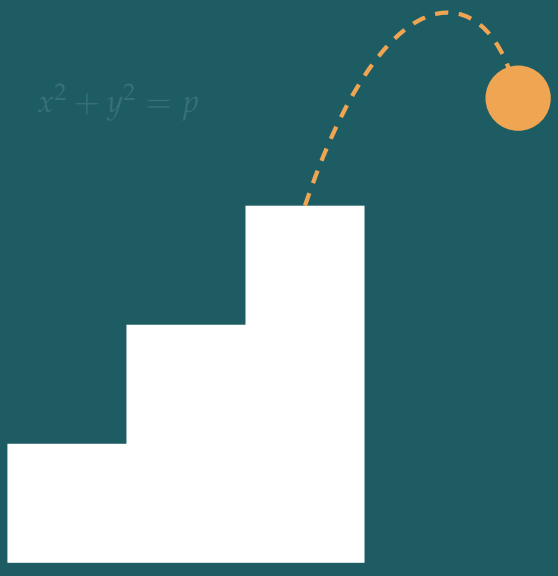
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

样本试卷详细解答



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP I, II 及 III

样本试卷详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

样本试卷详细解答

STEP 样本试卷（STEP I, II 及 III）全部 48 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 判别式与部分分式 纯数学 . . . 3

第 2 题 三角学与最优化 纯数学 . . . 7

第 3 题 二阶微分方程 纯数学 . . . 10

第 4 题 代换积分法 纯数学 . . . 13

第 5 题 数字计数与阶乘 纯数学 . . . 16

第 6 题 弓形的面积 纯数学 . . . 19

第 7 题 复指数级数 纯数学 . . . 23

第 8 题 积分方程 纯数学 . . . 26

第 9 题 绝对值不等式 纯数学 . . . 29

第 10 题 功率与阻力 力学 . . . 33

第 11 题 摩擦力与滚动接触 力学 . . . 36

第 12 题 涉及摩擦力的碰撞 力学 . . . 40

第 13 题 圆周运动 力学 . . . 44

第 14 题 期望值 概率与统计 . . . 47

第 15 题 随机变量之和 概率与统计 . . . 50

第 16 题 连续随机变量 概率与统计 . . . 53

STEP II

第 1 题 部分分式 纯数学 . . . 57

第 2 题 对称联立方程 纯数学 . . . 60

第 3 题 隶美弗定理 纯数学 . . . 64

第 4 题 双曲函数 纯数学 . . . 67

第 5 题 级数与积分界限 纯数学 . . . 71

第 6 题 由常微分方程推导加法公式 纯数学 . . . 75

第 7 题 阶梯函数的积分 纯数学 78
第 8 题 模群中元素的阶 纯数学 82
第 9 题 球面、平面与反射 纯数学 86
第 10 题 逻辑谜题 纯数学 90
第 11 题 抛体与碰撞 力学 93
第 12 题 金属线上的振动 力学 97
第 13 题 平衡与力矩 力学 101
第 14 题 刚体的转动 力学 105
第 15 题 期望值与样本平均值 概率与统计 109
第 16 题 淘汰赛的概率 概率与统计 113

STEP III

第 1 题 数学归纳法 纯数学 117
第 2 题 二项式不等式与极限 纯数学 120
第 3 题 复数的三角不等式 纯数学 123
第 4 题 抛物线的切线 纯数学 126
第 5 题 方程根的计数 纯数学 129
第 6 题 双曲函数归约公式 纯数学 133
第 7 题 降阶法 纯数学 136
第 8 题 向量与柯西–施瓦兹不等式 纯数学 139
第 9 题 矩阵群 纯数学 142
第 10 题 模算术 纯数学 145
第 11 题 阶微分方程 力学 148
第 12 题 杆的平衡 力学 152
第 13 题 滚动圆环 力学 155
第 14 题 转动碰撞 力学 159
第 15 题 随机网络的连通性 概率与统计 163
第 16 题 二项分布 概率与统计 166

详细解答

STEP I 样本试卷，第 1 题

纯数学 · 判别式与部分分式

STEP I 样本试卷，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/spec-s1-q1

关键思路

在第 (i) 部分，对每个固定的 x，方程都是 y 的二次方程（quadratic equation），而只有当判别式（discriminant）非负时，才存在实数 y；这个条件正好就是对 x 的限制。把两者的角色对调，方程同样是 x 的二次方程，由此得出对 y 的限制。第 (ii) 部分是一道标准的部分分式（partial fractions）题目：把分式拆成若干项，而每一项的二项展开式（binomial expansion）都是我们已知的，然后把 x^n 的系数相加。

解答

第 (i) 部分

把方程看成 y 的二次方程，其系数取决于 x：

$$y^2 + 16x\,y + (4x^2 + 24x) = 0.$$

对于给定的实数 x，存在实数 y 满足此方程，当且仅当判别式非负：

$$\begin{aligned}(16x)^2 - 4(4x^2 + 24x) &= 256x^2 - 16x^2 - 96x \\ &= 240x^2 - 96x = 48x(5x - 2) \geq 0.\end{aligned}$$

乘积 $x(5x-2)$ 非负，正好是在两个因子同号（或其中一个为零）的时候，即 $x \leq 0$ 或 $x \geq \frac{2}{5}$。因此任何实数解都满足

$$x \leq 0 \quad \text{或} \quad x \geq \tfrac{2}{5},$$

这正是所要证明的。

现在把方程看成 x 的二次方程：

$$4x^2 + (16y + 24)x + y^2 = 0.$$

存在实数 x，当且仅当

$$\begin{aligned}(16y + 24)^2 - 16y^2 &= 256y^2 + 768y + 576 - 16y^2 \\ &= 240y^2 + 768y + 576 \\ &= 48(5y^2 + 16y + 12) = 48(5y + 6)(y + 2) \geq 0.\end{aligned}$$

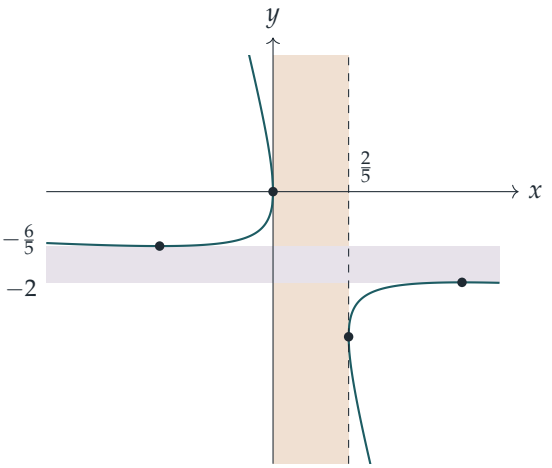
$(5y + 6)(y + 2)$ 的根是 $y = -2$ 和 $y = -\frac{6}{5}$，而这个二次式只在两根之间取负值。因此

$$y \leq -2 \quad \text{或} \quad y \geq -\tfrac{6}{5}.$$

想一想 这些限制是否已是最佳的？真正的取值范围会不会更小？

这些限制是精确的。判别式的论证是“当且仅当”的关系：每一个满足 $x \leq 0$ 或 $x \geq \frac{2}{5}$ 的 x 都至少对应一个实数 y，对 y 亦然。边界值同样可以取到，由重根（double root）给出：对于 x 的界限，是 $(x, y) = (0, 0)$ 和 $(\frac{2}{5}, -\frac{16}{5})$；对于 y 的界限，是 $(1, -2)$ 和 $(-\frac{3}{5}, -\frac{6}{5})$。你可以验证每一点都满足原方程。

从几何上看，这条曲线是双曲线（hyperbola）（因为 $16^2 > 4 \cdot 4 \cdot 1$），而被排除的两条带状区域就是它两支之间的空隙。



第 (ii) 部分

设

$$\frac{9}{(2-x)^2(1+x)} = \frac{A}{1+x} + \frac{B}{2-x} + \frac{C}{(2-x)^2},$$

于是 $9 = A(2-x)^2 + B(2-x)(1+x) + C(1+x)$。代入 $x=-1$ 得 $9 = 9A$，故 $A=1$；代入 $x=2$ 得 $9 = 3C$，故 $C=3$；代入 $x=0$ 得 $9 = 4A + 2B + C = 7 + 2B$，故 $B=1$。因此

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

现在把每一项展开，先抽出因子 2，使二项级数（binomial series）以 $\frac{x}{2}$ 的幂表示：

$$\frac{1}{1+x} = \sum_{n\geq 0} (-1)^n x^n, \qquad |x|<1,$$

$$\frac{1}{2-x} = \frac{1}{2}\Big(1-\frac{x}{2}\Big)^{-1} = \sum_{n\geq 0} \frac{x^n}{2^{n+1}}, \qquad |x|<2,$$

$$\frac{3}{(2-x)^2} = \frac{3}{4}\Big(1-\frac{x}{2}\Big)^{-2} = \frac{3}{4}\sum_{n\geq 0}(n+1)\frac{x^n}{2^n} = \sum_{n\geq 0}\frac{3(n+1)}{2^{n+2}}x^n, \qquad |x|<2,$$

其中用到当 $|u|<1$ 时 $(1-u)^{-2} = \sum_{n\geq 0}(n+1)u^n$。相加后，$x^n$ 的系数是

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

三个级数同时收敛（converge），正好是在 $|x|<1$ 的时候（第一个需要 $|x|<1$，其余两个只需要 $|x|<2$）。当 $1\leq |x|<2$ 时，后两个级数仍然收敛，但第一个不收敛，所以它们的和也不可能收敛。因此展开式成立的范围是

$$-1 < x < 1.$$

验算

在第 (i) 部分，点 $(1,-2)$ 给出 $4-32+4+24=0$，而 $(-\frac{3}{5},-\frac{6}{5})$ 给出 $\frac{36}{25}+\frac{288}{25}+\frac{36}{25}-\frac{360}{25}=0$，确认 y 的两个界限都可以取到。在第 (ii) 部分，$n=0$ 给出 $1+\frac{5}{4}=\frac{9}{4}$，与 $x=0$ 时的 $\frac{9}{4\cdot 1}$ 吻合；而 $n=1$ 给出 $-1+\frac{8}{8}=0$，这是正确的，因为分母 $(2-x)^2(1+x) = 4-3x^2+x^3$ 没有 x 项。从数值上看，在 $x=0.9$ 处把级数求和得 $3.914745\ldots$，与 $9/(1.1^2\times 1.9)$ 一致。

常见错误

- 把不等号的方向弄反，答成 $-2\leq y\leq -\frac{6}{5}$。判别式在两根之间为负，所以被排除的正是这个区间（而且 $(0,0)$ 在曲线上，其 $y=0$ 位于该区间之外）。
- 把 $(2-x)^{-1}$ 当作 $(1-x)^{-1}$ 来展开，忘记抽出因子 $\frac{1}{2}$（平方项则是 $\frac{1}{4}$）。
- 把成立范围写成 $|x|<2$：展开式只在每一项都收敛的地方才成立，所以是 $|x|<1$。

STEP 之外

第 (i) 部分的曲线是一条圆锥曲线（conic），它是椭圆、抛物线还是双曲线，取决于其二次项 $ax^2 + bxy + cy^2$ 的 $b^2 - 4ac$ 的正负号。在大学里，这会通过二次型（quadratic form）的对称矩阵（symmetric matrix）及其特征值（eigenvalue）来理解。第 (ii) 部分触及收敛半径（radius of convergence）的概念：展开式一直成立到函数趋向无穷大的最近一点为止，在这里就是 $x = -1$。Strang 的 *Introduction to Linear Algebra*（线性代数入门）讲述二次型，而 Abbott 的 *Understanding Analysis*（数学分析导论）则讲述幂级数（power series）及其收敛性。

详细解答

STEP I 样本试卷，第 2 题

纯数学 · 三角学与最优化

STEP I 样本试卷，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/spec-s1-q2

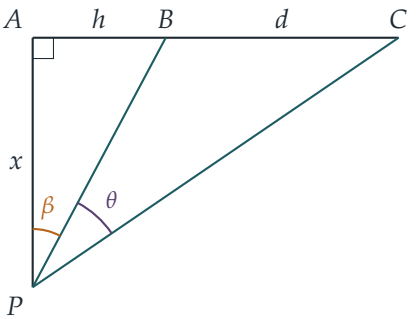
关键思路

角 θ 是以 P 为顶点的两个角之*差*，而这两个角各自位于一个直角在 A 的直角三角形内。它们的正切值可以直接从图中读出，因此只要用上 tan 的减法公式（subtraction formula）便可。至于求最大值，可把 $\tan\theta$ 改写成 x 只在一处出现的形式，即 $x + (\text{常数})/x$，再用算术平均—几何平均不等式（AM–GM inequality）便大功告成，完全不必动用微积分。

解答

求 $\tan\theta$

设 $\alpha = \angle APC$ 及 $\beta = \angle APB$。由于 B 位于 A 与 C 之间，故 $\theta = \alpha - \beta$。



在直角三角形 PAC 和 PAB 中（直角在 A），

$$\tan\alpha = \frac{AC}{AP} = \frac{h+d}{x}, \qquad \tan\beta = \frac{AB}{AP} = \frac{h}{x}.$$

由减法公式，

$$\tan\theta = \tan(\alpha-\beta) = \frac{\tan\alpha - \tan\beta}{1+\tan\alpha\tan\beta} = \frac{\dfrac{h+d}{x} - \dfrac{h}{x}}{1+\dfrac{h(h+d)}{x^2}}.$$

分子和分母同乘以 x^2，得

$$\boxed{\tan\theta = \frac{dx}{x^2+h(h+d)}}$$

（分母 $1+\tan\alpha\tan\beta$ 为正数，所以这条公式永远不会失效。）

求 θ 的最大值

> ? **想一想** 使 $\tan\theta$ 最大，是否等同于使 θ 最大?

在这里是的。α 和 β 都严格介乎 0 与 $\frac{\pi}{2}$ 之间，而且 $\alpha > \beta$，所以 $0 < \theta < \frac{\pi}{2}$。在这个区间上 tan 是严格递增（strictly increasing）的，因此 $\tan\theta$ 最大时 θ 也最大。

把分子和分母同除以 $x > 0$，得

$$\tan\theta = \frac{d}{x + \dfrac{h(h+d)}{x}}.$$

由于 d 是固定的，当分母 $x + h(h+d)/x$ 最小时，$\tan\theta$ 便最大。两项皆为正数，故由算术平均—几何平均不等式，

$$x + \frac{h(h+d)}{x} \;\geq\; 2\sqrt{x\cdot\frac{h(h+d)}{x}} = 2\sqrt{h(h+d)},$$

等号成立当且仅当两项相等，即 $x = h(h+d)/x$，亦即 $x^2 = h(h+d)$。所以分母取得最小值、θ 取得最大值之处为

$$\boxed{x = \sqrt{h(h+d)}}$$

此时 $\tan\theta = \dfrac{d}{2\sqrt{h(h+d)}}$。

*或用微积分。*设 $k = h(h+d)$ 及 $f(x) = dx/(x^2+k)$，由商法则（quotient rule）得

$$f'(x) = \frac{d\,(k-x^2)}{(x^2+k)^2},$$

它在 $0 < x < \sqrt{k}$ 时为正，在 $x > \sqrt{k}$ 时为负，所以 f 先上升至 $x = \sqrt{k}$ 处的唯一极大值，然后下降。算术平均—几何平均不等式的做法较快，而且一眼便看出这个极大值是全局最大值（global maximum）。

验算

- 极限：当 $x \to 0$ 或 $x \to \infty$ 时，公式给出 $\tan\theta \to 0$。这是合理的：从非常接近 A 的位置，或从非常远的地方看，线段 BC 在 P 所对的角几乎为零。所以两者之间必定有一个最大值。
- 对称性：把 x 换成 $h(h+d)/x$ 时，$\tan\theta$ 保持不变，所以最大值必定出现在这个对换的不动点（fixed point），即 $x = \sqrt{h(h+d)}$。取 $h = 2$、$d = 1$，则 $x = 2$ 和 $x = 3$ 两点给出相同的角（约 $11.31°$），而最大值在 $x = \sqrt{6}$ 处，约为 $11.54°$。
- 取 $h = 1$、$d = 3$：$x = 2$，而 $\tan\theta = 6/8 = \frac{3}{4}$，是 3–4–5 三角形中的角，即 $36.87°$。随机抽取 20 000 组 (h, d, x)，直接由坐标计算 θ，结果与公式相符至 10^{-14}；以数值方法搜索使角最大的 x，每次都与 $\sqrt{h(h+d)}$ 吻合。

常见错误

- 写成 $\tan\theta = d/x$，仿佛 θ 位于一个直角三角形内。事实并非如此：直角在 A，不在 BC 上。
- 求 $\tan\theta$ 的最大值，却没有说明为何这样便使 θ 最大；这需要 θ 一直处于 tan 递增的区间内。

STEP 之外

这个答案有一个漂亮的几何解释。使 BC 所对的角为定值的点，都位于通过 B 和 C 的圆的弧上；在仍然与直线 AP 相交的这类圆之中，最小的一个给出最大的角，而这个圆正是在 P 与该直线相切的那一个。切割线定理（tangent–secant theorem，即点 A 的幂）随即给出 $AP^2 = AB \cdot AC$，这正是 $x^2 = h(h+d)$。这就是雷格蒙塔努斯（Regiomontanus）问题，常见的问法是“应该站在哪里，才能把基座上的雕像看得最清楚?”。要优化某个量，先转而考虑另一个较易处理、但与它同升同降的量——这个想法贯穿大学一年级的分析学（analysis）；Siklos 的 *Advanced Problems in Mathematics* 一书中还有很多同类的题目。

详细解答

STEP I 样本试卷，第 3 题

纯数学 · 二阶微分方程

> **STEP I 样本试卷，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/spec-s1-q3

关键思路

二阶方程难解，两个一阶方程却容易。题目建议的代换（substitution）$z = y' - y$（以 y' 表示 $\mathrm{d}y/\mathrm{d}x$）经过特意挑选，使方程左边化为只含 z 和 z' 的式子。这样便得到一个关于 z 的一阶线性方程（first-order linear equation），然后再得到一个关于 y 的一阶线性方程；两者都可以用积分因子（integrating factor）解决。

解答

降阶

设 $z = y' - y$，则 $z' = y'' - y'$。我们把方程左边相应地拆开：

$$y'' + 2y' - 3y = (y'' - y') + 3(y' - y) = z' + 3z.$$

于是方程变成

$$\frac{\mathrm{d}z}{\mathrm{d}x} + 3z = 2\mathrm{e}^x.$$

由初始条件（initial conditions）得 $z(0) = y'(0) - y(0) = 2 - 1 = 1$。

> ? **想一想** 你会怎样想到 $(y'' - y') + 3(y' - y)$ 这个拆法?

我们需要 $y'' + 2y' - 3y$ 等于 $z' + kz = y'' + (k-1)y' - ky$，其中 k 为某个常数。比较 y' 项得 $k = 3$，而此时 $-ky = -3y$ 亦吻合。（这个做法之所以可行，是因为 1 是辅助方程（auxiliary equation）$m^2 + 2m - 3 = (m-1)(m+3) = 0$ 的一个根；另一个根 -3 正是 $z' + 3z$ 中出现的数。）

求 z

乘以积分因子 e^{3x}，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(z\mathrm{e}^{3x}\right) = 2\mathrm{e}^{4x} \quad \Longrightarrow \quad z\mathrm{e}^{3x} = \tfrac{1}{2}\mathrm{e}^{4x} + A.$$

当 $x = 0$ 时，$1 = \frac{1}{2} + A$，所以 $A = \frac{1}{2}$，而

$$z = \tfrac{1}{2}\left(\mathrm{e}^{x} + \mathrm{e}^{-3x}\right).$$

求 y

现在 $y' - y = z$，即

$$\frac{\mathrm{d}y}{\mathrm{d}x} - y = \tfrac{1}{2}\left(\mathrm{e}^{x} + \mathrm{e}^{-3x}\right).$$

积分因子是 e^{-x}：

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y\mathrm{e}^{-x}\right) = \tfrac{1}{2} + \tfrac{1}{2}\mathrm{e}^{-4x},$$

$$y\mathrm{e}^{-x} = \tfrac{1}{2}x - \tfrac{1}{8}\mathrm{e}^{-4x} + B.$$

当 $x = 0$ 时，$1 = -\frac{1}{8} + B$，所以 $B = \frac{9}{8}$。各项乘以 e^{x}，

$$\boxed{y = \left(\frac{x}{2} + \frac{9}{8}\right)\mathrm{e}^{x} - \frac{1}{8}\mathrm{e}^{-3x} = \frac{(4x+9)\mathrm{e}^{x} - \mathrm{e}^{-3x}}{8}}$$

留意我们在这里只用了 $y(0) = 1$；条件 $y'(0) = 2$ 在求 $z(0)$ 时已经用过，所以两个条件都已用上。

验算

- $y(0) = (9-1)/8 = 1$，而由 $y' = \left((4x+13)\mathrm{e}^{x} + 3\mathrm{e}^{-3x}\right)/8$ 得 $y'(0) = 16/8 = 2$。
- 代入验证：e^{-3x} 是齐次方程（homogeneous equation）的解（$9 - 6 - 3 = 0$），e^{x} 也是（$1 + 2 - 3 = 0$）。对于 $x\mathrm{e}^{x}$，$y'' + 2y' - 3y = (x+2) + 2(x+1) - 3x = 4$ 乘以 e^{x}，所以 $\frac{1}{2}x\mathrm{e}^{x}$ 刚好贡献 $2\mathrm{e}^{x}$。
- 数值上，用龙格—库塔法（Runge–Kutta method）由 $x = 0$ 起逐步计算原方程，得 $y(1) \approx 4.41098$，与 $(13\mathrm{e} - \mathrm{e}^{-3})/8$ 相符。

常见错误

- 在最后再用一次 $y'(0) = 2$，或者忘记把它转换成 $z(0) = 1$。
- 改用“其他方法”而尝试特别积分（particular integral）$c\,\mathrm{e}^{x}$：由于 e^{x} 本身已是齐次方程的解，这样做行不通，需要改用 $cx\mathrm{e}^{x}$。代换法会自动处理这个问题。

STEP 之外

这个代换其实是把微分算子（differential operator）作因式分解：以 D 表示 $\mathrm{d}/\mathrm{d}x$，方程就是 $(D+3)(D-1)y = 2\mathrm{e}^x$，而 $z = (D-1)y$。这样把算子分解，可以解释常系数线性方程的整套理论，包括为何重根或“共振”（resonant）根会带来额外的因子 x。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）对此有系统的阐述。

详细解答

STEP I 样本试卷，第 4 题

纯数学 · 代换积分法

STEP I 样本试卷，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/spec-s1-q4

关键思路

每当见到二次式的平方根，应先配方（complete the square）。这里根号内的二次式化为 $(x+1)^2-3$，而根号外的因子亦是 $x+1$，所以平移至 $u=x+1$ 后，便得到简洁的形式 $\dfrac{1}{u\sqrt{u^2-3}}$。这种形状（一个变量乘以“其平方减去常数”的平方根）正是正割代换（secant substitution）的用武之地，用倒数代换（reciprocal substitution）$t=1/u$ 亦同样可行。两种代换都能把无限的积分范围化为有限的范围。

解答

配方

我们有

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

当 $x\geq 1$ 时此式为正，因为此时 $(x+1)^2\geq 4>3$，所以被积函数（integrand）在整个 $[1,\infty)$ 上都有定义。设 $u=x+1$，则 $\mathrm{d}u=\mathrm{d}x$，当 $x=1$ 时 $u=2$，而当 $x\to\infty$ 时 $u\to\infty$：

$$I=\int_1^\infty \frac{\mathrm{d}x}{(x+1)\sqrt{x^2+2x-2}}=\int_2^\infty \frac{\mathrm{d}u}{u\sqrt{u^2-3}}.$$

当 u 很大时，被积函数的表现近似 $1/u^2$，所以这个积分收敛（converges）。

正割代换

设 $u=\sqrt{3}\sec\theta$，其中 $0\leq\theta<\frac{\pi}{2}$。则

$$\mathrm{d}u=\sqrt{3}\sec\theta\tan\theta\,\mathrm{d}\theta,\qquad \sqrt{u^2-3}=\sqrt{3(\sec^2\theta-1)}=\sqrt{3}\tan\theta,$$

这里我们用了在此范围内 $\tan\theta \geq 0$ 这一事实，从而去掉绝对值（modulus）符号。

> **想一想** 新的积分上下限是甚么？

在 $u = 2$ 时，我们需要 $\sec\theta = \frac{2}{\sqrt{3}}$，即 $\cos\theta = \frac{\sqrt{3}}{2}$，所以 $\theta = \frac{\pi}{6}$。当 $u \to \infty$ 时，$\sec\theta \to \infty$，所以 $\theta \to \frac{\pi}{2}$（从下方趋近）。这个代换在 $[\frac{\pi}{6}, \frac{\pi}{2})$ 上是递增的，所以上下限的次序保持不变。

现在所有东西都约去了：

$$\begin{aligned} I &= \int_{\pi/6}^{\pi/2} \frac{\sqrt{3}\sec\theta\tan\theta}{\sqrt{3}\sec\theta\cdot\sqrt{3}\tan\theta}\,\mathrm{d}\theta \\ &= \frac{1}{\sqrt{3}}\int_{\pi/6}^{\pi/2}\mathrm{d}\theta = \frac{1}{\sqrt{3}}\left(\frac{\pi}{2}-\frac{\pi}{6}\right) = \frac{\pi}{3\sqrt{3}}. \end{aligned}$$

因此

$$\int_1^\infty \frac{\mathrm{d}x}{(x+1)\sqrt{x^2+2x-2}} = \frac{\pi}{3\sqrt{3}} = \frac{\sqrt{3}\,\pi}{9}.$$

另一做法：倒数代换

假如一时想不起正割，$t = 1/u$ 同样奏效，而且立刻把积分范围变为有限。取 $u = 1/t$，则 $\mathrm{d}u = -\mathrm{d}t/t^2$，而且对于 $t > 0$

$$u\sqrt{u^2-3} = \frac{1}{t}\cdot\frac{\sqrt{1-3t^2}}{t} = \frac{\sqrt{1-3t^2}}{t^2}.$$

上下限 $u = 2$ 和 $u \to \infty$ 变为 $t = \frac{1}{2}$ 和 $t \to 0$，而负号把两者的次序对调：

$$I = \int_0^{1/2}\frac{\mathrm{d}t}{\sqrt{1-3t^2}} = \left[\frac{1}{\sqrt{3}}\arcsin\left(\sqrt{3}\,t\right)\right]_0^{1/2} = \frac{1}{\sqrt{3}}\arcsin\frac{\sqrt{3}}{2} = \frac{1}{\sqrt{3}}\cdot\frac{\pi}{3},$$

与第一种方法的结果一致。

验算

- 对原函数（antiderivative）$F(x) = \frac{1}{\sqrt{3}}\operatorname{arcsec}\left(\frac{x+1}{\sqrt{3}}\right)$ 求导，便可还原出被积函数；于是 $F(\infty) - F(1) = \frac{1}{\sqrt{3}}\left(\frac{\pi}{2} - \frac{\pi}{6}\right)$。
- 粗略估计大小：在 $[1,\infty)$ 上，被积函数比 $1/(x+1)^2$ 稍大，而后者的积分是 $\frac{1}{2}$。我们的答案是 $\pi/(3\sqrt{3}) \approx 0.6046$，略高于 $\frac{1}{2}$，正如所料。
- 对原积分作数值积分（numerical quadrature）（把范围分段，并以 $x = 1/w$ 消去无限长的尾段），得 0.604600，与 $\pi/(3\sqrt{3}) = 0.6045998\ldots$ 吻合。

常见错误

- 忘记更改上下限：下限是 $\theta = \frac{\pi}{6}$，而不是 0；用 0 会得出错误答案 $\pi/(2\sqrt{3})$。
- 写下 $\sqrt{\sec^2\theta - 1} = \tan\theta$，却没有检查 $\tan\theta$ 在所用范围内的正负号。
- 把答案留作 $\frac{1}{\sqrt{3}}\operatorname{arcsec}(\ldots)$“在无穷远处”取值，而没有正确地取极限。

STEP 之外

这是一个广义积分（improper integral）：严格来说，它的意思是 $\lim_{X\to\infty}\int_1^X$，而大学一年级的分析课程会精确说明这类极限何时存在（与 $\int^\infty u^{-2}\,\mathrm{d}u$ 作比较是标准的判别法）。这里的代换亦是一套一般理论的起点：对于有理函数（rational function）R，求 $R\left(x, \sqrt{ax^2+bx+c}\right)$ 的积分，总可以化为求有理函数的积分。关于广义积分和比较判别法（comparison test），可参阅 Abbott, *Understanding Analysis*。

详细解答

STEP I 样本试卷，第 5 题

纯数学 · 数字计数与阶乘

> **STEP I 样本试卷，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/spec-s1-q5

关键思路

两部分考的都是仔细计数，而不是计算。在 (i) 中，先把数字按区段计数（一位数、两位数、三位数），准确找出第 1000 个数字（digit）落在哪里，然后逐个数位计数 7 的数目：个位、十位、百位。在 (ii) 中，每个末尾零（trailing zero）都对应一个因数 $10 = 2 \times 5$，而 5 比 2 稀少，所以只需计数 365! 中因数 5 的个数。

解答

第 (i) 部分

在哪里停笔？ 数 1 至 9 用了 9 个数字。90 个两位数 10 至 99 用了 180 个数字，至此合共 189 个。于是还剩下

$$1000 - 189 = 811 = 3 \times 270 + 1$$

个数字。首 270 个三位数是 $100, 101, \ldots, 369$，它们用了 810 个数字。最后一个数字是 370 的首位数字，即 3。所以我们完整写出了 $1, 2, \ldots, 369$，之后再写了单独一个 3。

> ? **想一想** 370 这个数里面有一个 7。它算不算在内？

不算。我们写完它的首位数字 3 便停笔，所以它的 7 根本没有写出来。这正是我们必须把停笔的位置准确定到某一个数字的原因。

计数 1 至 369 中的 7。我们按位值（place value）计数。每个 7 都恰好在一个数位上，所以不会重复计数；像 77 这样的数会贡献两个，这是正确的。

▶ *个位数字。*在每个完整的十个数区段 $10k, \ldots, 10k + 9$ 中，恰好有一个数以 7 结尾。0

至 359 这些数组成 36 个完整区段，给出 36 个 7：7, 17, . . . , 357。不完整的区段 360 至 369 再多添一个，即 367。合计：37。

- 十位数字。在每个完整的一百个数之中，恰好有十个数的十位数字是 7：即 70–79、170–179 和 270–279。区段 370–379 已超出 369。合计：30。
- 百位数字。不超过 369 的每个数，百位数字最大是 3（或者根本没有百位数字）。合计：0。

最后那个单独的 3 没有任何贡献。因此数字 7 共用了

$$37 + 30 + 0 = 67 \text{ 次。}$$

第 (ii) 部分

N 在十进制（base 10）下末尾零的个数，是满足 $10^k \mid N$ 的最大的 k。由于 $10 = 2 \times 5$，这个数等于 N 的质因数分解（prime factorisation）中 2 的幂次与 5 的幂次两者中较小的一个。

计数 5 的个数。整除 365! 的 5 的幂次，等于对 $j = 1, \ldots, 365$ 把 j 中因数 5 的个数加起来的总和。5 的倍数至少贡献一个因数，25 的倍数贡献第二个，125 的倍数贡献第三个，如此类推。逐层分开计数：

$$\begin{aligned}
\left\lfloor \tfrac{365}{5} \right\rfloor &= 73 && \text{（5 的倍数），} \\
\left\lfloor \tfrac{365}{25} \right\rfloor &= 14 && \text{（25 的倍数），} \\
\left\lfloor \tfrac{365}{125} \right\rfloor &= 2 && \text{（125 的倍数：125, 250），} \\
\left\lfloor \tfrac{365}{625} \right\rfloor &= 0. &&
\end{aligned}$$

一个恰好被 5^m 整除的数，在首 m 行中各被计数一次，所以它的贡献恰好是 m，正合所需。因此 365! 中 5 的幂次是

$$73 + 14 + 2 = 89.$$

2 并不是瓶颈。对 2 作同样的论证，可知因数 2 至少有 $\lfloor 365/2 \rfloor = 182 > 89$ 个。（事实上共有 $182 + 91 + 45 + 22 + 11 + 5 + 2 + 1 = 359$ 个。）所以 $2^{89} \cdot 5^{89} = 10^{89}$ 整除 365!，但 10^{90} 不能，因为 $5^{90} \nmid 365!$。

因此 365! 的末尾恰好有 **89** 个零。

验算

对于 (i)，可以快速检查停笔的位置：$9 + 180 + 3 \times 270 = 999$ 个数字刚好写到 369 的末尾，而下一个数字就是 370 的首位数字 3。用电脑写出首 1000 个数字再计数，可确

认共有 67 个 7，其中 20 个在 1–99 之中，47 个在 100–369 之中。对于 (ii)，准确计算 365!（它有 779 位数字），再计数它末尾的零。结果是 89 个，与 5 的个数吻合。可以用一个小例子亲手检验这个方法：25! 的末尾应有 $\lfloor 25/5 \rfloor + \lfloor 25/25 \rfloor = 6$ 个零，而事实上 $25! = 15\,511\,210\,043\,330\,985\,984\,000\,000$。

常见错误

- 把 370（甚至 371–379）当作已经写出来，于是得出 68 或更大的答案。
- 只用 $\lfloor 365/5 \rfloor = 73$，忘记了 25、50、… 含有两个 5，而 125、250 含有三个 5。
- 按数而不是按数字来计数 7，以致 77、177 和 277 只计数了一次而不是两次。

STEP 之外

(ii) 中的逐层计数就是勒让德公式（Legendre's formula）：质数 p 在 $n!$ 中的幂次是 $\sum_{k\geq 1}\lfloor n/p^k \rfloor$，它亦等于 $(n - s_p(n))/(p-1)$，其中 $s_p(n)$ 是 n 在 p 进制下各位数字之和。这里 $365 = 2430_5$，数字和是 9，而 $(365-9)/4 = 89$。这条公式是库默尔定理（Kummer's theorem）的基础，该定理讨论二项式系数（binomial coefficients）的质因数。Davenport 的 *The Higher Arithmetic*（《高等算术》）是进一步探索这类初等数论（elementary number theory）的好去处。

详细解答

STEP I 样本试卷，第 6 题

纯数学 • 弓形的面积

> **STEP I 样本试卷，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/spec-s1-q6

关键思路

这枚硬币是一个正七边形（regular heptagon），每条边外面各凸出一个圆弧形的“帽子”。因此它的面积等于七边形的面积加上七个弓形（circular segment）的面积，而每个弓形就是一个扇形（sector）减去一个三角形。于是整个问题归结为两个角：每段弧在其圆心所对的角（用圆周角定理（inscribed angle theorem）立刻可得），以及七边形的长对角线与其外接圆（circumcircle）半径之间的夹角。

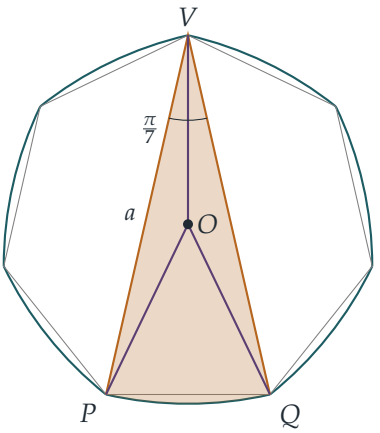
解答

形状

每个顶点（vertex）与它对面那段弧的两个端点的距离都是 a（两个端点都在以该顶点为圆心、半径为 a 的圆上）。所以七条“长”对角线（diagonal），即每条由一个顶点连到沿周界数过去第三个顶点的线段，长度全是 a。半径相等的等弧在各自圆心所对的角相等，所以七条边（即各段弧的弦（chord））也都相等。

把长对角线依次连起来，便得到一个各边长度均为 a 的七角星。它在每个尖角处的角，就是由该顶点出发的两条对角线之间的夹角；而这两条对角线连同对面的那条边，组成一个三角形，三边分别是 a、a 和共同的边长。这样的七个三角形全部全等（congruent），所以这个星形不但等边，而且等角，因而是正星形。由此可知，七个顶点就是一个正七边形的顶点，设其中心为 O，外接圆半径（circumradius）为 R。

把其中一个顶点记为 V，并设 P 和 Q 是以 V 为圆心的那段弧的两个端点；P 和 Q 是七边形的相邻顶点。



每段弧所对的角

边 PQ 在中心 O 所对的角是 $2\pi/7$。由于 V 在外接圆上，并且位于优弧（major arc）PQ 上，由圆周角定理得

$$\angle PVQ = \tfrac{1}{2} \cdot \frac{2\pi}{7} = \frac{\pi}{7}.$$

所以每段弧都是半径为 a 的圆的一部分，在其圆心所对的角为 $\pi/7$，而扇形 VPQ（阴影部分）的面积是

$$\tfrac{1}{2}a^2 \cdot \frac{\pi}{7} = \frac{\pi a^2}{14}.$$

把面积加起来

硬币的币面由七边形和七个弓形组成，每条边上各有一个弓形。七边形可分成七个像 OPQ 那样的三角形，而 PQ 上的弓形就是扇形 VPQ 减去三角形 VPQ。因此

$$\begin{aligned} \text{面积} &= 7\,[OPQ] + 7\big(\text{扇形}VPQ - [VPQ]\big) \\ &= \frac{\pi a^2}{2} - 7\big([VPQ] - [OPQ]\big), \end{aligned}$$

其中 $[\,\cdot\,]$ 表示面积。

想一想 怎样才能不必分别计算两个三角形，就求出 $[VPQ] - [OPQ]$?

