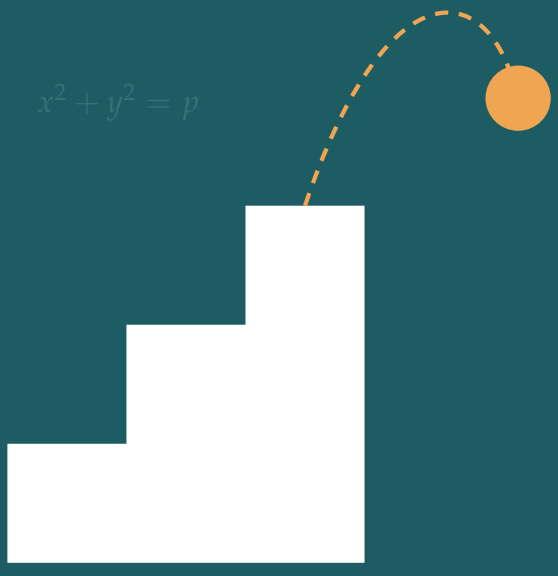
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

2012 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

2012 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

2012 年详细解答

2012 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 以微积分求最优值 纯数学 . . . 3

第 2 题 曲线描绘与根 纯数学 . . . 7

第 3 题 由面积推导不等式 纯数学 . . . 12

第 4 题 切线与法线 纯数学 . . . 16

第 5 题 代换积分法 纯数学 . . . 20

第 6 题 三角不等式 纯数学 . . . 23

第 7 题 周期性递推关系 纯数学 . . . 27

第 8 题 齐次微分方程 纯数学 . . . 31

第 9 题 抛体运动 力学 . . . 35

第 10 题 由落体与声音 力学 . . . 39

第 11 题 滑轮与摩擦力 力学 . . . 42

第 12 题 条件概率 概率与统计 . . . 46

第 13 题 以指示变量求期望值 概率与统计 . . . 49

STEP II

第 1 题 二项式级数的系数 纯数学 . . . 52

第 2 题 多项式的复合 纯数学 . . . 55

第 3 题 代换积分法 纯数学 . . . 59

第 4 题 对数级数不等式 纯数学 . . . 62

第 5 题 倒数函数的描绘 纯数学 . . . 66

第 6 题 圆内接四边形的面积 纯数学 . . . 71

第 7 题 向量与圆 纯数学 . . . 75

第 8 题 非线性递推关系 纯数学 . . . 78

第 9 题 抛体运动 力学 . . . 82

第 10 题 摩擦力与力矩 力学 . . . 86
第 11 题 非弹性碰撞 力学 . . . 90
第 12 题 条件概率 概率与统计 . . . 93
第 13 题 放松过程的最近距离 概率与统计 . . . 97

STEP III

第 1 题 非线性微分方程 纯数学 . . . 100
第 2 题 无穷乘积与级数 纯数学 . . . 104
第 3 题 相切曲线与四次方程 纯数学 . . . 108
第 4 题 无穷级数求和 纯数学 . . . 113
第 5 题 圆锥曲线上的有理点 纯数学 . . . 117
第 6 题 复数根的轨迹 纯数学 . . . 121
第 7 题 耦合微分方程 纯数学 . . . 126
第 8 题 斐波那契裂项级数 纯数学 . . . 130
第 9 题 转动滑轮的动力学 力学 . . . 134
第 10 题 圆周运动 力学 . . . 138
第 11 题 悬落的对折绳 力学 . . . 142
第 12 题 几何概率 概率与统计 . . . 146
第 13 题 截尾正态分布 概率与统计 . . . 150

详细解答

STEP I 2012，第 1 题

纯数学 · 以微积分求最优值

STEP I 2012，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：

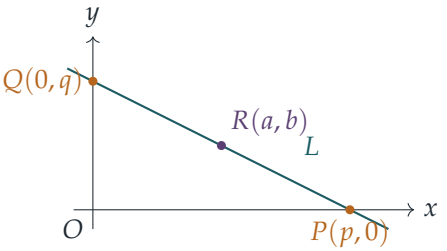
step.maths.org/questions/12-s1-q1

关键思路

一切都由一个数控制：斜率（gradient）。一旦规定 L 必须经过 R，截距（intercepts）p 和 q 便是 m 的简单函数（function），而两个要求最小化的距离也都变成只依赖 m 的函数。和 $p+q$ 可拆成一个常数加上形如 $am+b/m$ 的项，这正是算术－几何平均不等式（AM–GM）的典型用武之地（或者求一次导数也可以）。至于长度 PQ，留意到 $q=mp$ 会很有帮助，因为这样在求导之前 PQ^2 就能漂亮地因式分解。

解答

求 p 和 q



由于 $R(a,b)$ 在 $y=c-mx$ 上，我们有 $b=c-ma$，所以 $c=b+ma$。直线在 $x=0$ 处与 y 轴相交，所以 $q=c$；在 $y=0$ 处与 x 轴相交，所以 $p=c/m$（因为 $m>0$，这是可以的）。因此

$$p=a+\frac{b}{m}, \qquad q=b+ma.$$

对每个 $m > 0$，两者都是正数，符合题目要求。另外留意

$$q = b + ma = m\left(a + \frac{b}{m}\right) = mp,$$

这其实只是说 L 的斜率是 $-q/p$。下面会用到这一点。

最小化 $OP + OQ$

由于 $p, q > 0$，P 和 Q 到 O 的距离就是 p 和 q 本身，所以当 R 固定而 m 在 $m > 0$ 范围内变化时，我们要最小化

$$S(m) = p + q = a + b + \left(am + \frac{b}{m}\right).$$

括号内是两个正数之和，而它们的积 ab 与 m 无关。由算术－几何平均不等式，

$$am + \frac{b}{m} \geq 2\sqrt{am \cdot \frac{b}{m}} = 2\sqrt{ab},$$

等号成立当且仅当 $am = b/m$，即 $m = \sqrt{b/a}$（这是正数，所以这条直线是容许的）。因此

$$S(m) \geq a + b + 2\sqrt{ab} = (a^{\frac{1}{2}} + b^{\frac{1}{2}})^2,$$

而且这个下界可以达到。所以距离之和的最小值是 $(a^{\frac{1}{2}} + b^{\frac{1}{2}})^2$，正如所求。

（如果你偏好用微积分：$S'(m) = a - b/m^2$，它只在 $m = \sqrt{b/a}$ 处为零，代入后得到相同的值。）

最小化 PQ

由毕氏定理（Pythagoras），并利用 $q = mp$，

$$PQ^2 = p^2 + q^2 = p^2(1 + m^2) = \left(a + \frac{b}{m}\right)^2 (1 + m^2).$$

由于 $PQ > 0$，最小化 PQ 等同于最小化 $f(m) = PQ^2$。用乘积法则（product rule）和链式法则（chain rule）求导，

$$\begin{aligned} f'(m) &= 2\left(a + \frac{b}{m}\right)\left(-\frac{b}{m^2}\right)(1 + m^2) + 2m\left(a + \frac{b}{m}\right)^2 \\ &= 2\left(a + \frac{b}{m}\right)\left[-\frac{b}{m^2} - b + am + b\right] \\ &= 2\left(a + \frac{b}{m}\right)\left(am - \frac{b}{m^2}\right). \end{aligned}$$

? **想一想** $-b$ 在哪里被抵消了？

在方括号内，$-\frac{b}{m^2}(1+m^2) = -\frac{b}{m^2} - b$，而 $m(a+\frac{b}{m}) = am + b$；两个 b 互相抵消。正是因为这个抵消，我们才把 PQ^2 写成 $p^2(1+m^2)$，而不是直接展开 p^2+q^2。

第一个因子是正数，所以 $f'(m) = 0$ 当且仅当 $am = b/m^2$，即 $m^3 = b/a$，得到唯一的驻点（stationary point）

$$m = \left(\frac{b}{a}\right)^{\frac{1}{3}} = \frac{b^{\frac{1}{3}}}{a^{\frac{1}{3}}}.$$

（题目说我们可以假设它是极小值；事实上，m 较小时 $f' < 0$，m 较大时 $f' > 0$，所以它是 $m > 0$ 上的全局最小值（global minimum）。）

现在代入。取这个 m，

$$a + \frac{b}{m} = a + b\cdot\frac{a^{\frac{1}{3}}}{b^{\frac{1}{3}}} = a + a^{\frac{1}{3}}b^{\frac{2}{3}} = a^{\frac{1}{3}}(a^{\frac{2}{3}} + b^{\frac{2}{3}}), \qquad 1 + m^2 = 1 + \frac{b^{\frac{2}{3}}}{a^{\frac{2}{3}}} = \frac{a^{\frac{2}{3}} + b^{\frac{2}{3}}}{a^{\frac{2}{3}}}.$$

因此

$$PQ^2 = a^{\frac{2}{3}}(a^{\frac{2}{3}} + b^{\frac{2}{3}})^2 \cdot \frac{a^{\frac{2}{3}} + b^{\frac{2}{3}}}{a^{\frac{2}{3}}} = (a^{\frac{2}{3}} + b^{\frac{2}{3}})^3,$$

所以 P 与 Q 之间的最短距离是

$$PQ_{\min} = (a^{\frac{2}{3}} + b^{\frac{2}{3}})^{\frac{3}{2}}.$$

这就是题目所要求的“相似形式”：第一个答案括号内的指数是 $\frac{1}{2}$、括号外是 2，这个则是括号内 $\frac{2}{3}$、括号外 $\frac{3}{2}$。

验算

- 对称性：交换 a 和 b（即对 $y = x$ 作反射）必须令两个最小值都不变，而事实确实如此。
- 数值例子：取 $a = 1$、$b = 8$。第二个结果预测 $m = 2$，得 $p = 5$、$q = 10$ 及 $PQ = \sqrt{125} = 5\sqrt{5} \approx 11.180$，而 $(1+4)^{3/2} = 5\sqrt{5}$，果然吻合。至于距离之和，$m = \sqrt{8}$ 给出 $p + q = 9 + 4\sqrt{2} \approx 14.657 = (1+2\sqrt{2})^2$。
- 我们对 2000 组随机的 (a, b)，把两个最小值与对 m 的暴力搜索结果比较；两者相差约 10^{-8}。

常见错误

- 把 p^2+q^2 完全展开再求导：这样做没有错，但很繁琐，而且容易看漏因子 $am - b/m^2$。先看出 $q = mp$ 可以省下大量功夫。
- 忘记最小化 PQ 与最小化 PQ^2 之所以等价，只是因为 $PQ > 0$；以及在化简 $a + b/m$ 时把分数指数弄错。

STEP 之外

第二个结果其实是经典的“梯子转角”问题的变身：$(a^{2/3}+b^{2/3})^{3/2}$ 是能够水平地搬过宽度分别为 a 和 b 的两条走廊之间转角的最长梯子长度。长度固定的线段 PQ 所成的族的包络线（envelope）是一条星形线（astroid）$x^{2/3}+y^{2/3}=\ell^{2/3}$，而 R 恰好在包络线上，当且仅当 ℓ 等于这个最小值。包络线和约束最佳化（constrained optimisation）（以及之后的拉格朗日乘数（Lagrange multipliers））都会在大学一年级的微积分课程中出现；Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多同类型的题目。

详细解答

STEP I 2012，第 2 题

纯数学 · 曲线描绘与根

STEP I 2012，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/12-s1-q2

关键思路

数 $p(x)=0$ 的实根（real roots）个数，等同于数 $y=p(x)$ 的图像与 x 轴相交多少次；又或者把常数项移到另一边之后，数一条固定曲线与一条水平直线相交多少次。因此整道题的关键在于掌握一条四次曲线（quartic）的形状：它在哪里转向，以及转向时有多高或多低。第 (i) 部分中，这个四次多项式是一个完全平方；第 (ii) 部分中，关于导数的条件令 $\frac{\mathrm{d}y}{\mathrm{d}x}$ 可以因式分解并带有一个重根（repeated root），结果只剩下一个转向点（turning point）。

解答

第 (i) 部分

首先留意

$$y=x^4-6x^2+9=(x^2-3)^2,$$

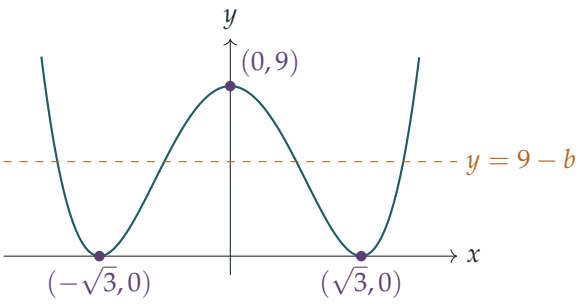
所以处处都有 $y\geq 0$，而且 $y=0$ 恰好在 $x=\pm\sqrt{3}$ 时成立。这条曲线是偶函数（even function）的图像（关于 y 轴对称），并在 $(0,9)$ 与 y 轴相交。求导，得

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

它在 $x=0$ 及 $x=\pm\sqrt{3}$ 时等于零。因此驻点（stationary points）为

$$(0,9)\text{（局部极大值）}, \qquad (\sqrt{3},0) \text{ 及 } (-\sqrt{3},0)\text{（极小值）}.$$

极小值是显然的，因为 $y\geq 0$ 而在这两点 $y=0$；又 $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=12x^2-12$ 在 $x=0$ 时等于 $-12<0$，所以该点是极大值。当 $x\to\pm\infty$ 时，$y\to+\infty$。这条曲线是一个“W”形，在 $x=\pm\sqrt{3}$ 处与 x 轴相切。



至于根的个数，在等式两边同时加上 9：

$$x^4 - 6x^2 + b = 0 \iff (x^2 - 3)^2 = 9 - b.$$

因此 n 就是水平直线 $y = 9 - b$ 与刚才所绘曲线的交点个数。让这条直线由高至低移动，从图中读出：

- $9 - b > 9$，即 $b < 0$：直线在极大值之上，只与外侧两支相交，所以 $n = 2$。
- $9 - b = 9$，即 $b = 0$：直线穿过极大值点，得到 $x = 0$ 及外侧两点；事实上 $x^2(x^2 - 6) = 0$ 给出 $x = 0, \pm\sqrt{6}$。所以 $n = 3$。
- $0 < 9 - b < 9$，即 $0 < b < 9$：直线与四段单调（monotonic）部分各相交一次，所以 $n = 4$。
- $9 - b = 0$，即 $b = 9$：直线只在两个极小值点 $x = \pm\sqrt{3}$ 处与曲线相切，所以 $n = 2$。
- $9 - b < 0$，即 $b > 9$：直线位于曲线下方，所以 $n = 0$。

综合以上：

(a) $n = 0 \iff b > 9$；　(b) $n = 1$：不存在这样的b；　(c) $n = 2 \iff b < 0$ 或$b = 9$；

(d) $n = 3 \iff b = 0$；　(e) $n = 4 \iff 0 < b < 9$.

$n = 1$ 不可能出现，也可由对称性得出：若 x 是根，则 $-x$ 也是根，所以唯一的根必须是 $x = 0$，这迫使 $b = 0$，但此时共有三个根。

第 (ii) 部分

现在 $y = x^4 - 6x^2 + ax + b$，所以

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 4x^3 - 12x + a, \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = 12x^2 - 12.$$

二阶导数只在 $x = \pm 1$ 时等于零。在 $x = 1$ 时一阶导数为 $a - 8$，在 $x = -1$ 时则为 $a + 8$。所以两者在同一点同时等于零，当且仅当

$$a = 8 \text{ (在}x = 1\text{)} \qquad \text{或} \qquad a = -8 \text{ (在}x = -1\text{)}.$$

情况 $a = 8$。此时

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

我们之所以找到因式 $(x-1)^2$，是因为我们刻意令 $x = 1$ 同时是 y' 和 y'' 的根，即 y' 的重根。

> ? **想一想** 这条曲线实际上有多少个转向点?

由于 $(x-1)^2 \geq 0$，$\frac{\mathrm{d}y}{\mathrm{d}x}$ 的正负号与 $x+2$ 的正负号相同。所以 y 在 $x < -2$ 时严格递减，在 $x > -2$ 时严格递增（斜率在 $x = 1$ 时为零，但不变号：那是一个驻拐点（stationary point of inflection），而不是转向点）。曲线只有一个转向点，就是在 $x = -2$ 处的全局极小值（global minimum），其值为

$$y(-2) = 16 - 24 - 16 + b = b - 24.$$

一条先下降到唯一的极小值再上升的曲线，若极小值在 x 轴之下，便与 x 轴相交两次；若极小值刚好在轴上，便相交一次（相切）；若在轴之上，则完全不相交。因此相异实根的个数为

$$\begin{cases} 2 & \text{若} b < 24, \\ 1 & \text{若} b = 24, \\ 0 & \text{若} b > 24. \end{cases}$$

（当 $b = 24$ 时，可以直接看到重根：$x^4 - 6x^2 + 8x + 24 = (x+2)^2(x^2 - 4x + 6)$，而该二次式的判别式（discriminant）为 $16 - 24 < 0$。）

情况 $a = -8$。以 $-x$ 代替 x，可把 $x^4 - 6x^2 - 8x + b$ 变成 $x^4 - 6x^2 + 8x + b$，所以这条曲线是前一条曲线关于 y 轴的反射。唯一的极小值现在位于 $x = 2$，其值为 $b - 24$，而根的个数相同：$b < 24$ 时为 2，$b = 24$ 时为 1，$b > 24$ 时为 0。

第 (iii) 部分

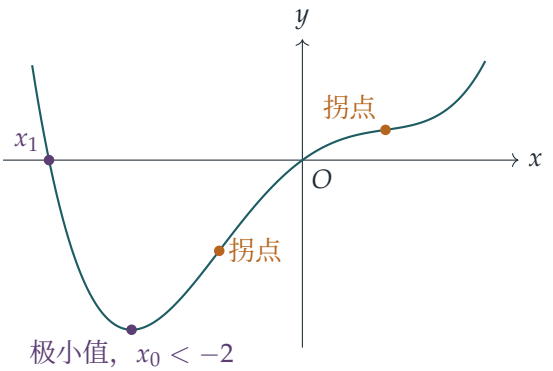
取 $y = x^4 - 6x^2 + ax$，其中 $a > 8$。

驻点。 $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$ 即 $x^3 - 3x = -\frac{a}{4}$。三次函数 $g(x) = x^3 - 3x$ 有局部极大值 $g(-1) = 2$ 及局部极小值 $g(1) = -2$，而且 $g(-2) = -2$。由于 $-\frac{a}{4} < -2$，高度为 $-\frac{a}{4}$ 的水平直线位于 g 的局部极小值之下，所以它与 g 恰好相交一次，交点在 $x < -2$ 的左支上。因此恰好有一个驻点，位于某个 $x_0 < -2$。该点左边斜率为负，右边斜率为正，所以它是极小值。

与 x 轴的交点。 $y = x(x^3 - 6x + a)$，所以 $x = 0$ 是其中一个交点，其余交点是 $h(x) = x^3 - 6x + a$ 的根。这个三次函数在 $x = \sqrt{2}$ 处取局部极小值，而 $h(\sqrt{2}) = a - 4\sqrt{2} > 0$，因为 $a > 8 > 4\sqrt{2}$。所以 h 只有一个实根，又因 $h(0) = a > 0$，这个根是负数。事实上 $h(-2) = a + 4 > 0$，而 h 在 $x < -\sqrt{2}$ 时递增，所以这个根 x_1 满足 $x_1 < -2$。曲线在 x_1 和原点与 x 轴相交，两者之间曲线取负值；极小值就位于这个区间内。

其他特征。在原点处斜率为 $a > 0$。二阶导数 $12(x^2-1)$ 在 $x = \pm 1$ 处变号，所以在 $(-1, -a-5)$ 及 $(1, a-5)$ 有拐点（points of inflection）；这两点的斜率分别为 $a+8$ 及 $a-8$，均为正数，所以曲线经过这两点时仍在上升。曲线在 $-1 < x < 1$ 上向下凹（concave down），在此区间以外向上凹。当 $x \to \pm\infty$ 时，$y \to +\infty$。

下图展示 $a = 10$ 的情况，此时 $x_0 \approx -2.05$，极小值约为 -28，而 $x_1 \approx -3.05$。



验算

- 在第 (i) 部分，$b = 5$ 应给出四个根：$x^2 = 3 \pm 2$，所以 $x = \pm 1, \pm\sqrt{5}$。结果吻合。
- 在第 (ii) 部分，$b = 24$ 给出因式分解 $(x+2)^2(x^2-4x+6)$，证实在极小值处只有一个（二重）根。
- 数值上，我们对 0、9 及 24 两侧的许多 b 值求出这些四次多项式的全部复数根，并数出其中相异的实根；每一个结果都吻合。对于 8 与 200 之间的随机 a，第 (iii) 部分的曲线总是只有一个驻点及两个与 x 轴的交点，并且 $x_1 < x_0 < -2$。当 $a = 10$ 时，$x_0 \approx -2.054$，$y(x_0) \approx -28.05$。

常见错误

- 把 $x = \pm\sqrt{3}$ 处的重根（当 $b = 9$ 时）算作四个根：题目问的是相异的 x 值。
- 把第 (ii) 部分中 $y' = y'' = 0$ 的点当作转向点。由于 y' 在该处有一个平方因式，其正负号不会改变，所以它是驻拐点。
- 遗漏 $a = -8$：$y'' = 0$ 有两个解，每个解都给出一个 a 值。

STEP 之外

当 b 经过 0、9 或 24 时实根个数的跳变，是分岔（bifurcation）的一个小例子：只有当水平直线经过某个驻值时，答案才会改变。第 (ii) 部分中 $y' = y'' = 0$ 这个特殊情况，正是两个驻点合并并消失之处，这是最简单的“突变”（catastrophe），即折叠型突变（fold），也是为甚么一旦 $a > 8$，第 (iii) 部分的四次曲线便只有一个转向点。同样的推理，借助介值定理（intermediate value theorem）和单调性加以严格化，就是实分析（real analysis）的起点；Abbott 的 *Understanding Analysis*（《理解分析》）是看到这套推理被仔细展开的好地方。

详细解答

STEP I 2012，第 3 题

纯数学 · 由面积推导不等式

STEP I 2012，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/12-s1-q3

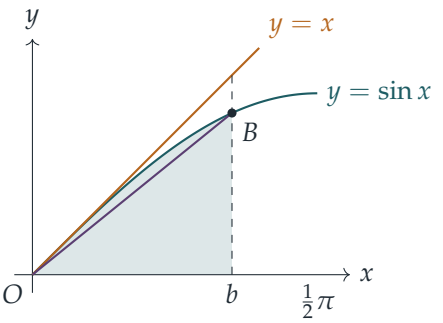
关键思路

一条始终朝同一方向弯曲的曲线，总是位于它每一条切线（tangent）的同一侧，并位于它每一条弦（chord）的另一侧。因此，曲线下的面积被夹在切线下的面积与弦下的面积之间，而这两个面积都不过是三角形或梯形（trapezium）的面积。把每个面积准确地写出来，不等式便自然得出；真正需要花心思的，只是决定该用哪一条切线和哪一条弦。

解答

第 (i) 部分

在 $0 < x < \frac{1}{2}\pi$ 上，我们有 $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\sin x = -\sin x < 0$，所以曲线 $y = \sin x$ 在该区间上是凹（concave）的：它向下弯曲。原点处的切线斜率（gradient）为 $\cos 0 = 1$，所以切线就是直线 $y = x$。由 O 到 $B = (b, \sin b)$ 的弦是直线 $y = \dfrac{\sin b}{b}\,x$。



凹性的意思是：在 O 与 B 之间，曲线严格位于弦的上方；而当 $x > 0$ 时，曲线严格位于 O 处切线的下方。所以，当 $0 < x < b$ 时，

$$\frac{\sin b}{b}\, x \;<\; \sin x \;<\; x.$$

现在比较在 $0 \le x \le b$ 上的三个面积。

- 曲线下的面积（阴影部分）：$\displaystyle\int_0^b \sin x \,\mathrm{d}x = 1 - \cos b.$
- 切线下的面积：一个底为 b、高为 b 的直角三角形，面积为 $\frac{1}{2}b^2$。
- 弦下的面积：一个底为 b、高为 $\sin b$ 的三角形，面积为 $\frac{1}{2}b\sin b$。

由于在 $0 < x < b$ 上曲线严格位于两条直线之间（而且三者都是连续的），这些面积满足严格不等式：

$$\tfrac{1}{2}b\sin b \;<\; 1 - \cos b \;<\; \tfrac{1}{2}b^2.$$

右边的不等式可整理为 $\cos b > 1 - \frac{1}{2}b^2$，左边的则可整理为 $\cos b < 1 - \frac{1}{2}b\sin b$。合起来便得

$$1 - \tfrac{1}{2}b^2 < \cos b < 1 - \tfrac{1}{2}b\sin b,$$

正是所求。

第 (ii) 部分

设 $k = \ln a$；由于 $a > 1$，所以 $k > 0$，而且 $a^x = \mathrm{e}^{kx}$。于是

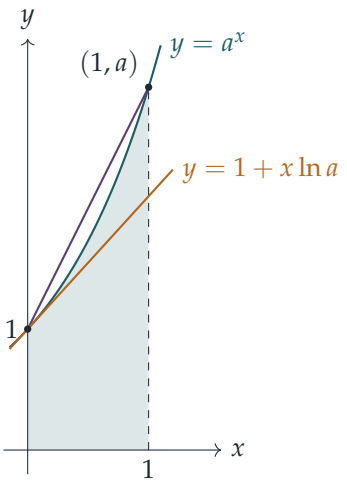
$$\frac{\mathrm{d}}{\mathrm{d}x}a^x = k\,\mathrm{e}^{kx}, \qquad \frac{\mathrm{d}^2}{\mathrm{d}x^2}a^x = k^2\mathrm{e}^{kx} > 0,$$

所以这条曲线是凸（convex）的：它向上弯曲。现在两者的角色对调：曲线严格位于它的弦的下方，并严格位于它的切线的上方（切点除外）。

想一想 我们应该在哪个区间上比较面积?

答案应该只涉及 $\ln a$ 和 a，而不涉及其他量。区间 $0 \le x \le 1$ 恰好做到这一点：曲线的两个端点是 $(0,1)$ 和 $(1,a)$，而曲线下的面积是

$$\int_0^1 \mathrm{e}^{kx}\,\mathrm{d}x = \frac{\mathrm{e}^k - 1}{k} = \frac{a-1}{\ln a}.$$



下界（弦）。由 $(0,1)$ 到 $(1,a)$ 的弦在 $0<x<1$ 上位于曲线上方，而它下方的区域是一个宽为 1、两条平行边分别为 1 和 a 的梯形，面积为 $\frac{1}{2}(1+a)$。因此

$$\frac{a-1}{\ln a} < \frac{a+1}{2} \quad\Longrightarrow\quad \ln a > \frac{2(a-1)}{a+1},$$

这里我们两边同乘正数 $2\ln a/(a+1)$，所以不等号方向不变。

上界（切线）。$(0,1)$ 处的切线斜率为 $k=\ln a$，所以切线是 $y=1+kx$。当 $x>0$ 时它位于曲线下方，而它在 $[0,1]$ 上的下方面积是一个两条平行边分别为 1 和 $1+k$ 的梯形，面积为 $1+\frac{1}{2}k$。因此

$$1+\tfrac{1}{2}k < \frac{a-1}{k}.$$

两边乘以 $2k>0$：

$$2k+k^2 < 2a-2 \quad\Longleftrightarrow\quad k^2+2k+1 < 2a-1 \quad\Longleftrightarrow\quad (k+1)^2 < 2a-1.$$

? **想一想** 我们可以直接开平方吗?

可以，但只因为两边都是正数：$2a-1>1>0$，而且由于 $k>0$，所以 $k+1>0$。对正数而言，$u^2<v^2$ 可推出 $u<v$，所以 $k+1<\sqrt{2a-1}$，即

$$\frac{2(a-1)}{a+1} < \ln a < -1+\sqrt{2a-1},$$

证明完毕。

验算

- 第 (i) 部分取 $b=1$：$1-\frac{1}{2}=0.5$，$\cos 1\approx 0.5403$，而 $1-\frac{1}{2}\sin 1\approx 0.5793$。大小次序正如所述。

- 第 (ii) 部分取 $a = 2$：$\frac{2}{3} \approx 0.6667$，$\ln 2 \approx 0.6931$，而 $\sqrt{3} - 1 \approx 0.7321$。次序同样正确。
- 当 $a \to 1$ 时，第 (ii) 部分的三个式子都趋于 0，而且彼此愈夹愈紧：在 $a = 1.01$ 时，它们准确至小数点后五位都相同。这很合理，因为在很短的范围内，一条光滑曲线几乎是直的。
- 我们在数以千计的随机值上验算了两串不等式，并用辛普森法则（Simpson's rule）验算了面积公式。

常见错误

- 把位置的上下弄反：正弦曲线是凹的（位于切线下方、弦上方），而 a^x 是凸的（情况相反）。应该用二阶导数（second derivative）而不是草图来证明这一点。
- 用 $\ln a$ 乘或除不等式时，忘记 $\ln a > 0$ 这一条件。
- 对 $(k+1)^2 < 2a - 1$ 开平方时，没有检查 $k + 1 > 0$。

STEP 之外

这些弦的界是一个一般原理的特例：对于凸函数，它在一个区间上的平均值介乎区间中点的函数值与两端点函数值的平均之间（这就是埃尔米特（Hermite）–阿达马（Hadamard）不等式；若改用中点处而非端点处的切线，便可得到更精确的下界）。第 (ii) 部分的下界说明，a 与 1 的对数平均（logarithmic mean）$(a-1)/\ln a$ 小于它们的算术平均（arithmetic mean）；这个事实在整个分析学以至物理学中都屡屡出现。Bartle 和 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）对凸函数和积分估计有细致的论述。

详细解答

STEP I 2012，第 4 题

纯数学 · 切线与法线

> **STEP I 2012，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/12-s1-q4

关键思路

一切都源于 $y = \frac{1}{2x}$ 的斜率（gradient），再加上细心的消元（elimination）。当我们求两条只差在把 p 与 q 互换的直线的交点时，把两条方程相减，总会得出一个因子 $p - q$；由于两点相异，我们可以把它约去。在特殊情况下，值得留意条件对 Q 的影响：它令 Q 成为 P 关于直线 $y = x$ 的镜像（mirror image），而这种对称性（symmetry）正好解释了为何 T 和 N 都落在这条直线上。

解答

点 T

在 C 上有 $y = \dfrac{1}{2x}$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -\dfrac{1}{2x^2}$。因此 P 点的切线（tangent）斜率为 $-\frac{1}{2p^2}$，方程为

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

两边乘以 $2p^2$，得 $2p^2y - p = -x + p$，即

$$x + 2p^2y = 2p. \tag{1}$$

同理，Q 点的切线为

$$x + 2q^2y = 2q. \tag{2}$$

以 (1) 减去 (2)，

$$2(p^2 - q^2)\,y = 2(p - q).$$

由于两点相异，$p \neq q$，所以可以除以 $2(p-q)$，得 $(p+q)y = 1$。因为 $p, q > 0$，所以 $p+q \neq 0$，故

$$y = \frac{1}{p+q}.$$

代入 (1)，

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

所以 T 是点 $\left(\frac{2pq}{p+q}, \frac{1}{p+q}\right)$，正如所求。

当 $pq = \frac{1}{2}$ 时的点 N

P 点的法线（normal）与切线垂直，所以其斜率为 $2p^2$，方程为

$$y = 2p^2 x - 2p^3 + \frac{1}{2p}. \tag{3}$$

同样地，Q 点的法线为 $y = 2q^2 x - 2q^3 + \frac{1}{2q}$。令两个 y 的表达式相等：

$$2(p^2 - q^2)\, x = 2(p^3 - q^3) - \left(\frac{1}{2q} - \frac{1}{2p}\right) = 2(p^3 - q^3) + \frac{p-q}{2pq}.$$

除以 $2(p-q) \neq 0$，

$$(p+q)\, x = p^2 + pq + q^2 + \frac{1}{4pq}.$$

这对任何相异的 p 和 q 都成立。现在运用 $pq = \frac{1}{2}$，于是 $\frac{1}{4pq} = \frac{1}{2} = pq$，右边变成

$$p^2 + 2pq + q^2 = (p+q)^2.$$

因此 $x = p + q$。代入 (3)，

$$y = 2p^2(p+q) - 2p^3 + \frac{1}{2p} = 2p^2 q + \frac{1}{2p} = p\,(2pq) + q = p + q,$$

这里两次用到 $2pq = 1$：$2p^2 q = p(2pq) = p$，以及 $\frac{1}{2p} = q$。因此

$$N = (p+q,\ p+q).$$

（注意 $p \neq q$ 加上 $pq = \frac{1}{2}$，意味着 $p \neq \frac{1}{\sqrt{2}}$，这是对 p 的唯一限制。）

想一想 为甚么 T 和 N 偏偏会落在 $y = x$ 上?

当 $pq = \frac{1}{2}$ 时，有 $\frac{1}{2q} = p$ 及 $\frac{1}{2p} = q$，所以 $P = (p, q)$、$Q = (q, p)$。曲线 $xy = \frac{1}{2}$ 在关于 $y = x$ 的反射下保持不变，而这个反射把 P 和 Q 互换。因此它把两条切线互换，也把两条法线互换，于是每个交点都被映射到自身；而被这个反射固定的点必定在 $y = x$ 上。代数计算正好印证了这一点。

T 与 N 在 $y = x$ 上，以及距离之积

由于 $2pq = 1$，T 的结果给出

$$T = \left(\frac{1}{p+q}, \frac{1}{p+q} \right),$$

它在 $y = x$ 上；而我们已求得 $N = (p+q, p+q)$，它同样在 $y = x$ 上。

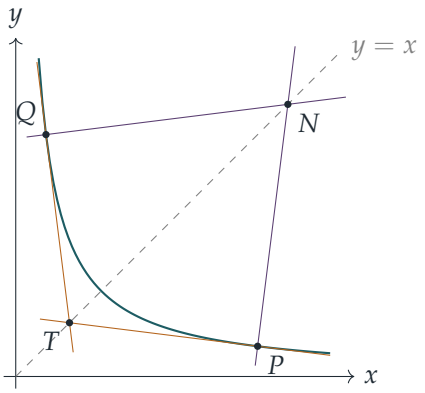
一点 (a, a)（其中 $a > 0$）与原点的距离为 $\sqrt{2}\,a$。由于 $p + q > 0$，

$$OT = \frac{\sqrt{2}}{p+q}, \qquad ON = \sqrt{2}\,(p+q),$$

所以

$$OT \cdot ON = 2,$$

这是常数（与 p 无关），正如所求。



图中显示 $p = 2$、$q = \frac{1}{4}$ 的情况：琥珀色为切线，梅紫色为法线。

验算

取 $p = 2$、$q = \frac{1}{4}$，则 $pq = \frac{1}{2}$。切线为 $x + 8y = 4$ 和 $8x + y = 4$，交于 $\left(\frac{4}{9}, \frac{4}{9}\right)$，与 $\frac{1}{p+q} = \frac{4}{9}$ 吻合。法线为 $y = 8x - \frac{63}{4}$ 和 $y = \frac{x}{8} + \frac{63}{32}$，交于 $\left(\frac{9}{4}, \frac{9}{4}\right)$，与 $p + q = \frac{9}{4}$ 吻合。于是 $OT \cdot ON = \frac{4\sqrt{2}}{9} \cdot \frac{9\sqrt{2}}{4} = 2$。我们亦以精确的有理数运算，对 2000 组随机选取的 p 和 q 验证了 T 的一般公式，以及特殊情况下的两个结果。

常见错误

- 除以 $p - q$ 时没有说明为何可以这样做：“相异的点”这个条件正是用在这里。

- 求 N 时，在 $\frac{1}{2q} - \frac{1}{2p} = \frac{p-q}{2pq}$ 中弄错正负号。
- 在声称 T 位于 $y = x$ 上之前，忘记用 $pq = \frac{1}{2}$ 化简 T。

STEP 之外

直角双曲线（rectangular hyperbola）$xy = c^2$ 是一种圆锥曲线（conic）；而 $OT \cdot ON$ 为常数这一事实，说明了 T 和 N 是关于圆 $x^2 + y^2 = 2$ 的反演点（inverse points）。关于圆的反演（inversion）是几何学以及复分析（complex analysis）中一种威力强大的变换；在复分析中，它以映射 $z \mapsto 1/\bar{z}$ 的形式出现。Needham 的 *Visual Complex Analysis*（《可视化复分析》）对反演及其几何性质有精彩的论述。

详细解答

STEP I 2012，第 5 题

纯数学 · 代换积分法

STEP I 2012，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/12-s1-q5

关键思路

头两个积分都可以用标准方法处理。把 $\sin 2x = 2\sin x\cos x$ 写出来，便看出 $\sin x$ 是 $\cos x$ 的导数（derivative）的相反数，所以代换（substitution）$u = \cos x$ 行得通。至于第二个积分，对数（logarithm）是麻烦的因子，因此我们用分部积分（integration by parts），通过求导把它消去。最后一个积分中，$\cos x + \sin x$ 其实是一个“乔装”的余弦：$\sqrt{2}\cos(x - \frac{1}{4}\pi)$。把变量平移 $\frac{1}{4}\pi$，积分区间便变成 $[0, \frac{1}{4}\pi]$，而被积函数（integrand）则变成刚才求得的两个积分的组合。

解答

第一个积分

记第一个积分为 I_1。由于 $\sin 2x = 2\sin x\cos x$，代入 $u = \cos x$，于是 $\mathrm{d}u = -\sin x\,\mathrm{d}x$。当 $x = 0$ 时 $u = 1$；当 $x = \frac{1}{4}\pi$ 时 $u = \frac{1}{\sqrt{2}}$。所以

$$I_1 = \int_0^{\frac{1}{4}\pi} 2\cos x \ln(\cos x)\,\sin x\,\mathrm{d}x = \int_{1/\sqrt{2}}^{1} 2u\ln u\,\mathrm{d}u,$$

其中 $\mathrm{d}u$ 带来的负号已用来对调上下限。用分部积分（对 $\ln u$ 求导，对 $2u$ 积分），

$$\int 2u\ln u\,\mathrm{d}u = u^2\ln u - \int u^2\cdot\frac{1}{u}\,\mathrm{d}u = u^2\ln u - \tfrac{1}{2}u^2.$$

在 $u = 1$ 处，其值为 $-\frac{1}{2}$。在 $u = \frac{1}{\sqrt{2}}$ 处，有 $u^2 = \frac{1}{2}$ 及 $\ln u = -\frac{1}{2}\ln 2$，所以其值为 $-\frac{1}{4}\ln 2 - \frac{1}{4}$。因此

$$I_1 = -\tfrac{1}{2} - \left(-\tfrac{1}{4}\ln 2 - \tfrac{1}{4}\right) = \tfrac{1}{4}\ln 2 - \tfrac{1}{4} = \tfrac{1}{4}(\ln 2 - 1),$$

与所求相符。

第二个积分

记第二个积分为 I_2。我们用分部积分，对 $\ln(\cos x)$ 求导，对 $\cos 2x$ 积分。$\cos 2x$ 的一个方便的原函数（antiderivative）是 $\frac{1}{2}\sin 2x = \sin x \cos x$，而 $\ln(\cos x)$ 的导数是 $-\sin x / \cos x$。所以

$$\begin{aligned} I_2 &= \Big[\sin x \cos x \, \ln(\cos x)\Big]_0^{\frac{1}{4}\pi} - \int_0^{\frac{1}{4}\pi} \sin x \cos x \cdot \left(-\frac{\sin x}{\cos x}\right) \mathrm{d}x \\ &= \tfrac{1}{2}\ln\left(\tfrac{1}{\sqrt{2}}\right) - 0 + \int_0^{\frac{1}{4}\pi} \sin^2 x \, \mathrm{d}x. \end{aligned}$$

想一想 为甚么选用 $\sin x \cos x$，而不直接写成 $\frac{1}{2}\sin 2x$?

