只有 A_j 在运动，速度为 v。那么由 (M) 和 (E) 得 $m_j v = mu$ 及 $m_j v^2 = mu^2$。两式相除，得 $v = u$，从而 $m_j = m$。

由于 $v = u > 0$，次序条件 (O) 要求 A_j 前面每一个粒子的速度都至少为 u。但它们全部静止，所以前面根本没有粒子，即 $A_j = A_n$。然而 A_n 的质量是 $\lambda m \neq m$，因为 $\lambda > 1$。这是一个矛盾（contradiction），所以最终不可能恰好只有一个粒子在运动。

第 (ii) 部分

假设只有 A_{n-1} 和 A_n 在运动，速度分别为 p 和 q。把 (M) 和 (E) 除以 m，

$$p + \lambda q = u, \qquad p^2 + \lambda q^2 = u^2.$$

把 $p = u - \lambda q$ 代入第二条方程，得 $-2\lambda uq + \lambda^2 q^2 + \lambda q^2 = 0$，即

$$\lambda q\big((1+\lambda)q - 2u\big) = 0.$$

因为 A_n 在运动，我们需要 $q \neq 0$，所以

$$q = \frac{2u}{1+\lambda}, \qquad p = u - \frac{2\lambda u}{1+\lambda} = -\frac{(\lambda-1)u}{1+\lambda} < 0.$$

所以 A_{n-1} 会向后运动。但 $n \geq 2$，所以它后面有一个粒子 A_{n-2}，而按假设该粒子是静止的。于是 $v_{n-2} = 0 > p = v_{n-1}$，这与 (O) 矛盾：$A_{n-1}$ 会撞上 A_{n-2}。所以这个最终状态同样不可能出现。

第 (iii) 部分

这里的主要工具是以下事实。

引理（lemma）。在运动结束时，必有某个粒子向后运动。

证明。假设相反，即对每个 k 都有 $v_k \geq 0$。由 (O)，v_n 是最大的速度。(E) 中的每一项都满足 $m_k v_k^2 \leq v_n m_k v_k$，因为 $0 \leq v_k \leq v_n$ 而 $m_k > 0$。把这些不等式相加并利用 (M)，

$$mu^2 = \sum m_k v_k^2 \leq v_n \sum m_k v_k = v_n\, mu, \qquad \text{所以} \qquad v_n \geq u.$$

另一方面，(M) 中的每一项都是非负的，所以单独一项 $\lambda m v_n$ 不会超过总和：

$$\lambda m v_n \leq mu, \qquad \text{所以} \qquad v_n \leq \frac{u}{\lambda} < u,$$

这里用到 $\lambda > 1$ 和 $u > 0$。这两个界互相矛盾。 □

想一想 题目从来没有给出 $A_1, \ldots, A_{n-2}$ 的质量。为甚么这并不要紧?

引理只用到 (M)、(E)、(O) 以及 A_n 的质量，所以无论其他质量是多少，它都成立。正因如此，这道题才有可能解答。

现在假设最终只有 A_{n-2}、A_{n-1} 和 A_n 在运动。

若 $n \geq 3$，则 A_{n-2} 后面有一个静止的粒子 A_{n-3}。由 (O)，$v_{n-2} \geq v_{n-3} = 0$，从而 $v_{n-1}, v_n \geq v_{n-2} \geq 0$。每个粒子的速度都 ≥ 0，而这是引理所不容许的。

若 $n = 2$，这三个粒子就是 A_0、A_1、A_2，亦即全部粒子，而整个运动是完全确定的。当 A_0 撞上 A_1 时（两者质量相等），速度互相交换：A_0 停下，而 A_1 以速率 u 继续前进。然后 A_1 撞上 A_2，按第 (ii) 部分的计算（碰撞前的速度为 u 和 0），它以速度 $-\frac{(\lambda-1)u}{\lambda+1}$ 反弹，而 A_2 则以 $\frac{2u}{\lambda+1}$ 离开。最后 A_1 撞上静止的 A_0，质量相等的两者再次交换速度。此后，A_0 向后运动，A_1 静止，A_2 向前运动，所以 (O) 成立，不会再有碰撞。A_1 最终静止，所以三个粒子全部都在运动同样是不可能的。

第 (iv) 部分

假设最终恰好有两个粒子在运动。由 (M)，总动量为 $mu > 0$，所以其中至少一个向前运动。由引理，至少一个向后运动。所以其中一个，设为 A_i，有 $v_i < 0$，而另一个 A_j 有 $v_j > 0$。

想一想 这两个粒子可以是哪两个?

由 (O)，A_i 后面每一个粒子的速度都 $\leq v_i < 0$，所以它在运动。但另外在运动的只有 A_j，而它是向前运动的，所以 A_i 后面没有任何粒子：$A_i = A_0$。同样地，A_j 前面每一个粒子的速度都 $\geq v_j > 0$，所以前面没有粒子，即 $A_j = A_n$。

所以在运动的是 A_0（质量 m）和 A_n（质量 λm），速度分别为 $a < 0$ 和 $b > 0$。方程 (M) 和 (E) 成为

$$a + \lambda b = u, \qquad a^2 + \lambda b^2 = u^2,$$

与第 (ii) 部分的方程完全一样。由于 $b \neq 0$，

$$b = \frac{2u}{\lambda+1}, \qquad a = -\frac{(\lambda-1)u}{\lambda+1}.$$

因此，在运动的两个粒子是 A_0，以速率 $\dfrac{(\lambda-1)u}{\lambda+1}$ 向后运动，以及 A_n，以速率 $\dfrac{2u}{\lambda+1}$ 向前运动。

这种情况确实可以发生：第 (iii) 部分中 $n = 2$ 的情形正是以这个状态告终。

验算

- 取 $\lambda = 3$ 和 $u = 4$，两个速率是 2 和 2。于是动量为 $m(-2) + 3m(2) = 4m$，能量为 $\frac{1}{2}m(4) + \frac{1}{2}(3m)(4) = 8m = \frac{1}{2}m(16)$，与应有的结果相符。
- 当 $\lambda \to 1$ 时，A_0 的最终速率趋向 0，而 A_n 的最终速率趋向 u：当所有质量都相等时，速度只是沿着整列粒子逐一传递下去，就像牛顿摆（Newton's cradle）一样。
- 我们对 60,000 个随机设置模拟了整个碰撞过程（n 由 2 至 6，$A_1, \ldots, A_{n-2}$ 的质量随机，λ 和间距亦随机）。A_0 最终总是向后运动，(i)–(iii) 中的最终状态从未出现，而每当恰好有两个粒子在运动时，它们都是 A_0 和 A_n，速率亦与上文所得相同。

常见错误

- 试图追踪每一次碰撞。中间各粒子的质量是未知的，所以只能分析最终状态。
- 忘记次序条件 (O)。单凭守恒定律，第 (ii) 部分的状态是容许的；把它排除的正是次序条件。
- 假定 A_{n-2} 不是 A_0。当 $n = 2$ 时它正是，而第 (iii) 部分需要另作一个（简短的）论证。

STEP 之外

这个引理是两个守恒量之间的一个不等式：当所有速度都非负时，能量不能超过（最大速度）×（动量）。这类论证利用守恒量和次序关系来排除某些最终状态，而无须解出整个动力学过程，在多粒子系统的研究中十分典型。下一步不妨阅读 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论），书中对碰撞、质心参考系（centre-of-mass frame）和守恒定律都有细致的处理。

详细解答

STEP I 2006，第 12 题

概率与统计 · 条件概率

STEP I 2006，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/06-s1-q12

关键思路

一条道路要畅通，必须路上每一段都畅通；而两镇要被隔绝，则必须每一条道路都受阻。因此我们逐条道路处理，利用独立性（independence），并且每一步都借助补事件（complement）来计算。第二部分是一个条件概率（conditional probability）问题，当中选了哪条路是有影响的：只有一段的道路比有两段的道路更可能畅通，所以发现自己所选的路畅通，其实透露了你选的是哪条路的一些信息。我们用定义 $\mathrm{P}(B \mid A) = \mathrm{P}(A \cap B)/\mathrm{P}(A)$ 来处理，并按所选的道路分情况计算。

解答

被隔绝的概率

把两条有两段的道路称为 R_1 和 R_2，只有一段的道路称为 R_3。设 $q = 1 - p$ 为单独一段畅通的概率。

一条有两段的道路畅通，当且仅当它的两段都畅通；由独立性，其概率为 $q^2 = (1-p)^2$。所以

$$\mathrm{P}(R_1 \text{ 受阻}) = \mathrm{P}(R_2 \text{ 受阻}) = 1 - (1-p)^2 = p(2-p),$$
$$\mathrm{P}(R_3 \text{ 受阻}) = p.$$

三条道路共用到五个不同的路段，所以“R_1 受阻”、“R_2 受阻”和“R_3 受阻”这三个事件各自取决于互不相交的几组独立路段，因而互相独立。Oxtown 与 Camville 被隔绝，当且仅当三条道路全部受阻，所以

$$\mathrm{P}(\text{被隔绝}) = p(2-p) \cdot p(2-p) \cdot p = p^3(2-p)^2,$$

正如所求。

条件概率

设 C_k 为我选择道路 R_k 的事件，则对每个 k 都有 $\mathrm{P}(C_k) = \frac{1}{3}$，而且我的选择与水浸情况独立。设 O 为我所选的路畅通的事件，B 为另外两条路都受阻的事件。我们要求的是 $\mathrm{P}(B \mid O) = \mathrm{P}(B \cap O)/\mathrm{P}(O)$。

> **想一想** 根据独立性，答案不就是 P(另外两条路受阻) 吗?

假如我们知道自己选了哪条路，答案确是如此。但“另外两条路”是哪两条，要视乎我们的选择而定；而且观察到自己的路畅通之后，选中较短的 R_3 的可能性就比原先高了。所以我们必须按所选的道路分情况，并配上正确的权重，而下面的公式做的正是这件事。

分母。对所选的道路运用全概率公式（law of total probability），

$$\begin{aligned}\mathrm{P}(O) &= \tfrac{1}{3}(1-p)^2 + \tfrac{1}{3}(1-p)^2 + \tfrac{1}{3}(1-p) \\ &= \tfrac{1}{3}(1-p)\big[2(1-p)+1\big] = \tfrac{1}{3}(1-p)(3-2p).\end{aligned}$$

分子。如果我选了 R_1，就需要 R_1 畅通而 R_2、R_3 受阻；由各条道路的独立性，其概率为 $(1-p)^2 \cdot p(2-p) \cdot p$。选 R_2 的结果相同。如果我选了 R_3，就需要 R_3 畅通而两条有两段的道路都受阻：$(1-p) \cdot p^2(2-p)^2$。因此

$$\begin{aligned}\mathrm{P}(B \cap O) &= \tfrac{1}{3}\Big[2p^2(2-p)(1-p)^2 + p^2(2-p)^2(1-p)\Big] \\ &= \tfrac{1}{3}\,p^2(1-p)(2-p)\big[2(1-p)+(2-p)\big] \\ &= \tfrac{1}{3}\,p^2(1-p)(2-p)(4-3p).\end{aligned}$$

答案。两式相除，并约去因子 $\frac{1}{3}(1-p)$（当 $p < 1$ 时它不等于零），得

$$\mathrm{P}(B \mid O) = \frac{p^2(2-p)(4-3p)}{3-2p}.$$

因子 p、$2-p$、$4-3p$ 和 $3-2p$ 全都是一次式，所以这已经是完全因式分解的形式。

极限 $p \to 1$

当 $p \to 1$ 时，每个因子都趋向 1，所以

$$\mathrm{P}(B \mid O) \to \frac{1 \cdot 1 \cdot 1}{1} = 1.$$

这是合理的。当水浸几乎必然发生时，我自己的路畅通纯属侥幸；而其他道路是否水浸与我的路独立，它们几乎肯定没有同样的运气：知道我的路畅通，并不能说明它们是否畅通。所以我应该几乎肯定，换了任何其他选择都会行不通。（在 $p = 1$ 本身，作为条件的事件概率为零，条件概率没有定义；只有极限才有意义。）

验算

- 当 $p = 0$ 时，没有任何路段会受阻，而公式给出 0，正该如此。
- 当 $p = \frac{1}{2}$ 时，公式给出 $\dfrac{\frac{1}{4} \cdot \frac{3}{2} \cdot \frac{5}{2}}{2} = \frac{15}{32}$。把路段受阻的全部 2^5 种情况和三种道路选择逐一精确列举，得到的数值相同；对其他几个 p 值也是如此，例如 $p = \frac{2}{3}$ 时得 $\frac{32}{45}$。
- 对 $0 \leq p < 1$，答案总是介乎 0 与 1 之间：例如当 $p = 0.99$ 时，它约为 0.9996。

常见错误

- 把有两段的道路受阻的概率当作 p^2（两段都受阻），而不是 $1 - (1 - p)^2$（至少一段受阻）。
- 忘记取条件：另外两条路受阻的概率，与已知我随机选中的路畅通时它的条件概率，并不相同。
- 在观察之后仍把三个选择看成权重相等；相等的是先验概率（prior probability）$\frac{1}{3}$，而不是后验概率（posterior probability）。

STEP 之外

多条道路并联、每条道路又由若干路段串联而成，是可靠性网络（reliability network）最简单的例子；一个网络能把两点连通的概率，其一般情况是渗流理论（percolation theory）的研究对象，而当 p 变化时网络行为出现的急剧转变，正是该理论的核心课题。第二部分其实是乔装了的贝叶斯定理（Bayes' theorem）：“我的路畅通”这项观察，更新了选中各条道路的概率。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）正是以这种思路讲解条件概率和独立性。

详细解答

STEP I 2006，第 13 题

概率与统计 · 泊松混合分布

STEP I 2006，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/06-s1-q13

关键思路

钻石以固定的平均率独立地散布，这就是一个泊松过程（Poisson process），因此 m 公斤朱古力中的钻石数目服从平均值为 m 的泊松分布（Poisson distribution）。朱古力的份量本身也是随机的，所以我们以它为条件：先就固定的份量求出答案，再利用全概率定理（law of total probability）（以及它在期望值方面的对应版本）对掷骰结果取平均。这样每一部分都归结为级数求和：(i) 是有限级数，(ii) 是几何级数（geometric series）。

解答

设 D 为我找到的钻石数目。由于钻石是以每公斤平均一颗的比率独立地藏在朱古力中，质量为 m kg 的一排朱古力所含的钻石数目服从平均值为 m 的泊松分布。特别地，已知质量为 m kg 时，

$$\mathrm{P}(D=0)=\mathrm{e}^{-m}, \qquad \mathrm{E}(D)=m.$$

以下一律把 $100k$ 克写成 $0.1k$ 公斤。

第 (i) 部分

设 N 为骰子掷出的点数，则对 $n=1,\ldots,6$ 有 $\mathrm{P}(N=n)=\frac{1}{6}$。已知 $N=n$ 时，我买 $0.1n$ kg，所以 $D \mid N=n$ 服从平均值为 $0.1n$ 的泊松分布。由全概率定理，

$$\begin{aligned}\mathrm{P}(D=0) &= \sum_{n=1}^{6} \mathrm{P}(N=n)\,\mathrm{P}(D=0 \mid N=n) = \frac{1}{6}\sum_{n=1}^{6} \mathrm{e}^{-0.1n} \\ &= \frac{1}{6}\left(\mathrm{e}^{-0.1}+\mathrm{e}^{-0.2}+\cdots+\mathrm{e}^{-0.6}\right).\end{aligned}$$

这是一个几何级数，首项为 $\mathrm{e}^{-0.1}$，公比（common ratio）为 $\mathrm{e}^{-0.1}$，共有六项，所以

$$\mathrm{P}(D=0)=\frac{\mathrm{e}^{-0.1}}{6}\left(\frac{1-\mathrm{e}^{-0.6}}{1-\mathrm{e}^{-0.1}}\right),$$

正是所求。

至于期望值（expectation），我们用同样的方法取条件：

$$\mathrm{E}(D)=\sum_{n=1}^{6}\mathrm{P}(N=n)\,\mathrm{E}(D\mid N=n)=\frac{1}{6}\sum_{n=1}^{6}0.1n=\frac{0.1\times 21}{6}=0.35.$$

换言之，$\mathrm{E}(D)=0.1\,\mathrm{E}(N)=0.1\times 3.5$，这正是我们要证明的结果。

第 (ii) 部分

现在设 T 为首次掷出 6 点时的掷骰次数。各次掷骰互相独立，而 $T=t$ 表示先掷出 $t-1$ 次非 6 点，然后掷出一次 6 点，所以

$$\mathrm{P}(T=t)=\left(\tfrac{5}{6}\right)^{t-1}\tfrac{1}{6},\qquad t=1,2,3,\ldots$$

已知 $T=t$ 时，我买 $0.1t$ kg，所以 $\mathrm{P}(D=0\mid T=t)=\mathrm{e}^{-0.1t}$。因此

$$\mathrm{P}(D=0)=\sum_{t=1}^{\infty}\left(\tfrac{5}{6}\right)^{t-1}\tfrac{1}{6}\,\mathrm{e}^{-0.1t}=\frac{\mathrm{e}^{-0.1}}{6}\sum_{t=1}^{\infty}\left(\tfrac{5}{6}\mathrm{e}^{-0.1}\right)^{t-1}.$$

? **想一想** 为甚么我们可以把这个级数求和至无穷?

这是一个公比为 $r=\frac{5}{6}\mathrm{e}^{-0.1}$ 的几何级数，而 $0<r<1$，所以它收敛（converge），其和为 $1/(1-r)$。于是

$$\mathrm{P}(D=0)=\frac{\mathrm{e}^{-0.1}}{6}\cdot\frac{1}{1-\frac{5}{6}\mathrm{e}^{-0.1}}=\frac{\mathrm{e}^{-0.1}}{6-5\mathrm{e}^{-0.1}},$$

其中分子和分母同乘以 6。这就是所求的结果。

至于钻石数目的期望值，

$$\mathrm{E}(D)=\sum_{t=1}^{\infty}\mathrm{P}(T=t)\,\mathrm{E}(D\mid T=t)=\sum_{t=1}^{\infty}\left(\tfrac{5}{6}\right)^{t-1}\tfrac{1}{6}\times 0.1t=\frac{0.1}{6}\sum_{t=1}^{\infty}t\left(\tfrac{5}{6}\right)^{t-1}.$$

题目的提示在这里派上用场。当 $|x|<1$ 时，由二项展开式（binomial expansion）可得

$$(1-x)^{-2}=1+2x+3x^2+4x^3+\cdots=\sum_{t=1}^{\infty}tx^{t-1}.$$

取 $x=\frac{5}{6}$，这个和就是 $\left(1-\frac{5}{6}\right)^{-2}=36$，所以

$$\mathrm{E}(D)=\frac{0.1}{6}\times 36=0.6.$$

这同样等于 $0.1\,\mathrm{E}(T)$，其中 $\mathrm{E}(T)=6$ 就是大家熟悉的、掷出一次 6 点所需掷骰次数的平均值。

验算

- 按数值计算，(i) 的概率约为 0.7150，(ii) 的概率约为 0.6131。以泊松分布产生钻石数目，模拟 2×10^5 位顾客，所得的结果为 0.7151 和 0.6136，而钻石数目的平均值则为 0.350 和 0.599。
- (ii) 的概率比 (i) 的小。这是合理的：我平均买 600 g，而不是 350 g。
- 两个期望值都是那个随机乘数的平均值的 0.1 倍。钻石数目的期望值等于公斤数的期望值，因为已知质量时，钻石数目的期望值是质量的线性函数。

常见错误

- 以 $\mathrm{e}^{-0.35}$（或 $\mathrm{e}^{-0.6}$）作为没有钻石的概率。这样只用到平均质量，但 e^{-m} 并不是 m 的线性函数，所以必须先取 e^{-m} 再求平均，而不是先求平均。期望值是线性的，所以这条捷径对 $\mathrm{E}(D)$ 确实行得通。
- 忘记把克换算成公斤，结果得到平均值 $100n$ 而不是 $0.1n$。
- 写成 $\mathrm{P}(T = t) = (\frac{5}{6})^t \frac{1}{6}$，这样便多算了一次非 6 点。

STEP 之外

这里泊松分布的参数本身是随机的，这使 D 成为一个混合（mixture）分布。所用的工具，即全概率定理和重期望定律（tower law）$\mathrm{E}(D) = \mathrm{E}\big(\mathrm{E}(D \mid T)\big)$，都是大学一年级概率论的核心内容。概率母函数（probability generating function）可以俐落地处理这类问题：$\mathrm{P}(D = 0)$ 就是 T 的母函数在 $\mathrm{e}^{-0.1}$ 处的值。先定出随机的试验次数，再在每次试验中取一个随机的数目，由此便可进一步引出复合分布（compound distribution）和分支过程（branching process）。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）对以上各项都有详细阐述。

详细解答

STEP I 2006，第 14 题

概率与统计 · 不放回抽样

STEP I 2006，第 14 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/06-s1-q14

关键思路

在第 r 次才取到红色糖果，即是先取到 $r-1$ 颗蓝色糖果，然后取到红色那一颗。若每次抽取后放回（with replacement），每次抽取的情况都相同，所以概率只是几个幂的简单乘积；之后把 n 当作连续变量，对其对数求导，便可求出使概率最大的 n。若抽取后不放回（without replacement），最简洁的看法是想像把所有糖果都抽出来：红色那一颗落在任何一个位置的可能性都相等。

解答

以下 $n \geq 0$ 均为整数，而 R 表示首次取到红色糖果时是第几次抽取。

第 (i) 部分

由于糖果会放回袋中并摇匀，各次抽取互相独立（independent），而每次抽取时，取到红色糖果的概率是 $\frac{1}{n+1}$，取到蓝色糖果的概率是 $\frac{n}{n+1}$。事件 $R=r$ 表示首 $r-1$ 次抽到蓝色，而第 r 次抽到红色，所以根据独立性，

$$P(R=r)=\left(\frac{n}{n+1}\right)^{r-1}\frac{1}{n+1}=\frac{n^{r-1}}{(n+1)^r},\qquad r=1,2,3,\ldots$$

（这就是成功概率为 $\frac{1}{n+1}$ 的几何分布（geometric distribution）。）

现在固定 r，把上式看成 n 的函数：

$$f(n)=\frac{n^{r-1}}{(n+1)^r}.$$

若 $r=1$，则 $f(n)=\frac{1}{n+1}$，它在 $n=0$ 时最大（没有蓝色糖果时，第一颗必定是红色）。

因此假设 $r \geq 2$。这时 $f(0) = 0$，而当 $n > 0$ 时 $f(n) > 0$，所以我们可以在 $n > 0$ 上取对数，并把 n 当作实变量：

$$\ln f(n) = (r-1)\ln n - r\ln(n+1), \qquad \frac{\mathrm{d}}{\mathrm{d}n}\ln f(n) = \frac{r-1}{n} - \frac{r}{n+1} = \frac{(r-1)-n}{n(n+1)}.$$

想一想 n 必须是整数。那么求导是否合理?

合理，因为我们只是利用导数（derivative）来找出 f 在哪里递增、在哪里递减。当 $n > 0$ 时，分母 $n(n+1)$ 是正数，所以导数在 $0 < n < r-1$ 时为正，在 $n > r-1$ 时为负。因此 f 在 $[0, r-1]$ 上严格递增（strictly increasing），在 $[r-1, \infty)$ 上严格递减。由于 $r-1$ 本身是整数，它胜过其他所有整数，无须再比较相邻的整数。

所以在所有情况下（包括 $r = 1$），使概率最大的是

$$n = r - 1,$$

最大值为 $\dfrac{(r-1)^{r-1}}{r^r}$（当 $r = 1$ 时理解为 1）。例如，当 $r = 3$ 时，最理想的袋中有 $n = 2$ 颗蓝色糖果，所得概率为 $\frac{4}{27}$。

第 (ii) 部分

现在各次抽取不再独立，但有两个简单的做法。

直接相乘。要 $R = r$，首 $r-1$ 次每次都要抽到蓝色，然后抽到红色。这只有在蓝色糖果足够多时才可能发生，即 $r - 1 \leq n$。已知之前抽到的全是蓝色，第 k 次抽取时袋中有 $n + 2 - k$ 颗糖果，其中一颗是红色。所以

$$P(R = r) = \frac{n}{n+1} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{n-r+2}{n-r+3} \cdot \frac{1}{n-r+2}.$$

每个分子都与下一个分母约去，乘积逐项相消（telescope），得

$$P(R = r) = \begin{cases} \dfrac{1}{n+1}, & 1 \leq r \leq n+1, \\ 0, & r > n+1. \end{cases}$$

对称性。抽到红色糖果便停止，并不影响它所在的位置。所以不妨想像一直抽下去，直至袋中空无一物。这些（可区分的）糖果共有 $(n+1)!$ 种排列次序，每种的可能性都相等，因此红色糖果落在 $n+1$ 个位置中任何一个的可能性都相等。这样得出相同的答案。

至于求最大值，先固定 r。若 $n < r-1$，概率为 0。若 $n \geq r-1$，概率为 $\frac{1}{n+1}$，它随 n 增大而严格递减。所以使概率最大的，是符合条件的最小的袋，

$$n = r - 1,$$

而最大概率为 $\frac{1}{r}$。（当 $n = r-1$ 时，红色糖果是 r 颗糖果中任何一颗的可能性都相等，而最后才抽到它，正是 r 个等可能结果的其中一个。）

验算

- 在 (i) 中，$\sum_{r\geq 1} P(R=r) = \frac{1}{n+1}\cdot\dfrac{1}{1-n/(n+1)} = 1$；在 (ii) 中，$n+1$ 个相等的值 $\frac{1}{n+1}$ 之和为 1。
- 在 (i) 中取 $r=3$：$n=1,2,3$ 分别给出 $\frac{1}{8}$、$\frac{4}{27}\approx 0.148$ 和 $\frac{9}{64}\approx 0.141$，所以 $n=2$ 胜出。
- 我们就 $n\leq 6$ 列出糖果的所有排列次序，验证了 (ii)。就 (i) 取 $n=3$、$r=2$ 进行模拟，得出 0.188，而准确值为 $\frac{3}{16}=0.1875$。

常见错误

- 在 (i) 中写成 $\left(\frac{n}{n+1}\right)^r\frac{1}{n+1}$：在红色之前只有 $r-1$ 颗蓝色。
- 在 (ii) 中，忘记了当 $r>n+1$ 时答案是 0。正是这一点，令“把 n 取得愈小愈好”不致一直滑落到 $n=0$。

STEP 之外

这两个求最大值的问题，其实都是*最大似然估计*（maximum likelihood estimation）。假如你看到红色糖果在第 r 次首次出现，却不知道 n 是多少，那么在两个实验中，令你的观察结果最有可能出现的 n 值都是 $n=r-1$。这是大学一年级统计学的基石之一。(ii) 中的对称性论证是*可交换性*（exchangeability）的一个例子，它令许多不放回抽样的问题变得容易。这两个概念在 Blitzstein 与 Hwang 所著的 *Introduction to Probability*（概率论入门教科书）中都有详细阐述。

详细解答

STEP II 2006，第 1 题

纯数学 · 周期性递推关系

STEP II 2006，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/06-s2-q1

关键思路

由于每一项都决定了下一项，只要 $u_{p+1} = u_1 = 2$，数列（sequence）便会以周期（period）p 重复。因此 (i) 的每一部分，归根究底都是把头几项写成 k 的函数，然后解一条多项式方程。周期为 4 的方程必定包含周期 1 和周期 2 的答案作为因式（周期为 1 或 2 的数列同样满足 $u_5 = u_1$），事先知道这一点，因式分解（factorisation）便轻而易举。第 (ii) 部分是一行便可完成的归纳法（induction），之后再用惯常的“在递推关系（recurrence）中令 $n \to \infty$”的论证。

