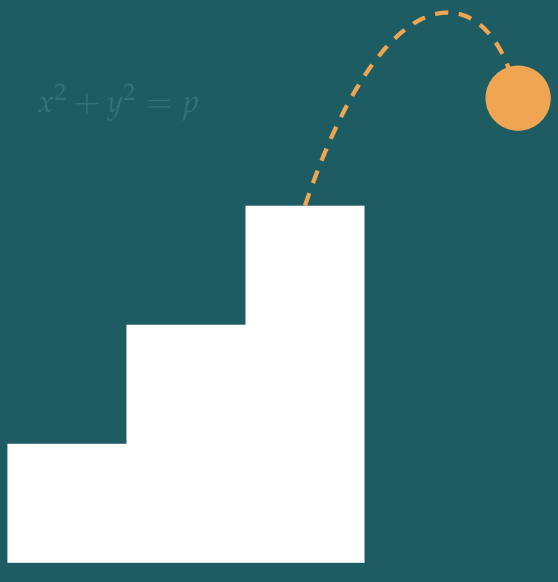
STEP 备试 · 大学数学

Beyond STEP

STEP I, II 及 III

1992 年详细解答



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

STEP I, II 及 III

1992 年详细解答

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

1992 年详细解答

1992 年 STEP 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 模 7 的日历算术 纯数学 . . . 3

第 2 题 幻方与计数 纯数学 . . . 6

第 3 题 奇函数与偶函数 纯数学 . . . 10

第 4 题 向根图上的轨迹 纯数学 . . . 13

第 5 题 多项式的重因式 纯数学 . . . 17

第 6 题 小角近似 纯数学 . . . 21

第 7 题 整值多项式 纯数学 . . . 25

第 8 题 微积分基本定理 纯数学 . . . 28

第 9 题 微分方程 纯数学 . . . 31

第 10 题 反冲与动量 力学 . . . 35

第 11 题 弹性绳的平衡 力学 . . . 38

第 12 题 涉及摩擦力的静力学 力学 . . . 42

第 13 题 逸速率与径向下落 力学 . . . 46

第 14 题 统计论断的诠释 概率与统计 . . . 50

第 15 题 期望等候时间 概率与统计 . . . 54

第 16 题 网络可靠度 概率与统计 . . . 58

STEP II

第 1 题 数列的极限 纯数学 . . . 62

第 2 题 克莱罗微分方程 纯数学 . . . 65

第 3 题 内摆线与面积 纯数学 . . . 69

第 4 题 微分法的公理 纯数学 . . . 73

第 5 题 矩阵群 纯数学 . . . 76

第 6 题 切线与根的计数 纯数学 . . . 80

第 7 题 三次方程的重根 纯数学 84
第 8 题 积分技巧 纯数学 87
第 9 题 向量三重积 纯数学 90
第 10 题 复根图上的轨迹 纯数学 94
第 11 题 抛体运动 力学 98
第 12 题 刚杆与弹性绳 力学 102
第 13 题 摩擦力与连接质点 力学 106
第 14 题 连接质点与滑轮 力学 110
第 15 题 随机漫步 概率与统计 114
第 16 题 条件概率 概率与统计 118

STEP III

第 1 题 麦克劳林级数 纯数学 122
第 2 题 矩阵的幂与数学归纳法 纯数学 126
第 3 题 曲线与多项式的根 纯数学 129
第 4 题 正交轨线 纯数学 133
第 5 题 三维向量几何 纯数学 137
第 6 题 三角函数归约公式 纯数学 141
第 7 题 向量几何 纯数学 144
第 8 题 以方程的根求三角函数之和 纯数学 148
第 9 题 椭圆的极坐标方程 纯数学 151
第 10 题 弧长与面积 纯数学 155
第 11 题 碰撞与恢复系数 力学 159
第 12 题 动量守恒 力学 163
第 13 题 绕固定轴的转动 力学 167
第 14 题 解绕绳的动力学 力学 171
第 15 题 几何概率 概率与统计 175
第 16 题 蓄水池疏化 概率与统计 179

详细解答

STEP I 1992，第 1 题

纯数学 • 模 7 的日历算术

> **STEP I 1992，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/92-s1-q1

关键思路

星期每 7 日循环一次，所以我们只需要知道两个日期之间相隔多少日，再取模（modulo）7 的余数。令计算变得轻松的窍门是 $365 = 52 \times 7 + 1$：平年会令星期向后移一日，闰年（leap year）则移两日。因此我们先数整年和闰日，最后才处理余下的零碎月份。所谓“仔细”，是指要准确判断哪些 2 月 29 日真正落在这段期间之内，并留意那个可能令我们出错的世纪年。

解答

我们把 1905 年 4 月 3 日至 1992 年 6 月 26 日这段期间分成两段：

$$1905 \text{ 年 } 4 \text{ 月 } 3 \text{ 日} \longrightarrow 1992 \text{ 年 } 4 \text{ 月 } 3 \text{ 日} \longrightarrow 1992 \text{ 年 } 6 \text{ 月 } 26 \text{ 日}.$$

整年：1905 年 4 月 3 日至 1992 年 4 月 3 日

这段时间共有 $1992 - 1905 = 87$ 个整年。每个这样的年份（由 4 月 3 日至下一个 4 月 3 日）有 365 日，若当中包含一个 2 月 29 日，则多加一日。

哪些 2 月 29 日落在 1905 年 4 月 3 日与 1992 年 4 月 3 日之间？这个范围内的闰年就是可被 4 整除的年份，即

$$1908,\ 1912,\ \ldots,\ 1988,\ 1992.$$

1904 年的 2 月 29 日在 1905 年 4 月 3 日之前，所以不计算在内；1992 年的 2 月 29 日在 1992 年 4 月 3 日之前，所以要计算在内。于是共有

$$\frac{1992 - 1908}{4} + 1 = 21 + 1 = 22$$

个闰日。

> **想一想** 这个范围内每个可被 4 整除的年份，真的都是闰年吗?

在格里历（Gregorian calendar）中，世纪年只有在可被 400 整除时才是闰年，所以 1900 年不是闰年（而 2000 年是）。不过这两个年份都不在我们的期间之内，所以简单的“可被 4 整除”法则（即口诀中的那一条）在这里完全正确。假如题目追溯到 1900 年之前，我们便要剔除 1900 年 2 月 29 日。

因此，由 1905 年 4 月 3 日至 1992 年 4 月 3 日的日数是

$$87 \times 365 + 22 = 31\,755 + 22 = 31\,777.$$

以模 7 计算会更快：由于 $365 \equiv 1 \pmod 7$，这个数等于

$$87 + 22 = 109 = 15 \times 7 + 4 \equiv 4 \pmod 7.$$

零碎月份：1992 年 4 月 3 日至 1992 年 6 月 26 日

根据口诀，四月有 30 日，五月有 31 日，所以

$$\begin{aligned} &4\text{ 月 }3\text{ 日} \to 5\text{ 月 }3\text{ 日}: 30\text{ 日}, \\ &5\text{ 月 }3\text{ 日} \to 6\text{ 月 }3\text{ 日}: 31\text{ 日}, \\ &6\text{ 月 }3\text{ 日} \to 6\text{ 月 }26\text{ 日}: 23\text{ 日}, \end{aligned}$$

合共 $30 + 31 + 23 = 84 = 12 \times 7$ 日，即 $\equiv 0 \pmod 7$。可喜的是，1992 年 4 月 3 日与 1992 年 6 月 26 日恰好是同一个星期几。

综合结果

总相隔日数为

$$31\,777 + 84 = 31\,861 = 4551 \times 7 + 4\text{ 日},$$

所以 1905 年 4 月 3 日在星期的次序上比 1992 年 6 月 26 日早 4 日。由星期五倒数四日：星期四、星期三、星期二、星期一。因此

$$\text{1905 年 4 月 3 日是星期一。}$$

验算

- 用电脑逐月直接计算两个日期之间的日数，得出 31 861 日，即 $4 \pmod 7$；电脑日历亦显示 1905 年 4 月 3 日是星期一（而 1992 年 6 月 26 日是星期五）。

- 检查方向是否合理：由星期一向前数 4 日得到星期五，正如所料。
- 用另一个方法核对闰日数目：87 年大约包含 $87/4 \approx 21.75$ 个闰日，所以 22 大致正确。

常见错误

- 把 1904 年 2 月 29 日（在起始日期之前）计算在内，或漏掉 1992 年 2 月 29 日（在结束日期之前）：两者都会令答案相差一日。
- 应该向前数时倒数，或应该倒数时向前数：余数 4 表示在星期的循环中，1905 年那一天是在星期五之前四日，而不是之后。
- 把这个方法推广到更早的日期时，将 1900 年当作闰年。

STEP 之外

这就是模 7 的算术：关键事实 $365 \equiv 1 \pmod 7$ 把一个庞大的数目化为一个细小的数目。同余（congruence），以及它在历法谜题中的应用，例如用来求任何日期是星期几的蔡勒公式（Zeller's congruence），都是数论（number theory）入门课程的标准内容。Davenport 的 *The Higher Arithmetic*（数论入门名著）是一本介绍同余及其他众多课题的佳作。

详细解答

STEP I 1992，第 2 题

纯数学 · 幻方与计数

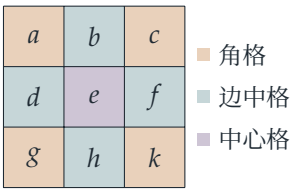
> **STEP I 1992，第 2 题**
>
> 继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：
>
> step.maths.org/questions/92-s1-q2

关键思路

一切都来自巧妙地把各条线加起来。把三个横行全部相加，每个数恰好用了一次，由此得出 n。把通过中心的四条线相加，每个数用了一次，而中心则用了四次，这就确定了 e。此后，题目就变成一个规模不大、但须小心处理的分情况讨论（case analysis）：一旦知道 e，每条通过中心的线都把两个数配成一对，其和固定，于是留给大数的空间就非常有限。最后，正方形的对称性（symmetry）（即把它旋转）便替我们完成计数。

解答

先为三类方格命名会比较方便：四个角格 a,c,g,k，四个边中格 b,d,f,h，以及中心格 e。



第 (i) 部分

三个横行合起来，九个数每个恰好出现一次。因此，把三个横行的和相加，

$$3n = (a+b+c)+(d+e+f)+(g+h+k) = 1+2+\cdots+9 = 45,$$

故 $n = 15$。

第 (ii) 部分

通过中心的线共有四条：中间横行、中间直行和两条对角线。把它们相加，

$$(d+e+f)+(b+e+h)+(a+e+k)+(c+e+g)=4n=60.$$

在左边，除 e 以外的每个数恰好出现一次，而 e 出现四次，所以左边等于 $45+3e$。因此 $45+3e=60$，得 $e=5$。

由此得到一个有用的推论：由于每条通过中心的线之和都是 15，而 $e=5$，所以隔着中心相对的两个方格，其和必为 10：

$$a+k=c+g=b+h=d+f=10.$$

第 (iii) 部分

因为 $e=5$，所以 9 不在中心格。于是我们只须证明 9 不可能在角格。

用反证法（proof by contradiction）：假设 9 在某个角格。对每个角格，论证都是一样的，所以取 $a=9$。那么对角的角格是 $k=10-9=1$。最上横行和最左直行都通过 a，所以

$$b+c=15-9=6 \qquad 及 \qquad d+g=15-9=6.$$

b,c,d,g 这四个数互不相同，并且取自用去 $1,5,9$ 之后余下的数，即 $\{2,3,4,6,7,8\}$。要把 6 写成这个集合中两个不同的数之和，唯一的方法是 $2+4$（因为 $1+5$ 用了已经填入的数，而 $3+3$ 则重复）。所以 $\{b,c\}=\{2,4\}$ 且 $\{d,g\}=\{2,4\}$，这就需要 2 和 4 各用两次，与各数互不相同矛盾。

> ? **想一想** 为甚么只处理角格 a 就已足够?

对任何一个角格，同样的三项事实都成立：对角的角格必定是 1，而沿正方形的边通过该角格的两条线，各自余下的两个数之和都须为 6。所以会出现同样的矛盾。（换个说法：把一个幻方（magic square）旋转四分之一圈，会得到另一个幻方，而旋转可把 a 带到其余每一个角格。）

所以 9 既不在中心格，也不在角格，它必定是 b,d,f,h 其中之一。

第 (iv) 部分

设 $b=9$。那么 $h=10-9=1$，而最上横行给出 $a+c=6$。一如 (iii)，未用的数是 $\{2,3,4,6,7,8\}$，其中和为 6 的一对相异数只有 $\{2,4\}$。所以或者 $a=2,\ c=4$，或者 $a=4,\ c=2$。

一旦 a 和 c 确定，其余各数都由相对的数对和左右两个直行完全决定：

- $a = 2$，$c = 4$：则 $k = 10 - a = 8$，$g = 10 - c = 6$，$d = 15 - a - g = 7$ 及 $f = 15 - c - k = 3$。
- $a = 4$，$c = 2$：则 $k = 6$，$g = 8$，$d = 15 - 4 - 8 = 3$ 及 $f = 15 - 2 - 6 = 7$。

推出这些必然的值，只说明它们是仅有的候选；我们仍须验证它们确实可行。在两种情况中，所填的数都是 $1, \ldots, 9$ 各一次，而且每个横行、直行和对角线之和都是 15：

2	9	4
7	5	3
6	1	8

4	9	2
3	5	7
8	1	6

（例如在第一个之中：横行 $15, 15, 15$；直行 $2+7+6$、$9+5+1$、$4+3+8$；对角线 $2+5+8$、$4+5+6$。）所以 $b = 9$ 的幻方恰好有两个，而第二个是第一个以中间直行为轴的镜像。

第 (v) 部分

把一个幻方顺时针旋转四分之一圈，会得到另一个幻方：它的横行是原来的直行，它的直行是原来的横行（次序倒转），而两条对角线互换。四分之一圈的旋转是可逆的（逆时针转回去），所以它给出幻方之间的一一对应（one-to-one correspondence）。

顺时针旋转四分之一圈，会把方格 b 移到方格 f。所以 $f = 9$ 的幻方，正好就是 $b = 9$ 的幻方旋转四分之一圈所得的，而且恰好有 2 个。再转一次，$h = 9$ 的恰好有 2 个，$d = 9$ 的也恰好有 2 个。

由 (iii)，每个幻方的 9 都恰好位于 b, f, h, d 其中一个位置，所以这四组互不相交，并且涵盖了所有幻方。因此共有

$$4 \times 2 = 8$$

个不同的 3×3 幻方。它们就是 (iv) 中的两个幻方，连同它们旋转四分之一圈、半圈和四分之三圈所得的幻方；换个说法，它们是同一个幻方的八个旋转和反射（reflection）。

验算

用电脑对 $1, \ldots, 9$ 的全部 $9! = 362\,880$ 种排列作一次简短的搜索，恰好找到 8 个幻方。它们的各线之和全是 15，中心全是 5，9 全都在边中格，而 $b = 9$ 的两个正是 (iv) 中找到的幻方。你也可以亲手验证：在每个幻方中，1 都在 9 的对面，而偶数 $2, 4, 6, 8$ 占满四个角格。

常见错误

- 在 (iv) 中，忘记验证推出来的方阵确实是幻方。那段推论只说明其他填法都不可行。

- 在 (v) 中，只答“2”或“4”，却没有解释为甚么旋转会产生新的、真正不同的幻方，或者没有利用 (iii) 说明四种情况涵盖一切而又互不重叠。

STEP 之外

第 (v) 部分让你初尝*群作用*（group action）的滋味：正方形的八个对称（四个旋转和四个反射）构成一个群（group），而全部幻方在这个群的作用下组成单一个轨道（orbit）。像这样在对称意义下计算物件的数目，会在伯恩赛德引理（Burnside's lemma）中得到进一步发展，你会在群论（group theory）的入门课程中遇到它。Allenby 的 *Rings, Fields and Groups*（《环、域与群》）是一本平易近人的入门书。

详细解答

STEP I 1992，第 3 题

纯数学 · 奇函数与偶函数

STEP I 1992，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/92-s1-q3

关键思路

这里每个积分都是在对称区间 $[-\pi, \pi]$ 上计算的，所以对每个被积函数（integrand），首先要问的是：它是偶函数（even function）还是奇函数（odd function）？奇函数在对称区间上的积分为零，而偶函数的积分则等于在 $[0, \pi]$ 上积分的两倍，而在这个区间上绝对值（modulus）符号可以去掉。看出对称性之后，四个部分各自最多只需一步简短的计算。

解答

我们会用到两个标准结果。若 $f(-x) = f(x)$（偶函数），则

$$\int_{-a}^{a} f(x)\,\mathrm{d}x = 2\int_{0}^{a} f(x)\,\mathrm{d}x,$$

而若 $f(-x) = -f(x)$（奇函数），则 $\int_{-a}^{a} f(x)\,\mathrm{d}x = 0$。两者都可以用代换（substitution）$x = -u$ 处理 $[-a, 0]$ 上的积分而得出，该积分会变成 $\int_{0}^{a} f(-u)\,\mathrm{d}u$。

我们还需要 $\int_{0}^{\pi} \sin x\,\mathrm{d}x = \left[-\cos x\right]_{0}^{\pi} = 1 - (-1) = 2$。

第 (i) 部分

由于 $|\sin(-x)| = |-\sin x| = |\sin x|$，被积函数是偶函数。在 $[0, \pi]$ 上有 $\sin x \geq 0$，所以在该处 $|\sin x| = \sin x$。因此

$$\int_{-\pi}^{\pi} |\sin x|\,\mathrm{d}x = 2\int_{0}^{\pi} \sin x\,\mathrm{d}x = 2 \times 2 = 4.$$

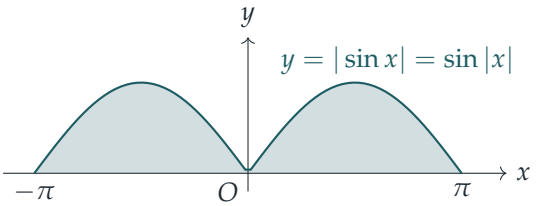
第 (ii) 部分

由于 $|-x| = |x|$，函数 $\sin|x|$ 是偶函数。在 $[0, \pi]$ 上，$|x| = x$，所以

$$\int_{-\pi}^{\pi} \sin|x|\,\mathrm{d}x = 2\int_{0}^{\pi} \sin x\,\mathrm{d}x = 4.$$

想一想 (i) 和 (ii) 的答案相同，是巧合吗?

不是:在这个区间上,两个被积函数其实是同一个函数。当 $0 \le x \le \pi$ 时,两者都等于 $\sin x$。当 $-\pi \le x \le 0$ 时,我们有 $\sin x \le 0$,所以 $|\sin x| = -\sin x$,而 $\sin|x| = \sin(-x) = -\sin x$ 亦然。两条曲线要到 $[-\pi, \pi]$ 之外才分道扬镳：例如在 $x = 3\pi/2$ 处，$|\sin x| = 1$，但 $\sin|x| = -1$。



阴影区域由正弦曲线的两个拱形组成，每个拱形的面积为 2。

第 (iii) 部分

这里 $(-x)\sin(-x) = (-x)(-\sin x) = x\sin x$，所以被积函数是偶函数（两个奇函数之积）。因此

$$\int_{-\pi}^{\pi} x\sin x\,\mathrm{d}x = 2\int_{0}^{\pi} x\sin x\,\mathrm{d}x.$$

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

$$\begin{aligned}\int_{0}^{\pi} x\sin x\,\mathrm{d}x &= \left[-x\cos x\right]_{0}^{\pi} + \int_{0}^{\pi} \cos x\,\mathrm{d}x \\ &= \left(-\pi\cos\pi - 0\right) + \left[\sin x\right]_{0}^{\pi} \\ &= \pi + 0 = \pi.\end{aligned}$$

因此

$$\int_{-\pi}^{\pi} x\sin x\,\mathrm{d}x = 2\pi.$$

第 (iv) 部分

现在 $(-x)^{10}\sin(-x) = x^{10}(-\sin x) = -x^{10}\sin x$，所以被积函数是奇函数（偶函数乘以奇函数）。于是它在对称区间上的积分为

$$\int_{-\pi}^{\pi} x^{10}\sin x\,\mathrm{d}x = 0.$$

根本不必做十次分部积分！来自 $[-\pi, 0]$ 和 $[0, \pi]$ 的贡献大小相等、符号相反，刚好完全抵消。

验算

- 从面积判断：在 (i) 中，被积函数非负且最大为 1，所以答案必定介乎 0 与 $2\pi \approx 6.28$ 之间；4 符合这个要求。
- 在 (iii) 中，$x\sin x \geq 0$ 在 $[-\pi, \pi]$ 上成立，所以答案必定为正数；2π 确是正数。对 $\sin x - x\cos x$ 求导得 $x\sin x$，由此确认原函数（antiderivative）无误。
- 以森逊法则（Simpson's rule）取 20 000 个子区间计算，四个积分分别得 4.0000000、4.0000000、6.2831853 和 0.0000000。在 (iv) 中，这个零并非只是一个很小的数：$|x^{10}\sin x|$ 在同一区间上的积分约为 13 275，可见抵消是精确而彻底的。

常见错误

- 写成 $\int_{-\pi}^{\pi} |\sin x|\,\mathrm{d}x = \left[-\cos x\right]_{-\pi}^{\pi} = 0$，忽略了绝对值。一个永不取负值的被积函数，除非几乎处处（almost everywhere）为零，否则积分不可能是 0。
- 因为 $\sin x$ 是奇函数，便以为 $\sin|x|$ 也是奇函数。任何函数与 $|x|$ 复合，得出的必定是偶函数。
- 在 (iii) 的分部积分中弄错正负号：记住 $\cos\pi = -1$，所以 $-\pi\cos\pi = +\pi$。

STEP 之外

把函数分拆成偶部和奇部，$f(x) = \frac{1}{2}\big(f(x) + f(-x)\big) + \frac{1}{2}\big(f(x) - f(-x)\big)$，正是傅里叶级数（Fourier series）的出发点。在 $[-\pi, \pi]$ 上，偶部展开成余弦项，奇部展开成正弦项，而像 (iii) 和 (iv) 这类积分，正是用来计算各项系数（coefficients）的积分。例如，x 的傅里叶级数中 $\sin x$ 的系数，就是来自 (iii) 的积分。Boyce 与 DiPrima 合著的 *Elementary Differential Equations*（常微分方程入门教材）对傅里叶级数及其在求解微分方程（differential equations）中的应用，有一段浅白易读的介绍。

详细解答

STEP I 1992，第 4 题

纯数学 · 阿根图上的轨迹

STEP I 1992，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/92-s1-q4

关键思路

把每个条件改写成关于 x 和 y 的叙述（其中 $z = x + \mathrm{i}y$），或者关于 z 的模（modulus）和辐角（argument）的叙述，所求的区域通常便一目了然。第 (i) 及 (ii) 部分只需要做到这一步。第 (iii) 部分则不同：它描述的是一个由和组成的集合，所以我们要证明两件事。每一个和都落在我们所声称的区域之内，而该区域的每一点也确实是这样的一个和。

解答

以下一律设 $z = x + \mathrm{i}y$，其中 x, y 为实数。

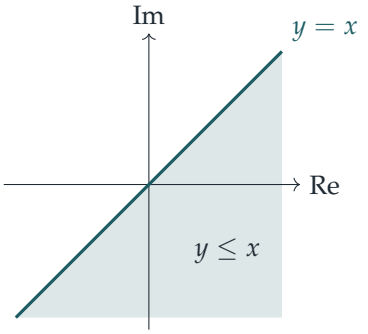
第 (i) 部分

展开得

$$(1+\mathrm{i})z = (1+\mathrm{i})(x+\mathrm{i}y) = (x-y) + \mathrm{i}(x+y),$$

所以 $\mathrm{Re}\big((1+\mathrm{i})z\big) = x - y$。条件就是 $x - y \geq 0$，即 $y \leq x$。这个集合是直线 $y = x$ 上及其下方的闭半平面（closed half-plane），包括直线本身。

从几何角度看，乘以 $1 + \mathrm{i} = \sqrt{2}\,\mathrm{e}^{\mathrm{i}\pi/4}$ 相当于旋转 $\frac{\pi}{4}$ 并放大 $\sqrt{2}$ 倍。因此条件是说，把 z 逆时针旋转 $\frac{\pi}{4}$ 之后，它会落在右半平面 $\mathrm{Re} \geq 0$ 之内。这表示 $-\frac{3\pi}{4} \leq \arg z \leq \frac{\pi}{4}$（或 $z = 0$），正是同一个半平面。



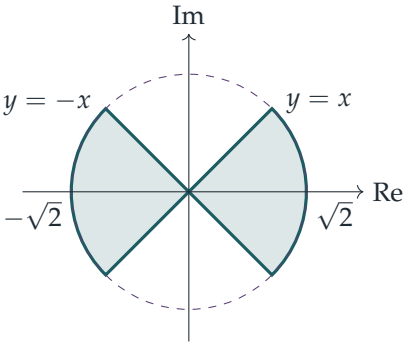
第 (ii) 部分

由于 $|z^2| = |z|^2$，第一个条件就是 $|z|^2 \leq 2$，即 $|z| \leq \sqrt{2}$：这是以原点为圆心、半径为 $\sqrt{2}$ 的闭圆盘（closed disc）。

至于第二个条件，$z^2 = (x^2 - y^2) + 2\mathrm{i}xy$，所以 $\mathrm{Re}(z^2) = x^2 - y^2$，条件就是 $x^2 \geq y^2$，即 $|y| \leq |x|$。这是直线 $y = x$ 与 $y = -x$ 之间、包含实轴的一对相对的闭扇形（sector）区域。

用极式（polar form）来做会更快：若 $z = r\mathrm{e}^{\mathrm{i}\theta}$，则 $z^2 = r^2\mathrm{e}^{2\mathrm{i}\theta}$，所以我们需要 $r \leq \sqrt{2}$ 及 $\cos 2\theta \geq 0$，由此得 $-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$ 或 $\frac{3\pi}{4} \leq \theta \leq \frac{5\pi}{4}$（或 $z = 0$）。

所求集合是两者的交集（intersection）：两个闭的四分之一圆盘（各为圆心角 $\frac{\pi}{2}$、半径 $\sqrt{2}$ 的扇形），一个向右张开，一个向左张开，在原点相接。整条边界都包括在内。



第 (iii) 部分

设 S 为所有这些和组成的集合。我们声称 S 就是闭圆环（annulus）$\{z : 1 \leq |z| \leq 3\}$。

每一个和都落在圆环之内。若 $|z_1| = 2$ 及 $|z_2| = 1$，由三角不等式（triangle inequality）得

$$|z_1 + z_2| \leq |z_1| + |z_2| = 3, \qquad |z_1 + z_2| \geq |z_1| - |z_2| = 1.$$

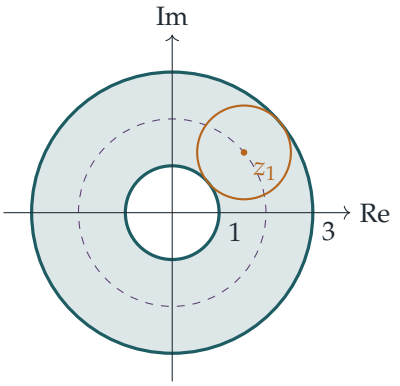
想一想 这样就足够了吗? 我们怎样知道整个圆环都被填满, 而不只是圆环内的某条曲线?

这两个不等式只说明 S 位于圆环之内，所以我们还要证明反方向的包含关系。任取一个满足 $1 \leq |z| \leq 3$ 的 z。我们想找一个模为一的复数 z_2，使得 $|z - z_2| = 2$，因为这样 $z_1 = z - z_2$ 的模便是 2，而且 $z = z_1 + z_2$。

从几何角度看, z_2 必须同时位于以 0 为圆心的单位圆上, 以及以 z 为圆心、半径为 2 的圆上。两个半径分别为 1 和 2 的圆, 若圆心相距 $d = |z|$, 则它们相交当且仅当 $2 - 1 \leq d \leq 2 + 1$, 而这正是我们的条件 $1 \leq |z| \leq 3$。

如果不想直接引用这个关于圆的事实, 可以这样看: 设 $z = re^{i\alpha}$, 并考虑 $f(\varphi) = |z - e^{i\varphi}|$, 它是 φ 的连续（continuous）函数。当 $\varphi = \alpha$ 时，点 $e^{i\varphi}$ 与 z 方向相同，所以 $f(\alpha) = |r - 1| = r - 1 \leq 2$。当 $\varphi = \alpha + \pi$ 时, 它指向相反方向, 所以 $f(\alpha + \pi) = r + 1 \geq 2$。根据中间值定理（intermediate value theorem），存在某个 φ 使得 $f(\varphi) = 2$，于是取 $z_2 = e^{i\varphi}$ 即可。

因此 S 恰好是圆 $|z| = 1$ 与圆 $|z| = 3$ 之间的闭圆环，两个边界圆都包括在内。有一个很好的想像方法：把 z_1 固定在半径为 2 的圆上，$z_1 + z_2$ 便描出一个以 z_1 为圆心的单位圆。当 z_1 绕圆一周时，这些小圆便扫出整个圆环。



琥珀色的圆显示了当 z_1 固定在虚线圆 $|z_1| = 2$ 上某一点时所得的各个和。

验算

- 在 (i) 中，$z = 1$ 给出 $\mathrm{Re}(1 + \mathrm{i}) = 1 \geq 0$，而 $z = \mathrm{i}$ 给出 $\mathrm{Re}(-1 + \mathrm{i}) = -1 < 0$。这与 $y \leq x$ 吻合。
- 在 (ii) 中，$z = 1$ 属于该集合而 $z = \mathrm{i}$ 不属于（$\mathrm{i}^2 = -1$），所以两个扇形应该包含实轴，与图中所画一致。

- 在 (iii) 中，$z_1 = 2$、$z_2 = -1$ 给出 1，而 $z_1 = 2$、$z_2 = 1$ 给出 3，所以两个边界圆都可以达到。我们用 200 000 个随机点测试了全部三个描述。随机所得的和，其模介乎 1.0000 与 3.0000 之间；而对于圆环内的 1000 个随机点，都能找到合适的 z_2，误差在 2×10^{-4} 以内。

常见错误

- 在 (ii) 中，只保留扇形 $|\arg z| \le \frac{\pi}{4}$ 而忘记了左边的那一个。z 与 $-z$ 的平方是相同的。
- 在 (iii) 中，用完三角不等式便停下来。这只说明 S 位于圆环之内，并未说明 S 就是整个圆环。
- 没有交代边界是否包括在内。这里每一个不等式都是非严格的，所以每一条边界都包括在内。

STEP 之外

第 (iii) 部分是两个集合的闵可夫斯基和（*Minkowski sum*）$A + B = \{a + b : a \in A,\ b \in B\}$ 的一个小例子，这个概念贯穿几何学和凸分析（convex analysis）。双向包含的论证（先证 $S \subseteq T$，再证 $T \subseteq S$）是证明两个集合相等的标准方法，你在大学里会经常用到。若想从几何角度理解复数乘法就是旋转加缩放（这会令 (i) 和 (ii) 几乎一望而知），可参阅 Needham 的 *Visual Complex Analysis*（以图像方式讲解复分析的著作）。

详细解答

STEP I 1992，第 5 题

纯数学 · 多项式的重因式

> **STEP I 1992，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/92-s1-q5

本题在 *Beyond STEP* 第 9.7 节相等的幂和中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch09/equal-power-sums.html

关键思路

若一个多项式（polynomial）有重因式（repeated factor）$(1-x)^m$，则对它求导（再乘以 x）之后，仍会留下因式 $(1-x)^{m-1}$。因此 1 是这个多项式的根（root），也是这条链中随后几个多项式的根。利用各多项式的系数（coefficient）计算它们在 $x=1$ 的值，会得到形如 $\sum k^r c_k$ 的和。当每个系数都是 ± 1 时，每一个这样的和都表示：两组整数的 r 次幂之和相等。

解答

因式 $(1-x)^3$

两次运用平方差（difference of two squares），

$$1-x^2=(1-x)(1+x), \qquad 1-x^4=(1-x^2)(1+x^2)=(1-x)(1+x)(1+x^2),$$

所以

$$p_0(x)=(1-x)^3(1+x)^2(1+x^2),$$

故 $(1-x)^3$ 是 $p_0(x)$ 的因式，正如题目所求。

$p_0'(1)=0$ 与 $p_1'(1)=0$

我们先证明一个小结果，然后重复使用。

引理（*Lemma*）。若 $\mathrm{p}(x) = (1-x)^m \mathrm{q}(x)$，其中 $m \geq 1$，而 q 是多项式，则 $\mathrm{p}'(x)$ 有因式 $(1-x)^{m-1}$，因而 $x\mathrm{p}'(x)$ 亦然。

证明。根据积法则（product rule），

$$\mathrm{p}'(x) = -m(1-x)^{m-1}\mathrm{q}(x) + (1-x)^m \mathrm{q}'(x) = (1-x)^{m-1}\big[(1-x)\mathrm{q}'(x) - m\,\mathrm{q}(x)\big],$$

而方括号内是一个多项式。乘以 x 并不会除去这个因式。□

取 $m = 3$ 应用此引理：$\mathrm{p}_0'(x)$ 有因式 $(1-x)^2$，所以 $\mathrm{p}_0'(1) = 0$。同时 $\mathrm{p}_1(x) = x\mathrm{p}_0'(x)$ 也有因式 $(1-x)^2$。再取 $m = 2$ 应用引理，$\mathrm{p}_1'(x)$ 有因式 $1-x$，所以 $\mathrm{p}_1'(1) = 0$。

想一想 为甚么不直接对三个括号的乘积求导，然后代入 $x = 1$?

这样做也可以，但题目要求我们从因式出发论证，而且因式的论证才是可以推广的那一种。重数（multiplicity）为 m 的根，经过 $m-1$ 次求导后仍然是根，而这正是最后一部分所需要的。

两个恒等式

展开得

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

然后

$$\mathrm{p}_0(x) = 1 - x - x^2 + x^3 - x^4 + x^5 + x^6 - x^7.$$

所以 $c_0 = c_3 = c_5 = c_6 = 1$，而 $c_1 = c_2 = c_4 = c_7 = -1$。

现在

$$\mathrm{p}_0'(x) = \sum_{k=1}^{7} k\,c_k x^{k-1}, \qquad \mathrm{p}_1(x) = \sum_{k=0}^{7} k\,c_k x^k, \qquad \mathrm{p}_1'(x) = \sum_{k=1}^{7} k^2 c_k x^{k-1}.$$

代入 $x = 1$，并利用 $\mathrm{p}_0'(1) = 0$:

$$-1 - 2 + 3 - 4 + 5 + 6 - 7 = 0, \quad \text{即} \quad 1 + 2 + 4 + 7 = 3 + 5 + 6.$$

利用 $\mathrm{p}_1'(1) = 0$:

$$-1^2 - 2^2 + 3^2 - 4^2 + 5^2 + 6^2 - 7^2 = 0, \quad \text{即} \quad 1^2 + 2^2 + 4^2 + 7^2 = 3^2 + 5^2 + 6^2.$$

（两边都等于 14；平方的情况则都等于 70。）c_0 一项没有任何贡献，因为它乘上了 0 或 0^2。

最后一部分

再多走一步。设

$$\mathrm{P}_0(x) = (1-x)(1-x^2)(1-x^4)(1-x^8) = \mathrm{p}_0(x)(1-x^8).$$

由于 $1-x^8=(1-x)(1+x)(1+x^2)(1+x^4)$，$\mathrm{P}_0(x)$ 有因式 $(1-x)^4$。对 $j=0,1,2$，定义 $\mathrm{P}_{j+1}(x)=x\mathrm{P}_j'(x)$。根据引理，$\mathrm{P}_1$、$\mathrm{P}_2$ 和 P_3 分别有因式 $(1-x)^3$、$(1-x)^2$ 和 $(1-x)$，所以

$$\mathrm{P}_1(1)=\mathrm{P}_2(1)=\mathrm{P}_3(1)=0.$$

记 $\mathrm{P}_0(x)=\sum_{k=0}^{15}d_kx^k$。每施行一次 $x\frac{\mathrm{d}}{\mathrm{d}x}$，$x^k$ 的系数便乘以 k，所以 $\mathrm{P}_r(x)=\sum_{k=0}^{15}k^rd_kx^k$，于是

$$\sum_{k=1}^{15}k^rd_k=0\qquad\text{其中}r=1,2,3.$$

要求出各个 d_k，把 $\mathrm{p}_0(x)$ 乘以 $1-x^8$，得

$$\mathrm{P}_0(x)=\mathrm{p}_0(x)-x^8\mathrm{p}_0(x).$$

第二项给出 $x^8,\ldots,x^{15}$ 各项，其正负号与 p_0 的相反。两部分没有重叠，所以每个 d_k 都是 ±1：

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d_k	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+

（事实上 $d_k=(-1)^{s(k)}$，其中 $s(k)$ 是 k 的二进制展开式（binary expansion）中 1 的个数，因为 x^k 是按 k 的二进制数字，从对应的括号中选取 $-x^{2^i}$ 而得出的。）

撇除 $k=0$，当 $k\in\{1,2,4,7,8,11,13,14\}$（八个数）时系数 $d_k=-1$，当 $k\in\{3,5,6,9,10,12,15\}$（七个数）时 $d_k=+1$。这些数合起来恰好是 $1,2,\ldots,15$。于是取

$$(a_1,\ldots,a_8)=(1,2,4,7,8,11,13,14),\qquad(a_9,\ldots,a_{15})=(3,5,6,9,10,12,15).$$

那么 $\sum k^rd_k=0$ 所说的正是

$$a_1^r+\cdots+a_8^r=a_9^r+\cdots+a_{15}^r\qquad(r=1,2,3),$$

这就是我们要证明的结果。

想一想 为甚么这个分法一边有八个数，另一边只有七个?

$+1$ 的一边其实还包含 $k=0$，而当 $r\geq1$ 时它的贡献是 $0^r=0$，所以可以略去。这就是两边的项数不必相同的原因。（若 $r=0$，恒等式便需要这个 0：这时是 $8=8$。）

验算

直接计算：$r=1$ 时两边之和为 $60=60$，$r=2$ 时为 $620=620$，$r=3$ 时为 $7200=7200$。我们用电脑展开了 P_0，确认了它各项的正负号，以及根 $x=1$ 的重数为 4，并验证了全部三个等式。当 $r=4$ 时等式不成立（该根的重数只有 4），这正好显示上述论证已是最佳的。

常见错误

- 忘记了 $k = 0$ 一项，因而以为两边的项数必须相同。
- 单凭 $\mathrm{p}_0''(1) = 0$ 便断定 $\mathrm{p}_1'(1) = 0$。由于 $\mathrm{p}_1' = \mathrm{p}_0' + x\mathrm{p}_0''$，你需要 $\mathrm{p}_0'(1) = 0$ 和 $\mathrm{p}_0''(1) = 0$ 两者同时成立。

STEP 之外

这就是 Prouhet–Tarry–Escott 问题：按二进制数字和的奇偶性（parity）把 $\{0, 1, \ldots, 2^n - 1\}$ 分成两组（即 Thue–Morse 数列），所得的两个集合对所有 $r < n$ 都有相等的 r 次幂之和。它背后的工具——重数为 m 的根同时是首 $m - 1$ 阶导数（derivative）的根——在任何域（field）上的多项式代数中都占有核心地位。关于多项式及其根的代数，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）。至于更广泛的数论（number theory），可参阅 Davenport 的 *The Higher Arithmetic*（高等算术）。

详细解答

STEP I 1992，第 6 题

纯数学 • 小角近似

STEP I 1992，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/92-s1-q6

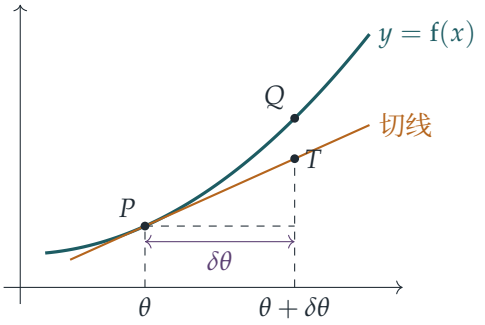
关键思路

远处那一点 C 的角恰好是 ϕ，因为 $\theta + \phi$ 是三角形 ABC 的外角（exterior angle）。于是由正弦定理（sine rule）可求得 AC 的准确值。由于 C 非常遥远，ϕ 必定极小，而小角近似（small-angle approximation，其实就是第一部分的线性近似换了个面貌）可把准确公式化成简单的式子。至于误差的问题，不过是把同一个线性近似再用一次，这次以 ϕ 为变量。

解答

线性近似

在点 $P = (\theta, \mathrm{f}(\theta))$ 附近，f 的图像非常贴近它的切线（tangent），而切线的斜率（gradient）是 $\mathrm{f}'(\theta)$。



向右移动一段距离 $\delta\theta$。沿着切线走，高度上升了（斜率）×（水平距离）$= \mathrm{f}'(\theta)\,\delta\theta$，到

达 T。沿着曲线走，则上升至 Q，高度为 $\mathrm{f}(\theta+\delta\theta)$。当 $\delta\theta\to 0$ 时，间距 QT 缩小得比 $\delta\theta$ 本身还要快，因为

$$\frac{\mathrm{f}(\theta+\delta\theta)-\mathrm{f}(\theta)}{\delta\theta}\to \mathrm{f}'(\theta)$$

正正就是导数（derivative）的定义。所以当 $\delta\theta$ 很小时，

$$\mathrm{f}(\theta+\delta\theta)\approx \mathrm{f}(\theta)+\mathrm{f}'(\theta)\,\delta\theta.$$

C 的距离

在三角形 ABC 中，A 处的角是 θ，而 B 处的外角是 $\theta+\phi$。外角等于两个内对角之和，所以

$$\angle ACB=\phi.$$

B 处的内角是 $\pi-(\theta+\phi)$，而 $\sin(\pi-x)=\sin x$，所以由正弦定理得出以下准确结果：

$$AC=\frac{a\sin(\theta+\phi)}{\sin\phi},\qquad h=AC\sin\theta=\frac{a\sin\theta\,\sin(\theta+\phi)}{\sin\phi}.$$

想一想 “h 远大于 a”这个条件究竟在哪里用上？

它告诉我们 ϕ 很小。把准确公式重新整理，得 $\sin\phi=a\sin\theta\sin(\theta+\phi)/h\le a/h$，所以 $h\gg a$ 迫使 $\sin\phi$ 极小，因而 ϕ 也极小。（从非常遥远的 C 望过来，基线（baseline）AB 所张的角几乎是零。）现在取 $\mathrm{f}=\sin$，应用第一部分的结果：

$$\sin(\theta+\phi)\approx\sin\theta+\phi\cos\theta\approx\sin\theta,\qquad \sin\phi\approx\sin 0+\phi\cos 0=\phi.$$

因此

$$h\approx\frac{a\sin^2\theta}{\phi},$$

这正是所求的结果，并且

$$AC=\frac{h}{\sin\theta}\approx\frac{a\sin\theta}{\phi}.$$

（若保留 $\phi\cos\theta$ 一项，会得到 $AC\approx a\sin\theta/\phi+a\cos\theta$。多出来的部分只有 a 的大小，与量级为 h 的距离相比可以忽略不计。）

ϕ 的误差所造成的影响

把 AC 看成函数 $g(\phi)=a\sin\theta\,\phi^{-1}$，其中 θ 固定不变，因为它的量度是准确的。根据第一部分，误差 $\delta\phi$ 会令估计值改变

$$\delta(AC)\approx g'(\phi)\,\delta\phi=-\frac{a\sin\theta}{\phi^2}\,\delta\phi.$$

所以 AC 的误差大小约为 $\dfrac{a\sin\theta}{\phi^2}\,|\delta\phi|$；高估 ϕ 就会低估距离。

想一想 如果把 a 加倍，公式的分子有 a，误差岂不是也加倍?