两者是同一个函数，但写成 $\sin x \cos x$ 的形式时，因子 $\cos x$ 会与 $\ln(\cos x)$ 的导数分母中的 $\cos x$ 互相抵消，剩下容易计算的 $\sin^2 x$ 的积分。

边界项（boundary term）是 $\frac{1}{2}\cdot(-\frac{1}{2}\ln 2) = -\frac{1}{4}\ln 2$。利用 $\sin^2 x = \frac{1}{2}(1-\cos 2x)$，

$$\int_0^{\frac{1}{4}\pi} \sin^2 x \, \mathrm{d}x = \Big[\tfrac{1}{2}x - \tfrac{1}{4}\sin 2x\Big]_0^{\frac{1}{4}\pi} = \tfrac{\pi}{8} - \tfrac{1}{4}.$$

所以

$$I_2 = -\tfrac{1}{4}\ln 2 + \tfrac{\pi}{8} - \tfrac{1}{4} = \tfrac{1}{8}(\pi - 2\ln 2 - 2) = \tfrac{1}{8}(\pi - \ln 4 - 2),$$

与所求相符。

最后一个积分

记最后一个积分为 I。由复角公式（compound-angle formula），

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

这提示我们使用代换 $t = x - \frac{1}{4}\pi$，它把区间 $[\frac{1}{4}\pi, \frac{1}{2}\pi]$ 变成 $[0, \frac{1}{4}\pi]$，即 I_1 和 I_2 的积分区间。由 $2x = 2t + \frac{1}{2}\pi$，

$$\cos 2x = \cos\left(2t + \tfrac{1}{2}\pi\right) = -\sin 2t, \qquad \sin 2x = \sin\left(2t + \tfrac{1}{2}\pi\right) = \cos 2t,$$

以及

$$\ln(\cos x + \sin x) = \ln\left(\sqrt{2}\cos t\right) = \tfrac{1}{2}\ln 2 + \ln(\cos t),$$

这是成立的，因为在 $[0, \frac{1}{4}\pi]$ 上 $\cos t > 0$。因此

$$\begin{aligned} I &= \int_0^{\frac{1}{4}\pi} (\cos 2t - \sin 2t)\left(\tfrac{1}{2}\ln 2 + \ln(\cos t)\right) \mathrm{d}t \\ &= \tfrac{1}{2}\ln 2 \int_0^{\frac{1}{4}\pi} (\cos 2t - \sin 2t)\, \mathrm{d}t + I_2 - I_1. \end{aligned}$$

余下的积分等于零：

$$\int_0^{\frac{1}{4}\pi} (\cos 2t - \sin 2t)\,\mathrm{d}t = \left[\tfrac{1}{2}\sin 2t + \tfrac{1}{2}\cos 2t\right]_0^{\frac{1}{4}\pi} = \tfrac{1}{2} - \tfrac{1}{2} = 0.$$

所以，利用上面两个结果，

$$I = I_2 - I_1 = \tfrac{1}{8}(\pi - 2\ln 2 - 2) - \tfrac{1}{8}(2\ln 2 - 2) = \frac{\pi - 4\ln 2}{8} = \frac{\pi}{8} - \frac{1}{2}\ln 2.$$

验算

用辛普森法则（Simpson's rule）算得 $I_1 \approx -0.076713$、$I_2 \approx -0.030588$ 及 $I \approx 0.046125$，与 $\frac{1}{4}(\ln 2 - 1)$、$\frac{1}{8}(\pi - \ln 4 - 2)$ 及 $\frac{1}{8}(\pi - 4\ln 2)$ 吻合至小数点后十位。对第一个积分作一个快速的正负号检查：在 $(0, \frac{1}{4}\pi)$ 上有 $\sin 2x > 0$ 及 $\ln(\cos x) < 0$，所以 I_1 必定是负数，而 $\ln 2 - 1 < 0$ 正好如此。

常见错误

- 在 I_2 的分部积分中漏掉边界项。它在 $x = \frac{1}{4}\pi$ 处并不为零。
- 把 $\cos(2t + \frac{1}{2}\pi)$ 和 $\sin(2t + \frac{1}{2}\pi)$ 的正负号弄反，结果得到 $I_1 - I_2$ 或 $I_1 + I_2$，而不是 $I_2 - I_1$。
- 把 $\ln\sqrt{2}$ 带来的常数 $\frac{1}{2}\ln 2$ 直接丢掉，却没有检查它的系数积分后是否为零。

STEP 之外

$\ln(\cos x)$ 和 $\ln(\sin x)$ 的积分是经典题材：同样的对称技巧可以得出著名的结果 $\int_0^{\pi/2} \ln(\sin x)\,\mathrm{d}x = -\frac{1}{2}\pi\ln 2$；而在一般的积分区间上，它们会引出克劳森函数（Clausen function）和卡塔兰常数（Catalan's constant）。在大学一年级的分析（analysis）课程中，你还要检查当对数在端点处趋于无穷时，这类积分是否有定义，这使它们成为反常积分（improper integral）。如想多做这类练习，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）。

详细解答

STEP I 2012，第 6 题

纯数学 · 三角不等式

STEP I 2012，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/12-s1-q6

关键思路

这里每个仰角（angle of elevation）的对边都是同一条旗杆，所以每个余切（cotangent）都只是一段水平距离除以旗杆高度。这样，第一个结果就变成关于圆中线段长度的事实：$DC^2 = AD \cdot DB$，因为半圆上的圆周角（angle in a semicircle）是直角。三角恒等式（trigonometric identity）把 $\cot p \cot q - \cot^2 \frac{1}{2}(p+q)$ 化成一个能直接看出正负号的式子。得到这个不等式后，几何部分几乎不费吹灰之力就给出最后的结果。

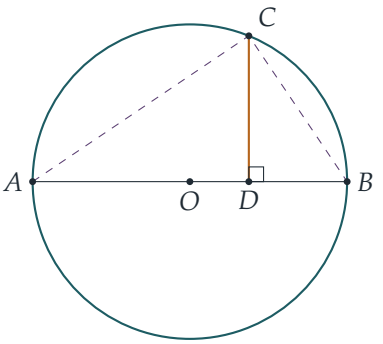
解答

几何结果

设旗杆高度为 $h > 0$，顶端为 T，因此 DT 是铅垂的。A、B、C 都在地面上，所以三角形 TDA、TDB、TDC 在 D 处都是直角。因此

$$\cot\alpha = \frac{AD}{h}, \qquad \cot\beta = \frac{DB}{h}, \qquad \cot\phi = \frac{DC}{h}.$$

由于 α 和 β 都是锐角，D 既不是 A 也不是 B（从这两点中任何一点看，仰角都会是直角），所以 D 严格位于 A 与 B 之间，且 $DC > 0$。



现在看地面平面图。设 O 为圆心，r 为半径，$x = OD$（以指向 B 的方向为正，所以 $-r < x < r$）。于是 $AD = r + x$，$DB = r - x$，而在三角形 ODC 中应用毕氏定理（Pythagoras′ theorem），得 $DC^2 = OC^2 - OD^2 = r^2 - x^2$。因此

$$AD \cdot DB = (r + x)(r - x) = r^2 - x^2 = DC^2.$$

（等价地说：$\angle ACB$ 是半圆上的圆周角，所以是直角，而 CD 是直角三角形 ACB 的高（altitude）。）两边除以 h^2，

$$\cot\alpha \cot\beta = \frac{AD \cdot DB}{h^2} = \frac{DC^2}{h^2} = \cot^2\phi,$$

得证。

恒等式

记 $s = \frac{1}{2}(p + q)$，并利用二倍角公式（double-angle formulae）的形式 $\sin^2 s = \frac{1}{2}\big(1 - \cos(p + q)\big)$ 及 $\cos^2 s = \frac{1}{2}\big(1 + \cos(p + q)\big)$。左边变成

$$\begin{aligned}
&\tfrac{1}{2}\cos p\cos q\big(1 - \cos(p + q)\big) - \tfrac{1}{2}\sin p\sin q\big(1 + \cos(p + q)\big)\\
&\quad = \tfrac{1}{2}\big(\cos p\cos q - \sin p\sin q\big) - \tfrac{1}{2}\cos(p + q)\big(\cos p\cos q + \sin p\sin q\big)\\
&\quad = \tfrac{1}{2}\cos(p + q) - \tfrac{1}{2}\cos(p + q)\cos(p - q),
\end{aligned}$$

这里用了 $\cos(p + q)$ 和 $\cos(p - q)$ 的复角公式（compound-angle formulae）。此式对所有 p 和 q 都成立。

不等式

把恒等式的右边因式分解：

$$\cos p\cos q\sin^2 s - \sin p\sin q\cos^2 s = \tfrac{1}{2}\cos(p + q)\big(1 - \cos(p - q)\big).$$

现设 $p, q > 0$ 且 $p + q \le \frac{1}{2}\pi$。于是 $0 < p + q \le \frac{1}{2}\pi$，所以 $\cos(p + q) \ge 0$；而 $1 - \cos(p - q) \ge 0$ 永远成立。所以右边非负，从而

$$\cos p\cos q\sin^2 s \;\ge\; \sin p\sin q\cos^2 s.$$

想一想 我们可以放心地把两边除以 $\sin p \sin q \sin^2 s$ 吗?

可以：由于 p 和 q 都是正数，且两者之和不超过 $\frac{1}{2}\pi$，所以 p、q 和 s 都在 $(0, \frac{1}{2}\pi)$ 之内，因此 $\sin p$、$\sin q$ 和 $\sin s$ 都严格为正。除以正数 $\sin p \sin q \sin^2 s$ 不会改变不等号的方向：

$$\cot p \cot q \ \geq \ \cot^2 s = \cot^2 \tfrac{1}{2}(p+q).$$

等号成立当且仅当 $p = q$ 或 $p + q = \frac{1}{2}\pi$。

最后结果

取 $p = \alpha$、$q = \beta$。两者都是正数（它们是一条高度为正的旗杆的仰角），且 $\alpha + \beta \leq \frac{1}{2}\pi$，所以上述不等式适用。结合第一个结果，

$$\cot^2 \phi = \cot \alpha \cot \beta \ \geq \ \cot^2 \tfrac{1}{2}(\alpha+\beta).$$

ϕ 和 $\frac{1}{2}(\alpha+\beta)$ 都在 $(0, \frac{1}{2}\pi)$ 之内（ϕ 是因为 $DC > 0$ 且 $h > 0$），所以两个余切都是正数，可以开平方：$\cot \phi \geq \cot \frac{1}{2}(\alpha+\beta)$。

想一想 为甚么这给出的是 $\phi \leq \frac{1}{2}(\alpha+\beta)$，而不是 $\geq$?

因为 cot 在 $(0, \frac{1}{2}\pi)$ 上是严格递减（decreasing）的：余切越大，角就越小。因此

$$\phi \leq \tfrac{1}{2}(\alpha+\beta) \qquad \text{只要} \alpha+\beta \leq \tfrac{1}{2}\pi.$$

验算

- 若 $D = O$，则 $AD = DB = DC = r$，所以 $\alpha = \beta = \phi$：第一个结果的两边相等，而最后的不等式取等号，与 $p = q$ 的情况吻合。
- 数值测试：取 $r = 1$、$OD = \frac{1}{2}$、$h = 1$，得 $\cot \alpha = 1.5$、$\cot \beta = 0.5$、$\cot^2 \phi = 0.75$，与预期一致。这里 $\alpha + \beta \approx 0.588 + 1.107 = 1.695 > \frac{1}{2}\pi$，所以最后结果不一定适用；事实上 $\phi \approx 0.857 > \frac{1}{2}(\alpha+\beta) \approx 0.848$，说明条件是必要的。
- 我们在 10^5 组随机的 (p, q) 上验证了恒等式，在容许范围内 2×10^5 组随机数对上验证了不等式，并在约 8×10^4 个满足 $\alpha + \beta \leq \frac{1}{2}\pi$ 的随机配置中验证了 $\phi \leq \frac{1}{2}(\alpha+\beta)$，全部没有出错。

常见错误

- 忘记 cot 是递减的，因而得出 $\phi \geq \frac{1}{2}(\alpha+\beta)$。

- 把不等式除以 $\sin p \sin q \sin^2 s$，却没有说明它为何是正数。
- 忽略条件 $p + q \leq \frac{1}{2}\pi$：正是这个条件保证 $\cos(p + q) \geq 0$；没有它，不等式可以不成立。

STEP 之外

这个不等式说明 $\ln \cot t$ 在 $(0, \frac{1}{4}\pi]$ 上是凸函数（convex function），因为它表示 $\ln \cot p$ 与 $\ln \cot q$ 的平均值不小于两者平均值处的 $\ln \cot$ 值。凸性与琴生不等式（Jensen's inequality）是大学一年级分析学中不等式的一个强大来源。关系式 $DC^2 = AD \cdot DB$ 正是欧几里得（Euclid）作几何平均数（geometric mean）的方法：$\cot \phi$ 是 $\cot \alpha$ 与 $\cot \beta$ 的几何平均数。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（《实分析导论》）讨论了凸函数，而 Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）则有更多这类风格的题目。

详细解答

STEP I 2012，第 7 题

纯数学 · 周期性递推关系

STEP I 2012，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/12-s1-q7

关键思路

一个以周期（period）1、2 或 3 重复的数列（sequence），由其首几项完全决定，所以递推关系（recurrence relation）只需对一个周期内少数几组连续三项成立（再加上首尾相接的那几组）。把这些条件逐一写出，便得到一组关于 x、y、z 的小型联立线性方程（simultaneous linear equations）。这组方程有一个不全为零的解，正是这一点迫使 p 和 q 必须满足所求的条件。整个过程中，$x \neq 0$ 这个事实是我们能够约去的关键。

解答

第 (i) 部分

若对每个 n 都有 $t_n = x$，则递推关系给出 $x = px + qx$，即 $(p+q-1)x = 0$。由于 $x \neq 0$，得 $p + q = 1$。

反过来，若 $p + q = 1$ 且 $t_0 = t_1 = x$，则 $t_2 = (p+q)x = x$，由归纳法（induction）可知每一项都等于 x。x 不受任何限制，因此它可以是任何非零实数。

第 (ii) 部分

模式 $x, y, x, y, \ldots$ 满足递推关系，当且仅当两组连续三项 (x, y, x) 和 (y, x, y) 满足递推关系：

$$x = py + qx, \qquad \text{即} \quad (1-q)x = py, \tag{1}$$

$$y = px + qy, \qquad \text{即} \quad (1-q)y = px. \tag{2}$$

把 (1) 和 (2) 相加及相减，得

$$(1-q-p)(x+y)=0, \qquad (1-q+p)(x-y)=0.$$

$x+y$ 和 $x-y$ 不可能同时为零，否则 $x=0$。若 $x+y\neq 0$，则 $p+q=1$；若 $x-y\neq 0$，则 $q-p=1$。因此 $q+p=1$ 和 $q-p=1$ 之中至少有一个成立，得证。

接着推导 x 与 y 之间的关系。

- 若 $q+p=1$，则 $1-q=p$，(1) 变成 $px=py$。所以 $x=y$，前提是 $p\neq 0$。
- 若 $q-p=1$，则 $1-q=-p$，(1) 变成 $-px=py$。所以 $x=-y$，同样前提是 $p\neq 0$。

唯一的例外是 $p=0$，此时两种情况都给出 $q=1$。当 $p=0$ 且 $q=1$ 时，递推关系为 $t_{n+2}=t_n$，任何模式 $x,y,x,y,\ldots$ 都满足它，因此 x 与 y 之间没有任何关系。故除非 $p=0$ 且 $q=1$，否则 $x=y$ 或 $x=-y$。

第 (iii) 部分

现在模式是 $x,y,z,x,y,z,\ldots$，我们需要三组连续三项 (x,y,z)、(y,z,x) 和 (z,x,y)：

$$z=py+qx, \tag{3}$$
$$x=pz+qy, \tag{4}$$
$$y=px+qz. \tag{5}$$

利用 (3) 从另外两条方程中消去 z：

$$\begin{aligned} &(4):\quad x=p(py+qx)+qy &&\Longrightarrow\quad (pq-1)x+(p^2+q)y=0,\\ &(5):\quad y=px+q(py+qx) &&\Longrightarrow\quad (p+q^2)x+(pq-1)y=0. \end{aligned}$$

把第一条方程乘以 $pq-1$，第二条乘以 p^2+q，然后相减。含 y 的项互相抵消，剩下

$$\left[(pq-1)^2-(p^2+q)(p+q^2)\right]x=0.$$

想一想 为甚么括号内的式子必须为零？这组方程会不会根本没有解？

它们确实有解：数列存在，所以 x 和 y 同时满足两条方程。而 $x\neq 0$，所以可以除以 x。展开括号，

$$\begin{aligned} (pq-1)^2-(p^2+q)(p+q^2) &= p^2q^2-2pq+1-\left(p^3+p^2q^2+pq+q^3\right)\\ &= 1-3pq-p^3-q^3, \end{aligned}$$

所以 $p^3+q^3+3pq-1=0$，得证。

因式分解（factorisation）。这个式子是 $a^3+b^3+c^3-3abc$，其中 $a=p$、$b=q$、$c=-1$，而标准恒等式（identity）

$$a^3+b^3+c^3-3abc=(a+b+c)\cdot\tfrac{1}{2}\big[(a-b)^2+(b-c)^2+(c-a)^2\big]$$

给出

$$p^3+q^3+3pq-1=\tfrac{1}{2}(p+q-1)\big[(p-q)^2+(q+1)^2+(p+1)^2\big].$$

（如果记不起这个恒等式，可以展开来验证：方括号等于 $2(p^2+q^2+1-pq+p+q)$，再乘以 $p+q-1$ 便还原为原来的三次式。）乘积为零，所以 $p+q=1$ 或 $(p-q)^2+(p+1)^2+(q+1)^2=0$。

情况 $(p-q)^2+(p+1)^2+(q+1)^2=0$。实数平方之和为零，当且仅当每个平方都为零，所以 $p=q=-1$。此时 (3) 变成 $z=-y-x$，即 $x+y+z=0$。

情况 $p+q=1$。把 $q=1-p$ 代入 (3)、(4) 和 (5)，并把每条整理成差的形式：

$$z-x=p(y-x),\qquad x-y=p(z-y),\qquad y-z=p(x-z).$$

设 $a=y-x$、$b=z-y$、$c=x-z$，于是 $a+b+c=0$。第一条和第三条方程表示 $c=-pa$ 及 $b=-pc=p^2a$。于是

$$0=a+b+c=a(1-p+p^2).$$

> ? **想一想** $1-p+p^2$ 可以为零吗?

不可以：对每个实数 p，$1-p+p^2=\left(p-\frac{1}{2}\right)^2+\frac{3}{4}>0$。所以 $a=0$，由此 $c=-pa=0$，进而 $b=0$。即 $x=y=z$。

两种情况不会重叠，因为 $p=q=-1$ 给出 $p+q=-2\neq 1$。所以 $x=y=z$（当 $p+q=1$ 时）或 $x+y+z=0$（当 $p=q=-1$ 时）。两者都确实会出现：在第二种情况中，递推关系 $t_{n+2}=-t_{n+1}-t_n$ 对任何初始值都给出 $x,\ y,\ -x-y,\ x,\ y,\ \ldots$。

验算

- 取 $p=q=-1$、$t_0=2$ 及 $t_1=5$，数列为 $2,5,-7,2,5,-7,\ldots$，而 $2+5-7=0$。
- 取 $p=3$ 及 $q=4$（所以 $q-p=1$），由 $1,-1$ 开始得 $t_2=-3+4=1$ 及 $t_3=3-4=-1$，正如第 (ii) 部分所预测般交替出现。
- 当 $p=2$、$q=-1$ 时，三次式给出 $8-1-6-1=0$，与 $p+q=1$ 一致。我们亦在 1000 组随机 (p,q) 上验证了展开式、因式分解，以及三次式与 (3)–(5) 的 3×3 行列式（determinant）一致。模拟递推关系亦证实了上述每一种情况。

常见错误

- 在第 (ii) 部分断定 $p+q=1$ 及 $q-p=1$ 同时成立。论证只表明至少有一个成立；两者同时成立只在例外情况 $p=0$、$q=1$ 中发生。
- 在第 (ii) 部分除以 p，却没有察觉 $p=0$ 正是题目所问的例外情况。
- 忘记首尾相接的方程 (4) 和 (5)。每个周期只用一条方程，不足以迫出所求的条件。

STEP 之外

处理 $t_{n+2}=pt_{n+1}+qt_n$ 的标准方法是通过特征方程（characteristic equation）$\lambda^2=p\lambda+q$：一般而言 $t_n=A\alpha^n+B\beta^n$，其中 α 和 β 是它的根。一个数列每 k 步重复一次，当且仅当它所用到的根满足 $\lambda^k=1$。常数数列需要根 1（所以 $p+q=1$），交替数列需要根 -1（所以 $q-p=1$）。情况 $p=q=-1$ 给出 $\lambda^2+\lambda+1=0$，其根是复数单位立方根（cube roots of unity），这正是为甚么它们的和 $1+\omega+\omega^2=0$ 会以 $x+y+z=0$ 的形式重现。这其实是乔装了的线性代数（linear algebra）（作用于数列上的移位映射（shift map）），而斯特朗（Strang）的 *Introduction to Linear Algebra*（《线性代数导论》）正是以这种方式，通过特征值（eigenvalues）来处理递推关系。

详细解答

STEP I 2012，第 8 题

纯数学 · 齐次微分方程

STEP I 2012，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/12-s1-q8

关键思路

两条方程都是齐次（homogeneous）的：除了一个整体的 x 幂次之外，每一项关于 x 和 y 的总次数都相同。若设 $y = xv$，那个 x 的幂次便会消去，方程随即分离成 v 的部分和 x 的部分。第 (i) 部分会一步步带你完成这个过程。第 (ii) 部分则留给你自己处理：把方程除以 x，便可看出它是同一类方程，而之后关于 v 的积分需要用到部分分式（partial fractions）。

解答

第 (i) 部分

若 $y = xv$，由乘积法则（product rule）得 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = v + x\dfrac{\mathrm{d}v}{\mathrm{d}x}$。代入，得

$$x(xv)\left(v + x\frac{\mathrm{d}v}{\mathrm{d}x}\right) + x^2v^2 - 2x^2 = x^3v\frac{\mathrm{d}v}{\mathrm{d}x} + 2x^2v^2 - 2x^2 = 0.$$

由于 $x \neq 0$，我们可以除以 x^2，得

$$xv\frac{\mathrm{d}v}{\mathrm{d}x} + 2v^2 - 2 = 0,$$

此即所求。

接着我们分离变量（separate the variables）。方程表明 $xv\,\dfrac{\mathrm{d}v}{\mathrm{d}x} = 2(1 - v^2)$。只要 $v^2 \neq 1$，

$$\int \frac{v}{1 - v^2}\,\mathrm{d}v = \int \frac{2}{x}\,\mathrm{d}x \quad \Longrightarrow \quad -\tfrac{1}{2}\ln|1 - v^2| = 2\ln|x| + c.$$

两边乘以 -2 再取指数，得 $|1 - v^2| = Ax^{-4}$，其中 $A > 0$ 为某常数，因此

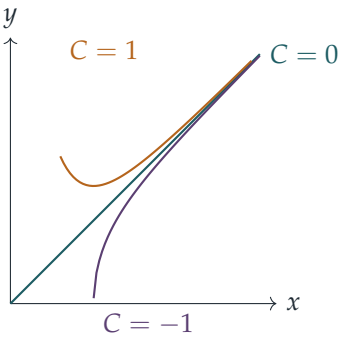
$$x^4(v^2 - 1) = C,$$

其中 $C = \pm A$ 是非零常数。先前搁置的情况 $v = \pm 1$（即 $y = \pm x$）也是解：此时 $\mathrm{d}v/\mathrm{d}x = 0$ 且 $2v^2 - 2 = 0$。它们恰好就是 $C = 0$ 的情况，所以 C 可以是任何常数。

最后，代入 $v = y/x$：

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

这正是所要求的形式。



上图展示了第一象限中的几条解曲线。正如 $y^2 = x^2 + C/x^2$ 所提示，当 $x \to \infty$ 时，它们都趋近直线 $y = x$。

第 (ii) 部分

$6x + 5y$ 中每一项的次数都是 1，$y\,\mathrm{d}y/\mathrm{d}x$ 的次数也是 1。这提示我们用同样的代换 $y = xv$：

$$xv\left(v + x\frac{\mathrm{d}v}{\mathrm{d}x}\right) + 6x + 5xv = 0.$$

除以 $x \neq 0$ 并整理各项，得

$$xv\frac{\mathrm{d}v}{\mathrm{d}x} = -(v^2 + 5v + 6) = -(v+2)(v+3).$$

想一想 怎样求 $\dfrac{v}{(v+2)(v+3)}$ 的积分？

用部分分式。设 $v = P(v+3) + Q(v+2)$。代入 $v = -2$ 得 $P = -2$，代入 $v = -3$ 得 $Q = 3$。所以

$$\frac{v}{(v+2)(v+3)} = \frac{3}{v+3} - \frac{2}{v+2}.$$

只要 $v \neq -2, -3$，分离变量得

$$\int\left(\frac{3}{v+3} - \frac{2}{v+2}\right)\mathrm{d}v = -\int\frac{\mathrm{d}x}{x} \quad\Longrightarrow\quad 3\ln|v+3| - 2\ln|v+2| = -\ln|x| + c.$$

因此 $|x|\,|v+3|^3/(v+2)^2$ 是常数。去掉绝对值符号（在一条 $v \neq -3$ 的解曲线上，$x(v+3)^3$ 的正负号不会改变），得

$$\frac{x(v+3)^3}{(v+2)^2} = K.$$

现在代入 $v = y/x$。分子是 $x(y+3x)^3/x^3$，分母是 $(y+2x)^2/x^2$，所以 x 的幂次完全消去：

$$\boxed{(y+3x)^3 = K(y+2x)^2},$$

其中 K 是常数。

我们排除了的直线也是解。若 $y = -3x$，则 $y\,\mathrm{d}y/\mathrm{d}x = 9x$ 而 $6x+5y = -9x$，所以方程成立。这条直线就是方框中形式在 $K = 0$ 时的情况。若 $y = -2x$，则 $y\,\mathrm{d}y/\mathrm{d}x = 4x$ 而 $6x+5y = -4x$，所以这条直线也是解。它不对应任何有限的 K，而是对应 $K \to \infty$。若把解写成 $(y+2x)^2 = k(y+3x)^3$，则可改为以 $k = 0$ 涵盖它。

验算

- 第 (i) 部分：对 $x^2y^2 - x^4 = C$ 作隐函数微分（implicit differentiation），得 $2xy^2 + 2x^2y\,y' - 4x^3 = 0$。除以 $2x$ 便还原为原方程。
- 第 (ii) 部分：我们从多个随机起点，以数值方法（Runge–Kutta 方法）逐步求解方程。沿每条解曲线，$(y+3x)^3/(y+2x)^2$ 这个量都保持不变，误差约为 10^{-14}。例如，通过 $(1,1)$ 的曲线有 $K = 64/9$。
- 每条方程中通过原点的直线解 $y = \pm x$ 和 $y = -2x,\ -3x$，分别来自 $2v^2 - 2 = 0$ 和 $v^2 + 5v + 6 = 0$ 的根。这是一个好迹象，显示关于 v 的方程是正确的。

常见错误

- 忘记乘积法则：$\mathrm{d}y/\mathrm{d}x$ 是 $v + x\,\mathrm{d}v/\mathrm{d}x$，而不是 $x\,\mathrm{d}v/\mathrm{d}x$。
- 部分分式算错（例如写成 $2/(v+2) - 3/(v+3)$）。可以代入一个值（例如 $v = 0$）检验分解是否正确。
- 除以 $1 - v^2$ 或 $(v+2)(v+3)$ 时没有指出这样会遗失某些直线解。

STEP 之外

形如 $\mathrm{d}y/\mathrm{d}x = F(y/x)$ 的方程称为齐次方程。缩放变换 $(x,y) \mapsto (\lambda x, \lambda y)$ 会把它的解曲线互相映射，这正是 $v = y/x$ 奏效的原因。这是索菲斯·李（Sophus Lie）提出的一个深刻思想的最简单例子：微分方程（differential equation）的对称性（symmetry）可以

用来降低方程的阶（order）。直线解 $y = mx$ 就是在缩放下保持不变的曲线。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（《初等微分方程》）讲解了齐次方程及其他一阶方程。

详细解答

STEP I 2012，第 9 题

力学 • 抛体运动

STEP I 2012，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：

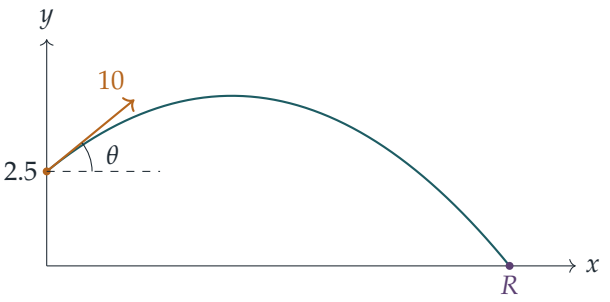
step.maths.org/questions/12-s1-q9

关键思路

这是一个从高处抛出的抛射体（projectile），所以飞行时间（time of flight）由一条关于 t 的二次方程（quadratic equation）给出，而题目的重点其实是把它的根改写成以 $c = \cos 2\theta$ 表示。倍角公式（double-angle formulae）$\sin^2\theta = \frac{1}{2}(1-c)$ 和 $\cos^2\theta = \frac{1}{2}(1+c)$ 包办了所有工作。一旦射程（range）写成单一变量 c 的函数（function），求它的最大值就只是普通的微积分，只需在平方时稍加小心。

解答

我们把铅球模拟为只受重力作用而自由运动的质点（particle），不计空气阻力（air resistance）。取原点在抛射点正下方的地面上，x 轴水平，y 轴铅直向上，并假设 $0 \le \theta < 90°$（铅球向前及向上抛出），因此 $\sin\theta \ge 0$ 而 $\cos\theta > 0$。



飞行时间

水平速度保持为 $10\cos\theta$，而铅直方向的运动初速为 $10\sin\theta$、加速度为 -10，所以在时间 t 时

$$x = 10t\cos\theta, \qquad y = 2.5 + 10t\sin\theta - 5t^2.$$

铅球在 $y = 0$ 时落地，即当

$$5t^2 - 10t\sin\theta - 2.5 = 0 \quad\Longleftrightarrow\quad t^2 - 2t\sin\theta - \tfrac{1}{2} = 0.$$

两根之积为 $-\frac{1}{2} < 0$，所以恰有一个正根，就是取加号的那一个：

$$T = \sin\theta + \sqrt{\sin^2\theta + \tfrac{1}{2}}.$$

现在利用 $\sin^2\theta = \frac{1}{2}(1-\cos 2\theta) = \frac{1}{2}(1-c)$。由于 $\sin\theta \geq 0$，我们可以写 $\sin\theta = \sqrt{\frac{1}{2}(1-c)}$，而 $\sin^2\theta + \frac{1}{2} = \frac{1}{2}(2-c)$。因此

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

这就是所求的时间（以秒为单位）。

射程

同样地，$\cos^2\theta = \frac{1}{2}(1+c)$，且 $\cos\theta > 0$，所以 $\cos\theta = \sqrt{\frac{1}{2}(1+c)}$。射程为

$$\begin{aligned} R = 10T\cos\theta &= 10\sqrt{\frac{1+c}{2}}\cdot\frac{1}{\sqrt{2}}\left(\sqrt{1-c} + \sqrt{2-c}\right) \\ &= 5\left(\sqrt{1-c^2} + \sqrt{(1+c)(2-c)}\right) = 5\left(\sqrt{1-c^2} + \sqrt{2+c-c^2}\right)\ \text{m}. \end{aligned}$$

当 θ 取遍 $[0°, 90°)$ 时，$c = \cos 2\theta$ 取遍 $(-1, 1]$，所以我们要在这个区间上求 R 的最大值。

求射程的最大值

对于 $-1 < c < 1$，求导得

$$\frac{\mathrm{d}R}{\mathrm{d}c} = 5\left(\frac{-c}{\sqrt{1-c^2}} + \frac{1-2c}{2\sqrt{2+c-c^2}}\right).$$

驻点（stationary point）须满足

$$\frac{c}{\sqrt{1-c^2}} = \frac{1-2c}{2\sqrt{2+c-c^2}}. \qquad (*)$$

想一想 我们可以直接把 $(*)$ 平方吗？可能会出甚么问题？

平方可能会引入增根（extraneous roots），因为它丢失了正负号的信息。两个分母都是正数，所以在 $(*)$ 中，c 与 $1-2c$ 必须同号（或同时为零，但这是不可能的）。这迫使 $0<c<\frac{1}{2}$，我们要记住这一点。平方并去分母：

$$\begin{aligned}4c^2(2+c-c^2) &= (1-2c)^2(1-c^2)\\ 8c^2+4c^3-4c^4 &= 1-4c+3c^2+4c^3-4c^4\\ 5c^2+4c-1 &= 0\\ (5c-1)(c+1) &= 0.\end{aligned}$$

根 $c=-1$ 不在 $0<c<\frac{1}{2}$ 之内（它对应铅直向上抛），所以唯一的驻点是 $c=\frac{1}{5}$。它确实满足 $(*)$：由于 $\sqrt{1-\frac{1}{25}}=\frac{2\sqrt{6}}{5}$ 及 $\sqrt{2+\frac{1}{5}-\frac{1}{25}}=\frac{3\sqrt{6}}{5}$，两边都等于 $\frac{1}{2\sqrt{6}}$。

在这一点

$$R\left(\tfrac{1}{5}\right)=5\left(\frac{2\sqrt{6}}{5}+\frac{3\sqrt{6}}{5}\right)=5\sqrt{6}\approx 12.25.$$

要确认这是最大值，可与区间的端点比较：当 $c\to -1$ 时 $R\to 0$，而 $R(1)=5\sqrt{2}\approx 7.07$。由于 R 在 $[-1,1]$ 上连续（continuous），并在区间内部可微（differentiable），其最大值必在端点或驻点取得；而唯一的驻点给出这些值中最大的一个。所以当 $c=\frac{1}{5}$ 时射程最大，得证。（这对应 $\theta=\frac{1}{2}\arccos\frac{1}{5}\approx 39.2^\circ$，比 45° 略小。）

多出的距离

在 45° 时，$c=\cos 90^\circ=0$，所以

$$R(0)=5\left(1+\sqrt{2}\right).$$

因此多出的距离是

$$5\sqrt{6}-5\left(1+\sqrt{2}\right)=5\left(\sqrt{6}-\sqrt{2}-1\right)\ \text{m},$$

得证。

验算

- 在 $\theta=0$（$c=1$）时，时间为 $\frac{1}{\sqrt{2}}$，与一个从 2.5 m 高的平台滚落的球所满足的 $2.5=5t^2$ 吻合。
- 当 $c\to -1$（铅直向上抛）时，射程趋向 0，理应如此。

- 以数值计算，多出的距离为 $5(2.4495 - 1.4142 - 1) \approx 0.176\,\text{m}$，只有约 18 cm：最佳点附近非常平坦。对飞行过程作逐步数值模拟，以及对 $R(c)$ 作精细的网格搜索（grid search），两者都与公式吻合，并得出最大值位于 $c = 0.2$。

常见错误

- 在二次公式中取了减号，得出负的时间。
- 把 $(*)$ 平方后保留了 $c = -1$ 而没有验证，或者忘记论证 $c = \frac{1}{5}$ 是最大值而非最小值。
- 假设最佳角度是 45°：这只在抛出点与落地点高度相同时才成立。

STEP 之外

从高处抛出时，最佳角度总是低于 45°；一般而言，最佳角度满足 $\tan\theta = u/\sqrt{u^2 + 2gh}$，在本题中给出 $\tan\theta = \sqrt{2/3}$，与 $\cos 2\theta = \frac{1}{5}$ 等价。对带参数的函数求最优值，并留意最大值附近有多平坦，是敏感度分析（sensitivity analysis）的初步体验：在最大值附近，参数的微小误差只会造成二阶（second-order）的损失。若想进一步了解有空气阻力及无空气阻力的抛射体运动，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP I 2012，第 10 题

力学 · 自由落体与声音

STEP I 2012，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/12-s1-q10

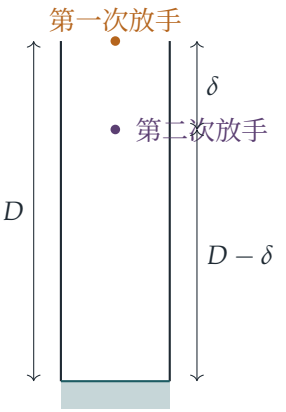
关键思路

每个量得的时间都由两部分组成：石头由静止开始自由下落的时间，以及水花声以等速传回上方的时间。把两次的时间都写出来再相减，便立即得到 t。至于“由此”（hence）部分，两个平方根正是两次的下落时间；我们知道它们的差，而它们的平方差又很简单，所以因式分解（factorisation）$a^2 - b^2 = (a-b)(a+b)$ 便把它们的和也交到我们手上，于是两者都可以分别求出。

解答

t 的公式

我们把石头模型化为只受重力作用、由静止开始下落的质点（particle）（不计空气阻力（air resistance）），并假设声音以等速 u 沿井笔直向上传播。



由静止放下的石头在时间 τ 内下落距离 h，其中 $h = \frac{1}{2}g\tau^2$，所以 $\tau = \sqrt{2h/g}$。之后声音要走同样的距离 h 传回我这里，需时 h/u。所以从水面以上高度 h 处放下石头时，我量得的时间是

$$\sqrt{\frac{2h}{g}} + \frac{h}{u}.$$

从井口放下时，$h = D$；爬下去之后，我（以及石头的起点）位于水面以上高度 $D - \delta$ 处，而声音只需传到那里便会被我听到。因此

$$\begin{aligned} t &= \left(\sqrt{\frac{2D}{g}} + \frac{D}{u}\right) - \left(\sqrt{\frac{2(D-\delta)}{g}} + \frac{D-\delta}{u}\right) \\ &= \sqrt{\frac{2D}{g}} - \sqrt{\frac{2(D-\delta)}{g}} + \frac{\delta}{u}, \end{aligned}$$

得证。

D 的公式

记两次的下落时间为 $a = \sqrt{2D/g}$ 及 $b = \sqrt{2(D-\delta)/g}$。刚才证得的结果说明

$$a - b = t - \frac{\delta}{u} = \beta.$$

由于 $\delta > 0$，我们有 $a > b$，所以 $\beta > 0$；特别地，$\beta \neq 0$。此外

$$a^2 - b^2 = \frac{2D}{g} - \frac{2(D-\delta)}{g} = \frac{2\delta}{g}.$$

想一想 我们知道 $a - b$。怎样才能单独求出 a?