中心 O 位于三角形 VPQ 之内，并在它的对称轴上，所以从三角形 VPQ 中除去三角形 OPQ，余下的就是两个全等三角形 OVP 和 OVQ：

$$[VPQ] - [OPQ] = [OVP] + [OVQ] = 2 \cdot \tfrac{1}{2}\, R\, a \sin \angle OVP.$$

根据对称性，VO 平分 $\angle PVQ$，所以 $\angle OVP = \pi/14$。三角形 OVP 是等腰三角形（isosceles

triangle），其中 $OV = OP = R$，所以两个底角都是 $\pi/14$；由 O 向 VP 作垂线，便得

$$a = VP = 2R\cos\frac{\pi}{14}, \qquad \text{所以} \qquad R = \frac{a}{2\cos(\pi/14)}.$$

因此

$$[VPQ] - [OPQ] = R\,a\sin\frac{\pi}{14} = \frac{a^2\sin(\pi/14)}{2\cos(\pi/14)} = \frac{a^2}{2}\tan\frac{\pi}{14},$$

而

$$\text{面积} = \frac{\pi a^2}{2} - \frac{7a^2}{2}\tan\frac{\pi}{14} = \frac{a^2}{2}\left(\pi - 7\tan\frac{\pi}{14}\right),$$

这正是所要证明的结果。

验算

- 数值上，$\frac{1}{2}(\pi - 7\tan\frac{\pi}{14}) \approx 0.7719$，所以面积约为 $0.772a^2$。在七段弧上密集取点，再用鞋带公式（shoelace formula）计算，结果吻合至小数点后六位；用蒙地卡罗（Monte Carlo）方法估算，结果也一样（一个点在硬币之内，当且仅当它与全部七个顶点的距离都不超过 a）。
- 这枚硬币在每个方向上的宽度都是 a，就像一个直径为 a 的圆盘，而圆盘的面积是 $\pi a^2/4 \approx 0.785a^2$。在所有定宽为 a 的图形之中，以圆盘的面积最大，所以我们预期答案会略小一些，而 $0.772a^2 < 0.785a^2$，正合预期。
- 把同样的论证用于勒洛（Reuleaux）n 边形（n 为奇数），可得 $\frac{a^2}{2}(\pi - n\tan\frac{\pi}{2n})$；当 $n \to \infty$ 时，$n\tan\frac{\pi}{2n} \to \frac{\pi}{2}$，面积趋向 $\pi a^2/4$，即圆盘的面积。取 $n = 3$ 则得 $\frac{1}{2}(\pi - \sqrt{3})a^2$，就是大家熟悉的勒洛三角形（Reuleaux triangle）。

常见错误

- 把弧所对的角当作 $2\pi/7$（在中心 O 所对的角），而不是 $\pi/7$（在对面顶点所对的角，那里才是该段弧真正的圆心）。
- 把 a 与外接圆半径 R 混淆：a 是长对角线，$a = 2R\cos(\pi/14)$。

STEP 之外

50 便士硬币是一条定宽曲线（curve of constant width）：每一对互相平行的切线之间的距离都同样是 a，所以无论硬币怎样转动，投币口都能用同一方法量度它。巴比尔定理（Barbier's theorem）指出，每一条这样的曲线的周界都是 πa（这里是 $7 \cdot a \cdot \pi/7$），而布拉施克–勒贝格定理（Blaschke–Lebesgue theorem）则指出，其中以勒洛三角形的面积

最小。想进一步了解这类几何，可参阅柯朗（Courant）与罗宾斯（Robbins）的 *What Is Mathematics?*，书中有一节谈定宽曲线，写得十分精彩。

详细解答

STEP I 样本试卷，第 7 题

纯数学 · 复指数级数

STEP I 样本试卷，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/spec-s1-q7

关键思路

每当见到 $1+\mathrm{e}^{\mathrm{i}\theta}$，就抽出“半角”因子 $\mathrm{e}^{\mathrm{i}\theta/2}$：余下的是 $\mathrm{e}^{-\mathrm{i}\theta/2}+\mathrm{e}^{\mathrm{i}\theta/2}$，它是实数。至于那个级数，二项式系数（binomial coefficients）$\frac{n!}{r!(n-r)!}=\binom{n}{r}$ 提示我们使用二项式定理（binomial theorem）。因此我们把每个正弦写成一个指数的虚部（imaginary part），于是该和就变成 $1+\mathrm{e}^{2\mathrm{i}\alpha}$ 的某次幂的虚部，而第一部分已经把它化成极式（polar form）。

解答

模与辐角

抽出因子 $\mathrm{e}^{\mathrm{i}\alpha}$，

$$\begin{aligned}1+\mathrm{e}^{2\mathrm{i}\alpha} &= \mathrm{e}^{\mathrm{i}\alpha}\left(\mathrm{e}^{-\mathrm{i}\alpha}+\mathrm{e}^{\mathrm{i}\alpha}\right)\\ &= \mathrm{e}^{\mathrm{i}\alpha}\left[(\cos\alpha-\mathrm{i}\sin\alpha)+(\cos\alpha+\mathrm{i}\sin\alpha)\right]\\ &= 2\cos\alpha\,\mathrm{e}^{\mathrm{i}\alpha}.\end{aligned}$$

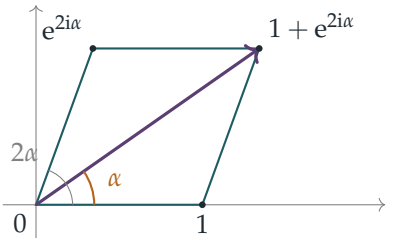
想一想 我们现在可以直接读出“模（modulus）为 $2\cos\alpha$，辐角（argument）为 α”吗？对 α 的限制有甚么用？

$r\,\mathrm{e}^{\mathrm{i}\theta}$ 这个形式，只有在 $r>0$ 时才给出模 r 和辐角 θ。这里 $-\frac{1}{2}\pi<\alpha<\frac{1}{2}\pi$，所以 $\cos\alpha>0$，而且这个范围也落在主值区间 $(-\pi,\pi]$ 之内。因此

$$\left|1+\mathrm{e}^{2\mathrm{i}\alpha}\right|=2\cos\alpha,\qquad \arg\left(1+\mathrm{e}^{2\mathrm{i}\alpha}\right)=\alpha.$$

（假如 $\cos\alpha$ 是负数，模就会是 $-2\cos\alpha$，而辐角则会相差 π。）

这里还有一个漂亮的几何图像。点 0、1 和 $1+\mathrm{e}^{2\mathrm{i}\alpha}$ 构成一个边长为 1 的菱形，余下的一个顶点是 $\mathrm{e}^{2\mathrm{i}\alpha}$。由 0 出发的对角线平分角 2α，长度为 $2\cos\alpha$。



级数

由于对实数 θ 有 $\sin\theta = \operatorname{Im}\mathrm{e}^{\mathrm{i}\theta}$，而二项式系数都是实数，

$$\begin{aligned} S &= \sum_{r=0}^{n}\binom{n}{r}\sin(2r+1)\alpha = \operatorname{Im}\sum_{r=0}^{n}\binom{n}{r}\mathrm{e}^{\mathrm{i}(2r+1)\alpha} \\ &= \operatorname{Im}\left[\mathrm{e}^{\mathrm{i}\alpha}\sum_{r=0}^{n}\binom{n}{r}\left(\mathrm{e}^{2\mathrm{i}\alpha}\right)^r\right] = \operatorname{Im}\left[\mathrm{e}^{\mathrm{i}\alpha}\left(1+\mathrm{e}^{2\mathrm{i}\alpha}\right)^n\right], \end{aligned}$$

其中先用了 $\mathrm{e}^{\mathrm{i}(2r+1)\alpha} = \mathrm{e}^{\mathrm{i}\alpha}\left(\mathrm{e}^{2\mathrm{i}\alpha}\right)^r$，再用二项式定理。现在利用第一部分的结果，配合棣美弗定理（de Moivre's theorem）$\left(\mathrm{e}^{\mathrm{i}\alpha}\right)^n = \mathrm{e}^{\mathrm{i}n\alpha}$：

$$\mathrm{e}^{\mathrm{i}\alpha}\left(1+\mathrm{e}^{2\mathrm{i}\alpha}\right)^n = \mathrm{e}^{\mathrm{i}\alpha}\,2^n\cos^n\alpha\,\mathrm{e}^{\mathrm{i}n\alpha} = 2^n\cos^n\alpha\,\mathrm{e}^{\mathrm{i}(n+1)\alpha}.$$

因为 $2^n\cos^n\alpha$ 是实数，取虚部便得

$$\boxed{\sum_{r=0}^{n}\frac{n!}{r!\,(n-r)!}\sin(2r+1)\alpha = 2^n\cos^n\alpha\,\sin(n+1)\alpha.}$$

因式分解 $1+\mathrm{e}^{2\mathrm{i}\alpha} = 2\cos\alpha\,\mathrm{e}^{\mathrm{i}\alpha}$ 对每一个实数 α 都成立。我们只是在指明模和辐角时才需要 $\cos\alpha > 0$，所以这条求和公式对所有实数 α 都成立，而不限于题目所给的范围。

验算

- $n=0$：两边都是 $\sin\alpha$。$n=1$：根据和化积公式（sum-to-product formula），左边是 $\sin\alpha+\sin 3\alpha = 2\sin 2\alpha\cos\alpha$，与 $2\cos\alpha\sin 2\alpha$ 吻合。
- $n=3$，$\alpha=\frac{\pi}{6}$：左边是 $\frac{1}{2}+3+\frac{3}{2}-\frac{1}{2}=\frac{9}{2}$，右边是 $8\left(\frac{\sqrt{3}}{2}\right)^3\cdot\frac{\sqrt{3}}{2}=\frac{9}{2}$。
- $\alpha=\frac{1}{2}\pi$：每一项 $\sin(2r+1)\frac{\pi}{2}=(-1)^r$，所以当 $n\geq 1$ 时，和为 $(1-1)^n=0$，与 $\cos\frac{\pi}{2}=0$ 吻合。
- 我们以数值方法验证过，对 $0\leq n\leq 24$ 及随机选取的 α 值，公式都成立。

常见错误

- 把模写成 $2\cos\alpha$，却没有指出 α 的范围保证它是正数。模不可能是负数。
- 过早取虚部。$\mathrm{Im}(zw) \neq \mathrm{Im}\, z \cdot \mathrm{Im}\, w$，所以要一直保留复数形式，到最后一步才取虚部。
- 把 $\mathrm{e}^{\mathrm{i}(2r+1)\alpha}$ 写成 $\left(\mathrm{e}^{2\mathrm{i}\alpha}\right)^r$，漏掉了额外的因子 $\mathrm{e}^{\mathrm{i}\alpha}$。这样会令 $\sin(n+1)\alpha$ 变成 $\sin n\alpha$。

STEP 之外

同一个想法，即把 $\sum \binom{n}{r}\mathrm{e}^{\mathrm{i}r\theta}$ 求和成闭合形式（closed form），也出现在傅里叶分析（Fourier analysis）之中。在那里，$\cos^n$ 类型的式子会展开成倍角之和，或由倍角之和收拢而成，而二项式权重 $\binom{n}{r}/2^n$ 就成为二项分布（binomial distribution）。抽出半角因子也是求 $1-\mathrm{e}^{\mathrm{i}\theta}$ 的模的标准方法，你会在等比级数（geometric series）和单位根（roots of unity）中遇到它。Needham 的 *Visual Complex Analysis* 把这些运算的几何观点阐述得十分精彩，当中也包括那幅菱形图。

详细解答

STEP I 样本试卷，第 8 题

纯数学 · 积分方程

STEP I 样本试卷，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：

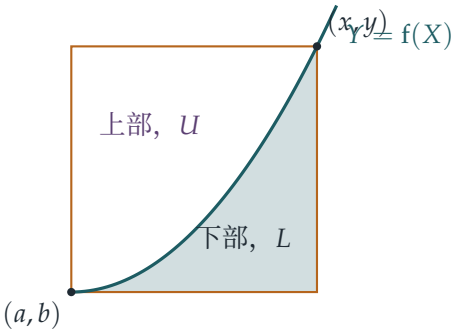
step.maths.org/questions/spec-s1-q8

关键思路

面积条件对 (x,y) 的每一个位置都成立，所以它其实是两个以 x 为自变量的函数之间的等式：一边是 f 的积分，另一边则涉及 $\mathrm{f}(x)$ 本身。对 x 求导便可消去积分，得出一个一阶微分方程（first-order differential equation），这个方程可以分离变量（或者更简洁地，化为“某个商的导数等于零”）。

解答

由于 $\mathrm{f}'(x) \geq 0$，曲线由矩形的一角 (a,b) 上升至对角 (x,y)，所以矩形的阔度为 $x-a$，高度为 $y-b=\mathrm{f}(x)-b\geq 0$。下部是曲线与底边 $Y=b$ 之间的区域（下图阴影部分）；上部则是余下的部分。



以 t 作为积分变量（因为 x 现在是积分上限），

$$L=\int_a^x \big(\mathrm{f}(t)-b\big)\,\mathrm{d}t, \qquad L+U=(x-a)\big(\mathrm{f}(x)-b\big).$$

条件 $L = \frac{1}{2}U$ 即是 $U = 2L$，所以整个矩形的面积为 $3L$：

$$3\int_a^x \big(\mathrm{f}(t) - b\big)\,\mathrm{d}t = (x-a)\big(\mathrm{f}(x) - b\big) \qquad \text{对所有} x \geq a. \tag{$*$}$$

想一想 怎样由一条含有积分的等式，得出一条可以解出 f 的方程?

由于 $(*)$ 对每一个 $x \geq a$ 都成立，两边是同一个以 x 为自变量的函数，所以它们的导数相等。函数 f 可微（differentiable）（题目提到 f′），因而连续（continuous），所以根据微积分基本定理（fundamental theorem of calculus），左边的导数是 $3(\mathrm{f}(x) - b)$。用积法则（product rule）对右边求导，

$$3\big(\mathrm{f}(x) - b\big) = \big(\mathrm{f}(x) - b\big) + (x-a)\,\mathrm{f}'(x),$$

即

$$(x-a)\,\mathrm{f}'(x) = 2\big(\mathrm{f}(x) - b\big).$$

为了整理这条式，设 $g(x) = \mathrm{f}(x) - b$ 及 $u = x - a$，于是 $g' = \mathrm{f}'$，方程变成 $u\,g' = 2g$。我们可以分离变量（separate variables），得 $g'/g = 2/u$，但这样要除以 g，而它有可能等于零。更干净利落的做法是留意到，当 $x > a$（即 $u > 0$）时，

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{g}{u^2}\right) = \frac{u^2 g' - 2u\,g}{u^4} = \frac{u\,g' - 2g}{u^3} = 0.$$

所以 g/u^2 是一个常数，设为 k，在 $x > a$ 上成立，于是

$$\mathrm{f}(x) - b = k(x-a)^2 \qquad (x > a),$$

而这在 $x = a$ 时同样成立，因为 (a, b) 在曲线上。由于 $\mathrm{f}(x) \geq b$ 对 $x \geq a$ 成立（函数由 $\mathrm{f}(a) = b$ 开始非递减），我们有 $k \geq 0$，而且 $k > 0$，除非曲线是水平直线 $y = b$；在这情况下每个矩形都退化成面积为零，条件变得空洞。

因此曲线是 $y - b = k(x-a)^2$，其中 $x \geq a$：这是以 $x = a$ 为铅垂对称轴、顶点（vertex）位于 (a, b) 的抛物线（parabola）的一半。由于 $x \geq a$，我们只得到右边的一支，所以它是该抛物线的一段弧，正如题目所要求。

反过来，这条曲线确实满足条件：$L = \int_a^x k(t-a)^2\,\mathrm{d}t = \frac{1}{3}k(x-a)^3$，而矩形的面积为 $(x-a)\cdot k(x-a)^2 = k(x-a)^3$，所以 $U = \frac{2}{3}k(x-a)^3 = 2L$。

验算

取 $a = b = 0$、$k = 1$ 及 $x = 3$：矩形为 3×9，面积为 27，而 $\int_0^3 t^2\,\mathrm{d}t = 9$，所以两部分的面积分别为 9 和 18，比为 1 : 2。更一般地，对于 $y - b = k(x-a)^p$，下部与上部之比为 $1/p$（我们用辛普森法则（Simpson's rule）对 $p = \frac{1}{2}, 1, \frac{3}{2}, 2, 3$ 作了数值验算），所以比值 $\frac{1}{2}$ 正好对应 $p = 2$；直线（$p = 1$）把矩形平分，理应如此。

常见错误

- 把 x 同时用作积分上限和积分变量，令求导一步变得混乱。
- 把两部分调转，得出 $U = \frac{1}{2}L$，结果得到 $(x-a)\mathrm{f}' = \frac{1}{2}(\mathrm{f}-b)$ 及曲线 $y - b = k\sqrt{x-a}$，那是一条对称轴为水平的抛物线，方向不对。
- 不加说明便除以 $\mathrm{f}(x) - b$；上面利用商的论证避开了这个问题。

STEP 之外

方程 $(*)$ 是一条*积分方程*（integral equation）：未知函数出现在积分之内。通过求导把它化为微分方程，并记住原来的方程同时决定了初始值（initial value），是标准的第一步；而答案的唯一性，则建基于这个事实：在一个区间上导数为零的函数必为常数（这是中值定理（mean value theorem）的推论）。Abbott 的 *Understanding Analysis* 严谨地证明了微积分基本定理和中值定理；Boyce 与 DiPrima 的 *Elementary Differential Equations* 则涵盖了像 $u\,g' = 2g$ 这类方程。

详细解答

STEP I 样本试卷，第 9 题

纯数学 · 绝对值不等式

STEP I 样本试卷，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/spec-s1-q9

关键思路

两部分的重点都是去掉绝对值符号（modulus sign），方法是准确找出符号内的式子在哪里变号，或在哪里穿过某个水平。在 (i) 中，设 $u = x + 1$，式子便化为 $u - 2/u$，它在 $u = 0$ 的两侧各自递增，因此不等式（inequality）只是在每一分支上把 u 限制在某个二次方程的两个根之间。在 (ii) 中，把 $\cos x - \sin x$ 写成单一个余弦，便可看出它在哪里变号，然后在该处把积分（integral）分段。

解答

第 (i) 部分

要使式子有定义，必须有 $x \neq -1$。设 $u = x + 1$，则 $u \neq 0$，且

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

把它记作 $f(u)$。不等式即为 $-2 < f(u) < 2$。

? **想一想** 为甚么不直接把各项乘以 $x + 1$?

因为 $x + 1$ 可以是负数，而乘以负数会令不等号反向；不小心处理的话，正是丢失一半答案的典型原因。我们改为利用 f 的形状。它的导数（derivative）是 $f'(u) = 1 + 2/u^2 > 0$，所以 f 在 $u < 0$ 上严格递增（strictly increasing），在 $u > 0$ 上亦各自严格递增。在每一分支上，它都由 $-\infty$ 升至 $+\infty$：当 $u \to 0^-$ 时，$-2/u \to +\infty$；当 $u \to 0^+$ 时，$-2/u \to -\infty$；而当 $|u| \to \infty$ 时，$f(u) \sim u$。因此在每一分支上，f 取 -2 和 2 这两个值各恰好一次，而 $-2 < f < 2$ 正好在这两点之间成立。

对于 $u \neq 0$，乘以 u 可得两个等价的方程：

$$f(u) = 2 \iff u^2 - 2u - 2 = 0 \iff u = 1 \pm \sqrt{3},$$
$$f(u) = -2 \iff u^2 + 2u - 2 = 0 \iff u = -1 \pm \sqrt{3}.$$

分支 $u > 0$。正根是 $u = -1 + \sqrt{3}$（此处 $f = -2$）和 $u = 1 + \sqrt{3}$（此处 $f = 2$）。由于 f 递增，$-2 < f(u) < 2$ 恰好在 $-1 + \sqrt{3} < u < 1 + \sqrt{3}$ 时成立，即 $-2 + \sqrt{3} < x < \sqrt{3}$。

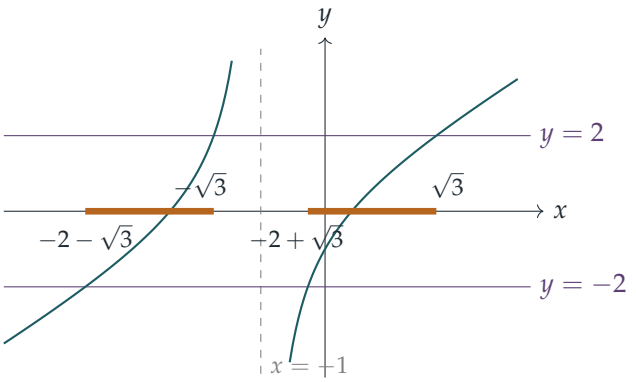
分支 $u < 0$。负根是 $u = -1 - \sqrt{3}$（此处 $f = -2$）和 $u = 1 - \sqrt{3}$（此处 $f = 2$）。同样由单调性（monotonicity）可知，$-2 < f(u) < 2$ 恰好在 $-1 - \sqrt{3} < u < 1 - \sqrt{3}$ 时成立，即 $-2 - \sqrt{3} < x < -\sqrt{3}$。

两个区间都不包含 $x = -1$（留意 $-\sqrt{3} < -1 < -2 + \sqrt{3}$），所以解集（solution set）是

$$\boxed{(-2 - \sqrt{3},\ -\sqrt{3}) \cup (-2 + \sqrt{3},\ \sqrt{3})}$$

每个端点都符合题目要求的 $a + b\sqrt{3}$ 形式。端点不包括在内，因为在该处绝对值恰好等于 2。

把 $y = x + (x-1)/(x+1)$ 的图像与直线 $y = \pm 2$ 画在一起，可见曲线位于两条直线之间的两个区间（以琥珀色标示）。



第 (ii) 部分

把两项合并成一个余弦：

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

当 $0 \leq x \leq \pi$ 时，有 $\pi/4 \leq x + \pi/4 \leq 5\pi/4$，而余弦在 $[\pi/4, \pi/2)$ 上为正，在 $(\pi/2, 5\pi/4]$ 上为负。因此当 $0 \leq x \leq \pi/4$ 时 $\cos x - \sin x \geq 0$，当 $\pi/4 \leq x \leq \pi$ 时则 ≤ 0。在 $\pi/4$ 处把积分分段，并利用 $\cos x - \sin x$ 的一个原函数（antiderivative）是

$\sin x + \cos x$：

$$\begin{aligned}\int_0^{\pi} |\cos x - \sin x|\,\mathrm{d}x &= \int_0^{\pi/4} (\cos x - \sin x)\,\mathrm{d}x - \int_{\pi/4}^{\pi} (\cos x - \sin x)\,\mathrm{d}x \\ &= \Big[\sin x + \cos x\Big]_0^{\pi/4} - \Big[\sin x + \cos x\Big]_{\pi/4}^{\pi} \\ &= \left(\sqrt{2} - 1\right) - \left(-1 - \sqrt{2}\right) \\ &= 2\sqrt{2}.\end{aligned}$$

验算

- 在 (i) 中抽查数值：当 $x = 0$ 时，式子的值是 -1，在绝对值的界限之内，而 $0 \in (-2+\sqrt{3}, \sqrt{3})$；当 $x = -2$ 时，值是 $-2+3 = 1$，而 $-2 \in (-2-\sqrt{3}, -\sqrt{3})$；当 $x = -1.5$ 时，值是 $-1.5+5 = 3.5$，在界限之外，而 $-1.5 \in (-\sqrt{3}, -1)$ 确实不在任何一个区间内。
- 在一个细密的网格上，以及在 $[-50, 50]$ 内的 2×10^5 个随机点上测试不等式，结果处处与这两个区间吻合，而在全部四个端点上，绝对值都等于 2。
- 在 (ii) 中，辛普森法则（Simpson's rule）给出 2.828427…，与 $2\sqrt{2} \approx 2.828427$ 相符；两段的值分别是 $\sqrt{2} - 1 \approx 0.4142$ 和 $\sqrt{2} + 1 \approx 2.4142$。作为合理性检查，被积函数（integrand）最大为 $\sqrt{2}$，所以积分最大为 $\sqrt{2}\,\pi \approx 4.44$。

常见错误

- 把 $x + \frac{x-1}{x+1} < 2$ 的两边乘以 $x+1$ 而没有考虑其正负，结果丢失了（或错误地包括了）$x = -1$ 左边的区间。
- 只解 $f < 2$，或只解 $f > -2$：绝对值不等式要求两者同时成立。
- 在 (ii) 中，直接在整个 $[0, \pi]$ 上对 $\cos x - \sin x$ 积分，得出 -2（绝对值的积分得出负数，是一个警号）。

STEP 之外

(i) 的窍门，是把一个代数不等式换成关于单调函数（monotonic function）的陈述：若 f 在某区间上严格递增，则 $a < f(u) < b$ 恰好在 $f = a$ 与 $f = b$ 的两点之间成立。要严格证明这一点，需要用到介值定理（intermediate value theorem），它是大学分析课程中最早出现的重要结果之一；而带绝对值的不等式，例如 $|x - a| < \varepsilon$，正是该课程的日常语言。Abbott 的 *Understanding Analysis*（一本分析学入门书）亲切易读，两者都可以在书

中见到。

详细解答

STEP I 样本试卷，第 10 题

力学 • 功率与阻力

STEP I 样本试卷，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：

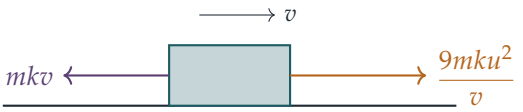
step.maths.org/questions/spec-s1-q10

关键思路

功率（power）等于力乘以速率，所以以恒定功率 P 运作的机车，其牵引力（driving force）为 P/v。由牛顿第二定律（Newton's second law）便得到一条关于 v 的一阶微分方程（first-order differential equation），而且可以分离变量（separation of variables）。关键在于选对加速度的形式：题目问时间时用 $\mathrm{d}v/\mathrm{d}t$，问距离时用 $v\,\mathrm{d}v/\mathrm{d}x$。

解答

把机车看成一个质量为 m 的质点（particle），在水平路轨上行驶，因此它的重量与法向反作用力（normal reaction）互相抵消，只需考虑水平方向的力。取运动方向为正。



最大速率

速率为 v 时，牵引力为 $9mku^2/v$，所以由牛顿第二定律得

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{9mku^2}{v} - mkv \quad\Longrightarrow\quad \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{k\,(9u^2 - v^2)}{v}.$$

当 $v < 3u$ 时，右边为正，所以机车会加速。只有在牵引力与阻力（resistance）平衡，即 $9u^2 = v^2$ 时，它才会停止加速。因此可能达到的最大速率为

$$v_{\max} = 3u.$$

想一想 只要 $v < 3u$，机车就一直在加速，那为甚么它到不了 $3u$？

当 $v \to 3u$ 时，加速度会减小至零。分离变量后，由速率 v_1 加速至速率 v_2 所需的时间为

$$t = \frac{1}{k}\int_{v_1}^{v_2} \frac{v}{9u^2 - v^2}\,\mathrm{d}v = \frac{1}{2k}\ln\left(\frac{9u^2 - v_1^2}{9u^2 - v_2^2}\right),$$

而当 $v_2 \to 3u$ 时，这个时间趋向无限大。所以 $3u$ 是终端速率（terminal speed）：机车的速率可以任意接近它，却永远不会在有限时间内真正达到它。（换个说法，$v = 3u$ 是这条微分方程的一个平衡点（equilibrium），而根据解的唯一性（uniqueness of solutions），从它下方出发的解不可能到达它。）

由 u 到 $2u$

在上面的公式中代入 $v_1 = u$ 和 $v_2 = 2u$：

$$t = \frac{1}{2k}\ln\left(\frac{9u^2 - u^2}{9u^2 - 4u^2}\right) = \frac{1}{2k}\ln\left(\frac{8u^2}{5u^2}\right) = \frac{1}{2k}\ln\left(\frac{8}{5}\right),$$

正是所求。

切断动力

现在只剩下两个阻力 mkv 和 mkB，两者都与运动方向相反。我们要求的是距离，所以把加速度写成 $v\,\mathrm{d}v/\mathrm{d}x$：

$$mv\frac{\mathrm{d}v}{\mathrm{d}x} = -mk(v + B) \quad\Longrightarrow\quad \frac{v}{v + B}\,\mathrm{d}v = -k\,\mathrm{d}x.$$

写成 $\dfrac{v}{v+B} = 1 - \dfrac{B}{v+B}$，然后积分，其中 v 由 $2u$ 下降至 0，而 x 由 0 增加至 D：

$$\begin{aligned} kD &= \int_0^{2u}\left(1 - \frac{B}{v+B}\right)\mathrm{d}v \\ &= \Big[v - B\ln(v+B)\Big]_0^{2u} \\ &= 2u - B\ln\left(\frac{2u+B}{B}\right). \end{aligned}$$

所以切断动力后行驶的距离为

$$D = \frac{1}{k}\left[2u - B\ln\left(1 + \frac{2u}{B}\right)\right].$$

这一次机车确实会在有限时间内停下：$\mathrm{d}v/\mathrm{d}t = -k(v+B) \le -kB < 0$，所以减速度（deceleration）永远不会小于 kB。这正是需要额外加上一个恒定阻力的原因。

验算

- 当 $B \to 0$ 时，$B\ln(1+2u/B) \to 0$，所以 $D \to 2u/k$。这与只有阻力 mkv 的情况吻合：此时 $v\,\mathrm{d}v/\mathrm{d}x = -kv$，即 v 随距离以 k 的速率线性下降。
- $D > 0$，因为当 $y > 0$ 时 $\ln(1+y) < y$，由此得 $B\ln(1+2u/B) < 2u$。而且 D 随 B 增大而减小，合乎预期。
- 数值上，取 $k = u = B = 1$，把运动方程逐步推算，得出停止距离为 0.90139，与 $2 - \ln 3 \approx 0.90139$ 吻合。由 $v = 1$ 起，把有动力的阶段逐步推算至时间 $\frac{1}{2}\ln\frac{8}{5} \approx 0.2350$，最后得 $v = 2.0000$。

常见错误

- 把牵引力当作常数。在恒定功率下，牵引力是 P/v，会随机车加速而减小。
- 在最后一部分使用 $\mathrm{d}v/\mathrm{d}t$。这样求得的是停止所需的时间，而不是距离。
- 声称机车“最终”会达到 $3u$。时间积分是发散（diverge）的，所以它永远达不到。

STEP 之外

速率 $3u$ 是自治方程（autonomous equation）$\dot{v} = f(v)$ 的一个稳定平衡点（stable equilibrium）。解会以指数方式趋近它，却永远不会与它相遇，因为性质良好的微分方程的解不可能相交或合并。这让你初次领略唯一性定理（uniqueness theorem）和相线分析（phase-line analysis），两者都是大学微分方程入门课程的内容。Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教材）对这两者都有讲解。想进一步了解功率和阻力，可参考 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）。

详细解答

STEP I 样本试卷，第 11 题

力学 · 摩擦力与滚动接触

STEP I 样本试卷，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/spec-s1-q11

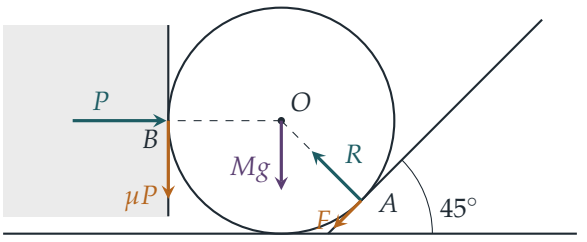
关键思路

在圆柱体即将移动的一刻，地面不再施力，因此只需考虑三个作用：重量、垂直面和斜面。“滚动”表示圆柱体在斜面上不打滑，所以它开始绕着与斜面的接触点转动。只要找出与垂直面接触的那一点此时如何运动，便会发现圆柱体必须沿垂直面滑动。因此，摩擦力（friction）正是在那里达到极限。再对圆心取力矩（moment），便可毫不费力地求得斜面的摩擦力。

解答

模型。圆柱体是半径为 r 的均匀刚体（rigid body），我们在通过其中心 O 的垂直截面上处理这个问题。设备的面是垂直的，并沿水平方向移动。设 B 为圆柱体与该面的接触点，它与 O 处于同一高度；A 为圆柱体与斜面的接触点。由于斜面的倾角为 45°，OA 与向下的铅垂线成 45°。取 x 轴向右（指向斜面），y 轴向上，并以 O 为原点，于是

$$B = (-r,\, 0), \qquad A = \tfrac{r}{\sqrt{2}}(1,\, -1).$$



哪一处的摩擦力达到极限

当圆柱体即将移动时，它正要离开地面，所以地面的法向反作用力（normal reaction）为零，地面的摩擦力亦随之为零。要滚上斜面，圆柱体在 A 点必须不打滑，所以它的运动以绕 A 的转动开始，而中心则沿 $(1,1)$ 方向向斜面上方移动。这个转动是顺时针方向的。设角速率（angular speed）为 ω，则圆柱体上一点 X 的速度等于 ω 乘以 $X-A$ 顺时针旋转 90° 后所得的向量，即把 $\omega\,(y,-x)$ 用于 $X-A=(x,y)$。对于 $X=B$，

$$B-A=r\Big(-1-\tfrac{1}{\sqrt{2}},\,\tfrac{1}{\sqrt{2}}\Big)\quad\Longrightarrow\quad \mathbf{v}_B=\omega r\Big(\tfrac{1}{\sqrt{2}},\,1+\tfrac{1}{\sqrt{2}}\Big).$$

该面只作水平移动，但 B 的速度有向上的分量，所以 B 沿该面向上滑动。圆柱体不可能同时在两个表面上滚动，所以它必定在垂直面上打滑，因此**摩擦力在垂直面上达到极限**。摩擦力与滑动方向相反，所以它作用于圆柱体的方向是向下，大小为 μP。在斜面上没有打滑，所以斜面摩擦力 F 的大小由平衡（equilibrium）条件决定，只须满足 $F\le\nu R$。（我们以 ν 表示斜面的摩擦系数（coefficient of friction），题目中印作 v。）

> ? **想一想** 为甚么不可以是斜面上的摩擦力达到极限?

假如圆柱体在 A 点打滑，它便是沿斜面向上滑动，而不是滚动。滚动的意思就是在 A 点不打滑。上面的计算显示，滚动必然导致 B 点打滑，所以 B 才是摩擦力达到法向反作用力的 μ 倍的接触点。

求 P

设 R 为斜面的法向反作用力，沿 AO 方向（方向为 $\frac{1}{\sqrt{2}}(-1,1)$），并设 F 为斜面摩擦力，以沿斜面向下为正（方向为 $\frac{1}{\sqrt{2}}(-1,-1)$）。

对 O 取力矩。重量、P 和 R 的作用线全部通过 O。B 点的摩擦力 μP 的力臂为 r，使圆柱体逆时针转动。A 点的摩擦力 F 沿切线方向，所以力臂同样为 r；当它指向斜面下方时，会使圆柱体顺时针转动。因此

$$\mu P\,r=Fr\quad\Longrightarrow\quad F=\mu P.$$

所以斜面摩擦力确实沿斜面向下，而且大小与垂直面上的摩擦力相同。

沿水平方向分解：

$$P-\frac{R}{\sqrt{2}}-\frac{\mu P}{\sqrt{2}}=0\quad\Longrightarrow\quad R=\sqrt{2}\,P-\mu P=P(\sqrt{2}-\mu).\tag{1}$$

沿垂直方向分解：

$$\frac{R}{\sqrt{2}}-\frac{\mu P}{\sqrt{2}}-\mu P-Mg=0.\tag{2}$$

把由 (1) 得出的 $R/\sqrt{2} = P - \mu P/\sqrt{2}$ 代入 (2)，

$$P - \frac{\mu P}{\sqrt{2}} - \frac{\mu P}{\sqrt{2}} - \mu P = Mg$$
$$P\left(1 - \mu - \sqrt{2}\,\mu\right) = Mg,$$

其中用到 $2/\sqrt{2} = \sqrt{2}$。因此

$$P = \frac{Mg}{1 - \mu(1+\sqrt{2})},$$

正是所求的结果。

μ 和 ν 须满足的条件

关于 μ。 P 是推力的大小，而 $Mg > 0$，所以我们需要 $1 - \mu(1+\sqrt{2}) > 0$，即

$$\mu < \frac{1}{1+\sqrt{2}} = \frac{\sqrt{2}-1}{(\sqrt{2}+1)(\sqrt{2}-1)} = \sqrt{2} - 1.$$

若 $\mu \geq \sqrt{2} - 1$，无论推力多大，都无法把圆柱体滚上斜面：垂直面上额外的向下摩擦力随 P 增长的速度，至少与上升力的增长一样快，于是圆柱体被卡死。

关于 ν。 圆柱体在斜面上不打滑，所以 $F \leq \nu R$。利用 $F = \mu P$ 和 (1)，并留意到由于 $\mu < \sqrt{2} - 1 < \sqrt{2}$，故 $R > 0$，

$$\mu P \leq \nu P(\sqrt{2} - \mu) \quad \Longleftrightarrow \quad \nu \geq \frac{\mu}{\sqrt{2} - \mu},$$

其中我们除以了 $P(\sqrt{2} - \mu) > 0$。至此全题完成。

验算

- *光滑的面*，$\mu = 0$：此时 $P = Mg$，而且不需要斜面摩擦力。这与一个无摩擦的圆柱体楔在垂直墙壁与 45° 斜面之间的情况吻合：斜面反作用力的水平分量和垂直分量相等，所以它必须同等地平衡 P 和 Mg。
- *数值*：取 $\mu = 0.2$，则 $P \approx 1.934\,Mg$，而我们需要 $\nu \geq 0.165$。把这些数值代回去，水平方向的力、垂直方向的力以及对 O 的力矩全部平衡（在舍入误差之内）；对随机选取的 M 和 μ 亦然。
- 当 $\mu \to \sqrt{2} - 1 \approx 0.414$ 时，$P \to \infty$，这与上文所述的卡死现象一致。

常见错误

- 把斜面上的摩擦力当作达到极限。“滚动”已排除了该处打滑的可能。

- 保留了地面的反作用力。在即将滚上斜面的一刻，圆柱体正要离地，所以地面的反作用力为零。
- 弄错垂直面上摩擦力的方向。B 相对于该面向上移动，所以作用于圆柱体的摩擦力是向下的。这正是 P 大于 Mg 的原因。

STEP 之外

本题的关键一步，是借着观察瞬时转动中心（instantaneous centre of rotation）来判断哪一个接触点打滑，这正是刚体运动学（rigid-body kinematics）的起点。在大学一年级的课程中，它会引伸出一般法则 $\mathbf{v}_X = \mathbf{v}_O + \boldsymbol{\omega} \times (X - O)$，以及多个粗糙接触点互相竞争的问题。在 $\mu \geq \sqrt{2} - 1$ 时出现的“卡死”现象，即怎样加大推力也无济于事，是一种自锁（self-locking）效应，工程师会在楔子和夹具中刻意加以利用。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门）正是以这种严谨的态度处理刚体运动和摩擦力。

详细解答

STEP I 样本试卷，第 12 题

力学 · 涉及摩擦力的碰撞

STEP I 样本试卷，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/spec-s1-q12

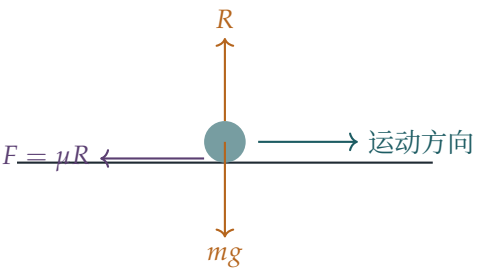
关键思路

这一题的主角是摩擦力（friction）。圆盘在水平冰面上滑行时，受到恒定的阻力，因此减速度（deceleration）不变，各事件之间的运动其实只是“匀加速运动”换了个面貌。我们由终点倒推回去：B 的停止距离决定了它在碰撞后的速率，碰撞定律把这个速率转化为 A 在碰撞前一刻的速率，再用一次匀加速运动的公式，便可回到起点 O。

解答

模型

把每个圆盘视为质量为 m 的质点（particle），置于摩擦系数（coefficient of friction）为 μ 的粗糙水平面上。圆盘滑行时，所受的力有：向下的重量 mg、向上的法向反作用力（normal reaction）R，以及与运动方向相反的极限摩擦力（limiting friction）F。



沿垂直方向分解：垂直方向没有加速度，所以 $R = mg$。圆盘正在滑动，故摩擦力达到极限值：$F = \mu R = \mu mg$。沿水平方向分解，$m\ddot{x} = -\mu mg$，所以每个滑行中的圆盘都以恒

定的减速度

$$a = \mu g,$$

减速，与其质量无关。我们又假设碰撞在瞬间完成，摩擦力在碰撞过程中来不及起作用；这时只需考虑两圆盘之间的碰撞力，而沿 OC 方向的动量（momentum）守恒。

B 在碰撞后的速率

B 由距离 O 为 d 的一点出发，停在距离 O 为 $2d$ 的 C 点，所以它滑行的距离是 d。设它在碰撞后一刻的速率为 w。利用 $v^2 = u^2 + 2as$，其中末速率为 0，加速度为 $-\mu g$，位移为 d，

$$0 = w^2 - 2\mu g d \quad \Longrightarrow \quad w = \sqrt{2\mu g d},$$

这正是所要求的结果。

A 在碰撞前一刻的速率

设 A 以速率 u 撞向 B，并设 A 在碰撞后一刻的速度为 v_A（两者均以 $O \to C$ 的方向为正）。由于两圆盘质量同为 m，动量守恒（conservation of momentum）给出

$$mu = mv_A + mw \quad \Longrightarrow \quad u = v_A + w,$$

而牛顿（Newton）实验定律给出

$$w - v_A = eu.$$

两式相加可消去 v_A：$2w = (1+e)u$，所以

$$u = \frac{2w}{1+e} = \frac{2\sqrt{2\mu g d}}{1+e}.$$

（若改为两式相减，则得 $v_A = \frac{1}{2}(1-e)u \geq 0$，所以 A 会以较慢的速率沿原来方向继续前进。）

> ? **想一想** A 会不会再次追上 B，令以上的计算失效?

不会。碰撞之后，两个圆盘都以相同的减速度 μg 减速，所以只要两者仍在运动，它们的速率差 $w - v_A = eu$ 便保持不变。B 永远不会比 A 慢，所以一直在前面；而 A 起步较慢，会先停下来。因此碰撞恰好只有一次，B 由 d 到 $2d$ 的一段路程不受干扰。（当 $e = 0$ 时，两个圆盘一起运动直至停下，同样不会再有碰撞。）

初速率

A 由 O 滑行距离 d 到达 B，期间以 μg 减速，所以再用一次 $v^2 = u^2 + 2as$，

$$u^2 = U^2 - 2\mu g d \quad \Longrightarrow \quad U^2 = u^2 + 2\mu g d.$$

代入 u 的表达式，

$$\begin{aligned} U^2 &= \frac{4\cdot 2\mu g d}{(1+e)^2} + 2\mu g d = \frac{2\mu g d}{(1+e)^2}\left(4+(1+e)^2\right) \\ &= \frac{2\mu g d}{(1+e)^2}\left(5+2e+e^2\right). \end{aligned}$$

由于 $U>0$ 且 $1+e>0$，取正平方根，

$$U = \frac{\sqrt{2\mu g d}}{1+e}\left(5+2e+e^2\right)^{\frac{1}{2}},$$

即为所求。

验算

- 完全弹性（perfectly elastic），$e=1$：两个相同的圆盘会互换速度，所以 A 立即停下，而 B 则以 A 抵达时的速率 $u=w$ 离开。公式给出 $U=\sqrt{2}\sqrt{2\mu g d}=\sqrt{2\mu g(2d)}$，恰好是单独一个圆盘滑完整段 $2d$ 到达 C 所需的速率。这很合理：动量原封不动地传了过去。
- 完全非弹性(perfectly inelastic),$e=0$:公式给出 $U=\sqrt{5}\sqrt{2\mu g d}$,比 $e=1$ 时大,这是应有的结果,因为碰撞中有动能(kinetic energy)损失。更一般地,$(5+2e+e^2)/(1+e)^2 = 1+4/(1+e)^2$ 随 e 增大而减小。
- 量纲（dimensions）：μ 和 e 都是纯数，而 $\sqrt{gd}$ 是速率。
- 数值上，取 $\mu=0.1$、$g=9.8\,\mathrm{m\,s^{-2}}$、$d=1\,\mathrm{m}$ 和 $e=0.5$，公式给出 $U=\frac{7}{3}\approx 2.33\,\mathrm{m\,s^{-1}}$。由这个 U 出发，以细小的时间步长逐步推算运动（并随机选取多组 μ、d 和 e），结果 B 停下的位置与 O 的距离和 $2d$ 相差不超过 0.001%。

常见错误

- 对碰撞前后使用能量守恒。除非 $e=1$，否则碰撞中会有动能损失；碰撞时应使用动量和恢复定律（restitution law），而功能原理（work–energy，或 $v^2=u^2+2as$）只可用于各事件之间。
- 忘记 A 在碰撞前同样要滑行距离 d，误写成 $U=u$。
- 把恢复定律写反了：应该是分离速率 $w-v_A$ 等于 e 乘以接近速率 u。

STEP 之外

这一题的巧妙之处，在于摩擦力令每个圆盘都有相同的减速度，与质量无关；因此在一个以 μg 减速的参考系（frame）中，只要两个圆盘都在滑行，冰面便等同于没有摩擦，而碰撞看起来就和教科书上的一模一样。选取合适的参考系是大学本科力学的核心技巧，而摩擦定律 $F = \mu R$ 本身只是一个经验模型，它的适用范围值得了解。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）对碰撞和摩擦都有细致的处理。

详细解答

STEP I 样本试卷，第 13 题

力学 • 圆周运动

STEP I 样本试卷，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/spec-s1-q13

关键思路

这是一个弦线可以伸长的圆锥摆（conical pendulum）。胡克定律（Hooke's law）把张力（tension）与弦线的长度联系起来，而沿垂直和水平方向分解力，则把张力与 g 和 ω 联系起来。由于弦线与垂直线的夹角不为零，张力必须严格大于 mg，而两个不等式都是由这一项事实推导出来的。

解答

把质点视为系于一条轻弹性弦线（light elastic string）上的点质量，弦线遵从胡克定律，并且不计空气阻力。设 λ 为弹性模量（modulus of elasticity）。

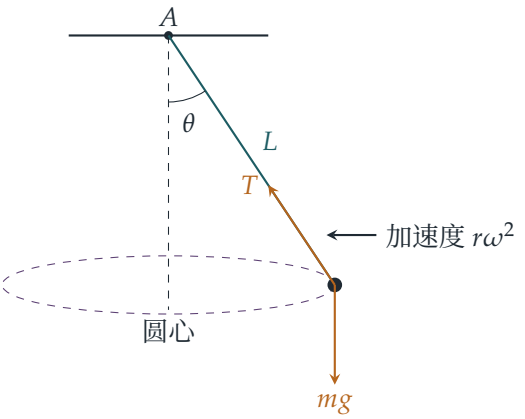
弹性模量。 质点静止悬挂时，弦线的伸长量（extension）为 $2a - a = a$，而张力与重量平衡：

$$mg = \frac{\lambda a}{a} \quad \Longrightarrow \quad \lambda = mg.$$

因此当弦线的长度为 L 时，其张力为

$$T = \frac{mg(L-a)}{a}.$$

圆周运动。 假设质点在 A 下方的一个水平圆上运动，长度为 L 的弦线与向下的垂直线成 θ 角。由于这是一个位于 A 所在水平之下的真正的圆周运动，我们有 $0 < \theta < \frac{\pi}{2}$，所以 $\sin\theta > 0$ 且 $0 < \cos\theta < 1$。圆心在 A 的正下方，半径为 $r = L\sin\theta$。



质点绕通过 A 的垂直线的角速率（angular speed）为 ω，所以它的加速度为 $r\omega^2 = L\omega^2 \sin\theta$，方向水平指向圆心。垂直方向没有加速度。

沿垂直方向分解：

$$T\cos\theta = mg. \tag{1}$$

沿水平方向（指向圆心）分解：

$$T\sin\theta = mL\omega^2\sin\theta \quad\Longrightarrow\quad T = mL\omega^2, \tag{2}$$

其中因为 $\sin\theta \neq 0$，我们可以除以 $\sin\theta$。

> **想一想** “这是一个真正的圆周运动（而不只是质点静止悬挂）”这项事实，在哪里用到了？

它用到了两次：在 (2) 中除以 $\sin\theta$，以及在下文指出 $\cos\theta < 1$。若 $\theta = 0$，质点只是静止悬挂在轴上，(2) 没有提供任何信息，而任何 ω 都“行得通”。

弦线长于 $2a$。 由 (1)，因为 $0 < \cos\theta < 1$，

$$T = \frac{mg}{\cos\theta} > mg.$$

利用胡克定律，这即是说

$$\frac{mg(L-a)}{a} > mg \quad\Longrightarrow\quad L - a > a \quad\Longrightarrow\quad L > 2a,$$

所以弦线被拉长至大于 $2a$ 的长度。

ω 的范围。 把由胡克定律和由 (2) 得到的两个 T 的表达式等同起来：

$$mL\omega^2 = \frac{mg(L-a)}{a} \quad\Longrightarrow\quad \omega^2 = \frac{g}{a}\left(1 - \frac{a}{L}\right).$$

由于 $L > 2a$，我们有 $0 < \frac{a}{L} < \frac{1}{2}$，所以 $\frac{1}{2} < 1 - \frac{a}{L} < 1$。因此

$$\frac{g}{2a} < \omega^2 < \frac{g}{a},$$

这正是我们要证明的结果。

逆命题（converse）（题目没有要求，但令人安心）。若 ω 位于这个范围内，取 $L = \dfrac{a}{1 - a\omega^2/g}$。它是正数，满足 $L > 2a$，并给出 $\cos\theta = \dfrac{mg}{T} = \dfrac{a}{L-a}$，其值严格介乎 0 与 1 之间。所以对于每一个这样的 ω，确实存在一个圆锥运动。

验算

- 极限：当 $L \to 2a$ 时，张力趋向 mg，$\theta \to 0$ 且 $\omega^2 \to g/(2a)$，这正是长度为 $2a$ 的单摆（simple pendulum）作小幅摆动的频率。当 $L \to \infty$ 时，$\omega^2 \to g/a$。
- 数值方面：取 $\omega^2 = \frac{3}{4}\,g/a$，我们得到 $L = 4a$、$T = 3mg$ 和 $\cos\theta = \frac{1}{3}$。从这些初始条件出发，以数值方法逐步求解完整的三维运动方程，计算两周，质点始终留在所预测的圆上，误差不超过 10^{-13} 个单位。
- 单位：g/a 的单位是 s^{-2}，与 ω^2 应有的单位相符。

常见错误

- 把圆的半径当作 L 而不是 $L\sin\theta$。这样 (2) 中的 $\sin\theta$ 便无法约去。
- 忘记了 $T > mg$ 需要 $\cos\theta < 1$，即 $\theta \neq 0$。严格不等式正是取决于这一点。
- 以自然长度（natural length）a（或平衡时的长度 $2a$）而不是拉长后的长度 L 作为半径。

STEP 之外

下限 $\omega^2 = g/(2a)$ 是绕垂直线作小幅振动的频率，而这并非巧合：靠近轴的稳定圆锥运动，其实就是两个互相垂直的小幅摆动，彼此相位相差四分之一个周期。上限则是弦线的刚度（stiffness）无论伸长多少都再也无法提供向心力（centripetal force）之处。研究稳定运动及其稳定性（stability），例如探讨质点被轻轻推离圆形轨道后会发生甚么事，是大学一年级动力学的标准课题。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对圆锥摆以及旋转系统中运动的分析，有细致的论述。

详细解答

STEP I 样本试卷，第 14 题

概率与统计 · 期望值

STEP I 样本试卷，第 14 题

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

step.maths.org/questions/spec-s1-q14

关键思路

一切都取决于一个表：两颗公平四面骰子点数总和的分布（distribution）。有了这个表，无论猜哪个 x，期望收益（expected gain）都只是“奖金乘以胜出的机会，减去罚款乘以落败的机会”。有趣的矛盾在于：猜大的数字奖金高，但机会小，所以要把所有猜测逐一比较，才看得出哪个最好。

解答

设 S 为总点数。每颗骰子掷出 $1,2,3,4$ 的概率（probability）各为 $\frac{1}{4}$，而且两颗骰子互相独立（independent），所以 16 个有序对（ordered pairs）出现的可能性均等。数出每个总和对应的有序对数目，得

s	2	3	4	5	6	7	8
$\mathrm{P}(S=s)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

（例如，$S=5$ 来自 $(1,4),(2,3),(3,2),(4,1)$）。

若 B 猜 x，并设 $p=\mathrm{P}(S=x)$，则他的收益 G（以便士计）有 p 的概率为 $100x^2$，有 $1-p$ 的概率为 $-50x$，所以

$$\begin{aligned}\mathrm{E}(G) &= 100x^2p - 50x(1-p)\\ &= 50x\big[(2x+1)p-1\big].\end{aligned}$$

写成 $p=\frac{k}{16}$，其中 k 是得出总和 x 的方法数目，则

$$\mathrm{E}(G)=\frac{25x\big[k(2x+1)-16\big]}{8}.$$

第 (i) 部分

只有 $(4,4)$ 的总和是 8，所以 $p = \frac{1}{16}$，而

$$\mathrm{E}(G) = 6400 \cdot \frac{1}{16} - 400 \cdot \frac{15}{16} = 400 - 375 = 25,$$

因此 B 猜 8 时的期望收益是 25 便士，正如题目所求。

第 (ii) 部分

猜 $2, 3, \ldots, 8$ 以外的数永远不会胜出，所以（取 x 为正数，因为点数必为正数）只会输钱。对于七个可能的总和，上面的公式给出

x	2	3	4	5	6	7	8
k	1	2	3	4	3	2	1
$\mathrm{E}(G)$（便士）	−68.75	−18.75	137.5	437.5	431.25	306.25	25

例如，当 $x = 5$ 时：$\mathrm{E}(G) = \frac{25 \cdot 5\,(4 \cdot 11 - 16)}{8} = \frac{125 \cdot 28}{8} = 437.5$。

想一想 奖金按 x^2 的速度增长，那么为甚么猜 7 或 8 这样的大数并不是最好?