不会, 因为 ϕ 并非固定: 它取决于基线的长度。点 C（因而 h）保持不变, 而由 $h \approx a\sin^2\theta/\phi$ 可得 $\phi \approx a\sin^2\theta/h$，所以把 a 加倍，ϕ 也大约加倍。（由于 C 非常遥远，θ 只有极轻微的改变。）以固定的量 h 来表示误差:

$$\frac{a\sin\theta}{\phi^2}\,|\delta\phi| \approx a\sin\theta\cdot\frac{h^2}{a^2\sin^4\theta}\,|\delta\phi| = \frac{h^2}{a\sin^3\theta}\,|\delta\phi|.$$

对于相同的仪器误差 $\delta\phi$，AC 的误差与 a 成反比。**把两台望远镜之间的距离加倍，误差大约减半**。

验算

- 取 $a = 1$、$\theta = \pi/6$、$\phi = 10^{-4}$。准确公式给出 $h \approx 2500.43$ 和 $AC \approx 5000.87$；近似公式则给出 2500 和 5000。AC 的差别约为 $0.87 \approx a\cos\theta$，与额外一项所预测的完全吻合。
- 直接把 C 当作两条视线的交点来计算, 对数以千计随机选取的 a、θ 和细小的 ϕ, 结果与两个近似式的差距都在 0.1% 以内, 而细小的 $\delta\phi$ 所引致的 AC 变化亦与 $-a\sin\theta\,\delta\phi/\phi^2$ 相符。
- 对于固定的远处一点 C，分别以基线 a 和 $2a$ 重新计算 ϕ，再对两者加上相同的扰动 $\delta\phi$，所得的误差比为 2。
- 量纲（dimensions）：$a\sin^2\theta/\phi$ 是长度，因为以弧度（radian）表示的角是无量纲的。

常见错误

- ϕ 以度为单位时使用 $\sin\phi \approx \phi$。这个近似式与 sin 的 f′ 一样，必须使用弧度。
- 把 ϕ 当作固定，从而断定误差随 a 加倍。目标固定的是 h，而不是 ϕ。
- 没有察觉 $\angle ACB = \phi$，转而建立坐标来计算。这样做也行得通，但较为费时。

STEP 之外

这就是利用视差（parallax）进行的三角测量（triangulation），天文学家正是用这个方法，以地球轨道为基线，量度邻近恒星的距离。这里的误差分析让我们初尝误差传递（error propagation）的滋味：在一阶近似下，量度值的误差 δx 会在任何由它计算出来的量中变成误差 $g'(x)\,\delta x$。这个想法经带余项的泰勒定理（Taylor's theorem）严格化之后，是

大学一年级分析学（analysis）的核心内容；参阅 Abbott, *Understanding Analysis*（分析学入门教材）。

详细解答

STEP I 1992，第 7 题

纯数学 · 整值多项式

STEP I 1992，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/92-s1-q7

关键思路

不要试图证明各系数（coefficient）都是整数（integer），因为对二次函数（quadratic function）而言，它们不一定是整数：$\frac{1}{2}x^2+\frac{1}{2}x$ 在每个整数处都取整数值。正确的做法是用已知的整数值表示各系数，再代回原式，使 $\mathrm{f}(n)$ 写成 $\mathrm{f}(-1)$、$\mathrm{f}(0)$ 和 $\mathrm{f}(1)$ 的组合，而各项的乘数都是整数。至于最后一部分，只要把原点平移，便是把中间部分再做一次。

解答

线性函数

由于 $\mathrm{g}(0)=b$ 及 $\mathrm{g}(1)=a+b$，我们有 $b=\mathrm{g}(0)$ 及 $a=\mathrm{g}(1)-\mathrm{g}(0)$，两者都是整数。因此对任意整数 n，

$$\mathrm{g}(n)=an+b=\mathrm{g}(0)+n\big(\mathrm{g}(1)-\mathrm{g}(0)\big),$$

这是若干个整数乘积之和，所以是整数。

二次函数

设 $p=\mathrm{f}(-1)$，$q=\mathrm{f}(0)$，$r=\mathrm{f}(1)$，三者都是整数。于是

$$p=A-B+C,\qquad q=C,\qquad r=A+B+C,$$

解之，得

$$C=q,\qquad B=\frac{r-p}{2},\qquad A=\frac{p+r-2q}{2}.$$

所以 A 和 B 不一定是整数；它们可以是奇数的一半。把这些结果代入 $\mathrm{f}(n) = An^2 + Bn + C$，并按 p、q 和 r 整理各项：

$$\begin{aligned}\mathrm{f}(n) &= \frac{p+r-2q}{2}\,n^2 + \frac{r-p}{2}\,n + q \\ &= \frac{n(n-1)}{2}\,p + (1-n^2)\,q + \frac{n(n+1)}{2}\,r.\end{aligned}$$

? **想一想** p 和 r 的乘数，分母都有一个 2。为甚么它们是整数?

对于整数 n，$n-1$ 和 n 是两个连续整数，其中必有一个是偶数，所以 $n(n-1)$ 是偶数。同理，$n(n+1)$ 也是偶数。因此 $\frac{1}{2}n(n-1)$ 和 $\frac{1}{2}n(n+1)$ 都是整数，而 $1-n^2$ 也是整数。故此 $\mathrm{f}(n)$ 是若干个整数乘积之和，所以对每个整数 n 它都是整数，这正是所要证明的。

（这三个乘数都是二次式，各自在 $-1, 0, 1$ 的其中一点取值 1，而在另外两点取值 0。这就是拉格朗日（Lagrange）插值法（Lagrange interpolation）。）

最后一部分

设 α 为实数，并定义 $\mathrm{h}(x) = \mathrm{f}(\alpha + x)$。展开后得

$$\mathrm{h}(x) = A(\alpha+x)^2 + B(\alpha+x) + C = Ax^2 + (2A\alpha + B)x + (A\alpha^2 + B\alpha + C),$$

所以 h 同样是形如 $A'x^2 + B'x + C'$ 的函数，而且系数都是实数。中间部分的论证除了用到 A、B、C 是实数之外，并没有用到它们的其他性质。现在

$$\mathrm{h}(-1) = \mathrm{f}(\alpha - 1), \qquad \mathrm{h}(0) = \mathrm{f}(\alpha), \qquad \mathrm{h}(1) = \mathrm{f}(\alpha + 1)$$

按假设都是整数。因此，把中间部分的结果应用于 h，可知对每个整数 n，$\mathrm{h}(n) = \mathrm{f}(\alpha + n)$ 都是整数。具体写出来就是

$$\mathrm{f}(\alpha+n) = \frac{n(n-1)}{2}\,\mathrm{f}(\alpha-1) + (1-n^2)\,\mathrm{f}(\alpha) + \frac{n(n+1)}{2}\,\mathrm{f}(\alpha+1).$$

验算

- 把 $n = -1, 0, 1$ 代入 $\mathrm{f}(n)$ 的公式：各乘数分别变成 $(1,0,0)$、$(0,1,0)$ 和 $(0,0,1)$，正好还原出 p、q、r。
- 以 $\mathrm{f}(x) = \frac{1}{2}x^2 + \frac{1}{2}x$ 为例，有 $p = 0$，$q = 0$，$r = 1$，而公式给出 $\mathrm{f}(n) = \frac{1}{2}n(n+1)$，即三角形数（triangular numbers）。
- 数值验算：取 $\alpha = \sqrt{2}$ 及 $\mathrm{f}(\alpha - 1) = 3$，$\mathrm{f}(\alpha) = -7$，$\mathrm{f}(\alpha + 1) = 11$，则通过这三点的二次函数给出 $\mathrm{f}(\alpha + 5) = 363.000\ldots$，与 $10 \cdot 3 - 24 \cdot (-7) + 15 \cdot 11 = 363$ 吻合。

常见错误

- 声称 A、B、C 必定是整数。它们不一定是整数，任何依赖这一点的论证都是错的。
- 在最后一部分，试图证明 α 必定是整数或有理数。它可以是任何实数；重要的只是平移。

STEP 之外

在每个整数处都取整数值的多项式（polynomial），恰好就是二项式系数（binomial coefficients）$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}$ 的整数组合。对二次多项式而言，这些就是 1、x 和 $\frac{1}{2}x(x-1)$，而这个结果可用有限差分（finite differences）证明：一个 d 次多项式，只要在 $d+1$ 个连续整数处取整数值，它就是整值（integer-valued）多项式。同样的想法也是数值分析（numerical analysis）中牛顿（Newton）前向差分公式（forward-difference formula）的基础。关于整值函数与整除性（divisibility）的更多内容，可参阅 Davenport 的 *The Higher Arithmetic*（一本数论入门经典）。

详细解答

STEP I 1992，第 8 题

纯数学 · 微积分基本定理

STEP I 1992，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/92-s1-q8

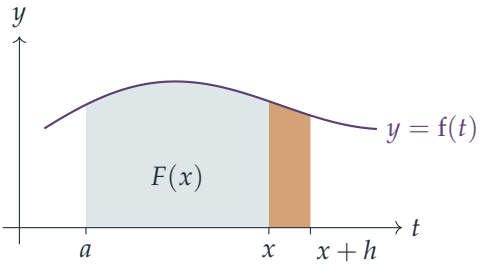
关键思路

上限为变量的积分，是一块随上限移动而增长的面积，而它的增长率正好等于曲线在该上限处的高度。单凭这一个事实（微积分基本定理，Fundamental Theorem of Calculus），就可以把题目中每一条积分方程化为微分方程（differential equation）。求导会失去有关常数的资料，所以我们每次都要把一个方便的 x 值（通常是 $x = 0$）代入原来的方程，把它找回来。

解答

积分的导数

设 $F(x) = \int_a^x \mathrm{f}(t)\,\mathrm{d}t$，即 $y = \mathrm{f}(t)$ 之下由 $t = a$ 至 $t = x$ 的面积。把 x 增加一个微小的量 h，就会多出一条阔度为 h 的窄条：



这条窄条的面积是 $F(x+h) - F(x)$。若 f 是连续（continuous）的，则在 $[x, x+h]$ 上，曲线与 $\mathrm{f}(x)$ 的差距不会超过一个微小的量 $\varepsilon(h)$，而当 $h \to 0$ 时 $\varepsilon(h) \to 0$。因此窄条夹在高

度为 $\mathrm{f}(x) \pm \varepsilon(h)$ 的两个长方形之间：

$$\big(\mathrm{f}(x) - \varepsilon(h)\big)h \le F(x+h) - F(x) \le \big(\mathrm{f}(x) + \varepsilon(h)\big)h.$$

除以 $h > 0$（$h < 0$ 的情况相同，只是窄条在左边），再令 $h \to 0$，得

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x \mathrm{f}(t)\,\mathrm{d}t = \lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \mathrm{f}(x).$$

留意结果中并没有 a：改变下限只会令 F 加上一个常数。

方程 $\mathrm{f}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t + 1$

根据上面的结果，右边是可导（differentiable）的，所以 f 也可导，而两边求导可得

$$\mathrm{f}'(x) = \mathrm{f}(x).$$

把 $x = 0$ 代入原来的方程，积分的区间长度为零，所以 $\mathrm{f}(0) = 1$。

想一想 我们怎样知道 $\mathrm{f}' = \mathrm{f}$ 除了 e^x 的倍数之外，再没有其他解？

考虑 $g(x) = \mathrm{e}^{-x}\mathrm{f}(x)$。由积法则（product rule），

$$g'(x) = \mathrm{e}^{-x}\big(\mathrm{f}'(x) - \mathrm{f}(x)\big) = 0,$$

所以 g 是常数，而 $g(0) = \mathrm{f}(0) = 1$。因此 $\mathrm{e}^{-x}\mathrm{f}(x) = 1$，即 $\mathrm{f}(x) = \mathrm{e}^x$，正如题目所求。（反过来，$\int_0^x \mathrm{e}^t\,\mathrm{d}t + 1 = \mathrm{e}^x - 1 + 1 = \mathrm{e}^x$，所以它确实是一个解。）

方程 $\mathrm{f}(x) = \int_0^x \mathrm{f}(t)\,\mathrm{d}t$

与之前完全一样，$\mathrm{f}' = \mathrm{f}$，所以 $\mathrm{e}^{-x}\mathrm{f}(x)$ 是常数。但这次代入 $x = 0$ 得 $\mathrm{f}(0) = 0$，所以该常数是 0，而唯一的解是

$$\mathrm{f}(x) = 0 \quad \text{对所有} x.$$

上一条方程中的“+1”正是决定指数函数大小的关键；没有了它，就只剩下零函数。

最后部分

把两边对 x 求导。在右边，x 现在是下限，而

$$\int_x^1 t^2\mathrm{f}(t)\,\mathrm{d}t = -\int_1^x t^2\mathrm{f}(t)\,\mathrm{d}t, \qquad \text{所以} \qquad \frac{\mathrm{d}}{\mathrm{d}x}\int_x^1 t^2\mathrm{f}(t)\,\mathrm{d}t = -x^2\mathrm{f}(x).$$

因此

$$\mathrm{f}(x) = -x^2\mathrm{f}(x) + 1 - x^4,$$
$$(1+x^2)\,\mathrm{f}(x) = 1 - x^4 = (1-x^2)(1+x^2).$$

由于 $1+x^2>0$，我们可以把两边除以它，得

$$\mathrm{f}(x)=1-x^2.$$

要求 C，把 $x=0$ 代入题目给出的方程。左边以及 $x-\frac{x^5}{5}$ 各项都变为零，余下

$$0=\int_0^1 t^2(1-t^2)\,\mathrm{d}t+C=\left[\tfrac{t^3}{3}-\tfrac{t^5}{5}\right]_0^1+C=\tfrac{2}{15}+C,$$

所以 $C=-\frac{2}{15}$，正如题目所求。

想一想 求导只告诉我们两边的导数相同。$\mathrm{f}(x)=1-x^2$ 和 $C=-\frac{2}{15}$ 是否真的对每一个 x 都满足方程?

是的。取这个 f 时，两边的导数对所有 x 都相等，所以两边只相差一个常数；而我们选取 C，正是要令两边在 $x=0$ 时相等。因此两边处处相等。反过来，任何解都必须满足求导后的方程，所以 $1-x^2$ 是唯一可能的 f。

验算

- 当 $x=1$：左边是 $\int_0^1(1-t^2)\,\mathrm{d}t=\frac{2}{3}$，右边是 $0+1-\frac{1}{5}-\frac{2}{15}=\frac{10}{15}=\frac{2}{3}$。两者吻合。
- 数值上，用辛普森法则（Simpson's rule）可得 $\int_0^1 t^2(1-t^2)\,\mathrm{d}t\approx 0.133333=\frac{2}{15}$，而最后一条方程的两边在 $x=-1.3,\ 0.25,\ 0.5,\ 0.9,\ 2$ 处的相差都在 10^{-10} 之内。

常见错误

- 当变量是下限时忘记负号：$\frac{\mathrm{d}}{\mathrm{d}x}\int_x^1 g(t)\,\mathrm{d}t=-g(x)$。
- 解出 $\mathrm{f}'=\mathrm{f}$ 后停在 Ae^x：要靠原来的积分方程在 $x=0$ 时的情况，才能确定 A。
- 写出 $\mathrm{f}(x)=\int_0^x \mathrm{f}(x)\,\mathrm{d}x$ 这类式子，把哑变量（dummy variable）t 与上限 x 混为一谈。

STEP 之外

未知函数出现在积分之内的方程，称为*积分方程*（integral equations）；本题这类带有可变积分限的方程是沃尔泰拉方程（Volterra equations），它们等价于微分方程的初值问题（initial-value problem）。在大学的第一个分析课程中，你会从黎曼积分（Riemann integral）的定义出发，严谨地证明微积分基本定理，而上面的窄条论证就会变成一个 ε–δ 证明。Abbott 的 *Understanding Analysis* 把这部分写得十分易读。

详细解答

STEP I 1992，第 9 题

纯数学 · 微分方程

STEP I 1992，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/92-s1-q9

关键思路

一切都源自一条方程：水的体积的变化率等于（流入率）减（流出率）。利用相似三角形（similar triangles），体积是水深 x 的一个简单函数，于是链式法则（chain rule）便把这条方程化为关于 x 的微分方程（differential equation）。在第一部分，这条方程可以干净利落地分离变量（separate variables）；在第二部分，我们不去求解，而是从 $\mathrm{d}x/\mathrm{d}t$ 的正负号，以及它等于零时的水深，直接读出水位的变化情况。

解答

圆锥排空

当水深为 x cm 时，水占据一个较小的圆锥，它与整个滤器相似，高为 x，半径为 r。由相似三角形得 $r/x = a/H$，所以 $r = ax/H$，而水的体积为

$$V = \tfrac{1}{3}\pi r^2 x = \frac{\pi a^2 x^3}{3H^2}.$$

由链式法则，

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\mathrm{d}V}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\pi a^2 x^2}{H^2}\frac{\mathrm{d}x}{\mathrm{d}t}.$$

水只出不入，流出率为 Ax，所以 $\mathrm{d}V/\mathrm{d}t = -Ax$。因此

$$\frac{\pi a^2 x^2}{H^2}\frac{\mathrm{d}x}{\mathrm{d}t} = -Ax \quad\Longrightarrow\quad x\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{AH^2}{\pi a^2},$$

这里我们除以了 x，只要圆锥内仍然有水（$x > 0$），这样做是容许的。左边就是 $\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}x^2\right)$，所以 $\frac{1}{2}x^2$ 以恒定的速率 $AH^2/(\pi a^2)$ 减少。它由 $\frac{1}{2}h^2$ 开始，到达 0 时圆锥便排空，所以所

需时间为

$$T = \frac{\frac{1}{2}h^2}{AH^2/(\pi a^2)} = \frac{\pi a^2 h^2}{2AH^2} \text{ 秒},$$

正如题目所求。（等价的做法是分离变量：$\int_h^0 x\,\mathrm{d}x = -\frac{AH^2}{\pi a^2}\int_0^T \mathrm{d}t$。）

同时注水

现在 $\mathrm{d}V/\mathrm{d}t = B - Ax$，所以

$$\frac{\pi a^2 x^2}{H^2}\frac{\mathrm{d}x}{\mathrm{d}t} = B - Ax, \qquad \text{即} \qquad \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{H^2(B-Ax)}{\pi a^2 x^2}.$$

由于 $x > 0$，因子 $H^2/(\pi a^2 x^2)$ 是正数，所以 $\mathrm{d}x/\mathrm{d}t$ 的正负号与 $B - Ax$ 相同。记 $c = B/A$。那么

$$\frac{\mathrm{d}x}{\mathrm{d}t} > 0 \text{ 当} x < c, \qquad \frac{\mathrm{d}x}{\mathrm{d}t} = 0 \text{ 当} x = c, \qquad \frac{\mathrm{d}x}{\mathrm{d}t} < 0 \text{ 当} x > c.$$

所以 $x = c$ 是一个平衡（equilibrium）水深，此时流入量与流出量刚好相抵，而水位总是*朝着* c 移动。这个水深只有在 $c \le H$ 时才可能实际出现在滤器之内，所以各种情况取决于 c 与 h 及 H 的大小比较。

> ? **想一想** 如果水位正朝着 c 移动，它最终会否真的到达该处，甚至越过它？

它不可能越过：要越过 c，就必须先等于 c，而在该处 $\mathrm{d}x/\mathrm{d}t = 0$。事实上，它在有限时间内永远到不了 c。把 t 看成 x 的函数，水位由 h 变到介乎 h 与 c 之间的某个水深 X 所需的时间为

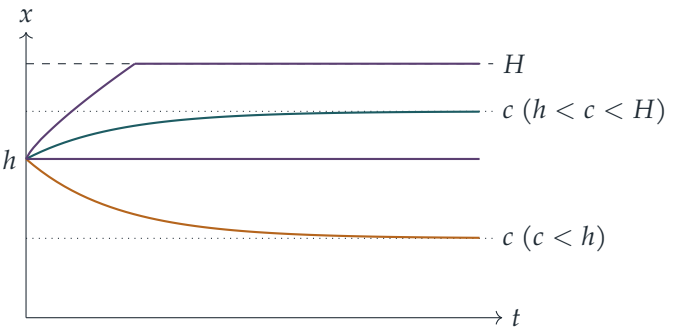
$$t = \int_h^X \frac{\pi a^2 x^2}{H^2(B-Ax)}\,\mathrm{d}x = \int_h^X \frac{\pi a^2 x^2}{AH^2(c-x)}\,\mathrm{d}x.$$

在 $x = c$ 附近，被积函数（integrand）的表现有如 $\frac{\pi a^2 c^2}{AH^2}\cdot\frac{1}{c-x}$，其积分是 $-\ln(c-x)$ 的倍数；当 $X \to c$ 时，它趋向无穷大。所以水位愈来愈慢地逼近 c，要“经过无限长的时间”才到达。无论从哪一边逼近，同样的论证都适用。

因此各种情况如下：

- $B < Ah$（即 $c < h$）。起初 $\mathrm{d}x/\mathrm{d}t < 0$：水位下降，愈降愈慢，并从上方趋向 $c = B/A$ 而不会到达。特别要注意，与第一部分不同，圆锥永远不会排空（因为 $c > 0$）。
- $B = Ah$（即 $c = h$）。这时 $\mathrm{d}x/\mathrm{d}t = 0$ 从一开始就成立：水流入的速率与排出的速率相同，水位保持在 h。
- $Ah < B < AH$（即 $h < c < H$）。水位上升，愈升愈慢，并从下方趋向 B/A。水位始终低于 H，所以滤器永远不会满溢。
- $B = AH$（即 $c = H$）。水位朝着边缘上升并趋向 H，但根据上面的论证，它永远不会真正到达，所以（在这个模型中）滤器永远不会满溢。

- $B > AH$（即 $c > H$）。现在对所有 $x \le H$ 都有 $B - Ax \ge B - AH > 0$，而且 $x^2 \le H^2$，所以
$$\frac{\mathrm{d}x}{\mathrm{d}t} \ge \frac{H^2(B - AH)}{\pi a^2 H^2} = \frac{B - AH}{\pi a^2} > 0.$$
上升速率有一个正的常数作为下界，所以水位在有限时间内到达边缘，最多需时 $\pi a^2(H - h)/(B - AH)$ 秒。此后滤器满溢：水位保持在 H，而多出的 $B - AH\,\mathrm{cm^3\,s^{-1}}$ 从顶部溢出。



草图显示了几种典型情况：下降至 c（琥珀色）、停留在 h 或在有限时间后满溢（紫红色），以及上升至 c（蓝绿色）。

验算

- 代入 $B = 0$ 得 $c = 0$：水位朝着 0 下降，与第一部分吻合（在那里水位确实到达 0，因为流出率 Ax 与 x 成正比，而体积则随 x^3 变化，所以 $\mathrm{d}x/\mathrm{d}t = -AH^2/(\pi a^2 x)$ 并不趋向 0）。
- 以数值验证：取 $a = 3$、$H = 8$、$A = 0.7$、$h = 5$，公式给出 $T \approx 7.889\,\mathrm{s}$，而按时间逐步推算 x^2 亦得到相同结果。对第二条方程作数值积分，可以确认每一种情况；例如当 $B/A = 9 > H$ 时，水位约在 41.5 s 后到达 H，而当 $B/A = 7$ 时，水位与 7 相差不超过 ε 所需的时间，在 ε 每除以 10 时便增加约 71 s，正是上文所预测的对数式增长。

常见错误

- 使用圆柱体积 $\pi a^2 x$，而不是高为 x 的相似圆锥的体积。
- 声称水位在有限时间内到达 B/A，或声称水位会越过它然后来回振荡。
- 忘记了边缘：平衡水深 B/A 只有在 $B/A \le H$ 时才会达到；否则滤器会满溢。

STEP 之外

第二部分让你初尝微分方程的定性（qualitative）理论：对于 $\mathrm{d}x/\mathrm{d}t = f(x)$，你可以找出平衡点（$f$ 的零点），根据 f 在两侧的正负号判断它们的稳定性（stability），再利用解的唯一性（uniqueness）证明解只会逼近平衡点而永远不会越过，整个过程完全不用求解任何方程。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）正是以这种精神处理自治方程（autonomous equations）与稳定性。

详细解答

STEP I 1992，第 10 题

力学 · 反冲与动量

STEP I 1992，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/92-s1-q10

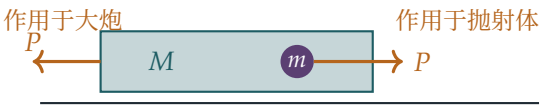
关键思路

力 P 把抛射体（projectile）向前推，但根据牛顿第三定律（Newton's third law），抛射体同时以同样大小的力 P 把大炮向后推，而水平方向上没有任何东西阻止大炮后座（recoil）。因此，抛射体在炮管内时，相对于地面走过的距离并不是 l：它是相对于大炮走过了 l。两个物体的加速度都是常数，所以最简洁的做法是考虑相对运动（relative motion）。

解答

模型

我们把抛射体和大炮都看成沿同一条水平直线运动的质点（particle），把地面看成光滑（大炮可“自由滑动”），并把抛射体与炮管之间沿炮管方向的力看成常数 P。在铅垂方向上，重量与法向反作用力（normal reaction）互相平衡（地面作用于大炮，炮管作用于抛射体），所以它们对水平运动没有影响。取发射方向为正方向。



两个加速度

抛射体在炮管内时，它所受的唯一水平力是向前的 P，而大炮所受的唯一水平力是向后的反作用力 P（牛顿第三定律；地面光滑）。根据牛顿第二定律（Newton's second law），两者的加速度为

$$a_{\text{proj}} = \frac{P}{m}, \qquad a_{\text{cannon}} = -\frac{P}{M}.$$

两者都是常数，而且两个物体都由静止开始运动。

> **想一想** 抛射体在多长的距离内加速：是 l，还是别的距离？

炮管的长度是 l，但炮管本身在向后移动。当抛射体相对于大炮的位移（displacement）达到 l 时，它便离开炮管。相对于大炮，抛射体由静止开始运动，其加速度为常数

$$a_{\text{rel}} = \frac{P}{m} - \left(-\frac{P}{M}\right) = \frac{P}{m} + \frac{P}{M} = \frac{P(M+m)}{mM}.$$

在炮管内的时间

对相对运动使用 $s = \frac{1}{2}at^2$，抛射体在时刻 T 离开炮管，其中

$$l = \frac{1}{2} \cdot \frac{P(M+m)}{mM} T^2, \qquad \text{所以} \qquad T^2 = \frac{2mMl}{P(M+m)}.$$

出口速率

相对于地面，抛射体由静止开始以恒加速度 P/m 运动，所以在时刻 T 它的速率 v 满足

$$\begin{aligned} v^2 = \left(\frac{P}{m}T\right)^2 &= \frac{P^2}{m^2} \cdot \frac{2mMl}{P(M+m)} \\ &= \frac{2PMl}{m(M+m)}. \end{aligned}$$

由于 $v > 0$，

$$v = \sqrt{\frac{2PMl}{m(M+m)}},$$

这正是所求的结果。

另一途径：能量与动量

利用守恒定律（conservation laws）也可得到同一答案。水平方向上没有外力，所以水平动量（momentum）守恒：若抛射体离开炮管的一刻大炮向后的速率为 V，则 $mv = MV$。若抛射体在炮管内的期间，抛射体向前移动了 x，大炮向后移动了 X，则 $x + X = l$。作用于抛射体的力所作的功（work）为 Px，而作用于大炮的反作用力与大炮的运动方向相同，所作的功为 PX；两者合共作功 $P(x+X) = Pl$。因此

$$Pl = \tfrac{1}{2}mv^2 + \tfrac{1}{2}MV^2 = \tfrac{1}{2}mv^2 + \tfrac{1}{2}M\left(\frac{mv}{M}\right)^2 = \tfrac{1}{2}mv^2\,\frac{M+m}{M},$$

由此同样得到 $v^2 = \dfrac{2PMl}{m(M+m)}$。

验算

- 重型大炮：当 $M \to \infty$ 时，因子 $M/(M+m) \to 1$，而 $v \to \sqrt{2Pl/m}$，即大炮固定不动时的速率（所作的功 Pl 全部归于抛射体）。大炮后座时，所得的速率总会略小一些，这是理所当然的，因为部分能量用于后座。
- 单位：Pl 是能量，而 Pl/m 的单位是 $\mathrm{m^2\,s^{-2}}$，所以 v 是速率。
- 数值：取 $m = 0.1$、$M = 0.9$、$P = 9$ 及 $l = 0.5$（国际单位制，SI units），公式给出 $v = 9\ \mathrm{m\,s^{-1}}$，而大炮以 $1\ \mathrm{m\,s^{-1}}$ 后座。于是动量为 $0.1 \times 9 = 0.9 \times 1$，动能（kinetic energy）为 $4.05 + 0.45 = 4.5 = Pl$。对两个物体作逐步数值模拟，在两者的相对位移达到 l 时停止，所得结果在 m、M、P 和 l 取多组随机数值时均与公式吻合。

常见错误

- 使用 $v^2 = 2(P/m)l$，即把 l 当作抛射体相对于地面走过的距离，忽略了后座。
- 忘记牛顿第三定律，让大炮保持静止；大炮可自由滑动，所以必定后座。
- 使用能量方法时，只把所作的功取为 P 乘以抛射体相对于地面的位移；其实作用于大炮的反作用力也有作功。

STEP 之外

这是一个二体问题（two-body problem），而相对加速度 $P(M+m)/(mM)$ 正是 P 除以约化质量（reduced mass）$\mu = mM/(m+M)$。任何两个只与对方相互作用的物体，彼此之间的相对运动都如同一个质量为 μ 的单一物体；同一想法可用来处理行星轨道、双星和粒子碰撞，也是火箭问题和后座问题的基础。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教科书）对这里的约化质量和动量两方面的想法都有详细阐述。

详细解答

STEP I 1992，第 11 题

力学 · 弹性绳的平衡

STEP I 1992，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/92-s1-q11

关键思路

每个质点都处于平衡（equilibrium）状态，所以紧接其上方的张力比紧接其下方的张力大 mg。这就把三个张力联系起来，而虎克定律（Hooke's law）则把张力化为伸长量。我们还需要的另一项事实是几何上的：三条拉长后的弦线，长度之和必须等于房间的高度。一旦求得各高度，“$CD = a$”只不过是说最低的那条弦线没有张力。

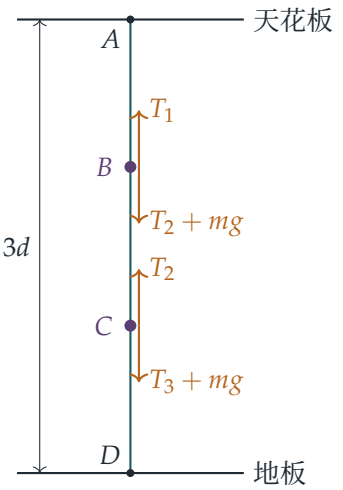
解答

模型

各弦线为轻质弹性弦线（light elastic string），拉长时遵从虎克定律（张力 $= \lambda \times$ 伸长量 $/a$）；各质点静止不动。设 T_1、T_2、T_3 为 AB、BC、CD 的张力（tension），x_1、x_2、x_3 为它们的伸长量（extension），于是 $T_i = \lambda x_i / a$。为方便起见，我们给单靠重量 mg 所产生的伸长量起一个名字：

$$e = \frac{mga}{\lambda}.$$

我们先假设三条弦线全部拉紧（taut），稍后再检查这在甚么情况下成立。设 y_B、y_C 为由地板量起的高度。



B 和 C 的高度

对每个质点沿铅垂方向分解力（上方的弦线向上拉，下方的弦线和重量向下拉）：

$$B: \quad T_1 = T_2 + mg, \qquad C: \quad T_2 = T_3 + mg.$$

两式除以 λ/a，便成为关于伸长量的关系式：

$$x_1 = x_2 + e, \qquad x_2 = x_3 + e, \qquad \text{所以} \qquad x_2 = x_3 + e, \quad x_1 = x_3 + 2e.$$

三条弦线的总长度就是房间的高度：

$$3a + x_1 + x_2 + x_3 = 3d \quad \Longrightarrow \quad 3x_3 + 3e = 3d - 3a \quad \Longrightarrow \quad x_3 = d - a - e.$$

于是 $x_2 = d - a$，$x_1 = d - a + e$。由下方各弦线的长度即可求得高度：

$$y_C = CD = a + x_3 = d - e = d - \frac{mga}{\lambda},$$
$$y_B = y_C + BC = (d - e) + (a + d - a) = 2d - \frac{mga}{\lambda}.$$

用 AB 验算一下：$3d - y_B = d + e = a + x_1$，正如所料。

想一想 这些答案究竟在甚么情况下才成立?

我们假设了每条弦线都拉紧，所以每个伸长量都必须至少为零。最严格的条件落在最低的弦线上：$x_3 \geq 0$，即

$$\frac{mga}{\lambda} \leq d - a.$$

（这时 $x_2 = d - a \geq 0$ 和 $x_1 > 0$ 自动成立。）CD 最先松弛（slack）是合理的：它不用支撑任何东西，只是靠房间把弦线拉开才保持拉紧。若 $mga/\lambda > d - a$，弦线 CD 便会松弛，

整个系统只不过是两个质点悬挂在 A 之下：这时 $T_2 = mg$，$T_1 = 2mg$，所以 $AB = a + 2e$，$BC = a + e$，从而 $y_B = 3d - a - 2e$，$y_C = 3d - 2a - 3e$，前提是这个位置仍在地板之上。题目所指的是所有弦线都拉紧的主要情况。

何时可以有 $CD = a$?

弦线 CD 的长度等于其自然长度（natural length），当且仅当它刚好拉紧而没有张力，即 $T_3 = 0$，亦即 $x_3 = 0$。这正是上述两种情况的分界，而用任何一组公式都可得

$$y_C = a \iff d - \frac{mga}{\lambda} = a \iff m = \frac{\lambda(d-a)}{ga}.$$

从物理上看，这时位于 C 的质点完全由 BC 支撑，$T_2 = mg$，而 $T_1 = 2mg$。

合适的质量存在，当且仅当这个 m 的值为正数，这需要 $d > a$。反过来说，若 $d \le a$，则没有任何正数 m 可行：假如所有弦线都拉紧，我们便需要 $x_2 = d - a \ge 0$ 和 $x_3 = d - a - e < 0$，这是不可能的，所以 CD 松弛，C 悬于高度 $3d - 2a - 3e < 3a - 2a = a$（或停在地板上），因此 $CD < a$。故所求的值的集合为

$$\{\, d : d > a \,\}.$$

对这样的 d，我们得到 $y_C = a$，$y_B = d + a$ 和 $AB = 2d - a$。

验算

- 单位：mga/λ 是力乘以长度再除以力，所以是长度；它要从 d 中减去，理应如此。
- 当 $m = 0$ 时，三条弦线平均分担伸长：每条的长度都是 d，C 位于 d，B 位于 $2d$。加上质量后，两个质点都下降相同的距离 e，因为 BC 的伸长量保持为 $d - a$。
- 数值上，对 250 组随机选取的 a、λ、d、m，在 $0 \le y_C \le y_B \le 3d$ 的范围内求位能（potential energy）的最小值（重力位能加弹性位能，并容许弦线松弛），所得高度在拉紧和 CD 松弛两种情况下都与上述结果吻合，误差在 10^{-14} 以内。取 $a = 1$、$\lambda = 10$、$d = 2$ 和 $m = \lambda(d-a)/(ga) \approx 1.020$，得出 $CD = 1.000000$ 和 $y_B = 3.000000$；当 $d \le a$ 时，没有任何正数 m 能使 $CD = a$。

常见错误

- 把张力写成 $\lambda\times$（拉长后的长度）$/a$，而不是 $\lambda\times$ 伸长量 $/a$。
- 忘记总长度由房间决定：没有 $\sum(a + x_i) = 3d$，单靠平衡方程并不能确定各伸长量。
- 假设每条弦线都必定有张力。弦线是可以松弛的，而 $CD = a$ 恰好发生在松弛的分界上。

STEP 之外

这是一个小型的静不定结构（statically indeterminate structure）：单靠质点的平衡并不能确定各张力，我们还需要一个相容条件（compatibility condition，即各长度之和固定）。同一个想法写成矩阵形式，就是工程师分析由弹簧和杆件组成的构架的方法。另一个切入点是求总位能的最小值，上面的数值验算正是这样做的；这种变分（variational）观点在往后的力学中占有核心地位。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）对弹簧、平衡和位能有仔细的讲解。

详细解答

STEP I 1992，第 12 题

力学 · 涉及摩擦力的静力学

STEP I 1992，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/92-s1-q12

关键思路

两把梯子都是轻质的，所以要为其中一把梯子列出有用的方程，唯一的办法就是对铰链（hinge）A 取力矩（moments）：这样可以消去另一把梯子所施的未知力，以及 A 处负重所施的力。梯脚有向外滑开的趋势，而这股向外的推力，一部分由摩擦力（friction）抵消，一部分由木杆抵消。在梯子倒塌的一刻，摩擦力已经用尽，余下的就要由木杆承担。

解答

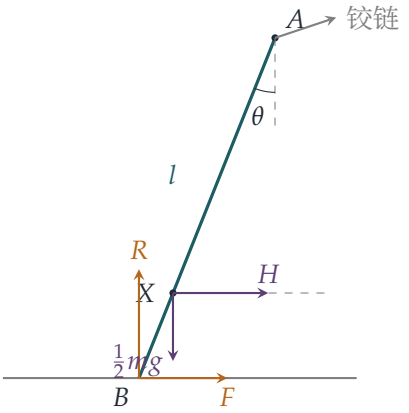
模型

两把梯子是轻杆，在 A 处以光滑铰链连接。木杆 XY 是均匀的，质量为 m，并在 X 和 Y 处以销钉接合（pin-jointed）。质量 M 视为位于 A 的质点（particle）。设 $t = \tan\theta$。由于整个装置左右对称，两只梯脚所受的法向反作用力（normal reaction）相等，设为 R；对整座折梯沿铅垂方向分解力，可得

$$2R = (M+m)g \quad \Longrightarrow \quad R = \tfrac{1}{2}(M+m)g.$$

由于梯子向外倾斜，梯脚有向外滑开的趋势，所以每只梯脚所受的摩擦力 F 都指向内。

木杆所施的力。 木杆是水平的，而且本身有重量，所以它两端所受的力并不沿着木杆的方向。根据对称性，两端各承受木杆重量的一半，所以梯子在 X 处对木杆施加向上的力 $\frac{1}{2}mg$，同时施加一个向外的水平力 H（即张力（tension））。根据牛顿第三定律（Newton's third law），木杆在 X 处对梯子 AB 施加向内的水平力 H 和向下的铅垂力 $\frac{1}{2}mg$。



对梯子 AB 以 A 为支点取力矩

从 A 量起，梯脚 B 的水平距离为 $l\sin\theta$，铅垂距离为向下 $l\cos\theta$；而 X 的水平距离为 $\frac{3}{4}l\sin\theta$，铅垂距离为向下 $\frac{3}{4}l\cos\theta$。对 A 取力矩（作用于 A 的力全部没有力矩）：

- R 使梯子向一个方向转动，力矩为 $Rl\sin\theta$；
- F 和 H（两者都指向内，作用点在 A 的下方）使梯子向另一个方向转动，力矩分别为 $Fl\cos\theta$ 和 $H\cdot\frac{3}{4}l\cos\theta$；
- 作用于 X 的向下力 $\frac{1}{2}mg$ 与 R 在同一侧，但方向相反，其力矩 $\frac{1}{2}mg\cdot\frac{3}{4}l\sin\theta$ 与 R 的力矩相抗。

因此

$$Fl\cos\theta+\tfrac{3}{4}Hl\cos\theta=Rl\sin\theta-\tfrac{3}{8}mgl\sin\theta,$$

两边除以 $l\cos\theta$，再代入 R 的值，得

$$F+\tfrac{3}{4}H=t\left(\tfrac{1}{2}(M+m)g-\tfrac{3}{8}mg\right)=t\left(\tfrac{1}{2}M+\tfrac{1}{8}m\right)g. \qquad (*)$$

? **想一想** 摩擦力和木杆都在把梯脚向内拉住。那么 F 应该取哪个值?

方程 $(*)$ 表明，摩擦力和张力要合力分担一个固定的总量：摩擦力承担得愈多，木杆需要承担的就愈少。所以，只要有某个容许的摩擦力（$F\le\mu R$）能令 $H\le T$，折梯就不会倒塌；而最小的张力出现在摩擦力最大的时候，即 $F=\mu R$。这正是题目容许我们假设的、倒塌一刻的极限状态（limiting state）。把 $F=\mu R=\frac{1}{2}\mu(M+m)g$ 代入 $(*)$：

$$\begin{aligned}\tfrac{3}{4}H&=t\left(\tfrac{1}{2}M+\tfrac{1}{8}m\right)g-\tfrac{1}{2}\mu(M+m)g\\&=\tfrac{1}{2}(t-\mu)Mg+\tfrac{1}{8}(t-4\mu)mg,\end{aligned}$$

所以

$$H = \tfrac{2}{3}(t-\mu)Mg + \tfrac{1}{6}(t-4\mu)mg.$$

由于 $t > \mu$，H 随 M 增加而增加：假如没有木杆，单靠摩擦力连负重本身也支撑不住（当 $m = 0$ 时，我们需要 $t \le \mu$）。

最大的安全质量

只要 $H \le T$，木杆就不会折断，即

$$\tfrac{2}{3}(t-\mu)Mg \le T - \tfrac{1}{6}(t-4\mu)mg,$$

两边除以正数 $\tfrac{2}{3}(t-\mu)g$，得

$$\boxed{M \le \frac{6T - (\tan\theta - 4\mu)mg}{4(\tan\theta - \mu)g}}$$

所以，可以安全地放在 A 处的最大质量是 $\dfrac{6T - (\tan\theta - 4\mu)mg}{4(\tan\theta - \mu)g}$。

因自身重量而损毁

当 A 处没有放任何东西（$M = 0$）时，所需的张力为 $H = \tfrac{1}{6}(t-4\mu)mg$，所以木杆会在以下情况折断：

$$\tfrac{1}{6}(\tan\theta - 4\mu)mg > T \iff \tan\theta > \frac{6T}{mg} + 4\mu,$$

这正是题目要求证明的结果。在这个情况下，上面 M 的上界是负数：任何负重都不安全，因为折梯连自身也支撑不住。

验算

- 如果 T 极大，安全质量也极大；如果 $\tan\theta$ 只是略大于 μ，分母就极小，同样几乎任何负重都安全，因为单靠摩擦力已差不多足够。
- 我们对完整的静力学（statics）问题（两把梯子、木杆和 A 处的销钉，共十二个未知的分力，两只梯脚的摩擦力均达极限值）作数值求解，随机选取 2000 组配置，所得结果与张力公式吻合，误差约为 10^{-12}。例如，当 $\theta = 45°$、$\mu = 0.2$、$m = 2\,\mathrm{kg}$ 及 $g = 9.8\,\mathrm{m\,s^{-2}}$ 时，$M = 0$ 的木杆张力为 $\tfrac{1}{6}(0.2)(19.6) \approx 0.653\,\mathrm{N}$。

常见错误

- 把木杆当作轻质拉杆，以为 X 处的力完全沿着 XY 的方向。木杆是有重量的，所以每一端还要在铅垂方向承受 $\tfrac{1}{2}mg$，而 $\tfrac{1}{8}m$ 和 4μ 正是由此而来。

- 把梯脚的摩擦力画成向外。由于 $\tan\theta > \mu$，梯脚有向外张开的趋势，所以摩擦力指向内，与木杆一同把梯脚拉住。

STEP 之外

像这样的框架，单凭平衡（equilibrium）方程并不足以确定每一个力（在这里，摩擦力和张力可以此消彼长），称为静不定（statically indeterminate）。工程师会借助额外的资料来处理这类情况：例如本题所用的极限条件，或者构件的弹性（elasticity）。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教科书）是加深理解力、力矩（torque）和平衡的好去处。

详细解答

STEP I 1992，第 13 题

力学 • 逃逸速率与径向下落

STEP I 1992，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/92-s1-q13

关键思路

力只取决于距离 x，所以最自然的工具是把运动方程（equation of motion）写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$ 的形式，一经积分便立即得到联系速率与距离的能量方程（energy equation）。在第一部分，v^2 永远不可能是负数这一事实，把该方程变成 x 的一个上限。在第二部分，能量方程把 $\frac{\mathrm{d}x}{\mathrm{d}t}$ 表示为 x 的函数，再用分离变量法（separating the variables）便可求出时间。

解答

我们把太阳当作固定不动，把彗星当作质点（particle），除了题目给出的引力外不受其他力。由于引力沿着连接彗星与太阳的直线作用，一颗正在直接远离太阳（或处于静止）的彗星会一直沿着该直线运动。因此 x 是一条直线上的坐标，我们设 $v=\frac{\mathrm{d}x}{\mathrm{d}t}$，并以 v 为正表示远离太阳。



能量方程

沿该直线应用牛顿第二定律（Newton's second law），力指向太阳（即 x 减少的方向），得

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}=-\frac{Gm}{x^2}.$$

把加速度写成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$ 并约去 m,

$$v\frac{\mathrm{d}v}{\mathrm{d}x} = -\frac{G}{x^2} \qquad \Longrightarrow \qquad \tfrac{1}{2}v^2 = \frac{G}{x} + C$$

其中 C 为某个常数。无论彗星朝哪个方向运动，这条式在整个运动过程中都成立。（这就是每单位质量的能量守恒（conservation of energy）：动能（kinetic energy）$\frac{1}{2}v^2$ 加上势能（potential energy）$-G/x$。）

x 可变得任意大的条件

在给定的一刻 $x = h$ 而 $v = V$，所以 $C = \frac{1}{2}V^2 - \frac{G}{h}$，而在其后任何时刻，

$$v^2 = V^2 - \frac{2G}{h} + \frac{2G}{x}.$$

我们证明其逆否命题（contrapositive）：若 $V^2 < 2G/h$，则 x 有界（bounded）。

想一想 原则上，彗星往后可以有各式各样的运动方式。单凭一条方程，怎能在所有时刻都限制住 x?

因为无论彗星在做甚么，每一瞬间都有 $v^2 \geq 0$。假设 $V^2 < 2G/h$，并设 $k = \dfrac{2G}{h} - V^2 > 0$。那么能量方程说明

$$\frac{2G}{x} = v^2 + k \;\geq\; k, \qquad \text{所以} \qquad x \leq \frac{2G}{k} = \frac{2Gh}{2G - hV^2}$$

在其后任何时刻都成立（由于 $x > 0$ 且 $k > 0$，我们可以把两边乘开）。所以 x 永远不能超过固定值 $2Gh/(2G - hV^2)$：彗星上升到该距离，在那里 $v = 0$，然后往回落。因此，除非 $V^2 \geq 2G/h$，否则 x 不可能变得任意大，这正是所要证明的。

（逆命题（converse）也成立。若 $V^2 \geq 2G/h$，则彗星向外运动时 $v^2 \geq 2G/x > 0$，所以它永不停下；又因为只要 $x \leq X$ 便有 $v^2 \geq 2G/x \geq 2G/X$，它不可能永远停留在任何距离 X 以内。这就是逃逸速率（escape speed）。）

坠入太阳所需的时间

现在彗星在极远的距离 D 处由静止开始运动。由静止释放后，它笔直地坠向太阳。由能量方程，以及在 $x = D$ 处 $v = 0$,

$$\tfrac{1}{2}v^2 = \frac{G}{x} - \frac{G}{D}.$$

想一想 “极远的距离”对那个常数来说意味着甚么?

