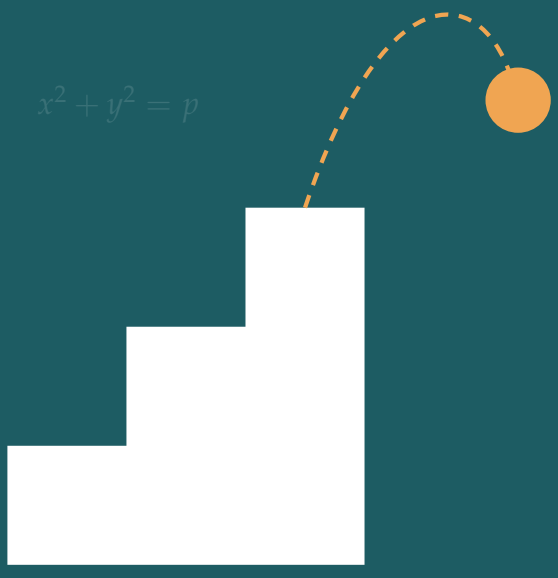
STEP 備試 · 大學數學

Beyond STEP

STEP I, II 及 III

1993 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

1993 年詳細解答

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

1993 年詳細解答

1993 年 STEP 試卷（STEP I, II 及 III）全部 48 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 骰子點數總和的計數 純數學 . . . 3

第 2 題 等比級數 純數學 . . . 7

第 3 題 丟番圖方程 純數學 . . . 11

第 4 題 代換積分法 純數學 . . . 14

第 5 題 絕對值與三角不等式 純數學 . . . 17

第 6 題 曲線描繪 純數學 . . . 21

第 7 題 三次方程的根 純數學 . . . 25

第 8 題 球面與向量 純數學 . . . 29

第 9 題 一階微分方程 純數學 . . . 33

第 10 題 彈及彈簧的平衡 力學 . . . 37

第 11 題 心 力學 . . . 41

第 12 題 體的攔截 力學 . . . 44

第 13 題 與功率 力學 . . . 48

第 14 題 數分佈的等候時間 概率與統計 . . . 52

第 15 題 葉斯定理 概率與統計 . . . 56

第 16 題 何概率 概率與統計 . . . 59

STEP II

第 1 題 組合博弈策略 純數學 . . . 62

第 2 題 三角函數與代數函數的積分 純數學 . . . 66

第 3 題 里卡蒂微分方程 純數學 . . . 70

第 4 題 偏斜線與最接近距離 純數學 . . . 74

第 5 題 極坐標曲線與弧長 純數學 . . . 78

第 6 題 矩陣代數的反例 純數學 . . . 82

第 7 題 模 5 的無窮遞降法 純數學 . . . 86
第 8 題 以數學歸納法證明算術幾何平均不等式 純數學 . . . 89
第 9 題 複數的輻角 純數學 . . . 93
第 10 題 次型與圓錐曲線 純數學 . . . 97
第 11 題 二次阻力的運動 力學 . . . 101
第 12 題 摩擦力與剛體 力學 . . . 105
第 13 題 簡諧運動 力學 . . . 109
第 14 題 彈與恢復係數 力學 . . . 113
第 15 題 心極限定理 概率與統計 . . . 117
第 16 題 松過程 概率與統計 . . . 121

STEP III

第 1 題 弧長與體積 純數學 . . . 125
第 2 題 點與極坐標面積 純數學 . . . 128
第 3 題 陣的 LU 分解 純數學 . . . 132
第 4 題 數求和 純數學 . . . 136
第 5 題 群與四元數 純數學 . . . 140
第 6 題 阿根圖上的軌跡 純數學 . . . 144
第 7 題 雙曲函數聯立方程 純數學 . . . 148
第 8 題 三維空間中的平面與直線 純數學 . . . 152
第 9 題 平方和不等式 純數學 . . . 156
第 10 題 陣變換 純數學 . . . 160
第 11 題 偶與轉動慣量 力學 . . . 164
第 12 題 性繩與能量 力學 . . . 168
第 13 題 質量鏈條 力學 . . . 172
第 14 題 垂圓周運動 力學 . . . 176
第 15 題 候時間的分佈 概率與統計 . . . 180
第 16 題 尾正態分佈 概率與統計 . . . 183

詳細解答

STEP I 1993，第 1 題

純數學 • 骰子點數總和的計數

> **STEP I 1993，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/93-s1-q1

本題在 *Beyond STEP* 第 9.6 節奇怪的骰子中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch09/strange-dice.html

關鍵思路

每一面的顏色都不同，所以擲出某個總和的一種「方法」，就是從每顆骰子各選一面。若其中一顆骰子有 a_k 面標着 k，另一顆有 b_k 面，則擲出總和 n 的方法數目為

$$N(n) = \sum_{i+j=n} a_i\, b_j.$$

在第二部分，我們由最小的總和開始，順着往上讀表。每個新的總和只會帶來一兩個未見過的未知數，所以各個點數的面數會逐一被確定下來，我們只需排除途中出現的少數幾個分支。

解答

總和為 6 的情況

第一顆骰子的面數為 $a_1 = 1$、$a_2 = 2$、$a_3 = 2$、$a_6 = 1$。第二顆的面數為 $b_1 = 1$、$b_3 = 2$、$b_4 = 1$、$b_5 = 1$、$b_6 = 1$。要得到總和 6，第一顆骰子須擲出 1、2 或 3；若擲出 6，第二顆骰子便要有 0 才行。所以

$$N(6) = a_1b_5 + a_2b_4 + a_3b_3 = 1\cdot 1 + 2\cdot 1 + 2\cdot 2 = 7,$$

這正是題目要求的七種方法。

找出新的編號方式

把兩顆骰子稱為 P 和 Q。設 P 有 a_k 面標着 k，Q 有 b_k 面，其中 $k = 1, \ldots, 6$。這些都是非負整數，並且

$$a_1 + \cdots + a_6 = b_1 + \cdots + b_6 = 6.$$

（作為簡單的核對，表中各項加起來是 $36 = 6 \times 6$。）

總和 2 和 3。 $N(2) = a_1 b_1 = 0$，所以兩顆骰子不可能同時有 1。但 $N(3) = a_1 b_2 + a_2 b_1 = 2$，所以總有一顆骰子有 1。把這顆骰子稱為 P。於是 $a_1 \geq 1$ 而 $b_1 = 0$，並且

$$N(3) = a_1 b_2 = 2, \qquad \text{所以} \qquad (a_1, b_2) = (2,1) \text{ 或 } (1,2).$$

$a_1 = 2$、$b_2 = 1$ 的情況不可能成立。 由於 Q 沒有 1，$N(7) \geq a_1 b_6 = 2b_6$，而由 $N(7) = 3$ 可得 $b_6 \leq 1$。在這個情況下，Q 有一面標着 2，我們來驗證它沒有 3 或 4：

- $N(4) = a_1 b_3 + a_2 b_2 = 2b_3 + a_2 = 1$，所以 $2b_3 \leq 1$，這迫使 $b_3 = 0$，繼而 $a_2 = 1$。
- $N(5) = a_1 b_4 + a_2 b_3 + a_3 b_2 = 2b_4 + 0 + a_3 = 1$，所以 $b_4 = 0$ 而 $a_3 = 1$。

因此 Q 有一個 2，沒有 1、3 或 4，其餘五面都是 5 和 6：$b_5 + b_6 = 5$。由於 $b_6 \leq 1$，可得 $b_5 \geq 4$，於是

$$N(6) \geq a_1 b_5 = 2b_5 \geq 8,$$

這與 $N(6) = 4$ 矛盾。所以這個情況不可能出現。

所以 $a_1 = 1$ 而 $b_2 = 2$。 現在我們順着表往上推，並記住 $b_1 = 0$。

- $N(4) = a_1 b_3 + a_2 b_2 = b_3 + 2a_2 = 1$。由於 $2a_2$ 是偶數，$a_2 = 0$ 而 $b_3 = 1$。
- $N(5) = a_1 b_4 + a_2 b_3 + a_3 b_2 = b_4 + 0 + 2a_3 = 1$。同理，$a_3 = 0$ 而 $b_4 = 1$。
- $N(6) = a_1 b_5 + a_2 b_4 + a_3 b_3 + a_4 b_2 = b_5 + 2a_4 = 4$。

到目前為止，Q 已確定了 $b_1 + b_2 + b_3 + b_4 = 0 + 2 + 1 + 1 = 4$ 面，所以 $b_5 + b_6 = 2$。由 $b_5 = 4 - 2a_4$ 可知 b_5 是偶數，所以 $b_5 = 0$ 或 $b_5 = 2$。

> ? **想一想** 我們怎樣在 $b_5 = 0$ 與 $b_5 = 2$ 之間作出取捨？

我們利用下一個總和。P 沒有 2 或 3，而 Q 沒有 1，所以

$$N(7) = a_1 b_6 + a_4 b_3 + a_5 b_2 = b_6 + a_4 + 2a_5 = 3.$$

- 若 $b_5 = 0$，則 $a_4 = 2$ 而 $b_6 = 2$。這時 $N(7) \geq b_6 + a_4 = 4 > 3$，這是不可能的。

- 若 $b_5 = 2$，則 $a_4 = 1$ 而 $b_6 = 0$。這時由 $N(7) = 0 + 1 + 2a_5 = 3$ 得 $a_5 = 1$。P 已確定了 $a_1 + a_4 + a_5 = 3$ 面，所以 $a_6 = 3$。

這個表已把一切都確定下來：

$$P:\ 1,\ 4,\ 5,\ 6,\ 6,\ 6 \qquad\qquad Q:\ 2,\ 2,\ 3,\ 4,\ 5,\ 5.$$

核對表中其餘各項。 我們只用了總和 2 至 7，所以仍須確認這個編號方式能給出其餘各項。取 $a_1 = a_4 = a_5 = 1$、$a_6 = 3$ 以及 $b_2 = 2$、$b_3 = b_4 = 1$、$b_5 = 2$：

$$\begin{aligned} N(8) &= a_4b_4 + a_5b_3 + a_6b_2 = 1 + 1 + 6 = 8, \\ N(9) &= a_4b_5 + a_5b_4 + a_6b_3 = 2 + 1 + 3 = 6, \\ N(10) &= a_5b_5 + a_6b_4 = 2 + 3 = 5, \\ N(11) &= a_6b_5 = 6, \\ N(12) &= a_6b_6 = 0. \end{aligned}$$

以上全部與表相符。所以兩顆骰子的編號必定是 $1,4,5,6,6,6$ 和 $2,2,3,4,5,5$，而這個編號方式確實能給出該表。除了把兩顆骰子的名稱對調之外，這是唯一可行的編號方式。

驗算

- 畫出 $1,4,5,6,6,6$ 對 $2,2,3,4,5,5$ 的 6×6 總和方格表，數一數每個總和出現的次數。你應該得到 $0,2,1,1,4,3,8,6,5,6,0$。
- 用電腦搜尋以 1 至 6 為兩顆骰子編號的所有方式（每顆骰子有 462 個多重集（multi-set）），結果恰好只找到這一對，別無其他，這確認了解是唯一的。

常見錯誤

- 猜出一個符合該表的編號方式便停下來。題目要求的是推理，所以你需要證明該表迫使編號只能如此。
- 忘記了 $a_1 = 2$、$b_2 = 1$ 的情況，或者沒有給出理由便把它排除。
- 由較小的總和確定了各點數的面數後，卻始終沒有核對總和 8 至 12。

STEP 之外

要把 $N(n)$ 包裝得簡潔俐落，可以用生成函數（generating function）。若 $p(x) = \sum a_k x^k$ 而 $q(x) = \sum b_k x^k$，則 $p(x)q(x)$ 中 x^n 的係數正好是 $N(n)$，所以這道題其實是要你把一

個給定的多項式（polynomial）因式分解成兩個具有特定形狀的因式。同樣的想法可以得出 *Sicherman* 骰子（Sicherman dice），即 $1,2,2,3,3,4$ 和 $1,3,4,5,6,8$，它們的總和分佈與兩顆普通骰子相同。Herbert Wilf 的 *generatingfunctionology* 是介紹這個方法的一本生動入門書。

詳細解答

STEP I 1993，第 2 題

純數學 • 等比級數

> **STEP I 1993，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/93-s1-q2

關鍵思路

每個級數（series）都是抽走了部分項的等比數列（geometric progression）。$S_n(r)$ 中的指數是所有不是 3 的倍數的數。$T_n(r)$ 中的指數是偶數或 3 的倍數的數。因此，我們把每個和寫成若干個完整等比數列的組合：取一個完整的數列再減去不要的項，或者把兩個數列相加，再減去重複計算了的項。

解答

偶數次冪之和

左邊是一個首項為 1、公比（common ratio）為 r^2、共有 $n+1$ 項的等比數列。把它記作 E。那麼

$$E - r^2E = (1 + r^2 + \cdots + r^{2n}) - (r^2 + r^4 + \cdots + r^{2n+2}) = 1 - r^{2n+2},$$

因為其餘各項全部互相抵銷。若 $|r| \neq 1$，則 $r^2 \neq 1$，所以可以除以 $1 - r^2$：

$$1 + r^2 + r^4 + \cdots + r^{2n} = \frac{1 - r^{2n+2}}{1 - r^2},$$

正是所要證明的結果。（對實數 r 而言，條件 $r^2 \neq 1$ 恰好就是 $|r| \neq 1$。）

級數 $S_n(r)$

指數 $1,2,4,5,7,8,10,\ldots,3n-1$ 是由 1 至 $3n-1$ 之間所有不是 3 的倍數的整數。它們可以配成一對對的 $3k+1,\ 3k+2$，其中 $k=0,1,\ldots,n-1$。所以

$$S_n(r)=\sum_{k=0}^{n-1}\left(r^{3k+1}+r^{3k+2}\right)=(r+r^2)\sum_{k=0}^{n-1}r^{3k}.$$

最後的那個和是一個公比為 r^3、共有 n 項的等比數列。對實數 r 而言，$r\neq 1$ 即保證 $r^3\neq 1$，所以

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

另一個等價的做法，是取所有冪 $r,r^2,\ldots,r^{3n}$，再剔除指數為 3 的倍數的項：

$$S_n(r)=\frac{r(1-r^{3n})}{1-r}-\frac{r^3(1-r^{3n})}{1-r^3}.$$

求極限時要用的，正是這第二個形式。

極限（limit）。若 $|r|<1$，則當 $n\to\infty$ 時 $|r^{3n}|=|r|^{3n}\to 0$。因此

$$S_n(r)\to\frac{r}{1-r}-\frac{r^3}{1-r^3}.$$

現在 $\dfrac{r}{1-r}=\dfrac{1}{1-r}-1$，而 $\dfrac{r^3}{1-r^3}=\dfrac{1}{1-r^3}-1$。兩個 -1 互相抵銷，餘下

$$S_n(r)\to\frac{1}{1-r}-\frac{1}{1-r^3},$$

正是所要證明的結果。

級數 $T_n(r)$

指數 $0,2,3,4,6,8,9,10,12,\ldots,6n$ 是由 0 至 $6n$ 之間所有偶數或可被 3 整除（或兩者皆是）的整數。

想一想 如果把偶數次冪與指數為 3 的倍數的冪相加，哪些項被計算了兩次？

一個指數既是偶數又是 3 的倍數，當且僅當它是 6 的倍數。所以我們運用容斥原理（inclusion–exclusion）：加上偶數次冪，加上指數為 3 的倍數的冪，再減去指數為 6 的倍數的冪。在 $6n$ 或以下，有

- 偶數次冪 $r^0,r^2,\ldots,r^{6n}$，它們組成一個公比為 r^2、共有 $3n+1$ 項的等比數列；
- 指數為 3 的倍數的冪 $r^0,r^3,\ldots,r^{6n}$，它們組成一個公比為 r^3、共有 $2n+1$ 項的等比數列；

- 指數為 6 的倍數的冪 $r^0, r^6, \ldots, r^{6n}$，它們組成一個公比為 r^6、共有 $n+1$ 項的等比數列。

把第一個結果中的 n 換成 $3n$，再對公比 r^3 和 r^6 運用同樣的論證，便得到

$$\mathrm{T}_n(r) = \frac{1-r^{6n+2}}{1-r^2} + \frac{1-r^{6n+3}}{1-r^3} - \frac{1-r^{6n+6}}{1-r^6}.$$

由於 $|r| \neq 1$，每個分母都不等於零。這就是為甚麼題目在這裏要求 $|r| \neq 1$，而對 S_n 只要求 $r \neq 1$：$r = -1$ 會令 $1-r^2$ 和 $1-r^6$ 變成零。

極限。若 $|r| < 1$，則 r^{6n+2}、r^{6n+3} 和 r^{6n+6} 全部趨向 0。所以

$$\mathrm{T}_n(r) \to \frac{1}{1-r^2} + \frac{1}{1-r^3} - \frac{1}{1-r^6}.$$

利用 $1-r^6 = (1-r^2)(1+r^2+r^4) = (1-r^3)(1+r^3)$，以 $1-r^6$ 為公分母通分，便可化簡為：

$$\lim_{n\to\infty} \mathrm{T}_n(r) = \frac{(1+r^2+r^4)+(1+r^3)-1}{1-r^6} = \frac{1+r^2+r^3+r^4}{1-r^6}.$$

$r > 1$、$r = 1$、$r = -1$ 的情況

我們先數一數項數。在每一組六個指數 $6k, 6k+1, \ldots, 6k+5$ 之中，是偶數或可被 3 整除的有 $6k, 6k+2, 6k+3, 6k+4$，即四個。$k = 0, \ldots, n-1$ 的各組涵蓋了直至 $6n-1$ 的指數，而最後一項是 r^{6n}。所以 $\mathrm{T}_n(r)$ 共有 $4n+1$ 項。

- $r > 1$：每一項 r^e 都滿足 $r^e \geq 1$，所以 $\mathrm{T}_n(r) \geq 4n+1$，因此 $\mathrm{T}_n(r) \to \infty$。
- $r = 1$：每一項都等於 1，所以 $\mathrm{T}_n(1) = 4n+1 \to \infty$。公式在這裏失效，因為它的分母都是零。
- $r = -1$：當 e 是偶數時 $(-1)^e$ 為 $+1$，當 e 是奇數時則為 -1。在每一組中，指數 $6k, 6k+2, 6k+4$ 是偶數，而 $6k+3$ 是奇數，所以每一組的貢獻是 $1+1-1+1=2$。因此

$$\mathrm{T}_n(-1) = 2n+1 \to \infty.$$

在這三種情況下，$\mathrm{T}_n(r)$ 都發散（diverge）至 $+\infty$。留意 $r = -1$ 時同樣發散：指數為 3 的奇數倍的項不足以抵銷那些偶數次項。試與 $1-1+1-\cdots$ 這類簡單的交錯級數（alternating series）比較，後者只是振盪（oscillate）而已。

驗算

- 在 S_n 的公式中代入 $n=1$：$\frac{r(1+r)(1-r^3)}{1-r^3} = r+r^2$，結果正確。

- 在 T_n 的極限中代入 $r = \frac{1}{2}$。公式給出 $\frac{1+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}}{1-\frac{1}{64}} = \frac{92}{63} \approx 1.4603$，而直接把 $\mathrm{T}_n(\frac{1}{2})$ 的幾百項加起來，也得到相同的數值。
- 當 $r \to 1^-$ 時，T_n 的極限的表現近似 $\frac{4}{6(1-r)} = \frac{2}{3(1-r)}$。這與每六個冪之中有四個出現的事實吻合，即約為完整級數 $\frac{1}{1-r}$ 的三分之二。

常見錯誤

- 忘記減去指數為 6 的倍數的項，以致把 $1, r^6, r^{12}, \ldots$ 計算了兩次。
- 數錯各個等比數列的項數。由 0 至 $6n$ 共有 $3n+1$ 個偶數指數，而不是 $3n$ 個。
- 沒有實際把一組的項加起來，便假設 $\mathrm{T}_n(-1)$ 會像交錯級數那樣振盪。

STEP 之外

要數算一個併集（union）中的項，可以把兩個集合相加，再減去它們的交集（intersection），這就是容斥原理。它與生成函數（generating function）結合起來，是組合數學（combinatorics）和數論（number theory）中的標準工具：我們求得的極限，正是與 6 有公因數的非負整數的生成函數。同樣的篩法（sieve）思想，若對每一個質數（prime）都運用一次，便得出黎曼 ζ 函數（Riemann zeta function）的歐拉（Euler）乘積。大學本科的分析學（analysis）會把「當 $|r| < 1$ 時 $r^n \to 0$」這一步嚴格化，並探討無窮級數在甚麼情況下可以這樣重新排列。Abbott 的 *Understanding Analysis*（《理解分析》）對此有很好的講解。

詳細解答

STEP I 1993，第 3 題

純數學 • 丟番圖方程

> **STEP I 1993，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/93-s1-q3

關鍵思路

由於 $p \le q \le r$，倒數（reciprocal）$\frac{1}{p}$ 是三項之中最大的一項。所以三項之和最多是 $\frac{3}{p}$，這樣 p 就只剩下兩三個可能的值。p 一經確定，對餘下兩項再用同一招便可限制 q 的範圍，然後 r 就別無選擇。整個想法是利用大小次序，把一條含三個未知數的方程（或不等式（inequality））化為一次簡短的有限驗證。

解答

第 (i) 部分

由於 $1 \le p \le q \le r$，我們有 $\frac{1}{r} \le \frac{1}{q} \le \frac{1}{p}$。所以

$$1 = \frac{1}{p} + \frac{1}{q} + \frac{1}{r} \le \frac{3}{p},$$

由此得 $p \le 3$。另外 $p \neq 1$，否則 $\frac{1}{q} + \frac{1}{r} = 0$，而這對正數 q, r 是不可能的。所以 $p = 2$ 或 $p = 3$。

情況 $p = 2$。這時 $\frac{1}{q} + \frac{1}{r} = \frac{1}{2}$。由於 $\frac{1}{r} \le \frac{1}{q}$，

$$\frac{1}{2} = \frac{1}{q} + \frac{1}{r} \le \frac{2}{q} \quad \Longrightarrow \quad q \le 4,$$

又因為 $\frac{1}{r} > 0$，必須有 $\frac{1}{q} < \frac{1}{2}$，所以 $q \ge 3$。

- $q = 3$：$\frac{1}{r} = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$，所以 $r = 6$。
- $q = 4$：$\frac{1}{r} = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$，所以 $r = 4$。

情況 $p=3$。這時 $\frac{1}{q}+\frac{1}{r}=\frac{2}{3}\le\frac{2}{q}$，所以 $q\le 3$。連同 $q\ge p=3$，只可能是 $q=3$，於是 $\frac{1}{r}=\frac{1}{3}$，所以 $r=3$。

所有情況都已考慮，因此僅有的解是

$$(p,q,r)=(2,3,6),\quad (2,4,4),\quad (3,3,3),$$

再沒有其他解。

第 (ii) 部分

若 $p=1$，則無論 q 和 r 取甚麼值，和都是 $1+\frac{1}{q}+\frac{1}{r}>1$。這些就是題目已告訴我們的解。（按次序 $p\le q\le r$，這些解就是 $q=n$、$r=m$，其中 $1\le n\le m$。題目在 m 與 n 之間所寫的不等式方向反了，但這並無影響：$p=1$ 對任何 q 和 r 都成立。）所以我們要找的是 $p\ge 2$ 的解。

和之前一樣，$\frac{1}{p}$ 是最大的一項，所以

$$1<\frac{1}{p}+\frac{1}{q}+\frac{1}{r}\le\frac{3}{p}\quad\Longrightarrow\quad p<3.$$

連同 $p\ge 2$，得 $p=2$。現在我們需要

$$\frac{1}{q}+\frac{1}{r}>\frac{1}{2}.$$

? **想一想** 為甚麼這次 q 的上界是嚴格（strict）不等式？這有沒有影響？

由於 $\frac{1}{r}\le\frac{1}{q}$，我們有 $\frac{1}{2}<\frac{1}{q}+\frac{1}{r}\le\frac{2}{q}$，所以 $q<4$。正是這個嚴格不等式排除了 $q=4$。（在第 (i) 部分，$q=4$ 是容許的，並給出等號成立的情況 $(2,4,4)$。）連同 $q\ge p=2$，得 $q=2$ 或 $q=3$。

- $q=2$：和是 $\frac{1}{2}+\frac{1}{2}+\frac{1}{r}=1+\frac{1}{r}>1$，對每一個 r 都成立。所以對每個整數 $r\ge 2$，$(2,2,r)$ 都是解。
- $q=3$：我們需要 $\frac{1}{r}>\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$，所以 $r<6$。連同 $r\ge q=3$，得 $r=3,4,5$。

所以，除了 $p=1$ 的那一族解之外，其餘的解是

$$(p,q,r)=(2,2,r)\text{ 其中 }r\text{ 為任何整數，}r\ge 2,\qquad (2,3,3),\quad (2,3,4),\quad (2,3,5).$$

以上每一步都是別無選擇的，所以再沒有其他解。

驗算

直接代入。例如 $\frac{1}{2}+\frac{1}{3}+\frac{1}{5}=\frac{31}{30}>1$，但 $\frac{1}{2}+\frac{1}{3}+\frac{1}{6}=1$ 剛好相等。這說明了為甚麼 $(2,3,5)$ 是 $q=3$ 那一族的最後一個，而 $(2,3,6)$ 則是第 (i) 部分的邊界情況。同樣，

$(2,4,4)$ 給出的和剛好是 1，所以 $p=2$、$q=4$ 的組合全都進不了第 (ii) 部分。用電腦搜尋所有 $1 \le p \le q \le r < 400$，結果與兩張清單完全吻合。

常見錯誤

- 在第 (ii) 部分漏掉無窮多個解的那一族 $(2,2,r)$。一旦 $p=q=2$，頭兩項加起來已經是 1。
- 在第 (ii) 部分把 $p<3$ 寫成 $p \le 3$，然後在 $p=3$ 上浪費時間（這時和最多只是 1）。
- 只列出解，卻沒有說明每種情況都已檢查。「再沒有其他解」這一句佔了一半分數。

STEP 之外

$\frac{1}{p}+\frac{1}{q}+\frac{1}{r} > 1$ 的解，正好就是那些能令內角為 $\frac{\pi}{p}, \frac{\pi}{q}, \frac{\pi}{r}$ 的三角形存在於球面上的三元組。它們與三維空間中的有限對稱羣（symmetry group）互相對應：二面體羣（dihedral）$(2,2,r)$、四面體羣（tetrahedral）$(2,3,3)$、八面體羣（octahedral）$(2,3,4)$ 和二十面體羣（icosahedral）$(2,3,5)$。第 (i) 部分則給出平面的三種正則鑲嵌（regular tiling）。像這樣要求整數解的方程，稱為丟番圖方程（Diophantine equation）。Davenport 的 *The Higher Arithmetic* 是很好的入門書，而 Allenby 的 *Rings, Fields and Groups* 則講述了這些對稱羣。

詳細解答

STEP I 1993，第 4 題

純數學 • 代換積分法

STEP I 1993，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/93-s1-q4

關鍵思路

代換（substitution）$t = \pi - x$ 把區間 $[0, \pi]$ 反射到它自身，而且不會改變 $\sin x$，因為 $\sin(\pi - t) = \sin t$。於是那個礙事的因子 x 變成 $\pi - t$，而我們要求的積分會帶着負號在右邊再次出現，這樣就可以把它解出來。至於其後的積分，主要工作是把 $1 + \cos^2 x$ 用 $\sin x$ 表示。在 $[\pi, 2\pi]$ 上，平移 π 就可以回到 $[0, \pi]$。

解答

一般結果

設 $I = \int_0^\pi x\,\mathrm{f}(\sin x)\,\mathrm{d}x$，並令 $t = \pi - x$，則 $\mathrm{d}x = -\mathrm{d}t$。當 $x = 0$ 時 $t = \pi$，當 $x = \pi$ 時 $t = 0$。另外 $\sin x = \sin(\pi - t) = \sin t$。因此

$$\begin{aligned} I &= \int_\pi^0 (\pi - t)\,\mathrm{f}(\sin t)\,(-\mathrm{d}t) = \int_0^\pi (\pi - t)\,\mathrm{f}(\sin t)\,\mathrm{d}t \\ &= \pi \int_0^\pi \mathrm{f}(\sin t)\,\mathrm{d}t - \int_0^\pi t\,\mathrm{f}(\sin t)\,\mathrm{d}t. \end{aligned}$$

最後一個積分又是 I，因為積分變量（variable of integration）叫甚麼名字並無關係。所以 $2I = \pi \int_0^\pi \mathrm{f}(\sin t)\,\mathrm{d}t$，由此得

$$\int_0^\pi x\,\mathrm{f}(\sin x)\,\mathrm{d}x = \frac{\pi}{2} \int_0^\pi \mathrm{f}(\sin x)\,\mathrm{d}x,$$

這正是所要證明的結果。（我們需要這些積分存在，例如 f 在 $[-1, 1]$ 上連續（continuous）即可。）

$[0,\pi]$ 上的積分

想一想 被積函數（integrand）含有 $\cos^2 x$，而不只是 $\sin x$。它真的是 $x\,\mathrm{f}(\sin x)$ 的形式嗎？

是的。由於 $\cos^2 x = 1-\sin^2 x$，我們有 $1+\cos^2 x = 2-\sin^2 x$，所以

$$\frac{x\sin x}{1+\cos^2 x} = x\,\mathrm{f}(\sin x) \qquad \text{其中} \qquad \mathrm{f}(s) = \frac{s}{2-s^2},$$

而且 f 在 $[-1,1]$ 上連續。根據一般結果，

$$\int_0^\pi \frac{x\sin x}{1+\cos^2 x}\,\mathrm{d}x = \frac{\pi}{2}\int_0^\pi \frac{\sin x}{1+\cos^2 x}\,\mathrm{d}x.$$

餘下的積分很容易，因為 $\sin x$ 是 $\cos x$ 的導數（derivative）（差一個負號）。代入 $u=\cos x$，其中 $\mathrm{d}u = -\sin x\,\mathrm{d}x$，

$$\int_0^\pi \frac{\sin x}{1+\cos^2 x}\,\mathrm{d}x = \int_1^{-1}\frac{-\mathrm{d}u}{1+u^2} = \int_{-1}^1 \frac{\mathrm{d}u}{1+u^2} = \Big[\arctan u\Big]_{-1}^1 = \frac{\pi}{4} - \Big(-\frac{\pi}{4}\Big) = \frac{\pi}{2}.$$

因此

$$\int_0^\pi \frac{x\sin x}{1+\cos^2 x}\,\mathrm{d}x = \frac{\pi}{2}\cdot\frac{\pi}{2} = \frac{\pi^2}{4}.$$

$[0,2\pi]$ 上的積分

一般結果只適用於 $[0,\pi]$，所以我們把積分範圍拆開：

$$\int_0^{2\pi} = \int_0^\pi + \int_\pi^{2\pi}.$$

想一想 怎樣才能把 $\int_\pi^{2\pi}$ 化回 $[0,\pi]$ 上的積分？

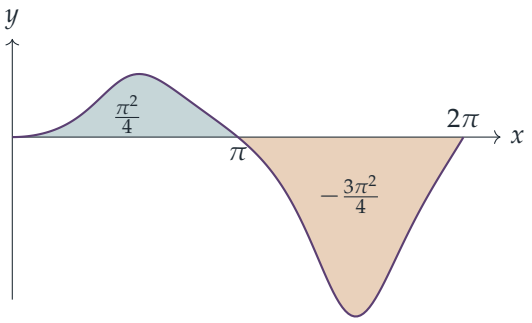
平移 π：令 $x = u+\pi$，於是 u 由 0 變到 π。這時 $\sin x = -\sin u$ 而 $\cos x = -\cos u$，所以 $\cos^2 x = \cos^2 u$。因此

$$\begin{aligned}\int_\pi^{2\pi} \frac{x\sin x}{1+\cos^2 x}\,\mathrm{d}x &= \int_0^\pi \frac{(u+\pi)(-\sin u)}{1+\cos^2 u}\,\mathrm{d}u \\ &= -\int_0^\pi \frac{u\sin u}{1+\cos^2 u}\,\mathrm{d}u - \pi\int_0^\pi \frac{\sin u}{1+\cos^2 u}\,\mathrm{d}u \\ &= -\frac{\pi^2}{4} - \pi\cdot\frac{\pi}{2} = -\frac{3\pi^2}{4},\end{aligned}$$

這裏用了上一部分的兩個結果。把兩部分相加，

$$\int_0^{2\pi} \frac{x\sin x}{1+\cos^2 x}\,\mathrm{d}x = \frac{\pi^2}{4} - \frac{3\pi^2}{4} = -\frac{\pi^2}{2}.$$

從圖像可以看出答案為甚麼是負數。在 $[\pi,2\pi]$ 上，因子 $\sin x$ 是負的，而且那裏的因子 x 比在 $[0,\pi]$ 上的大。所以軸下方的面積多於軸上方的面積。



驗算

- 取 $f \equiv 1$：一般結果給出 $\int_0^\pi x\,dx = \frac{\pi}{2}\cdot\pi = \frac{\pi^2}{2}$，這是正確的。
- 被積函數在 $(0,\pi)$ 上為正，所以第一個答案必須是正數，而 $\pi^2/4 \approx 2.467$ 確實如此。用辛普森法則（Simpson's rule）計算，第一個積分得 2.4674011，第二個得 -4.9348022，與 $\pi^2/4$ 和 $-\pi^2/2$ 吻合。
- 隨機選取幾個 f 作數值檢驗，一般恆等式（identity）同樣成立。

常見錯誤

- 直接在 $[0,2\pi]$ 上套用一般結果。這個結果依賴反射 $x \mapsto \pi - x$ 把區間映射到它自身，而在 $[0,2\pi]$ 上這樣做會得出錯誤答案 $\frac{\pi}{2}\cdot 0 = 0$（那裏正確的反射是 $x \mapsto 2\pi - x$，而 $\sin x$ 在這個反射下會變號）。
- 漏掉 $dx = -dt$ 中的負號，或者忘記把積分上下限調換回來。
- 寫成 $\arctan(-1) = 3\pi/4$：反正切（arctangent）的取值範圍是 $(-\pi/2, \pi/2)$，所以 $\arctan(-1) = -\pi/4$。

STEP 之外

這裏的竅門是利用定義域（domain）的一種對稱性（symmetry），而被積函數幾乎保持這種對稱：把 x 換成 $\pi - x$，再把積分的兩個版本取平均。用同樣的想法，完全不必積分就可以得到 $\int_0^{\pi/2} \frac{\sin x}{\sin x + \cos x}\,dx = \frac{\pi}{4}$。在大學裏，這類對稱性論證在分析學（analysis）中隨處可見，也出現在傅立葉級數（Fourier series）中，在那裏函數的奇偶性（parity）決定哪些項得以保留。Siklos 的 *Advanced Problems in Mathematics* 有更多這類代換的 STEP 式練習，而 Abbott 的 *Understanding Analysis* 則嚴格地建立了變量代換法則（change-of-variables rule）。

詳細解答

STEP I 1993，第 5 題

純數學 • 絕對值與三角不等式

STEP I 1993，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/93-s1-q5

關鍵思路

一切都建基於三角不等式（triangle inequality）：和的長度不會超過各項長度之和。逐項運用這個不等式，便可用一個關於 $|z|$ 的等比級數（geometric series）來界定多項式的大小；而當 $|z| \leq \frac{1}{2}$ 時，這個級數始終嚴格小於 1。若要成為根（root），多項式部分必須等於 -1，而它的模（modulus）恰好是 1，所以根不可能落在圓盤（disc）$|z| \leq \frac{1}{2}$ 之內。

解答

模與乘積法則

設 $z = x + \mathrm{i}y$，其中 x, y 為實數，則模就是原點到點 (x, y) 的距離：

$$|z| = \sqrt{x^2 + y^2} \qquad \text{(取非負平方根)}.$$

設 $z_1 = a + \mathrm{i}b$ 及 $z_2 = c + \mathrm{i}d$，其中 a, b, c, d 為實數。那麼 $z_1 z_2 = (ac - bd) + \mathrm{i}(ad + bc)$，所以

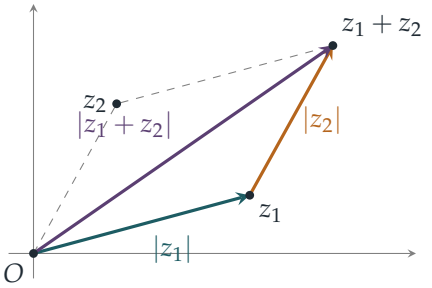
$$\begin{aligned} |z_1 z_2|^2 &= (ac - bd)^2 + (ad + bc)^2 \\ &= a^2c^2 - 2abcd + b^2d^2 + a^2d^2 + 2abcd + b^2c^2 \\ &= a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 \\ &= (a^2 + b^2)(c^2 + d^2) = |z_1|^2 |z_2|^2. \end{aligned}$$

$|z_1 z_2|$ 和 $|z_1||z_2|$ 都是非負數，所以兩邊取平方根便得 $|z_1 z_2| = |z_1||z_2|$，這正是所要證明的結果。

三角不等式

在阿根圖（Argand diagram）上，設 O 為原點，A 為點 z_1，C 為點 z_1+z_2。那麼 $\overrightarrow{AC}$ 代表 z_2（它與由 O 到 z_2 的向量相同），所以 OAC 是一個三角形，其邊長為

$$OA = |z_1|, \qquad AC = |z_2|, \qquad OC = |z_1+z_2|.$$



兩點之間以直線最短：由 O 直接走到 C，不會比繞道 A 更長。因此

$$|z_1+z_2| \le |z_1| + |z_2|,$$

等號成立，當且僅當 O、A、C 三點共線而 A 位於另外兩點之間，也就是說，z_1 與 z_2 指向同一方向（或其中一個為 0）。

如果你較喜歡代數方法：以 $\bar{z}$ 表示共軛（conjugate），則

$$\begin{aligned} |z_1+z_2|^2 &= |z_1|^2 + |z_2|^2 + 2\,\mathrm{Re}(z_1\bar{z}_2) \\ &\le |z_1|^2 + |z_2|^2 + 2|z_1||z_2| = (|z_1|+|z_2|)^2, \end{aligned}$$

這是因為複數的實部不會超過它的模，而根據乘積法則，$|z_1\bar{z}_2| = |z_1||z_2|$。

重複運用這個不等式（對項數作一個簡單的歸納法（induction）），便得

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

對任意複數 $w_1, \ldots, w_n$ 均成立。

多項式的上界

假設對每個 j 都有 $|a_j| \le 1$，並且 $|z| \le \frac{1}{2}$。那麼

$$\begin{aligned} \left|a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z\right| &\le |a_n||z|^n + |a_{n-1}||z|^{n-1} + \cdots + |a_1||z| \\ &\le |z|^n + |z|^{n-1} + \cdots + |z| \\ &\le \left(\tfrac{1}{2}\right)^n + \left(\tfrac{1}{2}\right)^{n-1} + \cdots + \tfrac{1}{2} \\ &= 1 - \left(\tfrac{1}{2}\right)^n \\ &< 1. \end{aligned}$$

第一行用了 n 項的三角不等式，再配合乘積法則(重複使用，從而得出 $|a_k z^k| = |a_k||z|^k$)。第二行用了 $|a_j| \le 1$，第三行用了 $0 \le |z| \le \frac{1}{2}$，第四行則是把有限等比級數求和。這就是所要求的嚴格不等式。

想一想 那個嚴格不等式是從哪裏來的？上面每一步都只是「$\le$」。

它來自有限等比級數：$\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n}$ 與 1 之間恰好相差 $\frac{1}{2^n} > 0$。接下來的推論所需要的，正是這個差距。

根

(試卷最後一行寫的是 $|x| > \frac{1}{2}$；這顯然是 $|w| > \frac{1}{2}$ 的誤印，因為所討論的根是 w。)

設 w 為一個根，則

$$a_n w^n + a_{n-1} w^{n-1} + \cdots + a_1 w = -1, \qquad \text{因此} \qquad |a_n w^n + \cdots + a_1 w| = |-1| = 1.$$

用反證法（proof by contradiction），假設 $|w| \le \frac{1}{2}$。那麼在上一部分取 $z = w$，便得 $|a_n w^n + \cdots + a_1 w| < 1$，這與模恰好等於 1 矛盾。因此每個根都滿足 $|w| > \frac{1}{2}$。

驗算

- 這個上界在極限情況下可以達到：當每個 $a_j = 1$ 而 $z = \frac{1}{2}$ 時，左邊恰好等於 $1 - 2^{-n}$。
- 關於根的結論同樣是最佳的。取每個 $a_j = -1$，得到 $1 - z - z^2 - \cdots - z^n = 0$，它的正根在 $n = 1$ 時是 1，在 $n = 2$ 時是 0.618（即黃金比的共軛值），在 $n = 5$ 時是 0.5087，在 $n = 10$ 時是 0.50025：總是大於 $\frac{1}{2}$，卻趨向於它。
- 數值方面，我們測試了 50 000 組隨機選取的數據，其中 a_j 取自單位圓盤，z 取自半徑為 $\frac{1}{2}$ 的圓盤（所得的最大值為 0.79）；又求出了 2000 個這類隨機多項式的所有根：根的模最小為 0.562。

常見錯誤

- 寫成 $|a_1 z + \cdots + a_n z^n| = |a_1||z| + \cdots + |a_n||z|^n$。三角不等式只能給出「$\le$」。
- 使用無窮級數之和 $\frac{1}{2} + \frac{1}{4} + \cdots = 1$，這樣只能得到 ≤ 1，令推論無法成立。應該用有限和。
- 忘記說明為何 $|z^k| = |z|^k$：它可由你剛剛證明的乘積法則推出。

STEP 之外

這是多項式根的界的入門版本，而這類界正是數值求根方法的基礎：柯西（Cauchy）界把多項式的每個根都限制在一個圓盤之內，其半徑只取決於係數的大小；而同樣的「主導項勝出」論證，也是複分析（complex analysis）中魯歇（Rouché）定理的核心。三角不等式本身就是度量（metric）的定義性質，而度量這個抽象的距離概念，正是大學一年級分析課程的核心。若想看一個以幾何觀點出發、處理得極為優美的講法，可讀 Needham 的 *Visual Complex Analysis*。

詳細解答

STEP I 1993，第 6 題

純數學 • 曲線描繪

STEP I 1993，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/93-s1-q6

關鍵思路

一切取決於當 $N = 10^{100}$ 時 x^N 的表現。若 $|x| < 1$，即使只比 1 小一點點，x^N 也小得難以想像；若 $|x| > 1$，即使只比 1 大一點點，它也大得難以想像；而在 $x = \pm 1$ 時，它恰好等於 1（因為 N 是偶數）。所以分式 $x^N/(1+x^N)$ 的表現就像一個開關：在 $(-1,1)$ 之內是 0，在其外是 1。至於 h，思路剛好相反：$10^{-N}x$ 實在太小，以致 sin 與它的自變量根本無從分辨。

解答

以下一律記

$$s(x) = \frac{x^N}{1+x^N} = \frac{1}{1+x^{-N}} \quad (x \neq 0), \qquad s(0) = 0,$$

於是 $\mathrm{f}(x) = 2 + s(x)$，而 $\mathrm{g}(x) = x\,s(x)$。

f 的特徵

- 對稱性與值域（*range*）。N 是偶數，所以 $x^N = (-x)^N \geq 0$。因此 f 是偶函數（even function），而由 $0 \leq s(x) < 1$ 可得 $2 \leq \mathrm{f}(x) < 3$。
- 當 $|x| < 1$ 時的水平 $y = 2$。這時 $x^N \to 0$ 的速度快得驚人：例如 $0.99^N = 10^{N \log_{10} 0.99} \approx 10^{-4.4\times 10^{97}}$。所以 $s(x) \approx x^N$ 遠遠小於任何圖像所能顯示的程度，在任何看得見的精確度下 $\mathrm{f}(x) = 2$。準確而言，$\mathrm{f}(0) = 2$。
- 當 $|x| > 1$ 時的水平 $y = 3$。現在 x^{-N} 同樣微小，所以在任何看得見的精確度下 $s(x) = 1/(1+x^{-N})$ 就是 1，而 $\mathrm{f}(x) = 3$，並且是從下方趨近。（事實上，當 $|x| \to \infty$ 時，直線 $y = 3$ 是一條漸近線（asymptote）。）

- 在 $x = \pm 1$ 處的跳躍。在 $x = \pm 1$ 處，$x^N = 1$，所以 $\mathrm{f}(\pm 1) = \frac{5}{2}$，正好在半途。這個函數是連續（continuous）而且可微（differentiable）的，所以草圖中的垂直線段其實是極之陡峭的攀升。求導可得

$$\mathrm{f}'(x) = \frac{Nx^{N-1}}{(1+x^N)^2}, \qquad \mathrm{f}'(1) = \frac{N}{4} = 2.5 \times 10^{99}.$$

想一想 圖像中「垂直」的部分究竟有多窄？

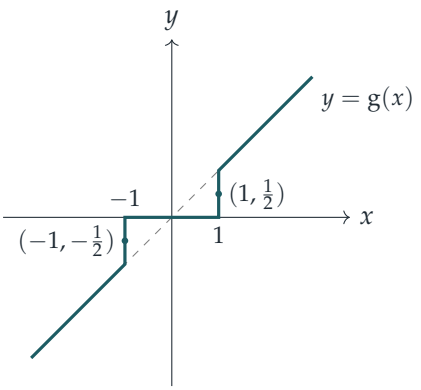
設 $x = 1 + \delta$，其中 δ 很小，則 $x^N = \mathrm{e}^{N\ln(1+\delta)} \approx \mathrm{e}^{N\delta}$，而 $s(x) \approx 1/(1+\mathrm{e}^{-N\delta})$。當 $N\delta$ 由 $-\ln 99$ 增至 $\ln 99$ 時，它由 0.01 攀升至 0.99，也就是說，整個過程只發生在一個寬度約為 9×10^{-100} 的區間內。在任何圖上，這都是一條垂直線。根據對稱性，在 $x = -1$ 處也是如此。（f′ 的正負號還告訴我們，f 在 $x > 0$ 時遞增，在 $x < 0$ 時遞減，因為 x^{N-1} 與 x 同號。）

g 的圖像

由於 $\mathrm{g}(x) = x\,s(x)$，開關 s 現在乘在直線 $y = x$ 上：

- 當 $|x| < 1$ 時，$s(x)$ 可以忽略，所以在任何看得見的精確度下 $\mathrm{g}(x) = 0$（在 $x = 0$ 處恰好是 0）；
- 當 $|x| > 1$ 時，在任何看得見的精確度下 $s(x)$ 就是 1，所以 $\mathrm{g}(x) = x$；準確而言，$x - \mathrm{g}(x) = \dfrac{x}{1+x^N}$，這個差可以忽略；
- 在 $x = \pm 1$ 處，$\mathrm{g}(\pm 1) = \pm\frac{1}{2}$，正好在每個跳躍的半途。

g 是奇函數（odd function），因為 s 是偶函數而 x 是奇函數。它也是遞增的：$\mathrm{g}'(x) = s(x) + x\,s'(x)$，而這兩項都 ≥ 0，因為 x 與 $s'(x)$ 同號。在 $x = 1$ 處斜率是 $\mathrm{g}'(1) = \frac{1}{2} + \frac{N}{4}$，所以我們再次看到一個看似垂直的跳躍，這次是在 $x = 1$ 處由 0 升至 1，以及在 $x = -1$ 處由 -1 升至 0。