奖金确实增长得很快，但胜出的机会在 5 的两侧逐渐下降，而猜错的可能性大得多，一旦猜错就要付罚款。在 $x = 8$ 时，胜出的机会已跌至 $\frac{1}{16}$，几乎刚好抵销了高额奖金。这两种效应在中间附近取得最佳平衡：$x = 6$ 的奖金比 $x = 5$ 高，但胜出的方法只有三种而不是四种，结果仅仅稍逊一筹。

所以我们建议 B 猜 $\boxed{x = 5}$，期望收益为 437.5 便士，即每局 £4.375（约 £4.38）。

验算

- 表中各概率之和为 $\frac{16}{16} = 1$。
- 第 (i) 部分的结果刚好是 25，印证了分布表和公式。
- 模拟 400 000 局猜 5 的游戏，平均收益约为 437.2 便士，而准确值是 437.5。

常见错误

- 把总和 2 至 8 当作可能性均等：其实不然，因为可能性均等的结果是那 16 个有序对。
- 忘记 B 是以 $1 - p$ 的概率输掉 $50x$，而不是以概率 1 输掉。
- 因为奖金较高，就停在 $x = 6$ 或 $x = 7$：5 和 6 的结果很接近，所以每个值都必须计算。

STEP 之外

这是一个小型的决策问题：选择令期望回报最大的行动。在大学里，同样的想法会在决策论（decision theory）中出现，也会在研究独立随机变量（random variables）之和时出现；在后者，和的分布是用卷积（convolution）求出的（在这里就是分布表的三角形状），而当骰子数目很多时，根据中心极限定理（central limit theorem），这个分布会趋近正态曲线（normal curve）。Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）涵盖了这两个课题。

详细解答

STEP I 样本试卷，第 15 题

概率与统计 · 随机变量之和

> **STEP I 样本试卷，第 15 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/spec-s1-q15

关键思路

C 这个看似古怪的分布，其实正是把一个价钱四舍五入至最接近的 10 便士时所产生的误差，前提是价钱的最后一位数字等可能地取 $0, 1, \ldots, 9$ 中任何一个。于是，购物者的总误差就是 36 个独立而与 C 同分布的变量之和，其平均值和方差（variance）可立即求得，而中心极限定理（Central Limit Theorem）则告诉我们，20 便士的差额有多令人意外。

解答

C 的期望值与方差

由于对每个 i 都有 $\mathrm{P}(C = i) = \mathrm{P}(C = -i)$，这个分布关于 0 对称。因此各项 $i\,\mathrm{P}(C = i)$ 两两抵消，得

$$\mathrm{E}(C) = 0.$$

（各概率之和确实是 1：$9 \cdot \frac{1}{10} + 2 \cdot \frac{1}{20} = 1$。）既然平均值是 0，方差就只是 $\mathrm{E}(C^2)$：

$$\begin{aligned}\mathrm{Var}(C) = \mathrm{E}(C^2) &= 2 \cdot \frac{25}{20} + \frac{2}{10}\left(1^2 + 2^2 + 3^2 + 4^2\right) \\ &= \frac{5}{2} + \frac{2 \cdot 30}{10} = \frac{17}{2}.\end{aligned}$$

购物者的估算

模型。 对于一件货品，设 C 为实际价钱减去舍入后的价钱，以便士为单位。它只取决于价钱的最后一位数字 d。我们假设 d 等可能地取 $0, 1, \ldots, 9$ 中任何一个；在这里，“随机”选购理应是这个意思，不过这毕竟是对现实价钱所作的一个假设。

- $d = 0$：无须舍入，所以 $C = 0$。
- $d = 1, 2, 3, 4$：价钱向下舍入了 d，所以 $C = d$。
- $d = 6, 7, 8, 9$：价钱向上舍入了 $10 - d$，所以 $C = -(10 - d)$，即 $-4, -3, -2, -1$。
- $d = 5$：价钱以相等的概率向下或向上舍入，所以 $C = 5$ 或 -5，各自的概率为 $\frac{1}{2} \cdot \frac{1}{10} = \frac{1}{20}$。

$-4, \ldots, 4$ 中每个值都恰好由一个数字产生，概率为 $\frac{1}{10}$，而 ± 5 的概率各为 $\frac{1}{20}$。这正正是第一部分的分布，题目如此设置的原因也在于此。

总误差。 设 $C_1, \ldots, C_{36}$ 为 36 件货品各自的误差，并设

$$S = C_1 + C_2 + \cdots + C_{36}$$

为真正的账单金额减去购物者的估算。由于货品是随机选购的，我们视各个 C_j 为独立（independent），且每个都与 C 同分布。于是

$$\mathrm{E}(S) = 36 \cdot 0 = 0, \qquad \mathrm{Var}(S) = 36 \cdot \tfrac{17}{2} = 306,$$

其中计算方差时用到了独立性。因此 S 的标准差（standard deviation）是 $\sqrt{306} \approx 17.5$ 便士。

> **想一想** 20 便士的差额算大还是算小?

这个差额必须与标准差比较，而不是与单次舍入误差的大小比较。收银员的账单比估算高出 20 便士，即 $S = 20$。这比平均值高出

$$\frac{20}{\sqrt{306}} \approx 1.14$$

个标准差。由于 S 是 36 个独立同分布变量之和，中心极限定理指出 S 近似服从 $\mathrm{N}(0, 306)$。采用连续性校正（continuity correction）（因为 S 只取整数值），

$$\mathrm{P}(S \geq 20) \approx \mathrm{P}\left(Z \geq \frac{19.5}{\sqrt{306}}\right) = \mathrm{P}(Z \geq 1.11) \approx 0.13,$$

其中 $Z \sim \mathrm{N}(0, 1)$。至少这么大的误差（不论方向）大约有 26% 的机会出现。

所以购物者**不**应怀疑账单出错。20 便士的差额完全在她的舍入误差的正常波动范围之内，大约是一个标准差，而且大约每七、八次购物就会出现一次。只有差额超过大约两个标准差，例如 35 便士或以上，她才有真正的理由起疑。

验算

- 把价钱四舍五入至最接近的 10 便士，最多只会令它改变 5 便士，所以 36 件货品在最坏情况下的误差是 180 便士。典型误差约为 17 便士，远小于此，这也合乎预期：正误差与负误差大致互相抵消。
- 电脑验证了舍入误差的分布与 C 完全相同，且 $\mathrm{E}(C) = 0$、$\mathrm{Var}(C) = \frac{17}{2}$。接着它以分数对 36 个相同的分布作精确卷积（convolution）：$\mathrm{Var}(S) = 306$，$\mathrm{P}(S \geq 20) \approx 0.1330$；相比之下，正态近似（normal approximation）给出 0.1325，而 2×10^5 次模拟购物则给出 0.132。

常见错误

- 把 $\mathrm{Var}(S)$ 当作 $36^2\,\mathrm{Var}(C)$。那其实是 $36C$ 的方差，即同一个误差重复 36 次；对于独立的误差，方差只须直接相加。
- 以 S 恰好等于 20 的概率（约 0.012）来衡量意外的程度。S 的每一个个别取值本身都相当不可能出现；真正重要的，是误差至少与观察所得一样大的概率。
- 忘记说明所作的假设：价钱的最后一位数字服从均匀分布（uniform distribution），而且各件货品的误差互相独立。

STEP 之外

这就是假设检验（hypothesis test）背后的推理：先假定一切正常（在这里，即 S 纯粹是舍入误差），然后问：出现至少与观察所得同样极端的结果，可能性有多大？同样的舍入误差分析也用于数值分析（numerical analysis），在那里，对许多数字作舍入所产生的误差，通常随数字个数的平方根增长，而非线性增长。Blitzstein 与 Hwang 的 *Introduction to Probability*（概率论入门）对随机变量（random variable）之和及中心极限定理有清晰的讲解。

详细解答

STEP I 样本试卷，第 16 题

概率与统计 · 连续随机变量

STEP I 样本试卷，第 16 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/spec-s1-q16

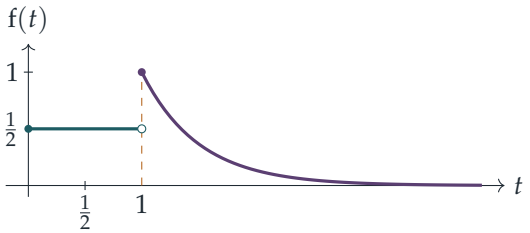
关键思路

未知数 k 由“总概率为 1”这一事实决定，其余各项只要把密度函数的两段分开积分便可得出。至于费用，窍门是把连续变量 T 化为离散变量：硬币的数目只取决于已过了多少个完整的半分钟，所以我们需要求 T 落在每个半分钟区间内的概率。余下的工作交给一个几何型的级数便可完成。

解答

求 k 及平均值

密度函数在 $[0,1)$ 上是高度为 $\frac{1}{2}$ 的水平线段，其后接上一条指数尾部（exponential tail），并在 $t=1$ 处有一个跳跃。



第一段贡献的概率为 $\frac{1}{2}$，所以尾部必须贡献余下的 $\frac{1}{2}$：

$$k\int_1^\infty \mathrm{e}^{-2t}\,\mathrm{d}t = k\left[-\tfrac{1}{2}\mathrm{e}^{-2t}\right]_1^\infty = \tfrac{1}{2}k\mathrm{e}^{-2} = \tfrac{1}{2}, \qquad \text{所以} \qquad k=\mathrm{e}^2.$$

（因此在 $t\geq 1$ 上的密度函数为 e^{2-2t}，它在 $t=1$ 处等于 1，与草图相符。）

至于平均值，取 $u = t$、$\mathrm{d}v = \mathrm{e}^{-2t}\,\mathrm{d}t$ 作分部积分（integration by parts）：

$$\int_1^\infty t\,\mathrm{e}^{-2t}\,\mathrm{d}t = \Big[-\tfrac{1}{2}t\,\mathrm{e}^{-2t}\Big]_1^\infty + \tfrac{1}{2}\int_1^\infty \mathrm{e}^{-2t}\,\mathrm{d}t = \tfrac{1}{2}\mathrm{e}^{-2} + \tfrac{1}{4}\mathrm{e}^{-2} = \tfrac{3}{4}\mathrm{e}^{-2},$$

其中用到当 $t \to \infty$ 时 $t\,\mathrm{e}^{-2t} \to 0$。因此

$$\mathrm{E}(T) = \int_0^1 \tfrac{1}{2}t\,\mathrm{d}t + \mathrm{e}^2 \cdot \tfrac{3}{4}\mathrm{e}^{-2} = \tfrac{1}{4} + \tfrac{3}{4} = 1,$$

所以通话时间的期望值（expected value）是一分钟，正如题目所求。

累积分布函数

记 $\mathrm{F}(t) = \mathrm{P}(T \le t)$，即累积分布函数（cumulative distribution function）。当 $t < 0$ 时它等于 0；当 $0 \le t < 1$ 时它等于 $\int_0^t \tfrac{1}{2}\,\mathrm{d}s = \tfrac{1}{2}t$；而当 $t \ge 1$ 时，

$$\mathrm{F}(t) = \tfrac{1}{2} + \int_1^t \mathrm{e}^{2-2s}\,\mathrm{d}s = \tfrac{1}{2} + \tfrac{1}{2}\big(1 - \mathrm{e}^{2-2t}\big) = 1 - \tfrac{1}{2}\mathrm{e}^{2-2t}.$$

所以

$$\mathrm{F}(t) = \begin{cases} 0 & t < 0, \\ \tfrac{1}{2}t & 0 \le t < 1, \\ 1 - \tfrac{1}{2}\mathrm{e}^{2-2t} & t \ge 1. \end{cases}$$

留意虽然 f 在 $t = 1$ 处有跳跃，F 在该处却是连续的（两条公式都给出 $\tfrac{1}{2}$），而且当 $t \to \infty$ 时 $\mathrm{F}(t) \to 1$。下文会用到这个形式：当 $t \ge 1$ 时，$\mathrm{P}(T > t) = \tfrac{1}{2}\mathrm{e}^{2-2t}$。

费用的期望值

通话开始时投入一枚硬币，其后每过完一个半分钟再投入一枚。所以若通话历时 T 分钟，完整的半分钟数目为 $\lfloor 2T \rfloor$，而

$$C = 10\big(1 + \lfloor 2T \rfloor\big).$$

（若通话恰好在某个半分钟的整点结束，那一刻有没有投入硬币并不要紧：T 是连续的，所以这种情况的概率为 0。）

想一想 怎样由连续的 T 得出 C 的分布？

事件 $C = 10(n+1)$ 正正就是事件 $\tfrac{n}{2} \le T < \tfrac{n+1}{2}$，其概率可直接由 F 读出：

- $n = 0$：$\mathrm{P}(C = 10) = \mathrm{F}(\tfrac{1}{2}) = \tfrac{1}{4}$；
- $n = 1$：$\mathrm{P}(C = 20) = \mathrm{F}(1) - \mathrm{F}(\tfrac{1}{2}) = \tfrac{1}{4}$；

- $n \geq 2$：两个端点都不小于 1，所以

$$\mathrm{P}\big(C = 10(n+1)\big) = \tfrac{1}{2}\mathrm{e}^{2-n} - \tfrac{1}{2}\mathrm{e}^{2-(n+1)} = \tfrac{1}{2}\big(1-\mathrm{e}^{-1}\big)\mathrm{e}^{2-n}.$$

设 $r = \mathrm{e}^{-1}$ 及 $m = n-2$，则 $\mathrm{e}^{2-n} = r^m$ 而 $10(n+1) = 10(m+3)$。于是

$$\begin{aligned}\mathrm{E}(C) &= 10\cdot\tfrac{1}{4} + 20\cdot\tfrac{1}{4} + \sum_{m=0}^{\infty} 10(m+3)\cdot\tfrac{1}{2}(1-r)\,r^m \\ &= \tfrac{15}{2} + 5(1-r)\left(\sum_{m=0}^{\infty} m r^m + 3\sum_{m=0}^{\infty} r^m\right).\end{aligned}$$

由于 $0 < r < 1$，由几何级数（geometric series）得 $\sum r^m = \dfrac{1}{1-r}$，把它逐项求导（再乘以 r）便得 $\sum m r^m = \dfrac{r}{(1-r)^2}$。因此

$$\mathrm{E}(C) = \tfrac{15}{2} + 5(1-r)\left(\frac{r}{(1-r)^2} + \frac{3}{1-r}\right) = \tfrac{15}{2} + \frac{5r}{1-r} + 15.$$

最后，$\dfrac{r}{1-r} = \dfrac{\mathrm{e}^{-1}}{1-\mathrm{e}^{-1}} = \dfrac{1}{\mathrm{e}-1}$，所以

$$\mathrm{E}(C) = 22\tfrac{1}{2} + \frac{5}{\mathrm{e}-1},$$

正如题目所求。其数值约为 25.41 便士。

评论

若按每分钟 20 便士计算，平均长度为一分钟的通话平均应花费 20 便士，但 $\mathrm{E}(C) \approx 25.4$。出现这个差距，是因为那人要为每个半分钟预先付款：每枚硬币买下完整的半分钟，但最后一枚几乎从来不会用尽。事实上，由于 $\lfloor 2T \rfloor > 2T - 1$，每一次通话都有 $C = 10(1 + \lfloor 2T \rfloor) > 20T$；费用把通话长度向上舍入至下一个半分钟，所以 $\mathrm{E}(C)$ 比 $20\,\mathrm{E}(T)$ 多出的数额，等于已付款而未用的时间的期望值的二十倍（在这里约为 5.4 便士）。

验算

- C 的各个概率之和为 1：$\frac{1}{4} + \frac{1}{4} + \sum_{m\geq 0} \frac{1}{2}(1-r)r^m = \frac{1}{2} + \frac{1}{2} = 1$。
- 由界限 $20T < C \leq 20T + 10$ 得 $20 < \mathrm{E}(C) \leq 30$，而 25.41 稳妥地落在其中。假如最后一个半分钟内未用的时间是均匀分布（uniform distribution）的，我们会预期多付约 5 便士；长尾令这个数额稍为多一点。
- 数值上，辛普森法则（Simpson's rule）给出总概率为 1、平均值为 1；另外模拟 10^6 次通话（一半在 $[0,1)$ 上均匀分布，一半为 1 加上率为 2 的指数分布），所得平均费用为 25.41 便士，与 $22.5 + 5/(\mathrm{e}-1) = 25.4099$ 吻合。

常见错误

- 把硬币数目当作 $\lfloor 2T \rfloor$ 而非 $1 + \lfloor 2T \rfloor$，忘记了开始时投入的那一枚；这样答案会刚好少了 10 便士。
- 在 $t = \frac{1}{2}$ 处套用 F 在 $t \geq 1$ 时的公式：区间 $[0, \frac{1}{2})$ 和 $[\frac{1}{2}, 1)$ 必须用均匀分布的那一段来处理。

STEP 之外

留意 T 的尾部具有无记忆性（memoryless property）：已知 $T \geq 1$ 时，剩余时间 $T - 1$ 服从率为 2 的指数分布，这正是 $n \geq 2$ 时各个概率构成几何数列（geometric sequence）的原因。记 $N = \lfloor 2T \rfloor$，则尾和公式（tail-sum formula）$\mathrm{E}(N) = \sum_{n \geq 1} \mathrm{P}(N \geq n)$ 能更快地得出相同的答案。像这样混合了离散与连续特性的分布，在 Grimmett 与 Welsh 的 *Probability: An Introduction* 中有仔细的论述，而 Blitzstein 与 Hwang 的 *Introduction to Probability* 则载有更多关于无记忆性的例子。

详细解答

STEP II 样本试卷，第 1 题

纯数学 · 部分分式

STEP II 样本试卷，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/spec-s2-q1

关键思路

在动用惯常的部分分式（partial fractions）那一套方法之前，先看看分子：它正正就是分母的导数（derivative）。由于分母可以分解为线性因式的乘积，f 是一个对数导数（logarithmic derivative），而乘积的对数导数等于各因式的对数导数之和。至于第二部分，方程变成三个单位分数（unit fractions）之和等于 1，单凭大小的考虑就足以迫使 x 非常小。

解答

部分分式

分母可以因式分解：

$$D(x) = x^3 + (a+b)x^2 + abx = x\big(x^2 + (a+b)x + ab\big) = x(x+a)(x+b),$$

求导后得

$$D'(x) = 3x^2 + 2(a+b)x + ab,$$

这恰好就是分子。所以 $\mathrm{f}(x) = D'(x)/D(x)$。

一般情况，$a \neq b$。三个因式 x、$x+a$、$x+b$ 互不相同（因为 $a \neq 0$、$b \neq 0$ 及 $a \neq b$），所以我们预期会有三个简单的部分分式。利用积法则（product rule），

$$D'(x) = (x+a)(x+b) + x(x+b) + x(x+a),$$

再逐项除以 $D(x) = x(x+a)(x+b)$，得

$$\boxed{\mathrm{f}(x) = \frac{1}{x} + \frac{1}{x+a} + \frac{1}{x+b}} \qquad (a \neq b).$$

（等价地，$\mathrm{f}(x) = \dfrac{\mathrm{d}}{\mathrm{d}x}\ln|x(x+a)(x+b)|$，而乘积的对数可拆成和。）当然，掩盖法（cover-up rule）也会得出同样的结果：在 $x=0$ 处留数（residue）为 $ab/(ab)=1$，在 $x=-a$ 处则为 $\big(3a^2-2a^2-2ab+ab\big)/\big((-a)(b-a)\big) = a(a-b)/\big(a(a-b)\big) = 1$，在 $x=-b$ 处亦同理。

> ? **想一想** 当 $a=b$ 时，这个形式会出甚么问题？

特殊情况，$a=b$。这时分母是 $x(x+a)^2$，带有一个重复因式（repeated factor），所以标准形式应为 $\frac{A}{x}+\frac{B}{x+a}+\frac{C}{(x+a)^2}$。但分子同样有因式 $x+a$：

$$3x^2+4ax+a^2 = (3x+a)(x+a),$$

所以它可以约去，剩下

$$\mathrm{f}(x) = \frac{3x+a}{x(x+a)} = \boxed{\frac{1}{x}+\frac{2}{x+a}} \qquad (a=b),$$

其中最后一步用了掩盖法（$x=0$ 给出 $a/a=1$；$x=-a$ 给出 $-2a/(-a)=2$）。式中根本没有 $(x+a)^{-2}$ 项，所以 $C=0$。这与把 b 换成 a 后的一般公式一致，但值得留意的是，重复因式是消失了，而不是产生一个平方项。

由于 a 和 b 均不为零，因式 x 永远不会与 $x+a$ 或 $x+b$ 重合，所以只有以上这些情况。

当 $\mathrm{f}(x)=1$ 时

现在设 x、a、b 为正整数。在上述两种情况中（$a=b$ 的公式不过是一般公式取 $b=a$），

$$\mathrm{f}(x) = \frac{1}{x}+\frac{1}{x+a}+\frac{1}{x+b},$$

而三项全为正数。因为 $a,b\geq 1$，后两项各自严格小于 $\frac{1}{x}$。

> ? **想一想** 如果三个各不超过 $\frac{1}{x}$ 的分数加起来等于 1，x 最大可以是多少？

- 若 $x=1$，则 $\mathrm{f}(x) = 1+(\text{正数}) > 1$。
- 若 $x\geq 4$，则 $\mathrm{f}(x) < \frac{3}{x} \leq \frac{3}{4} < 1$。
- 若 $x=3$，则 $x+a\geq 4$ 及 $x+b\geq 4$，所以 $\mathrm{f}(x) \leq \frac{1}{3}+\frac{1}{4}+\frac{1}{4} = \frac{5}{6} < 1$。

所以 $\mathrm{f}(x)=1$ 迫使 $x=2$，这时我们需要

$$\frac{1}{2+a}+\frac{1}{2+b} = \frac{1}{2}.$$

两边乘以 $2(2+a)(2+b)$ 得 $2(2+b)+2(2+a) = (2+a)(2+b)$，即 $8+2a+2b = 4+2a+2b+ab$，所以

$$ab = 4.$$

由于 a、b 为正整数，仅有的可能是 $(a,b)=(1,4)$、$(2,2)$、$(4,1)$。因此 $\mathrm{f}(x)=1$ 恰好在以下情况成立：

$$(x,a,b)=(2,1,4),\quad (2,2,2),\quad (2,4,1),$$

这是有限个（三个）值，正如题目所求。注意 $(2,2,2)$ 属于特殊情况 $a=b$：$\frac{1}{2}+\frac{2}{4}=1$。

验算

这三个解就是我们熟悉的单位分数恒等式 $\frac{1}{2}+\frac{1}{3}+\frac{1}{6}=1$ 和 $\frac{1}{2}+\frac{1}{4}+\frac{1}{4}=1$。数值方面，我们以精确的有理数运算，在 2000 个随机有理点上把部分分式形式与原来的 f 作比较（$a\neq b$ 和 $a=b$ 两种情况都有），另外对所有 $1\le x,a,b<200$ 作穷举搜索，结果恰好找到上述三组数，再无其他。

常见错误

- 在 $a=b$ 的情况中带着 $\frac{C}{(x+a)^2}$ 项埋头硬算，却没有留意到可以约分；你应该得出 $C=0$，而看出分子中的因式 $(x+a)$ 就可以省却这番工夫。
- 忘记排除 $x=1$ 和 $x=3$；粗略的上界 $\mathrm{f}<3/x$ 只能处理 $x\ge 4$。
- 停在“$ab=4$”而没有列出各个解，或者只列出 $(1,4)$ 而遗漏 $(4,1)$ 和 $(2,2)$。

STEP 之外

恒等式 $D'/D=\sum 1/(x-r_i)$（其中 r_i 是多项式 D 的根，重根按重数计算）用途极广：它是关于根的幂和的牛顿恒等式（Newton's identities）的基础，在复分析（complex analysis）中又是辐角原理（argument principle）的基础；辐角原理借着沿一条曲线对 f'/f 积分，数出函数在曲线内的零点数目。像 $\frac{1}{x}+\frac{1}{y}+\frac{1}{z}=1$ 这类求正整数解的方程属于埃及分数（Egyptian fraction）问题，是丢番图方程（Diophantine equations）中一个经典的角落，在那里像我们这样的估界论证就足以解决整个问题。Needham 的 *Visual Complex Analysis* 对辐角原理有优美的几何阐述，而 Davenport 的 *The Higher Arithmetic* 则是入门丢番图方程的好起点。

详细解答

STEP II 样本试卷，第 2 题

纯数学 · 对称联立方程

> **STEP II 样本试卷，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/spec-s2-q2

关键思路

三条方程具有循环（cyclic）对称性，所以很自然会把它们两两相减，而每个差都可以因式分解（factorise），并以 $x+y+z$ 为公因式（common factor）。接下来，各差的平方和有双重用途：以 x、y、z 表示时，它是 $a+b+c$ 的倍数；而第一个结果则可把它用 $a-b$、$b-c$、$c-a$ 和 $x+y+z$ 表示。一旦知道 $x+y+z$ 和两两之差，余下的工作只需解三条线性方程（linear equations）便可完成。

解答

相等的比

把第一条方程减去第二条方程：

$$\begin{aligned} a-b &= x^2-y^2-yz+zx \\ &= (x-y)(x+y)+z(x-y) \\ &= (x-y)(x+y+z). \end{aligned}$$

用完全相同的方法（或作循环代换 $x \to y \to z \to x$，$a \to b \to c \to a$），可得

$$b-c=(y-z)(x+y+z), \qquad c-a=(z-x)(x+y+z).$$

设 $S=x+y+z$。由于 $a \neq b$，乘积 $(x-y)S$ 不等于零，所以 $S \neq 0$（同理，x、y、z 互不相等）。因为 a、b、c 互不相等，我们可以除以 $a-b$、$b-c$、$c-a$ 以及 S，得出

$$\frac{y-z}{b-c}=\frac{z-x}{c-a}=\frac{x-y}{a-b}=\frac{1}{x+y+z},$$

这正是所要证明的结果。

求 $x+y+z$

把题目提示的和展开：

$$\begin{aligned}(y-z)^2+(z-x)^2+(x-y)^2 &= 2\big(x^2+y^2+z^2-xy-yz-zx\big)\\ &= 2\big[(x^2-yz)+(y^2-zx)+(z^2-xy)\big]\\ &= 2(a+b+c).\end{aligned}$$

另一方面，由第一部分可知 $y-z=(b-c)/S$，其余类推，所以同一个和亦等于

$$\frac{(b-c)^2+(c-a)^2+(a-b)^2}{S^2}=\frac{2\big(a^2+b^2+c^2-bc-ca-ab\big)}{S^2}.$$

令两个表达式相等，再除以 $2(a+b+c)$（因为 a、b、c 均为正数，所以它是正数），得

$$S^2=\frac{a^2+b^2+c^2-bc-ca-ab}{a+b+c}.$$

因此 $x+y+z=\triangle$，其中 $\triangle^2$ 就是上述的量，这正是所要证明的结果。注意 $\triangle^2>0$，因为它的分子是 $(b-c)^2+(c-a)^2+(a-b)^2$ 的一半，而当 a、b、c 互不相等时，这个和是正数。所以有两个可能，即 $\triangle=\pm\sqrt{\triangle^2}$，而到目前为止，没有任何条件能从中选定其一。

解方程

现在我们知道

$$x+y+z=\triangle,\qquad x-y=\frac{a-b}{\triangle},\qquad x-z=\frac{a-c}{\triangle}.$$

把三式相加，得

$$3x=\triangle+\frac{2a-b-c}{\triangle}=\frac{\triangle^2+2a-b-c}{\triangle}.$$

这个式子可以化简得很漂亮。设 $s=a+b+c$ 和 $Q=a^2+b^2+c^2-bc-ca-ab$，于是 $\triangle^2=Q/s$。那么

$$\begin{aligned}s\big(\triangle^2+2a-b-c\big) &= Q+(a+b+c)(2a-b-c)\\ &= \big(a^2+b^2+c^2-bc-ca-ab\big)+\big(2a^2+ab+ca-b^2-c^2-2bc\big)\\ &= 3\big(a^2-bc\big).\end{aligned}$$

因此 $x=\dfrac{a^2-bc}{s\triangle}$，而 y、z 可由循环代换得出。现在 $(s\triangle)^2=sQ=a^3+b^3+c^3-3abc$，这是根据以下的标准因式分解：

$$a^3+b^3+c^3-3abc=(a+b+c)\big(a^2+b^2+c^2-bc-ca-ab\big).$$

设 $K=\sqrt{a^3+b^3+c^3-3abc}>0$，则 $s\triangle=\pm K$，而方程组的解为

$$(x,y,z)=\pm\frac{1}{K}\left(a^2-bc,\ b^2-ca,\ c^2-ab\right),$$

其中三个分量取相同的正负号。

想一想 我们只证明了任何解都必定是这个形式。这两组真的都是解吗?

是的。以上所有步骤只是一连串的推论，所以我们应该把结果代回原方程验证。由于方程中只出现未知数的平方和两个未知数的乘积，整体的正负号并无影响，而

$$\begin{aligned}(a^2-bc)^2-(b^2-ca)(c^2-ab)&=a^4-2a^2bc+b^2c^2-\left(b^2c^2-ab^3-ac^3+a^2bc\right)\\&=a\left(a^3+b^3+c^3-3abc\right)=aK^2.\end{aligned}$$

除以 K^2 便得 $x^2-yz=a$；另外两条方程可由循环代换得出。所以方程组恰好有两组解，其中一组是另一组的相反数。

验算

- 取 $a=1$，$b=2$，$c=3$。那么 $a^3+b^3+c^3-3abc=18$，$K=3\sqrt{2}$，而取正号的解是 $(x,y,z)=\frac{1}{3\sqrt{2}}(-5,1,7)$。例如 $x^2-yz=\frac{25-7}{18}=1$，以及 $x+y+z=\frac{1}{\sqrt{2}}$，其平方为 $\frac{1}{2}=\frac{14-11}{6}=\triangle^2$。
- 以电脑测试了 20,000 组随机选取、互不相等的正数 a、b、c，并且两种正负号都测试过：公式满足全部三条方程以及相等的比，相对误差小于 10^{-11}。

常见错误

- 只取 $\triangle$ 的正平方根：题目中的 $\triangle$ 只是通过 $\triangle^2$ 来定义，而两种正负号都给出真正的解。
- 除以 $x+y+z$ 时没有说明它为何不等于零。在这里，这是由 $a\neq b$ 推出的。

STEP 之外

恒等式 $a^3+b^3+c^3-3abc=(a+b+c)(a^2+b^2+c^2-ab-bc-ca)$ 是以 (a,b,c)、(c,a,b)、(b,c,a) 为各行的循环矩阵（circulant matrix）的行列式（determinant），而它在复数（complex numbers）范围内可以完全分解为 $(a+b+c)(a+\omega b+\omega^2c)(a+\omega^2b+\omega c)$，其中 $\omega=e^{2\pi i/3}$。事实上，映射 $(x,y,z)\mapsto(x^2-yz,\ y^2-zx,\ z^2-xy)$ 就是循环矩阵的伴随矩阵（adjugate），这解释了为何它那么容易求逆。循环矩阵及其特征向量（eigenvectors）

是初次接触离散傅里叶变换（discrete Fourier transform）的一个平易入门；想进一步了解，Strang 的 *Introduction to Linear Algebra*（线性代数入门教科书）是很好的读物。

详细解答

STEP II 样本试卷，第 3 题

纯数学 · 棣美弗定理

STEP II 样本试卷，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/spec-s2-q3

关键思路

棣莫弗定理（De Moivre's theorem）用数学归纳法（induction）证明最为自然：根据复角公式（compound-angle formulae），每多乘一个因子 $\cos\theta + \mathrm{i}\sin\theta$，角度就增加 θ。至于那两条方程，$\cos ky$、$\sin ky$ 伴随着 x^k 出现，这个规律分明在提示“实部（real part）与虚部（imaginary part）”。设 $z = x(\cos y + \mathrm{i}\sin y)$，把第一条方程看成一条复数方程的实部，第二条看成它的虚部，这对方程便化为 z 的一条三次方程（cubic），而且有一个容易找到的根。

解答

棣莫弗定理

我们对 n 用归纳法。当 $n = 1$ 时，两边都是 $\cos\theta + \mathrm{i}\sin\theta$，所以结果成立。

假设结果对某个 $n = k \geq 1$ 成立。则

$$\begin{aligned}(\cos\theta + \mathrm{i}\sin\theta)^{k+1} &= (\cos k\theta + \mathrm{i}\sin k\theta)(\cos\theta + \mathrm{i}\sin\theta)\\ &= (\cos k\theta\cos\theta - \sin k\theta\sin\theta) + \mathrm{i}(\sin k\theta\cos\theta + \cos k\theta\sin\theta)\\ &= \cos(k+1)\theta + \mathrm{i}\sin(k+1)\theta,\end{aligned}$$

其中第一行用了归纳假设（induction hypothesis），第二行用了 $\mathrm{i}^2 = -1$，第三行用了 $\cos(A+B)$ 和 $\sin(A+B)$ 的复角公式。所以结果对 $n = k+1$ 成立，由归纳法可知，结果对所有正整数 n 都成立。

解方程

设 x 和 y 为实数，并令

$$z = x(\cos y + \mathrm{i}\sin y).$$

由棣莫弗定理，对 $n = 1, 2, 3$，

$$z^n = x^n(\cos y + \mathrm{i}\sin y)^n = x^n \cos ny + \mathrm{i}\, x^n \sin ny.$$

（无论 x 是正是负，这都成立：实数因子 x 直接提出来成为 x^n。）由于 x 和 y 是实数，两条方程的左边正好是 $z^3 + 2z^2 + 2z$ 的实部和虚部。所以这对实数方程等价于一条复数方程

$$z^3 + 2z^2 + 2z = -1, \qquad \text{即} \qquad z^3 + 2z^2 + 2z + 1 = 0.$$

想一想 为甚么一条复数方程等同于两条实数方程?

两个复数相等，当且仅当它们的实部相同、虚部也相同。右边的 -1 的实部是 -1，虚部是 0，正是题目给出的两个右边的值。所以 (x, y) 同时满足两条实数方程，当且仅当 z 满足该三次方程。

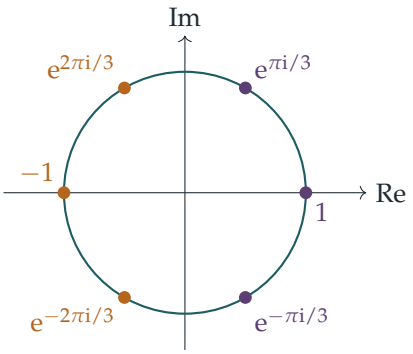
现在 $z = -1$ 是一个根，因为 $-1 + 2 - 2 + 1 = 0$，除去这个因式后得

$$z^3 + 2z^2 + 2z + 1 = (z + 1)(z^2 + z + 1).$$

这个二次式的根是 $z = \frac{1}{2}(-1 \pm \mathrm{i}\sqrt{3})$。所以三个根是

$$z = -1 = \cos \pi + \mathrm{i}\sin \pi, \qquad z = \cos \tfrac{2\pi}{3} \pm \mathrm{i}\sin \tfrac{2\pi}{3},$$

全部位于单位圆（unit circle）上。



余下要做的，是把每个根还原为实数对 (x, y)。由于 $|z| = |x|$，每个解都有 $|x| = 1$；特别地，$x = 0$ 是不可能的（否则会得出 $0 = -1$）。

- 若 $x = 1$，则 $\cos y + \mathrm{i}\sin y = z$，所以 y 是三个根之一的辐角（argument）：$y = \pi$、$\frac{2\pi}{3}$ 或 $-\frac{2\pi}{3}$，再加上 2π 的任意整数倍。
- 若 $x = -1$，则 $\cos y + \mathrm{i}\sin y = -z$，即 1、$\cos\frac{\pi}{3} \mp \mathrm{i}\sin\frac{\pi}{3}$ 其中之一（图中梅子色的点）。所以 $y = 0$、$-\frac{\pi}{3}$ 或 $\frac{\pi}{3}$，再加上 2π 的任意整数倍。

以上每一个都给出三次方程的一个根，所以全部都是解。设 k 为任意整数，解为

$$x = 1, \quad y = \pi + 2k\pi \text{ 或 } y = \pm\tfrac{2\pi}{3} + 2k\pi; \qquad x = -1, \quad y = 2k\pi \text{ 或 } y = \pm\tfrac{\pi}{3} + 2k\pi.$$

在 $[0, 2\pi)$ 内共有六对：$(1, \frac{2\pi}{3})$、$(1, \pi)$、$(1, \frac{4\pi}{3})$、$(-1, 0)$、$(-1, \frac{\pi}{3})$、$(-1, \frac{5\pi}{3})$。

验算

试 $x = -1$、$y = 0$：第一条方程给出 $-1 + 2 - 2 = -1$，第二条是 $0 = 0$。试 $x = 1$、$y = \frac{2\pi}{3}$：各余弦值为 $1, -\frac{1}{2}, -\frac{1}{2}$，得 $1 - 1 - 1 = -1$；各正弦值为 $0, -\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}$，得 $0 - \sqrt{3} + \sqrt{3} = 0$。用电脑在 $x \in [-3, 3]$、$y \in [0, 2\pi)$ 的细密网格上搜索，只在上述六点找到近似解（在舍入误差之内），所以没有遗漏。

常见错误

- 以为 x 是模（modulus），因而 $x \geq 0$。题目容许 x 为任意实数，$x = -1$ 的解同样有效；它们是同一批点 z，只是描述时 y 平移了 π。
- 忘记 $+2k\pi$：y 只确定到相差 2π 的整数倍，所以有无限多对 (x, y)。

STEP 之外

把两条实数方程合并成一条复数方程，是一个标准手法：以 $z = re^{\mathrm{i}\theta}$ 为变量的多项式，其实部和虚部正好满足这类三角方程组，而因式分解 $z^3 + 2z^2 + 2z + 1 = (z + 1)(z^2 + z + 1)$ 则把各根放在单位圆上。在大学的入门课程中，你会见到 $e^{\mathrm{i}\theta}$ 以幂级数（power series）定义，这令棣莫弗定理对所有整数都成立（并赋予非整数次幂意义，但须小心处理分支（branches））。Needham 的 *Visual Complex Analysis* 以优美的几何角度阐述了这一切。

详细解答

STEP II 样本试卷，第 4 题

纯数学 · 双曲函数

STEP II 样本试卷，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/spec-s2-q4

关键思路

这些恒等式（identity）每一条所讲的，都只是一个角（或一个双曲“角”）。先为右边的式子起一个名字，读出它的其中一个函数值，再利用毕氏恒等式（Pythagorean identities）$1+\tan^2\theta=\sec^2\theta$ 和 $1+\sinh^2 u=\cosh^2 u$ 求出另一个。唯一真正要小心的是平方根的正负号，而这正是主值（principal value）范围发挥作用的地方——到了最后一部分，也正是恒等式不再成立的地方。

解答

第 (i) 部分

设 $\theta=\tan^{-1}(\sinh x)$，则 $\tan\theta=\sinh x$，而且由于它是主值，$-\frac{1}{2}\pi<\theta<\frac{1}{2}\pi$。于是

$$\sec^2\theta=1+\tan^2\theta=1+\sinh^2 x=\cosh^2 x.$$

在 $(-\frac{1}{2}\pi,\frac{1}{2}\pi)$ 上有 $\cos\theta>0$，所以 $\sec\theta>0$；而 $\cosh x>0$ 恒成立。取正平方根，得 $\sec\theta=\cosh x$，因此

$$\sin\theta=\tan\theta\cos\theta=\frac{\sinh x}{\cosh x}=\tanh x.$$

由于 θ 位于 $[-\frac{1}{2}\pi,\frac{1}{2}\pi]$ 之内，即 $\sin^{-1}$ 的主值范围，这正好表示 $\theta=\sin^{-1}(\tanh x)$。因此对所有实数 x 都有 $\sin^{-1}(\tanh x)=\tan^{-1}(\sinh x)$，这就是所要证明的结果。

? **想一想** 为甚么得到 $\sin\theta=\tanh x$ 还不算完成论证?

