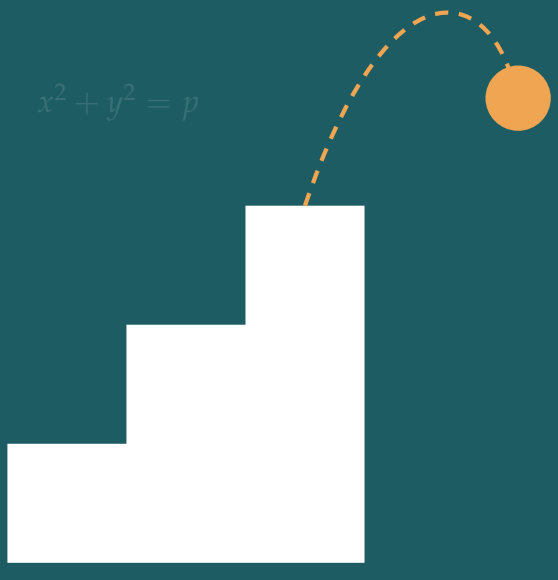
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2015 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

2015 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2015 年详细解答

2015 年 STEP 试卷（STEP I, II 及 III）全部 39 题的详细解答。自样本试卷以来每一道 STEP 题目的详细解答、*Beyond STEP* 全书以及各章的单行本，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

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

由机器从英文翻译而成；尚未经审阅。

2026 年 11 月首次于网上发表。

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

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

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

目录

STEP I

第 1 题 曲线描绘 纯数学 . . . 3

第 2 题 三次方程的三角解法 纯数学 . . . 7

第 3 题 坐标几何中的可见性 纯数学 . . . 11

第 4 题 参数曲线与面积 纯数学 . . . 15

第 5 题 含参数的积分 纯数学 . . . 18

第 6 题 几何中的向量 纯数学 . . . 22

第 7 题 区间上的最大值 纯数学 . . . 25

第 8 题 幂和的整除性 纯数学 . . . 28

第 9 题 抛体运动 力学 . . . 32

第 10 题 相对速度 力学 . . . 35

第 11 题 摩擦力与力矩 力学 . . . 39

第 12 题 泊松分布的稀疏化 概率与统计 . . . 43

第 13 题 条件概率 概率与统计 . . . 47

STEP II

第 1 题 不等式与级数 纯数学 . . . 51

第 2 题 三角形几何 纯数学 . . . 54

第 3 题 计数与数学归纳法 纯数学 . . . 57

第 4 题 反正切函数的分支 纯数学 . . . 61

第 5 题 反正切级数 纯数学 . . . 67

第 6 题 利用对称性积分 纯数学 . . . 71

第 7 题 相交圆 纯数学 . . . 75

第 8 题 向量与相切圆 纯数学 . . . 79

第 9 题 刚体转动 力学 . . . 83

第 10 题 约束运动 力学 . . . 87
第 11 题 动量与能量 力学 . . . 90
第 12 题 硬币序列游戏 概率与统计 . . . 95
第 13 题 指数分布的矩 概率与统计 . . . 99

STEP III

第 1 题 归约公式与代换 纯数学 . . . 103
第 2 题 分析学 纯数学 . . . 107
第 3 题 极坐标曲线与面积 纯数学 . . . 111
第 4 题 三次方程的根 纯数学 . . . 115
第 5 题 以递降法证明无理性 纯数学 . . . 118
第 6 题 根的对称函数 纯数学 . . . 121
第 7 题 微分算子 纯数学 . . . 125
第 8 题 极坐标微分方程 纯数学 . . . 128
第 9 题 非线性弹性振动 力学 . . . 133
第 10 题 转动的刚杆 力学 . . . 137
第 11 题 圆周运动 力学 . . . 141
第 12 题 概率生成函数 概率与统计 . . . 145
第 13 题 均匀分布变量的函数 概率与统计 . . . 149

详细解答

STEP I 2015，第 1 题

纯数学 · 曲线描绘

> **STEP I 2015，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/15-s1-q1

关键思路

能因式分解（factorise）的都先因式分解。这个二次式可以分解，所以零点（zeros）一目了然；而导数（derivative）同样是 e^x 乘以一个可分解的二次式，所以转向点（turning points）也一目了然。草图画对之后，要数“曲线 $= k$”有多少个解，只需把一条水平线上下移动，看它与曲线相交多少次。至于第 (ii) 部分，留意新函数其实就是把旧函数在 x^2 处取值，所以几乎所有结果都可以沿用。

解答

第 (i) 部分

设 $f(x) = e^x(2x^2 - 5x + 2) = e^x(2x-1)(x-2)$。

零点与正负号。由于 $e^x > 0$，$f(x) = 0$ 当且仅当 $x = \frac{1}{2}$ 或 $x = 2$。此外，f 在 $x < \frac{1}{2}$ 及 $x > 2$ 时为正，在 $(\frac{1}{2}, 2)$ 上为负。曲线与 y 轴相交于 $f(0) = 2$。

转向点。由乘积法则（product rule），

$$f'(x) = e^x(2x^2 - 5x + 2) + e^x(4x - 5) = e^x(2x^2 - x - 3) = e^x(2x-3)(x+1).$$

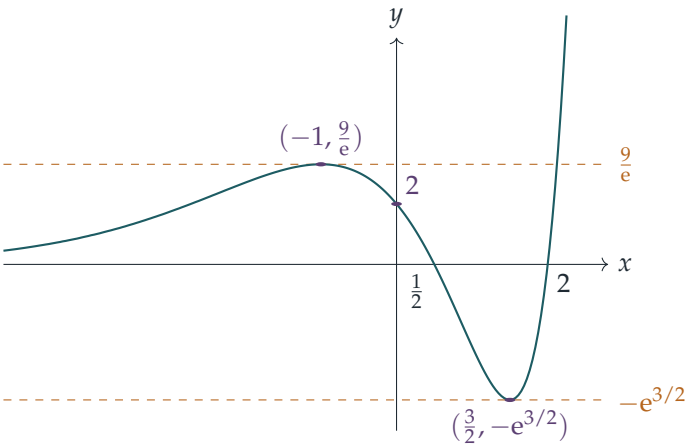
所以 $f'(x) = 0$ 于 $x = -1$ 及 $x = \frac{3}{2}$。f' 的正负号与 $(2x-3)(x+1)$ 相同：$x < -1$ 时为正，在 $(-1, \frac{3}{2})$ 上为负，$x > \frac{3}{2}$ 时为正。因此 f 先递增，后递减，再递增，并且

$$f(-1) = e^{-1}(2 + 5 + 2) = \frac{9}{e} \approx 3.31 \qquad \text{（局部极大值，local maximum），}$$

$$f(\tfrac{3}{2}) = e^{3/2}(\tfrac{9}{2} - \tfrac{15}{2} + 2) = -e^{3/2} \approx -4.48 \qquad \text{（局部极小值，local minimum）.}$$

两端的走势。当 $x \to \infty$ 时，两个因式都增大，所以 $f(x) \to \infty$。当 $x \to -\infty$ 时，由题目

给出的事实（取 $n = 2, 1, 0$）可知 $2x^2e^x$、$-5xe^x$ 及 $2e^x$ 各自趋于 0，所以 $f(x) \to 0$。由于 $f > 0$ 对 $x < \frac{1}{2}$ 成立，曲线从上方趋近负 x 轴：x 轴是一条渐近线（asymptote）。



数算 $f(x) = k$ 的解。解的数目就是水平线 $y = k$ 与曲线相交的次数。逐一考虑曲线的三段单调（monotonic）部分：

- 在 $(-\infty, -1]$ 上，f 严格递增，取遍 $(0, \frac{9}{e}]$ 中每个值恰好一次（它永远达不到 0）；
- 在 $[-1, \frac{3}{2}]$ 上，f 由 $\frac{9}{e}$ 严格递减至 $-e^{3/2}$，取遍其间每个值恰好一次；
- 在 $[\frac{3}{2}, \infty)$ 上，f 由 $-e^{3/2}$ 严格递增至 ∞，取遍每个值恰好一次。

把各段的贡献加起来（并小心不要把 $x = -1$ 或 $x = \frac{3}{2}$ 重复计算）：

$k < -e^{3/2}$	无解
$k = -e^{3/2}$	一个解（$x = \frac{3}{2}$）
$-e^{3/2} < k \leq 0$	两个解
$0 < k < \frac{9}{e}$	三个解
$k = \frac{9}{e}$	两个解（$x = -1$ 及一个 $x > 2$ 的解）
$k > \frac{9}{e}$	一个解

想一想 为甚么 $k = 0$ 时有两个解，而不是三个？

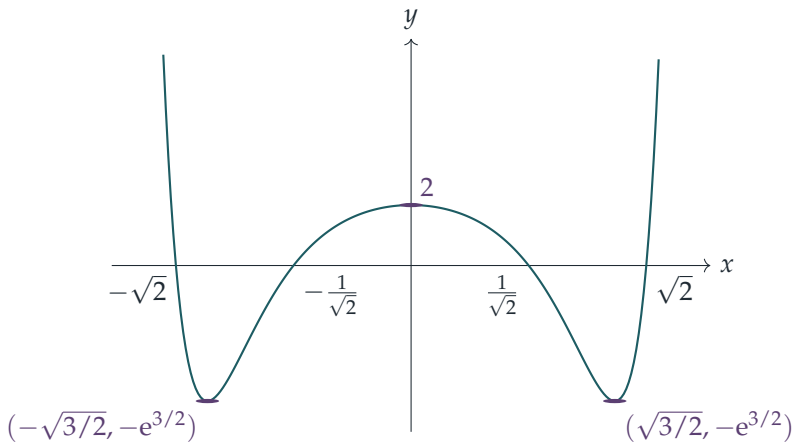
x 轴是左方的渐近线，而且对每个 $x < \frac{1}{2}$ 都有 $f(x) > 0$，所以最左边的分支永远不会真正到达 0。$f(x) = 0$ 的解只有 $x = \frac{1}{2}$ 和 $x = 2$。不过，对任何很小的 $k > 0$，左边的分支确实能到达高度 k，所以解的数目会跳升至三个。

第 (ii) 部分

设 $g(x) = \mathrm{e}^{x^2}(2x^4 - 5x^2 + 2)$。那么 $g(x) = f(x^2)$，所以 g 是偶函数（even function），其图像关于 y 轴对称。当 x 取遍 $[0, \infty)$ 时，$u = x^2$ 递增地取遍 $[0, \infty)$，所以对 $x \geq 0$，g 的图像只是重走 f 的图像中 $u \geq 0$ 的部分：

- $g(0) = f(0) = 2$；
- 零点在 $x^2 = \frac{1}{2}$ 或 $x^2 = 2$ 处，即 $x = \pm\frac{1}{\sqrt{2}}$ 及 $x = \pm\sqrt{2}$；
- 由链式法则（chain rule），$g'(x) = 2x\,f'(x^2) = 2x\,\mathrm{e}^{x^2}(2x^2 - 3)(x^2 + 1)$，它只在 $x = 0$ 及 $x = \pm\sqrt{3/2}$ 处为零。（f 在 $u = -1$ 处的转向点消失了，因为 $x^2 = -1$ 没有实数解。）

当 $0 < x < \sqrt{3/2}$ 时，$g'(x) < 0$；当 $x > \sqrt{3/2}$ 时，$g'(x) > 0$。所以 $(0, 2)$ 是局部极大值点（由对称性，在 0 的左边附近 $g' > 0$），而 $x = \pm\sqrt{3/2}$ 是极小值点，极小值为 $f(\frac{3}{2}) = -\mathrm{e}^{3/2}$。最后，当 $x \to \pm\infty$ 时 $g(x) \to \infty$，而且由于因子 e^{x^2}，增长得非常快。曲线呈“W”形：



验算

- $f(-1) = 9/\mathrm{e} \approx 3.311$ 及 $f(1.5) = -\mathrm{e}^{1.5} \approx -4.482$；在附近几点计算 f 的值，可确认它们分别是极大值和极小值。
- 在 $x = 0$ 附近，$g(x) = f(x^2) \approx f(0) + f'(0)x^2 = 2 - 3x^2$，是一条开口向下的抛物线（parabola），确认了 $(0, 2)$ 是局部极大值点。
- 已用电脑对每个范围内的样本 k 值核实根的数目（例如 $k = 2$ 有三个根，$k = -1$ 有两个根）。

- $f(-40) \approx 1.4 \times 10^{-14}$：正数，而且极小，与渐近线的预测一致。

常见错误

- 忘记曲线在左方只是趋近 x 轴，因而在 $k = 0$ 时数出三个解。
- 漏掉边界情况 $k = \frac{9}{\mathrm{e}}$ 及 $k = -\mathrm{e}^{3/2}$；在这些情况下，水平线与曲线相切（tangent），两个解合而为一。
- 在第 (ii) 部分，保留了对应 $x^2 = -1$ 的转向点。

STEP 之外

上面的数解论证建基于介值定理（intermediate value theorem）加上严格单调性（strict monotonicity）：在区间上连续且严格递增的函数，会取遍两端值之间的每个值恰好一次。大学的数学分析课程会严谨地证明这些事实，以及题目容许你假设的极限 $x^n \mathrm{e}^x \to 0$。Abbott 的 *Understanding Analysis*（《理解分析》）是一本非常易读的书，可以看看这些证明是怎样做的。

详细解答

STEP I 2015，第 2 题

纯数学 · 三次方程的三角解法

> **STEP I 2015，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/15-s1-q2

关键思路

三次式（cubic）$4x^3 - 3x$ 正正就是三倍角公式（triple-angle formula）所给出的形式：$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha$。因此，这种形式的三次方程可以通过猜一个余弦值来求解。一旦知道其中一个根（root），便可以因式分解（factorise），再解余下的二次方程（quadratic）。在最后一部分，代换 $y = 2x$ 把题目给出的三次方程化成这种形式，而第 (i) 部分则提供了我们所需的准确值。

解答

第 (i) 部分

把 $15^\circ = 45^\circ - 30^\circ$ 写出，并使用复角公式（compound-angle formulae）：

$$\cos 15^\circ = \cos 45^\circ \cos 30^\circ + \sin 45^\circ \sin 30^\circ = \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{3}}{2} + \frac{1}{\sqrt{2}} \cdot \frac{1}{2} = \frac{\sqrt{3}+1}{2\sqrt{2}},$$

此即所求。同样地，

$$\sin 15^\circ = \sin 45^\circ \cos 30^\circ - \cos 45^\circ \sin 30^\circ = \frac{\sqrt{3}-1}{2\sqrt{2}}.$$

把分母有理化（rationalising the denominator）后，得到等价的形式

$$\cos 15^\circ = \frac{\sqrt{6}+\sqrt{2}}{4}, \qquad \sin 15^\circ = \frac{\sqrt{6}-\sqrt{2}}{4},$$

这两个形式在第 (iii) 部分会更方便使用。

第 (ii) 部分

设 $c = \cos\alpha$ 及 $s = \sin\alpha$。展开 $\cos(2\alpha + \alpha)$，

$$\begin{aligned}\cos 3\alpha &= \cos 2\alpha \cos\alpha - \sin 2\alpha \sin\alpha \\ &= (2c^2 - 1)c - 2s^2 c \\ &= 2c^3 - c - 2(1 - c^2)c \\ &= 4c^3 - 3c.\end{aligned}$$

所以 $4c^3 - 3c - \cos 3\alpha = 0$：$x = \cos\alpha$ 是一个根。

至于其余的根，把 $\cos 3\alpha$ 换成 $4c^3 - 3c$，然后因式分解：

$$\begin{aligned}4x^3 - 3x - (4c^3 - 3c) &= 4(x^3 - c^3) - 3(x - c) \\ &= (x - c)\big(4x^2 + 4cx + 4c^2 - 3\big).\end{aligned}$$

? **想一想** 该如何化简这个二次因式的判别式?

利用 $c^2 + s^2 = 1$，判别式（discriminant）为 $16c^2 - 16(4c^2 - 3) = 48(1 - c^2) = 48s^2$。它的平方根是 $4\sqrt{3}\,|s|$，但由于我们同时取 $\pm$ 两个符号，所以可以略去绝对值。二次公式（quadratic formula）给出

$$x = \frac{-4c \pm 4\sqrt{3}\,s}{8},$$

因此另外两个根是

$$x = -\tfrac{1}{2}\cos\alpha + \tfrac{\sqrt{3}}{2}\sin\alpha \quad \text{及} \quad x = -\tfrac{1}{2}\cos\alpha - \tfrac{\sqrt{3}}{2}\sin\alpha.$$

它们就是 $\cos(\alpha - 120^\circ)$ 和 $\cos(\alpha + 120^\circ)$。这是合理的：$3(\alpha \pm 120^\circ) = 3\alpha \pm 360^\circ$，所以这两个角的 $\cos 3\theta$ 与 α 的相同。

第 (iii) 部分

? **想一想** 题目给出的三次式是 $y^3 - 3y - \sqrt{2}$，而不是 $4x^3 - 3x - \cdots$。该如何把两者配合起来?

我们希望三次项与一次项的系数（coefficient）之比为 4 比 3，与第 (ii) 部分一样。设 $y = 2x$：

$$8x^3 - 6x - \sqrt{2} = 0 \iff 4x^3 - 3x - \frac{\sqrt{2}}{2} = 0.$$

这正是第 (ii) 部分的方程，其中 $\cos 3\alpha = \frac{\sqrt{2}}{2} = \cos 45^\circ$，所以可以取 $\alpha = 15^\circ$。第 (ii) 部分于是给出三个根 x。它们互不相同（见下文），因此就是这个三次方程的全部根，而由

$y = 2x$ 得

$$y = 2\cos\alpha, \qquad y = -\cos\alpha \pm \sqrt{3}\sin\alpha, \qquad \alpha = 15^\circ.$$

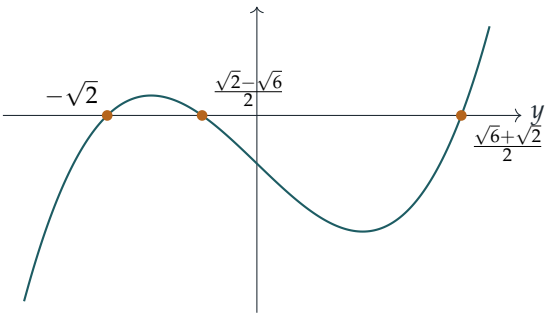
现在代入第 (i) 部分的值。

- $y = 2\cos 15^\circ = \dfrac{\sqrt{3}+1}{\sqrt{2}} = \dfrac{\sqrt{6}+\sqrt{2}}{2}$。
- $y = -\cos 15^\circ + \sqrt{3}\sin 15^\circ = \dfrac{-(\sqrt{6}+\sqrt{2})+(\sqrt{18}-\sqrt{6})}{4}$，化简后得 $\dfrac{2\sqrt{2}-2\sqrt{6}}{4} = \dfrac{\sqrt{2}-\sqrt{6}}{2}$。
- $y = -\cos 15^\circ - \sqrt{3}\sin 15^\circ = \dfrac{-(\sqrt{6}+\sqrt{2})-(3\sqrt{2}-\sqrt{6})}{4} = -\sqrt{2}$。

这里用了 $\sqrt{3}\sqrt{6} = \sqrt{18} = 3\sqrt{2}$。这三个值显然互不相同。所以解为

$$y = \frac{\sqrt{6}+\sqrt{2}}{2}, \qquad y = \frac{\sqrt{2}-\sqrt{6}}{2}, \qquad y = -\sqrt{2}.$$

以角度表示，它们分别是 $2\cos 15^\circ$、$2\cos 255^\circ$ 及 $2\cos 135^\circ$。



验算

- $y^3 - 3y - \sqrt{2}$ 的 y^2 系数为零，所以各根之和必为零。的确，$\frac{\sqrt{6}+\sqrt{2}}{2} + \frac{\sqrt{2}-\sqrt{6}}{2} - \sqrt{2} = 0$。
- 各根之积应为 $\sqrt{2}$。由于 $(\sqrt{2}+\sqrt{6})(\sqrt{2}-\sqrt{6}) = 2-6 = -4$，各根之积为 $(-1)\cdot(-\sqrt{2}) = \sqrt{2}$。
- 根 $y = -\sqrt{2}$ 很容易直接验证：$-2\sqrt{2}+3\sqrt{2}-\sqrt{2} = 0$。以数值计算，$\frac{\sqrt{6}+\sqrt{2}}{2} \approx 1.93185$，而 $(1.93185)^3 - 3(1.93185) \approx 1.41421 = \sqrt{2}$。
- 在第 (ii) 部分中，$\alpha = 0$ 给出方程 $4x^3 - 3x - 1 = 0$，即 $(x-1)(2x+1)^2 = 0$。公式给出的根为 $1, -\frac{1}{2}, -\frac{1}{2}$，与预期相符。

常见错误

- 在第 (iii) 部分忘记作伸缩代换，结果解了 $4x^3 - 3x - \sqrt{2} = 0$。这个方程需要 $\cos 3\alpha = \sqrt{2} > 1$，这是不可能的。
- 给出 x 的答案，而不是 $y = 2x$ 的答案。
- 写出 $\sqrt{3(1-c^2)} = \sqrt{3}\sin\alpha$，却没有指出 $\pm$ 号已照顾到 $\sin\alpha < 0$ 的情况。

STEP 之外

这是解有三个实根的三次方程的三角方法，归功于韦达（Viète）。任何三次方程都可以经平移消去平方项，得到 $t^3 + pt + q = 0$。当 $p < 0$ 时，伸缩代换 $t = 2\sqrt{-p/3}\,x$ 把它化成 $4x^3 - 3x = k$。在这种情况下，卡尔达诺（Cardano）公式帮不上忙，因为即使三个根全是实数，它也要经过负数的平方根。弄明白为何会这样，以及为何余弦技巧能避开这个问题，便会自然地引向复数（complex numbers）和单位根（roots of unity）。Needham 的 *Visual Complex Analysis*（《直观复分析》）对此给出了优美的几何图像。

详细解答

STEP I 2015，第 3 题

纯数学 • 坐标几何中的可见性

STEP I 2015，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：

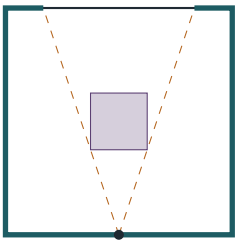
step.maths.org/questions/15-s1-q3

关键思路

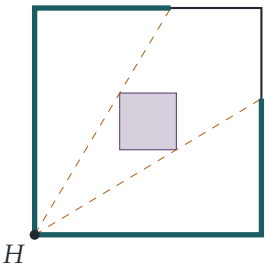
守卫的视线会被建筑物挡住；由于建筑物是一个凸（convex）正方形，刚好擦过建筑物某个角的直线，就标示出墙上可见部分的终点。因此，对每个位置，我们画出这些擦角线，找出它们与周界（perimeter）的交点，再把可见的长度加起来。所有工作都靠相似三角形（similar triangles）（或直线方程）完成。数值 $3a$ 的出现，是因为当 b 取这个值时，从墙中点出发的擦角线刚好到达庭院的一个角。

解答

把原点放在中心，于是周界墙是边在 $x = \pm\frac{b}{2}$、$y = \pm\frac{b}{2}$ 上的正方形，而建筑物是正方形 $|x| \le \frac{a}{2}$、$|y| \le \frac{a}{2}$，其中 $0 < a < b$。周界的总长度是 $4b$。守卫能看见墙上的点 Q，当且仅当由守卫到 Q 的线段不穿过建筑物。



墙的中点（$b > 3a$）



角落

墙上加粗（青色）的部分是守卫能看见的部分；虚线擦过建筑物的角。

守卫在墙的中点

设守卫在 $G = (0, -\frac{b}{2})$，即底墙的中点。整幅底墙都可见，因为建筑物严格位于它的上方：长度为 b。由于图形关于 y 轴对称（symmetry），我们只需看右边。

一条向右上方的视线不碰到建筑物，当且仅当它从最接近它的角 $K = (\frac{a}{2}, -\frac{a}{2})$ 的右边经过。（这条直线先遇到 $y = -\frac{a}{2}$，而且它的 x 坐标随高度上升而增加；所以如果它在高度 $-\frac{a}{2}$ 处位于建筑物右边，就会一直留在右边；如果它在该处位于 K 的左边，它便会进入建筑物，因为在高度 $-\frac{a}{2}$ 处它的 x 坐标介乎 0 与 $\frac{a}{2}$ 之间。）因此 G 所见范围的边界是直线 GK，它在水平距离 $\frac{a}{2}$ 内上升 $\frac{b-a}{2}$，所以斜率（gradient）是 $\frac{b-a}{a}$。

这条直线在以下高度与右墙 $x = \frac{b}{2}$ 相交：

$$y_1 = -\frac{b}{2} + \frac{b}{2} \cdot \frac{b-a}{a} = -\frac{b}{2} + \frac{b(b-a)}{2a}.$$

它到达该墙顶端（$y_1 \geq \frac{b}{2}$）当且仅当 $\frac{b-a}{a} \geq 2$，即 $b \geq 3a$。

> **想一想** 为甚么 b 与 $3a$ 的比较会在这里出现？

它决定擦角线 GK 撞上哪一幅墙：若 $b < 3a$，撞上右墙；若 $b > 3a$，则撞上远端（顶部）的墙。两种情况给出不同的可见长度公式。

*情况 $b < 3a$。*守卫看见右墙从底角至高度 y_1 的部分，长度为

$$y_1 + \frac{b}{2} = \frac{b(b-a)}{2a},$$

此外便看不见其他部分（任何连向该墙更高点或顶墙的直线，都从 K 的左边经过）。加上左边对应的一段，总可见长度为

$$M = b + 2 \cdot \frac{b(b-a)}{2a} = b + \frac{b(b-a)}{a} = \frac{b^2}{a}.$$

*情况 $b > 3a$。*这时整幅右墙（长度 b）都可见，而直线 GK 上升 b 后与顶墙 $y = \frac{b}{2}$ 相交，交点为

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

由于 $2a < b - a$，这个值小于 $\frac{b}{2}$。守卫看见顶墙由该点至角落的部分，长度为 $\frac{b}{2} - \frac{ab}{b-a} = \frac{b(b-3a)}{2(b-a)}$。把底墙、两幅侧墙和顶墙的两段加起来：

$$M = 3b + \frac{b(b-3a)}{b-a} = \frac{3b(b-a) + b(b-3a)}{b-a} = \frac{b(4b-6a)}{b-a}.$$

守卫在角落

设守卫在 $H = (-\frac{b}{2}, -\frac{b}{2})$。经过 H 的两幅墙完全可见：长度为 $2b$。图形关于对角线 $y = x$ 对称，所以只看右墙。按照与前面相同的论证，可见部分由经过 H 与角 $K = (\frac{a}{2}, -\frac{a}{2})$ 的

直线截断，其斜率为

$$\frac{\frac{b}{2}-\frac{a}{2}}{\frac{b}{2}+\frac{a}{2}}=\frac{b-a}{b+a}.$$

在到达墙 $x=\frac{b}{2}$ 的水平距离 b 内，它上升 $\frac{b(b-a)}{a+b}$，这个值小于 b，所以无论 a 和 b 取甚么值，这条直线总是撞上右墙。守卫看见右墙长度 $\frac{b(b-a)}{a+b}$ 的部分，顶墙亦看见同样长度，两幅墙的其余部分都看不见。因此

$$C=2b+\frac{2b(b-a)}{a+b}=\frac{2b(a+b)+2b(b-a)}{a+b}=\frac{4b^2}{a+b}.$$

比较两个位置

情况 $b<3a$。

$$C-M=\frac{4b^2}{a+b}-\frac{b^2}{a}=\frac{b^2\big(4a-(a+b)\big)}{a(a+b)}=\frac{b^2(3a-b)}{a(a+b)}>0,$$

所以**角落**较佳。

情况 $b>3a$。

$$M-C=\frac{b(4b-6a)}{b-a}-\frac{4b^2}{a+b}=\frac{b\big[(4b-6a)(a+b)-4b(b-a)\big]}{(b-a)(a+b)}=\frac{2ab(b-3a)}{(b-a)(a+b)}>0,$$

因为 $(4b-6a)(a+b)-4b(b-a)=4b^2-2ab-6a^2-4b^2+4ab=2ab-6a^2$。所以**墙的中点**较佳。

当 $b=3a$ 时，M 的两条公式都给出 $9a$，而 $C=\frac{36a^2}{4a}=9a$ 亦然，所以两个位置一样好。

验算

- 取 $a=1$、$b=2$：$M=4$，$C=\frac{16}{3}\approx 5.33$，所以角落胜出。取 $a=1$、$b=5$：$M=\frac{5\cdot 14}{4}=17.5$，$C=\frac{100}{6}\approx 16.67$，所以中点胜出。这些数值已由电脑验算证实，该验算测试了通往墙上数千个点的视线。
- 若 $b\to a$，建筑物填满整个庭院：$M\to a$（只有守卫所在的墙）而 $C\to 2a$（在角落相交的两幅墙），正如预期。
- 当 $b\to\infty$ 时，建筑物变得微不足道：$M=\frac{b(4b-6a)}{b-a}$ 与 $C=\frac{4b^2}{a+b}$ 都趋近整个周界 $4b$，而且都一直小于它。

常见错误

- 忘记当 $b>3a$ 时，在墙中点的守卫除了两幅侧墙，还能看见对面墙的一部分。

- 擦错建筑物的角：从下方看，起作用的是底部的两个角 $(\pm\frac{a}{2}, -\frac{a}{2})$。
- 总长度中漏掉守卫自己所在的墙（或两幅墙）。

STEP 之外

关于从某点能看见区域哪些部分的问题，属于计算几何（computational geometry）的范畴：经典的“美术馆定理”（art gallery theorem）指出，要监看一个有 n 幅墙的多边形展览馆的每一点，$\lfloor n/3 \rfloor$ 名守卫总是足够，而有时亦确实需要这么多。像这里那样求出可见区域（visibility region），是这类问题的基本步骤；在电脑图形学中亦然，隐藏面消除（hidden-surface removal）就是以同样方式运作。如想做更多这类风格的题目，Siklos 的 *Advanced Problems in Mathematics*（《数学高阶问题》）是很好的下一步。

详细解答

STEP I 2015，第 4 题

纯数学 · 参数曲线与面积

STEP I 2015，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/15-s1-q4

关键思路

以曲线的斜率（gradient）作为参数。若中点处的切线（tangent）与 x 轴成角 θ，则该处的斜率为 $\tan\theta$，由此便确定了中点的位置。棒的每一端都沿单位向量（unit vector）$(\cos\theta, \sin\theta)$ 方向与中点相距 b。至于面积，半条棒 AP 正好落在 P 点的切线上，所以所求区域其实就是抛物线（parabola）与它的一条切线之间的空隙，而对应的被积函数（integrand）是一个完全平方（perfect square）。

解答

一端的轨迹

设中点为 P，并设棒与 x 轴正方向成角 θ。由于 $y = \frac{1}{4}x^2$ 有 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{2}x$，而棒在 P 点与曲线相切，故

$$\tfrac{1}{2}x_P = \tan\theta, \qquad \text{故} \qquad P = \left(2\tan\theta,\ \tan^2\theta\right).$$

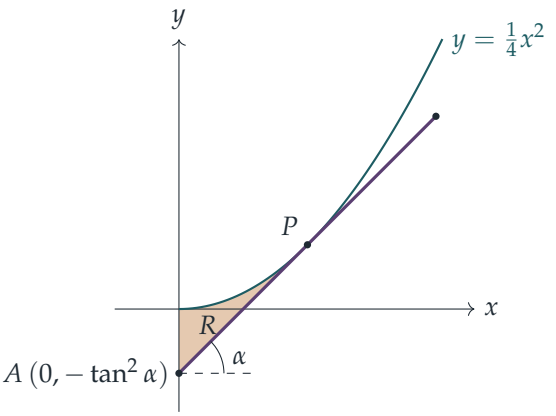
条件 $x \geq 0$ 给出 $\tan\theta \geq 0$，因此可取 $0 \leq \theta < \frac{1}{2}\pi$。当 θ 在这个范围内变动时，$2\tan\theta$ 取遍 $[0, \infty)$ 中的每一个值，所以中点的每一个位置都恰好对应一个 θ。

沿棒方向的单位向量是 $(\cos\theta, \sin\theta)$，而每一端与 P 相距 b，所以两端位于 $P \pm b(\cos\theta, \sin\theta)$。取负号（即较低、靠左的一端），得

$$x = 2\tan\theta - b\cos\theta,$$
$$y = \tan^2\theta - b\sin\theta,$$

正如所求。（另一端所描出的轨迹满足相同的方程，只是把 $-b$ 换成 $+b$。）

R 的面积



到达 y 轴的是满足 $x = 2\tan\theta - b\cos\theta$ 的那一端：另一端的 $x = 2\tan\theta + b\cos\theta > 0$ 恒成立。所以当 $\theta = \alpha$ 时，

$$2\tan\alpha = b\cos\alpha.$$

A 的 y 坐标是 $\tan^2\alpha - b\sin\alpha$，利用上一条方程可消去 b：

$$b\sin\alpha = (b\cos\alpha)\tan\alpha = 2\tan^2\alpha, \qquad \text{故} \qquad A = \big(0, -\tan^2\alpha\big).$$

想一想 A 相对于原点在甚么位置？直线 AP 的方程是甚么？

A 位于原点下方，而直线 AP 正是抛物线在 P 点的切线。记 $t = \tan\alpha$。在 $P = (2t, t^2)$ 的切线斜率为 t，所以其方程为

$$y = t^2 + t(x - 2t) = tx - t^2,$$

它确实经过 $A = (0, -t^2)$。抛物线位于这条切线上方，因为

$$\tfrac{1}{4}x^2 - (tx - t^2) = \big(\tfrac{1}{2}x - t\big)^2 \geq 0.$$

因此 R 是满足 $0 \leq x \leq 2t$、并介乎直线 AP（下方）与曲线（上方）之间的点的集合，其左边界是 y 轴上由 A 向上至 O 的线段。所以

$$\begin{aligned} \text{面积}R &= \int_0^{2t} \Big(\tfrac{1}{4}x^2 - (tx - t^2)\Big)\,\mathrm{d}x = \int_0^{2t} \big(\tfrac{1}{2}x - t\big)^2\,\mathrm{d}x \\ &= \Big[\tfrac{2}{3}\big(\tfrac{1}{2}x - t\big)^3\Big]_0^{2t} = 0 - \tfrac{2}{3}(-t)^3 = \tfrac{2}{3}t^3. \end{aligned}$$

所以 R 的面积是 $\frac{2}{3}\tan^3\alpha$，正如所求。

留意 b 完全消失了：面积只取决于 P 点的切线与 y 轴相交的位置。当然，b 与 α 之间有关系式 $2\sin\alpha = b\cos^2\alpha$，所以 b 决定了 α。

验算

- 取 $\alpha = \frac{1}{4}\pi$（如图所示）。由 $b\cos\alpha = 2$ 得 $b = 2\sqrt{2}$，此时 $P = (2,1)$，$A = (0,-1)$。面积为 $\int_0^2 \left(\frac{1}{4}x^2 - x + 1\right)\mathrm{d}x = \frac{2}{3} - 2 + 2 = \frac{2}{3} = \frac{2}{3}\tan^3\frac{1}{4}\pi$。
- 当 $b = 2$ 时，以数值方法解 $2\tan\alpha = 2\cos\alpha$ 得 $\alpha \approx 0.6662$，而用数值积分求得的面积是 0.32391，与 $\frac{2}{3}\tan^3\alpha$ 吻合。
- 当 $b \to 0$ 时，$\alpha \to 0$，区域缩小至消失，与公式的预测一致。

常见错误

- 在 x 轴处把区域分开，却忘记了 x 轴下方的三角形。在切线与曲线之间取单一个积分，便能一次过处理两部分。
- 用错了棒的一端：带 $+b\cos\theta$ 的那一端永远不会到达 y 轴。
- 答案中留下了 b，而没有利用 $b\cos\alpha = 2\tan\alpha$ 把它消去。

STEP 之外

中点描出抛物线，而每一端则描出一条沿切线方向保持固定距离 b 的曲线。由旧曲线的切线构造新曲线，是平面曲线微分几何（differential geometry）的一个常见主题（渐伸线（involute）、包络线（envelope）和曳物线（tractrix）都是经典例子），大学一年级课程中会与弧长（arc length）和曲率（curvature）一并学到。这个面积的简洁答案令人想起阿基米德（Archimedes）对抛物线的求积：他证明了抛物线弓形的面积是某个内接三角形面积的 $\frac{4}{3}$ 倍。如想做更多同类题目，可参考 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP I 2015，第 5 题

纯数学 • 含参数的积分

> **STEP I 2015，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/15-s1-q5

关键思路

两个积分在把 x 视为固定参数（parameter）时，对 t 都有初等的原函数（antiderivative），所以题目中“通过计算积分”这句指示已经是全部提示：先求出关于 x 的闭式（closed form），再作其草图。要小心的是细节。在 (i) 中，被积函数（integrand）在 $t = 1$ 处可能无界（unbounded）；在 (ii) 中，平方根计算出来是绝对值（*absolute values*），这使 g 成为一个分段函数（piecewise function），而其形状出奇地简单。

解答

第 (i) 部分

对固定的 $x > 0$，$(t-1)^{x-1}$ 的一个原函数是 $\frac{1}{x}(t-1)^x$。因此

$$f(x) = \left[\frac{(t-1)^x}{x}\right]_1^3 = \frac{2^x - 0}{x} = \frac{2^x}{x}.$$

当 $0 < x < 1$ 时，被积函数 $(t-1)^{x-1}$ 在 $t \to 1^+$ 时趋于无穷，所以这是一个反常积分（improper integral）；但由于 $x > 0$，当 $t \to 1^+$ 时 $(t-1)^x \to 0$，所以积分收敛（converges），上面的值仍然成立。

现在作 $y = 2^x/x$（$x > 0$）的草图。

- 当 $x \to 0^+$ 时，$2^x \to 1$，所以 $f(x) \to +\infty$：y 轴是一条渐近线（asymptote）。
- 当 $x \to \infty$ 时，指数函数的增长胜过幂函数，所以 $f(x) \to \infty$（而且比任何直线都快）。

- 求导，利用 $\frac{\mathrm{d}}{\mathrm{d}x}2^x = 2^x \ln 2$，

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

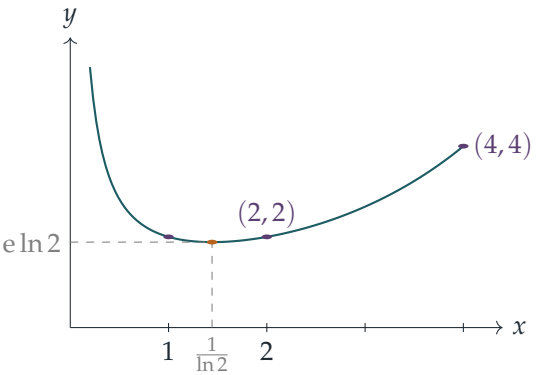
它在 $x < 1/\ln 2$ 时为负，在 $x > 1/\ln 2$ 时为正。所以只有一个驻点（stationary point），是一个极小点（minimum），位于 $x = 1/\ln 2 \approx 1.44$。

- 在该点，$2^{1/\ln 2} = \mathrm{e}^{(\ln 2)/\ln 2} = \mathrm{e}$，所以极小值为

$$f\left(\frac{1}{\ln 2}\right) = \mathrm{e} \ln 2 \approx 1.88.$$

- 有用的点：$f(1) = f(2) = 2$，$f(4) = 4$。

这条曲线也是凸（convex）的：$\ln f(x) = x \ln 2 - \ln x$ 是凸函数（其二阶导数为 $1/x^2 > 0$），而凸函数的指数仍是凸函数。所以草图是一条单一的 U 形曲线。



第 (ii) 部分

首先，被积函数是有意义的：对 $-1 \le t \le 1$，

$$1 - 2xt + x^2 \ge 1 - 2|x| + x^2 = (1 - |x|)^2 \ge 0,$$

等号只在 $|x| = 1$ 且 $t = \pm 1$（端点）时成立。这个情况我们稍后再处理。

当 $x = 0$ 时，被积函数为 1，所以 $g(0) = 2$。当 $x \ne 0$ 时，把 x 视为常数：由于 $\frac{\mathrm{d}}{\mathrm{d}t}(1 - 2xt + x^2) = -2x$，一个原函数是 $-\frac{1}{x}\sqrt{1 - 2xt + x^2}$。因此

$$g(x) = -\frac{1}{x}\left[\sqrt{1 - 2xt + x^2}\right]_{-1}^{1} = \frac{\sqrt{(1+x)^2} - \sqrt{(1-x)^2}}{x} = \frac{|1+x| - |1-x|}{x}.$$

想一想 为甚么不直接写 $\sqrt{(1+x)^2} = 1 + x$ 和 $\sqrt{(1-x)^2} = 1 - x$?