因式分解：$a^2 - b^2 = (a-b)(a+b) = \beta(a+b)$，而由于 β 是正数，我们可以除以 β。所以

$$a + b = \frac{2\delta}{\beta g}.$$

把此式与 $a - b = \beta$ 相加再除以二，得

$$a = \tfrac{1}{2}\beta + \frac{\delta}{\beta g} = T.$$

最后，$a = \sqrt{2D/g}$ 表示 $D = \frac{1}{2}ga^2$，所以

$$D = \tfrac{1}{2}gT^2,$$

得证。（所以 T 其实就是石头从井口下落整个深度所需的时间。）

数值情况

取 $u = 300\,\mathrm{m\,s^{-1}}$、$g = 10\,\mathrm{m\,s^{-2}}$、$t = \frac{1}{5}\,\mathrm{s}$ 及 $\delta = 10\,\mathrm{m}$,

$$\beta = \frac{1}{5} - \frac{10}{300} = \frac{1}{5} - \frac{1}{30} = \frac{1}{6}\,\mathrm{s},$$
$$T = \frac{1}{12} + \frac{10}{\frac{1}{6}\cdot 10} = \frac{1}{12} + 6 = \frac{73}{12}\,\mathrm{s},$$
$$D = \tfrac{1}{2}\cdot 10\cdot\left(\frac{73}{12}\right)^2 = \frac{5\times 5329}{144} = \frac{26645}{144} \approx 185.03\,\mathrm{m}.$$

所以这口井深约 185 m。

验算

- 把 $D = 26645/144$ 及 $\delta = 10$ 代回 t 的公式：两次下落时间分别是 $73/12 \approx 6.083\,\mathrm{s}$ 及 $\sqrt{2(175.03)/10} \approx 5.917\,\mathrm{s}$，相差 $1/6\,\mathrm{s}$，再加上 $\delta/u = 1/30\,\mathrm{s}$，恰好得到 $t = 0.2\,\mathrm{s}$。
- 若声音瞬间传到（$u \to \infty$），则 $\beta = t$，理应如此：只有下落时间才有影响。
- 以数值方法逐步模拟石头下落，再加上声音的延迟，可重现 t 的公式至 $10^{-4}\,\mathrm{s}$ 的精度；而反求公式 $D = \frac{1}{2}gT^2$ 在 10^4 个随机情况中都能还原 D。

常见错误

- 忘记了我爬下去之后声音只需传播 $D - \delta$，因而漏掉 δ/u 这一项。
- 试图直接把方程 $a - b = \beta$ 平方；这会导致繁琐的代数运算，而 $a^2 - b^2 = (a-b)(a+b)$ 只需一行便给出 $a + b$。
- 完全忽略声音：取 $\beta = t = 0.2$，同一方法会给出 $T = 5.1\,\mathrm{s}$ 及 $D \approx 130\,\mathrm{m}$，错得很远。

STEP 之外

这里所用的方法是一种*差分测量*（differential measurement）：我们不依赖单一次计时，而是比较两个相差已知量的实验，使未知量互相抵消。最后一个常见错误显示答案对数据的微小变化有多敏感，因为 t 是两个大时间之间的微小差；物理学家和工程师会用*误差传播*（error propagation）和*条件数*（condition number）来量化这种敏感度，这些内容会在大学一年级的实验方法及数值分析（numerical analysis）课程中学到。如想进一步了解如何仔细地为简单运动建立模型，可阅读 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP I 2012，第 11 题

力学 · 滑轮与摩擦力

STEP I 2012，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/12-s1-q11

关键思路

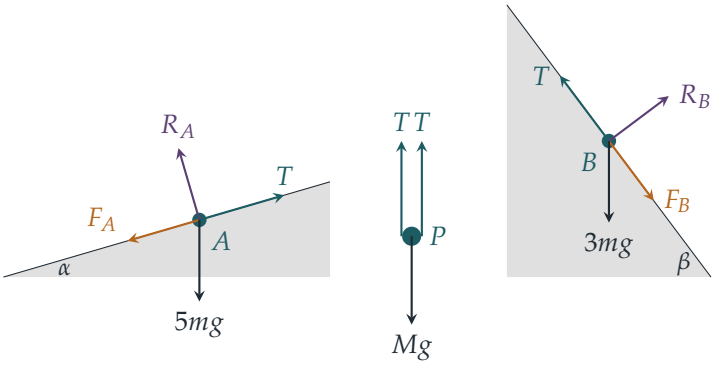
由于每个滑轮都是光滑（smooth）的，而绳子是轻（light）的，绳上各处的张力（tension）都相同。因此每个质点（particle）各给出一条把张力和摩擦力（friction）联系起来的方程，滑轮 P 则给出第三条。题中的角度取自 $7,24,25$ 和 $3,4,5$ 三角形，所以所有正弦和余弦都是准确的分数。至于运动的情况，我们还需要多一样东西：绳长固定如何把 A、B 和 P 的加速度联系起来。

解答

我们把 A、B 视为质点，把绳子视为轻而不可伸长（inextensible）的。滑轮 Q 和 R 光滑、轻且固定，而 P 光滑，质量为 M。设 $\alpha = \arctan\frac{7}{24}$ 及 $\beta = \arctan\frac{4}{3}$。由边长为 $7,24,25$ 和 $3,4,5$ 的直角三角形，

$$\sin\alpha = \frac{7}{25}, \quad \cos\alpha = \frac{24}{25}, \qquad \sin\beta = \frac{4}{5}, \quad \cos\beta = \frac{3}{5}.$$

设 T 为张力。由于所有滑轮都光滑，每段绳子中的 T 都相同。



在两部分中，质点都是即将或正在沿斜面向上运动，所以摩擦力如图所示沿各斜面向下。沿垂直于斜面的方向分解（resolve），

$$R_A = 5mg\cos\alpha = \tfrac{24}{5}mg, \qquad R_B = 3mg\cos\beta = \tfrac{9}{5}mg.$$

第 (i) 部分

两个质点都即将滑动，所以摩擦力达到极限（limiting）：$F_A = \mu R_A$ 及 $F_B = \mu R_B$。沿各斜面分解（平衡）：

$$\begin{aligned} A: &\quad T = 5mg\sin\alpha + \mu R_A = \tfrac{mg}{5}(7+24\mu), \\ B: &\quad T = 3mg\sin\beta + \mu R_B = \tfrac{mg}{5}(12+9\mu). \end{aligned}$$

两式相等，得 $7+24\mu = 12+9\mu$，所以

$$\mu = \tfrac{1}{3},$$

于是 $T = \frac{mg}{5}(7+8) = 3mg$。

滑轮 P 处于平衡，有两段竖直的绳子把它向上拉，而其重量把它向下拉，所以

$$2T = Mg \quad\Longrightarrow\quad M = \frac{2\cdot 3mg}{g} = 6m,$$

此即所求。

第 (ii) 部分

现在 $M = 9m$，比刚好能保持极限平衡（limiting equilibrium）的 $6m$ 更重，所以我们预期 P 会下降，并把两个质点沿各自的斜面向上拉。

? **想一想** 三个加速度之间有甚么关系?

设 A 和 B 沿各自斜面向上的加速度分别为 a 和 b，并设 P 向下的加速度为 p。绳长 $AQ+QP+PR+RB$ 固定不变。若 A 沿其斜面向上移动距离 x，B 沿其斜面向上移动距离 y，则 AQ 和 RB 分别缩短 x 和 y；若 P 下降 z，则 QP 和 PR 各伸长 z。因此在任何时刻都有 $2z=x+y$，求导两次，得

$$2p=a+b.$$

由于 $\mu=\frac{1}{3}$，向上滑动时 A 所受的摩擦力为 $\frac{1}{3}R_A=\frac{8}{5}mg$，$B$ 所受的为 $\frac{1}{3}R_B=\frac{3}{5}mg$。对两个质点沿各自斜面、对 P 沿竖直方向应用牛顿第二定律（Newton's second law），得

$$\begin{aligned} A:&\quad T-\tfrac{7}{5}mg-\tfrac{8}{5}mg=5ma, && \text{故}T-3mg=5ma,\\ B:&\quad T-\tfrac{12}{5}mg-\tfrac{3}{5}mg=3mb, && \text{故}T-3mg=3mb,\\ P:&\quad 9mg-2T=9mp. \end{aligned}$$

设 $T=3mg+X$。于是 $a=\dfrac{X}{5m}$、$b=\dfrac{X}{3m}$，而绳子的条件给出

$$p=\tfrac{1}{2}(a+b)=\frac{4X}{15m}.$$

代入 P 的方程：

$$3mg-2X=\frac{12X}{5}\quad\Longrightarrow\quad X=\frac{15mg}{22}.$$

因此

$$a=\frac{3g}{22}\text{ 沿斜面向上},\qquad b=\frac{5g}{22}\text{ 沿斜面向上},\qquad p=\frac{2g}{11}\text{ 竖直向下},$$

而张力为 $T=\frac{81}{22}mg$。

想一想 我们怎样知道两个质点真的都向上运动，而不是其中一个保持不动？

由上面的方程可见，质点只有在 $T>3mg$ 时才会沿斜面向上滑动，而只有在 $T\le 3mg$ 时才会保持静止（或会向下滑）。两个质点的临界值同样是 $3mg$，所以要么两者都向上运动，要么两者都不向上运动。若两者都不向上运动，则 $a+b\le 0$，所以 $p\le 0$；但 $T\le 3mg$ 给出 $9mg-2T\ge 3mg>0$，迫使 $p>0$。这是矛盾，所以两者都向上运动，而我们的解（其中 $a,b>0$）是自洽的。

验算

- 在第 (i) 部分，把 $\mu=\frac{1}{3}$ 代入两个质点的方程，得出相同的张力 $3mg$。
- 在第 (ii) 部分，$2p=\frac{4g}{11}=\frac{3g}{22}+\frac{5g}{22}=a+b$，符合绳子的要求；而 $T=\frac{81}{22}mg\approx 3.68mg$ 介乎 $3mg$（滑动的临界值）与 $4.5mg$（P 没有加速度时的值）之间。
- 以数值模拟逐步推进三条运动方程，绳长在舍入误差范围内保持不变，且 $p\approx 0.182g$。

- 若 M 再次为 $6m$，同样的方法给出 $X = 0$，所以所有加速度都为零，与第 (i) 部分吻合。

常见错误

- 以为 $a = b = p$：滑轮 P 的移动速率是两个质点移动速率总和的一半，因为有两段绳子分担它的运动。
- 把摩擦力定为沿斜面向上：两个质点都正被拉向上，所以摩擦力与此相反，沿斜面向下。
- 忘记 P 上的两段绳子都带有张力 T，所以作用在它上面的向上力是 $2T$。

STEP 之外

关系式 $2p = a + b$ 是一个约束（constraint）：它把独立坐标的数目由三个减至两个。在大学一、二年级会学到的拉格朗日力学（Lagrangian mechanics）中，你会选取自动满足约束的坐标，从而根本不必求张力。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）通过约束方程仔细处理了许多滑轮系统，是通往这种观点的良好桥梁。

详细解答

STEP I 2012，第 12 题

概率与统计 • 条件概率

STEP I 2012，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/12-s1-q12

关键思路

一切取决于灭火筒在哪一年出现故障，所以先把概率密度函数（probability density function）化为累积分布函数（cumulative distribution function），从而读出故障在第一、第二及第三年出现的概率。一旦知道是哪一年，各次测试就只是一连串互相独立（independent）的“通过”，最后接一次“不通过”。最后一部分是条件概率（conditional probability），而“在 18 个月时已有故障”这个事件，同样是关于 T 落在哪个区间的陈述。

解答

T 的分布

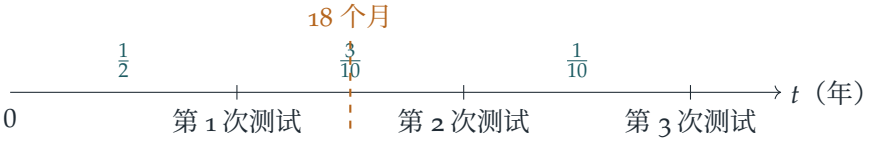
对 $t \geq 0$，代入 $u = 1 + s^2$（于是 $\mathrm{d}u = 2s\,\mathrm{d}s$），

$$F(t) = \mathrm{P}(T \leq t) = \int_0^t \frac{2s}{(1+s^2)^2}\,\mathrm{d}s = \left[-\frac{1}{1+s^2}\right]_0^t = 1 - \frac{1}{1+t^2} = \frac{t^2}{1+t^2}.$$

所以 $F(1) = \frac{1}{2}$，$F(2) = \frac{4}{5}$，$F(3) = \frac{9}{10}$，而

$$\mathrm{P}(0 < T < 1) = \tfrac{1}{2}, \qquad \mathrm{P}(1 < T < 2) = \tfrac{4}{5} - \tfrac{1}{2} = \tfrac{3}{10}, \qquad \mathrm{P}(2 < T < 3) = \tfrac{9}{10} - \tfrac{4}{5} = \tfrac{1}{10}.$$

由于 T 是连续（continuous）随机变量，故障恰好在测试当日出现的概率为零，所以不必担心端点处用的是 $<$ 还是 $\leq$。



恰好在制造后三年被销毁

设 D 为灭火筒在制造后三年的测试中被销毁这个事件。要发生这件事，灭火筒必须在第三次测试前已有故障（即 $T<3$），在此之前处于故障状态时所接受的每次测试都必须通过，而第三次测试则必须不通过。没有故障的灭火筒必定通过测试，所以只有时间 T 之后的测试才有影响，而这些测试互相独立，每次不通过的概率为 p。

- 若 $0<T<1$：它在第 1、2、3 次测试时都有故障，所以必须通过、通过、不通过：概率为 $(1-p)^2p$。
- 若 $1<T<2$：它自动通过第 1 次测试，然后必须通过第 2 次测试，并在第 3 次测试不通过：概率为 $(1-p)p$。
- 若 $2<T<3$：它自动通过第 1 及第 2 次测试，并必须在第 3 次测试不通过：概率为 p。

若 $T>3$，它在第 3 次测试时并无故障，因此不会在那时被销毁。根据全概率定律（law of total probability），

$$\begin{aligned}\mathrm{P}(D) &= \tfrac{1}{2}(1-p)^2p + \tfrac{3}{10}(1-p)p + \tfrac{1}{10}p \\ &= \frac{p}{10}\Big(5(1-2p+p^2)+3(1-p)+1\Big) \\ &= \frac{p\,(5p^2-13p+9)}{10},\end{aligned}$$

此即所求。

已知在三年时被销毁，求在 18 个月时已有故障的概率

设 E 为灭火筒在制造后 18 个月时已有故障这个事件，即 $T<\frac{3}{2}$。我们要求

$$\mathrm{P}(E\mid D)=\frac{\mathrm{P}(E\cap D)}{\mathrm{P}(D)}.$$

想一想 区间 $(0,\frac{3}{2})$ 横跨了一次测试。我们该如何把它分拆?

对测试而言，重要的是故障在哪一年出现，所以把 E 分拆为 $0<T<1$ 和 $1<T<\frac{3}{2}$。我们有 $F(\frac{3}{2})=\dfrac{9/4}{13/4}=\dfrac{9}{13}$，所以

$$\mathrm{P}\big(1<T<\tfrac{3}{2}\big)=\tfrac{9}{13}-\tfrac{1}{2}=\tfrac{5}{26}.$$

已知 $0<T<1$ 时，D 的概率与之前一样是 $(1-p)^2p$；已知 $1<T<\frac{3}{2}$ 时，故障仍然在

第二年出现，所以 D 的概率是 $(1-p)p$。因此

$$\begin{aligned}\mathrm{P}(E\cap D) &= \tfrac{1}{2}(1-p)^2p + \tfrac{5}{26}(1-p)p \\ &= \frac{p(1-p)}{26}\big(13(1-p)+5\big) = \frac{p(1-p)(18-13p)}{26}.\end{aligned}$$

除以 $\mathrm{P}(D)$（当 $0<p\le 1$ 时它是正数，因为 $5p^2-13p+9$ 的判别式（discriminant）为 $169-180<0$，故此式恒为正），因子 p 约去，得

$$\mathrm{P}(E\mid D) = \frac{10(1-p)(18-13p)}{26(5p^2-13p+9)} = \frac{5(1-p)(18-13p)}{13(5p^2-13p+9)}.$$

验算

- 当 $p=1$ 时，每个有故障的灭火筒都会在故障后的第一次测试中被销毁，所以在第 3 年被销毁即表示故障在 $(2,3)$ 内出现：$\mathrm{P}(D)=\frac{1}{10}$ 而 $\mathrm{P}(E\mid D)=0$。两条公式都吻合。
- 当 $p\to 0$ 时，$\mathrm{P}(E\mid D)\to\dfrac{5\cdot 18}{13\cdot 9}=\dfrac{10}{13}$，这正是 $\mathrm{P}(T<\frac{3}{2}\mid T<3)=\dfrac{9/13}{9/10}$：当测试几乎检测不到故障时，被销毁这件事除了 $T<3$ 之外，几乎没有告诉我们任何信息。
- 以 $p=0.4$ 对 10^6 个灭火筒进行模拟（simulation），得到 $\mathrm{P}(D)\approx 0.1838$（准确值 0.184）及 $\mathrm{P}(E\mid D)\approx 0.642$（准确值 $\frac{192}{299}\approx 0.642$）。

常见错误

- 忘记有故障的灭火筒在第三次测试不通过之前，必须先通过较早的测试。
- 在最后一部分把 $\mathrm{P}(T<\frac{3}{2})=\frac{9}{13}$ 当作单一整体使用：在第一年出现的故障与在第二年上半年出现的故障，所要经历的测试次数并不相同。

STEP 之外

这里的 T 是一个连续的寿命，但我们只能通过离散而不完美的检查来观察它。这正是生存分析（survival analysis）和可靠性理论（reliability theory）中区间设限（interval-censored）数据的情境，而最后一部分则是一次小型的贝叶斯更新（Bayesian update）：T 的先验分布（prior distribution）因观察到的一次失效而得到修正。Grimmett 和 Welsh 所著的 *Probability: An Introduction*（《概率导论》）涵盖了连续分布以及这类条件化的内容。

详细解答

STEP I 2012，第 13 题

概率与统计 · 以指示变量求期望值

STEP I 2012，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/12-s1-q13

关键思路

首先要把条件样本空间（conditional sample space）弄清楚：一旦知道哪些数字被排除，这个数就只是一串五个数字，每个数字都从 $\{1,2,3,4,5\}$ 中选取，而每一串出现的可能性都相同。接着，与其求出“不同数字的个数”的整个分布（distribution），不如把这个个数写成指示变量（indicator variable）之和，每个可能出现的数字对应一个。期望值的线性性质（linearity of expectation）把问题化为求一个概率，而这个概率考虑其补事件（complement）便很容易求得。

解答

条件样本空间

由 10000 至 99999 的每个整数都是一串五个数字 $d_1d_2d_3d_4d_5$，其中 $d_1 \neq 0$，而每一串都恰好出现一次。条件说明每个数字都属于 $\{1,2,3,4,5\}$。由于 0 本来就已被排除，限制 $d_1 \neq 0$ 并没有增加任何约束，因此符合条件的整数恰好对应于

$$5^5 = 3125$$

串由 $\{1,2,3,4,5\}$ 中数字组成的五位数字串。

原本的选取是均匀的。以某事件为条件时，该事件内的各个结果仍然等可能，所以在给定条件下，这 3125 串中每一串的概率都是 $\frac{1}{3125}$。换句话说，这五个数字互相独立（independent），而且每个都在 $\{1,2,3,4,5\}$ 上均匀分布（uniform）：这 3125 个结果中，每一个的概率都是五个因子 $\frac{1}{5}$ 的乘积。

指示变量

设 N 为所选的数中不同数字的个数。对 $j = 1,2,3,4,5$，定义

$$I_j = \begin{cases} 1 & \text{若数字} j \text{ 至少出现一次,} \\ 0 & \text{否则。} \end{cases}$$

每个出现了的数字对总和恰好贡献 1，因此

$$N = I_1 + I_2 + I_3 + I_4 + I_5.$$

想一想 各个 I_j 并不互相独立。我们仍然可以把它们的期望值相加吗?

可以。期望值的线性性质 $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ 对任何随机变量（random variable）都成立，不论它们是否独立。因此

$$\mathrm{E}(N) = \sum_{j=1}^{5} \mathrm{E}(I_j) = \sum_{j=1}^{5} \mathrm{P}(\text{数字} j \text{ 出现}).$$

数字 j 不出现，当且仅当五个独立的数字每一个都是其余四个值之一，而每个数字发生这情况的概率都是 $\frac{4}{5}$。因此

$$\mathrm{P}(\text{数字} j \text{ 出现}) = 1 - \left(\tfrac{4}{5}\right)^5 = 1 - \frac{1024}{3125} = \frac{2101}{3125}.$$

由对称性，这个概率对 j 的五个值都相同，所以

$$\mathrm{E}(N) = 5 \cdot \frac{2101}{3125} = \frac{2101}{625} = 3.3616,$$

即为所求。

另一方法：完整分布

如果你较喜欢直接计数，恰好使用 k 个不同数字的字串数目，等于 $\binom{5}{k}$（选出那些数字）乘以填满五个位置而使所选的 k 个数字全部出现的方法数。后者可由容斥原理（inclusion–exclusion）求得。例如当 $k = 3$ 时，它等于 $3^5 - 3 \cdot 2^5 + 3 \cdot 1^5 = 150$。

k	1	2	3	4	5	总数
字串数	$5 \cdot 1$	$10 \cdot 30$	$10 \cdot 150$	$5 \cdot 240$	$1 \cdot 120$	
	5	300	1500	1200	120	3125

于是

$$\mathrm{E}(N) = \frac{1 \cdot 5 + 2 \cdot 300 + 3 \cdot 1500 + 4 \cdot 1200 + 5 \cdot 120}{3125} = \frac{10505}{3125} = 3.3616.$$

这方法行得通，但较长，出错的地方也较多。指示变量的方法只需要求一个概率。

验算

- 表中各数目之和为 $3125 = 5^5$，证实每一串都恰好被计算了一次。
- 答案合理：N 介乎 1 与 5 之间，而从五个值中抽取五次，预期会有相当多重复，因此一个略大于 3 的值是可信的。作为一般性的验算，若从 m 个值中抽取 n 次，同样的论证给出 $m\left(1-\left(1-\frac{1}{m}\right)^n\right)$，当 $n=1$ 时它等于 1，正如所料。
- 用电脑列举由 10000 至 99999 的所有整数，保留只使用数字 1–5 的那 3125 个，再求不同数字个数的平均值，得到的恰好是 $\frac{2101}{625} = 3.3616$。

常见错误

- 把样本空间当作全部 90000 个整数。条件把它缩减为 5^5 串等可能的字串。
- 以为期望值的线性性质需要独立性。其实不需要，而这正是指示变量如此有用的原因。
- 以对五个位置各加 $\frac{1}{5}$ 的方式计算 P(数字j 出现)，结果得出 1，因为这样会把 j 出现多于一次的字串重复计算了多次。应该使用补事件。

STEP 之外

这是经典的“占位”(occupancy)问题,又称集券问题(coupon-collector problem):把 n 个球随机投进 m 个盒子,N 为最终有球的盒子数目。指示变量的技巧 $\mathrm{E}(N) = m\left(1-\left(1-\frac{1}{m}\right)^n\right)$ 是大学一年级概率论中，以至分析杂凑（hashing）等算法时最有用的工具之一。精确的分布牵涉第二类斯特灵数（Stirling numbers of the second kind），它们已在上面的容斥计数中出现过。Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）全书贯彻使用指示随机变量，是绝佳的下一步读物。

详细解答

STEP II 2012，第 1 题

纯数学 · 二项式级数的系数

> **STEP II 2012，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/12-s2-q1

关键思路

每个因子 $(1-x^k)^{-1}$ 都是以 x^k 为公比的几何级数（geometric series），所以把这些因子相乘，再读出 x^n 的系数（coefficient），实际上就是在数算把 n 写成若干容许部分之和的方法数。一旦把系数看成计数，分情况讨论便会自然理顺；而乘以 $(1-x)^{-1}$，只不过是把一列系数变成它的累积和（running totals）。

解答

通项

由二项式展开（binomial expansion），当 $|y|<1$ 时，$(1-y)^{-2}=\sum_{r\geq 0}(r+1)y^r$（系数为 $\binom{-2}{r}(-1)^r=r+1$）。令 $y=x^6$，

$$(1-x^6)^{-2}=\sum_{r=0}^{\infty}(r+1)\,x^{6r},$$

因此通项（general term）是 $(r+1)x^{6r}$，而每个指数不是 6 的倍数的 x 幂，其系数都是 0。

第 (i) 部分

另外，$(1-x^3)^{-1}=\sum_{s\geq 0}x^{3s}$。把两个级数相乘，每一对满足 $6r+3s=n$ 的非负整数（non-negative integers）(r,s) 都会产生一项 x^n，贡献为 $(r+1)\cdot 1$。所以

$$\text{系数，对应}x^n=\sum_{\substack{r,s\geq 0\\ 6r+3s=n}}(r+1).$$

当 $n = 24$ 时，我们需要 $2r + s = 8$，所以 r 可以是 $0, 1, 2, 3, 4$（每个 r 都恰好对应一个 $s = 8 - 2r \geq 0$）。x^{24} 的系数为

$$1 + 2 + 3 + 4 + 5 = 15.$$

一般的 n。由于 $6r + 3s$ 必定是 3 的倍数，除非 $3 \mid n$，否则系数为 0。若 $n = 3m$，我们需要 $2r + s = m$，而对每个满足 $0 \leq 2r \leq m$ 的 r，都恰好有一个合适的 s。因此 r 由 0 取到 $\lfloor m/2 \rfloor$，系数为

$$1 + 2 + \cdots + (\lfloor m/2 \rfloor + 1).$$

按 n 除以 6 的余数分类：

- $n = 6k$（即 $m = 2k$，$\lfloor m/2 \rfloor = k$）：系数为 $\frac{1}{2}(k+1)(k+2)$；
- $n = 6k + 3$（即 $m = 2k + 1$，$\lfloor m/2 \rfloor = k$）：系数同样是 $\frac{1}{2}(k+1)(k+2)$；
- $n \not\equiv 0 \pmod 3$：系数为 0。

验算一下：$n = 24 = 6 \times 4$ 给出 $\frac{1}{2} \cdot 5 \cdot 6 = 15$，与上面所得相同。

第 (ii) 部分

以 b_n 表示第 (i) 部分求得的 x^n 系数，以 c_n 表示现在要求的系数。由于 $(1 - x)^{-1} = 1 + x + x^2 + \cdots$，

$$\Big(\sum_j b_j x^j\Big)(1 + x + x^2 + \cdots) = \sum_n \Big(\sum_{j=0}^{n} b_j\Big) x^n,$$

所以 $c_n = b_0 + b_1 + \cdots + b_n$，即第 (i) 部分系数的累积和。

? **想一想** 为甚么乘以 $(1 - x)^{-1}$ 会得到累积和?

要在乘积中得到 x^n，我们从第一个级数取 x^j，再从 $1 + x + x^2 + \cdots$ 取 x^{n-j}，而后者的系数永远是 1。所以每个满足 $0 \leq j \leq n$ 的 b_j 都恰好在 c_n 中出现一次，权重为 1。

只有 $3 \mid j$ 的 b_j 才不为零，而由第 (i) 部分可知它们成对相等：$b_{6k} = b_{6k+3} = T_{k+1}$，其中 $T_p = \frac{1}{2}p(p+1)$ 是第 p 个三角形数（triangular number）。所以当 $n = 24$ 时，

$$\begin{aligned} c_{24} &= (b_0 + b_3) + (b_6 + b_9) + (b_{12} + b_{15}) + (b_{18} + b_{21}) + b_{24} \\ &= 2(1 + 3 + 6 + 10) + 15 = 40 + 15 = 55, \end{aligned}$$

正如题目所要求。

当 $n = 25$ 时，由于 $3 \nmid 25$，多出来的一项 b_{25} 为 0，所以 $c_{25} = c_{24} = 55$。

当 $n = 66 = 6 \times 11$ 时，同样的规律给出

$$c_{66} = 2\,(T_1 + T_2 + \cdots + T_{11}) + T_{12}.$$

首 p 个三角形数之和是四面体数（tetrahedral number）$\frac{1}{6}p(p+1)(p+2)$（可用归纳法（induction）证明，或由曲棍球棒恒等式（hockey-stick identity）得 $T_1 + \cdots + T_p = \binom{p+2}{3}$）。所以 $T_1 + \cdots + T_{11} = \frac{1}{6} \cdot 11 \cdot 12 \cdot 13 = 286$，而 $T_{12} = 78$。因此

$$c_{66} = 2 \times 286 + 78 = 650.$$

同样的论证可得出一般公式：若 $n = 6K + r$ 且 $0 \le r \le 2$，则 $c_n = 2\binom{K+2}{3} + \binom{K+2}{2}$；若 $3 \le r \le 5$，则 $c_n = 2\binom{K+3}{3}$。取 $K = 4$，第一条公式给出 $40 + 15 = 55$；取 $K = 11$，则给出 $572 + 78 = 650$。

验算

第 (ii) 部分中 x^n 的系数，数算的是 $6a + 6b + 3c + d = n$ 的非负整数解的个数（每个因子对应一个几何级数，$(1 - x^6)^{-2}$ 算作两个）。用电脑直接数算这些解，对 $n = 24, 25, 66$ 分别得出 55、55 和 650。把级数逐项乘开至 x^{80}，亦证实了两部分的公式。用人手验算小的情况也很快：在第 (ii) 部分中，$c_0 = c_1 = c_2 = 1$，$c_3 = c_4 = c_5 = 2$，与 $K = 0$ 时的一般公式吻合。

常见错误

- 在第 (i) 部分遗漏 $n = 6k + 3$ 的情况：3 的奇数倍确实会出现，来自奇数个 x^3 因子。
- 把 $(1 - x^6)^{-2}$ 的通项写成 $r\,x^{6r}$ 而非 $(r+1)x^{6r}$；常数项是 1，不是 0。
- 把 x^{25} 当作全新的计算：它等于 x^{24} 的系数，因为第 (i) 部分的级数中没有指数为 25 的项。

STEP 之外

像这样用作计数工具的幂级数称为*生成函数*（generating functions）；这里的乘积数算的是用面值 6、6（另一款设计）、3 和 1 的硬币支付 n 便士（pence）的方法数。整数分拆（partitions）理论正是从这个想法发展出来的；而答案是随 n 模 6 而定的分段多项式（piecewise polynomials），这是一个普遍现象（“拟多项式”（quasi-polynomials））。若想读一本生动的入门书介绍这个方法，可阅读 Wilf 的 *generatingfunctionology*（《生成函数学》）。

详细解答

STEP II 2012，第 2 题

纯数学 · 多项式的复合

STEP II 2012，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/12-s2-q2

关键思路

在复合（composition）之下，次数（degree）会相乘，所以把 p 迭代几次，次数很快就大得无法与右方相配。比较次数就能确定 p 的次数；之后把 p 写成带未知系数（coefficient）的形式，每一部分都只剩下一小段代数运算。第 (ii) 部分有一条巧妙的捷径：$\mathrm{p}(\mathrm{p}(x))$ 本身就是一个关于 $\mathrm{p}(x)$ 的二次式，因此整个左方是 $\mathrm{p}(x)$ 的二次式，用配方法（completing the square）便可完成。

解答

复合函数的次数

设 $\mathrm{p}(x) = a_m x^m + \ldots$ 及 $\mathrm{q}(x) = b_n x^n + \ldots$，其中 $a_m, b_n \neq 0$。在 $\mathrm{p}(\mathrm{q}(x)) = a_m[\mathrm{q}(x)]^m + a_{m-1}[\mathrm{q}(x)]^{m-1} + \ldots$ 中，项 $a_m[\mathrm{q}(x)]^m$ 的次数为 mn，首项系数为 $a_m b_n^m \neq 0$，而其余每一项的次数至多为 $(m-1)n < mn$（当 $n \geq 1$ 时）。所以没有任何项能抵消最高次项，故

$$\deg \mathrm{p}(\mathrm{q}(x)) = mn.$$

（若 $n = 0$，则 $\mathrm{p}(\mathrm{q}(x))$ 是常数，而 $mn = 0$ 亦同样成立。）

第 (i) 部分

设 p 的次数为 m。

若 $m \geq 2$，则由上述结果，$\mathrm{p}(\mathrm{p}(\mathrm{p}(x)))$ 的次数为 m^3，而 $3\mathrm{p}(x)$ 的次数为 $m < m^3$。所以左方的次数为 $m^3 \geq 8$，但右方 $-2x$ 的次数为 1。这是不可能的。

若 $m = 0$，则对某常数 c 有 $\mathrm{p}(x) = c$，左方是常数 $c - 3c = -2c$，不可能对所有 x 都等

于 $-2x$。这也是不可能的。

所以 $m = 1$。设 $\mathrm{p}(x) = ax + b$，其中 $a \neq 0$。于是

$$\mathrm{p}(\mathrm{p}(x)) = a(ax + b) + b = a^2x + (a + 1)b,$$
$$\mathrm{p}(\mathrm{p}(\mathrm{p}(x))) = a^3x + (a^2 + a + 1)b,$$

而方程变为

$$(a^3 - 3a)x + (a^2 + a + 1 - 3)b = -2x \quad \text{对所有} x.$$

两个多项式（polynomial）对所有 x 都相等，当且仅当它们的系数相等，所以

$$a^3 - 3a + 2 = 0 \qquad \text{及} \qquad (a^2 + a - 2)\, b = 0.$$

由于 $a = 1$ 是第一式的根，它可因式分解为 $(a - 1)^2(a + 2) = 0$，得 $a = 1$ 或 $a = -2$。第二个条件是 $(a - 1)(a + 2)b = 0$，对这两个值都自动成立，不论 b 是甚么。因此解恰好是

$$\mathrm{p}(x) = x + b \qquad \text{及} \qquad \mathrm{p}(x) = -2x + b, \qquad b \text{ 为任意常数.}$$

验算第二族：$\mathrm{p}(\mathrm{p}(x)) = 4x - b$ 及 $\mathrm{p}(\mathrm{p}(\mathrm{p}(x))) = -8x + 3b$，所以左方为 $-8x + 3b + 6x - 3b = -2x$。

第 (ii) 部分

设 p 的次数为 m。左方三项的次数分别为 m^2、$2m$ 及 m。

- 若 $m \geq 3$，则 $m^2 > 2m > m$，所以左方的次数为 $m^2 \geq 9 \neq 4$。
- 若 $m = 1$，左方的次数至多为 2；若 $m = 0$，左方是常数。两者都不可能等于 x^4。

所以 $m = 2$。（此时 $m^2 = 2m = 4$，故 $2\mathrm{p}(\mathrm{p}(x))$ 与 $3[\mathrm{p}(x)]^2$ 的最高次项次数相同，我们必须检查它们的系数。）

设 $\mathrm{p}(x) = ax^2 + bx + c$，其中 $a \neq 0$；为了少写一点，以 P 代表 $\mathrm{p}(x)$。

> ? **想一想** 我们真的要把 $\mathrm{p}(\mathrm{p}(x))$ 完全展开吗?

不必。按定义 $\mathrm{p}(\mathrm{p}(x)) = aP^2 + bP + c$，所以左方为

$$2(aP^2 + bP + c) + 3P^2 - 4P = (2a + 3)P^2 + (2b - 4)P + 2c,$$

这是 P 的二次式。它的 x^4 系数是 $(2a + 3)a^2$，因为 P^2 的首项（leading term）是 a^2x^4，而 P 的次数只有 2。与 x^4 比较，得

$$2a^3 + 3a^2 - 1 = 0, \qquad \text{即} \qquad (a + 1)^2(2a - 1) = 0,$$

（看出根 $a=-1$ 后再作除法），所以 $a=-1$ 或 $a=\frac{1}{2}$。

现在我们用一个简单的事实：若 $S(x)$ 是多项式，且对某常数 K 有 $S(x)^2=x^4+K$，则 $(S-x^2)(S+x^2)=K$。当 S 的首项为 $\pm x^2$ 时，若两个因式都不是零多项式，则至少其中一个的次数为 2，乘积便不可能是常数。所以对首项为 $\pm x^2$ 的 S，必有一个因式为零，即恰好 $S=\pm x^2$。

情况 $a=-1$。此时 $2a+3=1$，配方得

$$P^2+(2b-4)P+2c=\big(P+b-2\big)^2-(b-2)^2+2c=x^4.$$

所以 $S=P+b-2=-x^2+bx+(c+b-2)$ 满足 $S^2=x^4+K$，其中 K 为常数，且首项为 $-x^2$。于是 $S+x^2=bx+(c+b-2)$ 的次数小于 2，而 $S-x^2$ 的次数为 2，所以上述事实迫使 $S+x^2=0$：即 $b=0$ 及 $c+b-2=0$。所以 $b=0$、$c=2$，于是 $K=(b-2)^2-2c=0$，前后一致。这给出

$$\mathrm{p}(x)=2-x^2.$$

情况 $a=\frac{1}{2}$。此时 $2a+3=4$，而

$$4P^2+(2b-4)P+2c=\left(2P+\tfrac{b-2}{2}\right)^2-\tfrac{(b-2)^2}{4}+2c=x^4.$$

现在 $S=2P+\frac{b-2}{2}=x^2+2bx+2c+\frac{b-2}{2}$ 的首项为 x^2，所以同理 $S-x^2=0$：$b=0$ 及 $2c-1=0$。所以 $b=0$、$c=\frac{1}{2}$，而 $K=\frac{(b-2)^2}{4}-2c=1-1=0$，前后一致。这给出

$$\mathrm{p}(x)=\tfrac{1}{2}(x^2+1).$$

由于每一步都可以逆推，两者确实都是解，但直接验算一下会更安心。对 $\mathrm{p}(x)=2-x^2$，$\mathrm{p}(\mathrm{p}(x))=2-(2-x^2)^2=-x^4+4x^2-2$，而

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

对 $\mathrm{p}(x)=\frac{1}{2}(x^2+1)$，设 $u=x^2+1$；则 $2\mathrm{p}(\mathrm{p}(x))=\frac{1}{4}u^2+1$，而左方为 $\frac{1}{4}u^2+1+\frac{3}{4}u^2-2u=(u-1)^2=x^4$。

所以满足 (ii) 中方程的多项式恰好是 $\mathrm{p}(x)=2-x^2$ 及 $\mathrm{p}(x)=\frac{1}{2}(x^2+1)$。

验算

- 在 (i) 中试 $b=1$：$\mathrm{p}(x)=x+1$ 给出 $(x+3)-3(x+1)=-2x$。
- 在 (ii) 中代入 $x=1$:对 $\mathrm{p}(x)=2-x^2$,得 $\mathrm{p}(1)=1$、$\mathrm{p}(\mathrm{p}(1))=1$,而 $2+3-4=1=1^4$。对 $\mathrm{p}(x)=\frac{1}{2}(x^2+1)$，取 $x=3$：$\mathrm{p}(3)=5$、$\mathrm{p}(5)=13$，而 $26+75-20=81=3^4$。
- 我们亦用精确运算验证了 (i) 中两族解对许多随机的 b 都成立，并在一个以 $\frac{1}{4}$ 为步长取 b,c 的二次式网格上搜索，结果只找到 (ii) 的两个解。

常见错误

- 在 (ii) 中假设左方的次数就是 m^2：当 $m = 2$ 时，$2\mathrm{p}(\mathrm{p}(x))$ 与 $3[\mathrm{p}(x)]^2$ 两项的次数都是 4，正正因为它们有可能互相抵消，所以 a 必须由 $2a^3 + 3a^2 = 1$ 求得。
- 在 (i) 中解释为何次数必须是 1 时，漏掉 $m = 0$ 的情况。
- 在 (i) 中求得 $a = 1$ 或 $a = -2$ 后便停下，并把 b 定为某个特定值：常数项的条件对每个 b 都成立。

STEP 之外

法则 $\deg(\mathrm{p} \circ \mathrm{q}) = \deg \mathrm{p} \cdot \deg \mathrm{q}$，与 $\deg(\mathrm{pq}) = \deg \mathrm{p} + \deg \mathrm{q}$ 一样，依赖于系数之间没有零因子（zero divisors）：在一个两个非零数相乘可以得零的环（ring）中，首项可能消失，这两条法则都会失效。像 $\mathrm{p}(\mathrm{p}(x)) = \ldots$ 这样的函数方程（functional equation）亦会引向迭代映射（iterated maps）与动力系统（dynamics）的研究。关于多项式环以及次数论证何时适用，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 2012，第 3 题

纯数学 · 代换积分法

STEP II 2012，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/12-s2-q3

关键思路

$x+\sqrt{1+x^2}$ 这个麻烦的式子，明显应该设为 t。令这个代换（substitution）奏效的关键，在于它的倒数（reciprocal）是 $\sqrt{1+x^2}-x$，因此把 t 与 $1/t$ 相减和相加，便可把 x 与 $\sqrt{1+x^2}$ 都写成 t 的简单函数（function）。之后，两个积分（integral）各自的重点，就是看出如何把被积函数（integrand）写成只含 $x+\sqrt{1+x^2}$ 的函数。

解答

恒等式

设 $t=x+\sqrt{1+x^2}$。乘以共轭式（conjugate），得

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

把这两条方程相减及相加，得

$$x=\tfrac{1}{2}\left(t-\frac{1}{t}\right),\qquad \sqrt{1+x^2}=\tfrac{1}{2}\left(t+\frac{1}{t}\right).$$

第一条式直接把 x 表示为 t 的函数，因此

$$\mathrm{d}x=\tfrac{1}{2}\left(1+\frac{1}{t^2}\right)\mathrm{d}t.$$

至于积分上下限：当 $x=0$ 时，$t=1$；而当 $x\to\infty$ 时，$t\to\infty$。此外，t 是 x 的严格递增函数（strictly increasing function）（当 $x\geq 0$ 时，x 与 $\sqrt{1+x^2}$ 皆递增），所以这个代换在 $[0,\infty)$ 上是一一对应（one-to-one）的。因此

$$\int_0^\infty f\left(x+\sqrt{1+x^2}\right)\mathrm{d}x=\frac{1}{2}\int_1^\infty\left(1+\frac{1}{t^2}\right)f(t)\,\mathrm{d}t,$$

得证。

第一个积分

> ? **想一想** 分母 $2x^2+1+2x\sqrt{x^2+1}$ 怎样才能写成 $x+\sqrt{1+x^2}$ 的函数?