因为 $\sin\theta=\tanh x$ 有无穷多个解 θ。正因为我们的 θ 位于 $[-\frac{1}{2}\pi,\frac{1}{2}\pi]$ 之内，而 sin 在这

个区间上是一对一（one-to-one）的，我们才可以断定 θ 就是主值 $\sin^{-1}(\tanh x)$。同一个范围也决定了 $\sec\theta$ 的正负号。

第 (ii) 部分

设 $u = \sinh^{-1}(\tan y)$；它有定义，因为当 $-\frac{1}{2}\pi < y < \frac{1}{2}\pi$ 时 $\tan y$ 是有限的。于是 $\sinh u = \tan y$，而在第 (i) 部分取 $x = u$，可得

$$\sin^{-1}(\tanh u) = \tan^{-1}(\sinh u) = \tan^{-1}(\tan y) = y,$$

其中最后一步用到 $-\frac{1}{2}\pi < y < \frac{1}{2}\pi$，即 $\tan^{-1}$ 的主值范围。对两边取 sin，得 $\tanh u = \sin y$。由于在这个区间上 $|\sin y| < 1$，$\tanh^{-1}(\sin y)$ 有定义，而且 tanh 是一对一的，所以 $u = \tanh^{-1}(\sin y)$。也就是说，

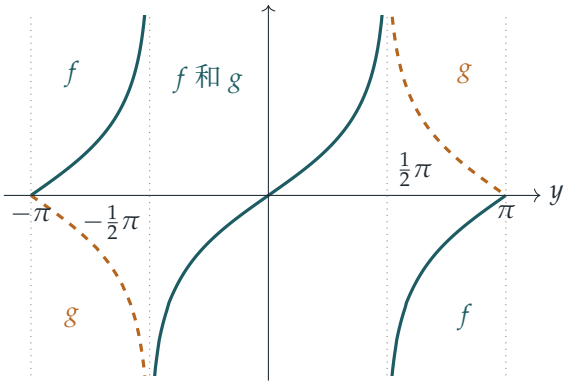
$$\sinh^{-1}(\tan y) = \tanh^{-1}(\sin y) \qquad \left(-\tfrac{1}{2}\pi < y < \tfrac{1}{2}\pi\right).$$

（直接计算：$\cosh u = \sqrt{1+\tan^2 y} = \sec y$，因为 $\cos y > 0$ 所以取正值，于是 $\tanh u = \tan y \cos y = \sin y$。）

草图 记 $f(y) = \sinh^{-1}(\tan y)$ 和 $g(y) = \tanh^{-1}(\sin y)$。两者在 $y = \pm\frac{1}{2}\pi$ 处都没有定义，因为在该处 $\tan y$ 为无穷大，而 $\sin y = \pm 1$。它们在其他地方的表现，可由 tan 和 sin 的对称性推出。

- 两者都是奇函数（odd function），并且都在 $y = 0$ 和 $y = \pm\pi$ 处取零值。
- f 的周期（period）为 π，因为 tan 也是如此。所以在 $(\frac{1}{2}\pi, \pi)$ 上，它重复在 $(-\frac{1}{2}\pi, 0)$ 上的取值：由 $-\infty$ 上升至 0。
- $\sin(\pi - y) = \sin y$，所以 g 关于 $y = \frac{1}{2}\pi$ 对称。在 $(\frac{1}{2}\pi, \pi)$ 上，它以镜像方式重复在 $(0, \frac{1}{2}\pi)$ 上的取值：由 $+\infty$ 下降至 0。
- 求导数，得 $f'(y) = \dfrac{\sec^2 y}{\sqrt{1+\tan^2 y}} = |\sec y|$ 和 $g'(y) = \dfrac{\cos y}{1-\sin^2 y} = \sec y$。所以 f 在每一个有定义的区间上都是递增的，而 g 在 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 上递增，在这个区间以外则递减。两者在原点的斜率都是 1；在 $y = \pi$ 处，f 的斜率为 1，g 的斜率为 -1。

在 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 上，两个图像是同一条曲线；在外侧的区间上，两者互为对方关于横轴的反射。



实线：$f(y) = \sinh^{-1}(\tan y)$。虚线：$g(y) = \tanh^{-1}(\sin y)$ 与 f 不同的部分。点线 $y = \pm\frac{1}{2}\pi$ 是渐近线（asymptote）。

$\frac{1}{2}\pi < y < \pi$ 时的关系 草图显示在该处 $f = -g$，而我们可以证明这一点。若 $\frac{1}{2}\pi < y < \pi$，则 $y - \pi$ 位于 $(-\frac{1}{2}\pi, 0)$ 之内，所以刚才证明的恒等式适用于 $y - \pi$：

$$\begin{aligned}\sinh^{-1}(\tan y) &= \sinh^{-1}\big(\tan(y-\pi)\big)\\ &= \tanh^{-1}\big(\sin(y-\pi)\big)\\ &= \tanh^{-1}(-\sin y) = -\tanh^{-1}(\sin y),\end{aligned}$$

这里用到 $\tan(y-\pi) = \tan y$、$\sin(y-\pi) = -\sin y$，以及 $\tanh^{-1}$ 是奇函数这个事实。所以当 $\frac{1}{2}\pi < y < \pi$ 时，

$$\sinh^{-1}(\tan y) = -\tanh^{-1}(\sin y).$$

（由奇函数的性质，同样的关系在 $-\pi < y < -\frac{1}{2}\pi$ 上也成立。）

想一想 当 $\frac{1}{2}\pi < y < \pi$ 时，第 (ii) 部分的论证在哪一步失效?

在 $\tan^{-1}(\tan y) = y$ 这一步：当 y 在 $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ 以外时，主值是 $y - \pi$，而不是 y。换一个说法，在直接计算的论证中，该处有 $\sqrt{1+\tan^2 y} = |\sec y| = -\sec y$，因为 $\cos y < 0$，而这个负号正是答案中的那个负号。

验算

- 在 $y = 1$：$\sinh^{-1}(\tan 1) = \tanh^{-1}(\sin 1) \approx 1.2262$。在 $y = 2$：$\sinh^{-1}(\tan 2) \approx -1.5235$，而 $\tanh^{-1}(\sin 2) \approx 1.5235$，确认了正负号的改变。
- 正负号：当 $\frac{1}{2}\pi < y < \pi$ 时，$\tan y < 0$ 但 $\sin y > 0$，所以两个式子必定异号；它们不可能相等。

- 第 (i) 部分在 x 很大时：两边都趋向 $\frac{1}{2}\pi$，与预期相符。

我们以数值方法在 10^5 个随机点上检验了两条恒等式，以及 $f = -g$ 在 $\frac{1}{2}\pi < |y| < \pi$ 上的关系。

常见错误

- 写出 $\sqrt{1+\tan^2 y} = \sec y$ 而没有检查 $\sec y$ 的正负号。这正是最后一部分的重点所在。
- 由 $\sin\theta = \tanh x$ 直接断定 $\theta = \sin^{-1}(\tanh x)$，却没有说明为甚么 θ 是主值。
- 把 g 在 $\frac{1}{2}\pi$ 以外的部分画成与 f 相同的形状：g 承袭的是 sin 关于 $\frac{1}{2}\pi$ 的对称性，而不是 tan 的周期性。

STEP 之外

函数 $\mathrm{gd}(x) = \tan^{-1}(\sinh x) = \sin^{-1}(\tanh x)$ 称为*古德曼函数*(Gudermannian function)。它无需借助复数（complex numbers），便把圆函数与双曲函数（hyperbolic functions）联系起来，而第 (ii) 部分说明它的反函数（inverse function）是 $\sinh^{-1}(\tan y)$。它出现在麦卡托（Mercator）地图投影之中，用来把地图上的纵坐标转换为纬度。反函数及其分支（branches）在这里令恒等式出现问题；当你研究复数 z 的 $\log z$ 和 $z^{1/2}$ 时，它们更会成为核心课题。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）是领略这个课题的绝佳读物。

详细解答

STEP II 样本试卷，第 5 题

纯数学 · 级数与积分界限

STEP II 样本试卷，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/spec-s2-q5

关键思路

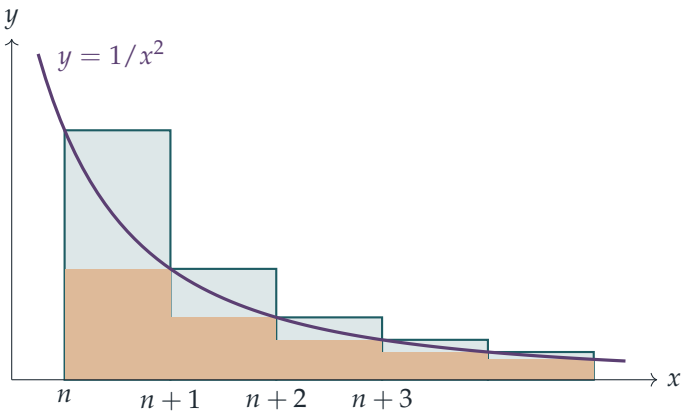
递减函数（decreasing function）各个取值之和，可以想像成一列阔度为一个单位的长方形，而积分（integral）则是曲线下的面积。把长方形放在每个取样点的右边，它们便会凸出于曲线之上；放在左边，它们便会收在曲线之下。这样就得到夹逼的不等式，其后的一切都只是按部就班的整理：级数（series）的尾项和被夹在两个简单的分数之间，所以只要加上对尾项和的一个猜测，估计值便会大幅改善。

解答

夹逼

设 $f(x) = 1/x^2$，它在 $x > 0$ 时为正，并且严格递减（strictly decreasing）。我们还需要

$$\int_n^\infty \frac{\mathrm{d}x}{x^2} = \lim_{M\to\infty}\left[-\frac{1}{x}\right]_n^M = \frac{1}{n}.$$



上方长方形（蓝绿色外框）。在 $[r, r+1]$ 上放一个高度为 $f(r) = 1/r^2$ 的长方形。由于 f 递减，当 $r < x \le r+1$ 时 $f(x) < f(r)$，所以这个长方形的面积 $1/r^2$ 严格大于 $\int_r^{r+1} f(x)\,\mathrm{d}x$。对 $r = n, n+1, \ldots$ 把这些式子相加，得

$$\sum_{r=n}^{\infty} \frac{1}{r^2} > \int_n^{\infty} \frac{\mathrm{d}x}{x^2}.$$

下方长方形（琥珀色）。在 $[r-1, r]$ 上放一个高度为 $f(r) = 1/r^2$ 的长方形。这次当 $r-1 \le x < r$ 时 $f(x) > f(r)$，所以这个长方形位于曲线之下，其面积 $1/r^2$ 严格小于 $\int_{r-1}^{r} f(x)\,\mathrm{d}x$。对 $r = n+1, n+2, \ldots$ 相加，而这些区间刚好填满 $[n, \infty)$，于是

$$\int_n^{\infty} \frac{\mathrm{d}x}{x^2} > \sum_{r=n+1}^{\infty} \frac{1}{r^2}.$$

（三个量都是有限的：积分等于 $1/n$，而两个和都以它为界。）这就是所要求的一对不等式（inequality），在此把它们明确写下来：

$$\sum_{r=n}^{\infty} \frac{1}{r^2} > \frac{1}{n} > \sum_{r=n+1}^{\infty} \frac{1}{r^2}. \qquad (*)$$

夹住尾项和

记 $S_n = \sum_{r=1}^{n} 1/r^2$。那么 $A - S_n$ 正好就是尾项和（tail）$T = \sum_{r=n+1}^{\infty} 1/r^2$。

由 $(*)$ 右边的不等式可直接得到 $T < 1/n$。

> **想一想** $(*)$ 左边的不等式中，求和由 $r = n$ 开始，但尾项和却由 $r = n+1$ 开始。我们怎样才能得到 T 的下界（lower bound）？

不等式 $(*)$ 对每一个正整数都成立，所以把 n 换成 $n+1$，再应用它左边的一半：

$$T = \sum_{r=n+1}^{\infty} \frac{1}{r^2} > \int_{n+1}^{\infty} \frac{\mathrm{d}x}{x^2} = \frac{1}{n+1}.$$

把两者合起来，

$$\frac{1}{n} > A - \sum_{r=1}^{n} \frac{1}{r^2} > \frac{1}{n+1},$$

这正是所要求的结果。

最小的 n

所有项都是正数，所以 $S_n < A$，而误差就是 $A - S_n$。

- 若 $n \le 9999$，则 $n+1 \le 10^4$，所以 $A - S_n > \dfrac{1}{n+1} \ge 10^{-4}$：误差太大。
- 若 $n = 10^4$，则 $A - S_n < \dfrac{1}{n} = 10^{-4}$：误差够小。

所以最小的值是 $n = 10\,000$。留意两个不等式都是严格不等式，这一点在这里十分关键：当 $n = 9999$ 时，下界刚好是 10^{-4}；而当 $n = 10^4$ 时，上界（upper bound）刚好是 10^{-4}。

改良的估计

在夹住尾项和的不等式中，每一部分都减去 $\frac{1}{n+1}$，得

$$\frac{1}{n} - \frac{1}{n+1} > A - \left(\frac{1}{n+1} + \sum_{r=1}^{n} \frac{1}{r^2}\right) > 0.$$

所以 $\frac{1}{n+1} + S_n$ 仍然低估了 A，但相差小于

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

当 $n = 10^4$ 时，这个值是 $\dfrac{1}{10^4 \times 10\,001} < \dfrac{1}{10^4 \times 10^4} = 10^{-8}$，所以改良后的估计值逼近 A 的误差小于 10^{-8}，这正是所要求的结果。

额外加上的一项 $1/(n+1)$ 是对尾项和的一个猜测；既然尾项和被困在一个阔度只有 $1/(n(n+1))$ 的区间之内，这个猜测就不可能错得太远。

验算

- 小数值的情况：取 $n = 1$，则 $S_1 = 1$，而 $A = \pi^2/6 \approx 1.6449$，所以 $A - S_1 \approx 0.645$，确实介乎 $\frac{1}{2}$ 与 1 之间。
- 利用 $A = \pi^2/6$ 并以数值方法求和：$A - S_{9999} \approx 1.00005 \times 10^{-4}$（仅仅太大）而 $A - S_{10000} \approx 0.99995 \times 10^{-4}$（仅仅够小），由此确认 $n = 10\,000$。
- 当 $n = 10^4$ 时,改良后的估计值相差约 5.0×10^{-9},低于界限 $1/(n(n+1)) \approx 1.0 \times 10^{-8}$；它大约是界限的一半，正如图中所示（曲线把每个长方形凸出的部分大致分成两半）。

常见错误

- 在 n 而不是 $n+1$ 处使用左边的不等式，这样得到的是 $\sum_{r=n}^{\infty}$ 的下界，而不是尾项和 $A - S_n$ 的下界。
- 把严格不等式当作非严格不等式，或只用了两个界限的其中一个，因而答 $n = 9999$ 或 $n = 10\,001$。你需要用下界来排除 $n \leq 9999$，并用上界来确定 $n = 10^4$ 可行。
- 把长方形画在每个取样点的错误一边，以致不等式的方向倒转。

STEP 之外

这就是判别收敛（convergence）的积分审敛法（integral test），再加上它在定量方面的额外收获：用来证明级数收敛的同一批长方形，也可以用来估计它的尾项和。把这个想法加以改进，即比较和与积分，再以两端的导数（derivative）作修正，便会引出欧拉–麦克劳林公式（Euler–Maclaurin formula）。在这里，该公式预测 $A - S_n \approx \frac{1}{n} - \frac{1}{2n^2} + \frac{1}{6n^3}$；欧拉（Euler）在证明 $\sum 1/r^2$ 等于 $\pi^2/6$ 之前，正是用这个方法把它计算到小数点后许多位。Abbott 的 *Understanding Analysis* 对级数及其审敛法有细致的论述。

详细解答

STEP II 样本试卷，第 6 题

纯数学 · 由常微分方程推导加法公式

STEP II 样本试卷，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/spec-s2-q6

关键思路

题目不容许我们说“f 就是 e^x”，所以一切都必须由微分方程（differential equation）本身推导出来。我们唯一可以动用的工具是：若一个函数的导数（derivative）处处为零，则该函数为常数。因此窍门在于构造一个导数恰好互相抵消为零的式子，然后代入 $x = 0$ 来确定那个常数。把 $a - x$ 处的值与 x 处的值配对是很自然的做法，因为这两个自变量之和恒为 a，而这正是加法公式（addition formula）的形状。

解答

指数函数的加法定律

固定 a，并设 $p(x) = \mathrm{f}(a-x)\,\mathrm{f}(x)$。由链式法则（chain rule），$\frac{\mathrm{d}}{\mathrm{d}x}[\mathrm{f}(a-x)] = -\mathrm{f}'(a-x) = -\mathrm{f}(a-x)$，其中用到 $\mathrm{f}' = \mathrm{f}$。于是由积法则（product rule）得

$$\begin{aligned} p'(x) &= -\mathrm{f}(a-x)\,\mathrm{f}(x) + \mathrm{f}(a-x)\,\mathrm{f}'(x) \\ &= -\mathrm{f}(a-x)\,\mathrm{f}(x) + \mathrm{f}(a-x)\,\mathrm{f}(x) = 0, \end{aligned}$$

这就是第一个结果。

由于对每个 x 都有 $p'(x) = 0$，函数 p 是常数。它在 $x = 0$ 处的值为 $\mathrm{f}(a)\,\mathrm{f}(0) = \mathrm{f}(a)$，因为 $\mathrm{f}(0) = 1$。因此

$$\mathrm{f}(a-x)\,\mathrm{f}(x) = \mathrm{f}(a) \qquad \text{对所有 } x \text{ 及所有 } a \text{ 成立。}$$

现在任取 x 和 y，并选 $a = x + y$。那么 $a - x = y$，而该恒等式变成

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

正是所求。

余弦的加法公式

我们沿用同一策略，所以需要找出 c 与 s 在 $a-x$ 和 x 处取值的一个组合，使其导数为零。

> **想一想** 我们应该对哪一个组合求导?

让答案来引导我们：取 $a=x+y$ 时，我们想得到 $\mathrm{c}(x)\mathrm{c}(y)-\mathrm{s}(x)\mathrm{s}(y)$，即 $\mathrm{c}(a-x)\mathrm{c}(x)-\mathrm{s}(a-x)\mathrm{s}(x)$。所以这就是值得一试的函数。

固定 a，并设

$$g(x)=\mathrm{c}(a-x)\,\mathrm{c}(x)-\mathrm{s}(a-x)\,\mathrm{s}(x).$$

由链式法则及题目给出的导数，

$$\frac{\mathrm{d}}{\mathrm{d}x}[\mathrm{c}(a-x)]=-\mathrm{c}'(a-x)=\mathrm{s}(a-x),\qquad \frac{\mathrm{d}}{\mathrm{d}x}[\mathrm{s}(a-x)]=-\mathrm{s}'(a-x)=-\mathrm{c}(a-x).$$

因此，由积法则，

$$\begin{aligned} g'(x) &= \big[\mathrm{s}(a-x)\,\mathrm{c}(x)-\mathrm{c}(a-x)\,\mathrm{s}(x)\big]-\big[-\mathrm{c}(a-x)\,\mathrm{s}(x)+\mathrm{s}(a-x)\,\mathrm{c}(x)\big]\\ &= 0. \end{aligned}$$

所以 g 是常数，其值为

$$g(0)=\mathrm{c}(a)\,\mathrm{c}(0)-\mathrm{s}(a)\,\mathrm{s}(0)=\mathrm{c}(a)\cdot 1-\mathrm{s}(a)\cdot 0=\mathrm{c}(a).$$

故此对所有 x 和 a，都有 $\mathrm{c}(a-x)\mathrm{c}(x)-\mathrm{s}(a-x)\mathrm{s}(x)=\mathrm{c}(a)$。任给 x 和 y，取 $a=x+y$，则 $a-x=y$：

$$\mathrm{c}(x+y)=\mathrm{c}(x)\,\mathrm{c}(y)-\mathrm{s}(x)\,\mathrm{s}(y),$$

这就是所求的结果。

以同样的方法处理 $h(x)=\mathrm{s}(a-x)\,\mathrm{c}(x)+\mathrm{c}(a-x)\,\mathrm{s}(x)$，可得 $h'=0$ 及 $h(0)=\mathrm{s}(a)$，所以 $\mathrm{s}(x+y)=\mathrm{s}(x)\mathrm{c}(y)+\mathrm{c}(x)\mathrm{s}(y)$ 亦成立。

验算

- 代入 $y=0$：利用 $\mathrm{f}(0)=\mathrm{c}(0)=1$ 及 $\mathrm{s}(0)=0$，两个结果分别化为 $\mathrm{f}(x)=\mathrm{f}(x)$ 和 $\mathrm{c}(x)=\mathrm{c}(x)$。
- 在余弦的结果中代入 $y=-x$：$1=\mathrm{c}(0)=\mathrm{c}(x)\mathrm{c}(-x)-\mathrm{s}(x)\mathrm{s}(-x)$，这其实就是换了个面貌的 $\cos^2 x+\sin^2 x=1$。

- 我们以数值方法解了这些微分方程（采用龙格—库塔法（Runge–Kutta method），不使用任何内置的指数函数或三角函数），并在 $[-3,3]$ 内随机抽取 60 对数值来检验两个恒等式；例如取 $x=0.7$ 和 $y=1.1$ 时，余弦结果的两边准确至四位小数均为 -0.2272。所有差值都小于 10^{-11}。

常见错误

- 对 $\mathrm{f}(a-x)$、$\mathrm{c}(a-x)$ 或 $\mathrm{s}(a-x)$ 求导时，忘记了链式法则带来的负号。
- 写到“导数为零”便停下来，没有在 $x=0$ 处求出常数的值，也没有接着选取 $a=x+y$。
- 直接引用 $\cos(x+y)$ 的公式或 $\mathrm{e}^{x+y}=\mathrm{e}^{x}\mathrm{e}^{y}$，而这是题目明确禁止的。

STEP 之外

在大学里，这才是定义指数函数和三角函数的严谨做法：把它们定义为微分方程的解（或定义为幂级数（power series）），然后由这个定义推导出每一个熟悉的性质。我们用到的事实，即导数处处为零的函数必为常数，是中值定理（Mean Value Theorem）的推论，而它同时说明了这些解是唯一的。Abbott 的 *Understanding Analysis*（《理解分析》）对这套理论有严谨细致的建构。

详细解答

STEP II 样本试卷，第 7 题

纯数学 · 阶梯函数的积分

> **STEP II 样本试卷，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/spec-s2-q7

关键思路

被积函数（integrand）$[\mathrm{e}^x]$ 是一个阶梯函数（step function）：在 e^x 保持于两个相邻整数之间的每个区间上，它都是常数。因此我们先找出这些区间，在每个区间上对常数积分，然后把结果加起来。这样得出的和看似难以处理，但只要把相同的项对齐，当中大部分对数（logarithm）便会成对抵消。

解答

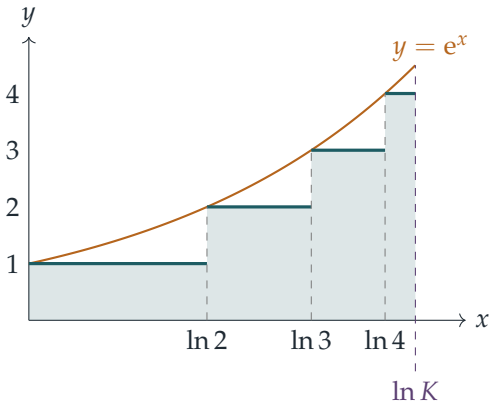
被积函数在哪里是常数

我们取 $K \geq 1$，于是 $\ln K \geq 0$ 且 $N = [K] \geq 1$。（若 $0 < K < 1$，则对于介乎 $\ln K$ 与 0 之间的 x，有 $0 < \mathrm{e}^x \leq 1$，而除了 $x = 0$ 这单独一点之外 $[\mathrm{e}^x] = 0$，所以 $I = 0$；同时 $N = 0$ 且 $N \ln K - \ln 0! = 0$。结果在这情况下同样成立。）

对于整数 $k \geq 1$，

$$k \leq \mathrm{e}^x < k+1 \iff \ln k \leq x < \ln(k+1),$$

因为 ln 是递增函数。所以在 $[\ln k, \ln(k+1))$ 上 $[\mathrm{e}^x] = k$。在积分范围 $0 \leq x \leq \ln K$ 上有 $1 \leq \mathrm{e}^x \leq K$，故 $[\mathrm{e}^x]$ 取值 $1, 2, \ldots, N$，而最后一个值 N 在 $[\ln N, \ln K]$ 上取得，因为在该处 $N \leq \mathrm{e}^x \leq K < N+1$。



阴影部分的面积就是 I（图中取 $K=4.5$，故 $N=4$）。

把各部分加起来

在每个区间上对常数积分，得

$$I=\sum_{k=1}^{N-1} k\big(\ln(k+1)-\ln k\big)+N\big(\ln K-\ln N\big).$$

（当 $N=1$ 时，这个和是空的；当 $K=N$ 时，最后一项为零。两种情况都无须另外处理。）

> **想一想** 怎样化简 $\sum k\big(\ln(k+1)-\ln k\big)$？照现在的形式，它并不能裂项相消（telescope）。

把它拆成两个和，并在第一个和中移动指标，令 $j=k+1$：

$$\sum_{k=1}^{N-1} k\ln(k+1)=\sum_{j=2}^{N}(j-1)\ln j.$$

现在两个和都是 $\ln j$ 的倍数之和，我们可以把每个 $\ln j$ 的系数合并起来。利用 $\ln 1=0$，

$$\begin{aligned}\sum_{k=1}^{N-1} k\big(\ln(k+1)-\ln k\big)&=\sum_{j=2}^{N}(j-1)\ln j-\sum_{j=2}^{N-1} j\ln j\\&=(N-1)\ln N+\sum_{j=2}^{N-1}\big((j-1)-j\big)\ln j\\&=(N-1)\ln N-\sum_{j=2}^{N-1}\ln j.\end{aligned}$$

加上最后一部分 $N\ln K - N\ln N$：

$$\begin{aligned} I &= N\ln K - \ln N - \sum_{j=2}^{N-1} \ln j \\ &= N\ln K - \sum_{j=1}^{N} \ln j \\ &= N\ln K - \ln(N!), \end{aligned}$$

这正是所要证明的结果。

更巧妙的看法

与其把阴影区域切成垂直的长条，不如把它切成水平的层。对于 $1 \le k \le N$，高度介乎 $k-1$ 与 k 之间的那一层，恰好出现在 $[\mathrm{e}^x] \ge k$ 的地方，也就是 $\mathrm{e}^x \ge k$ 的地方，即 $\ln k \le x \le \ln K$。该层的面积是 $\ln K - \ln k$，所以立刻得到

$$I = \sum_{k=1}^{N} (\ln K - \ln k) = N\ln K - \ln(N!)$$

。严格来说，这里用到了当 $1 \le y < N+1$ 时 $[y] = \sum_{k=1}^{N} \mathbf{1}[y \ge k]$，然后逐项积分。这与上面的指标移动是同一种重排，只不过是用几何方式完成。

验算

- $K = 1$：积分范围的长度为零，而公式给出 $1 \cdot 0 - \ln 1 = 0$。
- $K = \mathrm{e}$：这时 $N = 2$，直接计算得 $I = \ln 2 + 2(1 - \ln 2) = 2 - \ln 2 \approx 1.3069$，与 $2\ln \mathrm{e} - \ln 2$ 吻合。
- 当 $K = 10$ 时，公式给出 $10\ln 10 - \ln(10!) \approx 7.9214$，而对 $[\mathrm{e}^x]$ 在 $[0, \ln 10]$ 上用中点法则作数值积分，结果相差不超过 5×10^{-6}。在 $[1, 40]$ 中随机选取的 K 值亦同样吻合。

常见错误

- 把和一直加到 N，并使用区间 $[\ln N, \ln(N+1)]$。积分范围的上端是 $\ln K$，而不是 $\ln(N+1)$，所以最后一部分必须另外处理。
- 重新编排指标时弄错首尾两端的项：不妨把 $N = 3$ 时的各个和逐项写出来检查。

STEP 之外

把 $\sum k(a_{k+1}-a_k)$ 换成 $-\sum a_k$ 再加上边界项，这就是*分部求和法*（summation by parts，又称阿贝尔求和法（Abel summation）），它是分部积分法（integration by parts）的离散版本。取 $K=N$ 并利用 $\mathrm{e}^x-1<[\mathrm{e}^x]\leq \mathrm{e}^x$，可得 $N\ln N-N+1\leq \ln(N!)\leq N\ln N-N+1+\ln N$，这是迈向斯特灵近似公式（Stirling's approximation）的第一步。同样的想法写成黎曼–斯蒂尔吉斯积分（Riemann–Stieltjes integral）$\int f\,\mathrm{d}[t]$ 之后，在解析数论（analytic number theory）中随处可见，用来把和化为积分；参看 Apostol 的 *Introduction to Analytic Number Theory*（解析数论入门）。

详细解答

STEP II 样本试卷，第 8 题

纯数学 · 模群中元素的阶

> **STEP II 样本试卷，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/spec-s2-q8

关键思路

模（modulo）N 之下，$N-1$ 这个数就是 -1，而 $m+n=N$ 表示 $n\equiv -m$。所以 n 的每个幂都是 m 的幂，最多相差一个正负号：$n^k\equiv(-1)^k m^k$。k 是偶数还是奇数，决定了这个正负号是否有影响，而题目的三个部分正好对应两个阶（order）的奇偶性的三种组合方式。最后一部分则利用这些事实求出各元素的阶，而无须逐一计算每个幂。

解答

全文中，$\equiv$ 表示模 N 同余（congruent）。我们按题目的原意，取 G 的单位元（identity）为 1。（事实上，只要 $N-1\in G$，这一点就必然成立：若 e 是单位元，则 $e(N-1)\equiv N-1$，即 $-e\equiv -1$，所以 $e=1$。）

$N-1$ 的阶

我们有

$$(N-1)^2=N^2-2N+1\equiv 1,$$

而且由于 $N\geq 3$，故 $N-1\neq 1$。所以 $(N-1)^1$ 不是单位元，但 $(N-1)^2$ 是：$N-1$ 的阶为 2。

各部分的准备工作

由于 $m+n=N$，我们有 $n\equiv -m$，因此对每个正整数 k 都有

$$n^k\equiv(-1)^k m^k,\qquad \text{等价地}\qquad m^k\equiv(-1)^k n^k.$$

此外，$N-1 \equiv -1 \equiv n\,m^{-1}$ 属于 G，因为 G 是一个包含 n 和 m^{-1} 的群（group）。所以 -1 是 G 的元素，而且由于 $N \geq 3$，故 $-1 \neq 1$。我们还会用到以下的标准结果：若 x 的阶为 r 且 $x^k = 1$，则 $r \mid k$（写成 $k = ar + b$，其中 $0 \leq b < r$；于是 $x^b = 1$，由 r 的最小性得 $b = 0$）。

第 (i) 部分

设 p 和 q 都是偶数。那么

$$n^p \equiv (-1)^p m^p = m^p \equiv 1,$$

所以 $q \mid p$。同理，$m^q \equiv (-1)^q n^q = n^q \equiv 1$，所以 $p \mid q$。两个正整数互相整除，必定相等，所以 $p = q$。

第 (ii) 部分

设 p 是偶数而 q 是奇数。由于 p 是偶数，$n^p \equiv m^p \equiv 1$，所以 $q \mid p$。另外 $2q$ 是偶数，所以

$$m^{2q} \equiv n^{2q} = (n^q)^2 \equiv 1,$$

由此得 $p \mid 2q$。写成 $p = kq$；于是 $kq \mid 2q$，所以 $k \mid 2$，即 $k = 1$ 或 2。但若 $k = 1$，则 $p = q$ 是奇数，与假设不符。因此 $k = 2$，即 $p = 2q$。

> ? **想一想** 为甚么不能直接由 $m^q \equiv -n^q \equiv -1$ 推断“p 必定是 $2q$”?

这个计算是正确的，它说明 $m^q \neq 1$ 而 $m^{2q} \equiv 1$，但单凭这一点只能得出 $p \mid 2q$ 和 $p \nmid q$，并不足以确定 p 的值（例如原则上 p 可以是 2 而 $q = 3$）。要补上这个缺口，还需要另一个整除关系 $q \mid p$，而它正是来自 p 是偶数这个条件。

第 (iii) 部分

用反证法（proof by contradiction）：假设 p 和 q 都是奇数，并设 $r = pq$，它也是奇数。那么

$$n^r = (n^q)^p \equiv 1, \qquad \text{但同时} \qquad n^r \equiv (-1)^r m^r = -(m^p)^q \equiv -1.$$

所以 $1 \equiv -1$，即 $N \mid 2$，但 $N \geq 3$，这不可能。因此 p 和 q 不可能同时是奇数。

（根据对称性，第 (ii) 部分也说明：当 q 是偶数而 p 是奇数时，$q = 2p$。所以在任何情况下，m 和 $N - m$ 的阶要么相等且为偶数，要么其中一个是另一个的两倍，而后者是奇数。）

模 21 的群

2 和 5 的各次幂模 21 化简后，依次为

k	1	2	3	4	5	6
2^k	2	4	8	16	11	1
5^k	5	4	20	16	17	1

（例如 $2^5 = 32 \equiv 11$，$5^2 = 25 \equiv 4$，$5^4 \equiv 4^2 = 16$，$5^5 \equiv 80 \equiv 17$）。两行中的 1 都是在 $k = 6$ 时首次出现，所以 2 和 5 的阶都是 6。

现在从表中读出各元素的阶：若 x 的阶为 6，则 x^k 的阶为 $6/\gcd(k,6)$。所以

- $11 = 2^5$ 和 $17 = 5^5$ 的阶为 6；
- $4 = 2^2$ 和 $16 = 2^4$ 的阶为 3；
- $8 = 2^3$ 和 $20 = 5^3$ 的阶为 2（由于 $20 = N - 1$，题目的第一部分已经告诉我们这一点）；
- 1 的阶为 1。

余下的是 10、13 和 19，它们各自与一个已知的元素相加得 21：

- $13 = 21 - 8$，而 8 的阶是偶数 2。根据 (iii) 及其后的备注，13 的阶要么是 2（相等且为偶数），要么是 1（即 2 的一半）。它不是 1，因为 $13 \neq 1$，所以 13 的阶为 2。（验算：$13^2 = 169 = 8 \cdot 21 + 1$。）
- $19 = 21 - 2$ 及 $10 = 21 - 11$，而 2 和 11 的阶都是 6。所以 19 和 10 各自的阶是 6 或 3。

所以 G 的每个元素的阶都不超过 6。阶为 12 的循环群（cyclic group）有一个生成元（generator），即一个阶为 12 的元素，而同构（isomorphism）保持元素的阶不变。由于 G 没有阶为 12 的元素，它与阶为 12 的循环群并不同构。

还有另一种方法完成这一步：G 有三个阶为 2 的元素，即 8、13 和 20，而阶为 12 的循环群恰好只有一个（若 g 是它的生成元，则 g^k 的阶为 2 仅当 $k = 6$）。

验算

用电脑验算，可得 G 中各元素的阶的完整列表：1 的阶为 1；$8, 13, 20$ 的阶为 2；$4, 16$ 的阶为 3；而 $2, 5, 10, 11, 17, 19$ 的阶为 6。（直接计算，$10^3 = 1000 \equiv 13$ 及 $19^3 \equiv (-2)^3 = -8 \equiv 13$，所以两者的阶都不是 3。）各类元素的数目 $1 + 3 + 2 + 6 = 12$ 与 G 的大小吻合。我们亦对所有 $N < 60$，在单位群中由最多两个元素生成的每个子群（subgroup）里，就每一对 m、$N - m$ 检验了第 (i) 至 (iii) 部分：共近 1900 对，无一例外。

常见错误

- 假设 $N-1 \in G$ 而不说明原因。题目并没有给出这一点，但它可由 $m, n \in G$ 推出，因为 $N-1 \equiv n\,m^{-1}$。
- 在 (ii) 中只证明了 $p \mid 2q$；还需要多一个整除关系才能得到等式。
- 在古怪的情况下把单位元视为理所当然：例如模 6 之下的子集 $\{3\}$ 对乘法封闭，其单位元为 3，而取 $m = n = 3$ 时便会违反 (iii)。本题的论证依赖单位元是 1，而这显然正是题目的原意。

STEP 之外

G 是模 21 的单位群（group of units），根据中国剩余定理（Chinese remainder theorem），它同构于模 3 和模 7 的单位群的直积，即 $C_2 \times C_6$：这就是为甚么它有三个阶为 2 的元素，却没有阶为 12 的元素。判断模 N 的单位在甚么时候构成循环群（恰好是 $N = 2, 4, p^k, 2p^k$ 的情况，其中 p 是奇质数，亦即存在“原根”（primitive root）的情况），是大学一年级数论的经典结果。Davenport 的 *The Higher Arithmetic* 讲述了原根，而 Allenby 的 *Rings, Fields and Groups* 则讨论了相关的群论。

详细解答

STEP II 样本试卷，第 9 题

纯数学 · 球面、平面与反射

> **STEP II 样本试卷，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/spec-s2-q9

关键思路

两部分其实都围绕同一幅图：一个平面，以及向它所作的垂线。在 (i) 中，球面的球心在原点，所以一切都取决于由原点到平面的垂足（foot of the perpendicular）：它与 O 的距离决定两者是否相交，然后用毕氏定理（Pythagoras' theorem）便可求出那个圆。在 (ii) 中，把 $\mathbf{x}$ 分拆成沿 $\mathbf{c}$ 方向的部分和垂直于 $\mathbf{c}$ 的部分，再看看公式对每一部分做了甚么。

解答

第 (i) 部分

平面 $(\mathbf{x}-\mathbf{a})\cdot\mathbf{b}=0$ 通过点 $\mathbf{a}$，并以 $\mathbf{b}$ 为法向量（normal）。设 N 为由原点 O 到该平面的垂足。由于 ON 沿法向量方向，故对某个纯量（scalar）λ 有 $\mathbf{n}=\lambda\mathbf{b}$；又因为 N 在平面上，

$$(\lambda\mathbf{b}-\mathbf{a})\cdot\mathbf{b}=0 \quad\Longrightarrow\quad \lambda=\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2},$$

这一步是容许的，因为 $\mathbf{b}\neq\mathbf{0}$。所以

$$\mathbf{n}=\frac{(\mathbf{a}\cdot\mathbf{b})\,\mathbf{b}}{|\mathbf{b}|^2}, \qquad d=|\mathbf{n}|=\frac{|\mathbf{a}\cdot\mathbf{b}|}{|\mathbf{b}|},$$

其中 d 是由 O 到平面的距离。

现在取平面上任意一点 $\mathbf{x}$。向量 $\mathbf{x}-\mathbf{n}$ 位于平面之内，所以它垂直于 $\mathbf{b}$，因而也垂直于 $\mathbf{n}$。由毕氏定理，

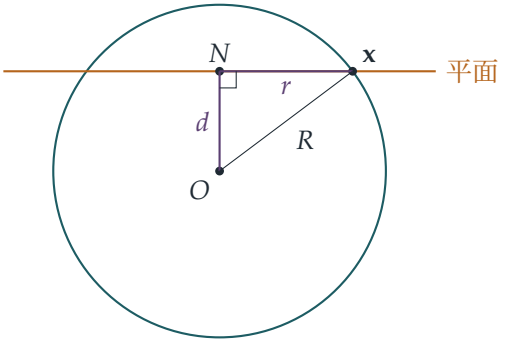
$$|\mathbf{x}|^2=|\mathbf{n}|^2+|\mathbf{x}-\mathbf{n}|^2=d^2+|\mathbf{x}-\mathbf{n}|^2. \qquad (*)$$

所以平面上的点都满足 $|\mathbf{x}|\geq d$，而且每一个满足 $|\mathbf{x}|\geq d$ 的值都会出现（只要在平面内

由 N 向外移动)。因此，球面 $|\mathbf{x}| = R$ 与平面相交，当且仅当 $R \geq d$，即

$$\boxed{|\mathbf{a}\cdot\mathbf{b}| \leq R\,|\mathbf{b}|} \qquad \text{或等价地} \qquad (\mathbf{a}\cdot\mathbf{b})^2 \leq R^2|\mathbf{b}|^2.$$

在等号成立的情况下，该平面是切平面（tangent plane），而那个“圆”则缩成单独一点 N。



当条件成立时，同时在两个曲面上的点满足 $|\mathbf{x}| = R$，并且由 $(*)$ 得

$$|\mathbf{x}-\mathbf{n}|^2 = R^2 - d^2.$$

所以这些点都在平面内，并且与 N 保持固定距离，而后者本身是平面上的一点：它们组成一个以 N 为圆心的圆。其圆心和半径为

$$\mathbf{n} = \frac{(\mathbf{a}\cdot\mathbf{b})\,\mathbf{b}}{|\mathbf{b}|^2}, \qquad r = \sqrt{R^2 - \frac{(\mathbf{a}\cdot\mathbf{b})^2}{|\mathbf{b}|^2}} = \frac{\sqrt{R^2|\mathbf{b}|^2 - (\mathbf{a}\cdot\mathbf{b})^2}}{|\mathbf{b}|}.$$

第 (ii) 部分

记 $\hat{\mathbf{c}} = \mathbf{c}/|\mathbf{c}|$，这是一个单位向量（unit vector）（由于 $\mathbf{c} \neq \mathbf{0}$，这样做没有问题）。于是该关系式变成

$$\mathbf{x}' = \mathbf{x} - 2(\mathbf{x}\cdot\hat{\mathbf{c}})\,\hat{\mathbf{c}}.$$

? **想一想** 从几何上看，$(\mathbf{x}\cdot\hat{\mathbf{c}})\,\hat{\mathbf{c}}$ 是甚么?