解答

第 (i) 部分

设 $f(u) = k - \dfrac{36}{u}$，则 $u_{n+1} = f(u_n)$。若 $u_{p+1} = u_1$，则反复应用 f 可得 $u_{p+2} = u_2$、$u_{p+3} = u_3$，如此类推，所以对每个 n 都有 $u_{n+p} = u_n$。反过来说，周期为 p 当然需要 $u_{p+1} = u_1 = 2$。因此，数列的周期恰好为 p，当且仅当 $u_{p+1} = 2$（而且各项均有定义，即永不为 0），并且没有更小的周期符合条件。

(a) 数列为常数数列，当且仅当 $u_2 = 2$，即 $k - 18 = 2$。所以 $k = 20$。

(b) 我们有 $u_2 = k - 18$，以及

$$u_3 = k - \frac{36}{k - 18} \qquad (k \neq 18).$$

令 $u_3 = 2$，得 $(k-2)(k-18) = 36$，所以 $k^2 - 20k = 0$，即 $k = 0$ 或 $k = 20$。$k = 20$ 给出常数数列，其周期为 1，而不是 2。当 $k = 0$ 时，数列为 $2, -18, 2, -18, \ldots$，周期确实

是 2。所以 $k = 0$。

(c) 我们还需要多计两项。当 $k \neq 18$ 时，

$$u_3 = \frac{k(k-18)-36}{k-18} = \frac{k^2-18k-36}{k-18},$$

然后，只要 $k^2 - 18k - 36 \neq 0$，便有

$$u_4 = k - \frac{36(k-18)}{k^2-18k-36} = \frac{k^3-18k^2-72k+648}{k^2-18k-36}.$$

条件 $u_5 = 2$ 表示 $k - 36/u_4 = 2$，即 $u_4 = \dfrac{36}{k-2}$（这里需要 $k \neq 2$；若 $k = 2$，则 $36/u_4 = 0$，这是不可能的）。交叉相乘，得

$$(k-2)(k^3-18k^2-72k+648) = 36(k^2-18k-36).$$

展开后把所有项移到左边，

$$k^4-20k^3-36k^2+792k-1296 = 36k^2-648k-1296,$$
$$k^4-20k^3-72k^2+1440k = 0.$$

想一想 怎样才可以不必反复试验，就把这样的四次多项式（quartic）分解因式？

动手之前，我们已经知道其中两个根。任何周期为 1 或 2 的数列同样满足 $u_5 = u_1$，所以 $k = 20$ 和 $k = 0$ 必定是根。先抽出 k，再抽出 $k - 20$，

$$k^4-20k^3-72k^2+1440k = k(k-20)(k^2-72).$$

所以 $k \in \{0, 20, 6\sqrt{2}, -6\sqrt{2}\}$。0 和 20 分别给出周期 2 和 1，所以新的值是 $k = \pm 6\sqrt{2}$。

最后，我们利用 $k^2 = 72$，验证当 $k = \pm 6\sqrt{2}$ 时各项均有定义：

- $u_2 = k - 18 \neq 0$；
- u_3 的分子是 $k^2 - 18k - 36 = 36 - 18k$，由于 $k \neq 2$，它不等于零；
- u_4 的分子是 $k(k^2-72) - 18k^2 + 648 = -1296 + 648 = -648 \neq 0$；
- 按构造 $u_5 = 2$，而且 $k \neq 2$、$k \neq 18$，符合推导过程中的要求。

由于 $k \neq 0, 20$，周期不是 1 或 2；而一个满足 $u_5 = u_1$ 的数列，若最小周期不是 1 或 2，其最小周期必定恰好是 4（最小周期必须整除 4）。所以数列的周期为 4，当且仅当 $k = \pm 6\sqrt{2}$。

第 (ii) 部分

取 $k = 37$。我们用归纳法证明 $u_n \geq 2$。当 $n = 1$ 时，由于 $u_1 = 2$，命题成立。若 $u_n \geq 2$，则 $u_n > 0$，所以 $0 < \dfrac{36}{u_n} \leq 18$，因此

$$u_{n+1} = 37 - \frac{36}{u_n} \geq 37 - 18 = 19 \geq 2.$$

所以对所有 n 都有 $u_n \geq 2$，这正是我们要证明的。（事实上，对所有 $n \geq 2$ 都有 $u_n \geq 19$。）

> **想一想** 极限（limit）所满足的方程有两个根。哪一个才是正确的？又为甚么可以在递推关系中令 $n \to \infty$？

由于 $u_n \to \ell$ 而且每个 $u_n \geq 2$，我们有 $\ell \geq 2$；特别地，$\ell \neq 0$，所以 $36/u_n \to 36/\ell$。在 $u_{n+1} = 37 - 36/u_n$ 的两边令 $n \to \infty$（左边同样趋向 ℓ），得

$$\ell = 37 - \frac{36}{\ell} \quad \Longrightarrow \quad \ell^2 - 37\ell + 36 = 0 \quad \Longrightarrow \quad (\ell - 1)(\ell - 36) = 0.$$

根 $\ell = 1$ 因 $\ell \geq 2$ 而被排除，所以

$$\ell = 36.$$

验算

- 当 $k = 6\sqrt{2} \approx 8.485$ 时，各项为 2, −9.515, 12.269, 5.551, 2, …；当 $k = -6\sqrt{2}$ 时，各项为 2, −26.485, −7.126, −3.433, 2, …。两者都在四步之后重复。
- 因式 $k^2 - 72$ 有一个简洁的含义：映射 $u \mapsto (ku - 36)/u$ 对应矩阵 $\begin{pmatrix} k & -36 \\ 1 & 0 \end{pmatrix}$，其迹（trace）为 k，行列式（determinant）为 36。它的特征值（eigenvalues）是 $6e^{\pm i\theta}$，其中 $k = 12\cos\theta$；当 $\theta = \pi/4$ 或 $3\pi/4$ 时，它的四次方是单位矩阵（identity）的倍数。这给出 $k = \pm 6\sqrt{2}$，而 $\theta = \pi/2$ 则给出周期为 2 的 $k = 0$。
- 当 $k = 37$ 时：$u_2 = 19$、$u_3 \approx 35.105$、$u_4 \approx 35.975$、$u_5 \approx 35.9993$，各项明显趋近 36，而不是 1。

常见错误

- 在 (b) 中答 $k = 0, 20$，或在 (c) 中把四个根全部写上。题目问的是周期 2 和周期 4，而按这个意思，常数数列的周期并不是 2。
- 除了 $\ell = 36$ 之外，把 $\ell = 1$ 也当作答案。(ii) 前半部分得出的界 $u_n \geq 2$，正是用来排除它的。
- 令 $n \to \infty$ 时没有指出 $\ell \neq 0$，而这是 $36/u_n \to 36/\ell$ 成立所必需的。

STEP 之外

映射 $u \mapsto (ku - 36)/u$ 是一个莫比乌斯变换（Möbius transformation），而这类映射的复合，对应 2×2 矩阵的乘法。这就是为甚么周期性归根究底取决于两个特征值之比是否单位根（root of unity）。当 $k = 37$ 时，两个不动点（fixed points）1 和 36 的表现并不相同：36 会吸引附近的项，而 1 则会排斥它们，这一点可以从 $|f'(\ell)| = 36/\ell^2$ 看出。这是不动点迭代（fixed-point iteration）和数列收敛（convergence）理论的起点，Abbott 的 *Understanding Analysis*（分析学入门教材）对此有严格的论述。Needham 的 *Visual Complex Analysis*（以几何直观讲解复分析）中，有一章精彩地讲述莫比乌斯变换及其分类。

详细解答

STEP II 2006，第 3 题

纯数学 · 根式与近似

> **STEP II 2006，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/06-s2-q3

关键思路

数 $5+\sqrt{24}$ 有一个“共轭数”（conjugate）$5-\sqrt{24}$，两者相乘等于 1。把一个数与其共轭数的同次幂相加，所有根式（surd）都会互相抵消，所以和是整数。共轭数很小，它的幂就更加微小，因此那个大的幂必定恰好略小于一个整数。第 (ii) 部分是同一个论证，只是把 5 换成 N。

解答

第 (i) 部分

设 $a=5+\sqrt{24}$ 及 $b=5-\sqrt{24}$。则

$$a+b=10, \qquad ab=25-24=1,$$

所以 $b=1/a$。因此我们要求的式子就是 a^4+b^4。两次运用 $a^2+b^2=(a+b)^2-2ab$：

$$\begin{aligned} a^2+b^2 &= 10^2-2=98, \\ a^4+b^4 &= (a^2+b^2)^2-2(ab)^2=98^2-2=9602. \end{aligned}$$

所以 $a^4+\dfrac{1}{a^4}=9602$，是一个整数。

至于各个不等式（inequality），记得 $\dfrac{1}{a}=b=5-\sqrt{24}$。

- $b>0.1 \iff \sqrt{24}<4.9 \iff 24<24.01$，这是成立的。（这里取平方是可逆的，因为两边都是正数。）

- $b < \frac{2}{19} \iff \sqrt{24} > 5 - \frac{2}{19} = \frac{93}{19} \iff 24 \cdot 19^2 > 93^2 \iff 8664 > 8649$，这是成立的。
- $\frac{2}{19} < 0.11 \iff 2 < 2.09$，这是成立的。

所以 $0.1 < \dfrac{1}{5+\sqrt{24}} < \dfrac{2}{19} < 0.11$。

想一想 这些上下界怎样把 a^4 确定至小数点后四位?

由第一个结果，$a^4 = 9602 - b^4$。把各项皆为正数的不等式 $0.1 < b < 0.11$ 取四次方，

$$0.0001 < b^4 < 0.11^4 = 0.00014641,$$

因此

$$9601.99985359 < a^4 < 9601.9999.$$

这个区间内的每一个数都介乎 9601.99985 与 9601.9999 之间，所以舍入后得

$$\left(5+\sqrt{24}\right)^4 = 9601.9999 \text{ （准确至小数点后四位）。}$$

（较精确的上界 $b < \frac{2}{19}$ 给出 $b^4 < \frac{16}{130321} \approx 0.000123$，这里用不着，但它说明了 $\frac{2}{19}$ 的来历：见“验算”。）

第 (ii) 部分

设 $A = N + \sqrt{N^2-1}$ 及 $B = N - \sqrt{N^2-1}$，由于 $N \geq 2$，两者皆为实数。与之前一样，

$$A + B = 2N, \qquad AB = N^2 - (N^2-1) = 1.$$

*第一步：$A^k + B^k$ 是整数。*用二项式定理（binomial theorem）展开这两个幂。含 $\sqrt{N^2-1}$ 偶数次幂的项都是整数，而且在两个展开式中完全相同；含奇数次幂的项则正负号相反，互相抵消。所以对每个正整数 k，$S_k = A^k + B^k$ 都是整数。（另一做法：A 和 B 是 $x^2 - 2Nx + 1 = 0$ 的根，所以 $S_k = 2NS_{k-1} - S_{k-2}$，其中 $S_0 = 2$，$S_1 = 2N$，再用归纳法（induction）便知它们全是整数。）

*第二步：B 很小。*我们断言 $0 < B < \dfrac{1}{2N - \frac{1}{2}}$。由于 $B = 1/A$ 且 $A > 0$，这等价于

$$A > 2N - \tfrac{1}{2} \iff \sqrt{N^2-1} > N - \tfrac{1}{2}.$$

两边都是正数（因为 $N \geq 2$），所以可以取平方：

$$N^2 - 1 > N^2 - N + \tfrac{1}{4} \iff N > \tfrac{5}{4},$$

而这是成立的，因为 $N \geq 2$。所以 $0 < B < \left(2N - \frac{1}{2}\right)^{-1}$，再取 k 次方（各项皆为正数），

$$0 < B^k < \left(2N - \tfrac{1}{2}\right)^{-k}.$$

第三步：作结。我们有 $A^k = S_k - B^k$，所以 A^k 与整数 S_k 的距离恰好是 B^k。最接近 A^k 的整数，与它的距离不会比 S_k 更远（事实上这个整数就是 S_k，因为 $B^k < \frac{1}{2}$）。因此

$$\left|A^k - (\text{最接近的整数})\right| \le B^k < \left(2N - \tfrac{1}{2}\right)^{-k},$$

这就是所要证明的结果。

验算

- 以数值计算，$(5+\sqrt{24})^4 \approx 9601.999896$，落在我们的区间之内，舍入后是 9601.9999。
- 取 $N = 5$，第 (ii) 部分给出 $(2N - \frac{1}{2})^{-1} = \frac{1}{9.5} = \frac{2}{19}$：正正是第 (i) 部分的上界。所以第 (i) 部分其实就是 $N = 5$，$k = 4$ 的情况，只是换了个面貌。
- 当 $N = 2$，$k = 1$：$2 + \sqrt{3} \approx 3.732$ 与 4 相距约 0.268，而 $(3.5)^{-1} \approx 0.286$，所以上界成立。我们亦用电脑核对过 N 至 1000、k 至 24 的所有情况。

常见错误

- 把 $\sqrt{N^2-1} > N - \frac{1}{2}$ 两边取平方，却没有指出两边都是正数。
- 写成 $a^4 \approx 9602.0000$：这个值比 9602 小，相差介乎 0.0001 与 0.00015 之间，而这已足以改变小数点后第四位。
- 忘记了排除 $N = 1$ 的情况是有原因的：那时 $B = 1$，整个论证便不成立。

STEP 之外

像 $5+\sqrt{24}$ 这样的数是环（ring）$\mathbb{Z}[\sqrt{6}]$ 中的单位（unit）：它们有同一形式的逆元 $5-\sqrt{24}$，而它们的幂给出佩尔方程（Pell equation）$x^2 - 24y^2 = 1$ 的所有解。这类单位的幂会越来越接近整数，这也是通往皮索数（Pisot number）理论的其中一条途径，皮索数的幂同样有这种接近整数的性质。Davenport 的 *The Higher Arithmetic*（一本初等数论名著）对佩尔方程和连分数（continued fraction）有十分精彩的论述。

详细解答

STEP II 2006，第 4 题

纯数学 · 代换积分法

STEP II 2006，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/06-s2-q4

关键思路

代换（substitution）$x = \pi - t$ 不会改变 $\sin x$，却会把因子 x 变成 $\pi - t$，于是原来的积分在等式右边再次出现，可以像解方程一样把它解出来。这样就完全消去了碍事的因子 x。此后，题目中每个积分都是 $\sin x$ 的函数乘以 $\sin x$（或 $\cos x$），再用代换 $u = \cos x$ 或 $u = \sin x$ 便可完成。真正要花心思的，是看出怎样把第 (ii) 和第 (iii) 部分化成这条恒等式（identity）适用的形式。

解答

恒等式

设 $I = \int_0^\pi x\, f(\sin x)\, \mathrm{d}x$。令 $x = \pi - t$，则 $\mathrm{d}x = -\mathrm{d}t$；当 $x = 0$ 时，$t = \pi$，而当 $x = \pi$ 时，$t = 0$。由于 $\sin(\pi - t) = \sin t$，

$$\begin{aligned} I &= \int_\pi^0 (\pi - t)\, f(\sin t)\, (-\mathrm{d}t) = \int_0^\pi (\pi - t)\, f(\sin t)\, \mathrm{d}t \\ &= \pi \int_0^\pi f(\sin t)\, \mathrm{d}t - \int_0^\pi t\, f(\sin t)\, \mathrm{d}t = \pi \int_0^\pi f(\sin x)\, \mathrm{d}x - I, \end{aligned}$$

因为积分变量叫甚么名字并无关系。因此 $2I = \pi \int_0^\pi f(\sin x)\, \mathrm{d}x$，这正是我们所需的结果：

$$\int_0^\pi x\, f(\sin x)\, \mathrm{d}x = \tfrac{1}{2}\pi \int_0^\pi f(\sin x)\, \mathrm{d}x.$$

第 (i) 部分

这里 $f(\sin x) = \dfrac{\sin x}{3+\sin^2 x}$，所以由恒等式可得

$$\int_0^{\pi} \frac{x\sin x}{3+\sin^2 x}\,\mathrm{d}x = \tfrac{1}{2}\pi J, \qquad J = \int_0^{\pi} \frac{\sin x}{3+\sin^2 x}\,\mathrm{d}x.$$

写成 $3+\sin^2 x = 4-\cos^2 x$，并代入 $u=\cos x$，$\mathrm{d}u = -\sin x\,\mathrm{d}x$，其中 u 由 1 递减至 -1：

$$J = \int_{-1}^{1} \frac{\mathrm{d}u}{4-u^2}.$$

利用部分分式（partial fractions），$\dfrac{1}{4-u^2} = \dfrac{1}{4}\left(\dfrac{1}{2-u}+\dfrac{1}{2+u}\right)$，所以

$$J = \frac{1}{4}\left[\ln\frac{2+u}{2-u}\right]_{-1}^{1} = \frac{1}{4}\left(\ln 3 - \ln\tfrac{1}{3}\right) = \tfrac{1}{2}\ln 3.$$

（在 $[-1,1]$ 上，$2\pm u$ 两者皆为正数，所以无须加上绝对值符号。）因此

$$\int_0^{\pi} \frac{x\sin x}{3+\sin^2 x}\,\mathrm{d}x = \frac{\pi}{4}\ln 3.$$

第 (ii) 部分

> ? **想一想** 恒等式只适用于 $[0,\pi]$。我们该怎样处理 $[\pi, 2\pi]$?

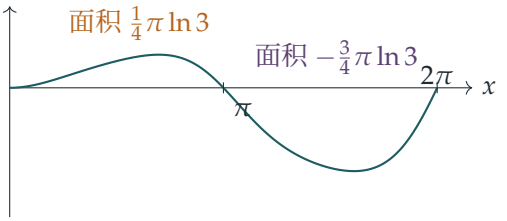
在 π 处把积分范围分成两段，再用 $x=\pi+t$ 把第二段平移回 $[0,\pi]$。这时 $\sin x = -\sin t$ 而 $\sin^2 x = \sin^2 t$，所以

$$\begin{aligned}\int_{\pi}^{2\pi} \frac{x\sin x}{3+\sin^2 x}\,\mathrm{d}x &= -\int_0^{\pi} \frac{(\pi+t)\sin t}{3+\sin^2 t}\,\mathrm{d}t = -\pi J - \int_0^{\pi} \frac{t\sin t}{3+\sin^2 t}\,\mathrm{d}t \\ &= -\tfrac{1}{2}\pi\ln 3 - \tfrac{1}{4}\pi\ln 3 = -\tfrac{3}{4}\pi\ln 3,\end{aligned}$$

当中用到 $J = \frac{1}{2}\ln 3$ 和第 (i) 部分的结果。加上第 (i) 部分在 $[0,\pi]$ 上的积分，

$$\int_0^{2\pi} \frac{x\sin x}{3+\sin^2 x}\,\mathrm{d}x = \frac{\pi}{4}\ln 3 - \frac{3\pi}{4}\ln 3 = -\frac{\pi}{2}\ln 3.$$

下面的草图说明了答案为何是负数：被积函数（integrand）在两半区间上的形状相同，但在 $[\pi, 2\pi]$ 上，它所乘的 x 值较大。



第 (iii) 部分

由于 $\sin 2x = 2\sin x\cos x$，而且在 $[0,\pi]$ 上 $\sin x \geq 0$，

$$|\sin 2x| = 2\sin x\,|\cos x| = 2\sin x\sqrt{1-\sin^2 x},$$

这纯粹是 $\sin x$ 的函数。

? **想一想** 为甚么这里出现的是 $|\cos x|$ 而不是 $\cos x$，这一点如此重要?

在 $[0,\pi]$ 上，$\cos x$ 并不是 $\sin x$ 的函数：$\cos(\pi - x) = -\cos x$，而 $\sin(\pi - x) = \sin x$。但 $|\cos x| = \sqrt{1-\sin^2 x}$ 却是，所以正是有了绝对值，我们才可以使用恒等式：

$$\int_0^\pi \frac{x|\sin 2x|}{3+\sin^2 x}\,\mathrm{d}x = \tfrac{1}{2}\pi\int_0^\pi \frac{|\sin 2x|}{3+\sin^2 x}\,\mathrm{d}x.$$

新的被积函数在 $x \mapsto \pi - x$ 之下保持不变，所以在 $[0,\pi]$ 上的积分是在 $[0,\frac{1}{2}\pi]$ 上的积分的两倍，而在后者的区间上 $\cos x \geq 0$，绝对值符号可以去掉：

$$\int_0^\pi \frac{|\sin 2x|}{3+\sin^2 x}\,\mathrm{d}x = 2\int_0^{\pi/2} \frac{2\sin x\cos x}{3+\sin^2 x}\,\mathrm{d}x = 2\Big[\ln(3+\sin^2 x)\Big]_0^{\pi/2} = 2\ln\tfrac{4}{3},$$

因为分子正是分母的导数。因此

$$\int_0^\pi \frac{x|\sin 2x|}{3+\sin^2 x}\,\mathrm{d}x = \tfrac{1}{2}\pi\cdot 2\ln\tfrac{4}{3} = \pi\ln\frac{4}{3}.$$

验算

- 用辛普森法则 (Simpson's rule) 计算, (i) 得 0.86285, (ii) 得 -1.72570, (iii) 得 0.90378, 与 $\frac{1}{4}\pi\ln 3$、$-\frac{1}{2}\pi\ln 3$ 和 $\pi\ln\frac{4}{3}$ 吻合。
- (ii) 的答案恰好是 (i) 的答案的 -2 倍。在 π 处分段便可解释这一点：两半分别是 $\frac{1}{4}\pi\ln 3$ 和 $-\frac{3}{4}\pi\ln 3$。
- 快速检验一下恒等式：取 $f \equiv 1$，两边都等于 $\frac{1}{2}\pi^2$。

常见错误

- 在 $[0,2\pi]$ 上套用类似的“恒等式”，这样会得出 $\pi\int_0^{2\pi} f(\sin x)\,\mathrm{d}x = 0$。证明需要用到 $\sin(2\pi - t) = \sin t$，而这是错的。
- 在 (iii) 中，在整个 $[0,\pi]$ 上都写 $|\sin 2x| = 2\sin x\cos x$，这样里面的积分会得出 0。

STEP 之外

让积分再次出现然后把它解出来的技巧，以及运用 $x \mapsto \pi - x$ 这类对称性（symmetry），都是分析学（analysis）中的标准工具，也常用于计算没有初等原函数（elementary antiderivative）的定积分。它们通向傅里叶级数（Fourier series），在那里，x 乘以三角函数在一个周期上的积分正是最基本的研究对象。想多了解这类论证手法，可参阅 Siklos 的 *Advanced Problems in Mathematics*；至于积分的严格理论，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP II 2006，第 5 题

纯数学 · 下取整函数与面积

STEP II 2006，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/06-s2-q5

关键思路

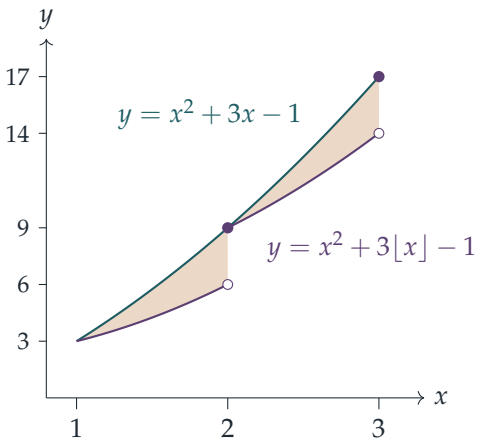
在两个相邻整数之间，即 $k \le x < k+1$ 时，下取整函数（floor function）是常数，其值为 k。因此在每一个这样的单位区间上，那条锯齿状的曲线其实只是一条普通的光滑曲线，而所求面积可以分拆成若干长条的面积之和，每个单位区间对应一条。先把 $[k, k+1]$ 上那一条的面积写成 k 的函数，再把 $k = 1, 2, \ldots, n-1$ 的各条加起来。

解答

以下一律记 $f(x) = x^2 + 3x - 1$。在 $1 \le x \le 3$ 上，它由 $f(1) = 3$ 经过 $f(2) = 9$ 升至 $f(3) = 17$；而且对所有 $x \ge 1$ 它都是递增的，因为在该范围内 $f'(x) = 2x + 3 > 0$。

第 (i) 部分

第二条曲线是 $y = x^2 + 3\lfloor x \rfloor - 1$。在 $[1, 2)$ 上有 $\lfloor x \rfloor = 1$，所以它是 $y = x^2 + 2$，由 3 升向 6；在 $[2, 3)$ 上有 $\lfloor x \rfloor = 2$，所以它是 $y = x^2 + 5$，由 9 升向 14；而在 $x = 3$ 处它等于 17。每一段都是抛物线 $y = f(x)$ 向下平移所得，而两条曲线在每个整数处相交。



两条曲线之间的垂直距离是

$$(x^2 + 3x - 1) - (x^2 + 3\lfloor x \rfloor - 1) = 3(x - \lfloor x \rfloor) \geq 0,$$

所以第一条曲线永远不会在第二条之下。在区间 $[k, k+1]$ 上（k 为整数），这个距离是 $3(x-k)$，只有 $x = k+1$ 这一点例外，而单独一点并不影响积分。因此 $[k, k+1]$ 上的长条面积为

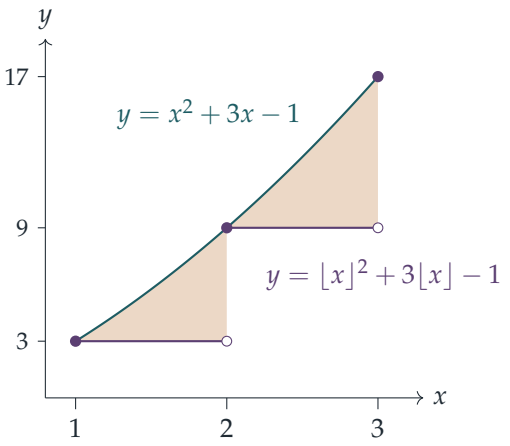
$$\int_k^{k+1} 3(x-k)\,\mathrm{d}x = \left[\tfrac{3}{2}(x-k)^2\right]_k^{k+1} = \tfrac{3}{2},$$

对每个 k 都相同。区间 $1 \leq x \leq n$ 由 $n-1$ 个单位区间 $[1,2], [2,3], \ldots, [n-1, n]$ 组成，所以两条曲线之间的面积是

$$\tfrac{3}{2}(n-1).$$

第 (ii) 部分

现在第二条曲线是 $y = f(\lfloor x \rfloor)$，这是一个阶梯函数（step function）：在 $[1,2)$ 上 $y = f(1) = 3$，在 $[2,3)$ 上 $y = f(2) = 9$，而在 $x = 3$ 处 $y = f(3) = 17$。每一级阶梯都由抛物线上某个整数点出发，水平向右延伸。



想一想 哪一条曲线在上方？它会一直在上方吗？

由于 f 在 $x \geq 1$ 时递增，而且 $\lfloor x \rfloor \leq x$，所以在整个 $1 \leq x \leq n$ 上都有 $f(\lfloor x \rfloor) \leq f(x)$。因此抛物线始终在上方，所求面积就是 $f(x) - f(\lfloor x \rfloor)$ 的积分，无须担心正负号改变。

在 $[k, k+1]$ 上（$k \geq 1$ 为整数），阶梯的高度是 $f(k)$（同样不理会 $x = k+1$ 这一点）。代入 $x = k + t$，其中 $0 \leq t \leq 1$，

$$\begin{aligned} f(k+t) - f(k) &= (k+t)^2 + 3(k+t) - 1 - (k^2 + 3k - 1) \\ &= 2kt + t^2 + 3t, \end{aligned}$$