因为 $\sqrt{u^2} = |u|$，而不是 u。粗心地这样做会得出对所有 x 都有 $g(x) = 2$，这在 $|x|$ 很大时肯定是错的，因为那时被积函数约为 $1/|x|$，所以 $g(x) \approx 2/|x|$。我们必须分情况讨论。

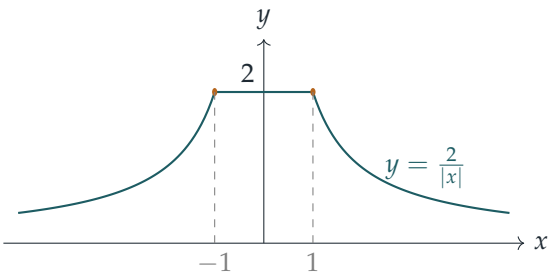
- $-1 \le x \le 1$，$x \ne 0$：$1+x$ 和 $1-x$ 都 ≥ 0，所以 $g(x) = \dfrac{(1+x)-(1-x)}{x} = 2$。这也与 $g(0) = 2$ 一致。
- $x > 1$：$|1+x| = 1+x$ 而 $|1-x| = x-1$，所以 $g(x) = \dfrac{2}{x}$。
- $x < -1$：$|1+x| = -1-x$ 而 $|1-x| = 1-x$，所以 $g(x) = \dfrac{-2}{x} = \dfrac{2}{|x|}$。

在 $x = \pm 1$ 时，被积函数在其中一个端点无界（当 $x = 1$ 时它是 $1/\sqrt{2-2t}$，在 $t \to 1$ 时趋于无穷），所以这是反常积分。原函数 $-\sqrt{2-2t}$ 直到 $t = 1$ 都是连续的，所以积分收敛，而公式仍然给出 $g(\pm 1) = 2$。

所以

$$g(x) = \begin{cases} 2, & |x| \le 1, \\ \dfrac{2}{|x|}, & |x| \ge 1. \end{cases}$$

这个函数是偶函数（even function）（这是必然的：把 t 换成 $-t$ 即可证明 $g(-x) = g(x)$）。它处处连续（continuous），在 $[-1, 1]$ 上为常数，并以 $2/|x|$ 的方式衰减，趋向渐近线 $y = 0$。在 $x = \pm 1$ 处有一个尖角：两侧的单侧斜率分别为 0（内侧）和 ∓ 2（外侧），所以 g 在该处不可微（differentiable）。



验算

- 第 (i) 部分取 $x = 1$：被积函数为 1，所以 $f(1) = 2$，与 $2^1/1$ 吻合。取 $x = 2$ 时积分为 $\int_1^3 (t-1)\,\mathrm{d}t = 2$，与 $2^2/2$ 吻合。
- 第 (ii) 部分取 $x = 0$：被积函数为 1，而 $g(0) = 2$。当 $|x|$ 很大时，被积函数在长度为 2 的区间上接近 $1/|x|$，与 $2/|x|$ 吻合。
- 数值上，$\int_1^3 (t-1)^{0.5}\,\mathrm{d}t = 1.8856\ldots = 2^{1.5}/1.5$；(ii) 中的积分在 $x = 1.2$ 时算得 $1.6667 = 2/1.2$，在 $x = 0.8$ 时算得 2.0000。

常见错误

- 对所有 x 都写 $\sqrt{(1-x)^2} = 1-x$，从而得出 g 是常数的结论。
- 忘记 (i) 中的积分在 $0 < x < 1$ 时是反常积分，或忘记 $(t-1)^x$ 在 $t = 1$ 处为零只是因为 $x > 0$。
- 把 g 在 $x = \pm 1$ 处画成光滑的：两段曲线以不同的斜率相接。

STEP 之外

第 (ii) 部分其实是乔装了的牛顿（Newton）壳层定理（shell theorem）。设 $t = \cos\theta$，则 $1/\sqrt{1-2xt+x^2}$ 是从距单位球面球心 x 的一点到球面上一点的距离的倒数，所以 g（相差一个常数）就是均匀球壳的引力势（gravitational potential）：在壳内为常数，在壳外则等同一个点质量的引力势。同一个式子也是勒让德多项式（Legendre polynomials）的生成函数（generating function）$\sum_n P_n(t)x^n$，而 $|x| < 1$ 时 $g(x) = 2$ 反映了一个事实：每个 $n \geq 1$ 的 P_n 在 $[-1,1]$ 上的积分都是零。第 (i) 部分把 x 当作参数来积分 t^{x-1}，正是通往伽玛函数（Gamma function）和贝塔函数（Beta function）的第一步。Mary Boas 的 *Mathematical Methods in the Physical Sciences*（《物理科学中的数学方法》）以恰到好处的程度介绍了勒让德多项式和伽玛函数。

详细解答

STEP I 2015，第 6 题

纯数学 · 几何中的向量

STEP I 2015，第 6 题

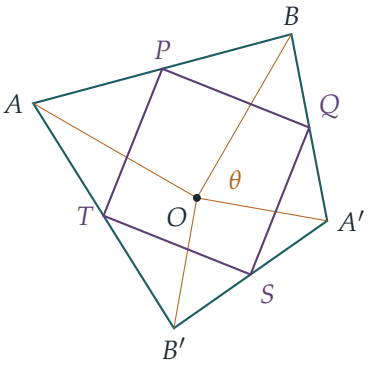
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/15-s1-q6

关键思路

以 O 为原点，条件 $OA = OB$ 及 $\angle AOB = 90°$ 表示 B 是 A 绕 O 旋转（rotation）一个直角所得的点；同样地，B' 是 A' 以相同方向旋转所得的点。把这个四分之一转记作 R。于是中点四边形的每一条边，原来都是前一条边经 R 作用后的结果，而这正正就是构成正方形的条件。一旦以 $\mathbf{a}$ 和 $\mathbf{a}'$ 表示出边长，面积就只是一道余弦定理（cosine rule）的计算。

解答



中点构成正方形

取 O 为原点，并以 $\mathbf{a}, \mathbf{b}, \mathbf{a}', \mathbf{b}'$ 表示 A, B, A', B' 的位置向量（position vector）。设 R 为绕 O 顺时针旋转 90°。它是线性（linear）的，保持长度不变，并把每个非零向量转过一个

直角。

由于顶点按顺时针方向排列，而 O 位于四边形内部，从 OA 到 OB 是顺时针转 $90°$。又因 $OA = OB$，得 $\mathbf{b} = R\mathbf{a}$。同理 $\mathbf{b}' = R\mathbf{a}'$。

设 P, Q, S, T 分别为 AB、BA'、$A'B'$、$B'A$ 的中点：

$$\mathbf{p} = \tfrac{1}{2}(\mathbf{a}+\mathbf{b}), \quad \mathbf{q} = \tfrac{1}{2}(\mathbf{b}+\mathbf{a}'), \quad \mathbf{s} = \tfrac{1}{2}(\mathbf{a}'+\mathbf{b}'), \quad \mathbf{t} = \tfrac{1}{2}(\mathbf{b}'+\mathbf{a}).$$

现在逐一考虑四条边。设 $\mathbf{u} = \overrightarrow{PQ}$：

$$\begin{aligned}
\overrightarrow{PQ} &= \tfrac{1}{2}(\mathbf{a}'-\mathbf{a}) = \mathbf{u},\\
\overrightarrow{QS} &= \tfrac{1}{2}(\mathbf{b}'-\mathbf{b}) = \tfrac{1}{2}R(\mathbf{a}'-\mathbf{a}) = R\mathbf{u},\\
\overrightarrow{ST} &= \tfrac{1}{2}(\mathbf{a}-\mathbf{a}') = -\mathbf{u},\\
\overrightarrow{TP} &= \tfrac{1}{2}(\mathbf{b}-\mathbf{b}') = -R\mathbf{u}.
\end{aligned}$$

这里 $\mathbf{u} \neq \mathbf{0}$，因为 $A \neq A'$。所以 PQ 与 ST 大小相等、方向相反，故 $PQST$ 是平行四边形（parallelogram）。它的邻边 QS 是把 $\mathbf{u}$ 转过一个直角而长度不变所得。邻边相等且互相垂直的平行四边形是正方形，因此 $PQST$ 是正方形，得证。

想一想 为甚么把 A 转到 B 的四分之一转，必定与把 A' 转到 B' 的是同一个？

两对点在顺时针标记中都是前后相连的，而 O 位于内部，所以两种情况下由第一条射线转到第二条射线都是顺时针方向。若其中一个是顺时针而另一个是逆时针，由 $\overrightarrow{QS}$ 得出 $R\mathbf{u}$ 的那一步便不成立，而一般来说所得图形也不会是正方形。

求面积的最大值

设 $OA = OB = p$ 及 $OA' = OB' = q$，两者皆固定，并设 $\theta = \angle BOA'$。O 处的四个角之和为 $360°$，其中两个是直角，所以 $\angle B'OA = 180° - \theta$。由于 O 位于内部，θ 和 $180° - \theta$ 都严格介乎 $0°$ 与 $180°$ 之间，故 $0 < \theta < 180°$。

正方形的边长为 $|\mathbf{u}| = \frac{1}{2}AA'$，所以面积为 $\frac{1}{4}AA'^2$。由 OA 经 OB 转到 OA'，$\mathbf{a}$ 与 $\mathbf{a}'$ 之间的夹角为 $90° + \theta$，而 $\cos(90° + \theta) = -\sin\theta$。由余弦定理（或展开 $|\mathbf{a}' - \mathbf{a}|^2$），

$$AA'^2 = p^2 + q^2 - 2pq\cos(90° + \theta) = p^2 + q^2 + 2pq\sin\theta,$$

所以

$$\text{面积} = \tfrac{1}{4}\left(p^2 + q^2 + 2pq\sin\theta\right).$$

在 p 和 q 固定的情况下，当 $\sin\theta$ 最大时此式最大。在 $0 < \theta < 180°$ 上，这恰好发生于

$$\angle BOA' = \theta = 90°,$$

而最大面积为 $\frac{1}{4}(p+q)^2$。此时 O 处的四个角全是直角，所以 A' 位于 AO 越过 O 的延长线上，而正方形的边长为 $\frac{1}{2}(p+q)$。

验算

- 取 $p = 2$、$q = 3$。面积公式给出 $\frac{1}{4}(13 + 12\sin\theta)$，在 $\theta = 0$ 附近为 3.25，在 $\theta = 90°$ 时升至 $6.25 = \frac{1}{4} \cdot 5^2$，然后回落至接近 3.25。以电脑对 θ 作精细扫描，得出峰值位于 90.0°。
- 对称情况 $p = q$、$\theta = 90°$：$ABA'B'$ 本身就是一个以 O 为中心、边长为 $p\sqrt{2}$ 的正方形，而其中点正方形的边长为 p，面积为 $p^2 = \frac{1}{4}(2p)^2$，与公式相符。
- 随机选取 p、q 和 θ，经电脑验证，各中点确实构成面积为 $\frac{1}{4}(p^2 + q^2 + 2pq\sin\theta)$ 的正方形。

常见错误

- 把 OA 与 OA' 之间的夹角当作 θ，而不是 $90° + \theta$。这会得出面积 $\frac{1}{4}(p^2 + q^2 - 2pq\cos\theta)$，并错误地推断 θ 应趋近 180°。
- 只证明四条边相等。这只能得出菱形（rhombus）。你还需要一个直角，而这里由 $\overrightarrow{QS} = R\,\overrightarrow{PQ}$ 便可直接得到。

STEP 之外

旋转一个直角是一个线性映射（linear map）$R(x, y) = (y, -x)$，在复平面（complex plane）上则是乘以 $-\mathrm{i}$。整个论证其实是一道关于线性映射的计算：$\mathbf{q} - \mathbf{p}$ 与 $\mathbf{s} - \mathbf{q}$ 相差一个 R，因为 R 与减法可交换（commute）。同样的思路可以迅速证明拿破仑定理（Napoleon's theorem）和 van Aubel 定理（van Aubel's theorem）等结果。Needham 的 *Visual Complex Analysis*（《直观复分析》）正是以这种方式，通过复数乘法来发展几何。

详细解答

STEP I 2015，第 7 题

纯数学 • 区间上的最大值

> **STEP I 2015，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/15-s1-q7

关键思路

光滑函数（smooth function）在闭区间（closed interval）上的最大值，必定在区间内部的驻点（stationary point）或在端点（endpoint）取得。这里的驻点是 $x = 0$ 和 $x = \frac{a}{3}$，而后者会随着 a 增大向右移动，最终离开区间。因此我们先列出所有候选值，再随着 a 变化逐一比较：三个范围正是来自比较结果转换之处，以及 $\frac{a}{3}$ 越过 $x = 1$ 之处。

解答

候选值

求导，得

$$f'(x) = 6ax - 18x^2 = 6x(a - 3x),$$

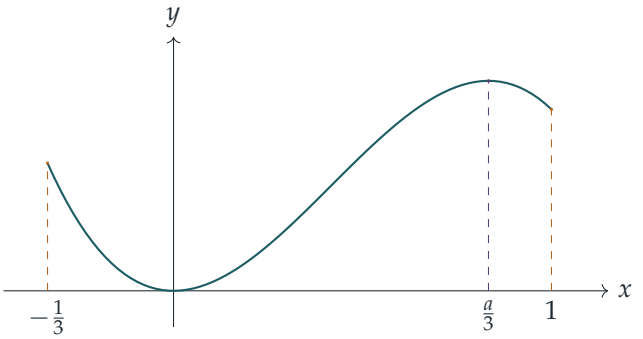
所以驻点是 $x = 0$ 和 $x = \frac{a}{3}$。当 $a \geq 0$ 时，f' 的正负号告诉我们函数的形状：

- 当 $x < 0$ 时，$6x < 0$ 而 $a - 3x > 0$，所以 $f'(x) < 0$，f 递减；
- 当 $0 < x < \frac{a}{3}$ 时，$f'(x) > 0$，f 递增；
- 当 $x > \frac{a}{3}$ 时，$f'(x) < 0$，f 递减。

所以 $x = 0$ 是局部极小值（local minimum）点，$x = \frac{a}{3}$ 是局部极大值（local maximum）点（当 $a = 0$ 时两者重合，而 $f = -6x^3$ 处处递减）。我们需要的函数值是

$$f\left(-\tfrac{1}{3}\right) = \frac{a}{3} + \frac{2}{9}, \qquad f\left(\tfrac{a}{3}\right) = \frac{a^3}{3} - \frac{2a^3}{9} = \frac{a^3}{9}, \qquad f(1) = 3a - 6.$$

由于 f 在 $[-\frac{1}{3}, 0]$ 上递减，这一段上的最大值是 $f(-\frac{1}{3})$；而 $f(0) = 0$ 永远不会是最大值。所以 $\mathrm{M}(a)$ 就是 $f(-\frac{1}{3})$ 与 f 在 $[0, 1]$ 上最大值两者中较大的一个。



上图显示 $a = \frac{5}{2}$ 的情况：三个候选值是两个端点（琥珀色）和局部极大值（紫红色），在这里胜出的是局部极大值。

情况 $0 \le a \le 3$

此时 $\frac{a}{3} \in [0,1]$，所以在 $[0,1]$ 上函数先上升至 $x = \frac{a}{3}$，然后下降。它在这一段上的最大值是 $f(\frac{a}{3}) = \frac{a^3}{9}$，我们要把它与 $f(-\frac{1}{3})$ 比较：

$$\frac{a^3}{9} - \left(\frac{a}{3} + \frac{2}{9}\right) = \frac{a^3 - 3a - 2}{9} = \frac{(a+1)^2(a-2)}{9}.$$

? **想一想** 怎样看出 $a^3 - 3a - 2$ 的因式分解（factorisation）?

试代入一些小整数：$a = -1$ 给出 $-1 + 3 - 2 = 0$，$a = 2$ 给出 $8 - 6 - 2 = 0$。三个根之和是 0（因为没有 a^2 项），所以第三个根又是 -1，于是得到 $(a+1)^2(a-2)$。

由于当 $a \ge 0$ 时 $(a+1)^2 > 0$，差的正负号与 $a - 2$ 相同。因此

$$\mathrm{M}(a) = \frac{a}{3} + \frac{2}{9} \quad \text{当} 0 \le a \le 2, \qquad \mathrm{M}(a) = \frac{a^3}{9} \quad \text{当} 2 \le a \le 3,$$

而两条公式在 $a = 2$ 时都给出 $\frac{8}{9}$。

情况 $a > 3$

此时 $\frac{a}{3} > 1$，所以 f 在整个 $[0,1]$ 上递增，它在这一段上的最大值是 $f(1) = 3a - 6$。与左端点比较，

$$(3a - 6) - \left(\frac{a}{3} + \frac{2}{9}\right) = \frac{8a}{3} - \frac{56}{9} = \frac{8}{9}(3a - 7),$$

由于 $a > 3 > \frac{7}{3}$，这个值是正数。所以当 $a \ge 3$ 时 $\mathrm{M}(a) = 3a - 6$；在 $a = 3$ 时它等于 $3 = \frac{27}{9}$，与中间那条公式吻合。

结论

$$\mathrm{M}(a)=\begin{cases}\dfrac{a}{3}+\dfrac{2}{9}, & 0\le a\le 2,\\[2mm] \dfrac{a^3}{9}, & 2\le a\le 3,\\[2mm] 3a-6, & a\ge 3.\end{cases}$$

换句话说：a 较小时左端点胜出，a 适中时局部极大值胜出，而一旦局部极大值离开了区间，就由右端点胜出。

验算

- M 是连续（continuous）的：在 $a=2$ 时两边都是 $\frac{8}{9}$，在 $a=3$ 时两边都是 3。这是必然的，因为最大值随 a 连续变化。
- $a=0$：$f(x)=-6x^3$ 递减，所以其最大值是 $f(-\frac{1}{3})=\frac{2}{9}$，与第一条公式一致。
- $a=\frac{5}{2}$：公式给出 $\frac{125}{72}\approx 1.736$，而两个端点分别给出 $f(-\frac{1}{3})\approx 1.056$ 和 $f(1)=1.5$，两者都较小。以电脑在精细的 x 网格上搜索，对 0 至 12 之间数百个 a 值进行验算，结果与公式的误差在 10^{-8} 以内。

常见错误

- 忘记了左端点 $x=-\frac{1}{3}$。由于 f 在 $x<0$ 时递减，这个端点有可能胜过局部极大值，而当 $a<2$ 时确实如此。
- 在 $a>3$ 时仍使用 $f(\frac{a}{3})$，但此时 $\frac{a}{3}$ 已在区间之外，因此不能采用。
- 以为局部极大值自然就是区间上的最大值。

STEP 之外

这里的 $\mathrm{M}(a)=\max_{x\in[-1/3,1]} f(x)$ 是通过对一族函数求最优值而定义的函数，这类函数通常是连续的，但只是分段光滑（piecewise smooth）：你可以验证 M 在 $a=2$ 处的左导数与右导数并不相等。至于闭有界区间上的连续函数必定能取得最大值，这就是极值定理（Extreme Value Theorem），会在第一门分析课中证明；参见 Abbott，*Understanding Analysis*（《理解分析》）。

详细解答

STEP I 2015，第 8 题

纯数学 · 幂和的整除性

> **STEP I 2015，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/15-s1-q8

关键思路

把和式写两次，一次顺序、一次倒序，然后相加：这是第 (i) 部分所用的高斯（Gauss）技巧，而当各项是 k 次幂且 k 为奇数时，第 (ii) 部分正是为此而设的。把 r^k 与 $(n-r)^k$ 配对，可证 $2S$ 是 n 的倍数；把 r^k 与 $(n+1-r)^k$ 配对，可证 $2S$ 也是 $n+1$ 的倍数。由于 n 与 $n+1$ 没有公因数，两个结果合起来便得出能被 $\frac{1}{2}n(n+1)$ 整除（divisible）。

解答

第 (i) 部分

设 $T = 1+2+\cdots+n$。把和式倒序写出，

$$T = 1+2+\cdots+(n-1)+n,$$
$$T = n+(n-1)+\cdots+2+1.$$

把各列相加，n 列中每一列之和都是 $n+1$，所以 $2T = n(n+1)$，即

$$1+2+\cdots+n = \tfrac{1}{2}n(n+1).$$

第 (ii) 部分

在模 N 下考虑，有 $N-m \equiv -m$，所以

$$(N-m)^k + m^k \equiv (-m)^k + m^k = -m^k + m^k = 0 \pmod{N},$$

其中因为 k 是奇数，所以 $(-m)^k = -m^k$。因此 N 整除 $(N-m)^k + m^k$。

如果你想避免使用同余（congruence）：当 k 为奇数时，有以下因式分解（factorisation）

$$a^k + b^k = (a+b)\left(a^{k-1} - a^{k-2}b + a^{k-3}b^2 - \cdots + b^{k-1}\right),$$

（符号正负交替，括号内共有 k 项，所以最后一个符号是 $+$）。取 $a = N - m$、$b = m$，便得 $a + b = N$ 乘以一个整数。这里完全不需要 $m \le N$：若 $m > N$，则 $N - m$ 是负数，但它仍然是整数。

被 n 或 $\frac{1}{2}n$ 整除

设 $S = 1^k + 2^k + \cdots + n^k$，其中 k 为奇数。我们仿照第 (i) 部分的技巧。由于 $0^k = 0$，可以写成

$$S = \sum_{r=0}^{n} r^k = \sum_{r=0}^{n} (n-r)^k,$$

第二种写法就是同一个和式倒序排列。相加得

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

由第 (ii) 部分（取 $N = n$、$m = r$），每个方括号都能被 n 整除，所以 n 整除 $2S$。设 $2S = nM$，其中 M 为整数。

- 若 n 为奇数，则 n 整除 $2S$ 且与 2 没有公因数，所以 n 整除 S。
- 若 n 为偶数，则 $S = \frac{1}{2}n \cdot M$，而 $\frac{1}{2}n$ 是整数，所以 $\frac{1}{2}n$ 整除 S。

> ? **想一想** 当 n 为偶数时，为甚么不能直接说“n 整除 S”?

因为由“$n \mid 2S$”推出“$n \mid S$”这一步，需要 n 与 2 互质（coprime）。事实上这一步可能不成立：取 $n = 2$、$k = 1$，得 $S = 3$，它不是 2 的倍数。对于偶数 n，我们只能说 n 的一半整除 S。

被 $1 + 2 + \cdots + n$ 整除

由第 (i) 部分，我们需要证明 $\frac{1}{2}n(n+1) \mid S$。换一种方式配对：把 $1, 2, \ldots, n$ 倒序排列，会把 r 变成 $n + 1 - r$，所以

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

而由第 (ii) 部分（取 $N = n + 1$、$m = r$），每个方括号都能被 $n + 1$ 整除。所以 $n + 1$ 整除 $2S$。

n 与 $n + 1$ 是相邻整数，所以除了 1 以外没有其他公因数；于是 n 的任何因数与 $n + 1$ 的任何因数也是如此。我们用到以下事实：若两个互质的整数各自整除 S，则它们的积也整除 S。

- n 为奇数。我们已证明 $n \mid S$。此外 $n+1$ 是偶数且 $n+1 \mid 2S$，所以写成 $2S = (n+1)L$，得 $S = \frac{1}{2}(n+1)L$，即 $\frac{1}{2}(n+1) \mid S$。由于 n 与 $\frac{1}{2}(n+1)$ 互质，它们的积 $\frac{1}{2}n(n+1)$ 整除 S。
- n 为偶数。我们已证明 $\frac{1}{2}n \mid S$。此外 $n+1$ 是奇数且整除 $2S$，所以（由于它与 2 互质）它整除 S。由于 $\frac{1}{2}n$ 与 $n+1$ 互质，它们的积 $\frac{1}{2}n(n+1)$ 整除 S。

两种情况下，S 都能被 $\frac{1}{2}n(n+1) = 1+2+\cdots+n$ 整除，证毕。

> ? **想一想** 为甚么可以把两个整除结果合而为一，把除数相乘？

若 $a \mid S$、$b \mid S$ 且 $\gcd(a,b) = 1$，写 $S = aq$。那么 $b \mid aq$，而 b 与 a 没有公共质因数，所以整除 b 的每个质数幂都必定整除 q；因此 $b \mid q$，从而 $ab \mid S$。没有互质这个条件，结论便不成立：$4 \mid 12$ 且 $6 \mid 12$，但 $24 \nmid 12$。正因如此，我们才小心地把 n 与 $\frac{1}{2}(n+1)$ 配对，或把 $\frac{1}{2}n$ 与 $n+1$ 配对，而不是在 n 为偶数时把 n 与 $n+1$ 配对。

验算

试取 $k = 3$、$n = 4$：$S = 1+8+27+64 = 100$，而 $1+2+3+4 = 10$ 整除 100。事实上 $1^3+\cdots+n^3 = \left(\frac{1}{2}n(n+1)\right)^2$，所以 $k = 3$ 时结果是显然的。取 $k = 5$、$n = 5$：$S = 1+32+243+1024+3125 = 4425 = 15 \times 295$。最终结果亦已用电脑验证，对所有 $n \le 119$ 及所有奇数 $k \le 15$ 均成立。当 k 为偶数时，命题不成立：$1^2+2^2 = 5$ 不能被 3 整除，这说明 k 为奇数这个条件确实用上了。

常见错误

- 在 n 为偶数时，由 $n \mid 2S$ 推出 $n \mid S$。
- 把不互质的除数相乘，例如声称 $n(n+1) \mid 2S$，然后不加小心地除以 2。
- 把 r^k 与 $(n-r)^k$ 配对时忘记 $r = 0$（或 $r = n$）这一项：把 $0^k = 0$ 包括在内，每一项便能整齐地配对。

STEP 之外

幂和（power sum）$1^k+\cdots+n^k$ 是 n 的 $k+1$ 次多项式，可由 Faulhaber 公式以伯努利数（Bernoulli numbers）表示；当 k 为奇数时，它们甚至可以写成 $\frac{1}{2}n(n+1)$ 的多项式，这从另一角度解释了本题的结果。这里的论证让你初尝运用同余与互质的滋味，而这些正是初等数论（elementary number theory）的基本功。Davenport 的 *The Higher Arithmetic*

(《高等算术》) 是一本很出色的入门书。

详细解答

STEP I 2015，第 9 题

力学 • 抛体运动

STEP I 2015，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/15-s1-q9

关键思路

“最后落地”关乎的是*落地时间*（landing time），而不是发射时间：较迟以较低仰角发射的子弹，在空中停留的时间较短，因此可能比较早发射的子弹先落地。把落地时间写成发射时间加上飞行时间（time of flight），再在整段发射期间内求其最大值。以角度 α 代替 t 作为变量，微积分会简洁得多；而 k 所满足的条件，决定了最大值出现在驻点（stationary point）还是端点。

解答

建立模型

把每颗子弹视为只受重力作用、没有空气阻力、从地面发射的质点。以仰角 α 发射的子弹，在发射后时间 s 的竖直位置为 $y = u\sin\alpha\, s - \frac{1}{2}gs^2$，所以它在经过

$$\frac{2u\sin\alpha}{g},$$

之后回到地面，而在这段时间内它的水平位移为

$$R(\alpha) = u\cos\alpha \cdot \frac{2u\sin\alpha}{g} = \frac{2u^2\sin\alpha\cos\alpha}{g}.$$

仰角为 α 的子弹在时间 $t = (\frac{1}{3}\pi - \alpha)/\lambda$ 发射；当 t 取遍 $[0, \frac{\pi}{6\lambda}]$ 时，α 取遍 $[\frac{1}{6}\pi, \frac{1}{3}\pi]$，每个值恰好取一次。因此这颗子弹的落地时间为

$$L(\alpha) = \frac{\frac{1}{3}\pi - \alpha}{\lambda} + \frac{2u\sin\alpha}{g}, \qquad \frac{1}{6}\pi \le \alpha \le \frac{1}{3}\pi.$$

最后落地的子弹，就是使 L 取最大值的那一颗。由于 $k=\dfrac{g}{2\lambda u}$，我们有 $\dfrac{2u}{g}=\dfrac{1}{\lambda k}$，以及

$$L(\alpha)=\frac{1}{\lambda}\left(\tfrac{1}{3}\pi-\alpha+\frac{\sin\alpha}{k}\right),\qquad L'(\alpha)=\frac{1}{\lambda}\left(\frac{\cos\alpha}{k}-1\right)=\frac{\cos\alpha-k}{\lambda k}.$$

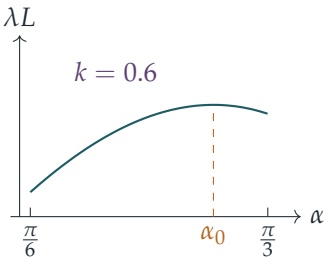
所以当 $\cos\alpha>k$ 时 L 递增，当 $\cos\alpha<k$ 时递减。在我们的区间上，$\cos\alpha$ 由 $\frac{1}{2}\sqrt{3}$（在 $\alpha=\frac{1}{6}\pi$ 处）递减至 $\frac{1}{2}$（在 $\alpha=\frac{1}{3}\pi$ 处）。

想一想 为甚么 α 越大，代表发射时间越早？这一点为何会影响 L' 的正负号？

枪管正在下降，所以仰角较大的子弹先发射。把 α 稍为增大，发射时间每弧度提早 $1/\lambda$，但飞行时间每弧度增加 $\frac{2u}{g}\cos\alpha$。$L'(\alpha)$ 正是这两种效应的此消彼长，其正负号告诉我们哪一方占优。

情况 $\frac{1}{2}\le k\le\frac{1}{2}\sqrt{3}$

此时 k 落在 $\cos\alpha$ 的值域之内，所以恰好有一个 $\alpha_0\in[\frac{1}{6}\pi,\frac{1}{3}\pi]$ 满足 $\cos\alpha_0=k$。由于 cos 在该区间上递减，当 $\alpha<\alpha_0$ 时 $\cos\alpha>k$，当 $\alpha>\alpha_0$ 时 $\cos\alpha<k$。所以 L 在 α_0 之前递增、之后递减，以仰角 α_0 发射的子弹最后落地。（若 $k=\frac{1}{2}$ 或 $k=\frac{1}{2}\sqrt{3}$，则 α_0 是端点，而 L 是单调的，但结论不变。）



（纵轴经过平移，好让落地时间的变化清楚可见。）

利用 $\sin\alpha_0=\sqrt{1-k^2}$（取正值，因为 α_0 是锐角），这颗子弹的射程为

$$R(\alpha_0)=\frac{2u^2\sin\alpha_0\cos\alpha_0}{g}=\frac{2ku^2\sqrt{1-k^2}}{g},$$

这就是所求的距离。

情况 $k<\frac{1}{2}$

此时对区间内每个 α 都有 $\cos\alpha\ge\frac{1}{2}>k$，所以 $L'(\alpha)>0$ 处处成立，L 严格递增。最大值出现在 $\alpha=\frac{1}{3}\pi$：第一颗发射的子弹最后落地。它的射程为

$$R\left(\tfrac{1}{3}\pi\right)=\frac{2u^2}{g}\cdot\frac{\sqrt{3}}{2}\cdot\frac{1}{2}=\frac{\sqrt{3}\,u^2}{2g}.$$

所以当 $k < \frac{1}{2}$ 时，最后一颗落地的子弹落在距离枪 $\dfrac{\sqrt{3}\,u^2}{2g}$ 之处。

（为求完整：若 $k > \frac{1}{2}\sqrt{3}$，则 L 递减，最后发射的子弹最后落地；由于 $\sin\frac{\pi}{3} = \sin\frac{2\pi}{3}$，距离同样是 $\frac{\sqrt{3}\,u^2}{2g}$。）

验算

- 连续性：在 $k = \frac{1}{2}$ 时，第一条公式给出 $\dfrac{2\cdot\frac{1}{2}u^2\cdot\frac{\sqrt{3}}{2}}{g} = \dfrac{\sqrt{3}\,u^2}{2g}$，与 $k < \frac{1}{2}$ 的答案一致；在 $k = \frac{1}{2}\sqrt{3}$ 时，它亦给出相同的值。
- 取 $u^2/g = 1$ 及 $k = 0.6$，最后落地的子弹满足 $\cos\alpha_0 = 0.6$，落在距离 $2(0.6)(0.8) = 0.96$ 之处。以电脑抽取两万个发射时间，并以数值方法逐步模拟该子弹的运动，结果吻合。
- 单位：k 是无量纲的（g 除以 λu，后者是速度每单位时间），而 u^2/g 是长度。

常见错误

- 只求射程或飞行时间的最大值，而不是落地时间 $t + \frac{2u\sin\alpha}{g}$ 的最大值。
- 想当然地以为最后发射的子弹就是最后落地的。
- 忘记 t 随 α 增大而减少，以致弄错 L' 的正负号。

STEP 之外

在闭区间上求函数的最大值，就是把区间内部的驻点与端点互相比较；而在这里，由一个参数决定哪一方胜出。这种“哪个约束生效”的切换，正是约束最优化（constrained optimisation）的起点，你会在大学一年级的数学方法课程中遇到。如想阅读更多这类风格的力学内容，可参考 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP I 2015，第 10 题

力学 · 相对速度

> **STEP I 2015，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/15-s1-q10

关键思路

雨点打在一个表面上，是一种通量（flux）：单位时间内打中的雨量 =（单位体积内的雨点数目）×（面积）×（雨的速度垂直于该表面的分量）。巴士行驶时，改在巴士的参考系（frame of reference）中考虑，这样只有雨速的水平分量会改变。然后，就整段行程而言，把速率乘以行程时间 D/u，再对 u 求最小值。绝对值（modulus）把问题分成两种情况，视乎巴士比雨的水平运动慢还是快。

解答

雨打中巴士的速率

把雨模拟成均匀分布于空气中的雨点，每单位体积有 n 点，全部以相同速度运动。设巴士的阔度为 b，并取北方为水平正方向。雨从南方吹来，即向北移动，所以其速度为

$$\text{水平分量} v\cos\theta\text{（向北）}, \qquad \text{铅直分量} v\sin\theta\text{（向下）}.$$

雨没有东西方向的分量，所以巴士两条长边的侧面保持干爽。

$u = 0$ 的情况。在时间 δt 内，打中车顶的雨点，就是位于一个以车顶为底、铅直高度为 $v\sin\theta\,\delta t$ 的斜棱柱（prism）之内的雨点，其体积为 $ab\,v\sin\theta\,\delta t$。打中车尾（南面，面积 hb）的雨点，则填满车尾后方一个水平深度为 $v\cos\theta\,\delta t$ 的棱柱。面向北方的车头则被遮挡。因此速率为

$$nb\left(h\,v\cos\theta + a\,v\sin\theta\right).$$

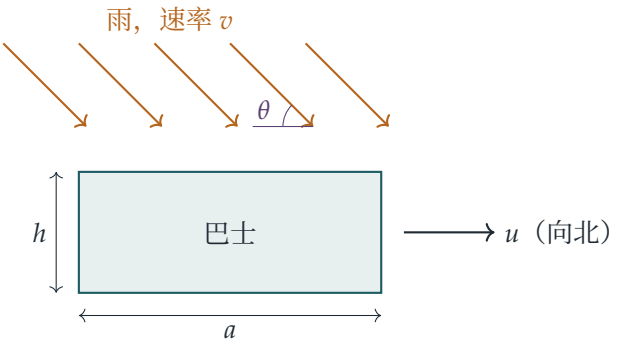
$u > 0$ 的情况。在以等速运动的巴士参考系中，雨的速度为

$$(v\cos\theta - u)\text{ 向北}, \qquad v\sin\theta\text{ 向下}.$$

巴士在这个参考系中静止，所以上述论证同样适用，只需把 $v\cos\theta$ 换成 $v\cos\theta - u$。铅直分量不变，所以车顶收集雨水的速率仍是 $nab\,v\sin\theta$。若 $v\cos\theta > u$，雨仍会追过巴士，以速率 $nhb\,(v\cos\theta - u)$ 打中车尾。若 $v\cos\theta < u$，则巴士迎头撞向雨，雨改为以速率 $nhb\,(u - v\cos\theta)$ 打中车头。无论哪种情况，总速率均为

$$nb\Big(h\,|v\cos\theta - u| + a\,v\sin\theta\Big),$$

与 $h|v\cos\theta - u| + av\sin\theta$ 成正比，正如所求。



选择车速

设行程长度为 D，所需时间为 D/u，所以收集到的总雨量与下式成正比：

$$R(u) = \frac{D}{u}\Big(h|v\cos\theta - u| + av\sin\theta\Big), \qquad 0 < u \le w.$$

记 $c = v\cos\theta$，并去掉绝对值符号。

若 $u \le c$：

$$R(u) = D\left(\frac{hv\cos\theta + av\sin\theta}{u} - h\right).$$

分子为正，所以 R 在 $0 < u \le c$ 上严格递减（strictly decreasing）。

若 $u \ge c$：

$$R(u) = D\left(h + \frac{v\,(a\sin\theta - h\cos\theta)}{u}\right).$$

它是递增还是递减，只取决于 $a\sin\theta - h\cos\theta$ 的正负号。

想一想 为甚么开得越快不一定越好?