试试平方：$\left(x+\sqrt{1+x^2}\right)^2 = x^2+2x\sqrt{1+x^2}+(1+x^2) = 2x^2+1+2x\sqrt{x^2+1}$，恰好就是分母。因此被积函数是 $f\left(x+\sqrt{1+x^2}\right)$，其中 $f(t)=1/t^2$，由恒等式得

$$\begin{aligned}\int_0^\infty \frac{\mathrm{d}x}{2x^2+1+2x\sqrt{x^2+1}} &= \frac{1}{2}\int_1^\infty \left(1+\frac{1}{t^2}\right)\frac{1}{t^2}\,\mathrm{d}t = \frac{1}{2}\int_1^\infty \left(t^{-2}+t^{-4}\right)\mathrm{d}t \\ &= \frac{1}{2}\left[-\frac{1}{t}-\frac{1}{3t^3}\right]_1^\infty = \frac{1}{2}\left(1+\frac{1}{3}\right) = \frac{2}{3}.\end{aligned}$$

第二个积分

设 $x=\tan\theta$，则 $\mathrm{d}x=\sec^2\theta\,\mathrm{d}\theta$，即 $\mathrm{d}\theta = \dfrac{\mathrm{d}x}{1+x^2}$。当 θ 由 0 变到 $\frac{1}{2}\pi$ 时，x 由 0 变到 ∞。由于 $0\le\theta<\frac{1}{2}\pi$，故 $\cos\theta>0$，所以 $\sec\theta=\sqrt{1+x^2}$，且

$$\sin\theta=\tan\theta\cos\theta=\frac{x}{\sqrt{1+x^2}},\qquad 1+\sin\theta=\frac{\sqrt{1+x^2}+x}{\sqrt{1+x^2}}=\frac{t}{\sqrt{1+x^2}},$$

其中同样有 $t=x+\sqrt{1+x^2}$。因此

$$\int_0^{\frac{1}{2}\pi}\frac{\mathrm{d}\theta}{(1+\sin\theta)^3} = \int_0^\infty \frac{(1+x^2)^{3/2}}{t^3}\cdot\frac{\mathrm{d}x}{1+x^2} = \int_0^\infty \frac{\sqrt{1+x^2}}{t^3}\,\mathrm{d}x.$$

> ? **想一想** 被积函数中仍有一个 $\sqrt{1+x^2}$。怎样把它写成 t 的式子?

我们在第一部分已经求得：$\sqrt{1+x^2}=\frac{1}{2}(t+1/t)$。所以被积函数是

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

只是 $t=x+\sqrt{1+x^2}$ 的函数。取 $f(t)=\frac{1}{2}(t^{-2}+t^{-4})$ 应用恒等式，

$$\begin{aligned}\int_0^{\frac{1}{2}\pi}\frac{\mathrm{d}\theta}{(1+\sin\theta)^3} &= \frac{1}{2}\int_1^\infty\left(1+\frac{1}{t^2}\right)\cdot\frac{1}{2}\left(\frac{1}{t^2}+\frac{1}{t^4}\right)\mathrm{d}t = \frac{1}{4}\int_1^\infty\left(t^{-2}+2t^{-4}+t^{-6}\right)\mathrm{d}t \\ &= \frac{1}{4}\left(1+\frac{2}{3}+\frac{1}{5}\right) = \frac{1}{4}\cdot\frac{28}{15} = \frac{7}{15}.\end{aligned}$$

验算

- 用一个简单的 f 测试恒等式：取 $f(t)=1/t^3$，右方为 $\frac{1}{2}\int_1^\infty(t^{-3}+t^{-5})\,\mathrm{d}t=\frac{1}{2}\left(\frac{1}{2}+\frac{1}{4}\right)=\frac{3}{8}$，而用数值求积（numerical quadrature）计算左方，同样得 0.375。

- 用辛普森法则（Simpson's rule）计算原来的两个积分，得 0.666667 及 0.466667，与 $\frac{2}{3}$ 及 $\frac{7}{15}$ 吻合。
- 数量级的合理性检查：在 $[0, \frac{1}{2}\pi]$ 上，被积函数 $1/(1+\sin\theta)^3$ 由 1 降至 $\frac{1}{8}$，所以积分值介乎 $\frac{1}{8}\cdot\frac{1}{2}\pi \approx 0.20$ 与 $\frac{1}{2}\pi \approx 1.57$ 之间。$\frac{7}{15} \approx 0.47$ 稳稳落在这个范围内。

常见错误

- 先求 $\mathrm{d}t/\mathrm{d}x$，再费力地把它化成 t 的式子。先把 x 写成 t 的函数，然后对它求导，会快得多。
- 在第二个积分中，漏掉了来自 $\mathrm{d}\theta$ 的因子 $1/(1+x^2)$，或者把 $\sqrt{1+x^2}$ 留在被积函数中，而没有把它化成 $\frac{1}{2}(t+1/t)$。

STEP 之外

代换 $t = x+\sqrt{1+x^2}$ 其实就是在 $x=\sinh u$ 时的 $t=\mathrm{e}^u$：事实上 $\sinh^{-1}x = \ln\bigl(x+\sqrt{1+x^2}\bigr)$，而公式 $x=\frac{1}{2}(t-1/t)$、$\sqrt{1+x^2}=\frac{1}{2}(t+1/t)$ 不过是用指数写出的 $\sinh u$ 与 $\cosh u$。这类双曲代换（hyperbolic substitution）是大学一年级处理含 $\sqrt{1+x^2}$ 的积分的标准工具。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）收录了更多同类风格的 STEP 积分题。

详细解答

STEP II 2012，第 4 题

纯数学 · 对数级数不等式

STEP II 2012，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/12-s2-q4

关键思路

一切都源自 $\ln(1+x)$ 的级数（series），以及 e^t 是递增函数（increasing function）这个事实。要比较 $(1+\frac{1}{n})^n$ 与 e，可以先取对数（logarithm）：我们需要比较 $n\ln(1+\frac{1}{n})$ 与 1。一个各项逐渐缩小的交错级数（alternating series）可以两项两项地加上括号，从而证明它小于首项；而组合 $\ln\frac{1+x}{1-x}$ 则把所有负项都消去，所以它大于首项。把数列夹在这两个界之间，便得出极限（limit）。

解答

第 (i) 部分

由于 $n>1$ 且 $k\geq 1$，我们有 $(k+1)n>n>1$，当然亦有 $(k+1)n>k$。两边乘以 $n^k>0$，得

$$(k+1)n^{k+1}>kn^k>0,$$

对这些正数取倒数（reciprocal），不等号方向会反转：

$$\frac{1}{(k+1)n^{k+1}}<\frac{1}{kn^k}.$$

现在在级数中代入 $x=\frac{1}{n}$；这是容许的，因为 $n>1$ 给出 $0<x<1$。把首项之后的各项两两分组，

$$\ln\left(1+\frac{1}{n}\right)=\frac{1}{n}-\left(\frac{1}{2n^2}-\frac{1}{3n^3}\right)-\left(\frac{1}{4n^4}-\frac{1}{5n^5}\right)-\cdots.$$

每个括号都具有 $\frac{1}{kn^k}-\frac{1}{(k+1)n^{k+1}}$ 的形式，其中 k 为偶数，因此由刚才证明的不等式可知它严格为正。

想一想 我们可以这样为一个无穷级数加上括号吗?

可以，因为这个级数收敛（converges）。项数为奇数的部分和（partial sums）构成全体部分和的一个子数列（subsequence），所以它们趋向同一个极限，即 $\ln(1+\frac{1}{n})$。加上括号后的级数，正正就是这些奇数项部分和组成的数列。

加括号后级数的部分和为 $\frac{1}{n}-b_1$、$\frac{1}{n}-b_1-b_2$、...，其中每个 $b_j>0$。它们递减，所以全部都不大于 $\frac{1}{n}-b_1$，因此它们的极限亦然:

$$\ln\left(1+\frac{1}{n}\right)\le\frac{1}{n}-b_1<\frac{1}{n},\qquad b_1=\frac{1}{2n^2}-\frac{1}{3n^3}>0.$$

(把 b_1 单独留出来，正是令最后的不等式成为严格不等式（strict inequality）的关键；否则一列小于 $\frac{1}{n}$ 的数，其极限有可能等于 $\frac{1}{n}$。)

两边乘以 $n>0$，得 $n\ln(1+\frac{1}{n})<1$，即 $\ln\left((1+\frac{1}{n})^n\right)<1$。由于指数函数（exponential function）是递增的，

$$\left(1+\frac{1}{n}\right)^n<\mathrm{e}.$$

第 (ii) 部分

设 $x=\frac{1}{2y}$。由于 $y>\frac{1}{2}$，我们有 $0<x<1$，所以 x 与 $-x$ 都落在级数成立的范围内:

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots,$$
$$\ln(1-x)=-x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}-\cdots.$$

逐项相减（对两个收敛级数而言这是合理的），偶次幂互相抵消，奇次幂则加倍:

$$\ln\left(\frac{1+x}{1-x}\right)=2\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots\right).$$

而

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

所以，利用 $2x=\frac{1}{y}$，

$$\ln\left(\frac{2y+1}{2y-1}\right)=\frac{1}{y}+\frac{2}{3(2y)^3}+\frac{2}{5(2y)^5}+\cdots=\frac{1}{y}\sum_{r=0}^{\infty}\frac{1}{(2r+1)(2y)^{2r}}.$$

首项之后的每一项都严格为正（因为 $y>0$），所以由第二个起的部分和全部都不小于 $\frac{1}{y}+\frac{2}{3(2y)^3}$，因此

$$\ln\left(\frac{2y+1}{2y-1}\right)\ge\frac{1}{y}+\frac{2}{3(2y)^3}>\frac{1}{y},$$

此即所求。

至于推论，选取 y 使分数变成 $1+\frac{1}{n}$。取 $y=n+\frac{1}{2}$（对每个正整数 n，它都大于 $\frac{1}{2}$），

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

所以

$$\ln\left(1+\frac{1}{n}\right)>\frac{1}{n+\frac{1}{2}}.$$

两边乘以 $n+\frac{1}{2}>0$，得 $\left(n+\frac{1}{2}\right)\ln(1+\frac{1}{n})>1$，再取指数（同样是递增函数），

$$\mathrm{e}<\left(1+\frac{1}{n}\right)^{n+\frac{1}{2}}.$$

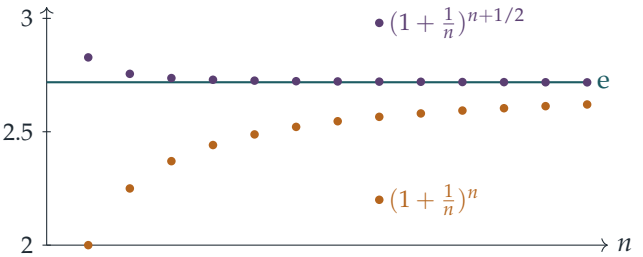
第 (iii) 部分

把第 (ii) 部分的结果除以 $(1+\frac{1}{n})^{1/2}>0$，并与第 (i) 部分结合，对每个整数 $n\geq 2$ 都有

$$\frac{\mathrm{e}}{\sqrt{1+\frac{1}{n}}}<\left(1+\frac{1}{n}\right)^{n}<\mathrm{e}.$$

当 $n\to\infty$ 时，$1+\frac{1}{n}\to 1$，所以 $\sqrt{1+\frac{1}{n}}\to 1$（平方根函数是连续（continuous）的），于是左边趋向 e。右边是常数 e。由夹逼定理（squeeze／sandwich theorem），

$$\left(1+\frac{1}{n}\right)^{n}\to\mathrm{e}.$$



验算

当 $n=10$：$(1.1)^{10}\approx 2.5937$，$(1.1)^{10.5}\approx 2.7203$，两者确实分处 $\mathrm{e}\approx 2.7183$ 的两侧。当 $n=1000$，下界 $\mathrm{e}/\sqrt{1.001}\approx 2.7169237$ 仅仅略低于 $(1.001)^{1000}\approx 2.7169239$，可见这个夹逼有多紧。我们亦对直至数千的每个 n 验证了两条严格不等式，并以数值方法验证了对随机选取的 $y>\frac{1}{2}$，第 (ii) 部分的级数之和等于 $\ln\frac{2y+1}{2y-1}$。

常见错误

- 在第 (i) 部分代入 $x = 1$：题目只给出 $|x| < 1$ 时的级数，这正是那里需要 $n > 1$ 的原因（第 (iii) 部分只用到很大的 n，所以并无损失）。
- 直接由一列严格不等式的极限得出严格不等式。极限只保持 $\leq$；要得到 $<$，须把一个正项单独留出来。
- 忘记取指数之所以保持不等式，只是因为 e^t 是递增的；应当明确写出这一点。

STEP 之外

这些界所显示的不止于极限：由于 $(1+\frac{1}{n})^{n+1/2}/(1+\frac{1}{n})^n = \sqrt{1+1/n} \approx 1 + \frac{1}{2n}$，误差 $\mathrm{e} - (1+\frac{1}{n})^n$ 的大小约为 $\mathrm{e}/(2n)$，所以这个数列收敛得颇慢。重排收敛级数、比较部分和以及夹逼定理，都是大学第一门分析（analysis）课程的基本功；在那里你亦会遇到其中的微妙之处（重排条件收敛（conditionally convergent）级数可以改变其和！）。Abbott 的 *Understanding Analysis*（《理解分析》）对这一切有精彩的论述。

详细解答

STEP II 2012，第 5 题

纯数学 · 倒数函数的描绘

STEP II 2012，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/12-s2-q5

关键思路

两条曲线都是倒数（reciprocal）形式，所以我们在脑中先想像分母的图像，再把它“倒过来”：分母为零之处会出现垂直渐近线（vertical asymptote），分母很大之处 y 接近 0，而分母的转向点（turning point）亦是曲线的转向点。在第 (ii) 部分，两个二次因式关于 a 与 b 的中点互为镜像，因此把原点移到该中点，g 便成为新变量的偶函数（even function），而驻点（stationary point）则可从一条关于 t^2 的二次方程直接得出。

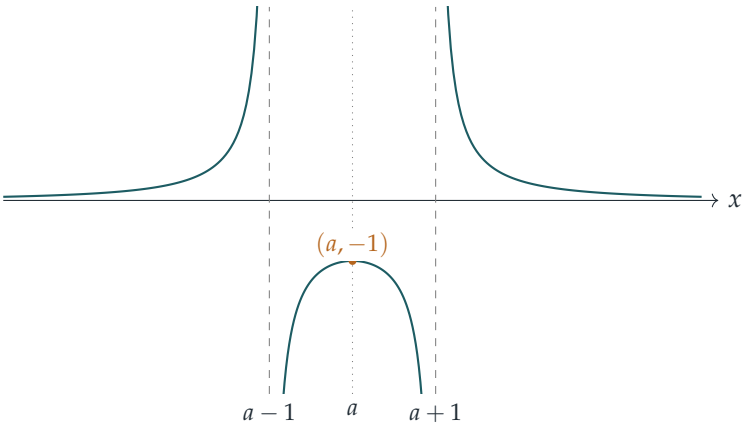
解答

第 (i) 部分

设 $D(x) = (x-a)^2 - 1$，于是 $f = 1/D$。分母 D 是一条抛物线，关于 $x = a$ 对称，在 $x = a$ 取最小值 -1，零点为 $x = a \pm 1$。由此读出倒数的性质：

- 曲线关于 $x = a$ 对称，垂直渐近线为 $x = a \pm 1$。
- 当 $|x-a| < 1$ 时，$-1 \le D < 0$，所以 $f \le -1$，等号只在 $x = a$ 成立。因此 $(a, -1)$ 是局部极大点（local maximum），而当 x 从内侧趋近 $a \pm 1$ 时，$f \to -\infty$。
- 当 $|x-a| > 1$ 时，$D > 0$，且 D 随 $|x-a|$ 增大而递增，所以 f 为正，并由 $+\infty$（在渐近线旁）递减趋向 0。x 轴是水平渐近线（horizontal asymptote），当 $|x|$ 很大时 $f \approx 1/x^2$。

事实上 $f'(x) = -2(x-a)/D^2$ 只在 $x = a$ 处为零。曲线与 y 轴相交于 $1/(a^2-1)$（当 $a \neq \pm 1$ 时），但其位置取决于 a，所以我们以 $x = a$ 为参照来作图。



第 (ii) 部分

设 $p=(x-a)^2-1$ 及 $q=(x-b)^2-1$，于是 $g=1/(pq)$。令

$$c=\tfrac{1}{2}(a+b),\qquad d=\tfrac{1}{2}(b-a)>0,\qquad t=x-c,$$

使得 $x-a=t+d$ 及 $x-b=t-d$。利用 $(t+d)(t-d)=t^2-d^2$ 及 $(t+d)^2+(t-d)^2=2t^2+2d^2$，得

$$\begin{aligned}pq&=\big((t+d)^2-1\big)\big((t-d)^2-1\big)\\&=(t^2-d^2)^2-(2t^2+2d^2)+1.\end{aligned}$$

这只取决于 t^2，所以 g 关于 $x=c$ 对称，即关于 a 与 b 的中点对称。记 $u=t^2$ 及 $h(u)=(u-d^2)^2-2u-2d^2+1$，则

$$\frac{\mathrm{d}(pq)}{\mathrm{d}t}=2t\,h'(u)=2t\big(2(u-d^2)-2\big)=4t\,(t^2-d^2-1).$$

由于 $g'=-(pq)'/(pq)^2$，g 的驻点就是 $(pq)'=0$ 之处：

$$t=0\qquad 或\qquad t=\pm\sqrt{d^2+1},$$

前提是这些点不是渐近线。渐近线位于 $t=\pm(d+1)$ 及 $t=\pm(d-1)$，而由于 $d<\sqrt{d^2+1}<d+1$，外侧的两个驻点必定是真正的驻点。以 a 和 b 表示，候选点为

$$x=\frac{a+b}{2}\qquad 及\qquad x=\frac{a+b\pm\sqrt{(b-a)^2+4}}{2}.$$

在这些点上的值为 $h(0)=(d^2-1)^2$ 及 $h(d^2+1)=1-2(d^2+1)-2d^2+1=-4d^2$。

? **想一想** 这些候选点中哪些真正成立？它们是极大点还是极小点？

这取决于 $t=0$ 是否与渐近线重合，即是否有 $d=1$；这正好对应 $b>a+2$（即 $d>1$）与 $b=a+2$（即 $d=1$）两种情况的区分。至于每个点的性质，我们不再求二阶导数，而是观察 g 在渐近线之间每个区间上的正负号，以及它在区间两端的行为。

情况 $b > a+2$。此时 $d > 1$，四条渐近线依次为

$$t = -(d+1),\ -(d-1),\ d-1,\ d+1, \quad \text{即} \quad x = a-1,\ a+1,\ b-1,\ b+1,$$

四者互不相同，且 c 严格位于 $a+1$ 与 $b-1$ 之间。

- 外侧，$|t| > d+1$：此处 $|x-a| > 1$ 且 $|x-b| > 1$，所以 $p, q > 0$，$g > 0$。这里没有驻点，所以 g 由渐近线处的 $+\infty$ 下降至无穷远处的 0，行为类似 $1/x^4$。
- 中间，$|t| < d-1$，即 $a+1 < x < b-1$：同样 $p, q > 0$，所以 $g > 0$，且在两端 $g \to +\infty$。唯一的驻点是 $t = 0$，因此它必定是极小点，其值为

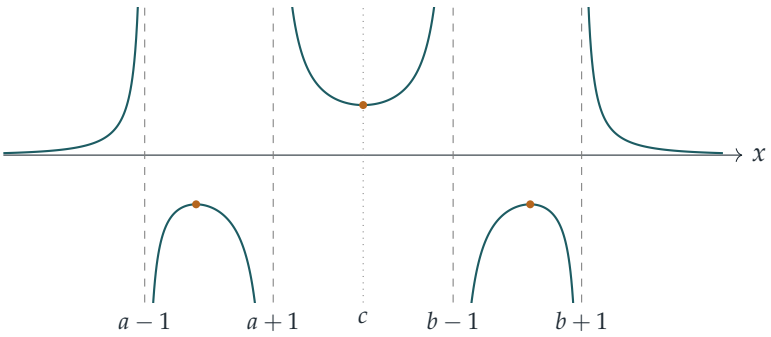
$$g(c) = \frac{1}{(d^2-1)^2} = \frac{16}{\big((b-a)^2-4\big)^2}.$$

- 两个间隙，$d-1 < |t| < d+1$：对于 $(d-1, d+1)$ 中的 t，有 $|x-b| < 1$，但 $x - a > 2d - 1 > 1$，所以 $q < 0 < p$，$g < 0$，且在两端 $g \to -\infty$。该处唯一的驻点是 $t = \sqrt{d^2+1}$，它必定是极大点，其值为

$$g = \frac{1}{-4d^2} = -\frac{1}{(b-a)^2}.$$

  间隙 $-(d+1) < t < -(d-1)$ 的情况与此成镜像。

因此共有三个驻点：在 $x = \frac{1}{2}(a+b)$ 的一个局部极小点，以及在 $x = \frac{1}{2}\big(a+b \pm \sqrt{(b-a)^2+4}\big)$ 的两个局部极大点，两者均有 $y = -1/(b-a)^2$。下图取 $b-a = 4.8$，并把 y 方向拉伸，使转向点清楚可见。



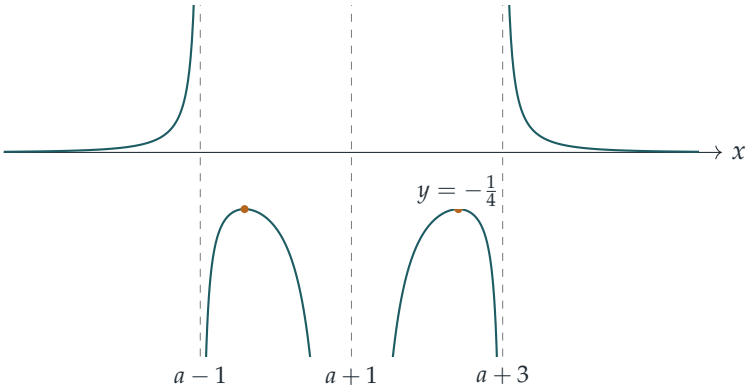
情况 $b = a+2$。此时 $d = 1$，所以 $a+1 = b-1 = c$，中间两条渐近线合而为一，位于 $x = a+1$。设 $t = x-a-1$，

$$pq = h(t^2) = t^4 - 4t^2 = t^2(t^2-4), \qquad g = \frac{1}{t^2(t^2-4)},$$

所以共有三条渐近线：$x = a-1$、$a+1$、$a+3$。候选点 $t = 0$ 现在是渐近线，而不是驻点，只剩下 $t = \pm\sqrt{2}$。

- 当 $|t|>2$ 时，$g>0$，且 g 由 $+\infty$ 下降至 0，与前相同。
- 当 $0<|t|<2$ 时，$g<0$，且无论 $t\to 0$（此处 $g\approx -1/(4t^2)$，所以两侧都向下）还是 $|t|\to 2$，都有 $g\to -\infty$。因此在 $t=\sqrt{2}$ 的驻点是极大点，其值为 $g=1/\big(2\cdot(-2)\big)=-\frac{1}{4}$；在 $t=-\sqrt{2}$ 亦然。

因此驻点位于 $x=a+1\pm\sqrt{2}$，两者都是局部极大点，且 $y=-\frac{1}{4}$。（这与一般公式吻合：当 $b-a=2$ 时，公式给出 $x=\frac{1}{2}(2a+2\pm\sqrt{8})$，以及 $-1/(b-a)^2=-\frac{1}{4}$。）此时曲线在 $x=a-1$ 与 $x=a+3$ 之间从不升到 x 轴以上。



验算

- 取 $a=0$、$b=5$：$g(2.5)=1/\big((6.25-1)^2\big)=1/27.5625$，而公式 $16/(25-4)^2=16/441$ 给出相同结果，约为 0.0363。在 $x=\frac{1}{2}(5+\sqrt{29})\approx 5.193$，直接代入计算得 $g\approx -0.04=-1/25$。
- 在第一种情况中令 $b\to a+2$，极小值 $16/\big((b-a)^2-4\big)^2$ 趋向无穷大，与中间那一支消失于合并后的渐近线的情况相符。
- 对随机的 a 和 b 以数值方法扫描 g' 的变号位置，发现当 $b>a+2$ 时恰有三个驻点，当 $b=a+2$ 时恰有两个，位置与上文一致。

常见错误

- 把 pq 展开成关于 x 的四次多项式（quartic），再尝试直接把三次式 $(pq)'$ 因式分解（factorise）：这样做行得通，但对称代换（substitution）$t=x-c$ 能令因式分解一目了然。
- 忘记在 $b=a+2$ 的情况中，点 $x=a+1$ 是渐近线（分母的二重零点），而不是转向点。

► 弄错 g 在间隙中的正负号：在该处 p、q 中恰有一个为负，所以 $g < 0$。

STEP 之外

由 D 描绘 $1/D$ 的图像，是定性地思考函数的初步体验：正负号、零点、对称性及在无穷远处的行为，在动用任何微积分之前已几乎告诉你一切。平移 $t = x - \frac{1}{2}(a+b)$ 是利用对称性把四次式化为关于 t^2 的二次式，这个技巧在代数和分析中反复出现。若想了解这类作图背后对极限和渐近线的严谨处理，可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）；若想做更多同类题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题》）。

详细解答

STEP II 2012，第 6 题

纯数学 • 圆内接四边形的面积

STEP II 2012，第 6 题

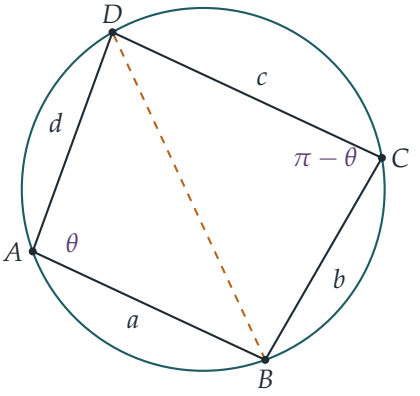
继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

step.maths.org/questions/12-s2-q6

关键思路

对角线（diagonal）BD 把四边形分成两个三角形，而 BD 可以用余弦定理（cosine rule）以两种方式表示。圆内接四边形（cyclic quadrilateral）的特别之处在于对角之和为 π，所以 C 处的角是 $\pi - \theta$。于是 $\cos C = -\cos\theta$ 而 $\sin C = \sin\theta$，余弦定理和面积公式 $\frac{1}{2}xy\sin(\text{夹角})$ 都只涉及同一个角 θ。之后，利用 $\sin^2\theta + \cos^2\theta = 1$ 再加上细心的因式分解（factorising），便可完成余下的工作。

解答



求 $\cos\theta$ 和 $\sin\theta$

由于 $ABCD$ 是圆内接四边形，对角 DAB 与 BCD 之和为 π，所以 $\angle BCD = \pi - \theta$。因此

$$\cos\angle BCD = -\cos\theta, \qquad \sin\angle BCD = \sin\theta.$$

对对角线 BD 应用余弦定理，先在三角形 ABD 中（两边为 a、d，夹角为 θ），再在三角形 CBD 中（两边为 b、c，夹角为 $\pi-\theta$）：

$$BD^2 = a^2 + d^2 - 2ad\cos\theta = b^2 + c^2 + 2bc\cos\theta.$$

把含 $\cos\theta$ 的项归在一起，得 $2(ad+bc)\cos\theta = a^2+d^2-b^2-c^2$，所以

$$\cos\theta = \frac{a^2+d^2-b^2-c^2}{2(ad+bc)}.$$

对角线 BD 把四边形分成这两个三角形，所以它们的面积之和为 Q：

$$Q = \tfrac{1}{2}ad\sin\theta + \tfrac{1}{2}bc\sin(\pi-\theta) = \tfrac{1}{2}(ad+bc)\sin\theta, \qquad \text{所以} \qquad \sin\theta = \frac{2Q}{ad+bc}.$$

（由于所有边长都是正数，我们可以除以 $ad+bc$。）

$16Q^2$ 的公式

把两个表达式代入 $\sin^2\theta + \cos^2\theta = 1$：

$$\frac{4Q^2}{(ad+bc)^2} + \frac{(a^2+d^2-b^2-c^2)^2}{4(ad+bc)^2} = 1.$$

两边乘以 $4(ad+bc)^2$，

$$16Q^2 = 4(ad+bc)^2 - (a^2+d^2-b^2-c^2)^2,$$

此即所求。

婆罗摩笈多公式

右方是平方差（difference of two squares）$X^2 - Y^2$，其中 $X = 2(ad+bc)$，$Y = a^2+d^2-b^2-c^2$。

? **想一想** 怎样把 $X-Y$ 和 $X+Y$ 化成一次因式的乘积?

把各项分组，使每一组再次成为平方差。把含 a,d 的项放在一起，含 b,c 的项放在一起：

$$X - Y = (b^2+2bc+c^2) - (a^2-2ad+d^2) = (b+c)^2 - (a-d)^2,$$
$$X + Y = (a^2+2ad+d^2) - (b^2-2bc+c^2) = (a+d)^2 - (b-c)^2.$$

再把两者各自因式分解，

$$\begin{aligned}16Q^2 &= (X-Y)(X+Y)\\ &= (b+c-a+d)(b+c+a-d)(a+d-b+c)(a+d+b-c).\end{aligned}$$

设 $2s = a+b+c+d$，则每个括号都等于 $2s$ 减去某一边的两倍：例如 $b+c+d-a = 2s-2a$。所以四个括号分别是 $2(s-a)$、$2(s-d)$、$2(s-b)$ 和 $2(s-c)$，而

$$16Q^2 = 16(s-a)(s-b)(s-c)(s-d), \qquad \text{即} \qquad Q^2 = (s-a)(s-b)(s-c)(s-d).$$

这就是婆罗摩笈多（Brahmagupta）公式。

三角形的面积

每个三角形都是圆内接的，因为它的三个顶点都在其外接圆（circumcircle）上。现在让边 DA 缩短至零，即 D 沿圆周滑向 A。这时 $ABCD$ 变成边长为 a、b、c 的三角形 ABC，而 $d = 0$。公式两边都随 D 连续变化，所以在极限下公式仍然成立。设 $s = \frac{1}{2}(a+b+c)$，则三角形的面积 Δ 满足

$$\Delta^2 = s(s-a)(s-b)(s-c), \qquad \text{所以} \qquad \Delta = \sqrt{s(s-a)(s-b)(s-c)}.$$

这就是海伦公式（Heron's formula）。

验算

- 单位正方形是圆内接四边形，$s = 2$，而 $(2-1)^4 = 1$ 正是它的面积。对于 3–4–5 三角形，$s = 6$，而 $\sqrt{6\cdot 3\cdot 2\cdot 1} = 6 = \frac{1}{2}\cdot 3\cdot 4$。
- 对于长方形，$a = c$ 且 $b = d$，所以 $\cos\theta = 0$；这是对的，因为长方形的每个角都是 $\frac{\pi}{2}$。
- 我们在圆上随机放置了 20 000 个圆内接四边形，把 $\cos\theta$、$\sin\theta$ 和 Q^2 的公式与实际的角及鞋带公式（shoelace formula）算出的面积比较，两者相差不超过 3×10^{-12}。海伦公式亦与 1000 个随机三角形的结果吻合。

常见错误

- 忘记 $\cos(\pi-\theta) = -\cos\theta$。这会把 $+2bc\cos\theta$ 变成减号，之后的一切都会出错。
- 试图把 $4(ad+bc)^2 - (a^2+d^2-b^2-c^2)^2$ 完全展开。把它当作平方差分解两次，快得多，也不容易出错。

STEP 之外

对于一般的（不一定是圆内接的）四边形，Bretschneider 公式可给出其面积。该公式减去一项 $abcd\cos^2\frac{\alpha+\gamma}{2}$，其中 α 和 γ 是一对对角，所以在边长固定时，圆内接四边形的面积最大。海伦公式的因式分解形式亦即时显示：三角形面积为零，当且仅当三角不等式（triangle inequality）取等号。如想看更多这种风格的几何，可参阅考克斯特（Coxeter）与 Greitzer 合著的 *Geometry Revisited*（《几何的新探索》）。

详细解答

STEP II 2012，第 7 题

纯数学 · 向量与圆

STEP II 2012，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：

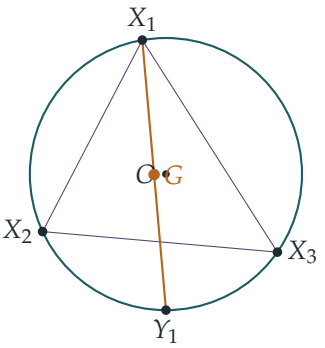
step.maths.org/questions/12-s2-q7

关键思路

点 Y_1 位于通过 X_1 和 G 的直线上，所以它的位置向量（position vector）是 $\mathbf{x}_1$ 与 $\mathbf{g}$ 的简单组合，只含一个未知纯量（scalar）。关于 Y_1，我们唯一知道的另一件事是它在单位圆（unit circle）上，因此 $|\overrightarrow{OY_1}|^2 = 1$ 给出一条关于 λ_1 的方程。这是一条二次方程（quadratic equation），但其中一个根事先已显而易见：该直线在 X_1 本身与圆相交，所以这个根可以先因式分解出来。

解答

设 $\mathbf{g} = \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)$ 及 $\lambda = \lambda_1$。由于每个 $\mathbf{x}_i$ 都是单位向量（unit vector），所以 $\mathbf{x}_i \cdot \mathbf{x}_i = 1$。



Y_1 的位置向量

由 $\overrightarrow{GY_1} = -\lambda\overrightarrow{GX_1}$ 得 $\mathbf{y}_1 - \mathbf{g} = -\lambda(\mathbf{x}_1 - \mathbf{g})$，所以

$$\begin{aligned}\overrightarrow{OY_1} = \mathbf{y}_1 &= (1+\lambda)\mathbf{g} - \lambda\mathbf{x}_1 \\ &= \tfrac{1}{3}(1+\lambda)(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3) - \lambda\mathbf{x}_1 \\ &= \tfrac{1}{3}\big((1-2\lambda)\mathbf{x}_1 + (1+\lambda)(\mathbf{x}_2 + \mathbf{x}_3)\big),\end{aligned}$$

因为 $\frac{1}{3}(1+\lambda) - \lambda = \frac{1}{3}(1-2\lambda)$。这就是所求的表达式。

求 λ_1

因为 Y_1 在圆上，$|\mathbf{y}_1|^2 = 1$，所以 $|3\mathbf{y}_1|^2 = 9$。我们需要

$$|\mathbf{x}_2 + \mathbf{x}_3|^2 = 1 + 2\alpha + 1 = 2 + 2\alpha, \qquad \mathbf{x}_1 \cdot (\mathbf{x}_2 + \mathbf{x}_3) = \beta + \gamma.$$

因此，把 $|3\mathbf{y}_1|^2$ 展开为它与自身的点积（dot product），得

$$(1-2\lambda)^2 + (1+\lambda)^2(2+2\alpha) + 2(1-2\lambda)(1+\lambda)(\beta+\gamma) = 9.$$

按 λ 的幂整理，再除以 2：

$$(3+\alpha-2\beta-2\gamma)\lambda^2 + (2\alpha-\beta-\gamma)\lambda - (3-\alpha-\beta-\gamma) = 0.$$

为了简洁，设

$$A = 3+\alpha-2\beta-2\gamma, \qquad B = 3-\alpha-\beta-\gamma,$$

并留意中间的系数正好是 $A - B$。所以方程为 $A\lambda^2 + (A-B)\lambda - B = 0$，可因式分解为：

$$(\lambda+1)(A\lambda - B) = 0.$$

> ? **想一想** 为甚么 $\lambda = -1$ 会是一个根？为甚么可以舍去它？

把 $\lambda = -1$ 代入 $\mathbf{y}_1 = (1+\lambda)\mathbf{g} - \lambda\mathbf{x}_1$，得 $\mathbf{y}_1 = \mathbf{x}_1$。通过 X_1 和 G 的直线上每一点都具有 $(1+\lambda)\mathbf{g} - \lambda\mathbf{x}_1$ 的形式，而这条直线除了在 Y_1 之外，也在 X_1 与圆相交，所以这条二次方程必然以 $\lambda = -1$ 为根。题目给出 $\lambda_1 > 0$，所以 $\lambda_1 \neq -1$，因此 $A\lambda_1 = B$。

还须确认可以除以 A。若 $A = 0$，则 $A\lambda_1 = B$ 迫使 $B = 0$。但

$$|\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3|^2 = 3 + 2(\alpha+\beta+\gamma), \qquad \text{所以} \qquad B = \tfrac{1}{2}\big(9 - |\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3|^2\big) = \tfrac{9}{2}\big(1 - |\mathbf{g}|^2\big).$$

由三角不等式（triangle inequality），$|\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3| \le 3$，而等号只在三个单位向量相等时成立；由于三点互不相同，所以 $B > 0$。因此 $A \neq 0$，且

$$\lambda_1 = \frac{B}{A} = \frac{3-\alpha-\beta-\gamma}{3+\alpha-2\beta-2\gamma},$$

得证。

最后一部分

由于 $\overrightarrow{GY_1} = -\lambda_1 \overrightarrow{GX_1}$ 且 $\lambda_1 > 0$，长度满足 $GY_1 = \lambda_1 GX_1$，所以

$$\frac{GX_1}{GY_1} = \frac{1}{\lambda_1} = \frac{3+\alpha-2\beta-2\gamma}{3-\alpha-\beta-\gamma}.$$

把标号轮换（$1 \to 2 \to 3 \to 1$，这会把 $\alpha \to \beta \to \gamma \to \alpha$），用同样的论证便得到另外两个比。分母 $B = 3-\alpha-\beta-\gamma$ 是对称的，所以三者的分母相同，而分子分别为

$$3+\alpha-2\beta-2\gamma, \qquad 3+\beta-2\gamma-2\alpha, \qquad 3+\gamma-2\alpha-2\beta.$$

三者之和为 $9+(1-2-2)(\alpha+\beta+\gamma) = 9-3(\alpha+\beta+\gamma) = 3B$。因此

$$\frac{GX_1}{GY_1} + \frac{GX_2}{GY_2} + \frac{GX_3}{GY_3} = \frac{3B}{B} = 3.$$

验算

- 等边三角形：此时 $G = O$，每条通过 G 的弦都是直径，所有比都等于 1。在公式中，$\alpha = \beta = \gamma = -\frac{1}{2}$，得 $\lambda_1 = \frac{9/2}{9/2} = 1$。
- 这里有一个几何解读：$A = \frac{1}{2}|\mathbf{x}_2 + \mathbf{x}_3 - 2\mathbf{x}_1|^2 = \frac{9}{2} GX_1^2$ 及 $B = \frac{9}{2}(1 - OG^2)$，所以 $\lambda_1 = B/A$ 表示 $GX_1 \cdot GY_1 = 1 - OG^2$，即相交弦定理（intersecting chords theorem），亦即圆幂（power of a point）定理。
- 数值验算：取 X_1, X_2, X_3 位于角度 0、1 和 2.5 弧度处，三个比 GX_i/GY_i 约为 1.13、0.31 和 1.57，其和在机器精度内等于 3。我们用同样方法测试了 20 000 个随机三角形。

常见错误

- 忘记 $|\mathbf{x}_2 + \mathbf{x}_3|^2$ 含有交叉项 2α，而不只是 $1+1$。
- 硬套二次方程求根公式，而没有看出根 $\lambda = -1$。
- 除以 $3+\alpha-2\beta-2\gamma$ 时没有说明它为何不等于零。

STEP 之外

这道题中隐藏的恒等式 $GX_1 \cdot GY_1 = 1 - OG^2$，就是一点对于圆的圆幂，而上述点积计算正是它的向量证明。用内积来表述几何，让长度和角度全都由 $\mathbf{u} \cdot \mathbf{v}$ 得出，这是大学第一个线性代数（linear algebra）课程中内积空间（inner-product space）的起点。斯特朗（Strang）的 *Introduction to Linear Algebra*（《线性代数导论》）就是从这个观点展开的。

详细解答

STEP II 2012，第 8 题

纯数学 · 非线性递推关系

STEP II 2012，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/12-s2-q8

关键思路

递推关系（recurrence）$u_{n+1}u_{n-1} = u_n^2 - q^2$ 是非线性（non-linear）的，但它对每一个 n 都成立，所以把它写成两个相邻 n 值的形式再相减，就能消去那个麻烦的常数 q^2。剩下的式子说明比值 $(u_{n+1} + u_{n-1})/u_n$ 永远不变，于是数列便化为一个常系数为 p 的线性递推关系（linear recurrence）。开首的不等式正正就是 $p > 2$ 这个命题，而它正是推动数列增长的原因。

解答

不等式

把所有项化为公分母（common denominator）$\alpha\beta$，

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

由于 $\beta - \alpha > q > 0$，两边平方（两边皆为正）得 $(\beta - \alpha)^2 > q^2$，所以分子为正。因为 $\alpha, \beta > 0$，分母 $\alpha\beta$ 亦为正。因此该式为正，得证。

第一个恒等式

把递推关系两边乘以 u_{n-1}：对每一个 $n \geq 1$，

$$u_{n+1}u_{n-1} = u_n^2 - q^2.$$

以 $n+1$ 代替 n（这仍然容许，因为 $n+1 \geq 1$），

$$u_{n+2}u_n = u_{n+1}^2 - q^2.$$

以第二式减去第一式，消去 q^2：

$$u_n u_{n+2} - u_{n-1}u_{n+1} = u_{n+1}^2 - u_n^2,$$

移项整理后得

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

对所有 $n \geq 1$ 成立，这就是所求的恒等式（identity）。

线性递推关系

想一想 刚证明的恒等式含有四个不同的项。它怎样能给出只含单一常数 p 的递推关系?