所以 $[k, k+1]$ 上的长条面积为

$$\int_0^1 (2kt + t^2 + 3t)\,\mathrm{d}t = k + \tfrac{1}{3} + \tfrac{3}{2} = k + \tfrac{11}{6}.$$

与第 (i) 部分不同，这次长条的面积随 k 增大，因为抛物线愈来愈陡。把 $k = 1, 2, \ldots, n-1$ 的长条加起来，并利用 $\sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1)$，

$$\begin{aligned} \text{面积} &= \sum_{k=1}^{n-1} \left(k + \tfrac{11}{6}\right) = \tfrac{1}{2}n(n-1) + \tfrac{11}{6}(n-1) \\ &= \tfrac{1}{6}(n-1)(3n+11), \end{aligned}$$

这正是所要求的结果。

验算

- 当 $n = 1$ 时根本没有区域，而两个答案都给出 0。
- 当 $n = 2$ 时，第 (ii) 部分给出 $\frac{1}{6} \cdot 1 \cdot 17 = \frac{17}{6}$。直接计算，$\int_1^2 (x^2 + 3x - 4)\,\mathrm{d}x = \frac{7}{3} + \frac{9}{2} - 4 = \frac{17}{6}$。

- 以中点法则（midpoint rule）对 $n = 2, \ldots, 7$ 的差距作数值积分，结果与 $\frac{3}{2}(n-1)$ 及 $\frac{1}{6}(n-1)(3n+11)$ 吻合至小数点后四位；例如 $n = 4$ 时第 (ii) 部分给出 11.5。

常见错误

- 数成 n 条长条而不是 $n-1$ 条：区间 $[1, n]$ 只包含 $n-1$ 个单位区间。
- 在第 (ii) 部分，误用 $\int_0^1 2kt \, \mathrm{d}t = 2k$ 而不是 k。
- 在整个区间上把 $\lfloor x \rfloor$ 当作 x 来积分（或干脆忽略它），而没有先在各整数处把区间分开。

STEP 之外

第 (ii) 部分比较的是曲线下的面积与曲线下方一列“阶梯”状矩形的面积，这正正是步长为 1 的黎曼下和（lower Riemann sum）。在大学第一门分析（analysis）课程中，积分的定义就是让阶梯愈来愈窄，把这样的下和与上和夹在一起；而像 $\lfloor x \rfloor$ 这类有跳跃的函数仍然可积（integrable），因为它们的跳跃点都是孤立点。Abbott 的 *Understanding Analysis* 对黎曼积分（Riemann integral）有非常易读的讲解。

详细解答

STEP II 2006，第 7 题

纯数学 · 椭圆的切线

STEP II 2006，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/06-s2-q7

关键思路

求切线（tangent）最快的方法，是用“把 x^2 换成 xx_0、把 y^2 换成 yy_0”的形式，用隐函数微分法（implicit differentiation）也同样可以。至于那个作图，只需求出 Q 的坐标和 PQ 的方程，再与一般点上的切线比较。题目中 $\tan(\alpha/2) = k$ 这几个字是很强的提示：它叫你用 $t = \tan(\alpha/2)$ 来表示 $\cot\alpha$ 和 $\operatorname{cosec}\alpha$。

解答

切线

把 $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ 对 x 作隐函数微分，得

$$\frac{2x}{a^2} + \frac{2y}{b^2}\frac{\mathrm{d}y}{\mathrm{d}x} = 0, \qquad \text{所以} \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{b^2x}{a^2y}.$$

在 $(a\cos\alpha, b\sin\alpha)$ 处，设 $\sin\alpha \neq 0$（否则切线是铅垂线，题目所给的方程便没有意义），斜率（gradient）为

$$-\frac{b^2 a\cos\alpha}{a^2 b\sin\alpha} = -\frac{b\cot\alpha}{a}.$$

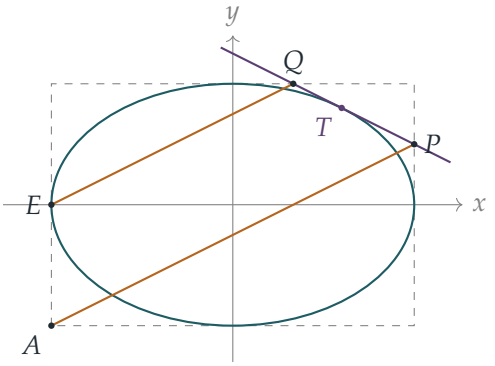
所以切线是

$$\begin{aligned} y - b\sin\alpha &= -\frac{b\cot\alpha}{a}(x - a\cos\alpha) \\ \implies \quad y &= -\frac{b\cot\alpha}{a}x + b\frac{\cos^2\alpha}{\sin\alpha} + b\sin\alpha = -\frac{b\cot\alpha}{a}x + \frac{b(\cos^2\alpha + \sin^2\alpha)}{\sin\alpha}, \end{aligned}$$

即 $y = -\dfrac{b\cot\alpha}{a}x + b\operatorname{cosec}\alpha$，这正是所求。

作图

下图取 $a=3$、$b=2$、$k=\frac{1}{2}$。椭圆（ellipse）位于以 $(\pm a,\pm b)$ 为顶点的长方形之内；A 是左下角的顶点，E 是左边的中点，P 在右边上，而 Q 在上边上。



直线 AP 连接 $(-a,-b)$ 与 (a,kb)，所以它的斜率是

$$\frac{kb-(-b)}{a-(-a)}=\frac{(1+k)b}{2a}.$$

通过 $E(-a,0)$ 的平行线是 $y=\dfrac{(1+k)b}{2a}(x+a)$。代入 $y=b$ 得 $x+a=\dfrac{2a}{1+k}$，所以

$$Q=\left(\frac{a(1-k)}{1+k},\ b\right).$$

于是 PQ 的斜率是

$$\frac{b-kb}{\dfrac{a(1-k)}{1+k}-a}=\frac{b(1-k)}{\dfrac{-2ak}{1+k}}=-\frac{b(1-k^2)}{2ak},$$

由于 $k\neq 0$，这个式子没有问题。利用点 $P(a,kb)$，直线 PQ 是

$$y=kb-\frac{b(1-k^2)}{2ak}(x-a)=-\frac{b(1-k^2)}{2ak}\,x+\frac{2k^2b+b(1-k^2)}{2k},$$

所以

$$PQ:\qquad y=-\frac{b}{a}\cdot\frac{1-k^2}{2k}\,x+b\cdot\frac{1+k^2}{2k}.$$

? **想一想** 我们怎样看出这是椭圆上某一点的切线?

把它与上面求得在 $(a\cos\alpha,b\sin\alpha)$ 处的切线比较。我们需要找一个 α，使得 $\cot\alpha=\dfrac{1-k^2}{2k}$ 及 $\operatorname{cosec}\alpha=\dfrac{1+k^2}{2k}$。这正是 t 公式（t-formulae）：设 $t=\tan(\alpha/2)$，则

$$\tan\alpha=\frac{2t}{1-t^2},\qquad \sin\alpha=\frac{2t}{1+t^2},\qquad \text{所以}\qquad \cot\alpha=\frac{1-t^2}{2t},\qquad \operatorname{cosec}\alpha=\frac{1+t^2}{2t}.$$

由于 $0 < k < 1$，我们可以取 $\alpha = 2\arctan k$，它位于 $(0, \frac{\pi}{2})$ 之内，所以 $\sin\alpha \neq 0$，而且 $\tan(\alpha/2) = k$。取 $t = k$，这两个表达式便与 PQ 的斜率和截距（intercept）都吻合。所以 PQ 就是直线 $y = -\frac{b\cot\alpha}{a}x + b\operatorname{cosec}\alpha$，亦即椭圆在 $(a\cos\alpha, b\sin\alpha)$ 处的切线，其中 $\tan(\alpha/2) = k$。这正是我们要证明的。

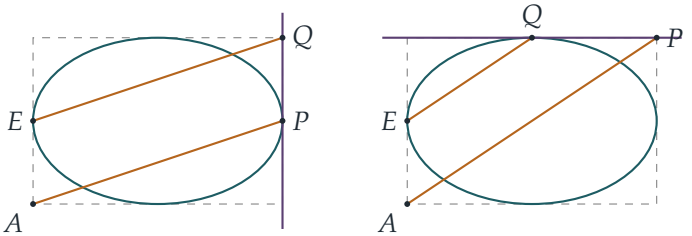
用坐标表示，切点（point of contact）是 $T = \left(\frac{a(1-k^2)}{1+k^2}, \frac{2kb}{1+k^2}\right)$。

$k = 0$ 与 $k = 1$ 的情况

上面的代数运算曾经除以 k，而切线公式又需要 $\sin\alpha \neq 0$，所以我们直接把作图重做一次。

$k = 0$。这时 $P = (a, 0)$，而 AP 的斜率是 $\frac{b}{2a}$。通过 E 的平行线是 $y = \frac{b}{2a}(x + a)$，它与 $y = b$ 相交于 $x = a$，所以 $Q = (a, b)$。于是 PQ 是铅垂线 $x = a$。这是椭圆在 $(a, 0)$ 处的切线，该点对应 $\alpha = 0$，而 $\tan(0/2) = 0 = k$。所以即使这条切线不能写成 $y = mx + c$ 的形式，结果确实仍然成立。

$k = 1$。这时 $P = (a, b)$，而 AP 是长方形的对角线，斜率为 $\frac{b}{a}$。通过 E 的平行线是 $y = \frac{b}{a}(x + a)$，它与 $y = b$ 相交于 $x = 0$，所以 $Q = (0, b)$。于是 PQ 是水平线 $y = b$，即椭圆最高点 $(0, b)$ 处的切线。这一点对应 $\alpha = \frac{\pi}{2}$，而 $\tan(\pi/4) = 1 = k$。所以结果在这里同样成立（而且公式也适用：由 $\cot\frac{\pi}{2} = 0$ 和 $\operatorname{cosec}\frac{\pi}{2} = 1$ 得出 $y = b$）。



验算

- 取 $a = 3$、$b = 2$、$k = \frac{1}{2}$。这时 $Q = (1, 2)$，而 PQ 是 $y = -\frac{1}{2}x + \frac{5}{2}$。另外 $\tan\alpha = \frac{4}{3}$，所以 $\cos\alpha = \frac{3}{5}$、$\sin\alpha = \frac{4}{5}$，而 $T = (\frac{9}{5}, \frac{8}{5})$。这一点在 PQ 上（$-\frac{9}{10} + \frac{5}{2} = \frac{8}{5}$），而该处的斜率是 $-\frac{b^2x}{a^2y} = -\frac{4\cdot 9/5}{9\cdot 8/5} = -\frac{1}{2}$。
- 把 PQ 代入椭圆方程，应该得到一条有重根（repeated root）的二次方程。我们已用一千组随机的 a、b、k 以数值方法验证了这一点，以及 Q 和 PQ 的公式。
- 当 $k \to 0^+$ 时，$T \to (a, 0)$；当 $k \to 1^-$ 时，$T \to (0, b)$。这与两个端点情况一致。

常见错误

- 弄错 PQ 斜率的正负号：分母 $\frac{a(1-k)}{1+k} - a$ 是负数。
- 因为 $\operatorname{cosec} 0$ 没有定义，便说结果在 $k = 0$ 时不成立。失效的只是公式，几何上并没有问题：PQ 仍然是 $\alpha = 0$ 处的切线。

STEP 之外

代换 $t = \tan(\alpha/2)$ 给出椭圆的一个有理参数化（rational parametrisation）：$\left(a\frac{1-t^2}{1+t^2}, b\frac{2t}{1+t^2}\right)$。把三角函数积分化为有理函数积分，以及求出所有毕氏三元数（Pythagorean triples），用的都是同一个想法。每一条圆锥曲线（conic）都可以这样作有理参数化，而像本题这种只靠平行线和交点完成的作图，正是射影几何（projective geometry）研究的对象。想初步了解这与数论（number theory）的关系，可以读 Davenport 的 *The Higher Arithmetic*。

详细解答

STEP II 2006，第 8 题

纯数学 · 直线的向量几何

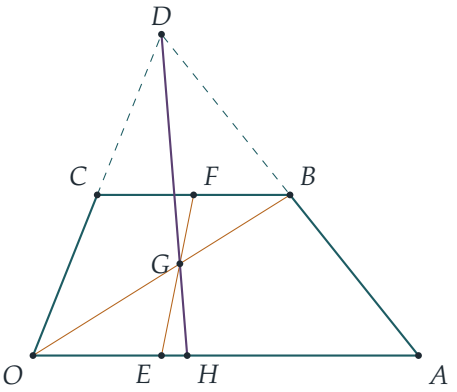
STEP II 2006，第 8 题

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/06-s2-q8

关键思路

题目中每一点都定义为两条直线的交点，所以整条题目只靠一套固定做法：把每条直线写成第一部分的形式，令两个表达式相等，然后比较 **a** 和 **c** 的系数（comparing coefficients）。这样比较是合理的，因为 **a** 与 **c** 不平行，所以一个向量写成 $p\mathbf{a}+q\mathbf{c}$ 的方式只有一种。



解答

直线的方程

通过位置向量（position vector）为 **x** 和 **y** 的两点的直线，经过 **x**，方向为 $\mathbf{y}-\mathbf{x}$，所以线上的点是

$$\mathbf{r}=\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x})=(1-\alpha)\mathbf{x}+\alpha\mathbf{y},$$

其中 α 取遍所有实数。留意 $\alpha = 0$ 给出 $\mathbf{x}$，$\alpha = 1$ 给出 $\mathbf{y}$，而两个系数之和恒为 1。这就是所要求的形式。

在整个解答中，$\mathbf{a}$ 与 $\mathbf{c}$ 都不平行（否则 O、A、C 三点共线（collinear），便不会有梯形（trapezium）），所以若 $p\mathbf{a} + q\mathbf{c} = p'\mathbf{a} + q'\mathbf{c}$，则 $p = p'$ 且 $q = q'$。

第 (i) 部分

由于 CB 平行于 OA，向量 $\overrightarrow{CB}$ 是 $\overrightarrow{OA} = \mathbf{a}$ 的纯量倍数（scalar multiple）；记它为 $\lambda\mathbf{a}$。于是

$$\mathbf{b} = \overrightarrow{OC} + \overrightarrow{CB} = \mathbf{c} + \lambda\mathbf{a}.$$

按次序绕 $OABC$ 一周，$\overrightarrow{CB}$ 与 $\overrightarrow{OA}$ 方向相同，所以 $\lambda > 0$；而 $CB < OA$ 给出 $\lambda < 1$。因此 $0 < \lambda < 1$，特别地 $\lambda \neq 1$，这一点稍后会用到。

第 (ii) 部分

和 E F。这两点可直接由比例得出：

$$\mathbf{e} = \tfrac{1}{3}\mathbf{a}, \qquad \mathbf{f} = \tfrac{1}{2}(\mathbf{c} + \mathbf{b}) = \tfrac{1}{2}\lambda\mathbf{a} + \mathbf{c}.$$

D。直线 OC 上的点是 $\mu\mathbf{c}$。直线 AB 上的点是

$$(1 - \beta)\mathbf{a} + \beta(\lambda\mathbf{a} + \mathbf{c}) = (1 - \beta + \beta\lambda)\mathbf{a} + \beta\mathbf{c}.$$

在 D 点，两者相等。比较系数得 $\beta = \mu$ 及 $1 - \beta(1 - \lambda) = 0$，所以 $\beta = \dfrac{1}{1 - \lambda}$（没有问题，因为 $\lambda \neq 1$）。因此

$$\mathbf{d} = \frac{1}{1 - \lambda}\,\mathbf{c}.$$

由于 $\frac{1}{1-\lambda} > 1$，D 确实位于 OC 越过 C 的延长线上，与预期相符。

G。OB 上的点是 $t(\lambda\mathbf{a} + \mathbf{c})$。$EF$ 上的点是

$$(1 - s)\tfrac{1}{3}\mathbf{a} + s\big(\tfrac{1}{2}\lambda\mathbf{a} + \mathbf{c}\big) = \big(\tfrac{1}{3} - \tfrac{1}{3}s + \tfrac{1}{2}\lambda s\big)\mathbf{a} + s\mathbf{c}.$$

比较 $\mathbf{c}$ 的系数得 $s = t$；再比较 $\mathbf{a}$ 的系数，得

$$\lambda t = \tfrac{1}{3} - \tfrac{1}{3}t + \tfrac{1}{2}\lambda t \quad\Longrightarrow\quad t\Big(\frac{3\lambda + 2}{6}\Big) = \frac{1}{3} \quad\Longrightarrow\quad t = \frac{2}{3\lambda + 2}.$$

所以

$$\mathbf{g} = \frac{2}{3\lambda + 2}\,(\lambda\mathbf{a} + \mathbf{c}).$$

H。直线 DG 上的点是

$$(1 - k)\mathbf{d} + k\mathbf{g} = \frac{2k\lambda}{3\lambda + 2}\,\mathbf{a} + \Big(\frac{1 - k}{1 - \lambda} + \frac{2k}{3\lambda + 2}\Big)\mathbf{c}.$$

想一想 怎样从这条直线上找出位于 OA 上的那一点?

直线 OA 上的点只是 $\mathbf{a}$ 的倍数, 所以我们需要 $\mathbf{c}$ 的系数等于零。两边乘以 $(1-\lambda)(3\lambda+2)$,

$$\begin{aligned}(1-k)(3\lambda+2)+2k(1-\lambda)&=0\\(3\lambda+2)-k\big(3\lambda+2-2+2\lambda\big)&=0\\k&=\frac{3\lambda+2}{5\lambda}.\end{aligned}$$

这时 $\mathbf{a}$ 的系数是

$$\frac{2\lambda}{3\lambda+2}\cdot\frac{3\lambda+2}{5\lambda}=\frac{2}{5},\qquad \text{所以}\qquad \mathbf{h}=\tfrac{2}{5}\,\mathbf{a}.$$

把答案整理如下:

$$\mathbf{d}=\frac{\mathbf{c}}{1-\lambda},\quad \mathbf{e}=\tfrac{1}{3}\mathbf{a},\quad \mathbf{f}=\tfrac{1}{2}\lambda\mathbf{a}+\mathbf{c},\quad \mathbf{g}=\frac{2(\lambda\mathbf{a}+\mathbf{c})}{3\lambda+2},\quad \mathbf{h}=\tfrac{2}{5}\mathbf{a}.$$

由于 $OH=\frac{2}{5}OA$ 及 $HA=\frac{3}{5}OA$,

$$OH:HA=2:3,$$

不论梯形是甚么形状, 结果都一样。

验算

- 取 $\mathbf{a}=(6,0)$、$\mathbf{c}=(1,2.5)$ 及 $\lambda=\frac{1}{2}$ (即图中的情况)。由公式得 $D=(2,5)$、$G\approx(2.286,1.429)$ 及 $H=(2.4,0)$; 由 D 到 G 的斜率是 $-3.571/0.286=-12.5$, 由 D 到 H 的斜率也是这个值, 即 $-5/0.4$。所以 D、G、H 确实共线。
- H 的答案与 λ、$\mathbf{a}$ 和 $\mathbf{c}$ 都无关。我们用电脑以精确的有理数运算, 对 2000 个随机梯形重复整个作图过程, 每次都得到 $OH:HA=2:3$。
- 当 $\lambda\to1$ 时, 边 OC 与 AB 趋于平行, 而 $\mathbf{d}=\mathbf{c}/(1-\lambda)$ 跑向无穷远, 与预期相符。

常见错误

- 两条不同的直线用了同一个参数 (parameter) 字母。每条直线都要有自己的参数, 然后才可令两式相等。
- 比较系数却没有说明为何可以这样做: 这方法行得通, 是因为 $\mathbf{a}$ 与 $\mathbf{c}$ 不平行。
- 写成 $\mathbf{b}=\mathbf{a}+\lambda\mathbf{c}$。与 $\mathbf{a}$ 平行的是 $\overrightarrow{CB}$, 而不是 $\overrightarrow{AB}$。

STEP 之外

答案 2 : 3 与梯形的形状无关，是因为每一步只用到直线、交点和直线上的比例，而仿射变换（affine map，即线性变换（linear map）之后再作平移）会保持这一切不变。所以你可以把梯形变换成一个方便计算的梯形，例如令 $OABC$ 中的 C 位于 O 的正上方，然后在那里计算。这就是*仿射几何*（affine geometry）的想法；系数形式 $(1-\alpha)\mathbf{x}+\alpha\mathbf{y}$ 正是重心坐标（barycentric coordinates）的起点。Strang 的 *Introduction to Linear Algebra* 讲解线性变换和坐标，而 Siklos 的 *Advanced Problems in Mathematics* 则有更多这类风格的向量几何题目。

详细解答

STEP II 2006，第 9 题

力学 · 摩擦力与倾倒

STEP II 2006，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：

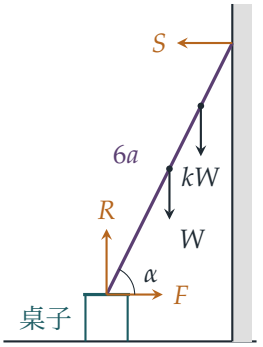
step.maths.org/questions/06-s2-q9

关键思路

把梯子和桌子视为两个独立的物体，两者靠梯脚处的接触连系起来。单看梯子，便可求出梯脚处的反作用力和摩擦力，而这两个力取决于油漆工站得多高。然后根据牛顿第三定律（Newton's third law），同样的力以相反方向作用于桌子。对于桌子，我们比较两个临界值：一个是令桌子滑动的推力，另一个是对远端桌脚的力矩（moment）大于桌上重量的力矩的推力。

解答

我们把梯子模拟为均匀杆（uniform rod），油漆工模拟为质点（particle），墙壁视为光滑（smooth），桌子则视为刚体（rigid body），桌面高度为 a（忽略桌面的厚度）。设 $\alpha = \arctan 2$，则 $\tan\alpha = 2$、$\sin\alpha = \frac{2}{\sqrt{5}}$ 及 $\cos\alpha = \frac{1}{\sqrt{5}}$。设油漆工站在沿梯子距梯脚 x 之处，其中 $0 \le x \le 6a$。



第 (i) 部分

作用于梯子的力有：作用于梯子中点的重量 W、油漆工的重量 kW、光滑墙壁给予的水平反作用力 S，以及梯脚处向上的法向反作用力（normal reaction）R 和指向墙壁的摩擦力（friction）F。（摩擦力必须指向墙壁，因为它是唯一能与 S 平衡的水平力。）

沿铅垂和水平方向分解力，得

$$R = (k+1)W, \qquad F = S.$$

对梯脚取力矩：梯顶比梯脚高 $6a\sin\alpha$，而两个重量的作用线与梯脚的水平距离分别为 $3a\cos\alpha$ 和 $x\cos\alpha$，所以

$$S \cdot 6a\sin\alpha = W \cdot 3a\cos\alpha + kW \cdot x\cos\alpha.$$

两边除以 $6a\sin\alpha$，并利用 $\cot\alpha = \frac{1}{2}$，得

$$F = S = \frac{W}{12}\Big(3 + \frac{kx}{a}\Big).$$

此式随 x 递增，所以当油漆工站在梯子最顶端，即 $x = 6a$ 时，所需的摩擦力最大：

$$F_{\max} = \frac{(3+6k)W}{12} = \frac{(2k+1)W}{4}.$$

可提供的最大摩擦力为

$$\tfrac{1}{2}R = \frac{(k+1)W}{2} = \frac{(2k+2)W}{4}.$$

由于 $2k+1 < 2k+2$，无论油漆工站在哪个位置，也无论 k 取何值，都有 $F \le F_{\max} < \frac{1}{2}R$。因此，在桌子固定的情况下，不论油漆工站得多高，梯子都不会滑动，这正是所要证明的。

第 (ii) 部分

现在 $k = 9$，所以 $R = 10W$，而

$$F = \frac{W}{4}\Big(1 + \frac{3x}{a}\Big),$$

它由 $x = 0$ 时的 $\frac{1}{4}W$ 增至 $x = 6a$ 时的 $\frac{19}{4}W$。第 (i) 部分的结论仍然适用，所以梯子不会在桌面上滑动。

根据牛顿第三定律，梯子在桌面中心对桌面施力：一个向下的力 $R = 10W$，以及一个背离墙壁的水平力 F。桌子本身的重量 $2W$ 亦作用于其中心，而地面则以法向反作用力和摩擦力作用于各桌脚。

? **想一想** 作用于桌面中心的水平推力，怎样能令桌子倾侧?

