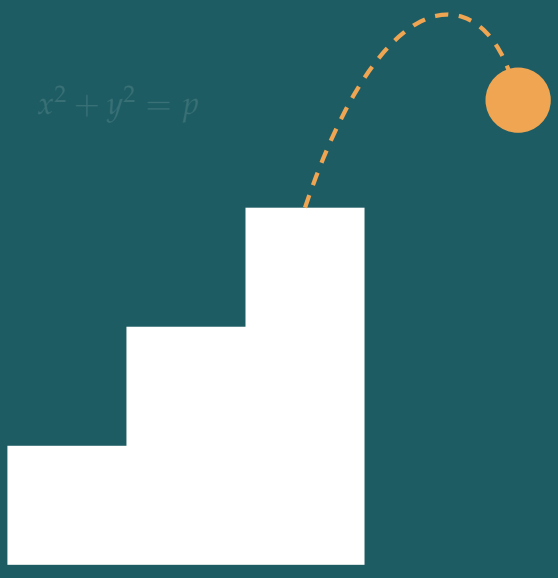
STEP 備試 · 大學數學

Beyond STEP

STEP I, II 及 III

2004 年詳細解答



KEITH McNULTY

Beyond STEP

以歷屆 STEP 試題為門
走進大學數學

STEP I, II 及 III

2004 年詳細解答

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

2004 年詳細解答

2004 年 STEP 試卷（STEP I, II 及 III）全部 42 題的詳細解答。自樣本試卷以來每一道 STEP 題目的詳細解答、*Beyond STEP* 全書以及各章的單行本，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

閱讀解答之前，請先認真嘗試解題。每個答案均已用數值方法核對，但解答數量如此之多，難免會有疏漏：如果發現有可疑之處，請相信自己的演算，並仔細檢查。

由機器從英文翻譯而成；尚未經審閱。

2026 年 11 月首次於網上發表。

本 PDF 內凡屬作者本人創作的內容，包括正文、解答、插圖、練習及答案，均按照 Creative Commons CC0 1.0 Universal 授權條款貢獻予公有領域：你可以為任何目的複製、改編及分享，毋須徵求許可。詳見 `creativecommons.org/publicdomain/zero/1.0`。

本書討論的 STEP 試題，版權屬 Cambridge University Press & Assessment 所有。本書並無轉載這些試題：每一題均附連結，通往官方來源的題目，即 STEP Support Programme 的試題資料庫或 OCR 的歷屆試卷。

如有錯漏，歡迎到 github.com/keithmcnulty/step_book_mcnulty 提交 issue 指正。

目錄

STEP I

第 1 題 根式與立方根 純數學 . . . 3

第 2 題 下取整函數的積分 純數學 . . . 6

第 3 題 二元多項式的因式分解 純數學 . . . 9

第 4 題 式換積分法 純數學 . . . 12

第 5 題 模 5 的剩餘 純數學 . . . 16

第 6 題 三角形的向量幾何 純數學 . . . 20

第 7 題 生成函數 純數學 . . . 24

第 8 題 單調數列與數學歸納法 純數學 . . . 27

第 9 題 拋體運動 力學 . . . 31

第 10 題 加速運動學 力學 . . . 35

第 11 題 摩擦力與鉸接梯子 力學 . . . 39

第 12 題 葉斯定理 概率與統計 . . . 43

第 13 題 勻分佈變量的最大值 概率與統計 . . . 46

第 14 題 排列的計數 概率與統計 . . . 50

STEP II

第 1 題 含根式的方程 純數學 . . . 53

第 2 題 二次不等式 純數學 . . . 56

第 3 題 曲線描繪 純數學 . . . 60

第 4 題 幾何中的最優化 純數學 . . . 65

第 5 題 積分方程 純數學 . . . 69

第 6 題 向量分解 純數學 . . . 72

第 7 題 二分法 純數學 . . . 75

第 8 題 數分方程的比較 純數學 . . . 79

第 9 題 重心 力學 83
第 10 題 斜面上的摩擦力 力學 87
第 11 題 振動数と抵抗力 力學 91
第 12 題 連続分布 概率與統計 95
第 13 題 最優停止 概率與統計 99
第 14 題 鴿巢原理 概率與統計 103

STEP III

第 1 題 二重積分 純數學 107
第 2 題 漸近線與曲線描繪 純數學 111
第 3 題 凸性與級數的界限 純數學 115
第 4 題 相切圓與級數 純數學 119
第 5 題 三角方程 純數學 123
第 6 題 遞推關係 純數學 127
第 7 題 歸約公式 純數學 131
第 8 題 微分方程 純數學 134
第 9 題 力矩與摩擦力 力學 138
第 10 題 彈性繩與簡諧運動 力學 142
第 11 題 體積排斥 力學 146
第 12 題 指示變量與錯排 概率與統計 150
第 13 題 幾何分布 概率與統計 154
第 14 題 假設檢驗的錯誤 概率與統計 157

詳細解答

STEP I 2004，第 1 題

純數學 • 根式與立方根

> **STEP I 2004，第 1 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q1

關鍵思路

形如 $a+b\sqrt{n}$（其中 a、b 為整數）的數，本身就像一個小小的數系：兩個這樣的數相乘，結果仍是同一形式；而兩個這樣的數相等，當且僅當它們的有理部分（rational part）相同，且 $\sqrt{n}$ 部分亦相同。第 (i) 部分是熱身，把這樣的一個數立方；第 (ii) 部分則透過比較各部分，把這個過程倒過來做。第 (iii) 部分是一條關於 x^3 的二次方程，而它的根原來正是第 (ii) 部分求得的立方。

解答

第 (i) 部分

由二項式定理（binomial theorem），並利用 $(2\sqrt{5})^2 = 20$ 及 $(2\sqrt{5})^3 = 40\sqrt{5}$，

$$\begin{aligned}\left(3+2\sqrt{5}\right)^3 &= 3^3 + 3\cdot 3^2\cdot 2\sqrt{5} + 3\cdot 3\cdot 20 + 40\sqrt{5}\\ &= 27 + 54\sqrt{5} + 180 + 40\sqrt{5}\\ &= 207 + 94\sqrt{5}.\end{aligned}$$

所以 $a = 207$，$b = 94$。

第 (ii) 部分

把 $\sqrt{5}$ 換成 $\sqrt{2}$，同樣展開，對任意 c 和 d 可得

$$\begin{aligned}\left(c-d\sqrt{2}\right)^3 &= c^3 - 3c^2d\sqrt{2} + 3c\cdot 2d^2 - 2d^3\sqrt{2}\\ &= \left(c^3+6cd^2\right) - \left(3c^2d+2d^3\right)\sqrt{2}.\end{aligned}$$

由於實立方根（real cube root）是唯一的，我們需要 $(c-d\sqrt{2})^3 = 99-70\sqrt{2}$。

想一想 為甚麼可以把有理部分和 $\sqrt{2}$ 部分分開比較？

若 $p+q\sqrt{2} = r+s\sqrt{2}$，其中 p,q,r,s 為整數且 $q \neq s$，則 $\sqrt{2} = (p-r)/(s-q)$ 便是有理數，但事實並非如此。所以 $q = s$，從而 $p = r$。因此我們需要

$$c^3 + 6cd^2 = 99, \qquad 3c^2d + 2d^3 = 70.$$

由於 c 和 d 均為正數，第一條方程迫使 $c^3 < 99$，故 $c \leq 4$。把每個值逐一代入 $6cd^2 = 99 - c^3$：

- $c = 1$：$6d^2 = 98$，沒有整數解；
- $c = 2$：$12d^2 = 91$，沒有；
- $c = 3$：$18d^2 = 72$，故 $d = 2$；
- $c = 4$：$24d^2 = 35$，沒有。

唯一的可能是 $c = 3$，$d = 2$，而它也滿足第二條方程：$3\cdot 9\cdot 2 + 2\cdot 8 = 54 + 16 = 70$。因此

$$\sqrt[3]{99-70\sqrt{2}} = 3-2\sqrt{2}, \qquad c = 3,\ d = 2.$$

（順帶檢查一下正負號：$2\sqrt{2} \approx 2.83 < 3$，所以 $3-2\sqrt{2}$ 是正數，這與 $99-70\sqrt{2} \approx 0.005$ 是正數相符。）

第 (iii) 部分

設 $y = x^3$。方程變成 $y^2 - 198y + 1 = 0$，所以

$$y = \frac{198 \pm \sqrt{198^2-4}}{2} = 99 \pm \sqrt{99^2-1} = 99 \pm \sqrt{9800} = 99 \pm 70\sqrt{2},$$

當中用到 $9800 = 4900\cdot 2$。兩個值都是正實數。

由第 (ii) 部分，$99-70\sqrt{2} = (3-2\sqrt{2})^3$。至於另一個根，留意該二次方程兩根之積為 1，所以

$$99+70\sqrt{2} = \frac{1}{99-70\sqrt{2}} = \frac{1}{(3-2\sqrt{2})^3} = \left(3+2\sqrt{2}\right)^3,$$

因為 $(3-2\sqrt{2})(3+2\sqrt{2}) = 9-8 = 1$。（同樣地，在第 (ii) 部分的展開式中把 d 的正負號改變，亦可直接得出 $(3+2\sqrt{2})^3 = 99+70\sqrt{2}$。）

每個實數恰好有一個實立方根，所以每個 y 值恰好給出一個實數 x。兩個實數解是

$$x = 3-2\sqrt{2} \qquad \text{和} \qquad x = 3+2\sqrt{2}.$$

這條六次方程（sextic）其餘四個解都是複數：即每個 y 餘下的立方根。

驗算

- 以數值計算，$(3+2\sqrt{5})^3 \approx 417.19$，而 $207+94\sqrt{5} \approx 417.19$。
- $3+2\sqrt{2} \approx 5.8284$，而 $x^6 - 198x^3 + 1$ 在此值處等於零（在捨入誤差範圍內）；在 $3-2\sqrt{2} \approx 0.1716$ 處亦然。
- 第 (iii) 部分兩個答案之積為 1，理應如此：若 x 是根，則 $1/x$ 也是根，因為把方程除以 x^6，便得到關於 $1/x$ 的同一條方程。

常見錯誤

- 展開時忘記 $(d\sqrt{2})^3 = 2\sqrt{2}\,d^3$，而不是 $\sqrt{2}\,d^3$。
- 由第一條方程求得 $c=3$，$d=2$，卻沒有驗證第二條方程。
- 在第 (iii) 部分給出六個「解」：每個正數只有一個立方根是實數。

STEP 之外

所有形如 $a+b\sqrt{2}$（a,b 為整數）的數組成的集合是一個環（ring），記作 $\mathbb{Z}[\sqrt{2}]$，而 $3+2\sqrt{2}$ 是其中的一個單位（unit）：它在同一個環內有逆元（inverse）$3-2\sqrt{2}$。這類單位與佩爾方程（Pell's equation）$a^2-2b^2=\pm1$ 息息相關（留意 $99^2-2\cdot70^2=1$），而 $\mathbb{Z}[\sqrt{2}]$ 的每個單位都是 $\pm(1+\sqrt{2})^n$。想進一步了解，可參閱 Davenport 的 *The Higher Arithmetic*，或 Allenby 的 *Rings, Fields and Groups*。

詳細解答

STEP I 2004，第 2 題

純數學 • 下取整函數的積分

> **STEP I 2004，第 2 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q2

關鍵思路

函數 $[x]$ 在每個區間 $r \le x < r+1$ 上都是常數，所以凡是由它構成的函數都是階梯函數（step function）。對階梯函數求積分，不過是把一個個長方形的面積加起來：在各整數處把 $[0, a]$ 分段，在每一段單位長度的區間上，被積函數（integrand）都是常數。當 a 不是整數時，唯一的新情況是最後多了一個較窄的長方形。

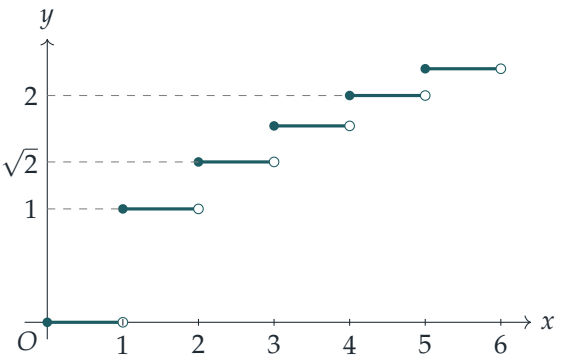
解答

第 (i) 部分

對於 $x \ge 0$，設 $r = [x]$。那麼 r 是非負整數，且 $r \le x < r+1$，所以在該區間上

$$y = \sqrt{[x]} = \sqrt{r}.$$

因此圖像呈階梯狀：在 $[0,1)$ 上 $y = 0$，在 $[1,2)$ 上 $y = 1$，在 $[2,3)$ 上 $y = \sqrt{2}$，如此類推。每一級都包含左端點（實心點），但不包含右端點（空心點），因為 $[x]$ 恰好在每個整數處向上跳升。由於 $\sqrt{r+1} - \sqrt{r} = 1/(\sqrt{r+1} + \sqrt{r})$ 遞減，各級升高的幅度愈來愈小。對於 $x < 0$，我們有 $[x] \le -1$，所以 $\sqrt{[x]}$ 在該處沒有定義，圖像只在 $x \ge 0$ 的範圍內出現。



現設 a 為正整數。在整數 $1, 2, \ldots, a-1$ 處把積分範圍分段：

$$\int_0^a \sqrt{[x]}\,\mathrm{d}x = \sum_{r=0}^{a-1} \int_r^{r+1} \sqrt{[x]}\,\mathrm{d}x.$$

在 $r \le x < r+1$ 上，被積函數等於常數 $\sqrt{r}$。（在 $x = r+1$ 這一點上它取另一個值，但改變函數在一點上的值並不會改變其積分。）所以每一段都是一個闊度為 1、高度為 $\sqrt{r}$ 的長方形的面積：

$$\int_r^{r+1} \sqrt{[x]}\,\mathrm{d}x = \sqrt{r}, \qquad \text{因此} \qquad \int_0^a \sqrt{[x]}\,\mathrm{d}x = \sum_{r=0}^{a-1} \sqrt{r},$$

這正是所要證明的結果。

第 (ii) 部分

完全相同的論證也適用於 $2^{[x]}$，它在 $r \le x < r+1$ 上等於 2^r。對於正整數 a，

$$\int_0^a 2^{[x]}\,\mathrm{d}x = \sum_{r=0}^{a-1} \int_r^{r+1} 2^r\,\mathrm{d}x = \sum_{r=0}^{a-1} 2^r = 1 + 2 + 4 + \cdots + 2^{a-1}.$$

這是一個等比級數（geometric series），首項為 1，公比為 2，共有 a 項，所以其和為

$$\frac{2^a - 1}{2 - 1} = 2^a - 1,$$

這就是所求的結果。

第 (iii) 部分

現設 $a > 0$ 不是整數，並令 $n = [a]$，那麼 n 是整數，且 $n < a < n+1$ 及 $n \ge 0$。

? **想一想** 積分到了哪裏，就不再是若干個完整單位長方形之和？

直到 $x = n$ 為止，一切都與第 (ii) 部分相同；餘下的一段是 $n \le x \le a$，這是一條闊度為 $a - n$ 的窄條，完全位於同一級之內，其上 $[x] = n$。（若 $n = 0$，第一段是空的，其貢獻為 $2^0 - 1 = 0$，所以同一條公式仍然適用。）因此

$$\begin{aligned}\int_0^a 2^{[x]}\,\mathrm{d}x &= \int_0^n 2^{[x]}\,\mathrm{d}x + \int_n^a 2^n\,\mathrm{d}x \\ &= \left(2^n - 1\right) + (a - n)\,2^n \\ &= 2^{[a]}\left(1 + a - [a]\right) - 1.\end{aligned}$$

驗算

- 若 $0 < a < 1$，則 $[a] = 0$，公式給出 $1 \cdot (1 + a) - 1 = a$：正確，因為被積函數在 $[0, a]$ 上等於 1。
- 當 a 從下方趨近整數 m 時，$[a] = m - 1$，公式趨向 $2^{m-1} \cdot 2 - 1 = 2^m - 1$，與第 (ii) 部分吻合。所以這個積分是 a 的連續函數（continuous function）；任何有界函數（bounded function）的積分，作為其積分上限的函數，都必定如此。
- 以數值驗算，取 $a = 2.5$，公式給出 $4 \times 1.5 - 1 = 5$，而用中點法則（midpoint rule）計算該積分，同樣得出 5.000。至於第 (i) 部分，取 $a = 5$，兩邊都是 $0 + 1 + \sqrt{2} + \sqrt{3} + 2 \approx 6.1463$。

常見錯誤

- 由 $r = 1$ 加到 a，而不是由 $r = 0$ 加到 $a - 1$：在 $[r, r + 1)$ 上，高度是 $\sqrt{r}$，即左端點處的值。
- 在第 (iii) 部分，遺漏了那條不完整的窄條，或把它的闊度當作 1 而不是 $a - [a]$。
- 用垂直線把各級連起來，或在兩端都畫上實心點：下取整函數（floor function）在整數處並不連續。

STEP 之外

把長方形面積相加來求階梯函數的積分，正是大學裏建立黎曼積分（Riemann integral）的方法：先為階梯函數定義積分，然後用階梯函數從上方和下方把一般的函數夾住。第 (i) 部分也是比較求和與積分的第一步，由此可得出諸如 $\sum_{r=1}^{n} \sqrt{r} \approx \frac{2}{3}n^{3/2}$ 的估計。Abbott 的 *Understanding Analysis*（一本實分析入門教科書）就是以這種方式建立黎曼積分的。

詳細解答

STEP I 2004，第 3 題

純數學 • 二元多項式的因式分解

> **STEP I 2004，第 3 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q3

關鍵思路

二元多項式（polynomial in two variables）可以看成只含其中一個變數的多項式，而係數則含有另一個變數。把它看成 x 的多項式時，因式定理（factor theorem）依然適用，長除法（long division）依然適用，而 x 的二次式只要判別式（discriminant）是完全平方（perfect square），就依然可以因式分解（factorise）。在第 (ii) 部分，x 的首項係數（leading coefficient）本身是 y 的一次式，這暗示其中一個因式只含 y。

解答

第 (i) 部分

把式子 $(*)$ 按 x 的冪次整理：

$$P(x) = x^3 + (2y-5)x^2 + (y^2 - 8y + 6)x + (6y - 3y^2).$$

代入 $x = 3$，

$$\begin{aligned} P(3) &= 27 + 9(2y-5) + 3(y^2 - 8y + 6) + 6y - 3y^2 \\ &= (27 - 45 + 18) + (18y - 24y + 6y) + (3y^2 - 3y^2) = 0 \end{aligned}$$

對任何 y 都成立，所以根據因式定理，$x - 3$ 是一個因式，正如題目所求。

要找出另一個因式，設 $P(x) = (x-3)(x^2 + px + q)$，其中 p 和 q 可以與 y 有關。展開右邊得 $x^3 + (p-3)x^2 + (q-3p)x - 3q$，再比較係數（comparing coefficients）：

$$\begin{aligned} x^2: &\quad p - 3 = 2y - 5, && \text{所以} p = 2y - 2, \\ x^0: &\quad -3q = 6y - 3y^2, && \text{所以} q = y^2 - 2y, \end{aligned}$$

而 x 的係數亦吻合：$q-3p=y^2-8y+6$，與應有的結果一致。所以二次因式是

$$x^2+(2y-2)x+y^2-2y=(x^2+2xy+y^2)-2(x+y)=(x+y)(x+y-2).$$

想一想 我們是怎樣看出這個二次式可以因式分解的？

$x^2+2xy+y^2$ 這幾項合起來是完全平方 $(x+y)^2$，而餘下的項 $-2x-2y$ 是 $x+y$ 的倍數。(如果一時看不出這個分組方法，可以計算 x 的判別式：$(2y-2)^2-4(y^2-2y)=4$，是完全平方，所以根是 $x=-y+1\pm1$，即 $x=-y$ 和 $x=2-y$。) 因此

$$(*)=(x-3)(x+y)(x+y-2),$$

所以 $a=1$，$b=0$，$c=1$，$d=-2$（或者把兩個因式對調：$a=1$，$b=-2$，$c=1$，$d=0$）。

第 (ii) 部分

把這個式子稱為 Q，同樣按 x 的冪次整理：

$$Q=(y+2)x^2+(12-4y-5y^2)x+(6y^3-y^2-21y+10).$$

x^2 的係數是 $y+2$。如果 Q 是三個一次因式的積，x^2 項就來自其中兩個因式含 x 的部分，所以很自然會問：$y+2$ 能否整除整個式子？設 $y=-2$：

$$12+8-20=0,\qquad -48-4+42+10=0,$$

所以餘下兩個係數在 $y=-2$ 時都等於零，根據 (關於 y 的) 因式定理，兩者都可被 $y+2$ 整除：

$$12-4y-5y^2=-(5y^2+4y-12)=-(y+2)(5y-6),$$
$$6y^3-y^2-21y+10=(y+2)(6y^2-13y+5).$$

另外 $6y^2-13y+5=(2y-1)(3y-5)$。因此

$$Q=(y+2)\big[x^2-(5y-6)x+(2y-1)(3y-5)\big].$$

想一想 方括號內的式子可以因式分解嗎？我們怎樣判斷？

它關於 x 的判別式是

$$\begin{aligned}(5y-6)^2-4(2y-1)(3y-5)&=25y^2-60y+36-(24y^2-52y+20)\\&=y^2-8y+16=(y-4)^2,\end{aligned}$$

是完全平方，所以可以分解。更快捷的做法是留意 $(2y-1)+(3y-5)=5y-6$，所以積為 $(2y-1)(3y-5)$、和為 $5y-6$ 的兩個數就是 $2y-1$ 和 $3y-5$，由此得根 $x=2y-1$ 和 $x=3y-5$。因此

$$Q=(y+2)(x-2y+1)(x-3y+5).$$

驗算

- 在第 (i) 部分代入 $x = y = 1$：$(*)$ 給出 $1-5+2+1-8-3+6+6 = 0$，而 $(1-3)(2)(0) = 0$。取 $x = 2$，$y = 1$：$(*)$ 給出 $8-20+8+2-16-3+12+6 = -3$，而 $(-1)(3)(1) = -3$。
- 在第 (ii) 部分代入 $x = y = 0$：$Q = 10$，而 $(2)(1)(5) = 10$。取 $x = 1$，$y = 1$：$Q = 6-1-21+2+12-4+1-5+10 = 0$，而的確 $x - 2y + 1 = 0$。
- 我們在 2000 個隨機點 (x, y) 上以精確的有理數運算核對了兩個因式分解，每次結果都一致。

常見錯誤

- 把式子看成 x 的多項式，卻忘記「常數」項是指所有不含 x 的項，在第 (i) 部分即 $6y - 3y^2$。
- 在第 (ii) 部分，嘗試找一個根 $x = k$，其中 k 是一個數：形如 $x - k$ 的因式並不存在。應該先找出的，是不含 x 的因式 $y + 2$。
- 把 $-(5y^2 + 4y - 12)$ 除以 $y + 2$ 時出現正負號錯誤；應展開驗算。

STEP 之外

多元多項式構成一個環（ring），而 $\mathbb{Z}[x, y]$ 可以看成 $(\mathbb{Z}[y])[x]$：即係數是 y 的多項式的 x 的多項式。這正是我們剛才的做法，亦是因式定理和除法依然適用的原因。這些環具有唯一分解性質（unique factorisation）（高斯引理，Gauss's lemma），所以我們找到的因式分解，除了次序和正負號之外是唯一的。Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）詳細闡述了這些概念。

詳細解答

STEP I 2004，第 4 題

純數學 • 代換積分法

> **STEP I 2004，第 4 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q4

關鍵思路

每當見到 $\sqrt{x^2-1}$，就應該想起恆等式 (identity) $\sec^2 t - 1 = \tan^2 t$：設 $x = \sec t$，平方根便化為 $\tan t$，而由於 $\mathrm{d}x = \sec t \tan t \, \mathrm{d}t$，這個 $\tan t$ 亦會一併約去。第 (ii) 和第 (iii) 部分看似不同，但只要對根號內的式子配方（completing the square），兩者都會化成同一個形狀 $\sqrt{u^2 - a^2}$，其中 $u = x + 2$，而根號外的因子 $x + 2$ 正是這個代換（substitution）所需要的。

解答

導數

把 $\sec t = (\cos t)^{-1}$ 寫出，再用鏈式法則（chain rule），得

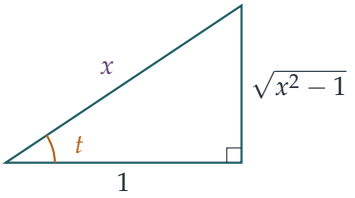
$$\frac{\mathrm{d}}{\mathrm{d}t} \sec t = -(\cos t)^{-2} \cdot (-\sin t) = \frac{\sin t}{\cos^2 t} = \sec t \tan t.$$

第 (i) 部分

設 $x = \sec t$，於是 $\mathrm{d}x = \sec t \tan t \, \mathrm{d}t$。當 $x = \sqrt{2}$ 時，$\cos t = 1/\sqrt{2}$，$t = \frac{\pi}{4}$；當 $x = 2$ 時，$\cos t = \frac{1}{2}$，$t = \frac{\pi}{3}$。對於 $[\frac{\pi}{4}, \frac{\pi}{3}]$ 內的 t，我們有 $\tan t > 0$，所以

$$\sqrt{x^2 - 1} = \sqrt{\sec^2 t - 1} = \sqrt{\tan^2 t} = \tan t$$

（無須加上絕對值符號）。下面的三角形顯示的是同一回事：斜邊為 x，鄰邊為 1，所以對邊是 $\sqrt{x^2 - 1}$。



積分變成

$$\begin{aligned}\int_{\sqrt{2}}^{2} \frac{1}{x^3\sqrt{x^2-1}}\,\mathrm{d}x &= \int_{\pi/4}^{\pi/3} \frac{\sec t\tan t}{\sec^3 t\,\tan t}\,\mathrm{d}t = \int_{\pi/4}^{\pi/3} \cos^2 t\,\mathrm{d}t\\ &= \int_{\pi/4}^{\pi/3} \tfrac{1}{2}(1+\cos 2t)\,\mathrm{d}t = \left[\tfrac{1}{2}t + \tfrac{1}{4}\sin 2t\right]_{\pi/4}^{\pi/3}\\ &= \left(\frac{\pi}{6}+\frac{\sqrt{3}}{8}\right) - \left(\frac{\pi}{8}+\frac{1}{4}\right) = \frac{\sqrt{3}-2}{8} + \frac{\pi}{24},\end{aligned}$$

其中用到 $\cos^2 t = \frac{1}{2}(1+\cos 2t)$、$\sin\frac{2\pi}{3} = \frac{\sqrt{3}}{2}$ 和 $\sin\frac{\pi}{2} = 1$。這就是所求的結果。

第 (ii) 部分

想一想 第 (ii) 部分裏並沒有 $\sqrt{x^2-1}$。第 (i) 部分可以怎樣幫得上忙？

把根號內的式子展開，然後配方：

$$(x+1)(x+3) = x^2+4x+3 = (x+2)^2 - 1.$$

所以設 $u = x+2$（而 $\mathrm{d}u = \mathrm{d}x$），積分便是

$$\int \frac{1}{u\sqrt{u^2-1}}\,\mathrm{d}u.$$

被積函數（integrand）只在 $u^2 > 1$ 時才有定義。先考慮 $u > 1$（即 $x > -1$），並設 $u = \sec t$，其中 $0 < t < \frac{\pi}{2}$，使得 $\tan t > 0$ 及 $\sqrt{u^2-1} = \tan t$。於是

$$\int \frac{\sec t\tan t}{\sec t\,\tan t}\,\mathrm{d}t = \int 1\,\mathrm{d}t = t + C,$$

而由於 $\sec t = u$，我們有 $t = \operatorname{arcsec} u$（等價地，即 $\arccos(1/u)$）。因此，對於 $x > -1$，

$$\int \frac{1}{(x+2)\sqrt{(x+1)(x+3)}}\,\mathrm{d}x = \operatorname{arcsec}(x+2) + C.$$

為求完整，被積函數在 $x < -3$ 時亦有定義，此時 $u < -1$。在這個情況下，設 $u = -v$，其中 $v > 1$：被積式 $\dfrac{1}{u\sqrt{u^2-1}}\,\mathrm{d}u$ 變成 $\dfrac{1}{(-v)\sqrt{v^2-1}}(-\mathrm{d}v) = \dfrac{1}{v\sqrt{v^2-1}}\,\mathrm{d}v$，其積分為 $\operatorname{arcsec} v$。兩種情況都可以用

$$\operatorname{arcsec}|x+2| + C = \arccos\frac{1}{|x+2|} + C,$$

來概括，而兩個區間上的常數可以不同。

第 (iii) 部分

再次配方：

$$x^2+4x-5=(x+2)^2-9,$$

所以設 $u=x+2$ 後，我們需要求 $\displaystyle\int \frac{1}{u\sqrt{u^2-9}}\,\mathrm{d}u$。這一次自然的代換是 $u=3\sec t$，這樣選是為了令 $u^2-9=9\tan^2 t$。對於 $u>3$（即 $x>1$），取 $0<t<\frac{\pi}{2}$，於是 $\sqrt{u^2-9}=3\tan t$，而 $\mathrm{d}u=3\sec t\tan t\,\mathrm{d}t$。於是

$$\int \frac{3\sec t\tan t}{3\sec t\cdot 3\tan t}\,\mathrm{d}t=\int \frac{1}{3}\,\mathrm{d}t=\frac{t}{3}+C,$$

而 $\sec t=u/3$。因此，對於 $x>1$，

$$\int \frac{1}{(x+2)\sqrt{x^2+4x-5}}\,\mathrm{d}x=\frac{1}{3}\operatorname{arcsec}\frac{x+2}{3}+C.$$

與第 (ii) 部分完全一樣，$x<-5$ 的情況（此時 $u<-3$）可用 $u=-v$ 處理，而兩個區間都可以用 $\frac{1}{3}\operatorname{arcsec}\left(\frac{1}{3}|x+2|\right)+C$ 來概括。

驗算

- 對第 (i) 部分的積分使用辛普森法則（Simpson's rule），得 $0.0974060\ldots$，與 $\frac{1}{8}(\sqrt{3}-2)+\frac{\pi}{24}$ 吻合。
- 對答案求導：當 $u>1$ 時，$\dfrac{\mathrm{d}}{\mathrm{d}u}\operatorname{arcsec}u=\dfrac{1}{u\sqrt{u^2-1}}$（設 $u=\sec t$，並使用開首求得的導數）。例如，第 (ii) 部分的被積函數的 $\int_0^3$ 數值上是 0.322241，而 $\operatorname{arcsec}5-\operatorname{arcsec}2=\arccos\frac{1}{5}-\frac{\pi}{3}$ 與之相符。
- 第 (iii) 部分應該是第 (ii) 部分經過縮放後的版本：設 $u=3w$，則 $\int\frac{\mathrm{d}u}{u\sqrt{u^2-9}}=\frac{1}{3}\int\frac{\mathrm{d}w}{w\sqrt{w^2-1}}$，這就解釋了因子 $\frac{1}{3}$ 的由來。

常見錯誤

- 在第 (i) 部分忘記更改積分上下限，或把上下限弄反：$\sec t=\sqrt{2}$ 給出 $t=\frac{\pi}{4}$，而不是 $\frac{\pi}{3}$。
- 寫下 $\sqrt{\tan^2 t}=\tan t$，卻沒有檢查 $\tan t$ 在所用範圍內的正負號。
- 在第 (iii) 部分代入 $u=\sec t$ 而不是 $u=3\sec t$，結果剩下 $\sqrt{\sec^2 t-9}$，沒有恆等式可以把它化簡。

STEP 之外

反正割函數（inverse secant）只是一個家族的其中一員：代換 $x = a\sin t$, $a\tan t$, $a\sec t$（或它們的雙曲函數（hyperbolic function）近親 $a\sinh t$, $a\cosh t$）可以馴服 $\sqrt{a^2 - x^2}$, $\sqrt{a^2 + x^2}$ 和 $\sqrt{x^2 - a^2}$，而配方可以把任何 $\sqrt{\text{二次式}}$ 化為其中一種。在正負號上，以及在原函數（antiderivative）定義於哪個區間上所需要的謹慎，正是初嘗分析（analysis）課程如何令積分變得嚴謹的滋味。Bartle 與 Sherbert 的 *Introduction to Real Analysis*（實分析入門）嚴格地處理了代換定理（substitution theorem），而 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題集）則有更多同類風格的 STEP 積分題。

詳細解答

STEP I 2004，第 5 題

純數學 • 模 5 的剩餘

> **STEP I 2004，第 5 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q5

關鍵思路

這五個等差數列（arithmetic progression）把正整數按除以 5 的餘數（remainder）分成五類。只要為每一類寫出像 $5n-3$ 這樣的通項（general term），就可以把和與平方展開，再抽出因子 5，從而追蹤它們落在哪一類。這樣便能完全掌握平方數可以落在哪些類別，而兩個「不可能」的證明都由此而來：等式右邊落在一個左邊永遠到不了的類別。

解答

通項

每個數列的公差（common difference）都是 5，所以對 $n=1,2,3,\ldots$，第 n 項分別是

$$A:\ 5n-4, \qquad B:\ 5n-3, \qquad C:\ 5n-2, \qquad D:\ 5n-1, \qquad E:\ 5n.$$

（代入 $n=1$，可驗證每條式子都給出題目所列的首項。）因此，一個正整數屬於 A、B、C、D 或 E，當且僅當它除以 5 的餘數分別是 1、2、3、4 或 0。

B 的一項與 C 的一項之和

任取 B 的一項 $5m-3$ 和 C 的一項 $5n-2$，其中 $m,n\geq 1$。那麼

$$(5m-3)+(5n-2)=5m+5n-5=5(m+n-1).$$

由於 $m+n-1\geq 1$ 是正整數，這就是 E 的第 $(m+n-1)$ 項，正如所求。

B 和 C 各項的平方

對 B 的一項 $5m-3$，

$$\begin{aligned}(5m-3)^2 &= 25m^2-30m+9\\ &= 5(5m^2-6m+2)-1.\end{aligned}$$

這裏 $k=5m^2-6m+2$ 是整數，而且對所有 $m\geq 1$ 都有 $k=5m^2-6m+2=m(5m-6)+2\geq 1$（當 $m=1$ 時它等於 1；當 $m\geq 2$ 時 m 和 $5m-6$ 都是正數）。所以 $(5m-3)^2=5k-1$ 是 D 的第 k 項。B 中每一項的平方都屬於 D。

關於 C 的類似命題是：*C 中每一項的平方也是 D 的一項*。事實上，對 C 的一項 $5m-2$，

$$\begin{aligned}(5m-2)^2 &= 25m^2-20m+4\\ &= 5(5m^2-4m+1)-1,\end{aligned}$$

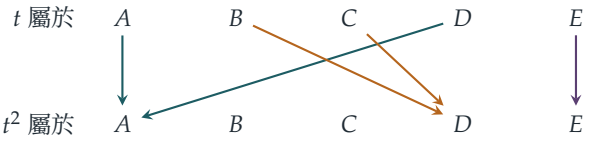
而且當 $m\geq 1$ 時 $k=5m^2-4m+1=m(5m-4)+1\geq 2$。所以 $(5m-2)^2$ 是 D 的第 k 項。

想一想 A、D 和 E 各項的平方又會落在哪裏？

第 (i) 和第 (ii) 部分都要用到這個結果，所以值得用同樣的方法處理餘下三類：

$$\begin{aligned}(5m-4)^2 &= 25m^2-40m+16 = 5(5m^2-8m+4)-4 && \text{（屬於 } A\text{）,}\\ (5m-1)^2 &= 25m^2-10m+1 = 5(5m^2-2m+1)-4 && \text{（屬於 } A\text{）,}\\ (5m)^2 &= 5(5m^2) && \text{（屬於 } E\text{）.}\end{aligned}$$

當 $m\geq 1$ 時，括號內的都是正整數（例如 $5m^2-8m+4=m(5m-8)+4\geq 1$，以及 $5m^2-2m+1\geq 4$）。所以平方運算把各類對應如下：



特別要留意，*沒有完全平方數（perfect square）屬於 B 或 C*：每個正的平方數都屬於 A、D 或 E。

第 (i) 部分

假設正整數 x 和 y 滿足 $x^2+5y=243\,723$。由於 $243\,723=5\times 48\,744+3$，等式右邊是 C 的一項 $5\times 48\,745-2$。於是

$$x^2=243\,723-5y=5(48\,745-y)-2.$$

這個數是正數（它等於 x^2），所以 $48\,745 - y$ 是正整數，而 x^2 是 C 的一項。但我們已證明沒有平方數屬於 C。這個矛盾（contradiction）說明這樣的正整數 x 和 y 並不存在。

第 (ii) 部分

四次方（fourth power）就是平方的平方。平方數 x^2 屬於 A、D 或 E；把 A 或 D 的一項平方，得到 A 的一項，而把 E 的一項平方，則得到 E 的一項。所以對任何正整數 x，

$$x^4 \text{ 屬於} A \text{ 或} E, \quad \text{也就是說，} \quad x^4 \text{ 的餘數是1 或0，這裏的除數是5.}$$

y^4 也是如此，所以 $2y^4$ 的餘數是 2 或 0。兩者相加，$x^4 + 2y^4$ 除以 5 的餘數等於以下其中一個數的餘數：

$$1 + 2 = 3, \qquad 1 + 0 = 1, \qquad 0 + 2 = 2, \qquad 0 + 0 = 0,$$

所以 $x^4 + 2y^4$ 屬於 C、A、B 或 E，而永遠不會屬於 D。

想一想 為甚麼可以直接把餘數相加？

如果 $x^4 = 5a + r$ 而 $2y^4 = 5b + s$，那麼 $x^4 + 2y^4 = 5(a + b) + (r + s)$，所以這個和與 $r + s$ 相差 5 的倍數，與 $r + s$ 屬於同一類。這正是上面計算 $B + C$ 時所做的運算，只是寫成了一般形式。

現在 $26\,081\,974 = 5 \times 5\,216\,395 - 1$ 的末位數字是 4，所以它是 D 的一項。因此，對任何正整數 x 和 y，它都不可能等於 $x^4 + 2y^4$，正如所求。

驗算

- 小數值的例子：$2^2 = 4$、$7^2 = 49$ 和 $12^2 = 144$ 的末位數字都是 4 或 9，所以屬於 D；$3^2 = 9$、$8^2 = 64$、$13^2 = 169$ 亦然。
- 四次方：$1, 16, 81, 256, 625, 1296$ 的末位數字都是 1、6 或 5，所以屬於 A 或 E。
- 用電腦搜尋所有 $x \le \sqrt{243\,723} \approx 493.7$，找不到任何 x 使 $243\,723 - x^2$ 是 5 的正倍數；再搜尋所有滿足 $x^4 < 26\,081\,974$ 的 x（即 $x \le 71$），也找不到任何 y 使 $2y^4 = 26\,081\,974 - x^4$。

常見錯誤

- 把 E 的通項寫成 $5n + 5$、把 A 的通項寫成 $5n + 1$ 並無不可，但這時 n 要由 0 開始；把兩種寫法混在一起用，項數就會差了一項。

- 在第 (i) 部分，忘記利用 $x^2 > 0$ 來說明 $48\,745 - y$ 是正整數；少了這一步，x^2 就不一定是 C 的一項，因為這些數列只包含正整數。
- 在第 (ii) 部分，只處理 x^4，卻忘記因子 2 會把 y^4 的餘數由 1 變成 2。

STEP 之外

這五個數列就是模 5 的剩餘類（residue classes），而整道題目其實是 $\mathbb{Z}/5\mathbb{Z}$ 中的算術：那裏的平方數是 $0, 1, 4$（即二次剩餘（quadratic residues)），四次方則是 $0, 1$，這正是費馬小定理（Fermat's little theorem）的一個例子：當 $5 \nmid a$ 時 $a^4 \equiv 1 \pmod 5$。要證明一條丟番圖方程（Diophantine equation）無解，首選的工具就是選一個合適的數，把方程對它取模。Davenport 的 *The Higher Arithmetic* 對同餘（congruences）和二次剩餘有十分精彩的闡述。

詳細解答

STEP I 2004，第 6 題

純數學 • 三角形的向量幾何

> **STEP I 2004，第 6 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q6

關鍵思路

以位置向量（position vector）代替分開處理的 x 坐標和 y 坐標：設 $\mathbf{a} = (p_1, q_1)$、$\mathbf{b} = (p_2, q_2)$、$\mathbf{c} = (p_3, q_3)$，則 H 的位置向量為 $\mathbf{a} + \mathbf{b} + \mathbf{c}$。這樣，題目中每一條直線都有簡短的向量描述，而「互相垂直」就變成「純量積（scalar product）為零」。出現的純量積是一個平方差（difference of two squares）：$(\mathbf{c} + \mathbf{b}) \cdot (\mathbf{c} - \mathbf{b})$，而題目的後半部分就全在於此。

解答

我們假設 A、B、C 不共線（collinear），因而構成一個真正的三角形（否則第一部分的直線可能重合，便沒有唯一的交點）。

中線相交

設 L、M、N 分別為 BC、CA、AB 的中點（midpoint），其位置向量為

$$\mathbf{l} = \tfrac{1}{2}(\mathbf{b} + \mathbf{c}), \qquad \mathbf{m} = \tfrac{1}{2}(\mathbf{a} + \mathbf{c}), \qquad \mathbf{n} = \tfrac{1}{2}(\mathbf{a} + \mathbf{b}).$$

直線 AL 上的一般點為

$$\mathbf{a} + s(\mathbf{l} - \mathbf{a}) = (1 - s)\,\mathbf{a} + \tfrac{s}{2}\,\mathbf{b} + \tfrac{s}{2}\,\mathbf{c},$$

而直線 BM 上的一般點為

$$\mathbf{b} + t(\mathbf{m} - \mathbf{b}) = \tfrac{t}{2}\,\mathbf{a} + (1 - t)\,\mathbf{b} + \tfrac{t}{2}\,\mathbf{c}.$$

只要 $\mathbf{a}$、$\mathbf{b}$ 和 $\mathbf{c}$ 的係數對應相等，這兩個表達式便一致：$1 - s = \frac{t}{2}$、$\frac{s}{2} = 1 - t$ 及 $\frac{s}{2} = \frac{t}{2}$。最後一式給出 $s = t$，於是由 $1 - s = \frac{s}{2}$ 得 $s = t = \frac{2}{3}$，這組值同時滿足三條等式。因此

兩條直線都通過點 G，其位置向量為

$$\mathbf{g} = \tfrac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c}), \qquad \text{即} \qquad G = \left(\tfrac{p_1+p_2+p_3}{3}, \tfrac{q_1+q_2+q_3}{3}\right).$$

想一想 比較係數找到了一個公共點。我們怎樣知道它就是唯一的交點？

兩條相異直線最多只有一個交點，所以我們只需確定 AL 和 BM 是不同的直線。假如它們是同一條直線，這條直線便包含 A、B 和 L，因而包含整條直線 BL，亦即直線 BC；那麼 A 就會在 BC 上，與我們的假設矛盾。所以 G 是唯一的交點。