把它除以 $u_n u_{n+1}$；由于数列没有任何一项为零，这是容许的：

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

所以量 $v_n = (u_{n+1} + u_{n-1})/u_n$ 对所有 $n \geq 1$ 满足 $v_{n+1} = v_n$：它是常数。设其公共值为 p；则 $u_{n+1} + u_{n-1} = pu_n$，即

$$u_{n+1} - pu_n + u_{n-1} = 0 \qquad (n \geq 1).$$

要求 p，取 $n = 1$。由递推关系得 $u_2 = (\beta^2 - q^2)/\alpha$，所以

$$p = \frac{u_2 + u_0}{u_1} = \frac{1}{\beta}\left(\frac{\beta^2 - q^2}{\alpha} + \alpha\right) = \frac{\alpha^2 + \beta^2 - q^2}{\alpha\beta}.$$

这正是第一部分中的式子，所以当 $\beta - \alpha > q$ 时，我们已经知道 $p > 2$。

当 $\beta > \alpha + q$ 时数列递增

设 $d_n = u_{n+1} - u_n$。利用线性递推关系的形式 $u_{n+1} = pu_n - u_{n-1}$，

$$d_n - d_{n-1} = u_{n+1} - 2u_n + u_{n-1} = (p-2)u_n.$$

想一想 我们知道 $p - 2 > 0$，但为甚么 $(p-2)u_n$ 会是正数？我们还未知道 $u_n > 0$。

正值性与增长必须一并证明，方法是归纳法（induction）。命题：对每一个 $n \geq 1$，$u_n > 0$ 且 $d_{n-1} > 0$。

当 $n = 1$：$u_1 = \beta > 0$ 且 $d_0 = \beta - \alpha > q > 0$。

假设对某个 $n \geq 1$ 有 $u_n > 0$ 及 $d_{n-1} > 0$。那么由于 $p > 2$,

$$d_n = d_{n-1} + (p-2)u_n > 0, \qquad u_{n+1} = u_n + d_n > 0,$$

这就是 $n+1$ 的情况。由归纳法，对每一个 $n \geq 0$ 都有 $d_n = u_{n+1} - u_n > 0$，所以数列严格递增（strictly increasing）。（事实上，差本身也在递增，所以增长比线性更快。由于每一项都是正数，在这情况下“没有任何一项为零”这个条件自动成立。）

$\beta = \alpha + q$ 的情况

此时 $(\beta - \alpha)^2 - q^2 = 0$，所以 $p = 2$ 刚好成立，且 $d_n - d_{n-1} = 0$。所有差都等于 $d_0 = \beta - \alpha = q$，所以数列是等差数列（arithmetic progression）

$$u_n = \alpha + nq.$$

由于 $q > 0$，它仍然严格递增，但现在只是线性增长。直接验证：$(\alpha + (n+1)q)(\alpha + (n-1)q) = (\alpha + nq)^2 - q^2$，所以这个数列确实满足原来的递推关系。

验算

- 取 $\alpha = 1$、$\beta = 4$、$q = 2$（于是 $\beta > \alpha + q$）。递推关系给出 $1, 4, 12, 35, \frac{407}{4}, \ldots$，而 $p = \frac{1+16-4}{4} = \frac{13}{4}$；确实 $\frac{13}{4} \cdot 4 - 1 = 12$ 及 $\frac{13}{4} \cdot 12 - 4 = 35$。
- 取 $\alpha = 1$、$q = 2$、$\beta = 3$，各项为 $1, 3, 5, 7, \ldots$，即最后一部分的等差数列。
- 我们以精确的有理数运算，对数百组随机的 α, β, q 验证了两个恒等式，并验证了每当 $\beta > \alpha + q$ 时数列都是递增的。

常见错误

- 未先知道 $u_n > 0$，便由 $p > 2$ 推出 $d_n > d_{n-1}$：这两个事实必须放在同一个归纳法中。
- 除以 $u_n u_{n+1}$ 时没有指出它不等于零。
- 忘记相减那一步用到的是指标为 $n+1$ 的递推关系，而这需要 $n+1 \geq 1$。

STEP 之外

一个暗中满足某线性递推关系的非线性递推关系，我们说它具有一个隐藏的不变量（invariant）：这里是 $(u_{n+1} + u_{n-1})/u_n$，或等价地 $u_n^2 - u_{n+1}u_{n-1} = q^2$，它是力学中守恒量（conserved quantity）的离散类比。线性递推关系 $u_{n+1} - pu_n + u_{n-1} = 0$ 的解为

$u_n = A\lambda^n + B\lambda^{-n}$，其中 $\lambda + \lambda^{-1} = p$；当 $p > 2$ 时它是实数且呈指数增长，而在重根（repeated root）的情况 $p = 2$ 下则变为 $A + Bn$，正正就是上面的二分情况（dichotomy）。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）收录了更多 STEP 风格的递推关系题目，而与微分方程（differential equations）的平行关系，则在 Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）中有所发展。

详细解答

STEP II 2012，第 9 题

力学 • 抛体运动

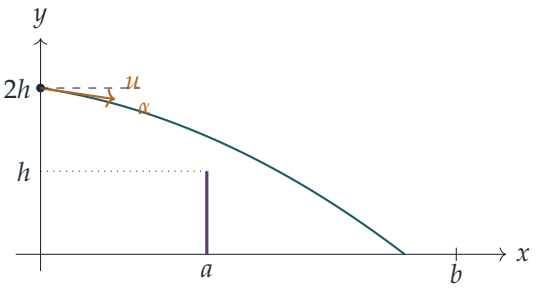
> **STEP II 2012，第 9 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/12-s2-q9

关键思路

把球的高度写成它与抛出点之间水平距离 x 的函数，也就是轨迹方程（trajectory equation）。这样，“越过球网”就是 $x = a$ 处的一个不等式，而“在 b 之前着地”则是关于飞行时间（time of flight）的一个不等式。最后一部分之所以行得通，是因为两个条件原来都是同一个量 $1/u^2$ 的界：一个是上界（upper bound），一个是下界（lower bound）。因此下界必须小于上界，u 便随之消去。

解答

我们把球视为只受重力作用而自由运动的质点（particle），忽略空气阻力（air resistance）和自旋（spin），并设 $0 \le \alpha < \frac{1}{2}\pi$。取原点为抛出点正下方的地面，$x$ 轴水平指向球网，y 轴铅直向上。



越过球网

初速度的水平分量为 $u\cos\alpha$，铅直分量为 $u\sin\alpha$，方向向下。经过时间 t 后，

$$x = ut\cos\alpha, \qquad y = 2h - ut\sin\alpha - \tfrac{1}{2}gt^2.$$

以 $t = x/(u\cos\alpha)$ 消去 t，得轨迹

$$y = 2h - x\tan\alpha - \frac{gx^2\sec^2\alpha}{2u^2}.$$

水平速度是正的常数，所以球恰好一次到达 $x = a$，而当该处 $y > h$ 时，球便越过球网：

$$2h - a\tan\alpha - \frac{ga^2\sec^2\alpha}{2u^2} > h \iff \frac{ga^2\sec^2\alpha}{2u^2} < h - a\tan\alpha.$$

两边乘以正数 $2/(ga^2\sec^2\alpha)$，得

$$\frac{1}{u^2} < \frac{2(h - a\tan\alpha)}{ga^2\sec^2\alpha},$$

此即所求。

在 b 之前着地

球下降了 $2h$ 时便着地，即 $\frac{1}{2}gt^2 + ut\sin\alpha - 2h = 0$ 之时。这个关于 t 的二次方程（quadratic equation）的两个根（roots）异号（两根之积为 $-4h/g < 0$），而着地时间是正根：

$$T = \frac{-u\sin\alpha + \sqrt{u^2\sin^2\alpha + 4gh}}{g}.$$

球在水平距离 b 之前着地，当且仅当 $uT\cos\alpha < b$，即 $T < \dfrac{b}{u\cos\alpha}$（因为 $u\cos\alpha > 0$）。两边乘以 g 再加上 $u\sin\alpha$，得

$$\sqrt{u^2\sin^2\alpha + 4gh} < \frac{bg}{u\cos\alpha} + u\sin\alpha,$$

这就是第二个结果。

最后的不等式

> ? **想一想** 两个涉及 u 的条件，怎样才能得出一个不含 u 的条件？

第一个结果是 $1/u^2$ 的上界。如果能把第二个结果化为 $1/u^2$ 的下界，那么无论 u 是多少，下界都必须小于上界。

第二个不等式两边皆为正（右边为正，是因为 $\sin\alpha \geq 0$ 且 $\cos\alpha > 0$），所以两边平方后不等号方向不变：

$$u^2\sin^2\alpha + 4gh < \frac{b^2g^2}{u^2\cos^2\alpha} + 2bg\tan\alpha + u^2\sin^2\alpha.$$

$u^2\sin^2\alpha$ 项互相抵消，剩下 $2g(2h - b\tan\alpha) < b^2g^2\sec^2\alpha/u^2$，即

$$\frac{2(2h - b\tan\alpha)}{gb^2\sec^2\alpha} < \frac{1}{u^2}.$$

与第一个结果合并，得

$$\frac{2(2h - b\tan\alpha)}{gb^2\sec^2\alpha} < \frac{1}{u^2} < \frac{2(h - a\tan\alpha)}{ga^2\sec^2\alpha}.$$

约去正因子 $2/(g\sec^2\alpha)$，再整条乘以 $a^2b^2 > 0$，

$$\begin{aligned} & a^2(2h - b\tan\alpha) < b^2(h - a\tan\alpha) \\ \iff\quad & (ab^2 - a^2b)\tan\alpha < h(b^2 - 2a^2) \\ \iff\quad & ab(b - a)\tan\alpha < h(b^2 - 2a^2). \end{aligned}$$

想一想 可以除以 $b - a$ 吗?

可以，但正负号很重要。在 $x = a$ 处，球仍高于高度 $h > 0$，所以它的着地距离大于 a；而它又在 b 之前着地。因此 $b > a$，故 $ab(b - a) > 0$，相除后不等号方向保持不变：

$$\tan\alpha < \frac{h(b^2 - 2a^2)}{ab(b - a)}.$$

验算

- 单位（units）：$a\tan\alpha$ 与 h 一样是长度，ga^2/u^2 也是长度，所以 $1/u^2$ 的每个界的单位都是 $\mathrm{s}^2\,\mathrm{m}^{-2}$，理应如此。
- 当 $\alpha = 0$ 时，最终结果变成 $b^2 > 2a^2$。事实上，水平抛出的球到 $x = a$ 时下降少于 h；由于下降量与 x^2 成正比，到 $x = \sqrt{2}\,a$ 时它下降少于 $2h$，所以它在 $\sqrt{2}\,a$ 之外着地，故 $b > \sqrt{2}\,a$。
- 数值上，取 $g = 9.8$、$u = 12$、$\alpha = 0.2$、$h = 1.5$，逐步模拟运动得着地距离为 6.7759，与 $uT\cos\alpha$ 吻合。在 20 000 个随机情况中，首两个不等式与“越过球网”及“在 b 之前着地”完全对应，而每当两者同时成立时，最后的不等式亦成立。

常见错误

- 把初速度的铅直分量当成 $+u\sin\alpha$：球是在水平线以下抛出的。
- 未检查两边是否皆为正，便把不等式平方。
- 除以 $b-a$ 时没有说明它为何是正数。

STEP 之外

最后一部分是在不等式之间消去参数（eliminating a parameter）的一个小例子：若 L 和 U 不含 u，则不等式组 $L<1/u^2<U$ 有解当且仅当 $L<U$。把同一想法系统地应用于大量线性不等式，便是 Fourier–Motzkin 消去法（Fourier–Motzkin elimination），它是线性规划（linear programming）的基础之一。至于力学方面，包括考虑空气阻力、轨迹不再是抛物线（parabola）的抛射体（projectile）运动，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP II 2012，第 10 题

力学 · 摩擦力与力矩

> **STEP II 2012，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/12-s2-q10

关键思路

杆是圆形横截面（cross-section）的一条弦（chord），所以每个法向反作用力（normal reaction）都直指圆心 O，而杆与每条半径之间的夹角 ϕ 由几何关系决定：$\cos\phi = a/r$。杆唯一可能的运动是绕 O 转动，因此两端会一同滑动，而我们也知道每一端的摩擦力（friction）朝哪个方向。对中点取力矩（moments）可消去重量，立即得出第一条关系式；再对 O 取力矩并分解（resolve）一次，便可求得 $\tan\theta$。

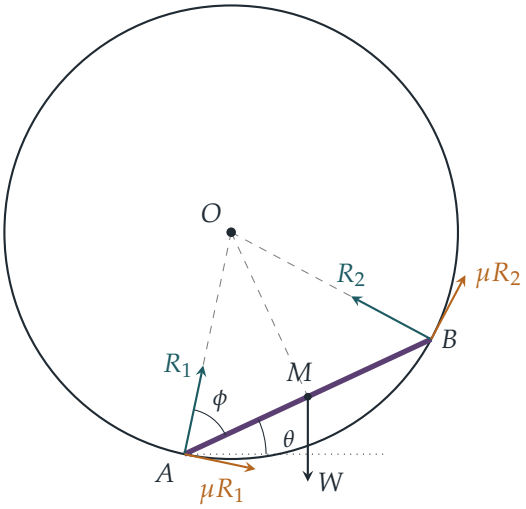
解答

建立模型

把杆视为重量为 W 的均匀刚性杆（uniform rigid rod），只在两端 A（较低）及 B（较高）与粗糙圆柱接触。设 O 为圆形横截面的圆心，M 为 AB 的中点。由于 $OA = OB = r$，三角形 OAB 是等腰三角形（isosceles triangle），所以 OM 垂直于 AB，且

$$\cos\phi = \frac{a}{r}, \qquad \sin\phi = \frac{\sqrt{r^2 - a^2}}{r}, \qquad OM = \sqrt{r^2 - a^2} = r\sin\phi,$$

其中 $0 < \phi < \frac{\pi}{2}$。杆与半径之间的夹角在两端都是 ϕ。



想一想 每一端的摩擦力朝哪个方向?

两端始终留在圆上，而 AB 的长度不变，所以 OM 的长度保持不变，并一直垂直于杆：唯一可能的运动是绕 O 的转动。由于 $\theta \neq 0$，M 并不在 O 的正下方，因此重量对 O 有一个力矩，使杆转向水平（在图中为顺时针：A 会沿壁向上移动，B 则向下移动）。反作用力都经过 O，所以只有摩擦力能抵抗这个转动。两端会一同移动，所以在即将滑动时，两端的摩擦力都达到极限（limiting friction）：μR_1 和 μR_2 分别沿 A 和 B 的切线（tangent）方向作用，两者相对于 O 都是逆时针方向，如图所示。

R_1 与 R_2 之间的关系

对 M 取力矩，从而消去重量。R_1 的作用线（line of action）是 AO，它与 AM 成角 ϕ，所以它与 M 的垂直距离是 $a\sin\phi$。A 处的摩擦力沿切线作用，而切线垂直于 OA，所以它与 M 的距离是 $a\cos\phi$。B 处的情况相同。在图中，R_1 使杆绕 M 顺时针转动，而 R_2 及两个摩擦力则使杆逆时针转动，所以

$$R_1\, a\sin\phi = R_2\, a\sin\phi + \mu R_1\, a\cos\phi + \mu R_2\, a\cos\phi.$$

除以 $a\cos\phi$ 并移项整理，得

$$\mu(R_1 + R_2) = (R_1 - R_2)\tan\phi,$$

此即所求。特别地，$R_1 > R_2$：较低的一端压得较用力。

求 $\tan\theta$

对 O 取力矩。反作用力都经过 O，而每个摩擦力的力臂都是 r。重量作用于 M，由于 OM 垂直于杆，它与铅垂线成角 θ，所以 M 与 O 的水平距离是 $OM\sin\theta = r\sin\phi\sin\theta$。因此

$$\mu(R_1+R_2)\,r = Wr\sin\phi\sin\theta \quad\Longrightarrow\quad \mu(R_1+R_2) = W\sin\phi\sin\theta. \tag{1}$$

沿垂直于杆的方向分解（指向 O 所在的一侧）。每个反作用力与杆成角 ϕ，所以贡献 $R_i\sin\phi$。每个摩擦力与这个垂直方向成角 ϕ；在 A 处它指离杆上 O 所在的一侧，在 B 处则指向该侧。重量贡献 $-W\cos\theta$。所以

$$(R_1+R_2)\sin\phi - \mu(R_1-R_2)\cos\phi = W\cos\theta. \tag{2}$$

记 $S = R_1+R_2$。由第一个结果得 $R_1-R_2 = \mu S\cot\phi$，所以 (2) 变为

$$S\,\frac{\sin^2\phi-\mu^2\cos^2\phi}{\sin\phi} = W\cos\theta,$$

而 (1) 表示 $\dfrac{\mu S}{\sin\phi} = W\sin\theta$。两式相除（杆是倾斜而非铅垂的，所以 $0<\theta<\frac{\pi}{2}$ 且 $\cos\theta\neq 0$），得

$$\tan\theta = \frac{\mu}{\sin^2\phi-\mu^2\cos^2\phi}.$$

分子分母同乘 r^2，并利用 $r^2\sin^2\phi = r^2-a^2$、$r^2\cos^2\phi = a^2$：

$$\tan\theta = \frac{\mu r^2}{r^2-a^2-\mu^2a^2} = \frac{\mu r^2}{r^2-a^2(1+\mu^2)},$$

此即所求。

推导 $\lambda<\phi$

> **想一想** 手上只有一条等式，不等式从何而来?

来自正负号。杆与水平成角 θ，且 $\theta\neq 0$，所以 $0<\theta<\frac{\pi}{2}$，$\tan\theta>0$。由于 $\mu>0$，分子 μr^2 为正，所以分母也必须为正：

$$r^2\sin^2\phi-\mu^2r^2\cos^2\phi>0 \quad\Longleftrightarrow\quad \tan^2\phi>\mu^2=\tan^2\lambda.$$

ϕ 和 λ 都在 $(0,\frac{\pi}{2})$ 内，而 tan 在此区间为正且递增，所以由 $\tan\phi>\tan\lambda$ 得 $\lambda<\phi$。

验算

- 光滑极限：当 $\mu\to 0$ 时，$\tan\theta\to 0$，所以光滑的杆只能水平静止，且 M 在 O 的正下方。正确。

- 数值：取 $r=5$、$a=3$、$\mu=\frac{1}{2}$，得 $\tan\phi=\frac{4}{3}>\frac{1}{2}$，以及 $\tan\theta=\frac{12.5}{25-11.25}=\frac{10}{11}$。在这个角度（以及数以千计随机选取的 r,a,μ）下，以向量形式解两条力的方程求 R_1,R_2，所得结果使对 O 的力矩为零，并在 10^{-12} 的误差内满足 $\mu(R_1+R_2)=(R_1-R_2)\tan\phi$。
- $\lambda<\phi$ 的物理意义：λ 是摩擦角（angle of friction）。若 $\lambda\geq\phi$，摩擦力足以在任何倾角下支撑着杆，所以杆永远不会处于即将滑动的状态，也就不存在这样的 θ。

常见错误

- 把反作用力取为垂直于杆，而不是沿半径方向：接触面是圆柱，所以法线是圆柱的法线。
- 把两个摩擦力沿杆取为相反方向。两端绕 O 的转动方向相同，所以两端的摩擦力相对于 O 都是逆时针方向。
- 为求第一个结果而沿杆及垂直于杆分解：两条方程都含有 W。对 M 取力矩则可避开它。

STEP 之外

推导 $\lambda<\phi$ 用的正是摩擦锥（friction cone）的概念：粗糙接触点的总反作用力必须落在与法线成角 λ 的范围之内。这里杆在两点接触，问题在于两个摩擦锥是否留有空间，让三个力交于一点。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）更全面地讨论摩擦力及刚体静力学（rigid-body statics），其中包括需要先判断即将发生的运动方向的题目。

详细解答

STEP II 2012，第 11 题

力学 · 非弹性碰撞

STEP II 2012，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：

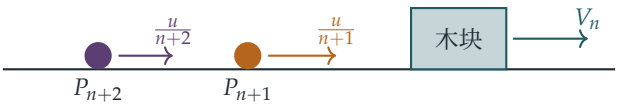
step.maths.org/questions/12-s2-q11

关键思路

留意每粒粒子所带的动量（momentum）：质量 im 乘以速率 u/i 等于 mu，每粒粒子都一样。因此吸收了 n 粒粒子之后，木块的动量就是 nmu，而动量守恒（conservation of momentum）立即给出它的速率。下一粒粒子会否到达，就只是比较两个速率；而能量损失则是“进入的动能减去最终的动能”。

解答

我们把木块和各粒子视为在光滑表面上沿同一直线运动的质点，因此没有水平方向的外力作用，每次碰撞中沿该直线的动量都守恒。由于粒子嵌入木块，每次碰撞都是完全非弹性碰撞（perfectly inelastic collision）。



第 n 次碰撞后的速率

粒子 P_i 的动量为 $im \cdot \frac{u}{i} = mu$，与 i 无关。

假设第 n 粒粒子与木块相撞。那么粒子 $P_1, \ldots, P_{n-1}$ 必定已在它之前相撞：各粒子按次序射出，而且每粒都比前一粒慢（u/i 递减），所以较后的粒子永远追不上较前的粒子，而任何粒子都不可能不撞上木块而越过它。因此在第 n 次碰撞刚完成时，木块包含了 $P_1, \ldots, P_n$，其质量为

$$km + m(1 + 2 + \cdots + n) = m\left(k + \tfrac{1}{2}n(n+1)\right).$$

木块起初静止，而每粒被吸收的粒子带来动量 mu，所以由动量守恒，此时总动量为 nmu。因此第 n 次碰撞后木块的速率 V_n 满足

$$m\Big(k+\tfrac{1}{2}n(n+1)\Big)V_n = nmu \quad\Longrightarrow\quad V_n = \frac{2nu}{2k+n(n+1)},$$

此即所求。

当 $2k = N(N+1)$ 时的碰撞次数

经过 n 次碰撞后，木块以恒定速率 V_n 运动（表面光滑）。位于木块后方的粒子 P_{n+1} 能追上木块，当且仅当它严格地较快：

$$\frac{u}{n+1} > \frac{2nu}{2k+n(n+1)} \iff 2k+n(n+1) > 2n(n+1) \iff 2k > n(n+1),$$

其中我们两边乘以正数 $n+1$ 及 $2k+n(n+1)$。（当 $n=0$ 时，此式为 $2k>0$，这是成立的：P_1 必定会撞上静止的木块。）

> **想一想** 一旦有一粒粒子追不上，较后的粒子还有可能撞上木块吗?

不可能。若 P_{n+1} 追不上木块，木块便永远保持速率 $V_n \geq \frac{u}{n+1}$，而之后的每粒粒子都更慢，因为对 $i > n+1$ 有 $\frac{u}{i} < \frac{u}{n+1}$。所以发生碰撞的正好是 $P_1, P_2, \ldots$，直至第一粒追不上的粒子为止。

现在取 $2k = N(N+1)$。P_{n+1} 撞上木块的条件变为 $n(n+1) < N(N+1)$，即 $n < N$，因为 $n(n+1)$ 在 $n \geq 0$ 时严格递增。所以 $P_1, \ldots, P_N$ 全部撞上木块，而 P_{N+1} 则不会：当 $n = N$ 时等号成立，而事实上

$$V_N = \frac{2Nu}{N(N+1)+N(N+1)} = \frac{u}{N+1},$$

恰好等于 P_{N+1} 的速率，因此它与木块之间始终保持固定距离。故此恰好有 N 次碰撞。

损失的动能

只有 $P_1, \ldots, P_N$ 参与碰撞，所以总损失等于这些粒子射出时的动能（kinetic energy）（木块起初静止）减去木块最终的动能。

起初，

$$\sum_{i=1}^{N} \tfrac{1}{2}(im)\Big(\frac{u}{i}\Big)^2 = \tfrac{1}{2}mu^2 \sum_{i=1}^{N} \frac{1}{i}.$$

最终木块的质量为 $m\big(k+\frac{1}{2}N(N+1)\big) = mN(N+1)$，速率为 $\frac{u}{N+1}$，所以其动能为

$$\tfrac{1}{2}\,mN(N+1)\,\frac{u^2}{(N+1)^2} = \tfrac{1}{2}mu^2\,\frac{N}{N+1} = \tfrac{1}{2}mu^2\Big(1-\frac{1}{N+1}\Big).$$

相减得

$$\begin{aligned}\text{损失的动能} &= \tfrac{1}{2}mu^2\left(\sum_{i=1}^{N}\frac{1}{i}-1+\frac{1}{N+1}\right)\\ &= \tfrac{1}{2}mu^2\left(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{N}+\frac{1}{N+1}\right)=\tfrac{1}{2}mu^2\sum_{n=2}^{N+1}\frac{1}{n},\end{aligned}$$

此即所求结果。

验算

- $N=1$（$k=1$）：一粒质量为 m、速率为 u 的粒子撞上质量为 m 的木块，得速率 $\frac{1}{2}u$，损失为 $\frac{1}{2}mu^2-\frac{1}{4}mu^2=\frac{1}{4}mu^2$，与 $\frac{1}{2}mu^2\cdot\frac{1}{2}$ 吻合。之后 P_2 的速率同样是 $\frac{1}{2}u$，永远不会到达。
- $N=2$（$k=3$）：公式给出损失 $\frac{1}{2}mu^2(\frac{1}{2}+\frac{1}{3})=\frac{5}{12}mu^2\approx 0.4167\,mu^2$。以分数精确地逐步模拟运动（位置及碰撞时刻），得到相同的数值、恰好两次碰撞，以及与公式吻合的速率 V_n；模拟亦确认了对每个不超过 11 的 N，都恰好有 N 次碰撞，且损失公式成立。
- 单位：损失是质量乘以速率的平方，正是能量应有的单位。

常见错误

- 只用第 n 粒粒子的动量，忘记木块已包含 $P_1,\ldots,P_{n-1}$（或忘记木块本身的质量 km）。
- 数成 $N+1$ 次碰撞：P_{N+1} 的速率恰好与木块相同，所以永远不会接触木块。
- 把 $P_{N+1},P_{N+2},\ldots$ 的动能计入损失；它们从不碰撞，所以并无损失任何动能。

STEP 之外

这木块是一个在扫集物质时质量不断增加的物体，与火箭、以及穿过云层下落的雨点属同类问题。对于这类变质量系统（variable-mass systems），我们把牛顿（Newton）第二定律以“动量的变化率等于力”的形式应用于整个系统，正如我们逐次碰撞所做的一样。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（《力学导论》）仔细地讨论了变质量系统和火箭运动。

详细解答

STEP II 2012，第 12 题

概率与统计 • 条件概率

> **STEP II 2012，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/12-s2-q12

关键思路

按一下开关会把它*切换*（toggle），所以泳池在走廊的开关最终处于“开”，要么是它本来就开着而没有被按，要么是它本来关着而被按了一下。厨房那一半其实是障眼法（red herring）：一个以概率 $\frac{1}{2}$ 处于“开”的开关，无论有没有被切换，仍然以概率 $\frac{1}{2}$ 处于“开”。第 (i) 部分得出一个条件概率（conditional probability）之后，七个星期一就构成一个二项（binomial）试验，而“最可能的值是 3”就转化为相邻二项概率之间的一对不等式。

解答

第 (i) 部分

定义以下事件（events）：

- H：客人按了走廊中泳池的开关，故 $\mathrm{P}(H) = \frac{1}{4}$；
- A：走廊的开关一开始全部处于“开”，故 $\mathrm{P}(A) = p$；
- L：最终泳池的灯亮着。

客人是随机选择开关的，所以 H 与 A 互相独立（independent）。

走廊。按过开关之后，泳池在走廊的开关处于“开”，当且仅当 A 发生而 H 不发生（它本来开着而没有被碰），或 A 不发生而 H 发生（它本来关着而被切换成开）。所以

$$\mathrm{P}(\text{走廊的泳池开关最终为开}) = p \cdot \tfrac{3}{4} + (1-p) \cdot \tfrac{1}{4} = \frac{1+2p}{4}.$$

厨房。泳池在厨房的开关一开始以概率 $\frac{1}{2}$ 处于“开”。若它没有被按，它保持原状；若它

被按，它就翻转，而一个以概率 $\frac{1}{2}$ 处于“关”的开关翻转后以概率 $\frac{1}{2}$ 处于“开”。无论客人在厨房做甚么，它最终处于“开”的概率都是 $\frac{1}{2}$。

厨房的开关以及客人在厨房的选择，与走廊的一切都互相独立，所以

$$\mathrm{P}(L) = \frac{1+2p}{4} \cdot \frac{1}{2} = \frac{1+2p}{8}.$$

至于条件概率，$H \cap L$ 要求客人按了泳池在走廊的开关、该开关原本处于“关”（这样按下去才会变成“开”），以及厨房的开关最终处于“开”：

$$\mathrm{P}(H \cap L) = \frac{1}{4} \cdot (1-p) \cdot \frac{1}{2} = \frac{1-p}{8}.$$

因此

$$\mathrm{P}(H \mid L) = \frac{(1-p)/8}{(1+2p)/8} = \frac{1-p}{1+2p},$$

得证。

第 (ii) 部分

记 $q = \dfrac{1-p}{1+2p}$。每个星期一我们都已知灯是亮着的，所以由第 (i) 部分，泳池在走廊的开关被按的概率是 q，而且每星期之间互相独立（客人各自独立行事，而每个星期一的初始状态都是重新产生的）。所以开关被按的日数 X 满足

$$X \sim \mathrm{B}(7, q).$$

我们也需要 $1-q$：

$$1 - q = \frac{(1+2p)-(1-p)}{1+2p} = \frac{3p}{1+2p}, \qquad \text{所以} \qquad \frac{q}{1-q} = \frac{1-p}{3p}.$$

（这里 $p > 0$；若 $p = 0$，则 $q = 1$，而 $X = 7$ 必然成立，所以 3 不可能是最可能的值。）

> **想一想** 怎样把“3 是最可能的值”变成可以计算的东西?