推力 F 作用于高度 a 处，所以它对离墙最远的两条桌脚的底部有力矩。如果桌子倾侧，它会绕着连接这两条桌脚底部的直线转动。各桌脚位于边长为 $\frac{1}{2}a$ 的正方形的四个顶点，所以该直线与桌子中心的水平距离为 $\frac{1}{4}a$。

倾侧。对连接远端桌脚底部的直线而言，地面作用于远端桌脚的力没有力矩，地面作用于任何一条桌脚的摩擦力也没有力矩（因为它作用于地面的高度）。作用于中心的向下的力 $10W + 2W = 12W$ 有恢复力矩 $12W \cdot \frac{1}{4}a = 3Wa$。推力 F 则有倾覆力矩 Fa。对这条直线取力矩，离墙最近的两条桌脚所受的法向反作用力（合计）N_1 满足

$$N_1 \cdot \tfrac{1}{2}a = 3Wa - Fa,$$

所以 $N_1 \geq 0$ 当且仅当 $F \leq 3W$。当 $N_1 = 0$ 时，桌子即将倾侧，也就是当

$$\frac{W}{4}\left(1 + \frac{3x}{a}\right) = 3W \quad \Longleftrightarrow \quad x = \frac{11a}{3}.$$

滑动。对桌子沿铅垂方向分解力，地面给予的法向反作用力合共为 $12W$，所以可提供的最大总摩擦力为 $\frac{1}{3} \cdot 12W = 4W$。沿水平方向分解力，所需的摩擦力为 F。所以桌子即将滑动的条件是

$$\frac{W}{4}\left(1 + \frac{3x}{a}\right) = 4W \quad \Longleftrightarrow \quad x = 5a.$$

这两个值都在梯子的长度 $6a$ 之内，而 F 随着油漆工向上爬而不断增加。它先达到 $3W$（在 $x = \frac{11}{3}a$ 时），然后才达到 $4W$（在 $x = 5a$ 时）。所以**桌子先倾侧**。这发生在油漆工沿梯子走了 $\frac{11}{3}a$ 之时，即在桌面之上 $\frac{11}{3}a \sin\alpha = \frac{22a}{3\sqrt{5}} \approx 3.28a$ 的高度。在那一刻，地面所需提供的摩擦力只是 $3W$，小于可提供的 $4W$，所以桌子并未滑动。

验算

- 梯上没有油漆工时（$x = 0$），$F = \frac{1}{4}W$，近墙的桌脚承受 $N_1 = \frac{11}{2}W$，而远端的桌脚承受 $\frac{13}{2}W$。远端桌脚承受的力稍大，这与推力背离墙壁的方向相符。
- 用电脑对多个 k 和 x 的值计算各个力。在 (i) 中，所需摩擦力与可提供摩擦力之比的最大值仅略小于 1（在梯顶时为 $\frac{2k+1}{2k+2}$）。在 (ii) 中，电脑求得近墙桌脚的反作用力在 $x \approx 3.667a$ 时变为零，而地面摩擦力在 $x = 5a$ 时达到极限。

常见错误

- 把倾侧的转轴取在距离中心 $\frac{1}{2}a$ 的桌边：桌面的边长是 $\frac{1}{2}a$，所以半阔度是 $\frac{1}{4}a$。
- 忘记了作用于桌子的推力位于高度 a 处，即桌脚的顶端，而不是在地面的高度。正是这个高度令推力产生倾覆力矩。

- 计算地面对桌子的法向反作用力时，漏掉了梯子和油漆工的 $10W$。只计桌子本身的重量 $2W$ 并不足够。

STEP 之外

这道题探讨的是由多个刚体组成的系统怎样以互相竞争的方式失效：滑动，或是翻倒。方法是求出每种方式的临界值，再看负载先达到哪一个。工程师把同样的想法应用于挡土墙、起重机和堆叠的货物；在静力学（statics）课程中，它会进一步引申到虚功原理（principle of virtual work）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对刚体平衡和摩擦力有细致的论述。

详细解答

STEP II 2006，第 10 题

力学 · 碰撞与恢复系数

> **STEP II 2006，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/06-s2-q10

关键思路

每一次碰撞都是一道两条方程的问题：沿直线方向的动量守恒（conservation of momentum），以及牛顿恢复定律（Newton's law of restitution）（分离速率等于 e 乘以接近速率）。先解 A–B 碰撞，再把 B 的新速度代入 B–C 碰撞。第二次 A–B 碰撞发生，当且仅当位于 B 后方的 A，其向前的速度最终大于 B 在 B 撞上 C 之后的速度。第 (ii) 部分于是成为一道匀速的追及问题。

解答

我们把 A、B 和 C 视为在光滑（smooth）平面上沿一条直线运动的质点（particle），因此除碰撞期间外，并无水平力作用。取 A 被抛射的方向为正方向。以下所有速度均带正负号。

第 (i) 部分

A 与 B 的碰撞。设 A 和 B 在碰撞刚结束时的速度分别为 v_A 和 v_B。由动量守恒和恢复定律（其中 $e = \frac{1}{2}$）得

$$mv_A + km\,v_B = mu, \qquad v_B - v_A = \tfrac{1}{2}u.$$

把 $v_A = v_B - \frac{1}{2}u$ 代入第一条方程，得 $(1+k)v_B = \frac{3}{2}u$，所以

$$v_B = \frac{3u}{2(1+k)}, \qquad v_A = v_B - \frac{u}{2} = \frac{(2-k)\,u}{2(1+k)}.$$

由于 $v_B > 0$，B 确实会继续向 C 移动；又由于 $v_A < v_B$，A 不可能在 B 到达 C 之前追上 B。

B 与 C 的碰撞。设 B 和 C 在碰撞刚结束时的速度分别为 w_B 和 w_C。取 $e = \frac{1}{4}$，

$$km\,w_B + 3m\,w_C = km\,v_B, \qquad w_C - w_B = \tfrac{1}{4}v_B.$$

代入 $w_C = w_B + \frac{1}{4}v_B$，得 $(k+3)w_B = \left(k - \frac{3}{4}\right)v_B$，所以

$$w_B = \frac{(4k-3)\,v_B}{4(k+3)} = \frac{3(4k-3)\,u}{8(1+k)(k+3)}.$$

此外 $w_C > w_B$，所以 B 和 C 互相分开，此后不再相遇。这次碰撞之后，在 A 或 B 身上再无其他作用，直至两者相遇为止（假如真的会相遇）。

> **想一想** A 与 B 再次碰撞的确切条件是甚么？

一旦 B 撞上 C，A 和 B 都以恒定速度运动，而 A 在 B 的后方（即位于 B 的负方向一侧）。两者之间的距离缩短，当且仅当 $v_A > w_B$，而这时距离会在有限时间内缩减至零。若 $v_A \le w_B$，距离永不减少。这一条不等式已涵盖所有正负号的情况：例如，若 A 正在向后运动，但 B 向后运动得更快，B 仍会追上 A，而这同样是 $v_A > w_B$。

所以我们需要

$$\frac{(2-k)u}{2(1+k)} > \frac{3(4k-3)u}{8(1+k)(k+3)}.$$

两边乘以 $8(1+k)(k+3)/u$（由于 $k>0$，此数为正），得到等价的不等式

$$\begin{aligned} 4(2-k)(k+3) &> 3(4k-3) \\ 24 - 4k - 4k^2 &> 12k - 9 \\ 4k^2 + 16k - 33 &< 0 \\ (2k-3)(2k+11) &< 0. \end{aligned}$$

由于 $2k+11>0$，上式成立当且仅当 $2k-3<0$。因为 k 是质量之比，$k>0$。因此 A 与 B 发生第二次碰撞，当且仅当

$$0 < k < \tfrac{3}{2}.$$

当 $k = \frac{3}{2}$ 时，我们得到 $v_A = w_B = \frac{u}{10}$：两者以相同速率一同前进，实际上永不再相遇，所以端点不包括在内。

第 (ii) 部分

取 $k=1$，由上面的公式得

$$v_A = \frac{u}{4}, \qquad v_B = \frac{3u}{4}, \qquad w_B = \frac{3u}{64}, \qquad w_C = \frac{15u}{64}.$$

把原点设在 A 与 B 首次碰撞之处（即 B 的起始位置），并由该次碰撞起计算时间 t。

直至 B–C 碰撞。B 以速率 $\frac{3}{4}u$ 走过与 C 相隔的距离 d，所需时间为

$$t_1 = \frac{d}{3u/4} = \frac{4d}{3u}.$$

在这段时间内，A 移动了 $\frac{u}{4}\cdot\frac{4d}{3u} = \frac{d}{3}$。所以 A 位于 $\frac{d}{3}$，而 B 位于 d：两者相距 $\frac{2d}{3}$。

追赶。现在 A 以 $\frac{u}{4} = \frac{16u}{64}$ 运动，而 B 以 $\frac{3u}{64}$ 运动，所以 A 以相对速率 $\frac{13u}{64}$ 拉近距离。所需时间为

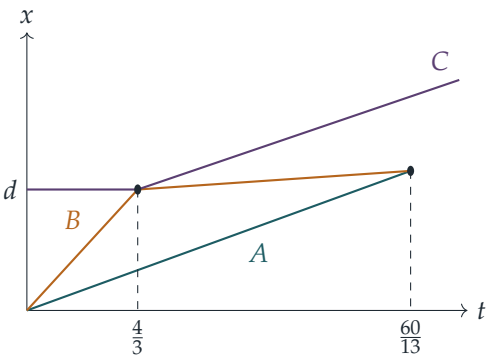
$$t_2 = \frac{2d/3}{13u/64} = \frac{128d}{39u}.$$

总时间。两次 A–B 碰撞之间相隔的时间为

$$t_1 + t_2 = \frac{4d}{3u} + \frac{128d}{39u} = \frac{52d + 128d}{39u} = \frac{180d}{39u} = \frac{60d}{13u},$$

正如所求。

下面的距离–时间图显示了整个运动过程，其中 $u = d = 1$。代表 A 和 B 的两条线在 $t = \frac{60}{13}$ 再次相交。



验算

- 动量：取 $k = 1$，第一次碰撞后 $m\cdot\frac{u}{4} + m\cdot\frac{3u}{4} = mu$，第二次碰撞后 $m\cdot\frac{3u}{64} + 3m\cdot\frac{15u}{64} = \frac{48}{64}mu = m\cdot\frac{3u}{4}$。两者均平衡。
- 边界 $k = \frac{3}{2}$：v_A 和 w_B 都等于 $\frac{u}{10}$，这印证了临界值。
- 合理性检查：当 k 很大时，A 向后反弹（$v_A < 0$），而 B 继续向前，所以显然没有第二次碰撞。这与所得范围吻合。
- 我们以数值方法验证过：对 10^5 个随机有理数 k，$v_A > w_B$ 成立当且仅当 $k < \frac{3}{2}$；另外，取 $u = d = 1$ 作逐步时间模拟，得出第二次碰撞发生于 $t \approx 4.6154$，与 $\frac{60}{13} \approx 4.6154$ 相符。

常见错误

- 在恢复定律中使用速率而非带正负号的速度。使用带正负号的速度时，分离速率就是“前者减后者”，而 v_A 的正负号自然会得到妥善处理。
- 把 $v_A > 0$ 当作条件的一部分。原则上 A 可以向后运动而仍被 B 追上（若 $w_B < v_A$）。唯一的条件 $v_A > w_B$ 才是正确的条件。
- 在第 (ii) 部分，由第一次碰撞起计算追赶时间，而不是由 B 的速度改变的一刻起计算，因而漏掉了 $\frac{d}{3}$ 这段 A 早已走过的领先距离。

STEP 之外

动量守恒与恢复定律合起来，使每一次碰撞都成为由碰撞前速度到碰撞后速度的线性映射（linear map）。一连串的碰撞，例如牛顿摆（Newton's cradle），或是计算两个方块与一面墙之间碰撞次数的著名问题，便可通过复合这些映射来研究。在大学一年级的力学中，这会引伸至质心参考系（centre-of-mass frame），在其中恢复定律变成相对速度的简单缩放，并进一步推展至二维和三维的碰撞。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）对碰撞和质心参考系有清晰的论述。

详细解答

STEP II 2006，第 14 题

概率与统计 · 连续随机变量

> **STEP II 2006，第 14 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/06-s2-q14

关键思路

整道题都建基于一个原函数 (antiderivative)。分母 $x\ln x$ 是 $\ln x$ 乘以 x，而 $1/x$ 正是 $\ln x$ 的导数，所以被积函数具有 $g'(x)/g(x)$ 的形式，其中 $g(x)=\ln x$。因此

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

此后，每一部分都只需运用“总概率为 1”，并小心留意正负号，以及密度（density）实际上在哪里不为零。

解答

草图

设 $y=\dfrac{1}{x\ln x}$，其中 $x>0$，$x\neq 1$。

正负号。由于 $x>0$，y 的正负号与 $\ln x$ 相同：在 $0<x<1$ 上 $y<0$，在 $x>1$ 上 $y>0$。曲线与两条坐标轴都不相交。

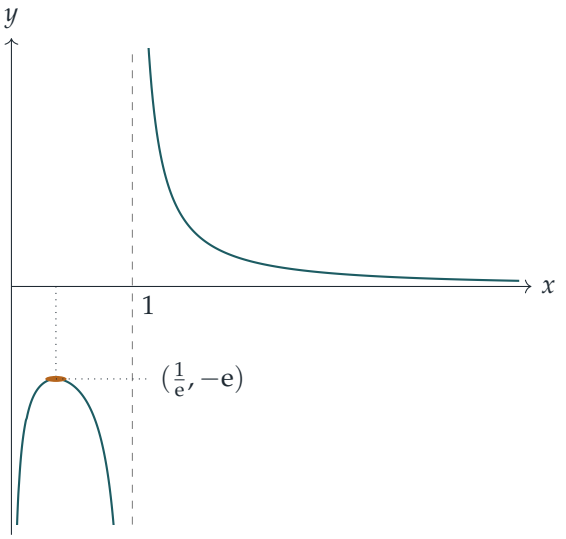
两端的变化。

- 当 $x\to 0^+$ 时：题目告诉我们 $x\ln x\to 0$，而且在该处 $x\ln x<0$，所以 $y\to -\infty$。
- 当 $x\to 1$ 时：$x\ln x\to 0$，从左边趋近时为负，从右边趋近时为正，所以当 $x\to 1^-$ 时 $y\to -\infty$，当 $x\to 1^+$ 时 $y\to +\infty$。直线 $x=1$ 是垂直渐近线（vertical asymptote）。
- 当 $x\to\infty$ 时：$x\ln x\to\infty$，所以 $y\to 0^+$，而 x 轴是渐近线。

转向点（*turning points*）。由商法则（quotient rule）或链式法则（chain rule），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\ln x + 1}{(x \ln x)^2},$$

它只在 $\ln x = -1$ 时为零，即 $x = 1/\mathrm{e}$，此时 $y = \dfrac{1}{(1/\mathrm{e})(-1)} = -\mathrm{e}$。当 $0 < x < 1/\mathrm{e}$ 时，我们有 $\ln x + 1 < 0$，所以 y 递增；当 $x > 1/\mathrm{e}$（且 $x \neq 1$）时，函数递减。因此 $(1/\mathrm{e}, -\mathrm{e})$ 是左边分支的局部极大值（local maximum）点，而右边分支则由 $+\infty$ 一直递减至 0。



第 (i) 部分

这里 $a = \frac{1}{4}$ 而 $b = \frac{1}{2}$，两者都在 $(0,1)$ 内，而在该区间上 $x \ln x < 0$。利用上面的原函数，

$$\begin{aligned}\int_{1/4}^{1/2} \frac{\mathrm{d}x}{x \ln x} &= \Big[\ln|\ln x|\Big]_{1/4}^{1/2} = \ln|\ln \tfrac{1}{2}| - \ln|\ln \tfrac{1}{4}| \\ &= \ln(\ln 2) - \ln(2 \ln 2) = -\ln 2.\end{aligned}$$

总概率必须为 1，所以 $\lambda \cdot (-\ln 2) = 1$，由此得

$$\lambda = -\frac{1}{\ln 2}.$$

这个负号正是我们所需要的：在 $[\frac{1}{4}, \frac{1}{2}]$ 上 $x \ln x < 0$，所以 λ 必须为负，f 在该处才会非负。

第 (ii) 部分

当 $\lambda = 1$ 时，密度必须非负，所以区间 $[a, b]$ 必须位于 $x \ln x > 0$ 的范围内；由于 $b > a > 1$，这一点确实成立。于是 $\ln a, \ln b > 0$，并且

$$1 = \int_a^b \frac{\mathrm{d}x}{x \ln x} = \ln(\ln b) - \ln(\ln a) = \ln \frac{\ln b}{\ln a}.$$

所以 $\dfrac{\ln b}{\ln a} = \mathrm{e}$，即 $\ln b = \mathrm{e} \ln a = \ln\left(a^{\mathrm{e}}\right)$；又因为 ln 是一对一（one-to-one）的，故 $b = a^{\mathrm{e}}$，正如所求。

第 (iii) 部分

取 $a = \mathrm{e}$，由第 (ii) 部分得 $b = \mathrm{e}^{\mathrm{e}}$。由于 $1 < \frac{3}{2} < 2 < \mathrm{e}$，整个区间 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 都位于 $[a, b] = [\mathrm{e}, \mathrm{e}^{\mathrm{e}}]$ 之内，所以

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = \Big[\ln(\ln x)\Big]_{\mathrm{e}^{3/2}}^{\mathrm{e}^2} = \ln 2 - \ln \tfrac{3}{2} = \ln \tfrac{4}{3}.$$

要把这个结果化成分数，可使用麦克劳林级数（Maclaurin series）$\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \frac{t^4}{4} + \cdots$，取 $t = \frac{1}{3}$，并保留四项：

$$\ln \tfrac{4}{3} \approx \frac{1}{3} - \frac{1}{18} + \frac{1}{81} - \frac{1}{324} = \frac{108 - 18 + 4 - 1}{324} = \frac{93}{324} = \frac{31}{108}.$$

想一想 我们怎样知道取到第四项就停下来是合理的?

这个级数正负项交替出现，而且各项的大小递减，所以取四项之后的误差小于第五项，即 $\frac{1}{5 \cdot 3^5} = \frac{1}{1215} < 0.001$。因此 $\mathrm{P} = \ln \frac{4}{3} \approx \frac{31}{108}$，正如所求。

第 (iv) 部分

现在 $a = \mathrm{e}^{1/2}$，所以由第 (ii) 部分得

$$b = a^{\mathrm{e}} = \mathrm{e}^{\mathrm{e}/2}.$$

想一想 在积分之前，先想想：密度实际上在哪里不为零?

由于 $\mathrm{e} < 3$，我们有 $\mathrm{e}/2 < \frac{3}{2}$，所以 $b = \mathrm{e}^{\mathrm{e}/2} < \mathrm{e}^{3/2}$。区间 $[\mathrm{e}^{3/2}, \mathrm{e}^2]$ 完全位于 b 的右边，而在那里 $\mathrm{f}(x) = 0$。因此

$$\mathrm{P}(\mathrm{e}^{3/2} \le X \le \mathrm{e}^2) = 0.$$

（若不加思索地把 $1/(x \ln x)$ 由 $\mathrm{e}^{3/2}$ 积分至 e^2，便会再次得到 $\ln \frac{4}{3}$，而这是错的。）

验算

- 以数值计算，$e^{-1} \approx 0.368$，而在 $(0,1)$ 上逐点检视 y 的值，可见它在该区间上的最大值 $\approx -2.718 = -e$。
- 辛普森法则（Simpson's rule）给出 $\int_{1/4}^{1/2} \frac{dx}{x \ln x} \approx -0.6931 = -\ln 2$，所以 $\lambda \approx -1.4427$。
- 对于若干个 $a > 1$ 的值（例如 $a = 2$、$a = 7.3$），$\int_a^{a^e} \frac{dx}{x \ln x}$ 的结果都是 1.0000。
- $\ln \frac{4}{3} \approx 0.28768$ 而 $\frac{31}{108} \approx 0.28704$，两者相差不超过 0.0007。
- 在 (iv) 中，$b = e^{e/2} \approx 3.89$，而 $e^{3/2} \approx 4.48$。

常见错误

- 在第 (i) 部分写成 $\ln(\ln x)$，但那里 $\ln x < 0$：原函数需要加上绝对值（modulus），即 $\ln|\ln x|$。
- 答 $\lambda = \frac{1}{\ln 2}$：正的 λ 会令“密度”变成负数。
- 在第 (iv) 部分，没有先检查 $[e^{3/2}, e^2]$ 是否位于支撑集（support）$[a,b]$ 之内，便直接在这个区间上积分。

STEP 之外

代换（substitution）$u = \ln x$ 把这个密度变成 $[\ln a, \ln b]$ 上的 $1/u$，而条件 $b = a^e$ 其实就是 $\int_{\ln a}^{\ln b} \frac{du}{u} = 1$。像这样对随机变量（random variable）作变换（这里是 $U = \ln X$），是大学一年级概率论的核心工具；在那里，随机变量的函数的密度是由链式法则求出的。同一个函数 $1/(x \ln x)$ 在分析学中也十分有名：$\sum 1/(n \ln n)$ 发散（diverges），这可以通过与这个积分比较而得出，尽管它的各项趋向零。关于随机变量的变换，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*。

详细解答

STEP III 2006，第 1 题

纯数学 · 曲线描绘

STEP III 2006，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/06-s3-q1

关键思路

先做除法：把曲线写成一条直线加上一个细小的修正项，便可得出斜渐近线（oblique asymptote），知道曲线位于渐近线的哪一侧，而且求导数（derivative）也变得容易。结果导数处处大于 2，而单凭这一个事实，便足以完成数根部分的大半工作：每条方程都可以改写成“曲线 = 某个简单的式子”，然后比较斜率。

解答

曲线

由于 $2x^3 - 10x = 2x(x^2 - 4) - 2x$，

$$y = \frac{2x(x^2-5)}{x^2-4} = 2x - \frac{2x}{x^2-4}.$$

有些特征可以即时看出：

- y 是奇函数（odd function），所以曲线关于原点有 2 重旋转对称（rotational symmetry）。
- 曲线在 $x = 0$ 和 $x = \pm\sqrt{5}$ 处与 x 轴相交。
- **渐近线**。分母在 $x = \pm 2$ 处为零（该处分子为 $\pm 4 \cdot (-1) \cdot 2 \neq 0$），所以 $x = 2$ 和 $x = -2$ 是铅垂渐近线（vertical asymptote）。当 $|x| \to \infty$ 时，修正项 $2x/(x^2-4) \to 0$，所以 $y = 2x$ 是斜渐近线。
- **在渐近线的哪一侧**。$y - 2x = -2x/(x^2-4)$。当 $x > 2$ 时，此式为负（曲线略低于 $y = 2x$）；当 $x < -2$ 时为正（略高于它）。当 $0 < x < 2$ 时为正，当 $-2 < x < 0$ 时为负。

- **在 $x = \pm 2$ 附近的表现**。当 $x \to 2^+$ 时，$-2x/(x^2-4) \to -\infty$，而当 $x \to 2^-$ 时它趋向 $+\infty$；由奇函数的性质，当 $x \to -2^-$ 时 $y \to +\infty$，当 $x \to -2^+$ 时 $y \to -\infty$。

对除法后的形式求导，

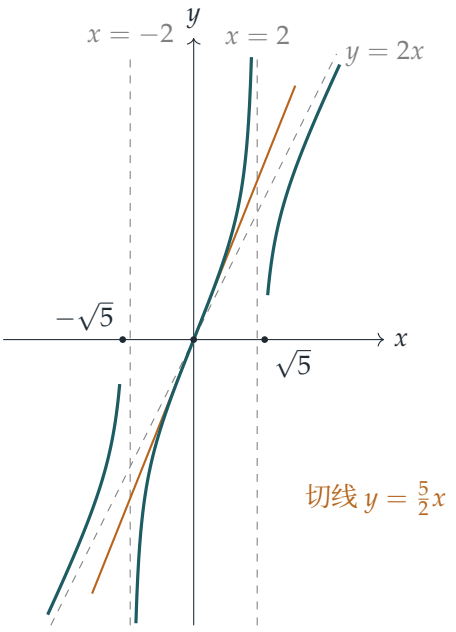
$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2 - \frac{2(x^2-4) - 2x \cdot 2x}{(x^2-4)^2} = 2 + \frac{2(x^2+4)}{(x^2-4)^2}.$$

所以对任何 $x \neq \pm 2$ 都有 $\mathrm{d}y/\mathrm{d}x > 2$：曲线没有驻点（stationary point），而且在 $x < -2$、$-2 < x < 2$、$x > 2$ 这三个区间的每一个上，曲线都严格递增（strictly increasing），由 $-\infty$ 升至 $+\infty$。

原点处的切线（tangent）。在 $x = 0$ 处斜率为 $2 + 8/16 = \frac{5}{2}$，所以切线是 $y = \frac{5}{2}x$。再求导一次，

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = -\frac{4x(x^2+12)}{(x^2-4)^3},$$

它只在 $x = 0$ 处为零，并且在该处变号：原点是拐点（point of inflection）。在 $0 < x < 2$ 上二阶导数（second derivative）为正（分子和分母皆为负），所以中间的分支在该处是凸（convex）的，位于其切线 $y = \frac{5}{2}x$ 的上方；在 $-2 < x < 0$ 上它是凹（concave）的，位于切线的下方。



小结。渐近线为 $x = 2$、$x = -2$ 和 $y = 2x$；原点处的切线为 $y = \frac{5}{2}x$。

在以下每一部分中，$x = \pm 2$ 都不会是根（该处左方不为零而右方为零），所以我们可以除以 $x^2 - 4$，然后数算曲线与另一个图像的交点。

第 (i) 部分

除以 $\frac{3}{2}(x^2-4)$，方程变成

$$\frac{2x(x^2-5)}{x^2-4} = \frac{2}{3}(x+3),$$

所以我们要找曲线与直线 $y = \frac{2}{3}x + 2$ 的交点，该直线的斜率为 $\frac{2}{3}$。

想一想 为甚么斜率小于 2 的直线与每个分支都恰好相交一次?

设 $h(x)$ 为曲线的 y 值减去直线的对应值。在每个分支上 $h'(x) > 2 - \frac{2}{3} > 0$，所以 h 严格递增，最多只有一个零点。在靠近铅垂渐近线的两端，$h \to \pm\infty$，与 y 的情况一样（直线保持有限）。在外侧的两端，$h \approx 2x - \frac{2}{3}x = \frac{4}{3}x$，当 $x \to -\infty$ 时趋向 $-\infty$，当 $x \to +\infty$ 时趋向 $+\infty$。所以在三个分支的每一个上，h 都由 $-\infty$ 走到 $+\infty$，根据中间值定理（intermediate value theorem），它恰好有一个零点。

因此方程 (i) 有三个实根。（事实上它就是三次方程 $2x^3 - 3x^2 - 11x + 12 = 0$，其中一个根是 $x = 1$。）

第 (ii) 部分

除以 $2(x^2-4)$，我们需要找曲线与 $y = \frac{5}{2}x - 1$ 的交点：这是把原点处的切线向下平移 1。它的斜率现在大于 2，所以 (i) 的论证不再适用，我们要逐个分支考察。

- $x > 2$：曲线位于 $y = 2x$ 的下方，而直线满足 $\frac{5}{2}x - 1 - 2x = \frac{1}{2}x - 1 > 0$，所以它位于 $y = 2x$ 的上方。没有交点。
- $x < -2$：曲线位于 $y = 2x$ 的上方，而由于 $\frac{1}{2}x - 1 < 0$，直线位于其下方。没有交点。
- $0 \le x < 2$：曲线位于其切线 $y = \frac{5}{2}x$ 之上或恰在其上（凸性），而切线又在直线的上方。没有交点。
- $-2 < x < 0$：当 $x \to -2^+$ 时，曲线趋向 $-\infty$，而直线趋向 -6；在 $x = 0$ 处曲线的值为 0，高于直线的 -1。所以至少有一个交点。因为这里 $y'' < 0$，y' 在 $(-2, 0)$ 上递减，所以 $y' > y'(0) = \frac{5}{2}$；“曲线减直线”这个差严格递增，故交点唯一。

因此方程 (ii) 恰好有一个实根，位于 -2 与 0 之间。（写成三次方程就是 $x^3 - 2x^2 + 8 = 0$，它的驻点值 8 和 $\frac{184}{27}$ 都是正数，由此确认只有一个实根，约在 $x \approx -1.51$ 附近。）

第 (iii) 部分

除以 $(x^2-4)^2$ 得

$$\left(\frac{2x(x^2-5)}{x^2-4}\right)^2 = x^2 + 1,$$

所以我们需要找曲线与等轴双曲线（rectangular hyperbola）$y^2 - x^2 = 1$ 的交点，即与 $y = \sqrt{x^2+1}$（上支）或 $y = -\sqrt{x^2+1}$（下支）的交点。

> ? **想一想** 可以利用对称性把工作减半吗?