至於第三條中線（median）CN，取點 $\mathbf{c}+u(\mathbf{n}-\mathbf{c})$，其中 $u=\frac{2}{3}$：

$$\mathbf{c}+\tfrac{2}{3}\left(\tfrac{1}{2}\mathbf{a}+\tfrac{1}{2}\mathbf{b}-\mathbf{c}\right) = \tfrac{1}{3}\mathbf{a}+\tfrac{1}{3}\mathbf{b}+\tfrac{1}{3}\mathbf{c} = \mathbf{g}.$$

所以 G 位於連接 C 與 AB 中點的直線上，正如所求。（每次都得到 $\frac{2}{3}$ 這個值，說明 G（即形心（centroid））位於每條中線上由頂點起計三分之二的位置。）

在 H 處的垂直關係

點 H 的位置向量為 $\mathbf{h}=\mathbf{a}+\mathbf{b}+\mathbf{c}$，所以

$$\overrightarrow{AH} = \mathbf{h}-\mathbf{a} = \mathbf{b}+\mathbf{c}, \qquad \overrightarrow{BC} = \mathbf{c}-\mathbf{b}.$$

若直線 AH 與 BC 垂直相交，則這兩個方向向量（direction vector）的純量積為零：

$$0 = (\mathbf{c}+\mathbf{b})\cdot(\mathbf{c}-\mathbf{b}) = \mathbf{c}\cdot\mathbf{c}-\mathbf{b}\cdot\mathbf{b} = (p_3^2+q_3^2)-(p_2^2+q_2^2),$$

其中交叉項 $\mathbf{c}\cdot\mathbf{b}$ 與 $\mathbf{b}\cdot\mathbf{c}$ 互相抵銷，因為純量積符合交換律（commutative）。因此

$$p_2^2+q_2^2 = p_3^2+q_3^2,$$

這就是所求的結果。用文字來說，B 和 C 與原點 O 的距離相等。

以完全相同的方法，$\overrightarrow{BH}=\mathbf{a}+\mathbf{c}$ 而 $\overrightarrow{AC}=\mathbf{c}-\mathbf{a}$，所以若 BH 與 AC 垂直相交，則 $(\mathbf{c}+\mathbf{a})\cdot(\mathbf{c}-\mathbf{a})=0$，由此得

$$p_1^2+q_1^2 = p_3^2+q_3^2.$$

第三條高線

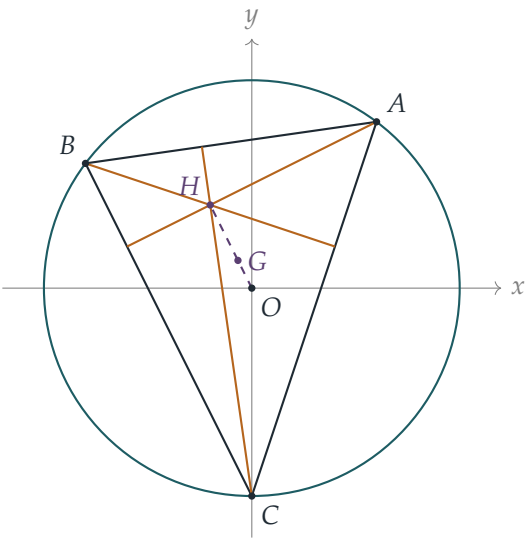
假設兩個垂直條件都成立。結合以上兩個結果，

$$p_1^2+q_1^2 = p_3^2+q_3^2 = p_2^2+q_2^2, \qquad \text{所以} \qquad \mathbf{a}\cdot\mathbf{a} = \mathbf{b}\cdot\mathbf{b}.$$

現在 $\overrightarrow{CH}=\mathbf{a}+\mathbf{b}$ 而 $\overrightarrow{AB}=\mathbf{b}-\mathbf{a}$，並且

$$\overrightarrow{CH}\cdot\overrightarrow{AB} = (\mathbf{b}+\mathbf{a})\cdot(\mathbf{b}-\mathbf{a}) = \mathbf{b}\cdot\mathbf{b}-\mathbf{a}\cdot\mathbf{a} = 0.$$

所以 CH 垂直於 AB（前提是 $H \neq C$，這樣 CH 才是一條真正的直線；若 $H = C$，這個命題便沒有內容）。推論至此完成。



上圖顯示最後一部分的情況：A、B、C 位於以原點 O 為圓心的圓上，三條高線（altitude）（琥珀色）都通過 H，而 O、G、H 共線，且 $\mathbf{h} = 3\mathbf{g}$。

驗算

- 取 $A = (3,4)$、$B = (-4,3)$、$C = (0,-5)$，三點與 O 的距離都是 5。於是 $H = (-1,2)$，而 $\overrightarrow{AH} \cdot \overrightarrow{BC} = (-4)(4) + (-2)(-8) = 0$；另外兩個純量積也是 0。形心為 $G = (-\frac{1}{3}, \frac{2}{3})$，即 H 的三分之一。
- 數值方面，對大約 2000 個頂點為整數點的隨機三角形，以分數精確計算出的兩條中線交點總是 $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$，並且位於第三條中線上；另外，對 1000 個內接於以 O 為圓心的圓的隨機三角形，關於 H 的三個純量積全部為零。

常見錯誤

- 證明了兩條中線都通過 G，卻沒有說明它們為何不可能是同一條直線。只要用一句話指出 A、B、C 不共線，問題便解決了。
- 把 $\overrightarrow{OH} = \mathbf{a}+\mathbf{b}+\mathbf{c}$ 當作 AH 的方向。正確的方向是 $\mathbf{h}-\mathbf{a} = \mathbf{b}+\mathbf{c}$。
- 以為這個結果對所有三角形都成立。題目的假設迫使 O 與 A、B、C 等距；若原點是任意選取的，H 便不是垂心（orthocentre）。

STEP 之外

最後一部分證明了三角形的三條高線共點（concurrent）：把原點放在外心（circumcentre）O（與各頂點等距），則點 $\mathbf{a}+\mathbf{b}+\mathbf{c}$ 同時位於三條高線上。由於 $\mathbf{h}=3\mathbf{g}$，這也證明了外心、形心和垂心位於同一條直線上，這條直線稱為歐拉線（Euler line），且 $OG:GH=1:2$。這種向量式的幾何，由不依賴坐標的恆等式擔起主要工作，正是大學一年級線性代數（linear algebra）處理幾何的方式；Strang 的 *Introduction to Linear Algebra*（線性代數入門教材）對此有詳細闡述，而 Siklos 的 *Advanced Problems in Mathematics* 則提供更多 STEP 風格的練習。

詳細解答

STEP I 2004，第 7 題

純數學 • 生成函數

> **STEP I 2004，第 7 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q7

關鍵思路

線性遞推關係（linear recurrence）與冪級數（power series）是同一個對象的兩種描述。把級數乘以 $1 - 7x + 10x^2$，每個係數 a_n 便會與 a_{n-1} 和 a_{n-2} 對齊，於是遞推關係令乘積中幾乎所有係數都變為零。剩下來的是一個多項式，所以 f 是有理函數（rational function）。接着利用部分分式（partial fractions）和幾何級數（geometric series），便可直接得出 a_n。第 (ii) 部分用的是同一套方法，只需先由已知數值求出 p 和 q。

解答

第 (i) 部分

當 $|x| < \frac{1}{5}$ 時，級數收斂（converge），所以我們可以把級數乘以 x 的冪，並逐項相加：

$$\begin{aligned} \mathrm{f}(x) &= a_0 + a_1 x + \sum_{n=2}^{\infty} a_n x^n, \\ -7x\,\mathrm{f}(x) &= -7a_0 x - \sum_{n=2}^{\infty} 7a_{n-1} x^n, \\ 10x^2 \mathrm{f}(x) &= \sum_{n=2}^{\infty} 10a_{n-2} x^n. \end{aligned}$$

（在第二行和第三行，我們重新標記了求和指標，使每個和式都是對 x^n（$n \geq 2$）求和。）相加得

$$\mathrm{f}(x) - 7x\,\mathrm{f}(x) + 10x^2 \mathrm{f}(x) = a_0 + (a_1 - 7a_0)x + \sum_{n=2}^{\infty} \left(a_n - 7a_{n-1} + 10a_{n-2}\right) x^n.$$

根據遞推關係，和式中每個括號都等於零，而 $a_0 = 2$，$a_1 - 7a_0 = 7 - 14 = -7$。所以

$$(1 - 7x + 10x^2)\,\mathrm{f}(x) = 2 - 7x.$$

現在 $1 - 7x + 10x^2 = (1 - 2x)(1 - 5x)$，當 $|x| < \frac{1}{5}$ 時不等於零，所以可以相除：

$$\mathrm{f}(x) = \frac{2 - 7x}{(1 - 2x)(1 - 5x)}.$$

分子正好是兩個因式之和，$(1 - 5x) + (1 - 2x) = 2 - 7x$，所以

$$\mathrm{f}(x) = \frac{(1 - 5x) + (1 - 2x)}{(1 - 2x)(1 - 5x)} = \frac{1}{1 - 2x} + \frac{1}{1 - 5x},$$

正如所求。

當 $|x| < \frac{1}{5}$ 時，我們有 $|2x| < 1$ 和 $|5x| < 1$，所以兩個分式都是收斂幾何級數之和：

$$\mathrm{f}(x) = \sum_{n=0}^{\infty} (2x)^n + \sum_{n=0}^{\infty} (5x)^n = \sum_{n=0}^{\infty} \left(2^n + 5^n\right) x^n.$$

想一想 為甚麼現在可以直接讀出 a_n？

我們有兩個冪級數，它們對區間 $|x| < \frac{1}{5}$ 內的每個 x 都相等。冪級數由它在 0 附近的值決定：它的係數可以由 $a_n = \mathrm{f}^{(n)}(0)/n!$ 還原出來。所以兩者的係數必定相同，於是

$$a_n = 2^n + 5^n \quad \text{對所有} n \geq 0.$$

（簡單驗算一下：$a_0 = 1 + 1 = 2$，$a_1 = 2 + 5 = 7$。）

第 (ii) 部分

首先要求出 p 和 q。把 $n = 2$ 和 $n = 3$ 代入遞推關係：

$$\begin{aligned} b_2 = pb_1 + qb_0: &\quad 40 = 10p + 5q, \quad \text{所以} 2p + q = 8; \\ b_3 = pb_2 + qb_1: &\quad 100 = 40p + 10q, \quad \text{所以} 4p + q = 10. \end{aligned}$$

兩式相減，得 $2p = 2$，所以 $p = 1$ 及 $q = 6$。遞推關係是 $b_n = b_{n-1} + 6b_{n-2}$。

與第 (i) 部分完全一樣，乘以 $1 - x - 6x^2$ 會消去由 x^2 起的所有係數：

$$(1 - x - 6x^2)\,\mathrm{g}(x) = b_0 + (b_1 - b_0)x = 5 + 5x.$$

由於 $1 - x - 6x^2 = (1 - 3x)(1 + 2x)$，當 $|x| < \frac{1}{3}$ 時不等於零，

$$\mathrm{g}(x) = \frac{5 + 5x}{(1 - 3x)(1 + 2x)} = \frac{A}{1 - 3x} + \frac{B}{1 + 2x},$$

其中 $5+5x = A(1+2x)+B(1-3x)$。代入 $x=\frac{1}{3}$ 得 $\frac{20}{3}=\frac{5}{3}A$，所以 $A=4$；代入 $x=-\frac{1}{2}$ 得 $\frac{5}{2}=\frac{5}{2}B$，所以 $B=1$。因此

$$\mathrm{g}(x)=\frac{4}{1-3x}+\frac{1}{1+2x}.$$

當 $|x|<\frac{1}{3}$ 時，我們有 $|3x|<1$ 和 $|2x|<1$，所以把兩者都展開成幾何級數，

$$\mathrm{g}(x)=\sum_{n=0}^{\infty}\left(4\cdot 3^n+(-2)^n\right)x^n,$$

再如前比較係數，

$$b_n=4\cdot 3^n+(-2)^n.$$

驗算

這些公式與已知數值吻合：$4+1=5$，$12-2=10$，$36+4=40$，$108-8=100$。由遞推關係逐項計算 a_n 和 b_n 直至 $n=39$，結果與 2^n+5^n 和 $4\cdot 3^n+(-2)^n$ 完全一致。數值上，在 $x=0.1$ 時，級數的部分和（partial sums）給出 f 的值為 3.25，g 的值為 6.547619…，與 $\frac{1}{0.8}+\frac{1}{0.5}=3.25$ 和 $\frac{4}{0.7}+\frac{1}{1.2}=\frac{275}{42}$ 相符。

常見錯誤

- 忘記遞推關係只由 $n=2$ 起成立：乘積中的 x^0 項和 x^1 項必須另行計算，而它們正好給出分子。
- 在第 (ii) 部分，在 p 和 q 仍然未知時便使用遞推關係，或把分子寫成 b_0+b_1x 而不是 $b_0+(b_1-pb_0)x$。
- 因式分解 $1-x-6x^2$ 時弄錯正負號，結果得出 3^n 和 2^n，而不是 3^n 和 $(-2)^n$。

STEP 之外

級數 f 稱為數列 (a_n) 的*生成函數*（generating function），而這個方法適用於任何常係數線性遞推關係：分母總是由遞推關係構造出來，而它的因式則給出「特徵根」（characteristic roots）（這裏是 2 和 5）。生成函數是組合數學（combinatorics）和概率論的核心工具；在概率論中，概率生成函數（probability generating functions）把隨機變量之和化為乘積。「相等的冪級數有相等的係數」這一步，要在大學第一門分析課程中才有嚴格的論證；關於冪級數及其唯一性，可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP I 2004，第 8 題

純數學 • 單調數列與數學歸納法

STEP I 2004，第 8 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s1-q8

關鍵思路

要證明一個數列（sequence）遞增，可把差 $t_{n+1} - t_n$ 看成 t_n 的函數。在兩部分中，這個差都可以因式分解，而所得的因式正好對應遞推關係（recurrence）的不動點（fixed point）2 和 3（即令 $t_{n+1} = t_n$ 的值）。於是差的正負只取決於 t_n 相對於 2 和 3 的位置，所以真正要下工夫的，是用歸納法（induction）證明每一項都留在適當的區域內。

解答

第 (i) 部分

對任意 n，

$$x_{n+1} - x_n = \frac{x_n^2 + 6}{5} - x_n = \frac{x_n^2 - 5x_n + 6}{5} = \frac{(x_n - 2)(x_n - 3)}{5}.$$

若 $x_n > 3$，則兩個因式皆為正，所以 $x_{n+1} - x_n > 0$，即 $x_{n+1} > x_n$。

餘下要驗證每一項都大於 3，我們用歸納法來證明。根據假設，命題 $x_n > 3$ 在 $n = 0$ 時成立。假設 $x_n > 3$。由上面的計算，$x_{n+1} > x_n > 3$，所以亦有 $x_{n+1} > 3$。因此對所有 $n \geq 0$ 都有 $x_n > 3$，從而對所有 $n \geq 0$ 都有 $x_{n+1} > x_n$：這個數列嚴格遞增（strictly increasing）。

? **想一想** 為甚麼只證明 $x_1 > x_0$ 並不足夠？

不等式 $x_{n+1} > x_n$ 是由 $x_n > 3$ 推導出來的，所以我們在每一步都需要這個條件，而不僅是第一步。在這裏，這個條件不費工夫便可得到：從大於 3 的位置再向上走一步，仍然大於 3。（留意若起始值恰好是 $x_0 = 3$，便會得到常數數列，所以嚴格不等式是必要的。）

第 (ii) 部分

首先留意，若 $2 < y_n < 3$，則 $y_n > 0$，所以 $y_{n+1} = 5 - 6/y_n$ 有定義。對這樣的 y_n，

$$y_{n+1} - y_n = 5 - \frac{6}{y_n} - y_n = -\frac{y_n^2 - 5y_n + 6}{y_n} = \frac{(y_n - 2)(3 - y_n)}{y_n},$$

此式為正，因為 $y_n - 2 > 0$、$3 - y_n > 0$ 及 $y_n > 0$。所以只要 $2 < y_n < 3$，便有 $y_{n+1} > y_n$。

與第 (i) 部分一樣，我們必須證明每一項都位於這個區間內。這次它不會自動成立，因為遞增的數列有可能越過 3。所以我們考察各項與 3 的距離：

$$3 - y_{n+1} = 3 - 5 + \frac{6}{y_n} = \frac{6 - 2y_n}{y_n} = \frac{2(3 - y_n)}{y_n}.$$

? **想一想** 我們怎樣知道這個數列永遠不會跳過 3？

上式說明，只要 $y_n > 0$，$3 - y_{n+1}$ 便與 $3 - y_n$ 同號。所以小於 3 的一項之後，緊接的一項必定也小於 3。

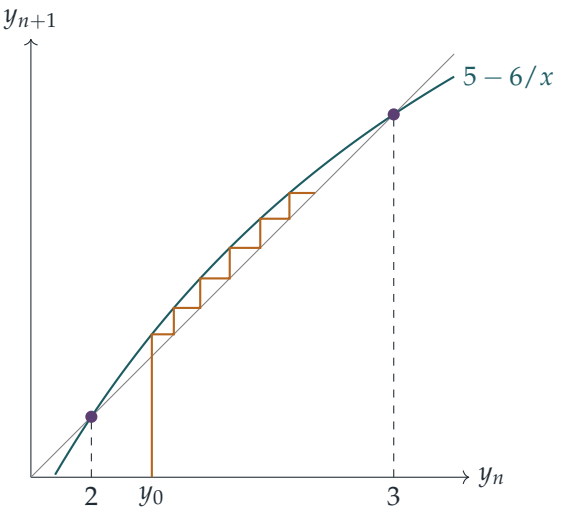
現在進行歸納。斷言：對所有 $n \geq 0$，$2 < y_n < 3$。根據假設，這在 $n = 0$ 時成立。假設 $2 < y_n < 3$。那麼：

- 由關於差的計算，$y_{n+1} > y_n > 2$；
- $3 - y_{n+1} = 2(3 - y_n)/y_n > 0$，因為 $3 - y_n > 0$ 及 $y_n > 0$，所以 $y_{n+1} < 3$。

因此 $2 < y_{n+1} < 3$，歸納完成。

由於對每個 n 都有 $2 < y_n < 3$，關於差的計算在每一步都適用，所以對所有 $n \geq 0$ 都有 $y_{n+1} > y_n$。這個數列嚴格遞增，但對所有 n 仍有 $y_n < 3$，正如所求。

蛛網圖（cobweb diagram）顯示了實際的情況：曲線 $y = 5 - 6/x$ 與直線 $y = x$ 相交於 $x = 2$ 和 $x = 3$，在兩者之間曲線位於直線上方，而各次迭代的值向着 3 攀升，卻始終達不到它。



驗算

- 在不動點處，遞推關係給出常數數列：$(3^2+6)/5 = 3$，以及 $5-6/3 = 3$、$5-6/2 = 2$，與因式 $(t-2)(t-3)$ 吻合。
- 由 $y_0 = 2.5$ 開始：$y_1 = 2.6$，$y_2 \approx 2.692$，$y_3 \approx 2.771$，遞增且小於 3；經過 200 步後，y_n 與 3 相差不超過 10^{-15}。由 $x_0 = 3.1$ 開始：$x_1 = 3.122$，$x_2 \approx 3.149$，遞增。
- 我們用電腦以精確分數對數百個隨機起始值檢驗了兩個歸納論證，並確認當 x_0 略小於 3 時會得到 $x_1 < x_0$，所以條件 $x_0 > 3$ 是必需的。

常見錯誤

- 在假設 $x_n > 3$ 之下證明了 $x_{n+1} > x_n$ 便停下來，沒有用歸納法證明每一項都大於 3。
- 在第 (ii) 部分，證明了 $y_{n+1} < 3$，卻忘記要繼續歸納還需要 $y_{n+1} > 2$（它可由 $y_{n+1} > y_n$ 推出）。
- 把不等式兩邊同乘以 y_n，卻沒有指出 $y_n > 0$。

STEP 之外

第 (ii) 部分的數列是有界的遞增數列，而數學分析中的單調收斂定理（monotone convergence theorem）指出，這樣的數列必定收斂；任何極限 L 都滿足 $L = 5-6/L$，所以這裏 $y_n \to 3$。在第 (i) 部分，數列遞增而且無上界，因為有限的極限必須是一個大於

3 的不動點，而這樣的不動點並不存在。迭代的值究竟被不動點吸引還是排斥，取決於映射在該點的斜率，這正是離散動力系統（discrete dynamical systems）理論的起點。Abbott 的 *Understanding Analysis* 對單調收斂有嚴謹細緻的論述。

詳細解答

STEP I 2004，第 9 題

力學 • 拋體運動

> **STEP I 2004，第 9 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s1-q9

關鍵思路

第一部分是標準的「在最高點垂直速度為零」的計算。在隧道中，題目沒有給出速率 u，但它也不是任意的：一旦選定 θ，要求路徑通過 P 就確定了 u。因此我們消去 u，把最大高度寫成只含 θ 的函數（以 $t = \tan\theta$ 表示最為方便），再問它何時保持在隧道頂之下。答案最後化為一個關於 t 的二次不等式（quadratic inequality）。

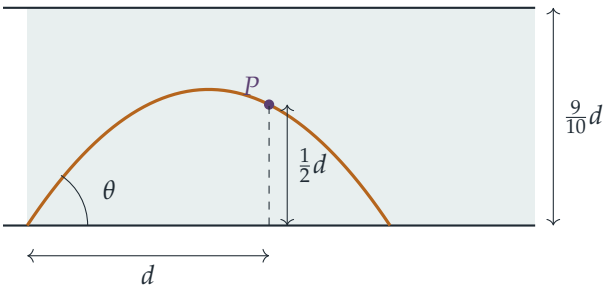
解答

最大高度

把粒子視為在重力作用下自由運動的質點，不計空氣阻力（air resistance）。取拋射點為原點建立坐標軸，x 軸水平，y 軸垂直向上。速度的垂直分量（vertical component）起初為 $u\sin\theta$，並有恆定加速度 $-g$。在最高點垂直速度為零，所以在垂直方向應用 $v^2 = v_0^2 + 2as$，得

$$0 = u^2\sin^2\theta - 2gH, \qquad \text{所以} \qquad H = \frac{u^2\sin^2\theta}{2g}.$$

隧道



粒子的路徑可由 $x = uT\cos\theta$ 與 $y = uT\sin\theta - \frac{1}{2}gT^2$ 消去時間 T 而得：

$$y = x\tan\theta - \frac{gx^2}{2u^2\cos^2\theta} = xt - \frac{gx^2(1+t^2)}{2u^2},$$

其中 $t = \tan\theta$，並用了 $\sec^2\theta = 1 + t^2$。

通過 P。 代入 $x = d$、$y = \frac{1}{2}d$，再除以 d：

$$\frac{1}{2} = t - \frac{gd(1+t^2)}{2u^2}, \qquad \text{所以} \qquad \frac{gd(1+t^2)}{2u^2} = t - \frac{1}{2}.$$

左邊為正數，所以必須有 $t > \frac{1}{2}$。（這是合理的：粒子起初朝 θ 方向運動，而重力只會把路徑向下彎，所以只有瞄準的方向高於通往 P 的直線時才能到達 P，而該直線的斜率為 $\frac{1}{2}$。）對每個這樣的 θ，恰好有一個速率符合要求：

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

想一想 題目沒有給出 u 的值。條件怎麼可能只與 θ 有關？

因為速率並不由我們選擇：一旦 θ 固定，「通過 P」這個條件就把 u 確定下來。所以最大高度其實只是 θ 的函數。

不碰隧道頂。 把 $\sin^2\theta = \dfrac{t^2}{1+t^2}$ 代入 H 的公式，

$$H = \frac{u^2}{2g}\cdot\frac{t^2}{1+t^2} = \frac{d(1+t^2)}{2(2t-1)}\cdot\frac{t^2}{1+t^2} = \frac{dt^2}{2(2t-1)}.$$

粒子不碰隧道頂，當且僅當 $H < \frac{9}{10}d$。由於 $2t-1 > 0$，我們可以把不等式兩邊同乘 $10(2t-1)/d$ 而不必改變不等號的方向：

$$\begin{aligned}\frac{dt^2}{2(2t-1)} < \frac{9d}{10} &\iff 5t^2 < 9(2t-1)\\ &\iff 5t^2 - 18t + 9 < 0\\ &\iff (5t-3)(t-3) < 0\\ &\iff \tfrac{3}{5} < t < 3.\end{aligned}$$

這個區間內的每個 t 都滿足 $t > \frac{1}{2}$，所以先前的條件沒有帶來額外限制。若 $t = \frac{3}{5}$ 或 $t = 3$，粒子剛好觸及隧道頂，這種情況被排除。由於粒子始終沒有碰到隧道頂，它會落回地面，並在隧道內着地。

最後，θ 介乎 0 與 $\frac{\pi}{2}$ 之間，而 tan 在此範圍內遞增，所以

$$\arctan\tfrac{3}{5} < \theta < \arctan 3,$$

這正是所要證明的結果。

驗算

- 當 $t = 3$ 時：$H = \dfrac{9d}{2\cdot 5} = \frac{9}{10}d$；當 $t = \frac{3}{5}$ 時：$H = \dfrac{(9/25)d}{2\cdot(1/5)} = \frac{9}{10}d$。兩個端點都剛好擦過隧道頂，正如所料。
- 當 $t \to \frac{1}{2}^{+}$ 時，所需速率 $u \to \infty$，且 $H \to \infty$：粒子幾乎沿着通往 P 的直線運動，所以它必須非常快，其後亦會升得非常高。
- 數值方面，取 $d = 3$ 和 $g = 9.8$，我們測試了 10^5 個隨機角度。對每個角度，我們求出 u，確認路徑通過 P，並發現 $H < \frac{9}{10}d$ 當且僅當 $0.5404 < \theta < 1.2490$（即以弧度表示的 $\arctan 0.6$ 和 $\arctan 3$）。逐步模擬亦確認了 $H = u^2\sin^2\theta/(2g)$。

常見錯誤

- 把 u 當作另一個獨立的未知數，結果做不下去：必須利用在 P 的條件把它消去。
- 把不等式兩邊同乘 $2t-1$ 之前，忘記檢查它的正負。
- 只檢查 P 是否在隧道頂之下。路徑的最高點通常出現在 P 之後，所以真正重要的是最大高度。

STEP 之外

把軌跡（trajectory）以 $\tan\theta$ 表示，可將拋體（projectile）問題化為關於 t 的二次式。同一技巧可求出以給定速率拋射的所有軌跡的「包絡線」（envelope，即作為邊界的拋物線），並給出擊中目標的兩個可能角度。像這樣的問題，部分參數由約束條件確定，其餘參數則構成一個容許區域，由此可通往最優化（optimisation）和變分法（calculus of variations）。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）是進一步學習這類力學的極佳讀物。

詳細解答

STEP I 2004，第 10 題

力學 • 勻加速運動學

STEP I 2004，第 10 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

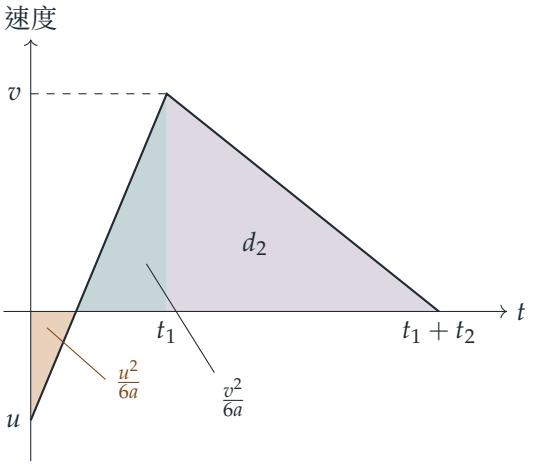
step.maths.org/questions/04-s1-q10

關鍵思路

這裏全部都是勻加速運動（constant acceleration），所以單靠 $v^2 = u^2 + 2as$ 和 $v = u + at$ 便足以解決整條題目。唯一的陷阱是距離（distance）這個詞。當 $u < 0$ 時，質點（particle）先向後運動，停下來，然後才向前加速，所以距離 d_1 是兩段路程之和，而不是單一個位移（displacement）。畫一幅速度—時間圖（velocity–time graph）便一目了然：距離就是圖線與時間軸之間的總面積，軸下方的面積也當作正數計算。

解答

取最終速度 v 的方向為正方向。在第一階段，加速度是 $+3a$（它令速度由 u 上升至 $v > u$）；在第二階段，加速度是 $-a$，直至速度變為 0。全程均有 $a > 0$。



上圖所示為 $u < 0$ 的情況：軸下方的小三角形就是向後運動的那段路程。

第二階段（兩種情況皆適用）。 質點以減速度（deceleration）a 由速率 v 減至靜止，期間從不掉頭，所以

$$0 = v^2 - 2ad_2 \quad \Longrightarrow \quad d_2 = \frac{v^2}{2a}.$$

第 (i) 部分

若 $u > 0$，速度由 u 上升至 v，一直保持正值，所以距離等於位移：

$$v^2 = u^2 + 2(3a)d_1 \quad \Longrightarrow \quad d_1 = \frac{v^2 - u^2}{6a}.$$

因此

$$3d_1 = \frac{v^2 - u^2}{2a} < \frac{v^2}{2a} = d_2,$$

因為 $u^2 > 0$。所以 $3d_1 < d_2$，正如所求。

第 (ii) 部分

現在 $u < 0$，第一階段要分成兩段。

> ? **想一想** 為甚麼不能再直接使用 $v^2 = u^2 + 6ad_1$？

這條公式給出的是位移，即 $\frac{v^2-u^2}{6a}$。但質點是先向後運動的。在加速度 $3a$ 之下，它由速率 $|u|$ 減慢至靜止，走過 $\frac{u^2}{6a}$（由 $0 = u^2 - 2(3a)s$ 得出）。然後它由靜止加速至 v，走過 $\frac{v^2}{6a}$。距離是兩者之和：

$$d_1 = \frac{u^2 + v^2}{6a}, \qquad \text{所以} \qquad 3d_1 = \frac{u^2 + v^2}{2a}.$$

與 $d_2 = \frac{v^2}{2a}$ 和 $2d_2 = \frac{2v^2}{2a}$ 比較：

- $3d_1 - d_2 = \dfrac{u^2}{2a} > 0$，因為 $u \neq 0$；
- $2d_2 - 3d_1 = \dfrac{v^2 - u^2}{2a} > 0$，因為 $v > |u|$。

因此 $d_2 < 3d_1 < 2d_2$，正如所求。

平均速率

所需時間與方向無關。在第一階段，速度以變率 $3a$ 改變了 $v - u$；在第二階段，速度以變率 a 改變了 v，所以

$$t_1 = \frac{v - u}{3a}, \qquad t_2 = \frac{v}{a}, \qquad t_1 + t_2 = \frac{4v - u}{3a},$$

不論 u 的正負如何。

$u>0$ 的情況。總距離是 $\frac{v^2-u^2}{6a}+\frac{3v^2}{6a}=\frac{4v^2-u^2}{6a}$，所以平均速率（average speed）是

$$S_+=\frac{4v^2-u^2}{6a}\cdot\frac{3a}{4v-u}=\frac{4v^2-u^2}{2(4v-u)}.$$

$u<0$ 的情況。設 $u=-w$，其中 $0<w<v$。總距離是 $\frac{v^2+w^2}{6a}+\frac{3v^2}{6a}=\frac{4v^2+w^2}{6a}$，所需時間是 $\frac{4v+w}{3a}$，所以

$$S_-=\frac{4v^2+w^2}{2(4v+w)}.$$

要比較這兩個數，可把它們分別與 $\frac{1}{2}v$ 比較：

$$S_+-\tfrac{1}{2}v=\frac{4v^2-u^2-v(4v-u)}{2(4v-u)}=\frac{u(v-u)}{2(4v-u)},$$
$$S_--\tfrac{1}{2}v=\frac{4v^2+w^2-v(4v+w)}{2(4v+w)}=-\frac{w(v-w)}{2(4v+w)}.$$

在第一種情況中 $0<u<v$，第一個式子的分子和分母皆為正數，所以 $S_+>\frac{1}{2}v$。在第二種情況中 $0<w<v$，第二個式子是負數，所以 $S_-<\frac{1}{2}v$。因此

$$S_-<\tfrac{1}{2}v<S_+,$$

即 $u>0$ 時平均速率較大，正如所求。留意這個結論對兩種情況中任何 $|u|$ 的值都成立，而不限於兩者相等的時候：只要 v 和 a 相同，$\frac{1}{2}v$ 便把每一段 $u>0$ 的旅程與每一段 $u<0$ 的旅程分隔開。

驗算

- 取 $v=2$，$a=1$。當 $u=1$：$d_1=\frac{1}{2}$，$d_2=2$，所以 $3d_1=1.5<2$，所需時間是 $\frac{7}{3}$，得出 $S_+=\frac{2.5}{7/3}\approx 1.071>1$。當 $u=-1$：$d_1=\frac{5}{6}$，所以 $2<3d_1=2.5<4$，所需時間是 3，得出 $S_-=\frac{17/6}{3}\approx 0.944<1$。以數值方法逐步模擬這個運動，也得出相同的數值。
- 極限（limit）：當 $u\to 0$ 或 $|u|\to v$ 時，兩個平均速率都趨向 $\frac{1}{2}v$，這正是由靜止起步、最後停下的簡單旅程的平均速率。在容許範圍的邊界上，各個不等式也變成等式：在兩個部分中，當 $u\to 0$ 時 $3d_1\to d_2$；在第 (ii) 部分中，當 $u\to -v$ 時 $3d_1\to 2d_2$。

常見錯誤

- 在第 (ii) 部分使用 $d_1 = \frac{v^2-u^2}{6a}$。這是位移，而且會再次得出 $3d_1 < d_2$，與題目要求證明的結果矛盾。
- 以為 $u < 0$ 時，時間也需要分段計算。其實不必：$v = u + 3at_1$ 在經過折返點前後一直成立。

STEP 之外

位移與距離之別，就是 $\int v\,\mathrm{d}t$ 與 $\int |v|\,\mathrm{d}t$ 之別。後者是路徑的弧長（arc length），也就是汽車里程錶所量度的數值。在大學的力學課程中，弧長成為描述彎曲軌跡的自然參數，由此引出加速度的切向分量（tangential component）和法向分量（normal component）。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）正是以這種方式建立運動學（kinematics）的。

詳細解答

STEP I 2004，第 11 題

力學 • 摩擦力與鉸接梯子

STEP I 2004，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s1-q11

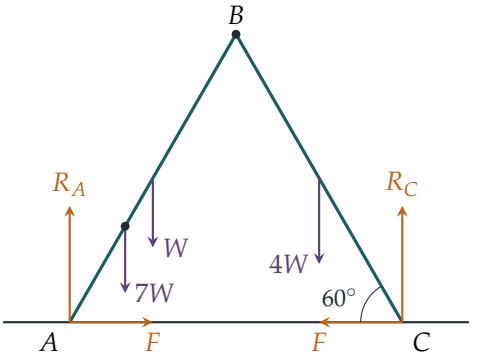
關鍵思路

在任何一處滑動之前，整個梯架處於平衡（equilibrium），所以我們可以由整個系統求出兩個法向反作用力（normal reaction），再單獨考慮其中一把梯子來求摩擦力（對鉸鏈取力矩，便可消去未知的鉸鏈力）。水平方向的外力只有兩個摩擦力，所以它們大小相等。法向反作用力較小的那隻梯腳，所需的比值 F/R 較大，因此最先滑動的就是那隻梯腳。

解答

我們把每把梯子視為長度為 $2a$ 的均勻桿（uniform rod），把 B 處的鉸鏈（hinge）視為光滑，並把裝修工人視為質點（particle）。由於 $AB = BC$ 且 $\angle ABC = 60°$，三角形 ABC 是等邊三角形：$AC = 2a$，而每把梯子與地面成 $60°$ 角。於是 B 與每隻梯腳的水平距離為 a，離地面的高度為 $a\sqrt{3}$。

當裝修工人沿 AB 爬到三分之一處時，她與 A 的水平距離為 $\frac{1}{3}a$。AB 的中心與 A 的水平距離為 $\frac{1}{2}a$，而 BC 的中心與 A 的水平距離為 $\frac{3}{2}a$。



法向反作用力

設 A 和 C 處的法向反作用力分別為 R_A 和 R_C。對整個系統沿鉛垂方向分解力，得

$$R_A + R_C = W + 4W + 7W = 12W.$$

對整個系統以 A 為支點取力矩（moments）（鉸鏈力是內力，所以不會出現）：

$$R_C \cdot 2a = W \cdot \tfrac{1}{2}a + 7W \cdot \tfrac{1}{3}a + 4W \cdot \tfrac{3}{2}a = \tfrac{53}{6}Wa,$$

所以

$$R_C = \tfrac{53}{12}W, \qquad R_A = 12W - \tfrac{53}{12}W = \tfrac{91}{12}W.$$

摩擦力

兩隻梯腳有向外張開的趨勢，所以每隻梯腳所受的摩擦力（friction）都指向內側。對整個系統沿水平方向分解力，摩擦力是僅有的水平外力，所以兩者大小相同；記這個大小為 F。

要求出 F，我們需要單獨考慮其中一把梯子。取沒有裝修工人的 BC，並以 B 為支點取力矩，使未知的鉸鏈力消失。以 B 為支點，反作用力 R_C 的力臂（lever arm）為 a，重量 $4W$ 的力臂為 $\frac{1}{2}a$，而摩擦力 F（水平方向，作用於地面的高度）的力臂為 $a\sqrt{3}$。重量和 F 使 BC 向一方轉動，而 R_C 使它向另一方轉動，所以

$$R_C \cdot a = 4W \cdot \tfrac{1}{2}a + F \cdot a\sqrt{3} \quad \Longrightarrow \quad F = \frac{R_C - 2W}{\sqrt{3}} = \frac{29W}{12\sqrt{3}}.$$

作為驗算，對 AB 以 B 為支點作同樣的計算，得 $R_A \cdot a = W \cdot \frac{1}{2}a + 7W \cdot \frac{2}{3}a + Fa\sqrt{3}$，所以 $F\sqrt{3} = \frac{91}{12}W - \frac{31}{6}W = \frac{29}{12}W$，與上面的結果一致。

哪一把梯子會滑動？

想一想 兩隻梯腳所受的摩擦力相同，那麼我們怎樣判斷哪一隻會先失守？

當所需的摩擦力等於 μ 乘以該處的法向反作用力時，梯腳便處於即將滑動的狀態。兩隻梯腳所需的 F 相同，所以法向反作用力較小的梯腳所需的比值較大：

$$\frac{F}{R_A} = \frac{29}{91\sqrt{3}} \approx 0.184, \qquad \frac{F}{R_C} = \frac{29}{53\sqrt{3}} \approx 0.316.$$

所以 C 最先達到極限摩擦（limiting friction）：滑動的是**較重的梯子** BC，儘管裝修工人是在 AB 上。她的重量把 A 牢牢壓在地面上，而 C 所承受的負荷則少得多。

我們還應確定在此之前沒有任何一處已經滑動。若她爬到全長的 t 倍處，同樣的計算給出 $R_C = (\frac{13}{4} + \frac{7}{2}t)W$ 和 $F\sqrt{3} = R_C - 2W = (\frac{5}{4} + \frac{7}{2}t)W$，所以

$$\frac{F}{R_C} = \frac{1}{\sqrt{3}} \cdot \frac{\frac{5}{4} + \frac{7}{2}t}{\frac{13}{4} + \frac{7}{2}t} = \frac{1}{\sqrt{3}}\left(1 - \frac{2}{\frac{13}{4} + \frac{7}{2}t}\right),$$

它隨 t 遞增。同樣，F/R_A 也隨 t 遞增（它的分子上升，而分母 $R_A = 12W - R_C$ 下降）。所以隨着她向上爬，所需的比值不斷增大，而第一次滑動恰好發生在 F/R_C 達到 μ 之時，即 $t = \frac{1}{3}$。因此

$$\mu = \frac{29}{53\sqrt{3}} = \frac{29\sqrt{3}}{159} \approx 0.316.$$

驗算

- 對 AB 以 B 為支點取力矩，所得的 F 與對 BC 以 B 為支點取力矩所得的相同（見上文），所以各反作用力互相吻合。
- 完全沒有裝修工人時，由於 BC 較重，R_C 應大於 R_A；事實上，此時以 A 為支點取力矩得 $R_C = \frac{13}{4}W$ 和 $R_A = \frac{7}{4}W$，兩者之和為 $5W$。
- 當 $t = \frac{11}{14}$ 時 $R_A = R_C$；只有超過這一點，滑動的才會是 AB，所以在 $t = \frac{1}{3}$ 時滑動的必定是 BC。
- 數值上，由坐標重新計算每一個力（並對兩把梯子檢查以 B 為支點的力矩），得 $\mu \approx 0.31591$。

常見錯誤

- 以為載着裝修工人的那把梯子會滑動。她的重量增大了 A 處的法向反作用力，使那隻梯腳更加穩固。

- 對單獨一把梯子以梯腳為支點取力矩，這樣會引入未知的鉸鏈力。對一把梯子以 B 為支點取力矩便可避開它。
- 忘記摩擦力對鉸鏈的力臂為 $a\sqrt{3}$（即 B 的高度）。

STEP 之外

這是處理鉸接結構的靜力學（statics）基本策略：先把整個結構當作一個整體來求外部反作用力，然後在某個接合處把它切開，求出其中一個構件必須提供的力。哪一個接觸點最先失效這類問題，會引向框架的分析，以及虛功原理(principle of virtual work)；在虛功原理中，平衡是透過考慮符合約束條件的微小位移來求得的。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）對剛體靜力學和摩擦有細緻的論述。

詳細解答

STEP I 2004，第 12 題

概率與統計 • 貝葉斯定理

STEP I 2004，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s1-q12

關鍵思路

從第 1 組抽出的樣本，大致告訴我們：已知一塊晶片屬於第 1 組，它來自機器 A 的條件概率（conditional probability）是多少。因此，我們利用貝葉斯定理（Bayes' theorem），以 λ、p 和 q 表示這個條件概率，令它等於觀察所得的比例，再解出 λ。分揀出錯只改變了每部機器所生產的晶片落入第 1 組的概率；方法則完全相同。

解答

從工廠的產品中隨機抽出一塊晶片。設 A 為該晶片由機器 A 製造的事件（故 $\mathrm{P}(A) = \lambda$ 及 $\mathrm{P}(A') = 1 - \lambda$），$P$ 為該晶片完美無瑕的事件，G 為該晶片被放入第 1 組的事件。題目給出 $\mathrm{P}(P \mid A) = p$ 及 $\mathrm{P}(P \mid A') = q$。

從第 1 組隨機抽出的一塊晶片，就是在 G 發生的條件下，從全部產品中隨機抽出的一塊晶片。由於樣本很大，觀察所得的比例 $\frac{2}{5}$ 是 $\mathrm{P}(A \mid G)$ 的良好估計，而我們對 λ 的估計值，就是使 $\mathrm{P}(A \mid G)$ 恰好等於 $\frac{2}{5}$ 的那個值。

分揀完全準確

此時 G 與 P 是同一事件。根據貝葉斯定理，

$$\mathrm{P}(A \mid P) = \frac{\mathrm{P}(P \mid A)\,\mathrm{P}(A)}{\mathrm{P}(P \mid A)\,\mathrm{P}(A) + \mathrm{P}(P \mid A')\,\mathrm{P}(A')} = \frac{\lambda p}{\lambda p + (1-\lambda)q}.$$