这表示我们令 $D \to \infty$，于是 $G/D \to 0$，常数为零：彗星的总能量为零。（若 D 很大但有限，彗星只会稍为慢一点，所需时间稍为长一点，而相差的量当 $D \to \infty$ 时趋向零。）所以在距离 x 处它的速率是 $\sqrt{2G/x}$，而且由于 x 正在减少，

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\sqrt{\frac{2G}{x}} \qquad \Longrightarrow \qquad \sqrt{x}\,\mathrm{d}x = -\sqrt{2G}\,\mathrm{d}t.$$

设 T 为由 $x = h$ 到 $x = 0$ 所需的时间。积分得

$$\sqrt{2G}\,T = \int_0^h \sqrt{x}\,\mathrm{d}x = \tfrac{2}{3}h^{3/2},$$

所以

$$T = \frac{2h^{3/2}}{3\sqrt{2G}} = \frac{h}{3}\sqrt{\frac{2h}{G}}.$$

（被积函数（integrand）$\sqrt{x}$ 在 $x = 0$ 处完全没有问题，所以尽管彗星的速率无止境地增大，它确实在有限时间内到达太阳。）

验算

- 单位：力 Gm/x^2 是质量乘以加速度，所以 G 的单位是 $\mathrm{m}^3\,\mathrm{s}^{-2}$。于是 V^2 和 $2G/h$ 的单位都是 $\mathrm{m}^2\,\mathrm{s}^{-2}$，而 $h^{3/2}/\sqrt{G}$ 的单位是秒。
- 在 $x = h$ 处彗星的速率是 $\sqrt{2G/h}$。假如它保持这个速率，所需时间便是 $h/\sqrt{2G/h} = \frac{h}{2}\sqrt{2h/G}$，这比 T 长；彗星一直在加速，所以理应如此。事实上 T 是这个时间的三分之二。
- 数值验证，取 $G = h = 1$：由 $x = 1$、$v = \sqrt{1.5}$ 出发逐步推算运动，彗星在 $x = 4.000$ 处掉头，与 $2Gh/(2G - hV^2) = 4$ 的预测相符。由 $x = 1$、$v = -\sqrt{2}$ 出发，它经过 0.4714 后到达太阳，与 $T = \frac{1}{3}\sqrt{2} \approx 0.47140$ 吻合。若改为由 $D = 10^6$ 而非无穷远处从静止出发，T 只在小数点后第七位才有变化。

常见错误

- 漏掉 $\frac{\mathrm{d}x}{\mathrm{d}t} = -\sqrt{2G/x}$ 中的负号，结果得到负的时间，或者由 0 积分到 h 却没有说明原因。
- 只证明彗星会减速：每一颗向外运动的彗星都会减速，即使是能够逃逸的那些也一样。重点在于，当 $V^2 < 2G/h$ 时，v^2 在超过某个距离后便会变成负数。

STEP 之外

$\frac{1}{2}v^2 - G/x$ 这个量是牛顿引力场（Newtonian gravitational field）中每单位质量的能量，它的正负决定了轨道是束缚轨道（椭圆）还是逃逸轨道（抛物线和双曲线）。最后一部分的零能量坠落，是抛物线轨道退化成直线的特殊情况，而 $T \propto h^{3/2}$ 则与开普勒第三定律（Kepler's third law）一脉相承。大学一年级有关轨道的课程，正是由同样的能量方程和角动量（angular momentum）方程出发，推导出这一切；Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对有心力（central forces）和轨道有透彻的论述。

详细解答

STEP I 1992，第 14 题

概率与统计 • 统计论断的诠释

STEP I 1992，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/92-s1-q14

本题在 *Beyond STEP* 第 13.1 节解读新闻标题中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch13/headlines.html

关键思路

每则标题都是拿着一个数字，就一跃跳到结论。对每一则标题，都要问两个问题。第一，它是否在拿同类的东西作比较：同一样事物，用同一种方法量度，并按承受风险的程度（exposure）作出调整？第二，这个差异会不会纯属偶然？对于第二个问题，一把很自然的尺是泊松模型（Poisson model），它描述罕有而互相独立的事件的发生次数，其标准差（standard deviation）等于平均值的平方根。懂得这个模型何时适用、何时不适用，正是 (ii) 至 (iv) 的核心。

解答

以下所说的“平均”，一律指 1974–1989 年的年平均数：每年 1080 名行人死亡，180 人死于空难。

第 (i) 部分

“六倍”之说来自比率 $1080/180 = 6$，但这个比率比较的是原始总数，而不是风险。风险是指每单位风险承担量的死亡人数：按每程、每小时或每公里计算。在 Poldavia，几乎人人每天都在马路附近步行，而乘飞机的人相对很少，而且只是偶一为之。要比较两者

的安全程度，我们需要的是类似

$$\frac{\text{行人死亡人数}}{\text{步行的总时数（或总公里数）}} \quad \text{对比} \quad \frac{\text{空难死亡人数}}{\text{飞行的总时数（或总公里数）}},$$

这样的比较，而标题却没有给出任何一个分母。视乎所用的量度方法，答案甚至可能倒转过来：按每公里计，飞行看来非常安全；但按每小时或每程计，就未必那么安全了。

此外还有其他问题。这里所谓的“步行”，其实是指徒步时被车辆撞倒，这个数字完全没有反映行人在僻静地方的情况。空难死亡人数除了乘客，还包括机组人员。而且一段时期的平均数根本无法“证明”1991 年的安全情况；它们充其量只是关于 1974–1989 年的证据。

第 (ii) 部分

乍看之下，125 远低于平均数 180。这个跌幅是否超出了偶然的范围？

假如死亡事件是互相独立、逐宗发生的，那么采用平均值为 180 的泊松模型便算合理。它的标准差是 $\sqrt{180} \approx 13.4$，因此 125 大约比平均值低 4.1 个标准差，纯粹出于偶然的可能性极低（$P(X \leq 125) \approx 10^{-5}$）。

但空难死亡并不是逐宗发生的，而是成群出现：一次坠机就可以夺去数十以至数百人的性命。较自然的模型是：*致命坠机*的次数服从泊松分布，而每次坠机的死亡人数是随机的。这样一来，变异程度便大大增加。举例说，假设平均每年有 6 宗致命坠机，每宗有 30 人死亡。那么全年死亡人数是 $30N$，其中 $N \sim \text{Po}(6)$，其标准差为 $30\sqrt{6} \approx 73$，而

$$P(\text{死亡人数} \leq 125) = P(N \leq 4) \approx 0.29.$$

按这些（仅作说明用的）数字，一年有 125 人死亡实在平常不过：只需要比平常少一两宗坠机便会出现。如果连每宗坠机的死亡人数也容许变动，离散程度还会更大。

所以单凭一年的数字，我们所知甚少。我们还想知道 1990 年的飞行量有多少（如果航空交通减少了，死亡人数本来就应该较少），以及好几年的数字，才能看出是否真的有趋势。标题的说法并无根据。

第 (iii) 部分

这里有两个不同的毛病。

单凭一天不能显示上升。“上升”是关于一段时间内趋势的说法，而单独一天的数字根本无从比较。我们需要一段长时间内每天（或每年）的数字。

*七人根本不算异常。*我们手上唯一的道路数字是行人的：每年 1080 人，即

$$\lambda = \frac{1080}{365} \approx 2.96\,(\text{每天}).$$

用泊松分布来模拟行人死亡相当合适：这类事件罕有，大多是每宗意外一人死亡，而且大致互相独立。设 $X \sim \mathrm{Po}(2.96)$ 为某一天死亡的行人数目。那么

$$P(X \geq 7) = 1 - \sum_{k=0}^{6} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \approx 0.031,$$

所以我们预期一年之中大约有 $365 \times 0.031 \approx 11$ 天，会有最少 7 名行人死亡。况且标题说的是“死亡人数”，那还包括司机、乘客和骑单车的人。每天道路死亡总人数的平均率必定高于 2.96，于是 7 就更加不足为奇了。

想一想 某一天的死亡人数是平均数的两倍有多，难道不令人震惊吗？

数目小的时候并不然。当平均值约为 3 时，泊松标准差是 $\sqrt{3} \approx 1.7$，与平均值相比已经很大，所以每天之间按比例计出现大幅波动是正常的。专挑最差的一天，也是一种选择效应（selection effect）：一年有 365 天，报章总能找到差的一天。那天是星期几也有关系，因为交通流量和外出步行的人数会随星期几和季节而变化。速率恒定的泊松模型只是一个粗略的初步近似。

第 (iv) 部分

这一则标题确实有值得认真回应之处。先看统计，再谈诠释。

*1350 是否超出了偶然的范围？*基于 (iii) 所述的理由，泊松模型在这里是个不错的起点。平均值为 1080 时，标准差是 $\sqrt{1080} \approx 32.9$，而

$$\frac{1350 - 1080}{32.9} \approx 8.2.$$

在这个模型下，一个比平均值高出 8 个标准差的数值，其概率的数量级为 10^{-15}。即使实际上每年之间的离散程度是泊松值的两三倍（比方说因为严冬），相对于 1080 这个稳定的速率，1350 仍然非常异常。所以极有可能有些东西改变了。问题是：改变的是甚么？

*与甚么比较？*标题是拿 1990 年与一个 16 年的平均数比较，而不是与 1989 年比较。假如行人死亡人数在这段时期内一直稳步上升，那么平均数 1080 描述的就是这段时期的中段（约 1981–82 年），而 1990 年理应远高于这个数。例如，由 1974 年约 855 人起，每年稳定增加 30 人，平均数便是 1080，而 1990 年则约为 1335 人。在这种情况下，1990 年根本算不上甚么“急升”，只是长期趋势的延续。这个趋势或许值得上头条，但应该是另一则标题。我们需要逐年的数字，才能分辨这两种说法。

*风险承担量有没有改变？*人口增加、车辆增加或步行增加，都会令死亡人数上升，却不代表按每程计步行变得更危险。公平的比较应该用比率，例如每 100 000 人口的死亡人数，或每步行一公里的死亡人数。

*量度方法有没有改变？*死亡个案的记录或分类方法一旦改变，数字便会出现表面上的跳

升。例子包括：意外发生后多久之内死亡仍算作道路死亡的准则有变、谁算是行人的定义有变，或者呈报情况有所改善。任何这类改变都必须先予排除。

用词是否恰当？“伤亡”通常是指死亡*或受伤*的人。所引述的数字只计死亡，因此标题夸大了这个数字所能说明的事。受伤的人数多得多，而且可能呈现截然不同的规律。

总括而言：升至 1350 的幅度实在太大，不能用围绕恒定速率 1080 的泊松波动来解释。但在称之为“令人震惊”之前，我们需要近年的年度数字（以区分突然上升与长期趋势），以及人口和交通数字（以把数目化为比率）。我们还须查证各项定义没有改变。

验算

各个概率已通过精确地求泊松级数之和，以及进行 2×10^5 次试验的模拟来核对。在 (iii) 中，当 $\lambda = 1080/365$ 时，$P(X \geq 7) = 0.0315$（模拟结果：0.0317）。在 (ii) 中，若每年有 6 宗坠机、每宗 30 人死亡，则 $P(N \leq 4) = 0.285$。若让每宗坠机的死亡人数在 1 至 59 之间均匀变动，模拟所得的标准差为 84，而 $P(\text{死亡人数} \leq 125) \approx 0.28$。对泊松计数有一条简便的经验法则：与平均值相差约两个平均值平方根以内的数值，都不足为奇。就一天而言，$3 + 2 \times 1.7 \approx 6.4$，所以 7 只是刚刚超出这个范围，而且有 365 天可供挑选，这样的日子必然会出现。就一年而言，$1080 + 2 \times 33 = 1146$，与 1350 相距甚远。

常见错误

- 在 (i) 中比较总数而不是比率：风险永远需要一个分母。
- 在 (ii) 中使用泊松标准差 $\sqrt{180}$。空难中的死亡是成群出现的，所以离散程度大得多。
- 在 (iv) 中说到“8 个标准差，所以是真正的上升”便停下来，而没有追问：与 16 年平均数比较这一点、人口或定义的变化，能否解释这个升幅。

STEP 之外

第 (ii) 部分谈的是*过度离散*（overdispersion）。一个由泊松数目的群组合成、而每个群组大小随机的总数，是一个*复合泊松*（compound Poisson）变量，它的方差（variance）是 $\lambda\,\mathrm{E}[S^2]$，而不是 λ。第 (iv) 部分则通往假设检定（hypothesis testing），以及一个困难得多的问题：在统计上显著的变化，究竟意味着甚么。两者都会在大学概率与统计课程的初段出现。Blitzstein 与 Hwang 合著的 *Introduction to Probability*（概率论入门教材），是阅读泊松模型及其局限的好去处。

详细解答

STEP I 1992，第 15 题

概率与统计 · 期望等候时间

STEP I 1992，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/92-s1-q15

关键思路

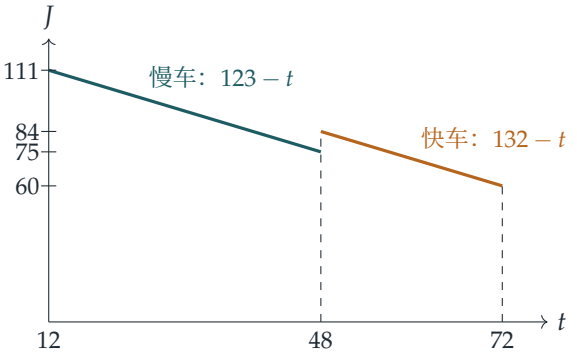
一切每小时重复一次，所以只需考虑一名乘客，其到达时间在一小时内均匀分布（uniformly distributed）。这一小时分成两段间隔：在某班列车之前那段间隔内到达的乘客，全部都会乘搭该班列车，而每人的行程时间就是“等候时间加行车时间”。在同一段间隔内，等候时间是均匀分布的，所以其平均值是间隔长度的一半。这样，整条题目就化为一个以其中一段间隔的长度为变量的二次式（quadratic）。

解答

现行时间表

以正午之后的分钟数量度时间 t（任何一个小时都可以），并设乘客的到达时间 T 在长一小时的区间 $(12, 72]$ 上均匀分布。这段时间内的列车有：在 48 开出的慢车，于 $48 + 75 = 123$ 抵达伦敦；以及在 72（即下一个小时的 12 分）开出的快车，于 $72 + 60 = 132$ 抵达。

- 若 $12 < T \leq 48$，最先可乘搭的是慢车，所以行程时间为 $J = 123 - T$。
- 若 $48 < T \leq 72$，最先可乘搭的是快车，所以 $J = 132 - T$。



由于 T 在 $(12,72]$ 上的概率密度（density）为 $\frac{1}{60}$，

$$\mathrm{E}(J)=\frac{1}{60}\left[\int_{12}^{48}(123-t)\,\mathrm{d}t+\int_{48}^{72}(132-t)\,\mathrm{d}t\right].$$

两个被积函数（integrand）都是线性的，所以每个积分都等于区间长度乘以函数在区间中点的值：

$$\int_{12}^{48}(123-t)\,\mathrm{d}t=36\times(123-30)=3348,\qquad \int_{48}^{72}(132-t)\,\mathrm{d}t=24\times(132-60)=1728.$$

因此

$$\mathrm{E}(J)=\frac{3348+1728}{60}=\frac{5076}{60}=84.6\text{ 分钟},$$

正如所求。

重新编排快车的开出时间

现在设快车在每小时的 x 分开出，慢车则仍在 48 分开出。设 d 为由一班快车开出至下一班慢车开出所相隔的分钟数，所以 $0\le d<60$ 且 $d\equiv 48-x\pmod{60}$。（在原来的时间表中，$d=36$。）于是这一小时分成

- 每班慢车之前一段长度为 d 的间隔：这些乘客的等候时间 W 在 $[0,d)$ 上均匀分布，然后行车 75 分钟；
- 每班快车之前一段长度为 $60-d$ 的间隔：这些乘客的等候时间在 $[0,60-d)$ 上均匀分布，然后行车 60 分钟。

乘客落在第一段间隔的概率是 $\frac{d}{60}$，落在第二段的概率是 $\frac{60-d}{60}$，而在长度为 g 的间隔上均匀分布的等候时间，其平均值为 $\frac{g}{2}$。所以，按乘客在哪一段间隔到达来取条件（conditioning），

可得

$$\begin{aligned}60\,\mathrm{E}(J) &= d\left(\frac{d}{2}+75\right)+(60-d)\left(\frac{60-d}{2}+60\right)\\ &= \frac{d^2}{2}+75d+\frac{3600-120d+d^2}{2}+3600-60d\\ &= d^2-45d+5400.\end{aligned}$$

作为验算，$d=36$ 给出 $1296-1620+5400=5076$，与第一部分吻合。

想一想 为甚么最佳的间隔不是 $d=30$，即把一小时平均分成两半？

在慢车那段间隔内到达的每名乘客，都要因为行车较慢而多付 15 分钟的代价。所以我们希望那段间隔内的乘客少一些，也就是要把它缩短至少于半小时，但又不能短得令快车那段间隔（以及那些乘客的等候时间）变得太长。这个二次式正好平衡了这两种效应。

配方（completing the square）后，

$$d^2-45d+5400=\left(d-\tfrac{45}{2}\right)^2+5400-\tfrac{2025}{4},$$

它在 $d=22.5$ 时取得最小值，而这个值在 $[0,60)$ 之内。此时 $60\,\mathrm{E}(J)=\frac{19575}{4}$，所以平均行程时间的最小值为

$$\mathrm{E}(J)=\frac{19575}{240}=\frac{1305}{16}=81\tfrac{9}{16}\approx 81.56 \text{ 分钟}.$$

快车必须比慢车早 22.5 分钟开出，所以

$$x=48-22.5=25.5,$$

即快车应在每小时的 $25\frac{1}{2}$ 分开出。若开出时间必须是整数分钟，则 $x=25$ 与 $x=26$（$d=23$ 与 $d=22$）不分高下，两者都给出 $\frac{2447}{30}\approx 81.57$ 分钟。

验算

- 当 $d=0$ 时，两班列车同时开出；人人都乘搭快车，平均等候 30 分钟，所以 $\mathrm{E}(J)=90$。公式给出 $\frac{5400}{60}=90$。
- 当 $d\to 60$ 时，慢车刚好在快车之前开出，几乎接载了所有乘客：平均等候 30 分钟，行车 75 分钟，合共 105 分钟。公式给出 $\frac{3600-2700+5400}{60}=105$。
- 以步长 0.1 的网格取遍每个 x，对一小时内的行程时间作数值积分（numerical integration），所得结果是：$x=12$ 时为 84.6，而最小值 81.5625 出现在 $x=25.5$。

常见错误

- 把在每小时 48 分至 60 分之间到达的乘客算作乘搭“12 分”开出的快车，却没有加上一小时：他们的列车是在下一个小时的 12 分开出的。
- 以列车的班数而非间隔的长度作为权重，去平均两个行车时间；决定一班列车接载多少乘客的，是该班列车之前那段间隔的长度。

STEP 之外

这里计算的量，是周期性时间表下的平均等候时间；而“较长的间隔会接载较多乘客”这个观察，正是检查悖论（inspection paradox）的一种形式：随机到达的乘客，落在长间隔的机会比落在短间隔的大。这是更新理论（renewal theory）和排队论（queueing theory）的核心概念，这两门理论研究的是随机的到达与服务。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门）是认识取条件与期望值（expectation）这些基础工具的好起点。

详细解答

STEP I 1992，第 16 题

概率与统计 • 网络可靠度

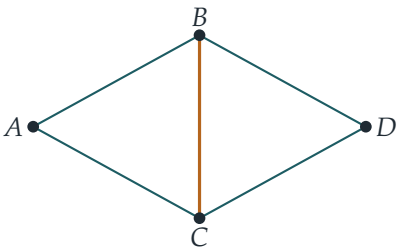
STEP I 1992，第 16 题

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/92-s1-q16

关键思路

这个道路网络是一座“桥”：由 A 到 D 有两条路线，一条经过 B，一条经过 C，而道路 CB 在中间把两者连接起来。处理这座桥最干净利落的方法，是按 CB 是否畅通来取条件（condition），因为一旦知道这一点，一切都化为串联（series）和并联（parallel）的道路。为一条道路加设防雪设施，效果等同把它的两个端点合并成一个市镇，所以我们可以考察两个方案各自产生的较小网络，从而比较两者。



解答

以下一律以 $q = 1 - p$ 表示某条道路畅通的概率。各条道路是否被封是互相独立（independent）的，所以任何一种特定的“畅通与被封”组合，其概率都是若干个 p 和 q 的乘积。由 A 到 D 而不重复经过任何市镇的路线共有四条：

$$ABD, \qquad ACD, \qquad ABCD, \qquad ACBD.$$

唯一畅通的路线是 $ABCD$

要 $ABCD$ 畅通，AB、BC 和 CD 必须全部畅通。要它成为唯一畅通的路线，ABD 必须被封，而由于 AB 畅通，这就迫使 BD 被封。同理，ACD 必须被封，而由于 CD 畅通，这就迫使 AC 被封。这样 $ACBD$ 也自动被封（它同时用到 AC 和 BD）。

所以这个事件（event）恰好就是“AB、BC、CD 畅通，而 AC、BD 被封”。由独立性，其概率为

$$q^3p^2 = p^2(1-p)^3,$$

正如所求。

现时能够通行的概率

设 E 为 A 与 D 有未被封的道路相连这个事件，并按道路 CB 的状况取条件。

若 CB 畅通，则 B 和 C 就如同一个市镇。除非 AB 和 AC 同时被封，否则我们可以由 A 到达这个市镇；除非 BD 和 CD 同时被封，否则可以由这个市镇到达 D。这两件事涉及不同的道路，所以

$$\mathrm{P}(E \mid CB \text{ 畅通}) = (1-p^2)^2.$$

若 CB 被封，可行的路线只有 ABD 和 ACD，各自畅通的概率为 q^2，而且两者没有共用的道路，所以互相独立。因此

$$\mathrm{P}(E \mid CB \text{ 被封}) = 1-(1-q^2)^2.$$

改为计算 A 与 D 被隔绝的概率会较为方便。由于 $1-(1-p^2)^2 = p^2(2-p^2)$ 及 $1-q^2 = p(2-p)$，

$$\begin{aligned}\mathrm{P}(\text{被隔绝}) &= q\,p^2(2-p^2) + p\cdot p^2(2-p)^2\\ &= p^2[(1-p)(2-p^2) + p(4-4p+p^2)]\\ &= p^2\,(2+2p-5p^2+2p^3).\end{aligned}$$

应为哪条道路加设防雪设施

为 CB 加设防雪设施。这时 CB 永远畅通，所以我们永远处于上面的第一种情况：

$$\mathrm{P}(\text{被隔绝}) = 1-(1-p^2)^2 = 2p^2-p^4.$$

为 AB 加设防雪设施。这时 A 和 B 就如同一个市镇，而我们与 D 被隔绝，当且仅当 BD 被封，并且经 C 也无路可通。经 C 通行需要 CD 畅通，并且 AC、CB 之中至少有一条畅通，所以此路不通的概率为 $p+q\,p^2$。因此

$$\mathrm{P}(\text{被隔绝}) = p\,(p+qp^2) = p^2+p^3-p^4.$$

想一想 两个答案都以 p^2 项开头。怎样才能判断对每一个 p 而言哪一个较小?

把两者相减:

$$(2p^2 - p^4) - (p^2 + p^3 - p^4) = p^2 - p^3 = p^2(1-p) > 0 \quad \text{其中} 0 < p < 1.$$

所以为 AB 加设防雪设施，被隔绝的机会总是较小，不论 p 的值是多少。政府应该选择 AB；两个选择永远不会同样好。

这与直觉吻合。修好 CB 只在需要它作为绕道的一部分时才有帮助，但修好 AB 却令 A 和 B 成为一个市镇，这个市镇既有直路 BD 通往 D，也有经过 C 的路径可走。

最后一部分

当 $p = \frac{1}{10}$ 时，现时被隔绝的概率为

$$\frac{1}{100}\left(2 + \frac{2}{10} - \frac{5}{100} + \frac{2}{1000}\right) = \frac{1}{100} \times 2.152 = 0.02152.$$

这大于 $\frac{1}{100}$，所以政府会落实计划（而且根据上一部分，它会为 AB 加设防雪设施）。

单看主项 $2p^2 = 0.02$ 已足以定论：只需两条道路被封便被隔绝的情况有两种，即 $\{AB, AC\}$ 和 $\{BD, CD\}$。

验算

- 当 $p \to 0$ 时，被隔绝的概率约为 $2p^2$，这正好数算了那两对道路：其中一对被封，便会把 A 或 D 孤立起来。为 AB 加设防雪设施后，只剩下 $\{BD, CD\}$ 这一对，概率约为 p^2，与我们的公式相符。为 CB 加设防雪设施后，两对仍然存在，概率约为 $2p^2$。
- 当 $p = 1$ 时，现时的公式给出 $1 \cdot (2 + 2 - 5 + 2) = 1$：若每条道路都被封，我们必定被隔绝。
- 把道路畅通与被封的全部 2^5 种组合逐一穷举，对多个 p 值都证实了以上每一条公式。当 $p = \frac{1}{10}$ 时，穷举给出现时的概率为 0.02152，为 AB 加设防雪设施后为 0.0109，为 CB 加设防雪设施后为 0.0199。模拟 (simulation) 2×10^5 个冬天所得的结果是 0.0215。

常见错误

- 把四条路线畅通的概率相加。这些路线有共用的道路，所以这些事件互有重叠，相加会重复计算。
- 在第一部分，忘记了“唯一畅通的路线”会迫使 AC 和 BD 被封，结果只写出 $(1-p)^3$。

- 只在 $p = \frac{1}{10}$ 时比较两个方案。两者之差 $p^2(1-p)$ 说明对每一个 p 而言都是 AB 较佳。

STEP 之外

这是网络可靠性（network reliability）的一个小例子，而按一条边取条件（或把它的两端合并）这个技巧，正是贯穿图论（graph theory）的删除—收缩（deletion–contraction）思想。对于大型的随机网络，同样的问题会引向渗流理论（percolation theory），当中 p 有一个分明的临界值，一旦越过，长距离的连通便会突然消失。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）是打好这里所用的条件概率工具的好地方。

详细解答

STEP II 1992，第 1 题

纯数学 · 数列的极限

> **STEP II 1992，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/92-s2-q1

关键思路

每个极限（limit）都只需要一个标准手法，功夫在于选对手法。遇到有理函数（rational function），把分子和分母同除以 n 的最高次幂。遇到有界（bounded）的量除以不断增大的量，就用夹逼（squeeze）。遇到乔装了的已知极限，就作代换（substitution），让伪装自行脱落。至于两个平方根之差，则乘以共轭式（conjugate）：这样可以把两个几乎相等的量之差，化成一个一目了然的商。

解答

第 (i) 部分

把分子和分母同除以 n，

$$\frac{n}{n+1} = \frac{1}{1+\frac{1}{n}} \longrightarrow \frac{1}{1+0} = 1,$$

因为 $\frac{1}{n} \to 0$。极限是 1。

第 (ii) 部分

最高次幂是 n^2，它只在分母出现。把分子和分母同除以 n^2，

$$\frac{5n+1}{n^2-3n+4} = \frac{\frac{5}{n}+\frac{1}{n^2}}{1-\frac{3}{n}+\frac{4}{n^2}} \longrightarrow \frac{0}{1} = 0.$$

极限是 0：分母的增长速度与 n^2 相若，而分子只与 n 相若。

第 (iii) 部分

对于大的整数 n，我们对 $\sin n$ 所知的有用资料只有一项，就是它有界：$|\sin n| \leq 1$。这已经足够，因为

$$\left|\frac{\sin n}{n}\right| \leq \frac{1}{n} \longrightarrow 0.$$

所以 $\dfrac{\sin n}{n}$ 被夹在 $-\frac{1}{n}$ 和 $\frac{1}{n}$ 之间，而两者都趋向 0，故极限是 0。

第 (iv) 部分

想一想 这不是和 (iii) 一样吗?

不一样：这里的角 $1/n$ 是缩小至 0，而不是不断增大。设 $x = 1/n$。当 $n \to \infty$ 时，我们有 $x \to 0^+$，而所求的式子就是 $\dfrac{\sin x}{x}$，它在 $x \to 0$ 时的极限是以下标准结果

$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$

（其中一个理解方法：这正是 sin 在 0 的导数（derivative），即 $\cos 0 = 1$。又或者利用 $\sin x = x - \frac{x^3}{6} + \cdots$，于是 $\frac{\sin x}{x} = 1 - \frac{x^2}{6} + \cdots$。）极限是 1。

第 (v) 部分

当 $n \to \infty$ 时，$\arctan n \to \frac{\pi}{2}$，因为当 $\theta \to \frac{\pi}{2}$（由下方趋近）时 $\tan\theta \to +\infty$，而且 arctan 是连续（continuous）且递增的。极限 $\frac{\pi}{2}$ 不是零，所以我们可以取倒数（reciprocal）：

$$(\arctan n)^{-1} \longrightarrow \frac{1}{\pi/2} = \frac{2}{\pi}.$$

留意 $(\arctan n)^{-1}$ 的意思是 $1/\arctan n$，而不是 arctan 的反函数（inverse function）（那样只会得出 tan）。

第 (vi) 部分

分子和分母都趋向 0，所以不能直接分别取极限。把两者各自乘以其共轭式：

$$\sqrt{n+1} - \sqrt{n} = \frac{(n+1) - n}{\sqrt{n+1} + \sqrt{n}} = \frac{1}{\sqrt{n+1} + \sqrt{n}}, \qquad \sqrt{n+2} - \sqrt{n} = \frac{2}{\sqrt{n+2} + \sqrt{n}}.$$

因此

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

这里把分子和分母同除以 $\sqrt{n}$。当 $n \to \infty$ 时，两个平方根都趋向 1（因为 $\sqrt{\ }$ 是连续的），所以极限是

$$\frac{1+1}{2(1+1)} = \frac{1}{2}.$$

想一想 为甚么答案应该是 $\frac{1}{2}$，而不是例如 1?

在一个大的 n 附近，$\sqrt{x}$ 的图像几乎是一条斜率（gradient）为 $\frac{1}{2\sqrt{n}}$ 的直线。所以 $\sqrt{n+1} - \sqrt{n} \approx \frac{1}{2\sqrt{n}}$（步长为 1），而 $\sqrt{n+2} - \sqrt{n} \approx \frac{2}{2\sqrt{n}}$（步长为 2）：两者之比约为 $\frac{1}{2}$，正是步长之比。

验算

代入一个大的 n。取 $n = 10^6$：(i) 得 0.999999；(ii) 得 5×10^{-6}；(iii) 约得 -3.5×10^{-7}；(iv) 准确至七个小数位得 1.0000000；(v) 得 0.636620，对照 $2/\pi \approx 0.636620$；而 (vi) 以有理化（rationalised）后的形式计算，得 0.50000012。在 (vi) 中，也应取小的 n 核对有理化后的形式与原式是否一致：当 $n = 1$ 时，两者都等于 $(\sqrt{2}-1)/(\sqrt{3}-1) \approx 0.5658$。

常见错误

- 把 (iii) 和 (iv) 当作同一回事：在 (iii) 中，正弦值保持有界而分母不断增大；在 (iv) 中，分子和分母都缩小至 0，而两者之比趋向 1。
- 把 (v) 中的 $(\arctan n)^{-1}$ 理解为反函数。
- 在 (vi) 中取极大的 n，直接按原式用计算机计算：两个几乎相等的数相减会损失精确度。应先有理化。

STEP 之外

在大学里，“$a_n \to L$” 有精确的定义：对每一个 $\varepsilon > 0$，都存在一个 N，使得对所有 $n > N$ 均有 $|a_n - L| < \varepsilon$。上面用过的每一个手法，例如 (iii) 中的夹逼、(v) 中对非零极限取倒数，以及 (vi) 中把极限移入连续函数之内，在第一个分析（analysis）课程中都会成为由这个定义出发证明的定理。Abbott 的 *Understanding Analysis*（一本分析学入门书）把这些概念阐述得既严谨又易读。

详细解答

STEP II 1992，第 2 题

纯数学 · 克莱罗微分方程

STEP II 1992，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/92-s2-q2

关键思路

这条方程的形式是 $y = xp - f(p)$，其中 $p = \mathrm{d}y/\mathrm{d}x$，称为克莱罗方程（Clairaut equation）。求导一次后，麻烦的 p 项会互相抵消，余下的式子可以因式分解成两条简单得多的方程。要留意的是，求导会丢失信息（常数消失了），所以每个候选解都必须代回 $(*)$，才能定出当中的常数。

解答

设 $p = \dfrac{\mathrm{d}y}{\mathrm{d}x}$，则 $(*)$ 可写成 $y = xp - \cosh p$。

求导

把等式两边对 x 求导，对 xp 用积法则（product rule），对 $\cosh p$ 用链式法则（chain rule）：

$$p = p + x\frac{\mathrm{d}p}{\mathrm{d}x} - \sinh p\,\frac{\mathrm{d}p}{\mathrm{d}x}.$$

两边的 p 互相抵消，余下

$$\frac{\mathrm{d}p}{\mathrm{d}x}\left(x - \sinh p\right) = 0, \qquad \text{即} \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2}\left(x - \sinh\frac{\mathrm{d}y}{\mathrm{d}x}\right) = 0.$$

乘积为零，当且仅当其中一个因子为零，所以 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2} = 0$ 或 $x - \sinh\left(\dfrac{\mathrm{d}y}{\mathrm{d}x}\right) = 0$，正是题目所求。

两条方程

第一条方程。把 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}=0$ 积分两次，得通解（general solution）

$$y=Ax+B,\qquad A,B\text{ 为任意常数。}$$

第二条方程。这里 $\sinh p=x$。由于 sinh 是由 $\mathbb{R}$ 到 $\mathbb{R}$ 的双射（bijection），所以对每个 x 都有 $p=\sinh^{-1}x$。用分部积分法（integration by parts），取 $u=\sinh^{-1}x$，$\mathrm{d}v=\mathrm{d}x$：

$$\begin{aligned}y&=\int\sinh^{-1}x\,\mathrm{d}x=x\sinh^{-1}x-\int\frac{x}{\sqrt{1+x^2}}\,\mathrm{d}x\\&=x\sinh^{-1}x-\sqrt{1+x^2}+C,\qquad C\text{ 为任意常数。}\end{aligned}$$

$(*)$ 的解

$(*)$ 的每个解都满足上面推导出的两条方程之一，但反过来却不成立，因为求导时丢掉了一个条件。所以我们要把每一族解代回 $(*)$。

> **想一想** 推导出的两条方程分别有两个和一个任意常数。为甚么这些解不可能全部都满足像 $(*)$ 这样的一阶方程?

一阶方程的解应该是单参数族，而上面每一族解不是常数太多，就是有一个尚未验证的常数。把 $(*)$ 求导，只能告诉我们 $y-xp+\cosh p$ 是常数；我们还需要这个常数等于零。

直线。当 $y=Ax+B$ 时，$p=A$，于是 $(*)$ 变成

$$Ax+B=Ax-\cosh A\iff B=-\cosh A.$$

所以我们得到单参数的直线族

$$y=Ax-\cosh A,\qquad A\in\mathbb{R}.$$

另一条曲线。当 $y=x\sinh^{-1}x-\sqrt{1+x^2}+C$ 时，$p=\sinh^{-1}x$，所以 $\sinh p=x$，而 $\cosh p=\sqrt{1+\sinh^2 p}=\sqrt{1+x^2}$（因为 $\cosh p>0$，故取正根）。于是 $(*)$ 的右边是

$$xp-\cosh p=x\sinh^{-1}x-\sqrt{1+x^2},$$

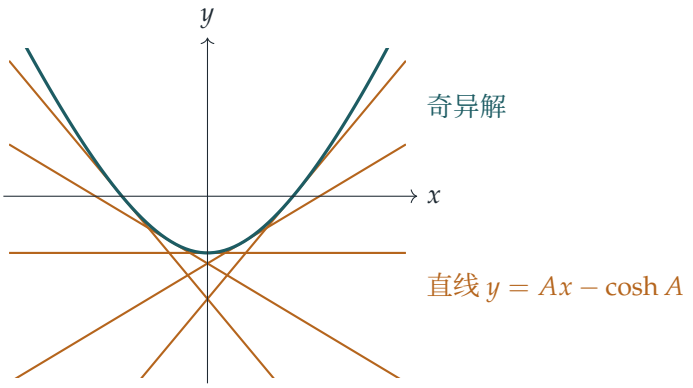
它等于 y，当且仅当 $C=0$。这样便多得一个解，当中不含任意常数：

$$y=x\sinh^{-1}x-\sqrt{1+x^2}.$$

因此，$(*)$ 的解是直线 $y=Ax-\cosh A$（A 为任意实数），加上单独一条曲线 $y=x\sinh^{-1}x-\sqrt{1+x^2}$。

两者如何配合

斜率为 A 的直线在 $x = \sinh A$ 处与曲线相切。在该处，曲线的斜率是 $\sinh^{-1}(\sinh A) = A$，高度是 $A\sinh A - \cosh A$，与直线的相同。曲线的 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2} = \dfrac{1}{\sqrt{1+x^2}} > 0$，所以它是凸（convex）的，而每一条这样的切线都位于它的下方。这条曲线是该直线族的包络线（envelope），称为奇异解（singular solution）。



这也解释了一个微妙之处：由于条件 $\dfrac{\mathrm{d}^2y}{\mathrm{d}x^2}(x - \sinh p) = 0$ 是逐点成立的，一个解可以在两种类型之间切换。例如，沿着曲线走到 $x = \sinh A$，然后改沿该处的切线继续走，所得的函数有连续的导数，而且仍然处处满足 $(*)$。这类分段的解都是由上面找到的解拼合而成的，而题目预期的答案是直线族加上奇异曲线。

验算

- 当 $A = 0$ 时，直线是 $y = -1$，而曲线有 $y(0) = 0 - 1 = -1$，斜率为 $\sinh^{-1} 0 = 0$：两者在 $(0, -1)$ 相切，与预期相符。
- 数值上，取 $x = 0.7$：$\sinh^{-1} 0.7 \approx 0.6527$，所以曲线给出 $y \approx 0.4569 - 1.2207 = -0.7638$，而 $xp - \cosh p \approx 0.4569 - 1.2207$，两者吻合。在 y 上加上任何 $C \neq 0$，都会刚好相差 C。
- 在随机选取的点上用有限差分（finite differences）检验 $(*)$，每一条直线 $y = Ax - \cosh A$ 和奇异曲线的残差都只有舍入误差的大小。

常见错误

- 做到 $y = Ax + B$ 或 $y = x\sinh^{-1} x - \sqrt{1+x^2} + C$ 便停下来：这些只是求导后的方程的解，而 $(*)$ 本身要求 $B = -\cosh A$ 和 $C = 0$。

▶ 把因子 $\dfrac{\mathrm{d}p}{\mathrm{d}x}$ “约去”，以致完全漏掉奇异解。

STEP 之外

形如 $y = xp - f(p)$ 的方程就是克莱罗方程。它们的通解永远是直线族 $y = cx - f(c)$，而奇异解就是这个直线族的包络线。这里 $f = \cosh$，而包络线 $y = x\sinh^{-1} x - \sqrt{1+x^2}$ 是 cosh 的勒让德变换（Legendre transform），这种构造在力学（由拉格朗日量（Lagrangian）过渡到哈密顿量（Hamiltonian））和热力学中都会出现。这个例子也说明，当存在唯一性定理（existence and uniqueness theorem）的条件不成立时，一阶方程的解不一定唯一。Boyce 与 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科书）涵盖一阶方程和这条唯一性定理。

详细解答

STEP II 1992，第 3 题

纯数学 · 内摆线与面积

> **STEP II 1992，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/92-s2-q3

关键思路

一切都源自*纯滚动而不滑动*（rolling without slipping）：小圆沿大圆滚过的弧长，等于小圆上曾经与大圆接触的弧长。这就决定了小圆转了多少，于是 B 不过是“圆心 P 加上一个方向正确的单位向量（unit vector）”。至于描绘曲线，值得留意的是，利用三倍角公式（triple-angle formulae），坐标可以化简得非常简洁。

解答

求 B 的坐标

由于小圆（半径为 1）一直与大圆（半径为 4）的内侧相切，它的圆心与 O 的距离恒为 $4-1=3$。设 ϕ 为角 AOP，则

$$P=(3\cos\phi,\ 3\sin\phi).$$

设 T 为当前的切点（point of contact）。它位于射线 OP 的延长线上，所以 $T=(4\cos\phi, 4\sin\phi)$，而由 P 到 T 的方向与 x 轴成角 ϕ。

起初 B 位于切点 A。由于没有滑动，大圆上的弧 AT 与小圆上的弧 BT 长度相同：

$$4\phi = 1\times\angle BPT \quad\Longrightarrow\quad \angle BPT = 4\phi.$$

> **想一想** 在 P 处量度，B 所在的角度是 $\phi+4\phi$ 还是 $\phi-4\phi$?

