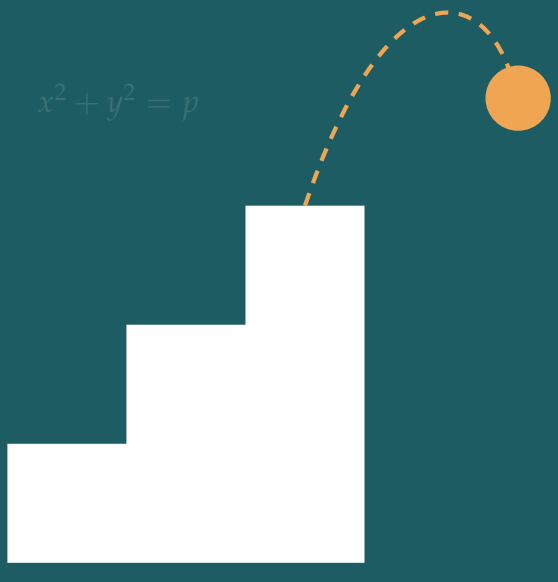
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2017 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

2017 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2017 年详细解答

2017 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 代换积分法 纯数学 . . . 3

第 2 题 对不等式积分 纯数学 . . . 6

第 3 题 抛物线的切线 纯数学 . . . 10

第 4 题 二次方程与等比级数 纯数学 . . . 14

第 5 题 以微分法求最优值 纯数学 . . . 18

第 6 题 积分与符号变化 纯数学 . . . 22

第 7 题 拿破仑定理 纯数学 . . . 26

第 8 题 递推关系与数学归纳法 纯数学 . . . 30

第 9 题 抛体运动 力学 . . . 33

第 10 题 连锁碰撞 力学 . . . 37

第 11 题 涉及摩擦力的刚体静力学 力学 . . . 41

第 12 题 期望值与条件化 概率与统计 . . . 45

第 13 题 概率的递推关系 概率与统计 . . . 48

STEP II

第 1 题 归约公式 纯数学 . . . 52

第 2 题 周期数列 纯数学 . . . 55

第 3 题 曲线描绘 纯数学 . . . 59

第 4 题 柯西–施瓦兹积分不等式 纯数学 . . . 64

第 5 题 抛物线的法线与圆 纯数学 . . . 67

第 6 题 不等式与数学归纳法 纯数学 . . . 71

第 7 题 对数微分法 纯数学 . . . 75

第 8 题 向量与垂心 纯数学 . . . 79

第 9 题 摩擦力与刚体 力学 . . . 82

第 10 题 克服阻力所作的功 力学 86
第 11 题 抛体与能量 力学 90
第 12 题 放松过程 概率与统计 94
第 13 题 期望值与无穷级数 概率与统计 98

STEP III

第 1 题 裂项相消级数的界限 纯数学 102
第 2 题 复平面上的旋转 纯数学 105
第 3 题 多项式的根 纯数学 109
第 4 题 函数的几何平均 纯数学 112
第 5 题 极坐标曲线 纯数学 116
第 6 题 代换积分法 纯数学 120
第 7 题 椭圆的切线 纯数学 124
第 8 题 分部求和法 纯数学 128
第 9 题 连接质点与能量 力学 131
第 10 题 转动的刚体 力学 135
第 11 题 动量守恒 力学 139
第 12 题 离散联合分布 概率与统计 143
第 13 题 期望值与方差 概率与统计 146

详细解答

STEP I 2017，第 1 题

纯数学 · 代换积分法

> **STEP I 2017，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/17-s1-q1

关键思路

只要把三角函数的分式化简，这几个被积函数（integrand）其实都暗藏 $f'(x)/f(x)$ 的形式（最后一题则是 $f'(x)/f(x)^2$）。$x\sin x+\cos x$ 的导数（derivative）特别简洁，因为两个 $\sin x$ 项互相抵消，而其余的代换（substitution）都建基于同样的抵消。所以策略是：把每个被积函数改写成以 sin 和 cos（或 sec 和 tan）表示，找出导数与分子（numerator）吻合的分母（denominator），然后代换。

解答

第 (i) 部分

设 $u=x\sin x+\cos x$，由乘积法则（product rule）得

$$\frac{\mathrm{d}u}{\mathrm{d}x}=\sin x+x\cos x-\sin x=x\cos x.$$

接着把被积函数的分子和分母同乘以 $\cos x$：

$$\frac{x}{x\tan x+1}=\frac{x\cos x}{x\sin x+\cos x}.$$

分子恰好是 $\mathrm{d}u/\mathrm{d}x$，所以 $x\cos x\,\mathrm{d}x=\mathrm{d}u$，从而

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

第二个积分的做法相同，只是以 $\sin x$ 代替 $\cos x$。把分子和分母同乘以 $\sin x$，

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

设 $v = x\cos x - \sin x$，则

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \cos x - x\sin x - \cos x = -x\sin x,$$

所以分子是 $-\mathrm{d}v/\mathrm{d}x$，从而

$$\int \frac{x}{x\cot x - 1}\,\mathrm{d}x = -\int \frac{1}{v}\,\mathrm{d}v = -\ln|x\cos x - \sin x| + C.$$

不要漏掉负号：它来自对 $\cos x$ 求导。

第 (ii) 部分

第一个被积函数的分母已经值得一试。设 $w = x\sec^2 x - \tan x$。利用 $\frac{\mathrm{d}}{\mathrm{d}x}\sec^2 x = 2\sec^2 x\tan x$，

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \sec^2 x + 2x\sec^2 x\tan x - \sec^2 x = 2x\sec^2 x\tan x.$$

多余的项再一次互相抵消，而分子是 $\frac{1}{2}\,\mathrm{d}w/\mathrm{d}x$。因此

$$\int \frac{x\sec^2 x\,\tan x}{x\sec^2 x - \tan x}\,\mathrm{d}x = \frac{1}{2}\int \frac{1}{w}\,\mathrm{d}w = \tfrac{1}{2}\ln|x\sec^2 x - \tan x| + C.$$

? **想一想** 最后一个积分里没有 sec 或 tan。它怎样才能与 $w = x\sec^2 x - \tan x$ 拉上关系?

从分母的括号中提取因子 $\cos^2 x$：

$$x - \sin x\cos x = \cos^2 x\left(x\sec^2 x - \frac{\sin x}{\cos x}\right) = \cos^2 x\,(x\sec^2 x - \tan x) = w\cos^2 x.$$

所以 $(x - \sin x\cos x)^2 = w^2\cos^4 x$，而

$$\frac{x\sin x\cos x}{(x - \sin x\cos x)^2} = \frac{x\sin x\cos x}{w^2\cos^4 x} = \frac{x\sec^2 x\tan x}{w^2},$$

因为 $\dfrac{\sin x\cos x}{\cos^4 x} = \dfrac{\sin x}{\cos x}\cdot\dfrac{1}{\cos^2 x}$。分子同样是 $\frac{1}{2}\,\mathrm{d}w/\mathrm{d}x$，所以用同一个代换，

$$\begin{aligned}\int \frac{x\sin x\cos x}{(x - \sin x\cos x)^2}\,\mathrm{d}x &= \frac{1}{2}\int \frac{1}{w^2}\,\mathrm{d}w = -\frac{1}{2w} + C\\ &= -\frac{1}{2(x\sec^2 x - \tan x)} + C = -\frac{\cos^2 x}{2(x - \sin x\cos x)} + C,\end{aligned}$$

其中最后一步是把分子和分母同乘以 $\cos^2 x$。如果你喜欢用倍角（double angles），答案可写成 $-\dfrac{\cos^2 x}{2x - \sin 2x} + C$。

验算

- 把每个答案求导，确认能得回被积函数。对最后一题，用商法则（quotient rule）处理 $-\cos^2 x/\big(2(x-\sin x\cos x)\big)$ 时要用到 $\frac{\mathrm{d}}{\mathrm{d}x}(x-\sin x\cos x)=1-\cos 2x=2\sin^2 x$，而各个 $\sin^2 x\cos^2 x$ 项会整齐地合并。
- 数值计算：在 $[0.3, 1.2]$ 上，第一个积分的值是 $0.34953\ldots$，而 $\ln|x\sin x+\cos x|$ 的改变量相同；最后一个积分在同一区间上的值是 $25.7364\ldots$，与 $-\cos^2 x/\big(2(x-\sin x\cos x)\big)$ 吻合。四个答案都已用电脑这样核实过。

常见错误

- 未先消去 $\tan x$ 或 $\cot x$ 就代换：只有在整式乘以 $\cos x$ 或 $\sin x$ 之后，分子才会与 $\mathrm{d}u/\mathrm{d}x$ 吻合。
- 漏掉第 (i) 部分第二个积分中的负号，或第 (ii) 部分中的因子 $\frac{1}{2}$。
- 在最后一个积分中出于习惯就用对数：$1/w^2$ 的积分是 $-1/w$。

STEP 之外

这里的技巧是逆向设计：选取代换，使分母的导数在抵消之后恰好等于分子。背后的道理是 $\int f'/f=\ln|f|$，即对数导数（logarithmic derivative）；它在复分析（complex analysis）中再次出现（用来数函数的零点），在微分方程（differential equations）中则以积分因子（integrating factor）的形式出现。不少 STEP 积分题都是先对一个精心挑选的函数求导，再把答案藏起来而设计出来的；Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）有大量练习，助你学会识破它们。

详细解答

STEP I 2017，第 2 题

纯数学 · 对不等式积分

STEP I 2017，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/17-s1-q2

关键思路

积分（integration）保持不等式（inequality）：若在区间 $[a,b]$（其中 $a<b$）上 $f(t)\le g(t)$，则 $\int_a^b f \le \int_a^b g$。唯一要小心的是积分的方向：当 $x<1$ 时，区间 $[1,x]$ 是倒过来走的，不等式便会反转。最后一部分再次用到同一个想法：把一个非负（non-negative）函数由 1 积分到 y，所得结果与 $y-1$ 同号，而这个符号正是我们安全地除去该因子所需要的。

解答

第 (i) 部分

情况 $x\ge 1$。当 $1\le t\le x$ 时，有 $\frac{1}{t}\le 1$，而且区间是正向的，所以

$$\int_1^x \frac{1}{t}\,\mathrm{d}t \le \int_1^x 1\,\mathrm{d}t, \qquad \text{即} \qquad \ln x \le x-1.$$

情况 $0<x\le 1$。现在我们在 $x\le t\le 1$ 上积分，此时 $t\le 1$，故 $\frac{1}{t}\ge 1$。于是

$$\int_x^1 \frac{1}{t}\,\mathrm{d}t \ge \int_x^1 1\,\mathrm{d}t, \qquad \text{即} \qquad -\ln x \ge 1-x.$$

乘以 -1 会令不等式反转，再次得到 $\ln x\le x-1$。所以 $(*)$ 对所有 $x>0$ 成立。

第 (ii) 部分

情况 $x\ge 1$。当 $t\ge 1$ 时，有 $\frac{1}{t^2}\le\frac{1}{t}$，所以

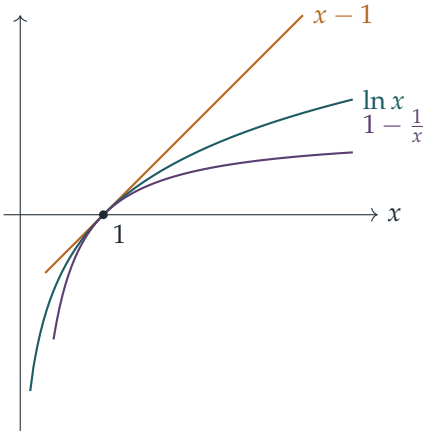
$$\int_1^x \frac{1}{t^2}\,\mathrm{d}t \le \int_1^x \frac{1}{t}\,\mathrm{d}t, \qquad \text{即} \qquad 1-\frac{1}{x}\le \ln x.$$

情况 $0 < x \leq 1$。当 $x \leq t \leq 1$ 时，不等式反转为 $\frac{1}{t^2} \geq \frac{1}{t}$（把 $t \leq 1$ 乘以正数 $\frac{1}{t^2}$）。在 $[x,1]$ 上积分，

$$\frac{1}{x} - 1 \geq -\ln x,$$

再乘以 -1 便得 $\ln x \geq 1 - \frac{1}{x}$。所以 $(**)$ 对所有 $x > 0$ 成立。

（另一方法：$(**)$ 可由 $(*)$ 直接推出：在 $(*)$ 中以 $1/x$ 代替 x，即得 $-\ln x \leq \frac{1}{x} - 1$。）



三条曲线在 $x = 1$ 处相切，每条在该点的斜率（gradient）都是 1；$\ln x$ 被夹在另外两条曲线之间。

第 (iii) 部分

$(*)$ 和 $(**)$ 都是说某个函数在 $x > 0$ 上非负：

$$f(x) = x - 1 - \ln x \geq 0, \qquad g(x) = \ln x - 1 + \frac{1}{x} \geq 0.$$

我们需要一个关于积分这类函数的简单事实。

> **想一想** 若对所有 $x > 0$ 都有 $h(x) \geq 0$，那么当 $y < 1$ 时，$\int_1^y h(x)\,\mathrm{d}x$ 的符号是甚么？

若 $y > 1$，积分是在正向区间上进行，所以结果 ≥ 0。若 $0 < y < 1$，则 $\int_1^y h = -\int_y^1 h \leq 0$。无论哪种情况，都有

$$(y-1)\int_1^y h(x)\,\mathrm{d}x \;\geq\; 0.$$

这让我们可以把 $y > 1$ 和 $0 < y < 1$ 两种情况一并处理。

我们需要 $\int \ln x\,\mathrm{d}x = x\ln x - x$（以 $u = \ln x$ 用分部积分法（integration by parts）求得），所以

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

*右边的不等式。*对 f 积分，

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

由上述事实，

$$(y-1)\left[\frac{(y-1)(y+1)}{2} - y\ln y\right] \geq 0 \quad\Longrightarrow\quad \frac{(y-1)^2(y+1)}{2} \geq y(y-1)\ln y.$$

由于 $y > 0$ 且 $y \neq 1$，$y(y-1)^2$ 是正数，除以它不会改变不等式的方向：

$$\frac{y+1}{2y} \geq \frac{\ln y}{y-1}.$$

*左边的不等式。*对 g 积分，

$$\int_1^y g(x)\,\mathrm{d}x = (y\ln y - y + 1) - (y-1) + \ln y = (y+1)\ln y - 2(y-1).$$

由上述事实，

$$(y-1)\left[(y+1)\ln y - 2(y-1)\right] \geq 0 \quad\Longrightarrow\quad (y+1)(y-1)\ln y \geq 2(y-1)^2.$$

现在除以 $(y+1)(y-1)^2 > 0$：

$$\frac{\ln y}{y-1} \geq \frac{2}{y+1}.$$

把两者合起来，

$$\frac{2}{y+1} \leq \frac{\ln y}{y-1} \leq \frac{y+1}{2y} \qquad \text{对} y > 0,\ y \neq 1,$$

得证。

想一想 为甚么要先乘以 $y-1$ 再除，而不是直接除？

把不等式除以 $y-1$，在 $y < 1$ 时会令不等式反转，迫使我们分开两种情况处理。先乘以 $y-1$，所得的式子含有 $(y-1)^2$，无论我们在 1 的哪一边它都是正数，因此一个论证便可涵盖两种情况。

验算

- 当 $y = 2$：$\frac{2}{3} \approx 0.667 \leq \ln 2 \approx 0.693 \leq \frac{3}{4} = 0.75$。
- 当 $y = \frac{1}{2}$：$\frac{4}{3} \approx 1.333 \leq 2\ln 2 \approx 1.386 \leq \frac{3}{2}$。

- 当 $y \to 1$ 时，三个式子都趋向 1（中间那个是 ln 在 1 处的斜率），与图中曲线相切的情况一致。
- 以 $1/y$ 代替 y，再全式除以 y，会得回完全相同的一对不等式，所以 $0 < y < 1$ 的结果与 $y > 1$ 的结果是等价的。

常见错误

- 当 $x < 1$ 时在 $[1, x]$ 上积分，却没有留意到不等式会反转。
- 除以 $y - 1$ 时没有考虑它的符号，结果只证明了 $y > 1$ 时的第 (iii) 部分。

STEP 之外

中间的量 $\dfrac{\ln y}{y-1}$ 是 1 与 y 的对数平均（logarithmic mean）的倒数，所以左边的不等式是说对数平均不大于算术平均（arithmetic mean）$\frac{1+y}{2}$。事实上，对数平均亦不小于几何平均（geometric mean）$\sqrt{y}$，由此得 $\frac{\ln y}{y-1} \leq \frac{1}{\sqrt{y}}$，这是比 $\frac{y+1}{2y}$ 更精确的上界。对不等式积分这个技巧，以及支持它的积分单调性（monotonicity），都是大学第一门数学分析（analysis）课程的基本工具。Abbott 的 *Understanding Analysis*（《理解分析》）对积分有细致的处理。

详细解答

STEP I 2017，第 3 题

纯数学 · 抛物线的切线

> **STEP I 2017，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/17-s1-q3

关键思路

这里的一切都可以用参数（parameter）p 和 q 表示，所以计划是先求出 R、S、T 的坐标，然后直接计算两个面积。三角形 RST 特别容易处理，因为它的边 ST 落在 y 轴上：它的高就是 R 到该轴的距离。至于三角形 OPQ，由于它有一个顶点在原点，公式 $\frac{1}{2}|x_1y_2 - x_2y_1|$ 一行就能完成计算。

解答

P 处的切线

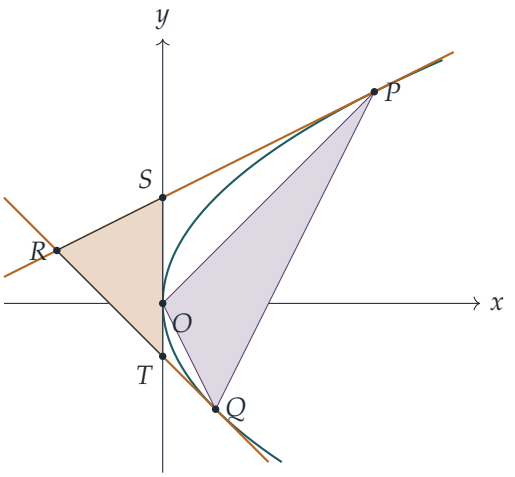
对 $y^2 = 4ax$ 作隐函数微分（implicit differentiation），得 $2y\dfrac{\mathrm{d}y}{\mathrm{d}x} = 4a$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{2a}{y}$。在 P 点，$y = 2ap \neq 0$（因为 $a, p > 0$），斜率（gradient）为

$$\frac{2a}{2ap} = \frac{1}{p}.$$

因此通过 $P(ap^2, 2ap)$ 的切线（tangent）是

$$y - 2ap = \frac{1}{p}\left(x - ap^2\right) \quad \Longrightarrow \quad y = \frac{1}{p}x - ap + 2ap = \frac{1}{p}x + ap,$$

正如所求。同理，由于 Q 的形式相同，只是以 q 代替 p（且 $q \neq 0$），Q 处的切线是 $y = \dfrac{1}{q}x + aq$。



草图展示的是 $a=1$、$p=2$、$q=-1$ 的情况：R 位于 y 轴左方，高度介乎 S 与 T 之间。

点 R、S 和 T

在每条切线中代入 $x=0$，得

$$S=(0,\,ap),\qquad T=(0,\,aq).$$

两条切线在 R 相交，所以

$$\frac{x}{p}+ap=\frac{x}{q}+aq\quad\Longrightarrow\quad x\left(\frac{1}{p}-\frac{1}{q}\right)=a(q-p)\quad\Longrightarrow\quad x\cdot\frac{q-p}{pq}=a(q-p).$$

由于 $p>0>q$，有 $q-p\neq 0$，因此可以除以它（这也说明了两条切线并不平行，R 确实存在）。于是 $x=apq$，再得 $y=\dfrac{apq}{p}+ap=a(p+q)$：

$$R=\big(apq,\ a(p+q)\big).$$

想一想 R 在 y 轴的哪一边？为甚么这对计算面积很重要？

因为 $p>0$ 而 $q<0$，乘积 pq 是负数，所以 R 的 x 坐标 apq 是负的：R 在轴的左方，与轴的距离为 $a|pq|=-apq$。面积必须是正数，所以我们要留意正负号，而不是把 apq 写成“高”。

三角形 RST 的面积

以 ST 为底。它落在 y 轴上，长度为 $ap-aq=a(p-q)$，由于 $p>q$，这是正数。由 R 到 y 轴的垂直高度是 $-apq$。所以

$$[RST]=\tfrac{1}{2}\cdot a(p-q)\cdot(-apq)=-\tfrac{1}{2}a^2pq(p-q).$$

三角形 OPQ 的面积

对于顶点为 O、(x_1, y_1) 和 (x_2, y_2) 的三角形，面积是 $\frac{1}{2}|x_1y_2 - x_2y_1|$。取 $P(ap^2, 2ap)$ 和 $Q(aq^2, 2aq)$，

$$x_1y_2 - x_2y_1 = ap^2 \cdot 2aq - aq^2 \cdot 2ap = 2a^2pq(p-q).$$

这里 $a^2 > 0$、$pq < 0$ 且 $p - q > 0$，所以这个值是负的，因此

$$[OPQ] = \tfrac{1}{2} \cdot \left(-2a^2pq(p-q)\right) = -a^2pq(p-q).$$

结论

比较两个结果，

$$[OPQ] = -a^2pq(p-q) = 2\left(-\tfrac{1}{2}a^2pq(p-q)\right) = 2\,[RST],$$

所以三角形 OPQ 的面积是三角形 RST 面积的两倍，正如所求。

验算

在草图的情况 $a = 1$、$p = 2$、$q = -1$ 中，得 $P(4,4)$、$Q(1,-2)$、$R(-2,1)$、$S(0,2)$ 和 $T(0,-1)$。于是 $[RST] = \frac{1}{2} \cdot 3 \cdot 2 = 3$，而 $[OPQ] = \frac{1}{2}|4 \cdot (-2) - 1 \cdot 4| = 6$，正好是两倍。你也可以验证 R 同时满足两条切线方程：$1 = \frac{-2}{2} + 2$ 及 $1 = \frac{-2}{-1} - 1$。这个恒等式亦已用电脑以精确算术，对 2000 组随机选取的 a、p、q 验证过。

常见错误

- 把三角形 RST 的高写成 apq：这是负数，因为 R 位于 y 轴上与曲线相反的一侧。
- 在 $\frac{1}{2}|x_1y_2 - x_2y_1|$ 中漏掉绝对值（modulus），结果得出 OPQ 的面积为负数。
- 除以 $q - p$ 时没有说明它为何不等于零。

STEP 之外

参数化（parametrisation）$(at^2, 2at)$ 使抛物线（parabola）上每一点都对应一个数 t，而 $t = p$ 和 $t = q$ 处的切线总是相交于 $(apq, a(p+q))$：即两个参数的“积”与“和”。这类结果，连同本题的 $2:1$ 面积比，都属于圆锥曲线（conics）的古典射影几何（projective geometry），在这门学问中，切线和弦都是借助参数和行列式（determinant）来研究的。三角形面积的行列式公式，是把面积看成一种有向（signed）、线性的量的起点，这方面的发展可参阅 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）；Siklos 的

Advanced Problems in Mathematics（《数学进阶问题》）则有更多关于参数曲线的 STEP 式练习。

详细解答

STEP I 2017，第 4 题

纯数学 · 二次方程与等比级数

> **STEP I 2017，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/17-s1-q4

关键思路

一切取决于开区间（open interval）$(-1,1)$ 内有多少个 r 值会给出某个特定的 p（或 q）值。这其实是关于一条限制在 $-1<r<1$ 上的抛物线（parabola）的问题：一条水平直线与它相交一次、两次或完全不相交，而草图会告诉我们是哪一种情况。一旦知道 r，和便随之确定；又因为和是 r 的一一对应（one-to-one）函数，两个不同的 r 值必定给出两个不同的和。要找出这些和所满足的方程，我们把 r 用和来表示，然后代入。

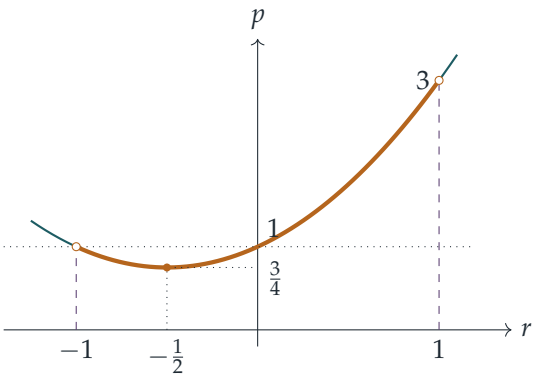
解答

第 (i) 部分

配方（completing the square），

$$1+r+r^2=\left(r+\tfrac{1}{2}\right)^2+\tfrac{3}{4},$$

因此图像是一条抛物线，其最小点（minimum point）为 $\left(-\tfrac{1}{2},\tfrac{3}{4}\right)$。它经过 $(-1,1)$、$(0,1)$ 和 $(1,3)$。



在 $-1 < r < 1$ 上（粗线那段弧），函数由 1（取不到，因为 $r = -1$ 被排除）递减至 $r = -\frac{1}{2}$ 处的最小值 $\frac{3}{4}$，然后递增趋向 3（同样取不到，因为 $r = 1$ 被排除）。因此

$$\frac{3}{4} \leq p < 3.$$

当 $1 < p < 3$ 时。给出这个 p 的 r 值是以下方程的根（roots）：

$$r^2 + r + (1 - p) = 0.$$

两根之积为 $1 - p < 0$，所以有一个正根和一个负根。两根之和为 -1，所以负根等于 -1 减去正根，即小于 -1，因此不容许。正根为

$$r = \frac{-1 + \sqrt{4p - 3}}{2},$$

又因为 $1 < 4p - 3 < 9$，我们有 $0 < r < 1$，所以这个根是容许的。（从图像看：高度为 $p \in (1, 3)$ 的直线只在上升部分 $0 < r < 1$ 与粗线弧相交。）所以 p 唯一确定 r，从而唯一确定 $S = 1/(1 - r)$。

当 $\frac{3}{4} < p < 1$ 时。此时判别式（discriminant）$1 - 4(1 - p) = 4p - 3$ 为正，所以这个二次方程（quadratic）有两个相异实根。两根之积 $1 - p$ 为正，两根之和为 -1，所以两根皆为负。每个根都等于 -1 减去另一个根，而另一个根是负数，所以每个根都大于 -1。因此两根都位于 $(-1, 0)$ 内，两者皆容许。

想一想 有两个 r 值，没问题；但为甚么它们一定给出两个不同的 S 值?

因为 $S = \dfrac{1}{1 - r}$ 可以反过来写成 $r = 1 - \dfrac{1}{S}$。因此不同的 r 值给出不同的 S 值，所以 S 恰好有两个可能值。

要找出它们满足的方程，把 $r = 1 - u$（其中 $u = 1/S$）代入 $r^2 + r + 1 - p = 0$：

$$(1 - u)^2 + (1 - u) + 1 - p = 0$$
$$u^2 - 3u + (3 - p) = 0.$$

代入 $u = 1/S$ 并乘以 S^2（它不为零），得

$$(3-p)S^2 - 3S + 1 = 0,$$

正如所求。由于两个 S 值都来自容许的根 r，它们都满足这个方程。

第 (ii) 部分

再次配方，

$$q = 1 + 2r + 3r^2 = 3\left(r + \tfrac{1}{3}\right)^2 + \tfrac{2}{3}.$$

最小值为 $r = -\frac{1}{3}$ 处的 $\frac{2}{3}$；在 $r = -1$ 处的值为 2，在 $r = 1$ 处的值为 6。所以在 $-1 < r < 1$ 上：

- 在下降部分 $-1 < r \le -\frac{1}{3}$ 上，q 取 $\left[\frac{2}{3}, 2\right)$ 中每个值恰好一次；
- 在上升部分 $-\frac{1}{3} \le r < 1$ 上，q 取 $\left[\frac{2}{3}, 6\right)$ 中每个值恰好一次。

两部分只共有 $r = -\frac{1}{3}$ 这一点。因此一个 q 值来自

- 恰好一个 r：若 $q = \frac{2}{3}$（即 $r = -\frac{1}{3}$）或 $2 \le q < 6$（只在上升部分）；
- 恰好两个 r 值：若 $\frac{2}{3} < q < 2$（每部分各一个）；
- 其他情况下没有任何 r。

例如在 $q = 2$ 时，方程 $3r^2 + 2r - 1 = 0$ 可因式分解为 $(3r-1)(r+1) = 0$，而根 $r = -1$ 被排除，只剩下 $r = \frac{1}{3}$。

和之前一样，T 确定 r：由于 $1 - r > 0$，我们有 $1 - r = 1/\sqrt{T}$。所以不同的 r 给出不同的 T。因此

$$T \text{ 唯一确定，当} q = \tfrac{2}{3} \text{ 或} 2 \le q < 6,$$

以及

$$T \text{ 有两个可能值，当} \tfrac{2}{3} < q < 2.$$

二次方程。设 $r = 1 - u$，其中 $u = 1 - r = 1/\sqrt{T} > 0$。则

$$q = 1 + 2(1-u) + 3(1-u)^2 = 3u^2 - 8u + 6.$$

由于 $u^2 = 1/T$，上式即 $\dfrac{3}{T} + 6 - q = \dfrac{8}{\sqrt{T}}$。乘以 T，

$$(6-q)T + 3 = 8\sqrt{T}.$$

这已经是一个关于 $\sqrt{T}$ 的二次方程。要得到关于 T 本身的方程，把两边平方：

$$(6-q)^2T^2 + 6(6-q)T + 9 = 64T,$$

即

$$(6-q)^2\,T^2-2(3q+14)\,T+9=0.$$

T 的两个可能值都满足这个方程。这里平方不可能引入多余的根。由于 $q<6$ 且 $T>0$，左边 $(6-q)T+3$ 为正，所以另一种可能 $(6-q)T+3=-8\sqrt{T}$ 没有 $T>0$ 的解。此外，关于 T 的二次方程两根之积 $9/(6-q)^2$ 为正，两根之和亦为正，所以两根皆为正。因此它的两个根正好就是 T 的两个值。

验算

- 取 $p=\frac{7}{8}$。则 $r^2+r+\frac{1}{8}=0$ 给出 $r=\frac{-2\pm\sqrt{2}}{4}$，两者皆在 $(-1,0)$ 内。相应的和 $S\approx 0.873$ 及 $S\approx 0.540$ 满足 $\frac{17}{8}S^2-3S+1=0$，其两根之和为 $\frac{24}{17}\approx 1.412$，积为 $\frac{8}{17}$。
- 取 $q=1$。则 $3r^2+2r=0$ 给出 $r=0$ 或 $r=-\frac{2}{3}$，所以 $T=1$ 或 $T=\frac{9}{25}$。二次方程变成 $25T^2-34T+9=(T-1)(25T-9)$，与预期相符。
- 在边界 $q=\frac{2}{3}$ 处，二次方程为 $\frac{256}{9}T^2-32T+9=0$，判别式为 $1024-1024=0$。它有重根 $T=\frac{9}{16}=1/(1+\frac{1}{3})^2$，与唯一的值 $r=-\frac{1}{3}$ 吻合。
- 我们用电脑检验了各范围内数以千计的随机 p 和 q 值：根的个数及两条二次方程都一致。

常见错误

- 忘记 $r=\pm 1$ 被排除，以致错误地把 $p=1$ 归入“两个值”的情况，并把 $q=2$ 排除在“唯一”的集合之外。
- 遗漏孤立的值 $q=\frac{2}{3}$（顶点（vertex）），在该处 T 同样唯一。
- 平方后没有检查是否引入了多余的根。

STEP 之外

这道题其实是在问函数何时可逆（invertible）：映射 $r\mapsto p$ 在定义域（domain）的一部分上是一一对应的，在其余部分上则是二对一的，而答案取决于定义域，不只取决于公式。级数 $\sum r^n$ 和 $\sum nr^{n-1}$ 分别是一个幂级数（power series）及其逐项求导所得的导数（term-by-term derivative）。为甚么在收敛半径（radius of convergence）之内可以逐项求导，是大学一年级分析（analysis）课程的标准结果。Abbott 的 *Understanding Analysis*（《理解分析》）对幂级数和一致收敛（uniform convergence）的处理既严谨又易读。

详细解答

STEP I 2017，第 5 题

纯数学 • 以微分法求最优值

STEP I 2017，第 5 题

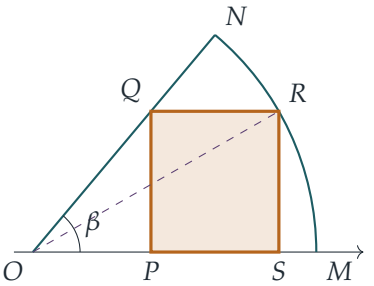
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/17-s1-q5

关键思路

顶点的位置一旦定下来，一切就只由单一变量 x 决定：Q 位于 P 的正上方，在直线 $y = x\tan\beta$ 上；R 与 Q 同高，位于圆上。利用毕氏定理（Pythagoras′ theorem）可把 s 以 x 表示，于是面积变成一个普通的单变量函数，可以直接求导。最后的角度条件则可由 tan 的半角公式（half-angle formula）直接得出。

解答



面积

由于 $PQRS$ 是长方形，且 P 和 S 都在 x 轴上，所以边 PQ 和 SR 都是铅垂的。因此 Q 的 x 坐标是 x。它位于 ON 上，即直线 $y = x\tan\beta$ 上，所以 $Q = (x,\, x\tan\beta)$，而 $R = (s,\, x\tan\beta)$。长方形的阔为 $PS = s - x$（这是正数，因为 S 比 P 离 O 更远），高为 $PQ = x\tan\beta$，所以

$$A = x(s - x)\tan\beta,$$

正如所求。

求 s 及 $\frac{\mathrm{d}A}{\mathrm{d}x}$

由于 R 在圆上，$OR = a$，所以 $s^2 + x^2\tan^2\beta = a^2$。因为 $s > 0$，

$$s = \sqrt{a^2 - x^2\tan^2\beta}.$$

对 $s^2 = a^2 - x^2\tan^2\beta$ 作隐函数求导（implicit differentiation），得 $2s\dfrac{\mathrm{d}s}{\mathrm{d}x} = -2x\tan^2\beta$，所以 $\dfrac{\mathrm{d}s}{\mathrm{d}x} = -\dfrac{x\tan^2\beta}{s}$。现在把面积写成 $A = (xs - x^2)\tan\beta$，并对 xs 使用乘积法则（product rule）：

$$\begin{aligned}\frac{\mathrm{d}A}{\mathrm{d}x} &= \left(s + x\frac{\mathrm{d}s}{\mathrm{d}x} - 2x\right)\tan\beta \\ &= (s - 2x)\tan\beta - \frac{x^2}{s}\tan^3\beta,\end{aligned}$$

这就是所求的结果。

最大面积

首先考虑 x 的范围。当 x 从 0 开始增加时，Q 沿 ON 向上移动，R 则沿圆弧向 N 移动；当 R 到达 N 时，长方形收缩成一条铅垂线段，即 $x = a\cos\beta$ 时（此时 $s = \sqrt{a^2 - a^2\sin^2\beta} = a\cos\beta = x$）。所以 $0 < x < a\cos\beta$，而 A 是 $[0, a\cos\beta]$ 上的连续函数，两端的值都是 $A = 0$，中间则 $A > 0$。因此它的最大值必定出现在区间内部的一个驻点（stationary point）。

设 $\frac{\mathrm{d}A}{\mathrm{d}x} = 0$，除以 $\tan\beta > 0$ 再乘以 $s > 0$：

$$\begin{aligned}s^2 - 2xs &= x^2\tan^2\beta \\ s^2 - 2xs + x^2 &= x^2(1 + \tan^2\beta) = x^2\sec^2\beta \\ (s - x)^2 &= x^2\sec^2\beta.\end{aligned}$$

想一想 应该取哪一个平方根？我们又怎样知道这是最大值？

$s - x$ 和 $x\sec\beta$ 都是正数（因为 $0 < \beta < \frac{\pi}{2}$），所以 $s - x = x\sec\beta$；另一个根会令 $s - x$ 为负数。因此在 $0 < x < a\cos\beta$ 内恰好只有一个驻点，由

$$s = x(1 + \sec\beta),$$

给出，而根据上一段，它必定给出最大面积。

要求出该面积，把它代入 $s^2 + x^2\tan^2\beta = a^2$。利用 $\tan^2\beta = \sec^2\beta - 1$，

$$(1 + \sec\beta)^2 + \tan^2\beta = 2\sec\beta + 2\sec^2\beta = 2\sec\beta\,(1 + \sec\beta),$$

所以

$$x^2 = \frac{a^2}{2\sec\beta\,(1+\sec\beta)} = \frac{a^2\cos^2\beta}{2(1+\cos\beta)}.$$

在这一点，$s - x = x\sec\beta$，所以

$$A = x\cdot x\sec\beta\cdot\tan\beta = x^2\,\frac{\sin\beta}{\cos^2\beta} = \frac{a^2\sin\beta}{2(1+\cos\beta)}$$
$$= \frac{a^2\cdot 2\sin\frac12\beta\cos\frac12\beta}{2\cdot 2\cos^2\frac12\beta} = \tfrac12 a^2\tan\tfrac12\beta,$$

这里用了 $\sin\beta = 2\sin\frac12\beta\cos\frac12\beta$ 及 $1+\cos\beta = 2\cos^2\frac12\beta$。

角 ROS

在直角三角形 ORS 中，$\tan\angle ROS = \dfrac{SR}{OS} = \dfrac{x\tan\beta}{s}$。在最大值处，

$$\tan\angle ROS = \frac{\tan\beta}{1+\sec\beta} = \frac{\sin\beta}{\cos\beta+1} = \tan\tfrac12\beta,$$

用的是同样的两条半角恒等式。$\angle ROS$ 和 $\frac12\beta$ 都位于 $(0, \frac{\pi}{2})$ 内，而 tan 在此区间上是一一对应（one-to-one）的，所以 $\angle ROS = \frac12\beta$：在最佳的长方形中，$OR$ 平分扇形。

验算

- 取 $a = 1$、$\beta = \frac{\pi}{3}$。则 $x^2 = \frac{1/4}{3} = \frac{1}{12}$，$s = 3x$，而 $A = x\cdot 2x\cdot\sqrt{3} = \frac{\sqrt{3}}{6} \approx 0.2887$，正好等于 $\frac12\tan\frac{\pi}{6}$。以电脑对大量随机的 a 和 β，在 x 的细密网格上搜索，得到相同的最大值和相同的最佳 x。
- 当 $\beta\to 0$ 时，扇形收缩，而 $\frac12 a^2\tan\frac12\beta\to 0$，理应如此。
- 单位：A 是 a^2 乘以一个纯数，所以它确实是面积。

常见错误

- 忘记 s 依赖于 x，把 s 当作常数去对 $x(s-x)\tan\beta$ 求导。
- 把 $s - x = -x\sec\beta$ 也一并接受，或没有说明为何该驻点是最大值。

STEP 之外

答案如此整齐——OR 正好沿扇形的角平分线——背后有一个漂亮的解释：若 $R = (a\cos\theta, a\sin\theta)$，简短计算可得 $A = \frac12 a^2\big(\cos(2\theta-\beta) - \cos\beta\big)/\sin\beta$，显然在 $\theta = \frac12\beta$ 时

取得最大值。选对变量，可以令最优化（optimisation）问题几乎变得轻而易举。在大学里，你会学到多变量的约束最优化（constrained optimisation），届时不必消去 s，而可以用拉格朗日（Lagrange）乘数法（Lagrange multipliers），在 $s^2 + x^2 \tan^2 \beta = a^2$ 的约束下求 $x(s - x)$ 的最大值。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）收录了更多同类题目。

详细解答

STEP I 2017，第 6 题

纯数学 · 积分与符号变化

STEP I 2017，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/17-s1-q6

关键思路

一个从不变号、又不恒等于零的连续函数（continuous function），其积分（integral）必定不为零：它在某个小区间上与零保持距离，单是这个区间的贡献已经不为零。第 (i) 部分把这个想法严格化，而第 (ii) 及 (iii) 部分则把它应用到一些巧妙选取的函数上。在 (ii) 中，被积函数（integrand）$(x-\alpha)^2 g(x)$ 与 g 同号；在 (iii) 中，分部积分（integration by parts）把关于 h 的资料转化为 h' 满足 (ii) 的条件。

解答

第 (i) 部分

用反证法（proof by contradiction），假设 f 在 $[0,1]$ 上并非同时取正值和负值。那么，要么对 $[0,1]$ 中所有 x 都有 $f(x) \geq 0$，要么对所有这些 x 都有 $f(x) \leq 0$。如有需要，以 $-f$ 代替 f（这既不改变假设 $\int_0^1 f(x)\,\mathrm{d}x = 0$，也不改变结论），我们可假设 $f(x) \geq 0$ 处处成立。

题目告诉我们，对某个 $c \in [0,1]$ 有 $f(c) \neq 0$，因此 $f(c) > 0$。由于 f 在 c 处连续，存在 $\delta > 0$，使得对 $[0,1]$ 中所有满足 $|x-c| < \delta$ 的 x，都有 $f(x) > \frac{1}{2} f(c)$。这个集合包含一个长度 $q-p$ 为正的区间 $[p,q] \subseteq [0,1]$（若 $c \leq \frac{1}{2}$，取 $[c, c+\eta]$；若 $c > \frac{1}{2}$，取 $[c-\eta, c]$，其中 $0 < \eta < \delta$ 且 $\eta \leq \frac{1}{2}$）。由于 $f \geq 0$ 处处成立，

$$\int_0^1 f(x)\,\mathrm{d}x \;\geq\; \int_p^q f(x)\,\mathrm{d}x \;\geq\; \tfrac{1}{2} f(c)\,(q-p) \;>\; 0,$$

这与 $\int_0^1 f(x)\,\mathrm{d}x = 0$ 矛盾。因此 f 在 $[0,1]$ 上同时取正值和负值。

想一想 为甚么单凭“$f \geq 0$ 且 $f(c) > 0$，所以积分是正数”并不足够?