可以。映射 $(x, y) \mapsto (-x, -y)$ 把曲线映回自身（因为它是奇函数），并把双曲线的两支互换。所以曲线与下支的交点和它与上支的交点一一对应，我们只需数算后者。

设 $u(x) = \sqrt{x^2+1}$。那么 $|u'(x)| = |x|/\sqrt{x^2+1} < 1$，所以在曲线的每个分支上，差 $y - u$ 的导数都大于 $2 - 1 = 1 > 0$，每个分支最多给出一个交点。现在检查各端：

- $-2 < x < 2$：$y - u$ 由 $-\infty$ 走到 $+\infty$。一个根。
- $x > 2$：当 $x \to 2^+$ 时 $y - u \to -\infty$，而当 $x \to \infty$ 时 $y - u \approx 2x - x \to +\infty$。一个根。
- $x < -2$：当 $x \to -\infty$ 时 $y - u \approx 2x - |x| = 3x \to -\infty$，而当 $x \to -2^-$ 时 $y - u \to +\infty$。一个根。

所以曲线与上支相交三次，而根据对称性，与下支也相交三次。两支没有公共点（$u \geq 1$），所以这六个点有六个不同的 x 值，方程 (iii) 有**六**个实根。由于它是一条 6 次多项式方程（即 $3x^6 - 33x^4 + 92x^2 - 16 = 0$），它的所有根都是实数。

验算

- 在 $x = 0$ 附近，$y \approx 2x(-5)/(-4) = \frac{5}{2}x$，与切线吻合。
- 在 $x = 1000$ 处，$y - 2x \approx -0.002$：数值细小而且为负，与预期相符。
- 对原来的方程以数值方法搜索变号位置，可找到 (i) 的根 $x \approx -2.212,\ 1,\ 2.713$；(ii) 只有 $x \approx -1.510$；而 (iii) 的根为 $x \approx \pm 0.431,\ \pm 2.155,\ \pm 2.484$。

常见错误

- 在除以 $x^2 - 4$ 之前，忘记检查 $x = \pm 2$ 不是根。
- 在 (iii) 中，只数算与 $y = \sqrt{x^2+1}$ 的交点，遗漏了负平方根。
- 在 (ii) 中，像 (i) 那样假设“直线够斜，所以每个分支一个根”：当斜率大于 2 时，外侧的分支需要另作论证。

STEP 之外

这里的数根方法，是中间值定理与单调性（monotonicity）的结合：一个连续、严格递增而且由 $-\infty$ 走到 $+\infty$ 的函数，恰好有一个零点。在大学一年级的分析课程中，这两个事实都会从实数的完备性（completeness）出发严格证明，而同样的推理也是“每个奇数次实系数多项式都有一个实根”这类结果的基础。Abbott 的 *Understanding Analysis* 对此有非常精彩的阐述。

详细解答

STEP III 2006，第 2 题

纯数学 · 利用对称性求定积分

STEP III 2006，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/06-s3-q2

关键思路

区间 $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ 是对称的，所以把 θ 换成 $-\theta$ 便得到 I 的第二个表达式；两式相加，分母中那个碍事的 $\sin\theta$ 就会消失。一旦把所有式子都用 $\tan\theta$ 表示，积分（integral）I 和 J 便有相同的分母，而两者经适当选取的组合，其被积函数（integrand）恰好是 1。整个过程中，要时刻留意 $\cos 2\alpha$ 的正负号：最后一部分的变化正在于此。

解答

设 $s = \sin 2\alpha$ 及 $c = \cos 2\alpha$，则 $s^2 + c^2 = 1$。在所考虑的两个 α 范围内（$\alpha \neq \frac{1}{4}\pi$），都有 $0 < s < 1$，所以 $1 \mp s\sin\theta \geq 1 - s > 0$：$I$ 的被积函数在整个区间上连续（continuous）。

第 (i) 部分

作代换（substitution）$\theta = -\phi$，则 $\mathrm{d}\theta = -\mathrm{d}\phi$，而上下限互换。由于 $\cos^2(-\phi) = \cos^2\phi$ 及 $\sin(-\phi) = -\sin\phi$，

$$I = \int_{\frac{1}{2}\pi}^{-\frac{1}{2}\pi} \frac{\cos^2\phi}{1 + \sin\phi\, s}\,(-\mathrm{d}\phi) = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{\cos^2\phi}{1 + s\sin\phi}\,\mathrm{d}\phi,$$

这就是第一个结果（把 ϕ 改写为 θ 即可）。

把 I 的两个表达式相加，

$$2I = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos^2\theta \left(\frac{1}{1 - s\sin\theta} + \frac{1}{1 + s\sin\theta}\right)\mathrm{d}\theta = \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \frac{2\cos^2\theta}{1 - s^2\sin^2\theta}\,\mathrm{d}\theta.$$

现在把分子和分母同时除以 $\cos^2\theta$（在区间内部不为零）：

$$\frac{2\cos^2\theta}{1-s^2\sin^2\theta}=\frac{2}{\sec^2\theta-s^2\tan^2\theta}=\frac{2}{1+\tan^2\theta\,(1-s^2)}=\frac{2}{1+c^2\tan^2\theta},$$

其中用到 $\sec^2\theta=1+\tan^2\theta$ 及 $1-\sin^2 2\alpha=\cos^2 2\alpha$。因此

$$2I=\int_{-\frac12\pi}^{\frac12\pi}\frac{2}{1+\tan^2\theta\cos^2 2\alpha}\,\mathrm{d}\theta,$$

正如所求。（在 $\theta=\pm\frac12\pi$ 处，这个被积函数趋向 0，所以在两个端点不会出问题。）

第 (ii) 部分

分子中的 $\sec^2\theta$ 提示我们采用 $t=\tan\theta$，$\mathrm{d}t=\sec^2\theta\,\mathrm{d}\theta$。当 θ 由 $-\frac12\pi$ 变到 $\frac12\pi$ 时，t 走遍整条实数线：

$$J=\int_{-\infty}^{\infty}\frac{\mathrm{d}t}{1+c^2t^2}.$$

由于 $0<\alpha<\frac14\pi$，我们有 $0<2\alpha<\frac12\pi$，所以 $c>0$。设 $u=ct$，

$$J=\frac1c\int_{-\infty}^{\infty}\frac{\mathrm{d}u}{1+u^2}=\frac1c\Big[\arctan u\Big]_{-\infty}^{\infty}=\frac{\pi}{c}=\pi\sec 2\alpha.$$

第 (iii) 部分

由第 (i) 部分，$I=\displaystyle\int_{-\frac12\pi}^{\frac12\pi}\frac{\mathrm{d}\theta}{1+c^2\tan^2\theta}$，它与 J 有相同的分母。所以

$$s^2I+c^2J=\int_{-\frac12\pi}^{\frac12\pi}\frac{s^2+c^2\sec^2\theta}{1+c^2\tan^2\theta}\,\mathrm{d}\theta.$$

想一想 为甚么偏偏这个组合会比较容易处理?

看看分子：$s^2+c^2\sec^2\theta=s^2+c^2+c^2\tan^2\theta=1+c^2\tan^2\theta$。它恰好就是分母，所以被积函数是 1，而

$$s^2I+c^2J=\pi.$$

利用第 (ii) 部分的 $J=\pi/c$，

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

其中用到 $s^2=1-c^2=(1-c)(1+c)$，并约去了 $1-c\neq0$（因为 $c<1$）。最后，$1+\cos 2\alpha=2\cos^2\alpha$，所以

$$I=\frac{\pi}{2\cos^2\alpha}=\tfrac12\pi\sec^2\alpha.\qquad\checkmark$$

第 (iv) 部分

现在 $\frac14\pi<\alpha<\frac12\pi$，所以 $\frac12\pi<2\alpha<\pi$：仍然有 $0<s<1$，但这次 $c<0$。第 (i) 部分只用到 $s^2+c^2=1$，所以依然成立；第 (iii) 部分的恒等式（identity）$s^2I+c^2J=\pi$ 亦然。唯一改变的是第 (ii) 部分。

> **想一想** 第 (ii) 部分究竟在哪里用到了 $\alpha<\frac14\pi$?

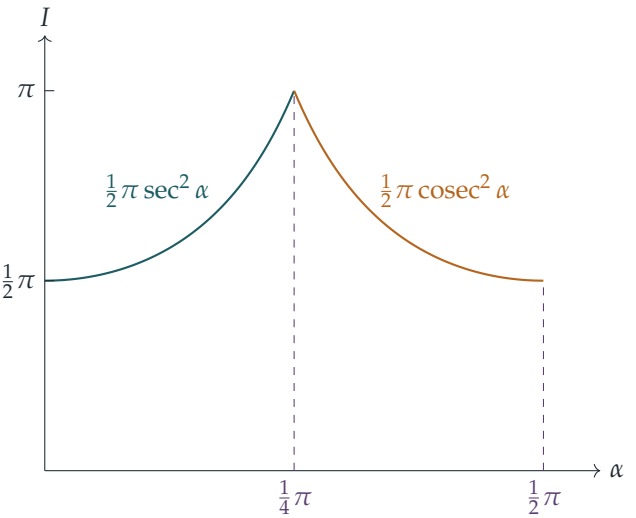
就在代换 $u=ct$ 之中：若 $c<0$，u 的上下限会变成由 $+\infty$ 到 $-\infty$，把它们换回来便多出一个负号。换句话说，$J=\pi/|c|=-\pi/c$。（这是必然的：J 是一个正函数的积分，所以不可能等于负数 $\pi\sec 2\alpha$。）于是

$$s^2I=\pi-c^2\Big(-\frac{\pi}{c}\Big)=\pi(1+c)\quad\Longrightarrow\quad I=\frac{\pi(1+c)}{(1-c)(1+c)}=\frac{\pi}{1-c},$$

其中约去了 $1+c\neq0$（因为 $c>-1$）。由于 $1-\cos 2\alpha=2\sin^2\alpha$，

$$I=\frac{\pi}{2\sin^2\alpha}=\tfrac12\pi\operatorname{cosec}^2\alpha\qquad\big(\tfrac14\pi<\alpha<\tfrac12\pi\big).$$

有一个快捷的方法可以核实：I 只通过 $\sin 2\alpha$ 依赖于 α，而 $\sin 2(\frac12\pi-\alpha)=\sin 2\alpha$。所以 $I(\alpha)=I(\frac12\pi-\alpha)=\frac12\pi\sec^2(\frac12\pi-\alpha)=\frac12\pi\operatorname{cosec}^2\alpha$。$I$ 的图像关于 $\alpha=\frac14\pi$ 对称，而两条公式在该处都给出 π：



验算

- 在 $\alpha=\frac14\pi$（$c=0$）时，第 (i) 部分的 2I 形式的被积函数是 2，得出 $I=\pi$，与两条公式都吻合。

- 当 $\alpha \to 0$ 时，$I \to \int \cos^2\theta \,\mathrm{d}\theta = \frac{1}{2}\pi$，与 $\frac{1}{2}\pi \sec^2 0$ 吻合。
- 数值计算方面（对原来的积分使用辛普森法则（Simpson's rule）），$\alpha = 0.5$ 给出 $I \approx 2.039595$，等于 $\frac{1}{2}\pi \sec^2 0.5$；而 $\alpha = 1.1$ 给出 $I \approx 1.977709 = \frac{1}{2}\pi \operatorname{cosec}^2 1.1$。在所有测试过的情况中，$s^2 I + c^2 J$ 的结果都是 3.141593。

常见错误

- 在第 (iv) 部分仍然写 $J = \pi \sec 2\alpha$：这个值在该范围内是负数，而正函数的积分不可能是负数。
- 忘记了代换 $t = \tan\theta$ 会把有限区间变成整条实数线，所以上下限变成 $\pm\infty$。

STEP 之外

第 (i) 部分的技巧，即取 $f(\theta)$ 与 $f(-\theta)$ 的平均，正是把函数分解为偶部和奇部（even and odd parts）的起点：在对称区间上积分，只有偶部会留下来。$\sin\theta$ 和 $\cos\theta$ 的有理函数（rational function）在一个完整周期上的积分，是大学第一门复分析（complex analysis）课程中围道积分（contour integration，即留数定理 the residue theorem）的经典应用，像本题这样的结果在那里寥寥数行便可得出；参阅 Needham 的 *Visual Complex Analysis*（《可视化复分析》）。

详细解答

STEP III 2006，第 3 题

纯数学 · 幂级数的系数

STEP III 2006，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/06-s3-q3

关键思路

我们完全不必反复对 $\tan x$ 求导。每一条三角恒等式（trigonometric identity）都把其中两个函数连系起来，因而化为它们的系数（coefficient）之间的一条关系式。只要比较 x^n 的系数，三个级数就全由 b_n 这一个数列决定。接着，毕氏恒等式（Pythagorean identity）$\cot^2 x + 1 = \operatorname{cosec}^2 x$ 给出 b_n 必须满足的方程，我们可以逐个解出。

解答

第 (i) 部分

若一个幂级数（power series）在 0 附近收敛（converge）至某函数，它就是该函数的麦克劳林级数（Maclaurin series），所以 $a_n = \tan^{(n)}(0)/n!$。而 tan 是奇函数（odd function）：$\tan(-x) = -\tan x$。对奇函数求导得到偶函数（even function），对偶函数求导则得到奇函数。因此 tan 的每个偶数阶导数都是奇函数，而奇函数在 0 处的值为零。故此当 n 为偶数时 $\tan^{(n)}(0) = 0$，于是当 n 为偶数时 $a_n = 0$。换言之，这个级数只含 x 的奇数次幂。同样的论证适用于 $\cot x - \frac{1}{x} = \sum b_n x^n$，它也是奇函数，所以当 n 为偶数时 $b_n = 0$。我们稍后亦会从下面的关系式得出这个结论。

至于恒等式，把各项通分，以 $\sin x \cos x$ 为分母：

$$\cot x - \tan x = \frac{\cos^2 x - \sin^2 x}{\sin x \cos x} = \frac{\cos 2x}{\frac{1}{2}\sin 2x} = 2\cot 2x.$$

现在取 $0 < x < \frac{1}{4}\pi$，使 $2x$ 同样落在 $(0, \frac{1}{2}\pi)$ 之内，于是可以使用 $\cot 2x$ 的级数，即以 $2x$

代替 x：

$$2\cot 2x = 2\left(\frac{1}{2x} + \sum_{n=0}^{\infty} b_n 2^n x^n\right) = \frac{1}{x} + \sum_{n=0}^{\infty} 2^{n+1} b_n x^n.$$

恒等式的左方是

$$\cot x - \tan x = \frac{1}{x} + \sum_{n=0}^{\infty} (b_n - a_n) x^n.$$

两边的 $\frac{1}{x}$ 项互相抵消，余下两个幂级数，它们对 $(0, \frac{1}{4}\pi)$ 内的每个 x 都相等。

想一想 为甚么我们可以比较两个无穷级数的系数?

若两个幂级数在整个区间上有相同的和，它们的系数便一一相等。把两者相减：所得的差是一个在整个区间上恒等于零的幂级数。于是它的各阶导数在该区间上都是零，再令 $x \to 0^+$，便知每个系数都是零。所以

$$b_n - a_n = 2^{n+1} b_n, \qquad \text{即} \qquad a_n = (1 - 2^{n+1})\, b_n,$$

这正是所要证明的。由于 $1 - 2^{n+1} \neq 0$，我们有 $b_n = a_n/(1 - 2^{n+1})$，所以每当 $a_n = 0$ 时 $b_n = 0$，也就是当 n 为偶数时。

第 (ii) 部分

同理,

$$\cot x + \tan x = \frac{\cos^2 x + \sin^2 x}{\sin x \cos x} = \frac{1}{\frac{1}{2}\sin 2x} = 2\operatorname{cosec} 2x.$$

当 $0 < x < \frac{1}{4}\pi$ 时，在 cosec 的级数中以 $2x$ 代替 x，得

$$2\operatorname{cosec} 2x = \frac{1}{x} + \sum_{n=0}^{\infty} 2^{n+1} c_n x^n, \qquad \cot x + \tan x = \frac{1}{x} + \sum_{n=0}^{\infty} (b_n + a_n) x^n.$$

比较系数得 $b_n + a_n = 2^{n+1} c_n$。代入第 (i) 部分的 $a_n = (1 - 2^{n+1}) b_n$,

$$2^{n+1} c_n = (2 - 2^{n+1})\, b_n \quad \Longrightarrow \quad c_n = \frac{2 - 2^{n+1}}{2^{n+1}}\, b_n = (2^{-n} - 1)\, b_n,$$

这就是所求的结果。特别地，当 n 为偶数时同样有 $c_n = 0$。

第 (iii) 部分

把 $\cot x$ 和 $\operatorname{cosec} x$ 的级数乘以 x,

$$1 + x\sum_{n=0}^{\infty} b_n x^n = x\cot x, \qquad 1 + x\sum_{n=0}^{\infty} c_n x^n = x\operatorname{cosec} x.$$

所以题目所述的等式就是 $x^2\cot^2 x + x^2 = x^2\operatorname{cosec}^2 x$。这是恒等式 $\cot^2 x + 1 = \operatorname{cosec}^2 x$ 乘以 x^2 的结果，而该恒等式则由 $\cos^2 x + \sin^2 x = 1$ 除以 $\sin^2 x$ 得出。等式由此得证。

现在把它展开。下标为偶数的 b_n 和 c_n 都是零，所以

$$x\cot x = 1 + b_1x^2 + b_3x^4 + \cdots, \qquad x\operatorname{cosec} x = 1 + c_1x^2 + c_3x^4 + \cdots,$$

而等式变成

$$\left(1 + b_1x^2 + b_3x^4 + \cdots\right)^2 + x^2 = \left(1 + c_1x^2 + c_3x^4 + \cdots\right)^2.$$

把两边平方展开，

$$1 + (2b_1 + 1)x^2 + (2b_3 + b_1^2)x^4 + \cdots = 1 + 2c_1x^2 + (2c_3 + c_1^2)x^4 + \cdots.$$

由第 (ii) 部分，$c_1 = (\frac{1}{2} - 1)b_1 = -\frac{1}{2}b_1$ 及 $c_3 = (\frac{1}{8} - 1)b_3 = -\frac{7}{8}b_3$。

x^2 的系数：

$$2b_1 + 1 = 2c_1 = -b_1 \quad\Longrightarrow\quad b_1 = -\tfrac{1}{3}.$$

于是由第 (i) 部分得 $a_1 = (1 - 2^2)b_1 = (-3)(-\frac{1}{3}) = 1$，正如所求。

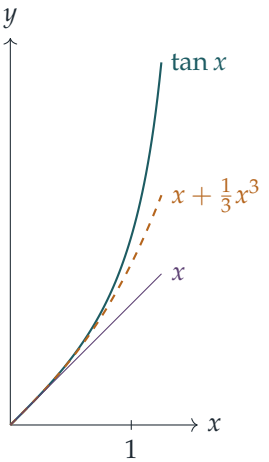
x^4 的系数：代入 $b_1 = -\frac{1}{3}$ 及 $c_1 = \frac{1}{6}$，

$$2b_3 + \tfrac{1}{9} = -\tfrac{7}{4}b_3 + \tfrac{1}{36} \quad\Longrightarrow\quad \tfrac{15}{4}b_3 = -\tfrac{3}{36} = -\tfrac{1}{12} \quad\Longrightarrow\quad b_3 = -\tfrac{1}{45}.$$

因此

$$a_3 = (1 - 2^4)\,b_3 = (-15)\left(-\tfrac{1}{45}\right) = \tfrac{1}{3}.$$

所以 $\tan x = x + \frac{1}{3}x^3 + \cdots$。



验算

- 在 $x = 0.3$ 处，$\tan 0.3 \approx 0.309336$，而 $0.3 + \frac{1}{3}(0.3)^3 = 0.309$。两者相差约 3×10^{-4}，正好是下一项 $\frac{2}{15}x^5$ 的大小。

- $b_1 = -\frac{1}{3}$ 这个值与熟悉的 $\cot x \approx \frac{1}{x} - \frac{x}{3}$ 吻合。此外 $c_1 = \frac{1}{6}$ 与 $\operatorname{cosec} x \approx \frac{1}{x} + \frac{x}{6}$ 吻合，后者可由 $\sin x \approx x - \frac{x^3}{6}$ 得出。
- 我们把 sin 和 cos 的级数相除，以有理数运算精确地验证了三条关系式对 48 以内的每个 n 都成立。我们亦验证了在 $(0, 1)$ 内随机选取的点上，部分和（partial sum）与 tan、cot 和 cosec 的值相符。

常见错误

- 比较系数时忘记了 $\frac{1}{x}$ 项。这些项确实会抵消，但全靠 $2\cot 2x$ 和 $2\operatorname{cosec} 2x$ 中的因子 2。
- 在第 (iii) 部分展开平方时，忘记了 $b_0 = b_2 = 0$。若保留 b_0，便会多出本不存在的 x 的奇数次幂，你还得另行证明它们的系数为零。
- 计算 $c_3 = -\frac{7}{8}b_3$ 时弄错正负号或 2 的幂次。留意 $2^{-3} - 1 = -\frac{7}{8}$。

STEP 之外

把这个方法推演到更高阶，就会发现这些系数原来是伯努利数（Bernoulli numbers）B_{2k}：$x\cot x = \sum_k (-1)^k 2^{2k} B_{2k}\, x^{2k}/(2k)!$。第 (i) 部分的因子 $1 - 2^{n+1}$ 随即给出 $\tan x$ 各系数的一般公式。为甚么 $\tan x$ 的级数到了 $|x| = \frac{1}{2}\pi$ 就不再收敛？复分析（complex analysis）回答了这个问题：幂级数的收敛范围一直延伸到函数趋向无穷的最近一点，对 tan 而言，就是位于 $\pm\frac{1}{2}\pi$ 的极点（pole）。Needham 的 *Visual Complex Analysis* 对此有精彩的解说，而 Abbott 的 *Understanding Analysis* 则为幂级数的逐项比较提供了严格的论证。

详细解答

STEP III 2006，第 4 题

纯数学 · 函数方程

STEP III 2006，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/06-s3-q4

关键思路

开首的恒等式（identity）就是柯西函数方程（Cauchy's functional equation）。代入 $y = x$，便得到 f 在 x 与在 $2x$ 的值之间的关系；把这个关系求导 n 次，再令 $x = 0$，麦克劳林级数（Maclaurin series）的系数便几乎全部消失。其后每一部分都是同一条方程的变奏：只要作一个变量代换（change of variable）（对数、指数、正切），新的恒等式就会变回 $(*)$，于是可以直接套用已经得到的答案。

解答

基本方程

在 $(*)$ 中代入 $y = x$：$\mathrm{f}(x) + \mathrm{f}(x) = \mathrm{f}(2x)$，即

$$2\mathrm{f}(x) \equiv \mathrm{f}(2x).$$

这是关于 x 的恒等式，所以两边可以求导。根据链式法则（chain rule），每对 $\mathrm{f}(2x)$ 求导一次，就多出一个因子 2，因此求导 n 次之后得

$$2\mathrm{f}^{(n)}(x) \equiv 2^n \mathrm{f}^{(n)}(2x).$$

令 $x = 0$，得 $(2 - 2^n)\,\mathrm{f}^{(n)}(0) = 0$。取 $n = 2$，此式即 $-2\mathrm{f}''(0) = 0$，所以 $\mathrm{f}''(0) = 0$，正是所求。

同样的论证对每个 $n \neq 1$ 都成立：因子 $2 - 2^n$ 只在 $n = 1$ 时才等于零，所以

$$\mathrm{f}(0) = 0 \quad \text{及} \quad \mathrm{f}^{(n)}(0) = 0 \ \text{对所有} n \geq 2.$$

（$n=0$ 的情况就是 $2\mathrm{f}(0)=\mathrm{f}(0)$。）麦克劳林级数中只有一次项留下来：

$$\mathrm{f}(x)=\mathrm{f}(0)+\mathrm{f}'(0)\,x+\tfrac{1}{2!}\mathrm{f}''(0)\,x^2+\cdots=kx,\qquad k=\mathrm{f}'(0).$$

反过来，$kx+ky=k(x+y)$，所以每个这样的函数都满足方程。$(*)$ 的通解是 $\mathrm{f}(x)=kx$，其中 k 是任意常数。

第 (i) 部分

要取对数，必须有 $\mathrm{g}>0$，所以先处理正负号的问题。代入 $x=y=\frac{1}{2}s$，得 $\mathrm{g}(s)=\left(\mathrm{g}(\frac{1}{2}s)\right)^2\geq 0$。若对某个 a 有 $\mathrm{g}(a)=0$，则对每个 x，

$$\mathrm{g}(x)=\mathrm{g}(a)\,\mathrm{g}(x-a)=0,$$

所以 $\mathrm{g}\equiv 0$，而它确实满足恒等式。否则对所有 x 都有 $\mathrm{g}(x)>0$，于是 $\mathrm{G}=\ln\mathrm{g}$ 处处有定义。对 $\mathrm{g}(x)\mathrm{g}(y)=\mathrm{g}(x+y)$ 取对数，得

$$\mathrm{G}(x)+\mathrm{G}(y)=\mathrm{G}(x+y),$$

这正是 $(*)$。所以 $\mathrm{G}(x)=kx$，而

$$\mathrm{g}(x)=\mathrm{e}^{kx}$$

其中 k 为常数；它显然满足恒等式，因为 $\mathrm{e}^{kx}\mathrm{e}^{ky}=\mathrm{e}^{k(x+y)}$。除了平凡解 $\mathrm{g}\equiv 0$ 之外，这些就是全部的解。

第 (ii) 部分

每个正数 x 都可以唯一地写成 e^u，其中 u 是实数，即 $u=\ln x$。所以我们可以写 $x=\mathrm{e}^u$ 和 $y=\mathrm{e}^v$，其中 u 和 v 取遍所有实数。这时 $xy=\mathrm{e}^{u+v}$，恒等式变成

$$\mathrm{H}(u)+\mathrm{H}(v)=\mathrm{h}(\mathrm{e}^u)+\mathrm{h}(\mathrm{e}^v)=\mathrm{h}(\mathrm{e}^{u+v})=\mathrm{H}(u+v)$$

对所有实数 u、v 成立。这又是 $(*)$，所以 $\mathrm{H}(u)=ku$。由于 $u=\ln x$，

$$\mathrm{h}(x)=k\ln x\qquad(x>0),$$

而它满足恒等式，因为 $\ln x+\ln y=\ln(xy)$。

第 (iii) 部分

式子 $\dfrac{x+y}{1-xy}$ 其实是乔装了的正切加法公式（tangent addition formula）。设 $x=\tan u$ 和 $y=\tan v$，其中 $u,v\in(-\frac{\pi}{2},\frac{\pi}{2})$，于是 $u=\arctan x$ 及 $v=\arctan y$。那么

$$z=\frac{\tan u+\tan v}{1-\tan u\tan v}=\tan(u+v).$$

在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上定义 $\mathrm{T}(u) = \mathrm{t}(\tan u)$。只要 $u+v$ 也落在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 之内，恒等式就给出

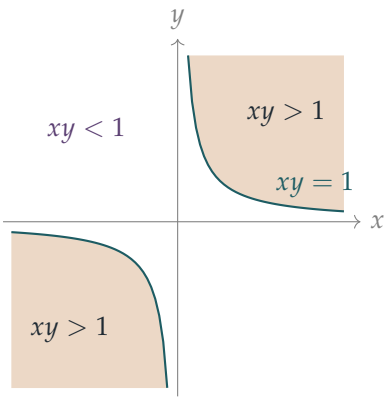
$$\mathrm{T}(u) + \mathrm{T}(v) = \mathrm{t}(x) + \mathrm{t}(y) = \mathrm{t}\big(\tan(u+v)\big) = \mathrm{T}(u+v).$$

想一想 处理 $(*)$ 的论证用到恒等式对所有 x 和 y 成立。如果我们只知道它在 u、v 和 $u+v$ 都属于 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 时成立，论证还行得通吗?