当 P 绕 O 逆时针移动时，小圆沿大圆的内侧滚动，因此是*顺时针*转动的（见图中箭头）。点 B 是较早时的切点，所以它落在新切点 T 的后面：沿小圆由 T 到 B 是顺时针转了 4ϕ。

因此方向 $\overrightarrow{PB}$ 与 x 轴成角 $\phi - 4\phi = -3\phi$，又因为 $PB = 1$，

$$\begin{aligned} B &= P + (\cos(-3\phi),\ \sin(-3\phi)) \\ &= (3\cos\phi + \cos 3\phi,\ 3\sin\phi - \sin 3\phi), \end{aligned}$$

正如所求。（验算：$\phi = 0$ 给出 $(4, 0) = A$。）

描绘 C

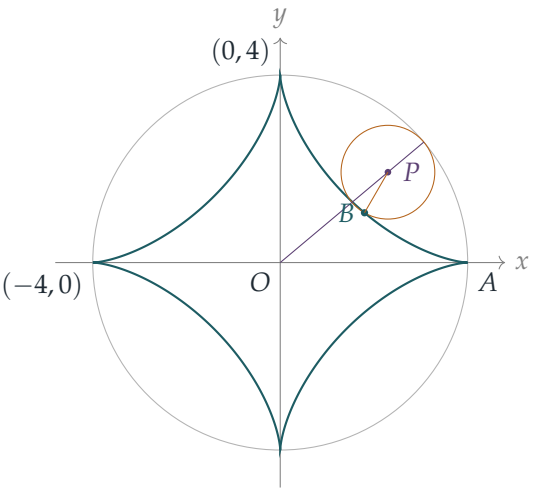
利用 $\cos 3\phi = 4\cos^3\phi - 3\cos\phi$ 及 $\sin 3\phi = 3\sin\phi - 4\sin^3\phi$，

$$x = 4\cos^3\phi, \qquad y = 4\sin^3\phi.$$

所以 C 是曲线 $x^{2/3} + y^{2/3} = 4^{2/3}$，即星形线（astroid）。它有以下特征：

- B 在大圆上，当且仅当 $\cos^6\phi + \sin^6\phi = 1$，即 $\phi = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$，对应的点为 $(\pm 4, 0)$、$(0, \pm 4)$。这时 B 就是切点，所以它瞬时静止，曲线在该处有一个尖点（cusp）。
- 曲线对两条坐标轴都对称（把 ϕ 换成 $-\phi$ 或 $\pi - \phi$）。
- 在第一象限，$\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{12\sin^2\phi\cos\phi}{-12\cos^2\phi\sin\phi} = -\tan\phi$，所以曲线沿 x 轴方向离开 $(4, 0)$，并沿 y 轴方向到达 $(0, 4)$。曲线向内弯曲：当 $\phi = \frac{\pi}{4}$ 时它通过 $(\sqrt{2}, \sqrt{2})$，远在直线 $x + y = 4$ 之内。

P 绕行一周（ϕ 由 0 到 2π）之后，小圆相对于 OP 转了 3 圈，而 B 回到 A，所以 C 是一条有四个尖点的闭曲线（closed curve）。



面积

我们有 $x = 3\cos\phi + \cos 3\phi$ 及

$$\frac{\mathrm{d}y}{\mathrm{d}\phi} = 3\cos\phi - 3\cos 3\phi,$$

所以

$$x\frac{\mathrm{d}y}{\mathrm{d}\phi} = 9\cos^2\phi - 6\cos\phi\cos 3\phi - 3\cos^2 3\phi.$$

在一个完整周期上，对任何整数 $k \neq 0$ 都有 $\int_0^{2\pi}\cos^2 k\phi\,\mathrm{d}\phi = \pi$（把被积函数写成 $\frac{1}{2}(1+\cos 2k\phi)$），而

$$\int_0^{2\pi}\cos\phi\cos 3\phi\,\mathrm{d}\phi = \frac{1}{2}\int_0^{2\pi}(\cos 4\phi + \cos 2\phi)\,\mathrm{d}\phi = 0.$$

因此

$$\text{面积} = \int_0^{2\pi} x\frac{\mathrm{d}y}{\mathrm{d}\phi}\,\mathrm{d}\phi = 9\pi - 0 - 3\pi = 6\pi.$$

答案为正数，这是应有的结果：随着 ϕ 增加，B 沿 C 逆时针绕行。

验算

- 由于 C 向内弯曲，所围区域位于以四个尖点为顶点的正方形之内，该正方形的面积为 32；而确实 $6\pi \approx 18.85 < 32$。
- 数值上，对该积分使用辛普森法则（Simpson's rule），以及用鞋带公式（shoelace formula）计算 C 的内接 200 000 边形的面积，两者都给出 $18.8496\ldots = 6\pi$。
- 标准的星形线公式（面积为 $\frac{3}{8}\pi a^2$），取 $a = 4$，同样给出 6π。
- 在每个尖点处，B 的速率的数值结果为零，印证了 B 碰到大圆时处于静止状态，与滚动的要求完全吻合。

常见错误

- 把 $\overrightarrow{PB}$ 的角度取为 $+5\phi$ 或 $+3\phi$：小圆的转动方向与 P 绕 O 运动的方向相反。
- 计算 P 的路径时用了半径 4 而不是 3。
- 把尖点画成圆滑的末端，或把弧画成向外凸出。

STEP 之外

一个圆在另一个圆内滚动时，圆上一点所描出的曲线称为*内摆线*（hypocycloids）；在外面滚动则得到*外摆线*（epicycloids）。当半径之比为 4 : 1 时，我们得到星形线；比为 3 : 1 时，则得到有三个尖点的三尖瓣线（deltoid）。面积公式 $\int x\,\mathrm{d}y$ 是格林定理（Green's theorem）的特殊情况，该定理把沿闭曲线的积分与曲线所围区域上的二重积分（double integral）联系起来；它是大学一年级向量微积分（vector calculus）的一大亮点。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）利用 $z = 3e^{i\phi} + e^{-3i\phi}$ 优雅地处理了这些滚动曲线。

详细解答

STEP II 1992，第 4 题

纯数学 · 微分法的公理

> **STEP II 1992，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/92-s2-q4

本题在 *Beyond STEP* 第 6.5 节从公理出发的微分法中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch06/derivations.html

关键思路

这四条规则正是我们熟悉的求导（differentiation）性质：乘积法则（product rule）、线性（linearity）以及 $\frac{\mathrm{d}}{\mathrm{d}x}x = 1$。题目要我们证明，单凭这些性质，就足以完全确定这个运算在多项式（polynomial）上的作用。处理常数的窍门，是把一个我们早已知道结果的乘积交给乘积法则，例如 $1 \times 1 = 1$。之后，乘积法则会逐步建立 x 的每一个幂，余下的则由线性处理。

在整个过程中，我们只可使用题目给出的四条规则。不能因为 $\Diamond$ 看起来像求导，就假设它具有任何性质；这正是我们要证明的事。

解答

常数

> ? **想一想** 我们对 $\Diamond 1$ 一无所知。这些规则怎样能告诉我们它是甚么？

把乘积法则 (i) 应用于多项式 $\mathrm{f}(x) = \mathrm{g}(x) = 1$。它们的乘积仍是 1，所以

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

两边同时减去 $\Diamond 1$，得 $\Diamond 1 = 0$，即零多项式（zero polynomial）。

现设 $\mathrm{f}(x) = c$ 为任意常数多项式。则 $\mathrm{f}(x) = c \cdot 1$，而 c 是常数，所以由规则 (iv) 得

$$\Diamond c = \Diamond(c \cdot 1) = c \Diamond 1 = c \cdot 0 = 0.$$

这也涵盖了 $c = 0$ 的情况，所以对每一个常数多项式都有 $\Diamond \mathrm{f}(x) = 0$，正如所求。

（求得 $\Diamond 1 = 0$ 的另一途径：写成 $x = x \cdot 1$，则由规则 (i) 和规则 (iii) 得 $1 = 1 \cdot 1 + x \Diamond 1$，因此 $x \Diamond 1 = 0$，而一个多项式乘以 x 后得 0，它必定是零多项式。）

幂 x^2 和 x^3

取 $\mathrm{f}(x) = \mathrm{g}(x) = x$ 并使用乘积法则，再用规则 (iii)，

$$\Diamond x^2 = \Diamond(x \cdot x) = x \Diamond x + x \Diamond x = x + x = 2x.$$

接着，$x^3 = x^2 \cdot x$，所以由乘积法则、规则 (iii) 和刚刚得到的结果，

$$\Diamond x^3 = x \Diamond x^2 + x^2 \Diamond x = x \cdot 2x + x^2 \cdot 1 = 3x^2.$$

任意多项式

*第 1 步：x 的幂。*我们用归纳法（induction）证明

$$\Diamond x^n = nx^{n-1} \qquad \text{对每一个整数} n \geq 1.$$

当 $n = 1$ 时，这就是规则 (iii)。假设它对某个 $n = k \geq 1$ 成立。写成 $x^{k+1} = x^k \cdot x$ 并使用乘积法则，再用规则 (iii) 和归纳假设（induction hypothesis），

$$\Diamond x^{k+1} = x \Diamond x^k + x^k \Diamond x = x \cdot kx^{k-1} + x^k = (k+1)x^k,$$

所以结果对 $n = k+1$ 成立。由归纳法，它对所有 $n \geq 1$ 成立。当 $n = 0$ 时，$x^0 = 1$ 是常数，所以由第一部分得 $\Diamond x^0 = 0$。在每一种情况下都有 $\Diamond x^n = \frac{\mathrm{d}}{\mathrm{d}x} x^n$。

*第 2 步：和。*规则 (ii) 只是就两个多项式而言，但对项数作归纳，便可把它推广至任意有限和：若 $\Diamond(\mathrm{f}_1 + \cdots + \mathrm{f}_m) = \Diamond \mathrm{f}_1 + \cdots + \Diamond \mathrm{f}_m$，则把规则 (ii) 应用于 $(\mathrm{f}_1 + \cdots + \mathrm{f}_m) + \mathrm{f}_{m+1}$，便得到 $m+1$ 项时的同一命题。

*第 3 步：综合起来。*任意多项式都可写成 $\mathrm{h}(x) = a_0 + a_1 x + \cdots + a_n x^n$，其中系数 a_k 为常数。先用第 2 步，再对每一项用规则 (iv)，然后用第 1 步，

$$\begin{aligned}
\Diamond \mathrm{h}(x) &= \Diamond(a_0) + \Diamond(a_1 x) + \cdots + \Diamond(a_n x^n) \\
&= a_0 \Diamond 1 + a_1 \Diamond x + \cdots + a_n \Diamond x^n \\
&= 0 + a_1 + 2a_2 x + \cdots + n a_n x^{n-1} \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\big(\mathrm{h}(x)\big),
\end{aligned}$$

这正是我们要证明的结果。

验算

值得确认一下求导确实满足全部四条规则，否则这个结果便会流于空泛；它确实满足（乘积法则、线性以及 $\frac{\mathrm{d}}{\mathrm{d}x}x = 1$）。我们亦按照解答中的做法，只用这四条规则在电脑上构作了 $\Diamond$，并以 500 个次数不超过 7、系数为分数的随机多项式，把它与普通的导数（derivative）作比较。两者每次都一致，并且与数值差商（difference quotient）吻合；例如由这些规则可得 $\Diamond x^3 = 3x^2$。

常见错误

- 说“$\Diamond c = 0$，因为常数的导数是零”。这是假设了我们要证明的事；这个零必须由规则推出。
- 在最后一部分遗漏常数项。归纳法只由 $n = 1$ 开始，所以 a_0 需要用到关于常数的结果。
- 对多项之和使用规则 (ii)，却没有指出它是藉归纳法由两项推广而来。

STEP 之外

满足规则 (i)、(ii) 和 (iv) 的运算称为导子（derivation），代数和几何中处处都会研究导子。本题说明，在多项式上，导子完全由它在 x 上的值决定。更一般地，$\Diamond \mathrm{h} = \mathrm{p}(x)\,\mathrm{h}'(x)$，其中 $\mathrm{p} = \Diamond x$；在微分几何（differential geometry）中，这个想法成为向量场（vector field）的定义。至于多项式环（polynomial ring）这个代数背景，可参阅 Allenby 的 *Rings, Fields and Groups*（《环、域与群》）。

详细解答

STEP II 1992，第 5 题

纯数学 · 矩阵群

> **STEP II 1992，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/92-s2-q5

关键思路

这里的一切都建基于条件 $wz - xy = 1$，它令 S 中矩阵（matrix）的逆矩阵（inverse）特别简单。处理阶为 3 的部分，最好的方法是留意到每个 2×2 矩阵都满足一条由其迹（trace）和行列式（determinant）构成的二次方程。对于 S 中的矩阵，我们可借此把 $\mathbf{A}^2$，继而把 $\mathbf{A}^3$，用 $\mathbf{A}$ 和 $\mathbf{I}$ 表示出来，于是条件 $\mathbf{A}^3 = \mathbf{I}$ 就变成关于迹 $w + z$ 的条件。

（试卷上把 $\mathbf{A}$ 右下角的元素印成 x；但最后一部分提到 z，可见原意显然是该位置为 z，以下一律采用此写法。）

解答

元素的阶

$g \in G$ 的阶（order）是使 $g^n = e$ 成立的最小正整数 n，其中 e 是 G 的单位元（identity）。若 g 的任何正整数次幂都不等于 e，则 g 的阶为无限。单位元是唯一阶为 1 的元素。

逆矩阵

设

$$\mathbf{A} = \begin{pmatrix} w & x \\ y & z \end{pmatrix}, \qquad wz - xy = 1.$$

2×2 逆矩阵的常用公式要除以行列式，而这里行列式为 1，所以

$$\mathbf{A}^{-1} = \begin{pmatrix} z & -x \\ -y & w \end{pmatrix}.$$

乘开即可验证：$\mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} wz - xy & 0 \\ 0 & wz - xy \end{pmatrix} = \mathbf{I}$，同理 $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$。

S 是一个群

我们逐一检验群（group）的四条公理。

- 封闭性（*closure*）。若 $\mathbf{A}, \mathbf{B} \in S$，则 $\det(\mathbf{AB}) = \det\mathbf{A}\,\det\mathbf{B} = 1 \times 1 = 1$，所以 $\mathbf{AB} \in S$。
- 结合律（*associativity*）。矩阵乘法的结合律可视为已知。
- 单位元。$\det\mathbf{I} = 1$，所以 $\mathbf{I} \in S$，而且对每个 $\mathbf{A}$ 都有 $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$。
- 逆元。上面求得的矩阵 $\mathbf{A}^{-1}$ 的行列式为 $wz - (-x)(-y) = wz - xy = 1$，所以它属于 S。

因此 S 在矩阵乘法下是一个群。

自逆元素与阶为 2 的元素

把 $\mathbf{A}^{-1}$ 与 $\mathbf{A}$ 逐个元素比较，可知 $\mathbf{A}^{-1} = \mathbf{A}$ 当且仅当

$$z = w, \qquad -x = x, \qquad -y = y,$$

即 $x = y = 0$ 且 $z = w$。这时行列式条件变成 $w^2 = 1$，所以 $w = \pm 1$。因此自逆（self-inverse）元素为

$$\mathbf{A} = \mathbf{I} \quad \text{和} \quad \mathbf{A} = -\mathbf{I}.$$

当 $\mathbf{A}^2 = \mathbf{I}$（等价于 $\mathbf{A}^{-1} = \mathbf{A}$）且 $\mathbf{A} \neq \mathbf{I}$ 时，元素的阶为 2。所以阶为 2 的元素只有 $-\mathbf{I}$。

阶为 3 的元素

记 $t = w + z$ 为 $\mathbf{A}$ 的迹。平方得

$$\mathbf{A}^2 = \begin{pmatrix} w^2 + xy & x(w+z) \\ y(w+z) & xy + z^2 \end{pmatrix}.$$

由于 $xy = wz - 1$，左上角的元素是 $w^2 + wz - 1 = tw - 1$，右下角的元素是 $tz - 1$。因此

$$\mathbf{A}^2 = t\mathbf{A} - \mathbf{I}. \tag{$*$}$$

> **?** **想一想** $(*)$ 对计算 $\mathbf{A}^3$ 有甚么帮助?

把 $(*)$ 乘以 $\mathbf{A}$，再用一次 $(*)$ 消去 $\mathbf{A}^2$：

$$\mathbf{A}^3 = t\mathbf{A}^2 - \mathbf{A} = t(t\mathbf{A} - \mathbf{I}) - \mathbf{A} = (t^2 - 1)\mathbf{A} - t\mathbf{I}.$$

所以 $\mathbf{A}$ 的任何次幂都是 $\mathbf{A}$ 和 $\mathbf{I}$ 的组合，我们可以比较系数。

若 $w+z+1=0$：则 $t=-1$，所以 $t^2-1=0$，从而 $\mathbf{A}^3=\mathbf{I}$。另外 $\mathbf{A}\neq\mathbf{I}$，因为 $\mathbf{I}$ 的迹是 $2\neq-1$。$\mathbf{A}$ 的阶整除 3 而又不是 1，所以是 3。

反过来，假设 $\mathbf{A}$ 的阶为 3。则 $\mathbf{A}^3=\mathbf{I}$ 且 $\mathbf{A}\neq\mathbf{I}$，所以

$$(t^2-1)\mathbf{A}=(1+t)\mathbf{I}.$$

- 若 $t^2\neq1$，则 $\mathbf{A}=\dfrac{1+t}{t^2-1}\mathbf{I}=\dfrac{1}{t-1}\mathbf{I}$，是单位矩阵的倍数 $\lambda\mathbf{I}$。它的行列式为 $\lambda^2=1$，所以 $\lambda=\pm1$。但 $\mathbf{A}=\mathbf{I}$ 已被排除，而 $\mathbf{A}=-\mathbf{I}$ 的阶是 2，不是 3。所以这个情况不可能出现。
- 若 $t=1$，等式变成 $\mathbf{0}=2\mathbf{I}$，这并不成立。

余下唯一的可能是 $t=-1$，即 $w+z+1=0$。这就证明了“当且仅当”（if and only if）。

一个例子。取 $w=0$ 和 $z=-1$，并选 $x=-1$、$y=1$，使得 $wz-xy=0+1=1$：

$$\mathbf{A}=\begin{pmatrix}0&-1\\1&-1\end{pmatrix},\qquad \mathbf{A}^2=\begin{pmatrix}-1&1\\-1&0\end{pmatrix},\qquad \mathbf{A}^3=\mathbf{I}.$$

验算

- 把上面的例子乘开，直接看出 $\mathbf{A}^3=\mathbf{I}$。
- 旋转 120° 的矩阵 $\begin{pmatrix}-\frac{1}{2}&-\frac{\sqrt{3}}{2}\\\frac{\sqrt{3}}{2}&-\frac{1}{2}\end{pmatrix}$ 的行列式为 1，迹为 -1，而它的阶当然是 3。
- 我们用电脑对 S 中 20 000 个随机有理数矩阵验证了 $\mathbf{A}^2=t\mathbf{A}-\mathbf{I}$。我们亦搜索了 S 中所有元素均为 $\frac{1}{2}$ 的倍数且介乎 -4 与 4 之间的矩阵：阶为 3 的恰好就是迹为 -1 的那 48 个，而 $-\mathbf{I}$ 是唯一阶为 2 的矩阵。

常见错误

- 把 $\mathbf{I}$ 算作阶为 2 的元素。它是自逆的，但它的阶是 1。
- 只证明了阶为 3 这个结果的其中一个方向。逆命题（converse）需要排除 $t^2\neq1$ 和 $t=1$ 这两个情况。
- 验证群公理时，忘记检查逆矩阵本身的行列式也是 1。

STEP 之外

群 S 就是特殊线性群（special linear group）SL_2，是数学中最重要的群之一。恒等式 $\mathbf{A}^2 - (\operatorname{tr}\mathbf{A})\mathbf{A} + (\det\mathbf{A})\mathbf{I} = \mathbf{0}$ 是凯莱–哈密顿定理（Cayley–Hamilton theorem）的 2×2 情形：每个方阵都满足它自己的特征方程（characteristic equation）。它亦显示出单凭迹就能告诉我们多少关于 SL_2 中元素的信息：上面的论证可以推广，证明阶为有限数 n 的元素，其迹必为 $2\cos(2\pi k/n)$，其中 k 为某个整数。想进一步了解矩阵群，可参阅 Allenby 的 *Rings, Fields and Groups*（环、域与群）；至于凯莱–哈密顿定理和特征值（eigenvalue），可参阅 Strang 的 *Introduction to Linear Algebra*（线性代数导论）。

详细解答

STEP II 1992，第 6 题

纯数学 · 切线与根的计数

> **STEP II 1992，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/92-s2-q6

关键思路

方程 $kx = \ln(2\sec x)$ 的解，就是通过原点的直线 $y = kx$ 与曲线 $y = \ln(2\sec x)$ 相交的点。这条曲线从原点上方出发，是凸（convex）的，并且急升至无穷大，所以通过原点的直线要么与它不相交，要么与它相切，要么与它相交两次。特殊值 k_0 就是由原点所作切线的斜率（gradient）。相切（tangency）的意思是两条曲线在该点高度相同、斜率也相同，而其余一切都可以由这两条方程推出。

解答

两个图像

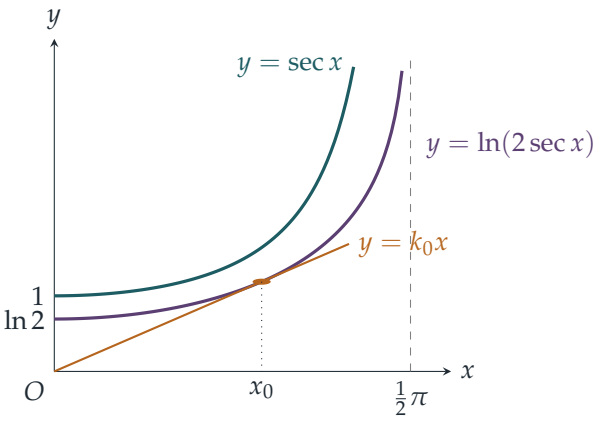
设 $f(x) = \ln(2\sec x) = \ln 2 - \ln\cos x$。则

$$f'(x) = \frac{\sin x}{\cos x} = \tan x, \qquad f''(x) = \sec^2 x > 0.$$

所以 $f(0) = \ln 2 \approx 0.69$，$f'(0) = 0$，$f$ 在 $(0, \frac{1}{2}\pi)$ 上递增，它是凸的（向上凹），而且当 $x \to \frac{1}{2}\pi^-$ 时 $f(x) \to \infty$，因为 $\cos x \to 0^+$。

同样地，$y = \sec x$ 满足 $\sec 0 = 1$，斜率为 $\sec x \tan x$，在 $x = 0$ 处为 0，其后为正，而且当 $x \to \frac{1}{2}\pi^-$ 时 $\sec x \to \infty$。两条曲线都以 $x = \frac{1}{2}\pi$ 为铅垂渐近线（vertical asymptote），并且在该处都没有定义。

对数曲线永远位于正割曲线的下方。设 $u = \sec x \geq 1$，则有 $\ln(2u) < u$，因为 $\ln(2u) - u$ 的导数（derivative）是 $\frac{1}{u} - 1 \leq 0$，所以它不会超过它在 $u = 1$ 处的值，即 $\ln 2 - 1 < 0$。对数曲线的增长也慢得多：当 $x \to \frac{1}{2}\pi$ 时，它的表现就像正割的 ln。



k 小时无解，k 大时有两个解

由于 f 递增，在 $[0, \frac{1}{2}\pi)$ 上有 $f(x) \geq \ln 2$，而在该处 $kx < \frac{1}{2}\pi k$。若 $k \leq 2\ln 2/\pi \approx 0.44$，则处处都有 $kx < \ln 2 \leq f(x)$，所以直线从曲线下方经过，方程无解。

当 k 大时，直线起初在曲线下方（在 $x = 0$ 处它的高度是 $0 < \ln 2$），但它够陡，会在某处升到曲线之上：在 $x = \frac{1}{4}\pi$ 处曲线的高度是 $\ln(2\sqrt{2}) = \frac{3}{2}\ln 2 \approx 1.04$，只要 $k > \frac{6}{\pi}\ln 2 \approx 1.32$，直线在该处就较高。在 $\frac{1}{2}\pi$ 附近，曲线趋向无穷大，而直线的高度始终不超过 $\frac{1}{2}\pi k$，所以曲线最终又回到直线之上。直线先向上穿过曲线，然后又向下穿过一次：有两个解。

解不可能多于两个。函数 $f(x) - kx$ 的二阶导数（second derivative）是 $\sec^2 x > 0$，所以它是凸的，而凸函数最多只有两个零点。

为何 k_0 存在

由于 $f(0) = \ln 2 \neq 0$，每个解都满足 $x > 0$，所以我们可以除以 x，把方程改写成

$$k = g(x), \qquad \text{其中} g(x) = \frac{\ln(2\sec x)}{x}, \quad 0 < x < \tfrac{1}{2}\pi.$$

所以解的数目就是 g 的图像到达高度 k 的次数。现在

$$g'(x) = \frac{xf'(x) - f(x)}{x^2} = \frac{h(x)}{x^2}, \qquad h(x) = x\tan x - \ln(2\sec x),$$

而 $h'(x) = \tan x + x\sec^2 x - \tan x = x\sec^2 x > 0$。由于 $h(0) = -\ln 2 < 0$，而且当 $x \to \frac{1}{2}\pi$ 时 $h(x) \to \infty$，h 恰好有一个零点 x_0。所以 g 在 $(0, x_0)$ 上递减，在 $(x_0, \frac{1}{2}\pi)$ 上递增，而且在两端都有 $g \to \infty$（当 $x \to 0$ 时分子趋向 $\ln 2$，当 $x \to \frac{1}{2}\pi$ 时分子趋向 ∞）。因此，取 $k_0 = g(x_0)$，即 g 的最小值，便有：

- $k < k_0$：无解；
- $k = k_0$：恰好有一个解，$x = x_0$；

▶ $k > k_0$：有两个解。

从图像来看，$y = k_0 x$ 就是 $y = \ln(2\sec x)$ 的通过原点的切线。这印证了草图的论证，并准确地确定了临界值。

> ? **想一想** 为甚么“恰好有一个解”就表示直线是切线?

假如直线在 x_0 处横截地穿过曲线，那么把它绕 O 稍微转动（令 k 稍为减小或增大），附近仍然会有一个交点，而上面的凸性论证又会给出第二个交点。一个单一而孤立的交点，在 k 减小时变成没有交点，这正正就是切点。用代数来说，这就是 $h(x_0) = 0$，即 $g'(x_0) = 0$。

x_0 所满足的方程

在切点处，直线与曲线的高度相同，斜率也相同：

$$k_0 x_0 = \ln(2\sec x_0), \qquad k_0 = f'(x_0) = \tan x_0.$$

（把两式合起来，就是上面的 $h(x_0) = 0$。）把第二式代入第一式，得 $x_0 \tan x_0 = \ln(2\sec x_0)$，而由于 $0 < x_0 < \frac{1}{2}\pi$，$\tan x_0 \neq 0$，所以

$$x_0 = \cot x_0 \, \ln(2\sec x_0),$$

这正是所要证明的。

求 x_0 和 k_0

我们很自然会想直接用 $x_{n+1} = \cot x_n \ln(2\sec x_n)$ 来迭代（iterate）。这样做收敛（converge）得很慢：右边在 x_0 处的导数是 $1 - 2x_0/\sin 2x_0 \approx -0.89$，所以迭代值会来回振荡，每一步只能逼近约 11%。对 $h(x) = x\tan x - \ln(2\sec x)$ 使用牛顿法（Newton's method）则好得多，而且 $h'(x) = x\sec^2 x$ 十分简单：

$$x_{n+1} = x_n - \frac{x_n \tan x_n - \ln(2\sec x_n)}{x_n \sec^2 x_n}.$$

由 $x_1 = 1$ 开始（全部以弧度（radian）计算）：

$$x_2 \approx 0.9274, \qquad x_3 \approx 0.9147, \qquad x_4 \approx 0.91440.$$

为确认准确至两位小数，我们检查符号有否改变：$h(0.905) \approx -0.023 < 0$ 而 $h(0.915) \approx +0.0015 > 0$，所以 $0.905 < x_0 < 0.915$，于是

$$x_0 \approx 0.91.$$

然后 $k_0 = \tan x_0 \approx \tan 0.9144 \approx 1.30$。等价地，$k_0 = \ln(2\sec x_0)/x_0$，所得的值相同。

验算

- k_0 的两个表达式互相吻合：$\tan(0.91440) \approx 1.2981$ 而 $\ln(2\sec 0.91440)/0.91440 \approx 1.2981$。
- 以数值方法在 $(0, \frac{1}{2}\pi)$ 上扫描 $g(x) = \ln(2\sec x)/x$，得最小值为 1.2981，位于 $x \approx 0.914$。
- 数算 $kx - \ln(2\sec x)$ 的符号改变次数：$k = 1.29$ 时没有根，$k = 1.31$ 时有两个根。
- 答案符合先前求得的粗略界限：$0.44 < k_0 < 1.32$。

常见错误

- 使用角度制。这里每一个 x 的值，包括牛顿法的迭代值和 $\tan x_0$，都必须以弧度计算。
- 只列出一条相切的方程。与曲线相交并不足够：斜率也必须相等，而正是这第二个条件确定了 k_0。
- 未经检查就声称准确至两位小数。应展示 h 在 $[0.905, 0.915]$ 两端符号相反，或者一直迭代至第二位小数稳定下来。

STEP 之外

通过考察 $f(x)/x$ 的值域（range）来数算 $kx = f(x)$ 的根的数目，是一个重要思想的一小部分。通过原点的切线，加上 f 的凸性，正是勒让德变换（Legendre transform）和凸对偶（convex duality）的基础，在分析学和力学中都是如此。牛顿法在这里收敛得这么快，是因为它在单根（simple root）附近是二次收敛（quadratically convergent）的，而不动点迭代（fixed-point iteration）只是线性收敛，比率为 $|\varphi'(x_0)|$。这两个事实在分析学或数值方法的入门课程中都会证明。Bartle 与 Sherbert 的 *Introduction to Real Analysis*（实分析入门）涵盖了凸函数以及牛顿法的误差分析。

详细解答

STEP II 1992，第 7 题

纯数学 · 三次方程的重根

> **STEP II 1992，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/92-s2-q7

关键思路

把 $(x-a)(x-b)(x-c)$ 与该三次多项式比较系数（comparing coefficients），便可把 p、q 和 r 表示成根的对称函数（symmetric functions）。在 (i) 中，有一个重根（repeated root），再加上“各根之和为零”，根的形式就只能是 $\alpha, \alpha, -2\alpha$，其余一切皆由此而来。至于 (ii)，题目中的式子看起来就像 (i) 的答案，只是 q 和 r 换成了较复杂的东西。这提示我们把 x 平移以消去 x^2 项，然后对新的三次多项式应用 (i)。

解答

由于 a, b, c 是根，而首项系数为 1，

$$x^3 - px^2 + qx - r = (x-a)(x-b)(x-c) = x^3 - (a+b+c)x^2 + (ab+bc+ca)x - abc$$

对所有 x 成立。比较系数，得

$$p = a+b+c, \qquad q = ab+bc+ca, \qquad r = abc.$$

第 (i) 部分

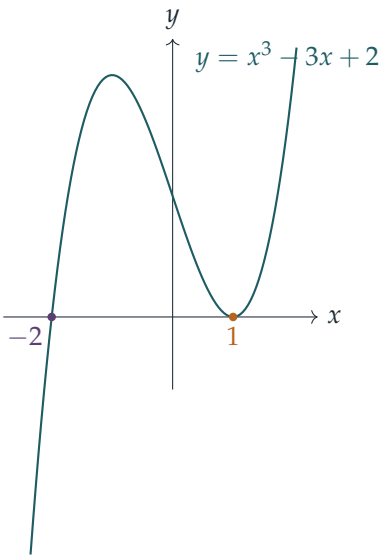
把各根标记成重根为 $a = b = \alpha$。那么由 $p = 0$ 得 $c = -(a+b) = -2\alpha$，所以三个根是 $\alpha, \alpha, -2\alpha$。因此

$$q = \alpha^2 + \alpha(-2\alpha) + (-2\alpha)\alpha = -3\alpha^2,$$
$$r = \alpha \cdot \alpha \cdot (-2\alpha) = -2\alpha^3.$$

于是

$$4q^3 + 27r^2 = 4(-27\alpha^6) + 27(4\alpha^6) = -108\alpha^6 + 108\alpha^6 = 0,$$

这正是所要证明的。这里完全不需要 α 是实数，而三重根的情况（$\alpha = 0$，所有系数为零）亦已包括在内。



例如，$x^3 - 3x + 2 = (x-1)^2(x+2)$ 有 $\alpha = 1, q = -3, r = -2$，而 $4(-27) + 27 \cdot 4 = 0$。二重根在图像上表现为曲线与横轴相切。

第 (ii) 部分

现在 p 不一定为零。原来的三次多项式各根之和为 p，所以它们的平均值是 $p/3$。按平均值平移，设

$$x = y + \tfrac{p}{3}.$$

那么 $y = x - \frac{p}{3}$，所以关于 y 的新三次多项式的根是 $a - \frac{p}{3}, b - \frac{p}{3}, c - \frac{p}{3}$。它们的和是 $p - p = 0$，而且若 a, b, c 之中有两个相等，这三个数之中亦有两个相等。

> **想一想** 关于 y 的三次多项式究竟是甚么？

把每一项展开，

$$\begin{aligned}
\left(y + \tfrac{p}{3}\right)^3 &= y^3 + py^2 + \tfrac{p^2}{3}y + \tfrac{p^3}{27}, \\
-p\left(y + \tfrac{p}{3}\right)^2 &= -py^2 - \tfrac{2p^2}{3}y - \tfrac{p^3}{9}, \\
q\left(y + \tfrac{p}{3}\right) &= qy + \tfrac{pq}{3},
\end{aligned}$$

再把这些连同 $-r$ 相加，y^2 项互相抵消（这正是平移的目的），得

$$x^3 - px^2 + qx - r = y^3 + \left(q - \frac{p^2}{3}\right)y - \left(\frac{2p^3}{27} - \frac{pq}{3} + r\right),$$

当中用到 $\frac{p^3}{27} - \frac{p^3}{9} = -\frac{2p^3}{27}$。这个关于 y 的三次多项式具有 $y^3 - Py^2 + Qy - R$ 的形式，其中

$$P = 0, \qquad Q = q - \frac{p^2}{3}, \qquad R = \frac{2p^3}{27} - \frac{pq}{3} + r.$$

它的 y^2 系数为零，并且有两个相等的根，所以第 (i) 部分适用，得出 $4Q^3 + 27R^2 = 0$，即

$$4\left(q - \frac{p^2}{3}\right)^3 + 27\left(\frac{2p^3}{27} - \frac{pq}{3} + r\right)^2 = 0,$$

这就是所求的结果。

验算

- 当 $p = 0$ 时，(ii) 的结果简化为 $4q^3 + 27r^2 = 0$，即第 (i) 部分。
- 对于 $(x-1)^2(x-2) = x^3 - 4x^2 + 5x - 2$，我们有 $p = 4, q = 5, r = 2$，所以 $Q = -\frac{1}{3}$ 而 $R = \frac{128}{27} - \frac{20}{3} + 2 = \frac{2}{27}$。于是 $4Q^3 + 27R^2 = -\frac{4}{27} + \frac{4}{27} = 0$。
- 我们亦以精确的有理数运算，在数百个随机选取、带有重根的三次多项式上测试了这两个恒等式，并在随机选取的点上检验了平移恒等式。

常见错误

- 韦达定理（Vieta's formulae）中的正负号出错：本题选用的符号是 $-p, +q, -r$，因此 p, q, r 三者以根表示时全部带正号。
- 按 $p/3$ 平移时弄错方向。新三次多项式的根是 $a - \frac{p}{3}$ 等等，正因如此它们的和才是零。
- 在 (ii) 中硬用 a, a, c 把所有式子展开。这样做行得通，但过程冗长又容易出错；平移则可不费工夫地重用 (i)。

STEP 之外

对一般的三次多项式，(ii) 的左方等于 $-(a-b)^2(b-c)^2(c-a)^2$，即判别式（discriminant）的相反数。它恰好在有两个根重合时等于零，而它的正负号则告诉你一个实系数三次多项式是有三个实根还是只有一个。平移 $x = y + \frac{p}{3}$ 亦是以卡尔达诺公式（Cardano's formula）解三次方程的第一步。根的对称函数、判别式，以及哪些方程可用根式求解（solved by radicals）的问题，都通往伽罗瓦理论（Galois theory）；Allenby 的 *Rings, Fields and Groups*（环、域与群）是很好的入门读物。

详细解答

STEP II 1992，第 8 题

纯数学 · 积分技巧

STEP II 1992，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/92-s2-q8

关键思路

三个标准技巧，每部分各用一个。在 (i) 中，我们把分母彻底因式分解（factorise），并留意到有一个重复因式，因此部分分式（partial fractions）需要包含平方项。在 (ii) 中，代换（substitution）$t = \tan\frac{x}{2}$ 可把任何关于 $\sin x$ 和 $\cos x$ 的有理函数（rational function）化为 t 的有理函数，而这里所得的二次式刚好可以漂亮地分解。在 (iii) 中，我们把 $\sinh x$ 写成指数形式，再代入 $u = \mathrm{e}^x$，结果同样归结为部分分式。

解答

第 (i) 部分

由于 $x^2 - 1 = (x-1)(x+1)$，分母是 $(x-1)^2(x+1)$。因式 $x-1$ 重复出现，所以我们设

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

即

$$x = A(x-1)(x+1) + B(x+1) + C(x-1)^2.$$

代入 $x = 1$ 得 $1 = 2B$，所以 $B = \frac{1}{2}$。代入 $x = -1$ 得 $-1 = 4C$，所以 $C = -\frac{1}{4}$。比较 x^2 的系数得 $0 = A + C$，所以 $A = \frac{1}{4}$。因此

$$\begin{aligned}\int \frac{x}{(x-1)(x^2-1)}\,\mathrm{d}x &= \int \left(\frac{1}{4(x-1)} + \frac{1}{2(x-1)^2} - \frac{1}{4(x+1)}\right)\mathrm{d}x \\ &= \tfrac{1}{4}\ln|x-1| - \frac{1}{2(x-1)} - \tfrac{1}{4}\ln|x+1| + c \\ &= \frac{1}{4}\ln\left|\frac{x-1}{x+1}\right| - \frac{1}{2(x-1)} + c.\end{aligned}$$

第 (ii) 部分

设 $t = \tan\frac{x}{2}$。则

$$\cos x = \frac{1-t^2}{1+t^2}, \qquad \sin x = \frac{2t}{1+t^2}, \qquad \mathrm{d}x = \frac{2\,\mathrm{d}t}{1+t^2},$$

最后一式是因为 $\mathrm{d}t/\mathrm{d}x = \frac{1}{2}\sec^2\frac{x}{2} = \frac{1}{2}(1+t^2)$。所以

$$3\cos x + 4\sin x = \frac{3-3t^2+8t}{1+t^2} = \frac{(1+3t)(3-t)}{1+t^2},$$

而因式 $1+t^2$ 互相约去：

$$\int \frac{\mathrm{d}x}{3\cos x + 4\sin x} = \int \frac{2\,\mathrm{d}t}{(1+3t)(3-t)}.$$

? **想一想** 我们怎样知道二次式 $3+8t-3t^2$ 可以因式分解？

它的判别式（discriminant）是 $64+36 = 100 = 10^2$，是一个完全平方数，所以它的根 $t = 3$ 和 $t = -\frac{1}{3}$ 都是有理数。（这并非巧合：$3^2+4^2 = 5^2$，这也正是下文的谐波形式（harmonic form）振幅为整数的原因。）

现在写 $\dfrac{2}{(1+3t)(3-t)} = \dfrac{a}{1+3t} + \dfrac{b}{3-t}$，于是 $2 = a(3-t)+b(1+3t)$。代入 $t=3$ 得 $b = \frac{1}{5}$；代入 $t = -\frac{1}{3}$ 得 $\frac{10}{3}a = 2$，所以 $a = \frac{3}{5}$。因此

$$\begin{aligned}\int \frac{\mathrm{d}x}{3\cos x + 4\sin x} &= \int \left(\frac{3}{5(1+3t)} + \frac{1}{5(3-t)}\right)\mathrm{d}t = \tfrac{1}{5}\ln|1+3t| - \tfrac{1}{5}\ln|3-t| + c \\ &= \frac{1}{5}\ln\left|\frac{1+3\tan\frac{x}{2}}{3-\tan\frac{x}{2}}\right| + c.\end{aligned}$$

另一方法：写 $3\cos x + 4\sin x = 5\sin(x+\beta)$，其中 $\cos\beta = \frac{4}{5}$ 及 $\sin\beta = \frac{3}{5}$。则该积分为 $\frac{1}{5}\int \operatorname{cosec}(x+\beta)\,\mathrm{d}x = \frac{1}{5}\ln\left|\tan\frac{1}{2}(x+\beta)\right| + c$。两个答案是一致的：由于 $\tan\frac{\beta}{2} = \frac{1}{3}$（因为 $\tan\frac{\beta}{2} = \frac{\sin\beta}{1+\cos\beta} = \frac{3/5}{9/5}$），由复角公式（addition formula）得

$$\tan\tfrac{1}{2}(x+\beta) = \frac{t+\frac{1}{3}}{1-\frac{1}{3}t} = \frac{1+3t}{3-t}.$$

第 (iii) 部分

由于 $\sinh x = \frac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x})$，把分子和分母同乘以 e^x 得

$$\frac{1}{\sinh x} = \frac{2\mathrm{e}^x}{\mathrm{e}^{2x}-1}.$$

代入 $u = \mathrm{e}^x$，于是 $\mathrm{d}u = \mathrm{e}^x\,\mathrm{d}x$：

$$\int \frac{\mathrm{d}x}{\sinh x} = \int \frac{2\,\mathrm{d}u}{u^2-1} = \int \left(\frac{1}{u-1} - \frac{1}{u+1}\right)\mathrm{d}u = \ln\left|\frac{u-1}{u+1}\right| + c$$
$$= \ln\left|\frac{\mathrm{e}^x-1}{\mathrm{e}^x+1}\right| + c = \ln\left|\tanh\frac{x}{2}\right| + c,$$

其中最后一步是把分子和分母同除以 $\mathrm{e}^{x/2}$：$\dfrac{\mathrm{e}^{x/2}-\mathrm{e}^{-x/2}}{\mathrm{e}^{x/2}+\mathrm{e}^{-x/2}} = \tanh\frac{x}{2}$。留意这与 (ii) 的另一方法中用到的 $\int \operatorname{cosec} x\,\mathrm{d}x = \ln|\tan\frac{x}{2}| + c$ 何其相似。

验算

把每个答案求导。例如在 (iii) 中，$\frac{\mathrm{d}}{\mathrm{d}x}\ln\tanh\frac{x}{2} = \frac{\frac{1}{2}\operatorname{sech}^2\frac{x}{2}}{\tanh\frac{x}{2}} = \frac{1}{2\sinh\frac{x}{2}\cosh\frac{x}{2}} = \frac{1}{\sinh x}$。数值上，用辛普森法则（Simpson's rule）计算，(i) 的被积函数（integrand）的 $\int_2^5$ 为 0.548287，(ii) 的被积函数的 $\int_0^1$ 为 0.234276，而 $\int_{1/2}^2 \operatorname{cosech} x\,\mathrm{d}x$ 为 1.134488，各自与原函数（antiderivative）所得的值吻合至最少 12 个小数位。

常见错误

- 在 (i) 中，看漏了重复因式而只尝试 $\frac{A}{x-1} + \frac{C}{x+1}$，又或者根本忘记了 x^2-1 可以因式分解。
- 漏掉对数内的绝对值符号：当 $x-1$、$1+3t$ 和 $\tanh\frac{x}{2}$ 取负值时，被积函数同样有定义。
- 在 (ii) 中，忘记了来自 $\mathrm{d}x$ 的因子 $\frac{2}{1+t^2}$。

STEP 之外

代换 $t = \tan\frac{x}{2}$（有时称为魏尔斯特拉斯代换，Weierstrass substitution）说明了每一个关于 $\sin x$ 和 $\cos x$ 的有理函数都有初等原函数（elementary antiderivative），因为部分分式可以处理 t 的任何有理函数。它背后是以有理函数对单位圆作出的参数化（parametrisation），而正是这个参数化生成了所有毕氏三元数组（Pythagorean triples）。至于究竟哪些函数才有初等原函数，这类问题会引向刘维尔定理（Liouville's theorem）和微分代数（differential algebra）；关于积分的分析基础，可参阅 Abbott 的 *Understanding Analysis*。

详细解答

STEP II 1992，第 9 题

纯数学 · 纯量三重积

> **STEP II 1992，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/92-s2-q9

关键思路

把 $\mathbf{r}$ 的表达式重新整理成 $\mathbf{a}$ 加上 $\mathbf{b}-\mathbf{a}$ 与 $\mathbf{c}-\mathbf{a}$ 的组合，便可看出它是通过 A、B 和 C 的平面。该平面的一个法向量（normal）是平面上两个向量的向量积（cross product），而把这个向量积展开，恰好得出题目中的三个三重积（triple product）。此后，由于恒等式对*所有* λ 和 μ 都成立，我们可以选取方便的数值。

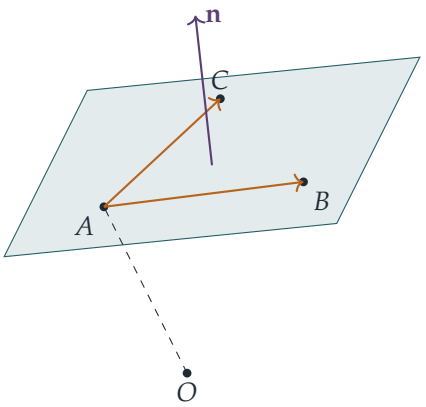
解答

轨迹

把方程改写成

$$\mathbf{r}=\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})+\mu(\mathbf{c}-\mathbf{a}).$$

这就是点 A 加上向量 $\overrightarrow{AB}=\mathbf{b}-\mathbf{a}$ 与 $\overrightarrow{AC}=\mathbf{c}-\mathbf{a}$ 的任意组合。这两个向量并不平行：假如平行，A、B、C 便会共线，而一条直线连同 O 必定位于某个平面上，这与题设不符。因此，当 λ 和 μ 取遍所有实数时，$\mathbf{r}$ 扫出的就是**通过 A、B 和 C 的平面**。取 $(\lambda,\mu)=(0,0)$、$(1,0)$、$(0,1)$ 便分别得到 A、B、C 三点本身。由于 O 不在平面 ABC 上，这个平面不通过原点。



方程 $\mathbf{r}\cdot\mathbf{n}=p$

平面的法向量同时垂直于 $\overrightarrow{AB}$ 和 $\overrightarrow{AC}$，所以我们取

$$\mathbf{n}=(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{a}),$$

由于这两个向量不平行，它不是零向量。展开，并利用 $\mathbf{a}\times\mathbf{a}=\mathbf{0}$ 和 $\mathbf{u}\times\mathbf{v}=-\mathbf{v}\times\mathbf{u}$，得

$$\begin{aligned}\mathbf{n}&=\mathbf{b}\times\mathbf{c}-\mathbf{b}\times\mathbf{a}-\mathbf{a}\times\mathbf{c}+\mathbf{a}\times\mathbf{a}\\&=\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b}.\end{aligned}$$

平面上每一点与 A 相差的向量都垂直于 $\mathbf{n}$，所以 $\mathbf{r}\cdot\mathbf{n}=\mathbf{a}\cdot\mathbf{n}$。由于 $\mathbf{a}$ 垂直于 $\mathbf{c}\times\mathbf{a}$，亦垂直于 $\mathbf{a}\times\mathbf{b}$，

$$p=\mathbf{a}\cdot\mathbf{n}=\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}).$$

因此平面的方程是 $\mathbf{r}\cdot(\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}+\mathbf{a}\times\mathbf{b})=\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$。

? **想一想** 把 $\mathbf{r}$ 的参数形式（parametric form）代入后，$\mathbf{r}\cdot\mathbf{n}$ 是甚么？

我们需要 $\mathbf{a}\cdot\mathbf{n}$、$\mathbf{b}\cdot\mathbf{n}$ 和 $\mathbf{c}\cdot\mathbf{n}$。在每一个之中，三项里都有两项消失，因为一个向量必定垂直于任何包含它的向量积：

$$\mathbf{a}\cdot\mathbf{n}=\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}),\qquad\mathbf{b}\cdot\mathbf{n}=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}),\qquad\mathbf{c}\cdot\mathbf{n}=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}).$$

对于所有 λ、μ，点 $\mathbf{r}=(1-\lambda-\mu)\mathbf{a}+\lambda\mathbf{b}+\mu\mathbf{c}$ 都在平面上，所以由 $\mathbf{r}\cdot\mathbf{n}=p$ 得

$$(1-\lambda-\mu)\,\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})+\lambda\,\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})+\mu\,\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})=\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}).$$

两边同时减去 $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$，得

$$-(\lambda+\mu)\,\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})+\lambda\,\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})+\mu\,\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})=0$$

对所有纯量 λ、μ 都成立，这正是所要证明的。

推论

由于恒等式对所有 λ 和 μ 都成立，可选取能把个别项分离出来的数值。

- $\lambda = 1, \mu = 0$：$-\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) + \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) = 0$。
- $\lambda = 0, \mu = 1$：$-\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) + \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = 0$。

两式合起来便得

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}).$$

（换一个等价的说法：左边是 λ 和 μ 的线性式，而且恒等于零，所以 λ 的系数和 μ 的系数都必须为零。）

当 A、B、C 和 O 共面时

这时 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 全都位于同一个通过 O 的平面上。$\mathbf{b}\times\mathbf{c}$ 垂直于该平面（或为零向量），所以 $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = 0$，其余两个三重积同样为零。从几何上看，以 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 为棱的平行六面体（parallelepiped）是扁平的，体积为零。结果依然成立，但只是平凡地成立：$0 = 0 = 0$。若 A、B、C 不共线，轨迹仍然是一个平面，但它通过原点，所以 $p = 0$，其方程为 $\mathbf{r}\cdot\mathbf{n} = 0$。若 A、B、C 共线，则 $\mathbf{n} = \mathbf{0}$，轨迹只是一条直线（若三点重合，则只是一个点），所以这个论证得不出任何信息。

验算

- 取 $\mathbf{a} = \mathbf{i}$、$\mathbf{b} = \mathbf{j}$、$\mathbf{c} = \mathbf{k}$，三个三重积都等于 1，而 $\mathbf{n} = \mathbf{i}+\mathbf{j}+\mathbf{k}$，$p = 1$：平面为 $x + y + z = 1$，它确实包含这三点。
- 我们以精确分数运算测试了 2000 组随机整数向量 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$：$\mathbf{n}$ 的展开式、p 的值、对随机 λ、μ 的方程 $\mathbf{r}\cdot\mathbf{n} = p$，以及三个三重积相等，全部精确成立。对于共面的向量组，每个三重积都是 0。

常见错误

- 展开 $(\mathbf{b}-\mathbf{a})\times(\mathbf{c}-\mathbf{a})$ 时出现正负号错误：$-\mathbf{b}\times\mathbf{a}$ 一项变成 $+\mathbf{a}\times\mathbf{b}$，而 $-\mathbf{a}\times\mathbf{c}$ 变成 $+\mathbf{c}\times\mathbf{a}$。
- 忘记说明为何 $\mathbf{n}\neq\mathbf{0}$，而这正是用上不共面条件（经由 A、B、C 不共线）的地方。

STEP 之外

纯量三重积（scalar triple product）$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ 就是以 **a**、**b**、**c** 为各行的 3×3 行列式（determinant），而这里证明的循环对称性，正是“把各行作循环置换（即两次对换）不会改变行列式的值”这一事实。它的值是以这三个向量为棱的平行六面体的有向体积，并且恰好在三个向量线性相关（linearly dependent）时为零，也就是共面的情况。在大学一年级的课程中，这会发展成把行列式视为体积缩放因子的一般理论；Strang 的 *Introduction to Linear Algebra*（线性代数入门）对此有清晰的阐述。

详细解答

STEP II 1992，第 10 题

纯数学 • 阿根图上的轨迹

> **STEP II 1992，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/92-s2-q10

关键思路

把式子写成 $\dfrac{z-1}{z} = \dfrac{1-z}{0-z}$：它的辐角（argument）就是由方向 $z \to 0$ 逆时针转到方向 $z \to 1$ 的角。因此，固定辐角就等于固定由 0 到 1 的线段在 z 处所对的角，接下来便可交给圆的定理（circle theorems）处理（同弓形内的圆周角相等、圆内接四边形对角互补）。最后一部分用的是同一个想法，只是把由 0 到 1 的线段换成弦 z_4z_2。

解答

（题目的叙述有一处误植：条件应为 $0 < \alpha \leq \frac{1}{2}\pi$。）

整个解答全靠两个事实。第一，由于 $\arg\dfrac{z-1}{z} = \arg(1-z) - \arg(0-z)$，这个辐角量度的是在 z 处由直线 $z0$ 逆时针转到直线 $z1$ 的角。第二，它告诉我们 z 位于实轴（real axis）的哪一侧：

$$\operatorname{Im}\frac{z-1}{z} = \operatorname{Im}\left(1-\frac{1}{z}\right) = \operatorname{Im}\left(-\frac{\bar{z}}{|z|^2}\right) = \frac{\operatorname{Im} z}{|z|^2}.$$

所以 $\arg\frac{z-1}{z} \in (0,\pi)$ 当且仅当 $\operatorname{Im} z > 0$，而它落在 $(-\pi,0)$ 内当且仅当 $\operatorname{Im} z < 0$。实轴上的点给出的商是实数（辐角为 0 或 π），而 $z = 0, 1$ 则根本没有辐角。

第 (i) 部分

这里 $\operatorname{Im} z > 0$，而由 0 到 1 的线段在 z 处所对的角为 α。在实轴上方，使一条固定的弦（chord）对着固定角度的所有点，组成一条通过弦两端的圆弧（同弓形内的圆周角相等）。反过来，该弧上每一点看这条弦的角都是 α，而且因为该点位于弦的上方，这个角是逆时针方向转的，所以整条弧都属于所求的轨迹（locus）。

要确定这个圆：圆心角为 2α，所以长度为 1 的弦给出半径 $\dfrac{1}{2\sin\alpha}$，而圆心位于 $\operatorname{Re} z = \frac{1}{2}$ 上，高度为 $\frac{1}{2}\cot\alpha \geq 0$。轨迹就是圆

$$\left|z - \tfrac{1}{2} - \tfrac{\mathrm{i}}{2}\cot\alpha\right| = \frac{1}{2\sin\alpha}$$

位于实轴上方的那段弧，但不包括端点 0 和 1。由于 $\alpha \leq \frac{1}{2}\pi$，这是优弧（major arc）。当 $\alpha = \frac{1}{2}\pi$ 时，它就是以 0 到 1 为直径的上半圆（半圆上的圆周角）。

第 (ii) 部分

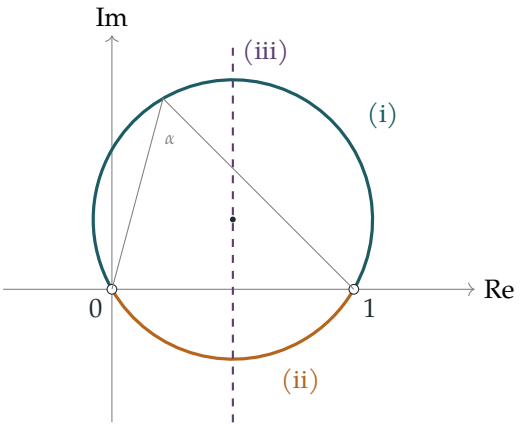
现在辐角 $\alpha - \pi$ 落在 $(-\pi, -\frac{1}{2}\pi]$ 内，所以 $\operatorname{Im} z < 0$，而弦在 z 处所对的角大小为 $\pi - \alpha$，方向为顺时针。

想一想 在实轴下方，弦所对的角为 $\pi - \alpha$ 的是哪一个圆?