比较相邻的概率。对于 $1 \le k \le 7$,

$$\frac{\mathrm{P}(X=k)}{\mathrm{P}(X=k-1)} = \frac{\binom{7}{k} q^k (1-q)^{7-k}}{\binom{7}{k-1} q^{k-1} (1-q)^{8-k}} = \frac{8-k}{k} \cdot \frac{q}{1-q}.$$

因子 $\frac{8-k}{k}$ 随 k 增大而严格递减，所以当这个比值大于 1 时概率上升，一旦比值小于 1 概率便下降。因此 3 是（唯一的）最可能值，当且仅当概率由 $k=2$ 到 $k=3$ 上升，并由 $k=3$ 到 $k=4$ 下降：

$$\frac{\mathrm{P}(X=3)}{\mathrm{P}(X=2)} = \frac{5}{3} \cdot \frac{q}{1-q} > 1 \qquad \text{及} \qquad \frac{\mathrm{P}(X=4)}{\mathrm{P}(X=3)} = \frac{4}{4} \cdot \frac{q}{1-q} < 1.$$

（单凭这两个条件已足够证明我们需要的方向：若 3 是最可能的值，则必然有 $\mathrm{P}(X=3) > \mathrm{P}(X=2)$ 及 $\mathrm{P}(X=3) > \mathrm{P}(X=4)$。）

代入 $\frac{q}{1-q} = \frac{1-p}{3p}$，并在两边乘以正数 $3p$：

- 第二个条件给出 $1-p < 3p$，即 $p > \frac{1}{4}$；
- 第一个条件给出 $\frac{5}{3}(1-p) > 3p$，即 $5-5p > 9p$，所以 $p < \frac{5}{14}$。

因此 $\frac{1}{4} < p < \frac{5}{14}$，得证。

验算

- 当 $p=0$ 时，走廊的开关全部处于“关”，所以灯要亮，泳池的开关就必须被按过：第 (i) 部分给出 $\mathrm{P}(H \mid L) = 1$，正如所料。当 $p=1$ 时，灯要亮，该开关就必须没有被按，而公式给出 0。
- 两个端点分别对应 $q=\frac{1}{2}$（在 $p=\frac{1}{4}$）及 $q=\frac{3}{8}$（在 $p=\frac{5}{14}$）。当 $q=\frac{1}{2}$ 时，$\mathrm{B}(7,\frac{1}{2})$ 是对称的，所以 3 与 4 打和；当 $q=\frac{3}{8}$ 时，$(n+1)q=3$，2 与 3 打和。所以严格不等式恰到好处。
- 我们以分数精确地列举所有初始状态及所有开关选择，验证了第 (i) 部分；又以步长 $\frac{1}{2800}$ 的网格取 p 值，计算 $\mathrm{B}(7,q)$ 的概率，验证了第 (ii) 部分：唯一的众数（mode）是 3，当且仅当 $\frac{1}{4} < p < \frac{5}{14}$。

常见错误

- 忘记按一个已经处于“开”的开关会把它关掉：走廊的计算需要考虑通往“开”的两条途径。
- 明明答案就是 $\frac{1}{2}$，却把厨房分情况计算；分情况讨论虽然无害，却很容易出现算术错误。
- 在第 (ii) 部分用 p 而不是 q 作为二项分布的参数：正是“已知灯亮着”这个条件，把按该开关的概率由 $\frac{1}{4}$ 改变为 q。

STEP 之外

第 (i) 部分正是贝叶斯定理（Bayes' theorem）的实际应用：一个观察结果（灯亮着）把客人按了该开关的概率由 $\frac{1}{4}$ 更新为 $\frac{1-p}{1+2p}$。第 (ii) 部分则把这个想法倒过来运用，问未知参数 p 取甚么值时，才会与给定的最可能结果相符，这正是统计推断（statistical inference）

和最大似然估计（maximum likelihood estimation）的萌芽。用相邻项之比来找出众数的技巧，适用于许多离散分布（泊松分布（Poisson）、超几何分布（hypergeometric））。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）对条件概率及这些分布有详尽的论述。

详细解答

STEP II 2012，第 13 题

概率与统计 • 泊松过程的最近距离

STEP II 2012，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/12-s2-q13

本题在 *Beyond STEP* 第 12.19 节平面上的随机点中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/spatial-poisson.html

关键思路

“最近的超级市场距离至少为 y”这件事，恰好等同于“以我们所在点为圆心、半径为 y 的圆内没有任何超级市场”，而泊松模型（Poisson model）直接给出这件事的概率。由此得到 Y 的累积分布函数（cumulative distribution function），再求导便得到概率密度函数（probability density function）。之后各阶矩（moments）都是高斯（Gauss）型积分（Gaussian-type integrals）：一次分部积分法（integration by parts）就能把 $\mathrm{E}(Y)$ 化为题目容许我们假设的积分，而代换（substitution）$u = y^2$ 则使 $\mathrm{E}(Y^2)$ 变得十分初等。

解答

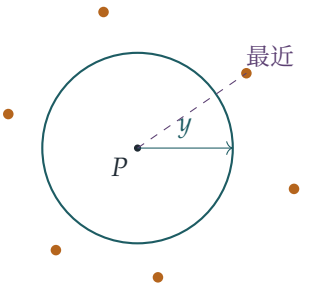
在整个解答中，我们把区域视为足够大，可以忽略边界效应（edge effects），因此对所有相关的 y，以所选点为圆心、半径为 y 的圆都位于区域之内。

圆内没有超级市场

设 N 为半径为 y 的圆内超级市场的数目。该圆的面积为 πy^2，所以 $N \sim \mathrm{Po}(\pi k y^2)$，并且

$$\mathrm{P}(N = 0) = \mathrm{e}^{-\pi k y^2}.$$

Y 的分布



设 P 为随机选取的点。最近的超级市场与 P 的距离至少为 y，当且仅当以 P 为圆心、半径为 y 的开圆内没有超级市场。因此，对 $y \geq 0$，

$$\mathrm{P}(Y \geq y) = \mathrm{e}^{-\pi k y^2}, \qquad \text{因此} \qquad \mathrm{P}(Y < y) = 1 - \mathrm{e}^{-\pi k y^2}.$$

（由于 Y 是连续随机变量（random variable），写 $<$ 还是 $\leq$ 并无分别。）这就是 Y 的累积分布函数 $F(y)$；而当 $y < 0$ 时 $F(y) = 0$，因为距离不可能是负数。对 $y \geq 0$ 求导，

$$f(y) = F'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left(1 - \mathrm{e}^{-\pi k y^2}\right) = 2\pi k y\, \mathrm{e}^{-\pi k y^2},$$

这正是所求的概率密度函数。简单验证一下：当 $y \to \infty$ 时 $F(y) \to 1$，所以 f 的积分等于 1。

期望值

$$\mathrm{E}(Y) = \int_0^\infty y \cdot 2\pi k y\, \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y.$$

想一想 指数函数前面有一个 y^2，我们怎样才能利用题目给出的积分？

把被积函数拆成 $y \times \left(2\pi k y\, \mathrm{e}^{-\pi k y^2}\right)$：第二个因子正是 $-\mathrm{e}^{-\pi k y^2}$ 的导数。取 $u = y$ 进行分部积分，

$$\mathrm{E}(Y) = \left[-y\, \mathrm{e}^{-\pi k y^2}\right]_0^\infty + \int_0^\infty \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y = \int_0^\infty \mathrm{e}^{-\pi k y^2}\, \mathrm{d}y,$$

因为当 $y \to \infty$ 时 $y\, \mathrm{e}^{-\pi k y^2} \to 0$（指数函数占上风），而且它在 $y = 0$ 时等于零。现在把指数配成 $-\frac{1}{2}x^2$：设 $\pi k y^2 = \frac{1}{2}x^2$，即 $y = x/\sqrt{2\pi k}$，$\mathrm{d}y = \mathrm{d}x/\sqrt{2\pi k}$。积分上下限仍然是 0 和 ∞，所以

$$\mathrm{E}(Y) = \frac{1}{\sqrt{2\pi k}} \int_0^\infty \mathrm{e}^{-x^2/2}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi k}} \sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{k}}.$$

方差

计算 $\mathrm{E}(Y^2)$ 时完全不需要题目给出的积分，因为 y 的奇次幂让我们可以代换 $u = \pi k y^2$，$\mathrm{d}u = 2\pi k y\,\mathrm{d}y$：

$$\mathrm{E}(Y^2) = \int_0^\infty y^2 \cdot 2\pi k y\,\mathrm{e}^{-\pi k y^2}\,\mathrm{d}y = \int_0^\infty \frac{u}{\pi k}\,\mathrm{e}^{-u}\,\mathrm{d}u = \frac{1}{\pi k}\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = \frac{1}{\pi k},$$

这里用到 $\int_0^\infty u\,\mathrm{e}^{-u}\,\mathrm{d}u = 1$（可用分部积分求得，或者认出它是参数为 1 的指数分布（exponential distribution）的期望值）。因此

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - \big(\mathrm{E}(Y)\big)^2 = \frac{1}{\pi k} - \frac{1}{4k} = \frac{4-\pi}{4\pi k},$$

正如所求。

验算

- 标度。若超级市场的密度变为 k 倍，长度应缩小 $\sqrt{k}$ 倍，所以 $\mathrm{E}(Y) \propto k^{-1/2}$，而 $\mathrm{Var}(Y) \propto k^{-1}$。两个答案都符合这个形式。
- 正值。$4 - \pi > 0$，所以方差是正数，理应如此。
- 数值。取 $k = 1$，对概率密度函数作中点求积法（midpoint quadrature）得出 $\mathrm{E}(Y) = 0.5000$ 及 $\mathrm{Var}(Y) = 0.06831$，与 $(4-\pi)/(4\pi) \approx 0.06831$ 吻合。对某点周围超级市场的泊松随机散布进行 10^5 次模拟，得出平均值 0.4996 及方差 0.0681。

常见错误

- 写成 $\mathrm{P}(Y < y) = \mathrm{e}^{-\pi k y^2}$：这其实是距离 y 以内*没有*超级市场的概率，即 $\mathrm{P}(Y \geq y)$。
- 代换时用了 $x = \sqrt{\pi k}\,y$ 而不是 $x = \sqrt{2\pi k}\,y$，结果得到 e^{-x^2} 而非题目假设的 $\mathrm{e}^{-x^2/2}$，答案相差 $\sqrt{2}$ 倍。

STEP 之外

这里的超级市场构成平面上的一个泊松点过程（Poisson point process），这是“完全空间随机性”（complete spatial randomness）的基本模型，可用于描述森林中的树木、天上的星星以及流动电话发射站。Y 的分布是瑞利分布（Rayleigh distribution），它也是两个分量均为独立正态随机变量的向量之长度所服从的分布；而我们的分部积分所揭示的简洁尾部公式 $\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y \geq y)\,\mathrm{d}y$，对任何非负随机变量都成立。Grimmett 和 Welsh 所著的 *Probability: An Introduction*（《概率导论》）涵盖了泊松过程以及这些尾部积分的概念。

详细解答

STEP III 2012，第 1 题

纯数学 · 非线性微分方程

STEP III 2012，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/12-s3-q1

关键思路

两条微分方程（differential equation）都是二阶（second order）而且非线性（non-linear）的，所以没有现成的标准解法。窍门在于：只要适当地选取 n，$\mathrm{d}z/\mathrm{d}x$ 中的括号就可以与每条方程的左边吻合。这样，方程便表示 $\mathrm{d}z/\mathrm{d}x$ 等于某个可以立即积分的式子，从而得到一条可以分离变量（separation of variables）的一阶（first-order）方程。

解答

预备结果

记 $y' = \mathrm{d}y/\mathrm{d}x$ 及 $y'' = \mathrm{d}^2y/\mathrm{d}x^2$。由乘积法则（product rule），并对 y^n 及 $(y')^2$ 使用链式法则（chain rule），

$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}x} &= ny^{n-1}y'\,(y')^2 + y^n \cdot 2y'y'' \\ &= y^{n-1}y'\big(n(y')^2 + 2yy''\big),\end{aligned}$$

这正是所求的结果。

第 (i) 部分

左边是 $(y')^2 + 2yy''$，即 $n = 1$ 时的括号。因此取 $z = y(y')^2$，则

$$\frac{\mathrm{d}z}{\mathrm{d}x} = y'\big((y')^2 + 2yy''\big) = y'\sqrt{y} = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\tfrac{2}{3}y^{3/2}\Big).$$

积分得 $y(y')^2 = \frac{2}{3}y^{3/2} + C$。在 $x = 0$ 时有 $y = 1$ 及 $y' = 0$，所以 $0 = \frac{2}{3} + C$，于是

$$y\,(y')^2 = \tfrac{2}{3}\big(y^{3/2} - 1\big). \qquad (*)$$

想一想 $(*)$ 只告诉我们 $(y')^2$。开平方时应取哪个符号？而 $y\equiv 1$（它使两边都等于零）又是否可能？

常数 $y\equiv 1$ 并不是原方程的解：代入会得到 $0=1$。事实上，把 $x=0$ 代入方程得 $0+2y''(0)=1$，所以 $y''(0)=\frac12>0$。因此 y 在 $x=0$ 处取极小值，而对细小的 $x>0$ 有 $y'>0$。此外，方程在 $x\mapsto -x$ 下保持不变，所以解是偶函数（even function），我们可以只考虑 $x>0$、$y'>0$、$y>1$ 的情况。

把 $(*)$ 乘以 y，取正平方根，再分离变量：

$$\frac{\sqrt{y}\,y'}{\sqrt{y^{3/2}-1}}=\sqrt{\tfrac23}.$$

代换 $w=y^{3/2}-1$，则 $\mathrm{d}w=\frac32 y^{1/2}\,\mathrm{d}y$。左边变成 $\frac23 w^{-1/2}\,\mathrm{d}w/\mathrm{d}x$，由 $x=0$（此时 $w=0$）起积分，

$$\tfrac43\sqrt{w}=\sqrt{\tfrac23}\,x\quad\Longrightarrow\quad w=\tfrac{9}{16}\cdot\tfrac23\,x^2=\tfrac38x^2.$$

所以 $y^{3/2}=\frac38x^2+1$，而由偶函数性质，此式对 $x<0$ 亦成立。因此

$$y=\left(\tfrac38x^2+1\right)^{2/3},$$

得证。

直接验算：设 $u=\frac38x^2+1$，则 $y'=\frac23u^{-1/3}\cdot\frac34x=\frac12xu^{-1/3}$，所以 $y(y')^2=u^{2/3}\cdot\frac14x^2u^{-2/3}=\frac14x^2$。另一方面 $\frac23(y^{3/2}-1)=\frac23\cdot\frac38x^2=\frac14x^2$，所以 $(*)$ 成立。

第 (ii) 部分

现在左边是 $(y')^2-yy''$。我们需要 $n(y')^2+2yy''$ 是它的倍数，所以 $n:2=1:-1$，得 $n=-2$。取 $z=y^{-2}(y')^2$（这是可以的，因为 y 接近 1，故不为零），

$$\frac{\mathrm{d}z}{\mathrm{d}x}=y^{-3}y'\left(-2(y')^2+2yy''\right)=-2y^{-3}y'\left((y')^2-yy''\right)=-2y^{-3}y'\left(-y^2\right)=\frac{2y'}{y},$$

这里用了方程的形式 $(y')^2-yy''=-y^2$。积分得 $z=2\ln y+C$，而 $x=0$ 给出 $0=0+C$。所以

$$\left(\frac{y'}{y}\right)^2=2\ln y.$$

设 $v=\ln y$，则 $v'=y'/y$，方程变成 $(v')^2=2v$，且 $v(0)=0$。

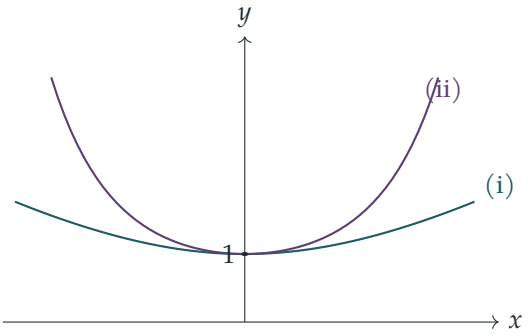
与第 (i) 部分一样，$v\equiv 0$（即 $y\equiv 1$）满足这条一阶方程，却不满足原方程（代入会得到 $1=0$）。在 $x=0$ 时，原方程给出 $0-y''(0)+1=0$，所以 $y''(0)=1>0$。方程同样在 $x\mapsto -x$ 下不变，所以我们取 $x>0$，此时 $v'>0$。于是

$$\frac{v'}{\sqrt{2v}}=1\quad\Longrightarrow\quad\sqrt{2v}=x\quad\Longrightarrow\quad v=\tfrac12x^2,$$

而由偶函数性质，此式对所有 x 成立。因此

$$y = \mathrm{e}^{x^2/2}.$$

验算：$y' = xy$ 及 $y'' = (1+x^2)y$，所以 $(y')^2 - yy'' + y^2 = y^2\left(x^2 - 1 - x^2 + 1\right) = 0$。此外 $y(0) = 1$ 及 $y'(0) = 0$。



验算

- 两个解都是偶函数，并在 $x = 0$ 处取极小值 1，与 $y''(0)$ 的值所预示的一致。
- 第 (ii) 部分有一条更快的途径：直接把 $y = \mathrm{e}^v$ 代入方程，得 $y' = v'\mathrm{e}^v$ 及 $y'' = (v'' + (v')^2)\mathrm{e}^v$，所有项消去后只剩 $v'' = 1$。所以立即得 $v = \frac{1}{2}x^2$。
- 数值上，以龙格－库塔法（Runge–Kutta method）由 $y(0) = 1$、$y'(0) = 0$ 求解 (i)，得 $y(2) \approx 1.842016$，与 $(\frac{5}{2})^{2/3}$ 吻合。对 (ii) 则得 $y(1) \approx 1.648721 = \mathrm{e}^{1/2}$。

常见错误

- 忘记 z 的积分常数(constant of integration)。在第 (i) 部分，初始条件(initial conditions)使它不为零：$C = -\frac{2}{3}$。
- 在第 (ii) 部分照搬类比而选 $n = 1$。$(y')^2$ 与 yy'' 的系数之比必须吻合，这就迫使 $n = -2$。
- 从一阶方程接受 $y \equiv 1$，却没有用原方程加以检验。

STEP 之外

这里的方法是二阶方程的积分因子(integrating factor)。把 $(y')^2 + 2yy''$ 乘以 y'，就变成一个恰当导数（exact derivative），正如把 $\ddot{x} = -\omega^2 x$ 乘以 $\dot{x}$ 便得到能量守恒（conservation

of energy）一样。像这两条不显含 x 的方程，总可以把 $p = y'$ 看成 y 的函数，从而降为一阶。对 $y \equiv 1$ 的顾虑，让我们窥见解的非唯一性（non-uniqueness）：$v' = \sqrt{2v}$ 在 $v = 0$ 处不满足利普希茨条件（Lipschitz condition）。Boyce 与 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖了降阶法（reduction of order）以及存在唯一性定理（existence and uniqueness theorem）。

详细解答

STEP III 2012，第 2 题

纯数学 · 无穷乘积与级数

STEP III 2012，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/12-s3-q2

关键思路

第 (i) 部分的乘积可以利用平方差（difference of two squares）逐个括号地合并，因此会像伸缩（telescoping）一样化简成非常简单的式子。把 $\frac{1}{1-x}$ 写成无穷乘积（infinite product）之后，取对数（logarithm）就把乘积变成和，再把这个和逐项微分（differentiate term by term），便得到分式组成的级数（series）。第 (ii) 部分是同一技巧的变奏：$1+x+x^2$ 是两个 $1-t$ 形式的式子之商，所以第 (i) 部分的结果可以分别应用到两者身上。

解答

全文假设 $|x|<1$，而且按题目所允许，我们不理会收敛（convergence）的问题。

第 (i) 部分

从左边开始合并括号，每次都用 $(1-a)(1+a)=1-a^2$：

$$\begin{aligned}(1-x)(1+x)&=1-x^2,\\(1-x^2)(1+x^2)&=1-x^4,\\(1-x^4)(1+x^4)&=1-x^8,\end{aligned}$$

如此类推。每一步都令幂次加倍，而最后一个括号 $1+x^{2^n}$ 把 $1-x^{2^n}$ 变成 $1-x^{2^{n+1}}$。所以

$$(1-x)(1+x)(1+x^2)(1+x^4)\cdots(1+x^{2^n})=1-x^{2^{n+1}}.$$

由于 $x\neq 1$，我们可以除以 $1-x$：

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

移项后即得题目所要求的 $\frac{1}{1-x}$ 的表达式。

现在令 $n \to \infty$。因为 $|x| < 1$，所以 $x^{2^{n+1}} \to 0$，余项消失，于是

$$\frac{1}{1-x} = \prod_{r=0}^{\infty}(1 + x^{2^r}).$$

每个因子都是正数（因为 $x^{2^r} > -1$），$1 - x$ 亦然，所以可以取对数。乘积的对数等于各对数之和：

$$-\ln(1-x) = \sum_{r=0}^{\infty} \ln\left(1 + x^{2^r}\right),$$

两边乘以 -1 即得所求结果。

想一想 怎样由对数之和得到分式之和?

两边对 x 求导。根据连锁法则（chain rule），$\ln(1 + x^{2^r})$ 的导数（derivative）是

$$\frac{2^r x^{2^r-1}}{1 + x^{2^r}}.$$

左边则有 $\frac{\mathrm{d}}{\mathrm{d}x}\left(-\ln(1-x)\right) = \frac{1}{1-x}$。逐项微分（这正是题目叫我们不必理会的收敛问题）得

$$\frac{1}{1-x} = \sum_{r=0}^{\infty} \frac{2^r x^{2^r-1}}{1 + x^{2^r}} = \frac{1}{1+x} + \frac{2x}{1+x^2} + \frac{4x^3}{1+x^4} + \cdots,$$

其中列出的是 $r = 0, 1, 2$ 的项。这就是所要求的级数。

第 (ii) 部分

左边是 $\ln(1 + x + x^2)$ 的导数，因为 $\frac{\mathrm{d}}{\mathrm{d}x}(1 + x + x^2) = 1 + 2x$。因此我们要为 $1 + x + x^2$ 找一个对数和的表达式，而因式分解 $1 - x^3 = (1 - x)(1 + x + x^2)$ 正好提供了一个：

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

由于 $|x^3| < 1$ 同样成立，第 (i) 部分的对数结果在以 x^3 代替 x 后仍然成立。记 $y_r = x^{2^r}$，使得 $(x^3)^{2^r} = y_r^3$，则

$$\ln(1 + x + x^2) = -\sum_{r=0}^{\infty} \ln\left(1 + y_r^3\right) + \sum_{r=0}^{\infty} \ln\left(1 + y_r\right) = -\sum_{r=0}^{\infty} \ln \frac{1 + y_r^3}{1 + y_r}.$$

立方和（sum of cubes）可分解为 $1 + y^3 = (1 + y)(1 - y + y^2)$，所以每个分式都等于 $1 - y_r + y_r^2$，从而

$$\ln(1 + x + x^2) = -\sum_{r=0}^{\infty} \ln\left(1 - x^{2^r} + x^{2^{r+1}}\right).$$

（每个 $1 - y + y^2 = (y - \frac{1}{2})^2 + \frac{3}{4}$ 都是正数，所以这些对数都有意义。）

想一想 求导时，负号到哪里去了？

连锁法则给出

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

而这个负号正好抵消和式前面的负号。两边逐项微分，得

$$\frac{1+2x}{1+x+x^2}=\sum_{r=0}^{\infty}\frac{2^r x^{2^r-1}\left(1-2x^{2^r}\right)}{1-x^{2^r}+x^{2^{r+1}}}.$$

首三项为

$$\begin{aligned}
r=0:&\quad \frac{1-2x}{1-x+x^2},\\
r=1:&\quad \frac{2x(1-2x^2)}{1-x^2+x^4}=\frac{2x-4x^3}{1-x^2+x^4},\\
r=2:&\quad \frac{4x^3(1-2x^4)}{1-x^4+x^8}=\frac{4x^3-8x^7}{1-x^4+x^8},
\end{aligned}$$

这正是题目要我们证明的级数。

另一途径。恒等式 $(1+y+y^2)(1-y+y^2)=1+y^2+y^4$ 直接为 $1+x+x^2$ 提供一个伸缩乘积，就像第 (i) 部分那个一样：$1+x+x^2=\left(1+x^{2^{n+1}}+x^{2^{n+2}}\right)\big/\prod_{r=0}^{n}\left(1-x^{2^r}+x^{2^{r+1}}\right)$，而分子趋向 1。

验算

- 代入 $x=0$：每条级数的两边都等于 1。
- 当 $x=\frac{1}{2}$ 时，第 (ii) 部分的左边是 $\frac{2}{7/4}=\frac{8}{7}\approx 1.142857$。右边首四项之和为 1.142369，而由于幂 x^{2^r} 缩小得极快，误差也缩小得极快。
- 我们用分数精确验证了伸缩后的乘积，并取 30 项，在 $(-0.95,0.95)$ 内随机抽取 1000 个 x，以数值方法验证了两条级数及各对数恒等式。

常见错误

- 用连锁法则时漏掉因子：x^{2^r} 的导数是 $2^r x^{2^r-1}$，而不是 $2^r x^{2^r}$。
- 第 (ii) 部分的整体符号弄错：$1-y+y^2$ 的导数是 $-1+2y$，正是这个负号抵消了和式前面的负号。

STEP 之外

乘积 $\prod(1+x^{2^r})$ 表明每个正整数都可以唯一地写成若干个不同的 2 的幂之和，这就是以生成函数（generating function）表达的二进制（binary）记数法。“不理会收敛”其实掩盖了一些真正的问题：甚么时候可以对无穷乘积取对数？甚么时候可以把级数逐项微分？大学第一个分析（analysis）课程会用一致收敛（uniform convergence）来回答这些问题，而在 $|x| \leq c < 1$ 上，这里的一切都是一致收敛的。清晰的讲解可参阅 Abbott 的 *Understanding Analysis*（《理解分析》）。

详细解答

STEP III 2012，第 3 题

纯数学 · 相切曲线与四次方程

> **STEP III 2012，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/12-s3-q3

关键思路

把 $y = 4 - x^2$ 代入第二条方程，得到一条关于 x 的四次方程（quartic），它的根就是两曲线交点的 x 坐标。“相切”即是有一个重根（repeated root）。与其把 m 和 k 当作未知数，不如用切点的 x 坐标 a 来标记一切：这样 m、k 以及其余所有交点，结果都是 a 的简单函数。由此可以准确知道四幅草图各自的样子，而第 (ii) 部分就是 $m = 12$ 的特殊情况。

解答

准备工作：两曲线可在哪里相切？

把两条曲线记作 $C_1\colon y = 4 - x^2$ 和 $C_2\colon mx = k - y^2$。C_1 是大家熟悉的开口向下的抛物线（parabola），经过 $(\pm 2, 0)$ 和 $(0, 4)$。C_2 即 $x = (k - y^2)/m$：这是一条关于 x 轴对称、开口向左（因为 $m > 0$）的抛物线，顶点（vertex）为 $(k/m, 0)$。当 $k > 0$ 时，它与 y 轴相交于 $(0, \pm\sqrt{k})$；当 $k < 0$ 时，它与 y 轴完全不相交。

C_1 上每一点都是 $(x, 4 - x^2)$，因此它在 C_2 上当且仅当

$$F(x) = x^4 - 8x^2 + mx + 16 - k = 0,$$

因为 $(4 - x^2)^2 = x^4 - 8x^2 + 16$。

> ? **想一想** 为甚么“相切”意味着 F 有重根？

在公共点上，C_1 的方向为 $(1, -2x)$，而 C_2（即 $mx + y^2$ 的等值曲线（level curve））的法向量（normal）为 $(m, 2y)$。当方向与法向量垂直时，两曲线相切：$m - 4xy = 0$，即 $m - 4x(4 - x^2) = 0$。而 $F'(x) = 4x^3 - 16x + m$ 正好就是这个式子。所以两曲线在 $x = a$

相切，当且仅当 $F(a) = F'(a) = 0$，即重根。

由 $F'(a) = 0$ 得 $m = 4a(4 - a^2)$。代入 $F(a) = 0$,

$$k = a^4 - 8a^2 + 4a^2(4 - a^2) + 16 = -3a^4 + 8a^2 + 16 = (4 - a^2)(3a^2 + 4),$$

因此

$$\frac{k}{m} = \frac{3a^2 + 4}{4a}.$$

由于 $m > 0$，我们需要 $0 < a < 2$ 或 $a < -2$。把二重因式除去,

$$F(x) = (x - a)^2\left(x^2 + 2ax + 3a^2 - 8\right),$$

这可以展开验证：x^3 项互相抵消，x^2 的系数是 $3a^2 - 8 - 4a^2 + a^2 = -8$，$x$ 的系数是 $2a^3 - 2a(3a^2 - 8) = 4a(4 - a^2) = m$，常数项是 $a^2(3a^2 - 8) = 16 - k$。这个二次因式的判别式（discriminant）为 $8(4 - a^2)$，所以只有当 $0 < a < 2$ 时才有其他交点，它们位于

$$x_\pm = -a \pm \sqrt{8 - 2a^2}.$$

现在可以按 a 把四种情况分类：

- $k < 0$ 当且仅当 $a^2 > 4$，故 $a < -2$。
- 当 $0 < a < 2$ 时，$k - 16 = a^2(8 - 3a^2)$，故 $k > 16$ 当且仅当 $a < \sqrt{8/3}$。
- 当 $0 < a < 2$ 时，$\frac{k}{m} < 2$ 当且仅当 $3a^2 - 8a + 4 < 0$，即 $(3a - 2)(a - 2) < 0$，即 $a > \frac{2}{3}$。

所以 (a) 对应 $a < -2$，(c) 对应 $0 < a < \frac{2}{3}$，(d) 对应 $\frac{2}{3} < a < \sqrt{8/3}$，(b) 对应 $\sqrt{8/3} < a < 2$。（没有任何相切的情形同时满足 $0 < k < 16$ 和 $k/m > 2$，这就是题目没有列出这种情况的原因。）

第 (i) 部分

在每种情况下，C_1 与坐标轴的交点都是 $(\pm 2, 0)$ 和 $(0, 4)$；C_2 与坐标轴的交点是 $(k/m, 0)$，当 $k > 0$ 时还有 $(0, \pm\sqrt{k})$。

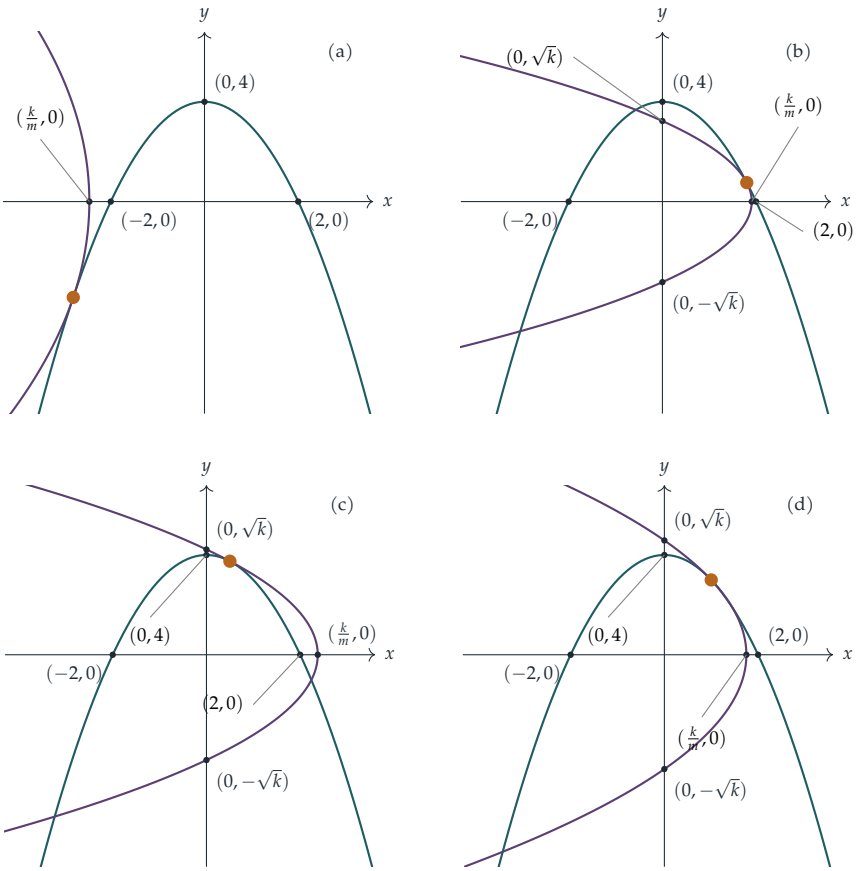
(a) 这里 $a < -2$，所以两曲线在第三象限（quadrant）的 $(a, 4 - a^2)$ 相切，而且在其他地方没有交点。顶点位于 $k/m = -(3a^2 + 4)/(4|a|)$，它小于 -2，因为 $(3|a| - 2)(|a| - 2) > 0$。所以 C_2 整条位于 $(-2, 0)$ 的左方，它的下半支贴靠在 C_1 的左支上。

(b) 现在 $\sqrt{8/3} < a < 2$，所以切点在第一象限，靠近 $(2, 0)$。顶点满足 $\sqrt{3} \leq k/m < 2$（由算术–几何平均不等式（AM–GM），$3a^2 + 4 \geq 4\sqrt{3}\,a$），而且 $\sqrt{k} < 4$，所以 C_2 的顶点和两个 y 截距都位于 C_1 之下。上半支从下方与 C_1 相切，然后在 x_+ 与它相交，其中 $-2 < x_+ < 0$（第二象限中的一点）。下半支在 $x_- < -2$ 与 C_1 相交，位于第三象限。

(c) 这里 $0 < a < \frac{2}{3}$。顶点在 $(2,0)$ 的右方，而且 $\sqrt{k} > 4$，所以 C_2 绕过 C_1 的顶部，在第一象限从外侧与它相切。此外 $x_+ > 2$，因为 $\sqrt{8-2a^2} > 2+a \iff (3a-2)(a+2) < 0$，而 $x_- < -2$。所以下半支在第四象限与 C_1 相交，在第三象限再相交一次。

(d) 这里 $\frac{2}{3} < a < \sqrt{8/3}$。顶点在 0 与 2 之间，位于 C_1 之下，但 $\sqrt{k} > 4$，所以上半支从 $(0,4)$ 上方经过。它在第一象限与 C_1 相切，并且也在第一象限于 $x_+ \in (0,2)$ 与它相交（因为 $x_+ > 0 \iff a^2 < \frac{8}{3}$）。下半支在 $x_- < -2$ 相交。

对 $0 < a < 2$，每种情况都有 $x_- < -2$，因为 $a + \sqrt{8-2a^2} > 2$（把 $\sqrt{8-2a^2} > 2-a$ 两边平方即可）。各草图分别取 $a = -2.8,\ 1.8,\ 0.5,\ 1$。C_1 为青色，C_2 为梅紫色，切点为琥珀色。



第 (ii) 部分

当 $m = 12$ 时，把 $y = 4 - x^2$ 代入 $12x = k - y^2$，得 $12x = k - 16 + 8x^2 - x^4$，即

$$x^4 - 8x^2 + 12x + 16 - k = 0,$$

正如所求。

在切点 $x=a$，这条四次方程有重根（见上面的方框），所以它的导数在该处也等于零：

$$4a^3-16a+12=0 \quad\Longrightarrow\quad a^3-4a+3=0.$$

由于 $a=1$ 是一个根，$a^3-4a+3=(a-1)(a^2+a-3)$，因此

$$a=1 \qquad \text{或} \qquad a=\frac{-1\pm\sqrt{13}}{2}.$$

为求 k，在四次方程中代入 $x=a$，并利用 $a^3=4a-3$ 化简，从而 $a^4=4a^2-3a$：

$$k=a^4-8a^2+12a+16=(4a^2-3a)-8a^2+12a+16=-4a^2+9a+16.$$

现在逐一看三种情况：

- $a=1$：$k=21$，所以 $k>16$ 且 $k/m=\frac{7}{4}<2$。这是情况 **(d)**。
- $a=\frac{1}{2}(-1+\sqrt{13})\approx 1.303$：这里 $a^2=3-a$，所以 $k=-4(3-a)+9a+16=13a+4=\frac{1}{2}(13\sqrt{13}-5)\approx 20.94$。它大于 16，因为 $13\sqrt{13}>37$（平方：$2197>1369$）；它小于 24，因为 $13\sqrt{13}<53$（$2197<2809$）。所以 $k>16$ 且 $k/m<2$，又是情况 **(d)**。
- $a=\frac{1}{2}(-1-\sqrt{13})\approx -2.303$：$k=13a+4=-\frac{1}{2}(5+13\sqrt{13})\approx -25.94<0$。这是情况 **(a)**。

所以只会出现草图 (a) 和 (d)。这与上面按 a 所作的分类一致：1 和 1.303 位于 $(\frac{2}{3},\sqrt{8/3})\approx(0.667,1.633)$ 之内，而 $-2.303<-2$。注意 $a^3-4a+3=0$ 其实就是 $m=4a(4-a^2)$ 取 $m=12$ 的情形。

两幅 (d) 图略有不同。对 $a=1$，其他交点为 $x=-1\pm\sqrt{6}$；在 $x\approx 1.45$ 的交点位于切点的*右方*，上半支从上方与 C_1 相切（如草图所示）。对 $a\approx 1.303$，交点在 $x\approx 0.84$，位于切点左方，上半支从下方相切。

验算

- 当 $m=12$、$k=21$ 时，四次式可因式分解为 $(x-1)^2(x^2+2x-5)$。展开得 $x^4-8x^2+12x-5$，正是 $16-k=-5$ 时的四次式。
- 在 $(1,3)$，C_1 的斜率为 -2；对 $12x=21-y^2$ 作隐函数微分（implicit differentiation），得 $\mathrm{d}y/\mathrm{d}x=-6/y=-2$。所以两曲线确实在该处相切。
- 数值上，对 $(0,2)$ 内及 $a<-2$ 的随机 a，公式 $m=4a(4-a^2)$ 和 $k=(4-a^2)(3a^2+4)$ 均令 $F(a)$ 和 $F'(a)$ 等于零，而各情况的分界正好落在 $a=\frac{2}{3}$ 和 $a=\sqrt{8/3}$。对 $a\approx 1.3028$，求根程序找到其他交点位于 $x\approx 0.8433$ 和 $x\approx -3.4488$，与 $x_\pm$ 的预测相符。

常见错误

- 要求四次方程只有一个根。“恰好相切一次”并不排除其他交点：在 $m = 12$ 的情况 (d) 中就有两个。
- 把 C_2 画成开口向右。当 $m > 0$ 时，$x = (k - y^2)/m$ 随 $|y|$ 增大而减小。
- 在 (a) 中，忘记 C_2 没有 y 截距，或把它的顶点放在 -2 与 0 之间。顶点必须在 -2 的左方。

STEP 之外

用切点 a 来标记各种构形，让我们初步窥见*判别式*和*包络线*（envelope）。使 F 有重根的 (m,k) 所组成的集合，就是曲线 $\big(4a(4-a^2),\,(4-a^2)(3a^2+4)\big)$；穿越这条曲线时，实交点的数目会改变二。这是平面曲线的代数几何（algebraic geometry）的起点，也是奇点理论（singularity theory，又称“突变理论”（catastrophe theory））的起点，后者用来研究解如何随参数变化。Siklos 的 *Advanced Problems in Mathematics*（《数学进阶问题》）收录了许多同类风格、STEP 式的曲线题目。

详细解答

STEP III 2012，第 4 题

纯数学 · 无穷级数求和

STEP III 2012，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/12-s3-q4

关键思路

一切都建基于两个标准级数（series）：$\sum x^n/n! = \mathrm{e}^x$ 及 $\sum x^n/n = -\ln(1-x)$。在第 (i) 部分，窍门是把每个关于 n 的多项式（polynomial）改写成能与 $n!$ 相消的形式，即使用 $n(n-1)$ 和 $n(n-1)(n-2)$ 这类乘积，而不是 n 的幂。在第 (ii) 部分，部分分式（partial fractions）把那个棘手的有理因子拆成 $1/(n+k)$ 形式的若干项，而每一项都是一个平移了的对数级数。

解答

第 (i) 部分

两个事实便足以完成所有工作。第一，$\sum_{n=0}^{\infty} 1/n! = \mathrm{e}$，所以

$$\sum_{n=1}^{\infty} \frac{1}{n!} = \mathrm{e} - 1.$$

第二，对任何整数 $k \geq 1$，当 $n < k$ 时，项 $n(n-1)\cdots(n-k+1)/n!$ 等于零；当 $n \geq k$ 时，它等于 $1/(n-k)!$。所以

$$\sum_{n=1}^{\infty} \frac{n(n-1)\cdots(n-k+1)}{n!} = \sum_{m=0}^{\infty} \frac{1}{m!} = \mathrm{e} \qquad (k \geq 1). \qquad (*)$$

这里所有级数都绝对收敛（absolutely converge）（每一个都是关于 n 的多项式除以 $n!$），因此我们可以随意把它们拆开再重新组合。

第一个和。 由于 $n+1=n+1$，利用 $(*)$ 并取 $k=1$，

$$\sum_{n=1}^{\infty}\frac{n+1}{n!}=\mathrm{e}+(\mathrm{e}-1)=2\mathrm{e}-1,$$

正如所求。

第二个和。 我们写成 $(n+1)^2=n^2+2n+1=n(n-1)+3n+1$。于是由 $(*)$ 取 $k=2$ 及 $k=1$，得

$$\sum_{n=1}^{\infty}\frac{(n+1)^2}{n!}=\mathrm{e}+3\mathrm{e}+(\mathrm{e}-1)=5\mathrm{e}-1,$$

正如所求。

第三个和。

? **想一想** 怎样把 $(2n-1)^3$ 化成下降乘积 $n(n-1)(n-2)$、$n(n-1)$、n?

展开得 $(2n-1)^3=8n^3-12n^2+6n-1$。接着我们利用

$$n^3=n(n-1)(n-2)+3n(n-1)+n,\qquad n^2=n(n-1)+n.$$

你可以展开右边来验证，或者代入 $n=0,1,2,3$ 检查。代入后，

$$\begin{aligned}(2n-1)^3&=8\big[n(n-1)(n-2)+3n(n-1)+n\big]-12\big[n(n-1)+n\big]+6n-1\\&=8\,n(n-1)(n-2)+12\,n(n-1)+2n-1.\end{aligned}$$

（快速验证：当 $n=1$ 时两边都是 1，当 $n=2$ 时两边都是 27。）现在由 $(*)$ 取 $k=3,2,1$，得

$$\sum_{n=1}^{\infty}\frac{(2n-1)^3}{n!}=8\mathrm{e}+12\mathrm{e}+2\mathrm{e}-(\mathrm{e}-1)=21\mathrm{e}+1.$$

第 (ii) 部分

部分分式。 分子与分母的次数相同，所以先做除法：

$$\frac{n^2+1}{(n+1)(n+2)}=1-\frac{3n+1}{(n+1)(n+2)}.$$

设 $\dfrac{3n+1}{(n+1)(n+2)}=\dfrac{A}{n+1}+\dfrac{B}{n+2}$，利用遮盖法（cover-up rule）得 $A=\dfrac{-2}{1}=-2$（取 $n=-1$）及 $B=\dfrac{-5}{-1}=5$（取 $n=-2$）。因此

$$\frac{n^2+1}{(n+1)(n+2)}=1+\frac{2}{n+1}-\frac{5}{n+2}.$$

三个部分。 我们令 $x=\frac{1}{2}$，并利用对于 $|x|<1$ 成立的 $-\ln(1-x)=\sum_{m=1}^{\infty} x^m/m$。下面三个级数各自收敛（各项缩小的速度至少与 2^{-n} 一样快），所以把和拆开是合理的。

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} = 2,$$

$$\sum_{n=0}^{\infty} \frac{x^n}{n+1} = \frac{1}{x}\sum_{m=1}^{\infty} \frac{x^m}{m} = \frac{-\ln(1-x)}{x} = 2\ln 2,$$

$$\sum_{n=0}^{\infty} \frac{x^n}{n+2} = \frac{1}{x^2}\sum_{m=2}^{\infty} \frac{x^m}{m} = \frac{-\ln(1-x)-x}{x^2} = 4\left(\ln 2-\tfrac{1}{2}\right) = 4\ln 2-2.$$

在第二行和第三行中，我们分别令 $m=n+1$ 及 $m=n+2$。第三行是完整的对数级数减去其 $m=1$ 项，即 x。

> **想一想** 为甚么第三部分需要额外的 $-x$?