一个函数可以只在单一点上为正，而积分完全察觉不到；单一点是没有宽度的。正是连续性把 c 处的正值扩展到一个长度为正的区间上，这才迫使积分为正。

第 (ii) 部分

展开平方并利用三个条件 $(*)$，

$$\begin{aligned}\int_0^1 (x-\alpha)^2 g(x)\,\mathrm{d}x &= \int_0^1 x^2 g(x)\,\mathrm{d}x - 2\alpha\int_0^1 x g(x)\,\mathrm{d}x + \alpha^2\int_0^1 g(x)\,\mathrm{d}x \\ &= \alpha^2 - 2\alpha\cdot\alpha + \alpha^2\cdot 1 = 0.\end{aligned}$$

现在用反证法，假设对每个 $x \in [0,1]$ 都有 $g(x) \neq 0$。那么 g 不可能在 $[0,1]$ 上同时取两种符号（否则根据题目开首的假设，它会在某处等于零），所以 g 要么处处严格为正，要么处处严格为负。

设 $f(x) = (x-\alpha)^2 g(x)$，它是连续的。由于 $(x-\alpha)^2 \geq 0$，凡 f 不为零之处，它都与 g 同号，所以 f 不会同时取两种符号。但 f 在 $[0,1]$ 中某处不为零：$(x-\alpha)^2$ 只在 $x = \alpha$ 处等于零，所以除了（可能的）α 之外，对每个 $x \in [0,1]$ 都有 $f(x) \neq 0$。而且 $\int_0^1 f(x)\,\mathrm{d}x = 0$。根据第 (i) 部分，$f$ 必须同时取两种符号，矛盾。因此对 $[0,1]$ 中某个 x，有 $g(x) = 0$。

一个线性例子。设 $g(x) = a + bx$，则条件 $(*)$ 变为

$$a + \tfrac{1}{2}b = 1, \qquad \tfrac{1}{2}a + \tfrac{1}{3}b = \alpha, \qquad \tfrac{1}{3}a + \tfrac{1}{4}b = \alpha^2.$$

由第一式得 $a = 1 - \frac{1}{2}b$；代入第二式得 $\frac{1}{2} + \frac{1}{12}b = \alpha$，所以

$$b = 12\alpha - 6, \qquad a = 4 - 6\alpha.$$

于是第三个条件变为

$$\tfrac{1}{3}(4 - 6\alpha) + \tfrac{1}{4}(12\alpha - 6) = \alpha - \tfrac{1}{6} = \alpha^2,$$

即 $6\alpha^2 - 6\alpha + 1 = 0$，所以

$$\alpha = \frac{3 \pm \sqrt{3}}{6}.$$

留意 α 并非可以随意选取：线性的 g 只在这两个值下才成立。取 $\alpha = \frac{3+\sqrt{3}}{6}$，得 $b = 2\sqrt{3}$ 及 $a = 1 - \sqrt{3}$，所以

$$g(x) = 1 - \sqrt{3} + 2\sqrt{3}\,x,$$

它在

$$x = \frac{\sqrt{3}-1}{2\sqrt{3}} = \frac{3-\sqrt{3}}{6} \approx 0.211,$$

处等于零，而这点位于 $[0,1]$ 内，正如第 (ii) 部分所说必然如此。（另一个选择 $\alpha = \frac{3-\sqrt{3}}{6}$ 给出 $g(x) = 1 + \sqrt{3} - 2\sqrt{3}\,x$，它在 $x = \frac{3+\sqrt{3}}{6} \approx 0.789$ 处等于零，同样位于 $[0,1]$ 内。）

第 (iii) 部分

我们证明 $g = h'$ 满足 $(*)$，其中 $\alpha = 1 - \beta$。这里 h' 是连续的，符合第 (ii) 部分的要求。首先，

$$\int_0^1 h'(x)\,\mathrm{d}x = h(1) - h(0) = 1.$$

以 $u = x$、$\mathrm{d}v = h'(x)\,\mathrm{d}x$ 进行分部积分，

$$\int_0^1 xh'(x)\,\mathrm{d}x = \Big[xh(x)\Big]_0^1 - \int_0^1 h(x)\,\mathrm{d}x = 1 - \beta.$$

再进行一次分部积分，这次取 $u = x^2$，

$$\begin{aligned}\int_0^1 x^2h'(x)\,\mathrm{d}x &= \Big[x^2h(x)\Big]_0^1 - 2\int_0^1 xh(x)\,\mathrm{d}x \\ &= 1 - \beta(2-\beta) = 1 - 2\beta + \beta^2 = (1-\beta)^2.\end{aligned}$$

? **想一想** 那个古怪的值 $\frac{1}{2}\beta(2-\beta)$ 是从哪里来的?

它正是为了令第三个积分等于第二个积分的平方而特意选取的。经过两次分部积分后，第二个积分是 $1-\beta$，所以我们需要第三个积分等于 $(1-\beta)^2$，而 $\frac{1}{2}\beta(2-\beta)$ 恰好做到这一点。

所以 h' 是连续的，并且满足 $(*)$，其中 $\alpha = 1 - \beta$。根据第 (ii) 部分，对 $[0,1]$ 中某个 x，有 $h'(x) = 0$。

验算

- 对于 $g(x) = 1 - \sqrt{3} + 2\sqrt{3}\,x$，直接积分得 $\int_0^1 g = 1$、$\int_0^1 xg = \frac{3+\sqrt{3}}{6} \approx 0.7887$ 及 $\int_0^1 x^2 g \approx 0.6220 = (0.7887)^2$。
- 第 (iii) 部分可用 $h(x) = (1-\sqrt{3})x + \sqrt{3}\,x^2$ 来测试，它的导数 (derivative) 正是上面的 g。那么 $h(0) = 0$、$h(1) = 1$、$\beta = \frac{3-\sqrt{3}}{6} \approx 0.2113$，以及 $\int_0^1 xh = \frac{1}{2}\beta(2-\beta) \approx 0.1890$。它的导数在 $x \approx 0.211$ 处等于零，所以 h 在 $[0,1]$ 内有一个转向点 (turning point)，与预测相符。

常见错误

- 在 (i) 中，不用连续性便论证“$f \geq 0$，所以 $\int f > 0$”。一个非负函数可以积分为零而又不处处为零（例如只在一点上不为零），所以连续性是不可或缺的。
- 在 (ii) 中，把 (i) 应用到 g 本身。我们对 $\int g$ 除了知道它等于 1 之外一无所知；第 (i) 部分必须应用到 $(x-\alpha)^2 g$ 上，而且我们必须验证这个函数在某处不为零。

► 在 (iii) 中，忘记了 h' 必须连续，(ii) 才可应用。正因如此，题目才说 h 有连续的导数。

STEP 之外

条件 $(*)$ 表示：如果 g 是一个概率密度（probability density），它的平均值（mean）便是 α，而它的方差（variance）$\int (x-\alpha)^2 g = 0$。真正的概率密度不可能有零方差，这就是为甚么 g 必须在某处跌至零以下。第 (i) 部分是实分析（real analysis）中的一条标准引理（lemma）：一个连续、非负且积分为零的函数必定恒等于零。它经常被使用，例如在变分法（calculus of variations）中。Abbott 的 *Understanding Analysis*（《理解分析》）对连续性和黎曼积分（Riemann integral）的发展足够严谨，既可证明这条引理，也可证明题目容许你假设的介值定理（intermediate value theorem）。

详细解答

STEP I 2017，第 7 题

纯数学 · 拿破仑定理

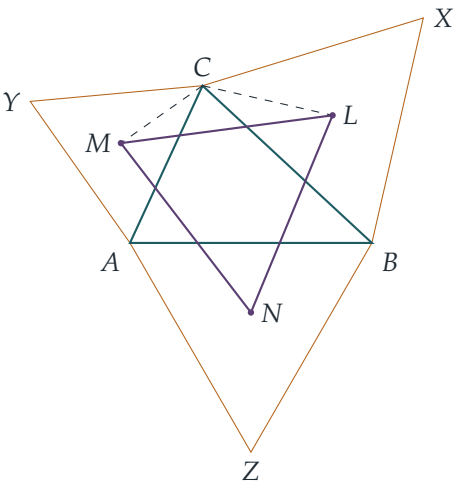
STEP I 2017，第 7 题

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/17-s1-q7

关键思路

一切都发生在 ABC 的顶点上。等边三角形（equilateral triangle）的中心（centre）位于其每个角的角平分线（angle bisector）上，所以由 C 到 L 及由 C 到 M 的直线，各自与 ABC 的一条边成 30°。由此可以用 C 表示角 LCM。再在三角形 LCM 中应用余弦公式（cosine rule），便得到 $|LM|$。结果原来是关于 a、b、c 对称（symmetric）的，而正是这种对称性令 LMN 成为等边三角形。



解答

第 (i) 部分

在边长为 b 的等边三角形 CYA 中，中心 M 位于全部三条中线（median）上，而这些中线同时也是高（altitude）和角平分线。由 C 出发的中线长度为 $\frac{\sqrt{3}}{2}b$，而中心位于由顶点起沿中线三分之二的位置。所以

$$|CM| = \tfrac{2}{3} \cdot \tfrac{\sqrt{3}}{2}\, b = \frac{b}{\sqrt{3}}.$$

同理，三角形 BXC 的边长为 a，所以

$$|CL| = \frac{a}{\sqrt{3}}.$$

第 (ii) 部分

由于 CM 平分 60° 的角 YCA，直线 CM 与 CA 成 30°，且位于远离 B 的一侧。同样地，CL 与 CB 成 30°，且位于远离 A 的一侧。绕 C 由 L 转到 M，先经过 30°，再经过三角形的角 C，然后再经过 30°。所以

$$\angle LCM = C + 60^\circ.$$

? **想一想** 如果 $C > 120^\circ$，以致 $C + 60^\circ$ 大于 180°，那又怎样?

这时三角形 LCM 在 C 的角是另一边的角，即 $360^\circ - (C + 60^\circ)$。但 $\cos(360^\circ - \theta) = \cos\theta$，所以下面的余弦公式在所有情况下都给出相同结果。

在三角形 LCM 中应用余弦公式，并代入第 (i) 部分的长度：

$$\begin{aligned} |LM|^2 &= \frac{a^2}{3} + \frac{b^2}{3} - \frac{2ab}{3}\cos(C + 60^\circ) \\ &= \frac{a^2 + b^2}{3} - \frac{2ab}{3}\left(\tfrac{1}{2}\cos C - \tfrac{\sqrt{3}}{2}\sin C\right) \\ &= \frac{a^2 + b^2}{3} - \frac{ab\cos C}{3} + \frac{\sqrt{3}\,ab\sin C}{3}. \end{aligned}$$

现在我们运用关于三角形 ABC 的两个事实。余弦公式给出 $2ab\cos C = a^2 + b^2 - c^2$，而面积为 $\Delta = \frac{1}{2}ab\sin C$，所以 $ab\sin C = 2\Delta$。两边乘以 6，

$$\begin{aligned} 6|LM|^2 &= 2(a^2 + b^2) - 2ab\cos C + 2\sqrt{3}\,ab\sin C \\ &= 2a^2 + 2b^2 - (a^2 + b^2 - c^2) + 4\sqrt{3}\,\Delta \\ &= a^2 + b^2 + c^2 + 4\sqrt{3}\,\Delta, \end{aligned}$$

得证。

若把 a、b、c 置换（permute），右方保持不变，而 Δ 也与三角形顶点的标记方式无关。在顶点 A（求 $|MN|$）及顶点 B（求 $|NL|$）重复同样的论证，会得到相同的式子。所以

$$|LM| = |MN| = |NL|,$$

因此 LMN 是等边三角形。

面积相等。边长为 s 的等边三角形面积为 $\frac{\sqrt{3}}{4}s^2$。设 $S = a^2 + b^2 + c^2$。那么

$$[LMN] = \frac{\sqrt{3}}{4}|LM|^2 = \frac{\sqrt{3}}{24}\big(S + 4\sqrt{3}\,\Delta\big) = \frac{\sqrt{3}}{24}S + \frac{\Delta}{2}.$$

因此

$$[LMN] = \Delta \iff \frac{\sqrt{3}}{24}S = \frac{\Delta}{2} \iff S = \frac{12\Delta}{\sqrt{3}} = 4\sqrt{3}\,\Delta.$$

每一步都可逆，所以这正是所要求的“当且仅当”。

第 (iii) 部分

利用 $c^2 = a^2 + b^2 - 2ab\cos C$ 及 $\Delta = \frac{1}{2}ab\sin C$，把所有量都用 a、b 和 C 表示。下面每一行都与上一行等价：

$$\begin{aligned} & a^2 + b^2 + c^2 = 4\sqrt{3}\,\Delta \\ \iff \quad & 2a^2 + 2b^2 - 2ab\cos C = 2\sqrt{3}\,ab\sin C \\ \iff \quad & a^2 + b^2 = ab\big(\cos C + \sqrt{3}\sin C\big) \\ \iff \quad & a^2 + b^2 = 2ab\cos(C - 60^\circ), \end{aligned}$$

因为 $\cos C + \sqrt{3}\sin C = 2\big(\cos C\cos 60^\circ + \sin C\sin 60^\circ\big) = 2\cos(C - 60^\circ)$。两边减去 $2ab$，

$$a^2 - 2ab + b^2 = 2ab\cos(C - 60^\circ) - 2ab \iff (a-b)^2 = -2ab\big(1 - \cos(C - 60^\circ)\big).$$

所以两个条件等价。

> ? **想一想** 一个平方怎么可能等于一个前面带负号的量?

只有两边都是零时才可以。左方 $(a-b)^2$ 至少是 0。右方则有 $ab > 0$ 及 $1 - \cos(C - 60^\circ) \geq 0$，所以右方至多是 0。因此等式成立若且仅若

$$a = b \quad \text{及} \quad \cos(C - 60^\circ) = 1.$$

由于 $0^\circ < C < 180^\circ$，我们有 $-60^\circ < C - 60^\circ < 120^\circ$，而在这范围内余弦等于 1 的角只有 0°。所以 $C = 60^\circ$。一个 $a = b$ 而顶角为 60° 的三角形，其余两角相等，且总和为 120°，所以三个角都是 60°，三角形是等边三角形。反之，等边三角形满足 $a = b$ 及 $C = 60^\circ$，所以它符合该条件。

结合第 (ii) 部分：$[LMN] = [ABC] \iff a^2 + b^2 + c^2 = 4\sqrt{3}\,\Delta \iff ABC$ 是等边三角形。

验算

- 边长为 1 的等边三角形 ABC：此时 $\Delta = \frac{\sqrt{3}}{4}$，而 $6|LM|^2 = 3+3 = 6$，所以 $|LM| = 1$，两个面积相等，与第 (iii) 部分的预测一致。
- 边长为 3–4–5 的直角三角形：$S = 50$ 及 $\Delta = 6$，所以 $|LM|^2 = (50+24\sqrt{3})/6 \approx 15.26$。于是 $[LMN] \approx 6.61$，大于 $[ABC] = 6$。用电脑以坐标计算点 L、M、N，得到相同的数值，并在数以千计的随机三角形（包括钝角三角形）上验证了 $6|LM|^2$ 的公式。
- 第 (iii) 部分亦显示对所有三角形都有 $a^2+b^2+c^2 \geq 4\sqrt{3}\,\Delta$，所以 $[LMN] \geq [ABC]$ 恒成立。

常见错误

- 把 $|CM|$ 当作整条中线 $\frac{\sqrt{3}}{2}b$，或其三分之一，而不是三分之二。
- 用了 $\angle LCM = C + 30^\circ$ 或 C，忘记了两侧各有一个 30°。
- 在第 (iii) 部分只得出“$a = b$”便停下来。你还需要 $C = 60^\circ$，并且要利用 C 的范围来排除其他余弦为 1 的角。

STEP 之外

这就是拿破仑定理（Napoleon's theorem）：在任意三角形的各边上向外作等边三角形，这些等边三角形的中心构成一个等边三角形。第 (iii) 部分得到的不等式 $a^2+b^2+c^2 \geq 4\sqrt{3}\,\Delta$ 是 Weitzenböck 不等式（Weitzenböck's inequality）。复数（complex numbers）可以给出拿破仑定理一个非常简洁的证明：旋转 60° 就是乘以 $e^{i\pi/3}$。Needham 的 *Visual Complex Analysis*（《可视化复分析》）把这种处理几何的方法发展得十分优美。

详细解答

STEP I 2017，第 8 题

纯数学 · 递推关系与数学归纳法

> **STEP I 2017，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/17-s1-q8

关键思路

表达式 $a_n^2 + 2a_nb_n - b_n^2$ 在递推关系（recurrence）的每一步下都保持不变，因此它永远等于初始值。把它除以 b_n^2，便得到一条关于比值 c_n 的精确方程，其中带有一个很小的误差项 $1/b_n^2$。对这条方程配方（completing the square）几乎完成了所有工作：它给出极限、比值位于 $\sqrt{2}-1$ 的哪一侧，以及 $\sqrt{2}$ 的有理数上下界。

解答

归纳法

当 $n = 1$ 时，$a_1^2 + 2a_1b_1 - b_1^2 = 1 + 4 - 4 = 1$，所以 $(*)$ 成立。

假设 $(*)$ 对某个 $n \geq 1$ 成立。记 $a = a_n$，$b = b_n$，

$$\begin{aligned} a_{n+1}^2 &= a^2 + 4ab + 4b^2, \\ 2a_{n+1}b_{n+1} &= 2(a+2b)(2a+5b) = 4a^2 + 18ab + 20b^2, \\ b_{n+1}^2 &= 4a^2 + 20ab + 25b^2. \end{aligned}$$

把前两式相加再减去第三式，得

$$a_{n+1}^2 + 2a_{n+1}b_{n+1} - b_{n+1}^2 = a^2 + 2ab - b^2 = 1.$$

因此 $(*)$ 对 $n+1$ 成立，由归纳法（induction）可知它对所有 $n \geq 1$ 成立。

第 (i) 部分

首先，每个 a_n 和 b_n 都是正数：这在 $n=1$ 时成立，而递推关系以正系数由 a_n 和 b_n 构成 a_{n+1} 和 b_{n+1}。因此

$$b_{n+1} = 2a_n + 5b_n > 5b_n,$$

因为 $a_n > 0$。由 $b_1 = 2 = 2\times 5^0$ 开始，每一步至少乘以 5，便得到对所有 $n \geq 1$ 有 $b_n \geq 2\times 5^{n-1}$（一行归纳即可）。

现在把 $(*)$ 除以非零的 b_n^2：

$$c_n^2 + 2c_n - 1 = \frac{1}{b_n^2}, \qquad \text{即} \qquad (c_n+1)^2 = 2 + \frac{1}{b_n^2}. \tag{†}$$

> **想一想** 应该取哪一个平方根?

由于 $c_n > 0$，数 $c_n + 1$ 是正数，所以我们取正平方根：

$$c_n + 1 = \sqrt{2 + \frac{1}{b_n^2}}.$$

当 $n \to \infty$ 时，$0 < 1/b_n^2 \leq \frac{1}{4}25^{-(n-1)} \to 0$，所以 $c_n + 1 \to \sqrt{2}$（平方根函数是连续（continuous）的）。因此 $c_n \to \sqrt{2} - 1$，得证。

第 (ii) 部分

由 (†)，$(c_n+1)^2 = 2 + 1/b_n^2 > 2$。由于 $c_n + 1 > 0$ 且 $\sqrt{2} > 0$，这给出 $c_n + 1 > \sqrt{2}$，即 $c_n > \sqrt{2} - 1$。

至于另一个界，以 2 分别除以 $c_n + 1 > \sqrt{2}$ 的两边；两边都是正数，所以不等号反转：

$$\frac{2}{c_n+1} < \frac{2}{\sqrt{2}} = \sqrt{2}.$$

合起来，

$$\frac{2}{c_n+1} < \sqrt{2} < c_n + 1.$$

> **想一想** 哪一个 n 值会给出 99/70?

我们需要 $c_n + 1 = (a_n + b_n)/b_n$ 等于 99/70，所以要找 $b_n = 70$。逐步计算递推关系：

$$(a_1, b_1) = (1, 2), \quad (a_2, b_2) = (5, 12), \quad (a_3, b_3) = (29, 70).$$

所以 $c_3 = \frac{29}{70}$，$c_3 + 1 = \frac{99}{70}$，而

$$\frac{2}{c_3+1} = \frac{140}{99}.$$

把 $n = 3$ 代入这个双向不等式，得

$$\frac{140}{99} < \sqrt{2} < \frac{99}{70}.$$

验算

- $(a_3, b_3) = (29, 70)$ 满足 $(*)$：$841 + 4060 - 4900 = 1$。
- $\frac{140}{99} \approx 1.414141$，$\frac{99}{70} \approx 1.414286$，而 $\sqrt{2} \approx 1.414214$ 位于两者之间。这个区间的宽度是 $\frac{1}{6930}$。
- 下一项 $(a_4, b_4) = (169, 408)$ 给出 $c_4 \approx 0.4142157$，与 $\sqrt{2} - 1$ 的差已在 2.2×10^{-6} 以内，这与误差大约按 $1/b_n^2$ 下降的情况吻合。
- 平方验证：$140^2 = 19600 < 2 \times 99^2 = 19602$，而 $99^2 = 9801 > 2 \times 70^2 = 9800$。

常见错误

- 展开 $n+1$ 时的 $(*)$ 而没有仔细合并同类项：ab 和 b^2 的系数只是刚好抵消（$4 + 18 - 20 = 2$ 及 $4 + 20 - 25 = -1$）。
- 声称 $b_{n+1} \geq 5b_n$，却没有指出 $a_n > 0$。
- 忘记说明为何取正平方根，或在取倒数时反转不等号，却没有说明两边都是正数。

STEP 之外

这个递推关系就是乘以矩阵（matrix）$\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$，其行列式（determinant）为 1，并保持二次型（quadratic form）$x^2 + 2xy - y^2$ 不变；寻找这类不变二次型正是佩尔方程（Pell's equation）理论的核心。数 $c_n + 1$，即 $\frac{3}{2}, \frac{17}{12}, \frac{99}{70}, \ldots$，是 $\sqrt{2}$ 的连分数渐近分数（continued-fraction convergent）中每隔一个所取的项；渐近分数是在其分母大小范围内最佳的有理逼近。Davenport 的 *The Higher Arithmetic*（《高等算术》）对连分数和佩尔方程有精彩的阐述。

详细解答

STEP I 2017，第 9 题

力学 · 抛体运动

STEP I 2017，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：

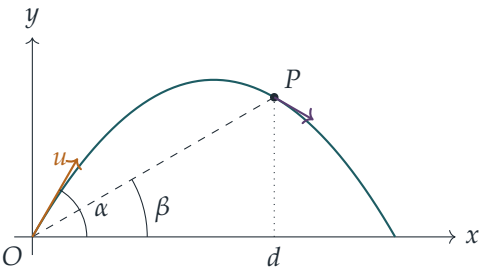
step.maths.org/questions/17-s1-q9

关键思路

先写出轨迹方程（trajectory equation），并令它通过 P。这样便得到只以投射角为变量的 u^2 表达式。最方便处理的变量是 $T = \tan\alpha$，因为 $\sec^2\alpha = 1 + T^2$ 能把一切化成 T 的有理函数（rational function）。求这个函数的最小值只是普通的微积分，而驻点（stationary point）方程原来是一条经过伪装的倍角公式（double-angle formula）。

解答

把质点视为只受重力作用而自由运动（不计空气阻力），以 O 为原点，x 轴水平指向 P，y 轴铅直向上。设 $T = \tan\alpha$ 及 $k = \tan\beta > 0$。



第 (i) 部分

水平速度为 $u\cos\alpha$，因此质点在时间 $t = x/(u\cos\alpha)$ 到达水平距离 x，此时其高度为

$$y = x\tan\alpha - \frac{gx^2}{2u^2\cos^2\alpha} = xT - \frac{gx^2(1+T^2)}{2u^2}.$$

通过 $P=(d,\,dk)$ 即表示

$$dk = dT - \frac{gd^2(1+T^2)}{2u^2} \quad\Longrightarrow\quad u^2 = \frac{gd}{2}\cdot\frac{1+T^2}{T-k}.$$

右方必须为正，所以只有 $T>k$，即 $\alpha>\beta$ 的角度才可行。这很合理：我们必须瞄准 P 的上方。因此我们要求以下函数的最小值：

$$f(T)=\frac{1+T^2}{T-k}\qquad (T>k).$$

由商法则（quotient rule），

$$f'(T)=\frac{2T(T-k)-(1+T^2)}{(T-k)^2}=\frac{T^2-2kT-1}{(T-k)^2}.$$

分子在 $T=k\pm\sqrt{k^2+1}$ 时为零，其中只有取 $+$ 号的根大于 k。由于当 $T\to k^+$ 及 $T\to\infty$ 时都有 $f(T)\to\infty$，而两者之间只有一个驻点，所以该点给出最小值。因此最佳的 α 满足

$$T^2-2kT-1=0. \tag{$*$}$$

想一想 不代入根式的话，怎样在最小值处化简 $f(T)$?

利用 $(*)$ 改写各式。由它得 $k=\dfrac{T^2-1}{2T}$，故

$$T-k=\frac{2T^2-T^2+1}{2T}=\frac{1+T^2}{2T},\qquad \text{因此}\qquad f(T)=\frac{1+T^2}{(1+T^2)/(2T)}=2T.$$

所以

$$u^2=\frac{gd}{2}\cdot 2T=gd\tan\alpha,$$

正如所求。

至于角度，把 $(*)$ 写成 $k=\dfrac{T^2-1}{2T}$ 的形式，即为

$$\tan\beta=-\frac{1-\tan^2\alpha}{2\tan\alpha}=-\cot 2\alpha=\tan(2\alpha-90^\circ),$$

这里用了 $\tan 2\alpha=\dfrac{2\tan\alpha}{1-\tan^2\alpha}$ 及 $-\cot\theta=\tan(\theta-90^\circ)$。（若 $\alpha=45^\circ$，则 $(*)$ 给出 $k=0$，但由于 $\beta>0$，这情况已被排除，所以 $\cot 2\alpha$ 有定义。）由于 $0^\circ<\alpha<90^\circ$，角 $2\alpha-90^\circ$ 位于 $(-90^\circ,90^\circ)$ 内，β 亦然，而 tan 在该区间上是一一对应（one-to-one）的。所以 $\beta=2\alpha-90^\circ$，即

$$2\alpha=\beta+90^\circ.$$

换言之，最佳投射方向平分直线 OP 与铅直向上方向之间的夹角。

第 (ii) 部分

在 P 点，水平速度仍为 $u\cos\alpha$。所需时间为 $d/(u\cos\alpha)$，因此铅直速度为

$$v_y = u\sin\alpha - \frac{gd}{u\cos\alpha}.$$

若质点在 P 点的运动方向与水平成仰角 θ，则

$$\tan\theta = \frac{v_y}{u\cos\alpha} = \tan\alpha - \frac{gd}{u^2\cos^2\alpha} = T - \frac{gd(1+T^2)}{gdT} = T - \frac{1+T^2}{T} = -\frac{1}{T} = -\cot\alpha,$$

这里用了第 (i) 部分的 $u^2 = gdT$。所以 $\tan\theta = \tan(\alpha - 90^\circ)$，又由于 θ 介乎 -90° 与 90° 之间，故 $\theta = \alpha - 90^\circ$。

质点通过 P 时正在向下运动，方向与水平成俯角 $90^\circ - \alpha$。它在 P 点的方向与投射方向互相垂直。

验算

- 取 $\beta = 30^\circ$。则 $\alpha = 60^\circ$ 及 $u^2 = \sqrt{3}\,gd$。直接计算，$f(T)$ 在 $T = \sqrt{3}$ 的值为 $4/(\sqrt{3} - \frac{1}{\sqrt{3}}) = 2\sqrt{3}$，与 $2T$ 吻合。最终方向为水平以下 30°，与图中一致。
- 令 $\beta \to 0$，得 $\alpha \to 45^\circ$ 及 $u^2 \to gd$。这正是熟悉的事实：在水平地面上，以 45° 投射可得最大射程 u^2/g。此时质点以水平以下 45° 著地，同样与发射方向垂直。
- 对 200 组随机选取的 d 和 β，在 α 上作数值搜索，每次所得的最小化角度都是 $\frac{1}{2}(\beta + 90^\circ)$，最小值都是 $gd\tan\alpha$。

常见错误

- 忽略条件 $T > k$，保留了根 $k - \sqrt{k^2+1}$。该根是负数，意味着向下投射。
- 由正切值相等推出 $\beta = 2\alpha - 90^\circ$，却没有说明为何两个角都位于 tan 为一一对应的区间内。
- 在 (ii) 中给出角度 $90^\circ - \alpha$，却没有说明质点正在下降。

STEP 之外

对于固定的速率 u，能到达的点都位于一条抛物线之内，这条抛物线就是该族轨迹的包络线（envelope），又称“安全抛物线”（parabola of safety）。到达 P 的最小速率轨迹，正是使 P 落在该包络线上的那一条，而 (ii) 中的垂直关系是这条曲线的几何性质之一。曲线

族的包络线以及带约束的最优化（constrained optimisation），都会在大学一年级课程中进一步发展。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）是一本好书，可以看到如何以这种严谨态度处理抛射体问题。

详细解答

STEP I 2017，第 10 题

力学 · 连锁碰撞

STEP I 2017，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/17-s1-q10

关键思路

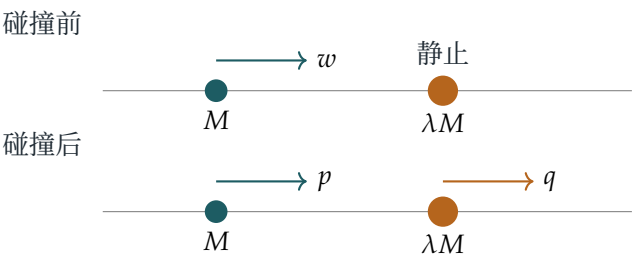
链中每一次碰撞的情况都一样：一粒质量为 M、以速率 w 运动的质点，撞向一粒质量为 λM 的静止质点。因此我们只需以比值 λ 表示，把这一次碰撞解一遍，整条碰撞链便可以重复套用这个结果得出，并得到等比数列（geometric sequence）。之后每一部分的关键，就是比较各质点碰撞后的速度，看看有没有质点追上前面的质点。

解答

把所有质点视为沿 x 轴运动的点，速度以 x 正方向为正。

一次典型的碰撞

质量为 M、以速率 w 运动的质点 A，撞向质量为 λM 的静止质点 B。设碰撞后两者的速度分别为 p 和 q。



由动量守恒（conservation of momentum）及牛顿（Newton）碰撞实验定律（Newton's

experimental law），得

$$Mw = Mp + \lambda Mq, \qquad q - p = ew.$$

把 $p = q - ew$ 代入第一条方程，得 $w = q - ew + \lambda q$，因此

$$q = \frac{1+e}{1+\lambda}\, w, \qquad p = q - ew = \frac{1-\lambda e}{1+\lambda}\, w.$$

记 $r = \dfrac{1+e}{1+\lambda}$ 及 $s = \dfrac{1-\lambda e}{1+\lambda}$。留意

$$q - p = ew > 0,$$

所以每次这样的碰撞刚完结时，两粒质点都会分开。

第 (i) 部分

第一次碰撞是 P（质量 m，速率 u）撞向 P_1（质量 λm）。取 $w = u$ 及 $M = m$，得

$$u_1 = \frac{1+e}{1+\lambda}\, u,$$

即所求结果，而 P 碰撞后的速度为 su。

接着 P_1 向前运动，撞向质量为 $\lambda^2 m = \lambda \cdot (\lambda m)$ 的 P_2。质量比同样是 λ，所以这是同一类碰撞，只是 $w = u_1$。一般而言，P_n（质量 $\lambda^n m$）以速率 u_n 撞向 P_{n+1}（质量 $\lambda^{n+1} m$），所以 $u_{n+1} = ru_n$，而 P_n 碰撞后的速度为 su_n。因此

$$u_n = \left(\frac{1+e}{1+\lambda}\right)^n u.$$

这里 P_n 的第一次碰撞是被 P_{n-1} 撞击，第二次则是它撞向 P_{n+1}。（第一次碰撞刚完结时，P_n 正远离 P_{n-1}，所以接下来它会遇上的事，就是到达 P_{n+1}，前提是后面没有质点先追上它。这正是题目所设想的事件次序，而在第 (ii)–(iv) 部分中，我们会验证情况确实如此。）所以

$$v_n = su_n = \frac{(1-\lambda e)(1+e)^n}{(1+\lambda)^{n+1}}\, u,$$

这是以 x 正方向为正的速度。作为速率，它是上式的绝对值；只有当 $\lambda e > 1$，即 P_n 反弹时，这一点才有影响。

第 (ii) 部分

设 $e > \lambda$。那么 $\lambda < e < 1$，所以 $\lambda e < 1$ 及 $s > 0$；此外 $r > 1$。

碰撞依次发生：先是 P 与 P_1，然后 P_1 与 P_2，再然后 P_2 与 P_3，如此类推。当 P_n 撞向 P_{n+1} 之后，由队尾起计，$P, P_1, \ldots, P_n, P_{n+1}, P_{n+2}, \ldots$ 的速度为

$$su,\ su_1,\ su_2,\ \ldots,\ su_n,\ u_{n+1},\ 0,\ 0,\ \ldots$$

由于 $u_k = r^k u$，其中 $r > 1$ 及 $s > 0$，

$$0 < su < su_1 < su_2 < \cdots < su_n,$$

而且 $u_{n+1} > su_n$，因为 $u_{n+1} - su_n = (r-s)u_n = eu_n > 0$。所以在 $P, P_1, \ldots, P_{n+1}$ 之中，每粒质点都严格地比后面那粒快，任何两粒都不可能再相遇。唯一尚未发生的碰撞，是 P_{n+1} 到达静止的 P_{n+2}；之后把 n 换成 $n+1$，同样的论证便再次适用。

所以每粒 P_n（$n \geq 1$）先被 P_{n-1} 撞击一次，然后撞向 P_{n+1} 一次，此后再不参与任何碰撞：每粒恰好经历两次碰撞。（这也证实了第 (i) 部分所假设的事件次序。）

第 (iii) 部分

若 $e = \lambda$，则 $r = 1$ 及 $s = \dfrac{1-e^2}{1+e} = 1 - e$。所以

$$u_n = u \quad \text{且} \quad v_n = (1-e)u \text{ 对每个} n,$$

而 P 碰撞后的速度也是 $(1-e)u$。整个运动就像一个沿直线传播的脉冲（pulse）：最前面的质点总是以速率 u 运动，撞上下一粒质点时，把速率 u 传下去，自己则降至 $(1-e)u$。后面所有质点都以 $(1-e)u$ 一起运动，所以它们之间不会再有碰撞，而前面以速率 u 运动的质点则离它们愈来愈远。每粒 P_n 恰好经历两次碰撞。

经过 N 次碰撞后，$P, P_1, \ldots, P_{N-1}$ 以 $(1-e)u$ 运动，而 P_N（质量 $\lambda^N m = e^N m$）则以 u 运动。动能（kinetic energy）为

$$\begin{aligned}\tfrac{1}{2}m(1-e)^2u^2\left(1 + e + \cdots + e^{N-1}\right) + \tfrac{1}{2}e^N mu^2 &= \tfrac{1}{2}mu^2\left[(1-e)^2\,\frac{1-e^N}{1-e} + e^N\right] \\ &= \tfrac{1}{2}mu^2\left[(1-e)(1-e^N) + e^N\right] \\ &= \tfrac{1}{2}mu^2\left[1 - e + e^{N+1}\right].\end{aligned}$$

除以初始能量 $\frac{1}{2}mu^2$，经过 N 次碰撞后损失的比例为

$$1 - \left(1 - e + e^{N+1}\right) = e\left(1 - e^N\right) \to e \quad \text{当} N \to \infty,$$

因为 $0 < e < 1$。这就是所求结果。

第 (iv) 部分

若 $\lambda e = 1$，则 $s = 0$ 且

$$r = \frac{1+e}{1+1/e} = \frac{e(1+e)}{e+1} = e.$$

所以每次碰撞都令运动中的质点完全停下：P 停下，然后 P_1 以 eu 向前运动，撞向 P_2 后停下，如此类推，其中 $u_n = e^n u$ 及 $v_n = 0$。每粒 P_n 恰好经历两次碰撞，最终各质点都静止，而且每粒都紧贴下一粒，同时只有单独一粒质点把运动带向前，而且愈来愈慢。由于 P_n 需要时间 $1/(e^n u)$ 才到达 P_{n+1}，相邻两次碰撞之间的时间间隔会无限增大。

想一想 动量守恒，那么能量又怎会流失?