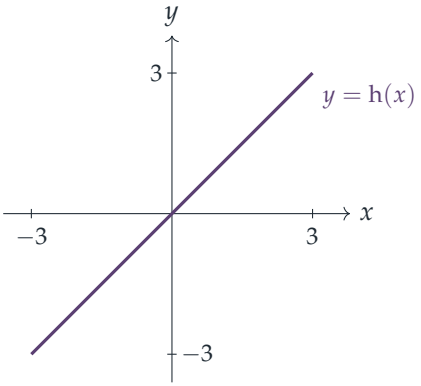
g 在 $[-3,3]$ 上的草圖：在 $(-1,1)$ 上為零，在其外是直線 $y=x$（虛線），兩者之間由經過 $(\pm1,\pm\frac{1}{2})$ 的近乎垂直的攀升連接起來。

h 的圖像

設 $u=10^{-N}x$。當 $|x|\le 3$ 時，我們有 $|u|\le 3\times 10^{-N}$，這是極其微小的。對於細小的 u，$\sin u = u-\frac{u^3}{6}+\cdots$，而當 $u>0$ 時我們有上下界 $u-\frac{u^3}{6}\le \sin u\le u$。乘以 10^N，當 $0\le x\le 3$ 時，

$$0\le x-\mathrm{h}(x)\le \frac{x^3}{6}10^{-2N}\le \frac{27}{6}\times 10^{-2N},$$

而 h 是奇函數，所以當 $-3\le x\le 0$ 時同樣的結論成立，只是正負號相反。因此在 $[-3,3]$ 上，在任何想像得到的精確度下，h 的圖像就是經過原點的直線 $y=x$，在 $x>0$ 時比它低一絲，在 $x<0$ 時比它高一絲。它的斜率是 $\mathrm{h}'(x)=\cos(10^{-N}x)$，與 1 的差距約在 10^{-2N} 之內。



正弦波仍然存在，只是被極度拉闊了：它的第一個極大值位於 $x=\frac{\pi}{2}\times 10^N$，高度為 10^N，所以在窗口 $[-3,3]$ 內只看得見經過原點、看似筆直的一小段。

驗算

- 以 $s=1/(1+\mathrm{e}^{-N\ln|x|})$ 直接計算：$s(0.999)$ 和 $s(1.001)$ 在機器精確度下分別是 0 和 1，而 $s(\pm1)=\frac{1}{2}$ 是準確值。
- 取一個不太大的 N，例如 $N=10$ 或 200，數值微分證實 $s'(1)=N/4$ 和 $\mathrm{g}'(1)=\frac{1}{2}+N/4$，而細密的格點顯示 g 在 $[-3,3]$ 上遞增。增大 N 只會令階梯更加陡峭分明。
- 對於隨機選取的 $x\in[-3,3]$，$10^{100}\sin(10^{-100}x)$ 與 x 在完整的雙精度下吻合，而按比例縮小的類比 $M\sin(x/M)$ 在 $M=1000$ 時符合上下界 $0<x-M\sin(x/M)\le x^3/(6M^2)$。

常見錯誤

- 忘記 N 是偶數：當 $x < -1$ 時，x^N 極大而且是正數，所以 f 在那裏是 3，而不是別的數值。
- 把階梯說成真正的不連續點。三個函數全都是連續的（其實還是可微的）；那些攀升只是陡峭得畫不出來。
- 把 h 當成一個振盪的圖像。在 $[-3,3]$ 上，它離任何轉向點都還遠得很。

STEP 之外

函數 $x^n/(1+x^n)$ 在 $n \to \infty$（只取偶數值）時逐點收斂（pointwise convergence）於一個階梯函數（step function），它在 $(-1,1)$ 之內是 0，在 ± 1 處是 $\frac{1}{2}$，在其外是 1：這是一列光滑函數，其極限卻不連續。這是一個經典例子，說明為何分析學家需要一致收斂（uniform convergence），而這正是實分析（real analysis）入門課程的核心概念。函數 h 則展示了另一面：當 $M \to \infty$ 時，$M\sin(x/M) \to x$ 在有界區間上一致收斂。Abbott 的 *Understanding Analysis*（實分析入門讀本）把這兩個概念都講得十分易讀。

詳細解答

STEP I 1993，第 7 題

純數學 • 三次方程的根

> **STEP I 1993，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/93-s1-q7

關鍵思路

首項係數為正的三次函數（cubic）由 $-\infty$ 上升至 $+\infty$，所以關於它的根（root）的一切，都由它的兩個轉向點（turning point）決定。如果曲線的極大值（local maximum）在橫軸之上，而極小值（local minimum）在橫軸之下，圖像必定與橫軸相交三次；否則相交的次數較少。對於 $x^3 + ax^2 + b$，轉向點很容易準確求出，因為它的導數（derivative）可以因式分解。

解答

轉向點

對於 $\mathrm{f}(x) = x^3 + Ax^2 + B$，

$$\mathrm{f}'(x) = 3x^2 + 2Ax = x(3x + 2A),$$

所以駐點（stationary point）位於 $x = 0$ 和 $x = -\frac{2A}{3}$。函數在這兩點的值是 $\mathrm{f}(0) = B$ 和

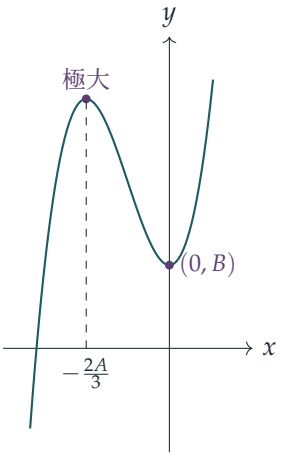
$$\mathrm{f}\left(-\tfrac{2A}{3}\right) = -\frac{8A^3}{27} + \frac{4A^3}{9} + B = \frac{4A^3}{27} + B.$$

此外 $\mathrm{f}''(x) = 6x + 2A$，所以 $\mathrm{f}''(0) = 2A$ 而 $\mathrm{f}''(-\frac{2A}{3}) = -2A$。當 $A > 0$ 時，位於 $x = 0$ 的點是極小點，位於 $x = -\frac{2A}{3}$ 的點是極大點；當 $A < 0$ 時，兩者的角色互換。

$A > 0$、$B > 0$ 時的草圖

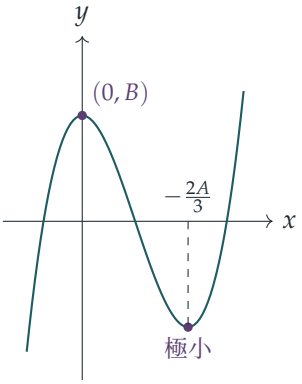
極大點位於 $\left(-\frac{2A}{3},\, B + \frac{4A^3}{27}\right)$，在 y 軸的左方，而且比 B 高；極小點位於 $(0, B)$，在正 y 軸上。兩個轉向點都在 x 軸之上，所以曲線恰好與它相交一次，交點在極大點的左方。（圖

中取 $A = 3$，$B = 2$：極大點 $(-2, 6)$，極小點 $(0, 2)$。)



$A < 0$、$B > 0$ 時的草圖

現在極大點位於 $(0, B)$，而極小點位於 $\left(-\frac{2A}{3}, B + \frac{4A^3}{27}\right)$，在 y 軸的右方。由於 $A^3 < 0$，極小值小於 B，它可以是正數、零或負數。必定有一個根滿足 $x < 0$，因為 $\mathrm{f}(0) = B > 0$，而當 $x \to -\infty$ 時 $\mathrm{f} \to -\infty$。如果極小點跌到橫軸之下，就另有兩個根滿足 $x > 0$（如圖所示，$A = -3$，$B = 2$：極小點 $(2, -2)$，根為 $1 - \sqrt{3}$、1、$1 + \sqrt{3}$）。如果極小點在橫軸上，曲線就在該處與橫軸相切。如果極小點在橫軸之上，那個負根就是唯一的根。



數算根的數目

現在考慮 $\mathrm{g}(x) = x^3 + ax^2 + b$。如果 $a = 0$，則 g 嚴格遞增（strictly increasing）（它的導數 $3x^2$ 只在 0 處為零），所以它恰好有一個實根。因此假設 $a \neq 0$。由上文的結果，g 有兩個相異的轉向點，函數值分別是 b 和 $b + \frac{4a^3}{27}$。

想一想 兩個轉向點的函數值怎樣告訴我們根的數目？

設轉向點為 $x_1 < x_2$。這個三次函數在 $(-\infty, x_1]$ 上嚴格遞增，在 $[x_1, x_2]$ 上嚴格遞減，在 $[x_2, \infty)$ 上嚴格遞增，所以在每一段上最多有一個根。如果 $\mathrm{g}(x_1) > 0 > \mathrm{g}(x_2)$，由介值定理（intermediate value theorem）可知在 $(-\infty, x_1)$、(x_1, x_2) 和 (x_2, ∞) 中各有一個根（這裏用到當 $x \to \mp\infty$ 時 $\mathrm{g} \to \mp\infty$），所以共有三個相異的根。相反，如果兩個值同號，或者其中一個為零，則至少有一段的內部沒有根，根的數目最多是兩個。極大值永遠是兩個值之中較大的一個，所以「一正一負」等同於「兩者之積為負」。因此 g 有三個相異實根，當且僅當

$$b\left(b + \frac{4a^3}{27}\right) < 0 \quad \Longleftrightarrow \quad 27b^2 + 4a^3 b < 0.$$

（乘以 $27 > 0$ 不會改變不等式的方向。）這個結論也涵蓋 $a = 0$ 的情況，此時 $27b^2 + 4a^3 b = 27b^2 \geq 0$。

第一個條件

假設 $27b^2 + 3a^3 b < 0$。那麼 $3a^3 b < -27b^2 \leq 0$，所以 $a^3 b < 0$，因此

$$27b^2 + 4a^3 b = \left(27b^2 + 3a^3 b\right) + a^3 b < 0 + 0 = 0.$$

根據上述判別準則，方程有三個相異實根，這正是所要證明的。

第二個條件

按照試卷上所印的寫法，第二個條件重複了不等式 $27b^2 + 4a^3 b < 0$，而用「$<$」的話，這個命題是錯的：取 $a = -3$，$b = 1$，得 $27 - 108 = -81 < 0$，但 $x^3 - 3x^2 + 1$ 在 $x = 0$ 處取值 1，在 $x = 2$ 處取值 -3，所以它有三個相異實根。題目原意的不等式必定是方向相反的那一個。判別準則顯示，如果

$$27b^2 + 4a^3 b > 0$$

（其實只要 $27b^2 + 4a^3 b \geq 0$），方程的相異實根就少於三個。至此全題完成。

所以準確的分界線是 $27b^2 + 4a^3 b = 0$。含 $3a^3$ 的條件對於有三個根是充分但非必要的：取 $a = -2$，$b = 1$，得 $27b^2 + 3a^3 b = 3 > 0$，但 $27b^2 + 4a^3 b = -5 < 0$，而 $x^3 - 2x^2 + 1 = (x-1)(x^2 - x - 1)$ 有三個根 1 和 $\frac{1 \pm \sqrt{5}}{2}$。

驗算

取 $A = -3$，$B = 2$，得 $x^3 - 3x^2 + 2 = (x-1)(x^2 - 2x - 2)$，它的根 1、$1 \pm \sqrt{3}$ 與草圖吻合，而且 $27 \cdot 4 + 4(-27)(2) = -108 < 0$，正如判別準則所預測。數值方面，對於約

80,000 對滿足第一個條件的隨機有理數 (a, b)，兩個駐點值總是異號。在 200,000 對數值中，「有三個相異根」與 $27b^2 + 4a^3b < 0$ 完全吻合。

常見錯誤

- 忘記 $a = 0$ 的情況，此時兩個轉向點合併成一個拐點（point of inflection）。
- 把轉向點的函數值為零當作有三個根：這種情況得到的是重根（repeated root），所以只有兩個相異的根。
- 嘗試完全按照所印的寫法去證明第二個命題，而沒有留意到它與判別準則矛盾。

STEP 之外

$-(4a^3b + 27b^2)$ 這個量是 $x^3 + ax^2 + b$ 的判別式（discriminant），是 $b^2 - 4ac$ 在三次方程中的對應物。它等於各根之差的平方之積，所以有三個相異實根時它是正數，有重根時是零，有兩個複數根時是負數。判別式和根的對稱函數（symmetric functions）會在伽羅瓦理論（Galois theory）中進一步發展；Allenby 的 *Rings, Fields and Groups*（環、域與羣）是一本很好的入門書。

詳細解答

STEP I 1993，第 8 題

純數學 • 球面與向量

> **STEP I 1993，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/93-s1-q8

關鍵思路

在第 (i) 部分，把原點和坐標軸放在具有對稱性的位置。如果球心位於 z 軸上，而平面是 $z = 0$，那麼方程幾乎會替我們完成所有工作。兩個球面的情況也不會更難，因為把兩者的方程相減，所有平方項都會抵消，剩下的是一個平面的方程。在第 (ii) 部分，坐標同樣把一切化為純量積（scalar product）的計算。把 B 放在原點，這樣兩根桿的頂端都可以由 E 量起。

解答

第 (i) 部分

球面與平面。設球面（sphere）的球心為 O'，半徑為 $r > 0$，並設 N 為由 O' 至平面的垂足（foot of the perpendicular）。取 N 為原點，平面為 $z = 0$，並令 z 軸沿 NO' 方向。於是 $O' = (0, 0, d)$，其中 $d \geq 0$ 是球心到平面的距離。球面的方程是

$$x^2 + y^2 + (z - d)^2 = r^2,$$

所以一點同時位於球面和平面上，當且僅當 $z = 0$ 並且

$$x^2 + y^2 = r^2 - d^2.$$

- 若 $d < r$，這是平面上以 N 為圓心、半徑為 $\sqrt{r^2 - d^2}$ 的圓。
- 若 $d = r$，則 $x^2 + y^2 = 0$ 迫使 $x = y = 0$，所以交集只有 N 一點（該平面是切平面（tangent plane））。
- 若 $d > r$，右方為負數，所以沒有任何點符合條件，交集是空集。

這幾種情況涵蓋了所有可能，第一個命題因而得證。

兩個球面。設兩個球面的半徑為 $r, s > 0$，球心相異，相距 $c > 0$。把第一個球心放在原點，第二個放在 $(0, 0, c)$：

$$x^2 + y^2 + z^2 = r^2, \qquad x^2 + y^2 + (z - c)^2 = s^2.$$

? **想一想** 兩個二次方程的解集，怎麼可能只是一個圓？

把第一個方程減去第二個方程。所有平方項都會抵消，剩下

$$2cz - c^2 = r^2 - s^2 \quad \Longleftrightarrow \quad z = h, \qquad \text{其中} h = \frac{r^2 - s^2 + c^2}{2c}.$$

由於兩個球心相異，所以可以除以 c。一點同時位於兩個球面上，當且僅當它位於第一個球面上並且滿足相減所得的方程，所以

$$(\text{球面 1}) \cap (\text{球面 2}) = (\text{球面 1}) \cap (\text{平面} z = h).$$

根據第一個結果，這是一個圓、一點或空集。（明確地說，當 $|h| < r$ 時，它是平面 $z = h$ 上的圓 $x^2 + y^2 = r^2 - h^2$；當 $|h| = r$ 時是單獨一點；當 $|h| > r$ 時是空集。）這個圓所在的平面垂直於連心線（line of centres），而圓心就在這條線上。

第 (ii) 部分

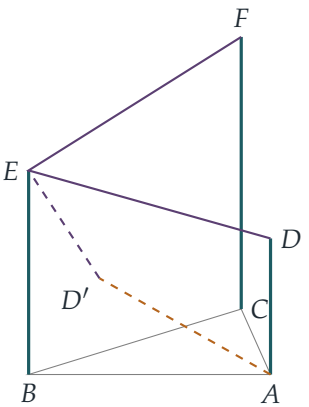
關於底部位置的假設。題目沒有說明 A、B 和 C 的位置。我們假設它們位於水平地面上一個邊長為 2 m 的等邊三角形（equilateral triangle）的頂點。這個佈局能得出題目所述的角（見「驗算」）。

取 B 為原點，z 軸為鉛垂方向：

$$B = (0, 0, 0), \quad A = (2, 0, 0), \quad C = (1, \sqrt{3}, 0),$$

因此各桿的頂端是

$$D = (2, 0, 2), \quad E = (0, 0, 3), \quad F = (1, \sqrt{3}, 4).$$



圖中的 D'（虛線）是破壞者移動桿子之後 D 的位置。

原來的雕塑。使用由 E 出發的向量（vector），

$$\overrightarrow{ED} = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \qquad \overrightarrow{EF} = \begin{pmatrix} 1 \\ \sqrt{3} \\ 1 \end{pmatrix},$$

所以 $\overrightarrow{ED} \cdot \overrightarrow{EF} = 2 + 0 - 1 = 1$，而 $|\overrightarrow{ED}| = |\overrightarrow{EF}| = \sqrt{5}$。因此

$$\cos \angle DEF = \frac{1}{\sqrt{5}\sqrt{5}} = \frac{1}{5}, \qquad \text{所以} \qquad \angle DEF = \cos^{-1} \tfrac{1}{5},$$

正如所求。

破壞者動手之後。桿 AD 的長度仍然是 2，底端仍然在 A。它現在位於通過 A 和 B 的鉛垂平面（即平面 $y = 0$）之內，向 B 傾斜，與地面成 $\frac{1}{4}\pi$ 角。由 A 走向 B 就是 $-x$ 的方向，所以

$$D' = \left(2 - 2\cos \tfrac{\pi}{4},\, 0,\, 2\sin \tfrac{\pi}{4}\right) = \left(2 - \sqrt{2},\, 0,\, \sqrt{2}\right).$$

由於 $2 - \sqrt{2} \approx 0.59$ 介乎 0 與 2 之間，點 D' 確實在線段 AB 的上方。現在

$$\overrightarrow{ED'} = \begin{pmatrix} 2 - \sqrt{2} \\ 0 \\ \sqrt{2} - 3 \end{pmatrix}, \qquad \overrightarrow{ED'} \cdot \overrightarrow{EF} = (2 - \sqrt{2}) + (\sqrt{2} - 3) = -1,$$

並且

$$|\overrightarrow{ED'}|^2 = (2 - \sqrt{2})^2 + (3 - \sqrt{2})^2 = (6 - 4\sqrt{2}) + (11 - 6\sqrt{2}) = 17 - 10\sqrt{2}.$$

? **想一想** 點積（dot product）是負數。這告訴我們這個角有甚麼特點？

純量積為負，表示這個角是鈍角（obtuse angle），所以我們需要的是介乎 $\frac{1}{2}\pi$ 與 π 之間的 $\cos^{-1}$ 值：

$$\cos \angle D'EF = \frac{-1}{\sqrt{5}\sqrt{17 - 10\sqrt{2}}} = \frac{-1}{\sqrt{85 - 50\sqrt{2}}} \approx -0.264542,$$

$$\angle D'EF \approx 1.839 \text{ 弧度（準確至四位有效數字）}.$$

驗算

- 若底部是邊長為 s 的等邊三角形，用同樣的方法可得 $\cos \angle DEF = \dfrac{s^2/2 - 1}{s^2 + 1}$。只有當 $s = 2$ 時，這個值才等於 $\frac{1}{5}$，這支持了我們對底部所作的假設。數值驗算確認當 $s = 2$ 時 $\cos \angle DEF = 0.2$。

- 已用數值方法驗算 $|AD'| = 2$ 和 $\angle BAD' = \frac{1}{4}\pi$。新的角算得為 1.838525… 弧度（約 105.3°），與上面的精確形式吻合。
- 第 (i) 部分：以隨機選取的球面作測試，並在預測的圓上取點；這些點都在兩個曲面上，誤差在 10^{-12} 以內。

常見錯誤

- 在兩個球面的情況中，聲稱兩個方程之差就是交集。該平面只是包含了交集；我們仍然要求出這個平面與其中一個球面的交集。
- 把 D' 放在 A 的另一側，即遠離 B 的一方。該點並不在線段 AB 的上方，而且會得出不同的角。
- 以度數作答，或者因為漏掉負號而把答案寫成銳角 $\cos^{-1}(0.2645)$。

STEP 之外

把兩個球面的方程相減而得到一個平面，正是兩個圓或兩個球面的*根平面*（radical plane；在二維的情況則是根軸（radical axis））背後的想法。同樣的技巧可以把 GPS 式的三邊測量（trilateration）化為線性問題。更一般地說，任何二次曲面（quadric surface）與平面的交集都是圓錐曲線（conic），而為這類交集分類，是利用二次型（quadratic form）學習線性代數（linear algebra）的一個很好的入門練習。Strang 的 *Introduction to Linear Algebra*（線性代數入門）建立了所需的工具，而 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題）則有更多 STEP 風格的坐標幾何題目。

詳細解答

STEP I 1993，第 9 題

純數學 • 一階微分方程

> **STEP I 1993，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/93-s1-q9

關鍵思路

這是一道混合問題（mixing problem）：缸內 A 的份量的變化率等於（流入率）減（流出率），而由於混合得很快，流出的液體與缸內液體的濃度相同。首先要留意的是，每個缸的總體積始終不變，所以濃度不過是「A 的份量除以一個常數」。於是每個缸各給出一條一階線性方程（first-order linear equation），而第一個缸的輸出就成為第二個缸的強迫項（forcing term）。

解答

第一個缸

液體以總流率每秒 $a + (b - a) = b$ 升流入第一個缸，並以每秒 b 升流出，所以對所有 $t \geq 0$，缸內體積一直保持在初始值 K 升。

由於混合得很快，在時間 t 缸內液體是均勻的，其中化學品 A 所佔的比例為 $V(t)/K$。在一段短時間 δt 內，有 $a\,\delta t$ 升的 A 流入，並有 $b\,\delta t$ 升的混合液流出，其中 $\frac{V}{K}\,b\,\delta t$ 升是 A。因此

$$\delta V \approx a\,\delta t - \frac{bV}{K}\,\delta t,$$

兩邊除以 δt，再令 $\delta t \to 0$，得

$$\dot{V}(t) = -\frac{bV(t)}{K} + a,$$

這正是所求的方程。

這是一條常係數線性方程。把它寫成 $\dot{V} + \frac{b}{K}V = a$，再乘以積分因子（integrating factor）

$\mathrm{e}^{bt/K}$，

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(V\mathrm{e}^{bt/K}\right) = a\,\mathrm{e}^{bt/K} \quad\Longrightarrow\quad V\mathrm{e}^{bt/K} = \frac{aK}{b}\mathrm{e}^{bt/K} + C.$$

最初缸內只有化學品 B，所以 $V(0) = 0$，由此得 $C = -\frac{aK}{b}$。因此

$$V(t) = \frac{aK}{b}\left(1 - \mathrm{e}^{-bt/K}\right), \qquad t \geq 0,$$

正如所求。

第二個缸

第二個缸每秒從第一個缸接收 b 升液體，並以每秒 b 升排出，所以它的體積保持在 L 升。設 $W(t)$ 為時間 t 時缸內 A 的體積，則 $W(0) = 0$。

> **想一想** 化學品 A 以甚麼速率流入第二個缸？

它流入的速率恰好就是它離開第一個缸的速率，即 $\frac{bV}{K} = a\left(1 - \mathrm{e}^{-bt/K}\right)$。根據同樣的快速混合論證，它以速率 $\frac{bW}{L}$ 流出。所以

$$\dot{W} + \frac{b}{L}W = a - a\,\mathrm{e}^{-bt/K}.$$

乘以積分因子 $\mathrm{e}^{bt/L}$，

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(W\mathrm{e}^{bt/L}\right) = a\,\mathrm{e}^{bt/L} - a\,\mathrm{e}^{bt(1/L - 1/K)}.$$

因為 $L \neq K$，指數 $b\left(\frac{1}{L} - \frac{1}{K}\right) = \frac{b(K-L)}{KL}$ 不等於零，所以第二項可以當作指數函數來積分：

$$W\mathrm{e}^{bt/L} = \frac{aL}{b}\mathrm{e}^{bt/L} - \frac{aKL}{b(K-L)}\mathrm{e}^{b(K-L)t/(KL)} + D.$$

全式乘以 $\mathrm{e}^{-bt/L}$，並注意到 $\frac{b(K-L)}{KL} - \frac{b}{L} = -\frac{b}{K}$，

$$W = \frac{aL}{b} - \frac{aKL}{b(K-L)}\mathrm{e}^{-bt/K} + D\mathrm{e}^{-bt/L}.$$

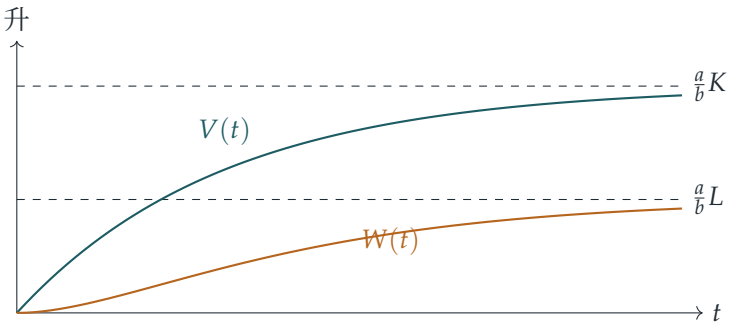
代入 $W(0) = 0$：

$$D = \frac{aKL}{b(K-L)} - \frac{aL}{b} = \frac{aL}{b}\cdot\frac{K - (K-L)}{K-L} = \frac{aL^2}{b(K-L)}.$$

所以在時間 $t \geq 0$，第二個缸內有

$$\boxed{W(t) = \frac{aL}{b}\left(1 - \frac{K\mathrm{e}^{-bt/K} - L\mathrm{e}^{-bt/L}}{K-L}\right)}$$

升化學品 A。



草圖取 $a=1$、$b=2$、$K=10$、$L=5$。留意 W 起初的斜率為零：一開始從第一個缸流來的液體中幾乎沒有 A。

驗算

- 在 $t=0$ 時，括號內的值為 $1-\frac{K-L}{K-L}=0$，理應如此。另外 $\dot{W}(0)=0$，與 A 以速率 $bV(0)/K=0$ 流入這一事實吻合。
- 當 $t\to\infty$ 時，$V\to\frac{a}{b}K$ 而 $W\to\frac{a}{b}L$：兩個缸內 A 的濃度最終都是 a/b，即流入液體的濃度。
- 除了前面的因子 L 之外，公式對 K 和 L 是對稱的，這正好可以用來檢查代數運算是否前後一致。
- 取 $a=1$、$b=2$、$K=10$、$L=5$、$t=5$：$V=5(1-\mathrm{e}^{-1})\approx 3.1606$，而 $W=2.5(1-\mathrm{e}^{-1})^2\approx 0.99894$。對兩個缸進行數值模擬（同時追蹤體積），在多組隨機選取的 a,b,K,L 之下，結果與兩條公式相符，誤差約為 10^{-13}。

常見錯誤

- 把 A 的流出率當作 $b-a$ 或 a：流出率應是總流出率乘以濃度。
- 把 A 流入第二個缸的速率當作 a：只有離開第一個缸的 A 才會到達第二個缸。
- 忘記為何需要 $L\neq K$：若 $L=K$，強迫項 $\mathrm{e}^{-bt/K}$ 是齊次方程（homogeneous equation）的解，答案會改為含有 $t\,\mathrm{e}^{-bt/K}$ 一項。

STEP 之外

一連串充分混合的缸是標準的房室模型（compartment model），可用來描述藥物劑量在體內的流動，以及放射性衰變鏈。把它寫成方程組 $\dot{\mathbf{x}}=M\mathbf{x}+\mathbf{f}$ 後，衰減率 b/K 和 b/L

就是 M 的特徵值（eigenvalue），而 $L = K$ 的情況就是重特徵值，也就是產生 $t\,\mathrm{e}^{\lambda t}$ 形式的解的那種「共振」（resonance）。Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科書）正是以這種方式處理混合問題和線性方程組。

詳細解答

STEP I 1993，第 10 題

力學 • 涉及彈簧的平衡

STEP I 1993，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：

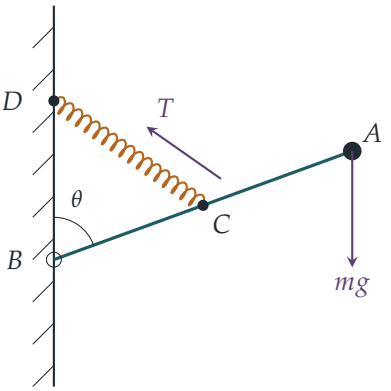
step.maths.org/questions/93-s1-q10

關鍵思路

對鉸鏈 B 取力矩（moments）：這樣可以完全消去 B 處未知的反作用力，只留下一條把彈簧張力與桿的角度聯繫起來的方程。這裏的幾何其實比表面看來容易處理，因為 $BD = BC = a$，所以三角形 BCD 是等腰三角形，當中每一個長度都可以用 B 處角度的一半來表示。

解答

我們把燈看成質點（particle），把桿看成輕而剛硬，把 B 處的鉸鏈（hinge）看成光滑，並把彈簧看成輕而且遵從胡克定律（Hooke's law）。設 $\theta = \angle DBA$，即桿與向上的鉛垂線 BD 之間的夾角。「離開牆壁」的意思是 $0 < \theta < \pi$：當 $\theta = 0$ 或 $\theta = \pi$ 時，桿貼着牆壁。



彈簧的長度

由於 $BC = BD = a$，而兩者之間的夾角是 θ，三角形 BCD 是等腰三角形。由 B 向 CD 作垂線，這條垂線既平分角 θ，亦平分邊 CD，所以

$$CD = 2a\sin\tfrac{\theta}{2}, \qquad \text{由 } B \text{ 到直線 } CD \text{ 的距離} = a\cos\tfrac{\theta}{2}.$$

根據胡克定律，彈簧的張力（tension）為

$$T = \frac{\lambda}{a}\left(2a\sin\tfrac{\theta}{2} - a\right) = \lambda\left(2\sin\tfrac{\theta}{2} - 1\right),$$

其中 T 若為負數，即表示彈簧受壓縮並向外推。

對 B 取力矩

作用於「桿連燈」這個系統的力有：在 A 的重量 mg、在 C 沿 CD 方向的彈簧力 T，以及在 B 的反作用力（reaction），而後者對 B 沒有力矩。重量沿鉛垂方向作用，與牆壁的水平距離為 $2a\sin\theta$，它使桿向下轉動（令 θ 增大）。張力則把 C 拉向 D，使桿轉回牆壁的方向。要達至平衡（equilibrium），須有

$$T \cdot a\cos\tfrac{\theta}{2} = mg \cdot 2a\sin\theta = 4mga\sin\tfrac{\theta}{2}\cos\tfrac{\theta}{2}.$$

當 $0 < \theta < \pi$ 時，我們有 $\cos\frac{\theta}{2} > 0$，所以可以把它約去：

$$T = 4mg\sin\tfrac{\theta}{2}.$$

特別地，$T > 0$：彈簧必定被拉長，這合乎常理，因為總要有東西把燈撐起來。

> **想一想** 現在我們有兩條關於 T 的表達式。把兩者相等會得到甚麼？所得的答案必須符合甚麼條件，才能描述桿的一個真實位置？

記 $s = \sin\frac{\theta}{2}$。把 T 的兩條表達式相等，

$$\lambda(2s-1) = 4mgs \quad\Longrightarrow\quad s(2\lambda - 4mg) = \lambda \quad\Longrightarrow\quad s = \frac{\lambda}{2(\lambda - 2mg)},$$

前提是 $\lambda \neq 2mg$（若 $\lambda = 2mg$，方程便變成 $0 = \lambda$，這是不可能的）。由於 $0 < \theta < \pi$，真實的位置需要滿足 $0 < s < 1$。

- 若 $\lambda < 2mg$，則 $s < 0$：無解。
- 若 $\lambda > 2mg$，則 $s > 0$，而且

$$s < 1 \iff \lambda < 2\lambda - 4mg \iff \lambda > 4mg.$$

所以當 $\lambda > 4mg$ 時，數值 $s = \lambda/\big(2(\lambda - 2mg)\big)$ 嚴格介乎 0 與 1 之間，從而在 $\theta \in (0, \pi)$ 內恰好有一個角度使力矩平衡。（事實上 $s > \frac{1}{2}$，所以彈簧確實是被拉長的。）這時合力由 B 處的反作用力承受，所以燈可以離開牆壁懸掛而處於平衡，正如題目所要求。反過來說，若 $\lambda \le 4mg$，則沒有這樣的位置。

該角度為

$$\angle DBA = \theta = 2\arcsin\left(\frac{\lambda}{2(\lambda - 2mg)}\right),$$

或者等價地，由於 $\cos\theta = 1 - 2s^2$，

$$\cos\angle DBA = 1 - \frac{\lambda^2}{2(\lambda - 2mg)^2}.$$

驗算

- 當 $\lambda \to \infty$ 時，彈簧無法伸長，所以 $CD \to a$，三角形 BCD 變成等邊三角形。公式給出 $s \to \frac{1}{2}$，所以 $\theta \to 60°$。✓
- 當 $\lambda \to 4mg$ 由上方趨近時，$s \to 1$ 而 $\theta \to 180°$：彈簧太弱，燈一直下沉，最後貼着牆壁懸掛。
- 取 $\lambda = 6mg$：$s = \frac{3}{4}$，$\cos\theta = -\frac{1}{8}$，所以 $\theta \approx 97.2°$。把這個角度代入完整的向量力矩方程（彈簧力由坐標計算），所得淨力矩為零。在 $0 < \theta < \pi$ 的範圍內對淨力矩作數值掃描，當 $\lambda > 4mg$ 時只找到一個根（例如 $\lambda = 5mg$ 時為 112.9°），而當 $\lambda \le 4mg$ 時則沒有根。

常見錯誤

- 以整條桿的長度 $2a$，而不是 a，來計算由 B 到彈簧力作用線的距離。垂直距離是 $a\cos\frac{\theta}{2}$。
- 分解力而不取力矩：這樣會引入 B 處反作用力的兩個未知分量，卻毫無得益。
- 忘記檢查 $0 < \sin\frac{\theta}{2} < 1$，而條件 $\lambda > 4mg$ 正正是由此而來。

STEP 之外

這道題只問平衡是否存在。順理成章的下一個問題是：這個平衡是否穩定？輕輕一碰之後，擾動會逐漸消失還是愈來愈大？要回答這個問題，可以寫出勢能（potential energy）$V(\theta)$（重力勢能加彈性勢能），再考察 $V''(\theta)$ 在平衡位置的正負號；這個想法貫穿大學一年級的力學，其後更貫穿整個動力系統（dynamical systems）理論。Kleppner 與

Kolenkow 所著的 *An Introduction to Mechanics*（力學入門）對能量方法和穩定性有仔細的論述。

詳細解答

STEP I 1993，第 11 題

力學 • 質心

> **STEP I 1993，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/93-s1-q11

關鍵思路

一個剛體（rigid body）從一點自由懸掛時，只受兩個力作用：支承點的拉力和它本身的重量。兩個力要互相平衡，必須作用在同一直線上，所以質心（centre of mass）必定懸在支承點的正下方。因此，整個問題歸結為求出彎曲鐵線的質心 G，然後留意到：在兩個位置之間，物體轉過的角度正是直線 GA 與 GB 之間的夾角。

解答

質心

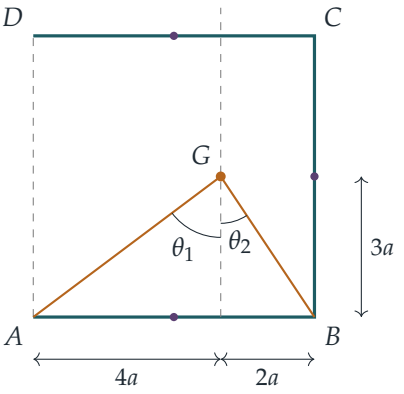
把鐵線看成一根均勻的剛性幼棒，彎成 AB、BC、CD 三條邊，每條邊的重量同為 W，並掛在一口光滑的釘上或一條輕繩上。為了使數字都是整數，取正方形的邊長為 $6a$，並在鐵線所在的平面上建立坐標軸，使得

$$A = (0,0), \quad B = (6a,0), \quad C = (6a,6a), \quad D = (0,6a).$$

每條邊都是均勻的幼棒，所以它的重量作用在中點：AB 的是 $(3a,0)$，BC 的是 $(6a,3a)$，CD 的是 $(3a,6a)$。三個重量相等，所以 G 就是三個中點的平均：

$$G = \left(\frac{3a+6a+3a}{3}, \frac{0+3a+6a}{3}\right) = (4a,\ 3a).$$

（鐵線對稱於通過 BC 和 AD 兩個中點的直線，因此 G 必定位於高度 $3a$ 處；需要計算的只有它的水平位置。）



圖中通過 G 的垂直虛線與 BC 平行（亦與缺少了的邊 AD 平行）。設 θ_1 為 GA 與這條虛線所成的角，θ_2 為 GB 在另一側與它所成的角。從圖中可以看出，

$$\tan\theta_1 = \frac{4a}{3a} = \frac{4}{3}, \qquad \tan\theta_2 = \frac{2a}{3a} = \frac{2}{3}.$$

（事實上 $AG = 5a$：這是一個 3–4–5 三角形。）

先從 A 懸掛，再從 B 懸掛

當鐵線從 A 懸掛並處於平衡（equilibrium）時，作用在它上面的力只有 A 處的反作用力（reaction），以及作用在 G 的總重量 $3W$。要使兩者平衡，它們對 A 的力矩（moment）必須互相抵消，所以重量的作用線必須通過 A：直線 AG 是垂直的，而 G 在 A 的下方。由於 BC 與圖中的虛線平行，此時 BC 與垂直方向成 θ_1 的傾角。

同樣地，當鐵線從 B 懸掛時，直線 BG 是垂直的，而 BC 則向另一邊傾斜，與垂直方向成 θ_2 的傾角。

? **想一想** 為甚麼 BC 的兩個方向之間的夾角是 $\theta_1 + \theta_2$，而不是 $\theta_1 - \theta_2$？

從第一個位置轉到第二個位置時，鐵線是整體剛性地轉動的，所以它上面的每一條直線都轉過相同的角度。換一個說法：讓鐵線在紙上固定不動，改為讓「向下」的方向轉動。在第一個位置，向下的方向沿着 GA；在第二個位置，則沿着 GB。因此轉過的角度就是兩者之間的夾角 AGB。在圖中，A 和 B 分別位於通過 G 的虛線的兩側，所以

$$\angle AGB = \theta_1 + \theta_2.$$

利用複角公式（addition formula），

$$\tan(\theta_1 + \theta_2) = \frac{\tan\theta_1 + \tan\theta_2}{1 - \tan\theta_1 \tan\theta_2} = \frac{\frac{4}{3} + \frac{2}{3}}{1 - \frac{8}{9}} = \frac{2}{\frac{1}{9}} = 18.$$

θ_1 和 θ_2 都是銳角，所以 $0 < \theta_1 + \theta_2 < 180°$；又因為它的正切是正數，這個和其實是銳角。因此 BC 的兩個方向之間的夾角是 $\tan^{-1} 18$，正如題目所求。

驗算

- 以數值計算，$\theta_1 = \tan^{-1}\frac{4}{3} \approx 53.13^\circ$，$\theta_2 = \tan^{-1}\frac{2}{3} \approx 33.69^\circ$，兩者相加得 86.82°，而 $\tan^{-1} 18 \approx 86.82^\circ$ 亦然。
- 把鐵線換成 90 000 小段相等的部分，把它們的位置取平均而得 $G = (4a, 3a)$，然後每次把整個物體旋轉，使 G 位於支承點下方，所得 BC 的兩個像之間的夾角與上述結果相同，準確至小數點後九位。
- 檢查 G 是否合理：它較接近 BC，而離缺少了的邊 AD 較遠，這是理所當然的，因為右邊有一整條邊的鐵線，左邊卻完全沒有。

常見錯誤

- 把 G 取在正方形的中心 $(3a, 3a)$。如果四條邊齊全，這是對的；但邊 AD 並不存在，沒有東西把質心拉向左邊。
- 把兩個傾角相減而不是相加：從 A 懸掛時 BC 向一邊傾斜，從 B 懸掛時則向另一邊傾斜。

STEP 之外

「自由懸掛的物體靜止時，質心位於支承點的下方」這個原理，正是實際上找出不規則平面圖形質心的方法：把它先後從兩點懸掛，看看兩條鉛垂線在哪裏相交。在大學一年級的課程中，同一個想法會以勢能（potential energy）的形式出現：平衡位置就是 G 的高度的駐點（stationary point），而穩定的平衡位置就是最低的那一個。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）討論質心和剛體靜力學（statics），當中還有更多類似的例子。

詳細解答

STEP I 1993，第 12 題

力學 • 拋體的攔截

> **STEP I 1993，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/93-s1-q12

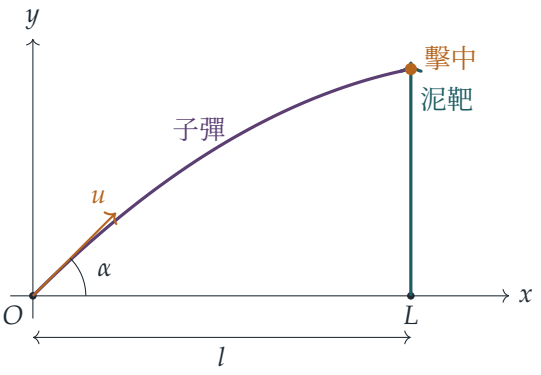
關鍵思路

擊中的意思是子彈與泥靶在同一時刻位於同一位置。子彈到達通過發射器的鉛垂線的時刻，只由它的水平運動決定，所以我們先求出這個時刻，再令兩者的高度相等。在子彈飛行期間，重力把兩個物體向下拉的距離完全相同，都是 $\frac{1}{2}gt^2$，因此這一項互相抵消，剩下一個簡潔的線性條件。當 $T = 0$ 時，這個抵消更讓我們可以用一個巧妙的角度看通整道題。

解答

模型

我們把泥靶和子彈都看成只在重力作用下自由運動的質點 (particle)，不計空氣阻力 (air resistance)，並假設參賽者在包含發射器的鉛垂平面內朝發射器的方向射擊。把參賽者放在原點 O，x 軸水平指向位於 $x = l$ 的發射器 L，y 軸鉛垂向上。時間 t 由步槍發射的一刻起計，所以泥靶此時已在空中飛行了時間 T，而在時刻 t 它已飛行了 $T + t$。



擊中的條件

發射後的時刻 t，子彈位於

$$x_B = (u\cos\alpha)\,t, \qquad y_B = (u\sin\alpha)\,t - \tfrac{1}{2}gt^2,$$

而泥靶在時間 T 之前以速率 v 向上射出，此時位於 $x = l$，高度為

$$y_C = v(T+t) - \tfrac{1}{2}g(T+t)^2 = \left(vT - \tfrac{1}{2}gT^2\right) + (v - gT)\,t - \tfrac{1}{2}gt^2.$$

子彈只可能在直線 $x = l$ 上與泥靶相遇，而它到達該處時 $(u\cos\alpha)\,t = l$，即

$$t = \frac{l}{u\cos\alpha}.$$

（這裏 $\cos\alpha > 0$：由於 $l > 0$，槍口垂直向上的步槍永遠打不到發射器那邊。）要擊中，就需要在這一刻有 $y_B = y_C$。兩邊的 $-\frac{1}{2}gt^2$ 項互相抵消，所以

$$(u\sin\alpha)\,t = \left(vT - \tfrac{1}{2}gT^2\right) + (v - gT)\,t.$$

代入 $t = l/(u\cos\alpha)$ 得

$$\frac{l\sin\alpha}{\cos\alpha} = \left(vT - \tfrac{1}{2}gT^2\right) + \frac{(v - gT)\,l}{u\cos\alpha},$$

再把全式乘以 $\cos\alpha$，

$$l\sin\alpha - \left(vT - \tfrac{1}{2}gT^2\right)\cos\alpha = \frac{v - gT}{u}\,l,$$

這就是所要求的結果。

$T = 0$ 的情況

當 $T = 0$ 時，條件變成 $l\sin\alpha = vl/u$，而由於 $l > 0$，

$$\sin\alpha = \frac{v}{u}.$$

想一想 為甚麼這必然導致 $v < u$，而不只是 $v \le u$？

我們需要 $\sin\alpha \le 1$，所以 $v \le u$。但 $v = u$ 就需要 $\alpha = \frac{\pi}{2}$，這時 $\cos\alpha = 0$：子彈從 O 垂直向上飛，永遠到不了直線 $x = l$。所以只有當 $v < u$ 時才可能擊中。這時

$$\cos\alpha = \sqrt{1 - \frac{v^2}{u^2}} = \frac{\sqrt{u^2 - v^2}}{u},$$

因為 $0 < \alpha < \frac{\pi}{2}$，故取正根，而擊中發生於

$$t = \frac{l}{u\cos\alpha} = \frac{l}{\sqrt{u^2 - v^2}}.$$

泥靶由 $t = 0$ 起在空中，直至 $vt - \frac{1}{2}gt^2 = 0$ 再次成立為止，即直至 $t = 2v/g$。因此，要在它落地之前擊中它，就需要

$$\frac{l}{\sqrt{u^2 - v^2}} < \frac{2v}{g} \quad\Longleftrightarrow\quad \frac{2v\sqrt{u^2 - v^2}}{g} > l,$$

其中我們乘上了正數 $\sqrt{u^2 - v^2}$。連同 $v < u$，這正是我們要證明的結論。

$T = 0$ 的情況有一個很漂亮的看法。當 $\sin\alpha = v/u$ 時，子彈速度的鉛垂分量是 $u\sin\alpha = v$，與泥靶的相同，而兩者所受的重力也一樣。所以子彈和泥靶*在每一瞬間都處於同一高度*，唯一的問題是：子彈的水平速率 $\sqrt{u^2 - v^2}$ 能否在泥靶落下之前走完距離 l。

驗算

- *無重力的極限情況。*若 $g = 0$ 且 $T = 0$，兩個物體都沿直線運動，而條件 $\sin\alpha = v/u$ 正是勻速攔截時常見的「提前量」瞄準條件。
- *最大射程。*對固定的 u，上界 $2v\sqrt{u^2 - v^2}/g$ 在 $v = u/\sqrt{2}$ 時最大，其值為 u^2/g，即以速率 u 發射的拋體（projectile）的最大射程。這是合理的：當 $v = u/\sqrt{2}$ 時我們得到 $\alpha = 45°$，而擊中恰好發生在子彈本應落回地面的一刻。
- *數值。*取 $g = 9.8$、$v = 20$、$T = 0.7$、$u = 120$ 及 $l = 60$（國際單位制），用數值方法解出擊中所需的角度，得 $\alpha \approx 0.2987$ rad，而第一條恆等式的兩邊相符。把兩個運動按時間逐步推算，子彈到達發射器時，子彈和泥靶的高度都是 17.13 m。對於 $T = 0$、$v = 20$ 及 $u = 100$，上界約為 400 m，而 $\alpha \approx 11.5°$。

常見錯誤

- 對兩個物體使用同一個時間變量，卻沒有考慮延遲 T：泥靶的高度必須用 $T + t$ 計算。

- 忘記排除 $v = u$ 的情況，這時 $\alpha = 90°$，根本不可能擊中。
- 只檢查子彈能否到達發射器，卻沒有檢查那一刻泥靶是否仍在空中。

STEP 之外

$\frac{1}{2}gt^2$ 的抵消，其實是一個深刻思想的特例：在以加速度 g 自由下落的參考系（frame）中，重力消失了，每個拋體都以恆定速度沿直線運動。這是把伽利略相對性（Galilean relativity）應用於加速參考系，而到了愛因斯坦（Einstein）手中，它成為廣義相對論（general relativity）核心的等效原理（equivalence principle）。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學導論）對相對運動和非慣性參考系（non-inertial frames）有細緻的闡述。