平移 $m=n+2$ 令求和由 $m=2$ 开始，所以这个级数是缺了首项的 $-\ln(1-x)$。忘记这一点是这里最容易出错的地方。

合并。

$$\sum_{n=0}^{\infty} \frac{(n^2+1)2^{-n}}{(n+1)(n+2)} = 2+2(2\ln 2)-5(4\ln 2-2) = 12-16\ln 2.$$

验算

- 计算至 $n=60$ 的部分和（partial sums），(i) 中三个和分别得 4.436564、12.591409 及 58.083918。它们在机器精度内与 $2e-1$、$5e-1$ 及 $21e+1$ 吻合。
- 在 (ii) 中，$12-16\ln 2\approx 0.909645$，与部分和一致。首几项是 $\frac{1}{2}$、$\frac{1}{6}$、$\frac{5}{48}$、$\frac{1}{16}$，所以总和略小于 1 是合理的。
- (ii) 的答案必须是正数，而它的确是：$16\ln 2\approx 11.09<12$。

常见错误

- 使用 $\sum_{n=1}^{\infty} 1/n! = e$，忘记了缺少 $n=0$ 这一项。
- 把 $\sum n^2/n!$ 写成 $\sum n/(n-1)!$ 便停下来。你仍须拆开 $n=(n-1)+1$。
- 在 (ii) 中，为 $\sum x^n/(n+2)$ 平移指标时漏掉了 $-x$。

STEP 之外

下降乘积 $n(n-1)\cdots(n-k+1)$ 称为下降阶乘（falling factorials），而用它们来表示多项式就要用到第二类斯特林数（Stirling numbers of the second kind）。$(*)$ 中的和表明，平均值为 1 的泊松分布（Poisson distribution）的第 k 阶阶乘矩（factorial moment）等于 1。正因如此，例如 $\sum n^k/n!$ 等于 e 乘以第 k 个贝尔数（Bell number）。这里的重新排列之所以合理，是基于绝对收敛，这是大学一年级分析（analysis）的一个重要主题：参见 Abbott，*Understanding Analysis*（《理解分析》）。至于概率方面，参见 Blitzstein 与 Hwang，*Introduction to Probability*（《概率导论》）。

详细解答

STEP III 2012，第 5 题

纯数学 · 圆锥曲线上的有理点

> **STEP III 2012，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/12-s3-q5

关键思路

(i)(b) 中的恒等式（identity）其实是一个经过伪装的旋转（rotation）：点 $(\cos\theta + \sqrt{m}\sin\theta,\ \sin\theta - \sqrt{m}\cos\theta)$ 正是把点 $(1, -\sqrt{m})$ 旋转角度 θ 后所得的点，因此它仍然位于半径为 $\sqrt{1+m}$ 的圆上。若 $\cos\theta$ 和 $\sin\theta$ 都是有理数（例如 (i)(a) 中得到的 $\frac{3}{5}$ 和 $\frac{4}{5}$），这个旋转就会把一个“好”点变成另一个“好”点。第 (ii) 部分的所有内容都是同一个技巧，只是从一个合适的整数点出发；至于双曲线（hyperbola），我们则把旋转换成它的双曲版本。

解答

第 (i) 部分

(a) 点 $(1,0)$ 是 $x^2+y^2=1$ 上的一个整数有理点（integer rational point）。要找非整数的有理点，我们利用 3,4,5 三角形：$\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2 = \frac{9+16}{25} = 1$，所以 $\left(\frac{3}{5}, \frac{4}{5}\right)$ 符合要求，而且两个坐标都不是整数。

(b) 点 $(1,1)$ 是 $x^2+y^2=2$ 上的一个整数有理点。

展开后，交叉项（cross terms）为 $+2\sqrt{m}\sin\theta\cos\theta$ 和 $-2\sqrt{m}\sin\theta\cos\theta$，两者互相抵消，所以

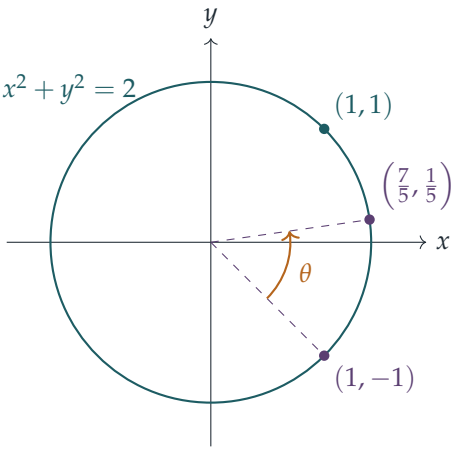
$$\begin{aligned}(\cos\theta + \sqrt{m}\sin\theta)^2 + (\sin\theta - \sqrt{m}\cos\theta)^2 &= \cos^2\theta + m\sin^2\theta + \sin^2\theta + m\cos^2\theta \\ &= (1+m)(\cos^2\theta + \sin^2\theta) = 1 + m.\end{aligned}$$

取 $m=1$，使右方等于 2，并选取 θ 使 $\cos\theta = \frac{3}{5}$、$\sin\theta = \frac{4}{5}$（这是可行的，因为 $\left(\frac{3}{5}, \frac{4}{5}\right)$

位于单位圆（unit circle）上）。于是

$$x = \tfrac{3}{5} + \tfrac{4}{5} = \tfrac{7}{5}, \qquad y = \tfrac{4}{5} - \tfrac{3}{5} = \tfrac{1}{5}$$

满足 $x^2 + y^2 = 2$；事实上 $\frac{49+1}{25} = 2$。所以 $\left(\frac{7}{5}, \frac{1}{5}\right)$ 是 $x^2 + y^2 = 2$ 上的一个非整数有理点。



上图显示了个中过程：把 $(1, -1)$ 沿逆时针方向旋转 θ（其中 $\cos\theta = \frac{3}{5}$、$\sin\theta = \frac{4}{5}$），便落在 $\left(\frac{7}{5}, \frac{1}{5}\right)$ 上。

第 (ii) 部分

(a) 由于 $1^2 + (\sqrt{2})^2 = 3$，点 $(1, \sqrt{2})$，即 $p = 1$、$q = 0$、$r = 0$、$s = 1$，是 $x^2 + y^2 = 3$ 上的一个整数 2-有理点。

要找非整数的点，就用 (i)(b) 的恒等式，取 $m = 2$ 使 $1 + m = 3$，并再次取 $\cos\theta = \frac{3}{5}$、$\sin\theta = \frac{4}{5}$：

$$x = \tfrac{3}{5} + \tfrac{4}{5}\sqrt{2}, \qquad y = \tfrac{4}{5} - \tfrac{3}{5}\sqrt{2}.$$

这里 $p = \frac{3}{5}$、$q = \frac{4}{5}$、$r = \frac{4}{5}$、$s = -\frac{3}{5}$，全都不是整数，而由恒等式可知 $x^2 + y^2 = 3$。直接验证：$x^2 = \frac{9+32+24\sqrt{2}}{25}$，$y^2 = \frac{16+18-24\sqrt{2}}{25}$，两者相加得 $\frac{75}{25} = 3$。

> ? **想一想** 恒等式只给出 $1 + m$，而 m 又必须与 $\sqrt{2}$ 有关，那么我们怎样才能得到 11?

把起点放大。恒等式源自旋转 $(1, -\sqrt{m})$；我们同样可以旋转任何点 $(a, -b\sqrt{2})$。交叉项同样会抵消，得到

$$(a\cos\theta + b\sqrt{2}\sin\theta)^2 + (a\sin\theta - b\sqrt{2}\cos\theta)^2 = (a^2 + 2b^2)(\cos^2\theta + \sin^2\theta) = a^2 + 2b^2.$$

所以我们需要满足 $a^2+2b^2=11$ 的整数，而 $a=3$、$b=1$ 正好可行（$9+2=11$）。换句话说，我们从 $x^2+y^2=11$ 上的整数 2-有理点 $(3,\sqrt{2})$ 出发。

(b) 取 $a=3$、$b=1$、$\cos\theta=\frac{3}{5}$、$\sin\theta=\frac{4}{5}$：

$$x=\tfrac{9}{5}+\tfrac{4}{5}\sqrt{2},\qquad y=\tfrac{12}{5}-\tfrac{3}{5}\sqrt{2}.$$

所以 $p=\frac{9}{5}$、$q=\frac{4}{5}$、$r=\frac{12}{5}$、$s=-\frac{3}{5}$，全都不是整数，且 $x^2+y^2=11$。直接验证：$x^2=\frac{81+32+72\sqrt{2}}{25}$，$y^2=\frac{144+18-72\sqrt{2}}{25}$，两者相加得 $\frac{275}{25}=11$。

(c) 对于双曲线，我们需要一个保持 x^2-y^2 而非 x^2+y^2 不变的“旋转”。若 $c^2-s^2=1$（可以想像 $c=\cosh t$、$s=\sinh t$），则

$$\begin{aligned}(ac+b\sqrt{2}\,s)^2-(as+b\sqrt{2}\,c)^2&=a^2c^2+2b^2s^2-a^2s^2-2b^2c^2\\&=(a^2-2b^2)(c^2-s^2)=a^2-2b^2,\end{aligned}$$

因为交叉项 $2\sqrt{2}\,abcs$ 再次互相抵消。我们需要 $a^2-2b^2=7$：取 $a=3$、$b=1$（所以 $(3,\sqrt{2})$ 是双曲线上的一个整数 2-有理点）。

接着我们需要有理数 c、s 使 $c^2-s^2=(c-s)(c+s)=1$。选取 $c+s=2$ 及 $c-s=\frac{1}{2}$，得 $c=\frac{5}{4}$、$s=\frac{3}{4}$。于是

$$x=\tfrac{15}{4}+\tfrac{3}{4}\sqrt{2},\qquad y=\tfrac{9}{4}+\tfrac{5}{4}\sqrt{2},$$

其中 $p=\frac{15}{4}$、$q=\frac{3}{4}$、$r=\frac{9}{4}$、$s=\frac{5}{4}$，全都不是整数。直接验证：$x^2=\frac{225+18+90\sqrt{2}}{16}$，$y^2=\frac{81+50+90\sqrt{2}}{16}$，所以 $x^2-y^2=\frac{112}{16}=7$。这就是 $x^2-y^2=7$ 上的一个非整数 2-有理点。

验算

每一个点都可以像上文那样用分数准确验证，也可以用数值验证：对于 (ii)(c)，$x\approx 4.81066$、$y\approx 4.01777$，得 $x^2-y^2\approx 7.0000$；对于 (ii)(b)，$x\approx 2.93137$、$y\approx 1.55147$，得 $x^2+y^2\approx 11.0000$。最后务必逐一检查 p,q,r,s，确认没有一个是整数。

常见错误

- 忘记 0 是整数：像 $\left(\frac{7}{5}\sqrt{2},\ldots\right)$ 这样的点有 $p=0$，所以它不是非整数 2-有理点。
- 在 (c) 中，一个看似无害的选择也可能失败：$c=\frac{5}{3}$、$s=\frac{4}{3}$ 同样满足 $c^2-s^2=1$，但会得出 $x=5+\frac{4}{3}\sqrt{2}$，其中 $p=5$ 是整数。每个系数都要检查。

STEP 之外

形如 $p+q\sqrt{2}$ 的数构成一个域（field）$\mathbb{Q}(\sqrt{2})$，而这道题其实是在该域上寻找圆锥曲线（conic）上的点。把已知点旋转一个有理角度的技巧，是一套优美理论的起点：只要一条圆锥曲线上有一个有理点，经过该点的直线便能把其余所有有理点参数化（parametrise）；而 $x^2-2y^2=1$（佩尔方程（Pell's equation））上的有理点，在 (c) 所用的那种双曲“旋转”之下正好构成一个群（group）。Davenport 的 *The Higher Arithmetic*（《高等算术》）讲述了佩尔方程与平方和，而 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）则介绍了像 $\mathbb{Q}(\sqrt{2})$ 这样的域。

详细解答

STEP III 2012，第 6 题

纯数学 · 复数根的轨迹

STEP III 2012，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/12-s3-q6

关键思路

一个复数（complex number）为零，当且仅当其实部（real part）与虚部（imaginary part）同时为零。因此，关于 $z = x + \mathrm{i}y$ 的一条复数方程，会变成关于 x、y 和 p 的两条实数方程。虚部方程总能因式分解（factorise）出因子 y，于是每个轨迹（locus）都分成“实根”和“非实根”两部分。在每种情况下，我们先消去（eliminate）p，得到 (x, y) 平面上的一条曲线，然后检验该曲线上哪些点确实来自某个实数 p。

解答

第一条方程 $z^2 + pz + 1 = 0$

设 $z = x + \mathrm{i}y$，其中 x, y 为实数。则

$$z^2 + pz + 1 = (x^2 - y^2 + px + 1) + \mathrm{i}\,(2xy + py).$$

由于 p 是实数，两个括号分别是实部和虚部，两者都必须为零：

$$x^2 - y^2 + px + 1 = 0, \qquad (2x + p)y = 0,$$

此即所求。

由第二条方程，得 $p = -2x$ 或 $y = 0$。若 $y = 0$，第一条方程变为 $x^2 + px + 1 = 0$。此时 $x \neq 0$，因为 $x = 0$ 会得出 $1 = 0$，所以可以除以 x，得 $p = -(x^2 + 1)/x$。因此，要么 $p = -2x$，要么 $p = -(x^2 + 1)/x$ 且 $x \neq 0$（而在第二种情况下 $y = 0$）。

根的轨迹。情况 $p = -2x$。代入实部方程，得 $x^2 - y^2 - 2x^2 + 1 = 0$，即

$$x^2 + y^2 = 1.$$

反过来（converse），在单位圆（unit circle）上任取一点 $x+\mathrm{i}y$，并设 $p=-2x$，这是实数。则 $(2x+p)y=0$，且 $x^2-y^2+px+1=1-x^2-y^2=0$，所以该点是一个根。因此整个单位圆都在轨迹之内。

情况 $y=0$，$p=-(x^2+1)/x$。对每个实数 $x\neq 0$，这个 p 都是实数，而且 x 是 $x^2+px+1=0$ 的根。点 $x=0$ 永远不是根，因为 $0^2+p\cdot 0+1=1$。

所以根的轨迹是去掉原点的实轴，连同单位圆 $|z|=1$。（点 ± 1 同时位于两者之上。）

第二条方程 $pz^2+z+1=0$

设 $z=x+\mathrm{i}y$，由实部和虚部得

$$p(x^2-y^2)+x+1=0,\qquad (2px+1)y=0.$$

原点永远不是根，因为 $z=0$ 会得出 $1=0$。

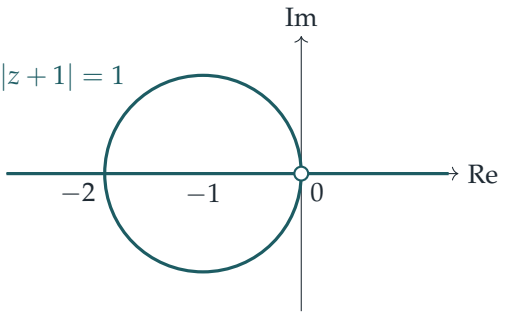
情况 $y=0$。此时 $px^2+x+1=0$ 且 $x\neq 0$，所以 $p=-(x+1)/x^2$。对每个实数 $x\neq 0$，这都是实数，所以实轴上除原点外的每一点都在轨迹之内。（当 $x=-1$ 时得 $p=0$，方程变成一次方程 $z+1=0$，其根的确是 -1。）

情况 $2px+1=0$。此时 $x\neq 0$ 且 $p=-1/(2x)$。代入实部方程并乘以 $-2x$：

$$(x^2-y^2)-2x^2-2x=0 \iff x^2+y^2+2x=0 \iff (x+1)^2+y^2=1.$$

反过来，对这个圆上除原点以外的任何一点，都有 $x\neq 0$，所以 $p=-1/(2x)$ 是实数，而把步骤倒推回去，便可证明两条方程都成立。

所以根的轨迹是实轴连同圆 $|z+1|=1$（圆心 -1，半径 1），但要去掉同时位于两者之上的原点。



空心圆表示缺失的点 $z=0$。

想一想 为甚么这个轨迹看起来像是把第一个轨迹“里外翻转”了？

若 z 是 $pz^2+z+1=0$ 的根，则 $w=1/z$ 满足 $w^2+w+p=0$。映射 $z\mapsto 1/z$ 把实轴（不含 0）映到自身，又把直线 $\operatorname{Re} w=-\frac{1}{2}$（即 $w^2+w+p=0$ 的非实根所在之处）映到通过原点的圆 $|z+1|=1$。原点之所以缺失，是因为它对应于 $w=\infty$。

第三条方程 $pz^2+p^2z+2=0$

若 $p=0$，方程变为 $2=0$，没有根，所以下文始终有 $p\neq 0$。由实部和虚部得

$$p(x^2-y^2)+p^2x+2=0,\qquad p(2x+p)y=0,$$

由于 $p\neq 0$，第二条化为 $(2x+p)y=0$。

情况 $p=-2x$。由于 $p\neq 0$，故 $x\neq 0$。代入得

$$-2x(x^2-y^2)+4x^3+2=0\quad\Longleftrightarrow\quad x(x^2+y^2)=-1.$$

反过来，这条曲线上任何一点都有 $x\neq 0$，所以 $p=-2x$ 是非零实数，而把步骤倒推回去，便可证明该点是一个根。由于 $x^2+y^2>0$，曲线上 $x<0$。把它写成

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

我们需要 $-1/x\geq x^2$；乘以 $x<0$ 会令不等式（inequality）反向，得 $-1\leq x^3$，所以 $-1\leq x<0$。曲线通过 -1（该处 $y=0$），关于实轴对称，而且当 $x\to 0^-$ 时 $y^2\to\infty$，所以虚轴是一条渐近线（asymptote）。此外，对 $x<0$ 有 $\mathrm{d}(y^2)/\mathrm{d}x=x^{-2}-2x>0$，所以当 x 由 -1 增至 0 时，$|y|$ 一直递增。以极坐标形式（polar form）表示，曲线为 $r^3\cos\theta=-1$。

情况 $y=0$。此时，实数 x 在轨迹之内，当且仅当存在实数 p 使 $xp^2+x^2p+2=0$。当 $x=0$ 时，这就是 $2=0$，所以 0 不在轨迹之内。当 $x\neq 0$ 时，这是关于 p 的二次方程，它有实数解当且仅当其判别式（discriminant）非负：

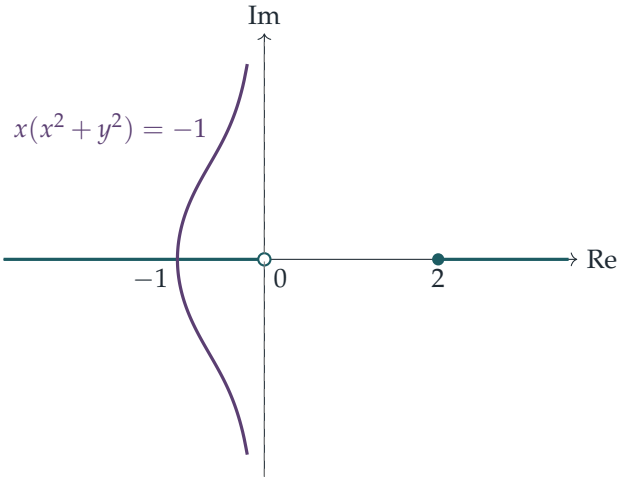
$$x^4-8x=x(x^3-8)\geq 0.$$

当 $x<0$（两个因子皆为负）或 $x\geq 2$（两个因子皆非负）时成立，而当 $0<x<2$ 时不成立。以这种方法求得的任何实数 p 必然非零，因为 $p=0$ 会得出 $2=0$。例如 $x=2$ 需要 $2p^2+4p+2=0$，故 $p=-1$，而确实有 $-z^2+z+2=-(z-2)(z+1)$。

? **想一想** 为甚么这里要解出 p，而不是解出 x?

我们想知道哪些点 x 可以达到，所以 x 是已知的，而 p 才是未知数。这条方程是关于 p 的二次方程，而存在实根恰好就是关于 x 的判别式条件。

所以根的轨迹是负实轴、实轴上的射线 $x\geq 2$，以及曲线 $x(x^2+y^2)=-1$；这条曲线从 -1 出发，向上下两个方向伸展，趋近虚轴。



曲线只在 -1 处与实轴相交，该处的切线是铅垂的；当 $|y|\to\infty$ 时，曲线趋近虚轴。

验算

- 对于 $z^2+pz+1=0$，两根之积为 1，所以非实根是一对共轭（conjugate）复数 $z,\bar{z}$，且 $|z|^2=1$：再次得到单位圆。
- 在第三条方程中，$p=-1$ 给出根 2 和 -1，分别是射线的端点和曲线的顶点。取 $p=1$ 得 $z^2+z+2=0$，其根为 $-\frac{1}{2}\pm\frac{\sqrt{7}}{2}\mathrm{i}$，而 $x(x^2+y^2)=-\frac{1}{2}\cdot 2=-1$，正如所料。
- 在数值上，我们以 10^5 个随机 p 值求解全部三条方程，确认每个根都位于所述的轨迹上。我们亦确认了各轨迹上的随机点都能以所述的 p 达到；而对 p 作精细扫描后，并未发现第三条方程在 $(0,2)$ 内有任何实根，最小的正实根是 2.0。

常见错误

- 忘记检验反方向：消去 p 只得到一条包含轨迹的曲线，你仍须证明每一点都能由某个实数 p 达到。在第三条方程中，$(0,2)$ 内的实数点虽然满足 $y=0$，却永远不能达到。
- 除以 x 或 p 时没有说明它为何非零。缺失的原点正是由此而来。
- 草率处理 $p=0$：在第二条方程中它给出真正的根 -1，但在第三条方程中它完全没有根。

STEP 之外

根轨迹（root locus）是控制工程（control engineering）中的标准工具：当增益（gain）参数改变时，系统的极点（poles）沿着曲线移动，而系统的稳定性（stability）取决于这些极点会否越过进入右半平面（right half-plane）。这道题亦暗示了映射 $z \mapsto 1/z$ 的几何性质：它把直线和圆映成直线和圆。这正是莫比乌斯变换（Möbius transformations）理论的开端，Needham 的 *Visual Complex Analysis*（《直观复分析》）对此有精彩的图解。

详细解答

STEP III 2012，第 7 题

纯数学 · 耦合微分方程

> **STEP III 2012，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/12-s3-q7

关键思路

两条耦合的一阶方程（coupled first-order equations），只要消去一个变量，就会变成一条二阶方程（second-order equation）：先对 $\dot{y}$ 的方程求导，再利用第一条方程本身，把 z 写成 y 和 $\dot{y}$ 的式子。之后，每一部分都只是选取常数 A 和 B 而已。在第 (iii) 部分，无穷和是两个几何级数（geometric series），把结果与第 (ii) 部分对比，就能确定 c。从物理角度看，z_2 是每隔一个单位时间注射同一剂量后所达到的稳态（steady state）。

解答

通解

对 $\dot{y} = -2y + 2z$ 求导，再以第二条方程代入 $\dot{z}$，得

$$\ddot{y} = -2\dot{y} + 2\dot{z} = -2\dot{y} + 2(-\dot{y} - 3z) = -4\dot{y} - 6z.$$

由第一条方程可把 z 用 y 表示：

$$z = y + \tfrac{1}{2}\dot{y}, \qquad \text{所以} \qquad 6z = 6y + 3\dot{y}.$$

代入后得

$$\ddot{y} + 7\dot{y} + 6y = 0.$$

辅助方程（auxiliary equation）$m^2 + 7m + 6 = (m+1)(m+6) = 0$ 的根为 $m = -1$ 和 $m = -6$，因此

$$y = Ae^{-t} + Be^{-6t}$$

其中 A 和 B 为任意常数。然后由 $z = y + \frac{1}{2}\dot{y}$ 得出 z：

$$z = Ae^{-t}\left(1 - \tfrac{1}{2}\right) + Be^{-6t}(1 - 3) = \tfrac{1}{2}Ae^{-t} - 2Be^{-6t},$$

此即所求。

> **想一想** 我们用了两条方程来得出 y 的方程，但求 z 时只用了第一条。我们找到的这一对函数，真的满足第二条方程吗?

确实满足。由 $z = y + \frac{1}{2}\dot{y}$，有 $\dot{z} = \dot{y} + \frac{1}{2}\ddot{y} = \dot{y} + \frac{1}{2}(-7\dot{y} - 6y) = -\frac{5}{2}\dot{y} - 3y$。另一方面，$-\dot{y} - 3z = -\dot{y} - 3y - \frac{3}{2}\dot{y} = -\frac{5}{2}\dot{y} - 3y$，两者相同。所以方程组的每个解都具有这个形式，而每对具有这个形式的函数都是解。

第 (i) 部分

在 $t = 0$ 时，$y(0) = A + B$，$z(0) = \frac{1}{2}A - 2B$。由条件得

$$\tfrac{1}{2}A - 2B = 0, \qquad A + B = 5,$$

所以 $A = 4B$，于是 $5B = 5$。因此 $B = 1$，$A = 4$:

$$y = 4\mathrm{e}^{-t} + \mathrm{e}^{-6t}, \qquad z_1(t) = 2\mathrm{e}^{-t} - 2\mathrm{e}^{-6t}.$$

药物起初在血液中，脑内的药量由零开始上升，然后随着被吸收而衰减。

第 (ii) 部分

设 $a = \frac{1}{2}A$ 及 $b = 2B$，使 $z = a\mathrm{e}^{-t} - b\mathrm{e}^{-6t}$。条件为

$$a - b = c, \qquad a\mathrm{e}^{-1} - b\mathrm{e}^{-6} = c.$$

以第一式减第二式，得 $a(1 - \mathrm{e}^{-1}) = b(1 - \mathrm{e}^{-6})$。把这个共同的值记作 k，于是

$$a = \frac{k}{1 - \mathrm{e}^{-1}}, \qquad b = \frac{k}{1 - \mathrm{e}^{-6}}, \qquad \text{及} \qquad c = a - b = kC,$$

其中

$$C = \frac{1}{1 - \mathrm{e}^{-1}} - \frac{1}{1 - \mathrm{e}^{-6}}.$$

由于 $\mathrm{e}^{-1} > \mathrm{e}^{-6}$，有 $1 - \mathrm{e}^{-1} < 1 - \mathrm{e}^{-6}$，所以 $C > 0$。特别地，$C \neq 0$，因此可以除以它: $k = c/C$。（等价地说，这两条线性方程的行列式（determinant）为 $\mathrm{e}^{-1} - \mathrm{e}^{-6} \neq 0$，所以解是唯一的。）因此

$$z_2(t) = \frac{c}{C}\left(\frac{\mathrm{e}^{-t}}{1 - \mathrm{e}^{-1}} - \frac{\mathrm{e}^{-6t}}{1 - \mathrm{e}^{-6}}\right),$$
$$y_2(t) = \frac{c}{C}\left(\frac{2\mathrm{e}^{-t}}{1 - \mathrm{e}^{-1}} + \frac{\mathrm{e}^{-6t}}{2(1 - \mathrm{e}^{-6})}\right),$$

其中 y_2 由 $A = 2a$ 及 $B = \frac{1}{2}b$ 得出。

第 (iii) 部分

设 $m = -n$，则 m 取遍 $0, 1, 2, \ldots$，且 $t - n = t + m$。利用第 (i) 部分，

$$\sum_{n=-\infty}^{0} z_1(t-n) = \sum_{m=0}^{\infty} \left(2\mathrm{e}^{-t}(\mathrm{e}^{-1})^m - 2\mathrm{e}^{-6t}(\mathrm{e}^{-6})^m\right).$$

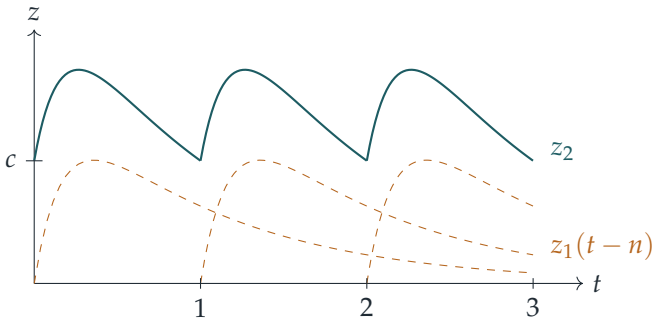
两部分都是几何级数，公比分别为 e^{-1} 和 e^{-6}，都严格介于 0 与 1 之间，所以两者各自收敛（converge），可以分开求和：

$$\sum_{n=-\infty}^{0} z_1(t-n) = \frac{2\mathrm{e}^{-t}}{1-\mathrm{e}^{-1}} - \frac{2\mathrm{e}^{-6t}}{1-\mathrm{e}^{-6}}.$$

这正好是 $z_2(t)$ 中括号内的式子乘以 2。所以该恒等式在 $0 \le t \le 1$ 上成立，当且仅当 $c/C = 2$，即

$$c = 2C = \frac{2}{1-\mathrm{e}^{-1}} - \frac{2}{1-\mathrm{e}^{-6}} \approx 1.159.$$

（若 c 取其他值，两边相差同一个非零函数的非零倍数，所以这个值是唯一的。）



这幅图展示了其中的含义。在时间 n 向血液注射剂量 5，会产生脑内曲线 $z_1(t-n)$（虚线）。若从遥远的过去开始，每个整数时间都注射一剂，脑内的药量就是所有这些响应之和。这个和每隔一个单位时间重复一次（实线），而在 $[0,1]$ 上，它就是 c 取这个值时的 z_2。

验算

- 在第 (i) 部分，$z_1(0) = 2 - 2 = 0$，$y(0) = 4 + 1 = 5$。
- 在第 (iii) 部分，$z_2(0)$ 和 $z_2(1)$ 都应等于 c。在求和式中，$t = 1$ 的级数就是 $t = 0$ 的级数去掉首项，而该项为 $z_1(0) = 0$。这样不用任何代数运算就能确认 $z(0) = z(1)$。
- 当 $c = 2C$ 时，$y_2(0) - y_2(1) = 4 + 1 = 5$：每个周期内血液恰好损失一剂，而下一次注射正好补回。

- 数值上，$c \approx 1.15898$。在 $t = 0.37$ 处把 $z_1(t+m)$ 的 200 项相加，得 1.967693…，与 $z_2(0.37)$ 吻合至小数点后 12 位。

常见错误

- 忘记 z 已经通过 $z = y + \frac{1}{2}\dot{y}$ 由 y 决定，反而另外解一条 z 的方程，并引入新的、独立的常数。
- 在第 (iii) 部分弄错平移的方向：$n \le 0$ 意味着 $t - n \ge t$，所以只有过去的剂量才有贡献。

STEP 之外

这是一个线性房室模型（linear compartment model），可写成 $\dot{\mathbf{x}} = M\mathbf{x}$，其中 M 是一个 2×2 矩阵。衰减率 1 和 6 是 M 的特征值（eigenvalue）的相反数，这正是你会在大学一年级课程中遇到的、以矩阵指数（matrix exponential）看待线性系统的观点。第 (iii) 部分是叠加原理（superposition）的一个简单例子：对重复脉冲的响应，是同一个脉冲响应（impulse response）的多个平移副本之和。这个想法引向格林函数（Green's function）和卷积（convolution）。Boyce 和 DiPrima 的 *Elementary Differential Equations*（《初等微分方程》）涵盖了线性方程组和脉冲响应。

详细解答

STEP III 2012，第 8 题

纯数学 · 斐波那契裂项级数

> **STEP III 2012，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/12-s3-q8

关键思路

第 (ii) 部分的式子每当 n 增加一，就会改变正负号，而第 (i) 部分正是在提示我们去寻找这种规律。一旦知道它在 n 为奇数时的值，第 (iii) 部分的反正切恒等式（identity）其实只是 $\tan(A-B)$ 公式换了个样子。把这条恒等式重新整理，就能把级数（series）的每一项写成两个反正切的*差*，于是求和便可以伸缩相消（telescope）。

解答

首几项为

$$F_0, F_1, \ldots, F_8 = 0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21.$$

第 (i) 部分

利用这些数值，

$$F_0F_3 - F_1F_2 = 0\cdot 2 - 1\cdot 1 = -1, \qquad F_2F_5 - F_3F_4 = 1\cdot 5 - 2\cdot 3 = -1,$$

所以两个式子相等，正如所求。

第 (ii) 部分

设 $G_n = F_nF_{n+3} - F_{n+1}F_{n+2}$。第 (i) 部分说明 $G_0 = G_2$，这启发我们比较 G_{n+1} 与 G_n。把递推关系（recurrence）写成 $F_{n+4} = F_{n+3} + F_{n+2}$ 的形式，得

$$\begin{aligned} G_{n+1} &= F_{n+1}F_{n+4} - F_{n+2}F_{n+3} \\ &= F_{n+1}F_{n+3} + F_{n+1}F_{n+2} - F_{n+2}F_{n+3} \\ &= F_{n+1}F_{n+2} - F_{n+3}\left(F_{n+2} - F_{n+1}\right) \\ &= F_{n+1}F_{n+2} - F_{n+3}F_n \\ &= -G_n, \end{aligned}$$

其中第四行再次用了递推关系 $F_{n+2} - F_{n+1} = F_n$。

所以 G_n 每一步都改变正负号。由于第 (i) 部分给出 $G_0 = -1$，可得（对 n 用归纳法（induction））$G_n = (-1)^{n+1}$。两种情况为

$$F_nF_{n+3} - F_{n+1}F_{n+2} = \begin{cases} -1 & \text{若 } n \text{ 为偶数,} \\ 1 & \text{若 } n \text{ 为奇数。} \end{cases}$$

例如，$n = 1$ 给出 $1 \cdot 3 - 1 \cdot 2 = 1$，而 $n = 4$ 给出 $3 \cdot 13 - 5 \cdot 8 = -1$。

第 (iii) 部分

固定 $r \geq 1$，并设

$$\alpha = \arctan \frac{1}{F_{2r}}, \qquad \beta = \arctan \frac{1}{F_{2r+1}}.$$

所有 $k \geq 1$ 的 F_k 都是正整数，所以这两个量都有定义。由复角公式（compound-angle formula），

$$\tan(\alpha - \beta) = \frac{\dfrac{1}{F_{2r}} - \dfrac{1}{F_{2r+1}}}{1 + \dfrac{1}{F_{2r}F_{2r+1}}} = \frac{F_{2r+1} - F_{2r}}{F_{2r}F_{2r+1} + 1} = \frac{F_{2r-1}}{F_{2r}F_{2r+1} + 1},$$

这里用了 $F_{2r+1} - F_{2r} = F_{2r-1}$。现在以奇数值 $n = 2r - 1$ 应用第 (ii) 部分：

$$F_{2r-1}F_{2r+2} - F_{2r}F_{2r+1} = 1, \qquad \text{所以} \qquad F_{2r}F_{2r+1} + 1 = F_{2r-1}F_{2r+2}.$$

由于 $F_{2r-1} \geq 1$，它不为零，可以约去：

$$\tan(\alpha - \beta) = \frac{F_{2r-1}}{F_{2r-1}F_{2r+2}} = \frac{1}{F_{2r+2}}.$$

想一想 正切相等不一定代表角相等。为甚么我们可以断定 $\alpha - \beta = \arctan(1/F_{2r+2})$?