经过 N 次碰撞后，只有 P_N 在运动，其质量为 $\lambda^N m = e^{-N}m$，速率为 $e^N u$。它的动量是 $e^{-N}m \cdot e^N u = mu$，理应如此；但它的动能是

$$\tfrac{1}{2}\, e^{-N}m\,(e^N u)^2 = e^N \cdot \tfrac{1}{2}mu^2.$$

动量由愈来愈重、愈来愈慢的质点承载；在动量 p 固定时，动能 $p^2/(2M)$ 会随质量 M 增大而趋于零。剩余的比例 e^N 趋于 0，所以最终初始动能会全部损失。

验算

- 取 $e = 1$ 及 $\lambda = 1$（质量相等、完全弹性），得 $u_1 = u$ 且 P 停下：这正是大家熟悉的牛顿摆（Newton's cradle）交换速度现象。
- 在第 (iii) 部分中,单次碰撞($N = 1$)损失的比例为 $e(1-e)$。直接计算:P 最终以 $(1-e)u$ 运动, 而质量为 em 的 P_1 则以 u 运动, 所以剩余能量是原来的 $(1-e)^2 + e = 1 - e + e^2$, 两者吻合。
- 以精确分数逐步进行碰撞的电脑模拟, 证实了 u_n 和 v_n 的公式、$e > \lambda$ 时每粒质点各经历两次碰撞的规律、损失比例 $e(1-e^N)$（当 $e = \frac{1}{2}$，经过 25 次碰撞后为 0.49999998），以及 $\lambda e = 1$ 时的剩余比例 e^N。

常见错误

- 把牛顿定律写成 $q - p = -ew$，或把次序倒转：分离速率等于接近速率的 e 倍，而这里两者都是正数。
- 忘记每一次碰撞的质量比都是 λ，而正是这一点令答案成为等比数列。
- 未经验证便假设事件的次序。当 $e < \lambda$ 且 $\lambda e < 1$ 时，落在后面的质点可能在前面那粒到达下一粒质点之前就追上它，简单的规律便不再成立。

STEP 之外

第 (iv) 部分显示动量和能量的表现截然不同：动量完全守恒，但能量却因为由愈来愈重的质点承载而完全流失。互相碰撞的质点链，是脉冲在密度变化的介质中传播的一个简单模型，并可引伸至离散晶格（discrete lattice）中的波动传播。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）就以同样严谨的态度讨论碰撞及质心参考系（centre-of-mass frame）。

详细解答

STEP I 2017，第 11 题

力学 • 涉及摩擦力的刚体静力学

STEP I 2017，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/17-s1-q11

关键思路

一切都发生在箱子的一个铅垂截面内，而这个截面是一个倾斜了 α 的正方形。因此自然的坐标轴应沿箱底和沿箱侧向上，而不是水平和铅垂方向：这样两个法向反作用力（normal reaction）都沿坐标轴，只有重量需要分解。两次分解力给出法向反作用力之间的一条关系，对中点取力矩（moment）的方程（重量在此不产生力矩）给出第二条关系，再利用 tan 的复角公式（compound-angle formula）比较两者，便得出答案。

解答

模型与受力

把杆视为均匀刚性杆（uniform rigid rod），并在包含它的铅垂平面内讨论。在这个平面内，箱底沿最大斜率线（line of greatest slope）放置，与水平成角 α，而经过 O 的箱侧与箱底垂直。在 O 取坐标轴：

- x 沿箱底，指向斜面上方（朝向 A）；
- y 沿箱侧，垂直于斜面（朝向 B）。

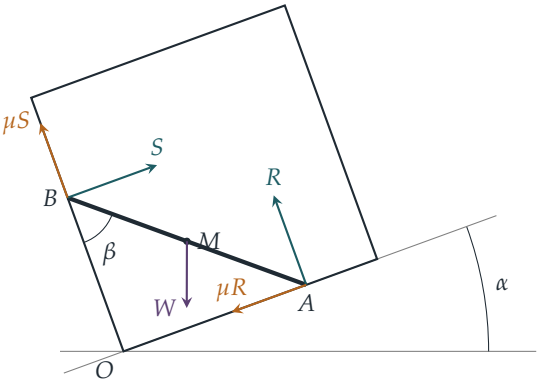
箱侧与箱底垂直，所以三角形 OAB 在 O 处为直角，而 B 处的角为 β。因此

$$OA = 2L\sin\beta, \qquad OB = 2L\cos\beta.$$

设 $\mu = \tan\gamma$。杆所受的力为：

- 在 A：沿 $+y$ 方向的法向反作用力 R，以及朝向 O（即 $-x$ 方向）的摩擦力 μR，因为 A 即将向远离 O 的方向滑动，而摩擦力处于极限状态（limiting）；

- 在 B：沿 $+x$ 方向的法向反作用力 S，以及远离 O（即 $+y$ 方向）的摩擦力 μS，因为 B 即将朝 O 滑动；
- 作用于中点 M、铅垂向下的重量 W。y 轴与向上的铅垂方向成角 α，所以重量沿 x 的分量为 $-W\sin\alpha$，沿 y 的分量为 $-W\cos\alpha$。



分解力

沿箱底（x）及沿箱侧向上（y）：

$$S-\mu R-W\sin\alpha=0, \tag{1}$$

$$R+\mu S-W\cos\alpha=0. \tag{2}$$

由于 $R+\mu S=W\cos\alpha\neq 0$，把 (1) 除以 (2) 得

$$\frac{S-\mu R}{R+\mu S}=\tan\alpha. \tag{3}$$

对中点取力矩

重量作用于 M，所以它对 M 没有力矩。相对于 M，点 A 位于 $(L\sin\beta,\,-L\cos\beta)$，而 B 位于 $(-L\sin\beta,\,L\cos\beta)$。按图中逆时针方向取力矩（作用于相对位置 (p,q) 的力 (F_x,F_y) 的力矩为 pF_y-qF_x）：

$$\text{在}A:\quad L\sin\beta\cdot R-(-L\cos\beta)(-\mu R)=LR(\sin\beta-\mu\cos\beta),$$

$$\text{在}B:\quad -L\sin\beta\cdot\mu S-L\cos\beta\cdot S=-LS(\cos\beta+\mu\sin\beta).$$

两者之和必须为零，所以

$$R(\sin\beta-\mu\cos\beta)=S(\cos\beta+\mu\sin\beta).$$

现在代入 $\mu = \tan\gamma$，并把整条式乘以 $\cos\gamma$（它是正数）：

$$R(\sin\beta\cos\gamma - \cos\beta\sin\gamma) = S(\cos\beta\cos\gamma + \sin\beta\sin\gamma),$$
$$\text{即}\quad R\sin(\beta-\gamma) = S\cos(\beta-\gamma).$$

想一想 我们可以除以 R 和 $\cos(\beta-\gamma)$ 吗?

两个法向反作用力都是推力，所以 $R>0$ 且 $S>0$。于是 $\sin(\beta-\gamma)$ 与 $\cos(\beta-\gamma)$ 不可能同时为零；又因为 $S\cos(\beta-\gamma) = R\sin(\beta-\gamma)$ 且 $R,S>0$，两者同号，所以两者都不为零，并且 $\tan(\beta-\gamma)>0$。由于 β 和 γ 都在 $(0,\frac{1}{2}\pi)$ 内，$\beta-\gamma$ 在 $(-\frac{1}{2}\pi,\frac{1}{2}\pi)$ 内，因此事实上

$$\frac{S}{R} = \tan(\beta-\gamma), \qquad 0<\beta-\gamma<\tfrac{1}{2}\pi. \tag{4}$$

综合起来

把 (3) 左边的分子和分母都除以 R，并记 $k = S/R = \tan(\beta-\gamma)$ 及 $\mu = \tan\gamma$：

$$\tan\alpha = \frac{k-\mu}{1+k\mu} = \frac{\tan(\beta-\gamma)-\tan\gamma}{1+\tan(\beta-\gamma)\tan\gamma} = \tan\big((\beta-\gamma)-\gamma\big) = \tan(\beta-2\gamma),$$

这里用了复角公式（分母 $1+k\mu$ 是正数，所以没有任何无定义的情况）。由 (4) 及 $0<\gamma<\frac{1}{2}\pi$，

$$\beta-2\gamma = (\beta-\gamma)-\gamma \in \left(-\tfrac{1}{2}\pi,\ \tfrac{1}{2}\pi\right),$$

而且 $\alpha\in(0,\frac{1}{2}\pi)$ 亦成立。正切函数在 $(-\frac{1}{2}\pi,\frac{1}{2}\pi)$ 上是一对一（one-to-one）的，所以 $\tan\alpha = \tan(\beta-2\gamma)$ 迫使

$$\beta = \alpha + 2\gamma,$$

得证。这与题目所给的条件 $\alpha<\beta$ 相符；它亦告诉我们，这种状态只有在 $\alpha+2\gamma<\frac{1}{2}\pi$ 时才可能出现。

验算

- 光滑箱子：当 $\gamma=0$ 时，结果表示 $\beta=\alpha$。可以直接验证：没有摩擦时，两个法向反作用力相交于 P，即以 AB 为对角线、边沿箱子方向的长方形的第四个顶点。三个力处于平衡时必须共点（concurrent），所以重量经过 M 的作用线必须经过 P，即 MP 是铅垂的。对角线 MP 平行于把 OB 向远离 O 的方向转过 β 后的方向，而 OB 本身朝 O 的一侧与铅垂成 α 倾斜；所以 MP 与铅垂成 $\beta-\alpha$，并且恰好在 $\beta=\alpha$ 时是铅垂的。
- 数值例子：取 $\alpha=20°$、$\gamma=15°$，所以 $\beta=50°$。于是 (1) 和 (2) 给出 $R\approx 0.791W$ 及 $S\approx 0.554W$（两者皆为正），$S/R\approx 0.700\approx\tan 35°$，而对 M 的总力矩算出来为零。

常见错误

- 沿水平和铅垂方向分解力：这样做可行，但四个接触力每一个都需要分解，代数运算会繁重得多。
- 弄错摩擦力的方向。摩擦力与即将发生的运动方向相反：在 A 朝向 O，在 B 远离 O（沿箱侧向上）。
- 由 $\tan\alpha = \tan(\beta - 2\gamma)$ 直接下结论，却没有说明为何两个角都位于 tan 为一对一的区间内。

STEP 之外

把摩擦力写成 μR 并令 $\mu = \tan\gamma$，正是*摩擦角*（angle of friction）的概念：在极限平衡（limiting equilibrium）时，总接触力位于一个以法线为轴、半顶角为 γ 的“摩擦锥”（friction cone）的边缘上。关于刚体倚靠粗糙表面的问题，于是变成关于这些锥体以及三力共点的几何问题；同一幅图像亦是楔子、梯子以及叠放物体稳定性等问题的基础。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）从基本原理出发，细致地发展了刚体静力学和力矩。

详细解答

STEP I 2017，第 12 题

概率与统计 · 期望值与条件化

STEP I 2017，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/17-s1-q12

关键思路

主办者收取的款项永远相同，都是 Nc；唯一的不确定之处，在于究竟需不需要派发奖金。因此期望利润（expected profit）等于收入减去 J 乘以“有人选中中奖号码”的概率，而整个问题最终归结为其互补事件（complementary event），即没有人选中。这个概率是一个幂 $(1-p)^N$，而近似式 $(*)$ 正是为处理这类式子而设的。

解答

第 (i) 部分

设 W 为中奖号码，并设 E 为没有参加者选中 W 的事件。无论 W 结果是甚么，每位参加者都独立（independently）地以概率 $1-\frac{1}{N}$ 避开它，所以

$$\mathrm{P}(E)=\left(1-\frac{1}{N}\right)^N.$$

在任何情况下，主办者都收到 Nc。若 E 发生，则无人中奖，利润为 Nc；否则奖金由中奖者瓜分，但不论中奖者有多少人，派出的总额都是 J，所以利润为 $Nc-J$。因此

$$\text{期望利润}=Nc-J\,\mathrm{P}(E^c)=Nc-J\left[1-\left(1-\frac{1}{N}\right)^N\right].$$

在 $(*)$ 中取 $a=1$，得 $(1-\frac{1}{N})^N\approx \mathrm{e}^{-1}$，所以期望利润约为 $Nc-J(1-\mathrm{e}^{-1})$。若 $J=2Nc$，则变为

$$Nc-2Nc\left(1-\mathrm{e}^{-1}\right)=Nc\left(\frac{2}{\mathrm{e}}-1\right).$$

由于 $\mathrm{e}>2$，我们有 $2/\mathrm{e}<1$，所以此值为负：主办者预期会亏蚀。

第 (ii) 部分

共有 γN 个热门号码和 $(1-\gamma)N$ 个冷门号码。每位参加者恰好选一个号码，所以全部 N 个号码的概率之和为 1：

$$\gamma N \cdot \frac{a}{N} + (1-\gamma)N \cdot \frac{b}{N} = 1 \quad \Longrightarrow \quad \gamma a + (1-\gamma)b = 1.$$

主办者仍然均匀地（uniformly）选取 W，所以 W 是热门号码的概率为 γ，是冷门号码的概率为 $1-\gamma$。

想一想 参加者不再均匀地选号。这会否改变我们求 $\mathrm{P}(E)$ 的方法?

只会通过避开 W 的概率产生影响，而这个概率现在取决于 W 属于哪一类号码。因此我们以此作条件（condition）。若 W 是热门号码，每位参加者独立地以概率 $1-\frac{a}{N}$ 避开它；若是冷门号码，则以概率 $1-\frac{b}{N}$ 避开。由全概率定律（law of total probability）及 $(*)$，

$$\mathrm{P}(E) = \gamma\left(1-\frac{a}{N}\right)^N + (1-\gamma)\left(1-\frac{b}{N}\right)^N \approx \gamma\,\mathrm{e}^{-a} + (1-\gamma)\,\mathrm{e}^{-b}.$$

与第 (i) 部分一样，期望利润为 $Nc - J\big(1-\mathrm{P}(E)\big)$，所以

$$\text{期望利润} \approx \gamma J\,\mathrm{e}^{-a} + (1-\gamma)J\,\mathrm{e}^{-b} + (Nc - J),$$

这正是所要求的形式，其中

$$A = \gamma J, \qquad B = (1-\gamma)J, \qquad C = Nc - J.$$

情况 $\gamma = \frac{1}{8}$，$a = 9b$。由上述关系得 $\frac{1}{8}\cdot 9b + \frac{7}{8}b = 1$，即 $2b = 1$，所以

$$b = \tfrac{1}{2}, \qquad a = \tfrac{9}{2}.$$

当 $J = 2Nc$ 时，我们有 $A = \frac{1}{4}Nc$、$B = \frac{7}{4}Nc$ 及 $C = -Nc$，所以期望利润约为

$$Nc\left(\tfrac{1}{4}\mathrm{e}^{-9/2} + \tfrac{7}{4}\mathrm{e}^{-1/2} - 1\right).$$

括号内第一项为正，所以只需证明 $\frac{7}{4}\mathrm{e}^{-1/2} > 1$，即 $\sqrt{\mathrm{e}} < \frac{7}{4}$。由于 $\mathrm{e} < 3$，

$$\sqrt{\mathrm{e}} < \sqrt{3} < \tfrac{7}{4},$$

其中最后一步成立，是因为两边皆为正，且 $3 < \frac{49}{16}$（即 $48 < 49$）。因此括号内的值为正，主办者预期会获利。奖金与第 (i) 部分相同，但由于中奖号码往往是冷门号码，奖金无人认领的情况更为常见。

验算

- 当 $J = 2Nc$ 时，第 (i) 部分给出的期望利润约为 $(2/\mathrm{e} - 1)Nc \approx -0.264\,Nc$，而最后一部分给出 $\left(\frac{1}{4}\mathrm{e}^{-4.5} + \frac{7}{4}\mathrm{e}^{-0.5} - 1\right)Nc \approx 0.064\,Nc$。正如所述，是一笔小额利润。
- 当 $N = 800$ 时，用精确的幂 $(1 - a/N)^N$ 计算得 $0.0640\,Nc$，可见近似式 $(*)$ 在此非常准确。
- 若 $\gamma = 1$（每个号码都是“热门”的），上述关系迫使 $a = 1$，而第 (ii) 部分的公式便化为第 (i) 部分的 $J\mathrm{e}^{-1} + Nc - J$。
- 以电脑对 $N = 4$ 作完全枚举（参加者选号的全部 4^4 种方式），结果与两条未经近似的公式皆吻合。

常见错误

- 纠结于奖金如何分配。只要至少有一人中奖，不论中奖人数多少，主办者派出的总额都是 J。
- 误以为中奖号码是热门号码的概率取决于 a 或 b。主办者是均匀选号的，所以它是热门号码的概率为 γ。
- 把关系式写成 $a + b = 1$ 或 $\gamma a + b = 1$，忘记计算每类号码各有多少个。

STEP 之外

选中中奖号码的参加者人数服从参数为 N 和 p 的二项分布（binomial distribution），而 $(*)$ 正是理解以下事实的第一步：当 N 很大而 Np 固定时，它近似于泊松分布（Poisson distribution）——$(1 - p)^N \to \mathrm{e}^{-Np}$ 不过就是泊松分布取值为零的概率。现实中设有奖金累积（roll-over）的彩票正是以这种方式分析的，而选择冷门号码的玩家确实有较高的期望回报。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）一书，详细发展了泊松极限，以及类似本题所用的条件论证。

详细解答

STEP I 2017，第 13 题

概率与统计 • 概率的递推关系

> **STEP I 2017，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/17-s1-q13

关键思路

逐片面包去看，问它是怎样被用掉的：它可以做多士，可以是三文治的第一片，也可以是三文治的第二片。三者恰好发生其一，而前一片面包的情况决定了哪几种是可能的。若第 $r-1$ 片完成了一份小食，第 r 片便开始一份新的小食；若没有完成，第 r 片必然是某份三文治的后半。这样便得到一个一阶递推关系（first-order recurrence），我们用惯常的技巧——减去不动点（fixed point）——来求解。

解答

设置

按使用次序把面包编号为 $1, 2, \ldots, n$。对每一片面包 r，以下情况恰好发生其一：

- T_r：第 r 片用来做多士；
- A_r：第 r 片是某份三文治两片面包中的第一片；
- S_r：第 r 片是某份三文治两片面包中的第二片。

因此 $t_r = \mathrm{P}(T_r)$、$s_r = \mathrm{P}(S_r)$，而 $\mathrm{P}(A_r) = 1 - s_r - t_r$。若 T_r 或 S_r 发生，我们说第 r 片完成一份小食；其概率为 $s_r + t_r$。

第一片面包开始第一份小食，所以它不可能是三文治的第二片：

$$s_1 = 0.$$

（此外，若 $n \geq 2$ 则 $t_1 = p$；若 $n = 1$ 则 $t_1 = 1$。）

两条方程

关于 t_r 的方程。第 r 片开始一份新小食，当且仅当第 $r-1$ 片完成一份小食，其概率为 $s_{r-1}+t_{r-1}$。已知第 r 片开始一份小食，余下的是第 $r, r+1, \ldots, n$ 片；当 $r \le n-1$ 时至少有两片，所以无论之前发生了甚么，我选择多士的概率都是 p。若第 $r-1$ 片没有完成小食，第 r 片便完成一份三文治，不是多士。因此

$$t_r = (s_{r-1} + t_{r-1})\, p \qquad (2 \le r \le n-1).$$

这条式在 $r = n$ 时不成立，因为从最后一片开始的小食必定是多士。

关于 s_r 的方程。第 r 片是三文治的第二片，当且仅当第 $r-1$ 片是三文治的第一片，即第 $r-1$ 片没有完成小食。所以

$$s_r = 1 - (s_{r-1} + t_{r-1}) \qquad (2 \le r \le n).$$

这里不需要对余下片数加任何条件，所以此式一直成立至 $r = n$。

s_r 的递推关系

把第二条方程中的 r 换成 $r+1$，得 $s_r + t_r = 1 - s_{r+1}$，所以第一条方程可写成

$$t_r = p\,(1 - s_r) \qquad (2 \le r \le n-1).$$

当 $n \ge 2$ 时，此式对 $r = 1$ 亦成立，因为 $t_1 = p$ 且 $s_1 = 0$。把 $t_{r-1} = p(1 - s_{r-1})$（在 $1 \le r-1 \le n-1$ 时成立）代入 s_r 的方程：

$$\begin{aligned} s_r &= 1 - s_{r-1} - p(1 - s_{r-1}) \\ &= (1-p)(1 - s_{r-1}) \\ &= q\,(1 - s_{r-1}), \end{aligned}$$

此式对 $2 \le r \le n$ 成立，特别地对 $2 \le r \le n-1$ 成立，正如所求。

解递推关系

> **想一想** 怎样解 $s_r = q - qs_{r-1}$？哪个常数数列满足它？

常数解 $s_r = c$ 需要 $c = q - qc$，所以 $c = \dfrac{q}{1+q}$。从递推关系减去 $c = q - qc$，

$$s_r - \frac{q}{1+q} = -q\left(s_{r-1} - \frac{q}{1+q}\right),$$

所以 $s_r - \frac{q}{1+q}$ 每一步都乘以 $-q$。由 $s_1 - \frac{q}{1+q} = -\frac{q}{1+q}$ 开始，

$$s_r - \frac{q}{1+q} = (-q)^{r-1}\left(-\frac{q}{1+q}\right) = \frac{(-q)^r}{1+q},$$

从而得到

$$s_r = \frac{q + (-q)^r}{1+q} \qquad (1 \le r \le n-1).$$

（由于递推关系一直成立至 $r = n$，同一公式其实对 $r = n$ 亦成立。）

至于 t_r，我们利用 $t_r = p(1 - s_r)$ 及

$$1 - s_r = \frac{1 + q - q - (-q)^r}{1+q} = \frac{1 - (-q)^r}{1+q},$$

所以

$$t_r = \frac{p\big(1 - (-q)^r\big)}{1+q} \qquad (1 \le r \le n-1).$$

最后一片

如上所述，s_r 的方程及递推关系 $s_r = q(1 - s_{r-1})$ 对 $r = n$ 亦成立，所以公式延续下去：

$$s_n = \frac{q + (-q)^n}{1+q}.$$

最后一片不可能是三文治的第一片，因为之后再没有面包。所以它要么是多士，要么是三文治的第二片，故 $t_n = 1 - s_n$：

$$t_n = \frac{1 - (-q)^n}{1+q}.$$

与 $r < n$ 时的 t_r 相比，因子 p 消失了：一份小食若从最后一片开始，必定是多士。（这些公式在 $n = 1$ 时亦成立：它们给出 $s_1 = 0$ 及 $t_1 = 1$。）

验算

- 取 $n = 3$ 及 $p = q = \frac{1}{2}$。可能的次序有：多士—多士—多士（概率 $\frac{1}{4}$）、先多士后三文治（概率 $\frac{1}{4}$）、先三文治后多士（概率 $\frac{1}{2}$）。所以 $s_3 = \frac{1}{4}$ 而 $t_3 = \frac{3}{4}$。公式给出 $s_3 = \dfrac{\frac{1}{2} - \frac{1}{8}}{\frac{3}{2}} = \frac{1}{4}$ 及 $t_3 = \dfrac{1 + \frac{1}{8}}{\frac{3}{2}} = \frac{3}{4}$。
- 当 $q = 0$（永远做多士）时，对每个 r 都得 $s_r = 0$ 及 $t_r = 1$，理应如此。
- 当 r 增大时，$s_r \to \frac{q}{1+q}$。这是合理的：长远而言，有比例 q 的小食是三文治，每份有一片“第二片”，而每份小食平均用 $1 + q$ 片面包，所以有比例 $\frac{q}{1+q}$ 的面包是第二片。
- 我们用电脑对 $n \le 12$ 及若干 p 值，精确列举所有小食序列，验证了全部公式。

常见错误

- 使用 $t_n = p(s_{n-1} + t_{n-1})$：当一份小食从最后一片开始时，它是多士的概率是 1，而不是 p。
- 以为事件“第 $r-1$ 片是多士”与“第 r 片是多士”互相独立（independent）。其实并不独立：关键在于第 r 片是否开始一份小食，而这取决于之前的历史。
- 解递推关系时丢失负号：公比是 $-q$，所以答案围绕 $\frac{q}{1+q}$ 交替摆动。

STEP 之外

第 r 片是否开始一份新小食，构成一个简单的*更新过程*（renewal process）：各份小食的长度互相独立，每份是 1 或 2，而极限 $\frac{q}{1+q}$ 正是更新定理（renewal theorem）的一个例子（长远的更新率等于平均间隔的倒数）。等价地，每片面包的类型构成一条马可夫链（Markov chain），而 $(-q)^r$ 显示它遗忘起始状态的速度。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）同时涵盖马可夫链及这类递推关系。

详细解答

STEP II 2017，第 1 题

纯数学 · 归约公式

STEP II 2017，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/17-s2-q1

关键思路

被积函数（integrand）中难处理的部分是 $\arctan x$，但它的导数（derivative）是一个有理函数（rational function），所以用分部积分法（integration by parts），取 $u = \arctan x$，就能把它换成 $\frac{1}{1+x^2}$。接着，(ii) 的窍门是把两个积分配对，使分子合并成 $x^{n+1}(1+x^2)$，从而消去麻烦的分母。这样便得到一条令 n 每次增加 2 的递推公式（reduction formula）；用两次就令 n 增加 4，而这正是 (iii) 的归纳法（induction）所需要的。

解答

第 (i) 部分

用分部积分法，取 $u = \arctan x$ 及 $\mathrm{d}v = x^n\,\mathrm{d}x$，于是 $v = \frac{x^{n+1}}{n+1}$（由于 $n \geq 0$，这没有问题）：

$$\begin{aligned} I_n &= \left[\frac{x^{n+1}}{n+1}\arctan x\right]_0^1 - \int_0^1 \frac{x^{n+1}}{n+1}\cdot\frac{1}{1+x^2}\,\mathrm{d}x \\ &= \frac{1}{n+1}\cdot\frac{\pi}{4} - \frac{1}{n+1}\int_0^1 \frac{x^{n+1}}{1+x^2}\,\mathrm{d}x, \end{aligned}$$

这里用了 $\arctan 1 = \frac{\pi}{4}$，以及边界项（boundary term）在 $x = 0$ 处为零这一事实。两边乘以 $n+1$，便得到所需结果：

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

当 $n = 0$ 时，余下的积分是 $\int_0^1 \frac{x}{1+x^2}\,\mathrm{d}x = \left[\frac{1}{2}\ln(1+x^2)\right]_0^1 = \frac{1}{2}\ln 2$，所以

$$I_0 = \frac{\pi}{4} - \frac{1}{2}\ln 2.$$

第 (ii) 部分

把 (i) 中的 n 换成 $n+2$，再与原本的 (i) 相加：

$$\begin{aligned}(n+3)I_{n+2}+(n+1)I_n &= \frac{\pi}{2}-\int_0^1 \frac{x^{n+3}+x^{n+1}}{1+x^2}\,\mathrm{d}x\\ &= \frac{\pi}{2}-\int_0^1 x^{n+1}\,\mathrm{d}x\\ &= \frac{\pi}{2}-\frac{1}{n+2},\end{aligned}$$

因为 $x^{n+3}+x^{n+1}=x^{n+1}(1+x^2)$。

> **想一想** 这条公式把 I_4 与 I_2 联系起来，又把 I_2 与 I_0 联系起来。我们需要哪些 n 值?

我们需要 $n=0$ 及 $n=2$：

$$\begin{aligned}n=0:&\quad 3I_2+I_0=\frac{\pi}{2}-\frac{1}{2},\\ n=2:&\quad 5I_4+3I_2=\frac{\pi}{2}-\frac{1}{4}.\end{aligned}$$

以第二式减去第一式，消去 I_2：

$$5I_4=I_0+\frac{1}{4}=\frac{\pi}{4}+\frac{1}{4}-\frac{1}{2}\ln 2,\qquad \text{所以}\qquad I_4=\frac{\pi+1-2\ln 2}{20}.$$

第 (iii) 部分

为简便起见，记 $S_n=\sum_{r=1}^{2n}\frac{(-1)^r}{r}$。

求 A。当 $n=1$ 时，$S_1=-1+\frac{1}{2}=-\frac{1}{2}$，所以命题变成 $5I_4=A+\frac{1}{4}$。由 (ii) 得 $5I_4=\frac{\pi}{4}+\frac{1}{4}-\frac{1}{2}\ln 2$，因此必须取

$$A=\frac{\pi}{4}-\frac{1}{2}\ln 2\ (=I_0),$$

而取这个值时，命题对 $n=1$ 成立。

归纳步骤。假设对某个 $k\geq 1$，有 $(4k+1)I_{4k}=A-\frac{1}{2}S_k$。我们要推到 I_{4k+4}，即用 (ii) 的公式走两步。分别取 $n=4k+2$ 及 $n=4k$，得

$$\begin{aligned}(4k+5)I_{4k+4}+(4k+3)I_{4k+2}&=\frac{\pi}{2}-\frac{1}{4k+4},\\ (4k+3)I_{4k+2}+(4k+1)I_{4k}&=\frac{\pi}{2}-\frac{1}{4k+2}.\end{aligned}$$

> **想一想** 怎样才能消去多余的 I_{4k+2}?

它在两条方程中的系数相同，所以以第一式减去第二式便可把它消去：

$$(4k+5)I_{4k+4} = (4k+1)I_{4k} + \frac{1}{4k+2} - \frac{1}{4k+4}.$$

另一方面，S_{k+1} 比 S_k 多出两项，即 $r = 2k+1$（奇数，所以是负的）和 $r = 2k+2$（偶数，所以是正的）：

$$-\frac{1}{2}S_{k+1} = -\frac{1}{2}S_k - \frac{1}{2}\Big(-\frac{1}{2k+1} + \frac{1}{2k+2}\Big) = -\frac{1}{2}S_k + \frac{1}{4k+2} - \frac{1}{4k+4}.$$

利用归纳假设（inductive hypothesis），

$$(4k+5)I_{4k+4} = A - \frac{1}{2}S_k + \frac{1}{4k+2} - \frac{1}{4k+4} = A - \frac{1}{2}S_{k+1},$$

这正是 $n = k+1$ 时的命题。由于命题对 $n = 1$ 成立，由归纳法可知它对所有 $n \geq 1$ 均成立，其中 $A = \frac{\pi}{4} - \frac{1}{2}\ln 2$。

验算

- 数值上，$I_0 \approx 0.438825$，$I_4 \approx 0.137765$；对原来的积分使用辛普森法则（Simpson's rule），结果在小数点后十位都吻合。
- A 等于 I_0，所以 (iii) 的公式对 $n = 0$ 也成立（此时和式是空和），这是个好兆头。
- 当 $n \to \infty$ 时，(i) 中的积分趋向 0（被积函数不超过 x^{n+1}），所以 $(4n+1)I_{4n} \to \frac{\pi}{4}$。同时 $S_n \to -\ln 2$，所以右方趋向 $\frac{\pi}{4} - \frac{1}{2}\ln 2 + \frac{1}{2}\ln 2 = \frac{\pi}{4}$。两边在极限下吻合。

常见错误

- 写了 $v = \frac{x^{n+1}}{n+1}$，却忘了把第二个积分也除以 $n+1$。
- 弄错交错和中两个新项的正负号：$r = 2k+1$ 是奇数，所以 $(-1)^r = -1$。
- 未先由基本情况（base case）确定 A，就去证明归纳步骤。

STEP 之外

最后一项验算用到了交错调和级数（alternating harmonic series）$1 - \frac{1}{2} + \frac{1}{3} - \cdots = \ln 2$，而这道题实际上就证明了它：(iii) 的恒等式，加上 $(4n+1)I_{4n} \to \frac{\pi}{4}$ 这一事实，迫使部分和（partial sums）收敛（converge）到 $\ln 2$。利用像 $\int_0^1 \frac{x^{n+1}}{1+x^2}\,\mathrm{d}x$ 这类趋向零的积分来控制级数的误差，是一种标准方法，你在第一个数学分析（analysis）课程中便会遇到。Abbott 的 *Understanding Analysis*（《理解数学分析》）浅白易读，是了解如何严格处理这类极限的好起点。

详细解答

STEP II 2017，第 2 题

纯数学 · 周期数列

STEP II 2017，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/17-s2-q2

关键思路

把映射与自身复合并化简：递推关系（recurrence relation）走两步，得到的是另一个形状完全相同的“一次式除以一次式”映射，所以走四步其实就是把“走两步”再做两次。每当我们问 $x_{n+k} = x_n$ 是否成立时，消去分母都会得到一条关于 x_n 的二次方程，而这条二次式可以因式分解（factorise）。其中一个因式总是 $x^2 + (b-a)x + 1$，它的根正好就是递推关系的*不动点*（fixed points），即令数列成为常数数列的那些值。看出这一点，两部分便都迎刃而解。

解答

记 $f(x) = \dfrac{ax-1}{x+b}$，使得 $x_{n+1} = f(x_n)$。于是

$$x_{n+2} = \frac{a\dfrac{ax_n-1}{x_n+b} - 1}{\dfrac{ax_n-1}{x_n+b} + b}.$$

分子分母同乘以 $x_n + b$（按假设不为零），得

$$x_{n+2} = \frac{a(ax_n-1) - (x_n+b)}{(ax_n-1) + b(x_n+b)} = \boxed{\frac{(a^2-1)x_n - (a+b)}{(a+b)x_n + b^2 - 1}}$$

现在值得多记下一个事实。一个值 x 是不动点（$f(x) = x$）当且仅当 $ax - 1 = x^2 + bx$，即

$$q(x) \equiv x^2 + (b-a)x + 1 = 0. \qquad (*)$$

若 x_0 是 q 的根，则 $x_1 = x_0$，如此类推：数列是常数数列，周期（period）为 1。

第 (i) 部分

假设数列的周期为 2。那么特别地 $x_2 = x_0$。在 x_{n+2} 的公式中代入 $x = x_0$ 并消去分母，

$$\begin{aligned} 0 &= (a^2-1)x - (a+b) - x\big[(a+b)x + b^2 - 1\big] \\ &= -(a+b)x^2 + (a^2-b^2)x - (a+b) \\ &= -(a+b)\big[x^2 + (b-a)x + 1\big], \end{aligned}$$

这里用了 $a^2 - b^2 = (a+b)(a-b)$。所以 $(a+b)\,q(x_0) = 0$。

> **想一想** 为甚么不能以“$q(x_0) = 0$ 只对应 x_0 的一个值”作结?

因为 x_0 只是一个数，没有任何东西阻止它是 q 的根。真正排除这种情况的是 $(*)$ 的含义：若 $q(x_0) = 0$，则 x_0 是不动点，所以 $x_1 = x_0$，并且对所有 n 都有 $x_{n+1} = x_n$。这是周期 1，而周期 2 的定义不容许这种情况（取 $m = 1$）。因此 $q(x_0) \neq 0$，从而

$$a + b = 0$$

是周期 2 的必要条件。

（反过来，若 $b = -a$，两步公式便简化为 $x_{n+2} = \dfrac{(a^2-1)x_n}{a^2-1} = x_n$。这里不可能出现 $a^2 = 1$ 的情况，因为当 $a = 1$、$b = -1$ 时 $x_1 = 1$，当 $a = -1$、$b = 1$ 时 $x_1 = -1$；无论哪种情况都有 $x_1 + b = 0$，而这是被排除的。）

第 (ii) 部分

记 $A = a^2 - 1$、$B = b^2 - 1$ 及 $c = a + b$，使得 $x_{n+2} = g(x_n)$，其中 $g(x) = \dfrac{Ax - c}{cx + B}$。把 g 作用两次，并再次将分子分母同乘以 $cx + B$，

$$x_{n+4} = \frac{A(Ax_n - c) - c(cx_n + B)}{c(Ax_n - c) + B(cx_n + B)} = \frac{(A^2 - c^2)x_n - c(A+B)}{c(A+B)x_n + B^2 - c^2}.$$

那么 $x_{n+4} = x_n$ 成立当且仅当

$$\begin{aligned} 0 &= (A^2-c^2)x_n - c(A+B) - x_n\big[c(A+B)x_n + B^2 - c^2\big] \\ &= -c(A+B)x_n^2 + (A^2 - B^2)x_n - c(A+B) \\ &= -(A+B)\big[c\,x_n^2 + (B-A)x_n + c\big]. \end{aligned}$$

由于 $B - A = b^2 - a^2 = (b-a)(a+b) = (b-a)c$，方括号内的式子等于 $c\,q(x_n)$，所以

$$x_{n+4} = x_n \iff (a^2 + b^2 - 2)\,(a+b)\,q(x_n) = 0.$$

必要性。假设周期为 4。数列不是常数数列，所以 x_0 不是不动点，$q(x_0) \neq 0$。若 $a + b = 0$，则由第 (i) 部分的逆命题得 $x_2 = x_0$，而且实际上对所有 n 都有 $x_{n+2} = x_n$，与周期 4 矛

盾。所以 $a+b\neq 0$，于是取 $n=0$ 时的因式分解迫使

$$a^2+b^2=2.$$

充分性。假设 $a^2+b^2=2$。那么 $A+B=0$，所以对每个 n 都有 $x_{n+4}=x_n$。我们必须验证没有更小的 m 能成立。

- $a+b\neq 0$：否则 $a^2=b^2=1$，并如第 (i) 部分所述，$x_1+b=0$，而这是被排除的。
- 没有实数不动点：q 的判别式（discriminant）是 $(b-a)^2-4=(a^2+b^2)-2ab-4=-2(1+ab)$，而 $|ab|\leq\frac{1}{2}(a^2+b^2)=1$，且只有当 $(a,b)=\pm(1,-1)$ 时才有 $ab=-1$，这情况我们刚刚已经排除。所以 $ab>-1$，判别式为负，对每个实数 x_n 都有 $q(x_n)\neq 0$。因此 $m=1$ 不成立。
- $m=2$：由第 (i) 部分，$x_2=x_0$ 需要 $(a+b)q(x_0)=0$，但这是错的。所以 $m=2$ 不成立。
- $m=3$：若对所有 n 都有 $x_{n+3}=x_n$，则对所有 n 都有 $x_{n+1}=x_{(n+1)+3}=x_{n+4}=x_n$，得出周期 1，而这我们刚刚已经排除。

因此，对实数 a、b 及 x_0（分母均不为零，如题目所容许），数列的周期为 4 当且仅当

$$\boxed{a^2+b^2=2.}$$

起始值 x_0 在这个条件中不起任何作用。

验算

取 $a=\sqrt{2}$、$b=0$ 及 $x_0=1$。数列依次为

$$1,\quad \sqrt{2}-1,\quad -1,\quad \sqrt{2}+1,\quad 1,\ \ldots$$

其周期恰好为 4。例如 $f(-1)=(-\sqrt{2}-1)/(-1)=\sqrt{2}+1$。电脑验算以数以千计的随机有理数 a、b、x_0 确认了两步及四步公式完全正确。它亦对圆 $a^2+b^2=2$ 上约二千个随机点确认了周期为 4，而对不在该圆上的随机 (a,b)，则发现四步之后都没有回到起点。

常见错误

- 由 $(a+b)q(x_0)=0$ 推出 $a+b=0$，却没有解释为甚么 $q(x_0)\neq 0$。不动点的诠释正是个中原因。
- 给出 $a^2+b^2=2$ 作为周期 4 的答案，却没有验证周期其实不是 1 或 2。
- 在第 (ii) 部分的方括号中遗漏了因式 $a+b$，以致周期 2 的情况 $a+b=0$ 被错误地看成给出周期 4。

STEP 之外

形如 $x \mapsto \dfrac{\alpha x+\beta}{\gamma x+\delta}$ 的映射是一个莫比乌斯变换（Möbius transformation）。复合两个这样的映射，相当于把它们对应的 2×2 矩阵相乘，在这里该矩阵为 $\begin{pmatrix} a & -1 \\ 1 & b \end{pmatrix}$。映射的周期为 k 当且仅当矩阵的 k 次幂是单位矩阵（identity matrix）的倍数，而这由比值 $(\text{迹})^2/\det = (a+b)^2/(ab+1)$ 决定。周期 4 对应于值 2，其实是一个乔装了的“旋转 $\pi/4$”。Needham 的 *Visual Complex Analysis*（《直观复分析》）有精彩的一章讲述莫比乌斯变换，而 Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）则涵盖了其背后的特征值（eigenvalue）概念。

详细解答

STEP II 2017，第 3 题

纯数学 · 曲线描绘

STEP II 2017，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/17-s2-q3

关键思路

方程 $\sin y = k$ 从来不会只有一个解：若 y_0 是一个解，则 $\pi - y_0$ 也是，而与这两者相差 2π 的任何值亦是。因此这里每一幅草图，都是同一条基本曲线的几个复本，按正弦函数的对称性（symmetry）摆放而成。第 (i) 部分用直线展示这个规律；在 (ii) 中，微积分确定了基本部分 $y = \arcsin(\frac{1}{2}\sin x)$ 的形状，其余的交由对称性完成。第 (iii) 部分则是把 $\cos y$ 写成正弦之后，再做一次 (ii)。

解答

第 (i) 部分

$\sin y = \sin x$ 成立，当且仅当

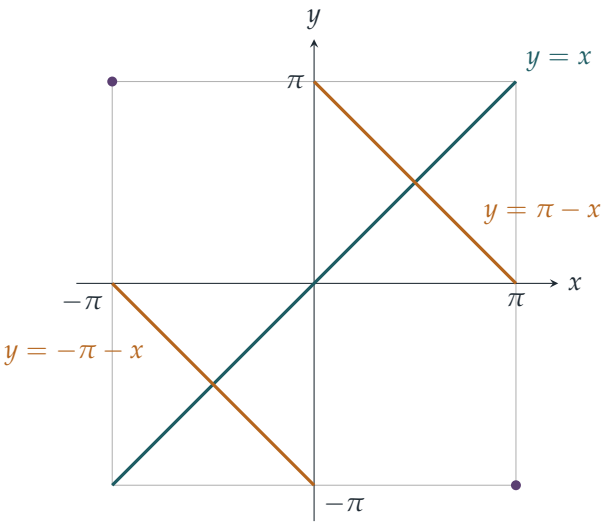
$$y = x + 2k\pi \qquad 或 \qquad y = \pi - x + 2k\pi \qquad (k \in \mathbb{Z}).$$

现在看看这些直线中哪些与正方形 $-\pi \le x, y \le \pi$ 相交。

- $y = x$ 贯穿整条对角线，由 $(-\pi, -\pi)$ 至 (π, π)。
- $y = x \pm 2\pi$ 要求 $y - x = \pm 2\pi$。在正方形内，$|y - x| \le 2\pi$，且等号只在角点成立，所以这两条直线只给出两个孤立点 $(-\pi, \pi)$ 和 $(\pi, -\pi)$。
- $x + y = \pi$ 是由 $(0, \pi)$ 至 $(\pi, 0)$ 的线段，而 $x + y = -\pi$（即 $k = -1$ 的情况）是由 $(-\pi, 0)$ 至 $(0, -\pi)$ 的线段。
- $x + y = 3\pi$ 或 -3π 都不可能，因为 $|x + y| \le 2\pi$。

所以这个集合由对角线 $y = x$、$x + y = \pm\pi$ 的两条线段，以及两个角点 $(-\pi, \pi)$ 和

$(\pi, -\pi)$ 组成。



孤立的角点很容易被遗漏。它们确实是解，因为 $\sin(\pm\pi) = 0$。

第 (ii) 部分

对 $\sin y = \frac{1}{2}\sin x$ 进行隐函数微分（implicit differentiation），得 $y' \cos y = \frac{1}{2}\cos x$。当 $0 \le y \le \frac{1}{2}\pi$ 时，$\cos y \ge 0$，所以

$$\cos y = \sqrt{1 - \sin^2 y} = \sqrt{1 - \tfrac{1}{4}\sin^2 x} = \tfrac{1}{2}\sqrt{4 - \sin^2 x},$$

它永远不为零。因此

$$y' = \frac{\cos x}{\sqrt{4 - \sin^2 x}}.$$

设 $s = \sin x$。利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\sqrt{4 - s^2} = \dfrac{-s\cos x}{\sqrt{4 - s^2}}$，由商法则（quotient rule）得

$$y'' = \frac{-\sin x\sqrt{4 - s^2} - \cos x \cdot \dfrac{-s\cos x}{\sqrt{4 - s^2}}}{4 - s^2} = \frac{-s(4 - s^2) + s\cos^2 x}{(4 - s^2)^{3/2}}$$

$$= \frac{s\left(-4 + s^2 + 1 - s^2\right)}{(4 - s^2)^{3/2}} = -\frac{3\sin x}{(4 - \sin^2 x)^{3/2}},$$

其中用了 $\cos^2 x = 1 - s^2$。这就是所求的结果。

基本部分。在这个四分之一正方形内，曲线是 $y = f(x)$，其中 $f(x) = \arcsin(\frac{1}{2}\sin x)$。它由 $(0, 0)$ 出发，终于 $(\frac{1}{2}\pi, \frac{1}{6}\pi)$，因为 $\arcsin\frac{1}{2} = \frac{1}{6}\pi$，所以它从不接近上边 $y = \frac{1}{2}\pi$。它在原点的斜率是 $\frac{1}{2}$，到 $x = \frac{1}{2}\pi$ 时降至 0。此外，当 $0 < x \le \frac{1}{2}\pi$ 时 $y'' < 0$，而在 $x = 0$ 时 $y'' = 0$，所以曲线是上凸（concave）且递增的，并在原点有一个拐点（point of inflection）。

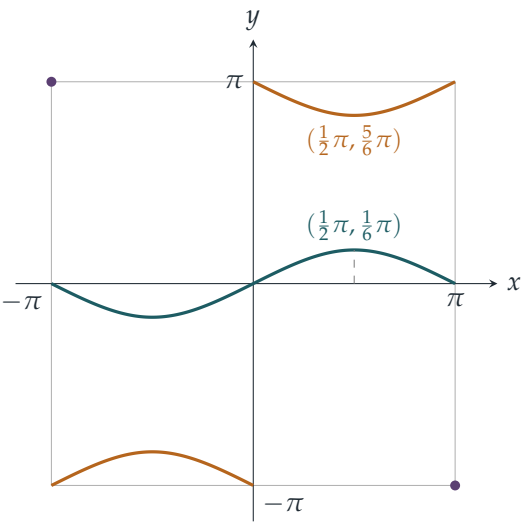
想一想 不再做任何微积分，怎样由这一小段弧得出整个正方形内的图像?