它是 $\mathbf{x}$ 沿 $\mathbf{c}$ 方向的分量（component），亦即 $\mathbf{x}$ 在通过 O 而方向为 $\mathbf{c}$ 的直线上的投影（projection）。所以可作分拆

$$\mathbf{x} = \mathbf{x}_{\parallel} + \mathbf{x}_{\perp}, \qquad \mathbf{x}_{\parallel} = (\mathbf{x}\cdot\hat{\mathbf{c}})\,\hat{\mathbf{c}}, \qquad \mathbf{x}_{\perp} = \mathbf{x} - \mathbf{x}_{\parallel},$$

其中 $\mathbf{x}_{\perp}\cdot\mathbf{c} = \mathbf{x}\cdot\mathbf{c} - (\mathbf{x}\cdot\hat{\mathbf{c}})(\hat{\mathbf{c}}\cdot\mathbf{c}) = 0$。关系式变成

$$\mathbf{x}' = \mathbf{x}_{\perp} - \mathbf{x}_{\parallel}.$$

设 Π 为通过 O 并垂直于 $\mathbf{c}$ 的平面（即平面 $\mathbf{x}\cdot\mathbf{c}=0$），则位于这个平面内的分量保持不变，而垂直于 Π 的分量则反转方向。这正正是镜子的作用：**$\mathbf{x}'$ 是 $\mathbf{x}$ 关于平面 $\mathbf{x}\cdot\mathbf{c}=0$ 的反射（reflection）**，即关于通过原点、以 $\mathbf{c}$ 为法向量的平面的反射。

以下按反射的定义加以确认：

- $\mathbf{x}'-\mathbf{x}=-2\mathbf{x}_{\parallel}$ 是 $\mathbf{c}$ 的倍数，所以由 $\mathbf{x}$ 到 $\mathbf{x}'$ 的线段垂直于 Π；
- 中点 $\frac{1}{2}(\mathbf{x}+\mathbf{x}')=\mathbf{x}_{\perp}$ 满足 $\mathbf{x}_{\perp}\cdot\mathbf{c}=0$，所以它在 Π 上。

以下几个推论可令这幅图像更具体。Π 上的点（即满足 $\mathbf{x}\cdot\mathbf{c}=0$ 的点）固定不动；$\mathbf{c}$ 本身变成 $-\mathbf{c}$；$|\mathbf{x}'|=|\mathbf{x}|$，因为 $|\mathbf{x}'|^2=|\mathbf{x}_{\perp}|^2+|\mathbf{x}_{\parallel}|^2$；而把这个映射连续施行两次便会得回 $\mathbf{x}$，反射两次理应如此。若把 $\mathbf{c}$ 换成任何非零倍数 $k\mathbf{c}$，关系式并无改变，这也很合理：起作用的只有那个平面，而不是其法向量的长度或指向。

验算

在 (i) 中，取 $\mathbf{a}=\mathbf{b}=(0,0,1)$ 及 $R=2$：平面是 $z=1$，它与球面 $x^2+y^2+z^2=4$ 相交于圆 $x^2+y^2=3$，$z=1$。公式给出圆心 $(0,0,1)$ 和半径 $\sqrt{4-1}=\sqrt{3}$，与预期相符。在 (ii) 中，取 $\mathbf{c}=(0,0,1)$：公式把 (x,y,z) 映射到 $(x,y,-z)$，即关于 $z=0$ 的反射。我们亦以 20,000 组随机选取的 $\mathbf{a}$、$\mathbf{b}$、R 以及 $\mathbf{x}$、$\mathbf{c}$ 对两部分作了数值测试：在所预测的圆上构作的点，都在球面和平面上，误差不超过 10^{-9}；而 (ii) 中的映射使 Π 内的向量固定不动，保持长度和点积（dot product）不变，并且是它自己的逆映射。

常见错误

- 忘记加上绝对值（modulus）：条件是 $|\mathbf{a}\cdot\mathbf{b}|\le R|\mathbf{b}|$，而不是 $\mathbf{a}\cdot\mathbf{b}\le R|\mathbf{b}|$；位于 O 另一侧的平面有 $\mathbf{a}\cdot\mathbf{b}<0$。
- 把 $\mathbf{a}$ 当作圆心。点 $\mathbf{a}$ 只不过是平面上的某一点；圆心是由 O 所作垂线的垂足。
- 在 (ii) 中，说成“关于 $\mathbf{c}$ 的反射”或“关于沿 $\mathbf{c}$ 的直线的反射”。关于该直线的反射应是 $\mathbf{x}\mapsto 2\mathbf{x}_{\parallel}-\mathbf{x}$，亦即 $-\mathbf{x}'$。

STEP 之外

(ii) 中的映射是一个豪斯霍尔德反射（*Householder reflection*），$\mathbf{x}\mapsto(I-2\hat{\mathbf{c}}\hat{\mathbf{c}}^{\mathsf{T}})\mathbf{x}$，它是一个线性映射（linear map），既正交（orthogonal）又对称（symmetric），行列式（determinant）为 -1。空间中的每一个旋转都是两个这类反射的乘积，而数值分析学者利用它们来稳定地分解矩阵（QR 分解，QR decomposition）。Strang 的 *Introduction to Linear Algebra*（线

性代数入门）正是从这个观点讲解投影、反射和正交矩阵。

详细解答

STEP II 样本试卷，第 10 题

纯数学 • 逻辑谜题

> **STEP II 样本试卷，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/spec-s2-q10

关键思路

一切都建基于一个转译：说话者是蟾蜍，当且仅当他所说的话为真。因此，对每个角色我们都问“他可能是蟾蜍吗?”和“他可能是青蛙吗?”，然后剔除任何令其陈述的真值（truth value）出错的选择。尤其要留意那些关于说话者自己的陈述：其中有些是某一种生物根本不可能说出口的，而这通常就是突破口。

解答

在整个解答中，“蟾蜍”指说真话者，“青蛙”指说谎者，所以

$$\text{说话者是蟾蜍} \iff \text{其陈述为真.}$$

在每个场景中，我们找出所有符合这条规则的指派（assignment），并验证恰好只有一个能够成立。

第 (i) 部分

假设 A 是蟾蜍。那么 A 的陈述为真，于是 A 便是青蛙，这是一个矛盾（contradiction）。因此 A 是青蛙。

所以 A 的陈述为假：A 和 B 并非都是青蛙。既然 A 是青蛙，这句话要不成立，唯一的可能就是 B 是蟾蜍。

验算：若 A 是青蛙而 B 是蟾蜍，则“我们两个都是青蛙”为假，正符合青蛙的陈述必须为假的要求。所以 **A 是青蛙，B 是蟾蜍**。

第 (ii) 部分

这是第 (i) 部分的镜像。假设 C 是青蛙。那么“D 和我之中至少有一个是青蛙”便为真（单凭 C 已足够），但青蛙不可能说出真话。因此 C 是蟾蜍。

所以 C 的陈述为真：C、D 之中至少有一个是青蛙。既然 C 是蟾蜍，D 必定就是那只青蛙。

验算：若 C 是蟾蜍而 D 是青蛙，则该陈述为真，正符合蟾蜍的陈述必须为真的要求。所以 **C 是蟾蜍，D 是青蛙**。

第 (iii) 部分

以 S 表示 E 的陈述“G 和 H 都是蟾蜍”。那么 G 断言 S 为真，而 H 断言 S 为假。

> ? **想一想** 我们对任何一个角色都还一无所知。到底可以从哪里入手?

不要先看 E，而是把 G 和 H 放在一起看。他们两个的陈述恰好相反，所以其中恰有一个为真。因此，不论 S 最终是真是假，G、H 之中恰有一个是蟾蜍，另一个是青蛙。

这就确定了 S 的真假：既然 G、H 之中有一个是青蛙，他们便不可能都是蟾蜍，所以 S 为假。于是：

- E 说的话为假，所以 E 是青蛙；
- G 声称 S 为真，而这是假的，所以 G 是青蛙；
- H 声称 S 为假，而这是真的，所以 H 是蟾蜍。

验算：G 是青蛙，所以“G 和 H 都是蟾蜍”确实为假，与每一个结论都吻合。所以 **E 和 G 是青蛙，H 是蟾蜍**。

第 (iv) 部分

这里的“我们”是指 I、J、K 三个。

I 是青蛙。假如 I 是蟾蜍，“我们全部都是青蛙”便为真，于是 I 是青蛙，矛盾。所以 I 是青蛙，而 I 的陈述为假：三个并非全是青蛙，故此

$$\text{在} J,\ K \text{ 之中至少有一个是蟾蜍.} \qquad (*)$$

J 是蟾蜍。反过来假设 J 是青蛙。那么由 $(*)$ 可知 K 是蟾蜍，三者分别是 I 青蛙、J 青蛙、K 蟾蜍：恰好有一只蟾蜍。这使 J 的陈述为真，而青蛙不可能说出真话。所以 J 是蟾蜍。

K 是青蛙。既然 J 是蟾蜍，J 的陈述便为真：三个之中恰好有一只蟾蜍。那只蟾蜍就是 J，所以 K 是青蛙。

想一想 现在知道 K 是青蛙，$(*)$ 还成立吗?

成立：$(*)$ 只要求 J、K 之中有一个是蟾蜍，而 J 正是。所以完整的验算如下：I 的“我们全部都是青蛙”为假（J 是蟾蜍），而 J 的“我们之中恰好有一个是蟾蜍”为真（只有 J）。两个陈述的真值都正确，所以 **I 和 K 是青蛙，J 是蟾蜍**。

验算

每个场景只有 2^2 或 2^3 种可能的指派，所以完整的验算不过是一张简短的真值表（truth table）：对每一种指派，检验是否每个说话者的陈述都恰好在该说话者是蟾蜍时为真。我们对全部四个场景都做了这样的穷举搜索，每个场景中都恰好只有一种指派能够成立，就是上面找到的那一种。（唯一性很重要：题目要求我们判定，所以我们必须排除其他指派，而不只是找出一个可行的。）

常见错误

- 找到一个没有矛盾的指派就停下来。你还必须证明其他指派都不可行，而这正是那些“假设他是...”的反证所起的作用。
- 在第 (iii) 部分，把 G 的“这是真的”当作关于 G 自己的陈述，而不是关于 E 的说法的陈述。
- 在第 (iv) 部分，忘记了 K 虽然一言不发，却仍然包括在“我们全部”和“我们之中恰好有一个”之内。

STEP 之外

这类谜题其实是命题逻辑（propositional logic）的练习：每只生物是一个布尔变量（Boolean variable）t，每个陈述是一条公式 P，而每个说话者都施加了约束条件 $t \iff P$。判定这样一组约束条件是否有解，就是可满足性（satisfiability）问题，它是理论计算机科学的核心课题；而第 (i) 部分那句自我否定的陈述，则是说谎者悖论（liar paradox）的一位温和的近亲。斯穆里安（Raymond Smullyan）的 *What Is the Name of This Book?*（《这本书叫甚么名字?》）是一本妙趣横生的骑士与无赖（knights and knaves）谜题集，一步步引领读者走向哥德尔（Gödel）的不完备定理（incompleteness theorem）。

详细解答

STEP II 样本试卷，第 11 题

力学 • 抛体与碰撞

STEP II 样本试卷，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：

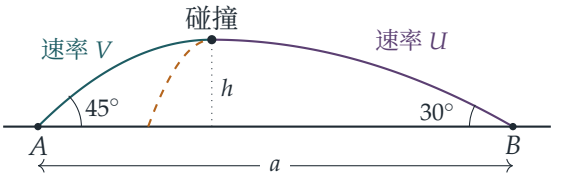
step.maths.org/questions/spec-s2-q11

关键思路

“两个质点相撞时各自位于最高点”其实包含了两个条件：两个质点（particle）必须用相同的时间到达最高点，而且那一刻必须位于同一位置。时间条件把 U 和 V 联系起来，而两段水平距离之和就是 a。碰撞之后，最要紧的观察是：两个质点在最高点时都沿水平方向运动，所以合并后的质点起初完全没有垂直速度。

解答

我们把两个质点都当作只受重力作用（不计空气阻力）的抛体（projectile），在包含 AB 的垂直平面内运动。水平方向取由 A 到 B 为正，垂直方向取向上为正。



证明 $U^2 = 2V^2$

由 A 抛出的质点，初速度的水平分量为 $\frac{V}{\sqrt{2}}$，垂直分量为 $\frac{V}{\sqrt{2}}$。由 B 抛出的质点，水平分量为 $\frac{\sqrt{3}}{2}U$（指向 A），垂直分量为 $\frac{1}{2}U$。

抛体的垂直速度为零时，它正位于最高点。若初始垂直速率为 w，这会在时间 w/g 发生。两个质点在同一瞬间抛出，所以如果它们在碰撞的一刻都位于各自的最高点，这两个时

间必须相等：

$$\frac{V}{\sqrt{2}\,g} = \frac{U}{2g} \quad\Longrightarrow\quad U = \sqrt{2}\,V \quad\Longrightarrow\quad U^2 = 2V^2,$$

正如所求。把这个共同的时间记为 $T = \dfrac{V}{\sqrt{2}\,g}$。

想一想 我们不是也应该检查两个最高点的高度是否相同吗?

应该：时间相等是必要条件，但两个质点还必须相遇。初始垂直速率为 w 的抛体会上升到高度 $w^2/(2g)$。对 A 的质点而言，这是 $\frac{V^2}{2}/2g = \frac{V^2}{4g}$；对 B 的质点而言，这是 $\frac{U^2}{4}/2g = \frac{2V^2}{8g} = \frac{V^2}{4g}$。两者一致，所以只要 $U^2 = 2V^2$，高度便自动相同。（事实上这是必然的：两者在同一时间 T 的垂直速度都是零，而且加速度相同，所以两者的高度都是 $\frac{1}{2}gT^2$。）记

$$h = \frac{V^2}{4g}$$

为碰撞时的高度。

求 a

在时间 T，两者的水平位置也必须一致。水平速度保持不变，所以在时间 T 内，由 A 抛出的质点向 B 移动了 $\frac{V}{\sqrt{2}}T$，由 B 抛出的质点向 A 移动了 $\frac{\sqrt{3}}{2}UT$。两者相遇，所以这两段距离之和为 a：

$$\begin{aligned} a &= \left(\frac{V}{\sqrt{2}} + \frac{\sqrt{3}}{2}\cdot\sqrt{2}\,V\right)\frac{V}{\sqrt{2}\,g} \\ &= \frac{V^2}{g}\left(\frac{1}{2} + \frac{\sqrt{3}}{2}\right) = \frac{V^2}{2g}(1+\sqrt{3}), \end{aligned}$$

这就是所求的结果。

碰撞之后

在碰撞的一瞬间，每个质点都位于自己的最高点，所以都沿水平方向运动：A 的质点速度为 $+\frac{V}{\sqrt{2}}$，B 的质点速度为 $-\frac{\sqrt{3}}{2}U = -\frac{\sqrt{3}\,V}{\sqrt{2}}$。

想一想 哪一个动量守恒？在哪个方向守恒?

在极短的碰撞过程中，唯一的巨大作用力是两个质点之间大小相等、方向相反的力；重力大小有限，而且作用时间可以忽略。因此总动量（momentum）守恒。在水平方向，设质量为 $2m$ 的合并质点的速度为 v，则

$$2m\,v = m\frac{V}{\sqrt{2}} - m\frac{\sqrt{3}\,V}{\sqrt{2}} \quad\Longrightarrow\quad v = -\frac{(\sqrt{3}-1)V}{2\sqrt{2}}.$$

负号表示合并质点朝 A 的方向往回运动，速率为 $\dfrac{(\sqrt{3}-1)V}{2\sqrt{2}}$。在垂直方向，两个质点的垂直速度都是零，所以合并质点起初的垂直速度也是零。

合并质点现在是一个由高度 $h = \frac{V^2}{4g}$ 水平抛出的抛体。它的水平速度维持在 $|v|$。它到达平面时的垂直速率 w 可由 $w^2 = 0 + 2gh$ 求得：

$$w^2 = 2g \cdot \frac{V^2}{4g} = \frac{V^2}{2}, \qquad w = \frac{V}{\sqrt{2}}.$$

因此，速度与水平方向所成的角 ϕ 满足

$$\begin{aligned}\tan\phi = \frac{w}{|v|} &= \frac{V}{\sqrt{2}} \cdot \frac{2\sqrt{2}}{(\sqrt{3}-1)V} = \frac{2}{\sqrt{3}-1} \\ &= \frac{2(\sqrt{3}+1)}{(\sqrt{3}-1)(\sqrt{3}+1)} = \frac{2(\sqrt{3}+1)}{2} = 1+\sqrt{3},\end{aligned}$$

其中第二行把分母有理化（rationalising the denominator）。至此全题完成。

验算

- *单位。* V^2/g 的单位是 $(\mathrm{m\,s^{-1}})^2/(\mathrm{m\,s^{-2}}) = \mathrm{m}$，正是距离应有的单位；而 $\tan\phi$ 是两个速率之比，所以没有量纲。留意 $\tan\phi$ 完全不取决于 V 或 g，这是个好兆头。
- *几何上是否合理。* A 的质点到达最高点时，与 A 的水平距离为 $\frac{V^2}{2g}$；B 的质点到达最高点时，与 B 的水平距离为 $\sqrt{3}\,\frac{V^2}{2g}$。以较大速率、较平的 30° 角抛出的质点走得较远，而两段距离之和为 a。
- *数值。* 取 $V = 7.3\ \mathrm{m\,s^{-1}}$ 和 $g = 9.8\ \mathrm{m\,s^{-2}}$，我们把两条轨迹按时间逐步推算，发现在 $t = T$ 时两者位于同一点 $(2.719, 1.359)$（以米为单位），而且两者的垂直速度都是零。再把合并质点逐步推算至落到平面，得出 $\tan\phi \approx 2.73205$，与 $1+\sqrt{3} = 2.732050\ldots$ 吻合。

常见错误

- 在垂直方向应用动量守恒时，用了非零的垂直速度。在每条轨迹的最高点，垂直速度都是零，这正是这次碰撞如此简单的原因。
- 把水平动量相加而不是相减：两个质点正朝相反方向运动。
- 用了 $\tan\phi = |v|/w$（与垂直方向所成的角），而不是 $w/|v|$。

STEP 之外

质点合而为一，是完全非弹性碰撞（perfectly inelastic collision）最简单的例子：动量守恒，但动能（kinetic energy）并不守恒（这里损失了相当多，因为两个质点原本朝相反方向运动）。到了大学，这会发展成在质心参考系（centre-of-mass frame）中研究碰撞：在这个参考系中总动量为零，而损失的能量就是相对于质心的全部动能。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学导论）正是以这种方式处理动量和碰撞，并附有许多出色的习题。

详细解答

STEP II 样本试卷，第 12 题

力学 • 金属线上的振动

STEP II 样本试卷，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：

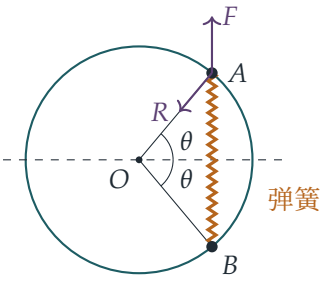
step.maths.org/questions/spec-s2-q12

关键思路

金属线是光滑的，所以只有弹簧的力做功，机械能（mechanical energy）守恒。我们只需要用 θ 表示弹簧的长度（圆的一条弦（chord）），以及每颗珠子的速率。有两件事要稍作思考。为甚么平分角 AOB 的直线会保持固定？还有在第 (ii) 部分，如果珠子有足够的能量到达直径的两端，会发生甚么事？

解答

把珠子看成质点，它们套在水平面上的光滑圆形金属线上，并由一条遵从胡克定律（Hooke's law）的轻弹簧连接。重力沿铅垂方向，与金属线反作用力（reaction）的铅垂分量互相平衡，所以对水平方向的运动没有影响。



图中，F 是弹簧对 A 的推力（沿 BA 方向，因为弹簧起初处于压缩状态），R 是金属线对 A 的反作用力的水平分量。

建立能量方程

用圆上的角度来量度每颗珠子的位置：A 位于角 ϕ_A，B 位于角 ϕ_B。

> **想一想** 为甚么可以假设两颗珠子以相同的速率运动，并且关于一条固定的直线对称?

对通过 O 的铅垂轴取力矩（moments）。弹簧作用于 A 和 B 的力大小相等、方向相反，并且沿同一条直线 AB 作用，所以它们的力矩互相抵销。每个金属线反作用力的水平分量 R 都垂直于金属线，因此沿着通过 O 的半径方向，没有力矩。铅垂方向的力对铅垂轴没有力矩。所以角动量（angular momentum）$ma^2(\dot\phi_A+\dot\phi_B)$ 守恒。由于珠子从静止释放，角动量起初为零，因此 $\phi_A+\phi_B$ 保持不变：角 AOB 的平分线是固定的。取它为图中的虚线，使得 $\phi_A=\theta$ 及 $\phi_B=-\theta$。于是每颗珠子的速率都是 $a\dot\theta$。

弹簧就是弦 AB，长度为 $2a\sin\theta$。它的自然长度（natural length）是 $2a\sin\alpha$，所以伸长量（extension）是 $x=2a(\sin\theta-\sin\alpha)$，而弹性势能（elastic energy）是

$$\frac{\lambda x^2}{2\cdot 2a\sin\alpha}=\frac{4\lambda a^2(\sin\theta-\sin\alpha)^2}{4a\sin\alpha}=\frac{\lambda a(\sin\theta-\sin\alpha)^2}{\sin\alpha}.$$

这条公式在 $x<0$（压缩）时同样成立，因为 x 取了平方。动能（kinetic energy）是 $2\cdot\frac12 m(a\dot\theta)^2=ma^2\dot\theta^2$。金属线是光滑的，它的反作用力始终垂直于运动方向，所以不做功，能量守恒。释放时 $\theta=\beta$ 而 $\dot\theta=0$，所以

$$ma^2\dot\theta^2+\frac{\lambda a(\sin\theta-\sin\alpha)^2}{\sin\alpha}=\frac{\lambda a(\sin\beta-\sin\alpha)^2}{\sin\alpha}.$$

全式乘以 $\sin\alpha/a$，得

$$ma\dot\theta^2\sin\alpha+\lambda(\sin\theta-\sin\alpha)^2=\lambda(\sin\beta-\sin\alpha)^2,$$

这正是所求的结果。注意弹簧起初处于压缩状态，所以 $\beta<\alpha$。

第 (i) 部分

把能量方程对 t 求导：

$$2ma\sin\alpha\,\dot\theta\ddot\theta+2\lambda(\sin\theta-\sin\alpha)\cos\theta\,\dot\theta=0.$$

除以 $2\dot\theta$（只要珠子在运动，这一步就成立，而所得结果可凭连续性推广至所有 t），得运动方程

$$ma\sin\alpha\,\ddot\theta=-\lambda(\sin\theta-\sin\alpha)\cos\theta.$$

平衡位置（equilibrium）在 $\theta=\alpha$，此时弹簧处于自然长度。如果 $\beta-\alpha$ 很小，则 θ 会一直接近 α（由能量方程可得 $|\sin\theta-\sin\alpha|\le|\sin\beta-\sin\alpha|$），所以可写 $\theta=\alpha+\varepsilon$，其中

ε 很小。准确至一阶，$\sin\theta - \sin\alpha \approx \varepsilon\cos\alpha$，而 $\cos\theta \approx \cos\alpha$，后者所乘的量本身已经很小。所以

$$\ddot{\varepsilon} \approx -\frac{\lambda\cos^2\alpha}{ma\sin\alpha}\,\varepsilon,$$

这是简谐运动（simple harmonic motion），其中 $\omega^2 = \lambda\cos^2\alpha/(ma\sin\alpha)$。因此周期（period）是

$$T = \frac{2\pi}{\omega} \approx 2\pi\sqrt{\frac{ma\sin\alpha}{\lambda\cos^2\alpha}}.$$

第 (ii) 部分

设 $k = \sqrt{ma\sin\alpha/\lambda}$。由能量方程，

$$\dot{\theta}^2 = \frac{1}{k^2}\Big[(\sin\alpha - \sin\beta)^2 - (\sin\theta - \sin\alpha)^2\Big].$$

珠子在 $\sin\theta - \sin\alpha = \pm(\sin\alpha - \sin\beta)$ 时停下，即当 $\sin\theta = \sin\beta$ 或 $\sin\theta = 2\sin\alpha - \sin\beta$ 时。由 $\theta = \beta$ 出发，θ 会增加（受压缩的弹簧把两颗珠子推开），而只要 $\sin\beta < \sin\theta < 2\sin\alpha - \sin\beta$，就有 $\dot{\theta}^2 > 0$。

? **想一想** 是甚么决定我们处于两种情况中的哪一种?

关键在于 θ 能否在 $2\theta = \pi$ 之前到达第二个停止点；在该处 A 和 B 位于直径的两端，而 $\sin\theta$ 达到其最大值 1。这取决于 $2\sin\alpha - \sin\beta$ 是小于还是大于 1。

(a) $\sin\beta > 2\sin\alpha - 1$。这时 $2\sin\alpha - \sin\beta < 1$，所以存在锐角 γ 使得 $\sin\gamma = 2\sin\alpha - \sin\beta$，并且 $\gamma > \alpha$。珠子在 $\theta = \gamma$ 时停下（此时弹簧的伸长量与原来的压缩量相同），而 θ 在 β 与 γ 之间振荡。去程和回程所需的时间相等，所以

$$T = 2\sqrt{\frac{ma\sin\alpha}{\lambda}}\int_\beta^\gamma \frac{\mathrm{d}\theta}{\sqrt{(\sin\alpha - \sin\beta)^2 - (\sin\theta - \sin\alpha)^2}}, \qquad \gamma = \arcsin(2\sin\alpha - \sin\beta).$$

(b) $\sin\beta < 2\sin\alpha - 1$。这时 $2\sin\alpha - \sin\beta > 1 \geq \sin\theta$，所以当 θ 介乎 β 与 $\frac{1}{2}\pi$ 之间时，$\dot{\theta}$ 永不为零：珠子经过直径两端时仍在运动。此后弦再次缩短。能量方程只通过 $\sin\theta$ 与 θ 有关，而前者关于 $\frac{1}{2}\pi$ 对称，所以由 $\frac{1}{2}\pi$ 到 $\pi - \beta$ 的运动，是由 β 到 $\frac{1}{2}\pi$ 的运动的镜像。珠子在 $\theta = \pi - \beta$ 停下，此时弹簧再次被压缩至原来的长度，然后把珠子沿原路推回去。（在 $\theta = \pi - \beta$ 时，运动方程给出 $\ddot{\theta} < 0$，因为 $\sin\theta - \sin\alpha < 0$ 而 $\cos\theta < 0$。）一次完整的振荡是 $\beta \to \pi - \beta \to \beta$，相当于把由 β 到 $\frac{1}{2}\pi$ 的四分之一段重复四次：

$$T = 4\sqrt{\frac{ma\sin\alpha}{\lambda}}\int_\beta^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{(\sin\alpha - \sin\beta)^2 - (\sin\theta - \sin\alpha)^2}}.$$

由于 $\sin\alpha < 1$，我们有 $2\sin\alpha - 1 < \sin\alpha$，所以两种情况都与 $\beta < \alpha$ 相容。情况 (b) 只会在 $2\sin\alpha - 1 > 0$ 时出现，即当 $\frac{1}{6}\pi < \alpha < \frac{1}{4}\pi$ 时。

验算

- 在情况 (a) 中，代换 $\sin\theta = \sin\alpha - (\sin\alpha - \sin\beta)\cos u$ 把积分化为 $\int_0^{\pi} \mathrm{d}u / \cos\theta$。当 $\beta \to \alpha$ 时，它趋向 $\pi / \cos\alpha$，从而得回第 (i) 部分的周期。
- 我们以 $m = 1.3$、$a = 0.7$、$\lambda = 2.1$ 模拟了两颗珠子在金属线上各自独立运动的情况。能量方程在舍入误差范围内成立，平分线亦保持固定。取 $\alpha = 0.6$、$\beta = 0.59$ 时，周期为 3.766，而第 (i) 部分给出 3.7657。取 $\alpha = 0.7$、$\beta = 0.4$（情况 (a)）时，周期为 4.8847，而积分给出 4.8846。取 $\alpha = 0.75$、$\beta = 0.1$（情况 (b)）时，周期为 7.0553，而积分给出 7.0553，并且 θ 一如所料在 $\pi - \beta$ 达到最大值。

常见错误

- 用 $\lambda x^2/(2a)$ 作为弹性势能。分母应是*自然长度*的两倍，即 $2 \cdot 2a\sin\alpha$。
- 在情况 (b) 中，把积分停在 $\theta = \frac{1}{2}\pi$，或声称珠子在该处停下。珠子经过直径两端时速率不为零，所以周期是 $4\int_{\beta}^{\pi/2}$，而不是 $2\int_{\beta}^{\pi/2}$。

STEP 之外

第 (ii) 部分的周期积分无法以初等函数表示。它们与*椭圆积分*（elliptic integrals）属同一家族，后者给出单摆（simple pendulum）的精确周期。留意第 (i) 部分的小幅振荡周期与振幅无关，而精确周期则不然。线性化振子与非线性振子之间的这种对比，是经典力学入门课程的核心主题之一。内力作用下的系统的角动量守恒，在本题中令平分线保持固定；Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教材）对此有严谨的论述。

详细解答

STEP II 样本试卷，第 13 题

力学 • 平衡与力矩

STEP II 样本试卷，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/spec-s2-q13

关键思路

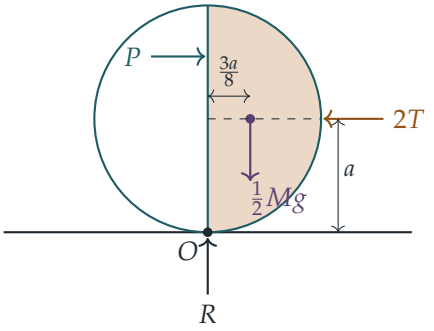
橙子若要散开，各瓣会向外翻倒，每一瓣都绕著球面与桌面接触的那一点转动。因此我们沿其中一个切面把橙子切成两半，然后对其中一半，绕通过该最低点的一条水平线取力矩（moment）。这一半的重量要把它推倒；能把它拉住的只有丝带，别无其他。另一半施加的力只会令情况更坏，“至少”二字正是由此而来。

解答

模型

我们把朱古力视为均匀的刚体（rigid body），把丝带视为轻质、整圈张力（tension）同为 T，并假设各瓣大小相等、数目为偶数（真正的朱古力橙有二十瓣），于是有一个通过轴线的切面把橙子分成两半，每一半都是半球形、夹角为 π 的楔形体。到最后我们会看到，只要切面光滑，这个论证对任何一组橙瓣都成立。

球面与桌面只在一点 O 接触，即轴线的底端。以 O 为原点建立坐标轴，z 铅垂向上，x 为水平方向，由轴线指向我们所研究的那一半（图中右边的一半）。分界平面是 $x = 0$。设 ℓ 为通过 O 且位于该平面内的水平线（即 y 轴）；我们绕 ℓ 取力矩，并规定若力矩倾向使这一半向外翻倒、远离轴线，则该力矩为正。



其中一半所受的力

右边的一半质量为 $\frac{1}{2}M$，它所受的外力如下。

- **重量** $\frac{1}{2}Mg$，作用于质心（centre of mass）。在题目给出的结果中代入 $2\theta = \pi$，可知质心与轴线的距离为
$$\frac{3\pi a \sin(\pi/2)}{16(\pi/2)} = \frac{3a}{8}$$
并位于 $x > 0$ 的一侧。它绕 ℓ 的力矩为 $\frac{1}{2}Mg \cdot \frac{3a}{8} = \frac{3}{16}Mga$，属翻倒方向。
- **桌面**：一个法向反作用力（normal reaction），另外可能有摩擦力（friction），全部作用于 O。由于 O 在 ℓ 上，这些力绕 ℓ 的力矩为零。
- **丝带**，绕著这一半的赤道。窄丝带沿法向压在表面上，每单位长度的力为 T/a（情况就如一条绳绕过光滑曲面）。在角度由 $-\frac{\pi}{2}$ 至 $\frac{\pi}{2}$ 的弧上，各 x 分量之和为
$$\int_{-\pi/2}^{\pi/2} \frac{T}{a} \cos\phi \, a \, \mathrm{d}\phi = 2T$$
方向指向轴线，而各 y 分量则因对称而互相抵消。（另一个等价的看法：在丝带穿过分界平面之处把它剪断，每个断口都以张力 T 沿 $-x$ 方向拉扯，同样得出 $2T$。）这个力完全是水平的，作用于高度 a，所以它绕 ℓ 的力矩为 $2T \cdot a$，属回复方向。
- **另一半**，作用于整个平面 $x = 0$ 上。这是唯一棘手的力，而我们无须知道它如何分布。

? **想一想** 为甚么另一半施加的力只会助长这一半翻倒?

把切面上每一小块所受的力分解为垂直于切面的部分和沿切面的部分（若切面粗糙，后者就是摩擦力）。

- 垂直的部分沿 x 方向。两半只是互相压著，并没有黏在一起，所以这部分是推力，沿 $+x$ 方向。它作用于某个高度 $z \geq 0$，所以它绕 ℓ 的力矩等于 z 乘以它的大小，永不为负：它倾向使这一半向外翻倒。

- 沿切面的部分位于平面 $x = 0$ 内，而这个平面也包含 ℓ。若一个力的作用线（line of action）与 ℓ 位于同一平面内，则作用线要么与 ℓ 相交，要么与它平行，所以这个力绕 ℓ 的力矩为零。

因此，无论接触的细节如何，另一半所贡献的总力矩 $P_{\text{mom}} \geq 0$，属翻倒方向。

绕 ℓ 取力矩

这一半处于平衡（equilibrium），所以绕 ℓ 的总力矩为零：

$$2Ta = \tfrac{3}{16}Mga + P_{\text{mom}}.$$

由于 $P_{\text{mom}} \geq 0$，

$$2Ta \;\geq\; \tfrac{3}{16}Mga, \qquad \text{所以} \qquad T \;\geq\; \tfrac{3}{32}Mg,$$

这正是所要证明的结果。等号成立，当且仅当两半之间完全没有互相施压，即两半仅仅能够维持在一起。

橙瓣数目不是偶数的情况

假如切面是光滑的，同样的计算对任何由完整橙瓣组成、夹角为 2θ 的楔形体都适用。取 x 沿它的对称平面，并取 ℓ 为通过 O 且垂直于该平面的水平线。楔形体的重量为 $\frac{\theta}{\pi}Mg$，作用于距离 $\frac{3\pi a \sin\theta}{16\theta}$ 处，产生的翻倒力矩为 $\frac{3}{16}Mga\sin\theta$。丝带在它的弧上以 $2T\sin\theta$ 向内拉，作用于高度 a。光滑的切面只会产生垂直于切面的水平推力，而这些推力在 $+x$ 方向的分量不会是负数，所以它们同样只会增加翻倒力矩。于是 $2Ta\sin\theta \geq \frac{3}{16}Mga\sin\theta$，约去 $\sin\theta$ 后再次得出 $T \geq \frac{3}{32}Mg$。无论我们选取哪一组橙瓣，所得的下界都相同。

验算

- 题目给出的质心结果在 $2\theta = \pi$ 时给出 $\frac{3a}{8}$，正是大家熟悉的实心半球体质心与其平面的距离。以蒙地卡罗（Monte Carlo）方法估算，这一半的结果为 $0.376a$，而对其他楔形夹角，估算值亦与公式吻合。
- 以数值方法把丝带在夹角为 2θ 的弧上的法向力加起来，得出 $2T\sin\theta$，例如对半个橙子就是 $2T$。
- 对每一个检验过的楔形夹角，所得的下界都同样是 $\frac{3}{32}Mg \;\approx\; 0.094Mg$（$2\theta = 0.6,\ \pi,\ 2.4,\ 5$）。下界只取决于整个橙子，而与我们选择怎样切开它无关，这令人放心。
- 单位：力矩是力乘以长度，而 a 会约去，余下的是一个力。

常见错误

- 绕球心或绕赤道上的一点取力矩。这样一来，桌面施加的未知力，或另一半施加的未知力，就会进入方程。选取 O 点，正是因为桌面的力绕通过该点的直线没有力矩。
- 说丝带对这一半的力是 T。丝带在这一半的弧的两端都在拉，所以合力（resultant）是 $2T$。
- 没有理据就把两半之间的力设为零。那样得出的只是等号成立的情况。“至少”二字，来自留意到这个力只能是推力。

STEP 之外

这是一道含有未知内力的静力学（statics）题目，处理方法是先选取一条轴，消去我们无法求出的未知量，然后利用一个不等式（接触力只会推，不会拉）来应付余下的部分。分析砖石拱门和穹顶时，背后也是同一套推理：结构能够屹立，当且仅当能找到一条始终留在材料内部的推力线。质心的结果本身是球极坐标（spherical polar coordinates）下一个标准的三重积分（triple integral），很值得亲自推导一次。想进一步了解刚体平衡，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP II 样本试卷，第 14 题

力学 • 刚体的转动

STEP II 样本试卷，第 14 题

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

step.maths.org/questions/spec-s2-q14

关键思路

圆盘和质点都绕同一条固定的铅垂轴转动，所以最自然的工具是转动的运动方程（rotational equation of motion）：力矩（torque）= 转动惯量（moment of inertia）× 角加速度（angular acceleration），并分别应用于两个物体。两者之间的摩擦力是一对内力：它令其中一个加快多少（以角动量（angular momentum）计），就令另一个减慢多少。由此立即得出守恒的组合；再把两条方程相减，便得到一条关于相对角速率的一阶方程，而这个相对角速率按指数衰减（exponential decay）。

解答

模型

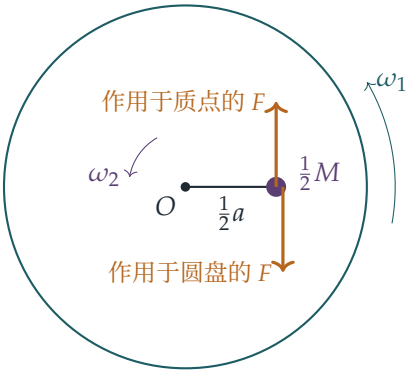
圆盘是一块均匀薄片（uniform lamina），绕通过其中心 O 的光滑固定轴转动，所以它对该轴的转动惯量为

$$I_1 = \tfrac{1}{2}Ma^2.$$

质点的质量为 $\frac{1}{2}M$，由一条轻而刚硬的金属线固定在距离轴 $\frac{1}{2}a$ 之处，所以它沿半径为 $\frac{1}{2}a$ 的圆运动，对该轴的转动惯量为

$$I_2 = \tfrac{1}{2}M\left(\tfrac{1}{2}a\right)^2 = \tfrac{1}{8}Ma^2.$$

起初 $\omega_1 = \Omega$ 而 $\omega_2 = 0$，所以在接触点处圆盘比质点运动得快。因此摩擦力沿质点的运动方向作用于质点（沿切线方向，位于半径 $\frac{1}{2}a$ 处），而根据牛顿第三定律（Newton's third law），在同一点上有一个大小相等、方向相反的力作用于圆盘。



（俯视图；$F = Mak(\omega_1 - \omega_2)$。金属线的张力沿径向，所以它对轴没有力矩。）

运动方程

摩擦力 $F = Mak(\omega_1 - \omega_2)$ 作用于距离轴 $\frac{1}{2}a$ 之处，并与半径垂直，所以它对轴的力矩为

$$F \cdot \tfrac{1}{2}a = \tfrac{1}{2}Ma^2k(\omega_1 - \omega_2).$$

质点的重量和各法向反作用力（normal reaction）都是铅垂的，所以它们对铅垂轴没有力矩，而轴本身是光滑的。对圆盘而言，摩擦力阻碍转动：

$$\tfrac{1}{2}Ma^2\frac{\mathrm{d}\omega_1}{\mathrm{d}t} = -\tfrac{1}{2}Ma^2k(\omega_1 - \omega_2) \quad\Longrightarrow\quad \frac{\mathrm{d}\omega_1}{\mathrm{d}t} = -k(\omega_1 - \omega_2).$$

对质点而言，摩擦力推动转动：

$$\tfrac{1}{8}Ma^2\frac{\mathrm{d}\omega_2}{\mathrm{d}t} = \tfrac{1}{2}Ma^2k(\omega_1 - \omega_2) \quad\Longrightarrow\quad \frac{\mathrm{d}\omega_2}{\mathrm{d}t} = 4k(\omega_1 - \omega_2).$$

（对质点，你同样可以沿切线方向应用牛顿第二定律：$\frac{1}{2}M \cdot \frac{1}{2}a\,\dot{\omega}_2 = F$，这是同一条方程。）

守恒的组合

把第一条方程乘以 4，再加上第二条，

$$4\frac{\mathrm{d}\omega_1}{\mathrm{d}t} + \frac{\mathrm{d}\omega_2}{\mathrm{d}t} = -4k(\omega_1 - \omega_2) + 4k(\omega_1 - \omega_2) = 0,$$

所以 $4\omega_1 + \omega_2$ 是常数。在 $t = 0$ 时它等于 $4\Omega + 0$，因此

$$4\omega_1 + \omega_2 = 4\Omega,$$

正如所求。这就是对该轴的角动量守恒（conservation of angular momentum）：$I_1\omega_1 + I_2\omega_2 = \frac{1}{8}Ma^2(4\omega_1 + \omega_2)$，而仅有的外力（轴上的反作用力、重力和铅垂的反作用力）对该轴都没有力矩。

求解 ω_1

想一想 我们有两条耦合的方程。哪一个量能够独自满足一条方程?