开快一点会缩短行程，这对车顶总是有利：车顶所占的份额是 $Dav\sin\theta/u$。可是一旦 $u > c$，巴士便迎头撞向雨，而车头所占的份额 $Dh(u - c)/u$ 会随 u 增大。$a\sin\theta - h\cos\theta$ 的正负号告诉我们哪个效应占上风。

两条公式在 $u = c$ 时相等，所以 R 是连续（continuous）的。现在：

- 若 $w < v\cos\theta$，则所有容许的车速都满足 $u \le c$，而 R 在此递减，所以最佳选择是 $u = w$。
- 若 $a\sin\theta > h\cos\theta$，则 R 在 $u \le c$ 上递减，在 $u \ge c$ 上亦递减。所以 R 在整个 $(0, w]$ 上递减，最佳选择同样是 $u = w$。

这就证明了第一个结论。

若 $w > v\cos\theta$ 且 $a\sin\theta < h\cos\theta$：$R$ 在 $(0, c]$ 上递减，在 $[c, w]$ 上严格递增。所以最小值出现在

$$u = v\cos\theta,$$

即刚好以雨的水平速率行驶。此时雨相对于巴士铅直落下，只有车顶会淋湿，总量为 $R(c) = Da\tan\theta$（以我们的单位计）。

若 $a\sin\theta = h\cos\theta$：当 $u \ge c$ 时，$R(u) = Dh$，与 u 无关。若 $w \ge v\cos\theta$，则 $v\cos\theta$ 与 w 之间的任何车速都同样好，而低于 $v\cos\theta$ 的车速都较差。（若 $w < v\cos\theta$，则适用第一种情况，取 $u = w$。）从物理角度看，超过 $u = c$ 之后，开快一点在车顶省下的雨量，刚好等于车头多淋的雨量。这个条件即 $\tan\theta = h/a$：从路面看，雨沿着与巴士侧面一条对角线平行的方向落下。

回程

现在巴士向南行驶，迎着雨前进。按我们以北为正的约定，其速度为 $-u$，所以雨的相对水平速度为 $v\cos\theta + u > 0$，雨总是打中车头。总量与下式成正比：

$$R(u) = \frac{D}{u}\Big(h(v\cos\theta + u) + av\sin\theta\Big) = D\left(h + \frac{v\,(h\cos\theta + a\sin\theta)}{u}\right),$$

它对 u 严格递减。所以回程时，无论 a、h 和 θ 取甚么值，司机都应尽量开快，即 $u = w$。

验算

- 取 $a = 1.3$、$h = 0.8$、$v = 2$、$\theta = 0.6$ 及 $u = 0.7$，公式给出速率 2.229（以单位雨点密度及单位阔度计）。用电脑模拟随机分布的雨点穿过巴士截面，得出 2.232。在 $u = 0$、$u > v\cos\theta$ 以及 u 为负（回程）的情况下，吻合程度同样良好。
- 取 $\theta \to \frac{1}{2}\pi$（雨铅直落下）。此时 $a\sin\theta > h\cos\theta$，所以应全速行驶。这很合理：只有车顶会淋湿，在雨中的时间越短，淋到的雨就越少。
- 在 $\tan\theta = h/a$ 的相等情况下，总量 Dh 在 $u = v\cos\theta$、$1.5v\cos\theta$ 及 $3v\cos\theta$ 时都相同，与预测一致。

常见错误

- 求速率的最小值，而不是总量的最小值：必须把速率乘以行程时间 D/u。
- 忘记当 $u > v\cos\theta$ 时，雨打中的是车头而不是车尾。这正是需要绝对值的原因。
- 回程时再次使用 $|v\cos\theta - u|$；此时相对速率应为 $v\cos\theta + u$。

STEP 之外

“密度 × 面积 × 法向速度”这种计算，就是向量场（vector field）穿过一个曲面的通量。在向量微积分中，它成为曲面积分（surface integral）$\iint \rho\,\mathbf{v}\cdot\mathbf{n}\,\mathrm{d}S$，这是流体力学和电磁学中守恒定律（conservation laws）的出发点。改用巴士的参考系，就是一次伽利略变换（Galilean transformation），它是相对论在经典力学中的先驱。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）同时讨论了相对运动和这类通量论证。

详细解答

STEP I 2015，第 11 题

力学 · 摩擦力与力矩

STEP I 2015，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/15-s1-q11

关键思路

圆柱体上的每一个法向反作用力（normal reaction）都沿半径方向作用，因此必定通过圆心。所以，对每个圆心取力矩（moments），重力和所有法向反作用力都会消去，只剩下摩擦力（friction）；而这些摩擦力与圆心的距离全部相同。这样第 (i) 部分几乎不费吹灰之力便可得出。至于第 (ii) 部分，先对每个圆柱体沿斜面方向分解力，再在题目中两条不等式所涉及的两个接触点应用摩擦定律。

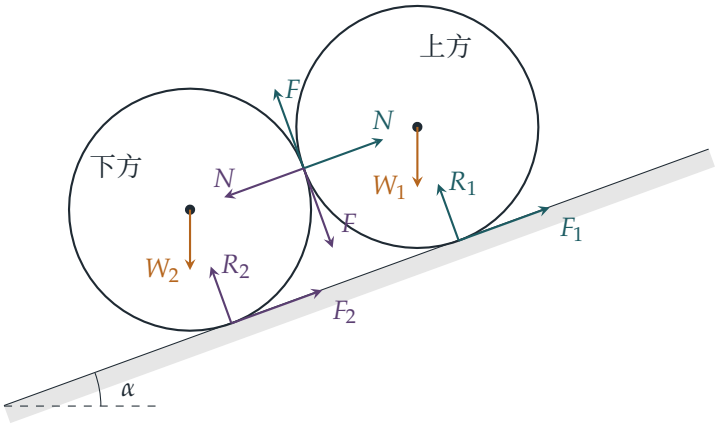
解答

我们把两个圆柱体视为半径同为 r 的均匀刚体（rigid body）圆柱，两者都与斜面接触，而上方的圆柱体位于斜面较高的位置。由于两者半径相等，连接两圆心的直线与斜面平行，所以两个圆柱体的接触点位于斜面上方高度 r 处，与两个圆心同高。两圆柱体之间的法向力沿圆心连线（与斜面平行）作用，而它们之间的摩擦力则垂直于斜面。

取 x 方向为沿斜面向上，y 方向为垂直于斜面并指离斜面。设

- R_1、R_2 为斜面对上方及下方圆柱体的法向反作用力；
- F_1、F_2 为斜面对上方及下方圆柱体的摩擦力，以沿斜面向上为正（我们容许它们带正负号，负值即表示沿斜面向下）；
- N 为下方圆柱体对上方圆柱体的法向力，方向沿斜面向上；
- F 为下方圆柱体对上方圆柱体的摩擦力，以指离斜面为正（同样带正负号）。

根据牛顿（Newton）第三定律，上方圆柱体对下方圆柱体施加沿斜面向下的力 N，以及指向斜面的力 F。



作用于上方圆柱体的力以蓝绿色表示，作用于下方圆柱体的力则以紫红色表示。

第 (i) 部分

对上方圆柱体的圆心取力矩。其重力作用于圆心，而 R_1 和 N 都沿半径作用，所以这些力全都没有力矩。剩下的只有两个摩擦力，各自沿切线方向作用，与圆心的距离均为 r：

- F_1 在圆柱体最低点沿斜面向上作用，产生（按图中方向）逆时针的力矩 rF_1；
- F 在接触点指离斜面作用，而该接触点位于圆心靠斜面下方的一侧，产生顺时针的力矩 rF。

要达至平衡（equilibrium），两者必须互相抵消，所以 $rF_1 = rF$，即 $F_1 = F$。

> ? **想一想** 为甚么对圆心取力矩时可以不理会法向反作用力？

圆上的法向反作用力垂直于接触点的切线，因此沿半径方向，其作用线通过圆心，对圆心的力矩为零。正正因为这样，在这里对圆心取力矩才是正确的选择。

同理，对下方圆柱体而言，只有 F_2（在其最低点沿斜面向上，产生逆时针力矩 rF_2）以及在接触点指向斜面的摩擦力 F（该接触点位于其圆心靠斜面上方的一侧，产生顺时针力矩 rF）有力矩。所以 $F_2 = F$。

因此作为带正负号的量，$F = F_1 = F_2$；特别地，它们的大小相等，得证。

第 (ii) 部分

对每个圆柱体沿平行于斜面的方向分解力：

$$\begin{aligned} \text{上方：} &\quad F_1 + N = W_1 \sin\alpha, \\ \text{下方：} &\quad F_2 = N + W_2 \sin\alpha. \end{aligned}$$

两式相加，并利用 $F_1 = F_2 = F$，得

$$2F = (W_1 + W_2)\sin\alpha, \qquad \text{所以} \qquad F = \tfrac{1}{2}(W_1 + W_2)\sin\alpha > 0,$$

这亦说明每个摩擦力的方向都与我们图中所画的一致。代回原式，

$$N = W_1\sin\alpha - F = \tfrac{1}{2}(W_1 - W_2)\sin\alpha,$$

由于 $W_1 > W_2$ 且 $\alpha > 0$，这是正数，所以两个圆柱体确实互相压着。对上方圆柱体沿垂直于斜面的方向分解力，

$$R_1 + F = W_1\cos\alpha, \qquad \text{所以} \qquad R_1 = W_1\cos\alpha - F.$$

想一想 我们实际上需要哪些摩擦不等式?

在平衡状态下，每个接触点的摩擦力最多只能是摩擦系数（coefficient of friction）乘以该处的法向反作用力。第一个结果来自两个圆柱体之间的接触点，第二个结果则来自上方圆柱体与斜面之间的接触点。（下方的接触点会给出涉及 μ_2 的第三个条件，但题目没有要求。）

*两个圆柱体之间。*我们需要 $F \le \mu N$:

$$\tfrac{1}{2}(W_1 + W_2)\sin\alpha \le \mu \cdot \tfrac{1}{2}(W_1 - W_2)\sin\alpha.$$

除以正数 $\frac{1}{2}(W_1 - W_2)\sin\alpha$，得

$$\mu \ge \frac{W_1 + W_2}{W_1 - W_2},$$

得证。

*上方圆柱体与斜面之间。*我们需要 $F \le \mu_1 R_1 = \mu_1(W_1\cos\alpha - F)$，即

$$(1+\mu_1)F \le \mu_1 W_1\cos\alpha \quad \Longleftrightarrow \quad \tfrac{1}{2}(1+\mu_1)(W_1+W_2)\sin\alpha \le \mu_1 W_1\cos\alpha.$$

斜面的倾角满足 $0 < \alpha < \frac{1}{2}\pi$，所以 $\cos\alpha > 0$，而且 $(1+\mu_1)(W_1+W_2) > 0$。两者同时除去，得

$$\tan\alpha \le \frac{2\mu_1 W_1}{(1+\mu_1)(W_1+W_2)},$$

得证。（注意由于 $\mu_1 R_1 \ge F > 0$，这亦迫使 $R_1 > 0$，所以上方圆柱体会一直留在斜面上。）

验算

- 取 $W_1 = 3$、$W_2 = 1$，我们需要 $\mu \ge 2$；若再取 $\mu_1 = 1$，则需要 $\tan\alpha \le \frac{3}{4}$。把这类数值代入六条平衡方程（每个圆柱体两条分解方程和一条力矩方程）并直接求解，所得的 F、N 和 R_1 与上面的结果相同。

- 当 $W_1 \to W_2$ 时，μ 的下界趋向无穷大。这是合理的：此时 $N \to 0$，两个圆柱体几乎没有互相压着，但它们之间的摩擦力仍然必须是 $\frac{1}{2}(W_1 + W_2)\sin\alpha$。
- 当 $\mu_1 \to \infty$ 时，该上界变成 $\tan\alpha \le 2W_1/(W_1 + W_2)$，这正正就是条件 $R_1 \ge 0$：一旦上方圆柱体开始离开斜面，即使斜面完全粗糙也无补于事。

常见错误

- 以为两个圆柱体之间的摩擦力沿斜面方向作用。其实它沿接触点的公切线作用，而在这里公切线垂直于斜面。
- 忘记力矩方程。单靠分解力只得出四条方程，却有六个未知数；确定摩擦力的正是那两条力矩方程。
- 计算法向反作用力 R_1 时漏掉 F。下方圆柱体的摩擦力把上方圆柱体推离斜面，所以 R_1 小于 $W_1\cos\alpha$。

STEP 之外

刚体处于平衡，当且仅当合力和合力矩同时为零；而巧妙地选择取力矩的点，是刚体静力学（rigid-body statics）的主要技巧。在这题中，方程的数目恰好等于未知数的数目；若未知数比方程多，问题便是*超静定*（statically indeterminate）的，必须为物体如何变形建立模型才能完成求解。Kleppner 和 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）对刚体静力学和力矩有细致的论述。

详细解答

STEP I 2015，第 12 题

概率与统计 · 泊松分布的稀疏化

STEP I 2015，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/15-s1-q12

关键思路

以到达人数作条件（conditioning），需要做手术的人数便成为一个二项（binomial）计数。然后对所有可能的到达人数求和（即全概率定律（law of total probability）），便得出需要做手术人数的分布，而这个和式会化简为一个指数级数（exponential series）。至于最后一部分，关键事实是：若两个平均值相同的独立（independent）泊松（Poisson）计数以其总和作条件，则两者之间的分配服从二项分布。

解答

设 X 为某一天到达的伤者人数，Y 为其中需要做手术的人数。

第 (i) 部分

已知 $X = n$，这 n 名伤者各自以概率 $\frac{1}{4}$ 需要做手术，且彼此独立。所以在此条件下 $Y \sim \mathrm{B}(n, \frac{1}{4})$，而对于 $0 \le r \le n$，

$$\mathrm{P}(Y = r \mid X = n) = \binom{n}{r}\left(\frac{1}{4}\right)^r\left(\frac{3}{4}\right)^{n-r} = \binom{n}{r}\frac{3^{n-r}}{4^n}.$$

（若 $r > n$，则概率为 0。）

第 (ii) 部分

固定 $r \geq 0$。由于 $Y = r$ 只可能在 $X \geq r$ 时发生，全概率定律给出

$$\begin{aligned} \mathrm{P}(Y = r) &= \sum_{n=r}^{\infty} \mathrm{P}(X = n)\,\mathrm{P}(Y = r \mid X = n) \\ &= \sum_{n=r}^{\infty} \frac{\mathrm{e}^{-8} 8^n}{n!} \cdot \frac{n!}{r!\,(n-r)!} \left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r}. \end{aligned}$$

$n!$ 互相抵消。接着把 $8^n = 8^r \cdot 8^{n-r}$ 拆开，使得 $8^r (\frac{1}{4})^r = 2^r$ 及 $8^{n-r}(\frac{3}{4})^{n-r} = 6^{n-r}$：

$$\begin{aligned} \mathrm{P}(Y = r) &= \frac{\mathrm{e}^{-8}\, 2^r}{r!} \sum_{n=r}^{\infty} \frac{6^{n-r}}{(n-r)!} \\ &= \frac{\mathrm{e}^{-8}\, 2^r}{r!} \sum_{k=0}^{\infty} \frac{6^k}{k!} \qquad (k = n - r) \\ &= \frac{\mathrm{e}^{-8}\, 2^r}{r!}\, \mathrm{e}^6 = \frac{\mathrm{e}^{-2}\, 2^r}{r!}. \end{aligned}$$

此式对每个 $r = 0, 1, 2, \ldots$ 都成立，所以每天需要做手术的人数服从平均值为 $8 \times \frac{1}{4} = 2$ 的泊松分布（Poisson distribution）。

> ? **想一想** e^6 从何而来?

把所有只与 r 有关的部分移到和式之外后，剩下的就是 $\sum 6^k / k!$，即 e^6 的指数级数。把 8^n 拆成 $8^r \cdot 8^{n-r}$，正正是令余下的和式只依赖于 $n - r$ 的关键。

第 (iii) 部分

设 M 和 T 分别为星期一和星期二需要做手术的人数。由第 (ii) 部分，两者均为 $\mathrm{Po}(2)$。由于不同日子的到达人数互相独立，而伤者是否需要做手术亦彼此独立，所以 M 与 T 独立。于是

$$\mathrm{P}(M = 8 \mid M + T = 12) = \frac{\mathrm{P}(M = 8,\ T = 4)}{\mathrm{P}(M + T = 12)}.$$

由独立性，分子为

$$\mathrm{P}(M = 8)\,\mathrm{P}(T = 4) = \frac{\mathrm{e}^{-2} 2^8}{8!} \cdot \frac{\mathrm{e}^{-2} 2^4}{4!} = \frac{\mathrm{e}^{-4}\, 2^{12}}{8!\,4!}.$$

至于分母，独立泊松随机变量之和仍服从泊松分布，其平均值相加，所以 $M + T \sim \mathrm{Po}(4)$。若不想直接引用这个结果，亦可直接推导：

$$\mathrm{P}(M + T = 12) = \sum_{m=0}^{12} \frac{\mathrm{e}^{-2} 2^m}{m!} \cdot \frac{\mathrm{e}^{-2} 2^{12-m}}{(12-m)!} = \frac{\mathrm{e}^{-4} 2^{12}}{12!} \sum_{m=0}^{12} \binom{12}{m} = \frac{\mathrm{e}^{-4}\, 4^{12}}{12!},$$

当中用了 $\sum_m \binom{12}{m} = 2^{12}$。两者相除，

$$\begin{aligned}\mathrm{P}(M = 8 \mid M + T = 12) &= \frac{\mathrm{e}^{-4}\,2^{12}}{8!\,4!} \cdot \frac{12!}{\mathrm{e}^{-4}\,4^{12}} \\ &= \binom{12}{8}\left(\frac{1}{2}\right)^{12} = \frac{495}{4096}.\end{aligned}$$

由于 $495 = 3^2 \cdot 5 \cdot 11$ 是奇数，而 $4096 = 2^{12}$，此分数已是最简分数（lowest terms）。所以所求概率为 $\dfrac{495}{4096}$。

想一想 为甚么答案看起来像一个二项概率?

已知总数为 12，由于两天的平均值相同，这 12 宗手术个案中的每一宗来自星期一或星期二的可能性均等。所以星期一的人数服从 $\mathrm{B}(12, \frac{1}{2})$。留意原本的平均值 8 已完全消失。

验算

- 第 (ii) 部分的平均值与 $\mathrm{E}(Y) = \mathrm{E}\big(\mathrm{E}(Y \mid X)\big) = \mathrm{E}(X/4) = 2$ 一致。
- 对 $r = 0, \ldots, 14$ 以数值方法计算第 (ii) 部分的级数,结果在舍入误差范围内与 $\mathrm{e}^{-2}2^r/r!$ 相同。例如 $\mathrm{P}(Y = 0) = \sum_n \mathrm{e}^{-8}8^n(\frac{3}{4})^n/n! = \mathrm{e}^{-8}\mathrm{e}^{6}$。
- 对 $m = 0, \ldots, 12$，条件概率 $\binom{12}{m}/2^{12}$ 之和为 1，而 $495/4096 \approx 0.121$。以电脑模拟整个模型（泊松到达，然后每人各以概率 $\frac{1}{4}$ 需要做手术），所得数值与此相符。

常见错误

- 在第 (ii) 部分中由 $n = 0$ 而非 $n = r$ 开始求和：若到达人数少于 r，就不可能有 r 名伤者需要做手术。
- 在第 (iii) 部分中每天都用 $\mathrm{Po}(8)$。条件是针对手术个案，所以相关的平均值是 2。（在这里平均值无论如何都会抵消，但推导过程仍应正确。）
- 题目要求最简分数时，却把答案写成 $\binom{12}{8}/2^{12}$ 或 $1980/16384$。

STEP 之外

第 (ii) 部分称为*稀疏化*（thinning）：若泊松过程（Poisson process）中的每一点各自独立地以概率 p 保留，则被保留的点构成一个速率乘以 p 的泊松过程，而且，或许令人意外的是，它与被舍弃的点互相独立。第 (iii) 部分则是其逆向：以总和作条件时，独立的泊松计数按多项分布（multinomial distribution）分配。这两个事实是泊松过程理论的基础，

你会在大学第一个概率课程中学到。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）对两者均有论述。

详细解答

STEP I 2015，第 13 题

概率与统计 • 条件概率

STEP I 2015，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/15-s1-q13

关键思路

以第一个 6 出现的时间作条件（conditioning）。若已知第一个 6 出现在第 n 次投掷，则之前的 $n-1$ 次投掷各自独立，且每次同样可能是 $1,\ldots,5$ 中任何一个；于是关于第一个 6 之前出现的 4 和 5 的问题，就变成普通的二项分布（binomial distribution）问题。再对 n 求和，便得到几何级数（geometric series）及其导数，而这正是提示中二项展开式（binomial expansion）派上用场之处。

解答

各次投掷互相独立，每一面的概率均为 $\frac{1}{6}$。记第一个 6 出现在第 n 次投掷这事件为 A，并写 $p_n = \mathrm{P}(A)$。

第 (i) 部分

第一个 6 出现在第 n 次投掷，当且仅当第 $1,\ldots,n-1$ 次投掷都不是 6，而第 n 次投掷是 6。由独立性，

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

条件化的关键事实。 在 A 发生的条件下，对第 $1,\ldots,n-1$ 次投掷，我们只知道它们“不是 6”。由于骰子公平而各次投掷独立，这 $n-1$ 次投掷互相独立，且每次都在 $\{1,2,3,4,5\}$ 上均匀分布。之后每一部分都会用到这一点，并配合全概率定律（law of total probability）$\mathrm{P}(X) = \sum_n \mathrm{P}(X \mid A)\, p_n$。

第 (ii) 部分

较容易的做法是求 $\mathrm{P}(B')$，即在第一个 6 之前没有出现 5 的概率。在 A 的条件下，之前 $n-1$ 次投掷每次避开 5 的概率为 $\frac{4}{5}$，所以

$$\mathrm{P}(B') = \sum_{n=1}^{\infty} \left(\tfrac{4}{5}\right)^{n-1} \left(\tfrac{5}{6}\right)^{n-1} \tfrac{1}{6} = \tfrac{1}{6} \sum_{n=1}^{\infty} \left(\tfrac{2}{3}\right)^{n-1} = \tfrac{1}{6} \cdot \frac{1}{1-\frac{2}{3}} = \tfrac{1}{2}.$$

因此 $\mathrm{P}(B) = \frac{1}{2}$。

第 (iii) 部分

由容斥原理（inclusion–exclusion principle）（以补集形式表示），

$$\mathrm{P}(B \cap C) = 1 - \mathrm{P}(B') - \mathrm{P}(C') + \mathrm{P}(B' \cap C').$$

由 4 和 5 两面之间的对称性，$\mathrm{P}(C') = \mathrm{P}(B') = \frac{1}{2}$。至于 $B' \cap C'$，在 A 的条件下，之前每次投掷都必须同时避开 4 和 5，概率为 $\frac{3}{5}$：

$$\mathrm{P}(B' \cap C') = \sum_{n=1}^{\infty} \left(\tfrac{3}{5}\right)^{n-1} \left(\tfrac{5}{6}\right)^{n-1} \tfrac{1}{6} = \tfrac{1}{6} \cdot \frac{1}{1-\frac{1}{2}} = \tfrac{1}{3}.$$

所以 $\mathrm{P}(B \cap C) = 1 - \frac{1}{2} - \frac{1}{2} + \frac{1}{3} = \frac{1}{3}$。

第 (iv) 部分

在 A 的条件下，之前 $n-1$ 次投掷中 5 的个数服从 $\mathrm{B}(n-1, \frac{1}{5})$，所以

$$\mathrm{P}(D \mid A) = (n-1)\tfrac{1}{5} \left(\tfrac{4}{5}\right)^{n-2} \qquad (n \geq 2),$$

而当 $n = 1$ 时此概率为 0。因此

$$\mathrm{P}(D) = \sum_{n=2}^{\infty} (n-1)\tfrac{1}{5} \left(\tfrac{4}{5}\right)^{n-2} \left(\tfrac{5}{6}\right)^{n-1} \tfrac{1}{6}.$$

抽出一个因子 $\frac{5}{6}$，其余两两配对：$\frac{1}{5} \cdot \frac{5}{6} \cdot \frac{1}{6} = \frac{1}{36}$ 及 $\frac{4}{5} \cdot \frac{5}{6} = \frac{2}{3}$。设 $m = n - 1$，

$$\mathrm{P}(D) = \tfrac{1}{36} \sum_{m=1}^{\infty} m \left(\tfrac{2}{3}\right)^{m-1} = \tfrac{1}{36} \left(1 - \tfrac{2}{3}\right)^{-2} = \tfrac{9}{36} = \tfrac{1}{4},$$

这里用了题目所给、指数为 -2 的展开式：$(1-x)^{-2} = 1 + 2x + 3x^2 + \cdots$。

想一想 怎样才能求得 $\mathrm{P}(D \cup E)$，而无须逐一列出其中一个或两个事件同时发生的所有情况?

第 (v) 部分

利用 $\mathrm{P}(D\cup E)=\mathrm{P}(D)+\mathrm{P}(E)-\mathrm{P}(D\cap E)$，其中由 4 与 5 的对称性得 $\mathrm{P}(E)=\mathrm{P}(D)=\frac{1}{4}$。所以只需求 $\mathrm{P}(D\cap E)$。

在 A 的条件下，之前 $n-1$ 次投掷中恰有一个 4 和恰有一个 5，即要选定哪一次是 5、哪一次是 4（共 $(n-1)(n-2)$ 种有序选法），而其余 $n-3$ 次投掷全部落在 $\{1,2,3\}$ 中：

$$\mathrm{P}(D\cap E\mid A)=(n-1)(n-2)\left(\tfrac{1}{5}\right)^2\left(\tfrac{3}{5}\right)^{n-3}\qquad(n\geq 3).$$

于是

$$\mathrm{P}(D\cap E)=\sum_{n=3}^{\infty}(n-1)(n-2)\tfrac{1}{25}\left(\tfrac{3}{5}\right)^{n-3}\left(\tfrac{5}{6}\right)^{n-1}\tfrac{1}{6}.$$

抽出 $\left(\frac{5}{6}\right)^2$，得常数 $\frac{1}{25}\cdot\frac{25}{36}\cdot\frac{1}{6}=\frac{1}{216}$，并有 $\frac{3}{5}\cdot\frac{5}{6}=\frac{1}{2}$。设 $k=n-3$，

$$\begin{aligned}\mathrm{P}(D\cap E)&=\tfrac{1}{216}\sum_{k=0}^{\infty}(k+2)(k+1)\left(\tfrac{1}{2}\right)^k=\tfrac{2}{216}\sum_{k=0}^{\infty}\frac{(k+2)!}{k!\,2!}\left(\tfrac{1}{2}\right)^k\\&=\tfrac{1}{108}\left(1-\tfrac{1}{2}\right)^{-3}=\tfrac{8}{108}=\tfrac{2}{27},\end{aligned}$$

这里用了题目所给、指数为 -3 的展开式，其 x^k 项的系数为 $\frac{(k+2)!}{k!\,2!}$。因此

$$\mathrm{P}(D\cup E)=\tfrac{1}{4}+\tfrac{1}{4}-\tfrac{2}{27}=\tfrac{27-4}{54}=\tfrac{23}{54}.$$

验算

有一个更巧妙的方法可以看出每一个答案，用来验算再好不过。把所有掷出 1、2、3 的投掷（以及在 (ii) 和 (iv) 中连 4 也一并）忽略掉：它们从不影响这些事件。余下的相关投掷组成一个序列，其中每个相关的面出现的机会均等。

- (ii)：第一次掷出 5 或 6 时，两者机会均等，所以 $\mathrm{P}(B)=\frac{1}{2}$。
- (iii)：在 4、5、6 三面之中，每一面同样可能是最后一个首次出现的，而 $B\cap C$ 正是说 6 最后出现，所以概率为 $\frac{1}{3}$。
- (iv)：D 表示由 5 和 6 组成的序列以 5, 6 开始：概率为 $\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$。
- (v)：$D\cap E$ 表示由 4、5、6 组成的序列以 4, 5, 6 或 5, 4, 6 开始：概率为 $2\left(\frac{1}{3}\right)^3=\frac{2}{27}$。

全部五个答案亦已通过电脑模拟 3×10^5 个投掷序列加以确认，模拟得出 $\mathrm{P}(D\cup E)\approx 0.426$，而理论值为 $\frac{23}{54}\approx 0.4259$。

常见错误

- 把之前的投掷当作在全部六面上均匀分布：在 A 的条件下，它们不可能是 6，所以每次是 5 的概率为 $\frac{1}{5}$，而非 $\frac{1}{6}$。
- 在 (v) 中漏了因子 $(n-1)(n-2)$，或者用了 $\binom{n-1}{2}$，后者没有区分两次投掷中哪一次是 4。
- 在 (v) 中直接把 $\mathrm{P}(D)+\mathrm{P}(E)$ 相加，却没有减去重叠部分。

STEP 之外

像本题这样，以一个随机停时（stopping time）作条件再求和，正是全概率定律最实用的形式；而你遇到的级数，则是负二项分布（negative binomial distribution）的生成函数（generating function）。“忽略无关投掷”这个验算方法，体现了一个威力强大的想法：只在有相关事情发生的时刻观察一个过程，往往比原来的过程简单得多。Grimmett 和 Welsh 所著的 *Probability: An Introduction*（《概率导论》）对这两个想法都有深入发展。

详细解答

STEP II 2015，第 1 题

纯数学 · 不等式与级数

STEP II 2015，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/15-s2-q1

关键思路

每一条函数与对数（logarithm）之间的不等式，只要代入一个精心选取的 x 值，就会变成级数（series）中某一项与两个对数之差之间的不等式。把这些差相加，几乎所有项都会互相抵消（即该和式“伸缩相消”（telescopes）），只剩下一两个对数。技巧在于选取适当的代换，使对数能拆成一个差。

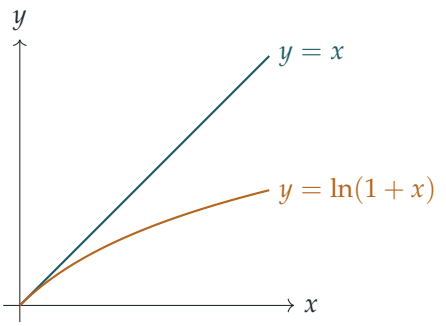
解答

第 (i) 部分

对 $x \geq 0$，设 $f(x) = x - \ln(1+x)$。那么 $f(0) = 0 - \ln 1 = 0$，而且

$$f'(x) = 1 - \frac{1}{1+x} = \frac{x}{1+x} > 0 \qquad \text{对} x > 0.$$

所以 f 在 $[0, \infty)$ 上严格递增(strictly increasing),因此对每个 $x > 0$,都有 $f(x) > f(0) = 0$。即对所有正数 x，$x > \ln(1+x)$，正是所求。



现在对正整数 k 取 $x = \frac{1}{k}$：

$$\frac{1}{k} > \ln\left(1+\frac{1}{k}\right) = \ln\left(\frac{k+1}{k}\right) = \ln(k+1) - \ln k.$$

把 $k = 1, 2, \ldots, n$ 的各式相加，右方便伸缩相消：

$$\begin{aligned}\sum_{k=1}^{n} \frac{1}{k} &> \big(\ln 2 - \ln 1\big) + \big(\ln 3 - \ln 2\big) + \cdots + \big(\ln(n+1) - \ln n\big) \\ &= \ln(n+1) - \ln 1 = \ln(n+1).\end{aligned}$$

这就是所求的结果。（附带一提，由于 $\ln(n+1) \to \infty$，调和级数（harmonic series）发散（diverges）。）

第 (ii) 部分

对 $0 \le x < 1$，设 $g(x) = x + \ln(1-x)$。那么 $g(0) = 0$，而且

$$g'(x) = 1 - \frac{1}{1-x} = \frac{-x}{1-x} < 0 \qquad \text{对} 0 < x < 1,$$

因为在该区间内 $1 - x > 0$。所以 g 严格递减，且对 $0 < x < 1$ 有 $g(x) < 0$；即

$$x < -\ln(1-x) \qquad (0 < x < 1).$$

? **想一想** 应该代入哪个 x 值？为甚么不能用 $k = 1$？

我们希望左方是 $\frac{1}{k^2}$，所以取 $x = \frac{1}{k^2}$。这要求 $x < 1$，故 $k \ge 2$；当 $k = 1$ 时，对数 $\ln(1-1)$ 没有定义。这就是为甚么要把 $\frac{1}{1} = 1$ 这一项另外处理，而它正是上界 $1 + \ln 2$ 中的那个 1。对 $k \ge 2$，关键在于 $1 - \frac{1}{k^2}$ 可以因式分解：

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

所以

$$\frac{1}{k^2} < -\ln\frac{(k-1)(k+1)}{k^2} = \big(\ln k - \ln(k-1)\big) - \big(\ln(k+1) - \ln k\big).$$

记 $a_k = \ln k - \ln(k-1)$，则右方为 $a_k - a_{k+1}$。由 $k = 2$ 加至 N，和式伸缩相消：

$$\sum_{k=2}^{N} \frac{1}{k^2} < a_2 - a_{N+1} = \ln 2 - \ln\left(\frac{N+1}{N}\right) < \ln 2,$$

因为 $\ln\frac{N+1}{N} > 0$。

所以部分和（partial sums）$\sum_{k=1}^{N} \frac{1}{k^2}$ 随 N 递增，而且全部小于 $1 + \ln 2$。上有界（bounded above）的递增数列必定收敛（converges），因此这个无穷级数收敛，而其和至多为 $1 + \ln 2$。

要使不等式严格成立，须注意取极限只能保留 $\leq$，所以我们要预留一点空间。$k=2$ 时的不等式 $\frac{1}{4} < \ln\frac{4}{3}$ 是严格的；设 $\delta = \ln\frac{4}{3} - \frac{1}{4} > 0$。那么对每个 $N \geq 2$，同样的伸缩相消给出

$$\sum_{k=2}^{N} \frac{1}{k^2} < \ln 2 - \ln\left(\frac{N+1}{N}\right) - \delta < \ln 2 - \delta,$$

所以取极限得 $\sum_{k=2}^{\infty} \frac{1}{k^2} \leq \ln 2 - \delta < \ln 2$。加上 $k=1$ 的一项，

$$\sum_{k=1}^{\infty} \frac{1}{k^2} < 1 + \ln 2,$$

这就是所求的结果。

验算

- 第 (i) 部分取 $n=1$：$1 > \ln 2 \approx 0.693$。取 $n=10$：$\sum_{k=1}^{10} \frac{1}{k} \approx 2.929$，而 $\ln 11 \approx 2.398$。
- 第 (ii) 部分：真正的和是 $\frac{\pi^2}{6} \approx 1.6449$，稳稳低于 $1+\ln 2 \approx 1.6931$。事实上，我们的推导给出稍为更紧的上界 $1+\ln 2 - \delta \approx 1.6555$，其中 $\delta \approx 0.0377$。
- $\sum_{k=2}^{N} \frac{1}{k^2}$ 的伸缩相消上界 $\ln 2 - \ln\frac{N+1}{N}$，已用电脑对所有不超过 10^5 的 N 验证。

常见错误

- 连 $k=1$ 也代入 $x = \frac{1}{k^2}$：此时 $\ln(1-x)$ 没有定义；而忘记把第一项另外处理，就会漏掉 $1+\ln 2$ 中的 1。
- 只证明 $f'(x) > 0$ 而没有指出 $f(0) = 0$：一个函数可以是递增的，但仍然是负数。
- 对严格不等式取极限，却不加说明便声称极限仍是严格不等式；极限只能给出 $\leq$。

STEP 之外

把和式与对数（或积分）比较，正是积分判别法（integral test）的起点；而把第 (i) 部分加以精炼，可以证明 $\sum_{k=1}^{n} \frac{1}{k} - \ln n$ 收敛至欧拉常数（Euler's constant）$\gamma \approx 0.5772$。精确值 $\sum \frac{1}{k^2} = \frac{\pi^2}{6}$ 是欧拉（Euler）对巴塞尔问题（Basel problem）的解答，大学一年级课程常以傅里叶级数（Fourier series）推导出来。我们所用的单调收敛（monotone convergence）论证，即有界递增数列必定收敛，是实分析（real analysis）的基石之一；Abbott 的 *Understanding Analysis*（《理解分析》）对此有细致的阐述。

详细解答

STEP II 2015，第 2 题

纯数学 · 三角形几何

STEP II 2015，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：

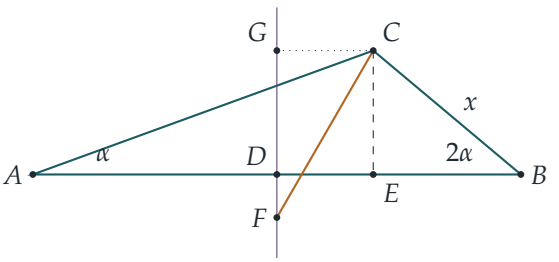
step.maths.org/questions/15-s2-q2

关键思路

第一个结果只需用正弦定理（sine rule）配合 $\sin 3\alpha$ 的三倍角公式（triple-angle formula）。至于 DE，就把所有长度都沿 AB 从其中一端量起：CE 两侧的直角三角形（right-angled triangle）直接给出 AE 和 BE。DE 的答案出奇地简单，而正是这份简单令三等分（trisection）得以成立：它告诉我们 C 与垂直平分线（perpendicular bisector）相距多远，再与 $CF = x$ 比较，便能确定一个角。

解答

全文中，C 处的角为 $\angle ACB = \pi - 3\alpha$，又因 2α 是锐角（acute angle），故 $0 < \alpha < \frac{\pi}{4}$。



长度 AB

由正弦定理，并利用 $\sin(\pi - 3\alpha) = \sin 3\alpha$，得

$$\frac{AB}{\sin 3\alpha} = \frac{x}{\sin \alpha}.$$

再用三倍角公式 $\sin 3\alpha = 3\sin\alpha - 4\sin^3\alpha$，并除以 $\sin\alpha \neq 0$，得

$$AB = x\,\frac{3\sin\alpha - 4\sin^3\alpha}{\sin\alpha} = (3 - 4\sin^2\alpha)\,x,$$

此即所求。

长度 DE

A 和 B 处的角都是锐角，所以由 C 所作垂线的垂足（foot of the perpendicular）E 位于 A 与 B 之间。在直角三角形 CEB 中，

$$BE = x\cos 2\alpha, \qquad CE = x\sin 2\alpha.$$

因此

$$AE = AB - BE = x\big(3 - 4\sin^2\alpha - (1 - 2\sin^2\alpha)\big) = x(2 - 2\sin^2\alpha) = 2x\cos^2\alpha,$$
$$AD = \tfrac{1}{2}AB = x\big(\tfrac{3}{2} - 2\sin^2\alpha\big).$$

两式相减，

$$AE - AD = x\big(2 - 2\sin^2\alpha - \tfrac{3}{2} + 2\sin^2\alpha\big) = \tfrac{1}{2}x.$$

此值为正，所以 E 位于 D 与 B 之间，且

$$DE = \tfrac{1}{2}x,$$

与 α 无关。

三等分

设 G 为由 C 向 AB 的垂直平分线 DF 所作垂线的垂足。由于 CE 和 GD 都垂直于 AB，$CGDE$ 是一个矩形（rectangle），所以 $CG = DE = \frac{1}{2}x$，而且 CG 平行于 AB，即垂直于 CE。

? **想一想** 怎样把 $CF = x$ 转化为一个角?

三角形 CGF 在 G 处为直角，斜边（hypotenuse）$CF = x$，一边 $CG = \frac{1}{2}x$。所以

$$\cos\angle GCF = \frac{CG}{CF} = \frac{1}{2}, \qquad \text{所以} \qquad \angle GCF = \frac{\pi}{3}.$$

垂直平分线上与 C 相距 x 的点有两个，分别在 G 的两侧。从射线 CE（向下指向 AB）起，朝 A 的方向量度，射线 CG 所成的角为 $\frac{\pi}{2}$，所以两条候选射线 CF 所成的角分别为 $\frac{\pi}{2} - \frac{\pi}{3} = \frac{\pi}{6}$（$G$ 下方的点）和 $\frac{\pi}{2} + \frac{\pi}{3} = \frac{5\pi}{6}$（$G$ 上方的点）。

现在判断哪一点与 B 位于 AC 的同一侧。在直角三角形 ACE 中，$\angle ACE = \frac{\pi}{2} - \alpha$，所以射线 CA 在同一侧与 CE 所成的角为 $\frac{\pi}{2} - \alpha$。由 C 出发、位于直线 AC 上 B 那一侧的射

线，就是按这种量法角度小于 $\frac{\pi}{2}-\alpha$ 的射线（或位于 CE 另一侧、朝向 B 的射线）。由于 $\alpha<\frac{\pi}{4}$，

$$\frac{\pi}{6}<\frac{\pi}{4}<\frac{\pi}{2}-\alpha<\frac{5\pi}{6},$$

所以 G 下方的点在 B 那一侧，而 G 上方的点则不在。因此 F 是下方的点，$\angle ECF=\frac{\pi}{6}$，而射线 CF 位于角 ACE 之内，因而也在 $\angle ACB$ 之内。所以

$$\angle ACF=\angle ACE-\angle ECF=\left(\frac{\pi}{2}-\alpha\right)-\frac{\pi}{6}=\frac{\pi}{3}-\alpha=\frac{1}{3}(\pi-3\alpha)=\frac{1}{3}\angle ACB.$$

故此 CF 在靠近 CA 的一边截出 $\angle ACB$ 的三分之一（而 $\angle FCB=\frac{2}{3}\angle ACB$）：直线 FC 三等分角 ACB。

验算

- 取 $\alpha=\frac{\pi}{6}$，三角形的三个角为 30°、60°、90°，而 $AB=(3-1)x=2x$，正是熟悉的斜边长度。于是 $AE=2x\cos^2 30^\circ=\frac{3}{2}x$，$AD=x$，所以 $DE=\frac{1}{2}x$。C 处的角为 90°，而 $\angle ACF=60^\circ-30^\circ=30^\circ$，正好是它的三分之一。
- 取 $\alpha=20^\circ$ 及 $x=1$，得 $AB\approx 2.532$ 及 $\angle ACB=120^\circ$；直接用坐标构作 F，得 $\angle ACF=40^\circ$。三个结果亦已用电脑对数以千计的随机三角形逐一验证。

常见错误

- 把 F 取在 G 上方。那一点与 C 的距离同样是 x，但它位于 AC 的另一侧，这正是题目要指明哪一侧的原因。
- 在把 DE 写成两段长度之差以前，忘记检查 E 是否位于 A 与 B 之间（以及位于 D 的哪一侧）。

STEP 之外

用尺规（ruler and compasses）三等分一般的角是不可能的，这个事实可用域论（field theory）证明：三等分 60° 需要构作一个不可约三次多项式（irreducible cubic）的根，而可构作（constructible）的长度只会满足一层层的二次方程。在本题中，三角形提供了额外信息（B 处的角是 A 处的两倍），正是这一点让一个简单的构作也能完成任务。想看更多优美的三角形几何，可参阅考克斯特（Coxeter）与 Greitzer 合著的 *Geometry Revisited*（《几何再探》）；至于不可能性的证明，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 2015，第 3 题

纯数学 · 计数与数学归纳法

STEP II 2015，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/15-s2-q3

关键思路

按最长的棒把三角形分类。被 T_n 计入但不被 T_{n-1} 计入的三角形，正好是那些用上长度为 n 的棒的三角形，所以 $T_n - T_{n-1}$ 就是满足 $a < b < n$ 及 $a + b \geq n$ 的数对个数。对中间那枝棒 b 的每个选择，这都是一个简单的计数，而这些计数构成一个等差数列（arithmetic progression）。把两个相邻的差相加便得到 $T_{2m} - T_{2m-2}$，这正是归纳步骤（inductive step）所需要的。

解答

计算最长棒为 c 的三角形

被 T_n 计入的三角形，其棒长为 $a < b < c \leq n$。当中 $c \leq n-1$ 的，恰好就是被 T_{n-1} 计入的三角形，所以

$$T_n - T_{n-1} = \#\{(a,b) : 1 \leq a < b < n,\ a + b \geq n\}.$$

固定中间的棒 b。我们需要 $n - b \leq a \leq b - 1$（由于 $b < n$，$a \geq 1$ 自动成立），所以当 $2b - n > 0$ 时有 $2b - n$ 个选择，否则没有。因此 b 由大于 $\frac{n}{2}$ 的最小整数取到 $n - 1$。

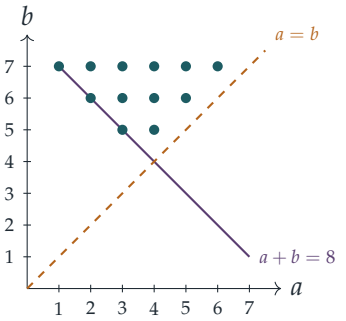
$n = 8$ 的情况

当 $c = 8$ 时，中间的棒可以是 $b = 5, 6, 7$，分别给出 a 的 2、4 及 6 个选择：

$$b = 5 :\ a \in \{3,4\}, \qquad b = 6 :\ a \in \{2,3,4,5\}, \qquad b = 7 :\ a \in \{1,\ldots,6\}.$$

（$b = 4$ 需要 $4 \leq a \leq 3$，这是不可能的。）因此 $T_8 - T_7 = 2 + 4 + 6 = 12$，正如所求。

下图显示这 12 个数对 (a,b)：它们是位于直线 $a+b=8$ 之上或其上方，并严格位于对角线 $a=b$ 上方的点。



至于 T_7-T_6，最长的棒是 7，中间的棒 $b=4,5,6$ 分别给出 $2b-7=1,3,5$ 个选择，所以 $T_7-T_6=1+3+5=9$。因此

$$T_8-T_6=(T_8-T_7)+(T_7-T_6)=12+9=21.$$

一般的差

当 $n=2m$ 时，中间的棒取遍 $b=m+1,\ldots,2m-1$，对应有 $2b-2m=2,4,\ldots,2(m-1)$ 个选择：

$$T_{2m}-T_{2m-1}=2+4+\cdots+2(m-1)=m(m-1).$$

当 $n=2m-1$ 时，中间的棒取遍 $b=m,\ldots,2m-2$，对应有 $2b-(2m-1)=1,3,\ldots,2m-3$ 个选择，而首 $m-1$ 个奇数之和是 $(m-1)^2$。相加得

$$T_{2m}-T_{2m-2}=m(m-1)+(m-1)^2=(m-1)(2m-1).$$

（取 $m=4$，这两条式分别给出 12 和 21，与上面的数值吻合。）只要较小的下标至少是 3，这些式子便成立，即第二条公式要求 $m\geq 3$。

归纳法

设 $f(m)=\frac{1}{6}m(m-1)(4m+1)$。我们用归纳法（induction）证明对所有 $m\geq 2$，$T_{2m}=f(m)$。

基本情况（*base case*），$m=2$。从棒 $1,2,3,4$ 中选三枝，共有四种选法：$(1,2,3)$、$(1,2,4)$、$(1,3,4)$、$(2,3,4)$，除 $(1,2,4)$ 外全部满足 $a+b\geq c$。所以 $T_4=3$，而 $f(2)=\frac{1}{6}\cdot 2\cdot 1\cdot 9=3$。

归纳步骤。假设对某个 $k\geq 3$，$T_{2k-2}=f(k-1)$。由差的公式，

$$T_{2k}=f(k-1)+(k-1)(2k-1),$$

所以只需证明 $f(k)-f(k-1)=(k-1)(2k-1)$。

想一想 怎样才能整理 $f(k)-f(k-1)$，而不用把两个三次式完全展开?