就是 (i) 中的同一个圆。若 P 在 (i) 的弧上，而 Q 在该圆的另一段弧上，则 $0PQ1$（按适当的次序排列）是一个圆内接四边形（cyclic quadrilateral），其中 P 与 Q 是对角，所以 P 和 Q 处的角之和为 π。因此，实轴下方的劣弧（minor arc）正正就是角为 $\pi - \alpha$ 的地方。轨迹是 (i) 中的圆位于实轴下方的那段弧，同样不包括 0 和 1。(i) 和 (ii) 两部分合起来，就是除去 0 和 1 之后的整个圆。（若 $\alpha = \frac{1}{2}\pi$，(ii) 就是下半圆。）

第 (iii) 部分

$\left|\frac{z-1}{z}\right| = 1$ 即 $|z-1| = |z|$ 且 $z \neq 0$：z 与 0 和 1 等距。轨迹是垂直平分线（perpendicular bisector），即直线 $\operatorname{Re} z = \frac{1}{2}$。



草图所示为 $\alpha = \frac{1}{3}\pi$ 的情况：蓝绿色的优弧是 (i)，琥珀色的劣弧是 (ii)，紫红色的虚线是 (iii)，空心圆圈标示被剔除的点 0 和 1。

最后一部分

设 $f(z)=\dfrac{z-z_2}{z-z_4}$。由于 $(z_1-z_4)(z_3-z_2)=(z_4-z_1)(z_2-z_3)$，

$$\frac{f(z_1)}{f(z_3)}=\frac{(z_1-z_2)(z_3-z_4)}{(z_1-z_4)(z_3-z_2)}=\frac{(z_1-z_2)(z_3-z_4)}{(z_4-z_1)(z_2-z_3)}=w,$$

所以 $\arg w=\arg f(z_1)-\arg f(z_3)$（模 2π）。

函数 f 其实就是 (i) 和 (ii) 两部分的那个函数，只是换了坐标。设 $z=z_4+(z_2-z_4)u$，这是先旋转（rotation）和放大（enlargement），再平移（translation），把 $u=0,1$ 分别送到 z_4,z_2。于是

$$f(z)=\frac{(z_2-z_4)(u-1)}{(z_2-z_4)u}=\frac{u-1}{u}.$$

通过这四点的圆，对应于 u 平面上一个通过 0 和 1 的圆。

> ? **想一想** 为甚么 z_1 和 z_3 位于不同的弧上?

这些点沿圆周依次为 z_1,z_2,z_3,z_4，所以弦 z_2z_4 把圆分成两段弧，一段包含 z_1，另一段包含 z_3。在 u 平面上，其中一段位于实轴上方，另一段位于实轴下方。

由 (i) 和 (ii) 可知，$\frac{u-1}{u}$ 的辐角在上方的弧上取常数值 $\beta\in(0,\pi)$，在下方的弧上取值 $\beta-\pi$。这里我们并没有假设 $\beta\le\frac12\pi$：上面运用圆的定理的论证从未用过这个限制。因此

$$\arg w=\arg f(z_1)-\arg f(z_3)=\pm\big(\beta-(\beta-\pi)\big)=\pm\pi\pmod{2\pi},$$

所以 w 是实数，而且其实是负数。 ■

验算

对于正方形 $z_1=1$、$z_2=\mathrm{i}$、$z_3=-1$、$z_4=-\mathrm{i}$，我们得到 $w=\dfrac{(1-\mathrm{i})(-1+\mathrm{i})}{(-1-\mathrm{i})(1+\mathrm{i})}=\dfrac{-(1-\mathrm{i})^2}{-(1+\mathrm{i})^2}=\dfrac{-2\mathrm{i}}{2\mathrm{i}}=-1$，确实是负实数。至于 (i)，取 $\alpha=\frac12\pi$ 和 $z=\frac12+\frac12\mathrm{i}$。则 $\frac{z-1}{z}=\frac{-1+\mathrm{i}}{1+\mathrm{i}}=\mathrm{i}$，其辐角为 $\frac12\pi$。我们亦以数值方法验证过：对若干个 α 的值，两段弧上的辐角确如上文所述；并且在随机的圆上依次随机取四点，共 10^4 组，w 都是负实数。

常见错误

- 把整个圆当作 (i) 的答案。只有一段弧的辐角是对的；另一段弧给出的是 $\alpha-\pi$，那正是第 (ii) 部分。
- 忘记剔除 0 和 1，在这两点辐角没有定义。

- 因为处理 $(z_4 - z_1)(z_2 - z_3)$ 中的负号时出错，而弄错 $\arg w$ 的正负号。像上面那样把它们配对，就可避免这个问题。

STEP 之外

w 这个量（除了正负号和次序上的差别外）就是四个点的交比（cross-ratio）。它在每一个莫比乌斯变换（Möbius transformation）$z \mapsto \frac{az+b}{cz+d}$ 之下都保持不变，而四点共圆或共线，当且仅当它们的交比是实数。由于莫比乌斯映射把圆和直线送到圆和直线，这就提供了一个证明四点共圆（concyclicity）的巧妙方法。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）把这种几何观点发挥得淋漓尽致。

详细解答

STEP II 1992，第 11 题

力学 • 抛体运动

> **STEP II 1992，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/92-s2-q11

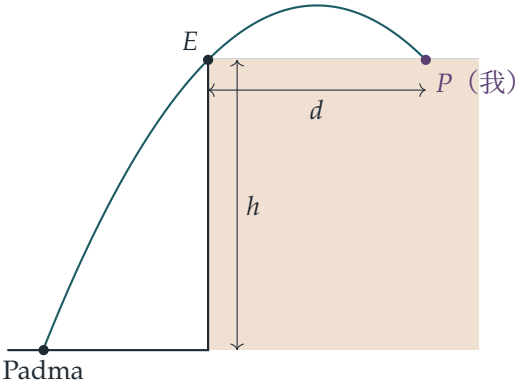
关键思路

这一掷的难处不在于把银包抛到高度 h，而在于让它越过崖顶：由崖边到我所站之处，水平距离至少为 d，银包在这一整段路程中的高度都不得低于 h。因此，我们只考虑银包首次到达高度 h 那一刻之后的运动。由那一刻起，它就是一个在水平地面（即崖顶所在的水平面）上的普通抛体（projectile），其速率可由能量守恒（conservation of energy）求得，余下的工作交给大家熟悉的最大射程（maximum range）u^2/g 便可。

解答

模型

把银包视为一个只受重力作用而自由运动的质点（particle），不计空气阻力，并设它由地面抛出（Padma 的身高可忽略不计）。设 E 为悬崖的顶边，P 为我所站的位置，因此 P 的高度为 h，并在 E 之后水平距离 d 处。崖面是铅垂的，而沙滩在崖面的另一侧，与我分处两边，所以无论 Padma 站在哪里，银包都必须越过 E（最差的情况是刚好擦过它），然后落到 P。



以下会反复用到一个事实。若一个质点在某一水平面上的速率为 u，运动方向与水平线成仰角 α，则它经过时间 $2u\sin\alpha/g$ 后回到该水平面，其间的水平位移为

$$\frac{2u^2\sin\alpha\cos\alpha}{g}=\frac{u^2\sin 2\alpha}{g}\le\frac{u^2}{g},$$

等号成立当且仅当 $\alpha=45°$。

条件是必要的

假设某一掷能把银包送到 P。银包必定到达高度 h，设 Q 为轨迹上首个高度为 h 的点。取崖脚为势能（potential energy）的零点，由能量守恒可得银包在 Q 的速率 u：

$$u^2=V^2-2gh.$$

想一想 为甚么 Q 与 P 的水平距离必定至少为 d?

当银包的水平位置与崖面相同时，它的高度必须至少为 h，否则便会撞上崖面（或在崖边之下经过）。它的高度由 0 开始连续上升，所以在此之前它已经到达高度 h：Q 位于 E 靠沙滩的一侧，或就在 E。由于 P 在 E 之后水平距离 d 处，Q 到 P 的水平距离至少为 d。

在 Q 与 P 之间，银包以速率 u 出发，由高度 h 回到高度 h。根据上述关于射程的事实，它所走过的水平距离至多为 u^2/g。因此

$$d\le\frac{u^2}{g}=\frac{V^2-2gh}{g},\qquad\text{即}\qquad V^2\ge g(2h+d).$$

条件是充分的

现在假设 $V^2\ge g(2h+d)$。则 $u^2=V^2-2gh\ge gd>0$，所以 $gd/u^2\le 1$，于是可以选取一个角 α，满足 $0<\alpha\le 45°$ 及

$$\sin 2\alpha=\frac{gd}{u^2}.$$

考虑这样一条抛物线（parabola）轨迹：它以速率 u 经过 E，以仰角 α 朝远离沙滩的方向运动。根据关于射程的事实，它在 E 之后水平距离 $u^2 \sin 2\alpha / g = d$ 处回落到高度 h，那正是 P。在 E 与 P 之间，它高于高度 h，所以能越过崖顶。

把这条轨迹由 E 起在时间上倒推回去：在较早的时刻，银包的位置较低，而且在沙滩上方离崖面较远（当时它的铅垂速度较大，而且向上），所以轨迹会在沙滩上某一点与地面相交，该点与崖脚的水平距离为正数。在到达 E 之前，轨迹完全位于崖面靠沙滩的一侧，所以途中没有任何阻碍；又因为沙滩宽阔，Padma 可以站在该点。由能量守恒，银包在该点的速率为 $\sqrt{u^2 + 2gh} = V$。因此，若 Padma 在该点以速率 V 沿轨迹的方向抛出银包，银包便会沿着这条轨迹到达我手上。

（当 $V^2 = g(2h + d)$ 恰好成立时，$\alpha = 45°$，银包刚好擦过崖边 E；若 V 更大，则在 $(0°, 90°)$ 内有两个角可行，而且两者都令银包严格地在崖边上方经过。）

综合以上两部分，Padma 能把银包抛给我，当且仅当 $V^2 \geq g(2h + d)$，这正是所要证明的。

验算

- 若 $d = 0$，条件变成 $V^2 \geq 2gh$，这正是铅垂上抛到达高度 h 所需的条件。
- 若 $h = 0$，条件变成 $V^2 \geq gd$，即水平地面上常见的最大射程。
- 若不理会崖边，Padma 最有利的做法是站在崖脚，而所有轨迹的包络抛物线（bounding parabola）只给出 $V^2 \geq g\left(h + \sqrt{h^2 + d^2}\right)$；由于 $\sqrt{h^2 + d^2} \leq h + d$，这个条件较弱。所以崖边确实是起决定作用的约束。
- 数值上，取 $g = 9.8$、$h = 20$ 及 $d = 15$，临界速率为 $\sqrt{539} \approx 23.2\ \mathrm{m\,s^{-1}}$；Padma 站在离崖脚约 11.4 m 处，而对这一掷进行逐步模拟，银包落在崖边之后 15.0 m 处，并擦过 E。对随机选取的 h、V 扫描所有抛射角，结果证实轨迹上两个高度为 h 的点之间的最大水平间距为 $(V^2 - 2gh)/g$。

常见错误

- 只对由崖脚抛往点 P 的情况使用包络抛物线。这样做忽略了崖边，所得的条件是必要但不充分的。
- 只证明一个方向。“若”的方向需要明确构造出一掷，并检查它由沙滩出发，而且不会撞上悬崖。

STEP 之外

在高度 h 重新开始考虑运动这一招之所以可行，是因为运动方程（equations of motion）不随时间改变，在水平平移下亦保持不变，所以轨迹上任何一点都可以当作一次全新的抛射。“哪些目标可以到达?” 这类问题，讨论的是一族曲线的包络线（envelope）；若途中有障碍物，答案便取决于哪一个约束最先生效，这个想法在最优化（optimisation）和变分法（calculus of variations）中会再次出现。Kleppner 与 Kolenkow 合著的 *An Introduction to Mechanics*（力学入门教材）是进一步学习抛体运动和能量方法的好选择。

详细解答

STEP II 1992，第 12 题

力学 · 力矩与弹性绳

STEP II 1992，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/92-s2-q12

关键思路

每个轮子都可以自由转动，所以它处于平衡（equilibrium）的唯一条件，就是对轮心的力矩（moment）互相抵销：悬挂的质点（particle）要把轮子向一个方向转，而弹性绳（elastic string）则把它向另一个方向拉回去。由几何关系可以用 θ 表示弹性绳的伸长量（extension），于是由胡克定律（Hooke's law）得出张力（tension），再取力矩便得到一条关于 θ 的方程。因此，数算平衡位置的数目，就是数算这条方程在真正有意义的角度范围内有多少个解。

解答

建立模型

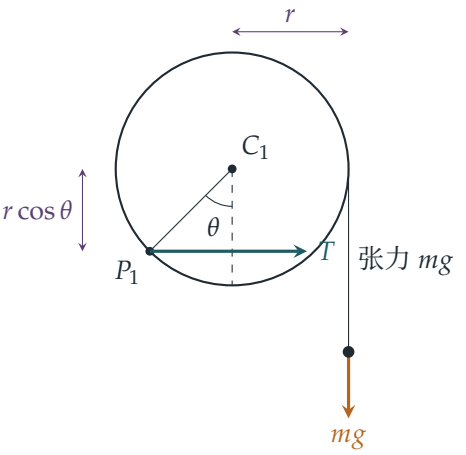
我们把各质点视为质点，把悬挂它们的绳视为轻（light）而不可伸长（inextensible）的绳，并把弹性绳视为轻而且服从胡克定律。轮轴是光滑（smooth）的，所以每个轮心处的反作用力对该轮心没有力矩。

取 C_1C_2 为水平。在图中，P_1 位于 C_1 外侧（左方）$r\sin\theta$ 处，并在其下方 $r\cos\theta$ 处；P_2 则是它的镜像。所以 P_1 和 P_2 处于同一高度，弹性绳是水平的，长度为

$$P_1P_2 = d + 2r\sin\theta.$$

它的自然长度（natural length）是 d，所以伸长量是 $2r\sin\theta$，而根据胡克定律，张力为

$$T = \frac{\lambda\,(2r\sin\theta)}{d}.$$



W_1 上的质点悬挂着并处于平衡，所以系着它的绳的张力是 mg；这条绳在较接近 W_2 的一侧沿切线方向离开轮子，与 C_1 的水平距离为 r。弹性绳以力 T 把 P_1 水平地拉向 W_2，而它的作用线在 C_1 下方 $r\cos\theta$ 处。

平衡条件

对 C_1 取力矩：重量使 W_1 顺时针转动（如图所示），力矩为 mgr；张力则使它逆时针转动，力矩为 $Tr\cos\theta$。所以

$$mgr = Tr\cos\theta = \frac{2\lambda r\sin\theta}{d}\,r\cos\theta = \frac{\lambda r^2\sin 2\theta}{d},$$

其中用了 $2\sin\theta\cos\theta = \sin 2\theta$。两边除以 λr^2，得

$$\sin 2\theta = \frac{mgd}{\lambda r},$$

这正是所要证明的结果。根据对称性，对 C_2 取力矩也会得到同一条方程。

哪些角度是容许的?

? **想一想** 为甚么 θ 不可以是任意角度，使得 $\sin 2\theta = k$ 有无限多个解?

张力的力矩必须与重量的力矩方向相反。若 $\sin\theta \le 0$，则 P_1 与 P_2 两点相距不超过 d，于是弹性绳松弛，$T = 0$，没有东西可以抵销 mgr。若 $\cos\theta \le 0$，则弹性绳与 C_1 同高或在其上方，所以它对 C_1 的力矩是零，或者与重量一样使轮子向同一方向转动。无论哪种情况都不会有平衡。因此，平衡必须满足

$$0 < \theta < \tfrac{\pi}{2}, \qquad \text{即} \qquad 0 < 2\theta < \pi.$$

（两个轮子可否处于不同的角度？张力对每个轮心的力矩，等于 T 乘以该轮心到直线 P_1P_2 的垂直距离，而两个力矩都必须等于 mgr。所以 C_1 和 C_2 与这条直线的距离相同，并且在直线的同一侧，而直线在两个轮心的下方。因此这条直线平行于 C_1C_2，所以 P_1 和 P_2 处于同一深度，两个角度亦相同，因为假如两个角度异号，弹性绳的长度便刚好是 d，因而松弛。所以每一个平衡位置都属于图中所示的对称类型。）

设 $k = \dfrac{mgd}{\lambda r} > 0$。在 $0 < 2\theta < \pi$ 上，$\sin 2\theta$ 由 0 上升，在 $2\theta = \frac{\pi}{2}$ 时达到 1，然后回落至 0，其间恰好两次取得 $(0,1)$ 内的每一个值，并恰好一次取得 1 这个值。所以方程 $\sin 2\theta = k$

- 在 $k > 1$ 时无解；
- 在 $k = 1$ 时恰有一个解 $\theta = \frac{\pi}{4}$；
- 在 $0 < k < 1$ 时恰有两个解 $\theta = \alpha$ 和 $\theta = \frac{\pi}{2} - \alpha$，其中 $\alpha = \frac{1}{2}\arcsin k \in (0, \frac{\pi}{4})$。

由于 $k > 1 \iff \lambda < \dfrac{mgd}{r}$，答案如下：

第 (i) 部分

当 $\lambda < \dfrac{mgd}{r}$ 时没有平衡位置：弹性绳太弱，在任何角度都无法拉住质点。

第 (ii) 部分

当 $\lambda = \dfrac{mgd}{r}$ 时只有一个平衡位置，此时 $\theta = \frac{\pi}{4}$。

第 (iii) 部分

当 $\lambda > \dfrac{mgd}{r}$ 时恰有两个平衡位置，分别在 $\theta = \alpha$ 和 $\theta = \frac{\pi}{2} - \alpha$。

第 (iv) 部分

没有任何 λ 的值可以给出多于两个平衡位置，因为 θ 被限制在 $(0, \frac{\pi}{2})$ 之内。

验算

- 量纲：λ 是力，mg 也是力，而 d/r 是两个长度之比，所以 $mgd/(\lambda r)$ 是无量纲的，正弦值理应如此。
- 极限情况：当 $\lambda \to \infty$ 时，弹性绳变得不可伸长，两个解分别趋向 $\theta \to 0$（伸长量极小，张力很大）和 $\theta \to \frac{\pi}{2}$（弹性绳几乎与轮心同高，所以力臂极小）。两者都合理。

- 数值验算：取 $m=1$、$d=4$、$r=1$、$g=9.8$ 和 $\lambda=3mgd/r$，于是 $k=\frac{1}{3}$；一个包含两个轮子的完整模型（容许两个角度不同）恰有两个平衡位置，分别在 $\theta\approx 0.1699$ 和 $\theta\approx 1.4009$；它们就是 $\frac{1}{2}\arcsin\frac{1}{3}$ 以及 $\frac{\pi}{2}$ 减去这个值。

常见错误

- 把伸长量当作 $r\sin\theta$ 而不是 $2r\sin\theta$：弹性绳的两端都向外移动了。
- 以 $r\sin\theta$ 作为张力的力臂。张力是水平的，所以它的力臂是垂直距离 $r\cos\theta$。
- 在 (iv) 中答“无限多个”，忘记了只有 $0<\theta<\frac{\pi}{2}$ 在物理上才有可能。

STEP 之外

研究平衡位置的数目如何随参数改变而变化，正是分岔理论（*bifurcation theory*）的起点。在这里，$\lambda>mgd/r$ 时的两个平衡位置在 $\lambda=mgd/r$ 时合而为一，然后消失，这称为鞍结分岔（saddle-node bifurcation），又称折叠分岔（fold bifurcation）。考察势能（potential energy）便会发现，这一对平衡位置中一个是稳定的，另一个是不稳定的，这正是折叠分岔的典型特征。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics* 讲述了力矩、刚体平衡，以及如何由势能判断稳定性。

详细解答

STEP II 1992，第 13 题

力学 · 摩擦力与连接质点

STEP II 1992，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：

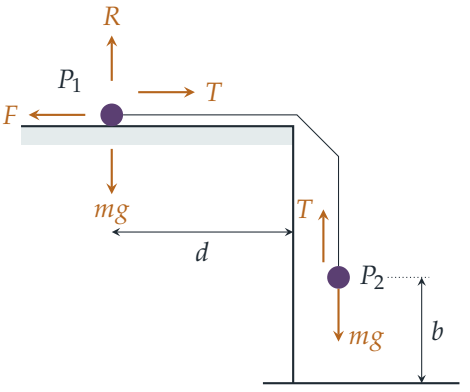
step.maths.org/questions/92-s2-q13

关键思路

整个运动分为两个阶段。P_2 下坠期间，绳子拉紧，两个质点（particle）以同一个加速度一起运动。P_2 一撞到地面，绳子便松弛，P_1 只会在摩擦力（friction）作用下滑行至停下。一切都取决于碰撞一刻的速率：第一部分是把停下所需的距离与距离桌边尚余的距离作比较，第二部分则是比较两段时间。

解答

我们把 P_1 和 P_2 视为质点，并把绳子视为轻而不可伸长（light and inextensible），绕过光滑的桌边，因此整条绳子的张力（tension）T 处处相同。桌面对 P_1 而言是粗糙的。我们假设 $b < d$，使得 P_2 在 P_1 到达桌边之前已经着地（绳子的长度此后再无作用）。



第一部分

第一阶段：绳子拉紧。对 P_1 沿竖直方向分解力，得 $R = mg$。它正在滑动，所以摩擦力达到极限值（limiting friction）：$F = \mu R = \mu mg$，方向与运动相反。设共同加速度为 a，

$$P_2 \text{ (向下)}: \quad mg - T = ma,$$
$$P_1 \text{ (朝向桌边)}: \quad T - \mu mg = ma.$$

两式相加，$2ma = mg(1-\mu)$，所以 $a = \frac{1}{2}g(1-\mu)$；由于 $\mu < 1$，这是正数，系统确实会开始运动。P_2 由静止下坠距离 b，所以它到达地面时的速率 V 满足

$$V^2 = 2ab = gb(1-\mu).$$

此刻 P_1 的速率同样是 V，而且它已移动了距离 b，所以它距离桌边 $d-b$。

第二阶段：绳子松弛。碰撞之后，P_2 在地面上或正在向上弹起，而 P_1 继续向桌边移动，所以绳子松弛，$T = 0$。

> **想一想** 为甚么 P_2 弹跳期间绳子不会再次拉紧?

碰撞一刻绳子刚好拉紧。其后 P_1 向桌边靠近，让更多绳子越过桌边垂下，而 P_2 不会低于地面，所以它所需的绳子不会比碰撞时多。绳子自此一直保持松弛。

因此 P_1 只以 μg 减速，并在再移动距离

$$s = \frac{V^2}{2\mu g} = \frac{b(1-\mu)}{2\mu},$$

之后停下，前提是它没有先到达桌边。故此 P_1 掉下桌子，当且仅当（if and only if）$s > d-b$：

$$\begin{aligned} \frac{b(1-\mu)}{2\mu} > d-b &\iff b - b\mu > 2\mu d - 2\mu b \qquad (\text{两边乘以} 2\mu > 0) \\ &\iff b > \mu(2d-b) \\ &\iff \mu < \frac{b}{2d-b}, \end{aligned}$$

其中最后一步是除以 $2d-b$，因为 $d > b$，这是正数。每一步都可逆，所以这正是所要求的“当且仅当”。（若 μ 等于 $b/(2d-b)$，P_1 刚好停在桌边，不会掉下。）

第二部分

现在 $\mu > b/(2d-b)$，所以 P_1 会停在桌上。由 P_2 第一次撞到地面的一刻起计时。在那一刻两个质点的速率都是 V。

P_1：由速率 V 起以 μg 减速，经过时间

$$t_1 = \frac{V}{\mu g}.$$

后停下。

P_2：根据牛顿恢复定律（Newton's law of restitution），它以速率 eV 反弹。由于绳子松弛，它只在重力作用下自由运动，所以当 $eVt - \frac{1}{2}gt^2 = 0$ 时它返回地面，即经过时间

$$t_2 = \frac{2eV}{g}.$$

（碰撞是瞬时的，所以第一次弹跳不耗时间。）

P_1 在第二次弹跳之前停下，正正就是 $t_1 < t_2$ 的时候：

$$\frac{V}{\mu g} < \frac{2eV}{g} \iff \frac{1}{\mu} < 2e \iff e > \frac{1}{2\mu},$$

这里除以了 $V/g > 0$。所以若 $e > 1/(2\mu)$，P_1 会在 P_2 第二次弹跳之前停下，正如所求。事实上逆命题（converse）亦成立。

验算

- 极限情况：当 $\mu \to 0$ 时，界限 $b/(2d-b)$ 显然得到满足，而 P_1 在光滑桌面上当然会滑出桌边。若 $b = d$，界限变成 1，所以在 $\mu < 1$ 之下它必定掉下；这是对的，因为 P_2 着地时 P_1 已在桌边，而且仍在运动。
- 由于 $\mu < 1$，第二个条件需要 $e > \frac{1}{2}$。地面几乎没有弹性时，弹跳很短，所以 P_2 会先再次着地，与我们的预期一致。
- 数值上，取 $d = 2$、$b = 1$、$\mu = 0.5$ 和 $e = 0.8$（取 $g = 9.8$），逐步模拟运动得出 $t_1 \approx 0.452\,\mathrm{s}$ 和 $t_2 \approx 0.361\,\mathrm{s}$，与 $V/(\mu g)$ 和 $2eV/g$ 在 $V^2 = 4.9$ 时的值吻合。这里 $e = 0.8 < 1/(2\mu) = 1$，所以 P_2 先再次弹跳。在逐步模拟中对两个判别条件所作的随机测试，每一次结果都吻合。

常见错误

- 在第二阶段仍保留张力：P_2 一着地，绳子便松弛，P_1 只受摩擦力减速。
- 忘记 P_1 已经移动了距离 b，把 s 与 d 比较，而不是与 $d-b$ 比较。
- 把系统的加速度写成 $g(1-\mu)$，忘记了 $mg(1-\mu)$ 这个力要令两个质量一起加速。

STEP 之外

这里的绳子是一个单侧约束（unilateral constraint）：它只能拉，不能推，所以绳子一松弛，运动方程便会转换。支配方程在碰撞时或失去接触时改变的系统，称为分段光滑

（piecewise-smooth）系统或混合（hybrid）系统，在工程动力学和机械人学中都有深入的研究。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门）对绳子、摩擦和碰撞有仔细的论述，是很好的下一步读物。

详细解答

STEP II 1992，第 14 题

力学 · 连接质点与滑轮

STEP II 1992，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/92-s2-q14

关键思路

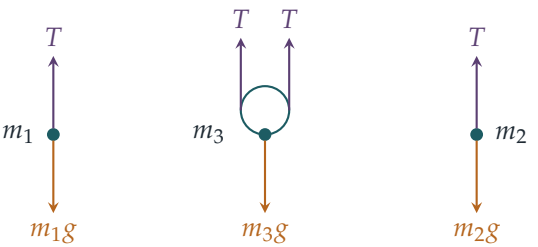
系统中有三个运动的物体，但只有一条绳，所以整条绳的张力（tension）都是同一个 T，而三个加速度（acceleration）之间只有一个约束条件。对每个质点写出牛顿第二定律（Newton's second law）（滑轮 P_3 与挂在它上面的质点一起运动），再写出绳长条件，便得到四个未知数的四条线性方程。至于静止的条件，不过是说 P_3 的加速度为零。

解答

模型

滑轮（pulley）光滑而轻，绳轻而不可伸长（inextensible），所以整条绳上的张力处处相同，都等于 T。由于 P_3 是轻滑轮，它与质量为 m_3 的质点合成一个质量为 m_3 的物体，由两段竖直的绳吊着，每段绳都以力 T 把它向上拉。

所有位移一律以向下为正：设 x_1、x_2、x_3 分别为质点 m_1、质点 m_2 和滑轮 P_3 在固定滑轮以下的深度，对应的加速度为 a_1、a_2、a_3（向下为正）。



运动方程

对每个物体沿竖直方向分解（向下为正）：

$$\begin{aligned} m_1 a_1 &= m_1 g - T, \\ m_2 a_2 &= m_2 g - T, \\ m_3 a_3 &= m_3 g - 2T. \end{aligned}$$

绳的约束条件

略去绕在滑轮上的固定长度不计，这条绳由以下各段组成：由 P_1 垂下至 m_1 的一段（长度 x_1）、由两个固定滑轮垂下至 P_3 的两段（每段长度 x_3），以及由 P_2 垂下至 m_2 的一段（长度 x_2）。由于绳不可伸长，

$$x_1 + x_2 + 2x_3 = \text{常数}, \qquad \text{所以} \qquad a_1 + a_2 + 2a_3 = 0$$

后一式由前一式求导两次而得。

> ? **想一想** 为甚么 x_3 的系数是 2 而不是 1？

当 P_3 下降距离 d 时，吊着它的两段绳都各自长了 d，所以共用去 $2d$ 的绳。这些绳只能来自外侧的两段，因此 $x_1 + x_2$ 减少 $2d$。漏掉这个因子，是这道题最常见的出错原因。

张力

把每条运动方程除以对应的质量，

$$a_1 = g - \frac{T}{m_1}, \qquad a_2 = g - \frac{T}{m_2}, \qquad a_3 = g - \frac{2T}{m_3}.$$

代入 $a_1 + a_2 + 2a_3 = 0$：

$$4g - T\left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{4}{m_3}\right) = 0.$$

全式乘以 $m_1 m_2 m_3$，再解出 T，

$$\boxed{T = \frac{4 m_1 m_2 m_3\, g}{m_3(m_1 + m_2) + 4 m_1 m_2}}$$

这个结果在整个运动过程中都成立（直至有质点碰到滑轮为止），因为方程中没有任何一项与位置有关。

P_3 保持静止的情况

系统由静止开始运动，所以 P_3 保持静止，当且仅当 $a_3 = 0$，即 $2T = m_3 g$。代入上面求得的 T：

$$\begin{aligned} \frac{8m_1 m_2 m_3\, g}{m_3(m_1+m_2)+4m_1m_2} &= m_3 g \\ \iff \quad 8m_1m_2 &= m_3(m_1+m_2)+4m_1m_2 \\ \iff \quad 4m_1m_2 &= m_3(m_1+m_2), \end{aligned}$$

其中我们除以了 $m_3 g > 0$，并乘以了为正数的分母。所以若 $4m_1m_2 = m_3(m_1+m_2)$，则 $a_3 = 0$，P_3 保持静止，正是所要证明的。（每一步都可逆，所以这个条件既是充分条件，也是必要条件。）

要留意，外侧的两个质点不一定静止：这时 $T = \frac{1}{2}m_3 g$，而 $a_1 = g - T/m_1 = -a_2$，除非 $m_1 = m_2$，否则不为零。一个质点上升，另一个以相同的快慢下降，绳在 P_3 下面滑过，却没有带动它。

验算

- 对称情况 $m_1 = m_2 = m$：条件变成 $4m^2 = 2mm_3$，即 $m_3 = 2m$。这不过是静力平衡，$2T = 2mg = m_3 g$：结果显而易见，也令人放心。
- 若 $m_3 \to \infty$，则 $T \to \dfrac{4m_1m_2 g}{m_1+m_2}$，是大家熟悉的阿特伍德机（Atwood machine）张力 $\dfrac{2m_1m_2 g}{m_1+m_2}$ 的两倍。这是合理的：很重的 P_3 近乎自由下落，$a_3 \to g$，所以各段绳加速得很快。若 $m_3 \to 0$，则 $T \to 0$，因为无质量的滑轮不能承受任何净力。
- 数值上，取 $m_1 = 2$、$m_2 = 3$、$m_3 = 4$ 及 $g = 9.8$（国际单位制），公式给出 $T \approx 21.38\,\mathrm{N}$。另外用能量方法（拉格朗日方法，Lagrangian method）独立求解，得到相同的加速度 $a_1 \approx -0.891$、$a_2 \approx 2.673$、$a_3 \approx -0.891\,\mathrm{m\,s^{-2}}$。取 $m_1 = 1$、$m_2 = 3$、$m_3 = 3$ 时，静止条件成立，而果然 $a_3 = 0$，同时 $a_1 = -a_2 = -4.9\,\mathrm{m\,s^{-2}}$。

常见错误

- 把约束条件写成 $a_1 + a_2 + a_3 = 0$，忘记了 P_3 是挂在两段绳上的。
- 忘记了作用在 P_3 上的力是 $2T$，而不是 T。
- 以为 P_3 静止，整个系统就必定静止。除非 $m_1 = m_2$，否则并非如此。

STEP 之外

约束条件 $x_1 + x_2 + 2x_3 =$ 常数 把系统的自由度（degrees of freedom）减至两个，而处理这类系统最简洁的方法是拉格朗日力学（Lagrangian mechanics）：只要用独立的坐标写出能量，就完全用不着张力（它是一种“约束力”（constraint force））。我们的数值验算正是这样做的。你会在大学一年级的动力学课程中学到它。想看看牛顿定律怎样严谨地应用于滑轮这类受约束的系统，Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）是个好去处。

详细解答

STEP II 1992，第 15 题

概率与统计 • 随机漫步

STEP II 1992，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/92-s2-q15

关键思路

若该点在 s 步之中有 R 步向正方向走，它最后会停在 $x = R - (s - R) = 2R - s$，而 R 服从二项分布（binomial distribution）。因此，这里每一个概率都是若干个二项式系数（binomial coefficients）之和。奇偶性（parity）帮了很大的忙：走了偶数步之后，只可能到达 $x = 2$ 和 $x = 4$；走了奇数步之后，只可能到达 $x = 3$ 和 $x = 5$，所以每个 $\mathrm{P}(s)$ 都只是两个相邻二项式系数之和。帕斯卡法则（Pascal's rule）随即把这个和化简为单一个系数。至于最大值，则要考察相邻两项之比（ratio），并把 s 为偶数和奇数的情况分开处理。

解答

设 R 为首 s 步中向正方向走的步数。各步互相独立（independent），每一步向正方向走的概率为 $\frac{1}{2}$，所以 $R \sim \mathrm{B}(s, \frac{1}{2})$，而且

$$\mathrm{P}(R = r) = \binom{s}{r}\left(\frac{1}{2}\right)^s.$$

走了 s 步之后的位置是 $X = 2R - s$，所以 $X = x$ 当且仅当 $R = \frac{1}{2}(s + x)$。特别地，X 的奇偶性与 s 相同。

$\mathrm{P}(5)$ 与 $\mathrm{P}(6)$ 的值

当 $s = 5$ 时，只有 $x = 3$ 和 $x = 5$ 可能出现，分别需要 $R = 4$ 和 $R = 5$：

$$\mathrm{P}(5) = \left[\binom{5}{4} + \binom{5}{5}\right]\frac{1}{2^5} = \frac{5+1}{32} = \frac{3}{16}.$$

当 $s=6$ 时，只有 $x=2$ 和 $x=4$ 可能出现，分别需要 $R=4$ 和 $R=5$：

$$\mathrm{P}(6)=\left[\binom{6}{4}+\binom{6}{5}\right]\frac{1}{2^6}=\frac{15+6}{64}=\frac{21}{64}.$$

一般的偶数情况

当 $s=2k$ 时，可能到达的目标点是 $x=2$ 和 $x=4$，分别需要 $R=k+1$ 和 $R=k+2$。根据帕斯卡法则 $\binom{n}{r}+\binom{n}{r+1}=\binom{n+1}{r+1}$，

$$\mathrm{P}(2k)=\left[\binom{2k}{k+1}+\binom{2k}{k+2}\right]\left(\frac{1}{2}\right)^{2k}=\binom{2k+1}{k+2}\left(\frac{1}{2}\right)^{2k}=\binom{2k+1}{k-1}\left(\frac{1}{2}\right)^{2k},$$

最后一步用了 $\binom{n}{r}=\binom{n}{n-r}$，其中 $(2k+1)-(k+2)=k-1$。这就是所要求的结果。它在 $k=1$ 时同样成立：这时 $\binom{2}{3}=0$，所以 $x=4$ 的那一项根本不存在，而公式给出 $\frac{1}{4}$，正是两步都向正方向走的概率。

奇数情况

当 $s=2k+1$ 时，目标点是 $x=3$ 和 $x=5$，分别需要 $R=k+2$ 和 $R=k+3$。同理，

$$\mathrm{P}(2k+1)=\left[\binom{2k+1}{k+2}+\binom{2k+1}{k+3}\right]\left(\frac{1}{2}\right)^{2k+1}=\binom{2k+2}{k-1}\left(\frac{1}{2}\right)^{2k+1}.$$

当 $k=2$ 时，这给出 $\binom{6}{1}/32=\frac{3}{16}$，与 $\mathrm{P}(5)$ 吻合。

最大值

对 $k\geq 1$ 记 $a_k=\mathrm{P}(2k)$，并对 $k\geq 1$ 记 $b_k=\mathrm{P}(2k+1)$（留意 $b_0=\mathrm{P}(1)=0$）。这些数全是正数，所以我们可以用比值来比较相邻两项。

> **想一想** 两个二项式系数之比应该怎样化简？

把两者都写成阶乘（factorials）再约简。对于偶数项，

$$\frac{\binom{2k+3}{k}}{\binom{2k+1}{k-1}}=\frac{(2k+3)!}{k!\,(k+3)!}\cdot\frac{(k-1)!\,(k+2)!}{(2k+1)!}=\frac{(2k+3)(2k+2)}{k(k+3)},$$

所以

$$\frac{a_{k+1}}{a_k}=\frac{(2k+3)(2k+2)}{4k(k+3)}=\frac{4k^2+10k+6}{4k^2+12k}.$$

由于分母是正数，这个比值大于 1 当且仅当 $10k+6>12k$，即 $k<3$；当 $k=3$ 时它等于 1；而当 $k>3$ 时它小于 1。因此

$$a_1<a_2<a_3=a_4>a_5>a_6>\cdots,$$

而偶数步的最大值是 $a_3 = a_4 = \mathrm{P}(6) = \mathrm{P}(8) = \frac{21}{64}$。（验算：$\mathrm{P}(8) = \binom{9}{3}/2^8 = 84/256 = 21/64$。）

对于奇数项，同样的约简给出

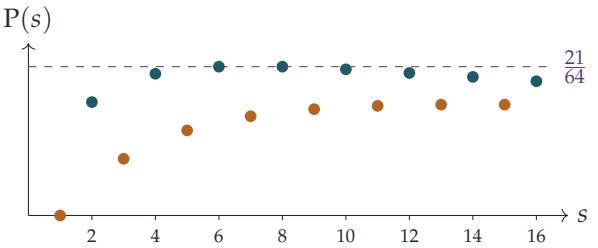
$$\frac{b_{k+1}}{b_k} = \frac{(2k+4)(2k+3)}{4k(k+4)} = \frac{4k^2+14k+12}{4k^2+16k},$$

它在 $k < 6$ 时大于 1，在 $k = 6$ 时等于 1，在 $k > 6$ 时小于 1。所以奇数步的最大值是

$$b_6 = b_7 = \mathrm{P}(13) = \mathrm{P}(15) = \binom{14}{5}\frac{1}{2^{13}} = \frac{2002}{8192} = \frac{1001}{4096}.$$

最后，$\frac{21}{64} = \frac{1344}{4096} > \frac{1001}{4096}$，所以没有任何奇数步的值能超过偶数步的最大值。因此 $\mathrm{P}(s)$ 取得最大值 $\frac{21}{64}$ 的时候，是

$$s = 6 \quad 及 \quad s = 8.$$



蓝绿色的点对应偶数 s，琥珀色的点对应奇数 s：这是两个独立的数列，各自先升后降，而偶数的那个始终在上方。

验算

- 用手计算小的情况：$\mathrm{P}(2) = \frac{1}{4}$（两步都向右），$\mathrm{P}(3) = \frac{1}{8}$（三步都向右）；而取 $k = 1$ 时，公式给出 $\binom{3}{0}/2^2 = \frac{1}{4}$ 和 $\binom{4}{0}/2^3 = \frac{1}{8}$，正如所料。
- 用电脑对这个漫步作精确穷举（以分数运算，$s \le 60$），结果证实两条公式都正确，$\mathrm{P}(6) = \mathrm{P}(8) = \frac{21}{64} \approx 0.328$ 是整体的最大值，而 $\mathrm{P}(13) = \mathrm{P}(15) = \frac{1001}{4096} \approx 0.244$ 是奇数步的最大值。

常见错误

- 对每个 s 都把全部四个目标点加起来：基于奇偶性，其中一半根本不可能到达。
- 求出偶数步的最大值便停下来。奇数步的值构成另一个独立的数列，你必须验证它的峰值较低。

- 停在“比值 > 1 对 $k < 3$ 成立”便只答 $s = 6$：比值在 $k = 3$ 时等于 1，所以 $s = 8$ 与它并列。

STEP 之外

这就是简单对称随机漫步（simple symmetric random walk），是概率论的核心对象之一。走了 s 步之后，它的位置扩散的幅度大约与 $\sqrt{s}$ 成正比，所以它落在任何一个固定有限范围内的概率会先上升，最终下降至零；$s = 7$ 附近的峰值，正是扩散幅度与该范围大小相若的时候。同样的二项式计数方法还会引出反射原理（reflection principle），以及一维漫步以概率 1 返回原点这个事实。Grimmett and Welsh 的 *Probability: An Introduction*（概率论入门教材）中有一章讲随机漫步，写得很好。