我们利用对称性。由于 $\sin(\pi - x) = \sin x$，而且 f 是奇函数 (odd function)，$y = f(x)$ 在 $[-\pi, \pi]$ 上的图像关于 $x = \frac{1}{2}\pi$ 对称，并且关于原点有旋转对称 (rotational symmetry)。它是一条平缓的波，经过 $(-\pi, 0)$、$(0, 0)$ 和 $(\pi, 0)$，最大点为 $(\frac{1}{2}\pi, \frac{1}{6}\pi)$，最小点为 $(-\frac{1}{2}\pi, -\frac{1}{6}\pi)$。它在原点的斜率是 $\frac{1}{2}$，在 $x = \pm\pi$ 的斜率是 $-\frac{1}{2}$。

与 (i) 一样，完整的解集由 $y = f(x) + 2k\pi$ 或 $y = \pi - f(x) + 2k\pi$ 给出。在正方形内剩下：

- 对所有 $-\pi \leq x \leq \pi$ 的波 $y = f(x)$；
- $y = \pi - f(x)$，它只在 $f(x) \geq 0$ 之处，即 $0 \leq x \leq \pi$ 时，位于正方形内。这是一条由 $(0, \pi)$ 垂到 (π, π) 的拱形，最低点为 $(\frac{1}{2}\pi, \frac{5}{6}\pi)$；
- $y = -\pi - f(x)$，它只在 $f(x) \leq 0$ 之处，即 $-\pi \leq x \leq 0$ 时，位于正方形内。这是一条由 $(-\pi, -\pi)$ 拱起至 $(0, -\pi)$ 的拱形，最高点为 $(-\frac{1}{2}\pi, -\frac{5}{6}\pi)$；
- 孤立的角点 $(-\pi, \pi)$ 和 $(\pi, -\pi)$，分别来自 $x = -\pi$ 时的 $y = \pi - f(x)$ 和 $x = \pi$ 时的 $y = -\pi - f(x)$。

k 的其他值都给出 $|y| > \pi$。拱形以斜率 $\pm\frac{1}{2}$ 与直线 $y = \pm\pi$ 相交，它们是那条波旋转半周后再向上或向下平移所得。与 (i) 比较：对角线被压平成波，两条线段则弯成了拱形。



第 (iii) 部分

想一想 怎样把 $\cos y$ 变成我们已经画过的东西?

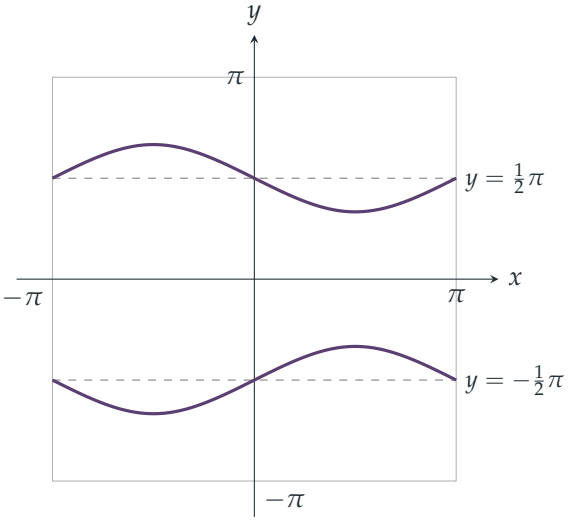
写 $\cos y = \sin(\frac{1}{2}\pi - y)$。于是 $\cos y = \frac{1}{2}\sin x$ 表示 $u = \frac{1}{2}\pi - y$ 满足 $\sin u = \frac{1}{2}\sin x$，所以 $u = f(x) + 2k\pi$ 或 $u = \pi - f(x) + 2k\pi$。即是

$$y = \tfrac{1}{2}\pi - f(x) - 2k\pi \qquad \text{或} \qquad y = f(x) - \tfrac{1}{2}\pi - 2k\pi.$$

由于 $|f(x)| \le \frac{1}{6}\pi$，只有 $k = 0$ 给出在 $[-\pi, \pi]$ 内的值，而且对每一个 x 都如此。所以这个集合恰好由两条完整的波组成:

- $y = f(x) - \frac{1}{2}\pi$: 把 (ii) 的波向下平移 $\frac{1}{2}\pi$，在 $-\frac{2}{3}\pi$ 与 $-\frac{1}{3}\pi$ 之间振荡;
- $y = \frac{1}{2}\pi - f(x)$: 它在 x 轴中的镜像（这是必然的，因为 $\cos y$ 是 y 的偶函数）。它的最小点为 $(\frac{1}{2}\pi, \frac{1}{3}\pi)$，最大点为 $(-\frac{1}{2}\pi, \frac{2}{3}\pi)$。

这次没有孤立点。



验算

- 取样点: $(\frac{1}{2}\pi, \frac{1}{6}\pi)$ 给出 $\sin\frac{1}{6}\pi = \frac{1}{2} = \frac{1}{2}\sin\frac{1}{2}\pi$，而 $(\frac{1}{2}\pi, \frac{5}{6}\pi)$ 亦成立。在 (iii) 中，$(\frac{1}{2}\pi, \frac{1}{3}\pi)$ 给出 $\cos\frac{1}{3}\pi = \frac{1}{2}$。
- 在 $x = 0$，公式给出 $y' = \frac{1}{2}$，与小角近似（small-angle approximation）$y \approx \frac{1}{2}x$ 吻合。在 $x = \frac{1}{2}\pi$，公式给出 $y' = 0$，在最大点理应如此。

- 我们用电脑以有限差分（finite differences）在 $(0, \frac{1}{2}\pi)$ 内的随机点验证了 y' 和 y'' 的公式。电脑亦对许多随机的 x 求出所有解 $y \in [-\pi, \pi]$，结果与上述各分支（branches）相符：在 (ii) 中当 $x \neq 0, \pm\pi$ 时有两个解，在 (iii) 中则总是有两个解。

常见错误

- 在 (i) 和 (ii) 中遗漏角点 $(-\pi, \pi)$ 和 $(\pi, -\pi)$。
- 在 (ii) 中把拱形画满整个 $[-\pi, \pi]$：一旦 $f(x) < 0$，$y = \pi - f(x)$ 便离开正方形。
- 取 $\cos y = -\frac{1}{2}\sqrt{4 - \sin^2 x}$：符号已由 $0 \leq y \leq \frac{1}{2}\pi$ 确定。

STEP 之外

形如 $\{(x, y) : F(x, y) = 0\}$ 的集合称为水平集（level set）。隐函数定理（implicit function theorem）指出，凡 $\partial F/\partial y \neq 0$ 之处，它在局部都是一个光滑的函数图像。在这里，这条件恰好在 $\cos y = 0$ 之处失效；而由于当 $|\sin y| \leq \frac{1}{2}$ 时这永远不会发生，(ii) 中的曲线都是关于 x 的光滑图像。在 (i) 中，交叉点正是 $\sin y - \sin x$ 的两个偏导数（partial derivatives）同时为零之处，而水平集亦正是在这些点上不再是一条曲线。Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）建立了单变量的基础；Abbott 的 *Understanding Analysis*（《理解分析》）则是一本生动的辅助读物。

详细解答

STEP II 2017，第 4 题

纯数学 · 柯西–施瓦兹积分不等式

STEP II 2017，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/17-s2-q4

关键思路

施瓦茨不等式（Schwarz inequality）以平方的积分 $\int f^2$ 和 $\int g^2$ 为乘积的积分 $\int fg$ 提供一个上界（upper bound）。因此，整个技巧在于选取 f 和 g，使得每一个积分都能算出来，唯一可能例外的就是我们想估计的那一个。如果想估计的积分是 $\int fg$，我们得到的是上界；如果它是 $\int f^2$ 或 $\int g^2$，两边相除后得到的便是下界（lower bound）。

解答

以下一律把 $(*)$ 写成 $\left(\int fg\right)^2 \le \int f^2 \cdot \int g^2$。

第 (i) 部分

目标式中含有 e^t，所以取 $f(x)=1$、$g(x)=\mathrm{e}^x$、$a=0$ 及 $b=t$。于是

$$\int_0^t \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^t - 1, \qquad \int_0^t 1\,\mathrm{d}x = t, \qquad \int_0^t \mathrm{e}^{2x}\,\mathrm{d}x = \tfrac{1}{2}\left(\mathrm{e}^{2t}-1\right),$$

因此 $(*)$ 给出

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

由于 $t>0$，有 $\mathrm{e}^t-1>0$，而且 $\mathrm{e}^t+1>0$。除以正数 $(\mathrm{e}^t-1)(\mathrm{e}^t+1)$ 不会改变不等号的方向：

$$\frac{\mathrm{e}^t-1}{\mathrm{e}^t+1} \le \frac{t}{2},$$

得证。

第 (ii) 部分

我们希望 $\int_0^1 \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x$ 以 $\int g^2$ 的形式出现，所以在 $[0,1]$ 上取 $g(x)=\mathrm{e}^{-\frac{1}{4}x^2}$ 及 $f(x)=x$。乘积 $x\,\mathrm{e}^{-\frac{1}{4}x^2}$ 正是我们能够积分的那一类（指数外面的 x 幂，在相差一个常数的意义下，正是指数的导数）：

$$\int_0^1 x\,\mathrm{e}^{-\frac{1}{4}x^2}\,\mathrm{d}x=\left[-2\,\mathrm{e}^{-\frac{1}{4}x^2}\right]_0^1=2\left(1-\mathrm{e}^{-\frac{1}{4}}\right),\qquad \int_0^1 x^2\,\mathrm{d}x=\frac{1}{3}.$$

因此 $(*)$ 给出

$$4\left(1-\mathrm{e}^{-\frac{1}{4}}\right)^2\le\frac{1}{3}\int_0^1 \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x,$$

两边乘以 3，得

$$\int_0^1 \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x\ge 12\left(1-\mathrm{e}^{-\frac{1}{4}}\right)^2.$$

第 (iii) 部分

记 $J=\int_0^{\frac{1}{2}\pi}\sqrt{\sin x}\,\mathrm{d}x$。这里所有积分都在 $[0,\frac{1}{2}\pi]$ 上进行，该区间内 $\sin x\ge 0$ 且 $\cos x\ge 0$，所以所有根式都是实数。

上界。这里 J 应当是 $\int fg$。取 $f(x)=1$ 及 $g(x)=\sqrt{\sin x}$，使得 $g^2=\sin x$ 容易积分：

$$J^2\le\int_0^{\frac{1}{2}\pi}1\,\mathrm{d}x\cdot\int_0^{\frac{1}{2}\pi}\sin x\,\mathrm{d}x=\frac{\pi}{2}\cdot 1.$$

由于 $J>0$，取正平方根得 $J\le\sqrt{\pi/2}$。

下界。现在 J 必须是其中一个平方的积分，所以取 $f(x)=(\sin x)^{1/4}$，使得 $f^2=\sqrt{\sin x}$。

> ? **想一想** 哪一个 g 能使 $\int fg$ 和 $\int g^2$ 都可以算出来?

我们需要 g^2 容易积分，同时 $(\sin x)^{1/4}g(x)$ 也容易积分。选取 $g(x)=\cos x$ 两者兼顾：$(\sin x)^{1/4}\cos x$ 是 $\sin x$ 的幂乘以它的导数。答案的形式也是一条线索：分母中的 π 暗示 $\int g^2=\int_0^{\frac{1}{2}\pi}\cos^2 x\,\mathrm{d}x=\frac{\pi}{4}$。事实上

$$\int_0^{\frac{1}{2}\pi}(\sin x)^{1/4}\cos x\,\mathrm{d}x=\left[\tfrac{4}{5}(\sin x)^{5/4}\right]_0^{\frac{1}{2}\pi}=\frac{4}{5}$$

以及

$$\int_0^{\frac{1}{2}\pi}\cos^2 x\,\mathrm{d}x=\int_0^{\frac{1}{2}\pi}\tfrac{1}{2}(1+\cos 2x)\,\mathrm{d}x=\frac{\pi}{4}.$$

因此 $(*)$ 给出

$$\frac{16}{25}\le J\cdot\frac{\pi}{4},\qquad \text{即}\qquad J\ge\frac{64}{25\pi}.$$

把两部分合起来，

$$\frac{64}{25\pi} \le \int_0^{\frac{1}{2}\pi} \sqrt{\sin x}\,\mathrm{d}x \le \sqrt{\frac{\pi}{2}}.$$

附注。其他选取方法也行得通，有些效果更好。取 $f = (\sin x)^{1/4}$、$g = (\sin x)^{3/4}$，得 $1 \le J\int_0^{\frac{1}{2}\pi}(\sin x)^{3/2}\,\mathrm{d}x$；再以 $f = \sin x$、$g = \sqrt{\sin x}$ 第二次应用 $(*)$，可证 $\int_0^{\frac{1}{2}\pi}(\sin x)^{3/2}\,\mathrm{d}x \le \sqrt{\pi/4}$。两者结合，得到更精确的下界 $J \ge 2/\sqrt{\pi} \approx 1.128$。

验算

- (i)：当 $t \to 0$ 时，左边是 $\tanh(t/2) \approx t/2$，所以对于细小的 t，这个界是紧的。当 t 很大时，左边趋于 1，而 $t/2$ 则无限增大。
- (ii)：$\int_0^1 \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x \approx 0.8556$，而 $12(1-\mathrm{e}^{-1/4})^2 \approx 0.5871$。
- (iii)：数值上 $J \approx 1.1981$，介乎 $64/(25\pi) \approx 0.8149$ 与 $\sqrt{\pi/2} \approx 1.2533$ 之间。上界相当接近，因为 $\sqrt{\sin x}$ 在区间的大部分范围内都颇为平坦（接近 1），而取 $f = 1$ 时，施瓦茨不等式取等号当且仅当 g 是常数。

常见错误

- 弄错界的方向。若未知积分是 $\int fg$，施瓦茨不等式只能给出它的上界，所以要得到下界，未知积分必须是 $\int f^2$ 或 $\int g^2$。
- 在 (i) 中除以 $\mathrm{e}^t - 1$ 时没有说明它是正数。
- 在 (iii) 的上界中，忘记利用 $J > 0$ 取正平方根。

STEP 之外

施瓦茨（柯西－施瓦茨，Cauchy–Schwarz）不等式指出，内积（inner product）$\langle f, g\rangle = \int fg$ 满足 $|\langle f, g\rangle| \le \|f\|\,\|g\|$，且等号成立当且仅当 f 与 g 成比例。它在每一个内积空间（inner product space）中都成立，由 $\mathbb{R}^n$ 中的向量到函数空间皆然；正是它使“两个函数之间的夹角”这个概念有意义。这个想法是傅里叶级数（Fourier series）和最小二乘逼近（least-squares approximation）的基础。Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）阐述了内积与正交性（orthogonality），而 Abbott 的 *Understanding Analysis*（《理解分析》）则展示了如何运用这类不等式来控制积分。

详细解答

STEP II 2017，第 5 题

纯数学 · 抛物线的法线与圆

STEP II 2017，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/17-s2-q5

关键思路

一切都建基于抛物线（parabola）$x = at^2, y = 2at$ 的一个事实：连接参数（parameter）为 t_1 和 t_2 两点的弦（chord），其斜率（gradient）为 $\dfrac{2}{t_1 + t_2}$。有了这个事实，“法线与曲线再次相交”和“半圆上的圆周角是直角”都变成参数之间的简单方程，几何问题便化为简短的代数运算。

解答

一个有用的事实：弦的斜率

对于相异参数 $t_1 \neq t_2$，连接 $(at_1^2, 2at_1)$ 与 $(at_2^2, 2at_2)$ 的弦的斜率为

$$\frac{2a(t_1 - t_2)}{a(t_1^2 - t_2^2)} = \frac{2}{t_1 + t_2},$$

这里约去了 $t_1 - t_2 \neq 0$。（若 $t_1 + t_2 = 0$，则该弦是铅垂的。）令 $t_2 \to t_1$，可知在 t 处的切线（tangent）斜率为 $\frac{1}{t}$；直接计算亦得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{2a}{2at} = \dfrac{1}{t}$。

第 (i) 部分

在 P 处切线斜率为 $\frac{1}{p}$，故法线（normal）斜率为 $-p$（这里用到 $p \neq 0$）。点 N（其中 $n \neq p$）在这条法线上，当且仅当弦 PN 的斜率为 $-p$：

$$\frac{2}{p + n} = -p \iff p + n = -\frac{2}{p} \iff n = -\left(p + \frac{2}{p}\right).$$

（弦 PN 不可能是铅垂的，因为 P 处的法线斜率 $-p$ 是有限的。）注意 $n \neq p$，因为 $n-p=-2\left(p+\frac{1}{p}\right)$，而对实数 p 而言 $p+\frac{1}{p} \neq 0$，所以这的确是另一个点。这就是所求的结果。

第 (ii) 部分

利用 $p+n=-\frac{2}{p}$ 及 $p-n=2p+\frac{2}{p}=\dfrac{2(p^2+1)}{p}$，

$$\begin{aligned}|PN|^2 &= a^2(p^2-n^2)^2+4a^2(p-n)^2\\ &= a^2(p-n)^2\left[(p+n)^2+4\right]\\ &= a^2\cdot\frac{4(p^2+1)^2}{p^2}\cdot\left(\frac{4}{p^2}+4\right)=a^2\cdot\frac{4(p^2+1)^2}{p^2}\cdot\frac{4(p^2+1)}{p^2}\\ &= 16a^2\,\frac{(p^2+1)^3}{p^4},\end{aligned}$$

正如所求。

为求最小值，设 $u=p^2>0$ 及 $f(u)=\dfrac{(u+1)^3}{u^2}$。由商法则（quotient rule），

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

当 $u>0$ 时，因子 $\frac{(u+1)^2}{u^3}$ 为正，所以当 $0<u<2$ 时 $f'(u)<0$，当 $u>2$ 时 $f'(u)>0$。因此 f 先递减后递增，其在 $u>0$ 上的最小值出现在 $u=2$，即 $p^2=2$。由于 $|PN|^2=16a^2f(p^2)$，这亦使 $|PN|$ 取得最小值。（最小值为 $|PN|^2=16a^2\cdot\frac{27}{4}=108a^2$，故 $|PN|=6\sqrt{3}\,a$。）

第 (iii) 部分

由于 PN 是直径（diameter），而 Q 是圆上异于 P 和 N 的点，所以角 PQN 是直角（半圆上的圆周角（angle in a semicircle））。因此 $QP \perp QN$。

> ? **想一想** QP 或 QN 会否是铅垂的，以致我们不能把斜率相乘?

若 QP 是铅垂的，则 QN 是水平的，于是 Q 和 N 的 y 坐标相同，得 $q=n$，但由于 $Q \neq N$，这是不可能的。同理，若 QN 是铅垂的，则 QP 必为水平，得 $q=p$。所以两个斜率都是有限且非零的，而它们的乘积为 -1：

$$\frac{2}{q+p}\cdot\frac{2}{q+n}=-1 \iff (q+p)(q+n)=-4.$$

展开，并利用第 (i) 部分的 $p+n=-\frac{2}{p}$ 及 $pn=-p^2-2$，

$$q^2+(p+n)q+pn=q^2-\frac{2q}{p}-p^2-2=-4,$$

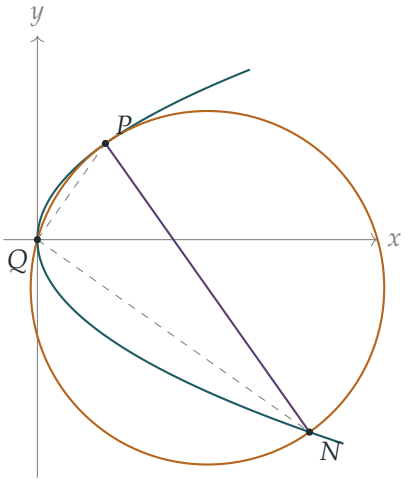
整理后得

$$2 = p^2 - q^2 + \frac{2q}{p},$$

正如所求。

推论。代入 $q = 0$：关系式变成 $2 = p^2$。反之，若 $p^2 = 2$，关系式给出 $q^2 - \frac{2q}{p} = 0$，即 $q = 0$ 或 $q = \frac{2}{p}$；但 $\frac{2}{p} = \frac{p^2}{p} = p$，这就是点 P 本身，所以 Q 必有 $q = 0$。因此 Q 位于原点当且仅当 $p^2 = 2$，而由第 (ii) 部分，这恰好就是 $|PN|$ 取最小值的情况。

一般而言，该关系式是一条二次方程 $q^2 - \frac{2}{p}q + (2 - p^2) = 0$，其两根之和为 $\frac{2}{p}$；其中一根必定是 $q = p$，所以另一根为 $q = \dfrac{2 - p^2}{p}$，它为零当且仅当 $p^2 = 2$。



上图显示取最小值的情况 $a = 1, p = \sqrt{2}$：法线 PN（梅紫色）、以 PN 为直径的圆（琥珀色），以及在 $Q = O$ 处的直角。

验算

- 取 $a = 1, p = 1$：$n = -3$，所以 $P = (1, 2)$ 及 $N = (9, -6)$。于是 $|PN|^2 = 64 + 64 = 128$，而公式给出 $16 \cdot 8/1 = 128$。
- 取 $p = \sqrt{2}, a = 1$：$P = (2, 2\sqrt{2})$，$N = (8, -4\sqrt{2})$，且 $\overrightarrow{OP} \cdot \overrightarrow{ON} = 16 - 16 = 0$，所以原点确实在以 PN 为直径的圆上。此时 $|PN| = 6\sqrt{3} \approx 10.392$。
- 对多个 p 值以数值方法找出圆与抛物线的交点，除 P 外，所得的参数恰好就是 n 和 $\frac{2-p^2}{p}$。

常见错误

- 忘记 $q = p$ 亦满足最后的关系式，因而错误地断定当 $p^2 = 2$ 时 Q 有两个可能位置。
- 把法线斜率写成 $-\frac{1}{p}$：切线斜率是 $\frac{1}{p}$，所以法线斜率是 $-p$。

STEP 之外

一个圆与一条抛物线一般相交于四点；把 $x = at^2, y = 2at$ 代入圆的方程，会得到一条关于 t 且没有 t^3 项的四次方程（quartic），所以四个参数之和为零。这里 $n + q = -\frac{2}{p} - p + \frac{2}{p} = -p$，所以第四个参数又是 p：圆在 P 处与抛物线相切。这是必然的，因为圆在该处的半径正沿着法线方向。像这样按重数（multiplicity）计算交点数目，正是代数几何（algebraic geometry）的起点（贝祖定理（Bézout's theorem））。如想做更多这类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶难题》）。

详细解答

STEP II 2017，第 6 题

纯数学 · 不等式与数学归纳法

STEP II 2017，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/17-s2-q6

关键思路

两部分都是把和 S_n 与 $2\sqrt{n}$ 比较，而 $2\sqrt{n}$ 大约等于 $\int_0^n x^{-1/2}\,\mathrm{d}x$。在第 (i) 部分的归纳法（induction）中，由 n 推到 $n+1$ 的一步，归结为比较 $1/\sqrt{n+1}$ 与 $2(\sqrt{n+1}-\sqrt{n})$，而把这个差有理化（rationalising）便可解决。在第 (ii) 部分，最小的 C 就是右方（不计 C）与 S_n 之间差距的最大值。因此我们研究这个差距由 n 到 $n+1$ 如何变化，而题目给出的不等式（inequality）正好告诉我们这个差距总是在缩小。

解答

第 (i) 部分

设 $P(n)$ 为命题 $S_n \le 2\sqrt{n}-1$。

基本情况。$S_1 = 1$ 而 $2\sqrt{1}-1=1$，所以 $P(1)$ 成立（等号成立）。

归纳步骤。假设对某个 $k \ge 1$，$P(k)$ 成立。那么

$$S_{k+1} = S_k + \frac{1}{\sqrt{k+1}} \le 2\sqrt{k} - 1 + \frac{1}{\sqrt{k+1}},$$

所以只需证明

$$2\sqrt{k} + \frac{1}{\sqrt{k+1}} \le 2\sqrt{k+1}, \qquad \text{即是} \qquad \frac{1}{\sqrt{k+1}} \le 2\left(\sqrt{k+1}-\sqrt{k}\right).$$

有理化后，

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

因为 $\sqrt{k} \le \sqrt{k+1}$ 使分母不大于 $2\sqrt{k+1}$。因此 $S_{k+1} \le 2\sqrt{k+1} - 1$，这正是 $P(k+1)$。由于 $P(1)$ 为真，而且对每个 $k \ge 1$ 都有 $P(k) \Rightarrow P(k+1)$，所以对所有正整数 n，$S_n \le 2\sqrt{n} - 1$。

第 (ii) 部分：不等式

当 $k \ge 0$ 时，$(4k+1)\sqrt{k+1} > (4k+3)\sqrt{k}$ 的两边都是非负的，而且左方是正数，所以这个不等式等价于两边平方后的不等式。现在

$$\begin{aligned}(4k+1)^2(k+1) - (4k+3)^2 k &= (16k^2+8k+1)(k+1) - (16k^2+24k+9)k \\ &= (16k^3+24k^2+9k+1) - (16k^3+24k^2+9k) \\ &= 1 > 0.\end{aligned}$$

所以 $(4k+1)^2(k+1) > (4k+3)^2 k$，取正平方根便得对所有 $k \ge 0$，$(4k+1)\sqrt{k+1} > (4k+3)\sqrt{k}$，得证。

第 (ii) 部分：最小的 C

定义

$$T_n = 2\sqrt{n} + \frac{1}{2\sqrt{n}} - S_n.$$

不等式 $S_n \ge 2\sqrt{n} + \frac{1}{2\sqrt{n}} - C$ 与 $C \ge T_n$ 相同。所以它对每个 n 都成立，当且仅当 C 不小于每一个 T_n；而最小的这样的 C 就是 T_n 的最大值（如果存在的话）。

想一想 在不知道 S_n 的公式的情况下，怎样找出最大的 T_n？

我们考虑变化量 $T_{n+1} - T_n$，在其中 S_n 会消去。由于 $S_{n+1} - S_n = 1/\sqrt{n+1}$，

$$\begin{aligned}T_{n+1} - T_n &= 2\sqrt{n+1} - 2\sqrt{n} + \frac{1}{2\sqrt{n+1}} - \frac{1}{2\sqrt{n}} - \frac{1}{\sqrt{n+1}} \\ &= 2\sqrt{n+1} - 2\sqrt{n} - \frac{1}{2\sqrt{n+1}} - \frac{1}{2\sqrt{n}}.\end{aligned}$$

两边乘以正数 $2\sqrt{n}\sqrt{n+1}$：

$$\begin{aligned}2\sqrt{n}\sqrt{n+1}\left(T_{n+1} - T_n\right) &= 4(n+1)\sqrt{n} - 4n\sqrt{n+1} - \sqrt{n} - \sqrt{n+1} \\ &= (4n+3)\sqrt{n} - (4n+1)\sqrt{n+1}.\end{aligned}$$

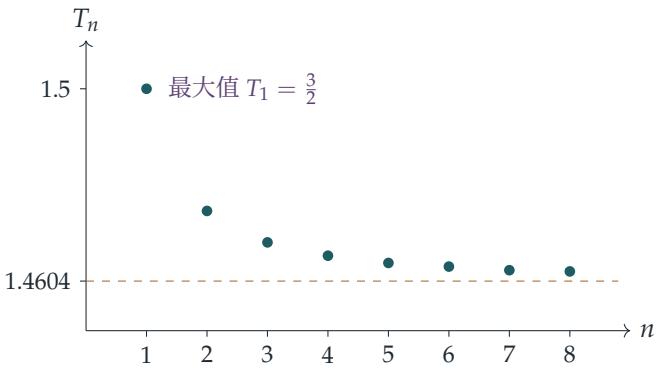
根据刚才证明的不等式（取 $k = n$），这是负数。因此对每个 $n \ge 1$ 都有 $T_{n+1} < T_n$：数列（sequence）T_n 严格递减（strictly decreasing），其最大值就是第一项，

$$T_1 = 2 + \tfrac{1}{2} - 1 = \tfrac{3}{2}.$$

所以对所有 $n \geq 1$，$S_n \geq 2\sqrt{n} + \frac{1}{2\sqrt{n}} - \frac{3}{2}$。任何更小的 C 都不行，因为在 $n = 1$ 时不等式变成 $1 \geq \frac{5}{2} - C$，这要求 $C \geq \frac{3}{2}$。所求的最小数是

$$C = \tfrac{3}{2}.$$

T_n 的首几个值显示它稳定下来的速度有多快：



（纵轴经过拉伸：它的范围大约由 1.45 至 1.51。）

验算

- 当 $n = 4$：$S_4 = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{2} \approx 2.7845$，它介乎 $2\sqrt{4} + \frac{1}{4} - \frac{3}{2} = 2.75$ 与 $2\sqrt{4} - 1 = 3$ 之间，符合两个结果的要求。
- 用电脑检查首 200 000 个值，确认了两个界都成立，而且 T_n 递减；T_n 由 1.5 下降，趋向大约 1.46035，并且永不回到 1.5。
- 恒等式 $(4k+1)^2(k+1) - (4k+3)^2 k = 1$ 很容易在 $k = 1$ 时验证：$25 \cdot 2 - 49 = 1$。

常见错误

- 以 T_n 的极限（大约 1.46）作答，而不是它的最大值。小于 $\frac{3}{2}$ 的 C 在 $n = 1$ 时不成立。
- 把不等式平方之前，没有先指出两边都是非负的。
- 忘记证明 $C = \frac{3}{2}$ 确实是最小的，即没有指出等号成立的情况。

STEP 之外

数列 T_n 递减至一个极限，而这个极限是 $-\zeta(\frac{1}{2}) \approx 1.46035$，其中 ζ 是黎曼 zeta 函数（Riemann zeta function），并已延拓至其定义级数收敛的区域以外。把和与积分比较，再

以像 $\frac{1}{2\sqrt{n}}$ 这样的项作修正，正是欧拉－麦克劳林公式（Euler–Maclaurin formula）的起点。背后的基本事实——有界的单调数列必定收敛——就是大学一年级分析课中的单调收敛定理（monotone convergence theorem）；Abbott 的 *Understanding Analysis*（《理解分析》）对此有仔细的论述。

详细解答

STEP II 2017，第 7 题

纯数学 · 对数微分法

STEP II 2017，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/17-s2-q7

关键思路

这里每一个幂都是通过对数（logarithm）来定义的，即 $a^b = \mathrm{e}^{b\ln a}$，所以比较 x 的两个幂，就等于比较它们的对数；由于 exp 是递增（increasing）函数，两种比较是等价的。唯一要留意的是 $\ln x$ 的正负号：当 $0 < x < 1$ 时它是负数，而除以负数会令不等式反向。取对数亦令求导数（derivative）变得直截了当。

解答

全文中，$\ln \mathrm{f}(x) = x\ln x$ 及 $\ln \mathrm{g}(x) = \mathrm{f}(x)\ln x$。

第 (i) 部分

设 $0 < x < 1$，则 $\ln x < 0$ 且 $x - 1 < 0$。于是

$$\ln \mathrm{f}(x) - \ln x = x\ln x - \ln x = (x-1)\ln x > 0,$$

因为这是两个负数的乘积。由于 exp 是递增的，所以 $\mathrm{f}(x) > x$。

此外 $x\ln x < 0$，所以 $\mathrm{f}(x) = \mathrm{e}^{x\ln x} < 1$。现在再比较对数：

$$\begin{aligned} \ln \mathrm{g}(x) - \ln x &= \big(\mathrm{f}(x) - 1\big)\ln x > 0, && \text{因为} \mathrm{f}(x) < 1 \text{ 且 } \ln x < 0, \\ \ln \mathrm{f}(x) - \ln \mathrm{g}(x) &= \big(x - \mathrm{f}(x)\big)\ln x > 0, && \text{因为} \mathrm{f}(x) > x \text{ 且 } \ln x < 0. \end{aligned}$$

所以 $\ln x < \ln \mathrm{g}(x) < \ln \mathrm{f}(x)$，由此得出当 $0 < x < 1$ 时 $x < \mathrm{g}(x) < \mathrm{f}(x)$，正如所求。

当 $x > 1$ 时，同样的差仍然适用，但此时 $\ln x > 0$、$x - 1 > 0$ 及 $x\ln x > 0$，所以 $\mathrm{f}(x) > 1$。

于是由 $(x-1)\ln x > 0$ 得 $\mathrm{f}(x) > x$，而由 $(x-\mathrm{f}(x))\ln x < 0$ 得 $\mathrm{g}(x) > \mathrm{f}(x)$。所以

$$x < \mathrm{f}(x) < \mathrm{g}(x) \qquad \text{当} x > 1,$$

而在 $x = 1$ 时三者皆等于 1。

第 (ii) 部分

对 $\ln \mathrm{f}(x) = x \ln x$ 求导，得 $\mathrm{f}'(x)/\mathrm{f}(x) = \ln x + 1$，所以

$$\mathrm{f}'(x) = x^x(\ln x + 1).$$

由于 $x^x > 0$，$\mathrm{f}'(x) = 0$ 当且仅当 $\ln x = -1$，即 $x = \mathrm{e}^{-1}$。（f′ 在该处由负变正，所以这是极小值（minimum），其值为 $\mathrm{e}^{-1/\mathrm{e}} \approx 0.692$。）

第 (iii) 部分

当 $x \to 0$ 时，$x \ln x \to 0$，而 exp 是连续（continuous）的，所以

$$\mathrm{f}(x) = \mathrm{e}^{x \ln x} \to \mathrm{e}^0 = 1.$$

接着 $\ln \mathrm{g}(x) = \mathrm{f}(x) \ln x$，其中 $\mathrm{f}(x) \to 1$ 而 $\ln x \to -\infty$，所以 $\ln \mathrm{g}(x) \to -\infty$，因此

$$\lim_{x \to 0} \mathrm{g}(x) = 0.$$

第 (iv) 部分

对 $x > 0$，设 $h(x) = x^{-1} + \ln x$。则

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

它在 $0 < x < 1$ 时为负，在 $x > 1$ 时为正。所以 h 先递减后递增，其最小值为 $h(1) = 1$。因此对所有 $x > 0$，$x^{-1} + \ln x \geq 1$。

现在用积法则（product rule）对 $\ln \mathrm{g}(x) = \mathrm{f}(x) \ln x$ 求导，并利用第 (ii) 部分：

$$\frac{\mathrm{g}'(x)}{\mathrm{g}(x)} = \mathrm{f}'(x) \ln x + \frac{\mathrm{f}(x)}{x} = x^x \left[(\ln x + 1) \ln x + \frac{1}{x} \right] = x^x \left[(\ln x)^2 + \ln x + \frac{1}{x} \right].$$

想一想 刚证明的不等式怎样帮助处理这个括号?