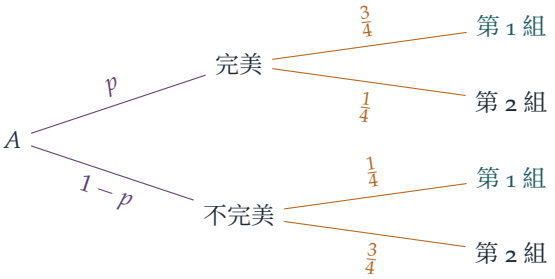
令上式等於 $\frac{2}{5}$，然後交叉相乘（分母為正數）：

$$\begin{aligned} 5\lambda p &= 2\lambda p + 2(1-\lambda)q \\ 3\lambda p + 2\lambda q &= 2q \\ \lambda &= \frac{2q}{3p+2q}, \end{aligned}$$

這就是所求的估計值，因為 $3p + 2q > 0$。

分揀出錯

現在，一塊晶片進入第 1 組有兩種途徑：本身完美而被正確分揀，或本身不完美而被錯誤分揀。我們假設分揀出錯與晶片由哪部機器製造互相獨立 (independent)，也就是說，在已知晶片是否完美的情況下，它如何被分揀與它來自哪部機器無關。下面的樹形圖顯示來自機器 A 的晶片的各個分支；機器 B 的情況相同，只需把 p 換成 q。



把通往第 1 組的兩個分支相加，

$$\mathrm{P}(G \mid A) = \tfrac{3}{4}p + \tfrac{1}{4}(1-p) = \frac{1+2p}{4}, \qquad \mathrm{P}(G \mid A') = \tfrac{3}{4}q + \tfrac{1}{4}(1-q) = \frac{1+2q}{4}.$$

於是，再次運用貝葉斯定理（因子 $\frac{1}{4}$ 互相約去），

$$\mathrm{P}(A \mid G) = \frac{\lambda(1+2p)}{\lambda(1+2p) + (1-\lambda)(1+2q)}.$$

想一想 證據有沒有改變？$\frac{2}{5}$ 還應該沿用嗎？

應該沿用：樣本確實是從第 1 組抽出的，而其中確實有 $\frac{2}{5}$ 來自機器 A。改變了的，只是我們對晶片如何進入第 1 組所建立的模型。因此，我們令 $\mathrm{P}(A \mid G)$ 的新表達式等於 $\frac{2}{5}$：

$$\begin{aligned} 5\lambda(1+2p) &= 2\lambda(1+2p) + 2(1-\lambda)(1+2q) \\ \lambda\big[3(1+2p) + 2(1+2q)\big] &= 2(1+2q) \\ \lambda &= \frac{2(1+2q)}{5+6p+4q}. \end{aligned}$$

這就是 λ 的新估計值。留意它其實就是第一個答案，只是把 p 和 q 分別換成 $\frac{1+2p}{4}$ 和 $\frac{1+2q}{4}$，即進入第 1 組的新概率。

驗算

- 若 $p = q$，兩部機器在品質上無從分辨，因此第 1 組的組成應與整體產量的比例一致，而 λ 應為 $\frac{2}{5}$。兩條公式都給出這個結果：$\frac{2p}{5p} = \frac{2}{5}$ 及 $\frac{2(1+2p)}{5+10p} = \frac{2}{5}$。
- 兩個估計值都嚴格介乎 0 與 1 之間，因為兩者的分子均為正數，且小於分母。
- 以 $\lambda = 0.3$、$p = 0.6$、$q = 0.8$ 及分揀出錯的情況模擬 4×10^5 塊晶片，結果第 1 組中有 0.2667 來自機器 A，而公式給出的數值是 0.2661；把模擬所得的比例代入公式反過來求解，得回 $\lambda \approx 0.301$。

常見錯誤

- 令 $\frac{2}{5}$ 等於 $\mathrm{P}(G \mid A)$ 或 λp，而不是等於條件概率 $\mathrm{P}(A \mid G)$。
- 在第二部分，忘記了被錯誤放入第 1 組的不完美晶片，即只用了 $\frac{3}{4}p$。

STEP 之外

選取能使觀察所得數據「完全吻合」的參數值，這種做法稱為*矩估計法*（method of moments）；在本題中，它與最大似然估計（maximum likelihood estimation）的結果一致，而後者你會在大學的第一門統計學課程中遇到。出錯的分揀機是量度誤差（measurement error）的一個簡單模型，這種誤差又稱錯誤分類（misclassification），在醫學檢測中是個嚴重的實際問題：假陽性（false positive）和假陰性（false negative）正是以這種方式扭曲觀察所得的比例。想進一步了解貝葉斯定理和條件化，可參閱 Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）。

詳細解答

STEP I 2004，第 13 題

概率與統計 • 均勻分佈變量的最大值

STEP I 2004，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s1-q13

關鍵思路

若干個獨立（independent）的數的最大值不超過 x，當且僅當其中每一個數都不超過 x。這樣一來，棘手的「最大值」就化為若干個相同概率的乘積，一行便可寫出 X 的分佈函數（distribution function）。其餘一切都由此而來：求導可得密度（density）和期望值（expectation），而令分佈函數等於 0.05，便可求出檢驗的臨界值（critical value）。

解答

第 (i) 部分

設 U_1, U_2, U_3 為該三個數，它們互相獨立，且各自在 $[0,1]$ 上服從均勻分佈（uniform distribution），所以當 $0 \le x \le 1$ 時，$\mathrm{P}(U_i \le x) = x$。由於 $X = \max(U_1, U_2, U_3)$，

$$X \le x \iff U_1 \le x,\ U_2 \le x \text{ 及 } U_3 \le x.$$

由獨立性，

$$F(x) = \mathrm{P}(X \le x) = x^3 \qquad (0 \le x \le 1).$$

特別地，$\mathrm{P}(X \le 0.8) = 0.8^3 = 0.512$，正是所求。

求導後，X 在 $[0,1]$ 上的密度為 $f(x) = F'(x) = 3x^2$，所以

$$\mathrm{E}(X) = \int_0^1 x \cdot 3x^2 \,\mathrm{d}x = \left[\tfrac{3}{4}x^4\right]_0^1 = \frac{3}{4}.$$

這個結果合情合理：三個數中最大的一個，理應遠高於單獨一個數的平均值 $\frac{1}{2}$。

第 (ii) 部分

現在有 N 個獨立的數，各自在 $[0,a]$ 上服從均勻分佈，所以當 $0 \le x \le a$ 時，$\mathrm{P}(U_i \le x) = x/a$，而同樣的論證給出

$$\mathrm{P}(X \le x) = \left(\frac{x}{a}\right)^N \qquad (0 \le x \le a),$$

而當 $x \ge a$ 時，$\mathrm{P}(X \le x) = 1$。由於 X 是連續的，$\mathrm{P}(X < x)$ 與此相同。

臨界值。 當 $x < c$ 時，檢驗拒絕 H_0；而 5% 的顯著性水平（significance level）表示，在 H_0 為真（$a = 1$）時拒絕它的概率為 0.05：

$$\mathrm{P}(X < c \mid a = 1) = c^N = 0.05, \qquad \text{所以} \qquad c = 0.05^{1/N}.$$

x 的值偏小，是支持 $a < 1$ 的證據，這就是拒絕域（rejection region）為 $x < c$ 的原因。

? **想一想** 「$x \le 0.8$ 必然導致拒絕」怎樣化為一個關於 N 的不等式？

每一個 $x \le 0.8$ 都被拒絕，當且僅當每一個這樣的 x 都小於 c，即 $0.8 < c$。由於 $t \mapsto t^N$ 在 $t > 0$ 時是遞增的，這等價於

$$0.8^N < c^N = 0.05.$$

所以我們要找滿足 $0.8^N < 0.05$ 的最小整數 N。（留意當 N 增大時，c 會向 1 遞增，所以 N 愈大，檢驗就愈傾向於拒絕。）

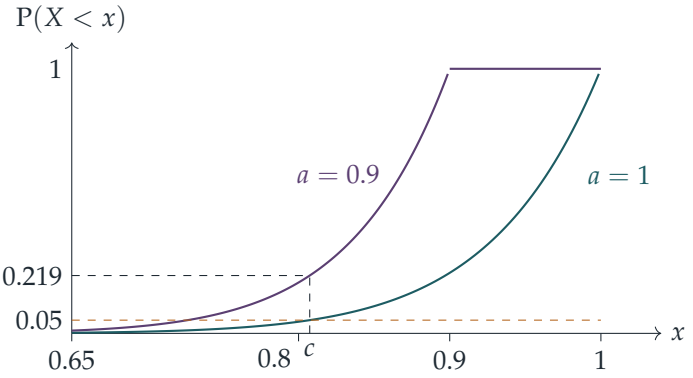
利用 $2^{10} \approx 10^3$。 寫成 $0.8 = 2^3/10$，於是 $0.8^N = 2^{3N}/10^N$。

- $N = 13$： $0.8^{13} = \dfrac{2^{39}}{10^{13}} = \dfrac{2^9 \cdot 2^{30}}{10^{13}} \approx \dfrac{512 \times 10^9}{10^{13}} = 0.0512 > 0.05$。
- $N = 14$： $0.8^{14} = \dfrac{2^{42}}{10^{14}} = \dfrac{2^2 \cdot 2^{40}}{10^{14}} \approx \dfrac{4 \times 10^{12}}{10^{14}} = 0.04 < 0.05$。

由於 0.8^N 隨 N 增大而遞減，每一個 $N \le 13$ 同樣不符合要求。所以最小的值是

$$N = 14.$$

（這個近似會略為低估 2 的冪，因為 $2^{10} = 1024$；真實數值 $0.8^{13} \approx 0.0550$ 和 $0.8^{14} \approx 0.0440$ 會得出相同的結論。）



當 $N = 14$ 時，臨界值為 $c = 0.05^{1/14} \approx 0.807$，僅僅高於 0.8。

若 $a = 0.8$。 所有的數都落在 $[0, 0.8]$ 內，所以必然有 $X \leq 0.8 < c$。零假設（null hypothesis）被拒絕的概率為 1。

若 $a = 0.9$。 此時 $c \approx 0.807 < 0.9$，所以

$$\mathrm{P}(\text{拒絕}) = \mathrm{P}(X < c \mid a = 0.9) = \left(\frac{c}{0.9}\right)^{14} = \frac{c^{14}}{0.9^{14}} = \frac{0.05}{0.9^{14}}.$$

這裏 $0.9^2 = 0.81$，$0.9^4 = 0.6561$，$0.9^8 \approx 0.4305$，因此 $0.9^{14} = 0.9^8 \cdot 0.9^4 \cdot 0.9^2 \approx 0.2288$。於是

$$\mathrm{P}(\text{拒絕} \mid a = 0.9) = \frac{0.05}{0.9^{14}} \approx 0.219.$$

所以即使有 14 個觀測值，這個檢驗通常也察覺不到 a 是 0.9 而不是 1。

驗算

- 蒙地卡羅模擬（Monte Carlo simulation）（2×10^5 次試驗）在第 (i) 部分得出 $\mathrm{P}(X \leq 0.8) \approx 0.512$ 和 $\mathrm{E}(X) \approx 0.750$。
- 一般而言，$[0, 1]$ 上 n 個均勻分佈的數的最大值，其平均值為 $\frac{n}{n+1}$；當 $n = 1$ 時這是 $\frac{1}{2}$，理應如此，而當 $n = 3$ 時則是 $\frac{3}{4}$。
- 取 $N = 14$ 和 $c = 0.05^{1/14}$，模擬得出的拒絕率在 $a = 1$ 時約為 0.0505（即檢驗的大小），在 $a = 0.9$ 時約為 0.218，與 $0.05/0.9^{14} = 0.2186$ 吻合。

常見錯誤

- 以 $1 - 0.8^N$ 代替 0.8^N：最大值偏小的概率是乘積 0.8^N，而不是它的補數。

- 把 $a = 0.9$ 時的拒絕概率當作 $(0.8/0.9)^{14}$。檢驗是在 $x < c$ 時拒絕，而 c 並不是 0.8。
- 因為 0.0512「大約是 0.05」而停在 $N = 13$。我們需要的是嚴格小於 0.05。

STEP 之外

最大值是一個順序統計量（*order statistic*），而對於 $[0, a]$ 上的均勻分佈，它是 a 的充分統計量（*sufficient statistic*）：它包含了樣本中關於 a 的全部資訊。把拒絕概率看成真實 a 的函數，就是該檢驗的檢定力函數（*power function*）。在這裏，當 $a \leq 1$ 時它是 $\min\{1, 0.05/a^N\}$，而統計學家正是透過比較檢定力函數來在不同的檢驗之間作出選擇。Grimmett 與 Welsh 的 *Probability: An Introduction* 涵蓋了最大值和最小值的分佈；Blitzstein 與 Hwang 的 *Introduction to Probability* 則詳細討論了順序統計量。

詳細解答

STEP I 2004，第 14 題

概率與統計 • 排列的計數

STEP I 2004，第 14 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s1-q14

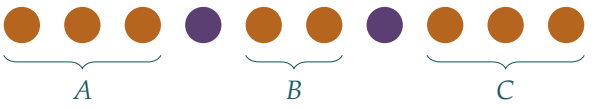
關鍵思路

想像把整個寶箱裏的硬幣逐枚取出，排成一行。海盜其實只是在兩枚鉛幣的位置把這一行切開：第一個海盜得到第一枚鉛幣之前的金幣，第二個海盜得到兩枚鉛幣之間的金幣，第三個海盜得到第二枚鉛幣之後的金幣。硬幣的每一種排列次序出現的可能性都相同，所以整個問題歸結為數算兩枚鉛幣可以落在哪些位置。

解答

設想全部 $n+2$ 枚硬幣按隨機次序逐一抽出，儘管第三個海盜實際上只是把剩下的一手撈走。這並不改變誰得到哪些硬幣。抽取是隨機的，所以從 $n+2$ 個位置中為鉛幣選出兩個位置的 $\binom{n+2}{2}$ 種選法，每一種的可能性都相同。

設 A、B 和 C 分別為第一、第二和第三個海盜取得的金幣數目。那麼 $A+B+C=n$，而這一行硬幣看起來是這樣的：



圖中琥珀色的是金幣，紫紅色的是鉛幣。鉛幣的位置決定了 (A,B,C)，而每一個滿足 $A+B+C=n$ 的非負整數三元組，都恰好來自一種位置選法。所以這 $\binom{n+2}{2}=\frac{1}{2}(n+1)(n+2)$ 個三元組 (A,B,C) 的可能性全部相同。

第 (i) 部分

第一個海盜得不到金幣，當且僅當 $A = 0$，即抽出的第一枚硬幣就是鉛幣。由於 $n+2$ 枚硬幣中每一枚最先被抽出的可能性都相同，這個事件的概率（probability）是 $\frac{2}{n+2}$。因此

$$\mathrm{P}(A \geq 1) = 1 - \frac{2}{n+2} = \frac{n}{n+2}.$$

（用數算的方法也得到相同答案：$A = 0$ 時餘下 $B + C = n$，它有 $n+1$ 個解，而 $(n+1)\big/\frac{1}{2}(n+1)(n+2) = \frac{2}{n+2}$。）

第 (ii) 部分

> **想一想** 第二個海盜是在一個隨機的時刻才開始抽的。她的處境真的不比第一個海盜差嗎？

是的，她的處境並不較差。我們數算 $B = 0$ 的三元組：這時 $A + C = n$，它有 $n+1$ 個解，與第 (i) 部分完全一樣。所以

$$\mathrm{P}(B \geq 1) = 1 - \frac{n+1}{\frac{1}{2}(n+1)(n+2)} = \frac{n}{n+2}.$$

事實上，由於所有三元組 (A, B, C) 的可能性都相同，三人所得的份額是**可交換的**（exchangeable）：每個海盜得到金幣的機會都同樣是 $\frac{n}{n+2}$，連第三個海盜也不例外。

第 (iii) 部分

我們需要 A、B 和 C 全部至少為 1。設 $A = A' + 1$、$B = B' + 1$ 和 $C = C' + 1$。於是我們想要的三元組就對應於 $A' + B' + C' = n - 3$ 的非負整數解。根據隔板法（stars and bars），這樣的解共有 $\binom{n-1}{2}$ 個，這一點也可以直接看出：在相鄰金幣之間的 $n-1$ 個空隙中，選出兩個來放鉛幣。因此，對於 $n \geq 1$，

$$\mathrm{P}(A \geq 1,\ B \geq 1,\ C \geq 1) = \frac{\binom{n-1}{2}}{\binom{n+2}{2}} = \frac{(n-1)(n-2)}{(n+1)(n+2)}.$$

當 $n = 1$ 和 $n = 2$ 時，這個概率是 0，這是必然的，因為金幣不足三枚時，三個海盜不可能人人都有金幣。（當 $n = 0$ 時，概率顯然是 0；公式在這裏並不適用，因為 $\binom{n-1}{2}$ 所數算的空隙根本不存在。）

驗算

- 當 $n = 1$ 時，硬幣的排列有 GLL、LGL 和 LLG 三種，可能性全部相同。第一個海盜只在 GLL 中得到金幣，其概率為 $\frac{1}{3} = \frac{1}{1+2}$；第二個海盜只在 LGL 中得到金幣，概率同樣是 $\frac{1}{3}$。

- 當 $n = 3$ 時，唯一合乎要求的排列是 GLGLG，它是 $\binom{5}{2} = 10$ 種排列中的一種，而公式給出 $\frac{2\cdot 1}{4\cdot 5} = \frac{1}{10}$。
- 對 $n = 1, \ldots, 14$ 逐一窮舉，結果與三條公式全部吻合。以 $n = 5$ 模擬 2×10^5 個寶箱，得出 0.715、0.714 和 0.286，而理論值分別為 $\frac{5}{7} \approx 0.714$、$\frac{5}{7}$ 和 $\frac{12}{42} = \frac{2}{7} \approx 0.286$。

常見錯誤

- 把第二個海盜的抽取當作從一個裝滿的寶箱重新開始。留給她的金幣有多少，取決於第一個海盜抽得怎樣。
- 在第 (iii) 部分把 $\mathrm{P}(A \geq 1)$、$\mathrm{P}(B \geq 1)$ 和 $\mathrm{P}(C \geq 1)$ 相乘。各人的份額並非獨立（independent），因為它們的總和必須是 n。

STEP 之外

鉛幣位置與 $A + B + C = n$ 的解之間的對應，就是「隔板法」的雙射（bijection）。它說明了隨機放置的標記之間的間隔是可交換的。它的連續版本，是把點均勻地投在一根棒上，由此引出均勻間距（uniform spacings）理論和狄利克雷分佈（Dirichlet distribution），而這類次序統計量（order statistics）在大學一年級的概率論中隨處可見。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）對這種數算方法以及相關的對稱性論證，都有十分出色的講解。

詳細解答

STEP II 2004，第 1 題

純數學 • 含根式的方程

> **STEP II 2004，第 1 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q1

關鍵思路

每一條方程都含有 $\sqrt{x}$ 和 $\sqrt{3x^2+1}$，所以設 $t=\sqrt{x}$（其中 $t\geq 0$），把前者化為多項式（polynomial）。然後把餘下的平方根單獨移到一邊，平方一次。這樣會得到一條關於 t 的四次方程（quartic），而且很容易因式分解（factorise），因為 $t=0$ 和 $t=1$ 必定是根。真正要花工夫的是最後一步：平方可能引入增根（false roots），所以我們只保留那些令被平方的一邊不是負數的根。

解答

三條方程都需要 $x\geq 0$，$\sqrt{x}$ 才是實數，所以我們寫 $x=t^2$，其中 $t\geq 0$。於是 $\sqrt{x}=t$，而 $3x^2+1=3t^4+1$。

平方這一步只在一個方向上成立。對於實數 A 和 B，

$$\sqrt{A}=B \iff A=B^2 \text{ 且} B\geq 0.$$

所以平方之後，我們必須檢查右邊在每一個根處是否非負。

第 (i) 部分

把根式單獨移到一邊，得

$$\sqrt{3t^4+1}=2t^2-t+1.$$

右邊恆為正數：它的判別式(discriminant)是 $1-8<0$，而首項係數(leading coefficient)是正數。所以在這裏平方不會有任何損失，平方後所得方程的每一個根都是真正的解。

平方得

$$3t^4+1=(2t^2-t+1)^2=4t^4-4t^3+5t^2-2t+1,$$
$$0=t^4-4t^3+5t^2-2t=t\,(t^3-4t^2+5t-2).$$

這個三次式在 $t=1$ 時等於零（因為 $1-4+5-2=0$），除去因式 $t-1$ 後得

$$t^3-4t^2+5t-2=(t-1)(t^2-3t+2)=(t-1)^2(t-2).$$

所以 $t\,(t-1)^2(t-2)=0$，得 $t=0,1,2$。三個值都是非負的，所以

$$x=0,\quad x=1,\quad x=4.$$

驗算一下：當 $x=4$ 時，我們得到 $\sqrt{49}+2-8-1=0$。

第 (ii) 部分

現在我們有

$$\sqrt{3t^4+1}=-t^2+2t+1.$$

平方得

$$3t^4+1=(-t^2+2t+1)^2=t^4-4t^3+2t^2+4t+1,$$
$$0=2t^4+4t^3-2t^2-4t=2t\,(t^3+2t^2-t-2).$$

把三次式的各項分組為 $t^2(t+2)-(t+2)$，得

$$2t\,(t+2)(t-1)(t+1)=0.$$

由於 $t\geq 0$，只剩下 $t=0$ 和 $t=1$。右邊 $-t^2+2t+1$ 在這兩處的值分別是 1 和 2，都是正數，所以兩個根都是真正的解：

$$x=0,\quad x=1.$$

第 (iii) 部分

這一次

$$\sqrt{3t^4+1}=t^2+2t-1.$$

平方得

$$3t^4+1=(t^2+2t-1)^2=t^4+4t^3+2t^2-4t+1,$$
$$0=2t^4-4t^3-2t^2+4t=2t\,(t^3-2t^2-t+2).$$

把三次式分組為 $t^2(t-2)-(t-2)$，得

$$2t\,(t-2)(t-1)(t+1)=0.$$

非負的根是 $t=0,1,2$。

想一想 這三個值是否全都滿足原來的方程？

不是。當 $t=0$ 時，右邊 t^2+2t-1 等於 -1，但平方根不可能是負數。所以 $t=0$ 是增根，是由平方引入的。（直接驗證：當 $x=0$ 時，(iii) 的左邊是 $1-0-0+1=2\neq 0$。）當 $t=1$ 和 $t=2$ 時，右邊分別是 2 和 7，都是正數，所以這兩個根是真正的解。因此

$$x=1,\quad x=4.$$

(ii) 和 (iii) 之間有一個巧妙的聯繫。把 t 換成 $-t$，$-t^2+2t+1$ 就變成 $-(t^2+2t-1)$，而平方會消去負號。所以 (iii) 中平方後的方程，就是把 (ii) 中平方後的方程的 t 換成 $-t$ 所得的方程，它的根 $0,\pm1,2$ 正是 (ii) 的根 $0,\pm1,-2$ 的相反數。

驗算

把每個答案代入原來的方程。例如，在 (iii) 中代入 $x=4$，得 $\sqrt{49}-4-4+1=0$。在 $0\leq x\leq 50$ 的範圍內對每條方程的左邊作細密的數值掃描，沒有發現其他零點。在 (i) 中，左邊在 $x=1$ 處觸及零但不變號，這與二重因式 $(t-1)^2$ 吻合。

常見錯誤

- 忘記檢查被平方那一邊的正負號，因而在 (iii) 中保留了 $x=0$。
- 除以 t（或除以 $\sqrt{x}$），因而在 (i) 和 (ii) 中漏掉了 $x=0$ 這個根。
- 一邊仍有兩個平方根時就平方。應先把 $\sqrt{3x^2+1}$ 單獨移到一邊，這樣平方一次便能把它消去。

STEP 之外

映射（map）$A\mapsto A^2$ 在 $\mathbb{R}$ 上不是單射（injective），所以兩邊平方得到的只是蘊涵關係（implication），而不是等價關係（equivalence）。留意哪些步驟可逆，是每一個嚴謹解方程過程背後的習慣；到了分析學（analysis），這就變成「函數在甚麼時候有反函數（inverse）、在哪個定義域（domain）上有反函數」的問題。Bartle 與 Sherbert 的 *Introduction to Real Analysis*（《實分析導論》）對此有細緻的闡述。

詳細解答

STEP II 2004，第 2 題

純數學 • 二次不等式

> **STEP II 2004，第 2 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q2

關鍵思路

絕對值（modulus）是唯一棘手的地方，而只要設 $t = |x|$，它就會消失。由於 $x^2 = |x|^2$，原不等式便化為一個關於 t 的普通二次不等式（quadratic inequality），並附帶條件 $t \geq 0$。接着只要配方（completing the square），或者考察判別式（discriminant），一切便一目了然。t 的每一個容許區間都對應 x 的兩個區間，分別位於原點兩側。

解答

以下一律設 $t = |x| \geq 0$。由於 $x^2 = t^2$，不等式 $(*)$ 正好等價於

$$q(t) = t^2 - \alpha t + 2 < 0, \qquad t = |x| \geq 0.$$

當 $|\alpha| < 2\sqrt{2}$ 時無解

配方得

$$q(t) = \left(t - \frac{\alpha}{2}\right)^2 + 2 - \frac{\alpha^2}{4}.$$

若 $|\alpha| < 2\sqrt{2}$，則 $\alpha^2 < 8$，故 $2 - \frac{\alpha^2}{4} > 0$。平方項永不為負，所以對任何實數 t 都有 $q(t) > 0$，對 $t = |x|$ 自然也成立。因此對所有 x 均有 $x^2 - \alpha|x| + 2 > 0$，沒有任何 x 的值滿足 $(*)$。

$\alpha = 3$ 的情況

此時 $q(t) = t^2 - 3t + 2 = (t-1)(t-2)$，它恰好在 $1 < t < 2$ 時為負。所以 $(*)$ 成立當且僅當 $1 < |x| < 2$，解集為

$$\{x : -2 < x < -1\} \cup \{x : 1 < x < 2\}.$$

長度之和 S

設 $\alpha > 2\sqrt{2}$。則 $\alpha^2 > 8$，故 q 有兩個相異實根

$$t_1 = \frac{\alpha - \sqrt{\alpha^2 - 8}}{2}, \qquad t_2 = \frac{\alpha + \sqrt{\alpha^2 - 8}}{2},$$

而且由於 q 的首項係數為正，$q(t) < 0$ 恰好在 $t_1 < t < t_2$ 時成立。

> **想一想** 區間 $t_1 < t < t_2$ 會不會有一部分落在 t 為負的範圍？我們需要擔心這一點嗎？

不需要。兩根之積為 $t_1 t_2 = 2 > 0$，兩根之和為 $t_1 + t_2 = \alpha > 0$，所以兩根皆為正數。因此整個區間 (t_1, t_2) 都位於 $t > 0$ 之內，其中每一個 t 值都確實是 $|x|$ 可取的值。

所以 $(*)$ 成立當且僅當 $t_1 < |x| < t_2$，即

$$-t_2 < x < -t_1 \qquad \text{或} \qquad t_1 < x < t_2.$$

這是兩個互不相交的區間（因為 $t_1 > 0$），長度各為 $t_2 - t_1 = \sqrt{\alpha^2 - 8}$。因此

$$S = 2\sqrt{\alpha^2 - 8}.$$

由於 $\alpha^2 - 8 < \alpha^2$ 且 $\alpha > 0$，取正平方根得 $\sqrt{\alpha^2 - 8} < \alpha$，所以

$$S < 2\alpha,$$

這正是所要證明的。

S 對 α 的圖像

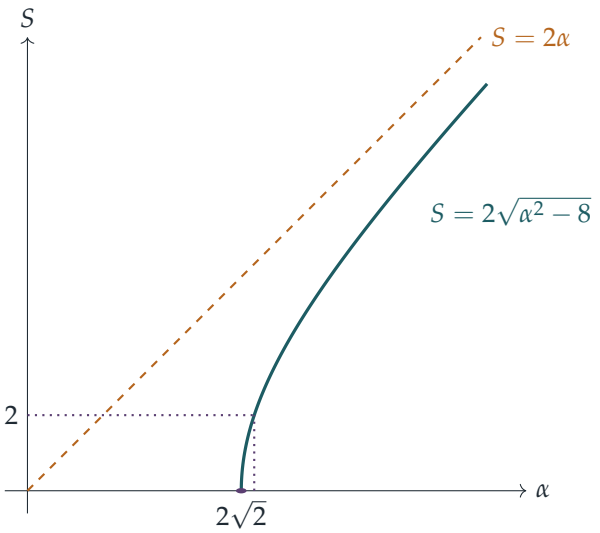
函數 $S = 2\sqrt{\alpha^2 - 8}$ 的定義域為 $\alpha > 2\sqrt{2}$（當 $\alpha \le 2\sqrt{2}$ 時根本沒有任何區間：在 $\alpha = 2\sqrt{2}$ 時有 $q(t) = (t - \sqrt{2})^2 \ge 0$，而只要 $\alpha \le 0$，對所有 $t \ge 0$ 都有 $q(t) \ge 2$）。它的主要特徵如下：

- 當 $\alpha \to 2\sqrt{2}^{+}$ 時 $S \to 0$，而 $\dfrac{\mathrm{d}S}{\mathrm{d}\alpha} = \dfrac{2\alpha}{\sqrt{\alpha^2 - 8}}$ 在該處趨向無窮大：曲線由 $(2\sqrt{2}, 0)$ 出發，該點的切線是鉛垂的。
- S 是遞增的，而且 $\dfrac{\mathrm{d}^2 S}{\mathrm{d}\alpha^2} = -\dfrac{16}{(\alpha^2 - 8)^{3/2}} < 0$，所以曲線是凹的（concave）。

- 直線 $S = 2\alpha$ 是漸近線（asymptote），曲線從下方趨近它：由於 $(2\alpha - S)(2\alpha + S) = 4\alpha^2 - 4(\alpha^2 - 8) = 32$，

$$2\alpha - S = \frac{32}{2\alpha + S} \to 0 \quad \text{當} \alpha \to \infty.$$

事實上 $\frac{\alpha^2}{8} - \frac{S^2}{32} = 1$，所以這個圖像是雙曲線(hyperbola)的四分之一，漸近線為 $S = 2\alpha$。



驗算

- 當 $\alpha = 3$ 時，公式給出 $S = 2\sqrt{1} = 2$，與 $(-2, -1)$ 和 $(1, 2)$ 這兩個長度各為 1 的區間吻合（即圖中以虛點線標示、位於 $\alpha = 3$ 的那一點）。
- 當 $\alpha = 5$ 時，$S = 2\sqrt{17} \approx 8.246$；以數值方法量度 $[-20, 20]$ 內滿足 $x^2 - 5|x| + 2 < 0$ 的 x 所組成的集合，所得的值相同，而且明顯小於 $2\alpha = 10$。
- 當 α 很大時，$2\alpha - S \approx 8/\alpha$：在 $\alpha = 100$ 時，兩者相差 $0.0800\ldots$。

常見錯誤

- 忘記 $|x|$ 的每一個區間對應 x 的兩個區間，因而得出 $S = \sqrt{\alpha^2 - 8}$。
- 把 $\alpha = 3$ 的答案只寫成 $1 < x < 2$，漏掉了負的那一半。
- 使用判別式時，沒有指出關於 t 的兩個根皆為正數；當 $\alpha < 0$ 時，同樣的判別式條件卻完全沒有解，因為兩個根都是負數。

STEP 之外

函數 $x \mapsto x^2 - a|x| + 2$ 是偶函數（even function），所以它的解集是對稱的，而長度 S 就是該函數取負值的集合的測度（measure）。量度由不等式所定義的集合的大小，正是測度論（measure theory）和勒貝格積分（Lebesgue integration）的起點，大學二年級的分析課程會學到這些內容。若想細讀本題用到的極限、導數和不等式的嚴謹論述，可參閱 Abbott 的 *Understanding Analysis*（一本實分析入門書）。

詳細解答

STEP II 2004，第 3 題

純數學 • 曲線描繪

> **STEP II 2004，第 3 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q3

關鍵思路

一切都源自同一條曲線。對於 C，先把導數(derivative)徹底因式分解：重複因式 $(x-2)$ 在求導之後仍然保留，餘下的只是一個二次式。其餘兩條曲線都是由 C 建構出來的，毋須從頭做起：在 (i) 中取 $y = \pm\sqrt{f(x)}$，在 (ii) 中則把 x^2 代入 f。要留意 f 在哪裏取正值、在哪裏有單根（simple root）、在哪裏有重根（repeated root）；曲線的形狀正是由這些細節決定的。

解答

設 $f(x) = x(x+1)(x-2)^4$，則 C 就是 $y = f(x)$。

C 的駐點

利用乘積法則（product rule），並抽出公因式 $(x-2)^3$，

$$\begin{aligned} f'(x) &= (2x+1)(x-2)^4 + 4x(x+1)(x-2)^3 \\ &= (x-2)^3\left[(2x+1)(x-2) + 4x^2 + 4x\right] \\ &= (x-2)^3(6x^2 + x - 2) \\ &= (x-2)^3(3x+2)(2x-1). \end{aligned}$$

所以駐點（stationary point）位於 $x = -\frac{2}{3}$、$x = \frac{1}{2}$ 和 $x = 2$，而

$$f\left(-\tfrac{2}{3}\right) = \left(-\tfrac{2}{3}\right)\left(\tfrac{1}{3}\right)\left(-\tfrac{8}{3}\right)^4 = -\frac{8192}{729}, \qquad f\left(\tfrac{1}{2}\right) = \tfrac{1}{2}\cdot\tfrac{3}{2}\cdot\left(-\tfrac{3}{2}\right)^4 = \frac{243}{64}, \qquad f(2) = 0.$$

要判別這些駐點的類型，可考察 f' 各因式的正負號：

	$x < -\frac{2}{3}$	$-\frac{2}{3} < x < \frac{1}{2}$	$\frac{1}{2} < x < 2$	$x > 2$
$(x-2)^3$	$-$	$-$	$-$	$+$
$(3x+2)(2x-1)$	$+$	$-$	$+$	$+$
$f'(x)$	$-$	$+$	$-$	$+$

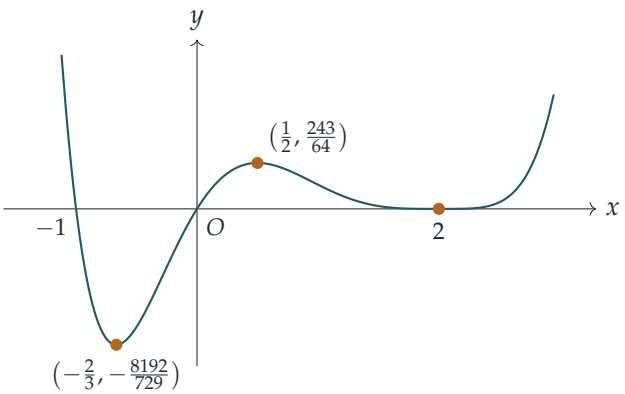
因此

- $(-\frac{2}{3}, -\frac{8192}{729})$ 是極小點（minimum）（f' 由 $-$ 變為 $+$）；
- $(\frac{1}{2}, \frac{243}{64})$ 是極大點（maximum）（f' 由 $+$ 變為 $-$）；
- $(2,0)$ 是極小點（f' 由 $-$ 變為 $+$）。

最後一個即使不用微積分也顯而易見：在 $x = 2$ 附近 $f(x) \geq 0$，因為在那裏 $x(x+1) > 0$，而 $(x-2)^4 \geq 0$。

描繪 C

曲線在 $x = -1$ 和 $x = 0$ 處穿過 x 軸（兩者都是單根，斜率分別為 $f'(-1) = -81$ 和 $f'(0) = 16$），並在 $x = 2$ 處與該軸相切。在 $x = 2$ 附近，$f(x) \approx 2 \cdot 3\,(x-2)^4 = 6(x-2)^4$，所以曲線在那裏非常平坦。由於 f 是一個 6 次多項式，而且首項係數為正，$y \to +\infty$ 當 $x \to \pm\infty$。草圖中的縱軸比例經過壓縮。



第 (i) 部分

現在 $y^2 = f(x)$，所以 $y = \pm\sqrt{f(x)}$。以下三點足以確定曲線的形狀。

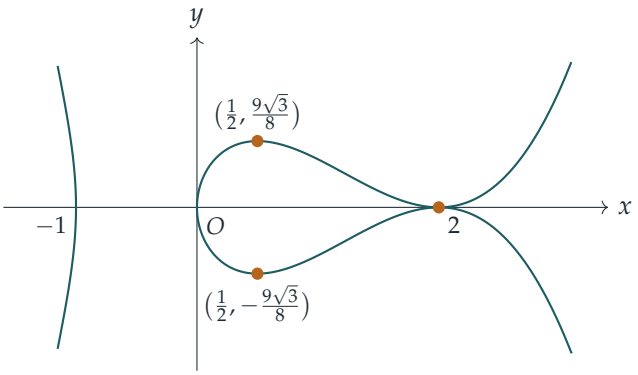
- 曲線存在的範圍。我們需要 $f(x) \geq 0$，由 C 可知這表示 $x \leq -1$ 或 $x \geq 0$。在 $-1 < x < 0$ 的範圍內沒有曲線。
- 對稱性。把 y 換成 $-y$，方程不變，所以曲線關於 x 軸對稱。只需描繪上半部分 $y = \sqrt{f(x)} = (x-2)^2\sqrt{x(x+1)}$，再作反射即可。

- 駐點。對於 $y=\sqrt{f}$，當 $f>0$ 時有 $\mathrm{d}y/\mathrm{d}x = f'/(2\sqrt{f})$，所以上半部分的駐點恰好出現在 $f'=0$ 且 $f>0$ 之處。這樣只剩下 $x=\frac{1}{2}$（$x=-\frac{2}{3}$ 這一點在定義域之外），於是在 $\left(\frac{1}{2},\sqrt{243/64}\right) = \left(\frac{1}{2},\frac{9\sqrt{3}}{8}\right)$ 有一個極大點，而在下半部分的 $\left(\frac{1}{2},-\frac{9\sqrt{3}}{8}\right)$ 有一個極小點。

想一想 在 $f=0$ 的三點 $(-1,0)$、$(0,0)$ 和 $(2,0)$，曲線會怎樣？

在單根處，例如 $x=0$，$f(x)\approx 16x$，所以在原點附近 $y^2\approx 16x$：這是一條橫臥的拋物線，曲線穿過橫軸時的切線是鉛垂的。在 $x=-1$ 處情況相同，那裏 $y^2\approx 81(-1-x)$。在重根 $x=2$ 處，$y^2\approx 6(x-2)^4$，所以 $y\approx\pm\sqrt{6}\,(x-2)^2$：這是兩段形如拋物線的弧，一段在橫軸之上，一段在橫軸之下，兩者都在 $(2,0)$ 與 x 軸相切。因此上下兩半在該處相遇，並有共同的水平切線；在 $(2,0)$ 處，上半部分取得極小值，下半部分取得極大值。

當 $|x|$ 很大時，$y\approx\pm|x|^3$。所以曲線包括一個由 $(0,0)$ 到 $(2,0)$ 的閉合環，它在 $(2,0)$ 處收窄成一點，然後延伸為兩條向右張開的分支；另外還有獨立的一段，經過 $(-1,0)$ 並向左張開。



第 (ii) 部分

這裏 $y=x^2(x^2+1)(x^2-2)^4=f(x^2)$。以下兩點觀察已解決了大部分問題。

- 曲線關於 y 軸對稱，因為 x 只以 x^2 的形式出現。
- 當 x 走遍 $[0,\infty)$ 時，$u=x^2$ 亦遞增地走遍 $[0,\infty)$，所以對於 $x\geq 0$，這條曲線就是 C 中 $x\geq 0$ 的部分，只是重新參數化而已。特別是處處都有 $y\geq 0$，而 C 中對應負 x 的一段（包括其極小點）完全不起作用。

由鏈式法則（chain rule），

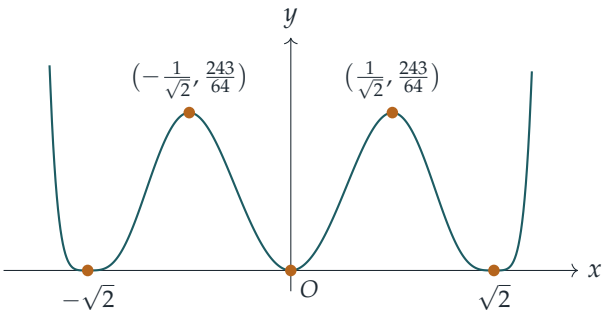
$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2x\,f'(x^2) = 2x\,(x^2-2)^3(3x^2+2)(2x^2-1),$$

當 $x=0$、$x^2=\frac{1}{2}$ 或 $x^2=2$ 時，上式等於零（因式 $3x^2+2$ 永不為零）。在 $u\geq 0$ 上，f

由 0 上升至 $u = \frac{1}{2}$ 處的極大值，再下降至 $u = 2$ 處的 0，之後一直上升；而對於 $x \geq 0$，$u = x^2$ 保持這個次序。所以駐點為

- 位於 $(0,0)$ 的極小點，該處 $y \approx 16x^2$；
- 位於 $\left(\pm\frac{1}{\sqrt{2}}, \frac{243}{64}\right)$ 的極大點；
- 位於 $(\pm\sqrt{2}, 0)$ 的極小點，曲線在該處非常平坦地與橫軸相切：$y \approx 384\,(x \mp \sqrt{2})^4$。

當 $|x|$ 很大時，$y \approx x^{12}$，上升得極為陡峭。



驗算

先對導數作一個簡單的合理性檢查：在 $x = 0$ 處，$f(x) = x(x+1)(x-2)^4 \approx 16x$，而確實有 $f'(0) = (-8)(2)(-1) = 16$。數值驗算方面，在隨機選取的點上把 f' 的因式分解與有限差分（finite difference）導數比較，結果吻合；又得出 $f(-\frac{2}{3}) \approx -11.237$ 和 $f(\frac{1}{2}) = 3.796875$，並求得 $\sqrt{f}$ 在 $[0,2]$ 上的最大值為 $1.94856\ldots = \frac{9\sqrt{3}}{8}$，在 $x = 0.5$ 處取得。至於 (ii)，$\mathrm{d}y/\mathrm{d}x$ 在 $x \approx \pm 0.707$ 和 ± 1.414（以及 0）處變號，與預期一致。

常見錯誤

- 先把 f 展開才求導。這樣 f' 會是一個看不出任何因式的五次式；把 $(x-2)^3$ 留在括號外，就只需解一個二次方程。
- 在 (i) 中，於 $(0,0)$ 和 $(-1,0)$ 畫出尖點（cusp）或折角。在 f 的單根處，$y^2 = f(x)$ 的切線是鉛垂的。
- 在 (ii) 中，照搬 C 在 $x = -\frac{2}{3}$ 處的極小點。由於 $x^2 \geq 0$，只會用到 C 中 $x \geq 0$ 的部分，而 y 永不為負。

STEP 之外

$y^2 = f(x)$ 在 f 的根附近的表現，正是研究代數曲線（algebraic curve）奇點（singular point）的起點：單根給出一個帶鉛垂切線的光滑點，而重根則給出兩條分支相遇之處，如 $(2,0)$ 的情況。當 f 是三次式時，曲線 $y^2 = f(x)$ 稱為橢圓曲線（*elliptic curves*），是現代數論的核心對象。若想仔細了解曲線描繪背後的微積分（單調性（monotonicity）、導數的正負號和鏈式法則），可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP II 2004，第 4 題