詳細解答

STEP I 1993，第 13 題

力學 • 功與功率

STEP I 1993，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：

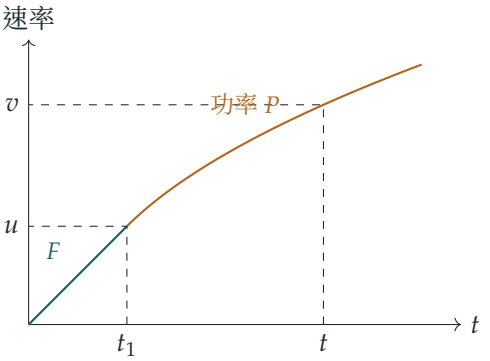
step.maths.org/questions/93-s1-q13

關鍵思路

運動分為兩個階段，每個階段要用牛頓第二定律（Newton's second law）的不同形式。當力恆定時，加速度也恆定，所以 *suvat* 方程已足以解決一切。引擎一旦改以恆定功率（constant power）運作，驅動力便是 P/v，它會隨速率而改變；訣竅是把加速度寫成 $v\frac{\mathrm{d}v}{\mathrm{d}t}$（用來求時間）或 $v\frac{\mathrm{d}v}{\mathrm{d}x}$（用來求距離），使變數得以分離。

解答

把機車和列車合起來看成一個質量為 M 的質點（particle），沿筆直的水平路軌運動，且不受阻力。水平方向唯一的力是牽引力（tractive force）（重力和法向反作用力（normal reaction）都是垂直方向的，互相抵消）。下圖是速率對時間的草圖：先是一條直線，升至速率 u，然後是一條曲線，其斜率隨速率上升而下降。



第一階段：恆力

力恆為 F 時，加速度是 F/M。由靜止出發，速率達到 u 的時刻是

$$t_1 = \frac{Mu}{F} = \frac{Mu^2}{P},$$

這裏用了 $F = P/u$。由 $v^2 = u^2 + 2as$（這裏速率由 0 增至 u），所走的距離是

$$s_1 = \frac{u^2}{2F/M} = \frac{Mu^2}{2F} = \frac{Mu^3}{2P}.$$

第二階段：恆定功率

功率等於力乘以速度，所以速率為 w 時，牽引力是 P/w。留意當 $w = u$ 時，這個力是 $P/u = F$，所以引擎轉換模式時，力不會突然跳變。由牛頓第二定律得

$$M\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{P}{w}.$$

時間。 分離變數（separating the variables），得 $Mw\,\mathrm{d}w = P\,\mathrm{d}t$。由速率 u、時刻 t_1 積分至速率 v、時刻 t，

$$\frac{M}{2}\left(v^2 - u^2\right) = P\,(t - t_1).$$

（這其實就是功能原理（work–energy principle）：引擎所作的功是 $P(t - t_1)$，全部轉化為動能（kinetic energy））所以

$$t = \frac{Mu^2}{P} + \frac{M}{2P}\left(v^2 - u^2\right) = \frac{M}{2P}\left(2u^2 + v^2 - u^2\right) = \frac{M}{2P}\left(u^2 + v^2\right),$$

正是所求。由於 $v \geq u$，速率 v 是在第二階段之中（或剛好在其開始的一刻）達到的，所以我們可以使用第二階段的方程。

想一想 求距離時，自變數應該是 x，而不是 t。我們應怎樣改寫加速度？

由連鎖律（chain rule），$\dfrac{\mathrm{d}w}{\mathrm{d}t} = \dfrac{\mathrm{d}w}{\mathrm{d}x}\dfrac{\mathrm{d}x}{\mathrm{d}t} = w\dfrac{\mathrm{d}w}{\mathrm{d}x}$。於是運動方程變成

$$Mw\frac{\mathrm{d}w}{\mathrm{d}x} = \frac{P}{w}, \qquad \text{即} \qquad Mw^2\,\mathrm{d}w = P\,\mathrm{d}x.$$

距離。 由 $w = u$ 積分至 $w = v$，第二階段所走的距離 s_2 滿足

$$\frac{M}{3}\left(v^3 - u^3\right) = P\,s_2, \qquad s_2 = \frac{M}{3P}\left(v^3 - u^3\right).$$

因此總距離是

$$\begin{aligned} s_1 + s_2 &= \frac{Mu^3}{2P} + \frac{M}{3P}\left(v^3 - u^3\right) \\ &= \frac{M}{6P}\left(3u^3 + 2v^3 - 2u^3\right) \\ &= \frac{M}{6P}\left(2v^3 + u^3\right). \end{aligned}$$

試卷上的一處誤印。 試卷印出的答案把 u^3 寫成了 u^2。這不可能是對的：Mv^3/P 是長度（功率的單位是 $\mathrm{kg\,m^2\,s^{-3}}$，而 Mv^3 的單位是 $\mathrm{kg\,m^3\,s^{-3}}$），但 Mu^2/P 卻是時間，所以含 u 的一項也必須是三次方，兩項才可以相加。正確的距離是 $\dfrac{M}{6P}\left(2v^3 + u^3\right)$，也就是我們上面證明了的結果。

驗算

- 取 $v = u$：時間公式給出 $Mu^2/P = Mu/F$，距離公式給出 $Mu^3/(2P) = Mu^2/(2F)$，正好是第一階段的結果，理應如此。
- 單位：Mv^2/P 是能量除以功率，即時間；Mv^3/P 是長度。誤印正是這樣露出馬腳的。
- 取 $M = F = u = 1$（於是 $P = 1$）及 $v = 2$，公式給出 $t = 2.5$，距離為 $17/6 \approx 2.833$。以數值方法逐步模擬這個運動（速率低於某值時力為 F，該值就是 u，高於它時力為 P/w），所得的兩個數值都與上述相同；而且對隨機選取的 M、F、u 和 $v \geq u$，數值結果與公式的誤差小於 10^{-7}；試卷上印有 u^2 的版本則並不吻合。

常見錯誤

- 忘記第一階段，直接由靜止開始對恆定功率的方程積分。第二階段的方程只在 $w \geq u$ 時才成立；若由 $w = 0$ 開始使用，就等於說力是無限大。
- 對 x 積分時卻用了 $\mathrm{d}w/\mathrm{d}t$：求距離時，加速度必須寫成 $w\,\mathrm{d}w/\mathrm{d}x$。
- 把第一階段的時間寫成 u/F：加速度是 F/M，不是 F。

STEP 之外

恆定功率階段是一條一階微分方程（first-order differential equation），以分離變數法求解；而把自變數由 t 換成 x，正是一般能量方法所用的同一訣竅：$w\,\mathrm{d}w/\mathrm{d}x$ 積分後得出動能，由此引出功能定理（work–energy theorem），再往後就是位勢（potentials）和守

恆量（conserved quantities）。低速時力設有上限，也正是現實中牽引的運作方式，因為功率固定的摩打在靜止時不可能提供無限大的力。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學導論）正是循這個思路闡述功、能量和功率。

詳細解答

STEP I 1993，第 14 題

概率與統計 • 指數分佈的等候時間

STEP I 1993，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/93-s1-q14

關鍵思路

重要的不是蟾蜍先生的速率，而是他走完最後一公里所需的時間 T，而 $T = 1/V$。V 的密度函數看來不好處理，但它的累積分佈函數（cumulative distribution function）卻非常簡單；把它轉化為關於 T 的敘述，便可看出 T 服從參數為 α 的指數分佈（exponential distribution）。此後，每個綠燈時段的概率都是兩個指數函數之差，而把所有綠燈時段加起來，就得到一個等比級數（geometric series）。

解答

到達時間的分佈

設 V 為速率，單位為公里每分鐘，並設 $T = 1/V$ 為他到達交通燈的時間，以分鐘為單位，由交通燈轉綠的一刻起計。留意到

$$\frac{\mathrm{d}}{\mathrm{d}v}\,\mathrm{e}^{-\alpha/v} = \alpha v^{-2}\mathrm{e}^{-\alpha/v},$$

以及當 $v \to 0^+$ 時 $\mathrm{e}^{-\alpha/v} \to 0$，可知 V 的分佈函數為

$$\mathrm{P}(V \le v) = \int_0^v \alpha u^{-2}\mathrm{e}^{-\alpha/u}\,\mathrm{d}u = \mathrm{e}^{-\alpha/v} \qquad (v > 0).$$

（當 $v \to \infty$ 時它趨向 1，所以這確實是一個密度函數。）由於 $V > 0$，對 $t > 0$ 我們有 $T > t \iff 1/V > t \iff V < 1/t$，所以

$$\mathrm{P}(T > t) = \mathrm{P}\left(V < \tfrac{1}{t}\right) = \mathrm{e}^{-\alpha t}.$$

因此 T 服從參數為 α 的指數分佈。

? 想一想 為甚麼要用 T 而不用 V 來計算？

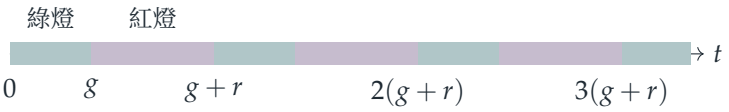
交通燈是以時間來描述的，所以我們關心的每個事件，都是 T 的取值所在的一個區間。若要直接把 V 的密度函數換算成 T 的密度函數，便須作變數代換（change of variable）；改為經由 $\mathrm{P}(T > t)$ 入手，則可避開這一步，一行便得出答案。

一個綠燈時段

在時間 0，交通燈轉為綠燈。綠燈時段為 $[0, g)$，紅燈時段為 $[g, g+r)$，然後綠燈時段為 $[g+r, (g+r)+g)$，如此類推。第 $(n+1)$ 個綠燈時段是 $n(g+r) \le t < n(g+r)+g$。由於 T 是連續的，是否包括端點並無分別，而

$$\begin{aligned}\mathrm{P}\big(n(g+r) < T < n(g+r)+g\big) &= \mathrm{P}\big(T > n(g+r)\big) - \mathrm{P}\big(T > n(g+r)+g\big)\\ &= \mathrm{e}^{-\alpha n(g+r)} - \mathrm{e}^{-\alpha(n(g+r)+g)},\end{aligned}$$

正如題目所求。（原試卷把第二個指數中的右括號印錯了位置；該指數應為 $-\alpha\big(n(g+r)+g\big)$，如上所示。）以上推算對 $n = 0$ 同樣成立，那就是第一個綠燈時段，此時交通燈還未曾轉過紅燈。



遇上綠燈的概率

他在綠燈時通過，當且僅當 T 落在其中一個綠燈時段之內。這些時段互不相交（disjoint），所以我們把它們的概率相加，包括 $n = 0$：

$$\begin{aligned}\mathrm{P}(\alpha) &= \sum_{n=0}^{\infty}\left(\mathrm{e}^{-\alpha n(g+r)} - \mathrm{e}^{-\alpha n(g+r)}\mathrm{e}^{-\alpha g}\right) = \left(1-\mathrm{e}^{-\alpha g}\right)\sum_{n=0}^{\infty}\left(\mathrm{e}^{-\alpha(g+r)}\right)^n\\ &= \frac{1-\mathrm{e}^{-\alpha g}}{1-\mathrm{e}^{-\alpha(g+r)}},\end{aligned}$$

這裏求了等比級數之和；該級數收斂（converges），因為 $0 < \mathrm{e}^{-\alpha(g+r)} < 1$（由於 $\alpha > 0$ 及 $g+r > 0$）。

極限

當 $\alpha \to \infty$ 時，$\mathrm{e}^{-\alpha g} \to 0$ 和 $\mathrm{e}^{-\alpha(g+r)} \to 0$ 同時成立（因為 $g > 0, g+r > 0$），所以

$$\mathrm{P}(\alpha) \to \frac{1-0}{1-0} = 1.$$

這是合理的：α 很大表示蟾蜍先生開得非常快，幾乎每次都能在第一個綠燈時段結束前通過。

當 $\alpha \to 0$ 時，分子和分母都趨向 0，所以要更小心處理。在分子和分母中分別寫成 $1 - \mathrm{e}^{-x} = \mathrm{e}^{-x}(\mathrm{e}^{x} - 1)$：

$$\mathrm{P}(\alpha) = \frac{\mathrm{e}^{-\alpha g}\,(\mathrm{e}^{\alpha g} - 1)}{\mathrm{e}^{-\alpha(g+r)}\,(\mathrm{e}^{\alpha(g+r)} - 1)} = \mathrm{e}^{\alpha r} \cdot \frac{\dfrac{\mathrm{e}^{\alpha g} - 1}{\alpha g}}{\dfrac{\mathrm{e}^{\alpha(g+r)} - 1}{\alpha(g+r)}} \cdot \frac{g}{g+r},$$

其中我們把分子和分母同除以 α，並補上因子 g 和 $g+r$。當 $\alpha \to 0$ 時，$\mathrm{e}^{\alpha r} \to 1$，而且 $\alpha g \to 0$ 和 $\alpha(g+r) \to 0$ 同時成立，所以由 $(\mathrm{e}^{x} - 1)/x \to 1$ 可知兩個分式各自趨向 1。因此

$$\mathrm{P}(\alpha) \to \frac{g}{r+g} \qquad \text{當} \alpha \to 0.$$

這時蟾蜍先生開得非常慢，他的到達時間分散在很多個週期之中，所以遇上綠燈的機會就只是每個週期中綠燈所佔的比例。

驗算

- 取 $\alpha = 0.7, g = 1, r = 2$，公式給出 $\mathrm{P}(\alpha) \approx 0.5737$。模擬 4×10^5 次車程，以求 $\mathrm{e}^{-\alpha/v}$ 的反函數的方法對 V 抽樣，結果為 0.5744，而 $n = 0, 1, 2$ 各個綠燈時段的概率亦一一吻合。
- 取 $g = 1, r = 2$：$\mathrm{P}(10^{-4}) \approx 0.33337$，接近 $\frac{1}{3}$，而 $\mathrm{P}(10) \approx 0.99995$。
- 若 $r = 0$，交通燈永遠是綠燈，而公式對每個 α 都給出 $\mathrm{P}(\alpha) = 1$，正該如此。

常見錯誤

- 漏掉 $n = 0$。第一個綠燈時段在任何紅燈之前，所佔的概率最大。
- 照抄試卷上印的指數 $-\alpha(n(g+r)) + g$，那是 $-\alpha\big(n(g+r) + g\big)$ 的誤印。
- 把 V 本身當作到達時間，或者嘗試求 $\mathrm{E}(T)$。事實上這裏的 $\mathrm{E}(V)$ 是無限大，但題目完全用不着它。

STEP 之外

透過分佈函數把 V 的分佈轉化為 $T = g(V)$ 的分佈，是變換隨機變量（random variable）的標準方法。這裏 V 服從逆指數分佈（inverse-exponential distribution）。極限 $\alpha \to 0$ 是

一個普遍想法的小例子：一個分散在許多個週期之中的隨機時間，落在各個階段的機會與該階段的長度成正比，這正是更新理論（renewal theory）的起點。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）對隨機變量的變換和指數分佈都有論述。

詳細解答

STEP I 1993，第 15 題

概率與統計 • 貝葉斯定理

STEP I 1993，第 15 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/93-s1-q15

關鍵思路

這是一道披着故事外衣的貝葉斯定理（Bayes' theorem）題目。竅門在於以眞正的方向為條件，然後問：在每種情況下，那位土著給出這一連串答案的可能性有多大。Biscii 族人永遠說謊，所以 Biscii 族人重複作答時，答案必定一致，而且必定指向錯誤的方向。因此，答案重複出現，會令你更有可能正在跟 Biscii 族人說話，這正是多問一次便可能推翻你原來決定的原因。

解答

只考慮 Ascii 族人

Ascii 族人每次回答問題時，說真話的概率是 $\frac{3}{4}$，且與之前的答案獨立（independent）。這條問題只有兩個可能的答案（左或右），所以兩次答案相同，當且僅當兩次都是真話或兩次都是假話：

$$\mathrm{P}(\text{兩次答案相同}) = \left(\tfrac{3}{4}\right)^2 + \left(\tfrac{1}{4}\right)^2 = \tfrac{9}{16} + \tfrac{1}{16} = \tfrac{10}{16} = \tfrac{5}{8},$$

正如所求。其中「兩次都是真話」的情況佔了 $\frac{9}{16}$，所以根據條件概率（conditional probability）的定義，

$$\mathrm{P}(\text{說真話} \mid \text{兩次答案相同}) = \frac{9/16}{10/16} = \frac{9}{10},$$

這就是第二個結果。

兩次回答「左」

設 L 為「左是正確方向」這個事件（event），R 為「右是正確方向」這個事件。設 A 和 B 分別為土著是 Ascii 族人和 Biscii 族人的事件，其中 $\mathrm{P}(A) = \frac{11}{16}$，$\mathrm{P}(B) = \frac{5}{16}$，而土著屬於哪一族與地理情況獨立。船長起初對方向一無所知，所以我們取 $\mathrm{P}(L) = \mathrm{P}(R) = \frac{1}{2}$。設 E_2 為土著兩次都回答「左」的事件。

> **想一想** 為甚麼求 $\mathrm{P}(E_2 \mid L)$ 比求 $\mathrm{P}(L \mid E_2)$ 容易？

一旦知道真正的方向，每一族的行為便完全確定，所以答案出現的概率很容易寫出來。然後貝葉斯定理便會把這些概率倒轉過來。

若左是正確的，Ascii 族人要兩次都說真話才會兩次回答「左」，概率是 $\frac{9}{16}$。Biscii 族人會說謊而回答「右」，所以不可能產生 E_2。因此

$$\mathrm{P}(E_2 \mid L) = \tfrac{11}{16} \cdot \tfrac{9}{16} + \tfrac{5}{16} \cdot 0 = \tfrac{99}{256}.$$

若右是正確的，Ascii 族人要兩次都說謊才會兩次回答「左」，概率是 $\frac{1}{16}$，而 Biscii 族人則必定兩次都回答「左」。因此

$$\mathrm{P}(E_2 \mid R) = \tfrac{11}{16} \cdot \tfrac{1}{16} + \tfrac{5}{16} \cdot 1 = \tfrac{11+80}{256} = \tfrac{91}{256}.$$

L 和 R 的先驗概率（prior）相等，所以在貝葉斯定理中互相抵消：

$$\mathrm{P}(L \mid E_2) = \frac{\frac{1}{2} \cdot \frac{99}{256}}{\frac{1}{2} \cdot \frac{99}{256} + \frac{1}{2} \cdot \frac{91}{256}} = \frac{99}{99+91} = \frac{99}{190}.$$

這個值大於 $\frac{1}{2}$，所以左是較有可能的方向，船長應該向左走，而他走對的概率是 $\frac{99}{190}$。

三次回答「左」

設 E_3 為三次都回答「左」的事件。同樣的推論仍然適用，只是現在 Ascii 族人給出三個獨立的答案：

$$\mathrm{P}(E_3 \mid L) = \tfrac{11}{16}\left(\tfrac{3}{4}\right)^3 + \tfrac{5}{16} \cdot 0 = \tfrac{11}{16} \cdot \tfrac{27}{64} = \tfrac{297}{1024},$$
$$\mathrm{P}(E_3 \mid R) = \tfrac{11}{16}\left(\tfrac{1}{4}\right)^3 + \tfrac{5}{16} \cdot 1 = \tfrac{11}{1024} + \tfrac{320}{1024} = \tfrac{331}{1024}.$$

由於 $\mathrm{P}(E_3 \mid R) > \mathrm{P}(E_3 \mid L)$，而且先驗概率相等，現在右才是較有可能的方向。確切地說，

$$\mathrm{P}(R \mid E_3) = \frac{331}{297+331} = \frac{331}{628} > \frac{1}{2}.$$

所以船長應該向右走，此時他走對的概率是 $\frac{331}{628}$。

想一想 第三個「左」怎麼會令左變得更不可能？

每多一個「左」，Ascii 族人對 $\mathrm{P}(E \mid L)$ 的貢獻便乘以 $\frac{3}{4}$。在 $\mathrm{P}(E \mid R)$ 中，Ascii 族人的一項會縮小至原來的 $\frac{1}{4}$，但 Biscii 族人的一項始終保持在 $\frac{5}{16}$。一長串完全相同的答案，正是 Biscii 族人會給出的，所以證據愈來愈傾向於說「這是個說謊者」。說真話的 Ascii 族人也會前後一致，但概率只有 $(\frac{3}{4})^n$，而這個值會逐漸趨向零。

驗算

- 只回答一次「左」之後，似然（likelihood）分別是 $\frac{11}{16} \cdot \frac{3}{4} = \frac{33}{64}$（已知 L）和 $\frac{11}{16} \cdot \frac{1}{4} + \frac{5}{16} = \frac{31}{64}$（已知 R），所以 $\mathrm{P}(L \mid \text{左}) = \frac{33}{64}$。數列 $\frac{33}{64} \approx 0.516$、$\frac{99}{190} \approx 0.521$、$\frac{297}{628} \approx 0.473$ 先升後跌，與上面的解釋吻合。
- 回答次數很多時，$\mathrm{P}(L \mid E_n) \to 0$，因為 Biscii 族人的一項佔主導。
- 我們以分數作精確窮舉，並模擬了 10^6 次對話，核對了全部四個概率：模擬所得的值是 0.5214（精確值 $\frac{99}{190} \approx 0.5211$）和 0.5266（精確值 $\frac{331}{628} \approx 0.5271$）。

常見錯誤

- 忘記說明 $\mathrm{P}(L) = \mathrm{P}(R) = \frac{1}{2}$ 這個假設。若對方向沒有任何先驗概率，這條問題便無法作答。
- 把第一部分的 $\frac{9}{10}$ 套用到整個族羣上。這個結果只適用於 Ascii 族人；Biscii 族人同樣會兩次給出相同的答案，而且每次都如此。
- 以為不論土著是哪一族人，他的三個答案都是獨立的。只有在知道他屬於哪一族（以及真正的方向）之後，這些答案才是獨立的，這就是我們要先取條件的原因。

STEP 之外

這個計算是貝葉斯更新（Bayesian updating）的一個例子：比值 $\mathrm{P}(E \mid R)/\mathrm{P}(E \mid L)$ 稱為似然比（likelihood ratio）或貝葉斯因子（Bayes factor），用來量度證據有多支持其中一個假設。由於土著屬於哪一族是所有答案共有的隱藏變量（hidden variable），這些答案只是條件獨立（conditionally independent），看似指向一方的證據，最終可能指向另一方。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門教材）對條件概率以及正是這一類的悖論有非常出色的論述。

詳細解答

STEP I 1993，第 16 題

概率與統計 • 幾何概率

> **STEP I 1993，第 16 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/93-s1-q16

本題在 *Beyond STEP* 第 12.20 節蒲豐投針中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch12/buffon.html

關鍵思路

這就是蒲豐投針問題（Buffon's needle）。針會否跨過最近的那條線，只取決於針尖朝着該線伸出多遠，而這段伸出的距離正是半根針在垂直方向上的投影（projection）。先固定針的中心與線的距離，單憑角度求出條件概率（conditional probability），然後再對距離取平均。第一部分的積分正是這個取平均的過程所產生的積分，這就是題目由它開始的原因。

解答

積分

設 $y = \cos^{-1} t$，則 $t = \cos y$，$\mathrm{d}t = -\sin y \,\mathrm{d}y$。當 $t = 0$ 時，$y = \frac{\pi}{2}$；當 $t = 1$ 時，$y = 0$。因此

$$\int_0^1 \cos^{-1} t \,\mathrm{d}t = \int_{\pi/2}^0 y\,(-\sin y)\,\mathrm{d}y = \int_0^{\pi/2} y \sin y \,\mathrm{d}y.$$

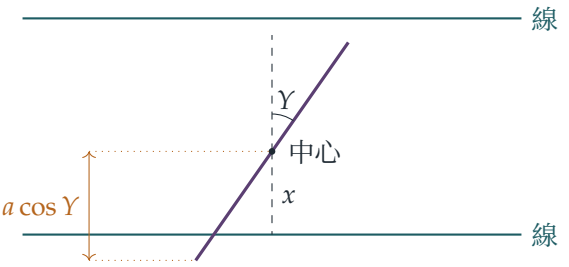
取 $u = y$ 和 $\mathrm{d}v = \sin y \,\mathrm{d}y$，以分部積分法（integration by parts）計算，

$$\begin{aligned}\int_0^{\pi/2} y \sin y \,\mathrm{d}y &= \Big[-y\cos y\Big]_0^{\pi/2} + \int_0^{\pi/2} \cos y \,\mathrm{d}y \\ &= 0 + \Big[\sin y\Big]_0^{\pi/2} = 1,\end{aligned}$$

這正是所求的結果。

（另一做法：這個積分是 $y = \cos^{-1} t$ 與兩條坐標軸之間的面積，而這個區域與 $t = \cos y$ 在 $0 \le y \le \frac{\pi}{2}$ 之下的區域相同，面積即 $\int_0^{\pi/2} \cos y \, \mathrm{d}y = 1$。）

已知 $X = x$ 時的概率



針的每一半長度為 a，並與各線的垂直線成角 Y，所以針的兩端分別位於中心兩側，與中心的垂直距離為 $a\cos Y$。最近的線與中心相距 x，故此針跨過該線，當且僅當

$$a\cos Y \ \ge \ x.$$

（等號成立時，針尖剛好碰到線，這種情況的概率為零，所以是否把它計算在內並無影響。）

> **想一想** 怎樣把 $a\cos Y \ge x$ 化為只涉及 Y 的條件？

由於 cos 在 $[0, \frac{\pi}{2}]$ 上是遞減的，答案視乎 x/a 這個值能否取到。

- 若 $x \ge a$，則 $a\cos Y \le a \le x$，而等號只在 $Y = 0$ 且 $x = a$ 時成立，所以概率為 0。
- 若 $0 \le x < a$，則 $\cos Y \ge x/a$ 等價於 $Y \le \cos^{-1}(x/a)$，其中 $\cos^{-1}(x/a) \in (0, \frac{\pi}{2}]$。由於 Y 在 $[0, \frac{\pi}{2}]$ 上均勻分佈（uniform），

$$\mathrm{P}(\text{跨線} \mid X = x) = \frac{\cos^{-1}(x/a)}{\pi/2} = \frac{2}{\pi}\cos^{-1}\frac{x}{a}.$$

這裏用到了 X 與 Y 的獨立性（independence）：即使知道 $X = x$，Y 仍然在 $[0, \frac{\pi}{2}]$ 上均勻分佈。

一般投擲的概率

現設 $a < b$。X 的概率密度（density）在 $[0, b]$ 上為 $1/b$。因為 $a < b$，有可能跨線的區間 $[0, a)$ 位於 $[0, b]$ 之內，而條件概率在 $[a, b]$ 上為零。根據全概率公式（law of total probability），

$$\mathrm{P}(\text{跨線}) = \int_0^b \mathrm{P}(\text{跨線} \mid X = x)\,\frac{1}{b}\,\mathrm{d}x = \frac{1}{b}\int_0^a \frac{2}{\pi}\cos^{-1}\frac{x}{a}\,\mathrm{d}x.$$

代入 $x = at$，$\mathrm{d}x = a\,\mathrm{d}t$，並利用第一部分的結果，

$$\mathrm{P}(\text{跨線}) = \frac{2a}{\pi b}\int_0^1 \cos^{-1} t\,\mathrm{d}t = \frac{2a}{\pi b},$$

這正是所求的結果。由於 $a < b$，針比線與線之間的間距 $2b$ 短，絕不可能碰到第二條線，所以「跨過某一條線」與「跨過最近的線」是同一事件。

驗算

- 當 $a \to b$ 時，答案趨向 $2/\pi \approx 0.637$，小於 1，這是必然的：即使針與間距一樣長，它仍然經常近乎平行於各線而落下。
- 答案與 a 成正比：把一根短針的長度加倍，跨線的機會也加倍，這與「跨線次數的期望值（expected number）是長度的線性函數」這一事實吻合。
- 按實際的幾何情況進行模擬（中心在一條帶狀區域內均勻分佈，角度均勻分佈），取 $a = 1$、$b = 2$，投擲 4×10^5 次後得出 0.3183，而 $1/\pi \approx 0.31831$。以中點法則作數值積分（midpoint quadrature），得 $\int_0^1 \cos^{-1} t\,\mathrm{d}t = 1.0000000$。

常見錯誤

- 把 $a\cos Y$ 誤寫成 $a\sin Y$：這裏的 Y 是由各線的垂直線量起，而不是由各線本身量起。
- 以公式 $\frac{2}{\pi}\cos^{-1}(x/a)$ 在整個 $[0, b]$ 上對條件概率積分；這個公式在 $x > a$ 時沒有定義，而條件概率在該處為 0。
- 對 X 取平均時忘記乘上密度 $1/b$。

STEP 之外

蒲豐（Buffon）在十八世紀提出這個問題，它是幾何概率（geometric probability）的經典例子：把概率當作參數空間中的面積來計算，這裏的參數空間是矩形 $[0, b] \times [0, \frac{\pi}{2}]$。由於答案涉及 π，投針便可以為 π 提供一個（非常緩慢的）蒙地卡羅（Monte Carlo）估計。這個問題有一個漂亮的推廣，稱為「蒲豐麵條」（Buffon's noodle）：利用期望值的線性性質（linearity of expectation）可以證明，任何形狀的曲線，其跨線次數的期望值只取決於曲線的長度。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率入門教科書）涵蓋了這類連續聯合分佈（continuous joint distributions）與取條件的方法。

詳細解答

STEP II 1993，第 1 題

純數學 • 組合博弈策略

> **STEP II 1993，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/93-s2-q1

關鍵思路

一切都建基於一個點算上的事實。兩位玩家都恰好有 9 可供分配，所以 Betty 不可能在每一塊場地上都與 Adam 打和或勝過他，同時又在其中一塊場地上明確取勝：那樣她至少要用上 $9+1$。因此，每當 Betty 勝出，她必定是恰好贏兩塊場地而輸掉第三塊。掌握了這一點，第一部分就變成一套固定做法，而兩個三元組的部分則變成簡短的分情況討論（case analysis）。

解答

全文中，「Betty 贏得場地 i」指 $b_i > a_i$，「Adam 贏得場地 i」指 $a_i > b_i$，而 $a_i = b_i$ 的場地則為打和。

一條引理：Betty 怎樣才能勝出

設 Betty 的三元組（triple）擊敗 Adam 的三元組 (a_1, a_2, a_3)，其中 $a_1 + a_2 + a_3 = 9$。

- 她不可能三塊場地全贏，因為那樣需要對每個 i 都有 $b_i \geq a_i + 1$，於是 $b_1 + b_2 + b_3 \geq 12$。
- 假設她恰好贏一塊場地。由於她的得分比 Adam 多，Adam 的得分是 0，所以其餘兩塊場地不是打和就是 Betty 贏。於是對每個 i 都有 $b_i \geq a_i$，而且至少有一處是嚴格不等式（strict inequality），因此 $b_1 + b_2 + b_3 \geq 9 + 1 = 10$。這是不可能的。

她的得分比 Adam 多，所以她至少得一分。因此她恰好贏兩塊場地。所以 Betty 擊敗 (a_1, a_2, a_3) 當且僅當（if and only if）對某一對場地 i, j 有

$$b_i \geq a_i + 1, \qquad b_j \geq a_j + 1,$$

並且她輸掉餘下的場地 k。留意輸掉場地 k 是自動成立的，因為 $b_k = 9 - b_i - b_j \leq 9 - a_i - a_j - 2 = a_k - 2$。

Betty 總能勝出

把場地重新編號，使得 $a_1 \geq a_2 \geq a_3$。那麼 $3a_1 \geq a_1 + a_2 + a_3 = 9$，所以 $a_1 \geq 3$。Betty 放棄 Adam 最強的那塊場地，而在另外兩塊場地上各比他多一：

$$(b_1, b_2, b_3) = (a_1 - 2,\ a_2 + 1,\ a_3 + 1).$$

這些都是非負整數（因為 $a_1 - 2 \geq 1$），而且和為 9。Betty 贏得場地 2 和場地 3，並輸掉場地 1，因為 $a_1 - 2 < a_1$。她得 2 分而 Adam 得 1 分，所以她勝出。

Adam 可以確保自己永不勝出

取 $(a_1, a_2, a_3) = (9, 0, 0)$。Adam 不可能在場地 2 或場地 3 上得分，因為 $b_2, b_3 \geq 0$。所以他最多得一分，而且只能靠贏得場地 1 來取得，這即是說 $b_1 \leq 8$。在這情況下 $b_2 + b_3 = 9 - b_1 \geq 1$，所以 Betty 在場地 2 或場地 3 上有一個正數項，因而至少得一分。要麼 Adam 得 0 分，要麼他得 1 分而 Betty 至少得 1 分。無論哪種情況，他的得分都不會多於 Betty，所以不論 Betty 怎樣做，他都永不勝出。

（以窮舉搜尋（brute-force search）可知，$(9, 0, 0)$ 及其各種重新排列是具有這個性質的僅有三元組。面對任何其他三元組，Betty 只要作出某個愚蠢的選擇，就會把勝利拱手讓給 Adam。）

兩個三元組：$(5, 3, 1)$ 和 $(3, 1, 5)$

我們需要證明沒有任何三元組 (b_1, b_2, b_3) 能同時擊敗這兩者。根據引理，Betty 面對每個三元組都需要贏兩塊場地。

對 $(5, 3, 1)$。贏得場地 1 和場地 2 要用上 $6 + 4 = 10 > 9$，這是不可能的。所以她必須贏得場地 3，再加上場地 1 或場地 2：

$$b_3 \geq 2 \quad \text{及} \quad (b_1 \geq 6 \ \text{或}\ b_2 \geq 4). \tag{A}$$

對 $(3, 1, 5)$。贏得場地 1 和場地 3 要用上 $4 + 6 = 10 > 9$，這是不可能的。所以她必須贏得場地 2，再加上場地 1 或場地 3：

$$b_2 \geq 2 \quad \text{及} \quad (b_1 \geq 4 \ \text{或}\ b_3 \geq 6). \tag{B}$$

想一想 兩個條件都涉及場地的選擇。我們怎樣把它們合併起來而不遺漏任何情況？

按 (B) 中的第二個選擇分情況討論。

- 若 $b_3 \geq 6$，則 $b_1 + b_2 \leq 3$，所以 $b_1 \geq 6$ 和 $b_2 \geq 4$ 都不可能成立，這與 (A) 矛盾。
- 若 $b_1 \geq 4$，則連同 (B) 中的 $b_2 \geq 2$ 和 (A) 中的 $b_3 \geq 2$，再加上 (A) 中的二擇其一，可得

$$b_1 + b_2 + b_3 \geq 6 + 2 + 2 = 10 \qquad \text{或} \qquad b_1 + b_2 + b_3 \geq 4 + 4 + 2 = 10.$$

  兩者都不可能，因為總和是 9。

每種情況都導致矛盾，所以沒有單一個三元組能同時擊敗 $(5,3,1)$ 和 $(3,1,5)$。Adam 憑這一對三元組勝出，正如所求。

驗算

總共只有 $\binom{11}{2} = 55$ 個三元組，所以一切都可以用窮舉法檢查。電腦證實，Adam 的 55 個三元組每一個都被 Betty 上述的做法擊敗，而且每一個致勝的回應都恰好贏兩塊場地。電腦亦列出擊敗 $(5,3,1)$ 的 13 個三元組和擊敗 $(3,1,5)$ 的 13 個三元組，並發現兩張清單沒有共同的三元組。舉一個具體例子，$(4,2,3)$ 以 2–1 擊敗 $(3,1,5)$，但面對 $(5,3,1)$ 時它只贏得場地 3，並輸掉場地 1 和場地 2，所以它以 1–2 落敗。

常見錯誤

- 以 $(3,3,3)$ 作為 Adam「永不勝出」的三元組。Betty 可以出 $(2,2,5)$ 而以 1–2 落敗，所以 Adam 憑它是可以勝出的。
- 忘記了 Betty 若只贏一塊場地而勝出，就需要 Adam 一分不得，繼而沒有利用總和 9 來排除這個情況。
- 在兩個三元組的部分，只檢查幾個自然想到的回應，例如 $(6,0,3)$ 或 $(0,2,7)$，而沒有證明任何回應都行不通。

STEP 之外

布洛托上校博弈（Colonel Blotto）是博弈論（game theory）中的經典。由於 Betty 後行並掌握全部資訊，她總能勝出，所以有趣的版本是讓兩位玩家同時行動。這時通常沒有

最佳的純策略（pure strategy），而解答是一個混合策略（mixed strategy）：三元組上的一個概率分佈（probability distribution），其存在由馮紐曼（von Neumann）的極小極大定理（minimax theorem）保證。本題中兩個三元組的變體是朝這個方向邁出的第一步，因為 Adam 的最佳防守就是把他的選擇分散到幾個三元組上。想做同類精神的題目，可參閱 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題集）。至於支撐混合策略的概率工具，可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率入門）。

詳細解答

STEP II 1993，第 2 題

純數學 • 三角函數與代數函數的積分

STEP II 1993，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/93-s2-q2

關鍵思路

在第 (i) 部分，餘弦的乘積不易積分，但和卻很容易，所以我們用積化和差公式（factor formula）把乘積化為和。之後一切都取決於一個事實：除非 $k = 0$，否則 $\cos kx$ 在一個完整週期（period）上的積分為零。特殊情況正正就是 $m - n$ 或 $m + n$ 為零的情況。在第 (ii) 部分，我們把平方根內的式子合併成單一分數，再進行有理化（rationalise），從而把被積函數（integrand）拆成兩個標準部分。

解答

第 (i) 部分

由積化和差公式 $2\cos A\cos B = \cos(A-B) + \cos(A+B)$ 得

$$\cos(mx)\cos(nx) = \tfrac{1}{2}\big[\cos\big((m-n)x\big) + \cos\big((m+n)x\big)\big].$$

因此我們需要對整數 k 求 $\int_0^{2\pi} \cos(kx)\,\mathrm{d}x$：

$$\int_0^{2\pi} \cos(kx)\,\mathrm{d}x = \begin{cases} \left[\dfrac{\sin kx}{k}\right]_0^{2\pi} = \dfrac{\sin 2k\pi}{k} = 0, & k \neq 0, \\ \displaystyle\int_0^{2\pi} 1\,\mathrm{d}x = 2\pi, & k = 0. \end{cases}$$

這裏用到 k 是整數，所以 $\sin 2k\pi = 0$。當 $k = 0$ 時不能除以 k，因此這個情況要分開處理。

? **想一想** 共有多少個不同的情況？它們會否重疊？

$k = m - n$ 和 $k = m + n$ 各自或為零，或不為零，而兩者同時為零只會在 $m = n = 0$ 時發生。於是有三個情況：

- $m \neq n$ 且 $m \neq -n$，即 $|m| \neq |n|$：兩個積分都等於零，答案是 0。
- $m - n$、$m + n$ 之中恰好有一個為零，即 $|m| = |n| \neq 0$：其中一項貢獻 $\frac{1}{2} \cdot 2\pi$，另一項貢獻 0，所以答案是 π。
- $m = n = 0$：兩項各貢獻 $\frac{1}{2} \cdot 2\pi$。（直接來看，被積函數就是 1。）答案是 2π。

總括而言，

$$\int_0^{2\pi} \cos(mx)\cos(nx)\,\mathrm{d}x = \begin{cases} 0, & |m| \neq |n|, \\ \pi, & |m| = |n| \neq 0, \\ 2\pi, & m = n = 0. \end{cases}$$

留意 $m = -n$ 與 $m = n$ 屬於同一類，因為 cos 是偶函數（even function），$\cos(-nx) = \cos(nx)$。

第 (ii) 部分

被積函數 $\sqrt{1 + 1/x} = \sqrt{(x+1)/x}$ 只在 $(x+1)/x \geq 0$ 時才是實數，即 $x > 0$ 或 $x \leq -1$。我們先處理 $x > 0$。

$x > 0$ 的情況。 把根號內的分子和分母同乘以 $x + 1$：

$$\sqrt{\frac{x+1}{x}} = \sqrt{\frac{(x+1)^2}{x(x+1)}} = \frac{x+1}{\sqrt{x^2+x}},$$

其中利用了 $x + 1 > 0$ 來取 $(x+1)^2$ 的平方根。現在把分子拆開，使其中一部分是 $x^2 + x$ 的導數的一半：

$$\frac{x+1}{\sqrt{x^2+x}} = \frac{1}{2} \cdot \frac{2x+1}{\sqrt{x^2+x}} + \frac{1}{2} \cdot \frac{1}{\sqrt{x^2+x}}.$$

? **想一想** 怎樣對 $1/\sqrt{x^2+x}$ 積分？

我們用配方法（completing the square）：$x^2 + x = \left(x + \frac{1}{2}\right)^2 - \frac{1}{4}$。然後標準結果 $\int (u^2 - a^2)^{-1/2}\,\mathrm{d}u = \ln\left(u + \sqrt{u^2 - a^2}\right)$（在 $u > a > 0$ 時成立）可以應用，取 $u = x + \frac{1}{2}$ 和 $a = \frac{1}{2}$：

$$\int \frac{\mathrm{d}x}{\sqrt{x^2+x}} = \ln\left(x + \tfrac{1}{2} + \sqrt{x^2+x}\right) + \text{常數} = \ln\left(2x + 1 + 2\sqrt{x^2+x}\right) + \text{常數}.$$

第一部分是 $\int \frac{1}{2}(2x+1)(x^2+x)^{-1/2}\,\mathrm{d}x = \sqrt{x^2+x}$。所以，當 $x > 0$ 時，

$$\int \sqrt{1 + \frac{1}{x}}\,\mathrm{d}x = \sqrt{x^2+x} + \frac{1}{2}\ln\left(2x + 1 + 2\sqrt{x^2+x}\right) + C.$$

由於 $2x+1+2\sqrt{x^2+x} = \left(\sqrt{x}+\sqrt{x+1}\right)^2$，這也可以寫成

$$\sqrt{x(x+1)} + \ln\left(\sqrt{x}+\sqrt{x+1}\right) + C.$$

用代換（substitution）$x = \sinh^2 t$ 可以直接得出這第二個形式。被積函數變成 $2\cosh^2 t = 1+\cosh 2t$，其積分為 $t + \sinh t \cosh t$。

$x \le -1$ 的情況。 這時 $x+1 \le 0$，所以 $\sqrt{(x+1)^2} = -(x+1)$，而

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

同樣的拆法仍然適用，但現在 $x+\frac{1}{2} < 0$，所以我們需要 $\int (u^2-a^2)^{-1/2}\,\mathrm{d}u = \ln\left|u+\sqrt{u^2-a^2}\right|$，此式在 $|u| > a$ 時成立。當 $x < -1$ 時，

$$\int \sqrt{1+\frac{1}{x}}\,\mathrm{d}x = -\sqrt{x^2+x} - \frac{1}{2}\ln\left|2x+1+2\sqrt{x^2+x}\right| + C.$$

因為 $\left(2x+1+2\sqrt{x^2+x}\right)\left(2x+1-2\sqrt{x^2+x}\right) = 1$，這個對數項也可以寫成 $+\frac{1}{2}\ln\left|2x+1-2\sqrt{x^2+x}\right|$。

驗算

- 第 (i) 部分：取 $m = n = 1$，被積函數是 $\cos^2 x$，其平均值為 $\frac{1}{2}$，所以積分是 $\frac{1}{2}\cdot 2\pi = \pi$，與預期相符。對每一對 $-6 \le m, n \le 6$ 用數值方法（辛普森法則，Simpson's rule）計算，結果都與三個情況的公式吻合。
- 第 (ii) 部分：對答案求導。$\sqrt{x^2+x}$ 的導數是 $\dfrac{2x+1}{2\sqrt{x^2+x}}$。$\frac{1}{2}\ln(\cdots)$ 的導數是 $\dfrac{1}{2\sqrt{x^2+x}}$。兩者相加得 $\dfrac{x+1}{\sqrt{x^2+x}}$，正如所求。
- 數值上，$\int_1^3 \sqrt{1+1/x}\,\mathrm{d}x \approx 2.48547$，用辛普森法則和用公式 $\sqrt{12} - \sqrt{2} + \frac{1}{2}\ln\dfrac{7+4\sqrt{3}}{3+2\sqrt{2}}$ 計算，結果一致。

常見錯誤

- 忘記 $m = -n \neq 0$ 的情況；它與 $m = n \neq 0$ 一樣，答案是 π。
- 在 $k = 0$ 時寫出 $\dfrac{\sin kx}{k}$；這個情況需要另行處理。
- 沒有檢查正負號就寫 $\sqrt{(x+1)^2} = x+1$，這會令 $x < -1$ 時的答案出錯。

STEP 之外

第 (i) 部分說明函數 $1, \cos x, \cos 2x, \ldots$ 在 $[0, 2\pi]$ 上是正交（orthogonal）的。這正是傅立葉級數（Fourier series）背後的事實：若 $f(x) = \sum a_n \cos nx$，乘以 $\cos mx$ 再積分，就能抽出單獨一個係數 a_m。同樣的想法，即以積分扮演點積（dot product）的角色，是函數空間上許多線性代數（linear algebra）內容的基礎。Strang 的 *Introduction to Linear Algebra*（線性代數入門）講解正交性，而 Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程）則展示了傅立葉級數如何實際用於求解熱方程（heat equation）。

詳細解答

STEP II 1993，第 3 題

純數學 • 里卡蒂微分方程

STEP II 1993，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/93-s2-q3

關鍵思路

兩條方程都因為 y^2 項而成為非線性（nonlinear）方程，所以標準的一階方程解法都不能直接套用。題目給出的代換（substitution）正是為了消去這個 y^2 項而設計的。在 (i) 中，把 y 寫成 u'/u 的倍數，平方項便會互相抵消，剩下一條常係數的二階線性方程。在 (ii) 中，$y = 1/x$ 剛好是一個精確解，而寫成 $y = 1/x + 1/v$ 後，方程便化為關於 v 的一階線性方程，可用積分因子（integrating factor）求解。

解答

第 (i) 部分

設 $y = -\dfrac{u'}{3u}$，其中 $u' = \dfrac{\mathrm{d}u}{\mathrm{d}x}$。由商法則（quotient rule），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{u''}{3u} + \frac{(u')^2}{3u^2}, \qquad \text{而} \qquad 3y^2 = 3 \cdot \frac{(u')^2}{9u^2} = \frac{(u')^2}{3u^2}.$$

於是那些麻煩的項互相抵消：

$$\frac{\mathrm{d}y}{\mathrm{d}x} - 3y^2 = -\frac{u''}{3u}.$$

因此方程 $\dfrac{\mathrm{d}y}{\mathrm{d}x} - y - 3y^2 = -2$ 變成

$$-\frac{u''}{3u} + \frac{u'}{3u} = -2.$$

兩邊乘以 $-3u$（其中 $u \neq 0$，這本來就是 y 有定義所必需的），得

$$u'' - u' - 6u = 0.$$

輔助方程（auxiliary equation）為 $m^2 - m - 6 = (m-3)(m+2) = 0$，所以 $m = 3$ 或 $m = -2$，而

$$u = Ae^{3x} + Be^{-2x}$$

其中 A 和 B 為常數，且不同時為零。於是 $u' = 3Ae^{3x} - 2Be^{-2x}$，而

$$y = -\frac{u'}{3u} = \frac{2Be^{-2x} - 3Ae^{3x}}{3\left(Ae^{3x} + Be^{-2x}\right)}.$$

想一想 一階方程應該只有一個任意常數（arbitrary constant）。為甚麼我們會有兩個？

其實只有 A 與 B 的比才有影響：把 u 乘以一個非零常數，並不會改變 u'/u。若 $A \neq 0$，把分子和分母同時除以 Ae^{3x}，並記 $K = B/A$：

$$y = \frac{2Ke^{-5x} - 3}{3\left(1 + Ke^{-5x}\right)} = \frac{2K - 3e^{5x}}{3\left(e^{5x} + K\right)},$$