括号内正好包含 $\frac{1}{x} + \ln x$，而它至少是 1。所以

$$(\ln x)^2 + \ln x + \frac{1}{x} \;\geq\; (\ln x)^2 + 1 \;>\; 0.$$

由于 $g(x) > 0$ 及 $x^x > 0$，我们得到

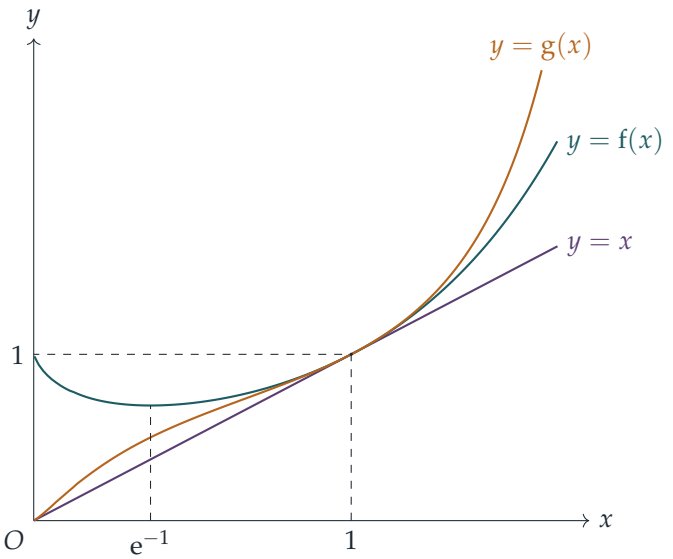
$$g'(x) = g(x)\, x^x \left[(\ln x)^2 + \ln x + \frac{1}{x} \right] > 0 \qquad \text{对所有} x > 0,$$

正如所求。

草图

把各项事实汇集起来：

- 当 $x \to 0$ 时，$y = f(x)$ 从 1 开始，下降至 $x = e^{-1}$ 处的极小值 $e^{-1/e}$，然后上升经过 $(1,1)$，并增长得非常快。在 0 附近它非常陡峭，因为 $f'(x) = x^x(\ln x + 1) \to -\infty$。
- $y = g(x)$ 从原点开始，并一直递增（第 (iv) 部分），经过 $(1,1)$。
- 当 $0 < x < 1$ 时，次序为 $x < g < f$；当 $x > 1$ 时，次序为 $x < f < g$。所以 g 与 f 在 $(1,1)$ 相交，而三条曲线都在该处相遇。
- 在 $x = 1$：$f'(1) = 1$ 及 $g'(1) = 1 \cdot 1 \cdot (0 + 0 + 1) = 1$，所以两条曲线都在 $(1,1)$ 与 $y = x$ 相切。
- 在原点附近 $g(x)/x = e^{(f(x)-1)\ln x} \to 1$（因为 $f(x) - 1 \approx x \ln x$，指数的表现类似 $x(\ln x)^2 \to 0$），所以 $y = g(x)$ 沿着 $y = x$ 离开原点，并略高于它。



验算

- 在 $x = \frac{1}{2}$：$\mathrm{f} = \frac{1}{\sqrt{2}} \approx 0.707$ 及 $\mathrm{g} = (\frac{1}{2})^{0.707} \approx 0.613$，所以 $0.5 < 0.613 < 0.707$。在 $x = 2$：$\mathrm{f} = 4$ 及 $\mathrm{g} = 2^4 = 16$，所以 $2 < 4 < 16$。
- $\mathrm{f}(0.3) \approx 0.697$、$\mathrm{f}(\mathrm{e}^{-1}) \approx 0.6922$ 及 $\mathrm{f}(0.4) \approx 0.693$，与在 e^{-1} 处取极小值相符。
- g' 的公式给出 $\mathrm{g}'(\frac{1}{2}) \approx 0.774$，与 g 在该处的差商（difference quotient）吻合；电脑验算亦确认了在 $(0,3)$ 上 $\mathrm{g}' > 0$。

常见错误

- 忘记在 $(0,1)$ 上 $\ln x < 0$，因而除以它时没有把不等式反向；这正正是次序在 $x = 1$ 处对调的原因。
- 把 x^x 的导数写成 $x \cdot x^{x-1}$（把指数（index）当作常数），或写成 $x^x \ln x$（把底（base）当作常数）；两部分都是需要的。
- 把 g 的极限（limit）写成 $1^{-\infty}$ 然后靠猜；应改为处理 $\ln \mathrm{g} = \mathrm{f} \ln x$。

STEP 之外

幂塔 $x^{x^{x^{\cdots}}}$ 延续了这个规律：当 $0 < x < 1$ 时，奇数层的幂塔 x, g, … 与偶数层的幂塔 f, … 互相交错；而无穷幂塔收敛（converges）当且仅当 $\mathrm{e}^{-\mathrm{e}} \le x \le \mathrm{e}^{1/\mathrm{e}}$，这是关于单调数列（monotone sequences）和不动点（fixed points）的一道好练习。像 $x \ln x \to 0$ 这类极限，以及通过 exp 和 ln 严谨地定义 a^b，都会在大学第一个分析（analysis）课程中得到妥善处理；可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

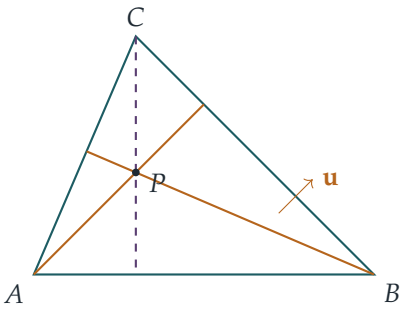
STEP II 2017，第 8 题

纯数学 · 向量与垂心

> **STEP II 2017，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/17-s2-q8

关键思路

这是用向量证明三角形三条高（altitude）相交于同一点，即垂心（orthocentre）。要求两条直线 $\mathbf{a}+\lambda\mathbf{u}$ 与 $\mathbf{b}+\mu\mathbf{v}$ 的交点，窍门是把方程与一个垂直于 $\mathbf{v}$ 的向量取纯量积（scalar product）：这样一下子就消去了 μ。这里这样的向量已经现成给了我们，因为 $\mathbf{v}$ 垂直于 CA，所以 $\mathbf{c}-\mathbf{a}$ 正好合用。



解答

第一条高

通过 A 且垂直于 BC 的直线经过 $\mathbf{a}$，方向为 $\mathbf{u}$（我们身处平面之中，所以每个垂直于 BC 的向量都是 $\mathbf{u}$ 的倍数）。其方程为

$$\mathbf{r}=\mathbf{a}+\lambda\mathbf{u}.$$

同理，通过 B 且垂直于 CA 的直线是 $\mathbf{r}=\mathbf{b}+\mu\mathbf{v}$。

求 P

这两条直线并不平行：若 $\mathbf{u}$ 与 $\mathbf{v}$ 平行，那么在同一平面上分别垂直于它们的 BC 与 CA 便会平行，这对三角形来说是不可能的。因此两线恰好相交于一点 P，在该点有

$$\mathbf{a}+\lambda\mathbf{u}=\mathbf{b}+\mu\mathbf{v}.$$

把两边都与 $\mathbf{c}-\mathbf{a}$ 取纯量积。由于 $\mathbf{v}$ 垂直于 CA，$\mathbf{v}\cdot(\mathbf{c}-\mathbf{a})=0$，于是 μ 消失：

$$\mathbf{a}\cdot(\mathbf{c}-\mathbf{a})+\lambda\,\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=\mathbf{b}\cdot(\mathbf{c}-\mathbf{a}),\qquad\text{所以}\qquad\lambda\,\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=(\mathbf{b}-\mathbf{a})\cdot(\mathbf{c}-\mathbf{a}).$$

> ? **想一想** 我们可以除以 $\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})$ 吗?

可以。若 $\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=0$，则 $\mathbf{u}$ 既垂直于 BC，亦垂直于 CA；在平面上，这迫使 CA 与 BC 平行，但它们并不平行。因此

$$\mathbf{p}=\mathbf{a}+\frac{(\mathbf{b}-\mathbf{a})\cdot(\mathbf{c}-\mathbf{a})}{\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})}\,\mathbf{u}.$$

CP 垂直于 AB

我们需要证明 $(\mathbf{p}-\mathbf{c})\cdot(\mathbf{b}-\mathbf{a})=0$。以 λ 记上面求得的系数，则

$$(\mathbf{p}-\mathbf{c})\cdot(\mathbf{b}-\mathbf{a})=(\mathbf{a}-\mathbf{c})\cdot(\mathbf{b}-\mathbf{a})+\lambda\,\mathbf{u}\cdot(\mathbf{b}-\mathbf{a}).$$

有用的观察是：由于 $\mathbf{u}$ 垂直于 BC，$\mathbf{u}\cdot(\mathbf{b}-\mathbf{c})=0$，所以

$$\mathbf{u}\cdot(\mathbf{b}-\mathbf{a})=\mathbf{u}\cdot(\mathbf{b}-\mathbf{c})+\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=\mathbf{u}\cdot(\mathbf{c}-\mathbf{a}).$$

因此 $\lambda\,\mathbf{u}\cdot(\mathbf{b}-\mathbf{a})=\lambda\,\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=(\mathbf{b}-\mathbf{a})\cdot(\mathbf{c}-\mathbf{a})$，而

$$(\mathbf{p}-\mathbf{c})\cdot(\mathbf{b}-\mathbf{a})=-(\mathbf{c}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a})+(\mathbf{b}-\mathbf{a})\cdot(\mathbf{c}-\mathbf{a})=0.$$

> ? **想一想** 纯量积为零，真的能给出一条直线 CP 吗?

只有在 $P\neq C$ 时才行，而这正是用到“非直角”这个条件的地方。若 $P=C$，则 C 位于通过 A 且垂直于 BC 的直线上，于是 $AC\perp BC$，C 处的角是直角，而这已被排除。所以 $\mathbf{p}-\mathbf{c}\neq\mathbf{0}$；又因 $\mathbf{b}-\mathbf{a}\neq\mathbf{0}$，纯量积为零正好表示 CP 垂直于 AB。因此三条高都通过 P。

验算

取 $A=(0,0)$、$B=(5,0)$、$C=(1.5,3.5)$ 及 $\mathbf{u}=(1,1)$，它垂直于 $\mathbf{c}-\mathbf{b}=(-3.5,3.5)$。于是 $(\mathbf{b}-\mathbf{a})\cdot(\mathbf{c}-\mathbf{a})=7.5$，$\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})=5$，所以 $\lambda=1.5$，$P=(1.5,1.5)$。这一点位于 C 的正

下方, 所以 CP 是铅垂的, 垂直于 AB; 而 $(\mathbf{p}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{a}) = (-3.5)(1.5)+(1.5)(3.5) = 0$, 所以 P 亦在由 B 出发的高上。一般结果亦已用电脑在大量随机三角形上验证过。另请留意，若把 $\mathbf{u}$ 换成 $k\mathbf{u}$，$\mathbf{p}$ 的公式不变，这是理所当然的。

常见错误

- 与 $\mathbf{v}$ 或 $\mathbf{u}$ 而非与 $\mathbf{c}-\mathbf{a}$ 作点积：这样方程中两个参数都会留下。
- 除以 $\mathbf{u}\cdot(\mathbf{c}-\mathbf{a})$ 时没有说明它为何不等于零。
- 漏掉 $\mathbf{u}\cdot(\mathbf{b}-\mathbf{a}) = \mathbf{u}\cdot(\mathbf{c}-\mathbf{a})$ 这一步，而最后能够相消，靠的正是这一步。

STEP 之外

任何三角形的垂心 H、形心（centroid）G 与外心（circumcentre）O 都位于同一直线上, 即欧拉线（Euler line), 且 $\overrightarrow{OH} = 3\overrightarrow{OG}$; 若以 O 为原点, 这就是简洁的命题 $\mathbf{h} = \mathbf{a}+\mathbf{b}+\mathbf{c}$, 你可以用同样的纯量积思路来证明它。这类纯量积正是大学线性代数（linear algebra）入门课程中所研究的内积空间（inner-product space）的起点；Strang 的 *Introduction to Linear Algebra*（《线性代数导论》）是看它们如何发展下去的好地方。

详细解答

STEP II 2017，第 9 题

力学 · 摩擦力与刚体

STEP II 2017，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/17-s2-q9

关键思路

每个圆柱都是刚体（rigid body），而且只在两处受力：木板和地面。因此第一步自然是对圆柱的中心取力矩（moments）：两个法向反作用力（normal reaction）都通过中心，剩下的只有两个摩擦力（friction），而它们必须相等。之后只需分别对圆柱和木板分解力（resolving），再代入接触点的摩擦定律 $G \leq \frac{1}{2}R$ 和 $F \leq \frac{1}{2}N$。最后才用几何关系把 θ 与 a 和 r 联系起来。

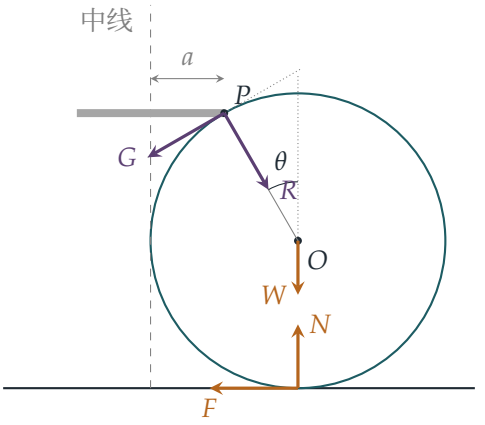
解答

模型与设置。 把横截面看作两个半径为 r 的圆盘和一根水平杆，所有接触点都是粗糙的。两个圆柱之间的距离可忽略，所以它们的中心分别位于对称铅垂线两侧、水平距离为 r 之处。木板宽度为 $2a < 2r$，所以它以两条长边搁在两个圆柱内侧的上部。由对称性，我们只需考虑右边的圆柱。

设 P 为木板与该圆柱的接触点。P 处的切线（tangent）与水平的木板成 θ 角，所以通往 P 的半径与铅垂线成 θ 角。作用在圆柱上的力为：

- 其重量 W，作用于中心 O；
- 在地面处，向上的法向反作用力 N 和摩擦力 F（指向中线，因为木板把圆柱向外推）；
- 在 P 处，沿 PO 方向的法向反作用力 R，以及沿切线方向、指向下方和内侧的摩擦力 G。

根据牛顿第三定律（Newton's third law），木板受到沿 OP 方向的 R（向上、背离中线），以及沿切线方向、指向上方和外侧的 G。



通过 P 的虚线是切线。在水平方向上，P 比中心 O 更靠内 $r\sin\theta$，而 O 与中线的距离为 r。

第 (i) 部分

对圆柱取关于 O 的力矩。W、N 和 R 都通过 O。两个摩擦力都与圆相切，所以各自的力矩为（力）$\times r$，而且转动方向相反：F 在底部，指向内侧；G 在 P 处，指向下方和内侧。因此

$$Fr = Gr \quad \Longrightarrow \quad G = F.$$

对圆柱作水平分解（以向外为正）：R 的水平分量为向外的 $R\sin\theta$，G 的分量为向内的 $G\cos\theta$，而地面摩擦力 F 向内，所以

$$R\sin\theta - G\cos\theta - F = 0.$$

由于 $G = F$，上式即

$$R\sin\theta = F(1+\cos\theta),$$

此即所求。

推论。木板处没有滑动，所以该处的摩擦力满足 $G \le \frac{1}{2}R$，即 $F \le \frac{1}{2}R$。于是

$$R\sin\theta = F(1+\cos\theta) \le \tfrac{1}{2}R(1+\cos\theta),$$

两边除以 $\frac{1}{2}R > 0$（木板压在圆柱上），得 $2\sin\theta \le 1+\cos\theta$。

第 (ii) 部分

对木板作铅垂分解。每条边都受到 R，其向上分量为 $R\cos\theta$；又受到沿切线向上的 $G = F$，其向上分量为 $F\sin\theta$。所以

$$2(R\cos\theta + F\sin\theta) = kW.$$

由第 (i) 部分，$R = \dfrac{F(1+\cos\theta)}{\sin\theta}$，所以

$$R\cos\theta + F\sin\theta = \frac{F\left[(1+\cos\theta)\cos\theta + \sin^2\theta\right]}{\sin\theta} = \frac{F(1+\cos\theta)}{\sin\theta},$$

这是因为 $\cos^2\theta + \sin^2\theta = 1$。因此

$$\tfrac{1}{2}kW = \frac{(1+\cos\theta)}{\sin\theta}F.$$

（由对称性，木板所受的水平力互相抵消，力矩亦然。）

对圆柱作铅垂分解。木板施加的反作用力向下推，大小为 $R\cos\theta + G\sin\theta = \frac{1}{2}kW$，所以

$$N = W + \tfrac{1}{2}kW = \left(\frac{2}{k}+1\right)\tfrac{1}{2}kW = \left(1+\frac{2}{k}\right)\left(\frac{1+\cos\theta}{\sin\theta}\right)F.$$

这就是所求的表达式。（这不过是显而易见的事实：地面承受总重量 $2W + kW$ 的一半。）

地面处没有滑动。条件为

$$F \le \tfrac{1}{2}N.$$

由 N 的表达式，

$$\frac{N}{F} = \left(1+\frac{2}{k}\right)\frac{1+\cos\theta}{\sin\theta}.$$

由于 θ 为锐角时 $\sin\theta > 0$，第 (i) 部分给出 $\dfrac{1+\cos\theta}{\sin\theta} \ge 2$；又因 $k > 0$，有 $1 + \frac{2}{k} > 1$。所以 $N/F > 2$，即 $F < \frac{1}{2}N$。因此只要木板不滑动，地面的条件便自动成立，对 θ 没有额外限制。

> ? **想一想** 两个接触点的摩擦力同样是 F，为何地面接触点的条件会“免费”成立？

因为地面压得更用力。比较 R 和 N 的公式，得 $N = \left(1+\frac{2}{k}\right)R > R$，因为地面还要承受圆柱本身的重量。两个接触点的摩擦力相同，而地面的法向力较大，所以总是木板接触点先滑动。

第 (iii) 部分

把第 (i) 部分的不等式移项后平方。两边都是正数，所以平方后不等号方向不变：

$$4\sin^2\theta \le (1+\cos\theta)^2 \iff 4(1-\cos\theta)(1+\cos\theta) \le (1+\cos\theta)^2.$$

除以 $1+\cos\theta > 0$，得 $4 - 4\cos\theta \le 1 + \cos\theta$，即 $\cos\theta \ge \frac{3}{5}$。由于 θ 为锐角，

$$\sin\theta = \sqrt{1-\cos^2\theta} \le \sqrt{1-\tfrac{9}{25}} = \tfrac{4}{5}.$$

（等价地，设 $t = \tan\frac{1}{2}\theta$，则不等式变为 $2t \le 1$，而 $\sin\theta = \frac{2t}{1+t^2}$ 在 $t = \frac{1}{2}$ 时最大为 $\frac{4}{5}$。）

几何关系。木板的边 P 与中线的水平距离为 a。中心 O 与中线的距离为 r，而 P 比 O 更接近中线 $r\sin\theta$。所以

$$a = r - r\sin\theta = r(1-\sin\theta) \geq r\left(1-\tfrac{4}{5}\right) = \tfrac{1}{5}r,$$

由此得 $r \leq 5a$。

验算

- 在极端情况 $\sin\theta = \frac{4}{5}$、$\cos\theta = \frac{3}{5}$ 下，取 $W = 1$ 及 $k = 1$。木板的方程给出 $F = \frac{1}{4}$，于是 $R = \frac{1}{2}$、$N = \frac{3}{2}$。所以 $G/R = \frac{1}{2}$ 恰好成立（木板处于将要滑动的状态），而 $F/N = \frac{1}{6}$，远低于 $\frac{1}{2}$，与第 (ii) 部分的预测一致。
- 四条平衡（equilibrium）方程（力矩方程，以及对两个物体的分解）都经电脑以 10^5 组随机的 W、k 和 θ 验证。同时亦验证了：$G \leq \frac{1}{2}R$ 成立当且仅当 $\sin\theta \leq \frac{4}{5}$，而此时 $F < \frac{1}{2}N$ 总是成立。
- 极限情况的合理性：宽木板（$a \to r$）搁在接近圆柱顶部的位置，所以 $\theta \to 0$，几乎不需要摩擦力。窄木板则深深卡在两个圆柱之间，最终会从两者之间滑下去。

常见错误

- 漏掉力矩方程，因而没有发现 $G = F$。没有这一点，未知数就太多了。
- 弄错木板处摩擦力的方向：两个圆柱有向两旁滚开的趋势，所以作用在圆柱上的摩擦力指向下方和内侧。
- 写成 $a = r\sin\theta$ 而不是 $a = r(1-\sin\theta)$：a 是从中线量起的，而不是从圆柱的中心量起。

STEP 之外

这是一个小型的静定（statically determinate）系统：三个物体，方程数目足以确定每一个力，而摩擦力给出的不等式则划定了可能的位置。在大学一年级的课程中，同样的概念会以刚体平衡条件（合力为零及合力矩为零）的形式再次出现，之后又会在楔入与卡死（wedging and jamming）的研究中出现。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对刚体静力学和摩擦力有仔细的论述。

详细解答

STEP II 2017，第 10 题

力学 · 克服阻力所作的功

> **STEP II 2017，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/17-s2-q10

关键思路

牛顿第二定律（Newton's second law）给出每一点的驱动力（driving force），而引擎所做的功（work）就是这个力对距离的积分（integral）。在等加速度的情况下，$v\frac{\mathrm{d}v}{\mathrm{d}x}=a$ 可以把对 x 的积分转化为对 v 的积分。唯一微妙之处在于引擎只能推、不能拉：当所需的力变成负数时，便由煞车负责，引擎毫无贡献，所以我们只在驱动力为正的范围内积分。

解答

把汽车视为沿 x 轴运动的质点（particle），引擎提供的驱动力为 F。汽车向前行驶时，阻力（resistance）Av^2+R 向后作用。

前半段。 由牛顿第二定律得 $F-(Av^2+R)=ma$，所以

$$F=ma+R+Av^2,$$

它在整段过程中都是正数。力 F 使其作用点移动 $\mathrm{d}x$ 时所做的功为 $F\,\mathrm{d}x$，因此引擎在前半段所做的功为

$$\int_0^d (ma+R+Av^2)\,\mathrm{d}x,$$

正如所求。至于变量代换（change of variable），加速度为 $v\frac{\mathrm{d}v}{\mathrm{d}x}=a$，所以 $\mathrm{d}x=\frac{v}{a}\,\mathrm{d}v$。由于 $v^2=2ax$（由静止开始以等加速度运动），v 由 $x=0$ 时的 0 变到 $x=d$ 时的 $\sqrt{2ad}=w$。因此所做的功为

$$\int_0^w \frac{(ma+R+Av^2)\,v}{a}\,\mathrm{d}v,$$

正如所求。计算这个积分，并利用 $w^2 = 2ad$：

$$W_1 = \frac{1}{a}\left[\frac{(ma+R)v^2}{2} + \frac{Av^4}{4}\right]_0^w = \frac{(ma+R)w^2}{2a} + \frac{Aw^4}{4a} = (ma+R)d + Aad^2.$$

后半段。 此时加速度为 $-a$，所以 $F - (Av^2 + R) = -ma$，得

$$F = Av^2 + R - ma.$$

汽车由 w 减速至 0，其中 $v\frac{\mathrm{d}v}{\mathrm{d}x} = -a$，所以 $\mathrm{d}x = -\frac{v}{a}\,\mathrm{d}v$。在任何一段 $F > 0$、而速率由 v_1 降至 v_2（其中 $v_1 > v_2$）的路程上，引擎所做的功为

$$\int_{v_1}^{v_2} F\left(-\frac{v}{a}\right)\mathrm{d}v = \int_{v_2}^{v_1} \frac{(Av^2 + R - ma)\,v}{a}\,\mathrm{d}v. \qquad (*)$$

第 (i) 部分

若 $R > ma$，则对所有 v 皆有 $F = Av^2 + (R - ma) > 0$，所以引擎在整个后半段都在做功。由 $(*)$，取 $v_1 = w$、$v_2 = 0$：

$$W_2 = \frac{1}{a}\left[\frac{(R-ma)v^2}{2} + \frac{Av^4}{4}\right]_0^w = (R-ma)d + Aad^2.$$

两者相加，$ma\,d$ 项互相抵消，总功为

$$W_1 + W_2 = 2Aad^2 + 2Rd,$$

正如所求。

第 (ii) 部分

现在 $R < ma$，所以 $F = Av^2 - (ma - R)$ 恰好在以下情况等于零：

$$v = u, \qquad \text{其中} u^2 = \frac{ma - R}{A} > 0.$$

若 $u < w$，即 $\frac{ma-R}{A} < 2ad$，这个速率便真的会在后半段出现，而这正正就是条件 $R > ma - 2Aad$。因此，当汽车由 w 减速至 0 时，F 在 $v > u$ 时为正，在速率为 u 时降至零，而在 $v < u$ 时为负（煞车）。这就证明了第一个结论。

> **想一想** 为甚么不能直接沿用 (i) 的答案?

(i) 中的积分把整个后半段的 $F\,\mathrm{d}x$ 加起来。这里 F 在 $v < u$ 时为负，那一段会从总和中减去。但引擎在那里并不是做负功，而是甚么也没做，由煞车接手。所以我们在 $v = u$ 处

停止积分。由 $(*)$，取 $v_1 = w$、$v_2 = u$：

$$W_2 = \frac{1}{a}\left[\frac{(R-ma)v^2}{2} + \frac{Av^4}{4}\right]_u^w$$
$$= (R-ma)d + Aad^2 - \frac{1}{a}\left(\frac{(R-ma)u^2}{2} + \frac{Au^4}{4}\right).$$

代入 $Au^2 = ma - R$，括号内的式子为

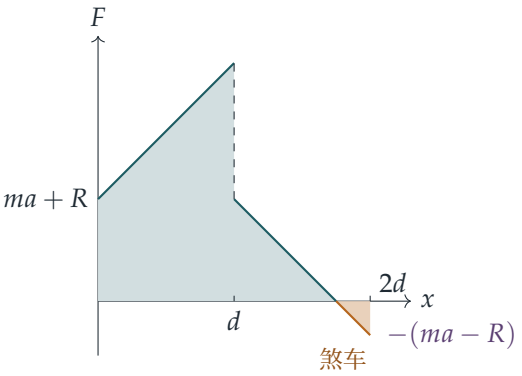
$$-\frac{(ma-R)^2}{2A} + \frac{(ma-R)^2}{4A} = -\frac{(ma-R)^2}{4A},$$

所以 $W_2 = (R-ma)d + Aad^2 + \frac{(ma-R)^2}{4Aa}$。前半段不变（那里 $F > 0$ 恒成立），所以总功为

$$W_1 + W_2 = 2Aad^2 + 2Rd + \frac{(ma-R)^2}{4Aa},$$

正如所求。

图解。由于前半段 $v^2 = 2ax$，后半段 $v^2 = 2a(2d-x)$，F 在每一半都是 x 的线性函数（linear function），而引擎所做的功就是图像正值部分之下的面积。在情况 (ii) 中，图像在接近终点时跌到轴下方。缺去的三角形高为 $ma - R$；由于 F 每单位距离下降 $2Aa$，其底为 $\frac{ma-R}{2Aa}$。它的面积 $\frac{(ma-R)^2}{4Aa}$ 恰好就是那个额外的项。



验算

- 在边界 $R = ma$ 时，额外的项消失，所以 (i) 和 (ii) 的答案一致，这是理所当然的，因为此时 F 只在最后一刻才降至零。
- 在另一边界 $R = ma - 2Aad$ 时，我们有 $u = w$，引擎在后半段完全不做功，总功应为 $W_1 = (ma+R)d + Aad^2$。把 $ma - R = 2Aad$ 代入 (ii) 的答案，得 $2Aad^2 + 2Rd + Aad^2 = 2Rd + 3Aad^2$，而 $(ma+R)d + Aad^2 = 2Rd + 2Aad^2 + Aad^2$ 亦然，所以两者一致。

- 数值上，取 $m = a = 1$、$A = \frac{1}{2}$、$d = 2$、$R = \frac{1}{2}$（即上图所画的情况），(ii) 的公式给出 $4 + 2 + \frac{1}{8} = 6.125$，而直接对整段行程中 F 的正值部分作数值积分，亦得出相同的数值。
- 单位：Aad^2、Rd 和 $(ma - R)^2/(Aa)$ 都是力乘以距离，所以全都是能量。

常见错误

- 忘记后半段有 $\mathrm{d}x = -\frac{v}{a}\,\mathrm{d}v$，这会使积分上下限对调；这里的正负号错误会令后半段的功变成负数。
- 在 (ii) 中把 F 的负值部分也积分进去，结果只会重复得出 (i) 的答案。
- 没有检查速率 u 是否真的在汽车停下之前出现：这正是 R 的下界所保证的。

STEP 之外

力所做的功是一个线积分（line integral）$\int \mathbf{F} \cdot \mathrm{d}\mathbf{r}$，而这里所用的变量代换就是功能定理（work–energy theorem）的一维版本。像 Av^2 这类阻力并不是保守力（conservative force），所以所做的功取决于行程是怎样驾驶的，而不只取决于起点和终点。引擎这种“单向”的特性，让我们初步领略到最优控制（optimal control）问题中出现的约束，例如寻找行驶指定距离最省燃料的方法。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）对功、能量及与速度有关的力有细致的讨论。

详细解答

STEP II 2017，第 11 题

力学 · 抛体与能量

STEP II 2017，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：

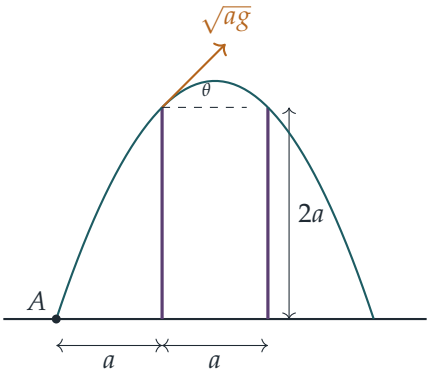
step.maths.org/questions/17-s2-q11

关键思路

当质点升到与第一道墙顶同高时，余下的飞行就是一道全新的抛射体（projectile）问题：它从高度 $2a$ 出发，速率已知，而在水平距离 a 之后，高度仍须至少为 $2a$。利用能量（energy）可以在不知道抛射角的情况下求出这个速率。把轨迹（trajectory）以 $\tan\theta$ 表示，就能把“它能否越过第二道墙?”转化为“一个关于 $\tan\theta$ 的二次方程（quadratic equation）是否有实根?”，这样两个部分便可一并处理。

解答

把每个质点视为只受重力作用、自由运动的质点（point mass），不计空气阻力（air resistance）。取 x 为水平方向（朝向墙的一方），y 为铅直向上。



第 (i) 部分

越过第一道墙时的速率。只有重力做功，所以动能（kinetic energy）与势能（potential energy）之和保持不变。质点从地面升至高度 $2a$，设它在墙顶时的速率为 v，则

$$\tfrac{1}{2}m(5ag) = \tfrac{1}{2}mv^2 + mg(2a) \quad \Longrightarrow \quad v^2 = 5ag - 4ag = ag,$$

因此质点以速率 $\sqrt{ag}$ 越过第一道墙。

角度。设质点越过第一道墙时，速度与水平方向成仰角 θ；由于质点稍后（越过第二道墙时）会再次处于高度 $2a$，它此时必定仍在上升，所以 $\theta > 0$。以第一道墙顶为起点量度：经过水平距离 X 后，质点高于墙顶水平的高度为

$$Y = X\tan\theta - \frac{gX^2}{2v^2\cos^2\theta} = X\tan\theta - \frac{X^2}{2a}\,(1+\tan^2\theta),$$

这里用了 $v^2 = ag$ 及 $\sec^2\theta = 1 + \tan^2\theta$。恰好越过第二道墙，即在 $X = a$ 时 $Y = 0$：

$$a\tan\theta - \tfrac{a}{2}(1+\tan^2\theta) = 0 \iff \tan^2\theta - 2\tan\theta + 1 = 0 \iff (\tan\theta - 1)^2 = 0.$$

所以 $\tan\theta = 1$，$\theta = 45^\circ$，正如所求。（另一种说法：从高度 $2a$ 回到高度 $2a$，相当于在水平地面上的射程（range）为 $v^2\sin 2\theta/g = a\sin 2\theta$，而它只在 $\sin 2\theta = 1$ 时等于 a。）

A 的距离。速度的水平分量始终不变，在第一道墙上方时为 $\sqrt{ag}\cos 45^\circ = \sqrt{ag/2}$；该处的铅直分量同样是 $\sqrt{ag/2}$，方向向上。在 A 点，铅直分量 u 满足 $u^2 = \tfrac{1}{2}ag + 2g(2a) = \tfrac{9}{2}ag$，所以 $u = 3\sqrt{ag/2}$。（验算一下：$\tfrac{1}{2}ag + \tfrac{9}{2}ag = 5ag$，与发射速率吻合。）从 A 到第一道墙所需的时间为

$$t = \frac{u - \sqrt{ag/2}}{g} = \frac{2\sqrt{ag/2}}{g},$$

在这段时间内走过的水平距离为

$$\sqrt{\tfrac{ag}{2}}\cdot\frac{2\sqrt{ag/2}}{g} = \frac{2\cdot\frac{1}{2}ag}{g} = a.$$

因此 A 与第一道墙脚的距离为 a。

第 (ii) 部分

现在质点在高度 $2a + h$ 处越过第一道墙。由能量守恒，它在该处的速率 w 满足

$$w^2 = 5ag - 2g(2a + h) = g(a - 2h).$$

质点朝墙的方向运动，所以水平速度不为零，故 $w^2 > 0$；因此 $h < \tfrac{1}{2}a$。

设它在该处的运动方向与水平方向成角 ϕ，并记 $T = \tan\phi$。以质点在第一道墙上方的位置为起点量度，经过水平距离 X 后，它高于该点的高度为

$$Y = XT - \frac{gX^2}{2w^2}(1+T^2) = XT - kX^2(1+T^2)/a^2, \qquad \text{其中} k = \frac{ga^2}{2w^2} = \frac{a^2}{2(a-2h)}.$$

在第二道墙处 $X = a$，所以 $Y = aT - k(1+T^2)$。要越过这道墙，质点的高度须至少为 $2a$，即 $Y \geq -h$：

$$aT - k(1+T^2) + h \geq 0 \iff kT^2 - aT + (k-h) \leq 0.$$

? **想一想** 我们不知道 ϕ 的值。怎样才能证明没有任何方向可行?

由于 $k > 0$，二次式 $kT^2 - aT + (k-h)$ 的图像开口向上。它只有在判别式（discriminant）非负时，才会对某个实数 T 取得 ≤ 0 的值。这个判别式为

$$\Delta = a^2 - 4k(k-h) = a^2 - \frac{a^4}{(a-2h)^2} + \frac{2a^2h}{a-2h}.$$

乘以正数 $(a-2h)^2/a^2$，得

$$\frac{(a-2h)^2}{a^2}\Delta = (a-2h)^2 - a^2 + 2h(a-2h) = -2ah.$$

由于 $a > 0$ 且 $h > 0$，此式为负，所以 $\Delta < 0$。于是对所有实数 T 都有 $kT^2 - aT + (k-h) > 0$；换言之，无论质点越过第一道墙时的方向如何，它到达第二道墙时的高度都低于 $2a$。它无法越过第二道墙。

留意 $h = 0$ 时 $\Delta = 0$，并得到重根（double root）$T = a/(2k) = 1$：这正是第 (i) 部分的 45°。

验算

- 在第 (i) 部分，A 点的速度分量为 $\sqrt{ag/2}$ 和 $3\sqrt{ag/2}$，两者平方之和为 $5ag$。取 $a = 1.7\,\mathrm{m}$ 及 $g = 9.81\,\mathrm{m\,s^{-2}}$，从 A 点起逐步向前推算运动，得出质点在两道墙处的高度都是 $2.000a$。
- 轨迹关于最高点对称；在第 (i) 部分，最高点位于两墙正中，与 A 的水平距离为 $1.5a$。所以质点落在距 A 为 $3a$ 之处，即第二道墙之后距离 a 的地方：整幅图是对称的。
- 在第 (ii) 部分，用电脑以速率 $\sqrt{5ag}$ 检验了数以千计随机的发射角和发射点，没有一个既能在第一道墙上方经过，又能越过第二道墙。

常见错误

- 忘记第 (i) 部分的 θ 必须是越过第一道墙时的角度，而不是在 A 点的抛射角（后者满足 $\tan\alpha = 3$）。
- 在第 (ii) 部分只检验某一个特定角度（例如 45°）。题目的断言涉及所有可能的方向，所以才需要用判别式。

STEP 之外

第 (ii) 部分说明：从第一道墙上方的一点出发，具有这个能量的质点只能恰好到达第二道墙顶；换言之，该墙顶位于可到达范围的边界上。从某一给定点以给定速率出发可到达的所有点，都位于一条称为包络线（envelope）或安全抛物线（safety parabola）的抛物线之内，而抛物线上的点恰好只由一条轨迹到达，这就是这里所见的重根。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）以大学一年级程度讲解抛射体及能量方法。

详细解答

STEP II 2017，第 12 题

概率与统计 • 泊松过程

STEP II 2017，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/17-s2-q12

关键思路

两条互相独立（independent）的泊松（Poisson）鱼流合并后，成为一条单一的泊松鱼流，其速率为两个速率之和；这正是第 (i) 部分中的卷积（convolution）所显示的。至于有关时间的问题，要把关于第一条鱼何时上钓的陈述，转化为关于钓到多少条鱼的陈述：“亚当的第一条鱼在时间 t 之后才上钓”的意思恰恰就是“亚当在 $[0,t]$ 内没有钓到任何鱼”，而泊松公式可以直接给出这个概率。

解答

全文中，X 和 Y 分别是亚当和夏娃在长度为 T 的时段内钓到的鱼的数目，因此 $X \sim \mathrm{Po}(\lambda T)$ 及 $Y \sim \mathrm{Po}(\mu T)$，两者互相独立。

第 (i) 部分

对整数 $r \geq 0$，事件 $X+Y=r$ 可分拆为互斥事件 $\{X=j,\ Y=r-j\}$，其中 $j=0,1,\ldots,r$。利用独立性，

$$\begin{aligned}\mathrm{P}(X+Y=r) &= \sum_{j=0}^{r} \mathrm{P}(X=j)\,\mathrm{P}(Y=r-j) = \sum_{j=0}^{r} \frac{e^{-\lambda T}(\lambda T)^j}{j!} \cdot \frac{e^{-\mu T}(\mu T)^{r-j}}{(r-j)!} \\ &= \frac{e^{-(\lambda+\mu)T}}{r!} \sum_{j=0}^{r} \binom{r}{j} (\lambda T)^j (\mu T)^{r-j} = \frac{e^{-(\lambda+\mu)T}\big((\lambda+\mu)T\big)^r}{r!},\end{aligned}$$

这里我们先乘以再除以 $r!$，然后运用二项式定理（binomial theorem）。这正是 $\mathrm{Po}\big((\lambda+\mu)T\big)$ 的概率，因此在时间 T 内的总渔获服从参数为 $(\lambda+\mu)T$ 的泊松分布。

第 (ii) 部分

对 $0 \le r \le k$,

$$\mathrm{P}(X = r \mid X+Y = k) = \frac{\mathrm{P}(X=r,\ Y=k-r)}{\mathrm{P}(X+Y=k)} = \frac{\dfrac{e^{-\lambda T}(\lambda T)^r}{r!} \cdot \dfrac{e^{-\mu T}(\mu T)^{k-r}}{(k-r)!}}{\dfrac{e^{-(\lambda+\mu)T}\big((\lambda+\mu)T\big)^k}{k!}},$$

其中分母用了第 (i) 部分的结果。指数项互相抵消，T 的幂也一样（分子有 $r+(k-r)=k$ 次，分母也有 k 次），剩下

$$\mathrm{P}(X = r \mid X+Y = k) = \binom{k}{r}\left(\frac{\lambda}{\lambda+\mu}\right)^r\left(\frac{\mu}{\lambda+\mu}\right)^{k-r}.$$

所以，在总数为 k 的条件下，亚当的渔获服从 $\mathrm{B}\big(k,\ \frac{\lambda}{\lambda+\mu}\big)$。实际上，这 k 条鱼中的每一条，都各自独立地以概率 $\frac{\lambda}{\lambda+\mu}$ 属于亚当。留意 T 已经消失了。

第 (iii) 部分

设 A 和 E 分别为亚当和夏娃钓到第一条鱼的时间，从开始时起计。亚当的第一条鱼在时间 t 之后才上钓，当且仅当他在 $[0,t]$ 内甚么也没钓到，因此

$$\mathrm{P}(A > t) = \mathrm{P}\big(\mathrm{Po}(\lambda t) = 0\big) = e^{-\lambda t}, \qquad \text{同理} \qquad \mathrm{P}(E > t) = e^{-\mu t}.$$

因此 A 的概率密度函数（probability density function）为 $\lambda e^{-\lambda t}$，其中 $t \ge 0$（把 $1-e^{-\lambda t}$ 求导即得），而由于两人的渔获互相独立，A 和 E 也互相独立。

想一想 我们该如何把 A 的密度与 E 的概率结合起来?

