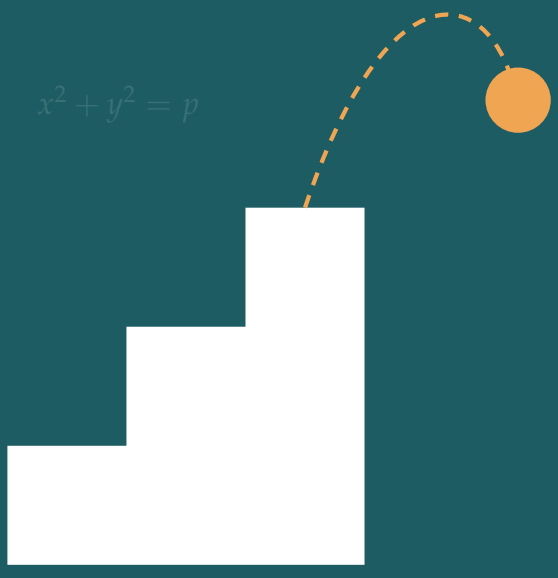
STEP 備試 · 大學數學

Beyond STEP

STEP I, II 及 III

2002 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

2002 年詳細解答

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

2002 年詳細解答

2002 年 STEP 試卷（STEP I, II 及 III）全部 42 題的詳細解答。自樣本試卷以來每一道 STEP 題目的詳細解答、*Beyond STEP* 全書以及各章的單行本，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

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

由機器從英文翻譯而成；尚未經審閱。

2026 年 11 月首次於網上發表。

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

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

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

目錄

STEP I

第 1 題 圓與橢圓 純數學 . . . 3

第 2 題 曲線描繪 純數學 . . . 6

第 3 題 以微積分證明不等式 純數學 . . . 10

第 4 題 切線與不等式 純數學 . . . 14

第 5 題 多項式的整數根 純數學 . . . 18

第 6 題 角錐的幾何 純數學 . . . 21

第 7 題 三角函數積分 純數學 . . . 25

第 8 題 複利貸款 純數學 . . . 29

第 9 題 斜面上的傾倒 力學 . . . 32

第 10 題 彈性碰撞 力學 . . . 36

第 11 題 彈性斜碰撞 力學 . . . 40

第 12 題 幾何概率 概率與統計 . . . 43

第 13 題 條件期望值 概率與統計 . . . 47

第 14 題 概率中的等比級數 概率與統計 . . . 50

STEP II

第 1 題 代換積分法 純數學 . . . 53

第 2 題 倒數多項式方程 純數學 . . . 56

第 3 題 費馬數與質數 純數學 . . . 60

第 4 題 由積分推導不等式 純數學 . . . 63

第 5 題 羅輯斯諦映射的迭代 純數學 . . . 66

第 6 題 斜面上的直線 純數學 . . . 70

第 7 題 直線的向量幾何 純數學 . . . 73

第 8 題 一階微分方程 純數學 . . . 77

第 9 題 拋體運動 力學 . . . 81
第 10 題 運動學與最優化 力學 . . . 85
第 11 題 摩擦力與極限平衡 力學 . . . 89
第 12 題 順態近似與泊松近似 概率與統計 . . . 93
第 13 題 變換後的分佈函數 概率與統計 . . . 97
第 14 題 混合分佈與貝葉斯定理 概率與統計 . . . 100

STEP III

第 1 題 廣義積分與體積 純數學 . . . 104
第 2 題 反正切級數 純數學 . . . 107
第 3 題 曲線描繪 純數學 . . . 110
第 4 題 丟番圖不等式 純數學 . . . 114
第 5 題 四次方程的因式分解 純數學 . . . 117
第 6 題 非線性微分方程 純數學 . . . 120
第 7 題 坐標幾何恆等式 純數學 . . . 124
第 8 題 托勒密定理 純數學 . . . 128
第 9 題 質心與傾倒 力學 . . . 132
第 10 題 滑及摩擦力的轉動 力學 . . . 136
第 11 題 牆壁的斜碰撞 力學 . . . 140
第 12 題 條件化求期望值 概率與統計 . . . 145
第 13 題 離數分佈變量的最小值 概率與統計 . . . 149
第 14 題 差與協方差 概率與統計 . . . 152

詳細解答

STEP I 2002，第 1 題

純數學 • 圓與橢圓

> **STEP I 2002，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/02-s1-q1

關鍵思路

先找出兩個橢圓（ellipse）實際相交之處。兩條方程都只透過 y^2 含有 y，所以消去 y^2 後便得到一條 x 的二次方程，而結果顯示兩個橢圓只相交於兩個實點。通過兩個已知點的圓，其圓心必在這兩點的垂直平分線（perpendicular bisector）上，這樣便把圓的方程確定為一個單參數族（one-parameter family）。

解答

求交點

展開後，兩個橢圓的方程為

$$x^2 + 4x + 2y^2 = 14,$$
$$9x^2 - 18x + 16y^2 = 16.$$

把第一式乘以 8，再減去第二式，便可消去 y^2：

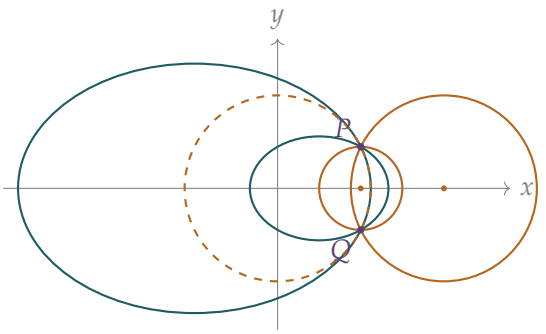
$$8x^2 + 32x - 9x^2 + 18x = 112 - 16 \implies x^2 - 50x + 96 = 0 \implies (x-2)(x-48) = 0.$$

> ? **想一想** 為甚麼不能把兩個根都保留？

每個根仍須給出實數的 y。由第一個橢圓得 $2y^2 = 14 - x^2 - 4x$。

- $x = 48$：$2y^2 = 14 - 2304 - 192 < 0$，所以沒有實點。（無論如何，第一個橢圓只涵蓋 $-2 - 3\sqrt{2} \le x \le -2 + 3\sqrt{2}$。）
- $x = 2$：$2y^2 = 14 - 4 - 8 = 2$，所以 $y = \pm 1$。

因此兩個橢圓恰好相交於兩點，即 $P = (2,1)$ 和 $Q = (2,-1)$。（兩點亦滿足第二條方程：$9 + 16 = 25$。）



圖中所示為兩個橢圓（藍綠色），以及通過 P 和 Q 的其中三個圓（琥珀色），圓心分別為 $(0,0)$、$(2,0)$ 和 $(4,0)$。

通過 P 和 Q 的圓

任何通過 P 和 Q 的圓，其圓心與這兩點等距，所以圓心位於 PQ 的垂直平分線上。線段 PQ 是鉛垂的，中點為 $(2,0)$，所以它的垂直平分線就是 x 軸。因此圓心為 $(a,0)$，其中 a 為某個實數，而半徑滿足

$$r^2 = (2-a)^2 + 1^2.$$

於是圓的方程為

$$\begin{aligned}(x-a)^2 + y^2 &= (2-a)^2 + 1\\ x^2 - 2ax + a^2 + y^2 &= 4 - 4a + a^2 + 1\\ x^2 - 2ax + y^2 &= 5 - 4a,\end{aligned}$$

這正是所求的形式。

反過來說，對每個實數 a，這條方程都表示一個真正的圓，圓心為 $(a,0)$，半徑為 $\sqrt{(a-2)^2+1} > 0$，而且它通過 $(2,\pm 1)$，因為 $4 - 4a + 1 = 5 - 4a$。所以這個圓族恰好就是通過兩個交點的所有圓所組成的集合。

想一想 那麼，把兩條方程組合起來這個慣用技巧又如何？

取 $7\times$（第一式）$-$（第二式），可令 x^2 和 y^2 的係數相等，得出 $x^2 - 23x + y^2 + 41 = 0$，這是我們的圓族中 $a = \frac{23}{2}$ 的那一個成員。但這只是一個圓。兩條方程的組合，是通過全部四個（複數）公共點的圓錐曲線（conic），當中包括 $x = 48$ 的兩個非實點；而一個圓只需通過那兩個實點。這就是本題必須採用直接方法的原因。

驗算

- $a = 2$ 給出 $x^2 - 4x + y^2 = -3$，即 $(x-2)^2 + y^2 = 1$：這是以 PQ 為直徑的圓，正如所料。
- $a = 0$ 給出 $x^2 + y^2 = 5$，而確實 $2^2 + 1^2 = 5$。
- 在 $-6 \leq x \leq 6$ 的範圍內以數值方法搜尋同時位於兩個橢圓上的點，只找到 $x = 2$。在 $(2,1)$ 處，兩個橢圓的法線（normal）方向分別為 $(8,4)$ 和 $(18,32)$，兩者並不平行，所以兩個橢圓在該點是真正相交，而不是相切。

常見錯誤

- 保留 $x = 48$，卻沒有檢查 y 是否實數。
- 由兩條橢圓方程的組合得出單獨一個圓後便停下來：題目問的是通過這些點的*任何*圓。

STEP 之外

兩條圓錐曲線的組合 $\lambda E_1 + \mu E_2 = 0$ 稱為*曲線束*（pencil）：它由通過兩者四個公共點的圓錐曲線組成，而這些公共點是在複射影平面（complex projective plane）上計算的。這道題目正好提醒我們，這些點當中有些可能是非實點，而圓是特殊的（所有圓都通過兩個固定的複數點，即「無窮遠圓點」（circular points at infinity））。這些概念屬於射影幾何（projective geometry）和代數幾何（algebraic geometry）的範疇；若想以這種觀點初步認識平面曲線和圓錐曲線，可參閱 Reid 的 *Undergraduate Algebraic Geometry*（一本平易近人的入門書），而 Siklos 的 *Advanced Problems in Mathematics* 則有更多 STEP 風格的練習。

詳細解答

STEP I 2002，第 2 題

純數學 • 曲線描繪

> **STEP I 2002，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/02-s1-q2

關鍵思路

多項式 $x^m(x-1)^n$ 在 $x = 0$ 和 $x = 1$ 處為零，所以曲線在兩者之間必定要轉向。用乘積法則（product rule）求導，並且把所有公因式（*common factor*）都提出來：導數（derivative）於是分解成幾個簡單的部分，各部分的正負號一目了然。整個草圖由奇偶性（parity）決定：某個冪是偶數還是奇數，決定了曲線是與軸相切還是穿過軸，也決定了曲線在軸的哪一側。

解答

駐點

由乘積法則，

$$\begin{aligned} \mathrm{f}'(x) &= mx^{m-1}(x-1)^n + nx^m(x-1)^{n-1} \\ &= x^{m-1}(x-1)^{n-1}\big[m(x-1) + nx\big] \\ &= x^{m-1}(x-1)^{n-1}\big[(m+n)x - m\big]. \end{aligned}$$

所以 $\mathrm{f}'(x) = 0$ 當且僅當 $x = 0$、$x = 1$ 或

$$x = x_0 = \frac{m}{m+n}.$$

由於 $m, n > 0$，我們有 $0 < m < m+n$，所以 $0 < x_0 < 1$。這就是所求的駐點(stationary point)。(由於 $m, n \geq 2$，冪 $m-1$ 和 $n-1$ 都至少是 1，所以 $x = 0$ 和 $x = 1$ 也是駐點。畫草圖時會用到它們。)

駐點的性質

把導數寫成 $\mathrm{f}'(x) = g(x)\,h(x)$，其中 $g(x) = x^{m-1}(x-1)^{n-1}$，$h(x) = (m+n)x - m$。那麼

$$\mathrm{f}''(x) = g'(x)h(x) + g(x)h'(x) = g'(x)h(x) + (m+n)g(x).$$

> **?** **想一想** 我們真的需要把 g 完整地求導嗎？

不需要。在 x_0 處我們有 $h(x_0) = 0$，所以無論 g' 是甚麼，第一項都會消失：

$$\mathrm{f}''(x_0) = (m+n)\,x_0^{m-1}\,(x_0-1)^{n-1}.$$

這裏 $m+n > 0$，$x_0^{m-1} > 0$，但 $x_0 - 1 < 0$。所以 $\mathrm{f}''(x_0)$ 不為零，其正負號與 $(-1)^{n-1}$ 相同：

- 若 n 是偶數，則 $n-1$ 是奇數，$\mathrm{f}''(x_0) < 0$，x_0 是**極大點（maximum）**；
- 若 n 是奇數，則 $n-1$ 是偶數，$\mathrm{f}''(x_0) > 0$，x_0 是**極小點（minimum）**。

因為 $\mathrm{f}''(x_0) \neq 0$，二階導數判別法（second-derivative test）可以下定論。這正是我們要證明的。

草圖

四種情況由 m 和 n 的奇偶性決定。要畫出每一種情況，我們先整理以下事實。

*f 的正負號。*當 $x > 1$ 時，兩個因式都是正的，所以 $\mathrm{f} > 0$。當 $0 < x < 1$ 時，$x^m > 0$，而 $(x-1)^n$ 的正負號與 $(-1)^n$ 相同。當 $x < 0$ 時，正負號是 $(-1)^m(-1)^n = (-1)^{m+n}$。

*在根（root）附近的表現。*在 $x = 0$ 附近，我們有 $\mathrm{f}(x) \approx (-1)^n x^m$。若 m 是偶數，f 在 0 的兩側保持同號，所以曲線在該處與軸相切（一個轉向點（turning point））。若 m 是奇數，f 變號，又因為 $\mathrm{f}'(0) = 0$，曲線會水平地穿過軸（一個駐拐點（stationary point of inflection））。把同樣的推理用於 n，便得到 $x = 1$ 處的情況。

*兩端。*當 $|x|$ 很大時，$\mathrm{f}(x)$ 的表現與 x^{m+n} 相似，所以當 $x \to +\infty$ 時 $\mathrm{f} \to +\infty$。當 $x \to -\infty$ 時，若 $m+n$ 是偶數則 $\mathrm{f} \to +\infty$，若 $m+n$ 是奇數則 $\mathrm{f} \to -\infty$。

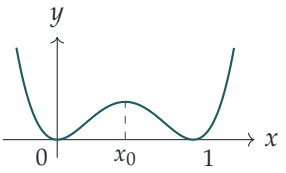
*沒有其他駐點。*駐點只有 $x = 0$、x_0 和 $x = 1$。

綜合以上各點：

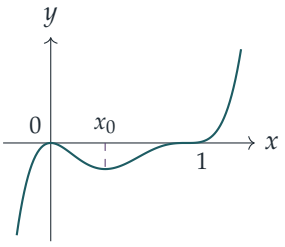
- **m 為偶數，n 為偶數：**處處皆有 $\mathrm{f} \geq 0$。在 0 和 1 處有位於軸上的極小點，在 x_0 處有一個極大點：呈 W 形。
- **m 為偶數，n 為奇數：**曲線由 $-\infty$ 上升，在 0 處從下方與軸相切（一個極大點），下降至 x_0 處的極小點，然後上升，經過 1 處的駐拐點，一直升向 $+\infty$。

- **m 為奇數，n 為偶數：**曲線由 $-\infty$ 上升，經過 0 處的駐拐點，到達 x_0 處的極大點，下降至 1 處與軸相切（一個極小點），然後升向 $+\infty$。
- **m 為奇數，n 為奇數：**曲線由 $+\infty$ 下降，經過 0 處的駐拐點，到達 x_0 處的極小點，再回升經過 1 處的駐拐點，一直升向 $+\infty$。

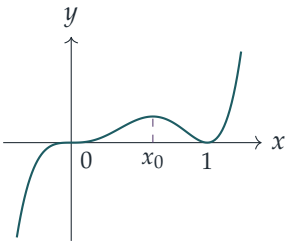
以下草圖採用最小的情況 $(m,n) = (2,2),(2,3),(3,2),(3,3)$，每幅圖各有自己的縱軸比例。在 x_0 處的值都很小：分別是 $\frac{1}{16}$、$-\frac{108}{3125}$、$\frac{108}{3125}$ 和 $-\frac{1}{64}$。



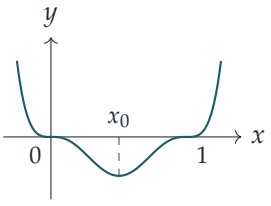
m 為偶數，n 為偶數



m 為偶數，n 為奇數



m 為奇數，n 為偶數



m 為奇數，n 為奇數

驗算

- 取 $m = n = 2$：$\mathrm{f}(x) = \big(x(x-1)\big)^2$ 關於 $x = \frac{1}{2}$ 對稱，而 $x_0 = \frac{1}{2}$ 確實是極大點，其值為 $\frac{1}{16}$。
- 代換 $x \mapsto 1-x$ 把 $x^m(x-1)^n$ 變成 $(-1)^{m+n}x^n(x-1)^m$。所以把 m 和 n 對調，相當於把圖像沿 $x = \frac{1}{2}$ 反射（若 $m+n$ 是奇數，還要再沿 x 軸反射）。這就是兩幅奇偶性相異的草圖互為鏡像的原因。
- 我們把 f' 與數值導數作了比對，又把 $\mathrm{f}''(x_0)$ 的正負號與有限差分（finite difference）作了比對，範圍涵蓋所有 $2 \le m,n \le 8$。例如當 $(m,n) = (2,3)$ 時，$\mathrm{f}''(\frac{2}{5}) = 5 \cdot \frac{2}{5} \cdot \frac{9}{25} = 0.72 > 0$，是極小點，與預測相符。

常見錯誤

- 把 $f''(x)$ 完整展開。在 x_0 處，因式 $(m+n)x - m$ 為零，所以只有一項留下來。
- 忘記 $x = 0$ 和 $x = 1$ 也是駐點（因為 $m, n \geq 2$），結果把曲線畫成在該處斜斜地穿過軸。
- 弄錯 $x < 0$ 時的正負號。它取決於 $m + n$，而不是單單取決於 m。

STEP 之外

多項式在重數（multiplicity）為 k 的根附近，局部形狀與 $c(x-a)^k$ 相似，而 k 的奇偶性決定了曲線是相切還是穿過。這是邁向泰勒定理（Taylor's theorem）和駐點的高階導數判別法的第一步：若駐點處第一個不為零的導數是 k 階的，則當 k 是偶數時該點是轉向點，當 k 是奇數時該點是拐點。Abbott 的 *Understanding Analysis* 為這些想法建立了嚴格的基礎。

詳細解答

STEP I 2002，第 3 題

純數學 • 以微積分證明不等式

> **STEP I 2002，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/02-s1-q3

關鍵思路

把兩個平方根記作 $\sqrt{P}$ 和 $\sqrt{Q}$。根號下的兩個量關係密切：$P+Q=a^2+b^2$ 是常數，而且它們的導數（derivative）大小相等、正負相反。因此 $\mathrm{d}y/\mathrm{d}\theta$ 可以乾淨利落地因式分解（factorise），其正負只由一個三角因子決定。只要知道 y 的最大值和最小值，那個雙重不等式（double inequality）不過就是「最小值 $\le y \le$ 最大值」，而第一部分則告訴我們兩端的大小次序正確。

解答

第一個不等式

展開得

$$2a^2+2b^2-(a+b)^2=a^2-2ab+b^2=(a-b)^2\ge 0,$$

所以 $(a+b)^2\le 2a^2+2b^2$，等號成立當且僅當 $a=b$。

駐點

設 $c=\cos\theta$，$s=\sin\theta$，並且

$$P=a^2c^2+b^2s^2,\qquad Q=a^2s^2+b^2c^2,\qquad y=\sqrt{P}+\sqrt{Q}.$$

先看兩個特殊情況。若 $a^2=b^2$，則對每一個 θ 都有 $P=Q=a^2$，所以 $y=2|a|$ 是常數，每一點都（平凡地）是駐點（stationary point）；這個情況暫且擱下。若 $a=0$ 或 $b=0$（但不同時為零），則 P 或 Q 會在某些 θ 處等於零，而 y 在該處不可微（not differentiable）；這個情況留待最後處理。所以現在先假設 $ab\neq 0$ 且 $a^2\neq b^2$。於是

$P \geq \min(a^2, b^2)(c^2 + s^2) > 0$，同理 $Q > 0$，所以 y 處處可微。

求導得

$$\frac{\mathrm{d}P}{\mathrm{d}\theta} = -2a^2cs + 2b^2sc = (b^2 - a^2)\sin 2\theta, \qquad \frac{\mathrm{d}Q}{\mathrm{d}\theta} = (a^2 - b^2)\sin 2\theta.$$

因此

$$\frac{\mathrm{d}y}{\mathrm{d}\theta} = \frac{(b^2 - a^2)\sin 2\theta}{2\sqrt{P}} - \frac{(b^2 - a^2)\sin 2\theta}{2\sqrt{Q}} = \frac{(b^2 - a^2)\sin 2\theta\left(\sqrt{Q} - \sqrt{P}\right)}{2\sqrt{PQ}}.$$

要看出 $\sqrt{Q} - \sqrt{P}$ 的正負，可以把它有理化（rationalise）：

$$\sqrt{Q} - \sqrt{P} = \frac{Q - P}{\sqrt{Q} + \sqrt{P}}, \qquad Q - P = (a^2 - b^2)(s^2 - c^2) = (b^2 - a^2)\cos 2\theta.$$

代入上式，並利用 $2\sin 2\theta \cos 2\theta = \sin 4\theta$，得

$$\frac{\mathrm{d}y}{\mathrm{d}\theta} = \frac{(a^2 - b^2)^2 \sin 4\theta}{4\sqrt{PQ}\left(\sqrt{P} + \sqrt{Q}\right)}.$$

想一想 為甚麼值得花工夫把 $\sqrt{Q} - \sqrt{P}$ 有理化？

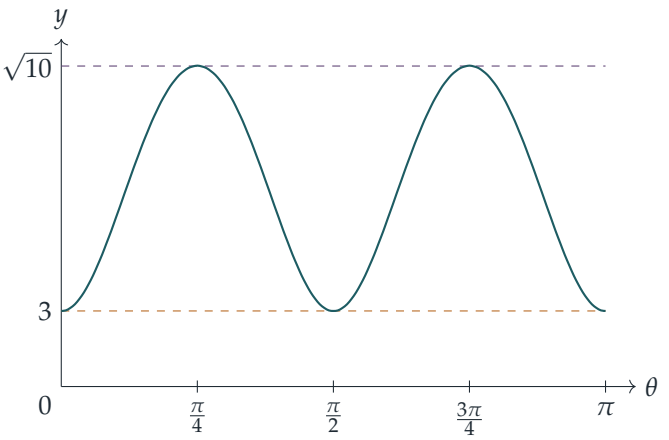
因為這樣一來，除了 $\sin 4\theta$ 之外，其餘各項一眼就看得出是正數：$(a^2 - b^2)^2 > 0$，而 P、Q 都是正數。所以 $\mathrm{d}y/\mathrm{d}\theta$ 的正負與 $\sin 4\theta$ 完全相同，這既告訴我們駐點在哪裏，也告訴我們它們屬於哪一類。

駐點出現在 $\sin 4\theta = 0$ 之處，即 $\theta = \frac{1}{4}n\pi$，其中 n 為整數。

- 當 n 為偶數時，$\theta = \frac{1}{2}k\pi$，所以 $\{c^2, s^2\} = \{1, 0\}$，$\{P, Q\} = \{a^2, b^2\}$。此時 $y = |a| + |b|$（記住 $\sqrt{a^2} = |a|$）。
- 當 n 為奇數時，$\theta = \frac{1}{4}\pi + \frac{1}{2}k\pi$，所以 $c^2 = s^2 = \frac{1}{2}$，$P = Q = \frac{1}{2}(a^2 + b^2)$。此時 $y = 2\sqrt{\frac{1}{2}(a^2 + b^2)} = \sqrt{2a^2 + 2b^2}$。

駐點的性質。當 θ 遞增經過 $\frac{1}{2}k\pi$ 時，$\sin 4\theta$ 由負變正，所以 y 由遞減轉為遞增：這些點是**極小點（minima）**。當 θ 遞增經過 $\frac{1}{4}\pi + \frac{1}{2}k\pi$ 時，$\sin 4\theta$ 由正變負，所以這些點是**極大點（maxima）**。

下圖顯示 $a = 2$、$b = 1$ 的情況，縱軸的比例經過拉長：y 在 3 與 $\sqrt{10} \approx 3.162$ 之間振盪，週期（period）為 $\frac{1}{2}\pi$。



最後的不等式

把 θ 換成 $\theta+\frac{1}{2}\pi$，c^2 與 s^2 便會對調，於是 P 與 Q 也對調，所以 y 的週期為 $\frac{1}{2}\pi$。在一個週期 $[0,\frac{1}{2}\pi]$ 上，y 在 $[0,\frac{1}{4}\pi]$ 上遞增，在 $[\frac{1}{4}\pi,\frac{1}{2}\pi]$ 上遞減（由 $\sin 4\theta$ 的正負可知），所以它的最小值是 $|a|+|b|$，在兩個端點取得；最大值是 $\sqrt{2a^2+2b^2}$，在 $\frac{1}{4}\pi$ 取得。根據週期性，這兩個值就是對所有 θ 而言的最小值和最大值，所以

$$|a|+|b| \;\le\; y \;\le\; \sqrt{2a^2+2b^2}.$$

這兩個界並無矛盾：把第一部分的結果應用於 $|a|$ 和 $|b|$，得 $(|a|+|b|)^2 \le 2a^2+2b^2$，所以下界確實不會超過上界。

*先前擱下的情況。*若 $a^2=b^2$，則 $y=2|a|$，而 $|a|+|b|=2|a|$，$\sqrt{2a^2+2b^2}=\sqrt{4a^2}=2|a|$，所以不等式處處以等號成立。若 $b=0$，則 $y=|a|\big(|\cos\theta|+|\sin\theta|\big)$。由於每一項都不超過 1，故 $|\cos\theta|+|\sin\theta| \ge \cos^2\theta+\sin^2\theta=1$，而由第一部分可知 $(|\cos\theta|+|\sin\theta|)^2 \le 2$；所以 $|a| \le y \le \sqrt{2}\,|a|$，這正是所求的不等式。$a=0$ 的情況相同，只需把 a 與 b 對調；而 $a=b=0$ 的情況是平凡的。這就證明了結果對所有 a、b 和 θ 都成立。 ■

驗算

- 取 $a=2$、$b=1$：$y(0)=2+1=3$，$y(\frac{1}{4}\pi)=2\sqrt{5/2}=\sqrt{10}\approx 3.1623$；以數值方法對大量 θ 值掃描 y，結果都落在這兩個值之間。
- 導數公式與 $\mathrm{d}y/\mathrm{d}\theta$ 的有限差分（finite-difference）估計值相符，在隨機選取的 a、b、θ 下誤差約為 10^{-9}。
- 當 $a=b$ 時，兩個界都退化為 $2|a|$，與 y 為常數的情況吻合。

常見錯誤

- 寫成 $\sqrt{a^2} = a$：最小值是 $|a| + |b|$，這正是答案中出現絕對值（modulus）的原因。
- 只做到 $\sin 2\theta = 0$ 就停下來，漏掉了位於 $\frac{1}{4}\pi$ 的奇數倍之處的駐點，在那裏 $\sqrt{P} = \sqrt{Q}$。
- 毫無理由就斷定某些點是極大點或極小點；利用 $\sin 4\theta$ 的正負，一行就能給出理由。

STEP 之外

從幾何上看，$\sqrt{P}$ 是橢圓（ellipse）$x = a\cos\phi$，$y = b\sin\phi$ 的中心到橢圓上某一點的距離，而 $\sqrt{Q}$ 是中心到參數再轉四分之一圈之後那一點的距離，所以這道題探討的是這兩條「共軛」半徑如何攤分固定的總和 $a^2 + b^2$。這兩個界同時也是向量 (a, b) 的 ℓ^1 範數（norm）和（經縮放的）ℓ^2 範數，而像這樣比較範數，是大學一年級分析（analysis）和線性代數（linear algebra）中的常用工具。想進一步了解如何利用導數證明不等式，可參閱 Abbott 的 *Understanding Analysis*（數學分析入門教材）。

詳細解答

STEP I 2002，第 4 題

純數學 • 切線與不等式

> **STEP I 2002，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/02-s1-q4

關鍵思路

一條通過 $(0,1)$ 並在較遠處與曲線相切的直線，在這兩點之間必定位於曲線的下方，因為曲線在該處是凹的（concave），其後才變得平緩。直線下方的區域是一個梯形（trapezium），把它繞 y 軸旋轉會得到一個圓柱體和一個圓錐體，所以毋須用微積分也能輕易求出它的面積和體積。把它們與曲線下方的準確面積和體積比較，便得到那兩個不等式。

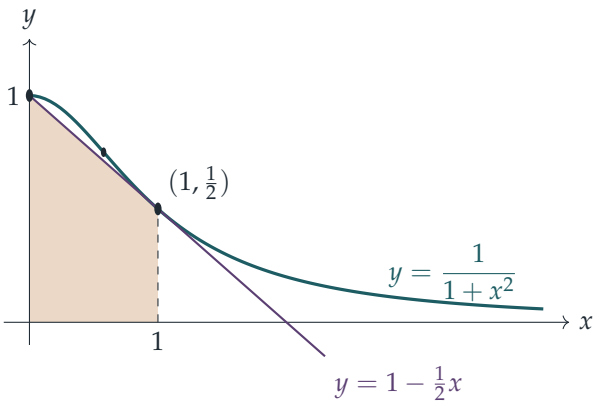
解答

略圖

設 $f(x) = \dfrac{1}{1+x^2}$。則 $f(0) = 1$，$f(x) > 0$，而且當 $x \to \infty$ 時 $f(x) \to 0$，所以 x 軸是漸近線（asymptote）。求導得

$$f'(x) = -\frac{2x}{(1+x^2)^2}, \qquad f''(x) = \frac{6x^2-2}{(1+x^2)^3}.$$

所以 $f'(0) = 0$（曲線在 $(0,1)$ 處以水平方向出發），而當 $x > 0$ 時 $f'(x) < 0$，故曲線一直遞減。此外，當 $0 \le x < 1/\sqrt{3}$ 時 $f''(x) < 0$，當 $x > 1/\sqrt{3}$ 時 $f''(x) > 0$，所以在 $(1/\sqrt{3}, \frac{3}{4})$ 有一個拐點（point of inflection）：曲線在 y 軸附近是凹的（向下彎），在較遠處則是凸的（convex）。



陰影部分的梯形是 $0 \le x \le 1$ 時直線下方的區域；它位於曲線下方的區域之內。

直線

直線通過 $(0, f(0)) = (0,1)$，所以它的方程是 $y = 1 + mx$，其中 m 是某個斜率（gradient）。假設它在 $x = a > 0$ 處與曲線相切。在該處必須滿足兩個條件。

它與曲線相交：$1 + ma = \dfrac{1}{1+a^2}$，所以 $ma = -\dfrac{a^2}{1+a^2}$，而由於 $a \neq 0$，

$$m = -\frac{a}{1+a^2}.$$

它的斜率與曲線的斜率相同：$m = f'(a) = -\dfrac{2a}{(1+a^2)^2}$。

令兩式相等，再除以 $-a/(1+a^2)$（因為 $a > 0$，這樣做是容許的），得 $1 = \dfrac{2}{1+a^2}$，所以 $a^2 = 1$，而由於 $a > 0$，故 $a = 1$。於是 $m = -\frac{1}{2}$，直線是

$$y = 1 - \tfrac{1}{2}x,$$

它在 $(1, \frac{1}{2})$ 處與曲線相切。

再無其他交點

把直線減去曲線：

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

這個式子只在 $x = 0$ 或 $x = 1$ 時等於零。所以直線與曲線相交於 $x = 0$ 和切點 $x = 1$（重因式 $(x-1)^2$ 正是相切在代數上的標誌），此外再無其他交點：在 $x \ge 0$ 時沒有，事實上在 x 為負數時也沒有。

我們還可以從中讀出下一步需要的結果。當 $x > 0$ 且 $x \neq 1$ 時，這個式子嚴格小於零，所以

$$0 < 1 - \tfrac{1}{2}x < \frac{1}{1+x^2} \qquad \text{當} 0 < x < 1. \tag{$*$}$$

（左邊的不等式成立，是因為在 $[0,1]$ 上 $1 - \frac{1}{2}x \geq \frac{1}{2}$。）所以在 $[0,1]$ 上，除了端點之外，直線嚴格位於曲線下方，正如略圖所示。

$\pi > 3$

把 $(*)$ 在 $[0,1]$ 上積分。兩個函數都是連續的，而且在開區間上曲線嚴格位於直線上方，所以積分之後不等式仍然是嚴格不等式：

$$\int_0^1 \frac{1}{1+x^2}\,\mathrm{d}x > \int_0^1 \Big(1 - \frac{x}{2}\Big)\mathrm{d}x.$$

左邊是 $\frac{\pi}{4}$（即 $\big[\arctan x\big]_0^1$）。右邊是陰影梯形的面積，梯形的兩條平行邊長為 1 和 $\frac{1}{2}$，寬為 1，所以面積是 $\frac{1}{2}\big(1 + \frac{1}{2}\big) = \frac{3}{4}$。因此 $\frac{\pi}{4} > \frac{3}{4}$，即

$$\pi > 3.$$

$\ln 2 > 2/3$

現在把同樣的兩個區域（$0 \leq x \leq 1$）繞 y 軸旋轉 2π。根據 $(*)$，梯形位於曲線下方的區域之內，所以它的旋轉體（solid of revolution）位於另一個旋轉體之內，體積也較小。

想一想 如果曲線是以 x 表示 y 的形式給出，我們怎樣求繞 y 軸旋轉的旋轉體體積？

使用圓柱殼法（cylindrical shells）：位於 x 處、寬度為 δx、高度為 y 的窄條，旋轉後掃出一個體積約為 $2\pi x\, y\, \delta x$ 的薄殼，所以體積是 $\int_0^1 2\pi x\, y\,\mathrm{d}x$。（另一個方法是沿水平方向切片；見「驗算」。）

對於曲線，

$$V_{\text{curve}} = \int_0^1 \frac{2\pi x}{1+x^2}\,\mathrm{d}x = \pi\big[\ln(1+x^2)\big]_0^1 = \pi \ln 2,$$

因為分子 $2x$ 是分母的導數。

對於直線，旋轉體是一個半徑為 1、高為 $\frac{1}{2}$ 的圓柱體（來自梯形在 $y = \frac{1}{2}$ 以下的部分），上面再加一個半徑為 1、高為 $\frac{1}{2}$ 的圓錐體（來自其上方的三角形）。所以

$$V_{\text{line}} = \pi \cdot 1^2 \cdot \tfrac{1}{2} + \tfrac{1}{3}\pi \cdot 1^2 \cdot \tfrac{1}{2} = \frac{\pi}{2} + \frac{\pi}{6} = \frac{2\pi}{3}.$$

用圓柱殼法驗算：$\int_0^1 2\pi x\big(1 - \frac{1}{2}x\big)\mathrm{d}x = 2\pi\big(\frac{1}{2} - \frac{1}{6}\big) = \frac{2\pi}{3}$。

被積函數 $2\pi x\left(\frac{1}{1+x^2}-1+\frac{x}{2}\right)$ 是連續的，在 $[0,1]$ 上非負，而且根據 $(*)$ 在 $(0,1)$ 上嚴格為正，所以 $V_{\text{curve}} > V_{\text{line}}$：

$$\pi \ln 2 > \frac{2\pi}{3} \quad \Longrightarrow \quad \ln 2 > \frac{2}{3},$$

這正是所要證明的。

驗算

- 以數值計算，$\pi/4 \approx 0.7854 > 0.75$，$\ln 2 \approx 0.6931 > 0.6667$，所以兩個不等式都成立，而且兩者都不算寬鬆。
- 改為沿水平方向切片：當 $\frac{1}{2} \le y \le 1$ 時，曲線滿足 $x^2 = \frac{1}{y} - 1$，所以體積是圓柱體的 $\pi \cdot 1^2 \cdot \frac{1}{2}$ 加上 $\int_{1/2}^{1} \pi\left(\frac{1}{y}-1\right)\mathrm{d}y = \pi\left(\ln 2 - \frac{1}{2}\right)$，同樣得到 $\pi \ln 2$。
- 快速檢驗因式分解：在 $x = 2$ 時，直線給出 0，曲線給出 $\frac{1}{5}$，兩者相差 $-\frac{1}{5}$；公式給出 $-\frac{2\cdot 1}{2\cdot 5} = -\frac{1}{5}$。

常見錯誤

- 求出一般點上的切線，卻忘記了切線須通過 $(0,1)$ 這個條件（或者反過來）：兩個條件都要用上，才能確定 a。
- 止步於「$\ge$」：要得到嚴格不等式，直線必須在某個區間上嚴格位於曲線下方，而這正是因式 $x(x-1)^2$ 所提供的。
- 誤把區域繞 x 軸旋轉，這樣會得到 $\int \pi y^2\,\mathrm{d}x$ 和不同的數值。

STEP 之外

這個論證是一個比較原理（comparison principle）：如果在某個區間上 $g \le h$，在某處有嚴格不等式成立，而且兩者都連續，那麼 $\int g < \int h$。把「在某處嚴格成立」說得精確，是大學第一門分析課程中的典型習題，在那裏積分是透過上和與下和（upper and lower sums）來定義的。以簡單圖形的面積來估計 π 的上下界，可以追溯至阿基米德（Archimedes），而這道題還有精密得多的版本（例如 $\int_0^1 x^4(1-x)^4/(1+x^2)\,\mathrm{d}x = \frac{22}{7} - \pi$），值得查閱一下。至於分析方面，可以試讀 Abbott 的 *Understanding Analysis*（分析學入門教材）。

詳細解答

STEP I 2002，第 5 題

純數學 • 多項式的整數根

> **STEP I 2002，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/02-s1-q5

關鍵思路

寫成一次因式（linear factor）乘積的多項式（polynomial），在特殊點上很容易求值：代入 $x = 0$、$x = 1$ 或 $x = -1$，因式分解後的形式就變成 k_i、$k_i + 1$ 或 $k_i - 1$ 的乘積，而展開後的形式則給出係數（coefficient）之和。對於題中的四次多項式（quartic），這些值都是固定的整數，而根（root）是整數就意味着每個 $k_i \pm 1$ 都必須是某個已知數的因數（divisor）。把這些數分解成質因數之後，可供選擇的餘地就所剩無幾。

解答

三個恆等式

由於 $f(x)$ 的兩個表達式是同一個多項式，它們在 x 的每一個值上都相等。

代入 $x = 0$：展開式給出 $f(0) = a_n$，因為其餘各項都含有因子 x。因式分解後的形式給出 $f(0) = k_1 k_2 \cdots k_n$。因此

$$k_1 k_2 \cdots k_n = a_n.$$

代入 $x = 1$：x 的每個冪都變成 1，所以展開式給出 $f(1) = 1 + a_1 + a_2 + \cdots + a_n$，而因式分解後的形式給出 $(1 + k_1)(1 + k_2) \cdots (1 + k_n)$。因此

$$(k_1 + 1)(k_2 + 1) \cdots (k_n + 1) = 1 + a_1 + a_2 + \cdots + a_n,$$

正如所求。

代入 $x = -1$：因式分解後的形式給出 $(k_1 - 1)(k_2 - 1) \cdots (k_n - 1)$。在展開式中，$a_j x^{n-j}$ 一項變成 $(-1)^{n-j} a_j = (-1)^n (-1)^j a_j$（這裏用到 $(-1)^{-j} = (-1)^j$），所以

$$(k_1 - 1)(k_2 - 1) \cdots (k_n - 1) = (-1)^n \big(1 - a_1 + a_2 - a_3 + \cdots + (-1)^n a_n\big).$$

當 n 為偶數時，這就是 $1-a_1+a_2-\cdots+a_n$；當 n 為奇數時，每個符號都要反轉。

四次多項式

把這個四次多項式寫成 $f(x) = (x+k_1)(x+k_2)(x+k_3)(x+k_4)$，於是它的根是 $-k_1,\ldots,-k_4$，而且都是整數。取 $n=4$ 應用上述恆等式：

$$k_1k_2k_3k_4 = 576 = 2^6\cdot 3^2,$$
$$(k_1+1)(k_2+1)(k_3+1)(k_4+1) = 1+22+172+552+576 = 1323 = 3^3\cdot 7^2,$$
$$(k_1-1)(k_2-1)(k_3-1)(k_4-1) = 1-22+172-552+576 = 175 = 5^2\cdot 7.$$

每個 k_i 都至少是 2。所有係數都是正數，所以對每個 $x\geq 0$ 都有 $f(x)>0$；因此每個根都是負數，每個 k_i 都是正整數。此外，沒有任何 k_i 等於 1，否則第三個乘積會是 0 而不是 175。所以對每個 i 都有 $k_i\geq 2$，而 k_i-1 和 k_i+1 全都是正整數。

想一想 怎樣把三個乘積化為一張簡短的清單，列出 k_i 的可能值？

每個 k_i-1 都是 175 的正因數，每個 k_i+1 都是 1323 的正因數。175 的因數是 $1,5,7,25,35,175$，由此得

$$k_i\in\{2,6,8,26,36,176\},\qquad k_i+1\in\{3,7,9,27,37,177\}.$$

在 k_i+1 的這些值當中，只有 $3,7,9,27$ 能整除 $1323=3^3\cdot 7^2$（因為 37 是質數，而 $177=3\cdot 59$）。所以每個 k_i 都是 $2,6,8,26$ 其中之一，每個 k_i+1 都是 $3,7,9,27$ 其中之一。

現在看乘積 $\prod(k_i+1)=3^3\cdot 7^2$。

- 因子 7^2 只能來自等於 7 的因子，所以 k_i+1 當中恰好有兩個等於 7：有兩個 k_i 是 6。
- 另外兩個因子是取自 $\{3,9,27\}$ 的 3 的冪，乘積為 $3^3=27$。兩者都至少是 3，所以不可能取 27（否則另一個就要是 1），唯一的選擇是 3×9。所以餘下的 k_i 是 2 和 8。