其中 K 為任意常數。$A = 0$ 的情況給出常數解 $y = \frac{2}{3}$，它是這族解當 $K \to \infty$ 時的極限。取 $K = 0$ 則得另一個常數解 $y = -1$。作為驗算，常數解滿足 $-y - 3y^2 = -2$，即 $(3y-2)(y+1) = 0$。

第 (ii) 部分

設 $y = \frac{1}{x} + \frac{1}{v}$。則

$$\frac{dy}{dx} = -\frac{1}{x^2} - \frac{v'}{v^2},$$

因此

$$\begin{aligned} x^2\frac{dy}{dx} &= -1 - \frac{x^2v'}{v^2}, \\ xy &= 1 + \frac{x}{v}, \\ x^2y^2 &= x^2\left(\frac{1}{x^2} + \frac{2}{xv} + \frac{1}{v^2}\right) = 1 + \frac{2x}{v} + \frac{x^2}{v^2}. \end{aligned}$$

把這三行相加，方程 $x^2y' + xy + x^2y^2 = 1$ 變成

$$1 - \frac{x^2v'}{v^2} + \frac{3x}{v} + \frac{x^2}{v^2} = 1, \qquad \text{即} \qquad -\frac{x^2v'}{v^2} + \frac{3x}{v} + \frac{x^2}{v^2} = 0.$$

兩邊乘以 v^2/x^2（其中 $x \neq 0$，這是方程本身的要求），得線性方程

$$\frac{dv}{dx} - \frac{3}{x}v = 1.$$

積分因子為 $\exp\left(-\int \frac{3}{x}\,\mathrm{d}x\right) = x^{-3}$（在 $x > 0$ 或 $x < 0$ 其中一個區間上；我們可以略去絕對值符號，因為積分因子乘以一個常數後同樣適用）。故此

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{v}{x^3}\right) = \frac{1}{x^3} \quad\Longrightarrow\quad \frac{v}{x^3} = -\frac{1}{2x^2} + c \quad\Longrightarrow\quad v = cx^3 - \tfrac{1}{2}x = \frac{x\left(Kx^2 - 1\right)}{2},$$

其中 c 為任意常數，$K = 2c$。所以

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

其中 K 為任意常數。取 $K = 0$ 得 $y = -1/x$。這個代換假設了 $1/v$ 存在，所以遺漏了 $y = 1/x$ 本身（「$v = \infty$」）。它同樣是一個解，因為 $x^2\left(-\frac{1}{x^2}\right) + 1 + 1 = 1$，而且它是這族解當 $K \to \infty$ 時的極限。

驗算

- 在 (i) 中，當 $x = 0$ 時，解的值為 $y(0) = \dfrac{2K - 3}{3(1 + K)}$。當 $x \to +\infty$ 時，$y \to -1$；當 $x \to -\infty$ 時，$y \to \frac{2}{3}$（當 $K \neq 0$）。所以解由一個常數解走向另一個常數解，正如我們所料。
- 在 (ii) 中，取 $K = 1$ 及 $x = 2$：$y = \frac{5}{6}$，$y'(2) = -\frac{31}{36}$。於是 $x^2y' + xy + x^2y^2 = -\frac{31}{9} + \frac{5}{3} + \frac{25}{9} = 1$，符合要求。
- 我們以有限差分（finite differences）在數百個隨機點及隨機的 K 值上，對兩個通解（general solution）作了數值驗算。代入原方程所得的殘差，在捨入誤差範圍內為零。

常見錯誤

- 在 (i) 的答案中保留兩個常數 A、B，而沒有留意到只有它們的比才有影響。
- 在 (ii) 中忘記了 x^2y^2 交叉項的因子 2，以致把 $3v/x$ 變成 $2v/x$，得出錯誤的積分因子。
- 遺漏了代換所無法得到的解 $y = 1/x$（或 (i) 中的 $y = \frac{2}{3}$）。

STEP 之外

形如 $y' = p(x) + q(x)y + r(x)y^2$ 的方程稱為里卡蒂方程（Riccati equation）。本題的兩個技巧都是普遍適用的。代換 $y = -u'/(ru)$ 總能把里卡蒂方程化為二階線性方程。而如果你知道一個特解（particular solution）y_1，則 $y = y_1 + 1/v$ 總能給出一條關於 v 的一階線性方程。里卡蒂方程與二階線性方程之間的這種聯繫，在量子力學（quantum

mechanics）和控制理論（control theory）中會再次出現。Boyce 與 DiPrima 的 *Elementary Differential Equations*（微分方程入門教科書）中有關於里卡蒂方程的習題。

詳細解答

STEP II 1993，第 4 題

純數學 • 偏斜線與最接近距離

STEP II 1993，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/93-s2-q4

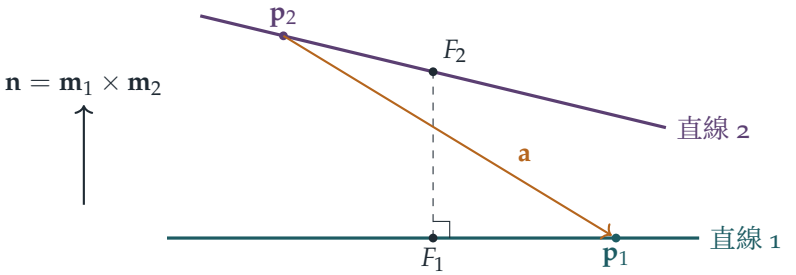
關鍵思路

由一條直線上的點連到另一條直線上的點的任何向量，沿公共法向量（common normal）$\mathbf{m}_1 \times \mathbf{m}_2$ 的分量（component）都相同，因為沿任何一條直線移動，只會加上 $\mathbf{m}_1$ 和 $\mathbf{m}_2$ 的倍數，而這兩個向量都與該法向量垂直。這個固定的分量就是最短距離。至於飛機的部分，竅門是在 A_2 的參考系（frame）中考慮：由 A_2 指向 A_1 的向量沿單一一條直線移動，所以最接近時的距離就是一點到一條直線的距離。

解答

兩條直線之間的最短距離

設 $\mathbf{a} = \mathbf{p}_1 - \mathbf{p}_2$ 及 $\mathbf{n} = \mathbf{m}_1 \times \mathbf{m}_2$，由於兩條直線不平行，後者不是零向量，而且同時垂直於 $\mathbf{m}_1$ 和 $\mathbf{m}_2$。



草圖顯示了這兩條直線（它們是異面直線（skew lines），所以不相交）、由 $\mathbf{p}_2$ 指向 $\mathbf{p}_1$ 的向量 $\mathbf{a}$，以及與 $\mathbf{n}$ 平行的公垂線（common perpendicular）F_1F_2。

由直線 2 上一點指向直線 1 上一點的一般向量是

$$\mathbf{d} = \mathbf{a} + t_1\mathbf{m}_1 - t_2\mathbf{m}_2.$$

由於 $\mathbf{m}_1 \cdot \mathbf{n} = \mathbf{m}_2 \cdot \mathbf{n} = 0$，它沿單位法向量（unit normal）$\hat{\mathbf{n}} = \mathbf{n}/|\mathbf{n}|$ 的分量是

$$\mathbf{d} \cdot \hat{\mathbf{n}} = \mathbf{a} \cdot \hat{\mathbf{n}},$$

無論 t_1, t_2 怎樣選取，結果都一樣。向量的長度不會小於它的任何一個分量，所以 $|\mathbf{d}| \geq |\mathbf{a} \cdot \hat{\mathbf{n}}|$，而等號恰好在 $\mathbf{d}$ 與 $\mathbf{n}$ 平行時成立。等號確實可以成立：$\mathbf{m}_1, \mathbf{m}_2, \mathbf{n}$ 不共面，所以它們構成一組基（basis），我們可以寫成 $\mathbf{a} = \alpha\mathbf{m}_1 + \beta\mathbf{m}_2 + \gamma\mathbf{n}$；取 $t_1 = -\alpha$、$t_2 = \beta$，便剩下 $\mathbf{d} = \gamma\mathbf{n}$。這個選擇給出草圖中的垂足 F_1, F_2。從圖中看，最短距離就是 $\mathbf{a}$ 在 $\mathbf{n}$ 方向上的「高度」。因此最短距離是

$$|\mathbf{a} \cdot \hat{\mathbf{n}}| = \frac{|(\mathbf{p}_1 - \mathbf{p}_2) \cdot (\mathbf{m}_1 \times \mathbf{m}_2)|}{|\mathbf{m}_1 \times \mathbf{m}_2|}.$$

第 (i) 部分

這裏 $\mathbf{a} = (1, 1, 1)$。把 $\mathbf{m}_1$ 和 $\mathbf{m}_2$ 換成它們的任何非零倍數，公式都不變，因為這些因子在分子和分母中互相約去，所以我們可以用 $(4, 3, 0)$ 和 $(0, -3, 1)$：

$$(4, 3, 0) \times (0, -3, 1) = (3, -4, -12), \qquad |(3, -4, -12)| = \sqrt{9 + 16 + 144} = 13.$$

於是 $(1, 1, 1) \cdot (3, -4, -12) = -13$，所以最短距離是 $13/13 = 1$。

第 (ii) 部分

在時間 t，兩架飛機分別位於 $\mathbf{p}_1 + v_1 t\,\mathbf{m}_1$ 和 $\mathbf{p}_2 + v_2 t\,\mathbf{m}_2$，所以由 A_2 指向 A_1 的向量是

$$\mathbf{r}(t) = \mathbf{a} + t\mathbf{w}, \qquad \mathbf{w} = v_1\mathbf{m}_1 - v_2\mathbf{m}_2.$$

由於 $\mathbf{m}_1, \mathbf{m}_2$ 不平行，$\mathbf{w} \neq \mathbf{0}$（除非兩個速率都是零）。我們把飛行過程理解為包括所有時間，即 $t = 0$ 之前和之後。於是

$$|\mathbf{r}(t)|^2 = |\mathbf{a}|^2 + 2t\,\mathbf{a} \cdot \mathbf{w} + t^2|\mathbf{w}|^2,$$

這是 t 的二次式，首項係數為正。配方（completing the square）後得

$$|\mathbf{r}(t)|^2 = |\mathbf{w}|^2\Big(t + \frac{\mathbf{a} \cdot \mathbf{w}}{|\mathbf{w}|^2}\Big)^2 + |\mathbf{a}|^2 - \frac{(\mathbf{a} \cdot \mathbf{w})^2}{|\mathbf{w}|^2},$$

所以最小值在 $t = -\mathbf{a} \cdot \mathbf{w}/|\mathbf{w}|^2$ 時取得，其值為

$$d^2 = \frac{|\mathbf{p}_1 - \mathbf{p}_2|^2\,|v_1\mathbf{m}_1 - v_2\mathbf{m}_2|^2 - [(\mathbf{p}_1 - \mathbf{p}_2) \cdot (v_1\mathbf{m}_1 - v_2\mathbf{m}_2)]^2}{|v_1\mathbf{m}_1 - v_2\mathbf{m}_2|^2},$$

這正是所求。從幾何上看，d 是原點到直線 $\mathbf{r} = \mathbf{a} + t\mathbf{w}$ 的距離，即 $|\mathbf{a}| \sin\theta$，其中 θ 是 $\mathbf{a}$ 與 $\mathbf{w}$ 之間的夾角。

第 (iii) 部分

先給出一個不需要任何代數運算的下界：在每一刻，A_1 都在直線 1 上，A_2 都在直線 2 上，所以由第 (i) 部分可知 $d \geq 1$。等號成立的話，兩架飛機必須在同一時間分別位於垂足 F_1、F_2。

> **想一想** A_2 的機師真的可以做到這一點嗎？

先求垂足。要求 $\mathbf{a} + t_1\mathbf{m}_1 - t_2\mathbf{m}_2$ 同時垂直於 $\mathbf{m}_1$ 和 $\mathbf{m}_2$，並利用 $\mathbf{a}\cdot\mathbf{m}_1 = \frac{7}{5}$、$\mathbf{a}\cdot\mathbf{m}_2 = -\frac{2}{\sqrt{10}}$ 及 $\mathbf{m}_1\cdot\mathbf{m}_2 = -\frac{9}{5\sqrt{10}}$，解這兩條線性方程便得

$$t_1 = -\tfrac{20}{13}, \qquad t_2 = \tfrac{\sqrt{10}}{13}.$$

A_1 在時間 $t_1/v_1 < 0$ 位於 F_1，但當 $v_2 > 0$ 時，A_2 在時間 $t_2/v_2 > 0$ 才位於 F_2。這兩個時間永遠不可能相同：F_1 在 $\mathbf{p}_1$ 的後方，而 F_2 在 $\mathbf{p}_2$ 的前方。所以若 A_2 沿 $\mathbf{m}_2$ 的方向飛行，$d = 1$ 這個值是達不到的。要知道可以達到甚麼值，我們利用第 (ii) 部分。

設 $v_2 = \sqrt{10}\,\lambda v_1$，所以 λ 是經過縮放的速率比。於是

$$\mathbf{w} = v_1\big(\tfrac{1}{5}(4,3,0) - \lambda(0,-3,1)\big) = \tfrac{v_1}{5}(4,\ 3+15\lambda,\ -5\lambda),$$

而因子 $\frac{v_1}{5}$ 在 d^2 的公式中會約去。利用 $|\mathbf{a}|^2 = 3$，

$$\mathbf{a}\cdot(4,3+15\lambda,-5\lambda) = 7+10\lambda, \qquad |(4,3+15\lambda,-5\lambda)|^2 = 25+90\lambda+250\lambda^2,$$

所以

$$d^2 = 3 - \frac{(7+10\lambda)^2}{25+90\lambda+250\lambda^2} = \frac{26+130\lambda+650\lambda^2}{25+90\lambda+250\lambda^2} = \frac{26}{5}\cdot\frac{1+5\lambda+25\lambda^2}{5+18\lambda+50\lambda^2}.$$

對最後一個分式 N/D 求導，導數的分子可以化簡：

$$\begin{aligned} N'D - ND' &= (5+50\lambda)(5+18\lambda+50\lambda^2) - (1+5\lambda+25\lambda^2)(18+100\lambda) \\ &= 7+150\lambda+200\lambda^2 = (1+20\lambda)(7+10\lambda). \end{aligned}$$

所以，把 d^2 看成所有實數 λ 的函數時，它在 $\lambda = -\frac{7}{10}$ 取得極大值（此時 $\mathbf{a} \perp \mathbf{w}$ 而 $d = \sqrt{3}$），並在 $\lambda = -\frac{1}{20}$ 取得極小值，此時

$$d^2 = \frac{26}{5}\cdot\frac{1-\frac{1}{4}+\frac{1}{16}}{5-\frac{9}{10}+\frac{1}{8}} = \frac{26}{5}\cdot\frac{13/16}{169/40} = 1.$$

這就是下界 $d = 1$，在 $v_2 = -\frac{v_1}{2\sqrt{10}}$ 時達到，與由垂足求得的 t_2/t_1 吻合。

所以兩架飛機最接近時可以相距 $d = 1$，即兩條直線之間的最短距離，而不能再近。要做到這一點，A_2 的機師須沿直線 2 以 $-\mathbf{m}_2$ 的方向、以速率 $\frac{v_1}{2\sqrt{10}}$ 飛行：用題目的記號，

就是 $v_2 = -\frac{v_1}{2\sqrt{10}}$，這是沿 $\mathbf{m}_2$ 的速度而不是速率。這樣兩架飛機便會在同一刻到達公垂線的兩個垂足。這重現了第 (i) 部分的答案，相信這正是考官的原意。留意答案只取決於比值 v_2/v_1。

*關於正速率的備註。*如果 v_2 必須是真正的速率，即 $v_2 > 0$ 及 $\lambda > 0$，那麼 $(1+20\lambda)(7+10\lambda) > 0$，而 d^2 對 v_2 嚴格遞增：

$$d^2 \to \frac{26}{25} \quad \text{當} v_2 \to 0^+, \qquad d^2 \to \frac{13}{5} \quad \text{當} v_2 \to \infty.$$

按這種理解，並不存在最佳速率：機師飛得愈慢，效果愈好，最接近時的距離可以任意接近 $\frac{\sqrt{26}}{5} \approx 1.020$，即 $\mathbf{p}_2$ 到直線 $\imath$ 的距離，但永遠達不到這個值。

驗算

- (i) 中的公垂線：取 $t_1 = -\frac{20}{13}$、$t_2 = \frac{\sqrt{10}}{13}$ 時，連接兩點的向量是 $\frac{1}{13}(-3,4,12)$，它與 $(3,-4,-12)$ 平行，長度恰好是 1。
- 用電腦對隨機選取的速率以直接搜尋法求 $|\mathbf{r}(t)|$ 的最小值，結果與 (ii) 中的公式吻合，誤差在 10^{-6} 以內。電腦亦在 $[-2,2]$ 上掃描了 v_2/v_1：最小值是 $d^2 = 1.0000$，在 $v_2/v_1 \approx -0.1581 = -1/(2\sqrt{10})$ 時取得，而當 $v_2 > 0$ 時，d^2 由 1.04 開始上升。
- 極限情況：$v_2 = 0$ 時 A_2 是固定在 $\mathbf{p}_2$ 的一點，而 $3 - (\frac{7}{5})^2 = \frac{26}{25}$ 確實是 $\mathbf{p}_2$ 到直線 $\imath$ 的距離的平方。

常見錯誤

- 以「1，由第 (i) 部分可得」回答 (iii)，卻沒有檢查兩架飛機能否在同一刻到達公垂線的垂足。
- 忘記了 (ii) 中的公式不受 $\mathbf{w}$ 的縮放影響，結果帶着 v_1 做了好幾頁代數運算。
- 在直線距離公式中漏掉絕對值，得出 -1。

STEP 之外

第 (ii) 部分是*正交投影*（orthogonal projection）的一個小例子：d 是 $\mathbf{a}$ 垂直於 $\mathbf{w}$ 的分量的長度，而 $|\mathbf{a}|^2|\mathbf{w}|^2 - (\mathbf{a}\cdot\mathbf{w})^2 = |\mathbf{a}\times\mathbf{w}|^2$ 是格拉姆行列式（Gram determinant），即以 $\mathbf{a}$ 和 $\mathbf{w}$ 為邊的平行四邊形面積的平方。大學第一門線性代數課程中的最小二乘法（least squares）擬合，正是建基於同一個投影概念。Strang 的 *Introduction to Linear Algebra*（線性代數入門）對此有十分精彩的闡述。

詳細解答

STEP II 1993，第 5 題

純數學 • 極坐標曲線與弧長

STEP II 1993，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/93-s2-q5

關鍵思路

R 處的直角讓我們只需用 OQ 和小角 $\delta\theta$，便可從直角三角形 ORQ 讀出小三角形 PQR 的每一條邊長。取 $Q \to P$ 時的極限（limit），弦（chord）PQ 便變成切線（tangent）。對於這條螺線，$\mathrm{d}r/\mathrm{d}\theta$ 是 r 的常數倍，所以 $r\,\mathrm{d}\theta/\mathrm{d}r$ 處處相同，而弧長（arc length）的被積函數（integrand）只是一個指數函數。

解答

曲線與半徑之間的夾角

點 R 位於直線 OP（的延長線）上，而 $\angle ORQ$ 是直角。在直角三角形 ORQ 中，斜邊（hypotenuse）是 $OQ = r + \delta r$，O 處的角是 $\delta\theta$，所以

$$QR = (r+\delta r)\sin\delta\theta, \qquad OR = (r+\delta r)\cos\delta\theta.$$

由於 $OP = r$，我們有 $PR = OR - OP = (r+\delta r)\cos\delta\theta - r$。三角形 PRQ 的直角在 R，所以

$$\tan\angle QPR = \frac{QR}{PR} = \frac{(r+\delta r)\sin\delta\theta}{(r+\delta r)\cos\delta\theta - r}.$$

當 $\delta\theta$ 很小時，我們有 $\sin\delta\theta = \delta\theta + O(\delta\theta^3)$ 和 $\cos\delta\theta = 1 + O(\delta\theta^2)$，所以

$$\tan\angle QPR = \frac{r\,\delta\theta + \delta r\,\delta\theta + \cdots}{\delta r + O(\delta\theta^2)} = \frac{r\dfrac{\delta\theta}{\delta r} + \delta\theta + \cdots}{1 + O\left(\dfrac{\delta\theta^2}{\delta r}\right)},$$

其中省略號代表與已列出的項相比很小的項。當 $Q \to P$ 時，δr 和 $\delta\theta$ 都趨向 0，且 $\delta\theta/\delta r \to \mathrm{d}\theta/\mathrm{d}r$。所以 $\delta\theta^2/\delta r = \delta\theta \cdot (\delta\theta/\delta r) \to 0$，同時 $\delta\theta \to 0$。因此

$$\tan\angle QPR \to r\frac{\mathrm{d}\theta}{\mathrm{d}r}.$$

想一想 為甚麼 $\angle QPR$ 的極限就是曲線與 OP 相交的角度？

角 QPR 是弦 PQ 與通過 P 的半徑線之間的夾角。當 $Q \to P$ 沿曲線移動時，弦 PQ 便變成 P 處的切線。所以在極限之下，$\angle QPR$ 變成切線與 OP 之間的夾角，這正是「曲線與 OP 相交的角度」的意思。因此該角是

$$\tan^{-1}\left(r\frac{\mathrm{d}\theta}{\mathrm{d}r}\right),$$

正如所求。

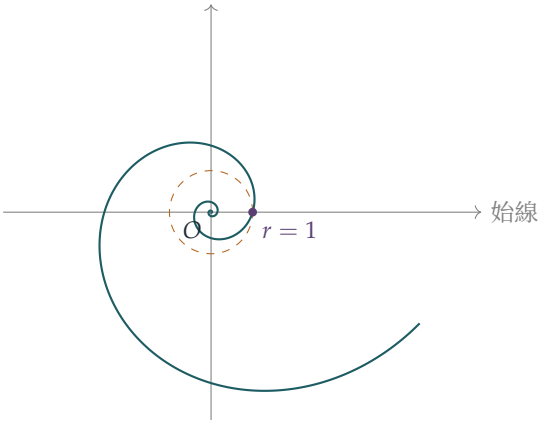
等角螺線

對於 $r = \mathrm{e}^{\theta\cot\alpha}$，

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = \cot\alpha\,\mathrm{e}^{\theta\cot\alpha} = r\cot\alpha, \qquad \text{所以} \qquad r\frac{\mathrm{d}\theta}{\mathrm{d}r} = \frac{r}{r\cot\alpha} = \tan\alpha.$$

因為 $0 < \alpha < \frac{1}{2}\pi$，角 α 位於 $\tan^{-1}$ 的值域之內，所以 $\tan^{-1}(\tan\alpha) = \alpha$。這個結果與 θ 無關，所以螺線在每一點都以角 α 與每條半徑線相交。

草圖。由於 $\cot\alpha > 0$，r 隨 θ 穩定地增加。每轉一整圈，r 便乘以固定的因子 $\mathrm{e}^{2\pi\cot\alpha}$。當 $\theta \to -\infty$ 時，$r \to 0$，所以曲線繞原點無限多圈，卻永遠到不了原點。當 $\theta \to +\infty$ 時，它以逆時針方向向外盤旋。當 $\theta = 0$ 時，它在 $r = 1$ 處與始線（initial line）相交。草圖採用 $\cot\alpha = 0.3$。



虛線圓是 $r = 1$。螺線在圓內的部分就是那無限向內捲繞的一段。

弧長

利用題目給出的公式和 $\mathrm{d}r/\mathrm{d}\theta = r\cot\alpha$，

$$\left(\frac{\mathrm{d}s}{\mathrm{d}\theta}\right)^2 = r^2 + r^2\cot^2\alpha = r^2\operatorname{cosec}^2\alpha, \qquad \text{所以} \qquad \frac{\mathrm{d}s}{\mathrm{d}\theta} = \frac{\mathrm{e}^{\theta\cot\alpha}}{\sin\alpha},$$

這裏取正根，因為 s 隨 θ 增加，而且 $\sin\alpha > 0$。

由 $r = 1$ 開始向外走的完整一圈，是從 $\theta = 0$ 到 $\theta = 2\pi$。它的長度是

$$\int_0^{2\pi} \frac{\mathrm{e}^{\theta\cot\alpha}}{\sin\alpha}\,\mathrm{d}\theta = \frac{1}{\sin\alpha}\cdot\frac{\mathrm{e}^{2\pi\cot\alpha}-1}{\cot\alpha} = \sec\alpha\left(\mathrm{e}^{2\pi\cot\alpha}-1\right).$$

$r \le 1$ 的部分對應 $\theta \le 0$，即 $-\infty < \theta \le 0$。它的長度是反常積分（improper integral）

$$\int_{-\infty}^{0} \frac{\mathrm{e}^{\theta\cot\alpha}}{\sin\alpha}\,\mathrm{d}\theta = \lim_{T\to\infty}\frac{1-\mathrm{e}^{-T\cot\alpha}}{\sin\alpha\cot\alpha} = \sec\alpha,$$

因為 $\cot\alpha > 0$ 使得 $\mathrm{e}^{-T\cot\alpha} \to 0$。所以即使這部分繞原點無限多圈，它的總長度仍是有限的，就是 $\sec\alpha$。

驗算

- 當 $\alpha \to \frac{1}{2}\pi$ 時，我們有 $\cot\alpha \to 0$，螺線變成單位圓，而每條半徑都與它成直角相交。一圈的長度是 $\sec\alpha(\mathrm{e}^{2\pi\cot\alpha}-1) \approx \sec\alpha\cdot 2\pi\cot\alpha = 2\pi\operatorname{cosec}\alpha \to 2\pi$，即圓周。
- 對 $\sec\alpha$ 有一個漂亮的幾何驗算：沿螺線由任意一點到原點的長度是 $r\sec\alpha$，這正是由該點到通過 O 且垂直於 OP 的直線的切線段長度。
- 數值上，取 $\alpha = 0.9$，把 2×10^5 條弦加起來，得出一圈的長度為 233.8349，內側部分的長度為 1.608726。這些數值與 $\sec\alpha(\mathrm{e}^{2\pi\cot\alpha}-1)$ 和 $\sec\alpha$ 吻合，至少準確至八位有效數字。

常見錯誤

- 寫下 $PR = \delta r$ 和 $QR = r\,\delta\theta$ 而不解釋原因。兩者都只在一階近似下成立，而極限論證正是這道題要考核的地方。
- 計算內側部分時在 $[0, 2\pi]$ 上積分，而不是在 $(-\infty, 0]$ 上積分。$r \le 1$ 的部分包含無限多圈，而不是一圈。

STEP 之外

等角螺線（equiangular spiral，又稱對數螺線）是自相似（self-similar）的：把它按 $e^{2\pi\cot\alpha}$ 或任何因子縮放，都會得到旋轉後的同一條曲線。用複數表示，它是 $z = e^{(\cot\alpha + i)\theta}$，即一條直線在指數映射下的像，而夾角恆定這一點，源於 exp 是保角（conformal）的（即它保持角度不變）。Needham 的 *Visual Complex Analysis* 把這幅圖像闡述得非常精彩。一條無限捲繞的曲線竟有有限的長度，這讓人初次嘗到大學第一門分析課程中會遇到的反常積分與收斂（convergence）問題的滋味。

詳細解答

STEP II 1993，第 6 題

純數學 • 矩陣代數的反例

> **STEP II 1993，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/93-s2-q6

本題在 *Beyond STEP* 第 7.1 節對矩陣不成立的法則中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch07/matrix-rules.html

關鍵思路

每一個斷言對實數而言都是成立的，所以這道題其實是在問：普通代數裏的哪些習慣，換成矩陣(matrix)之後仍然管用。有兩件事會失效：矩陣乘法不一定可交換(commute)，而且兩個非零矩陣的乘積可以是零矩陣。只要精心挑選一對非常簡單的矩陣，就能解決 (i) 和 (ii)；伴隨矩陣（adjugate matrix）解決 (iii)；而一個重「根」再加上一點冪零（nilpotent）的變化，就解決 (iv)。

解答

以下一律以 $\mathbf{I}$ 表示 2×2 單位矩陣（identity matrix），以 $\mathbf{0}$ 表示零矩陣。

第 (i) 部分

錯誤。取

$$\mathbf{A}=\begin{pmatrix}1&0\\0&0\end{pmatrix},\qquad \mathbf{B}=\begin{pmatrix}0&0\\1&0\end{pmatrix},$$

兩者皆非零。於是

$$\mathbf{AB}=\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix},\qquad \mathbf{BA}=\begin{pmatrix}0&0\\1&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\1&0\end{pmatrix}\neq\mathbf{0}.$$

想一想 怎樣才能找出這樣的反例（counterexample），而不是單靠瞎猜？

我們想要 $\mathbf{AB} = \mathbf{0}$，所以 $\mathbf{B}$ 的每一列都必須被 $\mathbf{A}$ 化為零。上面的矩陣 $\mathbf{A}$ 會把任何形如 $(0, y)^{\mathsf{T}}$ 的向量化為零，因此我們可以把 $\mathbf{B}$ 的各列都取成這個形式。然後只需檢查 $\mathbf{BA}$ 會不會碰巧也等於零；採用最顯而易見的選擇時，它並不等於零。

第 (ii) 部分

錯誤。小心展開，並保持每個乘積的次序，

$$(\mathbf{A} - \mathbf{B})(\mathbf{A} + \mathbf{B}) = \mathbf{A}^2 + \mathbf{AB} - \mathbf{BA} - \mathbf{B}^2,$$

所以該斷言恰好在 $\mathbf{AB} = \mathbf{BA}$ 時成立。第 (i) 部分的兩個矩陣不可交換，因此它們就是一個反例：在那裏 $\mathbf{AB} - \mathbf{BA} = -\mathbf{BA} \neq \mathbf{0}$。具體計算，

$$(\mathbf{A} - \mathbf{B})(\mathbf{A} + \mathbf{B}) = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \qquad \mathbf{A}^2 - \mathbf{B}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

其中用到 $\mathbf{A}^2 = \mathbf{A}$ 和 $\mathbf{B}^2 = \mathbf{0}$。

第 (iii) 部分

正確。我們證明兩個方向。

*若 $\mathbf{AX} = \mathbf{0}$ 有非零解，則 $\det \mathbf{A} = 0$。*我們證明其逆否命題（contrapositive）。假設 $\det \mathbf{A} \neq 0$。那麼 $\mathbf{A}^{-1}$ 存在，而由 $\mathbf{AX} = \mathbf{0}$ 可得 $\mathbf{X} = \mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0}$。所以唯一的解是零矩陣。

*若 $\det \mathbf{A} = 0$，則存在非零解。*設

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

（即 $\mathbf{X}$ 是 $\mathbf{A}$ 的伴隨矩陣）。乘出來，

$$\mathbf{AX} = \begin{pmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{pmatrix} = (\det \mathbf{A})\,\mathbf{I} = \mathbf{0}.$$

$\mathbf{X}$ 的元素就是 $\mathbf{A}$ 的元素重新排列，並改變了其中一些正負號，而 $\mathbf{A} \neq \mathbf{0}$，所以 $\mathbf{X} \neq \mathbf{0}$。證明至此完成。請留意這裏用到了 $\mathbf{A} \neq \mathbf{0}$ 這個假設：正是它保證伴隨矩陣非零。

第 (iv) 部分

錯誤。對數而言，二次方程（quadratic）最多有兩個根，但這個事實的證明用到「若 $pq = 0$，則 $p = 0$ 或 $q = 0$」，而第 (i) 部分剛剛說明了這一點對矩陣並不成立。因此，我們應

該找一個可以因式分解成完全平方的二次式。取 $\mathbf{A} = -2\mathbf{I}$ 和 $\mathbf{B} = \mathbf{I}$，兩者皆非零。由於 $\mathbf{I}$ 與任何矩陣都可交換，

$$\mathbf{X}^2 - 2\mathbf{X} + \mathbf{I} = (\mathbf{X} - \mathbf{I})^2,$$

所以我們需要 $(\mathbf{X} - \mathbf{I})^2 = \mathbf{0}$。任何滿足 $\mathbf{N}^2 = \mathbf{0}$ 的矩陣 $\mathbf{N}$ 都可以用作 $\mathbf{X} - \mathbf{I}$，而這樣的矩陣有無限多個。對每個實數 t，設

$$\mathbf{X}_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

於是

$$\mathbf{X}_t^2 - 2\mathbf{X}_t + \mathbf{I} = \begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 2t \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{0}.$$

每個 $\mathbf{X}_t$ 都非零，而不同的 t 值給出不同的矩陣。所以這個方程有無限多個解，而不是最多兩個。

想一想 $\mathbf{A} = -2\mathbf{I}$、$\mathbf{B} = \mathbf{I}$ 有甚麼特別之處嗎？

其實沒有。取 $\mathbf{A} = -2\lambda\mathbf{I}$ 和 $\mathbf{B} = \lambda^2\mathbf{I}$（$\lambda \neq 0$ 為任意數），則每個 $\mathbf{X} = \lambda\mathbf{I} + \mathbf{N}$（其中 $\mathbf{N}^2 = \mathbf{0}$）都是解。冪零矩陣本身就構成一個龐大的族：對所有實數 p、q，$\begin{pmatrix} pq & -p^2 \\ q^2 & -pq \end{pmatrix}$ 的平方都是 $\mathbf{0}$。關鍵只在於這個二次式可以分解成兩個矩陣的乘積，而這樣的乘積等於零的方式，遠比數的情況多。

驗算

親手把每個反例乘出來：在 (i) 中，$\mathbf{AB}$ 的每個元素都是 0，而 $\mathbf{BA}$ 的左下角有一個 1。恆等式 $(\mathbf{A} - \mathbf{B})(\mathbf{A} + \mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2 + \mathbf{AB} - \mathbf{BA}$ 已用 1000 對隨機整數矩陣測試過，而 $\mathbf{A}\operatorname{adj}\mathbf{A} = \mathbf{0}$ 且 $\operatorname{adj}\mathbf{A} \neq \mathbf{0}$ 亦已用 2000 個隨機奇異矩陣（singular matrix）測試過。在 (iv) 中，試取 $t = 3$：$\mathbf{X}_3^2 = \begin{pmatrix} 1 & 6 \\ 0 & 1 \end{pmatrix}$，且 $\mathbf{X}_3^2 - 2\mathbf{X}_3 + \mathbf{I} = \mathbf{0}$。

常見錯誤

- 展開 $(\mathbf{A} - \mathbf{B})(\mathbf{A} + \mathbf{B})$ 時，當作 $\mathbf{AB} = \mathbf{BA}$ 來處理。在 (ii) 中只寫「錯誤」而不給出具體反例，得分甚少：題目要求的正是反例。
- 在 (iii) 中，只證明了「當且僅當」（if and only if）的其中一個方向。
- 在 (iv) 中，用求根公式或因式分解去「解」這個二次方程，然後斷定 $\mathbf{X} = \mathbf{I}$ 是唯一的根。因式分解本身沒有問題，但 $\mathbf{P}^2 = \mathbf{0}$ 並不意味着 $\mathbf{P} = \mathbf{0}$。

STEP 之外

2×2 實矩陣構成一個帶有零因子的非交換環（*non-commutative ring with zero divisors*），而這道題的每一部分，都在考驗一條在域（field）中成立、在這樣的環中卻失效的定律。第 (iii) 部分是線性代數（linear algebra）一條核心定理的 2×2 情形：一個方陣有非平凡的零空間（null space），當且僅當它的行列式（determinant）等於零。像 (iv) 中那些滿足 $\mathbf{N}^2 = \mathbf{0}$ 的矩陣稱為冪零矩陣，是若爾當標準形（Jordan normal form）的關鍵組成部分。線性代數方面可參考 Strang 的 *Introduction to Linear Algebra*，環論（ring theory）方面可參考 Allenby 的 *Rings, Fields and Groups*。

詳細解答

STEP II 1993，第 7 題

純數學 • 模 5 的無窮遞降法

> **STEP II 1993，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/93-s2-q7

關鍵思路

等式右邊是 5 的倍數，所以改以模（modulo）5 來考慮，方程便化為一個小規模的有限驗算：平方數模 5 只可能是 0、1 或 4 $(\mathrm{mod}\ 5)$。一旦迫使 a 和 b 都是 5 的倍數，方程便迫使 c 也是 5 的倍數，而把各數全部除以 5 就得到一個更小的解。非零整數不可能無止境地被 5 整除，這正是費馬（Fermat）的無窮遞降法（infinite descent）。

解答

a 和 b 可被 5 整除

把模 5 的每個剩餘（residue）逐一平方：

$x\ (\mathrm{mod}\ 5)$	0	1	2	3	4
$x^2\ (\mathrm{mod}\ 5)$	0	1	4	4	1
$2x^2\ (\mathrm{mod}\ 5)$	0	2	3	3	2

所以 $b^2 \equiv 0, 1$ 或 4，而 $2a^2 \equiv 0, 2$ 或 $3\ (\mathrm{mod}\ 5)$。由於 $5c^2 \equiv 0\ (\mathrm{mod}\ 5)$，我們需要

$$2a^2 + b^2 \equiv 0 \pmod 5.$$

現在列出所有組合：

$2a^2 + b^2$	$b^2 \equiv 0$	$b^2 \equiv 1$	$b^2 \equiv 4$
$2a^2 \equiv 0$	**0**	1	4
$2a^2 \equiv 2$	2	3	1
$2a^2 \equiv 3$	3	4	2

表中唯一等於 0 的一格，就是 $2a^2 \equiv 0$ 且 $b^2 \equiv 0$ 的那一格。由第一個表可見，只有當 $x \equiv 0$ 時才有 $x^2 \equiv 0 \pmod 5$（換句話說，5 是質數（prime），所以 $5 \mid x^2$ 蘊含 $5 \mid x$）。因此 $a \equiv 0$ 且 $b \equiv 0 \pmod 5$：a 和 b 都可被 5 整除，這正是所要證明的。

唯一的解是 $a = b = c = 0$

設 $a = 5p$ 及 $b = 5q$，其中 p, q 為整數。代入後得

$$25(2p^2 + q^2) = 5c^2 \quad \Longrightarrow \quad c^2 = 5(2p^2 + q^2).$$

所以 $5 \mid c^2$，而由於 5 是質數，故 $5 \mid c$。設 $c = 5r$。於是 $25r^2 = 5(2p^2 + q^2)$，即

$$2p^2 + q^2 = 5r^2.$$

這又是原來的方程。所以若 (a, b, c) 是一個解，則三個數都可被 5 整除，而且 $(a/5,\, b/5,\, c/5)$ 也是一個整數解。

> **想一想** 為甚麼「可以一直除以 5」就迫使所有數都等於零？

重複運用這一步。從一個解 (a, b, c) 出發，經過 n 步之後，我們得到整數解 $(a/5^n,\, b/5^n,\, c/5^n)$，所以 5^n 整除 a、b 和 c 每一個，而且這對每一個 $n \geq 1$ 都成立。

然而，非零整數 m 只能被 5 整除有限次：若 $5^n \mid m$ 且 $m \neq 0$，則 $5^n \leq |m|$，而當 n 足夠大時這便不成立。所以 a、b、c 沒有一個可以是非零的。

同一個論證也可以寫成遞降的形式：假設存在 $c \neq 0$ 的解，並從中選取 $|c|$ 為正數而又盡可能小的一個。那麼 $(a/5, b/5, c/5)$ 也是一個解，並且 $0 < |c/5| < |c|$，矛盾。所以 $c = 0$，於是 $2a^2 + b^2 = 0$ 迫使 $a = b = 0$，因為兩項都是非負的。

因此，唯一的整數解是 $a = b = c = 0$。

驗算

- 逐一檢查全部 25 對剩餘 $(a, b) \bmod 5$：只有 $(0, 0)$ 使得 $2a^2 + b^2 \equiv 0$。
- 以電腦搜尋所有 $|a|, |b| \leq 300$，發現 $2a^2 + b^2 = 5c^2$ 除了 $a = b = c = 0$ 之外並無其他解。
- 試與 $a^2 + b^2 = 5c^2$ 比較，這個方程*確實*有非平凡解，例如 $(1, 2, 1)$：在這情況下 $1 + 4 \equiv 0 \pmod 5$，所以第一部分的剩餘論證不再成立，而這正是應有的結果。

常見錯誤

- 沒有說明理由，就由 $5 \mid a^2$ 推斷 $5 \mid a$。這個結論是對的，因為 5 是質數（或可由表得出），但必須寫明。
- 忘記證明 c 同樣可被 5 整除。缺少這一步，$(a/5, b/5, c/5)$ 就未必是整數解，遞降也就無從開始。
- 寫到「可以再除以 5」就停筆。必須說明為甚麼重複整除會迫使各數為零，證明才算完整。

STEP 之外

這就是費馬的**無窮遞降法**，同一方法可以證明 $\sqrt{2}$ 是無理數，以及 $x^4 + y^4 = z^2$ 沒有非平凡解。像 $2a^2 + b^2 = 5c^2$ 這類方程，在數論中有系統的研究：勒讓德（Legendre）定理以二次剩餘（quadratic residue）為工具，給出 $Ax^2 + By^2 = Cz^2$ 有非平凡整數解的確切條件。在本題中，障礙在於 -2 不是模 5 的平方數。Davenport 的 *The Higher Arithmetic* 一書涵蓋同餘（congruence）、二次剩餘和遞降法。

詳細解答

STEP II 1993，第 8 題

純數學 • 以數學歸納法證明算術幾何平均不等式

> **STEP II 1993，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/93-s2-q8

本題在 *Beyond STEP* 第 3.2 節倒過來的 *AM–GM* 中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch03/am-gm-backwards.html

關鍵思路

兩項的不等式（inequality）其實只是說一個平方必定非負。要由 2^m 項推到 2^{m+1} 項，可把整列數分成長度各為 2^m 的兩半，對每一半使用歸納假設（induction hypothesis），然後對兩個半列的平均數使用兩項的不等式。這樣只證明了項數是 2 的冪時的算術平均—幾何平均不等式（AM–GM）；最後一部分用一個巧妙的技巧（出自柯西（Cauchy））補上缺口：用這列數本身的平均數（mean）來填充這列數，這樣既不改變平均數，也不影響乘積的任何要緊之處。

解答

兩項的情況

展開得

$$\left(\frac{a_1+a_2}{2}\right)^2 - a_1a_2 = \frac{a_1^2+2a_1a_2+a_2^2-4a_1a_2}{4} = \left(\frac{a_1-a_2}{2}\right)^2 \geq 0,$$

所以 $a_1a_2 \leq \left(\frac{a_1+a_2}{2}\right)^2$，而且等號恰好在 $a_1 = a_2$ 時成立。留意這個結果對任何兩個正數都成立；下文我們會把它用在 a_1 和 a_2 以外的數上。

歸納法

設 $P(m)$ 為以下命題：$(*)$ 對每一列 2^m 個正數都成立。$m = 1$ 的情況就是剛剛證明的兩項不等式。

現在假設 $P(m)$ 成立，並取 2^{m+1} 個正數 $a_1, \ldots, a_{2^{m+1}}$。記 $N = 2^m$，把它們分成各有 N 項的兩半，其平均數分別為

$$X = \frac{a_1 + \cdots + a_N}{N}, \qquad Y = \frac{a_{N+1} + \cdots + a_{2N}}{N}.$$

對每一半應用 $P(m)$，得

$$a_1 \cdots a_N \le X^N \qquad 及 \qquad a_{N+1} \cdots a_{2N} \le Y^N.$$

這裏所有的量都是正數，所以可以把這兩個不等式相乘：

$$a_1 \cdots a_{2N} \le X^N Y^N = (XY)^N.$$

由於 X 和 Y 都是正數，兩項的不等式給出 $XY \le \left(\frac{X+Y}{2}\right)^2$。因為 $t \mapsto t^N$ 在 $t > 0$ 時是遞增（increasing）的，把兩邊同時取 N 次方後，不等號的方向不變：

$$(XY)^N \le \left(\frac{X+Y}{2}\right)^{2N}.$$

最後，$\dfrac{X+Y}{2} = \dfrac{(a_1 + \cdots + a_N) + (a_{N+1} + \cdots + a_{2N})}{2N}$ 正是全部 $2N = 2^{m+1}$ 個數的平均數。把以上各個不等式合起來，得

$$a_1 \cdots a_{2^{m+1}} \le \left(\frac{a_1 + \cdots + a_{2^{m+1}}}{2^{m+1}}\right)^{2^{m+1}},$$

這就是 $P(m+1)$。根據歸納法（induction），$(*)$ 對所有正整數 m 都成立。

想一想 像這樣把兩個不等式相乘，是否合理？

若 $0 < p \le q$ 且 $0 < r \le s$，則 $pr \le qr \le qs$，當中先後用到 r 和 q 是正數。沒有正數這個條件，結論便可能不成立（例如 $-3 \le 1$ 及 $-3 \le 1$，但 $9 > 1$），這就是條件 $a_i > 0$ 如此重要的原因。

填充至 2 的冪

題目印出的定義是 $b_1 = a_2$，這必定是 $b_1 = a_1$ 的誤印：我們取 $b_i = a_i$（$1 \le i \le n$）及 $b_i = A$（$n < i \le 2^m$）。每個 b_i 都是正數（因為 A 是若干正數的平均數）。它們的和是

$$b_1 + \cdots + b_{2^m} = (a_1 + \cdots + a_n) + (2^m - n)A = nA + (2^m - n)A = 2^m A,$$

所以各 b_i 的平均數恰好是 A。它們的乘積是 $a_1 \cdots a_n\, A^{2^m-n}$。因此對各 b_i 應用 $(*)$，便得

$$a_1 \cdots a_n\, A^{2^m-n} \le A^{2^m},$$

這正是所要證明的。

算術平均—幾何平均不等式

給定任何正整數 n，選取 m 使得 $2^m > n$（由於 $2^m \to \infty$，這是辦得到的），於是上一步適用。兩邊除以正數 A^{2^m-n}，得

$$a_1 \cdots a_n \le A^n.$$

兩邊都是正數，而 $t \mapsto t^{1/n}$ 在 $t > 0$ 上遞增，所以取 n 次方根，得

$$(a_1 \cdots a_n)^{1/n} \le A = \frac{a_1 + \cdots + a_n}{n},$$

這就是算術平均—幾何平均不等式。（假如 n 剛好是 2 的冪，我們也可以直接引用 $(*)$，但選取 $2^m > n$ 便毋須另外分情況處理。）

驗算

試取 $n = 3$，$a = (1,2,6)$：$A = 3$，而取 $m = 2$ 時，填充後的一列數是 $(1,2,6,3)$，其平均數是 3，乘積是 $36 \le 3^4 = 81$；除以 3 得 $12 \le 27 = A^3$，所以 $\sqrt[3]{12} \approx 2.29 \le 3$。此外，當所有 a_i 都相等時，所用到的每個不等式都變成等式，這與預期相符。用電腦以精確的有理數運算，對數千列隨機數檢驗 $m \le 6$ 時的 $(*)$ 以及 $n \le 40$ 時的最終不等式，並沒有發現任何反例。

常見錯誤

- 把歸納假設一次過用在全部 2^{m+1} 項上：歸納假設只涵蓋長度為 2^m 的數列，所以必須分成兩半。
- 忘記檢查填充後的一列數的平均數仍然是 A；這正是用 A 而不用其他數來填充的全部意義所在。
- 取 n 次方根或作除法時，沒有指出有關的量都是正數。

STEP 之外

這種「向前—向後」歸納法就是柯西對算術平均—幾何平均不等式的證明：先對 2 的冪證明，再回頭降到每一個 n。算術平均—幾何平均不等式也是琴生不等式（Jensen's inequal-

ity）應用於凹函數（concave function）ln 時的特殊情況，而一整族的冪平均（power-mean）不等式、赫爾德（Hölder）不等式和閔可夫斯基（Minkowski）不等式，都是由同樣的想法發展出來的；這些都是大學一年級分析學（analysis）的基本工具。Bartle 與 Sherbert 的 *Introduction to Real Analysis*（實分析入門）從一開始就運用這類不等式，而 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題）則提供更多 STEP 風格的練習。