行得通。麦克劳林级数的论证只涉及 0 附近的情况。当 $|u| < \frac{\pi}{4}$ 时可以取 $v = u$（这时 $|2u| < \frac{\pi}{2}$），于是在 0 附近有 $2\mathrm{T}(u) = \mathrm{T}(2u)$。在 $u = 0$ 处求导 n 次，与之前完全一样得到 $(2 - 2^n)\mathrm{T}^{(n)}(0) = 0$，所以 $\mathrm{T}(u) = ku$。由于 $u = \arctan x$，

$$\mathrm{t}(x) = k \arctan x.$$

究竟对哪些 x 和 y 这个答案才真正成立? 若 $xy < 1$，则 $u + v \in (-\frac{\pi}{2}, \frac{\pi}{2})$。当 u 和 v 异号时，这是必然的。当两者皆为正数时，$\tan u \tan v < 1$ 表示 $\tan v < \cot u = \tan(\frac{\pi}{2} - u)$，所以 $u + v < \frac{\pi}{2}$；两者皆为负数的情况由对称性同理可得。因此 $\arctan x + \arctan y = \arctan z$，恒等式成立。然而若 $xy > 1$，$u + v$ 便落在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 之外，这时 $\arctan z = u + v \mp \pi$。例如 $x = y = 2$ 给出 $z = -\frac{4}{3}$，而 $2\arctan 2 - \arctan(-\frac{4}{3}) = \pi$。



所以 $\mathrm{t}(x) = k \arctan x$ 对每一对满足 $xy < 1$ 的数都符合恒等式，也就是图中包含原点的区域。这就是题目想要的答案。在 $xy > 1$ 的阴影区域内，等式两边相差 $\pm k\pi$。如果我们坚持恒等式要对每一对满足 $xy \neq 1$ 的数都毫无例外地成立，那就只剩下 $k = 0$，唯一的解是 $\mathrm{t} \equiv 0$。

验算

- 对于第 (iii) 部分，把恒等式对 y 求导，然后代入 $y = 0$。由于 $\partial z/\partial y = (1 + x^2)/(1 - xy)^2$，在 $y = 0$ 时等于 $1 + x^2$，我们得到 $\mathrm{t}'(0) = \mathrm{t}'(x)(1 + x^2)$。所以 $\mathrm{t}'(x) = k/(1 + x^2)$，再配合 $\mathrm{t}(0) = 0$，积分便得 $k \arctan x$。对 $(*)$ 用同样的技巧，立即得到 $\mathrm{f}'(x) = \mathrm{f}'(0)$。

- 我们对全部四个答案各用 10^4 对随机数作了数值检验。在第 (iii) 部分，$k\arctan x$ 在约 5×10^4 对满足 $xy<1$ 的随机数上全部符合恒等式，而在每一对满足 $xy>1$ 的数上，误差的大小都恰好是 $k\pi$。

常见错误

- 忘记处理常数项：$\mathrm{f}(0)=0$ 来自 $n=0$ 的情况，所以答案是 kx，而不是 $kx+c$。
- 没有检查 $\mathrm{g}>0$ 就取 $\ln\mathrm{g}$。零函数是第 (i) 部分名副其实的一个解。
- 以为 $\arctan x+\arctan y=\arctan\dfrac{x+y}{1-xy}$ 总是成立。它需要 $xy<1$。

STEP 之外

即使不假设麦克劳林级数存在，柯西方程 $(*)$ 所需的条件其实少得多：只要 f 连续（continuous）（甚至只需在某个区间上有界（bounded）），仍然可以推出 $\mathrm{f}(x)=kx$，出发点是对有理数 q 有 $\mathrm{f}(q)=q\,\mathrm{f}(1)$。不过，利用选择公理（axiom of choice），可以构造出并非线性的怪异解，它们的图像在平面上稠密（dense）。第 (i) 部分还说明，指数函数本质上是由加法到乘法的唯一连续同态（homomorphism），这个想法在群论（group theory）中反复出现。想看这类连续性与稠密性论证的严谨处理，Abbott 的 *Understanding Analysis* 是很好的入门读物。

详细解答

STEP III 2006，第 5 题

纯数学 · 复数与几何

STEP III 2006，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/06-s3-q5

关键思路

一个三角形是等边三角形（equilateral triangle），当且仅当从某个顶点看，其中一条边是另一条边旋转 60° 的结果。在复数（complex number）中，这个旋转就是乘以 $\mathrm{e}^{\pm \mathrm{i}\pi/3}$，而这两个数正好是 $t^2 - t + 1 = 0$ 的根。把原点移到其中一个顶点，那个对称式便恰好化为这个二次式。此后，三次方程的部分用韦达公式（Vieta's formulae）即可，而最后一部分只需留意 $z \mapsto pw + q$ 只是把图形旋转、缩放和平移。

解答

等边三角形的条件

设 $u = \beta - \alpha$ 和 $v = \gamma - \alpha$，即由顶点 α 出发的两条边。由于各点互不相同，$u \neq 0$、$v \neq 0$ 且 $u \neq v$。首先，

$$\alpha^2 + \beta^2 + \gamma^2 - \beta\gamma - \gamma\alpha - \alpha\beta = \tfrac{1}{2}\left[(\alpha-\beta)^2 + (\beta-\gamma)^2 + (\gamma-\alpha)^2\right],$$

把右边展开即可验证。现在 $\alpha - \beta = -u$、$\gamma - \alpha = v$ 及 $\beta - \gamma = u - v$，所以方括号内的式子是 $u^2 + (u-v)^2 + v^2 = 2(u^2 - uv + v^2)$。因此

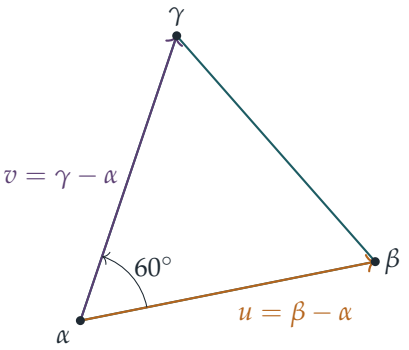
$$\alpha^2 + \beta^2 + \gamma^2 - \beta\gamma - \gamma\alpha - \alpha\beta = u^2 - uv + v^2 = u^2\left(t^2 - t + 1\right), \qquad t = \frac{v}{u}.$$

由于 $u \neq 0$，这个式子等于零当且仅当 $t^2 - t + 1 = 0$，即

$$t = \frac{1 \pm \mathrm{i}\sqrt{3}}{2} = \mathrm{e}^{\pm \mathrm{i}\pi/3}.$$

想一想 为甚么 $\dfrac{\gamma-\alpha}{\beta-\alpha}=\mathrm{e}^{\pm \mathrm{i}\pi/3}$ 恰好表示这个三角形是等边三角形?

v/u 的模（modulus）是 $|\gamma-\alpha|/|\beta-\alpha|$，而它的辐角（argument）是在 α 处由边 $\alpha\beta$ 转到边 $\alpha\gamma$ 的角。所以 $v/u=\mathrm{e}^{\pm \mathrm{i}\pi/3}$ 的意思是：由 α 出发的两条边相等，而它们之间的夹角是 $60°$。这样的等腰三角形，底角是 $\frac{1}{2}(180°-60°)=60°$，所以它是等边三角形。反过来，如果三角形是等边的，则 $|v|=|u|$，而 α 处的角是 $60°$，至于向哪一边转则视乎三角形的定向（orientation），所以 $v/u=\mathrm{e}^{\mathrm{i}\pi/3}$ 或 $\mathrm{e}^{-\mathrm{i}\pi/3}$。



因此，这些互不相同的点（不论次序）构成等边三角形，当且仅当该式等于零，正如题目所要求。

三次方程

设 $(*)$ 的根为 α,β,γ。根据韦达公式，$\alpha+\beta+\gamma=-a$ 及 $\alpha\beta+\beta\gamma+\gamma\alpha=b$。所以

$$\begin{aligned}\alpha^2+\beta^2+\gamma^2-\beta\gamma-\gamma\alpha-\alpha\beta&=(\alpha+\beta+\gamma)^2-3(\alpha\beta+\beta\gamma+\gamma\alpha)\\&=a^2-3b.\end{aligned}$$

根据第一个结果，这些（互不相同的）根构成等边三角形，当且仅当 $a^2-3b=0$，即 $a^2=3b$。

有一点要留意：第一个结果需要各点互不相同。如果有两个根重合，例如 $\alpha=\beta$，该式便化为 $(\gamma-\alpha)^2$，所以只有当三个根全部相等时它才等于零。因此，在有重根（repeated root）的情况下，$a^2=3b$ 只会在三重根时出现，即 $z^3+az^2+bz+c=(z+\frac{a}{3})^3$，亦即同时有 $c=a^3/27$。除了这个只有一点的退化“三角形”之外，$a^2=3b$ 正是所求的条件。

变换

把 $z = pw + q$ 代入 $(*)$，再除以 $p^3 \neq 0$：

$$w^3 + \frac{3q + a}{p}\, w^2 + \frac{3q^2 + 2aq + b}{p^2}\, w + \frac{q^3 + aq^2 + bq + c}{p^3} = 0.$$

所以

$$A = \frac{3q + a}{p}, \qquad B = \frac{3q^2 + 2aq + b}{p^2},$$

因而

$$A^2 - 3B = \frac{(9q^2 + 6aq + a^2) - (9q^2 + 6aq + 3b)}{p^2} = \frac{a^2 - 3b}{p^2}.$$

如果 $(*)$ 的根构成等边三角形，则它们互不相同，而且 $a^2 = 3b$，所以 $A^2 = 3B$。$(**)$ 的根是 $w = (z - q)/p$，其中 z 是 $(*)$ 的根；由于 $p \neq 0$，这个映射是一对一（one-to-one）的，所以这些根同样互不相同。根据上一个结果，$(**)$ 的根构成等边三角形。

从几何角度看，这并不意外：$w = (z - q)/p$ 是平移 $-q$、旋转 $-\arg p$ 再缩放 $1/|p|$ 倍，而这些运算都不会破坏等边三角形。用第一部分的说法，每条边都除以 p，所以比 v/u 不变。

验算

- 单位立方根（cube roots of unity）$1, \omega, \omega^2$ 构成等边三角形，而且是 $z^3 - 1 = 0$ 的根，其中 $a = b = 0$，所以 $a^2 = 3b$ 成立。
- 取 $\alpha = 0$、$\beta = 1$ 及 $\gamma = \frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i}$，该式为 $1 + \gamma^2 - \gamma$，它等于零，因为 $\gamma = \mathrm{e}^{\mathrm{i}\pi/3}$ 是 $t^2 - t + 1$ 的根。
- 以 1000 组随机三元组作数值验算，证实了对于以这些数为根的三次方程，$\sum \alpha^2 - \sum \alpha\beta = a^2 - 3b$ 成立。同一验算亦证实了 $A^2 - 3B = (a^2 - 3b)/p^2$，以及在随机选取的 p 和 q 之下，构成等边三角形的根仍然构成等边三角形。

常见错误

- 只检查一种定向：$\mathrm{e}^{\mathrm{i}\pi/3}$ 和 $\mathrm{e}^{-\mathrm{i}\pi/3}$ 都会出现，而“当且仅当”需要两者兼顾。
- 忘记了正是因为 $u \neq 0$（各点互不相同），我们才可以除以 u^2。
- 代入时把 B 算错：w 的系数包括来自立方项的 $3q^2$、来自平方项的 $2aq$，以及 b，全部再除以 p^2。

STEP 之外

映射 $z \mapsto pz + q$ 就是平面上的正向相似变换（direct similarity）。在这些映射下保持不变的量，例如比 $(\gamma - \alpha)/(\beta - \alpha)$，或者 $a^2 - 3b$ 是否等于零，称为不变量（invariant），这是几何和代数中的核心概念。$a^2 - 3b$ 这个式子（除了一个常数因子外）就是 f' 的判别式（discriminant），其中 $f(z) = z^3 + az^2 + bz + c$。这又连系到一个漂亮的事实——Marden 定理（Marden's theorem）：以 f 的根为顶点的三角形，有一个椭圆与它的三边相切于各边中点，而 f' 的根正是这个椭圆的焦点。对于等边三角形，这个椭圆是一个圆，它的两个焦点合而为一，成为 f' 的二重根。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）正是以这种精神来阐述复数几何。

详细解答

STEP III 2006，第 6 题

纯数学 · 极坐标曲线与反射

> **STEP III 2006，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/06-s3-q6

关键思路

斜率（gradient）公式不过是把链式法则（chain rule）应用于 $x = r\cos\theta, y = r\sin\theta$ 而已。至于镜面，把极点（pole）设在 O，并令始线（initial line）沿着 L：这样，在点 P 反射的光线便沿着径向直接返回，因此其方向只由 θ 决定。反射定律（law of reflection）指出法线（normal）平分两条光线之间的夹角，于是法线（以至切线（tangent））便可用 θ 确定下来。这样一来，几何问题就化为关于 $r(\theta)$ 的可分离变量微分方程（separable differential equation）。

解答

极坐标下的斜率

由 $r = r(\theta)$ 给出的曲线，其直角坐标（Cartesian coordinates）为 $x = r\cos\theta, y = r\sin\theta$，其中 r 是 θ 的函数。利用积法则（product rule）对 θ 求导，

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

根据链式法则，凡是 $\mathrm{d}x/\mathrm{d}\theta \neq 0$ 之处，

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}\theta}{\mathrm{d}x/\mathrm{d}\theta} = \frac{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\sin\theta + r\cos\theta}{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\cos\theta - r\sin\theta} = \frac{\dfrac{\mathrm{d}r}{\mathrm{d}\theta}\tan\theta + r}{\dfrac{\mathrm{d}r}{\mathrm{d}\theta} - r\tan\theta},$$

这里把分子和分母同除以 $\cos\theta$（当 $\cos\theta \neq 0$ 时可以这样做）。这就是所求的斜率。

未经相除的形式也值得保留：向量

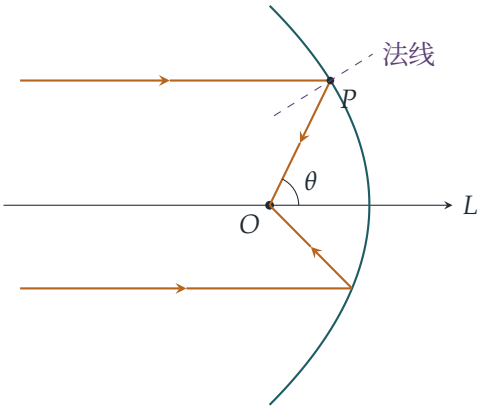
$$\mathbf{t} = \left(\tfrac{\mathrm{d}r}{\mathrm{d}\theta}\cos\theta - r\sin\theta,\ \tfrac{\mathrm{d}r}{\mathrm{d}\theta}\sin\theta + r\cos\theta\right)$$

在曲线上每一点都沿着切线方向，包括 $\tan\theta$ 或斜率无定义的地方。

镜面

任取一个包含 L 的平面 Π，并设 P 为镜面与 Π 相交的一点。在 P 的入射光线（incident ray）平行于 L，所以它位于 Π 之内；反射光线（reflected ray）通过 O，而该点在 L 上，所以它同样位于 Π 之内。反射定律亦要求在 P 的法线位于两条光线所在的平面内，因此法线位于 Π 之内，而交线在 P 的切线就是 Π 内与法线垂直的那条直线。所以一切都发生在平面 Π 之内，我们可以在二维中处理问题。

在 Π 内采用极坐标（polar coordinates），以 O 为极点，始线沿着 L，并选定方向使入射光沿 $\theta = 0$ 的方向行进（即图中由左至右）。设 P 为 (r,θ)，其中 $-\pi < \theta < \pi$。（$\theta = \pi$ 所对应的点位于 L 上、O 的后方，光线在那里会直接穿过 O，根本不需要镜面。）



法线的方向。 在 P 取两个沿着光线、背向 P 的单位向量（unit vectors）：

$$\mathbf{a} = (-1,\, 0)\ \text{（沿入射光线往回走）}, \qquad \mathbf{b} = (-\cos\theta,\, -\sin\theta)\ \text{（指向 } O\text{）}.$$

法线与 $\mathbf{a}$ 和 $\mathbf{b}$ 所成的角相等，所以它沿着两者夹角的平分线，亦即 $\mathbf{a}+\mathbf{b}$ 的方向（两个单位向量所构成的菱形，其对角线平分它们的夹角）。现在

$$\mathbf{a}+\mathbf{b} = -(1+\cos\theta,\, \sin\theta) = -2\cos\tfrac{\theta}{2}\left(\cos\tfrac{\theta}{2},\, \sin\tfrac{\theta}{2}\right),$$

这里用了 $1+\cos\theta = 2\cos^2\frac{\theta}{2}$ 和 $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$。由于 $-\pi < \theta < \pi$，$\cos\frac{\theta}{2} > 0$，所以这个向量非零，而在 P 的法线沿着

$$\mathbf{n} = \left(\cos\tfrac{\theta}{2},\, \sin\tfrac{\theta}{2}\right):$$

法线与 L 所成的角为 $\frac{\theta}{2}$。

想一想 为甚么不直接令斜率公式等于切线的斜率?

这样做是可以的：切线垂直于 $\mathbf{n}$，所以其斜率为 $-\cot\frac{\theta}{2}$，令它等于极坐标公式并去分母，便会得到与下文相同的方程。不过这条路要除以 $\tan\theta$、$\cot\frac{\theta}{2}$ 以及公式中的分母，所以在 $\theta=0$ 和 $\theta=\pm\frac{\pi}{2}$ 处行不通。采用切向量 $\mathbf{t}$ 则可避开这一切麻烦。

微分方程。 切线垂直于法线，所以 $\mathbf{t}\cdot\mathbf{n}=0$：

$$\frac{\mathrm{d}r}{\mathrm{d}\theta}\left(\cos\theta\cos\tfrac{\theta}{2}+\sin\theta\sin\tfrac{\theta}{2}\right)+r\left(\cos\theta\sin\tfrac{\theta}{2}-\sin\theta\cos\tfrac{\theta}{2}\right)=0.$$

根据复角公式（compound-angle formulae），两个括号分别等于 $\cos\left(\theta-\frac{\theta}{2}\right)=\cos\frac{\theta}{2}$ 和 $\sin\left(\frac{\theta}{2}-\theta\right)=-\sin\frac{\theta}{2}$。因此

$$\frac{\mathrm{d}r}{\mathrm{d}\theta}\cos\tfrac{\theta}{2}=r\sin\tfrac{\theta}{2},\qquad\text{即}\qquad\frac{1}{r}\frac{\mathrm{d}r}{\mathrm{d}\theta}=\tan\tfrac{\theta}{2},$$

这里除以了 $r\cos\frac{\theta}{2}$，由于 $P\neq O$ 且 $-\pi<\theta<\pi$，它是正数。

求解。 这个方程是可分离变量的。对 θ 积分，并注意到 $\int\tan\frac{\theta}{2}\,\mathrm{d}\theta=-2\ln\cos\frac{\theta}{2}$（因为 $\cos\frac{\theta}{2}>0$，此式成立），

$$\ln r=-2\ln\cos\tfrac{\theta}{2}+C\quad\Longrightarrow\quad r\cos^2\tfrac{\theta}{2}=k,$$

其中 $k=\mathrm{e}^C>0$ 沿着整条曲线是常数（曲线是区间 $-\pi<\theta<\pi$ 上单独一段连通的 $r=r(\theta)$）。利用 $2\cos^2\frac{\theta}{2}=1+\cos\theta$，

$$r(1+\cos\theta)=2k.$$

转换为直角坐标。 以 $x=r\cos\theta$ 和 $r=\sqrt{x^2+y^2}$ 代入，上式变成 $\sqrt{x^2+y^2}=2k-x$。两边平方（两边均为非负，因为 $2k-x=r\geq0$），

$$x^2+y^2=4k^2-4kx+x^2\qquad\Longrightarrow\qquad y^2=4k(k-x).$$

这是一条抛物线（parabola），以 L 为轴，顶点（vertex）在 $(k,0)$，焦点（focus）在原点 O，开口朝向入射光的来向。由于 Π 是任意一个包含 L 的平面，镜面与每一个这样的平面的交线都是抛物线，这正是所要证明的。■

验算

- 倒过来检验：在 $y^2=4k(k-x)$ 上，法线沿着 $(4k,2y)$；把方向 $(1,0)$ 对这条法线作反射，便得到一条通过 O 的光线。我们以数值方法在 10^4 个随机点上验证了这一点

（取 $k = 1.3$），并且用有限差分（finite differences）验证了 $r = 2k/(1+\cos\theta)$ 满足 $r'/r = \tan\frac{\theta}{2}$。

- 对于随机选取的曲线 $r(\theta)$，斜率公式与有限差分所得的斜率相符，误差约为 10^{-8}。
- 在 $\theta = 0$ 处，法线沿着 L，而切线垂直于 L：这就是顶点 $(k,0)$，沿 L 射来的光线在那里会直接反射回来并通过 O。

常见错误

- 以外角平分线（external bisector）$\mathbf{a} - \mathbf{b}$ 作为法线。这个方向与两条光线所成的角是互补而非相等，并会导致 $r'/r = -\cot\frac{\theta}{2}$，其解 $r(1-\cos\theta) = 2k$ 是一条开口朝相反方向的抛物线，所聚焦的是从相反方向射来的光。
- 忘记解释为何问题是二维的：法线必须位于包含 L 和 P 的平面内，否则交线的切线未必就是我们所用的那一条。
- 把 $\sqrt{x^2+y^2} = 2k - x$ 两边平方时，没有指出两边均为非负。

STEP 之外

这块镜面是一个旋转抛物面（paraboloid of revolution），而这里所证明的性质其实就是焦点－准线（focus–directrix）性质的另一种面貌：$r(1+\cos\theta) = 2k$ 是说，与 O 的距离等于与直线 $x = 2k$ 的距离。这正是卫星碟形天线、汽车车头灯和反射望远镜都采用抛物面的原因。更一般地说，所有圆锥曲线（conics）都具有极坐标形式 $r(1+e\cos\theta) = \ell$，而这恰恰就是牛顿（Newton）求解行星轨道时所出现的式子；Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）由平方反比定律（inverse-square law）把它推导出来。

详细解答

STEP III 2006，第 7 题

纯数学 · 一阶微分方程

STEP III 2006，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/06-s3-q7

关键思路

两条微分方程（differential equation）都是关于 $\mathrm{d}y/\mathrm{d}x$ 的二次方程。我们不急于把它们当作微分方程来处理，而是先就导数（derivative）解出每条二次方程，这样原方程便分拆成两条普通的一阶方程。利用双曲恒等式（hyperbolic identities），两个根会化简成很简单的形式，而 $x = 0$ 处的条件则告诉我们应该跟随哪一个根。

解答

第 (i) 部分

由二次公式（quadratic formula），

$$u = -\sinh x \pm \sqrt{\sinh^2 x + 1} = -\sinh x \pm \cosh x,$$

因为 $\sinh^2 x + 1 = \cosh^2 x$ 而且 $\cosh x > 0$。利用 $\cosh x + \sinh x = \mathrm{e}^x$ 和 $\cosh x - \sinh x = \mathrm{e}^{-x}$，两个根是

$$u = \mathrm{e}^{-x} \qquad 或 \qquad u = -\mathrm{e}^x.$$

微分方程就是这条二次方程，其中 $u = \mathrm{d}y/\mathrm{d}x$，所以在每一个 x 处必有

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{e}^{-x} \qquad 或 \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\mathrm{e}^x.$$

第一个根恒为正，第二个根恒为负，所以两者永不相等。因此，导数连续（continuous）的解不可能由一个根跳到另一个根，而 $x = 0$ 处 $\mathrm{d}y/\mathrm{d}x > 0$ 这个条件，就决定了自始至终都取第一个根。积分后得 $y = -\mathrm{e}^{-x} + c$，再由 $x = 0$ 时 $y = 0$ 得 $c = 1$：

$$y = 1 - \mathrm{e}^{-x}.$$

第 (ii) 部分

设 $p = \mathrm{d}y/\mathrm{d}x$。若 $\sinh y = 0$，即 $y = 0$，方程便化为 $2p = 0$。$y \equiv 0$ 就是这样成为一个解的，而题目正是叫我们把这个解撇开不理。在 $y \neq 0$ 之处，由二次公式得

$$p = \frac{-1 \pm \sqrt{1 + \sinh^2 y}}{\sinh y} = \frac{-1 \pm \cosh y}{\sinh y}.$$

利用半角形式（half-angle forms）$\cosh y - 1 = 2\sinh^2 \frac{y}{2}$、$\cosh y + 1 = 2\cosh^2 \frac{y}{2}$ 和 $\sinh y = 2\sinh \frac{y}{2} \cosh \frac{y}{2}$，两个根是

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \tanh \frac{y}{2} \qquad 或 \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\coth \frac{y}{2}.$$

这两个根永不相等，因为它们的积是 -1，而若两根相等，其积应为 $\tanh^2 \frac{y}{2} \geq 0$。

想一想 两条方程之中，哪一条可以给出一个通过原点而又不恒等于零的解?

第一条方程 $y' = \tanh \frac{y}{2}$ 有解 $y \equiv 0$。它的右边是 y 的光滑函数（smooth function），所以通过 $(0,0)$ 的解就只有这一个：这又是那个平凡解（trivial solution）。第二条方程在 $y = 0$ 处是奇异（singular）的，该处 $|\mathrm{d}y/\mathrm{d}x| \to \infty$，所以它的解曲线可以带着铅垂切线与 $y = 0$ 相交。这正是我们要找的解。分离变量（separating the variables），

$$\int \tanh \frac{y}{2} \, \mathrm{d}y = -\int \mathrm{d}x \quad \Longrightarrow \quad 2\ln\cosh \frac{y}{2} = -x + C.$$

代入 $x = 0$ 时 $y = 0$，得 $C = 0$，所以 $\cosh^2 \frac{y}{2} = \mathrm{e}^{-x}$。由于 $\cosh y = 2\cosh^2 \frac{y}{2} - 1$，

$$\cosh y = 2\mathrm{e}^{-x} - 1.$$

我们可以直接验证这个结果。求导得 $\sinh y \, y' = -2\mathrm{e}^{-x} = -(\cosh y + 1)$，这正是 $y' = -\coth \frac{y}{2}$。无论 y 是正是负，这都成立，所以整条曲线，包括上半和下半，都是解。由于 $\cosh y \geq 1$，曲线只在 $x \leq 0$ 时存在。它关于 x 轴对称。在原点附近 $1 + \frac{1}{2}y^2 \approx 1 - 2x$，所以曲线在那里看起来像抛物线 $y^2 = -4x$。

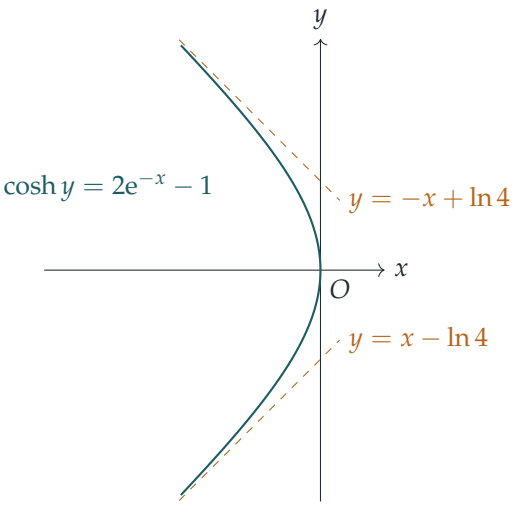
渐近线（asymptotes）。当 $x \to -\infty$ 时，$\cosh y \to \infty$，所以 $|y| \to \infty$。设 $w = 2\mathrm{e}^{-x} - 1$，则

$$|y| = \operatorname{arcosh} w = \ln\bigl(w + \sqrt{w^2 - 1}\bigr),$$

所以

$$|y| - (-x + \ln 4) = \ln \frac{w + \sqrt{w^2 - 1}}{4\mathrm{e}^{-x}} = \ln\Bigl(\frac{w}{4\mathrm{e}^{-x}}\Bigl(1 + \sqrt{1 - w^{-2}}\Bigr)\Bigr).$$

当 $x \to -\infty$ 时，$w/(4\mathrm{e}^{-x}) = \frac{1}{2} - \frac{1}{4}\mathrm{e}^{x} \to \frac{1}{2}$，而 $1 + \sqrt{1 - w^{-2}} \to 2$，所以对数内的式子趋向 1，上述差值趋向 0。因此曲线的上半趋近 $y = -x + \ln 4$，而由对称性，下半趋近 $y = -(-x + \ln 4)$。渐近线是 $y = \pm(-x + \ln 4)$，正如题目所求。（当 $x \to 0^-$ 时曲线保持有界，所以渐近线就只有这两条。）



验算

- 在 (i) 中，$y = 1 - \mathrm{e}^{-x}$ 的导数是 $y' = \mathrm{e}^{-x}$，而 $\mathrm{e}^{-2x} + 2\mathrm{e}^{-x}\sinh x - 1 = \mathrm{e}^{-2x} + (1 - \mathrm{e}^{-2x}) - 1 = 0$。
- 在 (ii) 中，由原点附近开始，以数值方法逐步求解 $y' = -\coth\frac{y}{2}$ 直到 $x = -6$，得 $y \approx 7.38505$，与 $\operatorname{arcosh}(2\mathrm{e}^6 - 1)$ 吻合。在 $x = -10$ 处，曲线与 $-x + \ln 4$ 之间的差距约为 2×10^{-5}。

常见错误

- 在 (i) 中取了 $-\sinh x - \cosh x = -\mathrm{e}^x$ 这个根：这样导数在 $x = 0$ 处是负数。
- 在 (ii) 中因为 $y' = \tanh\frac{y}{2}$ 在 $y = 0$ 处“性质良好”而选了它。通过原点时，它只给出 $y \equiv 0$。
- 忘记了曲线的下半；这一半有 $y < 0$，第二条渐近线正是由它而来。