两项都含有因子 $k-1$，所以先把它抽出来:

$$\begin{aligned} f(k)-f(k-1) &= \tfrac{1}{6}(k-1)\big[k(4k+1)-(k-2)(4k-3)\big] \\ &= \tfrac{1}{6}(k-1)\big[(4k^2+k)-(4k^2-11k+6)\big] \\ &= \tfrac{1}{6}(k-1)(12k-6) = (k-1)(2k-1). \end{aligned}$$

因此 $T_{2k}=f(k)$。由归纳法，对每个 $m\geq 2$，$T_{2m}=\frac{1}{6}m(m-1)(4m+1)$。

奇数枝棒

利用 $T_{2m-1}=T_{2m}-m(m-1)$,

$$T_{2m-1}=\tfrac{1}{6}m(m-1)(4m+1)-m(m-1)=\tfrac{1}{6}m(m-1)(4m-5) \qquad (m\geq 2).$$

若改把棒的数目写成 $2m+1$（以 $m+1$ 代替 m），则得

$$T_{2m+1}=\tfrac{1}{6}m(m+1)(4m-1) \qquad (m\geq 1).$$

例如 $T_3=\frac{1}{6}\cdot 1\cdot 2\cdot 3=1$，即唯一的三角形 $(1,2,3)$。

验算

- 由公式得 $T_6=\frac{1}{6}\cdot 3\cdot 2\cdot 13=13$、$T_7=\frac{1}{6}\cdot 3\cdot 4\cdot 11=22$ 及 $T_8=\frac{1}{6}\cdot 4\cdot 3\cdot 17=34$，所以 $T_8-T_7=12$ 及 $T_8-T_6=21$，与直接计算的结果相同。
- 用电脑逐一数算 $\{1,\ldots,n\}$ 中所有三元组，对 3 至 79 的每个 n，结果都与两条闭合形式（closed form）吻合。
- 设 $n=2m$，偶数情况的公式约为 $\frac{2}{3}m^3=\frac{n^3}{12}$，奇数情况的公式亦有相同表现。这是全部 $\binom{n}{3}\approx\frac{n^3}{6}$ 个可能三元组的一半，所以当 n 很大时，大约一半的三枝棒选法能组成三角形。

常见错误

- 忘记各枝棒长度互不相同，所以 $a<b$ 是严格的：容许 $a=b$ 会多算。
- 漏掉面积为零的（退化，degenerate）三角形：条件是 $a+b\geq c$，所以 $(1,2,3)$ 也要计算在内。
- 从 $m=1$ 开始归纳：T_2 没有定义，因为我们需要 $n\geq 3$。

STEP 之外

像上图那样计算一个区域内的格点（lattice points），是分拆（partitions）理论和生成函数（generating functions）的起点：T_n 有一个有理生成函数，而 n 分为偶数和奇数两种情况，反映了其分母中的因子 $1 - x^2$。像这样的结果，即计数是 n 的多项式、但该多项式取决于 n 的余数（residue），称为拟多项式（quasi-polynomial）。如欲进一步了解递推关系（recurrence）和计数论证，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）。

详细解答

STEP II 2015，第 4 题

纯数学 · 反正切函数的分支

> **STEP II 2015，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/15-s2-q4

关键思路

知道 $\tan\theta$ 的值，只能把 θ 确定至相差 π 的整数倍，因此每条方程都有无穷多个解的分支（branch），彼此之间只是沿铅直方向平移 π 的倍数。连续性（continuity）和 $x=0$ 处的值决定选取哪一个分支；而每当右方趋向 $\pm\infty$ 时，连续性亦迫使我们转换分支：那正是角度经过 $\frac{1}{2}\pi$ 的奇数倍的时刻。所以在每一部分，我们先描绘右方的图像，然后连续地追踪角度的变化。

全文中 arctan 指主值（principal value），即 $-\frac{1}{2}\pi < \arctan u < \frac{1}{2}\pi$。

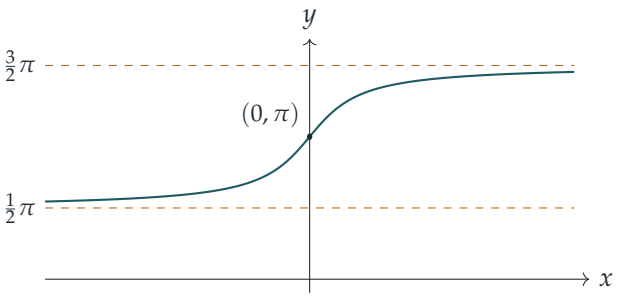
解答

第 (i) 部分

$\tan\theta = x$ 的每一个解都是 $\theta = \arctan x + n\pi$，其中 n 为整数。函数 $\arctan x$ 在整条实数线上连续，所以 $\mathrm{f}(x) - \arctan x$ 是一个只取 $\{n\pi\}$ 中的值的连续函数；连续函数不能在这些孤立的值之间跳跃，因此它是常数。由于 $\mathrm{f}(0) = \pi$ 而 $\arctan 0 = 0$，该常数为 π：

$$\mathrm{f}(x) = \pi + \arctan x.$$

所以这条曲线就是一般的 arctan 曲线向上平移 π：它严格递增，以斜率 1 通过 $(0,\pi)$（因为 $\mathrm{f}'(x) = \frac{1}{1+x^2}$），并有水平渐近线（asymptote）：当 $x \to -\infty$ 时为 $y = \frac{1}{2}\pi$，当 $x \to \infty$ 时为 $y = \frac{3}{2}\pi$。由于 $\mathrm{f}(-x) = 2\pi - \mathrm{f}(x)$，它关于 $(0,\pi)$ 具有 2 阶旋转对称（rotational symmetry）。



第 (ii) 部分

曲线 $y = \dfrac{x}{1+x^2}$。记之为 $q(x)$。它是奇函数（odd function），只在 $x = 0$ 处为零，且与 x 同号。当 $x \to \pm\infty$ 时，$q(x) \approx 1/x \to 0$，所以 x 轴是一条渐近线，在右方从上方趋近，在左方从下方趋近。求导，得

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

所以驻点（stationary point）位于 $x = \pm 1$：极大点 $(1, \frac{1}{2})$ 和极小点 $(-1, -\frac{1}{2})$。（等价地，由算术—几何平均不等式（AM–GM）得 $1 + x^2 \geq 2|x|$，故 $|q(x)| \leq \frac{1}{2}$。）在原点处的斜率为 1。

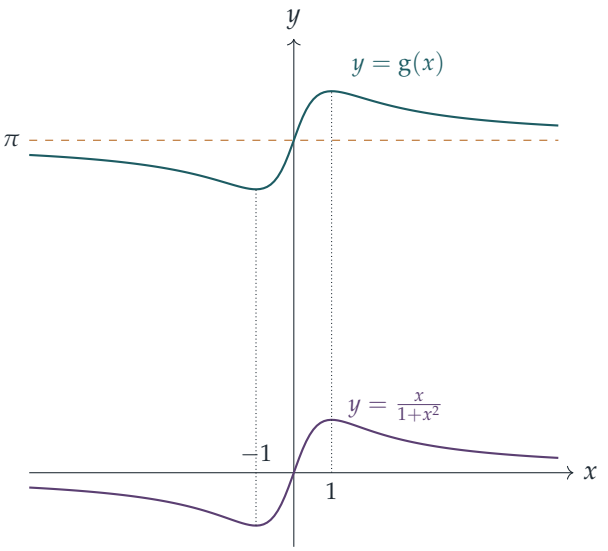
曲线 $y = \mathrm{g}(x)$。现在右方 $q(x)$ 连续且有界，所以与第 (i) 部分完全相同的论证给出

$$\mathrm{g}(x) = \pi + \arctan \frac{x}{1+x^2}.$$

由于 arctan 是递增的，g 的升降位置与 q 完全相同。因此

- 当 $x \to \pm\infty$ 时 $\mathrm{g}(x) \to \pi$（在右方从上方趋近，在左方从下方趋近）；
- g 在 $x = 1$ 处有极大值 $\pi + \arctan \frac{1}{2} \approx 3.61$，在 $x = -1$ 处有极小值 $\pi - \arctan \frac{1}{2} \approx 2.68$；
- 它以斜率 $\dfrac{q'(0)}{1+q(0)^2} = 1$ 通过 $(0, \pi)$。

它的形状就是 q 的形状稍为压扁（因为当 $u > 0$ 时 $\arctan u < u$），再向上平移 π；它从不接近 $\frac{1}{2}\pi$ 或 $\frac{3}{2}\pi$，所以从不需要转换分支。



第 (iii) 部分

曲线 $y = \dfrac{x}{1-x^2}$。记之为 $r(x)$。它是奇函数，只在 $x = 0$ 处为零，并在 $x = \pm 1$ 处有铅直渐近线。当 $x \to \pm\infty$ 时，$r(x) \approx -1/x \to 0$。它的导数是

$$r'(x) = \frac{(1-x^2)+2x^2}{(1-x^2)^2} = \frac{1+x^2}{(1-x^2)^2} > 0,$$

所以 r 在其有定义的三个区间上各自严格递增。由 x 和 $1-x^2$ 的正负号可读出：

- 在 $x < -1$ 上，r 由 0^+（当 $x \to -\infty$）上升至 $+\infty$（当 $x \to -1^-$）；
- 在 $-1 < x < 1$ 上，r 由 $-\infty$ 上升至 $+\infty$，并以斜率 1 通过原点；
- 在 $x > 1$ 上，r 由 $-\infty$（当 $x \to 1^+$）上升至 0^-（当 $x \to \infty$）。

曲线 $y = \mathrm{h}(x)$。在每个区间上分别而言，r 是连续的，所以 $\mathrm{h}(x) - \arctan r(x)$ 在该区间上是 π 的某个固定倍数，但在不同区间上这个倍数可以不同。我们逐个区间求出它。

在 $-1 < x < 1$ 上，条件 $\mathrm{h}(0) = \pi$ 给出 $\mathrm{h}(x) = \pi + \arctan r(x)$。当 $x \to 1^-$ 时 $r \to +\infty$，所以 $\mathrm{h}(x) \to \pi + \frac{1}{2}\pi = \frac{3}{2}\pi$；当 $x \to -1^+$ 时 $r \to -\infty$，而 $\mathrm{h}(x) \to \frac{1}{2}\pi$。在 $x = \pm 1$ 处的连续性迫使 $\mathrm{h}(1) = \frac{3}{2}\pi$ 及 $\mathrm{h}(-1) = \frac{1}{2}\pi$。

> **想一想** 在 $x = 1$ 的右边紧贴处，$r(x)$ 是很大的负数，所以 $\arctan r(x)$ 接近 $-\frac{1}{2}\pi$。我们要加上 π 的哪一个倍数，才能使 h 在 $x = 1$ 处连续?

当 $x > 1$ 时有 $\mathrm{h}(x) = n\pi + \arctan r(x)$，而当 $x \to 1^+$ 时它趋向 $n\pi - \frac{1}{2}\pi$。要使它等于 $\frac{3}{2}\pi$，须有 $n = 2$。同理，当 $x < -1$ 时有 $\mathrm{h}(x) = m\pi + \arctan r(x) \to m\pi + \frac{1}{2}\pi$（当

$x \to -1^-$），要与 $\frac{1}{2}\pi$ 吻合须有 $m = 0$。因此

$$\mathrm{h}(x) = \begin{cases} \arctan\dfrac{x}{1-x^2}, & x < -1, \\ \pi + \arctan\dfrac{x}{1-x^2}, & -1 < x < 1, \\ 2\pi + \arctan\dfrac{x}{1-x^2}, & x > 1, \end{cases} \qquad \mathrm{h}(\pm 1) = \pi \pm \tfrac{1}{2}\pi.$$

所以当 r 急升至 $+\infty$ 再从 $-\infty$ 重新出现时，角度 h 只是继续递增，越过 $\frac{3}{2}\pi$（在 $x = -1$ 处同样越过 $\frac{1}{2}\pi$）。

现在看草图的各项特征。

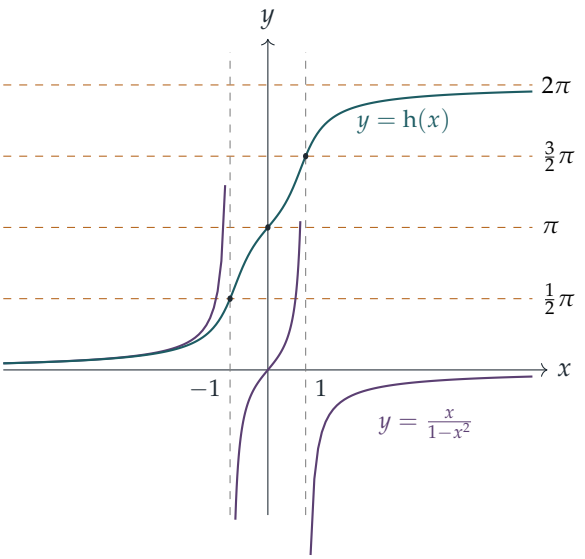
- 渐近线。当 $x \to -\infty$ 时，$r \to 0^+$，所以 $\mathrm{h}(x) \to 0$，从上方趋近；当 $x \to \infty$ 时，$r \to 0^-$，所以 $\mathrm{h}(x) \to 2\pi$，从下方趋近。
- 单调（*monotonic*）。在 ± 1 以外，由链式法则（chain rule），

$$\mathrm{h}'(x) = \frac{r'(x)}{1 + r(x)^2} = \frac{1 + x^2}{(1-x^2)^2 + x^2} = \frac{1 + x^2}{x^4 - x^2 + 1} > 0.$$

  （分母等于 $(x^2 - \frac{1}{2})^2 + \frac{3}{4} > 0$。）这个表达式在 $x = \pm 1$ 处完全没有问题，其值为 2，所以 h 在该处是光滑的：曲线以斜率 2 通过 $(1, \frac{3}{2}\pi)$ 和 $(-1, \frac{1}{2}\pi)$，并没有尖角。
- 关键点。$(0, \pi)$，斜率 1；$(\pm 1, \pi \pm \frac{1}{2}\pi)$，斜率 2。
- 对称性。$\mathrm{h}(-x) = 2\pi - \mathrm{h}(x)$，所以曲线关于 $(0, \pi)$ 具有旋转对称。

所以 h 由 0 稳定地递增至 2π，是一条与第 (i) 部分相似的 S 形曲线，但覆盖的高度是其两倍。在草图中，r 的渐近线 $x = \pm 1$ 恰好对准 h 越过 $\frac{1}{2}\pi$ 和 $\frac{3}{2}\pi$ 的点。

有一个简洁的方法可把三种情况一并写出：$\mathrm{h}(x) - \pi$ 是点 $(1 - x^2, x)$ 在 $(-\pi, \pi)$ 内的极角（polar angle）。该点从不到达原点或负水平轴（当 $1 - x^2 \le 0$ 时有 $x \ne 0$），这正是角度连续变化的原因。



验算

- 每条公式都满足定义方程：例如 $\tan(2\pi+\theta)=\tan\theta$，而且在 0 处的值全都是 π。
- 代入个别数值：$g(1)=\pi+\arctan\frac{1}{2}\approx 3.605$，以及 $h(2)=2\pi+\arctan(-\frac{2}{3})\approx 5.695$，正如草图所要求，它介乎 $\frac{3}{2}\pi\approx 4.712$ 与 $2\pi\approx 6.283$ 之间。
- h 的连续性：$h(0.999)\approx 4.710$ 及 $h(1.001)\approx 4.714$，分处 $\frac{3}{2}\pi\approx 4.712$ 两侧，与 $x=1$ 处斜率为 2 一致。以电脑在 $[-10,10]$ 上逐步检查，确认 f、g、h 都没有任何跳跃。

常见错误

- 对所有 x 都写 $h(x)=\pi+\arctan\frac{x}{1-x^2}$：这在 $x=1$ 处向下跳 π，在 $x=-1$ 处向上跳 π，所以并不连续。
- 只因 $\frac{x}{1-x^2}$ 在 $x=\pm1$ 处有铅直渐近线，便在 $y=h(x)$ 的图像中也画上；角度本身保持有限，只是它的正切值趋向无穷。
- 在 $y=h(x)$ 的 $x=\pm1$ 处画出尖角；该处左右两侧的斜率都是 2。

STEP 之外

我们所做的，是为一个多值反函数（many-valued inverse function）选取一个连续的分支；而“每当正切值经过无穷大便加上 π”这条规则，让我们初尝把路径提升（lifting）到覆

叠空间（covering space）的滋味。同一想法应用于复对数（complex logarithm）和 $\arg z$，正是复分析（complex analysis）中环绕数（winding number）和辐角原理（argument principle）的基础。Needham 的 *Visual Complex Analysis*（《直观复分析》）以几何方式发展这套理论，读来趣味盎然。

详细解答

STEP II 2015，第 5 题

纯数学 · 反正切级数

> **STEP II 2015，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/15-s2-q5

关键思路

正切和角公式（tangent addition formula）能把若干个反正切（arctangent）之和化为单一分式，因此关于 $\tan S_n$ 的归纳法（induction）证明只需细心处理代数运算。微妙之处在于由 $\tan S_n$ 推出 S_n 本身这一步：知道一个角的正切值，只能确定该角至相差 π 的整数倍，所以我们必须确定 S_n 落在哪个区间。在第 (ii) 部分，角 α_n 原来正是第 (i) 部分中的那些反正切，而无穷级数（infinite sum）就是 S_n 的极限（limit）。

解答

第 (i) 部分

设 $\theta_m = \arctan\left(\frac{1}{2m^2}\right)$，于是 $S_n = \theta_1 + \cdots + \theta_n$，且 $\tan\theta_m = \frac{1}{2m^2}$。

$\tan S_n$ 的公式。 当 $n = 1$ 时，$\tan S_1 = \frac{1}{2} = \frac{1}{1+1}$，所以结果成立。

假设对某个 $n \geq 1$ 有 $\tan S_n = \frac{n}{n+1}$。由于 $S_{n+1} = S_n + \theta_{n+1}$，和角公式给出

$$\tan S_{n+1} = \frac{\tan S_n + \tan\theta_{n+1}}{1 - \tan S_n \tan\theta_{n+1}} = \frac{\dfrac{n}{n+1} + \dfrac{1}{2(n+1)^2}}{1 - \dfrac{n}{2(n+1)^3}}.$$

分母不为零，因为乘积 $\frac{n}{n+1} \cdot \frac{1}{2(n+1)^2}$ 小于 1。将分子和分母同乘以 $2(n+1)^3$，得

$$\tan S_{n+1} = \frac{(n+1)\big(2n(n+1)+1\big)}{2(n+1)^3 - n} = \frac{(n+1)(2n^2+2n+1)}{2n^3+6n^2+5n+2}.$$

想一想 我们怎样知道分母能够漂亮地因式分解?

我们的目标是 $\frac{n+1}{n+2}$，所以分母理应是 $(n+2)(2n^2+2n+1)$。展开即可证实：$(n+2)(2n^2+2n+1) = 2n^3+6n^2+5n+2$。约去正因式 $2n^2+2n+1$，得

$$\tan S_{n+1} = \frac{n+1}{n+2},$$

这正是 $n+1$ 时的结果。由归纳法，对所有 $n \geq 1$，$\tan S_n = \frac{n}{n+1}$。

S_n 的公式。 单凭 $\tan S_n = \frac{n}{n+1}$，我们只能说对某个整数 k，$S_n = \arctan\frac{n}{n+1} + k\pi$。为了证明 $k=0$，我们再次用归纳法证明

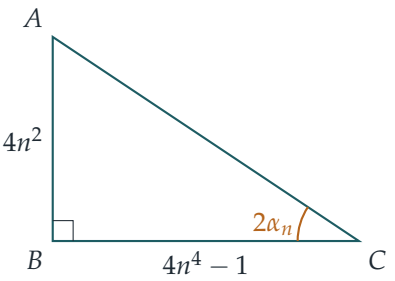
$$S_n = \arctan\frac{n}{n+1}.$$

当 $n=1$ 时，$S_1 = \arctan\frac{1}{2}$，符合要求。假设结果对某个 n 成立。由于 $0 < \frac{n}{n+1} < 1$，我们有 $0 < S_n < \arctan 1 = \frac{1}{4}\pi$。又 $0 < \frac{1}{2(n+1)^2} < 1$，所以 $0 < \theta_{n+1} < \frac{1}{4}\pi$。相加得

$$0 < S_{n+1} = S_n + \theta_{n+1} < \tfrac{1}{2}\pi.$$

因此 S_{n+1} 位于 $[0, \frac{1}{2}\pi)$，这恰好是 arctan 在非负数上的值域；而由第一个结果，$\tan S_{n+1} = \frac{n+1}{n+2}$。由于 tan 在此区间上是一一对应的（one-to-one），故 $S_{n+1} = \arctan\frac{n+1}{n+2}$。归纳法至此完成。

第 (ii) 部分



由于 B 处的角是直角，角 C 的对边是 AB，邻边是 BC（注意当 $n \geq 1$ 时 $4n^4-1 > 0$）。所以

$$\tan 2\alpha_n = \frac{AB}{BC} = \frac{4n^2}{4n^4-1}.$$

第 (i) 部分提示我们试试 $\tan\alpha_n = \frac{1}{2n^2}$。设 $t = \frac{1}{2n^2}$，倍角公式（double-angle formula）给出

$$\frac{2t}{1-t^2} = \frac{\frac{1}{n^2}}{1-\frac{1}{4n^4}} = \frac{4n^2}{4n^4-1},$$

两者吻合。为了确保这是正确的那个值，而不仅仅是其中一个可能的值，设 $t = \tan\alpha_n$。方程 $\frac{2t}{1-t^2} = \frac{4n^2}{4n^4-1}$ 可整理为二次方程（quadratic equation）

$$4n^2t^2 + 2(4n^4 - 1)t - 4n^2 = 0,$$

其两根之积为 -1。其中一根是 $\frac{1}{2n^2}$，所以另一根是 $-2n^2 < 0$。但 $2\alpha_n$ 是直角三角形中直角以外的一个角，所以 $0 < 2\alpha_n < \frac{1}{2}\pi$，从而 $0 < \alpha_n < \frac{1}{4}\pi$，且 $t > 0$。因此 $\tan\alpha_n = \frac{1}{2n^2}$；又由于 α_n 位于 arctan 的值域内，

$$\alpha_n = \arctan\frac{1}{2n^2}.$$

想一想 这里的无穷级数是甚么意思？第 (i) 部分又怎样帮上忙？

无穷级数就是其部分和（partial sums）的极限，而由第 (i) 部分，这些部分和是

$$\sum_{n=1}^{N}\alpha_n = S_N = \arctan\frac{N}{N+1}.$$

当 $N \to \infty$ 时，$\frac{N}{N+1} \to 1$；又由于 arctan 是连续（continuous）函数，$\arctan\frac{N}{N+1} \to \arctan 1 = \frac{1}{4}\pi$。因此

$$\sum_{n=1}^{\infty}\alpha_n = \tfrac{1}{4}\pi,$$

符合要求。

验算

- 当 $n = 1$ 时，第 (ii) 部分的三角形就是我们熟悉的 3–4–5 三角形，其中 $\tan 2\alpha_1 = \frac{4}{3}$。而 $\tan\alpha_1 = \frac{1}{2}$ 满足 $\frac{2\cdot\frac{1}{2}}{1-\frac{1}{4}} = \frac{4}{3}$。
- $S_2 = \arctan\frac{1}{2} + \arctan\frac{1}{8}$，而 $\frac{\frac{1}{2}+\frac{1}{8}}{1-\frac{1}{16}} = \frac{10}{15} = \frac{2}{3}$，与公式的预测一致。
- 从数值上看，把首 2000 项相加得 $S_{2000} \approx 0.785148$，而 $\frac{1}{4}\pi \approx 0.785398$。两者相差约 $\frac{1}{4000}$，这正合预期：由差角公式，恰有 $\frac{1}{4}\pi - \arctan\frac{N}{N+1} = \arctan\frac{1}{2N+1}$。

常见错误

- 由 $\tan S_n = \frac{n}{n+1}$ 直接得出 $S_n = \arctan\frac{n}{n+1}$，而没有证明 S_n 位于 $[0, \frac{1}{2}\pi)$。
- 凭侥幸猜中 $\tan\alpha_n = \frac{1}{2n^2}$，却没有排除二次方程的另一个根。
- 把无穷级数当作 S_∞，却没有说明它是 S_N 的极限。

STEP 之外

这其实是一个经过伪装的伸缩级数（telescoping series）：由于 $\frac{1}{2m^2} = \frac{(2m+1)-(2m-1)}{1+(2m+1)(2m-1)}$，每一项都等于 $\arctan(2m+1) - \arctan(2m-1)$，于是总和坍缩为 $\arctan(2n+1) - \frac{1}{4}\pi$，而这等于 $\arctan\frac{n}{n+1}$。把无穷级数定义为部分和的极限，并懂得何时可以让极限穿过连续函数，正是大学第一门数学分析（analysis）课程的起点。想进一步阅读，Abbott 的 *Understanding Analysis*（《理解分析》）是个好起点。

详细解答

STEP II 2015，第 6 题

纯数学 • 利用对称性积分

STEP II 2015，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/15-s2-q6

关键思路

把 $\sin x$ 写成 $\cos\left(\frac{1}{2}\pi - x\right)$，就能把 $1+\sin x$ 变成二倍角（double-angle）表达式，接着 $\sec^2$ 便可直接积分。代换（substitution）$y = \pi - x$ 把 $[0, \pi]$ 关于其中点作反射（reflection），而 $\sin x$ 对这个反射毫无反应。因此，$\mathrm{f}(\sin x)$ 前面的任何多项式（polynomial）因子，都可以换成它反射后的版本。在第 (iii) 部分，那个三次式（cubic）是特意挑选的，令这种交换一下子消去所有 x 的幂。

解答

第 (i) 部分

由于 $\sin x = \cos\left(\frac{1}{2}\pi - x\right)$ 及 $1+\cos 2\theta = 2\cos^2\theta$，取 $\theta = \frac{1}{4}\pi - \frac{1}{2}x$，得

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

取倒数再乘以 2，得 $\sec^2\left(\frac{1}{4}\pi - \frac{1}{2}x\right) = \dfrac{2}{1+\sin x}$，此即所求。（两边在相同的点上都没有定义，即 $\sin x = -1$ 之处。）

因此

$$\int \frac{1}{1+\sin x}\,\mathrm{d}x = \frac{1}{2}\int \sec^2\left(\tfrac{1}{4}\pi - \tfrac{1}{2}x\right)\mathrm{d}x = -\tan\left(\tfrac{1}{4}\pi - \tfrac{1}{2}x\right) + C,$$

因为根据连锁法则（chain rule），对 $\tan\left(\frac{1}{4}\pi - \frac{1}{2}x\right)$ 求导会产生一个因子 $-\frac{1}{2}$。

第 (ii) 部分

设 $I = \int_0^\pi x\,\mathrm{f}(\sin x)\,\mathrm{d}x$。令 $y = \pi - x$，于是 $\mathrm{d}x = -\mathrm{d}y$，上下限 $x = 0, \pi$ 变成 $y = \pi, 0$，而且 $\sin x = \sin(\pi - y) = \sin y$。于是

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

因为 y 只是哑变量（dummy variable）。所以 $2I = \pi \int_0^\pi \mathrm{f}(\sin x)\,\mathrm{d}x$，即所求结果。

想一想 未知的积分 I 在等式两边都出现了。这是个问题吗?

不是，这正是我们想要的。代换本身并没有求出 I 的值，而是给出一条关于 I 的线性方程（linear equation），我们解这条方程便可。

现在取 $\mathrm{f}(t) = \dfrac{1}{1+t}$，它在 $t \in [0,1]$ 上连续（continuous），而这正是 $\sin x$ 在 $[0,\pi]$ 上的取值范围。由第 (i) 部分，

$$\int_0^\pi \frac{1}{1+\sin x}\,\mathrm{d}x = \Big[-\tan\left(\tfrac{1}{4}\pi - \tfrac{1}{2}x\right) \Big]_0^\pi = -\tan\left(-\tfrac{1}{4}\pi\right) + \tan\tfrac{1}{4}\pi = 1 + 1 = 2,$$

而且 $\tan\left(\frac{1}{4}\pi - \frac{1}{2}x\right)$ 在 $[0,\pi]$ 上连续，因为自变量一直留在 $[-\frac{1}{4}\pi, \frac{1}{4}\pi]$ 之内。因此

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

第 (iii) 部分

使用同样的反射。设 $g(x) = 2x^3 - 3\pi x^2$。展开，

$$\begin{aligned} g(\pi - x) &= 2(\pi^3 - 3\pi^2 x + 3\pi x^2 - x^3) - 3\pi(\pi^2 - 2\pi x + x^2) \\ &= -\pi^3 + 3\pi x^2 - 2x^3 = -\pi^3 - g(x). \end{aligned}$$

设 $J = \int_0^\pi \dfrac{g(x)}{(1+\sin x)^2}\,\mathrm{d}x$。完全仿照第 (ii) 部分作代换 $y = \pi - x$，得

$$J = \int_0^\pi \frac{g(\pi - y)}{(1+\sin y)^2}\,\mathrm{d}y = -\pi^3 K - J, \qquad \text{其中} K = \int_0^\pi \frac{\mathrm{d}x}{(1+\sin x)^2}.$$

所以 $J = -\frac{1}{2}\pi^3 K$。

想一想 为甚么这个特定的三次式这么好用?

以 $u = x - \frac{1}{2}\pi$（即与中点的距离）表示，$g(x) = 2u^3 - \frac{3}{2}\pi^2 u - \frac{1}{2}\pi^3$。由于 $\sin x = \cos u$，函数 $1/(1+\sin x)^2$ 是 u 的偶函数（even function），所以奇次项 $2u^3$ 和 $-\frac{3}{2}\pi^2 u$ 在对称区间上的积分为零。剩下的只有常数 $-\frac{1}{2}\pi^3$，正是我们得到的结果。第 (ii) 部分其实是同一回事，只是取 $x = u + \frac{1}{2}\pi$。

要求 K，把第 (i) 部分的恒等式平方：$\dfrac{1}{(1+\sin x)^2} = \frac{1}{4}\sec^4\left(\frac{1}{4}\pi - \frac{1}{2}x\right)$。代入 $u = \frac{1}{4}\pi - \frac{1}{2}x$，于是 $\mathrm{d}x = -2\,\mathrm{d}u$，而 $x = 0, \pi$ 变成 $u = \frac{1}{4}\pi, -\frac{1}{4}\pi$。写成 $\sec^4 u = (1+\tan^2 u)\sec^2 u$，其积分为 $\tan u + \frac{1}{3}\tan^3 u$，于是

$$\begin{aligned} K &= \int_{\pi/4}^{-\pi/4} \tfrac{1}{4}\sec^4 u\,(-2\,\mathrm{d}u) = \frac{1}{2}\int_{-\pi/4}^{\pi/4} (1+\tan^2 u)\sec^2 u\,\mathrm{d}u \\ &= \frac{1}{2}\Big[\tan u + \tfrac{1}{3}\tan^3 u\Big]_{-\pi/4}^{\pi/4} = \frac{1}{2}\left(\tfrac{4}{3} + \tfrac{4}{3}\right) = \frac{4}{3}. \end{aligned}$$

因此

$$\int_0^{\pi} \frac{2x^3 - 3\pi x^2}{(1+\sin x)^2}\,\mathrm{d}x = -\frac{\pi^3}{2}\cdot\frac{4}{3} = -\frac{2\pi^3}{3}.$$

验算

- 对 $-\tan\left(\frac{1}{4}\pi - \frac{1}{2}x\right)$ 求导，得 $\frac{1}{2}\sec^2\left(\frac{1}{4}\pi - \frac{1}{2}x\right) = 1/(1+\sin x)$。
- 正负号：第 (ii) 部分的被积函数为正，而 $\pi \approx 3.14$ 是合理的，因为由 $\frac{1}{2} \le 1/(1+\sin x) \le 1$ 可得上下界 $\frac{1}{4}\pi^2 \approx 2.47$ 及 $\frac{1}{2}\pi^2 \approx 4.93$。第 (iii) 部分中，$2x^3 - 3\pi x^2 = x^2(2x - 3\pi) < 0$ 在 $(0, \pi]$ 上成立，所以答案为负是对的。
- 数值积分（numerical quadrature）给出 $\int_0^{\pi}(1+\sin x)^{-2}\,\mathrm{d}x \approx 1.33333$，而第 (iii) 部分的积分 ≈ -20.6709，分别与 $\frac{4}{3}$ 及 $-\frac{2}{3}\pi^3$ 吻合。

常见错误

- 在第 (i) 部分漏掉连锁法则带来的因子 $-\frac{1}{2}$，结果得出 $+2\tan(\dots)$ 或 $-\frac{1}{2}\tan(\dots)$。
- 忘记对调上下限会抵消 $\mathrm{d}x = -\mathrm{d}y$ 带来的负号。
- 在第 (iii) 部分尝试连续三次分部积分（integration by parts）。这样做是可行的，但反射只需一行就能完成。

STEP 之外

第 (ii) 部分的技巧运用了对称性（symmetry）。如果一个权函数（weight function）在某个反射下保持不变，那么被积函数中只有关于反射中心为偶的部分才有贡献。你会在大

学一年级数学中处处遇到这个想法：在傅里叶级数（Fourier series）中，奇函数没有余弦项；在概率中，关于 μ 对称的分布（distribution），其平均值（mean）就在 μ。对于任何 $\sin x$ 与 $\cos x$ 的有理函数（rational function），代换 $t = \tan\frac{1}{2}x$ 是一个有系统的替代方法。想找更多这类积分，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）。

详细解答

STEP II 2015，第 7 题

纯数学 · 相交圆

> **STEP II 2015，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/15-s2-q7

关键思路

当两个圆相交于两点时，通过这两点的直线（即公共弦（common chord））垂直于连心线（line of centres）。圆 D 平分（bisects）C，当且仅当这条公共弦是 C 的直径（diameter），即公共弦通过 C 的圆心。由两个圆心及其中一个交点组成一个直角三角形，对它应用毕氏定理（Pythagoras′ theorem），便可把“平分”化为一条方程：（圆心之间的距离）$^2 = r^2 - a^2$。

解答

第 (i) 部分

设 D 的圆心为 Q、半径为 r，并设 O 为原点。

*一般条件。*假设 D 平分 C，与 C 相交于 P 及 $P' = -P$，其中 $|P| = a$。两点都在 D 上，所以 $QP = QP' = r$。于是 Q 与 P、P' 等距，因此 Q 位于 PP' 的垂直平分线（perpendicular bisector）上，而这条垂直平分线通过 O（即 PP' 的中点）。因此 $OQ \perp OP$，对三角形 OQP 应用毕氏定理，得

$$r^2 = OQ^2 + a^2. \qquad (*)$$

反过来，若 D 是任何满足 $OQ^2 = r^2 - a^2$ 及 $OQ > 0$ 的圆，在 C 上取一点 P 使 $OP \perp OQ$。那么 $QP^2 = OQ^2 + a^2 = r^2$，对 $-P$ 亦同理，所以 D 通过 C 上两个径向相对（diametrically opposite）的点。两个不同的圆最多相交于两点，所以这两点就是仅有的交点，因此 D 平分 C。

*当 $r > a$ 时的存在性。*把圆心放在 x 轴上，与 O 相距 $\sqrt{r^2 - a^2} > 0$：

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

直接验证：从这条方程减去 $x^2+y^2=a^2$，得 $-2x\sqrt{r^2-a^2}+(r^2-a^2)=r^2-a^2$，所以 $x=0$（因为 $r\neq a$），于是 $y=\pm a$。两圆恰好相交于 $(0,a)$ 及 $(0,-a)$，而这两点在 C 上径向相对。所以这个半径为 r 的圆平分 C。

*当 $r\leq a$ 时的不可能性。*若某个半径为 r 的圆平分 C，则由 $(*)$ 得 $r^2=OQ^2+a^2\geq a^2$。所以 $r<a$ 是不可能的。若 $r=a$，则 $(*)$ 迫使 $OQ=0$，所以 D 就是 C 本身，它与 C 相交于无限多点，而不是恰好两点。因此没有半径 $r\leq a$ 的圆能平分 C。

第 (ii) 部分

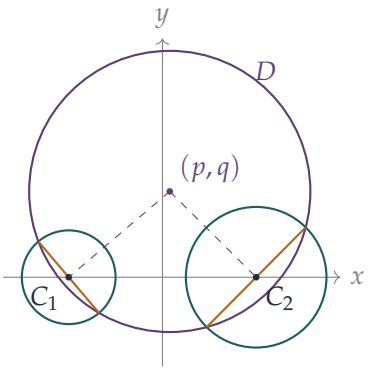
设 D 的圆心为 (p,q)。对每个圆应用 $(*)$（把原点平移至相应的圆心），

$$(p+d)^2+q^2=r^2-a_1^2,$$
$$(p-d)^2+q^2=r^2-a_2^2.$$

以第一式减去第二式，q^2 和 p^2 项互相抵消，剩下 $4pd=a_2^2-a_1^2$，所以

$$p=\frac{a_2^2-a_1^2}{4d},$$

即为所求。



在草图中（$d=2$，$a_1=1$，$a_2=1.5$，$r=3$），每条公共弦（琥珀色）都是 C_1 或 C_2 的直径，并垂直于连接圆心的虚线。

为求 q，把 p 代回第一条方程。首先，

$$p+d=\frac{4d^2+a_2^2-a_1^2}{4d},$$

所以

$$q^2=r^2-a_1^2-\frac{(4d^2+a_2^2-a_1^2)^2}{16d^2}=\frac{16d^2r^2-16d^2a_1^2-(4d^2+a_2^2-a_1^2)^2}{16d^2}.$$

分子中被减去的项展开为

$$\begin{aligned}16d^2a_1^2+(4d^2+a_2^2-a_1^2)^2 &= 16d^4+8d^2(a_2^2-a_1^2)+16d^2a_1^2+(a_2^2-a_1^2)^2\\ &= 16d^4+8d^2(a_1^2+a_2^2)+(a_2^2-a_1^2)^2\\ &= \big(4d^2+(a_2-a_1)^2\big)\big(4d^2+(a_2+a_1)^2\big),\end{aligned}$$

最后一行成立，是因为 $(a_2-a_1)^2+(a_2+a_1)^2=2(a_1^2+a_2^2)$ 及 $(a_2-a_1)^2(a_2+a_1)^2=(a_2^2-a_1^2)^2$。因此

$$q=\pm\frac{\sqrt{16r^2d^2-\big(4d^2+(a_2-a_1)^2\big)\big(4d^2+(a_2+a_1)^2\big)}}{4d}.$$

> **想一想** 不等式（inequality）从何而来？到目前为止，似乎没有任何东西像是不等式。

圆心 (p,q) 是实点，所以 $q^2\geq 0$。由于 $16d^2>0$，这表示

$$16r^2d^2\geq\big(4d^2+(a_2-a_1)^2\big)\big(4d^2+(a_2+a_1)^2\big),$$

这正是所要求关于 r 的条件。

我们还应检查第 (i) 部分的要求 $r>a_1$（及 $r>a_2$）并不是额外的条件。由于 $d>a_1$，我们有 $4d^2+a_2^2-a_1^2>0$，所以 $p+d>0$；于是 $r^2-a_1^2=(p+d)^2+q^2>0$。同理，因为 $d>a_2$，所以 $d-p>0$，因此 $r^2-a_2^2=(p-d)^2+q^2>0$。所以求得的圆 D 确实同时平分 C_1 和 C_2。

验算

- 对称情况 $a_1=a_2=a$：此时 $p=0$，正如关于 y 轴的对称性所要求的；条件变为 $16r^2d^2\geq 4d^2(4d^2+4a^2)$，即 $r^2\geq d^2+a^2$。这恰好是圆心在原点时的 $(*)$，而原点正是同时最接近 $(\pm d,0)$ 的可能圆心。
- 在草图的例子中，$p=0.15625$，$q\approx 1.830$，而 D 与 C_1 的交点约为 $(-1.353,-0.762)$ 及 $(-2.647,0.762)$，其中点为 $(-2,0)$，即 C_1 的圆心。在该例中 r 的最小可能值约为 2.377。
- 对调 C_1 和 C_2（$a_1\leftrightarrow a_2$，$d\to -d$），p 的形式不变，不等式亦不变，理应如此。

常见错误

- 在第 (i) 部分忘记 $r=a$ 的情况：$(*)$ 容许 $OQ=0$，你必须说明为何 $D=C$ 不算数。

- 把条件写成 $OQ^2 = r^2 + a^2$，这是把平分与直角相交（正交圆（orthogonal circles））混淆了。画出那个三角形：直角位于 C 的圆心，所以斜边（hypotenuse）是 D 的半径 r。
- 只给出 $q = +\sqrt{\cdots}$：关于 x 轴反射会得到第二个圆。

STEP 之外

对于点 (p, q) 及圆心为 (x_0, y_0)、半径为 r 的圆，量 $(p - x_0)^2 + (q - y_0)^2 - r^2$ 称为该点关于该圆的幂（power）；而关于两个圆具有相等幂的点的集合是一条直线，称为根轴（radical axis）；像我们那样把两条圆方程相减，便能求出它。在本题中，平分条件表示 C 的圆心关于 D 的幂等于 $-a^2$。这些概念，以及相关的圆反演（inversion），在 Needham 的 *Visual Complex Analysis*（《直观复分析》）中有精彩的阐述；而许多 STEP 风格的应用可见于 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）。

详细解答

STEP II 2015，第 8 题

纯数学 · 向量与相切圆

STEP II 2015，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/15-s2-q8

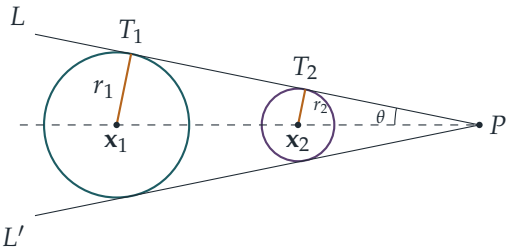
关键思路

每个圆心与两条切线（tangent）的距离相同（即其自身的半径），所以两个圆心都位于平分 L 与 L' 之间夹角的直线上，即角平分线（angle bisector）上。沿着这条直线，与 P 的距离和半径成正比，由此立即得出 (i) 的公式。至于 (ii)，把公式除以 $r_1 r_2$ 后会变得对称得多：以 $\mathbf{x}_i/r_i$ 和 $1/r_i$ 表示时，三个焦点（focus）满足一条系数之和为零的线性关系，而这正正就是共线（collinear）的意思。

解答

第 (i) 部分

设 θ 为 L 与 L' 在 P 处夹角的一半，取的是包含两个圆的那个角。



圆的切线与圆心的垂直距离等于半径。所以 $\mathbf{x}_1$ 与 L 和 L' 的距离都是 r_1，而与两条相交直线等距的点，必位于 P 处两条角平分线的其中一条上。由于两个圆位于 L 的同一侧，亦位于 L' 的同一侧，两个圆心都在 L 与 L' 分割出的四个角区域中的同一个区域内，所以它们位于从 P 出发的同一条半直线上：即该区域的角平分线。设 $\mathbf{u}$ 为沿这条半直线、由

P 指向两个圆的单位向量（unit vector）。

设 T_1 为 L 与 C_1 相切的点，则三角形 PT_1X_1（其中 X_1 为圆心）在 T_1 处为直角，所以 $r_1 = PX_1 \sin\theta$。同理 $r_2 = PX_2 \sin\theta$。因此

$$\mathbf{x}_1 - \mathbf{p} = \frac{r_1}{\sin\theta}\,\mathbf{u}, \qquad \mathbf{x}_2 - \mathbf{p} = \frac{r_2}{\sin\theta}\,\mathbf{u},$$

其中 $\mathbf{p}$ 为 P 的位置向量（position vector）。消去 $\mathbf{u}/\sin\theta$，得

$$r_2(\mathbf{x}_1 - \mathbf{p}) = r_1(\mathbf{x}_2 - \mathbf{p}) \quad \Longrightarrow \quad (r_1 - r_2)\,\mathbf{p} = r_1\mathbf{x}_2 - r_2\mathbf{x}_1.$$

两个圆大小不同，所以 $r_1 \neq r_2$，可以相除：

$$\mathbf{p} = \frac{r_1\mathbf{x}_2 - r_2\mathbf{x}_1}{r_1 - r_2},$$

即为所求。（若半径相等，L 与 L' 便会平行，根本没有焦点。）

第 (ii) 部分

> **想一想** $\mathbf{p}$ 的公式并不具有明显的对称性。能否把它改写，使三个焦点的形式看起来一样?