以 A 的值为条件：若亚当的第一条鱼在时间 t 上钓，那么他领先恰恰就是 $E > t$ 的情况，由独立性可知其概率为 $e^{-\mu t}$。对所有 t 求和，

$$\mathrm{P}(A < E) = \int_0^\infty \lambda e^{-\lambda t}\, e^{-\mu t}\,\mathrm{d}t = \lambda\left[\frac{-e^{-(\lambda+\mu)t}}{\lambda+\mu}\right]_0^\infty = \frac{\lambda}{\lambda+\mu}.$$

（由于 E 服从连续分布，两人同时钓到的概率为零。）所以第一条鱼属于亚当的概率是 $\dfrac{\lambda}{\lambda+\mu}$，与第 (ii) 部分的二项分布参数相同。

第 (iv) 部分

我们要求 $\mathrm{E}(M)$，其中 $M = \max(A,E)$ 是两人都钓到鱼的时间。记 $m = \min(A,E)$，即两人之中任何一人钓到第一条鱼的时间。对任意两个数，max + min 等于它们的和，因此

$$M = A + E - m \qquad \Longrightarrow \qquad \mathrm{E}(M) = \mathrm{E}(A) + \mathrm{E}(E) - \mathrm{E}(m).$$

$\mathrm{E}(A)$ 与 $\mathrm{E}(E)$。从开始时起计，亚当的渔获所构成的过程，与从任何一次钓到鱼的时刻起计的过程属同一类，因此他等待第一条鱼的期望（expectation）时间，与题目给出的期望间隔相同，即 $\mathrm{E}(A) = \lambda^{-1}$；同样 $\mathrm{E}(E) = \mu^{-1}$。（直接计算：利用分部积分（integration by parts），$\mathrm{E}(A) = \int_0^\infty t\lambda e^{-\lambda t}\,\mathrm{d}t = \lambda^{-1}$。）

> **想一想** “两人之中任何一人钓到的第一条鱼”属于哪一类过程?

由第 (i) 部分，在任何长度为 t 的时段内，两人的合计渔获服从 $\mathrm{Po}\big((\lambda+\mu)t\big)$。所以合计渔获与亚当的渔获是完全同类的过程，只是速率由 λ 换成 $\lambda+\mu$，因此等待第一条鱼的期望时间为 $\mathrm{E}(m) = (\lambda+\mu)^{-1}$。（直接计算：$\mathrm{P}(m > t) = \mathrm{P}(A > t)\,\mathrm{P}(E > t) = e^{-(\lambda+\mu)t}$。）因此

$$\mathrm{E}(M) = \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu} = \frac{\lambda^2+\lambda\mu+\mu^2}{\lambda\mu(\lambda+\mu)}.$$

验算

- 若 $\lambda = \mu$，第 (iii) 部分给出 $\frac{1}{2}$，正如对称性（symmetry）所要求；第 (iv) 部分给出 $\frac{3}{2\lambda}$：先等 $\frac{1}{2\lambda}$ 得到第一条鱼，再根据同样的“重新开始”推理，另一人还要多等 $\frac{1}{\lambda}$。
- 若 $\mu \to \infty$（夏娃即时钓到鱼），第 (iv) 部分的答案趋向 $\frac{1}{\lambda}$，即亚当独自所需的时间。
- 第 (iv) 部分的答案至少为 $\max(\frac{1}{\lambda}, \frac{1}{\mu})$，这是必然的，因为 $\frac{1}{\mu} > \frac{1}{\lambda+\mu}$。
- 取 $\lambda = 1.3$、$\mu = 0.7$，公式给出 0.65 和 1.6978；以电脑模拟 2×10^5 次钓鱼之旅，得到 0.651 和 1.700。

常见错误

- 在第 (iv) 部分，把 $\frac{1}{\lambda} + \frac{1}{\mu}$ 相加，或取 $\max\big(\frac{1}{\lambda}, \frac{1}{\mu}\big)$：最大值的期望并非两者之一。
- 在第 (ii) 部分，忘记除以 $\mathrm{P}(X+Y=k)$，而正是这一步把联合概率（joint probability）变成条件概率（conditional probability）。

STEP 之外

你在这里遇到的是泊松过程（Poisson process）：事件之间的等待时间服从指数分布，并具有无记忆性（memoryless）；互相独立的泊松过程叠加（superpose）后成为另一个泊松过程；而每个事件都独立地以与其速率成正比的概率被标记（反方向的操作，即“稀疏化”（thinning），同样成立）。这些事实是排队论（queueing theory）和连续时间马可夫链（continuous-time Markov chains）的基础。Grimmett 与 Welsh 合著的 *Probability:*

An Introduction（《概率导论》）对泊松过程和指数分布（exponential distribution）的处理，正好可以作为这道题目的自然延续。

详细解答

STEP II 2017，第 13 题

概率与统计 · 期望值与无穷级数

STEP II 2017，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/17-s2-q13

关键思路

无论哪一种情况，门在第 k 次尝试时打开，当且仅当前 $k-1$ 次尝试都失败而第 k 次成功，所以所求概率是若干个“失败”因子乘以一个“成功”因子。各种情况之间的分别，只在于每次抽取时袋中有多少条钥匙。把概率写出来之后，期望值（expectation）就是一些求和：(i) 是一个经微分的等比级数（geometric series），(ii) 是一个等差级数（arithmetic series），而 (iii) 则是一个裂项相消（telescoping）的和，最后剩下调和级数（harmonic series）的一部分。

解答

全文中，设 X 为门被打开时的尝试次数，并设 $q = 1 - \frac{1}{n}$。

第 (i) 部分

每次都从装有全部 n 条钥匙的袋中抽取；由于钥匙会放回，每次抽取与之前的抽取互相独立（independent）。因此每次尝试成功的概率都是 $\frac{1}{n}$，而

$$\mathrm{P}(X = k) = q^{k-1} \cdot \frac{1}{n}, \qquad k = 1, 2, 3, \ldots$$

由二项展开（binomial expansion），因 $0 < q < 1$ 而成立，

$$(1-q)^{-2} = 1 + 2q + 3q^2 + \cdots = \sum_{k=1}^{\infty} k q^{k-1}.$$

因此

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} k\, q^{k-1} \frac{1}{n} = \frac{1}{n}(1-q)^{-2} = \frac{1}{n} \cdot n^2 = n.$$

尝试次数的期望值是 n。

第 (ii) 部分

现在每次失败后，袋中就少了一条错误的钥匙，所以第 j 次尝试时袋中有 $n-j+1$ 条钥匙。要门在第 k 次尝试时打开，需要先失败 $k-1$ 次，然后成功一次：

$$\mathrm{P}(X=k)=\frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\cdot\frac{1}{n-k+1}=\frac{1}{n},$$

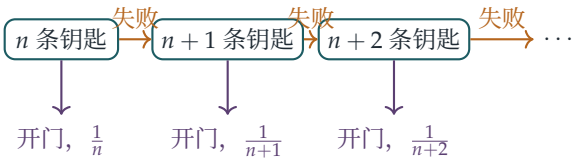
因为这个乘积会逐项相消。（这一点也可以直接看出：逐条不放回地抽出钥匙，正确的钥匙在 n 个位置中的位置是均匀随机的。）所以 X 在 $\{1,2,\ldots,n\}$ 上服从均匀分布（uniform distribution），而

$$\mathrm{E}(X)=\frac{1}{n}\sum_{k=1}^{n}k=\frac{1}{n}\cdot\frac{n(n+1)}{2}=\frac{n+1}{2}.$$

第 (iii) 部分

概率。 袋中起初有 n 条钥匙，每次失败后就多加一条错误的钥匙，而袋中始终恰好有一条正确的钥匙。所以第 j 次尝试时袋中有 $n+j-1$ 条钥匙，而

$$\mathrm{P}(\text{第}j\text{ 次尝试失败}\mid\text{之前的尝试均失败})=\frac{n+j-2}{n+j-1}.$$



把 $k-1$ 次失败及其后一次成功的条件概率（conditional probability）相乘，

$$\begin{aligned}\mathrm{P}(X=k)&=\frac{n-1}{n}\cdot\frac{n}{n+1}\cdots\frac{n+k-3}{n+k-2}\cdot\frac{1}{n+k-1}\\&=\frac{n-1}{n+k-2}\cdot\frac{1}{n+k-1}=\frac{n-1}{(n+k-1)(n+k-2)},\end{aligned}$$

因为每个分子都与前一个分母相消，只剩下第一个分子 $n-1$ 和最后一个失败因子的分母 $n+k-2$。当 $k=1$ 时没有失败因子，而公式给出 $\frac{n-1}{n(n-1)}=\frac{1}{n}$，与预期相符。这就是所求的结果。

门一定会打开。 在计算平均值之前,值得先验证 X 以概率 1 为有限。由于 $\dfrac{n-1}{(n+k-1)(n+k-2)} = (n-1)\left(\dfrac{1}{n+k-2} - \dfrac{1}{n+k-1}\right)$，这个和会裂项相消：

$$\sum_{k=1}^{N} \mathrm{P}(X=k) = (n-1)\left(\frac{1}{n-1} - \frac{1}{n+N-1}\right) \to 1 \quad \text{当} N \to \infty.$$

期望值。 我们需要求 $\sum k\,\mathrm{P}(X=k)$,所以把 $\dfrac{k}{(n+k-1)(n+k-2)}$ 拆成部分分式(partial fractions)。设

$$k = A(n+k-2) + B(n+k-1)$$

代入 $k = 1-n$ 得 $1-n = -A$，所以 $A = n-1$；代入 $k = 2-n$ 得 $B = 2-n$。因此

$$k\,\mathrm{P}(X=k) = (n-1)\left(\frac{n-1}{n+k-1} - \frac{n-2}{n+k-2}\right).$$

想一想 这个式子不能干净地裂项相消，因为系数 $n-1$ 和 $n-2$ 并不相同。那么剩下来的是甚么?

设 S_N 为首 N 项之和。第一个分数的分母取遍 $m = n, \ldots, n+N-1$；第二个分数的分母则取遍 $m = n-1, \ldots, n+N-2$。在共同范围 $m = n, \ldots, n+N-2$ 上，每个 $\frac{1}{m}$ 的系数都是 $(n-1)-(n-2) = 1$，而两端则剩了下来：

$$S_N = (n-1)\left(\sum_{m=n}^{n+N-2} \frac{1}{m} + \frac{n-1}{n+N-1} - \frac{n-2}{n-1}\right).$$

最后两项是有界的（中间一项为正，最后一项是一个固定的数），而

$$\sum_{m=n}^{n+N-2} \frac{1}{m} = \sum_{m=1}^{n+N-2} \frac{1}{m} - \sum_{m=1}^{n-1} \frac{1}{m} \to \infty$$

当 $N \to \infty$ 时成立，因为调和级数发散（diverges），而我们只减去一个固定的有限值。由于 $n-1 > 0$，所以 $S_N \to \infty$。所有项 $k\,\mathrm{P}(X=k)$ 都是正数，因此定义 $\mathrm{E}(X)$ 的级数发散至无穷：尝试次数的期望值是无穷大，尽管门最终必定会打开。

验算

- 当 $n = 2$ 时，三个答案分别是 2、$\frac{3}{2}$ 和 ∞。第二种情况效率最高，这是理所当然的，因为它从不重试同一条钥匙；第一种情况则介乎两者之间。
- 在 (iii) 中取 $n = 3$，公式对 $k = 1, 2, 3$ 给出 $\frac{1}{3}, \frac{1}{6}, \frac{1}{10}$；直接计算则得 $\frac{1}{3}$、$\frac{2}{3}\cdot\frac{1}{4} = \frac{1}{6}$ 和 $\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{5} = \frac{1}{10}$。以电脑模拟 2×10^5 局，得到的频率为 0.334、0.168 和 0.099。

- 同样取 $n = 3$，当 $N = 10^4$ 时部分和 S_N 约为 15.6，而且它会像 $2\ln N$ 般不断增长：并没有极限。

常见错误

- 在 (iii) 中，把第 k 次尝试时袋中钥匙的数目当作 $n + k$ 而不是 $n + k - 1$：在第 k 次抽取之前，只加入了 $k - 1$ 条钥匙。
- 因为平均值是无穷大，就断定门可能永远打不开。各概率之和为 1；只是其尾部像 $1/k^2$ 般衰减，太“重”了，令 $\sum k\,\mathrm{P}(X = k)$ 无法收敛。
- 把部分分式当作两个系数相等般裂项相消，因而得出一个有限的答案。

STEP 之外

第三种情况给出一个以概率 1 为有限、但平均值为无穷大的随机变量（random variable），原因是 $\mathrm{P}(X = k)$ 的表现像 $1/k^2$。像这样的重尾分布（heavy-tailed distribution），以及著名的圣彼得堡悖论（St Petersburg paradox），说明了为何大学一年级的概率课程坚持：只有当 $\sum k\,\mathrm{P}(X = k)$ 收敛时，$\mathrm{E}(X)$ 才存在。每次失败后多加一条错误钥匙的袋子，与波利亚（Pólya）罐子模型是近亲；后者是描述“富者愈富”过程的标准模型。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（《概率导论》）讲述了离散随机变量的期望值以及这些微妙之处。

详细解答

STEP III 2017，第 1 题

纯数学 · 裂项相消级数的界限

STEP III 2017，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/17-s3-q1

关键思路

第 (i) 部分的恒等式（identity）把每个二项式系数（binomial coefficient）的倒数写成一个常数乘以相邻两项的*差*，于是整个和式可以伸缩相消（telescoping）。在第 (ii) 部分，$1/n^3$ 不容易求和，但二项式系数的倒数却可以：$\binom{n+1}{3}$ 和 $\binom{n+2}{5}$ 都是 n 的多项式，分别与 n^3 和 n^5 的倍数相近，所以我们可以把 $1/n^3$ 夹在两个能用第 (i) 部分精确求和的式子之间。

解答

第 (i) 部分

记 $\binom{m}{k} = \dfrac{m!}{k!\,(m-k)!}$。于是

$$\begin{aligned}\frac{1}{\binom{n+r-1}{r}} - \frac{1}{\binom{n+r}{r}} &= \frac{r!\,(n-1)!}{(n+r-1)!} - \frac{r!\,n!}{(n+r)!} \\ &= \frac{r!\,(n-1)!}{(n+r)!}\big[(n+r)-n\big] = \frac{r\cdot r!\,(n-1)!}{(n+r)!},\end{aligned}$$

其中我们利用 $(n+r)! = (n+r)\,(n+r-1)!$ 和 $n! = n\,(n-1)!$，提出了公因子 $r!\,(n-1)!/(n+r)!$。乘以 $\frac{r+1}{r}$，得

$$\frac{r+1}{r}\left(\frac{1}{\binom{n+r-1}{r}} - \frac{1}{\binom{n+r}{r}}\right) = \frac{(r+1)!\,(n-1)!}{(n+r)!} = \frac{1}{\binom{n+r}{r+1}},$$

因为 $(n+r)-(r+1) = n-1$。这就是所求的恒等式。

求和。设 $a_n = 1/\binom{n+r-1}{r}$，则恒等式可写成 $1/\binom{n+r}{r+1} = \frac{r+1}{r}(a_n - a_{n+1})$。由 $n=1$ 加到 N，

除了首项和末项外，其余全部相消：

$$\sum_{n=1}^{N}\frac{1}{\binom{n+r}{r+1}}=\frac{r+1}{r}(a_1-a_{N+1})=\frac{r+1}{r}\left(1-\frac{1}{\binom{N+r}{r}}\right),$$

因为 $a_1=1/\binom{r}{r}=1$。当 $N\to\infty$ 时，$\binom{N+r}{r}\geq N+r\to\infty$（对 $r\geq 1$ 而言），所以

$$\sum_{n=1}^{\infty}\frac{1}{\binom{n+r}{r+1}}=\frac{r+1}{r}.$$

推论。取 $r=2$，由 $n=1$ 起的和为 $\frac{3}{2}$。$n=1$ 那一项是 $1/\binom{3}{3}=1$，所以

$$\sum_{n=2}^{\infty}\frac{1}{\binom{n+2}{3}}=\frac{3}{2}-1=\frac{1}{2}.$$

第 (ii) 部分

第一条不等式。对 $n\geq 3$，

$$\frac{1}{\binom{n+1}{3}}=\frac{3!}{(n+1)n(n-1)}=\frac{3!}{n^3-n}>\frac{3!}{n^3},$$

因为 $0<n^3-n<n^3$（左边的不等式成立是因为 $n\geq 2$）。

第二条不等式。我们有 $\binom{n+1}{3}=\frac{1}{6}n(n^2-1)$ 及 $\binom{n+2}{5}=\frac{1}{120}(n+2)(n+1)n(n-1)(n-2)=\frac{1}{120}n(n^2-1)(n^2-4)$。所以

$$\frac{20}{\binom{n+1}{3}}-\frac{1}{\binom{n+2}{5}}=\frac{120}{n(n^2-1)}-\frac{120}{n(n^2-1)(n^2-4)}=\frac{120\,(n^2-5)}{n(n^2-1)(n^2-4)}.$$

当 $n\geq 3$ 时，因子 n、n^2-1、n^2-4 全部为正，所以此式小于 $120/n^3=5!/n^3$ 当且仅当

$$n^2(n^2-5)<(n^2-1)(n^2-4),\quad \text{即}\quad n^4-5n^2<n^4-5n^2+4,$$

而这显然成立。

> ? **想一想** 我们需要第 (i) 部分中的哪些和式？它们由哪一项开始？

我们需要 $\sum_{n\geq 3}1/\binom{n+1}{3}$ 和 $\sum_{n\geq 3}1/\binom{n+2}{5}$。在第一个和式中设 $m=n-1$，它变成 $\sum_{m\geq 2}1/\binom{m+2}{3}=\frac{1}{2}$，正正就是第 (i) 部分的推论。在第二个和式中设 $m=n-2$，它变成 $\sum_{m\geq 1}1/\binom{m+4}{5}$，即第 (i) 部分取 $r=4$ 的情况，得 $\frac{5}{4}$。

所涉及的级数全部收敛（convergent）（它们都是正项级数，并且如上述不等式所示，被收敛级数所界）。把收敛级数之间的严格不等式逐项相加，得到的仍是严格不等式，因为各个正差之和至少等于其首项，而首项是正数。记 $T=\sum_{n\geq 3}1/n^3$。

上界。把第一条不等式对 $n \geq 3$ 求和：

$$6T < \sum_{n=3}^{\infty} \frac{1}{\binom{n+1}{3}} = \frac{1}{2}, \qquad \text{所以} \qquad T < \frac{1}{12}.$$

下界。把第二条不等式对 $n \geq 3$ 求和：

$$120\,T > 20 \cdot \frac{1}{2} - \frac{5}{4} = \frac{35}{4}, \qquad \text{所以} \qquad T > \frac{35}{480} = \frac{7}{96}.$$

最后，$\sum_{n\geq 1} 1/n^3 = 1 + \frac{1}{8} + T$，而 $1 + \frac{1}{8} = \frac{108}{96}$。因此

$$\frac{108+7}{96} = \frac{115}{96} < \sum_{n=1}^{\infty} \frac{1}{n^3} < \frac{108}{96} + \frac{8}{96} = \frac{116}{96},$$

证毕。

验算

- 以 $n = r = 1$ 测试恒等式：$1/\binom{2}{2} = 1$，而 $2\,(1/\binom{1}{1} - 1/\binom{2}{1}) = 2 \cdot \frac{1}{2} = 1$。
- 当 $r = 1$ 时，和为 $\sum 1/\binom{n+1}{2} = \sum \frac{2}{n(n+1)} = 2$，这是大家熟悉的伸缩和，与 $\frac{r+1}{r} = 2$ 吻合。
- 数值上 $\sum 1/n^3 \approx 1.20206$，而 $\frac{115}{96} \approx 1.19792$、$\frac{116}{96} \approx 1.20833$，所以上下界成立，并把该值确定在约 0.01 的范围之内。

常见错误

- 移位后弄错起始指标：$\sum_{n\geq 3} 1/\binom{n+1}{3}$ 由 $\binom{4}{3}$ 开始，这是第 (i) 部分推论中 $n = 2$ 的那一项，而不是 $n = 1$ 的那一项。
- 忘记加上 $n = 1, 2$ 时的项 $1 + \frac{1}{8}$，这两项并不在不等式涵盖的范围内。

STEP 之外

数 $\sum 1/n^3$ 就是 $\zeta(3)$，称为阿佩里常数（Apéry's constant）；与 $\sum 1/n^2 = \pi^2/6$ 不同，它没有已知的闭合形式。阿佩里（Apéry）在 1978 年证明它是无理数（irrational），所用的正是与这里相似的快速收敛级数。把一个级数与伸缩级数比较，以及把不等式逐项相加，都是大学第一个分析（analysis）课程的标准工具。如想仔细了解收敛性和比较判别法（comparison test），可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2017，第 2 题

纯数学 · 复平面上的旋转

STEP III 2017，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/17-s3-q2

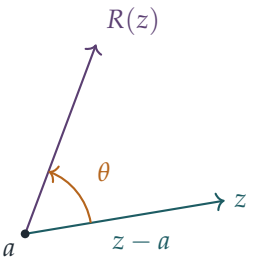
关键思路

要描述绕 a 的旋转（rotation），最简单的做法是把原点移到 a：先减去 a，再乘以 $\mathrm{e}^{\mathrm{i}\theta}$，然后把 a 加回去。因此每个旋转都具有 $z \mapsto \mathrm{e}^{\mathrm{i}\alpha}z + k$ 的形式；反过来，只要 $\mathrm{e}^{\mathrm{i}\alpha} \neq 1$，这样的映射就是绕其不动点（fixed point）转 α 的旋转。所以要辨认 SR，只需把两条公式复合（compose），再求出不动点。利用半角因式分解（half-angle factorisation）$1 - \mathrm{e}^{\mathrm{i}\alpha} = -2\mathrm{i}\,\mathrm{e}^{\mathrm{i}\alpha/2}\sin\frac{1}{2}\alpha$，便可把答案化成所要求的形式。

解答

第 (i) 部分

由中心 a 指向 P 的向量（vector）是 $z - a$。绕 a 逆时针旋转 θ，就是把这个向量转 θ，在复数中即乘以 $\mathrm{e}^{\mathrm{i}\theta}$；由于 $|\mathrm{e}^{\mathrm{i}\theta}| = 1$，长度保持不变。



所以 P 的像 w 满足 $w - a = \mathrm{e}^{\mathrm{i}\theta}(z - a)$，即

$$w = \mathrm{e}^{\mathrm{i}\theta}z + a\left(1 - \mathrm{e}^{\mathrm{i}\theta}\right),$$

正如所求。同理，$S(z) = \mathrm{e}^{\mathrm{i}\phi}z + b\big(1 - \mathrm{e}^{\mathrm{i}\phi}\big)$。

第 (ii) 部分

先作用 R，再作用 S：

$$\begin{aligned} SR(z) &= \mathrm{e}^{\mathrm{i}\phi}\Big(\mathrm{e}^{\mathrm{i}\theta}z + a\big(1 - \mathrm{e}^{\mathrm{i}\theta}\big)\Big) + b\big(1 - \mathrm{e}^{\mathrm{i}\phi}\big) \\ &= \mathrm{e}^{\mathrm{i}(\theta+\phi)}z + a\,\mathrm{e}^{\mathrm{i}\phi}\big(1 - \mathrm{e}^{\mathrm{i}\theta}\big) + b\big(1 - \mathrm{e}^{\mathrm{i}\phi}\big). \end{aligned}$$

由第 (i) 部分，绕 c 转 $\theta + \phi$ 的旋转是 $z \mapsto \mathrm{e}^{\mathrm{i}(\theta+\phi)}z + c\big(1 - \mathrm{e}^{\mathrm{i}(\theta+\phi)}\big)$。两者 z 的系数已经相同，所以 SR 恰好是这个旋转，当且仅当

$$c\big(1 - \mathrm{e}^{\mathrm{i}(\theta+\phi)}\big) = a\,\mathrm{e}^{\mathrm{i}\phi}\big(1 - \mathrm{e}^{\mathrm{i}\theta}\big) + b\big(1 - \mathrm{e}^{\mathrm{i}\phi}\big). \qquad (*)$$

? **想一想** 怎样才能把括号 $1 - \mathrm{e}^{\mathrm{i}\alpha}$ 化成半角的正弦?

提取因子 $\mathrm{e}^{\mathrm{i}\alpha/2}$，使括号内变得对称：

$$1 - \mathrm{e}^{\mathrm{i}\alpha} = \mathrm{e}^{\mathrm{i}\alpha/2}\big(\mathrm{e}^{-\mathrm{i}\alpha/2} - \mathrm{e}^{\mathrm{i}\alpha/2}\big) = -2\mathrm{i}\,\mathrm{e}^{\mathrm{i}\alpha/2}\sin\tfrac{1}{2}\alpha.$$

分别取 $\alpha = \theta + \phi$、θ 和 ϕ 代入，方程 $(*)$ 变成

$$-2\mathrm{i}\,c\,\mathrm{e}^{\mathrm{i}(\theta+\phi)/2}\sin\tfrac{1}{2}(\theta+\phi) = -2\mathrm{i}\,a\,\mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{\mathrm{i}\theta/2}\sin\tfrac{1}{2}\theta - 2\mathrm{i}\,b\,\mathrm{e}^{\mathrm{i}\phi/2}\sin\tfrac{1}{2}\phi.$$

两边除以 $-2\mathrm{i}\,\mathrm{e}^{\mathrm{i}(\theta+\phi)/2}$（它永不为零），得

$$c\sin\tfrac{1}{2}(\theta+\phi) = a\,\mathrm{e}^{\mathrm{i}\phi/2}\sin\tfrac{1}{2}\theta + b\,\mathrm{e}^{-\mathrm{i}\theta/2}\sin\tfrac{1}{2}\phi.$$

若 $\theta + \phi \neq 2n\pi$，则 $\sin\frac{1}{2}(\theta+\phi) \neq 0$，因此上式唯一确定 c，而取这个 c 时，映射 SR 就是绕 c 转 $\theta + \phi$ 的旋转。这正是所要求的结果。（这个条件是必要的：若 $\sin\frac{1}{2}(\theta+\phi) = 0$，一般而言根本没有可供旋转的不动点。）

$\theta + \phi = 2\pi$ 的情况。此时 $\mathrm{e}^{\mathrm{i}(\theta+\phi)} = 1$ 且 $\mathrm{e}^{\mathrm{i}\phi} = \mathrm{e}^{-\mathrm{i}\theta}$，所以

$$\begin{aligned} SR(z) &= z + a\,\mathrm{e}^{-\mathrm{i}\theta}\big(1 - \mathrm{e}^{\mathrm{i}\theta}\big) + b\big(1 - \mathrm{e}^{-\mathrm{i}\theta}\big) \\ &= z + a\big(\mathrm{e}^{-\mathrm{i}\theta} - 1\big) + b\big(1 - \mathrm{e}^{-\mathrm{i}\theta}\big) \\ &= z + (b-a)\big(1 - \mathrm{e}^{-\mathrm{i}\theta}\big). \end{aligned}$$

所以 SR 是平移（*translation*）$(b-a)\big(1 - \mathrm{e}^{-\mathrm{i}\theta}\big)$，平移距离为 $2|b-a|\,|\sin\frac{1}{2}\theta|$。当 $a = b$（绕同一点先转 θ 再转 $2\pi - \theta$），或当 θ 是 2π 的倍数时，它是恒等映射（identity）；否则它是一个真正的平移，没有不动点。

第 (iii) 部分

在 SR 的公式中把 (a,θ) 与 (b,ϕ) 的角色互换，得

$$RS(z) = \mathrm{e}^{\mathrm{i}(\theta+\phi)}z + b\,\mathrm{e}^{\mathrm{i}\theta}\left(1-\mathrm{e}^{\mathrm{i}\phi}\right) + a\left(1-\mathrm{e}^{\mathrm{i}\theta}\right).$$

两个映射都把 z 乘以 $\mathrm{e}^{\mathrm{i}(\theta+\phi)}$，所以 $RS = SR$ 当且仅当常数项相同。两者之差为

$$\begin{aligned} SR(z) - RS(z) &= a\left(1-\mathrm{e}^{\mathrm{i}\theta}\right)\left(\mathrm{e}^{\mathrm{i}\phi}-1\right) + b\left(1-\mathrm{e}^{\mathrm{i}\phi}\right)\left(1-\mathrm{e}^{\mathrm{i}\theta}\right) \\ &= (b-a)\left(1-\mathrm{e}^{\mathrm{i}\theta}\right)\left(1-\mathrm{e}^{\mathrm{i}\phi}\right). \end{aligned}$$

乘积为零当且仅当其中一个因子为零，所以 $RS = SR$ 当且仅当

- $a = b$（两个旋转的中心相同），或
- $\mathrm{e}^{\mathrm{i}\theta} = 1$，即 θ 是 2π 的倍数，所以 R 是恒等映射，或
- $\mathrm{e}^{\mathrm{i}\phi} = 1$，即 ϕ 是 2π 的倍数，所以 S 是恒等映射。

换句话说：两个真正的旋转只有在中心相同时才可交换（commute）。

验算

- 取 $a = b$，(ii) 的公式给出 $c\sin\frac{1}{2}(\theta+\phi) = a\left(\mathrm{e}^{\mathrm{i}\phi/2}\sin\frac{1}{2}\theta + \mathrm{e}^{-\mathrm{i}\theta/2}\sin\frac{1}{2}\phi\right)$，把括号展开得 $\sin\frac{1}{2}\theta\cos\frac{1}{2}\phi + \cos\frac{1}{2}\theta\sin\frac{1}{2}\phi = \sin\frac{1}{2}(\theta+\phi)$（虚部互相抵消），所以 $c = a$，理应如此。
- 取 $\phi = 0$（S 为恒等映射），则 $c\sin\frac{1}{2}\theta = a\sin\frac{1}{2}\theta$，所以 $c = a$。
- 取 $a = 0$、$b = 1$、$\theta = \phi = \frac{1}{2}\pi$。则 $R(z) = \mathrm{i}z$, $S(z) = \mathrm{i}z + 1 - \mathrm{i}$, 所以 $SR(z) = -z + 1 - \mathrm{i}$, 即绕 $\frac{1}{2}(1-\mathrm{i})$ 的半周旋转。公式给出 $c\cdot 1 = \mathrm{e}^{-\mathrm{i}\pi/4}\cdot\frac{1}{\sqrt{2}} = \frac{1}{2}(1-\mathrm{i})$。两者吻合。
- 三个结果亦已用电脑在数以千计的随机中心、角度和点上验证, 验证时以几何方式（按模和辐角）进行旋转，而非使用上述公式；误差均小于 10^{-12}。

常见错误

- 复合次序弄错：SR 表示先作用 R，所以是把 $R(z)$ 代入 S。
- 除以 $1-\mathrm{e}^{\mathrm{i}(\theta+\phi)}$ 时，没有指出它正是因为 $\theta+\phi \neq 2n\pi$ 才不为零。
- 在 (iii) 中，遗漏了其中一个旋转是恒等映射的平凡情况。

STEP 之外

满足 $|\alpha| = 1$ 的映射 $z \mapsto \alpha z + \beta$，恰好就是平面上所有保持定向的等距变换（orientation-preserving isometries），而它们构成一个群（group）：旋转与平移合在一起，在复合运算下封闭。第 (iii) 部分是“这个群不可交换”这一事实的一个小例子，而交换子（commutator）$RSR^{-1}S^{-1}$ 原来是一个平移。Needham 的 *Visual Complex Analysis*（《直观复分析》）把复数乘法的这种几何观点发展得非常漂亮，而 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则涵盖相关的群论。

详细解答

STEP III 2017，第 3 题

纯数学 · 多项式的根

STEP III 2017，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/17-s3-q3

本题在 *Beyond STEP* 第 6.4 节解四次方程中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch06/quartic.html

关键思路

任何多项式的系数都是其根的对称函数（symmetric functions），所以三次方程的系数 A 是其三个根之和的相反数，而这个和正是 α、β、γ、δ 的一个熟悉的对称函数。对于题中指定的四次方程，其对应的三次方程有一个整数根，因此可以完全因式分解。知道 $\alpha\beta+\gamma\delta$ 的值后，便可把四次方程拆成两个二次因式，一个以 α、β 为根，另一个以 γ、δ 为根，而其余的系数则确定这两个因式。

解答

求 A

全程我们使用四次方程的韦达定理（Vieta's relations）：

$$\alpha+\beta+\gamma+\delta=-p, \qquad \alpha\beta+\alpha\gamma+\alpha\delta+\beta\gamma+\beta\delta+\gamma\delta=q,$$
$$\alpha\beta\gamma+\alpha\beta\delta+\alpha\gamma\delta+\beta\gamma\delta=-r, \qquad \alpha\beta\gamma\delta=s.$$

该三次方程是首一（monic）的，其根为 $y_1=\alpha\beta+\gamma\delta$、$y_2=\alpha\gamma+\beta\delta$ 及 $y_3=\alpha\delta+\beta\gamma$，所以

$$y^3+Ay^2+\cdots=(y-y_1)(y-y_2)(y-y_3),$$

比较 y^2 的系数，得 $A = -(y_1 + y_2 + y_3)$。但 $y_1 + y_2 + y_3$ 恰好包含六个由两个不同根组成的乘积各一次，所以它等于 q。因此

$$A = -q.$$

第 (i) 部分

当 $p = 0$、$q = 3$、$r = -6$、$s = 10$ 时，三次方程的系数为

$$A = -3, \qquad pr - 4s = -40, \qquad 4qs - p^2 s - r^2 = 120 - 0 - 36 = 84,$$

所以方程为

$$y^3 - 3y^2 - 40y + 84 = 0.$$

任何整数根都必整除 84。试代入较小的因数，$y = 2$ 给出 $8 - 12 - 80 + 84 = 0$，所以 $y - 2$ 是一个因式：

$$y^3 - 3y^2 - 40y + 84 = (y - 2)(y^2 - y - 42) = (y - 2)(y - 7)(y + 6).$$

各根为 2、7 及 -6；最大的是 7，所以

$$\alpha\beta + \gamma\delta = 7.$$

第 (ii) 部分

展开得

$$(\alpha + \beta)(\gamma + \delta) = \alpha\gamma + \alpha\delta + \beta\gamma + \beta\delta,$$

即全部六个两两乘积之和减去 $\alpha\beta$ 与 $\gamma\delta$。所以

$$(\alpha + \beta)(\gamma + \delta) = q - (\alpha\beta + \gamma\delta) = 3 - 7 = -4.$$

接着，$\alpha\beta$ 与 $\gamma\delta$ 之和为 7，积为 $\alpha\beta\gamma\delta = s = 10$。因此它们是以下方程的根：

$$t^2 - 7t + 10 = (t - 5)(t - 2) = 0,$$

又因为 $\alpha\beta > \gamma\delta$，所以 $\alpha\beta = 5$ 及 $\gamma\delta = 2$。

第 (iii) 部分

设 $u = \alpha + \beta$。由于 $p = 0$，四个根之和为零，所以 $\gamma + \delta = -u$，而第 (ii) 部分给出 $-u^2 = -4$，所以 $u = \pm 2$。

? **想一想** u 应取哪个符号？我们怎样判断？

我们还未用到 r。把四个三元乘积两两分组，

$$\alpha\beta\gamma+\alpha\beta\delta+\alpha\gamma\delta+\beta\gamma\delta=\alpha\beta(\gamma+\delta)+\gamma\delta(\alpha+\beta)=5(-u)+2u=-3u.$$

这等于 $-r=6$，所以 $u=-2$。（取另一个符号会得到 $r=+6$，即另一个四次方程。）因此

$$\alpha+\beta=-2,\quad \alpha\beta=5,\qquad \gamma+\delta=2,\quad \gamma\delta=2.$$

所以 α、β 是 $x^2+2x+5=0$ 的根，得 $x=-1\pm 2\mathrm{i}$；而 γ、δ 是 $x^2-2x+2=0$ 的根，得 $x=1\pm\mathrm{i}$。因此四次方程

$$x^4+3x^2-6x+10=(x^2+2x+5)(x^2-2x+2)=0$$

的根为

$$x=-1+2\mathrm{i},\quad -1-2\mathrm{i},\quad 1+\mathrm{i},\quad 1-\mathrm{i}.$$

验算

- 展开 $(x^2+2x+5)(x^2-2x+2)$：x^3 项互相抵消，得到 $x^4+3x^2-6x+10$，正是原来的四次式。
- 以这些根计算，$\alpha\gamma+\beta\delta$ 及 $\alpha\delta+\beta\gamma$（按某个次序）应为三次方程的另外两个根 2 及 -6。它们之和为 $q-7=-4=2+(-6)$，与预期相符。
- 公式 $A=-q$ 亦已用电脑在一千个根为随机复数的四次方程上验证过。

常见错误

- 误取 $A=q$：y^2 的系数是各根之和的*相反数*。
- 停在 $u=\pm 2$ 并同时给出两个可能：只有一个符号与 $r=-6$ 相容。
- 预期有实根：这个四次式对所有实数 x 都是正数，所以四个根全是复数。

STEP 之外

这里的三次方程是该四次方程的*预解三次方程*（resolvent cubic），也是费拉里（Ferrari）方法的核心：解任何四次方程都可化为解一个三次方程，再解若干二次方程。这方法为何奏效，可由伽罗瓦理论（Galois theory）解释；该理论研究根的置换（permutations）如何作用于像 $\alpha\beta+\gamma\delta$ 这样的式子（三个式子各自被 24 个置换中一个含八个置换的子群（subgroup）固定）。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是入门这些概念的好起点。

详细解答

STEP III 2017，第 4 题

纯数学 · 函数的几何平均

STEP III 2017，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/17-s3-q4

关键思路

几何平均（geometric mean）其实就是普通平均值的指数：$\ln F(y)$ 是 $\ln f$ 在 $[0,y]$ 上的平均值。因此 F 的每一个性质，都来自对数的某个性质（换底公式（change of base）、$\ln(fg)=\ln f+\ln g$），再配合积分的线性（linearity）。至于第 (iv) 部分的逆命题（converse），条件会变成一条含积分的方程，对它求导后，便化为一条简单的微分方程（differential equation）。

解答

全文中记

$$\ln F(y)=\frac{1}{y}\int_0^y \ln f(x)\,\mathrm{d}x \qquad (y>0),$$

这就是把定义两边取对数后的形式。

第 (i) 部分

由于 $a>0$ 且 $a\neq 1$，$\log_a$ 有定义，而且 $\ln a\neq 0$；由换底公式得 $\log_a f(x)=\dfrac{\ln f(x)}{\ln a}$。此外，对任何实数 t，都有 $a^t=\mathrm{e}^{t\ln a}$。因此

$$\begin{aligned} a^{\frac{1}{y}\int_0^y \log_a f(x)\,\mathrm{d}x} &= \exp\left(\ln a\cdot\frac{1}{y}\int_0^y \frac{\ln f(x)}{\ln a}\,\mathrm{d}x\right)\\ &= \exp\left(\frac{1}{y}\int_0^y \ln f(x)\,\mathrm{d}x\right)=F(y), \end{aligned}$$