於是 $\{k_1,k_2,k_3,k_4\}=\{2,6,6,8\}$。用第一個乘積驗證：$2\cdot 6\cdot 6\cdot 8=576$，正如所求。展開亦可確認：

$$(x+2)(x+8)=x^2+10x+16,\qquad (x+6)^2=x^2+12x+36,$$

以及

$$(x^2+10x+16)(x^2+12x+36)=x^4+22x^3+172x^2+552x+576.$$

所以各根是

$$x=-2,\quad x=-6\text{（重根）},\quad x=-8.$$

驗算

- 各 k_i 之和是 $2+6+6+8=22$，與 x^3 的係數吻合。
- $f(-6)=1296-4752+6192-3312+576=0$，而 $f'(x)=4x^3+66x^2+344x+552$ 在 $x=-6$ 處同樣等於零，確認 -6 是二重根。
- 用數值快速檢驗這些恆等式：對 $f(x)=(x+2)(x-3)=x^2-x-6$，我們得到 $(2-1)(-3-1)=-4$ 和 $1-(-1)+(-6)=-4$。

常見錯誤

- 在 $x=-1$ 的結果中弄錯符號：當 n 為奇數時，因子 $(-1)^n$ 不能忽略。
- 忘記重根是容許的：一旦在 1323 中見到 7^2，就必定有兩個 k_i 相等。
- 沒有先證明每個 $k_i \geq 2$ 就只列出正因數；正是這一步保證了所有 $k_i \pm 1$ 都是正數。

STEP 之外

這裏的恆等式是韋達定理（Vieta's formulae）的特殊情況；韋達定理把多項式的係數表示為其根的基本對稱函數（elementary symmetric functions）。至於因數的論證，則與有理根定理（rational root theorem）一脈相承，也與高斯（Gauss）關於整係數多項式因式分解的引理（Gauss's lemma）相近，後者你會在環（ring）論的入門課程中遇到。Allenby 的 *Rings, Fields and Groups*（《環、域與羣》）對這些想法有細緻的闡述。

詳細解答

STEP I 2002，第 6 題

純數學 • 角錐的幾何

STEP I 2002，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：

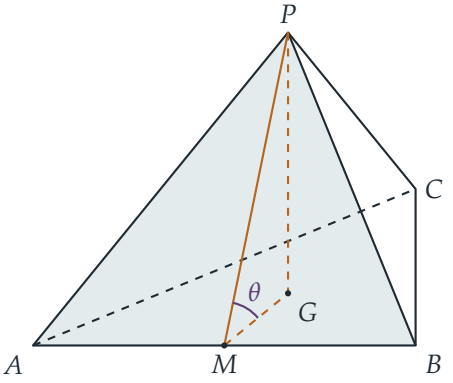
step.maths.org/questions/02-s1-q6

關鍵思路

根據對稱性，頂點位於等邊底面中心的正上方，因此只需用一次畢氏定理（Pythagoras' theorem）便可求得高。知道體積之後，我們根本不必直接求新的高：無論哪一個面放在地上，體積都是一樣的，所以 $V = \frac{1}{3} \times$ (新底面的面積) $\times h$。最後一部分所求的角，是兩個面沿公共棱所成的夾角；要看清楚這個角，最簡潔的方法是考慮通過頂點並垂直於該棱的截面（cross-section）。

解答

把底面記為 ABC，邊長為 a，頂點記為 P，且 $PA = PB = PC = b$。設 G 為 ABC 的中心，M 為 AB 的中點。



體積

設 H 為 P 在底面之上的高度，F 為由 P 所作垂線的垂足。直角三角形 PFA、PFB、PFC 有公共邊 $PF = H$，而斜邊均為 b，所以

$$FA = FB = FC = \sqrt{b^2 - H^2}.$$

因此 F 與三個頂點等距：它是 ABC 的外心（circumcentre），而等邊三角形的外心就是其中心 G。ABC 的中線（median）長度為 $\frac{\sqrt{3}}{2}a$，而 G 位於中線上由頂點起計三分之二的位置，所以 $GA = \frac{2}{3}\cdot\frac{\sqrt{3}}{2}a = \frac{a}{\sqrt{3}}$。在三角形 PGA 中運用畢氏定理，得

$$H^2 = b^2 - \frac{a^2}{3} = \frac{3b^2 - a^2}{3}.$$

底面的面積為 $\frac{1}{2}a^2\sin 60^\circ = \frac{\sqrt{3}}{4}a^2$，所以

$$V = \frac{1}{3}\cdot\frac{\sqrt{3}}{4}a^2\cdot\frac{\sqrt{3b^2 - a^2}}{\sqrt{3}} = \frac{1}{12}a^2\left(3b^2 - a^2\right)^{\frac{1}{2}},$$

正是所要求的結果。（這種形狀的棱錐只在 $3b^2 > a^2$ 時才存在，而這正好是上式為正實數的條件。）

新的高

現在把面 PAB 放在地上。它是等腰三角形，$PA = PB = b$ 而 $AB = a$，所以 PM 垂直於 AB；在三角形 PMA 中運用畢氏定理，得

$$PM^2 = b^2 - \frac{a^2}{4} = \frac{4b^2 - a^2}{4}, \qquad \text{面積}PAB = \frac{1}{2}a\cdot PM = \frac{a}{4}\sqrt{4b^2 - a^2}.$$

想一想 要求 h，我們真的需要找出新的最高頂點的位置嗎？

不需要。立體本身沒有改變，所以它的體積仍然是 V；而體積公式對任何一個面作為底面都成立，只要高是垂直於該面量度的。所以

$$V = \frac{1}{3}\cdot\frac{a}{4}\sqrt{4b^2 - a^2}\cdot h.$$

與第一部分的結果比較，

$$\frac{a}{12}\sqrt{4b^2 - a^2}\,h = \frac{a^2}{12}\sqrt{3b^2 - a^2} \quad\Longrightarrow\quad h = \frac{a\sqrt{3b^2 - a^2}}{\sqrt{4b^2 - a^2}},$$

兩邊平方得

$$h^2 = \frac{a^2(3b^2 - a^2)}{4b^2 - a^2},$$

正是所要求的結果。這裏 $4b^2 - a^2 > 0$，因為 $4b^2 > 3b^2 > a^2$。

等邊面的傾角

在新的位置，地面就是 PAB 所在的平面，而等邊面 ABC 與它沿棱 AB 相交。兩個平面之間的夾角，是指在兩個平面上各取一條直線，於同一點垂直於公共棱，這兩條直線所成的角。在 M 點，面 PAB 上的直線 MP 垂直於 AB（因為 PAB 是等腰三角形），面 ABC 上的直線 MC 亦垂直於 AB（因為 ABC 是等邊三角形），而且 G 在 MC 上。所以所求的角是 $\theta = \angle PMC = \angle PMG$。

三角形 PMG 在 G 處為直角，因為 PG 垂直於整個底面。此外，GM 是中線的三分之一，$GM = \frac{1}{3}\cdot\frac{\sqrt{3}}{2}a = \frac{a}{2\sqrt{3}}$。因此

$$\tan\theta = \frac{PG}{GM} = \frac{\sqrt{3b^2-a^2}}{\sqrt{3}}\cdot\frac{2\sqrt{3}}{a} = \frac{2\sqrt{3b^2-a^2}}{a},$$

所以等邊三角形與水平面之間的夾角為

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

這個角是銳角，因為 G 位於底面之內，所以它確實是兩個平面之間的夾角，而不是其補角。

另一個等價的做法是利用新的高：頂點 C 在地面之上的高度為 h，而沿着該面量度，它與棱 AB 的距離為 $MC = \frac{\sqrt{3}}{2}a$，所以 $\sin\theta = h\big/\frac{\sqrt{3}}{2}a$，由此得

$$\sin\theta = \frac{2\sqrt{3b^2-a^2}}{\sqrt{3(4b^2-a^2)}}, \qquad \cos\theta = \frac{a}{\sqrt{3(4b^2-a^2)}},$$

與上面 $\tan\theta$ 的值吻合。

驗算

對於正四面體（regular tetrahedron），$a = b = 1$：公式給出 $V = \frac{\sqrt{2}}{12}$，$h^2 = \frac{2}{3}$（與原來的高相同，根據對稱性亦必然如此），以及 $\cos\theta = \frac{1}{3}$，即大家熟悉的正四面體二面角（dihedral angle）$\theta \approx 70.53°$。當 $b \to a/\sqrt{3}$ 時，棱錐變得扁平，$\theta \to 0$；當 $b \to \infty$ 時，它變成又高又尖的錐體，$\theta \to 90°$。我們亦用坐標為 20,000 組隨機數對 (a, b) 建構了這個棱錐，並以三重積計算 V，以 C 到平面 PAB 的距離計算 h，以法向量計算 θ；三者均與公式吻合，誤差約為 10^{-13}。例如 $a = 2$、$b = 3$ 給出 $h \approx 1.6956$ 及 $\theta \approx 78.22°$。

常見錯誤

- 把 GA 當作中線的一半，或在需要用外接圓半徑（circumradius）$\frac{a}{\sqrt{3}}$ 的地方用了內切圓半徑（inradius）$\frac{a}{2\sqrt{3}}$（在最後一部分則剛好相反）。

► 量度該面與地面之間的夾角時，用了一條不垂直於公共棱的直線，例如 AC 或 PC。

STEP 之外

以不同的面作底來計算同一個體積，讓我們初次窺見這樣一個想法：體積並不取決於我們怎樣切割立體或怎樣擺放它。在大學一年級的課程中，四面體的體積會寫成 $\frac{1}{6}|\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})|$，即一個純量三重積（scalar triple product，亦即一個 3×3 行列式（determinant）），而兩個面之間的夾角就是它們的法向量（normal vector）之間的夾角，我們的數值驗算正是這樣做的。Strang 的 *Introduction to Linear Algebra*（線性代數入門）把行列式作為體積來闡述。

詳細解答

STEP I 2002，第 7 題

純數學 • 三角函數積分

STEP I 2002，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/02-s1-q7

關鍵思路

單獨來看，I 和 J 都不易處理，但兩者之和與兩者之差卻都很好計算。把它們相加，分母（denominator）便出現在分子（numerator）上，於是被積函數（integrand）就是 1。把它們相減，分子上出現的是分母的導數（derivative），於是積分是一個對數（logarithm）。後面兩個積分用的是同一技巧的較一般形式：把分子拆成分母的某個倍數，加上分母的導數的某個倍數。

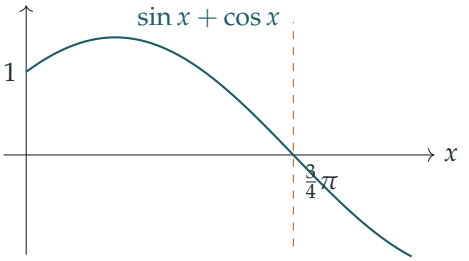
解答

主要結果

首先，為甚麼要對 a 加上限制？我們可以寫成

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

它在 $-\frac{\pi}{4} < x < \frac{3\pi}{4}$ 時為正，並在 $x = \frac{3}{4}\pi$ 處首次等於零。因此當 $0 \le a < \frac{3}{4}\pi$ 時，分母在整個 $[0, a]$ 上都是正數。這樣兩個積分都是普通的積分，而下面取對數的也是一個正數。



把兩個積分相加，

$$I + J = \int_0^a \frac{\cos x + \sin x}{\sin x + \cos x}\,\mathrm{d}x = \int_0^a 1\,\mathrm{d}x = a.$$

把它們相減，

$$I - J = \int_0^a \frac{\cos x - \sin x}{\sin x + \cos x}\,\mathrm{d}x.$$

分子 $\cos x - \sin x$ 正好是分母 $\sin x + \cos x$ 的導數，所以被積函數具有 $f'(x)/f(x)$ 的形式，其中在 $[0, a]$ 上 $f > 0$。因此

$$I - J = \Big[\ln(\sin x + \cos x)\Big]_0^a = \ln(\sin a + \cos a) - \ln 1 = \ln(\sin a + \cos a).$$

把兩個結果相加，得

$$2I = a + \ln(\sin a + \cos a),$$

這正是所要求的結果。（把兩者相減，還可以順帶得到 $2J = a - \ln(\sin a + \cos a)$。）

第 (i) 部分

由於 $p, q > 0$，而 $\sin x$ 和 $\cos x$ 在 $[0, \frac{\pi}{2}]$ 上都非負，且不會同時為零，所以分母 $D(x) = p\sin x + q\cos x$ 在該區間上是正數。它的導數是 $D'(x) = p\cos x - q\sin x$。

想一想 當分母帶有係數 p 和 q 時，「考慮 $I + J$ 與 $I - J$」這個做法應該怎樣推廣？

先前取和與差之所以奏效，是因為兩個分子相加得出分母，相減則得出分母的導數。所以現在我們要找常數 A 和 B，使得

$$\cos x = A\,(p\sin x + q\cos x) + B\,(p\cos x - q\sin x) \qquad \text{對所有} x.$$

比較 $\sin x$ 和 $\cos x$ 的係數：

$$pA - qB = 0, \qquad qA + pB = 1.$$

由第一式得 $B = pA/q$；代入第二式得 $qA + p^2A/q = 1$，所以

$$A = \frac{q}{p^2 + q^2}, \qquad B = \frac{p}{p^2 + q^2}.$$

現在積分拆成容易的一部分和對數的一部分：

$$\begin{aligned}\int_0^{\pi/2} \frac{\cos x}{p\sin x + q\cos x}\,\mathrm{d}x &= \int_0^{\pi/2} \left(A + B\,\frac{D'(x)}{D(x)}\right)\mathrm{d}x \\ &= \frac{\pi}{2}A + B\Big[\ln D(x)\Big]_0^{\pi/2} = \frac{\pi}{2}A + B\ln\frac{p}{q},\end{aligned}$$

其中用到 $D(\frac{\pi}{2}) = p$ 和 $D(0) = q$，兩者皆為正數。因此

$$\int_0^{\pi/2} \frac{\cos x}{p\sin x + q\cos x}\,\mathrm{d}x = \frac{\frac{1}{2}\pi q + p\ln(p/q)}{p^2 + q^2}.$$

第 (ii) 部分

現在分母是 $D(x) = 3\sin x + 4\cos x + 25$，而 $D'(x) = 3\cos x - 4\sin x$。分母是正數，因為 $|3\sin x + 4\cos x| \le 5 < 25$。$D$ 中有常數 25，表示我們還須令常數項相符，所以要找 A 和 B，使得

$$\cos x + 4 = A\,(3\sin x + 4\cos x + 25) + B\,(3\cos x - 4\sin x).$$

比較係數：

$$\sin x:\ 3A - 4B = 0, \qquad \cos x:\ 4A + 3B = 1, \qquad \text{常數項}:\ 25A = 4.$$

由前兩式得 $A = \frac{4}{25}$ 和 $B = \frac{3}{25}$，而常數項的方程 $25 \cdot \frac{4}{25} = 4$ 也同樣成立。三條方程只有兩個未知數，這並非必然成立：分子中的「$+4$」是特意選來令它成立的。所以

$$\begin{aligned}\int_0^{\pi/2} \frac{\cos x + 4}{3\sin x + 4\cos x + 25}\,\mathrm{d}x &= \int_0^{\pi/2} \left(\frac{4}{25} + \frac{3}{25}\,\frac{D'(x)}{D(x)}\right)\mathrm{d}x \\ &= \frac{4}{25}\cdot\frac{\pi}{2} + \frac{3}{25}\Big[\ln D(x)\Big]_0^{\pi/2} \\ &= \frac{2\pi}{25} + \frac{3}{25}\ln\frac{28}{29},\end{aligned}$$

因為 $D(\frac{\pi}{2}) = 3 + 25 = 28$ 而 $D(0) = 4 + 25 = 29$。

驗算

- 取 $p = q = 1$ 時，第 (i) 部分應與主要結果在 $a = \frac{\pi}{2}$ 時的情況一致；後者給出 $2I = \frac{\pi}{2} + \ln 1$，所以 $I = \frac{\pi}{4}$。公式給出 $\frac{\pi/2+0}{2} = \frac{\pi}{4}$。兩者吻合。
- 當 $a \to \frac{3}{4}\pi$ 時，$\ln(\sin a + \cos a) \to -\infty$。這與以下事實相符：在分母快要變成零之前，$I$ 的被積函數是絕對值很大的負數，因為在那裏 $\cos x < 0$。
- 以數值計算，第 (ii) 部分是 $\frac{2\pi}{25} + \frac{3}{25}\ln\frac{28}{29} \approx 0.247116$，而對該積分使用辛普森法則（Simpson's rule）所得的結果與它吻合至小數點後十二位。主要結果和第 (i) 部分也分別對 200 個隨機的 a 值，以及隨機的 p 和 q 值，與數值積分（quadrature）的結果吻合。

常見錯誤

- 在第 (i) 部分得出 $\ln(q/p)$ 而不是 $\ln(p/q)$。在上限 $\frac{\pi}{2}$ 處，留下來的是 sin 那一項，所以得出 p。
- 在第 (ii) 部分忘記常數項，只比較 $\sin x$ 和 $\cos x$ 的係數。這樣做碰巧行得通，但你應該加以驗證，而不是假定它成立。

STEP 之外

把分子寫成「A 乘以分母，加上 B 乘以分母的導數」，其實是一小段線性代數（linear algebra）。函數 $\sin x$ 和 $\cos x$ 張成一個二維空間，而只要 $p^2 + q^2 \neq 0$，D 和 D' 就構成這個空間的另一組基（basis）。同一個想法，即找出合適的基來展開，也是部分分式（partial fractions）和求解微分方程（differential equations）的待定係數法（method of undetermined coefficients）背後的原理，你會在大學一年級的數學方法課程中學到這些內容。Siklos 的 *Advanced Problems in Mathematics* 中還有更多可以用這類對稱技巧解決的積分。

詳細解答

STEP I 2002，第 8 題

純數學 • 複利貸款

STEP I 2002，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/02-s1-q8

關鍵思路

先寫下一年之內發生的事：債務先乘以因子 $1+\alpha$，然後減去還款額。把這個一步遞推關係（recurrence）逐層展開，便可看出每筆還款 R_k 到第 n 年年底時「值」$(1+\alpha)^{n-k}R_k$，因為它令這麼多的利息無從累積。其餘的一切都只是等比級數（geometric series）。至於最後的近似值，用二項式定理（binomial theorem）展開 $1+\alpha$ 的冪，保留剛好足夠的項即可。

解答

C_{n+1} 的一般公式

在第 k 年年底，加上利息後債務 C_k 變成 $(1+\alpha)C_k$，然後償還 R_k。餘下的就是第 $k+1$ 年年初的債務：

$$C_{k+1} = (1+\alpha)C_k - R_k.$$

重複應用這條關係式，

$$\begin{aligned} C_{n+1} &= (1+\alpha)C_n - R_n \\ &= (1+\alpha)^2 C_{n-1} - (1+\alpha)R_{n-1} - R_n \\ &\vdots \\ &= (1+\alpha)^n C_1 - (1+\alpha)^{n-1}R_1 - (1+\alpha)^{n-2}R_2 - \cdots - R_n, \end{aligned}$$

因此，由於 $C_1 = C$，

$$\boxed{C_{n+1} = (1+\alpha)^n C - \sum_{k=1}^{n}(1+\alpha)^{n-k}R_k}.$$

（要令這個論證嚴謹，可驗證它在 $n=1$ 時成立，此時它說的是 $C_2=(1+\alpha)C-R_1$；並驗證若它對 n 成立，則乘以 $1+\alpha$ 再減去 R_{n+1} 便得出 $n+1$ 的情況。）

還款額 $R_k=(1+\alpha)^k r$

在 N 年內還清貸款，即是 $C_{N+1}=0$。當 $R_k=(1+\alpha)^k r$ 時，和式中的每一項都相同：

$$(1+\alpha)^{N-k}R_k=(1+\alpha)^{N-k}(1+\alpha)^k r=(1+\alpha)^N r.$$

因此

$$0=C_{N+1}=(1+\alpha)^N C-N(1+\alpha)^N r=(1+\alpha)^N\,(C-Nr).$$

由於 $(1+\alpha)^N\neq 0$，我們得到 $C=Nr$，即 $r=C/N$，正是所求。

這個結果合乎情理：第 k 筆還款是「以今天的幣值計」的 C/N，再加上 k 年的利息增長，所以你合共償還的是本金（principal）C，另加恰好是它所生的利息。

等額還款 R

現在對每個 k 都有 $R_k=R$，而 $C_{N+1}=0$ 給出

$$(1+\alpha)^N C=R\sum_{k=1}^{N}(1+\alpha)^{N-k}=R\big(1+(1+\alpha)+\cdots+(1+\alpha)^{N-1}\big).$$

括號內是一個等比級數，首項為 1，公比為 $1+\alpha$，共 N 項，所以（因為 $\alpha>0$，公比不等於 1）它等於

$$\frac{(1+\alpha)^N-1}{(1+\alpha)-1}=\frac{(1+\alpha)^N-1}{\alpha}.$$

於是 $(1+\alpha)^N C=R\,\dfrac{(1+\alpha)^N-1}{\alpha}$，整理後得

$$\frac{R}{C}=\frac{\alpha(1+\alpha)^N}{(1+\alpha)^N-1},$$

這就是所求的結果。

$\alpha=\frac{1}{50}$、$N=4$ 的情況

> ? **想一想** 二項展開式應該保留多少項？為甚麼分子需要的項數比分母少？

分母 $(1+\alpha)^4-1$ 由 α 的一階項開始，而分子 $\alpha(1+\alpha)^4$ 則是 α 乘以一個由 1 開始的式子。所以約去因子 α 之後，分子是 $(1+\alpha)^4$，分母是 $\big((1+\alpha)^4-1\big)/\alpha$。要令兩者都準確至 α 的一階，分子需要 $(1+\alpha)^4\approx 1+4\alpha$，但分母需要 $(1+\alpha)^4\approx 1+4\alpha+6\alpha^2$（除以 α 之後，$6\alpha^2$ 變成 6α）。因此

$$\frac{R}{C}=\frac{(1+\alpha)^4}{\big((1+\alpha)^4-1\big)/\alpha}=\frac{1+4\alpha+6\alpha^2+\ldots}{4+6\alpha+4\alpha^2+\alpha^3}\approx\frac{1+4\alpha}{4+6\alpha}.$$

代入 $\alpha = \frac{1}{50}$，

$$\frac{R}{C} \approx \frac{1 + \frac{4}{50}}{4 + \frac{6}{50}} = \frac{54}{206} = \frac{27}{103},$$

正是所求。我們捨去的項，相對大小約為 α^2，在這裏只是百分之一的一小部分。

驗算

- 當 $N = 1$ 時，公式給出 $R/C = 1 + \alpha$：借入 C，連同一年的利息一併償還。正確。
- 當 $\alpha \to 0$ 時，$(1+\alpha)^N - 1 \approx N\alpha$，所以 $R/C \to 1/N$：沒有利息時，每年只需償還 C/N。這亦與 $r = C/N$ 吻合。
- 準確值是 $R/C = \dfrac{6765201}{25760050} \approx 0.26262$，而 $27/103 \approx 0.26214$；兩者相差約 0.2%。取 $R = 0.26262\,C$，把債務逐年推算下去，四年後（在捨入誤差之內）再無欠款。

常見錯誤

- 弄錯和式中的指數：R_k 是在第 k 年年底償還的，所以它少計的利息是往後的 $n - k$ 年，而不是 $n - k + 1$ 年。
- 在分母中也用 $(1+\alpha)^4 \approx 1 + 4\alpha$ 作近似：這樣會得出 $R/C \approx 27/100$，是一個較粗略的答案，與題目所求不符。

STEP 之外

$\sum (1+\alpha)^{-k} R_k$ 這個量是各筆還款的*現值*（present value），而貸款剛好還清，正是還款的現值等於借款額之時；R 的公式就是標準的年金（annuity，即按揭）公式。像 $C_{k+1} = (1+\alpha)C_k - R_k$ 這樣的遞推關係是一階線性差分方程（first-order linear difference equation），可說是微分方程（differential equation）$y' = \alpha y - R(t)$ 的離散版近親，兩者都可用同一個「積分因子」（integrating factor）的想法求解。如想多做一些這類把代數與建模（modelling）結合起來的題目，可參閱 Siklos 的 *Advanced Problems in Mathematics*（進階數學難題集）。

詳細解答

STEP I 2002，第 9 題

力學 • 斜面上的傾倒

STEP I 2002，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：

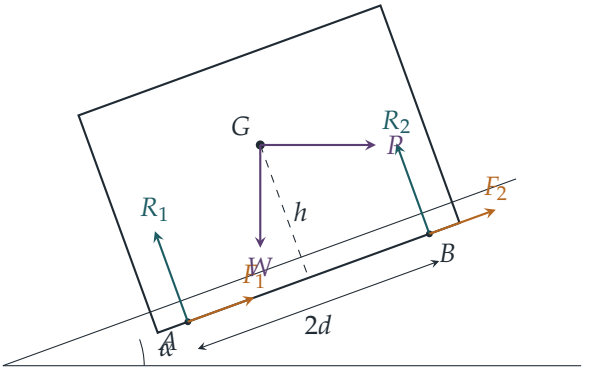
step.maths.org/questions/02-s1-q9

關鍵思路

貨車是處於平衡（equilibrium）狀態的剛體（rigid body），所以我們有三條方程：沿斜面方向分解力、沿垂直於斜面的方向分解力，以及取力矩（moments）。繞 G 取力矩是聰明的選擇，因為重量和 P 都通過 G，兩者一併消失，只剩下車輪上的力。這一條方程把兩個法向反作用力（normal reaction）之差與總摩擦力（friction）聯繫起來，幾乎包辦了全部工作。於是「不翻倒」的意思不過是：兩個法向反作用力都不是負數。

解答

我們把貨車視為剛體，其車輪與斜面在兩點接觸，即 A（較低）和 B（較高），兩點相距 $2d$。設 A 處的法向反作用力為 R_1，B 處的為 R_2，並設摩擦力為 F_1 和 F_2，兩者均以沿斜面向上為正（任何一個最後都可能是負數）。我們假設摩擦力足以防止滑動；題目只關乎翻倒。如圖所示，P 是水平的，指向斜面，所以它有沿斜面向上的分量 $P\cos\alpha$，以及壓向斜面的分量 $P\sin\alpha$。



沿垂直於斜面的方向和沿斜面方向分解力（以沿斜面向上為正）：

$$R_1 + R_2 = W\cos\alpha + P\sin\alpha, \tag{1}$$

$$F_1 + F_2 = W\sin\alpha - P\cos\alpha. \tag{2}$$

現在繞 G 取力矩。重量和 P 的作用線都通過 G，所以它們沒有任何貢獻。每個法向反作用力與 G 的垂直距離為 d，而每個摩擦力沿斜面作用，與 G 的垂直距離為 h。以逆時針方向為正（如圖所示，斜面向右上升），R_2 使貨車繞 G 逆時針轉動，R_1 使它順時針轉動，而沿斜面向上的摩擦力作用於 G 的下方，使它逆時針轉動。所以

$$d(R_2 - R_1) + h(F_1 + F_2) = 0. \tag{3}$$

第 (i) 部分

若 $R_1 = R_2$，方程 (3) 變成 $h(F_1 + F_2) = 0$。由於 $h > 0$，可得 $F_1 + F_2 = 0$：摩擦力之和為零，正是所要證明的。

（由 (2) 可知，這種情況恰好在 $P = W\tan\alpha$ 時出現：P 沿斜面向上的部分剛好抵銷重量沿斜面向下的部分，所以斜面無須提供任何淨摩擦力。）

第 (ii) 部分

把 (2) 代入 (3)，

$$R_2 - R_1 = \frac{h}{d}(P\cos\alpha - W\sin\alpha),$$

再與 (1) 結合，

$$2R_1 = W\cos\alpha + P\sin\alpha - \frac{h}{d}(P\cos\alpha - W\sin\alpha),$$

$$2R_2 = W\cos\alpha + P\sin\alpha + \frac{h}{d}(P\cos\alpha - W\sin\alpha).$$

? **想一想** 用這些方程來說，「不翻倒」是甚麼意思？

地面可以推車輪，卻不能拉住它，所以每個法向反作用力都必須滿足 $R \geq 0$。假如 R_2 的公式算出來是負數，較高的車輪就要靠外力壓住才行；實際上它會離地，貨車繞 A 轉動。同樣，$R_1 < 0$ 表示繞 B 轉動。所以貨車保持直立，當且僅當 $R_1 \geq 0$ 且 $R_2 \geq 0$。

以下我們一直使用 $\frac{d}{h} = \tan\beta$，其中 $0 < \beta < \frac{\pi}{2}$，並且 $0 < \alpha < \frac{\pi}{2}$。

較高的車輪不離地（$R_2 \geq 0$）。把 $2R_2 \geq 0$ 乘以 $d/h > 0$，

$$\tan\beta\,(W\cos\alpha + P\sin\alpha) + P\cos\alpha - W\sin\alpha \geq 0.$$

乘以 $\cos\beta > 0$ 並整理各項，

$$P(\cos\alpha\cos\beta + \sin\alpha\sin\beta) \geq W(\sin\alpha\cos\beta - \cos\alpha\sin\beta),$$

即

$$P\cos(\alpha-\beta) \geq W\sin(\alpha-\beta).$$

由於 $-\frac{\pi}{2} < \alpha-\beta < \frac{\pi}{2}$，我們有 $\cos(\alpha-\beta) > 0$，所以可以除以它而無須把不等式（inequality）的方向倒轉：

$$P \geq W\tan(\alpha-\beta).$$

較低的車輪不離地（$R_1 \geq 0$）。同樣地，由 $2R_1 \geq 0$ 可得

$$\tan\beta\,(W\cos\alpha + P\sin\alpha) - P\cos\alpha + W\sin\alpha \geq 0,$$

再乘以 $\cos\beta > 0$，

$$P(\cos\alpha\cos\beta - \sin\alpha\sin\beta) \leq W(\sin\alpha\cos\beta + \cos\alpha\sin\beta),$$

即

$$P\cos(\alpha+\beta) \leq W\sin(\alpha+\beta).$$

只要 $\alpha+\beta < \frac{\pi}{2}$（題目給出的答案其實暗中假設了這一點，否則 $\tan(\alpha+\beta)$ 是負數或者沒有定義），便有 $\cos(\alpha+\beta) > 0$，相除後得

$$P \leq W\tan(\alpha+\beta).$$

若 $\alpha+\beta \geq \frac{\pi}{2}$，則當 $P \geq 0$ 時左邊 ≤ 0，而右邊是正數，所以指向斜面的水平推力無論多大，都不能令貨車繞較高的車輪翻倒。

把兩個條件合起來，貨車不翻倒當且僅當

$$W\tan(\alpha-\beta) \leq P \leq W\tan(\alpha+\beta),$$

這就是所要求的結果。推力太小，貨車會繞 A 向斜面下方翻倒；推力太大，它會繞 B 被推得向斜面上方翻倒。

驗算

- 幾何圖像。重量和 P 合成為通過 G 的單一個力，它與鉛垂線成角 θ，其中 $\tan\theta = P/W$。當這個合力（resultant）通過車輪的接觸點時，貨車便處於即將翻倒的狀態。直線 GA 與斜面的法線成角 β，而法線與鉛垂線成角 α，所以合力在 $\theta = \alpha-\beta$ 時通過 A，在 $\theta = \alpha+\beta$ 時通過 B。這正好就是區間的兩個端點。
- 沒有斜坡。取 $\alpha = 0$，條件變成 $|P| \leq Wd/h$：這正是大家熟悉的條件，即平地上的箱子在高度 h 處受橫向推力而不翻倒的條件。
- 數值。取 $W = 3$、$\alpha = 0.4$、$d = 1$、$h = 1.5$，區間為 $-0.571 \leq P \leq 4.551$。對 20 000 個隨機配置直接檢查 $R_1, R_2 \geq 0$ 並與這個區間比對，結果完全吻合，而且在下端有 $R_2 = 0$，在上端有 $R_1 = 0$。

常見錯誤

- 繞車輪而不是繞 G 取力矩。這樣做也行得通，但重量和 P 的力臂（moment arm）就變得很麻煩，正負號也容易出錯。
- 忘記摩擦力對 G 有力矩。摩擦力作用於斜面上，在 G 下方距離 h 處，h 正是由此進入答案的。
- 沒有檢查 $\cos(\alpha \pm \beta)$ 的正負號就除以它，而不等式的方向正是由這個正負號決定的。

STEP 之外

車輪與地面之間的摩擦力是靜不定（statically indeterminate）的：平衡條件只能確定 $F_1 + F_2$，永遠不能分別確定 F_1 和 F_2，要知道兩者如何分配，就需要一個描述輪胎如何變形的模型。工程師經常遇到這種情況，例如支承數目多於嚴格所需的結構。上面的幾何驗算，用的就是合力作用線（line of action of the resultant）和支承多邊形（support polygon）的想法，步行機械人和疊起的積木的穩定性，背後也是同一個道理。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）仔細講解了剛體靜力學（statics）。

詳細解答

STEP I 2002，第 10 題

力學 • 彈性碰撞

STEP I 2002，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/02-s1-q10

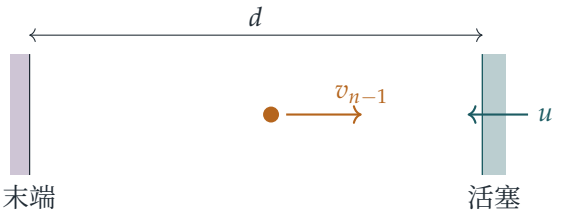
本題在 *Beyond STEP* 第 10.4 節一個分子與一個活塞中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch10/piston.html

關鍵思路

質點與一面遠比它重的牆發生完全彈性碰撞（perfectly elastic collision）時，相對速度會反向，所以最簡單的做法是在活塞（piston）的參考系（reference frame）中思考：在那裏，質點只是從一面靜止的牆反彈回來。因此，每次與活塞碰撞，速率都會增加 $2u$，而與固定端的碰撞則不會帶來任何改變。至於距離，留意在兩次與活塞碰撞之間，質點和活塞合共走過的長度有固定的規律，由此可得一條以 d_n 表示 d_{n+1} 的線性方程。

解答

我們把空氣模擬為一個質點（particle），把活塞視為一面以恆定速率 u 移動、不受碰撞影響的牆，並把圓筒的末端視為一面固定的牆。所有運動都沿着軸線進行，所以我們取由圓筒末端指向活塞的方向為正方向。



速率

假設質點以速率 w 趨近活塞。活塞的速度為 $-u$，所以質點相對於活塞的速度為 $w+u$（朝向活塞）。在與一面不會被撞偏的牆發生完全彈性碰撞時，根據牛頓實驗定律（Newton's experimental law），取 $e=1$，相對速度會反向，所以碰撞後質點相對於活塞的速度為 $-(w+u)$。它的實際速度於是為

$$-u-(w+u)=-(w+2u),$$

所以它以速率 $w+2u$ 離開活塞。

在圓筒的固定端，同一條定律（此時牆是靜止的）只會令速度反向：速率不變。所以在相鄰兩次與活塞碰撞之間，質點的速率保持不變。

第一次碰撞之前，質點以速率 v 朝活塞運動，所以 $v_1 = v+2u$。剛完成第 n 次碰撞後，它的速率為 v_n；它從末端反彈而速率不變，再以速率 v_n 遇上活塞，離開時速率為 $v_{n+1}=v_n+2u$。由歸納法（induction）可得

$$v_n = v+2nu,$$

正是所求。

距離

在第 n 次與第 $(n+1)$ 次碰撞之間（歷時 t_n），活塞以速率 u 向內移動，所以

$$d_n - d_{n+1} = u\,t_n.$$

在同一段時間內，質點以恆定速率 v_n 運動，由活塞走到末端（距離為 d_n，因為末端與活塞原來的位置相距 d_n），再返回活塞（距離為 d_{n+1}，即活塞此刻所在的位置）。所以

$$d_n + d_{n+1} = v_n t_n.$$

> **想一想** 怎樣才能消去 t_n？

把一條方程除以另一條：由於 $t_n>0$，所以

$$\frac{d_n+d_{n+1}}{d_n-d_{n+1}} = \frac{v_n}{u}.$$

交叉相乘，得 $ud_n+ud_{n+1}=v_nd_n-v_nd_{n+1}$，所以

$$(v_n+u)\,d_{n+1} = (v_n-u)\,d_n.$$

由 $v_n=v+2nu$，可得 $v_n-u=v+(2n-1)u$ 及 $v_n+u=v+(2n+1)u$，所以

$$d_{n+1} = \frac{v+(2n-1)u}{v+(2n+1)u}\,d_n,$$

這就是所求的結果。

d_n 的公式

重複應用這個遞推關係（recurrence），

$$d_n = d_1 \prod_{k=1}^{n-1} \frac{v+(2k-1)u}{v+(2k+1)u} = d_1 \cdot \frac{v+u}{v+3u} \cdot \frac{v+3u}{v+5u} \cdots \frac{v+(2n-3)u}{v+(2n-1)u}.$$

這個乘積可以逐項相消（telescope）：每個分母都與下一個分子約去，餘下

$$d_n = \frac{v+u}{v+(2n-1)u}\, d_1.$$

當 $n = 1$ 時這條式同樣成立，此時它就是 $d_1 = d_1$。

$v = u$ 的情況

代入 $v = u$，得

$$d_n = \frac{2u}{2nu}\, d_1 = \frac{d_1}{n}.$$

然後，由 $ut_n = d_n - d_{n+1}$，

$$ut_n = \frac{d_1}{n} - \frac{d_1}{n+1} = \frac{d_1}{n(n+1)},$$

正是所求。

驗算

- 當 $u = 0$ 時，活塞是一面固定的牆：公式給出對所有 n 都有 $v_n = v$ 及 $d_n = d_1$，與預期相符。
- 取 $u = 1$、$v = 3$、$d_1 = 75$。公式預測 $d_2 = 50$、$d_3 = 37.5$、$d_4 = 30$，而速率為 $5, 7, 9, 11$。以有理數運算逐步精確模擬反彈過程，所得結果與此完全相同；隨機選取 200 組 u、v 和起始位置，結果亦同樣吻合。
- 當 $v = u$ 時，$\sum ut_n = d_1 \sum \left(\frac{1}{n} - \frac{1}{n+1}\right) = d_1$：活塞合共恰好移動了 d_1，到達圓筒的末端，這與 $d_n = d_1/n \to 0$ 一致。

常見錯誤

- 錯用 $d_n - d_{n+1}$，例如把質點來回一趟的路程當作 $2d_n$：活塞在這段時間內已經移動了，所以回程只有 d_{n+1}。
- 每次碰撞只加上 u 而不是 $2u$。相對速度會反向，所以質點速度的改變量是活塞所帶來的趨近速率的兩倍。

STEP 之外

每次碰撞都令質點的速率改變一個固定的量，但隨着間隙收窄，碰撞會愈來愈頻密，而當 n 很大時，乘積 $v_n d_n$ 大致保持不變（這裏當 n 很大時 $v_n d_n \approx (v+u)d_1$）。這種緩慢變化的守恆量稱為絕熱不變量（adiabatic invariant）；對大量分子取平均後，它解釋了為何氣體被迅速壓縮時會變熱：對於一維的緩慢壓縮，$T \propto 1/L^2$。Kleppner 與 Kolenkow 合著的 *An Introduction to Mechanics*（力學入門）是進一步學習碰撞和參考系的好選擇。

詳細解答

STEP I 2002，第 11 題

力學 • 彈性斜碰撞

STEP I 2002，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：

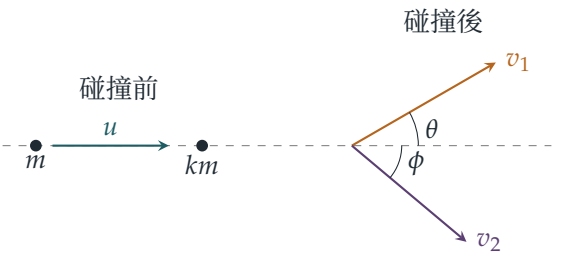
step.maths.org/questions/02-s1-q11

關鍵思路

一次碰撞給我們兩條守恆定律：動量（momentum）守恆（動量是向量，所以有兩條分量方程），以及動能（kinetic energy）守恆（因為沒有能量損失）。兩條動量方程對於碰撞後的未知速率而言是線性的，所以我們可以解出這些速率，並以角度表示。再代入能量方程，便可消去所有速率，只剩下角度與 k 之間的關係式。

解答

我們把 P_1 和 P_2 看成在平面上運動的質點（particle），並假設碰撞期間沒有外力作用。設 P_1 以速率 $u > 0$ 沿 x 軸趨近，並設碰撞後的速率分別為 v_1（P_1 的速率，方向在軸的上方，與軸成角 θ）和 v_2（P_2 的速率，方向在軸的下方，與軸成角 ϕ），其中 $0 < \theta < \pi$ 及 $0 < \phi < \pi$。



恆等式

沿原來的運動方向及與之垂直的方向應用動量守恆，（約去 m 後）得

$$v_1\cos\theta + kv_2\cos\phi = u,$$
$$v_1\sin\theta - kv_2\sin\phi = 0,$$

其中出現負號，是因為 P_2 向軸的另一側離開。由動能守恆，約去 $\frac{1}{2}m$ 後得

$$v_1^2 + kv_2^2 = u^2. \qquad (*)$$

兩條動量方程是關於兩個未知量 v_1 和 kv_2 的線性方程。把第一式乘以 $\sin\phi$，第二式乘以 $\cos\phi$，然後相加：含 kv_2 的項互相抵消，於是

$$v_1(\sin\theta\cos\phi + \cos\theta\sin\phi) = u\sin\phi, \qquad \text{所以} \qquad v_1 = \frac{u\sin\phi}{\sin(\theta+\phi)}.$$