詳細解答

STEP II 1993，第 9 題

純數學 • 複數的輻角

> **STEP II 1993，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/93-s2-q9

關鍵思路

按照 $0 \le \arg z < 2\pi$ 的約定，「虛部（imaginary part）為正」即是「輻角（argument）嚴格介乎 0 與 π 之間」。複數（complex number）相加就是向量（vector）相加，而若干個全部指向同一個開半平面（open half-plane）的向量，其和仍然留在該半平面內；只要旋轉一下，使 z_1 或 z_3 落在實軸上，題目的不等式就正正是這個事實。至於最後一部分，積的輻角等於各輻角之和（可相差 2π 的倍數），而剛證明的不等式把兩者之差嚴格困在 0 與 2π 之間。

解答

虛部為正

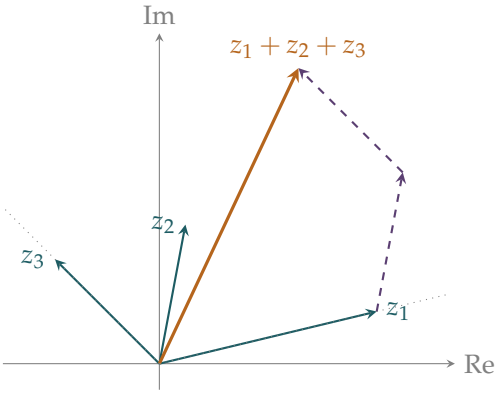
設 $z = r(\cos\theta + \mathrm{i}\sin\theta)$，其中 $r > 0$ 且 $0 \le \theta < 2\pi$。那麼 $\operatorname{Im} z = r\sin\theta > 0$ 當且僅當 $\sin\theta > 0$，而在 $[0, 2\pi)$ 上，這當且僅當

$$0 < \arg z < \pi.$$

所以 z 位於上半平面（upper half-plane），而它的輻角永不等於 0 或 π（否則 z 便是實數）。

圖示

設 $\theta_k = \arg z_k$，於是 $0 < \theta_1 < \theta_2 < \theta_3 < \pi$。把 z_1、z_2、z_3 畫成由原點出發的箭嘴；然後把它們首尾相接，便到達 $z_1 + z_2 + z_3$。



路徑的每一步相對於前一步都是逆時針轉向（因為 $\theta_1 < \theta_2 < \theta_3$），而且每一步都指向上半平面。路徑起初沿着角度為 θ_1 的射線前進，然後向它的左方彎曲，所以終點嚴格位於該射線的逆時針方向；從另一端回望，路徑的最後一步角度為 θ_3，而在此之前的一切都位於角度為 θ_3 的射線的順時針方向。所以和位於角度為 θ_1 與 θ_3 的兩條射線之間的楔形區域內。

要把這個想法嚴格寫出來，可乘以 $\mathrm{e}^{-\mathrm{i}\theta_1}$，即把所有東西順時針旋轉 θ_1：

$$\mathrm{e}^{-\mathrm{i}\theta_1}(z_1+z_2+z_3) = r_1 + r_2\mathrm{e}^{\mathrm{i}(\theta_2-\theta_1)} + r_3\mathrm{e}^{\mathrm{i}(\theta_3-\theta_1)},$$

其中 $r_k = |z_k| > 0$。第一項是正實數，其餘兩項的角度都在 $(0,\pi)$ 內，所以虛部為正。因此右邊的虛部為正，它的輻角 α 滿足 $0 < \alpha < \pi$。於是 $\theta_1 + \alpha$ 是 $z_1+z_2+z_3$ 的一個輻角，而由於 $\theta_1 < \theta_1+\alpha < \theta_1+\pi < 2\pi$，這正是落在 $[0,2\pi)$ 範圍內的那個輻角。所以 $\arg(z_1+z_2+z_3) > \theta_1$。

同樣地，乘以 $\mathrm{e}^{-\mathrm{i}\theta_3}$ 得 $r_1\mathrm{e}^{\mathrm{i}(\theta_1-\theta_3)} + r_2\mathrm{e}^{\mathrm{i}(\theta_2-\theta_3)} + r_3$，其中首兩個角度都在 $(-\pi,0)$ 內，所以這個數的虛部為負，可以寫成 $\rho\,\mathrm{e}^{-\mathrm{i}\beta}$，其中 $0 < \beta < \pi$。於是 $z_1+z_2+z_3 = \rho\,\mathrm{e}^{\mathrm{i}(\theta_3-\beta)}$，其中 $-\pi < \theta_3-\beta < \theta_3 < \pi$。和的虛部為正，所以它的輻角在 $(0,\pi)$ 內，並且與 $\theta_3-\beta$ 相差 2π 的倍數；由於兩個數都在長度為 2π 的區間 $(-\pi,\pi)$ 內，它們必定相等。所以 $\arg(z_1+z_2+z_3) = \theta_3-\beta < \theta_3$，綜合起來就是

$$\arg z_1 < \arg(z_1+z_2+z_3) < \arg z_3.$$

積與和

設 $\varphi = \arg(z_1+z_2+z_3)$，由上一部分可知 $\theta_1 < \varphi < \theta_3$。複數相乘時輻角相加，所以

$$z_1z_2z_3 = r_1r_2r_3\,\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2+\theta_3)},$$

而 $\arg(z_1z_2z_3)$ 就是 $[0,2\pi)$ 內與 $\theta_1+\theta_2+\theta_3$ 相差 2π 的倍數的那個數。

想一想 和 $\theta_1+\theta_2+\theta_3$ 可以是 $(0,3\pi)$ 內的任何數，所以我們不知道應該減去 2π 的哪一個倍數。怎樣才可以避開這個問題，根本不必知道？

我們不必知道：只需要知道，假如 $\arg(z_1z_2z_3)=\varphi$，那麼

$$D=\theta_1+\theta_2+\theta_3-\varphi$$

是 2π 的倍數。現在利用 $\theta_1<\varphi<\theta_3$ 來估計 D 的上下界：

$$D>\theta_1+\theta_2+\theta_3-\theta_3=\theta_1+\theta_2>0,$$
$$D<\theta_1+\theta_2+\theta_3-\theta_1=\theta_2+\theta_3<2\pi,$$

第一行用到 $\theta_k>0$，第二行用到 $\theta_k<\pi$。所以 $0<D<2\pi$，而嚴格介乎 0 與 2π 之間並沒有 2π 的倍數。因此 $\arg(z_1z_2z_3)$ 永不等於 $\arg(z_1+z_2+z_3)$。

驗算

以圖中的例子 $z_1=2.5+0.6\mathrm{i}$、$z_2=0.3+1.6\mathrm{i}$、$z_3=-1.2+1.2\mathrm{i}$ 來說，三個輻角約為 0.236、1.385 和 2.356，而 $\arg(z_1+z_2+z_3)=\arg(1.6+3.4\mathrm{i})\approx1.131$，確實介乎第一個與最後一個之間。以 200 000 組隨機選取的三個複數（模（modulus）介乎 e^{-4} 與 e^{4} 之間）作測試，兩條不等式和 $0<D<2\pi$ 每一次都成立，而積與和從未指向同一方向。

要明白上半平面這個條件為何重要，可試試 $z_1=z_2=z_3=\mathrm{e}^{2\pi\mathrm{i}/3}$，即把大小次序放寬至容許相等：積是 1，而和指向 $\mathrm{e}^{2\pi\mathrm{i}/3}$ 的方向，與 $D=4\pi/3$ 吻合。若讓三個角度都趨近 π，D 就會被推向 2π，可見 $\theta_k<\pi$ 這個上界確實發揮了作用。

常見錯誤

- 直接寫 $\arg(z_1z_2z_3)=\theta_1+\theta_2+\theta_3$：只要這個和等於或大於 2π，這就是錯的。我們只能說兩者在模 2π 的意義下相等。
- 聲稱 $\arg(z_1+z_2+z_3)$ 是各個 θ_k 的某種平均值。事實並非如此（模的大小有影響）；真正成立的只是它嚴格介乎 θ_1 與 θ_3 之間。

STEP 之外

第一個事實其實是說，開半平面(事實上任何角度小於 π 的楔形區域)是一個凸錐（*convex cone*）：其中的向量的正係數組合仍然留在其中。凸錐在線性規劃（linear programming）和最優化（optimisation）中佔有核心地位；而最後一部分要設法繞過的 arg 的多值性（multivaluedness），在複分析（complex analysis）的入門課程中會引出支割線（branch

cut）和複對數（complex logarithm）。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）以優美的幾何方式闡述了這兩個概念。

詳細解答

STEP II 1993，第 10 題

純數學 • 二次型與圓錐曲線

STEP II 1993，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/93-s2-q10

關鍵思路

一條關於 x 和 y 的二次方程之所以繁複，原因有二：一是 xy 交叉項（cross term），二是一次項。矩陣 $\mathbf{P}$ 處理前者，因為 $\mathbf{PAP}$ 是對角矩陣（diagonal matrix）。平移向量 $\mathbf{a}$ 處理後者：我們選取它，把原點移到圓錐曲線（conic）的中心，使所有關於 $\mathbf{x}_1$ 的一次項互相抵消。餘下的就是純粹的「平方和 = 常數」。

解答

對角化

首先，

$$\mathbf{AP} = \begin{pmatrix} -1 & 8 \\ 8 & 11 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 15 & -10 \\ 30 & 5 \end{pmatrix},$$

所以

$$\mathbf{PAP} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 15 & -10 \\ 30 & 5 \end{pmatrix} = \begin{pmatrix} 75 & 0 \\ 0 & -25 \end{pmatrix},$$

這是對角矩陣，正如所求。另外留意 $\mathbf{P}$ 是對稱（symmetric）矩陣，所以 $\mathbf{P}^{\mathrm{T}} = \mathbf{P}$。這一點稍後便會用到。

代入

設 $\mathbf{x} = \mathbf{P}\mathbf{x}_1 + \mathbf{a}$。則 $\mathbf{x}^{\mathrm{T}} = \mathbf{x}_1^{\mathrm{T}}\mathbf{P} + \mathbf{a}^{\mathrm{T}}$，而且

$$\begin{aligned}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} &= \mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{A}\mathbf{P}\mathbf{x}_1 + \mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{A}\mathbf{a} + \mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{P}\mathbf{x}_1 + \mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a} \\ &= \mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{A}\mathbf{P}\mathbf{x}_1 + 2\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{P}\mathbf{x}_1 + \mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a}.\end{aligned}$$

中間兩項相等，因為它們各自都是 1×1 矩陣，所以等於自身的轉置（transpose），而 $(\mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{A}\mathbf{a})^{\mathrm{T}} = \mathbf{a}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{x}_1 = \mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{P}\mathbf{x}_1$，這裏用到 $\mathbf{A}$ 和 $\mathbf{P}$ 都是對稱矩陣。此外 $\mathbf{b}^{\mathrm{T}}\mathbf{x} = \mathbf{b}^{\mathrm{T}}\mathbf{P}\mathbf{x}_1 + \mathbf{b}^{\mathrm{T}}\mathbf{a}$。於是方程變成

$$\mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{A}\mathbf{P}\mathbf{x}_1 + \left(2\mathbf{a}^{\mathrm{T}}\mathbf{A} + \mathbf{b}^{\mathrm{T}}\right)\mathbf{P}\mathbf{x}_1 + \left(\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a} + \mathbf{b}^{\mathrm{T}}\mathbf{a} - 11\right) = 0.$$

想一想 「一個合適的向量 $\mathbf{a}$」需要做到甚麼？

目標方程 $3x_1^2 - y_1^2 = c$ 沒有只含 x_1 或 y_1 的一次項，所以中間的括號必須為零。由於 $\mathbf{A}$ 是對稱矩陣，$2\mathbf{a}^{\mathrm{T}}\mathbf{A} + \mathbf{b}^{\mathrm{T}} = (2\mathbf{A}\mathbf{a} + \mathbf{b})^{\mathrm{T}}$，所以我們需要

$$\mathbf{A}\mathbf{a} = -\tfrac{1}{2}\mathbf{b} = \begin{pmatrix} -9 \\ -3 \end{pmatrix}.$$

這裏 $\det \mathbf{A} = -11 - 64 = -75 \neq 0$，所以恰好有一個解：

$$\mathbf{a} = -\frac{1}{75}\begin{pmatrix} 11 & -8 \\ -8 & -1 \end{pmatrix}\begin{pmatrix} -9 \\ -3 \end{pmatrix} = -\frac{1}{75}\begin{pmatrix} -75 \\ 75 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

（驗算：$\mathbf{A}\mathbf{a} = \begin{pmatrix} -1 - 8 \\ 8 - 11 \end{pmatrix} = \begin{pmatrix} -9 \\ -3 \end{pmatrix}$。）

常數項

取這個 $\mathbf{a}$，便有 $\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a} = \mathbf{a}^{\mathrm{T}}\left(-\tfrac{1}{2}\mathbf{b}\right) = -\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\mathbf{a}$，所以常數項是

$$\tfrac{1}{2}\mathbf{b}^{\mathrm{T}}\mathbf{a} - 11 = \tfrac{1}{2}(18 - 6) - 11 = -5.$$

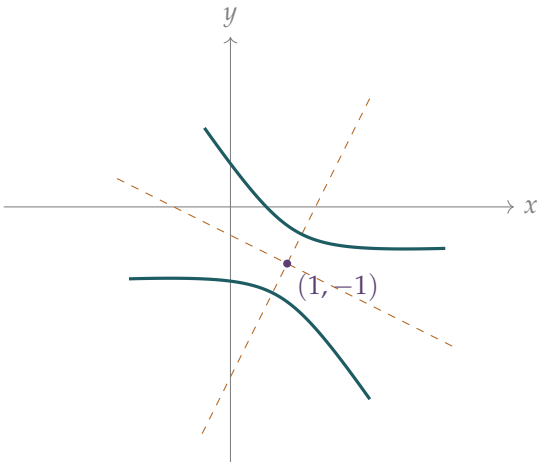
二次項是

$$\mathbf{x}_1^{\mathrm{T}}\begin{pmatrix} 75 & 0 \\ 0 & -25 \end{pmatrix}\mathbf{x}_1 = 75x_1^2 - 25y_1^2.$$

所以方程變成 $75x_1^2 - 25y_1^2 - 5 = 0$。兩邊除以 25，得

$$3x_1^2 - y_1^2 = \tfrac{1}{5}, \qquad \text{所以} c = \tfrac{1}{5}.$$

把代換逐項寫出來，就是 $x = x_1 + 2y_1 + 1$，$y = 2x_1 - y_1 - 1$。原方程是 $-x^2 + 16xy + 11y^2 + 18x + 6y - 11 = 0$，所以這條曲線是一條雙曲線（hyperbola），中心在 $(1, -1)$。$\mathbf{P}$ 的兩個列向量 $(1, 2)$ 和 $(2, -1)$ 互相垂直，所以它們給出雙曲線的對稱軸。



虛線是對稱軸，方向分別為 $(1,2)$ 和 $(2,-1)$。

> **想一想** 為甚麼 **P** 不是旋轉（rotation）？

它的兩個列向量互相垂直，但長度各為 $\sqrt{5}$，所以 $\mathbf{P}^2 = 5\mathbf{I}$，而不是 $\mathbf{I}$。這就是 **PAP** 的對角元素是 75 和 -25 的原因：它們是 **A** 的特徵值（eigenvalue）15 和 -5 的 5 倍。這個比例因子在這裏無關重要，因為它只是把二次部分乘以 5。

驗算

- **A** 的跡（trace）和行列式（determinant）分別是 10 和 -75，所以它的特徵值是 15 和 -5。這與 $75/5$ 和 $-25/5$ 吻合。
- 取 $x_1 = y_1 = 0$。則 $(x,y) = (1,-1)$，原方程左方為 $-1-16+11+18-6-11 = -5$，與 $75x_1^2 - 25y_1^2 - 5$ 一致。
- 取 $x_1 = 1$，$y_1 = 1$。則 $(x,y) = (4,0)$，而 $-16+72-11 = 45 = 75-25-5$。

我們亦在 1000 個隨機有理點上，精確地驗證了恆等式 左方 $= 25\left(3x_1^2 - y_1^2 - \frac{1}{5}\right)$。

常見錯誤

- 忘記交叉項 $\mathbf{x}_1^{\mathrm{T}}\mathbf{P}\mathbf{a}$ 和 $\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{P}\mathbf{x}_1$ 相等，因而漏掉因子 2，得出 $\mathbf{A}\mathbf{a} = -\mathbf{b}$。
- 停在 $75x_1^2 - 25y_1^2 = 5$ 便答 $c = 5$，沒有把全式除到題目要求的形式。

STEP 之外

這是主軸定理（principal axes theorem）的縮影。每個實對稱矩陣都可以用一個正交矩陣（orthogonal matrix）對角化（這裏是 $\mathbf{P}/\sqrt{5}$）。所以每條圓錐曲線，以及三維空間中的每個二次曲面（quadric surface），都可以透過一次旋轉（或反射）和一次平移化為標準形式。同樣的想法亦可藉黑塞矩陣（Hessian matrix）為多元函數的駐點（stationary point）分類。Strang 的 *Introduction to Linear Algebra*（線性代數入門教科書）有一章講對稱矩陣和正定性（positive definiteness），寫得很好。

詳細解答

STEP II 1993，第 11 題

力學 • 有二次阻力的運動

STEP II 1993，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：

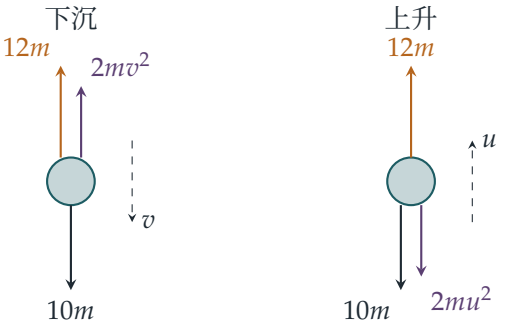
step.maths.org/questions/93-s2-q11

關鍵思路

物體一旦進入液體，浮力（buoyancy）便大於重力，因此除了與 v^2 成正比的阻力（resistance）之外，還有一個恆定的向上淨力。阻力永遠與運動方向相反，所以物體掉頭時阻力的方向亦隨之反轉，下沉和上升兩個階段須各自列出運動方程（equation of motion）。兩條方程都可以乾淨利落地分離變數。求時間時用 $\frac{\mathrm{d}v}{\mathrm{d}t}$，求距離時用 $v\frac{\mathrm{d}v}{\mathrm{d}x}$。

解答

我們把物體模擬為一個質點（particle）。在液體中，物體的重量為 $10m$ N，方向向下，而浮力為 $\frac{6}{5} \times 10m = 12m$ N，方向向上，故兩者合起來是 $2m$ N 的淨力，方向向上，不論物體向哪個方向運動都是如此。阻力 $2mv^2$ N 的方向則與速度相反。



下沉階段

取向下為正，設 x 為液面以下的深度，v 為向下的速率。由牛頓第二定律（Newton's second law）得

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = 10m - 12m - 2mv^2 \qquad \Longrightarrow \qquad \frac{\mathrm{d}v}{\mathrm{d}t} = -2(1+v^2).$$

設物體以速率 $V > 0$ 進入液體。分離變數（separating the variables），並由入水一刻（$t = 0$，$v = V$）積分至停止下沉的一刻（$t = T_1$，$v = 0$）：

$$\int_V^0 \frac{\mathrm{d}v}{1+v^2} = -2\int_0^{T_1} \mathrm{d}t \qquad \Longrightarrow \qquad T_1 = \tfrac{1}{2}\arctan V.$$

由於對任何有限的 V 都有 $\arctan V < \frac{\pi}{2}$，故 $T_1 < \frac{1}{4}\pi$：無論物體入水時有多快，它都會在入水後不足 $\frac{1}{4}\pi$ 秒內停止下沉。（此外，整個過程中 $\frac{\mathrm{d}v}{\mathrm{d}t} \le -2$，所以它必定會停下來。）

$V = 1$ 時到達的深度

此時 $T_1 = \frac{1}{2}\arctan 1 = \frac{\pi}{8}$。要求深度，可把加速度寫成 $v\frac{\mathrm{d}v}{\mathrm{d}x}$：

$$v\frac{\mathrm{d}v}{\mathrm{d}x} = -2(1+v^2) \qquad \Longrightarrow \qquad \int_1^0 \frac{v\,\mathrm{d}v}{1+v^2} = -2\int_0^D \mathrm{d}x,$$

所以 $-\frac{1}{2}\ln 2 = -2D$，最大深度為

$$D = \tfrac{1}{4}\ln 2 \text{ 米},$$

正如題目所求。

上升階段

在深度 D 處，物體瞬時靜止，但 $2m$ N 的向上淨力依然存在，所以物體開始上升。這時阻力的方向向下。取向上為正，設 y 為由最低點起計上升的高度，u 為向上的速率：

$$m\frac{\mathrm{d}u}{\mathrm{d}t} = 12m - 10m - 2mu^2 \qquad \Longrightarrow \qquad \frac{\mathrm{d}u}{\mathrm{d}t} = 2(1-u^2).$$

想一想 為甚麼不能沿用下沉階段的方程，只讓 v 變成負數？

不論 v 的正負，$-2mv^2$ 這一項永遠指向上方。物體一旦上升，阻力必須指向下方，所以方程也要改變。（若要用單一條方程，便須寫成 $-2mv|v|$。）還要留意，整個上升過程中 $u < 1$：u 由 0 開始，而只有當 $u < 1$ 時才有 $\frac{\mathrm{d}u}{\mathrm{d}t} > 0$，所以 u 只能趨近終端速率（terminal speed）1 ms^{-1}。這保證了下文的 $1-u^2$ 為正數。

回到液面時的速率。利用 $u\frac{\mathrm{d}u}{\mathrm{d}y} = 2(1-u^2)$，由 $y = 0$，$u = 0$ 積分至 $y = D$，$u = U$：

$$\int_0^U \frac{u\,\mathrm{d}u}{1-u^2} = 2\int_0^D \mathrm{d}y \qquad \Longrightarrow \qquad -\tfrac{1}{2}\ln(1-U^2) = 2D = \tfrac{1}{2}\ln 2.$$

所以 $1-U^2=\frac{1}{2}$，物體回到液面時的速率為 $U=\dfrac{1}{\sqrt{2}}\,\mathrm{ms}^{-1}$。

上升所需時間。利用 $\frac{\mathrm{d}u}{\mathrm{d}t}=2(1-u^2)$，並以部分分式（partial fractions）$\dfrac{1}{1-u^2}=\dfrac{1}{2}\left(\dfrac{1}{1+u}+\dfrac{1}{1-u}\right)$ 處理：

$$T_2=\frac{1}{2}\int_0^U\frac{\mathrm{d}u}{1-u^2}=\frac{1}{4}\Big[\ln\frac{1+u}{1-u}\Big]_0^U=\frac{1}{4}\ln\frac{1+\frac{1}{\sqrt{2}}}{1-\frac{1}{\sqrt{2}}}=\frac{1}{4}\ln\frac{\sqrt{2}+1}{\sqrt{2}-1},$$

最後一步是把分子和分母同乘以 $\sqrt{2}$。

總時間。物體在進入水池後

$$T_1+T_2=\frac{\pi}{8}+\frac{1}{4}\ln\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right)\ \text{秒}$$

再次到達液面，正如題目所求。

驗算

- 由於 $\dfrac{\sqrt{2}+1}{\sqrt{2}-1}=(\sqrt{2}+1)^2$，上升時間為 $T_2=\frac{1}{2}\ln(1+\sqrt{2})\approx 0.4407\,\mathrm{s}$，總時間 $\approx 0.3927+0.4407=0.8334\,\mathrm{s}$。
- 以數值方法逐步計算這個運動（龍格－庫塔法（Runge–Kutta），阻力寫成 $-2v|v|$），得出停止下沉的時間為 $0.39270\,\mathrm{s}$，深度為 $0.17329\,\mathrm{m}=\frac{1}{4}\ln 2$，回到液面的速率為 $0.70711\ \mathrm{ms}^{-1}$，回到液面的時間為 $0.83339\,\mathrm{s}$，全部吻合。
- 回到液面的速率小於入水速率，這是必然的：阻力在來回兩段路程上都作負功。

常見錯誤

- 物體回升時仍把阻力當作向上，這樣會得出 $2m+2mu^2$ 的向上淨力，上升也會快得多。
- 忘記浮力大於重力，以致把淨力（以向下為正）寫成 $+2m$，而非 $-2m$。
- 求深度時用 $\frac{\mathrm{d}v}{\mathrm{d}t}$，或求時間時用 $v\frac{\mathrm{d}v}{\mathrm{d}x}$，結果卡在積分上。

STEP 之外

二次阻力（quadratic drag）令運動方程變成非線性，但方程仍然是可分離變數的，這就是為何下沉時出現 arctan，而上升時出現 $\tanh^{-1}$（事實上 $T_2=\frac{1}{2}\tanh^{-1}U$）。上升時物

體趨近 1 ms^{-1} 的終端速率，此時阻力與向上淨力互相平衡。大學一年級的課程會有系統地處理這類模型，包括像 $\dot{u} = 2(1 - u^2)$ 這類方程的相線（phase line）圖像。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）討論取決於速度的力，而 Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程）則討論可分離變數方程及自治方程（autonomous equations）。

詳細解答

STEP II 1993，第 12 題

力學 • 摩擦力與剛體

STEP II 1993，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/93-s2-q12

關鍵思路

球體被卡在牆與斜面之間，所以它一旦移開，就必定與其中一方失去接觸。如果它在兩個接觸點同時滑動，球心便保持不動，球體只是繞着球心自轉。弄清楚它會向哪個方向轉動，就知道每個摩擦力（friction）的方向。之後只需處理三條方程（兩條分解力的方程和一條力矩方程），未知數為 N_1、N_2 和 m。至於第二部分，只要選對取力矩的點，一行就能得出答案。

解答

模型與摩擦力的方向

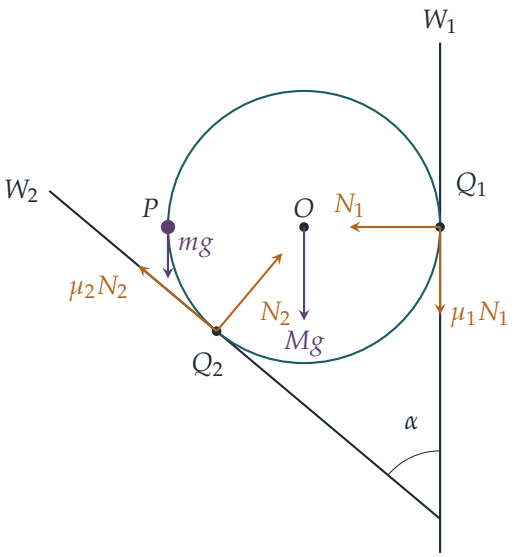
我們把球體視為均勻的剛體（rigid body）球，把質點（particle）視為固定在 P 的點質量。Q_1 和 Q_2 兩處的接觸面都是粗糙的。設法向反作用力（normal reaction）為 N_1（水平，指離牆壁）和 N_2（垂直於斜面，通過球心 O）。

由於 PQ_1 是水平直徑，質點位於與 O 水平距離為 r 之處，在遠離牆壁的一側。因此它的重量對 O 的力矩（moment）為 mgr，使球體朝 P 下降的方向轉動，而接觸點上的力必須與之平衡。

? **想一想** 每個接觸點上的摩擦力朝哪個方向？

如果球體在兩點都滑動，它仍與兩個面保持接觸，所以 O 固定不動，球體繞 O 朝 P 下降的方向轉動。於是與 P 相對的點 Q_1 沿牆壁向上移動，而 Q_2 則沿斜面向下移向牆角。摩擦力阻礙滑動，所以在 Q_1 處沿牆壁向下，在 Q_2 處沿斜面向上。由於兩點都即將滑動，兩處的摩擦力都達到極限（limiting）：$F_1 = \mu_1 N_1$ 和 $F_2 = \mu_2 N_2$。兩個摩擦力對 O

的力矩都與質點的力矩相反，這正是應有的情況。



兩個接觸點同時滑動

斜面與鉛垂線成角 α，所以它的法線與水平線成角 α。於是 N_2 的分量為指向牆壁的 $N_2\cos\alpha$ 和向上的 $N_2\sin\alpha$，而沿斜面向上的摩擦力 $\mu_2 N_2$ 的分量為指離牆壁的 $\mu_2 N_2\sin\alpha$ 和向上的 $\mu_2 N_2\cos\alpha$。

沿水平方向分解：

$$N_1 = N_2\cos\alpha - \mu_2 N_2\sin\alpha = N_2(\cos\alpha - \mu_2\sin\alpha). \tag{1}$$

沿鉛垂方向分解：

$$N_2\sin\alpha + \mu_2 N_2\cos\alpha - \mu_1 N_1 = (M+m)g. \tag{2}$$

對 O 取力矩：兩個法向反作用力都通過 O，Mg 也是。兩個摩擦力沿切線作用，與 O 的距離為 r。所以

$$mgr = \mu_1 N_1 r + \mu_2 N_2 r \quad\Longrightarrow\quad mg = \mu_1 N_1 + \mu_2 N_2. \tag{3}$$

把 (1) 代入 (3) 和 (2)：

$$mg = N_2\big(\mu_2 + \mu_1\cos\alpha - \mu_1\mu_2\sin\alpha\big),$$
$$(M+m)g = N_2\big(\sin\alpha + \mu_2\cos\alpha - \mu_1\cos\alpha + \mu_1\mu_2\sin\alpha\big).$$

從第二式減去第一式，得

$$\begin{aligned} Mg &= N_2\big(\sin\alpha + \mu_2\cos\alpha - 2\mu_1\cos\alpha + 2\mu_1\mu_2\sin\alpha - \mu_2\big) \\ &= N_2\big((2\mu_1\mu_2+1)\sin\alpha + (\mu_2 - 2\mu_1)\cos\alpha - \mu_2\big). \end{aligned}$$

把 mg 的方程除以這條方程，N_2 和 g 便會約去：

$$m = \frac{\mu_2 + \mu_1\cos\alpha - \mu_1\mu_2\sin\alpha}{(2\mu_1\mu_2+1)\sin\alpha + (\mu_2 - 2\mu_1)\cos\alpha - \mu_2}\,M,$$

這正是所要求的結果。

要令這個結果在物理上有意義，我們需要 $N_1 > 0$，由 (1) 可知這即是 $\mu_2 < \cot\alpha$。如果 μ_2 比這個值大，Q_2 處的摩擦力會把球體拉離牆壁，那就是最後一部分的情況。

繞 Q_2 滾動

現在球體在 Q_2 處不滑動，而是即將繞 Q_2 轉動（沿斜面向上滾動），所以球心會遠離牆壁，Q_1 處的接觸即將消失。在即將滾動的一刻，$N_1 = 0$，因此 Q_1 處的摩擦力也是零。

> ? **想一想** Q_2 處的摩擦力現在是未知的，我們該怎樣處理？

我們對 Q_2 取力矩，這樣 Q_2 處的反作用力和摩擦力都會消失，只剩下兩個重量。從 Q_2 起量度水平距離。由於 OQ_2 垂直於斜面，它與水平線成角 α，所以 O 與 Q_2 的水平距離為 $r\cos\alpha$，在靠牆的一側。質點在 O 之外再遠 r，朝遠離牆壁的方向，所以它與 Q_2 的水平距離為 $r - r\cos\alpha$，在另一側。要令力矩平衡，

$$Mgr\cos\alpha = mgr(1-\cos\alpha) \quad\Longrightarrow\quad m = \frac{M\cos\alpha}{1-\cos\alpha} = \frac{M}{\sec\alpha - 1},$$

其中最後一步是把分子和分母同除以 $\cos\alpha$。這就是所求的結果。

驗算

- 在兩種情況的分界上，即 $N_1 = 0$，亦即 $\mu_2 = \cot\alpha$ 時，兩個答案應該一致。把 $\mu_2 = \cos\alpha/\sin\alpha$ 代入第一條公式，不論 μ_1 是甚麼，分子都變成 $\cot\alpha$，分母都變成 $(1-\cos\alpha)/\sin\alpha$。所以 $m = M\cos\alpha/(1-\cos\alpha)$，與第二個答案吻合。
- 當 $\alpha = 60°$ 時，第二個答案給出 $m = M$。當 $\alpha = 60°$、$\mu_1 = 0.2$、$\mu_2 = 0.3$ 和 $M = 1$ 時，第一個答案給出 $m \approx 0.561$。我們用電腦就這些數值以及 2000 個隨機個案，檢查過力和力矩的平衡。
- 如果兩個接觸面都是光滑的（$\mu_1 = \mu_2 = 0$），第一條公式給出 $m = 0$：沒有任何東西能抵抗質點的力矩，一如所料。

常見錯誤

- 弄錯摩擦力的方向。先判斷球體會怎樣轉動，再判斷每個接觸點會向哪個方向滑動。

- 忘記兩個反作用力都通過球心，所以它們不會出現在對 O 的力矩方程中。
- 在最後一部分仍保留牆壁的摩擦力。球體一旦即將離開牆壁，$N_1 = 0$，所以該處的摩擦力是零。

STEP 之外

這類題目涉及多個可能滑動或滾動的接觸點，是帶約束（constraints）的剛體力學的起點。每個接觸點都給出一組不等式（$|F| \leq \mu N$、$N \geq 0$），而你要找出其中哪些是以等式成立的。同樣的分情況討論也出現在機械人學（robotics）中，用來判斷機械手能否抓穩物件。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）對靜力學（statics）和剛體的轉動有細緻的講解。

詳細解答

STEP II 1993，第 13 題

力學 • 簡諧運動

> **STEP II 1993，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/93-s2-q13

關鍵思路

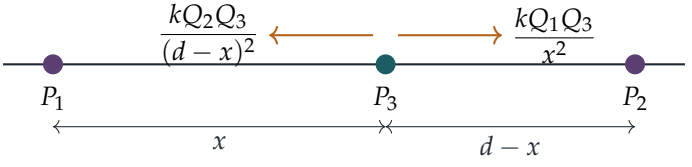
為運動中的小珠設定坐標，寫出它所受的淨力（net force）：兩個方向相反的平方反比（inverse-square）斥力。兩力互相抵銷之處就是平衡位置（equilibrium position）。至於振動，令小珠偏離平衡位置一個微小的 y，對兩個力都使用題目給出的近似式，常數項便會因平衡條件而互相抵銷，只剩下一個與 y 成正比的回復力（restoring force）。這正是簡諧運動（simple harmonic motion）的方程。

（題目說 P_3 在「P_1 與 P_3 之間」移動；這是試卷的筆誤，應為「P_1 與 P_2 之間」。）

解答

模型。各小珠均視為質點（particle）；金屬線光滑且水平，所以 P_3 的重量與法向反作用力（normal reaction）都垂直於金屬線，對沿線的運動毫無影響。沿金屬線方向的力，就只有作用於 P_3 的兩個電斥力（固定的小珠已被固定在原位，所以毋須理會它們所受的力）。

由 P_1 起向 P_2 方向量度 x，於是 P_3 與 P_1 的距離為 x，與 P_2 的距離為 $d-x$，其中 $0 < x < d$。



P_1 把 P_3 推向 P_2（正方向），而 P_2 則把它推回去，所以沿金屬線應用牛頓第二定律（New-

ton's second law），得

$$m\ddot{x} = \frac{kQ_1Q_3}{x^2} - \frac{kQ_2Q_3}{(d-x)^2}. \qquad (*)$$

平衡位置

在平衡位置 $x = a$，$(*)$ 的右邊等於零。由於 k 和 Q_3 均為正數，可以兩邊除以 kQ_3：

$$\frac{Q_1}{a^2} = \frac{Q_2}{(d-a)^2} \quad \Longrightarrow \quad \left(\frac{d-a}{a}\right)^2 = \frac{Q_2}{Q_1}.$$

a 和 $d-a$ 都是正數（小珠位於兩顆固定小珠之間），所以取正平方根：

$$\frac{d-a}{a} = \frac{\sqrt{Q_2}}{\sqrt{Q_1}} \quad \Longrightarrow \quad \sqrt{Q_1}\,(d-a) = \sqrt{Q_2}\,a \quad \Longrightarrow \quad a = \frac{d\sqrt{Q_1}}{\sqrt{Q_1}+\sqrt{Q_2}},$$

這正是所要求的結果。為方便起見，記 $S = \sqrt{Q_1} + \sqrt{Q_2}$ 及 $b = d - a$，於是

$$a = \frac{d\sqrt{Q_1}}{S}, \qquad b = \frac{d\sqrt{Q_2}}{S}.$$

留意小珠停在較接近較弱電荷的一邊，這是合理的。

微小振動

設 $x = a + y$，其中 y 很小，於是 $d - x = b - y$ 而 $\ddot{x} = \ddot{y}$。利用題目給出的近似式（以及把 y 換成 $-y$、a 換成 b 後的同一近似式），

$$\frac{1}{(a+y)^2} \approx \frac{1}{a^2} - \frac{2y}{a^3}, \qquad \frac{1}{(b-y)^2} \approx \frac{1}{b^2} + \frac{2y}{b^3}.$$

代入 $(*)$：

$$\begin{aligned} m\ddot{y} &\approx kQ_3\left[Q_1\left(\frac{1}{a^2} - \frac{2y}{a^3}\right) - Q_2\left(\frac{1}{b^2} + \frac{2y}{b^3}\right)\right] \\ &= kQ_3\left(\frac{Q_1}{a^2} - \frac{Q_2}{b^2}\right) - 2kQ_3\left(\frac{Q_1}{a^3} + \frac{Q_2}{b^3}\right)y. \end{aligned}$$

想一想 不含 y 的各項會怎樣？

它們正好就是平衡條件，所以 $\frac{Q_1}{a^2} - \frac{Q_2}{b^2} = 0$，這些項互相抵銷。在平衡位置附近展開時，這種情況必然出現：按定義，零階的力就是零。

現在利用 a 和 b 的表達式化簡 y 的係數：

$$\frac{Q_1}{a^3} = \frac{Q_1 S^3}{d^3 Q_1^{3/2}} = \frac{S^3}{d^3\sqrt{Q_1}}, \qquad \frac{Q_2}{b^3} = \frac{S^3}{d^3\sqrt{Q_2}},$$

所以

$$\frac{Q_1}{a^3}+\frac{Q_2}{b^3}=\frac{S^3}{d^3}\cdot\frac{\sqrt{Q_2}+\sqrt{Q_1}}{\sqrt{Q_1Q_2}}=\frac{S^4}{d^3\sqrt{Q_1Q_2}}.$$

因此

$$\ddot{y}\approx-\omega^2 y,\qquad \omega^2=\frac{2kQ_3S^4}{md^3\sqrt{Q_1Q_2}}>0.$$

這就是簡諧運動的方程，所以在位移很小時，P_3 圍繞 $x=a$ 作近似的簡諧運動。（若在 $y=y_0$ 處由靜止釋放，其運動為 $y\approx y_0\cos\omega t$。）週期（period）為

$$\begin{aligned}T&=\frac{2\pi}{\omega}=2\pi\sqrt{\frac{md^3\sqrt{Q_1Q_2}}{2kQ_3S^4}}=\frac{2\pi d}{S^2}\sqrt{\frac{md\sqrt{Q_1Q_2}}{2kQ_3}}\\&=\frac{\pi d}{S^2}\sqrt{\frac{4md\sqrt{Q_1Q_2}}{2kQ_3}}=\frac{\pi d}{(\sqrt{Q_1}+\sqrt{Q_2})^2}\sqrt{\frac{2md\sqrt{Q_1Q_2}}{kQ_3}},\end{aligned}$$

這就是所要求的結果。

驗算

- **對稱性：**當 $Q_1=Q_2$ 時，得 $a=d/2$，即中點，理應如此。週期公式對 Q_1 和 Q_2 是對稱的，這也是應有的性質（對調兩個電荷只是把整幅圖反射過來）。
- **量綱（dimensions）：**kQQ'/r^2 是力，所以 $kQ_3\sqrt{Q_1Q_2}$ 的量綱是力 $\times$ 長度^2。於是 $md\sqrt{Q_1Q_2}/(kQ_3)\cdot d^2/(\sqrt{Q_1}+\sqrt{Q_2})^4$ 的量綱為 質量 $\times$ 長度/力 $=$ 時間^2。沒有問題。
- **數值：**取 $Q_1=4$、$Q_2=1$、$d=3$ 及 $k=Q_3=m=1$，得 $a=2$（驗證：$4/2^2=1/1^2$）及 $T=2\pi/\sqrt{3}\approx 3.628$。對多組隨機選取的常數，由一個極小的位移出發以數值方法對 $(*)$ 積分，所得週期與公式吻合，誤差小於 10^4 分之一。

常見錯誤

- 只展開兩個力的其中一個：兩個力都取決於 y，而且兩者都對回復力有貢獻（符號相同）。
- 弄錯 $1/(b-y)^2$ 中 y 項的符號：在題目給出的近似式中把 y 換成 $-y$，得到的是 $+2y/b^3$。
- 在平衡條件中取負平方根，這樣小珠便會落在線段 P_1P_2 之外。

STEP 之外

我們所做的就是線性化（linearisation）：在勢能（potential）$V(x)$ 的穩定平衡點附近，運動近似為簡諧運動，其中 $\omega^2 = V''(a)/m$。這裏 $V = kQ_3\big(Q_1/x + Q_2/(d-x)\big)$，你可以驗證 $V''(a)$ 給出相同的 ω^2。把這個想法推廣到多個坐標，就會引出振動的簡正模式（normal modes），那是大學一年級動力學的核心課題。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）對平衡點附近的微小振動有清晰的講解。

詳細解答

STEP II 1993，第 14 題

力學 • 反彈與恢復係數

STEP II 1993，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/93-s2-q14

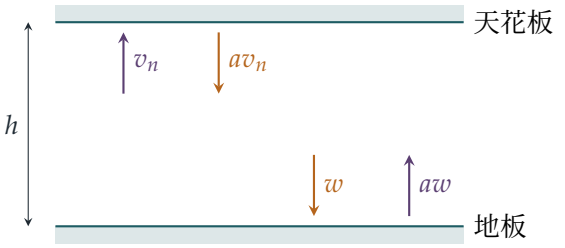
關鍵思路

在兩次碰撞之間，球只受重力作用而自由運動，因此 $v^2 = u^2 \pm 2gh$ 把它在地板和天花板處的速率聯繫起來，而每次碰撞都把速率乘以 a。追蹤一個完整循環（天花板、地板、再回到天花板），便得到球在天花板處的動能（kinetic energy）所滿足的線性遞推關係（linear recurrence），我們把它平移到不動點（fixed point）來求解。球能第 n 次到達天花板，當且僅當該能量為正數，再取對數便得出所求的上界。

解答

我們把球視為沿鉛垂方向運動的質點，不計空氣阻力，並取 $0 < a < 1$（上界的分母含有 $\ln a$，而且若 $a = 1$，球會永遠不停地撞擊天花板）。設 v_n 為球第 n 次撞擊天花板前一刻的速率，並設

$$E_n = \frac{v_n^2}{2gh}, \qquad \text{於是} \qquad \tfrac{1}{2}mv_n^2 = mgh\,E_n.$$



第一次撞擊天花板

球離開地板時 $u^2 = 2kgh$，並克服重力上升高度 h，所以

$$v_1^2 = 2kgh - 2gh = 2gh(k-1), \qquad E_1 = k - 1 > 0,$$

由於 $k > 1$，這是正數；因此球必定至少到達天花板一次。

一個完整循環

第 n 次撞擊天花板後的一刻，球以速率 av_n 向下運動（對固定的天花板應用牛頓實驗定律（Newton's experimental law））。它下落高度 h，因此以速率 w 到達地板，其中

$$w^2 = a^2 v_n^2 + 2gh.$$

它以速率 aw 向上反彈，並上升高度 h。若它能回到天花板，它在該處的速率滿足

$$v_{n+1}^2 = a^2 w^2 - 2gh = a^4 v_n^2 + 2gh\,(a^2 - 1).$$

除以 $2gh$，得

$$E_{n+1} = a^4 E_n + (a^2 - 1).$$

只要右邊為正數，此式便成立；若右邊為零或負數，球便再也回不到天花板。

> **想一想** 形如 $E_{n+1} = a^4 E_n + c$ 的遞推關係應怎樣求解？

先找出滿足 $E = a^4 E + (a^2 - 1)$ 的常數解 E：

$$E = \frac{a^2 - 1}{1 - a^4} = \frac{a^2 - 1}{(1 - a^2)(1 + a^2)} = -\frac{1}{1 + a^2}.$$

減去這個不動點便可消去常數項：設 $F_n = E_n + \dfrac{1}{1 + a^2}$，則 $F_{n+1} = a^4 F_n$。因此 $F_n = a^{4(n-1)} F_1$，而由於 $F_1 = (k-1) + \dfrac{1}{1 + a^2}$，

$$E_n = a^{4n-4}\left(k - 1 + \frac{1}{a^2 + 1}\right) - \frac{1}{a^2 + 1} = a^{4n-4}(k-1) + \frac{a^{4n-4} - 1}{a^2 + 1}.$$

乘以 mgh，可知第 n 次撞擊天花板前一刻的動能為

$$mgh\left(a^{4n-4}(k-1) + \frac{a^{4n-4} - 1}{a^2 + 1}\right),$$

正如所求。

撞擊天花板的次數

球只有在到達天花板時速率為正，才會第 n 次撞擊天花板，即必須有 $E_n > 0$。（若球第 n 次撞擊天花板，則之前每一次它也都撞擊了天花板，所以上面的公式適用。）把 $E_n > 0$ 乘以 $a^2 + 1 > 0$，

$$a^{4n-4}\big((k-1)(a^2+1)+1\big) > 1 \quad \Longleftrightarrow \quad a^{4n-4}\big(a^2(k-1)+k\big) > 1.$$

兩邊皆為正數，所以可以取對數：

$$(4n-4)\ln a + \ln\big[a^2(k-1)+k\big] > 0.$$

> **想一想** 除以 $4\ln a$ 時，不等號的方向應怎樣？

由於 $0 < a < 1$，我們有 $\ln a < 0$，所以除以 $4\ln a$ 時不等號要反轉：

$$n - 1 < -\frac{\ln\big[a^2(k-1)+k\big]}{4\ln a}, \qquad \text{即} \qquad n < 1 - \frac{\ln\big[a^2(k-1)+k\big]}{4\ln a}.$$

因此，若球撞擊天花板 N 次，則不等式在 $n = N$ 時成立，於是

$$N < 1 - \frac{\ln[a^2(k-1)+k]}{4\ln a}.$$

特別地，撞擊天花板的次數不會超過這個值，這正是我們要證明的。（事實上 E_n 隨 n 遞減，所以 N 恰好等於嚴格小於該上界的正整數的個數。）

驗算

- $n = 1$：公式給出 $mgh(k-1)$，與由能量守恆（energy conservation）得出的 $\frac{1}{2}m(2kgh) - mgh$ 吻合。
- n 很大時：$E_n \to -\frac{1}{1+a^2} < 0$，所以球最終確實無法再到達天花板；既然每次碰撞都損失能量，結果必然如此。
- 數值驗算：取 $k = 6$、$a = 0.9$、$h = 2\,\text{m}$、$m = 1\,\text{kg}$ 及 $g = 9.8\,\text{m s}^{-2}$，逐步模擬得出在天花板處的動能依次為 98、60.57、36.02 J，與公式相符。球撞擊天花板 6 次，而上界約為 6.48。