純數學 • 幾何中的最優化

STEP II 2004，第 4 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

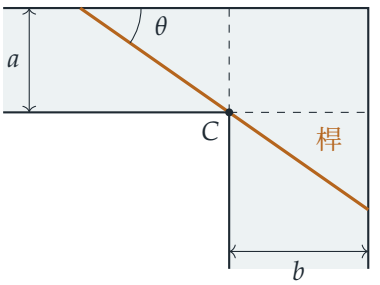
step.maths.org/questions/04-s2-q4

關鍵思路

無論怎樣搬，桿（或桌子）都要由指向一條走廊的方向，轉到指向另一條走廊的方向，所以在某一刻，它與牆壁所成的角必會取遍兩者之間的每一個角 θ。對固定的角度而言，空間最寬裕的位置就是剛好擦過內角的那條直線，而這段長度用直角三角形（right-angled triangle）很容易算出。因此，物件能夠轉過彎角，當且僅當它的長度不超過這些長度在所有角度之中的最小者，而微積分（calculus）正好可以求出這個最小值（minimum）。

解答

以下一律把外角放在原點（origin），外牆分別沿 $y = 0$（闊度為 a 的走廊，向左延伸）和 $x = 0$（闊度為 b 的走廊，向下延伸）。於是內角位於 $C = (-b, -a)$。



第 (i) 部分

假設在某一刻，桿與 a 走廊的牆壁成銳角（acute angle）θ，並如圖所示由外牆 $y = 0$ 向下斜向外牆 $x = 0$。

*角度爲 θ 時最長的線段。*取通過 C 而傾角為 θ 的直線。由 C 向上到牆 $y = 0$，垂直升高了 a，所以這一段的長度是 $a\operatorname{cosec}\theta$；由 C 橫向到牆 $x = 0$，水平移動了 b，所以這一段的長度是 $b\sec\theta$。因此，這條直線夾在兩道外牆之間的弦（chord）的長度為

$$f(\theta) = a\operatorname{cosec}\theta + b\sec\theta.$$

任何其他傾角為 θ 而由一條走廊伸入另一條走廊的直線，都必須在 C 的外側經過（否則它會穿過內牆）。這樣的直線與兩道外牆圍出一個較小的相似（similar）直角三角形，所以它在外牆之間的弦比 $f(\theta)$ 短。另外，假如此刻桿完全位於 a 走廊之內，它就容納在一條闊度為 a 的長帶裏，所以 $L\sin\theta \le a$，由此得 $L \le a\operatorname{cosec}\theta < f(\theta)$；同理，假如它完全位於 b 走廊之內，則 $L \le b\sec\theta < f(\theta)$。無論哪一種情況，都有

$$L \le f(\theta) \quad \text{對桿所經過的每一個角 } \theta \text{ 均成立。}$$

當桿由沿着第一條走廊轉到沿着第二條走廊時，θ 由接近 0 連續地變到接近 $\frac{1}{2}\pi$，所以它會取遍其間的每一個值。因此 $L \le f(\theta)$ 對每一個 $\theta \in (0, \frac{1}{2}\pi)$ 都成立，也就是說，L 不超過 f 的最小值。

*求 f 的最小值。*求導（differentiate）得

$$f'(\theta) = -\frac{a\cos\theta}{\sin^2\theta} + \frac{b\sin\theta}{\cos^2\theta} = \frac{b\sin^3\theta - a\cos^3\theta}{\sin^2\theta\cos^2\theta}.$$

這個式子等於零，當且僅當 $\tan^3\theta = a/b$；由於 tan 在 $(0, \frac{1}{2}\pi)$ 上遞增，這個方程在該區間內只有一個解 α。f' 的正負號與 $b\tan^3\theta - a$ 相同，在 α 之前為負，之後為正，所以 α 給出最小值。（此外，在區間兩端均有 $f \to \infty$。）在 $L \le f(\theta)$ 中取 $\theta = \alpha$，得

$$L \le a\operatorname{cosec}\alpha + b\sec\alpha, \qquad \tan^3\alpha = \frac{a}{b},$$

這正是所要證明的。

第 (ii) 部分

現在由桌子的長邊擔當桿的角色。桌子的內側長邊位於一條傾角為 θ 的直線上，外側長邊則位於與之平行、向外角方向再遠 w 的直線上。桌子外側的兩個角都必須位於外牆之內。

*外側長邊可用的空間。*設一條傾角為 θ 的直線與外角 O 的距離為 d，它在 O 處截出一個直角三角形，其斜邊（hypotenuse）為 c（即兩牆之間的弦），兩條直角邊為 $c\cos\theta$ 和

$c\sin\theta$，由 O 所作的高為 d。這個三角形面積的兩倍既等於 $c\cos\theta\cdot c\sin\theta$，又等於 cd，所以

$$c=\frac{d}{\sin\theta\cos\theta}=2d\operatorname{cosec}2\theta.$$

可見弦長與直線到 O 的距離成正比。因此，把直線向內（即向 O 的方向）移動 w，它的弦便會縮短 $2w\operatorname{cosec}2\theta$。

想一想 為甚麼外側長邊必須容納在兩道外牆之間，而且是在比通過 C 的直線再靠內 w 的那條直線上？

內側長邊不能越過內角，所以它位於通過 C 的直線上，或在該直線的外側。於是外側長邊所在的直線至少再向外 w，其弦長至多為 $f(\theta)-2w\operatorname{cosec}2\theta$。（兩條短邊由外側的兩個角向內伸展，與長邊垂直，並且留在走廊之內，所以不會帶來額外的限制。）假如整張桌子都位於 a 走廊之內，則它橫跨長帶的高度滿足 $l\sin\theta+w\cos\theta\le a$。由此得出的上界 $l\le(a-w\cos\theta)\operatorname{cosec}\theta$ 比下文的 $g(\theta)$ 小了 $(b-w\sin\theta)\sec\theta>0$，這是因為 $w<b$。b 走廊的情況亦相同。所以，與第 (i) 部分完全一樣，要成功搬過彎角，必須有 $l\le g(\theta)$ 對所有 $\theta\in(0,\frac12\pi)$ 成立，其中

$$g(\theta)=a\operatorname{cosec}\theta+b\sec\theta-2w\operatorname{cosec}2\theta.$$

（桌子必須比兩條走廊都窄，即 $w<a$ 且 $w<b$，否則它連沿走廊移動也辦不到。）

求 g 的最小值。由於 $\dfrac{\mathrm{d}}{\mathrm{d}\theta}\operatorname{cosec}2\theta=-\dfrac{2\cos2\theta}{\sin^2 2\theta}$，

$$g'(\theta)=-\frac{a\cos\theta}{\sin^2\theta}+\frac{b\sin\theta}{\cos^2\theta}+\frac{4w\cos2\theta}{\sin^2 2\theta}.$$

兩邊乘以 $\sin^2\theta\cos^2\theta=\frac14\sin^2 2\theta>0$。於是 $g'(\theta)=0$ 當且僅當

$$-a\cos^3\theta+b\sin^3\theta+w\cos2\theta=0,\qquad\text{即}\qquad w\cos2\theta=\cos^3\theta\,(a-b\tan^3\theta).$$

利用 $\cos2\theta=\cos^2\theta\,(1-\tan^2\theta)$ 改寫，再除以 $\cos^2\theta$。假如 $\theta=\frac14\pi$，方程便會變成 $0=\frac{1}{2\sqrt2}(a-b)$，但 $a\ne b$，所以這不可能。因此在駐點（stationary point）處 $\tan^2\theta\ne1$，我們可以除以 $1-\tan^2\theta$，得到

$$w=\left(\frac{a-b\tan^3\theta}{1-\tan^2\theta}\right)\cos\theta.$$

當 $\theta\to0^+$ 時，$g(\theta)\approx(a-w)/\theta\to+\infty$，因為 $w<a$；同理，當 $\theta\to\frac12\pi^-$ 時 $g\to+\infty$，因為 $w<b$。所以連續（continuous）函數 g 在區間內部的某一點取得最小值，而在該點 $g'=0$。把這一點記為 β；它滿足上面列出的方程。在 $l\le g(\theta)$ 中取 $\theta=\beta$，得

$$l\le a\operatorname{cosec}\beta+b\sec\beta-2w\operatorname{cosec}2\beta,$$

這正是所要證明的。（事實上，對方程的任何一個根 β 都有 $l\le\min g\le g(\beta)$，所以題目指的是哪一個根並無影響。）

驗算

- 在第 (ii) 部分中取 $w = 0$，便得 $\tan^3\beta = a/b$ 以及第 (i) 部分的上界，與預期相符。
- 在第 (i) 部分中，$\sin\alpha = a^{1/3}/\sqrt{a^{2/3}+b^{2/3}}$ 而 $\cos\alpha = b^{1/3}/\sqrt{a^{2/3}+b^{2/3}}$，所以上界是 $\left(a^{2/3}+b^{2/3}\right)^{3/2}$。當 $a = 1$，$b = 8$ 時，這個值是 $5^{3/2} \approx 11.18$，而以數值方法搜尋 f 的最小值，結果亦與此吻合。
- 我們隨機選取了數百組 $a \neq b$ 和 $w < \min(a,b)$，以窮舉法求出 g 的最小值，並確認取得最小值的角度滿足關於 β 的方程。我們亦直接用坐標核對了各弦的長度。

常見錯誤

- 求了 f 的最大值而不是最小值。桿必須捱得過最差的角度，所以上界是可用長度的最小值。
- 以為受外牆限制的是桌子的內側長邊。其實被擠壓的是外側長邊，它位於距離外角近 w 的那條直線上。
- 除以 $1-\tan^2\beta$ 時，沒有留意到這一步需要 $a \neq b$。

STEP 之外

第 (i) 部分是經典的「梯子轉彎」問題，其答案 $\left(a^{2/3}+b^{2/3}\right)^{3/2}$ 與星形線（astroid）有關：通過內角的各條直線掃出一條包絡線（envelope）。第 (ii) 部分則是搬沙發問題（moving sofa problem）的一個小特例：在闊度為一個單位的直角彎角中，能夠搬過去的圖形之中，面積最大的是哪一個？這個問題由 Leo Moser 於 1966 年提出，在其後的大部分歲月裏一直未能完全解決。如果想做更多這類幾何與微積分交匯的題目，可以試試 Siklos 的 *Advanced Problems in Mathematics*。

詳細解答

STEP II 2004，第 5 題

純數學 • 積分方程

> **STEP II 2004，第 5 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q5

關鍵思路

未知函數藏在積分(integral)之內，看起來很嚇人，但只要用複角公式(compound-angle formula) 把 $\sin(x+t)$ 展開，t 就可以移到積分之外，而留在積分內的只是兩個固定的數。因此 f 必定是 t 加上 $\sin t$ 與 $\cos t$ 的一個組合，整個問題於是歸結為由兩條線性方程求出兩個常數。第一部分的兩個積分，正是那條方程所需要的。

解答

兩個積分

運用分部積分法（integration by parts），取 $u = x$ 及 $\mathrm{d}v = \sin x\,\mathrm{d}x$：

$$\int_0^\pi x\sin x\,\mathrm{d}x = \Big[-x\cos x\Big]_0^\pi + \int_0^\pi \cos x\,\mathrm{d}x = (-\pi\cos\pi - 0) + \Big[\sin x\Big]_0^\pi = \pi.$$

同樣地，取 $u = x$ 及 $\mathrm{d}v = \cos x\,\mathrm{d}x$：

$$\int_0^\pi x\cos x\,\mathrm{d}x = \Big[x\sin x\Big]_0^\pi - \int_0^\pi \sin x\,\mathrm{d}x = 0 - \Big[-\cos x\Big]_0^\pi = -(1+1) = -2.$$

我們還需要在 $[0,\pi]$ 上的三個標準積分：

$$\int_0^\pi \sin^2 x\,\mathrm{d}x = \int_0^\pi \cos^2 x\,\mathrm{d}x = \frac{\pi}{2}, \qquad \int_0^\pi \sin x\cos x\,\mathrm{d}x = 0.$$

頭兩個可由 $\sin^2 x = \frac{1}{2}(1-\cos 2x)$ 及 $\cos^2 x = \frac{1}{2}(1+\cos 2x)$ 得出，因為 $\cos 2x$ 在一整個周期上的積分為零；第三個是 $\frac{1}{2}\int_0^\pi \sin 2x\,\mathrm{d}x = 0$，理由相同。

f 的形式

展開 $\sin(x+t) = \sin x\cos t + \cos x \sin t$。積分是對 x 進行的，所以 $\sin t$ 和 $\cos t$ 都是常數，可以移到積分之外：

$$\begin{aligned} \mathrm{f}(t) &= t + \int_0^\pi \mathrm{f}(x)\big(\cos x \sin t + \sin x \cos t\big)\,\mathrm{d}x \\ &= t + \sin t \int_0^\pi \mathrm{f}(x)\cos x\,\mathrm{d}x + \cos t \int_0^\pi \mathrm{f}(x)\sin x\,\mathrm{d}x \\ &= t + A\sin t + B\cos t, \end{aligned}$$

其中 A 和 B 就是題目給出的兩個積分。由於它們與 t 無關，所以是真正的常數。這就是所要求的形式。

求 A 和 B

想一想 我們並不知道 f 是甚麼，那又怎樣求出 A 和 B 呢？

我們其實知道它的形狀：$\mathrm{f}(x) = x + A\sin x + B\cos x$。把它代回 $(*)$，積分就變成可以用 A 和 B 表示並計算出來的式子，再比較等式兩邊，便得到這兩個常數所滿足的方程。

代入 $(*)$ 的右邊，並利用上面 $\sin(x+t)$ 的展開式：

$$t + \sin t \int_0^\pi (x + A\sin x + B\cos x)\cos x\,\mathrm{d}x + \cos t \int_0^\pi (x + A\sin x + B\cos x)\sin x\,\mathrm{d}x.$$

利用先前求得的積分，

$$\begin{aligned} \int_0^\pi (x + A\sin x + B\cos x)\cos x\,\mathrm{d}x &= -2 + A\cdot 0 + B\cdot\frac{\pi}{2} = -2 + \frac{\pi}{2}B, \\ \int_0^\pi (x + A\sin x + B\cos x)\sin x\,\mathrm{d}x &= \pi + A\cdot\frac{\pi}{2} + B\cdot 0 = \pi + \frac{\pi}{2}A. \end{aligned}$$

所以 $(*)$ 表示，對於每一個 t，

$$t + A\sin t + B\cos t = t + \left(-2 + \frac{\pi}{2}B\right)\sin t + \left(\pi + \frac{\pi}{2}A\right)\cos t.$$

比較 $\sin t$ 和 $\cos t$ 的係數（這樣做是合理的，因為此恆等式對所有 t 都成立：先代入 $t = \frac{\pi}{2}$，再代入 $t = 0$，便可分別抽出每個係數），得

$$A = -2 + \frac{\pi}{2}B, \qquad B = \pi + \frac{\pi}{2}A.$$

把第二式代入第一式，

$$A = -2 + \frac{\pi^2}{2} + \frac{\pi^2}{4}A \quad\Longrightarrow\quad A\left(1 - \frac{\pi^2}{4}\right) = \frac{\pi^2}{2} - 2, \quad \text{即} \quad A\cdot\frac{4-\pi^2}{4} = -\frac{4-\pi^2}{2}.$$

由於 $\pi^2 \neq 4$，我們可以除以 $4-\pi^2$，得 $A = -2$。於是 $B = \pi + \frac{\pi}{2}(-2) = 0$。因此

$$A = -2, \qquad B = 0, \qquad \text{所以} \qquad \mathrm{f}(t) = t - 2\sin t.$$

驗算

直接用 A 和 B 的定義來核對答案。取 $\mathrm{f}(x) = x - 2\sin x$，

$$\int_0^{\pi} \mathrm{f}(x)\cos x\,\mathrm{d}x = -2 - 2\cdot 0 = -2 = A, \qquad \int_0^{\pi} \mathrm{f}(x)\sin x\,\mathrm{d}x = \pi - 2\cdot\frac{\pi}{2} = 0 = B.$$

數值方面，用辛普森法則(Simpson's rule)可得 $\int_0^{\pi} x\sin x\,\mathrm{d}x \approx 3.14159$ 及 $\int_0^{\pi} x\cos x\,\mathrm{d}x \approx -2.00000$。我們又取 $\mathrm{f}(t) = t - 2\sin t$，在五十個隨機的 t 值上計算 $(*)$ 的右邊，每一次所得的結果與 $\mathrm{f}(t)$ 的差距都在 10^{-9} 之內。

常見錯誤

- 忘記 $\int_0^{\pi} x\cos x\,\mathrm{d}x$ 中的負號：邊界項為零，而 $-\int_0^{\pi}\sin x\,\mathrm{d}x = -2$。
- 弄錯哪一個積分與 $\sin t$ 相乘。$\sin(x+t)$ 中 $\cos x\sin t$ 的部分給出 $A\sin t$，其中 $A = \int \mathrm{f}(x)\cos x\,\mathrm{d}x$，而不是 $\int \mathrm{f}(x)\sin x\,\mathrm{d}x$。
- 除以 $1 - \frac{\pi^2}{4}$ 時，沒有指出它不等於零。

STEP 之外

方程 $(*)$ 是一條*第二類弗雷德霍姆積分方程*（*Fredholm integral equation of the second kind*）。它的核（kernel）$\sin(x+t) = \sin x\cos t + \cos x\sin t$ 是有限個「x 的函數乘以 t 的函數」之和（稱為「退化」（degenerate）核或「可分離」（separable）核），這正是問題能夠化約為一個 2×2 線性方程組（linear system）的原因。條件 $\pi^2 \neq 4$ 所說的，就是這個方程組的行列式（determinant）不等於零；假如行列式為零，方程便可能無解，或有無限多個解，這就是奇異矩陣（singular matrix）在積分方程中的對應情況。至於這背後的線性代數，可參閱 Strang 的 *Introduction to Linear Algebra*。

詳細解答

STEP II 2004，第 6 題

純數學 • 向量分解

> **STEP II 2004，第 6 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q6

關鍵思路

把一個向量（vector）分拆成沿 $\mathbf{a}$ 方向的部分和垂直（perpendicular）於 $\mathbf{a}$ 的部分，其實就是投影（projection）：由於 $\mathbf{a}$ 是單位向量（unit vector），任何向量 $\mathbf{v}$ 沿它的分量（component）就是 $(\mathbf{v}\cdot\mathbf{a})\,\mathbf{a}$。第二部分是同一個想法，只是高了一個維度。向量 $\mathbf{a}$ 和 $\mathbf{q}$ 是 Π 內兩個互相垂直的方向，而 $\mathbf{R}$ 則提供第三個與兩者垂直的方向。因此，我們既可以逐一向每個方向投影，也可以先設未知倍數，再用點積（dot product）寫出垂直的條件。我們採用第二種做法，並以第一種做法作驗算。

解答

向量 p 與 q

由於 $\mathbf{p}$ 平行（parallel）於 $\mathbf{a}$，設 $\mathbf{p}=\lambda\mathbf{a}$。於是

$$\mathbf{q}=\mathbf{a}+\mathbf{b}-\lambda\mathbf{a}=(1-\lambda)\mathbf{a}+\mathbf{b},$$

它自然位於 Π 內，因為 $\mathbf{a}$ 和 $\mathbf{b}$ 都在其中。要使 $\mathbf{q}$ 垂直於 $\mathbf{a}$，我們需要

$$0=\mathbf{q}\cdot\mathbf{a}=(1-\lambda)\,|\mathbf{a}|^2+\mathbf{a}\cdot\mathbf{b}=(1-\lambda)+3,$$

這裏用了 $|\mathbf{a}|=1$ 和 $\mathbf{a}\cdot\mathbf{b}=3$。所以 $\lambda=4$，並且

$$\boxed{\mathbf{p}=4\mathbf{a},\qquad \mathbf{q}=\mathbf{b}-3\mathbf{a}.}$$

確實有 $\mathbf{q}\cdot\mathbf{a}=3-3=0$，以及 $\mathbf{p}+\mathbf{q}=\mathbf{a}+\mathbf{b}$。（換一個說法，$\mathbf{p}$ 就是投影 $\big((\mathbf{a}+\mathbf{b})\cdot\mathbf{a}\big)\mathbf{a}=4\mathbf{a}$。）

向量 P、Q 與 R

首先列出幾項稍後要用的事實。由 $|\mathbf{b}| = 5$，

$$|\mathbf{q}|^2 = |\mathbf{b}|^2 - 6\,\mathbf{a}\cdot\mathbf{b} + 9|\mathbf{a}|^2 = 25 - 18 + 9 = 16,$$

所以 $|\mathbf{q}| = 4$。特別地，$\mathbf{q} \neq \mathbf{0}$，故 $\mathbf{b}$ 不平行於 $\mathbf{a}$，而 $\mathbf{a}$ 和 $\mathbf{b}$ 張成（span）平面 Π。

設 $\mathbf{P} = \alpha\mathbf{a}$ 及 $\mathbf{Q} = \beta\mathbf{q} = \beta(\mathbf{b} - 3\mathbf{a})$。這樣 $\mathbf{R}$ 就必定是

$$\mathbf{R} = \mathbf{a} + \mathbf{b} + \mathbf{c} - \alpha\mathbf{a} - \beta(\mathbf{b} - 3\mathbf{a}) = \mathbf{c} + u\,\mathbf{a} + v\,\mathbf{b}, \qquad \text{其中} u = 1 - \alpha + 3\beta,\ v = 1 - \beta.$$

想一想 怎樣把「$\mathbf{R}$ 垂直於 Π」化成方程？

一個向量垂直於一個平面，當且僅當它垂直於該平面內兩個不平行的向量。我們剛剛看到 $\mathbf{a}$ 和 $\mathbf{b}$ 正是這樣的一對向量。所以條件就是 $\mathbf{R}\cdot\mathbf{a} = 0$ 和 $\mathbf{R}\cdot\mathbf{b} = 0$。利用 $\mathbf{a}\cdot\mathbf{c} = -2$、$\mathbf{b}\cdot\mathbf{c} = 2$、$\mathbf{a}\cdot\mathbf{b} = 3$、$|\mathbf{a}|^2 = 1$ 和 $|\mathbf{b}|^2 = 25$：

$$\mathbf{R}\cdot\mathbf{a} = -2 + u + 3v = 0,$$
$$\mathbf{R}\cdot\mathbf{b} = 2 + 3u + 25v = 0.$$

由第一式得 $u = 2 - 3v$。代入第二式得 $8 + 16v = 0$，所以

$$v = -\tfrac{1}{2}, \qquad u = \tfrac{7}{2}.$$

於是 $\beta = 1 - v = \frac{3}{2}$，而 $\alpha = 1 + 3\beta - u = 1 + \frac{9}{2} - \frac{7}{2} = 2$。因此

$$\boxed{\mathbf{P} = 2\mathbf{a}, \qquad \mathbf{Q} = \tfrac{3}{2}(\mathbf{b} - 3\mathbf{a}), \qquad \mathbf{R} = \mathbf{c} + \tfrac{7}{2}\mathbf{a} - \tfrac{1}{2}\mathbf{b}.}$$

把三者相加，$\mathbf{a}$ 的係數（coefficient）是 $2 - \frac{9}{2} + \frac{7}{2} = 1$，$\mathbf{b}$ 的係數是 $\frac{3}{2} - \frac{1}{2} = 1$，$\mathbf{c}$ 的係數是 1，所以 $\mathbf{P} + \mathbf{Q} + \mathbf{R} = \mathbf{a} + \mathbf{b} + \mathbf{c}$，符合要求。

最後，$\mathbf{R} \neq \mathbf{0}$：假如它是零向量，$\mathbf{c} = \frac{1}{2}\mathbf{b} - \frac{7}{2}\mathbf{a}$ 便會位於 Π 內，與 $\mathbf{c}$ 不平行於 Π 這個假設矛盾。正是這個假設，令 $\mathbf{R}$ 成為一個真正的法向量（normal vector）。

另一個角度：向互相垂直的軸投影

單位向量 $\mathbf{a}$、$\hat{\mathbf{q}} = \frac{1}{4}(\mathbf{b} - 3\mathbf{a})$ 和 $\mathbf{n}$（Π 的一個單位法向量）組成一個正交歸一組（orthonormal set）。所以 $\mathbf{P}$ 和 $\mathbf{Q}$ 就是 $\mathbf{v} = \mathbf{a} + \mathbf{b} + \mathbf{c}$ 在前兩個向量上的投影：

$$\mathbf{v}\cdot\mathbf{a} = 1 + 3 - 2 = 2, \qquad \text{所以} \mathbf{P} = 2\mathbf{a};$$
$$\mathbf{v}\cdot\hat{\mathbf{q}} = \tfrac{1}{4}\big(\mathbf{a}\cdot\mathbf{q} + \mathbf{b}\cdot\mathbf{q} + \mathbf{c}\cdot\mathbf{q}\big) = \tfrac{1}{4}(0 + 16 + 8) = 6, \qquad \text{所以} \mathbf{Q} = 6\hat{\mathbf{q}} = \tfrac{3}{2}(\mathbf{b} - 3\mathbf{a}),$$

這裏用了 $\mathbf{b}\cdot\mathbf{q} = 25 - 9 = 16$ 和 $\mathbf{c}\cdot\mathbf{q} = 2 + 6 = 8$。餘下的部分 $\mathbf{R}$，就是 $\mathbf{c}$ 沿 Π 法線方向的分量。這與上面的答案一致。

驗算

- 直接驗算：$\mathbf{R}\cdot\mathbf{a} = -2 + \frac{7}{2} - \frac{3}{2} = 0$，而 $\mathbf{R}\cdot\mathbf{b} = 2 + \frac{21}{2} - \frac{25}{2} = 0$。
- 一個具體例子：取 $\mathbf{a} = \mathbf{i}$、$\mathbf{b} = 3\mathbf{i} + 4\mathbf{j}$ 和 $\mathbf{c} = -2\mathbf{i} + 2\mathbf{j} + \mathbf{k}$，它們滿足所有已知條件。這時 $\mathbf{q} = 4\mathbf{j}$、$\mathbf{P} = 2\mathbf{i}$、$\mathbf{Q} = 6\mathbf{j}$ 及 $\mathbf{R} = \mathbf{k}$，而它們的和是 $\mathbf{a} + \mathbf{b} + \mathbf{c} = 2\mathbf{i} + 6\mathbf{j} + \mathbf{k}$。
- 我們亦用 2000 組滿足條件、方向隨機的向量組合測試過這些公式。每一次都有 $\mathbf{q}\cdot\mathbf{a} = 0$、$|\mathbf{q}| = 4$，$\mathbf{R}$ 是 Π 的法向量，而且三個向量之和等於 $\mathbf{a} + \mathbf{b} + \mathbf{c}$，誤差全部在 10^{-9} 以內。

常見錯誤

- 取 $\mathbf{p} = (\mathbf{a}\cdot\mathbf{b})\mathbf{a} = 3\mathbf{a}$，這只是 $\mathbf{b}$ 本身的投影。我們需要的是 $\mathbf{a} + \mathbf{b}$ 的投影，即 $4\mathbf{a}$。
- 只加上 $\mathbf{R}\cdot\mathbf{a} = 0$ 這一個條件。要垂直於一個平面，需要兩個獨立的條件。
- 假設 $\mathbf{P} = \mathbf{p}$ 及 $\mathbf{Q} = \mathbf{q}$。加入 $\mathbf{c}$ 之後，平面內的分量也會改變，因為 $\mathbf{c}$ 本身在 Π 內也有分量。

STEP 之外

這是格拉姆－施密特正交化過程（Gram–Schmidt process）的縮影。我們由 $\mathbf{a}$、$\mathbf{b}$、$\mathbf{c}$ 出發，構造出兩兩垂直的方向 $\mathbf{a}$、$\mathbf{b} - 3\mathbf{a}$ 和 $\mathbf{c} + \frac{7}{2}\mathbf{a} - \frac{1}{2}\mathbf{b}$，然後把一個向量沿這些方向分解。同樣的步驟在任何維數都適用。它是正交投影（orthogonal projection）、最小二乘擬合（least-squares fitting）和矩陣的 QR 分解（QR factorisation）的基礎。Strang 的 *Introduction to Linear Algebra*（線性代數入門教材）對這些課題都有詳細闡述。

詳細解答

STEP II 2004，第 7 題

純數學 • 二分法

> **STEP II 2004，第 7 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q7

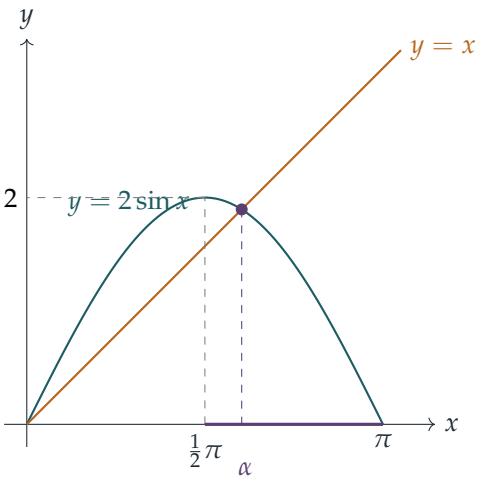
關鍵思路

方程 $2\sin x - x = 0$ 的根（root），就是曲線 $y = 2\sin x$ 與直線 $y = x$ 相交之處，所以圖像部分只涉及一條正弦曲線和一條直線。至於二分法（bisection）的每一步，歸根究底都只問同一個問題：在中點（midpoint）處，$2\sin c$ 比 c 大還是小？半角公式（half-angle formula）可把 $\frac{5}{8}\pi$ 和 $\frac{9}{16}\pi$ 的正弦值化為多重根式，而比較兩個正數的平方，則令計算只需用到題目給出的近似值。

解答

I_0 內恰有一個根

我們有 $f(x) = 0 \iff 2\sin x = x$，所以各根就是 $y = 2\sin x$ 與 $y = x$ 的交點的 x 坐標。



在 $[\frac{1}{2}\pi, \pi]$ 上：

- 在 $x = \frac{1}{2}\pi$ 處，曲線的高度為 2，直線的高度為 $\frac{1}{2}\pi < 2$（因為 $\pi < 4$），所以曲線在直線之上；
- 在 $x = \pi$ 處，曲線的高度為 0，直線的高度為 $\pi > 0$，所以曲線在直線之下；
- 在這個區間（interval）上，曲線 $y = 2\sin x$ 遞減，而直線 $y = x$ 遞增。

兩個圖像都是連續（continuous）的，所以曲線必定在區間內某處穿過直線；又因為一個下降、一個上升，曲線一旦落到直線之下，就不可能再回到直線之上。因此兩者恰好相交一次，即 $\mathrm{f}(x) = 0$ 在 I_0 內恰有一個根 α。用 f 來表達：$\mathrm{f}(\frac{1}{2}\pi) = 2 - \frac{1}{2}\pi > 0$，$\mathrm{f}(\pi) = -\pi < 0$，而且 f 在該區間上嚴格遞減（strictly decreasing）。

在以下每一步二分法中，我們都保留 f 變號的那一半。由於 $\mathrm{f}(\frac{1}{2}\pi) > 0$，一切都取決於 $\mathrm{f}(c_n)$ 的正負：若 $\mathrm{f}(c_n) < 0$，根在 c_n 的左邊；若 $\mathrm{f}(c_n) > 0$，根就在右邊。

求 I_1

這裏 $c_0 = \frac{3}{4}\pi$，而

$$\mathrm{f}(\tfrac{3}{4}\pi) = 2 \cdot \frac{1}{\sqrt{2}} - \tfrac{3}{4}\pi \approx 1.4 - \tfrac{3}{4}\sqrt{10}.$$

由於 $\sqrt{10} > 3$，我們有 $\frac{3}{4}\sqrt{10} > 2.25 > 1.4$，所以 $\mathrm{f}(c_0) < 0$。連同 $\mathrm{f}(a_0) > 0$，可得 $\mathrm{f}(a_0)\mathrm{f}(c_0) < 0$，所以

$$I_1 = [\tfrac{1}{2}\pi, \tfrac{3}{4}\pi].$$

求 I_2

現在 $c_1 = \frac{1}{2}(\frac{1}{2}\pi + \frac{3}{4}\pi) = \frac{5}{8}\pi$，我們需要知道 $2\sin\frac{5}{8}\pi - \frac{5}{8}\pi$ 的正負。

想一想 不用計算機，怎樣求 $\sin\frac{5}{8}\pi$？

利用 $\sin\frac{5}{8}\pi = \sin(\pi - \frac{5}{8}\pi) = \sin\frac{3}{8}\pi = \cos\frac{1}{8}\pi$，再以 $\theta = \frac{1}{8}\pi$ 代入倍角公式（double-angle formula）$\cos 2\theta = 2\cos^2\theta - 1$：

$$4\cos^2\tfrac{1}{8}\pi = 2\big(1+\cos\tfrac{1}{4}\pi\big) = 2+\sqrt{2}.$$

所以 $(2\sin\frac{5}{8}\pi)^2 = 2+\sqrt{2}$。$2\sin\frac{5}{8}\pi$ 和 $\frac{5}{8}\pi$ 都是正數，所以可以比較它們的平方：

$$(2\sin\tfrac{5}{8}\pi)^2 = 2+\sqrt{2} \approx 3.4, \qquad (\tfrac{5}{8}\pi)^2 \approx \tfrac{25}{64}\cdot 10 = \tfrac{250}{64} \approx 3.9.$$

因此 $2\sin\frac{5}{8}\pi < \frac{5}{8}\pi$，即 $\mathrm{f}(c_1) < 0$。同樣地，$\mathrm{f}(a_1) = \mathrm{f}(\frac{1}{2}\pi) > 0$，所以 $\mathrm{f}(a_1)\mathrm{f}(c_1) < 0$，於是

$$I_2 = [\tfrac{1}{2}\pi, \tfrac{5}{8}\pi],$$

正如題目所求。

求 I_3

現在 $c_2 = \frac{1}{2}(\frac{1}{2}\pi + \frac{5}{8}\pi) = \frac{9}{16}\pi$。與之前一樣，$\sin\frac{9}{16}\pi = \sin\frac{7}{16}\pi = \cos\frac{1}{16}\pi$，再以 $\theta = \frac{1}{16}\pi$ 代入倍角公式，得

$$(2\sin\tfrac{9}{16}\pi)^2 = 4\cos^2\tfrac{1}{16}\pi = 2\big(1+\cos\tfrac{1}{8}\pi\big) = 2+2\cos\tfrac{1}{8}\pi = 2+\sqrt{2+\sqrt{2}},$$

當中用到上一步得出的 $2\cos\frac{1}{8}\pi = \sqrt{2+\sqrt{2}}$。取 $\sqrt{2} \approx 1.4$，我們有 $\sqrt{2+\sqrt{2}} \approx \sqrt{3.4} > 1.8$（因為 $1.8^2 = 3.24$），所以

$$(2\sin\tfrac{9}{16}\pi)^2 > 3.8, \qquad (\tfrac{9}{16}\pi)^2 \approx \tfrac{81}{256}\cdot 10 = \tfrac{810}{256} < 3.2.$$

所以 $2\sin\frac{9}{16}\pi > \frac{9}{16}\pi$，即 $\mathrm{f}(c_2) > 0$。這一次 $\mathrm{f}(a_2)\mathrm{f}(c_2) > 0$，而 $\mathrm{f}(c_2)\mathrm{f}(b_2) = \mathrm{f}(\frac{9}{16}\pi)\,\mathrm{f}(\frac{5}{8}\pi) < 0$。因此

$$I_3 = [\tfrac{9}{16}\pi, \tfrac{5}{8}\pi].$$

這裏使用近似值是穩妥的：在每個中點處，所比較的兩個量相差約 0.5 或以上，遠大於 $\frac{1}{\sqrt{2}} \approx 0.7$ 或 $\pi \approx \sqrt{10}$ 所帶來的誤差。

驗算

- 以數值方法執行二分法，得到的區間相同，而根為 $\alpha \approx 1.8955$，位於 $\frac{9}{16}\pi \approx 1.767$ 與 $\frac{5}{8}\pi \approx 1.963$ 之間。
- 精確值與近似值吻合：$\mathrm{f}(\frac{3}{4}\pi) \approx -0.942$，$\mathrm{f}(\frac{5}{8}\pi) \approx -0.116$，以及 $\mathrm{f}(\frac{9}{16}\pi) \approx 0.194$。
- 每個區間的長度都是前一個的一半：$\frac{1}{2}\pi$、$\frac{1}{4}\pi$、$\frac{1}{8}\pi$、$\frac{1}{16}\pi$。

常見錯誤

- 只指出函數變號，卻沒有解釋為何只有一個根。單憑變號只能保證至少有一個根；要證明唯一性（uniqueness），還需要「一個遞減、一個遞增」的論證。
- 以平方來比較 $2\sin c$ 和 c，卻沒有先指出兩者都是正數。

STEP 之外

變號論證其實就是介值定理（intermediate value theorem），它是大學分析課程中最早遇到的重要定理之一；其證明需要用到實數的完備性（completeness），而二分法本質上就是把這個證明當作演算法來執行。二分法必定收斂，但速度緩慢（每一步只多得一個二進制位）。牛頓－拉弗森法（Newton–Raphson）等方法收斂得快得多，但有可能失敗，所以數值分析（numerical analysis）在很大程度上就是在穩妥與速度之間作取捨。有關理論可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP II 2004，第 8 題

純數學 • 微分方程的比較

STEP II 2004，第 8 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s2-q8

關鍵思路

當 $0 \leq x < 1$ 時，右邊為正數，所以 x 隨 t 遞增，我們可以把方程倒轉過來看：t 變成 x 的函數，即由 0 爬升到高度 x 所需的時間。在每個高度上斜率都較大，即每個高度都較早到達，所以題目給出的右邊各式之間的不等式（inequality）立即定出三條曲線的次序，一條在另一條之上。至於爬升到 $x = 1$ 所需的時間是有限還是無限，就變成一個瑕積分（improper integral）是否收斂（converge）的問題。

解答

以下一律以 $x_n(t)$ 表示指數為 n 時的解。在 $t = 0$ 時，$x = 0$，而對每個 n 斜率都是 $(1-0)^{1/n} = 1$，所以所有曲線都沿直線 $x = t$ 離開原點。當 $0 \leq x < 1$ 時，斜率嚴格為正，所以 x 遞增，而

$$\frac{\mathrm{d}t}{\mathrm{d}x} = \frac{1}{(1-x^n)^{1/n}}, \qquad \text{所以} \qquad t_n(x) = \int_0^x \frac{\mathrm{d}\xi}{(1-\xi^n)^{1/n}}$$

就是解到達高度 x 的時刻，其中 $0 \leq x < 1$。

第 (i) 部分

取 $n = 1$，方程為 $\dfrac{\mathrm{d}x}{\mathrm{d}t} = 1 - x$。分離變數（separating variables）（這是可行的，因為沿着解有 $1 - x > 0$），

$$t = \int_0^x \frac{\mathrm{d}\xi}{1-\xi} = -\ln(1-x), \qquad \text{所以} \qquad x = 1 - \mathrm{e}^{-t}.$$

圖像由原點出發，斜率為 1，一直遞增並且向下凹（concave）（其斜率 e^{-t} 遞減），當 $t \to \infty$ 時趨近漸近線（asymptote）$x = 1$ 但永遠不會到達：當 $x \to 1$ 時，$t_1(x) =$

$-\ln(1-x) \to \infty$。下圖中最低的一條曲線就是它。

第 (ii) 部分

不等式。對 $0 < x < 1$，$1-x$ 和 $(1-x^2)^{1/2}$ 都是正數，所以該不等式等價於兩邊平方後所得的不等式：

$$(1-x)^2 < 1-x^2 \iff 1-2x+x^2 < 1-x^2 \iff 2x^2 < 2x \iff x < 1,$$

其中最後一步是除以 $2x > 0$。由於 $x < 1$ 成立，所以 $1-x < (1-x^2)^{1/2}$ 也成立。

草圖。在 $(0,1)$ 內的每個高度 x，$n = 2$ 的曲線都比 $n = 1$ 的曲線陡。因此，對 $0 < x < 1$，

$$t_2(x) = \int_0^x \frac{\mathrm{d}\xi}{(1-\xi^2)^{1/2}} < \int_0^x \frac{\mathrm{d}\xi}{1-\xi} = t_1(x),$$

因為左邊的被積函數（integrand）在 $(0,x]$ 上嚴格較小。所以 $n = 2$ 的解嚴格較早到達每一個高度；又因為兩個解都是遞增的，所以在每個時刻 $t > 0$（只要 $x_2 < 1$），$n = 2$ 的曲線都嚴格位於 $n = 1$ 的曲線之上。兩者都以斜率 1 離開原點，所以在原點附近它們非常接近，然後隨着 t 增大而分開。

> **想一想** 兩條曲線在原點的斜率相同。為甚麼 $n = 1$ 的曲線不能在稍後悄悄升到上面？

要升到上面，它必須在某個高度 x 追上 $n = 2$ 的曲線，即在同一時刻到達 x。但對每個 $x \in (0,1)$ 都有 $t_2(x) < t_1(x)$：爬得較慢的那個總是嚴格地遲到。（等價的說法是：凡兩條曲線處於同一高度之處，$n = 2$ 的那條都較陡，所以它永遠不會被超越。）

求解。設 $x = \sin y$，其中 $0 \le y < \frac{1}{2}\pi$，且在 $t = 0$ 時 $y = 0$。於是 $\cos y > 0$，所以 $(1-x^2)^{1/2} = (\cos^2 y)^{1/2} = \cos y$，方程變成

$$\cos y \frac{\mathrm{d}y}{\mathrm{d}t} = \cos y \quad \Longrightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}t} = 1 \quad \Longrightarrow \quad y = t,$$

這裏用到在 $t = 0$ 時 $y = 0$。所以解為

$$x = \sin t \qquad \text{其中} 0 \le t < \tfrac{1}{2}\pi.$$

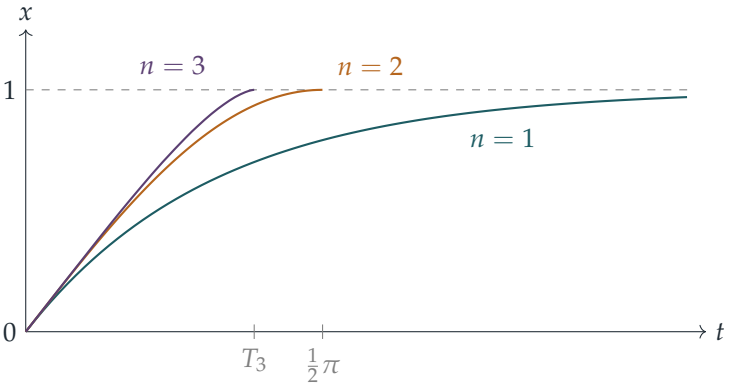
這與 $t_2(x) = \arcsin x$ 吻合。與 $n = 1$ 的情況不同，這個解在有限時間內到達 $x = 1$，時刻為 $t = \frac{1}{2}\pi$，而且到達時斜率為 $\cos\frac{1}{2}\pi = 0$，與直線 $x = 1$ 相切。