详细解答

STEP II 1992，第 16 题

概率与统计 · 条件概率

STEP II 1992，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/92-s2-q16

本题在 *Beyond STEP* 第 12.17 节巴拿赫的火柴盒中有完整的引导式讲解：https://beyondstep.keithmcnulty.org/zh-hans/ch12/matchboxes.html

关键思路

这是巴拿赫（Banach）火柴盒问题的一个变体。题目告诉了我们何时发现糖果不够（第 N 次伸手），所以这是一个条件概率（conditional probability）问题。因此我们要计算两个联合概率（joint probability）：第一次失败发生在第 N 次伸手而且拿的是拖肥糖，以及第一次失败发生在第 N 次伸手而且拿的是薄荷糖。两者都是“头 $N-1$ 次伸手中每种糖恰好各有若干次，然后下一次是指定的一种”这类计算，属于二项（binomial）概率的计算。

解答

假设各次伸手互相独立（independent），每次拿拖肥糖的概率为 p，拿薄荷糖的概率为 q，其中 $0 < p < 1$。设 T 为以下事件（event）：她在第 N 次伸手时想拿拖肥糖，却发现拖肥糖那包已经空了，而且之前各次伸手都没有发现任何一包是空的。设 M 为把拖肥糖换成薄荷糖后的相应事件。我们要求的是

$$\mathrm{P}(\text{拖肥糖} \mid \text{首次发现时的伸手次数为} N) = \frac{\mathrm{P}(T)}{\mathrm{P}(T) + \mathrm{P}(M)}.$$

T 的概率

要 T 发生，拖肥糖那包必须在 $N-1$ 次伸手之后已经空了。到目前为止没有一次伸手失败，所以这些伸手每一次都拿到了一颗糖。因此头 $N-1$ 次伸手中恰好有 n 次是拿拖肥糖（多一次就会失败），其余 $N-1-n$ 次是拿薄荷糖。然后第 N 次伸手是拿拖肥糖。

反过来，假设头 $N-1$ 次伸手中恰好有 n 次拿拖肥糖，以及 $N-1-n \le m$ 次拿薄荷糖，次序不限。那么之前不可能有任何一次伸手失败：在任何阶段，她伸手拿拖肥糖都不超过 n 次，拿薄荷糖都不超过 m 次。这样，如果第 N 次伸手是拿拖肥糖，T 就会发生。

> **想一想** 头 $N-1$ 次伸手的次序难道不要紧吗？不是要确保她不会提早把糖拿光吗？

不要紧。她只有在某一包已空而又伸手进去的时候，才会发现那一包是空的。提早把拖肥糖拿光（即拿第 n 颗拖肥糖之时）并无问题，只要她在第 N 次伸手之前不再伸手拿拖肥糖便可。这一点自动成立，因为头 $N-1$ 次伸手中只有 n 次是拿拖肥糖。所以全部 $\binom{N-1}{n}$ 种次序都可以，于是

$$\mathrm{P}(T) = \binom{N-1}{n} p^n q^{N-1-n} \cdot p = \binom{N-1}{n} p^{n+1} q^{N-1-n}, \qquad n \le N-1 \le n+m,$$

其他情况下 $\mathrm{P}(T)=0$。如果 $N-1<n$，她不可能已经把拖肥糖拿光。如果 $N-1>n+m$，她就需要多于 m 颗薄荷糖。

M 的概率

把两种糖的角色对调，用完全相同的方法可得

$$\mathrm{P}(M) = \binom{N-1}{m} q^{m+1} p^{N-1-m}, \qquad m \le N-1 \le n+m,$$

其他情况下 $\mathrm{P}(M)=0$。

答案

这次发现必定发生在第 $\min(n,m)+1$ 次至第 $n+m+1$ 次伸手之间。第 $n+m+2$ 次伸手是不可能的，因为到那时两包糖都已经空了，早就应该有一次伸手失败。

- 如果 $N \le m$（即 $\mathrm{P}(M)=0$ 但 $\mathrm{P}(T)>0$），她肯定是在拿拖肥糖：概率为 1。
- 如果 $N \le n$（即 $\mathrm{P}(T)=0$），概率为 0。

- 如果 $\max(n,m)+1 \le N \le n+m+1$，两种情况都有可能，而

$$\mathrm{P}(\text{拖肥糖}) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}}.$$

在最后一种情况，我们可以把分子和分母同时除以 $\mathrm{P}(T)$ 来化简。二项式系数（binomial coefficient）中的公因数 $(N-1)!$ 互相约去，而各个幂则合并成 $(q/p)^{n+m+2-N}$。由此得出

$$\mathrm{P}(\text{拖肥糖}) = \left[1 + \frac{n!\,(N-1-n)!}{m!\,(N-1-m)!} \left(\frac{q}{p}\right)^{n+m+2-N}\right]^{-1}.$$

验算

- 对称性：如果 $n=m$ 而且 $p=q=\frac{1}{2}$，两个式子相等，答案是 $\frac{1}{2}$，正如所料。
- 取 $n=2$、$m=3$、$p=\frac{1}{2}$、$N=5$。那么 $\mathrm{P}(T) = \binom{4}{2}/2^5 = \frac{3}{16}$，而 $\mathrm{P}(M) = \binom{4}{3}/2^5 = \frac{1}{8}$，所以答案是 $\frac{3/16}{5/16} = \frac{3}{5}$。拖肥糖较少，所以先把拖肥糖拿光的可能性较大。
- 对 $N=1,\ldots,n+m+1$ 求和，有 $\sum_N \big(\mathrm{P}(T)+\mathrm{P}(M)\big) = 1$，因为终究必有一次伸手会失败。我们已就所有 $n,m \le 6$ 及若干个 p 值，以逐步的精确枚举核对过这一点以及上述公式。以 $n=4$、$m=3$、$N=7$、$p=0.55$ 进行 2×10^5 次试验的模拟（simulation），结果是 0.527，而精确值是 0.528。

常见错误

- 漏掉第 N 次伸手本身所对应的最后一个因数 p（或 q）。
- 答了 $\mathrm{P}(T)$，而不是条件概率 $\mathrm{P}(T)/(\mathrm{P}(T)+\mathrm{P}(M))$。题目已告诉我们，发现是在第 N 次伸手时发生的。
- 以为所指的事件是“拖肥糖在第 N 次伸手时拿光”。拿走最后一颗拖肥糖并不算发现，下一次再去拿才是发现。

STEP 之外

第一次失败发生的时间，其分布是两个负二项（negative binomial）分布的混合（mixture）。负二项分布计算的是取得指定次数的成功所需的试验次数。巴拿赫原本的火柴盒问题（$n=m$，$p=\frac{1}{2}$）问的是：在发现的一刻，另一个盒子里还剩下多少根火柴。其答案可以用斯特灵公式（Stirling's formula）来近似，得出平均值（mean）约为 $2\sqrt{n/\pi}$。关

于负二项分布，可参阅 Grimmett 与 Welsh 的 *Probability: An Introduction*（概率入门教材）。费勒（Feller）的经典著作 *An Introduction to Probability Theory and Its Applications* 第 1 卷（概率论及其应用导论）中，有一段关于火柴盒问题的著名讨论。

详细解答

STEP III 1992，第 1 题

纯数学 · 麦克劳林级数

STEP III 1992，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/92-s3-q1

关键思路

在第 (i) 部分，那两个恒等式（identity）就是一部制造导数（derivative）的机器：一旦把 f'' 用 f' 表示出来，再逐次求导便可得到所有更高阶的导数，而在 $x = 0$ 处，一切都由 $f'(0) = \frac{1}{2}$ 构建而成。在第 (ii) 部分，麻烦在于因子 $\sinh x$，它在原点处等于零；乘以 x 便可消去这个零点，于是展开式就是两个可以徒手求出的级数（series）之积。对称性（奇函数（odd function）与偶函数（even function））预先告诉我们哪些幂次可能出现。

解答

第 (i) 部分

利用链式法则（chain rule）对 $f(x) = \ln(1 + e^x)$ 求导，

$$f'(x) = \frac{e^x}{1 + e^x} > 0,$$

所以它的对数有定义，并且

$$\ln[f'(x)] = \ln e^x - \ln(1 + e^x) = x - f(x),$$

这就是第一个结果。对这个恒等式求导，左边得出 $f''(x)/f'(x)$，所以

$$\frac{f''(x)}{f'(x)} = 1 - f'(x) \quad \Longrightarrow \quad f''(x) = f'(x) - [f'(x)]^2,$$

正是所求。把它写成 $f'' = f'(1 - f')$ 会较为方便。

? **想一想** 怎样才能不必对 $\ln(1 + e^x)$ 求导四次，就得到 f''' 和 f''''?

直接对恒等式 $\mathrm{f}'' = \mathrm{f}' - (\mathrm{f}')^2$ 本身求导，对 $(\mathrm{f}')^2$ 使用积法则（product rule）：

$$\mathrm{f}''' = \mathrm{f}'' - 2\mathrm{f}'\mathrm{f}'' = \mathrm{f}''(1 - 2\mathrm{f}'),$$
$$\mathrm{f}'''' = \mathrm{f}'''(1 - 2\mathrm{f}') - 2(\mathrm{f}'')^2.$$

现在代入 $x = 0$，此时 $\mathrm{f}(0) = \ln 2$ 而 $\mathrm{f}'(0) = \frac{1}{1+1} = \frac{1}{2}$：

$$\mathrm{f}''(0) = \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4},$$
$$\mathrm{f}'''(0) = \tfrac{1}{4}\left(1 - 2 \cdot \tfrac{1}{2}\right) = 0,$$
$$\mathrm{f}''''(0) = 0 - 2\left(\tfrac{1}{4}\right)^2 = -\tfrac{1}{8}.$$

根据麦克劳林定理（Maclaurin's theorem），$\mathrm{f}(x) = \sum \mathrm{f}^{(n)}(0)\, x^n/n!$，所以

$$\mathrm{f}(x) = \ln 2 + \frac{x}{2} + \frac{x^2}{8} - \frac{x^4}{192} + \cdots,$$

其中 x^3 项为零（而 $-\frac{1}{8} \div 24 = -\frac{1}{192}$）。

第 (ii) 部分

为何 g 没有这样的展开式。 一个只含非负幂次的级数 $a_0 + a_1x + a_2x^2 + \cdots$ 在 $x = 0$ 处取有限值 a_0，并且在该处连续（continuous）。但 $\sinh 0 = 0$ 而 $\cosh 0 = 1$，所以当 $x \to 0$ 时，分母 $\sinh x \cosh 2x$ 趋向 0；事实上，由于 x 很小时 $\sinh x \approx x$，我们有 $\mathrm{g}(x) \approx 1/x$，它在 0 附近是无界的（unbounded）。因此 $\mathrm{g}(x)$ 在原点附近不可能等于这样的级数。

为何 $x\mathrm{g}(x)$ 有。 写成

$$x\,\mathrm{g}(x) = \frac{x}{\sinh x} \cdot \frac{1}{\cosh 2x}.$$

由于 $\sinh x = x\left(1 + \frac{x^2}{6} + \frac{x^4}{120} + \cdots\right)$，因子 x 可以约去，于是

$$\frac{x}{\sinh x} = \frac{1}{1 + \frac{x^2}{6} + \frac{x^4}{120} + \cdots},$$

这是 1 除以一个常数项为 1 的级数；这样的倒数可以用二项式展开（binomial expansion）写成非负幂次的级数。同样，$\cosh 2x \geq 1$ 永不为零，而 $1/\cosh 2x$ 亦可用同样的方法展开。两个这样的级数之积仍是这样的级数。

为何只有偶数幂次。 sinh 是奇函数而 cosh 是偶函数，所以 $\mathrm{g}(-x) = -\mathrm{g}(x)$：g 是奇函数。因此 $h(x) = x\,\mathrm{g}(x)$ 满足 $h(-x) = (-x)(-\mathrm{g}(x)) = h(x)$，所以 h 是偶函数。若 $h(x) = \sum a_n x^n$，则 $h(x) - h(-x) = 2\sum_{n\text{ 为奇数}} a_n x^n$ 恒等于零，而由于恒等于零的幂级数（power series）所有系数皆为零，故对每个奇数 n 都有 $a_n = 0$。（你也可以直接看出这一点：上面两个因子都是 x^2 的级数。）

展开式。 利用 $(1+u)^{-1} = 1 - u + u^2 - \cdots$，取 $u = \frac{x^2}{6} + \frac{x^4}{120}$，并保留至 x^4 的各项，

$$\frac{x}{\sinh x} = 1 - \frac{x^2}{6} - \frac{x^4}{120} + \frac{x^4}{36} + \cdots = 1 - \frac{x^2}{6} + \frac{7x^4}{360} + \cdots,$$

因为 $\frac{1}{36} - \frac{1}{120} = \frac{10-3}{360}$。其次，$\cosh 2x = 1 + 2x^2 + \frac{2}{3}x^4 + \cdots$（来自 $\frac{(2x)^2}{2}$ 和 $\frac{(2x)^4}{24}$），所以取 $v = 2x^2 + \frac{2}{3}x^4$，

$$\frac{1}{\cosh 2x} = 1 - v + v^2 - \cdots = 1 - 2x^2 - \tfrac{2}{3}x^4 + 4x^4 + \cdots = 1 - 2x^2 + \tfrac{10}{3}x^4 + \cdots.$$

把两个级数相乘并按幂次整理：

$$\begin{aligned} x\,\mathrm{g}(x) &= \left(1 - \tfrac{1}{6}x^2 + \tfrac{7}{360}x^4\right)\left(1 - 2x^2 + \tfrac{10}{3}x^4\right) + \cdots \\ &= 1 - \left(2 + \tfrac{1}{6}\right)x^2 + \left(\tfrac{10}{3} + \tfrac{2}{6} + \tfrac{7}{360}\right)x^4 + \cdots \\ &= 1 - \frac{13}{6}x^2 + \frac{1327}{360}x^4 + \cdots, \end{aligned}$$

因为 $\frac{10}{3} + \frac{1}{3} = \frac{11}{3} = \frac{1320}{360}$。

验算

- 第 (i) 部分中 x^3 项消失并非偶然：$\mathrm{f}(x) - \frac{x}{2} = \ln(\mathrm{e}^{-x/2} + \mathrm{e}^{x/2}) = \ln(2\cosh\frac{x}{2})$ 是偶函数，所以在 $\frac{x}{2}$ 之后只会出现偶数幂次。
- 数值上，在 $x = 0.1$ 处：$\ln(1 + \mathrm{e}^x) \approx 0.7443966600$，与四次多项式 $\ln 2 + \frac{x}{2} + \frac{x^2}{8} - \frac{x^4}{192}$ 相差仅约 3.5×10^{-10}。
- 在 $x = 0.05$ 处，$x\,\mathrm{g}(x)$ 与 $1 - \frac{13}{6}x^2 + \frac{1327}{360}x^4$ 相差约 9×10^{-8}，大约是 $6x^6$，与下一项为 x^6 项时的预期相符。

常见错误

- 把 $1/(1+u)$ 只展开成 $1 - u$：u^2 项对 x^4 有贡献（即上面的 $\frac{1}{36}$ 和 $4x^4$）。
- 写成 $\cosh 2x = 1 + 2x^2 + \frac{x^4}{24}\cdots$，忘记了 $2x$ 的 4 次方是 $16x^4$。

STEP 之外

第 (ii) 部分谈的是带有*极点*（pole）的函数：g 像 $1/x$ 那样趋向无穷，而乘以 x 便可消除这个奇点（singularity）。在复分析（complex analysis）中，这样的函数有一个*洛朗级数*（Laurent series）$\frac{1}{x}\left(1 - \frac{13}{6}x^2 + \cdots\right)$，而 $1/x$ 的系数（即*留数*（residue），这里是 1）正

是计算许多实积分的关键。这个级数的收敛半径（radius of convergence）取决于分母最接近的另一个零点，即 $\cosh 2x = 0$ 之处 $x = \pm\frac{\pi}{4}\mathrm{i}$。Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》）把这些想法阐述得十分精彩。

详细解答

STEP III 1992，第 2 题

纯数学 · 矩阵的幂与数学归纳法

> **STEP III 1992，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/92-s3-q2

关键思路

一切都建基于一个事实：$\mathbf{J}^2 = 2\mathbf{J}$。这表示任何形如 $\mathbf{I} + p\mathbf{J}$ 的矩阵（matrix）相乘，所得的积仍然是这个形式，所以我们自始至终只需追踪一个数。只要留意到在乘法之下表现良好的是 $1 + 2p$ 而不是 p，那些看似累赘的式子 $\frac{1}{2}[(1+2a)^k - 1]$ 便一目了然。

解答

一条乘法法则

由于

$$\mathbf{J}^2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = 2\mathbf{J},$$

而且 $\mathbf{I}$ 与任何矩阵都可交换（commute），所以对任意实数 p 和 q，

$$(\mathbf{I} + p\mathbf{J})(\mathbf{I} + q\mathbf{J}) = \mathbf{I} + (p+q)\mathbf{J} + pq\mathbf{J}^2 = \mathbf{I} + (p + q + 2pq)\mathbf{J}. \qquad (*)$$

留意

$$1 + 2(p + q + 2pq) = (1 + 2p)(1 + 2q),$$

所以如果我们把数 $1 + 2p$ 配给矩阵 $\mathbf{I} + p\mathbf{J}$，矩阵相乘便对应于这些数相乘。对于 $\mathbf{A} = \mathbf{I} + a\mathbf{J}$，所配的数是 $1 + 2a$。

平方与立方

在 $(*)$ 中取 $p = q = a$，得

$$\mathbf{A}^2 = \mathbf{I} + (2a + 2a^2)\mathbf{J}, \qquad \text{及} \qquad \tfrac{1}{2}\left[(1+2a)^2 - 1\right] = \tfrac{1}{2}(4a + 4a^2) = 2a + 2a^2,$$

所以 $\mathbf{A}^2 = \mathbf{I} + \frac{1}{2}[(1+2a)^2 - 1]\mathbf{J}$，正如题目所求。

至于立方，写成 $\mathbf{A}^3 = \mathbf{A}^2\mathbf{A}$，并在 $(*)$ 中取 $p = 2a + 2a^2$，$q = a$：

$$\begin{aligned} p + q + 2pq &= 2a + 2a^2 + a + 2a(2a + 2a^2) \\ &= 3a + 6a^2 + 4a^3 \\ &= \tfrac{1}{2}(6a + 12a^2 + 8a^3) = \tfrac{1}{2}\big[(1+2a)^3 - 1\big], \end{aligned}$$

当中用到 $(1+2a)^3 = 1 + 6a + 12a^2 + 8a^3$。所以 $\mathbf{A}^3 = \mathbf{I} + \frac{1}{2}[(1+2a)^3 - 1]\mathbf{J}$，正如题目所求。

至于四次方，最快捷的做法是 $\mathbf{A}^4 = \mathbf{A}^2\mathbf{A}^2$。取 $p = q = \frac{1}{2}[(1+2a)^2 - 1]$，则 $1 + 2p = (1+2a)^2$，所以配给 $\mathbf{A}^4$ 的数是 $(1+2a)^2(1+2a)^2 = (1+2a)^4$。因此

$$\mathbf{A}^4 = \mathbf{I} + \tfrac{1}{2}\big[(1+2a)^4 - 1\big]\mathbf{J} = \mathbf{I} + (4a + 12a^2 + 16a^3 + 8a^4)\mathbf{J}.$$

一般次方

从以上规律可以猜想

$$p_k = \tfrac{1}{2}\big[(1+2a)^k - 1\big], \qquad \text{等价地} \qquad 1 + 2p_k = (1+2a)^k.$$

我们用归纳法（induction）证明，对所有整数 $k \geq 1$，$\mathbf{A}^k = \mathbf{I} + p_k\mathbf{J}$ 成立。

基础步骤。当 $k = 1$ 时，$p_1 = \frac{1}{2}(1 + 2a - 1) = a$，而 $\mathbf{A}^1 = \mathbf{I} + a\mathbf{J}$。（公式亦给出 $p_0 = 0$，与 $\mathbf{A}^0 = \mathbf{I}$ 吻合。）

归纳步骤。假设 $\mathbf{A}^k = \mathbf{I} + p_k\mathbf{J}$，其中 p_k 如上所述。那么由 $(*)$，

$$\mathbf{A}^{k+1} = \mathbf{A}^k\mathbf{A} = \mathbf{I} + (p_k + a + 2ap_k)\mathbf{J},$$

而

$$\begin{aligned} p_k + a + 2ap_k &= (1+2a)p_k + a \\ &= \tfrac{1}{2}(1+2a)\big[(1+2a)^k - 1\big] + a \\ &= \tfrac{1}{2}(1+2a)^{k+1} - \tfrac{1}{2}(1+2a) + a \\ &= \tfrac{1}{2}\big[(1+2a)^{k+1} - 1\big] = p_{k+1}. \end{aligned}$$

所以 $\mathbf{A}^{k+1} = \mathbf{I} + p_{k+1}\mathbf{J}$，由归纳法可知

$$\mathbf{A}^k = \mathbf{I} + \tfrac{1}{2}\big[(1+2a)^k - 1\big]\mathbf{J} \qquad \text{对所有整数} k \geq 1.$$

? **想一想** 假如题目没有先给出 $\mathbf{A}^2$ 和 $\mathbf{A}^3$，我又怎样猜到 p_k 的形式?

答案就在递推关系（recurrence）本身：归纳步骤显示 $p_{k+1} = (1+2a)p_k + a$，这是一个一阶线性递推关系。两边同时加上 $\frac{1}{2}$，得 $p_{k+1} + \frac{1}{2} = (1+2a)(p_k + \frac{1}{2})$，所以 $p_k + \frac{1}{2}$ 是等比数列（geometric sequence），公比为 $1+2a$，首项为 $a + \frac{1}{2}$。这正正就是 $p_k + \frac{1}{2} = \frac{1}{2}(1+2a)^k$。

负数次方（额外收获）。 若 $a \neq -\frac{1}{2}$，公式对 $k=-1$ 同样成立：$p_{-1} = -\dfrac{a}{1+2a}$，而由 $(*)$ 得 $a + p_{-1} + 2ap_{-1} = a - \dfrac{a(1+2a)}{1+2a} = 0$，所以 $(\mathbf{I}+a\mathbf{J})(\mathbf{I}+p_{-1}\mathbf{J}) = \mathbf{I}$。若 $a = -\frac{1}{2}$，则 $\mathbf{A} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$ 是奇异（singular）矩阵，而公式对每个 $k \geq 1$ 都给出 $p_k = -\frac{1}{2}$：事实上 $\mathbf{A}^2 = \mathbf{A}$。

验算

- 取 $a=1$：$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$，直接相乘得 $\mathbf{A}^2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$，$\mathbf{A}^4 = \begin{pmatrix} 41 & 40 \\ 40 & 41 \end{pmatrix}$。公式给出 $p_2 = \frac{1}{2}(9-1) = 4$ 及 $p_4 = \frac{1}{2}(81-1) = 40$。
- $\mathbf{A}$ 有特征向量（eigenvector）$(1,1)$，对应的特征值（eigenvalue）为 $1+2a$；另有特征向量 $(1,-1)$，对应的特征值为 1。所以 $\mathbf{A}^k$ 的特征值是 $(1+2a)^k$ 和 1；矩阵 $\mathbf{I}+p_k\mathbf{J}$ 的特征值是 $1+2p_k$ 和 1，两者一致。
- 我们以精确的有理数运算，对 $0 \leq k \leq 12$ 以及四十个随机选取的 a 值（包括 $a=-\frac{1}{2}$）验证了 $\mathbf{A}^k = \mathbf{I} + p_k\mathbf{J}$。

常见错误

- 用二项式定理（binomial theorem）展开 $(\mathbf{I}+a\mathbf{J})^k$ 是合理的（因为 $\mathbf{I}$ 与 $\mathbf{J}$ 可交换），但若忘记 $\mathbf{J}^r = 2^{r-1}\mathbf{J}$ 而误当作 $\mathbf{J}$，便会得出错误的系数。
- 在归纳步骤中，先写下关于 p_{k+1} 的结论再倒推回去，并不算是证明；应该由 $\mathbf{A}^k\mathbf{A}$ 出发。

STEP 之外

我们配给 $\mathbf{I}+p\mathbf{J}$ 的数 $1+2p$，正是该矩阵对应于向量 $(1,1)$ 的特征值；所有这类矩阵都有相同的特征向量 $(1,1)$ 和 $(1,-1)$，所以它们可以用同一个基底变换（change of basis）同时对角化，而计算次方时只需把特征值自乘 k 次。这就是对角化（diagonalisation）背后的想法，它把矩阵次方、矩阵指数和线性递推关系都化为纯量计算。Strang 的 *Introduction to Linear Algebra*（线性代数入门）以大量例题阐述这个想法。

详细解答

STEP III 1992，第 3 题

纯数学 · 曲线与多项式的根

STEP III 1992，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/92-s3-q3

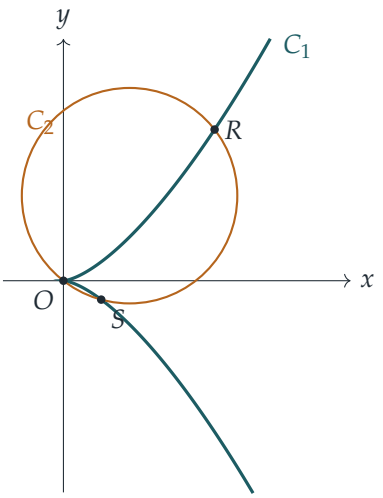
关键思路

把参数化（parametrisation）直接代入通过原点的圆的一般方程。这样会得到一个关于 t 的多项式方程，它的根就是各交点的参数，而原点则以因子 t^2 的形式出现。余下的是一个没有 t^3 项的四次方程（quartic），所以它的根之和是固定的，两两乘积之和也是固定的。既然已知其中两个根 r 和 s，另外两个根便完全确定。

解答

草图

由于 $x = t^2 \geq 0$，曲线位于半平面 $x \geq 0$ 内。把 t 换成 $-t$ 会把 (x, y) 变成 $(x, -y)$，所以曲线关于 x 轴对称。消去 t 得 $y^2 = x^3$，故 $y = \pm x^{3/2}$。当 $t = 0$ 时，$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}t} = 0$，而在原点附近 $|y| = x^{3/2}$ 远小于 x。因此两条分支离开 O 时都与 x 轴相切，曲线在 O 处有一个尖点（cusp）。当 t 很大时，两条分支上升得很陡，因为 $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{3t}{2}$。



曲线 C_1 以蓝绿色绘出。琥珀色的圆是 C_2 的一个例子，即 $r = 1$、$s = -\frac{1}{2}$ 时通过 O、R 和 S 的圆。

四次方程

通过原点的圆，其方程没有常数项：

$$x^2 + y^2 + 2gx + 2fy = 0,$$

其中 g 和 f 是不同时为零的实常数。参数为 t 的点位于圆上，当且仅当

$$t^4 + t^6 + 2gt^2 + 2ft^3 = 0, \qquad \text{即} \qquad t^2\left(t^4 + t^2 + 2ft + 2g\right) = 0.$$

因子 t^2 给出 $t = 0$，也就是原点。除 O 以外的每个交点都有 $t \neq 0$，所以它的参数满足

$$t^4 + t^2 + at + b = 0, \qquad a = 2f, \quad b = 2g.$$

特别地，r 和 s 都是这个方程的根，而 a、b 是实常数，正如题目所要求。

C_1 上每一点恰好对应一个参数，因为当 $x \neq 0$ 时 $t = y/x$（而当 $x = 0$ 时 $t = 0$）。所以相异的两点 R 和 S 有相异的参数，即 $r \neq s$，而 C_1 与 C_2 的交点恰好对应方程 $t^2(t^4 + t^2 + at + b) = 0$ 的实根。

另外两个根

设四次方程的四个根（实根或复根）为 r、s、p 和 q。比较

$$(t - r)(t - s)(t - p)(t - q) = t^4 + 0 \cdot t^3 + 1 \cdot t^2 + at + b$$

两边的系数，得

$$r+s+p+q=0,$$
$$rs+(r+s)(p+q)+pq=1.$$

（这里我们把六个两两乘积分组为 rs、pq 和四个交叉项 $rp+rq+sp+sq=(r+s)(p+q)$。）

> **想一想** 四个根之中我们只知道两个。在不知道 a 和 b 的情况下，怎样才能求出另外两个？

t^3 和 t^2 这两项的系数，即没有用字母表示的那两个，完全与圆无关。第一条方程给出 $p+q=-(r+s)$。代入第二条方程，

$$pq=1-rs+(r+s)^2=r^2+rs+s^2+1.$$

所以 p 和 q 是以下二次方程的根：

$$t^2+(r+s)\,t+\left(r^2+rs+s^2+1\right)=0,$$

其系数只取决于 r 和 s。C_1 与 C_2 若再有其他交点，其参数必须是这个二次方程的实根。

（顺带可得 $a=-(r+s)(r^2+s^2+1)$ 及 $b=rspq=rs(r^2+rs+s^2+1)$，所以通过 O、R 和 S 的圆由 r 和 s 确定。）

再没有其他实交点

这个二次方程的判别式（discriminant）是

$$\begin{aligned}\Delta&=(r+s)^2-4\left(r^2+rs+s^2+1\right)\\&=-3r^2-2rs-3s^2-4\\&=-\left[2(r^2+s^2)+(r+s)^2+4\right].\end{aligned}$$

方括号内的数值至少是 4，因为 $2(r^2+s^2)$ 和 $(r+s)^2$ 都是非负的。所以 $\Delta\le -4<0$，这个二次方程没有实根：p 和 q 是一对非实数的共轭复数（complex conjugates）。

因此 $t^2(t^4+t^2+at+b)=0$ 的实根只有 0、r 和 s。由于每个实根恰好对应 C_1 上的一点，C_1 与 C_2 的实交点就只有 O、R 和 S。这正是我们要证明的。

验算

取 $r=1$，$s=-\frac{1}{2}$。由上面的公式得 $a=-\frac{9}{8}$ 及 $b=-\frac{7}{8}$，所以 $f=-\frac{9}{16}$，$g=-\frac{7}{16}$。圆心是 $(\frac{7}{16},\frac{9}{16})$，而 $R=(1,1)$ 和 $S=(\frac{1}{4},-\frac{1}{8})$ 与圆心的距离平方都是 $\frac{130}{256}$，O 也是如此。余下

的二次式是 $t^2 + \frac{1}{2}t + \frac{7}{4}$，其判别式为 $\frac{1}{4} - 7 < 0$。这就是图中的圆，可以看到它与 C_1 只相交于 O、R 和 S。我们还用数值方法检验了 400 对随机选取的 (r, s)。对每一对，我们求出通过 O、R、S 的圆，确认了因式分解，并在 $-30 \le t \le 30$ 的范围内逐点检查：圆方程只在 $t = r$ 和 $t = s$ 处变号。

常见错误

- 忘记通过 O 的圆没有常数项。一般的圆有三个常数，而正是因为少了常数项，才会出现因子 t^2。
- 以为重根（double root）$t = 0$ 表示圆在 O 处与 C_1 相切。其实这个重根来自尖点：每一个通过 O 的圆都会给出因子 t^2。
- 只说判别式“显然是负数”，却没有把它写成“负的平方和加正常数”的形式。

STEP 之外

曲线 $y^2 = x^3$ 是奇异三次曲线（singular cubic curve）的标准例子。像本题这样连同重数（multiplicity）一起计算交点，正是贝祖定理（Bézout's theorem）的起点：次数分别为 m 和 n 的两条曲线，按适当方式计算，相交于 mn 点。一条三次曲线和一个圆给出 6 个交点，这与我们的结果吻合：O 处的两个，加上 R、S 和两个复数点。同一套韦达（Vieta）式的系数关系，也是椭圆曲线（elliptic curves）上群律（group law）的基础。Reid 的 *Undergraduate Algebraic Geometry* 以及 Silverman 与 Tate 的 *Rational Points on Elliptic Curves* 都把这些想法进一步发展下去。

详细解答

STEP III 1992，第 4 题

纯数学 · 正交轨线

> **STEP III 1992，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/92-s3-q4

关键思路

把 S_1 的每条曲线写成 $y = 1 + \dfrac{a}{x-a}$：它是一条等轴双曲线（rectangular hyperbola），渐近线（asymptote）为 $x = a$ 和 $y = 1$，而且每一条都通过原点。要找出整个曲线族的微分方程（differential equation），先求导，再利用曲线本身的方程消去 a。两条曲线正交，当且仅当它们的斜率（gradient）之积为 -1，因此把 $\mathrm{d}y/\mathrm{d}x$ 换成 $-1/\dfrac{\mathrm{d}y}{\mathrm{d}x}$，就可以把 S_1 的方程变成 S_2 的方程，而后者可以分离变量（separation of variables）。

解答

S_1 的曲线

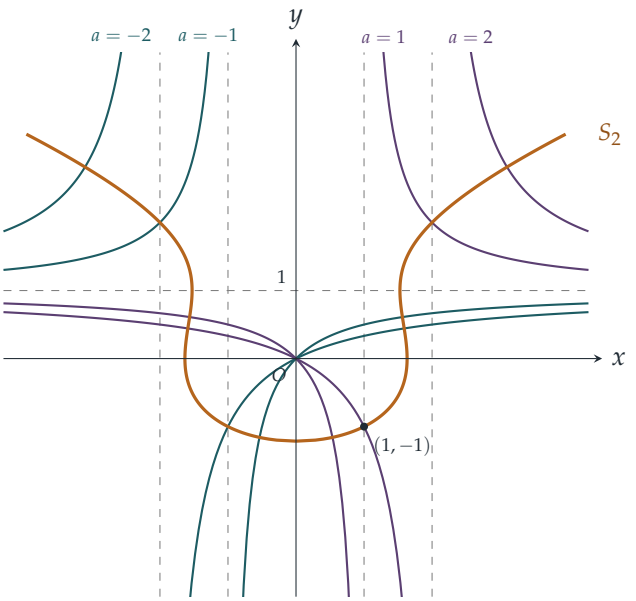
由于 $x = (x-a) + a$，

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

因此当 $a \neq 0$ 时，曲线是一条等轴双曲线，铅垂渐近线为 $x = a$，水平渐近线为 $y = 1$，并且通过原点。

- 若 $a > 0$，斜率恒为负。$x < a$ 的一支从 $y = 1$（当 $x \to -\infty$）向下走，经过 O，并在 $x \to a^-$ 时跌向 $-\infty$；$x > a$ 的一支则从 $+\infty$ 向下趋近 $y = 1$。
- 若 $a < 0$，斜率恒为正。$x > a$ 的一支从 $-\infty$ 上升，经过 O，趋近 $y = 1$；$x < a$ 的一支则位于 $y = 1$ 之上。

把 x 换成 $-x$，$\dfrac{x}{x-a}$ 就变成 $\dfrac{x}{x+a}$，所以对应 $-a$ 的曲线，是对应 a 的曲线沿 y 轴的反射。一个方便描绘的点：当 $x = \dfrac{3}{2}a$ 时，$y = 3$。



虚线是渐近线。琥珀色的曲线是 S_2 中通过 $(1,-1)$ 的那一条，它的形状我们会在下文说明理由。

S_1 的微分方程

由 $y(x-a)=x$ 可得 $x-a=\dfrac{x}{y}$ 及 $a=x-\dfrac{x}{y}=\dfrac{x(y-1)}{y}$。（在 S_1 的曲线上，只有在原点才有 $y=0$，所以在该点以外我们可以除以 y。）代入斜率的式子，

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

所以 S_1 的每一条曲线都满足

$$x\frac{\mathrm{d}y}{\mathrm{d}x}=y(1-y).$$

S_2 的微分方程

两条曲线正交，当且仅当它们在交点处的斜率之积为 -1。因此，若在某个公共点上 S_1 曲线的斜率为 p_1，S_2 曲线的斜率为 p_2，则 $p_1p_2=-1$，即 $p_1=-1/p_2$。

想一想 刚才求得的方程是关于 S_1 的。为甚么它能告诉我们有关 S_2 的事情?

通过每一点 (x,y)（其中 $x,y\neq 0$）的 S_1 曲线恰好有一条，就是 $a=x(y-1)/y$ 的那一条，而它在该点的斜率为 $-y(y-1)/x$。通过同一点的 S_2 曲线，斜率必须是 $p_2=-1/p_1$，

所以在 S_2 上

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{1}{-y(y-1)/x} = \frac{x}{y(y-1)}, \qquad \text{即} \qquad x = y(y-1)\frac{\mathrm{d}y}{\mathrm{d}x},$$

正如题目所求。

S_2 中通过 $(1,-1)$ 的曲线

这个方程可以分离变量：

$$\int (y^2 - y)\,\mathrm{d}y = \int x\,\mathrm{d}x \quad \Longrightarrow \quad \frac{y^3}{3} - \frac{y^2}{2} = \frac{x^2}{2} + C.$$

在 $(1,-1)$：$-\frac{1}{3} - \frac{1}{2} = \frac{1}{2} + C$，所以 $C = -\frac{4}{3}$。两边乘以 6，图中所画的曲线就是

$$3x^2 = 2y^3 - 3y^2 + 8.$$

要描绘这条曲线，设 $g(y) = 2y^3 - 3y^2 + 8$，则 $x = \pm\sqrt{g(y)/3}$，而曲线对称于 y 轴。由于 $g'(y) = 6y(y-1)$，g 有局部极大值（local maximum）$g(0) = 8$ 和局部极小值（local minimum）$g(1) = 7$。两者皆为正数，所以 g 只有一个实零点，位于 -2 与 -1 之间（因为 $g(-2) = -20$，$g(-1) = 3$）；其数值为 $y_0 \approx -1.214$。因此这条曲线：

- 最低点在 $(0, y_0)$，该处 $\mathrm{d}y/\mathrm{d}x = 0$；
- 在 $y = 0$（即 $x = \pm\sqrt{8/3} \approx \pm 1.63$）和 $y = 1$（即 $x = \pm\sqrt{7/3} \approx \pm 1.53$）处有铅垂切线，因为在这些位置 $y(y-1) = 0$；
- 在 $y = 0$ 与 $y = 1$ 之间稍微向轴弯回去（当 $x > 0$ 时，斜率 $x/(y(y-1))$ 在那里是负的），然后当 $|x|$ 很大时，大致按 $y \approx (\frac{3}{2}x^2)^{1/3}$ 增长。

为核对图形，通过 $(1,-1)$ 的 S_1 曲线满足 $-1 = \frac{1}{1-a}$，所以 $a = 2$，在该点的斜率为 -2；S_2 的斜率为 $\frac{1}{(-1)(-2)} = \frac{1}{2}$，而确实 $(-2)\cdot\frac{1}{2} = -1$。$y = 1$ 处的铅垂切线也合情合理：曲线在那里与 $y = 1$（即 S_1 中 $a = 0$ 的成员）正交。

验算

- 在 $(1,-1)$：$3(1)^2 = 3$，而 $2(-1)^3 - 3(-1)^2 + 8 = 3$。✓
- S_2 的曲线与 $a = 1$ 的曲线相交于 $(2,2)$：在该点 S_1 的斜率为 $-1/(2-1)^2 = -1$，S_2 的斜率为 $2/(2\cdot 1) = 1$，乘积为 -1。
- 以数值方法，我们找出了琥珀色曲线与图中四条双曲线的全部十二个交点（每条三个，例如 $a = -2$ 时的 $(1.5901,\, 0.4429)$），而斜率之积在每个交点都是 -1.00000000。

常见错误

- 微分方程中仍然留有 a：关键正正在于利用 $y = x/(x-a)$ 本身把它消去。
- 用了 $p_2 = -p_1$ 或 $p_2 = 1/p_1$，而不是 $p_1 p_2 = -1$。
- 把 S_2 的曲线画成简单的 U 形：它在 $y = 0$ 和 $y = 1$ 处有铅垂切线，两者之间还有一小段向后弯的部分。

STEP 之外

S_2 的曲线是 S_1 的正交轨线（orthogonal trajectories）。它们在物理学中以场线（field lines）与等势线（equipotentials）相交的形式出现，又或者是流线（streamlines）与等速度势线相交。这里的技巧（消去参数，再把 y' 换成 $-1/y'$）是标准方法，而由此得出的一阶方程，是常微分方程（ODE）入门课程的必备内容；参见 Boyce and DiPrima, *Elementary Differential Equations*（常微分方程入门教科书）。

详细解答

STEP III 1992，第 5 题

纯数学 · 三维向量几何

> **STEP III 1992，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/92-s3-q5

关键思路

对称性几乎包办了所有工作。以平面 $y+z=2$ 作反射（reflection），四面体（tetrahedron）会映射到自身，而 A 和 D 保持不动，所以这个反射必定把 B 和 C 对调。由此可知 BC 垂直于该平面，而且它的中点（midpoint）N 位于该平面上。之后的一切都只是毕氏定理（Pythagoras）和纯量积（scalar product），当中用到四个顶点与球心 Y 的距离全部相等这一事实。

解答

求 N 和半径

我们有 $\overrightarrow{AY}=(3,-2,4)-(0,4,-2)=(3,-6,6)$，所以球面（sphere）的半径（radius）是

$$YA=\sqrt{9+36+36}=9.$$

由于 $\overrightarrow{AY}=3\overrightarrow{YN}$，可得 $\overrightarrow{YN}=(1,-2,2)$，所以 $N=(4,-4,6)$ 而 $YN=3$。

留意 A、Y 和 N 全部满足 $y+z=2$：三点都位于对称平面（plane of symmetry）上。

为何 BC 垂直于该平面

设 Π 为平面 $y+z=2$。以 Π 作反射会把四面体映射到自身，所以它把各顶点重新排列（permute）。它固定 A 和 D，这两点都在 Π 上。B 和 C 都不在 Π 上：假如其中一点在该平面上，反射便会固定这一点，于是也必须固定第四个顶点，结果四个顶点全部落在 Π 上，根本构不成四面体。因此反射把 B 和 C 对调，即是说 C 是 B 的镜像。由此可知

- BC 垂直于 Π，所以它平行于法向量（normal）$(0,1,1)$；
- BC 的中点 N 位于 Π 上（与我们先前求得的结果吻合）。

> **想一想** 为甚么三角形 YNB 在 N 处成直角？

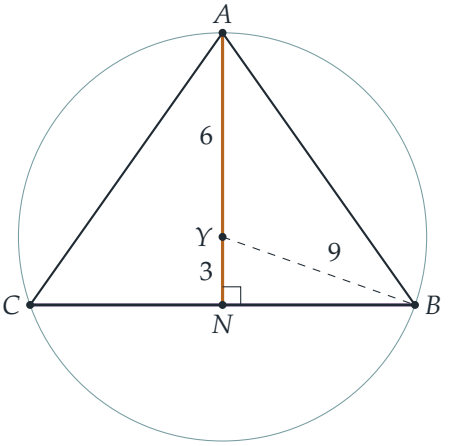
Y 和 N 都在 Π 上，所以 YN 是 Π 上的一条直线。但 NB 沿着 Π 的法线方向。法线垂直于平面上的每一条直线，所以 $\angle YNB = 90°$。

BN 与 B、C 的坐标

B 在球面上，所以 $YB = 9$。在三角形 YNB 中应用毕氏定理，

$$BN^2 = YB^2 - YN^2 = 81 - 9 = 72, \qquad \text{所以} \qquad BN = 6\sqrt{2},$$

正如题目所求。



上图按比例绘出平面 ABC，该平面包含 Y（因为 Y 在 AN 上）。琥珀色的线段 AN 是它与 Π 相交之处；圆则是它截球面所得的截口。

沿 BC 方向的单位向量（unit vector）是 $\frac{1}{\sqrt{2}}(0,1,1)$，所以

$$\overrightarrow{NB} = \pm 6\sqrt{2} \cdot \tfrac{1}{\sqrt{2}}(0,1,1) = \pm(0,6,6).$$

已知 $N = (4,-4,6)$，这条弦（chord）的两个端点是 $(4,2,12)$ 和 $(4,-10,0)$。所以

$$B = (4,2,12),\ C = (4,-10,0) \qquad \text{或} \qquad B = (4,-10,0),\ C = (4,2,12).$$

验算一下：$(4,2,12) - (3,-2,4) = (1,4,8)$ 的长度是 $\sqrt{1+16+64} = 9$，所以 B 确实在球面上。

D 的坐标

设 $\overrightarrow{YD} = (p, q, r)$。我们有三个条件。

- D 和 Y 都在 Π 上，所以 $\overrightarrow{YD}$ 垂直于法向量 $(0,1,1)$：$q + r = 0$。
- D 在球面上：$p^2 + q^2 + r^2 = 81$。
- 角度条件：$\overrightarrow{YA} \cdot \overrightarrow{YD} = YA \cdot YD \cos \angle AYD = 9 \cdot 9 \cdot \frac{1}{3} = 27$。

> **想一想** 三个未知数、三条方程，但其中一条是二次的。我们预期会有多少个解？

最多两个，因为最后会得到一条二次方程。从几何角度看，D 位于 Π 截球面所得的圆上（这是一个大圆（great circle），因为 $Y \in \Pi$），而该圆上与 YA 成指定角度的点，正是分处 A 两侧、互为镜像的两个位置。

利用 $\overrightarrow{YA} = (-3, 6, -6)$ 和 $r = -q$，角度条件给出

$$-3p + 6q + 6q = 27 \quad \Longrightarrow \quad p = 4q - 9.$$

代入球面方程，并利用 $r^2 = q^2$：

$$\begin{aligned} (4q - 9)^2 + 2q^2 &= 81 \\ 18q^2 - 72q &= 0 \\ 18q(q - 4) &= 0. \end{aligned}$$

- $q = 0$：这时 $p = -9$，$r = 0$，所以 $\overrightarrow{YD} = (-9, 0, 0)$，而 $D = (-6, -2, 4)$。
- $q = 4$：这时 $p = 7$，$r = -4$，所以 $\overrightarrow{YD} = (7, 4, -4)$，而 $D = (10, 2, 0)$。