把 $(r_1 - r_2)\mathbf{p} = r_1\mathbf{x}_2 - r_2\mathbf{x}_1$ 除以 r_1r_2。记 $s_i = 1/r_i$ 及 $\mathbf{y}_i = \mathbf{x}_i/r_i$，式子变成

$$(s_2 - s_1)\,\mathbf{p} = \mathbf{y}_2 - \mathbf{y}_1.$$

把第 (i) 部分应用于另外两对圆（这是容许的，因为 $r_3 \neq r_1$、$r_3 \neq r_2$，而且各圆互不重叠），并以 $\mathbf{q}$ 和 $\mathbf{r}$ 表示 Q 和 R 的位置向量，得

$$(s_3 - s_1)\,\mathbf{q} = \mathbf{y}_3 - \mathbf{y}_1, \qquad (s_3 - s_2)\,\mathbf{r} = \mathbf{y}_3 - \mathbf{y}_2.$$

右方满足 $(\mathbf{y}_2 - \mathbf{y}_1) + (\mathbf{y}_3 - \mathbf{y}_2) = \mathbf{y}_3 - \mathbf{y}_1$，所以

$$(s_2 - s_1)\,\mathbf{p} + (s_3 - s_2)\,\mathbf{r} = (s_3 - s_1)\,\mathbf{q}.$$

由于 $s_3 \neq s_1$，可写成

$$\mathbf{q} = \lambda\,\mathbf{p} + (1 - \lambda)\,\mathbf{r}, \qquad \text{其中} \lambda = \frac{s_2 - s_1}{s_3 - s_1},$$

因为两个系数 $\frac{s_2-s_1}{s_3-s_1}$ 与 $\frac{s_3-s_2}{s_3-s_1}$ 之和为 1。所以 $\mathbf{q} = \mathbf{r} + \lambda(\mathbf{p} - \mathbf{r})$，这是经过 R 和 P 的直线上的一点。因此 P、Q 和 R 共线。（若 P 与 R 刚好重合，则同样得出 $Q = P = R$，命题显然成立。）

第 (iii) 部分

由第 (ii) 部分，$\mathbf{q} = \lambda\mathbf{p} + (1-\lambda)\mathbf{r}$。假设 $P \neq R$，则点 $\lambda\mathbf{p} + (1-\lambda)\mathbf{r}$ 是中点 $\frac{1}{2}(\mathbf{p}+\mathbf{r})$ 若且唯若 $\lambda = \frac{1}{2}$（不同的 λ 值对应直线上不同的点）。所以条件为

$$\frac{s_2 - s_1}{s_3 - s_1} = \frac{1}{2} \iff 2s_2 = s_1 + s_3 \iff \frac{2}{r_2} = \frac{1}{r_1} + \frac{1}{r_3} \iff r_2 = \frac{2r_1 r_3}{r_1 + r_3}.$$

换言之，r_2 必须是 r_1 与 r_3 的调和平均数（harmonic mean）。（在退化情况 $P = R$ 下，三个焦点重合，无论半径为何，Q 都显然位于中点。）

验算

- 第 (i) 部分：令 $r_2 \to 0$，C_2 缩成一点，公式给出 $\mathbf{p} = \mathbf{x}_2$：从一点引出的切线当然经过该点。另外，$\mathbf{p}$ 等于 $\mathbf{x}_2 + \frac{r_2}{r_1 - r_2}(\mathbf{x}_2 - \mathbf{x}_1)$，当 $r_2 < r_1$ 时位于较小的圆之外，与图中一致。
- (iii) 的一个数值例子：取圆心 $(0,0)$、$(10,1)$、$(3,8)$，半径分别为 1、$\frac{3}{2}$、3，使得 $\frac{2}{3/2} = \frac{4}{3} = 1 + \frac{1}{3}$。于是 $P = (-20,-2)$、$R = (17,-6)$、$Q = (-\frac{3}{2},-4)$，而 Q 确实是 PR 的中点。
- 公式亦经电脑验证：对许多随机产生的圆对，直接求两条切线的交点，都得出相同的点 P。

常见错误

- 没有解释为何 P 位于连心线上，或为何 P 位于连接两圆心的线段之外（承载两个圆心的，是从 P 出发的同一条半直线）。
- 只显示 $\mathbf{q} - \mathbf{p}$ 与 $\mathbf{r} - \mathbf{p}$ 平行，却没有检查关系式的系数之和为零；关系式 $\alpha\mathbf{p} + \beta\mathbf{q} + \gamma\mathbf{r} = \mathbf{0}$ 只有在 $\alpha + \beta + \gamma = 0$ 时才蕴含共线。

STEP 之外

焦点 P 是把 C_1 映射到 C_2、比例因子为正数 r_2/r_1 的放大（enlargement，又称位似 homothety）的中心。第 (ii) 部分就是蒙日（Monge）的三圆定理：把以 Q 和 R 为中心的两个放大复合起来，便得到以 P 为中心的放大，而两个放大的复合的中心，位于经过这两个中心的直线上。(ii) 中的代数运算是重心坐标（barycentric coordinates，又称仿射坐标 affine coordinates）的初步体验；在这种坐标中，共线就是一条系数之和为零的线性关系。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）发展了背后的线性映

射理论，而 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）则收录了更多同类题目。

详细解答

STEP II 2015，第 9 题

力学 • 刚体转动

STEP II 2015，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/15-s2-q9

关键思路

这个系统的一切都由一个点控制：两个质点的质心（centre of mass）G。一个以 A 为枢轴的刚体（rigid body）静止悬挂时，G 位于 A 的正下方；而当且仅当它有足够能量把 G 送过最高点时，它才会作完整的旋转。因此，只要找出 G 的位置及其与 A 的距离，两部分便迎刃而解。

解答

我们把 B 和 C 视为以轻质刚性杆连接的质点，并设 A 处的转轴光滑，因此唯一做功的力是重力。

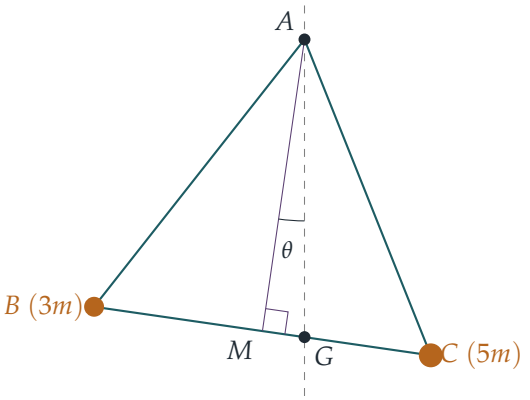
第 (i) 部分

由于杆是轻的，系统的质心就是位于 B 的 $3m$ 与位于 C 的 $5m$ 的质心。该点 G 位于 BC 上，且

$$BG = \frac{5}{8}a, \qquad CG = \frac{3}{8}a.$$

设 M 为 BC 的中点。于是 $MG = \frac{5}{8}a - \frac{1}{2}a = \frac{1}{8}a$，位于靠近 C 的一侧。由于三角形是等边三角形，$AM \perp BC$ 且 $AM = \frac{\sqrt{3}}{2}a$，因此在三角形 AMG 中应用毕氏定理（Pythagoras），

$$AG^2 = \frac{3a^2}{4} + \frac{a^2}{64} = \frac{49a^2}{64}, \qquad AG = \frac{7a}{8}.$$



在平衡状态下，作用于 G 的重量 $8mg$ 对 A 的力矩（moment）必须为零，所以 G 位于 A 的正下方：直线 AG 是铅垂的。而 BC 垂直于 AM，水平方向则垂直于铅垂线 AG。把一对直线同时旋转一个直角，并不会改变它们之间的夹角，因此

$$\theta = \angle GAM.$$

在直角三角形 AMG（直角在 M）中，

$$\sin\theta = \frac{MG}{AG} = \frac{a/8}{7a/8} = \frac{1}{7},$$

得证。注意较重的质点 C 是 BC 较低的一端，这正合理。

第 (ii) 部分

两个质点与转轴的距离都是 a，所以它们的速率永远相同。若三角形以角速率（angular speed）ω 转动，每个质点的速率都是 $a\omega$。冲量（impulse）作用后的一瞬间，B 的速率为 v，C 亦然，因此动能为

$$\tfrac{1}{2}(3m)v^2 + \tfrac{1}{2}(5m)v^2 = 4mv^2.$$

（等价地说，对 A 的转动惯量（moment of inertia）是 $8ma^2$，而 $\omega = v/a$。）冲量的方向并不重要，因为下面的能量方程只涉及速率。

> ? **想一想** 我们是否需要 v 足够大，使某些东西在最高点保持拉紧，就像系在绳上的质点那样?

不需要。杆是刚性的，既能拉也能推，所以唯一的要求是在系统转完一整圈之前，角速率永不降至零。这纯粹是一个能量条件。

系统的势能是 $8mg$ 乘以 G 的高度。由于 G 在以 A 为圆心、半径为 $AG = \frac{7}{8}a$ 的圆上运动，当 G 位于 A 的正上方时势能最大，此时它比起始位置高出

$$2 \cdot \frac{7a}{8} = \frac{7a}{4}$$

。根据能量守恒（conservation of energy），在 G 升高了 h 的位置，

$$4m(a\omega)^2 = 4mv^2 - 8mgh.$$

当 $h = \frac{7}{4}a$ 时，右方最小。若右方在该处为正，则全程 $\omega \neq 0$，所以 ω 永不变号，系统一直朝同一方向转动，一圈接一圈地旋转。若右方为负，则 ω 必定在 G 到达最高点之前降至零，系统便会摆回去。因此条件是

$$4mv^2 > 8mg \cdot \frac{7a}{4} \iff v^2 > \frac{7ag}{2}.$$

故此

$$v_0 = \sqrt{\frac{7ag}{2}}.$$

（当 $v = v_0$ 恰好成立时，G 会缓缓趋向 A 正上方的位置，却要花无限长的时间才能到达，这就是条件为严格不等式（strict inequality）的原因。）

验算

- 若两个质量相等，G 便会在 M，得出 $\theta = 0$ 及 $AG = \frac{\sqrt{3}}{2}a$：把 $MG = 0$ 代入公式正好得出这结果。
- 以数值方法在所有方位中求势能的最小值，得出 $\sin\theta = 0.142857\ldots = \frac{1}{7}$。
- 单位：ag 的单位是 $\mathrm{m^2\,s^{-2}}$，所以 v_0 是速率。当 $a = 1\,\mathrm{m}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$ 时，$v_0 \approx 5.86\,\mathrm{m\,s^{-1}}$。
- 以电脑逐步模拟运动，证实当 $v = 1.01v_0$ 时系统会越过最高点，而当 $v = 0.99v_0$ 时则会摆回去。

常见错误

- 把 G 的升幅当作 $2 \cdot \frac{\sqrt{3}}{2}a$（用了 AM 而不是 AG），结果得出 $v_0^2 = 2\sqrt{3}\,ag$。
- 把 B 和 C 分开处理，忘记冲量同样令 C 的速率为 v。
- 在最高点加上“绳子保持拉紧”的条件：杆可以承受推力（thrust），所以根本不需要这条件。

STEP 之外

这里的能量论证是复摆（compound pendulum）的刚体版本：一个在距离质心 d 处以枢轴支撑的物体，其行为就像单摆（simple pendulum），运动方程为 $I\ddot{\psi} = -Mgd\sin\psi$。

在大学一年级的课程中，你会学到这个方程的相平面（phase plane），其中临界速率 v_0 对应于分隔摆动与完整旋转的分界线（separatrix）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）详细讨论了刚体绕固定轴的转动以及物理摆（physical pendulum）。

详细解答

STEP II 2015，第 10 题

力学 • 约束运动

STEP II 2015，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：

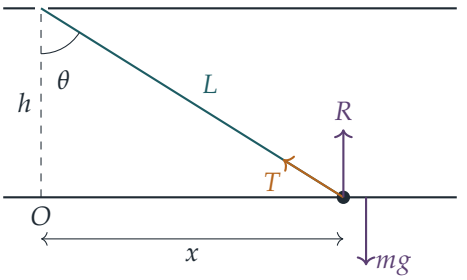
step.maths.org/questions/15-s2-q10

关键思路

这是一道约束（constraint）问题：质点的运动完全由绳子决定，所以先处理运动学（kinematics），再处理力。先把天花板下方的绳长用质点与小孔正下方那一点的水平距离表示出来，然后对时间求导；这段绳长的变化率是已知的。求得加速度后，在水平方向应用牛顿第二定律（Newton's second law）便得出张力（tension），再在铅垂方向分解力便得出法向反作用力（normal reaction）；质点留在地面期间，法向反作用力必须保持非负。

解答

我们把质点视为光滑水平地面上的一点，并把绳子视为轻而不可伸长（inextensible），因此整条绳子的张力 T 都相同。设 O 为小孔正下方地面上的一点，x 为质点与 O 的距离，L 为小孔与质点之间的绳长。



速率

由毕氏定理（Pythagoras）及图中的三角形，

$$L^2 = x^2 + h^2, \qquad x = h\tan\theta, \qquad \sin\theta = \frac{x}{L}.$$

绳子以速率 V 经小孔被拉入，所以 $\dot{L} = -V$。把 $L^2 = x^2 + h^2$ 对时间求导（h 为常数）：

$$2L\dot{L} = 2x\dot{x} \quad \Longrightarrow \quad \dot{x} = \frac{L}{x}\dot{L} = -\frac{VL}{x} = -\frac{V}{\sin\theta}.$$

因此质点以速率

$$\boxed{v = \frac{V}{\sin\theta} = V\operatorname{cosec}\theta}$$

向 O 移动，这个速率总是大于 V（以速率 V 移动的是绳子，而不是质点）。

> **想一想** 为甚么质点的速率不就是 V，或者 $V\sin\theta$?

只有质点速度*沿绳子方向*的分量，才必须等于绳子被拉入的速率。由于绳子与地面成角 $\frac{\pi}{2} - \theta$，这个分量是 $v\sin\theta$。令 $v\sin\theta = V$，便得出与上面相同的答案，而且清楚说明了为甚么 $v > V$。

加速度

利用商法则（quotient rule）、$\dot{L} = -V$ 及 $\dot{x} = -VL/x$，再对 $\dot{x} = -VL/x$ 求导一次：

$$\ddot{x} = -V\,\frac{\dot{L}\,x - L\,\dot{x}}{x^2} = -V\,\frac{-Vx + VL^2/x}{x^2} = -\frac{V^2(L^2 - x^2)}{x^3} = -\frac{V^2h^2}{x^3}.$$

代入 $x = h\tan\theta$，得 $\ddot{x} = -\dfrac{V^2}{h\tan^3\theta}$。由于 $\dot{x}$ 和 $\ddot{x}$ 都是负数，质点正在加速：其加速度的大小为

$$\boxed{a = \frac{V^2}{h\tan^3\theta} = \frac{V^2\cos^3\theta}{h\sin^3\theta}}$$

方向指向 O。（验证一下：当 θ 减小时，$v = V/\sin\theta$ 的确增大。）

张力，以及离开地面

作用于质点的力有：重量 mg、地面施加的法向反作用力 R（由于地面光滑，此力为铅垂方向），以及沿绳子方向的张力 T；张力的水平分量为指向 O 的 $T\sin\theta$，铅垂分量为向上的 $T\cos\theta$。

在水平方向，由牛顿第二定律，

$$T\sin\theta = ma = \frac{mV^2}{h\tan^3\theta} \quad \Longrightarrow \quad \boxed{T = \frac{mV^2}{h\sin\theta\tan^3\theta} = \frac{mV^2\cos^3\theta}{h\sin^4\theta}}$$

在铅垂方向，质点留在地面期间处于平衡状态，所以

$$R + T\cos\theta = mg \quad\Longrightarrow\quad R = mg - \frac{mV^2\cos\theta}{h\sin\theta\tan^3\theta} = mg - \frac{mV^2}{h\tan^4\theta}.$$

? **想一想** 质点被拉近时，R 怎样变化?

当质点接近 O 时，θ 减小，所以 $\tan^4\theta$ 减小，R 亦随之稳定地减小。开始时（θ 较大）R 接近 mg；只要 $R \geq 0$，质点便留在地面，而在 $R = 0$ 的第一刻离开地面，即当

$$\frac{mV^2}{h\tan^4\theta} = mg \quad\Longleftrightarrow\quad \tan^4\theta = \frac{V^2}{gh},$$

此即所求。对于更小的 θ，公式会给出 $R < 0$，这是不可能的，所以质点确实已离开地面。

验算

- 量纲（dimensions）：V^2/h 是加速度，mV^2/h 是力，而 $V^2/(gh)$ 是无量纲的，正如 $\tan^4\theta$ 必须是无量纲一样。
- 极限情况：当 $\theta \to \frac{\pi}{2}$（质点非常远）时，绳子几乎水平，$v \to V$ 且 $a \to 0$，符合预期。当 $\theta \to 0$ 时，速率和张力都趋于无限大，所以质点必定在到达 O 之前离开地面。
- 数值：取 $V = 2$、$h = 1$、$\theta = \frac{\pi}{4}$，得 $v = 2\sqrt{2} \approx 2.83$ 及 $a = 4$；以数值方法对 $x(t) = \sqrt{(L_0 - Vt)^2 - h^2}$ 求导，得出相同的数值。取 $g = 9.8$，质点在 $\theta \approx 38.6^\circ$ 时离开地面。

常见错误

- 把质点的速率当作 V 或 $V\sin\theta$，而不是 $V/\sin\theta$。
- 第二次对 $\dot{x} = -VL/x$ 求导时，忘记 L 也在变化。
- 因为绳子以恒定速率移动，便假设张力不变；其实质点在加速，所以张力并非不变。

STEP 之外

这里的质点只有一个自由度（degree of freedom）：一旦绳子的运动给定，质点在每一时刻的位置便已确定。带有这类约束的问题，可用拉格朗日力学（Lagrangian mechanics）系统地处理；这是大学课程第一或第二年会学到的内容，做法是把约束直接融入坐标的选择之中，之后才求出像 R 和 T 这样的约束力。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）收录了许多这类问题。

详细解答

STEP II 2015，第 11 题

力学 · 动量与能量

STEP II 2015，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：

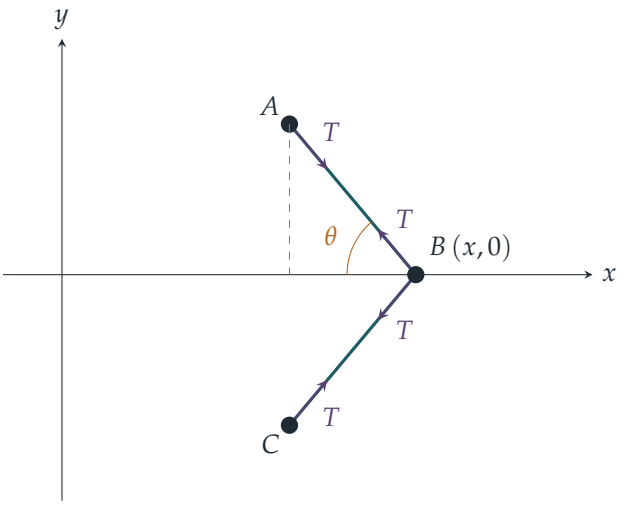
step.maths.org/questions/15-s2-q11

关键思路

由于运动始终关于 x 轴对称，两个坐标 x 和 θ 就足以描述整个系统。桌面光滑（smooth），绳子轻而不可伸长（light and inextensible），因此有两个量守恒（conserved）：x 方向的动量（momentum）和动能（kinetic energy）。这两条守恒定律直接给出 (ii) 中的两条方程；细看第二条，它告诉我们 $\dot{\theta}$ 永远不可能为零。单凭这一点，(iii) 和 (iv) 便迎刃而解。

解答

我们把 A、B、C 视为质点（particles），绳子视为轻而不可伸长。重力和法向反作用力（normal reaction）都是竖直的，所以唯一的水平力是张力（tension），而张力是系统的内力。



第 (i) 部分

A 与 B 相距 a，所在直线与 x 轴负方向成角 θ，因此

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

对 t 求导，并对 $\cos\theta$ 和 $\sin\theta$ 使用链式法则（chain rule），得

$$\dot{\mathbf{r}}_A = \left(\dot{x} + a\sin\theta\,\dot{\theta},\ a\cos\theta\,\dot{\theta}\right),$$

此即所求（而 $\dot{x} = v$ 是 B 的速率）。

整个装置和冲量（impulse）在关于 x 轴的反射下都是对称的，所以运动始终保持对称：$C = (x - a\cos\theta,\ -a\sin\theta)$，其速度为 $\left(\dot{x} + a\sin\theta\,\dot{\theta},\ -a\cos\theta\,\dot{\theta}\right)$。

第 (ii) 部分

首先看初始值。在冲量作用的一瞬间，两条绳子都沿 y 轴，与 B 的新速度垂直，因此绳子无需猛拉 A 或 C；它们由静止开始，而 $\dot{x} = u$。总 x 方向动量为 mu，动能为 $\frac{1}{2}mu^2$。

动量。没有外部水平力，所以总 x 方向动量一直等于 mu：

$$m\dot{x} + 2m\left(\dot{x} + a\sin\theta\,\dot{\theta}\right) = mu \quad\Longrightarrow\quad 3\dot{x} + 2a\dot{\theta}\sin\theta = u. \tag{1}$$

（A 和 C 的 y 方向动量互相抵消，理应如此。）这里右边是 u，即 B 的初速率。试卷印的是 v，但 v 已经用来表示 $\dot{x}$，所以这必定是 u 的误印。

能量。绳子轻而不可伸长，所以张力对整个系统所做的净功为零，而其他力亦不做功。在 A 与 C 相撞之前，动能守恒：

$$\dot{x}^2 + 2\left[\left(\dot{x} + a\sin\theta\,\dot{\theta}\right)^2 + a^2\cos^2\theta\,\dot{\theta}^2\right] = u^2. \tag{2}$$

由 (1)，记 $s=\sin\theta$、$c=\cos\theta$，得

$$\dot{x}=\tfrac{1}{3}\big(u-2as\dot{\theta}\big),\qquad \dot{x}+as\dot{\theta}=\tfrac{1}{3}\big(u+as\dot{\theta}\big).$$

代入 (2)，交叉项 $\mp 4uas\dot{\theta}$ 互相抵消：

$$\tfrac{1}{9}\big(u-2as\dot{\theta}\big)^2+\tfrac{2}{9}\big(u+as\dot{\theta}\big)^2+2a^2c^2\dot{\theta}^2=\tfrac{1}{9}\big(3u^2+6a^2s^2\dot{\theta}^2\big)+2a^2c^2\dot{\theta}^2=u^2.$$

所以 $a^2\dot{\theta}^2\big(\frac{2}{3}s^2+2c^2\big)=\frac{2}{3}u^2$，即 $a^2\dot{\theta}^2(s^2+3c^2)=u^2$。由于 $s^2+3c^2=3-2s^2$，

$$\dot{\theta}^2=\frac{u^2}{a^2(3-2\sin^2\theta)},$$

这就是第二个结果（同样把 v 理解为 u）。

第 (iii) 部分

因为 $3-2\sin^2\theta\le 3$，由第 (ii) 部分得

$$\dot{\theta}^2\ge\frac{u^2}{3a^2}>0.$$

所以 $\dot{\theta}$ 永不为零。在两次碰撞之间它是连续的，因此保持同一符号，θ 单调（monotonically）变化。

开始时 A 静止，所以由第 (i) 部分得 $\dot{x}+a\dot{\theta}=0$，即 $\dot{\theta}=-u/a<0$。因此 θ 从 $\frac{\pi}{2}$ 开始稳定地减小。A 和 C 的 x 坐标始终相同，而 y 坐标为 $\pm a\sin\theta$，所以它们恰好在 $\sin\theta=0$ 时相遇。第一次碰撞发生在 $\theta=0$，位于 B 的后方。

想一想 碰撞对 $\dot{x}$ 和 $\dot{\theta}$ 有甚么影响?

在 $\theta=0$ 时，两条绳子都沿 x 轴，A 和 C 的 x 方向速度相同，y 方向速度为 $\pm a\dot{\theta}$。它们沿 y 方向互相接近，所以两者之间的冲量沿 y 方向。这与绳子垂直，因此不会产生冲击张力，所有 x 方向速度都不变。在等质量物体的弹性碰撞（elastic collision）中，沿连心线的速度分量会互换。所以 A 的 y 方向速度由 $a\dot{\theta}$ 变为 $-a\dot{\theta}$：碰撞只是把 $\dot{\theta}$ 的符号反转，而 $\dot{x}$ 保持不变。

碰撞前后的动量和能量都一样，所以 (1) 和 (2)，以及由此得出的 (ii) 两个结果，在下一阶段的运动中仍然成立。此时 $\dot{\theta}>0$，而且和之前一样不可能变为零。所以 θ 从 0 开始稳定地增大，下一个使 $\sin\theta=0$ 的值是 $\theta=\pi$。由于 $\dot{\theta}\ge u/(a\sqrt{3})$，它在有限时间内到达该值，所需时间至多为 $\pi a\sqrt{3}/u$。第二次碰撞发生在

$$\theta=\pi,$$

此时 A 和 C 在 B 的*前方*相遇。

第 (iv) 部分

在整个运动过程中 $\theta \in [0, \pi]$，所以 $\sin\theta \geq 0$。若 $v = \dot{x} = 0$，则由 (1) 得 $2a\dot{\theta}\sin\theta = u$；两边平方并利用 (ii)，

$$4a^2\sin^2\theta \cdot \frac{u^2}{a^2(3-2\sin^2\theta)} = u^2 \iff 4\sin^2\theta = 3 - 2\sin^2\theta \iff \sin^2\theta = \frac{1}{2}.$$

所以可能的值是 $\theta = \frac{\pi}{4}$ 和 $\theta = \frac{3\pi}{4}$。

> **想一想** 平方丢失了一个符号。我们实际上需要哪个符号?

我们把 $2a\dot{\theta}\sin\theta = u$ 平方了，而由于 $u > 0$，我们还需要 $\dot{\theta}\sin\theta > 0$，即 $\dot{\theta} > 0$。在 $\sin\theta = \frac{1}{\sqrt{2}}$ 时，$\dot{\theta}^2 = \frac{u^2}{2a^2}$，所以 $\dot{\theta} = \pm\frac{u}{a\sqrt{2}}$，而 $2a\dot{\theta}\sin\theta = \pm u$。由 (1)：

- 若 $\dot{\theta} > 0$，则 $\dot{x} = \frac{1}{3}(u - u) = 0$；
- 若 $\dot{\theta} < 0$，则 $\dot{x} = \frac{1}{3}(u + u) = \frac{2}{3}u$。

所以答案是否定的。在第一次摆动中（θ 从 $\frac{\pi}{2}$ 减至 0），θ 经过 $\frac{\pi}{4}$ 时 $v = \frac{2}{3}u$。在第一次与第二次碰撞之间（θ 从 0 增至 π），在 $\theta = \frac{\pi}{4}$ 和 $\theta = \frac{3\pi}{4}$ 两处都有 $v = 0$。第二次碰撞后 $\dot{\theta}$ 再次为负，所以在从 π 摆回 0 的过程中，两个角度处都有 $v = \frac{2}{3}u$。之后这个规律不断重复：只有在 θ 增大的摆动中，这些角度处才有 $v = 0$。

验算

- 开始时 $\theta = \frac{\pi}{2}$，(ii) 给出 $\dot{\theta}^2 = u^2/a^2$，与直接由 A 静止得出的 $\dot{\theta} = -u/a$ 吻合。
- 由 (1)，$\dot{x} = \frac{u}{3}\left(1 - \frac{2a\dot{\theta}\sin\theta}{u}\right)$。在 $\theta = 0$ 时这等于 $\frac{u}{3}$：三个质点一起以 $\frac{u}{3}$ 运动，这正是质心（centre of mass）速度；当绳子沿 x 轴拉直时理应如此。在 θ 增大的摆动中，于 $\theta = \frac{\pi}{2}$ 处 $\dot{x} = -\frac{u}{3}$，所以 B 有一段时间向后移动，这解释了为何 v 两次经过 0。
- 用电脑模拟三个由极硬弹簧连接的质点，取 $a = u = 1$，结果重现了 (ii) 的两条方程。模拟显示第一次碰撞在 $\theta = 0$，第二次在 $\theta = \pi$；而 v 只在 θ 增大的摆动中、于 $\theta \approx 0.785$ 和 2.356 附近改变符号，该处 $\dot{\theta} \approx 0.707$。在第一次摆动中，模拟给出 $\theta = \frac{\pi}{4}$ 时 $v \approx 0.67$。

常见错误

- 把 (1) 的右边当作 $\dot{x}$：它其实是由冲量决定的常数 u。
- 以为碰撞会把 θ 送回 $\frac{\pi}{2}$，而第二次碰撞又在 $\theta = 0$ 发生。把 $\dot{\theta}$ 反转会使 θ 增大并越过 $\frac{\pi}{2}$，而且在到达 π 之前没有任何东西能使它停下，因为 $\dot{\theta}^2$ 有大于 0 的下界。

► 在第 (iv) 部分忘记平方会丢失 $\dot{\theta}$ 的符号。

STEP 之外

选取 x 和 θ 作为坐标，并用它们写出动能，正是拉格朗日力学（Lagrangian mechanics）的起点，你会在大学一年级学到。在那个框架中，由于 x 不出现在能量表达式中（它是一个“可忽略”坐标（ignorable coordinate）），x 方向动量守恒会自动出现。之后能量方程把问题化为一条关于 θ 的一阶方程，正如这里一样。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对质点系统的动量和能量方法有非常详尽的论述。

详细解答

STEP II 2015，第 12 题

概率与统计 • 硬币序列游戏

STEP II 2015，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/15-s2-q12

关键思路

这里每个序列的长度都是三，所以某一次抛掷会否结束游戏，只取决于该次抛掷及之前的两次。因此最自然的做法是追踪*最后两次抛掷*：可能的状态（state）只有四个（HH、HT、TH、TT），而从每个状态出发，下一次抛掷要么结束游戏，要么把我们带到另一个状态。有些状态一出现便决定了胜负；至于其余状态，我们以下一次抛掷作条件（conditioning），列出一对简单的方程。

解答

全文假设各次抛掷互相独立（independent），每次出现 H 或 T 的概率（probability）均为 $\frac{1}{2}$。头两次抛掷不可能完成任何序列，所以游戏必定会到达四个“两次抛掷”状态之一，每个的概率均为 $\frac{1}{4}$。另外要注意，游戏以概率 1 结束：无论处于甚么位置，接下来三次抛掷完成某个指定序列的概率是 $\frac{1}{8}$，所以游戏再持续超过 $3n$ 次抛掷的机会至多为 $(\frac{7}{8})^n \to 0$。

第 (i) 部分

这里只有 HHT（属于 A）和 THH（属于 B）计算在内。

若头两次抛掷是 HH：硬币会连续出现 H 一段时间，而第一个出现的 T 会完成 HHT，因为它之前至少有两个 H。在这个 T 之前从未出现过反面，所以 THH 不可能已经出现。因此 A 胜出。

若头两次抛掷中出现 T：考虑在那个 T 之后第一次出现两个连续正面的时刻。紧接这个 HH 之前的那次抛掷不可能是 H（否则在那个 T 之后便会有一个更早的 HH，或者这个 HH 就是开头的一对，而这已被排除），所以它是 T，于是 THH 已经完成。HHT 不可能

更早出现，因为它需要一个 HH，而这正是第一个。因此 B 胜出。

所以 A 胜出若且唯若头两次抛掷都是正面，从而

$$\mathrm{P}(A \text{ 胜出}) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4},$$

即为所求。

第 (ii) 部分

现在 HHT、THH、TTH、HTT 四个序列全部有效。以头两次抛掷作条件。

- **HH**。与第 (i) 部分一样，抛掷结果会是 H、H、...，直至第一个 T 出现并完成 HHT。其他序列都不包含“HH 之后接 H”，而且尚未出现任何反面来开始 THH、TTH 或 HTT。因此 A 胜出。
- **TT**。把 H 和 T 的角色互换，以同样的论证可知第一个 H 会完成 TTH，所以 C 胜出。
- **HT**。若下一次是 T，便得到 HTT，D 胜出。若是 H，最后两次抛掷是 TH（而 HTH 不完成任何序列）。
- **TH**。若下一次是 H，便得到 THH，B 胜出。若是 T，最后两次抛掷是 HT（而 THT 不完成任何序列）。

所以游戏一旦进入状态 HT 或 TH，便会在这两个状态之间来回弹跳，直至 D 或 B 胜出。设 d_1 和 d_2 分别为从 HT 和从 TH 出发时 D 胜出的概率。以下一次抛掷作条件，

$$d_1 = \tfrac{1}{2} + \tfrac{1}{2}d_2, \qquad d_2 = \tfrac{1}{2}d_1.$$

代入得 $d_1 = \frac{1}{2} + \frac{1}{4}d_1$，所以 $d_1 = \frac{2}{3}$，$d_2 = \frac{1}{3}$。从这些状态出发只有 B 或 D 能胜出，所以 B 从 HT 出发胜出的概率是 $\frac{1}{3}$，从 TH 出发则是 $\frac{2}{3}$。因此

$$\mathrm{P}(A) = \tfrac{1}{4}, \qquad \mathrm{P}(C) = \tfrac{1}{4},$$
$$\mathrm{P}(D) = \tfrac{1}{4}\cdot\tfrac{2}{3} + \tfrac{1}{4}\cdot\tfrac{1}{3} = \tfrac{1}{4}, \qquad \mathrm{P}(B) = \tfrac{1}{4}\cdot\tfrac{1}{3} + \tfrac{1}{4}\cdot\tfrac{2}{3} = \tfrac{1}{4}.$$

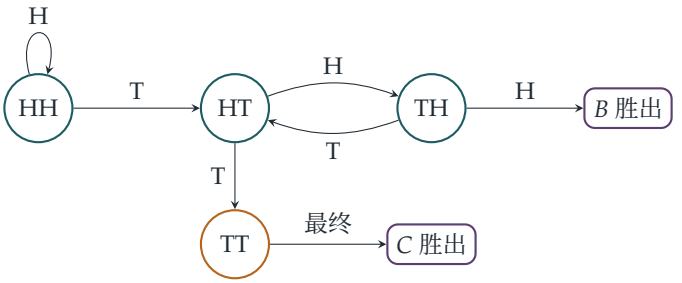
每位玩家胜出的概率都是 $\frac{1}{4}$。

第 (iii) 部分

现在只有 THH（属于 B）和 TTH（属于 C）计算在内。

头两次抛掷是 TT。之后的抛掷会是 T、T、...，直至第一个 H 出现并完成 TTH。THH 需要一个 TH 之后接 H，而 H 尚未出现过，所以 B 不可能先胜出。因此 C 以概率 1 胜出。

至于其他开局，在游戏尚未结束的前提下，看清状态（最后两次抛掷）如何转移会很有帮助：



（从 TT 出发，下一个 T 令我们留在 TT，下一个 H 则完成 TTH，所以我们可以把 TT 视为 C 胜出。）

设 p、q、r 分别为从 HT、TH、HH 出发时 C 胜出的概率：

从 HT 出发。T 会得到 TT，从那里 C 以概率 1 胜出；H 会得到 HTH，它不完成任何一个序列，并令我们处于状态 TH。所以

$$p = \tfrac{1}{2}\cdot 1 + \tfrac{1}{2}\cdot q = \tfrac{1}{2} + \tfrac{1}{2}q,$$

即为所求。

从 TH 出发。H 完成 THH，所以 B 胜出；T 得到 THT，不完成任何序列，而我们处于状态 HT。所以 $q = \tfrac{1}{2}\cdot 0 + \tfrac{1}{2}p = \tfrac{1}{2}p$。

想一想 为甚么一旦知道最后两次抛掷，之前的历史便无关重要?