同樣地，把第一式乘以 $\sin\theta$，第二式乘以 $\cos\theta$，然後相減，得

$$kv_2 = \frac{u\sin\theta}{\sin(\theta+\phi)}.$$

想一想 我們可以除以 $\sin(\theta+\phi)$ 嗎？

我們需要 $\theta+\phi \neq \pi$（它嚴格介乎 0 與 2π 之間）。若 $\theta+\phi = \pi$，則 $\sin\phi = \sin\theta$ 及 $\cos\phi = -\cos\theta$。第二條動量方程變成 $(v_1 - kv_2)\sin\theta = 0$，而 $\sin\theta \neq 0$，所以 $v_1 = kv_2$。這時第一條方程變成 $(v_1 - kv_2)\cos\theta = u$，即 $0 = u$，這是不可能的。因此 $\sin(\theta+\phi) \neq 0$，這個除法並無問題。

現在把結果代入 $(*)$，並把 kv_2^2 寫成 $(kv_2)^2/k$：

$$\frac{u^2\sin^2\phi}{\sin^2(\theta+\phi)} + \frac{u^2\sin^2\theta}{k\sin^2(\theta+\phi)} = u^2.$$

全式乘以 $k\sin^2(\theta+\phi)/u^2$（此數不等於零），得

$$\sin^2\theta + k\sin^2\phi = k\sin^2(\theta+\phi),$$

這正是所要證明的結果。

（從幾何角度看，動量方程表示動量 mv_1、kmv_2 和 mu 構成一個三角形，其內角為 ϕ、θ 和 $\pi-(\theta+\phi)$，而上面兩個表達式不過是這個三角形的正弦定理（sine rule）。）

兩條路徑互相垂直

若兩條路徑互相垂直，則 $\theta+\phi = \frac{\pi}{2}$，所以 $\sin^2(\theta+\phi) = 1$ 及 $\sin\theta = \cos\phi$。恆等式變成

$$\cos^2\phi + k\sin^2\phi = k, \qquad \text{即} \qquad \cos^2\phi = k(1-\sin^2\phi) = k\cos^2\phi.$$

這裏 $\cos\phi \neq 0$：否則 $\phi = \frac{\pi}{2}$，於是 $\theta = 0$，但題目已說明 θ 不等於零。兩邊除以 $\cos^2\phi$，得 $k = 1$，這正是所要證明的結果。

驗算

- 桌球中為人熟知的事實：兩個質量相等的球作彈性碰撞（elastic collision），而其中一個原本靜止，碰撞後兩球的運動方向互相垂直。我們的結果正是它的逆命題（converse）。
- 一個數值例子：取 $k = 2$ 及 $\phi = 30°$。沿該直線的光滑彈性碰撞會使 P_2 以速率 $\frac{2}{3}u\cos 30°$ 離開，而 P_1 則沿 $\theta = 90°$ 的方向運動。恆等式兩邊都等於 $\frac{3}{2}$。
- 我們亦模擬了 2×10^5 次隨機彈性碰撞（隨機選取 k、u 和碰撞線）。恆等式以及 v_1 和 kv_2 的兩條公式均成立，誤差在 10^{-12} 以內。

常見錯誤

- 把垂直方向的動量方程寫成 $v_1\sin\theta + kv_2\sin\phi = 0$：兩個質點向相反的兩側離開，所以兩個分量的符號相反。
- 在能量方程中漏掉 k，或在最後一步沒有說明為何 $\cos\phi \neq 0$。

STEP 之外

在隨質心（centre of mass）運動的參考系中，彈性碰撞只是把兩個速度旋轉，而大小保持不變。所有這些角度關係，都是把這個由各種可能結果組成的圓映射回實驗室參考系而得出的。這個觀點是處理散射（scattering）的標準方法，由桌球以至粒子物理學皆然。在粒子物理學中，人們利用量度所得的散射角來推算相撞粒子的質量比。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門）仔細闡述了以質心處理碰撞的方法。

詳細解答

STEP I 2002，第 12 題

概率與統計 • 幾何概率

> **STEP I 2002，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/02-s1-q12

關鍵思路

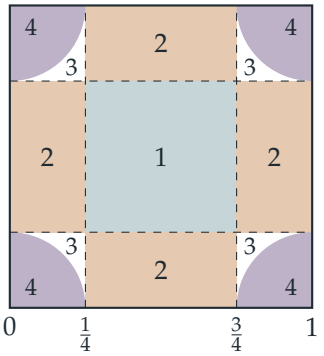
方格網不斷重複，所以「隨機放下馬蹄」的意思，就是馬蹄的中心在一個單位正方形上均勻分佈（uniformly distributed）。馬蹄碰到多少個方格，只取決於中心離該正方形的邊和角有多近，因此每個答案都對應正方形內的一個區域，而它的概率就是該區域的面積。求出四個概率之後，餘下的工作就是把這些概率及其平均值，與 Harry 的實際表現作比較。

解答

Harry 答案的分佈

設馬蹄的中心 C 位於單位正方形 $0 \leq x \leq 1$、$0 \leq y \leq 1$ 之內。若 Harry 隨機放下馬蹄，則由方格網的週期性（periodicity）可知 C 在這個正方形上均勻分佈，所以 C 落在某個區域內的概率就是該區域的面積。設 N 為被馬蹄部分覆蓋的方格數目。

馬蹄的半徑為 $\frac{1}{4}$，小於 $\frac{1}{2}$，所以它最多只能跨過一條垂直格線和一條水平格線。它跨過垂直格線，當且僅當 $x < \frac{1}{4}$ 或 $x > \frac{3}{4}$；它跨過水平格線，當且僅當 $y < \frac{1}{4}$ 或 $y > \frac{3}{4}$。（與格線的距離剛好等於 $\frac{1}{4}$ 的位置面積為零，所以可以不理。）



「跨過垂直格線」和「跨過水平格線」這兩個事件的概率各為 $\frac{1}{2}$，而且兩者獨立（independent），因為 x 和 y 是互相獨立的均勻坐標。所以：

- $N = 1$ 即馬蹄沒有跨過任何格線：此時 C 位於中央邊長為 $\frac{1}{2}$ 的正方形內，所以 $\mathrm{P}(N=1) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$。
- $N = 2$ 即馬蹄剛好跨過一條格線：$\mathrm{P}(N=2) = 2\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$。
- 當馬蹄跨過垂直和水平格線各一條時，C 位於四個邊長為 $\frac{1}{4}$ 的角落正方形其中之一，總概率為 $\frac{1}{4}$。這兩條格線相交於方格網的一個格點，而馬蹄覆蓋該格點周圍全部四個方格的一部分，當且僅當它覆蓋該格點本身。

想一想 在角落正方形內，為甚麼答案取決於與格點的距離？

考慮最接近原點的角落正方形，即 $0 < x < \frac{1}{4}$ 及 $0 < y < \frac{1}{4}$。馬蹄必定碰到 C 所在的方格、它左邊的方格和它下面的方格。至於對角的方格 $x < 0$、$y < 0$，其中最接近的點是原點，與 C 的距離為 $\sqrt{x^2+y^2}$，所以馬蹄進入這個方格，當且僅當 $\sqrt{x^2+y^2} < \frac{1}{4}$。因此，當 C 與某個格點的距離小於 $\frac{1}{4}$ 時 $N = 4$，否則 $N = 3$。

單位正方形內與其中一個頂點的距離小於 $\frac{1}{4}$ 的點，組成四個半徑為 $\frac{1}{4}$ 的四分之一圓盤，總面積為 $\pi\left(\frac{1}{4}\right)^2 = \frac{\pi}{16}$。因此

$$\mathrm{P}(N=1) = \tfrac{1}{4}, \qquad \mathrm{P}(N=2) = \tfrac{1}{2}, \qquad \mathrm{P}(N=3) = \tfrac{1}{4} - \tfrac{\pi}{16}, \qquad \mathrm{P}(N=4) = \tfrac{\pi}{16}.$$

目前的證據與「隨機的馬」是否吻合？

當正確答案是 $4 \div 4 = 1$ 時，一匹隨機的馬答對的概率是 $\mathrm{P}(N=1) = \frac{1}{4}$。當正確答案是 $2+2=4$ 時，牠答對的概率是 $\mathrm{P}(N=4) = \frac{\pi}{16} \approx 0.196$。這正正就是觀察所得的成功率，所以答案是肯定的：這些證據與我的懷疑完全吻合，即 Harry 只是隨機放下馬蹄。

答案的期望值，以及換一匹新馬

若 Harry 隨機放下馬蹄，他的答案便與題目無關，所以他的答案的期望值（expected value）就是 $\mathrm{E}(N)$（各個正確答案出現的頻率相同這一點，在這裏不起作用）：

$$\begin{aligned}\mathrm{E}(N) &= 1\cdot\tfrac{1}{4} + 2\cdot\tfrac{1}{2} + 3\left(\tfrac{1}{4} - \tfrac{\pi}{16}\right) + 4\cdot\tfrac{\pi}{16} \\ &= \tfrac{1}{4} + 1 + \tfrac{3}{4} + \tfrac{\pi}{16} \\ &= 2 + \tfrac{\pi}{16} \approx 2.20.\end{aligned}$$

把觀察所得的平均值 2 與兩個假設作比較。

- *Harry 計算正確*。正確答案是 $1, 2, 3, 4$，各自出現的次數相同，所以平均值應為 $\frac{5}{2}$。平均值 2 遠低於此，所以 Harry 肯定答錯了很多題。事實上，每個錯誤答案最多比正確答案小 3，而平均少了 $\frac{1}{2}$，所以他的答案中至少有 $\frac{1}{6}$ 是錯的。
- *Harry 是隨機的*。平均值應接近 $2 + \frac{\pi}{16} \approx 2.20$。單一答案的方差（variance）為 $\mathrm{E}(N^2) - \mathrm{E}(N)^2 = \frac{9}{2} + \frac{7\pi}{16} - \left(2 + \frac{\pi}{16}\right)^2 \approx 1.05$，所以做了 n 道題之後，平均值的標準差（standard deviation）約為 $1.02/\sqrt{n}$。題目「很多」的時候（例如 $n = 400$，標準差約為 0.05），平均值 2 低了大約四個標準差，純粹出於偶然的可能性極低。

所以，做了很多道題之後，平均值 2 即使對一匹隨機放下馬蹄的馬來說也太低：Harry 似乎偏好較小的答案，彷彿他下蹄很輕，或者瞄準方格的中央。無論如何，他並沒有在做計算，而且表現比一匹隨機的馬還要差。是的：是時候換一匹新馬了（不過也許該找一匹會做算術的）。

驗算

- 四個概率之和為一：$\frac{1}{4} + \frac{1}{2} + \frac{1}{4} - \frac{\pi}{16} + \frac{\pi}{16} = 1$，而 $\mathrm{P}(N = 3) = \frac{1}{4} - \frac{\pi}{16} \approx 0.054$ 為正數，這是必須的。
- 模擬（simulation）400 000 次隨機放下馬蹄，並數出與中心的距離小於 $\frac{1}{4}$ 的方格數目，所得的相對頻率為 0.2495、0.4990、0.0547、0.1968，平均值為 2.199，而理論值為 $2 + \frac{\pi}{16} \approx 2.196$。
- 合理性檢查：若馬蹄非常小，答案幾乎總是 1，平均值會趨向 1。馬蹄愈大，答案愈大，而我們的情況介乎兩者之間。

常見錯誤

- 以為馬蹄只要跨過兩條格線，就一定覆蓋四個方格。其實它還要碰到兩條格線相交的格點，而 π 正是因此而出現。

► 因為平均值 2「低於 2.5」，就把它當作隨機的證據。隨機的假設所預測的平均值約為 2.2，所以必須與這個數值比較，並考慮平均值的離散程度。

STEP 之外

這就是幾何概率（geometric probability）：當一點在某個區域上均勻分佈時，概率就是面積。同樣的想法可以用來處理蒲豐投針問題（Buffon's needle），其答案同樣涉及 π，十分著名。最後一部分讓你初嘗假設檢驗（hypothesis testing）的滋味：利用中心極限定理（central limit theorem）來判斷在某個模型之下，觀察所得的平均值是否合理。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）涵蓋了連續均勻分佈，以及這類論證背後的極限定理。

詳細解答

STEP I 2002，第 13 題

概率與統計 • 條件期望值

STEP I 2002，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/02-s1-q13

關鍵思路

(U, V) 只有六個可能的組合，所以整件事歸結為一個小小的聯合概率（joint probability）表，再對每種情況做少許代數運算。對於二次方程，V 的正負號決定一切，因為 U^2 最多是 1。對於三次方程，各項係數與 U 和 V 糾纏得令人起疑；這強烈暗示它可以因式分解（factorise），而一旦分解出來，各個根便一目了然。

解答

首先建立聯合分佈（joint distribution）。由於 $\mathrm{P}(U = u,\, V = v) = \mathrm{P}(U = u)\,\mathrm{P}(V = v \mid U = u)$，而對每個 u 都有 $\mathrm{P}(U = u) = \frac{1}{3}$，

	$U = 1$	$U = 0$	$U = -1$
$V = 1$	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{2}{9}$
$V = -1$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$

六個數值加起來是 1，理應如此。

第 (i) 部分

$x^2 + Ux + V = 0$ 的兩個根皆為實數，當且僅當判別式（discriminant）非負：$U^2 - 4V \geq 0$。

- 若 $V = -1$，判別式為 $U^2 + 4 > 0$，所以無論 U 取甚麼值，兩個根都是實數（而且相異）。
- 若 $V = 1$，由於 $U^2 \leq 1$，判別式為 $U^2 - 4 \leq 1 - 4 < 0$。所以兩個根永遠不是實數。

因此，兩個根皆為實數當且僅當 $V=-1$，而

$$\mathrm{P}(\text{兩根皆為實數})=\mathrm{P}(V=-1)=\tfrac{2}{9}+\tfrac{1}{6}+\tfrac{1}{9}=\tfrac{4+3+2}{18}=\tfrac{1}{2},$$

正是所求。

第 (ii) 部分

根據第 (i) 部分，以「兩根皆為實數」為條件，等同於以 $V=-1$ 為條件。把表中最下一行除以 $\mathrm{P}(V=-1)=\frac{1}{2}$，得

$$\mathrm{P}(U=1\mid V=-1)=\tfrac{4}{9},\qquad \mathrm{P}(U=0\mid V=-1)=\tfrac{1}{3},\qquad \mathrm{P}(U=-1\mid V=-1)=\tfrac{2}{9}.$$

? **想一想** 為甚麼這裏不能直接用 $\mathrm{P}(U=u)=\frac{1}{3}$？

因為 U 與 V 並非獨立（independent）：知道 $V=-1$ 之後，$U=1$ 的可能性便會提高（因為當 $U=1$ 時，$V=-1$ 的概率是 $\frac{2}{3}$）。我們要求的是條件期望值，所以權重必須是條件概率（conditional probability）。

當 $V=-1$ 時，方程為 $x^2+Ux-1=0$，其較大的根是

$$R=\frac{-U+\sqrt{U^2+4}}{2}.$$

所以當 $U=1$ 時 $R=\frac{1}{2}(\sqrt{5}-1)$，當 $U=0$ 時 $R=1$，當 $U=-1$ 時 $R=\frac{1}{2}(\sqrt{5}+1)$。因此

$$\begin{aligned}\mathrm{E}(R\mid\text{兩根為實數})&=\tfrac{4}{9}\cdot\frac{\sqrt{5}-1}{2}+\tfrac{1}{3}\cdot 1+\tfrac{2}{9}\cdot\frac{\sqrt{5}+1}{2}\\&=\frac{4\sqrt{5}-4+6+2\sqrt{5}+2}{18}\\&=\frac{6\sqrt{5}+4}{18}=\frac{3\sqrt{5}+2}{9}.\end{aligned}$$

第 (iii) 部分

常數項是 U，這提示我們有因式 $(x+U)$。果然

$$\begin{aligned}(x+U)(x^2-2Vx+1)&=x^3-2Vx^2+x+Ux^2-2UVx+U\\&=x^3+(U-2V)x^2+(1-2UV)x+U,\end{aligned}$$

這正是題目給出的三次式。所以方程的根是 $x=-U$，連同 $x^2-2Vx+1=0$ 的根。由於 $V=\pm1$ 給出 $V^2=1$，這個二次方程的判別式為 $4V^2-4=0$。所以它有重根（repeated root）$x=V$，而三次方程就是

$$(x+U)(x-V)^2=0,$$

其根為 $-U$、V、V。

三個根全為正數，當且僅當 $-U>0$ 且 $V>0$，即 $U=-1$ 且 $V=1$。（當 $U=0$ 時，其中一個根是 0，並非正數。）從表中可得

$$\mathrm{P}(\text{所有根皆為正數})=\mathrm{P}(U=-1,\,V=1)=\tfrac{1}{3}\cdot\tfrac{2}{3}=\tfrac{2}{9}.$$

驗算

- 條件概率 $\frac{4}{9},\frac{1}{3},\frac{2}{9}$ 加起來是 1。
- 第 (ii) 部分的答案約為 0.967。它必須介乎較大根的最小可能值 $\frac{1}{2}(\sqrt{5}-1)\approx 0.618$ 與最大可能值 $\frac{1}{2}(\sqrt{5}+1)\approx 1.618$ 之間，而且應該低於中間值 1，因為最小的根所佔的權重最大。結果確是如此。
- 在第 (iii) 部分，試代入 $U=-1$、$V=1$：三次式變成 $x^3-3x^2+3x-1=(x-1)^3$，根為 $1,1,1$，全是正數。再試 $U=1$、$V=1$：三次式變成 $x^3-x^2-x+1=(x+1)(x-1)^2$，它有負根 -1。
- 我們以分數對六種情況逐一精確枚舉，並在每種情況下以數值方法求出三次方程的根，從而確認了全部三個答案。

常見錯誤

- 在第 (ii) 部分使用無條件的權重 $\frac{1}{3}$，結果得出 $\frac{1}{3}\left(\sqrt{5}+1\right)\approx 1.079$。
- 在第 (iii) 部分把 $U=0$ 也計算在內：等於 0 的根並非正數。
- 試圖用判別式或正負號變化來分析三次方程，而不去尋找因式分解。

STEP 之外

這道題的核心是條件期望值（conditional expectation）$\mathrm{E}(X\mid A)$，計算方法是以條件概率重新加權。到了大學，這個概念會發展成本身就是一個隨機變量（random variable）的條件期望值 $\mathrm{E}(X\mid Y)$，以及塔式法則（tower law）$\mathrm{E}\big(\mathrm{E}(X\mid Y)\big)=\mathrm{E}(X)$，後者能令許多困難的期望值計算變得輕而易舉。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門教材）對此有清晰的闡述。

詳細解答

STEP I 2002，第 14 題

概率與統計 • 概率中的等比級數

STEP I 2002，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/02-s1-q14

關鍵思路

以我總共按了多少個號碼作為條件（conditioning）。這個數目一旦確定，警鐘響起與我取得款項之間相隔的時間也隨之確定，於是警察先到的概率就是對應該段時間的 $1-\mathrm{e}^{-\lambda t}$。再對嘗試次數取平均，便得出答案。在兩種方案中，這個平均都會化為以 $q=\mathrm{e}^{-\lambda/r}$ 為公比的等比級數（geometric series），所以值得及早為 q 命名。

解答

順序按鍵

設 K 為我的號碼在 $1,2,\ldots,n$ 這列數中的位置。那麼 K 就是我嘗試的次數，而且對 $k=1,\ldots,n$ 有 $P(K=k)=\frac{1}{n}$。

假設 $K=k$。我每分鐘按 r 個號碼，所以相鄰兩個號碼相隔 $1/r$ 分鐘。時間由我按下第一個號碼的一刻起計。

- 若 $k=1$，第一個號碼就是正確的，我從未輸入過錯誤號碼，警察也不會接到通知。警察先到的概率是 $0=1-\mathrm{e}^{0}$。
- 若 $k\geq 2$，第一個號碼是錯的，所以警鐘在時間 0 響起，而正確號碼是第 k 個，在時間 $(k-1)/r$ 輸入。如果警察在 $t=(k-1)/r$ 分鐘內到達，他們就會在我取得款項之前到達，其概率為 $1-\mathrm{e}^{-\lambda(k-1)/r}$。

因此公式 $1-\mathrm{e}^{-\lambda(k-1)/r}$ 適用於所有 k，包括 $k=1$。設 A 為「警察在我取得款項之前到

達」這個事件。根據全概率公式（law of total probability），

$$P(A) = \sum_{k=1}^{n} P(K=k)\,P(A \mid K=k) = \sum_{k=1}^{n} \frac{1-\mathrm{e}^{-\lambda(k-1)/r}}{n},$$

正如題目所求。

為了化簡，設 $q = \mathrm{e}^{-\lambda/r}$。由於 $\lambda, r > 0$，我們有 $0 < q < 1$，而且 $\mathrm{e}^{-\lambda(k-1)/r} = q^{k-1}$。所以

$$\begin{aligned} P(A) &= 1 - \frac{1}{n}\sum_{k=1}^{n} q^{k-1} = 1 - \frac{1-q^n}{n(1-q)} \\ &= 1 - \frac{1-\mathrm{e}^{-\lambda n/r}}{n\left(1-\mathrm{e}^{-\lambda/r}\right)}, \end{aligned}$$

這裏用了等比級數求和，因為 $q \neq 1$，所以這一步成立。

隨機按鍵

現在每次嘗試都是從 $1, \ldots, n$ 中均勻地重新抽出一個號碼，與之前的嘗試獨立（independent）（我不記得自己試過哪些號碼）。所以每次嘗試正確的概率都是 $\frac{1}{n}$，而且各次互相獨立。若 N 是直至按中正確號碼為止（包括該次）的嘗試次數，則 N 服從幾何分佈（geometric distribution）：

$$P(N=k) = \frac{1}{n}\left(\frac{n-1}{n}\right)^{k-1}, \qquad k = 1, 2, 3, \ldots$$

想一想 當號碼可以重複時，關於時間的論證是否仍然成立？

成立。警察在第一個錯誤號碼出現時接到通知，而若 $N = k \geq 2$，那正是第一次嘗試，發生在時間 0。款項在第 k 次嘗試時到手，即時間 $(k-1)/r$。重複號碼只會改變 N 的分佈，不會改變時間的計算。所以與之前完全一樣，$P(A \mid N=k) = 1 - q^{k-1}$，而這對 $k = 1$ 也是對的。

記 $p = \frac{n-1}{n}$。那麼

$$\begin{aligned} P(A) &= \sum_{k=1}^{\infty} \frac{1}{n} p^{k-1}\left(1 - q^{k-1}\right) = \frac{1}{n}\sum_{k=1}^{\infty} p^{k-1} - \frac{1}{n}\sum_{k=1}^{\infty} (pq)^{k-1} \\ &= \frac{1}{n}\cdot\frac{1}{1-p} - \frac{1}{n}\cdot\frac{1}{1-pq}. \end{aligned}$$

把和式拆開是可以的，因為兩個級數都收斂（converge）：$0 \leq p < 1$ 而且 $0 < pq < 1$。第一項是 $\frac{1}{n}\cdot n = 1$，這只不過是說 N 必定是有限的。至於第二項，

$$n(1-pq) = n - (n-1)q,$$

所以

$$P(A) = 1 - \frac{1}{n - (n-1)\mathrm{e}^{-\lambda/r}},$$

正如題目所求。

驗算

- 當 $n = 1$ 時，永遠不會出現錯誤號碼，而兩條公式都給出 0：第一條是因為 $1 - q^1 = 1 - q$，第二條是因為分母等於 1。
- 當 $\lambda \to \infty$（警察即時到達）時，$q \to 0$，兩個答案都趨向 $1 - \frac{1}{n}$。這正是第一次嘗試出錯的概率，理應如此。
- 隨機按鍵應該較差。取 $n = 5$、$\lambda = 0.7$、$r = 3$，順序按鍵得 0.3382，隨機按鍵得 0.4543。兩種方案各做 2×10^5 次模擬試驗，結果分別是 0.339 和 0.456。

常見錯誤

- 把時間寫成 k/r 而不是 $(k-1)/r$。計時由第一個（錯誤的）號碼開始，所以第 k 個號碼是在 $k-1$ 個間隔之後出現。
- 在隨機的情況下，給每個 k 的概率都是 $\frac{1}{n}$，或者求和只加到 n 為止。既然容許重複，N 服從幾何分佈，而且沒有上界。

STEP 之外

這裏警察的到達時間是一個指數隨機變量（exponential random variable），而整個計算就是對隨機的 X 求 $P(T < X)$，方法是以 X 為條件。當 X 服從幾何分佈時，答案有簡潔的閉合形式（closed form），因為 $E\left(q^X\right)$ 正是 X 的概率生成函數（probability generating function）在 q 處的值。事實上，兩個答案都是 $1 - E\left(q^{N-1}\right)$。生成函數和指數分佈是大學一年級概率論的核心內容。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）對兩者都有很好的講解。

詳細解答

STEP II 2002，第 1 題

純數學 • 代換積分法

STEP II 2002，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/02-s2-q1

關鍵思路

倍角公式（double-angle formula）把 $1-\cos 2\theta$ 化為 $2\sin^2\theta$，而 $\operatorname{cosec}^2\theta$ 是一眼就能積分的函數。題目建議的代換（substitution）$x=\sin 2\theta$ 是特意設計的，使 $\sqrt{1-x^2}$ 變成 $\cos 2\theta$，從而把第二個積分（integral）拉回第一個積分。至於最後一個積分，積分限（limits）1 和 $2/\sqrt{3}$ 正是積分限 1 和 $\sqrt{3}/2$ 的倒數（reciprocal），這強烈暗示應該試用 $y=1/x$。

解答

第一個積分

由於 $\cos 2\theta = 1-2\sin^2\theta$，我們有 $1-\cos 2\theta = 2\sin^2\theta$，它在 $[\frac{1}{6}\pi, \frac{1}{4}\pi]$ 上為正，所以被積函數（integrand）在該區間上性質良好。利用 $\frac{\mathrm{d}}{\mathrm{d}\theta}(\cot\theta) = -\operatorname{cosec}^2\theta$，

$$\begin{aligned}\int_{\pi/6}^{\pi/4} \frac{1}{1-\cos 2\theta}\,\mathrm{d}\theta &= \frac{1}{2}\int_{\pi/6}^{\pi/4} \operatorname{cosec}^2\theta\,\mathrm{d}\theta = \frac{1}{2}\Big[-\cot\theta\Big]_{\pi/6}^{\pi/4} \\ &= \frac{1}{2}\left(\cot\tfrac{\pi}{6} - \cot\tfrac{\pi}{4}\right) = \frac{1}{2}\left(\sqrt{3}-1\right) = \frac{\sqrt{3}}{2} - \frac{1}{2},\end{aligned}$$

正如所求。

第二個積分

設 $x=\sin 2\theta$，則 $\mathrm{d}x = 2\cos 2\theta\,\mathrm{d}\theta$。當 $\theta=\frac{1}{6}\pi$ 時，$x=\sin\frac{1}{3}\pi=\frac{\sqrt{3}}{2}$；當 $\theta=\frac{1}{4}\pi$ 時，$x=\sin\frac{1}{2}\pi=1$。當 θ 走遍 $[\frac{1}{6}\pi, \frac{1}{4}\pi]$ 時，x 由 $\frac{\sqrt{3}}{2}$ 穩定地遞增至 1。

想一想 $\sqrt{1-x^2}$ 等於 $\cos 2\theta$，還是等於 $-\cos 2\theta$？

我們有 $\sqrt{1-x^2} = \sqrt{\cos^2 2\theta} = |\cos 2\theta|$。在我們的範圍內 $2\theta \in [\frac{1}{3}\pi, \frac{1}{2}\pi]$，此處 $\cos 2\theta \geq 0$，所以 $\sqrt{1-x^2} = \cos 2\theta$。因此

$$\int_{\sqrt{3}/2}^{1} \frac{1}{1-\sqrt{1-x^2}}\,\mathrm{d}x = \int_{\pi/6}^{\pi/4} \frac{2\cos 2\theta}{1-\cos 2\theta}\,\mathrm{d}\theta.$$

為了引入第一個積分，把分子用分母來表示：

$$\frac{\cos 2\theta}{1-\cos 2\theta} = \frac{1-(1-\cos 2\theta)}{1-\cos 2\theta} = \frac{1}{1-\cos 2\theta} - 1.$$

於是，利用第一個結果，

$$\begin{aligned}\int_{\sqrt{3}/2}^{1} \frac{1}{1-\sqrt{1-x^2}}\,\mathrm{d}x &= 2\int_{\pi/6}^{\pi/4} \frac{1}{1-\cos 2\theta}\,\mathrm{d}\theta - 2\int_{\pi/6}^{\pi/4} 1\,\mathrm{d}\theta \\ &= 2\left(\frac{\sqrt{3}}{2} - \frac{1}{2}\right) - 2\left(\frac{\pi}{4} - \frac{\pi}{6}\right) \\ &= \sqrt{3} - 1 - \frac{\pi}{6},\end{aligned}$$

正如所求。

最後一個積分

設 $y = 1/x$，則 $\mathrm{d}y = -\mathrm{d}x/x^2$。當 $y = 1$ 時，$x = 1$；當 $y = 2/\sqrt{3}$ 時，$x = \sqrt{3}/2$。在這個範圍內 $x > 0$，所以

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

$$y\left(y - \sqrt{y^2-1}\right) = \frac{1}{x}\cdot\frac{1-\sqrt{1-x^2}}{x} = \frac{1-\sqrt{1-x^2}}{x^2}.$$

想一想 $\mathrm{d}y = -\mathrm{d}x/x^2$ 中的負號到哪裏去了？

這個代換把積分限的次序顛倒了，把它們調換回來時正好抵銷了負號：

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

x^2 這個因子剛好完全約去，這正是分母中要多放一個因子 y 的原因。

驗算

- 第二個和第三個積分的被積函數都是正的，所以答案必須為正。事實上 $\sqrt{3}-1-\frac{\pi}{6}\approx 0.73205-0.52360=0.20845$。
- 對三個積分分別使用辛普森法則（Simpson's rule），第一個得 $0.366025\ldots$（與 $\frac{\sqrt{3}}{2}-\frac{1}{2}$ 吻合），第二個和第三個得 $0.208452\ldots$（與 $\sqrt{3}-1-\frac{\pi}{6}$ 吻合）。
- 粗略估算：在第二個積分中，區間的長度為 $1-\frac{\sqrt{3}}{2}\approx 0.134$，而被積函數由 2（在 $x=\frac{\sqrt{3}}{2}$ 處）下降至 1（在 $x=1$ 處），所以積分值應介乎 0.134 與 0.268 之間。結果確實如此。

常見錯誤

- 沒有檢查 $\cos 2\theta$ 在該範圍內的正負號，就寫下 $\sqrt{1-\sin^2 2\theta}=\cos 2\theta$。
- 忘記更改積分限，或在 $y=1/x$ 令積分限次序顛倒時漏掉負號。
- 把 $\left[-\cot\theta\right]$ 算成 $\cot\frac{\pi}{4}-\cot\frac{\pi}{6}$，這樣得出的數值大小正確，但正負號錯了。

STEP 之外

最後一部分是一個普遍原則的小例子：代換是一種變數變換（change of variables），可以把一個積分準確地對應到另一個積分上，而留意到積分限互為倒數，就是 $y=1/x$ 能夠奏效的可靠提示。在大學的第一門分析（analysis）課程中，代換法會由鏈式法則（chain rule）和微積分基本定理（fundamental theorem of calculus）嚴格地推導出來，並要小心留意代換在何處單調（monotonic）、被積函數在何處連續（continuous），這些正是我們在這裏檢查過的條件。Abbott 的 *Understanding Analysis*（一本分析學入門教科書）對這些運算背後的理論有清晰的闡述。

詳細解答

STEP II 2002，第 2 題

純數學 • 倒數多項式方程

STEP II 2002，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/02-s2-q2

關鍵思路

由外至內觀察各項係數：先是 1 和 1，然後是 5 和 -5，最後是中間的 4。係數成對相等或成對異號，表示把方程除以 z 的中間次冪之後，只會剩下 $z^k + z^{-k}$ 和 $z^k - z^{-k}$ 的組合，而這些式子每一個都是 $w = z - z^{-1}$ 的多項式（polynomial）。這樣次數（degree）便減半。那個八次方程也有相同的結構，而且化簡之後，竟然再一次出現這種結構，所以這個技巧要用兩次。

解答

四次方程

由於 $z = 0$ 不是根（root）（常數項是 1），我們可以除以 z^2：

$$z^2 + 5z + 4 - 5z^{-1} + z^{-2} = 0, \qquad \text{即} \qquad \left(z^2 + z^{-2}\right) + 5\left(z - z^{-1}\right) + 4 = 0.$$

設 $w = z - z^{-1}$，則 $w^2 = z^2 - 2 + z^{-2}$，所以 $z^2 + z^{-2} = w^2 + 2$。代入得

$$(w^2 + 2) + 5w + 4 = 0 \quad \Longleftrightarrow \quad w^2 + 5w + 6 = 0,$$

正如題目所求。此式可因式分解（factorise）為 $(w + 2)(w + 3) = 0$，所以 $w = -2$ 或 $w = -3$。

要回到 z，把 $z - z^{-1} = w$ 兩邊乘以 z：

$$z^2 - wz - 1 = 0.$$

▶ $w = -2$：$z^2 + 2z - 1 = 0$，所以 $z = -1 \pm \sqrt{2}$。

► $w=-3$：$z^2+3z-1=0$，所以 $z=\frac{1}{2}\left(-3\pm\sqrt{13}\right)$。

這四個互不相同的值就是該四次方程的根：

$$z=-1\pm\sqrt{2},\qquad z=\frac{-3\pm\sqrt{13}}{2}.$$

八次方程

同樣，$z=0$ 不是根，所以我們除以 z^4，並把係數以中間項為中心對稱地配對：

$$2\left(z^4+z^{-4}\right)-3\left(z^3-z^{-3}\right)-12\left(z^2+z^{-2}\right)+12\left(z-z^{-1}\right)+22=0.$$

（核對各對係數：z^8 和 z^0 的係數都是 2；z^7 的係數是 -3，z^1 的係數是 $+3$；z^6 和 z^2 的係數都是 -12；z^5 的係數是 12，z^3 的係數是 -12。）

我們需要把每個括號都以 $w=z-z^{-1}$ 表示。已知 $z^2+z^{-2}=w^2+2$。取立方，

$$w^3=z^3-3z+3z^{-1}-z^{-3}\quad\Longrightarrow\quad z^3-z^{-3}=w^3+3w,$$

再把 z^2+z^{-2} 平方，

$$z^4+z^{-4}=\left(z^2+z^{-2}\right)^2-2=(w^2+2)^2-2=w^4+4w^2+2.$$

代入得

$$\begin{aligned}0&=2(w^4+4w^2+2)-3(w^3+3w)-12(w^2+2)+12w+22\\&=2w^4-3w^3-4w^2+3w+2.\end{aligned}$$

想一想 現在剩下一個 w 的四次方程。我們有甚麼得着嗎？

有：看看它的係數 2, −3, −4, 3, 2。它們的規律與之前完全一樣（最外面的一對相等，其次的一對異號），所以同一個技巧可以在下一層再用一次。由於 $w=0$ 不是根，除以 w^2：

$$2\left(w^2+w^{-2}\right)-3\left(w-w^{-1}\right)-4=0.$$

設 $v=w-w^{-1}$，於是 $w^2+w^{-2}=v^2+2$：

$$2(v^2+2)-3v-4=0\quad\Longleftrightarrow\quad 2v^2-3v=0\quad\Longleftrightarrow\quad v(2v-3)=0.$$

所以 $v=0$ 或 $v=\frac{3}{2}$。現在利用 $w^2-vw-1=0$ 逐層還原：

► $v=0$：$w^2=1$，所以 $w=\pm1$。

► $v=\frac{3}{2}$：$2w^2-3w-2=0$，即 $(2w+1)(w-2)=0$，所以 $w=2$ 或 $w=-\frac{1}{2}$。

最後利用 $z^2 - wz - 1 = 0$ 再還原一次：

- $w = 1$：$z^2 - z - 1 = 0$，所以 $z = \frac{1}{2}\left(1 \pm \sqrt{5}\right)$。
- $w = -1$：$z^2 + z - 1 = 0$，所以 $z = \frac{1}{2}\left(-1 \pm \sqrt{5}\right)$。
- $w = 2$：$z^2 - 2z - 1 = 0$，所以 $z = 1 \pm \sqrt{2}$。
- $w = -\frac{1}{2}$：$2z^2 + z - 2 = 0$，所以 $z = \frac{1}{4}\left(-1 \pm \sqrt{17}\right)$。

每個二次方程 $z^2 - wz - 1 = 0$ 的判別式（discriminant）都是 $w^2 + 4 > 0$，所以這些根全是實數，而且這八個值互不相同。由於八次多項式最多只有八個根，這些就是全部的根：

$$z = \frac{1 \pm \sqrt{5}}{2}, \qquad z = \frac{-1 \pm \sqrt{5}}{2}, \qquad z = 1 \pm \sqrt{2}, \qquad z = \frac{-1 \pm \sqrt{17}}{4}.$$

驗算

- 由 $z^2 - wz - 1 = 0$ 得出的每一對根，乘積都是 -1，可以用來快速檢查計算有沒有出錯：例如 $(1 + \sqrt{2})(1 - \sqrt{2}) = -1$。
- 對於八次方程，各根之和應為 $\frac{3}{2}$（即 z^7 的係數除以 z^8 的係數，再取負號）。四對根的貢獻是 $1 - 1 + 2 - \frac{1}{2} = \frac{3}{2}$。正確。
- 以數值計算，$z = \frac{1}{4}(-1 + \sqrt{17}) \approx 0.7808$ 代入八次多項式所得的值與零相差不超過 10^{-9}，其餘七個根以及四次方程的四個根亦然。

常見錯誤

- 寫成 $z^3 - z^{-3} = w^3$：交叉項會多出 $3w$。
- 得到 w 的四次方程後便停下來，嘗試用試根的方法把它因式分解，而沒有留意到它的係數具有相同的對稱性。
- 只給出 w（或 v）的值，忘記回到 z。

STEP 之外

如果一個多項式的係數倒過來讀仍然一樣（或者只差交錯的正負號），便稱為*倒數多項式*（reciprocal polynomial）或*迴文多項式*（palindromic polynomial）；它們的根成對出現，每對之間以 $z \mapsto 1/z$ 或 $z \mapsto -1/z$ 相聯繫，而本題的代換正是這種對稱性所容許的化簡。在伽羅瓦理論（Galois theory）中，這相當於觀察到方程在根的某一組變換所構成

的羣（group）之下保持不變，因此可以經由一連串較小的方程分階段求解，本題的情況正是如此：先是八次，然後四次，再到二次。Allenby 的 *Rings, Fields and Groups*（環、域與羣）是了解這些想法如何發展的好讀物。

詳細解答

STEP II 2002，第 3 題

純數學 • 費馬數與質數

> **STEP II 2002，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/02-s2-q3

本題在 *Beyond STEP* 第 1.1 節質數工廠中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch01/fermat-numbers.html

關鍵思路

寫出 $F_k - 2 = 2^{2^k} - 1$。把它乘以 $F_k = 2^{2^k} + 1$，便得到一個平方差（difference of two squares），而這正是歸納法（induction）所需要的那一步。乘積公式一旦成立，任何兩個費馬數（Fermat number）的公因數（common factor）都必須整除 2。由於每個費馬數都是奇數，這個因數只可能是 1。

解答

首四個費馬數

由於 $2^0 = 1$、$2^1 = 2$、$2^2 = 4$ 及 $2^3 = 8$，

$$F_0 = 2^1 + 1 = 3, \qquad F_1 = 2^2 + 1 = 5, \qquad F_2 = 2^4 + 1 = 17, \qquad F_3 = 2^8 + 1 = 257.$$

現在就 k 的首三個值驗證 $(*)$：

$$\begin{aligned} k = 1: &\quad F_0 = 3 = 5 - 2 = F_1 - 2, \\ k = 2: &\quad F_0 F_1 = 15 = 17 - 2 = F_2 - 2, \\ k = 3: &\quad F_0 F_1 F_2 = 15 \times 17 = 255 = 257 - 2 = F_3 - 2. \end{aligned}$$

證明 $(*)$ 對所有 $k \geq 1$ 成立

我們對 k 使用歸納法。$k = 1$ 的情況已在上面完成。

假設 $(*)$ 對某個 $k \geq 1$ 成立，即 $F_0F_1 \cdots F_{k-1} = F_k - 2$。那麼

$$\begin{aligned} F_0F_1 \cdots F_{k-1}F_k &= (F_k - 2)\, F_k \\ &= \left(2^{2^k} - 1\right)\left(2^{2^k} + 1\right) \\ &= \left(2^{2^k}\right)^2 - 1 \\ &= 2^{2\cdot 2^k} - 1 = 2^{2^{k+1}} - 1 \\ &= F_{k+1} - 2. \end{aligned}$$

所以 $(*)$ 對 $k+1$ 同樣成立，根據歸納法，它對每個 $k \geq 1$ 都成立。

想一想 為甚麼 $\left(2^{2^k}\right)^2$ 等於 $2^{2^{k+1}}$，而不是 $2^{2^{2k}}$ 或其他式子？

把一個冪平方，指數 (exponent) 便會加倍：$\left(2^a\right)^2 = 2^{2a}$。這裏 $a = 2^k$，而 $2a = 2 \cdot 2^k = 2^{k+1}$。歸納步驟的大部分工夫，就在於小心處理這類層層疊起的指數。