两者都给出真正的四面体：D 在 Π 上，但不在直线 AN 上（平面 ABC 只有这一部分落在 Π 内），所以 D 不在平面 ABC 上。因此

$$D = (-6, -2, 4) \qquad \text{或} \qquad D = (10, 2, 0).$$

验算

- 每个 D 都满足 $y + z = 2$，而且两种情况下均有 $|\overrightarrow{YD}|^2 = 81$（分别是 81 和 $49 + 16 + 16$）。
- 对于 $D = (10, 2, 0)$：$\overrightarrow{YA} \cdot \overrightarrow{YD} = -21 + 24 + 24 = 27 = 81 \cdot \frac{1}{3}$。✓
- B 和 C 互为镜像：把 (x, y, z) 以 $y + z = 2$ 作反射，得到 $(x,\ 2 - z,\ 2 - y)$，它把 $(4, 2, 12)$ 送到 $(4, -10, 0)$。

- 用电脑搜索 Π 上坐标不超过 20 的所有整数点，找出位于球面上且满足 $\cos\angle AYD = \frac{1}{3}$ 的点，结果恰好只有 $(-6,-2,4)$ 和 $(10,2,0)$。两种情况下四面体的体积都是 $1728/6 = 288$，所以它并非扁平。

常见错误

- 求错 N。$\overrightarrow{AY} = 3\overrightarrow{YN}$ 表示 Y 位于 AN 的四分之三处，所以 $N = Y + \frac{1}{3}\overrightarrow{AY}$，而不是 $A + \frac{1}{3}\overrightarrow{AY}$。
- 把 YN 或 AN 当作半径。半径是 $YA = 9$；$YN = 3$ 只是球心到弦 BC 的距离。
- 忘记 D 必须位于 Π 上。少了 $q + r = 0$ 这个条件，可能的点便有整整一个圆那么多。

STEP 之外

“反射必定把 B 和 C 对调”这个论证，其实是群论（group theory）的一小片段：一个图形的对称通过重新排列其顶点来作用，而反射的不动点（fixed point）必定位于镜面上。这个反射本身就是线性代数中的映射 $\mathbf{x} \mapsto \mathbf{x} - 2\frac{(\mathbf{x}\cdot\mathbf{n} - d)}{|\mathbf{n}|^2}\mathbf{n}$，是一个行列式（determinant）为 -1 的正交变换（orthogonal transformation）。Strang 的 *Introduction to Linear Algebra* 讨论反射和正交矩阵（orthogonal matrix），而 Allenby 的 *Rings, Fields and Groups* 则阐述对称群（symmetry group）的理论。

详细解答

STEP III 1992，第 6 题

纯数学 · 三角函数归约公式

STEP III 1992，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/92-s3-q6

关键思路

两个化简都建基于同一做法：把三角函数之差化为含有因子 $\sin x$ 的乘积，从而约去碍事的分母。对 I_n 而言，所用的工具是 $\sin^2 A - \sin^2 B = \sin(A+B)\sin(A-B)$；对 J_n 而言，则是 $\sin P - \sin Q$ 的和差化积公式（sum-to-product formula）。分母一旦消去，余下的只是初等积分，而两条递推公式（reduction formula）亦可裂项相消（telescope）。

解答

先谈谈被积函数（integrand）。两者的分母都有 $\sin x$，它在 $x = 0$ 和 $x = \pi$ 处为零，但这些奇点是可去奇点（removable singularity）：由于当 $x \to 0$ 时 $\sin(nx)/\sin x \to n$，而当 $x \to \pi$ 时 $\sin(nx)/\sin x \to \pm n$，两个被积函数都保持有界。因此下文所有积分都是连续函数的普通积分。

关系式 $I_n - I_{n-1} = J_n$

利用 $\sin^2 A - \sin^2 B = (\sin A - \sin B)(\sin A + \sin B)$ 及和差化积公式，

$$\sin^2 A - \sin^2 B = \left(2\cos\tfrac{A+B}{2}\sin\tfrac{A-B}{2}\right)\left(2\sin\tfrac{A+B}{2}\cos\tfrac{A-B}{2}\right) = \sin(A+B)\sin(A-B).$$

取 $A = nx$ 及 $B = (n-1)x$，便得

$$\sin^2(nx) - \sin^2\big((n-1)x\big) = \sin\big((2n-1)x\big)\sin x.$$

因此，对 $n \geq 1$，

$$I_n - I_{n-1} = \int_0^{\pi} \frac{x\left[\sin^2(nx) - \sin^2\big((n-1)x\big)\right]}{\sin^2 x}\,\mathrm{d}x = \int_0^{\pi} \frac{x\sin(2n-1)x}{\sin x}\,\mathrm{d}x = J_n,$$

正如所求。（当 $n = 1$ 时，我们把 I_0 理解为 $x \sin^2 0 / \sin^2 x = 0$ 的积分，故 $I_0 = 0$。）

J_n 的递推公式

现在比较 J_{n+1} 与 J_n。由和差化积公式 $\sin P - \sin Q = 2\cos\frac{P+Q}{2}\sin\frac{P-Q}{2}$，

$$\sin(2n+1)x - \sin(2n-1)x = 2\cos(2nx)\sin x,$$

于是 $\sin x$ 再次被约去：

$$J_{n+1} - J_n = 2\int_0^\pi x\cos(2nx)\,\mathrm{d}x.$$

取 $u = x$ 及 $\mathrm{d}v = \cos(2nx)\,\mathrm{d}x$ 作分部积分（integration by parts），

$$\begin{aligned}\int_0^\pi x\cos(2nx)\,\mathrm{d}x &= \Big[\frac{x\sin(2nx)}{2n}\Big]_0^\pi - \frac{1}{2n}\int_0^\pi \sin(2nx)\,\mathrm{d}x \\ &= 0 + \frac{1}{4n^2}\Big[\cos(2nx)\Big]_0^\pi = \frac{1}{4n^2}(1-1) = 0,\end{aligned}$$

这是因为对整数 n 有 $\sin(2n\pi) = 0$ 及 $\cos(2n\pi) = 1$。因此递推公式为

$$J_{n+1} = J_n \qquad (n \geq 1).$$

想一想 一条说 $J_{n+1} = J_n$ 的递推公式，看来简单得令人起疑。这真的是答案吗？

是的：式中出现的 $2\cos(2nx)$ 与 x 相乘后，在整数个周期上的积分为零，所以由一个 n 到下一个，数值毫无改变。要确定这个值，我们只需计算首项：

$$J_1 = \int_0^\pi \frac{x\sin x}{\sin x}\,\mathrm{d}x = \int_0^\pi x\,\mathrm{d}x = \frac{\pi^2}{2},$$

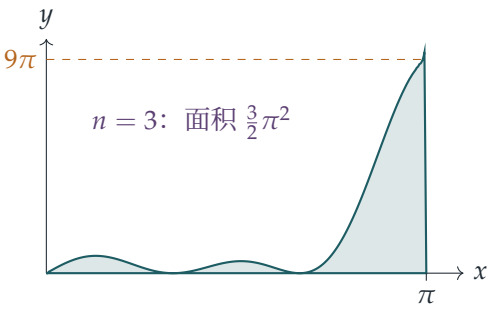
因此对每个正整数 n，都有 $J_n = \frac{1}{2}\pi^2$。

C 之下的面积

由于在 $(0, \pi)$ 上 $x \geq 0$ 且 $\sin^2(nx)/\sin^2 x \geq 0$，曲线永不低于 x 轴，所以 C 之下的面积正好是 I_n。把第一条关系式求和（各项裂项相消），并利用 $I_0 = 0$，

$$I_n = I_0 + \sum_{k=1}^{n}(I_k - I_{k-1}) = \sum_{k=1}^{n} J_k = n\cdot\frac{\pi^2}{2}.$$

所以 C 之下的面积是 $\frac{1}{2}n\pi^2$，正如所求。



验算

- $n=1$：曲线是 $y=x$，而该三角形的面积是 $\frac{1}{2}\pi^2$。✓
- 对 $n=1,\ldots,8$ 的积分使用辛普森法则(Simpson's rule),每次都得出 $J_n \approx 4.9348022 = \frac{1}{2}\pi^2$，又例如 $I_3 \approx 14.8044066 = \frac{3}{2}\pi^2$。
- 曲线在 $x=\pi$ 处升至 $y=n^2\pi$，但面积只随 n 线性增长：各个峰愈来愈高，却愈来愈窄（宽度约为 π/n），两者并无矛盾。

常见错误

- 忘记了 $I_0=0$（或从未查看 I_0 的意思），以致裂项求和没有起始值。
- 计算 $\int x\cos(2nx)\,\mathrm{d}x$ 时弄错正负号；较简洁的做法是留意到边界项为零，而余下的 $\sin(2nx)$ 在 $[0,\pi]$ 上的积分亦为零。

STEP 之外

函数 $\sin^2(nx)/\sin^2 x$ 除了相差一个因子 $1/n$ 之外，就是傅里叶分析（Fourier analysis）中的*费耶核*（*Fejér kernel*），而 $\sin\big((2n-1)x\big)/\sin x$ 基本上就是*狄利克雷核*（*Dirichlet kernel*）。你在这里发现的事实，即它们的积分与 n 无关，而函数本身却集中于区间的两端，正是它们能用来证明傅里叶级数（Fourier series）收敛的原因。Stein 与 Shakarchi 的 *Fourier Analysis: An Introduction*（傅里叶分析入门）对此有精彩的阐述；至于实分析（real analysis）的背景知识，可参看 Abbott 的 *Understanding Analysis*（实分析入门读本）。

详细解答

STEP III 1992，第 7 题

纯数学 · 向量几何

STEP III 1992，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/92-s3-q7

关键思路

题目中的一切都由平行四边形（parallelogram）以及按给定比例分割各边的点构成，而这正是位置向量（position vector）最擅长处理的情况。把原点设在 A，并用两个边向量 $\overrightarrow{AB}$ 和 $\overrightarrow{AD}$ 表示每一个点。然后求出其中两条直线的交点，再验证这个点位于第三条直线上。

解答

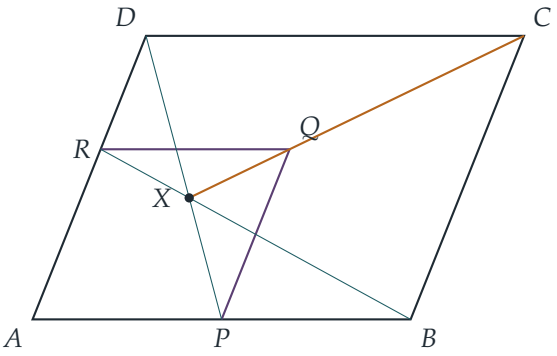
取 A 为原点，并设 $\mathbf{b} = \overrightarrow{AB}$，$\mathbf{d} = \overrightarrow{AD}$。由于 $ABCD$ 是一个真正的平行四边形，这两个向量并不平行。因此平面上每一点都可以写成 $\lambda\mathbf{b} + \mu\mathbf{d}$，其中实数对 (λ, μ) 恰好只有一组，于是我们可以比较系数（compare coefficients）。

由于 $ABCD$ 是平行四边形，C 的位置向量是 $\mathbf{b} + \mathbf{d}$。因为 P 在边 AB 上，而 R 在边 AD 上，我们可以写成

$$\mathbf{p} = p\,\mathbf{b}, \qquad \mathbf{r} = r\,\mathbf{d}, \qquad 0 \le p, r \le 1.$$

由于 $APQR$ 是平行四边形，$\overrightarrow{AQ} = \overrightarrow{AP} + \overrightarrow{AR}$，所以

$$\mathbf{q} = p\,\mathbf{b} + r\,\mathbf{d}.$$



BR 与 DP 的交点

直线 BR 上的一般点是

$$\mathbf{b}+s(r\mathbf{d}-\mathbf{b})=(1-s)\,\mathbf{b}+sr\,\mathbf{d},$$

而直线 DP 上的一般点是

$$\mathbf{d}+t(p\mathbf{b}-\mathbf{d})=tp\,\mathbf{b}+(1-t)\,\mathbf{d}.$$

在公共点上，$\mathbf{b}$ 的系数和 $\mathbf{d}$ 的系数必须分别相等：

$$1-s=tp,\qquad sr=1-t.$$

把 $s=1-tp$ 代入第二条方程，得 $r-rpt=1-t$，所以 $t(1-pr)=1-r$。

? **想一想** 我们可以除以 $1-pr$ 吗?

由于 $0\le p,r\le 1$，只有当 $p=r=1$ 时才有 $pr=1$，即 $P=B$ 且 $R=D$ 的情况。这时 $Q=C$，“直线”CQ 没有定义，也就没有甚么需要证明。因此我们排除这个情况，于是 $1-pr>0$。这亦说明 BR 与 DP 并不平行，所以它们恰好相交于一点。于是

$$t=\frac{1-r}{1-pr},\qquad s=1-tp=\frac{1-pr-p+pr}{1-pr}=\frac{1-p}{1-pr},$$

而交点 X 是

$$\mathbf{x}=tp\,\mathbf{b}+sr\,\mathbf{d}=\frac{p(1-r)\,\mathbf{b}+r(1-p)\,\mathbf{d}}{1-pr}.$$

X 位于 CQ 上

直线 CQ 上的一般点是

$$\mathbf{c}+u(\mathbf{q}-\mathbf{c})=(\mathbf{b}+\mathbf{d})+u\big((p-1)\mathbf{b}+(r-1)\mathbf{d}\big)=\big(1-u(1-p)\big)\mathbf{b}+\big(1-u(1-r)\big)\mathbf{d}.$$

试取 $u = \dfrac{1}{1-pr}$。**b** 的系数变成

$$1 - \frac{1-p}{1-pr} = \frac{1-pr-1+p}{1-pr} = \frac{p(1-r)}{1-pr},$$

而 **d** 的系数变成

$$1 - \frac{1-r}{1-pr} = \frac{1-pr-1+r}{1-pr} = \frac{r(1-p)}{1-pr}.$$

这正好就是 **x** 的系数。所以 X 也位于 CQ 上，三条直线 BR、CQ 和 DP 都通过 X。它们交于一点，正如所求。 ■

（当中的对称性令人放心：把 B, P, p 与 D, R, r 的角色互换，BR 与 DP 便会互换，CQ 保持不变，而 **x** 的公式亦不变。）

验算

- 若 $P = A$（$p = 0$），则 $Q = R$，而 $\mathbf{x} = r\mathbf{d}$，即是 R。事实上，DP 就是直线 DA，它与 BR 交于 R；而 CQ 就是 CR，它通过 R。
- 若 P 和 R 分别是 AB 和 AD 的中点，则 $\mathbf{x} = \frac{1}{3}(\mathbf{b} + \mathbf{d})$，位于由 A 到 C 的三分之一处。这是三角形 ABD 的形心（centroid），理应如此，因为 BR 和 DP 是它的其中两条中线（median）。
- 我们随机选取了 5000 个平行四边形以及 P 和 R 的随机位置，用分数精确地检验了共点性（concurrency）。每一个情况都与 **x** 的公式吻合。

常见错误

- 忘记了正因为 **b** 与 **d** 不平行，才可以比较这两个系数。
- 两条不同的直线用了同一个参数。每条直线都需要有自己的参数（s、t、u）。
- 没有提及退化情况（degenerate case）$P = B$、$R = D$，这时 $Q = C$，直线 CQ 没有定义。

STEP 之外

留意整个解答从未用过长度或角度，只用了平行线和直线上的比例。因此这个结果属于仿射几何（affine geometry）：仿射映射（affine map，即先作线性映射、再作平移）可以把任何平行四边形变成正方形，同时保持共点性，所以只需就正方形证明这个结果便已足够。关于直线相交的问题在射影几何（projective geometry）中也十分自然，塞瓦

（Ceva）定理和笛沙格（Desargues）定理等都属于这个范畴。如想广泛而轻松地游览这些概念，可参阅考克斯特（Coxeter）的 *Introduction to Geometry*（几何学入门），或 Brannan、Esplen 与 Gray 的 *Geometry*（几何学）。

详细解答

STEP III 1992，第 8 题

纯数学 · 以方程的根求三角函数之和

STEP III 1992，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/92-s3-q8

关键思路

棣美弗定理（De Moivre's theorem）把 $\sin(2n+1)\theta$ 化为 $\cos\theta$ 和 $\sin\theta$ 的多项式，再除以 $\sin^{2n+1}\theta$，便得到一个 $\cot^2\theta$ 的多项式。角 $k\pi/(2n+1)$ 令左边等于零，所以它们的余切平方就是一个已知多项式的根。只要知道了全部的根，那三个和其实不过是韦达定理（Vieta's formulas，即根之和与根的倒数之和）换了个样子而已。

解答

恒等式

设 $c=\cos\theta$、$s=\sin\theta$ 及 $N=2n+1$。由棣美弗定理和二项式定理（binomial theorem），

$$\cos N\theta+\mathrm{i}\sin N\theta=(c+\mathrm{i}s)^N=\sum_{j=0}^{N}\binom{N}{j}c^{N-j}(\mathrm{i}s)^j.$$

虚数项就是 j 为奇数的那些项。设 $j=2m+1$，则 $\mathrm{i}^{2m+1}=\mathrm{i}(-1)^m$，所以

$$\sin N\theta=\sum_{m=0}^{n}(-1)^m\binom{2n+1}{2m+1}c^{2n-2m}s^{2m+1}.$$

现在以 $r=n-m$ 把求和的次序倒转。这时 $2m+1=2n+1-2r$，而由二项式系数（binomial coefficients）的对称性，$\binom{2n+1}{2n+1-2r}=\binom{2n+1}{2r}$。因此

$$\begin{aligned}\sin(2n+1)\theta&=\sum_{r=0}^{n}(-1)^{n-r}\binom{2n+1}{2r}c^{2r}s^{2n+1-2r}\\&=\sin^{2n+1}\theta\sum_{r=0}^{n}(-1)^{n-r}\binom{2n+1}{2r}\cot^{2r}\theta,\end{aligned}$$

当中抽出了 s^{2n+1}，并用了 $c^{2r}/s^{2r}=\cot^{2r}\theta$。只要 $\sin\theta\neq0$，此式便成立，正如题目所求。

多项式的根

设

$$p(x)=\sum_{r=0}^{n}(-1)^r\binom{2n+1}{2r}x^r,$$

并对 $k=1,\dots,n$ 取 $\theta_k=k\pi/(2n+1)$。于是 $(2n+1)\theta_k=k\pi$，所以 $\sin(2n+1)\theta_k=0$，而 $0<\theta_k<\pi/2$ 则保证 $\sin\theta_k\neq 0$。把恒等式除以 $\sin^{2n+1}\theta_k$，再乘以 $(-1)^n$，得

$$\sum_{r=0}^{n}(-1)^r\binom{2n+1}{2r}\cot^{2r}\theta_k=0,\qquad \text{即}\qquad p\left(\cot^2\theta_k\right)=0.$$

> ? **想一想** 我们怎样知道这些就是全部的根？下面的求和全靠这一点。

角 $\theta_1<\theta_2<\cdots<\theta_n$ 全都位于 $(0,\pi/2)$ 之内，而 cot 在这个区间上为正，并且严格递减（strictly decreasing）。所以这 n 个数 $\cot^2\theta_k$ 互不相同。p 的首项系数（leading coefficient）是 $(-1)^n\binom{2n+1}{2n}=(-1)^n(2n+1)\neq 0$，所以 p 的次数恰好是 n，因而最多有 n 个根。由此可知，$p(x)=0$ 的根恰好就是 $\cot^2\theta_k$（$k=1,\dots,n$），而且每个都是单根（simple root）。

我们需要的系数是

$$[x^n]:\ (-1)^n\binom{2n+1}{2n}=(-1)^n(2n+1),$$
$$[x^{n-1}]:\ (-1)^{n-1}\binom{2n+1}{2n-2}=(-1)^{n-1}\binom{2n+1}{3},$$
$$[x^1]:\ -\binom{2n+1}{2}=-n(2n+1),\qquad [x^0]:\ 1.$$

第 (i) 部分

根之和等于 x^{n-1} 与 x^n 的系数之比的相反数：

$$\sum_{k=1}^{n}\cot^2\theta_k=-\frac{(-1)^{n-1}\binom{2n+1}{3}}{(-1)^n(2n+1)}=\frac{1}{2n+1}\cdot\frac{(2n+1)(2n)(2n-1)}{6}=\frac{n(2n-1)}{3}.$$

第 (ii) 部分

由于 $\tan^2\theta_k=1/\cot^2\theta_k$，我们要求的是各根 x_k 的倒数之和。这些根全都不是零（常数项是 1），而

$$\sum_{k=1}^{n}\frac{1}{x_k}=\frac{\sum_k\prod_{j\neq k}x_j}{\prod_k x_k},$$

根据韦达定理，其分子和分母分别是 $(-1)^{n-1}\cdot[x^1]/[x^n]$ 和 $(-1)^n\cdot[x^0]/[x^n]$。两者之比是 $-[x^1]/[x^0]$，所以

$$\sum_{k=1}^{n}\tan^2\theta_k=-\frac{-n(2n+1)}{1}=n(2n+1).$$

（换个说法：倒数 $1/x_k$ 是 $x^n p(1/x)$ 的根，也就是把系数次序倒转后所得的多项式的根，我们只须读出它的根之和。）

第 (iii) 部分

利用 $\operatorname{cosec}^2\theta = 1 + \cot^2\theta$ 和第 (i) 部分，

$$\sum_{k=1}^{n} \operatorname{cosec}^2\theta_k = n + \frac{n(2n-1)}{3} = \frac{3n+2n^2-n}{3} = \frac{2n(n+1)}{3}.$$

验算

当 $n=1$ 时，多项式是 $1-3x$，它的根是 $\frac{1}{3} = \cot^2(\pi/3)$。这时 (i) 给出 $\frac{1}{3}$，(ii) 给出 $3 = \tan^2(\pi/3)$，(iii) 给出 $\frac{4}{3} = \operatorname{cosec}^2(\pi/3)$，全部正确。当 $n=3$ 时，各角是 $\pi/7, 2\pi/7, 3\pi/7$，用计算机算得三个和分别是 5、21 和 8，与 $\frac{3\cdot5}{3}$、$3\cdot7$ 和 $\frac{2\cdot3\cdot4}{3}$ 吻合。我们还以数值方法，就随机选取的 θ 和 n 验证了恒等式，并就 $n \le 29$ 验证了全部三个和。

常见错误

- 忘记论证各个 $\cot^2\theta_k$ 互不相同，以及 p 的次数恰好是 n。缺了这一步，便不能把韦达定理用在 $\cot^2\theta_k$ 上。
- 系数的正负号出错：因子 $(-1)^n$ 在第 (i) 部分会约去，但前提是两个系数都带有这个因子。
- 把 k 一直取到 $2n$：当 $k > n$ 时，角 $k\pi/(2n+1)$ 只会重复相同的 $\cot^2$ 值，所以只有 $k \le n$ 才给出互不相同的根。

STEP 之外

把第 (iii) 部分与 $0 < \theta < \pi/2$ 时成立的不等式 $\cot^2\theta < 1/\theta^2 < \operatorname{cosec}^2\theta$ 结合起来，便可把 $\sum_{k=1}^{n} 1/k^2$ 夹在这里求得的两个和的倍数之间。令 $n \to \infty$，就得到欧拉（Euler）的著名结果 $\sum 1/k^2 = \pi^2/6$；这是一个漂亮的初等证明，大学一年级的分析（analysis）课程经常介绍。从多项式的系数读出各种和，这个想法同样是对称函数（symmetric functions）和牛顿恒等式（Newton's identities）的基础。分析方面，可参阅 Abbott 的 *Understanding Analysis*；想做更多同类风格的题目，可参阅 Siklos 的 *Advanced Problems in Mathematics*。

详细解答

STEP III 1992，第 9 题

纯数学 · 椭圆的极坐标方程

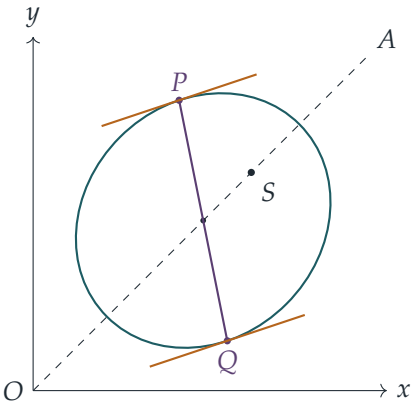
STEP III 1992，第 9 题

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/92-s3-q9

关键思路

极坐标（polar coordinates）以 S 为极点量度，始线沿 $y = x$ 的方向，因此极角为 θ 的点，从 S 看去在直角坐标（Cartesian coordinates）中的方向角是 $\theta + \frac{\pi}{4}$。把 x 和 y 写成 θ 的函数再求导，便得到斜率（gradient）。至于第二部分，直径（diameter）真正用得上的性质，是椭圆（ellipse）关于其中心对称：直径两端的切线（tangent）互相平行。于是我们令两个斜率相等，余下的工夫交给半角三角公式便可。



在草图中（$e = \frac{1}{2}$），PQ 是通过中心的直径，而 P 和 Q 处的切线互相平行。

解答

切线的斜率

由于 S 位于第一象限内的直线 $y=x$ 上，设 S 为点 (s,s)，其中 $s>0$。沿 SA 方向的单位向量（unit vector）是 $\frac{1}{\sqrt{2}}(1,1)$，与 x 轴成 $\frac{\pi}{4}$ 角。所以，以 S 为极点、极坐标为 (r,θ) 的点是

$$x=s+r\cos\phi,\qquad y=s+r\sin\phi,\qquad \text{其中}\phi=\theta+\tfrac{\pi}{4}.$$

在 E 上有 $r=\ell/(1+e\cos\theta)$，所以

$$\frac{\mathrm{d}r}{\mathrm{d}\theta}=\frac{\ell e\sin\theta}{(1+e\cos\theta)^2}=\frac{r\,e\sin\theta}{1+e\cos\theta}.$$

以 r' 表示 $\mathrm{d}r/\mathrm{d}\theta$，并利用 $\mathrm{d}\phi/\mathrm{d}\theta=1$，得

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

我们只关心两者之比，所以把两式同乘以正数 $(1+e\cos\theta)/r$：

$$\frac{1+e\cos\theta}{r}\frac{\mathrm{d}x}{\mathrm{d}\theta}=e\sin\theta\cos\phi-(1+e\cos\theta)\sin\phi=e\sin(\theta-\phi)-\sin\phi=-\tfrac{e}{\sqrt{2}}-\sin\phi,$$
$$\frac{1+e\cos\theta}{r}\frac{\mathrm{d}y}{\mathrm{d}\theta}=e\sin\theta\sin\phi+(1+e\cos\theta)\cos\phi=e\cos(\theta-\phi)+\cos\phi=\tfrac{e}{\sqrt{2}}+\cos\phi,$$

这是因为 $\theta-\phi=-\frac{\pi}{4}$。现在 $\sin\phi=\frac{1}{\sqrt{2}}(\sin\theta+\cos\theta)$，而 $\cos\phi=\frac{1}{\sqrt{2}}(\cos\theta-\sin\theta)$。全式乘以 $\sqrt{2}$，可知在 $\theta=\alpha$ 处的一个切向量（tangent vector）是

$$\mathbf{d}(\alpha)=\big(-(\sin\alpha+\cos\alpha+e),\ \cos\alpha-\sin\alpha+e\big).\qquad(*)$$

因此

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\cos\alpha-\sin\alpha+e}{-(\sin\alpha+\cos\alpha+e)}=\frac{\sin\alpha-\cos\alpha-e}{\sin\alpha+\cos\alpha+e},$$

正是所求。（当 $\sin\alpha+\cos\alpha+e=0$ 时，切线是铅垂的；$(*)$ 在该处依然成立。）

直径的两端

> **想一想** 我们还未求出中心，怎样才能把“PQ 通过中心”化为一条关于 α 和 β 的方程?

我们利用第一部分的结果。椭圆绕其中心 C 旋转半周（half-turn）后与自身重合。若 P 和 Q 是一条直径的两端，这个半周旋转会把两点互换，并把 P 处的切线映射到 Q 处的切线。半周旋转把每条直线都变成与之平行的直线，所以*P 和 Q 处的切线互相平行。*

因此，设 P 处 $\theta=\alpha$，Q 处 $\theta=\beta$，则由 $(*)$ 得出的向量 $\mathbf{d}(\alpha)$ 和 $\mathbf{d}(\beta)$ 互相平行。用向量而不用斜率来处理，即使切线是铅垂的，我们也永远不会除以零。记 $s_\alpha=\sin\alpha$，$c_\alpha=\cos\alpha$，余此类推。平行的条件是

$$(s_\alpha+c_\alpha+e)(c_\beta-s_\beta+e)=(s_\beta+c_\beta+e)(c_\alpha-s_\alpha+e).$$

展开后，$c_\alpha c_\beta$、$-s_\alpha s_\beta$、ec_α、ec_β 和 e^2 各项在等号两边都出现，互相抵消，剩下

$$s_\alpha c_\beta - c_\alpha s_\beta + es_\alpha - es_\beta = s_\beta c_\alpha - c_\beta s_\alpha + es_\beta - es_\alpha,$$

亦即

$$\sin(\alpha-\beta) + e(\sin\alpha - \sin\beta) = 0.$$

现在利用 $\sin(\alpha-\beta) = 2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha-\beta}{2}$ 和 $\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$：

$$2\sin\tfrac{\alpha-\beta}{2}\left(\cos\tfrac{\alpha-\beta}{2} + e\cos\tfrac{\alpha+\beta}{2}\right) = 0.$$

想一想 我们可以约去 $\sin\frac{\alpha-\beta}{2}$ 吗?

可以。P 和 Q 是不同的点，所以 α 与 β 之差不是 2π 的倍数，故 $\sin\frac{\alpha-\beta}{2} \neq 0$。因此

$$\cos\tfrac{\alpha-\beta}{2} + e\cos\tfrac{\alpha+\beta}{2} = 0.$$

以 $\frac{\alpha}{2}$ 和 $\frac{\beta}{2}$ 展开：

$$(1+e)\cos\tfrac{\alpha}{2}\cos\tfrac{\beta}{2} + (1-e)\sin\tfrac{\alpha}{2}\sin\tfrac{\beta}{2} = 0.$$

除以 $(1-e)\cos\frac{\alpha}{2}\cos\frac{\beta}{2}$（因为 E 是椭圆，所以这里 $1-e>0$），得

$$\tan\tfrac{\alpha}{2}\tan\tfrac{\beta}{2} = -\frac{1+e}{1-e},$$

正是所求。除非 $\cos\frac{\alpha}{2}$ 或 $\cos\frac{\beta}{2}$ 等于零，否则这一步除法是成立的。假如（例如）$\alpha=\pi$，方程便迫使 $\sin\frac{\beta}{2}=0$，所以 $\beta=0$。这条直径就是沿 OSA 的长轴（major axis），此时两个正切之积没有定义。其余每一条直径都满足这个结果。

验算

- 在 $\alpha=0$（较接近 S 的顶点）处，公式给出斜率 $\frac{-1-e}{1+e} = -1$。这是对的：长轴端点处的切线与 $y=x$ 垂直。
- 当 $e=\frac{1}{2}$ 时，乘积应为 -3。取 $\alpha=\frac{\pi}{2}$，则 $\tan\frac{\alpha}{2}=1$。于是 $\tan\frac{\beta}{2}=-3$，所以 $\sin\beta=-\frac{3}{5}$ 而 $\cos\beta=-\frac{4}{5}$。由 $(*)$ 得 $\mathbf{d}(\alpha)=(-\frac{3}{2},-\frac{1}{2})$ 和 $\mathbf{d}(\beta)=(\frac{9}{10},\frac{3}{10})$，两者互相平行。✓
- 数值上，随机选取 2000 组 ℓ、e、S 和 α，斜率公式都与以有限差分（finite difference）计算的导数吻合；而把 P 对中心（位于从 S 沿 AS 方向距离 $\ell e/(1-e^2)$ 之处）作反射，所得的点 Q 在 E 上，其 β 满足该恒等式，误差在 10^{-13} 以内。

常见错误

- 忘记了 $\frac{\pi}{4}$ 的旋转而使用 $x = r\cos\theta$，这样得出的是另一个坐标系中的斜率，并不是题目所求。
- 除以 $\sin\frac{\alpha-\beta}{2}$ 而没有说明它为何不等于零。
- 试图先求出中心，再直接加上“C 是 PQ 的中点”这个条件。这样做是行得通的，但远比利用平行切线繁复。

STEP 之外

直径两端的切线互相平行，属于*共轭直径*（conjugate diameters）理论的一部分：一族平行弦的中点都落在同一条直径上，而该直径两端的切线与这些弦平行。在一个把椭圆变成圆的仿射映射（affine map）之下，这一切都化为初等的圆几何，这正是仿射几何和射影几何（projective geometry）如此强大的原因之一。这里的半角正切，也是天体力学中描述轨道时所用的标准参数化（parametrisation）。想在大学一年级的程度踏入这门几何，不妨一读 Needham 的 *Visual Complex Analysis*（《复分析：可视化方法》），或 Siklos 的 *Advanced Problems in Mathematics*（《高等数学问题集》）。

详细解答

STEP III 1992，第 10 题

纯数学 · 弧长与面积

> **STEP III 1992，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/92-s3-q10

关键思路

两条曲线的周界（perimeter）都是弧长（arc length）积分，而且都无法用初等函数求出（它们是椭圆积分（elliptic integrals）），所以要证明两者相等，唯一的方法就是把其中一个积分化成另一个。先写出每条曲线的弧长元素，利用 $\cos^2 + \sin^2 = 1$ 化简各自的被积函数（integrand），再作一次代换（substitution），加上一点对称性，便可看出两者其实是同一个积分。面积则较容易处理，因为两者都可以直接计算出来。

（试卷上印出的范围以 $\frac{1}{4}\theta$ 结尾。这显然是 $\frac{1}{4}\pi$ 的误印，我们取 $-\frac{1}{4}\pi < \theta < \frac{1}{4}\pi$。）

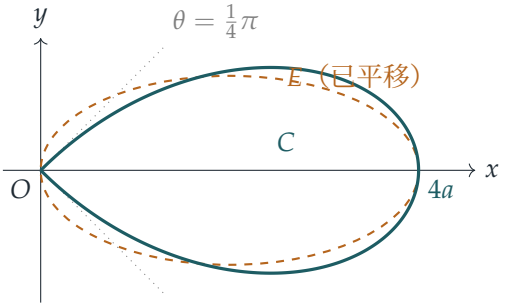
解答

描绘 C 的草图

当 $-\frac{1}{4}\pi < \theta < \frac{1}{4}\pi$ 时，我们有 $-\frac{1}{2}\pi < 2\theta < \frac{1}{2}\pi$，所以 $r = 4a\cos 2\theta > 0$ 恒成立（取 $a > 0$）。曲线有以下特征：

- $r(-\theta) = r(\theta)$，所以 C 关于始线（initial line）对称。
- 最大值 $r = 4a$ 在 $\theta = 0$ 处取得，即点 $(4a, 0)$。
- 当 $\theta \to \pm\frac{1}{4}\pi$ 时，$r \to 0$，所以曲线会进入原点，并在该处与直线 $\theta = \pm\frac{1}{4}\pi$ 相切。

所以 C 是单一个闭合环圈，也就是四瓣玫瑰线（rose）$r = 4a\cos 2\theta$ 的其中一瓣，沿正 x-轴伸展。下图把它与 E 的一个复本画在一起，该复本平移了 $2a$，使它横跨同一个区间 $0 \le x \le 4a$。



C 的周界

对于极坐标（polar）曲线，弧长元素满足 $\left(\frac{\mathrm{d}s}{\mathrm{d}\theta}\right)^2 = r^2 + \left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2$。这里 $\frac{\mathrm{d}r}{\mathrm{d}\theta} = -8a\sin 2\theta$，所以

$$\begin{aligned} r^2 + \left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^2 &= 16a^2\cos^2 2\theta + 64a^2\sin^2 2\theta \\ &= 16a^2\left(1 + 3\sin^2 2\theta\right). \end{aligned}$$

由于当 θ 走遍 $(-\frac{1}{4}\pi, \frac{1}{4}\pi)$ 时，环圈恰好被描画一次，因此

$$\text{周界}C = \int_{-\pi/4}^{\pi/4} 4a\sqrt{1 + 3\sin^2 2\theta}\,\mathrm{d}\theta.$$

代入 $\phi = 2\theta$，于是 $\mathrm{d}\theta = \frac{1}{2}\mathrm{d}\phi$，而积分上下限变成 $\pm\frac{1}{2}\pi$：

$$\text{周界}C = 2a\int_{-\pi/2}^{\pi/2}\sqrt{1 + 3\sin^2\phi}\,\mathrm{d}\phi = 4a\int_0^{\pi/2}\sqrt{1 + 3\sin^2\phi}\,\mathrm{d}\phi,$$

最后一步成立，是因为被积函数是 ϕ 的偶函数（even function）。

E 的周界

设 $x = 2a\cos\phi$，$y = a\sin\phi$，则

$$\left(\frac{\mathrm{d}x}{\mathrm{d}\phi}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\phi}\right)^2 = 4a^2\sin^2\phi + a^2\cos^2\phi = a^2\left(1 + 3\sin^2\phi\right),$$

而当 ϕ 由 0 走到 2π 时，E 恰好被描画一次，所以

$$\text{周界}E = a\int_0^{2\pi}\sqrt{1 + 3\sin^2\phi}\,\mathrm{d}\phi.$$

? **想一想** 为甚么 $\int_0^{2\pi}\sqrt{1 + 3\sin^2\phi}\,\mathrm{d}\phi$ 是在 $[0, \frac{1}{2}\pi]$ 上的积分的四倍?

记 $g(\phi)=\sqrt{1+3\sin^2\phi}$。因为 $\sin^2(\pi-\phi)=\sin^2\phi$，$g$ 在 $[\frac12\pi,\pi]$ 上的图像是它在 $[0,\frac12\pi]$ 上的图像的镜像，所以 $\int_0^{\pi}g=2\int_0^{\pi/2}g$。因为 $\sin^2(\phi+\pi)=\sin^2\phi$，$g$ 的周期为 π，所以 $\int_{\pi}^{2\pi}g=\int_0^{\pi}g$。（从几何上看，椭圆在四个象限内的部分长度相等。）因此

$$\text{周界}E=4a\int_0^{\pi/2}\sqrt{1+3\sin^2\phi}\,\mathrm{d}\phi,$$

这正好就是我们之前求得的 C 的周界的表达式。所以两个周界相等，而我们从头到尾都不必计算这个积分的值。

面积

极坐标曲线所围的面积是 $\frac12\int r^2\,\mathrm{d}\theta$。利用 $\cos^2 2\theta=\frac12(1+\cos 4\theta)$，

$$\begin{aligned}\text{所围面积}C&=\frac12\int_{-\pi/4}^{\pi/4}16a^2\cos^2 2\theta\,\mathrm{d}\theta=4a^2\int_{-\pi/4}^{\pi/4}(1+\cos 4\theta)\,\mathrm{d}\theta\\&=4a^2\Big[\theta+\tfrac14\sin 4\theta\Big]_{-\pi/4}^{\pi/4}=4a^2\cdot\frac{\pi}{2}=2\pi a^2.\end{aligned}$$

至于椭圆，一条以逆时针方向描画的闭合曲线所围的面积是 $\frac12\oint\left(x\frac{\mathrm{d}y}{\mathrm{d}\phi}-y\frac{\mathrm{d}x}{\mathrm{d}\phi}\right)\mathrm{d}\phi$，由此得

$$\text{所围面积}E=\frac12\int_0^{2\pi}\left(2a^2\cos^2\phi+2a^2\sin^2\phi\right)\mathrm{d}\phi=\frac12\cdot 2a^2\cdot 2\pi=2\pi a^2.$$

（同样地，这就是标准公式 $\pi\times$ 半轴 $\times$ 半轴 $=\pi(2a)(a)$。）所以两个面积相等，同为 $2\pi a^2$。

验算

- 取 $a=1.7$，以 200 000 边的多边形逼近每条曲线，得出两个周界均为 16.470362，与用森逊法则（Simpson's rule）算出的 $4a\int_0^{\pi/2}\sqrt{1+3\sin^2\phi}\,\mathrm{d}\phi$ 相等。所以每个周界约为 $9.6884a$。
- 同样的多边形给出两个面积均为 $18.158406=2\pi(1.7)^2$。
- 对 C 的面积作一个合理性检查：它位于长方形 $0\le x\le 4a$，$|y|\le 1.1a$ 之内，该长方形的面积约为 $8.8a^2$，而 $2\pi a^2\approx 6.3a^2$ 占其中一个合理的比例。

常见错误

- 代入 $\phi=2\theta$ 时，忘记了来自 $\mathrm{d}\theta=\frac12\mathrm{d}\phi$ 的因子 $\frac12$。因子 $4a$ 对 C 而言会变成 $2a$，要到偶函数那一步之后才再次变回 $4a$。
- 计算极坐标面积时，用了 $\int r\,\mathrm{d}\theta$ 而不是 $\frac12\int r^2\,\mathrm{d}\theta$。

▶ 把 $0 \leq \theta < 2\pi$ 范围内的整条玫瑰线 $r = 4a\cos 2\theta$ 当作 C。题目给定的范围只选出其中一瓣。

STEP 之外

积分 $\int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 \phi}\, \mathrm{d}\phi$ 称为第二类完全椭圆积分（complete elliptic integral of the second kind），它没有初等的原函数（antiderivative）。这就是题目不准计算积分的原因：椭圆的周界并没有闭合形式（closed form）。我们的积分是 $k^2 = -3$ 的情况。只要把 $\sin^2 \phi$ 换成 $1 - \cos^2 \phi$，再把 ϕ 换成 $\frac{1}{2}\pi - \phi$，便可把它改写成 $0 < k < 1$ 的标准形式。椭圆积分引出了椭圆函数（elliptic functions）和椭圆曲线（elliptic curves），而这些都是现代数论的核心。Siklos 的 *Advanced Problems in Mathematics*（《高等数学难题》）收录了更多 STEP 题目，同样以这种方式比较极坐标曲线和参数（parametric）曲线的弧长和面积。

详细解答

STEP III 1992，第 11 题

力学 · 碰撞与恢复系数

STEP III 1992，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：

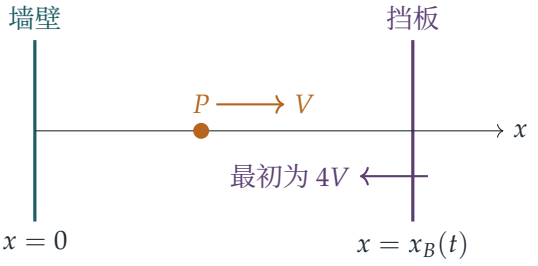
step.maths.org/questions/92-s3-q11

关键思路

所有位置一律由墙壁量起，这样质点和挡板与墙壁的距离各自都有一条简单的公式。所谓碰撞，不过是其中两条公式取值相同的一刻，于是每个关于时间的问题都化为一条二次方程（quadratic）。碰撞时，牛顿恢复定律（Newton's law of restitution）针对的是相对速度（relative velocity），所以挡板在该瞬间本身的速度至关重要。

解答

把 P 看成一个质点（particle），在光滑（smooth）桌面上沿着通过 O 并垂直于墙壁的直线运动。取 x 为与墙壁 AB 的距离，以朝向挡板的方向为正，并由抛射的一刻起计时间 t。速度以离开墙壁的方向为正。



第一次碰撞

挡板起初与墙壁相距 $\frac{75}{32}VT$，速度为 $-4V$，并有恒定加速度（constant acceleration）$+4V/T$，所以由匀加速运动公式可得

$$x_B(t) = \tfrac{75}{32}VT - 4Vt + \frac{2Vt^2}{T}, \qquad v_B(t) = -4V + \frac{4Vt}{T}.$$

在首次碰到任何东西之前，P 以恒定速率运动（因为桌面光滑），所以 $x_P = Vt$。两者在 $x_P = x_B$ 时相遇。设 $s = t/T$，则

$$2s^2 - 5s + \tfrac{75}{32} = 0 \iff 64s^2 - 160s + 75 = 0 \iff (8s-5)(8s-15) = 0.$$

首次相遇对应较小的根，所以碰撞发生于 $t = 5T/8$，正如题目所求。（在此之前，间距 $x_B - x_P$ 为正，因为它在 $s = 0$ 时为正，而其间并无任何根。）

此刻 P 位于 $x = \frac{5}{8}VT$，而挡板的速度为

$$v_B\left(\tfrac{5}{8}T\right) = -4V + \tfrac{5}{2}V = -\tfrac{3}{2}V,$$

所以挡板仍然朝着墙壁运动，速率为 $\frac{3}{2}V$。

碰撞后的速率

挡板的运动不受影响，所以只需用到牛顿实验定律（Newton's experimental law）。接近速率（speed of approach）为 $V - (-\frac{3}{2}V) = \frac{5}{2}V$。若 P 在紧接碰撞之后的速度为 w，则分离速率（speed of separation）为 $-\frac{3}{2}V - w$，而

$$-\tfrac{3}{2}V - w = \tfrac{1}{2}\cdot\tfrac{5}{2}V \implies w = -\tfrac{11}{4}V.$$

所以 P 以速率 $\dfrac{11V}{4}$ 朝墙壁反弹。

> **想一想** 为甚么反弹后的速率比 P 原先的速率 V 还要大?