第 (iii) 部分

與之前完全一樣，在 $(0,1)$ 上運用題目給出的不等式 $(1-x^2)^{1/2} < (1-x^3)^{1/3}$，可知 $n = 3$ 的曲線在每個高度都比 $n = 2$ 的曲線陡，所以

$$t_3(x) = \int_0^x \frac{\mathrm{d}\xi}{(1-\xi^3)^{1/3}} < \int_0^x \frac{\mathrm{d}\xi}{(1-\xi^2)^{1/2}} = \arcsin x \qquad (0 < x < 1).$$

所以 $n = 3$ 的曲線和其他曲線一樣以斜率 1 離開原點，並且在 $t > 0$ 時嚴格位於 $x = \sin t$ 之上。它的斜率 $(1 - x^3)^{1/3}$ 在 $x \to 1$ 時降至 0，所以它與直線 $x = 1$ 相遇時同樣會變得平坦。



在 $t = \frac{1}{2}\pi$ 之前到達 $x = 1$。草圖顯示 $n = 3$ 的曲線位於 $\sin t$ 之上，而後者本身在 $\frac{1}{2}\pi$ 到達 1；所以 $n = 3$ 的曲線必定不遲於該時刻到達 1。要看出它是嚴格較早到達，可在積分的比較中令 $x \to 1$。到達 $x = 1$ 所需的時間為

$$T_3 = \int_0^1 \frac{\mathrm{d}\xi}{(1 - \xi^3)^{1/3}}.$$

這個瑕積分收斂：由於在 $[0,1]$ 上 $1 - \xi^3 = (1 - \xi)(1 + \xi + \xi^2) \geq 1 - \xi$，被積函數至多為 $(1 - \xi)^{-1/3}$，而後者在 $[0,1]$ 上的積分為 $\frac{3}{2}$。又由於被積函數在 $(0,1)$ 上嚴格小於 $(1 - \xi^2)^{-1/2}$，

$$T_3 < \int_0^1 \frac{\mathrm{d}\xi}{(1 - \xi^2)^{1/2}} = \Big[\arcsin \xi\Big]_0^1 = \tfrac{1}{2}\pi.$$

所以 $n = 3$ 的解在時刻 $T_3 < \frac{1}{2}\pi$ 到達 $x = 1$，這正是所要證明的。（一旦到達，它就停留在 $x = 1$，因為斜率為 0，而且不可能變成負數。）

驗算

- 代回原式：$\frac{\mathrm{d}}{\mathrm{d}t}(1 - \mathrm{e}^{-t}) = \mathrm{e}^{-t} = 1 - x$，而對 $0 \leq t \leq \frac{1}{2}\pi$ 有 $\frac{\mathrm{d}}{\mathrm{d}t}\sin t = \cos t = (1 - \sin^2 t)^{1/2}$。以數值方法（龍格—庫塔法，Runge–Kutta）解各方程，所得結果與這些公式吻合。
- 數值上，$T_3 \approx 1.2092$，明顯小於 $\frac{1}{2}\pi \approx 1.5708$。事實上 $T_3 = \dfrac{2\pi}{3\sqrt{3}}$ 是準確值（見下文），兩者吻合至小數點後七位。
- 在高度 $x = 0.9$，三個時間分別為 $-\ln 0.1 \approx 2.30$、$\arcsin 0.9 \approx 1.12$ 和 $t_3(0.9) \approx 0.98$，次序與預測相符。

常見錯誤

- 只說「斜率較大，所以曲線較高」，卻沒有指出所比較的是同一高度上的斜率，而不是同一時刻的斜率。改為把 t 看成 x 的函數，論證才站得住腳。
- 寫出 $(1-\sin^2 y)^{1/2} = \cos y$ 而沒有說明為何 $\cos y \geq 0$：這需要 $0 \leq y \leq \frac{1}{2}\pi$。
- 把 $x = \sin t$ 延續到 $t = \frac{1}{2}\pi$ 之後：在那裏它會遞減，但方程規定 $\mathrm{d}x/\mathrm{d}t \geq 0$。

STEP 之外

解在有限時間內到達 $x = 1$，然後永遠停留在 $x = 1$，原因是 $(1-x^n)^{1/n}$ 在 $x = 1$ 處不滿足利普希茨（Lipschitz）條件；微分方程（differential equation）的唯一性定理（uniqueness theorem）正是在這裏失效（當 $n = 2$ 時，$x = 1$ 和 $x = \sin t$ 都通過 $(\frac{1}{2}\pi, 1)$）。到達時間是一個 Beta 函數（Beta function）：代入 $u = \xi^n$ 可得 $T_n = \frac{1}{n}B(\frac{1}{n}, 1-\frac{1}{n}) = \dfrac{\pi/n}{\sin(\pi/n)}$，當 $n = 2$ 時為 $\frac{1}{2}\pi$，當 $n = 3$ 時為 $\frac{2\pi}{3\sqrt{3}}$，並且當 $n \to \infty$ 時遞減趨向 1。Boyce 與 DiPrima 的 *Elementary Differential Equations*（基礎微分方程教科書）討論了解的存在性與唯一性，以及解的比較。

詳細解答

STEP II 2004，第 9 題

力學 • 重心

STEP II 2004，第 9 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s2-q9

關鍵思路

半球以曲面着地時，與平面接觸的是球面上的一點，而球面的法線必定通過球心。因此反作用力沿着通過底面圓心 O 的鉛垂線作用；平衡時，無論放在那裏的是甚麼物體，其重心（centre of gravity）都必定在同一條鉛垂線上。於是題目的每一部分都只是「求出重心，然後傾斜物體，直至重心在 O 的正下方」。採用固定在物體上的坐標，整個計算就輕鬆得多。

解答

模型與平衡條件

把半球視為剛體（rigid body），它與水平面只在一點 C 接觸；把質點（particle）視為固定在邊緣上的點質量。作用力只有各個重量和 C 處的接觸力。由於 C 在球面上，C 處的法線通過 O；又因為平面是水平的，這條法線是鉛垂的。所以 C 在 O 的正下方，而且無論物體怎樣滾動，O 的高度始終是 r。

平衡（equilibrium）時，接觸力必須與總重量互相抵銷，所以它是鉛垂的（沒有摩擦力），並且與重量沿同一直線作用。因此：

平衡時，系統的重心在通過 O 的鉛垂線上。

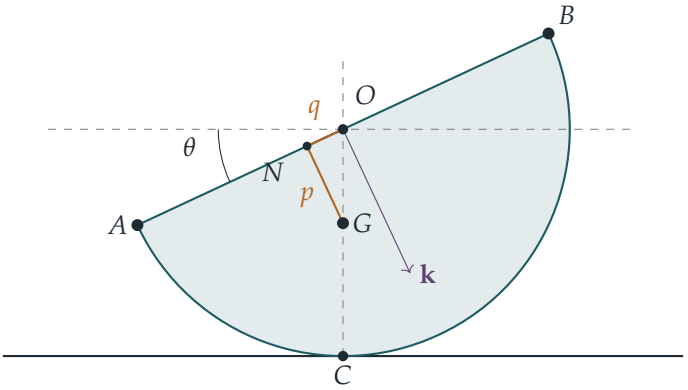
重心在 O 的下方，因為物體位於底面的曲面那一側。

物體坐標與初始位置

在物體上固定兩個互相垂直的單位向量（unit vector）：**k** 沿對稱軸，由 O 指向曲面；**e** 在底面所在的平面內。由於 G 與底面的距離是 p，與 O 的距離是 $\sqrt{p^2+q^2}$，它在底面上的垂足 N 與 O 的距離是 q。選取 **e** 由 O 指向 N，便得

$$\overrightarrow{OG} = p\,\mathbf{k} + q\,\mathbf{e}.$$

底面垂直於 **k**，所以底面與水平面的夾角等於 **k** 與鉛垂線的夾角。在初始位置，$\overrightarrow{OG}$ 鉛垂向下，所以這個角 θ 滿足 $\tan\theta = q/p$。



邊緣上哪一點最低？邊緣上的一點相對於 O 的位置是 $r(\cos\phi\,\mathbf{e} + \sin\phi\,\mathbf{f})$，其中 **f** 是第三個單位向量，與前兩者都垂直。這一點在 O 以下的深度，就是它沿向下的單位向量 $(p\mathbf{k}+q\mathbf{e})/\sqrt{p^2+q^2}$ 的分量，即

$$\frac{rq\cos\phi}{\sqrt{p^2+q^2}},$$

當 $\phi = 0$ 時最大（這裏取 $q > 0$；若 $q = 0$，底面是水平的，沒有唯一的最低點）。所以

$$\overrightarrow{OA} = r\,\mathbf{e}, \qquad \overrightarrow{OB} = -r\,\mathbf{e}.$$

用文字說明：A 與 G 在對稱軸的同一側，並且在通過 O 和 G 的鉛垂平面內，而整個問題始終留在這個平面內。

質點在 A

半球連同質點的重心 G_A 由下式給出：

$$(M+m)\,\overrightarrow{OG_A} = M(p\,\mathbf{k} + q\,\mathbf{e}) + mr\,\mathbf{e} = Mp\,\mathbf{k} + (Mq+mr)\,\mathbf{e}.$$

在新的平衡位置，$\overrightarrow{OG_A}$ 是鉛垂的，所以與求 θ 時完全一樣，**k** 與鉛垂線的夾角（亦即底面與水平面的夾角）就是 $\overrightarrow{OG_A}$ 與 **k** 的夾角：

$$\tan\alpha = \frac{Mq+mr}{Mp}, \qquad \text{所以} \qquad \alpha = \tan^{-1}\left(\frac{Mq+mr}{Mp}\right).$$

正如所料，A 處多出的重量把 A 拉得更低，於是 $\alpha > \theta$。

> **想一想** 為甚麼底面的傾角就是 $\overrightarrow{OG_A}$ 與對稱軸的夾角？

底面垂直於對稱軸 $\mathbf{k}$，而水平面垂直於鉛垂線。把一對直線各自旋轉一個直角，並不改變兩者之間的夾角，所以底面與水平面的夾角等於 $\mathbf{k}$ 與鉛垂線的夾角。平衡時，通過 O 的鉛垂線就是直線 OG_A，它沿 $\mathbf{k}$ 和 $\mathbf{e}$ 的分量分別是 Mp 和 $Mq + mr$（除以 $M + m$ 之後）。它與 $\mathbf{k}$ 所成角的正切就是這兩個分量之比。

質點在 B

現在質點位於 $-r\,\mathbf{e}$，所以新的重心 G_B 滿足

$$(M+m)\,\overrightarrow{OG_B} = Mp\,\mathbf{k} + (Mq - mr)\,\mathbf{e}, \qquad \text{由此得} \qquad \tan\beta = \frac{Mq - mr}{Mp}.$$

由於 m 很小，我們有 $mr < Mq$，所以 $\overrightarrow{OG_B}$ 仍然與先前一樣在對稱軸的同一側，底面向同一方向傾斜，傾角是較小的 β。

兩個位置之間的夾角

由複角公式（compound-angle formula），

$$\begin{aligned}\tan(\alpha-\beta) &= \frac{\tan\alpha - \tan\beta}{1 + \tan\alpha\tan\beta} = \frac{\dfrac{2mr}{Mp}}{1 + \dfrac{(Mq+mr)(Mq-mr)}{M^2p^2}} \\ &= \frac{2mMrp}{M^2p^2 + M^2q^2 - m^2r^2} = \frac{2mMrp}{M^2\left(p^2+q^2\right) - m^2r^2},\end{aligned}$$

其中第三步把分子和分母同乘以 M^2p^2。這就是所要求的結果。

驗算

- 若 $q = 0$（例如均勻半球），兩個位置互為鏡像：$\tan\alpha = mr/(Mp)$ 而 $\beta = -\alpha$。公式給出 $\tan 2\alpha = 2t/(1-t^2)$，其中 $t = mr/(Mp)$，正是二倍角公式，與預期相符。
- 若 $m \to 0$，則 α 和 β 都趨向 θ，而右邊趨向 0。
- 取 $M = 1$、$m = 0.1$、$r = 1$、$p = \frac{3}{8}$、$q = \frac{1}{8}$，得 $\tan\alpha = 0.6$ 和 $\tan\beta = \frac{1}{15}$，所以 $\alpha \approx 30.96°$，$\beta \approx 3.81°$。結果的兩邊都等於 $\frac{20}{39} \approx 0.5128$。我們亦以數值方法求出平衡角：對 2000 組隨機數值，把重心的高度減至最小，所得結果與 $\tan\alpha$ 和 $\tan\beta$ 的公式相符，誤差在 10^{-13} 以內。

常見錯誤

- 把 G 放在原來位置的接觸點的正下方，而不是 O 的正下方。接觸點會隨物體滾動而移動；O 才是始終在接觸點正上方的那一點。
- 把 A 取在對稱軸的錯誤一側。邊緣的最低點與 G 在同一側，這正是質點在 A 時傾角增大的原因。
- 當 $mr > Mq$ 時仍把 β 當作正數。這時底面向另一方向傾斜，公式只在把 β 算作負數時才成立；「m 很小」這個條件正是要排除這種情況。

STEP 之外

這裏所用的原理是：剛體在重力和單一接觸力作用下平衡時，其質心（centre of mass）在該力的作用線上。等價的說法是，質心處於高度的駐值（stationary）位置；而考察物體滾動時這個高度的變化，還可以判定平衡是否穩定。這就是判定穩定性的能量法（energy method）的起點，大學一年級力學會把這套方法發展下去，用來處理搖擺和滾動的物體。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）正是以這種觀點講述質心和剛體平衡。

詳細解答

STEP II 2004，第 10 題

力學 • 斜面上的摩擦力

STEP II 2004，第 10 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s2-q10

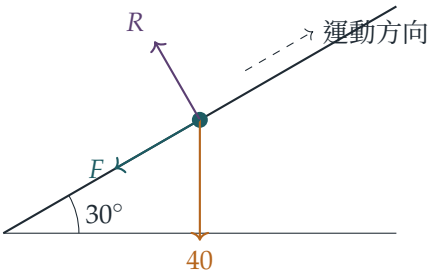
關鍵思路

摩擦力（friction）永遠與運動方向相反，所以質點（particle）沿斜面向上滑與向下滑時的加速度（acceleration）並不相同。先求出這兩個加速度：題目的數字經過挑選，結果十分整齊。然後 P 向上滑到停下，再滑回來，與仍在向上滑的 Q 相遇。把兩個質點的位置都寫成「在最高點以下的距離」，碰撞（collision）方程就成為一個平方差（difference of two squares）。

（印刷試卷的第 (i) 部分有一處誤植：碰撞時間中的「b」應為 T。）

解答

我們把每個質點視為粗糙斜面（rough plane）上的質點，並以 x 表示由 A 沿 ℓ 向上量度的位置，取沿斜面向上為正方向。



垂直於斜面的方向沒有加速度，所以法向反作用力（normal reaction）為

$$R = 4 \cdot 10 \cos 30^\circ = 20\sqrt{3}\ \text{N},$$

而由於每個質點都在滑動，摩擦力達到極限值（limiting friction）：

$$F = \mu R = \frac{1}{5\sqrt{3}} \cdot 20\sqrt{3} = 4\,\text{N}.$$

重量沿斜面向下的分量是 $40\sin 30^\circ = 20\,\text{N}$。

向上運動時：摩擦力沿斜面向下，所以 $4a = -(20+4)$，減速度為 $6\,\text{m}\,\text{s}^{-2}$。

向下運動時：摩擦力沿斜面向上，所以沿斜面向下的加速度為 $(20-4)/4 = 4\,\text{m}\,\text{s}^{-2}$。

由於 $20 > 4$，質點停下後不會保持靜止：極限摩擦力不足以把它留住，所以它會滑回下來。

第 (i) 部分

P 的運動。向上時，$v = 6 - 6t$，所以 P 在 $t = 1$ 停下，此時已走了 $\frac{1}{2} \cdot 6 \cdot 1 = 3\,\text{m}$。當 $0 \le t \le 1$ 時，$x_P = 6t - 3t^2 = 3 - 3(1-t)^2$。其後 P 由 $x = 3$ 從靜止開始以加速度 4 向下滑，所以

$$x_P = 3 - 2(t-1)^2 \qquad (t \ge 1),$$

此式一直成立，直至 P 回到 A，那時 $2(t-1)^2 = 3$，即 $t = 1 + \sqrt{3/2}$。

Q 的運動。與上述運動完全相同，只是時間推遲了 T：設 $u = t - T$，則

$$x_Q = 6u - 3u^2 = 3 - 3(1-u)^2 \qquad (0 \le u \le 1).$$

想一想 兩個質點會否在都仍向上運動時碰撞？

不會。當兩者都向上運動時，Q 所走的正是 P 在 T 秒之前走過的路徑，而沿這條路徑 x 是遞增的，所以 P 嚴格地在前面。它們只可能在 P 掉頭之後相遇，即 $t \ge 1$ 之時。此外，Q 也不可能先到達自己的最高點（$x = 3$，在 $u = 1$ 時），因為那時 P 早已離開 $x = 3$，位於其下方。所以碰撞時 P 正在下滑，而 Q 正在上滑。

碰撞。設 $\tau = t - 1 \ge 0$ 為 P 掉頭後所經過的時間。那麼 $1 - u = 1 - (t - T) = T - \tau$，而由 $x_P = x_Q$ 得

$$2\tau^2 = 3(T-\tau)^2.$$

τ 和 $T - \tau = 1 - u$ 都是非負的（後者是因為 Q 仍在向上運動），所以可以取正平方根：

$$\sqrt{2}\,\tau = \sqrt{3}\,(T - \tau)$$

$$\tau = \frac{\sqrt{3}\,T}{\sqrt{3}+\sqrt{2}} = \sqrt{3}(\sqrt{3}-\sqrt{2})\,T = (3-\sqrt{6})\,T,$$

其中利用 $(\sqrt{3}+\sqrt{2})(\sqrt{3}-\sqrt{2}) = 1$ 作有理化（rationalising）。因此碰撞發生在 P 拋出後

$$t = (3-\sqrt{6})\,T + 1$$

秒，正如所求。

餘下要看看關於 T 的條件在哪裏用到。這個時刻必須早於 P 回到 A 的時刻，所以需要 $\tau < \sqrt{3/2}$：

$$(3-\sqrt{6})\,T < \sqrt{\tfrac{3}{2}} \iff T < \frac{\sqrt{3/2}\,(3+\sqrt{6})}{3} = \sqrt{\tfrac{3}{2}} + \tfrac{\sqrt{9}}{3} = 1 + \sqrt{\tfrac{3}{2}},$$

因為 $\sqrt{3/2}\cdot\sqrt{6} = 3$。這正是題目給出的條件。（同一個不等式亦保證 $t \geq T$，即 Q 已經拋出：$(3-\sqrt{6})T + 1 \geq T \iff T \leq 1/(\sqrt{6}-2) = 1+\sqrt{3/2}$。）

第 (ii) 部分

當 $T = 1+\sqrt{2/3}$ 時，

$$\tau = (3-\sqrt{6})\left(1+\sqrt{\tfrac{2}{3}}\right) = 3 - \sqrt{6} + 3\sqrt{\tfrac{2}{3}} - \sqrt{6}\sqrt{\tfrac{2}{3}} = 3 - \sqrt{6} + \sqrt{6} - 2 = 1,$$

所以碰撞發生在 $t = 2$。這個 T 值小於 $1+\sqrt{3/2}$，所以第 (i) 部分的結果適用。

因摩擦而損失的能量等於摩擦力乘以滑行的總距離，因為摩擦力的大小恆為 4 N，而且永遠與運動方向相反。

- P 向上滑了 3 m，然後向下滑了 $2\tau^2 = 2$ m：合共 5 m。
- Q 尚未掉頭，並且與 P 位於同一點，即 $x = 3-2 = 1$：它滑行了 1 m。

所以損失的能量是

$$4 \times (5+1) = 24\ \text{J}.$$

驗算

用能量收支可以核實這個答案。起初的動能（kinetic energy）是 $2 \times \frac{1}{2}\cdot 4 \cdot 6^2 = 144$ J。在 $t = 2$ 時，P 的速率是 $4 \times 1 = 4$（動能 32 J）；Q 有 $u = 1-\sqrt{2/3}$，速率為 $6(1-u) = 6\sqrt{2/3}$，所以 $v^2 = 24$（動能 48 J）。兩者都在斜面上 1 m 處，即升高了 $\frac{1}{2}$ m，勢能（potential energy）增加了 $2 \times 4 \cdot 10 \cdot \frac{1}{2} = 40$ J。而 $144 - 32 - 48 - 40 = 24$ J。另外，對隨機選取的 $T < 1+\sqrt{3/2}$，以數值方法逐步模擬兩個質點的運動，所得的碰撞時間與 $(3-\sqrt{6})T+1$ 吻合（例如當 $T \approx 1.816$ 時為 ≈ 2.000），因摩擦而損失的能量為 24.000 J。

常見錯誤

- P 向下滑時沿用向上滑時的加速度：摩擦力的方向已經逆轉，所以應是 4，而不是 6。
- 忘記了 Q 所受的摩擦力：Q 也一直在滑動，它滑行的 1 m 貢獻了 4 J。
- 在 $2\tau^2 = 3(T-\tau)^2$ 中取錯平方根：正負號由 $\tau \geq 0$ 和 $T - \tau \geq 0$ 決定。

STEP 之外

摩擦力是非保守力（non-conservative force）最簡單的例子：它所作的功取決於路徑，而不僅取決於起點和終點，這就是為甚麼我們需要的是滑行的總距離，而不是位移。這也解釋了為甚麼機械能（mechanical energy）在某些系統中守恆，在另一些系統中卻不守恆；大學一年級的力學課程會用線積分（line integral）和勢函數（potential）把這個想法嚴格地表述出來。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）對功、保守力和摩擦力有仔細的論述。

詳細解答

STEP II 2004，第 11 題

力學 • 功率與阻力

STEP II 2004，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

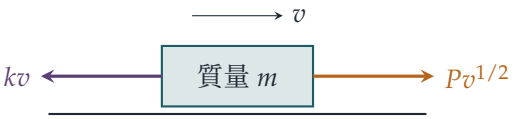
step.maths.org/questions/04-s2-q11

關鍵思路

功率 (power) 等於力乘以速率，所以引擎以全功率 $Pv^{3/2}$ 運作時，所施的推力為 $Pv^{1/2}$。由牛頓第二定律（Newton's second law）便得到一條含有 $\sqrt{v}$ 的微分方程（differential equation），而代換（substitution）$u = \sqrt{v}$ 可把它化為一階線性方程（linear first-order equation），馬上就能解出。引擎關掉之後，兩列火車的速率各自按指數衰減（exponential decay），而第 (iii) 部分的竅門就是比較這兩個衰減率。

解答

我們把每列火車看成沿水平直路軌運動的質點（particle）。引擎以全功率開動時，沿路軌方向的力有驅動力（driving force）和阻力（resistance）kv。



若引擎在速率為 v 時輸出的功率為 H，則其驅動力為 H/v。所以在全功率下，A 所受的驅動力為 $Pv^{3/2}/v = Pv^{1/2}$，而 B 所受的驅動力為 $2Pv^{1/2}$。

第 (i) 部分

對於 A，沿路軌方向應用牛頓第二定律，得

$$m\frac{\mathrm{d}v_A}{\mathrm{d}t} = Pv_A^{1/2} - kv_A.$$

設 $u = v_A^{1/2}$，則 $v_A = u^2$，且 $\dfrac{\mathrm{d}v_A}{\mathrm{d}t} = 2u\dfrac{\mathrm{d}u}{\mathrm{d}t}$。於是

$$2mu\frac{\mathrm{d}u}{\mathrm{d}t} = Pu - ku^2, \qquad \text{所以} \qquad 2m\frac{\mathrm{d}u}{\mathrm{d}t} = P - ku,$$

這裏兩邊除以 u；由於火車一直在運動（$u > 0$），這樣做是容許的。這是一條線性（而且可分離變數（separable））的方程：

$$\int \frac{2m\,\mathrm{d}u}{P - ku} = \int \mathrm{d}t \quad \Longrightarrow \quad -\frac{2m}{k}\ln(P - ku) = t + \text{常數},$$

所以 $P - ku = \mathrm{C}\mathrm{e}^{-kt/2m}$，其中 C 為某個常數。（開始時 u 極小，故 $P - ku > 0$；而且它一直保持為正，因為解永遠不會到達平衡值（equilibrium）$u = P/k$。）

想一想 「以極低的速率緩緩前進」告訴了我們甚麼？

它給出了初始條件（initial condition）。在 $t = 0$ 時，兩個速率都是正數（所以除以 u 並無問題），但小得可以忽略，因此我們取 $t = 0$ 時 $u = 0$。於是 $C = P$，而

$$u = \frac{P}{k}\left(1 - \mathrm{e}^{-kt/2m}\right), \qquad \text{所以} \qquad v_A = \frac{P^2\left(1 - \mathrm{e}^{-kt/2m}\right)^2}{k^2},$$

正是所要求的結果。（假如火車完全靜止，$v \equiv 0$ 也會滿足原來的方程，因為那時根本沒有驅動力；這就是題目要讓火車一開始已在緩慢移動的原因。）

對於 B，方程是 $2m\dfrac{\mathrm{d}v_B}{\mathrm{d}t} = 2Pv_B^{1/2} - kv_B$。用同一個代換可得 $4m\dfrac{\mathrm{d}u}{\mathrm{d}t} = 2P - ku$，這正是 A 的方程，只是把 P 換成 $2P$、m 換成 $2m$。所以

$$v_B = \frac{4P^2\left(1 - \mathrm{e}^{-kt/4m}\right)^2}{k^2}.$$

第 (ii) 部分

兩個速率都是正數，所以 $9v_A = 4v_B$ 等價於 $3\sqrt{v_A} = 2\sqrt{v_B}$。設 $x = \mathrm{e}^{-kt/4m}$，則 $\mathrm{e}^{-kt/2m} = x^2$，而且當 $t > 0$ 時 $0 < x < 1$。條件變成

$$\frac{3P}{k}\left(1 - x^2\right) = \frac{4P}{k}(1 - x) \quad \Longrightarrow \quad 3(1 - x)(1 + x) = 4(1 - x).$$

由於 $x \neq 1$，我們可以除以 $1 - x$，得 $3 + 3x = 4$，所以 $x = \frac{1}{3}$（即 $t = \frac{4m}{k}\ln 3$）。於是 $1 - x^2 = \frac{8}{9}$，$1 - x = \frac{2}{3}$，所以

$$v_A = \frac{64P^2}{81k^2}, \qquad v_B = \frac{16P^2}{9k^2}.$$

簡單驗算一下：$9v_A = \frac{64P^2}{9k^2} = 4v_B$。

第 (iii) 部分

從引擎關掉的一刻起計算時間 τ，並設 $V_A = \frac{64P^2}{81k^2}$ 和 $V_B = \frac{16P^2}{9k^2}$ 為該刻的速率。這時只有阻力起作用：

$$m\frac{\mathrm{d}v_A}{\mathrm{d}\tau} = -kv_A, \qquad 2m\frac{\mathrm{d}v_B}{\mathrm{d}\tau} = -kv_B,$$

所以

$$v_A = V_A\,\mathrm{e}^{-k\tau/m}, \qquad v_B = V_B\,\mathrm{e}^{-k\tau/2m}.$$

較重的火車 B 減速的指數率只有一半，所以把 v_B 平方，兩個指數項便一致了：

$$v_B^2 = V_B^2\,\mathrm{e}^{-k\tau/m} = \frac{V_B^2}{V_A}\,v_A.$$

最後

$$\frac{V_B^2}{V_A} = \frac{256P^4}{81k^4}\cdot\frac{81k^2}{64P^2} = \frac{4P^2}{k^2},$$

所以 $v_B^2 = \dfrac{4P^2}{k^2}v_A$，即在其後的所有時刻都有 $k^2v_B^2 = 4P^2v_A$，正是所要求的結果。

驗算

- 當 $t \to \infty$ 時，$v_A \to P^2/k^2$，這正是驅動力 $Pv^{1/2}$ 與阻力 kv 相等之處。同樣地，$v_B \to 4P^2/k^2$，那裏 $2Pv^{1/2} = kv$。
- 當 t 很小時，$v_A \approx P^2t^2/(4m^2)$，$v_B \approx P^2t^2/(4m^2)$：兩列火車起步時並駕齊驅，這是應有的結果，因為兩者每單位質量的驅動力都是 $Pv^{1/2}/m$。
- 取 $m = 1.7$、$P = 2.3$、$k = 0.9$，我們由一個極小的初速率出發，用數值方法（龍格—庫塔法（Runge–Kutta））逐步計算兩條運動方程，結果與第 (i) 部分的公式吻合至約 10^{-5}；在 $9v_A = 4v_B$ 的一刻，得 $\mathrm{e}^{-kt/4m} = 0.33333$、$v_A = 5.1602$ 和 $v_B = 11.6104$；關掉引擎後繼續逐步計算，$k^2v_B^2$ 一直與 $4P^2v_A$ 相等（例如兩秒後兩者都是 37.8739）。

常見錯誤

- 把 $Pv^{3/2}$ 當作力。它是功率；力等於功率除以速率。
- 忘記第 (ii) 部分的條件是一條關於 $x = \mathrm{e}^{-kt/4m}$ 的二次方程，除了 $x = \frac{1}{3}$ 之外，還有根 $x = 1$（即 $t = 0$，此時兩個速率皆為零）。

STEP 之外

代換 $u = \sqrt{v}$ 是求解伯努利方程（Bernoulli equation）$y' + p(t)y = q(t)y^n$ 的方法的一個特例；在該方法中，$u = y^{1-n}$ 總能給出一條線性方程。關於火車完全靜止的那個備註，讓我們初次窺見唯一性理論（uniqueness theory）：由於 $\sqrt{v}$ 在 $v = 0$ 處的斜率為無窮大，方程由 $v = 0$ 出發的解不止一個，皮卡－林德勒夫定理（Picard–Lindelöf theorem）解釋了這個現象。Boyce 與 DiPrima 的 *Elementary Differential Equations*（微分方程入門教材）涵蓋了這兩個課題；至於功率與阻力的力學，可參閱 Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教材）。

詳細解答

STEP II 2004，第 12 題

概率與統計 • 連續分佈

> **STEP II 2004，第 12 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q12

關鍵思路

一切都源自一個原函數（antiderivative）：除了一個常數因子之外，xe^{-ax^2} 就是 e^{-ax^2} 的導數。由此可求出歸一化常數(normalising constant)，並得到累積分佈函數(cumulative distribution function)的閉合形式，所以只需解一條指數方程便可求得中位數(median)。眾數（mode）就是曲線達到頂點的位置，前提是該頂點位於 $[0,1]$ 之內；否則概率密度在 $x=1$ 處仍在上升。至於條件概率（conditional probability），只需用到累積分佈函數在 m 和 h 的值。

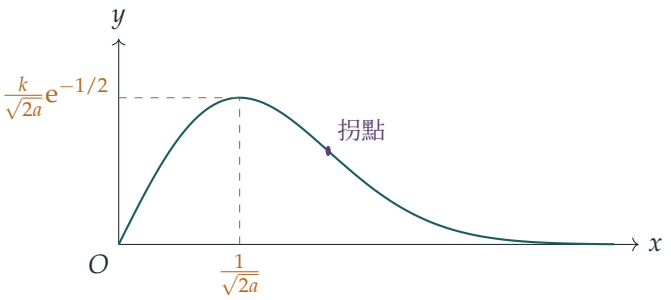
解答

草圖

設 $g(x)=kxe^{-ax^2}$，其中 $x\geq 0$。則 $g(0)=0$，當 $x>0$ 時 $g(x)>0$，而且由積法則（product rule）可得

$$g'(x)=ke^{-ax^2}\left(1-2ax^2\right),$$

它在 $0\leq x<1/\sqrt{2a}$ 時為正，其後為負。所以 g 由原點開始上升（在該處的斜率為 k），在 $x=1/\sqrt{2a}$ 達到唯一的極大值，此時 $y=\dfrac{k}{\sqrt{2a}}e^{-1/2}$，然後下降。當 $x\to\infty$ 時，指數因子壓倒線性因子，所以 $y\to 0$：x 軸是漸近線（asymptote）。（再求導一次可知曲線在 $x=\sqrt{3/(2a)}$ 有一個拐點（point of inflection），曲線在該處由凹（concave）轉為凸（convex）。）



常數 k

由於 $\dfrac{\mathrm{d}}{\mathrm{d}x}e^{-ax^2} = -2axe^{-ax^2}$，對於 $0 \le t \le 1$，

$$\int_0^t kxe^{-ax^2}\,\mathrm{d}x = \left[-\frac{k}{2a}e^{-ax^2}\right]_0^t = \frac{k}{2a}\left(1-e^{-at^2}\right).$$

總概率為 1，所以取 $t=1$ 可得 $\dfrac{k}{2a}(1-e^{-a}) = 1$，即

$$k = \frac{2a}{1-e^{-a}},$$

這正是所需的結果。（因為 $a > 0$，所以分母為正。）把這個值代回去，在 $[0,1]$ 上的累積分佈函數為

$$F(t) = P(X \le t) = \frac{1-e^{-at^2}}{1-e^{-a}}.$$

眾數

概率密度函數（density）就是我們剛才所畫的曲線，在 $x=1$ 處截斷。從草圖可見，在 $[0,1]$ 上，如果轉向點（turning point）位於區間之內，密度便在該點最大，否則在區間的右端點最大。

- 若 $a > \frac{1}{2}$，則 $2a > 1$，所以 $1/\sqrt{2a} < 1$：頂點位於區間之內，故 $m = \dfrac{1}{\sqrt{2a}}$。
- 若 $a < \frac{1}{2}$，則 $1/\sqrt{2a} > 1$，所以在整個 $[0,1)$ 上 $f'(x) > 0$：密度在整個區間上遞增，故 $m = 1$。

（當 $a = \frac{1}{2}$ 時，兩條公式都給出 $m = 1$。）

中位數

中位數 h 滿足 $F(h) = \frac{1}{2}$：

$$1-e^{-ah^2} = \tfrac{1}{2}\left(1-e^{-a}\right) \quad\Longleftrightarrow\quad e^{-ah^2} = \frac{1+e^{-a}}{2}.$$

取對數，並利用 $h > 0$，

$$h = \sqrt{\frac{1}{a}\ln\left(\frac{2}{1+\mathrm{e}^{-a}}\right)}.$$

由於 $1+\mathrm{e}^{-a} < 2$，對數為正，所以這是一個實數。

何時 $h > m$？

若 $a \le \frac{1}{2}$，則 $m = 1$，而 $h < 1$（一半的概率在 h 以下，而 1 以上則沒有概率），所以 $h < m$。因此只有當 $a > \frac{1}{2}$ 時才可能有 $h > m$，而此時 $m = 1/\sqrt{2a}$。比較兩者的平方（兩者皆為正），

$$\begin{aligned} h > m &\iff \frac{1}{a}\ln\left(\frac{2}{1+\mathrm{e}^{-a}}\right) > \frac{1}{2a} \iff \ln\left(\frac{2}{1+\mathrm{e}^{-a}}\right) > \frac{1}{2} \\ &\iff \frac{2}{1+\mathrm{e}^{-a}} > \mathrm{e}^{1/2} \iff 1+\mathrm{e}^{-a} < 2\mathrm{e}^{-1/2} \\ &\iff \mathrm{e}^{-a} < 2\mathrm{e}^{-1/2} - 1 \iff a > -\ln\left(2\mathrm{e}^{-1/2}-1\right). \end{aligned}$$

第一步我們乘以 $a > 0$，第三步用到 exp 是遞增函數，最後一步用到 ln 是遞增函數。最後一步還需要 $2\mathrm{e}^{-1/2} - 1 > 0$，這是成立的，因為 $\mathrm{e}^{1/2} < 2$（因為 $\mathrm{e} < 4$）。

> **想一想** 以上一連串的推導假設了 $a > \frac{1}{2}$。結論不是需要把它列為額外條件嗎？

不需要，而這正是下一部分要求證明它的原因。我們在下面會證明 $c = -\ln\left(2\mathrm{e}^{-1/2} - 1\right)$ 大於 $\frac{1}{2}$。所以若 $a > c$，則自動有 $a > \frac{1}{2}$，公式 $m = 1/\sqrt{2a}$ 適用，而上述推導說明 $h > m$。連同 $a \le \frac{1}{2}$ 的情況，這就證明了 $h > m$ 當且僅當 $a > c$。

該常數大於 $\frac{1}{2}$

由於 ln 是遞增函數，

$$-\ln\left(2\mathrm{e}^{-1/2}-1\right) > \tfrac{1}{2} \iff 2\mathrm{e}^{-1/2} - 1 < \mathrm{e}^{-1/2} \iff \mathrm{e}^{-1/2} < 1,$$

而最後一個不等式成立。（數值上，$c \approx 1.546$。）

條件概率

假設 $a > c$。則 $m < h$，所以事件 $\{X > m\} \cap \{X < h\}$ 就是事件 $\{m < X < h\}$。由於 $P(X < h) = \frac{1}{2}$，

$$P(X > m \mid X < h) = \frac{P(m < X < h)}{P(X < h)} = \frac{F(h) - F(m)}{\frac{1}{2}} = 1 - 2F(m).$$

當 $m = 1/\sqrt{2a}$ 時，$am^2 = \frac{1}{2}$，所以 $F(m) = \dfrac{1 - \mathrm{e}^{-1/2}}{1 - \mathrm{e}^{-a}}$。因此

$$P(X > m \mid X < h) = \frac{(1 - \mathrm{e}^{-a}) - 2(1 - \mathrm{e}^{-1/2})}{1 - \mathrm{e}^{-a}} = \frac{2\mathrm{e}^{-1/2} - \mathrm{e}^{-a} - 1}{1 - \mathrm{e}^{-a}},$$

這正是所需的結果。

驗算

- 當 $a = c$ 時，分子 $2\mathrm{e}^{-1/2} - \mathrm{e}^{-a} - 1$ 恰好為 0，與該處 $m = h$ 吻合；當 $a > c$ 時它為正，符合概率應有的性質。
- 當 $a \to \infty$ 時，答案趨向 $2\mathrm{e}^{-1/2} - 1 \approx 0.213$，小於 1。
- 取 $a = 2.5$：$k \approx 5.447$，$m \approx 0.447$，$h \approx 0.496$，而對密度作數值積分可得 $P(X > m \mid X < h) \approx 0.1427$，與公式一致。取 $a = 1$ 則得 $m \approx 0.707 > h \approx 0.616$，與 $1 < c$ 相符。

常見錯誤

- 在所有情況下都答 $m = 1/\sqrt{2a}$：當 $a < \frac{1}{2}$ 時，轉向點落在 $x = 1$ 之外，而該處的密度為零。
- 忘記 F 裏面的 k 必須是歸一化後的值，才能使 $F(1) = 1$。
- 寫成 $P(X > m \mid X < h) = P(X > m)/P(X < h)$；分子必須是兩個事件同時發生的概率。

STEP 之外

這個密度是截斷瑞利分佈（truncated Rayleigh distribution）：若不在 $x = 1$ 處截斷，取 $k = 2a$ 的 $kx\mathrm{e}^{-ax^2}$ 就是一個點與原點距離的分佈，而該點的兩個坐標是互相獨立的正態變量（normal variables）。比較平均值（mean）、中位數和眾數，以及截斷如何使它們移動，是大學第一門統計課程中反覆出現的主題。Grimmett and Welsh 的 *Probability: An Introduction*（概率入門教材）是進一步了解連續隨機變量（continuous random variables）及其密度函數和分佈函數的好去處。

詳細解答

STEP II 2004，第 13 題

概率與統計 • 最優停止

STEP II 2004，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s2-q13

關鍵思路

逐次抽球去想。假設她已經安全取得 n 個白球，再抽一次的話，要麼多得 £1，要麼把手上的 £n 全部輸掉，所以這一次抽球的期望（expected）效果很容易寫出來。以 N 個球為目標時，總期望獎金會上升，當且僅當多抽那一球的期望得益為正，因此最佳目標就是該得益仍然非負的最後一個目標。

解答

全題假設 $1 \le r < b$，即最少有一個紅球和最少有一個白球。（若 $r = 0$，她只需把所有球取出。）

下一次抽球的期望得益

假設她首 n 次抽到的都是白球。袋中現在有 $b-n$ 個球：r 個紅球和 $b-n-r$ 個白球，每個球在下一次被抽出的機會均等（要還有白球剩下，須有 $n \le b-r-1$；無論如何 $b-n \ge r \ge 1$）。

- 抽到白球的概率（probability）為 $\dfrac{b-n-r}{b-n}$，她的獎金由 £n 變為 £$(n+1)$：得益為 $+1$。
- 抽到紅球的概率為 $\dfrac{r}{b-n}$，她會失去本來可得的 £n：得益為 $-n$。

所以她下一次抽球的期望得益是

$$G_n = \frac{(b-n-r)\cdot 1 + r\cdot(-n)}{b-n} = \frac{b-r-n(r+1)}{b-n}.$$

分母為正，所以 $G_n = 0$ 當且僅當 $b - r = n(r+1)$，亦即

$$n = \frac{b-r}{r+1},$$

這正是所需的結果。更有用的是：當 $n < \dfrac{b-r}{r+1}$ 時 $G_n > 0$，而當 $n > \dfrac{b-r}{r+1}$ 時 $G_n < 0$。

最佳的取球數目

設 P_N 為首 N 次抽球全為白球的概率，並設 E_N 為她以取出 N 個球為目標時的總期望獎金。若該 N 個球全為白球，她得到 £N，否則一無所得，所以

$$E_N = NP_N.$$

想一想 單次抽球的期望得益與總期望獎金 E_N 有甚麼關係？

比較以 N 為目標和以 $N+1$ 為目標這兩個方案。只有當首 N 個球全為白球時，兩個方案才有分別（否則在兩個方案下她都已經輸掉），這情況出現的概率為 P_N；而在這情況下，多抽的一球使她的獎金改變，改變量的期望值為 G_N。因此

$$E_{N+1} - E_N = P_N G_N = P_N \frac{b-r-N(r+1)}{b-N}.$$

（亦可由 $P_{N+1} = P_N \cdot \frac{b-r-N}{b-N}$ 直接驗證。）

當 $N \le b-r$ 時，我們有 $P_N > 0$，所以 $E_{N+1} - E_N$ 的正負號與 $b-r-N(r+1)$ 相同。記 $k = N+1$，這就是下式的正負號：

$$(b+1) - k(r+1).$$

設 $m = \dfrac{b+1}{r+1}$。那麼當 $k < m$ 時 $E_k > E_{k-1}$，當 $k = m$ 時 $E_k = E_{k-1}$，而當 $k > m$ 時 $E_k < E_{k-1}$。所以數列（sequence）$E_0, E_1, E_2, \ldots$ 嚴格遞增至 $E_{\lfloor m \rfloor}$（若 m 為整數，則出現相等的情況 $E_{m-1} = E_m$），然後嚴格遞減，至少在 $k-1 \le b-r$ 的範圍內如此。超出這範圍後 $P_N = 0$ 且 $E_N = 0$，不會更好。

我們應該驗證 $\lfloor m \rfloor$ 是一個合理的目標，即 $\lfloor m \rfloor \le b-r$。若 $b = r+1$，則 $m = \frac{r+2}{r+1}$，所以 $\lfloor m \rfloor = 1 = b-r$。若 $b \ge r+2$，則

$$(b-r)(r+1) - (b+1) = r(b-r-1) - 1 \ge r - 1 \ge 0,$$

所以 $m \le b-r$。無論哪種情況，E_N 的最大值都在

$$n = \left\lfloor \frac{b+1}{r+1} \right\rfloor,$$

取得，即 $\frac{b+1}{r+1}$ 的整數部分，這正是所需的結果。

無論採用甚麼停止規則（stopping rule），這個目標也是她所能做到的最佳選擇：在決定是否繼續之前，她唯一能得知的就是至今抽出的每個球都是白球（否則遊戲已經結束），所以任何規則都等同於預先選定一個數 N。