STEP 之外

像这两条方程一样，对 y' 而言并非线性的方程，称为一阶隐式（first-order implicit）方程。就 y' 求解，会把它分拆成若干个分支。第 (ii) 部分亦显示解的唯一性（uniqueness）可以不成立：在 $y = 0$ 处，右边的 $-\coth\frac{y}{2}$ 不满足利普希茨（Lipschitz）条件，所以一个非平凡解才可以与 $y \equiv 0$ 从同一点出发。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）讲述了存在性与唯一性定理，以及这些定理在甚么情

况下适用。

详细解答

STEP III 2006，第 8 题

纯数学 · 公理化微分法

STEP III 2006，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/06-s3-q8

关键思路

这四条规则正是你日常使用的求导（differentiation）性质：线性（linearity）和积法则（product rule），再加上一个事实，就是 x“求导后得”1。每个多项式（polynomial）都是由数字 1 和字母 x 经过相乘和相加构成的，所以只要知道 $\triangle$ 对 1 和对 x 的作用，这些规则就令它在其他地方再无任何自由。把积法则应用于 $x \cdot x^n$，便可对幂次作归纳法（induction），最后再由线性完成余下的工作。

解答

以下“规则 (i)”至“规则 (iv)”均指题目所给的四条定义规则，次序与题目相同。

常数

先考虑常数多项式 1。由于 $1 = 1 \cdot 1$，在规则 (iv) 中取 $f(x) = g(x) = 1$，得

$$\triangle 1 = \triangle(1 \cdot 1) = 1 \cdot \triangle 1 + 1 \cdot \triangle 1 = 2\,\triangle 1.$$

两边同时减去 $\triangle 1$，得 $\triangle 1 = 0$（零多项式）。

? **想一想** 为甚么不直接说“$\triangle$ 作用于常数得零，就像导数一样”？

因为我们还不知道 $\triangle$ 就是导数（derivative）；这正是最后一部分要我们证明的。我们能用的只有那四条规则，而其中没有一条直接提及常数。窍门在于 1 等于它自己的平方，所以积法则迫使 $\triangle 1$ 等于它自身的两倍。

现设 $f(x)=\lambda$ 为任意常数。则 $\lambda=\lambda\cdot 1$，故由规则 (iii)

$$\triangle\lambda=\triangle(\lambda\cdot 1)=\lambda\,\triangle 1=\lambda\cdot 0=0.$$

所以对每个次数（degree）为零的多项式，都有 $\triangle f(x)=0$，正如所求。

平方与立方

把 $x^2=x\cdot x$ 写出，先用规则 (iv)，再用规则 (i)，

$$\triangle x^2=x\,\triangle x+x\,\triangle x=x\cdot 1+x\cdot 1=2x.$$

把 $x^3=x\cdot x^2$ 写出，先用规则 (iv)，再用刚得到的结果和规则 (i)，

$$\triangle x^3=x\,\triangle x^2+x^2\,\triangle x=x\cdot 2x+x^2\cdot 1=3x^2.$$

一般多项式

*步骤 1：x 的幂。*我们用归纳法证明

$$\triangle x^n=n\,x^{n-1}\qquad \text{对每个整数}n\geq 1.$$

当 $n=1$ 时，这就是 $\triangle x=1$，即规则 (i)。假设结果对某个 $n=k\geq 1$ 成立。由于 $x^{k+1}=x\cdot x^k$，规则 (iv) 给出

$$\begin{aligned}\triangle x^{k+1}&=x\,\triangle x^k+x^k\,\triangle x\\&=x\cdot k\,x^{k-1}+x^k\cdot 1\\&=(k+1)\,x^k,\end{aligned}$$

其中第二行用了归纳假设（induction hypothesis）和规则 (i)。所以结果对 $n=k+1$ 成立，由归纳法可知它对所有 $n\geq 1$ 成立。当 $n=0$ 时，$x^0=1$，而我们已证明 $\triangle 1=0$，这与 $\dfrac{\mathrm{d}}{\mathrm{d}x}1=0$ 一致。所以对每个 $n\geq 0$，

$$\triangle x^n=\frac{\mathrm{d}}{\mathrm{d}x}\,x^n.$$

*步骤 2：多项之和。*规则 (ii) 只是就两个多项式而言，但它可以推广至任意有限个。若 $\triangle(f_1+\cdots+f_m)=\triangle f_1+\cdots+\triangle f_m$，则把规则 (ii) 应用于 $f_1+\cdots+f_m$ 和 f_{m+1} 这两个多项式，得

$$\triangle(f_1+\cdots+f_{m+1})=\triangle(f_1+\cdots+f_m)+\triangle f_{m+1}=\triangle f_1+\cdots+\triangle f_{m+1},$$

所以对 m 作归纳法，可知它对任意多项都成立。

步骤 3：综合起来。任何多项式都可以写成 $h(x) = \sum_{k=0}^{n} a_k x^k$，其中 a_k 为常数。于是

$$\begin{aligned}
\triangle h(x) &= \sum_{k=0}^{n} \triangle\left(a_k x^k\right) && \text{根据规则 (ii)，并按步骤 2 推广} \\
&= \sum_{k=0}^{n} a_k \triangle x^k && \text{根据规则 (iii)} \\
&= \sum_{k=1}^{n} k\, a_k\, x^{k-1} && \text{根据步骤 1（}k = 0\text{ 的项为 0）} \\
&= \frac{\mathrm{d}h(x)}{\mathrm{d}x},
\end{aligned}$$

最后一行就是多项式通常的导数。这对每个多项式 $h(x)$ 都成立，所以 $\triangle h(x) \equiv \dfrac{\mathrm{d}h(x)}{\mathrm{d}x}$，正如所求。

验算

令人安心的是，普通的求导确实符合全部四条规则，所以这些规则并不互相矛盾，而 $\mathrm{d}/\mathrm{d}x$ 的确是题目所描述的那类运算。你也可以用公式来检验积法则：取 $f = x^2$ 和 $g = x^3$，规则 (iv) 给出 $x^2 \cdot 3x^2 + x^3 \cdot 2x = 5x^4$，这正是 $\triangle x^5$。我们在电脑上只用这些规则（如步骤 1 至 3）构造出 $\triangle$，并把它与导数作精确比较，共测试了 2000 个次数不超过 9 的随机多项式；例如 $\triangle\left(\frac{1}{2}x^4 - 3x + 7\right) = 2x^3 - 3$，这也与数值差商（difference quotient）吻合。

常见错误

- 断言 $\triangle(\text{常数}) = 0$，“因为它是导数”。这是预先假设了最终答案；这个结果必须由规则推导出来。
- 不加说明就把规则 (ii) 用于一长串的和。规则只是就两个多项式而言，所以项数更多时，需要用一句话交代归纳法。
- 没有说明哪一步用了哪条规则：题目明确要求这样做。

STEP 之外

符合线性和积法则的运算称为*导子*（derivation），而这里的论证显示，在多项式上，导子完全由它在 x 上的值决定。假如改为 $\triangle x = p(x)$，其中 p 为某个多项式，同样的证明会给出 $\triangle h = p(x)\, h'(x)$。求导可以纯粹用代数方法定义，完全不涉及极限，这个想法在抽象代数（abstract algebra）和微分几何（differential geometry）中占有核心地位；在微分几何中，向量场（vector field）正正就是函数上的导子。Allenby 的 *Rings, Fields and*

Groups（环、域与群）是认识以代数方式处理多项式环（polynomial ring）的好去处。

详细解答

STEP III 2006，第 9 题

力学 · 能量守恒

STEP III 2006，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：

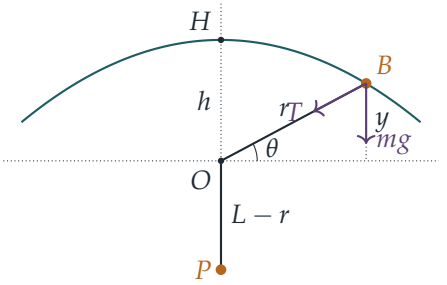
step.maths.org/questions/06-s3-q9

关键思路

绳的长度固定，所以小珠与 O 的距离增加多少，质点 P 就上升多少。这样便把两者的势能（potential energy）连系起来：总势能只取决于 $y + r$。“在任何位置都处于平衡（equilibrium）”表示势能并不偏好任何位置，所以它沿着金属线必定是常数。既然势能是常数，能量守恒（conservation of energy）便告诉我们总动能（kinetic energy）永不改变，余下的工作只是用极坐标（polar coordinates）写出小珠的速率。

解答

模型。P 和 B 是质量为 m 的质点；绳是轻而不可伸长（light and inextensible）的；小环和金属线都是光滑（smooth）的，所以绳的张力（tension）所作的净功为零，而金属线的法向反作用力（normal reaction）不作功。取 O 为原点，x 轴水平，y 轴垂直向上，并设绳长为 L。



金属线的形状

由 O 到 B 的一段绳长为 r，所以 OP 的长度是 $L-r$，即 P 位于 O 之下 $L-r$ 的距离。取通过 O 的水平面为势能的零点，总势能为

$$U = mgy - mg(L-r) = mg(y+r) - mgL.$$

这个系统只有一个自由度（degree of freedom），即小珠在金属线上的位置，而平衡位置就是 U 沿金属线的驻点（stationary point）。如果每一个位置都是平衡位置，那么沿着金属线处处都有 $\mathrm{d}U/\mathrm{d}s = 0$（其中 s 是弧长（arc length）），所以 U 沿金属线是常数。因此 $y+r$ 是常数。

在 H 处，小珠位于 O 的正上方，高度为 h，所以该处 $y = h$ 且 $r = h$，这个常数就是 $2h$。于是金属线的方程是

$$y + r = 2h,$$

正如题目所求。（写成直角坐标形式，由 $x^2 + y^2 = (2h-y)^2$ 得 $y = h - x^2/(4h)$：这是一条以 O 为焦点（focus）的抛物线（parabola），如图所示。你也可以直接验证平衡：由于 P 处于平衡，张力是 mg；沿 BO 方向的 mg 与垂直向下的 mg 的合力（resultant）平分这两个方向之间的夹角，而根据抛物线的反射性质（reflection property），这正好是抛物线在该点的法线（normal）方向。）

θ 的变化率

在金属线上 $y = 2h - r$，所以

$$\sin\theta = \frac{y}{r} = \frac{2h}{r} - 1.$$

对时间求导，

$$\cos\theta\,\dot{\theta} = -\frac{2h\dot{r}}{r^2}.$$

由于 $\theta = \arcsin(y/r)$ 位于 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 之内，我们有 $\cos\theta \geq 0$，所以

$$\cos\theta = \sqrt{1 - \left(\frac{2h-r}{r}\right)^2} = \frac{\sqrt{r^2 - (2h-r)^2}}{r} = \frac{\sqrt{4rh - 4h^2}}{r} = \frac{2\sqrt{rh-h^2}}{r}.$$

小珠离开 H 之后，我们有 $r > h$（因为 $r = 2h - y$，而在 H 以外 $y < h$），所以 $\cos\theta \neq 0$，可以放心相除：

$$\dot{\theta} = -\frac{2h\dot{r}}{r^2} \cdot \frac{r}{2\sqrt{rh-h^2}} = -\frac{h\dot{r}}{r\sqrt{rh-h^2}},$$

这就是所求的结果。

P 的速率

> **想一想** 为甚么总动能必须一直等于 $\frac{1}{2}mV^2$?

对系统作功的力只有重力和绳的张力（属内力）。张力所作的净功为零：绳不可伸长，小环又光滑，所以张力对 B 所作的功与对 P 所作的功互相抵销。金属线是光滑的，所以它的反作用力不作功。因此动能与势能之和守恒。但我们刚刚证明了势能 U 沿这条金属线是常数，所以动能也是常数。

开始时小珠在 H，该处 r 取最小值 h（因为 y 在该处最大）。所以最初 $\dot r = 0$，而 P 的速率是 $|\dot r|$，故它由静止开始运动。因此初始动能是 $\frac{1}{2}mV^2$。

在以 O 为极点的极坐标中，小珠速率的平方是 $\dot r^2 + r^2\dot\theta^2$。利用 $\dot\theta$ 的结果，

$$r^2\dot\theta^2 = \frac{h^2\dot r^2}{rh - h^2} = \frac{h\dot r^2}{r-h}, \qquad \text{所以} \qquad v_B^2 = \dot r^2\Big(1 + \frac{h}{r-h}\Big) = \frac{r\dot r^2}{r-h}.$$

由动能守恒，约去 $\frac{1}{2}m$ 之后得

$$\frac{r\dot r^2}{r-h} + \dot r^2 = V^2 \quad\Longrightarrow\quad \dot r^2\,\frac{2r-h}{r-h} = V^2.$$

由于 $r > h$，$r-h$ 和 $2r-h$ 都是正数，所以 P 的速率是

$$|\dot r| = V\Big(\frac{r-h}{2r-h}\Big)^{\frac12},$$

正如题目所求。

验算

- 在 H 处，$r = h$，公式给出 0：正确，因为 P 由静止开始。
- 当 $r \to \infty$ 时，P 的速率趋向 $V/\sqrt{2}$。在很远的地方，小珠几乎沿径向运动，所以 B 和 P 的速率几乎相等，两者平分动能。
- 数值验证：取 $h = 1$ 和 $V = 2$，我们沿 $y = 1 - \frac14 x^2$ 逐步计算小珠的运动方程（用 RK4），并把 $\dot\theta$ 和 $\dot r$ 的有限差分值与公式比较。在 $r \approx 4.043$ 时，P 的速率是 1.31064，与 $2\sqrt{3.043/7.087}$ 吻合至六位有效数字。

常见错误

- 把 P 的势能取为 $+mg(L-r)$：P 在 O 的下方，所以它的高度是 $-(L-r)$。
- 忘记 P 也在运动，于是写成 $\frac{1}{2}mv_B^2 = \frac{1}{2}mV^2$。
- 除以 $\cos\theta$ 时，没有指出它只在 H 处才等于零。

STEP 之外

寻找一条曲线，使系统在其上处处都处于“随遇”平衡（neutral equilibrium），是“平衡位置即势能的驻点”这个原理的一个小例子。这个原理通往拉格朗日力学（Lagrangian mechanics），在那里，像绳和金属线这样的约束（constraints）是通过选取合适的坐标而直接纳入其中的。在这一题，单一个坐标 r 已经足够，而能量守恒就是首次积分（first integral）。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）正是以这种精神讲解能量方法和极坐标。

详细解答

STEP III 2006，第 10 题

力学 · 角动量守恒

STEP III 2006，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/06-s3-q10

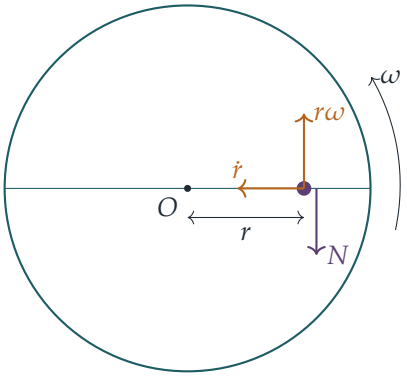
关键思路

没有任何外来作用令圆盘与质点组成的系统绕轴转动，而光滑凹槽所作的净功为零，所以有两条守恒定律：绕轴的角动量（angular momentum）守恒和动能（kinetic energy）守恒。由此可得 ω 和 $\dot{r}$ 均为只含 r 的函数，而那个把 V、Ω、a 和 k 连系起来、看似古怪的条件，正是为了令能量的表达式大幅化简。把 $\dot{r}$ 除以 $\dot{\theta} = \omega$ 便可消去时间，剩下一条描述路径的可分离变量（separable）方程。

解答

模型

我们把质点视为质量为 m 的点质量，并把凹槽视为光滑，所以质点所受的唯一水平力是垂直于凹槽的反作用力（reaction）N。圆盘绕固定的铅垂轴自由转动，所以轴承对该轴不施加力矩（torque）。重力和铅垂方向的反作用力都是铅垂的，不起任何作用。设 θ 为圆盘已转过的角度，故 $\omega = \dot{\theta}$，并设 r 为质点与轴的距离。由于质点一直留在凹槽内，它的速度有一个沿凹槽的分量 $\dot{r}$ 和一个垂直于凹槽的分量 $r\omega$。



角速率

反作用力 N 作用于质点，而 $-N$ 则作用于圆盘上的同一点，所以两者绕轴的力矩互相抵销。既然没有绕轴的外力矩，绕轴的总角动量便是常数。圆盘的贡献是 $mk^2\omega$，质点的贡献是 $mr\cdot r\omega = mr^2\omega$（只有其速度的横向分量才有矩）。在 $t=0$ 时，总角动量为 $m(k^2+a^2)\Omega$，所以

$$(k^2+r^2)\,\omega = (k^2+a^2)\,\Omega, \qquad \text{即} \qquad \omega = \frac{\Omega(k^2+a^2)}{k^2+r^2},$$

这就是第一个结果。

径向速率

> **想一想** 质点运动时，凹槽从侧面推着它。为甚么能量仍然守恒?

反作用力 N 垂直于凹槽。质点与圆盘上跟它接触的那一点，在垂直于凹槽方向上的速度分量相同，都是 $r\omega$；两者的差别只在于沿凹槽的滑动速度 $\dot r$，而这个速度垂直于 N。所以 N 对质点所作的功，恰好被 $-N$ 对圆盘所作的功抵销。轴上的轴承不作功，所以总动能是常数。

约去公因子 $\frac12 m$ 后，能量守恒给出

$$k^2\omega^2 + \dot r^2 + r^2\omega^2 = k^2\Omega^2 + V^2 + a^2\Omega^2.$$

现在把题目所给的条件 $V^2 = \Omega^2a^2(k^2+a^2)/k^2$ 用于右边：

$$(k^2+a^2)\Omega^2 + \frac{\Omega^2a^2(k^2+a^2)}{k^2} = \frac{\Omega^2(k^2+a^2)(k^2+a^2)}{k^2} = \frac{\Omega^2(k^2+a^2)^2}{k^2}.$$

至于左边，由第一个结果可得 $(k^2+r^2)\omega^2 = \Omega^2(k^2+a^2)^2/(k^2+r^2)$。因此

$$\begin{aligned}\dot{r}^2 &= \Omega^2(k^2+a^2)^2\left(\frac{1}{k^2}-\frac{1}{k^2+r^2}\right)\\ &= \frac{\Omega^2 r^2(k^2+a^2)^2}{k^2(k^2+r^2)},\end{aligned}$$

正是所求。

路径的方程

质点向轴运动时，$\dot{r}<0$，所以取负平方根，

$$\dot{r} = -\frac{\Omega r(k^2+a^2)}{k\sqrt{k^2+r^2}}.$$

由于 $\omega = \dot{\theta} > 0$，我们可以使用链式法则（chain rule），$\dfrac{\mathrm{d}r}{\mathrm{d}\theta} = \dot{r}\Big/\dot{\theta} = \dot{r}/\omega$：

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = -\frac{\Omega r(k^2+a^2)}{k\sqrt{k^2+r^2}}\cdot\frac{k^2+r^2}{\Omega(k^2+a^2)} = -\frac{r\sqrt{k^2+r^2}}{k},$$

所以 $k\dfrac{\mathrm{d}r}{\mathrm{d}\theta} = -r(k^2+r^2)^{1/2}$，正是所求。

求解 r

设 $u = k/r$，故 $u>0$。于是

$$\frac{\mathrm{d}u}{\mathrm{d}\theta} = -\frac{k}{r^2}\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{k}{r^2}\cdot\frac{r\sqrt{k^2+r^2}}{k} = \sqrt{\frac{k^2}{r^2}+1} = \sqrt{1+u^2}.$$

这条方程可分离变量：

$$\int\frac{\mathrm{d}u}{\sqrt{1+u^2}} = \int\mathrm{d}\theta \quad\Longrightarrow\quad \operatorname{arsinh}u = \theta+\alpha$$

其中 α 为某个常数。所以 $k/r = \sinh(\theta+\alpha)$，即

$$r\sinh(\theta+\alpha) = k.$$

在 $t=0$ 时，$\theta=0$ 而 $r=a$，所以 $\sinh\alpha = k/a$，正是所求。

质点永远不会到达轴

? **想一想** $r = k/\sinh(\theta+\alpha)$ 永不为零，但 θ 会否在有限时间内变成无穷大?

不会。由于 $r^2 \geq 0$，第一个结果给出 $\omega \leq \Omega(k^2+a^2)/k^2$，所以 $\theta \leq \Omega(k^2+a^2)t/k^2$ 在每一个有限的时刻 t 都是有限的。于是在任何时刻都有 $r = k/\sinh(\theta+\alpha) > 0$。

我们还应检查这些公式是否一直成立，即质点是否一直向轴运动。根据 $\dot{r}^2$ 的公式，$\dot{r}$ 只有在 $r = 0$ 时才可能等于零。所以只要 $r > 0$，$\dot{r}$ 便保持最初的负号，而上述论证对所有 $t > 0$ 都适用。质点沿螺旋线不断向轴靠近，只有当 $\theta \to \infty$ 时才有 $r \to 0$，但它永远不会到达轴。

验算

- 在 $r = a$ 时，$\dot{r}^2$ 的公式给出 $\Omega^2 a^2(k^2+a^2)/k^2 = V^2$，与初始条件吻合。
- 取 $k = a = \Omega = 1$（故 $V = \sqrt{2}$），我们以数值方法对运动方程 $\ddot{r} = r\omega^2$ 和 $(k^2+r^2)\omega =$ 常数作积分。当圆盘转过 $\theta \approx 4.4170$ 后，质点到达 $r = 0.01$，而 $\operatorname{arsinh} 100 - \operatorname{arsinh} 1 \approx 5.2983 - 0.8814 = 4.4169$ 与此相符。在整个计算过程中，能量、$\dot{r}^2$ 和 $r\sinh(\theta+\alpha) = k$ 全都成立，误差约为 10^{-11}。

常见错误

- 忘记质点本身的角动量 $mr^2\omega$，或它的横向动能 $\frac{1}{2}mr^2\omega^2$。
- 对 $\dot{r}$ 取正平方根，令 $\mathrm{d}r/\mathrm{d}\theta$ 的正负号出错。
- 单凭 $r \neq 0$ 便断定“永远不会到达轴”，而没有指出 θ 在有限时间内保持有限。

STEP 之外

质点不受任何径向力，但在圆盘的旋转参考系（rotating frame）中，它却似乎被向外推。这就是 $\ddot{r} = r\omega^2$ 中的离心（centrifugal）项，而侧向的反作用力 N 则与科里奥利力（Coriolis force）平衡。大学力学会用拉格朗日（Lagrange）方法处理这类问题：角度 θ 不出现在能量之中，所以它的共轭动量（conjugate momentum），在这里即角动量，便自动守恒。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门）以恰当的程度讲解旋转参考系和角动量。

详细解答

STEP III 2006，第 11 题

力学 · 连接系统中的碰撞

STEP III 2006，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/06-s3-q11

关键思路

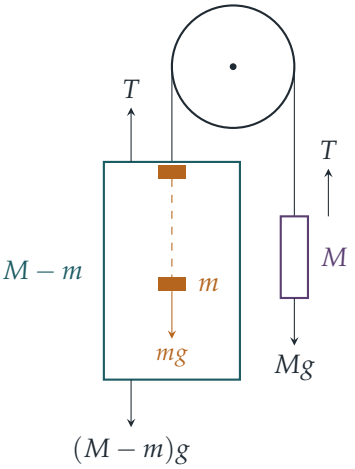
瓷砖一旦松脱，升降机便比平衡锤轻，于是升降机开始上升，而瓷砖则下坠：瓷砖要走过的，是相对于升降机的高度 h。至于碰撞，升降机与平衡锤可视为一个质量为 $2M - m$ 的单一物体，经地板与瓷砖相连；题目对缆索冲量所给的条件，正好让我们为这个物体和瓷砖写出一条动量式的守恒定律。这个“动量”在碰撞前一刻原来刚好是零，令能量的计算变得十分简洁。

解答

我们把升降机、平衡锤和瓷砖都当作质点（particle）处理（只考虑垂直方向的运动），缆索为轻而不可伸长（light and inextensible），滑轮为光滑（smooth）。瓷砖脱落之前，系统处于静止状态，所以连同瓷砖的升降机与平衡锤互相平衡：升降机连瓷砖的质量为 M，不连瓷砖则为 $M - m$。

下坠

对升降机和瓷砖，取向上为正。由于缆索不可伸长，当升降机以速率 u、加速度 a 向上运动时，平衡锤便以相同的速率和加速度向下运动。设张力（tension）为 T。



对升降机（向上）和平衡锤（向下）应用牛顿第二定律（Newton's second law），得

$$(M-m)a = T-(M-m)g, \qquad Ma = Mg - T.$$

两式相加，得 $(2M-m)a = mg$，所以

$$a = \frac{mg}{2M-m}.$$

瓷砖作自由落体运动（free fall），加速度为向下的 g（导轨光滑，亦没有其他东西碰到它）。因此，相对于升降机，瓷砖向下的加速度为

$$g + a = g + \frac{mg}{2M-m} = \frac{2Mg}{2M-m},$$

而它相对于升降机是由静止开始运动的。设它经过时间 t 后碰到地板，则 $\frac{1}{2}(g+a)t^2 = h$，所以

$$t^2 = \frac{(2M-m)h}{Mg}.$$

在那一刻，瓷砖的速率为

$$v = gt = \sqrt{g^2 \cdot \frac{(2M-m)h}{Mg}} = \sqrt{\frac{(2M-m)gh}{M}},$$

正如题目所求。此时升降机（和平衡锤）的速率为

$$U = at = \frac{m}{2M-m}\,gt = \frac{mv}{2M-m}.$$

在这段时间内，升降机上升了 $\frac{1}{2}at^2 = \frac{mh}{2M} < h$，所以题目对升降机与滑轮距离所设的条件，保证了升降机不会先碰到滑轮。

碰撞

设 J 为地板作用于瓷砖的冲量（impulse）（向上），又设 I 为缆索张力的冲量；根据题目所给的条件，这个冲量作用于升降机和作用于平衡锤的大小相同（两者都是向上）。碰撞在瞬间完成，重力在此期间不产生冲量。设碰撞后一刻瓷砖的速度为 w（向上），升降机的速度为 U'（向上），则平衡锤向下的速度同样是 U'。于是：

$$\begin{aligned} \text{瓷砖：} \quad & m\big(w-(-v)\big) = J, \\ \text{升降机：} \quad & (M-m)(U'-U) = I - J, \\ \text{平衡锤（向下）：} \quad & M(U'-U) = -I. \end{aligned}$$

三式相加，便消去 J 和 I：

$$mw + (2M-m)U' = -mv + (2M-m)U.$$

> ? **想一想** 平衡锤的方程以“向下”为正，为甚么可以把它与另外两条方程相加?