任何兩個費馬數都沒有公因數

取兩個不同的費馬數 F_m 和 F_n，其中 $m < n$。那麼 $n \geq 1$，而在 $k = n$ 時的 $(*)$ 中，F_m 是因數 $F_0, \ldots, F_{n-1}$ 的其中一個。所以

$$F_n - 2 = F_m \times (\text{某個整數}),$$

這表示 F_m 整除 $F_n - 2$。

現在假設正整數 d 同時整除 F_m 和 F_n。由於 $d \mid F_m$ 及 $F_m \mid F_n - 2$，我們有 $d \mid F_n - 2$。又由於 d 也整除 F_n，它便整除兩者之差：

$$d \mid F_n - (F_n - 2) = 2,$$

所以 $d = 1$ 或 $d = 2$。但 $F_m = 2^{2^m} + 1$ 是一個偶數加 1，所以是奇數（注意 $2^m \geq 1$）。因此 $d \neq 2$，從而 $d = 1$。由此可見，任何兩個費馬數都沒有大於 1 的公因數。

質數有無限多個

每個 F_n 都至少是 3，所以它至少有一個質因數（prime factor）。選取其中一個，稱它為 p_n。這樣，每個 $n = 0, 1, 2, \ldots$ 都對應一個質數。

想一想 兩個不同的費馬數會不會給出同一個質數？

不會。假如 $p_m = p_n$ 而 $m \neq n$，這個質數便會同時整除 F_m 和 F_n。它就成了一個大於 1 的公因數，而上一部分已排除了這個可能。所以質數 $p_0, p_1, p_2, \ldots$ 全部互不相同。由於費馬數有無限多個，互不相同的質數也有無限多個。

留意我們從來不需要 F_n 本身是質數，只需要每個費馬數都有某個質因數，是其他費馬數所沒有的。

驗算

- $k = 4$ 時的 $(*)$：$255 \times 257 = 65535 = 65537 - 2 = F_4 - 2$。我們亦用電腦以精確運算驗證了 $(*)$ 對直至 12 的所有 k 都成立。
- $F_0, \ldots, F_5$ 的最小質因數分別是 $3, 5, 17, 257, 65537, 641$，全部不同，正如論證所預測。$F_5 = 4294967297 = 641 \times 6700417$ 是第一個不是質數的費馬數。
- 我們以數值計算確認了對所有 $0 \leq m < n \leq 10$，都有 $\gcd(F_m, F_n) = 1$。

常見錯誤

- 把 2^{2^n} 讀成 $(2^2)^n = 4^n$。指數是 2^n，所以 $F_3 = 2^8 + 1$，而不是 $4^3 + 1$。
- 得出 $d \mid 2$ 之後，忘記排除 $d = 2$。正是「費馬數是奇數」這個事實，才令論證得以完成。
- 假設每個費馬數都是質數。最後一部分不需要這個假設，而且它根本是錯的（$641 \mid F_5$）。

STEP 之外

這個證明出自波利亞（Pólya）。它屬於「質數有無限多個」的一系列證明之一。這類證明都是構造一個兩兩互質（pairwise coprime）的無窮整數數列，於是每一項都貢獻一個屬於自己的新質數。歐幾里得（Euclid）的證明也可以這樣理解。費馬數在數論課程中還會再次出現：高斯（Gauss）證明了正 n 邊形可以用尺規作圖（ruler and compass construction），當且僅當 n 是 2 的冪乘以若干個互不相同的費馬質數（Fermat prime）之積，而至今仍未發現 F_4 之後的費馬質數。想了解更多，可閱讀 Davenport 的 *The Higher Arithmetic*（一本經典的數論入門書）。

詳細解答

STEP II 2002，第 4 題

純數學 • 由積分推導不等式

STEP II 2002，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/02-s2-q4

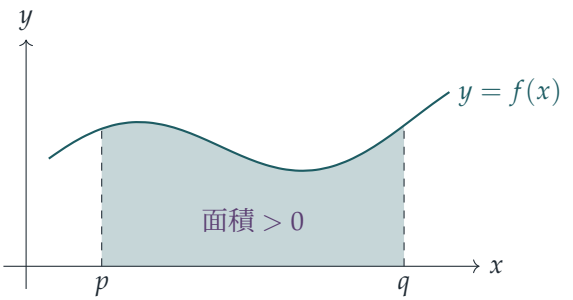
關鍵思路

二次式（quadratic）$as^2 - bs + c$ 若滿足 $a > 0$ 和 $b^2 < 4ac$，便沒有實根，而且首項係數為正，所以對每一個實數 s 都取正值。無論把 s 換成甚麼（x、$\sin x$ 或 $1/x$），所得的函數都是正的，因此它在任何區間上的積分（integral）都是正的。於是每一部分要做的，只是選好代換和區間，使積分剛好給出所求不等式（inequality）中的那些數。

解答

開首的略圖

積分 $\int_p^q f(x)\,\mathrm{d}x$ 是曲線 $y = f(x)$ 與 x 軸之間由 $x = p$ 至 $x = q$ 的有向面積（signed area）：在軸上方的面積算作正，在軸下方的面積算作負。若在整個 $p < x < q$ 上都有 $f(x) > 0$，整條曲線在該處便位於軸的上方，所以這個區域沒有負的部分，並且有真正（非零）的面積。因此積分嚴格為正。（本題所有函數都是連續（continuous）的，所以面積有明確定義。）



每一部分都會用到的事實

設 $g(s) = as^2 - bs + c$，其中 $a > 0$ 及 $b^2 < 4ac$。配方（completing the square）後得

$$g(s) = a\left(s - \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a}.$$

由於 $a > 0$，第一項至少為 0；又由於 $4ac - b^2 > 0$ 及 $a > 0$，第二項嚴格為正。所以對所有實數 s 都有 $g(s) > 0$。

第 (i) 部分

取 $f(x) = ax^2 - bx + c = g(x)$，它對所有 x 都是正的。我們可以在任何喜歡的區間上積分，而 $[0, 1]$ 給出的數最簡單：

$$0 < \int_0^1 (ax^2 - bx + c)\,\mathrm{d}x = \frac{a}{3} - \frac{b}{2} + c.$$

乘以 6 得 $2a - 3b + 6c > 0$，即 $3b < 2a + 6c$，這正是所要證明的。

第 (ii) 部分

現在 $f(x) = a\sin^2 x - b\sin x + c = g(\sin x)$。既然 g 在每一個實數上都取正值，它在 $\sin x$ 上當然也取正值，所以對所有 x 都有 $f(x) > 0$。

> **想一想** 我們應該在哪一個區間上積分？

目標不等式中，π 乘在 a 和 c 上，卻沒有乘在 b 上。在 $[0, \frac{\pi}{2}]$ 上，$\int \sin x\,\mathrm{d}x = 1$ 不含 π，而 $\int \sin^2 x\,\mathrm{d}x$ 和 $\int 1\,\mathrm{d}x$ 則含有。利用 $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$，

$$\int_0^{\pi/2} \sin^2 x\,\mathrm{d}x = \left[\frac{x}{2} - \frac{\sin 2x}{4}\right]_0^{\pi/2} = \frac{\pi}{4}, \qquad \int_0^{\pi/2} \sin x\,\mathrm{d}x = \Big[-\cos x\Big]_0^{\pi/2} = 1.$$

因此

$$0 < \int_0^{\pi/2} f(x)\,\mathrm{d}x = \frac{a\pi}{4} - b + \frac{c\pi}{2}.$$

乘以 4 得 $a\pi + 2c\pi - 4b > 0$，即 $4b < (a + 2c)\pi$，這正是所要證明的。

第 (iii) 部分

> **想一想** 哪一個函數的積分會給出 $\ln(q/p)$、$\frac{1}{p} - \frac{1}{q}$ 和 $q - p$？

它們分別是 $\int_p^q \frac{1}{x}\,\mathrm{d}x$、$\int_p^q \frac{1}{x^2}\,\mathrm{d}x$ 和 $\int_p^q 1\,\mathrm{d}x$。所以取

$$f(x) = \frac{a}{x^2} - \frac{b}{x} + c = g\left(\frac{1}{x}\right),$$

它在 $x > 0$ 時有定義而且連續，並且由於 g 處處為正，它在該處也是正的。因為 $q > p > 0$，區間 $[p,q]$ 避開了 $x = 0$，所以

$$0 < \int_p^q \left(\frac{a}{x^2} - \frac{b}{x} + c\right) \mathrm{d}x = \left[-\frac{a}{x} - b\ln x + cx\right]_p^q$$
$$= a\left(\frac{1}{p} - \frac{1}{q}\right) - b\ln\frac{q}{p} + c(q-p).$$

移項後得 $b\ln(q/p) < a\left(\frac{1}{p} - \frac{1}{q}\right) + c(q-p)$，這正是所要證明的。（條件 $p > 0$ 也讓我們可以寫出 $\ln q - \ln p = \ln(q/p)$，其中兩個對數都有定義。）

驗算

取 $a = b = c = 1$，則 $b^2 = 1 < 4 = 4ac$。這時 (i) 說 $3 < 8$，(ii) 說 $4 < 3\pi \approx 9.42$，而 (iii) 取 $p = 1$、$q = 2$ 時說 $\ln 2 \approx 0.693 < \frac{1}{2} + 1 = 1.5$。全部成立。我們還用 20 000 組符合條件的隨機三元組 (a,b,c) 測試了全部三個不等式，當中包括 b^2 僅略小於 $4ac$ 的情況，並以辛普森法則（Simpson's rule）核對了每個積分。留意若沒有這些條件，不等式可以不成立：$a = 1$、$b = 10$、$c = 1$ 會令 (i) 失效，因為 $30 > 8$。

常見錯誤

- 忘記說明為何 $f > 0$：條件 $a > 0$ 和 $b^2 < 4ac$ 正是令二次式取正值的原因，這一點必須寫出來。
- 在 (ii) 中，誤取 $\int_0^{\pi/2} \sin^2 x\,\mathrm{d}x = \frac{\pi}{2}$，而正確的值是 $\frac{\pi}{4}$。
- 在 (iii) 中，沒有提到 $p > 0$ 保證了 $x = 0$（f 在該處無定義）不在區間之內。

STEP 之外

「正的連續函數有正的積分」這個原理，是大學第一門分析課中為黎曼積分（Riemann integral）證明的基本性質之一。它與線性性質（linearity）結合起來，便是許多經典不等式的來源，其中包括柯西－施瓦茲不等式（Cauchy–Schwarz inequality）的積分形式：把平方 $(f + tg)^2 \geq 0$ 積分，再用上與本題相同的判別式（discriminant）論證，便可得出。Abbott 的 *Understanding Analysis*（一本分析學入門教科書）正是從這個角度仔細地建立積分理論。

詳細解答

STEP II 2002，第 5 題

純數學 • 邏輯斯諦映射的迭代

> **STEP II 2002，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/02-s2-q5

關鍵思路

一切都取決於拋物線（parabola）$y = kx(1-x)$，它在 $[0,1]$ 上的最大值是 $\frac{k}{4}$，在 $x = \frac{1}{2}$ 處取得。單憑這一點，就可以用歸納法（induction）解決第 (i) 部分。第 (ii) 和第 (iii) 部分先後要求找出不動點（fixed point）和週期（period）為二的點。第 (iii) 部分的竅門是把 x_1 用 b 表示：$x_1 = ka(1-a) = ab$。這樣 $x_2 = a$ 就變成一條關於 b 的三次方程（cubic），而這條三次方程有一個明顯的根（root），正好就是第 (ii) 部分的不動點。

解答

第 (i) 部分

我們對 n 使用歸納法。$n = 0$ 的情況是已知條件。假設 $0 < x_n < 1$，則 $x_n > 0$ 且 $1 - x_n > 0$，所以 $x_n(1-x_n) > 0$。利用配方法（completing the square），

$$x_n(1-x_n) = \tfrac{1}{4} - \left(x_n - \tfrac{1}{2}\right)^2 \leq \tfrac{1}{4}.$$

由於 $0 < k < 4$，各項乘以 k 後不等式（inequality）方向不變，於是

$$0 < x_{n+1} = kx_n(1-x_n) \leq \tfrac{k}{4} < 1.$$

所以 $0 < x_{n+1} < 1$，由歸納法可知對所有 n 都有 $0 < x_n < 1$。留意每個假設在哪裏用到：$k > 0$ 用於下界，$k < 4$ 用於上界。

第 (ii) 部分

要得到常數數列（constant sequence），只需 $x_1 = x_0$，即 $a = ka(1-a)$。由於 $a \neq 0$，可以除以 a 得 $1 = k(1-a)$；又由於 $a \neq 1$，可以除以 $1-a$：

$$k = \frac{1}{1-a}.$$

反過來，取這個 k 時有 $x_1 = a$，因此其後每一項也都是 a。

第 (iii) 部分

設 $b = k(1-a)$，則

$$x_1 = ka(1-a) = ab.$$

若 $x_2 = a$，數列便以週期二重複，所以整個條件就只是 $x_2 = a$：

$$k\,(ab)(1-ab) = a.$$

除以 $a \neq 0$，得 $kb(1-ab) = 1$。由於 $a \neq 1$，可以寫成 $k = \dfrac{b}{1-a}$，所以

$$\frac{b^2(1-ab)}{1-a} = 1 \quad \Longrightarrow \quad b^2 - ab^3 = 1-a \quad \Longrightarrow \quad ab^3 - b^2 + (1-a) = 0,$$

這就是所要求的結果。

想一想 這條三次方程有哪一個根可以一眼看出？它代表甚麼？

代入 $b = 1$ 得 $a - 1 + 1 - a = 0$，所以 $b = 1$ 是一個根。但 $b = 1$ 即是 $k(1-a) = 1$，這正是第 (ii) 部分的不動點：此時 $x_1 = ab = a$。所以額外條件 $x_1 \neq a$，即 $ab \neq a$，恰好就是 $b \neq 1$。除去因式 $b - 1$：

$$ab^3 - b^2 + (1-a) = (b-1)\big(ab^2 + (a-1)b + (a-1)\big),$$

你可以展開來驗證。所以 b 必須是

$$ab^2 - (1-a)b - (1-a) = 0$$

的一個不等於 1 的根。由於 $a \neq 0$，這確實是一條二次方程（quadratic），其根為

$$b = \frac{(1-a) \pm \sqrt{(1-a)^2 + 4a(1-a)}}{2a} = \frac{(1-a) \pm \sqrt{(1-a)(1+3a)}}{2a}.$$

要有實根，需要 $(1-a)(1+3a) \geq 0$，即 $-\frac{1}{3} \leq a \leq 1$。連同 $a \neq 0, 1$，餘下的範圍是 $-\frac{1}{3} \leq a < 1$，$a \neq 0$。對於 a 的任何其他值，都沒有實數 k 符合要求。

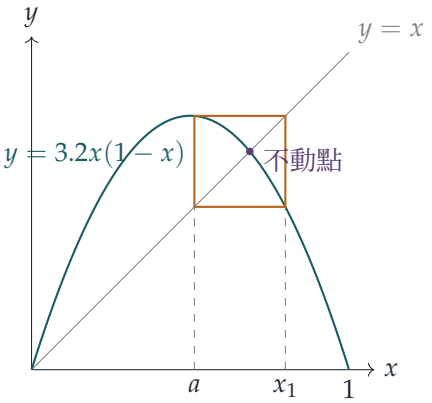
在這個範圍內 $1-a>0$，所以除以 $1-a$ 便得到形式簡潔的 $k=b/(1-a)$，當中用到 $\sqrt{(1-a)(1+3a)}/(1-a)=\sqrt{(1+3a)/(1-a)}$：

$$\boxed{k=\frac{1}{2a}\left(1\pm\sqrt{\frac{1+3a}{1-a}}\right),\qquad -\tfrac{1}{3}\le a<1,\ a\neq 0.}$$

最後還有一件事：我們必須確保這兩個值都不是被排除的根 $b=1$。這條二次方程只有在 $a+(a-1)+(a-1)=0$，即 $a=\frac{2}{3}$ 時，才以 $b=1$ 為根。這時 $\sqrt{(1+3a)/(1-a)}=3$，兩個值分別是 $k=3$（即 $b=1$，也就是不動點，故須排除）和 $k=-\frac{3}{2}$。所以：

- 當 $a=\frac{2}{3}$ 時，唯一的可能是 $k=-\frac{3}{2}$；
- 當 $a=-\frac{1}{3}$ 時，正負兩個符號給出相同的值，$k=-\frac{3}{2}$；
- 對於其餘所有容許的 a，兩個符號都給出真正的 2-循環（cycle）。

下圖顯示了其中一個這樣的循環，取 $k=3.2$。軌道（orbit）在 $a\approx 0.513$ 與 $x_1\approx 0.800$ 之間來回跳動，在拋物線與直線 $y=x$ 之間描出一個正方形。



驗算

- 取 $a=0.7$。公式給出 $k\approx 3.0104$ 或 $k\approx -1.5818$。取 $k\approx 3.0104$ 時，得 $x_1\approx 0.6322$，然後 $x_2\approx 0.7000$，所以軌道的週期確實是二。
- 當 $a=\frac{2}{3}$ 時，$k=3$ 給出 $x_1=3\cdot\frac{2}{3}\cdot\frac{1}{3}=\frac{2}{3}$，即不動點，與預期相符。取 $k=-\frac{3}{2}$ 時，得 $x_1=-\frac{1}{3}$ 和 $x_2=-\frac{3}{2}\cdot(-\frac{1}{3})\cdot\frac{4}{3}=\frac{2}{3}$，是真正的 2-循環。
- 數值上，對數以千計隨機選取的 $a\in[-\frac{1}{3},1)$，不論取哪個符號，$x_2=a$ 都成立。對於該區間以外的 a，在 k 的範圍內逐一搜尋，除了不動點 $k=1/(1-a)$ 之外找不到任何解。

常見錯誤

- 忘記三次方程必定包含不動點那個解。看出 $b = 1$ 這個根，正是令三次方程變得容易處理的關鍵。
- 遺漏實根條件 $-\frac{1}{3} \leq a < 1$，或沒有留意到當 $a = \frac{2}{3}$ 時，k 的兩個值之中有一個是不動點，必須捨去。
- 在第 (i) 部分只證明了 $x_{n+1} \leq \frac{1}{4}k$，卻沒有指出因為 $k < 4$，這個值嚴格小於 1。

STEP 之外

這就是邏輯斯諦映射（logistic map），是「規則簡單而行為極其豐富」的經典例子。當 k 在通常的範圍 $0 < k \leq 4$ 內，你找到的 2-循環恰好在 $k > 3$ 時存在：在 $k = 3$ 處，不動點失去穩定性，並分裂成兩個點。這是第一次*倍週期分岔*（*period-doubling bifurcation*）。其後還會繼續倍增，出現倍增的 k 值以幾何方式愈聚愈密（其比率由費根鮑姆（Feigenbaum）常數決定），直至映射進入混沌(chaotic)狀態。斯托加茨(Strogatz)的 *Nonlinear Dynamics and Chaos*（《非線性動力學與混沌》）以大學一年級的程度講述了整個故事。

詳細解答

STEP II 2002，第 6 題

純數學 • 斜面上的直線

> **STEP II 2002，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/02-s2-q6

關鍵思路

整道題都建基於斜面（inclined plane）的一個事實：若斜面上的一條直線與最大斜率線（line of greatest slope）成角 ψ，則它與水平面所成的角 γ 滿足

$$\sin\gamma = \cos\psi \, \sin\beta,$$

其中 β 是最大斜率線的傾角。有了這個結果，題目中每條直線都給出一條三角方程，而 l_1 與 l_3 互相垂直，則把一個餘弦變成正弦。之後就只是代數運算了。

解答

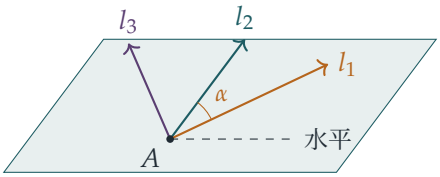
斜面上一條直線的斜度

以 A 為原點建立坐標軸，z 軸鉛垂向上。設 $\mathbf{h}$ 為位於 P 內的水平單位向量（unit vector），並設 $\mathbf{u}$ 為沿最大斜率線 l_2 向上的單位向量，它在 P 內與 $\mathbf{h}$ 垂直。由於 l_2 與水平面成角 β，我們可取

$$\mathbf{h} = (1,0,0), \qquad \mathbf{u} = (0, \cos\beta, \sin\beta).$$

在 P 內與 $\mathbf{u}$ 成角 ψ 的單位向量是 $\mathbf{d} = \cos\psi\,\mathbf{u} + \sin\psi\,\mathbf{h}$，因為 $\mathbf{u}$ 與 $\mathbf{h}$ 標準正交（orthonormal），並且張成（span）P。它的鉛垂分量是 $\cos\psi\sin\beta$。若一條直線的單位方向向量為 $\mathbf{d}$，則它與水平面所成的角 γ 滿足 $\sin\gamma =$ |鉛垂分量，取自$\mathbf{d}$|，所以

$$\sin\gamma = |\cos\psi| \, \sin\beta. \tag{$*$}$$



第一個位置

在 P 內，直線 l_1 與 l_2 成銳角（acute angle）α，而它與水平面所成的角是 $\pi/6$。由 $(*)$，

$$\cos\alpha\sin\beta = \sin\tfrac{\pi}{6} = \tfrac{1}{2},$$

正是所要證明的結果（這裏 $\cos\alpha > 0$，因為 α 是銳角）。

? **想一想** l_3 與 l_2 所成的角是多少？

三條直線都位於同一平面 P 內，並且都通過 A。由於 $l_3 \perp l_1$，l_3 的方向就是 l_1 的方向在 P 內轉過一個直角所得的方向，所以 l_3 所成的角是 $\alpha \pm \frac{\pi}{2}$，這是相對於 l_2 而言的（正負號取決於轉動的方向，但並不影響結果）。無論哪一種情況，都有 $|\cos(\alpha \pm \frac{\pi}{2})| = \sin\alpha$，因為 α 是銳角。於是把 $(*)$ 用於 l_3，得

$$\sin\alpha\sin\beta = \sin\tfrac{\pi}{4} = \tfrac{1}{\sqrt{2}}.$$

把兩個結果平方後相加，

$$\sin^2\beta\,(\cos^2\alpha + \sin^2\alpha) = \tfrac{1}{4} + \tfrac{1}{2}, \qquad \text{故} \qquad \sin^2\beta = \tfrac{3}{4}.$$

直線與水平面所成的角 β 位於 $[0, \frac{\pi}{2}]$ 之內，所以 $\sin\beta = \frac{\sqrt{3}}{2}$，即 $\beta = \pi/3$，正是所要證明的結果。（順帶一提，把兩條方程相除可得 $\tan\alpha = \sqrt{2}$，這確實對應一個銳角，所以這個位置是存在的。）

轉動後的位置

平面沒有改變，因此 l_2 和 $\beta = \pi/3$ 也沒有改變。現在 l_1 與 l_2 成銳角 θ，而 l_3 所成的角是 $\theta \pm \frac{\pi}{2}$，與之前完全一樣。把 $(*)$ 分別用於兩條直線，並代入 $\sin\beta = \frac{\sqrt{3}}{2}$：

$$\sin\phi = \tfrac{\sqrt{3}}{2}\cos\theta,$$
$$\sin 2\phi = \tfrac{\sqrt{3}}{2}\sin\theta.$$

由於 2ϕ 是與水平面所成的角，故 $0 \le 2\phi \le \frac{\pi}{2}$，而且 $\phi \ne 0$，因為 $\cos\theta > 0$。所以 $\sin\phi > 0$，我們可以把第二條方程除以第一條：

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

設 $t = \tan^2\theta$，則 $\cos^2\phi = t/4$ 及 $\sin^2\phi = 1 - t/4$。為了消去 θ，我們把 $\cos^2\theta = 1/(1+\tan^2\theta)$ 代入第一條方程的平方：

$$1 - \frac{t}{4} = \frac{3}{4}\cdot\frac{1}{1+t} \quad\Longrightarrow\quad (4-t)(1+t) = 3 \quad\Longrightarrow\quad t^2 - 3t - 1 = 0.$$

因此 $t = \frac{1}{2}\left(3 \pm \sqrt{13}\right)$。由於 $t = \tan^2\theta \geq 0$ 而 $\sqrt{13} > 3$，我們必須取正號：

$$\tan^2\theta = \frac{3+\sqrt{13}}{2},$$

正是所要證明的結果。

這個根確實可以達到：它給出 $\cos^2\phi = t/4 \approx 0.826$，所以 $\phi \approx 24.7°$，$2\phi \approx 49.4° < 90°$，而 $\theta \approx 61.2°$ 是銳角。

驗算

- 在第一個位置，$\tan\alpha = \sqrt{2}$ 給出 $\cos\alpha = 1/\sqrt{3}$，$\sin\alpha = \sqrt{2/3}$，再配合 $\sin\beta = \sqrt{3}/2$，兩個乘積分別是 $\frac{1}{2}$ 和 $1/\sqrt{2}$，與應有的結果相符。
- 以數值計算，$\tan^2\theta \approx 3.3028$，所以 $\tan\theta \approx 1.8174$，$\theta \approx 61.18°$；於是 $\sin\phi = \frac{\sqrt{3}}{2}\cos\theta \approx 0.4175$ 給出 $\phi \approx 24.68°$，而 $\frac{\sqrt{3}}{2}\sin\theta \approx 0.7589 \approx \sin(49.35°) = \sin 2\phi$。
- $(*)$ 的極限情況：$\psi = 0$ 給出 l_2 本身（所成的角為 β），而 $\psi = \frac{\pi}{2}$ 給出一條水平線（所成的角為 0）。

常見錯誤

- 誤用 $\tan\gamma = \cos\psi\tan\beta$ 而不是 $\sin\gamma = \cos\psi\sin\beta$。只有當 ψ 是在水平的平面圖上量度時，正切的版本才正確，但這裏 α 和 θ 都是在 P 內量度的角。
- 忘記捨去負根 $\frac{1}{2}(3-\sqrt{13})$：平方數不可能是負數。

STEP 之外

公式 $(*)$ 其實就是一個點積（dot product）：直線與水平面所成角的正弦，就是它的單位向量在鉛垂方向上的分量，而把這個向量用平面的一組標準正交基（orthonormal basis）表示出來，便完成了全部工作。把平面看成標準正交向量所張成的空間，把角看成點積，正是大學一年級線性代數（linear algebra）以及旋轉幾何的起點。Strang 的 *Introduction to Linear Algebra*（線性代數入門）對這些想法有細緻的闡述。

詳細解答

STEP II 2002，第 7 題

純數學 • 直線的向量幾何

> **STEP II 2002，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/02-s2-q7

關鍵思路

這裏的一切都歸結為點積（dot product）。一條通過原點、單位方向向量為 $\mathbf{u}$ 的直線，與 m_1 的夾角為 $\pi/4$，當且僅當 $|\mathbf{u}\cdot\mathbf{d}_1| = |\mathbf{d}_1|\cos\frac{\pi}{4}$，於是第一部分就化為聯立方程。至於後半部分，「垂直」即點積等於零；而空間中兩條直線只有在共面（coplanar）且不平行時才會相交，共面則等價於純量三重積（scalar triple product）等於零。出人意表的是，共面的那個情況原來正是平行的情況。

解答

直線 m_3 和 m_4

設 $\mathbf{d}_1 = \mathbf{i}+\mathbf{j}$ 及 $\mathbf{d}_2 = \mathbf{i}+\mathbf{k}$，兩者的長度均為 $\sqrt{2}$。設 $\mathbf{u} = a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ 為沿着一條通過 O 的直線的單位向量（unit vector）。兩條直線之間的夾角是指它們之間的銳角，所以條件是

$$|\mathbf{u}\cdot\mathbf{d}_1| = \sqrt{2}\cos\tfrac{\pi}{4} = 1, \qquad |\mathbf{u}\cdot\mathbf{d}_2| = 1,$$

即 $a+b = \pm 1$ 及 $a+c = \pm 1$。以 $-\mathbf{u}$ 取代 $\mathbf{u}$ 會得出同一條直線，並把兩個正負號同時反轉，所以只需考慮兩種情況。

同號：$a+b=1$，$a+c=1$。這時 $b=c=1-a$，而由 $a^2+b^2+c^2=1$ 得

$$a^2+2(1-a)^2 = 1 \iff 3a^2-4a+1=0 \iff (3a-1)(a-1)=0.$$

所以 $a=1$，得 $\mathbf{u}=\mathbf{i}$；或 $a=\frac{1}{3}$，得 $\mathbf{u}=\frac{1}{3}(\mathbf{i}+2\mathbf{j}+2\mathbf{k})$。

異號：$a+b=1$，$a+c=-1$。這時 $b=1-a$，$c=-1-a$，而由 $a^2+(1-a)^2+(1+a)^2=1$ 得 $3a^2+2=1$，此方程沒有實數解。

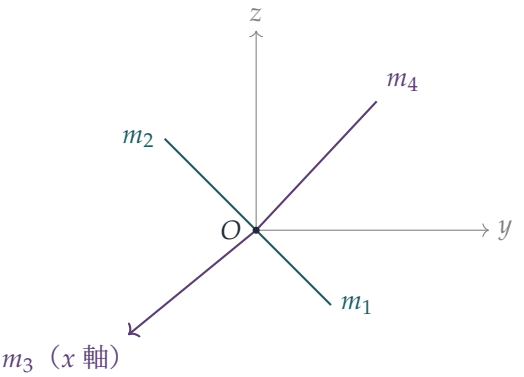
所以這兩條直線的方向為

$$m_3:\ \mathbf{i}, \qquad m_4:\ \mathbf{i}+2\mathbf{j}+2\mathbf{k}$$

（兩個標記對調亦無不可）。取單位向量 $\mathbf{e}_3=\mathbf{i}$ 及 $\mathbf{e}_4=\frac{1}{3}(\mathbf{i}+2\mathbf{j}+2\mathbf{k})$，則

$$\cos\theta=\mathbf{e}_3\cdot\mathbf{e}_4=\tfrac{1}{3},$$

此值為正數，所以這就是銳角：其餘弦為 $\frac{1}{3}$。



設定各點

由於 A 位於 m_1 上，且與 O 的距離為 $\lambda\sqrt{2}$，它就是 $\pm\lambda(1,1,0)$；坐標須為非負數，故我們必須取 $A=\lambda(1,1,0)$，其中 $\lambda\geq 0$（由於 $\lambda\sqrt{2}$ 是距離，我們可以全程取 $\lambda\geq 0$）。同樣地，$B=\lambda(1,0,1)$。在 m_3, m_4 上坐標為非負數的單位點是 $(1,0,0)$ 和 $\frac{1}{3}(1,2,2)$。

當中哪一點是 P、哪一點是 Q，取決於 m_3 和 m_4 的標記方式。但這並不要緊：反射 $(x,y,z)\mapsto(x,z,y)$ 把 A 和 B 對調，並保持 $(1,0,0)$ 和 $\frac{1}{3}(1,2,2)$ 兩點不變，所以它把這個構形的一種標記方式變為另一種（直線 AQ 和 BP 的角色互換）。反射保持垂直關係和相交關係，所以只需取

$$A=(\lambda,\lambda,0),\quad B=(\lambda,0,\lambda),\quad P=(1,0,0),\quad Q=\left(\tfrac{1}{3},\tfrac{2}{3},\tfrac{2}{3}\right).$$

於是

$$\overrightarrow{AQ}=\left(\tfrac{1}{3}-\lambda,\ \tfrac{2}{3}-\lambda,\ \tfrac{2}{3}\right),\qquad \overrightarrow{BP}=(1-\lambda,\ 0,\ -\lambda).$$

第 (i) 部分

AQ 垂直於 BP，當且僅當 $\overrightarrow{AQ}\cdot\overrightarrow{BP}=0$：

$$\begin{aligned}\overrightarrow{AQ}\cdot\overrightarrow{BP}&=\left(\tfrac{1}{3}-\lambda\right)(1-\lambda)+0-\tfrac{2}{3}\lambda\\&=\lambda^2-\tfrac{4}{3}\lambda+\tfrac{1}{3}-\tfrac{2}{3}\lambda=\lambda^2-2\lambda+\tfrac{1}{3}.\end{aligned}$$

當 $(\lambda-1)^2=\frac{2}{3}$ 時，上式等於零，即

$$\lambda=1\pm\sqrt{\tfrac{2}{3}}=1\pm\tfrac{\sqrt{6}}{3}.$$

兩個值都是正數（因為 $\sqrt{6}<3$），所以兩者都符合非負條件，而這個二次方程再沒有其他根。因此恰好有兩個 λ 的值使得 $AQ\perp BP$，正如題目所求。

第 (ii) 部分

> **想一想** 怎樣判斷三維空間中的兩條直線是否相交？

兩條相異的直線相交，當且僅當它們位於同一平面上而且不平行。直線 AQ 和 BP 位於同一平面上，當且僅當 A、B、P、Q 四點共面，而這又當且僅當 $\overrightarrow{AB}$、$\overrightarrow{AP}$、$\overrightarrow{AQ}$ 的純量三重積等於零。

這裏 $\overrightarrow{AB}=(0,-\lambda,\lambda)$ 及 $\overrightarrow{AP}=(1-\lambda,-\lambda,0)$，所以

$$\begin{vmatrix}0 & -\lambda & \lambda\\ 1-\lambda & -\lambda & 0\\ \frac{1}{3}-\lambda & \frac{2}{3}-\lambda & \frac{2}{3}\end{vmatrix}=\lambda\cdot\tfrac{2}{3}(1-\lambda)+\lambda\Big[(1-\lambda)\big(\tfrac{2}{3}-\lambda\big)+\lambda\big(\tfrac{1}{3}-\lambda\big)\Big]$$

$$=\lambda\Big[\tfrac{2}{3}-\tfrac{2}{3}\lambda+\tfrac{2}{3}-\tfrac{4}{3}\lambda\Big]=2\lambda\big(\tfrac{2}{3}-\lambda\big),$$

這是沿第一行展開的結果。因此，對於 $\lambda\neq0$，兩條直線只有在 $\lambda=\frac{2}{3}$ 時才有可能相交。

可是當 $\lambda=\frac{2}{3}$ 時，

$$\overrightarrow{AQ}=\big(-\tfrac{1}{3},\,0,\,\tfrac{2}{3}\big),\qquad\overrightarrow{BP}=\big(\tfrac{1}{3},\,0,\,-\tfrac{2}{3}\big)=-\overrightarrow{AQ},$$

所以兩條直線平行。它們並不是同一條直線，因為 $\overrightarrow{BA}=(0,\frac{2}{3},-\frac{2}{3})$ 不是 $\overrightarrow{BP}$ 的倍數，所以 A 不在 BP 上。兩條相異的平行線永不相交。

所以對於每一個非零的 λ，直線 AQ 和 BP 都不相交，正如題目所求。（當 $\lambda=0$ 時，$A=B=O$，兩條直線確實相交於 O，這正是題目把它排除在外的原因。）

驗算

- 一望而知，$\mathbf{i}$ 與 $\mathbf{i}+\mathbf{j}$ 的夾角為 $\pi/4$，這正好印證了二次方程的第一個根，令人安心。
- 當 $\lambda=1+\sqrt{2/3}\approx1.8165$ 時，點積 $\overrightarrow{AQ}\cdot\overrightarrow{BP}$ 的數值計算結果為 0（準確至 10^{-16}），當 $\lambda\approx0.1835$ 時亦然。
- 對 $\lambda=0.001,0.002,\ldots,5$ 計算直線 AQ 與 BP 之間的最短距離（並對 P 和 Q 的另一種標記方式作同樣計算），結果從不為零；只有當 $\lambda\to0$ 時，距離才趨向零。

常見錯誤

- 遺漏 $a+b=1$，$a+c=-1$ 這個情況：它雖然得不出任何解，但仍須加以排除。
- 寫到「三重積在 $\lambda = \frac{2}{3}$ 時等於零，所以兩線相交」便停筆。共面的直線可以是平行的，而這裏正是如此。

STEP 之外

純量三重積是一個 3×3 行列式（determinant），它等於零即表示三個向量線性相關（linearly dependent）；兩條歪斜線（skew lines）之間的最短距離 $|\overrightarrow{AB}\cdot(\mathbf{u}\times\mathbf{v})|/|\mathbf{u}\times\mathbf{v}|$，也是由同樣的組件構成的。在大學一年級的線性代數（linear algebra）課程中，這些概念會透過行列式、秩（rank）和子空間（subspace）的幾何統一起來。Strang 的 *Introduction to Linear Algebra*（線性代數入門）以大量三維圖像闡述這些內容。

詳細解答

STEP II 2002，第 8 題

純數學 • 一階微分方程

STEP II 2002，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/02-s2-q8

關鍵思路

每一條方程都是最簡單的一階線性方程（first-order linear equation）$y' = p(x)\,y$，所以在任何一個 p 由單一公式給出的區間上，解都是一個常數乘以 $\exp\left(\int p\,\mathrm{d}x\right)$。難處在於問題在 $x = 0$ 處被分成兩段：兩邊各有一個任意常數（arbitrary constant），其中一邊的常數由 y 的已知值確定，另一邊的常數則只能靠要求兩段在 $x = 0$ 處相接來確定。第一部分正是這個想法的預演，而第二部分則暗藏了一個絕對值（absolute value）。

解答

第一對方程

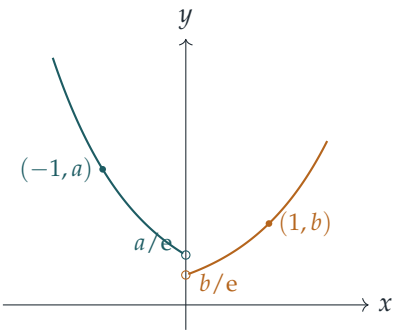
當 $x < 0$ 時，由 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = -y$ 得 $y = A\mathrm{e}^{-x}$，其中 A 為某個常數。（要看出沒有其他解，可留意 $\frac{\mathrm{d}}{\mathrm{d}x}\left(y\mathrm{e}^{x}\right) = \mathrm{e}^{x}(y' + y) = 0$，所以 $y\mathrm{e}^{x}$ 在區間 $x < 0$ 上是常數。）由條件「在 $x = -1$ 處 $y = a$」得 $A\mathrm{e} = a$，所以

$$y = a\,\mathrm{e}^{-x-1} \qquad (x < 0).$$

同樣地，當 $x > 0$ 時有 $y = B\mathrm{e}^{x}$，而在 $x = 1$ 處 $y = b$ 給出 $B = b/\mathrm{e}$，所以

$$y = b\,\mathrm{e}^{x-1} \qquad (x > 0).$$

當 $x \to 0^{-}$ 時，$y \to a/\mathrm{e}$；當 $x \to 0^{+}$ 時，$y \to b/\mathrm{e}$。一般而言這兩個值並不相同，曲線在 y 軸處會出現跳躍。下圖所示為 $a, b > 0$ 且 $a > b$ 的典型情況：左邊是一條衰減的指數曲線，愈接近該軸愈低；右邊是一條增長的指數曲線；空心圓圈標示兩個（不同的）極限（limit）值。



曲線沒有跳躍，當且僅當兩個極限相等，即 $a/\mathrm{e} = b/\mathrm{e}$，也就是

$$a = b.$$

在這情況下，曲線為 $y = a\,\mathrm{e}^{|x|-1}$，是一條對稱、形狀像「V」的曲線，在 $x = 0$ 處有一個尖角。

含 $|\mathrm{e}^x - 1|$ 的方程

由於 $\mathrm{e}^x > 1$ 當且僅當 $x > 0$，

$$|\mathrm{e}^x - 1| = \begin{cases} \mathrm{e}^x - 1, & x \geq 0, \\ 1 - \mathrm{e}^x, & x < 0. \end{cases}$$

所以這條方程和先前一樣，分成兩條普通的線性方程。

當 $x > 0$ 時： $y' = (\mathrm{e}^x - 1)y \circ \mathrm{e}^x - 1$ 的一個原函數（antiderivative）是 $\mathrm{e}^x - x$，所以 $\frac{\mathrm{d}}{\mathrm{d}x}\left(y\,\mathrm{e}^{-(\mathrm{e}^x - x)}\right) = 0$，於是

$$y = C\exp(\mathrm{e}^x - x).$$

在 $x = 1$ 處，$y = \mathrm{e}^{\mathrm{e}}$，所以 $C\,\mathrm{e}^{\mathrm{e}-1} = \mathrm{e}^{\mathrm{e}}$，得 $C = \mathrm{e}$。因此

$$y = \exp(\mathrm{e}^x - x + 1) \qquad (x \geq 0).$$

當 $x < 0$ 時： $y' = (1 - \mathrm{e}^x)y \circ 1 - \mathrm{e}^x$ 的一個原函數是 $x - \mathrm{e}^x$，所以

$$y = D\exp(x - \mathrm{e}^x).$$

想一想 在 $x < 0$ 的範圍內沒有已知的數據點。那麼 D 由甚麼確定？

答案是在 $x = 0$ 處的連續性（continuity）。從右邊趨近時，$y \to \exp(1 - 0 + 1) = \mathrm{e}^2$。從左邊趨近時，$y \to D\exp(0 - 1) = D/\mathrm{e}$。令兩者相等，得 $D = \mathrm{e}^3$，所以

$$y = \exp(x - \mathrm{e}^x + 3) \qquad (x < 0).$$

把兩段合起來，

$$y = \begin{cases} \exp(e^x - x + 1), & x \geq 0, \\ \exp(x - e^x + 3), & x < 0. \end{cases}$$

留意這一次在 $x = 0$ 處連尖角也沒有：兩條公式都給出 $y'(0) = 0$，與 $|e^0 - 1|\,y = 0$ 吻合。由於 $y > 0$ 且 $|e^x - 1| \geq 0$，這個解處處遞增，並在 $x = 0$ 處有一個駐點拐點（stationary point of inflection），該處 $y = e^2$。