$r = 1$ 的情況

只有一個紅球時，首 N 次抽球避開紅球，當且僅當紅球位於最後 $b - N$ 個位置之一，而各個位置的機會均等。所以

$$P_N = \frac{b-N}{b}, \qquad E_N = \frac{N(b-N)}{b}.$$

最佳目標是 $n = \left\lfloor \frac{b+1}{2} \right\rfloor$。

b 爲偶數。這時 $\frac{b+1}{2} = \frac{b}{2} + \frac{1}{2}$，所以 $n = \frac{b}{2}$，且

$$E_n = \frac{\frac{b}{2} \cdot \frac{b}{2}}{b} = \frac{b}{4}.$$

她的總期望獎金為 $£\frac{1}{4}b$，這正是所需的結果。

b 爲奇數。這時 $n = \frac{b+1}{2}$，所以 $b - n = \frac{b-1}{2}$，且

$$E_n = \frac{\frac{b+1}{2} \cdot \frac{b-1}{2}}{b} = \frac{b^2-1}{4b}.$$

她的總期望獎金為 $£\dfrac{b^2-1}{4b}$。（這裏 $m = \frac{b+1}{2}$ 是整數，所以以 $\frac{b-1}{2}$ 個球為目標會得到相同的值，由對稱的公式 $N(b-N)/b$ 可以看出。）

驗算

- 當 $r = 1$ 時，$E_N = N(b-N)/b$ 是關於 N 的開口向下的拋物線（parabola），頂點在 $N = b/2$，與兩個答案都吻合：b 為偶數時，在頂點處恰好是 $b/4$；b 為奇數時則略小，為 $\frac{b}{4} - \frac{1}{4b}$。
- 取 $b = 7$，$r = 2$：公式給出 $n = \lfloor 8/3 \rfloor = 2$，而直接計算得 $E_0, \ldots, E_5 = 0,\ \frac{5}{7},\ \frac{20}{21},\ \frac{6}{7},\ \frac{4}{7},\ \frac{5}{21}$，在 $N = 2$ 處達到最高。我們列舉了球的全部 7! 種排列次序來確認這一點，並就每個 $b \le 24$，以完整的最優停止計算核對了 n 的公式。
- 當 $b = 9$，$r = 1$ 時，答案是 $\frac{80}{36} = \frac{20}{9} \approx 2.22$，剛好略低於 $\frac{9}{4}$。

常見錯誤

- 忘記了抽到紅球會令她失去已取得的 $£n$，而把期望得益只寫成 $\frac{b-n-r}{b-n}$。

- 混淆「已抽出 n 個球」與「以 n 個球為目標」：得益 G_n 在 $n = \frac{b-r}{r+1}$ 時為零，所以最佳目標要再多一個，即 $\frac{b-r}{r+1} + 1 = \frac{b+1}{r+1}$。
- 遺漏了 $\frac{b+1}{r+1}$ 為整數時的相等情況：這時 n 和 $n-1$ 同為最優。

STEP 之外

這是一個小型的最優停止（optimal stopping）問題，而所用的論證就是「向前看一步」（one-step look-ahead）法則：一旦再走一步的期望得益為負，便立即停止。只要問題是單調（monotone）的，這個法則就是最優的；所謂單調，是指一旦繼續下去顯得不利，其後便一直不利，正如本題中 G_n 隨 n 遞減。同樣的想法會引伸到祕書問題（secretary problem），以及大學一、二年級概率論中的鞅（martingale）方法。Grimmett and Welsh 的 *Probability: An Introduction*（概率論入門教材）是打好背後條件期望（conditional expectation）工具基礎的好去處。

詳細解答

STEP II 2004，第 14 題

概率與統計 • 容斥原理

> **STEP II 2004，第 14 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s2-q14

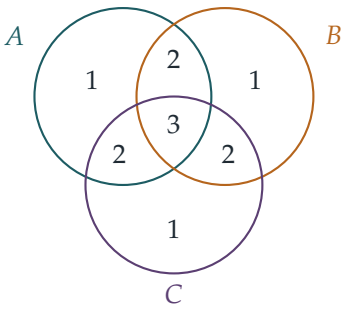
關鍵思路

容斥原理（inclusion–exclusion）是一種修正計數的方法：把 $\mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)$ 加起來時，重疊的部分被多算了，而兩兩相交與三者相交的項正好把計數修正過來。對於布甸這一題，考慮補事件（complement）會容易得多：設 E_i 為布甸 i 是空的這一事件（event）。某個布甸是空的概率是一個併集（union）的概率，這正是容斥原理所處理的情況，而且各個 E_i 的每一個交集（intersection）的概率都十分簡單。

解答

容斥公式

把 $A \cup B \cup C$ 分成溫氏圖（Venn diagram）中七個互不相交（disjoint）的區域。任何由若干個完整區域組成的事件，其概率等於這些區域的概率之和，所以只需驗證每個區域在等式右邊都恰好被計算一次。



圖中的數字表示每個區域落在 A、B、C 當中的多少個之內。取一個恰好落在三個事件

中 k 個之內的區域，其中 $k = 1$、2 或 3。

- 它出現在 $\mathrm{P}(A), \mathrm{P}(B), \mathrm{P}(C)$ 這些項當中的 k 項。
- 它出現在 $\binom{k}{2}$ 個兩兩交集之中，因為（舉例說）它落在 $A \cap B$ 之內，當且僅當它同時落在 A 和 B 之內。
- 只有當 $k = 3$ 時，它才出現在三者的交集之中，即出現 $\binom{k}{3}$ 次。

因此它被計算的淨次數為

$$k - \binom{k}{2} + \binom{k}{3} = \begin{cases} 1 - 0 + 0 = 1 & (k = 1), \\ 2 - 1 + 0 = 1 & (k = 2), \\ 3 - 3 + 1 = 1 & (k = 3). \end{cases}$$

$A \cup B \cup C$ 的每個區域都恰好被計算一次，而併集以外的區域則完全沒有被計算。所以等式右邊等於 $\mathrm{P}(A \cup B \cup C)$。

（另一個等價的做法：把兩個事件的法則 $\mathrm{P}(X \cup Y) = \mathrm{P}(X) + \mathrm{P}(Y) - \mathrm{P}(X \cap Y)$ 應用於 $X = A \cup B$ 和 $Y = C$，然後再應用於 $(A \cap C) \cup (B \cap C)$。）

每個布甸最少有一枚六便士硬幣

模型。徹底攪拌後再平均分成幾份，意思是每枚六便士硬幣（sixpence）落入布甸 1、2 或 3 的概率各為 $\frac{1}{3}$，並且與其他硬幣獨立（independent）。取 $r \geq 1$。

設 E_i 為布甸 i 內沒有六便士硬幣這一事件。

- 當全部 r 枚硬幣都沒有落入布甸 1 時，E_1 發生。每枚硬幣沒有落入布甸 1 的概率是 $\frac{2}{3}$，所以由獨立性可得 $\mathrm{P}(E_1) = \left(\frac{2}{3}\right)^r$，$E_2$、$E_3$ 亦同理。
- 當每枚硬幣都落入布甸 3 時，$E_1 \cap E_2$ 發生，所以 $\mathrm{P}(E_1 \cap E_2) = \left(\frac{1}{3}\right)^r$，其餘各對亦同理。
- $E_1 \cap E_2 \cap E_3$ 要求三個布甸全是空的，但由於 $r \geq 1$，這是不可能的。所以它的概率是 0。

由容斥原理，最少有一個布甸是空的概率為

$$p_r = \mathrm{P}(E_1 \cup E_2 \cup E_3) = 3\left(\tfrac{2}{3}\right)^r - 3\left(\tfrac{1}{3}\right)^r.$$

每個布甸最少有一枚六便士硬幣的概率就是其補事件的概率：

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

最少六枚六便士硬幣

我們要找出使 $p_r < \frac{1}{3}$ 的 r 值。直接計算可得

$$p_5 = \frac{3\cdot 32 - 3}{243} = \frac{93}{243} = \frac{31}{81} > \frac{27}{81} = \frac{1}{3}, \qquad p_6 = \frac{3\cdot 64 - 3}{729} = \frac{189}{729} = \frac{7}{27} < \frac{9}{27} = \frac{1}{3}.$$

為了確定沒有更小的 r 符合要求，我們證明 p_r 是遞減的。對於 $r \geq 1$，

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

當 $r = 1$ 時此式為 0，當 $r \geq 2$ 時為負數。所以 $p_1 = p_2 \geq p_3 \geq p_4 \geq p_5 = \frac{31}{81}$，而每個 $r \leq 5$ 都有 $p_r > \frac{1}{3}$。（當 $r = 0$ 時，每個布甸必然是空的。）因此，只有當 $r \geq 6$ 時，某個布甸沒有六便士硬幣的機會才會低於 $\frac{1}{3}$；又由於 p_r 持續遞減，每個 $r \geq 6$ 確實都有 $p_r \leq p_6 < \frac{1}{3}$。所以廚師必須使用 6 枚或以上的六便士硬幣，這正是所要證明的。

每個布甸有兩枚六便士硬幣

現在 $r = 6$。設 T 為每個布甸恰好有兩枚六便士硬幣這一事件，S 為每個布甸最少有一枚這一事件。由上一部分，

$$\mathrm{P}(S) = 1 - p_6 = 1 - \frac{7}{27} = \frac{20}{27}.$$

把六枚（可區分的）六便士硬幣分配到各個布甸，共有 $3^6 = 729$ 種等可能的方法。其中每個布甸得到兩枚的方法數目，就是把六件物件分成三個有標號的對的方法數目，即 $\dfrac{6!}{2!\,2!\,2!} = 90$。所以 $\mathrm{P}(T) = \frac{90}{729} = \frac{10}{81}$。

> **想一想** 為甚麼 $\mathrm{P}(T \cap S)$ 就是 $\mathrm{P}(T)$？

如果每個布甸都有兩枚六便士硬幣，那麼每個布甸當然最少有一枚，所以 $T \subseteq S$，而 $T \cap S = T$。因此

$$\mathrm{P}(T \mid S) = \frac{\mathrm{P}(T)}{\mathrm{P}(S)} = \frac{10/81}{20/27} = \frac{10}{81}\cdot\frac{27}{20} = \frac{1}{6}.$$

驗算

對於 $r = 1$ 和 $r = 2$，公式給出每個布甸都有六便士硬幣的概率為 0，這是對的：少於三枚硬幣不可能令三個布甸都有硬幣。對於 $r = 3$，公式給出 $\frac{27-24+3}{27} = \frac{2}{9}$，與直接計數所得的 $3!/3^3 = 6/27$ 吻合。對 $r \leq 9$ 逐一列舉全部 3^r 種分配方法，結果與公式一致，並確認 $p_5 = \frac{31}{81} \approx 0.383$ 和 $p_6 = \frac{7}{27} \approx 0.259$，同時得出最終答案 $\frac{1}{6} \approx 0.167$（模擬 200 000 次布甸製作所得的結果是 0.166）。

常見錯誤

- 把六便士硬幣當作不可區分，並把 $6 = a + b + c$ 的 28 種寫法視為等可能。它們並非等可能：$(2,2,2)$ 可以有 90 種出現方式，而 $(6,0,0)$ 只有一種。
- 只檢查 $r = 6$ 和 $r = 5$。「必須使用 6 枚或以上」還要求每個更小的 r 都不符合要求，這就是單調性（monotonicity）那一步之所以重要的原因。
- 忘記三者的交集是空集，或者不加說明就把它的概率寫成 $(0)^r$。

STEP 之外

對於 n 個事件的一般容斥原理，涉及所有 k 重交集的交錯和，其證明用的正是上面的計數論證，當中用到 $\sum_{j\geq 1}(-1)^{j-1}\binom{k}{j} = 1$。把它應用於 n 個布甸，所數算的就是滿射（surjection）的數目，並與第二類斯特靈數（Stirling numbers of the second kind）以及經典的優惠券收集問題（coupon-collector problem）相關。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）對容斥原理和這類佔位問題（occupancy problem）有詳細闡述。

詳細解答

STEP III 2004，第 1 題

純數學 • 雙曲函數積分

STEP III 2004，第 1 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q1

關鍵思路

兩個被積函數（integrand）都是一個雙曲函數（hyperbolic function）乘以由另一個雙曲函數構成的式子，而 $\sinh x\,\mathrm{d}x$ 和 $\cosh x\,\mathrm{d}x$ 正好分別是 $\cosh x$ 和 $\sinh x$ 的微分（differential）。因此，代換（substitution）$c = \cosh x$ 和 $s = \sinh x$ 可把它們化為有理函數的積分：一個要用部分分式（partial fractions），另一個則得出反正切（arctangent）。至於題目中「由此」（hence）的一步，關鍵在於留意到兩個分母其實是同一個函數，因為 $2\cosh^2 x - 1 = 1 + 2\sinh^2 x$。

解答

第一個積分

設 $c = \cosh x$，則 $\mathrm{d}c = \sinh x\,\mathrm{d}x$；積分限 $x = 0$ 和 $x = a$ 變為 $c = 1$ 和 $c = \cosh a$。於是

$$\int_0^a \frac{\sinh x}{2\cosh^2 x - 1}\,\mathrm{d}x = \int_1^{\cosh a} \frac{\mathrm{d}c}{2c^2 - 1}.$$

分母可因式分解為 $(\sqrt{2}\,c - 1)(\sqrt{2}\,c + 1)$，而

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

把右邊通分合併即可驗證：分子是 $\frac{1}{2}\big((\sqrt{2}c + 1) - (\sqrt{2}c - 1)\big) = 1$。每一項積分後都得出一個對數，並因鏈式法則（chain rule）而多出一個因子 $1/\sqrt{2}$。當 $c \geq 1$ 時，$\sqrt{2}c \pm 1$

兩者皆為正數，所以無須加上絕對值符號：

$$\begin{aligned}\int_1^{\cosh a} \frac{\mathrm{d}c}{2c^2-1} &= \frac{1}{2\sqrt{2}}\Big[\ln(\sqrt{2}\,c-1)-\ln(\sqrt{2}\,c+1)\Big]_1^{\cosh a}\\ &= \frac{1}{2\sqrt{2}}\ln\left(\frac{\sqrt{2}\cosh a-1}{\sqrt{2}\cosh a+1}\right)-\frac{1}{2\sqrt{2}}\ln\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right).\end{aligned}$$

由於 $-\ln\frac{\sqrt{2}-1}{\sqrt{2}+1}=\ln\frac{\sqrt{2}+1}{\sqrt{2}-1}$，這正是所求的結果。

第二個積分

現在設 $s=\sinh x$，則 $\mathrm{d}s=\cosh x\,\mathrm{d}x$，積分限為 $s=0$ 和 $s=\sinh a$：

$$\begin{aligned}\int_0^a \frac{\cosh x}{1+2\sinh^2 x}\,\mathrm{d}x &= \int_0^{\sinh a}\frac{\mathrm{d}s}{1+(\sqrt{2}\,s)^2} = \frac{1}{\sqrt{2}}\Big[\arctan(\sqrt{2}\,s)\Big]_0^{\sinh a}\\ &= \frac{1}{\sqrt{2}}\arctan\big(\sqrt{2}\sinh a\big).\end{aligned}$$

無窮積分

想一想 兩個被積函數的分母看來並不相同，我們怎樣才可以把它們相減？

利用 $\cosh^2 x=1+\sinh^2 x$，

$$2\cosh^2 x-1=2+2\sinh^2 x-1=1+2\sinh^2 x,$$

所以兩個分母完全相同，而

$$\begin{aligned}\int_0^a \frac{\cosh x-\sinh x}{1+2\sinh^2 x}\,\mathrm{d}x &= \frac{1}{\sqrt{2}}\arctan(\sqrt{2}\sinh a)-\frac{1}{2\sqrt{2}}\ln\left(\frac{\sqrt{2}\cosh a-1}{\sqrt{2}\cosh a+1}\right)\\ &\quad-\frac{1}{2\sqrt{2}}\ln\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right).\end{aligned}$$

現在令 $a\to\infty$。由於 $\sinh a\to\infty$，故 $\arctan(\sqrt{2}\sinh a)\to\frac{\pi}{2}$。由於 $\cosh a\to\infty$，

$$\frac{\sqrt{2}\cosh a-1}{\sqrt{2}\cosh a+1}=\frac{1-1/(\sqrt{2}\cosh a)}{1+1/(\sqrt{2}\cosh a)}\to 1,$$

而 ln 在 1 處連續（continuous），所以它的對數趨向 0。三項都有有限的極限，因此這個無窮積分收斂（converge），並且

$$\int_0^\infty \frac{\cosh x-\sinh x}{1+2\sinh^2 x}\,\mathrm{d}x=\frac{\pi}{2\sqrt{2}}-\frac{1}{2\sqrt{2}}\ln\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right),$$

正如所求。

最後一個積分

想一想 以 e^x 表示的話，$\cosh x - \sinh x$ 和 $1 + 2\sinh^2 x$ 會是甚麼樣子？

設 $u = e^x$，則 $\cosh x - \sinh x = e^{-x} = 1/u$；再利用二倍角公式（double-angle formula），

$$1 + 2\sinh^2 x = \cosh 2x = \frac{u^2 + u^{-2}}{2} = \frac{u^4 + 1}{2u^2}.$$

此外 $\mathrm{d}u = e^x\,\mathrm{d}x$，所以 $\mathrm{d}x = \mathrm{d}u/u$，而積分限 $x = 0$、$x \to \infty$ 變為 $u = 1$、$u \to \infty$。因此

$$\begin{aligned}\int_0^\infty \frac{\cosh x - \sinh x}{1 + 2\sinh^2 x}\,\mathrm{d}x &= \int_1^\infty \frac{1}{u} \cdot \frac{2u^2}{1 + u^4} \cdot \frac{\mathrm{d}u}{u} \\ &= 2\int_1^\infty \frac{\mathrm{d}u}{1 + u^4}.\end{aligned}$$

把上一個結果減半，

$$\int_1^\infty \frac{\mathrm{d}u}{1 + u^4} = \frac{\pi}{4\sqrt{2}} - \frac{1}{4\sqrt{2}} \ln\left(\frac{\sqrt{2} + 1}{\sqrt{2} - 1}\right).$$

這個式子還可以化簡：$\dfrac{\sqrt{2}+1}{\sqrt{2}-1} = \dfrac{(\sqrt{2}+1)^2}{(\sqrt{2}-1)(\sqrt{2}+1)} = (\sqrt{2}+1)^2$，所以該對數等於 $2\ln(1+\sqrt{2})$，而

$$\int_1^\infty \frac{\mathrm{d}u}{1 + u^4} = \frac{\pi}{4\sqrt{2}} - \frac{\ln(1+\sqrt{2})}{2\sqrt{2}} \approx 0.2437.$$

驗算

- 當 $a = 0$ 時，第一條公式給出 $\frac{1}{2\sqrt{2}}\left(\ln\frac{\sqrt{2}-1}{\sqrt{2}+1} + \ln\frac{\sqrt{2}+1}{\sqrt{2}-1}\right) = 0$，在空區間上的積分理應如此。
- 最後的答案是正數，並且小於 $\int_1^\infty u^{-4}\,\mathrm{d}u = \frac{1}{3}$；由於 $1 + u^4 > u^4$，這是合理的。
- 對 $\int_1^\infty \frac{\mathrm{d}u}{1+u^4}$ 使用辛普森法則(Simpson's rule)(經代換 $u = 1/t$ 後，它變為 $\int_0^1 \frac{t^2}{1+t^4}\,\mathrm{d}t$)，得出 0.243748，與公式吻合至小數點後十二位。以 x 表示的無窮積分，數值驗算結果為 0.487495，正是它的兩倍。

常見錯誤

- 對 $\frac{1}{\sqrt{2}c \pm 1}$ 或 $\frac{1}{1+2s^2}$ 積分時，漏掉因子 $\frac{1}{\sqrt{2}}$。
- 代換時沒有更改積分限，或者答案仍以 c 或 s 表示。

- 在最後一部分，忘記 $\mathrm{d}x = \mathrm{d}u/u$，結果得出 $2u/(1+u^4)$ 和一個發散（divergent）的積分。

STEP 之外

最後的結果是一個經典積分的一部分：$\int_0^\infty \frac{\mathrm{d}u}{1+u^4} = \frac{\pi}{2\sqrt{2}}$，而代換 $u \mapsto 1/u$ 顯示，在 $[0,1]$ 和 $[1,\infty)$ 上的兩部分分別是 $\int_0^1 \frac{1}{1+t^4}\,\mathrm{d}t$ 和 $\int_0^1 \frac{t^2}{1+t^4}\,\mathrm{d}t$，它們的和與差都可以用初等方法求出。在大學一年級的課程中，你還會遇到一條更巧妙的途徑，就是圍道積分（contour integration）和留數（residue），它可以一舉求出 $\int_0^\infty \frac{\mathrm{d}u}{1+u^n} = \frac{\pi/n}{\sin(\pi/n)}$。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）是學習這個想法的絕佳讀物。

詳細解答

STEP III 2004，第 2 題

純數學 • 漸近線與曲線描繪

STEP III 2004，第 2 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q2

關鍵思路

把 $f(x)$ 寫成直線 $y = x - 4$ 加上一個修正項。修正項的分子是完全平方（perfect square），所以它在某一點處有二重零點，而這個二重零點正正就是相切（tangency）的意思。至於 $f(x) = 0$ 的根，只要消去分母，便會得到一個平方差（difference of two squares），可以立即因式分解。

解答

在整個解答中，記

$$f(x) = (x-4) - \frac{16(2x+1)^2}{x^2(x-4)}.$$

第 (i) 部分

分母在 $x = 0$ 和 $x = 4$ 處為零。分子 $16(2x+1)^2$ 在這兩處都不為零，所以垂直漸近線（vertical asymptote）是

$$x = 0 \qquad 及 \qquad x = 4.$$

當 $|x|$ 很大時，分式的分子是二次式，分母是三次式：

$$\frac{16(2x+1)^2}{x^2(x-4)} = \frac{64x^2+64x+16}{x^3-4x^2} \approx \frac{64}{x} \to 0 \quad 當 x \to \pm\infty,$$

所以 $f(x) - (x-4) \to 0$，而斜漸近線（oblique asymptote）是

$$y = x - 4.$$

想一想 漸近線怎可能同時又是切線（tangent）？

漸近線只描述曲線在遠處的情況，所以沒有甚麼會阻止這條直線同時在某個有限的點與曲線相切。要找出直線與曲線的交點，設 $\mathrm{f}(x) = x - 4$：

$$\mathrm{f}(x) - (x-4) = -\frac{16(2x+1)^2}{x^2(x-4)} = 0 \iff x = -\tfrac{1}{2}.$$

由於因式 $(2x+1)$ 帶有平方，$x = -\frac{1}{2}$ 是 $\mathrm{f}(x) - (x-4) = 0$ 的二重根（double root）。這表示直線在該處與曲線相切。要直接驗證，設 $D(x) = \mathrm{f}(x) - (x-4) = (2x+1)^2\, h(x)$，其中 $h(x) = -16/\big(x^2(x-4)\big)$。函數 h 在 $x = -\frac{1}{2}$ 附近可導（differentiable），所以

$$D'(x) = 4(2x+1)h(x) + (2x+1)^2 h'(x),$$

它在 $x = -\frac{1}{2}$ 處等於 0。因此 $\mathrm{f}'(-\frac{1}{2}) = 1$，即漸近線的斜率（gradient）。此外 $\mathrm{f}(-\frac{1}{2}) = -\frac{9}{2}$，這一點位於 $y = x - 4$ 上。所以斜漸近線就是曲線在 $\big(-\frac{1}{2}, -\frac{9}{2}\big)$ 處的切線。

第 (ii) 部分

把所有項通分，公分母為 $x^2(x-4)$：

$$\mathrm{f}(x) = \frac{x^2(x-4)^2 - 16(2x+1)^2}{x^2(x-4)} = \frac{\big[x(x-4)\big]^2 - \big[4(2x+1)\big]^2}{x^2(x-4)}.$$

分子是一個平方差：

$$\begin{aligned}\big[x^2 - 4x - 4(2x+1)\big]\big[x^2 - 4x + 4(2x+1)\big] &= (x^2 - 12x - 4)(x^2 + 4x + 4)\\ &= (x^2 - 12x - 4)(x+2)^2.\end{aligned}$$

所以

$$\mathrm{f}(x) = \frac{(x+2)^2(x^2 - 12x - 4)}{x^2(x-4)}.$$

因式 $(x+2)^2$ 顯示 $x = -2$ 是 $\mathrm{f}(x) = 0$ 的二重根，而且 $x = -2$ 並非分母的零點，正如題目所求。另外兩個根來自 $x^2 - 12x - 4 = 0$：

$$x = 6 \pm 2\sqrt{10} \approx 12.32,\ -0.32.$$

第 (iii) 部分

我們先整理繪畫草圖所需的各項特徵。

*在 $x = 0$ 附近。*當 $x \to 0$ 時，$\mathrm{f}(x) \approx -\dfrac{16}{x^2 \cdot (-4)} = \dfrac{4}{x^2}$，所以從兩側趨近時都有 $\mathrm{f}(x) \to +\infty$。

在 $x = 4$ 附近。當 $x \to 4$ 時，$\mathrm{f}(x) \approx -\dfrac{16 \cdot 81}{16(x-4)} = -\dfrac{81}{x-4}$。所以當 $x \to 4^-$ 時 $\mathrm{f}(x) \to +\infty$，而當 $x \to 4^+$ 時 $\mathrm{f}(x) \to -\infty$。

曲線在斜漸近線的哪一側？$\mathrm{f}(x) - (x-4) = -16(2x+1)^2/\big(x^2(x-4)\big)$。當 $x < 4$ 時（$x = 0$ 和 $x = -\frac{1}{2}$ 除外）這個量為正，當 $x > 4$ 時為負。所以在 $x = 4$ 的左方，曲線位於 $y = x - 4$ 之上，只在 $x = -\frac{1}{2}$ 處與它相切；在右方，曲線則位於它之下。

f 的正負號。從因式分解後的形式可見，因式 $(x+2)^2$ 和 x^2 永不為負。所以 f 的正負號與 $(x^2 - 12x - 4)/(x-4)$ 的正負號相同：

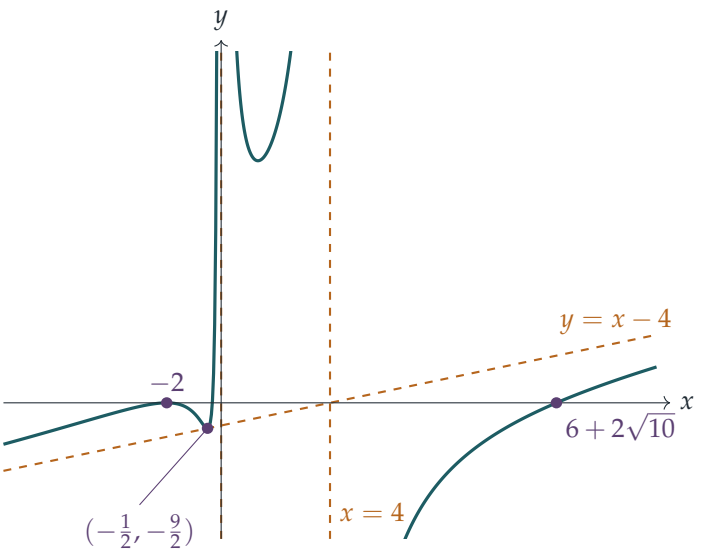
- 當 $x < 6 - 2\sqrt{10}$ 時（$\mathrm{f}(-2) = 0$ 除外），以及當 $4 < x < 6 + 2\sqrt{10}$ 時，$\mathrm{f} < 0$；
- 當 $6 - 2\sqrt{10} < x < 4$ 時（$x \neq 0$），以及當 $x > 6 + 2\sqrt{10}$ 時，$\mathrm{f} > 0$。

曲線在 $x = -2$ 的兩側都保持為負，並在該處與 x 軸相切，所以 $(-2, 0)$ 是局部極大點（local maximum）。

轉向點（*turning point*）。STEP 的草圖並不需要這些轉向點，但有了它們，草圖會更準確。求導可得

$$\mathrm{f}'(x) = 1 + \frac{16(2x+1)(2x^2+3x-8)}{x^3(x-4)^2}.$$

在 $(-2,\, 6 - 2\sqrt{10})$ 上，曲線的起點和終點高度都是 0，中間為負，所以在這區間內有一個極小點。以數值計算，這個極小點約在 $(-0.51,\, -4.50)$，剛好在切點的左方。在 $(0, 4)$ 上，曲線在兩端都趨向 $+\infty$，所以這裏也有一個極小點，約在 $(1.35,\, 42.7)$。在 $\mathrm{f}'(x) = 0$ 中消去分母，得 $(x+2)(x^4 - 10x^3 + 100x^2 - 72x - 64) = 0$。這個四次式恰好只有上述兩個實根，所以再沒有其他轉向點。特別地，f 在整個 $x > 4$ 的範圍內都是遞增的。



這幅草圖是按比例繪畫的，因此 $x = -2$ 與 $x \approx -0.32$ 之間的凹陷看起來很淺。那裏的最低點只在軸下約 4.5 處，而曲線在 $\left(-\frac{1}{2}, -\frac{9}{2}\right)$ 與漸近線相切。

驗算

- $f(-\frac{1}{2}) = -\frac{9}{2} - 0 = -\frac{9}{2}$，這一點在 $y = x - 4$ 上。在 $x = -\frac{1}{2}$ 處以數值方法求導數，得 1.000。
- 展開 $(x+2)^2(x^2 - 12x - 4)$，並與 $x^2(x-4)^2 - 16(2x+1)^2$ 比較：兩者都是 $x^4 - 8x^3 - 48x^2 - 64x - 16$。
- $f(1000) - (1000 - 4) \approx -0.064 \approx -64/1000$，與 $64/x$ 這個估算所預測的一致。

常見錯誤

- 忘記了在 $x = 0$ 附近，曲線在兩側都趨向 $+\infty$，原因是有 x^2 這個因式。兩條垂直漸近線的表現並不相同。
- 證明相切時，只證明了直線與曲線相交。你還需要證明斜率相同，或者等價地，證明有重根。
- 在第 (ii) 部分把四次的分子展開，然後嘗試憑觀察找出它的根，而沒有看出那是一個平方差。

STEP 之外

一條直線與一條曲線相交而出現「二重根」，恰好就是兩者相切的時候。這就是**相交重數**（multiplicity of intersection）的概念，它是代數幾何（algebraic geometry）的基礎，並引出一些結果，例如關於兩條曲線相交多少次的貝祖定理（Bézout's theorem）。有理函數（rational function）在無窮遠處可以用它的多項式部分來近似，而餘項的大小（這裏是 $64/x$）說明了曲線趨近漸近線的速度有多快。這是對漸近展開（asymptotic expansion）的初步體驗。Siklos 的 *Advanced Problems in Mathematics* 有更多這種風格的曲線描繪題目。

詳細解答

STEP III 2004，第 3 題

純數學 • 凸性與級數的界限

STEP III 2004，第 3 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q3

關鍵思路

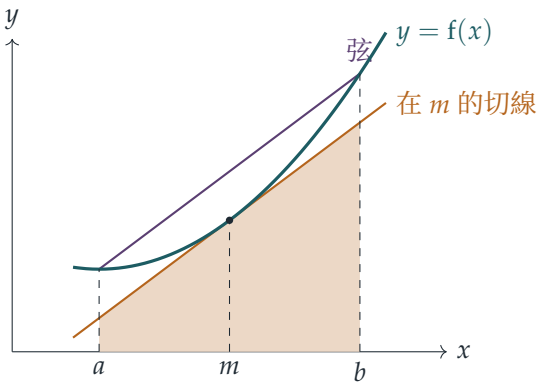
若 $\mathrm{f}'' > 0$，曲線向上彎。因此曲線位於每一條切線（tangent）之上，亦位於每一條弦（chord）之下。在中點的切線，以及連接兩端的弦，各自截出一個梯形（trapezium），其面積很容易寫出，而這兩個面積把積分夾在中間。取 $\mathrm{f}(x) = 1/x^2$ 及區間 $[n-1, n]$，積分便會裂項相消（telescope）。再把各不等式對 n 相加，便得出 $\sum 1/n^2$ 的上下界。

解答

積分的兩個界

設 $m = \frac{1}{2}(a+b)$。由於在 $[a,b]$ 上 $\mathrm{f}'' > 0$，斜率 f' 嚴格遞增，所以曲線是凸的（convex）。由此得到兩個事實。

- *曲線位於它在 m 的切線之上。*設 $t(x) = \mathrm{f}(m) + \mathrm{f}'(m)(x-m)$ 及 $g = \mathrm{f} - t$。則 $g(m) = g'(m) = 0$，且 $g'' = \mathrm{f}'' > 0$，所以 g' 遞增。因此在 m 的左邊 $g' < 0$，在右邊 $g' > 0$，故 g 在 m 取得嚴格極小值 0，而當 $x \neq m$ 時 $\mathrm{f}(x) > t(x)$。
- *曲線位於它的弦之下。*設 $c(x)$ 為通過 $(a, \mathrm{f}(a))$ 和 $(b, \mathrm{f}(b))$ 的直線，並設 $h = \mathrm{f} - c$。則 $h(a) = h(b) = 0$ 且 $h'' > 0$。假如 h 在 (a,b) 內取得某個 ≥ 0 的值，它便會在該區間內有一個極大值，而在該處 $h'' \leq 0$，這是不可能的。所以當 $a < x < b$ 時 $\mathrm{f}(x) < c(x)$。



現在比較 $[a,b]$ 上的面積。切線下的面積是一個闊度為 $b-a$ 的梯形。它的平均高度就是 t 在中點的高度 $t(m)=\mathrm{f}(m)$，所以面積是 $(b-a)\,\mathrm{f}(m)$（陰影部分）。弦下的面積是一個梯形，兩條平行邊的長度為 $\mathrm{f}(a)$ 和 $\mathrm{f}(b)$，所以面積是 $(b-a)\frac{1}{2}\big(\mathrm{f}(a)+\mathrm{f}(b)\big)$。除了個別孤立點外，曲線嚴格位於兩條直線之間。所以

$$(b-a)\,\mathrm{f}\Big(\frac{a+b}{2}\Big) < \int_a^b \mathrm{f}(x)\,\mathrm{d}x < (b-a)\,\frac{\mathrm{f}(a)+\mathrm{f}(b)}{2}.$$

（草圖所示的 f 是正的。但論證其實只用到 $\int(\mathrm{f}-t)>0$ 和 $\int(c-\mathrm{f})>0$，所以 f 的正負並無影響。）

選取 a、b 和 f

中間的一項 $\frac{1}{n-1}-\frac{1}{n}$ 看來像 $\big[-\frac{1}{x}\big]_{n-1}^{n}$，所以取 $\mathrm{f}(x)=1/x^2$、$a=n-1$、$b=n$。於是 $b-a=1$，而且因為 $n-1\geq 1>0$，在 $[n-1,n]$ 上有 $\mathrm{f}''(x)=6/x^4>0$。此外

$$\int_{n-1}^{n}\frac{\mathrm{d}x}{x^2}=\frac{1}{n-1}-\frac{1}{n},\qquad \mathrm{f}\Big(\frac{2n-1}{2}\Big)=\frac{4}{(2n-1)^2},$$

所以由第一部分得

$$\frac{4}{(2n-1)^2} < \frac{1}{n-1}-\frac{1}{n} < \frac{1}{2}\Big(\frac{1}{n^2}+\frac{1}{(n-1)^2}\Big)\qquad (n\geq 2),$$

這正是所要證明的。

推導關於 1 的兩個界

設 $S=\sum_{n=1}^{\infty}\frac{1}{n^2}$。這個級數收斂（converge），因為當 $n\geq 2$ 時有 $\frac{1}{n^2}<\frac{1}{n(n-1)}=\frac{1}{n-1}-\frac{1}{n}$，而這些項會裂項相消。把 $n=2,3,\ldots,N$ 的不等式相加。中間各項裂項相消後得 $1-\frac{1}{N}$，它趨向 1。

? **想一想** 當我們令 $N\to\infty$ 時，嚴格不等式通常會變成 $\leq$。怎樣才能保持嚴格？

看看當中的差距。對每個 $n \geq 2$，差距 $d_n = \left(\frac{1}{n-1} - \frac{1}{n}\right) - \frac{4}{(2n-1)^2}$ 是正數。所以這些差距的無窮和至少等於它的第一項，即 $d_2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} > 0$。這裏每個級數都收斂，所以可以逐項求和，而兩個和之差嚴格為正。同樣的論證亦適用於右邊的不等式。

在左邊，$n = 2, 3, 4, \ldots$ 給出 $2n - 1 = 3, 5, 7, \ldots$，所以

$$4\left(\frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots\right) < 1.$$

在右邊，

$$\sum_{n=2}^{\infty} \frac{1}{2}\left(\frac{1}{n^2} + \frac{1}{(n-1)^2}\right) = \frac{1}{2}(S-1) + \frac{1}{2}S = \frac{1}{2} + (S-1),$$

由此得

$$1 < \frac{1}{2} + \left(\frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots\right).$$

與奇數平方作比較

把左邊的級數每兩個相鄰的項分成一組，每組先是一個奇數項，然後是一個偶數項。對 $k \geq 2$，

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

因為 $2k > 2k - 1 > 0$。取 $k = 2, 3, 4, \ldots$，左邊的各組是 $(3^2, 4^2)$、$(5^2, 6^2), \ldots$，合起來剛好把 $\frac{1}{3^2}, \frac{1}{4^2}, \frac{1}{5^2}, \ldots$ 每一項各用一次。（可以這樣分組，是因為所有項都是正數，而且級數收斂。）右邊則是 $\frac{1}{3^2}, \frac{1}{5^2}, \ldots$。對 k 求和，並如前所述保持嚴格不等，便得

$$\frac{1}{2}\left(\frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \cdots\right) < \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots.$$

S 的上下界

下界（*lower bound*）。由右邊的不等式，$1 < \frac{1}{2} + (S-1)$，所以 $S > \frac{3}{2}$。

上界（*upper bound*）。設 $T = \frac{1}{3^2} + \frac{1}{4^2} + \cdots = S - 1 - \frac{1}{4}$。把剛才的比較與左邊的不等式結合，得

$$\frac{1}{2}T < \frac{1}{3^2} + \frac{1}{5^2} + \cdots < \frac{1}{4},$$

所以 $T < \frac{1}{2}$。因此 $S = 1 + \frac{1}{4} + T < \frac{7}{4}$。總括而言，

$$\frac{3}{2} < \sum_{n=1}^{\infty} \frac{1}{n^2} < \frac{7}{4}.$$

驗算

▶ 真確值是 $S = \pi^2/6 \approx 1.6449$，位於 $(1.5, 1.75)$ 之內。

- 數值上，$4\sum_{k\geq 2}(2k-1)^{-2} = 4(\pi^2/8-1) \approx 0.935 < 1$，而 $\frac{1}{2}+(S-1) \approx 1.145 > 1$。
- 至於那個比較，$\frac{1}{2}T \approx 0.197 < 0.234 \approx \sum_{k\geq 2}(2k-1)^{-2}$。
- 試代 $n=2$：$\frac{4}{9} \approx 0.444 < \frac{1}{2} < \frac{5}{8} = 0.625$。以精確分數對 $2 \leq n \leq 2000$ 逐一檢查，三個不等式全部成立。

常見錯誤

- 忘記了缺少的 $\frac{1}{4}$：括號由 $\frac{1}{3^2}$ 開始，所以 $T = S - \frac{5}{4}$，而不是 $S-1$。若用 $S-1$，便會得出 $S < \frac{3}{2}$ 這個錯誤結論。
- 把不等號的方向弄反。對於向上彎的曲線，切線（中點）估算偏小，而弦（梯形）估算偏大。
- 過渡到無窮和時，不加說明便保留「$<$」，沒有解釋為何它取極限後仍然成立。

STEP 之外

這兩個界就是中點法則（midpoint rule）和梯形法則（trapezium rule）。凸性告訴你每個法則的誤差偏向哪一邊，而泰勒定理（Taylor's theorem）則顯示兩者的誤差都以 $(b-a)^3$ 的速度縮小。精確值 $\sum 1/n^2 = \pi^2/6$ 是歐拉（Euler）對巴塞爾問題（Basel problem）的解答。大學的第一門分析課程會用傅立葉級數（Fourier series）證明它，也會把「嚴格不等式取極限」這一步嚴謹化。Abbott 的 *Understanding Analysis* 涵蓋了級數的收斂和凸函數這兩個課題。

詳細解答

STEP III 2004，第 4 題

純數學 • 相切圓與級數

STEP III 2004，第 4 題

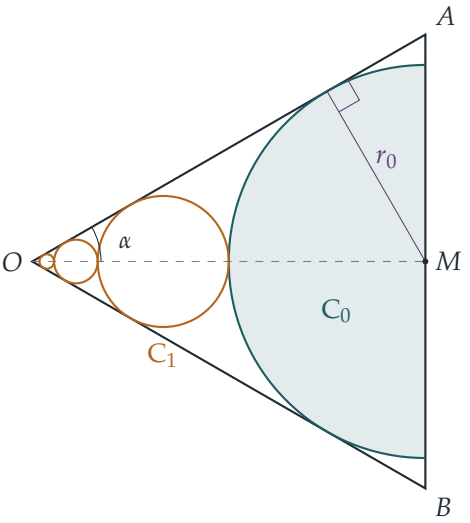
繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q4

關鍵思路

凡是同時與 OA 和 OB 相切的圓，圓心都在角 AOB 的角平分線（angle bisector）上，而圓心與 O 的距離由半徑決定：在 O 處夾角為 α 的直角三角形給出距離 $= r/\sin\alpha$。把「兩圓相切」寫成關於這些距離的方程，便把相鄰的半徑聯繫起來，從而得知各半徑構成等比數列（geometric progression）。於是各面積構成等比級數（geometric series），而最後一部分是單變量的求極大值問題（或者選一個合適的 α 值即可）。

解答



半徑之比

設 M 為 AB 的中點。由於 $OA = OB$，直線 OM 平分角 AOB，並且垂直於 AB；角 AOM 和角 BOM 各為 α。

任取一個半徑為 r、同時與直線 OA 和 OB 相切的圓。它的圓心與這兩條直線等距，所以位於角平分線 OM 上。設圓心為 P，圓與 OA 相切於 T，則 $PT \perp OA$，$PT = r$，且角 $POT = \alpha$，所以在直角三角形 OTP 中

$$OP = \frac{r}{\sin\alpha}.$$

這個結果適用於 C_0（圓心為 M），也適用於每一個 C_n。記 $d_n = r_n/\sin\alpha$ 為 O 到 C_n 圓心的距離。

圓 C_{n+1} 位於 C_n 與 O 之間的角落裏（它必定在角的內部，並在 C_n 靠近 O 的一側，因為它與邊 OA 和 OB 相切，而不是與它們的延長線相切）。兩圓外切（touch externally）時，圓心相距 $r_n + r_{n+1}$，而這裏兩個圓心都在 OM 上，所以

$$d_n - d_{n+1} = r_n + r_{n+1}.$$

代入 $d_n = r_n/\sin\alpha$，再乘以 $\sin\alpha$：

$$\begin{aligned} r_n - r_{n+1} &= (r_n + r_{n+1})\sin\alpha \\ r_n(1-\sin\alpha) &= r_{n+1}(1+\sin\alpha), \end{aligned}$$