因为有缆索相连，“平衡锤向下”与“升降机向上”其实是同一个坐标。我们所证明的是：$mw + (2M-m)U'$ 这个组合（瓷砖向上的动量，加上升降机与平衡锤沿缆索方向量度的动量）在碰撞前后保持不变。冲量相等这个条件正是用在这里：没有它，I 便不会抵消。

现在是关键的观察：由于 $U = \dfrac{mv}{2M-m}$，右边等于 $-mv + mv = 0$。所以

$$mw + (2M-m)U' = 0 = -mv + (2M-m)U.$$

在地板处应用牛顿实验定律（Newton's experimental law）：分离速率等于接近速率的 e 倍，即

$$w - U' = e\big(U - (-v)\big) = e(U+v).$$

碰撞前和碰撞后的速度都满足同一条“零动量”方程，所以 (w, U') 和 $(-v, U)$ 这两对数各自都是 $(2M-m,\ -m)$ 的倍数。恢复方程随即决定了这个倍数：$w = ev$ 及 $U' = -eU$。（验算：$mev - (2M-m)eU = e(mv - mv) = 0$，且 $w - U' = e(v+U)$。）

能量损失

碰撞前一刻，动能（kinetic energy）为

$$\tfrac{1}{2}mv^2 + \tfrac{1}{2}(2M-m)U^2 = \tfrac{1}{2}mv^2\left(1 + \frac{m}{2M-m}\right) = \frac{Mmv^2}{2M-m} = mgh,$$

其中先用了 $U = \frac{mv}{2M-m}$，再用了 $v^2 = \frac{(2M-m)gh}{M}$。碰撞后一刻，每个速度都乘上了 $\pm e$，所以动能为 $e^2\,mgh$。碰撞在瞬间完成，期间势能（potential energy）不变，因此系统的能量损失为

$$mgh - e^2 mgh = mgh(1-e^2),$$

正如题目所求。

评注

答案完全与 M 无关：它正好等于瓷砖在一部固定不动的升降机内下坠高度 h、再以恢复系数（coefficient of restitution）e 反弹时的能量损失。这是合理的。碰撞前一刻，系统的动能为 mgh，这正是所释放的势能（瓷砖与升降机之间的距离缩短了 h，而升降机上升的高度恰好被平衡锤下降的高度抵消）。又因为沿缆索方向的“动量”为零，这次碰撞就如同在零动量参考系（zero-momentum frame）中观察的碰撞，而在这个参考系中，损失的动能所占的比例永远是 $1-e^2$。特别地，若 $e=0$，全部动能都会损失，整个系统静止下来；若 $e=1$，则毫无损失。

验算

- 取 $M=5$、$m=2$、$h=3$、$g=10$、$e=\frac{1}{3}$。则 $v^2=48$ 及 $U^2=3$，所以碰撞前的动能为 $\frac{1}{2}\cdot 2\cdot 48+\frac{1}{2}\cdot 8\cdot 3=60=mgh$，而损失为 $60\cdot\frac{8}{9}=\frac{160}{3}$。
- 以数值方法逐步模拟这个运动（瓷砖自由下坠，升降机以 $\frac{mg}{2M-m}$ 向上加速），随机选取 200 组 M、m、h、e 的值，所得的碰撞速率 v、升降机速率 U 和能量损失 $mgh(1-e^2)$ 均与上述结果吻合。
- 当 $m\to 0$ 时，升降机几乎不动，而 $v\to\sqrt{2gh}$，即自由落体的速率，与预期相符。
- 张力的冲量 $I=M(U-U')=M(1+e)U$ 为正数，所以缆索在碰撞期间保持拉紧。

常见错误

- 瓷砖脱落之后，仍把升降机的质量当作 M。这样的话，甚么都不会加速，瓷砖的速率便会是 $\sqrt{2gh}$。
- 忘记地板会上升迎向瓷砖，所以瓷砖相对于地面下坠的距离小于 h。
- 只对瓷砖和升降机应用动量守恒（conservation of momentum），忘记了经缆索与升降机相连的平衡锤。

STEP 之外

碰撞中守恒的量并不是普通的动量，而是与描述缆索位置的那个单一坐标共轭的动量：这是初尝广义坐标（generalised coordinates）和拉格朗日力学（Lagrangian mechanics）的滋味。总动量为零的碰撞恰好损失动能的 $1-e^2$ 这个比例，这个事实正是贯穿整个碰撞理论的质心参考系（centre-of-mass frame）思想。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）对质心参考系中的动量和连结体系统都有细致的论

述。

详细解答

STEP III 2006，第 12 题

概率与统计 · 正态近似与期望值

STEP III 2006，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/06-s3-q12

关键思路

多加一辆巴士，其价值只在于它平均能赚到的钱，所以公司应考虑的牌照费，最多只能是每年额外利润的*期望值*（expected value）。在现有车队已足以应付的日子，新增的巴士一分钱也赚不到；而它载客再多，也不会超过车上的 32 个座位。因此我们要求的，是一个“被截断”的随机变量（random variable）的期望值。题目的数字经过刻意挑选：在正态近似（normal approximation）之下，新旧载客量刚好落在分布上非常整齐的位置。

解答

建立模型

考虑单独一次抵达，设 X 为 1024 名游客中被指示前往巴士站的人数。每名游客前往巴士站的概率为 $\frac{1}{2}$，并假设各游客获得的指示彼此独立（independent）。于是

$$X \sim \mathrm{B}\left(1024, \tfrac{1}{2}\right), \qquad \mathrm{E}(X) = 512, \qquad \mathrm{Var}(X) = 1024 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 256.$$

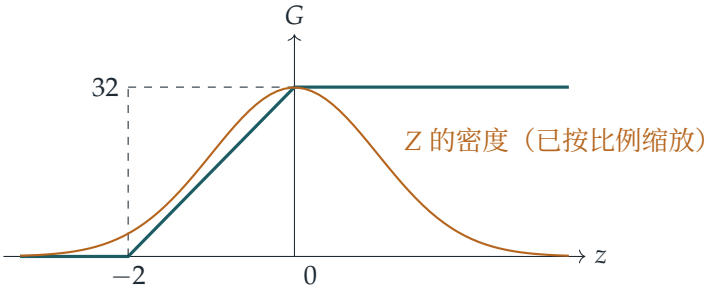
由于 n 很大而 $p = \frac{1}{2}$，我们采用正态近似 $X \approx \mathrm{N}(512, 16^2)$，并写成 $X = 512 + 16Z$，其中 $Z \sim \mathrm{N}(0,1)$。按题目指示，我们不作连续性校正（continuity correction）。

15 辆巴士可载 $15 \times 32 = 480$ 名乘客，所以现时每次抵达的利润是 $\pounds \min(X, 480)$。加上第 16 辆巴士后，载客量是 512，利润是 $\pounds \min(X, 512)$。因此每次抵达的额外利润是

$$G = \min(X,512) - \min(X,480) = \begin{cases} 0, & X \le 480, \\ X - 480, & 480 \le X \le 512, \\ 32, & X \ge 512. \end{cases}$$

以 Z 表示，两个分界点分别是 $X = 480 \iff Z = -2$ 和 $X = 512 \iff Z = 0$，而在中间的范围内 $X - 480 = 32 + 16Z$。所以

$$G = \begin{cases} 0, & Z \le -2, \\ 32 + 16Z, & -2 \le Z \le 0, \\ 32, & Z \ge 0. \end{cases}$$



额外利润的期望值

设 $\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}$ 为标准正态密度函数（standard normal density）。把期望值按三个范围拆开，得

$$\mathrm{E}(G) = 32\,\mathrm{P}(Z \ge 0) + \int_{-2}^{0} (32 + 16z)\,\phi(z)\,\mathrm{d}z.$$

第一项是 $32 \times \frac{1}{2} = 16$。至于积分，常数部分给出

$$32 \int_{-2}^{0} \phi(z)\,\mathrm{d}z = 32\big(\Phi(0) - \Phi(-2)\big) = 32\big(\tfrac{1}{2} - (1 - \Phi(2))\big) = 32\Phi(2) - 16,$$

其中用到了对称性 $\Phi(-2) = 1 - \Phi(2)$。

? **想一想** 没有数表，怎样求 $z\,\phi(z)$ 的积分?

对密度函数求导：$\phi'(z) = -z\,\phi(z)$。所以 $z\,\phi(z)$ 有精确的原函数（antiderivative）$-\phi(z)$，于是

$$16 \int_{-2}^{0} z\,\phi(z)\,\mathrm{d}z = 16\big[-\phi(z)\big]_{-2}^{0} = -16\big(\phi(0) - \phi(2)\big) = -\frac{16}{\sqrt{2\pi}}\big(1 - \mathrm{e}^{-2}\big).$$

（结果是负数，这正合道理：在整个范围内 $z < 0$。）

把三部分相加，两个 16 互相抵消：

$$\mathrm{E}(G) = 16 + \big(32\Phi(2) - 16\big) - \frac{16}{\sqrt{2\pi}}\big(1 - \mathrm{e}^{-2}\big) = 32\Phi(2) - \frac{16}{\sqrt{2\pi}}\big(1 - \mathrm{e}^{-2}\big).$$

牌照费

每年有 50 次抵达，所以根据期望值的线性性质（linearity of expectation），第 16 辆巴士每年带来的额外利润期望值是

$$50\,\mathrm{E}(G) = 1600\,\Phi(2) - \frac{800}{\sqrt{2\pi}}\left(1 - \mathrm{e}^{-2}\right) \text{ 英镑.}$$

若所付的牌照费高于此数，公司平均而言会亏本；而任何较低的费用都会留下期望收益。所以这就是值得考虑的最高年费（只是近似值，因为用了正态近似），亦即所求的结果。其数值约为 £1288。

验算

- 合理范围：新增的巴士每程赚取的金额介乎 0 与 32 之间，所以全年的数字必定介乎 0 与 1600 之间。我们的数值 $1600 \times 0.97725 - 319.15 \times 0.86466 \approx 1287.6$ 确实如此。
- 精确计算：以精确的 $\mathrm{B}(1024, \frac{1}{2})$ 概率对 $\min(k, 512) - \min(k, 480)$ 求和，再乘以 50，得 1287.69，与正态近似所得的 1287.64 相差不超过 6 便士。

常见错误

- 把额外利润当作 $32\,\mathrm{P}(X > 480)$，仿佛只要用得着新巴士，它就一定满座。其实在巴士乘客人数介乎 480 与 512 之间的日子，它只坐满了一部分。
- 计算 $\int_{-2}^{0} z\phi(z)\,\mathrm{d}z$ 时弄错正负号：被积函数是负的，所以答案必定是负数。

STEP 之外

$\mathrm{E}\big(\min(X, c)\big)$ 这个量，就是存货为 c 的报贩的期望销量；而如何选择 c 的“报贩问题”（newsvendor problem），是运筹学（operations research）的经典课题。同样的截断收益 $\min(\max(X - a, 0), b - a)$，在金融数学中称为“看涨期权价差”（call spread）。$\phi'(z) = -z\phi(z)$ 这个技巧，是截断正态分布的矩（truncated normal moments）理论的起点。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）涵盖了二项分布（binomial distribution）的正态近似，以及随机变量函数的期望值。

详细解答

STEP III 2006，第 13 题

概率与统计 • 几何概率

> **STEP III 2006，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/06-s3-q13

关键思路

周界由两段性质迥异的部分组成：一段长度为 π 的弧和一条长度为 2 的直径，所以我们按每一点落在哪一段来取条件（condition）。在三种情况的每一种之中，三角形的面积都有简单的公式 $\frac{1}{2}ab\sin C$，其中顶点 C 位于圆心，而各个条件期望值（conditional expectation）都只是简短的积分。最后用全期望定律（law of total expectation）把各部分重新合并起来。

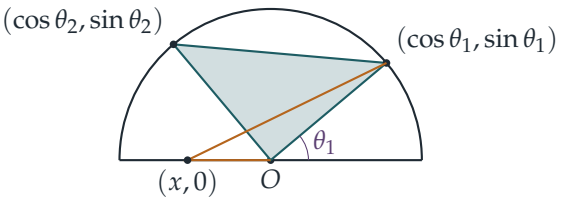
解答

设置

把直径的中点 O 放在原点，直径沿 x 轴由 $(-1,0)$ 伸展至 $(1,0)$，而弧则位于上半平面。“在周界上随机选取”的意思是沿整条边界按长度均匀分布（uniformly distributed），边界的总长度为 $\pi+2$。因此每一点各自独立地

- 以概率 $\dfrac{\pi}{\pi+2}$ 落在弧上，这时它是 $(\cos\theta,\sin\theta)$，其中 θ 在 $[0,\pi]$ 上均匀分布；
- 以概率 $\dfrac{2}{\pi+2}$ 落在直径上，这时它是 $(x,0)$，其中 x 在 $[-1,1]$ 上均匀分布。

（在弧上按长度均匀分布，等同于按角度均匀分布，因为单位圆上的弧长等于所对的角。）



设 D 为两点之中落在直径上的点的数目。那么 $D = 2$、1、0 的概率分别为

$$\frac{4}{(\pi+2)^2}, \qquad \frac{2\cdot 2\pi}{(\pi+2)^2} = \frac{4\pi}{(\pi+2)^2}, \qquad \frac{\pi^2}{(\pi+2)^2},$$

这里用到两点互相独立（independent）（中间一种情况的因子 2 是计算哪一点落在直径上）。在每一种情况下，三角形都有一个顶点在 O，所以它的面积是 $\frac{1}{2}\,|OP|\,|OQ|\sin\angle POQ$。

两点都在直径上

三个顶点全在同一直线上，所以三角形是退化（degenerate）的，$A = 0$。因此 $\mathrm{E}(A \mid D = 2) = 0$。

一点在弧上，一点在直径上

设这两点为 $P = (\cos\theta, \sin\theta)$ 和 $R = (x, 0)$。那么 $|OP| = 1$，$|OR| = |x|$，而底边 OR 沿着直径，P 则在它上方高度 $\sin\theta \geq 0$ 之处，所以

$$A = \tfrac{1}{2}\,|x|\sin\theta.$$

由于 x 与 θ 互相独立，

$$\mathrm{E}(A \mid D = 1) = \tfrac{1}{2}\,\mathrm{E}|x|\,\mathrm{E}(\sin\theta) = \tfrac{1}{2}\cdot\frac{1}{2}\int_{-1}^{1}|x|\,\mathrm{d}x\cdot\frac{1}{\pi}\int_0^{\pi}\sin\theta\,\mathrm{d}\theta = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\frac{2}{\pi} = \frac{1}{2\pi}.$$

两点都在弧上

现在 $P = (\cos\theta_1, \sin\theta_1)$ 和 $Q = (\cos\theta_2, \sin\theta_2)$，其中 θ_1, θ_2 互相独立，并在 $[0, \pi]$ 上均匀分布。两条半径的长度都是 1，而角 POQ 等于 $|\theta_1 - \theta_2|$，所以

$$A = \tfrac{1}{2}\,|\sin(\theta_1 - \theta_2)|.$$

想一想 要求出 $\mathrm{E}|\sin(\theta_1 - \theta_2)|$，我们是否需要知道 $\theta_1 - \theta_2$ 的分布？

不需要：先对 θ_2 取条件。对于固定的 θ_2，当 θ_1 走遍 $[0, \pi]$ 时，角 $\theta_1 - \theta_2$ 走遍一个长度为 π 的区间。函数 $|\sin u|$ 的周期（period）为 π，所以它在任何长度为 π 的区间上的积

分，都与在 $[0,\pi]$ 上的积分相同：

$$\int_0^{\pi} |\sin(\theta_1-\theta_2)|\,\mathrm{d}\theta_1 = \int_0^{\pi} \sin u\,\mathrm{d}u = 2.$$

所以无论 θ_2 是甚么，都有 $\mathrm{E}\big(|\sin(\theta_1-\theta_2)| \,\big|\, \theta_2\big) = 2/\pi$，因此

$$\mathrm{E}(A \mid D=0) = \tfrac{1}{2}\cdot\frac{2}{\pi} = \frac{1}{\pi}.$$

合并结果

根据全期望定律，

$$\begin{aligned}\mathrm{E}(A) &= \frac{4}{(\pi+2)^2}\cdot 0 + \frac{4\pi}{(\pi+2)^2}\cdot\frac{1}{2\pi} + \frac{\pi^2}{(\pi+2)^2}\cdot\frac{1}{\pi}\\ &= \frac{2+\pi}{(\pi+2)^2} = \frac{1}{2+\pi},\end{aligned}$$

这就是 $(2+\pi)^{-1}$，正如题目所求。

验算

- 三种情况的概率之和 $\dfrac{4+4\pi+\pi^2}{(\pi+2)^2}$ 等于 1。
- 每个条件期望值都不超过该情况下可能出现的最大面积：$\frac{1}{\pi}\approx 0.318 \le \frac{1}{2}$（两条半径互相垂直）以及 $\frac{1}{2\pi}\approx 0.159 \le \frac{1}{2}$。
- 用 10^6 对点进行蒙地卡罗模拟（Monte Carlo simulation），得出 $\mathrm{E}(A)\approx 0.1943$，而 $1/(2+\pi)\approx 0.1945$；数值积分（quadrature）亦证实了条件期望值 $1/\pi$ 和 $1/(2\pi)$。

常见错误

- 以各 $\frac{1}{2}$ 的概率选择每一点落在弧上或直径上。“在周界上随机选取”是按每一段的长度来加权的。
- 在 $\mathrm{P}(D=1)$ 中漏掉因子 2：两点中任何一点都可以是落在直径上的那一点。
- 写成 $A=\frac{1}{2}\sin(\theta_1-\theta_2)$ 而漏掉绝对值（modulus），这样由对称性得出的期望值会是 0。

STEP 之外

这是几何概率（geometric probability）的一个小例子，这门学问的经典例子包括蒲丰投针问题（Buffon's needle）和贝特朗悖论（Bertrand's paradox）。贝特朗悖论说明了为何“随机”一词必须界定清楚，正如我们以沿周界的长度作为测度（measure）那样。这里真正发挥作用的工具，即对一个划分（partition）取条件然后求平均（全期望定律），是大学一年级概率论的核心内容。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）对此有很好的讲解。

详细解答

STEP III 2006，第 14 题

概率与统计 · 协方差与独立性

STEP III 2006，第 14 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/06-s3-q14

关键思路

这里的一切都建基于两个事实：期望值(expectation)是线性的；而对于独立(independent)的变量，乘积的期望值等于期望值的乘积。方差(variance)可由 $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$ 求得，所以我们自始至终只需要计算期望值。要证明两个变量并非独立，只须找出独立性的一个推论不成立便已足够，而最容易检验的就是 $\mathrm{E}(UV) = \mathrm{E}(U)\,\mathrm{E}(V)$。

解答

两个事实

对于任意随机变量（random variable）及常数 a、b，

$$\mathrm{E}(aX_1 + bX_2) = a\,\mathrm{E}(X_1) + b\,\mathrm{E}(X_2),$$

而若 X_1 与 X_2 独立，则

$$\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\,\mathrm{E}(X_2).$$

第二个事实同样适用于这些变量的函数：若 X_1、X_2 独立，则 X_1^2 与 X_2、X_1 与 X_2^2，以及 X_1^2 与 X_2^2 亦各自独立。我们还会用到 $\mathrm{E}(X_i^2) = \sigma_i^2 + \mu_i^2$，这不过是把 $\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - \mu_i^2$ 移项而已。

周界与面积

周界为 $P = 2X_1 + 2X_2$，面积为 $A = X_1X_2$。由线性及独立性，

$$\mathrm{E}(P) = 2(\mu_1 + \mu_2), \qquad \mathrm{E}(A) = \mu_1\mu_2.$$

要求 P 的方差，我们从 $\mathrm{E}(P^2)$ 入手：

$$\begin{aligned}\mathrm{E}(P^2) &= 4\,\mathrm{E}(X_1^2 + 2X_1X_2 + X_2^2)\\ &= 4(\sigma_1^2 + \mu_1^2 + 2\mu_1\mu_2 + \sigma_2^2 + \mu_2^2)\\ &= 4(\sigma_1^2 + \sigma_2^2) + 4(\mu_1 + \mu_2)^2,\end{aligned}$$

其中用到线性及 $\mathrm{E}(X_1X_2) = \mu_1\mu_2$。减去 $[\mathrm{E}(P)]^2 = 4(\mu_1 + \mu_2)^2$，得

$$\mathrm{Var}(P) = 4(\sigma_1^2 + \sigma_2^2), \qquad \text{所以}P\text{ 的标准差（standard deviation）为}2\sqrt{\sigma_1^2 + \sigma_2^2}.$$

至于 A，由于 X_1^2 与 X_2^2 独立，

$$\mathrm{E}(A^2) = \mathrm{E}(X_1^2)\,\mathrm{E}(X_2^2) = (\sigma_1^2 + \mu_1^2)(\sigma_2^2 + \mu_2^2),$$

减去 $\mu_1^2\mu_2^2$ 后余下

$$\mathrm{Var}(A) = \sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2,$$

所以 A 的标准差为

$$\sqrt{\sigma_1^2\sigma_2^2 + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}.$$

P 与 A 并非独立

假如 P 与 A 独立，便会有 $\mathrm{E}(PA) = \mathrm{E}(P)\,\mathrm{E}(A)$。现在 $PA = 2X_1^2X_2 + 2X_1X_2^2$，所以由线性以及 X_1 与 X_2 的独立性，

$$\begin{aligned}\mathrm{E}(PA) &= 2(\sigma_1^2 + \mu_1^2)\mu_2 + 2\mu_1(\sigma_2^2 + \mu_2^2)\\ &= 2(\mu_1 + \mu_2)\mu_1\mu_2 + 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)\\ &= \mathrm{E}(P)\,\mathrm{E}(A) + 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2).\end{aligned}$$

长度和阔度都是正数, 所以 $\mu_1, \mu_2 > 0$, 而且 $\sigma_1, \sigma_2 \neq 0$, 因此多出来的一项 $2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)$ 严格为正。故此 $\mathrm{E}(PA) \neq \mathrm{E}(P)\,\mathrm{E}(A)$，即 P 与 A 并非独立。

变量 $Z = P - \alpha A$

同样地，若 Z 与 A 独立，则必有 $\mathrm{E}(ZA) = \mathrm{E}(Z)\,\mathrm{E}(A)$。由线性,

$$\begin{aligned}\mathrm{E}(ZA) - \mathrm{E}(Z)\,\mathrm{E}(A) &= \big[\mathrm{E}(PA) - \mathrm{E}(P)\,\mathrm{E}(A)\big] - \alpha\big[\mathrm{E}(A^2) - [\mathrm{E}(A)]^2\big]\\ &= 2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2) - \alpha(\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2),\end{aligned}$$

其中用到上面的两个结果。与 α 相乘的括号就是 $\mathrm{Var}(A)$, 它是正数 (当中含有 $\sigma_1^2\sigma_2^2 > 0$), 所以这个式子恰好只在 α 的一个值处等于零:

$$\alpha_0 = \frac{2(\mu_1\sigma_2^2 + \mu_2\sigma_1^2)}{\mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2 + \sigma_1^2\sigma_2^2}.$$

若 $\alpha \neq \alpha_0$，则 $\mathrm{E}(ZA) \neq \mathrm{E}(Z)\,\mathrm{E}(A)$，所以 Z 与 A 并非独立，这正是所要证明的。

想一想 $\alpha = \alpha_0$ 是否表示 Z 与 A 确实独立?

不是。协方差（covariance）为零是独立性的推论，却不能保证独立。当 $\alpha = \alpha_0$ 时，我们的检验根本提供不了任何信息，必须另觅论证。最后一部分说明，在一个具体例子中，独立性在该处同样不成立。

最后一部分

这里 $\mu_1 = \mu_2 = 2$ 而 $\sigma_1^2 = \sigma_2^2 = 1$，所以

$$\alpha_0 = \frac{2(2+2)}{4+4+1} = \frac{8}{9}.$$

由上一部分，对任何 $\alpha \neq \frac{8}{9}$，Z 与 A 都并非独立。若要用一个论证涵盖所有 α，可列出四个等可能的结果：

(X_1, X_2)	A	P	$Z = P - \alpha A$
$(1,1)$	1	4	$4 - \alpha$
$(1,3), (3,1)$	3	8	$8 - 3\alpha$
$(3,3)$	9	12	$12 - 9\alpha$

关键的观察是：P 完全由 A 决定，因而 Z 也是如此。设 $z = 4 - \alpha$。知道 $A = 1$ 便必有 $Z = z$，所以

$$\mathrm{P}(Z = z,\ A = 1) = \mathrm{P}(A = 1) = \tfrac{1}{4}.$$

若要独立，这个概率便须等于 $\mathrm{P}(Z = z)\,\mathrm{P}(A = 1)$，即 $\mathrm{P}(Z = z) = 1$，于是 Z 在每一个结果上都要取值 $4 - \alpha$。这需要 $8 - 3\alpha = 4 - \alpha$，即 $\alpha = 2$，同时又需要 $12 - 9\alpha = 4 - \alpha$，即 $\alpha = 1$。两者不可能同时成立，所以 $\mathrm{P}(Z = z) < 1$，独立性不成立。

因此，无论 α 取甚么值，Z 与 A 都并非独立。例如当 $\alpha = \frac{8}{9}$ 时，Z 的取值为 $\frac{28}{9}$、$\frac{16}{3}$ 和 4，而 $\mathrm{P}(Z = \frac{28}{9}, A = 1) = \frac{1}{4}$，但 $\mathrm{P}(Z = \frac{28}{9})\,\mathrm{P}(A = 1) = \frac{1}{16}$。

验算

- 若完全没有变异（$\sigma_1 = \sigma_2 = 0$），两个标准差都化为 0，理应如此。
- 在最后一部分，按公式得 $\mathrm{Var}(A) = 1 + 4 + 4 = 9$；直接计算，$A$ 取值 $1, 3, 9$ 的概率分别为 $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$，得出 $\mathrm{E}(A) = 4$、$\mathrm{E}(A^2) = 25$，方差为 9。
- 我们以分数作精确计算，在 200 个随机的离散分布上验证了平均值、方差和协方差的一般公式。我们亦确认了在 α_0 处 $\mathrm{E}(ZA) = \mathrm{E}(Z)\mathrm{E}(A)$，确认了最后一部分中 $\alpha_0 = \frac{8}{9}$，并确认了在该部分中，对于 $\alpha = \frac{8}{9}$ 及另外 500 个 α 值，Z 与 A 均不独立。

常见错误

- 写成 $\mathrm{E}(X_1^2) = \mu_1^2$。正是方差项 σ_1^2 令 $\mathrm{Var}(A)$ 和协方差不为零。
- 当 $\alpha = \alpha_0$ 时断定 Z 与 A 独立。协方差为零并不蕴涵独立。
- 忘记说明为何 $\mu_2\sigma_1^2 + \mu_1\sigma_2^2 \neq 0$。正是因为长度为正数，各项才不可能互相抵销。

STEP 之外

$\mathrm{E}(UV) - \mathrm{E}(U)\mathrm{E}(V)$ 这个量就是协方差，而特殊值 α_0 等于 $\mathrm{Cov}(P, A)/\mathrm{Var}(A)$，即 P 对 A 的最小二乘回归（least-squares regression）直线的斜率。取 $\alpha = \alpha_0$ 会令残差（residual）Z 与 A 不相关（uncorrelated），但并不令两者独立。对于联合正态（jointly normal）的变量，这两个概念是一致的，这也是正态分布如此方便的原因之一。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）对协方差、相关与独立性有细致的论述。

Beyond STEP

STEP 是大多数中学生一生中应考过最艰深的数学考试，也是最出色的考试之一：题目简短、陌生而深刻，当中最精彩的，其实是乔装了的大学数学小品。

本书精选最有趣的 STEP 试题，由 1987 年首次开考之前的样本试卷（specimen papers），到 2025 年的试卷，并以每一道题目作为一扇门。每道题目都会仔细解答，沿途提出优秀学生会问的问题，然后带你穿过这扇门：这个想法从何而来，到了大学会演变成甚么，以及接下来该读些甚么。

本书为正在准备 STEP、有天分又有志投身数学的学生而写。他们想要的不止是模拟试卷：他们想开始像数学系本科生那样思考，并一探前方有甚么正等着自己。

- 共 13 章，涵盖纯数学（pure mathematics）、力学（mechanics）、概率（probability）与统计（statistics）
- 136 道 STEP 历届试题，每题均附引导式解答
- 每道题目背后的大学数学
- 附答案的延伸练习，以及延伸阅读

2006 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。