極限

(i) 當 $x > 0$ 時，

$$y\exp(-e^x) = \exp(e^x - x + 1 - e^x) = e^{1-x} \to 0 \quad \text{當} x \to +\infty.$$

所以極限是 0：因子 $\exp(-e^x)$ 恰好把雙重指數增長完全抵消，只剩下衰減的 e^{1-x}。

(ii) 當 $x < 0$ 時，

$$y\,e^{-x} = \exp(x - e^x + 3 - x) = \exp(3 - e^x) \to e^3 \quad \text{當} x \to -\infty,$$

因為 $e^x \to 0$。所以極限是 e^3。換句話說，當 x 是絕對值很大的負數時，這個解的表現就像 e^3e^x，正如所料：在那裏 $|e^x - 1| \approx 1$，方程近乎 $y' = y$。

驗算

- 對答案求導：當 $x > 0$ 時，$y' = (e^x - 1)y$；當 $x < 0$ 時，$y' = (1 - e^x)y$。在 $x = 1$ 處，$\exp(e - 1 + 1) = e^e$。
- 數值上，用龍格－庫塔法（Runge–Kutta method）由 $y(1) = e^e \approx 15.154$ 逐步倒推至 $x = 0$，得 $y(0) \approx 7.38906 = e^2$；繼續推算至 $x = -5$，結果與 $\exp(x - e^x + 3)$ 吻合至六位有效數字（significant figures）。
- 就極限 (ii) 而言，在 $x = -10$ 處得 $ye^{-x} \approx 20.0846$，接近 $e^3 \approx 20.0855$；就 (i) 而言，在 $x = 10$ 處，$e^{1-x} \approx 1.2 \times 10^{-4}$。

常見錯誤

- 用 $x = 1$ 處的條件去確定 $x = 0$ 兩邊的常數。兩個區間是各自獨立的問題，左邊的常數只能由連續性得出。
- 丟掉模（modulus）符號，對所有 x 求解 $y' = (e^x - 1)y$；這樣在 $x < 0$ 時會得出錯誤的曲線，(ii) 的極限也會出錯。
- 寫出 $\int (e^x - 1)\,dx = e^x - x$ 之後，卻寫成 $\exp(e^x - x) = e^{e^x}/x$（正確的是 $e^{e^x}e^{-x}$）。

STEP 之外

方程 $y' = |e^x - 1|\, y$ 的係數在 $x = 0$ 處連續但不可微（differentiable）。關於 $y' = f(x, y)$ 的存在唯一性定理（existence and uniqueness theorem，即 Picard–Lindelöf 定理）只要求 f 對 x 連續、對 y 滿足利普希茨（Lipschitz）條件，所以在這裏單憑一個條件就能確定整條實數線上的解，而在 0 處的連續性是自動成立的。在第一部分，方程右邊本身在 $x = 0$ 處出現跳躍，所以才需要額外的條件把兩段接起來。Boyce 與 DiPrima 的 *Elementary Differential Equations*（微分方程入門教科書）討論了一階方程的這些存在性與唯一性問題。

詳細解答

STEP II 2002，第 9 題

力學 • 拋體運動

STEP II 2002，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：

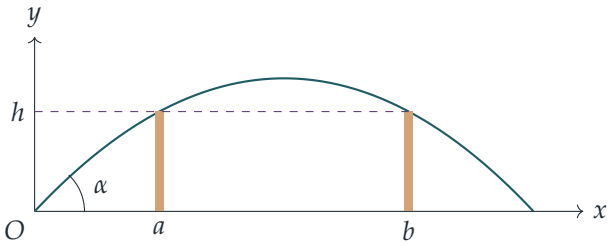
step.maths.org/questions/02-s2-q9

關鍵思路

先寫出軌跡（path）的方程，把 y 表示成 x 的函數。「剛好越過兩道牆」表示軌跡通過兩道牆的牆頂，所以方程 $y = h$ 的根恰好是 $x = a$ 和 $x = b$。這是一條已知兩根的二次方程（quadratic equation），把它與根之和及根之積（sum and product of roots）比較，便可一次過得出所有結果。至於第二部分，只需把兩條公式對 h 求導（differentiate）。

解答

我們把質點（particle）看成只受重力作用而運動的一點（不計空氣阻力（air resistance）），並取 O 為原點，x 軸水平，y 軸鉛垂向上。按圖中所示的情況，我們假設 $0 < a < b$、$h > 0$ 及 $0 < \alpha < \frac{\pi}{2}$。



求 $\tan\alpha$ 和 V

初速度的分量為 $V\cos\alpha$ 和 $V\sin\alpha$，經過時間 t 後

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

利用 $t = x/(V\cos\alpha)$ 消去時間，得軌跡

$$y = x\tan\alpha - \frac{gx^2}{2V^2\cos^2\alpha} = x\tan\alpha - kx^2, \qquad \text{其中} k = \frac{g\sec^2\alpha}{2V^2} = \frac{g(1+\tan^2\alpha)}{2V^2} > 0.$$

質點剛好越過每道牆，所以軌跡通過牆頂 (a,h) 和 (b,h)。因此 a 和 b 是以下方程的根：

$$x\tan\alpha - kx^2 = h, \qquad \text{即} \qquad kx^2 - x\tan\alpha + h = 0.$$

這是 x 的二次方程（因為 $k \neq 0$），所以它最多有兩個根，故此它的根恰好就是 a 和 b。與根之和及根之積比較，

$$a + b = \frac{\tan\alpha}{k}, \qquad ab = \frac{h}{k}.$$

由第二式得 $k = h/(ab)$，再由第一式得

$$\boxed{\tan\alpha = \frac{h(a+b)}{ab}.}$$

最後，由 k 的定義，

$$\begin{aligned}\frac{2V^2}{g} = \frac{1+\tan^2\alpha}{k} &= \frac{ab}{h}\left(1 + \frac{h^2(a+b)^2}{a^2b^2}\right)\\ &= \frac{ab}{h} + \frac{(a+b)^2 h}{ab},\end{aligned}$$

這正是所求的結果。

（值得留意的是，代入這些值後，軌跡就是 $y = \dfrac{h}{ab}x(a+b-x)$，一眼可見它在 $x = a$ 和 $x = b$ 時等於 h。）

把牆加高

第二個質點同樣剛好越過兩道牆，牆與原點的距離仍是 a 和 b，但高度變為 $h + \delta h$。所以把 h, α, V 換成 $h + \delta h, \alpha + \delta\alpha, V + \delta V$ 後，上面的公式依然成立：

$$\tan\alpha = \frac{a+b}{ab}h, \qquad \frac{2V^2}{g} = \frac{ab}{h} + \frac{(a+b)^2}{ab}h.$$

兩者都只是 h 的函數（a, b 固定），所以對於微小的變化，我們可以用 $\delta(\text{量}) \approx \dfrac{\mathrm{d}(\text{量})}{\mathrm{d}h}\delta h$。

角度。 把 $\tan\alpha = \frac{a+b}{ab}h$ 對 h 求導，得 $\sec^2\alpha\,\dfrac{\mathrm{d}\alpha}{\mathrm{d}h} = \dfrac{a+b}{ab}$，所以

$$\sec^2\alpha\,\delta\alpha \approx \frac{a+b}{ab}\delta h.$$

由於 $\sec^2\alpha > 0$、$a, b > 0$ 及 $\delta h > 0$，右邊是正數，所以 $\delta\alpha > 0$：質點必須以更陡的角度拋出。

想一想 這個「≈」只是一階（first-order）近似。我們真的可以憑它斷定 $\delta\alpha$ 的正負號嗎？

可以，而且其實完全不需要取近似。$\tan\alpha$ 的公式對 h 是線性的，所以以下等式準確成立：$\tan(\alpha+\delta\alpha)-\tan\alpha=\frac{a+b}{ab}\,\delta h>0$。由於 tan 在 $(0,\frac{\pi}{2})$ 上嚴格遞增（strictly increasing），這就迫使 $\alpha+\delta\alpha>\alpha$，所以 $\delta\alpha>0$。

速率。 把 $\dfrac{2V^2}{g}$ 對 h 求導，

$$\frac{4V}{g}\frac{\mathrm{d}V}{\mathrm{d}h}=-\frac{ab}{h^2}+\frac{(a+b)^2}{ab}=\frac{(a+b)^2}{abh^2}\left(h^2-\frac{a^2b^2}{(a+b)^2}\right),$$

所以

$$\delta V\approx\frac{g(a+b)^2}{4Vabh^2}\left(h-\frac{ab}{a+b}\right)\left(h+\frac{ab}{a+b}\right)\delta h.$$

除了 $h-\frac{ab}{a+b}$ 之外，每個因子都是正數（因為 $V,a,b,h,\delta h>0$）。因此 δV 的正負號與 $h-\frac{ab}{a+b}$ 相同：若 $h>ab/(a+b)$ 則為正，若 $h<ab/(a+b)$ 則為負，這正是所求的結果。（由於這個導數在 $h=ab/(a+b)$ 以外都不等於零，只要 δh 足夠小，一階項確實決定了正負號。）

詮釋。 與 $\tan\alpha=h(a+b)/(ab)$ 比較，條件 $h>ab/(a+b)$ 正好就是 $\tan\alpha>1$，即 $\alpha>45^\circ$。作為 h 的函數，$2V^2/g$ 具有 $p/h+qh$ 的形式，它在 $h=\sqrt{p/q}=ab/(a+b)$ 時最小，此時 $\alpha=45^\circ$ 而 $2V^2/g=2(a+b)$。所以：如果拋射角本來已經比 45° 陡，把牆加高便需要更大的速率；如果拋射角較平，把牆加高（同時令拋射角變陡，趨近 45°）所需的速率反而更小。

驗算

- 取 $a=3\,\mathrm{m}$、$b=7\,\mathrm{m}$、$h=2.5\,\mathrm{m}$。那麼 $\tan\alpha=25/21$，而 $2V^2/g=8.4+11.90\ldots=20.30\ldots\,\mathrm{m}$。用這些數值以數值方法逐步計算質點的運動，質點在 $x=3$ 和 $x=7$ 兩處的高度都是 $2.5\,\mathrm{m}$（準確至 10^{-11}）。
- 對稱性：答案對 a 和 b 是對稱的，這是必然的，因為軌跡是一條以 $x=\frac{1}{2}(a+b)$ 為對稱軸的拋物線（parabola）。
- 單位：ab/h 和 $(a+b)^2h/(ab)$ 都是長度，與 $2V^2/g$ 相符；而 $\tan\alpha$ 是無量綱（dimensionless）的。
- 當 $h=ab/(a+b)$ 時，我們得到 $\alpha=45^\circ$ 及 $2V^2/g=2(a+b)$；對於 $a=3$、$b=7$，這個值是 $20\,\mathrm{m}$，比上面的 $20.30\,\mathrm{m}$ 略小，與它是最小值相符。

常見錯誤

- 為兩道牆分別寫出兩條方程，然後硬解聯立方程：這樣做是可行的，但利用根之和與根之積，兩行便可完成。
- 把 k 換回 V 時，忘記了 $\sec^2\alpha = 1 + \tan^2\alpha$。
- 沒有把 $2V^2/g$ 的導數因式分解（factorise）便直接說出 δV 的正負號：那個簡潔的條件是由 $h^2 > a^2b^2/(a+b)^2$ 和 $h > 0$ 得出的。

STEP 之外

第二部分是一個小型的靈敏度分析（sensitivity analysis）：我們把映射 $h \mapsto (\alpha, V)$ 在某個解的附近線性化，然後讀出輸出如何隨之變化。線性化（linearisation），以及與之相關的軌跡族包絡線（envelope）的概念（即「安全拋物線」（parabola of safety），它是以某一給定速率所能到達的所有位置的邊界），都是大學一年級力學和微積分課程的標準內容。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）對拋體運動（projectile motion）和這些近似技巧都有細緻的論述。

詳細解答

STEP II 2002，第 10 題

力學 • 運動學與最優化

STEP II 2002，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/02-s2-q10

關鍵思路

總路程是固定的，所以完成比賽的時間 T 與轉換速度的時刻 t 由一條方程聯繫在一起。去分母後，它變成一條關於 t 的二次方程（quadratic），而某個時間 T 可以達到，當且僅當這條二次方程有實根。由判別式（discriminant）條件可直接得出 T 的最小值，完全不需要微積分。至於第二部分，一旦知道兩名跑手的位置如何隨時間變化，二人之間的距離便是一個分段二次函數（piecewise quadratic），我們用慣常的方法求其最大值即可。

解答

以下距離一律以 km 為單位，時間以小時為單位，並且 $42\frac{3}{8} = \frac{339}{8}$。

最短的比賽時間

第一名參賽者在最初 t 小時內跑了 $13t$，然後在餘下的 $T - t$ 小時內跑了 $\left(14 + \frac{2t}{T}\right)(T - t)$。因此

$$13t + \left(14 + \frac{2t}{T}\right)(T - t) = \frac{339}{8}.$$

展開後得 $13t + 14T - 14t + 2t - \dfrac{2t^2}{T} = \dfrac{339}{8}$，即

$$t + 14T - \frac{2t^2}{T} = \frac{339}{8}.$$

兩邊乘以 $T > 0$，並按 t 的冪次整理：

$$2t^2 - Tt + \left(\frac{339}{8}T - 14T^2\right) = 0. \qquad (*)$$

想一想 我們要求最小的 T，但 T 並不是以 t 的函數形式給出的。怎樣才能判斷哪些 T 值是有可能出現的呢？

某個 T 值能夠出現，當且僅當 $(*)$ 有實數解 t，所以它的判別式必須非負：

$$T^2 - 8\Big(\frac{339}{8}T - 14T^2\Big) \geq 0 \iff 113T^2 - 339T \geq 0 \iff 113T(T-3) \geq 0.$$

由於 $T > 0$，這就迫使 $T \geq 3$。

這個下界是可以達到的。當 $T = 3$ 時，判別式為零，$(*)$ 變成

$$2t^2 - 3t + \Big(\frac{1017}{8} - 126\Big) = 2t^2 - 3t + \frac{9}{8} = 2\Big(t - \frac{3}{4}\Big)^2 = 0,$$

所以 $t = \frac{3}{4}$。這個值介乎 0 與 T 之間，因此確實是一個可行的比賽方案：先以速率 13 跑 45 分鐘，然後在餘下的 $2\frac{1}{4}$ 小時以速率 $14 + \frac{2(3/4)}{3} = 14.5$ 跑完。事實上 $13 \cdot \frac{3}{4} + 14.5 \cdot \frac{9}{4} = 9.75 + 32.625 = 42.375$。因此 T 的最小可能值是 3，正如題目所要求。

（等價地，對 $(*)$ 作隱函數求導（implicit differentiation）並令 $\frac{\mathrm{d}T}{\mathrm{d}t} = 0$，可得 $T = 4t$，代回去便得 $T = 3$。判別式的論證有一個好處：它證明了 3 是全局最小值（global minimum），而不僅僅是一個駐值（stationary value）。）

最大的相隔距離

設 τ 為起跑後所經過的時間。第二名參賽者的速率是 $16 - k\tau$，其中 k 為某個常數，所以其所跑的距離是

$$s_2(\tau) = 16\tau - \tfrac{1}{2}k\tau^2.$$

由 3 小時跑畢全程得 $48 - \frac{9}{2}k = \frac{339}{8}$，所以 $\frac{9}{2}k = \frac{45}{8}$，$k = \frac{5}{4}$。於是 $s_2 = 16\tau - \frac{5}{8}\tau^2$，而速率由 16 下降至 12.25，一直保持為正。

第一名參賽者同樣用了 3 小時，根據第一部分，這表示 $t = \frac{3}{4}$。其所跑的距離是

$$s_1(\tau) = \begin{cases} 13\tau, & 0 \leq \tau \leq \frac{3}{4}, \\ \frac{39}{4} + \frac{29}{2}\big(\tau - \frac{3}{4}\big), & \frac{3}{4} \leq \tau \leq 3. \end{cases}$$

設 $d = s_2 - s_1$。第二名跑手起步較快，所以我們預期 $d \geq 0$。

當 $0 \leq \tau \leq \frac{3}{4}$：$d = 3\tau - \frac{5}{8}\tau^2$，而 $\frac{\mathrm{d}d}{\mathrm{d}\tau} = 3 - \frac{5}{4}\tau > 0$。所以 d 在這個區間上遞增，升至 $d\big(\frac{3}{4}\big) = \frac{9}{4} - \frac{45}{128} = \frac{243}{128}$。

當 $\frac{3}{4} \leq \tau \leq 3$：

$$d = 16\tau - \tfrac{5}{8}\tau^2 - \tfrac{39}{4} - \tfrac{29}{2}\tau + \tfrac{87}{8} = \tfrac{1}{8}\big(9 + 12\tau - 5\tau^2\big) = \tfrac{1}{8}(3-\tau)(3+5\tau).$$

這是一條開口向下的拋物線，頂點位於 $\tau = \frac{12}{10} = \frac{6}{5}$，落在 $\left[\frac{3}{4}, 3\right]$ 之內。（等價地，當兩人速率相等時，差距便停止擴大：$16 - \frac{5}{4}\tau = 14.5$。）在該處

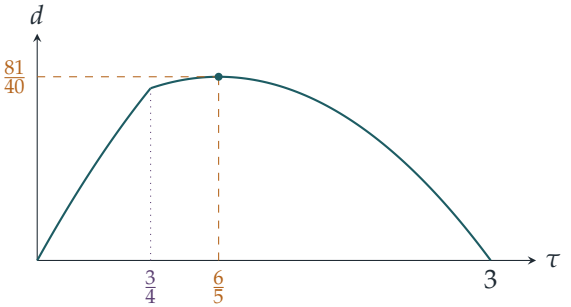
$$d\left(\tfrac{6}{5}\right) = \tfrac{1}{8} \cdot \tfrac{9}{5} \cdot 9 = \tfrac{81}{40}.$$

由因式分解後的形式可見，$d \geq 0$ 在整個區間上成立，並在終點降至 0，這是必然的。

比較兩段的結果，$\frac{81}{40} = 2.025 > \frac{243}{128} \approx 1.898$。所以兩名跑手之間的最大距離是

$$\boxed{\tfrac{81}{40}\ \text{km} = 2.025\ \text{km}},$$

在開賽後 1.2 小時（72 分鐘）達到，當時第二名參賽者領先。



驗算

- 兩名跑手都在 $\tau = 3$ 時衝線，而公式給出 $d(3) = 0$，這是必然的。
- 第二段的最大值恰好出現在兩人速率相等（同為 14.5）之時，這正是差距停止擴大的自然條件。
- 數值驗證：在一組細密的 t 值上解這條關於 T 的二次方程，得出最小的 T 為 3.000，出現在 $t = 0.75$；在 $[0,3]$ 內的 300 000 個時刻計算 d，得出最大值為 2.025，出現在 $\tau = 1.2$，並且 $d \geq 0$ 始終成立。

常見錯誤

- 直接套用第一名參賽者的最短時間方案，卻沒有留意到正是「跑步時間為 3 小時」這個條件確定了 $t = \frac{3}{4}$。
- 只在轉折點 $\tau = \frac{3}{4}$ 處求最大值：轉換速度之後差距仍會繼續擴大，因為直到 $\tau = \frac{6}{5}$ 之前，第二名跑手的速率仍然高於 14.5。
- 把第二名跑手所跑的距離寫成 $16\tau - k\tau^2$，忘記了對線性遞減的速率積分時所產生的 $\frac{1}{2}$。

STEP 之外

第一部分是在約束條件下求最優值（optimisation），而約束是由 $F(t,T)=0$ 以隱式給出的。提出「對哪些 T，解 t 才存在？」這個問題，正是隱函數定理（implicit function theorem）背後的想法，更一般地說，也是拉格朗日乘數（Lagrange multipliers）背後的想法：在最優點處，約束的等高線（level curve）滿足 $\partial F/\partial t=0$，而這正是本題中的重根條件。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學導論）全書以微積分處理運動學（kinematics），而 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題）則有更多 STEP 風格的最優化題目。

詳細解答

STEP II 2002，第 11 題

力學 • 摩擦力與極限平衡

STEP II 2002，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/02-s2-q11

關鍵思路

第一部分是直接的質心（centre of mass）積分；代換 $t = x/l$ 之後一行便可完成。至於梯子，「兩端同時處於極限狀態」令四個接觸力全部化為同一個未知量的倍數，因此分解兩次，再對 A 取力矩（moment），立即得出 $\tan\theta$。最後一部分還需要多一個想法：橫樑豎得愈陡便愈安全，所以「對每個 $\theta < \pi/4$ 都會滑動」即是說極限角至少為 $\pi/4$。

解答

質心

距離 B 為 x 處、長度為 δx 的一小段，其重量為 $\alpha W l^{-1}(x/l)^{\alpha-1}\,\delta x$。設 $t = x/l$（故 $\mathrm{d}x = l\,\mathrm{d}t$），總重量為

$$\int_0^l \frac{\alpha W}{l}\left(\frac{x}{l}\right)^{\alpha-1}\mathrm{d}x = \alpha W\int_0^1 t^{\alpha-1}\,\mathrm{d}t = W,$$

正如所料（由於 $\alpha > 0$，此積分收斂）。重量對 B 的力矩為

$$\int_0^l x\,\frac{\alpha W}{l}\left(\frac{x}{l}\right)^{\alpha-1}\mathrm{d}x = \alpha W l\int_0^1 t^{\alpha}\,\mathrm{d}t = \frac{\alpha W l}{\alpha+1}.$$

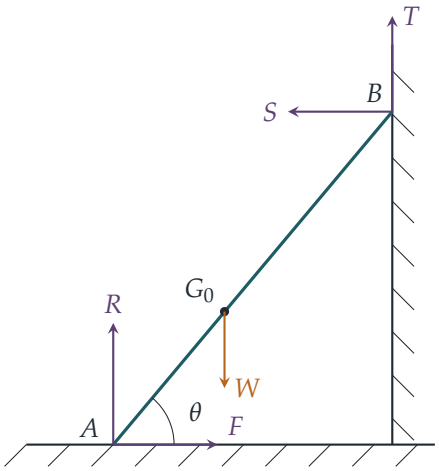
除以 W，可知質心 G_0 與 B 的距離為 $\alpha l/(\alpha+1)$，正是所求。因此它與 A 的距離為

$$d = l - \frac{\alpha l}{\alpha+1} = \frac{l}{\alpha+1}$$

稍後對 A 取力矩時，用的就是這個距離。

兩端同處極限平衡

把橫樑看成剛性桿（rigid rod），在 A 和 B 兩處均為粗糙接觸。若橫樑滑動，A 會遠離牆壁，B 會向下移動，所以摩擦力（friction）在 A 處指向牆壁，在 B 處向上。設地面在 A 處施加法向反作用力（normal reaction）R 和摩擦力 F，牆壁在 B 處施加法向反作用力 S 和摩擦力 T。重量作用於 G_0，與 A 相距 $l/(\alpha+1)$。



沿水平和垂直方向分解，並對 A 取力矩（重量的作用線與 A 的水平距離為 $d\cos\theta$；B 的高度為 $l\sin\theta$，水平距離為 $l\cos\theta$）：

$$F = S,$$
$$R + T = W,$$
$$S\,l\sin\theta + T\,l\cos\theta = W\,\frac{l}{\alpha+1}\cos\theta.$$

兩端同處極限狀態，即 $F=\mu R$ 及 $T=\mu S$。於是 $S=F=\mu R$，$T=\mu S=\mu^2 R$，而垂直方向的方程給出

$$R(1+\mu^2)=W, \qquad \text{所以} \qquad S=\frac{\mu W}{1+\mu^2}, \quad T=\frac{\mu^2 W}{1+\mu^2}.$$

把力矩方程除以 $l\cos\theta$ 再代入，

$$\frac{\mu W}{1+\mu^2}\tan\theta+\frac{\mu^2 W}{1+\mu^2}=\frac{W}{1+\alpha} \quad\Longrightarrow\quad \mu\tan\theta=\frac{1+\mu^2}{1+\alpha}-\mu^2=\frac{1-\alpha\mu^2}{1+\alpha},$$

即

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

正是所求。把這個角記作 θ_0，並設 $t_0=\tan\theta_0$。

μ 的最大值

想一想 為甚麼「對每個 $\theta < \pi/4$ 都會滑動」可以化為關於 θ_0 的條件？

我們證明：平衡（equilibrium）可能成立，當且僅當 $\tan\theta \geq t_0$。設 $t = \tan\theta$ 及 $k = 1/(1+\alpha)$，則力矩方程可寫成 $St + T = kW$。摩擦定律為 $F \leq \mu R$ 及 $|T| \leq \mu S$。

必要性（*necessity*）。由 $T \leq \mu S$ 及 $S = F \leq \mu R = \mu(W - T)$ 得 $T \leq \mu^2(W - T)$，所以 $T \leq \mu^2 W/(1+\mu^2)$。另外 $S \leq \mu(W - T)$，故

$$kW = St + T \leq \mu t(W - T) + T = \mu t W + (1 - \mu t)T.$$

若 $\mu t < 1$，利用 T 的上界可得 $kW \leq W(\mu t + \mu^2)/(1+\mu^2)$，整理後即 $t \geq t_0$。若 $\mu t \geq 1$，則無論如何都有 $t \geq 1/\mu > t_0$，因為 $\mu t_0 = (1 - \alpha\mu^2)/(1+\alpha) < 1$。兩種情況下，平衡都迫使 $t \geq t_0$。

充分性（*sufficiency*）。若 $t \geq t_0$（且 $t > 0$），取牆壁的摩擦力為極限值 $T = \mu S$，由力矩方程得 $S = kW/(t+\mu)$，並取 $F = S$，$R = W - T$。餘下唯一的條件是 $S \leq \mu R$，即 $S(1+\mu^2) \leq \mu W$，即 $k(1+\mu^2) \leq \mu(t+\mu)$，而這恰好就是 $t \geq t_0$。（此時亦有 $R > 0$。）

因此，橫樑在角 θ 時滑動，當且僅當 $\tan\theta < t_0$。「對每個 $\theta < \pi/4$ 都會滑動」即 $t_0 \geq 1$：若 $t_0 \geq 1$，每個 $\theta < \pi/4$ 都滿足 $\tan\theta < 1 \leq t_0$；若 $t_0 < 1$，任何滿足 $\max(t_0, 0) < \tan\theta < 1$ 的角都小於 $\pi/4$，卻處於平衡。

取 $\alpha = \frac{3}{2}$，

$$t_0 = \frac{1 - \frac{3}{2}\mu^2}{\frac{5}{2}\mu} = \frac{2 - 3\mu^2}{5\mu} \geq 1 \iff 3\mu^2 + 5\mu - 2 \leq 0 \iff (3\mu - 1)(\mu + 2) \leq 0,$$

其中兩邊乘以了 $5\mu > 0$。由於 $\mu + 2 > 0$，上式成立當且僅當 $\mu \leq \frac{1}{3}$。所以最大可能值為

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

驗算

- 當 $\alpha = 1$ 時橫樑是均勻的，公式變成 $\tan\theta = (1 - \mu^2)/(2\mu)$，正是兩端摩擦系數相同的均勻梯子的常見結果。
- 當 $\mu = \frac{1}{3}$、$\alpha = \frac{3}{2}$ 時：$t_0 = (2 - \frac{1}{3})/\frac{5}{3} = 1$，恰好等於 1，所以極限角為 $\pi/4$。數值上，$\mu = 0.332$ 給出 $t_0 \approx 1.004$，而 $\mu = 0.334$ 給出 $t_0 \approx 0.996$。
- 當 $\alpha \to \infty$ 時，質量集中於 B（$d \to 0$），對任何 $\mu > 0$，$\tan\theta$ 都變為負數：橫樑在任何角度都能豎立。這很合理，因為此時重量幾乎作用在牆壁上。

常見錯誤

- 在力矩方程中把 $\alpha l/(\alpha+1)$ 當作與 A 的距離；它其實是由 B 量起的。
- 把牆壁的摩擦力畫成向下：橫樑滑動時 B 向下滑，所以橫樑在 B 處所受的摩擦力向上。
- 把最後的不等式方向弄反：較陡的橫樑較安全，所以「在 $\pi/4$ 以下會滑動」需要 $\tan\theta_0 \geq 1$，而最大的 μ 令 θ_0 恰好等於 $\pi/4$。

STEP 之外

上面的必要性與充分性論證，其實是一個小型的線性規劃（linear programming）問題：摩擦定律界定了一個由可容許的力組成的凸區域（convex region），而當力矩直線與這個區域相交時，平衡便可能成立。像本題這類靜不定（statically indeterminate）問題，即單憑方程不足以確定各力的問題，是大學一年級力學和工程學的常見課題。Kleppner and Kolenkow, *An Introduction to Mechanics*（力學入門教材）正是以這種嚴謹的方式處理摩擦、力矩和剛體平衡。

詳細解答

STEP II 2002，第 12 題

概率與統計 • 正態近似與泊松近似

> **STEP II 2002，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/02-s2-q12

關鍵思路

這裏有三層隨機性，一層套一層：一枚硬幣之內的正面次數、一天之內的硬幣、一星期之內的日子。每一層都是一個二項計數，其「成功」概率來自下一層。公式的構造方法是：最底層用正態分佈（normal distribution）近似（並作連續性校正（continuity correction）），中間層先用泊松分佈（Poisson distribution）近似，再用正態分佈近似，而最頂層則保留為精確的二項分佈（binomial distribution）。一旦看出這三層結構，餘下的工作就是辨認公式中每個式子出自哪一個近似，然後再問：對於題目給定的數值，每個近似是否可靠。

解答

某一天的一枚硬幣

固定某一天和某一枚硬幣。設 X 為該硬幣 M 次投擲中正面的次數。各次投擲互相獨立，每次擲出正面的概率為 p，所以

$$X \sim \mathrm{B}(M, p), \qquad \mathrm{E}(X) = Mp, \qquad \mathrm{Var}(X) = Mp(1-p).$$

當 M 很大時，我們以正態變量 $X' \sim \mathrm{N}\big(Mp,\, Mp(1-p)\big)$ 近似 X。由於 X 取整數值，「正面少於 m 次」即 $X \le m-1$，作連續性校正後就變成 $X' < m - \frac{1}{2}$。所以

$$r = \mathrm{P}(X < m) = \mathrm{P}(X \le m-1) \approx \Phi\left(\frac{m - \frac{1}{2} - Mp}{\sqrt{Mp(1-p)}}\right) = \Phi\left(\frac{2m - 1 - 2Mp}{2\sqrt{Mp(1-p)}}\right).$$

與題目比較，這說明 $h \approx Lr$。

某一天的所有硬幣

設 Y 為 L 枚硬幣中，當天正面少於 m 次的硬幣數目。各枚硬幣的投擲互相獨立，而且每枚硬幣的表現相同，所以每枚硬幣各自獨立地以概率 r「成功」，於是

$$Y \sim \mathrm{B}(L, r), \qquad \mathrm{E}(Y) = Lr = h, \qquad \mathrm{Var}(Y) = Lr(1-r) = h(1-r).$$

> **想一想** q 的公式中，分母是 $\sqrt{h}$。為甚麼標準差（standard deviation）是 $\sqrt{h}$ 而不是 $\sqrt{h(1-r)}$？

這條公式是為 r 很小的情況而設的：對單獨一枚硬幣來說，正面少於 m 次是罕見事件，而 L 則很大。在這種情況下 $1-r \approx 1$，所以方差（variance）近似等於平均值（mean），這正是泊松近似 $Y \approx \mathrm{Po}(h)$ 適用的情形。如果 h 也相當大，$\mathrm{Po}(h)$ 又可以用 $\mathrm{N}(h, h)$ 近似。（等價的做法是：直接以正態分佈近似 $\mathrm{B}(L, r)$，再把 $1-r$ 換成 1。）

因此取 $Y' \sim \mathrm{N}(h, h)$。「多於 l 枚硬幣」即 $Y \geq l+1$，作連續性校正後就變成 $Y' > l + \frac{1}{2}$。於是

$$q = \mathrm{P}(Y > l) \approx 1 - \Phi\left(\frac{l + \frac{1}{2} - h}{\sqrt{h}}\right) = \Phi\left(\frac{h - l - \frac{1}{2}}{\sqrt{h}}\right) = \Phi\left(\frac{2h - 2l - 1}{2\sqrt{h}}\right),$$

當中用到標準正態分佈的對稱性 $1 - \Phi(z) = \Phi(-z)$。這就是題目給出的 q。

所有日子

如果某一天有多於 l 枚硬幣的正面次數少於 m，就稱這一天為**好日子**。每一天成為好日子的概率同樣是 q，而不同日子涉及不同的投擲，所以各天互相獨立。因此 K 天之中好日子的數目服從 $\mathrm{B}(K, q)$，而

$$\mathrm{P}(\text{恰好有 } k \text{ 個好日子}) = \binom{K}{k} q^k (1-q)^{K-k},$$

其中 q 按上文近似。這就是所要求的結果。

數值例子

這裏 $Mp = 60$，$Mp(1-p) = 24$，所以

$$\frac{2m - 1 - 2Mp}{2\sqrt{Mp(1-p)}} = \frac{96 - 1 - 120}{2\sqrt{24}} = \frac{-25}{9.798} \approx -2.55,$$

由此得 $r \approx \Phi(-2.55) \approx 0.0054$ 及 $h = 500r \approx 2.68$。然後

$$\frac{2h - 2l - 1}{2\sqrt{h}} \approx \frac{5.36 - 9}{3.27} \approx -1.11, \qquad q \approx \Phi(-1.11) \approx 0.133,$$

而近似公式給出 $\binom{7}{2} q^2 (1-q)^5 \approx 21 \times 0.0178 \times 0.490 \approx 0.18$。

想一想 每一個近似都是標準做法。當中哪些是在其可靠條件以外使用的？

逐層檢視這三層。

- **頂層**。$\mathrm{B}(7,q)$ 這一步是精確的（在 q 已知的前提下），所以它本身不會帶來誤差。但由於 $k=2$，它會把 q 的誤差平方。
- **底層**。$M=100$，$Mp=60$，$M(1-p)=40$，完全處於通常使用正態分佈近似二項分佈的範圍之內。但我們是在深入尾部 2.55 個標準差的位置使用它，用來求一個約為 0.005 的概率。就絕對誤差而言，正態近似在分佈的主體部分表現良好；但在尾部，它的相對誤差可以很大，尤其因為 $p\neq\frac{1}{2}$ 令二項分佈偏斜。r 的任何相對誤差都會直接傳到 $h=Lr$，而 q 對 h 十分敏感。
- **中間層**。以 $\mathrm{Po}(h)$ 取代 $\mathrm{B}(500,r)$ 效果極佳，因為 r 極小而 L 很大。但接着在 $h\approx 2.7$ 的情況下以正態分佈近似 $\mathrm{Po}(h)$，卻是沒有根據的：平均值小於 3 的泊松分佈明顯偏斜，而常用的經驗法則要求平均值至少約為 10（有人說 15 或 20）。

所以答案是否定的：我們不會預期這個近似準確。有疑問的步驟是對 $\mathrm{B}(100,0.6)$ 尾部的正態近似，以及對平均值很小的泊松分佈的正態近似。

事實上，r 的精確值是 $\mathrm{P}(X\leq 47)\approx 0.00576$，比正態估計值高出約 7%，所以 h 應約為 2.88；而精確的 $\mathrm{P}(Y>4)$ 則約為 0.165，而不是 0.133。最終的概率約為 0.23，而不是 0.18。（就這組特定的數值而言，對 $\mathrm{Po}(2.68)$ 的正態近似所給出的尾部概率，碰巧與泊松分佈本身幾乎相同，即 0.133 對 0.134，所以大部分的偏差來自底層。但你事前不可能指望有這種運氣。）

驗算

- 連續性校正的方向正確：「少於 m」即 $X\leq m-1$，所以我們在 $m-\frac{1}{2}$ 處截止；「多於 l」即 $Y\geq l+1$，所以我們由 $l+\frac{1}{2}$ 開始。題目給出的公式中正好含有 $2m-1$ 和 $2l+1$。
- 數值上，$\Phi(-2.5516)=0.005362$，$h=2.681$，$q=\Phi(-1.1108)=0.1333$，由此得 0.1825。精確的二項求和給出 $r=0.005761$，$q=0.1645$（已由模擬 2×10^4 天加以確認，所得結果為 0.164），最終答案為 0.2314。

常見錯誤

- 用了 m 而不是 $m-\frac{1}{2}$（或用了 l 而不是 $l+\frac{1}{2}$）：這樣公式就對不上。
- 用了 $\sqrt{h(1-r)}$，然後無法得出 $\sqrt{h}$。題目的公式略去了因子 $1-r$，因為它實際上是

在使用泊松近似，而你應該指出這一點。

- 只憑「$Mp > 5$ 及 $M(1-p) > 5$」來判斷準確度。這條法則針對的是分佈的中心部分，而不是一個 0.005 的尾部概率。

STEP 之外

這道題目所關乎的，是誤差如何在一連串近似之中傳播，以及尾部的絕對誤差與相對誤差之間的分別。在大學裏，中心極限定理（central limit theorem）會由 Berry–Esseen 定理進一步加強，該定理給出正態近似的絕對誤差的上界。尾部概率則需要另一些工具，例如切爾諾夫界（Chernoff bounds）和大偏差理論（large-deviation theory），它們描述尾部概率縮小得有多快。Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門）清楚地講解了泊松近似、正態近似以及切爾諾夫型不等式；Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論導引）則是一本很好的輔助讀物。

詳細解答

STEP II 2002，第 13 題

概率與統計 • 變換後的分佈函數

> **STEP II 2002，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/02-s2-q13

關鍵思路

這一題雖然放在力學部分，其實是一道關於變換分佈函數（distribution function）的概率題。一切都建基於一條公式：對 G 求導得 $G'(y) = f(y) \cdot \dfrac{2}{(2-F(y))^2}$，其中 $f = F'$ 是 X 的概率密度函數（density）。因此，Y 的密度就是 X 的密度乘以一個權重，而這個權重被夾在 $\frac{1}{2}$ 與 2 之間。有了這一點，Y 的每一個期望值（expectation）都是 X 的密度乘以一個在 $[\frac{1}{2}, 2]$ 內的權重的積分，只要比較被積函數（integrand），便可得出各個不等式。條件 $a > 0$ 的作用，是確保被積函數非負。

解答

G 的性質

我們按題目的原意，假設 X 在 $[a, b]$ 上有密度 $f(y) = F'(y) \geq 0$，因此 X 的全部概率都落在 $[a, b]$ 內。

由於 $F(a) = 0$ 及 $F(b) = 1$，

$$G(a) = \frac{0}{2-0} = 0, \qquad G(b) = \frac{1}{2-1} = 1.$$

由商法則（quotient rule），

$$G'(y) = \frac{F'(y)\big(2 - F(y)\big) - F(y)\big(-F'(y)\big)}{(2-F(y))^2} = \frac{2F'(y)}{(2-F(y))^2}.$$

分子非負，因為分佈函數是非遞減的，所以 $F'(y) \geq 0$。分母是 $2 - F(y) \geq 1$ 的平方，故為正數。因此 $G'(y) \geq 0$。

至於那個界，由於 F 是分佈函數，故 $0 \le F(y) \le 1$。於是

$$1 \le 2 - F(y) \le 2 \quad \Longrightarrow \quad 1 \le (2 - F(y))^2 \le 4 \quad \Longrightarrow \quad \tfrac{1}{2} \le \frac{2}{(2 - F(y))^2} \le 2,$$

其中各項皆為正數，所以取平方後不等號方向不變；取倒數會令不等號反向，再乘以 2 則只是把各項按比例放大。這就是所求的雙重不等式。

綜合而言，$G(a) = 0$、$G(b) = 1$ 以及 G 非遞減，說明 G 確實是 $[a, b]$ 上的一個分佈函數，所以隨機變量（random variable）Y 是有意義的。它的密度是

$$g(y) = G'(y) = w(y)\, f(y), \qquad \text{其中} w(y) = \frac{2}{(2 - F(y))^2} \in \left[\tfrac{1}{2}, 2\right].$$

$\mathrm{E}(Y)$ 的界

想一想 為甚麼我們可以把不等式 $\frac{1}{2} \le w(y) \le 2$ 乘以 $yf(y)$，而不等號方向保持不變？

因為在 $[a, b]$ 上 $yf(y) \ge 0$：我們有 $f(y) \ge 0$，而且 $y \ge a > 0$。假設 $a > 0$ 正是用在這裏。因此

$$\tfrac{1}{2}\, yf(y) \le y\, w(y) f(y) \le 2\, yf(y) \qquad \text{對於} a \le y \le b,$$

而在 $[a, b]$ 上積分後，不等式依然成立：

$$\tfrac{1}{2} \int_a^b yf(y)\, \mathrm{d}y \;\le\; \int_a^b y\, g(y)\, \mathrm{d}y \;\le\; 2 \int_a^b yf(y)\, \mathrm{d}y,$$

即

$$\tfrac{1}{2}\, \mathrm{E}(X) \le \mathrm{E}(Y) \le 2\, \mathrm{E}(X),$$

這正是所要證明的。

$\mathrm{Var}(Y)$ 的界

同樣地，$y^2 f(y) \ge 0$（這只需要 $f \ge 0$），所以把 $w(y) \le 2$ 乘以它再積分，得

$$\mathrm{E}(Y^2) = \int_a^b y^2\, w(y) f(y)\, \mathrm{d}y \le 2 \int_a^b y^2 f(y)\, \mathrm{d}y = 2\, \mathrm{E}(X^2).$$