两个目标序列的长度都是三，所以下一次抛掷会否结束游戏，只取决于该次抛掷和之前两次。更早的抛掷无关重要，而将来的抛掷与过去独立。所以从某个状态出发 C 胜出的机会，无论我们如何到达该状态都是一样的，正因如此 q 才能在 p 的方程右方再次出现。

从 HH 出发。H 令我们留在 HH；T 把我们带到 HT。所以 $r = \tfrac{1}{2}r + \tfrac{1}{2}p$，得 $r = p$。

求解：$p = \tfrac{1}{2} + \tfrac{1}{4}p$，所以

$$p = \tfrac{2}{3}, \qquad q = \tfrac{1}{3}, \qquad r = \tfrac{2}{3}.$$

最后，以头两次抛掷作条件（每种情况的概率均为 $\tfrac{1}{4}$），

$$\mathrm{P}(C \text{ 胜出}) = \tfrac{1}{4}(1 + p + q + r) = \tfrac{1}{4}\left(1 + \tfrac{2}{3} + \tfrac{1}{3} + \tfrac{2}{3}\right) = \tfrac{1}{4}\cdot\tfrac{8}{3} = \tfrac{2}{3}.$$

验算

- 在第 (ii) 部分，四个概率之和为 1，而且答案具有对称性：互换 H 和 T 会交换 $A \leftrightarrow C$ 及 $B \leftrightarrow D$，所以 $\mathrm{P}(A) = \mathrm{P}(C)$ 和 $\mathrm{P}(B) = \mathrm{P}(D)$ 是必然的。

- 在第 (iii) 部分，$q = \frac{1}{3} < p$ 是合理的：从 TH 出发，B 只差一次抛掷便胜出。
- 以电脑模拟（simulation）每类游戏各 200,000 局，得到第 (i) 部分 $\mathrm{P}(A) \approx 0.250$，第 (ii) 部分四位玩家都接近 0.25，第 (iii) 部分 $\mathrm{P}(C) \approx 0.667$，与沿马尔可夫链所作的精确计算吻合。

常见错误

- 假设两个长度相同的序列先出现的机会必定相等。HHT 和 THH 各自的期望等待时间（expected waiting time）都是 8 次抛掷，但 THH 每四次便有三次胜过 HHT。
- 忘记一次没有带来胜利的抛掷仍可把你带到有用的状态：从 HT 出发，抛出 H 并非单纯“失败”，而是把你带到 TH。
- 写下 $p = \frac{1}{2} + \frac{1}{2}q$，却没有解释为何 HTH 令你处于与从 TH 出发相同的位置。

STEP 之外

这里的方法是踏入马尔可夫链（Markov chains）的第一步：这是一种将来只取决于当前状态的过程，可以通过解线性方程求出吸收概率（absorption probabilities）来分析。这个游戏是彭尼游戏（Penney's game）的一个版本：无论对手选择哪个由三次抛掷组成的序列，你总能选出一个以至少 $\frac{2}{3}$ 的概率胜过它的序列，所以“哪个序列先出现”并不是一个传递关系（transitive relation）。Grimmett 和 Welsh 的 *Probability: An Introduction*（《概率导论》）正是以这个程度讨论马尔可夫链和吸收概率。

详细解答

STEP II 2015，第 13 题

概率与统计 · 指数分布的矩

STEP II 2015，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/15-s2-q13

关键思路

成本由固定金额 ky 加上一个随机金额组成，而除非洪水高于预备的高度，否则这个随机金额为零。因此一切都归结为 $W = \max(X - y, 0)$ 的矩（moments），即洪水超出预备高度的部分。指数分布（exponential distribution）具有无记忆性（memoryless）：在 $X > y$ 的条件下，超出量 $X - y$ 仍然服从参数相同的指数分布。这使期望值（expectation）和方差（variance）都变成只需一行的计算。

解答

记 $p = \mathrm{P}(X > y) = \int_y^\infty \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \mathrm{e}^{-\lambda y}$，并设

$$W = \begin{cases} X - y & \text{若} X > y, \\ 0 & \text{若} X \le y, \end{cases} \qquad \text{使得} \qquad C = ky + aW.$$

这里 $y \ge 0$ 是预先选定的常数，所以 ky 并非随机量。

第 (i) 部分

由于当 $X \le y$ 时 $W = 0$，

$$\begin{aligned} \mathrm{E}(W) &= \int_y^\infty (x - y)\,\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \int_0^\infty u\,\lambda \mathrm{e}^{-\lambda(u+y)}\,\mathrm{d}u \qquad (u = x - y) \\ &= \mathrm{e}^{-\lambda y} \int_0^\infty u\,\lambda \mathrm{e}^{-\lambda u}\,\mathrm{d}u = \mathrm{e}^{-\lambda y} \cdot \frac{1}{\lambda}, \end{aligned}$$

这里用到参数为 λ 的指数分布的平均值是 $\frac{1}{\lambda}$ 这个事实（或者用分部积分法（integration by parts）计算）。因此

$$\mathrm{E}(C) = ky + a\,\mathrm{E}(W) = ky + \frac{a}{\lambda}\mathrm{e}^{-\lambda y},$$

即为所求。

求最小值。对 $y \geq 0$，设 $g(y) = ky + \frac{a}{\lambda}\mathrm{e}^{-\lambda y}$。则

$$g'(y) = k - a\mathrm{e}^{-\lambda y}, \qquad g''(y) = a\lambda\mathrm{e}^{-\lambda y} > 0,$$

所以 g' 严格递增，由 $g'(0) = k - a$ 开始，当 $y \to \infty$ 时趋向 k。

> ? **想一想** 为甚么不能直接令 $g'(y) = 0$ 就了事?

因为 y 限制在 $y \geq 0$，而驻点（stationary point）$\mathrm{e}^{-\lambda y} = k/a$ 只有在 $k/a \leq 1$ 时才落在这个范围内。所以我们按 a/k 的值分情况讨论。

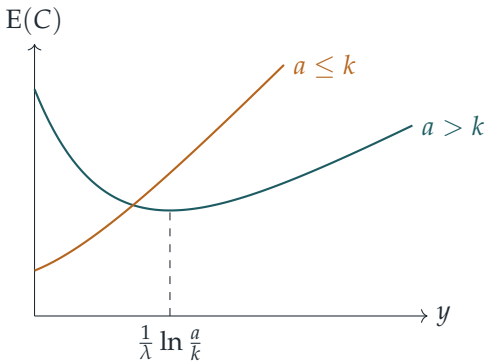
- $a/k > 1$。此时 $g'(0) = k - a < 0$，而 $g'(y) = 0$ 于

  $$y = \frac{1}{\lambda}\ln\frac{a}{k} > 0.$$

  由于 g' 递增，g 在此点之前递减、之后递增，所以这是 $y \geq 0$ 上的全域最小值。最小期望成本为 $\frac{k}{\lambda}\left(1 + \ln\frac{a}{k}\right)$。

- $a/k \leq 1$。此时对所有 $y > 0$ 都有 $g'(y) > g'(0) = k - a \geq 0$，所以 g 在 $y \geq 0$ 上严格递增，最小值出现在 $y = 0$：完全不作任何预备。此时期望成本为 $\frac{a}{\lambda}$。

这是合理的：每多预备一米，必定要付出 k，而只有在洪水本来会达到该高度时才能节省 a。若 $a \leq k$，节省下来的永远抵不上付出的。



第 (ii) 部分

加上常数 ky 不会改变方差，所以 $\mathrm{Var}(C) = a^2\,\mathrm{Var}(W)$。至于二阶矩，用同一个代换 $u = x - y$，

$$\mathrm{E}(W^2) = \int_y^\infty (x-y)^2 \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \mathrm{e}^{-\lambda y}\int_0^\infty u^2 \lambda \mathrm{e}^{-\lambda u}\,\mathrm{d}u = \mathrm{e}^{-\lambda y}\cdot\frac{2}{\lambda^2},$$

因为参数为 λ 的指数随机变量 U 满足 $\mathrm{E}(U^2) = \mathrm{Var}(U) + \mathrm{E}(U)^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2}$。因此，以 $p = \mathrm{e}^{-\lambda y}$，

$$\mathrm{Var}(C) = a^2\left(\frac{2p}{\lambda^2} - \frac{p^2}{\lambda^2}\right) = \frac{a^2}{\lambda^2}\,\mathrm{e}^{-\lambda y}\left(2 - \mathrm{e}^{-\lambda y}\right).$$

预先花费的金额是 ky，由于 $k > 0$，它随 y 递增；所以我们要证明 $\mathrm{Var}(C)$ 随 y 增加而减少。求导得

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(2\mathrm{e}^{-\lambda y} - \mathrm{e}^{-2\lambda y}\right) = -2\lambda \mathrm{e}^{-\lambda y} + 2\lambda \mathrm{e}^{-2\lambda y} = -2\lambda \mathrm{e}^{-\lambda y}\left(1 - \mathrm{e}^{-\lambda y}\right),$$

当 $y > 0$ 时这严格为负，因为此时 $0 < \mathrm{e}^{-\lambda y} < 1$。所以 $\mathrm{Var}(C)$ 在 $y \geq 0$ 上关于 y 严格递减：预备花费越多，成本的方差越小，证明完毕。（等价地，$p(2-p) = 1 - (1-p)^2$ 在 $[0,1]$ 上随 p 递增，而 $p = \mathrm{e}^{-\lambda y}$ 随 y 递减。）

验算

- 当 $y = 0$ 时，$C = aX$，所以 $\mathrm{E}(C) = a/\lambda$ 及 $\mathrm{Var}(C) = a^2/\lambda^2$，两者都与公式吻合。
- 当 $y \to \infty$ 时方差趋向 0：极高的防洪设施几乎消除了所有不确定性，只剩下固定成本 ky。
- 取 $k = 1$、$a = 3$、$\lambda = 0.5$、$y = 1.2$，公式给出 $\mathrm{E}(C) \approx 4.493$ 及 $\mathrm{Var}(C) \approx 28.67$；以电脑模拟一百万年得到 4.496 及 28.68。对这些数值，最佳高度为 $2\ln 3 \approx 2.197$，经电脑搜索确认。

常见错误

- 把 $a(x-y)\lambda\mathrm{e}^{-\lambda x}$ 从 0 而不是从 y 开始积分：当 $X \leq y$ 时并没有洪水损失。
- 忘记边界 $y \geq 0$，即使在 $a < k$ 时仍给出 $y = \frac{1}{\lambda}\ln\frac{a}{k}$，而此时它是负数。
- 把 ky 计入方差；它是常数，对方差毫无贡献。

STEP 之外

$\mathrm{E}\big(\max(X-y,0)\big)$ 这个量正是行使价（strike price）为 y 的认购期权（call option）的收益，而最优条件 $\mathrm{P}(X>y)=k/a$ 则是运筹学（operations research）中经典的“报贩”（newsvendor）法则：选择使预备的边际成本等于期望边际节省的水平。令答案如此简洁的无记忆性，在连续分布之中是指数分布所独有的。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）涵盖了指数分布以及随机变量函数的期望值。

详细解答

STEP III 2015，第 1 题

纯数学 · 归约公式与代换

STEP III 2015，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/15-s3-q1

关键思路

在第 (i) 部分，$I_n - I_{n+1}$ 可以合并成一个分子为 u^2 的积分。把 u^2 写成 $u \cdot u$，再用分部积分 (integration by parts)，就能回到 I_n。在第 (ii) 部分，把 x 换成 $1/x$ 时，式子 $(x - x^{-1})^2$ 保持不变，而这个代换 (substitution) 会产生因子 x^{-2}。把两种形式取平均，得到 $1 + x^{-2}$，这正好是 $x - x^{-1}$ 的导数 (derivative)，于是代换 $u = x - x^{-1}$ 便行得通。至于第 (iii) 部分，只要把分母写成以 $x - x^{-1}$ 表示的形式，它其实就是乔装了的第 (ii) 部分。

解答

第 (i) 部分

把两个被积函数（integrand）通分，

$$I_n - I_{n+1} = \int_0^\infty \frac{(1+u^2)-1}{(1+u^2)^{n+1}}\,\mathrm{d}u = \int_0^\infty u \cdot \frac{u}{(1+u^2)^{n+1}}\,\mathrm{d}u.$$

第二个因子的原函数（antiderivative）很容易求得：

$$\int \frac{u}{(1+u^2)^{n+1}}\,\mathrm{d}u = -\frac{1}{2n(1+u^2)^n}.$$

因此我们进行分部积分，对 u 求导，并对 $u/(1+u^2)^{n+1}$ 积分：

$$I_n - I_{n+1} = \left[-\frac{u}{2n(1+u^2)^n}\right]_0^\infty + \frac{1}{2n}\int_0^\infty \frac{1}{(1+u^2)^n}\,\mathrm{d}u.$$

边界项（boundary term）在 $u = 0$ 时为 0。当 $u \to \infty$ 时它亦趋向 0，因为由 $n \geq 1$ 可得 $0 \leq u/(1+u^2)^n \leq u/(1+u^2) < 1/u$。因此

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

正如所求。移项整理后，

$$I_{n+1} = \frac{2n-1}{2n} I_n \qquad (n \geq 1).$$

至于起始值，$I_1 = \left[\arctan u\right]_0^\infty = \frac{\pi}{2}$。反复应用这个递推关系（recurrence），

$$I_{n+1} = \frac{\pi}{2} \cdot \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)}.$$

想一想 怎样把奇数的乘积化成阶乘（factorials）？

把分子和分母同乘以偶数的乘积 $2 \cdot 4 \cdots (2n) = 2^n n!$。这样分子变成 $1 \cdot 2 \cdot 3 \cdots (2n) = (2n)!$，分母则变成 $(2^n n!)^2$。所以

$$I_{n+1} = \frac{\pi}{2} \cdot \frac{(2n)!}{2^{2n}(n!)^2} = \frac{(2n)!\,\pi}{2^{2n+1}(n!)^2},$$

正如所求。（若要严格证明，可用归纳法（induction）：公式在 $n = 0$ 时成立，此时它给出 $I_1 = \frac{\pi}{2}$。若公式对 I_n 成立，则乘以 $\frac{2n-1}{2n} = \frac{(2n-1)(2n)}{(2n)^2}$ 会把 $\frac{(2n-2)!}{2^{2n-1}((n-1)!)^2}$ 变成 $\frac{(2n)!}{2^{2n+1}(n!)^2}$。）

第 (ii) 部分

第一个等式。在 J 中代入 $x = 1/t$，于是 $\mathrm{d}x = -t^{-2}\,\mathrm{d}t$，而 x 由 0 变到 ∞ 对应于 t 由 ∞ 变到 0。此外

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

因为平方消除了正负号的改变。因此

$$J = \int_\infty^0 \mathrm{f}\big((t - t^{-1})^2\big)\left(-t^{-2}\right)\mathrm{d}t = \int_0^\infty t^{-2}\,\mathrm{f}\big((t - t^{-1})^2\big)\,\mathrm{d}t,$$

再把哑变量（dummy variable）t 改名为 x，便得到第一种形式。

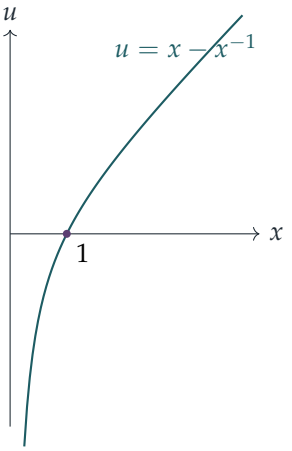
第二个等式。把 J 原来的表达式与这个表达式相加，再除以二，

$$J = \frac{1}{2}\int_0^\infty \left(1 + x^{-2}\right)\mathrm{f}\big((x - x^{-1})^2\big)\,\mathrm{d}x.$$

第三个等式。现在设 $u = x - x^{-1}$，于是 $\mathrm{d}u = (1 + x^{-2})\,\mathrm{d}x$。由于 $\mathrm{d}u/\mathrm{d}x > 0$，$u$ 随 x 严格递增，由 $-\infty$（当 $x \to 0^+$）增至 $+\infty$（当 $x \to \infty$）。所以这个代换在整个范围上是一一对应的变量变换：

$$J = \frac{1}{2}\int_{-\infty}^\infty \mathrm{f}(u^2)\,\mathrm{d}u = \int_0^\infty \mathrm{f}(u^2)\,\mathrm{d}u.$$

最后一步成立，是因为 $\mathrm{f}(u^2)$ 是 u 的偶函数（even function），所以在 $(-\infty, 0)$ 和 $(0, \infty)$ 上的积分相等。第 (ii) 部分至此完成。



这条曲线经过 $(1,0)$，并在 $x>0$ 上持续上升，恰好一次取遍 u 的每一个实数值。

第 (iii) 部分

目标是把被积函数改写成 $x^{-2}\,\mathrm{f}\big((x-x^{-1})^2\big)$ 的形式。由于 $(x-x^{-1})^2 = x^2-2+x^{-2}$，

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

因此

$$\frac{x^{2n-2}}{(x^4-x^2+1)^n} = \frac{x^{2n-2}}{x^{2n}\big(1+(x-x^{-1})^2\big)^n} = x^{-2}\,\mathrm{f}\big((x-x^{-1})^2\big), \qquad \text{其中} \mathrm{f}(y) = \frac{1}{(1+y)^n}.$$

根据第 (ii) 部分中 J 的第一种和最后一种形式，

$$\int_0^\infty \frac{x^{2n-2}}{(x^4-x^2+1)^n}\,\mathrm{d}x = \int_0^\infty \frac{1}{(1+u^2)^n}\,\mathrm{d}u = I_n.$$

在第 (i) 部分的公式中以 $n-1$ 代替 n（当 $n-1\geq 0$ 时成立，即对每个正整数 n 皆成立），得到

$$\int_0^\infty \frac{x^{2n-2}}{(x^4-x^2+1)^n}\,\mathrm{d}x = \frac{(2n-2)!\,\pi}{2^{2n-1}\big((n-1)!\big)^2}.$$

验算

- 当 $n=1$ 时答案是 $\frac{\pi}{2}$，所以 $\int_0^\infty \frac{\mathrm{d}x}{x^4-x^2+1} = \frac{\pi}{2} \approx 1.5708$。当 $n=2$ 时答案是 $\frac{\pi}{4}$，当 $n=3$ 时答案是 $\frac{3\pi}{16} \approx 0.5890$。对原积分进行数值积分，结果与以上各值吻合，直至 $n=8$ 的情况亦然。
- 递推关系给出 $I_2 = \frac{1}{2}I_1 = \frac{\pi}{4}$，与用 $u=\tan\theta$ 直接计算的结果一致：$I_2 = \int_0^{\pi/2}\cos^2\theta\,\mathrm{d}\theta = \frac{\pi}{4}$。

- 取 $\mathrm{f}(y) = \mathrm{e}^{-y}$，第 (ii) 部分表明 $\int_0^\infty \mathrm{e}^{-(x-1/x)^2}\,\mathrm{d}x = \int_0^\infty \mathrm{e}^{-u^2}\,\mathrm{d}u = \frac{\sqrt{\pi}}{2} \approx 0.88623$。对全部四种形式进行数值计算，都得到这个值。

常见错误

- 忘记检查边界项 $u/(1+u^2)^n$ 在无穷远处是否趋于零。这里它确实趋于零，但这只是因为 $n \geq 1$。
- 代入 $u = x - x^{-1}$ 后仍保留上下限 0 和 ∞。新的下限是 $-\infty$，正是 $\mathrm{f}(u^2)$ 的偶性把范围带回 $(0, \infty)$。
- 使用 I_{n+1} 的公式时没有移动下标：第 (iii) 部分的答案是 I_n，而不是 I_{n+1}。

STEP 之外

比值 $\frac{(2n)!}{2^{2n}(n!)^2} = \binom{2n}{n}/4^n$ 是一个中心二项概率（central binomial probability），而第 (i) 部分的递推关系，正是推导出 π 的沃利斯（Wallis）乘积的同一个递推关系。第 (ii) 部分的技巧称为柯西（Cauchy）— Schlömilch 变换（Cauchy–Schlömilch transformation）。它可以计算许多没有初等原函数的定积分，包括像验算中那个的高斯（Gauss）型积分。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）收录了更多可用巧妙代换解决的积分；Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）则严格地处理反常积分（improper integrals）和变量变换。

详细解答

STEP III 2015，第 2 题

纯数学 · 分析学

STEP III 2015，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/15-s3-q2

本题在 *Beyond STEP* 第 2.1 节“最终”是甚么意思中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch02/eventually.html

关键思路

关系 $(s_n) \leq (t_n)$ 的意思是“最终 (eventually) $s_n \leq t_n$”：我们可以不理会前面任意有限多项。因此，要证明一个命题为真，就要指出一个位置 m，保证由此以后不等式 (inequality) 必定成立；要证明一个命题为假，则要找出一对数列 (sequence)，使任何 m 的选择都行不通。时刻问自己：这个结果是从哪一个 n 开始成立的？

解答

第 (i) 部分

正确。对正整数 n，我们可以除以 $n > 0$，所以

$$1000n \leq n^2 \iff 1000 \leq n.$$

因此只要 $n \geq 1000$，便有 $1000n \leq n^2$，取 $m = 1000$ 即可。（更小的 m 都不行，因为 $1000 \cdot 999 > 999^2$。）

第 (ii) 部分

错误。陷阱在于“并非最终 $s_n \le t_n$”只告诉我们 $s_n > t_n$ 对无限多个 n 成立，而不是最终成立。两个数列可以不断互相超越。取

$$s_n = 2 + (-1)^n, \qquad t_n = 2,$$

于是 s_n 依次为 $1, 3, 1, 3, \ldots$，而两个数列的每一项都是正数。

- $(s_n) \le (t_n)$ 不成立：无论提出哪个正整数 m，总有偶数 $n \ge m$，而对这个 n 有 $s_n = 3 > 2 = t_n$。
- $(t_n) \le (s_n)$ 也不成立：无论提出哪个 m，总有奇数 $n \ge m$，而对这个 n 有 $t_n = 2 > 1 = s_n$。

所以命题的假设成立，但结论不成立。

第 (iii) 部分

正确。根据假设，存在正整数 m_1，使得对所有 $n \ge m_1$ 都有 $s_n \le t_n$；又存在正整数 m_2，使得对所有 $n \ge m_2$ 都有 $t_n \le u_n$。设

$$m = \max(m_1, m_2).$$

若 $n \ge m$，则 $n \ge m_1$ 且 $n \ge m_2$，所以两个不等式在 n 处都成立，得

$$s_n \le t_n \le u_n.$$

因此只要 $n \ge m$ 便有 $s_n \le u_n$，这正是 $(s_n) \le (u_n)$。

> ? **想一想** 为甚么不直接取 $m = m_1$，或者 $m = m_1 + m_2$?

取 $m = m_1$ 行不通，因为当 $m_1 \le n < m_2$ 时，我们对 t_n 与 u_n 的大小关系一无所知。我们需要 n 同时越过两个门槛（threshold），而 $\max(m_1, m_2)$ 是能保证这一点的最小选择。（$m_1 + m_2$ 也可以，但不必要地大。）

第 (iv) 部分

正确，取 $m = 4$。一个简短的表格显示应从哪里开始：

n	1	2	3	4	5	6
n^2	1	4	9	16	25	36
2^n	2	4	8	16	32	64

不等式在 $n = 1, 2$ 时成立，在 $n = 3$ 时不成立，从 $n = 4$ 起又再成立。所以 $m = 4$ 是最小的可能选择，我们用归纳法（induction）证明对所有 $n \geq 4$ 都有 $n^2 \leq 2^n$。

基本情况（*base case*）。当 $n = 4$ 时，$n^2 = 16 = 2^4$。

归纳步骤（*inductive step*）。假设 $n \geq 4$ 且 $n^2 \leq 2^n$。由于 $n \geq 3$，有 $n^2 - 2n - 1 = (n-1)^2 - 2 \geq 2 > 0$，所以 $2n + 1 < n^2$。因此

$$\begin{aligned}(n+1)^2 &= n^2 + 2n + 1 \\ &< n^2 + n^2 && \text{因为} 2n+1 < n^2 \\ &= 2n^2 \\ &\leq 2 \cdot 2^n && \text{由归纳假设（inductive hypothesis）} \\ &= 2^{n+1}.\end{aligned}$$

所以结果对 n 成立可推出对 $n+1$ 成立，由归纳法，对每个 $n \geq 4$ 都有 $n^2 \leq 2^n$。因此 $(n^2) \leq (2^n)$，取 $m = 4$。

想一想 归纳法究竟在哪一步需要 n 足够大?

只在 $2n + 1 < n^2$ 这一步，它需要 $n \geq 3$。这就是归纳法不能从 $n = 1$ 开始的原因：由 $n = 2$ 推到 $n = 3$ 的一步要用到 $5 < 4$，这是错的，而事实上 $3^2 > 2^3$。

验算

- 第 (i) 部分：当 $n = 1000$ 时两边都等于 10^6；当 $n = 999$ 时，$999\,000 > 998\,001$。
- 第 (iv) 部分：用电脑对 4 至 2000 的每个 n 检验 $n^2 \leq 2^n$，结果都成立；上面的表格亦确认了 $n = 3$ 时不成立。要感受 2^n 胜出得有多快：$20^2 = 400$，但 $2^{20} = 1\,048\,576$。
- 第 (ii) 部分：无论你尝试哪个 m，都会被下一个偶数 n（对其中一个次序）及下一个奇数 n（对另一个次序）推翻。

常见错误

- 把“并非最终 $\leq$”当成“最终 $>$”。“存在 m, 使得对所有 $n \geq m$……”的否定 (negation) 是“对每个 m，都存在某个 $n \geq m$ 使得 $s_n > t_n$”。
- 在 (iv) 中从 $n = 1$ 或 $n = 2$ 开始归纳：归纳步骤在那里确实不成立。
- 在 (iii) 中，两个假设都用同一个字母 m，仿佛两个门槛必须相同。

STEP 之外

这里的关系正是贯穿整个实分析（real analysis）的“最终”量词（quantifier）：一个数列收敛（converges），是指对每一个容许误差，它最终都落在其极限的该误差范围之内。第 (ii) 部分显示，与实数上的次序不同，数列的这种次序并不是全序（total order）：两个数列未必可以互相比较。电脑科学家在大 O 记号（big-O notation）中使用同一个概念，而 (iv) 就是“指数函数最终胜过多项式”这个熟悉的事实。Abbott 的 *Understanding Analysis*（《理解分析》）是一本亲切易读的书，可以从中看到这类量词论证如何被系统化。

详细解答

STEP III 2015，第 3 题

纯数学 · 极坐标曲线与面积

STEP III 2015，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/15-s3-q3

关键思路

曲线 $r = a\sec\theta$ 其实就是直线 $x = a$，因为 $r\cos\theta = x$。因此 $r - a\sec\theta$ 量度的是：沿着从原点出发、角度为 θ 的射线，一点与该射线和直线交点之间的距离；而方程表示这个距离等于 b。这正是蚌线（conchoid）的定义，其中定点就是极点（pole）。至于作图，可把绝对值（modulus）拆开成 $r = a\sec\theta + b$ 和 $r = a\sec\theta - b$，再逐一考虑哪些 θ 值使 $r \geq 0$。

解答

全题中，$\sec\theta$ 在 $\theta = \pm\frac{\pi}{2}$ 时无定义，因此这些角度须排除。去掉绝对值后，$(*)$ 表示

$$r = a\sec\theta + b \qquad 或 \qquad r = a\sec\theta - b,$$

而无论哪种情况，我们都需要 $r \geq 0$。

第 (i) 部分

为何 $\sec\theta > 0$。假设 $\sec\theta < 0$。由 $|\cos\theta| \leq 1$ 得 $\sec\theta \leq -1$，所以 $a\sec\theta \leq -a$。于是 r 的两个可能值都满足

$$r = a\sec\theta \pm b \leq -a + b < 0,$$

这里用了 $a > b$。这与 $r \geq 0$ 矛盾，所以每个满足 $(*)$ 的点都有 $\sec\theta > 0$，即 $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$。

蚌线。取 A 为极点 O，L 为直线 $x = a$（极坐标形式为 $r\cos\theta = a$，即 $r = a\sec\theta$），并取 $d = b$。注意由于 $a > 0$，A 不在 L 上。

设 P 满足 $(*)$，其极角为 θ，所以 $\cos\theta > 0$。直线 OP 不与 L 平行，它与 L 相交于同一射线上

的点 Q，且 $OQ = a\sec\theta$，即以直角坐标（Cartesian coordinates）表示为 $Q = (a,\, a\tan\theta)$。由于 O、Q 和 P 都在从 O 出发的同一条射线上，与 O 的距离分别为 $a\sec\theta$ 和 r，

$$PQ = |r - a\sec\theta| = b.$$

所以沿着一条经过 $A = O$ 的直线量度，P 与 L 的距离为 b：它位于以此 L、A 及 $d = b$ 构成的尼科梅德斯（Nicomedes）蚌线上。

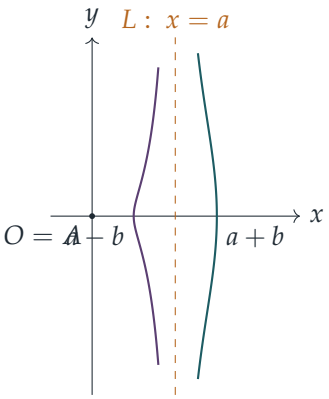
（反过来，该蚌线上的每一点都满足 $(*)$：在任何经过 O 并与 L 相交于 Q 的直线上，与 Q 距离为 b 的两点，沿射线 OQ 与 O 的距离为 $a\sec\theta \pm b$，由于 $a\sec\theta \geq a > b$，两者皆为正数。所以轨迹就是整条蚌线。）

作图。在直角坐标中，$x = r\cos\theta = a \pm b\cos\theta$，而 $y = r\sin\theta$。把 $r - a\sec\theta = \pm b$ 乘以 $\cos\theta$，得 $x - a = \pm b\cos\theta = \pm bx/r$，所以曲线为

$$(x-a)^2(x^2+y^2) = b^2x^2.$$

曲线有两个分支（branch），两者都关于 x 轴对称（以 $-\theta$ 代替 θ 即可见），而当 $\theta \to \pm\frac{\pi}{2}$ 时 $r \to \infty$，两者都以 L 为渐近线（asymptote）：

- 外分支 $r = a\sec\theta + b$，其中 $x = a + b\cos\theta \in (a,\, a+b]$，与轴相交于 $(a+b, 0)$；
- 内分支 $r = a\sec\theta - b$，其中 $x = a - b\cos\theta \in [a-b,\, a)$，与轴相交于 $(a-b, 0)$，该点位于 O 的右方。



（图中取 $a = 2$、$b = 1$：外分支以蓝绿色表示，内分支以梅紫色表示。）

第 (ii) 部分

现在 $a < b$。再逐一检查两种情况，看看何时 $r \geq 0$。

- $r = a\sec\theta + b$。若 $\sec\theta > 0$，它必为正数：这就是与之前相同、位于 L 右方的外分支。若 $\sec\theta < 0$，则 $r \geq 0$ 当且仅当 $-a\sec\theta \leq b$，即 $|\sec\theta| \leq b/a$，即 $\cos\theta \leq -a/b$。

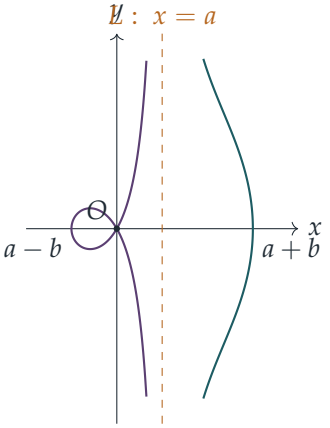
▶ $r = a\sec\theta - b$。若 $\sec\theta < 0$，它为负数，所以没有点。若 $\sec\theta > 0$，则 $r \geq 0$ 当且仅当 $\sec\theta \geq b/a$，即 $0 < \cos\theta \leq a/b$。

设 $\alpha = \arccos(a/b)$，由于 $0 < a/b < 1$，它位于 $(0, \frac{\pi}{2})$ 之内。

想一想 $\sec\theta < 0$ 的那一部分是甚么样子?

它是 $r = b + a\sec\theta$，其中 $\cos\theta \leq -a/b$，即 $\pi - \alpha \leq |\theta| \leq \pi$。在两端，$\cos\theta = -a/b$ 给出 $r = b - b = 0$，所以它在原点开始，也在原点结束。在 $\theta = \pi$ 时得 $r = b - a$，即负 x 轴上的点 $(a - b, 0)$。在两端之间 $r > 0$，而曲线关于 x 轴对称，所以这部分是原点左方的一个闭环，位于 $a - b \leq x \leq 0$ 之内（因为 $x = a + b\cos\theta$）。

内分支 $r = a\sec\theta - b$（其中 $\alpha \leq |\theta| < \frac{\pi}{2}$）同样从原点开始（该处 $\cos\theta = a/b$），并当 $|\theta| \to \frac{\pi}{2}$ 时伸向渐近线 L，一直保持在 $0 \leq x < a$ 之内。所以内分支与闭环合起来构成一条在 O 自身相交的曲线，在该处的切线方向为 $\theta = \pm\alpha$。



（图中取 $a = 1$、$b = 2$，所以 $\alpha = \frac{\pi}{3}$。外分支以蓝绿色表示；内分支连同闭环以梅紫色表示。）

$a = 1$、$b = 2$ 时闭环的面积。此时 $\cos\theta \leq -\frac{1}{2}$ 表示 $\frac{2\pi}{3} \leq |\theta| \leq \pi$，而在闭环上 $r = 2 + \sec\theta$。当 θ 走遍这个范围时，射线恰好扫过闭环一次，所以面积为 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$。利用关于 x 轴的对称性，我们取上半部分 $\frac{2\pi}{3} \leq \theta \leq \pi$ 的两倍：

$$\begin{aligned} \text{面积} &= 2\cdot\frac{1}{2}\int_{2\pi/3}^{\pi}(2+\sec\theta)^2\,\mathrm{d}\theta = \int_{2\pi/3}^{\pi}\left(4 + 4\sec\theta + \sec^2\theta\right)\mathrm{d}\theta \\ &= \Big[4\theta + 4\ln|\sec\theta + \tan\theta| + \tan\theta\Big]_{2\pi/3}^{\pi}. \end{aligned}$$

题目给出的 $\sec\theta$ 的原函数（antiderivative）在此适用，因为区间不包含 $\frac{\pi}{2}$。在 $\theta = \pi$：$\sec\theta = -1$，$\tan\theta = 0$，所以括号内的值为 $4\pi + 4\ln 1 + 0 = 4\pi$。在 $\theta = \frac{2\pi}{3}$：$\sec\theta = -2$，

$\tan\theta = -\sqrt{3}$，所以括号内的值为 $\frac{8\pi}{3} + 4\ln(2+\sqrt{3}) - \sqrt{3}$。相减，得

$$\text{闭环的面积} = \frac{4\pi}{3} + \sqrt{3} - 4\ln\left(2+\sqrt{3}\right).$$

验算

- 数值上，面积为 $4.1888 + 1.7321 - 5.2678 \approx 0.6530$。它是正数，理应如此；而且相比能容纳闭环的矩形 $-1 \le x \le 0$、$|y| \le 0.45$（面积为 0.9）也较小。直接以数值方法在 $[\frac{2\pi}{3}, \frac{4\pi}{3}]$ 上积分 $\frac{1}{2}(2+\sec\theta)^2$，同样得到 0.65301。
- 闭环经过 O 和 $(-1,0)$：在 $\theta = \frac{2\pi}{3}$，$r = 2 - 2 = 0$；在 $\theta = \pi$，$r = 2 - 1 = 1$。
- 在第 (i) 部分的曲线上取一点，例如 $a = 2$、$b = 1$、$\theta = \frac{\pi}{4}$：则 $r = 2\sqrt{2} + 1$，而射线与 L 相交于 $(2,2)$，与 O 的距离为 $2\sqrt{2}$，所以该点在交点之外 $1 = b$ 的位置。

常见错误

- 忘记条件 $r \ge 0$。正是这个条件迫使第 (i) 部分中 $\sec\theta > 0$，也决定了第 (ii) 部分中闭环的 θ 范围。
- 在 $\frac{\pi}{2} < \theta < \pi$ 上积分闭环，即穿过渐近线方向，而不是从 $r = 0$ 的 $\theta = \frac{2\pi}{3}$ 开始。
- 在第二象限处理 $\ln|\sec\theta + \tan\theta|$ 时出现正负号错误；该象限中 $\sec\theta$ 和 $\tan\theta$ 皆为负数。

STEP 之外

蚌线由尼科梅德斯约于公元前 200 年发明，用来三等分角和求立方根；这些问题单靠直尺和圆规是不可能解决的，而不可能性的证明来自域论（field theory），在大学第一个代数课程中便会接触到（见 Allenby，*Rings, Fields and Groups*《环、域与群》）。第 (ii) 部分中在原点的自交点是一个*结点*（node），是代数曲线（algebraic curve）上奇点（singular point）的一个例子，这里的代数曲线是四次曲线 $(x-a)^2(x^2+y^2) = b^2x^2$。更多 STEP 风格的曲线作图题，可参阅 Siklos，*Advanced Problems in Mathematics*《高等数学问题》。

详细解答

STEP III 2015，第 4 题

纯数学 · 三次方程的根

STEP III 2015，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/15-s3-q4

关键思路

第 (iii) 部分的三个条件看似一道三角谜题，但每一个其实都是某个量的虚部（imaginary part）：当 $z_k = r_k e^{i\theta_k}$ 时，$r_k^n \sin(n\theta_k)$ 正是 z_k^n 的虚部。因此，这些条件表示三个复数的幂和（power sum）S_1、S_2、S_3 都是实数。第 (ii) 部分把实数幂和转化为一个实系数（real coefficients）三次方程（cubic），而第 (i) 部分则随即给我们一个实根（real root）。

解答

第 (i) 部分

对实数 x，设 $f(x) = x^3 + ax^2 + bx + c$；这是一个连续（continuous）实函数。令 $R = 1 + |a| + |b| + |c|$。当 $x \geq R$ 时，有 $x \geq 1$，所以 $|x| \leq x^2$ 及 $1 \leq x^2$，从而 $|ax^2 + bx + c| \leq (|a| + |b| + |c|)x^2$。因此

$$f(x) \geq x^3 - (|a| + |b| + |c|)x^2 = x^2\big(x - |a| - |b| - |c|\big) > 0.$$

同理，当 $x \leq -R$ 时，得 $f(x) \leq -|x|^2\big(|x| - |a| - |b| - |c|\big) < 0$。所以 $f(-R) < 0 < f(R)$，由介值定理（intermediate value theorem）可知 f 在 $-R$ 与 R 之间有一个零点。这个零点就是 $(*)$ 的一个实根。

（等价的说法：实多项式（polynomial）的非实根必定成共轭（conjugate）对出现，而三个根不可能全部两两配对。）

第 (ii) 部分

由于 z_1、z_2、z_3 是方程的根，

$$z^3 + az^2 + bz + c = (z - z_1)(z - z_2)(z - z_3),$$

比较系数（comparing coefficients）得

$$a = -(z_1 + z_2 + z_3), \qquad b = z_1z_2 + z_2z_3 + z_3z_1, \qquad c = -z_1z_2z_3.$$

所以 $a = -S_1$。把和平方，得 $S_1^2 = S_2 + 2(z_1z_2 + z_2z_3 + z_3z_1) = S_2 + 2b$，故

$$a = -S_1, \qquad b = \tfrac{1}{2}\left(S_1^2 - S_2\right).$$

至于 c，与其展开 S_1^3，不如利用每个根都满足方程这一事实：对 $k = 1, 2, 3$，有 $z_k^3 + az_k^2 + bz_k + c = 0$。把这三条方程相加，得

$$S_3 + aS_2 + bS_1 + 3c = 0.$$

代入 a 和 b，再乘以 2：

$$\begin{aligned} 6c &= -2S_3 + 2S_1S_2 - S_1\left(S_1^2 - S_2\right) \\ &= -S_1^3 + 3S_1S_2 - 2S_3, \end{aligned}$$

即为所求。

第 (iii) 部分

对 $k = 1, 2, 3$，定义 $z_k = r_k\mathrm{e}^{\mathrm{i}\theta_k}$。由棣美弗定理（de Moivre's theorem），$z_k^n = r_k^n\left(\cos n\theta_k + \mathrm{i}\sin n\theta_k\right)$，所以题目给出的三个条件恰好表示

$$\operatorname{Im} S_1 = \operatorname{Im} S_2 = \operatorname{Im} S_3 = 0,$$

其中 S_1、S_2、S_3 是 z_1、z_2、z_3 的幂和，定义与第 (ii) 部分相同。所以 S_1、S_2、S_3 全是实数。

? **想一想** 第 (ii) 部分是从一个三次方程出发的。这里的三次方程在哪里?