两个变化率都只取决于差 $u = \omega_1 - \omega_2$，所以不妨考虑这个量。把两条方程相减，

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\mathrm{d}\omega_1}{\mathrm{d}t} - \frac{\mathrm{d}\omega_2}{\mathrm{d}t} = -ku - 4ku = -5ku,$$

其中 $u(0) = \Omega - 0 = \Omega$。分离变量（separating variables）（或认出这是指数衰减），

$$\omega_1 - \omega_2 = u = \Omega\,\mathrm{e}^{-5kt}.$$

把这条式子与 $4\omega_1 + \omega_2 = 4\Omega$ 相加，便消去 ω_2：

$$5\omega_1 = 4\Omega + \Omega\,\mathrm{e}^{-5kt}, \qquad \text{所以} \qquad \omega_1 = \frac{\Omega}{5}\left(4 + \mathrm{e}^{-5kt}\right),$$

正如所求。由此可得 $\omega_2 = 4\Omega - 4\omega_1 = \frac{4}{5}\Omega\left(1 - \mathrm{e}^{-5kt}\right)$。

另一个做法是把 $\omega_2 = 4\Omega - 4\omega_1$ 代入圆盘的方程，得到 $\dot{\omega}_1 = -5k\omega_1 + 4k\Omega$，这是一条线性方程，它满足 $\omega_1(0) = \Omega$ 的解与上面相同。

验算

- 在 $t = 0$ 时：$\omega_1 = \frac{1}{5}\Omega(4+1) = \Omega$ 而 $\omega_2 = 0$。正确。
- 当 $t \to \infty$ 时：$\omega_1, \omega_2 \to \frac{4}{5}\Omega$，两者趋向同一速率，这正是摩擦力最终必然造成的结果。最终的动能（kinetic energy）为 $\frac{1}{2}\cdot\frac{5}{8}Ma^2\cdot\frac{16}{25}\Omega^2 = \frac{1}{5}Ma^2\Omega^2$，而最初为 $\frac{1}{4}Ma^2\Omega^2$，所以有五分之一的能量因摩擦而损失，这与“完全非弹性”（perfectly inelastic）耦合的预期相符。
- 数值上，取 $M = a = k = \Omega = 1$，把两条运动方程逐步计算至 $t = 1$，得 $\omega_1 = 0.80135$ 和 $\omega_2 = 0.79461$，与 $\frac{1}{5}(4 + \mathrm{e}^{-5})$ 和 $\frac{4}{5}(1 - \mathrm{e}^{-5})$ 吻合，而 $4\omega_1 + \omega_2$ 自始至终保持为 4。

常见错误

- 把质点的转动惯量当作 $\frac{1}{2}Ma^2$（仿佛它位于圆盘边缘），而不是 $\frac{1}{2}M(\frac{1}{2}a)^2$。
- 在转动方程中使用摩擦力本身，而不是它的力矩 $F\cdot\frac{1}{2}a$。
- 在两条方程中给摩擦力相同的正负号：它令圆盘减慢，却令质点加快。

STEP 之外

这里的摩擦力是内力，而 $4\omega_1 + \omega_2$ 的守恒正是碰撞中动量守恒的转动版本；以指数方式趋向同一速率，与两个物体以黏滞阻力（viscous drag）耦合的情况是同一套数学。更一

般地说，线性系统（linear system）$\dot{\mathbf{x}} = A\mathbf{x}$ 以 A 的特征向量（eigenvector）表示时便会解耦（decouple）：在这里，特征向量就是组合 $4\omega_1 + \omega_2$（特征值（eigenvalue）为 0）和 $\omega_1 - \omega_2$（特征值为 $-5k$）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）细致地讲述刚体转动和角动量，而 Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程）则以特征向量处理线性系统。

详细解答

STEP II 样本试卷，第 15 题

概率与统计 · 期望值与样本平均值

STEP II 样本试卷，第 15 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/spec-s2-q15

关键思路

国王每天向极大量的小贩罚款，所以根据大数定律（law of large numbers），他每天从每名小贩身上得到的收入，实际上就等于罚款的期望值（expected value）。因此我们要求出每个方案下的期望罚款。头两个方案涉及一组正态变量的样本平均值（sample mean），其分布仍然是正态分布（normal distribution），只是方差（variance）较小。第三个方案是等待时间的问题，所以要用上几何分布（geometric distribution）。

解答

设 $W_1, W_2, \ldots$ 为被选中的那箱腌鱼各自的重量（以克为单位）。箱中的腌鱼基本上是随机选出的，所以各 W_i 互相独立（independent），而且每个 $W_i \sim \mathrm{N}(200, 10^2)$。从箱中随机抽取腌鱼，与牠们的重量无关，所以无论我们抽出哪些腌鱼、次序如何，牠们仍然是独立的 $\mathrm{N}(200, 10^2)$ 变量。

我们会用到以下的标准结果：若 $W_1, \ldots, W_m$ 是独立的 $\mathrm{N}(\mu, \sigma^2)$ 变量，则它们的平均值满足

$$\overline{W} = \frac{1}{m}\sum_{i=1}^{m} W_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{m}\right).$$

第 (i) 部分

这里 $m = 625$，所以 $\overline{W} \sim \mathrm{N}\left(200, \frac{100}{625}\right)$，标准差（standard deviation）为 $\frac{10}{25} = 0.4$。于是

$$\mathrm{P}(\overline{W} < 199) = \Phi\left(\frac{199 - 200}{0.4}\right) = \Phi(-2.5) = 1 - \Phi(2.5) \approx 1 - 0.99379 = 0.00621.$$

期望罚款为

$$1500 \times 0.00621 \approx 9.3 \text{ 克朗}.$$

第 (ii) 部分

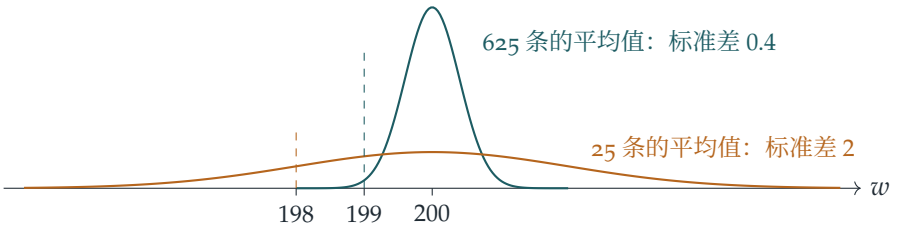
现在 $m = 25$，所以 $\overline{W} \sim \mathrm{N}\left(200, \frac{100}{25}\right)$，标准差为 2。于是

$$\mathrm{P}(\overline{W} < 198) = \Phi\left(\frac{198 - 200}{2}\right) = \Phi(-1) = 1 - \Phi(1) \approx 1 - 0.84134 = 0.15866,$$

而期望罚款为

$$100 \times 0.15866 \approx 15.9 \text{ 克朗}.$$

下图说明了为何金额较小的罚款反而胜出。对 625 条腌鱼的平均值来说，比平均值低一克的门槛距离平均值有 2.5 个标准差；但对 25 条腌鱼的平均值来说，比平均值低两克的门槛只相距一个标准差。



第 (iii) 部分

设 N 为取出的腌鱼数目。每条腌鱼重于 200 g 的概率是 $\frac{1}{2}$，因为 200 是一个对称分布的平均值，而且 $\mathrm{P}(W = 200) = 0$。抽出的腌鱼互相独立，所以 N 就是在成功概率为 $\frac{1}{2}$ 的独立试验中，直至首次成功为止（包括该次在内）的试验次数。换言之，N 服从几何分布：

$$\mathrm{P}(N = n) = \left(\tfrac{1}{2}\right)^n, \qquad n = 1, 2, 3, \ldots$$

（严格来说，抽完 625 条腌鱼后箱子就空了，但当中没有一条超过 200 g 的机会是 2^{-625}，完全可以忽略不计。）

想一想 怎样可以不先计算 $\mathrm{E}[N^2]$ 而求出 $\mathrm{E}[N(N-1)]$?

把几何级数（geometric series）求导两次。对于 $|x| < 1$，

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \Longrightarrow \quad \sum_{n=2}^{\infty} n(n-1)x^{n-2} = \frac{2}{(1-x)^3}.$$

乘以 x^2 并代入 $x=\frac{1}{2}$（$n=1$ 的一项反正是零）：

$$\mathrm{E}[N(N-1)]=\sum_{n=1}^{\infty} n(n-1)\left(\tfrac{1}{2}\right)^n=\left(\tfrac{1}{2}\right)^2\cdot\frac{2}{\left(\frac{1}{2}\right)^3}=4.$$

（等价地，取 $p=\frac{1}{2}$，则 $\mathrm{E}[N]=2$ 及 $\mathrm{Var}(N)=\frac{1-p}{p^2}=2$，所以 $\mathrm{E}[N^2]=6$，而 $\mathrm{E}[N^2]-\mathrm{E}[N]=4$。）因此期望罚款为

$$\mathrm{E}[3N(N-1)]=3\times 4=12 \text{ 克朗}.$$

结论

在方案 (i)、(ii) 和 (iii) 之下，每名小贩每天的期望罚款分别约为 9.3、15.9 和 12 克朗。由于受检查的箱子数量极多，国王的总收入非常接近检查次数乘以期望罚款。所以他应该选择**方案 (ii)**。它的收入比方案 (i) 多大约 70%，比方案 (iii) 多约三分之一。

验算

- 方案 (iii) 无法查表核对，但这个级数很容易逐项相加：$\frac{3\cdot 2}{4}+\frac{3\cdot 6}{8}+\frac{3\cdot 12}{16}+\cdots=1.5+2.25+2.25+1.875+\cdots$。部分和（partial sum）很快就攀升趋近 12。
- 模拟 10^5 箱各有 625 条正态腌鱼的箱子，并把三个方案都执行一遍，所得的平均罚款约为 10.2（来自 2×10^4 箱，所以误差较大）、15.9 和 11.9 克朗。这些数值与 9.31、15.87 和 12 相当接近。

常见错误

- 把样本平均值的标准差当作 $\frac{10}{625}$ 或 $\frac{100}{25}$。被 m 除的是*方差*，所以标准差是除以 $\sqrt{m}$。
- 把 N 当作由 0 开始（只数失败的次数）。这里 n 包括那条重的腌鱼，所以 $\mathrm{P}(N=1)=\frac{1}{2}$，而此时罚款为零。
- 比较被罚款的概率而不是比较期望罚款：方案 (iii) 有一半机会向小贩罚款（即 $n\geq 2$ 时），但通常只罚 6 克朗。

STEP 之外

“数量极多，所以用期望值”这个理据就是大数定律；而 $\overline{W}$ 服从方差为 σ^2/m 的正态分布这一事实，则与中心极限定理（central limit theorem）相伴而来，该定理令这个结论即使在个别重量并非正态分布时仍然近似成立。谨慎的国王或许还会考虑离散程度：方案 (i) 是罕有但巨额的罚款，其方差远比其他方案大。借着对级数求导来求出 $\mathrm{E}[N(N-1)]$ 这

类阶乘矩（factorial moment），这个想法会引伸出概率生成函数（probability generating function），它是大学一年级概率论的核心工具。Grimmett 与 Welsh 所著的 *Probability: An Introduction*（概率论入门）把这一切都讲解得十分清楚。

详细解答

STEP II 样本试卷，第 16 题

概率与统计 • 淘汰赛的概率

STEP II 样本试卷，第 16 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/spec-s2-q16

本题在 *Beyond STEP* 第 12.7 节淘汰赛中谁会遇上谁中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/knockout.html

关键思路

逐轮追踪两位选定的选手 A 和 B。在任何一轮中，只要两人仍未出局又未曾交手，就只有两件事要紧：A 是否抽中与 B 对赛？若否，两人是否都胜出？这两者都是简单的概率（probability），而由此得出的乘积会漂亮地逐项相消。至于最后一部分，还有一个巧妙的对称性（symmetry）论证：数一数整个淘汰赛中的比赛场数。

解答

记 $N = 2^n$，并称两位选定的选手为 A 和 B。经过 $r-1$ 轮后，余下 2^{n-r+1} 位选手，所以第 r 轮有 $M_r = 2^{n-r+1}$ 位选手，而第 n 轮（此时 $M_n = 2$）就是决赛。

我们会反复用到以下事实：若 A 和 B 同在 M 位随机配对的选手之中，则 A 的对手等可能地是其余 $M-1$ 位选手中的任何一位，所以

$$\mathrm{P}(A \text{ 抽中与 } B \text{ 对赛}) = \frac{1}{M-1}.$$

（决赛不设抽签，但代入 $M = 2$ 时公式给出 1，结果正确。）

第 (i) 部分

在第一轮中 $M_1 = N$，所以

$$\mathrm{P}(A \text{ 和 } B \text{ 在第一轮相遇}) = \frac{1}{2^n - 1}.$$

第 (ii) 部分

设 S_r 为“A 和 B 在第 r 轮开始时都仍在赛事之中”这个事件（event）（这必然表示两人尚未交手，因为每场比赛的负方即告出局）。在 S_r 发生的条件下，两人一同晋身第 $r+1$ 轮，当且仅当他们在第 r 轮没有抽中对赛，然后各自在（不同的）比赛中胜出。两场比赛的结果互相独立（independent），胜出的概率各为 $\frac{1}{2}$，所以

$$\mathrm{P}(S_{r+1} \mid S_r) = \frac{M_r - 2}{M_r - 1} \cdot \frac{1}{4}.$$

? **想一想** 怎样把它化成一个我们真正算得出来的乘积?

记 $M = 2^k$，并留意 $M - 2 = 2(2^{k-1} - 1)$，所以

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

某一轮的分子（撇除因子 2 不计）正是下一轮的分母。由于 S_1 是必然事件，把 $k = n, n-1, \ldots, n-r+2$ 各轮的结果相乘，便得到一个裂项相消的乘积（telescoping product）：

$$\mathrm{P}(S_r) = \frac{1}{4^{r-1}} \prod_{k=n-r+2}^{n} \frac{2\,(2^{k-1} - 1)}{2^k - 1} = \frac{2^{r-1}}{4^{r-1}} \cdot \frac{2^{n-r+1} - 1}{2^n - 1} = \frac{2^{n-r+1} - 1}{2^{r-1}\,(2^n - 1)}.$$

在 S_r 发生的条件下，两人在第 r 轮相遇的概率为 $1/(M_r - 1) = 1/(2^{n-r+1} - 1)$，所以

$$\mathrm{P}(A \text{ 和 } B \text{ 在第 } r \text{ 轮相遇}) = \frac{1}{2^{r-1}\,(2^n - 1)}.$$

这条公式简洁得令人欣喜：每过一轮，在该轮相遇的机会就减半。取 $r = 1$ 便重新得到第 (i) 部分的结果。至于决赛，取 $r = n$：

$$\mathrm{P}(A \text{ 和 } B \text{ 在决赛相遇}) = \frac{1}{2^{n-1}\,(2^n - 1)}.$$

再用对称性核对一下是否合理：赛制对每位选手一视同仁，所以决赛的两位选手等可能地是 $\binom{N}{2} = 2^{n-1}(2^n - 1)$ 对选手中的任何一对，由此得出相同的答案。

第 (iii) 部分

对 $r = 1, \ldots, n$，“在第 r 轮相遇”这些事件互斥（mutually exclusive）（两位选手最多只能交手一次，因为负方会离场），所以我们可利用等比级数（geometric series）把概率相

加：

$$\mathrm{P}(A \text{ 和 } B \text{ 相遇}) = \frac{1}{2^n-1}\sum_{r=1}^{n}\frac{1}{2^{r-1}} = \frac{1}{2^n-1}\cdot 2\left(1-\frac{1}{2^n}\right)$$
$$= \frac{2}{2^n}\cdot\frac{2^n-1}{2^n-1} = \frac{1}{2^{n-1}}.$$

? **想一想** 有没有方法可以完全不用求和就看出这个答案?

有。每场比赛恰好淘汰一位选手，而除冠军外所有人都会被淘汰，所以比赛总是恰好有 2^n-1 场。每场比赛由一对选手进行，而根据对称性，$\binom{N}{2}$ 对选手中每一对交手的概率都相同，设为 p。比赛场数的期望值（expected value）等于所有选手对的这些概率之和，所以

$$\binom{2^n}{2}p = 2^n-1 \quad\Longrightarrow\quad p = \frac{2(2^n-1)}{2^n(2^n-1)} = \frac{2}{2^n} = \frac{1}{2^{n-1}},$$

与求和所得的结果一致。

验算

- $n=1$：只有两位选手，他们必定在第一轮相遇，而这一轮就是决赛。三条公式都给出 1。
- $n=2$：四位选手。第 (i) 部分给出 $\frac{1}{3}$，决赛给出 $\frac{1}{6}$，而总和是 $\frac{1}{3}+\frac{1}{6}=\frac{1}{2}$，与第 (iii) 部分所说的相符。
- $n=3$：把每一种抽签结果和每一种赛果逐一列举，准确算出 $\frac{1}{7}$、$\frac{1}{28}$ 和 $\frac{1}{4}$。以 16 位选手模拟 2×10^5 次淘汰赛，得出 0.0664、0.0085 和 0.1247，而准确值则为 $\frac{1}{15}\approx0.0667$、$\frac{1}{120}\approx0.0083$ 和 $\frac{1}{8}$。

常见错误

- 忘记了两位选手若要在较后的一轮相遇，必须先在之前的每一轮都避开对方：因子 $\frac{M-2}{M-1}$ 很容易漏掉。
- 把第 (ii) 部分答作 $\left(\frac{1}{2^{n-1}}\right)^2$，即假如两人互相独立时双双打入决赛的机会。但他们并不独立：若两人早早抽中对赛，最多只有一人能够晋级。

STEP 之外

第 (iii) 部分的论证，其实是配合指示变量（indicator variable）使用的期望值的线性性质（linearity of expectation）：不论各事件是否独立，比赛场数的期望值都等于所有选手

对各自交手的概率之和。这是概率论和组合数学（combinatorics）中最强大的工具之一，也是用来证明具有指定性质的对象确实存在的“概率方法”（probabilistic method）的基础。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率入门教科书）有一整章精彩的例子，示范以指示变量立论的方法。

详细解答

STEP III 样本试卷，第 1 题

纯数学 · 数学归纳法

STEP III 样本试卷，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/spec-s3-q1

关键思路

第 (i) 部分：先算出最初几个乘积，再用合适的角度去看：只要把每个乘积都写成分母为 $2n$ 的形式，规律便一目了然。因式分解（factorisation）$1-\frac{1}{m^2}=\frac{(m-1)(m+1)}{m^2}$ 解释了个中原因，也令归纳步骤变成一行计算。第 (ii) 部分：固定 n，对 k 作归纳（induct）：在和式中多加一项，其实正是换了个样子的帕斯卡法则（Pascal's rule）。

解答

第 (i) 部分

记 $P_n=\left(1-\frac{1}{4}\right)\left(1-\frac{1}{9}\right)\cdots\left(1-\frac{1}{n^2}\right)$。最初几个值是

$$P_2=\frac{3}{4},\qquad P_3=\frac{3}{4}\cdot\frac{8}{9}=\frac{2}{3},\qquad P_4=\frac{2}{3}\cdot\frac{15}{16}=\frac{5}{8},\qquad P_5=\frac{5}{8}\cdot\frac{24}{25}=\frac{3}{5}.$$

分母 4、8 提示我们把所有分数都写成以 $2n$ 为分母：$P_2=\frac{3}{4}$，$P_3=\frac{4}{6}$，$P_4=\frac{5}{8}$，$P_5=\frac{6}{10}$。于是我们猜想

$$P_n=\frac{n+1}{2n}\qquad \text{对于} n\geq 2.$$

用数学归纳法（induction）证明。设 $S(n)$ 为命题 $P_n=\dfrac{n+1}{2n}$。

基础情况（*base case*）。$P_2=1-\frac{1}{4}=\frac{3}{4}=\frac{2+1}{2\cdot 2}$，所以 $S(2)$ 成立。

归纳步骤（*inductive step*）。假设 $S(n)$ 对某个 $n\geq 2$ 成立。由于 $P_{n+1}=P_n\left(1-\frac{1}{(n+1)^2}\right)$，

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

$$\begin{aligned}P_{n+1}&=\frac{n+1}{2n}\cdot\frac{n(n+2)}{(n+1)^2}\\&=\frac{n+2}{2(n+1)},\end{aligned}$$

这正是 $S(n+1)$。根据数学归纳法，$S(n)$ 对所有 $n\geq 2$ 成立。

（这个因式分解也解释了上面的猜想：$P_n=\prod_{m=2}^{n}\frac{m-1}{m}\cdot\prod_{m=2}^{n}\frac{m+1}{m}=\frac{1}{n}\cdot\frac{n+1}{2}$，两个乘积都可以逐项相消（telescoping）。）

第 (ii) 部分

固定一个正整数 n，并设 $T(k)$ 为命题

$$\sum_{r=0}^{k}(-1)^r\binom{n}{r}=(-1)^k\binom{n-1}{k}.$$

我们对 k 作归纳，证明 $T(k)$ 对 $k=0,1,\ldots,n-1$ 成立。

基础情况。当 $k=0$ 时，两边都等于 $\binom{n}{0}=\binom{n-1}{0}=1$。

归纳步骤。假设 $T(k)$ 对某个满足 $0\leq k\leq n-2$ 的 k 成立（这样 $k+1$ 仍在范围之内）。加上下一项，

$$\begin{aligned}\sum_{r=0}^{k+1}(-1)^r\binom{n}{r}&=(-1)^k\binom{n-1}{k}+(-1)^{k+1}\binom{n}{k+1}\\&=(-1)^{k+1}\left[\binom{n}{k+1}-\binom{n-1}{k}\right].\end{aligned}$$

> ? **想一想** $\binom{n}{k+1}-\binom{n-1}{k}$ 等于甚么？

帕斯卡法则指出 $\binom{n}{k+1}=\binom{n-1}{k}+\binom{n-1}{k+1}$：从 n 件物件中选出 $k+1$ 件，某一件指定的物件要么被选中（其余 k 件来自另外 $n-1$ 件），要么没有被选中（全部 $k+1$ 件都来自另外 $n-1$ 件）。所以方括号内的式子等于 $\binom{n-1}{k+1}$，于是

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

这正是 $T(k+1)$。根据数学归纳法，$T(k)$ 对 $k=0,1,\ldots,n-1$ 成立，证明完毕。

若 $n=1$，唯一的情况是 $k=0$，基础情况已经涵盖，所以这个论证在该处无须特别处理。

验算

- $P_{300} = \frac{301}{600} \approx 0.5017$；用电脑以分数把各因子精确相乘，结果对 2 至 300 的每一个 n 都与 $\frac{n+1}{2n}$ 吻合。当 $n \to \infty$ 时，乘积趋向 $\frac{1}{2}$。
- 当 $n = 4$ 时：部分和（partial sums）$1,\ 1-4=-3,\ -3+6=3,\ 3-4=-1$ 与 $\binom{3}{0}$，$-\binom{3}{1}$，$\binom{3}{2}$，$-\binom{3}{3}$ 吻合。电脑已对所有 $n \le 79$ 及所有 $k \le n-1$ 精确验证 (ii)。
- 取 $k = n-1$，再加上最后一项 $(-1)^n\binom{n}{n}$，得 $(-1)^{n-1} + (-1)^n = 0$，这就是大家熟悉的结果 $\sum_{r=0}^{n}(-1)^r\binom{n}{r} = (1-1)^n = 0$。

常见错误

- 根据约至最简的分数（$\frac{3}{4}, \frac{2}{3}, \frac{5}{8}, \frac{3}{5}$）来猜想，因而看不出规律：窍门在于把约分还原，写成统一的形式。
- 在 (ii) 中对 n 而不是对 k 作归纳（可行，但较为繁复），或者忘记了由 k 到 $k+1$ 的一步只需在 $k+1 \le n-1$ 时进行。

STEP 之外

第 (i) 部分是无穷乘积（infinite product）$\prod_{m\ge 2}\left(1-\frac{1}{m^2}\right) = \frac{1}{2}$ 的开端。它是欧拉（Euler）乘积 $\frac{\sin \pi x}{\pi x} = \prod_{m\ge 1}\left(1-\frac{x^2}{m^2}\right)$ 的特殊情况：把两边除以 $m=1$ 的因子 $1-x^2$，再令 $x \to 1$ 即可。怎样理解无穷乘积，以及它们何时收敛（converge），是大学一年级分析学（analysis）的课题。第 (ii) 部分是众多二项式恒等式（binomial identities）之一，这些恒等式可由帕斯卡法则或生成函数（generating functions）推出：它就是 $(1-x)^n \cdot \frac{1}{1-x} = (1-x)^{n-1}$ 中 x^k 的系数。想了解无穷和与无穷乘积的收敛，Abbott 的 *Understanding Analysis*（分析学入门书）是很好的读物。

详细解答

STEP III 样本试卷，第 2 题

纯数学 · 二项式不等式与极限

STEP III 样本试卷，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/spec-s3-q2

关键思路

设 $y = 1 + x$，其中 $x > 0$，则 y 的幂可写成二项式展开（binomial expansion），而展开式中每一项都是非负的。因此，我们可以舍去其余各项，只保留一项，所得的仍然是一个下界（lower bound）。保留含 x 的一项，得到的下界增长速度与 n 相若，足以超越任何固定的 K；若要在除以 n 之后仍能超越 K，便需要一个增长速度与 n^2 相若的下界，所以我们改为保留含 x^2 的一项。

解答

不等式

设 $x \geq 0$，并设 n 为正整数。根据二项式定理（binomial theorem），

$$(1+x)^n = 1 + nx + \binom{n}{2}x^2 + \cdots + x^n.$$

由于 $x \geq 0$，首两项之后的每一项都是非负的，所以

$$(1+x)^n \geq 1 + nx > nx.$$

最后一个不等式（inequality）是严格的，原因在于那个 1；正因如此，即使 $x = 0$，结果依然成立（此时不等式变成 $1 > 0$）。

推导 y^n 最终会超过 K

设 $y > 1$，并取 $x = y - 1$，于是 $x > 0$。由第一部分可知，对每个正整数 n，

$$y^n = (1+x)^n > n(y-1).$$

若 $K \leq 0$，则毋须再做甚么：对每个 n 都有 $y^n > 0 \geq K$，所以取 $N = 1$ 即可。若 $K > 0$，则选取一个整数 N，使得

$$N > \frac{K}{y-1},$$

这是办得到的，因为 $y - 1 > 0$，所以右边是一个固定的数。于是对每个整数 $n \geq N$，

$$y^n > n(y-1) \geq N(y-1) > K,$$

其中中间一步用到了 $y - 1 > 0$。这正是我们所需要的结果。注意同一个 N 对所有 $n \geq N$ 都适用，而不仅是对 $n = N$ 适用，因为下界 $n(y-1)$ 随 n 递增。

证明 y^n/n 最终会超过 K

由下界 $y^n > n(y-1)$ 只能得出 $y^n/n > y - 1$，而这是一个固定的数，对付不了很大的 K。我们需要为 y^n 找一个增长得比 n 更快的下界。

> ? **想一想** 这一次，我们应该保留二项式展开中的哪一项?

保留含 x^2 的一项。由于 $x = y - 1 > 0$，展开式中所有项都是正的，所以对每个正整数 n，

$$y^n = (1+x)^n \geq 1 + nx + \frac{n(n-1)}{2}x^2 > \frac{n(n-1)}{2}x^2.$$

（当 $n = 1$ 时，展开式中并没有含 x^2 的项，但最后一个不等式仍然成立，因为其右边等于 0。）两边除以 $n > 0$，得

$$\frac{y^n}{n} > \frac{(n-1)(y-1)^2}{2}.$$

现在右边会随 n 增大而无止境地增长，这正是我们所需要的。若 $K \leq 0$，则取 $N = 1$ 即可，因为 $y^n/n > 0$。若 $K > 0$，则选取一个整数 N，使得

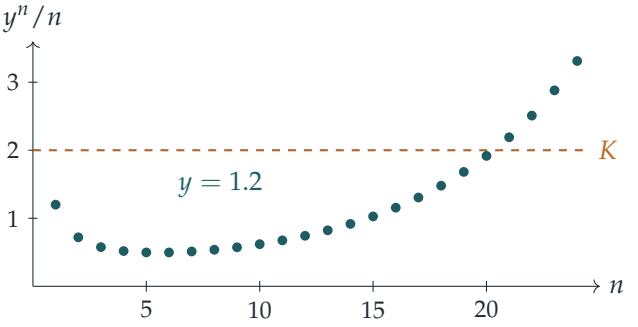
$$N > 1 + \frac{2K}{(y-1)^2}.$$

于是对每个整数 $n \geq N$，我们有 $n - 1 \geq N - 1 > 2K/(y-1)^2$，因此

$$\frac{y^n}{n} > \frac{(n-1)(y-1)^2}{2} > \frac{2K}{(y-1)^2} \cdot \frac{(y-1)^2}{2} = K,$$

这就是所要求的结果。

这个下界同样随 n 递增，所以论证对每个 $n \geq N$ 都成立。这一点在这里十分重要：y^n/n 本身并非总是递增的。当 y 接近 1 时，它起初会下降（例如，只要 $y < 2$，便有 $y^2/2 < y$），正如 $y = 1.2$ 的略图所示。因此，我们不能只找出一个满足 $y^n/n > K$ 的 n，然后声称其后各项都会更大。



验算

试取 $y = 1.01$ 和 $K = 1000$。按照我们的方法，对于 $y^n > K$ 得出 $N = 100\,001$，对于 $y^n/n > K$ 得出 $N = 20\,000\,002$。直接计算可知，其实对每个 $n \geq 1424$，$1.01^n/n > 1000$ 已经成立。我们的 N 远非最小的可能值，但题目只要求找出某一个 N。我们还以有理数（rational）x 作精确计算，对数千组随机的 x 和 n 检验了两个不等式，并对数千组随机的 $y > 1$ 和 K 检验了所选的 N（以及许多更大的 n）。

常见错误

- 只证明了对某一个 N 有 $y^N > K$，却没有解释为何对每个 $n \geq N$ 都有 $y^n > K$。
- 在最后一部分使用 $y^n > n(y-1)$：除以 n 之后只剩下一个常数，无法超越任意的 K。
- 忘记了 K 可以是负数或零；在这种情况下，$K/(y-1)$ 本身并不是一个合理的下界（不过此时任何 N 都适用）。

STEP 之外

这是大多数学生遇到的第一个严格的“ε–N”式论证：它证明了当 $y > 1$ 时，$y^n \to \infty$ 且 $y^n/n \to \infty$。运用同样的技巧，保留含 x^{k+1} 的一项，便可证明对任何固定的 k 都有 $y^n/n^k \to \infty$，所以指数函数（exponential）的增长胜过任何多项式（polynomial）。Abbott 的 *Understanding Analysis* 正是由这类推理出发，建立起数列（sequence）与极限（limit）的理论。

详细解答

STEP III 样本试卷，第 3 题

纯数学 · 复数的三角不等式

STEP III 样本试卷，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/spec-s3-q3

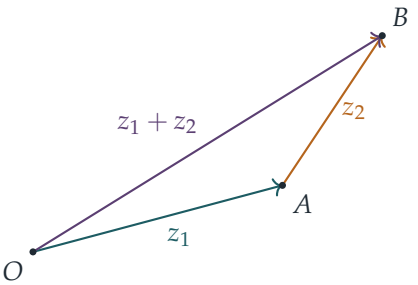
关键思路

题中的不等式就是三角不等式（triangle inequality）：三角形任何一边的长度都不会超过另外两边之和。反复应用这个不等式，便可以用各项模（modulus）之和来界定整个和的模。如果当 $|z| \leq \frac{1}{3}$ 时，方程左方的模总是严格小于 1，那么它就永远不可能等于 1；而这个上界只需用一个等比级数（geometric series）便可求得。

解答

几何意义

在阿根图（Argand diagram）中，把 z_1 看成由原点 O 到点 A 的位移，把 z_2 看成由 A 到点 B 的位移。那么 $z_1 + z_2$ 就是由 O 直接到 B 的位移。



三角形 OAB 三边的长度分别是 $|z_1|$、$|z_2|$ 和 $|z_1 + z_2|$。这个不等式说的是：由 O 直接走到 B，路程绝不会比途经 A 更长，也就是三角形的一边至多等于另外两边之和。等号成立，当且仅当三角形退化成一条线段，而 A 位于线段 OB 上，即 z_1 与 z_2 指向同一方向

（其中一个是另一个的非负实数倍，或者其中一个是零）。

利用归纳法（induction），同样的结论对任意有限多项都成立：

$$|w_1 + w_2 + \cdots + w_n| \le |w_1| + |w_2| + \cdots + |w_n|,$$

因为 $|w_1 + \cdots + w_{k+1}| \le |w_1 + \cdots + w_k| + |w_{k+1}|$。从几何上看，由 n 段直线组成的路径，其长度至少等于由起点到终点的直线距离。

在 $|z| \le \frac{1}{3}$ 内没有解

假设 $|z| \le \frac{1}{3}$，并且对每个 i 都有 $|a_i| \le 2$。先用推广后的三角不等式，再用 $|ab| = |a||b|$，得

$$\begin{aligned}\left|a_1 z + a_2 z^2 + \cdots + a_n z^n\right| &\le |a_1||z| + |a_2||z|^2 + \cdots + |a_n||z|^n \\ &\le 2\left(\tfrac{1}{3} + \tfrac{1}{3^2} + \cdots + \tfrac{1}{3^n}\right),\end{aligned}$$

其中第二步用到 $|a_i| \le 2$，以及对每个 k 都有 $|z|^k \le \left(\frac{1}{3}\right)^k$（所涉及的量全部非负，所以可以逐项估计上界）。

括号内是一个有限等比级数，首项为 $\frac{1}{3}$，公比为 $\frac{1}{3}$：

$$2\sum_{k=1}^{n} \frac{1}{3^k} = 2 \cdot \frac{\frac{1}{3}\left(1 - 3^{-n}\right)}{1 - \frac{1}{3}} = 1 - \frac{1}{3^n} < 1.$$

> ? **想一想** 上面每一步都只用了 $\le$，为甚么最后的不等式却是严格的?

严格性来自级数，而不是来自三角不等式。由于求和到第 n 项便停止，它与无穷级数的总和 1 相比，恰好少了 $3^{-n} > 0$。所以无论 a_i 取甚么值，也无论 z 是哪一个满足 $|z| \le \frac{1}{3}$ 的数，都有

$$\left|a_1 z + a_2 z^2 + \cdots + a_n z^n\right| \le 1 - 3^{-n} < 1.$$

但假如 z 是一个解，左方就会等于 $|1| = 1$。这是矛盾的，所以方程在 $|z| \le \frac{1}{3}$ 内没有解，这正是所要证明的。

验算

- 最坏的情况是 $a_1 = \cdots = a_n = 2$ 且 $z = \frac{1}{3}$，这时每个不等式都变成等式，左方等于 $1 - 3^{-n}$。当 $n = 10$ 时，这个值是 $0.99998\ldots$，仅仅略小于 1。
- 常数 $\frac{1}{3}$ 不能再改进。当每个 $a_i = 2$ 时，实函数 $2(x + x^2 + \cdots + x^n) - 1$ 在 $x = \frac{1}{3}$ 处取负值，而当 n 很大时，它在稍稍超过 $\frac{1}{3}$ 的地方取正值，所以存在与 $\frac{1}{3}$ 要多接近有多接近的实根。

- 以数值方法测试了 200 000 组随机选取的 $n \le 30$、满足 $|a_i| \le 2$ 的 a_i 以及满足 $|z| \le \frac{1}{3}$ 的 z，左方的模从未达到 1 或以上。

常见错误

- 用无穷级数来估计上界，只得到 ≤ 1，这并不能排除等号成立的可能。你需要的是严格不等式，而它正是来自级数只有有限项这一点。
- 写下 $|a_1 z + \cdots + a_n z^n| \le |a_1 z| + \cdots$ 却没有说明理由：这是推广到 n 项的三角不等式，题目设置第一部分正是为此。

STEP 之外

这是*无根区域*（root-free region）最简单的例子：只要多项式的系数有界，它的根就会被排除在某个圆盘之外。把同样的估计反过来用，便得到柯西（Cauchy）界：$z^n + c_{n-1}z^{n-1} + \cdots + c_0$ 的每个根的模都小于 $1 + \max |c_k|$；而在复分析（complex analysis）中，儒歇定理（Rouché's theorem）更把这类论证化为圆内根的个数的精确计算。想看一个极富几何味道的讲述，可以读 Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）。

详细解答

STEP III 样本试卷，第 4 题

纯数学 • 抛物线的切线

STEP III 样本试卷，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/spec-s3-q4

关键思路

一条直线与抛物线（parabola）相切，当且仅当把直线方程代入抛物线方程后，所得的二次方程有重根（repeated root）。因此第一部分只是一行的判别式（discriminant）计算，并且给出一族以 t 标记的切线（tangent）。至于第二部分，可把从某定点所作的切线，看成满足某个二次方程的 t 值：这样一来，有关 L 和 M 的一切都可以用两根之和与两根之积表示，而这两个量可直接从系数读出。

解答

切线

由直线方程得 $x = ty - at^2$。代入 $y^2 = 4ax$ 得

$$y^2 = 4aty - 4a^2t^2 \quad \Longleftrightarrow \quad y^2 - 4aty + 4a^2t^2 = 0 \quad \Longleftrightarrow \quad (y - 2at)^2 = 0.$$

这个关于 y 的二次方程有重根 $y = 2at$，所以直线与抛物线恰好交于一点，而且重数为二：直线与抛物线相切。（它并不是与抛物线的轴平行的直线——那种直线也只与抛物线相交一次——因为它不是水平线；当 $t = 0$ 时，它是直线 $x = 0$，即顶点处的切线。）对应的 x 为 $x = t(2at) - at^2 = at^2$，所以切点是

$$(at^2,\ 2at).$$

作为覆核，对 $y^2 = 4ax$ 作隐函数求导（implicit differentiation），得在该点 $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{2a}{y} = \frac{1}{t}$，这正是 $ty = x + at^2$ 的斜率。

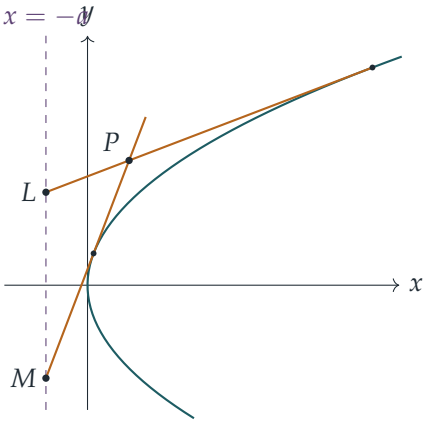
轨迹

抛物线上每一点都可写成 $(at^2, 2at)$，而且对应的 t 是唯一的，所以每条切线都是直线 $ty = x + at^2$ 的其中一条。设给定的点为 $P = (h, k)$。参数为 t 的切线通过 P，当且仅当

$$at^2 - kt + h = 0.$$

要有两条切线，就需要两个相异实根 t_1, t_2，即 $k^2 > 4ah$（P 位于抛物线之外）。这时

$$t_1 + t_2 = \frac{k}{a}, \qquad t_1 t_2 = \frac{h}{a}.$$



参数为 t 的切线与准线（directrix）$x = -a$ 相交于满足 $ty = -a + at^2$ 之处。只要 $t \neq 0$，便有

$$y = a\left(t - \frac{1}{t}\right).$$

（若 $t = 0$，切线是 $x = 0$，与准线平行，没有交点。这种情况只会在 $h = t_1 t_2\, a = 0$ 时出现，所以要 L 和 M 存在，必须有 $h \neq 0$，这时 $t_1, t_2 \neq 0$。）因此 L 和 M 分别是 $\left(-a,\, a(t_1 - \frac{1}{t_1})\right)$ 和 $\left(-a,\, a(t_2 - \frac{1}{t_2})\right)$，而由于两点位于同一条铅垂线上，

$$\begin{aligned} LM &= \left|a(t_1 - t_2) - a\left(\frac{1}{t_1} - \frac{1}{t_2}\right)\right| = \left|a(t_1 - t_2) + a\,\frac{t_1 - t_2}{t_1 t_2}\right| \\ &= |a|\,|t_1 - t_2|\left|\frac{t_1 t_2 + 1}{t_1 t_2}\right|. \end{aligned}$$

想一想 怎样不解二次方程而求得 $t_1 - t_2$？

它的平方是两根的对称式：

$$(t_1 - t_2)^2 = (t_1 + t_2)^2 - 4t_1 t_2 = \frac{k^2}{a^2} - \frac{4h}{a} = \frac{k^2 - 4ah}{a^2}.$$

另外

$$\frac{t_1t_2+1}{t_1t_2}=\frac{h/a+1}{h/a}=\frac{h+a}{h}.$$

把 LM 的表达式平方再代入，

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

若 $LM=\ell$，两边同乘 h^2 得

$$(h+a)^2(k^2-4ah)=\ell^2h^2,$$

所以 $P=(h,k)$ 位于 $(x+a)^2(y^2-4ax)=\ell^2x^2$ 上，这正是所要证明的。

验算

- 取 $a=1$ 及 $P=(1,3)$，如图所示。参数满足 $t^2-3t+1=0$，所以 $t=\frac{1}{2}(3\pm\sqrt{5})$，而 $t-\frac{1}{t}=\pm\sqrt{5}$（因为 $t_1t_2=1$，故 $1/t_1=t_2$）。因此 $LM=2\sqrt{5}$，$LM^2=20$，而公式给出 $(1+1)^2(9-4)/1^2=20$。
- 准线上的点，即 $h=-a$：公式给出 $LM=0$。事实上，从这样的点所作的切线，与准线的交点就是该点本身，所以 $L=M$。
- 我们用另一方法独立地求出切线（直线 $y=k+m(x-h)$，其中 $hm^2-km+a=0$），对超过 14,000 组随机的 a、h、k 作测试，计算所得的 LM^2 每次都与 $(h+a)^2(k^2-4ah)/h^2$ 吻合。例如 $a=1$、$P=(2,5)$ 时，两种方法都得出 $LM^2=38.25$。

常见错误

- 把切线的参数方程用反了：P 固定而 t 未知，得出的是二次方程 $at^2-kt+h=0$，而不是关于 h 和 k 的方程。
- 写成 $\frac{1}{t_1}-\frac{1}{t_2}=\frac{t_1-t_2}{t_1t_2}$；正确的是 $\frac{t_2-t_1}{t_1t_2}$，正是这个正负号令两项相加而非互相抵销。

STEP 之外

从一点向圆锥曲线（conic）所作的切线，以及连接两个切点的直线（即该点的极线（polar）），是射影几何（projective geometry）的起点。在射影几何中，极点（pole）、极线与准线都可纳入同一幅图像：准线就是焦点（focus）的极线。像本题这样通过根的对称函数（symmetric functions）来处理各根，也是迈向方程论（theory of equations）和伽罗瓦理论（Galois theory）的第一步。Siklos 的 *Advanced Problems in Mathematics* 收录了更多这类风格的坐标几何题目。

详细解答

STEP III 样本试卷，第 5 题

纯数学 · 方程根的计数

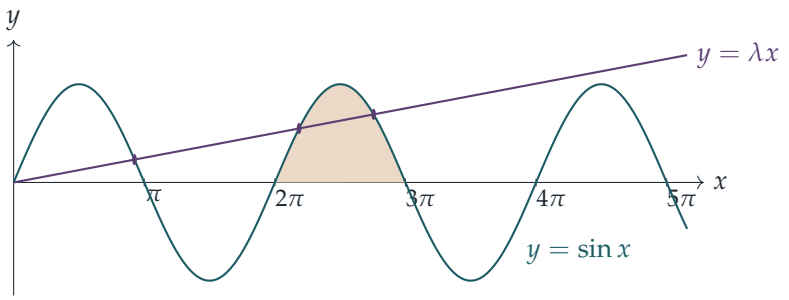
STEP III 样本试卷，第 5 题

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/spec-s3-q5

关键思路

把直线 $y = \lambda x$ 与曲线 $y = \sin x$ 画在一起：方程的解就是两者的交点。直线只能在 $\sin x > 0$ 的地方与曲线相交，即在正的“拱” $\big((2n-2)\pi, (2n-1)\pi\big)$ 上；又因为 $\sin x$ 在每个拱上都是凹（concave）的，直线与一个拱要么相交两次，要么完全不相交（第一个拱除外，因为它从原点开始）。至于 ϵ 很小时的估算，用麦克劳林级数（Maclaurin series）$\sin x \approx x - \frac{1}{6}x^3$ 就可以了。