因此

$$\frac{r_{n+1}}{r_n} = \frac{1-\sin\alpha}{1+\sin\alpha},$$

這正是所要證明的結果。$n = 0$ 時這個論證同樣成立：C_1 與半圓的弧形部分相切，而這部分正是 C_0 面向 O 的那一半。

總面積

記 $k = \dfrac{1-\sin\alpha}{1+\sin\alpha}$。由於 $0 < \sin\alpha < 1$，我們有 $0 < k < 1$，且 $r_n = k^n r_0$。半圓的面積為 $\frac{1}{2}\pi r_0^2$，而 C_n 的面積為 $\pi k^{2n} r_0^2$。級數 $\sum_{n\geq 1} k^{2n}$ 是公比為 $k^2 < 1$ 的等比級數，所以它收斂（converge）於 $k^2/(1-k^2)$。於是

$$S = \pi r_0^2\left(\frac{1}{2} + \frac{k^2}{1-k^2}\right).$$

現在

$$1 - k^2 = \frac{(1+\sin\alpha)^2 - (1-\sin\alpha)^2}{(1+\sin\alpha)^2} = \frac{4\sin\alpha}{(1+\sin\alpha)^2}, \qquad \text{所以} \qquad \frac{k^2}{1-k^2} = \frac{(1-\sin\alpha)^2}{4\sin\alpha}.$$

因此

$$S = \pi r_0^2 \frac{2\sin\alpha + 1 - 2\sin\alpha + \sin^2\alpha}{4\sin\alpha} = \frac{1+\sin^2\alpha}{4\sin\alpha}\pi r_0^2,$$

這就是我們要的結果。

與三角形比較

由第一部分，$OM = d_0 = r_0/\sin\alpha$。在直角三角形 OMA 中，$AM = OM\tan\alpha$，所以三角形 OAB 的面積為

$$T = OM\cdot AM = OM^2\tan\alpha = \frac{r_0^2}{\sin^2\alpha}\cdot\frac{\sin\alpha}{\cos\alpha} = \frac{r_0^2}{\sin\alpha\cos\alpha}.$$

（半圓確實放得進底邊上：$r_0 = OM\sin\alpha < OM\tan\alpha = AM$。）兩式相除，得

$$\frac{S}{T} = \frac{\pi}{4}(1+\sin^2\alpha)\cos\alpha.$$

想一想 我們應該試哪一個 α 值？

當 $\alpha \to 0$ 時，這個比趨向 $\pi/4 \approx 0.785$，僅僅略小於 0.8，所以非常瘦長的三角形並不可行。很自然的第一個嘗試，是令 O 處為直角，即 $\alpha = \pi/4$。這時 $\sin^2\alpha = \frac{1}{2}$，$\cos\alpha = 1/\sqrt{2}$，所以

$$\frac{S}{T} = \frac{\pi}{4}\cdot\frac{3}{2}\cdot\frac{1}{\sqrt{2}} = \frac{3\pi}{8\sqrt{2}}.$$

這個值大於 $\frac{4}{5}$，當且僅當 $15\pi > 32\sqrt{2}$。兩邊均為正數，所以可以平方：條件變成 $225\pi^2 > 2048$，即 $\pi^2 > 9.102\ldots$。由於 $\pi > 3.1$，我們有 $\pi^2 > 9.61$，所以不等式成立。因此當 $\alpha = \pi/4$ 時（並且根據連續性（continuity），對所有接近 $\pi/4$ 的 α 亦然），這些圓覆蓋了三角形超過五分之四的面積。

順帶一提，最佳的 α 可由求 $f(c) = (2-c^2)c$ 的極大值得出，其中 $c = \cos\alpha$：$f'(c) = 2 - 3c^2 = 0$ 給出 $\cos\alpha = \sqrt{2/3}$，即 $\sin\alpha = 1/\sqrt{3}$，而這個比的極大值為 $\pi\sqrt{6}/9 \approx 0.855$。

驗算

- 當 $\alpha \to \pi/2$ 時，三角形變得扁平，且 $k \to 0$：各圓 C_n 消失，$S \to \frac{1}{2}\pi r_0^2$，與 $\frac{1+1}{4}\pi r_0^2$ 吻合。
- 取 $\alpha = \pi/6$，則 $k = \frac{1}{3}$，$S = \pi r_0^2\left(\frac{1}{2} + \frac{1/9}{8/9}\right) = \frac{5}{8}\pi r_0^2$；公式給出 $\frac{1+1/4}{2}\pi r_0^2 = \frac{5}{8}\pi r_0^2$。
- 以數值方法驗證：對隨機選取的 α，直接按相切條件構造 3000 個圓，所得結果與兩條公式吻合，誤差約為 10^{-15}；而 $\alpha = \pi/4$ 時的比為 0.8330。

常見錯誤

- 把 C_0 當作整個圓來計算：它是半圓，這正是 $\frac{1}{2}$ 出現的原因。
- 使用 $d_n + d_{n+1} = r_n + r_{n+1}$ 或 $d_n - d_{n+1} = r_n - r_{n+1}$：這些圓是外切的，而且兩個圓心都在 O 的同一側。
- 在最後一部分嘗試很小的 α：這時比趨向 $\pi/4 < \frac{4}{5}$。

STEP 之外

這一串圓是由單一個以 O 為中心、比例因子為 k 的位似變換（dilation，即放大 enlargement）生成的：它把角落區域映射到自身，並把每個圓映射到下一個圓。這類自相似（self-similar）的構形通向分形（fractal）和阿波羅尼奧斯圓填充（Apollonian circle packing），其中互相相切的圓的曲率（curvature）滿足笛卡兒圓定理（Descartes' circle theorem）。反演（inversion）是一種把圓變成直線的變換，能令許多這類相切問題變得容易；Needham 的 *Visual Complex Analysis* 對此有精彩的論述。

詳細解答

STEP III 2004，第 5 題

純數學 • 三角方程

STEP III 2004，第 5 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q5

關鍵思路

兩個餘弦值相等，當且僅當兩個角在相差一個正負號及 2π 的整數倍之下相同。改用正切（tangent）之後，用來區分 θ 與 $\theta+\pi$ 的資訊便會丟失，所以這個推論只在一個方向成立，而這正是反例（counterexample）的用意所在。在第 (i) 及 (ii) 部分，我們把 $a\cos x + b\sin x$ 寫成 $R\cos(x-\alpha)$ 的形式，利用第一個結果以 $\tan\omega$ 表示 $\tan x$，然後根據各角的大小選出正確的象限（quadrant）。

解答

開首的結果

若 $\cos(x-\alpha)=\cos\beta$，則對某個整數 n 有 $x-\alpha=\pm\beta+2n\pi$，因為餘弦是偶函數（even function），且週期（period）為 2π。所以

$$x=\alpha+\beta+2n\pi \quad \text{或} \quad x=\alpha-\beta+2n\pi.$$

由於 tan 的週期為 π，2π 當然也是它的週期，由此可得 $\tan x=\tan(\alpha+\beta)$ 或 $\tan x=\tan(\alpha-\beta)$（只要這些正切值有定義）。

逆命題（converse）不成立，因為 tan 的週期為 π，而 cos 卻不是。取 $\alpha=\beta=0$ 及 $x=\pi$。這時 $\tan x=0=\tan(\alpha+\beta)$，但

$$\cos(x-\alpha)=\cos\pi=-1\neq 1=\cos\beta.$$

一般而言，$x=\alpha+\beta+\pi$ 會給出 $\cos(x-\alpha)=-\cos\beta$。

關於 ω 有兩項事實稍後會用得着。由於 $\tan\omega=\frac{4}{3}$，我們有 $\cos\omega=\frac{3}{5}$ 及 $\sin\omega=\frac{4}{5}$。又

由於 $\frac{4}{3} > 1$，我們亦有 $\frac{\pi}{4} < \omega < \frac{\pi}{2}$。

第 (i) 部分

這裏 $\sqrt{1^2+7^2} = 5\sqrt{2}$，所以我們把等式兩邊除以 $5\sqrt{2}$：

$$\tfrac{1}{5\sqrt{2}}\cos x - \tfrac{7}{5\sqrt{2}}\sin x = \tfrac{1}{\sqrt{2}}.$$

設 φ 為滿足 $\cos\varphi = \frac{1}{5\sqrt{2}}$ 及 $\sin\varphi = \frac{7}{5\sqrt{2}}$ 的銳角（acute angle），於是 $\tan\varphi = 7$。方程變成

$$\cos(x+\varphi) = \cos\tfrac{\pi}{4},$$

這正是開首的結果中取 $\alpha = -\varphi$ 及 $\beta = \frac{\pi}{4}$ 的情況。利用 $\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A \tan B}$，

$$\tan(\alpha+\beta) = \tan\left(\tfrac{\pi}{4} - \varphi\right) = \frac{1-7}{1+7} = -\frac{3}{4},$$
$$\tan(\alpha-\beta) = -\tan\left(\varphi + \tfrac{\pi}{4}\right) = -\frac{7+1}{1-7} = \frac{4}{3}.$$

想一想 這兩個正切值各自對應 $[0, 2\pi]$ 內的兩個角，但方程在該範圍內只有兩個解。是哪兩個呢？

正如反例所示，單憑正切值無法判斷，所以我們回到 $x = \alpha \pm \beta + 2n\pi$，找出 $\alpha \pm \beta$ 實際上位於何處。由於 $\tan\varphi = 7 > 1$，我們有 $\frac{\pi}{4} < \varphi < \frac{\pi}{2}$。

- $\alpha + \beta = \frac{\pi}{4} - \varphi$ 位於 $(-\frac{\pi}{4}, 0)$ 內，其正切值為 $-\frac{3}{4}$。而 $\tan(\frac{\pi}{2} - \omega) = \cot\omega = \frac{3}{4}$，且 $\frac{\pi}{2} - \omega \in (0, \frac{\pi}{4})$。因此 $\alpha + \beta = \omega - \frac{\pi}{2}$，加上 2π 便得 $x = \frac{3\pi}{2} + \omega$。
- $\alpha - \beta = -\varphi - \frac{\pi}{4}$ 位於 $(-\frac{3\pi}{4}, -\frac{\pi}{2})$ 內，其正切值為 $\frac{4}{3}$。角 $\omega - \pi$ 同樣位於這個區間內，正切值亦為 $\frac{4}{3}$。在長度小於 π 且不包含漸近線（asymptote）的區間上，正切函數是一對一（one-to-one）的，所以 $\alpha - \beta = \omega - \pi$。加上 2π 便得 $x = \pi + \omega$。

由於 $0 < \omega < \frac{\pi}{2}$，這兩個值都在 $[0, 2\pi]$ 內，而 n 取其他值時都不會得出此範圍內的值。因此

$$x = \pi + \omega \quad \text{或} \quad x = \tfrac{3\pi}{2} + \omega.$$

驗證一下：當 $x = \pi + \omega$ 時，$\cos x - 7\sin x = -\frac{3}{5} + \frac{28}{5} = 5$；當 $x = \frac{3\pi}{2} + \omega$ 時，$\cos x = \sin\omega$ 及 $\sin x = -\cos\omega$，由此得 $\frac{4}{5} + \frac{21}{5} = 5$。

第 (ii) 部分

這裏 $\sqrt{2^2+11^2} = \sqrt{125} = 5\sqrt{5}$，所以

$$\tfrac{2}{5\sqrt{5}}\cos x + \tfrac{11}{5\sqrt{5}}\sin x = \tfrac{2}{\sqrt{5}}.$$

設 α 為銳角，滿足 $\cos\alpha = \frac{2}{5\sqrt{5}}$ 及 $\sin\alpha = \frac{11}{5\sqrt{5}}$，於是 $\tan\alpha = \frac{11}{2}$。設 β 為銳角，滿足 $\cos\beta = \frac{2}{\sqrt{5}}$，於是 $\sin\beta = \frac{1}{\sqrt{5}}$ 及 $\tan\beta = \frac{1}{2}$。方程就是 $\cos(x-\alpha) = \cos\beta$，而

$$\tan(\alpha-\beta) = \frac{\frac{11}{2} - \frac{1}{2}}{1 + \frac{11}{4}} = \frac{5}{15/4} = \frac{4}{3},$$
$$\tan(\alpha+\beta) = \frac{\frac{11}{2} + \frac{1}{2}}{1 - \frac{11}{4}} = \frac{6}{-7/4} = -\frac{24}{7}.$$

第二個值就是 $\tan 2\omega$，因為

$$\tan 2\omega = \frac{2\cdot\frac{4}{3}}{1 - \frac{16}{9}} = \frac{8/3}{-7/9} = -\frac{24}{7}.$$

我們再次確定各角的位置。α 和 β 都是銳角，所以

- $\alpha - \beta \in (-\frac{\pi}{2}, \frac{\pi}{2})$，正切值為 $\frac{4}{3}$，所以 $\alpha - \beta = \omega$；
- $\alpha + \beta \in (0, \pi)$，正切值為負，所以它位於 $(\frac{\pi}{2}, \pi)$ 內。由於 $\frac{\pi}{4} < \omega < \frac{\pi}{2}$，角 2ω 同樣位於 $(\frac{\pi}{2}, \pi)$ 內，且正切值相同，所以 $\alpha + \beta = 2\omega$。

ω 和 2ω 都在 $[0, 2\pi]$ 內，而加上 $\pm 2\pi$ 便會超出這個範圍。因此方程的解為

$$x = \omega \quad 及 \quad x = 2\omega,$$

其中一個解是另一個的兩倍，正如題目所求。直接驗算：$x = \omega$ 給出 $\frac{6}{5} + \frac{44}{5} = 10$。當 $x = 2\omega$ 時，$\cos 2\omega = \frac{9-16}{25} = -\frac{7}{25}$ 及 $\sin 2\omega = \frac{24}{25}$，由此得 $-\frac{14}{25} + \frac{264}{25} = 10$。

驗算

- 每條方程在 $[0, 2\pi]$ 內應該恰好有兩個根，因為 $5 < 5\sqrt{2}$ 及 $10 < 5\sqrt{5}$，所以右邊的值嚴格介乎 $-R$ 與 R 之間。
- 以數值計算，$\omega \approx 0.9273$。用變號搜尋法可找到 (i) 的根為 4.0689 及 5.6397，即 $\pi + \omega$ 及 $\frac{3\pi}{2} + \omega$。對於 (ii)，則找到 0.9273 及 1.8546，即 ω 及 2ω。

常見錯誤

- 解 $\tan x = \frac{4}{3}$ 或 $\tan x = -\frac{3}{4}$，然後把 $[0, 2\pi]$ 內的四個角全部列出。其中兩個是增根，原因正是反例所揭示的那一點。
- 在 (i) 中把 φ 的正負號弄錯，例如把 $\cos x - 7\sin x$ 寫成 $\cos(x - \varphi)$ 而不是 $\cos(x + \varphi)$。

STEP 之外

這裏的困難在於 tan 不是單射（injective），所以「$\tan x = \tan y$」只能告訴你 x 模 π 的值。在複分析（complex analysis）中，同樣的問題出現在輻角（argument）以及 $\log z$ 的分支（branch）的定義之中：在那裏，角只能確定到相差 2π 的整數倍。在這個框架下，$a\cos x + b\sin x = \mathrm{Re}\left((a - \mathrm{i}b)\mathrm{e}^{\mathrm{i}x}\right)$，而 R 和 α 不過是 $a + \mathrm{i}b$ 的模（modulus）和輻角。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）把這種幾何觀點闡述得十分精彩。

詳細解答

STEP III 2004，第 6 題

純數學 • 遞推關係

STEP III 2004，第 6 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q6

關鍵思路

F_n 這個式子是附在數列每一對相鄰兩項上的量，而開首的恆等式（identity）正好告訴我們，由一對相鄰項到下一對時，它在甚麼情況下保持不變。對於 (i) 中的遞推關係（recurrence），第二個因式為零，所以 F_n 是一個不變量（invariant），我們只需求出它在起點的值。第 (ii) 部分把論證倒過來做：若 F_n 是常數，乘積必定為零，所以每一步兩個因式之中總有一個是零。於是每一步都有一個選擇，而週期（period）為 4 的數列（sequence）就是靠選得巧妙而得出的。

解答

恆等式

把兩項都寫出來，w_{n-1}^2 項互相抵消：

$$\begin{aligned} F_n - F_{n-1} &= \left(w_n^2 + w_{n-1}^2 - 4w_n w_{n-1}\right) - \left(w_{n-1}^2 + w_{n-2}^2 - 4w_{n-1}w_{n-2}\right) \\ &= \left(w_n^2 - w_{n-2}^2\right) - 4w_{n-1}\left(w_n - w_{n-2}\right) \\ &= (w_n - w_{n-2})(w_n + w_{n-2}) - 4w_{n-1}(w_n - w_{n-2}) \\ &= (w_n - w_{n-2})(w_n + w_{n-2} - 4w_{n-1}), \end{aligned}$$

此式對 $n \geq 2$ 成立（因為需要 w_{n-2} 存在），正是所求。

還有一點觀察，往後會一直用到：F_n 對 w_n 和 w_{n-1} 是對稱（symmetric）的。把兩者對調，$w_n^2 + w_{n-1}^2 - 4w_n w_{n-1}$ 保持不變。

第 (i) 部分

取 $w_n = u_n$，並設 F_n 為相應的量。當 $n \geq 2$ 時，遞推關係給出 $u_n + u_{n-2} - 4u_{n-1} = 0$，所以恆等式中的第二個因式為零，於是

$$F_n = F_{n-1} \quad (n \geq 2).$$

因此對所有 $n \geq 1$ 都有 $F_n = F_1$，而

$$F_1 = u_1^2 + u_0^2 - 4u_1u_0 = 4 + 1 - 8 = -3.$$

所以 $u_n^2 + u_{n-1}^2 - 4u_nu_{n-1} = -3$，即對所有 $n \geq 1$ 都有 $u_n^2 + u_{n-1}^2 = 4u_nu_{n-1} - 3$，正是所求。（這個數列是 $1, 2, 7, 26, 97, \ldots$。）

第 (ii)(a) 部分

把 $n = 1$ 和 $v_0 = 1$ 代入 $(*)$：

$$v_1^2 + 1 = 4v_1 - 3 \iff v_1^2 - 4v_1 + 4 = 0 \iff (v_1 - 2)^2 = 0,$$

所以 $v_1 = 2$。這裏沒有選擇的餘地，因為它是重根（repeated root）。

現在取 $w_n = v_n$。條件 $(*)$ 說的正是：對所有 $n \geq 1$，$F_n = -3$。所以當 $n \geq 2$ 時，F_n 和 F_{n-1} 都等於 -3，由恆等式得

$$0 = F_n - F_{n-1} = (v_n - v_{n-2})(v_n + v_{n-2} - 4v_{n-1}).$$

兩個實數的乘積為零，僅當其中一個為零，所以 $v_n = v_{n-2}$ 或 $v_n = 4v_{n-1} - v_{n-2}$，正是所求。

想一想 這樣就完了嗎？兩個選項是否真的都給出合乎條件的下一項？

是的。固定 v_{n-1}，條件 $(*)$ 就是以 $x = v_n$ 為未知數的二次方程(quadratic)$x^2 - 4v_{n-1}x + (v_{n-1}^2 + 3) = 0$。我們知道 $x = v_{n-2}$ 是一個根，因為 F 是對稱的，而且 $F_{n-1} = -3$。兩根之和為 $4v_{n-1}$，所以另一個根是 $4v_{n-1} - v_{n-2}$。因此 (a) 中的兩個選項恰好就是這兩個根。每一步我們都可以任選其一，所得結果總是滿足 $(*)$。（當兩根重合時，例如 $n = 1$ 的情況，就只有一個選項。）

第 (ii)(b) 部分

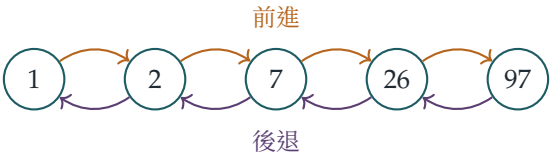
這裏 $v_0 = 1$，符合要求，而每一對相鄰項 (v_{n-1}, v_n) 不是 $(1, 2)$ 就是 $(2, 1)$。由於 F 是對稱的，兩者都給出

$$v_n^2 + v_{n-1}^2 - 4v_nv_{n-1} = 1 + 4 - 8 = -3,$$

所以 $(*)$ 對所有 $n \geq 1$ 成立。用 (a) 的說法，這個數列每一步都取 $v_n = v_{n-2}$ 這個選項。

第 (ii)(c) 部分

我們有 $v_0 = 1$ 和 $v_1 = 2$。由 (a)，其後每一項要麼等於前兩位的那一項（「後退」），要麼等於 $4\times$（前一項）$-$（再前一項）（「前進」）。由 $(1,2)$ 出發，前進得出 7，然後是 26、97，如此類推，這就是數列 u_n。後退則把最近一次前進抵消。



所以這個數列就是沿着這條鏈的一次遊走，由 1 出發，每次移到相鄰的一個數。由於 1 只有一個鄰居，遊走到達 1 之後必定立刻走到 2。

要在 $v_0 = 1$ 的情況下得到週期 4，我們需要 $v_4 = 1$。(b) 中週期為 2 的數列只是在 1 和 2 之間來回往返。要做出新的東西，我們多走一步：

$$v_0 = 1, \quad v_1 = 2, \quad v_2 = 4\cdot 2 - 1 = 7, \quad v_3 = v_1 = 2, \quad v_4 = v_2 \text{ 或 } 4\cdot 2 - 7 = 1.$$

選取 $v_4 = 1$，然後不斷重複，得出

$$v_n = \begin{cases} 1 & n \equiv 0 \pmod 4, \\ 2 & n \text{ 為奇數}, \\ 7 & n \equiv 2 \pmod 4, \end{cases} \qquad \text{即} \quad 1,2,7,2,1,2,7,2,\ldots$$

直接驗證 $(*)$：每一對相鄰項都是 $(1,2)$、$(2,7)$、$(7,2)$ 或 $(2,1)$。由對稱性，我們只需驗證

$$1 + 4 - 8 = -3 \quad \text{和} \quad 4 + 49 - 56 = -3,$$

兩式都成立。所以 $(*)$ 對所有 $n \geq 1$ 成立，而這個數列的週期為 4（而不是 2，因為 $v_2 \neq v_0$）。

事實上，週期恰好為 4 的答案只有這一個。在 1,2 之後，遊走必須有 $v_4 = 1$，所以它只能在 v_2 到達 7，然後必須立刻折返。

驗算

- 計算 u_n 直至 $n = 40$，確認每一次都有 $u_n^2 + u_{n-1}^2 - 4u_nu_{n-1} = -3$。例如 $(26,97)$：$676 + 9409 - 10088 = -3$。
- 用數以千計的隨機整數三元組檢驗開首的恆等式。
- 由 1,2 開始，列出 (a) 所容許的每一個數列，直至 v_8。當中自始至終滿足 $v_{k+4} = v_k$ 的只有 $1,2,1,2,\ldots$ 和 $1,2,7,2,\ldots$。

常見錯誤

- 在 (ii) 中斷定 $F_n - F_{n-1} = 0$ 卻沒有說明原因。原因是 $(*)$ 令每一個 F_n 都等於 -3。
- 忘記 v_1 是唯一確定的。該二次方程有重根 2，所以無需分情況討論。
- 在 (c) 中只檢查了一個週期之內的各對相鄰項，而忘記了「首尾相接」的那一對 $(v_3, v_4) = (2, 1)$。

STEP 之外

F_n 這個量是遞推關係的一個不變量：一個守恆量，就像力學中的能量。方程 $x^2 + y^2 - 4xy = -3$ 是一個馬爾可夫（Markov）型（或佩爾（Pell）型）的丟番圖方程（Diophantine equation）。把二次方程的一個根換成另一個根，即 $x \mapsto 4y - x$，正是經典的「韋達跳躍」（Vieta jumping）手法，它能由最小的解生成方程的全部整數解。Davenport 的 *The Higher Arithmetic*（高等算術）講述了佩爾方程，以及如何用遞推關係求出它的解。

詳細解答

STEP III 2004，第 7 題

純數學 • 歸約公式

STEP III 2004，第 7 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q7

關鍵思路

I_{n+1} 的被積函數（integrand）等於 I_n 的被積函數乘以 $\frac{t}{t+1}$，所以只要估計這一個因子的上界（upper bound），就能得到聯繫相鄰兩個積分的不等式。分部積分法（integration by parts）也給出一個精確的關係式；把不等式與等式並排放在一起，就能擠出一個只涉及 I_n 的上界。這個精確的關係式可以逐項相消（telescope），一直化到 I_1，而它是我們熟悉的對數；於是整件事就變成 $\ln 2$ 的一個級數（series），而且誤差項在我們掌握之中。

解答

不等式 $I_{n+1} < \frac{1}{2} I_n$

在 $0 \le t \le 1$ 上，

$$\frac{t}{t+1} = 1 - \frac{1}{t+1},$$

它隨 t 遞增，所以最大值是 $\frac{1}{2}$，而且只在 $t = 1$ 時取得。現在把 I_{n+1} 的被積函數寫成

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

因此

$$\frac{1}{2} I_n - I_{n+1} = \int_0^1 \left(\frac{1}{2} - \frac{t}{t+1} \right) \frac{t^{n-1}}{(t+1)^n} \, \mathrm{d}t.$$

這裏的被積函數是連續（continuous）的，永不為負，而且在 $0 < t < 1$ 時嚴格為正（兩個因子在該處均為正）。所以這個積分嚴格為正，即 $I_{n+1} < \frac{1}{2} I_n$。

? **想一想** 既然 $\frac{t}{t+1}$ 確實會達到 $\frac{1}{2}$，為甚麼不等式仍然是嚴格的？

它只在 $t = 1$ 這一點達到 $\frac{1}{2}$。一個連續的非負函數，只要在整個區間上為正，其積分就是正數，所以單單一點取等號並無影響。

遞推關係

對 I_{n+1} 作分部積分，取 $u = t^n$ 及 $\dfrac{\mathrm{d}v}{\mathrm{d}t} = (t+1)^{-(n+1)}$，於是 $v = -\dfrac{1}{n(t+1)^n}$（因為 $n \geq 1$，此式成立）：

$$\begin{aligned} I_{n+1} &= \Big[-\frac{t^n}{n(t+1)^n}\Big]_0^1 + \int_0^1 \frac{nt^{n-1}}{n(t+1)^n}\,\mathrm{d}t \\ &= -\frac{1}{n\,2^n} + I_n, \end{aligned}$$

這是因為方括號內的式子在 $t = 1$ 時等於 $-\frac{1}{n2^n}$，在 $t = 0$ 時等於 0。這就是所要求的結果，即題目中的遞推關係（recurrence）。

I_n 的上界

把遞推關係代入不等式：

$$I_n - \frac{1}{n\,2^n} = I_{n+1} < \frac{1}{2}I_n \quad\Longrightarrow\quad \frac{1}{2}I_n < \frac{1}{n\,2^n} \quad\Longrightarrow\quad I_n < \frac{1}{n\,2^{n-1}}.$$

$\ln 2$ 的級數

首先，

$$I_1 = \int_0^1 \frac{1}{t+1}\,\mathrm{d}t = \big[\ln(t+1)\big]_0^1 = \ln 2.$$

遞推關係指出，對每個 $r \geq 1$ 都有 $I_r - I_{r+1} = \dfrac{1}{r\,2^r}$。把 $r = 1, \ldots, n$ 的各式相加，左邊逐項相消：

$$I_1 - I_{n+1} = \sum_{r=1}^{n} \frac{1}{r\,2^r}, \qquad \text{所以} \qquad \ln 2 = \sum_{r=1}^{n} \frac{1}{r\,2^r} + I_{n+1}.$$

上下界 $\frac{2}{3} < \ln 2 < \frac{17}{24}$

關於餘項（remainder），我們知道兩件事：$I_{n+1} > 0$，因為它的被積函數在 $(0,1]$ 上為正；以及 $I_{n+1} < \dfrac{1}{(n+1)2^n}$，這是把上文的上界中的 n 換成 $n+1$ 而得。所以對每個 n，

$$\sum_{r=1}^{n} \frac{1}{r\,2^r} < \ln 2 < \sum_{r=1}^{n} \frac{1}{r\,2^r} + \frac{1}{(n+1)2^n}.$$

部分和（partial sums）依次為

$$\frac{1}{2}, \qquad \frac{1}{2} + \frac{1}{8} = \frac{5}{8}, \qquad \frac{5}{8} + \frac{1}{24} = \frac{16}{24} = \frac{2}{3}.$$

想一想 求上界和下界時，各應取哪個 n 值？

兩者並不相同。求下界（lower bound）時取 $n=3$：$\ln 2 > \frac{2}{3}$。求上界時取 $n=2$，此時誤差的上界是 $\frac{1}{3\cdot 4}=\frac{1}{12}$：

$$\ln 2 < \tfrac{5}{8}+\tfrac{1}{12}=\tfrac{15}{24}+\tfrac{2}{24}=\tfrac{17}{24}.$$

合起來便得 $\frac{2}{3}<\ln 2<\frac{17}{24}$，正是所求。

驗算

以數值計算，$\ln 2\approx 0.6931$，確實介乎 $\frac{2}{3}\approx 0.6667$ 與 $\frac{17}{24}\approx 0.7083$ 之間。用辛普森法則（Simpson's rule）可得 $I_3\approx 0.0681$，明顯小於上界 $\frac{1}{12}\approx 0.0833$，而且 $\frac{5}{8}+0.0681=0.6931$ 亦確實成立。我們也用數值方法驗算了遞推關係和兩個不等式，n 的取值直至 24。

常見錯誤

- 求上下界時用同一個 n：取 $n=3$ 會得到上界 $\frac{2}{3}+\frac{1}{32}=\frac{67}{96}$，這個上界是對的，而且更精確，但並不是題目要求的 $\frac{17}{24}$。
- 分部積分時弄錯邊界項的正負號；v 是負的。
- 忘記說明為甚麼 $I_{n+1}>0$，而嚴格的下界正需要這一點。

STEP 之外

這就是級數 $\ln 2=\sum_{r\ge 1}\frac{1}{r2^r}$，亦即 $-\ln(1-x)=\sum x^r/r$ 在 $x=\frac{1}{2}$ 時的情況，並且附有取 n 項之後餘項的明確上界。用積分來估計餘項，正是帶積分餘項的泰勒定理（Taylor's theorem）的做法，那是大學一年級分析課的基本內容。關於冪級數（power series）及其餘項的理論，可參閱 Abbott 的 *Understanding Analysis*。

詳細解答

STEP III 2004，第 8 題

純數學 • 微分方程

> **STEP III 2004，第 8 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/04-s3-q8

關鍵思路

由於有 $1/y$ 這一項，方程對 y 而言並非線性（linear）。但把方程兩邊同乘 y 後，$y\frac{\mathrm{d}y}{\mathrm{d}x}$ 就變成 y^2 的導數的一半，而 $1/y$ 項則變成純粹是 x 的函數。因此以 $u = y^2$ 表示時，方程是線性的（這是一條伯努利型（Bernoulli type）方程），再用積分因子（integrating factor）便可完成求解。三條曲線都是圓錐曲線（conics），辨認出它們之後，作圖就很容易。

解答

代換

設 $u = y^2$。由鏈式法則（chain rule），

$$\frac{\mathrm{d}u}{\mathrm{d}x} = 2y\frac{\mathrm{d}y}{\mathrm{d}x} = 2y\left(f(x)\,y + \frac{g(x)}{y}\right) = 2f(x)\,y^2 + 2g(x).$$

以 u 取代 y^2，

$$\frac{\mathrm{d}u}{\mathrm{d}x} - 2f(x)\,u = 2g(x),$$

這對 u 而言是線性的：u 和 $\frac{\mathrm{d}u}{\mathrm{d}x}$ 都只以一次冪出現，而係數只取決於 x。這正是我們要證明的。

解方程

這裏 $f(x) = \frac{1}{x}$ 而 $g(x) = -1$，所以線性方程是

$$\frac{\mathrm{d}u}{\mathrm{d}x} - \frac{2}{x}\,u = -2.$$

積分因子是 $\exp\left(-\int \frac{2}{x}\,\mathrm{d}x\right) = \mathrm{e}^{-2\ln|x|} = \dfrac{1}{x^2}$。把方程兩邊同乘這個因子，

$$\frac{1}{x^2}\frac{\mathrm{d}u}{\mathrm{d}x} - \frac{2}{x^3}u = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{u}{x^2}\right) = -\frac{2}{x^2}.$$

積分後得 $\dfrac{u}{x^2} = \dfrac{2}{x} + C$，其中 C 為常數，所以通解（general solution）是

$$y^2 = 2x + Cx^2.$$

（原方程要求 $x \neq 0$ 及 $y \neq 0$，所以每個解都定義在一個不含 0 的 x 區間上，在該區間上 $2x + Cx^2 > 0$，而且 y 始終保持同一符號。）

三條曲線

三個點都滿足 $x > 0$ 及 $y > 0$，所以每條解曲線（solution curve）都是 $y^2 = 2x + Cx^2$ 在包含該點的正 x 區間上 $y > 0$ 的那一部分。

通過 $(1,1)$：$1 = 2 + C$，所以 $C = -1$，$y^2 = 2x - x^2$。配方後得

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

即圓心為 $(1,0)$、半徑為 1 的圓。解曲線是上半圓 $y = \sqrt{2x - x^2}$，其中 $0 < x < 2$。

通過 $(2,2)$：$4 = 4 + 4C$，所以 $C = 0$，$y^2 = 2x$：這是頂點在原點、對稱軸沿 x 軸的拋物線。解曲線是 $y = \sqrt{2x}$，其中 $x > 0$。

通過 $(4,4)$：$16 = 8 + 16C$，所以 $C = \frac{1}{2}$，$y^2 = 2x + \frac{1}{2}x^2$。配方後得 $\frac{1}{2}(x+2)^2 - y^2 = 2$，即

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

這是一條雙曲線，中心為 $(-2,0)$，頂點為 $(0,0)$ 和 $(-4,0)$，漸近線（asymptotes）為 $y = \pm\frac{1}{\sqrt{2}}(x+2)$。解曲線是右支的上半部分，$y = \sqrt{2x + \frac{1}{2}x^2}$，其中 $x > 0$。

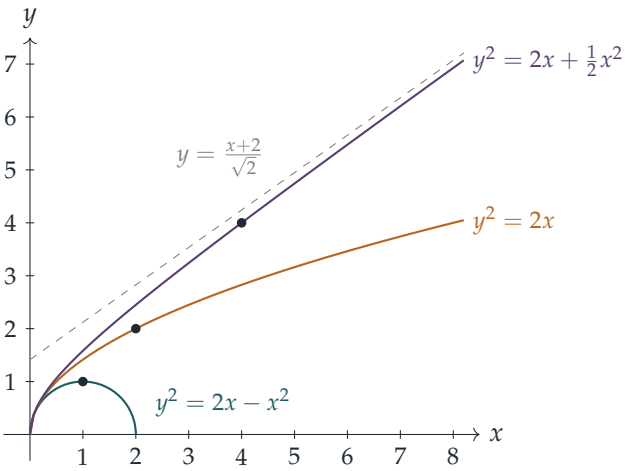
> **想一想** 為甚麼不包括雙曲線的左支，或者各曲線的下半部分？

微分方程的解必須在某個區間上可導，並在該區間上滿足方程，而方程在 $x = 0$ 或 $y = 0$ 時沒有定義。從 $(4,4)$ 出發向左走，曲線會到達 $(0,0)$，在那裏兩個條件都不成立，所以解不能延拓過這一點。左支（$x \leq -4$）和下半部分雖然滿足 $y^2 = 2x + \frac{1}{2}x^2$，但它們是另外的解，並不是通過 $(4,4)$ 的那一個。

略圖

三條曲線都從原點出發，在該處的切線是鉛垂的（在 $x = 0$ 附近，每條曲線的表現都近似 $y \approx \sqrt{2x}$），而在所給各點的斜率分別是 $\frac{1}{1} - \frac{1}{1} = 0$、$\frac{2}{2} - \frac{1}{2} = \frac{1}{2}$ 和 $\frac{4}{4} - \frac{1}{4} = \frac{3}{4}$。

當 x 較大時，圓會回落到 $(2,0)$，拋物線按 $\sqrt{x}$ 的速度增長，而雙曲線則趨近其漸近線 $y=\frac{1}{\sqrt{2}}(x+2)$。由於當 $x>0$ 時 $2x-x^2<2x<2x+\frac{1}{2}x^2$，圓位於拋物線之下，而拋物線又位於雙曲線之下。



驗算

- 代回原方程：若 $y^2=2x+Cx^2$，則 $2yy'=2+2Cx$，所以 $y'=\frac{1+Cx}{y}$，而 $\frac{y}{x}-\frac{1}{y}=\frac{y^2-x}{xy}=\frac{x+Cx^2}{xy}=\frac{1+Cx}{y}$。兩者一致。
- 用數值方法（龍格－庫塔法（Runge–Kutta））對原方程由 $(4,4)$ 逐步計算至 $x=8$，得 $y\approx 6.92820$，與 $\sqrt{16+32}=\sqrt{48}$ 吻合；由 $(1,1)$ 和 $(2,2)$ 出發作類似計算，結果與半圓和拋物線吻合，誤差約為 10^{-12}。

常見錯誤

- 把積分因子寫成 x^2 而不是 x^{-2}：u 的係數是 $-\frac{2}{x}$，所以指數是 $-2\ln x$。
- 漏掉 $\frac{\mathrm{d}u}{\mathrm{d}x}=2y\frac{\mathrm{d}y}{\mathrm{d}x}$ 中的因子 2，結果得出 $y^2=x+Cx^2$；這通不過代回原方程的驗算。
- 把整個圓、拋物線的上下兩半或雙曲線的兩支都畫出來，當作「解曲線」。

STEP 之外

這裏的技巧適用於任何伯努利方程 (Bernoulli equation) $y'=f(x)y+g(x)y^n$：代換 $u=y^{1-n}$ 總能把它化為線性方程（這裏 $n=-1$，所以 $u=y^2$）。在 $x=0$ 和 $y=0$ 處所需的謹慎，讓我們初次窺見存在與唯一性定理 (existence and uniqueness theorem)：凡在右邊

表現良好的點，都有且僅有一個解通過，而曲線族只能在右邊表現不良好的點相交，正如我們的所有曲線都在原點相交一樣。Boyce 與 DiPrima 所著的 *Elementary Differential Equations*（初等微分方程教科書）涵蓋了伯努利方程和這個定理。

詳細解答

STEP III 2004，第 9 題

力學 • 力矩與摩擦力

STEP III 2004，第 9 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q9

關鍵思路

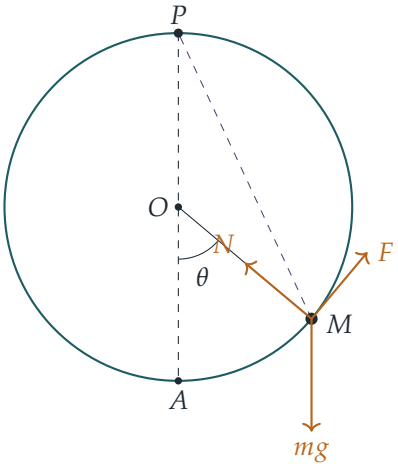
圓環的重量作用於轉軸正下方的 O，因此對 P 沒有力矩（moment）。除了轉軸處的反作用力之外，圓環所受的另一個力就只有老鼠施加的接觸力（contact force），所以圓環保持不動，當且僅當這個接觸力對 P 的力矩為零，也就是它的作用線（line of action）通過 P。整道題其實只在問：老鼠所受的接觸力何時可以沿 MP 的方向？這個問題要問兩次，一次求速率，一次求摩擦力的極限。

解答

模型

把老鼠視為質量為 m 的質點（particle），把圓環視為剛性（rigid）圓環（稍後可見它的質量並不影響結果），圓環可繞通過 P 的光滑水平軸自由轉動。設 $\theta = \angle AOM$，由向下的鉛垂線起量度，老鼠朝 θ 遞增的方向奔跑。圓環靜止時，老鼠在一個半徑為 a、圓心為 O 的固定圓上以恆定速率 u 運動。

老鼠所受的力有：重量 mg、圓環施加並指向 O 的法向反作用力（normal reaction）N，以及沿切線方向的摩擦力（friction）F，我們取 θ 遞增的方向（沿圓環向上）為正。



對老鼠沿切線方向（速率恆定，故沒有切向加速度）和沿半徑方向（加速度為 u^2/a，指向 O）分解：

$$F = mg\sin\theta,$$
$$N - mg\cos\theta = \frac{mu^2}{a}.$$

圓環何時處於平衡？

根據牛頓（Newton）第三定律，老鼠推圓環的力，與老鼠所受的接觸力（N 和 F）大小相等、方向相反。圓環的重量作用於 O，位於 P 的正下方，而轉軸的反作用力作用於 P，所以兩者對 P 都沒有力矩。因此，圓環保持平衡（equilibrium）且 POA 鉛垂，當且僅當接觸力對 P 的力矩為零，也就是它的作用線通過 P。

三角形 OPM 是等腰三角形（$OP = OM = a$），且 $\angle POM = \pi - \theta$，所以 $\angle OMP = \frac{1}{2}\theta$。於是直線 MP 與半徑 MO 的夾角為 $\frac{1}{2}\theta$，並偏向 θ 遞增的一方。因此，老鼠所受的接觸力沿 MP 的方向，當且僅當

$$\frac{F}{N} = \tan\tfrac{1}{2}\theta. \qquad (*)$$

（等價地，直接對 P 取力矩，老鼠施於圓環的力的力矩為 $a\big[N\sin\theta - F(1+\cos\theta)\big]$，而 $\sin\theta/(1+\cos\theta) = \tan\frac{1}{2}\theta$。）

想一想 怎樣把這個條件化為關於 u 的條件？

利用運動方程。老鼠所受的接觸力等於質量乘以加速度，再減去重量：

$$\mathbf{R} = m\cdot\frac{u^2}{a}\,\widehat{MO} + mg\,\hat{上}.$$

但 $\overrightarrow{MP} = \overrightarrow{MO} + \overrightarrow{OP} = a\,\widehat{MO} + a\,\hat{上}$，兩個係數相等。所以 $\mathbf{R}$ 平行於 MP 對每個 θ 都成

立，當且僅當 $u^2/a = g$，即

$$u = \sqrt{ag}.$$

用代數方法：當 $u^2 = ag$ 時，運動方程給出 $F = mg\sin\theta$ 和 $N = mg(1+\cos\theta)$，所以

$$\frac{F}{N} = \frac{\sin\theta}{1+\cos\theta} = \frac{2\sin\frac{1}{2}\theta\cos\frac{1}{2}\theta}{2\cos^2\frac{1}{2}\theta} = \tan\tfrac{1}{2}\theta,$$

這正是 $(*)$。對於一般的速率，對 P 的力矩算出來是 $m\sin\theta\,(u^2 - ag)$，它在每個位置都等於零，當且僅當 $u^2 = ag$。所以老鼠以速率 $\sqrt{ag}$ 奔跑，就能令圓環保持平衡，正如題目所要求，前提是圓環確實能夠提供所需的摩擦力。接下來我們就考察這一點。