其中我们把常数 $\frac{1}{\ln a}$ 提到积分号外，再与 $\ln a$ 约去。这就是所求的结果。（条件 $a\neq 1$ 正是容许我们除以 $\ln a$ 的原因。）

第 (ii) 部分

由于 $f(x)>0$ 且 $g(x)>0$，故 $h(x)>0$，所以 H 有定义，且 $\ln h(x)=\ln f(x)+\ln g(x)$。由积分的线性，

$$\ln H(y)=\frac{1}{y}\int_0^y\big(\ln f(x)+\ln g(x)\big)\,\mathrm{d}x=\ln F(y)+\ln G(y)=\ln\big(F(y)G(y)\big).$$

由于 ln 是一一对应的（one-to-one），故 $H(y)=F(y)G(y)$。

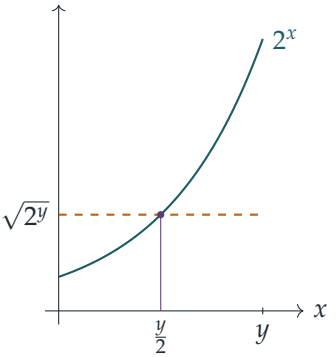
第 (iii) 部分

这里 $f(x)=b^x>0$，且 $\ln f(x)=x\ln b$，所以

$$\ln F(y)=\frac{1}{y}\int_0^y x\ln b\,\mathrm{d}x=\frac{1}{y}\cdot\frac{y^2}{2}\ln b=\frac{y}{2}\ln b.$$

因此 $F(y)=\mathrm{e}^{\frac{y}{2}\ln b}=b^{y/2}=\sqrt{b^y}$，即为所求。

下图展示 $b=2$ 时这个结果的意思：2^x 在 $[0,y]$ 上的几何平均，等于 2^x 在中点 $\frac{y}{2}$ 的值，因为 $\ln 2^x$ 是线性的，而线性函数的平均值就是它在中点的值。



第 (iv) 部分

正如题目隐含地假设的一样，我们假设 f 是连续（continuous）的，使各积分存在，并可使用微积分基本定理（fundamental theorem of calculus）。设 $\phi(x)=\ln f(x)$；由于 $f(x)>0$，它有定义，而且是连续的。取对数后，假设 $F(y)=\sqrt{f(y)}$ 表示

$$\frac{1}{y}\int_0^y\phi(x)\,\mathrm{d}x=\tfrac{1}{2}\phi(y)\quad\text{即}\quad\int_0^y\phi(x)\,\mathrm{d}x=\frac{y}{2}\phi(y)\qquad\text{对所有}y>0.\qquad(*)$$

想一想 怎样才能消去积分？我们又是否可以这样做？

对 y 求导。由微积分基本定理，$(*)$ 的左方是可微（differentiable）的，其导数为 $\phi(y)$。所以右方也可微，因此当 $y>0$ 时，$\phi(y)=\frac{2}{y}\times$（左方）也可微。对 $(*)$ 求导，右方使用乘积法则（product rule）：

$$\phi(y)=\tfrac{1}{2}\phi(y)+\tfrac{y}{2}\phi'(y)\quad\Longrightarrow\quad y\,\phi'(y)-\phi(y)=0.$$

由于 $y>0$，我们可以除以 y^2，并认出这是由商法则（quotient rule）得出的导数：

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{\phi(y)}{y}\right)=\frac{y\phi'(y)-\phi(y)}{y^2}=0.$$

所以 $\phi(y)/y$ 在区间 $y>0$ 上是常数，设为 k（由均值定理（mean value theorem），在一个区间上导数为零的函数是常数）。因此 $\ln f(x)=kx$，且

$$f(x)=\mathrm{e}^{kx}=b^x,\qquad\text{其中}b=\mathrm{e}^k>0.$$

这对每个 $x>0$ 都成立；单一点 $x=0$ 的值不影响任何积分，而且由连续性，无论如何都有 $f(0)=1=b^0$。结合第 (iii) 部分，这说明几何平均等于 $\sqrt{f(y)}$ 的函数，恰好就是函数 b^x。

验算

- 在第 (iii) 部分取 $b=4$、$y=2$：$\ln F(2)=\frac{1}{2}\int_0^2 x\ln 4\,\mathrm{d}x=\ln 4$，所以 $F(2)=4=\sqrt{4^2}$；数值积分的结果与此吻合。
- 在第 (iv) 部分，试一个不是 b^x 形式的函数，例如 $f(x)=3\cdot 2^x$。由第 (ii) 及 (iii) 部分，$F(y)=3\cdot 2^{y/2}$，而 $\sqrt{f(y)}=\sqrt{3}\cdot 2^{y/2}$，所以条件不成立，理应如此。我们的论证会得出 $\phi(y)/y=\ln 2+\frac{\ln 3}{y}$，它并非常数。
- 第 (i) 及 (ii) 部分已用电脑对随机的正函数、y 值，以及大于和小于 1 的底 a 作验证。

常见错误

- 在第 (iv) 部分对 $(*)$ 求导时，忘记对 $\frac{y}{2}\phi(y)$ 使用乘积法则，结果完全丢失了 ϕ' 项。
- 单凭第 (iii) 部分便断定 $f(x)=b^x$：第 (iii) 部分只说明 b^x 具有这个性质，并没有说明其他函数都没有。第 (iv) 部分是逆命题，需要独立的论证。
- 在第 (i) 部分没有指出哪里需要 $a\neq 1$：正是这个条件使 $\log_a$ 有意义，并使 $\ln a\neq 0$。

STEP 之外

连续几何平均是 $(x_1x_2\cdots x_n)^{1/n}$ 的积分版本；对凹函数（concave function）ln 应用琴生不等式（Jensen's inequality），可证明它永远不会超过普通平均值 $\frac{1}{y}\int_0^y f(x)\,\mathrm{d}x$，这就是

连续版本的算术–几何平均不等式（AM–GM inequality）。第 (iv) 部分是一个小例子，展示如何把函数方程（functional equation）化为微分方程来求解，而这需要小心处理可微性；在分析课程中，你会学到微积分基本定理究竟在甚么情况下容许这样做。如需严谨的理论背景，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2017，第 5 题

纯数学 · 极坐标曲线

STEP III 2017，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/17-s3-q5

关键思路

以 θ 为参数，把曲线写成参数式（parametric form）：$x = f(\theta)\cos\theta$，$y = f(\theta)\sin\theta$。这样斜率（gradient）便可由连锁法则（chain rule）求得，而“以直角相交”就变成“两斜率之积为 -1”。在最后一部分，曲线族 $r = a(1+\sin\theta)$ 经过 C 上的每一点，所以直角条件沿整条 C 都成立。于是条件便化为关于 f 的一阶微分方程（first-order differential equation）。

解答

斜率

在曲线上，$x = f(\theta)\cos\theta$ 及 $y = f(\theta)\sin\theta$。用乘积法则（product rule）对 θ 求导，

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = f'(\theta)\cos\theta - f(\theta)\sin\theta, \qquad \frac{\mathrm{d}y}{\mathrm{d}\theta} = f'(\theta)\sin\theta + f(\theta)\cos\theta.$$

由连锁法则，再把分子和分母同除以 $\cos\theta$（假设 $\cos\theta \neq 0$），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{f'(\theta)\sin\theta + f(\theta)\cos\theta}{f'(\theta)\cos\theta - f(\theta)\sin\theta} = \frac{f'(\theta)\tan\theta + f(\theta)}{f'(\theta) - f(\theta)\tan\theta}.$$

以直角相交的曲线

在交点处，两条曲线对应同一个 θ。设 $t = \tan\theta$，并以 f, f', g, g' 表示该处的值。若两条切线（tangent）都不是铅垂的，则两曲线垂直（perpendicular）当且仅当两斜率之积为 -1：

$$(f't + f)(g't + g) = -(f' - ft)(g' - gt).$$

把两边展开，含 $f'gt$ 和 $fg't$ 的项互相抵消，剩下

$$f'g't^2 + fg + f'g' + fgt^2 = 0 \iff (1+t^2)(f'g' + fg) = 0.$$

由于 $1 + t^2 = \sec^2\theta > 0$，得 $f'(\theta)g'(\theta) + f(\theta)g(\theta) = 0$，此即所求。

> **想一想** 若其中一条切线是铅垂的，或 $\cos\theta = 0$，以致斜率或 $\tan\theta$ 不存在，又会怎样?

改用方向向量（direction vector）代替斜率。$r = f(\theta)$ 的切线方向为 $\big(f'\cos\theta - f\sin\theta,\ f'\sin\theta + f\cos\theta\big)$，$g$ 的情况亦同理。两者的纯量积（scalar product）为

$$(f'\cos\theta - f\sin\theta)(g'\cos\theta - g\sin\theta) + (f'\sin\theta + f\cos\theta)(g'\sin\theta + g\cos\theta) = f'g' + fg,$$

因为交叉项互相抵消，且 $\cos^2\theta + \sin^2\theta = 1$。所以在任何情况下，垂直的条件都是同一个。

求 $f(\theta)$

对于 $r = a(1+\sin\theta)$，有 $g'(\theta) = a\cos\theta$。

> **想一想** 这个条件只在两曲线相交之处成立。为何我们可以在 C 的每一点上使用它?

任取 C 上一点，其角度为 θ，且 $f(\theta) > 0$、$\sin\theta \neq -1$。取 $a = f(\theta)/(1+\sin\theta) > 0$，便得到曲线族中一条在同一 θ 经过该点的曲线。依题设，这条曲线与 C 以直角相交，因此

$$f'(\theta)\, a\cos\theta + f(\theta)\, a(1+\sin\theta) = 0.$$

除以 $a > 0$，每个这样的 θ 都满足

$$f'(\theta)\cos\theta + f(\theta)(1+\sin\theta) = 0.$$

乘以 $1-\sin\theta$，并利用 $1 - \sin^2\theta = \cos^2\theta$:

$$\cos\theta\left[f'(\theta)(1-\sin\theta) + f(\theta)\cos\theta\right] = 0.$$

在 $\cos\theta \neq 0$ 之处，括号内的式子为零。由商法则（quotient rule），括号内的式子等于 $(1-\sin\theta)^2$ 乘以 $f(\theta)/(1-\sin\theta)$ 的导数:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{f(\theta)}{1-\sin\theta}\right) = \frac{f'(\theta)(1-\sin\theta) + f(\theta)\cos\theta}{(1-\sin\theta)^2} = 0.$$

所以在区间 $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$ 上，$f(\theta) = k(1-\sin\theta)$，其中 k 为常数。（等价地，我们也可以对 $f'/f = -(1+\sin\theta)/\cos\theta = -\cos\theta/(1-\sin\theta)$ 分离变量（separation of

variables）。）曲线经过 $r=4$、$\theta=-\frac{1}{2}\pi$，故由连续性（continuity）得 $k(1+1)=4$，即 $k=2$。在 $\frac{1}{2}\pi<\theta<\frac{3}{2}\pi$ 上，同样的论证亦适用，而 $\theta=\frac{3}{2}\pi$ 与 $-\frac{1}{2}\pi$ 是同一方向，所以同样有 $k=2$。因此

$$\boxed{f(\theta)=2(1-\sin\theta)}$$

这是另一条心形线（cardioid），即 $r=2(1+\sin\theta)$ 关于 x 轴的反射。

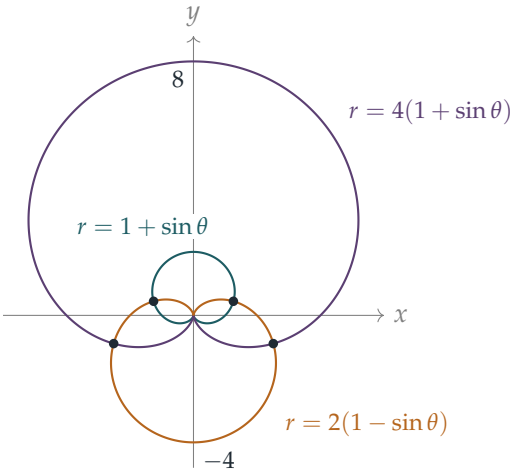
草图

曲线 $r=a(1+\sin\theta)$ 是一条尖点（cusp）在原点、关于 y 轴对称的心形线。它向上伸至 $(0,2a)$，并在 $(\pm a,0)$ 与 x 轴相交。曲线 $r=2(1-\sin\theta)$ 是形状相同但开口向下的心形线：其尖点在原点，在 $(\pm2,0)$ 与 x 轴相交，并向下伸至 $(0,-4)$。

交点可由解 $a(1+\sin\theta)=2(1-\sin\theta)$ 得出，结果为 $\sin\theta=\frac{2-a}{2+a}$：

- $a=1$：$\sin\theta=\frac{1}{3}$，$r=\frac{4}{3}$，交点为 $\left(\pm\frac{8\sqrt{2}}{9},\frac{4}{9}\right)$；
- $a=4$：$\sin\theta=-\frac{1}{3}$，$r=\frac{8}{3}$，交点为 $\left(\pm\frac{16\sqrt{2}}{9},-\frac{8}{9}\right)$。

三条曲线亦全都经过原点。



验算

- 在 $\theta=0$ 时，曲线 $r=2(1-\sin\theta)$ 位于 $(2,0)$，而 $r=2(1+\sin\theta)$ 亦经过同一点。该处 $f=2$、$f'=-2$、$g=2$、$g'=2$，所以 $f'g'+fg=-4+4=0$，与预期相符。
- 在 $\theta=-\frac{1}{2}\pi$ 时，$f=2(1+1)=4$，符合要求。
- 已用电脑确认在 $\left(\frac{8\sqrt{2}}{9},\frac{4}{9}\right)\approx(1.257,0.444)$ 的两条切线互相垂直；对许多随机取的 a 值，切线亦同样经确认为垂直。

常见错误

- 只在一个交点使用直角条件。“对每个正数 a”的用意，正是要令条件在 C 的每一点都成立。
- 忘记 $f(\theta)/(1-\sin\theta)$ 在 $\theta=\frac{1}{2}\pi$ 没有定义，因而在决定常数时不够谨慎。
- 把 $r=2(1-\sin\theta)$ 画成开口向上：由于在 $\theta=-\frac{1}{2}\pi$ 时 $r=4$，这条曲线主要位于 x 轴下方。

STEP 之外

若一族曲线与另一族曲线处处以直角相交，这两族曲线便称为一对**正交轨线**（orthogonal trajectories）。这里的心形线 $r=a(1+\sin\theta)$ 与 $r=b(1-\sin\theta)$ 就构成这样的一对。这个概念出现在大学一年级的微分方程课程中，亦在物理学里再次出现：场线（field lines）与等势线（equipotentials）以直角相交。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）讲解了如何求正交轨线。Needham 的 *Visual Complex Analysis*（《直观复分析》）则展示了共形映射（conformal map）如何产生这类曲线族。

详细解答

STEP III 2017，第 6 题

纯数学 · 代换积分法

STEP III 2017，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：

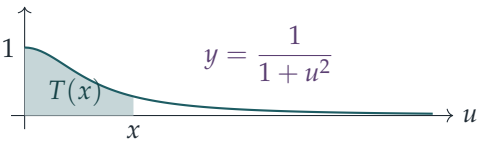
step.maths.org/questions/17-s3-q6

本题在 *Beyond STEP* 第 4.2 节由一个积分建立三角学中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch04/trig-from-integral.html

关键思路

我们只凭积分就要建立反正切函数（inverse tangent），所以每一个性质都必须由代换（substitution）得出。代换 $u \mapsto 1/u$ 把区间 $(0, x)$ 与 $(1/x, \infty)$ 互换；而默比乌斯（Möbius）代换 $v = \frac{u+a}{1-au}$ 使连同微分在内的被积函数（integrand）$\frac{\mathrm{d}u}{1+u^2}$ 保持不变，因此它只是平移了积分上下限。在最后一部分适当选取数值，使平移后的上下限落回一个已知的值，便得到一条可以解出的方程。

解答



$T(x)$ 是阴影部分的面积；T_∞ 是 $u \geq 0$ 时曲线下方的全部面积。

第 (i) 部分

在 $T(x)$ 的积分中代入 $u = 1/t$，于是 $\mathrm{d}u = -t^{-2}\,\mathrm{d}t$；当 u 由 0 变到 x 时，t 由 ∞ 下降到 $1/x$。于是

$$\begin{aligned} T(x) &= \int_{\infty}^{1/x} \frac{1}{1+t^{-2}}\left(-\frac{1}{t^2}\right)\mathrm{d}t = \int_{1/x}^{\infty} \frac{1}{t^2+1}\,\mathrm{d}t \\ &= \int_0^{\infty} \frac{\mathrm{d}t}{1+t^2} - \int_0^{1/x} \frac{\mathrm{d}t}{1+t^2} = T_\infty - T(x^{-1}). \end{aligned}$$

（由于 T_∞ 是有限值，从 0 开始的积分是确实存在的积分（proper integral），所以把积分区间拆开是合理的。）这正是所求的结果。

第 (ii) 部分

导数。由商法则（quotient rule），当 $u \neq a^{-1}$ 时，

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

另一方面，

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

除以 $1+u^2$，得 $\dfrac{1+v^2}{1+u^2} = \dfrac{1+a^2}{(1-au)^2} = \dfrac{\mathrm{d}v}{\mathrm{d}u}$，正如所求。

加法公式（addition formula）。取 $a > 0$ 及 $0 < x < \frac{1}{a}$。当 $0 \le u \le x$ 时，$1-au \ge 1-ax > 0$，所以 v 在整个区间上有定义且连续，并且 $\frac{\mathrm{d}v}{\mathrm{d}u} > 0$，因此 v 由 $v(0) = a$ 递增至 $v(x) = \frac{x+a}{1-ax}$。

> **想一想** 为甚么 $x < \frac{1}{a}$ 这个条件重要?

若 x 达到 $\frac{1}{a}$，v 便会在积分区间内趋于无穷大，代换就失去意义。在 $x < \frac{1}{a}$ 的情况下，它是 $[0, x]$ 上一个光滑而递增的变量代换。

导数的结果正好表示 $\dfrac{1}{1+u^2} = \dfrac{1}{1+v^2}\dfrac{\mathrm{d}v}{\mathrm{d}u}$，因此

$$\begin{aligned} T(x) = \int_0^x \frac{\mathrm{d}u}{1+u^2} &= \int_a^{\frac{x+a}{1-ax}} \frac{\mathrm{d}v}{1+v^2} \\ &= \int_0^{\frac{x+a}{1-ax}} \frac{\mathrm{d}v}{1+v^2} - \int_0^a \frac{\mathrm{d}v}{1+v^2} = T\left(\frac{x+a}{1-ax}\right) - T(a). \end{aligned}$$

这里 T 的两个自变量都是正数，所以一切都有定义。

第一个推论。记 $X = \frac{x+a}{1-ax}$。先后对 x 和 a 应用第 (i) 部分，

$$T(x^{-1}) = T_\infty - T(x) = T_\infty - T(X) + T(a) = T_\infty - T(X) + T_\infty - T(a^{-1}),$$

即 $T(x^{-1}) = 2T_\infty - T\left(\frac{x+a}{1-ax}\right) - T(a^{-1})$。

第二个推论。已知 $b > 0$ 及 $y > \frac{1}{b}$，设 $x = \frac{1}{y}$、$a = \frac{1}{b}$。于是 $x > 0$、$a > 0$，而 $y > \frac{1}{b}$ 等价于 $x < b = \frac{1}{a}$，所以上一个结果适用。分子分母同乘以 by，

$$\frac{x+a}{1-ax} = \frac{\frac{1}{y} + \frac{1}{b}}{1 - \frac{1}{by}} = \frac{b+y}{by-1},$$

又 $x^{-1} = y$、$a^{-1} = b$，所以

$$T(y) = 2T_\infty - T\left(\frac{y+b}{by-1}\right) - T(b).$$

第 (iii) 部分

$T(\sqrt{3})$。在最后一个结果中取 $y = b = \sqrt{3}$；由于 $\sqrt{3} > \frac{1}{\sqrt{3}}$，条件 $y > \frac{1}{b}$ 成立。此时

$$\frac{y+b}{by-1} = \frac{2\sqrt{3}}{3-1} = \sqrt{3},$$

所以 $T(\sqrt{3}) = 2T_\infty - T(\sqrt{3}) - T(\sqrt{3})$，即 $3T(\sqrt{3}) = 2T_\infty$，故

$$T(\sqrt{3}) = \tfrac{2}{3}T_\infty.$$

先求 $T(1)$。在第 (i) 部分取 $x = 1$，得 $T(1) = T_\infty - T(1)$，所以 $T(1) = \frac{1}{2}T_\infty$。

> ? **想一想** 当 $x = a$ 时，a 取甚么值才能使 $\frac{x+a}{1-ax}$ 等于 1?

我们需要 $\frac{2a}{1-a^2} = 1$，即 $a^2 + 2a - 1 = 0$，其正根为 $a = \sqrt{2} - 1$。

$T(\sqrt{2}-1)$。于是在加法公式中取 $x = a = \sqrt{2} - 1 > 0$。条件 $x < \frac{1}{a}$ 即 $\sqrt{2} - 1 < \sqrt{2} + 1$，显然成立。由于 $a^2 = 3 - 2\sqrt{2}$，

$$\frac{x+a}{1-ax} = \frac{2(\sqrt{2}-1)}{1-(3-2\sqrt{2})} = \frac{2(\sqrt{2}-1)}{2(\sqrt{2}-1)} = 1,$$

所以 $T(\sqrt{2}-1) = T(1) - T(\sqrt{2}-1)$，得

$$T(\sqrt{2}-1) = \tfrac{1}{2}T(1) = \tfrac{1}{4}T_\infty.$$

验算

当然，T 其实就是 arctan，而 $T_\infty = \frac{\pi}{2}$；题目不准我们使用这些事实，但我们大可以用它们来核对：$\frac{\pi}{3} = \frac{2}{3} \cdot \frac{\pi}{2}$，$\frac{\pi}{8} = \frac{1}{4} \cdot \frac{\pi}{2}$。用数值方法（辛普森法则（Simpson's rule））计算这些积分，得 $T_\infty \approx 1.570796$、$T(\sqrt{3}) \approx 1.047198$ 及 $T(\sqrt{2}-1) \approx 0.392699$，与 $\frac{2}{3}T_\infty$ 和 $\frac{1}{4}T_\infty$ 吻合。第 (ii) 部分的加法公式其实就是大家熟悉的 $\tan(A+B)$ 公式换了个样子，其中 $x = \tan A$、$a = \tan B$。

常见错误

- 代入 $u = 1/t$ 后忘记把上下限对调：正是 $\mathrm{d}u = -t^{-2}\,\mathrm{d}t$ 中的负号把 $\int_{\infty}^{1/x}$ 变成 $\int_{1/x}^{\infty}$。
- 在第 (iii) 部分暗中使用 $\tan\frac{\pi}{3} = \sqrt{3}$ 或 arctan 的性质：题目禁止这样做，重点在于一切都要由第 (i) 和第 (ii) 部分推出。
- 没有就所用的具体数值检查条件 $x < \frac{1}{a}$ 和 $y > \frac{1}{b}$。

STEP 之外

用积分定义一个函数，再从定义推导出它的全部性质，正是分析（analysis）课程严格建立 ln、exp、arctan 甚至 π 的方法：在这里，π 可以直接定义为 $2T_{\infty}$。代换 $v = \frac{u+a}{1-au}$ 是一个默比乌斯变换（Möbius transformation），而它保持 $\frac{\mathrm{d}u}{1+u^2}$ 不变这一事实，反映了圆的旋转。分析方面可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）；默比乌斯映射的几何方面，可参阅 Needham 的 *Visual Complex Analysis*（《直观复分析》）。

详细解答

STEP III 2017，第 7 题

纯数学 · 椭圆的切线

STEP III 2017，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/17-s3-q7

关键思路

参数（parameter）t 把“在 T 点的切线（tangent）通过 (X, Y)”这个条件化为一条关于 t 的二次方程（quadratic equation）。每一个根（root）都对应一个切点（point of contact），所以数从一点出发的切线条数，就等于数实根的个数，也就是判别式（discriminant）的问题。在第 (ii) 部分，两个切点的参数 p 和 q 正好是该二次方程的两个根，因此它们的和与积可以直接从系数读出。

解答

该点位于椭圆上

代入 T 的坐标，

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{(1-t^2)^2+4t^2}{(1+t^2)^2}=\frac{1+2t^2+t^4}{(1+t^2)^2}=1,$$

所以对每一个实数 t，T 都位于椭圆（ellipse）上。

第 (i) 部分

对 t 求导，

$$\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{-4at}{(1+t^2)^2},\qquad \frac{\mathrm{d}y}{\mathrm{d}t}=\frac{2b(1-t^2)}{(1+t^2)^2}.$$

与其把两者相除（在 $t=0$ 时会出问题），不如使用椭圆在 (x_0, y_0) 处的标准切线方程 $\frac{xx_0}{a^2}+\frac{yy_0}{b^2}=1$；它的法向量（normal）方向 $\left(\frac{x_0}{a^2}, \frac{y_0}{b^2}\right)$ 与上面的速度向量垂直，因为

$\frac{1-t^2}{a}\cdot(-4at)+\frac{2t}{b}\cdot 2b(1-t^2)=0$。因此在 T 点的切线 L 为

$$\frac{x(1-t^2)}{a(1+t^2)}+\frac{2ty}{b(1+t^2)}=1 \iff bx(1-t^2)+2aty=ab(1+t^2).$$

代入 $(x,y)=(X,Y)$，并按 t 的幂次整理：

$$bX-bXt^2+2aYt-ab-abt^2=0 \iff (a+X)bt^2-2aYt+b(a-X)=0,$$

即为所求。

*两条切线。*反过来看：参数为 t 的点处的切线通过 (X,Y)，当且仅当 t 满足这条方程。由于 $X\neq -a$ 且 $b\neq 0$，t^2 的系数不为零，所以这确实是一条二次方程，其判别式为

$$4a^2Y^2-4b^2(a+X)(a-X)=4\big(a^2Y^2-(a^2-X^2)b^2\big).$$

若 $a^2Y^2>(a^2-X^2)b^2$，判别式为正，所以有两个相异实根 $t_1\neq t_2$。不同的 t 值给出椭圆上不同的点（映射 $t\mapsto T$ 是一一对应的（one-to-one）；令 $t=\tan\frac{\theta}{2}$，它就是熟悉的 $(a\cos\theta,b\sin\theta)$，其中 $-\pi<\theta<\pi$）。不同的切点给出不同的切线，因为椭圆的切线只在其切点与椭圆相交。所以通过 (X,Y) 有两条不同的切线。

? 想一想 会不会遗漏了一条参数 t 取不到的切线?

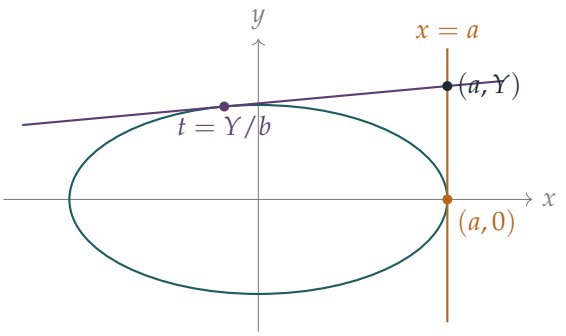
椭圆上唯一不能写成 $(*)$ 形式的点是 $(-a,0)$（即 $t\to\infty$ 时的极限），其切线为 $x=-a$。这条直线只在 $X=-a$ 时才通过 (X,Y)，而这情况已被排除。所以该二次方程涵盖了所有通过 (X,Y) 的切线。

*几何意义。*把条件除以 $a^2b^2>0$，得

$$\frac{Y^2}{b^2}>1-\frac{X^2}{a^2} \iff \frac{X^2}{a^2}+\frac{Y^2}{b^2}>1,$$

即 (X,Y) 位于椭圆之外。所以与圆的情况一样，从椭圆外任何一点都可以作两条切线。

*$X^2=a^2$ 的情况。*此时条件变为 $a^2Y^2>0$，即 $Y\neq 0$：该点位于其中一条铅垂线 $x=\pm a$ 上，但不在顶点（vertex）$(\pm a,0)$。其中一条切线就是铅垂线 $x=\pm a$ 本身，它在 $(\pm a,0)$ 与椭圆相切。另一条则在别处与椭圆相切：当 $X=a$ 时，二次方程变为 $2abt^2-2aYt=0$，其根为 $t=0$（铅垂切线）及 $t=Y/b\neq 0$；当 $X=-a$ 时，它变为 $-2aYt+2ab=0$，得 $t=b/Y$，而第二条切线是 $x=-a$，即在缺失的点 $(-a,0)$ 处的切线。下图展示 $X=a$、$Y>0$ 的情况。



所以当 $X^2 = a^2$ 时结果仍然成立：只要 $Y \neq 0$，即该点位于椭圆之外，通过 (X, Y) 就有两条不同的切线。

第 (ii) 部分

在 P 点的切线通过 (X, Y)，所以由第 (i) 部分，p 是

$$(a + X)bt^2 - 2aYt + b(a - X) = 0,$$

的一个根，q 亦然。由于 $X \neq -a$，这是一条二次方程，而且 $p \neq q$（因为 $P \neq Q$），所以 p 和 q 就是它的两个根。因此

$$pq = \frac{b(a - X)}{b(a + X)} = \frac{a - X}{a + X} \quad \Longrightarrow \quad (a + X)pq = a - X,$$

即为所求，而且

$$p + q = \frac{2aY}{b(a + X)}.$$

最后部分

在切线 $bx(1 - t^2) + 2aty = ab(1 + t^2)$ 中令 $x = 0$，得 $2aty = ab(1 + t^2)$。两条切线确实与 y 轴相交，所以 $p, q \neq 0$（在 $t = 0$ 时切线为 $x = a$），并且

$$y_1 = \frac{b(1 + p^2)}{2p}, \qquad y_2 = \frac{b(1 + q^2)}{2q}.$$

写成 $\frac{1+p^2}{p} = p + \frac{1}{p}$，

$$y_1 + y_2 = \frac{b}{2}\left(p + q + \frac{1}{p} + \frac{1}{q}\right) = \frac{b}{2}(p + q)\left(1 + \frac{1}{pq}\right) = \frac{b(p + q)(1 + pq)}{2pq}.$$

由上面的结果，$1 + pq = \dfrac{2a}{a + X}$，所以利用 $pq = \dfrac{a - X}{a + X} \neq 0$，

$$y_1 + y_2 = \frac{b}{2} \cdot \frac{2aY}{b(a + X)} \cdot \frac{2a}{a + X} \cdot \frac{a + X}{a - X} = \frac{2a^2Y}{(a + X)(a - X)} = \frac{2a^2Y}{a^2 - X^2}.$$

令它等于 $2b$，并乘以 $(a^2-X^2)/(2a^2b)$（由于 $X^2\neq a^2$，它不为零）：

$$\frac{Y}{b}=\frac{a^2-X^2}{a^2}=1-\frac{X^2}{a^2}\iff\frac{X^2}{a^2}+\frac{Y}{b}=1,$$

即为所求。所以交点位于一条通过 $(\pm a,0)$ 和 $(0,b)$ 的抛物线（parabola）上。

验算

- 取 $a=2$、$b=1$ 及 $(X,Y)=(4,1)$（在椭圆外，因为 $4+1>1$），二次方程为 $6t^2-4t-2=0$，其根为 $t=1$ 及 $t=-\frac{1}{3}$：分别对应在 $(0,1)$ 和 $(1.6,-0.6)$ 的切线。这两条切线 $y=1$ 和 $2x-3y=5$ 都确实通过 $(4,1)$。此时 $pq=-\frac{1}{3}=\frac{2-4}{2+4}$，$p+q=\frac{2}{3}=\frac{4}{6}$，与公式的预测一致。
- 在中心 $(0,0)$，判别式为 $-4a^2b^2<0$：没有切线，正如从椭圆内部出发所预期的。
- 用电脑检验了大量随机的椭圆及点对 P、Q，直接求两条切线的交点；pq、$p+q$ 和 y_1+y_2 的公式全部吻合，而且当 $y_1+y_2=2b$ 时，交点满足 $X^2/a^2+Y/b=1$。

常见错误

- 用 $\frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t}$ 求斜率（gradient），却忘记它在 $t=0$ 时失效，因为那里的切线是铅垂的。
- 在 $X=-a$ 时仍把 $(a+X)bt^2-2aYt+b(a-X)=0$ 当作二次方程；这正是 $X^2=a^2$ 的情况需要另行处理的原因。
- 只写到“判别式为正”便停下，而没有说明为何两个根会给出两条不同的直线。

STEP 之外

圆锥曲线（conic）的有理参数化（rational parametrisation）$\left(\frac{1-t^2}{1+t^2},\frac{2t}{1+t^2}\right)$ 是代数几何（algebraic geometry）的起点：正因如此，每条拥有一个有理点的圆锥曲线都有无穷多个有理点，而它亦是毕氏三元组（Pythagorean triples）分类的基础。从外点作的一对切线引出极线（polar）的概念，并通往射影几何（projective geometry）；在射影几何中，$t=\infty$ 处那个令人烦恼的缺失点变成一个普通的点。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）收录了更多这类题目。

详细解答

STEP III 2017，第 8 题

纯数学 · 分部求和法

STEP III 2017，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/17-s3-q8

关键思路

开首的恒等式是分部积分（integration by parts）的离散版本，即“分部求和”（summation by parts）：它把一个涉及 b 的差分的和，换成一个涉及 a 的差分的和。要把它用于三角和，我们选取 b_m，使 $b_{m+1}-b_m$ 产生我们想要的项；和差化积公式（sum-to-product formula）可把正弦或余弦之差化为单一乘积。然后选 a_m 为 1 或 m，便能令另一边变得容易处理。

解答

恒等式

$a_{n+1}b_{n+1}-a_1b_1$ 这个量可以伸缩相消（telescopes）：

$$a_{n+1}b_{n+1}-a_1b_1=\sum_{m=1}^{n}\left(a_{m+1}b_{m+1}-a_mb_m\right).$$

在每个括号内加上再减去 a_mb_{m+1}，得

$$a_{m+1}b_{m+1}-a_mb_m=a_m(b_{m+1}-b_m)+b_{m+1}(a_{m+1}-a_m).$$

对 $m=1,\ldots,n$ 求和并移项，便正好得到所求的恒等式。

在各个三角部分中，我们一直假设 $\sin\frac{1}{2}x\neq 0$（即 x 不是 2π 的倍数），使 $\operatorname{cosec}\frac{1}{2}x$ 有定义。

第 (i) 部分

取 $b_m=\sin mx$，以 $A=(m+1)x$、$B=mx$ 代入和差化积公式，得

$$b_{m+1}-b_m=\sin(m+1)x-\sin mx=2\cos\left(m+\tfrac{1}{2}\right)x\,\sin\tfrac{1}{2}x.$$

现在对每个 m 取 $a_m = 1$，于是 $a_{m+1} - a_m = 0$，恒等式中最后的和便等于零。恒等式变成

$$2\sin\tfrac{1}{2}x \sum_{m=1}^{n} \cos\big(m+\tfrac{1}{2}\big)x = b_{n+1} - b_1 = \sin(n+1)x - \sin x.$$

两边除以 $2\sin\frac{1}{2}x \neq 0$，

$$\sum_{m=1}^{n} \cos\big(m+\tfrac{1}{2}\big)x = \tfrac{1}{2}\big(\sin(n+1)x - \sin x\big)\operatorname{cosec}\tfrac{1}{2}x,$$

此即所求。

第 (ii) 部分

因子 m 提示我们取 $a_m = m$，使 $a_{m+1} - a_m = 1$。这样我们便需要 $b_{m+1} - b_m$ 是 $\sin mx$ 的倍数。

> ? **想一想** 选取哪个 b_m，才能令 $b_{m+1} - b_m$ 成为 $\sin mx$ 的倍数?