常見錯誤

- 忘記速率在天花板和地板兩處都要乘以 a，結果每個循環得出 a^2 而非 a^4。
- 除以 $\ln a$ 時沒有把不等號反轉。
- 把恢復系數（restitution）用在動能而非速率上：每次碰撞時能量是乘以 a^2，而不是乘以 a。

STEP 之外

遞推關係 $E_{n+1} = a^4 E_n + c$ 是一階線性差分方程（first-order linear difference equation）。減去不動點這個技巧，正是為線性常微分方程（ODE）求特別積分（particular integral）的離散版本，它在迭代映射（iterated maps）及其穩定性的研究中會再次出現。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教材）對碰撞和能量損失有更一般的論述。

詳細解答

STEP II 1993，第 15 題

概率與統計 • 中心極限定理

STEP II 1993，第 15 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/93-s2-q15

本題在 *Beyond STEP* 第 13.6 節 累積起來的誤差 中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch13/errors-add-up.html

關鍵思路

總誤差是許多微小而獨立的誤差之和，每個數各有一個。因此，先求出單獨一個數的誤差的平均值（mean）和方差（variance），再把所有數的結果加起來，然後利用中心極限定理（Central Limit Theorem）把總誤差視為近似服從正態分佈（normal distribution）。兩部電腦之間的全部差別都來自平均值：一部機器的誤差平均而言互相抵銷，另一部的誤差卻全都偏向同一方。

解答

一個數的誤差

設某數為 $a = m + U$，其中 $m = \lfloor a \rfloor$ 是整數，U 是它的小數部分（fractional part），且 $U \sim \mathrm{U}[0,1)$。我們把誤差定義為（近似值）−（真實值）；正負號的約定並不影響誤差的大小。

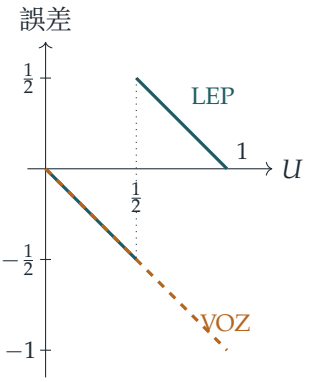
LEP（四捨五入，*rounding*）。若 $U < \frac{1}{2}$，該數會被換成 m，誤差為 $-U$；若 $U \geq \frac{1}{2}$，該數會被換成 $m+1$，誤差為 $1-U$。所以誤差為

$$X = \begin{cases} -U, & 0 \leq U < \frac{1}{2}, \\ 1-U, & \frac{1}{2} \leq U < 1, \end{cases}$$

它以恆定的密度取 $(-\frac{1}{2}, \frac{1}{2}]$ 內的每一個值：X 在 $[-\frac{1}{2}, \frac{1}{2}]$ 上服從均勻分佈（uniform dis-

tribution）（即圖中的鋸齒形）。

VOZ（截尾，*truncation*）。該數會被換成 m，所以誤差為 $Y = -U$，在 $(-1, 0]$ 上服從均勻分佈。



在長度為 1 的區間上的均勻分佈，不論區間位於何處，方差都是 $\frac{1}{12}$。所以

$$\mathrm{E}(X) = 0, \quad \mathrm{Var}(X) = \tfrac{1}{12}; \qquad \mathrm{E}(Y) = -\tfrac{1}{2}, \quad \mathrm{Var}(Y) = \tfrac{1}{12}.$$

（順帶記下：$\mathrm{Var}(U) = \int_0^1 u^2 \,\mathrm{d}u - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$。）

把 1500 個數相加

答案的誤差等於各個數的誤差之和，因為電腦把近似值相加時是完全準確的。各個小數部分互相獨立（independent），所以各個誤差亦互相獨立，它們的方差可以相加。對於 $n = 1500$ 個數，

$$\mathrm{Var}(\text{總誤差}) = \frac{1500}{12} = 125, \qquad \text{標準差} \sqrt{125} = 5\sqrt{5} \approx 11.18.$$

根據中心極限定理，1500 個獨立同分佈（independent, identically distributed）的誤差之和非常接近正態分佈。

LEP。總誤差 E_L 近似服從 $\mathrm{N}(0, 125)$，所以

$$\begin{aligned}\mathrm{P}(|E_L| > 15) &\approx \mathrm{P}\left(|Z| > \frac{15}{\sqrt{125}}\right) = \mathrm{P}(|Z| > 1.342) \\ &= 2\big(1 - \Phi(1.342)\big) \approx 2(1 - 0.9102) \approx 0.18.\end{aligned}$$

這裏毋須作連續性校正（continuity correction）：這些誤差是連續隨機變量（continuous random variable），而不是整數。

VOZ。總誤差 E_V 近似服從 $\mathrm{N}(-750, 125)$。只有當 $E_V \geq -15$ 時，誤差的大小才不超過 15，而這個值比平均值高出

$$\frac{-15 - (-750)}{\sqrt{125}} = \frac{735}{11.18} \approx 66$$

個標準差（standard deviation）。這個概率完全可以忽略不計，所以

$$\mathrm{P}(|E_V| > 15) \approx 1.$$

截尾永遠向下偏，平均每次偏 $\frac{1}{2}$，所以加了 1500 個數之後，VOZ 的答案大約相差 750。

想一想 兩部機器所產生的誤差方差相同。為甚麼答案卻相差這麼遠？

總誤差的離散程度是一樣的，大約是 ± 11，但它的中心卻不一樣。LEP 的誤差以 0 為中心，所以總誤差在 0 附近徘徊。VOZ 的誤差平均值全都是 $-\frac{1}{2}$，所以偏差（bias）與 n 成正比地增長，而隨機的離散程度只按 $\sqrt{n}$ 的速度增長。偏差輕易勝出。

LEP 可以做多少次加法？

假設 LEP 把 n 個數相加。它的總誤差近似服從 $\mathrm{N}\left(0, \frac{n}{12}\right)$，而我們需要

$$\mathrm{P}(|E_L| < 10) = 2\Phi\left(\frac{10}{\sqrt{n/12}}\right) - 1 \;\geq\; 0.9 \quad\Longleftrightarrow\quad \Phi\left(\frac{10}{\sqrt{n/12}}\right) \geq 0.95.$$

由於 Φ 是遞增函數，而且 $\Phi(1.6449) = 0.95$，上式即

$$\frac{10}{\sqrt{n/12}} \geq 1.6449 \iff n \leq 12\left(\frac{10}{1.6449}\right)^2 \approx 443.5.$$

這個概率隨 n 增加而下降，所以當 $n \leq 443$ 時它仍然不小於 0.9，到 $n = 444$ 時才首次跌破 0.9。因此 LEP 可以把 **443** 個數相加。（如果你把一次「加法」理解為使用一次 $+$ 號，那麼把 443 個數相加需要用 442 次；只要說明你採用哪一種理解，兩種都可以接受。）

驗算

- 1.342 這個數值合理嗎？總誤差 15 比一個標準差（11.18）略多，而正態分佈中約有 18% 的部分距離平均值超過 1.34 個標準差。這大致吻合。
- 數值檢驗。模擬 10^5 批、每批 1500 個四捨五入後的數，得出 $\mathrm{P}(|E_L| > 15) \approx 0.179$。按均勻分佈之和的精確分佈計算得 0.1797，而正態近似所得的亦是 0.1797。至於第二部分，精確概率在 $n = 443$ 時為 0.90019，在 $n = 444$ 時為 0.89981，印證了這個分界點。

常見錯誤

- 把 LEP 的誤差當作在 $[0, \frac{1}{2}]$ 上均勻分佈：四捨五入可以向上或向下，所以誤差是在 $[-\frac{1}{2}, \frac{1}{2}]$ 上均勻分佈。

- 忘記 VOZ 每個數的誤差平均值是 $-\frac{1}{2}$，因而得出與 LEP 相同的答案。
- 在最後一部分用了 $\Phi(z) = 0.9$ 而不是 0.95：事件 $|E| < 10$ 是雙側的，所以每邊尾部各佔 0.05。

STEP 之外

這是浮點運算（floating-point arithmetic）中捨入誤差（rounding error）的經典分析：無偏的捨入法令累積誤差按 $\sqrt{n}$ 的速度增長，而有偏的規則則令它按 n 的速度增長。這正是 IEEE 754 標準採用「逢半取偶」（round half to even）而不是逢半必進的原因。n 個均勻分佈隨機變量之和有一個精確的分佈（Irwin–Hall 分佈），觀察它變成正態分佈的速度有多快，是認識中心極限定理的一個好途徑；Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門教材）證明了這條定理，並對它加以討論。

詳細解答

STEP II 1993，第 16 題

概率與統計 • 泊松過程

STEP II 1993，第 16 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/93-s2-q16

關鍵思路

這裏有兩個互相獨立的隨機來源：巴士帶來一個幾何分佈（geometric distribution）的等候時間，而乘客則構成一個泊松過程（Poisson process）。先對第一班巴士到站的時間取條件（condition）。一旦知道這個時間，候車人數就只是一段已知長度時間內的泊松計數。第 (iii) 部分接着運用全概率定理（law of total probability），而真正的問題是：只憑統計表和計算機，怎樣把所得的無窮和計算至兩位有效數字。

解答

以 8:oo 之後的分鐘數來量度時間。設 T 為第一班巴士到站的時間，並設 X_t 為在 $(0,t]$ 內到達的乘客人數。乘客過程是速率為每分鐘 4 人的泊松過程，所以 $X_t \sim \mathrm{Po}(4t)$。巴士只可能在 $t = 1,2,3,\ldots$ 到站（「剛過八時」表示可能在 8:oo 到站的那班巴士已經來過並開走了），而且巴士與乘客互相獨立。

第 (i) 部分

第一班巴士在 8 時 n 分到站，當且僅當第 $1,2,\ldots,n-1$ 分鐘都沒有巴士到站，而第 n 分鐘有巴士到站。這些事件互相獨立，所以

$$\mathrm{P}(T = n) = \left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}, \qquad n = 1,2,3,\ldots$$

因此 T 服從參數為 $\frac{1}{4}$ 的幾何分佈。（驗算一下：這些概率之和為 $\frac{1}{4}\big/\left(1-\frac{3}{4}\right) = 1$。）

第 (ii) 部分

在 8:05 候車的人，正是在 $(0,5]$ 內到達的那些人。沒有人離開過，因為 8:05 之前沒有巴士到站。所以我們要求的是在 $T=5$ 的條件下 X_5 的分佈。

> **想一想** 知道第一班巴士在 8:05 到站，會否改變乘客人數的分佈？

不會。事件 $T=5$ 只取決於巴士，而巴士與乘客互相獨立。所以對 $T=5$ 取條件並不改變 X_5 的分佈，即 $X_5 \sim \mathrm{Po}(4\times 5)=\mathrm{Po}(20)$。因此

$$\mathrm{P}(m \text{ 人候車} \mid T=5)=\frac{\mathrm{e}^{-20}\,20^m}{m!},\qquad m=0,1,2,\ldots$$

第 (iii) 部分

巴士未能接載所有人，當且僅當候車人數超過 25，即 $X_T \geq 26$。設 F 為這個事件。對 T 取條件，並如 (ii) 般運用獨立性，

$$\mathrm{P}(F)=\sum_{n=1}^{\infty}\mathrm{P}(T=n)\,\mathrm{P}(X_n\geq 26)=\sum_{n=1}^{\infty}\left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}\,\mathrm{P}\big(\mathrm{Po}(4n)\geq 26\big).$$

現在我們需要一個方案，把這個無窮和化為有限的工作量，而誤差要遠低於 0.005。我們要的是兩位有效數字，而答案最後會是 0.1 這個數量級。

第 1 步：改為處理補事件，令尾項容易處理。 記 $w_n=\left(\tfrac{3}{4}\right)^{n-1}\tfrac{1}{4}$ 及 $q_n=\mathrm{P}(\mathrm{Po}(4n)\leq 25)$。由於 $\sum w_n=1$，

$$\mathrm{P}(F)=1-\sum_{n=1}^{\infty}w_n\,q_n.$$

這個形式較好，因為當 n 很大時 q_n 極小。當 $n\geq 11$ 時，平均值 $4n\geq 44$ 遠高於 25。由於 q_n 隨 n 增加而遞減（時間愈長，乘客愈多），

$$\sum_{n\geq 11}w_nq_n\leq q_{11}\sum_{n\geq 11}w_n=q_{11}\left(\tfrac{3}{4}\right)^{10}.$$

用正態近似（normal approximation）可得 $q_{11}\approx\Phi\big((25.5-44)/\sqrt{44}\big)=\Phi(-2.79)\approx 0.003$，所以這個尾項小於 0.0002，可以略去。這樣便只剩下十項。

第 2 步：最初幾項。 當 $n=1$ 和 $n=2$ 時（平均值為 4 和 8），有 26 名或以上乘客的可能性極低，在遠超所需的準確度之內 $q_1=q_2=1$，所以這兩項的貢獻是 $w_1+w_2=\frac{1}{4}+\frac{3}{16}$。

第 3 步：中間各項查表。 當 $n = 3, 4, 5$ 時（平均值為 12、16、20），直接從累積泊松分佈表讀出 $q_n = \mathrm{P}(\mathrm{Po}(4n) \le 25)$；這類表通常涵蓋至平均值 20。如果表中也列出平均值 24 和 28，則 $n = 6, 7$ 亦可查表。

第 4 步：較大的平均值用正態近似。 對於餘下直至 10 的 n（平均值最大為 40），使用 $\mathrm{Po}(\mu) \approx \mathrm{N}(\mu, \mu)$ 並作連續性校正（continuity correction）：

$$q_n \approx \Phi\left(\frac{25.5 - 4n}{\sqrt{4n}}\right),$$

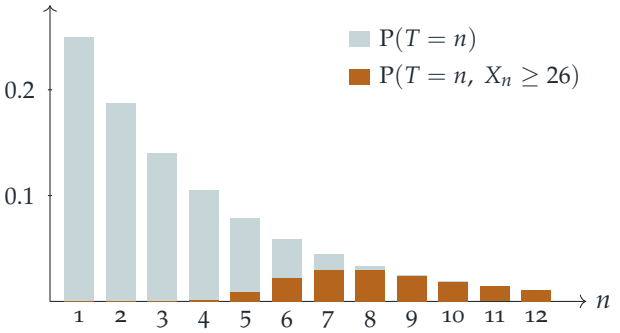
其中 Φ 的值從正態分佈表讀出。這些平均值已經夠大，近似效果良好。更重要的是，權重 w_n 本身已經很小（$w_8 \approx 0.033$，$w_{10} \approx 0.019$），所以 q_n 即使有千分之幾的誤差，對 $\mathrm{P}(F)$ 的影響也小於 0.001。

第 5 步：合併。 用計算機算出 $w_1, \ldots, w_{10}$，並計算

$$\mathrm{P}(F) \approx 1 - \sum_{n=1}^{10} w_n q_n,$$

然後捨入至兩位有效數字。每個被略去或作近似的部分最多約為 0.001，所以除非總和剛好非常接近捨入的分界，否則捨入是穩妥的。萬一真的如此，這個方法也清楚顯示應該改進哪幾項。

題目並沒有要求我們實際計算，但算出來的結果是 $\mathrm{P}(F) \approx 0.19$。下圖顯示這個概率來自何處：大部分來自第一班巴士在 8:06 至 8:10 到站的情況，這時預期的候車人數為 24 至 40，接近或超過載客量。



驗算

- 在 (i) 中，各概率之和為 1，而且 $\mathrm{P}(T = 1) = \frac{1}{4}$，與預期相符。
- 在 (ii) 中，候車人數的平均值為 20，即每分鐘四人、共五分鐘。

- 在 (iii) 中，精確的和（由電腦計算）為 0.1904。以蒙地卡羅模擬（Monte Carlo simulation）模擬 2×10^5 個早上，得 0.189。上述查表加正態近似的方法得 0.1912。三者捨入後都是 0.19。

常見錯誤

- 分界值差了一。只有當候車人數為 26 或以上時，巴士才未能接載所有人，所以我們需要的概率是 $1 - \mathrm{P}(\mathrm{Po}(4n) \leq 25)$，而不是 $1 - \mathrm{P}(\mathrm{Po}(4n) \leq 24)$。
- 把候車人數當作 $\mathrm{Po}(4\,\mathrm{E}[T]) = \mathrm{Po}(16)$。等候時間是隨機的，所以必須把泊松概率對 T 取平均，而不是把平均時間代入。
- 讓和式保持無窮，卻沒有說明為何可以截斷。你必須證明被略去的尾項小得不足以影響第二位有效數字。

STEP 之外

候車人數 X_T 是一個平均值本身隨機的泊松計數，即泊松分佈的一個混合（mixture）。由於 T 服從幾何分佈，X_T 的母函數（generating function）原來有一個簡潔的閉合形式；而像這樣的取條件論證，正是排隊論（queueing theory）的基礎，也是大學第一門概率課程中複合分佈（compound distribution）和混合分佈（mixed distribution）的基礎。怎樣截斷一個和式並估計誤差的上界，這個問題就是數值分析（numerical analysis）的起點。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）以合適的程度講解泊松過程和母函數。

詳細解答

STEP III 1993，第 1 題

純數學 • 弧長與體積

STEP III 1993，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/93-s3-q1

關鍵思路

二倍角公式（double-angle formula）$\cos 2\theta = 1 - 2\sin^2\theta$ 可把這條參數曲線（parametric curve）直接化為直角坐標（Cartesian）形式，而 θ 的取值範圍則清楚告訴我們所得的是拋物線（parabola）的哪一段。至於長度，$\sqrt{1+16x^2}$ 是採用雙曲代換（hyperbolic substitution）的典型信號，因為 $1+\sinh^2 u = \cosh^2 u$。只要留意到 P 的兩端都落在旋轉軸上，體積便只是一個普通的旋轉體體積（volume of revolution）。

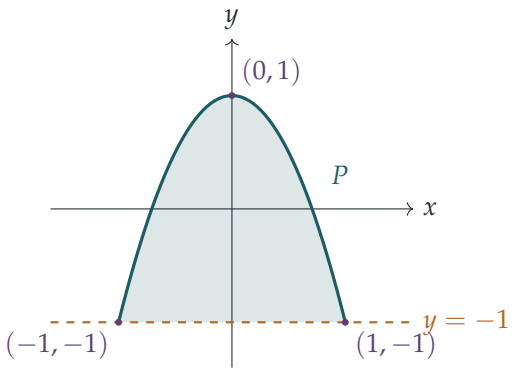
解答

曲線及其略圖

由於 $\cos 2\theta = 1 - 2\sin^2\theta$ 而 $x = \sin\theta$，

$$y = 1 - 2x^2.$$

因此 P 上每一點都在拋物線 $y = 1 - 2x^2$ 上。當 θ 由 $-\pi/2$ 變到 $\pi/2$ 時，$x = \sin\theta$ 由 -1 穩定地遞增至 1，每個值恰好取一次。所以 P 正是該拋物線上 $-1 \le x \le 1$ 的那段弧。它由 $(-1,-1)$ 上升至頂點（vertex）$(0,1)$，再下降至 $(1,-1)$，並且對稱於 y 軸。



P 的長度

以直角坐標形式計算，$\frac{\mathrm{d}y}{\mathrm{d}x} = -4x$，所以

$$\text{長度} = \int_{-1}^{1} \sqrt{1+16x^2}\,\mathrm{d}x = 2\int_{0}^{1} \sqrt{1+16x^2}\,\mathrm{d}x,$$

這裏用到被積函數（integrand）是偶函數（even function）這個事實。（若用參數形式的公式 $\int \sqrt{\dot{x}^2+\dot{y}^2}\,\mathrm{d}\theta$，代入 $x=\sin\theta$ 後會得到同一個積分。）

想一想 哪一個代換可以令 $\sqrt{1+16x^2}$ 化簡？

我們希望 $16x^2$ 成為一個 $\sinh^2$，所以設 $4x = \sinh u$。於是 $4\,\mathrm{d}x = \cosh u\,\mathrm{d}u$，而 $\sqrt{1+16x^2} = \cosh u$（取正值，因為 $\cosh u > 0$）。積分限 $x = 0, 1$ 變為 $u = 0$ 和 $u = \alpha$，其中 $\alpha = \sinh^{-1} 4$。注意 $\sinh\alpha = 4$ 及 $\cosh\alpha = \sqrt{1+16} = \sqrt{17}$。利用 $\cosh^2 u = \frac{1}{2}(1+\cosh 2u)$，

$$\begin{aligned} 2\int_0^1 \sqrt{1+16x^2}\,\mathrm{d}x &= 2\int_0^\alpha \cosh u \cdot \tfrac{1}{4}\cosh u\,\mathrm{d}u = \tfrac{1}{4}\int_0^\alpha (1+\cosh 2u)\,\mathrm{d}u \\ &= \tfrac{1}{4}\Big[u + \tfrac{1}{2}\sinh 2u\Big]_0^\alpha = \tfrac{1}{4}\big(\alpha + \sinh\alpha\cosh\alpha\big) \\ &= \tfrac{1}{4}\big(\sinh^{-1}4 + 4\sqrt{17}\big). \end{aligned}$$

所以 P 的長度是 $\sqrt{17} + \frac{1}{4}\sinh^{-1} 4$，正如題目所求。

體積

P 的兩端都在直線 $y = -1$ 上，所以當 P 繞這條直線旋轉時，會掃出一個封閉曲面。它所包圍的立體，正是把 P 與 $y = -1$ 之間的陰影區域旋轉所得的立體。在位置 x 處的切片是一個圓盤，其半徑為

$$y - (-1) = 2 - 2x^2 = 2(1-x^2)\ (\geq 0 \text{ 當} |x| \leq 1).$$

因此

$$V = \pi \int_{-1}^{1} 4(1-x^2)^2 \,\mathrm{d}x = 8\pi \int_{0}^{1} \left(1 - 2x^2 + x^4\right) \mathrm{d}x$$
$$= 8\pi\left(1 - \tfrac{2}{3} + \tfrac{1}{5}\right) = 8\pi \cdot \tfrac{8}{15} = \frac{64\pi}{15}.$$

驗算

- 當 $\theta = 0$ 時得 $(0,1)$，當 $\theta = \pm\pi/2$ 時得 $(\pm 1, -1)$，與 $y = 1 - 2x^2$ 吻合。
- 長度 $\sqrt{17} + \frac{1}{4}\sinh^{-1} 4 \approx 4.6468$ 應該略大於由 $(\pm 1, -1)$ 到 $(0,1)$ 的兩條直線弦的總長度，即 $2\sqrt{5} \approx 4.472$。事實確是如此。用中點法則（midpoint rule）在 $[-\pi/2, \pi/2]$ 上計算 $\int \sqrt{\cos^2\theta + 4\sin^2 2\theta}\,\mathrm{d}\theta$，同樣得到 4.64678。
- 帕普斯定理（Pappus's theorem）：該區域的面積為 $\int_{-1}^{1}(2 - 2x^2)\,\mathrm{d}x = \frac{8}{3}$，其形心（centroid）在旋轉軸上方 $\frac{4}{5}$ 處。所以 $V = 2\pi \cdot \frac{4}{5} \cdot \frac{8}{3} = \frac{64\pi}{15} \approx 13.404$，這也與數值積分的結果相符。

常見錯誤

- 誤把曲線繞 x 軸旋轉，用了半徑 y 而不是 $y + 1$。
- 忘記了 $\mathrm{d}x = \frac{1}{4}\cosh u\,\mathrm{d}u$ 中的因子 $\frac{1}{4}$，或者沒有把 $\sinh 2\alpha$ 化簡，未有寫成 $2\sinh\alpha\cosh\alpha = 8\sqrt{17}$。

STEP 之外

弧長（arc length）讓我們初次窺見曲線的微分幾何（differential geometry）。如果以弧長 s 來量度曲線上的位置，速率便恆為 1，而方向隨 s 轉動的快慢就是**曲率**（curvature）。對拋物線 $y = 1 - 2x^2$ 而言，曲率在頂點處最大。像上面那樣的雙曲代換，在大學一年級的微積分中隨處可見。Siklos 的 *Advanced Problems in Mathematics*（《高等數學難題》）收錄了更多這種風格的題目。

詳細解答

STEP III 1993，第 2 題

純數學 • 拐點與極坐標面積

STEP III 1993，第 2 題

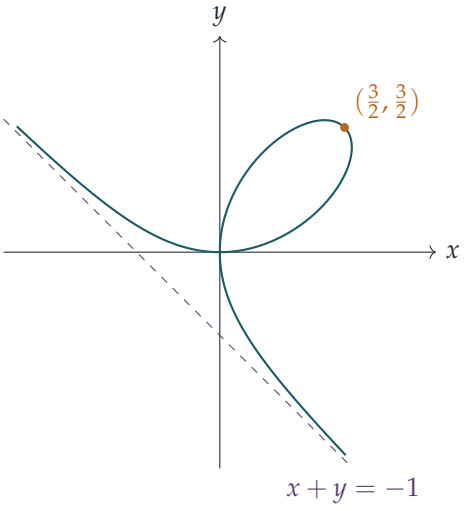
繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/93-s3-q2

關鍵思路

在拐點（point of inflection）處，二階導數必須為零（並且變號），所以 (i) 的計劃是用隱函數微分法（implicit differentiation）求出 $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$，再看它何時可以為零。代數運算看似繁重，但曲線本身的方程會把分子化簡成非常簡潔的式子。至於 (ii)，三次項與 xy 項的次數不同，所以改用極坐標（polar coordinates）後可以約去因子 r^2，而 r 就成為 θ 的顯函數。

解答



草圖顯示了這條曲線（笛卡兒葉形線，folium of Descartes）：第一象限內有一個環圈，

第二和第四象限內各有一條分支趨近漸近線（asymptote）$x+y=-1$。

第 (i) 部分

把 $x^3+y^3=3xy$ 對 x 作隱函數微分，

$$3x^2+3y^2\frac{\mathrm{d}y}{\mathrm{d}x}=3y+3x\frac{\mathrm{d}y}{\mathrm{d}x}, \qquad \text{所以} \qquad \frac{\mathrm{d}y}{\mathrm{d}x}=\frac{N}{D}, \quad N=y-x^2,\ D=y^2-x,$$

只要 $D\neq 0$ 即成立。對這個商求導，並利用 $\frac{\mathrm{d}N}{\mathrm{d}x}=y'-2x$ 和 $\frac{\mathrm{d}D}{\mathrm{d}x}=2yy'-1$，其中 $y'=N/D$，

$$\begin{aligned}\frac{\mathrm{d}^2y}{\mathrm{d}x^2}&=\frac{(y'-2x)D-N(2yy'-1)}{D^2}=\frac{(N-2xD)-\left(2yN^2/D-N\right)}{D^2}\\&=\frac{2\left(ND-xD^2-yN^2\right)}{D^3}.\end{aligned}$$

現在把括號展開：

$$\begin{aligned}ND&=x^3+y^3-x^2y^2-xy,\\xD^2&=xy^4-2x^2y^2+x^3,\\yN^2&=y^3-2x^2y^2+x^4y,\end{aligned}$$

所以

$$ND-xD^2-yN^2=3x^2y^2-xy-xy^4-x^4y=-xy\left(1+x^3+y^3-3xy\right).$$

在 C 上有 $x^3+y^3-3xy=0$，所以括號內的值就是 1，於是

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2}=\frac{-2xy}{(y^2-x)^3}.$$

想一想 這個式子在哪裏可能為零？公式失效的那些點又如何處理？

二階導數只有在 $x=0$ 或 $y=0$ 時才為零。在 C 上，$x=0$ 迫使 $y^3=0$，而 $y=0$ 迫使 $x^3=0$；無論哪種情況，我們都身處原點，而我們可以假設原點不是拐點。在原點和滿足 $y^2=x$ 的點以外，二階導數連續而且不為零，所以它不可能變號：這些地方沒有拐點。

餘下的是滿足 $D=y^2-x=0$ 的點，在這些點上 y 不是 x 的可微函數。假如 N 在那裏也是零，我們就有 $x=y^2$ 和 $y=x^2$，所以 $x=x^4$，得出 $(0,0)$ 或 $(1,1)$；但 $(1,1)$ 不在 C 上，因為 $1+1\neq 3$。因此在原點以外的每一個這樣的點上都有 $N\neq 0$：切線是鉛垂的，我們應該把 x 看成 y 的函數。C 的方程對 x 和 y 是對稱的，所以把上面運算中兩個字母的角色互換，便得

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

這個式子有定義（因為 $x^2 - y = -N \neq 0$），而且不為零（因為在原點以外 $xy \neq 0$）。所以曲線在這些點上同樣不會改變彎曲的方向。

因此 C 沒有拐點。

第 (ii) 部分

設 $x = r\cos\theta$，$y = r\sin\theta$。那麼

$$r^3(\cos^3\theta + \sin^3\theta) = 3r^2\cos\theta\sin\theta,$$

而當 $r \neq 0$ 時，我們可以除以 r^2：

$$r = \frac{3\cos\theta\sin\theta}{\cos^3\theta + \sin^3\theta}.$$

當 $0 \leq \theta \leq \frac{\pi}{2}$ 時，分子和分母都非負（而且分母為正），所以 $r \geq 0$，並且恰好在 $\theta = 0$ 和 $\theta = \frac{\pi}{2}$ 時 $r = 0$。因此當 θ 由 0 變到 $\frac{\pi}{2}$ 時，該點繞環圈走了一周，起點和終點都是原點。面積是

$$A = \frac{1}{2}\int_0^{\pi/2} r^2\,\mathrm{d}\theta = \frac{9}{2}\int_0^{\pi/2} \frac{\cos^2\theta\sin^2\theta}{(\cos^3\theta + \sin^3\theta)^2}\,\mathrm{d}\theta.$$

把分子和分母同時除以 $\cos^6\theta$（當 $0 \leq \theta < \frac{\pi}{2}$ 時為正），

$$A = \frac{9}{2}\int_0^{\pi/2} \frac{\tan^2\theta\sec^2\theta}{(1 + \tan^3\theta)^2}\,\mathrm{d}\theta.$$

代入 $u = \tan\theta$，$\mathrm{d}u = \sec^2\theta\,\mathrm{d}\theta$，其中 u 由 0 變到 ∞，

$$A = \frac{9}{2}\int_0^{\infty} \frac{u^2}{(1+u^3)^2}\,\mathrm{d}u = \frac{9}{2}\left[-\frac{1}{3(1+u^3)}\right]_0^{\infty} = \frac{9}{2}\cdot\frac{1}{3} = \frac{3}{2}.$$

這個廣義積分（improper integral）收斂，因為當 $u \to \infty$ 時原函數（antiderivative）趨向 0。所以環圈的面積是 $\frac{3}{2}$。

驗算

- 環圈的頂端位於 $y = x$ 之處：由 $2x^3 = 3x^2$ 得 $(\frac{3}{2}, \frac{3}{2})$，與原點的距離為 $\frac{3}{\sqrt{2}} \approx 2.12$。一個長度如此、闊度約為 1 的環圈，面積應該在 1 至 2 之間，與結果吻合。
- 利用參數式（parametrisation）$x = \frac{3t}{1+t^3}$，$y = \frac{3t^2}{1+t^3}$（直線 $y = tx$ 與 C 相交於該點），我們在隨機選取的點上比較了 $\frac{-2xy}{(y^2-x)^3}$ 與由參數式求得的二階導數，並檢查了曲率（curvature）在任何一條分支上都不變號。
- 對 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 用中點法則作數值積分，以及沿參數化的環圈用鞋帶公式（shoelace formula）求和，兩者都得出 1.500000。

常見錯誤

- 把二階導數完全展開，卻始終沒有留意到 $x^3 + y^3 - 3xy$ 就藏在其中；這一步化簡正是全題的關鍵。
- 忘記了切線鉛垂的那些點，在那裏 $\frac{\mathrm{d}^2y}{\mathrm{d}x^2}$ 是沒有定義，而不是等於零。
- 計算面積時把 $\frac{1}{2}\int r^2\,\mathrm{d}\theta$ 誤寫成 $\frac{1}{2}\int r\,\mathrm{d}\theta$。

STEP 之外

隱函數微分法的依據是隱函數定理（implicit function theorem）：在任何滿足 $\partial F/\partial y \neq 0$ 的點附近，方程 $F(x,y)=0$ 確實把 y 定義為 x 的光滑函數。在原點，兩個偏導數（partial derivative）都為零，那是一個奇點（singular point），有兩條分支在該處相交，這就是它需要特別處理的原因。葉形線也是一條有理曲線（rational curve），可以用通過其二重點（double point）的直線 $y = tx$ 來參數化，這個想法通向代數幾何（algebraic geometry）。至於這一切背後的分析學，可參閱 Bartle 與 Sherbert 的 *Introduction to Real Analysis*（實分析入門）。

詳細解答

STEP III 1993，第 3 題

純數學 • 矩陣的 LU 分解

STEP III 1993，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/93-s3-q3

關鍵思路

把 **M** 寫成一個下三角矩陣（lower-triangular matrix）乘以一個上三角矩陣（upper-triangular matrix）（即 *LU* 分解（LU decomposition）），整個問題就化為一連串只含一個未知數的方程。把 **AB** 乘開，再按合理的次序逐項對照：這樣每條新方程都只含一個未知數，既求出各個數值，也同時得出它們的唯一性（uniqueness）。三角矩陣的逆矩陣（inverse）亦容易求出，而 $\mathbf{Mx} = \mathbf{v}$ 則可拆成兩個容易求解的三角方程組。

先提一個警告。試卷上印的是 $e = 1$，但這是印刷錯誤：正確的值是 $e = -1$。若 $e = 1$，**AB** 最底一行就會是 (3, 10, 1)，而不是 (3, 8, 7)。以下一律採用 $e = -1$。

解答

求 A 和 B

乘開後得

$$\mathbf{AB} = \begin{pmatrix} a & ap & aq \\ b & bp + c & bq + cr \\ d & dp + e & dq + er + f \end{pmatrix}.$$

令上式等於 **M**，並按以下次序逐一處理各元素，使每條方程都只剩下一個未知數。

- 元素 $(1,1)$：$a = 1$。
- 元素 $(1,2)$、$(1,3)$：$ap = 3$ 及 $aq = 2$，所以 $p = 3$，$q = 2$（因為 $a = 1$）。
- 元素 $(2,1)$、$(3,1)$：$b = 4$，$d = 3$。
- 元素 $(2,2)$：$bp + c = 13$，所以 $c = 13 - 12 = 1$。

- 元素 $(2,3)$：$bq+cr=5$，所以 $8+r=5$，得 $r=-3$（用到 $c=1$）。
- 元素 $(3,2)$：$dp+e=8$，所以 $9+e=8$，得 $e=-1$。
- 元素 $(3,3)$：$dq+er+f=7$，所以 $6+3+f=7$，得 $f=-2$。

每個未知數都是由一條其餘各量均已知的方程迫出來的，而過程中只曾除以 $a=1$ 和 $c=1$，兩者皆不為零。因此，若 $\mathbf{M}=\mathbf{AB}$，各未知數就必須取這些值。反過來，這些值確實滿足全部九條方程。所以解存在而且唯一：

$$\mathbf{A}=\begin{pmatrix}1&0&0\\4&1&0\\3&-1&-2\end{pmatrix},\qquad \mathbf{B}=\begin{pmatrix}1&3&2\\0&1&-3\\0&0&1\end{pmatrix},$$

其中 $a=1$，$b=4$，$c=1$，$d=3$，$e=-1$，$f=-2$，$p=3$，$q=2$，$r=-3$。

> **想一想** 唯一性從何而來？我們只是找到了一個解而已。

唯一性來自推導的次序。整個過程中沒有任何一步需要作出選擇。每條方程都是關於單一個新未知數的線性方程，而且係數不為零，所以恰好只有一個解。因此，任何符合題目所給形式並滿足 $\mathbf{AB}=\mathbf{M}$ 的 $\mathbf{A},\mathbf{B}$，都必定是上面那一對。（$\mathbf{B}$ 對角線上的 1 十分重要：沒有它們，我們就可以把 $\mathbf{A}$ 放大、同時把 $\mathbf{B}$ 縮小。）

逆矩陣

要求 $\mathbf{A}$ 的逆矩陣，我們由上而下解 $\mathbf{Au}=\mathbf{v}$：

$$\begin{aligned}u_1&=v_1,\\ u_2&=v_2-4u_1=-4v_1+v_2,\\ -2u_3&=v_3-3u_1+u_2=-7v_1+v_2+v_3\implies u_3=\tfrac{1}{2}(7v_1-v_2-v_3).\end{aligned}$$

要求 $\mathbf{B}$ 的逆矩陣，我們由下而上解 $\mathbf{Bw}=\mathbf{u}$：

$$\begin{aligned}w_3&=u_3,\\ w_2&=u_2+3w_3=u_2+3u_3,\\ w_1&=u_1-3w_2-2w_3=u_1-3u_2-11u_3.\end{aligned}$$

讀出各係數，得

$$\mathbf{A}^{-1}=\begin{pmatrix}1&0&0\\-4&1&0\\\frac{7}{2}&-\frac{1}{2}&-\frac{1}{2}\end{pmatrix},\qquad \mathbf{B}^{-1}=\begin{pmatrix}1&-3&-11\\0&1&3\\0&0&1\end{pmatrix}.$$

乘開即可確認 $\mathbf{AA}^{-1}=\mathbf{I}$ 及 $\mathbf{BB}^{-1}=\mathbf{I}$。留意下（上）三角矩陣的逆矩陣仍然是下（上）三角矩陣。

聯立方程

這組聯立方程（simultaneous equations）就是 $\mathbf{M}\mathbf{x} = \mathbf{v}$，其中 $\mathbf{x} = (x, y, z)^{\mathrm{T}}$，$\mathbf{v} = (7, 18, 25)^{\mathrm{T}}$。由於 $\mathbf{M} = \mathbf{A}\mathbf{B}$，

$$\mathbf{x} = \mathbf{M}^{-1}\mathbf{v} = \mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{v}.$$

注意次序：$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$。首先，

$$\mathbf{u} = \mathbf{A}^{-1}\mathbf{v} = \begin{pmatrix} 7 \\ -28 + 18 \\ \frac{1}{2}(49 - 18 - 25) \end{pmatrix} = \begin{pmatrix} 7 \\ -10 \\ 3 \end{pmatrix},$$

然後

$$\mathbf{x} = \mathbf{B}^{-1}\mathbf{u} = \begin{pmatrix} 7 + 30 - 33 \\ -10 + 9 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}.$$

所以唯一解是

$$x = 4, \qquad y = -1, \qquad z = 3.$$

解之所以唯一，是因為 $\mathbf{M}$ 可逆（invertible）：$\det \mathbf{M} = \det \mathbf{A} \det \mathbf{B} = (1 \cdot 1 \cdot (-2))(1) = -2 \neq 0$。

驗算

代回原式：$4 - 3 + 6 = 7$，$16 - 13 + 15 = 18$，$12 - 8 + 21 = 25$。三條方程全部成立。我們亦以精確分數驗證了 $\mathbf{A}\mathbf{B} = \mathbf{M}$，以及 $\mathbf{B}^{-1}\mathbf{A}^{-1}$ 確實是 $\mathbf{M}^{-1}$；它最底一行是 $(\frac{7}{2}, -\frac{1}{2}, -\frac{1}{2})$。若採用試卷所印的 $e = 1$，乘積 $\mathbf{A}\mathbf{B}$ 最底一行是 $(3, 10, 1)$，由此可確認那是印刷錯誤。

常見錯誤

- 相信試卷所印的 $e = 1$，然後硬要 $(3, 2)$ 和 $(3, 3)$ 兩個元素吻合。務必把你求得的 $\mathbf{A}$ 和 $\mathbf{B}$ 重新相乘核對。
- 用了 $\mathbf{A}^{-1}\mathbf{B}^{-1}$ 而不是 $\mathbf{B}^{-1}\mathbf{A}^{-1}$。
- 只證明了這些值可行，卻沒有論證不可能有其他值。

STEP 之外

這就是 *LU* 分解，也是電腦實際求解線性方程組（linear systems）的方法。高斯消去法（Gaussian elimination）其實就是作出分解 $\mathbf{M} = \mathbf{LU}$ 的過程。分解完成後，每換一個新的右方向量，只需做一次前向代入（forward substitution）和一次回代（back substitution），毋須重新消元。一般而言，可能要先作行交換（row swap），這樣得到的是 $\mathbf{PM} = \mathbf{LU}$。Strang 的 *Introduction to Linear Algebra* 用了一整章圍繞這個想法展開。

詳細解答

STEP III 1993，第 4 題

純數學 • 級數求和

STEP III 1993，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/93-s3-q4

關鍵思路

每個級數（series）其實都是我們熟悉的級數，只是換了個樣子。求級數之和的慣常做法，是認出它是某個標準展開式（對數、等比級數（geometric series）或其導數（derivative））在某一點的取值，又或者把通項拆開，使各項在求和時逐項相消（telescope）。因此在每一部分，我們都要問：它可能是哪一個標準級數？要乘上甚麼、怎樣求導或怎樣重排，才能把它化成那個形式？

解答

第 (i) 部分

奇數分母 $1,3,5,\ldots$ 令人想到只剩下奇次冪的對數級數。當 $|x|<1$ 時，

$$\ln(1+x)=x-\tfrac{x^2}{2}+\tfrac{x^3}{3}-\cdots,\qquad \ln(1-x)=-x-\tfrac{x^2}{2}-\tfrac{x^3}{3}-\cdots,$$

兩式相減便消去偶次冪：

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

題目的級數是 $\sum x^{2n}/(2n+1)$，其中 $x=\frac{1}{2}$。這裏的冪是偶數次，所以我們除以 x（因為 $x\neq 0$，這樣做是容許的）：

$$\sum_{n=0}^{\infty}\frac{x^{2n}}{2n+1}=\frac{1}{2x}\ln\frac{1+x}{1-x}.$$

當 $x = \frac{1}{2}$ 時，此點位於收斂區間（interval of convergence）之內，$\frac{1+x}{1-x} = \frac{3/2}{1/2} = 3$ 而 $\frac{1}{2x} = 1$，所以

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

第 (ii) 部分

係數以 4 為週期重複：$x^{4k}, x^{4k+1}, x^{4k+2}, x^{4k+3}$ 的係數分別是 $2, -1, 0, -1$。於是我們把每四項歸為一組：

$$\sum_{k=0}^{\infty} x^{4k}\left(2 - x - x^3\right) = \frac{2 - x - x^3}{1 - x^4},$$

這是公比為 x^4 的等比級數，由於 $|x^4| < 1$，它是收斂的。這樣分組是合理的，因為原來的級數在 $|x| < 1$ 時絕對收斂（converges absolutely）：它各項的絕對值不超過 $2|x|^m$。

現在化簡。分子在 $x = 1$ 時等於零，所以 $1 - x$ 是它的因式：

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

約去 $1 - x$（因為 $|x| < 1$，它不等於零），

$$\frac{x^2 + x + 2}{(1 + x)(1 + x^2)} = \frac{A}{1 + x} + \frac{Bx + C}{1 + x^2}.$$

代入 $x = -1$ 得 $2 = 2A$，所以 $A = 1$。於是 $x^2 + x + 2 - (1 + x^2) = x + 1 = (Bx + C)(1 + x)$，所以 $B = 0$ 及 $C = 1$。因此

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

這個結果很容易驗證：$1 - x + x^2 - x^3 + x^4 - \cdots$ 加上 $1 - x^2 + x^4 - \cdots$，所得的係數正是 $2, -1, 0, -1, 2, \ldots$，符合要求。

第 (iii) 部分

把通項寫成

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

> ? **想一想** 我們以前在哪裏見過 $\sum r y^{r-1}$？

它是等比級數的導數。當 $|y| < 1$ 時，把 $\sum_{r=0}^{\infty} y^r = \frac{1}{1-y}$ 逐項求導（在收斂區間之內這樣做是成立的），得

$$\sum_{r=1}^{\infty} r y^{r-1} = \frac{1}{(1 - y)^2}.$$

取 $y=\frac{2}{3}$，其值為 $\frac{1}{(1/3)^2}=9$。我們的和由 $r=2$ 開始，所以要減去 $r=1$ 的一項，即 1：

$$\sum_{r=2}^{\infty}\frac{r\,2^{r-2}}{3^{r-1}}=\frac{1}{2}(9-1)=4.$$

第 (iv) 部分

由於 $r(r^2-1)=(r-1)r(r+1)$ 是三個連續整數之積，我們嘗試把它寫成兩個相鄰的「成對」分數之差：

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

因此求和時各項逐項相消。由 $r=2$ 加至 N，除了首尾兩項之外，其餘各項全部抵消：

$$\sum_{r=2}^{N}\frac{2}{r(r^2-1)}=\frac{1}{1\cdot 2}-\frac{1}{N(N+1)}.$$

當 $N\to\infty$ 時，最後一項趨向 0，所以

$$\sum_{r=2}^{\infty}\frac{2}{r(r^2-1)}=\frac{1}{2}.$$

驗算

- (i)：首四項之和為 $1+0.0833+0.0125+0.0022\approx 1.098$，已經很接近 $\ln 3\approx 1.0986$。
- (ii)：當 $x=0$ 時，兩邊都等於 2。當 $x=\frac{1}{2}$ 時，閉合形式（closed form）給出 $\frac{2}{3}+\frac{4}{5}=\frac{22}{15}$，與部分和（partial sums）的數值吻合。
- (iii)：最初幾項是 $\frac{2}{3},\frac{2}{3},\frac{16}{27},\ldots$；部分和緩緩上升，趨向 4。
- (iv)：加至 $N=4$ 的部分和是 $\frac{1}{3}+\frac{1}{12}+\frac{1}{30}=\frac{9}{20}=\frac{1}{2}-\frac{1}{20}$，與公式吻合。

常見錯誤

- 在 (i) 中忘記除以 x：$\sum x^{2n+1}/(2n+1)$ 在 $x=\frac{1}{2}$ 的值是 $\frac{1}{2}\ln 3$，而不是 $\ln 3$。
- 在 (iii) 中由 $r=1$ 開始求和，卻忘記減去原本沒有的第一項，結果得出 $\frac{9}{2}$。
- 在 (iv) 中誤用 $\frac{1}{r-1}-\frac{1}{r+1}$ 來相消：這樣拆開的是 $\frac{2}{r^2-1}$，而不是 $\frac{2}{r(r^2-1)}$。

STEP 之外

我們剛才不假思索便做了的幾步（在 (ii) 中把各項重新分組，以及在 (iii) 中把冪級數（power series）逐項求導），正是大學第一門分析（analysis）課程要嚴格論證的內容。冪

級數在其收斂半徑（radius of convergence）之內絕對收斂，並且可以逐項求導；但若一個級數只是條件收斂（converges conditionally），重排它的各項便可能改變它的和（黎曼（Riemann）重排定理）。Abbott 的 *Understanding Analysis* 對這兩點都有仔細的論述。

詳細解答

STEP III 1993，第 5 題

純數學 • 群與四元數

> **STEP III 1993，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/93-s3-q5

本題在 *Beyond STEP* 第 6.7 節四元數章中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch06/quaternions.html

關鍵思路

這個運算看似隨意，但第一部分的四個乘積顯示了它的結構：形如 $(z,0)$ 的數對相乘時與普通複數（complex number）無異，而第二個位置則像一個滿足 $j^2=-1$ 的新符號 j，凡被它越過的數都會變成共軛（conjugate）。有了這四個乘積，S_1 的八個元素就只是 $(u,0)$ 和 $(0,u)$，其中 $u\in\{1,i,-1,-i\}$。因此，封閉性（closure）、逆元（inverse）和階（order）都可由幾行一般的論證得出，毋須列出有 64 項的運算表。