由于 tan 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上是一一对应的（one-to-one），只需验证两个角都在这个区间内。这里 $0 < \beta \le \alpha \le \frac{\pi}{4}$，因为 $1 \le F_{2r} \le F_{2r+1}$，所以 $0 \le \alpha - \beta < \frac{\pi}{4}$。另外，按定义 $\arctan(1/F_{2r+2})$ 位于 $(0, \frac{\pi}{2})$ 内。因此

$$\arctan\frac{1}{F_{2r}} = \arctan\frac{1}{F_{2r+1}} + \arctan\frac{1}{F_{2r+2}} \qquad (r = 1, 2, 3, \ldots),$$

这就是所求的恒等式。

求和。把恒等式重新整理，得

$$\arctan\frac{1}{F_{2r+1}} = \arctan\frac{1}{F_{2r}} - \arctan\frac{1}{F_{2r+2}},$$

所以部分和（partial sums）伸缩相消：

$$\begin{aligned}\sum_{r=1}^{N}\arctan\frac{1}{F_{2r+1}} &= \left(\arctan\frac{1}{F_2} - \arctan\frac{1}{F_4}\right) + \left(\arctan\frac{1}{F_4} - \arctan\frac{1}{F_6}\right) + \cdots \\ &\quad + \left(\arctan\frac{1}{F_{2N}} - \arctan\frac{1}{F_{2N+2}}\right) \\ &= \arctan\frac{1}{F_2} - \arctan\frac{1}{F_{2N+2}} = \frac{\pi}{4} - \arctan\frac{1}{F_{2N+2}},\end{aligned}$$

因为 $F_2 = 1$。当 $n \ge 2$ 时，有 $F_{n+1} = F_n + F_{n-1} \ge F_n + 1$，所以 $F_n \to \infty$。于是 $1/F_{2N+2} \to 0$，而由于 arctan 连续（continuous）且 $\arctan 0 = 0$，令 $N \to \infty$ 便得

$$\sum_{r=1}^{\infty}\arctan\frac{1}{F_{2r+1}} = \frac{\pi}{4}.$$

验算

- $r = 1$：恒等式说 $\arctan 1 = \arctan\frac{1}{2} + \arctan\frac{1}{3}$。这是经典的 $\frac{\pi}{4}$ 恒等式，因为右方的 tan 为 $\left(\frac{1}{2} + \frac{1}{3}\right)/\left(1 - \frac{1}{6}\right) = 1$。
- 首三项 $\arctan\frac{1}{2} + \arctan\frac{1}{5} + \arctan\frac{1}{13} \approx 0.4636 + 0.1974 + 0.0768 = 0.7378$，与 $\frac{\pi}{4} - \arctan\frac{1}{21} \approx 0.7378$ 吻合。
- 数值上，级数的 89 项给出 $0.785398163\ldots$，在机器精度内与 $\frac{\pi}{4}$ 一致。第 (ii) 部分的公式对每个 $n < 150$ 亦准确成立。

常见错误

- 在第 (iii) 部分用了第 (ii) 部分的偶数情况。那里的下标是 $2r-1$, $2r$, $2r+1$, $2r+2$，所以 $n=2r-1$ 是奇数，值为 $+1$。
- 由正切相等直接推出角相等，而没有验证两个角都位于 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 内。
- 忘记伸缩求和后的首项：它是 $\arctan(1/F_2)=\frac{\pi}{4}$，而不是 $\arctan(1/F_1)$。

STEP 之外

第 (ii) 部分与卡西尼（Cassini）恒等式 $F_{n-1}F_{n+1}-F_n^2=(-1)^n$ 是近亲。两者最好通过矩阵（matrix）$\begin{pmatrix}1&1\\1&0\end{pmatrix}$ 来理解：它的 n 次幂的元素正是斐波那契数（Fibonacci numbers），所以取行列式（determinant）便立即得到卡西尼恒等式。伸缩级数（telescoping series），以及由部分和过渡到极限时所需的谨慎，是大学第一个分析（analysis）课程的必备内容。Abbott 的 *Understanding Analysis*（《理解分析》）是阅读这些内容的好起点，而 Siklos 的 *Advanced Problems in Mathematics*（《数学进阶题》）则有更多关于递推关系的 STEP 风格练习。

详细解答

STEP III 2012，第 9 题

力学 • 转动滑轮的动力学

STEP III 2012，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：

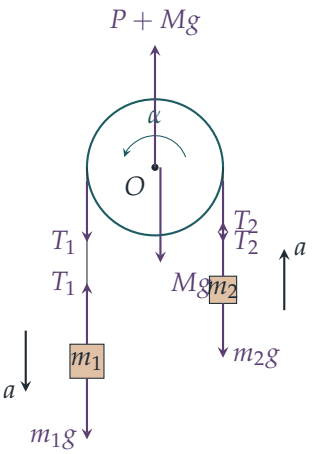
step.maths.org/questions/12-s3-q9

关键思路

测得的轴上作用力就是线索。滑轮的质心（centre of mass）不会移动，所以轴必须承托滑轮的重量，同时平衡两段绳的张力（tension）。由此 P 就是两个张力之和，从而可以求出两个质点的加速度。知道加速度之后，滑轮的转动方程就能给出转动惯量（moment of inertia）。摩擦力偶（friction couple）只会改变这一条方程，其他一概不变。

解答

模型。两个质点视为质点，绳是轻（light）而不可伸长（inextensible）的。绳在滑轮上不打滑，所以绳的速率等于圆盘边缘的速率。我们假设系统由静止释放，因此较重的质点 m_1 下降。设 a 为 m_1 向下的加速度，这也是 m_2 向上的加速度。于是滑轮的角加速度（angular acceleration）为 $\alpha = a/r$。m_1 一侧绳的张力为 T_1，m_2 一侧为 T_2（两者不同，因为滑轮有惯性）。设 I 为滑轮绕轴的转动惯量。



对每个质点应用牛顿第二定律（Newton's second law），得

$$m_1 g - T_1 = m_1 a, \qquad T_2 - m_2 g = m_2 a,$$

所以

$$T_1 = m_1(g - a), \qquad T_2 = m_2(g + a).$$

轴上的作用力。滑轮的质心位于固定点 O，所以滑轮所受的合力为零。作用在滑轮上的力有：它的重量 Mg、两个张力（两者都向下拉，因为两侧的绳都是竖直悬垂的），以及轴施加的力。摩擦力偶（我们在第 (ii) 部分才加入）没有合力（resultant），所以不会在这里出现。因此轴上的竖直作用力为 $Mg + T_1 + T_2$，与测量值比较，得

$$P = T_1 + T_2 = (m_1 + m_2)g - (m_1 - m_2)a,$$

所以

$$a = \frac{(m_1 + m_2)g - P}{m_1 - m_2}. \tag{1}$$

这只用到测量值和两个质点的方程，所以在两部分中都成立。

第 (i) 部分

轴上没有摩擦时，只有张力对 O 有力矩（moment）。对轴取力矩，

$$(T_1 - T_2)r = I\alpha = \frac{Ia}{r}, \qquad \text{所以} \qquad I = \frac{(T_1 - T_2)r^2}{a}.$$

而 $T_1 - T_2 = (m_1 - m_2)g - (m_1 + m_2)a$，所以

$$I = r^2\left(\frac{(m_1 - m_2)g}{a} - (m_1 + m_2)\right).$$

把 (1) 代入 a：

$$\begin{aligned} I &= r^2\left(\frac{(m_1-m_2)^2g}{(m_1+m_2)g-P}-(m_1+m_2)\right)\\ &= r^2\,\frac{(m_1-m_2)^2g-(m_1+m_2)^2g+(m_1+m_2)P}{(m_1+m_2)g-P}\\ &= \frac{\big((m_1+m_2)P-4m_1m_2g\big)r^2}{(m_1+m_2)g-P}, \end{aligned}$$

其中用到 $(m_1+m_2)^2-(m_1-m_2)^2=4m_1m_2$。这就是 $(*)$，证毕。

第 (ii) 部分

把新的计算值记为 I'。摩擦力偶 C 阻碍转动，而转动方向是由 T_1 驱动的方向，所以力矩方程变成

$$(T_1-T_2)r-C=\frac{I'a}{r}.$$

加速度仍然由 (1) 给出，因为所用的是同一个测量值 P，而关于轴上作用力的论证并不涉及 C。所以 T_1-T_2 和 a 与第 (i) 部分完全相同，于是

$$I'=\frac{(T_1-T_2)r^2}{a}-\frac{Cr}{a}=(*)-\frac{Cr}{a}.$$

想一想 我们怎样知道 a 的正负？

系统由静止开始，滑轮转动时 m_1 向下移动，所以 m_1 必定有正的向下加速度。（如果 a 为零或负数，m_1 根本不会开始向下移动。）所以 $a>0$。由于 $C>0$ 且 $r>0$，得 $Cr/a>0$，因此**考虑摩擦后算出的值比 $(*)$ 小**。这是合理的：张力产生的净扭矩（torque）有一部分用来克服摩擦，所以只需较小的惯性就能解释同一个观测到的加速度。

C 的上界。把张力代入力矩方程：

$$\big((m_1-m_2)g-(m_1+m_2)a\big)r-C=\frac{I'a}{r},$$

整理后得

$$(m_1-m_2)rg-C=a\left(\frac{I'}{r}+(m_1+m_2)r\right).$$

右边 $a>0$、$I'\geq 0$（转动惯量不可能是负数）而且 $(m_1+m_2)r>0$，所以右边严格为正。因此

$$C<(m_1-m_2)rg,$$

证毕。换句话说，摩擦力偶必须小于两边重量不平衡所产生的力偶 $(m_1-m_2)gr$。否则系统根本不会开始移动。

验算

- 轻滑轮：若 $I = 0$ 且 $C = 0$，则 $a = \frac{(m_1 - m_2)g}{m_1 + m_2}$（熟悉的阿特伍德机（Atwood machine）结果），而 $P = \frac{4m_1 m_2 g}{m_1 + m_2}$。此时 $(*)$ 的分子为零，正如预期。
- 静止的滑轮：当 $a \to 0$ 时，$P \to (m_1 + m_2)g$，$(*)$ 的分母趋于零，$(*)$ 趋于无限大。无限重的滑轮不会产生加速度。
- 数值验算：取 $g = 9.8$、$m_1 = 3$、$m_2 = 1$、$r = 0.5$、$I = 0.7$ 及 $C = 2$（国际单位制），真实运动中 $a \approx 2.29\,\mathrm{m\,s^{-2}}$，$P \approx 34.6\,\mathrm{N}$。公式 $(*)$ 给出 $I \approx 1.14$，恰好多了 $Cr/a \approx 0.44$，而考虑摩擦的模型则还原出 0.7。这里 $C = 2 < (m_1 - m_2)rg = 9.8$。我们用同样方法测试了 10^4 个随机例子。

常见错误

- 以为两侧的张力相同。这只对轻滑轮或无摩擦滑轮成立，而在这里会令 I 从问题中消失。
- 计算轴上作用力时忘记滑轮本身的重量。这正是测量值写成 $P + Mg$ 的原因。
- 以为摩擦力偶会改变轴上的作用力。力偶的合力为零，所以两部分中都有 $P = T_1 + T_2$。

STEP 之外

这是从测量推断隐藏参数的一个小例子。答案取决于你所假设的模型，而忽略摩擦会令估计值朝一个可预测的方向偏差。在大学一年级的力学中，刚体会用线动量和角动量的完整理论来处理，而对于绕一般轴转动的物体，转动惯量会变成一个张量（tensor）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）深入讲解定轴转动，以及滑轮有质量的阿特伍德机。

详细解答

STEP III 2012，第 10 题

力学 · 圆周运动

STEP III 2012，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：

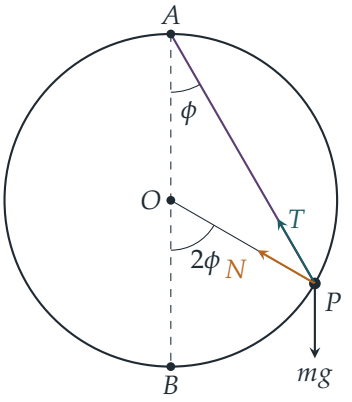
step.maths.org/questions/12-s3-q10

关键思路

一切都归结为一个等腰三角形（isosceles triangle）的几何：它由圆环的圆心、圆环的最高点和小环三点组成。只要看出绳长是 $2a\cos\phi$，而小环所在的半径与竖直向下方向成 2ϕ 角，整道题就只需一条能量方程（energy equation）（求 λ 和速率）和一条径向运动方程（radial equation of motion）（求反作用力）。

解答

我们把小环视为在光滑圆周上运动的质点，并把绳子视为遵守胡克定律（Hooke's law）的轻弹性绳。设 O 为圆环的圆心，A 为其最高点，B 为其最低点，P 为小环，并记 $c = \cos\phi$。



几何

三角形 OAP 中 $OA = OP = a$，所以它是等腰三角形，在 A 和 P 的两个角都是 ϕ。因此

$$AP = 2a\cos\phi, \qquad \angle AOP = \pi - 2\phi, \qquad \angle BOP = 2\phi.$$

于是 P 在 O 下方的深度是 $a\cos 2\phi$，而它在 B 上方的高度是 $a(1-\cos 2\phi) = 2a\sin^2\phi$。绳子（拉紧时）的伸长量（extension）为 $x = 2a\cos\phi - a = a(2c-1)$；当 $2c-1=0$，即 $\phi = \frac{\pi}{3}$ 时，绳子变得松弛。

弹性模量

圆环是光滑的，反作用力（reaction）与运动方向垂直，所以机械能守恒（conservation of mechanical energy）。在 B（$\phi = 0$）时，小环静止，伸长量为 a，所以弹性势能为 $\lambda a^2/(2a) = \frac{1}{2}\lambda a$。在 $\phi = \frac{\pi}{3}$ 时，小环再次静止，绳子刚好变得松弛（没有弹性势能），而小环已上升了 $2a\sin^2\frac{\pi}{3} = \frac{3}{2}a$。因此

$$\tfrac{1}{2}\lambda a = mg\cdot\tfrac{3}{2}a \quad\Longrightarrow\quad \lambda = 3mg.$$

这就是弹性模量（modulus of elasticity）。

速率

在一般角度 ϕ 而绳子拉紧时，由 B 起应用能量守恒，得

$$\begin{aligned}\tfrac{1}{2}mv^2 &= \tfrac{1}{2}\lambda a - \frac{\lambda a^2(2c-1)^2}{2a} - 2mga\sin^2\phi\\ &= \tfrac{3}{2}mga\left[1-(2c-1)^2\right] - 2mga(1-c^2)\\ &= \tfrac{3}{2}mga(4c-4c^2) - 2mga + 2mgac^2\\ &= mga(-4c^2+6c-2),\end{aligned}$$

所以

$$\frac{mv^2}{a} = mg(-8c^2+12c-4) = -4mg(2c-1)(c-1).$$

此式在 $c=1$ 和 $c=\frac{1}{2}$ 时为零，在两者之间为正，这与小环在 $\phi = -\frac{\pi}{3}$ 与 $\phi = \frac{\pi}{3}$ 之间来回运动相符：小环静止后会滑回并经过 B，到达另一边的对称点，如此往复。整个过程中 $\frac{1}{2} \le c \le 1$，绳子一直拉紧（只在两个端点刚好松弛），张力为

$$T = \frac{\lambda x}{a} = 3mg(2c-1).$$

反作用力

设 N 为反作用力指向 O 的分量，如图所示（结果可能是负数）。我们沿 PO 指向圆心的方向分解各力：

- 张力沿 PA 作用，而 PA 与 PO 成 ϕ 角（即等腰三角形的底角），所以张力指向 O 的分量是 $T\cos\phi$；
- OP 与竖直向下方向成 2ϕ 角，所以重力有一个背离 O 的分量 $mg\cos 2\phi$。

> **想一想** 为甚么这里重力的分量是向外而不是向内?

当 $\phi < \frac{\pi}{4}$ 时，小环在 O 的下方，重力把它拉离圆心；单一表达式 $-mg\cos 2\phi$（指向 O）已涵盖两种情况，因为当 P 越过 O 的水平高度时，$\cos 2\phi$ 会变号。

沿指向圆心的方向应用牛顿第二定律（Newton's second law），得

$$N + T\cos\phi - mg\cos 2\phi = \frac{mv^2}{a}.$$

代入 T、mv^2/a 及 $\cos 2\phi = 2c^2 - 1$：

$$\begin{aligned} N &= mg(-8c^2 + 12c - 4) - 3mg(2c-1)c + mg(2c^2 - 1) \\ &= mg(-8c^2 + 12c - 4 - 6c^2 + 3c + 2c^2 - 1) \\ &= -mg(12c^2 - 15c + 5). \end{aligned}$$

因此反作用力的大小为

$$R = (12\cos^2\phi - 15\cos\phi + 5)\, mg,$$

正如所求，前提是括号内的式子永不为负，下面我们便证明这一点。（负号告诉我们，在整个运动中，圆环都是把小环向外推，即背离 O。）

R 永不为零

二次式 $12c^2 - 15c + 5$ 的判别式（discriminant）为 $15^2 - 4\cdot 12\cdot 5 = 225 - 240 = -15 < 0$，而首项系数（leading coefficient）为正，所以它对每个实数 c 都严格为正。配方（completing the square）得

$$12c^2 - 15c + 5 = 12\left(c - \tfrac{5}{8}\right)^2 + \tfrac{5}{16} \geq \tfrac{5}{16},$$

所以在整个运动中 $R \geq \frac{5}{16}mg > 0$，最小值出现在 $\cos\phi = \frac{5}{8}$ 时（此值位于 $[\frac{1}{2}, 1]$ 内，所以确实能取到）。

验算

- 在 B（$c = 1$）：小环静止，$T = 3mg$ 向上，重力 mg 向下，所以圆环必须以 $2mg$ 向下推。公式给出 $(12 - 15 + 5)mg = 2mg$。
- 在松弛点（$c = \frac{1}{2}$）：没有张力，速率为零，P 在 O 上方 $\frac{a}{2}$ 处，所以径向反作用力与向内的重力分量 $mg\cos\frac{\pi}{3} = \frac{1}{2}mg$ 平衡。公式给出 $(3 - 7.5 + 5)mg = \frac{1}{2}mg$。

- 我们以数值方法积分运动方程（四阶龙格－库塔法（RK4），从稍为偏离 B 的位置出发）：小环在距 B 角度 $2.094 \approx \frac{2\pi}{3}$ 处停下，计算所得的反作用力与公式的误差在 10^{-7} 以内，其最小值为 $0.3125\,mg = \frac{5}{16}mg$。

常见错误

- 把小环的角位置当作 ϕ 而不是 2ϕ：圆周角是圆心角的一半。
- 在胡克定律或弹性势能中，用了整条绳长 $2a\cos\phi$，而不是伸长量 $2a\cos\phi - a$。
- 假设反作用力指向内，然后对得到负数答案感到困惑；其实题目只要求反作用力的大小。

STEP 之外

小环只有一个自由度（degree of freedom），处理这类系统最简洁的方法，是用单一坐标写出能量（或拉格朗日（Lagrangian））表述，之后再由牛顿定律求出像 R 这样的约束力（constraint force）。这种分工——由能量得出运动方程，由分解力得出约束力——正正是大学一年级力学课程处理“金属线上的珠子”问题的方法。Kleppner 和 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）便以这种思路讲解带约束的圆周运动和能量方法。

详细解答

STEP III 2012，第 11 题

力学 • 下落的对折绳

STEP III 2012，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/12-s3-q11

关键思路

一切都源于绳圈的几何。只要知道质点下落了 x 时，静止部分和移动部分各有多长，就可以写出损失的势能（potential energy）（利用每段直绳的质心（centre of mass））和增加的动能（kinetic energy）（只有质点和移动部分具有速率）。由能量守恒（conservation of energy）便可得出 v^2 作为 x 的函数，而加速度则可由 $\frac{1}{2}v^2$ 对 x 求导得出，因为 $v\,\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{\mathrm{d}v}{\mathrm{d}t}$。

解答

模型与几何

我们把绳子视为一条均匀、不可伸长（inextensible）、完全柔软的线，其单位长度质量为

$$\rho = \frac{2M}{2L} = \frac{M}{L},$$

把质点视为点质量，并假设绳圈底部的弯曲部分大小可以忽略。深度由 P 向下量度。

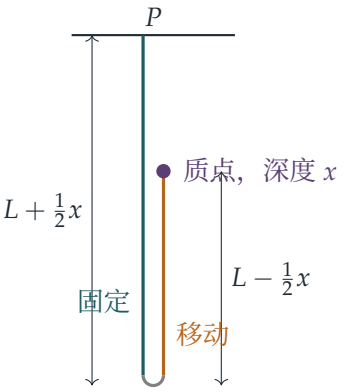
当质点下落了距离 x 时，绳子由两段竖直部分组成，两者在绳圈底部相接，设该处深度为 d：

- 固定部分，由 P 向下延伸至底部，处于静止，长度为 d；
- 移动部分，由质点（深度 x）向下延伸至底部，长度为 $d - x$。

总长度为 $2L$，所以 $d + (d - x) = 2L$，得

$$d = L + \tfrac{1}{2}x, \qquad \text{移动部分长度} = L - \tfrac{1}{2}x.$$

（开始时 $x = 0$，两部分的长度都是 L。）



损失的势能

每段直绳都是均匀的，所以其质心位于中点。在深度 h 处，质量 μ 的势能为 $-\mu gh$。设 $D(x)$ 为整个系统中各部分“质量 × 深度”之和：

$$\begin{aligned} D(x) &= mx + \rho d \cdot \frac{d}{2} + \rho(d-x)\cdot\frac{x+d}{2} \\ &= mx + \frac{\rho}{2}(2d^2 - x^2) \\ &= mx + \rho\left(L^2 + Lx + \tfrac{1}{4}x^2 - \tfrac{1}{2}x^2\right) = mx + \rho\left(L^2 + Lx - \tfrac{1}{4}x^2\right). \end{aligned}$$

开始时 $D(0) = \rho L^2$，所以损失的势能为

$$g\big(D(x) - D(0)\big) = gx\left(m + \rho L - \tfrac{1}{4}\rho x\right) = \frac{gx}{L}\left(mL + ML - \tfrac{1}{4}Mx\right),$$

这里用了 $\rho L = M$。

增加的动能

固定部分处于静止。质点以及移动部分上的每一点都具有速率 v（题目要求我们如此假设），所以移动中的质量为 $m + \rho\big(L - \frac{1}{2}x\big)$，而

$$\text{动能} = \tfrac{1}{2}\left(m + M - \frac{Mx}{2L}\right)v^2 = \frac{1}{2L}\left(mL + ML - \tfrac{1}{2}Mx\right)v^2.$$

系统由静止开始，所以由能量守恒得动能 = 损失的势能：

$$\frac{1}{2L}\left(mL + ML - \tfrac{1}{2}Mx\right)v^2 = \frac{gx}{L}\left(mL + ML - \tfrac{1}{4}Mx\right),$$

即

$$v^2 = \frac{2gx\big(mL + ML - \frac{1}{4}Mx\big)}{mL + ML - \frac{1}{2}Mx},$$

此即所求。当 $x < 2L$ 时分母永不为零，因为此时 $ML - \frac{1}{2}Mx > 0$，所以 $mL + ML - \frac{1}{2}Mx > mL \geq 0$。

加速度

为简洁起见，记

$$A = mL + ML - \tfrac{1}{4}Mx, \qquad B = mL + ML - \tfrac{1}{2}Mx,$$

于是 $\frac{1}{2}v^2 = gxA/B$，且 $A' = -\frac{1}{4}M$、$B' = -\frac{1}{2}M$（对 x 求导）。

> **想一想** 已知 v^2 是 x 的函数，怎样才能得出加速度?

由链式法则（chain rule），$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}v^2\right) = v\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{\mathrm{d}v}{\mathrm{d}t}$，这正是加速度。只要质点在向下移动，即 x 随 t 递增，这个做法便成立。

由乘积法则（product rule）和商法则（quotient rule），

$$\begin{aligned}\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{gxA}{B}\right) &= g\left[\frac{A}{B} + x\,\frac{A'B - AB'}{B^2}\right] \\ &= g\left[\frac{A}{B} + x\,\frac{-\frac{1}{4}MB + \frac{1}{2}MA}{B^2}\right] \\ &= g\left[\frac{A - \frac{1}{4}Mx}{B} + \frac{MxA}{2B^2}\right].\end{aligned}$$

关键的化简在于 $A - \frac{1}{4}Mx = mL + ML - \frac{1}{2}Mx = B$，所以第一个分数恰好等于 1。因此

$$\frac{\mathrm{d}v}{\mathrm{d}t} = g + \frac{Mgx\left(mL + ML - \frac{1}{4}Mx\right)}{2\left(mL + ML - \frac{1}{2}Mx\right)^2},$$

此即所求。

加速度大于 g

释放后，$0 < x < 2L$。此时

- 如上所示，$B = mL + M\left(L - \frac{1}{2}x\right) > 0$，所以 $B^2 > 0$；
- $A = B + \frac{1}{4}Mx > B > 0$；
- $M > 0$、$g > 0$ 且 $x > 0$。

所以额外项 $\frac{MgxA}{2B^2}$ 严格为正，对于 $0 < x < 2L$ 中的每个 x，加速度都严格大于 g。（在释放的一刻，$x = 0$，加速度恰好是 g。）

验算

- 若 $M = 0$（轻绳），两条公式都化为自由落体（free fall）的情况：$v^2 = 2gx$，加速度为 g。

- 取 $m = 0.3$、$M = L = 1$ 及 $x = 1$，并设 $g = 9.8$：则 $A = 1.05$、$B = 0.8$，所以 $v^2 = 25.725$，而加速度为 $9.8 + 9.8 \times 1.05/1.28 \approx 17.84\,\mathrm{m\,s^{-2}}$，远大于 g。在这一点对 $\frac{1}{2}v^2$ 作数值求导，得出相同的值；而由静止开始逐步推进运动方程，亦能重现 v^2 的公式。
- 单位：mL、ML 和 Mx 的单位都是“质量 × 长度”，所以各分数都是无量纲（dimensionless）的，而 v^2 的单位与 gx 相同。

常见错误

- 把移动部分的长度弄错：绳圈底部向下移动 $\frac{1}{2}x$，而不是 x，因为下落一侧所收入的绳子由两侧平分。
- 给固定部分加上动能，或忘记固定部分的势能同样会改变（它变长了，所以其质心下降）。
- 对 v^2 而不是 $\frac{1}{2}v^2$ 求导，结果得出加速度的两倍。

STEP 之外

这是经典的“下落折叠链”问题，是一个运动中的质量随时间改变的系统。这里采用的能量守恒模型预测链端的加速度大于 g；若 $m = 0$，当 $x \to 2L$ 时速率会无限增大，这正是挥鞭时发出爆响背后的原理。在弯曲处能量是否真的守恒，在研究文献中一直有争论；而以真实链条进行的实验显示，加速度确实如这里所述般大于 g。变质量系统（variable-mass systems），以及如何审慎地选用能量或动量（momentum）方法，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（《力学导论》）。

详细解答

STEP III 2012，第 12 题

概率与统计 · 几何概率

> **STEP III 2012，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/12-s3-q12

关键思路

把三角形分成三个以 P 为公共顶点的小三角形，就可看出三段距离之和必定等于高。至于最小值，最简洁的切入点是尾概率（tail probability）$\mathrm{P}(X > x)$。“每段距离都大于 x”这个事件对应一个与原三角形相似（similar）的较小三角形，而相似图形的面积按长度比的平方缩放。在三维空间中，同一思路改用体积亦同样奏效，而体积按立方缩放。

解答

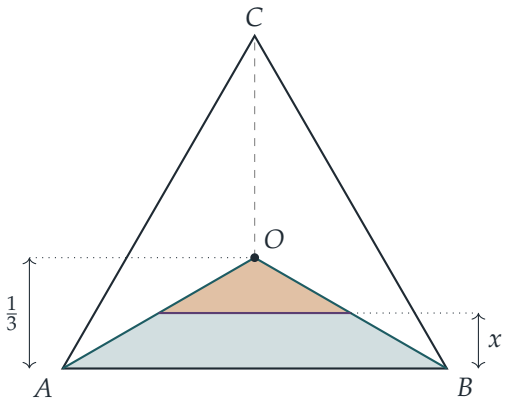
第 (i) 部分

设三角形的边长为 s。由于高为 1，ABC 的面积为 $\frac{1}{2}s$。把 P 与三个顶点连起来，便把 ABC 分成三角形 PAB、PBC 和 PCA。它们的底就是三条边，长度均为 s，而它们的高分别是 x_1、x_2 和 x_3。把面积相加，

$$\tfrac{1}{2}sx_1 + \tfrac{1}{2}sx_2 + \tfrac{1}{2}sx_3 = \tfrac{1}{2}s, \qquad \text{所以} \qquad x_1 + x_2 + x_3 = 1,$$

得证。

X_1 **为最小值的区域**。与 AB 和 CA 等距的点都在角 A 的角平分线（angle bisector）上。在这条角平分线靠 B 的一侧，P 与 AB 的距离比与 CA 的距离小。同理，P 与 AB 的距离比与 BC 的距离小，当且仅当 P 位于角 B 的角平分线靠 A 的一侧。在等边三角形中，角平分线同时也是中线（median），三者相交于中心 O。因此，$X_1 = \min(X_1, X_2, X_3)$ 当且仅当 P 位于三角形 ABO 内（下图阴影部分）。O 到 AB 的距离是高的三分之一，即 $\frac{1}{3}$。



X 的概率密度函数。三个三角形 ABO、BCO 和 CAO 全等（congruent），所以各自的概率都是 $\frac{1}{3}$。（距离相等的情况只会出现在角平分线上，而角平分线的面积为零。）由对称性，X 的分布（distribution）与“在 P 位于 ABO 内的条件下”X_1 的分布相同。在这情况下 $X = X_1$，即 P 到 AB 的距离。

? **想一想** 在三角形 ABO 内，哪些点满足 $X_1 > x$?

它们是位于与 AB 平行、相距 x 的直线上方的点，亦即以 O 为顶点的琥珀色小三角形。它与 ABO 相似，高为 $\frac{1}{3} - x$，而 ABO 的高为 $\frac{1}{3}$，所以它的面积是 ABO 面积的 $\left((\frac{1}{3} - x)/\frac{1}{3}\right)^2 = (1-3x)^2$ 倍。因此，当 $0 \le x \le \frac{1}{3}$ 时，

$$\mathrm{P}(X > x) = (1-3x)^2, \qquad \mathrm{F}(x) = \mathrm{P}(X \le x) = 1 - (1-3x)^2.$$

此外，由于三段距离之和为 1，X 不可能超过 $\frac{1}{3}$。求导，

$$\mathrm{f}(x) = \mathrm{F}'(x) = 6(1-3x) \quad (0 \le x \le \tfrac{1}{3}),$$

其他情况下 $\mathrm{f}(x) = 0$，得证。

期望值（expected value）。

$$\mathrm{E}(X) = \int_0^{1/3} 6x(1-3x)\,\mathrm{d}x = \left[3x^2 - 6x^3\right]_0^{1/3} = \tfrac{1}{3} - \tfrac{2}{9} = \tfrac{1}{9}.$$

第 (ii) 部分

设该点到四个面的距离为 $Y_1, \ldots, Y_4$，并设 $Y = \min(Y_1, \ldots, Y_4)$。每个面的面积都是 S，而体积为 $\frac{1}{3}S \cdot 1$。把该点与四个顶点连起来，便把四面体分成四个较小的四面体。每个小四面体以一个面为底，以其中一个 Y_i 为高，所以

$$\tfrac{1}{3}S(Y_1 + Y_2 + Y_3 + Y_4) = \tfrac{1}{3}S, \qquad \text{所以} \qquad Y_1 + Y_2 + Y_3 + Y_4 = 1.$$

重心（centroid）到每个面的距离都是 $\frac{1}{4}$，所以内切球（inscribed sphere）的半径为 $\frac{1}{4}$。

想一想 现在集合 $\{Y > y\}$ 是甚么样子?

它是与四个面的距离都大于 y 的点所组成的集合。每个面平行地向内移动 y，得出一个重心相同的较小正四面体（regular tetrahedron）。重心到每个面的距离由 $\frac{1}{4}$ 降至 $\frac{1}{4} - y$，所以比例因子（scale factor）为 $(\frac{1}{4} - y)/\frac{1}{4} = 1 - 4y$。体积按比例因子的立方缩放，所以当 $0 \le y \le \frac{1}{4}$ 时，

$$\mathrm{P}(Y > y) = (1 - 4y)^3.$$

（把同一论证应用于第 (i) 部分的三角形，可直接得出 $(1 - 3x)^2$：内部三角形的内切圆半径（inradius）为 $\frac{1}{3} - x$。）

由于 $Y \ge 0$，我们可以把尾概率积分来求 $\mathrm{E}(Y)$：

$$\mathrm{E}(Y) = \int_0^{1/4} \mathrm{P}(Y > y)\,\mathrm{d}y = \int_0^{1/4} (1 - 4y)^3\,\mathrm{d}y = \left[-\tfrac{1}{16}(1 - 4y)^4\right]_0^{1/4} = \tfrac{1}{16}.$$

如果你较喜欢用概率密度函数，它在 $[0, \frac{1}{4}]$ 上是 $12(1 - 4y)^2$，而 $\int_0^{1/4} 12y(1 - 4y)^2\,\mathrm{d}y = \frac{1}{16}$，结果一样。

所以到最近一个面的距离的期望值是 $\frac{1}{16}$。

验算

- $\int_0^{1/3} 6(1 - 3x)\,\mathrm{d}x = 1$，所以 f 确实是一个概率密度函数。
- 这里有一个规律：在高为 1 的三角形中，$\mathrm{E}(X) = \frac{1}{9} = \frac{1}{3^2}$；在高为 1 的四面体中，$\mathrm{E}(Y) = \frac{1}{16} = \frac{1}{4^2}$。在一维情况下，$[0, 1]$ 中随机一点到较近端点的距离的平均值是 $\frac{1}{4} = \frac{1}{2^2}$，同样吻合。
- 在三角形内取 4×10^5 个随机点进行蒙地卡罗模拟（Monte Carlo simulation），得出平均值 0.1110（理论值 $\frac{1}{9} \approx 0.1111$），而直方图（histogram）与 $6(1 - 3x)$ 吻合。在四面体内取 3×10^5 个点，平均值为 0.0625（理论值 $\frac{1}{16} = 0.0625$）。

常见错误

- 对 $\mathrm{P}(X > x)$ 求导时漏了负号：概率密度函数是 $-\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{P}(X > x)$。
- 忘记内部三角形的高是 $1 - 3x$ 而非 $1 - x$：三条边各自向内移动了 x。
- 在第 (ii) 部分用了比例因子的平方。体积是按立方缩放的。

STEP 之外

这道题目中的两个工具，在大学第一个概率课程里经常出现：通过 $\mathrm{P}(\min > x)$ 求最小值的分布，以及对非负随机变量（random variable）计算 $\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y$。这个规律还可以延续下去：在高为 1 的 n 维单形（simplex）中，到最近一个面的距离的期望值是 $1/(n+1)^2$。Grimmett 与 Welsh 的 *Probability: An Introduction*（《概率导论》）对这两个工具都有细致的论述。

详细解答

STEP III 2012，第 13 题

概率与统计 · 截尾正态分布

STEP III 2012，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/12-s3-q13

关键思路

在某区间上的条件期望（conditional expectation），就是 $z\,f(z)$ 在该区间上的积分，再除以该区间的概率。对标准正态分布（standard Normal）而言，$z\,\mathrm{e}^{-z^2/2}$ 恰好是 $-\mathrm{e}^{-z^2/2}$ 的导数，因此这个积分可以精确求出。其余的工作只是标准化（standardising），即 $X = \mu + \sigma Z$，以及按 X 的正负号把 $|X|$ 拆开处理。

解答

全文以 $\phi(z) = \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}$ 表示 Z 的概率密度函数（probability density function），于是 $\Phi'(z) = \phi(z)$ 及 $\phi'(z) = -z\,\phi(z)$。

第 (i) 部分

在 $a < Z < b$ 的条件下，Z 的条件密度在 (a, b) 上为 $\phi(z)/\mathrm{P}(a < Z < b)$，而 $\mathrm{P}(a < Z < b) = \Phi(b) - \Phi(a)$。因此

$$\begin{aligned}
\mathrm{E}(Z \mid a < Z < b) &= \frac{1}{\Phi(b) - \Phi(a)} \int_a^b z\,\phi(z)\,\mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}\,\big(\Phi(b) - \Phi(a)\big)} \Big[-\mathrm{e}^{-z^2/2} \Big]_a^b \\
&= \frac{\mathrm{e}^{-a^2/2} - \mathrm{e}^{-b^2/2}}{\sqrt{2\pi}\,\big(\Phi(b) - \Phi(a)\big)},
\end{aligned}$$

因为 $\dfrac{\mathrm{d}}{\mathrm{d}z}\big(-\mathrm{e}^{-z^2/2}\big) = z\,\mathrm{e}^{-z^2/2}$。这就是所求的结果。由于该积分收敛（converges），当 $b = \infty$（或 $a = -\infty$）时结果同样成立，只需把 $\mathrm{e}^{-\infty} = 0$ 及 $\Phi(\infty) = 1$ 代入理解即可。

第 (ii) 部分

设 $X = \mu + \sigma Z$，其中 $Z \sim \mathrm{N}(0,1)$ 且 $\sigma > 0$。由于 $\sigma > 0$，

$$X > 0 \iff Z > -\mu/\sigma,$$

所以以 $X > 0$ 为条件，等同于以 $Z > -\mu/\sigma$ 为条件。期望在条件下同样是线性（linear）的，故

$$\mathrm{E}(X \mid X > 0) = \mathrm{E}(\mu + \sigma Z \mid Z > -\mu/\sigma) = \mu + \sigma\,\mathrm{E}(Z \mid Z > -\mu/\sigma),$$

此即所求。

$|X|$ 的平均值。令 $c = \mu/\sigma$。在第 (i) 部分中取 $a = -c$ 及 $b = \infty$，得

$$\mathrm{E}(Z \mid Z > -c) = \frac{\mathrm{e}^{-c^2/2}}{\sqrt{2\pi}\,\big(1 - \Phi(-c)\big)} = \frac{\phi(c)}{p}, \qquad \text{其中} p = \mathrm{P}(X > 0) = 1 - \Phi(-c).$$

把上一条式子乘以 p，得

$$\mathrm{E}(X;\, X > 0) := \mathrm{E}(X \mid X > 0)\,\mathrm{P}(X > 0) = \mu p + \sigma\phi(c).$$

> ? **想一想** 怎样才能不再计算第二个积分，就得到 $X < 0$ 部分的贡献?

两部分加起来就是整个平均值：$\mathrm{E}(X;\, X > 0) + \mathrm{E}(X;\, X < 0) = \mathrm{E}(X) = \mu$（事件 $X = 0$ 的概率为 0）。所以 $\mathrm{E}(X;\, X < 0) = \mu - \mu p - \sigma\phi(c)$。在 $X < 0$ 时有 $|X| = -X$，因此

$$\begin{aligned} m = \mathrm{E}|X| &= \mathrm{E}(X;\, X > 0) - \mathrm{E}(X;\, X < 0) \\ &= 2\big(\mu p + \sigma\phi(c)\big) - \mu \\ &= \mu(2p - 1) + 2\sigma\phi(c). \end{aligned}$$

最后，$2p - 1 = 1 - 2\Phi(-\mu/\sigma)$ 及 $2\phi(c) = \sqrt{2/\pi}\,\mathrm{e}^{-\mu^2/(2\sigma^2)}$，得

$$m = \mu\big(1 - 2\Phi(-\mu/\sigma)\big) + \sigma\sqrt{2/\pi}\,\exp\big(-\tfrac{1}{2}\mu^2/\sigma^2\big),$$

此即所求。

$|X|$ 的方差。关键在于 $|X|^2 = X^2$，所以二阶矩（second moment）无须另作计算：

$$\mathrm{E}\big(|X|^2\big) = \mathrm{E}(X^2) = \mathrm{Var}(X) + \big(\mathrm{E}X\big)^2 = \sigma^2 + \mu^2.$$

因此

$$\mathrm{Var}|X| = \mu^2 + \sigma^2 - m^2.$$

验算

- 若 $\mu = 0$，则 $|X|$ 服从半正态分布（half-Normal），公式给出 $m = \sigma\sqrt{2/\pi}$，正是熟悉的数值；此时方差为 $\sigma^2(1 - 2/\pi)$，是正数。
- 若 $\mu \gg \sigma$，则 X 几乎总是正数，$\Phi(-\mu/\sigma) \approx 0$，而指数项极小，所以 $m \approx \mu$ 及 $\mathrm{Var}|X| \approx \sigma^2$，与预期相符。
- 把 μ 换成 $-\mu$ 时，m 的公式不变（利用 $1 - 2\Phi(-c) = -(1 - 2\Phi(c))$），与对称性 $|X| \sim |-X|$ 吻合。
- 数值上，取 $\mu = 1$、$\sigma = 2$，得 $m \approx 1.7912$ 及 $\mathrm{Var}|X| \approx 1.7917$，与直接数值积分（quadrature）以及 2×10^5 个数值的模拟（simulation）结果一致。

常见错误

- 在第 (i) 部分忘记除以 $\Phi(b) - \Phi(a)$：单是积分本身只是 $\mathrm{E}(Z;\, a < Z < b)$，并非条件期望。
- 写成 $\mathrm{E}|X| = \mathrm{E}(X \mid X > 0)$：条件期望必须分别以 $\mathrm{P}(X > 0)$ 及 $\mathrm{P}(X < 0)$ 加权。
- 用另一个积分去计算 $\mathrm{E}(X^2)$，其实 $|X|^2 = X^2$ 即可直接得出。

STEP 之外

$|X|$ 的分布称为**折叠正态分布**（folded Normal），而第 (i) 部分中的条件分布则是**截断正态分布**（truncated Normal）；比值 $\phi(c)/\big(1 - \Phi(-c)\big)$ 称为逆米尔斯比率（inverse Mills ratio），在计量经济学（econometrics）中处理删失（censored）或经筛选数据的模型里担当核心角色。技巧 $\phi'(z) = -z\phi(z)$ 亦是用于正态逼近的斯坦因方法（Stein's method）的起点。如想深入了解条件期望，可参阅 Blitzstein 与 Hwang 的 *Introduction to Probability*（《概率导论》）。

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

2012 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。