在第 (i) 部分中，于 $(m+1)x$ 和 mx 处两个正弦之差，给出了在中点 $(m+\frac{1}{2})x$ 处的余弦。要得到 $\sin mx$，我们需要一个余弦之差，而其两个自变量的中点是 mx，所以取 $b_m = \cos\big(m-\frac{1}{2}\big)x$。利用 $\cos A - \cos B = -2\sin\frac{A+B}{2}\sin\frac{A-B}{2}$（这亦可由题目中的第一个提示推出），

$$b_{m+1} - b_m = \cos\big(m+\tfrac{1}{2}\big)x - \cos\big(m-\tfrac{1}{2}\big)x = -2\sin mx\,\sin\tfrac{1}{2}x.$$

记 $S = \sum_{m=1}^{n} m\sin mx$。取 $a_m = m$，恒等式给出

$$-2\sin\tfrac{1}{2}x\,S = (n+1)\cos\big(n+\tfrac{1}{2}\big)x - \cos\tfrac{1}{2}x - \sum_{m=1}^{n}\cos\big(m+\tfrac{1}{2}\big)x.$$

最后的和正是第 (i) 部分的结果。代入后，两边乘以 $-2\sin\frac{1}{2}x$：

$$4\sin^2\tfrac{1}{2}x\,S = 2\sin\tfrac{1}{2}x\cos\tfrac{1}{2}x - 2(n+1)\cos\big(n+\tfrac{1}{2}\big)x\,\sin\tfrac{1}{2}x + \sin(n+1)x - \sin x.$$

现在化简两个乘积：

- $2\sin\frac{1}{2}x\cos\frac{1}{2}x = \sin x$，它与 $-\sin x$ 互相抵消；
- 由 $2\cos A\sin B = \sin(A+B) - \sin(A-B)$，取 $A = (n+\frac{1}{2})x$、$B = \frac{1}{2}x$，

$$2\cos\big(n+\tfrac{1}{2}\big)x\,\sin\tfrac{1}{2}x = \sin(n+1)x - \sin nx.$$

因此

$$\begin{aligned} 4\sin^2\tfrac{1}{2}x\,S &= -(n+1)\big(\sin(n+1)x - \sin nx\big) + \sin(n+1)x \\ &= -n\sin(n+1)x + (n+1)\sin nx. \end{aligned}$$

两边除以 $4\sin^2\frac{1}{2}x$，

$$\sum_{m=1}^{n} m\sin mx = \left(-\tfrac{n}{4}\sin(n+1)x + \tfrac{n+1}{4}\sin nx\right)\operatorname{cosec}^2\tfrac{1}{2}x,$$

所以

$$p = -\frac{n}{4}, \qquad q = \frac{n+1}{4}.$$

验算

- 当 $n = 1$ 时，利用 $1-\cos x = 2\sin^2\frac{1}{2}x$，公式给出 $\frac{1}{4}(2\sin x - \sin 2x)\operatorname{cosec}^2\frac{1}{2}x = \frac{1}{2}\sin x\,(1-\cos x)\operatorname{cosec}^2\frac{1}{2}x = \sin x$。这正是唯一的一项 $1\cdot\sin x$。
- 数值上，取 $n = 3$ 及 $x = 1$：$\sin 1 + 2\sin 2 + 3\sin 3 \approx 3.08343$，而 $\left(-\frac{3}{4}\sin 4 + \sin 3\right)\operatorname{cosec}^2\frac{1}{2} \approx 3.08343$ 亦然。
- 第 (ii) 部分亦可通过把 $\sum_{m=1}^{n}\cos mx$ 对 x 求导数（differentiating）得出，结果是 $-S$；这是一个有用的独立验算。

常见错误

- 在第 (ii) 部分选取 $b_m = \cos mx$：这样差便是 $\sin(m+\frac{1}{2})x$ 的倍数，而不是 $\sin mx$ 的倍数，方法因而卡住。
- 在 $\cos A - \cos B$ 中弄错正负号：它等于 $-2\sin\frac{A+B}{2}\sin\frac{A-B}{2}$，前面有负号。
- 在除以 $\sin\frac{1}{2}x \neq 0$ 中的那个量之前，忘记先说明它不为零。

STEP 之外

分部求和（阿贝尔求和，Abel summation）是分部积分的离散孪生兄弟，也是大学一年级分析中的主力工具。配合像第 (i) 部分那样的界——它显示 $\cos(m+\frac{1}{2})x$ 的部分和（partial sums）保持有界——它可以证明狄利克雷判别法（Dirichlet's test）：例如，$\sum\frac{\sin mx}{m}$ 对每个 x 都收敛（converges）。像第 (ii) 部分那样的和亦出现于傅里叶级数（Fourier series）中，这类闭合形式正是研究收敛性的起点。Abbott 的 *Understanding Analysis*（《理解分析》）对阿贝尔求和及狄利克雷判别法有清晰的论述。

详细解答

STEP III 2017，第 9 题

力学 · 连接质点与能量

STEP III 2017，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：

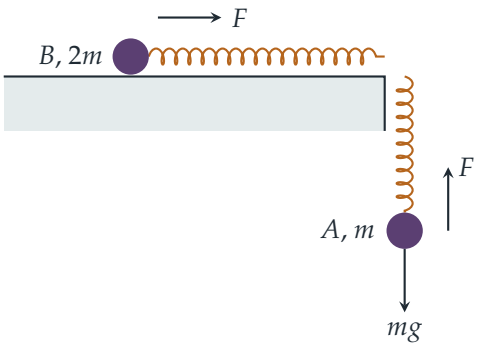
step.maths.org/questions/17-s3-q9

关键思路

弹簧的张力（tension）未知，而且随时间改变，但它以相同大小作用于两个质点上。所以，只要把两条运动方程（equations of motion）按适当比例相加，张力就会互相抵消，剩下一条可以直接积分（integrate）的关系式。在第二部分，这条积分后的关系告诉我们两个质点在时间 T 时的位置，以及它们的速率之间有何联系。接着，能量守恒（conservation of energy）给出第二条方程，两者合起来便可确定 B 的速率。

解答

我们把 A 和 B 视为质点，并把弹簧视为轻弹簧。桌面光滑，而弹簧光滑地绕过桌边。释放后，A 铅垂下落，B 则沿桌面向桌边滑动。设时间 t 时弹簧的张力为 F。当 $x < a$ 时，B 仍在桌面上。



关系式 $y+2x=\frac{1}{2}gt^2$

对 A 取向下为 y 的正方向，对 B 取向桌边为 x 的正方向。由牛顿（Newton）第二定律（Newton's second law）：

$$\begin{aligned} &\text{对}A\text{(铅垂方向)：} && m\ddot{y}=mg-F,\\ &\text{对}B\text{(水平方向)：} && 2m\ddot{x}=F. \end{aligned}$$

（对 B 而言，重量与法向反作用力（normal reaction）互相抵消，而且没有摩擦力（friction）。）把两条方程相加便消去 F：

$$m\ddot{y}+2m\ddot{x}=mg \quad\Longrightarrow\quad \ddot{y}+2\ddot{x}=g.$$

两个质点都由静止开始，且 $x=y=0$。积分一次得 $\dot{y}+2\dot{x}=gt$，由于 $t=0$ 时两个速度皆为零，所以没有常数项。再积分一次得

$$y+2x=\tfrac{1}{2}gt^2,$$

同样没有常数项，这正是所求的结果。

> **想一想** 我们不知道 F，因为它取决于不断改变的伸长量（extension）$y-x$。那我们怎样仍然能得出结论？

我们根本不需要 F。只要把两条方程乘以适当的因子（这里只需直接相加），F 就会以相反的符号出现。组合 $y+2x$ 是类似动量（momentum）的量 $my+2mx$ 的 $1/m$ 倍，而驱动它的唯一外力就是 A 的重量。

B 在时间 T 时的速率

在时间 $T=\sqrt{6a/g}$ 时，质点 B 到达桌边，所以 $x=a$。由于 $\frac{1}{2}gT^2=\frac{1}{2}g\cdot\frac{6a}{g}=3a$，由第一部分得

$$y+2a=3a \quad\Longrightarrow\quad y=a.$$

因此在时间 T 时，弹簧的伸长量为 $y-x=a-a=0$。弹簧再次回到自然长度（natural length），不储存任何弹性势能（elastic energy）。

设 u 和 v 分别为 B 和 A 在时间 T 时的速率。由积分一次后的关系式 $\dot{y}+2\dot{x}=gt$，在 $t=T$ 时得

$$v+2u=gT=\sqrt{6ag}. \tag{1}$$

能量。桌面光滑而弹簧是轻的，所以机械能（mechanical energy）守恒。开始时，两个质点都静止，弹簧未被拉伸。在时间 T 时，A 下降了距离 a，而弹簧再次处于未拉伸状态，所以失去的势能（potential energy）mga 全部转化为动能（kinetic energy）：

$$\tfrac{1}{2}mv^2+\tfrac{1}{2}(2m)u^2=mga \quad\Longrightarrow\quad v^2+2u^2=2ga. \tag{2}$$

想一想 两条方程，其中一条是二次的：我们不是应该预期有两个可能的速率吗？

把由 (1) 得出的 $v = \sqrt{6ag} - 2u$ 代入 (2)：

$$\begin{aligned} 6ag - 4\sqrt{6ag}\,u + 4u^2 + 2u^2 &= 2ga \\ 6u^2 - 4\sqrt{6ag}\,u + 4ag &= 0 \\ 3u^2 - 2\sqrt{6ag}\,u + 2ag &= 0. \end{aligned}$$

判别式（discriminant）为 $4 \cdot 6ag - 4 \cdot 3 \cdot 2ag = 0$，所以这个二次式是完全平方（perfect square），只有一个根：

$$\left(\sqrt{3}\,u - \sqrt{2ag}\right)^2 = 0 \quad \Longrightarrow \quad u = \sqrt{\frac{2ag}{3}}.$$

所以 B 到达桌边时的速率为 $\sqrt{2ag/3}$，正如所求。由 (1)，$v = \sqrt{6ag} - 2\sqrt{2ag/3} = \sqrt{2ag/3}$，两者亦相同。在那一刻，两个质点以相同速率运动，这与弹簧处于自然长度的情况吻合。

验算

- 重根（repeated root）是有意义的。当总“动量” $v + 2u$ 固定时，动能 $\frac{1}{2}mv^2 + mu^2$ 在 $v = u$ 时最小，而这里的最小值恰好是 mga。所以能量方程与 (1) 只有在 $u = v$ 时才能同时成立，此时由 (1) 直接得 $3u = \sqrt{6ag}$。
- 完整求解运动（题目叫我们不要这样做）会得到 $\omega T = 2\pi$，其中 $\omega^2 = 3\lambda/(2am)$：伸长量恰好经历一次完整的振动，然后回到零。以 $a = 1.3\,\mathrm{m}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$ 进行逐步数值积分（numerical integration），得 $x(T) = y(T) = a$，而两个速率都等于 $2.9143\,\mathrm{m\,s^{-1}} = \sqrt{2ag/3}$。
- 量纲（dimensions）：$\sqrt{ag}$ 是速率，$\sqrt{a/g}$ 是时间。

常见错误

- 忘记弹簧一般而言带有弹性势能。这个论证之所以成立，只是因为在时间 T 时 $x = y = a$，使伸长量为零。必须明确证明这一点。
- 只写出能量方程，然后尝试从中解出 u。这条方程有两个未知数，所以我们还需要关系式 (1)。

STEP 之外

量 $my + 2mx$ 的表现就像动量：内力（这里是弹簧张力）在其中互相抵消，只有外力才能改变它。这是质点系力学的起点。一般而言，我们把运动分拆为质心（centre of mass）部分和相对部分，而这里的相对坐标 $y - x$ 作简谐运动（simple harmonic motion）。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）细致地发展了这些概念。

详细解答

STEP III 2017，第 10 题

力学 • 转动的刚体

STEP III 2017，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/17-s3-q10

关键思路

在质点未滑动（slip）之前，杆与质点作为一个刚体（rigid body）一同绕铰链（hinge）转动，因此能量守恒（conservation of energy）可立即给出 $\dot{\theta}$，再求导便得 $\ddot{\theta}$。此时质点绕 P 作圆周运动，所以其加速度有熟悉的径向（radial）分量和横向（transverse）分量。单独对质点应用牛顿第二定律（Newton's second law），沿杆方向及垂直于杆方向分解，便得到摩擦力（friction）和法向反作用力（normal reaction）；当摩擦力达到反作用力的 μ 倍时，质点开始滑动。

解答

能量方程

把质点视为质点模型（point mass），并假设铰链光滑（smooth）。均匀杆绕其端点 P 的转动惯量（moment of inertia）为 $\frac{1}{3}m(3a)^2 = 3ma^2$，而与 P 相距 ℓ 的质点则贡献 $m\ell^2$。在没有滑动时，整个系统是一个以角速度（angular speed）$\dot{\theta}$ 绕 P 转动的刚体，其动能（kinetic energy）为

$$\tfrac{1}{2}\left(3ma^2 + m\ell^2\right)\dot{\theta}^2.$$

当杆转过角度 θ 时，其中点（与 P 相距 $\frac{3}{2}a$）下降了 $\frac{3}{2}a\sin\theta$，而质点下降了 $\ell\sin\theta$。铰链的反作用力不作功（其作用点固定不动），而摩擦力和法向反作用力是系统的内力，在没有滑动时不作净功。由静止开始，能量守恒给出

$$\tfrac{1}{2}m\left(3a^2 + \ell^2\right)\dot{\theta}^2 = mg\left(\tfrac{3}{2}a + \ell\right)\sin\theta,$$

两边乘以 $2/m$，即得

$$(3a^2 + \ell^2)\dot{\theta}^2 = g(3a + 2\ell)\sin\theta,$$

此即所求。

角加速度

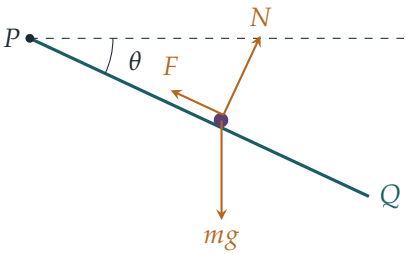
对 t 求导，得

$$2(3a^2+\ell^2)\dot\theta\,\ddot\theta = g(3a+2\ell)\cos\theta\,\dot\theta.$$

当 $t>0$ 时 $\dot\theta\neq 0$，故可两边除以 $2\dot\theta$（$t=0$ 时的结果可由连续性得出，或直接对 P 取力矩得出）：

$$\ddot\theta = \frac{g(3a+2\ell)\cos\theta}{2(3a^2+\ell^2)}.$$

作用于质点的力



杆位于水平线下方，与水平成角 θ。杆对质点施加垂直于杆的法向反作用力 N（方向背离杆，因为质点位于杆的上方），以及沿杆方向的摩擦力 F，我们设其指向 P。质点绕 P 作半径为 ℓ 的圆周运动，因此其加速度有一个垂直于杆、沿 θ 增加方向的分量 $\ell\ddot\theta$（向下并背离杆的上表面），以及一个沿杆指向 P 的分量 $\ell\dot\theta^2$。

在垂直于杆、沿 θ 增加的方向上，重力的分量为 $mg\cos\theta$，而 N 的方向相反：

$$mg\cos\theta - N = m\ell\ddot\theta.$$

沿杆指向 P 的方向上，重力的分量为 $-mg\sin\theta$（它把质点拉离 P）：

$$F - mg\sin\theta = m\ell\dot\theta^2.$$

法向反作用力

代入上面求得的 $\ddot\theta$，

$$N = mg\cos\theta - m\ell\ddot\theta = mg\cos\theta\left(1-\frac{\ell(3a+2\ell)}{2(3a^2+\ell^2)}\right)$$
$$= mg\cos\theta\,\frac{6a^2+2\ell^2-3a\ell-2\ell^2}{2(3a^2+\ell^2)} = \frac{3mga(2a-\ell)\cos\theta}{2(3a^2+\ell^2)}.$$

由于 $\ell<2a$，因子 $2a-\ell$ 为正；而当 θ 为锐角时 $\cos\theta>0$。所以 $N>0$；特别地，在 θ 为锐角期间，法向反作用力不为零，质点保持与杆接触。

开始滑动之时

由第二条方程及能量方程，

$$F = mg\sin\theta + m\ell\dot{\theta}^2 = mg\sin\theta\left(1 + \frac{\ell(3a+2\ell)}{3a^2+\ell^2}\right)$$
$$= mg\sin\theta\,\frac{3a^2+3a\ell+3\ell^2}{3a^2+\ell^2} = \frac{3mg(a^2+a\ell+\ell^2)\sin\theta}{3a^2+\ell^2}.$$

当 $0 < \theta < \frac{1}{2}\pi$ 时此式为正，所以摩擦力确实如假设般指向 P：若没有摩擦力，质点便会沿杆向外滑出。

> **想一想** 为甚么滑动的条件是 $F = \mu N$，而不是一条不等式?

在质点紧贴杆而不动时，摩擦力可提供不超过 μN 的任何大小的力，所以我们需要 $F \le \mu N$。当 $\theta = 0$ 时，$F = 0 < \mu N$。随着 θ 增加，F 增大（与 $\sin\theta$ 成正比），而 N 减小（与 $\cos\theta$ 成正比），因此 F/N（它是 $\tan\theta$ 的一个正倍数）稳定递增。质点在第一个使 $F = \mu N$ 的角度处即将滑动：

$$\frac{3mg(a^2+a\ell+\ell^2)\sin\theta}{3a^2+\ell^2} = \mu\,\frac{3mga(2a-\ell)\cos\theta}{2(3a^2+\ell^2)}.$$

约去 $3mg/(3a^2+\ell^2)$，再除以 $\cos\theta \neq 0$，得

$$\tan\theta = \frac{\mu a(2a-\ell)}{2(\ell^2+a\ell+a^2)},$$

此即所求。

若 $\ell > 2a$

N 的公式是在假设质点随杆一同运动的前提下推导的。若 $\ell > 2a$，公式在 $\theta = 0$ 时给出 $N < 0$，这是不可能的：杆只能推质点，不能拉质点。所以即使在释放的一刻，质点也不能随杆一同运动。

要了解实际发生的情况，可单独考虑杆。在释放的一刻对 P 取力矩，

$$3ma^2\,\ddot{\theta} = mg\cdot\tfrac{3}{2}a \quad\Longrightarrow\quad \ddot{\theta} = \frac{g}{2a},$$

因此杆上位于质点下方的点，起始时的向下加速度为 $\ell\ddot{\theta} = \frac{\ell}{2a}g$；当 $\ell > 2a$ 时，这大于 g。杆下落得比质点自由下落还要快。所以在释放的一刻，质点便离开杆（两者之间的反作用力为零），开始在重力作用下自由下落，而杆则独自摆下。

验算

- 对整个系统绕 P 取力矩，得 $(3ma^2+m\ell^2)\ddot{\theta} = mg\cos\theta\,(\frac{3}{2}a+\ell)$，与对能量方程求导所得的 $\ddot{\theta}$ 相同。

- 边界情况 $\ell = 2a$ 同时符合两种情形：释放时 $N = 0$，而杆上与 P 相距 $2a$ 的点起始加速度恰好为 g。
- 取 $a = \ell = 1$ 及 $\mu = 1$，滑动角满足 $\tan\theta = \frac{1}{6}$；在该角度，上述公式给出 $F/N = 6\tan\theta = 1 = \mu$。$N$ 和 F 的公式亦已用电脑验证：以数值方法逐步模拟运动，并直接由质点的位置计算其加速度。

常见错误

- 误用绕中点的转动惯量 $\frac{1}{12}m(3a)^2$，而非绕铰链的转动惯量；长度为 L 的杆绕一端的转动惯量是 $\frac{1}{3}mL^2$。
- 在摩擦力的方程中漏掉向心（centripetal）项 $m\ell\dot{\theta}^2$；正是这一项令分子成为 $a^2 + a\ell + \ell^2$，而不是 a^2。
- 在法向方程中把 $\ell\ddot{\theta}$ 的符号弄错：杆正以加速度离开质点，所以 N 小于 $mg\cos\theta$，而不是大于。

STEP 之外

最后一部分是一个经典的意外结果：下落中的铰接杆上的某一点，其加速度可以大于 g。这解释了为何放在下落间尺末端附近的硬币（或倾倒中的烟囱）会被“抛离”。要系统地求出 N 和 F 这类约束力（constraint forces），方法是在极坐标（polar coordinates）中写出运动方程；而日后学到的拉格朗日力学（Lagrangian mechanics）更会展示如何完全不涉及这些力而求得运动。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对刚体转动及这些极坐标下的加速度有细致的论述。

详细解答

STEP III 2017，第 11 题

力学 • 动量守恒

STEP III 2017，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/17-s3-q11

本题在 *Beyond STEP* 第 10.3 节火箭与枪炮中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch10/rockets.html

关键思路

系统不受任何外来的水平力，所以每次发射时水平动量（momentum）都守恒。题目中看似古怪的对地速率 $v - V$，其实说的是一件简单的事：抛射体（projectile）总是以相对于货车的速率 v 射出。至于动能（kinetic energy），不必每次都从头重新计算；应利用 $(*)$ 化简每次发射时的能量变化，然后再加起来。

解答

把货车、炮和抛射体都视为沿同一直线运动的质点（particle），并取向前方向（即货车运动的方向）为正。货车速率为 V 时发射的抛射体，其速度为 $-(v - V) = V - v$，即相对于货车的速度为 $-v$。发射属于系统的内部作用，而轨道是光滑的，所以每次发射时总水平动量都守恒。

第 (i) 部分

发射前一切静止，所以动量为零。由于 $V = 0$，每枚抛射体都以速度 $-v$ 射出，而货车（质量 M）以速率 u 向前运动：

$$Mu - nmv = 0 \quad \Longrightarrow \quad u = \frac{nmv}{M}.$$

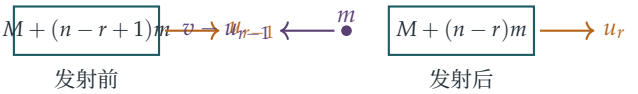
系统起初没有动能，所以所得的能量为

$$K = \tfrac{1}{2}Mu^2 + \tfrac{1}{2}nmv^2 = \tfrac{1}{2}\,\frac{n^2m^2v^2}{M} + \tfrac{1}{2}nmv^2 = \tfrac{1}{2}nmv^2\left(1 + \frac{nm}{M}\right),$$

即为所求。

第 (ii) 部分

考虑第 r 次发射。发射前一刻，货车载着 $n-r+1$ 枚未发射的抛射体，所以运动中的质量为 $M+(n-r+1)m$，速率为 u_{r-1}。发射后一刻，质量 $M+(n-r)m$ 以速率 u_r 运动，而一枚质量为 m 的抛射体的速度为 $u_{r-1}-v$。较早发射的抛射体不受影响，可以不予考虑。



由动量守恒得

$$\big(M+(n-r+1)m\big)u_{r-1} = \big(M+(n-r)m\big)u_r + m(u_{r-1}-v).$$

把左方写成 $\big(M+(n-r)m\big)u_{r-1} + mu_{r-1}$。于是 mu_{r-1} 项互相抵消，剩下

$$\big(M+(n-r)m\big)(u_r - u_{r-1}) = mv, \qquad \text{所以} \qquad u_r - u_{r-1} = \frac{mv}{M+(n-r)m},$$

这就是 $(*)$。

把 $(*)$ 对 $r = 1, \ldots, n$ 相加，左方伸缩相消（telescoping），且 $u_0 = 0$，所以

$$u_n = \sum_{r=1}^{n} \frac{mv}{M+(n-r)m}.$$

每个分母都满足 $M+(n-r)m \geq M$，所以每一项都不大于 $\dfrac{mv}{M}$。当 $r=1$ 时不等式严格成立，因为 $n>1$ 使 $(n-1)m>0$。因此

$$u_n < n\cdot\frac{mv}{M} = u.$$

想一想 为甚么逐枚发射的效果会较差?

在较早的发射中，炮不仅要推动货车，还要推动车上所有仍未发射的抛射体，所以每一发所带来的速率增量都较小。只有 $r=n$ 这一项（最后一发，车上已无其他负载）与同时发射时每枚所占的份额相同。

第 (iii) 部分

在第 r 次发射时，只有货车（连同其余负载）和刚发射的抛射体的速度会改变。所以

$$\begin{aligned} K_r - K_{r-1} &= \tfrac{1}{2}\big(M+(n-r)m\big)u_r^2 + \tfrac{1}{2}m(u_{r-1}-v)^2 - \tfrac{1}{2}\big(M+(n-r+1)m\big)u_{r-1}^2 \\ &= \tfrac{1}{2}\big(M+(n-r)m\big)(u_r^2-u_{r-1}^2) + \tfrac{1}{2}m\big[(v-u_{r-1})^2-u_{r-1}^2\big] \\ &= \tfrac{1}{2}\big(M+(n-r)m\big)(u_r-u_{r-1})(u_r+u_{r-1}) + \tfrac{1}{2}m(v^2-2vu_{r-1}). \end{aligned}$$

由 $(*)$，$\big(M+(n-r)m\big)(u_r-u_{r-1}) = mv$，所以第一项为 $\frac{1}{2}mv(u_r+u_{r-1})$。因此

$$\begin{aligned} K_r - K_{r-1} &= \tfrac{1}{2}mvu_r + \tfrac{1}{2}mvu_{r-1} + \tfrac{1}{2}mv^2 - mvu_{r-1} \\ &= \tfrac{1}{2}mv^2 + \tfrac{1}{2}mv(u_r-u_{r-1}), \end{aligned}$$

即为所求。

把上式对 $r=1,\dots,n$ 相加，并利用 $K_0=0$ 及 $u_0=0$，两边都伸缩相消：

$$K_n = \tfrac{1}{2}nmv^2 + \tfrac{1}{2}mv(u_n-u_0) = \tfrac{1}{2}nmv^2 + \tfrac{1}{2}mvu_n.$$

最后与第 (i) 部分比较。由于 $u = nmv/M$，

$$K = \tfrac{1}{2}nmv^2 + \tfrac{1}{2}\,\frac{n^2m^2v^2}{M} = \tfrac{1}{2}nmv^2 + \tfrac{1}{2}nmv\,u.$$

由第 (ii) 部分得 $u_n < u$，而因为 $u>0$ 且 $n>1$，所以 $u < nu$。由于 $mv>0$，

$$K_n = \tfrac{1}{2}nmv^2 + \tfrac{1}{2}mvu_n < \tfrac{1}{2}nmv^2 + \tfrac{1}{2}mvu < \tfrac{1}{2}nmv^2 + \tfrac{1}{2}nmvu = K.$$

所以 $K_n < K$：逐枚发射所释放的动能较少。

验算

- 取 $n=3$、$M=10$、$m=1$、$v=1$。则 $u=0.3$，而由 $(*)$ 得 $u_3 = \frac{1}{12}+\frac{1}{11}+\frac{1}{10} = \frac{181}{660} \approx 0.274$，比 u 小，正如预期。能量分别为 $K=1.95$ 及 $K_3 = \frac{2161}{1320} \approx 1.637$。
- 对 n、M、m 和 v 的数千组随机取值，在每一发之后直接计算每个物体的动量和动能，结果与 $(*)$ 及 K_n 的公式完全吻合；这已经由电脑验证。
- 极限情况：若 $m \ll M$，$(*)$ 的每一项都接近 mv/M，所以 $u_n \approx u$，而两个能量都接近 $\frac{1}{2}nmv^2$。当货车非常重时，发射的方式几乎没有影响。

常见错误

- 忘记未发射的抛射体随货车一起运动，在第 r 发之前用了质量 M 而不是 $M+(n-r+1)m$。

- 把抛射体的速度写成 $-v$ 而不是 $u_{r-1}-v$：题目规定的是它相对于地面的速率为 $v-V$，而不是 v。
- 在 K_r-K_{r-1} 中计入较早发射的抛射体的动能。它们不受影响，会互相抵消。

STEP 之外

这是火箭运动的离散版本。若把 n 发炮弹换成连续不断的废气喷流，u_n 的总和便会变成积分，并得出齐奥尔科夫斯基火箭方程（Tsiolkovsky rocket equation），$\Delta u = v\ln(m_0/m_1)$。火箭的处境就像逐枚发射的货车：它的燃料不可能在一瞬间全部烧完，所以要付出第 (ii) 部分所揭示的代价。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（《力学导论》）对变质量系统和火箭有详细的论述。

详细解答

STEP III 2017，第 12 题

概率与统计 • 离散联合分布

STEP III 2017，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/17-s3-q12

关键思路

一切都源自两条标准求和公式：$\sum_{r=1}^{n} r = \frac{1}{2}n(n+1)$ 及 $\sum_{r=1}^{n} r^2 = \frac{1}{6}n(n+1)(2n+1)$。首先令所有概率之和为 1，从而定出 k。然后对 y 求和，得出边际分布（marginal distribution）。至于协方差（covariance），我们需要 $\mathrm{E}(X)$ 和 $\mathrm{E}(XY)$。由于 $x+y$ 对 x 和 y 是对称的，这种对称性（symmetry）可以省去一半功夫。

解答

第 (i) 部分

所有 n^2 个数对 (x,y) 的概率之和必须为 1。由对称性，$\sum_{x,y} x = \sum_{x,y} y$。由于对 y 的 n 个取值，每个 x 都各出现一次，所以两者皆等于 $n\sum_{x=1}^{n} x = \frac{1}{2}n^2(n+1)$。因此

$$1 = k\sum_{x=1}^{n}\sum_{y=1}^{n}(x+y) = k\cdot 2\cdot \tfrac{1}{2}n^2(n+1) = k\,n^2(n+1), \qquad \text{故} \qquad k = \frac{1}{n^2(n+1)}.$$

求 X 的边际分布时，固定 x 并对 y 求和：

$$\begin{aligned}\mathrm{P}(X=x) &= k\sum_{y=1}^{n}(x+y) = k\Big(nx + \tfrac{1}{2}n(n+1)\Big) = \frac{kn\,(2x+n+1)}{2}\\ &= \frac{n+1+2x}{2n(n+1)},\end{aligned}$$

这正是所求结果。由于 $x+y$ 对 x 和 y 对称，Y 具有相同的分布：$\mathrm{P}(Y=y) = \dfrac{n+1+2y}{2n(n+1)}$。

? **想一想** 要证明 X 与 Y 不独立，是否需要逐一检查每个数对 (x,y)？

不需要。独立（independence）要求对每一个数对都有 $\mathrm{P}(X=x, Y=y)=\mathrm{P}(X=x)\,\mathrm{P}(Y=y)$，所以只要找到一个不成立的数对便足够。取 $x=y=1$：

$$\mathrm{P}(X=1,Y=1)=2k=\frac{2}{n^2(n+1)},\qquad \mathrm{P}(X=1)\,\mathrm{P}(Y=1)=\frac{(n+3)^2}{4n^2(n+1)^2}.$$

两者相等当且仅当 $8(n+1)=(n+3)^2$，即 $n^2-2n+1=0$，亦即 $(n-1)^2=0$。所以两者只在 $n=1$ 时相等。由于 $n\geq 2$，两个概率不相等，因此 X 与 Y **不独立**。

第 (ii) 部分

我们使用 $\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$，并留意由于两个变量分布相同，$\mathrm{E}(Y)=\mathrm{E}(X)$。

期望值（expectation）。利用第 (i) 部分的边际分布，

$$\begin{aligned}\mathrm{E}(X)&=\frac{1}{2n(n+1)}\sum_{x=1}^{n}\big((n+1)x+2x^2\big)=\frac{1}{2n(n+1)}\Big(\tfrac{1}{2}n(n+1)^2+\tfrac{1}{3}n(n+1)(2n+1)\Big)\\&=\frac{1}{2}\Big(\frac{n+1}{2}+\frac{2n+1}{3}\Big)=\frac{7n+5}{12}.\end{aligned}$$

乘积的期望值。这里 $xy(x+y)=x^2y+xy^2$，由对称性，两部分的二重和相等。此外，x^2y 的二重和可分解为两个单重和的乘积：

$$\begin{aligned}\mathrm{E}(XY)&=k\sum_{x=1}^{n}\sum_{y=1}^{n}xy(x+y)=2k\Big(\sum_{x=1}^{n}x^2\Big)\Big(\sum_{y=1}^{n}y\Big)=2k\cdot\frac{n(n+1)(2n+1)}{6}\cdot\frac{n(n+1)}{2}\\&=\frac{n^2(n+1)^2(2n+1)}{6\,n^2(n+1)}=\frac{(n+1)(2n+1)}{6}.\end{aligned}$$

协方差。以 144 为公分母，把结果合并：

$$\begin{aligned}\mathrm{Cov}(X,Y)&=\frac{(n+1)(2n+1)}{6}-\frac{(7n+5)^2}{144}=\frac{24(2n^2+3n+1)-(49n^2+70n+25)}{144}\\&=\frac{-n^2+2n-1}{144}=-\frac{(n-1)^2}{144}.\end{aligned}$$

由于 $n\geq 2$，$(n-1)^2>0$，所以 $\mathrm{Cov}(X,Y)=-\dfrac{(n-1)^2}{144}<0$。协方差为负，正如所求。协方差为负是合理的。联合概率只取决于 $x+y$，倾向于使总和较大。如果 X 已经很大，Y 便不必同样很大也能使总和变大，所以 Y 变大的“理由”稍为少了一点。

验算

- 取 $n=2$。此时 $k=\frac{1}{12}$，四个概率分别为 $\frac{2}{12},\frac{3}{12},\frac{3}{12},\frac{4}{12}$，总和为 1。边际概率为 $\mathrm{P}(X=1)=\frac{5}{12}$ 及 $\mathrm{P}(X=2)=\frac{7}{12}$，与公式吻合。于是 $\mathrm{E}(X)=\frac{19}{12}$，而 $\mathrm{E}(XY)=\frac{2+6+6+16}{12}=\frac{5}{2}$。由此得 $\mathrm{Cov}(X,Y)=\frac{5}{2}-\frac{361}{144}=-\frac{1}{144}$，与 $-(n-1)^2/144$ 一致。

- 边际概率之和为 1：$\sum_x (n+1+2x) = n(n+1) + n(n+1) = 2n(n+1)$。
- 令独立性检验失效的 $(n-1)^2$，正正也是令协方差为零的那个值。在两者中，$n=1$ 都是退化情况，此时必然有 $X = Y = 1$。
- 以电脑对 $n = 2, \ldots, 29$ 的精确分布作检验，确认了 k、边际分布、$\mathrm{E}(X)$、$\mathrm{E}(XY)$ 及协方差公式。

常见错误

- 忘记 $\sum_{x,y} x$ 中由对 y 求和而来的因子 n。这样会得出 $k = 1/(n(n+1))$，这是错误的。
- 试图“一般地”证明不独立，而不是给出一个明确的数对。单一个反例（counterexample）既已足够，也是最清晰的论证。
- 以为必须有非零协方差才能证明相依。在本题中非零协方差确实显示了相依，但协方差为零并不能证明独立。

STEP 之外

形如 $f(x,y) = g(x+y)$ 的联合分布（joint distribution），是一个简单的模型，描述通过总和而互相关联的变量。此时协方差的正负取决于 g 的形状。本题中 g 是线性的，协方差是一个很小的负数。在大学里，联合分布、边际分布、条件分布（conditional distribution）及相关（correlation）会一并学习，而连续情况则以二重积分代替二重和。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）以接近 STEP 的风格讲解这些内容。

详细解答

STEP III 2017，第 13 题

概率与统计 · 期望值与方差

STEP III 2017，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/17-s3-q13

关键思路

把 $X - x$ 写成 $(X - \mu) + (\mu - x)$。第一部分的平均值（mean）为零，所以平方后取期望值（expectation）时，交叉项（cross term）会消失。剩下的就是方差（variance）加上一个常数，这与“$\mathrm{E}(X^2) = \mathrm{Var}(X) + \mu^2$”是同一条恒等式。至于均匀分布（uniform distribution）的情况，Y 是 X 的函数，但并非一对一（one-to-one），所以求它的概率密度函数（probability density function）最稳妥的方法，是经由它的累积分布函数（cumulative distribution function）入手。

解答

函数 V

对任何固定的数 x，

$$\begin{aligned}\mathrm{V}(x) = \mathrm{E}\big((X-x)^2\big) &= \mathrm{E}\big(((X-\mu)+(\mu-x))^2\big)\\ &= \mathrm{E}\big((X-\mu)^2\big) + 2(\mu-x)\,\mathrm{E}(X-\mu) + (\mu-x)^2\\ &= \sigma^2 + (x-\mu)^2,\end{aligned}$$

这里用了期望值的线性性质（linearity of expectation）（此处 $\mu - x$ 是常数）以及 $\mathrm{E}(X - \mu) = 0$。

Y 的平均值

由于 $\mathrm{V}(x) = \sigma^2 + (x - \mu)^2$ 对每个实数 x 都成立，我们可以代入 $x = X$：

$$Y = \mathrm{V}(X) = \sigma^2 + (X-\mu)^2.$$

想一想 既然 V 是由期望值定义的，$\mathrm{V}(X)$ 岂不是只是一个数?

不是：V 是一个普通的实变量函数（一个关于 x 的二次式）。取期望值时 x 是固定不变的。把随机变量（random variable）X 代入这个二次式，便得到一个新的随机变量。所以

$$\mathrm{E}(Y) = \sigma^2 + \mathrm{E}\big((X-\mu)^2\big) = \sigma^2 + \sigma^2 = 2\sigma^2,$$

这就是 $(*)$。

均匀分布的情况

若 X 在 $[0,1]$ 上均匀分布，则 $\mu = \frac{1}{2}$，而 $\sigma^2 = \int_0^1 x^2\,\mathrm{d}x - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$。因此

$$\mathrm{V}(x) = \tfrac{1}{12} + \big(x - \tfrac{1}{2}\big)^2 = x^2 - x + \tfrac{1}{3}.$$

Y 的分布。 现在 $Y = \frac{1}{12} + (X - \frac{1}{2})^2$。当 X 取遍 $[0,1]$ 时，$|X - \frac{1}{2}|$ 取遍 $[0, \frac{1}{2}]$，所以 Y 的取值范围是 $[\frac{1}{12}, \frac{1}{12} + \frac{1}{4}] = [\frac{1}{12}, \frac{1}{3}]$。

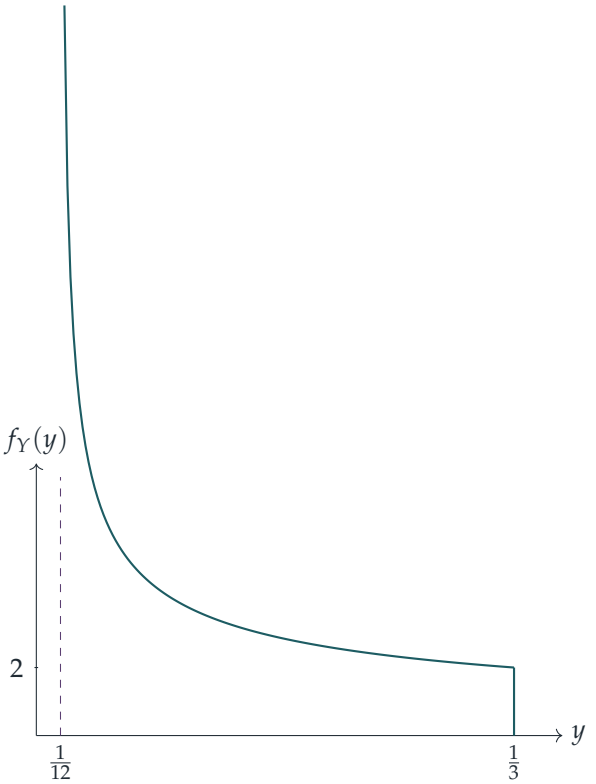
映射 $X \mapsto Y$ 是二对一的（点 $\frac{1}{2} \pm h$ 给出相同的 Y），所以我们改用分布函数处理。对于 $\frac{1}{12} \le y \le \frac{1}{3}$，

$$\begin{aligned}\mathrm{P}(Y \le y) &= \mathrm{P}\big((X-\tfrac{1}{2})^2 \le y - \tfrac{1}{12}\big) = \mathrm{P}\big(\tfrac{1}{2} - s \le X \le \tfrac{1}{2} + s\big), \quad \text{其中} s = \sqrt{y - \tfrac{1}{12}},\\ &= 2s = 2\sqrt{y - \tfrac{1}{12}},\end{aligned}$$

因为 $0 \le s \le \frac{1}{2}$，所以该区间位于 $[0,1]$ 之内，长度为 $2s$。（在 $y = \frac{1}{3}$ 时，这给出 $2 \cdot \frac{1}{2} = 1$，理应如此。）求导，得

$$f_Y(y) = \frac{1}{\sqrt{y - \frac{1}{12}}} \quad \text{当} \tfrac{1}{12} < y \le \tfrac{1}{3}, \qquad f_Y(y) = 0 \text{ 其他情况。}$$

这个概率密度函数在 $y = \frac{1}{12}$ 附近无界，但仍然可积（integrable）：Y 最有可能接近它的最小值，因为 $(X - \frac{1}{2})^2$ 在 $X = \frac{1}{2}$ 附近很平坦。



验证 $(*)$。 利用概率密度函数，并代换 $t = y - \frac{1}{12}$（于是 t 由 0 变到 $\frac{1}{4}$），

$$\mathrm{E}(Y) = \int_{1/12}^{1/3} \frac{y}{\sqrt{y - \frac{1}{12}}}\,\mathrm{d}y = \int_0^{1/4} \left(t^{1/2} + \tfrac{1}{12}t^{-1/2}\right)\mathrm{d}t = \left[\tfrac{2}{3}t^{3/2} + \tfrac{1}{6}t^{1/2}\right]_0^{1/4}$$
$$= \tfrac{2}{3}\cdot\tfrac{1}{8} + \tfrac{1}{6}\cdot\tfrac{1}{2} = \tfrac{1}{12} + \tfrac{1}{12} = \tfrac{1}{6}.$$

由于 $2\sigma^2 = 2\cdot\frac{1}{12} = \frac{1}{6}$，结果与 $(*)$ 吻合。

验算

- 概率密度函数的积分等于 1：$\int_0^{1/4} t^{-1/2}\,\mathrm{d}t = 2\cdot\frac{1}{2} = 1$。
- $\mathrm{V}(x)$ 在 $x = \mu$ 处取最小值，最小值为 σ^2：平均值正是使期望平方距离最小的那一点。在均匀分布的情况下，$\mathrm{V}(\frac{1}{2}) = \frac{1}{12}$，即 Y 取值范围的下端。
- 以 400 000 个均匀分布的数值进行电脑模拟，得到 Y 的样本平均值约为 0.1667，与 $\frac{1}{6}$ 吻合。

常见错误

- 把 $\mathrm{V}(X)$ 当作常数 σ^2，结果得出 $\mathrm{E}(Y) = \sigma^2$。
- 使用一对一的变量变换公式 $f_Y = f_X \left|\mathrm{d}x/\mathrm{d}y\right|$ 时只计算了一个分支，结果只得到正确概率密度函数的一半。其实 $X = \frac{1}{2} + s$ 和 $X = \frac{1}{2} - s$ 两者都有贡献。

STEP 之外

$\mathrm{V}(x) = \sigma^2 + (x - \mu)^2$ 是最简单的偏差–方差分解（bias–variance decomposition）。如果你用 x 去估计某个量，均方误差（mean squared error）就会分拆为数据本身的离散程度加上偏差的平方。结果 $\mathrm{E}(Y) = 2\sigma^2$ 说明，X 的两个独立副本之间的期望平方距离是 $2\sigma^2$。统计学家一再用到这个事实，例如在解释为何样本方差（sample variance）要除以 $n - 1$ 时。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（《概率导论》）涵盖了这两个概念，也讲解了如何求随机变量的非单调函数的概率密度函数。

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

2017 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。