對於被減去的那一項，我們需要的是 $(\mathrm{E}(Y))^2$ 的下界。由上一部分，$\mathrm{E}(Y) \ge \frac{1}{2}\mathrm{E}(X)$，而且因為 $X \ge a > 0$，所以 $\mathrm{E}(X) > 0$。兩邊都是正數，故可以放心取平方：

$$(\mathrm{E}(Y))^2 \ge \tfrac{1}{4}(\mathrm{E}(X))^2.$$

因此

$$\begin{aligned}\operatorname{Var}(Y) &= \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 \\ &\le 2\,\mathrm{E}(X^2) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2\big(\operatorname{Var}(X) + (\mathrm{E}(X))^2\big) - \tfrac{1}{4}(\mathrm{E}(X))^2 \\ &= 2\operatorname{Var}(X) + \tfrac{7}{4}(\mathrm{E}(X))^2,\end{aligned}$$

其中第三行用了 $\mathrm{E}(X^2) = \operatorname{Var}(X) + (\mathrm{E}(X))^2$。這就是所求的界。

驗算

取 X 為 $[1,3]$ 上的均勻分佈（uniform distribution），則 $F(y) = \frac{1}{2}(y-1)$，且 $\mathrm{E}(X) = 2$。由於 $Y \ge 1$，故 $\mathrm{E}(Y) = 1 + \int_1^3 (1 - G(y))\,\mathrm{d}y$；代入 $u = F(y)$ 得

$$\mathrm{E}(Y) = 1 + 2\int_0^1 \frac{2(1-u)}{2-u}\,\mathrm{d}u = 1 + 4(1 - \ln 2) \approx 2.227,$$

穩穩地落在 1 與 4 之間。直觀地說，w 在 F 小的地方較小，在 F 接近 1 的地方較大，所以 Y 偏向區間的上端，在這個例子中 $\mathrm{E}(Y) > \mathrm{E}(X)$。我們亦以數值方法驗證過：對於隨機區間上的 300 個隨機冪律分佈（power-law distribution），$\frac{1}{2}\mathrm{E}(X) \le \mathrm{E}(Y) \le 2\mathrm{E}(X)$ 與方差（variance）的界都成立；觀察所得的比值 $\mathrm{E}(Y)/\mathrm{E}(X)$ 大約介乎 0.92 與 1.31 之間。

常見錯誤

- 忘記了 $a > 0$ 為何重要：假如 X 可以取負值，$yf(y)$ 便可能是負數，乘以它會令不等號反向。
- 用 $\mathrm{E}(Y) \le 2\mathrm{E}(X)$ 來估計 $\operatorname{Var}(Y)$ 的界：$(\mathrm{E}(Y))^2$ 這一項是被減去的，所以你需要的是它的下界，而不是上界。
- 試圖把 $(y-\mu)^2$ 對 g 積分，直接比較 $\operatorname{Var}(Y)$ 與 $\operatorname{Var}(X)$：兩個方差所用的平均值並不相同，所以應改用 $\mathrm{E}(Y^2)$ 來處理。

STEP 之外

寫成 $g = wf$，是測度變換（change of measure）的一個簡單例子：新的分佈相對於舊的分佈有一個密度（稱為似然比（likelihood ratio），或 Radon–Nikodym 導數），而這個比值的界可以直接轉化為期望值的界。統計學中的重要性抽樣（importance sampling），以及金融數學中所用的概率測度變換，都以這個想法為基礎。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）講解分佈函數、密度與期望值，程度正好適合由此再深入探究。

詳細解答

STEP II 2002，第 14 題

概率與統計 • 混合分佈與貝葉斯定理

STEP II 2002，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/02-s2-q14

關鍵思路

這是一個分兩階段進行的試驗：先選出一個區域，然後該區域內發生意外。因此應以區域為條件來考慮。全期望定律（law of total expectation）給出平均值，全概率定律（law of total probability）給出分佈，最後由貝葉斯定理（Bayes' theorem）把問題反過來問。題目只給出平均值，所以我們以泊松分佈（Poisson distribution）作為某區域內意外宗數的模型；而要證明的公式帶有 $\mathrm{e}^{-\lambda}\lambda^k/k!$ 的形式，也印證了這個模型。

解答

設定

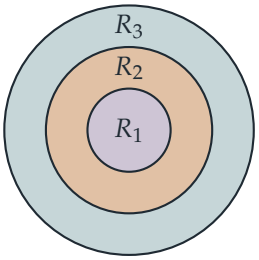
設 S 為所選區域的編號，$\lambda_n = 2 - \frac{n}{N}$ 為 R_n 內意外宗數的平均值。區域 R_n 是一個圓環（annulus），內外半徑分別為 $n-1$ 和 n，所以它的面積是

$$\pi n^2 - \pi(n-1)^2 = \pi(2n-1).$$

整個島的面積是 πN^2。（驗算一下：$\sum_{n=1}^{N}(2n-1) = N^2$。）因此

$$\mathrm{P}(S=n) = \frac{2n-1}{N^2}, \qquad n = 1, \ldots, N.$$

在 $S = n$ 的條件下，我們把 X 看成平均值為 λ_n 的泊松變量。由於對每個 n 都有 $\lambda_n \geq 1 > 0$，這些泊松分佈全都有意義。



區域面積：$R_n = \pi(2n-1)$

長期平均值

根據大數定律（law of large numbers），長遠而言，每天記錄到的平均宗數就是 $\mathrm{E}(X)$。以區域為條件，可得

$$\begin{aligned}\mathrm{E}(X) &= \sum_{n=1}^{N} \mathrm{P}(S=n)\,\mathrm{E}(X \mid S=n) = \sum_{n=1}^{N} \frac{2n-1}{N^2}\Big(2-\frac{n}{N}\Big)\\ &= 2\sum_{n=1}^{N}\frac{2n-1}{N^2} \;-\; \frac{1}{N^3}\sum_{n=1}^{N}(2n^2-n)\\ &= 2-\frac{1}{N^3}\Big(\frac{N(N+1)(2N+1)}{3}-\frac{N(N+1)}{2}\Big).\end{aligned}$$

第一個和等於 1，因為各區域被選中的概率加起來是 1。括號內我們用了題目給出的 $\sum n^2$ 公式，以及 $\sum n = \frac{1}{2}N(N+1)$。抽出 $\frac{1}{6}N(N+1)$ 後，餘下 $2(2N+1)-3 = 4N-1$，所以

$$\mathrm{E}(X) = 2-\frac{(N+1)(4N-1)}{6N^2} = 2-\frac{1}{6}\Big(1+\frac{1}{N}\Big)\Big(4-\frac{1}{N}\Big),$$

其中把 N^2 拆成 $N\cdot N$，分給兩個因子。這就是所要求的結果。

X 的分佈

根據全概率定律，對 $k = 0, 1, 2, \ldots$，

$$\mathrm{P}(X=k) = \sum_{n=1}^{N}\mathrm{P}(S=n)\,\mathrm{P}(X=k\mid S=n) = \sum_{n=1}^{N}\frac{2n-1}{N^2}\cdot\frac{\mathrm{e}^{-\lambda_n}\lambda_n^{k}}{k!}.$$

現在把每一部分用 N 表示：

$$\lambda_n^{k} = \Big(\frac{2N-n}{N}\Big)^{k} = N^{-k}(2N-n)^{k}, \qquad \mathrm{e}^{-\lambda_n} = \mathrm{e}^{-2+n/N} = \mathrm{e}^{-2}\,\mathrm{e}^{n/N}.$$

因子 e^{-2}、N^{-2}、N^{-k} 和 $1/k!$ 都與 n 無關，所以可以提到求和符號之外：

$$\mathrm{P}(X=k) = \frac{\mathrm{e}^{-2}N^{-k-2}}{k!}\sum_{n=1}^{N}(2n-1)(2N-n)^{k}\,\mathrm{e}^{n/N},$$

正如所求。

$N = 3$、兩宗意外的情況

我們要求 $\mathrm{P}(S = 2 \mid X = 2)$。根據貝葉斯定理，

$$\mathrm{P}(S = 2 \mid X = 2) = \frac{\mathrm{P}(S = 2)\,\mathrm{P}(X = 2 \mid S = 2)}{\mathrm{P}(X = 2)}.$$

想一想 我們需要把 $\mathrm{P}(X = 2)$ 完整地計算出來嗎？

不需要。分子正是 $\mathrm{P}(X = 2)$ 的求和式中 $n = 2$ 的那一項。因此公因子 $\mathrm{e}^{-2}N^{-k-2}/k!$ 可以約去，所求概率就是 $\sum(2n-1)(2N-n)^k\mathrm{e}^{n/N}$ 中 $n = 2$ 的一項除以整個和。

取 $N = 3$ 和 $k = 2$，各項 $(2n-1)(6-n)^2\mathrm{e}^{n/3}$ 分別是

$$n = 1:\ 1 \cdot 25\,\mathrm{e}^{1/3}, \qquad n = 2:\ 3 \cdot 16\,\mathrm{e}^{2/3} = 48\,\mathrm{e}^{2/3}, \qquad n = 3:\ 5 \cdot 9\,\mathrm{e} = 45\,\mathrm{e}.$$

於是，把分子和分母同除以 $\mathrm{e}^{2/3}$，得

$$\mathrm{P}(S = 2 \mid X = 2) = \frac{48\,\mathrm{e}^{2/3}}{25\,\mathrm{e}^{1/3} + 48\,\mathrm{e}^{2/3} + 45\,\mathrm{e}} = \frac{48}{48 + 45\,\mathrm{e}^{1/3} + 25\,\mathrm{e}^{-1/3}},$$

這就是所要求的結果。

驗算

- 當 $N = 1$ 時只有一個區域，平均值為 1，而公式給出 $2 - \frac{1}{6} \cdot 2 \cdot 3 = 1$。當 $N \to \infty$ 時，平均值趨向 $2 - \frac{4}{6} = \frac{4}{3}$。這與 $\int_0^1 2x(2-x)\,\mathrm{d}x = \frac{4}{3}$ 吻合，其中均勻選取的點位於相對半徑 x 處的密度為 $2x$。
- 當 $N = 3$ 時，平均值是 $\frac{32}{27} \approx 1.185$。模擬 4×10^5 天（在圓盤內均勻地選一點，然後抽取一個泊松計數），得出的平均值是 1.182。模擬同時得出 $\mathrm{P}(X = 2) \approx 0.209$（準確值為 0.2094），而在發生兩宗意外的日子當中，落在 R_2 的比例為 0.372。最後答案的數值是 0.3729。
- 把分佈公式對 k 求和，在我們試過的每個 N 之下，總概率都是 1。

常見錯誤

- 以相等的概率 $1/N$ 選取區域，而不是按面積 $2n-1$ 的比例選取。
- 把各個泊松分佈的平均值取平均，然後聲稱 X 是以該值為平均值的泊松變量。泊松分佈的混合並不是泊松分佈。只有期望值才能這樣簡單地通過取平均而得出。

STEP 之外

變量 X 服從一個混合分佈（mixture distribution）：區域充當隱變量（hidden variable），而最後一部分就是對它作貝葉斯推斷（Bayesian inference）。上面第二個錯誤可以用全方差定律（law of total variance）來解釋：$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X \mid S)) + \mathrm{Var}(\mathrm{E}(X \mid S))$，它說明混合分佈總比平均值相同的單一泊松分佈更為分散。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門）對條件期望（conditional expectation）和這類混合分佈有很好的講解。

詳細解答

STEP III 2002，第 1 題

純數學 • 廣義積分與體積

STEP III 2002，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：

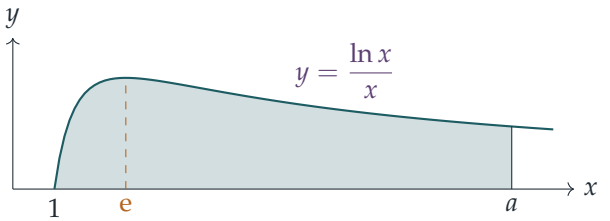
step.maths.org/questions/02-s3-q1

關鍵思路

只要留意到 $\frac{1}{x}$ 是 $\ln x$ 的導數（derivative），兩個積分都變得容易。面積的被積函數（integrand）是「一個函數乘以它自己的導數」，所以一眼就能積出來。體積的被積函數 $(\ln x)^2/x^2$ 多出了一個 x 的冪，因此我們要用兩次分部積分法（integration by parts），每次都對對數的冪求導，把它的次數降低；這裏所說的對數就是 $\ln x$。有趣之處在於兩個極限（limit）的對比：曲線衰減得剛好夠慢，令面積無止境地增長；但把它平方之後，衰減就夠快，令體積穩定下來。

解答

當 $x \geq 1$ 時，$\ln x \geq 0$，所以曲線位於 x 軸上或其上方，面積就是該積分本身。曲線由 $(1,0)$ 出發，上升至 $x = \mathrm{e}$ 處的極大值（該處 $y' = (1 - \ln x)/x^2 = 0$），然後緩緩地向橫軸衰減。



面積

由於 $\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{2}(\ln x)^2\right) = \dfrac{\ln x}{x}$，

$$A(a) = \int_1^a \frac{\ln x}{x}\,\mathrm{d}x = \left[\tfrac{1}{2}(\ln x)^2\right]_1^a = \tfrac{1}{2}(\ln a)^2,$$

因為 $\ln 1 = 0$。（等價的做法是用代換法（substitution）：設 $u = \ln x$，$\mathrm{d}u = \mathrm{d}x/x$，便得到 $\int_0^{\ln a} u\,\mathrm{d}u$。）

當 $a \to \infty$ 時，$\ln a \to \infty$，所以 $A(a) \to \infty$：面積是無界的，儘管它增長得非常緩慢。

體積

繞 x 軸旋轉，所得的體積為 $V(a) = \pi\displaystyle\int_1^a y^2\,\mathrm{d}x = \pi\int_1^a \frac{(\ln x)^2}{x^2}\,\mathrm{d}x$。

用分部積分法，取 $u = (\ln x)^2$ 及 $\mathrm{d}v = x^{-2}\,\mathrm{d}x$，於是 $\mathrm{d}u = \frac{2\ln x}{x}\,\mathrm{d}x$ 及 $v = -\frac{1}{x}$：

$$\int \frac{(\ln x)^2}{x^2}\,\mathrm{d}x = -\frac{(\ln x)^2}{x} + 2\int \frac{\ln x}{x^2}\,\mathrm{d}x.$$

再做一次同樣的步驟，取 $u = \ln x$，$\mathrm{d}v = x^{-2}\,\mathrm{d}x$：

$$\int \frac{\ln x}{x^2}\,\mathrm{d}x = -\frac{\ln x}{x} + \int \frac{1}{x^2}\,\mathrm{d}x = -\frac{\ln x}{x} - \frac{1}{x}.$$

把兩者合併，

$$\int \frac{(\ln x)^2}{x^2}\,\mathrm{d}x = -\frac{(\ln x)^2 + 2\ln x + 2}{x} + C.$$

在 $x = 1$ 處，括號內的值是 $0 + 0 + 2$，所以下限的貢獻是 $-(-2) = +2$，於是

$$V(a) = \pi\left(2 - \frac{(\ln a)^2 + 2\ln a + 2}{a}\right).$$

想一想 $(\ln a)^2/a$ 的分子和分母都趨向無窮大，我們怎樣知道它趨向 0？

對數的增長比任何正數次冪都要慢。要把這句話說得嚴謹，可由以下事實出發：對所有 $t > 0$ 都有 $\ln t < t$（直線 $y = t$ 位於 $y = \ln t$ 的上方；兩者的差距 $t - \ln t$ 在 $t = 1$ 處取得最小值 1）。代入 $t = a^{1/4}$：則 $\frac{1}{4}\ln a < a^{1/4}$，即 $\ln a < 4a^{1/4}$。所以當 $a > 1$ 時，

$$0 < \frac{(\ln a)^2}{a} < \frac{16\,a^{1/2}}{a} = \frac{16}{\sqrt{a}} \to 0, \qquad 0 < \frac{\ln a}{a} < \frac{4}{a^{3/4}} \to 0,$$

而 $2/a \to 0$ 自然不在話下。因此 $V(a)$ 中的分式趨向 0，而

$$V(a) \to 2\pi \quad \text{當} a \to \infty.$$

由於被積函數為正，$V(a)$ 隨 a 遞增，所以體積穩步上升並趨近 2π，但永遠不會達到這個值：即使被旋轉的區域面積無限，所得的立體卻只有有限的體積。

驗算

- 在 $a = 1$ 時，兩條公式都給出 0，這是必然的：$A(1) = 0$ 及 $V(1) = \pi(2 - 2) = 0$。
- 對 $-\big((\ln x)^2 + 2\ln x + 2\big)/x$ 求導，可還原出 $(\ln x)^2/x^2$（$2\ln x$ 和 2 這兩項與括號的導數互相抵消）。
- 數值上，對 $\pi \int_1^{10} (\ln x)^2/x^2 \,\mathrm{d}x$ 使用辛普森法則（Simpson's rule）得出 2.54247，與公式吻合；而設 $x = \mathrm{e}^t$ 後，無限延伸時的體積是 $\pi \int_0^{\infty} t^2 \mathrm{e}^{-t} \,\mathrm{d}t$，其數值計算結果為 $6.283185\ldots = 2\pi$。

常見錯誤

- 忘記下限：$\ln 1 = 0$ 令大部分的項在 $x = 1$ 處消失，但常數項 $-2/x$ 卻不會消失，而整個極限值 2π 正是由它提供的。
- 毫無理由地斷言 $(\ln a)^2/a \to 0$。只需寫一行比較，例如 $\ln a < 4a^{1/4}$，便已足夠。

STEP 之外

這是「加百列號角」(Gabriel's horn，即把 $y = 1/x$ 旋轉而成) 的近親，後者是一個體積有限但表面積無限的立體：而在這裏，連平面區域本身的面積也是無限的。代換 $x = \mathrm{e}^t$ 把體積化為 $\pi \int_0^{\infty} t^2 \mathrm{e}^{-t} \,\mathrm{d}t = \pi\,\Gamma(3) = 2!\,\pi$，讓我們初次窺見伽瑪函數（gamma function），$\Gamma(n+1) = n!$。怎樣理解瑕積分（improper integral），以及「對數輸給冪」這類比較，是大學一年級分析學（analysis）的核心課題；Abbott 的 *Understanding Analysis* 是閱讀這方面內容的好去處。

詳細解答

STEP III 2002，第 2 題

純數學 • 反正切級數

STEP III 2002，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/02-s3-q2

關鍵思路

第一個結果其實是喬裝了的正切加法公式（tangent addition formula）。真正要下工夫的，只是驗證等式兩邊的角都落在同一個區間內，而 tan 在該區間上是一對一（one-to-one）的。有了這條公式，歸納法（induction）那一步只需套用它一次，再做一些因式分解（factorising）。至於最後一個和，竅門在於留意到 $r^2 - r + 1$ 不過是把 $r^2 + r + 1$ 中的 r 平移一位。因此新的級數（series）就是原來的級數，在前面多加一項而已。

解答

加法公式

設 $A = \arctan a$ 及 $B = \arctan b$。則 $\tan A = a$ 及 $\tan B = b$，並且

$$\tan(A + B) = \frac{\tan A + \tan B}{1 - \tan A \tan B} = \frac{a + b}{1 - ab},$$

由於 $ab < 1$，此式有意義。這說明 $A + B$ 與 $\arctan\left(\frac{a+b}{1-ab}\right)$ 的正切值相同。但單憑這一點並不足以令兩者相等，因為 tan 每隔 π 便重複一次。所以我們需要兩個角都落在同一個區間內，而 tan 在該區間上是一對一的。

? **想一想** 為甚麼得到 $\tan(A + B) = \frac{a+b}{1-ab}$ 之後，證明還未完結？

證明還未完結，是因為 arctan 只會取 $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ 內的值。我們必須驗證 $A + B$ 也落在這個區間內。

- 由於 $0 < a < 1$，而 arctan 是遞增函數，故 $0 < A < \frac{\pi}{4}$。同理 $0 < B < \frac{\pi}{4}$，所以 $0 < A + B < \frac{\pi}{2}$。

► 由於 $a+b>0$ 及 $1-ab>0$，分數 $\frac{a+b}{1-ab}$ 為正數。所以它的 arctan 也落在 $(0,\frac{\pi}{2})$ 內。

函數 tan 在 $(0,\frac{\pi}{2})$ 上嚴格遞增（strictly increasing），因而是一對一的。所以該區間內正切值相同的兩個角必定相等，由此得

$$\arctan a+\arctan b=\arctan\left(\frac{a+b}{1-ab}\right). \qquad \checkmark$$

歸納法

記 $S_n=\sum_{r=1}^{n}\arctan\left(\frac{1}{r^2+r+1}\right)$。要證明的是 $S_n=\arctan\left(\frac{n}{n+2}\right)$。

基礎情況（*base case*）。當 $n=1$ 時，兩邊均為 $\arctan\frac{1}{3}$。

歸納步驟（*inductive step*）。假設對某個 $n\geq 1$，有 $S_n=\arctan\left(\frac{n}{n+2}\right)$。下一項中有

$$(n+1)^2+(n+1)+1=n^2+3n+3,$$

所以

$$S_{n+1}=\arctan\left(\frac{n}{n+2}\right)+\arctan\left(\frac{1}{n^2+3n+3}\right).$$

兩個自變量都嚴格介乎 0 與 1 之間，所以加法公式適用，其中 $a=\frac{n}{n+2}$ 及 $b=\frac{1}{n^2+3n+3}$。把 $\frac{a+b}{1-ab}$ 的分子和分母同乘以 $(n+2)(n^2+3n+3)$，得

$$\begin{aligned}\frac{a+b}{1-ab}&=\frac{n(n^2+3n+3)+(n+2)}{(n+2)(n^2+3n+3)-n}\\&=\frac{n^3+3n^2+4n+2}{n^3+5n^2+8n+6}.\end{aligned}$$

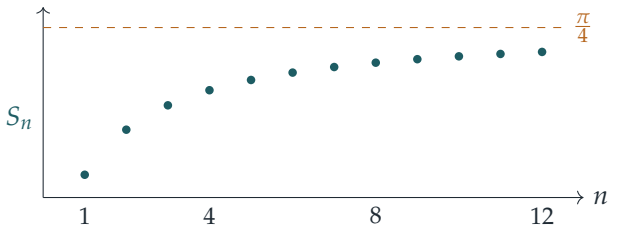
兩個三次式都在某個負整數處取值為零：分子在 $n=-1$，分母在 $n=-3$。因式分解後，

$$\frac{a+b}{1-ab}=\frac{(n+1)(n^2+2n+2)}{(n+3)(n^2+2n+2)}=\frac{n+1}{n+3},$$

這裏可以約去公因式，是因為 $n^2+2n+2=(n+1)^2+1>0$。因此 $S_{n+1}=\arctan\left(\frac{n+1}{(n+1)+2}\right)$，這正是 $n+1$ 時的命題。根據歸納法，結果對所有 $n\geq 1$ 成立。✓

無窮和。當 $n\to\infty$ 時，$\frac{n}{n+2}\to 1$。由於 arctan 是連續（continuous）的，

$$\sum_{r=1}^{\infty}\arctan\left(\frac{1}{r^2+r+1}\right)=\lim_{n\to\infty}\arctan\left(\frac{n}{n+2}\right)=\arctan 1=\frac{\pi}{4}.$$



最後的和

設 $s = r - 1$。則

$$r^2 - r + 1 = (s+1)^2 - (s+1) + 1 = s^2 + s + 1.$$

當 r 取遍 $1, 2, 3, \ldots$ 時，s 取遍 $0, 1, 2, \ldots$。$s = 0$ 的一項是 $\arctan 1 = \frac{\pi}{4}$，而其餘各項（$s \geq 1$）恰好組成我們剛剛求和的級數。因此

$$\sum_{r=1}^{\infty} \arctan\left(\frac{1}{r^2 - r + 1}\right) = \arctan 1 + \sum_{s=1}^{\infty} \arctan\left(\frac{1}{s^2 + s + 1}\right) = \frac{\pi}{4} + \frac{\pi}{4} = \frac{\pi}{2}. \qquad \checkmark$$

第二個級數收斂（converge），是因為第一個級數收斂，所以把首項分拆出來是合理的。

驗算

- 當 $n = 2$ 時：$\arctan\frac{1}{3} + \arctan\frac{1}{7}$ 的正切值為 $\frac{10/21}{20/21} = \frac{1}{2} = \frac{2}{4}$，與公式所預測的一致。
- 數值上，$S_{2000} \approx 0.784898$，而 $\arctan\frac{2000}{2002}$ 與之相差不超過 10^{-12}，另外 $\frac{\pi}{4} \approx 0.785398$。
- 把第二個級數求和至 $r = 10^6$，得 $1.5707953\ldots$，與 $\frac{\pi}{2}$ 相差不超過 10^{-6}。這與取 n 項後的差距 $\frac{\pi}{4} - \arctan\frac{n-1}{n+1} = \arctan\frac{1}{n} \approx \frac{1}{n}$ 吻合。

常見錯誤

- 寫到「兩邊的 tan 相等」便停筆。缺少關於取值範圍的論證，證明並不完整。（例如當 $a = b = 2$ 時公式不成立：左邊是正數，但 $\arctan(-\frac{4}{3})$ 是負數。）
- 在歸納步驟中使用公式之前，忘記驗證兩個自變量都落在 $(0, 1)$ 內。
- 把最後的和又寫成 $\frac{\pi}{4}$。平移 $r \mapsto r - 1$ 會在前面多加一項 $\arctan 1$。

STEP 之外

歸納法那部分還有一條更巧妙的路徑。由於 $\frac{1}{r^2+r+1} = \frac{(r+1)-r}{1+r(r+1)}$，每一項都等於 $\arctan(r+1) - \arctan r$。於是這個和可以裂項相消（telescope），得 $\arctan(n+1) - \frac{\pi}{4}$，再由減法公式可知它等於 $\arctan\frac{n}{n+2}$。看出裂項相消的結構，以及利用連續性來論證部分和（partial sum）的極限，都是大學第一門分析（analysis）課程中的日常工具。Abbott 的 *Understanding Analysis*（分析學入門教材）細緻地討論了級數與連續性，而 Siklos 的 *Advanced Problems in Mathematics*（進階數學難題集）則有更多這類 STEP 風格的練習。

詳細解答

STEP III 2002，第 3 題

純數學 • 曲線描繪

STEP III 2002，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/02-s3-q3

關鍵思路

一切都取決於圖像的形狀：圖像由 $x = b$ 出發，起點處有一條鉛垂切線（vertical tangent），曲線向下走，降至唯一的極小值（minimum），然後無止境地上升。只要知道起點的高度和極小值的高度，要數出 $\mathrm{f}(x) = c$ 有多少個解，只需在高度 c 處畫一條水平線，看看它與曲線相交多少次。其後兩條數值方程都是特殊情況，它們的 c^2 值分別落在這些範圍的邊界上和內部。

解答

圖像

在左端，$\mathrm{f}(b) = a\sqrt{b}$。對 $x > b$ 求導，得

$$\mathrm{f}'(x) = \frac{a}{2\sqrt{x}} - \frac{1}{2\sqrt{x-b}}.$$

當 $x \to b^+$ 時，第二項趨向 $-\infty$，所以曲線離開點 $(b, a\sqrt{b})$ 時是筆直向下的（鉛垂切線）。

由於兩個平方根均為正數，

$$\mathrm{f}'(x) < 0 \iff a\sqrt{x-b} < \sqrt{x} \iff a^2(x-b) < x \iff x < \frac{a^2 b}{a^2-1},$$

其中最後一步用到 $a^2 - 1 > 0$。因此 f 在 $[b, x_0]$ 上嚴格遞減（strictly decreasing），在 $[x_0, \infty)$ 上嚴格遞增（strictly increasing），其中

$$x_0 = \frac{a^2 b}{a^2-1}.$$

在極小點處，$x_0 - b = \dfrac{b}{a^2-1}$，所以

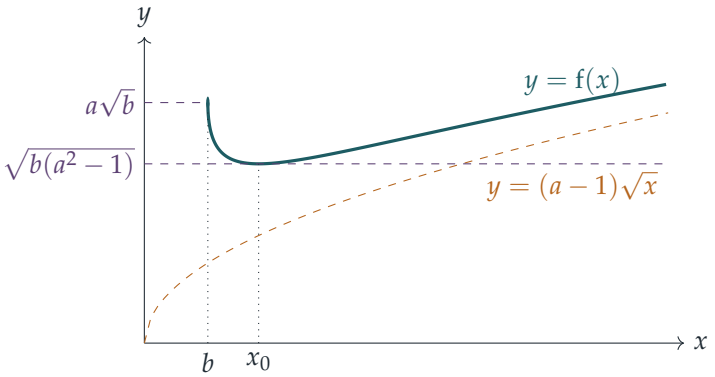
$$\mathrm{f}(x_0) = a\cdot\frac{a\sqrt{b}}{\sqrt{a^2-1}} - \frac{\sqrt{b}}{\sqrt{a^2-1}} = \frac{(a^2-1)\sqrt{b}}{\sqrt{a^2-1}} = \sqrt{b(a^2-1)}.$$

當 x 很大時，有理化（rationalising）後得

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

所以，因為 $a > 1$，$\mathrm{f}(x) \to \infty$，而且曲線位於 $y = (a-1)\sqrt{x}$ 稍上方，並隨 $x \to \infty$ 向它靠近。

最後，起點的高度高於極小值：$a^2b > (a^2-1)b$，所以 $a\sqrt{b} > \sqrt{b(a^2-1)}$。



數算解的個數

記極小值為 $m = \sqrt{b(a^2-1)}$，起點的值為 $M = a\sqrt{b}$，於是 $0 < m < M$。在兩段單調（monotonic）的部分上，f 各自連續（continuous）且嚴格單調，所以它對值域內的每個值都恰好取一次：

- 在 $[b, x_0]$ 上，f 對 $[m, M]$ 內的每個值恰好取一次；
- 在 $[x_0, \infty)$ 上，f 對 $[m, \infty)$ 內的每個值恰好取一次。

這兩段只有 x_0 一個公共點，該處的函數值為 m。

由於對所有 $x \geq b$ 均有 $\mathrm{f}(x) \geq m$，當 $c < m$ 時方程 $\mathrm{f}(x) = c$ 無解。對 $c > 0$ 而言，這等同於 $c^2 < m^2$，即

$$c^2 < b(a^2-1) \quad \Longrightarrow \quad \text{無解},$$

這正是要證明的結果。至於其餘情況：

? **想一想** 為甚麼 $c^2 = a^2b$ 時有兩個解，而不是一個？

當 $c = M$ 時，左邊一段提供的解正是端點 $x = b$ 本身；由於定義域（domain）包含 $x = b$，它是一個名正言順的解。右邊一段則在更遠處再提供一個解。只有當 $c > M$ 時，左邊一段才不再提供解。因此，對 $c > 0$：

- 當 $c^2 = b(a^2 - 1)$（只有 $x = x_0$，即直線剛好碰到極小點），或當 $c^2 > a^2 b$（只有上升的一支能到達這個高度）時，**恰有一個解**；
- 當 $b(a^2 - 1) < c^2 \le a^2 b$（兩支各有一個）時，**恰有兩個解**。

第 (i) 部分

這裏 $a = 3$，$b = 2$，$c = 4$，所以 $b(a^2 - 1) = 16 = c^2$：直線剛好碰到極小點，方程恰有一個解，

$$x = x_0 = \frac{9 \cdot 2}{8} = \frac{9}{4}.$$

確實有 $3 \cdot \frac{3}{2} - \sqrt{\frac{1}{4}} = \frac{9}{2} - \frac{1}{2} = 4$。直接計算亦可：把方程改寫成 $3\sqrt{x} - 4 = \sqrt{x - 2}$ 再兩邊平方，得 $8x - 24\sqrt{x} + 18 = 0$，即 $(2\sqrt{x} - 3)^2 = 0$。重根（repeated root）正是相切在代數上的標記。

第 (ii) 部分

這裏 $a = 3$，$b = 3$，$c = 5$，所以 $b(a^2 - 1) = 24 < 25 = c^2 < 27 = a^2 b$，方程恰有兩個解。要求出這兩個解，把方程寫成 $3\sqrt{x} - 5 = \sqrt{x - 3}$ 再兩邊平方：

$$9x - 30\sqrt{x} + 25 = x - 3 \quad \Longrightarrow \quad 4x - 15\sqrt{x} + 14 = 0.$$

把它看成 $\sqrt{x}$ 的二次方程，有 $(\sqrt{x} - 2)(4\sqrt{x} - 7) = 0$，所以 $\sqrt{x} = 2$ 或 $\sqrt{x} = \frac{7}{4}$，由此得

$$x = 4 \qquad 或 \qquad x = \frac{49}{16}.$$

平方可能會引入增根（spurious roots），所以我們把兩個值都代入原方程驗算。兩者都滿足 $x \ge 3$（因為 $\frac{49}{16} = 3.0625$），而且

$$3 \cdot 2 - \sqrt{1} = 5, \qquad 3 \cdot \tfrac{7}{4} - \sqrt{\tfrac{1}{16}} = \tfrac{21}{4} - \tfrac{1}{4} = 5.$$

所以兩個都是真正的解，與之前數出的兩個解吻合。一如所料，它們分別位於 $x_0 = \frac{27}{8} = 3.375$ 的兩側。

驗算

- 取 $a = \frac{3}{2}$，$b = 2$（即草圖所用的數值），則 $x_0 = \frac{18}{5}$，而 $\mathrm{f}(\frac{18}{5}) = \frac{3}{2}\sqrt{3.6} - \sqrt{1.6} = \sqrt{2.5} \approx 1.581$，低於 $\mathrm{f}(2) = \frac{3}{2}\sqrt{2} \approx 2.121$。

- 隨機選取 200 組 (a,b)，在細密的格點上對 f 作數值掃描，結果證實極小值 $\sqrt{b(a^2-1)}$ 出現在 x_0 處；而數算曲線與 $y = c$ 的交點，在 c^2 的三個範圍內分別得到 0、2 和 1 個解。
- 在 (i) 中，二重根與 c^2 恰好等於 $b(a^2-1)$ 這一點吻合；在 (ii) 中，$c^2 = 25$ 嚴格介乎 24 與 27 之間，而我們確實找到了兩個根。

常見錯誤

- 把 $c^2 = a^2 b$ 歸入「一個解」的情況：$x = b$ 在定義域內，所以要計算在內。
- 平方之後沒有驗算。這裏兩個根都保留下來，但你必須逐一檢驗 $3\sqrt{x} - 5 \geq 0$ 和 $x \geq 3$。
- 忘記了正是 $a > 1$ 才令 $\mathrm{f} \to \infty$；假如 $a \leq 1$，便不會有極小值，整個圖像會完全不同。

STEP 之外

這個數算解的個數的論證，建基於分析學（analysis）的兩個事實：區間上的連續函數會取遍兩端函數值之間的每一個值（介值定理，intermediate value theorem）；而嚴格單調函數對每個值最多取一次。兩者合起來，便把一幅草圖變成一個證明。Abbott 的 *Understanding Analysis* 對這兩點都有細緻的闡述，並說明為何連續性恰恰就是所需的假設。

詳細解答

STEP III 2002，第 4 題

純數學 • 丟番圖不等式

> **STEP III 2002，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/02-s3-q4

關鍵思路

方程 $x^3+x^2=y^3-y^2$ 的一邊增長得快，另一邊卻被 $-y^2$ 拖住，所以 y 必須「超越」x，但超出的幅度又不足 1。這兩個不等式（inequality）把 y 嚴格夾在 x 與 $x+1$ 之間，而這個空隙容不下任何整數。至於最後一部分，額外的解只可能來自零或負數的取值，而一個巧妙的對稱性（symmetry）——把 (m,n) 換成 $(-n,-m)$——可以把分類討論的工夫減半。

解答

第一個結果

設 $x,y>0$ 且 $x^3+x^2=y^3-y^2$，並用反證法（proof by contradiction）假設 $x\geq y$。由於立方和平方在正實數上都是遞增的，

$$x^3+x^2\geq y^3+y^2>y^3-y^2,$$

其中嚴格不等的一步用到 $y^2>0$。這與方程矛盾，所以 $x<y$。

第二個結果

現在設 $0<x\leq y-1$，於是 $y\geq x+1>1$。記 $f(t)=t^3-t^2$。那麼當 $t>\frac{2}{3}$ 時，$f'(t)=3t^2-2t=t(3t-2)>0$，所以 f 在 $[1,\infty)$ 上是遞增函數（increasing function）。

由於 $y \geq x+1$，而且兩者都在這個區間內，

$$\begin{aligned} y^3 - y^2 = f(y) \geq f(x+1) &= (x+1)^2\big((x+1)-1\big) \\ &= x(x+1)^2 = x^3 + 2x^2 + x \\ &= (x^3 + x^2) + (x^2 + x) \\ &> x^3 + x^2, \end{aligned}$$

因為當 $x > 0$ 時 $x^2 + x > 0$。所以 $x^3 + x^2 < y^3 - y^2$，這正是所要證明的。

把兩個結果合起來可知：對於正數 x，任何滿足 $x^3 + x^2 = y^3 - y^2$ 的正數 y 都符合

$$x < y < x+1.$$

沒有正整數對

假設正整數 a 和 b 滿足 $a^3 - b^3 = a^2 + b^2$。右邊是正數，所以 $a^3 > b^3$，即 $a > b$。移項後得

$$b^3 + b^2 = a^3 - a^2,$$

這正是上面的方程，其中 $x = b$、$y = a$，兩者皆為正數。所以 $b < a < b+1$。

想一想 整數這個條件究竟用在哪裏？

只用在這裏：沒有整數嚴格介乎 b 與 $b+1$ 之間。換個說法，$a > b$ 而 a, b 都是整數，就必有 $a \geq b+1$，即 $0 < b \leq a-1$，於是第二個結果給出 $b^3 + b^2 < a^3 - a^2$，與方程矛盾。所以這樣的正整數對並不存在。（若在實數範圍內，解多的是：對每個 $x > 0$ 都有一個 $y \in (x, x+1)$；只是整數剛好碰不上它們。）

所有整數對

現在要找出所有整數 m, n，使得

$$m^3 - n^3 = m^2 + n^2. \qquad (*)$$

以下兩點觀察可以省卻不少工夫。

- 若 (m, n) 是解，則 $(-n, -m)$ 也是解：事實上 $(-n)^3 - (-m)^3 = m^3 - n^3$，而平方和保持不變。
- $(*)$ 的右邊非負，所以 $m^3 \geq n^3$，即 $m \geq n$，而等號只在 $m^2 + n^2 = 0$ 時成立，即 $m = n = 0$。

其中一個是零。若 $n = 0$，則 $m^3 = m^2$，所以 $m = 0$ 或 $m = 1$。若 $m = 0$，則 $-n^3 = n^2$，所以 $n = 0$ 或 $n = -1$。這樣得到 $(0,0)$、$(1,0)$ 和 $(0,-1)$（最後一個是 $(1,0)$ 在上述對稱下的像）。

兩個都是正數。根據上一部分，這是不可能的。

兩個都是負數。這時 $-n$ 和 $-m$ 都是正數，而 $(-n,-m)$ 就會是一個兩項皆正的解，這是不可能的。

正負號相反。由於 $m>n$，我們有 $m>0>n$。記 $n=-p$，其中 $p\geq 1$。那麼 $(*)$ 變成

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

對於正整數，兩項都 ≥ 0，所以兩項都等於零，於是必有 $m=p=1$。這樣得到 $(1,-1)$，而確實 $1-(-1)=2=1+1$。

所以 $m^3-n^3=m^2+n^2$ 的整數解 (m,n) 恰好是

$$(m,n)=(0,0),\quad (1,0),\quad (0,-1),\quad (1,-1).$$

如果「整數對」的意思不包括平凡（trivial）的那一對，就剔除 $(0,0)$。

驗算

每一對都可以用手驗算：對於 $(1,0)$，$1-0=1=1+0$；對於 $(0,-1)$，$0-(-1)=1=0+1$；對於 $(1,-1)$，$2=2$。用電腦搜尋所有 $|m|,|n|\leq 300$ 的情況，找到的正好是這四個解。至於頭兩個結果的驗算：對隨機選取的 $x>0$ 以數值方法解 $y^3-y^2=x^3+x^2$，每次都得到 $x<y<x+1$；例如 $x=1$ 時 $y\approx 1.696$。

常見錯誤

- 在最後一部分忘記了零和負整數，而所有的解偏偏都在那裏。
- 想當然地認定兩個整數中哪一個排在「前面」。這裏立方差是正數，所以較大的數必須排在前面；這一點要寫明。
- 證明第二個結果時只是「展開」了事，沒有指出 t^3-t^2 在 $t\geq 1$ 時遞增；$f(y)\geq f(x+1)$ 這一步正需要這個性質。

STEP 之外

「實數解被嚴格夾在兩個相鄰整數之間，所以沒有整數解」這個論證，是研究丟番圖方程（Diophantine equation）的標準工具，面對整數多項式方程時，往往也是第一樣值得一試的方法。許多這類方程只有有限多個解，也是靠這樣給解定出界限來證明的，然後就可以用有限的搜尋把解全部找出來。Davenport 的 *The Higher Arithmetic*（一本數論入門書）是這類數論的絕佳入門讀物。

詳細解答

STEP III 2002，第 5 題

純數學 • 四次方程的因式分解

STEP III 2002，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/02-s3-q5

關鍵思路

一個二次式是某個完全平方（perfect square）的常數倍，當且僅當它的判別式（discriminant）等於零。第一部分的全部內容就是這一點，而它也是其餘部分的原動力：如果我們能選取 t，使得某個二次式的平方減去該四次式（quartic）後是一個完全平方的倍數，那麼這個四次式就是*兩個平方之差*（difference of two squares），可以立即因式分解（factorise）。這就是費拉里（Ferrari）解四次方程的經典方法，它把四次方程的問題換成一個關於 t 的三次方程（cubic）。