解答

結合律與交換律

記 $z=(z_1,z_2)$、$w=(w_1,w_2)$、$u=(u_1,u_2)$。以下會一直用到 $(ab)^*=a^*b^*$、$(a\pm b)^*=a^*\pm b^*$ 和 $a^{**}=a$。首先，

$$z\circ w=(z_1w_1-z_2w_2^*,\ z_1w_2+z_2w_1^*),$$

所以 $(z\circ w)\circ u$ 的兩個分量是

$$(z_1w_1-z_2w_2^*)u_1-(z_1w_2+z_2w_1^*)u_2^*=z_1w_1u_1-z_1w_2u_2^*-z_2w_1^*u_2^*-z_2w_2^*u_1,$$
$$(z_1w_1-z_2w_2^*)u_2+(z_1w_2+z_2w_1^*)u_1^*=z_1w_1u_2+z_1w_2u_1^*+z_2w_1^*u_1^*-z_2w_2^*u_2.$$

另一方面，$w \circ u = (w_1u_1 - w_2u_2^*,\ w_1u_2 + w_2u_1^*)$，所以 $z \circ (w \circ u)$ 的兩個分量是

$$z_1(w_1u_1 - w_2u_2^*) - z_2(w_1u_2 + w_2u_1^*)^* = z_1w_1u_1 - z_1w_2u_2^* - z_2w_1^*u_2^* - z_2w_2^*u_1,$$
$$z_1(w_1u_2 + w_2u_1^*) + z_2(w_1u_1 - w_2u_2^*)^* = z_1w_1u_2 + z_1w_2u_1^* + z_2w_1^*u_1^* - z_2w_2^*u_2.$$

兩組右式逐項相同，所以 $\circ$ 滿足結合律（associative）。

它並不滿足交換律（commutative）。例如，

$$(0,1) \circ (i,0) = (0-0,\ 0+1\cdot i^*) = (0,-i), \qquad (i,0) \circ (0,1) = (0-0,\ i\cdot 1+0) = (0,i).$$

四個乘積

把零代入定義，得

$$(z,0) \circ (w,0) = (zw,\ 0), \qquad (z,0) \circ (0,w) = (0,\ zw),$$
$$(0,z) \circ (w,0) = (0,\ zw^*), \qquad (0,z) \circ (0,w) = (-zw^*,\ 0).$$

集合 S_1

設 $U = \{1, i, -1, -i\}$。那麼

$$S_1 = \{(u,0) : u \in U\} \cup \{(0,u) : u \in U\},$$

其中 A, C, E, G 屬第一類，B, D, F, H 屬第二類。

封閉性。集合 U 在乘法、取複共軛（$i^* = -i$）和取負之下都是封閉的。根據四個乘積，S_1 中兩個元素的乘積是 $(zw,0)$、$(0,zw)$、$(0,zw^*)$ 或 $(-zw^*,0)$，其中 $z, w \in U$。zw 和 $\pm zw^*$ 都屬於 U，所以每個乘積都屬於 S_1。

單位元（identity）。對任意 $(w_1,w_2) \in S$，

$$(1,0) \circ (w_1,w_2) = (w_1,\ w_2), \qquad (w_1,w_2) \circ (1,0) = (w_1,\ w_2 \cdot 1^*) = (w_1,w_2),$$

所以 $A = (1,0)$ 是雙邊單位元，而且它屬於 S_1。

逆元與階。對第一類元素，$(z,0) \circ (w,0) = (zw,0)$，所以它們的表現與 U 中的複數完全一樣。因此 A 的階是 1，逆元是它本身；$E = (-1,0)$ 的階是 2，逆元是它本身；$C = (i,0)$ 和 $G = (-i,0)$ 的階都是 4，並且互為逆元。

想一想 第二類元素的平方是甚麼？

根據四個乘積中的最後一個，對 $u \in U$，

$$(0,u) \circ (0,u) = (-uu^*,\ 0) = (-|u|^2,\ 0) = (-1,0) = E.$$

所以 B, D, F, H 的平方都是 E。由於 E 的階是 2，這四個元素的階都是 4。至於逆元，$(0,u)\circ(0,w) = (-uw^*, 0)$ 等於 $(1,0)$，當且僅當 $w^* = -1/u = -u^*$（用到 $uu^* = 1$），即 $w = -u$。再檢查另一個次序，$(0,-u)\circ(0,u) = (uu^*, 0) = (1,0)$ 同樣成立。所以 $(0,u)$ 的逆元是 $(0,-u)$。

總結如下：

元素	A	B	C	D	E	F	G	H
逆元	A	F	G	H	E	B	C	D
階	1	4	4	4	2	4	4	4

羣（group）。至此四條公理都已齊備。S_1 在 $\circ$ 之下封閉。這個運算在整個 S 上滿足結合律，在 S_1 上自然也滿足。$A \in S_1$ 是單位元，而且每個元素在 S_1 中都有逆元。所以 S_1 在 $\circ$ 之下是一個羣，階為 8。它不是阿貝爾羣（abelian group）：$B\circ C = H$，但 $C\circ B = D$。

子羣與矩陣

利用以上的乘積，$B = (0,1)$ 的各次冪是

$$B^2 = E, \qquad B^3 = E\circ B = (-1,0)\circ(0,1) = (0,-1) = F, \qquad B^4 = E\circ E = A.$$

所以 $\{A, B, E, F\} = \{B^0, B^1, B^2, B^3\}$ 是由 B 生成的循環子羣（cyclic subgroup）。它是封閉的，因為 $B^m \circ B^n = B^{m+n}$，而指數可以模 4 化簡；它也包含 A，以及每個 B^k 的逆元 B^{4-k}。所以它是一個階為 4 的子羣。

設 $M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$。那麼

$$M^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I, \qquad M^3 = -M, \qquad M^4 = I,$$

所以由 M 生成的羣是 $\{I, M, -I, -M\}$，同樣是階為 4 的循環羣。任何兩個同階的循環羣都是同構（isomorphic）的。這裏的映射

$$A \mapsto I, \quad B \mapsto M, \quad E \mapsto -I, \quad F \mapsto -M,$$

即 $B^k \mapsto M^k$，是一個雙射（bijection），並且滿足 $B^m \circ B^n = B^{m+n} \mapsto M^{m+n} = M^m M^n$。所以這兩個羣*確實*同構。

驗算

► 元素的階整除羣的階：1, 2, 4 這些階全都整除 8；對於子羣，2 和 4 都整除 4。

- 從複數的角度看：$(z,0)\circ(w,0) = (zw,0)$，所以 $C\circ C = (i^2,0) = E$，正好對應 $i^2=-1$。
- 用電腦作了簡短的核對：以 10 000 組隨機的複數對三元組驗證了結合律；建立了完整的 8×8 運算表，確認了封閉性、單位元、上述的逆元和階，以及 $B\circ C = H \neq D = C\circ B$；並確認了 $B^k \mapsto M^k$ 保持乘法。

常見錯誤

- 忘記取共軛：$(0,z)\circ(w,0)$ 是 $(0,zw^*)$，不是 $(0,zw)$。不可交換性正是由此而來。
- 只用一個數值例子來說明結合律成立。一個例子可以說明某條定律*不成立*，但要證明它成立，就必須作一般的展開。
- 因為一個羣由複數對組成、另一個由矩陣組成，就斷定兩者「不同構」。同構關乎乘法的結構，而不是元素本身是甚麼。

STEP 之外

這個運算就是哈密頓（Hamilton）的*四元數*（quaternion）乘法。數對 (z_1,z_2) 代表 $z_1+z_2 j$，其中 $j^2=-1$ 而 $jz = z^*j$。羣 S_1 就是*四元數羣*（quaternion group）$Q_8=\{\pm 1,\pm i,\pm j,\pm k\}$。它是每個子羣都是正規子羣（normal subgroup）的最小非阿貝爾羣，而它與另一個階為 8 的非阿貝爾羣（正方形的對稱羣）的分別，在於它只有一個階為 2 的元素。Allenby 的 *Rings, Fields and Groups*（環、域與羣）涵蓋了低階羣和四元數。

詳細解答

STEP III 1993，第 6 題

純數學 • 阿根圖上的軌跡

> **STEP III 1993，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/93-s3-q6

關鍵思路

對圓而言，$|z-k|=r$ 是最自然的描述；利用 $|u|^2=uu^*$ 把它平方，便得到一條只含 z 和 z^*、不帶平方根的方程。軌跡（locus）一旦寫成這種形式，像 $w=z/(z-1)$ 這樣的莫比烏斯映射（Möbius map）就很容易處理：先求逆，把 z 用 w 表示，然後代入、消去分母，再與一般形式比較，讀出新的圓心和半徑。

解答

圓的方程

P 在圓上，當且僅當它與 K 的距離為 r，即 $|z-k|=r$。兩邊平方，並利用 $|u|^2=uu^*$ 及 $(z-k)^*=z^*-k^*$，

$$r^2=(z-k)(z^*-k^*)=zz^*-kz^*-k^*z+kk^*,$$

所以 $zz^*-kz^*-k^*z+kk^*-r^2=0$，正是所求。

反過來，任何形如 $ww^*-cw^*-c^*w+cc^*-s^2=0$（其中 $s>0$）的方程都表示 $|w-c|^2=s^2$，所以它是圓心為 c、半徑為 s 的圓。下文會用這一點來辨認圓。

P 的軌跡

以 OA 為直徑的圓，圓心為 i（即 0 與 $2i$ 的中點），半徑為 1。取 $k=i$、$k^*=-i$、$kk^*=1$ 及 $r=1$，由第一部分得

$$zz^*+iz-iz^*=0. \qquad (*)$$

Q_1 的軌跡 L

留意 $z = 1$ 不在圓上（它與 i 的距離是 $\sqrt{2}$），所以對每一點 P，$w_1 = z/(z-1)$ 都有定義。把 w_1 記作 w。從 $w(z-1) = z$ 解出 z，得

$$z = \frac{w}{w-1}, \qquad z^* = \frac{w^*}{w^*-1},$$

而且 $w \neq 1$，因為 $w = 1$ 就要求 $z = z - 1$。代入 $(*)$：

$$\frac{ww^*}{(w-1)(w^*-1)} + \frac{iw}{w-1} - \frac{iw^*}{w^*-1} = 0.$$

各項乘以 $(w-1)(w^*-1)$（它不等於零），

$$\begin{aligned} 0 &= ww^* + iw(w^*-1) - iw^*(w-1) \\ &= ww^* + iww^* - iw - iww^* + iw^* \\ &= ww^* - iw + iw^*. \end{aligned}$$

所以 L 的方程是

$$w_1 w_1^* - iw_1 + iw_1^* = 0.$$

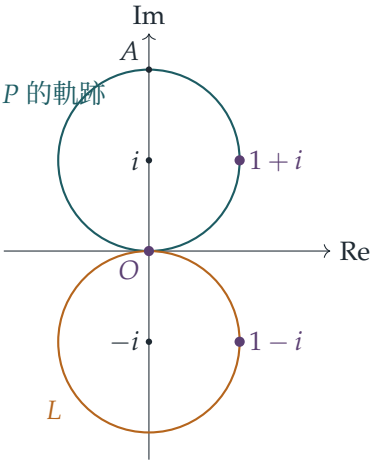
要辨認它，可與一般形式比較：我們需要 $-k^* = -i$ 及 $-k = i$，所以 $k = -i$；於是 $kk^* = 1$，而常數項 $kk^* - r^2$ 等於 0，所以 $r = 1$。因此 L 是圓心為 $-i$、半徑為 1 的圓，即 $|w_1 + i| = 1$。它是 P 所在的圓對實軸（real axis）的反射，而且同樣通過原點。

> **想一想** 這個圓上的每一點是否真的都能取到，還是只取到其中一部分？

以上每一步都可以逆推。若 $w \neq 1$ 在 $|w + i| = 1$ 上，則 $z = w/(w-1)$ 滿足 $(*)$，所以它是圓上的一點 P，並且 $w_1 = w$。唯一可能取不到的點是 $w = 1$，但它不在 L 上，因為 $|1 + i| = \sqrt{2}$。所以軌跡是整個圓。

Q_2 的軌跡

對 $|z - i| = 1$ 取共軛（conjugate），得 $|z^* + i| = 1$，因為 $|u^*| = |u|$ 而 $(z-i)^* = z^* + i$。所以 $w_2 = z^*$ 滿足 $|w_2 + i| = 1$。當 z 走遍以 i 為圓心的整個圓時，z^* 走遍以 $-i$ 為圓心的整個圓。換個說法，對 $(*)$ 取共軛得 $z^*z - iz^* + iz = 0$，這正是以 $w = z^*$ 代入的 L 的方程。所以 Q_2 的軌跡也是 L。



Q_1 與 Q_2 何時重合

我們需要 $\dfrac{z}{z-1} = z^*$，即

$$z = zz^* - z^*.$$

在圓上，由 $(*)$ 得 $zz^* = iz^* - iz$。代入，

$$z = iz^* - iz - z^* \quad \Longrightarrow \quad (1+i)z = (i-1)z^*.$$

其中一個解是 $z = 0$。否則 $z^* \neq 0$，而

$$\frac{z}{z^*} = \frac{i-1}{1+i} = \frac{(i-1)(1-i)}{(1+i)(1-i)} = \frac{2i}{2} = i.$$

設 $z = \rho e^{i\theta}$，其中 $\rho > 0$，上式即 $e^{2i\theta} = i$，所以 $\theta = \frac{\pi}{4}$ 或 $\frac{5\pi}{4}$：P 在直線 $y = x$ 上。這條直線與圓 $x^2 + (y-1)^2 = 1$（即 $x^2 + y^2 = 2y$）相交於 $2x^2 = 2x$ 之處，即 $x = 0$ 或 $x = 1$。非零的那一點是 $z = 1 + i$。

逐一驗證：$z = 0$ 給出 $w_1 = 0 = z^*$；而 $z = 1 + i$ 給出

$$w_1 = \frac{1+i}{i} = 1 - i = z^*.$$

所以 Q_1 與 Q_2 重合，當且僅當 P 是原點或點 $1 + i$。這正是 $y = x$ 與圓相交的兩點，而 $Q_1 = Q_2$ 的共同位置是 0 和 $1 - i$。

驗算

- 取一個樣本點：$P = 2i$（即點 A）給出 $w_1 = \frac{2i}{2i-1} = \frac{4-2i}{5}$，而 $|w_1 + i| = \left|\frac{4+3i}{5}\right| = 1$，所以它在 L 上。

- 數值上，我們在 $|z-i|=1$ 上隨機取了 10^4 個點，確認 $z/(z-1)$ 和 z^* 都滿足 $|w+i|=1$，誤差在 10^{-9} 以內。再沿圓周掃描 2×10^5 個點，尋找滿足 $\left|z/(z-1)-z^*\right|<10^{-3}$ 的點，結果只找到 $z=0$ 和 $z=1+i$。

常見錯誤

- 沒有先把 $z=0$ 分開處理就除以 z^*，因而漏掉 $P=O$ 這個解。
- 得出 $z/z^*=i$ 便就此停筆：這只確定了 P 的方向，你還要求出該直線與圓的交點。
- 把 $ww^*-iw+iw^*=0$ 與一般形式對照時，弄錯 k 的正負號。w^* 的係數是 $-k$，所以 $k=-i$，而不是 i。

STEP 之外

映射 $z\mapsto z/(z-1)$ 是一個莫比烏斯變換（Möbius transformation），這類映射總是把圓和直線映到圓和直線。這裏的像圓恰好是原來那個圓的鏡像，所以取共軛會得出同一個軌跡。取共軛本身會把定向（orientation）反轉，所以它不是莫比烏斯變換；但把它與莫比烏斯映射複合，便得到「反莫比烏斯」（anti-Möbius）映射，當中包括對圓的反射。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）以優美的幾何手法講解莫比烏斯映射、反演（inversion）和保圓性質。

詳細解答

STEP III 1993，第 7 題

純數學 • 雙曲函數聯立方程

> **STEP III 1993，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/93-s3-q7

關鍵思路

把兩個二倍角的 sinh 展開。第二條方程隨即直接以 y 表示 $\cosh x$，再用它去除第一條方程，便可同樣以 y 表示 $\sinh x$。利用恆等式（identity）$\cosh^2 x - \sinh^2 x = 1$ 可把 x 完全消去，剩下一條只含 $\sinh^2 y$ 的方程。之後要做的，就是找出三次方程（cubic）中合適的那個根，並以數值方法把它準確定出。

解答

消去 x

利用 $\sinh 2u = 2\sinh u \cosh u$，兩條方程變成

$$2\sinh x \cosh x = \cosh y, \qquad 2\sinh y \cosh y = 2\cosh x.$$

由第二條方程得

$$\cosh x = \sinh y \cosh y.$$

由於 $\cosh x \geq 1 > 0$，我們可以把第一條方程除以 $2\cosh x$：

$$\sinh x = \frac{\cosh y}{2\sinh y \cosh y} = \frac{1}{2\sinh y},$$

其中除以 $\cosh y$ 並無問題，因為 $\cosh y > 0$；而 $\sinh y \neq 0$，因為 $\sinh y \cosh y = \cosh x \neq 0$。

現在把兩者代入 $\cosh^2 x - \sinh^2 x = 1$。設 $t = \sinh^2 y$，則 $\cosh^2 y = 1 + t$：

$$t(1+t) - \frac{1}{4t} = 1.$$

兩邊乘以 $4t$（它不等於零），得 $4t^2 + 4t^3 - 1 = 4t$，即

$$4t^3 + 4t^2 - 4t - 1 = 0,$$

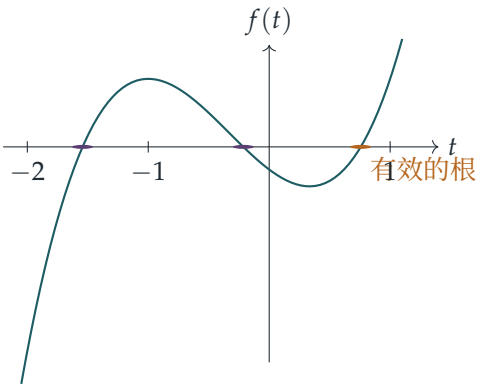
所以 $t = \sinh^2 y$ 是題目所給三次方程的一個根，正如所求。

最多只有一個有效解

設 $f(t) = 4t^3 + 4t^2 - 4t - 1$。由於 $t = \sinh^2 y$ 而 $\sinh y \neq 0$，有效的根必須滿足 $t > 0$。現在

$$f(-2) = -9, \quad f(-1) = 3, \quad f(0) = -1, \quad f(1) = 3,$$

所以根據介值定理（intermediate value theorem），這個三次方程在 $(-2, -1)$、$(-1, 0)$ 和 $(0, 1)$ 每個區間內各有一個根。三次方程最多有三個根，所以這些就是全部的根，而且恰好只有一個根是正數。



想一想 一個正的 t 給出兩個值 $\sinh y = \pm\sqrt{t}$。為甚麼只有一個能保留下來？

由 $\cosh x = \sinh y \cosh y$ 以及 $\cosh x > 0$、$\cosh y > 0$，可知必須有 $\sinh y > 0$。所以 $\sinh y = +\sqrt{t}$，而由於 sinh 是嚴格遞增（strictly increasing）的，這恰好給出一個 y，且 $y > 0$。因此，這個三次方程最多給出一個有效的 y（然後 $\sinh x = 1/(2\sinh y)$ 便唯一地確定 x，因為 sinh 是一對一（one-to-one）的）。

確定 t 的位置並計算其值

我們有

$$f(0.7) = 1.372 + 1.96 - 2.8 - 1 = -0.468 < 0,$$
$$f(0.8) = 2.048 + 2.56 - 3.2 - 1 = 0.408 > 0,$$

所以根據介值定理，正根位於 0.7 與 0.8 之間。

我們使用牛頓–拉弗森迭代法（Newton–Raphson iteration）

$$t_{n+1} = t_n - \frac{f(t_n)}{f'(t_n)}, \qquad f'(t) = 12t^2 + 8t - 4,$$

以中點 $t_0 = 0.75$ 為起始值。這裏 $f(0.75) = -0.0625$，$f'(0.75) = 8.75$，所以

$$\begin{aligned} t_1 &= 0.75 + \frac{0.0625}{8.75} = 0.7571429, \\ t_2 &= 0.7570685, \\ t_3 &= 0.75706846, \end{aligned}$$

而 t_4 與 t_3 在所示的各位數字上完全一致。為確認六位小數，我們檢查正負號的變化

$$f(0.7570680) < 0 < f(0.7570690),$$

所以根位於該區間內，而

$$t = 0.757068 \text{ （準確至六位小數）}.$$

（另一個較慢但完全可行的方法：把方程重新整理為 $t^2(4t+4) = 4t+1$，即 $t = \sqrt{(4t+1)/(4t+4)}$。在根的附近，右邊的導數（derivative）約為 0.16，所以這個迭代收斂（converge），每一步增加略少於一位準確的小數。）

求 y 和 x

取正平方根，$\sinh y = \sqrt{0.7570685} = 0.870097$，所以

$$y = \operatorname{arsinh}(0.870097) = \ln\bigl(0.870097 + \sqrt{1.757068}\bigr) = 0.786474,$$

並且

$$\sinh x = \frac{1}{2\sinh y} = 0.574649, \qquad x = \ln\bigl(0.574649 + \sqrt{1.330221}\bigr) = 0.546965.$$

代入原來的方程驗算：

$$\begin{aligned} \sinh 2x &= 1.325545 = \cosh y, \\ \sinh 2y &= 2.306704 = 2\cosh x, \end{aligned}$$

所以兩條方程都成立。準確至四位小數，

$$x = 0.5470, \qquad y = 0.7865.$$

驗算

- 三次方程的另外兩個根約為 -0.2140 和 -1.5431：兩者皆為負數，正如論證所要求。三個根的乘積是 $\frac{1}{4}$，這與 $0.757068 \times 0.2140 \times 1.5431 \approx 0.250$ 吻合。
- 直接在 $y \in [-5, 5]$ 上作數值搜尋，由第二條方程取 $x = \pm \operatorname{arcosh}(\frac{1}{2}\sinh 2y)$ 並檢驗第一條方程，結果只找到一個解，位於 $y \approx 0.7865$，且 $x > 0$。
- 若取負值 $\sinh y = -\sqrt{t}$，則 $\sinh y \cosh y < 0$，這絕不可能等於 $\cosh x$：這樣便可迅速確認它確實無效。

常見錯誤

- 過早平方（例如把原來的兩條方程都平方），之後卻忘記捨去不合適的正負號。
- 兩個迭代值一旦在六位小數上一致便停止迭代，而沒有檢查捨入區間兩端的正負號變化。
- 先把 t 捨入至六位小數，然後才取平方根和對數。應保留額外的位數，只在最後才把 x 和 y 捨入。

STEP 之外

這裏的兩種迭代收斂的速度相差甚遠。不動點迭代（fixed-point iteration）$t \mapsto g(t)$ 是線性收斂的，每一步誤差大約縮小至原來的 $|g'(\text{根})|$ 倍；而牛頓–拉弗森迭代法則是二次收斂（quadratic convergence）的，每一步正確的位數大約增加一倍。要把這一點說得嚴謹，需要用到均值定理（mean value theorem）和泰勒定理（Taylor's theorem），而這正是數值分析（numerical analysis）的起點。至於背後的分析，可參閱 Abbott 的 *Understanding Analysis*（一本實分析入門書）。

詳細解答

STEP III 1993，第 8 題

純數學 • 三維空間中的平面與直線

STEP III 1993，第 8 題

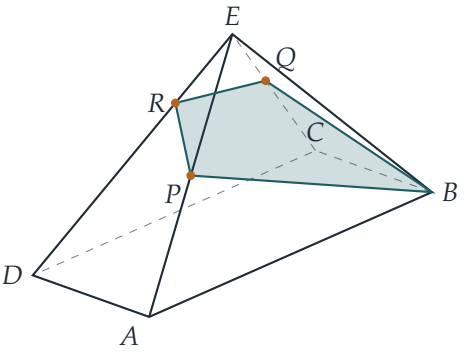
繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/93-s3-q8

關鍵思路

每條斜棱都是連接底面一個頂點與 E 的線段，所以斜棱上的每一點都有簡單的參數式（parametric form）。因此 P、Q 和 R 可以一下子全部寫出來，而通過 B、P 和 Q 的平面可以寫成一條線性方程，其係數由代入這幾點而得。至於第二部分，判斷平行四邊形（parallelogram）最簡潔的方法，是檢查兩條對角線是否互相平分（bisect）；只需看一個坐標，便足以說明這個條件不成立。

解答



求 γ

棱 AE 由所有形如 $(ta, 0, (1-t)a)$ 的點組成，其中 $0 \le t \le 1$，因為沿着這條棱有 $x+z=a$。同樣地，CE 是 $\{(-ta, 0, (1-t)a)\}$，而 DE 是 $\{(0, -ta, (1-t)a)\}$。所以

$$P = \big(\lambda a,\, 0,\, (1-\lambda)a\big), \qquad Q = \big(-\mu a,\, 0,\, (1-\mu)a\big), \qquad R = \big(0,\, -\gamma a,\, (1-\gamma)a\big).$$

我們設平面 BPQ 的方程為

$$px + qy + rz = a.$$

B、P、Q 三點不共線（collinear），所以它們確定唯一的一個平面。只要能找到滿足全部三個條件的 p, q, r，這條方程就是該平面的方程。逐點代入：

$$\begin{aligned} B:&\quad qa = a, && \text{所以} q = 1,\\ P:&\quad p\lambda + r(1-\lambda) = 1,\\ Q:&\quad -p\mu + r(1-\mu) = 1. \end{aligned}$$

要消去 p，可把 P 的方程乘以 μ，把 Q 的方程乘以 λ，然後相加：

$$r\big(\mu - \lambda\mu + \lambda - \lambda\mu\big) = \lambda + \mu, \qquad \text{所以} \qquad r = \frac{\lambda+\mu}{\lambda+\mu-2\lambda\mu}.$$

分母是 $\lambda(1-\mu) + \mu(1-\lambda) > 0$，所以這樣做沒有問題。（下文用不着，但順帶一提：$p = (\lambda-\mu)/(\lambda+\mu-2\lambda\mu)$。）

現在 R 位於這個平面上，所以

$$-\gamma + r(1-\gamma) = 1 \quad\Longrightarrow\quad \gamma(1+r) = r-1 \quad\Longrightarrow\quad \gamma = \frac{r-1}{r+1}.$$

把各項通分，公分母為 $\lambda+\mu-2\lambda\mu$：

$$r-1 = \frac{2\lambda\mu}{\lambda+\mu-2\lambda\mu}, \qquad r+1 = \frac{2(\lambda+\mu-\lambda\mu)}{\lambda+\mu-2\lambda\mu},$$

因此

$$\gamma = \frac{\lambda\mu}{\lambda+\mu-\lambda\mu},$$

這正是所求。

值得留意的是 $0 < \gamma < 1$，所以 R 確實位於棱 DE 上，並且嚴格地在 D 與 E 之間。它為正數是顯然的。至於 $\gamma < 1$，分母比分子大了 $\lambda+\mu-2\lambda\mu = \lambda(1-\mu) + \mu(1-\lambda) > 0$。

$BPRQ$ 不是平行四邊形

想一想 按給定次序排列的四個點是否構成平行四邊形，最快捷的判斷方法是甚麼？

四邊形 $BPRQ$ 是平行四邊形，當且僅當它的對角線 BR 與 PQ 互相平分，即

$$\overrightarrow{OB}+\overrightarrow{OR}=\overrightarrow{OP}+\overrightarrow{OQ}.$$

（這等價於 $\overrightarrow{BP}=\overrightarrow{QR}$，即有一組對邊相等且平行。）比較等式兩邊的 y 坐標：

$$\text{左邊：}a-\gamma a=(1-\gamma)a,\qquad \text{右邊：}0+0=0.$$

兩者必須相等，這就迫使 $\gamma=1$。但我們已證明 $\gamma<1$；事實上，$\gamma=1$ 會令 R 落在 E 上。所以兩條對角線永遠不會互相平分，$BPRQ$ 不可能是平行四邊形。 ■

為求完整，補充一點：四邊形的頂點確實是按 B,P,R,Q 的次序環繞排列的，因為平面依循環次序 EB,EA,ED,EC 與四條斜棱相交。即使把這幾點另行配對，也無濟於事。$B+P=Q+R$ 和 $B+Q=P+R$ 在 y 坐標上都需要 $a=-\gamma a$，而由於 $\gamma>0$，這是不可能的。

也可以從幾何角度看出同一結論：BP 位於面 ABE 上，而 QR 位於面 CDE 上。這兩個面相交於通過 E 且與 AB 平行的直線。分別位於這兩個平面上的兩條平行線，必定與這條公共直線平行。所以 $BP\parallel QR$ 會迫使 $BP\parallel AB$，而由於 $P\neq A$，這是不可能的。

驗算

- 對稱情況 $\lambda=\mu$：公式給出 $\gamma=\lambda/(2-\lambda)$。取 $\lambda=\mu=\frac{1}{2}$，結果是 $\frac{1}{3}$。直接計算：$p=0$，$r=2$，平面為 $y+2z=a$，它與 DE 相交於 $(0,-\frac{1}{3}a,\frac{2}{3}a)$，與預期相符。
- 極限 $\lambda,\mu\to1$：$P\to A$，$Q\to C$，平面變成底面，而 $\gamma\to1$，所以 $R\to D$，正如所料。
- 一個簡潔的改寫是 $\dfrac{1-\gamma}{\gamma}=\dfrac{1-\lambda}{\lambda}+\dfrac{1-\mu}{\mu}$，很容易用數值來檢驗。例如 $\lambda=\frac{1}{2}$，$\mu=\frac{2}{5}$ 給出 $\gamma=\frac{2}{7}\approx0.2857$。
- 我們隨機選取了 2000 組 a,λ,μ，以分數作精確計算，核實了公式、$0<\gamma<1$ 以及關於平行四邊形的結論。我們另外用獨立的方法求出 R，即利用叉積（cross product）求平面的法向量（normal）。

常見錯誤

- 把 AE 的參數式寫錯：在 AE 上，坐標滿足 $x+z=a$ 及 $y=0$，所以 P 的 z 坐標是 $(1-\lambda)a$，而不是 λa。

- 只在平面圖上用斜率來檢驗 $BP \parallel QR$。這個四邊形位於一個傾斜的平面上，所以應該用三維向量來處理。
- 忘記證明 $\gamma < 1$，而這正是平行四邊形的論證所需要的。

STEP 之外

求平面與直線的交點，是解線性方程組（linear system）最簡單的情況。由於 $EP : PA = (1-\lambda) : \lambda$，其他棱上的情況亦類似，所以結果說明

$$\frac{ER}{RD} = \frac{EP}{PA} + \frac{EQ}{QC},$$

這是一個不依賴坐標的命題，對任何通過 B 的平面都成立。因為它只用到直線上的比，所以對每一個以平行四邊形為底的棱錐（pyramid）都成立：任何這樣的棱錐都可以經仿射映射（affine map）變成本題的棱錐。Strang 的 *Introduction to Linear Algebra*（線性代數入門）對平面、法向量和線性方程組有完整而嚴謹的闡述。

詳細解答

STEP III 1993，第 9 題

純數學 • 平方和不等式

> **STEP III 1993，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/93-s3-q9

關鍵思路

這裏的每一條不等式（inequality），其實都是「平方和（sum of squares）非負」這句話的變裝。第 (i) 部分是最簡單的情形 $(a_1 - a_2)^2 \geq 0$，而其後各部分的做法，是把 $(a_i - a_j)^2$ 對所有數對求和，再數清楚每個平方項和每個乘積項各出現多少次。一旦得到一般形式 (iv)，把 $(\sum a_i)^2$ 展開成平方項加交叉項（cross terms），便可推出 (v)。

全文沿用以下記號：

$$S = \sum_{i=1}^{n} a_i^2, \qquad P = \sum_{i<j} a_i a_j,$$

其中第二個和式遍歷滿足 $1 \leq i < j \leq n$ 的 $\binom{n}{2}$ 個數對。

解答

第 (i) 部分

由於實數的平方非負，

$$0 \leq (a_1 - a_2)^2 = a_1^2 + a_2^2 - 2a_1 a_2,$$

所以 $a_1^2 + a_2^2 \geq 2a_1 a_2$，等號成立當且僅當 $a_1 = a_2$。

第 (ii) 部分

把 (i) 分別應用於三個數對，然後相加：

$$a_2^2 + a_3^2 \geq 2a_2a_3,$$
$$a_3^2 + a_1^2 \geq 2a_3a_1,$$
$$a_1^2 + a_2^2 \geq 2a_1a_2.$$

每個平方項恰好在其中兩行出現，所以相加後得

$$2(a_1^2 + a_2^2 + a_3^2) \geq 2(a_2a_3 + a_3a_1 + a_1a_2),$$

再除以 2 便得所求結果。換一個寫法，$a_1^2 + a_2^2 + a_3^2 - (a_2a_3 + a_3a_1 + a_1a_2) = \frac{1}{2}\left[(a_1 - a_2)^2 + (a_2 - a_3)^2 + (a_3 - a_1)^2\right] \geq 0$，等號成立當且僅當 $a_1 = a_2 = a_3$。

第 (iii) 部分

現在滿足 $1 \leq i < j \leq 4$ 的數對 (i, j) 共有六個。把 (i) 應用於每一個數對：

$$\sum_{1\leq i<j\leq 4} (a_i^2 + a_j^2) \;\geq\; 2 \sum_{1\leq i<j\leq 4} a_ia_j.$$

在左邊，a_1^2 出現了多少次？每一個含有下標 1 的數對都令它出現一次，也就是與其餘三個下標各配對一次。每個 a_k^2 的情況都相同，所以左邊等於 $3(a_1^2 + a_2^2 + a_3^2 + a_4^2)$，於是

$$3(a_1^2 + a_2^2 + a_3^2 + a_4^2) \geq 2(a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 + a_2a_4 + a_3a_4),$$

這正是所求。

第 (iv) 部分

(i) 至 (iii) 的規律是：左邊的係數比變數的個數少一，依次為 1、2（即第 (ii) 部分乘以 2）、3。

斷言。對任意 $n \geq 2$ 及實數 $a_1, \ldots, a_n$，

$$(n-1)\sum_{i=1}^{n} a_i^2 \;\geq\; 2 \sum_{1\leq i<j\leq n} a_ia_j, \qquad \text{即} \qquad (n-1)S \geq 2P,$$

等號成立當且僅當 $a_1 = a_2 = \cdots = a_n$。

> **想一想** 把 $a_i^2 + a_j^2$ 對每一個數對 $i < j$ 求和時，某個指定的 a_k^2 會出現多少次？

證明。考慮所有數對的差的平方之和：

$$D = \sum_{1\leq i<j\leq n} (a_i - a_j)^2 = \sum_{1\leq i<j\leq n} \left(a_i^2 + a_j^2\right) - 2 \sum_{1\leq i<j\leq n} a_ia_j.$$

下標 k 恰好屬於 $n-1$ 個數對（它與其餘 $n-1$ 個下標各配對一次），所以 a_k^2 在第一個和式中出現 $n-1$ 次。因此

$$D = (n-1)S - 2P.$$

D 的每一項都是實數的平方，所以 $D \geq 0$，由此得 $(n-1)S \geq 2P$。等號成立當且僅當每個 $(a_i - a_j)^2$ 都等於零，即當且僅當所有 a_i 都相等。 ■

取 $n = 2, 3, 4$，便分別重新得到 (i)、(ii) 的兩倍，以及 (iii)。

第 (v) 部分

這裏需要 $n \geq 2$，使得 $n - 1 > 0$。把和的平方展開，每個平方項出現一次，而每個 $i \neq j$ 的乘積 $a_i a_j$ 出現兩次（分別以 $a_i a_j$ 和 $a_j a_i$ 的形式），所以

$$\left(\sum_{i=1}^{n} a_i\right)^2 = S + 2P.$$

? **想一想** (iv) 給出的是 S 的界，它怎樣能幫助我們得到一個只提及 P 的界？

由於 $n - 1 > 0$，可以把 (iv) 除以 $n-1$ 而不必把不等號反轉，得 $S \geq \dfrac{2P}{n-1}$。把 S 的這個下界（lower bound）代入：

$$\begin{aligned}\left(\sum_{i=1}^{n} a_i\right)^2 &= S + 2P \\ &\geq \frac{2P}{n-1} + 2P \\ &= \frac{2P + 2(n-1)P}{n-1} \\ &= \frac{2n}{n-1} \sum_{1 \leq i < j \leq n} a_i a_j,\end{aligned}$$

這就是所求的不等式。等號成立當且僅當 (iv) 的等號成立，即所有 a_i 都相等。

留意我們從來不需要 P 為正數：整個論證只是在 $S \geq 2P/(n-1)$ 的兩邊同時加上 $2P$，無論 P 的正負如何，這一步都成立。

驗算

- 全部相等：若每個 $a_i = c$，則 $S = nc^2$ 及 $P = \binom{n}{2}c^2 = \frac{1}{2}n(n-1)c^2$，所以 $(n-1)S = n(n-1)c^2 = 2P$，而在 (v) 中兩邊都等於 n^2c^2。兩條不等式都取到等號，正如所料。
- 以數值檢驗 (iii)：取 $(a_1, a_2, a_3, a_4) = (1, 2, -3, 5)$，得 $S = 39$ 及 $P = -7$，所以

$3S = 117 \geq -14 = 2P$；而在 (v) 中，$(\sum a_i)^2 = 25 \geq \frac{8}{3}(-7) = -\frac{56}{3}$。（另外 $S + 2P = 39 - 14 = 25$，亦正如所料。）

- 我們亦以有理數運算，在 20 000 個長度由 2 至 9 的隨機數組上，以及一些非常接近等號情況的數組上，精確地檢驗了恆等式（identity）$D = (n-1)S - 2P$ 和兩條不等式，並沒有發現反例（counterexample）。

常見錯誤

- 數錯次數：每個 a_k^2 在 $\sum_{i<j}(a_i^2 + a_j^2)$ 中出現 $n-1$ 次，不是 n 次，也不是 $\binom{n}{2}$ 次。
- 寫成 $(\sum a_i)^2 = S + P$：每個交叉項在展開式中出現兩次。
- 把 (iv) 除以 $n-1$ 時沒有指出 $n - 1 > 0$，或忘記說明等號成立的情況。

STEP 之外

$D = (n-1)S - 2P$ 這個量等於各數與平均值之差的平方和的 n 倍，即 $n\sum(a_i - \bar{a})^2$，所以 (iv) 其實是在說方差（variance）非負。第 (v) 部分也可以看成柯西–施瓦茲不等式（Cauchy–Schwarz inequality）$(\sum a_i \cdot 1)^2 \leq n\sum a_i^2$ 的變裝。在線性代數（linear algebra）中，這是說二次型（quadratic form）$(n-1)S - 2P$ 是半正定（positive semi-definite）的，而它的核（kernel）由全一向量張成。Strang 的 *Introduction to Linear Algebra*（線性代數入門）講述二次型與正定性，而 Siklos 的 *Advanced Problems in Mathematics*（進階數學難題）則提供更多 STEP 風格的不等式練習。

詳細解答

STEP III 1993，第 10 題

純數學 • 矩陣變換

STEP III 1993，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：

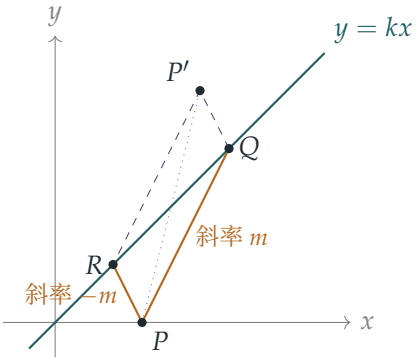
step.maths.org/questions/93-s3-q10

關鍵思路

這裏每一步都是求兩條直線的交點，所以各點的坐標只需每次解一條一次方程便可得到。平行四邊形（parallelogram）則讓我們不必再多做計算就得到 P'：在平行四邊形 $PQP'R$ 中，兩條對角線（diagonal）互相平分，所以以位置向量（position vector）表示時有 $P + P' = Q + R$。只要把 P' 寫成一個矩陣（matrix）乘以 P，面積的問題就只是計算一個行列式（determinant）。

解答

全題都需要 $m^2 \neq k^2$：若 $m = \pm k$，則通過 P 的兩條直線之中，有一條與 $y = kx$ 平行，永不與它相交（或者與它重合），這個作圖便不成立。



（圖中取 $m = 2$，$k = 1$ 及 $P = (1, 0)$。）

點 Q

通過 $P(x_1, y_1)$ 且平行於 $y = mx$ 的直線是 $y - y_1 = m(x - x_1)$。在 Q 點同時有 $y = kx$，所以

$$kx - y_1 = mx - mx_1 \quad \Longrightarrow \quad (m-k)x = mx_1 - y_1 \quad \Longrightarrow \quad x = \frac{mx_1 - y_1}{m-k},$$

其中除以了 $m - k \neq 0$。再由 $y = kx$ 得

$$Q = \left(\frac{mx_1 - y_1}{m-k}, \frac{k(mx_1 - y_1)}{m-k} \right),$$

這正是所求。

點 R

通過 P 且平行於 $y = -mx$ 的直線，計算方法完全相同，只需把 m 換成 $-m$：

$$x = \frac{-mx_1 - y_1}{-m-k} = \frac{mx_1 + y_1}{m+k}, \qquad R = \left(\frac{mx_1 + y_1}{m+k}, \frac{k(mx_1 + y_1)}{m+k} \right).$$

點 P'

想一想 怎樣可以不再求任何直線的交點，就找到平行四邊形的第四個頂點？

各頂點依次為 P, Q, P', R，所以 PP' 和 QR 是對角線。平行四邊形的對角線互相平分，因此兩者的中點相同：

$$\tfrac{1}{2}(P + P') = \tfrac{1}{2}(Q + R) \quad \Longrightarrow \quad P' = Q + R - P.$$

（等價地，$\overrightarrow{QP'} = \overrightarrow{PR}$。）$Q$ 和 R 都在 $y = kx$ 上，所以最簡單的做法是先把它們的 x 坐標相加。以 $m^2 - k^2$ 為公分母，

$$\begin{aligned} x_Q + x_R &= \frac{(mx_1 - y_1)(m+k) + (mx_1 + y_1)(m-k)}{m^2 - k^2} \\ &= \frac{(m^2x_1 + mkx_1 - my_1 - ky_1) + (m^2x_1 - mkx_1 + my_1 - ky_1)}{m^2 - k^2} \\ &= \frac{2(m^2x_1 - ky_1)}{m^2 - k^2}, \end{aligned}$$

並且 $y_Q + y_R = k(x_Q + x_R)$。因此，設 $P' = (x', y')$，則

$$x' = \frac{2(m^2x_1 - ky_1)}{m^2 - k^2} - x_1 = \frac{(m^2 + k^2)x_1 - 2ky_1}{m^2 - k^2},$$

$$y' = \frac{2k(m^2x_1 - ky_1)}{m^2 - k^2} - y_1 = \frac{2km^2x_1 - (m^2 + k^2)y_1}{m^2 - k^2}.$$

把 T 寫成矩陣

兩個坐標都是 x_1, y_1 的線性式，而且沒有常數項，所以

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{M} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \qquad \mathbf{M} = \frac{1}{m^2 - k^2} \begin{pmatrix} m^2 + k^2 & -2k \\ 2km^2 & -(m^2 + k^2) \end{pmatrix}.$$

面積

矩陣為 $\mathbf{M}$ 的線性映射（linear map）會把每個面積乘以 $|\det \mathbf{M}|$。從每一行抽出因子 $\frac{1}{m^2-k^2}$，

$$\det \mathbf{M} = \frac{-(m^2+k^2)^2 + 4k^2m^2}{(m^2-k^2)^2} = \frac{-(m^4 - 2m^2k^2 + k^4)}{(m^2-k^2)^2} = \frac{-(m^2-k^2)^2}{(m^2-k^2)^2} = -1.$$

所以 $|\det \mathbf{M}| = 1$，每個區域都被映射到面積相同的區域，這正是所求。負號告訴我們 T 會把定向（orientation）反轉，正如反射（reflection）一樣：一個以逆時針方向描出的圖形，其像是以順時針方向描出的。

> ? **想一想** 能否用一幅圖畫說明 T 在做甚麼？

可以。PP' 的中點等於 QR 的中點，而後者在 $y = kx$ 上；而且 $y = kx$ 上的點都固定不動（在那裏 $Q = R = P$）。解 $\mathbf{Mv} = -\mathbf{v}$ 得出一行 $2m^2x - 2ky = 0$，所以被 T 反轉的向量就是 (k, m^2) 的倍數。因此 T 是一個斜反射（skew reflection）：它以直線 $y = kx$ 為鏡作反射，但反射的方向是沿 (k, m^2)，而不是垂直於該直線。與任何反射一樣，它滿足 $\mathbf{M}^2 = \mathbf{I}$，這也與作圖吻合：從 P' 出發會得到相同的 Q 和 R，於是回到 P。

驗算

- 取 $m = 2$，$k = 1$，$P = (1, 0)$：$Q = (2, 2)$，$R = (\frac{2}{3}, \frac{2}{3})$，所以 $P' = (\frac{5}{3}, \frac{8}{3})$。矩陣是 $\frac{1}{3}\begin{pmatrix} 5 & -2 \\ 8 & -5 \end{pmatrix}$，它把 $(1, 0)$ 送到 $(\frac{5}{3}, \frac{8}{3})$，而它的行列式是 $\frac{1}{9}(-25 + 16) = -1$。
- 直線上的點 (t, kt) 應該固定不動：$x' = \frac{(m^2+k^2)t - 2k^2t}{m^2-k^2} = t$。✓
- $\mathbf{M}$ 的跡（trace）是 0，行列式是 -1，所以它的特徵值（eigenvalue）是 ± 1，與反射應有的情況相符。
- 我們以有理數精確運算，隨機選取約 3000 組 m、k 和 P，核對了以上所有公式，以及三角形面積保持不變這一點。

常見錯誤

- 取 $P' = Q + R$ 或 $P' = P + Q + R$，而不是 $Q + R - P$。字母的次序 $PQP'R$ 已告訴你哪些頂點互為對頂點。
- 由 $\det \mathbf{M} = -1$ 推斷面積變成負數。面積的比例因子是 $|\det \mathbf{M}|$；正負號只記錄定向。

STEP 之外

一個線性映射若把某條直線上的每一點都固定，並把另一個方向反轉，它就是一個對合（involution），即 $T^2 = I$。平常的正交反射是其中的特殊情況，即被反轉的方向垂直於該直線；在這裏，這種情況恰好在 (k, m^2) 垂直於 $(1, k)$ 時出現，也就是只有當 $k = 0$ 時（直線 $y = \pm mx$ 關於 x 軸對稱，而不是關於 $y = kx$ 對稱）。在大學裏，這屬於特徵值與對角化（diagonalisation）的課題，在那裏 T 在基（basis）$(1, k), (k, m^2)$ 之下可看成 $\mathrm{diag}(1, -1)$；它也屬於把行列式理解為帶正負號的體積比例因子的課題。斯特朗（Strang）的 *Introduction to Linear Algebra*（線性代數入門）對這兩個概念都有清晰的講解。

詳細解答

STEP III 1993，第 11 題

力學 • 力偶與轉動慣量

STEP III 1993，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/93-s3-q11

關鍵思路

在立方體上建立坐標，並把每個力寫成向量（vector）。六個力之和為零，所以它們的作用只是施加一個力偶（couple），而力偶對任何一點的力矩（moment）都相同。第 (i) 部分討論的就是這個力矩的方向。第 (ii) 部分只用到它沿 OO' 的分量，再配合轉動慣量（moment of inertia）；我們會利用點到直線的距離，逐根桿把轉動慣量計算出來。