（图中 $\lambda = 0.08$：第一个拱上有一个交点，涂上阴影的第二个拱上有两个交点，而直线在第三个拱到达顶点之前已升至 1 以上，所以 $N = 3$。）

解答

第 (i) 部分

$\lambda > 1$ **时无解**。曲线 $y = \sin x$ 的斜率（gradient）为 $\cos x \leq 1$，并且通过原点，而直线 $y = x$ 是它在原点的切线（tangent）。当 $x > 0$ 时，我们有 $\sin x < x$（函数 $x - \sin x$ 在

0 处的值为零，其导数（derivative）$1-\cos x \geq 0$ 只在孤立的点上等于零，所以它严格递增（strictly increasing））。从图像来看，当 $x>0$ 时，曲线严格位于其切线 $y=x$ 的下方，而若 $\lambda>1$，直线 $y=\lambda x$ 在该处位于 $y=x$ 的上方。因此

$$\sin x < x < \lambda x \qquad \text{对所有} x>0,$$

故此没有非零解。

$\lambda = 1-\epsilon^2$ 的情况。当 $x>0$ 时，方程等价于

$$\frac{\sin x}{x} = 1-\epsilon^2.$$

首先，解存在，而且只有一个。在 $(0,\pi]$ 上，函数 $\frac{\sin x}{x}$ 是连续（continuous）的，当 $x\to 0^+$ 时趋向 1，在 $x=\pi$ 处等于 0，所以它在 $(0,\pi)$ 内某处取值 $1-\epsilon^2$；它在该区间上严格递减（由原点连到凹拱上一点的直线，会随着该点沿拱移动而变得愈来愈平缓），所以这种情况只出现一次。当 $x\geq\pi$ 时再没有其他解：因为 ϵ 很小，所以 $\lambda x \geq (1-\epsilon^2)\pi > 1 \geq \sin x$。

现在来确定它的位置。利用 $\sin x = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \cdots$，

$$\frac{\sin x}{x} = 1 - \frac{x^2}{6} + \frac{x^4}{120} - \cdots = 1-\epsilon^2 \quad\Longrightarrow\quad \frac{x^2}{6} - \frac{x^4}{120} + \cdots = \epsilon^2.$$

> ? **想一想** 为甚么可以舍去 x^4 项？我们还未知道 x 是否很小。

当 $\epsilon\to 0$ 时，$\frac{\sin x}{x}\to 1$，而由于 $\frac{\sin x}{x}$ 在 $(0,\pi)$ 上由 1 开始严格递减，这个根必定趋向 0。所以当 ϵ 很小时，根也很小，于是 x^4 与 x^2 相比可以忽略。取首阶近似，$x^2/6=\epsilon^2$，所以

$$x \approx \epsilon\sqrt{6}.$$

（若保留下一项，$x^2 \approx 6\epsilon^2 + \frac{6}{120}(6\epsilon^2)^2$，可得更精确的结果 $x\approx \epsilon\sqrt{6}\left(1+\frac{3}{20}\epsilon^2\right)$。）

第 (ii) 部分

设 $g(x)=\sin x - \lambda x$，则方程的解就是 g 在 $x>0$ 时的零点。由于当 $x>0$ 时 $\lambda x>0$，每个解都满足 $\sin x>0$，所以它必定位于其中一个拱

$$A_n = \big((2n-2)\pi,\,(2n-1)\pi\big), \qquad n=1,2,3,\ldots$$

之上。

第一个拱。在原点处，直线的斜率为 $\lambda<1$，比曲线的斜率小，所以直线起初在曲线下方；在 $x=\pi$ 处，直线则在曲线上方（$\lambda\pi>0=\sin\pi$）。由于 $\sin x$ 在 A_1 上是凹的，而直线是直的，两者在该处恰好相交一次。所以 A_1 上恰好有一个解，它必定是 x_1，并且 $0<x_1<\pi$。

其后的拱。在 $A_{n+1} = (2n\pi, (2n+1)\pi)$ 上，其中 $n \geq 1$，有 $g''(x) = -\sin x < 0$，所以 g 是凹的，最多有两个零点。在拱的两端，$\sin x = 0 < \lambda x$，所以该处 $g < 0$。g 的最大值出现在 $g'(x) = \cos x - \lambda = 0$ 之处，即 $c_n = 2n\pi + \arccos\lambda$，而且 $c_n < (2n + \frac{1}{2})\pi$，因为当 $\lambda > 0$ 时 $\arccos\lambda < \frac{\pi}{2}$。由此可知，$g$ 在 $(2n\pi, c_n)$ 上递增，在 $(c_n, (2n+1)\pi)$ 上递减，所以：

- 若 $g(c_n) < 0$，直线与这个拱不相交；
- 若 $g(c_n) = 0$，直线与这个拱相切于一点；
- 若 $g(c_n) > 0$，则恰好有两个解，一个在 $(2n\pi, c_n)$ 内，另一个在 $(c_n, (2n+1)\pi)$ 内。

此外，直线一旦与某个拱不相交，便与其后所有的拱都不相交：若 $x \in A_{n+2}$，则 $x - 2\pi \in A_{n+1}$，所以 $\sin x = \sin(x - 2\pi) < \lambda(x - 2\pi) < \lambda x$。

所以，与直线相交的拱是 $A_1, A_2, \ldots$ 直至最后某一个，它们分别贡献 $1, 2, 2, \ldots$ 个解，其后可能还有一个切点（tangency）。由于 $N = 2R + 1$ 是奇数，所以没有切点，而与直线相交的拱恰好是 $A_1, \ldots, A_{R+1}$。把各个解由小至大排列，拱 A_{n+1}（$1 \leq n \leq R$）包含 x_{2n} 和 x_{2n+1}，并且

$$2n\pi < x_{2n} < c_n < (2n + \tfrac{1}{2})\pi, \qquad 2n\pi < x_{2n+1} < (2n+1)\pi.$$

第一个不等式给出题目所求不等式的第二行。把 x_{2n+1} 写成 $x_{2(n+1)-1}$，第二个不等式连同 $0 < x_1 < \pi$，便给出 $(2m-2)\pi < x_{2m-1} < (2m-1)\pi$，其中 $m = 1, \ldots, R+1$。这就是所求的结果。

估算 N。当 λ 很小时，直线与很多个拱相交，所以 $R \geq 1$。在 x_{2R} 处，我们有 $\lambda x_{2R} = \sin x_{2R} \leq 1$，而由上文可知 $x_{2R} > 2R\pi$，所以

$$2R\pi\lambda < 1.$$

另一方面，直线与 $A_{R+2} = \big((2R+2)\pi, (2R+3)\pi\big)$ 不相交。在这个拱的中点 $(2R + \frac{5}{2})\pi$ 处，我们有 $\sin x = 1$，所以必须有 $\lambda(2R + \frac{5}{2})\pi > 1$（否则该处 $g \geq 0$，直线便会与这个拱相交）。把两者合并，

$$\frac{1}{2\pi\lambda} - \frac{5}{4} < R < \frac{1}{2\pi\lambda}, \qquad \text{所以} \qquad \frac{1}{\pi\lambda} - \frac{3}{2} < N < \frac{1}{\pi\lambda} + 1.$$

当 λ 非常小时，$\frac{1}{\pi\lambda}$ 非常大，相比之下这个有界的误差可以忽略，所以

$$N \approx \frac{1}{\pi\lambda}.$$

这是合理的：直线在 $x = 1/\lambda$ 处到达高度 1，而在这一点之前，$\sin x$ 大约有 $\frac{1}{2\pi\lambda}$ 个完整的周期，每个周期贡献两个交点。

验算

- 取 $\epsilon = 0.01$（$\lambda = 0.9999$），用二分法（bisection）求得根为 0.0244953，而 $\epsilon\sqrt{6} = 0.0244949$；两者之比为 1.000015，正好就是 $1 + \frac{3}{20}\epsilon^2$ 这个修正。
- 当 $\lambda = 0.001$ 时，电脑找到 $N = 317$ 个根，而 $\frac{1}{\pi\lambda} \approx 318.3$；当 $\lambda = 0.01$ 时，电脑找到 31 个根，而估算值为 31.8。就多个 λ 值所找到的全部根，都符合 (ii) 的区间不等式，而 N 也落在上述的界限之内。
- 当 $\lambda = 1.0001$、1.1、2 或 5 时，找不到任何根。

常见错误

- 忘记了 x_1 独自位于第一个拱上：原点也是一个交点，所以这个拱只贡献一个非零解，而不是两个。
- 没有给出理由，便把 x_{2n} 放在所属拱的前半部分。关键在于 g 要过了 $c_n = 2n\pi + \arccos\lambda$ 之后才开始递减，而这一点位于 $\sin x$ 的顶点之前。

STEP 之外

近似式 $x \approx \epsilon\sqrt{6}$ 是渐近展开（asymptotic expansion）的第一步，而在截断级数之前先论证根必定很小，正是分析课程所要求的那种严谨。利用凹性、单调性（monotonicity）和中间值定理（intermediate value theorem）来数算解的个数，是大学一年级分析课的标准内容；参看 Abbott 的 *Understanding Analysis*。一个相关的经典例子是 $\tan x = x$，它的根趋近 $(n + \frac{1}{2})\pi$，并且出现在热传导和振动问题之中。

详细解答

STEP III 样本试卷，第 6 题

纯数学 · 双曲函数归约公式

STEP III 样本试卷，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/spec-s3-q6

关键思路

被积函数（integrand）$\operatorname{sech}^{n+2}u\,\sinh^2 u$ 可以很自然地拆成 $\sinh u$ 乘以 $\sinh u\operatorname{sech}^{n+2}u$，而第二个因子除了一个常数倍之外，正是 $\operatorname{sech}^{n+1}u$ 的导数（derivative）。这样一来，分部积分法（integration by parts）就是顺理成章的做法。至于归约公式（reduction formula），恒等式 $\cosh^2 u-\sinh^2 u=1$ 让我们可以把同一个积分改写成两个 I 之差。

解答

积分恒等式

首先求一个稍后要用的导数。由于 $\operatorname{sech}u=(\cosh u)^{-1}$，由链式法则（chain rule）得

$$\frac{\mathrm{d}}{\mathrm{d}u}\operatorname{sech}^{n+1}u=-(n+1)\cosh^{-n-2}u\,\sinh u=-(n+1)\operatorname{sech}^{n+2}u\,\sinh u.$$

因此我们取以下各项作分部积分：

$$f=\sinh u,\qquad g'=\operatorname{sech}^{n+2}u\,\sinh u,\qquad \text{于是}\qquad f'=\cosh u,\qquad g=-\frac{\operatorname{sech}^{n+1}u}{n+1}.$$

由此得

$$\begin{aligned}\int_0^\infty \operatorname{sech}^{n+2}u\,\sinh^2 u\,\mathrm{d}u&=\Big[-\frac{\sinh u\operatorname{sech}^{n+1}u}{n+1}\Big]_0^\infty+\frac{1}{n+1}\int_0^\infty\cosh u\operatorname{sech}^{n+1}u\,\mathrm{d}u\\&=-\frac{1}{n+1}\Big[\tanh u\operatorname{sech}^n u\Big]_0^\infty+\frac{1}{n+1}\int_0^\infty\operatorname{sech}^n u\,\mathrm{d}u,\end{aligned}$$

当中用到 $\sinh u\operatorname{sech}^{n+1}u=\tanh u\operatorname{sech}^n u$ 及 $\cosh u\operatorname{sech}^{n+1}u=\operatorname{sech}^n u$。

想一想 为甚么边界项（boundary term）会消失？$n > 0$ 这个条件又在哪里用到？

当 $u = 0$ 时，$\tanh 0 = 0$。当 $u \to \infty$ 时，$\tanh u \to 1$，而 $\operatorname{sech} u = 2/(e^u + e^{-u}) \to 0$，所以 $\operatorname{sech}^n u \to 0$，原因正是 $n > 0$。（若 $n = 0$，该项会趋向 1，而不是 0。）同一个事实 $\operatorname{sech}^n u < 2^n e^{-nu}$ 亦说明了 I_n 在 $n > 0$ 时收敛（converge）。因此边界项为零，而

$$\int_0^\infty \operatorname{sech}^{n+2} u \, \sinh^2 u \, \mathrm{d}u = \frac{1}{n+1} I_n,$$

这正是所要证明的结果。

归约公式

现在用第二种方法计算同一个积分。由于 $\sinh^2 u = \cosh^2 u - 1$，

$$\operatorname{sech}^{n+2} u \, \sinh^2 u = \operatorname{sech}^{n+2} u \cosh^2 u - \operatorname{sech}^{n+2} u = \operatorname{sech}^n u - \operatorname{sech}^{n+2} u.$$

两部分在 $[0, \infty)$ 上的积分都收敛（因为 $n > 0$），所以

$$\int_0^\infty \operatorname{sech}^{n+2} u \, \sinh^2 u \, \mathrm{d}u = I_n - I_{n+2}.$$

与第一个结果比较，

$$I_n - I_{n+2} = \frac{1}{n+1} I_n \quad \Longrightarrow \quad I_{n+2} = \left(1 - \frac{1}{n+1}\right) I_n = \frac{n}{n+1} I_n,$$

即 $(n+1) I_{n+2} = n I_n$，这正是所要证明的结果。

I_6 的值

归约公式每次把 n 增加 2，所以我们需要一个 n 为偶数的起始值。最容易计算的是

$$I_2 = \int_0^\infty \operatorname{sech}^2 u \, \mathrm{d}u = \Big[\tanh u\Big]_0^\infty = 1 - 0 = 1.$$

（我们不能由 I_0 开始：它发散（diverge），而且在公式中代入 $n = 0$ 会得出 $I_2 = 0$ 这个错误的结论。这正是 $n > 0$ 发挥作用的地方。）于是

$$I_4 = \frac{2}{3} I_2 = \frac{2}{3},$$
$$I_6 = \frac{4}{5} I_4 = \frac{4}{5} \cdot \frac{2}{3} = \frac{8}{15}.$$

所以

$$I_6 = \frac{8}{15}.$$



验算

- 数列 $I_2, I_4, I_6 = 1, \frac{2}{3}, \frac{8}{15}$ 是递减的，这是必然的：$0 < \operatorname{sech} u \leq 1$，所以幂次愈高，数值愈小。
- 若改由 $I_1 = \int_0^\infty \operatorname{sech} u \, \mathrm{d}u = \left[2 \arctan(\mathrm{e}^u)\right]_0^\infty = \frac{\pi}{2}$ 开始，公式给出 $I_3 = \frac{1}{2} I_1 = \frac{\pi}{4}$，与直接作数值积分的结果吻合。
- 在 $[0, 60]$ 上使用辛普森法则（Simpson's rule）得 $I_6 \approx 0.5333333$，与 $\frac{8}{15}$ 相符；对于非整数的 n，例如 $n = 0.5$ 和 $n = 2.5$，恒等式和归约公式同样通过数值检验。

常见错误

- 在 $\frac{\mathrm{d}}{\mathrm{d}u} \operatorname{sech}^{n+1} u$ 中弄错正负号或因子 $n+1$：$\operatorname{sech} u$ 在 $u > 0$ 时递减，所以导数必定是负数。
- 没有说明边界项为何在无穷远处消失。这是唯一需要用到条件 $n > 0$ 的地方。
- 把答案写成 $\frac{4}{5} \cdot \frac{2}{3} \cdot I_2$，却没有计算 I_2 的值。

STEP 之外

这类归约公式与沃利斯公式（Wallis's formula）结构相同：设 $t = \tanh u$（或 $\sin \theta = \tanh u$，即古德曼代换（Gudermannian substitution）），I_n 便化为 $\int_0^{\pi/2} \cos^{n-1} \theta \, \mathrm{d}\theta$，而所得的递推关系（recurrence）正是沃利斯（Wallis）的递推关系。到了大学，这些积分会以贝塔函数（Beta function）的值再次出现，$I_n = \frac{1}{2} B\left(\frac{n}{2}, \frac{1}{2}\right)$，从而把递推关系推广至所有实数 $n > 0$；而我们处理边界项时的谨慎态度，正是反常积分（improper integral）理论的起点。若想看严谨的论述，可参阅 Abbott 的 *Understanding Analysis*；若想找更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP III 样本试卷，第 7 题

纯数学 · 降阶法

> **STEP III 样本试卷，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/spec-s3-q7

关键思路

当一条二阶线性方程（second-order linear equation）含有 x^2y''、xy' 和 y 时，最自然的做法是试 x 的幂：这样每一项都会变成 x^α 的倍数，而方程就化为关于 α 的条件。一旦知道一个解 y_1，设 $y = y_1 v$ 就是*降阶法*（reduction of order）：只含 v 本身的各项会互相抵消，剩下一条关于 v' 的一阶方程。

解答

幂函数解

试 $y = x^\alpha$（取 $x > 0$，使 x^α 对任意实数 α 都有意义）。则 $y' = \alpha x^{\alpha-1}$ 及 $y'' = \alpha(\alpha-1)x^{\alpha-2}$，所以

$$xy' - y = (\alpha - 1)x^\alpha, \qquad x^2y'' = \alpha(\alpha-1)x^\alpha.$$

代入后得

$$x^2y'' + (x-2)(xy' - y) = \big[\alpha(\alpha-1) + (x-2)(\alpha-1)\big]x^\alpha = (\alpha-1)(\alpha + x - 2)\,x^\alpha.$$

要使上式对*所有* x 都等于零，括号内的式子必须恒等于零。因子 $\alpha + x - 2$ 取决于 x，不可能恒等于零；因此 $\alpha = 1$。

所以 $y = x$ 是一个解，而这一点可以直接看出：$y'' = 0$ 且 $xy' - y = x - x = 0$。于是任何倍数 $y = Ax$ 也是解，这就是所谓解“与 x^α 成正比”的意思。

通解

按照提示，设 $y = xv$，其中 v 是 x 的函数。则

$$y' = v + xv', \qquad y'' = 2v' + xv'',$$

而方程中出现的那个组合可以化简得很漂亮：

$$xy' - y = xv + x^2v' - xv = x^2v'.$$

代入方程，

$$\begin{aligned} 0 &= x^2(2v' + xv'') + (x-2)x^2v' \\ &= x^2\left(xv'' + 2v' + xv' - 2v'\right) \\ &= x^3\left(v'' + v'\right). \end{aligned}$$

想一想 为甚么所有只含 v 本身（不带导数）的项都消失了？

因为 $y = x$ 是一个解：假如 v 是常数，$y = xv$ 就是 x 的倍数，左边便等于零。所以只有 v' 和 v'' 能够留下来。降阶法之所以行得通，正是这个原因，而这也是检查代数运算的好方法。

在任何不包含 0 的区间上，我们可以除以 x^3，得到

$$v'' + v' = 0.$$

这是关于 $w = v'$ 的一阶方程：$w' = -w$，所以 $w = Be^{-x}$，其中 B 为常数。再积分一次（并把正负号并入常数之中），

$$v = A + Be^{-x}.$$

因此通解（general solution）为

$$\boxed{y = Ax + Bxe^{-x}}$$

其中 A 和 B 为任意常数。它含有两个任意常数，二阶线性方程的通解必须如此，而且 x 与 xe^{-x} 这两部分并非彼此的倍数。

直接验证新的那一部分：$y = xe^{-x}$ 给出 $y' = (1-x)e^{-x}$ 及 $y'' = (x-2)e^{-x}$，所以

$$x^2y'' + (x-2)(xy' - y) = (x-2)e^{-x}\left[x^2 + x(1-x) - x\right] = 0.$$

关于 $x = 0$ 的一点说明。y'' 的系数在 $x = 0$ 处等于零，所以我们只在远离 0 的地方除以 x^3。原则上，在 $x > 0$ 和 $x < 0$ 两边可以取不同的常数；但每一个这样的函数都有 $y(0) = 0$、$y'(0^{\pm}) = A_{\pm} + B_{\pm}$ 及 $y''(0^{\pm}) = -2B_{\pm}$，所以要使 y 在 0 处二阶可导，必须有 $B_+ = B_-$，继而 $A_+ = A_-$。因此 $y = Ax + Bxe^{-x}$ 在整条实数线上也是通解。

验算

- 如上所述，把 $y = x$ 和 $y = xe^{-x}$ 分别代回方程：两者都得出零。
- 数值验算：由 $x = 1$ 出发，以 $y(1) = y_0$、$y'(1) = y_1$ 逐步求解方程，公式预测 $A = y_1$、$B = (y_0 - y_1)e$。取 $y_0 = 2$、$y_1 = 1$，得 $y(4) = 4 + 4e^{-3} \approx 4.199$；以四阶龙格—库塔法（Runge–Kutta）作数值积分，对若干个随机初始值，结果的误差都小于 10^{-8}。

常见错误

- 由因子 $(\alpha + x - 2)$ 得出 $\alpha = 2 - x$：α 必须是常数，所以这个根不可取。
- 对 $y = xv$ 求 y'' 时出错（应为 $2v' + xv''$，而不是 $v' + xv''$）。如果含 v 的项没有抵消，就表示求导时出了错。
- 停在 $v' = Be^{-x}$，在积分求 v 时忘记第二个常数。

STEP 之外

降阶法适用于任何二阶线性方程 $y'' + p(x)y' + q(x)y = 0$：已知一个解 y_1，代换 $y = y_1 v$ 总会给出一条关于 v' 的一阶线性方程，而第二个解可以写成一个涉及 y_1 和 p 的积分。它与朗斯基行列式（Wronskian）和阿贝尔（Abel）公式关系密切，而在 $x = 0$ 这类点上试 x^{α}，正是以弗罗贝尼乌斯（Frobenius）方法求级数解的第一步。以上内容在 Boyce 与 DiPrima 的 *Elementary Differential Equations*（微分方程入门教科书）中都有论述。

详细解答

STEP III 样本试卷，第 8 题

纯数学 · 向量与柯西–施瓦兹不等式

STEP III 样本试卷，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/spec-s3-q8

关键思路

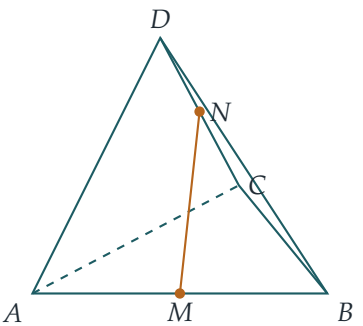
在第 (i) 部分，先为四个顶点设置位置向量（position vector）。我们需要的每一个长度，都是这些向量的某个简单组合的模（modulus），而长度的平方就是可以展开的点积（dot product）。于是两边都会化成同一个式子，而大部分的整理工作都可以靠对称性（symmetry）完成。在第 (ii) 部分，左边看起来像是与固定向量 $(3,2,6)$ 的点积，而这个向量的长度刚好是整数。这正是柯西－施瓦茨不等式（Cauchy–Schwarz inequality）大派用场的地方。

解答

第 (i) 部分

设四面体（tetrahedron）为 $ABCD$，各顶点相对于任意原点的位置向量为 $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$。它的六条棱可分成三对对棱（即没有公共顶点的棱）：

$$AB \text{ 与 } CD, \qquad AC \text{ 与 } BD, \qquad AD \text{ 与 } BC.$$



考虑 AB、CD 这一对：两条棱的中点分别是 $M = \frac{1}{2}(\mathbf{a}+\mathbf{b})$ 和 $N = \frac{1}{2}(\mathbf{c}+\mathbf{d})$，所以

$$MN^2 = \tfrac{1}{4}\,|\mathbf{c}+\mathbf{d}-\mathbf{a}-\mathbf{b}|^2.$$

另外两条中线（midline）可用同样方法求得，因此三条中线的平方和的四倍是

$$S = |-\mathbf{a}-\mathbf{b}+\mathbf{c}+\mathbf{d}|^2 + |-\mathbf{a}+\mathbf{b}-\mathbf{c}+\mathbf{d}|^2 + |-\mathbf{a}+\mathbf{b}+\mathbf{c}-\mathbf{d}|^2.$$

利用 $|\mathbf{u}|^2 = \mathbf{u}\cdot\mathbf{u}$ 展开每一项。每一项都给出 $|\mathbf{a}|^2+|\mathbf{b}|^2+|\mathbf{c}|^2+|\mathbf{d}|^2$，另加交叉项 $\pm 2\,\mathbf{u}\cdot\mathbf{v}$；若 $\mathbf{u}$ 和 $\mathbf{v}$ 在该项中同号，交叉项便取正号。

想一想 把三项的交叉项加起来，结果会怎样？

任取一对顶点，例如 $\mathbf{a}$ 和 $\mathbf{b}$。它们在第一项中同号，在另外两项中异号，所以它们的交叉项合共是 $2\,\mathbf{a}\cdot\mathbf{b} - 2\,\mathbf{a}\cdot\mathbf{b} - 2\,\mathbf{a}\cdot\mathbf{b} = -2\,\mathbf{a}\cdot\mathbf{b}$。每一对顶点都是如此：三项中的每一项都把四个顶点分成两对（一对顶点连同与它相对的一对），而每一对顶点恰好只在三项中的其中一项同号。因此

$$S = 3\big(|\mathbf{a}|^2+|\mathbf{b}|^2+|\mathbf{c}|^2+|\mathbf{d}|^2\big) - 2\big(\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}+\mathbf{a}\cdot\mathbf{d}+\mathbf{b}\cdot\mathbf{c}+\mathbf{b}\cdot\mathbf{d}+\mathbf{c}\cdot\mathbf{d}\big).$$

现在看各条棱。例如 $AB^2 = |\mathbf{b}-\mathbf{a}|^2 = |\mathbf{a}|^2+|\mathbf{b}|^2-2\,\mathbf{a}\cdot\mathbf{b}$。把六条棱全部加起来：每个顶点都在三条棱上，所以每个 $|\mathbf{a}|^2$ 出现三次，而每一对顶点给出一个 $-2\,\mathbf{u}\cdot\mathbf{v}$：

$$AB^2+AC^2+AD^2+BC^2+BD^2+CD^2 = 3\sum|\mathbf{a}|^2 - 2\sum_{\text{各对}}\mathbf{a}\cdot\mathbf{b} = S.$$

所以各棱的平方和，等于连接对棱中点的三条线段的平方和的四倍，这正是所要证明的结果。

第 (ii) 部分

取

$$\mathbf{a} = \begin{pmatrix}3\\2\\6\end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix}x\\y\\1\end{pmatrix}, \qquad \text{于是} \qquad \mathbf{a}\cdot\mathbf{b} = 3x+2y+6,$$

其中 $|\mathbf{a}| = \sqrt{9+4+36} = 7$，$|\mathbf{b}| = \sqrt{x^2+y^2+1}$。由于任何实数都不大于它的绝对值，

$$3x+2y+6 = \mathbf{a}\cdot\mathbf{b} \le |\mathbf{a}\cdot\mathbf{b}| \le |\mathbf{a}|\,|\mathbf{b}| = 7\sqrt{x^2+y^2+1},$$

这就是所求的不等式。

等号成立的情况。两步都必须取等号。

- $|\mathbf{a}\cdot\mathbf{b}| = |\mathbf{a}||\mathbf{b}|$ 恰好在 $\mathbf{a}$ 与 $\mathbf{b}$ 平行时成立，因为 $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta$，而我们需要 $|\cos\theta| = 1$。由于 $\mathbf{a} \neq \mathbf{0}$，这表示存在某个 λ 使得 $\mathbf{b} = \lambda\mathbf{a}$。比较第三个分量得 $1 = 6\lambda$，所以 $\lambda = \frac{1}{6}$，而 $x = \frac{1}{2}$，$y = \frac{1}{3}$。
- $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}\cdot\mathbf{b}|$ 需要 $\mathbf{a}\cdot\mathbf{b} \geq 0$。这里 $\mathbf{a}\cdot\mathbf{b} = \lambda|\mathbf{a}|^2 = \frac{49}{6} > 0$，所以这一步也成立。

直接验证：当 $x = \frac{1}{2}, y = \frac{1}{3}$ 时，左边是 $\frac{3}{2} + \frac{2}{3} + 6 = \frac{49}{6}$，右边是 $7\sqrt{\frac{1}{4} + \frac{1}{9} + 1} = 7\cdot\frac{7}{6} = \frac{49}{6}$。所以等号成立当且仅当

$$x = \tfrac{1}{2}, \qquad y = \tfrac{1}{3}.$$

验算

- 对于棱长为 1 的正四面体（regular tetrahedron），六条棱的平方和是 6。把它的顶点放在边长为 $\frac{1}{\sqrt{2}}$ 的正方体上相间的四个顶点，每条中线都连接一对相对的面的中心，所以长度是 $\frac{1}{\sqrt{2}}$，而 $4 \times 3 \times \frac{1}{2} = 6$。
- 以 10^4 个随机四面体作数值测试，结果与恒等式吻合，误差只在舍入误差范围之内。至于 (ii)，$7\sqrt{x^2+y^2+1} - (3x+2y+6)$ 在 2×10^5 个随机点上都是正数，而它在一个细密网格上的最小值出现在 $(0.5, 0.335)$，接近 $(\frac{1}{2}, \frac{1}{3})$，而在该点它的值是 0。

常见错误

- 漏掉 $\mathbf{a}\cdot\mathbf{b} \leq |\mathbf{a}\cdot\mathbf{b}|$ 这一步，或者在讨论等号成立的情况时忽略了它。在这题中因为 $\lambda > 0$，这一步不会出问题，但仍然必须检查：假如 $\lambda < 0$，两边的正负号就会不同。
- 用了 $(3,2,6)$ 和 $(x,y,1)$，却没有把 $|\mathbf{a}|$ 算成 7。这题是特意设计的，使 $3^2+2^2+6^2$ 是一个完全平方数。

STEP 之外

第 (i) 部分是平行四边形恒等式（parallelogram law）$|\mathbf{u}+\mathbf{v}|^2 + |\mathbf{u}-\mathbf{v}|^2 = 2|\mathbf{u}|^2 + 2|\mathbf{v}|^2$ 在三维空间中的近亲。事实上，三条中线相交于形心（centroid），并且互相平分。第 (ii) 部分就是柯西－施瓦茨不等式，它在任何内积空间（inner-product space）中都成立，包括由函数组成的空间，在那里它可以用来估计积分的上界。想在更广阔的背景下认识这两个想法，Strang 的 *Introduction to Linear Algebra*（线性代数入门）是很好的起点。

详细解答

STEP III 样本试卷，第 9 题

纯数学 · 矩阵群

> **STEP III 样本试卷，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/spec-s3-q9

关键思路

一切都源自一个计算：把两个这种形状的一般矩阵相乘，看看结果是甚么。原来乘积仍是同一形状，各元素由一条简单的公式给出，而这条公式几乎不费吹灰之力就给了我们封闭性（closure）、单位元（identity）、逆元（inverse）、整数子群（subgroup）和可交换（commute）的条件。记

$$M(x,y,z)=\begin{pmatrix}1&0&0\\x&1&0\\y&z&1\end{pmatrix}.$$

解答

G 是一个群

乘积公式。以行乘列，得

$$M(a,b,c)\,M(x,y,z)=\begin{pmatrix}1&0&0\\a&1&0\\b&c&1\end{pmatrix}\begin{pmatrix}1&0&0\\x&1&0\\y&z&1\end{pmatrix}=\begin{pmatrix}1&0&0\\a+x&1&0\\b+cx+y&c+z&1\end{pmatrix},$$

即

$$M(a,b,c)\,M(x,y,z)=M(a+x,\ b+cx+y,\ c+z). \tag{$*$}$$

封闭性。元素 $a+x$、$b+cx+y$ 和 $c+z$ 都是实数，所以乘积属于 G。

结合律（associativity）。题目容许我们假设矩阵乘法满足结合律。

单位元。单位矩阵是 $I = M(0,0,0)$，它属于 G，而且对每个矩阵 M 都有 $IM = MI = M$。

逆元。给定 $M(a,b,c)$，我们要找出 $M(x,y,z)$ 使乘积为 $M(0,0,0)$。由 $(*)$，我们需要 $a+x=0$、$c+z=0$ 和 $b+cx+y=0$，所以 $x=-a$、$z=-c$ 及 $y=-b-cx=ac-b$。由于矩阵乘法不满足交换律，另一个次序也要验证：

$$M(-a,\, ac-b,\, -c)\, M(a,b,c) = M\big(0,\ (ac-b)+(-c)a+b,\ 0\big) = M(0,0,0).$$

所以 $M(a,b,c)^{-1} = M(-a,\, ac-b,\, -c)$，它属于 G。

群（group）的四条公理全部成立，所以 G 在矩阵乘法下是一个群。

整数矩阵

设 H 为所有满足 $x,y,z \in \mathbb{Z}$ 的 $M(x,y,z)$ 所组成的集合。它是 G 的子集，并且包含 $I = M(0,0,0)$。若 a,b,c,x,y,z 都是整数，则 $a+x$、$b+cx+y$ 和 $c+z$ 也是整数，所以由 $(*)$ 可知 H 在乘法下封闭。逆元 $M(-a,\, ac-b,\, -c)$ 的元素也都是整数，所以 H 在取逆元下封闭。结合律则由 G 承袭而来。所以答案是**是**，H 是 G 的子群。

> ? **想一想** 为甚么单靠乘法下的封闭性并不足够?

对于无限子集，这是可以出错的：x,y,z 为非负整数的矩阵 $M(x,y,z)$ 在乘法下封闭，但 $M(1,0,0)$ 的逆元是 $M(-1,0,0)$，并不在该集合之内。在这里，逆元公式令我们始终留在整数之内，正是这一点使 H 成为子群。

与所有元素可交换的元素

取 $A = M(a,b,c)$ 和一般的 $B = M(x,y,z)$。由 $(*)$，按两个次序分别相乘，得

$$AB = M(a+x,\ b+cx+y,\ c+z), \qquad BA = M(x+a,\ y+za+b,\ z+c).$$

第一个和第三个元素必定相同，所以

$$AB = BA \iff cx = az.$$

（特别地，G 不是阿贝尔群（abelian group）：$M(1,0,0)$ 与 $M(0,0,1)$ 不可交换。）

要 A 与每一个 B 可交换，我们需要对所有实数 x 和 z 都有 $cx = az$。取 $x=1$、$z=0$ 得 $c=0$；取 $x=0$、$z=1$ 得 $a=0$。反过来，若 $a=c=0$，则对所有 x,z 都有 $cx = 0 = az$。所以与整个 G 可交换的元素恰好是

$$A = M(0,b,0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{pmatrix}, \qquad b \in \mathbb{R}.$$

有，这样的 A 存在：任何 $b \neq 0$ 都给出 $A \neq I$，例如 $A = M(0,1,0)$。由 $(*)$，对 G 中每一个 $B = M(x,y,z)$ 都有 $M(0,1,0)\, M(x,y,z) = M(x,\, 1+y,\, z) = M(x,y,z)\, M(0,1,0)$。

验算

- G 中每个矩阵都是下三角（lower triangular）矩阵，而且对角线上全是一，所以它的行列式（determinant）为 1，必定可逆：我们求得的逆矩阵必然存在，而且同样是这种形状。
- 数值验算：取 $A = M(2,3,5)$ 和 $B = M(1,0,0)$，由 $(*)$ 得 $AB = M(3,8,5)$，但 $BA = M(3,3,5)$，所以 A 与 B 不可交换，正如 $cx = 5 \neq 0 = az$ 所预测的一样。我们亦以分数作精确运算，用数千个随机矩阵验证了乘积公式、逆元公式和可交换的条件。

常见错误

- 只验证一边的逆元。在非阿贝尔的情况下，值得把两个乘积都确认一次（或者指出对方阵而言，单边逆元自动就是双边逆元）。
- 以“G 不是阿贝尔群，所以没有”来回答最后一部分。非阿贝尔群仍然可以有与所有元素可交换的非单位元元素，而这个群正是如此。
- 忘记逆命题：证明了 $a = c = 0$ 是必要条件之后，我们还须验证这样的 A 确实与每一个 B 可交换。

STEP 之外

这就是*海森堡群*（*Heisenberg group*），而与所有元素可交换的元素组成它的*中心*（*centre*）$Z(G)$，在这里就是矩阵 $M(0,b,0)$，相当于实数在加法下所成的群的一个复本。整数版本 H 是离散海森堡群，是群论和几何学中备受喜爱的例子；海森堡群本身则通过位置与动量的对易关系出现在量子力学之中。关于群、子群和中心，可参阅 Allenby 的 *Rings, Fields and Groups*。

详细解答

STEP III 样本试卷，第 10 题

纯数学 · 模算术

STEP III 样本试卷，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/spec-s3-q10

关键思路

两部分都与余数（remainder）有关，所以应选一个合适的数，以它为模（modulo）来计算。在 (i) 中，平方数模 8 只能取少数几个值，而这些值中任取三个，相加都不可能得到 7。在 (ii) 中，把各个幂改写，使两项在模 17 下有相同的底；这样两个系数之和便是 17 的倍数。

解答

第 (i) 部分

*奇数的平方。*把奇数写成 $2k+1$，其中 k 为整数。则

$$(2k+1)^2 = 4k^2 + 4k + 1 = 4k(k+1) + 1.$$

两个连续整数 k 与 $k+1$ 之中必有一个是偶数，所以 $k(k+1)$ 是偶数，而 $4k(k+1)$ 是 8 的倍数。因此，每个奇数的平方除以 8 的余数都是 1。

*偶数的平方。*把偶数写成 $2k$。则 $(2k)^2 = 4k^2$。

- 若 k 是偶数，设 $k = 2j$，则 $4k^2 = 16j^2$，是 8 的倍数：余数为 0。
- 若 k 是奇数，则 k^2 是奇数，设 $k^2 = 2j+1$，于是 $4k^2 = 8j + 4$：余数为 4。

所以每个偶数的平方除以 8 的余数是 0 或 4。

*推论。*每个平方数模 8 都同余（congruent）于 0、1 或 4（这包括 $0^2 = 0$，即我们容许三个平方数中有些是零）。假设 $N = a^2 + b^2 + c^2$。则 $N \bmod 8$ 是三个各取自 $\{0, 1, 4\}$ 的数之和的余数。按其中有多少个等于 4 来分类：

4 的个数	其余各数之和的可能值，以及总和模 8 的值
0	$0,1,2,3$
1	$4+\{0,1,2\}=4,5,6$
2	$8+\{0,1\}\equiv 0,1$
3	$12\equiv 4$

可能的余数是 $0,1,2,3,4,5,6$，而 7 从未出现。形如 $8n+7$ 的数除以 8 的余数是 7，所以它不可能是三个平方数之和。

> ? **想一想** 为甚么不能只验证 7 本身不是三个平方数之和?

因为这个命题涉及无穷多个数 $7,15,23,\ldots$，逐一验证永远做不完。以 8 为模来计算，就能一次过处理所有情况：无论 a、b、c 是甚么，要紧的只是它们的余数，而上表已涵盖每一种情形。

第 (ii) 部分

把各个幂拆开，使每一项都是常数乘以一个 n 次幂：

$$2^{3n+1}=2\cdot 8^n, \qquad 3\left(5^{2n+1}\right)=15\cdot 25^n.$$

由于 $25=17+8$，故 $25\equiv 8 \pmod{17}$，因此对所有 $n\geq 0$ 都有 $25^n\equiv 8^n \pmod{17}$（同余式可以相乘）。于是

$$\begin{aligned} 2^{3n+1}+3\left(5^{2n+1}\right) &= 2\cdot 8^n+15\cdot 25^n \\ &\equiv 2\cdot 8^n+15\cdot 8^n \\ &= 17\cdot 8^n\equiv 0 \pmod{17}. \end{aligned}$$

所以对所有整数 $n\geq 0$，17 整除 $2^{3n+1}+3(5^{2n+1})$，这正是所要证明的。

> ? **想一想** 如果我没有学过同余，怎么办?

同样的想法可以用归纳法（induction）来实现。设 $f(n)=2\cdot 8^n+15\cdot 25^n$。则 $f(0)=17$，并且

$$f(n+1)-8f(n)=15\cdot 25^{n+1}-15\cdot 8\cdot 25^n=15\cdot 25^n(25-8)=17\cdot 15\cdot 25^n,$$

所以 $f(n+1)=8f(n)+17\cdot 15\cdot 25^n$。若 $17\mid f(n)$，则 $17\mid f(n+1)$，由归纳法即得结论。

验算

试试较小的情况。$7=4+1+1+1$ 需要四个平方数，而 $15=9+4+1+1$ 同样需要四个。另一方面，$6=4+1+1$ 和 $14=9+4+1$ 则没有问题。电脑搜索证实，在 5000 以内

没有任何形如 $8n+7$ 的数是三个平方数之和。至于 (ii)：$n=0$ 时得 $2+15=17$，$n=1$ 时得 $16+375=391=17\times 23$。我们亦已验证，对由 0 至 299 的每个 n，该式都可被 17 整除。

常见错误

- 忘记 0 也是平方数，所以“三个平方数之和”其实可能只用了一个或两个非零的平方数。上面的论证已自动包括这种情况，因为 0 是其中一个剩余（residue）。
- 以“$4k^2+4k$ 可被 4 整除”为理由，说奇数的平方是 1 mod 8：这只能得出 1 mod 4。你还需要 $k(k+1)$ 所提供的因数 2。

STEP 之外

第 (i) 部分是勒让德（Legendre）三平方和定理（three-square theorem）中较容易的一半：一个正整数是三个平方数之和，当且仅当它不是 $4^a(8n+7)$ 的形式。逆命题（converse）的方向困难得多，需要用到二次型（quadratic forms）的理论。相比之下，拉格朗日（Lagrange）证明了每一个正整数都是四个平方数之和。第 (ii) 部分是 $\mathbb{Z}/17\mathbb{Z}$ 中的日常算术，大学一年级的课程会由此发展出群（groups）和环（rings）的研究。Davenport 的 *The Higher Arithmetic*（《高等算术》）对同余和平方和有十分精彩的讲解。

详细解答

STEP III 样本试卷，第 11 题

力学 • 一阶微分方程

STEP III 样本试卷，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/spec-s3-q11

本题在 *Beyond STEP* 第 11.13 节铲雪车问题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch11/snowplough.html

关键思路

每部铲雪车的速率都以积雪深度表示，而深度则取决于铲雪车何时到达某一点。因此，很自然会把时间看成位置的函数，并写出 $\mathrm{d}t/\mathrm{d}x$，即速率的倒数（reciprocal）。对第一部铲雪车而言，深度就是原有深度加上其后落下的雪；对第二部而言，关键的观察是：它前方的雪，只是从第一部铲雪车清理该处之后才开始累积的。

解答

把每部铲雪车看成沿路移动的一点，并由共同的起点量度距离。全文中，深度以米为单位，时间以秒为单位，距离以米为单位。

第 (i) 部分

在时间 t，第一部铲雪车遇到的雪自 $t = 0$ 起一直不受干扰地落下，所以其深度为

$$z = h + kt.$$

因此速率为

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{k}{\alpha(h+kt)}, \qquad \text{所以} \qquad \frac{\mathrm{d}t}{\mathrm{d}x} = \frac{\alpha(h+kt)}{k} = \alpha t + \frac{\alpha h}{k}.$$

（这里取倒数并无问题，因为速率为正，所以 x 是 t 的严格递增函数（strictly increasing function）。）这个方程对 t 是线性的，最简洁的写法是

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(t+\frac{h}{k}\Big)=\alpha\Big(t+\frac{h}{k}\Big).$$

由此得 $t+h/k=C\mathrm{e}^{\alpha x}$，其中 C 为常数。当 $x=0$ 时 $t=0$，所以 $C=h/k$，而行走 x 米所需的时间为

$$t_1(x)=\big(\mathrm{e}^{\alpha x}-1\big)\frac{h}{k}\text{ 秒},$$

正如所求。

第 (ii) 部分

建立方程。 设在时间 $t>T$，第二部铲雪车位于位置 y。第一部铲雪车在第 (i) 部分所得的时间 $t_1(y)$ 经过这一点，并把该处的雪全部清除。

想一想 第二部铲雪车前方的积雪深度是多少?