解答

二次式的條件

設 $p \neq 0$。配方（completing the square）得

$$px^2 + qx + r = p\left(x + \frac{q}{2p}\right)^2 + r - \frac{q^2}{4p}.$$

這個式子可寫成 $p(x+h)^2$（h 為某個數）的形式，當且僅當常數 $r - \frac{q^2}{4p}$ 等於零，即

$$q^2 = 4pr,$$

此時 $h = \frac{q}{2p}$。（反過來，展開 $p(x+h)^2 = px^2 + 2phx + ph^2$ 可得 $q^2 = 4p^2h^2 = 4pr$，所以這個條件既是充分條件，也是必要條件。）若 $p = 0$，唯一的可能是 $q = r = 0$；條件 $q^2 = 4pr$ 仍然迫使 $q = 0$，但我們還需要 $r = 0$。

t 所滿足的方程

展開得

$$\left(x^2+\tfrac{1}{2}bx+t\right)^2 = x^4+bx^3+\left(\tfrac{1}{4}b^2+2t\right)x^2+btx+t^2.$$

選用 $\frac{1}{2}b$，正是為了在減去四次式時令 x^3 項互相抵消（x^4 項也同樣抵消）：

$$\left(x^2+\tfrac{1}{2}bx+t\right)^2-\left(x^4+bx^3+cx^2+dx+e\right)=\left(\tfrac{1}{4}b^2+2t-c\right)x^2+(bt-d)x+\left(t^2-e\right).$$

這是一個二次式，其中 $p=\frac{1}{4}b^2+2t-c$、$q=bt-d$ 及 $r=t^2-e$，而我們希望它等於 $k(x+h)^2$，即一個完全平方的倍數（倍數就是 $k=p$）。根據第一部分，t 必須滿足

$$(bt-d)^2 = 4\left(\tfrac{1}{4}b^2+2t-c\right)\left(t^2-e\right).$$

這是一個關於 t 的三次方程（右邊 t^3 項的係數是 8，而左邊只是 t 的二次式），所以它必定有至少一個實根（real root）。

分解四次式

這裏 $b=6$、$c=9$、$d=-2$、$e=-7$，所以

$$\tfrac{1}{4}b^2+2t-c=2t,\qquad bt-d=6t+2,\qquad t^2-e=t^2+7.$$

t 所滿足的方程變成

$$\begin{aligned}(6t+2)^2 &= 8t\left(t^2+7\right)\\ 36t^2+24t+4 &= 8t^3+56t\\ 0 &= 2t^3-9t^2+8t-1.\end{aligned}$$

想一想 怎樣才能快速找出這個三次方程的一個根？

先試較小的數值：各係數之和是 $2-9+8-1=0$，所以 $t=1$ 是一個根。任何一個根都可以用，因為每個根都給出一個有效的因式分解；最簡單的根給出最簡單的分解。

取 $t=1$，所得的差是

$$2x^2+8x+8=2(x+2)^2,$$

所以 $k=2$ 及 $h=2$，而且確實有 $8^2=4\cdot 2\cdot 8$。因此

$$\left(x^2+3x+1\right)^2-\left(x^4+6x^3+9x^2-2x-7\right)=2(x+2)^2,$$

即

$$x^4+6x^3+9x^2-2x-7=\left(x^2+3x+1\right)^2-\left(\sqrt{2}\,(x+2)\right)^2.$$

這是兩個平方之差，所以

$$\begin{aligned} x^4+6x^3+9x^2-2x-7 &= \left(x^2+3x+1-\sqrt{2}(x+2)\right)\left(x^2+3x+1+\sqrt{2}(x+2)\right) \\ &= \left(x^2+(3-\sqrt{2})x+1-2\sqrt{2}\right)\left(x^2+(3+\sqrt{2})x+1+2\sqrt{2}\right). \end{aligned}$$

由於 $k=2>0$，它的平方根是實數，所以這是兩個實係數的二次因式，正如題目所要求。

三次方程的另外兩個根是由 $2t^2-7t+1=0$ 得出的 $t=\frac{1}{4}\left(7\pm\sqrt{41}\right)$，它們同樣滿足 $k=2t>0$，並給出另外兩種把四次式的四個（實）根配對成實二次因式的方法。這兩種分解的係數繁複得多，所以應該選用 $t=1$。

驗算

- 乘開：兩個因式之積的 x^3 係數是 $(3-\sqrt{2})+(3+\sqrt{2})=6$，常數項是 $(1-2\sqrt{2})(1+2\sqrt{2})=1-8=-7$，兩者皆正確。
- 當 $x=1$ 時，四次式的值是 $1+6+9-2-7=7$，而因式分解給出 $(5-\sqrt{2}-2\sqrt{2})(5+\sqrt{2}+2\sqrt{2})=25-18=7$。
- 兩個因式的判別式分別是 $7+2\sqrt{2}\approx 9.83$ 和 $7-2\sqrt{2}\approx 4.17$，兩者皆為正數，所以四次式有四個實根，這與三次方程的三個根全都給出實因式分解一致。我們亦在 200 個隨機點上以數值方法核對過這個因式分解。

常見錯誤

- 把 $(x^2+\frac{1}{2}bx+t)^2$ 的 x^2 係數算錯：它應是 $\frac{1}{4}b^2+2t$，因為除了 $(\frac{1}{2}bx)^2$ 之外，交叉項 $2\cdot x^2\cdot t$ 也有貢獻。
- 把判別式條件寫成 $q^2=pr$，或者忘記 k 就是首項係數（leading coefficient）$\frac{1}{4}b^2+2t-c$，而它必須是正數才能得到實因式分解。

STEP 之外

這個關於 t 的三次方程稱為**預解三次式**（resolvent cubic），而這就是費拉里在十六世紀提出的方法：每個四次方程都可以化為一個三次方程，再用卡爾達諾（Cardano）公式求解。這個方法行得通的深層原因，以及一般五次方程（quintic）沒有這類公式的原因，都屬於伽羅瓦理論（Galois theory）的範疇，該理論研究的是多項式的根可以怎樣置換（permute）。若想初步認識域（field）與多項式的代數，可以讀 Allenby 的 *Rings, Fields and Groups*（環、域與羣）。

詳細解答

STEP III 2002，第 6 題

純數學 • 非線性微分方程

> **STEP III 2002，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/02-s3-q6

關鍵思路

這條方程因為有四次方而顯得嚇人，但 $y\,\frac{\mathrm{d}y}{\mathrm{d}x}$ 正好是 y^2 的導數（derivative）的一半。因此作代換（substitution）$u = y^2 - 1$：方程隨即化簡為只含 u 和 u' 的方程，取平方根之後便可分離變數（separate variables）。唯一要小心的是正負號：u 的正負號把平面分成帶狀區域 $|y| < 1$ 和區域 $|y| > 1$，兩者給出不同的曲線族（family of curves）。

解答

設 $u = y^2 - 1$，則 $u' = 2y\,\frac{\mathrm{d}y}{\mathrm{d}x}$，而 $y^4\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^4 = \left(\frac{1}{2}u'\right)^4$。方程變成

$$\frac{1}{16}(u')^4 = u^2 \quad\Longleftrightarrow\quad (u')^2 = 4|u|,$$

這裏對兩邊取非負平方根（$(u')^2$ 和 $4|u|$ 都是非負的，所以沒有遺漏任何解）。

首先留意，解曲線（solution curve）永遠不會與 $y = 0$ 相遇：在那裏原方程的左邊是 0，右邊卻是 1。所以凡經過任何一個給定點的解，都處處滿足 $y > 0$。

> ? **想一想** 怎樣解 $(u')^2 = 4u$，而不必為 u' 該取哪個正負號而煩惱？

在 $u > 0$ 之處，可以寫 $u = v^2$，其中 $v = \sqrt{u} > 0$。於是 $u' = 2vv'$，所以 $(u')^2 = 4u$ 變成 $4v^2(v')^2 = 4v^2$，即 $(v')^2 = 1$。由於 v' 連續而且永不為零，它的正負號保持不變，所以 $v' = \pm 1$，並且 $v = \pm(x - a)$，其中 a 為某個常數。無論哪一種情況，都有

$$u = v^2 = (x - a)^2.$$

正負號的不確定性在平方之後自然消失。對 $u < 0$ 的情況，以 $u = -v^2$ 作同樣的論證即可。

第 (i) 部分：經過 $(0, \frac{1}{2}\sqrt{3})$

這裏 $y^2 = \frac{3}{4} < 1$，所以在這點附近 $u < 0$，即 $|y| < 1$。寫 $1 - y^2 = -u = v^2$，其中 $v > 0$；那麼 $(u')^2 = -4u$ 給出 $(v')^2 = 1$，與上面完全一樣，所以 $1 - y^2 = (x - a)^2$，也就是

$$(x - a)^2 + y^2 = 1 :$$

圓心在 x 軸上、半徑為 1 的圓。在 $x = 0$ 處需要 $a^2 = 1 - \frac{3}{4} = \frac{1}{4}$，所以 $a = \pm\frac{1}{2}$。由於 $y > 0$，經過 $(0, \frac{1}{2}\sqrt{3})$ 的解曲線是上半圓

$$y = \sqrt{1 - \left(x - \tfrac{1}{2}\right)^2} \qquad 及 \qquad y = \sqrt{1 - \left(x + \tfrac{1}{2}\right)^2}.$$

直接驗證：在 $(x - a)^2 + y^2 = 1$ 上，隱函數微分（implicit differentiation）給出 $y\frac{\mathrm{d}y}{\mathrm{d}x} = -(x - a)$，所以 $y^4\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^4 = (x - a)^4 = (1 - y^2)^2 = (y^2 - 1)^2$。✓

第 (ii) 部分：經過 $(0, \frac{1}{2}\sqrt{5})$

現在 $y^2 = \frac{5}{4} > 1$，所以 $u > 0$，上面的論證給出 $y^2 - 1 = (x - a)^2$，也就是

$$y^2 - (x - a)^2 = 1 :$$

中心為 $(a, 0)$ 的等軸雙曲線（rectangular hyperbola）。在 $x = 0$ 處，$a^2 = \frac{5}{4} - 1 = \frac{1}{4}$，所以同樣有 $a = \pm\frac{1}{2}$，而由於 $y > 0$，解曲線是上支

$$y = \sqrt{1 + \left(x - \tfrac{1}{2}\right)^2} \qquad 及 \qquad y = \sqrt{1 + \left(x + \tfrac{1}{2}\right)^2},$$

對所有 x 都有定義。直接驗證：$y\frac{\mathrm{d}y}{\mathrm{d}x} = x - a$，所以左邊是 $(x - a)^4 = (y^2 - 1)^2$。✓

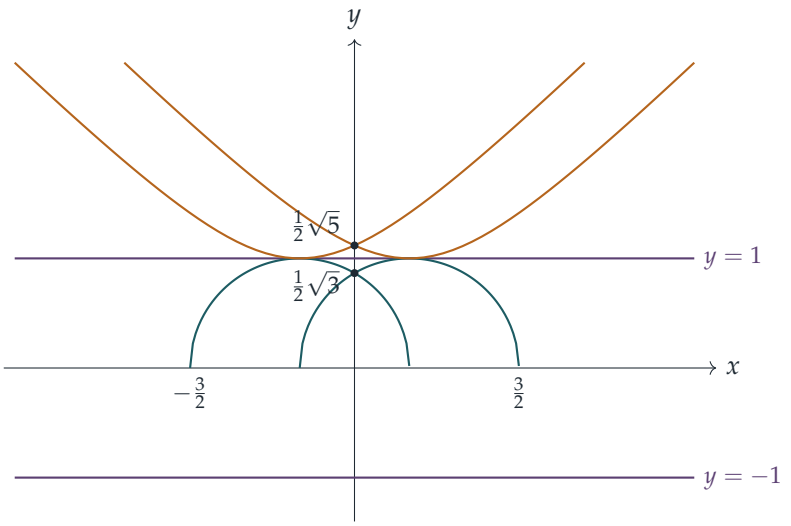
直線 $y = \pm 1$

若對所有 x 都有 $y = 1$ 或 $y = -1$，則 $\frac{\mathrm{d}y}{\mathrm{d}x} = 0$，所以左邊是 0；而 $y^2 - 1 = 0$，所以右邊也是 0。兩條直線都是解。

想一想 這兩條直線與那些圓和雙曲線有甚麼關係？

每個圓 $(x - a)^2 + y^2 = 1$ 和每條雙曲線 $y^2 - (x - a)^2 = 1$ 都在 $(a, 1)$ 以零斜率與 $y = 1$ 相切（同樣地，在 $(a, -1)$ 與 $y = -1$ 相切）。所以 $y = \pm 1$ 是這兩個曲線族的包絡線（envelope），而它們並不屬於任何一族：它們來自 $u = 0$，而我們除以 v 時已把這情況排除在外。由此得出一個結果：經過 $y = 1$ 上的點的解並不唯一。例如，沿着 $a = \frac{1}{2}$ 的圓由 $(0, \frac{1}{2}\sqrt{3})$ 走到它的頂點 $(\frac{1}{2}, 1)$，然後沿 $y = 1$ 走一段，再在 $x \geq b$ 時沿雙曲線 $y = \sqrt{1 + (x - b)^2}$ 向上走（任意 $b \geq \frac{1}{2}$ 皆可），便得到一條足夠光滑（斜率連續）而且處處滿足方程的曲線。上面四條曲線是經過給定點的「基本」解；經過這些點的其他每一個解，都是由這些曲線、$y = 1$，以及其他與 $y = 1$ 相切的圓和雙曲線的片段拼合而成。

略圖



半圓（藍綠色）的圓心為 $(\pm\frac{1}{2}, 0)$，半徑為 1，在 $x = \pm\frac{1}{2} \pm 1$ 處垂直地與 x 軸相遇，並在 $x = \pm\frac{1}{2}$ 處與 $y = 1$ 相切。雙曲線的分支（琥珀色）在同樣的兩點與 $y = 1$ 相切，其漸近線（asymptote）為 $y = \pm(x \mp \frac{1}{2})$，即它們的增長與 $|x|$ 相若。每一對曲線都關於 y 軸對稱，而每對的兩條曲線在該軸上的給定點相交。

驗算

- 在 $a = \frac{1}{2}$ 的圓上的點 $(0, \frac{1}{2}\sqrt{3})$：$y' = -(x-a)/y = 1/\sqrt{3}$，所以左邊是 $\frac{9}{16} \cdot \frac{1}{9} = \frac{1}{16}$，而右邊是 $(\frac{3}{4} - 1)^2 = \frac{1}{16}$。
- 由 $(0, \frac{1}{2}\sqrt{5})$ 出發，取正號對 $y' = \pm\sqrt{|y^2 - 1|}/y$ 作數值積分，得 $y(1.5) = 2.2360680$，與 $a = -\frac{1}{2}$ 的雙曲線上的 $\sqrt{1 + 2^2} = \sqrt{5}$ 吻合。

常見錯誤

- 在兩個區域都把平方根取成 $(u')^2 = 4u$；對 $|y| < 1$ 而言，這沒有實數解，於是那些圓就遺漏了。
- 在每一步只保留平方根的其中一個正負號，結果經過每個點只找到一條曲線，而不是兩條。
- 忘記了 $y = \pm 1$ 並不屬於任何一族，儘管它們與每一個成員都相切。

STEP 之外

直線 $y = \pm 1$ 是奇解（singular solution）：它們是通解（general solution）的包絡線，無論常數 a 取甚麼值都得不出來。它們恰恰出現在一階方程的唯一性定理（皮卡–林德勒夫定理，Picard–Lindelöf）失效之處，因為 $\sqrt{|y^2-1|}$ 在 $y = \pm 1$ 處不滿足利普希茨（Lipschitz）條件，這正是解可以在那裏分岔的原因。這類包絡線常見於對 y' 而言並非線性的一階方程（克萊羅（Clairaut）方程是經典例子）；Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科書）討論了存在性與唯一性，以及它們在甚麼情況下失效。

詳細解答

STEP III 2002，第 7 題

純數學 • 坐標幾何恆等式

STEP III 2002，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/02-s3-q7

關鍵思路

$s^2+b^2-a^2$ 這個式子應該似曾相識：在三角形 OAB 中（其中 $OA=s$、$OB=b$）使用餘弦定理（cosine rule），得出的正是這個式子。因此，第一個恆等式（identity）其實就是開首關於 $\cos^2\alpha+\cos^2\beta$ 的結果，應用於 OB 與兩條坐標軸所成的兩個角。其後，第二個恆等式純粹是建基於第一個恆等式的代數運算，而不等式（inequality）亦隨之而來，因為兩個平方之和必不小於其中任何一個平方。

解答

關於角的結果

首先假設 $\alpha+\beta=\frac{1}{2}\pi$。則 $\cos\beta=\cos(\frac{1}{2}\pi-\alpha)=\sin\alpha$，所以

$$\cos^2\alpha+\cos^2\beta=\cos^2\alpha+\sin^2\alpha=1.$$

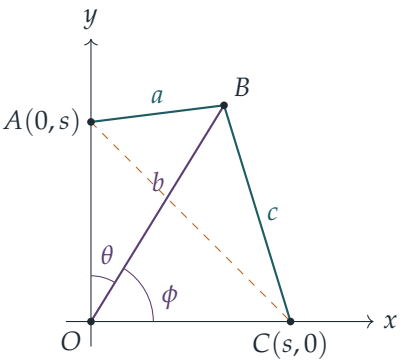
反過來，假設 $\cos^2\alpha+\cos^2\beta=1$。則 $\cos^2\beta=1-\cos^2\alpha=\sin^2\alpha$。$\alpha$ 和 β 都是銳角（acute angle），所以 $\cos\beta>0$ 且 $\sin\alpha>0$，取正平方根便得

$$\cos\beta=\sin\alpha=\cos(\tfrac{1}{2}\pi-\alpha).$$

現在 β 和 $\frac{1}{2}\pi-\alpha$ 都在 $(0,\frac{1}{2}\pi)$ 之內，而餘弦在這個區間上嚴格遞減（strictly decreasing），因此每個值只取一次。故此 $\beta=\frac{1}{2}\pi-\alpha$，即 $\alpha+\beta=\frac{1}{2}\pi$。這就證明了「當且僅當」（if and only if）。

第一個恆等式

設 $\theta=\angle AOB$ 及 $\phi=\angle BOC$。



在三角形 OAB（邊長 $OA = s$、$OB = b$、$AB = a$）和三角形 OCB（邊長 $OC = s$、$OB = b$、$CB = c$）中使用餘弦定理，得

$$\cos\theta = \frac{s^2 + b^2 - a^2}{2sb}, \qquad \cos\phi = \frac{s^2 + b^2 - c^2}{2sb}.$$

由於 B 位於開的第一象限（open first quadrant）內，射線 OB 嚴格地夾在正 y 軸（沿 OA 方向）與正 x 軸（沿 OC 方向）之間。所以 θ 和 ϕ 都是銳角，且 $\theta + \phi = \frac{1}{2}\pi$。根據上面「僅當」的方向，$\cos^2\theta + \cos^2\phi = 1$，即

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

兩邊乘以 $4s^2b^2$，便得第一個恆等式：

$$(s^2 + b^2 - a^2)^2 + (s^2 + b^2 - c^2)^2 = 4s^2b^2.$$

（若要直接驗證，可設 $B = (x, y)$。則 $a^2 = x^2 + (y - s)^2$、$b^2 = x^2 + y^2$ 及 $c^2 = (x - s)^2 + y^2$，所以 $s^2 + b^2 - a^2 = 2sy$ 及 $s^2 + b^2 - c^2 = 2sx$，而左方等於 $4s^2(x^2 + y^2) = 4s^2b^2$。）

第二個恆等式

為了令代數運算保持整潔，設

$$P = s^2 + b^2 - a^2, \qquad Q = s^2 + b^2 - c^2, \qquad S = s^2 + b^2, \qquad D = s^2 - b^2,$$

於是 $a^2 = S - P$ 及 $c^2 = S - Q$。第一個恆等式說的是 $P^2 + Q^2 = 4s^2b^2$。由於 $4s^2b^2 = (s^2 + b^2)^2 - (s^2 - b^2)^2 = S^2 - D^2$，它可以改寫成

$$D^2 = S^2 - P^2 - Q^2. \tag{†}$$

現在 $a^2 + c^2 = 2S - P - Q$，而 $2s^2 = S + D$、$2b^2 = S - D$，所以

$$2s^2 - a^2 - c^2 = (P + Q - S) + D, \qquad 2b^2 - a^2 - c^2 = (P + Q - S) - D.$$

想一想 為甚麼值得把它們拆成 $(P+Q-S)\pm D$ 的形式？

因為 $(u+D)^2+(u-D)^2=2u^2+2D^2$：交叉項互相抵銷，於是 D 只以 D^2 的形式出現，而 (†) 讓我們可以把它換成只涉及 P、Q 和 S 的式子。取 $u=P+Q-S$，

$$\begin{aligned}&(2s^2-a^2-c^2)^2+(2b^2-a^2-c^2)^2\\&\quad=2(P+Q-S)^2+2D^2\\&\quad=2(P^2+Q^2+S^2+2PQ-2PS-2QS)+2(S^2-P^2-Q^2)\\&\quad=4S^2-4PS-4QS+4PQ\\&\quad=4(S-P)(S-Q)\\&\quad=4a^2c^2,\end{aligned}$$

其中第二行用了 (†)。這就是第二個恆等式。

不等式

第二個恆等式左方的兩個平方都是非負的，所以

$$(2b^2-a^2-c^2)^2\le 4a^2c^2,\qquad \text{所以}\qquad |2b^2-a^2-c^2|\le 2ac,$$

這裏取了非負平方根（注意 $a,c>0$）。這即是說

$$-2ac\le 2b^2-a^2-c^2\le 2ac,$$

亦即

$$a^2-2ac+c^2\le 2b^2\le a^2+2ac+c^2,$$

也就是 $(a-c)^2\le 2b^2\le(a+c)^2$，正是所求。

等號恰好在 $2s^2=a^2+c^2$ 時成立。由於 $AC^2=2s^2$，根據畢氏定理的逆定理（converse of Pythagoras），這表示 $\angle ABC$ 是直角：B 位於以 AC 為直徑的圓上。

驗算

- 取 $s=2$ 及 $B=(1,3)$。則 $a^2=2$、$b^2=c^2=10$。第一個恆等式成為 $12^2+4^2=160=4\cdot 4\cdot 10$；第二個恆等式成為 $(-4)^2+8^2=80=4\cdot 2\cdot 10$；而不等式則是 $3.06\le 20\le 20.94$。
- 點 $B=(s,s)$ 是正方形的第四個頂點，所以 $\angle ABC$ 是直角，且 $a=c=s$、$b=s\sqrt{2}$：於是 $2b^2=4s^2=(a+c)^2$，正是等號成立的情況。
- 在 2×10^5 個點 B 上對全部三個結果作隨機測試，沒有發現任何不成立的情況。

常見錯誤

- 在逆命題中漏掉銳角這個條件：沒有這個條件，$\cos^2\beta = \sin^2\alpha$ 並不能推出 $\beta = \frac{1}{2}\pi - \alpha$（例如 $\beta = \frac{1}{2}\pi + \alpha$ 同樣滿足）。
- 不利用第一個恆等式，而以 a, b, c, s 硬生生地展開第二個恆等式。這樣做是可行的，但很容易漏掉某一項；採用 S、D 的代換，可令運算不致失控。
- 對 $(2b^2 - a^2 - c^2)^2 \leq 4a^2c^2$ 取平方根時漏掉絕對值（modulus），結果只得到不等式的其中一半。

STEP 之外

$2b^2 - a^2 - c^2$ 這個量等於 $2s(x + y - s)$，即三角形 ABC 面積的 ± 4 倍，所以第二個恆等式其實是把餘弦定理和面積公式 $\frac{1}{2}ac\sin B$ 合併成 $\cos^2 B + \sin^2 B = 1$。上界亦是托勒密不等式（Ptolemy's inequality）的一個特例：對任意四點，$OB \cdot AC \leq OA \cdot BC + OC \cdot AB$，而等號在圓內接四邊形（cyclic quadrilateral）的情況下成立，在這裏即是 $bs\sqrt{2} \leq s(a + c)$。這類距離之間的關係，可以引伸至度量空間（metric space）和距離幾何（distance geometry）的研究；Siklos 的 *Advanced Problems in Mathematics*（高等數學難題集）中有更多同類風格的題目。

詳細解答

STEP III 2002，第 8 題

純數學 • 托勒密定理

STEP III 2002，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/02-s3-q8

關鍵思路

單位圓上兩點之差，只要抽出「平均」角，便可漂亮地因式分解：$\mathrm{e}^{\mathrm{i}\alpha} - \mathrm{e}^{\mathrm{i}\beta} = \mathrm{e}^{\mathrm{i}(\alpha+\beta)/2}\left(\mathrm{e}^{\mathrm{i}(\alpha-\beta)/2} - \mathrm{e}^{-\mathrm{i}(\alpha-\beta)/2}\right)$，而括號內正是 $2\mathrm{i}$ 乘以一個正弦。這樣一來，每條弦（chord）的模（modulus）和輻角（argument）便同時求得。至於最後一部分，兩個輻角相同的複數指向同一方向，所以它們的模直接相加；餘下只需一條把三個乘積連繫起來的代數恆等式。

解答

$u_1 - u_2$ 的輻角

設 $u_k = \mathrm{e}^{\mathrm{i}\theta_k}$。抽出因子 $\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)/2}$，

$$\begin{aligned} u_1 - u_2 &= \mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)/2}\left(\mathrm{e}^{\mathrm{i}(\theta_1-\theta_2)/2} - \mathrm{e}^{-\mathrm{i}(\theta_1-\theta_2)/2}\right) \\ &= \mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)/2} \cdot 2\mathrm{i}\sin\tfrac{1}{2}(\theta_1-\theta_2) \\ &= 2\sin\tfrac{1}{2}(\theta_2-\theta_1)\cdot(-\mathrm{i})\,\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)/2} \\ &= 2\sin\tfrac{1}{2}(\theta_2-\theta_1)\,\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2-\pi)/2}, \end{aligned}$$

最後一行用了 $-\mathrm{i} = \mathrm{e}^{-\mathrm{i}\pi/2}$。

? **想一想** 為甚麼可以從這條式子直接讀出輻角？又為甚麼正弦要寫成 $\sin\frac{1}{2}(\theta_2-\theta_1)$ 而不是 $\sin\frac{1}{2}(\theta_1-\theta_2)$？

形如 $r\mathrm{e}^{\mathrm{i}\phi}$ 的數，只有在 $r>0$ 時，其輻角才是 ϕ（可相差 2π 的倍數）。這裏 $0<\theta_2-\theta_1<2\pi$，所以 $0<\frac{1}{2}(\theta_2-\theta_1)<\pi$，正弦嚴格為正。若反過來寫，正弦便是負數，輻角就會

相差 π。

因此 $|u_1-u_2| = 2\sin\frac{1}{2}(\theta_2-\theta_1)$，而 $\frac{1}{2}(\theta_1+\theta_2-\pi)$ 是 u_1-u_2 的一個輻角。取 arg 為 $(-\pi,\pi]$ 內的主值（principal value）：由於 $-\pi<\theta_1<\theta_2<\pi$，

$$-\tfrac{3\pi}{2} < \tfrac{1}{2}(\theta_1+\theta_2-\pi) < \tfrac{\pi}{2}.$$

若此值大於 $-\pi$，它本身已是主輻角（$n=0$）；否則它落在 $(-\frac{3\pi}{2},-\pi]$ 內，加上 2π 便回到範圍之內（$n=1$）。故此

$$\arg(u_1-u_2) = \tfrac{1}{2}(\theta_1+\theta_2-\pi)+2n\pi, \qquad n=0 \text{ 或}1,$$

正如題目所求。

兩個乘積的輻角相同

以上推導除了用到 $\theta_1<\theta_2$ 且兩者皆在 $(-\pi,\pi)$ 內之外，並沒有用到 u_1,u_2 的任何其他性質。所以對任意 $j<k$，

$$u_j-u_k = 2\sin\tfrac{1}{2}(\theta_k-\theta_j)\,\mathrm{e}^{\mathrm{i}(\theta_j+\theta_k-\pi)/2},$$
$$u_k-u_j = -(u_j-u_k) = 2\sin\tfrac{1}{2}(\theta_k-\theta_j)\,\mathrm{e}^{\mathrm{i}(\theta_j+\theta_k+\pi)/2},$$

其中用了 $-1=\mathrm{e}^{\mathrm{i}\pi}$，而兩式前面的實數因子都是正數。對 $j<k$，記 $s_{jk}=\sin\frac{1}{2}(\theta_k-\theta_j)>0$，並記 $\Sigma=\theta_1+\theta_2+\theta_3+\theta_4$。相乘（輻角相加），

$$(u_1-u_2)(u_4-u_3) = 4s_{12}s_{34}\,\mathrm{e}^{\mathrm{i}[(\theta_1+\theta_2-\pi)+(\theta_3+\theta_4+\pi)]/2} = 4s_{12}s_{34}\,\mathrm{e}^{\mathrm{i}\Sigma/2},$$
$$(u_1-u_4)(u_3-u_2) = 4s_{14}s_{23}\,\mathrm{e}^{\mathrm{i}[(\theta_1+\theta_4-\pi)+(\theta_2+\theta_3+\pi)]/2} = 4s_{14}s_{23}\,\mathrm{e}^{\mathrm{i}\Sigma/2}.$$

兩者都是 $\mathrm{e}^{\mathrm{i}\Sigma/2}$ 的正數倍，所以各自的輻角都是 $\frac{1}{2}\Sigma$（可相差 2π 的倍數）。因此它們的主輻角只可能相差 2π 的倍數：

$$\arg\big((u_1-u_2)(u_4-u_3)\big) = \arg\big((u_1-u_4)(u_3-u_2)\big)+2n\pi$$

其中 n 為某個整數。（事實上 $n=0$，因為兩個主輻角若相差 2π 的倍數，就必定相等；但題目只要求較弱的這個結論。）

最後一部分

把三個乘積記為

$$A=(u_1-u_2)(u_4-u_3), \qquad B=(u_1-u_4)(u_3-u_2), \qquad C=(u_1-u_3)(u_4-u_2).$$

展開得

$$\begin{aligned} A+B &= (u_1u_4-u_1u_3-u_2u_4+u_2u_3)+(u_1u_3-u_1u_2-u_3u_4+u_2u_4)\\ &= u_1u_4+u_2u_3-u_1u_2-u_3u_4,\\ C &= u_1u_4-u_1u_2-u_3u_4+u_2u_3, \end{aligned}$$

所以 $A+B=C$。這純粹是代數運算，對任意四個複數都成立。

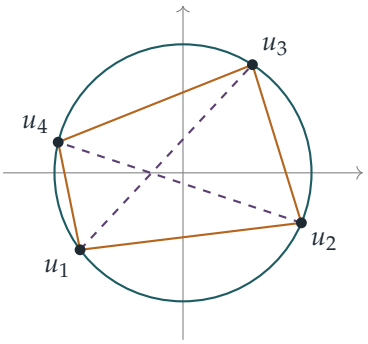
現在利用上一部分的結果：$A = a\,\mathrm{e}^{\mathrm{i}\Sigma/2}$ 及 $B = b\,\mathrm{e}^{\mathrm{i}\Sigma/2}$，其中 $a = 4s_{12}s_{34} > 0$，$b = 4s_{14}s_{23} > 0$。於是 $|A| = a$，$|B| = b$，並且

$$|C| = |A+B| = \left|(a+b)\,\mathrm{e}^{\mathrm{i}\Sigma/2}\right| = a+b = |A|+|B|,$$

因為 $a+b>0$ 而 $|\mathrm{e}^{\mathrm{i}\Sigma/2}| = 1$。即是說，

$$|(u_1-u_2)(u_4-u_3)| + |(u_1-u_4)(u_3-u_2)| = |(u_1-u_3)(u_4-u_2)|.$$

重點在於：三角不等式（triangle inequality）$|A+B| \le |A|+|B|$ 恰好在 A 與 B 指向同一方向時變成等式，而這正是中間部分所證明的。



從幾何角度看，$|u_j - u_k|$ 是一條弦的長度；又因為輻角遞增，$u_1u_2u_3u_4$ 是一個頂點依次排列的圓內接四邊形（quadrilateral inscribed in the circle）。這條恆等式說明：兩組對邊之積的和，等於兩條對角線（虛線）之積。這就是托勒密定理（Ptolemy's theorem）。

驗算

- 正方形：$\theta_k = -\frac{3\pi}{4}, -\frac{\pi}{4}, \frac{\pi}{4}, \frac{3\pi}{4}$。每條邊長 $\sqrt{2}$，每條對角線長 2，而確實 $2+2=4$。
- 取 $\theta = (-2.5, -0.4, 1.0, 2.9)$，得 $|A| \approx 2.8223$，$|B| \approx 1.1013$ 及 $|C| \approx 3.9236 = |A|+|B|$。
- 隨機選取 2×10^5 組依次排列的 θ_k，$\arg(u_1-u_2)$ 的公式每次都以 $n=0$ 或 1 成立（兩個值都有出現），而恆等式的誤差不超過 2×10^{-15}。

常見錯誤

- 寫出 $u_1-u_2 = 2\mathrm{i}\sin\frac{1}{2}(\theta_1-\theta_2)\,\mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)/2}$ 後，把輻角讀成 $\frac{1}{2}(\theta_1+\theta_2+\pi)$，忘記了正弦是負數。

- 把 $u_4 - u_3$ 當作與 $u_3 - u_4$ 有相同的輻角：把差的次序倒轉，輻角會增加 π。
- 斷言 $|A| + |B| = |A + B|$ 而不說明原因；這需要 A 與 B 的輻角相同。

STEP 之外

托勒密定理有逆定理（converse），也有更一般的形式：對平面上任意四點，$|A| + |B| \geq |C|$（只是把三角不等式應用於 $A + B = C$），等號恰好在四點依次位於同一個圓或同一條直線上時成立。這背後是交比（cross-ratio）$\frac{(u_1-u_2)(u_3-u_4)}{(u_1-u_4)(u_3-u_2)}$：它恰好在四點共圓（concyclic）時為實數，並且在莫比烏斯變換（Möbius transformations）下保持不變，是大學一年級複分析（complex analysis）與幾何的核心對象。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）以精彩的圖畫闡述了這些想法。

詳細解答

STEP III 2002，第 9 題

力學 • 質心與傾倒

STEP III 2002，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/02-s3-q9

關鍵思路

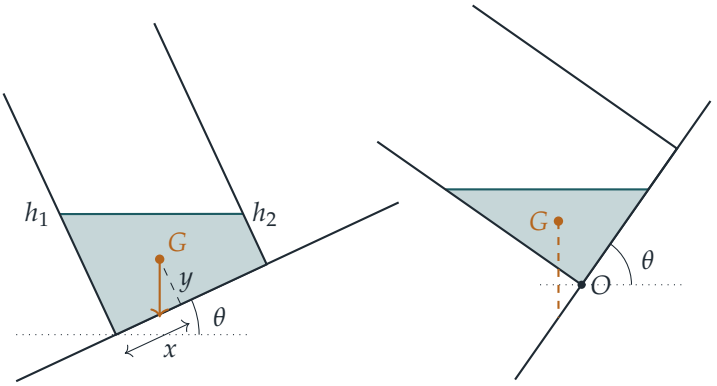
流體的自由表面（free surface）是水平的，所以在容器的參考系（容器傾斜了 θ 角）中，它是一條與底面成 θ 角的斜直線。因此，若液面與兩邊側壁都相交，流體的橫截面就是梯形（trapezium）；若液面未到達上方側壁便先碰到底面，橫截面就是直角三角形。這樣，求質心（centre of mass）便成為一個二維的形心（centroid）計算。至於傾倒（toppling），由於容器是輕的，一切取決於通過流體質心的鉛垂線與斜面相交於何處。

解答

建立模型

我們把容器看成一個輕的稜柱體，器壁厚度可忽略；把流體看成相對於容器靜止的均勻連續體；並假設斜面足夠粗糙，容器不會滑動。垂直於底面水平邊的每一個鉛垂橫截面都相同，而流體是均勻的，所以流體的質心就是其中一個橫截面的形心。該橫截面的寬度為 a，面積為 $ka^3/a = ka^2$。

沿底面由下方側壁起（沿斜面向上）量度 X，並垂直於底面量度 Y。隨着 X 增加，底面以 θ 角相對於水平面上升，所以水平的液面相對於底面以 θ 角下降：X 每增加一個單位，液面離底面的高度便減少 $\tan\theta$。記 $d = a\tan\theta$ 為整個寬度上的總下降量。



左圖顯示 $k > \frac{1}{2}\tan\theta$ 的情況（流體接觸兩邊側壁）；右圖顯示 $k < \frac{1}{2}\tan\theta$ 且 $\theta > 45°$ 的情況，此時通過 G 的鉛垂線落在底面之外。

$k > \frac{1}{2}\tan\theta$ 的情況

設下方側壁和上方側壁處的深度分別為 h_1 和 h_2。那麼 $h_1 - h_2 = d$，而梯形的面積為 $\frac{1}{2}a(h_1 + h_2) = ka^2$，所以 $h_1 + h_2 = 2ka$。因此

$$h_2 = ka - \tfrac{1}{2}d = a\big(k - \tfrac{1}{2}\tan\theta\big),$$

它為正數當且僅當 $k > \frac{1}{2}\tan\theta$。所以這正是液面與兩邊側壁都相交的情況。

把梯形分割成一個 $a \times h_2$ 的長方形，以及疊在其上的一個直角三角形，其兩條直角邊分別為 a（沿 $Y = h_2$）和 d（沿下方側壁向上）。

- 長方形：面積 ah_2，形心 $\big(\frac{1}{2}a,\ \frac{1}{2}h_2\big)$。
- 三角形：頂點為 $(0, h_2)$、(a, h_2)、$(0, h_2 + d)$，所以面積為 $\frac{1}{2}ad$，形心（各頂點的平均）為 $\big(\frac{1}{3}a,\ h_2 + \frac{1}{3}d\big)$。

對下方側壁取矩（taking moments），總面積為 ka^2，

$$ka^2\,x = ah_2 \cdot \tfrac{1}{2}a + \tfrac{1}{2}ad \cdot \tfrac{1}{3}a \quad\Longrightarrow\quad kx = \tfrac{1}{2}h_2 + \tfrac{1}{6}d = \tfrac{1}{2}ka - \tfrac{1}{4}d + \tfrac{1}{6}d = \tfrac{1}{2}ka - \tfrac{1}{12}d.$$

除以 ka 並利用 $d = a\tan\theta$，

$$\frac{x}{a} = \frac{1}{2} - \frac{\tan\theta}{12k}.$$

對底面取矩，

$$ka^2\,y = ah_2 \cdot \tfrac{1}{2}h_2 + \tfrac{1}{2}ad\big(h_2 + \tfrac{1}{3}d\big),$$
$$ka\,y = \tfrac{1}{2}h_2^2 + \tfrac{1}{2}dh_2 + \tfrac{1}{6}d^2 = \tfrac{1}{2}h_2(h_2 + d) + \tfrac{1}{6}d^2.$$

現在 $h_2 = ka - \frac{1}{2}d$ 而 $h_2 + d = ka + \frac{1}{2}d$，所以 $h_2(h_2 + d) = k^2a^2 - \frac{1}{4}d^2$，於是

$$ka\,y = \tfrac{1}{2}k^2a^2 - \tfrac{1}{8}d^2 + \tfrac{1}{6}d^2 = \tfrac{1}{2}k^2a^2 + \tfrac{1}{24}d^2.$$

除以 ka^2，

$$\frac{y}{a} = \frac{k}{2} + \frac{\tan^2\theta}{24k},$$

正如所求。

$k < \frac{1}{2}\tan\theta$ 的情況

這時 h_2 會是負數，所以液面在距離下方側壁 $b < a$ 的某處與底面相交，而流體的橫截面是一個直角三角形，其兩條直角邊分別為沿底面的 b 和沿下方側壁向上的 h。由液面的斜率得 $h = b\tan\theta$，由面積得 $\frac{1}{2}bh = ka^2$。所以 $b^2\tan\theta = 2ka^2$，於是

$$b = a\sqrt{\frac{2k}{\tan\theta}}, \qquad h = a\sqrt{2k\tan\theta}.$$

（事實上，$b < a$ 正是因為 $2k < \tan\theta$；容器很高，所以 h 不會造成問題。）三角形的形心位於每條直角邊的三分之一處，所以

$$\frac{x}{a} = \frac{1}{3}\sqrt{\frac{2k}{\tan\theta}} = \frac{1}{3}\sqrt{2k\cot\theta}, \qquad \frac{y}{a} = \frac{1}{3}\sqrt{2k\tan\theta}.$$

當 $k = \frac{1}{2}\tan\theta$ 時，兩組公式都給出 $x/a = \frac{1}{3}$ 和 $y/a = \frac{1}{3}\tan\theta$，所以兩種情況互相銜接。

傾倒

想一想 容器在放手的一刻會否翻倒，取決於甚麼？

容器是輕的，所以唯一的重量就是流體的重量，它鉛垂地作用於流體的質心 G。斜面只能對底面施力，所以法向反作用力（normal reaction）作用於底面上某處，即下方邊緣 O 與上方邊緣之間。假如通過 G 的鉛垂線與斜面相交於 O 的下坡一側，重量對 O 便有一個力矩（moment），而這樣的反作用力無法把它平衡，於是容器繞 O 轉動：它會傾倒。

取 O 為原點，x 沿斜面向上，y 垂直於斜面。在這組坐標軸中，鉛垂向下的方向的分量為 $(-\sin\theta, -\cos\theta)$，所以由 $G = (x, y)$ 沿鉛垂線向下走一段距離 $y\sec\theta$，便到達斜面上的

$$x - y\tan\theta.$$

所以若 $x < y\tan\theta$，容器便會傾倒。

假設 $k < \frac{1}{2}$ 且 $\theta > 45°$。那麼 $\tan\theta > 1 > 2k$，所以 $k < \frac{1}{2}\tan\theta$，我們處於三角形的情況。此時

$$y\tan\theta - x = \frac{a}{3}\left(\sqrt{2k\tan\theta}\,\tan\theta - \sqrt{\frac{2k}{\tan\theta}}\right) = \frac{a}{3}\sqrt{\frac{2k}{\tan\theta}}\left(\tan^2\theta - 1\right) > 0,$$