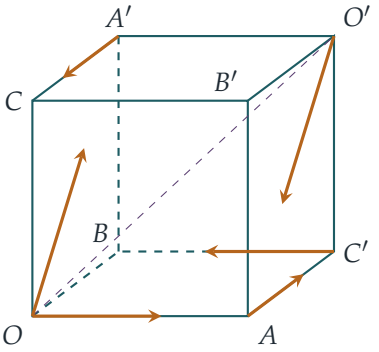
解答

我們把框架視為剛體（rigid body），由均勻細桿組成，並且不考慮重力。取 O 為原點，單位向量分別沿 OA、OB、OC，於是

$$A = (a,0,0), \quad B = (0,a,0), \quad C = (0,0,a),$$

而相對的頂點為

$$O' = (a,a,a), \quad A' = (0,a,a), \quad B' = (a,0,a), \quad C' = (a,a,0).$$



琥珀色箭頭顯示六條作用線，對角線 OO' 則以紫紅色虛線表示。

第 (i) 部分

每個力等於它的大小乘以沿其作用線的單位向量。例如，$O'A$ 是向量 $(0,-a,-a)$，長度為 $\sqrt{2}\,a$，所以沿它作用的力 $\sqrt{2}\,mg$ 就是 $mg(0,-1,-1)$。對全部六個力作同樣處理：

作用線	線上一點	力
OA	O	$mg(2,0,0)$
AC'	A	$mg(0,1,0)$
$O'A$	A	$mg(0,-1,-1)$
OA'	O	$mg(0,1,1)$
$C'B$	B	$mg(-2,0,0)$
$A'C$	C	$mg(0,-1,0)$

這些力之和為 $\mathbf{0}$。合力（resultant）為零的力系等價於一個力偶，而力偶對任何一點的力矩都相同。對 O 取力矩（沿 OA 和 OA' 的力通過 O，沒有貢獻）：

$$\begin{aligned}\mathbf{G} &= A\times mg(0,1,0) + A\times mg(0,-1,-1) + B\times mg(-2,0,0) + C\times mg(0,-1,0)\\ &= amg\big[(0,0,1)+(0,1,-1)+(0,0,2)+(1,0,0)\big]\\ &= amg\,(1,1,2).\end{aligned}$$

位於 O 的支點可以對框架施力，但這個力通過 O，所以對 O 沒有力矩。因此，框架由靜止開始，會繞着通過 O 並沿力偶軸線的直線轉動，其方向為

$$\begin{pmatrix}1\\1\\2\end{pmatrix},$$

正如題目所求。

*給好奇讀者的補充。*這是題目預期的答案，也是你應該寫下的答案。不過嚴格來說，對一個可在三維空間自由轉動的物體，角加速度（angular acceleration）是 $\mathbf{I}^{-1}\mathbf{G}$，其中 $\mathbf{I}$ 是對 O 的慣量張量（inertia tensor）。只有當 $\mathbf{G}$ 沿着主軸（principal axis）時，角加速度才會指向 $\mathbf{G}$ 的方向。就這個框架而言，對 O 來說，繞 OO' 的轉動慣量是 $14ma^2/3$（見第 (ii) 部分），而繞每一條通過 O 並垂直於 OO' 的軸的轉動慣量都是 $41ma^2/3$。由此得出初始轉軸的方向為 $(25,25,32)$，與 $(1,1,2)$ 相差約 12.6°。

第 (ii) 部分

設 $\mathbf{u} = \frac{1}{\sqrt{3}}(1,1,1)$ 為沿 OO' 的單位向量。由於 $\mathbf{p}\cdot\mathbf{u}$ 是 $\mathbf{p}$ 沿該直線的分量，根據畢氏定理（Pythagoras' theorem），點 $\mathbf{p}$ 與直線 OO' 的距離的平方為

$$d^2 = |\mathbf{p}|^2 - (\mathbf{p}\cdot\mathbf{u})^2$$

。

桿 OA。與 O 相距 x 的點是 $(x,0,0)$，所以 $d^2 = x^2 - \frac{x^2}{3} = \frac{2}{3}x^2$。長度為 $\mathrm{d}x$ 的小段，質量為 $\frac{m}{a}\,\mathrm{d}x$，所以

$$I_{OA} = \int_0^a \frac{m}{a}\cdot\frac{2x^2}{3}\,\mathrm{d}x = \frac{2ma^2}{9}.$$

中點 $(\frac{a}{2},0,0)$ 與 OO' 的距離的平方為 $\frac{2}{3}\cdot\frac{a^2}{4} = \frac{a^2}{6}$。根據平行軸定理（parallel axis theorem），繞通過中點的平行軸，

$$I_{\text{中點}} = \frac{2ma^2}{9} - m\cdot\frac{a^2}{6} = \frac{ma^2}{18}.$$

桿 $B'C$。這根桿由 $(a,0,a)$ 伸延至 $(0,0,a)$，所以與 OA 平行。它的長度也相同，與轉軸所成的角度亦相同。因此，繞通過它自身中點的平行軸，它的轉動慣量同樣是 $ma^2/18$。它的中點為 $\mathbf{p} = (\frac{a}{2},0,a)$，而

$$d^2 = \frac{a^2}{4} + a^2 - \frac{1}{3}\left(\frac{3a}{2}\right)^2 = \frac{5a^2}{4} - \frac{3a^2}{4} = \frac{a^2}{2}.$$

根據平行軸定理，

$$I_{B'C} = \frac{ma^2}{18} + \frac{ma^2}{2} = \frac{5ma^2}{9}.$$

想一想 我真的要對全部十二根桿逐一這樣計算嗎？

不用：對稱性（symmetry）會替我們完成。繞 OO' 旋轉 120°（即輪換 $x\to y\to z\to x$），以及對立方體中心的點反射 $\mathbf{p}\mapsto(a,a,a)-\mathbf{p}$，兩者都把框架和直線 OO' 映射到自身。所以它們不會改變任何繞 OO' 的轉動慣量。旋轉把 OA 帶到 OB 和 OC，而反射再把這三根桿帶到 $O'A'$、$O'B'$ 和 $O'C'$。這就是一端位於 O 或 O' 的六根桿，每根的轉動慣量都是 $2ma^2/9$。同樣地，$B'C$ 會被帶到環繞立方體中部的「之字形」AB'、$B'C$、CA'、$A'B$、BC'、$C'A$ 中的其餘五根桿，而這些桿每根的轉動慣量都是 $5ma^2/9$。因此

$$I_{\text{框架}} = 6\cdot\frac{2ma^2}{9} + 6\cdot\frac{5ma^2}{9} = \frac{42ma^2}{9} = \frac{14ma^2}{3},$$

正如題目所求。

角加速度。現在框架只能繞固定直線 OO' 轉動。軸承會提供垂直於該直線所需的任何力偶，所以只有外加力偶沿 OO' 的分量會令框架轉動。這些力仍然構成力偶

$\mathbf{G} = amg(1,1,2)$，而它對任何直線的力矩，就是它沿該直線的分量：

$$\mathbf{G}\cdot\mathbf{u} = \frac{amg(1+1+2)}{\sqrt{3}} = \frac{4amg}{\sqrt{3}}.$$

由繞固定軸轉動的方程 $I\ddot{\theta} = \mathbf{G}\cdot\mathbf{u}$ 可得

$$\ddot{\theta} = \frac{4amg/\sqrt{3}}{14ma^2/3} = \frac{12g}{14\sqrt{3}\,a} = \frac{2\sqrt{3}\,g}{7a}.$$

由 O' 望向 O 時，它的轉向是逆時針的，也就是繞方向 $\overrightarrow{OO'}$ 的右手旋轉。

驗算

- 六個力各有一個「拍檔」：OA 配 $C'B$，AC' 配 $A'C$，$O'A$ 配 OA'。每一對都是大小相等、方向相反，並且作用線平行，所以合力為零。
- 框架對其中心具有立方對稱性，所以繞每一條通過中心的軸，轉動慣量都相同。繞通過中心並平行於 OC 的軸，四根垂直的桿貢獻 $4m(a/\sqrt{2})^2 = 2ma^2$。八根水平的桿每根貢獻 $\frac{ma^2}{12} + \frac{ma^2}{4} = \frac{ma^2}{3}$。總和是 $\frac{14ma^2}{3}$，與上面的答案吻合，因為 OO' 同樣通過中心。
- 數值上，取 $a = m = g = 1$，對十二根桿以中點法求和，得出 $I = 4.66667$ 及 $\ddot{\theta} = 0.494872$，與 $2\sqrt{3}/7 \approx 0.494872$ 相符。

常見錯誤

- 把沿面對角線的力 $\sqrt{2}\,mg$ 化為分量時，忘記了因子 $1/\sqrt{2}$：每個力的分量大小是 mg，而不是 $\sqrt{2}\,mg$。
- 在第 (ii) 部分使用了整個力偶 $|\mathbf{G}| = \sqrt{6}\,amg$，而不是它沿固定軸的分量。
- 從一條不通過質心（centre of mass）的軸出發應用平行軸定理：這個定理必須經由質心，就像由 $2ma^2/9$ 到 $ma^2/18$ 的那一步。

STEP 之外

對於三維空間中的轉動，轉動慣量會變成一個 3×3 的對稱矩陣（symmetric matrix），即慣量張量，而角動量（angular momentum）為 $\mathbf{L} = \mathbf{I}\boldsymbol{\omega}$。它的特徵向量（eigenvector）就是主軸。正如第 (i) 部分的補充所示，物體一般不會繞着外加力偶的軸線開始轉動。這個立方體框架是個很好的例子：繞每一條通過其中心的軸，它的轉動慣量都相同，但對一個角來說卻並非如此。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門）闡述了三維空間中的剛體運動，包括慣量張量。

詳細解答

STEP III 1993，第 12 題

力學 • 彈性繩與能量

STEP III 1993，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：

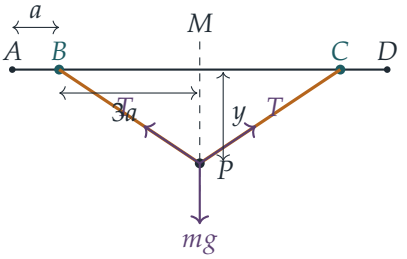
step.maths.org/questions/93-s3-q12

關鍵思路

由於釘子是光滑（smooth）的，而質點位於弦線的中點，弦線便分成兩段完全相同的半段，每段的自然長度（natural length）均為 a，而且每一半段上各處的張力（tension）相同。每一半段都由一個固定端出發，繞過一枚釘子，再向下（或向上）連到 P，所以它拉伸後的長度只取決於距離 BP。只要把伸長量（extension）用 P 的高度表示出來，平衡（equilibrium）條件便給出 k，而能量守恆（conservation of energy）則把 Q 與 R 聯繫起來。

解答

我們把 P 看作質點（particle），把弦線看作輕而有彈性（light and elastic），並把釘子看作光滑。根據對稱性，P 沿着通過 BC 中點 M 的鉛垂線運動。設 y 為 P 在 M 之下的深度；當 P 在 BC 之上時，它是負數。



求 k

半段 $A \to B \to P$ 的長度為 $a + BP = a + \sqrt{9a^2 + y^2}$，自然長度為 a，所以它的伸長量是

$$x = \sqrt{9a^2 + y^2}.$$

根據胡克定律（Hooke's law），每一半段的張力為

$$T = \frac{kmg\,x}{a}.$$

若 θ 是 BP 與水平線所成的角，則 $\sin\theta = |y|/x$。在 P 點沿鉛垂方向分解力，得

$$2T\sin\theta = mg \quad \Longrightarrow \quad 2 \cdot \frac{kmg\,x}{a} \cdot \frac{y}{x} = mg \quad \Longrightarrow \quad \frac{2ky}{a} = 1.$$

平方根已經約去，所以那個難看的數值 $BP = \frac{\sqrt{145}}{4}a$ 根本用不着。代入 $y = a/4$，得 $k/2 = 1$，所以

$$k = 2.$$

由 Q 到 R 的能量

每一半段儲存的彈性勢能（elastic energy）為 $\dfrac{kmg\,x^2}{2a}$，所以當 $k = 2$ 時，兩個半段合共儲存

$$E(y) = 2 \cdot \frac{2mg\,x^2}{2a} = \frac{2mg\,(9a^2 + y^2)}{a}.$$

? **想一想** 為甚麼 P 會在 R 停下來？P 最後升到釘子之上，為甚麼沒有問題？

P 在 Q 由靜止釋放，而 R 是它升到的最高點，所以它在那裏同樣是瞬時靜止的。在 BC 之上，兩個半段仍然處於拉伸狀態（每段仍有 $x \geq 3a > 0$），所以 E 的公式同樣適用。弦線始終不會鬆弛。

在 Q，$y = pa$；在 R，$y = -3a$。質點上升了 $(p+3)a$。由於釘子光滑，又沒有其他能量損失，而動能（kinetic energy）在兩端均為零，所以

$$\begin{aligned} E(pa) &= E(-3a) + mg(p+3)a \\ 2mga(9+p^2) &= 2mga(9+9) + mga(p+3) \\ 18 + 2p^2 &= 36 + p + 3, \end{aligned}$$

由此得

$$2p^2 - p - 21 = 0, \qquad \text{即} \qquad (2p+6)\left(p - \tfrac{7}{2}\right) = 0.$$

由於 $p > 0$，我們得到 $p = \frac{7}{2}$。

這與題目印出的 $2p^2 - p - 17 = 0$ 並不吻合。我們已仔細核對過這個模型（見「驗算」），包括直接進行數值模擬，而按題目所給的數據，得出的確實是 -21。$p = \frac{7}{2}$ 這個整齊的根顯示，21 才是原意中的常數，17 應是誤印。最後一部分並不取決於哪一個常數才正確。

要確認 P 的確是在 R 第一次停下來，可設 $y = ta$。在深度 ta 處的動能為

$$E(pa) - E(ta) - mg(p-t)a = mga\,(p-t)(2p+2t-1).$$

當 $p = \frac{7}{2}$ 時，這就是 $mga\,(\frac{7}{2} - t)(2t+6)$。它在 $-3 < t < \frac{7}{2}$ 時為正，並在 $t = -3$ 時首次變為零，而這正是 R。

比較張力

張力為 $T = 2mg\,x/a$，它隨 x 增加而增加，所以我們需要證明 x 在 R 的值比在 Q 的值小。能量方程毋須用到 p 的值便可做到這一點：

$$E(Q) - E(R) = mg(p+3)a > 0.$$

彈性勢能為 $\dfrac{2mg\,x^2}{a}$，所以 $x_R^2 < x_Q^2$，從而 $x_R < x_Q$，因此 $T_R < T_Q$。用文字來說，質點在上升途中獲得了勢能（potential energy），而這些能量只能來自弦線，所以弦線在 R 的拉伸程度必定較小。

具體而言，當 $p = \frac{7}{2}$ 時，$x_Q = \frac{\sqrt{85}}{2}a$ 而 $x_R = 3\sqrt{2}\,a$，所以

$$T_Q = \sqrt{85}\,mg \approx 9.22\,mg, \qquad T_R = 6\sqrt{2}\,mg \approx 8.49\,mg.$$

驗算

- 簡諧運動驗算。作用於 P 的向上的力為 $2T \cdot \dfrac{y}{x} = \dfrac{4mg\,y}{a}$，它是 y 的線性函數，在 BC 之上和之下都是如此。所以 P 圍繞 $y = a/4$（即上面求得的平衡位置）作簡諧運動（simple harmonic motion）。它在 $y = \frac{7}{2}a$ 由靜止釋放，振幅（amplitude）為 $\frac{13}{4}a$，所以它的最高點是 $y = \frac{1}{4}a - \frac{13}{4}a = -3a$。這正好就是 R，從而獨立地確認了 $p = \frac{7}{2}$。
- 取 $a = 1$、$k = 2$、$p = 3.5$，對運動方程作數值積分，所得的最高點為 $y = -3.0000$。
- 題目印出的方程會給出 $p = \frac{1+\sqrt{137}}{4} \approx 3.18$，而這樣能量收支便會相差 $4mga$。

常見錯誤

- 計算每一半段的長度時漏掉了 AB 這一段。半段的長度是 $AB + BP = a + BP$，而它的伸長量之所以是 BP，只是因為 AB 剛好等於它的自然長度 a。

- 把整條弦線當作一條彈簧，張力為 $kmg \cdot (\text{總伸長量})/(2a)$。這樣做的確會得到相同的答案，但前提是伸長量必須取總伸長量 $2x$。
- 忘記了當 P 在 BC 之上時，彈性勢能仍然存在。弦線在那裏仍然處於拉伸狀態。

STEP 之外

儘管弦線的幾何關係是非線性的，鉛垂方向的回復力（restoring force）卻恰好與位移成正比。這是因為，若張力與長度成正比，它的鉛垂分量便與鉛垂方向的距離成正比；自然長度為零的彈簧能成為理想的懸掛裝置，也是出於同一個原因（這正是 LaCoste 長周期地震儀背後的原理）。勢能若是坐標的二次函數，便會導出線性的運動方程，而這正是小振動（small oscillations）與簡正模式（normal modes）理論的起點。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）正是以這種精神處理彈簧、能量和振動的。

詳細解答

STEP III 1993，第 13 題

力學 • 變質量鏈條

> **STEP III 1993，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/93-s3-q13

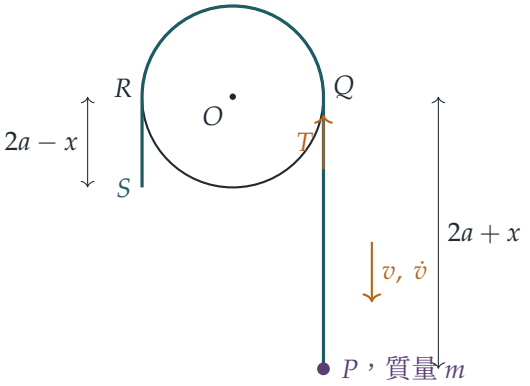
關鍵思路

由於鏈條不能滑動，圓盤、整條鏈條和質點都以相同的速率運動：未知量只有一個，就是 v。於是動能（kinetic energy）為 $\frac{1}{2}$(等效質量)v^2。在勢能（potential energy）方面，只有兩段懸垂部分會改變，因為繞過頂部的那段弧只是不斷被完全相同的鏈條取代。至於張力（tension），可考慮運動中的 PQ 段連同質點。然後仔細想想：在 Q 處加入這一段的鏈條，本身已經以多快的速率運動。

解答

模型

圓盤是均勻薄片（uniform lamina），繞通過 O 的光滑固定軸轉動，其轉動慣量（moment of inertia）為 $\frac{1}{2}(4m)a^2 = 2ma^2$。鏈條是均勻的，每單位長度的質量為 m/a，不可伸長，並且不會在盤邊上滑動。質點視為點質量。當 P 下降了 x 時，右邊的一段 QP 長度為 $2a + x$，左邊的一段 RS 長度為 $2a - x$，而弧 RQ 是長度為 πa 的半圓。鏈條上每一點都以速率 v 運動，由於沒有滑動，盤邊亦然，所以圓盤的角速率（angular speed）為 $\omega = v/a$。



速率

動能。圓盤的貢獻為 $\frac{1}{2}(2ma^2)(v/a)^2 = mv^2$，鏈條為 $\frac{1}{2}(4+\pi)mv^2$，質點為 $\frac{1}{2}mv^2$。所以

$$\text{動能} = \tfrac{1}{2}mv^2\big(2+4+\pi+1\big) = \tfrac{1}{2}(\pi+7)mv^2.$$

勢能。弧 RQ 始終是同樣的一段半圓形鏈條，所以它的勢能永不改變。取 O 的水平為零勢能面。一段長度為 ℓ、懸垂在該水平之下的鏈條，質量為 $m\ell/a$，其質心（centre of mass）位於深度 $\frac{1}{2}\ell$ 處，所以它的勢能為 $-\dfrac{mg\ell^2}{2a}$。因此，兩段懸垂部分的勢能損失為

$$\frac{mg}{2a}\Big[(2a+x)^2+(2a-x)^2-(2a)^2-(2a)^2\Big] = \frac{mg}{2a}\cdot 2x^2 = \frac{mgx^2}{a}.$$

質點則損失 mgx。

能量。軸是光滑的，而粗糙接觸面不作功，因為沒有滑動。所以能量守恆，而由靜止開始，

$$\tfrac{1}{2}(\pi+7)mv^2 = \frac{mgx^2}{a} + mgx \quad\Longrightarrow\quad (\pi+7)av^2 = 2g(x^2+ax),$$

正如所求。

加速度

對時間求導，並利用 $\dot{x} = v$：

$$2(\pi+7)av\dot{v} = 2g(2x+a)v \quad\Longrightarrow\quad \dot{v} = \frac{g(2x+a)}{(\pi+7)a} \qquad (v \neq 0).$$

當 $x = 2a$ 時，S 到達 R，這時

$$\dot{v} = \frac{5g}{\pi+7}, \qquad v^2 = \frac{2g(4a^2+2a^2)}{(\pi+7)a} = \frac{12ga}{\pi+7}.$$

Q 處的張力

設 T 為 Q 處的張力。我們選取的系統是質點連同懸垂在 Q 之下的鏈條。它的質量為

$$M = \frac{m}{a}(2a + x) + m,$$

隨着鏈條離開圓盤，這個質量以 $\dot{M} = \dfrac{m}{a}v$ 的速率增加。作用於它的外力是向下的重量 Mg 和向上的 T。整個系統都以速率 v 向下運動。

> **想一想** 正要加入 PQ 的鏈條，在加入前的一刻運動得有多快？

在 Q 處，盤邊是豎直的，所以到達該處的鏈條本身已經以速率 v 豎直向下運動。它把自己的動量（momentum）$v\,\delta M$ 一併帶來。於是變質量方程（variable-mass equation）（動量的變化率 = 力 + 每單位時間帶入的動量）為

$$\frac{\mathrm{d}}{\mathrm{d}t}(Mv) = Mg - T + v\dot{M}, \qquad \text{即} \qquad M\dot{v} = Mg - T.$$

當 S 到達 R 時，$M = 5m$，所以

$$T = 5m\left(g - \frac{5g}{\pi + 7}\right) = \frac{5(\pi + 2)}{\pi + 7}\,mg.$$

*用力矩覆核。*把圓盤和鏈條的弧 RQ 一併考慮。軸的作用力通過 O，弧的重量對 O 的總力矩（moment）為零（弧對通過 O 的豎直線對稱），而鏈條與圓盤之間的摩擦力是內力。弧上每一個質點都在距離 O 為 a 之處以速率 v 運動。所以角動量（angular momentum）為 $(2ma^2 + \pi ma^2)(v/a)$，而

$$(T - T_R)a = (2 + \pi)ma\,\dot{v}.$$

當 S 到達 R 時，左邊的一段已縮短至零，所以 $T_R = 0$。由此再次得出 $T = (2 + \pi)m \cdot \dfrac{5g}{\pi + 7} = \dfrac{5(\pi + 2)}{\pi + 7}mg$。

*試卷印出的答案。*試卷給出的值 $\dfrac{5\pi - 2}{\pi + 7}mg$，正是寫成 $\dfrac{\mathrm{d}}{\mathrm{d}t}(Mv) = Mg - T$ 時所得的結果，這樣寫等於把加入 PQ 的鏈條當作是靜止地到達：

$$T = M(g - \dot{v}) - \frac{m}{a}v^2 = \frac{5(\pi + 2)}{\pi + 7}mg - \frac{12}{\pi + 7}mg = \frac{5\pi - 2}{\pi + 7}mg.$$

對於一顆雨點沿途吸收靜止的霧氣，這條方程是對的。但在這裏並不正確，因為繞過盤邊而來的鏈條本身已經隨 PQ 一起運動。我們的兩種方法都得出 $\dfrac{5(\pi + 2)}{\pi + 7}mg$，所以我們相信印出的答案包含了這個錯誤。如果你需要得出印出的數值，可使用「靜止地加入」的方程，但要知道它把 T 少算了 mv^2/a。

驗算

- 在 $x = 0$ 時，公式給出 $v = 0$ 和 $\dot{v} = g/(\pi + 7)$。不平衡的力只有質點的重量 mg，作用於等效質量 $(\pi + 7)m$ 之上。
- 由靜止開始，以數值方法逐步求解運動方程 $(\pi + 7)m\dot{v} = mg(2x + a)/a$，可重現 $(\pi + 7)av^2 = 2g(x^2 + ax)$，誤差在 10^{-11} 之內。取 $m = a = 1$ 和 $g = 9.81$，動量法和力矩法都給出 $T/mg \approx 2.5349 = 5(\pi + 2)/(\pi + 7)$。「靜止地加入」的版本則給出 $1.3517 = (5\pi - 2)/(\pi + 7)$。
- $T < 5mg$，這是必然的，因為懸垂的質量 $5m$ 正在向下加速。

常見錯誤

- 忘記計算圓盤的動能，或者在計算其轉動慣量時用了 ma^2 而不是 $\frac{1}{2}(4m)a^2$。
- 把鏈條的弧計入勢能的變化之中。它的形狀和位置從不改變。
- 使用 $\frac{\mathrm{d}}{\mathrm{d}t}(Mv) = F$ 時，沒有問清楚加入的質量本身已有甚麼速度。

STEP 之外

變質量問題的真正重點，在於選定一個系統，並追蹤越過其邊界的動量。這就是控制體積（control volume）的概念，它是火箭運動和大部分流體力學（fluid mechanics）的基礎。這裏的常見錯誤，即假設 $F = \frac{\mathrm{d}}{\mathrm{d}t}(mv)$ 對任何變化中的質量都成立，是一個著名的陷阱：一般而言，這條方程甚至會隨參考系（frame of reference）而改變。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教科書）對變質量系統有嚴謹的論述。

詳細解答

STEP III 1993，第 14 題

力學 • 鉛垂圓周運動

STEP III 1993，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：

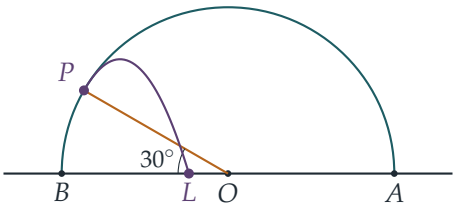
step.maths.org/questions/93-s3-q14

關鍵思路

由於繩子水平，而質點（particle）沿鉛垂方向拋出，初速度與繩子垂直，所以質點會在以 O 為圓心、半徑為 c 的鉛垂圓上運動，並落在與 A 位於同一直徑另一端的點 B。每次碰撞時速度都是鉛垂的，所以碰撞只是把速率減半。我們要找出質點在第幾次反彈後首次無法越過圓的頂點，然後利用能量和徑向方程（radial equation）求出張力（tension）變為零的位置。此後質點便是普通的拋體（projectile）。

解答

我們把質點看成質量為 m 的質點，繩子為輕而不可伸長（light and inextensible），桌面為光滑（smooth）且水平。取 O 為原點，並設 $A = (c, 0)$ 及 $B = (-c, 0)$。無論質點在哪一側上升，設它位於該側水平半徑之上、夾角為 θ 的位置。



兩次碰撞之間的運動

假設質點在 A 或 B 離開桌面，以速率 V 鉛垂向上運動。它的速度與拉緊的繩子垂直，所以只要繩子保持拉緊，它便在以 O 為圓心、半徑為 c 的圓上運動。根據能量守恆（con-

servation of energy），在角度 θ 時，它的速率 v 滿足

$$v^2 = V^2 - 2gc\sin\theta.$$

沿指向 O 的方向分解，設張力為 T，

$$T + mg\sin\theta = \frac{mv^2}{c} \quad\Longrightarrow\quad \frac{T}{m} = \frac{V^2}{c} - 3g\sin\theta.$$

當 $0 \le \theta \le 90^\circ$ 時，張力在頂點最小，所以繩子在整個半圓上保持拉緊，當且僅當 $V^2 \ge 3gc$。（過了頂點之後，$\sin\theta$ 再次減小，所以 T 只會增大。）在這情況下，質點會落到直徑另一端的桌面上，而根據能量守恆，它此時鉛垂向下運動，速率同樣是 V。

在該次碰撞中，速度與桌面垂直，所以反彈方向是鉛垂的，速率為 $\frac{1}{2}V$，而繩子仍然水平並且拉緊。因此整個過程會重複一次，只是 V 減半。

繩子在哪裏變鬆

質點起初的速率為 $V_0 = 64\sqrt{6gc}$，所以 $V_0^2 = 24576\,gc$。經過 k 次碰撞後，速率為 $V_k = V_0/2^k$，所以

$$V_k^2 = \frac{24576}{4^k}\,gc: \qquad V_6^2 = 6gc, \qquad V_7^2 = \tfrac{3}{2}gc.$$

因此，以速率 $V_0, \ldots, V_6$ 開始的各個半圓全部都能走完（每個都有 $V_k^2 \ge 3gc$），碰撞依次發生在 B, A, B, A, B, A, B。第七次碰撞在 B，而質點從 B 上升時 $V_7^2 = \frac{3}{2}gc < 3gc$。

想一想 質點會不會在繩子變鬆之前就已經往回落，而不是繩子先變鬆？

不會。速率會在 $V_7^2 = 2gc\sin\theta$ 處變為零，即 $\sin\theta = \frac{3}{4}$，但張力更早便變為零，位置滿足

$$\frac{V_7^2}{c} = 3g\sin\theta \quad\Longrightarrow\quad \tfrac{3}{2}gc = 3gc\sin\theta \quad\Longrightarrow\quad \sin\theta = \tfrac{1}{2}.$$

所以繩子在 $\theta = 30^\circ$ 時變鬆，位置是 B 那一側的點 P，離桌面的高度為 $c\sin\theta = c/2$，正如題目所求。

A 與 B 之間的第一次反彈

在 P 點有 $\theta = 30^\circ$，所以 $P = \left(-\frac{\sqrt{3}}{2}c,\ \frac{1}{2}c\right)$，並且

$$v^2 = \tfrac{3}{2}gc - 2gc\cdot\tfrac{1}{2} = \tfrac{1}{2}gc.$$

速度沿切線方向，與 OP 垂直，指向上方並朝向 O 的一邊；它與水平方向成 60° 角。所以它的分量為

$$v_x = \tfrac{1}{2}v, \qquad v_y = \tfrac{\sqrt{3}}{2}v, \qquad \text{其中} v_x^2 = \tfrac{1}{8}gc, \quad v_y^2 = \tfrac{3}{8}gc.$$

從這時起，質點就是一個拋體。當 $\frac{1}{2}c + v_y t - \frac{1}{2}gt^2 = 0$ 時它到達桌面；取正根，

$$t = \frac{v_y + \sqrt{v_y^2 + gc}}{g} = \frac{\sqrt{3}+\sqrt{11}}{\sqrt{8}}\sqrt{\frac{c}{g}}.$$

在這段時間內，它的水平位移為

$$v_x t = \sqrt{\frac{gc}{8}} \cdot \frac{\sqrt{3}+\sqrt{11}}{\sqrt{8}}\sqrt{\frac{c}{g}} = \frac{\sqrt{3}+\sqrt{11}}{8}\,c,$$

所以它落在點 L，其中

$$x_L = -\frac{\sqrt{3}}{2}c + \frac{\sqrt{3}+\sqrt{11}}{8}c = \frac{\sqrt{11}-3\sqrt{3}}{8}\,c \approx -0.235\,c.$$

（在整段飛行過程中，質點與 O 的距離始終不超過 c，所以繩子在到達 L 之前不會再次拉緊；特別地，$|x_L| < c$。）

由於 $\sqrt{11} < \sqrt{27} = 3\sqrt{3}$，我們有 $x_L < 0$：落點與 B 位於 O 的同一側。因為 O 是 AB 的中點，所以 A 與 B 之間的第一次反彈**較接近** B，與 B 相距約 $0.77c$，與 A 相距約 $1.23c$。

驗算

- 以 $V_k^2 \geq 3gc$ 作檢驗，可知 $k = 7$ 是第一個走不完的半圓，而 $V_7^2 = \frac{3}{2}gc$；繩子變鬆的條件給出 $\sin\theta = \frac{1}{2}$。
- 在 P 點，重力的徑向分量 $g\sin 30^\circ = \frac{1}{2}g$ 等於 $v^2/c = \frac{1}{2}g$，正好符合張力為零的要求。
- 逐步進行的數值模擬（取 $g = 9.8$，$c = 1.3$）顯示，繩子在高度 $0.5000c$ 處變鬆，落點在 $x \approx -0.2349c$，而飛行期間質點與 O 的距離從未超過 c。

常見錯誤

- 以 $V^2 \geq 4gc$（剛好到達頂點）作為走完半圓的條件。繩子與桿不同，需要 $V^2 \geq 3gc$。
- 弄不清第七次碰撞發生在哪一端：第奇數次碰撞都在 B。
- 把 P 點速度的方向弄錯：它與 OP 垂直，與水平方向成 60° 角，而不是 30°。

STEP 之外

繩子變鬆的一刻，就是一個約束（constraint）停止作用的一刻：張力正是用來維持 $|OP| = c$ 的拉格朗日乘數（Lagrange multiplier），當它必須變為負數才能維持約束

時，這個約束便不再起作用。像這樣帶有單邊約束（one-sided constraint）以及碰撞的問題，在大學本科的力學課程中會有系統地處理。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）是個好去處，可以看到能量與徑向方程怎樣這樣配合使用。

詳細解答

STEP III 1993，第 15 題

概率與統計 • 等候時間的分佈

> **STEP III 1993，第 15 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/93-s3-q15

關鍵思路

把注意力放在最後一次投擲上。實驗恰好在第 n 次投擲停止，當且僅當這一次投擲補上了仍然欠缺的那一面的第二次出現，因此首 $n-1$ 次投擲必須恰好有一次是該面，其餘全部是另一面。這樣，分佈（distribution）就是兩個簡單項之和。期望值（expectation）於是化為兩個形如 $\sum n(n-1)x^{n-2}$ 的級數，而這類級數可由等比級數（geometric series）求導兩次得到。

解答

我們假設 $0 < p < 1$（否則實驗永遠不會結束），並假設各次投擲互相獨立（independent）。

N 的分佈

我們至少需要兩次正面和兩次反面，所以 $N \geq 4$。固定 $n \geq 4$。在事件 $N = n$ 上，第 n 次投擲不是正面就是反面，而這兩種情況互斥（disjoint）。

*第 n 次投擲是正面。*在此之前目標尚未達到，而第 n 次投擲添上一次正面後便達到目標。所以首 $n-1$ 次投擲已經包含至少兩次反面，但只有一次正面：恰好一次正面和 $n-2$ 次反面（而且 $n-2 \geq 2$，因為 $n \geq 4$）。反過來說，任何這樣的序列後面接上一次正面，都恰好在第 n 次停止。那唯一的一次正面可以出現在 $n-1$ 個位置中的任何一個，所以這種情況的概率是

$$(n-1)\,p\,q^{n-2} \cdot p = (n-1)p^2q^{n-2}.$$

*第 n 次投擲是反面。*把正面和反面對調，用同樣的論證可知概率是 $(n-1)q^2p^{n-2}$。

兩者相加，

$$\boxed{\mathrm{P}(N=n)=(n-1)\big(p^2q^{n-2}+q^2p^{n-2}\big),\qquad n=4,5,6,\ldots}$$

而當 $n\le 3$ 時 $\mathrm{P}(N=n)=0$。

想一想 當最後一次投擲是正面時，為甚麼首 $n-1$ 次投擲不可能包含兩次正面？

假如包含兩次正面，那麼它們已經有兩次正面和（停止條件所需的）兩次反面，實驗便會在第 n 次投擲之前結束。正是這一點把正面的次數確定為恰好一次。

期望值

對於 $|x|<1$，把 $\sum_{n\ge 0}x^n=(1-x)^{-1}$ 求導兩次，得

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

我們的和式由 $n=4$ 開始，所以要減去 $n=2$ 和 $n=3$ 的兩項，即 2 和 $6x$：

$$\sum_{n=4}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}-2-6x.$$

現在

$$\mathrm{E}(N)=\sum_{n=4}^{\infty}n\,\mathrm{P}(N=n)=p^2\sum_{n=4}^{\infty}n(n-1)q^{n-2}+q^2\sum_{n=4}^{\infty}n(n-1)p^{n-2}.$$

先取 $x=q$（於是 $1-x=p$），再取 $x=p$（於是 $1-x=q$）代入上述級數；由於 $0<p,q<1$，兩者皆成立：

$$\begin{aligned}\mathrm{E}(N)&=p^2\Big(\frac{2}{p^3}-2-6q\Big)+q^2\Big(\frac{2}{q^3}-2-6p\Big)\\&=\frac{2}{p}+\frac{2}{q}-2(p^2+q^2)-6pq(p+q).\end{aligned}$$

由於 $p+q=1$，我們有 $\frac{1}{p}+\frac{1}{q}=\frac{p+q}{pq}=\frac{1}{pq}$ 以及 $p^2+q^2=(p+q)^2-2pq=1-2pq$。因此

$$\begin{aligned}\mathrm{E}(N)&=\frac{2}{pq}-2(1-2pq)-6pq\\&=\frac{2}{pq}-2-2pq=2\Big(\frac{1}{pq}-1-pq\Big),\end{aligned}$$

這正是所要證明的結果。

驗算

- 各概率之和為 1：同一個級數少一個因子的版本給出 $\sum_{n\geq 4}(n-1)p^2q^{n-2} = 1-p^2-2p^2q$，加上對調後的版本，餘下 $1-(p^2+q^2)-2pq(p+q) = 1-1+2pq-2pq = 1$，當中用到 $p^2+q^2 = 1-2pq$。
- 對於公平的硬幣，$pq = \frac{1}{4}$，而 $\mathrm{E}(N) = 2(4-1-\frac{1}{4}) = \frac{11}{2}$。直接計算：$\mathrm{P}(N=4) = \frac{3}{8}$，$\mathrm{P}(N=5) = \frac{1}{4}$，...，而 $n\,\mathrm{P}(N=n)$ 的部分和趨近 5.5。
- 當 $p \to 0$ 時，$\mathrm{E}(N) \approx 2/p$：我們要等待兩次正面出現，每次大約需要 $1/p$ 次投擲。
- 列舉所有長度不超過 10 的序列，所得結果與這個分佈完全吻合；另外以 $p = 0.3$ 進行模擬，得出的平均值為 7.10，而公式給出的是 7.1038。

常見錯誤

- 把 $n-1$ 寫成 $\binom{n}{2}$：最後一次投擲的位置是固定的，所以只有首 $n-1$ 次投擲中那唯一的一次正面（或反面）可以自由選擇位置。
- 使用標準級數時，忘記減去 $n = 2$ 和 $n = 3$ 的兩項。

STEP 之外

N 是兩個經平移的負二項（negative binomial）等待時間的混合，而對等比級數求導這個技巧，其實是概率生成函數（probability generating function）的手工版本。設 $G(s) = \mathrm{E}(s^N)$，便可同樣快捷地由 $G'(1)$ 得到平均值，並由 $G''(1)$ 得到方差（variance）。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門）詳細講述生成函數，以及像本題這類的等待時間問題。

詳細解答

STEP III 1993，第 16 題

概率與統計 • 截尾正態分佈

STEP III 1993，第 16 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/93-s3-q16

關鍵思路

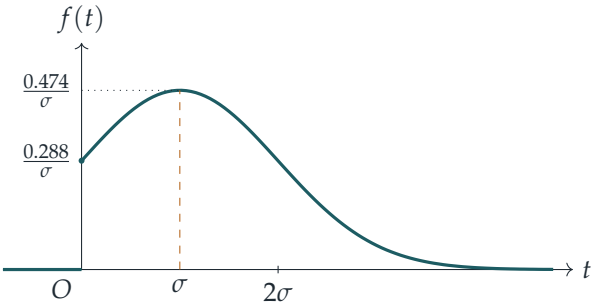
除了因子 $1/k$ 之外，這個密度函數正是平均值為 σ、標準差為 σ 的正態分佈（normal distribution）的密度函數，只是在 $t = 0$ 處被截斷。因此，凡是關於 T 的概率，都等於一個正態概率除以「未截斷的變量為非負」的概率。以 $z = (t - \sigma)/\sigma$ 作標準化（standardising），一切便化為關於 $Z \sim \mathrm{N}(0,1)$ 的命題，可查 Φ 的數值表求得。

解答

以下一律設 $Z \sim \mathrm{N}(0,1)$，其密度函數（density）為 $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$，分佈函數（distribution function）為 Φ。

略圖

當 $t \geq 0$ 時，圖像是以 $t = \sigma$ 為中心的正態鐘形曲線，再按 $1/k$ 的比例放大。它在 $t = \sigma$ 處取得極大值，$f(\sigma) = \frac{1}{k\sigma\sqrt{2\pi}} \approx \frac{0.474}{\sigma}$，並在 $t = 2\sigma$ 處有一個拐點（point of inflexion）。在截斷處，曲線由一個正的高度開始，$f(0) = \frac{e^{-1/2}}{k\sigma\sqrt{2\pi}} \approx \frac{0.288}{\sigma}$，而當 $t < 0$ 時則跳至 0。當 $t \to \infty$ 時，$f(t) \to 0$。



k 的值

總概率必須等於 1：

$$1 = \int_0^\infty \frac{1}{k\sigma\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}\left(\tfrac{t-\sigma}{\sigma}\right)^2\right) \mathrm{d}t.$$

代入 $z = (t-\sigma)/\sigma$，則 $\mathrm{d}t = \sigma\,\mathrm{d}z$，而 $t = 0$ 變成 $z = -1$，於是

$$1 = \frac{1}{k}\int_{-1}^\infty \phi(z)\,\mathrm{d}z = \frac{1-\Phi(-1)}{k} = \frac{\Phi(1)}{k}.$$

因此 $k = \Phi(1) \approx 0.8413$，即 $k \approx 0.841$，正如題目所求。

平均值

用同一個代換，$t = \sigma(1+z)$，於是

$$\mathrm{E}(T) = \frac{1}{k}\int_{-1}^\infty \sigma(1+z)\,\phi(z)\,\mathrm{d}z = \sigma\left[\frac{1}{k}\int_{-1}^\infty \phi(z)\,\mathrm{d}z + \frac{1}{k}\int_{-1}^\infty z\,\phi(z)\,\mathrm{d}z\right].$$

第一個積分等於 k，這是剛才已經證明的。

> **想一想** 既然 $\phi(z)$ 本身也積不出來，我們怎樣對 $z\,\phi(z)$ 求積分？

我們利用 $\phi'(z) = -z\,\phi(z)$ 這個事實，所以 $z\,\phi(z)$ 的原函數（antiderivative）是 $-\phi(z)$：

$$\int_{-1}^\infty z\,\phi(z)\,\mathrm{d}z = \Big[-\phi(z)\Big]_{-1}^\infty = \phi(-1) = \frac{e^{-1/2}}{\sqrt{2\pi}} \approx 0.24197.$$

因此

$$\mathrm{E}(T) = \sigma\left(1 + \frac{e^{-1/2}}{k\sqrt{2\pi}}\right) \approx \sigma\left(1 + \frac{0.24197}{0.84134}\right) \approx 1.288\,\sigma.$$

這個值比 σ 大，也理應如此：除去左尾之後，平均值會向右移。

求 σ

擬題時間的平均值是 3 小時，所以

$$\sigma = \frac{3}{1.2876} \approx 2.330 \text{ 小時}.$$

對任意 $a \geq 0$，用同一個代換可得

$$\mathrm{P}(T > a) = \frac{1}{k}\int_{(a-\sigma)/\sigma}^{\infty} \phi(z)\,\mathrm{d}z = \frac{1-\Phi\left(\frac{a}{\sigma}-1\right)}{\Phi(1)}. \qquad (*)$$

第 (i) 部分

取 $a = 4$：$\frac{4}{\sigma} - 1 \approx \frac{4}{2.330} - 1 \approx 0.717$，而 $\Phi(0.717) \approx 0.7633$。由 $(*)$，

$$\mathrm{P}(T > 4) \approx \frac{1 - 0.7633}{0.8413} \approx 0.2814, \qquad \mathrm{P}(T \leq 4) \approx 0.7186.$$

現在我們假設 16 道題目的擬題時間互相獨立（independent），而且各自服從 T 的分佈（這是唯一合理而可用的模型）。那麼

$$\mathrm{P}(\text{沒有題目需時超過 } 4 \text{ 小時}) = \mathrm{P}(T \leq 4)^{16} \approx 0.7186^{16} \approx 0.0051.$$

所以，每道題目都能在 4 小時內擬好的機會極微（大約二百分之一）。

第 (ii) 部分

> **想一想** 我們是以哪一個事件作為條件？

「2 小時後仍未令人滿意」表示該題目總共需要多於 2 小時，所以這是事件 $T > 2$。「再過 2 小時後仍然不獲接納」則是 $T > 4$。由於 $\{T > 4\} \subseteq \{T > 2\}$，

$$\mathrm{P}(T > 4 \mid T > 2) = \frac{\mathrm{P}(T > 4)}{\mathrm{P}(T > 2)}.$$

取 $a = 2$：$\frac{2}{\sigma} - 1 \approx -0.142$，所以 $1 - \Phi(-0.142) = \Phi(0.142) \approx 0.5563$。把兩個 $(*)$ 式相除時，因子 $\Phi(1)$ 互相抵消：

$$\mathrm{P}(T > 4 \mid T > 2) \approx \frac{1 - 0.7633}{0.5563} = \frac{0.2367}{0.5563} \approx 0.426.$$

驗算

- 直接對密度函數作數值積分（辛普森法則（Simpson's rule），取 $\sigma = 1$），在除以 k 之前得總面積為 0.84134，平均值為 1.2876σ。
- 模擬 4×10^5 個平均值和標準差均為 $\sigma = 2.330$ 的正態值，並捨去其中的負值，得樣本平均值 3.004，$\mathrm{P}(T > 4) \approx 0.282$ 及 $\mathrm{P}(T > 4 \mid T > 2) \approx 0.426$。
- 第 (i) 部分對捨入非常敏感：$0.719^{16} \approx 0.0051$，但 $0.718^{16} \approx 0.0050$。答案寫作約 0.005 即可。
- 即使沒有截斷，$\mathrm{P}(T > 4 \mid T > 2)$ 也會相同，因為因子 $1/k$ 會抵消。只有第 (i) 部分和平均值才受 k 影響。

常見錯誤

- 計算 $\mathrm{P}(T > 4)$ 時忘記除以 k：這樣會得到 0.237 而不是 0.281，並在 (i) 中得到 $0.7633^{16} \approx 0.013$，是正確值的兩倍有多。
- 把 T 的平均值當作 σ（即截斷前的平均值），這樣會得到 $\sigma = 3$。
- 在 (ii) 中求了 $\mathrm{P}(2 < T \leq 4)$ 或 $\mathrm{P}(T > 2)$，而不是條件概率（conditional probability）。

STEP 之外

平均值中出現的量 $\phi(1)/\Phi(1)$ 是*逆米爾斯比率*（inverse Mills ratio）的一個例子：一般而言，正態分佈 $\mathrm{N}(\mu, \sigma^2)$ 在 a 處截去下方之後，其平均值為 $\mu + \sigma\phi(\alpha)/(1 - \Phi(\alpha))$，其中 $\alpha = (a - \mu)/\sigma$。凡是數據只能在某個門檻以上才觀察得到的場合，都會出現截斷分佈（truncated distribution），例如存活分析（survival analysis）以及處理設限數據（censored data）的計量經濟模型。第 (ii) 部分所討論的是*存活函數*（survival function）$\mathrm{P}(T > t)$；指數分佈（exponential distribution）是唯一能令 $\mathrm{P}(T > s + t \mid T > s)$ 與 s 無關的情況。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）涵蓋連續分佈和條件概率。

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

1993 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。