不是 $h+kt$：第一部铲雪车驶过后，留下的是光秃秃的路面。在位置 y，雪自时间 $t_1(y)$ 起由零开始以速率 k 累积，所以深度为

$$z=k\big(t-t_1(y)\big).$$

（这需要 $t>t_1(y)$，即第二部铲雪车仍在第一部后面；在两车相遇之前，这一点都成立。）因此第二部铲雪车的速率为

$$\frac{\mathrm{d}y}{\mathrm{d}t}=\frac{k}{\alpha k\big(t-t_1(y)\big)}\qquad\Longrightarrow\qquad\frac{1}{\alpha}\frac{\mathrm{d}t}{\mathrm{d}y}=t-\big(\mathrm{e}^{\alpha y}-1\big)\frac{h}{k},$$

这正是题目给出的方程。第二部铲雪车在时间 T 由原点出发，所以初始条件（initial condition）是：当 $y=0$ 时 $t=T$。

验证解。 把题目提出的解写成

$$t=t_1(y)+g(y),\qquad g(y)=\Big(T-\frac{\alpha h y}{k}\Big)\mathrm{e}^{\alpha y}.$$

于是 $t-t_1(y)=g(y)$，所以我们需要 $\frac{1}{\alpha}\,\mathrm{d}t/\mathrm{d}y=g(y)$。求导，并对 g 使用积法则（product rule），得

$$\begin{aligned}\frac{\mathrm{d}t}{\mathrm{d}y}&=\frac{\alpha h}{k}\mathrm{e}^{\alpha y}\;-\;\frac{\alpha h}{k}\mathrm{e}^{\alpha y}\;+\;\alpha\Big(T-\frac{\alpha h y}{k}\Big)\mathrm{e}^{\alpha y}\\&=\alpha\,g(y),\end{aligned}$$

其中第一项是 $t_1'(y)$，其后两项是 $g'(y)$。所以微分方程（differential equation）成立。此外，当 $y=0$ 时，$t=0+T\cdot 1=T$，所以初始条件亦成立。

为甚么这就是所求的唯一解？这个方程是一阶线性方程：若 t 和 $\tilde{t}$ 都满足它，则两者之差 u 满足 $\mathrm{d}u/\mathrm{d}y = \alpha u$，所以 $u = u(0)\mathrm{e}^{\alpha y}$。两者都满足 $t(0) = T$，所以 $u(0) = 0$，从而 $u \equiv 0$。因此，通过 $(y,t) = (0,T)$ 的解恰好只有一个，而我们已经找到了它。

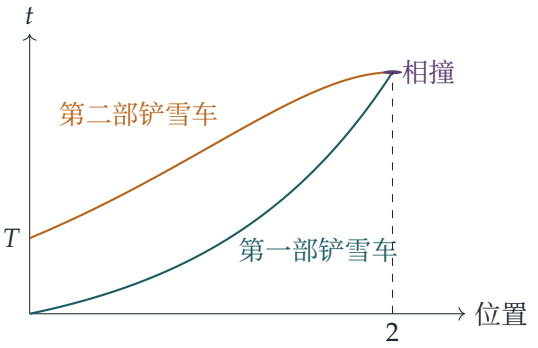
（如果你想直接把它推导出来：乘以积分因子（integrating factor）$\mathrm{e}^{-\alpha y}$ 可得 $\frac{\mathrm{d}}{\mathrm{d}y}\left(t\mathrm{e}^{-\alpha y}\right) = -\frac{\alpha h}{k}\left(1 - \mathrm{e}^{-\alpha y}\right)$，再由 0 积分至 y，便会得到同一条公式。）

相撞。 若两部铲雪车在同一时间到达某一点 y，它们便会在该点相撞；也就是说 $t = t_1(y)$，即 $g(y) = 0$。由于 $\mathrm{e}^{\alpha y} > 0$，这种情况恰好在以下条件成立时出现：

$$T - \frac{\alpha h y}{k} = 0 \qquad \Longleftrightarrow \qquad y = \frac{kT}{\alpha h}.$$

当 $0 \le y < kT/(\alpha h)$ 时，我们有 $g(y) > 0$（取 $\alpha, h, k > 0$，因为速率必须为正），所以第二部铲雪车到达每个这样的点的时间都严格迟于第一部，两车不会更早相遇。因此，两部铲雪车在行走了 $kT/(\alpha h)$ 米时相撞。 ■

下图显示当 $h = k = \alpha = 1$ 及 $T = 2$ 时，两条路径在 (y,t) 平面上的情况；两条曲线之间的时间差 $g(y)$ 在 $y = 2$ 处缩减至零。



验算

- 单位：α 的单位是 m^{-1}（因为 $k/(\alpha z)$ 是速率），所以 αx 是无量纲（dimensionless）的，h/k 是时间，而 $kT/(\alpha h)$ 是长度。
- 若 $T = 0$，两部铲雪车同时出发，公式给出在距离 0 处相撞；第一部铲雪车领先的时间 T 越大，相撞的地点便沿路推得越远，这与预期相符。
- 数值上，对多组随机选取的 h、k、α 和 T 的值，把每部铲雪车的运动逐步向前推算（对第二部而言，是由 $t = T$ 时 $y = 0$ 出发，按 $\mathrm{d}y/\mathrm{d}t = 1/\big(\alpha(t - t_1(y))\big)$ 推算），所得结果与公式相符，误差小于 10^{-8}。取 $h = k = \alpha = 1$ 及 $T = 2$，如图中所示，两部铲雪车都在时间 $\mathrm{e}^2 - 1 \approx 6.39\,\mathrm{s}$ 到达 $y = 2$。

常见错误

- 对第二部铲雪车也使用深度 $h+kt$。它所铲的路面在时间 $t_1(y)$ 已被清理过，所以深度是 $k(t-t_1(y))$。
- 验证解的时候忘记了初始条件：当 $y=0$ 时 $t=T$；单单是微分方程的解并不足够。
- 写成 $\mathrm{d}y/\mathrm{d}t$ 而不是 $\mathrm{d}t/\mathrm{d}y$：以时间作为因变量（dependent variable），方程便会变成线性，容易处理。

STEP 之外

这属于经典的“铲雪车问题”（snowplough problem）一类题目，而把因变量与自变量（independent variable）的角色互换、使方程变成线性的技巧，在大学一年级的微分方程课程中经常用到。上文的唯一性论证，是初值问题（initial value problem）的存在唯一性定理（existence and uniqueness theorem）的一个最简单特例。还要留意，当第二部铲雪车前方的积雪薄至全无时，它的速率会趋向无限大，这反映的是模型的局限，而不是现实情况。Boyce and DiPrima, *Elementary Differential Equations*（初等微分方程）涵盖一阶线性方程、积分因子和唯一性。

详细解答

STEP III 样本试卷，第 12 题

力学 • 杆的平衡

STEP III 样本试卷，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：

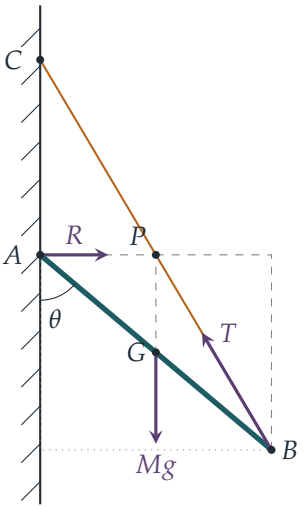
step.maths.org/questions/spec-s3-q12

关键思路

杆只受三个力作用：它的重量、光滑墙壁的反作用力（reaction），以及绳的张力（tension）。使物体保持平衡（equilibrium）的三个不平行的力必定通过同一点，而在这题中，这一点很容易找到。它完全确定了整个几何结构，之后只需分解一次力，便可求得张力。

解答

模型。杆是均匀的，所以它的重量 Mg 作用于中点 G。墙壁是光滑的，所以在 A 的反作用力 R 是水平的（垂直于墙壁）。绳是轻而不可伸长的，所以它以张力 T 拉着 B，方向沿 BC。



杆的倾斜角度

反作用力 R 沿着通过 A 的水平线作用，而重量则沿着通过 G 的铅垂线作用。这两条线相交于点 P，它与 A 同高，并位于 G 的正上方（或正下方）。由于三个力处于平衡，它们对 P 的力矩（moment）必须互相抵销；R 和 Mg 对 P 都没有力矩，所以 T 也没有，因此绳所在的直线必定通过 P。

> ? **想一想** 杆从 A 出发，是向上还是向下?

从墙壁起量度水平距离。B 与墙壁的距离是 $L\sin\theta$，其中 θ 是杆与铅垂线之间的夹角，而 G（因而 P 也一样）与墙壁的距离是它的一半，即 $\frac{1}{2}L\sin\theta$。绳从 B 成一直线伸延至墙上的 C，所以水平方向上走了一半的点，也就是绳上走了一半的点：P 是 BC 的中点。因此 P 的高度是 B 和 C 两点高度的平均值。但 P 与 A 同高，而 C 在 A 之上，所以 B 必定在 A 之下，而且低于它的距离恰好等于 AC。

所以杆是从 A 向下垂的。设 θ 为杆与向下铅垂线的夹角。那么 B 在 A 之下的深度是 $L\cos\theta$，于是 $AC = L\cos\theta$。相对于 B 来说，点 C 的水平距离是 $L\sin\theta$，高度则高出 $L\cos\theta + L\cos\theta = 2L\cos\theta$。由于绳长为 αL，由毕氏定理（Pythagoras' theorem）得

$$\alpha^2L^2 = L^2\sin^2\theta + 4L^2\cos^2\theta = L^2\left(1 + 3\cos^2\theta\right),$$

所以

$$\cos^2\theta = \frac{\alpha^2 - 1}{3}, \qquad \text{即} \qquad \theta = \cos^{-1}\sqrt{\frac{\alpha^2-1}{3}},$$

而杆是从 A 向下伸出，与铅垂线成这个角度。

α 的可能值

我们需要 C 严格地在 A 之上，所以 $\cos\theta > 0$。（若 $\cos\theta = 0$，则 C 与 A 重合，杆是水平的；这时 A、B、C 共线（collinear）。）我们亦需要 A、B、C 不共线，所以杆不是铅垂的：$\sin\theta \neq 0$，即 $\cos^2\theta < 1$。因此

$$0 < \frac{\alpha^2 - 1}{3} < 1 \iff 1 < \alpha^2 < 4 \iff 1 < \alpha < 2,$$

当中用到 $\alpha > 0$。反过来说，对于这个范围内的任何 α，上述角度都存在，而按这个构造所得的位形中，三个力共点（concurrent），所以力矩平衡，而下文分解力所得的 T 和 R 均为正数。可能值的集合是 $1 < \alpha < 2$。

张力

沿铅垂方向分解力（resolve）。绳与铅垂线所成角的余弦是（铅垂上升高度）/（长度）$= 2L\cos\theta/(\alpha L)$，而 R 是水平的，所以

$$T\cdot\frac{2\cos\theta}{\alpha}=Mg \quad\Longrightarrow\quad T=\frac{Mg\alpha}{2\cos\theta}=\frac{Mg\alpha}{2}\left(\frac{3}{\alpha^2-1}\right)^{\frac{1}{2}},$$

这里代入了 $\cos\theta=\sqrt{(\alpha^2-1)/3}$（取正根，因为 $\cos\theta>0$）。这就是所要求的结果。

为完整起见，沿水平方向分解力可得 $R=T\cdot\dfrac{L\sin\theta}{\alpha L}=\frac{1}{2}Mg\tan\theta>0$，所以墙壁确实把杆向外推，正如墙壁应有的表现。

验算

- 当 $\alpha\to 2^-$ 时，$\cos\theta\to 1$：杆铅垂地悬在 A 之下，绳沿着杆对折回来，而 $T\to Mg$，因为这时绳承担了全部重量。当 $\alpha\to 1^+$ 时，$\theta\to 90^\circ$ 且 $T\to\infty$：一根接近水平的杆需要极大的张力才能支撑得住。
- 取 $\alpha=1.5$，我们得到 $\cos\theta\approx 0.6455$（$\theta\approx 49.8^\circ$）及 $T\approx 1.162\,Mg$。我们以数值方法验证过：对于在 $(1,2)$ 内随机选取的数千个 α，对 A 的力矩恰好只在一个角度为零，这个角度与公式吻合，张力亦然；而当 α 在 $(1,2)$ 之外时，并不存在平衡状态。

常见错误

- 把杆画成从 A 向上倾斜，就像梯子问题那样。共点的论证迫使 B 在 A 之下，而图画错了的话，方程便没有相容的解。
- 忘记光滑墙壁的反作用力是水平的，因而引入了不必要的摩擦力或铅垂分量。
- 把 $\alpha=1$ 或 $\alpha=2$ 包括在内：这两个值给出共线（退化）的位形，而题目已把这种情况排除。

STEP 之外

三力原理是静力学（statics）中一个一般事实的特殊情况：一个力系处于平衡，当且仅当它的合力以及它对任意一点的合力矩都等于零。像我们选取 P 那样巧妙地选点，往往是最快捷的途径。工程力学中对桁架、缆索和框架的分析，背后也是同一个想法。若想深入而全面地了解力与力矩，可参阅 Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）。

详细解答

STEP III 样本试卷，第 13 题

力学 • 滚动圆环

STEP III 样本试卷，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：

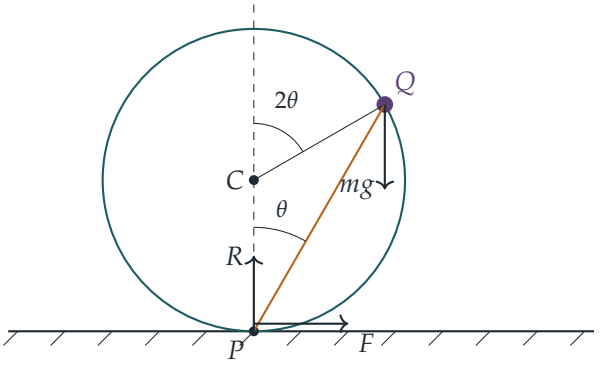
step.maths.org/questions/spec-s3-q13

关键思路

由于圆环是轻的，系统中唯一的动能就是质点的动能；又由于圆环作滚动，摩擦力不作功。因此能量守恒（energy conservation）可立即给出速率。滚动亦完全确定了几何关系：圆环转过 2θ 之后，其中心已移动了距离 $2a\theta$，由此可把 x 和 y 单用 θ 表示。最后，轻圆环不能承受任何净力，所以桌面的反作用力恰好就是质点的加速度所要求的力，而且它不能是负数。

解答

把圆环模型化为一个只滚不滑的轻刚性圆，把质点视为固定在环边上的点质量。设 P 为圆环与桌面的接触点，C 为圆心，Q 为质点。取圆环沿正 x 方向滚动，使 θ 由 0 开始增加。



这里 R 和 F 分别是桌面在 P 点作用于圆环的法向反作用力（normal reaction）和摩擦力（friction），而弦 PQ 与铅垂线的夹角为 θ（即圆心角 2θ 的一半）。

速率

R 和 F 都作用于 P 点，而由于圆环作滚动，该点在瞬间是静止的，所以两者都不作功。圆环是轻的，所以唯一的动能是质点的动能，唯一的势能变化也是质点的势能变化。质点由高度 $2a$ 出发，起初几乎静止；当半径 CQ 与向上的铅垂线成角 2θ 时，它的高度为 $a + a\cos 2\theta = 2a\cos^2\theta$。由能量守恒得

$$\tfrac{1}{2}mv^2 = mg\big(2a - 2a\cos^2\theta\big) = 2mga\sin^2\theta,$$

所以 $v^2 = 4ga\sin^2\theta$，而由于 $0 \le \theta < \tfrac{\pi}{2}$ 给出 $\sin\theta \ge 0$，

$$v = 2(ga)^{1/2}\sin\theta,$$

这正是所求。

位置与 $\dot\theta$

圆环已转过角 2θ，而只滚不滑意味着圆心向前移动的距离等于弧长 $a\cdot 2\theta$。相对于 C，质点在前方 $a\sin 2\theta$、上方 $a\cos 2\theta$ 处。所以

$$x = 2a\theta + a\sin 2\theta, \qquad y = a + a\cos 2\theta = 2a\cos^2\theta.$$

对时间求导，

$$\dot x = 2a\dot\theta\,(1 + \cos 2\theta) = 4a\dot\theta\cos^2\theta,$$
$$\dot y = -2a\dot\theta\sin 2\theta = -4a\dot\theta\sin\theta\cos\theta,$$

因此

$$v^2 = \dot x^2 + \dot y^2 = 16a^2\dot\theta^2\cos^2\theta\big(\cos^2\theta + \sin^2\theta\big) = 16a^2\dot\theta^2\cos^2\theta.$$

? **想一想** 有没有更快的方法看出 $v = 4a\dot\theta\cos\theta$?

有：P 是瞬时转动中心（instantaneous centre of rotation），而圆环以角速率 $2\dot\theta$ 转动。弦 PQ 在 C 所对的角为 $\pi - 2\theta$，所以 $PQ = 2a\cos\theta$，于是 $v = PQ\cdot 2\dot\theta = 4a\dot\theta\cos\theta$，与上面的计算吻合。

由于 $\dot\theta > 0$ 且 $\cos\theta > 0$，与由能量求得的速率比较，

$$4a\dot\theta\cos\theta = 2(ga)^{1/2}\sin\theta \quad\Longrightarrow\quad \dot\theta = \tfrac{1}{2}(g/a)^{1/2}\tan\theta,$$

这就是要求的第一个结果。（左边必须是变率 $\dot\theta$，而不是 θ：右边的单位是 s^{-1}。）

铅垂加速度

把 $\dot\theta$ 代入 $\dot y$,

$$\dot y = -4a\sin\theta\cos\theta\cdot\tfrac12(g/a)^{1/2}\frac{\sin\theta}{\cos\theta} = -2(ga)^{1/2}\sin^2\theta.$$

再求导一次，并再次使用 $\dot\theta$,

$$\ddot y = -4(ga)^{1/2}\sin\theta\cos\theta\,\dot\theta = -4(ga)^{1/2}\sin\theta\cos\theta\cdot\tfrac12(g/a)^{1/2}\frac{\sin\theta}{\cos\theta} = -2g\sin^2\theta,$$

这正是所求。同样地，$\dot x = 4a\cos^2\theta\,\dot\theta = (ga)^{1/2}\sin 2\theta$，而

$$\ddot x = 2(ga)^{1/2}\cos 2\theta\,\dot\theta = g\cos 2\theta\tan\theta.$$

甚么令滚动停止

圆环是轻的，所以作用于“圆环连质点”这个系统的合力等于 m 乘以质点的加速度。外力有向下的 mg、向上的 R 和水平方向的 F。沿铅垂方向分解,

$$R - mg = m\ddot y \quad\Longrightarrow\quad R = mg\big(1-2\sin^2\theta\big) = mg\cos 2\theta,$$

而沿水平方向则有 $F = m\ddot x = mg\cos 2\theta\tan\theta$，所以当 $R>0$ 时 $F/R = \tan\theta$。

> **想一想** 当 θ 超过 $\pi/4$ 时，会出甚么问题?

桌面只能推而不能拉，所以 $R\ge 0$。但 $R = mg\cos 2\theta$ 在 $\theta=\pi/4$ 时降至零，此时质点与圆心同高, 而要继续滚动下去就需要 $R<0$。换一个说法, 继续滚动需要 $\ddot y = -2g\sin^2\theta < -g$：质点向下的加速度要比自由下落还要大，单靠重力是做不到的。所以在 $\theta=\pi/4$ 时接触力消失，圆环离开桌面。此后质点作抛体运动（projectile motion），初速度的分量为向前 $(ga)^{1/2}$ 和向下 $(ga)^{1/2}$，而轻圆环只是被它带着走，其最低点升离桌面。

摩擦力可能更早介入。滚动需要 $|F|\le\mu R$，即 $\tan\theta\le\mu$，而当 $\theta\to\pi/4$ 时 $\tan\theta$ 递增至 1。所以若摩擦系数（coefficient of friction）满足 $\mu<1$，圆环在 $\tan\theta=\mu$ 时便开始滑动，未及到达 $\theta=\pi/4$。只有当 $\mu\ge 1$ 时，它才会一直滚到 $\theta=\pi/4$，然后跳起。

验算

- 当 $\theta=0$ 时，$\dot\theta=0$：最高位置是一个平衡位置，这就是为甚么圆环必须被轻微移离该位置。
- 当 $\theta=\pi/4$ 时：$v=(2ga)^{1/2}$，这是下落距离 a 所得的速率，而质点确实由 $2a$ 降到了 a。
- 电脑独立地对 $\phi=2\theta$ 的运动方程作数值积分（由拉格朗日量（Lagrangian）$ma^2(1+\cos\phi)\dot\phi^2 - mga(1+\cos\phi)$ 出发），并把有限差分求得的速度和加速度与公式比较。速

率、$\dot{\theta}$、$\ddot{y}$、$R = mg\cos 2\theta$ 和 $F/R = \tan\theta$ 全部吻合，误差在 10^{-6} 以内；例如在 $\theta = \pi/4$ 时，它给出 $R = 0$ 和 $F/R = 1$。

常见错误

- 以为圆心移动的距离是 $a\theta$ 而不是 $2a\theta$：圆环转过的角是 2θ，不是 θ。
- 把摩擦力计入能量方程。接触点是静止的，所以圆环滚动时摩擦力不作功。
- 只说质点“掉下来”。质点是固定在环上的；离开桌面的是圆环，因为反作用力否则就要变成负数。

STEP 之外

滚动物体的接触点就是它的瞬时转动中心，而只滚不滑是一个约束（constraint），它消去一个自由度（degree of freedom）而不作功。这类约束可由拉格朗日力学（Lagrangian mechanics）有系统地处理：只需用单一坐标（例如 θ）写出能量，运动方程便自动得出；上面的验算正是这样做的。约束力何时消失的问题（珠子离开球面、汽车离开路拱）是大学一年级动力学的典型题目。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对滚动运动和约束力有详细的论述。

详细解答

STEP III 样本试卷，第 14 题

力学 • 转动碰撞

STEP III 样本试卷，第 14 题

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

step.maths.org/questions/spec-s3-q14

关键思路

枢轴（pivot）对棒施加一个未知的冲量（impulse），所以线动量（linear momentum）并不守恒；但这个冲量对枢轴没有力矩（moment），因此对中点的角动量（angular momentum）守恒。再在接触点应用牛顿恢复定律（Newton's law of restitution）（由于碰撞是完全弹性（perfectly elastic）的，恢复系数为 1），一切便确定下来。至于第二部分，先找出质点接下来怎样运动，然后追踪两个物体：棒的另一半会转过来，到达质点的运动直线。

解答

模型。质点视为点质量；桌面光滑，所以两个物体都不受摩擦力；棒是均匀的，并在其中点 O 处光滑地枢接。棒对 O 的转动惯量（moment of inertia）为

$$I = \tfrac{1}{12}m(4a)^2 = \tfrac{4}{3}ma^2.$$

碰撞期间，冲量沿质点的运动直线作用（棒是光滑的，而质点垂直撞向棒），所以质点继续沿同一直线运动。设碰撞后质点的速度为 v，以其原来的运动方向为正。

角速率

对 O 的角动量。系统（质点和棒）所受的唯一外来冲量来自 O 处的枢轴，而它对 O 的力矩为零。碰撞前，质点对 O 的角动量为 mua，而棒静止，所以

$$mua = mva + \tfrac{4}{3}ma^2\omega \qquad \Longrightarrow \qquad u = v + \tfrac{4}{3}a\omega. \tag{1}$$

恢复定律。棒上被撞击的一点与 O 相距 a，所以刚碰撞后，它以速率 $a\omega$ 沿质点原来的方

向运动。恢复系数（coefficient of restitution）为 1，故分离速率等于接近速率：

$$a\omega - v = u. \tag{2}$$

想一想 牛顿恢复定律可以用于转动中的物体吗?

可以：它适用于两个接触点沿公法线方向的速度。如果你喜欢，也可以改用能量："完全弹性"表示动能（kinetic energy）守恒，$\frac{1}{2}mu^2 = \frac{1}{2}mv^2 + \frac{1}{2}\cdot\frac{4}{3}ma^2\omega^2$。把它写成 $m(u-v)(u+v) = \frac{4}{3}ma^2\omega^2$，并利用 (1)，即 $u - v = \frac{4}{3}a\omega$，便得 $u + v = a\omega$（两边除以 $\omega \neq 0$），这正是 (2)。

把 (1) 和 (2) 相加，消去 v：

$$2u = \tfrac{7}{3}a\omega \qquad \Longrightarrow \qquad a\omega = \frac{6u}{7},$$

这就是所要求的结果。然后由 (2)，

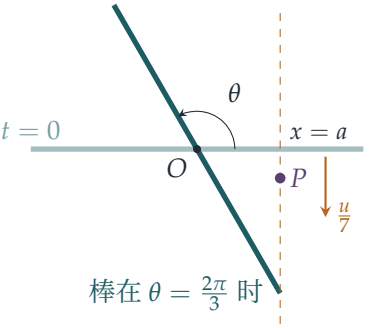
$$v = a\omega - u = -\frac{u}{7},$$

所以质点沿原来的直线以速率 $u/7$ 反弹。

第二次碰撞

在桌面上取坐标轴，以 O 为原点，x 轴沿棒的初始位置，而质点原本沿直线 $x = a$ 向正 y 方向运动，于是它在时间 $t = 0$ 时在 $(a, 0)$ 撞击棒。那么当 $t > 0$ 时：

- 质点位于 $\left(a,\ -\frac{u}{7}t\right)$，向负 y 方向匀速运动；
- 被撞击的一点向正 y 方向运动，所以棒逆时针转动，在时间 t 时它与 x 轴成角 $\theta = \omega t$。在两者再次碰撞之前，没有其他力作用于任何一个物体，所以两者的运动都是匀速的。



棒的半长为 $2a$，所以只有当棒所在直线上落在直线 $x = a$ 的那一点（它与 O 相距 $a/|\cos\theta|$）距离 O 不超过 $2a$ 时，棒才会到达该直线；也就是说，当 $|\cos\theta| \geq \frac{1}{2}$ 时。在这种情况下，棒在高度 $y = a\tan\theta$ 处与该直线相交。

- 当 $0 < \theta \le \frac{\pi}{3}$ 时，棒与该直线相交于 $y = a\tan\theta > 0$，在 x 轴上方，而质点在其下方：没有接触。
- 当 $\frac{\pi}{3} < \theta < \frac{2\pi}{3}$ 时，棒根本到不了该直线。
- 当 $\frac{2\pi}{3} \le \theta < \pi$ 时，棒的另一半与该直线相交于 $y = a\tan\theta < 0$，与质点在同一侧。

现在比较位置。由于 $t = \theta/\omega = \frac{7a\theta}{6u}$，在角度为 θ 时质点的 y 坐标为 $-\frac{u}{7}t = -\frac{a\theta}{6}$。考虑

$$f(\theta) = (\text{棒的}y\text{ 坐标，取于直线}x = a) - (\text{质点的}y) = a\tan\theta + \frac{a\theta}{6}, \qquad \frac{2\pi}{3} \le \theta < \pi.$$

- 在 $\theta = \frac{2\pi}{3}$ 时（棒的末端到达该直线）：$f = a\left(-\sqrt{3} + \frac{\pi}{9}\right) < 0$，因为 $\frac{\pi}{9} < 1 < \sqrt{3}$。质点位于 $y = -\frac{\pi}{9}a \approx -0.35a$，介乎 x 轴与位于 $y = -\sqrt{3}a$ 的棒末端之间：棒的末端已经转到质点之外。
- 当 $\theta \to \pi^-$ 时：$\tan\theta \to 0$，所以 $f \to \frac{\pi}{6}a > 0$。棒正向上扫回 x 轴，而质点大约位于 $-0.52a$。

f 在 $[\frac{2\pi}{3}, \pi)$ 上连续，所以根据介值定理（intermediate value theorem），在两者之间存在某个 θ 使得 $f(\theta) = 0$：在那一刻棒正处于质点的位置，两者再次碰撞。（由上面头两点可知，之前不可能有接触。）证明至此完成。

以数值计算，碰撞发生在 $\tan\theta = -\theta/6$ 之处，即 $\theta \approx 2.716$（约 155.6°），时间为 $t \approx 3.17\,a/u$，此时质点位于 $y \approx -0.45a$，沿棒量度距离 O 约 $1.10a$。

验算

- 能量：以 $a\omega = \frac{6}{7}u$ 和 $v = -\frac{1}{7}u$ 代入，最终动能为 $\frac{1}{2}m\left(\frac{1}{49} + \frac{4}{3}\cdot\frac{36}{49}\right)u^2 = \frac{1}{2}m\cdot\frac{49}{49}u^2$，恰好等于初始能量。
- 对一个转动惯量为 I 的物体，在距离 d 处撞击，重复以上论证可得 $d\omega = 2u/(1 + I/md^2)$。当 $I \to 0$ 时，被撞击的一点以 $2u$ 离开，正如与一个极轻的物体作弹性碰撞；当 $I \to \infty$ 时，棒几乎不转动，质点以速率 u 反弹。这里 $I/ma^2 = \frac{4}{3}$，得出 $\frac{6}{7}u$。
- 我们以精确分数验证了 $a\omega = \frac{6}{7}u$、$v = -\frac{1}{7}u$ 同时令角动量和能量守恒，然后把两个运动按时间逐步推进：碰撞后的首次接触发生在 $\theta \approx 2.7165$，与用二分法（bisection）解 $\tan\theta + \theta/6 = 0$ 所得的结果吻合。

常见错误

- 以为线动量守恒。枢轴会提供冲量，所以只有对枢轴的角动量守恒。
- 使用 $I = \frac{1}{3}m\ell^2$（这是对端点的值），或者忘记棒的长度是 $4a$ 而不是 $2a$。

- 在棒被撞击的那一半上寻找第二次碰撞。那一半会远离质点的运动直线；与质点相遇的是转过来的另一半。

STEP 之外

把对某固定点的角动量与能量条件或恢复条件配合使用，是处理刚体（rigid body）冲击运动的标准方法，并由此引出撞击中心（centre of percussion）的概念：即撞击一个枢接物体时，枢轴处不会产生冲量的那一点。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）仔细讲解了刚体碰撞和角动量，并收录了许多正是这一类型的习题。

详细解答

STEP III 样本试卷，第 15 题

概率与统计 · 随机网络的连通性

STEP III 样本试卷，第 15 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/spec-s3-q15

关键思路

先把几何抛开：唯一要紧的，是六条路之中哪几条仍然开通。任何一条村都能从其余各村到达，当且仅当开通的道路在四条村之间构成一个连通（connected）的网络。因此，我们把所有可能的开通道路集合按所含道路的数目分类，数出每种大小之中有多少个是连通的，再各自乘上相应的概率。由于只有四条村，这个计数过程既短又干净。

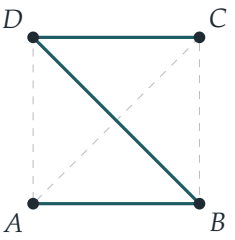
解答

设 $q = 1 - p$ 为某一条路开通的概率。设 K 为下雪之后每条村都能从其余各村到达这一事件。

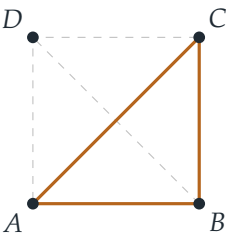
由于六条路互相独立（independent），任何一个指定的 k 条开通道路的集合（其余 $6 - k$ 条封闭）出现的概率都是 $q^k p^{6-k}$。所以

$$\mathrm{P}(K) = \sum_{k=0}^{6} N_k \, q^k p^{6-k},$$

其中 N_k 是能把四条村全部连通的 k 条道路集合的数目。



三条路：连通各村的树



三条路：三角形，D 被隔绝

计数连通的道路集合

最多两条路开通。每条开通的路连接两条村，所以从某一条村出发，每条路最多只能带来一条新的村。要到达其余三条村，至少需要三条路。因此 $N_0 = N_1 = N_2 = 0$。

三条路开通。三条路的集合共有 $\binom{6}{3} = 20$ 个。假设某个这样的集合未能连通各村。若有某条村被隔绝，则三条路全部落在其余三条村之间。但连接那三条村的路只有三条，所以这个集合恰好是一个三角形。若没有任何村被隔绝，则四条村分成两对，而最多只有两条路（每对之内各一条）可以开通，矛盾。所以失败的情况只有 4 个三角形，每个对应一条被隔绝的村，于是

$$N_3 = 20 - 4 = 16.$$

? **想一想** 为甚么有四条路开通时，网络绝不可能不连通？

四条或以上的路开通。沿用上面两种失败的方式。若有一条村被隔绝，所有开通的路都必须落在其余三条村之间，而那里只有三条路。若分成两对，则最多只容许两条路开通。有四条或以上的路开通时，两种情况都不可能发生，所以每个这样的集合都是连通的：

$$N_4 = \binom{6}{4} = 15, \qquad N_5 = \binom{6}{5} = 6, \qquad N_6 = 1.$$

所求概率

把各个数目合起来，

$$\mathrm{P}(K) = q^6 + 6q^5p + 15q^4p^2 + 16q^3p^3, \qquad q = 1 - p.$$

按 p 的幂展开，得到等价的形式

$$\mathrm{P}(K) = 1 - 4p^3 - 3p^4 + 12p^5 - 6p^6.$$

$p = \frac{1}{2}$ 的情况

此时 $p = q = \frac{1}{2}$，所以道路开通与封闭的 $2^6 = 64$ 种组合全部等可能，每种的概率为 $\frac{1}{64}$。因此

$$\mathrm{P}(K) = \frac{1 + 6 + 15 + 16}{64} = \frac{38}{64} = \frac{19}{32},$$

正是所求。

验算

- 极端值。当 $p = 0$ 时，每条路都开通，公式给出 1。当 $p = 1$ 时，每条路都封闭，公式给出 $1 - 4 - 3 + 12 - 6 = 0$。

- 补事件（*complement*）。网络不连通，当且仅当有某条村被隔绝，或者四条村分成两对而两对之间没有任何路相通。一条村被隔绝的概率是 p^3。对四条村运用容斥原理（inclusion–exclusion）（两条指定的村同时被隔绝需要五条路封闭，而三条村被隔绝即表示四条村全部被隔绝），得出 $4p^3 - 6p^5 + 3p^6$。分成两对的方法有 3 种，每种在两条对内的路都开通的情况下，再加上 p^4q^2。以一减去总和，结果与上面的公式吻合：当 $p = \frac{1}{2}$ 时，不连通的概率是 $\frac{32-12+3+3}{64} = \frac{13}{32}$。
- 穷举法（*brute force*）。用电脑列出全部 64 个道路集合，证实了 16, 15, 6, 1 这几个数目。以 $p = 0.3$ 进行模拟，得出 0.893，而准确值是 0.8925。

常见错误

- 把全部 20 个三条路的集合都当作连通，忘记了那四个三角形。
- 以为任何四条路的集合仍有可能令某条村孤立。一条村只有在它的三条路全部封闭时才会被隔绝。
- 写 $q^k p^{6-k}$ 时，把 p（封闭）和 $1 - p$（开通）混淆。

STEP 之外

这是随机图（random graph）的一个微型例子：网络的每条边各自独立地保留下来，然后我们问所得的结果是否连通。这个 p 的多项式称为该网络的可靠度多项式（reliability polynomial），而 16 这个数则是凯莱公式（Cayley's formula）的一个实例：连接 n 个有标号的点的树（tree）共有 n^{n-2} 棵。对于大型网络，艾狄胥（Erdős）和伦伊（Rényi）发现，当道路开通的概率越过某个临界值（threshold）时，连通性会异常突然地出现。若想进一步了解由独立事件构成的随机结构，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率论入门教材）。

详细解答

STEP III 样本试卷，第 16 题

概率与统计 • 负二项分布

STEP III 样本试卷，第 16 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/spec-s3-q16

关键思路

等到 r 个儿子出生，等同于先等一个儿子出生，然后重新开始等下一个，如此重复 r 次。因此 C 可以拆成 r 个分布相同的等候时间之和，余下的工作交给期望值的线性（linearity of expectation）便可。至于第 (ii) 部分，窍门是想像总统停止生育之后，婴儿仍然继续出生：这样一来，他在某个时刻之前是否已经停下，就变成一个关于固定数目的孩子的问题，再利用男孩与女孩之间的对称性（symmetry）便可解决。

解答

以下一律假设每次出生的婴儿是男孩的概率为 $\frac{1}{2}$，并且与其他各次出生互相独立（independent）。

第 (i) 部分

对于 $i = 1, 2, \ldots, r$，设 X_i 为第 $(i-1)$ 个儿子出生之后，直至第 i 个儿子出生为止（包括这个儿子在内）所生的孩子数目。（当 $i = 1$ 时，由最初开始数起。）于是

$$C = X_1 + X_2 + \cdots + X_r.$$

每个 X_i 所数的，都是在一轮全新的、独立而公平的出生之中，直至第一个男孩出现为止的出生次数，所以每个 X_i 的分布（distribution）都与 X 相同，其中

$$\mathrm{P}(X = k) = \left(\tfrac{1}{2}\right)^{k-1} \cdot \tfrac{1}{2} = \left(\tfrac{1}{2}\right)^{k}, \qquad k = 1, 2, 3, \ldots,$$

因为 $X = k$ 表示先有 $k - 1$ 个女孩，然后才有一个男孩。

要求出 $\mathrm{E}(X)$，可把几何级数（geometric series）$\sum_{k\geq 0} x^k = \dfrac{1}{1-x}$（适用于 $|x|<1$）求导，得到 $\sum_{k\geq 1} kx^{k-1} = \dfrac{1}{(1-x)^2}$。乘以 x 并代入 $x=\frac{1}{2}$，得

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} k\left(\tfrac{1}{2}\right)^k = \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} = 2.$$

这个结果合乎常理：如果所有婴儿之中有一半是男孩，那么平均要等两次出生才会有一个男孩。

根据期望值的线性（无论各个 X_i 是否独立，这个性质都成立，不过这里它们确实是独立的），

$$\mathrm{E}(C) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_r) = 2r.$$

第 (ii) 部分

想像有一个无穷无尽的婴儿序列，各个婴儿的性别早已预先决定，彼此独立而且公平；总统按着这个序列生孩子，直至第 r 个儿子出生为止。这样做完全不会改变 C，却让我们即使在他提早停下的情况下，仍然可以谈论“首 $2r-1$ 个婴儿”。

? **想一想** 怎样只用序列中的首 $2r-1$ 个婴儿，来描述事件 $C<2r$?

由于 C 是整数，$C<2r$ 即是 $C\leq 2r-1$，也就是说，第 r 个儿子在第 $2r-1$ 个位置或之前出现。这种情况发生，当且仅当至少有 r 个男孩出现在首 $2r-1$ 个婴儿之中。所以

$$\mathrm{P}(C<2r) = \mathrm{P}(\text{至少有}r\text{ 个男孩出现在首}2r-1\text{ 个婴儿之中}).$$

现在设 B 和 G 分别为这 $2r-1$ 个婴儿之中男孩和女孩的数目，于是 $B+G=2r-1$。因为 $2r-1$ 是奇数，B 与 G 不可能相等，而且两者之中恰好有一个不小于 r：

- 若 $B\geq r$，则 $G=2r-1-B\leq r-1$；
- 若 $B\leq r-1$，则 $G\geq r$。

所以事件 $\{B\geq r\}$ 与 $\{G\geq r\}$ 互斥（mutually exclusive），而且两者合起来涵盖了所有可能情况，由此得到

$$\mathrm{P}(B\geq r) + \mathrm{P}(G\geq r) = 1.$$

但每次出生时，男孩和女孩出现的机会均等，所以把“男孩”和“女孩”这两个标签对调，并不会改变任何概率，因此 $\mathrm{P}(B\geq r) = \mathrm{P}(G\geq r)$。故此两者各为 $\frac{1}{2}$，而

$$\mathrm{P}(C<2r) = \tfrac{1}{2},$$

这正是所要证明的。

由此可见，家庭人数的期望值是 $2r$，但家庭人数少于 $2r$ 的情况却恰好占一半。家庭人数的分布是偏斜（skewed）的：略小于 $2r$ 的数值十分常见，而偶尔出现的一长串女孩则把平均值拉高。

验算

- 取 $r = 1$：$\mathrm{E}(C) = 2$，而 $C < 2$ 表示第一个婴儿是男孩，其概率为 $\frac{1}{2}$。
- 取 $r = 2$：$C < 4$ 表示 $C = 2$ 或 3，概率分别为 $\frac{1}{4}$ 和 $2 \cdot \frac{1}{8} = \frac{1}{4}$，总和为 $\frac{1}{2}$。
- 对 $r = 1, \ldots, 8$ 把 $c\binom{c-1}{r-1}2^{-c}$ 精确地求和，每次都得到 $\mathrm{E}(C) = 2r$ 和 $\mathrm{P}(C < 2r) = \frac{1}{2}$；以 $r = 3$ 和 200 000 个家庭进行模拟，得出平均值为 6.00，以及 $\mathrm{P}(C < 6) \approx 0.499$。

常见错误

- 以“孩子之中有一半是男孩，所以 $C = 2r$”作为论证。这句话就平均而言是对的，但 C 本身是随机的；第 (i) 部分需要把它分解成各段等候时间，而不能只靠猜测。
- 在第 (ii) 部分只考虑总统实际所生的孩子。当 $C < 2r - 1$ 时，根本就没有 $2r - 1$ 个孩子；正是把出生序列延伸至停止点之后，计数的论证才得以成立。
- 忘记了 $2r - 1$ 是奇数。正是这一点排除了男孩与女孩数目相同的情况，并使两个事件互补（complementary）。

STEP 之外

C 的分布就是负二项（negative binomial）分布：$\mathrm{P}(C = c) = \binom{c-1}{r-1}2^{-c}$，其中 $c \geq r$。把一个随机变量（random variable）写成若干个较简单部分之和，再运用期望值的线性，是大学一年级概率论中最有力的工具之一；而第 (ii) 部分把“等候时间”与“固定试验次数中的计数”联系起来的窍门，正是连结负二项分布与二项（binomial）分布的那种对偶性（duality）（在连续时间的情况下，它连结的是伽玛（gamma）分布与泊松（Poisson）分布）。Blitzstein 与 Hwang 合著的 *Introduction to Probability* 以大量例子讲解了这两个想法。

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

样本试卷详细解答. 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。