我们自己构造它：以 $(z - z_1)(z - z_2)(z - z_3) = z^3 + az^2 + bz + c$ 定义 a、b、c。那么 z_1、z_2、z_3 就是这个三次方程的根，于是第 (ii) 部分适用，得

$$a = -S_1, \qquad b = \tfrac{1}{2}\left(S_1^2 - S_2\right), \qquad c = \tfrac{1}{6}\left(-S_1^3 + 3S_1S_2 - 2S_3\right),$$

它们全是实数。由第 (i) 部分，这个三次方程有一个实根，而它的根只有 z_1、z_2、z_3，所以某个 z_k 是实数。由于 $r_k > 0$，$z_k = r_k\mathrm{e}^{\mathrm{i}\theta_k}$ 是实数当且仅当 $\sin\theta_k = 0$，即 θ_k 是 π 的整数倍；在 $-\pi < \theta_k < \pi$ 范围内，唯一这样的值是 $\theta_k = 0$。所以至少有一个 k 使 $\theta_k = 0$。

现在设 $\theta_1 = 0$，于是 $z_1 = r_1$ 是非零实数。那么

$$z_2 + z_3 = S_1 - z_1 \quad 及 \quad z_2 z_3 = \frac{-c}{z_1}$$

都是实数。因此 z_2 和 z_3 是二次方程（quadratic）$z^2 - (z_2 + z_3)z + z_2 z_3 = 0$ 的根，而这个方程的系数都是实数。分两种情况。

- 若两根是非实数，则它们互为共轭复数：$z_3 = \overline{z_2} = r_2 \mathrm{e}^{-\mathrm{i}\theta_2}$。由于 $-\theta_2$ 同样位于 $(-\pi, \pi)$ 内，而在这个区间内辐角（argument）是唯一的，所以 $\theta_3 = -\theta_2$（且 $r_3 = r_2$）。
- 若两根是实数，则与上面完全相同的推理（每个根的模（modulus）为正，辐角位于 $(-\pi, \pi)$ 内）给出 $\theta_2 = \theta_3 = 0$，所以同样有 $\theta_2 = -\theta_3$。

无论哪种情况，都有 $\theta_2 = -\theta_3$，即为所求。

验算

- 用根 $1, 2, 3$ 验证第 (ii) 部分：$S_1 = 6$、$S_2 = 14$、$S_3 = 36$，而 $-216 + 252 - 72 = -36 = 6c$，其中 $c = -6$，正与 $(z-1)(z-2)(z-3)$ 相符。
- 反方向验证第 (iii) 部分：当 $\theta_1 = 0$、$\theta_3 = -\theta_2$ 及 $r_3 = r_2$ 时，每个和的第二项与第三项互相抵消，而第一项为零，所以三个条件全部成立。用电脑从数千个随机起点以数值方法求解这三个条件，找到的解全都是一个角为零、另外两个角互为相反数。

常见错误

- 在第 (i) 部分，未加论证便说“三次曲线必定穿过轴”；$|x|$ 很大时的变号需要一行推理来支持。
- 在第 (iii) 部分，忘记 $\theta_k = \pm\pi$ 同样会令 z_k 成为实数；正是开区间 $(-\pi, \pi)$ 排除了这个可能。
- 只从共轭的情况得出 $\theta_2 = -\theta_3$，却忽略了 z_2、z_3 都是实数的情况。

STEP 之外

多项式的系数是其根的基本对称函数（elementary symmetric functions），而第 (ii) 部分的恒等式正是把它们与幂和联系起来的牛顿（Newton）恒等式（Newton's identities）的头几条。大学的第一门代数课会证明：根的任何对称多项式（symmetric polynomial）都可以写成基本对称函数的多项式，而这正是伽罗瓦理论（Galois theory）的起点。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是阅读这方面内容的好起点。

详细解答

STEP III 2015，第 5 题

纯数学 • 以递降法证明无理性

STEP III 2015，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/15-s3-q5

本题在 *Beyond STEP* 第 1.3 节最小反例中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch01/well-ordering.html

关键思路

这是无穷递降法（infinite descent）的证明：假如 $\sqrt{2}$ 是有理数（rational），那么必定存在一个最小的正整数，使 $\sqrt{2}$ 乘以它之后变成整数；然后我们构造出一个更小的这样的数。乘以 $\sqrt{2}-1$ 正好能做到这一点，因为它是介乎 0 与 1 之间的数，而且乘了之后仍留在该集合之内。对于 2 的立方根（cube root），同一招数需要一个同时照顾 $2^{2/3}$ 的集合，这就是题目为何先要求你把这两个数联系起来。

解答

第 (i) 部分

第 3 步。根据第 1 步的假设，对某些整数 p 和 q（$q \neq 0$），有 $\sqrt{2} = \frac{p}{q}$。由于 $\sqrt{2} > 0$，我们可取 p 和 q 皆为正数。于是

$$q\sqrt{2} = p$$

是整数，而 q 是正整数，所以 $q \in S$。因此 S 非空。

第 5 步。由于 $k \in S$，数 $m = k\sqrt{2}$ 是整数。关于 $n = (\sqrt{2}-1)k$，我们需要验证两件事。

- n 是正整数：$n = k\sqrt{2} - k = m - k$ 是两个整数之差，而且由于 $k > 0$ 及 $\sqrt{2} > 1$，有 $n > 0$。

- $n\sqrt{2}$ 是整数：
$$n\sqrt{2} = (\sqrt{2}-1)k\sqrt{2} = 2k - k\sqrt{2} = 2k - m,$$
同样是两个整数之差。

所以 $(\sqrt{2}-1)k \in S$。

第 6 步。由于 $1 < \sqrt{2} < 2$（因为 $1 < 2 < 4$），我们有 $0 < \sqrt{2}-1 < 1$，因此由 $k > 0$ 得
$$(\sqrt{2}-1)k < k.$$
于是第 5 步给出了 S 中一个比 k 小的正整数，与第 4 步中选取 k 为 S 的最小元素相矛盾（contradiction）。（第 4 步是合理的，因为每个非空的正整数集合都有最小元素，而第 3 步已证明 S 非空。）整个论证唯一的假设是第 1 步，所以它必定是错的：$\sqrt{2}$ 是无理数（irrational）。

第 (ii) 部分

记 $a = 2^{1/3}$，则 $a^2 = 2^{2/3}$ 及 $a^3 = 2$。

等价性（equivalence）。若 a 是有理数，则 a^2 作为有理数的乘积亦是有理数。反过来，若 a^2 是有理数，则由于 $a^2 \neq 0$，
$$a = \frac{a^3}{a^2} = \frac{2}{a^2}$$
是有理数。所以 $2^{1/3}$ 是有理数当且仅当 $2^{2/3}$ 是有理数。

无理性。用反证法，假设 a 是有理数。根据等价性，a^2 亦是有理数。定义 T 为所有使 na 与 na^2 同时为整数的正整数 n 所组成的集合。

> ? **想一想** 为何不直接照搬第 (i) 部分，取 $S = \{n : na \text{ 是整数}\}$?

在第 (i) 部分中，验证 $(\sqrt{2}-1)k \in S$ 时需要 $k(\sqrt{2})^2 = 2k$ 是整数，而这是自动成立的。在这里，把 $(a-1)k$ 乘以 a 得到 $ka^2 - ka$，但条件“ka 是整数”并没有告诉我们 ka^2 是整数。所以我们把这一点写进集合的定义之中，而这正是需要本部分前半的地方。

T 非空。记 $a = \frac{p}{q}$，其中 p、q 为正整数。则 $q^2 a = pq$ 及 $q^2 a^2 = p^2$ 都是整数，所以 $q^2 \in T$。

递降。设 k 为 T 的最小元素，则 $u = ka$ 及 $v = ka^2$ 都是整数。考虑 $n = (a-1)k$。

- $n = u - k$ 是整数，而且由于 $a > 1$（因为 $1^3 < 2$），有 $n > 0$。
- $na = ka^2 - ka = v - u$ 是整数。
- 利用 $a^3 = 2$，$na^2 = ka^3 - ka^2 = 2k - v$ 是整数。

所以 $n \in T$。但 $a < 2$（因为 $2^3 = 8 > 2$），所以 $0 < a - 1 < 1$，从而 $n = (a-1)k < k$，与 k 的极小性（minimality）相矛盾。因此 $2^{1/3}$ 是无理数，而根据等价性，$2^{2/3}$ 亦是无

理数。

验算

- 在一个“差一点”的例子上试试递降。分数 $\frac{7}{5}$ 很接近 $\sqrt{2}$；第 5 步中的映射 $(k,m)\mapsto(m-k,\,2k-m)$ 把 $(5,7)$ 映到 $(2,3)$，而 $3^2-2\cdot 2^2=1=-(7^2-2\cdot 5^2)$。量 m^2-2k^2 只会改变正负号，所以假如它等于 0（即 $m=k\sqrt{2}$），它会一直保持为 0，而 k 却不断变小，这是不可能的。
- 数值上，$\sqrt{2}-1\approx 0.414$ 及 $2^{1/3}-1\approx 0.260$，两者都严格介乎 0 与 1 之间，正如递降所需。
- 用电脑搜索所有 $n\le 2\times 10^5$，找不到任何 n 使 $2n^2$ 为完全平方数，亦找不到任何 n 使 $2n^3$ 或 $4n^3$ 为完全立方数。

常见错误

- 忘记证明 $(\sqrt{2}-1)k$ 是正数：S 只包含正整数，所以这一步是必需的。
- 在第 (ii) 部分中，使用只以 na 定义的集合，然后毫无理据地断言 ka^2 是整数。
- 没有说明最小元素 k 为何存在：这需要 S（或 T）非空，而这正是第 3 步的目的。

STEP 之外

这个论证其实关乎正整数的良序性（well-ordering），也就是归纳法（induction）背后的原理，而且它可以推广：若 $\sqrt[r]{N}$ 不是整数，则它是无理数。要看出这一点，可取所有使 $n\alpha, n\alpha^2,\dots,n\alpha^{r-1}$ 全都是整数的 n 所组成的集合，再把其最小元素乘以 $\alpha-\lfloor\alpha\rfloor$，而这个数严格介乎 0 与 1 之间。在第一个代数课程中，这会变成以下命题：整数构成一个整闭（integrally closed）环——整数系数首一多项式（monic polynomial）的有理根必定是整数。Davenport 的 *The Higher Arithmetic*（《高等算术》）是进一步认识整数及其性质的绝佳读物。

详细解答

STEP III 2015，第 6 题

纯数学 · 根的对称函数

STEP III 2015，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/15-s3-q6

关键思路

关于 w、z 这一对数的一切，都由它们的和与积决定，因为 w 和 z 正是 $t^2-(w+z)t+wz=0$ 的两个根。于是第 (i) 部分其实是在问：一个实系数二次方程（quadratic equation）何时有实根；而条件 $u^2\le 2v$ 不过是乔装了的判别式（discriminant）。在第 (ii) 部分，头两条方程确定了积 wz，第三条方程随即变成只含 u 的三次方程（cubic equation），而且有一个一眼就能看出的根。

解答

第 (i) 部分

推动整个证明的恒等式（identity）是

$$2v-u^2=2(w^2+z^2)-(w+z)^2=w^2-2wz+z^2=(w-z)^2.$$

我们亦注意到 $wz=\frac{1}{2}\big((w+z)^2-(w^2+z^2)\big)=\frac{1}{2}(u^2-v)$。

若 w 和 z 是实数：那么 $u=w+z$ 和 $v=w^2+z^2$ 当然是实数，而由于 $w-z$ 是实数，$2v-u^2=(w-z)^2\ge 0$。所以 $u^2\le 2v$。

反过来，假设 u 和 v 是实数，且 $u^2\le 2v$。那么 $p=wz=\frac{1}{2}(u^2-v)$ 亦是实数。由于

$$(t-w)(t-z)=t^2-ut+p,$$

w 和 z 恰好是二次方程 $t^2-ut+p=0$ 的两个根，而这方程的系数都是实数。它的判别式是

$$u^2-4p=u^2-2(u^2-v)=2v-u^2\ge 0,$$

所以两个根都是实数。因此 w 和 z 是实数，等价关系（equivalence）的证明至此完成。

> **想一想** 为甚么单靠“u 和 v 是实数”并不足够?

因为一对共轭复数(conjugate pair)的和与平方和同样是实数。例如 $w=1+2\mathrm{i}$、$z=1-2\mathrm{i}$ 给出 $u=2$ 和 $v=-6$，两者都是实数，但 $u^2=4>-12=2v$。这条不等式正正就是用来排除非实数的共轭对。

第 (ii) 部分

化为三次方程。第一条方程说的是 $u=w+z$。记 $v=w^2+z^2$，第二条方程说 $v=u^2-\frac{2}{3}$，所以

$$wz=\tfrac{1}{2}(u^2-v)=\tfrac{1}{3}.$$

对于立方，我们用标准恒等式

$$w^3+z^3=(w+z)^3-3wz(w+z)=u^3-u.$$

因此第三条方程变成 $u^3-u-\lambda u=-\lambda$，即

$$u^3-(1+\lambda)u+\lambda=0.$$

代入 $u=1$ 得 $1-1-\lambda+\lambda=0$，所以 $u-1$ 是一个因式，而

$$u^3-(1+\lambda)u+\lambda=(u-1)(u^2+u-\lambda).$$

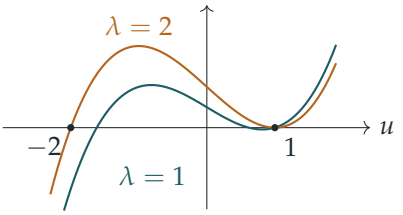
数算根的个数。由于 $\lambda>0$，二次因式的判别式 $1+4\lambda>0$。所以它有两个相异实根：

$$u=\frac{-1\pm\sqrt{1+4\lambda}}{2}.$$

其中一个根等于 1，当且仅当 $1^2+1-\lambda=0$，即 $\lambda=2$。所以：

- 若 $\lambda\neq 2$，u 的可能值为 1 及 $\frac{1}{2}\left(-1\pm\sqrt{1+4\lambda}\right)$：共三个相异的值，全部是实数；
- 若 $\lambda=2$，三次式为 $(u-1)^2(u+2)$，只有两个值：$u=1$ 和 $u=-2$。

这些值每一个都确实可以出现：给定 u，取 w 和 z 为 $t^2-ut+\frac{1}{3}=0$ 的根。那么 $w+z=u$ 且 $wz=\frac{1}{3}$，把上面的步骤倒过来推，便可知三条方程全部成立。所以例外的值是 $\lambda=2$。



这幅草图（已在垂直方向缩放）显示 $u^3-(1+\lambda)u+\lambda$ 在 $\lambda=1$ 时的图像，与横轴有三个分开的交点；以及在 $\lambda=2$ 时的图像，曲线在 $u=1$ 处刚好与横轴相切。

w 和 z 是否必然是实数? 不是。由于 u 是实数，而 $v=u^2-\frac{2}{3}$ 也是实数，第 (i) 部分告诉我们：w 和 z 是实数当且仅当 $u^2\le 2v=2u^2-\frac{4}{3}$，即当且仅当 $u^2\ge\frac{4}{3}$。无论 λ 取何值，根 $u=1$ 都存在，而 $1<\frac{4}{3}$，所以它总是给出非实数的 w 和 z。

具体来说，$u=1$ 给出 $t^2-t+\frac{1}{3}=0$，所以

$$w=\frac{1}{2}+\frac{\mathrm{i}}{2\sqrt{3}},\qquad z=\frac{1}{2}-\frac{\mathrm{i}}{2\sqrt{3}}.$$

验证一下：$w+z=1=u$；$wz=\frac{1}{4}+\frac{1}{12}=\frac{1}{3}$，所以 $w^2+z^2=1-\frac{2}{3}=u^2-\frac{2}{3}$；而 $w^3+z^3=u^3-3wz\,u=1-1=0$，所以对任何 $\lambda>0$ 都有 $w^3+z^3-\lambda u=-\lambda$。这是一个反例（counterexample）：$u$ 的三个值全是实数，但 w 和 z 未必是实数。

验算

- 当 $\lambda=1$ 时，u 的值为 1 及 $\frac{1}{2}(-1\pm\sqrt{5})\approx 0.618,\ -1.618$，每一个都满足 $u^3-2u+1=0$。
- 当 $\lambda=2$ 时，三次式 u^3-3u+2 在 $u=-2$ 处为零（$-8+6+2=0$），而它及其导数（derivative）$3u^2-3$ 在 $u=1$ 处都为零，证实了这是重根（double root）。
- 用电脑对大量 λ 值进行检查：由每个 u 构造出 w 和 z，确认三条方程在舍入误差范围内成立；检查亦发现除了 $\lambda=2$ 以外，u 都有三个相异的值。

常见错误

- 在第 (i) 部分只证明了一个方向。逆命题需要积 wz 是实数，而这正是用到 v 是实数的地方。
- 忘记检查二次因式的根是否与 1 不同，因而漏掉了例外值 $\lambda=2$。
- 以为 u 是实数，w 和 z 就必定是实数。w 和 z 的实值对称函数同样可以来自一对共轭复数。

STEP 之外

一对数（或任何一组数）由其对称函数决定——这个想法是对称多项式（symmetric polynomial）理论的起点：任何关于 w 和 z 的对称多项式都可以用 $w+z$ 和 wz 表示，而牛顿（Newton）恒等式则把 w^3+z^3 这类幂和（power sum）与它们联系起来。同样的推理，配

合判别式，可以判定一个实系数多项式的根何时全是实数。Needham 的 *Visual Complex Analysis*（《可视化复分析》）是学习复数的绝佳良伴，而 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则把对称多项式的讨论推得更远。

详细解答

STEP III 2015，第 7 题

纯数学 · 微分算子

STEP III 2015，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/15-s3-q7

关键思路

x 的幂是 $\mathrm{D} = x\frac{\mathrm{d}}{\mathrm{d}x}$ 的特征函数（eigenfunction）：把 D 作用于 x^a，结果只是乘以 a。因此 D^n 对多项式是逐项作用的，把 x^r 的系数乘以 r^n，而这正是最后那个和式中 r^n 的来源。至于整除性（divisibility），要留意每作用一次 D，因式 $(1-x)$ 的次方最多只会减少一。

解答

开首的结果

对任何常数 a（若 a 不是整数，则需 $x > 0$），

$$\mathrm{D}x^a = x \cdot ax^{a-1} = ax^a.$$

再作用一次 D，并利用这个结果，把常数 a 提出来，

$$\mathrm{D}^2x^a = \mathrm{D}\left(ax^a\right) = a\,\mathrm{D}x^a = a^2x^a,$$

即为所求。由于 D 是线性（linear）的（对常数 α, β，有 $\mathrm{D}(\alpha f + \beta g) = \alpha\,\mathrm{D}f + \beta\,\mathrm{D}g$），重复这一步，由归纳法（induction）可得

$$\mathrm{D}^n x^a = a^n x^a \qquad (n \geq 1).$$

第 (i) 部分

设 $\mathrm{P}(x) = \sum_{k=0}^{r} c_k x^k$，其中 $c_r \neq 0$。由线性及上述结果，

$$\mathrm{D}^n\mathrm{P}(x) = \sum_{k=0}^{r} c_k\, k^n x^k.$$

这是一个没有高于 x^r 的项的多项式，而其 x^r 项的系数为 $r^n c_r$。由于 $r \geq 1$，故 $r^n \neq 0$，又 $c_r \neq 0$，所以 $r^n c_r \neq 0$。因此 $\mathrm{D}^n\mathrm{P}(x)$ 是次数（degree）恰好为 r 的多项式。

（条件 $r \geq 1$ 是必要的：常数多项式会被映到 0，因为 $0^n = 0$。）

第 (ii) 部分

固定 m。我们对 n 用归纳法证明：当 $1 \leq n \leq m$ 时，

$$\mathrm{D}^n(1-x)^m = (1-x)^{m-n}\,\mathrm{Q}_n(x) \quad \text{对某个多项式} \mathrm{Q}_n.$$

基本情况。$\mathrm{D}(1-x)^m = x\cdot\left(-m(1-x)^{m-1}\right) = (1-x)^{m-1}\cdot(-mx)$，所以 $\mathrm{Q}_1(x) = -mx$。

归纳步骤。假设命题对某个满足 $n < m$ 的 n 成立，于是 $m - n \geq 1$。由乘积法则（product rule），

$$\begin{aligned}\mathrm{D}^{n+1}(1-x)^m &= x\frac{\mathrm{d}}{\mathrm{d}x}\Big[(1-x)^{m-n}\mathrm{Q}_n(x)\Big]\\ &= x\Big[-(m-n)(1-x)^{m-n-1}\mathrm{Q}_n(x) + (1-x)^{m-n}\mathrm{Q}_n'(x)\Big]\\ &= (1-x)^{m-n-1}\cdot x\Big[(1-x)\mathrm{Q}_n'(x) - (m-n)\mathrm{Q}_n(x)\Big].\end{aligned}$$

方括号内是一个多项式，所以上式具有所需的形式，其中 $\mathrm{Q}_{n+1}(x) = x\big[(1-x)\mathrm{Q}_n'(x) - (m-n)\mathrm{Q}_n(x)\big]$。

> ? **想一想** 在归纳步骤中，我们在哪里用到了 $n < m$?

只有当 $m - n - 1 \geq 0$，即 $n < m$ 时，提出因式 $(1-x)^{m-n-1}$ 作为多项式因式才有意义。到了 $n = m$，已经再没有 $(1-x)$ 的因式可以减少，所以归纳（理所当然地）在那里停止。

因此对每个正整数 $n < m$，$\mathrm{D}^n(1-x)^m$ 都能被 $(1-x)^{m-n}$ 整除。

第 (iii) 部分

用二项式定理（binomial theorem）展开：

$$(1-x)^m = \sum_{r=0}^{m}(-1)^r\binom{m}{r}x^r.$$

与第 (i) 部分完全一样，把 D^n 逐项作用，

$$\mathrm{D}^n(1-x)^m = \sum_{r=0}^{m}(-1)^r\binom{m}{r}r^n x^r.$$

由第 (ii) 部分，在 $n < m$ 时，左方等于 $(1-x)^{m-n}\mathrm{Q}_n(x)$，且 $m - n \geq 1$，所以它在 $x = 1$ 时等于零。因此在右方代入 $x = 1$，便得

$$\sum_{r=0}^{m}(-1)^r\binom{m}{r}r^n = 0,$$

即为所求。（$r = 0$ 那一项是 $0^n = 0$，与 D 消去常数项一致。）

验算

取 $m = 3$、$n = 2$：和式为 $0 - 3 \cdot 1 + 3 \cdot 4 - 1 \cdot 9 = 0$。取 $m = 4$、$n = 3$：$0 - 4 + 6 \cdot 8 - 4 \cdot 27 + 64 = 0$。我们亦已用电脑验证此恒等式对所有 $1 \le n < m \le 19$ 均成立，并验证了当 $m \le 11$ 时，$(1-x)^{m-n}$ 是整除 $\mathrm{D}^n(1-x)^m$ 的确切次方。作为对照，当 $n = m$ 时和式不再为零：当 $m = 2$ 时，它是 $0 - 2 + 4 = 2 = 2!$，一般而言则是 $(-1)^m m!$。

常见错误

- 把 $\mathrm{D}^2 x^a$ 写成 $x^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} x^a = a(a-1)x^a$：$\mathrm{D}^2$ 是把 D 作用两次，而不是 x^2 乘以二阶导数。
- 在第 (i) 部分，忘记说明为何首项系数（leading coefficient）$r^n c_r$ 不为零；条件 $r \ge 1$ 正是用在这里。
- 在第 (iii) 部分，在作用 D^n 之前就把 $x = 1$ 代入 $(1-x)^m$；重点在于因式 $(1-x)$ 在 $n < m$ 次作用后仍然保留。

STEP 之外

算子（operator）$x\frac{\mathrm{d}}{\mathrm{d}x}$ 有时称为欧拉（Euler）算子，而 x^a 是它的特征函数，对应的特征值（eigenvalue）为 a；它是处理欧拉—柯西方程（Euler–Cauchy equations）的自然工具，例如 $x^2y'' + bxy' + cy = 0$，在代换 $x = e^t$ 下会变成常系数方程。第 (iii) 部分的恒等式说的是：次数低于 m 的多项式，其 m 阶有限差分（finite difference）为零；它与斯特林数（Stirling numbers）以及用容斥原理（inclusion–exclusion）计算满射（surjection）个数密切相关。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）讲述了欧拉方程；Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）收录了更多同类风格的题目。

详细解答

STEP III 2015，第 8 题

纯数学 · 极坐标微分方程

STEP III 2015，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/15-s3-q8

本题在 *Beyond STEP* 第 8.4 节盘旋趋向圆中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch08/limit-cycle.html

关键思路

两条微分方程（differential equations）都由组合 $x\,\mathrm{d}y - y\,\mathrm{d}x$ 和 $x\,\mathrm{d}x + y\,\mathrm{d}y$ 构成，而这正是极坐标（polar coordinates）能够化简的式子：它们分别变成 $r^2\,\mathrm{d}\theta$ 和 $r\,\mathrm{d}r$。因此我们把 θ 视为沿曲线的参数（parameter），把 $\mathrm{d}y/\mathrm{d}x$ 写成两个对 θ 的导数（derivative）之比，让三角部分自行化简。在第 (ii) 部分，新增的项使方程变成关于 r 的伯努利方程（Bernoulli equation），而代换 $u = r^{-2}$ 可把它化为线性（linear）方程。

解答

记 r' 为 $\mathrm{d}r/\mathrm{d}\theta$，$c = \cos\theta$ 及 $s = \sin\theta$。沿曲线 $r = r(\theta)$，

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = r'c - rs, \qquad \frac{\mathrm{d}y}{\mathrm{d}\theta} = r's + rc, \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}\theta}{\mathrm{d}x/\mathrm{d}\theta}.$$

把任一方程两边乘以 $\mathrm{d}x/\mathrm{d}\theta$，形如 $P\dfrac{\mathrm{d}y}{\mathrm{d}x} = Q$ 的方程便变成

$$P\frac{\mathrm{d}y}{\mathrm{d}\theta} - Q\frac{\mathrm{d}x}{\mathrm{d}\theta} = 0.$$

全部工夫都靠以下两条恒等式：

$$x\frac{\mathrm{d}y}{\mathrm{d}\theta} - y\frac{\mathrm{d}x}{\mathrm{d}\theta} = rc(r's + rc) - rs(r'c - rs) = r^2(c^2 + s^2) = r^2,$$
$$x\frac{\mathrm{d}x}{\mathrm{d}\theta} + y\frac{\mathrm{d}y}{\mathrm{d}\theta} = rc(r'c - rs) + rs(r's + rc) = rr'.$$

第 (i) 部分

这里 $P = y + x$ 而 $Q = y - x$，所以

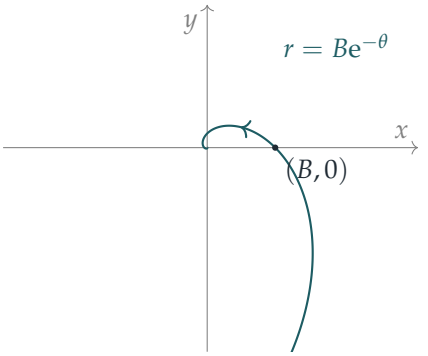
$$\begin{aligned}(y+x)\frac{\mathrm{d}y}{\mathrm{d}\theta} - (y-x)\frac{\mathrm{d}x}{\mathrm{d}\theta} &= \left(x\frac{\mathrm{d}y}{\mathrm{d}\theta} - y\frac{\mathrm{d}x}{\mathrm{d}\theta}\right) + \left(x\frac{\mathrm{d}x}{\mathrm{d}\theta} + y\frac{\mathrm{d}y}{\mathrm{d}\theta}\right)\\ &= r^2 + rr'.\end{aligned}$$

因此方程为 $r(r' + r) = 0$，由于 $r > 0$，可以除以 r，得

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} + r = 0,$$

即所求。

这是可分离变量 (separable) 的方程 (或直接认出它是指数衰减)：$r = B\mathrm{e}^{-\theta}$，其中 $B > 0$。这是一条等角螺线 (equiangular spiral)，亦称对数螺线 (logarithmic spiral)。当 θ 增加时，曲线以逆时针方向绕向原点，但永不到达；当 θ 减少时，它无界地向外盘旋。由于 $r/r' = -1$，曲线与每条由原点出发的射线都以相同角度相交，即与指向原点的径向成 45°。不同的 B 给出同一螺线的旋转，因为 $B\mathrm{e}^{-\theta} = \mathrm{e}^{-(\theta - \ln B)}$。



箭头表示 θ 增加的方向。螺线每绕一圈便缩小 $\mathrm{e}^{2\pi} \approx 535$ 倍，所以在任何比例下都只看得见大约一圈；其余部分都紧密地挤在原点周围。

第 (ii) 部分

现在 $P = y + x - x(x^2 + y^2)$ 而 $Q = y - x - y(x^2 + y^2)$，其中 $x^2 + y^2 = r^2$。新增的项贡献

$$-r^2\left(x\frac{\mathrm{d}y}{\mathrm{d}\theta} - y\frac{\mathrm{d}x}{\mathrm{d}\theta}\right) = -r^2 \cdot r^2 = -r^4,$$

所以利用第 (i) 部分处理原有的项，方程变成

$$r^2 + rr' - r^4 = 0 \quad \Longrightarrow \quad \frac{\mathrm{d}r}{\mathrm{d}\theta} = r^3 - r,$$

同样是除以 $r > 0$ 所得。

想一想 r^3 这一项令方程不再是线性的，我们该如何处理它?

这是一条伯努利方程。目标答案以 r^2 表示，提示我们使用 $u = r^{-2}$。于是

$$\frac{\mathrm{d}u}{\mathrm{d}\theta} = -2r^{-3}\frac{\mathrm{d}r}{\mathrm{d}\theta} = -2r^{-3}(r^3 - r) = -2 + 2u.$$

所以 $u' - 2u = -2$，这是一条常系数线性方程：$u = 1$ 是一个特解（particular solution），而 $A\mathrm{e}^{2\theta}$ 是齐次方程（homogeneous equation）的解，因此

$$\frac{1}{r^2} = 1 + A\mathrm{e}^{2\theta} \qquad \Longrightarrow \qquad r^2 = \frac{1}{1 + A\mathrm{e}^{2\theta}},$$

这正是所求的形式。（分离变量，并对 $\dfrac{1}{r(r-1)(r+1)}$ 使用部分分式（partial fractions），亦可得出相同结果，只是较费工夫。）曲线只在 $1 + A\mathrm{e}^{2\theta} > 0$ 的地方存在。

解的形状

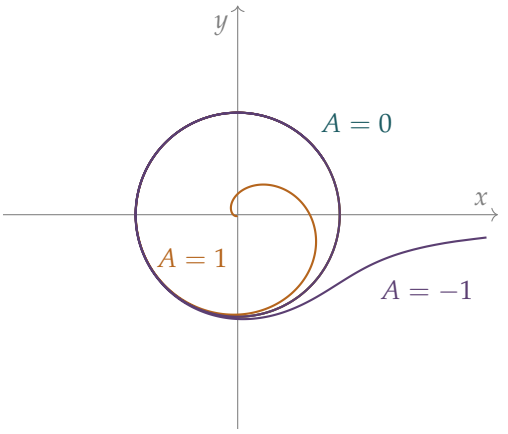
首先留意，若 $A = \pm\mathrm{e}^{-2\alpha}$，则 $A\mathrm{e}^{2\theta} = \pm\mathrm{e}^{2(\theta-\alpha)}$，所以对应 A 的曲线，就是对应 $A = \pm 1$ 的曲线旋转 α 所得。因此按 A 的正负号，实际上只有三种形状。

- $A = 0$：$r = 1$，即单位圆。
- $A > 0$：$1 + A\mathrm{e}^{2\theta} > 1$，所以 $r < 1$，而曲线对每个 θ 都有定义。由于 $r' = r(r^2 - 1) < 0$，r 稳定地递减。当 $\theta \to -\infty$ 时，$r \to 1$，所以曲线从内侧向外盘旋趋向单位圆；当 $\theta \to +\infty$ 时，$r \approx A^{-1/2}\mathrm{e}^{-\theta} \to 0$，所以它盘旋进入原点，看起来愈来愈像第 (i) 部分的螺线。
- $A < 0$：我们需要 $\mathrm{e}^{2\theta} < -1/A$，即 $\theta < \theta_0 = \frac{1}{2}\ln(-1/A)$。此时 $r > 1$ 且 $r' > 0$。当 $\theta \to -\infty$ 时，$r \to 1$ 从上方趋近，所以曲线从外侧向内盘旋趋向单位圆；当 $\theta \to \theta_0^-$ 时，$r \to \infty$。在 θ_0 附近，$1 + A\mathrm{e}^{2\theta} \approx 2(\theta_0 - \theta)$（因为 $A\mathrm{e}^{2\theta_0} = -1$），所以 $r \approx \big(2(\theta_0 - \theta)\big)^{-1/2}$，而曲线到直线 $\theta = \theta_0$ 的距离为

  $$r\sin(\theta_0 - \theta) \approx r(\theta_0 - \theta) \approx \frac{1}{2r} \to 0.$$

  因此曲线趋向无穷远，并以通过原点、倾角为 θ_0 的直线为渐近线（asymptote），从顺时针一侧趋近它。当 $A = -1$ 时，$\theta_0 = 0$，而对于大的 x，曲线最终贴近 $y = -\frac{1}{2x}$。

单位圆是一个极限环（limit cycle）：当 $\theta \to -\infty$ 时，其他每个解都盘旋趋向它；当 θ 增加时，则盘旋离开它。



紫红色曲线（$A=-1$）沿正 x 轴、紧贴其下方伸向无穷远。

验算

- 把 $r=B\mathrm{e}^{-\theta}$ 代回第 (i) 部分的原方程，在几个点上检查（换算成 x、y 和 $\mathrm{d}y/\mathrm{d}x$），结果为零；对第 (ii) 部分的公式，取若干正负不同的 A 值检查，结果亦然。
- $A=0$ 给出 $r=1$，而常数 $r=1$ 确实满足 $r'=r^3-r$；以 x、y 表示，圆 $x^2+y^2=1$ 使新增的项与原有的项恰好抵消。
- 从 $r(0)=\frac{1}{2}$ 出发（即 $A=3$），公式给出 $r(2)=1/\sqrt{1+3\mathrm{e}^4}\approx 0.0779$，与 $r'=r^3-r$ 的逐步数值解在许多位小数上吻合。
- 对于 $A=-1$，在 $\theta=-10^{-4}$ 处，$r\approx 70.71$，而到 x 轴的距离为 $0.00707\approx 1/(2r)$，印证了渐近线。

常见错误

- 忘记 $\mathrm{d}y/\mathrm{d}x$ 必须在 r 会变化的情况下以 $(\mathrm{d}y/\mathrm{d}\theta)/(\mathrm{d}x/\mathrm{d}\theta)$ 求得，即漏掉了含 r' 的项。
- 当 $A<0$ 时忽略了限制 $1+A\mathrm{e}^{2\theta}>0$，因而没有察觉这些曲线在有限的角度便趋向无穷远。
- 对每种 A 的正负号只画一条曲线，却没有说明其他曲线都是它的旋转。

STEP 之外

方程 $r' = r^3 - r$ 是一个平面动力系统（dynamical system）的极坐标描述，而单位圆是一个极限环：一条孤立的闭轨道，邻近的轨迹都盘旋趋向它或离开它。这类径向方程可以独立分离出来的系统，是非线性微分方程和相平面分析（phase-plane analysis）课程中的标准入门例子，并引向庞加莱–本迪克森定理（Poincaré–Bendixson theorem）。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）有一章很好地讲解非线性系统与极限环。

详细解答

STEP III 2015，第 9 题

力学 · 非线性弹性振动

STEP III 2015，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/15-s3-q9

关键思路

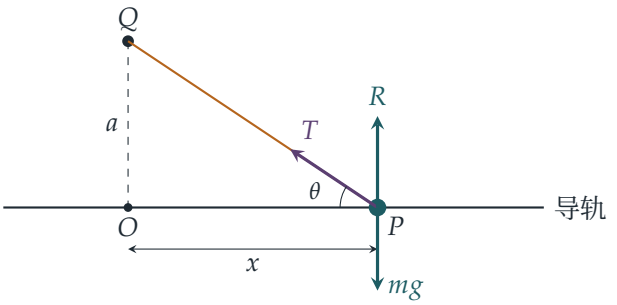
导轨光滑(smooth),法向反作用力(normal reaction)不做功,所以能量守恒(conservation of energy)，由此即可直接得到一阶方程（first-order equation）。由于在导轨上每一点都有 $PQ \geq a$，弦从不松弛（slack）。转向点（turning point）就是 $\dot{x} = 0$ 的位置。周期（period）是由中心走到 x_0 所需时间的四倍。当 v 很小时，质点一直停留在起点附近，所以 $\sqrt{x^2 + a^2} - a \approx x^2/(2a)$，而代换（substitution）$x = x_0 u$ 则可把积分整理好。

解答

能量方程

把 P 视为质点,弦视为轻而有弹性、遵从胡克定律(Hooke's law)。取导轨为 x 轴, P 从原点出发, 原点即由 Q 所作垂线的垂足。当 P 位于 x 时, 它与 Q 的距离为 $PQ = \sqrt{x^2 + a^2} \geq a$, 所以弦是拉紧的，其伸长量（extension）为

$$e = \sqrt{x^2 + a^2} - a \quad (\geq 0).$$



作用于 P 的力有其重量、导轨的法向反作用力 R（两者都与运动方向垂直，所以不做功），以及张力（tension）T。张力是保守力（conservative force），其弹性势能（elastic potential energy）为 $\lambda e^2/(2a)$。开始时 $e=0$，速率为 v，所以能量守恒给出

$$\tfrac{1}{2}m\dot{x}^2+\frac{\lambda}{2a}\left(\sqrt{x^2+a^2}-a\right)^2=\tfrac{1}{2}mv^2.$$

两边乘以 $2/m$，并记 $k^2=\lambda/(ma)$，得

$$\dot{x}^2=v^2-k^2\left(\sqrt{x^2+a^2}-a\right)^2,$$

此即所求。

最大位移及该处的加速度

伸长量 e 随 $|x|$ 增大而增大，所以 P 向外移动时 $\dot{x}^2$ 递减，而 P 会一直向外移动，直至 $\dot{x}=0$。在该点 $k\left(\sqrt{x_0^2+a^2}-a\right)=v$（由于 $e\geq 0$，取正根）。因此

$$\sqrt{x_0^2+a^2}=a+\frac{v}{k}\quad\Longrightarrow\quad x_0^2=\left(a+\frac{v}{k}\right)^2-a^2=\frac{2av}{k}+\frac{v^2}{k^2},$$

得

$$x_0=\frac{\sqrt{v\,(2ak+v)}}{k}.$$

求加速度时，沿导轨方向分解。设 θ 为 PQ 与导轨之间的夹角，则 $\cos\theta=x/\sqrt{x^2+a^2}$，而张力沿导轨的分量为 $T\cos\theta$，指向 O。由 $T=\lambda e/a$，

$$m\ddot{x}=-\frac{\lambda}{a}\left(\sqrt{x^2+a^2}-a\right)\frac{x}{\sqrt{x^2+a^2}},\qquad\text{所以}\qquad\ddot{x}=-k^2\left(\sqrt{x^2+a^2}-a\right)\frac{x}{\sqrt{x^2+a^2}}.$$

（把能量方程对 t 求导，再约去 $2\dot{x}$，亦得到相同结果。）在 $x=x_0$ 处有 $\sqrt{x_0^2+a^2}-a=v/k$ 及 $\sqrt{x_0^2+a^2}=(ak+v)/k$，所以

$$\ddot{x}=-k^2\cdot\frac{v}{k}\cdot\frac{k\,x_0}{ak+v}=-\frac{k^2v\,x_0}{ak+v}=-\frac{kv\sqrt{v\,(2ak+v)}}{ak+v}.$$

因此在 x_0 处加速度的大小为 $\dfrac{kv\sqrt{v(2ak+v)}}{ak+v}$，方向沿导轨指回初始位置。

周期

$\dot{x}^2$ 的方程只依赖于 x^2，所以运动关于 O 对称。质点在 $-x_0$ 与 x_0 之间振荡，而 $O\to x_0\to O\to -x_0\to O$ 这四段各需相同的时间。在第一段中 $\dot{x}>0$，所以 $\mathrm{d}t=\mathrm{d}x/\dot{x}$，周期为

$$\tau=4\int_0^{x_0}\frac{\mathrm{d}x}{\sqrt{v^2-k^2\left(\sqrt{x^2+a^2}-a\right)^2}}.$$

被积函数（integrand）在 x_0 处趋于无穷，但只是如 $(x_0-x)^{-1/2}$ 般增长，所以积分收敛（converge）。

> **想一想** $v \ll ka$ 告诉我们 x 的大小是怎样的?