因為 $\tan\theta > 1$。因此 $x < y\tan\theta$：通過 G 的鉛垂線落在下方邊緣之下，無論 $k < \frac{1}{2}$ 取甚麼值，容器都會傾倒。（在三角形的情況中，同樣的計算顯示若 $\theta < 45°$ 則容器是穩定的，所以 $45°$ 正是該情況下的臨界值。）這就是所要的結果。條件 $k < \frac{1}{2}$ 的作用，是保證當 $\theta > 45°$ 時必定處於三角形的情況。

驗算

- 當 $\theta = 0$ 時，第一種情況給出 $x = \frac{1}{2}a$ 和 $y = \frac{1}{2}ka$，即一塊深度為 ka 的流體的中心。
- 兩種情況在 $k = \frac{1}{2}\tan\theta$ 時吻合，此時梯形退化成三角形。
- 數值驗算：對 300 組隨機的 (k,θ) 直接對流體區域進行積分，結果與兩組公式的誤差都在 1.4×10^{-6} 之內（取 $a = 1$）。例如，$\theta = 55°$、$k = 0.3$ 給出 $x \approx 0.216a$、$y \approx 0.309a$，而 $y\tan\theta \approx 0.441a > x$，所以容器會傾倒。

常見錯誤

- 把液面畫成與底面平行，或者斜向錯誤的一邊：流體是堆向下方側壁的。
- 忘記容器是輕的，而為容器另加一個重量：整體的質心就是流體的質心。
- 把通過 G 的鉛垂線與底面的中心比較，而不是與下方邊緣 O 比較。

STEP 之外

「通過質心的鉛垂線是否與底面相交？」這個判別法，是穩定性分析（stability analysis）中最簡單的情形。浮在水面的船隻也是用同樣的方法處理，不過在那裏浮力（buoyancy）作用於被排開的水的質心，而這個點會隨船身傾側而移動；由此便引出定傾中心（meta-centre）的概念。流動中的流體（放手後的晃盪）則是流體動力學（fluid dynamics）中困難得多的問題。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）對剛體平衡（rigid-body equilibrium）及繞軸轉動有細緻的論述。

詳細解答

STEP III 2002，第 10 題

力學 • 涉及摩擦力的轉動

STEP III 2002，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/02-s3-q10

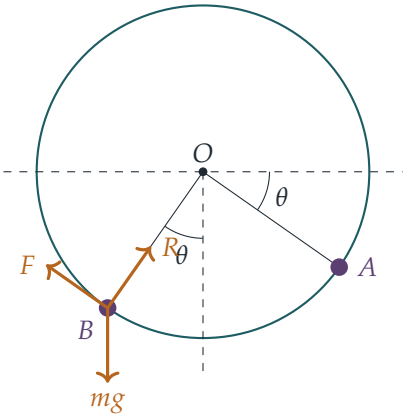
關鍵思路

只要第二個質點（particle）不滑動，它便隨圓柱一起運動，於是整個系統是一個繞固定軸轉動的剛體（rigid body）：兩個相等的質量位於半徑 a 處，彼此相隔四分之一圈。這樣，用能量（或對轉軸取力矩（moments））一行便可得出運動方程。要找出何時開始滑動，我們單獨考慮那個沒有固定的質點，沿表面方向和垂直於表面方向應用牛頓第二定律（Newton's second law），求出它所需的摩擦力（friction）F 和法向反作用力（normal reaction）R。當 F 首次達到 μR 時，它便開始滑動。

解答

建立模型

把兩個質點都視為輕圓柱上的點質量，輪軸光滑。設 O 為轉軸，A 為固定的質點，B 為沒有固定的質點。系統由靜止釋放後，A 下落，圓柱因而轉動（在圖中為順時針方向），並把 B 帶上遠離 A 的一側。轉過角度 θ 後，OA 在水平線以下成角 θ，而 OB 在另一側與向下的鉛垂線成角 θ。



運動方程

只要 B 不滑動，兩個質點都在半徑為 a 的圓上以角速率（angular speed）$\dot{\theta}$ 運動，而圓柱本身沒有質量。取 O 所在的水平面為勢能（potential energy）零點，則 A 的高度為 $-a\sin\theta$，B 的高度為 $-a\cos\theta$。系統在 $\theta = 0$ 時由靜止開始運動，光滑輪軸不作功，而 B 與圓柱之間沒有相對運動，所以摩擦力所作的淨功為零。因此能量守恆：

$$2\cdot\tfrac{1}{2}ma^2\dot{\theta}^2 - mga\sin\theta - mga\cos\theta = -mga,$$

即

$$\dot{\theta}^2 = \frac{g}{a}(\sin\theta + \cos\theta - 1). \qquad (*)$$

對時間求導，再除以 $2\dot{\theta}$（運動一旦開始，它便不等於零），得

$$\ddot{\theta} = \frac{g}{2a}(\cos\theta - \sin\theta),$$

這正是所要求的結果。另一個等價的做法是對 O 取力矩：A 的重量沿轉動方向的力矩為 $mga\cos\theta$，B 的重量逆轉動方向的力矩為 $mga\sin\theta$，而轉動慣量（moment of inertia）為 $2ma^2$。

B 何時滑動？

現在單獨考慮 B。作用於它的力有：重量 mg、指向 O 的法向反作用力 R，以及沿表面的摩擦力 F；我們取摩擦力沿 B 的運動方向（即沿側面向上）為正。B 在半徑為 a 的圓上運動，所以它有指向 O 的加速度 $a\dot{\theta}^2$ 和沿運動方向的加速度 $a\ddot{\theta}$。重量有一個大小為 $mg\cos\theta$、方向正背離 O 的分量，以及一個大小為 $mg\sin\theta$、與運動方向相反的分量。因此

$$\text{指向}O:\quad R - mg\cos\theta = ma\dot{\theta}^2,$$
$$\text{沿運動方向}:\quad F - mg\sin\theta = ma\ddot{\theta}.$$

利用 $(*)$ 和運動方程，

$$R = mg\cos\theta + mg(\sin\theta + \cos\theta - 1) = mg\,(2\cos\theta + \sin\theta - 1),$$
$$F = mg\sin\theta + \tfrac{1}{2}mg(\cos\theta - \sin\theta) = \tfrac{1}{2}mg\,(\sin\theta + \cos\theta).$$

當 $0 \le \theta \le \frac{\pi}{3}$ 時，兩者皆為正。（$R > 0$ 是因為當 $0 < \theta \le \frac{\pi}{3}$ 時，利用 $2\cos\theta \ge 1$ 可得 $2\cos\theta + \sin\theta - 1 \ge \sin\theta > 0$，而在 $\theta = 0$ 時 $R = mg$。）所以 B 保持接觸，而摩擦力把它沿側面向上推。只要 $F \le \mu R$，質點便不會滑動，也就是說，只要

$$\rho(\theta) = \frac{F}{R} = \frac{\sin\theta + \cos\theta}{2(2\cos\theta + \sin\theta - 1)} \le \mu.$$

在 $\theta = \frac{\pi}{3}$ 時，$\sin\theta = \frac{\sqrt{3}}{2}$ 而 $\cos\theta = \frac{1}{2}$，於是

$$\rho\left(\tfrac{\pi}{3}\right) = \frac{\frac{1}{2}(\sqrt{3}+1)}{2\cdot\frac{\sqrt{3}}{2}} = \frac{\sqrt{3}+1}{2\sqrt{3}} = \frac{3+\sqrt{3}}{6} = \mu.$$

想一想 為甚麼質點不可能在此之前已經開始滑動？

我們需要對所有 $0 \le \theta < \frac{\pi}{3}$ 都有 $\rho(\theta) < \mu$。記 $N = \sin\theta + \cos\theta$ 和 $D = 2\cos\theta + \sin\theta - 1$。則

$$\begin{aligned} N'D - ND' &= (\cos\theta - \sin\theta)(2\cos\theta + \sin\theta - 1) - (\sin\theta + \cos\theta)(\cos\theta - 2\sin\theta) \\ &= \cos^2\theta + \sin^2\theta + \sin\theta - \cos\theta = 1 + \sin\theta - \cos\theta, \end{aligned}$$

它在 $0 < \theta < \pi$ 時為正。由於在 $[0, \frac{\pi}{3}]$ 上 $D > 0$，故在該區間上 $\rho' = (N'D - ND')/(2D^2) > 0$，所以 ρ 由 $\rho(0) = \frac{1}{2}$ 起嚴格遞增。因此它一直小於 μ，直至在 $\theta = \frac{\pi}{3}$ 時恰好達到 μ。此外，由 $(*)$ 可知，當 $0 < \theta < \frac{\pi}{2}$ 時 $\dot\theta^2 > 0$，所以圓柱確實會一直轉動並到達這個角度。

所以，當圓柱轉過 $60°$ 時，第二個質點開始滑動，這正是所要求的結果。

驗算

- 在 $\theta = 0$ 時：B 靜止於最低點，但 $F = \frac{1}{2}mg \ne 0$。這是對的：A 的重量正試圖轉動圓柱，而摩擦力必須令 B 隨之加速。事實上，此時 $F = ma\ddot\theta = \frac{1}{2}mg$，而 $R = mg$，因為 B 還未有速率。
- 在 $\theta = \frac{\pi}{4}$ 時，$\ddot\theta = 0$，角速率達到最大值；兩個重量對 O 的力矩互相平衡時，正應如此。
- 以數值方法驗證：對運動方程作數值積分，並直接由 B 的加速度計算作用於它的接觸力，所得結果與 R 和 F 的公式吻合，而 F/R 首次達到 $(3+\sqrt{3})/6 \approx 0.7887$ 是在 $60.00°$。

常見錯誤

- 把圓柱的轉動慣量也計算在內：圓柱是輕的，所以只有兩個質點貢獻 $2ma^2$。
- 忘記了 B 的切向加速度（tangential acceleration）$a\ddot{\theta}$，這會得出 $F = mg\sin\theta$ 和錯誤的角度。
- 只驗證在 $60°$ 時 $F = \mu R$，卻沒有證明在此之前 F/R 較小。

STEP 之外

這是一個單自由度（one-degree-of-freedom）系統：一旦加上約束（constraint）條件（不滑動），單一個坐標 θ 便足以描述一切，而能量可直接給出運動。約束力是事後才求出的，而當約束所需的力超過表面所能提供的力時，約束便會失效。這與質點離開球面的問題是同一套邏輯，而在拉格朗日力學（Lagrangian mechanics）中，這套方法藉拉格朗日乘數（Lagrange multipliers）而變得有系統。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）正是以這種精神闡述剛體轉動和約束力。

詳細解答

STEP III 2002，第 11 題

力學 • 與牆壁的斜碰撞

STEP III 2002，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/02-s3-q11

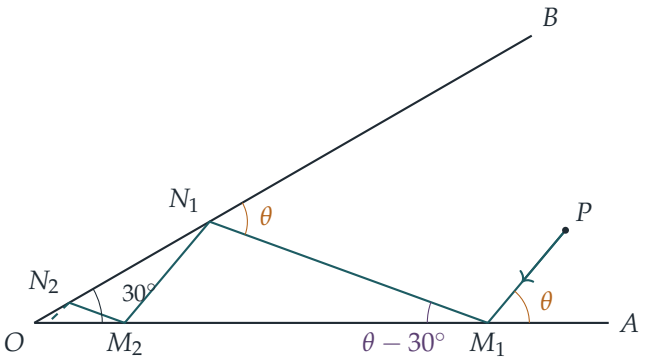
本題在 *Beyond STEP* 第 10.5 節有限時間內的無限次反彈中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch10/zeno-wedge.html

關鍵思路

在光滑（smooth）的牆上，衝量（impulse）垂直於牆。因此沿牆方向的速度分量保持不變，只有垂直方向的分量乘以 e。把這一點與三角形 OM_1N_1 的各個角結合起來，第一部分便化為一條三角方程。至於第二部分，每次反彈之後，運動都會以較小的規模重演一次。各段路程愈來愈短，但質點（particle）同時也在減慢，所以我們要相加的是時間，而不是距離。這個和是一個等比級數（geometric series）。

解答

我們把質點視為在光滑水平面上運動的一點，因此在兩次碰撞之間，它以恆定速率沿直線運動。兩面牆都是光滑的，而每次碰撞都遵從牛頓實驗定律（Newton's experimental law），其系數為 e。



求 e 及 v_1/u_1

由於角 $PM_1A = \theta$，質點到達 M_1 時正朝着 O 運動，與牆成角 θ。碰撞前的一刻，它的速度分量為

$$u_1\cos\theta \text{ 沿牆方向（朝向 } O\text{）}, \qquad u_1\sin\theta \text{ 垂直於牆（指向牆內）}.$$

牆是光滑的，所以質點所受的衝量垂直於 OA。因此：

$$\begin{aligned}&\text{沿牆方向：} u_1\cos\theta\text{（不變）},\\ &\text{垂直方向：} e\,u_1\sin\theta\text{（方向逆轉，並乘以 } e\text{）}.\end{aligned}$$

所以質點離開 M_1 時仍然朝着 O 運動，與牆成角 φ，其中

$$\tan\varphi = \frac{e\,u_1\sin\theta}{u_1\cos\theta} = e\tan\theta, \qquad v_1\cos\varphi = u_1\cos\theta. \tag{$*$}$$

接着我們利用三角形 OM_1N_1 的幾何關係求 φ。O、N_1、B 三點共線，而角 $M_1N_1B = \theta$，所以角 $ON_1M_1 = 180° - \theta$。位於 O 的角是 $30°$，故

$$\varphi = \angle OM_1N_1 = 180° - 30° - (180° - \theta) = \theta - 30°.$$

要 φ 為正，必須有 $\theta > 30°$，這是本題的佈局所隱含的假設。於是由 $(*)$ 得

$$\boxed{e = \frac{\tan(\theta - 30°)}{\tan\theta}}$$

以及

$$\boxed{\frac{v_1}{u_1} = \frac{\cos\theta}{\cos(\theta - 30°)} = \frac{2}{\sqrt{3} + \tan\theta}},$$

其中最後一個形式來自展開 $\cos(\theta - 30°) = \frac{\sqrt{3}}{2}\cos\theta + \frac{1}{2}\sin\theta$。你也可以寫成 $e = \dfrac{\sqrt{3}\tan\theta - 1}{\tan\theta\,(\sqrt{3} + \tan\theta)}$。由於 $0 < \theta - 30° < \theta < 90°$，我們得到 $0 < e < 1$，符合題目的要求。

在有限時間內到達 O

質點到達 N_1 時正朝着 O 運動，與牆 OB 成角 θ（因為角 $M_1N_1B = \theta$）。這與在 M_1 的情況完全相同，只是兩面牆的角色互換了。所以同樣的計算仍然適用：它反彈後與 N_1O 成角 $\theta - 30°$，而速率則乘以

$$r = \frac{\cos\theta}{\cos(\theta - 30°)}.$$

想一想 質點到達 M_2 時是否又再成角 θ，使這個規律真的重複下去？

是的。在三角形 ON_1M_2 中，位於 O 的角是 $30°$，位於 N_1 的角是 $\theta - 30°$。所以角 $OM_2N_1 = 180° - \theta$，即角 $N_1M_2A = \theta$。由歸納法（induction）可知，每次碰撞都是與牆成角 θ 發生的。每一段路程都以角 $\theta - 30°$ 離開牆。而每個三角形 OM_nN_n、ON_nM_{n+1} 的三個角都是 $30°$、$\theta - 30°$、$180° - \theta$，所以它們全部相似（similar）。

長度。在三角形 OM_1N_1 中應用正弦定理（sine rule），

$$\frac{ON_1}{\sin(\theta - 30°)} = \frac{OM_1}{\sin(180° - \theta)} = \frac{M_1N_1}{\sin 30°}.$$

所以每個新三角形都是前一個三角形按以下比例縮放而成：

$$k = \frac{ON_1}{OM_1} = \frac{\sin(\theta - 30°)}{\sin\theta}.$$

把各段路程 M_1N_1, N_1M_2, M_2N_2,... 稱為第 1,2,3,... 段。第 n 段的長度為 $\ell_1 k^{n-1}$，其中 $\ell_1 = M_1N_1 = \dfrac{OM_1}{2\sin\theta}$。

速率。質點以恆定速率 $v_1 r^{n-1}$ 走過第 n 段。

時間。因此第 n 段所需的時間為

$$\tau_n = \frac{\ell_1 k^{n-1}}{v_1 r^{n-1}} = \frac{\ell_1}{v_1} q^{n-1}, \qquad q = \frac{k}{r} = \frac{\sin(\theta - 30°)\cos(\theta - 30°)}{\sin\theta\cos\theta} = \frac{\sin(2\theta - 60°)}{\sin 2\theta}.$$

想一想 各段路程按等比縮短，但質點同時也在減慢。為甚麼單憑這一點還不足以下結論？

我們要相加的是時間，而時間之比 q 所比較的，正是路程的縮短（k）與質點的減慢（r）。我們需要 $q < 1$。現在假設 $30° < \theta < 60°$。那麼

$$0° < 2\theta - 60° < 60° \;\Rightarrow\; \sin(2\theta - 60°) < \tfrac{\sqrt{3}}{2}, \qquad 60° < 2\theta < 120° \;\Rightarrow\; \sin 2\theta > \tfrac{\sqrt{3}}{2},$$

所以 $0 < q < 1$。於是總時間是一個收斂（convergent）的等比級數：

$$T = \frac{PM_1}{u_1} + \sum_{n=1}^{\infty} \frac{\ell_1}{v_1} q^{n-1} = \frac{PM_1}{u_1} + \frac{\ell_1}{v_1(1-q)}.$$

這是有限的。要化簡，可利用 $\sin 2\theta - \sin(2\theta - 60^\circ) = 2\cos(2\theta - 30^\circ)\sin 30^\circ$，由此得 $1 - q = \dfrac{\cos(2\theta - 30^\circ)}{\sin 2\theta}$。於是由 M_1 起計的時間為

$$\frac{\ell_1}{v_1(1-q)} = \frac{OM_1\cos(\theta - 30^\circ)}{u_1\cos(2\theta - 30^\circ)},$$

當 $30^\circ < \theta < 60^\circ$ 時，它是正數而且有限。

最後，質點確實會到達 O。第 n 次碰撞之後，它與 O 的距離是 $OM_1k^n \to 0$，因為 $0 < k < 1$。在走每一段路程期間，它與 O 的距離不會超過該段兩個端點的距離中較大的一個。所以當 $t \to T$ 時，質點與 O 的距離趨向 0：它經過無限多次碰撞之後，在有限時間 T 到達 O。 ■

驗算

- 當 $\theta = 45^\circ$：$e = \tan 15^\circ = 2 - \sqrt{3} \approx 0.268$，$v_1/u_1 = \sqrt{3} - 1 \approx 0.732$，而 $q = \sin 30^\circ / \sin 90^\circ = \frac{1}{2}$，所以每一段所需的時間都是前一段的一半。取 $OM_1 = u_1 = 1$，由 M_1 到 O 的時間是 $\cos 15^\circ / \cos 60^\circ = 2\cos 15^\circ \approx 1.932$。逐次模擬各次碰撞（六十次反彈）得到 1.931852，兩者吻合。
- 當 $\theta = 60^\circ$：$q = 1$，所以每一段所需的時間都相同，總時間是無限的。當 $\theta > 60^\circ$ 時，$q > 1$。所以 60° 正好是分界線，而 T 的公式在該處因 $\cos(2\theta - 30^\circ) = \cos 90^\circ = 0$ 而趨向無限大。
- 當 $\theta \to 30^{\circ+}$ 時，$e \to 0$：質點幾乎不反彈，差不多筆直地滑向 O，正如你所預期的。

常見錯誤

- 把三角形弄錯：角 M_1N_1B 是位於 N_1 的外角（exterior angle），所以內角是 $180^\circ - \theta$，而在 M_1 的反彈角是 $\theta - 30^\circ$，不是 $\theta + 30^\circ$。
- 只因為距離 $OM_n \to 0$ 便斷言質點到達 O。無限多段路程需要一個收斂的時間級數，而質點在每次碰撞時都會減慢。
- 在碰撞時把切向分量（tangential component）逆轉。在光滑的牆上，只有法向分量（normal component）會改變。

STEP 之外

在有限時間內發生無限多次碰撞，稱為芝諾現象（Zeno behaviour）。大家熟悉的例子是一個以 $e < 1$ 在地板上彈跳的球，而這裏的楔形就扮演了地板的角色。研究「混合」系

統（hybrid systems，即把平滑運動與突然跳變混合起來的系統）的數學家，必須決定在芝諾時間之後會發生甚麼事。在顆粒物料中，同樣的效應稱為非彈性坍縮（inelastic collapse），會在模擬中造成實際的麻煩。若想看以大學一年級程度仔細處理碰撞與衝量的論述，可參閱 Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門）。

詳細解答

STEP III 2002，第 12 題

概率與統計 • 以條件化求期望值

> **STEP III 2002，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/02-s3-q12

關鍵思路

按遊戲如何開始來取條件（conditioning）。每一局遊戲的開頭不外乎三種：H、TH 或 TT。在前兩種情況下，並沒有取得任何有用的進展（連續反面已被打斷），所以遊戲等於從頭再來，而餘下部分的期望長度仍然是 E。這種「更新」（renewal）給出一條關於 E 的線性方程，解之即可。同樣的想法——按第一個正面何時出現來分類——也能處理連續 r 次反面的情況。

解答

全文中，p 是擲出正面的概率，$q = 1 - p$ 是擲出反面的概率，各次投擲互相獨立（independent）。我們假設 $0 < p < 1$，這樣遊戲以概率 1 結束，而且下文所有期望值（expectation）都是有限的。

解釋關於 E 的方程

考慮最初的一兩次投擲。以下三種情況恰好有一種發生：

- **先擲出 H**（概率為 p）。已用去一次投擲，而且沒有正在累積的連續反面。由於各次投擲互相獨立，遊戲餘下的部分就是整個遊戲的一個全新複本，期望長度為 E。期望總長度：$1 + E$。
- **先 T 後 H**（概率為 qp）。已用去兩次投擲，而正面再次把連續反面清零，所以遊戲重新開始。期望總長度：$2 + E$。
- **先 T 後 T**（概率為 q^2）。遊戲恰好在 2 次投擲後結束。

這三種情況互斥（exclusive）而且窮盡（exhaustive）所有可能（它們的概率之和為 $p + qp + q^2 = p + q = 1$）。根據全期望定律（law of total expectation），

$$E = p(1 + E) + (1 - p)p(2 + E) + 2(1 - p)^2,$$

這正是題目給出的方程。

求 E

把含 E 的項合併：

$$E\,(1 - p - qp) = p + 2qp + 2q^2.$$

左邊，$1 - p - qp = q - qp = q(1 - p) = q^2$。右邊，$p + 2q(p + q) = p + 2q = 1 + q$。由於 $q \neq 0$，

$$E = \frac{1 + q}{q^2} = \frac{2 - p}{(1 - p)^2}.$$

期望獎金

設 W 為正面次數的期望值，亦即以英鎊計的期望獎金。沿用同樣的三種情況。擲出 H 之後，我們已贏得 £1，遊戲重新開始；擲出 TH 之後，我們同樣已贏得 £1，遊戲重新開始；擲出 TT 之後，我們一無所獲，遊戲結束。所以

$$W = p(1 + W) + qp(1 + W) + q^2 \cdot 0.$$

? **想一想** 在 TH 的情況中，為甚麼是 $1 + W$ 而不是 W？

在這種情況下，第二次投擲*確實*是正面，所以在遊戲重新開始之前已經賺得 £1。這一題最容易出錯的地方，就是漏掉了它。

合併各項，得 $W(1 - p - qp) = p(1 + q)$，所以 $Wq^2 = p(1 + q)$，而

$$W = \frac{p(1 + q)}{q^2} = \frac{p(2 - p)}{(1 - p)^2} \text{ 英鎊}.$$

留意 $W = pE$：平均而言，所有投擲中有 p 這個比例是正面，即使遊戲停止的時刻取決於投擲的結果。

連續 r 次反面的遊戲

現在設 E 為遊戲在連續出現 r 次反面時停止的情況下，投擲次數的期望值。按第一個正面來取條件。要麼最初 r 次投擲全是反面，要麼第一個正面出現在第 $j + 1$ 次投擲，其中 j 滿足 $0 \leq j \leq r - 1$：

- j **次反面後一次正面**（概率為 $q^j p$）。這用去了 $j+1$ 次投擲，連續反面已被打斷，遊戲重新開始。期望總長度：$j+1+E$。
- r **次反面**（概率為 q^r）。遊戲恰好在 r 次投擲後結束。

所以

$$E = \sum_{j=0}^{r-1} q^j p\,(j+1+E) + rq^r.$$

右邊 E 的係數是 $p\sum_{j=0}^{r-1} q^j = p\cdot\dfrac{1-q^r}{1-q} = 1-q^r$，因為 $1-q=p$。把它移到左邊，

$$q^r E = p\sum_{j=0}^{r-1}(j+1)q^j + rq^r. \qquad (*)$$

> **想一想** 沒有公式表在手，怎樣求 $\sum(j+1)q^j$ 的和？

記 $S=\sum_{j=0}^{r-1}(j+1)q^j = 1+2q+3q^2+\cdots+rq^{r-1}$，然後減去 $qS = q+2q^2+\cdots+(r-1)q^{r-1}+rq^r$：

$$(1-q)S = 1+q+q^2+\cdots+q^{r-1}-rq^r = \frac{1-q^r}{p} - rq^r.$$

因此 $pS = \dfrac{1-q^r}{p} - rq^r$，代入 $(*)$ 後，rq^r 兩項互相抵消：

$$q^r E = \frac{1-q^r}{p} - rq^r + rq^r = \frac{1-q^r}{p}.$$

於是

$$E = \frac{1-q^r}{pq^r},$$

這就是所要證明的結果。當 $r=2$ 時，上式為 $\dfrac{(1-q)(1+q)}{pq^2} = \dfrac{1+q}{q^2}$，與第一個遊戲的結果一致。

驗算

- 若硬幣是公平的，$p=q=\frac{1}{2}$：第一個遊戲平均持續 $E=\frac{3/2}{1/4}=6$ 次投擲，贏得 $W=$£3，正好是 E 的一半，合乎預期。
- $r=1$：遊戲在第一次出現反面時停止，所以長度服從幾何分佈（geometric distribution），平均值為 $1/q$；公式給出 $\dfrac{1-q}{pq} = \dfrac{1}{q}$。
- 當 $q\to 0$（幾乎每次投擲都是正面）時，E 和 W 都趨向無窮大，這是理所當然的。

- 我們以分數作精確計算，對 $r = 1, \ldots, 7$ 及若干個 p 值解出「目前已連續出現 k 次反面」這一串狀態的方程，確認了這些公式。另外，以 $p = 0.4$ 模擬 2×10^5 局遊戲，在 $r = 2$ 時得出平均 4.438 次投擲（理論值 4.444）及獎金 £1.772（理論值 £1.778），在 $r = 3$ 時得出 9.076 次投擲（理論值 9.074）。

常見錯誤

- 在獎金的方程中，把 $qp(1 + W)$ 寫成 qpW，結果得出錯誤答案 p/q^2。
- 只按第一次投擲來取條件：擲出一次 T 之後，遊戲並不會重新開始，因為連續反面中的一次已經到手。這正是題目所給方程要往前看兩次投擲的原因。

STEP 之外

「一旦出現正面，未來的情形就與開局時完全一樣」這個論證，是馬可夫鏈（Markov chain）的*更新法*或*首步分析法*（first-step method），一般情況下的期望到達時間（expected hitting time）正是這樣計算的。$W = pE$ 這個簡潔的事實是沃爾德恆等式（Wald's identity）的一個特例：把隨機個數 N 個獨立同分佈（independent, identically distributed）的項相加，其和的期望值等於 $\mathbb{E}[N]$ 乘以單獨一項的平均值，前提是 N 為停時（stopping time）。這兩個課題都可以在 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門），或 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論導論）中讀到。

詳細解答

STEP III 2002，第 13 題

概率與統計 • 指數分佈變量的最小值

STEP III 2002，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/02-s3-q13

關鍵思路

「至少有一個小於 x」是「兩個都不小於 x」的補事件（complement），而後者的概率可憑獨立性（independence）分解成乘積。指數分佈（exponential distribution）的尾概率非常簡單，$\mathrm{P}(X_i \geq x) = \mathrm{e}^{-\lambda_i x}$，所以乘積仍然是一個指數尾概率，而關於最小值的證明就僅此而已。巴士一問則是直接應用：等候下一班巴士的時間，就是兩個指數間隔之中較小的一個。

解答

其中一個或兩個都小於 x 的概率

先求單一指數變量的尾概率。對於 $x \geq 0$，

$$\mathrm{P}(X_i \geq x) = \int_x^\infty \lambda_i \mathrm{e}^{-\lambda_i t}\,\mathrm{d}t = \left[-\mathrm{e}^{-\lambda_i t}\right]_x^\infty = \mathrm{e}^{-\lambda_i x}.$$

設 E 為 $X_1 < x$ 或 $X_2 < x$（或兩者皆是）這個事件。它的補事件是 $X_1 \geq x$ 與 $X_2 \geq x$ 同時成立。由於 X_1 與 X_2 獨立，

$$\mathrm{P}(X_1 \geq x \text{ 且 } X_2 \geq x) = \mathrm{P}(X_1 \geq x)\,\mathrm{P}(X_2 \geq x) = \mathrm{e}^{-\lambda_1 x}\mathrm{e}^{-\lambda_2 x},$$

所以，對於 $x \geq 0$，

$$\boxed{\mathrm{P}(E) = 1 - \mathrm{e}^{-(\lambda_1+\lambda_2)x}.}$$

（對於 $x < 0$，這個概率是 0，因為兩個變量都不可能取負值。）

? **想一想** 為甚麼不直接把 $\mathrm{P}(X_1 < x)$ 與 $\mathrm{P}(X_2 < x)$ 相加？

因為這兩個事件有重疊：兩個變量可以同時小於 x，相加便會把這種情況計算兩次。你也可以用容斥原理（inclusion–exclusion）來做，$(1-\mathrm{e}^{-\lambda_1 x})+(1-\mathrm{e}^{-\lambda_2 x})-(1-\mathrm{e}^{-\lambda_1 x})(1-\mathrm{e}^{-\lambda_2 x})$，化簡後得到相同的答案，但用補事件會快捷得多。

最小值服從指數分佈

設 $X=\min(X_1,X_2)$。兩個數中較小的一個小於 x，當且僅當其中至少有一個小於 x，所以 $\{X<x\}$ 就是上面的事件 E。因此 X 的分佈函數（distribution function）是

$$F(x)=\mathrm{P}(X\le x)=1-\mathrm{e}^{-(\lambda_1+\lambda_2)x}\qquad (x\ge 0),$$

而當 $x<0$ 時 $F(x)=0$。（對於連續變量，$\mathrm{P}(X<x)=\mathrm{P}(X\le x)$。）求導後，$X$ 的密度函數（density）是

$$f(x)=F'(x)=(\lambda_1+\lambda_2)\,\mathrm{e}^{-(\lambda_1+\lambda_2)x}\qquad (0\le x<\infty),$$

這正是參數為 $\lambda_1+\lambda_2$ 的指數密度函數。所以 X 服從指數分佈，這正是所要證明的。

巴士

參數為 λ 的指數分佈，其平均值為 $\int_0^\infty t\lambda\mathrm{e}^{-\lambda t}\,\mathrm{d}t=1/\lambda$（用分部積分法）。以分鐘為時間單位，平均間隔 15 分鐘和 30 分鐘所對應的參數是

$$\lambda_A=\tfrac{1}{15},\qquad \lambda_B=\tfrac{1}{30}.$$

兩條路線各有一班巴士剛剛開出，所以距離下一班 A 線巴士的時間，就是 A 線巴士之間一個完整的間隔 T_A，B 線同理為 T_B。兩者互相獨立，分別服從參數為 λ_A 和 λ_B 的指數分佈。兩條路線都可以送我到學院，所以我的等候時間是 $W=\min(T_A,T_B)$，而根據剛才證明的結果，W 服從指數分佈，參數為

$$\lambda_A+\lambda_B=\tfrac{1}{15}+\tfrac{1}{30}=\tfrac{1}{10}.$$

中位數（median）m 滿足 $\mathrm{P}(W\le m)=\frac{1}{2}$：

$$1-\mathrm{e}^{-m/10}=\tfrac{1}{2}\iff \mathrm{e}^{-m/10}=\tfrac{1}{2}\iff m=10\ln 2.$$

由於 $\ln 2\approx 0.693$，等候時間的中位數是 $m\approx 6.93$ 分鐘，即大約 7 分鐘，與題目要求相符。

驗算

- 若 $\lambda_2\to 0$（B 線巴士幾乎永遠不來），答案變成 $1-\mathrm{e}^{-\lambda_1 x}$，正好是 X_1 的分佈，合乎預期。

- 平均等候時間是 10 分鐘，比兩條路線各自的平均值 15 或 30 分鐘都要短；而中位數 $10\ln 2$ 低於平均值，這是右偏分佈（right-skewed distribution）的常見現象。
- 模擬 2×10^5 對平均值分別為 15 和 30 的獨立指數間隔，所得最小值的中位數約為 6.91 分鐘，而 $10\ln 2 = 6.931\ldots$。

常見錯誤

- 把平均值當作參數：平均值為 15 的指數分佈，其 $\lambda = \frac{1}{15}$，而不是 15。
- 把兩個概率相加，即 $\mathrm{P}(X_1 < x) + \mathrm{P}(X_2 < x)$，這樣會重複計算，而且結果可能超過 1。
- 取兩條路線中位數的平均，或只取 A 線的中位數（$15\ln 2 \approx 10.4$）：有兩條路線可選，等候時間會縮短。

STEP 之外

巴士之間的間隔服從指數分佈，意味着巴士的到達構成一個泊松過程（Poisson process），而這裏的結果說明：把兩個獨立的泊松過程合併，會得到另一個泊松過程，其速率是兩個速率之和。這個結果連同無記憶性（memoryless property）$\mathrm{P}(X > s+t \mid X > s) = \mathrm{P}(X > t)$，是排隊論（queueing theory）和連續時間馬可夫鏈（continuous-time Markov chains）的基礎，在大學一年級的概率課程中便會學到。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率入門教材）涵蓋了指數變量和泊松過程。

詳細解答

STEP III 2002，第 14 題

概率與統計 • 方差與協方差

STEP III 2002，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/02-s3-q14

關鍵思路

這條方差恆等式其實就是方差（variance）的定義，再利用期望值的線性性質（linearity of expectation）展開而已。國際象棋那部分比表面看來更巧妙：我們從頭到尾都沒有直接計算 $\mathrm{E}(WL)$。相反，我們由二項分佈（binomial distribution）得知 $W+L$ 的方差，同樣由二項分佈得知 $\mathrm{Var}(W)$ 和 $\mathrm{Var}(L)$，而剛證明的恆等式便把協方差（covariance）白白送到我們手上。

解答

方差恆等式

記 $\mu_X = \mathrm{E}(X)$ 及 $\mu_Y = \mathrm{E}(Y)$。我們採用以下定義

$$\mathrm{Var}(X) = \mathrm{E}\big[(X-\mu_X)^2\big], \qquad \mathrm{Cov}(X,Y) = \mathrm{E}\big[(X-\mu_X)(Y-\mu_Y)\big],$$

以及期望值的線性性質 $\mathrm{E}(aU+bV) = a\,\mathrm{E}(U) + b\,\mathrm{E}(V)$；對於離散隨機變量（discrete random variable），只要把和式 $\sum (au+bv)\,\mathrm{P}(U=u, V=v)$ 拆成兩個和式，並在每個和式中對另一個變量求和，便可得出這個性質。

根據線性性質，$\mathrm{E}(X+Y) = \mu_X + \mu_Y$，所以

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}\big[\big((X-\mu_X)+(Y-\mu_Y)\big)^2\big] \\
&= \mathrm{E}\big[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)\big] \\
&= \mathrm{E}\big[(X-\mu_X)^2\big] + \mathrm{E}\big[(Y-\mu_Y)^2\big] + 2\,\mathrm{E}\big[(X-\mu_X)(Y-\mu_Y)\big] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y),
\end{aligned}$$

其中第三行再次用到線性性質。這就是所需的結果。留意我們並不需要獨立性(independence)：這條恆等式對任何 X 和 Y 都成立。

第 (i) 部分

每一局棋不是勝、就是負、或者和，所以無論結果如何，都有 $W+L+D=m$。一個永遠取同一個值的隨機變量等於它的平均值，所以它的方差是 $\mathrm{E}\big[(m-m)^2\big]=0$。

第 (ii) 部分

如果一局棋分出了勝負（即勝或負），我們稱它為分勝負的一局。每一局分勝負的概率是 $\frac{1}{3}+\frac{1}{3}=\frac{2}{3}$，各局互相獨立，而且局數固定為 m。因此 $W+L$，即分勝負的局數，就是在 m 次成功概率相同的獨立試驗中成功的次數：

$$W+L\sim \mathrm{B}\big(m,\tfrac{2}{3}\big).$$

相關係數

同樣道理，W 是 m 次獨立試驗中勝出的次數，每次勝出的概率是 $\frac{1}{3}$，所以 $W\sim \mathrm{B}(m,\frac{1}{3})$，同理 $L\sim \mathrm{B}(m,\frac{1}{3})$。利用二項分佈的公式 $\mathrm{Var}=np(1-p)$，

$$\mathrm{Var}(W)=\mathrm{Var}(L)=m\cdot\tfrac{1}{3}\cdot\tfrac{2}{3}=\frac{2m}{9},\qquad \mathrm{Var}(W+L)=m\cdot\tfrac{2}{3}\cdot\tfrac{1}{3}=\frac{2m}{9}.$$

> ? **想一想** 怎樣可以不計算 $\mathrm{E}(WL)$ 而求出 $\mathrm{Cov}(W,L)$？

把恆等式倒過來用。取 $X=W$ 及 $Y=L$，

$$\frac{2m}{9}=\frac{2m}{9}+\frac{2m}{9}+2\,\mathrm{Cov}(W,L),\qquad \text{所以}\qquad \mathrm{Cov}(W,L)=-\frac{m}{9}.$$

因此，當 $m\geq 1$ 時（這樣方差才不是零），

$$\frac{\mathrm{Cov}(W,L)}{\sqrt{\mathrm{Var}(W)\,\mathrm{Var}(L)}}=\frac{-m/9}{2m/9}=-\frac{1}{2}.$$

答案與 m 無關。它是負數，這亦合乎情理：局數既然固定，勝的局數越多，留給負局的空間便越少。

第 (i) 部分提供了另一條途徑。把恆等式應用兩次，可知 $\mathrm{Var}(W+L+D)$ 等於三個方差之和，再加上三個兩兩協方差之和的兩倍。根據對稱性，三個協方差全部相等，設為 c，於是 $0=3\cdot\frac{2m}{9}+6c$，再次得出 $c=-\frac{m}{9}$。

驗算

- 取 $m = 1$。這時 W 和 L 都是指示變量（indicator），$\mathrm{E}(W) = \mathrm{E}(L) = \frac{1}{3}$，而且恆有 $WL = 0$，所以 $\mathrm{Cov}(W, L) = 0 - \frac{1}{9} = -\frac{1}{9}$，與 $-m/9$ 吻合。
- 兩條途徑（經第 (ii) 部分和經第 (i) 部分）的結果一致。
- 對 $m = 1, \ldots, 7$ 逐一窮舉全部 3^m 個結果，每次都得出 $\mathrm{Var}(W) = 2m/9$、$\mathrm{Cov}(W, L) = -m/9$，而相關係數（correlation）都恰好是 $-\frac{1}{2}$。

常見錯誤

- 因為各局棋互相獨立，便以為 W 和 L 也互相獨立。其實不然：它們與 D 合起來的總數是固定的。
- 取 $\mathrm{Var}(W + L) = \mathrm{Var}(W) + \mathrm{Var}(L) = 4m/9$，這樣做漏掉了協方差一項，而且與二項分佈得出的答案 $2m/9$ 矛盾。

STEP 之外

(W, L, D) 服從*多項分佈*（multinomial distribution），即把二項分佈自然地推廣至每次試驗有多於兩種結果的情況。一般而言，若各種結果的概率為 $p_1, \ldots, p_k$，則計數 N_i 與 N_j 的協方差為 $-mp_ip_j$，相關係數為 $-\sqrt{p_ip_j/\big((1-p_i)(1-p_j)\big)}$，在本題中就是 $-\frac{1}{2}$。協方差矩陣（covariance matrix）及協方差的雙線性（bilinearity），可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門），以及 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論導論）。

Beyond STEP

STEP 是大多數中學生一生中應考過最艱深的數學考試，也是最出色的考試之一：題目簡短、陌生而深刻，當中最精彩的，其實是喬裝了的大學數學小品。

本書精選最有趣的 STEP 試題，由 1987 年首次開考之前的樣本試卷（specimen papers），到 2025 年的試卷，並以每一道題目作為一扇門。每道題目都會仔細解答，沿途提出優秀學生會問的問題，然後帶你穿過這扇門：這個想法從何而來，到了大學會演變成甚麼，以及接下來該讀些甚麼。

本書為正在準備 STEP、有天分又有志投身數學的學生而寫。他們想要的不止是模擬試卷：他們想開始像數學系本科生那樣思考，並一探前方有甚麼正等着自己。

- 共 13 章，涵蓋純數學（pure mathematics）、力學（mechanics）、概率（probability）與統計（statistics）
- 136 道 STEP 歷屆試題，每題均附引導式解答
- 每道題目背後的大學數學
- 附答案的延伸練習，以及延伸閱讀

2002 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。