摩擦力的極限

只有當 $|F| \le \mu N$ 時，老鼠的腳才能抓緊而不打滑。取 $u = \sqrt{ag}$，並考慮圓環的下半部分，該處 $|\theta| < \frac{1}{2}\pi$，故 $\cos\frac{1}{2}\theta > 0$，這個條件就是

$$\tan\tfrac{1}{2}|\theta| \le \mu \quad\Longleftrightarrow\quad |\theta| \le 2\arctan\mu.$$

所以在 $\angle AOM = 2\arctan\mu$ 之前，一切都行得通。（只有當 $\mu < 1$ 時，這一點才位於圓環的下半部分；若 $\mu \ge 1$，老鼠永遠到不了這一點。）

過了這一點，無論選甚麼速率都無補於事。平衡需要 $F/N = \tan\frac{1}{2}\theta > \mu$，而摩擦力無法做到。不論速率是多少，只要摩擦力滿足 $|F| \le \mu N$，老鼠施於圓環的力對 P 的力矩都是

$$a\big[N\sin\theta - F(1+\cos\theta)\big] \;\ge\; aN(1+\cos\theta)\big(\tan\tfrac{1}{2}\theta - \mu\big) > 0.$$

會發生甚麼事

老鼠一旦越過 $\angle AOM = 2\arctan\mu$，就再也抓不夠緊。牠的腳開始打滑，摩擦力降至極限值（limiting value）μN，老鼠無法維持原有速率：牠開始向後滑。這時牠施於圓環的接觸力從 P 的旁邊經過，因而力矩不為零，圓環不再處於平衡。圓環開始繞 P 轉動。上面那個力矩的方向表明，老鼠的法向推力勝過已減弱的摩擦力，所以圓環擺動時，A 會移向老鼠所在的一側，也就是老鼠原本奔跑的方向。從這一刻起，運動不再是簡單的平衡情況：圓環繞 P 擺動，而老鼠則在圓環上滑動。

驗算

- 當 $\theta = 0$ 時，老鼠位於 A，接觸力 $N = 2mg$ 是鉛垂的，並通過 P，而且不需要摩擦力，與預期相符。
- 速率 $\sqrt{ag}$ 具有速率的單位，既不取決於 μ，也不取決於圓環的質量，而且它令法向反作用力等於 $mg(1+\cos\theta)$，在整個下半部分都是正數。

- 當 $\mu = 0.5$ 時，臨界角為 $2\arctan 0.5 \approx 0.927$ 弧度 $\approx 53.1°$。在該處 $\sin\theta = 0.8$，$1 + \cos\theta = 1.6$，所以 $F/N = 0.5 = \mu$，恰好相等。以 $u^2 = ag$ 在隨機位置對 P 的力矩作數值檢驗，結果為零，誤差在 $10^{-15}mga$ 以內。

常見錯誤

- 對 O 而不是對 P 取力矩。圓環是繞 P 轉動的，而法向反作用力（它通過 O）對 P 確實有力矩。
- 漏掉 N 中的向心（centripetal）項，結果得出 $N = mg\cos\theta$ 和錯誤的摩擦力條件。
- 聲稱老鼠只是「掉下來」。在下半部分，$N > 0$ 始終成立。失效的是抓地力，而不是接觸。

STEP 之外

這裏的巧妙之處在於：一個物體繞樞軸（pivot）處於平衡，當且僅當其餘各力的合力（resultant）通過樞軸。「作用線通過 P」這個論證，省去了一整頁的分解計算。題目的後半部分則讓人初嘗耦合系統（coupled systems）的滋味：打滑的接觸把一個物體的運動傳到另一個物體上。要分析打滑之後的運動，就需要圓環的轉動慣量（moment of inertia）和完整的剛體轉動方程。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門教科書）正是以這種風格講解繞固定軸的力矩（torque）和受摩擦力限制的運動。

詳細解答

STEP III 2004，第 10 題

力學 • 彈性繩與簡諧運動

STEP III 2004，第 10 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

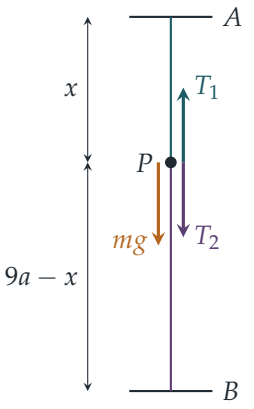
step.maths.org/questions/04-s3-q10

關鍵思路

每條弦線只有在拉緊時才遵從虎克定律（Hooke's law），所以要按哪些弦線處於拉緊狀態，把運動分成幾個階段。在每個階段中，加速度都是 $7a - x$ 的常數倍，因此質點以 $x = 7a$ 為中心作簡諧運動（simple harmonic motion），只是各階段的 ω 不同。我們先求出第一階段結束時的時間和速率，再以此作為第二階段的起始條件。

解答

我們把 P 視為質點（particle），並把弦線視為輕弦：伸長時遵從虎克定律，鬆弛時不施力。取 $x = AP$，由 A 向下量度，於是 $PB = 9a - x$，並以向下為正方向。



運動方程

當 $x > 6a$ 時，弦線 AP 的伸長量為 $x - 6a$，所以它把 P 向上拉的張力（tension）為

$$T_1 = \frac{6mg(x-6a)}{6a} = \frac{mg(x-6a)}{a}.$$

當 $x < 7a$ 時，弦線 BP 的伸長量為 $(9a - x) - 2a = 7a - x$，所以它把 P 向下拉向 B 的張力為

$$T_2 = \frac{6mg(7a-x)}{2a} = \frac{3mg(7a-x)}{a},$$

而當 $x \geq 7a$ 時，BP 鬆弛，$T_2 = 0$。

當 $6a < x < 7a$ 時，兩條弦線都拉緊，沿向下方向應用牛頓第二定律（Newton's second law），得

$$\begin{aligned} m\ddot{x} &= mg - \frac{mg(x-6a)}{a} + \frac{3mg(7a-x)}{a} \\ &= \frac{mg}{a}\big(a - x + 6a + 21a - 3x\big) = \frac{mg}{a}(28a - 4x), \end{aligned}$$

所以 $\ddot{x} = \dfrac{4g}{a}(7a - x)$，正是所需結果。

當 $7a < x < 9a$ 時，只有 AP 拉緊，所以

$$m\ddot{x} = mg - \frac{mg(x-6a)}{a} = \frac{mg}{a}(7a - x),$$

由此得 $\ddot{x} = \dfrac{g}{a}(7a - x)$，正是所需結果。

到達 B 所需的時間

第一階段：$6a \leq x \leq 7a$。設 $y = x - 7a$，則 $\ddot{y} = -\omega_1^2 y$，其中 $\omega_1^2 = 4g/a$，這是以 $x = 7a$ 為中心的簡諧運動。質點由 $y = -a$ 從靜止開始運動，故

$$x = 7a - a\cos\omega_1 t, \qquad \dot{x} = a\omega_1 \sin\omega_1 t.$$

質點在 $x = 6a$ 從靜止開始，向下加速（因為該處 $\ddot{x} > 0$），所以 x 由 $6a$ 起遞增，並一直處於這個階段，直至 $x = 7a$。這在 $\cos\omega_1 t = 0$ 時發生，首次出現於 $\omega_1 t = \pi/2$，即經過時間

$$t_1 = \frac{\pi}{2\omega_1} = \frac{\pi}{4}\sqrt{\frac{a}{g}},$$

此時速率為 $a\omega_1 = 2\sqrt{ag}$，方向向下。

> ? **想一想** 方程在 $x = 7a$ 處改變了。我們該怎樣繼續？

位置和速度在 $x = 7a$ 處都是連續的，所以第一階段的終結狀態就是第二階段的初始條件（starting conditions）。我們在該處重新計時。

第二階段：$7a \le x \le 9a$。這時 $\ddot{y} = -\omega_2^2 y$，其中 $\omega_2^2 = g/a$，質點由中心 $y = 0$ 以速率 $2\sqrt{ag}$ 出發。因此，設 τ 為自 $x = 7a$ 起計的時間，則

$$x = 7a + \frac{2\sqrt{ag}}{\omega_2} \sin \omega_2 \tau = 7a + 2a \sin \omega_2 \tau.$$

振幅（amplitude）剛好是 $2a$，所以當 $\sin \omega_2 \tau = 1$ 時，質點到達 $x = 9a$，亦即 B 點，首次出現於 $\omega_2 \tau = \pi/2$。在該刻 $\dot{x} = 2\sqrt{ag} \cos \omega_2 \tau = 0$，所以質點到達 B 時速率為零。這個階段需時

$$t_2 = \frac{\pi}{2\omega_2} = \frac{\pi}{2}\sqrt{\frac{a}{g}}.$$

因此，質點到達 B 所需的時間為

$$t_1 + t_2 = \frac{\pi}{4}\sqrt{\frac{a}{g}} + \frac{\pi}{2}\sqrt{\frac{a}{g}} = \frac{3\pi}{4}\sqrt{\frac{a}{g}}.$$

驗算

能量的計算可以確認 P 剛好到達 B 並在該處靜止。比較起點（BP 伸長了 a）和到達 B 的一刻（AP 伸長了 $3a$）。彈性勢能（elastic energy）由 $\frac{6mg\,a^2}{2\cdot 2a} = \frac{3}{2}mga$ 變為 $\frac{6mg\,(3a)^2}{2\cdot 6a} = \frac{9}{2}mga$。增加了 $3mga$，剛好等於下降 $3a$ 所失去的重力勢能（gravitational energy），所以在 B 處的動能（kinetic energy）為零。數值上，取 $a = 0.7$ 和 $g = 9.8$，由 $x = 6a$ 從靜止開始，對完整的運動方程逐步作數值計算，得出在 $t \approx 0.2099$ 時 $x = 7a$，在 $t \approx 0.6297$ 時到達最低點 $x = 9a$，而 $\frac{3\pi}{4}\sqrt{a/g} = 0.6297$。

常見錯誤

- 在 $x > 7a$ 時仍使用 T_2：弦線不能推物，所以一旦 BP 短於 $2a$，它便不施力。
- 整個運動都使用 ω_1，或者忘記以第一段簡諧運動所留下的速率 $2\sqrt{ag}$ 重新開始第二段簡諧運動。
- 沒有檢查振幅 $2a$ 是否真的能把 P 帶到 B，便把第二階段說成是四分之一個週期（period）。

STEP 之外

這是一個分段線性振子（piecewise-linear oscillator）：回復力（restoring force）是線性的，但勁度（stiffness）在 $x = 7a$ 處改變。兩個部分各有自己的相平面（phase plane）橢圓，而軌跡則由兩個不同橢圓的弧組成，在弦線變鬆之處相接。由彈跳的球到碰撞振子（impact oscillator），這類系統是非線性動力學（nonlinear dynamics）的標準入門例子。

Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學導論）深入討論了振動和能量方法。

詳細解答

STEP III 2004，第 11 題

力學 • 二體排斥

STEP III 2004，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q11

本題在 *Beyond STEP* 第 10.6 節二體化爲一體中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch10/two-body.html

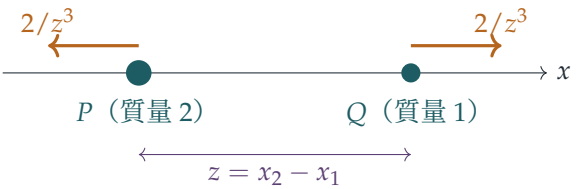
關鍵思路

兩個質點互相推開，會得出兩條運動方程（equation of motion），而只要把它們巧妙地組合兩次，問題便迎刃而解。*差* $x_2 - x_1$ 給出一條只涉及間距 z 的方程，利用 $v\,\mathrm{d}v/\mathrm{d}z$ 這個技巧便可積分一次（這其實就是能量守恆）。*加權和* $2x_1 + x_2$ 是質心（centre of mass）位置的三倍，而由於這些力都是內力，它完全不受任何淨力作用。

解答

z 的方程

把 P 和 Q 視為直線上的質點（particle），x 沿由 P 指向 Q 的方向遞增（因為 $x_2 > x_1$）。斥力（repulsion）把 P 推向負方向，把 Q 推向正方向，兩者的大小均為 $2/z^3$。



對每個質點應用牛頓第二定律（Newton's second law），得

$$2\ddot{x}_1 = -\frac{2}{z^3}, \qquad \ddot{x}_2 = \frac{2}{z^3}.$$

所以 $\ddot{x}_1 = -1/z^3$，兩式相減，得

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = \ddot{x}_2 - \ddot{x}_1 = \frac{2}{z^3} + \frac{1}{z^3} = \frac{3}{z^3},$$

正是所求。

求解 z

先看初始條件（initial conditions）。在 $t = 0$ 時，$z = 1$。P 朝着 Q 運動，即沿正方向運動，所以 $\dot{x}_1 = 1$，而 $\dot{x}_2 = 0$。因此在 $t = 0$ 時，$v = \dot{z} = 0 - 1 = -1$：兩者的間距正在縮小。

利用 $\ddot{z} = v\,\mathrm{d}v/\mathrm{d}z$，方程變成

$$v\frac{\mathrm{d}v}{\mathrm{d}z} = \frac{3}{z^3} \quad \Longrightarrow \quad \tfrac{1}{2}v^2 = -\frac{3}{2z^2} + C.$$

當 $z = 1$ 時，$v^2 = 1$，所以 $\frac{1}{2} = -\frac{3}{2} + C$，即 $C = 2$。因此

$$v^2 = 4 - \frac{3}{z^2}, \qquad \text{或} \qquad z^2\left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2 = 4z^2 - 3.$$

左邊是 $\left(\frac{1}{2}\,\mathrm{d}(z^2)/\mathrm{d}t\right)^2$，這提示我們設 $u = \sqrt{4z^2 - 3} \geq 0$。於是 $u^2 = 4z^2 - 3$，所以 $u\,\dot{u} = 4z\dot{z}$，而上面的方程表示 $(z\dot{z})^2 = u^2$，即 $z\dot{z} = \pm u$。因此

$$\dot{u} = \frac{4z\dot{z}}{u} = \pm 4.$$

起初 $u = 1$ 而 $z\dot{z} = -1 < 0$，所以取負號：$u = 1 - 4t$。

想一想 當 u 降至 0 時會發生甚麼事？$\dot{z}$ 的符號會一直保持為負嗎？

不會。在 $t = \frac{1}{4}$ 時 $u = 0$，所以 $z = \sqrt{3}/2$ 而 $v = 0$：兩個質點在這一刻最為接近。此時 $\ddot{z} = 3/z^3 > 0$，所以 z 開始增加，$\dot{z}$ 變為正數，改為取正號，於是當 $t \geq \frac{1}{4}$ 時 $u = 4t - 1$（在 $t = \frac{1}{4}$ 處連續）。在兩種情況下均有 $u^2 = (4t - 1)^2$，所以對所有 $t \geq 0$，

$$4z^2 - 3 = 16t^2 - 8t + 1 \quad \Longrightarrow \quad z^2 = 4t^2 - 2t + 1,$$

又因為 $z > 0$，

$$z = \sqrt{4t^2 - 2t + 1}.$$

這就是所求的結果。（注意 $4t^2 - 2t + 1 = 4(t - \frac{1}{4})^2 + \frac{3}{4} > 0$，所以平方根恆有定義。）

求 x_1 和 x_2

由兩條運動方程，

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(2x_1 + x_2) = 2\ddot{x}_1 + \ddot{x}_2 = -\frac{2}{z^3} + \frac{2}{z^3} = 0.$$

這就是動量守恆(conservation of momentum)：兩個力大小相等、方向相反。所以 $2x_1 + x_2$ 是 t 的線性函數。在 $t = 0$ 時它等於 $2 \cdot 0 + 1 = 1$，而它的變化率是 $2\dot{x}_1 + \dot{x}_2 = 2 \cdot 1 + 0 = 2$。因此

$$2x_1 + x_2 = 2t + 1, \qquad x_2 - x_1 = \sqrt{4t^2 - 2t + 1}.$$

第一式減去第二式得 $3x_1$，第一式加上第二式的兩倍得 $3x_2$：

$$x_1 = \frac{2t + 1 - \sqrt{4t^2 - 2t + 1}}{3}, \qquad x_2 = \frac{2t + 1 + 2\sqrt{4t^2 - 2t + 1}}{3}.$$

驗算

- 在 $t = 0$ 時，以上兩式給出 $x_1 = 0$ 和 $x_2 = 1$。求導後，$\dot{x}_1(0) = \frac{1}{3}\big(2 - \frac{1}{2} \cdot (-2)\big) = 1$ 及 $\dot{x}_2(0) = \frac{1}{3}(2 - 2) = 0$，符合要求。
- 求 z 的一條捷徑：設 $w = z^2$，則 $\ddot{w} = 2\dot{z}^2 + 2z\ddot{z} = 2\big(4 - 3/z^2\big) + 6/z^2 = 8$，所以由 $w(0) = 1$ 和 $\dot{w}(0) = 2z\dot{z} = -2$ 可得 $w = 4t^2 - 2t + 1$。這樣便完全不必為符號操心。
- 當 $t \to \infty$ 時，$z \approx 2t - \frac{1}{2}$，所以 $x_1 \to \frac{1}{2}$ 而 $\dot{x}_2 \to 2$：P 最終靜止，而 Q 帶走全部動量（$1 \times 2 = 2 \times 1$）和全部能量。勢能（potential energy）是 $1/z^2$，所以開始時的能量是 $1 + 1 = 2$，最終的能量是 $\frac{1}{2} \cdot 1 \cdot 2^2 = 2$。
- 以數值方法逐步計算原來的方程（龍格—庫塔法（Runge–Kutta），步長 10^{-4}）直至 $t = 5$，所得結果與 x_1、x_2 和 z 的公式相差不超過 10^{-11}，最接近時的間距為 $z = 0.8660\ldots = \sqrt{3}/2$。

常見錯誤

- 取 $\dot{z}(0) = +1$。P 正朝着 Q 運動，所以間距正在縮小。
- 寫出 $\mathrm{d}z/\mathrm{d}t = -\sqrt{4 - 3/z^2}$ 並以為它在所有時刻都成立，忽略了 $t = \frac{1}{4}$ 處的轉向點（turning point）。
- 在 P 的方程中漏掉質量 2，結果得出 $\ddot{z} = 4/z^3$。

STEP 之外

這是二體問題（two-body problem）的縮影：質心作勻速運動，而相對坐標的表現就像一個約化質量（reduced mass）為 $m_1m_2/(m_1+m_2)=\frac{2}{3}$ 的單一質點受着同一個力，因為 $\frac{2}{3}\ddot{z}=2/z^3$。同樣的分拆方法可把行星運動化為單一軌道的問題。平方反比以外，立方反比力（inverse-cube force）也很特別：z^2 原來恰好是 t 的二次函數，這是 $1/z^2$ 勢所隱藏的一種對稱性，屬於共形力學（conformal mechanics）的研究課題。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門）對約化質量和二體運動有細緻的論述。

詳細解答

STEP III 2004，第 12 題

概率與統計 • 指示變量與錯排

STEP III 2004，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q12

關鍵思路

把一個計數寫成指示變量（indicator variable）之和，是概率論中最有用的技巧之一。指示變量的期望值（expectation）其實就是概率，所以 $\mathrm{E}(N)$ 唾手可得；方差（variance）則需要用到每一對變量之間的協方差（covariance），這些協方差很小但不是零，因為球衣並不是獨立地分配的。一旦知道 N 的均值和方差都等於 1，參數為 1 的泊松分佈（Poisson distribution）自然就是值得一試的模型。

解答

單一指示變量 C_1

球衣的 $m!$ 種分配方式每一種都等可能，所以球衣 1 分給 m 名球員中任何一人的可能性都相同。因此 $\mathrm{P}(C_1 = 1) = \frac{1}{m}$，而由於 C_1 只取 0 和 1 兩個值，

$$\mathrm{E}(C_1) = 1 \cdot \mathrm{P}(C_1 = 1) = \frac{1}{m},$$

正如所求。因為 $C_1^2 = C_1$（因 $0^2 = 0$ 及 $1^2 = 1$），我們亦有 $\mathrm{E}(C_1^2) = \frac{1}{m}$，所以

$$\mathrm{Var}(C_1) = \frac{1}{m} - \frac{1}{m^2} = \frac{m-1}{m^2}.$$

C_1 與 C_2 的協方差

這裏我們需要 $m \geq 2$。乘積 C_1C_2 恰好在球員 1 和 2 都拿到自己的球衣時等於 1，否則等於 0。已知球員 1 拿到球衣 1，餘下的 $m-1$ 件球衣便隨機分給其餘 $m-1$ 名球員，所以

球員 2 拿到球衣 2 的概率是 $\frac{1}{m-1}$。因此

$$\mathrm{E}(C_1C_2) = \mathrm{P}(C_1 = 1,\, C_2 = 1) = \frac{1}{m} \cdot \frac{1}{m-1},$$

而

$$\begin{aligned}\mathrm{Cov}(C_1, C_2) &= \mathrm{E}(C_1C_2) - \mathrm{E}(C_1)\,\mathrm{E}(C_2) = \frac{1}{m(m-1)} - \frac{1}{m^2} \\ &= \frac{m-(m-1)}{m^2(m-1)} = \frac{1}{m^2(m-1)}.\end{aligned}$$

這個值是正數：知道球員 1 拿對了球衣，就等於從球員 2 可能拿到的球衣中剔除了一件錯的，令其拿對的機會略為提高。

? **想一想** 為甚麼協方差不是零？各球員的球衣不是分開選的嗎？

並不是。球衣是每人分配一件的，所以球員 1 拿到甚麼，會影響其他人剩下甚麼可拿。各個 C_i 是相依的，這正是計算 $\mathrm{Var}(N)$ 時必須計入協方差項的原因。

N 的均值與方差

根據對稱性，每個 C_i 的分佈都與 C_1 相同，而每一對 C_i, C_j（其中 $i \neq j$）的聯合分佈（joint distribution）都與 C_1, C_2 相同（把球員和球衣一併重新編號，並不會改變這個隨機分配）。利用期望值的線性性質（linearity of expectation），而這性質並不需要獨立性，

$$\mathrm{E}(N) = \sum_{i=1}^{m} \mathrm{E}(C_i) = m \cdot \frac{1}{m} = 1.$$

至於和的方差，我們把所有方差加起來，再加上所有有序對 $i \neq j$ 的協方差，這樣的有序對共有 $m(m-1)$ 個：

$$\begin{aligned}\mathrm{Var}(N) &= \sum_{i=1}^{m} \mathrm{Var}(C_i) + \sum_{i \neq j} \mathrm{Cov}(C_i, C_j) \\ &= m \cdot \frac{m-1}{m^2} + m(m-1) \cdot \frac{1}{m^2(m-1)} \\ &= \frac{m-1}{m} + \frac{1}{m} = 1.\end{aligned}$$

所以 $\mathrm{E}(N) = \mathrm{Var}(N) = 1$，正如所求。（若 $m = 1$，則必有 $N = 1$，於是 $\mathrm{Var}(N) = 0$；題目其實隱含假設了 $m \geq 2$，一如計算協方差時已經需要的那樣。）

用哪一種近似？

無論 m 取甚麼值，正態近似（Normal approximation）的效果都很差。N 只取整數值 $0, 1, 2, \ldots$，而由於均值為 1、標準差為 1，它的概率幾乎全部集中在 $0, 1, 2, 3$ 這三四個

值上。這個分佈在 0 處被截斷，那裏只比均值低一個標準差，所以它明顯向右偏斜；而 $\mathrm{N}(1,1)$ 分佈是對稱的，並且在 $-\frac{1}{2}$ 以下有不可忽視的概率（即使作了連續性校正（continuity correction），也約有 7%），但 N 根本不可能取到那裏的值。最關鍵的是，增大 m 也無濟於事：對每一個 m，均值和方差都維持在 1，所以分佈的形狀永遠不會展開成鐘形。

相反，泊松近似則合情合理。N 所數的，是在為數甚多的 m 個事件之中「成功」的個數，每個事件的概率 $\frac{1}{m}$ 都很小，而這些事件之間只有很弱的相依性（協方差 $\frac{1}{m^2(m-1)}$ 極小）。計數近似服從泊松分佈，正是在這種情況下發生的。此外，泊松變量的均值等於方差，而 N 恰好具有這個性質，所以 $\mathrm{Po}(1)$ 是自然的選擇。

$m = 4$ 的情況

共有 $4! = 24$ 種等可能的分配方式。把一種分配寫成球員 $1,2,3,4$ 所穿球衣的號碼序列，則沒有人穿上自己球衣的分配是

$$2143,\quad 2341,\quad 2413,\quad 3142,\quad 3412,$$
$$3421,\quad 4123,\quad 4312,\quad 4321.$$

（有系統地列舉的方法：球員 1 穿的是球衣 2、3 或 4，而每種情況下都恰好有三種方法完成餘下的分配，所以共有 $3 \times 3 = 9$ 種。）因此

$$\mathrm{P}(N = 0) = \frac{9}{24} = \frac{3}{8} = 0.375.$$

參數為 1 的泊松近似給出 $\mathrm{P}(N = 0) \approx \mathrm{e}^{-1}$。取 $\mathrm{e} \approx 2.72$，

$$\mathrm{e}^{-1} \approx \frac{100}{272} = \frac{25}{68}.$$

以真確值為基準的相對誤差（relative error）是

$$\frac{\frac{3}{8} - \frac{25}{68}}{\frac{3}{8}} = 1 - \frac{25}{68} \cdot \frac{8}{3} = 1 - \frac{200}{204} = \frac{1}{51} \approx 0.0196,$$

即大約 2%，與題目所述相符。（如果改為以近似值為基準來量度誤差，則得 $\frac{1}{50}$，恰好是 2%；無論用哪一種量法，對這麼小的 m 來說，吻合程度都好得驚人。）

驗算

- 當 $m = 2$ 時，只有兩種可能：兩人都拿對（$N = 2$）或兩人都拿錯（$N = 0$），概率各為 $\frac{1}{2}$：均值為 1，方差為 1。吻合。
- 用電腦枚舉 $m = 2, \ldots, 7$ 的所有排列（permutation），確認了 $\mathrm{E}(C_1) = \frac{1}{m}$、$\mathrm{Var}(C_1) = \frac{m-1}{m^2}$、$\mathrm{Cov}(C_1, C_2) = \frac{1}{m^2(m-1)}$ 及 $\mathrm{E}(N) = \mathrm{Var}(N) = 1$ 全部準確成立，並找到 4 件物件的錯排（derangement）共 9 個，所以 $\mathrm{P}(N = 0) = 0.375$，對比 $1/2.72 \approx 0.3676$。

- 留意 $N = m - 1$ 是不可能的：如果除一人外其他球員都拿對了球衣，最後那一人也必定拿對。這再次顯示 N 與正態分佈相去甚遠。

常見錯誤

- 把各個 C_i 當作獨立，結果得出 $\mathrm{Var}(N) = \frac{m-1}{m}$ 而不是 1。
- 只數了 $\binom{m}{2}$ 個協方差項而不是 $m(m-1)$ 個，即忘記了 $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$ 中的因子 2。
- 列舉錯排時粗心大意：有系統地數（3×3）可避免遺漏。

STEP 之外

這是經典的「配對」問題，又稱錯排問題。利用容斥原理（inclusion–exclusion）可以證明 $\mathrm{P}(N = 0) = \sum_{k=0}^{m} \frac{(-1)^k}{k!}$，這個值極快地趨向 e^{-1}，而事實上當 $m \to \infty$ 時，N 依分佈收斂（converges in distribution）於 $\mathrm{Po}(1)$。更廣泛的原理是：由許多罕見而弱相依的事件所構成的計數，近似服從泊松分佈；Chen–Stein 方法把這個原理嚴格地表述出來。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門教材）一書，反覆以指示變量和配對問題作為例子。

詳細解答

STEP III 2004，第 13 題

概率與統計 • 幾何分佈

STEP III 2004，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q13

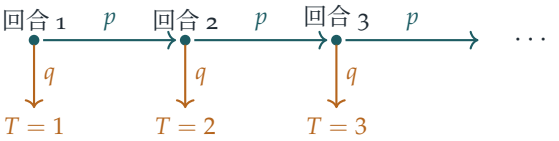
關鍵思路

暫且放下整場比賽，只跟着一位參賽者看。他出局的回合是一個幾何隨機變量（geometric random variable），而三位參賽者的出局回合互相獨立（independent）。於是每一部分都是關於三個獨立幾何變量的問題。原來大獎是否頒發，只取決於最遲的出局回合是否有人並列，所以最後一部分就是前兩部分合起來的補事件（complement）。

解答

設 $q = 1 - p$，並假設 $0 < p < 1$。對參賽者 j（$j = 1, 2, 3$），設 T_j 為他失敗的回合。他在第 k 回合失敗，當且僅當他通過了第 $1, \ldots, k-1$ 回合，然後在第 k 回合失敗，所以

$$\mathrm{P}(T_j = k) = p^{k-1}q, \qquad k = 1, 2, 3, \ldots$$



由於每一回合的通過與否互相獨立，而各參賽者的表現亦互相獨立，所以 T_1、T_2、T_3 互相獨立。另外留意

$$\mathrm{P}(T_j \leq r - 1) = 1 - \mathrm{P}(\text{通過回合} 1, \ldots, r-1) = 1 - p^{r-1}.$$

第 (i) 部分

根據獨立性，三人同在第 k 回合出局的概率是 $(p^{k-1}q)^3$。對 k 求這個等比級數（geometric series）之和（公比（common ratio）$p^3 < 1$），得

$$\mathrm{P}(T_1 = T_2 = T_3) = \sum_{k=1}^{\infty} q^3 p^{3(k-1)} = \frac{q^3}{1-p^3} = \frac{(1-p)^2}{1+p+p^2}.$$

第 (ii) 部分

較早出局的參賽者有 3 種選法。選定之後，其餘兩人同在第 r 回合失敗的概率是 $(p^{r-1}q)^2$，而較早出局的那一位在第 r 回合之前某一回合失敗的概率是 $1-p^{r-1}$。不同選法所對應的事件互斥（disjoint），所以

$$\mathrm{P}(\text{兩人在回合}r\text{, 第三人較早}) = 3p^{2r-2}q^2\left(1-p^{r-1}\right).$$

第 (iii) 部分

想一想 用 T_1, T_2, T_3 來表達，大獎究竟在甚麼情況下才會頒發？

假設 T_j 之中的最大值只由一位參賽者取得，例如 $T_3 > T_1, T_2$，並設 $m = \max(T_1, T_2)$。在第 m 回合，參賽者 3 仍然在場，並且通過該回合。當時仍在場的其他參賽者都有 $T_j = m$，所以都在該回合失敗。因此參賽者 3 在第 m 回合贏得大獎。

反過來，假設最大值有人並列，例如 $T_2 = T_3 = M \geq T_1$。在第 M 回合之前，參賽者 2 和 3 都仍然在場，而且每一回合都通過，所以沒有任何一個回合只剩一人過關。到了第 M 回合，餘下的人全部失敗，所以根本無人過關。因此大獎不會頒發。

由此可見，大獎不頒發，當且僅當最遲的兩個出局回合相等：要麼三者全部相等（第 (i) 部分），要麼其中兩者在某個第 r 回合相等，而第三者較早。第二種情況需要 $r \geq 2$，而它正是第 (ii) 部分的事件。把第 (ii) 部分的結果對 $r \geq 2$ 求和，並設 $s = r-1$，

$$\begin{aligned}\sum_{r=2}^{\infty} 3q^2p^{2r-2}\left(1-p^{r-1}\right) &= 3q^2\sum_{s=1}^{\infty}\left(p^{2s}-p^{3s}\right) = 3q^2\left(\frac{p^2}{1-p^2}-\frac{p^3}{1-p^3}\right)\\ &= 3q^2\,\frac{p^2(1-p^3)-p^3(1-p^2)}{(1-p^2)(1-p^3)} = 3q^2\,\frac{p^2(1-p)}{(1-p^2)(1-p^3)}\\ &= \frac{3p^2q^2}{(1+p)(1-p^3)}.\end{aligned}$$

加上第 (i) 部分的結果，並利用 $1-p^3 = q(1+p+p^2)$，

$$\mathrm{P}(\text{沒有大獎}) = \frac{q^3(1+p)+3p^2q^2}{(1+p)(1-p^3)} = \frac{q^2(1+2p^2)}{(1+p)(1-p^3)} = \frac{(1-p)(1+2p^2)}{(1+p)(1+p+p^2)}.$$

由於 $(1+p)(1+p+p^2) = 1+2p+2p^2+p^3$ 及 $(1-p)(1+2p^2) = 1-p+2p^2-2p^3$，兩者之差是 $3p+3p^3$。因此

$$\mathrm{P}(\text{頒發大獎}) = \frac{3p(1+p^2)}{(1+p)(1+p+p^2)}.$$

驗算

- 當 $p \to 0$ 時，人人都在第 1 回合失敗，而 (iii) 的答案趨向 0。當 $p \to 1$ 時，並列的機會變得很小，而答案趨向 $\frac{3\cdot 2}{2\cdot 3} = 1$。
- 直接計算：當其中一個 T_j 大於另外兩個時，大獎便會頒發，所以其概率是 $3\sum_k p^{k-1}q\,(1-p^{k-1})^2$。取 $p = \frac{1}{2}$ 作精確求和，得 $\frac{5}{7} \approx 0.714$，與 $\frac{3\cdot\frac{1}{2}\cdot\frac{5}{4}}{\frac{3}{2}\cdot\frac{7}{4}} = \frac{5}{7}$ 吻合。
- 直接模擬（simulation）比賽本身（取 $p = 0.6$，進行 2×10^5 次），結果有 77.9% 的次數頒發了大獎，而公式給出的是 0.7806。

常見錯誤

- 在第 (ii) 部分忘記了因數 3，即選擇哪一位參賽者較早出局的選法數目。
- 以為最後兩人必須在不同回合出局，大獎才會頒發。如果有兩人同在第 m 回合出局，而第三人通過了該回合，他仍然獲獎。只有在最高處出現並列，大獎才不會頒發。

STEP 之外

這道題目所問的，是獨立幾何隨機變量的最大值，以及這個最大值是否唯一。對 n 位參賽者而言，當 $n \to \infty$ 時，最高處出現並列的概率並不會穩定下來趨向一個極限。它會隨 $\log n$ 作極輕微的振盪，這是領袖選舉（leader election）等演算法分析中一個著名的意外現象。這裏用到的工具，即對一個變量取條件（conditioning）以及等比級數求和，在 Grimmett 與 Welsh 的 *Probability: An Introduction* 中有詳細闡述。

詳細解答

STEP III 2004，第 14 題

概率與統計 • 假設檢驗的錯誤

STEP III 2004，第 14 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/04-s3-q14

關鍵思路

在兩個假設之下，$\overline{X}$ 的中心同樣是 μ；兩者的分別只在於 $\overline{X}$ 有多分散。因此，否定 H_0 的證據就是 $\overline{X}$ 落在遠離 μ 的位置，不論在哪一邊，而對稱的雙側拒絕域（rejection region）正好能偵測這種情況。之後的一切，都歸結為把 $\overline{X}$ 標準化（standardise），在每個假設之下各做一次。

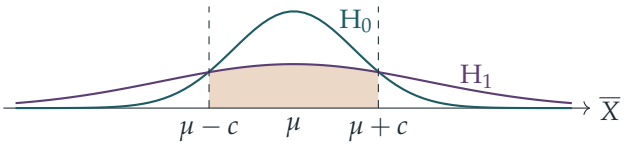
解答

為何採用雙側檢驗

由於 $X \sim N(\mu, \sigma^2)$，而且各樣本值互相獨立，樣本平均值（sample mean）滿足

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

無論在哪一個假設之下，$\overline{X}$ 都以已知的值 μ 為中心；唯一的分別是在 H_1 之下，它的標準差（standard deviation）$\sigma_1/\sqrt{n}$ 比 $\sigma_0/\sqrt{n}$ 大。$\overline{X}$ 接近 μ 的值在 H_0 之下相對較有可能出現，而遠離 μ 的值則在 H_1 之下相對較有可能出現（較闊的曲線尾部較厚；見附圖）。由於兩個分佈都關於 μ 對稱，一個很大的 $\overline{X}$ 值與一個同樣程度地小的值相比，並不是支持 H_1 的更強證據。所以，自然的檢驗方法是當 $|\overline{X} - \mu|$ 很大時拒絕 H_0，不論方向；而根據對稱性，接受區間應以 μ 為中心：當 $\mu - c < \overline{X} < \mu + c$ 時接受 H_0。在這裏採用單側檢驗（one-sided test）是錯的，因為這樣會忽略一半支持分散程度較大的證據。



陰影部分的面積（H_1 為真卻接受 H_0）是 β；兩條虛線以外的藍綠色曲線下面積是 α。

求 c

在 H_0 之下，$Z = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_0}$ 服從標準正態分佈（standard normal distribution）。當 $|\overline{X} - \mu| \geq c$ 時，即當 $|Z| \geq c\sqrt{n}/\sigma_0$ 時，我們接受 H_1。根據標準正態分佈的對稱性，

$$\alpha = \mathrm{P}\left(|Z| \geq \frac{c\sqrt{n}}{\sigma_0}\right) = 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right)\right), \qquad \text{所以} \qquad \Phi\left(\frac{c\sqrt{n}}{\sigma_0}\right) = 1 - \tfrac{1}{2}\alpha.$$

由於 Φ 嚴格遞增，這表示 $c\sqrt{n}/\sigma_0 = z_\alpha$，因此

$$c = \frac{z_\alpha\,\sigma_0}{\sqrt{n}}.$$

β 與 n 無關

在 H_1 之下，$W = \dfrac{\sqrt{n}\,(\overline{X} - \mu)}{\sigma_1}$ 服從標準正態分佈，而

$$\begin{aligned}\beta = \mathrm{P}(|\overline{X} - \mu| < c) &= \mathrm{P}\left(|W| < \frac{c\sqrt{n}}{\sigma_1}\right) = \mathrm{P}\left(|W| < \frac{z_\alpha\sigma_0}{\sigma_1}\right) \\ &= 2\,\Phi\left(\frac{z_\alpha\sigma_0}{\sigma_1}\right) - 1,\end{aligned}$$

其中我們代入了上一部分得到的 $c\sqrt{n} = z_\alpha\sigma_0$。$c$ 之中的 $\sqrt{n}$ 與標準化時出現的 $\sqrt{n}$ 互相抵消，所以 β 只取決於 α 和 σ_0/σ_1，與 n 無關。

想一想 樣本愈大，檢驗理應愈好吧？

通常確是如此，因為隨着 n 增大，兩個抽樣分佈（sampling distribution）會互相分開。但在這裏它們永遠不會分開：兩者都以 μ 為中心，而增大 n 會令兩個標準差按同一個因子 $\sqrt{n}$ 縮小。整幅圖只是按比例縮放，所以面積 α 和 β 都不變。用樣本平均值來檢驗方差（variance），是一個差劣的統計量；它把各個數值所包含的關於分散程度的資訊都丟掉了。

比值 σ_1/σ_0 的最小值

設 $r = \sigma_1/\sigma_0 > 1$。

條件 $\alpha < 0.05$。這表示 $\Phi(z_\alpha) = 1 - \frac{1}{2}\alpha > 0.975 \approx \Phi(1.960)$，所以，由於 Φ 遞增，$z_\alpha > 1.960$。

條件 $\beta < 0.05$。這表示 $2\Phi(z_\alpha/r) - 1 < 0.05$，即 $\Phi(z_\alpha/r) < 0.525 \approx \Phi(0.063)$，所以 $z_\alpha/r < 0.063$，亦即

$$r > \frac{z_\alpha}{0.063}.$$

綜合兩者，我們需要 $r > z_\alpha/0.063 > 1.960/0.063$。反過來說，若 $r > 1.960/0.063$，我們可以選取剛好大於 1.960 且滿足 $z_\alpha < 0.063r$ 的 z_α，這樣兩個條件都成立。因此這個比值必須大於

$$\frac{1.960}{0.063} \approx 31.1,$$

所以 σ_1/σ_0 的最小值約為 31。兩個標準差必須相差約三十倍，樣本平均值才能如此可靠地把它們區分開來，而且不論樣本大小如何都是這樣。

驗算

- 以 $\mu = 3$、$\sigma_0 = 1$、$\sigma_1 = 2.5$、$\alpha = 0.1$ 及 $n = 1, 4, 9, 25$ 進行模擬，在 H_0 之下的拒絕率每次都約為 0.10，而在 H_1 之下的接受率對每個 n 都約為 0.489，與 $2\Phi(1.645/2.5) - 1 \approx 0.489$ 吻合。
- 當 $r = 1.960/0.063$ 而 $\alpha = 0.05$ 恰好成立時，$\beta = 2\Phi(0.063) - 1 \approx 0.050$，一如所料正好在邊界上。若使用 Φ 的精確反函數值，會得到 $r \approx 31.3$，所以答案 31 不受題目所給表值的捨入誤差影響。
- 合理性檢查：若 $\sigma_1 = \sigma_0$，則 $\beta = 2\Phi(z_\alpha) - 1 = 1 - \alpha$，當兩個假設重合時必然如此。

常見錯誤

- 使用單側的拒絕域，例如 $\overline{X} > \mu + c$：在 H_1 之下平均值並沒有移動，所以沒有理由偏重其中一邊的尾部。
- 寫成 $\alpha = 1 - \Phi(c\sqrt{n}/\sigma_0)$，忘記了拒絕域有兩個尾部。
- 把關於 r 的不等式方向弄反：β 小需要 z_α/r 小，所以 r 要大。

STEP 之外

這道題展示了兩類錯誤之間的取捨，以及檢驗統計量（test statistic）的選擇為何重要。奈曼–皮爾遜引理（Neyman–Pearson lemma）指出，最強大的檢驗是比較兩個假設之下的似然（likelihood）；在這裏，這會引出統計量 $\sum(X_i - \mu)^2$，它服從的 χ^2_n 分佈確實會隨

着 n 增大而把兩個假設區分開來，這與 $\bar{X}$ 不同。想找一本易讀的入門書了解這些概念，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門），或 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率導論）。

Beyond STEP

STEP 是大多數中學生一生中應考過最艱深的數學考試，也是最出色的考試之一：題目簡短、陌生而深刻，當中最精彩的，其實是喬裝了的大學數學小品。

本書精選最有趣的 STEP 試題，由 1987 年首次開考之前的樣本試卷（specimen papers），到 2025 年的試卷，並以每一道題目作為一扇門。每道題目都會仔細解答，沿途提出優秀學生會問的問題，然後帶你穿過這扇門：這個想法從何而來，到了大學會演變成甚麼，以及接下來該讀些甚麼。

本書為正在準備 STEP、有天分又有志投身數學的學生而寫。他們想要的不止是模擬試卷：他們想開始像數學系本科生那樣思考，並一探前方有甚麼正等着自己。

- 共 13 章，涵蓋純數學（pure mathematics）、力學（mechanics）、概率（probability）與統計（statistics）
- 136 道 STEP 歷屆試題，每題均附引導式解答
- 每道題目背後的大學數學
- 附答案的延伸練習，以及延伸閱讀

2004 年詳細解答. 全書各章及
所有詳細解答均可免費閱讀、下載及分享，網址為
beyondstep.keithmcnulty.org

文字內容已貢獻予公有領域（CC0 1.0）。STEP 試題 © Cambridge University Press & Assessment。