碰撞的一刻，挡板正朝着 P 运动，就像球棒击球一样。恢复定律针对的是 P 相对于挡板的速率：相对于挡板，P 以 $\frac{5}{2}V$ 到达，以 $\frac{5}{4}V$ 离开。再加上挡板本身朝向墙壁的 $\frac{3}{2}V$，便得 $\frac{11}{4}V$。

从墙壁反弹与第二次碰撞

P 要以速率 $\frac{11}{4}V$ 走过距离 $\frac{5}{8}VT$，需时 $\frac{5}{8}\cdot\frac{4}{11}T = \frac{5}{22}T$。因此它撞上 AB 的时刻为

$$t_1 = \tfrac{5}{8}T + \tfrac{5}{22}T = \tfrac{75}{88}T.$$

挡板不可能在途中挡路。在这段时间内，它朝墙壁运动的速率始终比 P 慢（它的速率由 $\frac{3}{2}V$ 起不断下降），而且它根本不会到达墙壁：它最接近墙壁的时刻是 $t = T$，此时 $v_B = 0$，而 $x_B = \frac{75}{32}VT - 4VT + 2VT = \frac{11}{32}VT > 0$。

墙壁固定不动，所以当 $e=\frac{1}{2}$ 时，质点以速率 $\frac{11}{8}V$ 反弹，其后

$$x_P=\tfrac{11}{8}V\left(t-\tfrac{75}{88}T\right)=\tfrac{11}{8}Vt-\tfrac{75}{64}VT.$$

下一次与挡板碰撞，发生于 t_1 之后首次出现 $x_P=x_B$ 的时刻。以 $s=t/T$ 表示，

$$\tfrac{75}{32}-4s+2s^2=\tfrac{11}{8}s-\tfrac{75}{64}\quad\Longleftrightarrow\quad 128s^2-344s+225=0\quad\Longleftrightarrow\quad(8s-9)(16s-25)=0.$$

两根为 $s=\frac{9}{8}$ 和 $s=\frac{25}{16}$。由于 $\frac{75}{88}<1<\frac{9}{8}$，两者都在 t_1 之后。在 t_1 时，间距 $x_B-x_P=x_B(t_1)$ 为正，所以间距首次缩减至零的时刻是较小的根。因此下一次与挡板碰撞，是在开始后 $9T/8$ 的时刻，正如题目所求。（作为验算，届时两者都位于 $x=\frac{3}{8}VT$。）

想一想 为甚么必须取较小的根，而不是随便取一个根？

这条二次方程所描述的，是假设两者可以互相穿过时的运动。它的第二个根 $\frac{25}{16}T$，对应的是真正碰撞之后的一次“相遇”，但那时运动早已改变。实际的碰撞，是间距首次缩减至零的时刻。

验算

- 当 $t=\frac{5}{8}T$：$x_B=\frac{75}{32}-\frac{80}{32}+\frac{25}{32}=\frac{20}{32}=\frac{5}{8}$（以 VT 为单位），与 x_P 相等。
- 当 $t=\frac{9}{8}T$：$x_B=\frac{75}{32}-\frac{144}{32}+\frac{81}{32}=\frac{12}{32}=\frac{3}{8}$，而 $x_P=\frac{11}{8}\cdot\frac{24}{88}=\frac{3}{8}$。
- 取 $V=T=1$（步长 10^{-6}）作逐步时间模拟，得出碰撞时刻为 $t\approx0.6250$（反弹速度 -2.750）、$0.8523\approx\frac{75}{88}$ 及 1.1250。

常见错误

- 只对 P 本身的速率套用恢复定律，结果得出 $\frac{1}{2}V$ 而不是 $\frac{11}{4}V$。挡板正在运动，所以要用相对速度。
- 弄错加速度的正负号。挡板起初朝墙壁运动，但正在减速，并在 $t=T$ 时掉头。
- 忘记检查在 P 从 AB 反弹之前，挡板不会先碰到 P（或墙壁）。

STEP 之外

这里的巧妙之处，是在碰撞的瞬间改用挡板的参考系（frame of reference）来考虑。由于挡板的质量极大，恢复定律会把相对速度反向并按比例缩小，与固定的墙壁无异。这样转换参考系是大学一年级力学的核心技巧，并由此引伸出伽利略不变性（Galilean invariance），

以及二体碰撞中的质心参考系（centre-of-mass frame）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学导论）对这些概念有细致的阐述。

详细解答

STEP III 1992，第 12 题

力学 · 动量守恒

STEP III 1992，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：

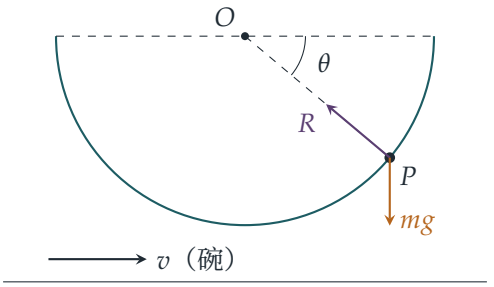
step.maths.org/questions/92-s3-q12

关键思路

碗与质点组成的系统并无受到任何水平方向的外力，而且除重力外没有其他力作功，所以我们有两条守恒定律：水平动量守恒（conservation of horizontal momentum）和能量守恒（conservation of energy）。动量把碗的速率与 $\dot{\theta}$ 联系起来；能量随后给出 v^2。至于反作用力（reaction），最简洁的做法是在碗的（加速）参考系（frame）中处理，在那里质点只是绕 O 作圆周运动；同时要记住，反作用力也是碗所受的唯一水平力。

解答

模型。碗是光滑的，所以 P 所受的反作用力 R 沿法线方向，即沿 PO 方向。小车是轻的，桌面是光滑的，所以碗（连同小车）所受的唯一水平力，就是质点推力的水平分量。设 P 从碗边的右端出发，于是当 P 向下滑时，θ 从 0 开始增加；碗则向右移动。



水平动量守恒

设碗的水平位置为 X，则 $\dot{X} = v$。相对于 O，质点位于 $(a\cos\theta,\ -a\sin\theta)$，所以它的位置是 $(X + a\cos\theta,\ -a\sin\theta)$，而它的速度是

$$\big(v - a\dot{\theta}\sin\theta,\ -a\dot{\theta}\cos\theta\big).$$

系统没有受到水平方向的外力，而且由静止开始，所以总水平动量始终为零：

$$2mv + m\big(v - a\dot{\theta}\sin\theta\big) = 0 \quad\Longrightarrow\quad 3v = a\dot{\theta}\sin\theta,$$

这正是所要求的结果。留意此时质点的水平速度是 $v - 3v = -2v$：它以碗的速率的两倍向左移动，这样才能抵销碗的动量 $2mv$。

能量

反作用力所作的净功为零（它是内力，而且表面光滑），桌面也不作功，所以能量守恒。当 OP 处于角度 θ 时，质点已下降了 $a\sin\theta$。利用 $a\dot{\theta} = 3v/\sin\theta$，质点的垂直速率为 $a\dot{\theta}\cos\theta = 3v\cot\theta$，所以

$$\begin{aligned} mga\sin\theta &= \tfrac{1}{2}(2m)v^2 + \tfrac{1}{2}m\big[(2v)^2 + 9v^2\cot^2\theta\big] \\ &= mv^2\Big(3 + \tfrac{9}{2}\cot^2\theta\Big). \end{aligned}$$

两边乘以 $2\sin^2\theta/m$：

$$2ga\sin^3\theta = v^2\big(6\sin^2\theta + 9\cos^2\theta\big) = v^2\big(9 - 3\sin^2\theta\big),$$

所以

$$v^2 = \frac{2ga\sin^3\theta}{3(3 - \sin^2\theta)},$$

这就是第二个结果。

反作用力

> **想一想** 碗正在加速。我们怎样写出质点沿 PO 方向的运动方程（equation of motion）？

用两条方程。第一，考虑碗：质点以大小为 R、方向由 O 指向 P 的力推它，这个力的水平分量是向右的 $R\cos\theta$。所以碗的加速度 A 满足

$$2mA = R\cos\theta. \tag{1}$$

第二，考虑质点，在碗的参考系中处理。在那里质点以速率 $a\dot{\theta}$ 在以 O 为圆心、半径为 a 的圆上运动，而除了 R 和 mg 之外，它还受到一个向左的假想力（fictitious force）mA。

沿指向 O 的方向（方向为 $(-\cos\theta, \sin\theta)$）分解，重力的贡献是 $-mg\sin\theta$，假想力的贡献是 $+mA\cos\theta$，所以

$$R - mg\sin\theta + mA\cos\theta = ma\dot{\theta}^2. \tag{2}$$

（等价地，在地面参考系中，质点的加速度等于碗的加速度 A 加上它相对于碗的圆周运动加速度；把它沿 PO 方向分解，会得到同一条方程。）

由头两个结果，

$$a\dot{\theta}^2 = \frac{9v^2}{a\sin^2\theta} = \frac{6g\sin\theta}{3-\sin^2\theta}.$$

把由 (1) 得到的 $mA = \frac{1}{2}R\cos\theta$ 代入 (2)，并把含 R 的项合并：

$$R\left(1 + \tfrac{1}{2}\cos^2\theta\right) = mg\sin\theta + \frac{6mg\sin\theta}{3-\sin^2\theta} = mg\sin\theta\,\frac{9-\sin^2\theta}{3-\sin^2\theta}.$$

由于 $1 + \frac{1}{2}\cos^2\theta = \frac{1}{2}(3-\sin^2\theta)$，

$$\boxed{R = \frac{2mg\sin\theta\,(9-\sin^2\theta)}{(3-\sin^2\theta)^2} = \frac{2mg\sin\theta\,(8+\cos^2\theta)}{(2+\cos^2\theta)^2}.}$$

验算

- 在起点，$\theta = 0$ 给出 $R = 0$，这是应有的结果：质点静止于碗边，没有任何东西把它压向碗。
- 在碗底，$\theta = \frac{1}{2}\pi$：碗的加速度为零（由 (1)），$a\dot{\theta}^2 = 3g$，而 $R - mg = 3mg$，所以 $R = 4mg$，与公式给出的 $2mg\cdot 8/2^2$ 吻合。
- 假如碗是固定的（无限重），通常的结果会是 $v_P^2 = 2ga\sin\theta$ 和 $R = 3mg\sin\theta$；碗的反冲令 R 在碗边附近变小（当 θ 很小时约为 $2mg\sin\theta$ 而不是 $3mg\sin\theta$，因为碗正在远离 P），但在碗底却变大（$4mg$ 而不是 $3mg$）。
- 以数值方法逐步求解完整的运动方程（取 $m = 1.3$、$a = 0.7$、$g = 9.81$），由碗边一直到碗底，都重现了 $3v = a\dot{\theta}\sin\theta$、$v^2$ 的公式和 R 的公式，误差在 10^{-7} 以内。

常见错误

- 写成 $R - mg\sin\theta = ma\dot{\theta}^2$，仿佛碗是固定的。碗在加速，所以质点的加速度并不只是它相对于碗的圆周运动加速度。
- 用 $\frac{1}{2}m(a\dot{\theta})^2$ 作为质点的动能（kinetic energy）：$a\dot{\theta}$ 是它相对于碗的速率，而不是相对于桌面的速率。

STEP 之外

这是一个二自由度（two-degree-of-freedom）系统，而最简洁、最有系统的处理方法是拉格朗日力学（Lagrangian mechanics）：把动能和势能（potential energy）用 X 和 θ 表示，动量守恒便会自然得出，因为 X 并不出现在拉格朗日量（Lagrangian）之中（它是一个“循环”（cyclic）坐标）。我们所用的假想力，是非惯性参考系（non-inertial frame）力学中最简单的情形；大学一年级的课程会把它推广到旋转参考系，并引入离心力（centrifugal force）和科里奥利力（Coriolis force）。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对非惯性参考系和相互作用物体的系统都有很出色的讲解。

详细解答

STEP III 1992，第 13 题

力学 • 绕固定轴的转动

STEP III 1992，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/92-s3-q13

关键思路

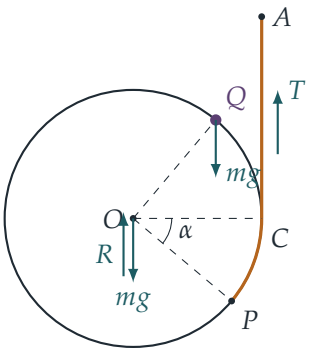
转轴固定，所以圆盘与质点构成一个刚体（rigid body），只需一个坐标，即转过的角度。由能量守恒（conservation of energy）（弹性势能、重力势能、动能）可得每个位置的角速率（angular speed），而“P 到达 C”的意思不过是动能在途中始终不会降至零。求得在 C 处的角速率后，把牛顿（Newton）第二定律应用于整个系统的质心（centre of mass），便可求出转轴上的反作用力（reaction）。

解答

模型与几何

我们把 Q 视为质点，并把弦线视为轻弦。我们假设盘边光滑，所以整条弦线上的张力 T 处处相同，而弦线对盘边的法向接触力都通过 O。因此，弦线对 O 的力矩（moment）只有在 P 处沿切线方向拉扯的张力所产生的力矩，即 $2bT$。（若把张力看成作用于弦线在 C 处离开盘边的位置，所得的力矩完全相同：张力沿 AC 铅垂向上，与 O 的距离为 $2b$。）

设 $\alpha = \angle POC$，其中 P 在 OC 下方。释放时 $\alpha = \frac{\pi}{2}$，而 P 到达 C 时 $\alpha = 0$。由于 $\angle POQ = 90^\circ$，OQ 在 OC 上方并与之成角 $\frac{\pi}{2} - \alpha$，所以 Q 的高度比 O 高 $2b\cos\alpha$。特别地，Q 起初与 O 同高（位于 C），而当 P 到达 C 时，它在 O 的正上方。



弦线的总长度为 $AC + \text{弧}CP = \pi b + 2b\alpha$，而其自然长度（natural length）为 πb，所以伸长量（extension）为 $2b\alpha$。由胡克定律（Hooke's law），

$$T = \frac{kmg \cdot 2b\alpha}{\pi b} = \frac{2kmg\alpha}{\pi},$$

而所储存的弹性势能（elastic energy）为

$$E(\alpha) = \frac{kmg\,(2b\alpha)^2}{2\pi b} = \frac{2kmgb\,\alpha^2}{\pi}.$$

对转轴的转动惯量（moment of inertia）等于圆盘的转动惯量加上盘边上 Q 的转动惯量：

$$I = \tfrac{1}{2}m(2b)^2 + m(2b)^2 = 6mb^2.$$

P 到达 C

当圆盘转动，使 P 向 C 上升（即 α 减小）时，弦线缩短并释出弹性势能，同时 Q 上升。转轴固定而且光滑，所以能量守恒。系统在 $\alpha = \frac{\pi}{2}$ 处由静止开始运动，此时 Q 相对于 O 的高度为 0，且 $E = \frac{1}{2}\pi kmgb$，于是

$$\begin{aligned}\tfrac{1}{2}I\dot{\alpha}^2 &= E\left(\tfrac{\pi}{2}\right) - E(\alpha) - mg \cdot 2b\cos\alpha \\ &= \frac{2kmgb}{\pi}\left(\frac{\pi^2}{4} - \alpha^2\right) - 2mgb\cos\alpha.\end{aligned}$$

圆盘确实会朝正确的方向开始转动：释放时弦线以 $T = kmg$ 在半径 $2b$ 处拉扯，而 Q 的重量则在水平距离 $2b$ 处产生反方向的力矩，所以净力矩为 $2(k-1)mgb > 0$，因为 $k > \frac{4}{\pi} > 1$。

P 能到达 C，只要对 $[0, \frac{\pi}{2})$ 内的每个 α，动能都保持为正，即只要

$$\frac{k}{\pi}\left(\frac{\pi^2}{4} - \alpha^2\right) > \cos\alpha \qquad \text{对于} 0 \le \alpha < \tfrac{\pi}{2}.$$

由于 $k > \frac{4}{\pi}$ 且 $\frac{\pi^2}{4} - \alpha^2 > 0$，左边大于

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

所以只需证明在 $[0, \frac{\pi}{2}]$ 上有 $g(\alpha) = 1 - \dfrac{4\alpha^2}{\pi^2} - \cos\alpha \geq 0$。

想一想 怎样在整个区间上比较一个二次函数与 $\cos\alpha$?

考虑二阶导数（second derivative）。我们有 $g(0) = g(\frac{\pi}{2}) = 0$，$g'(\alpha) = -\frac{8\alpha}{\pi^2} + \sin\alpha$，所以 $g'(0) = 0$，而 $g''(\alpha) = \cos\alpha - \frac{8}{\pi^2}$。由于 $\frac{8}{\pi^2} < 1$，而 $\cos\alpha$ 由 1 递减至 0，故存在唯一的 α_0，使得在 $[0, \alpha_0)$ 上 $g'' > 0$，在 $(\alpha_0, \frac{\pi}{2}]$ 上 $g'' < 0$。

- 在 $[0, \alpha_0]$ 上，g' 由 $g'(0) = 0$ 开始递增，所以 $g' \geq 0$，从而 $g \geq g(0) = 0$。
- 在 $[\alpha_0, \frac{\pi}{2}]$ 上，g 是凹（concave）的，所以它的图像位于连接两端点值 $g(\alpha_0) \geq 0$ 与 $g(\frac{\pi}{2}) = 0$ 的弦之上；因此在该处同样有 $g \geq 0$。

所以 $g \geq 0$ 处处成立，而严格不等式 $k > \frac{4}{\pi}$ 使动能在 $0 \leq \alpha < \frac{\pi}{2}$ 时严格为正。圆盘在 $\alpha = 0$ 之前不会静止下来，所以 P 到达 C。特别地，在 C 处（代入 $\alpha = 0$），

$$3mb^2\dot{\alpha}^2 = \frac{\pi k m g b}{2} - 2mgb \quad \Longrightarrow \quad \omega^2 = \frac{(\pi k - 4)g}{6b},$$

其中 ω 是 P 位于 C 时的角速率。（由此可见条件 $k > \frac{4}{\pi}$ 的来由：它正好保证 $\omega^2 > 0$。）

转轴上的反作用力

当 P 位于 C 时，有两件事同时发生。

- 弦线处于自然长度（$AC = \pi b$，而弧段已消失），所以 $T = 0$。
- Q 在 O 的正上方，所以它的重量对 O 没有力矩。

所以在该瞬间对转轴没有力矩，角加速度为零。圆盘的中心 O 固定不动，而 Q 沿半径为 $2b$ 的圆运动，角加速度为零，所以它的加速度纯粹是向心（centripetal）的：大小为 $2b\omega^2$，方向由 Q 指向 O，即铅垂向下。

想一想 求反作用力的方程中应包括哪些力?

对整个系统（圆盘连同 Q）在铅垂方向应用牛顿第二定律。外力包括两个重量，合共 $2mg$ 向下；转轴上的反作用力 R 向上；以及弦线的张力，它在此瞬间为零。Q 与圆盘之间的内力互相抵消。只有 Q 有加速度，所以

$$R - 2mg = m\,(-2b\omega^2),$$

因此

$$R = 2mg - 2mb \cdot \frac{(\pi k - 4)g}{6b} = 2mg - \frac{mg(\pi k - 4)}{3} = \frac{mg(10 - \pi k)}{3},$$

正如所求。（反作用力的水平分量在此瞬间为零，因为 Q 没有水平加速度。）

验算

- 在临界情况 $k = \frac{4}{\pi}$ 下，系统仅仅能够缓缓升到 C，此时 $\omega = 0$，而公式给出 $R = 2mg$，即静止时的值。两者吻合。
- 用电脑对运动方程 $6mb^2\ddot{\psi} = 2bT - 2mgb\cos\psi$（其中 $\psi = \frac{\pi}{2} - \alpha$）作数值积分。取 $m = g = b = 1$ 及 $k = 2$，系统到达 C 时 $\omega^2 = 0.38053$，与 $(2\pi - 4)/6$ 相符，且 $R = 1.23894 = (10 - 2\pi)/3$。对于 $k = 1.29,\ 1.5,\ 3$，结果同样准确至小数点后六位。
- 若 $k > \frac{10}{\pi}$，公式给出 $R < 0$：转轴必须向下拉，因为 Q 越过顶点时转得太快，所需的向心力超过了系统的重量。

常见错误

- 把伸长量自始至终当作 πb。事实上，当 P 到达 C 时，伸长量会缩减至零，这正是张力在该处消失的原因。
- 忘记圆盘本身的转动惯量 $\frac{1}{2}m(2b)^2$，只用了 Q 的转动惯量。
- 只证明在 C 处的动能为正。“P 到达 C”要求动能在整段路程上都保持为正。

STEP 之外

最后一步是“外力总和等于总质量乘以质心加速度”的刚体版本，而求支点上的反作用力正是它的经典应用。大学一年级的刚体动力学课程会把它与角动量方程 $\mathrm{d}\mathbf{L}/\mathrm{d}t = \boldsymbol{\tau}$ 一并发展。不等式 $\cos\alpha \le 1 - 4\alpha^2/\pi^2$ 是约当不等式（Jordan's inequality）$\sin x \ge 2x/\pi$ 的近亲，这里是用标准的凸性（convexity）论证来证明的。Kleppner 与 Kolenkow 的 *An Introduction to Mechanics*（力学入门教科书）对定轴转动及支点反作用力有非常清晰的讲解。

详细解答

STEP III 1992，第 14 题

力学 · 解绕绳的动力学

> **STEP III 1992，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/92-s3-q14

关键思路

整个问题只由一个角 θ 决定。绳子未贴着圆盘的部分在 Q 点与圆盘相切（tangent），其长度等于已从圆盘松开的那段绳子的弧长，因此很容易用 θ 表示 P 的位置（P 所描出的曲线是圆的渐伸线（involute））。求导之后，会发现 P 的速度（velocity）与绳子垂直。所以绳子的张力（tension）不作功，沿运动方向只有摩擦力（friction）起作用。于是沿 PQ 方向和垂直于它的方向分解，便可把问题干净利落地一分为二。

解答

P 的位置

把 P 视为水平桌面上的一个质点（particle），所受阻力只有题目给出的摩擦力，并把绳子视为轻而不可伸长（light and inextensible）。重力与法向反作用力（normal reaction）互相抵销，所以水平方向的力只有沿 PQ 的张力 T 和摩擦力。

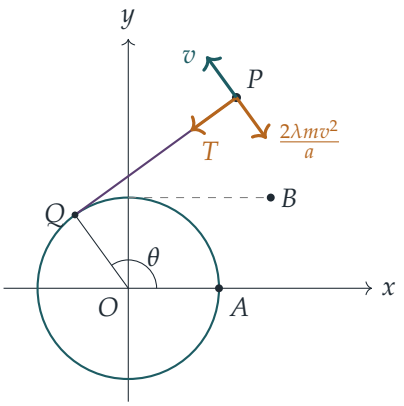
取 $A = (a, 0)$ 在 x 轴上，则切点为 $Q = (a\cos\theta,\, a\sin\theta)$。当 P 位于 A 时，整条绳子都绕在圆盘上；由于绳子不可伸长，松开的部分 QP 的长度等于已解开的弧 AQ 的长度，即 $a\theta$。绳子在 Q 点与圆盘相切，所以 QP 垂直于 OQ，方向为单位向量（unit vector）$(\sin\theta,\, -\cos\theta)$（当 $\theta = \frac{\pi}{2}$ 时，这向量是 $(1, 0)$，由圆盘顶点指向 B，如图所示）。因此

$$\overrightarrow{OP} = a(\cos\theta,\, \sin\theta) + a\theta(\sin\theta,\, -\cos\theta),$$

所以

$$x = a\cos\theta + a\theta\sin\theta, \qquad y = a\sin\theta - a\theta\cos\theta.$$

第一条式子就是所求的 x 坐标。在 B 点，松开的绳子与 OA 平行，并在圆盘顶点与圆盘相切，所以该处 $\theta = \frac{\pi}{2}$，而 $B = (\frac{\pi a}{2},\, a)$。



速度与加速度

设 $\mathbf{e} = (\cos\theta, \sin\theta)$ 为沿 OQ 方向的单位向量，$\mathbf{f} = (-\sin\theta, \cos\theta)$ 为由 P 指向 Q 的单位向量（与上文的绳子方向相反）。注意 $\dot{\mathbf{e}} = \dot\theta\,\mathbf{f}$ 及 $\dot{\mathbf{f}} = -\dot\theta\,\mathbf{e}$。

对 x 和 y 求导，

$$\dot x = (-a\sin\theta + a\sin\theta + a\theta\cos\theta)\,\dot\theta = a\theta\dot\theta\cos\theta,$$
$$\dot y = (a\cos\theta - a\cos\theta + a\theta\sin\theta)\,\dot\theta = a\theta\dot\theta\sin\theta,$$

所以速度为 $a\theta\dot\theta\,\mathbf{e}$：它与 OQ 平行，因而与绳子垂直。利用乘法法则（product rule）再求导一次，

$$\ddot{\mathbf{r}} = a\left(\dot\theta^2 + \theta\ddot\theta\right)\mathbf{e} + a\theta\dot\theta\,\dot{\mathbf{e}} = a\left(\dot\theta^2 + \theta\ddot\theta\right)\mathbf{e} + a\theta\dot\theta^2\,\mathbf{f}.$$

所以加速度沿 PQ 方向（指向 Q）的分量为 $a\theta\dot\theta^2$，而垂直于 PQ 方向（即 $\mathbf{e}$ 的方向）的分量为 $a\dot\theta^2 + a\theta\ddot\theta$，正如题目所要求。

运动方程

在 B 点，$\theta = \frac{\pi}{2}$，速度为 $a\cdot\frac{\pi}{2}\dot\theta\,(0,1)$。质点沿垂直于 OA 并远离 OA 的方向抛出，即表示速度 V 沿 y 轴正方向，所以

$$\dot\theta = \frac{2V}{\pi a} > 0 \quad \text{（起初）。}$$

只要 $\dot\theta > 0$，速率便是 $v = a\theta\dot\theta$，而运动方向为 $+\mathbf{e}$，所以大小为 $2\lambda mv^2/a = 2\lambda ma\theta^2\dot\theta^2$ 的摩擦力沿 $-\mathbf{e}$ 方向作用。张力沿 $\mathbf{f}$ 方向作用，没有 $\mathbf{e}$ 分量。

> ? **想一想** 要得到一条只含 θ 的方程，应该用加速度两个分量中的哪一个?

应该用垂直的那个：张力 T 是未知的，而它只出现在沿 PQ 方向的分量之中。在垂直于

PQ 的方向应用牛顿第二定律（Newton's second law），得

$$m\left(a\dot{\theta}^2 + a\theta\ddot{\theta}\right) = -2\lambda ma\theta^2\dot{\theta}^2.$$

两边除以 $ma\theta\dot{\theta}$（这是容许的，因为 $\theta > 0$ 且 $\dot{\theta} > 0$），

$$\frac{\ddot{\theta}}{\dot{\theta}} = -\left(\frac{1}{\theta} + 2\lambda\theta\right)\dot{\theta},$$

这就是所求的结果。

求 $\dot{\theta}$

左边是 $\frac{\mathrm{d}}{\mathrm{d}t}\ln\dot{\theta}$，而根据链式法则（chain rule），右边是 $\frac{\mathrm{d}}{\mathrm{d}t}\left(-\ln\theta - \lambda\theta^2\right)$。因此

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\ln(\theta\dot{\theta}) + \lambda\theta^2\right) = 0, \qquad \text{所以} \qquad \theta\dot{\theta}\,\mathrm{e}^{\lambda\theta^2} = \text{常数}.$$

在 B 点，这个常数为 $\frac{\pi}{2}\cdot\frac{2V}{\pi a}\,\mathrm{e}^{\lambda\pi^2/4} = \frac{V}{a}\,\mathrm{e}^{\lambda\pi^2/4}$。于是

$$\dot{\theta} = \frac{V}{a\theta}\,\exp\left(\lambda\left(\frac{\pi^2}{4} - \theta^2\right)\right).$$

这对任何 θ 都是正数，所以 $\dot{\theta}$ 永不为零，上文所用的假设 $\dot{\theta} > 0$ 自始至终成立。（答案除了 θ、λ 和 a 之外，必然还涉及 V。）等价地说，速率为 $v = a\theta\dot{\theta} = V\mathrm{e}^{\lambda(\pi^2/4-\theta^2)}$。

$\theta = \pi$ 时的张力

沿 PQ 方向只有 T 这一个力，它指向 Q，所以 $T = ma\theta\dot{\theta}^2$。当 $\theta = \pi$ 时，

$$\dot{\theta}^2 = \frac{V^2}{a^2\pi^2}\,\mathrm{e}^{2\lambda(\pi^2/4-\pi^2)} = \frac{V^2}{a^2\pi^2}\,\mathrm{e}^{-3\lambda\pi^2/2},$$

所以

$$T = ma\pi\cdot\frac{V^2}{a^2\pi^2}\,\mathrm{e}^{-3\lambda\pi^2/2} = \frac{mV^2}{\pi a}\,\mathrm{e}^{-3\lambda\pi^2/2}.$$

验算

- 松开的绳子 PQ 长度为 $a\theta$，而它正是渐伸线的曲率半径（radius of curvature），所以 Q 是曲率中心（centre of curvature），指向 Q 的加速度分量应为 $v^2/(a\theta) = a\theta\dot{\theta}^2$。结果确是如此。
- 当 $\lambda = 0$（没有摩擦力）时，速率保持为 V；这是必然的，因为张力与运动方向垂直，不作功。在 $\theta = \frac{\pi}{2}$ 时，$T = mV^2/(\frac{\pi}{2}a)$，这正是以 B 点的绳长 $r = \frac{\pi a}{2}$ 代入 mv^2/r 的结果。

- 数值方面，取 $a = 1.3$、$m = 0.7$、$V = 2.1$、$\lambda = 0.25$，我们用龙格－库塔法（Runge–Kutta method）逐步求解 θ 的方程：结果与 $\dot{\theta}$ 的公式吻合，误差在 10^{-14} 以内。以有限差分（finite difference）计算 (x, y) 的加速度，所得结果与两个分量及摩擦力均吻合，并在 $\theta = \pi$ 时得出 $T \approx 0.018667$，与公式一致。

常见错误

- 以 $\theta = 0$ 作为初始值：质点由 B 出发，该处 $\theta = \frac{\pi}{2}$，而不是由 A 出发。
- 沿 PQ 方向分解来求 $\dot{\theta}$：那条方程含有未知的张力，而摩擦力在该方向没有分量。
- 忘记摩擦力与速度方向相反，而速度是沿 OQ 方向，不是沿绳子方向。

STEP 之外

P 的路径是圆的渐伸线，即拉紧的绳子从线轴上松开时，绳端所描出的曲线，也是大多数齿轮轮齿所采用的形状。这里用到的事实（速度与绳子垂直，以及绳长就是曲率半径），都是曲线微分几何（differential geometry）中渐屈线（evolute）与渐伸线关系的特殊情况。从动力学的角度看，这是一个很好的例子，说明如何利用内禀（intrinsic）分量（即切向和法向分量）把已知的力与未知的约束力（constraint force）分开。Kleppner 与 Kolenkow 所著的 *An Introduction to Mechanics*（力学入门教材）正是以这种思路讨论极坐标（polar coordinates）和内禀坐标下的运动，以及受约束的运动。

详细解答

STEP III 1992，第 15 题

概率与统计 · 几何概率

> **STEP III 1992，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/92-s3-q15

关键思路

由于山羊的位置在田地上均匀分布（uniformly distributed），每一个概率都只是某块面积除以 a^2。所以整道题归根究底，就是要求出正方形内与角 O 的距离小于 r 的那部分的面积。当 r 超过 a 之后，四分之一圆会溢出正方形以外，而要求出仍留在正方形内的面积，最利落的做法是把它切成两个直角三角形和一个扇形（sector）。之后把分布函数微分便得到密度，而系绳的山羊则是同一个想法，只是换了一个区域。

解答

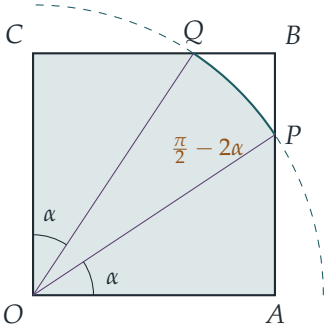
把 O 放在原点，OA 沿 x 轴，OC 沿 y 轴，于是田地就是 $0 \le x \le a$、$0 \le y \le a$。记 $F(r) = \mathrm{P}(R < r)$，它等于田地中满足 $x^2 + y^2 < r^2$ 的那部分的面积除以 a^2。

R 的分布

当 $0 < r \le a$ 时。以 O 为圆心、半径为 r 的四分之一圆盘完全落在田地之内，所以

$$F(r) = \frac{\frac{1}{4}\pi r^2}{a^2} = \frac{\pi r^2}{4a^2}.$$

当 $a \le r \le a\sqrt{2}$ 时。这时圆弧与边 AB（即直线 $x = a$）相交于 $P = \left(a, \sqrt{r^2 - a^2}\right)$，与边 CB（即直线 $y = a$）相交于 $Q = \left(\sqrt{r^2 - a^2}, a\right)$。设 $\alpha = \cos^{-1}(a/r)$，它就是角 AOP（因为 $\cos \angle AOP = a/r$）。根据对称性，角 COQ 也等于 α。



阴影区域（即田地中与 O 的距离小于 r 的部分）可分成三块：

- 三角形 OAP，直角在 A，两条直角边分别为 a 和 $\sqrt{r^2-a^2}$，所以面积为 $\frac{1}{2}a\sqrt{r^2-a^2}$；
- 三角形 OCQ，与它全等（congruent），面积相同；
- 扇形 OPQ，半径为 r，圆心角为 $\frac{\pi}{2}-2\alpha$，面积为 $\frac{1}{2}r^2\left(\frac{\pi}{2}-2\alpha\right)=\frac{\pi r^2}{4}-r^2\alpha$。

把三者相加再除以 a^2，

$$\begin{aligned}F(r)&=\frac{a\sqrt{r^2-a^2}}{a^2}+\frac{\pi r^2}{4a^2}-\frac{r^2}{a^2}\,\alpha\\&=\left(\frac{r^2}{a^2}-1\right)^{\frac{1}{2}}+\frac{\pi r^2}{4a^2}-\frac{r^2}{a^2}\cos^{-1}\left(\frac{a}{r}\right),\end{aligned}$$

这正是所求的结果。（当 $r\geq a\sqrt{2}$ 时，整块田地与 O 的距离都小于 r，所以 $F(r)=1$。在 $r=a\sqrt{2}$ 处，公式给出 $1+\frac{\pi}{2}-2\cdot\frac{\pi}{4}=1$，与预期相符。）

中位数

在 $r=a$ 处有 $F(a)=\pi/4\approx 0.785>\frac{1}{2}$，所以中位数（median）$m$ 落在 $0<m<a$ 的范围内，这里适用的是较简单的那条公式：

$$\frac{\pi m^2}{4a^2}=\frac{1}{2}\quad\Longrightarrow\quad m=a\sqrt{\frac{2}{\pi}}\approx 0.798a.$$

概率密度函数

概率密度函数（probability density function）为 $f(r)=F'(r)$。当 $0<r<a$ 时，$f(r)=\dfrac{\pi r}{2a^2}$。

当 $a<r<a\sqrt{2}$ 时，我们需要 $\cos^{-1}(a/r)$ 的导数：

$$\frac{\mathrm{d}}{\mathrm{d}r}\cos^{-1}\left(\frac{a}{r}\right)=-\frac{1}{\sqrt{1-a^2/r^2}}\cdot\left(-\frac{a}{r^2}\right)=\frac{a}{r\sqrt{r^2-a^2}}.$$

然后对 F 逐项微分（最后一项要用积法则（product rule）），

$$\begin{aligned} f(r) &= \frac{r}{a\sqrt{r^2-a^2}} + \frac{\pi r}{2a^2} - \frac{2r}{a^2}\cos^{-1}\left(\frac{a}{r}\right) - \frac{r^2}{a^2}\cdot\frac{a}{r\sqrt{r^2-a^2}} \\ &= \frac{r}{a^2}\left(\frac{\pi}{2} - 2\cos^{-1}\left(\frac{a}{r}\right)\right), \end{aligned}$$

因为第一项与最后一项互相抵消。所以

$$f(r) = \begin{cases} \dfrac{\pi r}{2a^2}, & 0 < r \le a, \\ \dfrac{r}{a^2}\left(\dfrac{\pi}{2} - 2\cos^{-1}\left(\dfrac{a}{r}\right)\right), & a < r \le a\sqrt{2}, \\ 0, & \text{其他情况。} \end{cases}$$

这在几何上也说得通：$f(r)\,\mathrm{d}r$ 等于（半径为 r 的圆在田地内的弧长）$\times\,\mathrm{d}r/a^2$，而该段弧所对的圆心角在 $r \le a$ 时为 $\frac{\pi}{2}$，在 $r > a$ 时为 $\frac{\pi}{2} - 2\alpha$。两段公式在 $r = a$ 处吻合（都给出 $\pi/(2a)$），而 f 在 $r = a\sqrt{2}$ 处降至 0，那里 $\alpha = \pi/4$。

系绳的山羊

山羊被系住之后，只能到达田地中 $R \le a$ 的点，即以 O 为圆心、半径为 a 的四分之一圆盘，其面积为 $\frac{1}{4}\pi a^2$。牠的位置仍然是均匀分布，只是现在分布在这个四分之一圆盘上。这恰好就是原来的模型以事件 $R \le a$ 为条件之后的情况。

> ? **想一想** 哪一个坐标量度的是与围栏 OC 的距离?

由于 OC 沿 y 轴，与 OC 的距离就是 x。所以我们需要求出四分之一圆盘中 $x > a/2$ 的那部分的面积：

$$\begin{aligned} \int_{a/2}^{a} \sqrt{a^2-x^2}\,\mathrm{d}x &= \left[\frac{x\sqrt{a^2-x^2}}{2} + \frac{a^2}{2}\sin^{-1}\frac{x}{a}\right]_{a/2}^{a} \\ &= \frac{\pi a^2}{4} - \left(\frac{\sqrt{3}\,a^2}{8} + \frac{\pi a^2}{12}\right) = \frac{\pi a^2}{6} - \frac{\sqrt{3}\,a^2}{8}. \end{aligned}$$

（从几何上看，这是一个圆心角为 $\pi/3$ 的扇形减去一个边长为 $a/2$ 和 $\sqrt{3}a/2$ 的三角形，因为直线 $x = a/2$ 与圆弧的交点和 OA 所成的角为 $\pi/3$。）除以面积 $\frac{1}{4}\pi a^2$，

$$\mathrm{P}\left(\text{距}OC > \tfrac{a}{2} \,\middle|\, R \le a\right) = \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \approx 0.391.$$

验算

- 连续性：F 的两条公式在 $r = a$ 处都给出 $\pi/4$，而且 $F(a\sqrt{2}) = 1$。

- 以单位正方形内 4×10^5 个均匀分布的点进行蒙地卡罗模拟（Monte Carlo simulation），结果在 $r = 0.3, 0.7, 1.1, 1.25, 1.4$ 处与 $F(r)$ 吻合至大约小数点后三位。模拟得出的样本中位数为 0.7986，而 $\sqrt{2/\pi} = 0.7979$；条件比例为 0.3914，而 $\frac{2}{3} - \frac{\sqrt{3}}{2\pi} = 0.3910$。
- 密度在 $[0, a\sqrt{2}]$ 上的数值积分为 1，并且在若干个点上与 F 的数值导数相符。
- 系绳情况的答案小于 $\frac{1}{2}$，这是应有的结果：四分之一圆盘在远离 OC 的地方较窄。

常见错误

- 求中位数时忘记了 F 在 $r > a$ 时要用另一条公式。值得先检查 $F(a) > \frac{1}{2}$，由此可知中位数落在容易处理的范围内。
- 把扇形的圆心角写成 $\frac{\pi}{2} - \alpha$ 而不是 $\frac{\pi}{2} - 2\alpha$。被切去的角位有两块，两边各一。
- 在系绳的部分除以 a^2。条件概率（conditional probability）应除以四分之一圆盘的面积 $\frac{1}{4}\pi a^2$。

STEP 之外

这里的 R 是在某个区域内均匀选取的一点的函数，而它的分布是通过量度 $R < r$ 的区域而得出的。这正是求随机变量（random variable）的函数的分布的一般方法，而密度公式 $f(r) = (\text{弧长})/a^2$ 则让人初尝“余面积”（coarea）的想法，即把一个区域切成一片片水平集（level sets）。同样的几何推理也是布丰投针（Buffon's needle）以及正方形内两个随机点之间的距离等问题的基础。Grimmett 与 Welsh 的 *Probability: An Introduction*，以及 Blitzstein 与 Hwang 的 *Introduction to Probability*，都有阐述这种处理连续随机变量的几何方法。

详细解答

STEP III 1992，第 16 题

概率与统计 · 泊松稀疏化

> **STEP III 1992，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/92-s3-q16

关键思路

一旦错字的总数固定下来，每个错字都是一次独立试验（independent trial），我要么发现它，要么看漏，所以第 (i) 部分是一个二项概率（binomial probability）。第 (ii) 部分是条件概率（conditional probability），而分母，即我恰好发现 m 个错字的概率，需要我们对所有可能的看漏数目逐一相加。这个和原来是一个指数级数（exponential series），正因如此，整个式子才会化简成简洁的泊松（Poisson）形式。

解答

全篇记 $q = 1 - p$。设 N 为该期报纸的错字总数，S 为我发现的错字数目，$U = N - S$ 为我未能发现的数目。题目告诉我们

$$\mathrm{P}(N = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

第 (i) 部分

已知 $N = m + n$ 时，这 $m + n$ 个错字各自以概率 p 被发现，彼此独立。因此，在 $N = m + n$ 的条件下，被发现的错字数目服从二项分布（binomial distribution）$\mathrm{B}(m + n, p)$。任意指定哪 m 个错字被发现、哪 n 个被看漏，其概率都是 $p^m q^n$，而这样的选法共有 $\binom{m+n}{m}$ 种，所以

$$\mathrm{P}(S = m \mid N = m + n) = \binom{m+n}{m} p^m (1-p)^n.$$

第 (ii) 部分

我们要求 $\mathrm{P}(U = n \mid S = m)$。事件 $\{S = m,\ U = n\}$ 与 $\{N = m+n,\ S = m\}$ 是同一事件，所以由第 (i) 部分可得

$$\begin{aligned}\mathrm{P}(S = m,\ U = n) &= \mathrm{P}(N = m+n)\,\mathrm{P}(S = m \mid N = m+n)\\ &= \frac{\mathrm{e}^{-\lambda}\lambda^{m+n}}{(m+n)!}\cdot\frac{(m+n)!}{m!\,n!}\,p^m q^n\\ &= \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\cdot\frac{(\lambda q)^n}{n!}.\end{aligned}$$

因子 $(m+n)!$ 刚好约去，这正是本题得以顺利解出的关键。

> **想一想** 我们不知道总共有多少个错字，那么怎样求 $\mathrm{P}(S = m)$ 呢?

如果我恰好发现 m 个，看漏的数目可以是 $0, 1, 2, \ldots$ 中的任何一个，而这些情况互斥（mutually exclusive）。因此我们把联合概率（joint probability）对 U 的所有取值 j 求和：

$$\begin{aligned}\mathrm{P}(S = m) &= \sum_{j=0}^{\infty}\mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\cdot\frac{(\lambda q)^j}{j!} = \mathrm{e}^{-\lambda}\,\frac{(\lambda p)^m}{m!}\,\mathrm{e}^{\lambda q}\\ &= \frac{(\lambda p)^m}{m!}\,\mathrm{e}^{-\lambda p},\end{aligned}$$

这里用到 $\sum_{j\ge 0} x^j/j! = \mathrm{e}^x$ 及 $-\lambda + \lambda q = -\lambda p$。（由此可见，我发现的错字数目本身也服从泊松分布（Poisson distribution），平均值（mean）为 λp。）这个概率是正数，所以可以用它作除数。于是

$$\begin{aligned}\mathrm{P}(U = n \mid S = m) &= \frac{\mathrm{P}(S = m,\ U = n)}{\mathrm{P}(S = m)} = \frac{\mathrm{e}^{-\lambda}\,\dfrac{(\lambda p)^m}{m!}\cdot\dfrac{(\lambda q)^n}{n!}}{\mathrm{e}^{-\lambda p}\,\dfrac{(\lambda p)^m}{m!}}\\ &= \frac{(\lambda q)^n}{n!}\,\mathrm{e}^{-\lambda+\lambda p} = \frac{(1-p)^n\lambda^n}{n!}\,\mathrm{e}^{-(1-p)\lambda},\end{aligned}$$

这正是所要证明的结果。

答案完全不依赖 m：知道我发现了多少个错字，并不能告诉我看漏了多少个。无论 m 是多少，看漏的数目都服从平均值为 $(1-p)\lambda$ 的泊松分布。

验算

- 把最后的答案对 $n = 0, 1, 2, \ldots$ 求和，得 $\mathrm{e}^{(1-p)\lambda}\mathrm{e}^{-(1-p)\lambda} = 1$，条件分布理应如此。
- 当 $p \to 1$ 时，我会发现所有错字，而答案在 $n = 0$ 时趋向 1，在 $n \ge 1$ 时趋向 0。吻合。
- 取 $\lambda = 3$、$p = 0.6$、$m = 2$、$n = 1$，公式给出 $1.2\,\mathrm{e}^{-1.2} \approx 0.3614$。模拟 400 000 份报纸，只保留恰好发现两个错字的那些，所得结果为 0.360。

常见错误

- 在第 (i) 部分漏掉二项式系数（binomial coefficient）$\binom{m+n}{m}$：被发现的 m 个错字可以是 $m+n$ 个之中的任意 m 个。
- 把 $\mathrm{P}(S=m)$ 当成 $\mathrm{P}(N=m)$，或当成第 (i) 部分的答案，而没有对所有可能的看漏数目求和。

STEP 之外

这就是泊松分布的稀疏化（thinning，又称分裂 splitting）性质：如果有 $\mathrm{Poisson}(\lambda)$ 个物件，每个各以概率 p 被保留，那么被保留的数目与被舍弃的数目是独立的泊松变量，平均值分别为 λp 和 $\lambda(1-p)$。我们求得的联合概率恰好分解成这两个部分。稀疏化是泊松过程（Poisson process）理论中的重要工具，它让你把一串随机事件分拆成若干串互相独立的子事件流。Grimmett 与 Welsh 合著的 *Probability: An Introduction*（概率论入门教材）有讲述这个性质，也介绍了只需一行便能证明它的母函数（generating function）方法。

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

1992 年详细解答. 全书各章及
所有详细解答均可免费阅读、下载及分享，网址为
beyondstep.keithmcnulty.org

文字内容已贡献予公有领域（CC0 1.0）。STEP 试题 © Cambridge University Press & Assessment。