由于 $x_0^2 = 2av/k + v^2/k^2$，若 $v/k \ll a$，则 $x_0^2 \approx 2av/k$，所以 $x_0^2/a^2 \approx 2v/(ka) \ll 1$。因此积分范围内每个 x 都满足 $|x| \ll a$，而二项式展开（binomial expansion）给出

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

周期变为

$$\tau \approx 4\int_0^{x_0} \frac{\mathrm{d}x}{\sqrt{v^2 - k^2x^4/(4a^2)}}, \qquad x_0 \approx \sqrt{\frac{2av}{k}},$$

而这个近似的 x_0 正好是新分母等于零的位置，因为 $k^2x_0^4/(4a^2) = v^2$。现在代入 $x = x_0u$，于是 $\mathrm{d}x = x_0\,\mathrm{d}u$ 及 $k^2x^4/(4a^2) = v^2u^4$：

$$\tau \approx 4\int_0^1 \frac{x_0\,\mathrm{d}u}{v\sqrt{1-u^4}} = \frac{4}{v}\sqrt{\frac{2av}{k}}\int_0^1 \frac{\mathrm{d}u}{\sqrt{1-u^4}} = \sqrt{\frac{32a}{kv}}\int_0^1 \frac{\mathrm{d}u}{\sqrt{1-u^4}},$$

因为 $\frac{16}{v^2}\cdot\frac{2av}{k} = \frac{32a}{kv}$。这就是所求的近似式。

验算

- 单位：$k^2 = \lambda/(ma)$ 的单位是 s^{-2}，所以 v/k 是长度（理应如此，因为它是 x_0 处的伸长量），而 $\sqrt{a/(kv)}$ 是时间。
- 取 $m=1$、$\lambda=2$、$a=1$、$v=1$，得 $k=\sqrt{2}$、$x_0 = \sqrt{1+2\sqrt{2}} \approx 1.3836$，以及 x_0 处的加速度 ≈ -1.1462。以数值方法逐步推进运动方程，得到相同的转向点，周期约为 7.612，与积分结果吻合。
- 取 $a=1$、$k=2$、$v=0.001$，精确的周期积分为 165.862，而近似式给出 165.833，其中 $\int_0^1 (1-u^4)^{-1/2}\,\mathrm{d}u \approx 1.3110$。当 $v \to 0$ 时，两者之比趋于 1。

常见错误

- 把伸长量当作 x 或 $\sqrt{x^2+a^2}$，而不是 $\sqrt{x^2+a^2}-a$；弦开始时处于自然长度（natural length）。
- 忘记分解张力：只有分量 $T\cos\theta$ 沿导轨作用。
- 把周期当作 $2\int_0^{x_0}$ 或 $\int_0^{x_0}$；振荡是在 $-x_0$ 与 x_0 之间进行的，所以需要四个四分之一周期。

STEP 之外

当 v 很小时，回复力（restoring force）大致与 x^3 而非 x 成正比，所以这是一个非线性振子（nonlinear oscillator），其周期依赖于振幅（amplitude）：与简谐运动（simple harmonic motion）不同，能量越小，周期越大，按 $v^{-1/2}$ 增长。积分 $\int_0^1 (1-u^4)^{-1/2}\,\mathrm{d}u$ 是一个椭圆积分（elliptic integral）（等于双纽线常数（lemniscate constant）的一半），摆（pendulum）的精确周期也会出现这类积分。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）以同样的精神讲解能量方法和振荡。

详细解答

STEP III 2015，第 10 题

力学 • 转动的刚杆

STEP III 2015，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：

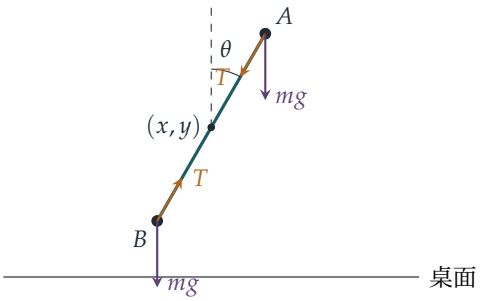
step.maths.org/questions/15-s3-q10

关键思路

用中点的运动和杆的转动来描述整个系统。以 x、y 和 θ 写出每个质点（particle）的位置，对时间求导两次，再对每个质点应用牛顿第二定律（Newton's second law）。把两条方程相加便消去张力（tension），得出质心（centre of mass）的运动。把两条方程相减，再沿垂直于杆的方向分解，又一次消去张力，剩下的就是转动。三个加速度都求出后，最后一部分就是抛射体运动（projectile motion）加上匀速转动。

解答

我们把两个质量模型化为由一根轻质刚性杆连接的质点。每个质点只受两个力：本身的重量，以及沿杆方向作用的杆的力。设 T 为张力，$T > 0$ 表示杆把每个质点拉向中点。（杆也可以推，此时 $T < 0$；以下的推导与 T 的正负无关。）



质点 A 的运动方程

设 A 为杆上起初位于中点上方那一端的质点，因此由中点指向 A 的向量（vector）是 $a(\sin\theta, \cos\theta)$。于是 A 的位置是

$$\mathbf{r}_A = \begin{pmatrix} x + a\sin\theta \\ y + a\cos\theta \end{pmatrix}, \qquad \text{所以} \qquad \dot{\mathbf{r}}_A = \begin{pmatrix} \dot{x} + a\dot{\theta}\cos\theta \\ \dot{y} - a\dot{\theta}\sin\theta \end{pmatrix}$$

这是由链式法则（chain rule）得出的，也就是题目要求的速度。再求导一次，对 $\dot{\theta}\cos\theta$ 和 $\dot{\theta}\sin\theta$ 使用积法则（product rule），得

$$\ddot{\mathbf{r}}_A = \begin{pmatrix} \ddot{x} + a\ddot{\theta}\cos\theta - a\dot{\theta}^2\sin\theta \\ \ddot{y} - a\ddot{\theta}\sin\theta - a\dot{\theta}^2\cos\theta \end{pmatrix}.$$

杆以大小为 T 的力，沿 $-(\sin\theta, \cos\theta)$ 方向把 A 拉向中点。重力贡献 $-mg(0,1)$。因此对 A 应用牛顿第二定律，正好得出题目所给的方程：

$$m\begin{pmatrix} \ddot{x} + a\ddot{\theta}\cos\theta - a\dot{\theta}^2\sin\theta \\ \ddot{y} - a\ddot{\theta}\sin\theta - a\dot{\theta}^2\cos\theta \end{pmatrix} = -T\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} - mg\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{A}$$

另一个质点

对于位于另一端的 B，把所有 a 换成 $-a$ 即可。此时杆沿 $+(\sin\theta, \cos\theta)$ 方向拉 B：

$$m\begin{pmatrix} \ddot{x} - a\ddot{\theta}\cos\theta + a\dot{\theta}^2\sin\theta \\ \ddot{y} + a\ddot{\theta}\sin\theta + a\dot{\theta}^2\cos\theta \end{pmatrix} = T\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} - mg\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{B}$$

这里 B 的速度是 $(\dot{x} - a\dot{\theta}\cos\theta,\ \dot{y} + a\dot{\theta}\sin\theta)$。

推导加速度

把 (A) 和 (B) 相加，所有含 a 或 T 的项都会抵消：

$$2m\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = -2mg\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \text{所以} \qquad \ddot{x} = 0, \quad \ddot{y} = -g.$$

以 (A) 减去 (B)，再除以 2：

$$ma\ddot{\theta}\cos\theta - ma\dot{\theta}^2\sin\theta = -T\sin\theta,$$
$$-ma\ddot{\theta}\sin\theta - ma\dot{\theta}^2\cos\theta = -T\cos\theta.$$

想一想 我们不知道 T，该怎样把它消去？

张力沿杆的方向作用，所以我们沿垂直于杆的方向 $(\cos\theta, -\sin\theta)$ 分解，即取第一条方程乘以 $\cos\theta$，减去第二条方程乘以 $\sin\theta$。右边得出 $-T\sin\theta\cos\theta + T\cos\theta\sin\theta = 0$，而 $\dot\theta^2$ 项亦以同样方式抵消。剩下的是

$$ma\ddot\theta(\cos^2\theta + \sin^2\theta) = 0,$$

由于 $m, a \neq 0$，得 $\ddot\theta = 0$。（若改为沿杆的方向分解，则得 $T = ma\dot\theta^2$，即绕中点作圆周运动所需的张力。）

最后一部分

在 $t = 0$ 时杆是铅直的，所以 $\theta = 0$。此时质点 A 在较高位置，速度为 $(\dot x + a\dot\theta,\ \dot y)$；$B$ 在较低位置，速度为 $(\dot x - a\dot\theta,\ \dot y)$。与题目所给的初速度比较：

$$\dot x + a\dot\theta = u, \qquad \dot x - a\dot\theta = 0, \qquad \dot y = v,$$

所以起初 $\dot x = \frac{1}{2}u$、$\dot\theta = \dfrac{u}{2a}$ 及 $\dot y = v$。

由上一部分，$\dot\theta$ 保持不变，所以 $\theta = \dfrac{ut}{2a}$。我们取 $u > 0$，使 θ 递增。此外 $\ddot y = -g$ 且 $y(0) = h$，所以

$$y = h + vt - \tfrac{1}{2}gt^2.$$

当 $\dfrac{ut}{2a} = \dfrac{\pi}{2}$，即 $t = \dfrac{\pi a}{u}$ 时，杆转过了 $\frac{1}{2}\pi$。此时杆是水平的，所以两个质点都位于中点的高度 y。两者在那一刻同时碰到桌面，当且仅当那时 $y = 0$：

$$h + v\cdot\frac{\pi a}{u} - \frac{1}{2}g\cdot\frac{\pi^2a^2}{u^2} = 0.$$

乘以 $2u^2$ 并整理，得

$$2hu^2 = \pi^2a^2g - 2\pi uva,$$

此即所求。

验算

- 当 $v = 0$ 时，下落时间为 $\sqrt{2h/g}$，这必须等于 $\pi a/u$。两边平方得 $2hu^2 = \pi^2a^2g$，与结果吻合。
- 单位：结果中每一项的量纲（dimensions）都是 $\mathrm{L^3T^{-2}}$。
- 我们用电脑模拟了两个由刚性杆连接的质点，随机选取多组 a、u、v，取 $g = 9.8$，并由公式给出 h。例如 $a \approx 1.754$ 时 $h \approx 17.93$。在每一种情况下，两个质点都在 $t = \pi a/u$ 时同时到达桌面，误差在 10^{-10} 以内，而且杆是水平的。

常见错误

- 把其中一个质点所受张力的方向弄错。杆把每个质点拉向中点，所以两个力大小相等、方向相反。
- 在两条方程中使用不同的 T。杆是轻质的，所以它所受的合力必须为零，两端的张力相同。
- 对 $\dot{\theta}\cos\theta$ 求导时漏掉 $-a\dot{\theta}^2$ 项。缺少这些项，相减那一步就不能干净地得出结果。

STEP 之外

这道题是关于质点系的一条一般定理的小小例子。质心的运动，就如同全部质量和全部外力都集中在质心一样。内力两两抵消，所以若对质心没有外力矩（torque），则绕质心的角动量（angular momentum）守恒。在大学一年级的力学课程中，你还会学到拉格朗日方法（Lagrangian approach），它把 x、y、θ 当作广义坐标（generalised coordinates），完全不需要用到 T。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对质心和刚体转动有详尽的论述。

详细解答

STEP III 2015，第 11 题

力学 · 圆周运动

STEP III 2015，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/15-s3-q11

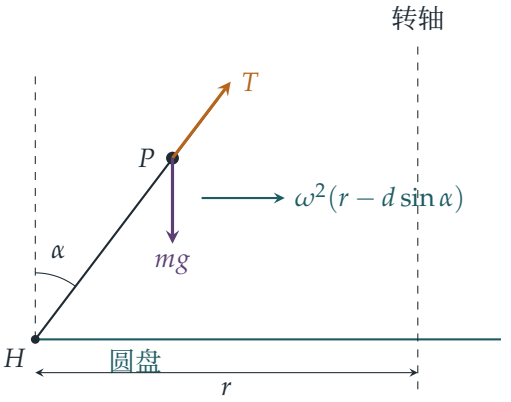
关键思路

轻杆（light rod）没有质量，所以即使它在运动，作用在它上面的力和力矩（moment）都必须完全平衡。在第 (i) 部分，只有两个力作用在杆上，而来自铰链（hinge）的力对铰链没有力矩，所以来自 P 的力必须正正指向铰链。之后，P 就是一个作水平圆周运动的质点，余下的只需分解力（resolving）便可完成。在第 (ii) 部分，两个质点施加的力不一定沿着杆的方向，但对铰链取力矩，仍然可以得到一条关于 β 的方程。

解答

模型。质点是点质量，杆是轻而刚性的，铰链是光滑的，所以它不会对自身的轴（即通过铰链、垂直于包含杆的铅垂平面的水平直线）施加力偶（couple）。在该平面上设立坐标轴，原点位于铰链 H，x 轴水平并指向转轴，y 轴铅垂向上。杆上与 H 距离为 s 的点位于 $(s\sin\theta,\ s\cos\theta)$，其中 θ 是杆与向上铅垂线所成的角，而该点与转轴的水平距离为 $r - s\sin\theta$。

第 (i) 部分



上图画出了作用在 P 上的力：它的重量，以及来自杆的力；我们稍后会证明后者沿着杆的方向作用。青色箭头表示 P 的加速度。

来自 P 的力沿着杆的方向。只有两个力作用在杆上：来自铰链、作用于 H 的力，以及来自 P、作用于 P 的力 $\mathbf{G}$。杆是轻的，所以无论它如何运动，它所受关于铰链轴的总力矩都是零（质量乘以角加速度等于零）。铰链力对 H 没有力矩，所以 $\mathbf{G}$ 对 H 也没有力矩。

其次，P 以转轴为圆心作水平圆周运动，所以它的加速度是水平的，并指向转轴，而转轴位于杆所在的铅垂平面上。P 的重量也在该平面上，所以杆对 P 施加的力，以及因此 $\mathbf{G}$（根据牛顿第三定律（Newton's third law），两者大小相等、方向相反），也位于该平面上。该平面上一个对 H 力矩为零的力，其作用线必通过 H。这条作用线也通过位于杆上的 P，所以 $\mathbf{G}$ 与杆平行，得证。

角度。因此，杆沿着杆的方向以力 T 推或拉 P；我们取 T 的方向为由 H 指向 P，即分量为 $(T\sin\alpha,\ T\cos\alpha)$。$P$ 与转轴的距离为 $r - d\sin\alpha$，由于 $d\sin\alpha < r$，这个距离是正数，所以 P 与 H 位于转轴的同一侧，而它的加速度 $\omega^2(r - d\sin\alpha)$ 指向 $+x$ 方向。对 P 应用牛顿第二定律（Newton's second law）：

$$\begin{aligned} &\text{铅垂方向：} && T\cos\alpha - mg = 0, \\ &\text{水平方向：} && T\sin\alpha = m\omega^2(r - d\sin\alpha). \end{aligned}$$

将第二式除以第一式（我们有 $\cos\alpha \neq 0$，否则 $T\cos\alpha = 0 \neq mg$）：

$$\tan\alpha = \frac{\omega^2(r - d\sin\alpha)}{g} = \frac{r - d\sin\alpha}{a}.$$

两边乘以 $a\cot\alpha$，得 $a = r\cot\alpha - d\cos\alpha$，即

$$r\cot\alpha = a + d\cos\alpha,$$

得证。

铰链力。第一式给出 $T = mg/\cos\alpha > 0$，所以杆把 P 推离铰链：杆处于受压（compression）状态。根据牛顿第三定律，P 沿着杆的方向以力 T 把杆推向 H。杆是轻的，所以它所受的力总和为零，而铰链对杆的力与此大小相等、方向相反：它**沿着杆的方向，由铰链指向** P（向上，并向内指向转轴），大小为

$$\frac{mg}{\cos\alpha} = mg\sec\alpha.$$

第 (ii) 部分

现在质量为 m_i $(i = 1, 2)$ 的质点位于 $(d_i\sin\beta,\ d_i\cos\beta)$，与转轴的距离为 $r - d_i\sin\beta > 0$。

> **想一想** 现在每个质点都以某个未知的方向拉杆。我们怎样才能避免逐一求出这些力？

我们不需要逐一求出它们：对每个质点应用牛顿第二定律，可得杆对它施加的力，然后对轻杆关于 H 列出力矩方程，便能消去未知的铰链力。设 $\mathbf{F}_i$ 为杆对质点 i 施加的力。质点 i 的加速度为 $+x$ 方向的 $\omega^2(r - d_i\sin\beta)$，所以

$$\mathbf{F}_i + (0, -m_ig) = \big(m_i\omega^2(r - d_i\sin\beta),\ 0\big) \quad\Longrightarrow\quad \mathbf{F}_i = \big(m_i\omega^2(r - d_i\sin\beta),\ m_ig\big).$$

根据牛顿第三定律，质点 i 在点 $(x_i, y_i) = (d_i\sin\beta,\ d_i\cos\beta)$ 对杆施加 $-\mathbf{F}_i$。作用于 (x, y) 的力 (X, Y) 关于 H 的逆时针力矩为 $xY - yX$，所以 $-\mathbf{F}_i$ 的力矩为

$$-d_i\sin\beta\cdot m_ig + d_i\cos\beta\cdot m_i\omega^2(r - d_i\sin\beta).$$

铰链力对 H 没有力矩，而杆是轻的，所以这两个力矩之和为零。除以 $\omega^2\sin\beta$（它不等于零：若 $\beta = 0$，力矩方程会变成 $\sum m_id_i\omega^2 r = 0$，这是错的），并利用 $g/\omega^2 = a$：

$$\sum_{i=1}^{2} m_id_i\big(-a + r\cot\beta - d_i\cos\beta\big) = 0,$$

即

$$(m_1d_1 + m_2d_2)\,r\cot\beta = (m_1d_1 + m_2d_2)\,a + (m_1d_1^2 + m_2d_2^2)\cos\beta.$$

除以 $m_1d_1 + m_2d_2 > 0$，

$$r\cot\beta = a + b\cos\beta, \qquad b = \frac{m_1d_1^2 + m_2d_2^2}{m_1d_1 + m_2d_2}.$$

验算

- 只有一个质点时（例如 $m_2 = 0$），得 $b = d_1$，与第 (i) 部分一致。若 $d_1 = d_2 = d$，则同样有 $b = d$：位于同一点的两个质点，作用就如一个质点。

- b 是 d_1 与 d_2 的加权平均（权为 $m_i d_i$），所以它介乎两者之间，理应如此。
- 当 $\omega \to 0$ 时，$a \to \infty$，所以 $\cot\alpha \to \infty$，杆趋向铅垂，这正是圆盘几乎不转动时我们所预期的。当 $\omega \to \infty$ 且 $d > r$ 时，$a \to 0$，方程趋向 $r = d\sin\alpha$：质点趋近转轴，在那里它几乎不需要向心力。
- 取 $r = 1\,\mathrm{m}$、$\omega = 5\,\mathrm{rad\,s^{-1}}$、$d = 0.6\,\mathrm{m}$ 及 $g = 9.81\,\mathrm{m\,s^{-2}}$，得 $a = 0.3924\,\mathrm{m}$，而方程给出 $\alpha \approx 0.9251$（约 $53.0°$）；于是 $r\cot\alpha \approx 0.7538 \approx a + d\cos\alpha$。以电脑模拟杆的运动，令它在这个角度由静止开始，它会一直停留在该处。

常见错误

- 在第 (ii) 部分假设每个质点对杆的力都沿着杆的方向。这只是单一质点时的特殊情况；有两个质点时，只有关于 H 的总力矩为零。
- 把质点圆周的半径误用为 r（或 $d\sin\alpha$），而不是 $r - d\sin\alpha$。
- 把铰链力说成指向铰链：杆处于受压状态，所以铰链把杆推向 P。

STEP 之外

第 (ii) 部分的力矩方程，正是有效势能（effective potential）$\sum m_i\left(gd_i\cos\theta - \frac{1}{2}\omega^2(r - d_i\sin\theta)^2\right)$ 取驻值的条件：在旋转参考系中，离心力的表现就像一个额外的势能。这是分析离心调速器（centrifugal governor）等旋转系统，以及探讨稳态是否稳定（stable）的起点。在这里，有效势能在 $\theta = \alpha$ 处取极大值，所以第 (i) 部分的稳态其实是不稳定的：只要轻轻推一下杆，它便会摆离该位置。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）涵盖了刚体力矩和旋转参考系。

详细解答

STEP III 2015，第 12 题

概率与统计 · 概率生成函数

> **STEP III 2015，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/15-s3-q12

关键思路

除以 m 的余数（remainder）只取决于指数模 m 的值，所以在余数的生成函数（generating function）中，凡是出现 x^m 的地方都可以换成 1。把独立（independent）的投掷结果相加，相当于把生成函数相乘，之后再把 x 的幂次约简一次。有了这条规则，整个计算就变成简单的多项式代数，其中 $x \cdot y$ 和 y^2 都会化为 y 的倍数。

解答

全文中，设 $X_1, X_2, \ldots$ 为各次投掷的结果，它们互相独立，且各自在 $\{1, \ldots, 6\}$ 上服从均匀分布（uniform）。

约简规则。 设 m 为正整数。若 U 与 V 互相独立，且取值于 $\{0, 1, \ldots, m-1\}$，则 $U+V$ 的概率生成函数（probability generating function, pgf）是 $G_U(x)G_V(x)$，这是一个多项式，其中 x^k 项的系数就是 $U+V=k$ 的概率。$U+V$ 除以 m 的余数等于 j，当且仅当 $U+V$ 等于 j 或 $j+m$。因此，只要对 $k \geq m$ 把每个 x^k 换成 x^{k-m}，换言之令 $x^m = 1$，便得到该余数的 pgf。由于 S_n 的余数就是（S_{n-1} 的余数）+（X_n 的余数）的余数，而这两者互相独立，所以每个余数的 pgf 都可由前一个相乘再令 $x^m = 1$ 得到。我们以 $\equiv$ 表示“在使用 $x^m = 1$ 后相等”。

第 (i) 部分

掷出 $1, \ldots, 5$ 的余数分别是 $1, \ldots, 5$，而掷出 6 的余数是 0。每种情况的概率都是 $\frac{1}{6}$，所以

$$G(x) = \tfrac{1}{6}\left(1 + x + x^2 + x^3 + x^4 + x^5\right) = \tfrac{1}{6}z, \qquad \text{其中} z = 1 + x + \cdots + x^5.$$

在 $x^6 = 1$ 之下，把 z 乘以 x 只是把每一项往前转一格：$xz = x + x^2 + \cdots + x^5 + x^6 \equiv z$。因此对每个 k 都有 $x^k z \equiv z$，而且

$$z^2 = \sum_{k=0}^{5} x^k z \equiv 6z.$$

所以 R_2 的 pgf 是

$$\tfrac{1}{36} z^2 \equiv \tfrac{1}{36} \cdot 6z = \tfrac{1}{6} z = \mathrm{G}(x),$$

正如所求。

同样的步骤在每一步都适用：若 R_{n-1} 的 pgf 是 $\mathrm{G}(x)$，则 R_n 的 pgf 是 $\mathrm{G}(x)^2 \equiv \mathrm{G}(x)$。由归纳法（induction），每个 R_n 的 pgf 都是 $\mathrm{G}(x) = \frac{1}{6} z$。$6 \mid S_n$ 的概率就是 $\mathrm{P}(R_n = 0)$，即 $\mathrm{G}(x)$ 的常数项（constant term）：

$$\mathrm{P}(S_n \text{ 可被整除，除数为6}) = \tfrac{1}{6}.$$

第 (ii) 部分

现在 $m = 5$。掷出 $1, 2, 3, 4$ 的余数是 $1, 2, 3, 4$；掷出 5 的余数是 0；掷出 6 的余数是 1。所以余数为 1 的概率是 $\frac{2}{6}$，其余各为 $\frac{1}{6}$：

$$\mathrm{G}_1(x) = \tfrac{1}{6}\left(1 + 2x + x^2 + x^3 + x^4\right) = \tfrac{1}{6}(x + y), \qquad y = 1 + x + x^2 + x^3 + x^4.$$

与第 (i) 部分完全一样，在 $x^5 = 1$ 之下有 $xy \equiv y$，所以对所有 k 都有 $x^k y \equiv y$，而且

$$y^2 = \sum_{k=0}^{4} x^k y \equiv 5y.$$

因此

$$\begin{aligned} \mathrm{G}_2(x) &\equiv \tfrac{1}{36}(x + y)^2 = \tfrac{1}{36}\left(x^2 + 2xy + y^2\right) \\ &\equiv \tfrac{1}{36}\left(x^2 + 2y + 5y\right) = \tfrac{1}{36}\left(x^2 + 7y\right), \end{aligned}$$

这就是所求的结果。（作为合理性检查，代入 $x = 1$ 得 $\frac{1}{36}(1 + 35) = 1$。）

一般的 pgf。 由约简规则，$\mathrm{G}_n(x) \equiv \mathrm{G}_1(x)^n = \frac{1}{6^n}(x + y)^n$。用二项式定理（binomial theorem）展开：

$$(x + y)^n = x^n + \sum_{k=1}^{n} \binom{n}{k} x^{n-k} y^k.$$

对 $k \geq 1$，利用 $y^k \equiv 5^{k-1} y$（由 $y^2 \equiv 5y$ 用归纳法得出），再利用 $x^{n-k} y \equiv y$，得

$$x^{n-k} y^k \equiv 5^{k-1} y.$$

所以

$$(x+y)^n \equiv x^n + y\sum_{k=1}^{n}\binom{n}{k}5^{k-1} = x^n + \frac{y}{5}\left[(1+5)^n - 1\right] = x^n + \frac{6^n-1}{5}y.$$

想一想 $(6^n-1)/5$ 是从哪里来的?

由二项式定理，和 $\sum_{k=0}^{n}\binom{n}{k}5^k$ 等于 $(1+5)^n = 6^n$；我们少了 $k=0$ 那一项（其值为 1），又多了一个因子 $\frac{1}{5}$。因此

$$G_n(x) = \frac{1}{6^n}\left(x^r + \frac{6^n-1}{5}y\right),$$

其中 r 是 n 除以 5 的余数（因为 $x^n \equiv x^r$）。所以每个余数的概率都是 $\dfrac{6^n-1}{5\cdot 6^n}$，唯独余数 r 额外多了 $\frac{1}{6^n}$。

被 5 整除。 我们需要 $P(T_n = 0)$，即 $G_n(x)$ 的常数项。y 的常数项是 1。

- 若 $5 \nmid n$，则 $r \neq 0$，x^r 对常数项没有贡献，所以

$$P(S_n \text{ 可被整除，除数为5}) = \frac{6^n-1}{5\cdot 6^n} = \frac{1}{5}\left(1 - \frac{1}{6^n}\right),$$

  正如所求。

- 若 $5 \mid n$，则 $r = 0$，$x^r = 1$ 会加上 $\frac{1}{6^n}$:

$$P(S_n \text{ 可被整除，除数为5}) = \frac{6^n-1}{5\cdot 6^n} + \frac{1}{6^n} = \frac{6^n+4}{5\cdot 6^n} = \frac{1}{5}\left(1 + \frac{4}{6^n}\right).$$

验算

- $n=1$: 公式给出 $\frac{1}{5}\cdot\frac{5}{6} = \frac{1}{6}$，而事实上只有掷出 5 才能被 5 整除。
- $n=2$: 公式给出 $\frac{1}{5}\cdot\frac{35}{36} = \frac{7}{36}$。直接计算，和为 5（四种方法）及 10（三种方法）共得 $\frac{7}{36}$，与 $\frac{1}{36}(x^2+7y)$ 的常数项吻合。
- $n=5$: 答案是 $\frac{6^5+4}{5\cdot 6^5} = \frac{389}{1944} \approx 0.2001$，略高于 $\frac{1}{5}$。用电脑对 $n \leq 8$ 穷举全部 6^n 种结果，证实了整个 pgf $G_n(x)$ 以及两种情况。
- 当 $n \to \infty$ 时，两个答案都趋向 $\frac{1}{5}$: 余数变成均匀分布。

常见错误

- 忘记掷出 6 除以 5 的余数是 1，因此 G_1 中 x 的系数是 $\frac{2}{6}$。

- 对所有 n 都把 x^n 约简为 x^0：其实 x^n 约简为 x^r，其中 r 是 n 的余数，这正是 $5 \mid n$ 的情况与别不同的原因。
- 用“=”写出 $G_n(x) = G_1(x)^n$，然后没有按 $x^5 = 1$ 约简就直接读取系数。

STEP 之外

令 $x^m = 1$ 相当于在模 $x^m - 1$ 的多项式环（ring）中运算，即 m 阶循环群（cyclic group）的群代数（group algebra）。一个等价的工具是单位根滤波（roots-of-unity filter）：$P(m \mid S_n) = \frac{1}{m} \sum_{\omega^m=1} G_S(\omega)$，这正是离散傅里叶变换（discrete Fourier transform）在发挥作用。这个观点亦解释了为何循环上随机步的总和会趋于均匀分布，而这正是马尔可夫链（Markov chains）与群上随机漫步（random walks）理论的起点。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）以合适的程度介绍了生成函数与随机漫步。

详细解答

STEP III 2015，第 13 题

概率与统计 • 均匀分布变量的函数

STEP III 2015，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/15-s3-q13

关键思路

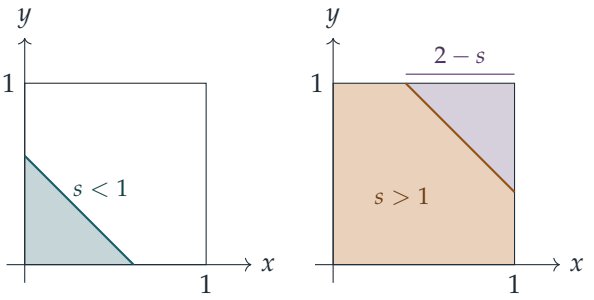
由于 X 与 Y 互相独立（independent），且都在 $[0,1]$ 上均匀分布（uniform distribution），数对 (X,Y) 便在单位正方形上均匀分布，因此关于 X 和 Y 的每一个概率，其实就是一个面积。像 $X+Y \le s$ 或 $Y \le rX$ 这类事件，都是用一条直线把正方形切开，而直线一侧的面积不是一个三角形，就是正方形减去一个三角形。之后，只要把 $1/(X+Y)$ 和 $X/(X+Y)$ 的事件改写成我们已经量度过的事件，就能得出它们的分布。

解答

全文中，(X,Y) 在面积为 1 的单位正方形 $[0,1]^2$ 上均匀分布，所以对正方形内任何区域 A，$\mathrm{P}\big((X,Y)\in A\big)$ 都等于 A 的面积。

第 (i) 部分

设 $S=X+Y$，其取值范围为 $[0,2]$，并设 $F(s)=\mathrm{P}(S\le s)$。事件 $S\le s$ 就是正方形中位于直线 $x+y=s$ 上或其下方的部分。



若 $0 \le s \le 1$，该直线与两轴相交于 $(s,0)$ 和 $(0,s)$，其下方的区域是一个两直角边长均为 s 的等腰直角三角形。若 $1 \le s \le 2$，则直线上方的区域是位于角落 $(1,1)$ 的三角形；它的两条直角边由直线延伸至该角，长度为 $1-(s-1)=2-s$。因此

$$F(s)=\mathrm{P}(X+Y\le s)=\begin{cases}0 & s<0,\\ \frac{1}{2}s^2 & 0\le s\le 1,\\ 1-\frac{1}{2}(2-s)^2 & 1\le s\le 2,\\ 1 & s>2.\end{cases}$$

中间两条公式在 $s=1$ 时一致（都等于 $\frac{1}{2}$），这是理所当然的。

第 (ii) 部分

设 $T=1/S$。由于 $0<S\le 2$（$S=0$ 的概率为零），T 的取值范围为 $[\frac{1}{2},\infty)$。对 $t>0$，因为 $s\mapsto 1/s$ 在 $s>0$ 上是递减的，

$$\mathrm{P}(T\le t)=\mathrm{P}\big(S\ge \tfrac{1}{t}\big)=1-F\big(\tfrac{1}{t}\big),$$

这里用到 S 是连续随机变量（continuous random variable），所以 $\mathrm{P}(S=1/t)=0$。

? **想一想** 应该用 F 的哪一条公式？这取决于 $1/t$ 落在哪里，而不是 t 落在哪里。

若 $t\ge 1$，则 $0<1/t\le 1$，所以用第一个分支：$\mathrm{P}(T\le t)=1-\frac{1}{2}t^{-2}$。若 $\frac{1}{2}\le t\le 1$，则 $1\le 1/t\le 2$，所以用第二个分支：$\mathrm{P}(T\le t)=\frac{1}{2}\big(2-t^{-1}\big)^2$。而当 $t<\frac{1}{2}$ 时，$\mathrm{P}(T\le t)=0$。

求导便得到概率密度函数（probability density function）。当 $t\ge 1$，

$$f(t)=\frac{\mathrm{d}}{\mathrm{d}t}\Big(1-\tfrac{1}{2}t^{-2}\Big)=t^{-3},$$

而当 $\frac{1}{2}\le t\le 1$，由链式法则（chain rule），

$$f(t)=\big(2-t^{-1}\big)\cdot t^{-2}=2t^{-2}-t^{-3},$$

其他情况下 $f(t)=0$。这就是所求的概率密度函数。

至于期望值（expectation），

$$\begin{aligned}\mathrm{E}\Big(\frac{1}{X+Y}\Big)=\int_{1/2}^{\infty} t f(t)\,\mathrm{d}t &= \int_{1/2}^{1}\big(2t^{-1}-t^{-2}\big)\,\mathrm{d}t+\int_{1}^{\infty} t^{-2}\,\mathrm{d}t\\ &=\Big[2\ln t+t^{-1}\Big]_{1/2}^{1}+\Big[-t^{-1}\Big]_{1}^{\infty}\\ &=(0+1)-(-2\ln 2+2)+1\\ &=2\ln 2.\end{aligned}$$

这个反常积分（improper integral）收敛，因为 t^{-2} 在 $[1,\infty)$ 上可积。

第 (iii) 部分

设 $R = Y/X$，其取值范围为 $[0,\infty)$（事件 $X = 0$ 的概率为零）。对 $r \geq 0$，事件 $R \leq r$ 即 $Y \leq rX$：正方形中位于通过原点的直线 $y = rx$ 上或其下方的部分。

若 $0 \leq r \leq 1$，这条直线从正方形右边的 $(1, r)$ 离开，其下方的区域是一个底为 1、高为 r 的三角形，面积为 $\frac{1}{2}r$。若 $r \geq 1$，直线从顶边的 $(1/r, 1)$ 离开，其上方的区域是一个沿顶边的底为 $1/r$、高为 1 的三角形，面积为 $\frac{1}{2r}$。所以

$$H(r) = \mathrm{P}\Big(\frac{Y}{X} \leq r\Big) = \begin{cases} 0 & r < 0, \\ \frac{1}{2}r & 0 \leq r \leq 1, \\ 1 - \dfrac{1}{2r} & r \geq 1. \end{cases}$$

现设 $Z = \dfrac{X}{X+Y} = \dfrac{1}{1+R}$，其取值范围为 $[0,1]$。对 $0 < z \leq 1$，由于 $r \mapsto 1/(1+r)$ 在 $r \geq 0$ 上是递减的，

$$\mathrm{P}(Z \leq z) = \mathrm{P}\Big(1 + R \geq \frac{1}{z}\Big) = \mathrm{P}\Big(R \geq \frac{1-z}{z}\Big) = 1 - H\Big(\frac{1-z}{z}\Big).$$

同样，用哪个分支取决于 $(1-z)/z$ 落在哪里：它不大于 1 当且仅当 $z \geq \frac{1}{2}$。

- 当 $\frac{1}{2} \leq z \leq 1$：$\mathrm{P}(Z \leq z) = 1 - \dfrac{1-z}{2z} = \dfrac{3}{2} - \dfrac{1}{2z}$。
- 当 $0 < z \leq \frac{1}{2}$：$\mathrm{P}(Z \leq z) = 1 - \Big(1 - \dfrac{z}{2(1-z)}\Big) = \dfrac{z}{2(1-z)}$。

两者在 $z = \frac{1}{2}$ 时都等于 $\frac{1}{2}$。求导（第二条可先写成 $\frac{z}{2(1-z)} = \frac{1}{2}\big(\frac{1}{1-z} - 1\big)$），得 Z 的概率密度函数为

$$g(z) = \begin{cases} \dfrac{1}{2(1-z)^2} & 0 \leq z \leq \frac{1}{2}, \\ \dfrac{1}{2z^2} & \frac{1}{2} \leq z \leq 1, \\ 0 & \text{其他情况.} \end{cases}$$

留意 $g(1-z) = g(z)$：这个概率密度函数关于 $\frac{1}{2}$ 对称。

期望值。 我们无须积分便可直接写出答案。由于 X 与 Y 互相独立且分布相同，数对 (X, Y) 与 (Y, X) 有相同的分布。所以 $\dfrac{X}{X+Y}$ 与 $\dfrac{Y}{X+Y}$ 的期望值相同。两者之和恰好是 1，所以

$$\mathrm{E}\Big(\frac{X}{X+Y}\Big) = \frac{1}{2}.$$

以积分验证。

$$\mathrm{E}(Z) = \int_0^{1/2} \frac{z}{2(1-z)^2}\,\mathrm{d}z + \int_{1/2}^{1} \frac{1}{2z}\,\mathrm{d}z.$$

第二个积分等于 $\frac{1}{2}\big[\ln z\big]_{1/2}^{1} = \frac{1}{2}\ln 2$。第一个积分中，代入 $u = 1 - z$，于是 $\mathrm{d}z = -\mathrm{d}u$，而 $z = 0, \frac{1}{2}$ 分别对应 $u = 1, \frac{1}{2}$：

$$\begin{aligned}\int_0^{1/2} \frac{z}{2(1-z)^2}\,\mathrm{d}z &= \frac{1}{2}\int_{1/2}^{1} \frac{1-u}{u^2}\,\mathrm{d}u = \frac{1}{2}\Big[-\frac{1}{u} - \ln u\Big]_{1/2}^{1} \\ &= \frac{1}{2}\big(-1 - (-2 + \ln 2)\big) = \frac{1}{2}(1 - \ln 2).\end{aligned}$$

相加得 $\mathrm{E}(Z) = \frac{1}{2}(1 - \ln 2) + \frac{1}{2}\ln 2 = \frac{1}{2}$，证实了上面的值。

验算

- 每个分布函数（distribution function）都由 0 上升至 1，并在各分支相接处连续：$F(1) = \frac{1}{2}$、$\mathrm{P}(T \le 1) = \frac{1}{2}$、$H(1) = \frac{1}{2}$、$\mathrm{P}(Z \le \frac{1}{2}) = \frac{1}{2}$。由于正方形关于对角线对称，这些值都是 $\frac{1}{2}$。
- 概率密度函数 f 的积分等于 1：$\int_{1/2}^{1}(2t^{-2} - t^{-3})\,\mathrm{d}t = 2 - \frac{3}{2} = \frac{1}{2}$，而 $\int_1^{\infty} t^{-3}\,\mathrm{d}t = \frac{1}{2}$。
- $2\ln 2 \approx 1.386$ 介乎 1 与 2 之间，合乎预期；而且它大于 $1/\mathrm{E}(X+Y) = 1$，正如对凸函数（convex function）$1/s$ 应用琴生不等式（Jensen's inequality）所预测的。直接以数值计算 $\iint (x+y)^{-1}\,\mathrm{d}x\,\mathrm{d}y$ 在正方形上的值，得 1.3863。各分布函数以及 $\mathrm{E}(Z) = \frac{1}{2}$ 亦已用电脑模拟证实。

常见错误

- 按 t 或 z 而不是按 $1/t$ 或 $(1-z)/z$ 来选择 F 或 H 的分支，结果把两条公式对调了。
- 忘记 $1/s$ 和 $1/(1+r)$ 是递减的，因而没有把不等号反转：$T \le t$ 的意思是 $S \ge 1/t$。
- 把第 (i) 部分角落三角形的直角边弄错：它是 $2 - s$，不是 $s - 1$。

STEP 之外

把关于 (X, Y) 的陈述化为面积，再对分布函数求导，这是求随机变量函数之分布的一般方法，称为*分布函数法*（distribution function method）。到了大学，还会学到联合概率密度函数的变量代换公式（雅可比行列式，Jacobian），它可以同时处理 $X+Y$ 与 $X/(X+Y)$；特别地，这里 $X+Y$ 的概率密度函数正是两个均匀分布概率密度函数的*卷积*（convolution）。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）涵盖了这两种方法，而 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）则有许多同类型的例子。

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

2015 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。

